

Supplemental Information for

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)

S1 Literature Review: 2-dimensional factorization of datasets that include size-distribution data

In this section, we discuss and summarize research that reports the application of mathematical techniques to datasets that include particle size information. We first present an overview of the mathematical techniques used in studies of 2-dimensional (2D) datasets. Then we summarize the arrangement of datasets containing particle size information into 2D matrices. The datasets include particle size information in one of two ways: (1) the dataset contains aerosol size distributions, and may also include simultaneous measurements of gas-phase and/or bulk aerosol-phase chemical constituents; *or* (2) the dataset contains size-resolved aerosol chemical composition, and the factors obtained from analysis of different size ranges were compared. Finally, we summarize the studies in the literature that have applied factorization methods to these datasets.

S1.1 Mathematical techniques for 2-dimensional factor analysis

Various mathematical techniques have been applied to datasets that include size distribution or size-resolved aerosol composition information. Collectively, the goal of these techniques is the same: to determine particle sources. Several terms have been used in the literature to describe these related techniques, including factor analysis, source apportionment, and matrix factorization. Each of these terms has a different technical definition and common usage. Although “factor analysis” technically refers to multivariate analyses that produce orthogonal factors (Malinowski, 1991), it is commonly used to refer to a variety of multivariate analyses, regardless of the orthogonality of the factors. In this work, we accept the common usage and refer to the class of multivariate analyses that have been applied the datasets of interest as “factor analysis.”

We now discuss how the 2D factorization model represents aerosol processes. A 2D matrix is deconvolved such that each factor is composed of two one-dimensional (1-D) vectors (i.e., a bilinear unmixing model). Several examples of this scheme are shown in Fig. S10. One factor vector describes how much of that factor is present in each sample and is called a time series. The other vector describes the composition of the aerosol in that factor and is called a profile. A

factor's profile shows the fractional contribution of the constituents contained in the rows of the input matrix. For example, in the matrix shown in Fig. S10a, each matrix row contains the concentrations of particles at a given size, and each factor profile also contains a size distribution. In contrast, the matrix rows in Fig. S10b contain concentrations of particles at given sizes and simultaneously measured concentrations of gas-phase species. The factor profiles from this matrix thus contain a size distribution and relative concentrations of the gas-phase species.

The data matrix is reconstructed by the linear combination of some number of factors, and each measured sample (matrix rows) can have contributions from more than one factor. This reconstruction is described by

$$x_{ij} = \sum_p a_{ip}c_{pj} + e_{ij} \quad (S1)$$

where i and j are the row and column indices for the matrix, respectively; p is the number of factors; x_{ij} is an element of the $m \times n$ data matrix \mathbf{X} to be factored; a_{ip} is an element of the $m \times p$ matrix \mathbf{A} , the columns of which contain the factor time series; c_{pj} is an element of the $p \times n$ matrix \mathbf{C} , the rows of which contain the factor profiles; and e_{ij} is an element of the $m \times n$ matrix \mathbf{E} of the residuals of the solution, i.e., the difference between the measured data and the reconstruction.

Several factor analytical methods are available to solve the bilinear unmixing model described in Eq. (S1). The methods differ by the requirements of a priori knowledge about the factor profiles and constraints placed on the factor profiles and time series. In principal component analysis (PCA, Malinowski, 1991) and absolute principal component analysis (APCA, Thurston and Spengler, 1985), no a priori knowledge about the profiles is required. In these methods, the factor profiles are orthogonal to each other, and the factor time series are also orthogonal to each other. The orthogonality requirement creates factor profiles and time series with both positive and negative values. Similarly, Positive Matrix Factorization (PMF, Paatero, 1997) does not require a priori knowledge about the profiles. However, orthogonal factors are not possible in PMF2 as the elements of the factor profiles and time series are constrained to be positive. The positivity requirement means that all profiles will have positive signals, and the time series will have positive mass. Consequently, PMF factors are usually more physically meaningful than

those from PCA. In contrast to these two methods, the factor profiles are completely prescribed in the Chemical Mass Balance approach (CMB, Friedlander, 1973), and no constraints are placed on the factor time series.

Instead of using factor analytical methods, samples can be grouped into clusters based on sample similarity. Cluster analysis groups measurements with similar characteristics, and these clusters can be associated with sources of measured aerosol (Murphy et al., 2003; Dillner et al., 2005; Marcolli et al., 2006; Beddows et al., 2009). The clusters can also be represented with two 1-D vectors: the cluster profile is the average composition of the samples in the cluster, and the time series is the number of samples in the cluster over average sampling periods. In most cluster analyses, an entire sample vector is assigned to a group, and therefore each total measurement can be assigned to only one cluster. A notable exception is fuzzy cluster analysis, in which sample vectors can be assigned to multiple clusters with an associated degree of membership to each group (Bezdek et al., 1981).

These mathematical techniques have been applied to 2D datasets that include particle size information. We now summarize these types of studies in order of increasing complexity of the data in the factorization matrix.

S1.2 Arrangements of datasets including particle size information for factor analysis

The datasets that include size distributions and whose factorization has been reported in the literature can be divided into two broad categories. In the first category, samples include one set of measurements at each time. These measurements may include any of the following: particle size distributions, particle-phase bulk chemical composition, or gas-phase chemical composition. The measurements are arranged as 2D matrix in which one dimension contains the measured data for each time step and the other dimension is time. This category of datasets contains four subcategories, depending on the type of measured data (Fig. S10a–d). All four subcategories include particle size distributions, i.e., the number or volume concentration of particles at multiple particle sizes. One subcategory uses only size distributions in the input matrix (Fig. S10a), while the other subcategories couple size distributions and different combinations of gas-phase composition and bulk aerosol chemical composition data (Fig. S10b–d).

In the second category, samples include two dependent measurements at each time: particle size, and the particles' chemical composition at each size, i.e., size-resolved composition data. The size-resolved composition data can be arranged in two forms for factorization (Fig. S10e–f). In the first arrangement, the chemical composition data from each particle size form a separate 2D matrix, and each 2D matrix is factored independently of the others (Fig. S10e). In this arrangement, a single source may contribute to different size ranges, but is not required to have the same chemical composition at every size. In contrast, the size-resolved composition data can be arranged such that the measured concentrations from different particles are appended as rows of a single 2D matrix (Fig. S10f). In this arrangement, the chemical composition of each factor must be the same for all particle sizes.

S1.3 Research reporting 2-dimensional factorizations using particle-size information

Research that has reported 2D factor analysis of aerosol number or mass size distributions, or datasets of size-resolved, aerosol chemical composition are summarized briefly in Table S1 and in greater detail in Tables S2–S7. Schematics of the factorization of six dataset types are shown in Fig. S10. Each of the six categories is discussed below.

The first category of studies applies factor analysis techniques to datasets of aerosol number distributions or combined aerosol number and mass distributions (Fig. S10a, Table S2). A good example from this category is the study by Costabile et al. (2009), who used PCA to analyze aerosol number distributions collected over two years from eight locations in and near Leipzig, Germany. The authors analyzed the data in two ways: (1) They factored the measurements from each site separately and (2) they factored combined measurements from selected subsets of the locations. Combining measurements from different locations allowed the authors to assess the temporal and spatial variation in the number size distribution. The authors identify factors in the nucleation mode, Aitken mode, and accumulation mode size ranges. The factors have names such as “fresh, roadside,” “fresh, background,” “rural”, “urban traffic”, and “continental.” The authors assigned some of their factors to sources, but noted that some aerosol modes have contributions from multiple sources, e.g., long-range-transported and primary urban aerosol. Another study in this category acknowledges that particle size distribution components may not be directly related to particle sources (Chan and Mozurkewich, 2007a), and so they combine

simplified representations of size distributions derived by APCA from three rural and urban locations near Toronto with gas-phase concentrations of CO, NO_x, SO₂, and O_x (NO₂ + O₃) and wind speed to associate the different aerosol modes with photochemical processing, regional pollution, boundary layer dynamics, local anthropogenic emissions, and processed nucleated particles (Chan and Mozurkewich, 2007b). However, most of the other studies in this category assign factors to sources such as stationary combustion sources, local traffic, spark-ignition gasoline emissions, and secondary aerosol, but only rarely note that such assignments can only be tentative without aerosol composition data (Kim et al., 2004). In addition to identifying aerosol modes, Costabile et al. (2009) presented a paradigm for transformation of the aerosol size distribution from the local to the regional scale, showing that aerosol near sources has high temporal and spatial variability, but that the size distributions become similar in both space and time as aerosol is aged and transported regionally.

The second category of studies pairs aerosol number distributions with the concentrations of simultaneously measured gas-phase species (Fig. S10b, Table S3). In one such study, Thimmaiah et al. (2009) propose the factorization of size distribution and gas-phase data as a cost-effective method for informing air-quality management programs. However, only four factors are identified in this PMF study (ozone-rich, transported ozone/ozone precursors; NO_x-rich diesel emissions; traffic-spark ignition vehicles; and local heating sources), and these factors are weakly supported by correlations between the factor time series and the time series of the species included in the factorization. Because the gas-phase species are important contributors to the factors themselves, it is not surprising that the correlations are high. For example, a factor attributed to transported ozone and ozone precursors has a strong correlation with ozone concentrations, and only this factor has an appreciable contribution from ozone. Stronger support for factor identification requires correlations with tracers external to the factor analysis (Zhang et al., 2005a; Lanz et al., 2007). The Thimmaiah et al. study does not convincingly hold up its claim for producing useful information. In the only other published study in this category, Wahlin et al. (2001) measured aerosol number distributions, CO, and NO_x at roadside-sampling locations, and applied PCA to these data with the goal of separating diesel and gasoline contributions to ultrafine particles. The diesel and gasoline contributions could not be completely separated, and one factor was attributed to diesel vehicles, while another was

164 attributed to both gasoline and diesel vehicles. The authors concluded that more reliable
165 separation of the gasoline and diesel contributions requires a specific tracer for diesel emissions;
166 furthermore, they noted that better understanding of particle sources and transformation could
167 come from combining particle size distributions and particle composition data.

168 The third category of studies pairs aerosol number distributions with chemically speciated
169 measurements of the bulk aerosol (Fig. S10c, Table S4). Such datasets must be constructed
170 carefully with respect to the particle size distribution used in the analysis. The aerosol number
171 distribution is dominated by very small particles, but the aerosol chemical composition comes
172 mainly from the large particles that contribute most of the particle mass (Seinfeld and Pandis,
173 1998). However, the particle mass and volume distributions usually agree well, and so including
174 these together should be satisfactory for factor analysis. To address the discrepancy and
175 differences in the uncertainties between the number size distribution and composition data,
176 Larson et al. (2006) decreased the weight of the size information by a factor of 10 in the
177 factorization of measured particle number distributions, volume distributions, and chemically
178 resolved species from PM_{2.5} (particulate matter with diameter ≤ 2.5 μm). In this study, the
179 degree of weighting of the size information had only a minor effect on the factors' size
180 distributions and average mass contribution, but the effect on the factors' chemical composition
181 was not described. Factors identified in this study include vegetative burning, aged sea salt, and
182 metals processing.

183 The final category of datasets that include non-chemically resolved size-distribution data
184 combines them with both gas-phase and bulk particle composition data (Fig. S10d, Table S5).
185 Like the research that combined particle size distribution and chemical composition data, in these
186 studies the input aerosol size distributions usually encompassed a smaller size range than the
187 bulk composition data. For example, Zhou et al. (2005a) applied PMF to a dataset that combined
188 aerosol number distributions, concentrations of particulate nitrate, sulfate, and 11 trace metals,
189 and concentrations of five gas-phase species. Factor names were assigned based mainly on the
190 chemical composition of the factors, which included two secondary nitrate factors, coal-fired
191 power plant, steel mill, and nucleation. In this study, sulfate (and to a lesser degree, nitrate) had
192 substantial contributions to the aerosol composition of all factors. Although each factor included
193 a size distribution, little information was given about how well the overall size distributions were

fit, and it was unclear how strongly the size distribution information influenced the factorization results.

We have now reviewed the studies that include particle size distributions in the factorization matrix. In general, the particle size information may have been too strongly linked to sources when little or no particle chemical composition was included, or undervalued when the analysis relied predominantly on bulk composition data for source identification. We now consider studies that use size-resolved aerosol composition datasets to explore the sources of ambient aerosol.

The remaining 2D factorization studies explore such size-resolved chemical composition datasets. In most of these studies, factor analysis is performed separately on the chemical species measured for each size range (Fig. S10e, Table S6). The studies that factor the composition from each size range separately fall into two main groups based on the size resolution of the data: (1) studies that use species from overlapping bulk PM sizes (e.g., PM_{2.5}, PM₁₀) and (2) those that use multistage samplers to separate particles by size. Two studies of the latter type are reviewed here. Kleeman et al. (2009) measured 8 molecular organic tracers from six stages of a MOUDI sampler (0.055–1.8 µm) at urban locations in California. This study used a custom source-apportionment algorithm to relate tracers for five sources (wood burning, meat cooking, motor oil, gasoline, and diesel fuel) to measured elemental and organic carbon concentrations. The dominant sources of elemental carbon were found to be gasoline and diesel vehicle exhaust. Little organic carbon mass was attributed to vehicle exhaust. The dominant sources of organic carbon were found to be wood burning and meat cooking. Organic carbon that could not be assigned to the selected sources was 10–24% of the PM_{1.8} mass and 0–58% of the PM_{0.1} mass and may come from oil and gas refining in the region or from SOA. In another study, Han et al. (2006) analyzed a dataset of 19 elements measured from an 8-stage DRUM sampler (0.07 to ~12 µm) collected at Gosan, Korea, in 2002. PMF was applied separately to the chemical composition data from each stage, i.e., particle size range. The authors identified fifteen sources in total, with four to eight sources contributing to any single size range. The chemical profiles identified for the same source in different size ranges were shown to have high similarity, and most sources had a strong size dependence. For example, local soil and sea salt

were present mainly in coarse sizes, while coal combustion particles and diesel vehicle particles were identified in only the fine and ultrafine sizes, respectively.

The final type of 2D factorization is based on the assumption that particles from a single source have the same composition across a broad size range. Instead of factoring size-resolved composition data separately for each particle size range, Amato et al. (2009) combined speciated chemical measurements from different overlapping size ranges (PM_{10} , $PM_{2.5}$, and PM_1) into one factorization matrix (Fig. S10f, Table S7). This dataset could not be factored using a 3-dimensional (3D) model because the measurements from different sizes were not collected simultaneously. Instead, the authors arranged their input matrix such that species from each size were appended as rows of the matrix. In this matrix arrangement, the time series of each factor includes the contribution of that factor at each size, and each factor must have the same chemical profile for all sizes. The assumption that the factors have the same chemical profile for all sizes was not tested directly because the data from each size were not factored separately. However, the contributions of the sources to each particle size matched previous knowledge about these sources. For example, aged sea salt was found with high concentrations in PM_{10} , lower concentrations in $PM_{2.5}$, and negligible mass contributions to PM_1 . In contrast, vehicle exhaust had the comparable contributions to all three size ranges.

In summary, multiple 2D factorization approaches have been applied to aerosol datasets that include particle size distributions. These studies attempt to understand the processes influencing ambient particle size distributions, but at best provide speculative assignments to sources in the absence of chemical information. The addition of some chemical information from gas-phase species gave insight to particle sources only in one carefully constructed study (Costabile et al., 2009). Combining aerosol size distributions with particle composition data enabled more complete attempts to characterize the size distribution of aerosol sources. The 2D factorization approaches that used size-resolved aerosol composition are the most promising of the 2D cases for obtaining a size-resolved source apportionment. However, these studies have coarse time and size resolution, and are unable to address the dynamic nature of the aerosol size distribution as it evolves through atmospheric processes.

S2 Research reporting 3-dimensional factorizations using particle-size information

Three peer-reviewed studies have reported 3D factorization of size-resolved aerosol composition datasets using the 3-vector model, or the vector-matrix model in which the vector contains the time series. The datasets include data from the submicron range to particles with diameters less than 10 μm (PM_{10}). Each is briefly discussed here.

Yakovleva et al. (1999) factored a dataset of particulate matter composition for PM_{10} $\text{PM}_{2.5}$ obtained during a study in Riverside, California in 1991. The goal of this study was to examine the relationship between daily activities and personal PM exposure. Particles were collected from five size-location combinations: PM_{10} samples were measured with personal, indoor, and outdoor monitors; and $\text{PM}_{2.5}$ samples were measured with indoor and outdoor monitors. The samples were analyzed for concentrations of 18 elements. The dataset was factored using the 2D model (applied separately to each particle size range) and the 3-vector model (Fig. 1a). The 2D and 3-vector analyses identified factors with different contributions to the size-location combinations. For example, sea salt contributed primarily to outdoor PM_{10} with small contributions to outdoor $\text{PM}_{2.5}$, while factors representing motor vehicle emissions and secondary sulfate were identified in all samples. In addition, particles generated by personal activities such as cooking, smoking, and vacuuming were identified in personal and indoor PM_{10} samples. The main difference between the factors from the 2D and 3-vector models was the separation of soil, identified as one factor in the 2D datasets, into three factors in the 3-vector results. These three soil factors had identical chemical profiles but different size distributions and were identified as ambient soil in outdoor PM_{10} samples, resuspended soil in personal PM_{10} samples, and indoor soil in personal and indoor PM_{10} samples.

Karanasiou et al. (2009) collected a dataset of coarse (PM_{10-2}) and fine (PM_2) particle samples at three sites in Athens, Greece, in 2002. The chemical speciation included 13 elements, black carbon (BC), and SO_4^{2-} . The dataset was factored with the 2D model (applied separately to each particle size range) and the 3-vector model. Using the 3-vector model enabled identification of some factors that could not be identified from the 2D factorizations. The factors identified only in the 3-vector model include motor vehicle exhaust contributions in the coarse fraction and a second road dust factor in both fractions. For some factors, different chemical profiles were

obtained for the coarse and fine fractions from the 2D and 3D factorizations. This result differs from the results of Han et al. (2006), who found that aerosol sources had very similar chemical profiles across different size ranges (Sect. S1.3 in the Supp. Info.).

Pere-Trepat et al. (2007) report the only study in the literature that used a vector-matrix model to factor a 3D size-resolved dataset of particle chemical composition. The dataset for this study was collected in Detroit, Michigan, on three stages of a DRUM sampler and included particles with diameters from 0.1–2.5 μm . The collected particles were analyzed for the concentrations of 27 elements. In addition, absorbances at four ultraviolet and visible wavelengths were included in the factorization matrix. The authors report that they attempted to use the 3-vector model for this dataset; however, the factor compositions were size dependent, and thus the assumptions of the trilinear model did not hold. The authors therefore factored their dataset using the vector-matrix model in which the vector contains the time series (Fig. 1d). This vector-matrix model finds the size-resolved chemical composition of a component with one time series, i.e., each factor represents the ensemble of particles that arrive simultaneously at the monitor, but the chemical composition of those particles may vary with size, though not with time. For example, the factor attributed to industrial metal works has the largest contribution from Fe at all sizes, but Ca appears only in the large particles and S occurs only in the larger two of the three size bins.

S3 m/z 's with organic signal omitted from the present study

The standard fragmentation matrix for organics does not assign organic signal to some m/z 's that are known to have organic signal but also have interferences from air or inorganic ion signals (m/z 14, 32, 33, 36, 39, 40, 46, and 47). This omission removes only a small part of the organic aerosol mass, about 7% of the total mass (Aiken et al., 2009). Of the omitted m/z 's, the greatest contribution to the organic mass comes from m/z 's 39 and 40. Each of these m/z 's contributes ~2% of the total organic mass of the MS mode data.

In addition to the masses normally omitted by the fragmentation matrix, we also omit m/z 's 12 and 30 from this analysis. These m/z 's can be examined in the high-resolution MS-mode data (Aiken et al., 2009) and appear to deviate from the assumptions of the standard fragmentation matrix. m/z 12 has contributions from an unidentified ion that can contribute up to 25% of the signal at this m/z , and we do not know how this unidentified ion might be represented in the size distribution. At m/z 30, interference from NO^+ requires estimation of the organic signal. The organic signal is predicted to come from the ^{13}C isotope of C_2H_5^+ , but is instead mainly from CH_2O^+ ions, for which a fragmentation ratio to another peak has not been characterized.

S4 Comparison of the PMF3 and ME-2 algorithms for solving the 3-vector model

The 3-vector model can be solved by PMF3 or ME-2, which use different convergence algorithms to minimize Q . Both algorithms iteratively minimize Q , and each step in the iteration tests a possible solution in the “ Q space”. The Q space is a function of the data and error matrices and the model used to fit the data, but is independent of the algorithm used to explore that space. We expect that both algorithms have landed in the same local minimum when they find solutions with the same factors and similar Q values.

ME-2 minimizes Q using the conjugate gradient algorithm (Paatero, 1999), which stops after encountering a sharp change in the Q gradient. Such a change marks the transition from a “wall” to the “floor” of a local minimum in the Q space. In contrast, PMF3 uses the Gauss-Newton algorithm (Paatero, 1999), which seeks the lowest point in the local minimum in the Q space. Logically, the lowest point in the minimum could be lower than the edge of the floor. This result is observed in solutions of the 3-vector model in this study (Fig. 3b). These solutions have downweighted Q/Q_{exp} values with differences similar to those between solutions in the same family solved by the same algorithm ($< 10^{-4} Q/Q_{exp}$ units). However, the unweighted Q/Q_{exp} values show somewhat larger differences (Fig. 3a). The difference in the unweighted Q/Q_{exp} values between the solutions indicates that the fits differ somewhat, and that differences in the fits occur mainly in the downweighted m/z -size combinations. In fact, comparing the best four-factor solution from each algorithm shows that the main difference in the fits is at m/z 100, where the PMF3 solution has a lower Q contribution. Since m/z 100 has low SNR at all sizes, the difference between the Q contributions is amplified when the downweighting is removed.

We do not believe that the lower Q/Q_{exp} values from PMF3 for this dataset imply that the PMF3 solution is better, or that PMF3 uses a better algorithm. Furthermore, the Gauss-Newton method used by PMF3 is less efficient than the conjugate-gradient algorithm used by ME-2 when solving large problems (Paatero, 1999). We observed this difference in the speed of computing solutions of the 3-vector model with six or more factors (Fig. S11). Therefore, we suggest using ME-2 to solve the 3-vector model for its speed advantage, and also when researchers plan to compare to other models (which cannot be computed by PMF3).

S5 Choice of solution of the 3-vector model

In this section we explore solutions of the 3-vector model with two or more factors to choose the best solutions. Each family of solutions is examined against the criteria outlined in Sect. 3.4.3 to determine whether the factors are physically meaningful.

S5.1 Solutions of the 3-vector model with two and three factors

The two-factor solution of the 3-vector model has only one family, in which one factor has a mass spectrum similar to OOA and the other has a mass spectrum that could be identified as HOA or BBOA. BBOA is generally differentiated from HOA in unit-mass-resolution data by BBOA's higher contributions from characteristic biomass-burning markers at m/z 's 60 and 73. However, the contribution from these m/z 's in this mass spectrum is higher than in the HR-MS HOA mass spectrum but considerably lower than in the HR-MS BBOA mass spectrum. This factor appears to be a mix of the BBOA and HOA contributions that cannot be separated with only two factors. Thus we consider solutions of the 3-vector model with three factors.

The three-factor solutions of the 3-vector model fall into two families. The family with the lower Q/Q_{exp} values has HOA, BBOA, and OOA factors. These factors appear to be physically meaningful; thus this solution could be acceptable. However, the solution lacks an LOA factor. The largest contribution to Q/Q_{exp} for this solution comes from m/z 58, which in the HR-MS solution was composed almost completely of organic-nitrogen fragments ($C_3H_8N^+$ and $C_2H_4NO^+$). Though LOA has a small contribution to the total aerosol mass, its chemical composition is distinctive, and we would expect to find such a factor in the PToF data.

The other 3-vector family with three factors has a Q/Q_{exp} that is 0.7% higher than the first family. We note that this difference is much larger than the final convergence criterion used to compute the solutions (0.001% Q/Q_{exp}), suggesting that the families indeed represent different local minima in the solution space. We do not know whether this increase in Q/Q_{exp} is meaningful (i.e., large enough to imply that the solution is bad) because comparisons between families of solutions have not been discussed in the literature of 3D aerosol factorization. The closest literature comparison is the 2D factorization of an AMS dataset by DeCarlo et al. (2010), in which the Q/Q_{exp} values for 50 seed solutions with four factors had a range of 1.5% above the

minimum Q/Q_{exp} . In this case, the family with the best solution included both the lowest and highest Q/Q_{exp} values. We might therefore expect a 0.7% increase in Q/Q_{exp} to be acceptable; however, it is not clear that the properties of 2D and 3D factorizations of different datasets should be similar. Therefore we make no judgment in the present case about the meaning of increases of Q/Q_{exp} .

The factors in the higher Q/Q_{exp} family are identified as OOA, BBOA, and LOA. The OOA and BBOA factors have mass spectra, time series, and size distribution that are similar to those of the lower Q/Q_{exp} family. The LOA, in contrast, is more difficult to recognize because its mass spectrum is distorted compared to the HR-MS LOA factor. The HR-MS LOA mass spectrum has significant contributions from m/z 's also prominent in the HOA and BBOA spectra. These peaks, at m/z 27 and 29, 41 and 43, 55 and 57, etc., form a characteristic “picket fence” pattern of peaks containing $C_nH_{2n-1}^+$ and $C_nH_{2n-1}CO^+$ ions at the lower m/z of each pair and $C_nH_{2n+1}^+$ and $C_nH_{2n+1}CO^+$ ions at the higher m/z of each pair (McLafferty and Turecek, 1993). However, these characteristic hydrocarbon peaks are much less prominent in the LOA mass spectrum found in this study. The HR-MS spectrum contains another characteristic peak at m/z 91, which is present in our LOA spectrum at similarly higher abundance relative to the neighboring peaks. Specifically, the HR-MS spectrum has a 3% contribution from m/z 91, which is ~10 times higher than the adjacent peaks in the spectrum, while our LOA has a 4.5% contribution from m/z 91, which is ~12 times higher than the adjacent peaks. The major difference between the HR-MS LOA spectrum and ours is at m/z 58. The nitrogen-containing fragments at m/z 58 contributed only 1% of the HR-MS LOA spectrum, but m/z 58 contributes 19% of the signal to LOA in this solution. The unusual appearance of our LOA spectrum would lead us to reject this factor as not physically meaningful if we did not know of its existence from the HR-MS solution. Nevertheless, the existence of the LOA factor in this solution demonstrates that LOA is identifiable in the PToF dataset, even though the mass spectrum is different than that in the HR-MS factor.

S5.2 Solutions of the 3-vector model with four factors

Since all of the HR-MS factors are found in the three-factor solutions, we expect to find solutions of the PToF dataset with four or more factors that include all four of the HR-MS

factors. Such a solution with four factors is found, but not in the family with the lowest Q/Q_{exp} value. The four-factor solutions are discussed in order of increasing Q/Q_{exp} value.

The four-factor solutions can be grouped into three families (Fig. S6). The family with the lowest Q/Q_{exp} values has two recognizable factors and two factors that do not appear to have physically meaningful mass spectra (Fig. S7). The recognizable factors represent HOA and BBOA and have mass spectra, size distribution, and time series similar to the three-factor solution with the lowest Q/Q_{exp} value. The mass spectra of the other factors each have large contributions from m/z 's that are not usually so prominent. One factor has a mass spectrum dominated by signals at m/z 44 and related peaks, as defined in the fragmentation table (m/z 's 28, 18, 17, and 16). The fraction of signal contributed by m/z 44 (f_{44}) is proportional to the aerosol oxygen-to-carbon ratio (O:C), which is also proportional to aerosol age, so high f_{44} represents highly oxidized, aged aerosol (Aiken et al., 2008). This factor might therefore represent OOA, and in fact has a high correlation with the OOA factor from the three-factor solution ($R = 0.86$). But this high- m/z -44 factor contributes only half of the mass of the OOA factor in the three-factor solution. Curiously, this mass spectrum has $f_{44} = 0.39$, and the group of related peaks contributes 90% of the signal in this mass spectrum. This spectrum represents aerosol with an elemental oxygen-to-carbon ratio (O:C) of 1.6 (Aiken et al., 2008) — much higher than O:C measured in ambient or laboratory aerosol by the AMS (Ng et al., 2010). O:C can be combined with the estimated hydrogen-to-carbon ratio (H:C, Ng et al., 2011a) to estimate the average carbon oxidation state (\overline{OS}_C) of the aerosol from this factor (Kroll et al., 2011). This factor has $\overline{OS}_C = +2.0$. However, the \overline{OS}_C of atmospheric aerosols rarely exceeds +1 because compounds with higher oxidation states tend to decompose (Kroll et al., 2011). Thus the factor in this solution with $f_{44} = 0.39$ is not physically meaningful, even though it improves the fit of the factorization, as evidenced by its appearance in the family of solutions with the lowest Q/Q_{exp} .

The other unrealistic factor in this solution has a mass spectrum that resembles HOA but is dominated by m/z 43. Compared to the HR-MS HOA, the mass spectrum of this factor has more signal at many m/z 's ≤ 45 and less signal at most m/z 's > 45 . Despite our factor's mass spectral HOA features, its time series is negatively correlated with that of HOA in the three-factor solution ($R = -0.12$). This factor's time series is actually correlated with the time series of OOA in the three-factor solution ($R = 0.77$) and accounts for 75% of the OOA mass from the three-

factor solution. Yet, the mass spectrum has no contribution from m/z 44, our most important marker for OOA. Thus the evidence for identifying this factor as HOA or OOA is contradictory. The factors dominated by m/z 43 or 44 likely represent a “splitting” of the OOA factor, i.e., the OOA factor has been divided into two factors (Ulbrich et al., 2009). This division can be demonstrated by summing the time series from the factors dominated by m/z 43 or m/z 44. The summed time series has very strong correlation with the time series of OOA in the three-factor solution ($R=0.95$) and similar mass contribution. The factor splitting can further be demonstrated by using the average mass contribution from these two factors as weights to calculate the weighted-average mass spectrum from these two factors. This weighted average mass spectrum has a strong correlation with the HR-MS OOA mass spectrum ($R=0.98$). From this evidence, we conclude that the factors dominated by m/z 43 or m/z 44 are not individually physically meaningful, and so we reject this family of solutions.

The next family of four-factor solutions has Q/Q_{exp} only 0.1% higher than the first family, but these factors are all recognizable aerosol types (Fig. 4). The four factors can be identified as OOA, HOA, BBOA, and LOA. The LOA mass spectrum has even higher fractions of m/z 58 and m/z 91 than in the three-factor solutions; however, these increases actually reflect a decreased contribution from the characteristic hydrocarbon m/z 's that also contribute to HOA and BBOA. The largest contribution to Q/Q_{exp} for this family is from m/z 43, and the remaining structure in the Q/Q_{exp} time series has only small features that have little correlation with factor time series or the total mass time series. All of the factors in this family of solutions are physically meaningful; thus, this solution could be acceptable.

The next family has Q/Q_{exp} 0.4% higher than the first family. The four factors in this family can be recognized from their mass spectra as OOA, HOA, and two factors representing BBOA (Fig. S8). The occurrence of two factors with the same aerosol type identification is rare in the solutions of the 3-vector model (Table 2), so the occurrence of two BBOA factors is intriguing. These two BBOA factors have very similar mass spectra ($R=0.94$), but quite different size distributions. One factor has a somewhat narrow size distribution with particle d_{va} that ranges ~40 nm to ~400 nm and has a mode at d_{va} ~180 nm. The other BBOA factor has a broader distribution with d_{va} that extends to 1150 nm and has a mode at a larger d_{va} near 370 nm. The difference in modes might suggest that the factor with the larger mode represents more aged

BBOA particles. However, that factor lacks signal at m/z 's 18 and 44, which usually represent oxidation, while the mass spectrum for the factor with the smaller mode has contributions from m/z 's 18 and 44 of 3.5%. Thus the interpretations based on the factor size distribution and mass spectra are contradictory, and we find that we cannot support this solution. The presence of two BBOA factors suggests that the BBOA factor has been split. Indeed, the sum of their time series has very high correlation with the time series of the BBOA factor in the previous family discussed ($R=0.98$) and has 21% more mass. In addition, the time series of these two factors are moderately correlated ($R=0.65$), a characteristic of split factors observed in previous studies (Ulbrich et al., 2009). The split of the BBOA factor is most likely because the 3-vector model is not able to fit the BBOA size distribution variability, which might be better fit in the vector-matrix model.

Another interesting feature of this solution is the broad size distribution of the HOA factor. The HOA factor has a bimodal size distribution that spans the entire diameter range (20 – 1300 nm) with modes at d_{va} of ~ 115 nm and ~ 410 nm (Fig. S8). The broad size distribution of this HOA factor is similar to the average size distribution of HOA in Pittsburgh aerosol (estimated as $12.3 * m/z\ 57 - 0.02 * m/z\ 44$, Zhang et al., 2005a; Zhang et al., 2005b). However, the size distribution of the Pittsburgh HOA is flatter, i.e., distinct modes are not easily identified. Since the other factors in this solution are physically meaningful, this solution may be acceptable. The only failing of this solution appears to be that it is missing an LOA factor, which is reflected by the high Q/Q_{exp} contribution from $m/z\ 58$.

In summary, the four-factor solutions of the 3-vector model included two families with potentially acceptable solutions. One of these families included all four of the HR-MS factors, while the other had two instances of the BBOA factor and lacked a factor for LOA. We now explore solutions with five or more factors to determine whether they might contain additional information, e.g., the four HR-MS factors and a new factor.

S5.3 Solutions of the 3-vector model with five or more factors

The five-factor solutions fall into three families (Fig. S6), and all contain factors that fail the criterion of being physically meaningful. Each family has the $m/z\ 43$ - and $m/z\ 44$ -dominated

factors that represent a splitting of the OOA factor. Only the family with the lowest Q/Q_{exp} value has HOA, BBOA, and LOA factors. The two families with the higher Q/Q_{exp} values include split BBOA or HOA factors (Table 2). Similar to the four-factor solution with split BBOA factors, the split HOA factor with the larger size distribution lacks signal in m/z 44, We therefore reject all of the five-factor solutions.

The solutions of the 3-vector model with six factors fall into five families. In all of these families, OOA is split into the factors dominated by m/z 43 and m/z 44. The solution with the lowest Q/Q_{exp} value also includes factors representing HOA, LOA, and the split BBOA factors. All of the families include split factors of HOA, BBOA, and/or the factor dominated by m/z 43. Some families include splits factors of both HOA and m/z 43, but no families have multiple instances of both BBOA and HOA factors. We reject these solutions because they do not contain physically meaningful factors, or because their factors do not provide new information compared to solutions with fewer factors.

The pattern of factors without additional physical meaning continues in solutions of the 3-vector model with seven or eight factors. These solutions have a higher-dimensional solution space, which has more possibilities for local minima. Accordingly, we observe more families than in the solutions with fewer factors, but these solutions include the same types of factors. Multiple occurrences of some factors are more common in the solutions with seven and eight factors. However, no new factor types are observed in any of these solutions.

In summary, the most acceptable solutions of the 3-vector model had four factors. One of the four-factor solutions included all four of the factors in the HR-MS solution. The other family included all of the HR-MS factors except LOA, but instead had two occurrences of the BBOA factor with similar mass spectra and time series but different size distributions. Because we do not have sufficient support for the validity of two BBOA factors, we choose the family that had the four factors in the HR-MS solution as the best solution of the 3-vector model.

S6 Solutions of the unconstrained vector-matrix model

Compared to the 3-vector solutions, solutions of the unconstrained vector-matrix model with the same number of factors have lower Q/Q_{exp} values (Fig. 3). The low Q/Q_{exp} values for the vector-matrix solutions imply that the vector-matrix model can fit more of the data than the 3-vector model. In other words, the two vectors in the 3-vector model fit the size-time information less well than the matrix in the vector-matrix model. The better fit of the vector-matrix model is consistent with this model's greater degrees of freedom [Eqns. (8) and (9)]. However, the large number of degrees of freedom in the vector-matrix model also allows the model to fit a large amount of noise. The ability to fit noise in this low SNR dataset distorts the mass spectra in the model solutions.

No solutions of the vector-matrix model have mass spectra that are all physically meaningful (Table 3). For example, in solutions with at least three factors, a single factor for OOA is obtained in only one solution. This solution contains OOA, HOA, and LOA factors, but not a BBOA factor. The omission of BBOA is surprising since this factor has a larger mass fraction than LOA when both are found in the same solution. In addition, this solution's increased Q/Q_{exp} may indicate a poor fit. However, in all other unconstrained vector-matrix solutions, the OOA is split into factors dominated by m/z 43 and m/z 44.

Other non-physical mass spectra appear in solutions with four or more factors. For example, a new, nonphysical factor not observed in the 3-vector solutions is dominated by m/z 's 67, 81, and 95 (Fig. S12). This series of m/z 's is present in mass spectra of dienes, alkynes, and cycloalkenes (McLafferty and Turecek, 1993), but the spectra of real compounds have a different distribution of the other peaks. Thus, these spectra are not known to represent a real aerosol type. No new, meaningful factors are identified in solutions with five or more factors. Consequently, we reject all of the unconstrained vector-matrix solutions.

S7 Constraining the vector-matrix model using multiple linear regression

Multiple linear regression solves the same model as Eq. (S1), but the factor compositions (i.e., the c_j values for p factors) must be provided a priori. In this work, the $p = 4$ factor compositions came from the four HR-MS mass spectra. The size-resolved mass spectrum measured at each time step was fit separately, i.e., 36×1366 fits are performed. Regressions were performed in Igor Pro v. 6.21 (Wavemetrics, Portland, OR) using the Levenberg-Marquard least-squares method. In addition, the fit coefficients (a_j) were constrained to be non-negative, matching the constraints of the solutions of the vector-matrix model solved by ME-2.

S7.1 Solutions of the multiple linear regression

The multiple linear regression of the HR-MS factors to the size-resolved composition data failed to fit most of the mass spectra. The regression was only successful (i.e., the matrix was not singular and the fit converged in the allowed number of steps) for 22.5% of the mass spectra. Most of these spectra are from particle size bins with diameters between 50 and 700 nm that have moderate mass spectral signal ($0.5\text{--}2 \mu\text{g}/\text{m}^3/\text{decade } d_{va}$). In contrast, the fit did not converge for the majority of the spectra (77%). The non-convergent fits include almost all of the data for particle size bins with $d_{va} < 50$ nm and $d_{va} > 700$ nm, and also particle size bins with d_{va} between 50 and 700 nm that have low signal ($< 0.5 \mu\text{g}/\text{m}^3/\log \text{ nm}$). Finally, the remaining 0.5% of mass spectral fits failed because the fitting matrix was singular (i.e., the mass spectra are linearly dependent, so an infinite number of solutions, are possible). The matrix singularity occurs during high-BBOA events, most likely because of near-colinearity between the BBOA and HOA spectra. Unfortunately, these failures occur when we are the most interested in the composition of the aerosol and whether the size distribution is evolving. In total, fitting the size-resolved mass spectra with the HR-MS mass spectra by multiple linear regression fails for 77.5% of the points. Furthermore, when the regression fits fail, no result is given for these points. Thus, this method is unable to provide factorization results from important, high-mass-loading events and cannot be used for this dataset.

S8 Choice of the number of factors in the solution of the constrained vector-matrix model

In this section we explore solutions of the fully constrained ($\beta = 0$) vector-matrix model with four or more factors to determine whether new, physically meaningful factors can be identified. Each family of solutions is examined against the criteria outlined in Sect. 3.4.3 to determine whether the factors are physically meaningful.

The four-factor solutions of the constrained vector-matrix model form one family. The four factors are the a priori factors specified for this model and are described in Sect. 4.2.

The five-factor solutions of the constrained vector-matrix model form only one family. In these solutions, the fifth factor resembles HOA, but has a greater contribution from m/z 43, and the characteristic “picket fence” pattern of HOA is shifted to m/z 's ≥ 43 (Fig. S9). This HOA-like spectrum may not be physically meaningful because it has only 6.5% of its signal in m/z 's < 43 , compared to 33% for the HR-MS HOA spectrum and 30% for a “standard” HOA spectrum derived from PMF analyses of fifteen urban AMS datasets (Ng et al., 2011b). Spectra with such a small fraction of signal in the low-mass fragments are not found in the AMS Spectral Database (Ulbrich et al., 2011) or in other electron impact spectra (McLafferty and Turecek, 1993). In addition, there is evidence that this factor is a splitting of the HOA factor. In fact, the sum of the size-distribution–time-series matrices from this factor and the HOA factor from this solution is very similar ($R = 0.99$) to the size-distribution–time-series matrix of HOA in the four-factor solution. Thus we conclude that this factor has split the HOA and does not constitute a new, useful factor.

The six-factor solution has two families. The family with the lower Q/Q_{exp} values has the four constrained factors, the split HOA factor found in the five-factor solution, and a factor dominated by m/z 43. This m/z 43-dominated factor is mainly a split of the OOA factor, as was observed in the 3-vector solutions. This factor also takes some mass from BBOA during the large BBOA event on 21 March. From this evidence we conclude that this factor is not physically meaningful. We therefore reject this family of solutions. In the other family of six-factor solutions, the m/z 43 spectrum is replaced by a spectrum dominated by m/z 44. The factor dominated by m/z 44 is also a split of the OOA factor, and its spectrum has even higher f_{44} (43%)

592 than the spectrum dominated by m/z 44 in the 3-vector solutions ($f_{44} = 0.39$) that was rejected as
593 nonphysical. Thus we conclude that this factor also fails the criterion of being physically
594 meaningful. We therefore reject the six-factor solutions. The seven- and eight-factor solutions
595 have more families, but also contain factors with unusual mass spectra that are not physically
596 meaningful. These solutions are also rejected.

597 Thus, only the four-factor solution has factors that are all physically meaningful. Therefore we
598 choose the four-factor solution and explore the effect of increasing β to relax the constraint on
599 the a priori spectra, as discussed in Sect.4.2.1.

600

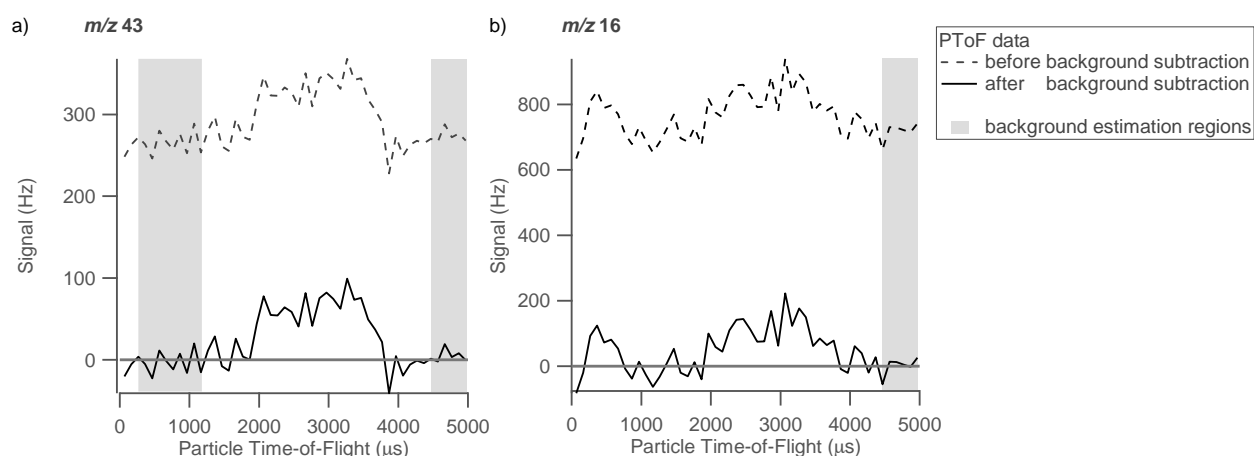


Fig. S1. Particle time-of-flight (PToF) data at **(a)** m/z 43 and **(b)** m/z 16. The raw data (dashed line) have a high background, estimated from the shaded region. The background estimation regions are selected for times before particles are expected to arrive and after no particles are expected to arrive at the vaporizer. Subtracting the average background level gives the solid line. When gas-phase signal is expected before particle arrival times, as for O^+ at m/z 16 in **(b)**, only the later background region is used for background subtraction. Data collected at **(a)** 28 March 2006, 10:32:31 LT, **(b)** 29 March 2006, 08:55:09 LT.

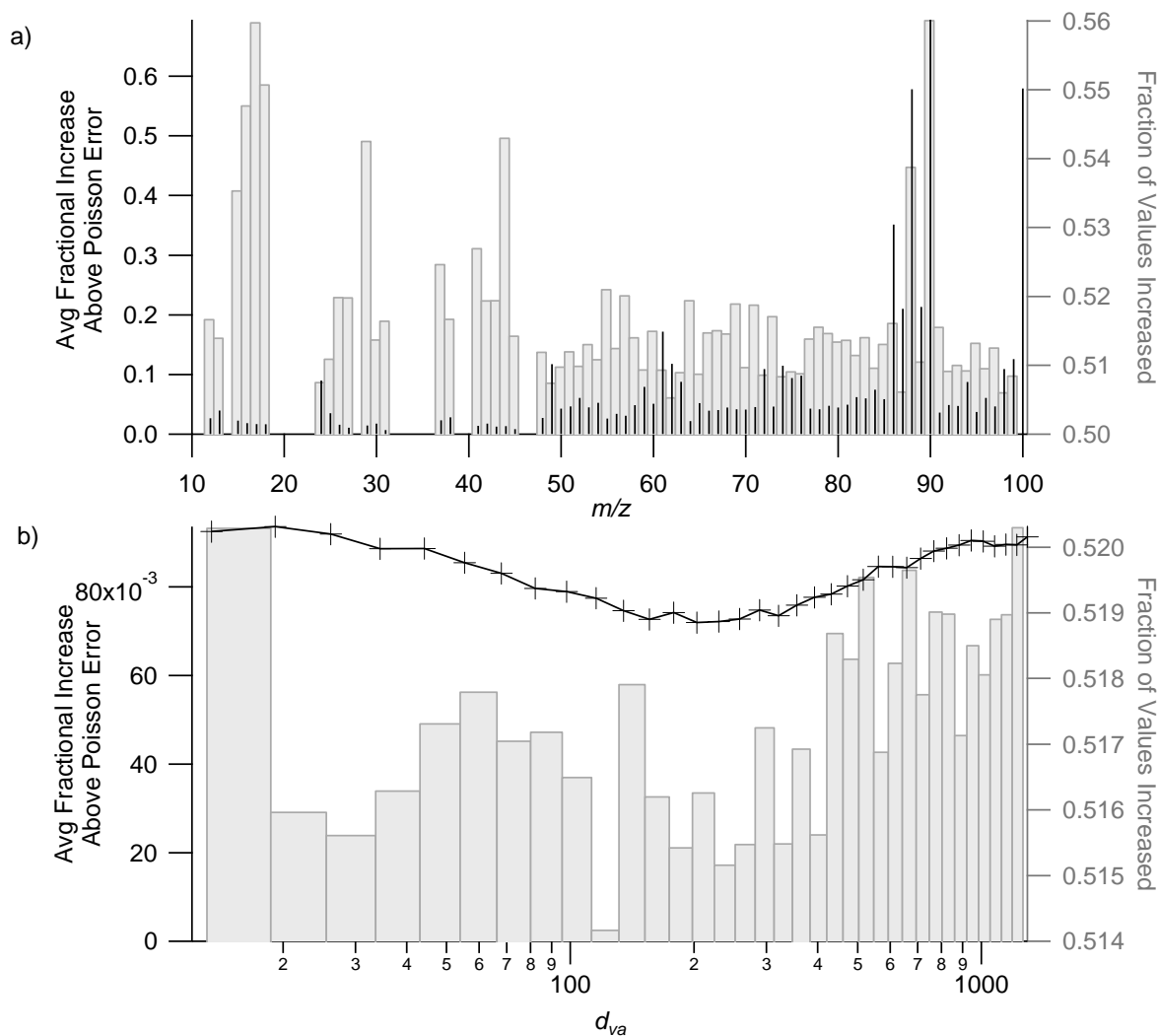


Fig. S2. Fraction of points in the 3D data matrix whose error was increased to a minimum value of 1 ion and the average increase of those values for each (a) m/z and (b) size bin. Average increases for m/z 's were highest for a few m/z 's > 85 (black lines). Errors were increased more frequently (grey bars) for particle size bins with the largest and smallest diameters.

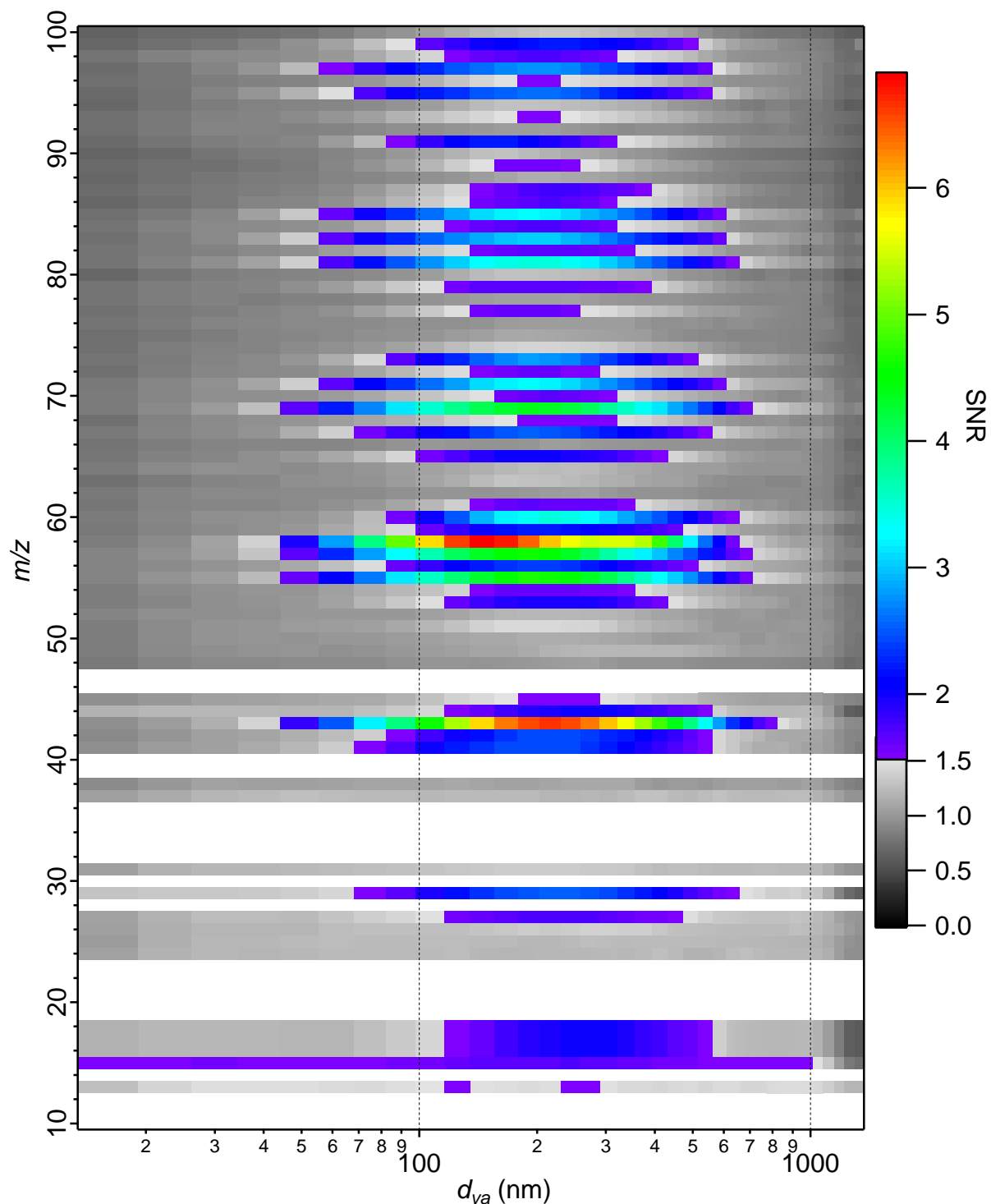
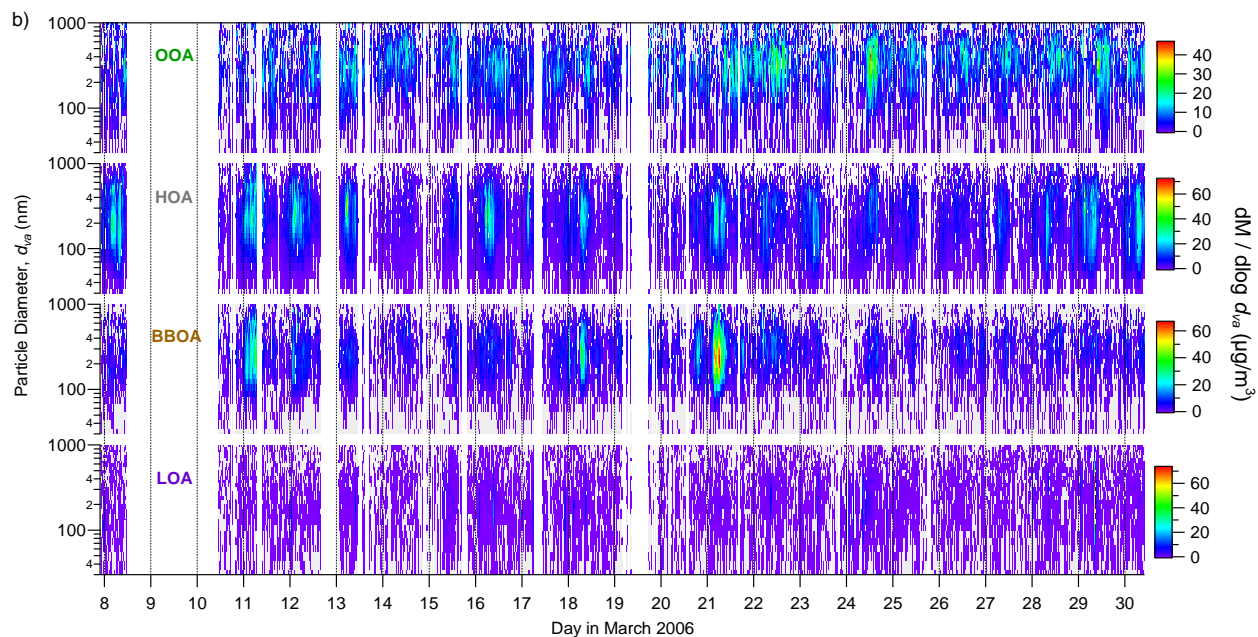


Fig. S3. Average signal-to-noise ratio (SNR) of each m/z (left axis) at each particle size (bottom axis). Size- m/z combinations are shaded by SNR. Combinations with $\text{SNR} < 1.5$ are considered “weak” and are shaded in grey. Combinations with $\text{SNR} \geq 1.5$ are considered “strong” and are shaded in color. White areas denote m/z ’s that are not assigned organic signal.

621



622

623 **Fig. S4.** Mass size distribution ($dM/d\log d_{va}$) for the best solution of the constrained vector-
 624 matrix model plotted vs. d_{va} on a log scale on the y-axis and vs. sampling date on the x-axis.
 625 Grey pixels have zero signal.

626

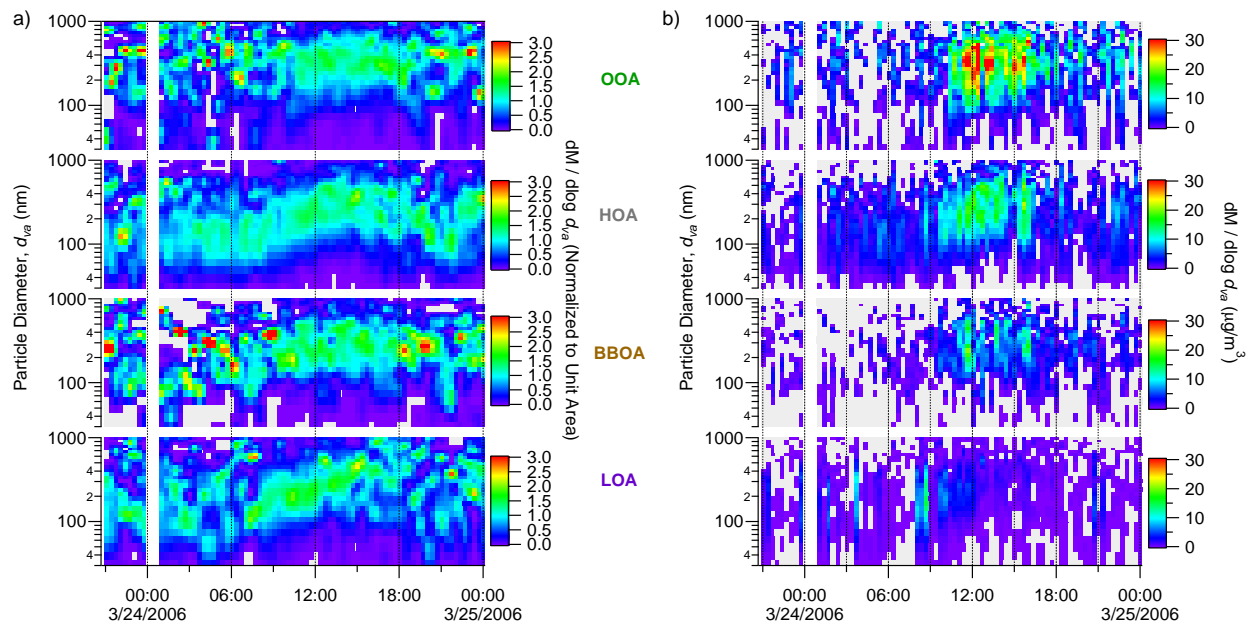
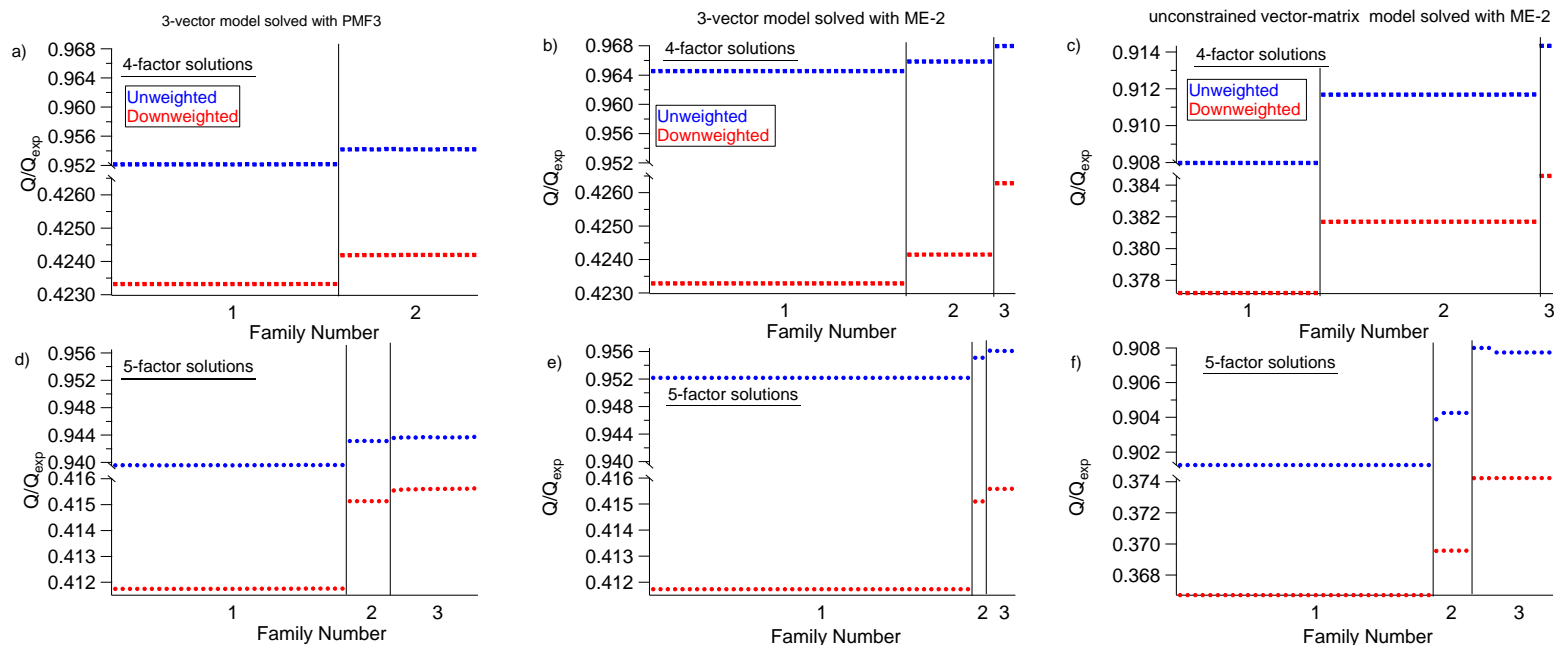


Fig. S5. Case study event on 24 March 2006. **(a)** Mass size distributions ($dM/d\log d_{va}$) of the 4 factors in the best solution of the vector-matrix model, normalized so that each size distribution has unit area, plotted vs. d_{va} on a log scale on the y-axis and vs. sampling time on the x-axis. The data have been binomially smoothed by one point each in time and size. **(b)** Mass size distributions ($dM/d\log d_{va}$) that have not been normalized, plotted vs. d_{va} on a log scale on the y-axis and vs. sampling time on the x-axis. In both panels, light-grey pixels have zero signal.



635

636 **Fig. S6.** Q/Q_{exp} values for 50 seed solutions of 4 factors (a–c) and 5 factors (d–f) grouped into families with similar factors.
 637 Unweighted Q/Q_{exp} values are in the top row of each panel (blue markers), and Q/Q_{exp} values as computed from the factorization are
 638 shown in the bottom row of each panel (red markers). Solutions were calculated for (a, d) the 3-vector model solved with PMF3, (b,
 639 e) the 3-vector model solved with ME-2, and (c, f) the unconstrained vector-matrix model solved with ME-2.

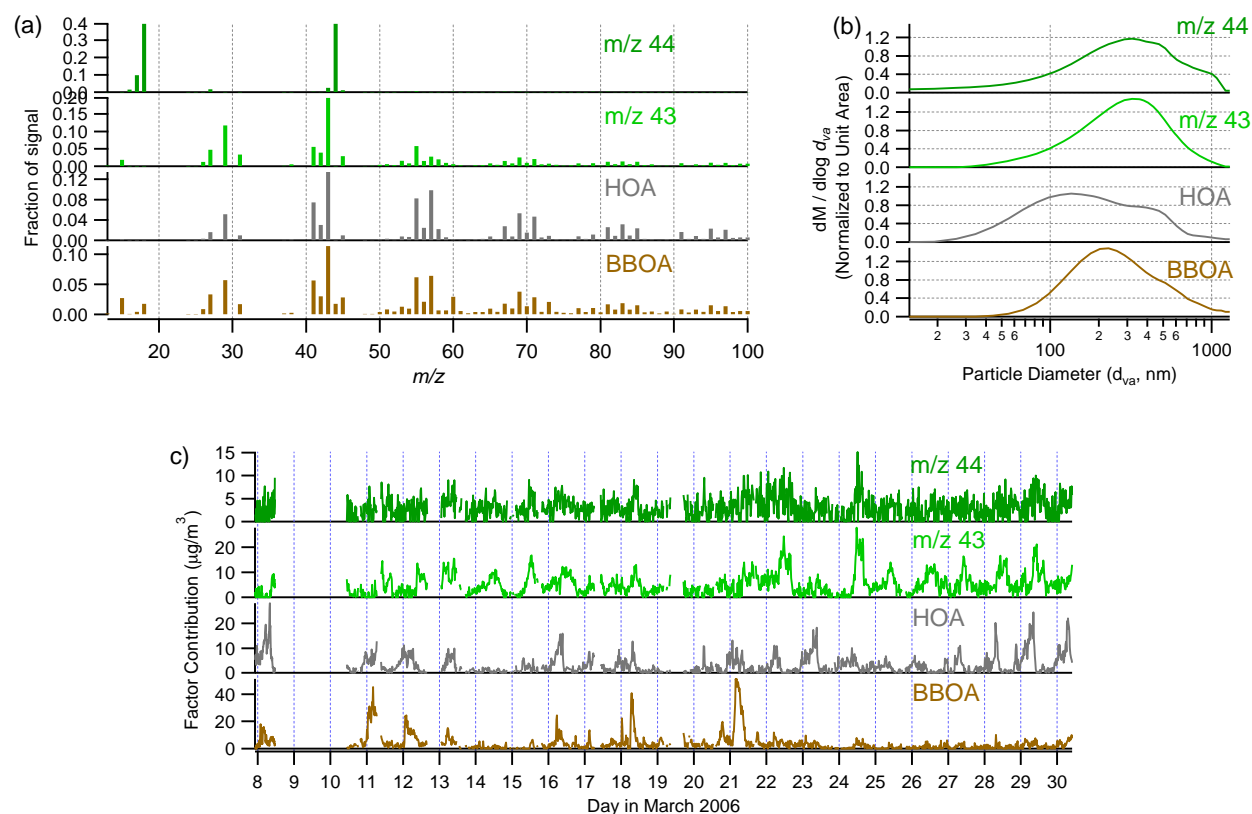


Fig. S7. Solution of the 3-vector model with 4 factors and the lowest Q/Q_{exp} values. Two of the factors have mass spectra dominated by m/z 's 44 and 43, and two of the factors represent HOA and biomass-burning organic aerosol BBOA. (a) Mass spectrum of each factor plotted vs. ion mass-to-charge ratio (m/z). Mass spectra are normalized to sum to 1. (b) Mass size distribution ($dM/d\log d_{va}$) plotted vs. particle vacuum-aerodynamic diameter (d_{va}) on a log scale. Size distributions are normalized so that the area under each curve sums to 1. (c) Mass contribution of each factor plotted vs. sampling date.

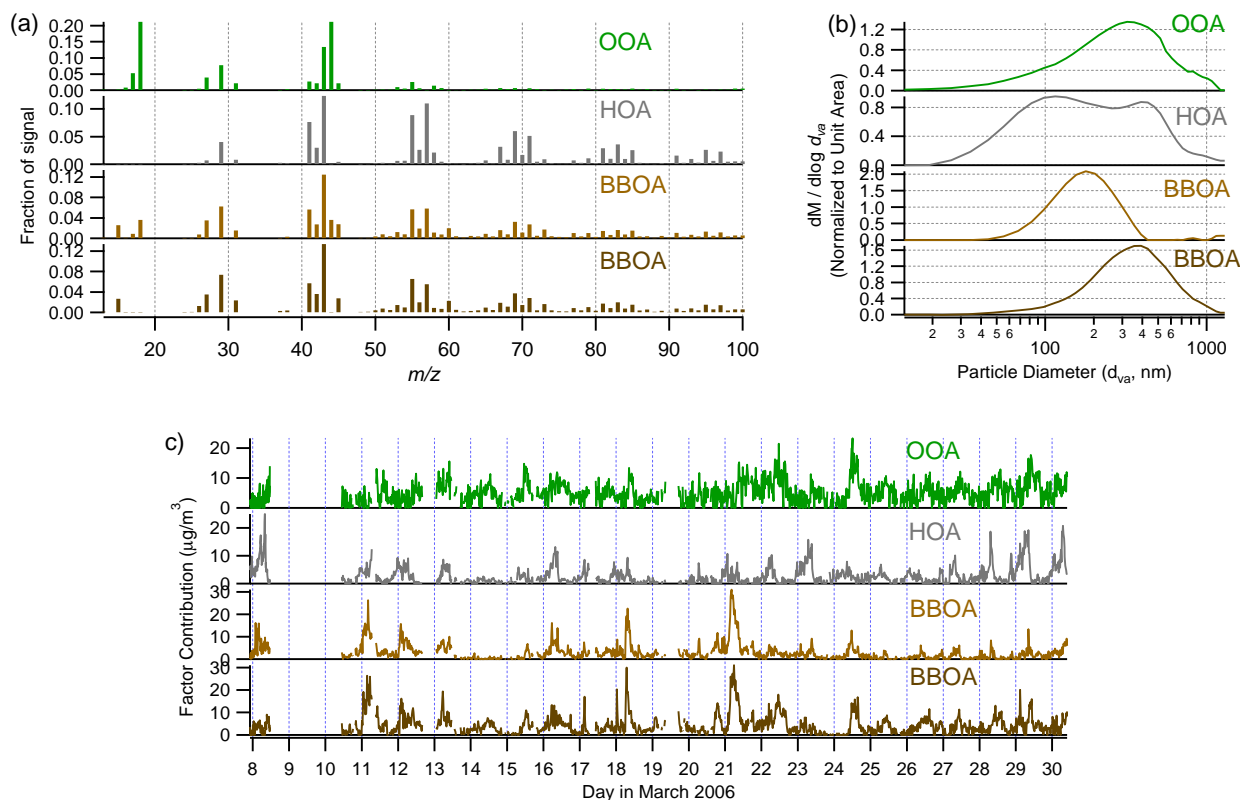


Fig. S8. Solution of the 3-vector model with 4 factors that include two with similar mass spectra. The four factors are OOA, HOA, and two BBOA factors. (a) Mass spectrum of each factor plotted vs. ion mass-to-charge ratio (m/z). Mass spectra are normalized to sum to 1. (b) Mass size distribution ($dM/d\log d_{va}$) plotted vs. particle vacuum-aerodynamic diameter (d_{va}) on a log scale. Size distributions are normalized so that the area under each curve sums to 1. (c) Mass contribution of each factor plotted vs. sampling date.

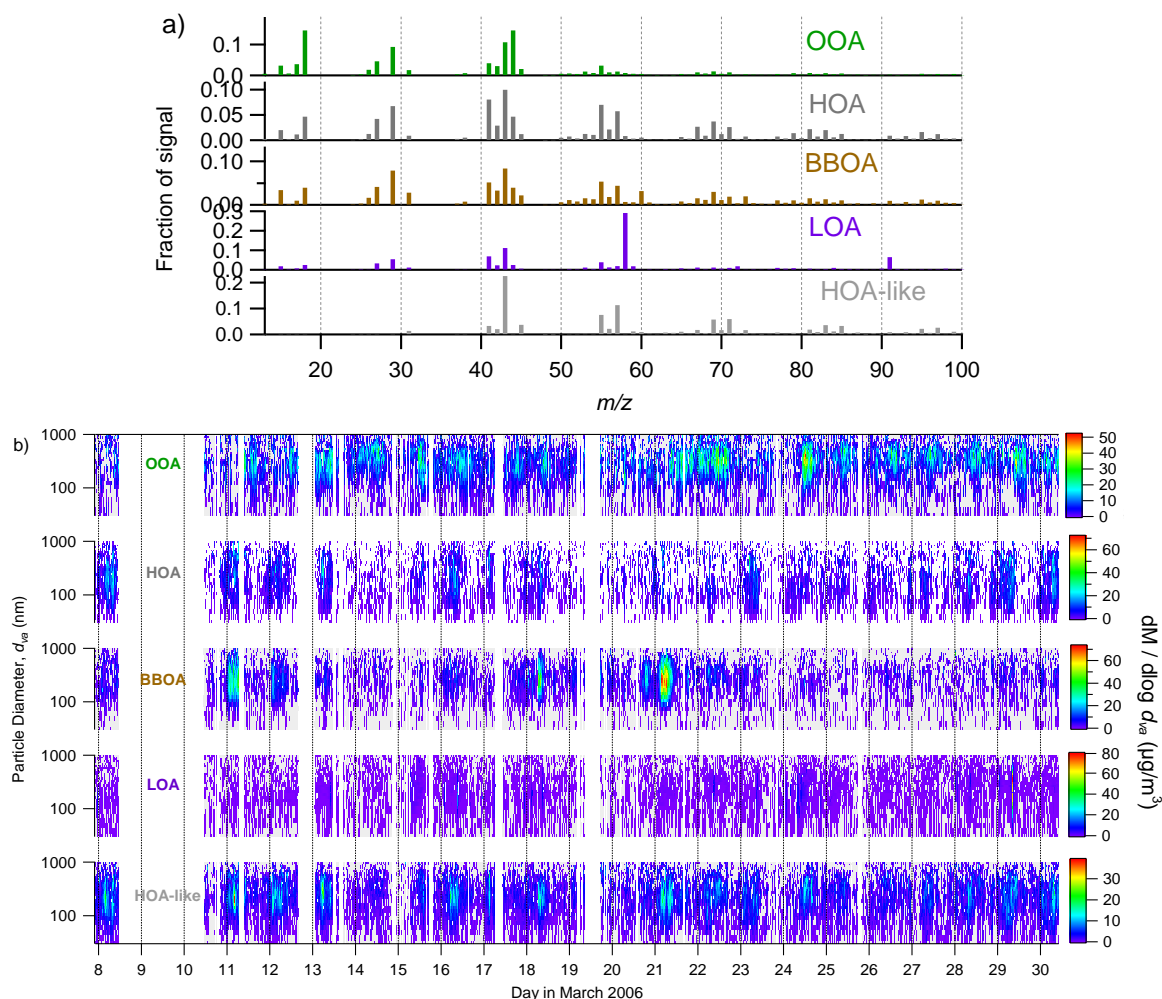


Fig. S9. Five-factor solution of the constrained vector-matrix model. Four a priori mass spectra were provided as starting guesses: OOA, HOA, and BBOA from the HR-MS solution, and LOA from the best solution of the 3-vector model (Fig. 3). The a priori spectra were not allowed to vary ($\beta = 0$). The fifth factor has a mass spectrum that is HOA-like, but with the signal shifted to m/z 's ≥ 43 . (a) Mass spectrum of each factor plotted vs. m/z . Mass spectra are normalized to sum to 1. (b) Mass size distribution ($dM/d\log d_{va}$) plotted vs. d_{va} on a log scale on the y-axis and vs. sampling date on the x-axis. Grey pixels have zero signal.

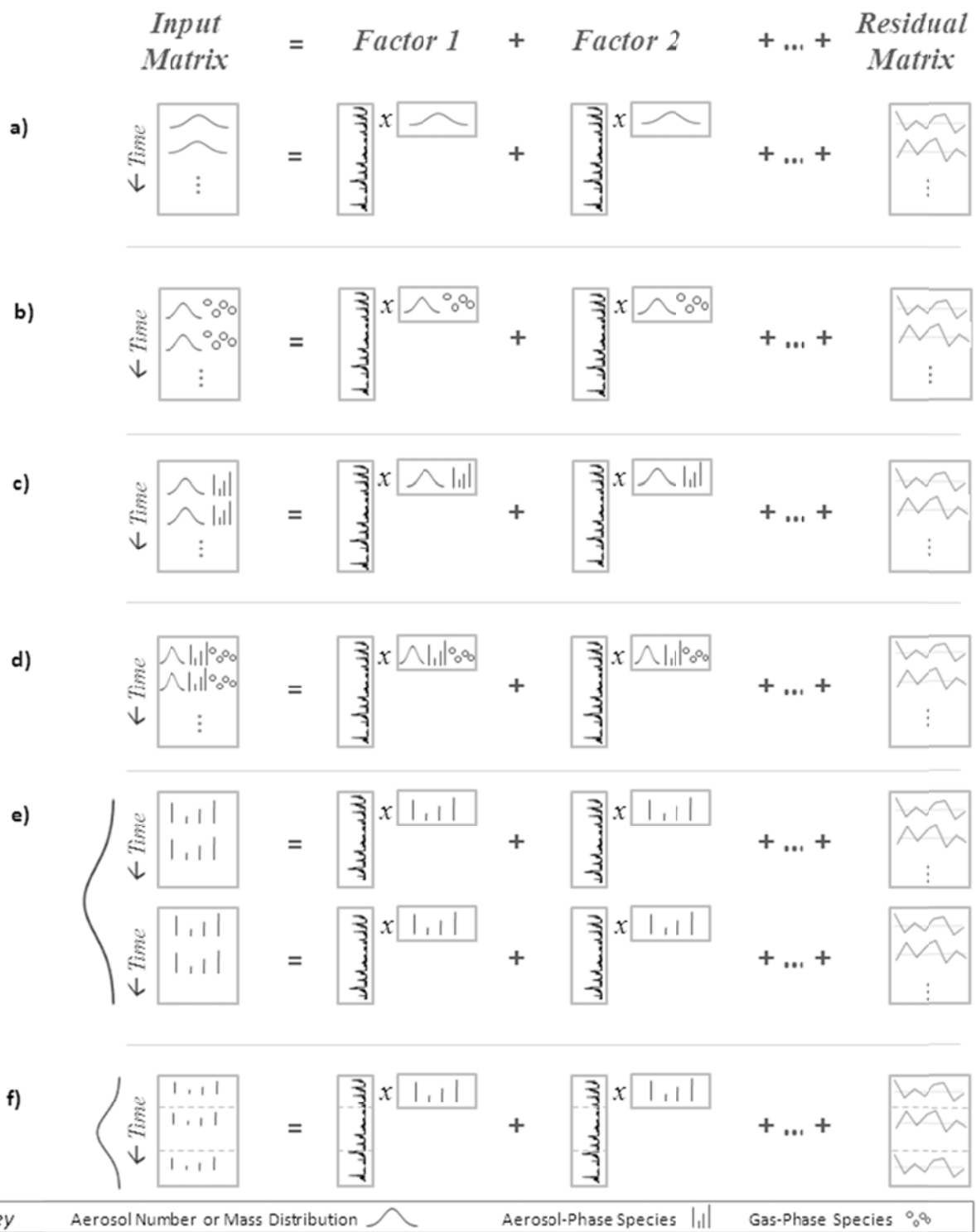


Fig. S10. Schematic representation of factorizations of datasets that include size distribution information (a–d) or size-resolved aerosol composition (e–f). Studies that included these factorizations are summarized in Table S1. Details are given in Tables S2–S7.

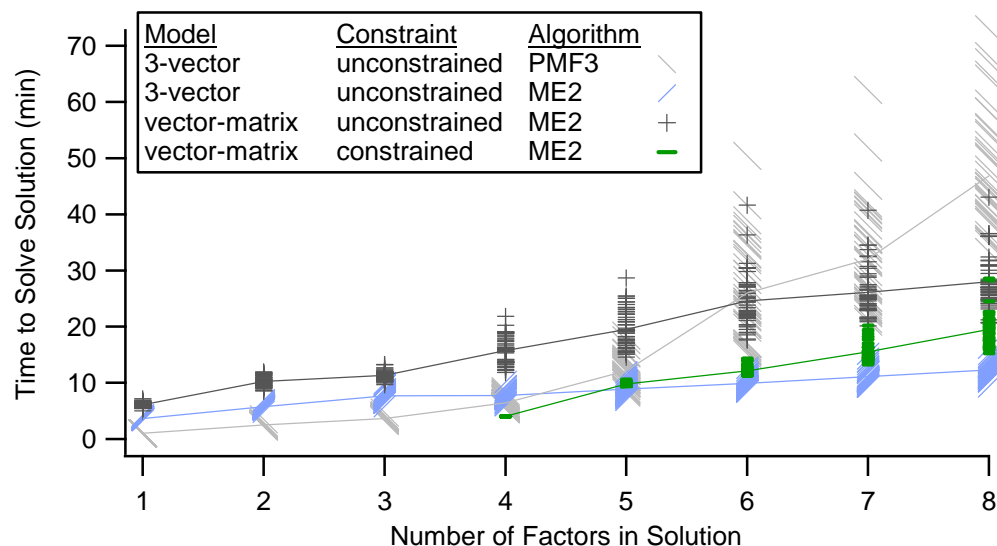


Fig. S11. Time to calculate each solution of the 3-vector and vector-matrix models of our 3.5×10^6 element matrix with 1 to 8 factors using the PMF3 or ME-2 algorithm. Each mark show the time to calculate a solution from a starting seed. The calculation of each solution begins with loading the data and error matrices; this takes ~ 1 minute. Details about the computer used for these calculations and the values of some ME-2 variables are shown in Table S8. Values are meant only as a rough guide because many elements of the script file can impact the speed of calculation, and we generally did not try to optimize the running speed.

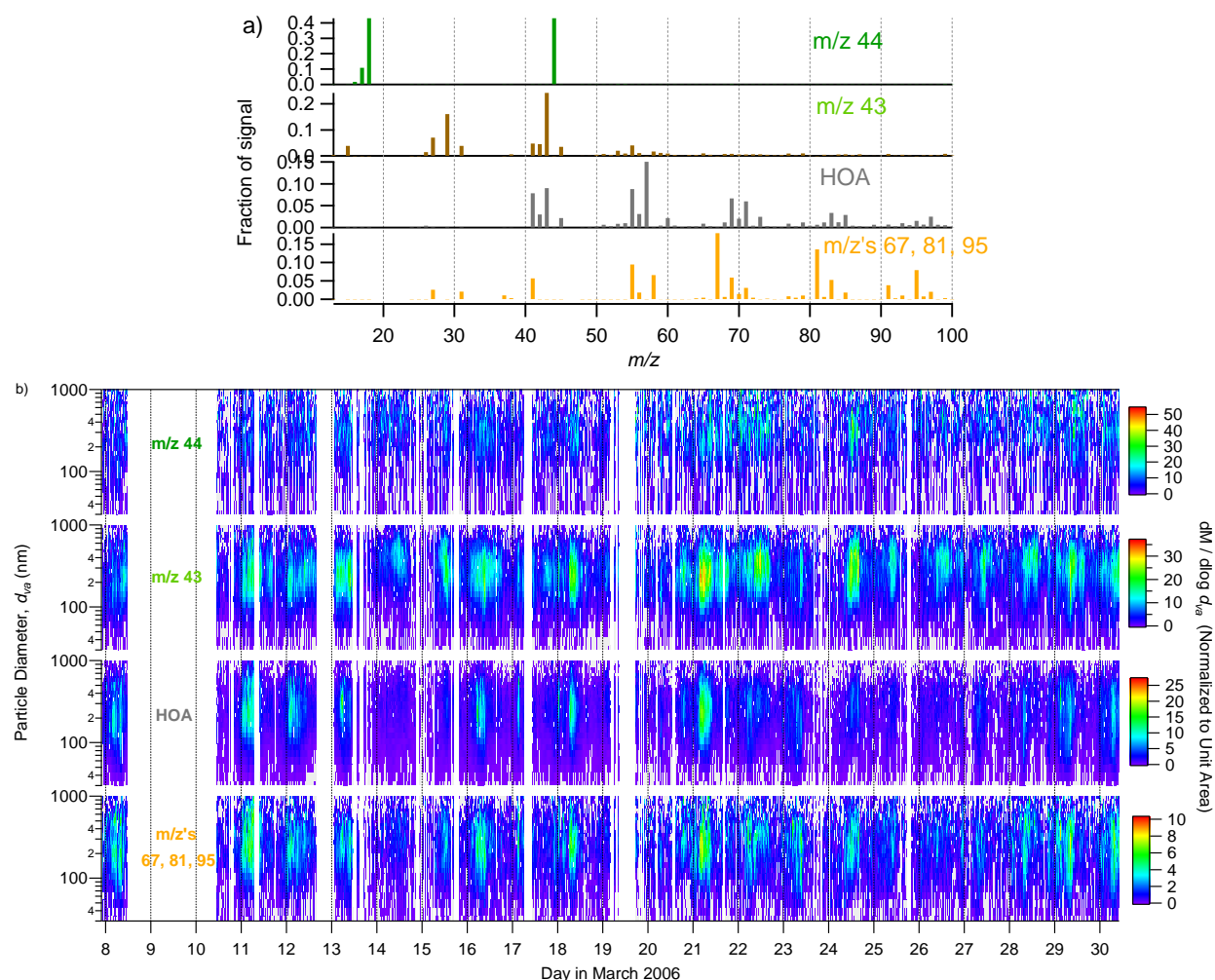











Fig. S12. Four-factor solution of the unconstrained vector-matrix model, including a factor with a mass spectrum dominated by m/z 's 67, 81, and 95. **(a)** Mass spectrum of each factor plotted vs. m/z . Mass spectra are normalized to sum to 1. **(b)** Mass size distribution ($dM/d\log d_{va}$) plotted vs. d_{va} on a log scale on the y-axis and vs. sampling date on the x-axis. Grey pixels have zero signal.

Table S1. Summary of research that applied 2-dimensional (2D) factorization techniques to aerosol size distributions (first four groups) or size-resolved chemical composition of aerosols (last two groups). Details are given in Tables S3–8, and schematic representations of the matrix factorization are shown in Fig. S10.

Matrix-Row Representation	Matrix-Row Contents	Citation
	Total size distribution (number, surface area, or mass)	(Kim et al., 2004) (Zhou et al., 2004) (Zhou et al., 2005b) (Ogulei et al., 2006a) (Ogulei et al., 2007b) (Chan and Mozurkewich, 2007a) (Yue et al., 2008) (Beddows et al., 2009) (Costabile et al., 2009)
	Total size distribution Gas-phase speciation	(Wahlin et al., 2001) (Thimmaiah et al., 2009)
	Total size distribution Chemical speciation of the bulk aerosol	(Ruuskanen et al., 2001) (Larson et al., 2006)
	Total size distribution Chemical speciation of the bulk aerosol Gas-phase speciation	(Zhou et al., 2005a) (Ogulei et al., 2006b) (Ogulei et al., 2007a)
	Chemical speciation of size-resolved aerosol, each size factored separately	(Yakovleva et al., 1999) ^a (Dillner et al., 2005) (Han et al., 2006) (Yatkin and Bayram, 2008) (Gietl and Klemm, 2009) (Karanasiou et al., 2009) ^a (Kleeman et al., 2009) (Srivastava et al., 2009)
	Chemical speciation of size-resolved aerosol, all sizes factored together	(Amato et al., 2009)

Key Aerosol Number or Mass Distribution  Aerosol-Phase Species  Gas-Phase Species 

^aIn these studies, the data were also arranged as a 3-dimensional (3D) matrix and factored using a 3D model. Details of these datasets and models are presented in Table 1.

Sampling Location and Year	Citation	Instrumentation	Chemical Speciation	Sizes (μm)	Sampling Time	Factorization Techniques
Seattle, WA 2000-2001	(Kim et al., 2004)	DMPS		0.02–0.4	1 hr	PMF, UNMIX
Pittsburgh, PA 2001	(Zhou et al., 2004)	SMPS APS		0.003–2.5	15 min	PMF
Pittsburgh, PA 2001-2002	(Zhou et al., 2005b)	SMPS APS		0.003–2.5	15 min	PMF
Reston, VA (indoors) 1999-2000	(Ogulei et al., 2006a)	SMPS APS		0.01–20	30 min	PMF
Egbert, ON Vancouver, BC Hamilton, ON Simcoe, ON 1999-2003	(Chan and Mozurkewich, 2007a)	SMPS		0.006–0.3	5 min	APCA
Buffalo, NY 2004	(Ogulei et al., 2007b)	EEPS		0.006–0.3	1-5 sec	PMF
Erfurt, Germany 1997-2001	(Yue et al., 2008)	DMPS OLAS		0.01 – 3.0	1 hr	PMF
London, UK 2005	(Beddows et al., 2009)	SMPS		0.012–0.437, 0.015–0.661	1 hr, 6 hr	Cluster Analysis
Leipzig, Germany 2005-2006	(Costabile et al., 2009)	TDMPS SMPS		0.003–0.9	30 min	PCA

b)

Acronym	Sampling Technique
APS	Aerodynamic Particle Sizer
DMPS	Differential Mobility Particle Sizer
EEPS	Engine Exhaust Particle Spectrometer
OLAS	Optical Laser Aerosol Spectrometer
SMPS	Scanning Mobility Particle Sizer
TDMPS	Twin Differential Mobility Particle Sizer

c)

Acronym	Factorization Technique
APCA	Absolute Principal Component Analysis
PCA	Principal Component Analysis
PMF	Positive Matrix Factorization
UNMIX	Unmix multivariate receptor model

Table S2. Details of research that reports application of 2D factorization techniques to aerosol size distributions. Table S2b expands the acronyms of instrumental techniques shown in Table S2a. Table S2c expands the acronyms of factorization techniques shown in Table S2a.

Sampling Location and Year	Citation	Instrumentation	Chemical Speciation	Sizes (µm)	Sampling Time	Factorization Techniques
Copenhagen and Odense, Denmark 1999	(Wahlin et al., 2001)	DMPS CPC	CO, NO _x	0.006–0.7	30 min	COPREM
Prague, Czech Republic 2008	(Thimmaiah et al., 2009)	SMPS	CO, NO _x , SO ₂ , O ₃ , CH ₄	0.018–0.723	1 hr	PMF

b)

Acronym	Sampling Technique
CPC	Condensation Particle Counter
DMPS	Differential Mobility Particle Sizer
SMPS	Scanning Mobility Particle Sizer

c)

Acronym	Factorization Technique
COPREM	Constrained Physical Receptor Model
PMF	Positive Matrix Factorization

Table S3. Details of research that reports application of 2D factorization techniques to datasets pairing aerosol size distributions and gas-phase composition. Table S3b expands the acronyms of instrumental techniques shown in Table S3a. Table S3c expands the acronyms of factorization techniques shown in Table S3a.

Sampling Location and Year	Citation	Instrumentation	Chemical Speciation	Sizes (μm)	Sampling Time	Factorization Techniques
Akjnaar, Netherlands Erfurt, Germany Helsinki, Finland 1996-1997	(Ruuskanen et al., 2001)	DMPS SMPS OLAS EAS CPC Harvard Impactor	PM _{2.5} mass Absorbance	0.01–10	1 hr	PCA
Seattle, WA 2000-2003	(Larson et al., 2006)	DMPS APS STN	PM _{2.5} mass NH ₄ ⁺ , NO ₃ ⁻ , SO ₄ ²⁻ , K ⁺ , Na ⁺ 17 elements EC, OC	0.02–5	24 hr	PMF with additional constraints, solved with ME-2

b)

Acronym	Sampling Technique
APS	Aerodynamic Particle Sizer
CPC	Condensation Particle Counter
DMPS	Differential Mobility Particle Sizer
EAS	Electrical Aerosol Spectrometer
OLAS	Optical Laser Aerosol Spectrometer
SMPS	Scanning Mobility Particle Sizer
STN	Speciation Trends Network

PCA	Principal Component Analysis
PMF	Positive Matrix Factorization
ME-2	Multilinear Engine 2

c)

Acronym	Factorization Technique
---------	-------------------------

Table S4. Details of research that reports application of 2D factorization techniques to datasets pairing aerosol size distributions and particle-phase composition. Table S4b expands the acronyms of instrumental techniques shown in Table S4a. Table S4c expands the acronyms of factorization techniques shown in Table S4a.

Sampling Location and Year	Citation	Instrumentation	Chemical Speciation	Sizes (μm)	Sampling Time	Factorization Techniques
Pittsburgh, PA 2001	(Zhou et al., 2005a)	SMPS APS	SO_4^{2-} , NO_3^- 11 elements O_3 , NO , NO_x , SO_2 , CO	0.003–2.5	15 min	PLS, PMF
Baltimore, MD 2002	(Ogulei et al., 2006b)	SMPS APS SEAS	$\text{PM}_{2.5}$ mass NO_3^- , SO_4^{2-} EC, OC 11 elements CO , NO , NO_2 , O_3	0.00965–2.458	1 hr	PLS, PMF
Rochester, NY 2004-2005	(Ogulei et al., 2007a)	SMPS	$\text{PM}_{2.5}$ mass CO , O_3 , SO_2	0.012–0.470	1 hr	PMF

b)

Acronym Sampling Technique

APS Aerodynamic Particle Sizer
 SMPS Scanning Mobility Particle Sizer
 SEAS Semi-continuous Elements in Aerosol

c)

Acronym Factorization Technique

PLS Partial Least Squares
 PMF Positive Matrix Factorization

Table S5. Details of research that reports application of 2D factorization techniques to datasets that combine aerosol size distributions with gas- and particle-phase composition. Table S5b expands the acronyms of instrumental techniques shown in Table S5a. Table S5c expands the acronyms of factorization techniques shown in Table S5a.

Sampling Location and Year	Citation	Instrumentation	Chemical Speciation	Sizes (μm)	Sampling Time	Factorization Techniques
Riverside, CA 1991	(Yakovleva et al., 1999) ^a	Stationary indoor monitors Stationary ambient monitors Personal exposure monitors	18 elements	PM _{2.5} , PM ₁₀ , personal PM ₁₀ (2 sizes, 5 types)	12 hr	PMF
Houston, TX 2000	(Dillner et al., 2005)	MOUDI	SO ₄ ²⁻ , NO ₃ ⁻ , Cl ⁻ , NH ₄ ⁺ EC, OM 32 elements	0.056–1.8 (6 ranges)	24 hr	Cluster Analysis
Gosan, Korea 2002	(Han et al., 2006)	DRUM	19 elements	0.07–12 (8 ranges)	3 hr	PMF
Izmir, Turkey 2004-2005	(Yatkin and Bayram, 2008)	Dichotomous sampler	16 elements	PM _{2.5} , PM ₁₀ (2 sizes)	24 hr	PMF, CMB
Münster, Germany 2006-2007	(Gietl and Klemm, 2009)	Berner impactor	SO ₄ ²⁻ , NO ₃ ⁻ , Cl ⁻ , NH ₄ ⁺ , Na ⁺ , Ca ²⁺ , Mg ²⁺ EC, OC	0.053–10 (5 ranges)	5-7.5 hr	PMF
Athens, Greece 2002	(Karanasiou et al., 2009) ^a	Custom impactors aethalometer	SO ₄ ²⁻ BC 13 elements	PM ₂ , PM ₁₀₋₂	24 hr	PMF

Sampling Location and Year	Citation	Instrumentation	Chemical Speciation	Sizes (μm)	Sampling Time	Factorization Techniques
Sacramento, Modesto, and Bakersfield, CA 2000-2001	(Kleeman et al., 2009)	MOUDI	8 molecular organic tracers	0.055–1.8 (6 ranges)	8, 12 hr	Custom source apportionment algorithm
Delhi, India 2005-2006	(Srivastava et al., 2009)	Cascade Impactor	11 metals	0.7–10.9 (2 ranges)	24 hr	CMB PCA

b)

Acronym	Sampling Technique
DRUM	Davis Rotating Unit for Monitoring
MOUDI	Micro-Orifice, Uniform Deposit Impactor

c)

Acronym	Factorization Technique
CMB	Chemical Mass Balance
PCA	Principal Component Analysis
PMF	Positive Matrix Factorization

Table S6. Details of research that reports application of 2D factorization techniques to size-resolved aerosol composition data. In these studies, the aerosol composition from each size range is factored separately. Table S6b expands the acronyms of instrumental techniques shown in Table S6a. Table S6c expands the acronyms of factorization techniques shown in Table S6a.

^a In these works, the data was also arranged as a 3D matrix and factored using a 3D model; details of these datasets and models are presented in Table 1.

Sampling Location and Year	Citation	Instrumentation	Chemical Speciation	Sizes (μm)	Sampling Time	Factorization Techniques
Barcelona, Spain 2003-2007	(Amato et al., 2009)	High-volume samplers	NO_3^- , Cl^- , NH_4^+ Total carbon 22 elements	PM_{10} , $\text{PM}_{2.5}$, PM_1	24 hr	PMF with pulling equations in ME-2

b)

Acronym	Factorization Technique
ME-2	Multilinear Engine 2
PMF	Positive Matrix Factorization

Table S7. Details of research that reports application of 2D factorization techniques to size-resolved aerosol composition data. In these studies, the aerosol composition from all size ranges is combined in one matrix and factored simultaneously. Table S7b expands the acronyms of factorization techniques shown in Table S7a.

Table S8. Configuration of the computer and ME-2 algorithm used to solve the factorizations. No attempt was made to adjust the settings in the ME-2 control file to increase the speed of the calculations.

Computer Specifications	
RAM	3.00 GB
Processor	Intel® Xeon™ 3.2 GHz
Operating System	Windows XP
ME-2 Control File Specifications	
Convergence test level 1	$Q_{exp} \times 10^{-4}$
Convergence test level 2	$Q_{exp} \times 2 \times 10^{-5}$
Convergence test level 3	$Q_{exp} \times 10^{-5}$
cgresets	10, 80, 1, 1, 2, 1
Precondition mode	5

References for Supplemental Information

- 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., Stone, E. A., Schauer, J. J., Volkamer, R., Fortner, E., De Foy, B., Wang, J., Laskin, A., Shutthanandan, V., Zheng, J., Zhang, R., Gaffney, J. S., 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, 6949–6981, doi:10.5194/acp-9-6633-2009, 2009.
- 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.
- Beddows, D. C. S., Dall'Osto, M., and Harrison, R. M.: Cluster Analysis of Rural, Urban, and Curbside Atmospheric Particle Size Data, *Environ. Sci. Technol.*, 43, 4694–4700, doi: 10.1021/es803121t, 2009.
- Bezdek, J. C., Coray, C., Gunderson, R., and Watson, J.: Detection and Characterization of Cluster Substructure .1. Linear Structure - Fuzzy C-Lines, *SIAM J. Appl. Math.*, 40, 339–357, 1981.
- Chan, T. W., and Mozurkewich, M.: Simplified Representation of Atmospheric Aerosol Size Distributions Using Absolute Principal Component Analysis, *Atmos. Chem. Phys.*, 7, 875–886, doi:10.5194/acp-7-875-2007, 2007a.
- Chan, T. W., and Mozurkewich, M.: Application of Absolute Principal Component Analysis to Size Distribution Data: Identification of Particle Origins, *Atmos. Chem. Phys.*, 7, 887–897, doi:10.5194/acp-7-887-2007, 2007b.
- Costabile, F., Birmili, W., Klose, S., Tuch, T., Wehner, B., Wiedensohler, A., Franck, U., Konig, K., and Sonntag, A.: Spatio-Temporal Variability and Principal Components of the Particle Number Size Distribution in an Urban Atmosphere, *Atmos. Chem. Phys.*, 9, 3163–3195, doi:10.5194/acp-9-3163-2009, 2009.
- 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.

- Friedlander, S. K.: Chemical Element Balances and Identification of Air-Pollution Sources, *Environ. Sci. Technol.*, 7, 235–240, 1973.
- Gietl, J. K., and Klemm, O.: Source Identification of Size-Segregated Aerosol in Munster, Germany, by Factor Analysis, *Aerosol Sci. Technol.*, 43, 828–837, doi: 10.1080/02786820902953923, 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.
- 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.
- Kim, E., Hopke, P. K., Larson, T. V., and Covert, D. S.: Analysis of Ambient Particle Size Distributions Using Unmix and Positive Matrix Factorization, *Environ. Sci. Technol.*, 38, 202–209, doi: 10.1021/es030310s, 2004.
- 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.
- Kroll, J. H., Donahue, N. M., Jimenez, J. L., Kessler, S. H., Canagaratna, M. R., Wilson, K. R., Altieri, K. E., Mazzoleni, L. R., Wozniak, A. S., Bluhm, H., Mysak, E. R., Smith, J. D., Kolb, C. E., and Worsnop, D. R.: Carbon Oxidation State as a Metric for Describing the Chemistry of Atmospheric Organic Aerosol, *Nature Chemistry*, 3, 133–139, doi:10.1038/nchem.948, 2011.
- Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prevot, A. S. H.: Source Apportionment of Submicron Organic Aerosols at an Urban Site by Factor Analytical Modelling of Aerosol Mass Spectra, *Atmos. Chem. Phys.*, 7, 1503–1522, doi:10.5194/acp-7-1503-2007, 2007.
- 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.
- Malinowski, E.: Factor Analysis in Chemistry, 2nd Ed. ed., Wiley & Sons, New York, 1991.
- Marcolli, C., Canagaratna, M. R., Worsnop, D. R., Bahreini, R., de Gouw, J. A., Warneke, C., Goldan, P. D., Kuster, W. C., Williams, E. J., Lerner, B. M., Roberts, J. M., Meagher, J. F., Fehsenfeld, F. C., Marchewka, M., Bertman, S. B., and Middlebrook, A. M.: Cluster Analysis of the Organic Peaks in Bulk Mass Spectra Obtained During the 2002 New England Air Quality Study with an Aerodyne Aerosol Mass Spectrometer, *Atmos. Chem. Phys.*, 6, 5649–5666, doi:10.5194/acp-6-5649-2006, 2006.
- McLafferty, F., and Turecek, F.: Interpretation of Mass Spectra, 4th Ed. ed., University Science Books, Mill Valley, CA, 1993.
- Murphy, D. M., Middlebrook, A. M., and Warshawsky, M.: Cluster Analysis of Data from the Particle Analysis by Laser Mass Spectrometry (Palms) Instrument, *Aerosol Sci. Technol.*, 37, 382–391, doi: 10.1080/02786820390125241, 2003.
- Ng, N. L., Canagaratna, M. R., Zhang, Q., Jimenez, J. L., Tian, J., Ulbrich, I. M., Kroll, J. H., Docherty, K. S., Chhabra, P. S., Bahreini, R., Murphy, S. M., Seinfeld, J. H., Hildebrandt,

- L., Donahue, N. M., DeCarlo, P. F., Lanz, V. A., Prevot, A. S. H., Dinar, E., Rudich, Y., and Worsnop, D. R.: Organic Aerosol Components Observed in Northern Hemispheric Datasets from Aerosol Mass Spectrometry, *Atmos. Chem. Phys.*, 10, 4625–4641, doi: 10.5194/acp-10-4625-2010, 2010.
- Ng, N. L., Canagaratna, M. R., Jimenez, J. L., Chhabra, P. S., Seinfeld, J. H., and Worsnop, D. R.: Changes in Organic Aerosol Composition with Aging Inferred from Aerosol Mass Spectra, *Atmos. Chem. Phys. Discuss.*, 11, 7095–7112, 10.5194/acpd-11-7095-2011, 2011a.
- 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, 2011b.
- Ogulei, D., Hopke, P. K., and Wallace, L. A.: Analysis of Indoor Particle Size Distributions in an Occupied Townhouse Using Positive Matrix Factorization, *Indoor Air*, 16, 204–215, doi: 10.1111/j.1600-0668.2006.00418.x, 2006a.
- Ogulei, D., Hopke, P. K., Zhou, L. M., Pancras, J. P., Nair, N., and Ondov, J. M.: Source Apportionment of Baltimore Aerosol from Combined Size Distribution and Chemical Composition Data, *Atmos. Environ.*, 40, S396–S410, doi: 10.1016/j.atmosenv.2005.11.075, 2006b.
- Ogulei, D., Hopke, P. K., Chalupa, D. C., and Utell, M. J.: Modeling Source Contributions to Submicron Particle Number Concentrations Measured in Rochester, New York, *Aerosol Sci. Technol.*, 41, 179–201, doi: 10.1080/02786820601116012, 2007a.
- Ogulei, D., Hopke, P. K., Ferro, A. R., and Jaques, P. A.: Factor Analysis of Submicron Particle Size Distributions near a Major United States-Canada Trade Bridge, *J. Air & Waste Manage. Assoc.*, 57, 190–203, 2007b.
- Paatero, P.: Least Squares Formulation of Robust Non-Negative Factor Analysis, *Chemom. Intell. Lab. Syst.*, 37, 23–35, 1997.
- Paatero, P.: The Multilinear Engine - a Table-Driven, Least Squares Program for Solving Multilinear Problems, Including the N-Way Parallel Factor Analysis Model, *J. Comput. Graph. Stat.*, 8, 854–888, 1999.
- 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.
- Ruuskanen, J., Tuch, T., Ten Brink, H., Peters, A., Khlystov, A., Mirme, A., Kos, G. P. A., Brunekreef, B., Wichmann, H. E., Buzorius, G., Vallius, M., Kreyling, W. G., and Pekkanen, J.: Concentrations of Ultrafine, Fine and PM_{2.5} Particles in Three European Cities, *Atmos. Environ.*, 35, 3729–3738, 2001.
- Seinfeld, J. H., and Pandis, S. N.: *Atmospheric Chemistry and Physics*, John Wiley & Sons, New York, 1998.
- 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.
- Thimmaiah, D., Hovorka, J., and Hopke, P. K.: Source Apportionment of Winter Submicron Prague Aerosols from Combined Particle Number Size Distribution and Gaseous Composition Data, *Aerosol Air Qual. Res.*, 9, 209–236, 2009.

- 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.
- Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of Organic Components from Positive Matrix Factorization of Aerosol Mass Spectrometric Data, *Atmos. Chem. Phys.*, 9, 2891–2918, doi:10.5194/acp-9-2891-2009, 2009.
- AMS Spectral Database: <http://cires.colorado.edu/jimenez-group/AMSsd>, access: 6 April, 2011.
- Wahlin, P., Palmgren, F., and Van Dingenen, R.: Experimental Studies of Ultrafine Particles in Streets and the Relationship to Traffic, *Atmos. Environ.*, 35, S63–S69, 2001.
- 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.
- Yatkin, S., and Bayram, A.: Source Apportionment of PM₁₀ and PM_{2.5} 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.
- Yue, W., Stolzel, M., Cyrys, J., Pitz, M., Heinrich, J., Kreyling, W. G., Wichmann, H. E., Peters, A., Wang, S., and Hopke, P. K.: Source Apportionment of Ambient Fine Particle Size Distribution Using Positive Matrix Factorization in Erfurt, Germany, *Sci. Total Environ.*, 398, 133–144, doi: 10.1016/j.scitotenv.2008.02.049, 2008.
- Zhang, Q., Alfarra, M. R., Worsnop, D. R., Allan, J. D., Coe, H., Canagaratna, M. R., and Jimenez, J. L.: Deconvolution and Quantification of Hydrocarbon-Like and Oxygenated Organic Aerosols Based on Aerosol Mass Spectrometry, *Environ. Sci. Technol.*, 39, 4938–4952, doi: 10.1021/es048568I, 2005a.
- Zhang, Q., Worsnop, D. R., Canagaratna, M. R., and Jimenez, J. L.: Hydrocarbon-Like and Oxygenated Organic Aerosols in Pittsburgh: Insights into Sources and Processes of Organic Aerosols, *Atmos. Chem. Phys.*, 5, 3289–3311, doi:10.5194/acp-5-3289-2005, 2005b.
- Zhou, L. M., Kim, E., Hopke, P. K., Stanier, C. O., and Pandis, S.: Advanced Factor Analysis on Pittsburgh Particle Size-Distribution Data, *Aerosol Sci. Technol.*, 38, 118–132, doi: 10.1080/02786820390229589, 2004.
- Zhou, L. M., Hopke, P. K., Stanier, C. O., Pandis, S. N., Ondov, J. M., and Pancras, J. P.: Investigation of the Relationship between Chemical Composition and Size Distribution of Airborne Particles by Partial Least Squares and Positive Matrix Factorization, *J. Geophys. Res. Atmos.*, 110, D07s18 doi: 10.1029/2004jd005050, 2005a.
- Zhou, L. M., Kim, E., Hopke, P. K., Stanier, C., and Pandis, S. N.: Mining Airborne Particulate Size Distribution Data by Positive Matrix Factorization, *J. Geophys. Res. Atmos.*, 110, D07s19 doi: 10.1029/2004jd004707, 2005b.