**Atmospheric
Measurement
Techniques**

Three-dimensional factorization of size-resolved organic aerosol mass spectra from Mexico City

I. M. Ulbrich[1,2], **M. R. Canagaratna**[3], **M. J. Cubison**[1,*], **Q. Zhang**[4], **N. L. Ng**[3,**], **A. C. Aiken**[1,2,***], **and J. L. Jimenez**[1,2]

[1]Cooperative Institute for Research in the Environmental Sciences (CIRES), Boulder, CO, USA
[2]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA
[3]Aerodyne Research, Inc., Billerica, MA, USA
[4]Department of Environmental Toxicology, University of California, Davis, CA, USA
*now at: TofWerk AG, Thun, Switzerland
**now at: School of Chemical and Biomolecular Engineering and School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA
***now at: Los Alamos National Laboratory, Los Alamos, NM, USA

Correspondence to: J. L. Jimenez (jose.jimenez@colorado.edu)

Abstract. A size-resolved submicron organic aerosol composition dataset from a high-resolution time-of-flight mass spectrometer (HR-ToF-AMS) collected in Mexico City during the MILAGRO campaign in March 2006 is analyzed using 3-dimensional (3-D) factorization models. A method for estimating the precision of the size-resolved composition data for use with the factorization models is presented here for the first time. Two 3-D models are applied to the dataset. One model is a 3-vector decomposition (PARAFAC model), which assumes that each chemical component has a constant size distribution over all time steps. The second model is a vector-matrix decomposition (Tucker 1 model) that allows a chemical component to have a size distribution that varies in time. To our knowledge, this is the first report of an application of 3-D factorization models to data from fast aerosol instrumentation, and the first application of this vector-matrix model to any ambient aerosol dataset. A larger number of degrees of freedom in the vector-matrix model enable fitting real variations in factor size distributions, but also make the model susceptible to fitting noise in the dataset, giving some unphysical results. For this dataset and model, more physically meaningful results were obtained by partially constraining the factor mass spectra using a priori information and a new regularization method. We find four factors with each model: hydrocarbon-like organic aerosol (HOA), biomass-burning organic aerosol (BBOA), oxidized organic aerosol (OOA), and a locally occurring organic aerosol (LOA). These four factors have previously been reported from 2-dimensional factor analysis of the high-resolution mass spectral dataset from this study. The size distributions of these four factors are consistent with previous reports for these particle types. Both 3-D models produce useful results, but the vector-matrix model captures real variability in the size distributions that cannot be captured by the 3-vector model. A tracer *m/z*-based method provides a useful approximation for the component size distributions in this study. Variation in the size distributions is demonstrated in a case study day with a large secondary aerosol formation event, in which there is evidence for the coating of HOA-containing particles with secondary species, shifting the HOA size distribution to larger particle sizes. These 3-D factorizations could be used to extract size-resolved aerosol composition data for correlation with aerosol hygroscopicity, cloud condensation nuclei (CCN), and other aerosol impacts. Furthermore, other fast and chemically complex 3-D datasets, including those from thermal desorption or chromatographic separation, could be analyzed with these 3-D factorization models. Applications of these models to new datasets requires careful construction of error estimates and appropriate choice of models that match the underlying structure of those data. Factorization studies with these 3-D datasets have the potential to provide further insights into organic aerosol sources and processing.

1 Introduction

Fine particles have important effects on human health, climate forcing, visibility, as well as deposition of nutrients and acids to crops and ecosystems. Some of the physical processes underlying these effects include deposition of toxic compounds into the lungs (Mauderly and Chow, 2008), cloud condensation-nucleus (CCN) activation (Andreae and Rosenfeld, 2008), scattering and absorbing of radiation (Watson, 2002), and particle settling and deposition (Feng, 2008). The extent and impact of these effects depend on both particle size and chemistry, with particles with submicron diameters being especially important. These aerosol effects are complex because aerosol size distributions are dynamic. Many processes can change the size distributions of aerosols, including creation of new particles by nucleation; growth by coagulation and condensation; decrease in size by evaporation of semivolatile species upon dilution, heating, or chemical reaction; and removal of particles by wet or dry deposition (Whitby, 1978). Thus, measured ambient size distributions do not reflect the original sources directly, but represent aerosols transformed by atmospheric processes. Many of the processes that change a particle's size result in simultaneous changes to its chemical composition. For example, when semivolatile organic compounds condense onto an ammonium sulfate particle, the particle size increases, and the particle becomes an internal mixture. Thus aerosol size and chemical composition are directly linked, and ideally should be measured simultaneously.

Determination of size-resolved aerosol composition is challenging and is possible with only a few techniques. The traditional method for measuring size-resolved chemistry is to collect particles separately in discrete size ranges and chemically analyze the size-fractioned samples offline. Particles are segregated by momentum (proportional to aerodynamic size) and impact onto sampling stages (e.g., DRUM sampler, Cahill et al., 1987; MOUDI sampler, Marple et al., 1991). These methods usually have a time resolution of a few hours to days, are limited by both the minimum amount of material that needs to be collected for analysis and labor and analysis costs, and cannot capture many dynamic changes of aerosol size distributions and chemical composition that occur over faster timescales.

In contrast, real-time measurements of size-resolved chemical composition can be made with aerosol mass spectrometry. Aerosol mass spectrometers can be divided into two main groups: those that measure single particles and those that measure the bulk aerosol. Single-particle aerosol mass spectrometers measure the size and chemical composition of individual aerosol particles (PALMS, Murphy and Thomson, 1995; ATOFMS, Gard et al., 1997; Thomson et al., 2000; Su et al., 2004; SPLAT, Zelenyuk and Imre, 2005). The single-particle mass spectra collected by these instruments can then be clustered into types with similar composition and the size distribution of the cluster can be

determined. These data give qualitative results about the refractory and non-refractory aerosol composition. In contrast, the Quadrupole Aerosol Mass Spectrometer (Q-AMS, Jayne et al., 2000) and Time-of-Flight Aerosol Mass Spectrometer (ToF-AMS, Drewnick et al., 2005; DeCarlo et al., 2006) quantitatively measures the size-resolved chemical composition of bulk non-refractory submicron aerosol ($\sim$30 000 particles in several minutes). The Q-AMS monitors only a few ion fragments when sampling size-resolved chemical composition. In this mode, the quadrupole is stepped slowly across the selected ion fragments; consequently, only ions at one mass-to-charge ratio (m/z) are measured for each particle. In contrast, the ToF-AMS measures all ion fragments for each particle in this sampling mode to give a more complete and precise representation of the size-resolved chemical composition of the aerosol.

Organic compounds contribute $\sim$20–70 % of the submicron aerosol mass (Zhang et al., 2007; Jimenez et al., 2009), so understanding their role in the atmosphere is important for understanding aerosol impacts. However, we have only begun to understand the complexity of atmospheric organic composition, reactions, and volatility. Organic aerosol (OA) enters the atmosphere by two mechanisms: it can be emitted directly (primary OA), or produced by secondary processes, including gas-phase chemical reactions that produce low-volatility products that condense onto the surfaces of existing particles, and aqueous-phase reactions that produce species that stay in the particle phase (secondary organic aerosol, SOA). Most primary submicron organic aerosol is the result of incomplete combustion from sources such as vehicle engines and biomass burning. These sources, and vegetation, also emit SOA precursors. Measurements of submicron OA by the AMS can provide insights into the contributions of different aerosol sources and processes. Primary organic aerosol is represented by aerosol components including hydrocarbon-like (HOA) and biomass-burning organic aerosol (BBOA, Ng et al., 2011). Secondary organic aerosol is observed as oxygenated organic aerosol (OOA) in the AMS (Dzepina et al., 2009; Jimenez et al., 2009). These aerosol components have been identified by analyzing AMS measurements with a family of mathematical techniques known as factor analysis.

Factor analysis aims to represent measured data with a reduced number of physically meaningful factors that describe underlying sources and processes controlling the variability of the original data. Possible factorization models depend on the contents of the data to be factored, how the data is arranged, and how the suspected underlying data structure can be mathematically described. Factor analytical methods have been applied to chemically speciated aerosol data for more than forty years to identify the sources of particles in urban and rural environments (e.g., Blifford and Meeker, 1967; Thurston and Spengler, 1985; Schauer et al., 1996; Chow and Watson, 2002; Engel-Cox and Weber, 2007; Reff et al., 2007; Viana et al., 2008). Size-resolved concentrations have

also been included in some analyses, but comparatively few researchers have simultaneously factored chemical composition data from particles in different size ranges (Yakovleva et al., 1999; Dillner et al., 2005; Han et al., 2006; Pere-Trepat et al., 2007; Yatkin and Bayram, 2008; Amato et al., 2009; Gietl and Klemm, 2009; Karanasiou et al., 2009; Kleeman et al., 2009; Srivastava et al., 2009). All of these studies used data with limited size or temporal resolution, typically 2–6 size ranges and 12–24-h average composition data; the study with the highest size and time resolution had 8 size ranges and 3-hour time resolution (Han et al., 2006). In fact, Larson et al. (2006) noted that their approach of combining number size distributions with bulk aerosol composition measurements in a factorization analysis was an interim approach until more size-resolved aerosol composition data from aerosol mass spectrometry was available. Size- and composition-resolved datasets such as those now available from the AMS have the potential to better characterize and constrain the changing size distribution of different aerosol species.

This study uses two 3-dimensional (3-D) factorization models to analyze a size-resolved organic composition dataset collected during the intensive MILAGRO campaign (Megacity Initiative: Local and Global Research Observations) in Mexico City in March 2006 (Molina et al., 2010). During the campaign, size-resolved submicron aerosol composition was measured near downtown Mexico City by high-resolution time-of-flight aerosol mass spectrometry (HR-ToF-AMS). Previous factor analysis of the bulk submicron organic composition dataset, i.e., without making use of the size-resolved data, identified contributions to the organic aerosol from HOA, BBOA, OOA, and a locally occurring organic aerosol (LOA, Aiken et al., 2009). The 3-D factorization techniques used in this study allow us to obtain robust estimates of the size distributions of these aerosol components and perhaps identify additional components. We first present a brief overview of 3-D factorization models for datasets that include aerosol size and composition information to determine sources of aerosol. We then discuss our methods, which include preparing the size-resolved aerosol composition measurements for factorization, and factoring these data using two 3-D models with two algorithms. We present the results of the two factorization models, then compare the solutions and discuss the advantages and disadvantages of each for this dataset. Finally, we discuss the implications of these results for future studies of aerosol effects, and the use of these models for factoring 3-D datasets from other emerging fast aerosol instrumentation.

2 Three-dimensional array factorization

Traditional factor analyses of 2-dimensional (2-D) matrices produce factors composed of two vectors. Several approaches to factor analysis of datasets that include aerosol size distributions or size-resolved aerosol composition datasets organized as 2-D matrices are presented in Sect. S1 of the Supplement. In contrast, the third dimension in 3-D arrays allows for more factorization model options (Fig. 1). The first of these models is known as "Parallel Factor Analysis" or "PARAFAC" (Harshman and Lundy, 1994) and extends the bilinear umixing model (Eq. S1 in the Supplement) to a trilinear model (Fig. 1a). In this model, each factor is described by three 1-D vectors, and we refer to it here as the "3-vector model." We arrange a $m \times n \times o$ data array X of size-resolved chemical composition such that each of the m rows holds a 2-D matrix of dimensions $n \times o$ (columns × layers of the 3-D array) that contains the measurements of o chemical components measured at each of the n sizes for one sample (time step). Equivalently, the $n \times o$ matrix could be thought of as the size distribution (across n size bins) of each of the o chemical components. The 3-D array X is reconstructed by a number of 3-vector factors as described by

$$x_{ijk} = \sum_p a_{ip} \, b_{jp} \, c_{kp} + e_{ijk}, \tag{1}$$

where i, j, and k are the row, column, and layer indices of a 3-D array; p is the number of factors; a_{ip} is an element of the $m \times p$ matrix $\mathbf{A}$, the columns of which contain the factor time series; b_{jp} is an element of the $n \times p$ matrix $\mathbf{B}$, the columns of which contain the factor size distributions; c_{kp} is an element of the $o \times p$ matrix $\mathbf{C}$, the rows of which contain the factor chemical profiles; and e_{ijk} is an element of the $m \times n \times o$ array $\mathbf{E}$ of the residuals of the solution, i.e., the difference between the measured data and the reconstruction. Note that the 3-vector model incorporates the assumption that each factor's size distribution and chemical composition are unchanging over the entire measurement period.

Another option for factoring 3-D arrays produces factors composed of a vector (1-D) and a matrix (2-D). This model is often called the "Tucker1" model (Tucker, 1966), but we refer to it here as the "vector-matrix model". Three such vector-matrix models are possible depending on which array dimension is chosen for the vector (Fig. 1b–d). The model in which the vector contains the chemical composition of one factor, and the matrix its time-varying size distribution (Fig. 1b), is described mathematically by

$$x_{ijk} = \sum_p d_{ijp} \, c_{kp} + e_{ijk}, \tag{2}$$

where x_{jik} and e_{ijk} are as described as in Eq. (1); c_{kp} is the same as in Eq. (1) and is an element of the factor composition profile; and d_{ijp} is an element of the $m \times n \times p$ array $\mathbf{D}$ in which each $m \times n$ layer represents the time-evolving size distribution of each factor. In this model, the chemical composition of each factor is assumed to be constant at all times, but the size distribution of the aerosol component can change over time.
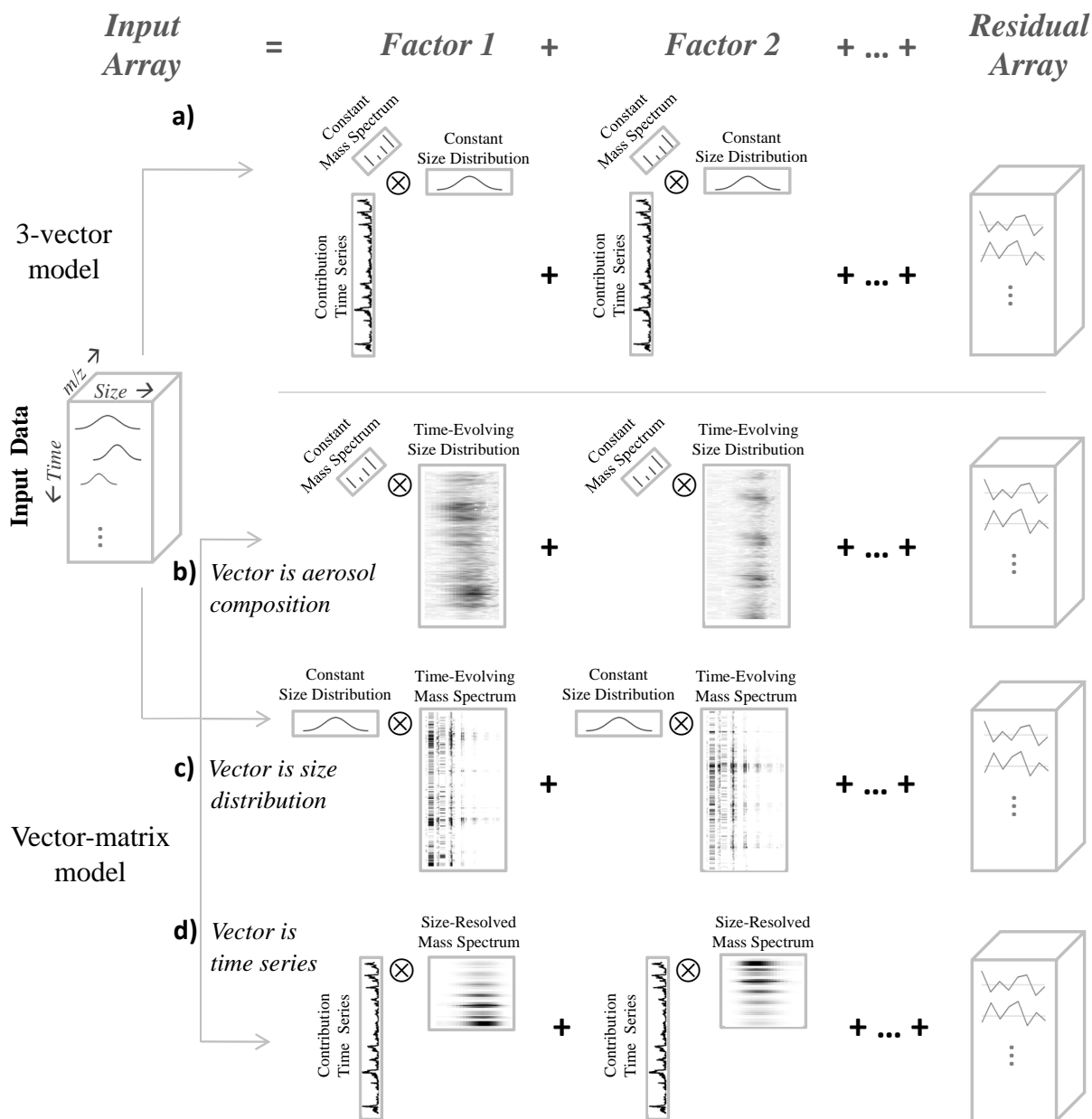
Fig. 1. Schematic representation of factorization methods of 3-D matrices. **(a)** In the 3-vector model, each factor is represented by three vectors (here a mass spectrum, a size distribution, and a time series of the contribution of the factor). **(b)**–**(d)** In the vector-matrix model, each factor is represented by one vector (representing one of the original array dimensions) and one matrix (representing the remaining array dimensions). The model can be arranged so that any of the original dimensions is represented by the vector: **(b)** the vector represents a mass spectrum, **(c)** the vector represents a size distribution; **(d)** the vector represents a time series.

Two alternate arrangements of the vector-matrix model incorporate different assumptions about aerosol properties. The vector-matrix model in which the vector contains the size distribution (Fig. 1c) could identify the changing composition of aerosol size modes. In the final vector-matrix model, the vector contains the time series, and the matrix represents size-resolved chemical composition (Fig. 1d). Thus each factor represents the size-resolved chemical composition of the

ensemble of particles that arrive simultaneously at the sampling location.

Three peer-reviewed studies that have reported factorization of 3-D size-resolved aerosol composition datasets using the 3-vector model (Fig. 1a; Yakovleva et al., 1999; Karanasiou et al., 2009) and the vector-matrix model in which the vector contains the factor time series (Fig. 1d; Pere-Trepat et al., 2007) are summarized in Table 1 and described in

Table 1. Details of research that applied 3-D factorization techniques to datasets of size-resolved aerosol chemical composition.

Citation	Yakovleva et al. (1999)[a]	Pere-Trepat et al. (2007)	Karanasiou et al. (2009)[a]	This work
Sampling location	Riverside, CA 1991	Detroit, MI 2002	Athens, Greece 2002	Mexico City, Mexico 2006
Instrumentation	Stationary indoor monitors Stationary ambient monitors Personal exposure monitors	3-stage DRUM impactor absorbance spectrometer	Custom impactors aethalometer	HR-ToF-AMS
Particle sizes	$PM_{2.5}$, PM_{10}, personal PM_{10} (2 sizes, 5 types)	0.1–2.5 µm (3 ranges)	PM_2, PM_{10-2}	0.01–1.2 µm[b] (36 ranges)
Chemical speciation	18 elements	27 elements σ_{abs} (four wavelengths)	SO_4^{2-} BC 13 elements	Organic fragments at 71 m/z's ≤ 100 amu
Sampling time	12 h	3 h	24 h	5 min, 3 times per h
Factorization model(s)	3-vector	Vector-matrix (time series as vector)	3-vector	– 3-vector – vector-matrix (mass spectrum as vector)
Algorithm(s)	PMF3[c]	ME-2[c]	PMF3	PMF3, ME-2

[a] In these studies, the data was also arranged as multiple 2-D matrices and factored using a 2-D model. These works are included in Table 1 and details are presented in Table S7 (in the Supplement).

[b] Particle transmission in the AMS begins at ∼0.04 µm, but smaller sizes were included in the factorization dataset to characterize noise level and provide a baseline for size distributions.

[c] Algorithms are Positive Matrix Factorization 3 (PMF3) and Multilinear Engine 2 (ME-2).

Sect. S2 of the Supplement. To the best of our knowledge, no published studies have applied the vector-matrix model in which the vector contains chemical compositions to aerosol datasets. Furthermore, we know of no studies that have applied 3-D factorization techniques to highly time- and size-resolved chemical composition data.

We choose to explore the 3-vector model (Fig. 1a) and the vector-matrix model in which the vector contains factor mass spectra and the matrix contains the time-varying factor size distributions (Fig. 1b). In particular, we choose this formulation of the vector-matrix model because we can recognize their chemical composition from prior work, and we expect the component size distributions to change more rapidly than their chemical composition as particles are transported from their sources (Cross et al., 2009; Canagaratna et al., 2010) and age in the atmosphere (Dzepina et al., 2009).

3 Methods

In this section we discuss the collection of size-resolved aerosol composition data using the HR-ToF-AMS, preparation of that data for factorization, and the factorization models and algorithms used in this study. Finally, we present the guidelines used to choose factorization solutions.

3.1 Mexico City measurements during the MILAGRO field campaign

Data were collected during the MILAGRO field campaign in Mexico City, Mexico, in March 2006 (Molina et al., 2010). The campaign urban supersite ("T0") was located at the Instituto Mexicano del Petroleo (IMP), 9 km NNE of the city center, near a combination of residential, commercial, and light industrial areas. A HR-ToF-AMS (referred to hereafter as AMS, Aerodyne Research Inc., Billerica, MA; DeCarlo et al., 2006) was located at T0, and data were collected from 10 to 31 March, 2006 (Aiken et al., 2009).

The sampling procedures for the AMS during the MILAGRO campaign have been described in detail by Aiken et al. (2009) and are briefly summarized here. We use the size distributions recorded in the V, i.e., single-reflectron-pass, mode of the AMS for this study. Because of multiplexing the ambient data with other experiments, size distributions were collected for 2.5 min each in two adjacent time steps, or "runs", of every eight time steps. Thus, 5 minutes of every 20 min were spent sampling in this mode, giving 6 chemically speciated size distributions per hour.

3.2 Measurement of chemically speciated size distributions with the AMS

The AMS generally alternates between two sampling modes during each time step: MS mode, in which the

ensemble-average mass spectrum of all particles is obtained; and Particle Time-of-Flight (PToF) mode, in which the size-resolved mass spectra, of particles are obtained (Jayne et al., 2000; Jimenez et al., 2003; Canagaratna et al., 2007). In both modes, particles and gases in ambient air enter the instrument and are focused into a beam by an aerodynamic lens. The lens transmits with nearly 100 % efficiency particles with a vacuum aerodynamic diameter (d_{va}) of 70–600 nm. It transmits with decreasing efficiency particles with d_{va} 35–70 nm and 600 nm to 1.5 μm (Liu et al., 1995a,b; Jayne et al., 2000; Zhang et al., 2004). Gases are also transmitted through the lens with 100 % efficiency, but are not focused at the exit of the lens. Particles and gases that are transmitted through the lens encounter a rotating chopper wheel (∼150 Hz) that is positioned differently for the MS and PToF sampling modes. Only the PToF mode is described here. In PToF mode, the chopper is positioned to partially block the particle beam, allowing a few particles through the slit during each chopper cycle. Particles and gases that pass the chopper impact a resistively heated surface (600 °C), and nonrefractory aerosol components flash vaporize. The particle vapors and gas molecules are ionized (70 eV electron impact), and the resulting cations are transmitted by ion optics to an extraction region. In the extraction region, the ions are extracted into the ToFMS and fly to the detector.

Particle size is determined from particle velocity in the particle-flight region. Particle velocity is calculated from the length of the particle-flight region from the chopper to the vaporizer (0.293 m) and particle time of flight (PToF), measured as the time between the midpoint of the opening of the chopper slit intersecting the particle beam and the pulsed extraction of ions into the mass spectrometer. This PToF includes the true PToF from the chopper to the vaporizer, and ion time of flight (IToF) from the extractor to the detector; however, since IToF ≪ PToF (<30 μs vs. ∼3000 μs, respectively), the approximation of PToF as measured is sufficient. The measured PToF and the length of the particle flight region (L_p) give the particle velocity (v_p), which is inversely related to the particle vacuum aerodynamic diameter (d_{va}), by

$$v_p = \frac{L_c}{\text{PToF}} = \frac{v_s - v_l}{1 + (d_{va}/d^*)^b} + v_l \qquad (3)$$

where v_s and v_l are the asymptotic velocities for very small and very large particles, respectively, and D^* and b are calibration constants (Jayne et al., 2000; Allan et al., 2003). Particle size (d_{va}) is calibrated to PToF using fast-vaporizing, monodisperse particles such as ammonium nitrate or polystyrene latex spheres (PSLs, at 800 °C).

Size-resolved mass spectra are obtained by recording many individual mass spectra as a function of particle flight time (Fig. 2a). The frequency of recording the mass spectra determines the achievable size resolution. Mass spectra are recorded, e.g., 100 times during each chopper cycle then reduced to 50 spectra by co-adding two adjacent spectra, to

obtain one spectrum every 50 μs in this study. Most of the mass spectra contain only noise because zero to a few particles enter the particle flight region during each chopper cycle, so only a few spectra contain particle signal. Size-resolved mass spectra from ∼10 000 chopper cycles during one 2.5-min time step are averaged into one size-resolved mass-spectral matrix for that time step (Fig. 2b). During one 2.5-min time step, sampling 50 % of the time in PToF mode with a flowrate of 2.0 cm³ s⁻¹, a 2 % chopper opening, and a typical urban concentration of 10 000 particles cm⁻³, this matrix contains the average size-resolved mass spectra of ∼30 000 particles.

This description of PToF sampling describes "ideal" PToF sampling in which particles stick to the vaporizer and vaporize quickly. However, the measured particle size can be affected if non-ideal behavior changes any part of the measured PToF. The measured PToF includes three components: the true PToF from the chopper to the vaporizer, the vaporization and ionization time, and the time for ions to transfer from the ionization region to the TOF mass spectrometer extraction region.

We discuss possible deviations from each component and their effects here. First, the true PToF can change if a particle bounces off the vaporizer, then impacts another surface and evaporates; thus the vaporization is "delayed". If the vaporized molecules from this particle are ionized and the ions are mass analyzed, they are recorded at a nominal size larger than the true particle size (Cross et al., 2009). Second, different aerosol species have different vaporization rates. "Fast" evaporation happens on the timescale of <200 μs for several aerosol constituents (NH_4NO_3 and $(NH_4)_2SO_4$ vaporize in 80 and 150 μs, respectively, at ∼600 °C, Cross et al., 2009). In contrast, some inorganic and organic compounds in ambient aerosol can have longer evaporation timescales (1/e decay), e.g., <10⁵ μs (0.1 s for PbCl, Salcedo et al., 2010), >5 × 10⁶ μs (5 s for ∼5 % of the total organic mass, Huffman et al., 2009) , or ∼1.4 × 10⁸ μs (2.4 min for other Pb compounds, Salcedo et al., 2010). Ions from these longer vaporization timescales are measured as part of the background signal in PToF mode, while ions from moderate vaporization timescales are recorded at larger nominal sizes than the true particle size. The final component of the measured PToF is the transfer time from the ionization region to the pulsed extraction region of the TOF mass spectrometer. For typical voltages of the transfer ion optics (of the order of a few tens of volts; we will use 20 V for this example calculation), the time to travel 96 mm between the ionization and extraction regions (Drewnick et al., 2005) is 6 and 27 μs for m/z's 17 and 300, respectively.

Thus, transfer time ≪ vaporization time ≪ measured PToF (∼3000 μs), and only variations in the vaporization time could potentially be large enough to shift the recorded particle size. Note that these deviations broaden the size distribution only towards nominally larger particle sizes. These possible deviations should be considered part of
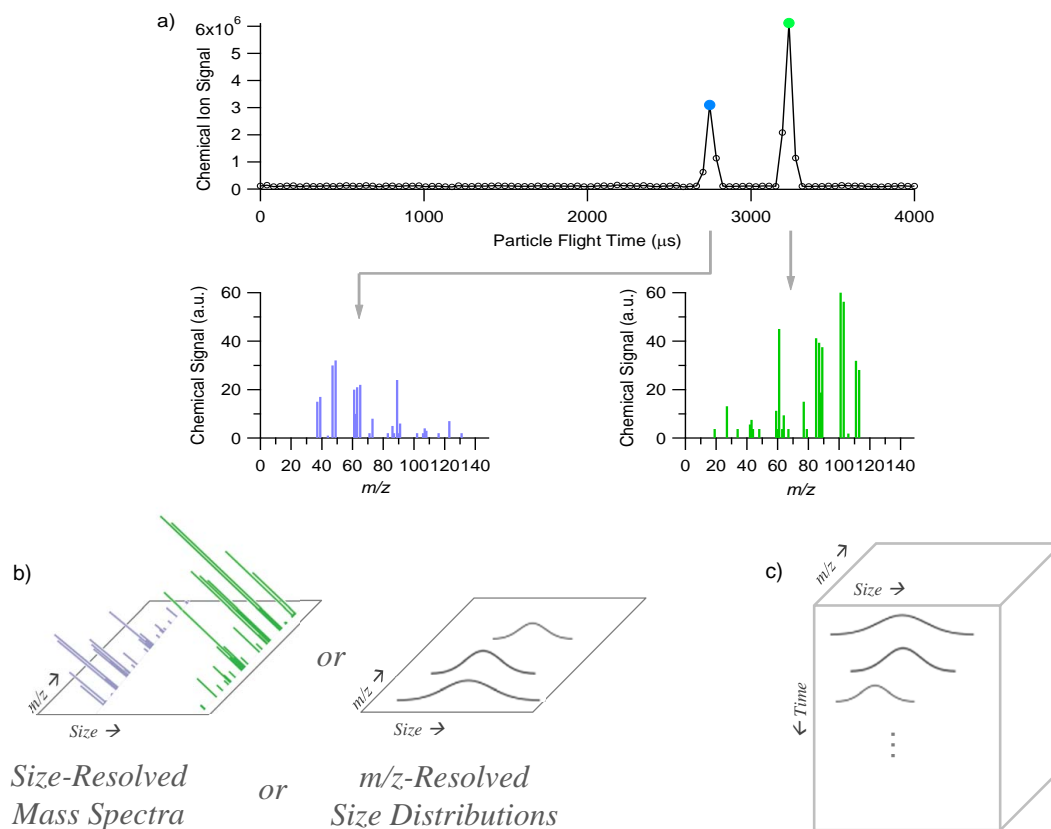
Fig. 2. Schematic representation of data collection for chemically resolved particle-size-distribution data. **(a)** Two particles have entered the particle flight region before being vaporized and ionized. Mass spectra are collected ∼50 times during the particle flight time (circles). Ions from the smaller particle arrive first (blue circle), and their mass spectrum is recorded (blue spectrum). Ions from the larger particle arrive second (green circle), and their mass spectrum is recorded (green spectrum). (Adapted from Cross et al., 2009 with permission.) **(b)** The size-resolved mass spectra can be arranged as a 2-D matrix in which one dimension represents the mass-to-charge (m/z) ratio, and one dimension represents particle size. The same data could be conceptualized as m/z-resolved size distributions if the points in each m/z row are connected. **(c)** Over many measurement periods, the 2-D samples can be arranged into a 3-D array.

the "transfer function" of the instrument, and could be further evaluated and corrected, although such techniques are outside the scope of this paper.

3.3 Particle time-of-flight data analysis and data pretreatment

Quantification of the size-resolved particle signal requires subtracting the background signal. However, the background signal is not explicitly measured in PToF mode. Instead, each PToF-mode sample (during one time step or "run") includes measurements before and after particle signals arrive at the detector, and these periods, or "DC regions," can be used to estimate the signal background (Allan et al., 2003). The signal from both periods is usually used to determine the background (Fig. S1a in the Supplement); however, some ions have high signal from gas-phase ions that arrive before the particles, and only the latter region is used to determine the background level (Fig. S1b in the Supplement). The signal

over the selected regions is averaged, and, analogous to a "DC offset", is subtracted from the measured signal. The estimated background subtraction is performed separately for each ion in every time step.

After subtracting the estimated background from the PToF matrix of the total aerosol and gas-phase signal, the organic ion fragments are separated from the bulk aerosol by applying a "fragmentation matrix" to the mass spectrum measured at each particle size (Allan et al., 2004). In this work we use the original Allan fragmentation matrix, not the updated matrix by Aiken et al. (2008). Details of changes to the fragmentation matrix for this study are included in Sect. S3 of the Supplement.

We now arrange the size-resolved organic aerosol mass spectra obtained over many sampling periods into a 3-D array of dimensions $m \times n \times o$, as described in Sect. 2.3 (Fig. 2c). The array is arranged so that the rows represent time steps (m time averages), the columns represent particle size (n sizes), and the "layers" or "pages" (the third dimension)

Table 2. Comparison of error preparation steps in Ulbrich et al. (2009) and this work. "X" marks the steps performed in each work.

Error preparation step	Ulbrich et al. (2009)	This work
1. Calculate σ_{ijk}	X	X
2. Apply minimum error	X	X
3. Smooth data	X (time dimension only)	X (time and size dimensions)
4. Remove datapoints without useful information		X
5. Downweight low-SNR data X	X	
6. Downweight m/z's directly proportional to m/z 44	X	X

represent ion fragments in a mass spectrum (o m/z's, Fig. 1). Each element of the array has units d$Signal$/d$PToF$, which are nonlinearly transformed to units of d$Signal$/dlogd_{va} (Allan et al., 2003). Units of d$Signal$/dlogd_{va} are converted to d$Mass$/dlogd_{va} by integrating d$Signal$/dlogd_{va} for each time step, applying AMS calibration factors (Jimenez et al., 2003), and normalizing the total PToF signal to the total mass of the same time step measured in MS mode.

3.3.1 Estimation of measurement precision of particle time-of-flight data

A method to estimate the precision of AMS PToF data has not been reported previously. This quantity is required for the factorization analyses, and so we have developed a method for its estimation here. The precision (or random error, σ, often termed "error" in studies using Positive Matrix Factorization, PMF) of the measured signal for one m/z at one size in each time step can be estimated by the sum of three terms in quadrature:

$$\sigma = \sqrt{\sigma_{IC}^2 + \sigma_{DC}^2 + \sigma_{elec,scat}^2}, \qquad (4)$$

where σ_{IC} is the precision from Poisson ion-counting statistics, calculated from the ion signal before DC-offset subtraction (i.e., $\sigma_{IC} = \sqrt{I/t_s}$, where I is the ion signal in ions per second and t_s is the time spent sampling that size bin and m/z in seconds, analogous to the method of Allan et al. (2003) for MS mode data); σ_{DC} is the standard error of the signals used to estimate the DC offset (estimated as $\sigma_{DC} = s/\sqrt{n}$, where s is the sample standard deviation of the points averaged for the DC offset and n is the number of points in that average – $n = 14$ for both DC regions and $n = 6$ for the later DC region in this study); and $\sigma_{elec,scat}$ is the error from electronic noise and scattered ion signals that are presumed to contribute to every measured ion signal (estimated as the standard deviation of the signal at very high m/z's at all particle sizes). In this study, $\sigma_{elec,scat}$ was estimated from the signals at m/z's > 400 for 200 time steps that did not appear to contain actual particle signal (defined as signal at least four times the average noise level). A fourth term could be added to the estimation of σ to reflect particle-counting statistics when particle number concentrations are low.

3.3.2 Further data and error treatments prior to factorization

Further treatments to the data and error matrices are needed prior to array factorization for three main purposes: (1) to decrease the influence of low signal-to-noise ratio (SNR) data in the factorization, (2) to remove array elements that do not contain useful particle information, and (3) to decrease the weight of duplicated information within the array. To implement these treatments, we follow the five-step procedure described by Ulbrich et al. (2009), with some additions specific to the PToF data (Table 2). The steps are described below, and numbered as in Table 2.

1. The first step in the procedure is to calculate the error for each measurement according to Eq. (4).

2. In the second step, we apply a minimum error threshold equal to the signal measured from one ion during one time step. This correction decreases the weight of points whose estimated error is smaller than one count. In the Ulbrich et al. (2009) study, this step had two functions: it replaced error values calculated as less than one ion, but also identified error values that were lower than the average of its neighbors in time and replaced the calculated error with that average. However, this second part of the correction cannot be applied here because only two adjacent measurements were made, so no point has two truly contiguous neighbors. The minimum-error correction increases the estimated σ only for very small signals. In this study, one ion is equivalent to 1.24 Hz or 0.03 µg m^{-3} decade^{-1} log(d_{va}), and 52 % of the data points included in the final array have their error increased in this step by ~8 % on average. The increases are larger for particular m/z's and for some particle sizes (Fig. S2 in the Supplement). Many ions with m/z > 85 required uncertainty increases of 20–70 %. The average increased uncertainty at each particle size was of smaller magnitude than the average increase by m/z and mainly affected particles that are transmitted with lower efficiency through the aerodynamic lens, i.e., d_{va} < 45 nm or > ~900 nm. The large fraction of points

in the array that require this correction is a consequence of the limited SNR for this PToF dataset.

3. The limited SNR of the PToF-mode data is partially the result of lower sampling duty cycle in PToF mode than in MS mode. In MS mode, the particle beam is sampled at least 50 % of the time during the "open" mode, but in PToF mode, the chopper allows particles to enter the sampling region during typically 2 % of the sampling time. The SNR of the PToF-mode data is further decreased because the ions produced from particles that reach the vaporizer are sampled over many nominal particle sizes instead of being averaged together as in MS mode. To reduce the impact of high-frequency noise on the PToF data, we follow the third step of the Ulbrich et al. (2009) procedure and smooth the data in the size and time dimensions. The particle size dimension was smoothed binomially by two points for each m/z in each time step. In the time dimension, two adjacent time steps were averaged together to obtain 5-min sample averages. The effect of these steps in the estimated precision was propagated in quadrature in the PToF uncertainties.

4. The minimum error step shows that many of the signals in the array are small; the small signals have two main causes. First, the array includes many points that we know do not contain useful information about particle size or composition. In the size dimension, some points represent signal before and after particles can arrive, based on the theoretical lens transmission; indeed, these are the sizes that were used to calculate the PToF background in Sect. 3.3. Because these data should not contain particle signal, we discard them and retain the data for nominal particle sizes $10 \, \text{nm} \leq d_{va} \leq 1200 \, \text{nm}$. Although we expect very little or no particle signal at the extremes of this range, we retain these edges so that both tails of the size distributions approach zero. Second, some m/z's have little chemical information. The signal from organic fragments is found predominantly at m/z's ≤ 100. For example, DeCarlo et al. (2008) report that 91 % of the organic signal was found below m/z 100 for an aircraft HR-ToF-AMS dataset during MILAGRO. In addition, ions at higher m/z's have lower SNR. Mass fragments with m/z's from 1 to 826 were measured in PToF mode during this study, but for this analysis we retain only m/z's ≤ 100 with organic fragments. Finally, 2 % of the time steps have total PToF mass concentration less than zero because of noise at low actual concentrations, and we omit these time steps from further analysis. These steps for removing particle sizes, ion fragments, and time steps with little organic particle information were not part of the Ulbrich et al. (2009) procedure.

After removing sections of the array with little organic particle information, the resultant array contains measurements at 1366 time steps (rows), 36 size bins (columns), and 71 m/z's (layers), or $\sim 3.5 \times 10^6$ data points. This is a 32-fold reduction from the original array size of $\sim 1 \times 10^8$ data points. Using this smaller array for the factorization analysis greatly reduces computer time, memory, and storage requirements while preserving the high time-, size-, and chemical resolution of the useful information.

5. Now that the data array contains only the points that will be used for further analysis, we assess the SNR to identify portions of the data with low-information content. The data with low-information content are reweighted, or "downweighted", to decrease their weight in the fit (Paatero and Hopke, 2003).

Metrics for assessing the information content of the data based on calculated SNR are discussed in detail by Paatero and Hopke (2003) for 2-D datasets, but these authors make no recommendations for 3-D datasets. In 2-D datasets, average SNR is calculated over all of the time steps for each variable; the variables in 2-D AMS datasets are the m/z's. In contrast, the average SNR for 3-D datasets can be calculated in more ways. Three examples are given here: (1) the SNR could be calculated for each m/z at each size, averaged across all of the time steps; (2) the SNR could be calculated for each size, averaged across all of the m/z's and time steps; or (3) the SNR could be calculated for each m/z, averaged across all sizes and time steps. The third method was used in the only published 3-D factorization study that downweighted low-SNR data (Pere-Trepat et al., 2007). In that study, the authors downweighted the data for selected chemical species at all sizes and all times by factors of 3 or 10, but no details were provided about the criteria used to determine which species should be downweighted. In our dataset, however, we know from applying the minimum error that the information content (SNR) in the array varies with m/z and particle size. For example, the minimum error was much larger than the calculated error at many of the higher m/z's, and the smallest and largest particle sizes contain little particle information regardless of m/z. Thus, we use the first method and calculate the SNR separately for each m/z-size combination by

$$\text{SNR}_{jk} = \sqrt{\sum_{i=1}^{m} x_{ijk}^2 \Big/ \sigma_{ijk}^2}. \tag{5}$$

We now use the average SNR values to downweight m/z-size combinations with low SNR. The threshold for low SNR recommended by Paatero and Hopke (2003) is 2, and the recommended threshold for very-low SNR

is 0.2; Paatero and Hopke call the low and very-low SNR data "weak" and "bad," respectively. No data are bad in the current dataset by that criterion. However, 89 % of the m/z-size combinations are weak when using the recommended threshold of SNR < 2. Since we have established that the SNR is generally weak for this data, and Paatero and Hopke note that the SNR threshold for "weak" variables is somewhat arbitrary, we set the threshold for weak data at SNR of 1.5. Furthermore, m/z 15 is the only m/z with "good" SNR at $d_{va} < 45$ nm (Fig. S3 in the Supplement), which is near the lower range of particles transmitted by the aerodynamic lens ($\sim$35 nm). Thus, this signal is unlikely to represent particles and may instead represent gas-phase contributions of ^{15}N. Therefore, we increase the "weak" SNR threshold for m/z 15 to 1.6 such that the signals of m/z 15 for $d_{va} < 125$ nm and $d_{va} > 412$ nm are also "weak." With this change, 77 % of the m/z-size combinations are still weak, but nearly all m/z's have strong signal for particles sizes that are transmitted through the instrument's aerodynamic lens with 100 % efficiency (Fig. S3 in the Supplement). We can therefore emphasize these high SNR data by downweighting the weak m/z-size combinations. We increase the calculated error for the weak m/z-size combinations (those with SNR $\leq$ 1.5, or SNR $\leq$ 1.6 for m/z 15) by a factor of 2 (Paatero and Hopke, 2003).

6. Finally, we downweight information that is repeated in the data array because of the application of the fragmentation matrix. In the fragmentation matrix, the organic signal and uncertainty at m/z's 16, 17, and 18 are defined to be proportional to the signal at m/z 44. Thus the information for m/z 44 is repeated in the data array four times. This repetition is chemically meaningful, but if these m/z's are used without modification, they have undue additional weight in the factorization analysis. We therefore downweight the signal at these four m/z's by the square root of 4 so that this information is weighted the same as any other single m/z (Ulbrich et al., 2009).

3.4 Array factorization

After performing these data analysis and pretreatment steps, the data array is ready for factorization. We now discuss the factorization models used in this study, the algorithms used to solve the models, and the guidelines used for choosing factorization solutions. All factorization results were examined using the PMF Evaluation Tool (PET) described previously (Ulbrich et al., 2009), with custom modifications for 3-D array factorization.

3.4.1 Models for factoring the 3-dimensional array

Four models for factoring 3-D arrays were presented in Sect. 2 (Fig. 1). This study applies two of these models to

the MILAGRO dataset. The first model is the 3-vector model (Fig. 1a). Recall that the 3-vector model is so named because each factor is composed of three vectors. When this model is applied to the present dataset, the vectors contain the factor's chemical composition (here a mass spectrum), size distribution, and mass concentration time series. The 3-vector model thus assumes that each factor's size distribution and chemical composition is unchanging over the entire measurement period. The second model used for this study is the vector-matrix model. Recall that in the vector-matrix model, each factor is composed of a vector and a matrix. In this study, we use a vector-matrix model in which the vector contains the aerosol composition, i.e., mass spectrum (Fig. 1b). Hereinafter we use "the vector-matrix model" to refer to this variant of the three possible vector-matrix models, unless otherwise noted. This vector-matrix model assumes that each factor's chemical composition is constant at all times, but the size distribution of each factor can change over time. To compare the factors from solutions of the 3-vector and vector-matrix models, we must match the shapes and units. In the factors from both models, we normalize mass spectra to sum to 1 (as is standard for AMS data, Ulbrich et al., 2009, but in contrast with the usual mass spectrometric practice of normalizing the m/z with the highest signal to 100). In the 3-vector model, we normalize the area under the size distribution to sum to 1, thus giving units of mass concentration (μg m^{-3}) to the time series. In contrast, in the vector-matrix model, the matrix represents dM/dlogd_{va} with units μg m^{-3}/decade^{-1} log(d_{va}). From this matrix, a factor's average normalized size distribution can be calculated by averaging all of the size distributions in the matrix, then normalizing to unit area. Similarly, the total time series can be calculated by summing the area under the size distribution from each time step. These two reductions of the matrix from the vector-matrix model can then be compared directly with the 3-vector model results.

When the model does not match the data

Both models used in this study – and the bilinear model in 2-D factorization studies – assume that there is no variability in the mass spectrum of each factor. However, as we discussed in the Introduction, aerosol undergoes chemical processing in the atmosphere and thus its spectrum can change, so the assumption of constant factor mass spectra cannot be strictly correct. These variations might appear in two forms in the factorization results. The variation could be fit by an additional factor, or the variation may not be fit well and appear in the residuals. For example, Ulbrich et al. (2009) reported a dataset in which the time series of the residuals and Q/Q_{exp} contributions were highly correlated with the time series of the semi-volatile OOA-II factor, suggesting that there were changes in the composition of OOA-II that could not be fit with additional factors. Unfortunately, we are not aware of criteria that might predict when changes in the underlying

data are significant enough to cause separate factors, and so must instead consider this possibility carefully when choosing a particular factorization solution.

Despite the limitations of the assumptions of the models used in this work, we believe that they are the most appropriate for the data being analyzed here. Models that allow spectral variation should be considered in future research.

Adding a priori constraints

Solving these models requires no a priori information about the factor mass spectra, size distributions, or time series. However, a priori knowledge can be incorporated to constrain the solution if the researcher deems the additional information appropriate and reliable. For selected analyses in this study we use the mass spectra obtained from previous analysis of the MS-mode data from this dataset (Aiken et al., 2009; henceforth referred to as "the HR-MS" factors, spectra, or solution) as starting guesses for solving the models. Thus, a priori information is available, but it may not directly correspond to the mass spectra obtained from factoring the PToF data. For example, different evaporation timescales of different organic components on the AMS vaporizer (Sect. 3.2) could lead to slightly different factors obtained from the mass spectral data compared to the characteristic mass spectra recorded in the PToF mode.

To allow for differences between the HR-MS and PToF mass spectra, we introduce a regularization parameter that allows the intensity, c, at each m/z in each reference factor to deviate from its starting value. Lanz et al. (2008) used a parameter, α, that allowed c in an a priori mass spectrum to vary by a fraction $\pm\alpha$ from its original value. With this parameter, the allowed range from the starting value is $c_0 - \alpha c_0 \leq c_0 \leq c_0 + \alpha c_0$, for $0 \leq \alpha \leq 1$. Thus when α is at its maximum, c may range from 0 to $2c_0$. However, if an a priori spectrum is too dissimilar from the latent spectra in the dataset – especially if the starting guesses have too-small values for important ion fragments – even an α of 1 may not allow the solution sufficient flexibility to find a good solution. Thus we define a different regularization parameter, β, which allows c to fractionally approach the limits 0 and 1. We implement β by

$$c_{low} = c_0 - \beta (c_0 - 0),$$

$$c_{high} = c_0 + \beta (1 - c_0), \tag{6}$$

where c_{low} and c_{high} are the low and high limits for c, respectively, and $0 \leq \beta \leq 1$. In this formulation, c_{low} is identical to its formulation by α, but c_{high} allows m/z's with small c_0 to grow substantially if necessary.

3.4.2 Algorithms for solving the 3-dimensional models

The two algorithms/software tools used in this study to solve the 3-vector and vector-matrix models are Positive Matrix

Factorization 3 (PMF3, Paatero, 1997b) and the Multilinear Engine 2 (ME-2, Paatero, 1999). PMF3 can only solve the 3-vector model, while ME-2 is a flexible tool that can solve both models, as well as other multilinear and quasi-multilinear models (Paatero, 1999). Both algorithms constrain the values in the factor matrices to be positive. Specifically, in PMF3 the values in factors are constrained to be positive (Paatero, 1997b), and in ME-2 the values of the factors are constrained by default to be non-negative, i.e., ≥ 0 (Paatero, 1999). The positivity constraint helps produce physically meaningful factors because real mass spectra, size distributions, and mass concentrations have all positive values, and negative values arise only because of noise.

Both algorithms evaluate potential solutions by minimizing a quality-of-fit parameter, Q, defined as the sum of the error-weighted residuals of the entire data array, or

$$Q = \sum_{i=0}^{m} \sum_{j=0}^{n} \sum_{k=0}^{0} \left(e_{ijk} / \sigma_{ijk}\right)^2. \tag{7}$$

A theoretical "expected" value of Q (Q_{exp}) is approximated by the number of points in the data array minus the degrees of freedom in the solution (i.e., the number of points in the solution arrays, Paatero et al., 2002), or

$$Q_{exp} = m\,n\,o - p\,(m + n + o) \text{ for the 3-vector model,} \tag{8}$$

and

$$Q_{exp} = m\,n\,o - p\,(m\,n + o) \text{ for the vector-matrix model.} \tag{9}$$

The change in Q_{exp} with the addition of each factor (i.e., increasing p by 1) is small for the 3-vector model, but more substantial for the vector-matrix model. For example, in the present dataset, each additional factor in the 3-vector model decreases Q_{exp} by 0.04 % of the array size $[(m+n+o)/mno]$, while each additional factor in the vector-matrix model decreases Q_{exp} by 1.4 % of the array size $[(mn+o)/mno]$. Because the vector-matrix model has more degrees of freedom, we expect it to fit substantially more of the array information with the vector-matrix model than the 3-vector model.

We use Q_{exp} to normalize the Q values for solutions of the models. If each point in the array is fit within its prescribed uncertainty, e_{ijk}/σ_{ijk} is ~ 1, $Q \sim$ the size of the array (mno), and $Q/Q_{exp} \sim 1$. However, if the uncertainty values have been calculated incorrectly, Q/Q_{exp} may be higher or lower than 1. Note that Q/Q_{exp} may be larger than 1, even if the errors have been specified correctly, if variations in the data do not behave according to the model. For example, if a factor's size distribution is not constant, it will not be fit well with the 3-vector model. Similarly, if a factor's mass spectrum varies in time in a way that cannot be fit well with an additional factor, it will be fit poorly by both models (Ulbrich et al., 2009). The Q/Q_{exp} values calculated with the error array used in the computations in this study are artificially low since so many of the points have been downweighted because of low

SNR, interference from gas-phase molecules, or repetition of chemical information. Thus we find it useful to recalculate the Q/Q_{exp} values using the error estimates calculated before downweighting (these error estimates do include application of the minimum error and uncertainty propagation of smoothing). We call these recalculated values "unweighted Q/Q_{exp} values." We present unweighted Q/Q_{exp} values in this study unless otherwise noted.

To solve the models, both the PMF3 and ME-2 algorithms begin by filling the factor matrices with random values determined from random seeds. Then the algorithms iteratively minimize Q. The two algorithms use different minimization algorithms. PMF3 uses a Gauss-Newton algorithm (Paatero, 1997b) while ME-2 uses the conjugate gradient method (Paatero, 1999). The difference in minimization algorithms means that PMF3 and ME-2 may not find identical solutions for the same problem, though solutions from the two algorithms should be similar.

Finally, we note several details about the configuration of the algorithms for this study. We run both algorithms in the "robust mode", in which outliers ($|e_{ijk}/\sigma_{ijk}| > 4$) are dynamically reweighted during the iteration so that they cannot pull the fit with weight >4. The algorithm uses three levels of convergence during the iteration; the convergence criteria for these levels were set to $Q_{exp} \times 10^{-4}$, $Q_{exp} \times 2 \times 10^{-5}$, and $Q_{exp} \times 10^{-5}$, respectively. Thus the final convergence criteria are ~35 and ~31 absolute Q-units for the 3-vector and vector-matrix models, respectively.

3.4.3   Guidelines for choosing a solution

The solutions to positively constrained unmixing models are not unique, and no set of mathematical criteria have proven sufficient to identify the best solution of a factor analytical model. Thus the modeler must choose the "best" solution from the set of possible solutions; this choice is unavoidably subjective. Paatero and Hopke (2009) note the importance of disclosing subjective decisions in publications of factor analyses so that the analyses can be repeated or modified by other researchers. In that spirit, we present here our guidelines for choosing the best solution from the 3-D factorizations, and discuss the acceptability of each candidate solution of the two models in two Appendices to the paper.

Choices regarding the best solution must be made in two main areas. First, the number of factors in the solution must be determined. Second, multiple solutions with this number of factors may exist, and one family must be chosen as the best solution. The choice of the best number of factors is discussed in more detail elsewhere (Paatero and Tapper, 1993; Paatero et al., 2002; Ulbrich et al., 2009) and is only briefly described here. We use the recommendations of Ulbrich et al. (2009) and consider these criteria for choosing a solution: (1) $Q/Q_{exp} \sim 1$, (2) decrease in the rate of change of Q/Q_{exp} with increasing number of factors, (3) little structure in the solution residuals, (4) strong correlation between component time series and diurnal cycles with those of tracers not included in the factorization array, and (5) plausibility of the factor mass spectra and their similarity to observed spectra of real-world sources. Criteria (4) and (5) form the basis for claiming that factors are "physically meaningful" – that is, that physically meaningful factors can be linked to atmospherically relevant sources or processes and assigned meaningful names that are chemically consistent with the factor mass spectra. For this dataset, we can also compare to the HR-MS solution, which identified four factors: HOA, BBOA, OOA, and LOA. We presume that the bulk aerosol has the same composition whether measured in the MS or PToF modes since the sampling modes are alternated every few seconds, i.e., much faster than aerosol sources or processes change at a fixed location. We therefore hypothesize that we should find the factors identified in the HR-MS analysis in the 3-D analysis, and may identify additional factors.

It is possible that the solution space for a given number of factors may contain multiple solutions; these solutions represent local minima in the Q space (Paatero, 2000, 2007). Solutions from different local minima usually have different Q/Q_{exp} values, but the solution with the lowest Q/Q_{exp} value is not necessarily the best solution. The possibility of solutions at local minima can be explored by varying the seed for the starting of the algorithm so that the algorithm begins from different parts of the Q space and might therefore encounter local minima. We calculate each solution from 50 different starting seeds and then compare these solutions (Ulbrich et al., 2009). The 50 solutions can be grouped into "families" of solutions by comparing Q/Q_{exp} values and the similarity of factors within the family (Allan et al., 2010; De-Carlo et al., 2010). Each family can then be represented by the average of the solutions in that family, and one family can be selected as the best solution. In some cases of factorization analysis there may not be sufficient evidence to support one solution over another that has a different physical interpretation. In such cases, researchers should report both solutions and the available supporting evidence, instead of choosing and reporting only one solution.

3.4.4   Uncertainties in the chosen solution

In addition to choosing the best solution, we would like to understand the uncertainty of the factors in this solution. Four main approaches to estimating the uncertainty of PMF and ME-2 solutions have been reported. First, PMF3 can report the standard deviations of the elements of one factor matrix (**A**, **B**, **C** in Eq. 1) from the diagonal of the joint covariance matrix of all factor elements (i.e., of all elements of **A**, **B**, and **C**) (Paatero, 2007). However, such estimates are not reported by ME-2 and thus could not be compared for the vector-matrix model; therefore we do not use this method. Second, bootstrapping with row replacement (Press et al., 2007; Norris et al., 2008; Ulbrich et al., 2009) has been used with 2-D input matrices to estimate uncertainty in the

factors; however, for 3-D input matrices more replacement schemes are possible and exploration of this complex topic is outside the scope of the current study. Third, the rotational uncertainty of 2-D solutions has been explored with a parameter known as FPEAK (Paatero, 1997a; Paatero and Hopke, 2009), and the range of rotated solutions has been used as an indication of the uncertainty in the factor solutions (Ulbrich et al., 2009). However, no such parameter is available for rotations in 3-D arrays. When 3-D matrices are factored with a 3-vector model, each solution usually has no rotational ambiguity (Hopke et al., 1998), even without non-negativity constraints. However, rotational ambiguity can occur between two factors with one dimension, such as a size profile, that is (almost) identical. In contrast, solutions of vector-matrix models usually have some degree of rotational ambiguity, especially between factors that have few or no zero values. In both cases, the rotations are partially constrained by non-negativity constraints. Further exploration of the rotational ambiguity of the solutions of vector-matrix models should be the focus of future research. Exploration of rotations is outside the scope of the current study and was not attempted.

Finally, uncertainty in the factors can be estimated by the variation in solutions in the same family. The variation amongst these solutions may depend strongly on the shape of the Q space near a local minimum and the algorithm's convergence method. Therefore these variations may better describe the solutions' mathematical variation than the physical variation that would help us evaluate uncertainty in aerosol properties. Nevertheless, this approach may give some insight into the mathematical uncertainties of the solution. We calculate the variation among the solutions in one family as the coefficient of variation (σ/μ, where μ is the mean) of average mass spectra over all m/z's, size distribution over all size bins, or time series over all times, after excluding points with very small means (below 0.002 fraction of mass-spectral signal, $0.006\,\mu\mathrm{g\,m^{-3}/decade^{-1}}\,\log(d_{\mathrm{va}})$ for size distributions, and $0.005\,\mu\mathrm{g\,m^{-3}}$ for time series).

4 Results

This section presents the results of the 3-vector and vector-matrix model factorizations, including the choice of the "best" solution for each model. We first present results that apply to both models, then discuss the choice of solution and physical interpretation of the factors for each model. Several results apply broadly to factorization with both the 3-vector and vector-matrix models for the current dataset and are discussed here.

Q/Q_{exp} values calculated by the algorithms using the downweighted error array were much smaller than the expected value for a good solution (~ 0.45 vs. 1, Fig. 3). The low Q/Q_{exp} (Fig. 3b) values reflect the large fraction of size-bin-m/z combinations that were downweighted because of low SNR. However, unweighted Q/Q_{exp} values are close
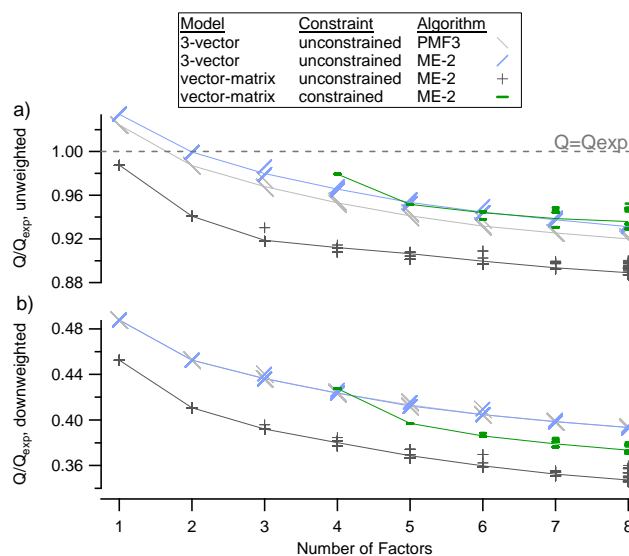

Fig. 3. Values of Q/Q_{exp} for the 3-vector and vector-matrix factorization models with 1 to 8 factors. Two algorithms were used to solve the two models: PMF3, which can only solve the 3-vector model; and ME-2, which can solve a variety of factorization models. Each mark shows the Q/Q_{exp} value of one of 50 seeds used to start the algorithm. Lines connect the average Q/Q_{exp} value of the 50 seeds for each model-algorithm combination. Marks that appear as an "X" are the intersection of "/" and "\". In the constrained vector-matrix case (green dashes), the mass spectra of four factors are fully constrained as described in the text. **(a)** Q/Q_{exp} values have been recalculated without downweighting. **(b)** Q/Q_{exp} values were calculated with the downweighted error array.

to 1 for solutions of both the 3-vector and vector-matrix models (Fig. 3a). We would expect unweighted Q/Q_{exp} values to be somewhat greater than 1 unless the model can fit all the real variability in the data. Or, the errors may be slightly overestimated for the PToF data. We expect that any small, systematic issues in our error estimate procedure do not depend strongly on m/z, size, and/or time. Thus, the weighting of the data in this study should be consistent and we believe that our results represent the actual structure of the dataset.

The suggested criterion of a steep change in the slope of Q/Q_{exp} vs. the number of factors in the solution for choosing a solution for factor analytic models was not observed in any of the models for this dataset; thus we cannot use this criterion to choose a solution and must rely upon the remaining criteria from Sect. 3.3.3: little structure in the solution residuals, correlation with the time series of tracers, and identification of the factors from the HR-MS solution.

Taking these criteria into account, we explore solutions of the models. In solutions with at least three factors, the 50-seed solutions do not all contain the same factors, but the solutions can be arranged by factor similarity into several families (Table 3). Although seed-dependent families have been reported previously for 2-D factorization of some AMS

Table 3. Types of factors identified in each solution family from 50 "seed" trials of the 3-vector and vector-matrix models solved by ME-2. Each row represents one family of solutions, and each X represents one factor in a solution. Multiple Xs in one entry denote multiple instances of this factor in the same solution. These factors usually have similar mass spectra but different size distributions and time series (Fig. S6 in the Supplement). The columns with grey shading represent factors related to oxygenated organic aerosol (OOA). Each solution contains one OOA factor or two factors dominated by m/z 44 and m/z 43. The column with darker grey shading represents factors that are dominated by m/z's 67, 81, and 95. These factors are not physically meaningful. The best solution of the 3-vector model has four factors and is shown in bold. No solutions of the vector-matrix model were considered physically meaningful.

Number of factors in solution	OOA	m/z 44	m/z 43	HOA	BBOA	LOA	m/z 67, 81, 95	Number of solutions in family	$\Delta Q/Q_{exp}/(Q/Q_{exp})_{min}$	Number of solutions in family	$\Delta Q/Q_{exp}/(Q/Q_{exp})_{min}$
				Factor type				3-vector solutions		Vector-matrix Solutions	
3	X			X	X			45	0		
	X				X	X		5	0.7 %		
		X	X		X					48	0
	X			X		X				2	1.4 %
4		X	X	X	X			35	0		
	X			**X**	**X**	**X**		**12**	**0.1 %**		
	X			X	XX			3	0.4 %		
		X	X		X	X				19	0
		X	X	X	X					29	0.4 %
		X	X	X			X			2	0.7 %
5		X	X	X	X	X		44	0	34	0
		X	X	X	XX			2	0.3 %		
		X	X	XX	X			4	0.4 %		
		X	X	X		X	X			5	0.3 %
		X	X	X	X		X			11	0.7 %

datasets (Allan et al., 2010; DeCarlo et al., 2010), our solutions within the same family show less variation than in the reported 2-D cases. The variation among the solutions in each family is quite small. The average coefficient of variation of the mass spectra, size distributions, or time series of a family in the 3-vector solutions solved by either algorithm is less than 2 %, in the unconstrained vector-matrix model is less than 0.8 %, and in the fully constrained vector-matrix model ($\beta = 0$) is less than 0.05 %.

The existence of multiple families for each number of factors complicates the choice of the best solution; now we must choose the best number of factors and the best family from that set of solutions. Criteria for choosing a family should be the same as for choosing the best number of factors: low Q/Q_{exp}, little structure in the residuals, and the factors must be physically meaningful. The families identify local minima in the Q surface being explored during the iterative minimization. It would be tempting to choose the family with the most solutions; however, the number of solutions in one family may have a stronger relationship to the probability of entering a region of a local minimum in the Q space and not have any intrinsic value. For example, during the iteration, the algorithm may enter a local minimum that has a large "opening" and therefore traps solutions that started from many seeds. Thus we reject a criterion for choosing a family based on the number of solutions in that family.

We now explore the solutions from the 3-vector model, choose the best solution, and identify the factors in this solution.

4.1 Results from the 3-vector model

The 3-vector model was solved by the PMF3 and ME-2 algorithms. The results from both algorithms were similar and are compared in Sect. S4 of the Supplement. Both algorithms found the same families at each number of factors in the solution. Since the results from the two algorithms are similar, we present the ME-2 solutions of the 3-vector case to later compare to solutions of the vector-matrix cases calculated with ME-2.

Because a steep change of slope of Q/Q_{exp} with the addition of each factor was not observed for solutions of the 3-vector model (Fig. 3), we therefore explore all solutions and use the criteria of finding physically meaningful factors
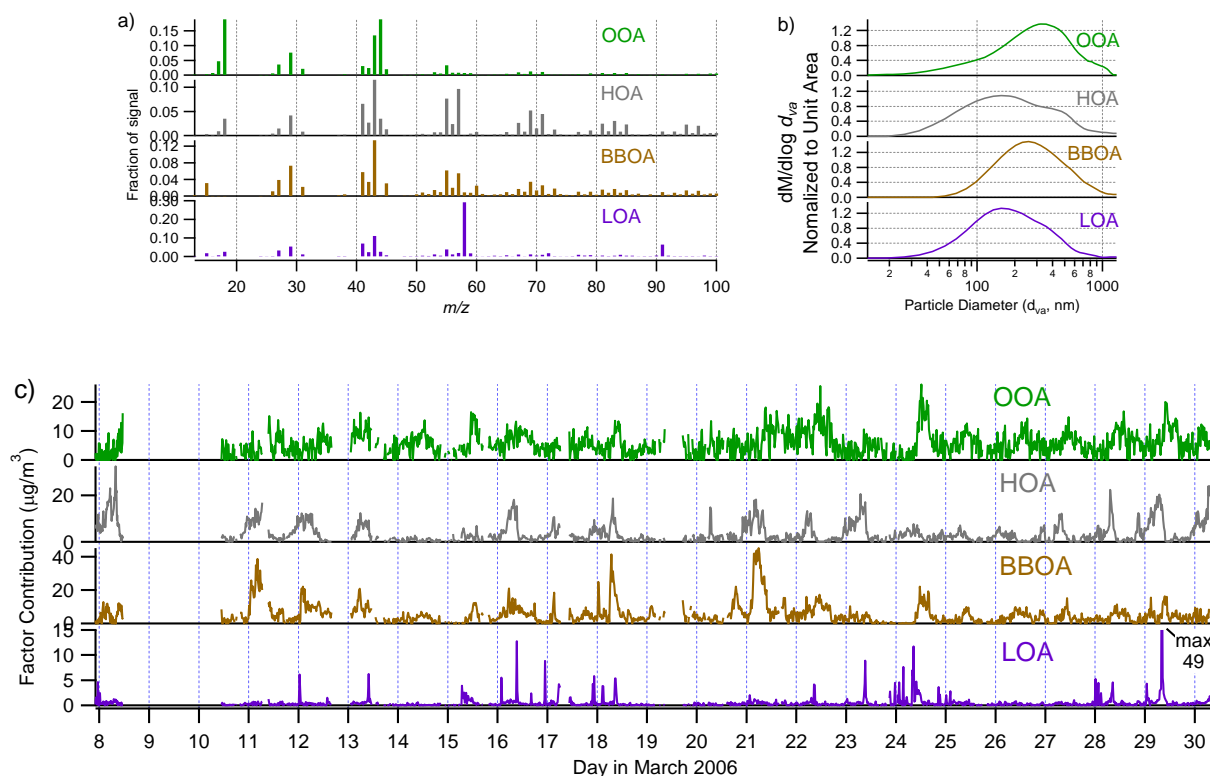
Fig. 4. The best 3-vector solution, which has four factors, solved with ME-2. The four factors are oxidized organic aerosol (OOA), hydrocarbonlike organic aerosol (HOA), biomass-burning organic aerosol (BBOA), and local organic aerosol (LOA). **(a)** Mass spectrum of each factor, normalized to sum to 1. **(b)** Mass size distribution ($dM/d\log d_{va}$), normalized to unit area. **(c)** Mass contribution of each factor. The scale for LOA has been expanded to show the structure during low-concentration periods.

to choose a good solution. Following the assumption that we will likely find the same factors from this dataset as Aiken et al. (2009) reported for the HR-MS data in that 2-D factorization, we expect to need at least four factors in the solution. However, we also explore solutions with more factors in case new factors appear. The choice of the best solution of the 3-vector model is described in Sect. S5 of the Supplement. The best solution of the 3-vector model has four factors, which are the factors from the HR-MS solution: OOA, HOA, BBOA, and LOA (Fig. 4). The features of these factors are discussed and compared to the factors from the vector-matrix model in Sect. 5.1.

4.2 Results from the vector-matrix model

Exploration of the vector-matrix model for this dataset is complicated by poor results in unconstrained solutions of the model. These poor results are manifested in non-physical factor mass spectra (Sect. S6 in the Supplement). To achieve factorization results with more physically meaningful mass spectra, we attempted two methods to constrain the mass spectra using a priori information. The first method was multiple linear regression (i.e., chemical mass balance), but the regression failed for most of the dataset (Sect. S7 in the

Supplement). However, physically meaningful mass spectra were obtained by constraining a priori spectra using the β parameter (Sect. 3.4.1) within the vector-matrix model.

The use of the β parameter with the vector-matrix model requires three additional choices: the number of a priori mass spectra, the source of those mass spectra, and the value of β. We are likely to identify at least four factors (OOA, HOA, BBOA, and LOA) in the dataset, based on the results of the 3-vector factorization. Since the 3-vector factorization had one solution with all four of these factors, we presume that all four factors should also be identifiable with the constrained vector-matrix model. Constraining between one and four factors gives fifteen possible combinations of a priori spectra, but exploring all of these combinations is beyond the scope of the present work. Because the unconstrained vector-matrix solutions produced non-physical factors, we have little confidence that constraining only one or two spectra will improve the solutions significantly. Thus we choose to constrain all four spectra.

Possible sources of the four a priori spectra include the HR-MS solution and the best solution of the 3-vector model, which had four factors. The two sources allow three combinations of a priori spectra: (1) all four spectra from the HR-MS solution, (2) all four from the 3-vector solution, or

(3) some spectra from each of these solutions. We try each of these three options. For the third, mixed-source option, we consider only the case taking OOA, HOA, and BBOA from the HR-MS solution and LOA from the 3-vector solution. We choose this combination since the OOA, HOA, and BBOA mass spectra from the 3-vector solution are very similar to those from the HR-MS solution, but the LOA mass spectrum from the 3-vector solution is less similar to the HR-MS LOA.

To guide the choice of the source of constrained spectra, we compare the Q/Q_{exp} values from the three combinations to the unconstrained vector-matrix solutions (Fig. 5). The constrained cases' higher Q/Q_{exp} values show that the a priori spectra strongly influence the solution. The increased Q/Q_{exp} values suggest that the a priori spectra are not wholly consistent with the mass spectral structure measured in PToF mode. For example, constraining the mass spectra to the HR-MS spectra gives the highest Q/Q_{exp} values of the constrained solutions. Replacing the HR-MS LOA mass spectrum with the LOA spectrum from the 3-vector solution decreases the Q/Q_{exp} value. When the constrained mass spectra come from the best 3-vector solution, the Q/Q_{exp} value is substantially lower than the other constrained solutions, but still somewhat higher than the unconstrained vector-matrix solutions with four factors. Thus, when using the same spectra, the vector-matrix model captures the real variability in the size distribution of each factor better than the 3-vector model. Of these three cases, the Q/Q_{exp} values suggest that the spectra from the 3-vector model give the best fit. But as the constraint on the solutions is relaxed (β is increased), the penalty decreases strongly and the Q/Q_{exp} values are similar for solutions of any of the source spectra.

In addition to the Q/Q_{exp} values of these solutions, our choice for the best a priori factors is also based on the meaningfulness of the factor mass spectra. The HR-MS spectra provide stronger spectral information because they are derived from data with much higher SNR and high-mass-resolution spectra, so we prefer to use the HR-MS spectra to constrain the solution. However, the 3-vector-model LOA spectrum is different from the HR-MS LOA. Although LOA is a small component, it is distinct enough to be resolved from the PToF data. Thus, the combined set of spectra with three factors from the HR-MS solution (OOA, HOA, BBOA) and the LOA spectrum from the 3-vector solution is a good compromise, using mainly HR-MS spectral information and a better starting guess for the most disparate spectrum. We therefore choose the mixed-source spectra as the best a priori information and explore solutions of the vector-matrix model in which we constrain these spectra. In the rest of this work, we only discuss results of the constrained vector-matrix model, unless otherwise noted.
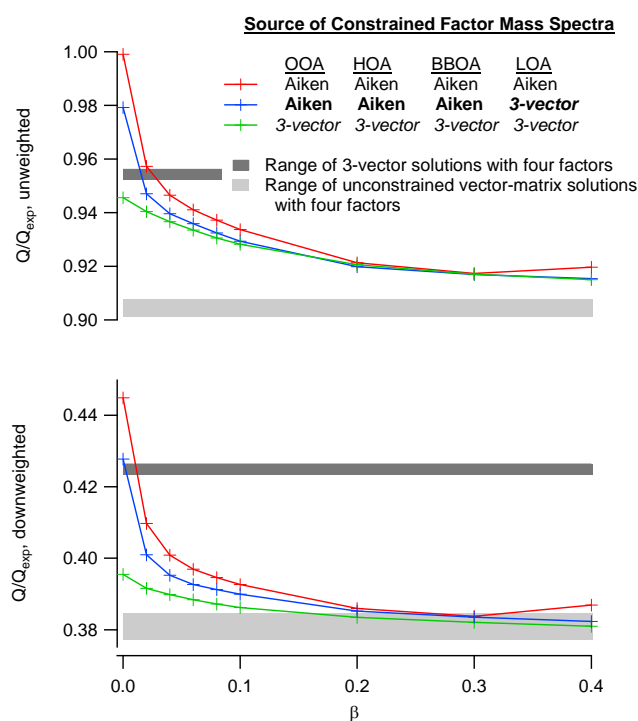

Fig. 5. Values of Q/Q_{exp} for four-factor solutions of the vector-matrix model in which the factor mass spectra have been constrained to a priori spectra vs. β. The β parameter can fully constrain the a priori spectra ($\beta = 0$) or allow them to relax, as described by Eq. (6). The downweighted Q/Q_{exp} values for solutions with $\beta \geq 0.3$ fall within the range of the unconstrained solutions (i.e., solved without a priori information about factor mass spectra) of the same model (lighter grey region). The solutions in which the constrained OOA, HOA, and BBOA spectra come from the high-resolution MS (HR-MS) solution and the LOA spectrum comes from the four-factor solution of the 3-vector model (blue curve) is used for further analysis.

4.2.1 Choosing a solution of the constrained vector-matrix model

To choose a solution of the constrained vector-matrix model, we must answer two questions. First, might we be able to identify additional, physically meaningful factors by constraining the first four factors and then fitting more free factors? Second, to what degree is it appropriate to relax the constraint on the spectra, i.e., to increase β? We first choose the best number of factors while fully constraining four spectra, and then explore the relaxation of the constraint on the four a priori spectra to find the best solution.

The choice of the best number of factors for the constrained vector-matrix model is described Sect. S7 in the Supplement. Four factors give the best fit of the vector-matrix model with physically-meaningful factors. In solutions with more than four factors, the additional factors are nonphysical splits of the HOA and BBOA factors. Therefore
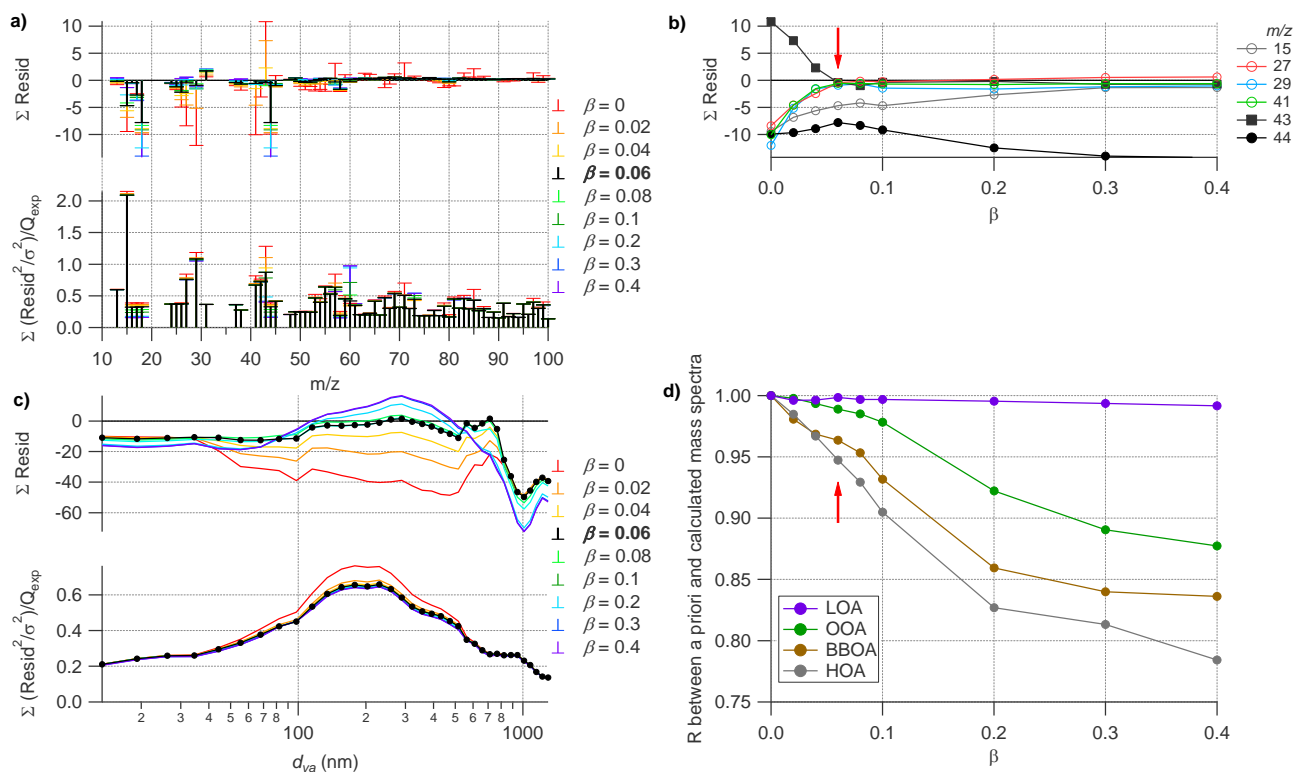
Fig. 6. Residuals and Q/Q_{exp} values for solutions of the constrained vector-matrix model with four factors. **(a)** Total residual (top panel) and Q/Q_{exp} contribution (bottom panel) from each m/z. **(b)** Total residual for selected m/z's as the solution is relaxed with increasing β. **(c)** Total residual (top panl) and Q/Q_{exp} contribution (bottom panel) at each particle size. **(d)** Correlation between the a priori and calculated mass spectra as the constraint is relaxed with increasing β. The solution selected as "best" has $\beta = 0.06$ – black lines in **(a)** and **(c)** and red arrows in **(b)** and **(d)**.

we continue to examine the four-factor solution and explore the effect of relaxing the constraint on the a priori spectra.

We know of no physical or statistical constraints that could be used to predict the optimal degree of relaxation (β) of the constrained factors. However, we posit that starting guesses that better match the dataset might allow smaller values of β, while constrained spectra that are less similar to the true factors would require a larger deviation from the original factors, i.e., larger β, to obtain a good solution. Thus, we take an empirical approach to determining the most appropriate value of β for this dataset. Two options for choosing β are considered here. First, we can compare our factors to a priori information. This was the approach taken by Lanz et al. (2008) when they constrained one mass spectrum in their factorization of a 2-D matrix. However, we have already imposed more a priori information on our solutions by constraining all of the mass spectra. Still, we could compare the time series of our factors to the HR-MS time series or to external tracers. However, the external tracers were already used to support the choice of the HR-MS solution from which we have taken the a priori spectra, so this approach would be somewhat circular. If possible, we would prefer a more independent approach to choosing a solution for this dataset. An

alternative option is to observe the fit residuals, which might indicate an appropriate degree of relaxation.

We examine two types of residuals metrics of the factorizations: the total residuals and the total Q/Q_{exp} contribution summed across two dimensions of the 3-D array, i.e., the total residual and total Q/Q_{exp} as summed to form a time series, mass spectrum, or size distribution. The residual and Q/Q_{exp} contributions summed to a time series are very similar across the range of β, and give no useful information about how to choose a solution. However, the residuals summed to form a mass spectrum show that many of the m/z's ≤ 44 have large negative residuals, i.e., the reconstruction assigns them more signal than is measured (Fig. 6a). But the large negative residuals of selected important m/z's approach zero as the constraint is relaxed; i.e., β is increased (Fig. 6b). In particular, two m/z's show changes that are useful for choosing a solution. The residual of m/z 44 is strongly negative in the fully constrained case, but becomes less negative as β is increased to 0.06. As β is increased further, the residual of m/z 44 becomes more negative. Thus the inflection point at $\beta = 0.06$ marks the best fit for this important marker. In the same solution, the negative residuals for most other m/z's tend toward zero values as β is increased, and

the high residual at m/z 43 is reduced to approximately zero. Based on these trends, the solution at $\beta = 0.06$ may be the best solution of the constrained vector-matrix model.

The choice of the $\beta = 0.06$ solution is supported by the residuals and Q/Q_{exp} contributions summed to a size distribution (Fig. 6c). The residuals summed as size distributions show that solutions with tighter constraints have large negative residuals over the size range of particles with the greatest transmission in the AMS ($d_{va} = 50$–700 nm). However, the residuals across this size region are near zero in the solution with $\beta = 0.06$. As β is increased further, the total residual becomes positive, indicating a worse fit in this region of the data. In addition, the Q/Q_{exp} contribution vs. size does not change as β is increased past 0.06. Thus the residuals and Q/Q_{exp} contributions summed as size distributions confirm our choice of the solution with $\beta = 0.06$ as the best solution of the constrained vector-matrix model.

Finally, we revisit the option of comparing the factor mass spectra to the a priori spectra. The factor mass spectra have not changed dramatically; in fact, the correlations between the factor and a priori spectra in the solution with $\beta = 0.06$ are greater than 0.94 (Fig. 6d). We note that the LOA spectrum changes the least compared to the a priori spectrum from the 3-vector solution, showing that the a priori LOA spectrum was a better representation of the PToF data than the a priori spectra for the other factors. The HOA and BBOA spectra from the HR-MS solution change the most. This change is not surprising, based on the results of the fully constrained solutions with five factors, in which the fifth factor has the characteristic HOA peaks shifted to higher m/z's and is thus described as HOA-like (Sect. S8 in the Supplement). The HOA-like factor implies that the a priori spectra do not completely match the HOA spectrum for the PToF data. But relaxing the constraint in the four-factor solution allows a better fit of the higher m/z peaks. However, the correlation between the solution spectra and the a priori spectra never show a dramatic change that might indicate that the solution has become so relaxed that the spectra are strongly distorted, and therefore do not suggest a particular value of β as an appropriate relaxation of the solution.

4.2.2 Factors in the best solution of the constrained vector-matrix model

The factors from the best solution of the constrained vector-matrix model show the changing size and concentration of the particles in the four factors (Figs. 7–8, S4 in the Supplement). The size distributions in the factor size-distribution–time-series matrices are normalized to unit area for each time step to highlight the shapes of the size distributions (Fig. 7). The contribution to the smallest and largest particle sizes, which have lower transmission into the AMS, is very noisy for all factors. In the middle size range, OOA and BBOA have somewhat narrower size distributions with smaller contributions to particles with $d_{va} < 100$ nm and more mass in

larger particles. In contrast, LOA appears in particle of all sizes, but most of its signal is between 80 and 600 nm. Finally, the HOA size distributions are also broad, and regularly include particles with diameters as small as 50 nm and larger than 700 nm. The HOA size distributions are less noisy than those of the other factors, especially at smaller sizes.

The size distributions of the factors do not usually change quickly, but an interesting exception occurs on 24 March (Fig. 8), when a cold surge brought clean air to the Mexico City basin (Aiken et al., 2009; Molina et al., 2010), decreasing the total submicron organic aerosol concentration to very low levels. This event is similar to a case during a 2003 campaign in Mexico City which has been studied in detail (Dzepina et al., 2009). Although most days in the Mexico City basin have a larger effect of background concentrations and advection, and the dynamic chemical and physical changes that are no doubt occurring in the aerosol (e.g., Hodzic et al., 2010) are less clear when combined with the meteorological effects, these case-study days from 2003 and 2006 are characterized by low initial background concentrations, and lower wind speed and boundary layer, which allow observation of the evolution of the emissions from the city. After midnight on 24 March 2006, the total organic concentration decreased to ~ 3 µg m^{-3}, the lowest concentration measured during the campaign. At 03:30 and 08:00 LT, separate LOA plumes are measured. As the second LOA plume arrives, the HOA concentration begins to increase. An hour later, a dramatic increase in the concentration of OOA begins with the onset of photochemistry. As the OOA concentration increases, the mode of the HOA size distribution shifts from ~ 75 nm to ~ 300 nm over the course of three hours.

HOA growth from smaller particles is not limited to this single event, but is also observed in diurnal average size distributions (Fig. 9). HOA shows the greatest variation in size distribution and concentration, with higher concentrations in the morning, and a shift to larger sizes through the afternoon. Normalizing the size distributions to unit area shows more clearly that HOA particles appear to grow to larger sizes throughout the morning (Fig. 9b). In contrast, the BBOA and OOA size distributions show little variation. In general, OOA has largest mode of all the aerosol components, and higher concentrations later in the day, whereas BBOA has higher concentrations in morning. BBOA particles transported from regional sources may have undergone chemical processing and have probably already grown to these larger sizes. Finally, LOA has the most dramatic changes in size distribution, but its typically low concentrations make the averages noisy and it is difficult to determine whether the average changes actually reflect particle growth.

5 Discussion

In this section we discuss three main points. First, we compare the best solutions of the 3-vector and vector-matrix
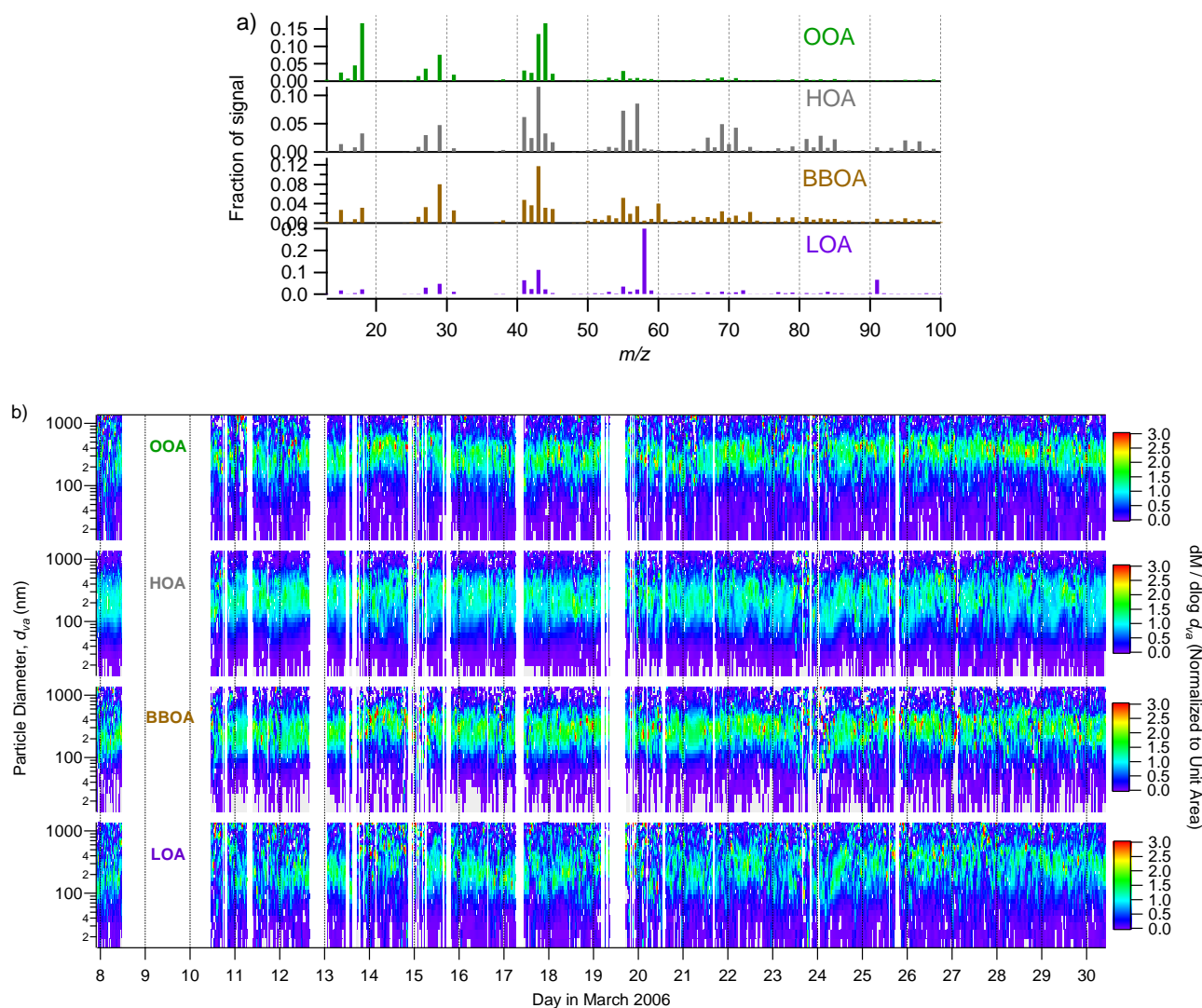
Fig. 7. The best constrained vector-matrix solution, which has four factors. Four a priori mass spectra were provided as starting guesses: OOA, HOA, and BBOA from the HR-MS solution, and LOA from the best solution of the 3-vector model (Fig. 4). The a priori spectra were allowed to vary with $\beta = 0.06$, as described in Eq. (6). **(a)** Mass spectrum of each factor, normalized to sum to 1. **(b)** Mass size distribution ($\mathrm{d}M/\mathrm{dlog}d_{va}$), normalized so that each size distribution has unit area. The factor size distributions have been binomially smoothed by one point each in time and size for improved visual clarity. Light-grey pixels have zero signal.

models to the HR-MS solution and a tracer method for estimating factor size distributions. Second, we discuss the insights gained from the size distributions of the factors and the potential effect of some details of particle vaporization and bounce on PToF sampling and the resulting factors. Finally, we discuss directions for future research on the application of 3-D factorization models to PToF and other datasets.

5.1 Evaluation of the assumptions of the 3-vector model

We compare the best solutions of the 3-vector and vector-matrix models to answer two related questions. First, is the 3-vector model appropriate for this dataset? Second, is one of these models better than the other for this dataset?

We first consider the appropriateness of the 3-vector model. Pere-Trepat et al. (2007) noted that in their size-resolved aerosol composition dataset, "it was found that there is sufficient size dependence in the [factor] compositions that the strict trilinear model does not hold." However, these authors provide no details to explain how they determined that the 3-vector method was inappropriate for their data. We seek such indicators in the solution Q/Q_{exp} values and residuals. Overall, Q/Q_{exp} values show that the best solution of the constrained vector-matrix model fits the data better than the best 3-vector solution, although the differences are small (0.94 vs. 0.95, respectively; Fig. 5). Furthermore, Q/Q_{exp} summed to size distributions is similar between the
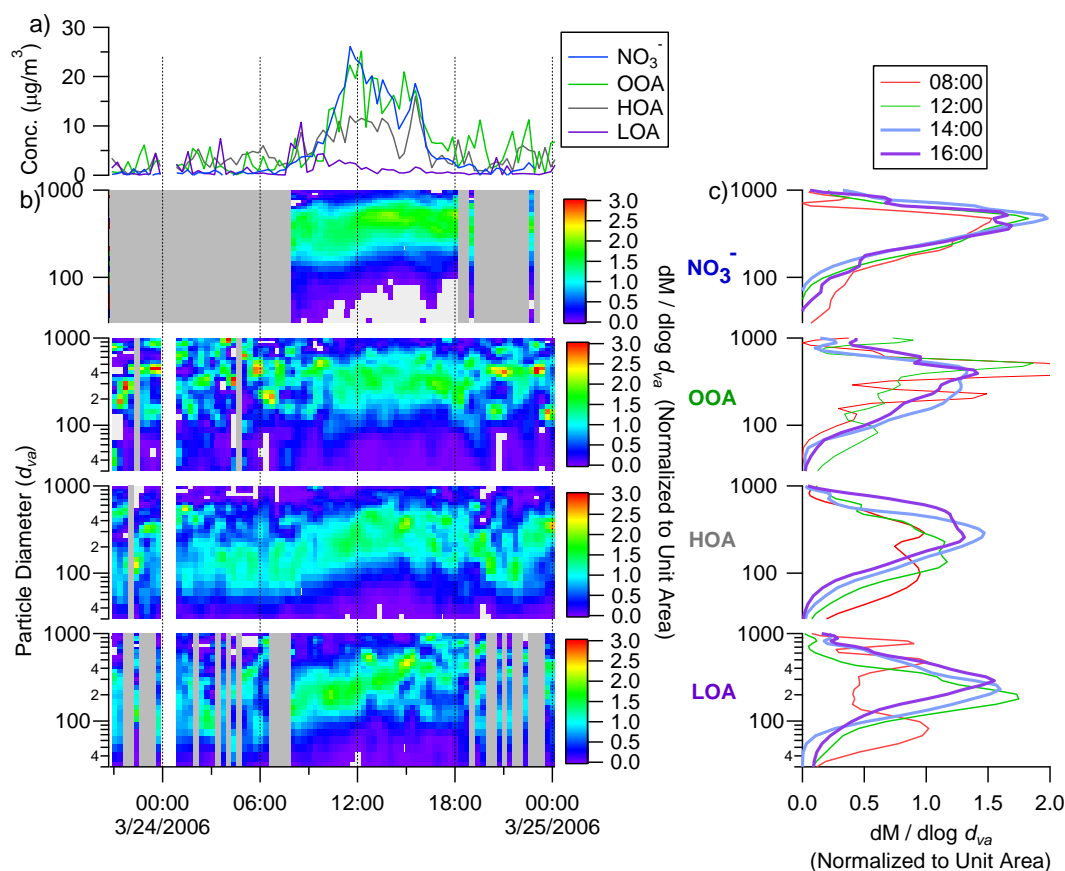
Fig. 8. Case study event on 24 March 2006. **(a)** Time series (LT) of mass concentrations of OOA, HOA, and LOA in the best solution of the vector-matrix model and NO_3^-. **(b)** Mass size distributions ($dM/dlog d_{va}$) of OOA, HOA, and LOA and NO_3^-, normalized so that each size distribution has unit area. The factor size distributions have been binomially smoothed by one point each in time and size for improved visual clarity. Light-grey pixels have zero signal. Medium grey bars indicate time steps with integrated species concentration less than $0.3 \,\mu g \, m^{-3}$ for organics and $1.5 \,\mu g \, m^{-3}$ for nitrate. **(c)** Size distributions from selected hours (LT), normalized to unit area to compare the shapes of the distributions.

two solutions. In contrast, examination of the residuals of these solutions shows that the constrained vector-matrix solution provides a better fit, but does not provide evidence that the assumptions of the 3-vector model fail for this dataset. The major difference in the residuals of the two solutions is seen by comparing the total residuals summed to form a size distribution. While the residual size distribution of the constrained vector-matrix solution is near zero or negative (Fig. 6), that of the 3-vector model is almost always positive, with larger residuals toward the larger sizes. The fact that the residual is always positive indicates that there are some real variations in the component size distributions that cannot be fit with the 3-vector model. Thus the residuals are distributed differently in the solutions of the two models. Yet, we have found no distinct characteristic of the residuals that indicates that one of the models is better or worse for this dataset.

We next compare the solutions from the 3-vector and vector-matrix models to understand the differences between these solutions. First, we compare the factor mass spectra

(Fig. 10a). The mass spectra of OOA, HOA, and BBOA are similar enough to each other to identify them as the same aerosol component by eye. The variation amongst the mass spectra of the same component from the HR-MS solution and the solutions of the 3-vector and vector-matrix models are within the observed variation for OOA, HOA, and BBOA factors in other urban datasets (Ng et al., 2011). However, the LOA mass spectrum shows a substantially different fraction of m/z 58 between the 3-D and HR-MS LOAs. Even with the high contribution of m/z 58 to the LOA spectra, the 3-D LOA spectra still retain the distinctive enhancement at m/z 91 and, to a lesser extent, characteristic HOA peaks that dominate the shape of the HR-MS LOA spectrum. Overall, the factor mass spectra are similar between the two 3-D factorization cases, and do not provide a clear reason to choose one solution over the other.

Second, we compare the factor time series in the 3-vector and vector-matrix solutions (Fig. 10c). The time series of OOA between the 3-vector and vector-matrix solutions, and
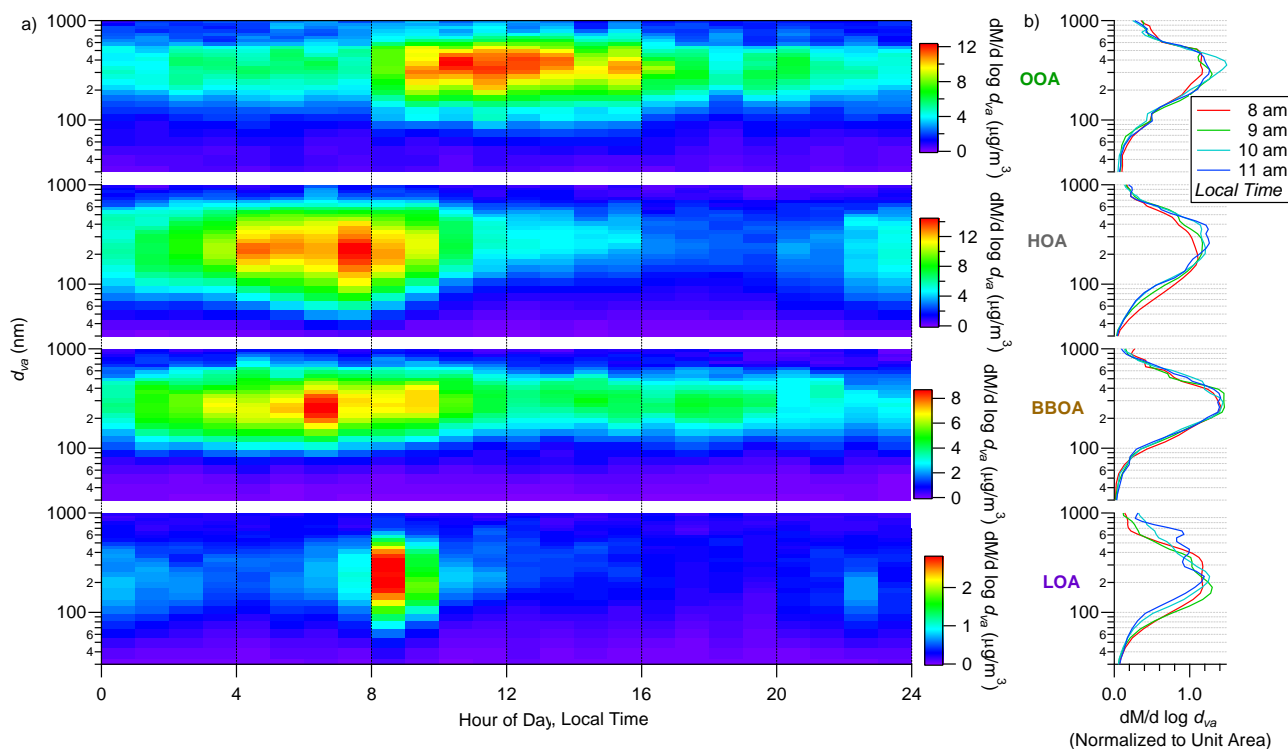
Fig. 9. Diurnal average size distributions of OOA, HOA, BBOA, and LOA from the best solution of the constrained vector-matrix model. **(a)** Image plots of diurnal average size distributions plotted ($dM/d\log d_{va}$), not normalized to unit area. **(b)** Size distributions from selected hours, normalized to unit area to compare the shapes of the distributions.

of LOA between the same solutions, have very high correlations ($R > 0.96$). In contrast, the HOA time series and BBOA time series are less similar between the 3-vector and vector-matrix solutions ($R > 0.89$). For the HOA and BBOA time series, we observe several instances in which mass is "traded" between these factors. We explain trading, which is also observed in 2-D PMF solutions, with an example. On 16 and 17 March, the 3-vector model attributes more mass to BBOA than does the vector-matrix model, but less mass to HOA than the vector-matrix model (Fig. 10c). Thus the HOA and BBOA factors have "traded mass" over this short time period. The trading of mass is the main difference between the HOA and BBOA time series for the two solutions, though in both models the general trends for the time series of both factors are very similar. HOA and BBOA have similar unit-resolution mass spectra, making them hard to separate in some datasets (Lanz et al., 2008); this similarity likely contributes to the trading in these factors' time series. Aiken et al. (2009) reported that factoring high-resolution spectra (i.e., the HR-MS dataset) substantially improved the separation of these two factors compared to factorization of the unit-mass-resolution spectral data. The improvement is the result of separating, before performing the factorization, the information from HR ions at the same nominal m/z that have different contributions to HOA and BBOA. For example, HOA and BBOA have similar contributions from m/z 57 in

the unit resolution mass spectra, but the HR-MS factors show that m/z 57 in HOA is almost exclusively from $C_4H_9^+$, while $C_3H_5O^+$ contributes a major fraction of the m/z 57 signal for BBOA. However, the high-resolution size distributions are not available, and thus these two factors are harder to separate. Nevertheless, the time series from our factorization of the unit-resolution PToF data with the vector-matrix model track the HR-MS time series more closely than do the time series from the 3-vector model.

We notice other interesting trends when comparing the time series from the 3-D solutions to those of the HR-MS solution. First, both 3-D models attribute more mass to BBOA than the HR-MS solution in the latter third of the campaign. Interestingly, this is the "low fire period" of the campaign, during which several non-AMS fire tracers and models indicate that fire activity was much reduced (Aiken et al., 2010). But comparison of the 3-D and HR-MS solutions shows that this BBOA signal is trading with the OOA signal. Because we believe, based on the reasons discussed above, that the HR-MS factorization should give more accurate results, the calculated BBOA during this period is likely an artifact of the factorization. Second, the OOA time series from the 3-D solutions are very noisy. However, binomial smoothing of the 3-D OOA time series by one point improves the correlation with the HR-MS OOA from $R = 0.68$ to $R = 0.81$. The noise in the OOA factors from the 3-D solutions is likely caused
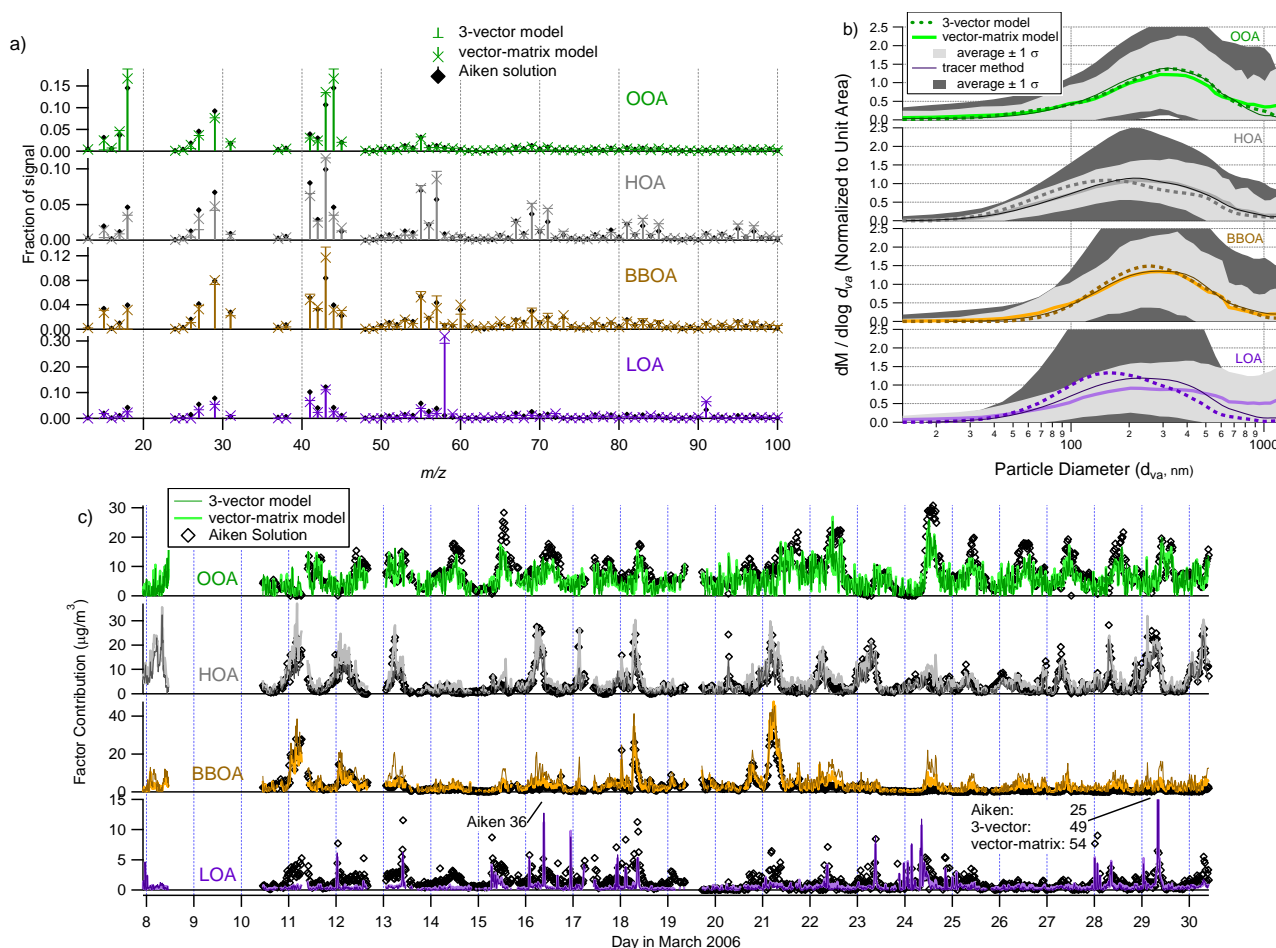
Fig. 10. Comparison of the factors from the best solutions of the 3-vector and vector-matrix models and the HR-MS solution for **(a)** mass spectra, **(b)** average size distributions, and **(c)** time series. The grey shaded region in **(b)** denotes ±1 standard deviation of the variation in time of the size distribution in the vector-matrix model. Standard deviations for the family of 3-vector solutions are less than 2 % and are not shown.

by the low SNR of m/z 44 in PToF mode. Third, there is only moderate correlation between the 3-D and HR-MS LOA time series ($R = 0.66$ and 0.62 for the 3-vector and vector-matrix solutions, respectively). The 3-D LOA factors are assigned less than half the average mass contribution of the HR-MS solution (0.52, 0.65, and 1.39 µg m^{-3} for the 3-vector, vector-matrix, and HR-MS solutions, respectively). However, the 3-D solution mass spectra for LOA have less contribution from the characteristic HOA peaks, and that mass may have been traded to the HOA factor. Overall, the results from the 3-D solutions are consistent with the general characteristics of the HR-MS solution.

Third, we can make two comparisons of the size distributions from the two 3-D factorizations. First, we can compare the static size distributions from the 3-vector model with the average size distributions from the vector-matrix model. Second, we can compare the dynamic diurnal average size distributions from the vector-matrix to size distributions estimated

from tracer m/z's (Zhang et al., 2005; Cubison et al., 2008; Wang et al., 2010).

The comparison of the average size distributions shows some differences between the 3-vector and vector-matrix models (Fig. 10b). The size distributions from the two models are very similar for OOA and BBOA, consistent with the constancy of their normalized size distributions observed in Fig. 9. In contrast, the HOA and LOA size distributions show more differences between the two models. The HOA size distribution from the vector-matrix model is shifted to higher particle sizes compared to the 3-vector model. The LOA size distributions are even more different between the two models. The average LOA size distribution from the vector-matrix model is strongly shifted to larger particles compared to the distribution from the 3-vector model, with about a third less particle mass between 50 and 300 nm in the vector-matrix solution. However, the size distributions from the 3-vector solution are within one standard deviation
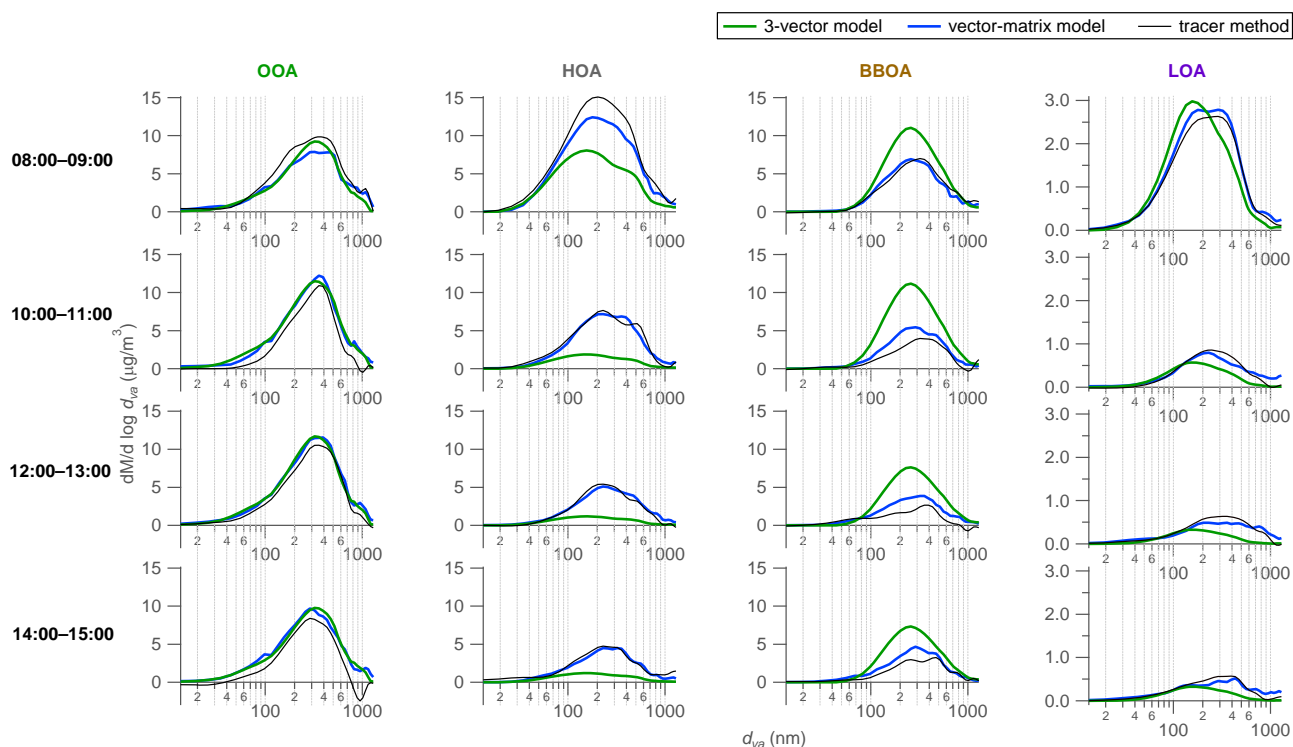
Fig. 11. Comparison of average size distributions as calculated from the 3-vector model, vector-matrix model, and tracer method for OOA, HOA, BBOA, and LOA. The evolution of the size distribution of one aerosol type over all days from 08:00–14:00 LT is shown in each column. The size distribution of each aerosol type at the same time is shown in each row. Tracers for each species are estimated from characteristic *m/z*'s as described in the text.

of the average from the vector-matrix model for all factors. Thus it appears that the 3-vector model is not fitting variations in the real size distributions that can be captured by the vector-matrix model.

Now we compare the dynamic estimates of the size distributions from the 3-vector solution, vector-matrix solution, and tracer methods. The size distributions of *m/z*'s 44, 57, 58, and 60 are used as tracers of OOA, HOA, LOA, and BBOA, respectively. The tracer size distribution smoothed in the same manner as the factorization data array (Sect. 3.3.2, Step 3) and scaled by the slope of the regression line between the MS-mode and PToF-mode time series for that *m/z*. Then an estimation formula between the tracer and the component is applied. In this work, we use tracer-component relationships derived from MS-mode data for OOA from Ng et al. (2011) and for HOA and BBOA from Aiken et al. (2009). The tracer-component relationship for LOA is estimated as the size distribution of *m/z* 58/ f_{58}, where f_{58} is the fraction of *m/z* 58 in the 3-D LOA factor mass spectrum. Size distributions from selected daytime hours for the four factors are shown in Fig. 11 and are not scaled to unit area. (Comparison of all 24 h can be viewed as movies in the Supplement.)

The dynamic OOA size distributions are quite similar for the two 3-D models. The OOA size distributions from the

tracer method are also similar to those from the 3-D models, but are generally narrower than the 3-D model size distributions. In addition, the tracer method OOA size distribution sometimes has negative concentrations for particles with diameters greater than 800 nm. Thus noise in the *m/z* 44 size distribution is carried through the tracer method, whereas this noise is tempered in the 3-D models by the contribution of other *m/z*'s present in the OOA mass spectrum.

In contrast, the HOA and BBOA size distributions in the 3-vector and vector-matrix models exhibit trading of mass between these two factors, as discussed above. HOA always has a larger mass contribution and BBOA a smaller mass contribution in the vector-matrix model. The vector-matrix model fits real variation in the size distributions and was determined in the discussion above to be more consistent with the HR-MS solution, and thus more likely to be correct. In addition to the differences in mass contribution, the modes of the size distributions differ between the models. The size distributions for both HOA and BBOA are shifted to larger particle diameters in the vector-matrix model, as discussed above. For HOA, the tracer size distribution usually matches the vector-matrix distribution well, though the tracer distribution has more mass at 08:00 LT than either of the 3-D models. Like the tracer HOA size distribution, the tracer BBOA size

distribution generally follows the vector-matrix size distribution, but has a slightly smaller mass concentration. The size distributions from the tracer method sometimes show negative concentrations for large particles, but less severely than for OOA.

Finally, the LOA size distributions from the 3-vector and vector-matrix models agree well at small particle sizes, but diverge for larger particle sizes. The vector-matrix model always attributes more LOA mass to larger particles than does the 3-vector model. This trend is more important in terms of the fraction of the mass in large particles as the day progresses.

In conclusion, it appears that both 3-D methods can represent much of the real variation in the time series and mass spectra of the components. While the 3-vector model captures the main features in the size distributions reasonably well and may be useful for simplified analysis and initial explorations of datasets, it cannot identify particle growth events that are resolved by the vector-matrix model. The vector-matrix model appears to be superior for more accurately capturing the factor concentrations (as evidenced by comparison with the HR-MS solution) and for identifying real variations of the size distributions. However, the full advantage of the vector-matrix model may not be apparent with this dataset because it includes only a few strong periods of rapid changes in the size distributions. Only the OOA growth event on 24 March 2006 shows fast changes in the HOA size distribution (Figs. 8, S5 in the Supplement). Furthermore, the appearance of LOA only in brief spikes makes it harder to understand this small component. Nevertheless, we prefer the vector-matrix model for exploring the size distributions of AMS components. Finally, the tracer method appears to provide a reasonable estimate of the size distributions for this dataset.

5.2 Insights into ambient aerosol and PToF sampling

The size distributions of OOA, HOA, and BBOA are consistent with previous interpretations of these factors in Mexico City and other studies. OOA has the largest mode diameter of $\sim 400\,\text{nm}$, and an asymmetric distribution with strong extension to smaller particles. This shape is consistent with the condensation of semivolatile gas-phase molecules onto smaller particles (Zhang et al., 2005). BBOA has a slightly smaller mode at $\sim 300\,\text{nm}$. Although BBOA is generated as primary aerosol from a combustion process, BBOA has a smaller fraction of ultrafine ($d_{\text{va}} < 100\,\text{nm}$) particles than OOA (McMeeking et al., 2005; Levin et al., 2010). In contrast, the size distribution of HOA is quite broad, with the largest fraction of ultrafine particles of any of these factors. The ultrafine fraction is consistent with the primary nature of this combustion source, and with past measurements of freshly emitted vehicle-exhaust particles. Fresh exhaust particles have a mode at $d_{\text{va}} \sim 100\,\text{nm}$, and a second mode with $d_{\text{va}} \sim 500\,\text{nm}$ is sometimes present (Canagaratna

et al., 2004). Thus the broad HOA size distribution found here may not represent only fresh HOA. A similar, broad distribution of HOA particles was observed at an urban Pittsburgh sampling location (Zhang et al., 2005). The broad size distributions could be explained by emission of many larger HOA particles and/or the growth of HOA particles as they are coated by condensation of secondary organic and inorganic species onto existing particle surface area during the day. For example, during the large OOA growth event on 24 March, the HOA size distribution grows as the OOA concentration increases (Sect. 4.2.2, Fig. 8). During this event, it is unlikely that the HOA particles remain externally mixed with other particles; rather, the HOA particles are most likely coated by OOA and secondary inorganics (ammonium nitrate and sulfate), and thus have an HOA core and coating of secondary material. When these particles are sampled in the AMS, they have mixed HOA-OOA mass spectra and are recorded at the particle's coated size. Therefore the HOA size distribution grows to larger particle sizes, even though each particle has a mixed composition. Thus, particle size distributions must be interpreted as showing the size distribution of particles *containing* an aerosol component, and not necessarily as the size distribution of externally mixed particles of any single component.

The size distribution of LOA gives some new insight into the source of this component previously identified by Aiken et al. (2009). These authors identified this factor as local based on its spiky time series, which correlates with single-particle measurements of nitrogen-containing organic carbon particles and lead-zinc-containing particles that appear to come from local industrial sources (Moffet et al., 2008). The HR mass spectrum of LOA includes N-containing peaks characteristic of aliphatic amines (m/z's 58 and 86) and stable $C_6H_5C_nH_{2n}^+$ ions characteristic of phenylalkyl compounds (m/z's 91 and 105, McLafferty and Turecek, 1993). The size distribution of LOA on the morning of 24 March, when LOA briefly has a large fraction of the organic mass is bimodal with modes at $d_{\text{va}} \sim 100\,\text{nm}$ and $\sim 270\,\text{nm}$. Since LOA appears to come from an individual, local source, it may have a more variable size distribution than, e.g., HOA, which represents the average size distribution of millions of vehicles and other combustion sources. Nitrogen-containing factors have been identified in some HR-AMS-PMF analyses of urban datasets (Docherty et al., 2011; Sun et al., 2011) and their direct sources are also uncertain.

Finally, we assess the calculated errors for the PToF data. The unweighted Q/Q_{exp} values of the solutions near 1 (Fig. 3) suggest that the errors have about the right magnitude. However, Q/Q_{exp} near 1 is not sufficient evidence that the error values are correct. Because 34 % of the data array elements have values less than 0, which cannot be fit with the positivity constraint, these elements must contribute to Q/Q_{exp}. Even if every positive array element is fit within its estimated uncertainty and contributes Q/Q_{exp} of ~ 1, the negative array elements must contribute additionally

to Q/Q_{exp}. Thus our Q/Q_{exp} values appear to be somewhat low, and the error estimates for the PToF data may be slightly overestimated.

5.3 Fitting of variations in mass spectra and size distributions

We explained in Sect. 3.4.1 that atmospheric processing of aerosol could change the mass spectrum, and if the processing resulted in slower particle vaporization, the apparent size distribution of OA components. We now discuss whether such changes were fit as additional factors in the solutions of either factorization model. The most likely place to find such additional factors is in solutions of the 3-vector model, whose requirements of constant mass spectrum and time series for each factor restrict this model's ability to fit variations. Only a few solutions of the 3-vector model included a pair of factors with similar mass spectra and time series but different size distributions, indicating a mathematical rather than physically meaningful separation.

In Sect. 3.4.1 we also noted a 2-D factorization case in which the time series of the residuals and Q/Q_{exp} contributions were highly correlated with the time series of the semi-volatile OOA-II factor, suggesting that there were changes in the composition of OOA-II that could not be fit with additional factors (Ulbrich et al., 2009). However, no such correlations were observed in the factorization of this dataset, nor did we observe other indicators that spectral variation was significant enough to interfere with the factorization using these models.

5.4 Directions for future research

In this section we first discuss the application of 3-D factorizations to PToF datasets from ToF-AMS instruments and our recommendations for applying these techniques. Finally, we suggest potential applications for 3-D factorization models to AMS instruments currently in development.

First, we reiterate that the analysis presented here has used the size distributions of unit-mass-resolution m/z's. The unit-mass-resolution data in this study hindered separation of HOA, BBOA, and LOA because these factors have contributions from a common series of m/z's. In contrast, the Aiken et al. (2009) study factored higher-resolution data, in which ions at each nominal m/z can be separated into the contrasting time series of individual ions. The time series of more ions might allow the separation of additional factors if the factorization were performed on the size distribution of high-resolution fragments. However, constructing that factorization array would require fitting high-resolution ions for each size-resolved mass spectrum, which is not yet part of the standard HR-AMS data analysis software (PIKA) and is a major project by itself. Nevertheless, this analysis might be possible with high-SNR datasets from locations with sufficient aerosol mass concentrations. Increasing the SNR of

the PToF data by reducing the range of m/z's sampled (De-Carlo et al., 2006), using an aerosol concentrator (Khlystov et al., 2005), or applying these techniques to C-ToF-AMS data (which has ~4 times more signal than the V-mode of the HR-ToF-AMS, DeCarlo et al., 2006) should also help improve factorization results.

We also reiterate that the two vector-matrix models not used in the present study (Fig. 1c and d) could be applied to size-resolved chemical-composition datasets to explore different questions than we have considered here. The vector-matrix model in which the vector contains a factor size distribution and the matrix shows how the chemical composition of that characteristic size distribution changes with time (Fig. 1c) would likely identify modes of submicron aerosol and the sources and processes affecting the those aerosol (Alfarra et al., 2004; Zhang et al., 2005). In contrast, the vector-matrix model in which the vector contains a factor time series and the matrix shows the size dependence of that factor's chemical composition (Fig. 1d) assumes that particles from a single source arrive at the receptor together and may have different chemical composition at each size, but the size-composition relationship does not change with time (Pere-Trepat et al., 2007). Application of these two vector-matrix models to size-resolved chemical-composition datasets is of interest and would allow quantitative evaluation of their appropriateness for describing these data.

We make two recommendations to researchers who wish to factor size-resolved AMS datasets. First, we suggest beginning with 2-D factorization of the high-resolution, mass spectral mode data. In our case, the factors from the HR-MS solution were critical for diagnosing the initial unsatisfactory solutions of the vector-matrix model. In addition, the 2-D factors were also useful for confirming when the 3-D factors had split and were not physically meaningful. Second, we recommend exploring the 3-vector model and at least one of the vector-matrix models shown in Fig. 1. Comparison of the results of two models enables the exploration of the appropriateness of each model's assumptions. Finally, we do not recommend constraining the factors in any model unless the unconstrained solutions are not useful. However, we recommend that researchers who constrain factors using a parameter such as α or β carefully examine residuals of the fit and other available metrics to choose an appropriate degree of relaxation for their datasets.

In addition to their application to size-resolved chemical composition datasets, these 3-D factorization models have the potential to analyze the structure of other 3-D datasets. The 3-vector and vector-matrix models can be applied to any appropriate 3-D data array for which the model assumptions are appropriate, not just the specific data type described here. For example, chemically resolved thermal-desorption datasets are inherently three-dimensional. The 3-D factorizations described in this work could be applied to thermal denuder-AMS datasets (previously factored using 2-D methods, Huffman et al., 2009), or data from any

of several thermal-desorption mass spectrometers, including the thermal-desorption particle-beam mass spectrometer (TDPBMS, Tobias et al., 2000), thermal-desorption chemical-ionization mass spectrometer (TD-CIMS, Smith and Rathbone, 2008), or the micro-orifice volatilization impactor coupled to a CIMS (MOVI-CIMS, Yatavelli and Thornton, 2010). As a second example, the themal-desorption aerosol GC/MS-FID, or TAG, uses chromatography to separate organic compounds from thermally desorbed ambient aerosol, also forming an inherently 3-D dataset (Williams et al., 2006, 2010). In the TAG dataset, the majority of the signal is present as an "unresolved complex mixture" which has not yet been analyzed in detail (Williams et al., 2010). In each of these cases, the third dimension of the data (thermal desorption temperature or chromatographic retention time) is expected to provide information distinct from the bulk mass spectra, and we expect that separation of additional factors should be possible from these datasets. Application of 3-D models to new datasets requires careful construction of error estimates and appropriate choice of models that match the underlying structure of that data.

6   Conclusions

We have applied two 3-D factorization models to three weeks of continuous HR-ToF-AMS size-resolved organic aerosol composition data from Mexico City. In the 3-vector model (Fig. 1a), each factor is composed of a characteristic chemical composition (mass spectrum), a characteristic size distribution, and the time series of the mass concentration of that component. In this model, the mass spectrum and size distribution are constant over the course of the measurements. In contrast, in the vector-matrix model in which the vector is a mass spectrum (Fig. 1b), the matrix shows the changing size distribution of that chemical component with time. The vector-matrix model has more degrees of freedom than the 3-vector model; the additional freedom provides greater ability to fit the dynamic nature of the data, but also to be distorted by noise in the dataset. Preparation of the data for factorization required a method for estimating the precision of the measured data, developed here for the first time. Noise hampered initial results of the vector-matrix model, but physically meaningful factors were obtained after partially constraining the mass spectra using a priori information and a new regularization parameter. For this dataset, four factors were identified that were consistent with factors obtained by Aiken et al. (2009) from factorization of the bulk (i.e., not size-resolved) HR-MS organic measurements from the same instrument. These factors represent oxidized organic aerosol (OOA), hydrocarbon-like organic aerosol (HOA), biomass-burning organic aerosol (BBOA) and a locally emitted organic aerosol (LOA). However, the mass spectra of these factors are not identical to those from the HR-MS solution. These differences may be due to noise and/or a fraction of

slowly vaporizing compounds whose ions are averaged into the total signal in MS mode but are recorded as part of the background in PToF mode.

The results of the vector-matrix model show diurnal cycles in the size distribution of HOA and suggest growth by condensation of secondary species onto pre-existing HOA particles, especially during an OOA growth event on 24 March 2006. The size distributions of HOA and BBOA are consistent with source size distributions, and the size distribution of OOA is consistent with that of more aged secondary aerosols. The size distribution of LOA supports the interpretation that this component is locally produced, probably by industrial combustion. In addition, these size distributions are less noisy and likely more robust than those obtained previously by tracer methods, and could be used for future cloud condensation nuclei (CCN) and hygroscopicity studies (Cubison et al., 2008; Gunthe et al., 2009; Wang et al., 2010).

The vector-matrix model appears to capture real variability in the size distributions that cannot be captured in the 3-vector model. While the 3-vector model captures the main modes in the size distributions reasonably well, it cannot identify particle growth events that are resolved by the vector-matrix method. The tracer m/z-based method provides a useful approximation for the component size distributions in this study. We suggest that others who apply 3-D factorization techniques first factor a 2-D version of the HR-MS data to understand the main trends in the dataset and as a basis for understanding factors from 3-D model solutions.

Other versions of the vector-matrix model are possible and could be applied to this type of dataset to explore other questions about aerosol evolution. Finally, these techniques, with appropriate error estimates and choice of 3-D model, can be applied to other 3-D datasets, especially those obtained by measuring thermal desorption aerosol mass spectra or chromatographically resolved aerosol composition.

Supplementary material related to this article is available online at: http://www.atmos-meas-tech.net/5/195/2012/ amt-5-195-2012-supplement.zip.

References

Aiken, A. C., Decarlo, P. F., Kroll, J. H., Worsnop, D. R., Huffman, J. A., Docherty, K. S., Ulbrich, I. M., Mohr, C., Kimmel, J. R., Sueper, D., Sun, Y., Zhang, Q., Trimborn, A., Northway, M., Ziemann, P. J., Canagaratna, M. R., Onasch, T. B., Alfarra, M. R., Prevot, A. S. H., Dommen, J., Duplissy, J., Metzger, A., Baltensperger, U., and 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., 42, 4478–4485, doi:10.1021/es703009q, 2008.

Aiken, A. C., Salcedo, D., Cubison, M. J., Huffman, J. A., DeCarlo, P. F., Ulbrich, I. M., Docherty, K. S., Sueper, D., Kimmel, J. R., Worsnop, D. R., Trimborn, A., Northway, M., Stone, E. A., Schauer, J. J., Volkamer, R. M., Fortner, E., de Foy, B., Wang, J., Laskin, A., Shutthanandan, V., Zheng, J., Zhang, R., Gaffney, J., Marley, N. A., Paredes-Miranda, G., Arnott, W. P., Molina, L. T., Sosa, G., and Jimenez, J. L.: Mexico City aerosol analysis during MILAGRO using high resolution aerosol mass spectrometry at the urban supersite (T0) – Part 1: Fine particle composition and organic source apportionment, Atmos. Chem. Phys., 9, 6633–6653, doi:10.5194/acp-9-6633-2009, 2009.

Aiken, A. C., de Foy, B., Wiedinmyer, C., DeCarlo, P. F., Ulbrich, I. M., Wehrli, M. N., Szidat, S., Prevot, A. S. H., Noda, J., Wacker, L., Volkamer, R., Fortner, E., Wang, J., Laskin, A., Shutthanandan, V., Zheng, J., Zhang, R., Paredes-Miranda, G., Arnott, W. P., Molina, L. T., Sosa, G., Querol, X., and Jimenez, J. L.: Mexico city aerosol analysis during MILAGRO using high resolution aerosol mass spectrometry at the urban supersite (T0) – Part 2: Analysis of the biomass burning contribution and the non-fossil carbon fraction, Atmos. Chem. Phys., 10, 5315–5341, doi:10.5194/acp-10-5315-2010, 2010.

Alfarra, M. R., Coe, H., Allan, J. D., Bower, K. N., Boudries, H., Canagaratna, M. R., Jimenez, J. L., Jayne, J. T., Garforth, A. A., Li, S. M., and Worsnop, D. R.: Characterization of Urban and Rural Organic Particulate in the Lower Fraser Valley Using Two Aerodyne Aerosol Mass Spectrometers, Atmos. Environ., 38, 5745–5758, doi:10.1016/j.atmosenv.2004.01.054, 2004.

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:10.1029/2002JD002358, 2003.

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 Worsnopf, 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, doi:10.1016/j.jaerosci.2004.02.007, 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., Pandolfi, M., Escrig, A., Querol, X., Alastuey, A., Pey, J., Perez, N., and Hopke, P. K.: Quantifying Road Dust Resuspension in Urban Environment by Multilinear Engine: A Comparison with PMF2, Atmos. Environ., 43, 2770–2780, doi:10.1016/j.atmosenv.2009.02.039, 2009.

Andreae, M. O. and Rosenfeld, D.: Aerosol-Cloud-Precipitation Interactions. Part 1. The Nature and Sources of Cloud-Active Aerosols, Earth-Sci. Rev., 89, 13–41, doi:10.1016/j.earscirev.2008.03.001, 2008.

Blifford, I. H. and Meeker, G. O.: A Factor Analysis Model of Large Scale Pollution, Atmos. Environ., 1, 147–157, 1967.

Cahill, T. A., Feeney, P. J., and Eldred, R. A.: Size Time Composition Profile of Aerosols Using the DRUM Sampler, Nucl. Instrum. Meth. B, 22, 344–348, 1987.

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. Tech., 38, 555–573, doi:10.1080/02786820490465504, 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.20115, 2007.

Canagaratna, M. R., Onasch, T. B., Wood, E. C., Herndon, S. C., Jayne, J. T., Cross, E. S., Miake-Lye, R. C., Kolb, C. E., and Worsnop, D. R.: Evolution of Vehicle Exhaust Particles in the Atmosphere, J. Air Waste Manage. Assoc., 60, 1192–1203, doi:10.3155/1047-3289.60.10.1192, 2010.

Chow, J. C. and Watson, J. G.: Review of $PM_{2.5}$ and PM_{10} Apportionment for Fossil Fuel Combustion and Other Sources by the Chemical Mass Balance Receptor Model, Energ. Fuel., 16, 222–260, doi:10.1021/ef0101715, 2002.

Cross, E. S., Onasch, T. B., Canagaratna, M., Jayne, J. T., Kimmel, J., Yu, X.-Y., Alexander, M. L., Worsnop, D. R., and Davidovits, P.: Single particle characterization using a light scattering module coupled to a time-of-flight aerosol mass spectrometer, Atmos. Chem. Phys., 9, 7769–7793, doi:10.5194/acp-9-7769-2009, 2009.

Cubison, M. J., Ervens, B., Feingold, G., Docherty, K. S., Ulbrich, I. M., Shields, L., Prather, K., Hering, S., and Jimenez, J. L.: The influence of chemical composition and mixing state of Los Angeles urban aerosol on CCN number and cloud properties, Atmos. Chem. Phys., 8, 5649–5667, doi:10.5194/acp-8-5649-2008, 2008.

DeCarlo, P. F., Kimmel, J. R., Trimborn, A., Northway, M. J., Jayne, J. T., Aiken, A. C., Gonin, M., Fuhrer, K., Horvath, T., Docherty, K. S., Worsnop, D. R., and Jimenez, J. L.: Field-Deployable, High-Resolution, Time-of-Flight Aerosol Mass Spectrometer, Anal. Chem., 78, 8281–8289, doi:10.1021/ac061249n, 2006.

DeCarlo, P. F., Dunlea, E. J., Kimmel, J. R., Aiken, A. C., Sueper, D., Crounse, J., Wennberg, P. O., Emmons, L., Shinozuka, Y., Clarke, A., Zhou, J., Tomlinson, J., Collins, D. R., Knapp, D., Weinheimer, A. J., Montzka, D. D., Campos, T., and Jimenez, J. L.: Fast airborne aerosol size and chemistry measurements above Mexico City and Central Mexico during the MILAGRO campaign, Atmos. Chem. Phys., 8, 4027–4048, doi:10.5194/acp-8-4027-2008, 2008.

DeCarlo, P. F., Ulbrich, I. M., Crounse, J., de Foy, B., Dunlea, E. J., Aiken, A. C., Knapp, D., Weinheimer, A. J., Campos, T., Wennberg, P. O., and Jimenez, J. L.: Investigation of the sources and processing of organic aerosol over the Central Mexican Plateau from aircraft measurements during MILAGRO, Atmos. Chem. Phys., 10, 5257–5280, doi:10.5194/acp-10-5257-2010, 2010.

Dillner, A. M., Schauer, J. J., Christensen, W. F., and Cass, G. R.: A Quantitative Method for Clustering Size Distributions of Elements, Atmos. Environ., 39, 1525–1537, doi:10.1016/j.atmosenv.2004.11.035, 2005.

Docherty, K. S., Aiken, A. C., Huffman, J. A., Ulbrich, I. M., DeCarlo, P. F., Sueper, D., Worsnop, D. R., Snyder, D. C., Peltier, R. E., Weber, R. J., Grover, B. D., Eatough, D. J., Williams, B. J., Goldstein, A. H., Ziemann, P. J., and Jimenez, J. L.: The 2005 Study of Organic Aerosols at Riverside (SOAR-1): instrumental intercomparisons and fine particle composition, Atmos. Chem. Phys., 11, 12387–12420, doi:10.5194/acp-11-12387-2011, 2011.

Drewnick, F., Hings, S. S., DeCarlo, P., Jayne, J. T., Gonin, M., Fuhrer, K., Weimer, S., Jimenez, J. L., Demerjian, K. L., Borrmann, S., and Worsnop, D. R.: A New Time-of-Flight Aerosol Mass Spectrometer (TOF-AMS) – Instrument Description and First Field Deployment, Aerosol Sci. Tech., 39, 637–658, doi:10.1080/02786820500182040, 2005.

Dzepina, K., Volkamer, R. M., Madronich, S., Tulet, P., Ulbrich, I. M., Zhang, Q., Cappa, C. D., Ziemann, P. J., and Jimenez, J. L.: Evaluation of recently-proposed secondary organic aerosol models for a case study in Mexico City, Atmos. Chem. Phys., 9, 5681–5709, doi:10.5194/acp-9-5681-2009, 2009.

Engel-Cox, J. A. and Weber, S. A.: Compilation and Assessment of Recent Positive Matrix Factorization and Unmix Receptor Model Studies on Fine Particulate Matter Source Apportionment for the Eastern United States, J. Air Waste Manage. Assoc., 57, 1307–1316, doi:10.3155/1047-3289.57.11.1307, 2007.

Feng, J.: A Size-Resolved Model and a Four-Mode Parameterization of Dry Deposition of Atmospheric Aerosols, J. Geophys. Res.-Atmos., 113, D12201, doi:10.1029/2007jd009004, 2008.

Gard, E., Mayer, J. E., Morrical, B. D., Dienes, T., Fergenson, D. P., and Prather, K. A.: Real-Time Analysis of Individual Atmospheric Aerosol Particles: Design and Performance of a Portable ATOFMS, Anal. Chem., 69, 4083–4091, 1997.

Gietl, J. K. and Klemm, O.: Source Identification of Size-Segregated Aerosol in Munster, Germany, by Factor Analysis, Aerosol Sci. Tech., 43, 828–837, doi:10.1080/02786820902953923, 2009.

Gunthe, S. S., King, S. M., Rose, D., Chen, Q., Roldin, P., Farmer, D. K., Jimenez, J. L., Artaxo, P., Andreae, M. O., Martin, S. T., and Pöschl, U.: Cloud condensation nuclei in pristine tropical rainforest air of Amazonia: size-resolved measurements and modeling of atmospheric aerosol composition and CCN activity, Atmos. Chem. Phys., 9, 7551–7575, doi:10.5194/acp-9-7551-2009, 2009.

Han, J. S., Moon, K. J., Lee, S. J., Kim, Y. J., Ryu, S. Y., Cliff, S. S., and Yi, S. M.: Size-resolved source apportionment of ambient particles by positive matrix factorization at Gosan background site in East Asia, Atmos. Chem. Phys., 6, 211–223, doi:10.5194/acp-6-211-2006, 2006.

Harshman, R. A. and Lundy, M. E.: PARAFAC – Parallel Factor-Analysis, Comput. Stat. Data An., 18, 39–72, 1994.

Hodzic, A., Jimenez, J. L., Madronich, S., Canagaratna, M. R., DeCarlo, P. F., Kleinman, L., and Fast, J.: Modeling organic aerosols in a megacity: potential contribution of semi-volatile and intermediate volatility primary organic compounds to secondary organic aerosol formation, Atmos. Chem. Phys., 10, 5491–5514, doi:10.5194/acp-10-5491-2010, 2010.

Huffman, J. A., Docherty, K. S., Aiken, A. C., Cubison, M. J., Ulbrich, I. M., DeCarlo, P. F., Sueper, D., Jayne, J. T., Worsnop, D. R., Ziemann, P. J., and Jimenez, J. L.: Chemically-resolved aerosol volatility measurements from two megacity field studies, Atmos. Chem. Phys., 9, 7161–7182, doi:10.5194/acp-9-7161-2009, 2009.

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. Tech., 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, J. Geophys. Res.-Atmos., 108, 8425, doi:10.1029/2001JD001213, 2003.

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, 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, doi:10.1126/science.1180353, 2009.

Karanasiou, A. A., Siskos, P. A., and Eleftheriadis, K.: Assessment of Source Apportionment by Positive Matrix Factorization Analysis on Fine and Coarse Urban Aerosol Size Fractions, Atmos. Environ., 43, 3385–3395, doi:10.1016/j.atmosenv.2009.03.051, 2009.

Khlystov, A., Zhang, Q., Jimenez, J. L., Stanier, C., Pandis, S. N., Canagaratna, M. R., Fine, P., Misra, C., and Sioutas, C.: In Situ Concentration of Semi-Volatile Aerosol Using Water-Condensation Technology, J. Aerosol Sci., 36, 866–880, doi:10.1016/j.jaerosci.2004.11.005, 2005.

Kleeman, M. J., Riddle, S. G., Robert, M. A., Jakober, C. A., Fine, P. M., Hays, M. D., Schauer, J. J., and Hannigan, M. P.: Source Apportionment of Fine ($PM_{1.8}$) and Ultrafine ($PM_{0.1}$) Airborne Particulate Matter During a Severe Winter Pollution Episode, Environ. Sci. Technol., 43, 272–279, doi:10.1021/es800400m, 2009.

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 Prevot, 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/es0707207, 2008.

Larson, T. V., Covert, D. S., Kim, E., Elleman, R., Schreuder, A. B., and Lumley, T.: Combining Size Distribution and Chemical Species Measurements into a Multivariate Receptor Model of PM$_{2.5}$, J. Geophys. Res.-Atmos., 111, D10s09, doi:10.1029/2005jd006285, 2006.

Levin, E. J. T., McMeeking, G. R., Carrico, C. M., Mack, L. E., Kreidenweis, S. M., Wold, C. E., Moosmuller, H., Arnott, W. P., Hao, W. M., Collett, J. L., and Malm, W. C.: Biomass Burning Smoke Aerosol Properties Measured During Fire Laboratory at Missoula Experiments (Flame), J. Geophys. Res.-Atmos., 115, D18210, doi:10.1029/2009jd013601, 2010.

Liu, P., Ziemann, P. J., Kittelson, D. B., and McMurry, P. H.: Generating Particle Beams of Controlled Dimensions and Divergence, 1. Theory of Particle Motion in Aerodynamic Lenses and Nozzle Expansions, Aerosol Sci. Tech., 22, 293–313, 1995a.

Liu, P., Ziemann, P. J., Kittelson, D. B., and McMurry, P. H.: Generating Particle Beams of Controlled Dimensions and Divergence, 2. Expeimental Evaluation of Particle Motion in Aerodynamic Lenses and Nozzle Expansions, Aerosol Sci. Tech., 22, 314–324, 1995b.

Marple, V. A., Rubow, K. L., and Behm, S. M.: A Microorifice Uniform Deposit Impactor (MOUDI) – Description, Calibration, and Use, Aerosol Sci. Tech., 14, 434–446, 1991.

Mauderly, J. L. and Chow, J. C.: Health Effects of Organic Aerosols, Inhal. Toxicol., 20, 257–288, doi:10.1080/08958370701866008, 2008.

McLafferty, F. and Turecek, F.: Interpretation of Mass Spectra, 4th Edn., University Science Books, Mill Valley, CA, 1993.

McMeeking, G. R., Kreidenweis, S. M., Carrico, C. M., Lee, T., Collett, J. L., and Malm, W. C.: Observations of Smoke-Influenced Aerosol During the Yosemite Aerosol Characterization Study: Size Distributions and Chemical Composition, J. Geophys. Res.-Atmos., 110, D09206, doi:10.1029/2004jd005389, 2005.

Moffet, R. C., de Foy, B., Molina, L. T., Molina, M. J., and Prather, K. A.: Measurement of ambient aerosols in northern Mexico City by single particle mass spectrometry, Atmos. Chem. Phys., 8, 4499–4516, doi:10.5194/acp-8-4499-2008, 2008.

Molina, L. T., Madronich, S., Gaffney, J. S., Apel, E., de Foy, B., Fast, J., Ferrare, R., Herndon, S., Jimenez, J. L., Lamb, B., Osornio-Vargas, A. R., Russell, P., Schauer, J. J., Stevens, P. S., Volkamer, R., and Zavala, M.: An overview of the MILAGRO 2006 Campaign: Mexico City emissions and their transport and transformation, Atmos. Chem. Phys., 10, 8697–8760, doi:10.5194/acp-10-8697-2010, 2010.

Murphy, D. M. and Thomson, D. S.: Laser Ionization Mass-Spectroscopy of Single Aerosol-Particles, Aerosol Sci. Tech., 22, 237–249, 1995.

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/es102951k, 2011.

Paatero, P.: Least Squares Formulation of Robust Non-Negative Factor Analysis, Chemom. Intell. Lab. Syst., 37, 23–35, 1997a.

Paatero, P.: A Weighted Non-Negative Least Squares Algorithm for Three-Way "PARAFAC" Factor Analysis, Chemom. Intell. Lab. Syst., 38, 223–242, 1997b.

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 the Multilinear Engine Program "ME2" for Fitting Multilinear and Quasi-Multilinear Models, University of Helsinki, Finland, 2000.

Paatero, P.: User's Guide for Positive Matrix Factorization Programs PMF2.Exe and PMF3.Exe, University of Helsinki, Finland, 2007.

Paatero, P. and Hopke, P. K.: Discarding or Downweighting High-Noise Variables in Factor Analytic Models, Anal. Chim. Acta, 490, 277–289, doi:10.1016/s0003-2670(02)01643-4, 2003.

Paatero, P. and Hopke, P. K.: Rotational Tools for Factor Analytic Models, J. Chemom., 23, 91–100, doi:10.1002/cem.1197, 2009.

Paatero, P. and Tapper, U.: Analysis of Different Modes of Factor Analysis as Least Squares Fit Problems, Chemom. Intell. Lab. Syst., 18, 183–194, 1993.

Paatero, P., Hopke, P. K., Song, X. H., and Ramadan, Z.: Understanding and Controlling Rotations in Factor Analytic Models, Chemom. Intell. Lab. Syst., 60, 253–264, 2002.

Pere-Trepat, E., Kim, E., Paatero, P., and Hopke, P. K.: Source Apportionment of Time and Size Resolved Ambient Particulate Matter Measured with a Rotating DRUM Impactor, Atmos. Environ., 41, 5921–5933, doi:10.1016/j.atmosenv.2007.03.022, 2007.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Statistical Descritption of Data, in: Numerical Recipes: The Art of Scientific Computing, 3rd Edn., Cambridge University Press, New York, USA, 720–772, 2007.

Reff, A., Eberly, S. I., and Bhave, P. V.: Receptor Modeling of Ambient Particulate Matter Data Using Positive Matrix Factorization: Review of Existing Methods, J. Air Waste Manage. Assoc., 57, 146–154, 2007.

Salcedo, D., Onasch, T. B., Aiken, A. C., Williams, L. R., de Foy, B., Cubison, M. J., Worsnop, D. R., Molina, L. T., and Jimenez, J. L.: Determination of particulate lead using aerosol mass spectrometry: MILAGRO/MCMA-2006 observations, Atmos. Chem. Phys., 10, 5371–5389, doi:10.5194/acp-10-5371-2010, 2010.

Schauer, J. J., Rogge, W. F., Hildemann, L. M., Mazurek, M. A., and Cass, G. R.: Source Apportionment of Airborne Particulate Matter Using Organic Compounds as Tracers, Atmos. Environ., 30, 3837–3855, 1996.

Smith, J. N. and Rathbone, G. J.: Carboxylic Acid Characterization in Nanoparticles by Thermal Desorption Chemical Ionization Mass Spectrometry, Int. J. Mass Spectrom., 274, 8–13, doi:10.1016/j.ijms.2008.04.008, 2008.

Srivastava, A., Gupta, S., and Jain, V. K.: Winter-Time Size Distribution and Source Apportionment of Total Suspended Particulate Matter and Associated Metals in Delhi, Atmos. Res., 92, 88–99, doi:10.1016/j.atmosres.2008.09.005, 2009.

Su, Y. X., Sipin, M. F., Furutani, H., and Prather, K. A.: Development and Characterization of an Aerosol Time-of-Flight Mass Spectrometer with Increased Detection Efficiency, Anal. Chem., 76, 712–719, doi:10.1021/ac034797z, 2004.

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 pro-
    cesses of organic and inorganic aerosols in New York city with
    a high-resolution time-of-flight aerosol mass apectrometer, At-
    mos. Chem. Phys., 11, 1581–1602, doi:10.5194/acp-11-1581-
    2011, 2011.
Thomson, D. S., Schein, M. E., and Murphy, D. M.: Particle Anal-
    ysis by Laser Mass Spectrometry Wb-57f Instrument Overview,
    Aerosol Sci. Tech., 33, 153–169, 2000.
Thurston, G. D. and Spengler, J. D.: A Quantitative Assessment of
    Source Contributions to Inhalable Particulate Matter Pollution in
    Metropolitan Boston, Atmos. Environ., 19, 9–25, 1985.
Tobias, H. J., Kooiman, P. M., Docherty, K. S., and Ziemann, P.
    J.: Real-Time Chemical Analysis of Organic Aerosols Using a
    Thermal Desorption Particle Beam Mass Spectrometer, Aerosol
    Sci. Tech., 33, 170–190, 2000.
Tucker, L. R.: Some Mathematical Notes on 3-Mode Factor Analy-
    sis, Psychometrika, 31, 279–311, 1966.
Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and
    Jimenez, J. L.: Interpretation of organic components from Posi-
    tive Matrix Factorization of aerosol mass spectrometric data, At-
    mos. Chem. Phys., 9, 2891–2918, doi:10.5194/acp-9-2891-2009,
    2009.
Viana, M., Kuhlbusch, T. A. J., Querol, X., Alastuey, A., Harri-
    son, R. M., Hopke, P. K., Winiwarter, W., Vallius, A., Szidat,
    S., Prevot, A. S. H., Hueglin, C., Bloemen, H., Wahlin, P., Vec-
    chi, R., Miranda, A. I., Kasper-Giebl, A., Maenhaut, W., and
    Hitzenberger, R.: Source Apportionment of Particulate Matter in
    Europe: A Review of Methods and Results, J. Aerosol Sci., 39,
    827–849, doi:10.1016/j.jaerosci.2008.05.007, 2008.
Wang, J., Cubison, M. J., Aiken, A. C., Jimenez, J. L., and Collins,
    D. R.: The importance of aerosol mixing state and size-resolved
    composition on CCN concentration and the variation of the im-
    portance with atmospheric aging of aerosols, Atmos. Chem.
    Phys., 10, 7267–7283, doi:10.5194/acp-10-7267-2010, 2010.
Watson, J. G.: Visibility: Science and Regulation, J. Air Waste
    Manage. Assoc., 52, 628–713, 2002.
Whitby, K. T.: Physical Characteristics of Sulfur Aerosols, Atmos.
    Environ., 12, 135–159, 1978.
Williams, B. J., Goldstein, A. H., Kreisberg, N. M., and Her-
    ing, S. V.: An in-Situ Instrument for Speciated Organic
    Composition of Atmospheric Aerosols: Thermal Desorption
    Aerosol GC/MS-FID (TAG), Aerosol Sci. Tech., 40, 627–638,
    doi:10.1080/02786820600754631, 2006.

Williams, B. J., Goldstein, A. H., Kreisberg, N. M., Hering, S. V.,
    Worsnop, D. R., Ulbrich, I. M., Docherty, K. S., and Jimenez, J.
    L.: Major components of atmospheric organic aerosol in south-
    ern California as determined by hourly measurements of source
    marker compounds, Atmos. Chem. Phys., 10, 11577–11603,
    doi:10.5194/acp-10-11577-2010, 2010.
Yakovleva, E., Hopke, P. K., and Wallace, L.: Receptor Modeling
    Assessment of Particle Total Exposure Assessment Methodology
    Data, Environ. Sci. Technol., 33, 3645–3652, 1999.
Yatavelli, R. L. N. and Thornton, J. A.: Particulate Or-
    ganic Matter Detection Using a Micro-Orifice Volatiliza-
    tion Impactor Coupled to a Chemical Ionization Mass Spec-
    trometer (MOVI-CIMS), Aerosol Sci. Tech., 44, 61–74,
    doi:10.1080/02786820903380233, 2010.
Yatkin, S. and Bayram, A.: Source Apportionment of PM10 and
    PM2.5 Using Positive Matrix Factorization and Chemical Mass
    Balance in Izmir, Turkey, Sci. Total Environ., 390, 109–123,
    doi:10.1016/j.scitotenv.2007.08.059, 2008.
Zelenyuk, A. and Imre, D.: Single Particle Laser Ablation Time-of-
    Flight Mass Spectrometer: An Introduction to SPLAT, Aerosol
    Sci. Tech., 39, 554–568, doi:10.1080/027868291009242, 2005.
Zhang, Q., Worsnop, D. R., Canagaratna, M. R., and Jimenez, J.
    L.: Hydrocarbon-like and oxygenated organic aerosols in Pitts-
    burgh: insights into sources and processes of organic aerosols,
    Atmos. Chem. Phys., 5, 3289–3311, doi:10.5194/acp-5-3289-
    2005, 2005.
Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Allan, J. D., Coe,
    H., Ulbrich, I., Alfarra, M. R., Takami, A., Middlebrook, A.
    M., Sun, Y. L., Dzepina, K., Dunlea, E., Docherty, K., DeCarlo,
    P., Salcedo, D., Onasch, T. B., Jayne, J. T., Miyoshi, T., Shi-
    mono, A., Hatakeyama, N., Takegawa, N., Kondo, Y., Schnei-
    der, J., Drewnick, F., Weimer, S., Demerjian, K. L., Williams,
    P. I., Bower, K. N., Bahreini, R., Cottrell, L., Griffin, R. J.,
    Rautianen, J., and Worsnop, D. R.: Ubiquity and Dominance of
    Oxygenated Species in Organic Aerosols in Anthropogenically-
    Influenced Northern Hemisphere Mid-Latitudes, Geophys. Res.
    Lett., 34, L13801, doi:10.1029/2007GL029979, 2007.
Zhang, X. F., Smith, K. A., Worsnop, D. R., Jimenez, J. L., Jayne, J.
    T., Kolb, C. E., Morris, J., and Davidovits, P.: Numerical Char-
    acterization of Particle Beam Collimation: Part II – Integrated
    Aerodynamic-Lens-Nozzle System, Aerosol Sci. Tech., 38, 619–
    638, doiL10.1080/02786820490479833, 2004.