



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

F. Canonaco, M. Crippa, J. G. Slowik, U. Baltensperger, and A. S. H. Prévôt

Paul Scherrer Institute, Laboratory of Atmospheric Chemistry, 5232 Villigen PSI, Switzerland

Correspondence to: A. S. H. Prévôt (andre.prevot@psi.ch)

Received: 24 April 2013 – Published in Atmos. Meas. Tech. Discuss.: 15 July 2013

Revised: 8 October 2013 – Accepted: 14 October 2013 – Published: 23 December 2013

Abstract. Source apportionment using the bilinear model through a multilinear engine (ME-2) was successfully applied to non-refractory organic aerosol (OA) mass spectra collected during the winter of 2011 and 2012 in Zurich, Switzerland using the aerosol chemical speciation monitor (ACSM). Five factors were identified: low-volatility oxygenated OA (LV-OOA), semivolatile oxygenated OA (SV-OOA), hydrocarbon-like OA (HOA), cooking OA (COA) and biomass burning OA (BBOA). A graphical user interface SoFi (Source Finder) was developed at PSI in order to facilitate the testing of different rotational techniques available within the ME-2 engine by providing a priori factor profiles for some or all of the expected factors. ME-2 was used to test the positive matrix factorization (PMF) model, the fully constrained chemical mass balance (CMB) model, and partially constrained models utilizing a values and pulling equations. Within the set of model solutions determined to be environmentally reasonable, BBOA and SV-OOA factor mass spectra and time series showed the greatest variability. This variability represents the uncertainty in the model solution and indicates that analysis of model rotations provides a useful approach for assessing the uncertainty of bilinear source apportionment models.

2002), ecosystems, and agricultural areas via acidification and eutrophication (Matson et al., 2002). Therefore, reliable source identification and quantification is essential for the development of effective political abatement strategies. Based on their formation processes, atmospheric aerosols can be roughly separated into primary and secondary aerosols, i.e. those directly emitted and those formed from gas-phase reactions of emitted precursor gases, respectively. However, the details of aerosol formation processes are still poorly understood; in particular the submicron organic fraction of particulate matter (PM₁) (Hallquist et al., 2009), which comprises 20–90 % of the total submicron aerosol mass depending on the measurement location (Jimenez et al., 2009), is poorly characterized.

The Aerodyne aerosol mass spectrometer (AMS) provides online quantitative mass spectra of the non-refractory (inorganic and organic) components of the submicron aerosol fraction with high time resolution, i.e., seconds to minutes (Canagaratna et al., 2007). Through knowledge of the typical mass spectral fragmentation patterns, these spectra can be assigned to several inorganic components and to the organic fraction (Allan et al., 2004). However, interpretation of the organic fraction is challenging due to the enormous number of possible compounds. Over the past years, numerous ambient studies have successfully exploited the positive matrix factorization (PMF) algorithm, apportioning the measured organic mass spectra in terms of source/process-related components (see Zhang et al., 2011 for a review). The statistical tool PMF (Paatero and Tapper, 1994; Paatero, 1997) within the bilinear algorithm represents the time series of measured organic mass spectra as a linear combination of static factor profiles (i.e. mass spectra) and their respective time series. However, if all measured variables (i.e., the mass to charge

1 Introduction

Atmospheric aerosols are of scientific and political interest due to their highly uncertain direct and indirect effects on the solar radiation balance of the Earth's atmosphere (IPCC, 2007). Moreover, aerosols have a strong negative effect on human health (Peng et al., 2005), visibility (Watson,

ratios (m/z) for the mass spectrometer exhibit temporal co-variation, e.g., due to meteorological events such as rainfall or boundary layer evolution, or if the model solution has high rotational ambiguity, then the apportionment with PMF can yield non-meaningful or mixed factors. Under such conditions, the bilinear model can be directed towards an optimal solution by utilizing a priori information in the form of the factor profiles and/or time series. If all factor profiles are predetermined, the approach is a so called chemical mass balance (CMB). At the other extreme, the factor profiles in PMF are calculated entirely by the algorithm. The multilinear engine algorithm (ME-2) is capable of solving both these extremes and all intermediate cases in accordance with the constraints provided by the user (Paatero, 1999; Paatero and Hopke, 2009). Several PM source apportionment studies in which PMF did not properly represent the measured data have utilized ME-2 to find acceptable solutions (e.g., Lanz et al., 2008; Amato and Hopke, 2012; Reche et al., 2012). However, such studies are scarce, possibly due to the need for manual configuration and analysis of the results of the powerful ME-2 package. Therefore, in order to facilitate the choice of the initial conditions for the ME-2 engine and the analysis of the results, we have written the graphical user interface SoFi (Source Finder) within the software package IGOR Pro (Wavemetrics, Inc., Portland, OR, USA). SoFi will be freely distributed to all interested ME-2/PMF users.

In this study, the ME-2 engine was successfully applied to organic mass spectra obtained with the recently developed aerosol chemical speciation monitor (ACSM) (Ng et al., 2011b), an instrument based on AMS technology and optimized for long-term sampling. The ACSM was deployed in downtown Zurich, Switzerland, from January 2011 to February 2012.

2 Materials and methods

2.1 Measurements

From January 2011 to February 2012, an ACSM (Aerodyne Research, Inc., Billerica, MA, USA) was deployed at Kaserne, Zurich (Switzerland), an urban background station in the center of a metropolitan area with about one million inhabitants. The ACSM is a compact, low-maintenance aerosol mass spectrometer designed for long-term measurements of non-refractory particulate matter with vacuum aerodynamic diameters smaller than $1\ \mu\text{m}$ (NR-PM1). The instrument is described in detail by Ng et al. (2011b), and the reader is referred to Jayne et al. (2000), Jimenez et al. (2003), Allan et al. (2003a,b, 2004) and Canagaratna et al. (2007) for a detailed description of the AMS technique.

At the Kaserne station in Zurich, ambient aerosol briefly entered the temperature-controlled room and was subsequently drawn to a cyclone (model SCC 1.829 cyclon from BGI, INC.) with a size cut-off of $2.5\ \mu\text{m}$, using a flow of

$5\ \text{L}\ \text{min}^{-1}$ for removing coarse mode particles. The resulting aerosol flow passed through a Nafion drier (MD-110-48S-4, PermaPure LLC, Toms River, NJ, USA) and a subsequent $\sim 2\ \text{m}$ long stainless steel sampling tube (6 mm OD) before reaching the ACSM inlet. In the ACSM, the dried aerosol particles are sampled continuously (with an averaging time of 30 min) through a $100\ \mu\text{m}$ aperture ($\sim 90\ \text{cm}^3\ \text{min}^{-1}$), and are then passed through an aerodynamic lens (~ 2 torr) where they are focused into a narrow beam. The particle beam impacts a surface that is resistively heated at $\sim 600\ ^\circ\text{C}$. Here the non-refractory fraction is flash vaporized. The resulting gas is ionized by hard electron impact (70 eV) and analyzed with a quadrupole mass spectrometer. The final aerosol signal is retrieved by subtracting the background signal from filtered air under the same sampling conditions.

To obtain quantitative mass concentrations for the ACSM, a collection efficiency parameter (CE) needs to be applied to account for the incomplete detection of the aerosol species (Middlebrook et al., 2012). The CE is a function of the lens system of the shape and bouncing of the aerosol particles on the vaporizer. Specially the bouncing effect of the aerosol particles was found to be influenced by several parameters, such as the mass fraction of ammonium nitrate, particle acidity, and water content (Matthew et al., 2008). Water content does not affect the present study because the particles are dried. The effects of the nitrate mass fraction and particle acidity on CE have recently been parameterized for ambient data (Middlebrook et al., 2012). However, for the present study, this parameterization underestimates the CE, as demonstrated by higher CE-corrected mass concentrations for the ACSM compared to simultaneous PM_{10} measurements by a tapered element oscillating microbalance (TEOM, FDMS 8500, Thermo Scientific). The CE will be investigated in detail in a future publication; here we assume $\text{CE} = 1$, which provides a lower limit for ACSM-measured mass concentrations. Note that since the CE is applied to all measured species, changes in the CE do not affect the relative intensity of m/z within a mass spectrum and hence also do not affect the ME-2 results reported in this manuscript.

The meteorological parameters and trace gases were measured with conventional instruments by the Swiss National Air Pollution Monitoring Network, NABEL (Empa, 2011). The time resolution of all these measurements was ten minutes. NO_x was measured by chemiluminescence. The technique involves a molybdenum converter that suffers from artifacts due to partial conversion of NO_x oxidation products (Steinbacher et al., 2007). These artifacts, however, are only expected to be important at low concentrations, for example during summer. Carbon monoxide was monitored by non-dispersive Fourier transform infrared spectroscopy (APNA 360, Horiba, Kyoto, Japan), UV absorption spectroscopy was employed to determine the temporal variation of ozone (Thermo Environmental Instruments (TEI) 49C, Thermo Electron Corp., Waltham, MA) and black carbon was estimated utilizing an aethalometer, AE 31 (Magee

Scientific Inc.), based on measured light absorption coefficients at different wavelengths. In addition, the measured absorption coefficients at wavelengths 470 and 880 nm were used in order to retrieve the related black carbon contributions from the traffic (BC_{traffic}) and wood burning (BC_{wb}) source (Sandradewi et al., 2008; Herich et al., 2011).

2.2 The multilinear engine (ME-2)

The organic mass spectra measured by the ACSM can be represented as a matrix \mathbf{X} where the columns j are the m/z 's and each row i represents a single mass spectrum. A frequently used method is to group the variables into distinct factors based on certain criteria. The simplest and most commonly used approach is to group the variables into two constant matrices, the so-called bilinear model, e.g., principal component analysis (PCA) (Wold et al., 1987) or positive matrix factorization (PMF) (Paatero and Tapper, 1994). The bilinear factor analytic model in matrix notation is defined as:

$$\mathbf{X} = \mathbf{GF} + \mathbf{E}, \quad (1)$$

where the measured matrix \mathbf{X} is approximated by the product of \mathbf{G} and \mathbf{F} and \mathbf{E} is the model residual. p is then defined as the number of factors of the chosen model solution, i.e., the number of columns of \mathbf{G} and at the same time the number of rows of \mathbf{F} . Each column j of the matrix \mathbf{G} represents the time series of a factor, and each row i of \mathbf{F} represents the profile (e.g., mass spectrum) of this factor. The differences between the bilinear models PCA and PMF are only due to the restrictions of the models. PCA imposes orthogonality of the factors, i.e., the scalar of two different rows of \mathbf{F} is zero and does not require non-negative entries. By contrast, PMF requires non-negative entries throughout \mathbf{G} and \mathbf{F} . This constraint makes the PMF algorithm particularly suitable when mass concentrations must always be positive (especially for chemometrics or environmental studies). In the literature there are two solvers, namely ME-2 and PMF2 that solve the PMF algorithm. The main difference is the enhanced control of rotations within the ME-2 solver, as described in the next subchapters.

However for both solvers, the entries in \mathbf{G} and \mathbf{F} are fit using a least squares algorithm that iteratively minimizes the quantity Q^m , defined as the sum of the squared residuals weighted by their respective uncertainties, where the uncertainty may contain the measurement and model uncertainty:

$$Q^m = \sum_{i=1}^m \sum_{j=1}^n \left(\frac{e_{ij}}{\sigma_{ij}} \right)^2. \quad (2)$$

Here, e_{ij} are the elements of the residual matrix \mathbf{E} and σ_{ij} are the measurement uncertainties for the input points ij . Data points where $\sigma_{ij} \ll e_{ij}$ constitute a large fraction of Q^m , and these points will have a high impact during the model iteration. Normally this ensures that data with high signal-to-noise has a higher impact than measurements near

the detection limit. However, $\sigma_{ij} \ll e_{ij}$ may also occur due to dominant and rare local events or electronic noise within the measurement equipment, when neither of these events should be considered by the model. To prevent the solution from being driven by a few strong outliers, the model is generally run in the “robust” mode, in which pulling of the solution by outliers is reduced. At each step of the solution process, outliers are defined based on the ratio of residuals to uncertainties:

$$\text{outlier} = \left| \frac{e_{ij}}{\sigma_{ij}} \right| > \alpha, \quad (3)$$

where α is the user-defined threshold value. A value of 4 is recommended as a defining criterion for outliers within the robust mode (Paatero, 1997). The residuals are reweighted dynamically to reduce, and ideally to remove, the dependence of the rate of change of Q^m on the rate of change of the residuals of the outliers:

$$\frac{dQ^m}{dE_{\text{outliers}}} \cong 0. \quad (4)$$

2.2.1 Normalizing Q by the expected value of Q (Q_{exp})

Normally, monitoring the total Q is not meaningful because the expected value depends on the size of the data matrix and on the number of chosen factors. One therefore normalizes Q^m by the degree of freedom of the model solution (called Q_{exp}) which is both a function of the size of the data matrix and of the number of factors.

$$Q_{\text{exp}} \cong n \cdot m - p \cdot (m + n). \quad (5)$$

Ideally, if the model entirely captured the variability of the measured data and all uncertainties were properly defined, a Q/Q_{exp} value of 1 would be expected. However, several reasons e.g. an erroneously chosen number of factors, transient sources that are not fully modeled, errors in the estimate of the measurement uncertainties, and the unknown model uncertainties – prevent the use of the absolute Q/Q_{exp} . Instead, one should investigate the relative change of this ratio across different model runs (large changes indicate significantly decreased residuals and suggest an improved solution), to assist in choosing reasonable model solutions.

2.2.2 Rotational ambiguity of the model solutions

Solutions given by the PMF algorithm may have a substantial degree of rotational ambiguity (Paatero et al., 2002). There are two different kinds of rotations that are allowed, namely the pure and the approximate rotations. For pure rotations, the object function Q^m does not change after the rotation:

$$\bar{\mathbf{G}} = \mathbf{GT} \text{ and } \bar{\mathbf{F}} = \mathbf{T}^{-1}\mathbf{F}, \quad (6)$$

where \mathbf{T} is a nonsingular matrix of dimension $p \times p$, \mathbf{T}^{-1} is its inverse, and $\bar{\mathbf{G}}$ and $\bar{\mathbf{F}}$ are the rotated matrices. The matrix multiplication of $\bar{\mathbf{G}}$ and $\bar{\mathbf{F}}$ leads to the same product as

for \mathbf{G} and \mathbf{F} , and therefore Q^m remains unchanged. If the transformation matrix \mathbf{T} does not fulfill Eq. (6), the rotation is called an approximate rotation and Q^m changes. For the PMF and the ME-2 solver, there is a user-specific parameter called *fpeak*, denoted by φ for the global control of such approximate rotations. For positive φ , elementary rotations or a series of elementary rotations are performed that increase columns of the matrix \mathbf{G} and decrease rows of the matrix \mathbf{F} while conserving mass. The opposite occurs for negative φ . However, the *fpeak* tool explores only rotations in one dimension of the multidimensional space, and if the entries of \mathbf{G} and \mathbf{F} are positive and more than one factor is chosen, then the rotational space is multidimensional and the corresponding ambiguity can be very large (e.g., for three factors, the rotational space is nine-dimensional). An advantage of the ME-2 solver compared to the PMF solver is improved rotational control, e.g., selected factors can be summed/subtracted together, so the entire matrix does not have to be transformed. Thus, the rotations can be studied in a more controlled environment. Normally, the user should explore the solution space, both since it is rare to find the environmental solution for the unrotated case, and since it allows the user to evaluate the rotational ambiguity of the chosen solution in the solution space. Alternatively, to reduce the rotational ambiguity, a priori information in the form of known rows of \mathbf{F} (factor profiles) or of known columns of \mathbf{G} (factor time series) can be added to the model (Paatero and Hopke, 2009). This a priori information creates a very specific rotated model solution that can be further investigated. Three main approaches can be exploited with the ME-2 solver, i.e. the chemical mass balance (CMB), the *a* value, and the pulling technique (described below).

The use of a priori information at the stage of the calculation of the model solution provides a more efficient and sensitive exploration of the model space than is possible with e.g., the global *fpeak* tool (Paatero and Hopke, 2009). For this reason, we developed a user-friendly interface (Fig. S1 in the Supplement), SoFi (Source Finder), to facilitate the testing of the different rotational techniques available within the ME-2 model. Three different approaches were exploited, i.e., the chemical mass balance (CMB), the *a* value, and the pulling technique, using the bilinear model based on the criterion of positive entries in \mathbf{G} and \mathbf{F} . The application of these techniques is described in detail in Paatero and Hopke (2009) and only a brief description is presented here. In addition, this interface allows the user to run the PMF algorithm with/without the above-mentioned constraining techniques for combined data sets, e.g. particle and gas-phase data, in the robust mode. This technique was first tested using a pseudo robust approach by Slowik et al. (2010). Crippa et al. (2013a) exploited this interface to perform a combined source apportionment on ambient AMS and PTR-MS (proton transfer reaction mass spectrometer) data from the Paris 2009/2010 campaign entirely in the robust mode.

2.2.3 Fully unconstrained matrices \mathbf{G} and \mathbf{F} : positive matrix factorization (PMF)

For a completely unconstrained PMF run, the algorithm models the entries of \mathbf{G} and \mathbf{F} autonomously.

2.2.4 Fully constrained matrix \mathbf{F} : chemical mass balance (CMB)

Within the chemical mass balance, all elements of the \mathbf{F} matrix, i.e., all factor profiles, are set to non-negative values by the user. The entries of the matrix \mathbf{G} remain variable and are evaluated by the model.

2.2.5 Constrained matrices \mathbf{F}/\mathbf{G} : *a* value approach (*a* value)

Here the elements of the \mathbf{F} matrix (factor profiles) and/or of the \mathbf{G} matrix (factor time series) can be constrained by the user. The user inputs one or more factor profiles (rows of \mathbf{F})/factor time series (columns of \mathbf{G}) and a constraint defined by the scalar *a* that can be applied to the entire profile/time series or to individual elements of the profile/time series only. The *a* value determines the extent to which the output \mathbf{F}/\mathbf{G} is allowed to vary from the input \mathbf{F}/\mathbf{G} , according to:

$$f_{j,\text{solution}} = f_j \pm a \cdot f_j, \quad (7)$$

$$g_{i,\text{solution}} = g_i \pm a \cdot g_i, \quad (8)$$

where *f* and *g* represent a row and the column of the matrices \mathbf{F} and \mathbf{G} , respectively. The index *j* varies between 0 and the number of variables and *i* varies between 0 and the number of measured points in time.

The situation of the chemical mass balance described in Sect. 2.2.4 is achieved by using the scalar *a* set to 0 for all factor profiles.

2.2.6 Constrained matrices \mathbf{F}/\mathbf{G} : pulling approach (pulling)

The user has the possibility of introducing pulling equations into the model that pull profile factor elements towards pre-defined anchor values (here shown for a row of the matrix \mathbf{F} only):

$$\bar{a}_j = \bar{f}_j + r_j. \quad (9)$$

In Eq. (9), *a_j* represents the anchor to which the model pulls the iterative value \bar{f}_j , and *r_j* represents the residual. The anchor is a known value introduced as a priori information by the user. The pulling equations create an additional auxiliary term Q^{aux} that is added to Q^m . Thus, if pulling equations are introduced, the model will minimize the argument of *Q* with respect to all entries in the matrices \mathbf{G} and \mathbf{F} :

$$\arg \min_{G,F}(Q) = \arg \min_{G,F}(Q^m + Q^{\text{aux}}). \quad (10)$$

The term of Q^{aux} has a conceptually similar aspect to Q^{m} :

$$Q^{\text{aux}} = \sum_k \left(\frac{r_k}{s_k} \right)^2. \quad (11)$$

The index j has been replaced by k in Eq. (11), since k denotes the index of the pulling equations added to the model (over many factor profiles/time series). The pulling parameter s_k specifies the softness of the pull. The smaller s_k becomes, the higher the impact of Q^{aux} of the k th pull within the iterative process. The pulling approach is a sensitive technique in that, if the pulling equation is not compatible with the specific data matrix, the decrease of Q^{aux} (Eq. 11) obtained as \bar{f}_j reaches its anchor value a_j (Eq. 9) is negligible compared to a larger increase of Q^{m} (Eq. 2); consequentially, the pull falls off. Adding known factor profiles/time series and using the pulling technique might be seen as a “soft” and self-regulating constraining technique. Generally, the user provides the total acceptable limits of Q^{m} , denoted as dQ . Changing these limits and the pulling parameter s_k allows the user to monitor the change in Q^{m} and to judge its acceptability.

2.2.7 The correct solution and number of factors

Generally, increasing the number of factors decreases Q and the ratio of Q to Q_{exp} , due to the additional degrees of freedom of the model, allowing a better fit to the measured matrix. However, these additional factors may not be physically meaningful. As a first metric in judging the correct number of factors, Paatero and Tapper (1993) recommended considering the size of the decrease of Q or Q/Q_{exp} as a function of added factors, rather than its absolute value. Changes in Q or Q/Q_{exp} over different model runs of a few percentages are acceptable, if the model solution is enhanced. If the difference is however, of tens of percentages, further investigation is required.

In addition to the Q – analysis, Paatero (2004) introduced another metric based on the estimation of the measurement variation explained by the factors. The explained variation (EV) is a dimensionless quantity that indicates how much variation in time or variation in each variable is explained by each factor. As an example, the equation for the explained variation for the i th point in time for the factor k is given by:

$$\text{EV}_{ik} = \frac{\sum_{j=1}^m (|g_{ik} \cdot f_{kj}| / \sigma_{ij})}{\sum_{j=1}^m ((\sum_{h=1}^P |g_{ih} \cdot f_{hj}| + |e_{ij}|) / \sigma_{ij})} \quad \text{for } k = 1, \dots, P. \quad (12)$$

Similar equations can be formulated for the unexplained variation (UEV) by replacing the product $g_{ik} \cdot f_{kj}$ in the numerator with e_{ij} . Expressing the explained and the unexplained variation for a variable j as EV_{jk} is done by simply replacing the sum over j in the ratio with the sum over i . If all variation is explained by the model, then $\text{EV} = 1$. According

to Paatero (2004) a variable should be regarded as explained only if the UEV for that variable is less than 25 %.

Besides these mathematical instruments, it is crucial to compare the model output with measurements or reference values that were not included in the model solution. This aids in the selection and verification of the factor solutions.

In order to test the aforementioned rotational approaches, we used a data matrix containing the winter data from downtown Zurich measured with the ACSM from both 2011 and 2012. The measurement error matrix was calculated according to the method of Allan et al. (2003a, 2004), the m/z 44-related m/z 's and weak and bad m/z 's were downweighted and a minimum error was considered for the uncertainty of all points in the data matrix as in Ulbrich et al. (2009). In addition, the measurement uncertainty for points with a signal to noise (S/N) smaller than 2 (weak variables) and a S/N smaller than 0.2 (bad variables) were increased by a factor of 3 and 10, respectively, as in Ulbrich et al. (2009).

3 Results

3.1 Unconstrained matrices G and F (PMF)

The first step in the source apportionment analysis was to perform the bilinear model without any a priori information in the modeled matrices (PMF) for different numbers of factors, e.g., one to ten factors, to estimate an environmentally reasonable number of factors. PMF analysis of aerosol mass spectra has previously been described in detail (e.g., Lanz et al., 2007, 2010; Ulbrich et al., 2009), and similar metrics for determining the appropriate number of factors were employed in this study. Specifically, the solution was chosen based on an analysis of the dependence of Q/Q_{exp} and the explained variation in the number of factors, as well as the correlation of the retrieved factor profiles and time series with reference spectra and collocated measurements. A five-factor solution was selected for further analysis. This solution is summarized below and additional details are provided in the Supplement (Sect. S6.2). PMF solutions with a higher number of factors are not considered, due to purely mathematical splits of the factor profiles leaving the EV almost untouched and hence not representing additional sources.

The five-factor solution consists of three primary factors and two secondary factors. The primary factors are hydrocarbon-like organic aerosol (HOA), cooking organic aerosol (COA) and biomass burning organic aerosol (BBOA), while the secondary factors are semi-volatile oxygenated organic aerosol (SV-OOA) and low-volatility oxygenated organic aerosol (LV-OOA). These factors have been identified in many previous studies and only a brief description of their most important characteristics is given here. Factor mass spectra are shown in Fig. S2, the time series are shown in Fig. S3 and the diurnal patterns are shown in Fig. S4. The HOA spectrum shows high signal at m/z

Table 1. Overview of the different model runs discussed in this study. The empty space indicates unconstrained information. The parameters listed in the table indicate the strength of the constraints for the corresponding model run.

Model run	Constraints	
	Secondary species	Primary species
PMF		
CMB	a value = 0	a value = 0
a value		a value (0.1, 0.2, 0.3)
pulling		$dQ = 100$, s (0.01, 0.015, 0.02, 0.05)

typical of aliphatic hydrocarbons (Canagaratna et al., 2004; Zhang et al., 2005). The time series and diurnal pattern of HOA are correlated with traffic-related species like NO_x , CO, and $\text{BC}_{\text{traffic}}$. The COA profile is qualitatively similar to HOA but has higher m/z 55 and less m/z 57, similar to previous results (Allan et al., 2010; He et al., 2010; Slowik et al., 2010; Sun et al., 2011; Mohr et al., 2012; Crippa et al., 2013b) and its diurnal cycle shows the characteristic lunch peak at noon. The BBOA profile has significantly higher contributions at m/z 60 and m/z 73. These fragments are characteristic of sugars such as levoglucosan (Alfarra et al., 2007) which are released during wood combustion processes. The BBOA diurnal pattern has higher contributions at night, consistent with domestic heating activities in winter. SV-OOA and LV-OOA have significantly higher contributions at m/z 44, which is typically dominated by the CO_2^+ ion. This ion results from the thermal decomposition and fragmentation of highly oxygenated species such as organic acids (Ng et al., 2010). Compared with SV-OOA, LV-OOA typically has a higher mass fraction at m/z 44, suggesting a more aged and less volatile aerosol. Their time series correlate with the time series of secondary species like sulfate, nitrate and ammonium aerosol.

Note that while features of the factors described above can be identified from the PMF analysis, there is no unequivocal apportionment of each factor to one specific source. Hence, the labeled factors in Sect. S6.2 in the Supplement are only indicative. For example, the characteristic COA peak at noon is visible but rather broad between 08:00 and 12:00 LT. The primary factors HOA and COA both contain signal from m/z 44 and m/z 60, suggesting that some biomass burning aerosol may be apportioned to these factors. These features reveal a mixed situation for the PMF factor solution. In order to retrieve an environmentally satisfactory model solution, further investigation of the multidimensional solution space is needed. One possible method is to make use of the global rotational parameter f_{peak} . Nonetheless, the outcome might not always be satisfactory, as was the case for this study. The ME-2 solver provides three alternative options for the exploration of the solution space:

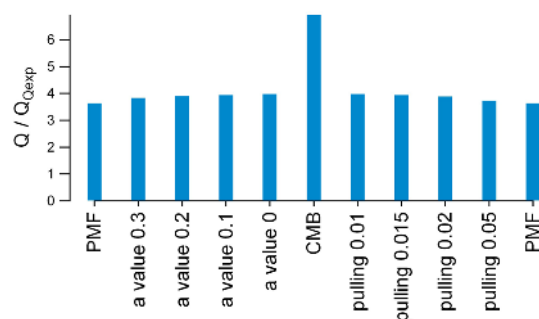


Fig. 1. Values of Q/Q_{exp} for different model runs. The CMB run, for which all factor profiles have been fixed, is almost double the other values.

1. application of user-specific rotations to search for solutions that better describe the measured data matrix
2. addition of specific pulling equations on e.g. retrieved factor profiles and/or time series from earlier unconstrained PMF solutions
3. utilization of a priori information, thus strongly reducing the rotational ambiguity.

This study investigates only the third approach, although the user-specific rotations and specific pulling equations are potentially valuable techniques and should be further investigated in future source apportionment studies.

3.2 Comparison of solutions constraining matrix \mathbf{F}

3.2.1 Overview

Besides the PMF run using unconstrained matrices \mathbf{G} and \mathbf{F} described in the last section, the subsequent model runs constraining the matrix \mathbf{F} , or parts of it, have been tested. The following runs are summarized in Table 1.

- CMB, with all five factors fixed (see Sect. 2.2.4).
- a value approach (see Sect. 2.2.5), where the primary factors HOA, COA and BBOA were constrained and the other factors left free. Different a values were tested, i.e., with a value 0 to 0.3 applied simultaneously to all constrained profiles. Note that an a value of 0 yields a “partial CMB” model where the primary factors are fully constrained and the secondary factors are fully free.
- pulling approach (see Sect. 2.2.6), where the primary factors HOA, COA and BBOA were constrained and the other factors left free. The parameters tested were $dQ = 100$ and softness (s) between 0.01 and 0.05. Since dQ stayed invariant, the only value reported for the pulling runs in the following graphs is the softness (s).

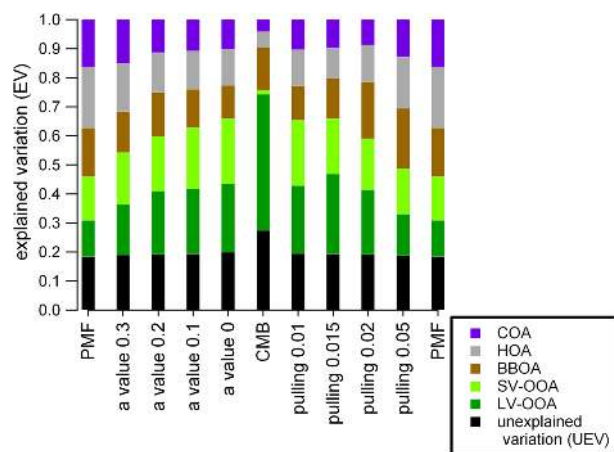


Fig. 2. Explained variation (EV) for each factor and total unexplained variation (UEV) during the different model runs.

The primary factors (HOA, COA, BBOA) employed have been taken from Crippa et al. (2013b), an unconstrained PMF analysis in which the primary sources have successfully been separated, and the secondary factors (SV-OOA, LV-OOA) were the averaged mass spectra reported by Ng et al. (2011a).

Figure 1 shows Q/Q_{exp} for the mentioned runs. This graph and the successive ones are structured such that model runs with weaker boundaries (i.e., with larger a values or softer pulling parameters) are on the outside, while runs with stronger constraints are inside. Therefore, PMF represents the outer edge and CMB is in the center. Note that for the a value approach, the value in the graph indicates the lower and upper limit. For the pulling runs, the value reported stands for the softness s of the pull (dQ is constant at 100).

The CMB result has the worst compatibility with the measured data matrix, as shown by the higher Q/Q_{exp} ratio. This is also reflected in the plot of the explained variation (EV) (Fig. 2), where the CMB run shows the highest unexplained variation (UEV). In general there is a considerable change in the distribution of EV between the different model configurations. In particular, the EV for the secondary species SV-OOA and LV-OOA varies significantly. The EV for the primary species COA, HOA and BBOA stays approximately constant for the a values runs between 0 and 0.2 and the pulling runs between 0.01 and 0.02.

The mean mass concentration of all factors as a function of all model runs is shown in Fig. 3. The black rectangles in Fig. 3 (and Fig. 4) denote environmentally reasonable solutions, as discussed later. The figure shows that the CMB approach lacks in representing the measured data, due to the very decreased explained mass compared to the other models, as already mentioned for Figs. 1 and 2. In addition, the continuous redistribution of the mass contributions to the five factors as the tightness of constraint changes is also apparent.

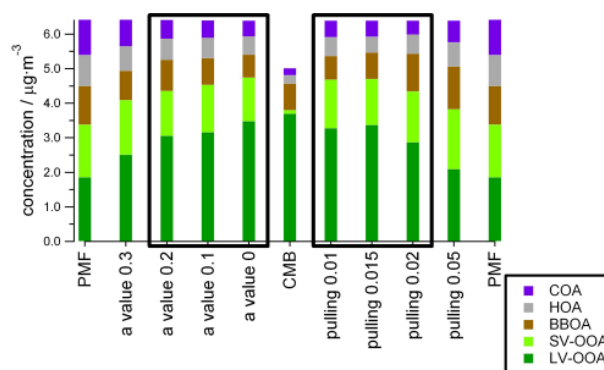


Fig. 3. Mean mass concentration for the five factors for the model runs. The results reported within the two rectangles represent environmentally reasonable solutions.

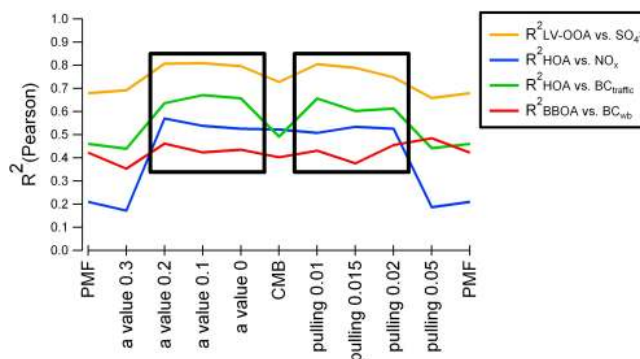


Fig. 4. Correlations R^2 (Pearson) between the time series of selected factors and the time series of external data as a function of the model runs. The results reported within the two rectangles represent environmentally reasonable solutions.

As mentioned in Sect. 2.2.7, an important criterion – along with the Q/Q_{exp} and the explained variation – for judging acceptable source apportionment solutions is the comparison with external information. Figure 4 lists R^2 (Pearson) for the correlations between the time series of HOA with the traffic species NO_x and $BC_{traffic}$ as well as between BBOA and $BC_{wood\ burning}$ and between LV-OOA and NR-PM1 sulfate as a function of the different model solutions. Acceptable correlation values fall within the black rectangle. Note that, although R^2 (Pearson) for BBOA with BC_{wb} is highest for the pulling model run with $s = 0.05$, this run is still rejected due to the other degraded correlations, in particular that for the traffic factor HOA. Further support for identifying the model solutions within the rectangles as environmentally reasonable is provided by the analysis of the diurnal cycle of HOA and COA (Sect. 3.2.3 and in the Sect. S6.4 in the Supplement), where the expected diurnal patterns for the traffic and cooking factors can be found. The absolute mean and relative mass concentrations for all selected solutions are shown in Table 2. The high standard deviation for BBOA indicates that the apportionment of this species is more uncertain, while COA and HOA show very little variation.

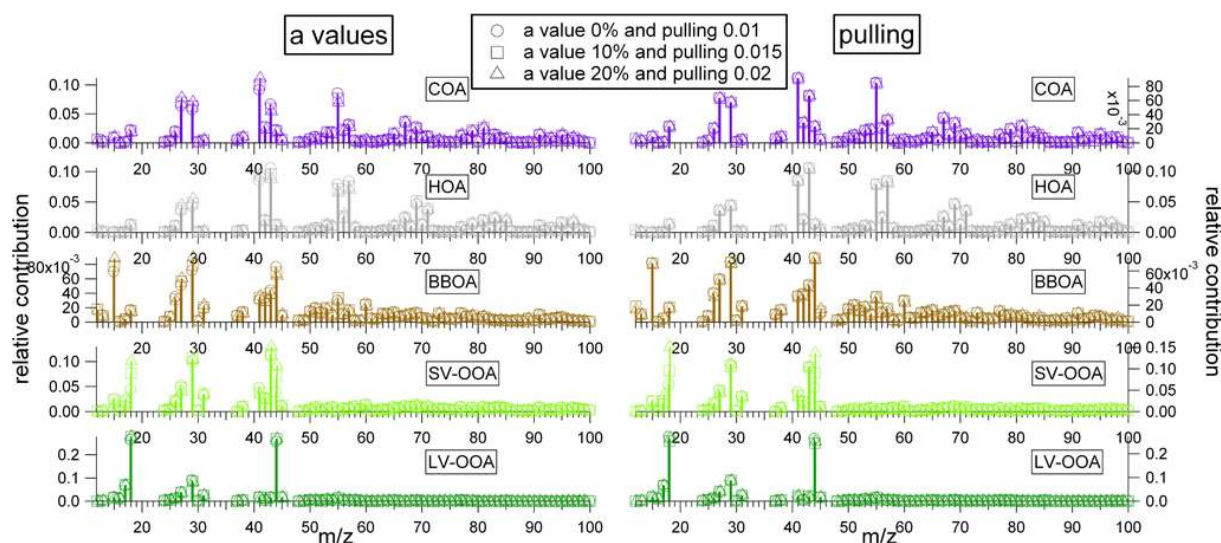


Fig. 5. Factor profiles of the five factor solutions HOA, COA, BBOA, SV-OOA and LV-OOA as a function of a value (left panel) and pulling strength (s) for solutions classified as environmentally reasonable. Different solutions are represented by different symbols, with circles and triangles being the most and least constrained, respectively.

Table 2. Absolute and relative mean factor mass concentrations averaged over all environmentally reasonable model runs.

Factor	mean \pm 1 std. dev./ $\mu\text{g m}^{-3}$	mean \pm 1 std. dev./ %	ratio of std. dev. to total/%
LV-OOA	3.20 ± 0.22	50.0 ± 3.5	7.0
SV-OOA	1.36 ± 0.08	21.2 ± 1.2	5.7
BBOA	0.82 ± 0.16	12.8 ± 2.5	19.5
COA	0.48 ± 0.05	7.5 ± 0.08	1.1
HOA	0.54 ± 0.05	8.5 ± 0.08	0.9

3.2.2 Comparison of factor profiles

Figure 5 shows the factor profiles of all environmentally reasonable model solutions. Models based on a values and pulling equations are shown in the left and right column, respectively. Different constraint levels are shown by different symbols. As noted in the previous section, the selected solutions lie in a relatively small range of a values (0–0.2) and strong pulling strengths (0.01–0.02).

As seen in Fig. 5, there is no significant variation of the primary factor profiles HOA, COA and BBOA as a function of the different model runs, due to the imposed tight constraint. By contrast, the unconstrained factors, especially SV-OOA, show more model-dependent variation. In particular, the high variation of m/z 43 in SV-OOA highlights the high uncertainty in apportioning this variable. Figure 3 highlights the fact that moving from a constrained run to a less constrained situation, apports less mass to LV-OOA and more in SV-OOA as well as to the three primary factors HOA, COA and, in particular, BBOA. This is evidenced in the factor profile with the increase of m/z 44 in SV-OOA for less constrained model runs.

3.2.3 Comparison of factor time series

Diurnal cycles for the environmentally reasonable model solutions are shown in Fig. 6. In addition, NO_x and $\text{BC}_{\text{traffic}}$ are plotted together with HOA, while BC_{wb} is plotted with BBOA.

The diurnal trends of HOA, NO_x , and $\text{BC}_{\text{traffic}}$ are highly correlated. The diurnal cycle of the cooking factor COA manifests a strong peak during meal activities, similar to COA factors found in other source apportionment studies conducted on NR-PM1 data in other cities such as Barcelona (Mohr et al., 2012) and Paris (Crippa et al., 2013b). In addition, the fact that the diurnal cycle over the weekends manifests only a small bump at noon for the cooking factor, while the morning traffic peak has totally disappeared, reinforces the interpretation of a successful separation of the two sources, HOA and COA (Sect. S6.4.3). BBOA is correlated with $\text{BC}_{\text{wood burning}}$ and is highest at night, consistent with domestic heating activities and previous measurements in Zurich (Lanz et al., 2008). The diurnal cycle of SV-OOA is anticorrelated with temperature, suggesting that the factor represents semivolatile material which is influenced by temperature-driven partitioning; the diurnal cycle of LV-OOA, by contrast, does not show strong diurnal trends.

3.3 Comparison with a previous source apportionment study

In the past, the ME-2 solver has already been used to constrain a hydrocarbon like factor HOA for a few weeks in winter in Zurich (Lanz et al., 2008). In that study, the a value leading to an acceptable solution was at 60 %. The lower and higher limits themselves depend on the constrained factor

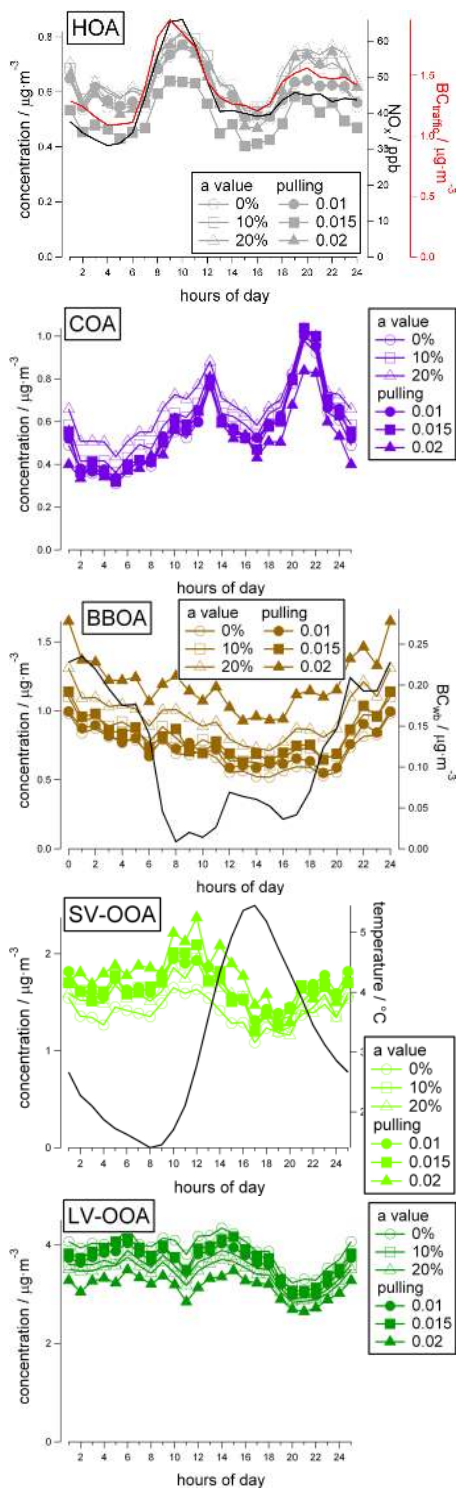


Fig. 6. Mean hourly factor mass concentrations for solutions classified as environmentally reasonable. Open and closed symbols denote a value and pulling solutions, respectively. Symbol shapes indicate the level of constraint, with circles being the most constrained and triangles the least. NO_x , $\text{BC}_{\text{traffic}}$ and $\text{BC}_{\text{wood burning}}$ are shown for comparison.

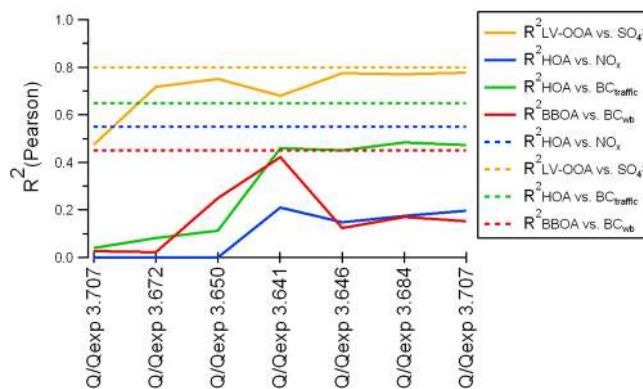


Fig. 7. Correlations R^2 (Pearson) between the time series of selected factors and the time series of external data as a function of the global f peak tool used for the exploration of the solution based on the unconstrained PMF run. The reported Q/Q_{exp} values are within the range of the constrained runs. The value in the middle represents the unrotated case ($f_{\text{peak}} = 0$).

profile. The profile used in our study has a substantial contribution (1.4 %) of m/z 44, the same variable which occurs in almost all factor profiles. Therefore, a large a value would easily lead to a mixing situation which is avoided by using only smaller a values. This was not the case for the constrained factor deployed in Lanz et al. (2008) where no m/z 44 was present at all (0%). In that study, the source apportionment over three weeks led to three factors, HOA, BBOA and OOA. The diurnal cycle of their HOA showed a lunch peak, revealing a possible contribution of the COA factor. This conclusion is reinforced by the fact that the mass contribution for their HOA was between 3 and 13 %. Our contribution when merging COA and HOA together varies between 7 and 18 % (Fig. 5). In addition, they could not separate OOA into SV-OOA and LV-OOA, most probably due to the small temperature range (-8 to 8°C) during those days. By contrast, the temperature range for this study (-13.5 to 18.1°C) was sufficient to allow a separation of SV-OOA and LV-OOA within the source apportionment.

4 Discussion

4.1 Uncertainty in bilinear model results

As discussed in the previous section, the bilinear PMF algorithm containing constraints in \mathbf{F} solved by the ME-2 engine yields a set of environmentally reasonable solutions which definitely improve the source apportionment. Note that the fully unconstrained PMF run did not even fall into the range of environmentally acceptable solutions. While the constrained ME-2 solutions have many features in common, the reported profiles and mass concentrations differ (see Table 2, Figs. 5 and 6). This variation reflects the model uncertainty for the bilinear system. Rotational techniques, such as the a value approach, pulling equations or individual

rotations as well as the frequently used global rotational tool fpeak are tools for the quantitative assessment of the bilinear model uncertainty.

An additional source of uncertainty in the model results, derives from the selection of the anchoring factor profiles and the magnitude of their constraints. The effects of the latter are evident from the CMB result (see Sect. 3.2.1). In particular, the mass contribution of the SV-OOA factor was almost negligible. This occurred because the profiles were completely fixed, leaving the model with no chance to adapt the SV-OOA profile to the measured data, and resulting in the highest unexplained variation (UEV) as shown in Fig. 2. However, a semi-volatile fraction has been modeled for the a value, pulling, and even the PMF approaches. This underlines the fact that if an anchor profile is too tight, a legitimate factor can be excluded from the model result.

The effect of choosing various factor profiles is highlighted by the comparison with the results of Lanz et al. (2008), in which winter AMS data from Zurich was analyzed using a constrained HOA factor profile with an a value of 60 %. This value is considerably higher than the maximum a value of 20 % selected in this study. The difference in the required a value for these two studies is likely due to the choice of HOA factor profile itself. In the present study, the employed HOA anchor profile has a non-zero contribution in m/z 44 (1.4 %). Most probably, a large a value would lead to a mixing situation based on the variable m/z 44, which we avoided in this study by using only smaller a values. This was not the case for the constrained factor used in Lanz et al. (2008), as no m/z 44 was present at all (0 %).

However, testing the influence of different anchoring profiles and the tightness of their constraints, before a solution fails to be environmentally interpretable, is ongoing and will be methodically discussed in a future study.

4.2 Recommendations for ME-2 analysis of aerosol mass spectra

Currently, there is very little research regarding the inclusion of a priori information in the source apportionment for aerosol mass spectrometer data. The scope of this work is to facilitate the source apportionment in this respect by testing, in a semi-automatic way, different rotational tools of the ME-2 solver with the user-friendly graphical user interface SoFi.

Generally, the user can anchor factor information (profiles or time series) and easily vary the tightness of the constraint while monitoring the various criteria for the evaluation of a solution (see Sect. 2.2.7). Based on the experience gained in this study, we recommend that one constrains the primary factors (HOA, COA, BBOA) for NR-PM1 source apportionments whenever the unconstrained PMF run reveals indications for such sources in the model result and/or in the corresponding residuals. In first runs, the secondary species can stay unconstrained, since they do not represent specific emis-

sions, but rather span the range of aging processes in a specific location during the measurement time. Thus, it is difficult to match this evolution with auxiliary data. However, this topic is under study and more information will be provided in future studies.

The user should in any case perform sensitivity tests on the tightness of the constrained factor profiles (a value or pulling parameters), to assess the environmentally reasonable solutions, and present this range rather than only a single solution. As stated in Sect. 4.1, these parameters highly depend on the anchoring profiles employed, and thus no limits for these values can be suggested at this stage. However, in general the increase of Q/Q_{exp} should not be larger than a few percentages as already stated in Paatero and Tapper (1993).

Crippa et al. (2013c) performed the source apportionment with the ME-2 engine on the AMS EUCAARI (European Integrated project on Aerosol, Cloud, Climate, and Air Quality Interactions) data measured in 2008/2009, with the aim of possibly standardizing the method of source apportionment on NR-PM1 data with ME-2.

4.3 Comparison between the PMF2 and the ME-2 solver

As discussed in Sect. 2.2.2, a fundamental difference between the PMF2 and ME-2 solvers their ability to explore the solution space. PMF2 utilizes the global fpeak tool, which allows rotations in only a single dimension of the multidimensional space. The limitation of this tool is summarized in Fig. 7. The R^2 values of LV-OOA, BBOA and HOA with their external tracers as a function of Q/Q_{exp} as the fpeak is varied, are reported in Fig. 7. The fpeak range was chosen as such to allow for an increase of Q/Q_{exp} similar to that of the constrained approach using ME-2. The dashed lines in this graph represent the mean R^2 values of the modeled ME-2 solutions in the boxed regions in Fig. 4. Figure 7 shows that the ME-2 solutions equal or outperform the best available PMF2 solutions for all factors. A particularly large improvement is evident for HOA, most likely due to the improved separation of HOA and COA by constraining these factors. Thus, there is no guarantee that the environmentally optimal solution can be retrieved through the PMF2 solver; it may be an inaccessible rotation in the PMF solution space. This limitation does not exist for the ME-2 solver, where all rotations are accessible.

The ME-2 solver provides the user with a tool for easily entering a priori information in form of, e.g. known factor profiles, similar to the tests conducted in this study with the primary factor profiles (HOA, BBOA and COA). Furthermore, it enables the full exploration of the solution space exploiting the individual fpeak tool or the pulling equations. We are actually systematically testing the exploration of the solution space based on the pulling equation for AMS data, and its advantages for AMS source apportionment studies will be presented in a future work (Canonaco et al., 2014).

5 Conclusions

Source apportionment using the bilinear model as implemented through the multilinear engine (ME-2) was successfully applied to non-refractory organic aerosol (OA) mass spectra measured during winter 2011 and 2012 in Zurich, Switzerland using the aerosol chemical speciation monitor (ACSM). The solutions have been analyzed exploiting the newly developed software source finder (SoFi). The selected solutions consist of two secondary factors and three primary factors. The secondary factors are a semi-volatile oxidized OA (SV-OOA) and a low-volatility oxidized OA (LV-OOA). The three primary factors are traffic-related hydrocarbon-like OA (HOA), cooking OA (COA) and biomass burning OA (BBOA).

Different rotational approaches were investigated employing the ME-2 engine. The tested implementations consisted of the unconstrained run (PMF), fully constrained chemical mass balance (CMB), and partially constrained models using the *a* value parameter and pulling equations. In addition, a sensitivity test on the constrained profiles was performed for the *a* value and pulling model runs. This allowed us to identify the set of environmentally reasonable solutions.

Moreover, such analysis provides insight into the uncertainty of the bilinear model solution, e.g., the primary factor BBOA and the secondary semivolatile factor SV-OOA show the highest variability across models (implying the highest model uncertainty), while COA and HOA have the least variability (smallest model uncertainty).

Finally, some recommendations for future NR-PM1 source apportionments exploiting ME-2 are reported.

Supplementary material related to this article is available online at <http://www.atmos-meas-tech.net/6/3649/2013/amt-6-3649-2013-supplement.pdf>.

Acknowledgements. The ACSM measurements were supported by the Swiss Federal Office for the Environment (FOEN). J. Slowik acknowledges support from the Swiss National Science Foundation (SNF) through the “Ambizione” program. We would like to thank P. Paatero for the valuable comments during the analysis of the data and the Environmental group of the Swiss Federal Laboratories for Materials and Testing (EMPA) for their support. Special thanks are also owed to Michele Canonaco for a critical reading of this manuscript.

Edited by: J. Abbatt

References

- Alfarra, M. R., Prévôt, A. S. H., Szidat, S., Sandradewi, J., Weimer, S., Lanz, V. A., Schreiber, D., Mohr, M., and Baltensperger, U.: Identification of the mass spectral signature of organic aerosols from wood burning emissions, *Environ. Sci. Technol.*, 41, 5770–5777, 2007.
- Allan, J. D., Alfarra, M. R., Bower, K. N., Williams, P. I., Gallagher, M. W., Jimenez, J. L., McDonald, A. G., Nemitz, E., Canagaratna, M. R., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer: 2. Measurements of fine particulate chemical composition in two UK cities, *J. Geophys. Res.-Atmos.*, 108, 4091, doi:10.1029/2002JD002359s, 2003a.
- Allan, J. D., Jimenez, J. L., Williams, P. I., Alfarra, M. R., Bower, K. N., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer: 1. Techniques of data interpretation and error analysis, *J. Geophys. Res.-Atmos.*, 108, 4090, doi:10.1029/2002JD002358s, 2003b.
- Allan, J. D., Delia, A. E., Coe, H., Bower, K. N., Alfarra, M. R., Jimenez, J. L., Middlebrook, A. M., Drewnick, F., Onasch, T. B., Canagaratna, M. R., Jayne, J. T., and Worsnop, D. R.: A generalised method for the extraction of chemically resolved mass spectra from aerodyne aerosol mass spectrometer data, *J. Aerosol Sci.*, 35, 909–922, 2004.
- Allan, J. D., Williams, P. I., Morgan, W. T., Martin, C. L., Flynn, M. J., Lee, J., Nemitz, E., Phillips, G. J., Gallagher, M. W., and Coe, H.: Contributions from transport, solid fuel burning and cooking to primary organic aerosols in two UK cities, *Atmos. Chem. Phys.*, 10, 647–668, doi:10.5194/acp-10-647-2010, 2010.
- Amato, F. and Hopke, P. K.: Source apportionment of the ambient PM_{2.5} across St. Louis using constrained positive matrix factorization, *Atmos. Environ.*, 46, 329–337, 2012.
- Canagaratna, M. R., Jayne, J. T., Ghertner, D. A., Herndon, S., Shi, Q., Jimenez, J. L., Silva, P. J., Williams, P., Lanni, T., Drewnick, F., Demerjian, K. L., Kolb, C. E., and Worsnop, D. R.: Chase studies of particulate emissions from in-use New York City vehicles, *Aerosol Sci. Technol.*, 38, 555–573, 2004.
- 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., and Worsnop, D. R.: Chemical and microphysical characterization of ambient aerosols with the Aerodyne aerosol mass spectrometer, *Mass Spectrom. Rev.*, 26, 185–222, doi:10.1002/mas.20115s, 2007.
- Canonaco, F., Crippa, M., Slowik, J., Baltensperger, U., and Prévôt, A. S. H.: Development of the pulling technique for AMS data based on the ME-2 solver using SoFi, in preparation, 2014.
- Crippa, M., Canonaco, F., Slowik, J. G., El Haddad, I., DeCarlo, P. F., Mohr, C., Heringa, M. F., Chirico, R., Marchand, N., Temime-Roussel, B., Abidi, E., Poulain, L., Wiedensohler, A., Baltensperger, U., and Prévôt, A. S. H.: Primary and secondary organic aerosol origin by combined gas-particle phase source apportionment, *Atmos. Chem. Phys.*, 13, 8411–8426, doi:10.5194/acp-13-8411-2013, 2013a.
- Crippa, M., DeCarlo, P. F., Slowik, J. G., Mohr, C., Heringa, M. F., Chirico, R., Poulain, L., Freutel, F., Sciare, J., Cozic, J., Di Marco, C. F., Elsasser, M., Nicolas, J. B., Marchand, N., Abidi, E., Wiedensohler, A., Drewnick, F., Schneider, J., Borrmann,

- S., Nemitz, E., Zimmermann, R., Jaffrezo, J.-L., Prévôt, A. S. H., and Baltensperger, U.: Wintertime aerosol chemical composition and source apportionment of the organic fraction in the metropolitan area of Paris, *Atmos. Chem. Phys.*, 13, 961–981, doi:10.5194/acp-13-961-2013, 2013b.
- Crippa, M., Canonaco, F., Lanz, V. A., Äijälä, M., Allan, J. D., Carbone, S., Capes, G., Dall'Osto, M., Day, D. A., DeCarlo, P. F., Di Marco, C. 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., 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., and Prévôt, A. S. H.: Organic aerosol components derived from 25 AMS datasets across Europe using a newly developed ME-2 based source apportionment strategy, *Atmos. Chem. Phys. Discuss.*, 13, 23325–23371, doi:10.5194/acpd-13-23325-2013, 2013c.
- Empa: Technischer Bericht zum Nationalen Beobachtungsnetz für Luftfremdstoffe (NABEL), available at: <http://www.empa.ch> (last access: 30 October 2013), 2011.
- Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, Th. F., Monod, A., Prévôt, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, *Atmos. Chem. Phys.*, 9, 5155–5236, doi:10.5194/acp-9-5155-2009, 2009.
- He, L.-Y., Lin, Y., Huang, X.-F., Guo, S., Xue, L., Su, Q., Hu, M., Luan, S.-J., and Zhang, Y.-H.: Characterization of high-resolution aerosol mass spectra of primary organic aerosol emissions from Chinese cooking and biomass burning, *Atmos. Chem. Phys.*, 10, 11535–11543, doi:10.5194/acp-10-11535-2010, 2010.
- Herich, H., Hueglin, C., and Buchmann, B.: A 2.5 year's source apportionment study of black carbon from wood burning and fossil fuel combustion at urban and rural sites in Switzerland, *Atmos. Meas. Tech.*, 4, 1409–1420, doi:10.5194/amt-4-1409-2011, 2011.
- IPCC: IPCC Fourth Assessment Report: The Physical Science Basis, working group I, final report, Geneva, Switzerland, 2007.
- Jayne, J. T., Leard, D. C., Zhang, X. F., Davidovits, P., Smith, K. A., Kolb, C. E., and Worsnop, D. R.: Development of an aerosol mass spectrometer for size and composition analysis of submicron particles, *Aerosol Sci. Technol.*, 33, 49–70, 2000.
- Jimenez, J. L., Jayne, J. T., Shi, Q., Kolb, C. E., Worsnop, D. R., Yourshaw, I., Seinfeld, J. H., Flagan, R. C., Zhang, X. F., Smith, K. A., Morris, J. W., and Davidovits, P.: Ambient aerosol sampling using the aerodyne aerosol mass spectrometer, *Geophys. Res. Atmos.*, 108, 8425, doi:10.1029/2001JD001213s, 2003.
- Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prévôt, A. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, E. J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of organic aerosols in the atmosphere, *Science*, 326, 1525–1529, 2009.
- Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prévôt, A. S. H.: Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra, *Atmos. Chem. Phys.*, 7, 1503–1522, doi:10.5194/acp-7-1503-2007, 2007.
- Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., Szidat, S., Wehrli, M. N., Wacker, L., Weimer, S., Caseiro, A., Puxbaum, H., and Prévôt, A. S. H.: Source attribution of submicron organic aerosols during wintertime inversions by advanced factor analysis of aerosol mass spectra, *Environ. Sci. Technol.*, 42, 214–220, doi:10.1021/es0707207s, 2008.
- Lanz, V. A., Prévôt, A. S. H., Alfarra, M. R., Weimer, S., Mohr, C., DeCarlo, P. F., Gianini, M. F. D., Hueglin, C., Schneider, J., Favez, O., D'Anna, B., George, C., and Baltensperger, U.: Characterization of aerosol chemical composition with aerosol mass spectrometry in Central Europe: an overview, *Atmos. Chem. Phys.*, 10, 10453–10471, doi:10.5194/acp-10-10453-2010, 2010.
- Matson, P., Lohse, K. A., and Hall, S. J.: The globalization of nitrogen deposition: Consequences for terrestrial ecosystems, *Ambio*, 31, 113–119, 2002.
- Matthew, B. M., Middlebrook, A. M., and Onasch, T. B.: Collection efficiencies in an aerodyne aerosol mass spectrometer as a function of particle phase for laboratory generated aerosols, *Aerosol Sci. Technol.*, 42, 884–898, doi:10.1080/02786820802356797s, 2008.
- Middlebrook, A. M., Bahreini, R., Jimenez, J. L., and Canagaratna, M. R.: Evaluation of composition-dependent collection efficiencies for the aerodyne aerosol mass spectrometer using field data, *Aerosol Sci. Technol.*, 46, 258–271, doi:10.1080/02786826.2011.620041s, 2012.
- Mohr, C., DeCarlo, P. F., Heringa, M. F., Chirico, R., Slowik, J. G., Richter, R., Reche, C., Alastuey, A., Querol, X., Seco, R., Peñuelas, J., Jiménez, J. L., Crippa, M., Zimmermann, R., Baltensperger, U., and Prévôt, A. S. H.: Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data, *Atmos. Chem. Phys.*, 12, 1649–1665, doi:10.5194/acp-12-1649-2012, 2012.
- Ng, N. L., Canagaratna, M. R., Zhang, Q., Jimenez, J. L., Tian, J., Ulbrich, I. M., Kroll, J. H., Docherty, K. S., Chhabra, P. S., Bahreini, R., Murphy, S. M., Seinfeld, J. H., Hildebrandt, L., Donahue, N. M., DeCarlo, P. F., Lanz, V. A., Prévôt, A. S. H., Dinar, E., Rudich, Y., and Worsnop, D. R.: Organic aerosol components observed in Northern Hemispheric datasets from Aerosol Mass Spectrometry, *Atmos. Chem. Phys.*, 10, 4625–4641, doi:10.5194/acp-10-4625-2010, 2010.
- Ng, N. L., Canagaratna, M. R., Jimenez, J. L., Zhang, Q., Ulbrich, I. M., and Worsnop, D. R.: Real-time methods for estimating organic component mass concentrations from aerosol mass spectrometer data, *Environ. Sci. Technol.*, 45, 910–916, doi:10.1021/es102951ks, 2011a.

- 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., and Jayne, J. T.: An aerosol chemical speciation monitor (ACSM) for routine monitoring of the composition and mass concentrations of ambient aerosol, *Aerosol Sci. Technol.*, 45, 770–784, doi:10.1080/02786826.2011.560211s, 2011b.
- Paatero, P.: Least squares formulation of robust non-negative factor analysis, *Chemometr. Intell. Lab.*, 37, 23–35, 1997.
- 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.*, 8, 854–888, 1999.
- Paatero, P.: User's guide for positive matrix factorization programs PMF2 and PMF3, 2004.
- Paatero, P. and Hopke, P. K.: Rotational tools for factor analytic models, *J. Chemometrics*, 23, 91–100, 2009.
- Paatero, P. and Tapper, U.: Analysis of different modes of factor analysis as least squares fit problems, *Chemometr. Intell. Lab.*, 18, 183–194, 1993.
- Paatero, P. and Tapper, U.: Positive matrix factorization – a nonnegative factor model with optimal utilization of error-estimates of data values, *Environmetrics*, 5, 111–126, 1994.
- Paatero, P., Hopke, P. K., Song, X. H., and Ramadan, Z.: Understanding and controlling rotations in factor analytic models, *Chemometr. Intell. Lab.*, 60, 253–264, 2002.
- Peng, R. D., Dominici, F., Pastor-Barriuso, R., Zeger, S. L., and Samet, J. M.: Seasonal analyses of air pollution and mortality in 100 US cities, *Am. J. Epidemiol.*, 161, 585–594, doi:10.1093/aje/kwi075s, 2005.
- Reche, C., Viana, M., Amato, F., Alastuey, A., Moreno, T., Hillamo, R., Teinila, K., Saarnio, K., Seco, R., Penuelas, J., Mohr, C., Prévôt, A. S. H., and Querol, X.: Biomass burning contributions to urban aerosols in a coastal Mediterranean City, *Sci. Total Environ.*, 427, 175–190, doi:10.1016/j.scitotenv.2012.04.012s, 2012.
- Sandradewi, J., Prévôt, A. S. H., Szidat, S., Perron, N., Alfarra, M. R., Lanz, V. A., Weingartner, E., and Baltensperger, U.: Using aerosol light absorption measurements for the quantitative determination of wood burning and traffic emission contributions to particulate matter, *Environ. Sci. Technol.*, 42, 3316–3323, doi:10.1021/es702253ms, 2008.
- Slowik, J. G., Vlasenko, A., McGuire, M., Evans, G. J., and Abbatt, J. P. D.: Simultaneous factor analysis of organic particle and gas mass spectra: AMS and PTR-MS measurements at an urban site, *Atmos. Chem. Phys.*, 10, 1969–1988, doi:10.5194/acp-10-1969-2010, 2010.
- Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., Prévôt, A. S. H., and Hueglin, C.: Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques, *Geophys. Res. Atmos.*, 112, D11307, doi:10.1029/2006jd007971s, 2007.
- Sun, Y.-L., Zhang, Q., Schwab, J. J., Demerjian, K. L., Chen, W.-N., Bae, M.-S., Hung, H.-M., Hogrefe, O., Frank, B., Rattigan, O. V., and Lin, Y.-C.: Characterization of the sources and processes of organic and inorganic aerosols in New York City with a high-resolution time-of-flight aerosol mass spectrometer, *Atmos. Chem. Phys.*, 11, 1581–1602, doi:10.5194/acp-11-1581-2011, 2011.
- Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data, *Atmos. Chem. Phys.*, 9, 2891–2918, doi:10.5194/acp-9-2891-2009, 2009.
- Watson, J. G.: Visibility: Science and regulation, *J. Air Waste Manage. Assoc.*, 52, 628–713, doi:10.1080/10473289.2002.10470813s, 2002.
- Wold, S., Esbensen, K., and Geladi, P.: Principal component analysis, *Chemometr. Intell. Lab.*, 2, 37–52, doi:10.1016/0169-7439(87)80084-9s, 1987.
- Zhang, Q., Alfarra, M. R., Worsnop, D. R., Allan, J. D., Coe, H., Canagaratna, M. R., and Jimenez, J. L.: Deconvolution and quantification of hydrocarbon-like and oxygenated organic aerosols based on aerosol mass spectrometry, *Environ. Sci. Technol.*, 39, 4938–4952, doi:10.1021/es0485681s, 2005.
- Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Ulbrich, I. M., Ng, N. L., Worsnop, D. R., and Sun, Y.: Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review, *Anal. Bioanal. Chem.*, 401, 3045–3067, doi:10.1007/s00216-011-5355-ys, 2011.