Seasonality and sources of organic particulate matter at selected IMPROVE sites in 2011 determined by infrared spectroscopy measurements

A. Kuzmiakova[a], A. M. Dillner[b], S. Takahama[a,*]

[a] *ENAC/IIE Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland*
[b] *University of California – Davis, Davis, CA, USA*

Abstract

National Ambient Air Quality Standards and the Regional Haze Rule mandate that states implement control strategies to reduce particulate matter (PM) emissions in order to maintain progress towards national air quality and visibility goals. The availability of long-term speciated aerosol datasets is very useful for investigating the aerosol composition over multiple seasons to provide guidance on how to effectively address air pollution at local levels. Unlike the inorganic fraction of PM, organic aerosol (OA) sources and their seasonality remain poorly characterized. This work approaches the problem by presenting a reference study from 6 Interagency Monitoring of PROtected Visual Environment (IMPROVE) sites containing 616 samples collected in 2011 and establishing a method for systematic interpretation of multi-site, multi-season source apportionment of organic matter (OM) from Fourier Transform Infrared (FT-IR) measurements. To confirm the validity of site aggregation, a cluster-based evaluation indicates that common factor profiles may be obtained at all 6 sites and seasons. Multi-site factor analysis resolves 4 factor profiles, Processed 1, Processed 2, Hydrocarbon, and Hydroxyl, which explain the major variations in OM across all sites and seasons, and were attributed to a common set of sources. Phoenix exhibited a strong seasonal cycle, with winter

*Corresponding author
Email address: `satoshi.takahama@epfl.ch` (S. Takahama)

OM maxima reaching 2.2 μg m^{-3}, generally dominated by emissions from local residential wood burning, traffic, and construction. OM in Trapper Creek (annual average of 0.22 μg m^{-3}) was dominated by sources not readily controllable by the local jurisdiction, including marine aerosol, volcanic activity, natural wildland fires, and international emissions from shipping lanes. The OM composition at Olympic (annual average of 0.43 μg m^{-3}) is affected by mobile sources, industrial point sources, and area sources at the Port of Seattle and its metropolitan area. Mobile sources, biomass burning, and vegetative emissions are important at Mesa Verde and Proctor Maple, while at St Marks emissions from prescribed fires and agricultural clearing are the most significant contributor to visibility impairment, reaching over 6 μg m^{-3} in summer. The multi-site analysis of 24-hour monitoring network measurements introduces several unique aspects. First, multiple factor-source associations can result if chemical similarity is shared among (anthropogenic and biogenic) sources, such as the backbone of alkane hydrocarbon precursor. Second, year-round and seasonal co-variation of Hydrocarbon and Processed factors suggest co-incident emission and atmospheric processing over the diurnal cycle at these locations. Third, varying contributions of two anthropogenic combustion factors to the local OM reveal differences in contribution of atmospheric processing that are specific to the given location.

Keywords: PM2.5, aerosols, organic matter, source apportionment

---

1. Introduction

The impact of aerosols on health, air quality, Earth's radiative budget, and compliance with National Ambient Air Quality Standards and the Regional Haze Rule is a growing concern. Sucessful air quality management strategies to reduce PM in the atmosphere require comprehensive undertanding of particulate matter sources, chemical processing, transport, and lifespan. For instance, the Regional Haze Rule mandates that states implement long-term enforceable plans for reducing visibility-impairing pollution in more than 100 Class I areas, such as National Parks and Wilderness Areas. These plans ar revised every 10 years (eg, in 2018, 2028) (EPA, 1999). One of the major difficulties in designing effective mitigation strategies is that there are a variety of sources responsible for producing primary and secondary aerosols. In this paper we focus on source apportionment of OA, which is a dominant contributor to air pollution in many regions of the US and can represent up to 90% of submicron ambient PM (Zhang et al., 2007). As opposed to the sources of inorganic fraction of PM, which are reasonably well identified (Kleindienst et al., 2007), the organic fraction remains challenging to characterize due to the vast number of directly emitted compounds (eg, > 1000; (Pio et al., 2001)) and products of atmospheric photooxidation. OA sources range from point sources, such as emissions from industrial combustion (Calvo et al., 2013), residential wood burning (Brown et al., 2007), or wildland fires (Corrigan et al., 2013), to mobile sources, such as emissions from windblown dust (Breider et al., 2014), motor vehicles, or on-road and non-road diesel engines (Kinney et al., 2000; Wu et al., 2007; Liu et al., 2012). Additionally, local air quality is often impacted by international emissions through a long-range transport. Transnational sources of aerosol, such as emissions from international shipping lanes (Brewer and Moore, 2009) or large-scale dust storms in East Asia (Takemura et al., 2002), are very complex to predict and constrain, because they are outside of state jurisdiction and thus remain ungovernable. Understanding the contribution of each source type to the local air pollution burden enables the development of policies to reach air quality and visibility goals.

Another difficulty in addressing air quality at local levels is that a substantial fraction of pollutants originates upwind and travels significant distances. For example, a work by Wagstrom and Pandis (2011) examined aerosol transport at 10 sites in the Eastern and Southeastern US and concluded that the average transport distance of primary aerosol species and elemental carbon is 175 km. Secondary OA and sulfate on average originated 350 km away from the receptor area, in some instances reaching up to 2000 km. Transport distance further depends on the location and altitude of the source, wind speed, and other meteorological variables (Rinaldi et al., 2015). As a result, background levels of pollution are especially challenging for state, local, and tribal planners to determine, highlighting the need for understanding the spatial and seasonal patterns in primary and secondary aerosols.

Long-term PM composition data from large-scale monitoring networks, such as IMPROVE or Chemical Speciation Network (CSN), are very useful for observing trends and quantifying sources to provide guidance to policy-makers in their efforts in reducing emissions. The IMPROVE and CSN networks have been collecting PM samples for speciated aerosol analysis for more than 30 and 15 years, respectively (Malm et al., 2002; Hand et al., 2013). The number of participating sites has exceeded 150 in 2016. Both networks have sampling, handling, analytical, and quality control protocols, which ensure the data is consistent and comparable across all sites within a each network.

The vast multi-site aerosol speciation data from the IMPROVE network has been used to investigate spatial and temporal trends in aerosol concentrations. For instance, the study by Hand et al. (2013) aggregated IMRPOVE and CSN data from 2007 to 2010 at over 300 sites to produce isopleths of seasonal mean OM and elemental carbon (EC) concentrations in winter, spring, summer, and fall. In spring and summer the highest OM concentrations were experienced in the Southeast (up to 6.6 $\mu g\,m^{-3}$) while in winter the highest OM was found along the West Coast (up to 6.6 $\mu g\,m^{-3}$). Similarly, several years before, Malm et al. (2004) examined nation-wide monthly concentrations of fine aerosol species (sulfates, nitrates, organics, light-absorbing carbon, dust, and coarse gravimetric mass) by aggreggating IMPROVE data from 2001. Additionally, Hand et al. (2014) consolidated OM and EC from IMPROVE and CSN data to infer urban influence on regional concentration during 2008-2011. Network data has also been used for source apportionment at individual sites, such as in Phoenix (Brown et al., 2007) or at southwestern Oregon (Hwang and Hopke, 2007). Although the IMPROVE and CSN datasets have been used to study trends in aerosol composition at many individual sites, the multi-site, multi-season nature of the datasets has not been utilized extensively for understanding OM sources.

OA source contributions are estimated from measurements using either a supervised or unsuperived method. In a "supervised" approach, source profiles are obtained from sampling various emissions (e.g., vehicles or biomass burning), and their contributions are obtained by regressing ambient sample spectra against these profiles. In an "unsupervised" approach, a matrix of ambient sample spectra is decomposed into a set of underlying spectra based on how variables co-vary over time or across samples (Paatero and Tapper, 1994; Henry, 2003); each spectrum is then associated with different sources (emissions or photochemical oxidation and partitioning) through interpretation. Less a priori information is assumed in the unspervised approach than in the supervised case.

Chemical Mass Balance (CMB) is a supervised approach, which uses ambient OC and speciated concentrations of tracer compounds from gas chromatography with mass spectrometry (GC-MS) measurements along with source profiles to estimate sources. Given the variations in chemical composition within source classes and range of subsequent chemical transformations that can occur, constructing an appropriate set of source profiles from emissions sampling and laboratory studies can pose significant challenges. Positive Matrix Factorization (PMF) (Paatero and Tapper, 1994) is a factor analytic technique which performs the desired form of decomposition by constraining component or "factor" profiles to be non-negative, and also permitting weighting of variables according to their uncertainty. PMF has enjoyed wide use in the mass spectrometry community (e.g., Ulbrich et al., 2009; Williams et al., 2010). Particularly for aerosol mass spectrometry (AMS), the large number of factor profiles collectively amassed by the community has permitted the application of a hybrid approach combining CMB and PMF (Lanz et al., 2007; Canonaco et al., 2013), where some factors are constrained by PMF components determined from archived field campaigns and the rest from new measurements.

Russell et al. (2009) introduced the application of PMF for FT-IR spectra of submicron PM samples, and since then it has been used in many studies. While AMS and other mass spectrometry online techniques take advantage of high time resolution (less than an hour) for resolving variations in contributions from sources, FT-IR (where samples are collected on the order of over several hours) produces feature-rich spectra that leads to complementary interpretations (Corrigan et al., 2013). PMF analyses conducted individually on field campaign FT-IR measurements sought solutions which had low inter-factor correlations such that each factor component was likely to be associated with different source classes, and factors were labeled with probable source classes by examining correlations of factor strengths with trace elements, geographic origin, and spectral features in factor profiles. A meta-analysis of 14 campaigns (Russell et al., 2011) identified factors as being from one of six categories: fossil fuel combustion (processed and less processed), marine biogenic (polluted and unpolluted), and terrestrial biogenic (burning and non-burning). These factors shared similar spectral features and source composition, which was identified by their relative functional group contribution. However, the applicability of supervised source apportionment approach for IMPROVE samples remains uncertain due to differences in particle size cut and sampling artifacts (Weakley et al., 2016).

Therefore, to address the question of seasonal OA source composition in the IMPROVE network, in this paper we utilize an unsupervised approach. We use six IMPROVE network sites (one urban and five rural) from 2011 previously characterized by Ruthenburg et al. (2014). Their study (which included an additional site that was operated until mid-2011) reported median contributions to OM over the entire year across sites to be 73% alkane CH, 15% alcohol OH, 4% carboxylic OH, and 7% carbonyl CO. Site-wide OM/OC ratios ranged between 1.5 and 2.0 (10th and 90th percentiles), with a median value of 1.7. These ratios varied by season but with different trends across sites, suggesting the differing role of local sources to OM. To determine the origins of this observed OM, it is possible to consider performing PMF analysis for each site or season, or pool all samples together in a single analysis, as depicted in Figure 1, which is the approach taken in this study. While PMF analysis on samples segregated by site and season may yield more precise factors for each locale, inclusion of a diverse set of related ambient aerosol mixtures increases the statistical likelihood that a few are dominated by a single source, which aids in separation of source contributions (Henry, 2003; Paatero et al., 2002). One implication of conducting a pooled analysis is that that sources impacting the sites are assumed to be approximately constant (or that factors profiles attributed to a source are averaged over their many variations). Furthermore, 24-hour time-integrated measurements of PM observed at each site is assumed to embody the combination of transported primary emissions and quasi-stationary atmospheric transformations (including secondary organic aerosol formation) that occur between origin and measurement site for the samples considered together (Zhou et al., 2005).

To evaluate the sensibility of aggregating available monitoring network samples within a single PMF analysis, initial cluster analysis of all 616 IMPROVE sample spectra was performed (Section 3.1). Results indicated that spectroscopic profiles are hardly unique to each site or season, supporting the notion that, to a first order, a common set of factors may be assumed to underlie these geographically and temporally diverse samples. Such an interpretation is also supported by prior conclusions of Russell et al. (2011), who reported similarities in FT-IR PMF factors obtained across a large number of geographically and temporally diverse short-term field campaigns. Therefore, pooling all 616 samples into one dataset, we apply factor analysis and present a method for exploring PMF solution space and facilitating solution selection in Section 3.3. Referencing specific mathematical and physical criteria: cluster analysis of PMF factors, explained variation, singular value decomposition, and physical basis of the solution, in Section 3.4 we select and present 4 chemical factors. We identify each factor (Processed 1, Processed 2, Hydrocarbon, and Hydroxyl) by referring to its similarity to profiles reported in earlier literature, the abundance of key functional groups, oxygenated content, and temporal profile (i.e., factor source strengths). By comparing the time series of factors with the time series of source markers to infer origins or atmospheric processes contributing to factor's emissions, in Section 3.5 we assign each factor with site-specific source labels and calculate their seasonal contributions to OM. We distinguish between natural and anthropogenic origins of OA sources. We close with a site-by-site summary of OA factors, sources, their relative contribution to the local OM, including examining OM uncertainty.

2. Methods

2.1. Spectral and functional group data

We use FT-IR spectra of 616 particulate matter ($\leq 2.5\mu$m in diameter, $PM_{2.5}$) samples collected between 1 January 2011 and 31 December 2011 at six IMPROVE network sites: Mesa Verde National Park (Colorado), Olympic National Park (Washington), Phoenix (Arizona), Proctor Maple Research Facility (Vermont), St Marks (Florida), and Trapper Creek (Alaska). All sampling sites but Phoenix are rural. Figure 2 shows the location of the sites and Table 1 details their geographical and meteorological characteristics, including elevation, annual temperature range, and precipitation available from The National Oceanic and Atmospheric Administration database (NOAA, 2011). The Phoenix site has co-located IMPROVE samplers but filters from only one sampler are used in this study. The original IMPROVE 2011 dataset includes 53 measurements from Sac and Fox, KS but the site was excluded from our analysis due to its discontinuation in summer 2011. In our evaluation we also exclude 36 samples (mostly from Proctor Maple and St Marks sites), which were identified as spectrally anomalous in (Ruthenburg et al., 2014). The Polytetrfluoroethylene (PTFE, Pall Corporation, 25 mm in diameter) filters used for FT-IR analysis were sampled every third day for 24 hours at a nominal flow rate of 22.8 $L\,min^{-1}$. Concentrations of elemental species are obtained via X-ray florescence. Particulate matter and aerosol composition data are available through a publicly hosted IMPROVE repository `http://views.cira.colostate.edu/fed/`.

Ruthenburg et al. (2014); Dillner and Takahama (2015b,a) detail the mechanics of spectra acquisition of PM constituents on PTFE filters by Fourier transform infrared (FT-IR) spectroscopy. Prior to factor analysis, all spectra were baseline corrected using the smoothing splines baseline correction algorithm formalized in Kuzmiakova et al. (2016) to minimize the PTFE interference. We use the mid-infrared wavenumber region between 4000 and 1500 cm^{-1}which contains quantifiable peaks of relevant functional groups. The carbon dioxide absorption band between 2500 and 2220 $cm^{-1}$is also removed using interpolation method described by Takahama et al. (2013b) to minimize the interference not associated with particulate matter composition. Finally, we exclude background regions (with nominally zero absorbance) between 4000 and 3710 $cm^{-1}$and between 2000 and 1820 cm^{-1}as they provide no useful information to the analysis. Ruthenburg et al. (2014) reported abundances of alkane CH, carbonyl CO, and carboxylic and alcohol hydroxyl OH groups. These groups typically represent the major fraction of organic aerosol content in ambient atmospheric samples (Russell et al., 2011). While remaining absorbing functional groups, such as alkenes, aromatics, or organonitrates, may account for a detectable OM contribution in some instances (e.g., Day et al., 2010), visual inspection of IMPROVE samples confirms their contribution may be below detection limit and therefore they are omitted from this study.

2.2. Cluster analysis

In this work, we used the hierarchical clustering algorithm of Ward (1963), which arranges data into a set of nested clusters organized as a tree and has previously been used to obtain meaningful cluster assignments in FT-IR spectra (Liu et al., 2009; Takahama et al., 2011; Corrigan et al., 2013). Cluster analysis reduces the dimensionality of the ambient FT-IR measurements and derive physically meaningful patterns for categorization and interpretation without any apriori knowledge. Clustering will categorize FT-IR samples into groups (known as clusters), each of which share distinct inter-cluster characteristics as a result of specific source composition, chemical properties, or extent of atmospheric processing. While factor analysis can resolve invidual sources that contribute to the ambient sample mixture, clustering may form sets of sample mixtures consisting of relatively similar proportion of factor components and their respective strengths. In the past, cluster categories were documented to provide complementary information for source apportionment results (Takahama et al., 2011; Corrigan et al., 2013). Some researchers (Takahama et al., 2011) suggest that the solution space of cluster analysis is somewhat better constrained than for PMF. Therefore, when applied to the entire dataset, clustering can be a useful starting point which helps decide whether PMF should be applied to all sites (or all seasons) or whether a specific site (or a season) should be examined separately. If samples from a specific site (or a season) are assigned to a single cluster category, this subspace most likely possesses distinct spectral features or history, including sources and extent of atmospheric processing, unlike the rest of the dataset. To identify the dominance of sources responsible for these unique patterns, it may be the best to analyze the single-site cluster in a separate PMF analysis. Otherwise, if aggregated with the rest of the samples, the distinct features may not be resolved completely. On the other hand, a uniform assignment of sites to multiple cluster categories implies that all sites contain sample mixtures with relatively similar proportion of contributing factors, as evidenced by intra-cluster similarity. The number of clusters is specified by the user. As a general rule, selecting a higher number of clusters may be more effective for discriminating against "atypical" spectral features as a result of "atypical" source composition while selecting a very low number of clusters may not permit sufficient distinction between individual sources of the existing aerosols.

2.3. Positive matrix factorization

Factor analysis was used to extract a set of common profiles that contribute in different proportions to the measured ambient PM FT-IR spectra. As a result, the factor analysis can linearly transform the measured dataset (a spectral matrix with rows representing time series of wavenumber variables) into several factor profiles while reducing the dimensionality of measurements and preserving most of the explained variance at the same time. Each extracted factor typically corresponds to molecular mixtures with specific functional group assignments and contains information about their sources, processing age, or chemical properties. Out of existing factor analysis techniques, in atmospheric sciences Positive Matrix Factorization (PMF) (Paatero and Tapper, 1994) has been widely adopted for source apportionment of atmospheric aerosol constituents. PMF has a long record in use for characterizing FT-IR spectra (Russell et al., 2010, 2011; Bahadur et al., 2010; Takahama et al., 2011; Takahama, 2015), X-ray absorption spectra (Liu et al., 2009; Takahama et al., 2010), aerosol mass fragment spectra (Zhang et al., 2011; Aiken et al., 2008; Canonaco et al., 2013), ambient particulate matter concentrations (Aguilera et al., 2015) and size distributions (Sowlat et al., 2016). While the detailed methods of PMF programs have been reported elsewhere (Paatero and Tapper, 1994; Paatero, 1997), in summary PMF generates factor solutions according to non-negativity constraints in factor profiles (chemical constituents) and their mass contributions, subject to weighting of sample and variable by uncertainties:

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij} \tag{1}$$

where x_{ij} refers to the spectra data matrix with i samples and j wavenumbers, f_{kj} is a representation of k^{th} factor profile at j^{th} wavenumber, and g_{ik} is a mass contribution of k^{th} factor towards i^{th} sample. e_{ij} refers to PMF residuals. g and f are found iteratively by minimizing a quantity Q defined as:

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

where s_{ij} represent the weights based on estimated measurement uncertainties specific to each sample and variable. Since the PMF method is a weighted least squares fit, the nature of the atmospheric aerosol data necessitates s_{ij} to be chosen judiciously to reflect the quality of spectral data and important physical implications from the FT-IR measurement process. Past studies (Takahama et al., 2011; Russell et al., 2011) worked with a simplistic representation of s_{ij} for FT-IR spectra, which consisted only of a term considering only wavenumber-dependent blank uncertainty. In actual FT-IR measurements, as with many instrumental signals, the analytical uncertainty increases with concentration of analyte. Thus, failing to account for this heteroscedastic behavior may place undue weight on the most prominent spectral features (e.g., those originating from functional groups with high absorption coefficients but not necessarily high abundance), and neglect more subtle features that can provide guidance for factor analytic decomposition. Furthermore, improved estimates of s_{ij} can better indicate the expected structure of residuals e_{ij} and enable alignment of Q values with the system degrees of freedom [$Q_{\text{exp}} = mn - p(m+n)$; Paatero et al., 2002] used as a reference for model evaluation.

2.3.1. FT-IR measurement uncertainty

For absorption-dominated interactions between sample and infrared radiation, the Beer-Lambert law describes the linear relationship between sample concentration c and observed absorbance x for wavenumber $\tilde{\nu}$ and substance r:

$$x(\tilde{\nu}) = \beta_r(\tilde{\nu}) c_r + \epsilon(\tilde{\nu}) \ . \tag{3}$$

β is the absorption coefficient, and the term ϵ represents the measurement error, often assumed to be normally distributed with mean of zero and standard deviation σ: $\epsilon \sim \mathcal{N}(0, \sigma)$ (Skoog et al., 2017). However, because measurement error increases with measured signal intensity (Sokal and Rohlf, 1981), we decompose the error term into a fixed term and concentration-dependent term, both of which are assumed to be normally distributed with a mean of zero (the notation for wavenumber dependence will henceforth be dropped to simplify the presentation):

$$\epsilon \sim \mathcal{N}(0, \sigma_0) + \mathcal{N}(0, \sigma(c)) \tag{4}$$

Assuming proportionality between the standard deviation and concentration in the second term through a wavenumber-dependent constant κ (e.g., Noblitt et al., 2016), the expected value and variance of the overall measurement error model can be described as follows:

$$\begin{aligned} \mathrm{E}[\epsilon|c] &= 0 \\ \mathrm{Var}[\epsilon|c] &= \sigma^2(c) = \sigma_0^2 + \kappa^2 c^2 \ . \end{aligned} \tag{5}$$

σ_0^2 is a fixed contribution from the variability of blank signal and $\kappa^2 c^2$ is a variable error contribution that grows with sample concentration. Therefore, Equation 5 conforms to a theoretical expectation of variance terms being additive. Rather than working with concentrations (which we do not know generally for ambient samples), we reformulate Equation 5 as a function of measured absorbance:

$$\tilde{\sigma}^2(x) = \sigma_0^2 + \tilde{\kappa}^2 x^2 \tag{6}$$

where $\tilde{\kappa} = \kappa/\beta$ (from equations 3 and 5).

Our approach is to estimate σ_0 and $\tilde{\kappa}$ directly from measurements, and obtain an expression for PMF uncertainty for use in equation 2:

$$s_{ij} = \sqrt{\sigma_{0,j}^2 + \tilde{\kappa}^2 x_{ij}^2} \ . \tag{7}$$

To obtain an estimate for $\sigma_{0,j}$, we apply the smoothing splines baseline correction algorithm of Kuzmiakova et al. (2016) (same as that applied to ambient sample spectra) on 54 blank PTFE sample spectra. While on average, the blank absorbances are zero at each wavenumber, the variability about the mean is used to determine the fixed variance contribution from instrumental signal, baseline correction, and blank signal ($\sigma_{0,j}^2$) (Russell et al., 2009).

To obtain an estimate for $\tilde{\kappa}$, we permit regression residuals from fitting equation 3 to measurements of reference standards to serve as surrogates for measurement errors, and develop a relationship between x^2 and $\tilde{\sigma}^2$ as described in equation 6. We use 238 laboratory standards prepared by Ruthenburg et al. (2014) from aqueous or ethanolic solutions of pure, atmospherically-relevant compounds, such as alcohol, sugars, and dicarboxylic acids. Each compound contains only a few relevant functional groups making it suitable to identify non-interfering, absorbing species necessary for developing analytical uncertainty models. All standard spectra are baseline corrected using the same smoothing splines algorithm as for ambient and blank sample spectra to isolate the absorption contributions for Equation 3 to be valid. We group standards by their compound type, and within each compound type we identify all functional group bands which do not overlap with peaks from other functional groups. These isolated bands were selected for the quantitative analysis and are summarized in Table 2. To minimize variability across different samples, we measure absorbance at specific wavenumbers (ν_1 and ν_2 in Table 2), which correspond to the centers of peaks where maximum absorbance intensities are expected. When two spectral bands from the same functional group absorb with no successive overlap, for example two isolated peaks in ammonium sulfate), we measure absorbance values from both peaks separately at their respective frequencies ν_1 and ν_2 to increase robustness of our models. Figure 3 summarizes results of fitting equation 3 to the dataset just described. An adaptive moving window containing seven successive points of ϵ, ordered according to magnitude of x, are used to pair the mean value of x and associated variance $\tilde{\sigma}^2$ within each window. As $\tilde{\sigma}^2$ contains the fixed uncertainty, the latter is subtracted prior to estimating $\tilde{\kappa}$ using equation 6. In this way, $\tilde{\kappa}$ are obtained for a few functional groups over several wavenumbers. Given that a precise value is not available over all wavenumbers, we calculate the global mean value from all estimated values of $\tilde{\kappa}$ for use in equation 7.

2.3.2. PMF solution space and source assignment

While the main objective of the PMF analysis is to explore underlying covariation of variables from FT-IR measurements to extract physically interpretable factors, which provide accurate information about OA sources, atmospheric processes, and chemical properties, PMF algorithm provides only mathematical solutions which necessitate careful selection, evaluation, and interpretation. PMF solution may vary depending on user's selection of free parameters. In this application, there are 3 degrees of freedom, which scope the PMF solution set: the number of factors, the rotational parameter (FPEAK), and seed value. FPEAK defines the linear combinations of factor profile and strength matrices which are constructed to characterize the possible solutions and therefore can indicate if there is rotational freedom in the solution. Seed values influence the likelihood that the solution will correspond to a global minimum of Q (Paatero et al., 2002; Brown et al., 2012). Thus to explore the factor solution space and facilitate solution selection, we use different mathematical and physical criteria, such as Q/Q_{exp} (defined in Section 2.3), factor cluster groupings, EV, and physical basis of the solutions.

After selecting the factor solution, we use functional group abundances for IMPROVE ambient samples estimated by Ruthenburg et al. (2014) (with minor revisions introduced by Takahama and Dillner (2015)) using linear model for calibration. In the past, the non-linear peak-fitting method of Takahama et al. (2013b) was applied to obtain such estimates (e.g., Russell et al., 2009). However, differences in abundances estimated by Takahama et al. (2013b) and Ruthenburg et al. (2014) can be expected on account of different reference standards and algorithms used for calibration. In this work, we formulate our OM and functional group composition estimation in PMF factors to explain those reported by Ruthenburg et al. (2014) (Section 2.1). Nominally, the relationship between functional group abundances in individual samples (y_{iz}) and PMF factors (y_{kz}) in such cases may be expressed through the following relationship:

$$y_{iz} = \tilde{x}_{ij}\tilde{b}_{jz} = g_{ik}\tilde{f}_{kj}\tilde{b}_{jz} = g_{ik}y_{kz} \ , \tag{8}$$

where b_{jz} are the coefficients obtained from partial least squares regression. However, the calibration model (embodied by regression coefficients) of Ruthenburg et al. (2014) was developed for raw spectra without baseline correction; tildes above symbols in equation 8 identify quantities associated with them (including $\tilde{f}$, which represents the spectral profiles resulting from a hypothetical bilinear decomposition of the raw spectra). As our PMF analysis is applied to baseline corrected spectra, the regression coefficients are not directly applicable to our factor profiles. Therefore, we find functional group abundances for each factor from those estimated for each ambient sample using a least squares approach to satisfy the relationship:

$$y_{iz} = g_{ik}\, y_{kz} + \epsilon_{iz}\ . \tag{9}$$

g_{ik} is the source contribution of k^{th} factor towards i^{th} sample obtained by PMF on baseline corrected spectra obtained from equation 1.

Two extensions to the estimates of Ruthenburg et al. (2014) are provided for y_{iz} used in this work: 1) aggregation of carboxylic COH and carbonyl CO to carboxyl functional groups, and 2) estimation of ammonium NH. Carboxyl groups (COOH) consist of carbon bonded to -OH and =O in the same functional group. While the abundances of the two vibrational modes (O-H stretching and double-bonded O stretching) are often quantified separately, it is sensible to determine how much of the carbonyl is associated with carboxyl groups how many are associated with others (e.g., ketone, aldehyde, ester). This apportionment can be achieved by comparing the molar abundance of carbonyl CO in excess of carboxylic COH (Takahama et al., 2013b), and assigning this to the non-carboxylic carbonyl. Our analysis indicates that the estimated carbonyl CO and carboxylic COH have nearly a 1:1 correspondence, suggesting that the carbonyl CO quantified in these samples belong to carboxylic COOH. The lack of additional non-carboxylic CO is surprising given their abundance in biogenic and biomass burning samples reported previously Schwartz et al. (2010); Takahama et al. (2011); Corrigan et al. (2013). However, analysis of carbonyl absorption bands in spectra (near 1720 cm^{-1}) indicates that relative peak heights in IMPROVE network samples are generally less than those found in submicron aerosols sampled over shorter intervals (Russell et al., 2011), supporting this interpretation. Ammonium NH (largely associated with inorganic salts) is considered an interferant for OM analysis as it can co-absorb over the same wavenumber range as many organic functional groups (Maria et al., 2003). This group is accounted for but not quantified in spectral analyses for quantification of OM (Takahama et al., 2013b; Ruthenburg et al., 2014). However, as its presence can be expected in the ambient FT-IR spectra and resulting PMF factor profiles, we estimate their abundances by assuming full neutralization of sulfate and nitrate anions measured by ion chromatography (Dillner and Takahama, 2015b).

The solutions to equation 9 are additionally examined for consistency between estimated chemical composition and spectroscopic profile — for instance, factors with significant abundances of carboxylic groups should are expected to exhibit absorbances (f_{kj}) in the carboxylic COH and carbonyl C=O absorption bands. This additional criterion (referred to as "chemical consistency" in this work) further provided guidance on the selection of the solution profiles.

In Section 3.4 we identify each factor profile based on its similarity to profiles reported in literature (Russell et al., 2011; Corrigan et al., 2013; Frossard et al., 2014), the abundance of key functional groups in each profile, oxygenated content, and factor's temporal profile. Finally, site-specific source labels were assigned by comparing the time series of factors with the time series of source markers to infer origins or atmospheric processes which gave rise to emissions contributing to chosen PMF factors in Section 3.5.

3. Results

3.1. Cluster analysis

Figure 5 presents cluster memberships for all 616 samples differentiated by site (horizontal panel) and season (left vertical panel). In this application, we considered selecting between 3 and 8 clusters to maintain consistency with Section 3.3 where we vary the number of factors from 3 to 8. We selected 8 clusters to gain advantage of more distinct clusters due to greater homogeneity within a group and greater difference between groups as mentioned in Section 2.2. In Figure 5 we notice that no site or season is singled out in a separate cluster, which would imply spectroscopic signature consistently distinct from the remaining dataset. All sites have been assigned to at least 5 different clusters, which contain members from all 6 sites. This confirms that intra-cluster spectroscopic features associated with similar contribution of sources contained in aerosol mixtures are present uniformly across all 6 sites. Therefore, uniform cluster assignment supports the multi-site application. Also, we notice no significant difference between urban samples (Phoenix) and rural samples (remaining sites) with the exception of clusters 7 and 8, which detected several unusual samples collected in fall and winter at Phoenix and St Marks sites. Figure S1 reveals that sources of these aerosols are dominated by biomass burning emissions and indicates similarities in atmospheric processing that may have occured during transport from their original locations in Arizona and Florida to their respective measurement sites, Phoenix and St Marks. Because these are only 10 samples ($< 2\%$ of total measurements) their spectral features may not be well represented by the PMF. More detailed analysis on IMPROVE clusters is outside the scope of this study and can be found in Ruthenburg et al. (2014). In our context, cluster-based evaluation is a first step towards data summarization and determining whether a multi-site or single-site source apportionment should be performed.

3.2. FT-IR measurement error model

This section presents results for the heteroscedastic component of FT-IR error, which is necessary to obtain a PMF uncertainty matrix, $s_{i,j}$ representative of FT-IR measurements. Figure 3 shows fitted linear calibration models to represent a relationship between reference concentration and an FT-IR instrument response (absorbance) for each functional group and compound type. To obtain calibration curves representative of ambient PM samples, we only work with standards containing absorbance values < 0.5, which corresponds to roughly twice the maximum absorbance found in our IMPROVE samples. Excellent agreements were obtained as coefficients of determination (R^2) in all models from Figure 3 were > 0.95. Regression residuals, ϵ (Equation 3), were used to determine variance, $\tilde{\sigma}^2$, using the moving average described in Section 2.3.1. Figure 4 shows fitted linear regression lines to relate x^2 (squared absorbance) and

$\tilde{\sigma}^2 - \sigma_0^2$ from Equation 6 for each functional group and relevant compound type. Since Equation 6 does not contain any additional physical terms, we performed regression through the origin (i.e. the fitted lines pass through (0,0)). Additionally, we exclude malonic acid (functional group: carbonyl), levoglucosane (alcohol), d-glucose (alcohol), arachidyl dodecanoate (alcohol), and 1-docosanol (alkane and alcohol) due to limited sample size or negligible variance values (i.e. when $\tilde{\sigma}^2$ is on the order of σ_0^2). The final scaling coefficient, $\tilde{\kappa}$, is determined as the square root of the mean of the 9 slopes of the regressed lines in Figure 4 and was found to be 0.054. The heteroscedastic component of error introduced in this work is found to be orders of magnitude larger than the fixed error term used previously (Russell et al., 2009).

3.3. PMF solution space

In this section, we systematically explore the PMF solution space of the three parameters: number of factors, FPEAK, and seed parameters. The following values were used: seed values = {1, 10, 100}; FPEAK = {-1.6, -1.2, -0.8, -0.4, 0, 0.4, 0.8, 1.2, 1.6}; and number of factors = {3, 4, 5, 6, 7, 8}. Therefore, the total number of PMF simulations was 162 ($3 \times 9 \times 6$). Figure S2 shows Q/Q_{exp} decreases from 3.5 in 3-factor solution to 0.8 in 8-factor solution. The overall range of Q/Q_{exp} is comparable to those from past studies (Takahama et al., 2013a) and is reflective of our FT-IR measurement uncertainty matrix, $s_{i,j}$. While a systematic comparison between the old and new methods for estimation of $s_{i,j}$ have been conducted in this study, the difference on Q/Q_{exp} is not immediately apparent. Because Q/Q_{exp} does not display a clear minimum and universally decreases with increase in the number of factors, this metric does not offer a method for selecting the correct number of factors. However, a large decrease in Q/Q_{exp} with the addition of the fourth factor (from 3.5 to 2.3) implies that the additional factor can explain a significant fraction of the OM variation that was unaccounted for by the previous three factors. In past studies (Paatero and Tapper, 1994) a large decrease in Q/Q_{exp} caused by an additional factor had been used as a metric for choosing a solution. This trend is consistent with an increase in explained variation in OM when we add a fourth factor to increase explained variation from 90 to 95% (Figure S3). Generally, in solutions with more than 4 factors the total OM was well apportioned and explained variation was > 95%. However, adding a fifth or sixth factor does not appear to change explained variation in the spectra and selecting a fewer number of factors leads to a parsimonious model that is less likely to be overfitted. Finally, using different seed and FPEAK values did not appear to yield additional variations in Q/Q_{exp} for a given number of factors, indicating robustness of solutions.

Additional method for examining the variance of the original aerosol sample matrix, x_{ij}, includes evolving factor analysis (Keller and Massart, 1992). We applied singular value decomposition (SVD) to the sample matrix (which does not account for the measurement uncertainty matrix s_{ij}) using a fixed-size moving window. Percentage of data recovery at each wavenumber using different numbers of components was estimated by normalizing the cumulative contribution of their singular values by the trace of the covariance matrix. Figure S4 shows that three components explain approximately 90% of the variation in the FT-IR measurements across most wavenumbers, consistent an explained variation of 90% from 3-factor solution in the PMF analysis. The percentage recovery signal is consistent with mean PMF residual structure, ϵ_j, in Figure S5.

Given the large number of solutions that are generated for the range of seed values and rotational parameters specified, we consider the possibility that these solutions may be reoccurring realizations of a few solution sets. Previous studies have shown consistent reports of spectral features associated with particular source classes (Russell et al., 2011). In Figure 6 we apply hierarchical clustering to assign factor profiles from all 162 solutions into groups based on their spectral similarity. Note that the number of factor groups and the number of factors in each solution are two parameters controlled independently by the user. The number of cluster groups can be smaller than the number of factors specified if resulting spectra appear very similar to each other (eg, when we use 6 factors; fourth row). Conversely, large spectral differences in factors across simulations warrant additional cluster groups (eg, when we use 4 factors; second row). The plot shows the PMF solution set follows reocurring solutions across a different number of factors. Observing trends across columns simulations with different seed and FPEAK parameters will lead to similar realizations of the same spectral profiles regardless of the number of factors used, thereby confirming the rotational stability of solutions (Paatero et al., 2002). For instance, cluster 2 in column 2 contains a hydrocarbon-like profile with visible methylene peaks in all simulations. However, solutions with > 5 factors begin to generate physically improbable, degenerate profiles, which either contained only a single organic functional group (eg, hydroxyl group in factor-cluster 6) or exhibited artificially jagged spectral features, which presented an unrealistic departure from smooth Gaussian peaks (Takahama et al., 2013b) (eg, factor-clusters 7 and 8). Such cases could be formally classified via a roughness metric but the implementation is outside the scope of our study. Additionally, in Figure S6, which compares pair-wise g-score correlations in each simulation, we see that increasing the number of factors beyond 4 leads to profiles which strongly correlate ($r > 0.65$) with other profiles in the given solution set. Strong correlation between two factors in time (eg, in our case factors from clusters 1 and 6) suggests they likely originated from the parent factor. This is consistent with "factor splitting" discussed in Ulbrich et al. (2009), suggesting that emissions from a single source are prescribed to two or more PMF factors.

According to our factor profile clustering scheme, the 162 solutions can be grouped into 22 categories (Figure S7), which provides a simplification in the PMF analysis. We note that as many samples from multiple periods and sites are grouped together in the analysis, the PMF factor profiles, while spectroscopically (and presumably chemically) similar, may be associated with different source classes. As shown in Figure S6, there exist factor pairs from a single solution which are grouped into the same factor-cluster (factor-clusters 1, 3, and 5), while maintaining factor strengths (g-scores) that are nearly orthogonal to one another (Figure S7). The interpretation of factor components which are spectroscopically similar but due to different sources is a topic of this work

(Figure 1), but additional approaches may be investigated in future works. For example, classifying solutions based not only on factor profiles (f-values) but their strengths (g-scores) can potentially be fruitful in differentiating patterns in reoccuring solutions.

As a result of the above evaluations, we constrain the number of factors to 4 and generate another set of PMF solutions over a wider seed range = {1, 3, 5, 10, 15, 20, 30, 50, 75, 100} (FPEAK range remains the same as above) to determine the set of profiles and their frequency. In total, these combinations lead to 90 PMF solutions divided into 3 solution classes, as previously described in Figured S7. The most frequently occurring solution class is plotted in Figure S8. Its factor profiles belong to factor-clusters 1, 2, 3, and 4, and occured in 81% of cases (73 out of 90). The remaining two solution classes occured in 12% (11 out of 90) and 7% of cases (6 out of 90), respectively, and are plotted in Figures S9a and S9b. These solution sets contain profiles from factor-clusters i) 1, 3, 4, 5, and ii) 1, 3, 5, 6. We reject these two solution sets for two reasons: i) using a frequentist approach (i.e., a consensus selection argument; Héberger, 2010) over a range of plausible seeds and rotation parameters, the solution appears in at most 12% cases, ii) the inverse functional group estimation for the fourth profile (with prominent carbonyl and methylyne peaks) would be inconsistent with its spectral profile. Therefore, we select the 81% solution in Figure S8 for our work with the expectation that each profile represents a chemically feasible factor with specific spectroscopic signature.

3.4. Spectral profiles

Four distinct spectral profiles were identified using PMF. Due to the apparent similarity in chemical composition in aerosol mixtures originating from different sources at several sites, we find that multiple source labels could plausibly be assigned to each spectroscopic (factor) profile (Figure 1). This marks a departure from previous studies where each PMF factor profile was attributed to a specific source. It is possible that this multi-site approach to PMF lumps distinct but similar chemical profiles into a single factor on account of resolvability.

PMF localized to specific sites may better able to determine more precise profiles. However, preliminary analysis of a single-site PMF case for Olympic, WA, suggests that profiles obtained for site-specific PMF may yield similar results to the multi-site PMF presented in the body of this work (Figure S10). Further comparisons of multi-site and site-specific PMF, alongside other factor analysis methods which target features that discriminate among the most source-relevant variations (rather than overall variation), are topics that can be investigated in future studies. For the remainder of this work, we describe our factor interpretations for the multi-site PMF. Table 3 summarizes their key characteristics and Figure 7 profiles the factors with their functional group composition.

Processed 1 and Processed 2 are two distinct anthropogenic fossil fuel combustion factors resulting from different degree of photochemical processing. One of the evident features in both combustion factors is substantial ammonium absorbance (2850 – 3300 cm^{-1}), similar to anthropogenic combustion factors reported in previous campaigns (Corrigan et al., 2013; Liu et al., 2012; Hawkins et al., 2010), which suggests these aerosols are secondary. Processed 1 contains roughly equal mass fractions of alkane (35%), alcohol (29%), and carboxylic acid (36%). It has the highest OM:OC and O:C ratios (2.5 and 1.0, respectively) amongst all factors, indicating it is heavily oxidized likely due to its formation in later generations (Jimenez et al., 2009; Aiken et al., 2008). Processed 2 contains a relatively large mass fraction of alkane (57%) with the remaining 43% of organic mass taken up by carboxylic acid. The large alkane mass fraction indicates that fossil fuel emissions captured in Processed 2 factor likely underwent less atmospheric processing than those found in Processed 1 (Frossard et al., 2014). Further, OM:OC and O:C ratios are lower for Processed 2 (1.7 and 0.4, respectively), suggesting lower oxygenated content in the less aged air masses. The oxidation state and aging of two secondary aerosol factors are consistent with previous studies on elemental ratios (Canagaratna et al., 2015; Aiken et al., 2008) where two secondary organic aerosol (SOA) components were reported: more oxidized SOA had OM:OC between 2.3 and 2.5 and less oxidized SOA had OM:OC between 1.8 and 2.0. Also interesting is a relatively substantial carboxylic acid contribution to OM (> 35% in both factors), which had been attributed to urban combustion sources (Russell et al., 2011). Hydrocarbon factor is dominated by alkane (81% of organic mass) with minor fractions of carboxylic acid (11%) and alcohol functional groups (8%). The large fraction of alkane and small oxygen content (OM:OC of 1.4) suggest some or most of the emissions originate from primary aerosol sources (Aiken et al., 2008). The prominent feature of this factor is a pair of alkane peaks (at around 2900 and 2850 cm^{-1}) associated with repeated methylene groups in long-chain hydrocarbons (Coates, 2000; Pavia et al., 2008). Repeating methylene units are derived from burning of vegetative detritus during forest fires (Hawkins and Russell, 2010), residential wood burning (Russell et al., 2011), and primary anthropogenic combustion (Liu et al., 2012).

Hydroxyl factor features broad organic hydroxyl absorption in the range between 3700 and 3300 cm^{-1}. The alcohol makes up the majority of organic mass (53%) with the rest being taken up by alkane (43%). Initial hypotheses regarding the origin of sources contributing to Hydroxyl factor can be inferred on the basis of the type of compounds where hydroxyl functional groups can be frequently found. First, hydroxyl groups may have originated from saccharides emitted from bubble bursting in surface seawater (Russell et al., 2011). The relatively high hydroxyl fraction in the PMF factor is consistent with the 80% carbohydrate fraction of total dissolved organic carbon at the ocean surface (Aluwihare et al., 1997). Additionally, the profile, functional group composition, and O:C ratio of 0.9 are very similar to those in marine PMF factors reported from previous shipboard and ground-based campaigns in coastal locations (Frossard et al., 2014; Bahadur et al., 2010; Russell et al., 2011). Yet, since primary marine biogenic sources are typically confined to coastal and marine regions, hydroxyl groups at continental sites may have been derived from alternative sources. Thus, the second likely origin of OH groups are mineral dust particles which had been found to be coated with organic OH (Takahama et al., 2013a; Hawkins et al., 2010). The substantial alcohol mass fraction is consistent with reported OM containing dust particles resuspended by vegetative detritus

(Ahlm et al., 2013) and with the composition of lignin and other carbohydrates in vegetative material (Bianchi et al., 1993). The shape and functional group composition of Hydroxyl factor in Figure 7 D is also consistent with Vegetative Detritus factor identified in organic aerosol source apportionment study at Bakersfield in 2010 (Liu et al., 2012). Finally, Hydroxyl profile shows signatures of methylene peaks attributable to biomass burning or potentially anthropogenic influences.

3.5. Site-specific sources

We focus on Phoenix, Trapper Creek, and Olympic and discuss remaining sites (Mesa Verde, Proctor Maple, and St Marks) only cursorily due to a lack of available literature on year-long aerosol characterization in these regions. Figure 8 presents the time series of PMF factors contributing to OM during 2011, Table 4 summarizes attributed site-specific sources together with their seasonal dominance, and Figure 9 presents the seasonal averages of those sources.

3.5.1. Phoenix

As the only urban site in our dataset, Phoenix shows the highest organic aerosol concentrations, with yearly average of 1.69 $\mu g\,m^{-3}$ in 2011. The site exhibits a distinct organic carbon seasonal cycle, which peaks in winter (2.2 $\mu g\,m^{-3}$) and shows its minimum in summer (1.3 $\mu g\,m^{-3}$), driven by seasonal meteorological and urban emissions variations. Phoenix is a city of 1.5 million people located in a larger metropolitan area with a total population of over 4.5 million. It is located in the central Arizona desert, a subtropical desert biome with extremely low annual precipitation (Table 1), high levels of solar radiation, and large differences between the annual lowest and highest temperatures. Airflows in the Phoenix metropolitan area are affected by local topography. The site is located in a broad valley at an altitude of 348 meters and surrounded by mountain ranges from north, east, and south. The mountains adjacent to the urban area rise to 900 meters above the valley leading to winter inversion layers, which trap locally-produced organics. Low inversion layers in winter and minimal atmospheric transport has also been identified to be responsible for unusually high $PM_{2.5}$ events in winter months (Brown et al., 2007).

In Figure 8 we see throughout the year OM composition in Phoenix is dominated by Hydrocarbon factor, which accounts for up to 90% of OM in winter. Figure 11 summarizes the magnitudes of seasonal correlations between Hydrocarbon factor and relevant tracer concentrations (EC, Br, Zn, Cu, Fe, Mn, Cr, K, and Cl), which point to evidence of a mixture of natural and anthropogenic urban sources. First, in winter the factor is highly correlated with EC (r=0.96), K (r=0.92), Br (r=0.75), and Cl (r=0.85) suggesting a strong influence of residential wood burning emissions originating from a a variety of biomass combustion appliances, such as open fireplaces or wood and pellet stoves. A previous work by Rau (1989) examining the composition of residential wood burning emissions reported that wood smoke particles had 20% to 60% carbon content (primarily elemental carbon) and high levels of K (11%) and Cl (3%). Residential wood combustion is evident only in winter as Cl, often used as an indicator for wood smoke particles (Khalil and Rasmussen, 2003), shows no correlations in remaining seasons. In 2011 there were 21 days (7 in January, 6 in February, and 8 in December) when minimum temperature in Phoenix reached below 0 degrees C, which suggests wood burning during the nighttime or even daytime, particularly during the holiday season between Christmas and New Year's Day. This is consistent with previous source apportionment studies, which identified residential wood burning as a major contributor to winter particulate matter in Phoenix area (Brown et al., 2007; Ramadan et al., 2000; Zielinska et al., 1998). Second, the factor is correlated with biomass burning tracers K ($0.41 < r < 0.92$) and Br ($0.48 < r < 0.75$) throughout the year, indicating the influence of forest fires and agricultural burning. In a previous Phoenix air quality (Ramadan et al., 2000) spanning two years (1996-1998) temporal profiles of biomass burning activities showed presence in all seasons but minor peaks in months of January and July. Third, year-round correlations with EC ($0.52 < r < 0.96$), Zn ($0.43 < r < 0.73$), and Cu ($0.50 < r < 0.90$) are associated with traffic emissions which include emissions from both motor vehicles and heavy-duty diesel trucks.

Specifically, Zn, Cu, and Cr are tracers associated with vehicle exhausts, tire and brake abrasion, oil combustion (Viana et al., 2008). Transportation-related emissions are important contributors to urban OC in all seasons because the Phoenix sampling site (Supersite) is located in a densely-populated area within 2 miles of a major freeway. In Figure S11 we look at daily measured concentrations of Zn and Cu, two main markers of vehicle exhaust emissions. Zn and Cu show maximum from October to February and minimum during summer. The winter peaks are consistent with a reported influx of visitors in fall and winter seasons (Brown et al., 2007). Higher rates of visitation are associated with higher rates of anthropogenic activities, such as driving and residence heating, and therefore a rise in locally-generated fossil fuel emissions is conceivable.

Hydroxyl factor is the second largest contributor of organic aerosol in Phoenix, accounting for 28% of OM in summer. Its correlations with major dust tracers, such as Si ($0.8 < r < 0.94$), Al ($0.90 < r < 0.94$), Mg ($0.78 < r < 0.90$), Ca ($0.85 < r < 0.90$), and Ti ($0.79 < r < 0.91$), confirm the presence of mineral dust, which is expected given the arid desert climate. In Figure 10 we notice concentrations of mineral dust elements follow a temperature trend: they peak in summer season and gradually decrease until winter. Therefore, likely sources of summertime dust in Phoenix area include resuspended dust from roads, construction sites, and other unpaved areas (Ramadan et al., 2000).

Finally, the remaining 10% of OM is attributed to Processed 1 and Processed 2 factors. Both processed factors are correlated with S ($0.78 < r < 0.80$ in Processed 1 and $0.60 < r < 0.65$ in Processed 2) and thus are associated with sulfur dioxide emissions from coal-fired power plants located southwest of Phoenix in Arizona, New Mexico, and Mexico (Brown et al., 2007; Ramadan et al., 2000). Organic contribution from Processed 1 factor appears relatively stable throughout the year (around 0.13 $\mu g\,m^{-3}$) suggesting that more processed or transported fossil fuel emissions are independent of photochemical activity. Organic contribution from Processed 2 factor shows minor peaks in winter providing some evidence of fresher, locally-produced fossil fuel emissions trapped in the inversion.

3.5.2. Trapper Creek

Trapper Creek has the lowest OM concentration from all 6 sites in our dataset; around 10 times less than Phoenix. Trapper Creek is also the only site in our study located in polar latitudes (north of 60 degrees) with distinct meteorological features, which include low levels of solar radiation, below- or near-zero mean temperatures throughout all seasons but summer, and low levels of precipitation (Table 1). The Arctic meteorology can influence seasonal sources, transport, and photochemical processing of OM, which ranges from 0.18 $\mu\mathrm{g\,m^{-3}}$ in fall and winter to 0.25 $\mu\mathrm{g\,m^{-3}}$ in summer (Figure 9). Observed OM concentrations at Trapper Creek peak in spring and early summer, contributing to a phenomenon commonly termed "Arctic haze" (Quinn et al., 2007). From November through April OM composition is dominated by Hydroxyl factor (Figure 8). Since aerosol concentrations are relatively very low and correlations with marine and mineral dust tracers are not conclusive, we examine seasonal tracer concentrations in Figure 10 to infer plausible Hydroxyl factor sources. In winter the factor represents marine aerosol source due to elevated Na concentrations. The ratio of Na and Si (the main mineral dust tracer) in winter (8:1) is similar to that in Olympic where Hydroxyl factor was identified as marine aerosol. In remaining seasons observed Na concentrations are a factor of 2-5 lower and proportionate to Si concentrations. Therefore, Hydroxyl factor most likely represents a mixture of oceanic and mineral dust sources. Si concentrations show a sharp maximum in spring (1.4 $\mu\mathrm{g\,m^{-3}}$). This is consistent with a previous long-term seasonal aerosol distribution study (Breider et al., 2014), which determined dust aerosol at Trapper Creek peaks in spring with major dust sources being the Sahara and the Taklaman and Gobi deserts. Additionally, in winter and spring we find the Arctic OM in Hydroxyl factor is mildly correlated with Fe, Mn, and Zn ($0.47 < r < 0.78$), indicating the presence of emissions from iron and steel industries and oil burning. This finding agrees with previous works (Frossard et al., 2014; Shaw et al., 2010), which report that international emissions from shipping lanes through the Bering Strait and oil industry contribute to the Arctic haze during springtime. From May to October OM composition in Trapper Creek is dominated by Hydrocarbon factor predominantly via episodic incidence, e.g. during events on May 30, July 17, or July 23, 2011 (Figure 8). The factor correlation with EC, K, and Br ($0.4 < r < 0.7$) indicates that emissions from biomass burning events account for the mass in this factor. A previous study by Shaw et al. (2010) characterizing a year-long aerosol composition in Barrow in northern Alaska reported that boreal forest fires in continental Alaska and central Siberia (west of Anadyr) were important sources of haze. In 2008 wildfire emissions from as far as Kazakhstan were known to affect Alaska air quality (Warneke et al., 2009). However, this long-range transport of biomass burning emissions in summer and early fall is the only source strongly affecting what is otherwise classified as pristine air masses (Hamilton et al., 2014). In winter Hydrocarbon factor is absent due to extensive snow and ice coverage, low solar radiation, and minimal biogenic activity in polar biomes. The remaining major fraction of Arctic OM throughout the year is accounted for by Processed 1 factor, which is highly correlated with S and sulfate ($0.8 < r < 1.0$). The OM contribution from Processed 1 factor shows a spring maximum (0.11 $\mu\mathrm{g\,m^{-3}}$), when it accounts for 42% of springtime OM, and gradually decreases throughout summer, fall, and winter. The factor seasonality is consistent with Trapper Creek sulfate aerosol concentrations (Breider et al., 2014), which were also the highest in spring months in 2008. In Alaska sources contributing to the processed factor are likely to originate from two source classes: i) anthropogenic and ii) natural (Breider et al., 2014). Anthropogenic sources include fossil fuel burning and smelting of sulphide ores in power plants in northeast Asia (Barrie, 1986; Polissar et al., 2001a). Natural sources include volcanic activities in Alaska Peninsula and Aleutian Islands. Specifically, volcanic emissions from Aniakchak, Okmok, and Cleveland volcanos, all of which were active in 2011 (McGimsey et al., 2014), may have contributed to elevated spring and summer masses in Processed 1 factor. Finally, Processed 2 factor accounts for around 12% of OM in spring. Mild correlations with Mn ($r = 0.51$), Fe ($r = 0.46$), Zn ($r = 0.54$), S ($r = 0.52$), and sulfate ($r = 0.56$) suggest regional diesel combustion emissions from power generators, trucks, cruise ships, and fishing boats.

3.5.3. Olympic

Referring to Figure 8, OM at Olympic site shows little seasonal cycle with organic concentrations usually less than 1.0 $\mu g\,m^{-3}$. In September and October OM was dominated by several high-pollution days causing concentrations to reach up to 4.0 $\mu g\,m^{-3}$ at the northwestern rural site. However, overall OM levels remained even throughout the whole year, suggesting influx of very stable organic sources which composition is independent of photochemical activity and precipitation. Although the Olympic site is in a national park, its air quality is affected by emissions from industrial regions along the Seattle metropolitan area and marine vessel traffic in the Strait of Juan De Fuca, both of which are less than 80 kilometers away from the park.

Unlike the rest of the sites, here Processed 1 and Hydrocarbon factors were highly correlated in time (r = 0.76, 0.59, 0.69, and 0.77 in winter, spring, summer, and fall, respectively), indicating they are associated with the same source but vary in their respective composition. We combined the 2 correlated factors into 1 factor called Processed Hydrocarbon (mass of the combined factor equals the sum of factor masses used in combination; Figure S12), leaving us with 3 linearly-independent factors (Hydroxyl, Processed Hydrocarbon, and Processed 2). The 3 factors explain the same fraction of OM variance as the 4 original factors in Figure 7 prior to their factor recombination. The combined g-score from Processed Hydrocarbon factor was used to infer correlations with source markers. The concept of factor combination was also used in previous organic aerosol characterization studies. For instance, Schwartz et al. (2010) reported a "summed biogenic factor" of two factors that represent different types of biogenic volatile organic compounds and processing, while Hawkins and Russell (2010) combine two minor factors that explain a small portion of the mass to a single one.

Figure 8 shows that Processed Hydrocarbon factor accounts for the majority of local OM, ranging from 69% in winter to 89% in summer. Year-round correlations with EC (r=0.76), S (r=0.9), Br (r=0.65), K (r=0.64), V (r=0.78), Ni (r=0.76), Fe (r=0.72), Mn (r=0.37), Si (r=0.68), Al (r=0.59), and Ti (r=0.68), summarized in Figure S13, imply mixed source combustion from regional anthropogenic and natural sources. First, the presence of substantial concentrations of V and Ni and their correlations at Olympic represent processed emissions from residual oil burning by large industrial sources and marine vessels at the Port of Seattle and greater metropolitan area (Wu et al., 2007). Ni and V are the two most abundant elements in petroleum (Barwise, 1990) and are therefore used as markers for the oil extraction and refinery operations. The contribution of oil combustion from marine transportation sources, such as ferries or container ships, is consistent with previous study by Kotchenruther (2013), who examined monthly average particulate matter attributed to marine vessels using residual fuel oil for 14 monitoring sites in the U.S. Pacific Northwest between 2007 and 2010. The authors found at Olympic marine vessel emissions showed a seasonal cycle with lower impacts in winter months and higher impacts in summer months which is similar to a trend identified in our study in Figure 9. Third, correlations with EC, K, and Br indicate the presence of vegetative burning emissions (including residential wood and wildfire burning). While the residential wood burning emissions may be higher during the heating season, which runs from October until February (Liu et al., 2003), the biomass burning emissions from forest fires tend be more prevalent in summer. In their IMPROVE speciation study, Malm et al. (2004) reports wild and prescribed fire seasons in late summer and early fall in the northwestern United States are responsible a pronounced increase in local OM concentrations. Additionally, the analysis from the aerosol measurement campaign at the Peak of Whistler Mountain, British Columbia, in 2009 confirmed that the mean OM concentration in summer is significantly higher than that in spring due to emissions from extensive local fire episodes (Takahama et al., 2011). Finally, correlations with Fe, Mn, EC, and S reveal emissions from diesel fuels used to operate commercial, transit, and passenger vehicles. Fe, Mn, and S in combination represent diesel exhaust tracers Calvo et al. (2013) and higher concentrations of sulfate and ammonium indicate more processed tail-pipe emissions from diesel vehicles (Wu et al., 2007). Main origins of diesel emissions are Seattle area freeways and highways, such as the Interstate-5 corridor, which support a large percentage of diesel truck traffic (Kim and Hopke, 2008). Additional contributors include diesel-powered locomotives and marine and port activities at Seattle and Tacoma ports.

The remaining OM at Olympic ranging from approximately 31% in winter to 5% in summer is accounted for by Hydroxyl factor. Its year-round correlations with Na (r=0.64) and elevated Na concentrations in Figure 10 (Na levels are consistently an order of magnitude higher than those of mineral dust markers) suggest the origin is marine aerosol. Na concentrations peak in fall, which corresponds to a season with high precipitation in the Pacific Northwest.

3.5.4. Remaining sites: Mesa Verde, Proctor Maple, and St Marks

All 3 sites (Mesa Verde, Proctor Maple, and St Marks) show a similar seasonal trend where organics peak in summer and decrease through fall and winter. Organic composition at Proctor Maple, a site located along the eastern seaboard, is nearly two times Mesa Verde, which can be an indication of year-long local fossil fuel emissions along the East Coast. OM at Mesa Verde is dominated by Hydrocarbon factor in summer when it accounts for 68% of local OM. Its correlation with EC, Zn, and Cu indicates motor vehicle emission source, similar to Phoenix. The high summer concentration of hydrocarbon source in Figure 10 may correspond to increased traffic during the holiday season. Other minor seasonal sources may include forest fires or agriculture waste burning. Hydroxyl is the second major contributor to OM at Mesa Verde where it makes up over 60% of OM in spring and 20% in remaining seasons. Throughout the year hydroxyl is correlated with major mineral dust tracers: Si (r=0.9), Al (r=0.9), and Ti (r=0.88), and diesel combustion tracers: Fe (r=0.93), Zn (r=0.68), and Mn (r=0.61). Figure 10 confirms elevated concentrations of Si, Al, and Ti, and ratio high ratio of Si:Na (10:1). The abundance profiles of mineral dust elements in Mesa Verde are similar to those in Phoenix where mineral dust was also identified as the major contributor to Hydroxyl factor. Presence of Fe, Zn, and Mn imply major sources of dust aerosols in Mesa Verde include fugitive dust and paved road dust resuspended by motor vehicles. Similarly, geological dust from paved and unpaved roads and open land was previously identified as one of the main contributors to $PM_{2.5}$ in Colorado (Watson et al., 2001).

Both Proctor Maple and St Mark exhibit distinct seasonal pattern where organic aerosol concentrations peak in summer. Their OM is dominated by Hydrocarbon factor, which accounts for 58% and 64% of the average annual OM. At Proctor Maple, Hydrocarbon factor is correlated with EC (r=0.73), K (r=0.71), and Br (r=0.67), which indicated wood smoke and motor vehicle emissions. In an earlier study on source apportionment at Underhill, VT (another monitoring site located < 20 km away from Proctor Maple), potential source contribution function (PSCF) revealed a strong local contribution from residential wood combustion in northern New England and southwestern Quebec (Polissar et al., 2001b). The contributions from the woodsmoke were present throughout the year but showed the highest concentrations during winter season. Our two combustion factors, Processed 1 and Processed 2, are correlated with S (r=0.85), Se (r=0.66), Zn (r=0.53), Ni (r=0.55), Cr (r=0.69), V (r=0.40), and Ti (r=0.44), which have been associated with emissions from coal-fired power plants, oil combustion sources, and smelters (Song et al., 2001). Figures 8 suggests their contributions are steady throughout the year, which is consistent with the fact that the East Coast region relies on the use of heavy oil as fuel for power generation and heating all-year around. The PSCF from (Polissar et al., 2001b) identified large potential source areas in upstate New York, Pennsylvania, and other midwestern states towards the coal combustion emissions at Underhill and areas along the East Coast and Mid-Atlantic states towards the oil combustion emissions. Similar to Olympic, at Proctor Maple we also observe correlations between two PMF factors, where Hydrocarbon and Processed 2 are correlated in summer (r=0.61). Lower correlations during the remaining seasons ($0.21 < r < 0.54$) indicate that the two factors did not covary over time to that extent. However, during summer, the combined factor is associated with summertime biogenic emissions as well as secondary aerosol production from Processed 2 combustion. Photochemistry has been known to enhance the sulfate production in the Northeast (Lioy et al., 1977), which could also be one of the reasons behind higher concentrations of the OM apportioned to the original Processed 2 during summer and its correlation with Hydrocarbon factor. As a result, this combined factor is labeled as "polluted biogenic" factor (Figure S14). Its factor profile is consistent with biogenic PMF factor identified during a field experiment at Appledore Island, New Hampshire, in 2004 (Bahadur et al., 2010).

At St Marks Hydrocarbon factor is correlated with only EC (r=0.71), K (r=0.44), and Br (r=0.60). Hydrocarbon emissions gradually increase from winter and peak in summer, which coincides with magnitudes and seasonal cycles of fire-related activities and terrestrial biogenic emissions. Prescribed burning is one of the largest contributor to aerosols in the Southeastern United States (Brewer and Moore, 2009). The region performs more then 50,000 prescribed fire treatments every year (Kobziar et al., 2015) with the goals to restore ecosystem and reduce wildfire hazard. The wildland activity, prescribed burning, and agricultural burning pick up in winter and spring (Zhang et al., 2010; Morris et al., 2006) which agrees with episodic events measured in St Marks in our study (Figure 9). While controlled prescribed fires are confined and smaller in scale than wildfires, prescribed fire emissions typically lead to a 50% increase in mean OC and EC concentrations in St Marks area (Zeng et al., 2008). Finally, summertime OC levels from Hydrocarbon factor can be enhanced by biogenic vegetation emissions from certain vegetation types, notably oaks. A study by Tanner and Zielinska (1994) identified biogenic hydrocarbons emitted from oak trees to be significant contributors to volatile organic compounds between July and August, providing potential biogenic precursors for formation of biogenic organic aerosols. The remaining OM (34%) at St Marks is attributed to two secondary aerosol factors, Processed 1 and Processed 2. Their stable presence throughout the year and correlations with S (r=0.84), Mn (r=0.64), and Zn (r=0.54) are linked to emissions from electric generating utilities and industrial activity (Brewer and Moore, 2009; Blanchard et al., 2013).

3.6. OM uncertainty

Uncertainties in PMF solutions can arise from random errors and rotational ambiguity (Paatero et al., 2014; Brown et al., 2015). In this work, we focus on characterizing the variations within our selected class of solutions, thereby focusing on a limited number of rotations that have similar interpretations to those presented in this work. To examine the variability associated with resulting OM apportioned to Hydrocarbon, Hydroxyl, Processed 1, and Processed 2 factors at each site, we use all 73 PMF solutions belonging to the first ("accepted") configuration (Figure S8). In our uncertainty calculation, we discard the second and third configurations (Figures S9a and S9b) for the reasons mentioned in Section 3.3. Therefore, we calculate the mean annual OM averaged over these 73 solutions apportioned to each factor at each site. Table 4 reports site-by-site mean annual OM estimates with their standard deviation, expressed in mass units and as a percentage. 4-factor PMF solutions with distinct seed and FPEAK parameters yield quite uniform OM results, as no standard deviation is $> 5\%$ of the total OM measured at the site. These estimates do not represent the overall uncertainty of solving the difficult inverse problem, but those due to the range of numerical realizations generated by PMF for the solution class selected for study in this work.

4. Conclusions

To facilitate future source analysis in light of available long-term speciated aerosol records, this work establishes a method for systematic interpretation of multi-site, multi-season source apportionment of OM with FT-IR measurements. Our results from this six-site reference study for 2011 IMPROVE monitoring network samples demonstrate that composition and sources of organic aerosols vary throughout the seasons. Four factor components (Processed 1, Processed 2, Hydrocarbon, and Hydroxyl) that explain the major variations in OM were observed across all sites and seasons, and were attributed to a common set of sources.

Phoenix experienced the highest organic aerosol loadings, up to 2.2 μg m^{-3} in winter. Its visibility impairment is dominated by emissions from local anthropogenic activities, such as residential wood burning, motor vehicle and truck traffic, and construction. OM in Trapper Creek was dominated by natural sources and anthropogenic sources not readily controllable by the local jurisdiction, including sea spray aerosol, volcanic activity, natural wildland fires, and international emissions from shipping lanes and power plant operations. The OM composition at Olympic is directly reflected by its location close to the Port of Seattle and greater metropolitan area. Port, industrial, and commercial activities are the major contributors to local visibility impairment. The emission sources range from mobile sources (including road vehicles, marine engines, locomotive engines, and engines from construction equipment), to industrial point sources and area sources (including local wood burning). The organic aerosol concentrations at remaining sites (Mesa Verde, Proctor Maple, and St Marks) show similar seasonal cycle, where highest OM concentrations occur in summer (or early spring in St Marks). Mobile sources, biomass burning, and natural vegetative emissions are important at Mesa Verde and Proctor Maple, while at St Marks emissions from prescribed fires and agricultural clearing are the most significant contributor to visibility impairment.

One of the important avenues for future research include extending multi-site source apportionment studies to different years or different networks (eg, CSN or SEARCH) to generalize man-made and natural contributions to existing air pollution and visibility impairments. Uniform cluster memberships confirm the potential for inclusion of sites that had been previously unexamined or unmonitored. The main steps for factor solution selection include i) select the number of factors, ii) sample over a wider seed range to select the final solution profile based on chemical consistency and a frequentist or consensus selection approach (after having enumerated a wide range of possible solutions). In our case, the two criteria converged to the same estimate.

Compared to single-site organic aerosol studies, such as those previously reported in intensive, short-term field campaigns (Russell et al., 2011), the multi-site analysis presents several unique aspects. The first one is the origin of multiple factor-source associations, as depicted in Figure 1. When samples from multiple sites and multiple seasons are aggregated in a single analysis, one factor can be attributed to multiple sources that occur in different places or at different times. For instance, in Phoenix Hydrocarbon factor encapsulated wood burning and traffic emissions (mixed source combustion B) while in St Marks it represented mostly emissions from agricultural and biomass burning activities. Similarly, Hydroxyl factor is attributed to either marine or mineral dust aerosol production depending on the location. The multiple factor-source correspondence provides evidence for the similarity in chemical formation, properties, and functional group composition of various anthropogenic and biogenic sources, such as the backbone of alkane hydrocarbon precursor. Second, OA sources at several locations were attributed to factors with high co-variation in time (eg, Processed 1 and Hydrocarbon at Olympic during all seasons and Processed 2 and Hydrocarbon at Proctor Maple during summer). The former suggests similarities in chemical formation and atmospheric processing of secondary OA and hydrocarbon compounds that occurred during their transport from the origin to the measurement site at Olympic. The latter implies the role of photochemistry in co-incident production and processing of biogenic and less oxygenated fossil fuel combustion aerosols at Proctor Maple. Third, the factor analysis identified two fossil fuel combustion factors with low correlation in time and associations with different combustion markers at each site. Their relative contributions to the local OM varied substantially amongst all sites; 9.5 - 33.5% for Processed 1 and 4.2 - 22.1 % for Processed 2, suggesting that the yields of secondary organic aerosol formation are specific to the given location.

On the whole, multi- and single-site factor analyses may provide qualitatively similar factor components. In the exploratory stage of our work, we performed the PMF analysis using samples from Olympic site only, which generated factor solutions with similar chemical profiles. For instance, a 4-factor solution from Olympic measurements also resolved two anthropogenic combustions factors, one hydroxyl factor, and one hydrocarbon-like factor. A brief summary is included in Supporting Information in Figure S10. The convergence of multi-site and single-site factor profiles confirms the robustness of our results and supports the findings of Liu et al. (2009) who also arrived at similar chemical profiles from "combined" and "individual" PMF analyses. Studies focusing on individual sites can additionally incorporate meteorological back-trajectory analyses (e.g., Seibert and Frank, 2004; Pekney et al., 2006; Stein et al., 2015) to further confirm sources impacting specific regions.

Finally, it is worth recognizing that 24-hour integrated FT-IR measurements in this study may obfuscate distinctions among individual sources of the existing aerosols at a very fine level, especially if they are chemically similar or if there are insufficient variations in source strengths across days. For example, obtaining statistically resolvable components of different fuel types (such as diesel, gasoline, ship, or motor oil) or burning emissions (such as wildfires, agricultural burning, or home wood burning) may be challenging since their chemical profiles are largely composed of long-chain alkane hydrocarbons. Discriminating features in spectra [e.g., that distinguish among fuel types (Guzman-Morales et al., 2014) or terrestrial emissions (Corrigan et al., 2013)] are not specifically targeted in this current inverse modeling strategy, but could be given higher weight or investigated in a supervised learning framework. The assumption of static source profiles can further be relaxed to obtain profiles "localized" in time (e.g., Baltensperger, 2016). Conversely, additional constraints can be placed on the constancy of seasonal profiles associated with each factor in three-way factor analyses (e.g., Tucker, 1966; Harshman and Lundy, 1994; de Juan et al., 1998; Hopke et al., 1998; Ulbrich et al., 2012) to further explore interpretations possible in such network measurements. This work establishes a base case interpretation against which results obtained by such approaches can be compared.

Acknowledgements. The authors acknowledge EPFL and the IMPROVE program (National Park Service cooperative agreement P11AC91045) for funding, and the IMPROVE team at UC Davis for performing the sample handling and site maintenance for all IMPROVE sites.

References

Aguilera, I., Eeftens, M., Meier, R., Ducret-Stich, R.E., Schindler, C., Ineichen, A., Phuleria, H.C., Probst-Hensch, N., Tsai, M.Y., Kunzli, N., 2015. Land use regression models for crustal and traffic-related PM2.5 constituents in four areas of the SAPALDIA study. Environmental Research 140, 377–384. doi:10.1016/j.envres.2015.04.011.

Ahlm, L., Shakya, K.M., Russell, L.M., Schroder, J.C., Wong, J.P.S., Sjostedt, S.J., Hayden, K.L., Liggio, J., Wentzell, J.J.B., Wiebe, H.A., Mihele, C., Leaitch, W.R., Macdonald, M., 2013. Temperature-dependent accumulation mode particle and cloud nuclei concentrations from biogenic sources during WACS 2010. Atmospheric Chemistry and Physics 13, 3393–3407. doi:10.5194/acp-13-3393-2013.

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., Jimenez, J.L., 2008. O/C and OM/OC ratios of primary, secondary, and ambient organic aerosols with high-resolution time-of-flight aerosol mass spectrometry. Environmental Science & Technology 42, 4478–4485. doi:10.1021/es703009q.

Aluwihare, L.I., Repeta, D.J., Chen, R.F., 1997. A major biopolymeric component to dissolved organic carbon in surface sea water. Nature 387, 166–169. doi:10.1038/387166a0.

Bahadur, R., Uplinger, T., Russell, L.M., Sive, B.C., Cliff, S.S., Millet, D.B., Goldstein, A., Bates, T.S., 2010. Phenol groups in Northeastern US submicrometer aerosol particles produced from seawater sources. Environmental Science & Technology 44, 2542–2548. doi:10.1021/es9032277.

Baltensperger, U., 2016. Spiers memorial lecture introductory lecture: chemistry in the urban atmosphere. Faraday Discussions 189, 9–29. doi:10.1039/c6fd00065g.

Barrie, L.A., 1986. Arctic air-pollution - an overview of current knowledge. Atmospheric Environment 20, 643–663. doi:10.1016/0004-6981(86)90180-0.

Barwise, A.J.G., 1990. Role of nickel and vanadium in petroleum classification. Energy & Fuels 4, 647–652. doi:10.1021/ef00024a005.

Bianchi, G., Gamba, A., Limiroli, R., Pozzi, N., Elster, R., Salamini, F., Bartels, D., 1993. The unusual sugar composition in leaves of the resurrection plant myrothamnus-flabellifolia. Physiologia Plantarum 87, 223–226. doi:10.1111/j.1399-3054.1993.tb00146.x.

Blanchard, C.L., Hidy, G.M., Tanenbaum, S., Edgerton, E.S., Hartsell, B.E., 2013. The Southeastern Aerosol Research and Characterization (SEARCH) study: Temporal trends in gas and PM concentrations and composition, 1999-2010. Journal of the Air & Waste Management Association 63, 247–259. doi:10.1080/10962247.2012.748523.

Breider, T.J., Mickley, L.J., Jacob, D.J., Wang, Q.Q., Fisher, J.A., Chang, R.Y.W., Alexander, B., 2014. Annual distributions and sources of arctic aerosol components, aerosol optical depth, and aerosol absorption. Journal of Geophysical Research-Atmospheres 119, 4107–4124. doi:10.1002/2013jd020996.

Brewer, P., Moore, T., 2009. Source contributions to visibility impairment in the Southeastern and Western United States. Journal of the Air & Waste Management Association 59, 1070–1081. doi:10.3155/1047-3289.59.9.1070.

Brown, S.G., Eberly, S., Paatero, P., Norris, G.A., 2015. Methods for estimating uncertainty in pmf solutions: Examples with ambient air and water quality data and guidance on reporting pmf results. Science of the Total Environment 518, 626–635. doi:10.1016/j.scitotenv.2015.01.022.

Brown, S.G., Frankel, A., Raffuse, S.M., Roberts, P.T., Hafner, H.R., Anderson, D.J., 2007. Source apportionment of fine particulate matter in Phoenix, AZ, using positive matrix factorization. Journal of the Air & Waste Management Association 57, 741–752. doi:10.3155/1047-3289.57.6.741.

Brown, S.G., Lee, T., Norris, G.A., Roberts, P.T., Collett, J.L., Paatero, P., Worsnop, D.R., 2012. Receptor modeling of near-roadway aerosol mass spectrometer data in las vegas, nevada, with epa pmf. Atmospheric Chemistry and Physics 12, 309–325. doi:10.5194/acp-12-309-2012.

Calvo, A.I., Alves, C., Castro, A., Pont, V., Vicente, A.M., Fraile, R., 2013. Research on aerosol sources and chemical composition: Past, current and emerging issues. Atmospheric Research 120, 1–28. doi:10.1016/j.atmosres.2012.09.021.

Canagaratna, M.R., Jimenez, J.L., Kroll, J.H., Chen, Q., Kessler, S.H., Massoli, P., Hildebrandt Ruiz, L., Fortner, E., Williams, L.R., Wilson, K.R., Surratt, J.D., Donahue, N.M., Jayne, J.T., Worsnop, D.R., 2015. Elemental ratio measurements of organic compounds using aerosol mass spectrometry: characterization, improved calibration, and implications. Atmospheric Chemistry and Physics 15, 253–272. doi:10.5194/acp-15-253-2015.

Canonaco, F., Crippa, M., Slowik, J.G., Baltensperger, U., Prevot, A.S.H., 2013. SoFi, an IGOR-based interface for the efficient use of the generalized multilinear engine (ME-2) for the source apportionment: ME-2 application to aerosol mass spectrometer data. Atmospheric Measurement Techniques 6, 3649–3661. doi:10.5194/amt-6-3649-2013.

Coates, J., 2000. Encyclopedia of Analytical Chemistry. John Wiley & Sons Ltd, Chichester. chapter Interpretation of Infrared Spectra, A Practical Approach. pp. 10815–10837.

Corrigan, A.L., Russell, L.M., Takahama, S., Äijälä, M., Ehn, M., Junninen, H., Rinne, J., Petäjä, T., Kulmala, M., Vogel, A.L., Hoffmann, T., Ebben, C.J., Geiger, F.M., Chhabra, P., Seinfeld, J.H., Worsnop, D.R., Song, W., Auld, J., Williams, J., 2013. Biogenic and biomass burning organic aerosol in a boreal forest at hyytiälä, Finland, during HUMPPA-COPEC 2010. Atmospheric Chemistry and Physics 13, 12233–12256. doi:10.5194/acp-13-12233-2013.

Day, D.A., Liu, S., Russell, L.M., Ziemann, P.J., 2010. Organonitrate group concentrations in submicron particles with high nitrate and organic fractions in coastal southern california. Atmospheric Environment 44, 1970–1979. doi:10.1016/j.atmosenv.2010.02.045.

Dillner, A.M., Takahama, S., 2015a. Predicting ambient aerosol thermal-optical reflectance measurements from infrared spectra: elemental carbon. Atmospheric Measurement Techniques 8, 4013–4023. doi:10.5194/amt-8-4013-2015.

Dillner, A.M., Takahama, S., 2015b. Predicting ambient aerosol thermal-optical reflectance (TOR) measurements from infrared spectra: organic carbon. Atmospheric Measurement Techniques 8, 1097–1109. doi:10.5194/amt-8-1097-2015.

EPA, 1999. Regional Haze State Implementation Milestones. Technical Report. United States Environmental Protection Agency.

Frossard, A.A., Russell, L.M., Burrows, S.M., Elliott, S.M., Bates, T.S., Quinn, P.K., 2014. Sources and composition of submicron organic mass in marine aerosol particles. Journal of Geophysical Research-Atmospheres 119, 12977–13003. doi:10.1002/2014jd021913.

Guzman-Morales, J., Frossard, A., Corrigan, A., Russell, L., Liu, S., Takahama, S., Taylor, J., Allan, J., Coe, H., Zhao, Y., Goldstein, A., 2014. Estimated contributions of primary and secondary organic aerosol from fossil fuel combustion during the CalNex and Cal-Mex campaigns. Atmospheric Environment 88, 330 – 340. doi:http://dx.doi.org/10.1016/j.atmosenv.2013.08.047.

Hamilton, D.S., Lee, L.A., Pringle, K.J., Reddington, C.L., Spracklen, D.V., Carslaw, K.S., 2014. Occurrence of pristine aerosol environments on a polluted planet. Proceedings of the National Academy of Sciences of the United States of America 111, 18466–18471. doi:10.1073/pnas.1415440111.

Hand, J.L., Schichtel, B.A., Malm, W.C., Frank, N.H., 2013. Spatial and temporal trends in PM2.5 organic and elemental carbon across the United States. Advances in Meteorology , 13doi:10.1155/2013/367674.

Hand, J.L., Schichtel, B.A., Malm, W.C., Pitchford, M., Frank, N.H., 2014. Spatial and seasonal patterns in urban influence on regional concentrations of speciated aerosols across the united states. Journal of Geophysical Research-Atmospheres 119, 12832–12849. doi:10.1002/2014jd022328.

Harshman, R.A., Lundy, M.E., 1994. PARAFAC: Parallel Factor Analysis. Comput. Stat. Data Anal. 18, 39–72. doi:10.1016/0167-9473(94)90132-5.

Hawkins, L.N., Russell, L.M., 2010. Oxidation of ketone groups in transported biomass burning aerosol from the 2008 Northern California Lightning Series fires. Atmospheric Environment 44, 4142–4154. doi:10.1016/j.atmosenv.2010.07.036.

Hawkins, L.N., Russell, L.M., Covert, D.S., Quinn, P.K., Bates, T.S., 2010. Carboxylic acids, sulfates, and organosulfates in processed continental organic aerosol over the southeast pacific ocean during VOCALS-REx 2008. Journal of Geophysical Research-Atmospheres 115, 16. doi:10.1029/2009jd013276.

Héberger, K., 2010. Sum of ranking differences compares methods or models fairly. TrAC Trends in Analytical Chemistry 29, 101–109. doi:10.1016/j.trac.2009.09.009.

Henry, R.C., 2003. Multivariate receptor modeling by N-dimensional edge detection. Chemometrics and Intelligent Laboratory Systems 65, 179–189. doi:10.1016/s0169-7439(02)00108-9.

Hopke, P.K., Paatero, P., Jia, H., Ross, R.T., Harshman, R.A., 1998. Three-way (parafac) factor analysis: examination and comparison of alternative computational methods as applied to ill-conditioned data. Chemometrics and Intelligent Laboratory Systems 43, 25 – 42. doi:http://dx.doi.org/10.1016/S0169-7439(98)00077-X.

Hwang, I., Hopke, P.K., 2007. Estimation of source apportionment and potential source locations of pm2.5 at a west coastal {IMPROVE} site. Atmospheric Environment 41, 506 – 518. doi:http://dx.doi.org/10.1016/j.atmosenv.2006.08.043.

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., Worsnop, D.R., 2009. Evolution of organic aerosols in the atmosphere. Science 326, 1525–1529. doi:10.1126/science.1180353.

de Juan, A., Rutan, S.C., Tauler, R., Massart, D.L., 1998. Comparison between the direct trilinear decomposition and the multivariate curve resolution-alternating least squares methods for the resolution of three-way data sets. Chemometrics and Intelligent Laboratory Systems 40, 19–32. doi:10.1016/S0169-7439(98)00003-3.

Keller, H.R., Massart, D.L., 1992. Evolving factor-analysis. Chemometrics and

Intelligent Laboratory Systems 12, 209–224. doi:10.1016/0169-7439(92)80002-L.

Khalil, M.A.K., Rasmussen, R.A., 2003. Tracers of wood smoke. Atmospheric Environment 37, 1211–1222. doi:10.1016/s1352-2310(02)01014-2.

Kim, E., Hopke, P.K., 2008. Source characterization of ambient fine particles at multiple sites in the Seattle area. Atmospheric Environment 42, 6047–6056. doi:10.1016/j.atmosenv.2008.03.032.

Kinney, P.L., Aggarwal, M., Northridge, M.E., Janssen, N.A.H., Shepard, P., 2000. Airborne concentrations of PM2.5 and diesel exhaust particles on harlem sidewalks: A community-based pilot study. Environmental Health Perspectives 108, 213–218. doi:10.2307/3454436.

Kleindienst, T.E., Jaoui, M., Lewandowski, M., Offenberg, J.H., Lewis, C.W., Bhave, P.V., Edney, E.O., 2007. Estimates of the contributions of biogenic and anthropogenic hydrocarbons to secondary organic aerosol at a southeastern US location. Atmospheric Environment 41, 8288–8300. doi:10.1016/j.atmosenv.2007.06.045.

Kobziar, L.N., Godwin, D., Taylor, L., Watts, A.C., 2015. Perspectives on trends, effectiveness, and impediments to prescribed burning in the southern u.s. Forests 6, 561–580. doi:10.3390/f6030561.

Kotchenruther, R.A., 2013. A regional assessment of marine vessel PM2.5 impacts in the U.S. Pacific Northwest using a receptor-based source apportionment method. Atmospheric Environment 68, 103–111. doi:10.1016/j.atmosenv.2012.11.067.

Kuzmiakova, A., Dillner, A.M., Takahama, S., 2016. An automated baseline correction protocol for infrared spectra of atmospheric aerosols collected on polytetrafluoroethylene (Teflon) filters. Atmospheric Measurement Techniques 9, 2615–2631. doi:10.5194/amt-9-2615-2016.

Lanz, V.A., Alfarra, M.R., Baltensperger, U., Buchmann, B., Hueglin, C., Prevot, A.S.H., 2007. Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra. Atmospheric Chemistry and Physics 7, 1503–1522.

Lioy, P.J., Wolff, G.T., Czachor, J.S., Coffey, P.E., Stasiuk, W.N., Romano, D., 1977. Evidence of high atmospheric concentrations of sulfates detected at rural sites in Northeast. Journal of Environmental Science and Health Part a-Environmental Science and Engineering & Toxic and Hazardous Substance Control 12, 1–14.

Liu, L.J.S., Box, M., Kalman, D., Kaufman, J., Koenig, J., Larson, T., Lumley, T., Sheppard, L., Wallace, L., 2003. Exposure assessment of particulate matter for susceptible populations in Seattle. Environmental Health Perspectives 111, 909–918. doi:10.1289/ehp.6011.

Liu, S., Ahlm, L., Day, D.A., Russell, L.M., Zhao, Y.L., Gentner, D.R., Weber, R.J., Goldstein, A.H., Jaoui, M., Offenberg, J.H., Kleindienst, T.E., Rubitschun, C., Surratt, J.D., Sheesley, R.J., Scheller, S., 2012. Secondary organic aerosol formation from fossil fuel sources contribute majority of summertime organic mass at Bakersfield. Journal of Geophysical Research-Atmospheres 117, 21. doi:10.1029/2012jd018170.

Liu, S., Takahama, S., Russell, L.M., Gilardoni, S., Baumgardner, D., 2009. Oxygenated organic functional groups and their sources in single and submicron organic particles in MILAGRO 2006 campaign. Atmospheric Chemistry and Physics 9, 6849–6863. doi:10.5194/acp-9-6849-2009.

Malm, W.C., Schichtel, B.A., Ames, R.B., Gebhart, K.A., 2002. A 10-year spatial and temporal trend of sulfate across the United States. Journal of Geophysical Research-Atmospheres 107, 20. doi:10.1029/2002jd002107.

Malm, W.C., Schichtel, B.A., Pitchford, M.L., Ashbaugh, L.L., Eldred, R.A., 2004. Spatial and monthly trends in speciated fine particle concentration in the United States. Journal of Geophysical Research-Atmospheres 109, 33. doi:10.1029/2003jd003739.

Maria, S.F., Russell, L.M., Turpin, B.J., Porcja, R.J., Campos, T.L., Weber, R.J., Huebert, B.J., 2003. Source signatures of carbon monoxide and organic functional groups in Asian Pacific Regional Aerosol Characterization Experiment (ACE-Asia) submicron aerosol types. Journal of Geophysical Research-atmospheres 108. doi:10.1029/2003JD003703.

McGimsey, R., Maharrey, J., Neal, C., 2014. 2011 Volcanic activity in Alaska - Summary of events and response of the Alaska Volcano Observatory: U.S. Geological Survey Scientific Investigations Report. Technical Report. U.S. Geological Survey.

Morris, R.E., Koo, B., Guenther, A., Yarwood, G., McNally, D., Tesche, T.W., Tonnesen, G., Boylan, J., Brewer, P., 2006. Model sensitivity evaluation for organic carbon using two multi-pollutant air quality models that simulate regional haze in the southeastern United States. Atmospheric Environment 40, 4960–4972. doi:10.1016/j.atmonsenv.2005.09.088.

NASA, 2010. Maps, population, landscape, and climate estimates. URL: http://sedac.ciesin.columbia.edu/data/set/nagdc-population-landscape-climate-estimates-v3/maps?facets=theme:population&facets=region:north%20america.

NOAA, 2011. Climate data online search. URL: https://www.ncdc.noaa.gov/cdo-web/search.

Noblitt, S.D., Berg, K.E., Cate, D.M., Henry, C.S., 2016. Characterizing nonconstant instrumental variance in emerging miniaturized analytical techniques. Analytica Chimica Acta 915, 64–73. doi:10.1016/j.aca.2016.02.023.

Paatero, P., 1997. Least squares formulation of robust non-negative factor analysis. Chemometrics and Intelligent Laboratory Systems 37, 23–35. doi:10.1016/s0169-7439(96)00044-5.

Paatero, P., Eberly, S., Brown, S.G., Norris, G.A., 2014. Methods for estimating uncertainty in factor analytic solutions. Atmospheric Measurement Techniques 7, 781–797. doi:10.5194/amt-7-781-2014.

Paatero, P., Hopke, P.K., Song, X.H., Ramadan, Z., 2002. Understanding and controlling rotations in factor analytic models. Chemometrics and Intelligent Laboratory Systems 60, 253–264. doi:10.1016/s0169-7439(01)00200-3.

Paatero, P., Tapper, U., 1994. Positive matrix factorization - a nonnegative factor model with optimal utilization of error-estimates of data values. Environmetrics 5, 111–126. doi:10.1002/env.3170050203.

Pavia, D., Lampman, G., Kriz, G., 2008. Introduction to Spectroscopy. Brooks/Cole Pub Co., Belmont, CA.

Pekney, N.J., Davidson, C.I., Zhou, L.M., Hopke, P.K., 2006. Application of pscf and cpf to pmf-modeled sources of pm2.5 in pittsburgh. Aerosol Science and Technology 40, 952–961. doi:10.1080/02786820500543324.

Pio, C., Alves, C., Duarte, A., 2001. Organic components of aerosols in a forested area of central Greece. Atmospheric Environment 35, 389–401. doi:10.1016/s1352-2310(00)00135-7.

Polissar, A.V., Hopke, P.K., Harris, J.M., 2001a. Source regions for atmospheric aerosol measured at Barrow, Alaska. Environmental Science & Technology 35, 4214–4226. doi:10.1021/es0107529.

Polissar, A.V., Hopke, P.K., Poirot, R.L., 2001b. Atmospheric aerosol over Vermont: Chemical composition and sources. Environmental Science & Technology 35, 4604–4621. doi:10.1021/es0105865.

Quinn, P.K., Shaw, G., Andrews, E., Dutton, E.G., Ruoho-Airola, T., Gong, S.L., 2007. Arctic haze: current trends and knowledge gaps. Tellus Series B-

Chemical and Physical Meteorology 59, 99–114. doi:10.1111/j.1600-0889.2006.00238.x.

Ramadan, Z., Song, X.H., Hopke, P.K., 2000. Identification of sources of Phoenix aerosol by positive matrix factorization. Journal of the Air & Waste Management Association 50, 1308–1320.

Rau, J.A., 1989. Composition and size distribution of residential wood smoke particles. Aerosol Science and Technology 10, 181–192. doi:10.1080/02786828908959233.

Rinaldi, M., Gilardoni, S., Paglione, M., Sandrini, S., Fuzzi, S., Massoli, P., Bonasoni, P., Cristofanelli, P., Marinoni, A., Poluzzi, V., Decesari, S., 2015. Organic aerosol evolution and transport observed at Mt. Cimone (2165 m a.s.l.), Italy, during the PEGASOS campaign. Atmospheric Chemistry and Physics 15, 11327–11340. doi:10.5194/acp-15-11327-2015.

Russell, L.M., Bahadur, R., Hawkins, L.N., Allan, J., Baumgardner, D., Quinn, P.K., Bates, T.S., 2009. Organic aerosol characterization by complementary measurements of chemical bonds and molecular fragments. Atmospheric Environment 43, 6100–6105. doi:10.1016/j.atmosenv.2009.09.036.

Russell, L.M., Bahadur, R., Ziemann, P.J., 2011. Identifying organic aerosol sources by comparing functional group composition in chamber and atmospheric particles. Proceedings of the National Academy of Sciences of the United States of America 108, 3516–3521. doi:10.1073/pnas.1006461108.

Russell, L.M., Hawkins, L.N., Frossard, A.A., Quinn, P.K., Bates, T.S., 2010. Carbohydrate-like composition of submicron atmospheric particles and their production from ocean bubble bursting. Proceedings of the National Academy of Sciences of the United States of America 107, 6652–6657. doi:10.1073/pnas.0908905107.

Ruthenburg, T.C., Perlin, P.C., Liu, V., McDade, C.E., Dillner, A.M., 2014. Determination of organic matter and organic matter to organic carbon ratios by infrared spectroscopy with application to selected sites in the improve network. Atmospheric Environment 86, 47–57. doi:10.1016/j.atmosenv.2013.12.034.

Schwartz, R.E., Russell, L.M., Sjostedt, S.J., Vlasenko, A., Slowik, J.G., Abbatt, J.P.D., Macdonald, A.M., Li, S.M., Liggio, J., Toom-Sauntry, D., Leaitch, W.R., 2010. Biogenic oxidized organic functional groups in aerosol particles from a mountain forest site and their similarities to laboratory chamber products. Atmospheric Chemistry and Physics 10, 5075–5088. doi:10.5194/acp-10-5075-2010.

Seibert, P., Frank, A., 2004. Source-receptor matrix calculation with a lagrangian particle dispersion model in backward mode. Atmospheric Chemistry and Physics 4, 51–63. doi:10.5194/acp-4-51-2004.

Shaw, P.M., Russell, L.M., Jefferson, A., Quinn, P.K., 2010. Arctic organic aerosol measurements show particles from mixed combustion in spring haze and from frost flowers in winter. Geophysical Research Letters 37, 5. doi:10.1029/2010gl042831.

Skoog, D., Holler, F., Crouch, S., 2017. Principles of Instrumental Analysis. 7th ed., Brooks/Cole Pub Co., Belmont, CA.

Sokal, R., Rohlf, F., 1981. Biometry: The Principles and Practice of Statistics in Biological Research. W. H. Freeman & Company, New York, NY.

Song, X.H., Polissar, A.V., Hopke, P.K., 2001. Sources of fine particle composition in the northeastern us. Atmospheric Environment 35, 5277–5286. doi:10.1016/s1352-2310(01)00338-7.

Sowlat, M.H., Hasheminassab, S., Sioutas, C., 2016. Source apportionment of ambient particle number concentrations in central los angeles using positive matrix factorization (pmf). Atmospheric Chemistry and Physics 16, 4849–4866. doi:10.5194/acp-16-4849-2016.

Stein, A.F., Draxler, R.R., Rolph, G.D., Stunder, B.J.B., Cohen, M.D., Ngan, F., 2015. NOAAs HYSPLIT Atmospheric Transport and Dispersion Modeling System. Bulletin of the American Meteorological Society 96, 2059–2077. doi:10.1175/BAMS-D-14-00110.1.

Takahama, S., 2015. PMF-tools. URL: https://github.com/stakahama/pmf-tools, doi:10.5281/zenodo.34249.

Takahama, S., Dillner, A.M., 2015. Model selection for partial least squares calibration and implications for analysis of atmospheric organic aerosol samples with mid-infrared spectroscopy. Journal of Chemometrics , n/a–n/adoi:10.1002/cem.2761.

Takahama, S., Johnson, A., Morales, J.G., Russell, L.M., Duran, R., Rodriguez, G., Zheng, J., Zhang, R., Toom-Sauntry, D., Leaitch, W.R., 2013a. Submicron organic aerosol in Tijuana, Mexico, from local and Southern California sources during the CalMex campaign. Atmospheric Environment 70, 500–512. doi:10.1016/j.atmosenv.2012.07.057.

Takahama, S., Johnson, A., Russell, L.M., 2013b. Quantification of carboxylic and carbonyl functional groups in organic aerosol infrared absorbance spectra. Aerosol Science and Technology 47, 310–325. doi:10.1080/02786826.2012.752065.

Takahama, S., Liu, S., Russell, L.M., 2010. Coatings and clusters of carboxylic acids in carbon-containing atmospheric particles from spectromicroscopy and their implications for cloud-nucleating and optical properties. Journal of Geophysical Research-Atmospheres 115, 21. doi:10.1029/2009jd012622.

Takahama, S., Schwartz, R.E., Russell, L.M., Macdonald, A.M., Sharma, S., Leaitch, W.R., 2011. Organic functional groups in aerosol particles from burning and non-burning forest emissions at a high-elevation mountain site. Atmospheric Chemistry and Physics 11, 6367–6386. doi:10.5194/acp-11-6367-2011.

Takemura, T., Uno, I., Nakajima, T., Higurashi, A., Sano, I., 2002. Modeling study of long-range transport of Asian dust and anthropogenic aerosols from East Asia. Geophysical Research Letters 29, 4. doi:10.1029/2002gl016251.

Tanner, R.L., Zielinska, B., 1994. Determination of the biogenic emission rates of species contributing to VOC in the San-Joaquin Valley of California. Atmospheric Environment 28, 1113–1120. doi:10.1016/1352-2310(94)90288-7.

Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31, 279–311. doi:10.1007/BF02289464.

Ulbrich, I.M., Canagaratna, M.R., Cubison, M.J., Zhang, Q., Ng, N.L., Aiken, A.C., Jimenez, J.L., 2012. Three-dimensional factorization of size-resolved organic aerosol mass spectra from mexico city. Atmospheric Measurement Techniques 5, 195–224. doi:10.5194/amt-5-195-2012.

Ulbrich, I.M., Canagaratna, M.R., Zhang, Q., Worsnop, D.R., Jimenez, J.L., 2009. Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data. Atmospheric Chemistry and Physics 9, 2891–2918. doi:10.5194/acp-9-2891-2009.

Viana, M., Kuhlbusch, T.A.J., Querol, X., Alastuey, A., Harrison, R.M., Hopke, P.K., Winiwarter, W., Vallius, A., Szidat, S., Prevot, A.S.H., Hueglin, C., Bloemen, H., Wahlin, P., Vecchi, R., Miranda, A.I., Kasper-Giebl, A., Maenhaut, W., Hitzenberger, R., 2008. Source apportionment of particulate matter in Europe: A review of methods and results. Journal of Aerosol Science 39, 827–849. doi:10.1016/j.jaerosci.2008.05.007.

Wagstrom, K.M., Pandis, S.N., 2011. Source-receptor relationships for fine particulate matter concentrations in the eastern united states. Atmospheric Environment 45, 347–356. doi:10.1016/j.atmosenv.2010.10.019.

Ward, Jr., J., 1963. Hierarchical grouping to optimize an objective function. Journal of the American statistical association 58, 236–244.

Warneke, C., Bahreini, R., Brioude, J., Brock, C.A., de Gouw, J.A., Fahey, D.W., Froyd, K.D., Holloway, J.S., Middlebrook, A., Miller, L., Montzka, S., Murphy, D.M., Peischl, J., Ryerson, T.B., Schwarz, J.P., Spackman, J.R., Veres, P., 2009. Biomass burning in Siberia and Kazakhstan as an important source for haze over the Alaskan Arctic in April 2008. Geophysical Research Letters 36, 6. doi:10.1029/2008gl036194.

Watson, J.G., Chow, J.C., Houck, J.E., 2001. PM2.5 chemical source profiles for vehicle exhaust, vegetative burning, geological material, and coal burning in Northwestern Colorado during 1995. Chemosphere 43, 1141–1151. doi:10.1016/s0045-6535(00)00171-5.

Weakley, A.T., Takahama, S., Dillner, A.M., 2016. Ambient aerosol composition by infrared spectroscopy and partial least-squares in the chemical speciation network: Organic carbon with functional group identification. Aerosol Science and Technology 50, 1096–1114. doi:10.1080/02786826.2016.1217389.

Williams, B.J., Goldstein, A.H., Kreisberg, N.M., Hering, S.V., Worsnop, D.R., Ulbrich, I.M., Docherty, K.S., Jimenez, J.L., 2010. Major components of atmospheric organic aerosol in southern California as determined by hourly measurements of source marker compounds. Atmospheric Chemistry and Physics 10, 11577–11603. doi:10.5194/acp-10-11577-2010.

Wu, C.F., Larson, T.V., Wu, S.Y., Williamson, J., Westberg, H.H., Liu, L.J.S., 2007. Source apportionment of PM2.5 and selected hazardous air pollutants in Seattle. Science of the Total Environment 386, 42–52. doi:10.1016/j.scitotenv.2007.07.042.

Zeng, T., Wang, Y.H., Yoshida, Y., Tian, D., Russell, A.G., Barnard, W.R., 2008. Impacts of prescribed fires on air quality over the Southeastern United States in spring based on modeling and ground/satellite measurements. Environmental Science & Technology 42, 8401–8406. doi:10.1021/es800363d.

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.F., Salcedo, D., Onasch, T., Jayne, J.T., Miyoshi, T., Shimono, A., Hatakeyama, S., Takegawa, N., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Williams, P., Bower, K., Bahreini, R., Cottrell, L., Griffin, R.J., Rautiainen, J., Sun, J.Y., Zhang, Y.M., Worsnop, D.R., 2007. Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenically-influenced Northern Hemisphere midlatitudes. Geophysical Research Letters 34, L13801. doi:10.1029/2007GL029979.

Zhang, Q., Jimenez, J.L., Canagaratna, M.R., Ulbrich, I.M., Ng, N.L., Worsnop, D.R., Sun, Y.L., 2011. Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review. Analytical and Bioanalytical Chemistry 401, 3045–3067. doi:10.1007/s00216-011-5355-y.

Zhang, X., Hecobian, A., Zheng, M., Frank, N.H., Weber, R.J., 2010. Biomass burning impact on PM2.5 over the southeastern us during 2007: integrating chemically speciated frm filter measurements, modis fire counts and pmf analysis. Atmospheric Chemistry and Physics 10, 6839–6853. doi:10.5194/acp-10-6839-2010.

Zhou, L.M., Kim, E., Hopke, P.K., Stanier, C., Pandis, S.N., 2005. Mining airborne particulate size distribution data by positive matrix factorization. Journal of Geophysical Research-atmospheres 110, D07S19. doi:10.1029/2004JD004707.

Zielinska, B., Fujita, E., Sagebiel, J., Harshfield, G., Uberna, E., Hayes, T., Keene, F., 1998. Arizona hazardous air pollutants monitoring program. Journal of the Air & Waste Management Association 48, 1038–1050.

Tables

Table 1: IMPROVE 2011 site characteristics

Site	Type	Latitude, longitude	Elevation (m)	Temperature range (C)	Precipitation (cm)
Mesa Verde, CO	Rural	37.2, -108.5	662	16.1 - 2.5	39.9
Olympic, WA	Rural	48.0, -123	182	14.5 - 4.0	128.8
Phoenix, AZ	Urban	33.5, -112.1	104	30.6 - 17.4	11.7
Proctor Maple, VT	Rural	44.5, -72.9	122	12.9 - 3.6	90.9
St Marks, FL	Rural	30.1, -84.2	2	28.5 - 12.6	95.5
Trapper Creek, AK	Rural	62.3, -150.3	47	6.2 - -0.83	40.6 (snow = 236.5)

Table 2: Standards grouped by functional groups and compound types

Functional Group	Wavenumber (cm^{-1})	Compound Types
Ammonium	$\nu_1 = 3227$, $\nu_2 = 3059$	Ammonium sulfate (49)
Alcohol	$\nu_1 = 3404$	1-Docosanol (3), D-Glucose (8), Levoglucosan (39), Fructose (17)
Alkane	$\nu_1 = 2917$, $\nu_2 = 2850$	1-Docosanol (3), Arachidyl dodecanoate (25), 12- Tricosanone (39)
Carbonyl	$\nu_1 = 1705$	Suberic acid (25), Malonic acid (13), Arachidyl dodecanoate (25), 12-Tricosanone (39)

Table 3: PMF factor profile characteristics

Chemical profile	Sources	OM:OC	O:C
Processed 1	Anthropogenic fossil fuel combustion (more aged)	2.5	1.0
Processed 2	Anthropogenic fossil fuel combustion (less aged)	1.7	0.4
Hydrocarbon	Burning of plant materials or fossil fuel combustion	1.4	0.2
Hydroxyl	Marine saccharides on sea salt particles or suspended dust	2.4	0.9

Table 4: Summary of site-by-site OA factors, sources, their relative contribution to the annual OM, and seasonal dominance. OM uncertainties are calculated for the selected solution class as described in Section 3.6.

Site:	Factor:	mean OM ($\mu g\, m^{-3}$) ± st. dev.:	% of OM ± st. dev.:	Attributed source:	Ion tracers:	Dominant season	Natural or anthropogenic origin?
Mesa Verde	Hydrocarbon	0.17 ± 0.00	40.1 ± 0.6	Motor vehicle traffic, biomass burning	EC, Zn, Cu	Summer	Mostly anthropogenic
	Hydroxyl	0.12 ± 0.00	28.1 ± 0.9	Mineral dust	Si, Al, Mg, Ti	All-year long	Natural
	Processed 1	0.12 ± 0.01	26.9 ± 1.6	Fossil fuel combustion (aged)	S, Br, Zn	All-year long	Anthropogenic
	Processed 2	0.02 ± 0.00	4.8 ± 0.5	Fossil fuel combustion (less aged)	S	All-year long	Anthropogenic
Olympic	Processed Hydrocarbon	0.28 ± 0.01	81.9 ± 2.2	Mixed source combustion A: i) residual oil burning by industrial sources and marine vessels, ii) soil and dust source, iii) wood burning, iv) diesel-powered locomotives and marine and port activities	EC, S, Br, K, V, Ni, Fe, Mn, Si, Al, and Ti	Summer, fall	Mostly anthropogenic
	Hydroxyl	0.05 ± 0.00	13.9 ± 0.4	Marine aerosol	Na	Fall, winter, spring	Natural
	Processed 2	0.02 ± 0.00	4.2 ± 0.6	Residual oil burning	S, V, Ti, Ni	Spring, summer, fall	Anthropogenic
Phoenix	Hydrocarbon	1.22 ± 0.02	70.6 ± 1.4	Mixed source combustion B: i) residential wood burning, ii) motor vehicles and heavy-duty diesel trucks	EC, Br, Zn, Cu, Fe, Mn, Cr, K, Cl	All-year long	Mostly anthropogenic
	Hydroxyl	0.25 ± 0.01	14.3 ± 0.6	Mineral and road dust	Si, Al, Mg, Ti	Spring, summer, fall	Mixed
	Processed 1	0.17 ± 0.01	9.9 ± 0.5	Fossil fuel combustion (aged)	S	All-year long	Anthropogenic
	Processed 2	0.09 ± 0.01	5.2 ± 0.7	Fossil fuel combustion (less aged)	S, Vi, Ti	All-year long	Anthropogenic
Proctor Maple	Hydrocarbon	0.36 ± 0.01	52.0 ± 2.0	Traffic emissions, biomass burning	EC, K, and Br	Summer	Mixed
	Processed 1	0.17 ± 0.01	24.7 ± 2.1	Fossil fuel combustion (aged)	S	All-year long	Anthropogenic
	Processed 2	0.12 ± 0.02	17.2 ± 2.9	Fossil fuel combustion (less aged)	S, Zn, Cu, Fe, Mn	All-year long	Anthropogenic
	Hydroxyl	0.04 ± 0.00	6.2 ± 0.2	Marine aerosol	Na	All-year long	Natural
St Marks	Hydrocarbon	0.88 ± 0.04	59.8 ± 2.4	Agricultural and biomass burning	EC, K, and Br	Winter, spring, summer	Mostly anthropogenic
	Processed 2	0.33 ± 0.05	22.1 ± 3.7	Fossil fuel combustion (less aged)	S, Zn, Fe, Mn	All-year long	Anthropogenic
	Processed 1	0.22 ± 0.02	15.3 ± 1.2	Fossil fuel combustion (aged)	S	All-year long	Anthropogenic
	Hydroxyl	0.04 ± 0.00	2.9 ± 0.1	Marine aerosol	Na	All-year long	Natural
Trapper Creek	Processed 1	0.07 ± 0.00	33.5 ± 2.4	Mixed source combustion C: i) fossil fuel burning and smelting of sulphide ores in power plants, ii) volcanic activity	S, Fe, Mn, Ca, Br, K	Spring, summer	Mostly anthropogenic
	Hydroxyl	0.06 ± 0.00	29.2 ± 1.2	Mainly marine aerosol	Na, also Si, Al, Ti	Winter	Mostly natural
	Hydrocarbon	0.06 ± 0.00	27.4 ± 1.2	Biomass burning	EC, K, and Br	Summer	Natural
	Processed 2	0.02 ± 0.00	9.3 ± 1.0	Fossil fuel combustion (less aged)	S, Fe, Mn	Spring	Anthropogenic

Figures

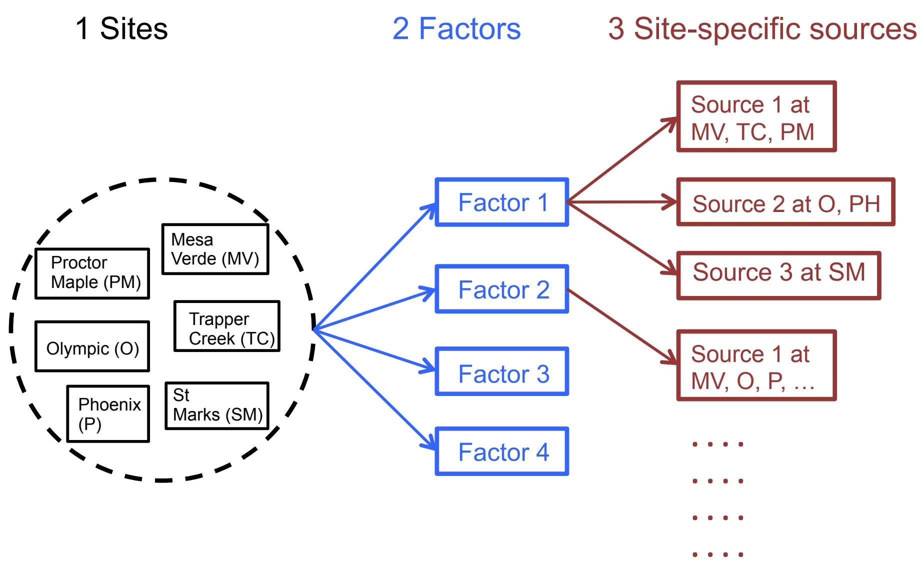

Figure 1: An interpretation of multi-site organic aerosol source apportionment results using FT-IR spectra

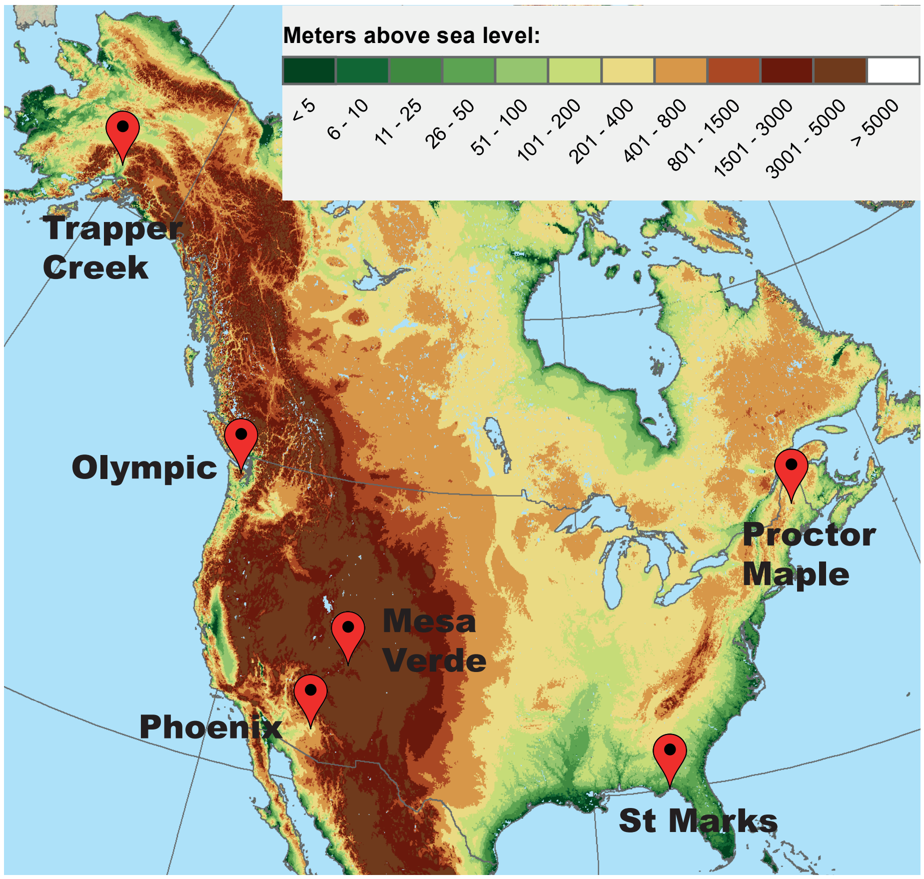

Figure 2: Location of IMPROVE sites. Map was obtained from Socioeconomic Data and Applications Center (NASA, 2010)

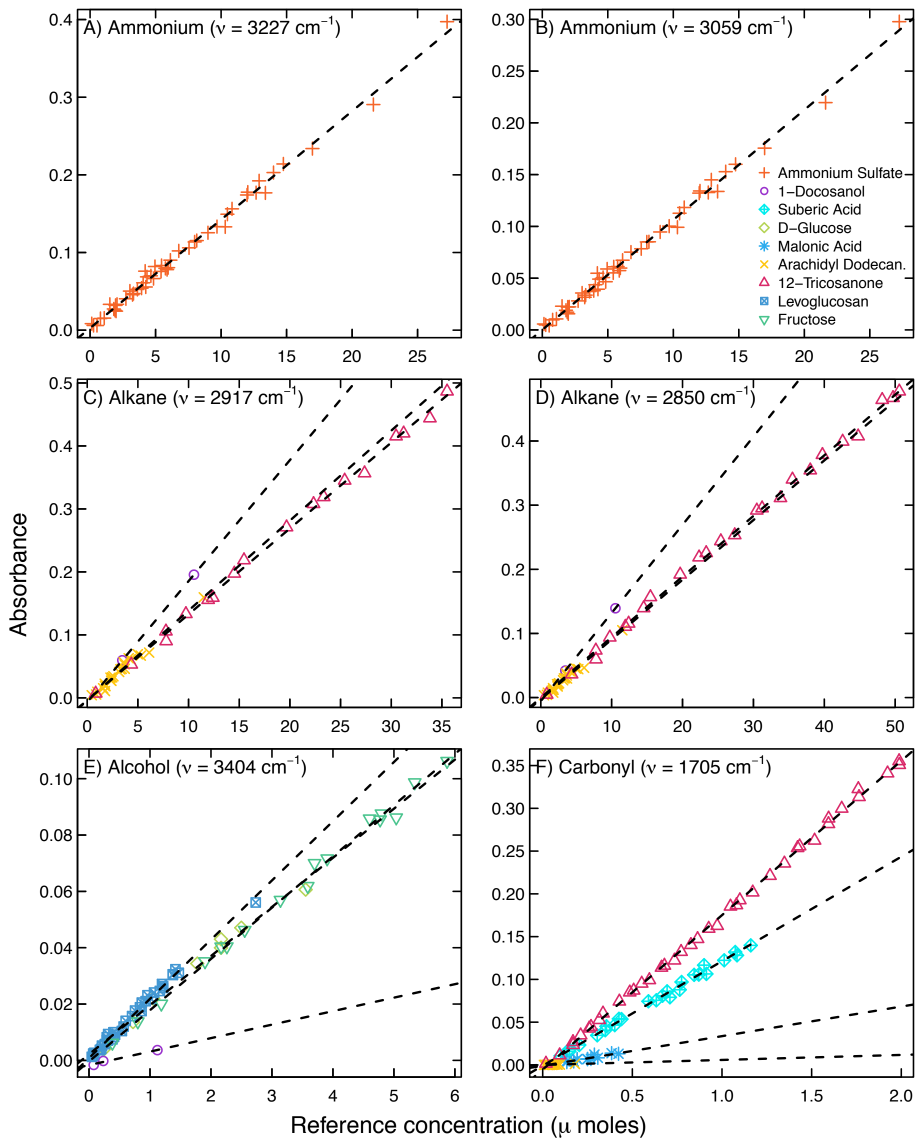

Figure 3: Fitted linear models to correlate reference concentration from different functional groups (A-F) from laboratory standards to measured absorbance. Colors and shapes in datapoints denote specific compound types.

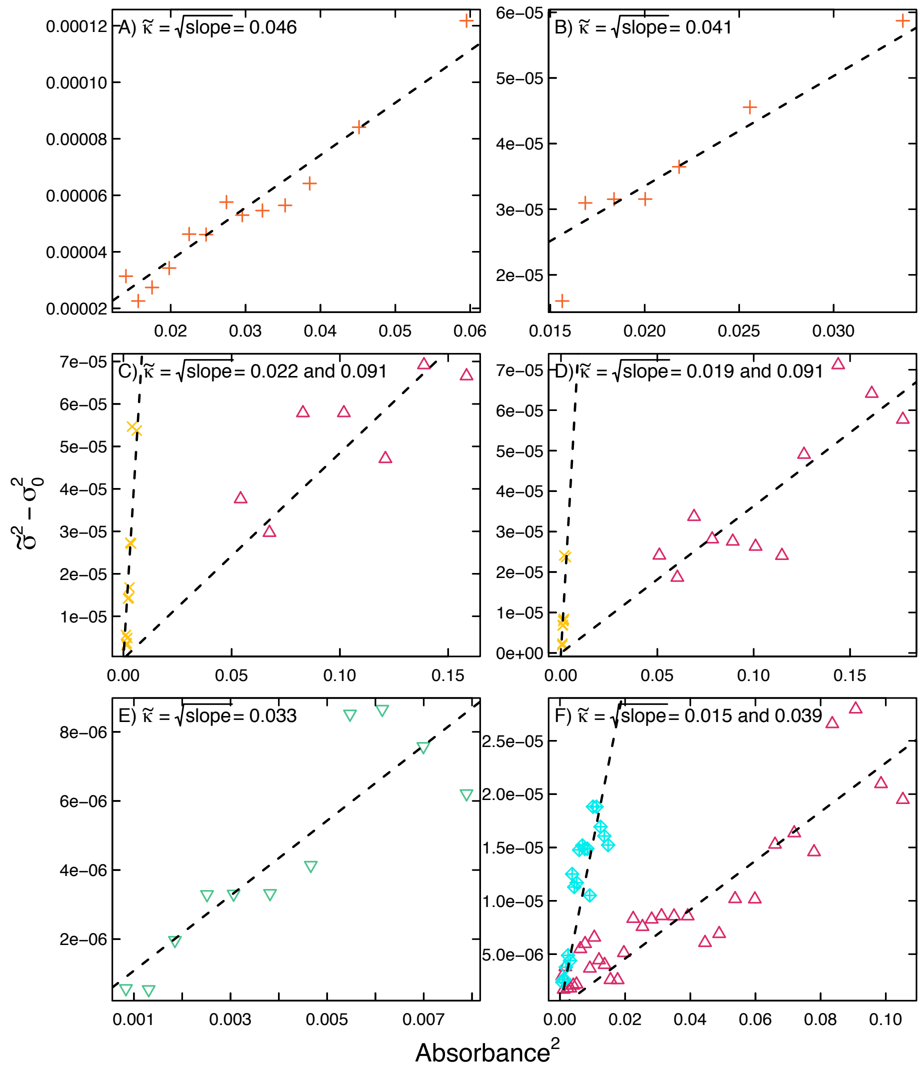

Figure 4: Fitted linear regression lines to relate x^2 ($Absorbance^2$) and $\tilde{\sigma}^2 - \sigma_0^2$. The legend for functional group and compound type assignment remains the same as in Figure 3.

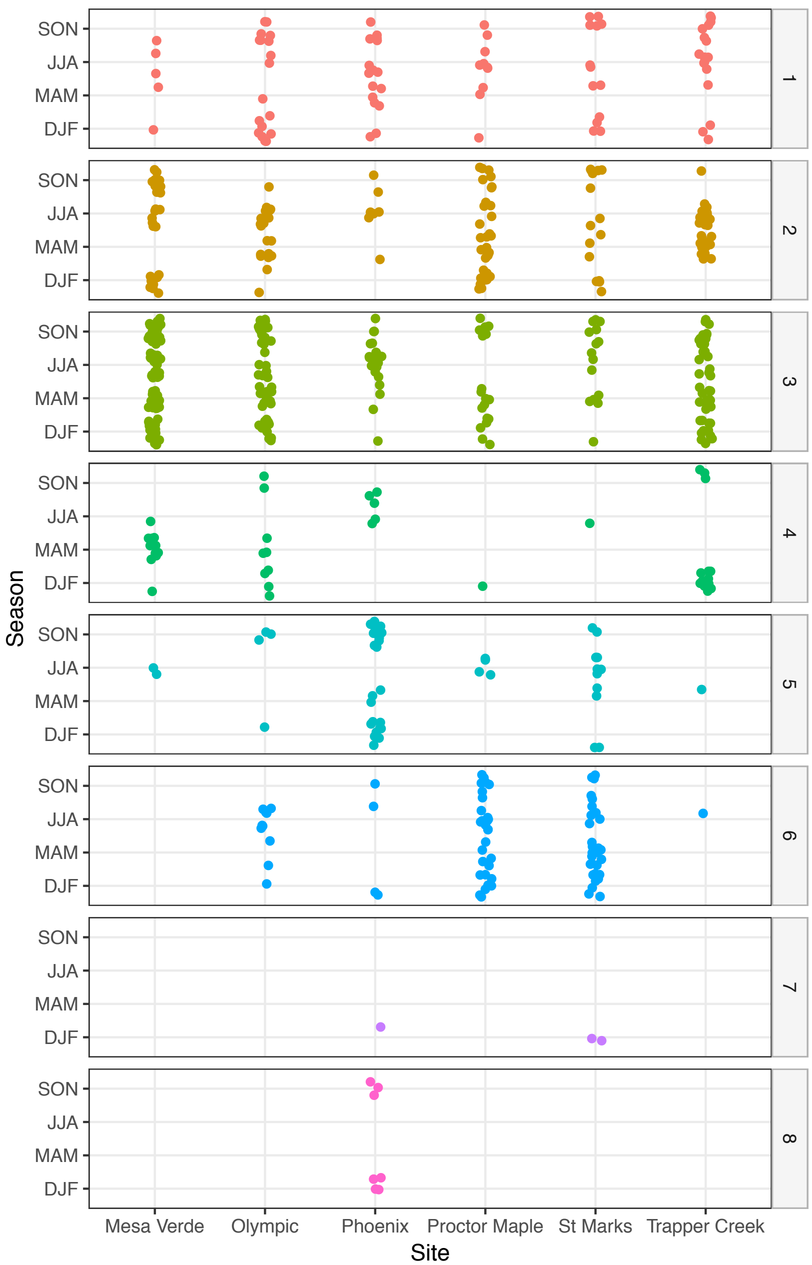

Figure 5: Cluster membership of 616 samples collected from 6 sites across four seasons. Cluster numbers are denoted by grey vertical panels and differentiated by color. To prevent overplotting, we added a small amount of random noise to the data (jitter).

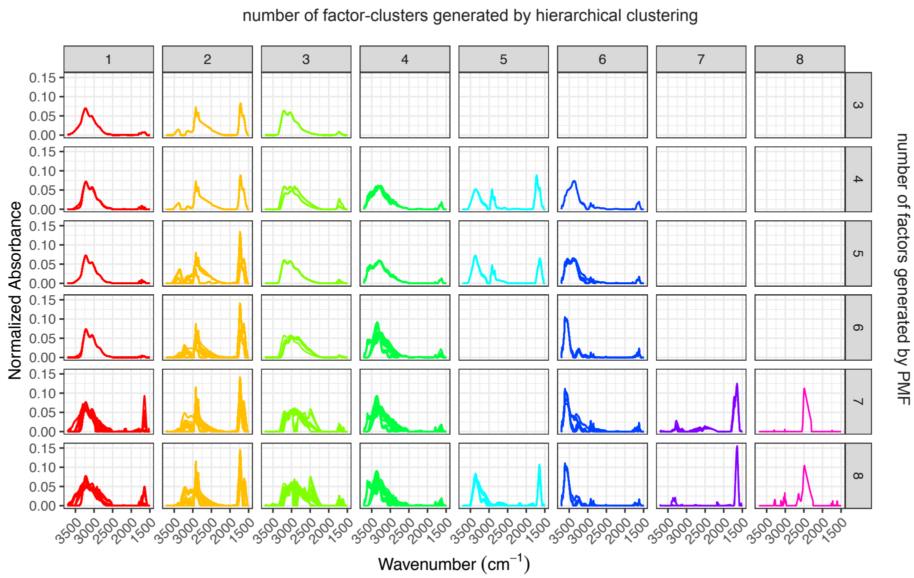

Figure 6: Factors from 162 solutions generated by varying seed, rotational parameter, and number of factors grouped into one of 8 clusters. Gray, horizontal panels along the top denote the number of factor-clusters generated by hierarchical clustering. Gray, vertical panels on the right denote the number of factors used in our PMF analyses.

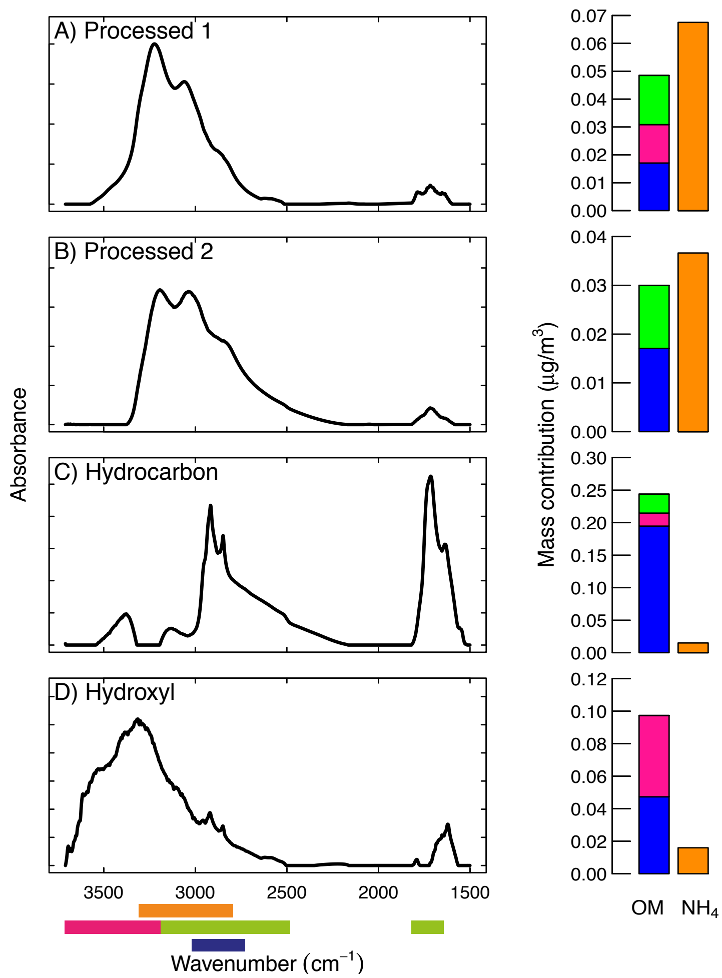

Figure 7: Left: Chemical profiles of factors derived from IMPROVE FT-IR measurements. Colored rectangles (bottom) denote the extent of wavenumber regions where specific functional groups absorb (orange, blue, green, and pink correspond to ammonium, alkane, carboxylic acid, and alcohol, respectively). Right: Bar charts show factor compositions in terms of organic mass (OM) content and inorganic ammounium (colors mapping to functional groups as specified above).

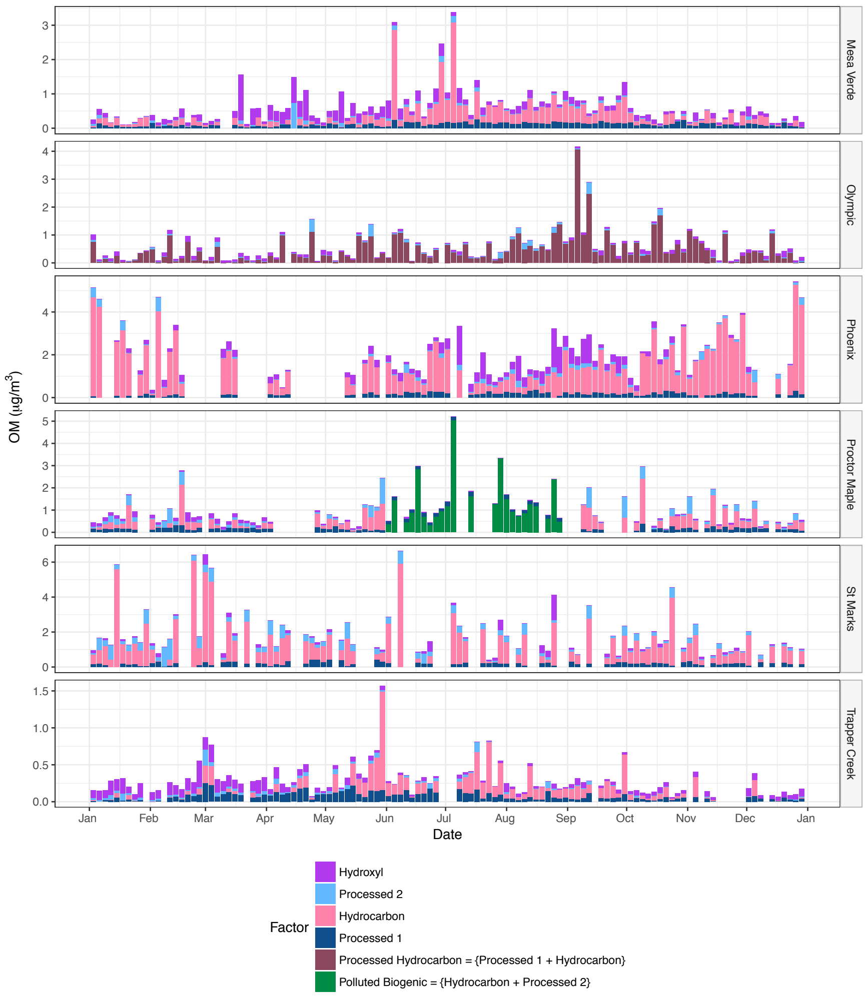

Figure 8: PMF factors contributing to organic mass during 2011.

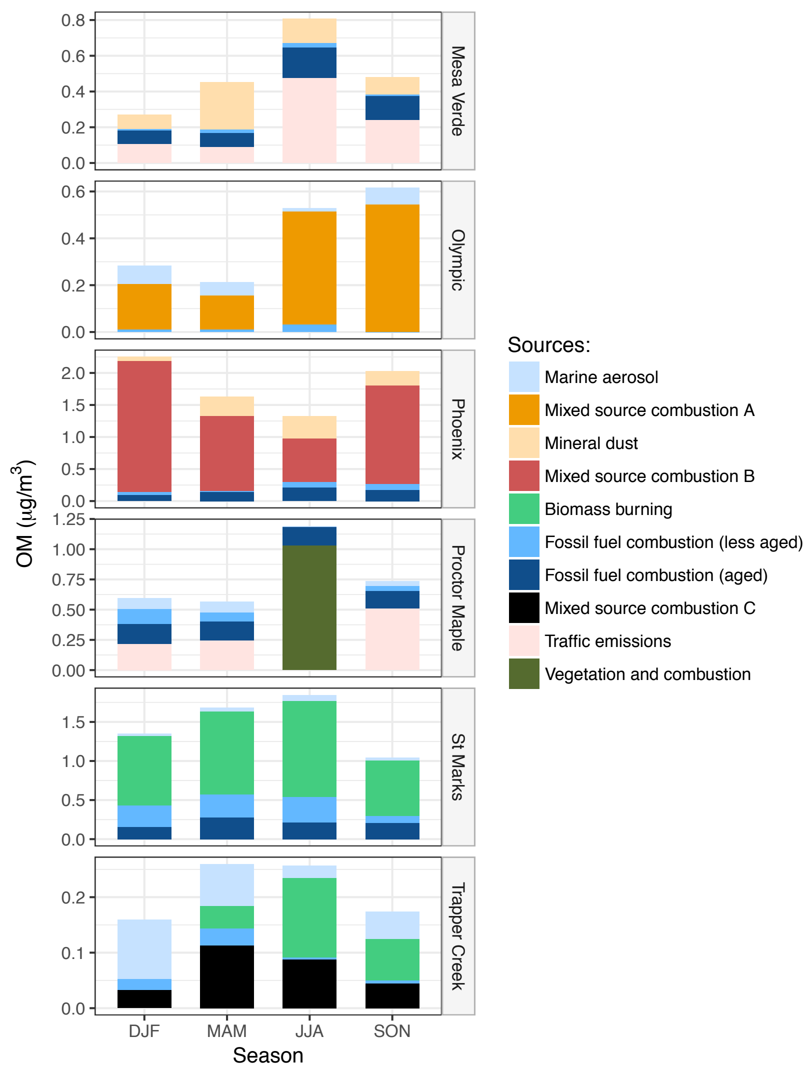

Figure 9: Seasonal averages of site-specific sources contributing to OM.

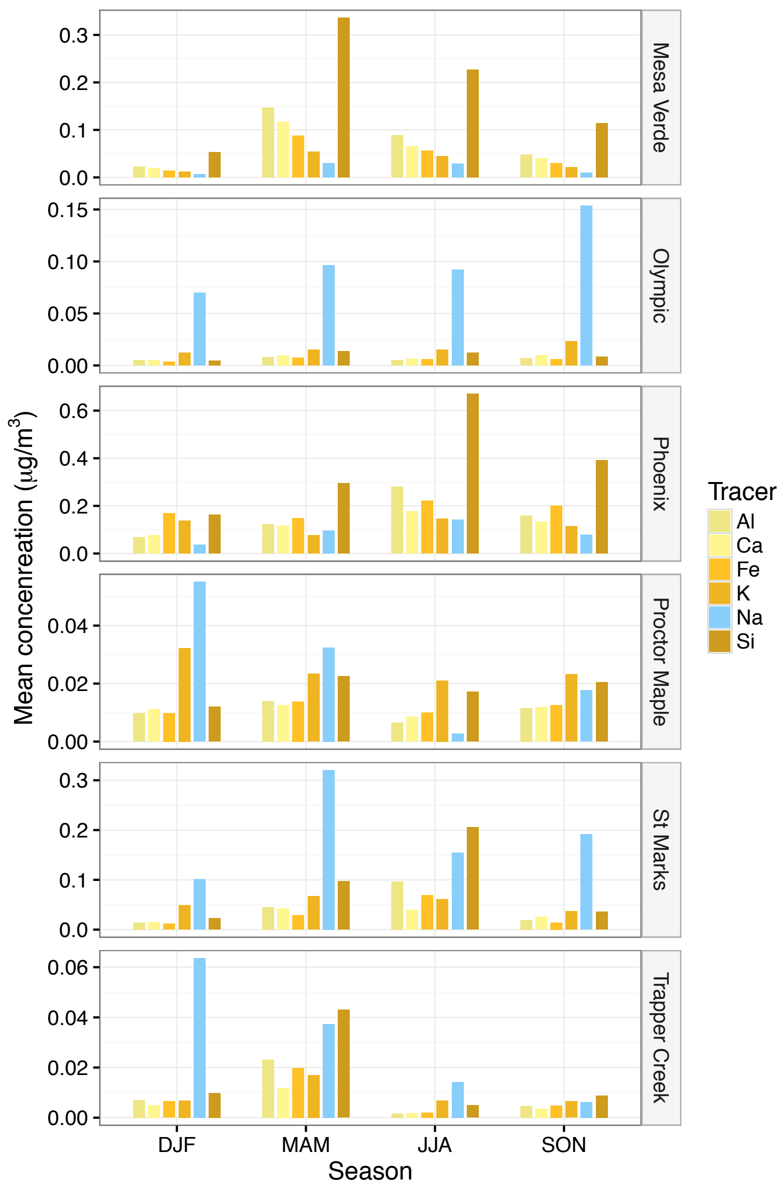

Figure 10: Average seasonal selected ion concentrations. Al, Ca, Fe, K, and Si are main mineral dust tracers whereas Na represents the main marine tracer.

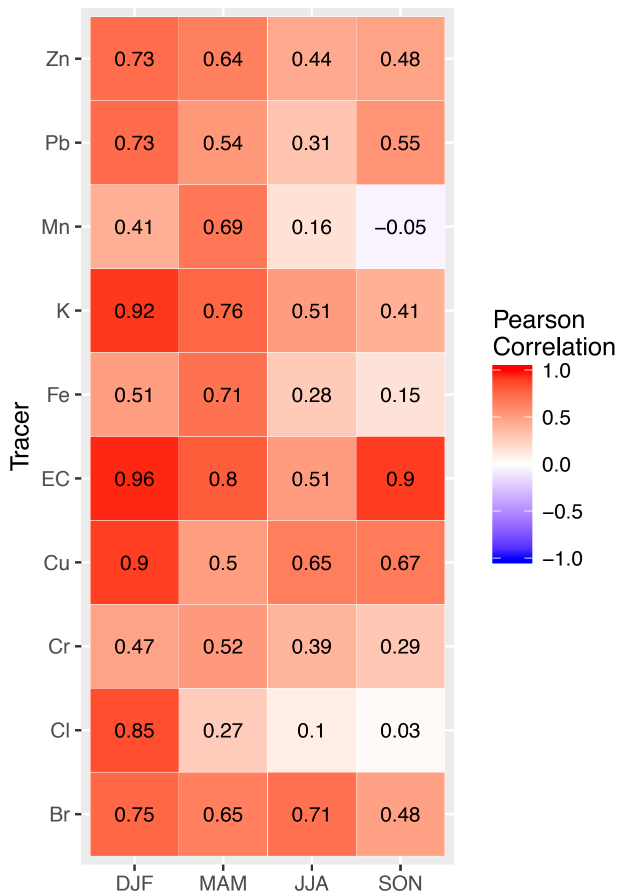

Figure 11: Pearson correlation coefficients (r) between selected ion seasonal concentrations measured at Phoenix site and strength of Hydrocarbon factor. The color spectrum denotes the magnitude of the correlation coefficient.

Supplemental Information

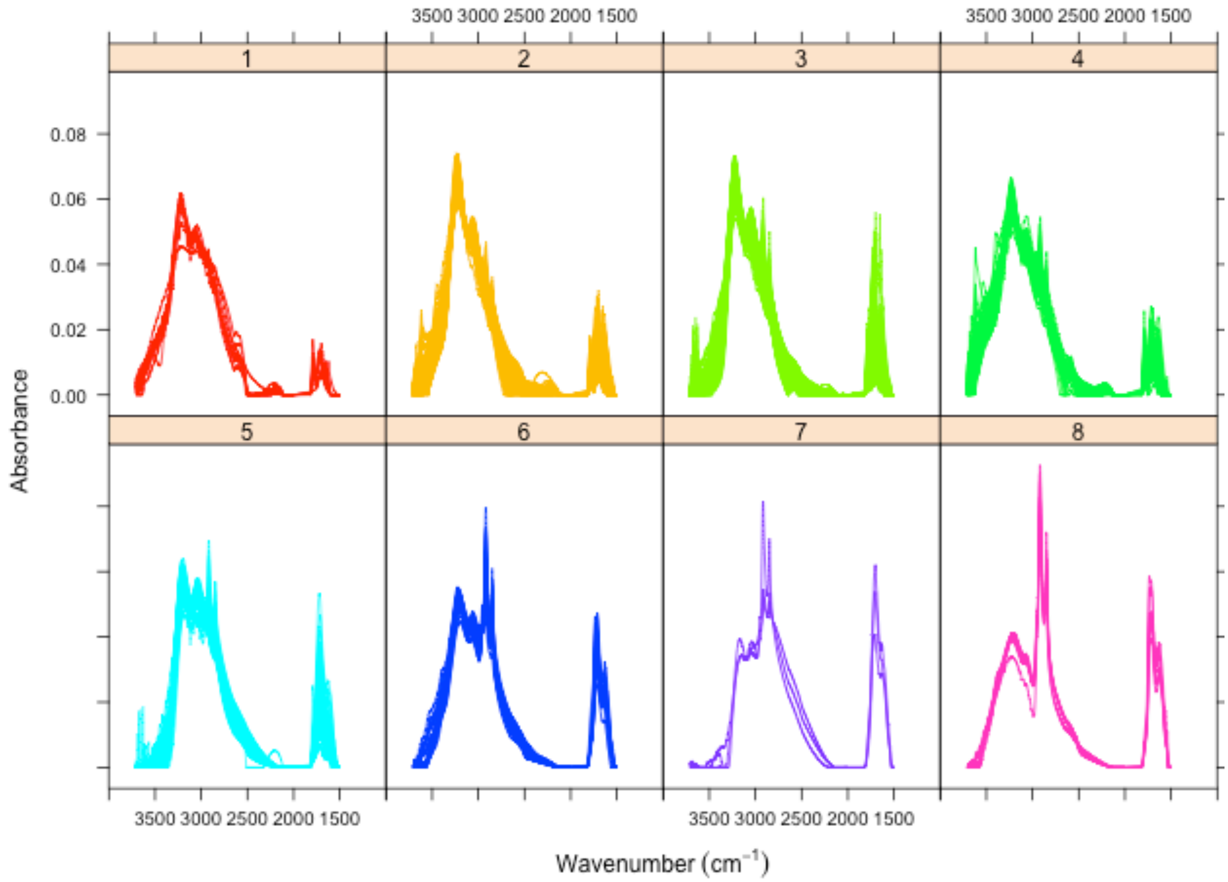

Figure S1: Spectral profiles from all 616 samples assigned to 8 clusters.

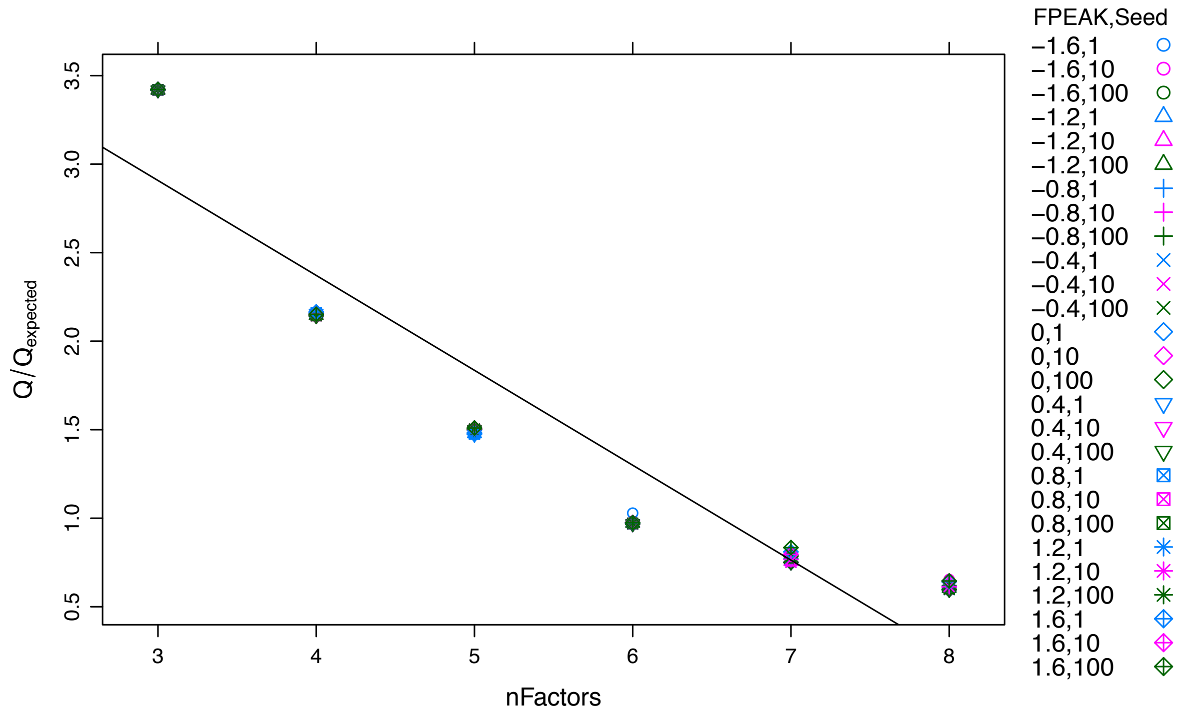

Figure S2: Relationship between Q/Q_{exp} and the number of PMF factors based on varying FPEAK and seed values

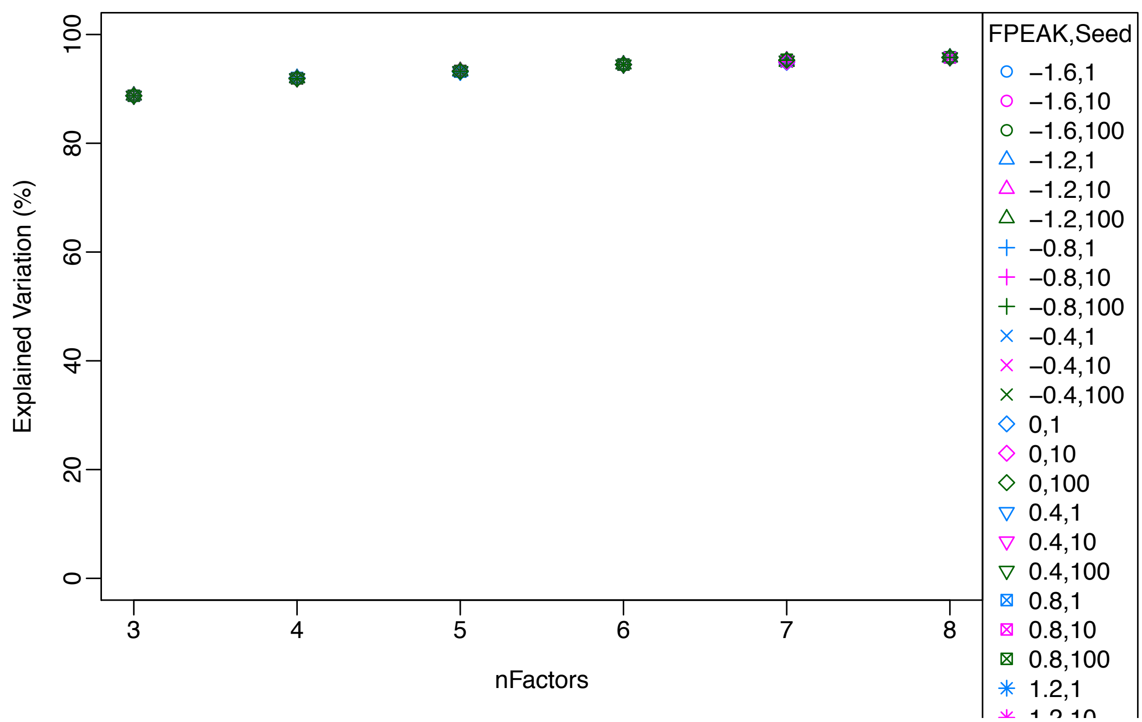

Figure S3: Explained variation as a function of the number of factors, FPEAK, and seed parameters. FPEAK varies from -1.6 to 1.6 and seed assumes values of 1, 10, and 100.

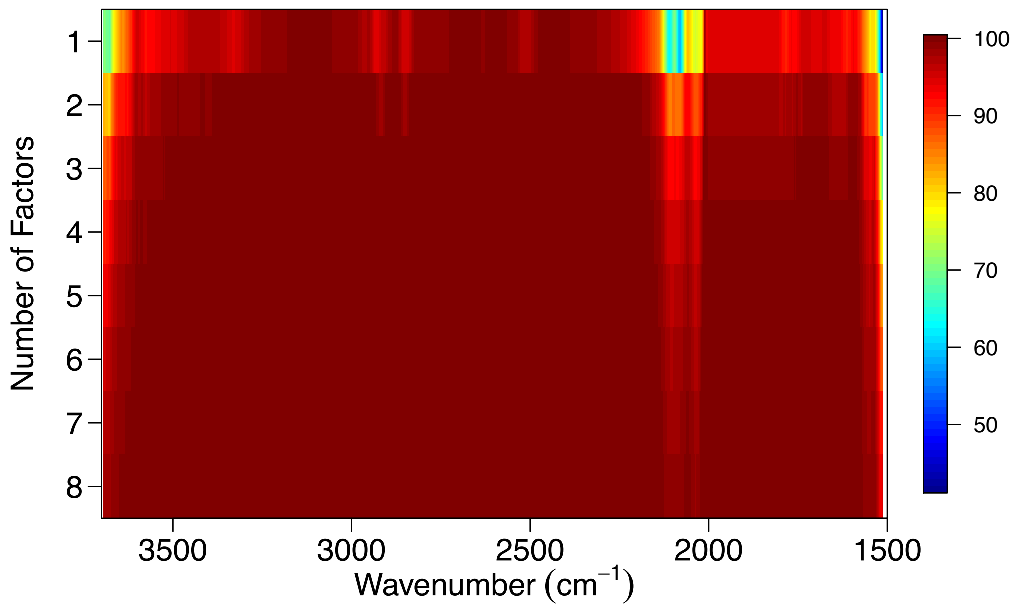

Figure S4: Percentage recovery of FT-IR samples across wavenumbers.

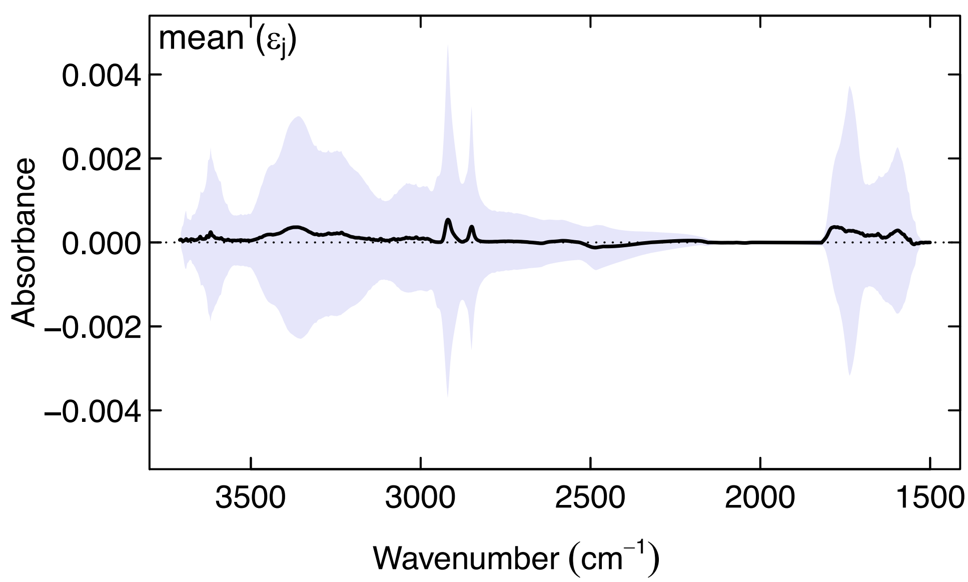

Figure S5: Black line represents mean ε_j and shaded areas denote one standard deviation.

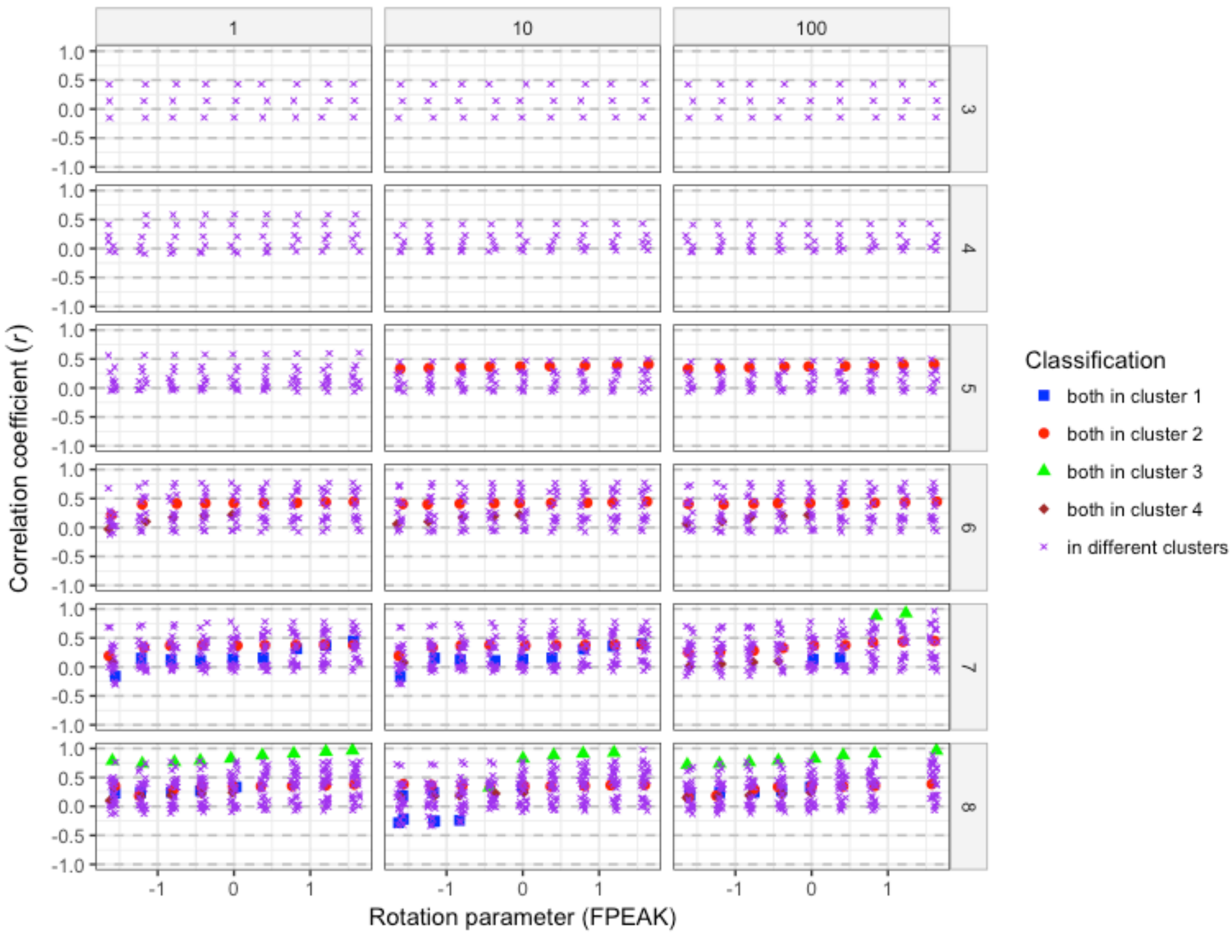

Figure S6: Pair-wise correlations of g-scores for each of the 126 solution. The gray, horizontal panels indicate the different seed values (1, 10, and 100) and the gray, vertical panels along the right represents groups with different number of factors (ranging from 3 to 8). The classification of symbols corresponds to the pairwise membership in factor-clusters (Figure 6 of main text).

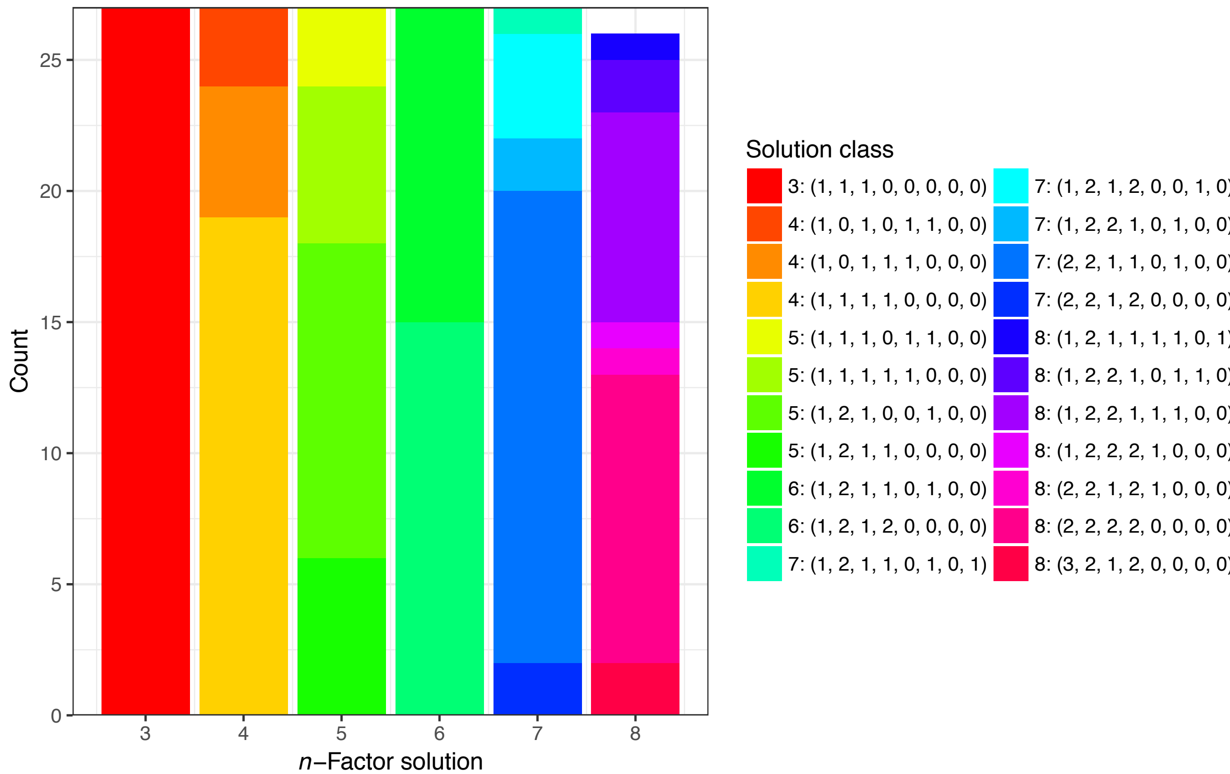

Figure S7. Classes of solutions obtained by grouping factor-clusters shown in Figure 6 of the main text. The solution class label (shown in the legend on the right) is an index of eight numbers indicating the number of factor profiles belonging to each of the eight factor clusters. For instance, (1, 0, 1, 1, 1, 0, 0, 0) is a solution for which its factor profiles belong to factor-clusters 1, 3, 4, and 5.

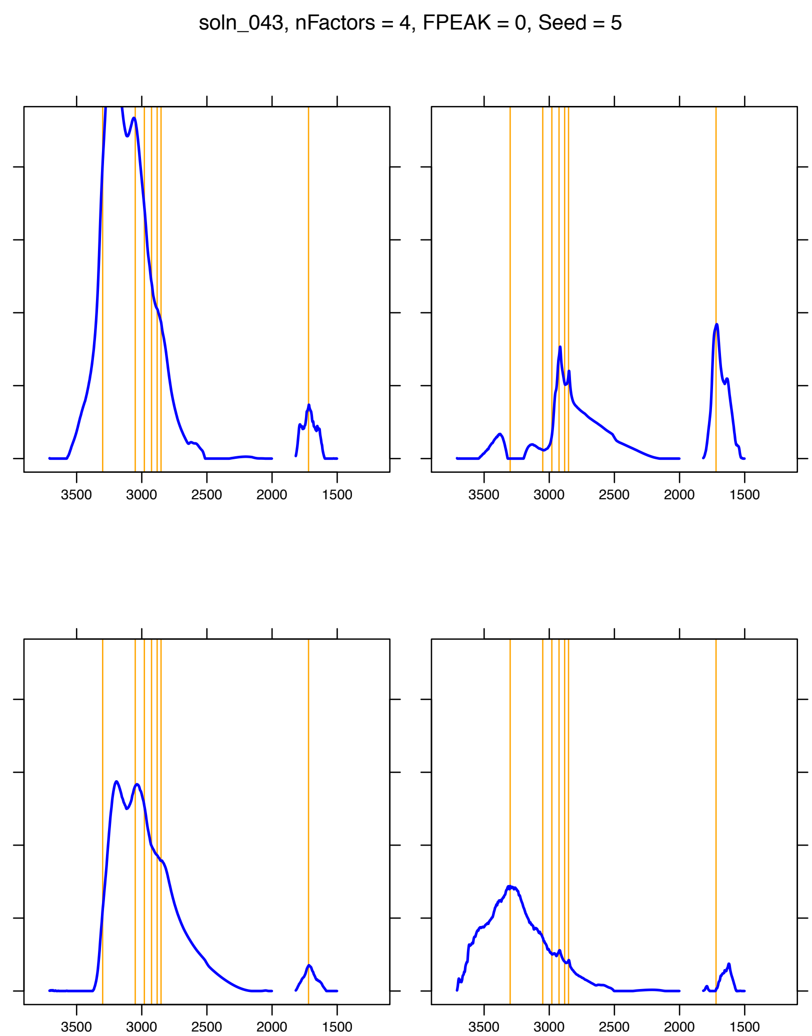

Figure S8: PMF solution class generated when number of factors = {4}, seed = {3, 5, 10, 15, 20, 30, 75, 100}, FPEAK = {-1.6, -1.2, -0.8, -0.4, 0, 0.4, 0.8, 1.2, 1.6}, and also when seed = {1} with FPEAK = {-1.6}.

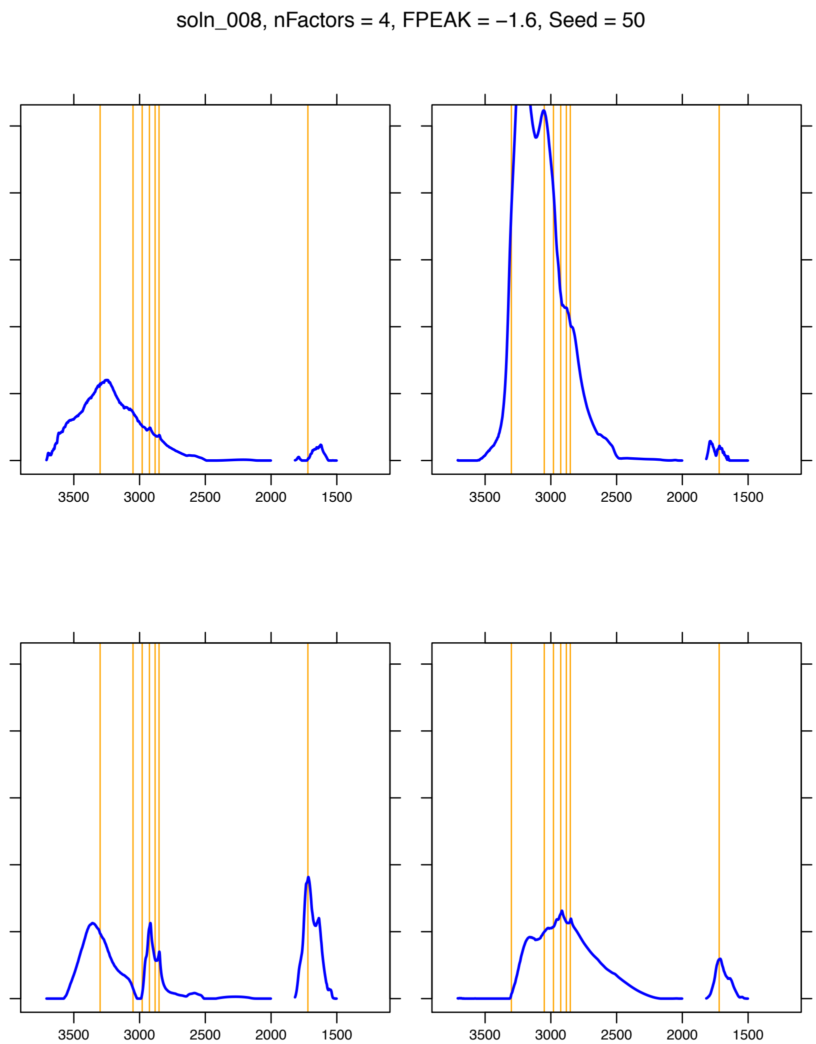

Figure S9a: PMF solution class generated when number of factors = {4}, seed = {1, 50}, FPEAK = {-1.2, -0.8, -0.4, 0, 0.4}, and also when seed {50} with FPEAK = {-1.6}.

soln_081, nFactors = 4, FPEAK = 1.6, Seed = 1

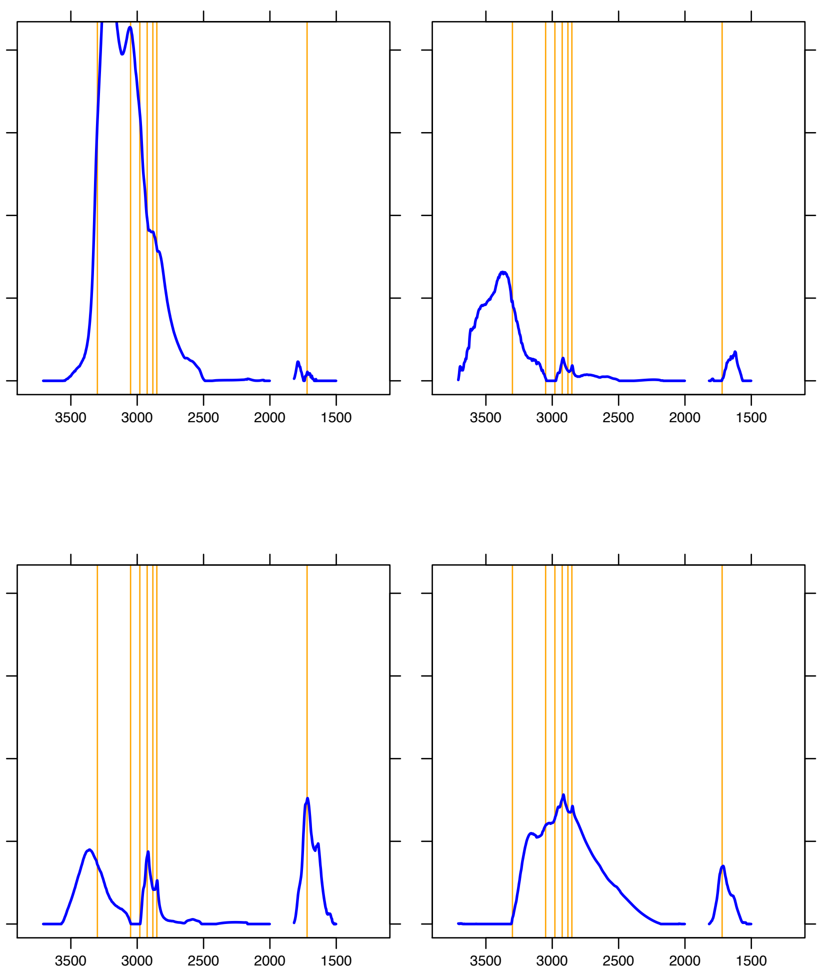

Figure S9b: PMF solution class generated when number of factors = {4}, seed = {1, 50}, FPEAK = {0.8, 1.2, 1.6}.

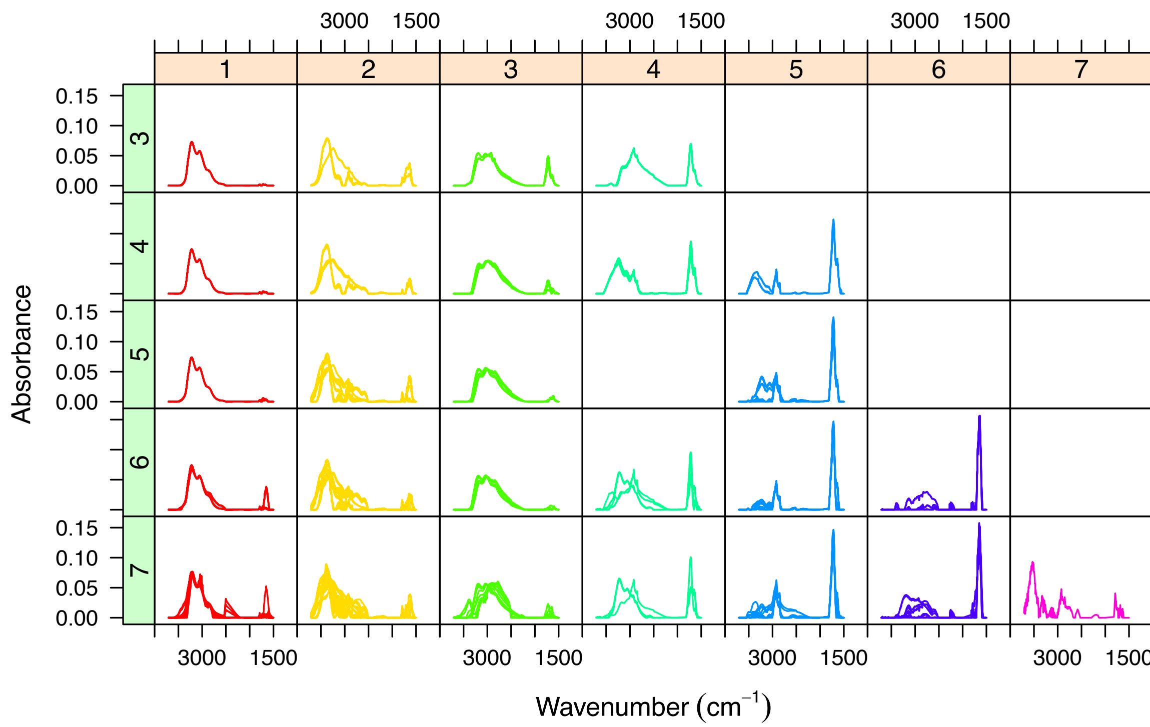

Figure S10: Factors from 105 solutions obtained from Olympic site measurements. Solutions were generated by varying seed {1, 10, 100}, rotational parameter {-0.8, -0.4, 0, 0.4, 0.8, 1.2, 1.6}, and number of factors {3, 4, 5, 6, 7} grouped into one of 7 clusters. Orange, horizontal panels along the top denote the number of factor-clusters generated by hierarchical clustering. Green, vertical panels on the left denote the number of factors used in our PMF analyses. The 3-factor solutions contains profiles in clusters 1 and 2 all the time, while profile in cluster 3 appears roughly 70% of the time and profile in cluster 4 appears the remaining 30% of the time. The 4-factor solution looks very similar to the solution presented in the main test except for Hydrocarbon profile, which occurs as the solution in either cluster 4 or 5, but not both. However, solutions in both cluster 4 & 5 do contain excess carbonyl, which we could not reconstruct from our ambient FT-IR measurements. Apart from differences in Hydrocarbon factors, the rest of the factors (Hydroxyl, Processed 1, and Processed 2) are contained in our multi-site result.

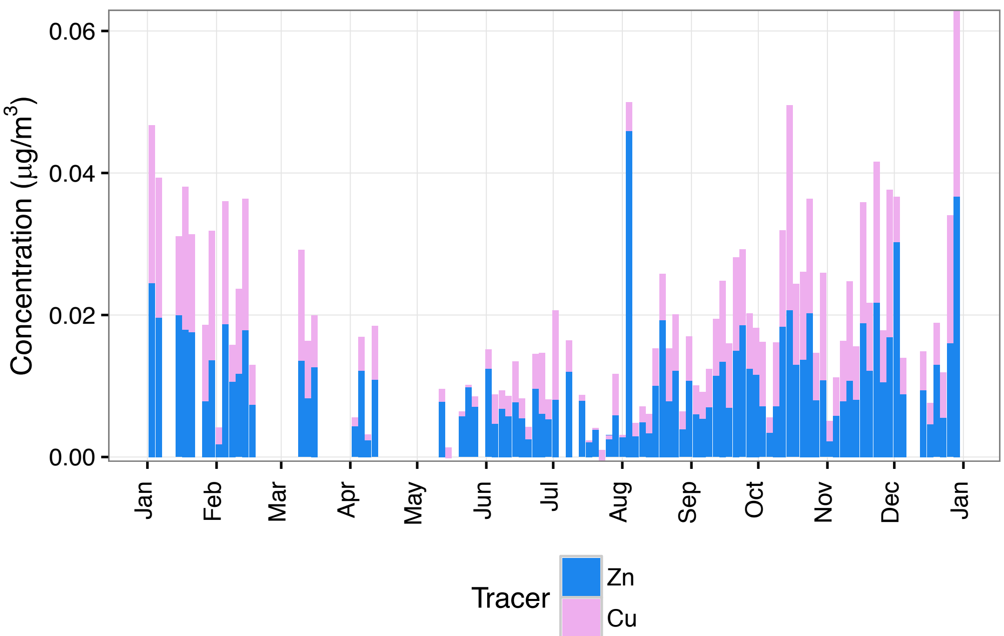

Figure S11: Daily measured concentrations of Zn and Cu at Phoenix site during 2011.

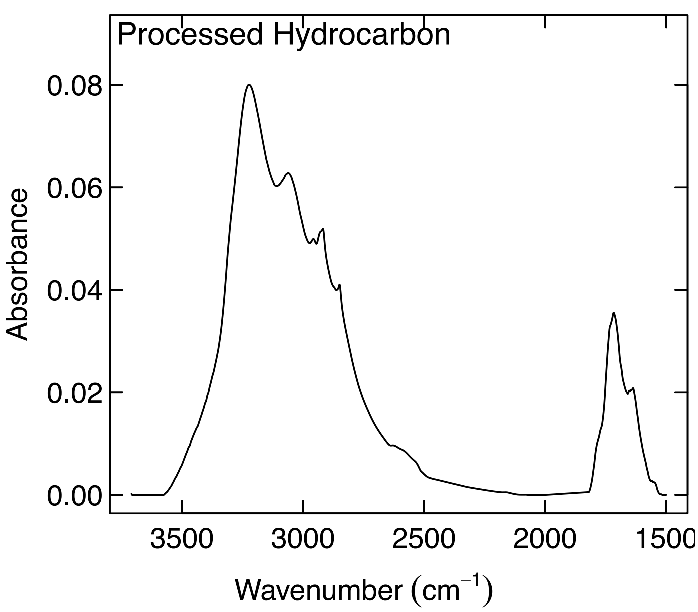

Figure S12: Processed Hydrocarbon factor profile.

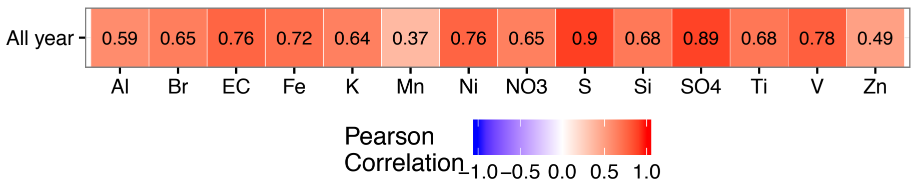

Figure S13: Pearson correlation coefficients (r) between selected ion annual concentrations measured at Olympic site and strength of Processed Hydrocarbon factor. The color spectrum denotes the magnitude of the correlation coefficient.

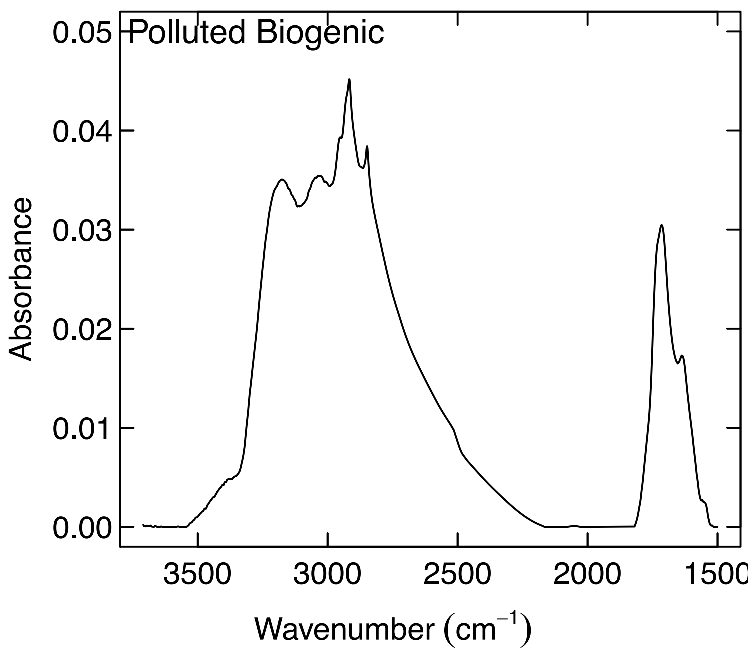

Figure S14: Polluted Biogenic factor profile.