Composition and Sources of Organic Aerosol Particles: Insights Gained from Positive Matrix Factorization (PMF).

Presented by Manjula Canagaratna
China AMS User Meeting
May 2017
Organic species are a dominant fraction of ambient PM.

**Goal:** Distinguish between different sources and processes of OA.

Slide adapted from Qi Zhang.
Signatures in AMS Organic Spectra

**Hydrocarbon-like Organic Aerosol (HOA)**

\[ C_{n}H_{2n+0,2} \rightarrow e^{-} C_{m}H_{2m\pm 1}^{+} \]

\[ 27, 29, 41, 43, 55, 57, 69, 71, ... \]

\[ \text{C}_2\text{H}_4^+, \text{C}_3\text{H}_7^+, \text{C}_4\text{H}_9^+ \]

**Oxidized Organic Aerosol (OOA)**

\[ C_{n}H_{m}O_{y} \rightarrow e^{-} \text{H}_2\text{O}^+, \text{CO}^+, \text{CO}_2^+, \text{C}_2\text{H}_3\text{O}^+ \]

\[ \Delta, \sim 600^\circ C \]

\[ 18, 28, 44, 43, 29, 55, ... \]

Alfarra et al.

*Different classes of organic molecules yield different ions (m/zs 57, 44)*
m/z 57 and 44 reflect primary and secondary sources
- Time Trends
- Size

FIGURE 1. Time trends of (a) m/z 57 and typical combustion emission tracers (CO, NOx, EC) and (b) m/z 44 and PM1 SO4^2− (both from the AMS) and (c) the average size distributions of m/z 44, m/z 57 and SO4^2− during the AMS deployment at the Pittsburgh Supersite.
Ambient measurements (urban vs. rural) have very different MS. Many ions show similar behaviour as 44 and 57.

AMS Approach
Analyze variability in ENTIRE AMS spectrum not just a few ions!!!!
A data matrix is decomposed into an **arbitrary number** of factors, each of which is represented by a **constant mass spectrum** and a **contribution time series**. There is usually some **residual** of fit.
Difference between PMF and Multilinear Engine (ME-2):

- PMF does not make any apriori assumptions about factor mass spectra or time series
- ME-2 constrains factors using apriori Information (More on this in SOFI session later today and tomorrow.)

A data matrix is decomposed into an arbitrary number of factors, each of which is represented by a constant mass spectrum and a contribution time series. There is usually some residual of fit.

Slide adapted from I. Ulbrich
The PMF Approach: Weighted Least Squares

Quantity that is minimized is $Q$, which is the sum over all elements of the weighted residual matrix.

$$Q = \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \frac{\text{resid}_{ij}}{\sigma_{ij}} \right)^2$$

Tool for decreasing, or discarding weights of ions in fit. Properly noise calculation important step in analysis! (AMS and CIMS use Poisson Counting Error).

- If all points are fit as well as we know them, the $Q$-contribution from each point is $\sim 1$

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

Ideally $Q$ is expected to be approx. $=mxn$

and the total $Q$ is $\sim$ the size of the matrix.

Monitor $Q/Q\text{exp}$ as to evaluate quality of fit.

Slide adapted from I. Ulbrich
The PMF Approach: Weighted Least Squares

Let’s normalize the total $Q$ to $Q_{\text{exp}}$, then if all points are fit to $\sim \sigma$, $Q/Q_{\text{exp}} \sim 1$

$$
Q/Q_{\text{exp}} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \frac{\sim \text{resid}_{ij}}{\sigma_{ij}} \right)^2 / m \times n
$$

Adding factors decreases $Q$, but may not give meaningful factors!

No a-priori information used BUT user interpretation needed in choice of factor

Slide adapted from I. Ulbrich
Applications: Field AMS Data

- AMS Data
- Lanz et al. 2007, Ulbrich et al., 2009

**Example:** PMF of Pittsburg, PA Q-AMS data

Ulbrich et al., 2009

- HOA MS dominated by hydrocarbon ions and TS correlates w/ vehicle emission tracers (POA)

- OOA-1 MS has high m/z 44 and TS correlates with SO$_4$ (SOA, aged regional)
- OOA-2 MS has high m/z 44 and m/z 43 and TS correlates with NO$_3$ (SOA, fresh, more volatile)
- observed first in Switzerland dataset by Lanz et al. 2007

- Process of choosing “best” solution can be ambiguous (i.e. many mathematically equivalent solutions). We want solution that identifies interpretable factors
- Use correlations with tracers from other instruments to help make decision.
Variability in time trends necessary for separation of more factors.

HR spectra can provide better separation of factors with small mass contributions.

NOTE:
Sources/processes with near identical time-trends and/or spectra cannot be easily resolved with PMF.

ME-2 to separate based on small variations.
ACSM shows that 30% of OA during summer in Atlanta is due to IEPOX-OA.
Common PMF components that have been observed in AMS/ACSM field studies

<table>
<thead>
<tr>
<th></th>
<th>Signature</th>
<th>Tracers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOA (Hydrocarbon-like OA)</td>
<td>57</td>
<td>CO, Nox</td>
</tr>
<tr>
<td>BBOA (Biomass-burning OA)</td>
<td>60,73</td>
<td>Acetonitrile, Levoglucosan, K</td>
</tr>
<tr>
<td>COA (Cooking OA)</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>CCOA (Coal combustion OA)</td>
<td>73, PAH</td>
<td></td>
</tr>
<tr>
<td>NOA (nitrogen OA)</td>
<td>N-peaks</td>
<td></td>
</tr>
<tr>
<td><strong>Secondary</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OOA (Oxidized OA)</td>
<td>44</td>
<td>O3</td>
</tr>
<tr>
<td>SV-OOA (Semi-volatile OOA)</td>
<td>44&lt;=43</td>
<td>NO3</td>
</tr>
<tr>
<td>LV-OOA (low volatility OOA)</td>
<td>44&gt;43</td>
<td>SO4</td>
</tr>
<tr>
<td>IEPOX-OA (Isoprene epoxide OA)</td>
<td>82</td>
<td>LC-MS Markers for isoprene oxidation</td>
</tr>
<tr>
<td>m/z 91 factor (biogenic factor)</td>
<td>91</td>
<td></td>
</tr>
</tbody>
</table>
Organic Aerosol (OA) separated into primary sources (HOA, BBOA, COA, Coal Burning) and secondary sources (SV-OOA, LV-OOA)
Resources: Getting started with PMF


• The website above, provides an executable that performs the PMF analysis. The key that allows the executable to be run, however, must be obtained directly from Penti Paatero (P. Paatero, Univ. Helsinki).

• Also provides free Igor-based program (PET, Ulbrich et al.2009) to view, diagnose, and manipulate
PET (Ulbrich et al. ACP 2009), Free software for analyzing data with PMF (works for AMS and CIMS).
Resources: Review article

Understanding atmospheric organic aerosols via factor analysis of aerosol mass spectrometry: a review

Qi Zhang · Jose L. Jimenez · Manjula R. Canagaratna · Ingrid M. Ulbrich · Nga L. Ng · Douglas R. Worsnop · Yele Sun

Anal Bioanal Chem (2011) 401:3045–3067

Examples of applications
Step-by-step guide for choosing solutions, Reporting results
Steps for preparing and analyzing AMS data

With PMF

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Calculate data and error matrices</td>
<td>[73]</td>
</tr>
<tr>
<td>2.</td>
<td>Further data and error treatment</td>
<td></td>
</tr>
<tr>
<td>2a.</td>
<td>Apply minimum error</td>
<td>[38]</td>
</tr>
<tr>
<td>2b.</td>
<td>Remove anomalous spikes, if desired</td>
<td>X</td>
</tr>
<tr>
<td>2c.</td>
<td>Smooth data, if desired</td>
<td>X</td>
</tr>
<tr>
<td>2d.</td>
<td>Downweight low-SNR data</td>
<td>X</td>
</tr>
<tr>
<td>2e.</td>
<td>Downweight repeated information (m/z 44 and related m/z values)</td>
<td>X</td>
</tr>
<tr>
<td>3.</td>
<td>Run PMF for a range of number of factors (P) and random starts (SEEDs)</td>
<td>[38]</td>
</tr>
<tr>
<td>3a.</td>
<td>Examine Q/\text{Q}_{\text{exp}} vs. P in solution (Fig. 5a). A steep change in slope indicates the minimum P to consider for a good solution</td>
<td></td>
</tr>
<tr>
<td>3b.</td>
<td>Examine solutions from different random starts for each P. Sort solutions by Q/\text{Q}_{\text{exp}} values and compare the factors in each solution. Are there multiple types of solutions (representing local minima in the solution space)?</td>
<td>[52, 62]</td>
</tr>
<tr>
<td>3c.</td>
<td>If not, proceed and use seed 0</td>
<td></td>
</tr>
<tr>
<td>3d.</td>
<td>If yes, can any of the solution types be excluded because the factors are not physically meaningful? Proceed, exploring seeds that have solutions with physically meaningful factors</td>
<td></td>
</tr>
<tr>
<td>3e.</td>
<td>Try to determine the optimum number of factors by examining multiple criteria:</td>
<td>[38]</td>
</tr>
<tr>
<td>3f.</td>
<td>Look for correlations between factor time series and time series of external tracers</td>
<td></td>
</tr>
<tr>
<td>3g.</td>
<td>Look for correlations between factor time series and time series of individual source number vs. mass</td>
<td></td>
</tr>
<tr>
<td>3h.</td>
<td>Q/\text{Q}_{\text{exp}} contribution for each time step</td>
<td></td>
</tr>
<tr>
<td>3i.</td>
<td>Q/\text{Q}_{\text{exp}} contribution for each fragment ion</td>
<td></td>
</tr>
</tbody>
</table>

**Diagnostics**

- **(a) Number of Factors (F) vs. Residual**
- **(b) D for FPEAK vs. D for S**
- **(c) Q/\text{Q}_{\text{exp}} vs. P in solution**
- **(d) D for FPEAK vs. FPEAK**
- **(e) Mass Ratio**
- **(f) Residual Mass Conc. (pg m⁻²)**
- **(g) Residual Mass Conc. (pg m⁻²)**
- **(h) Q/\text{Q}_{\text{exp}} contribution for each time step**
- **(i) Q/\text{Q}_{\text{exp}} contribution for each fragment ion**

*PMF Review: Zhang et al., Anal Bioanal Chem (2011) 401:3045–3067*
Thank you for your attention!

Acknowledgements

**CU Boulder**
Ingrid Ulbrich, Jose Jimenez

**UC Davis**
Qi Zhang

**Gtech**
Sally Ng

Entire AMS User community