Analysis of Sources of Ambient Organic Aerosol Particles by Positive Matrix Factorization Applied to Aerosol Mass Spectrometric Data

Kendra Joseph, Muhlenberg College

Mentor: Dr. Yin-Nan Lee,

Brookhaven National Laboratory

Outline

- Atmospheric Organic Aerosols (OA)
- Time-of-Flight Aerosol Mass Spectrometer (ToF-AMS)
- Positive Matrix Factorization (PMF)
- Megacity Initiative: Local and Global Research Observations (MILAGRO)
- Methods
- Results
- Discussion
- Acknowledgements

Atmospheric Organic Aerosol

Sources-

- Natural
  - -Wildfire
  - -Biogenic volatile organic compounds
- Anthropogenic
  - -Fossil fuel combustion

Picture taken from (http://www.ehponline.org/docs/2007/115-1/fire.jpg).

Atmospheric Organic Aerosol
(Continued)

Effects-

- Reduce visibility
- Adverse health effects
- Climate change, by affecting Earth's radiative budget*

Picture taken of Beijing from (http://chinadigitaltimes.net/china/olympics-air-pollution/).

*CCSP, U.S. Climate Change Science Program, Synthesis and Assessment Product, *Atmospheric Aerosol Properties and Climate Impacts* (http://www.climatescience.gov), 2009.

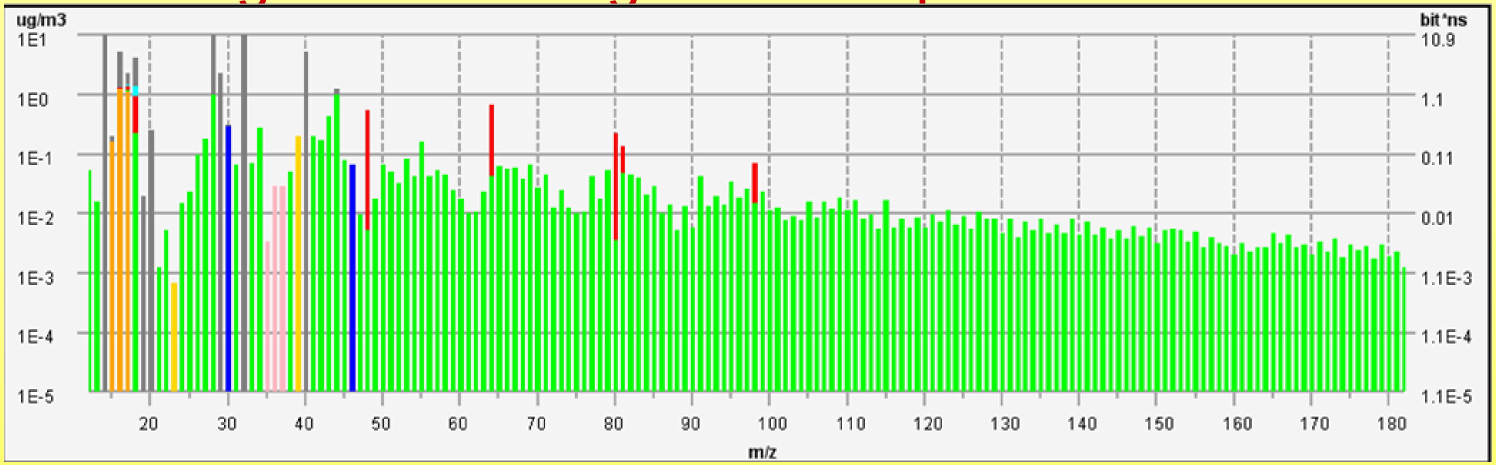
Atmospheric Organic Aerosol

(Continued)

- To assess their radiative effects and the role they play in climate change the sources of aerosols must be identified.
- By using an Aerodyne Time-of-Flight Aerosol Mass Spectrometer (ToF-AMS), composite mass spectra of OA are obtained with which different sources can be classified.

ToF-AMS

- Focuses particles into a narrow beam which travels through a chamber of a known length.
- Vaporizes the particles hitting on an oven heated at 600°C followed by electron impact ionization.
- Produces a composite mass spectra by detecting the ions using a time-of-flight mass spectrometer*.



Composite mass spectra produced on June 6, 2009. Green color represents OA

*I. M. Ulbrich, M.R. Canagaratna, Q. Zhang, D. R. Worsnop, J. L. Jimenez. “Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data”. *Atmospheric Chemistry and Physics*, vol. 9, 2891-2918, 2009.

PMF

- Individual compounds cannot be identified from mass spectrum produced by ToF-AMS.
- PMF uses factor analysis techniques to deduce the major types of sources (factors)*.

*P. Paatero, U. Tapper, "Least squares formulation of non-negative factor analysis".*Chemometrics and Intelligent Laboratory Systems,* vol. 37, pp 23-35, 1997.

PMF

(Continued)

- Techniques

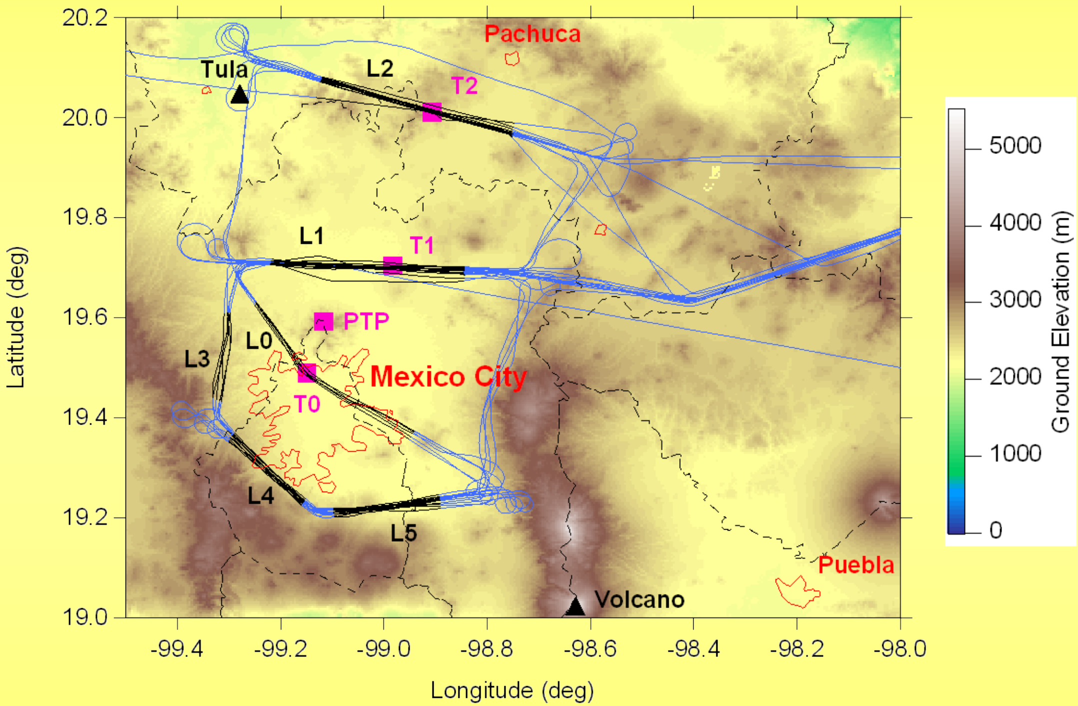
  - We used a version of PMF from the Environmental Protection Agency (http://www.epa.gov/heasd/products/pmf/pmf.htm).

- Data

  - Collected on March 18, 19, and 20 of 2006 during the MILAGRO campaign in Mexico City aboard DOE-G1.

  - Readings were taken every 12 seconds.

MILAGRO Campaign

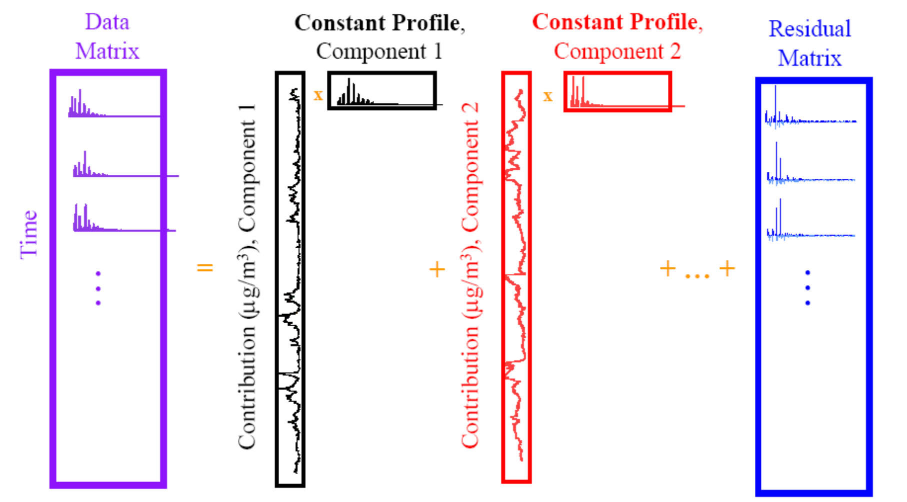


Part of G-1 sampling area, color coded by surface altitude. Map shows ground track of the 8 flights used during study, three of which were used for PMF analysis. Identified emission sources are Mexico City, Pueblo, Pachuca, the Tula industrial complex, and the Popacatapetl volcano.

Methods

PMF

- Solves equation for data matrix *X* such that *X = FG + E.*
- *F* and *G* are unknown *mxp* and pxn matrices constrained to non-negative values representing time series and mass spectrum of the factors.
- *E* represents an *mxn* matrix of residuals.
- Minimizes elements of *E* for each situation.



Schematic of the break down of data matrix, X, into factors. Provided by, I.M. Ulbirch, "Igor-PMF" , presented at the 9th AMS Users Meeting. University of Manchester, UK.

Q-Values

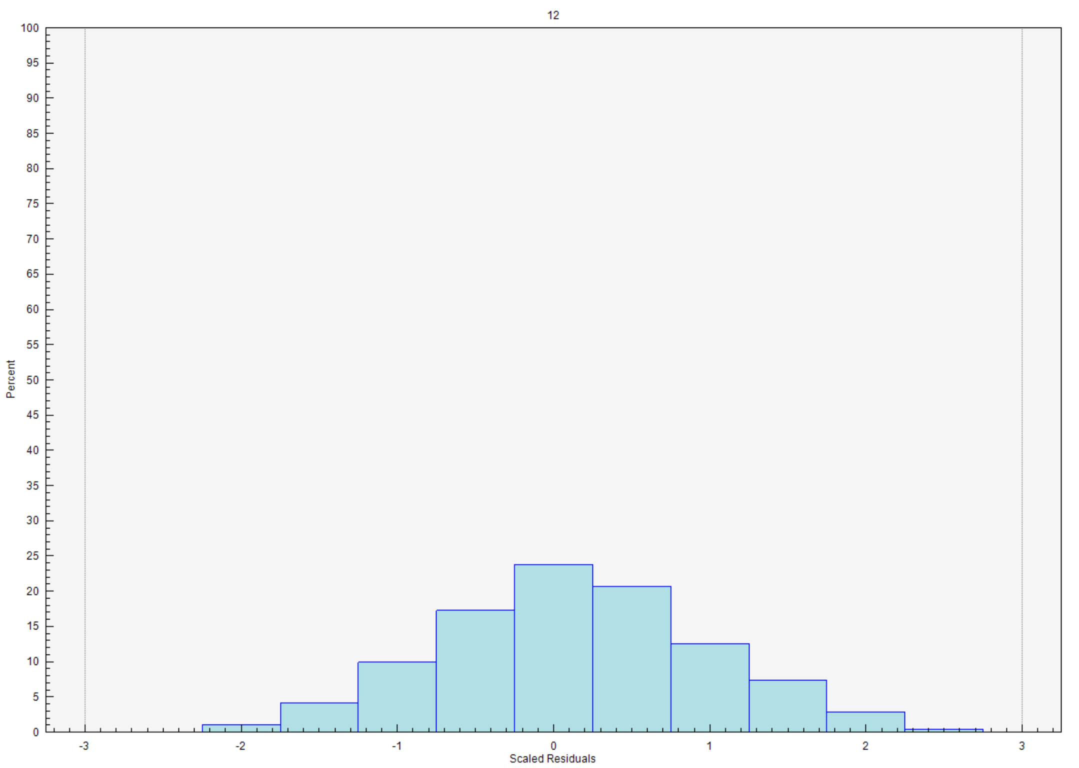
- Q_{true} and Q_{robust} are derived by a goodness-of-fit function, Q_{true} uses all data points while Q_{robust} excludes outliers.
- For a good fit of data Q_{true} should not exceed $1.5*Q_{robust}$.
- The ratio of the $Q_{robust}/Q_{expected}$ should have a value close to 1, meaning all points of the data matrix are within their expected error.
- Q values are minimized for the equation $Q = \sum_{i=1}^{m} \sum_{j=1}^{n} (e_{ij} / \sigma_{ij})^2$ where σ_{ij} is an element of an *mxn* matrix of standard deviations of each point of the data matrix.

Down-weighting Variables

- Variables with a S/N (signal to noise) ratio greater than 2 were down-weighted, their value in uncertainty matrix increased, to preserve accuracy of PMF.
- After initial analysis additional m/z values were down-weighted according to residual graphs and observed vs. predicted graphs.

Residual Graphs

- Created by PMF for each m/z to help assess in the accuracy of the solution.

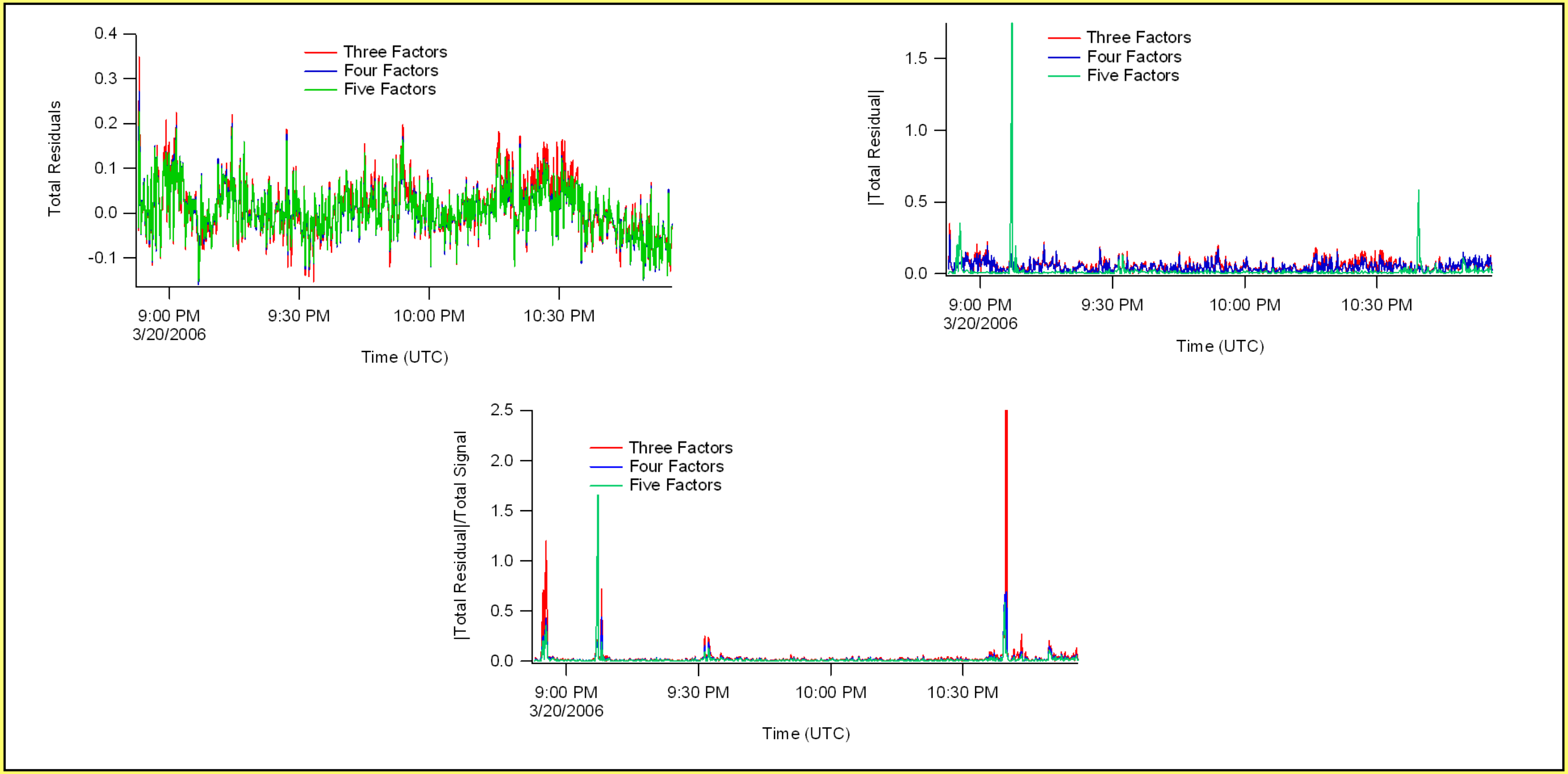


Provided by EPA PMF, m/z 12 residual graph, all values are within 3 standard deviations which indicates the solution is a good fit for m/z 12.

Residual Graphs
(Continued)

- Given the output data three residual graphs were created for each solution:
  - total residual vs. time
  - |total residual| vs. time
  - |total residual|/total signal vs. time (a weighted residual graph)
- A solution should reduce the residual matrix so that the residual graphs have values close to zero.

Residual Graphs
(Continued)



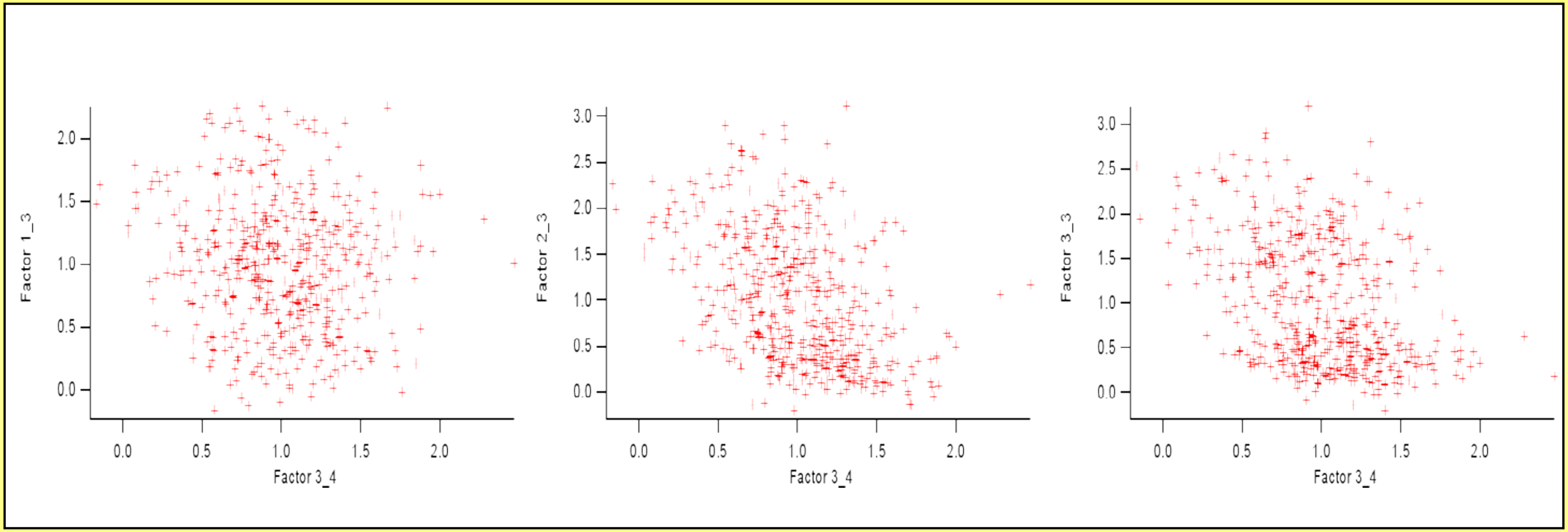
March 20th, 2006 residual graphs for MILAGRO data show evidence of stable four factor solution

Factor Comparisons

- Comparing factors of different solutions, i.e. a three factor solution and a four factor solution.
- Determines whether the additional factor is a real factor, or if it is a split of a factor within the three factor solution.

Factor Comparisons

(continued)



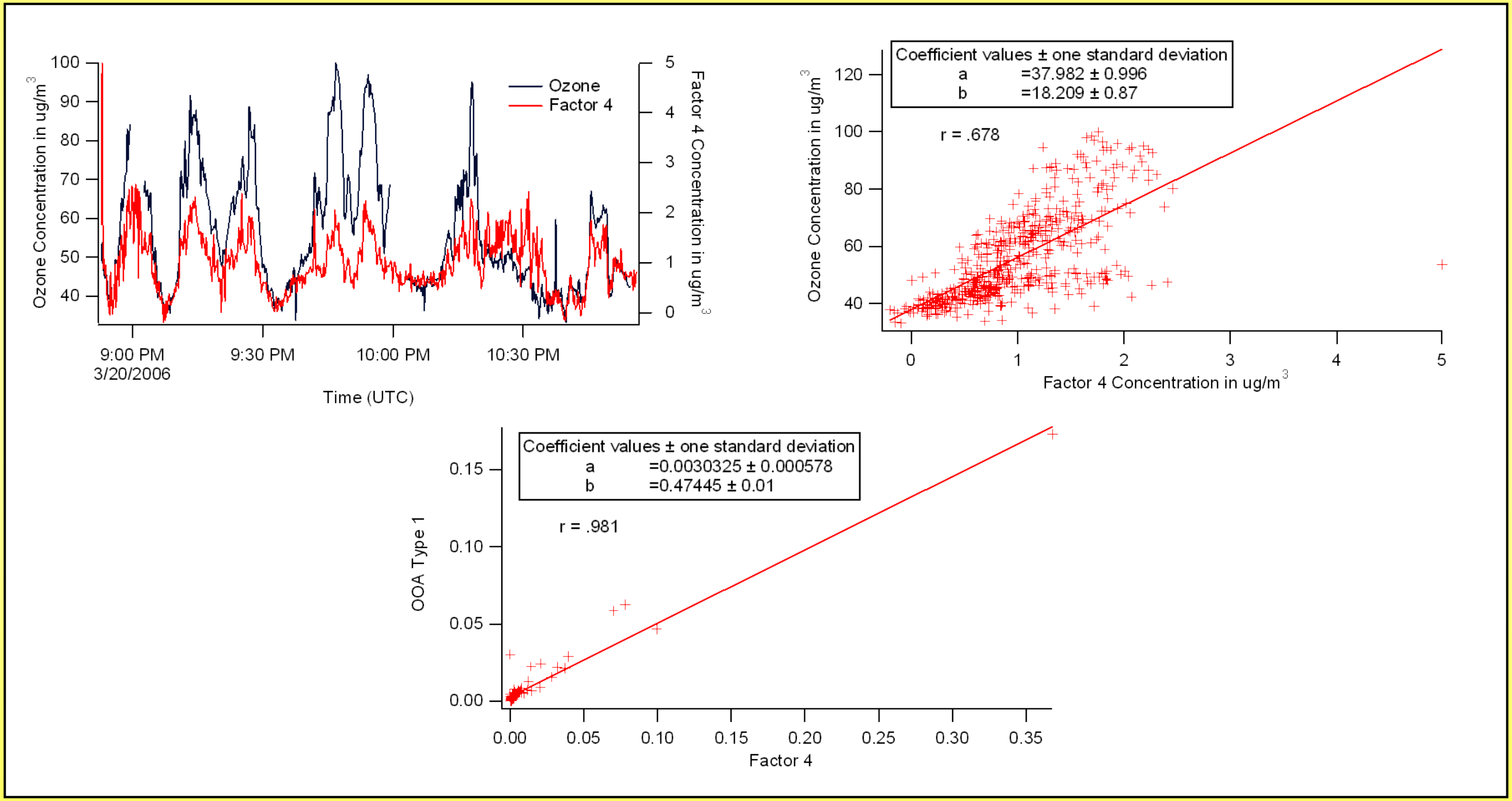
3-factor solution vs. factor three of the 4-factor solution (3_4) from MILAGRO data on 3/18/ 2006 clearly show no correlation which is a good indication that factor (3_4) is a real factor.

Comparing Factors to Tracers and Known Factors

- Indicates what factors given by PMF may represent.
- Factors were compared to OA spectra given by the AMS Spectral Database (http://cires.colorado.edu/jimenez-group/AMSsd).
- Factors were also compared to tracer compounds measured simultaneously.

Comparing Factors to Tracers and Known Factors

(Continued)

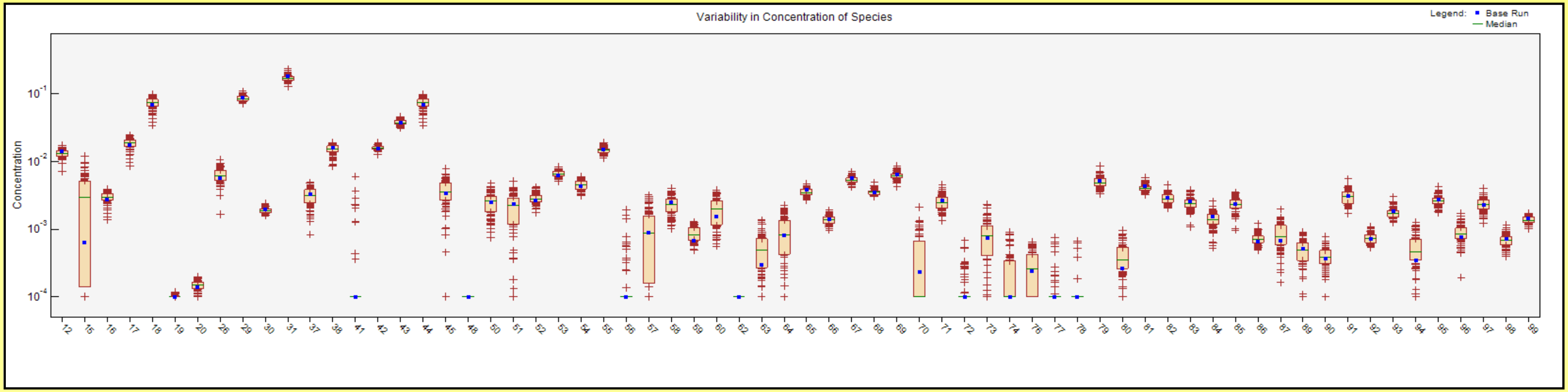


Factor 4 from 3/20/06 data set correlated well with OOA-I and ozone, a gas used as a tracer for OOA-I, the most oxidized OA

Bootstrapping

- To check the stability of the chosen number of factors.
- Randomly replaces ~10% runs within the data set to create a new data matrix.
- Checks how many of the columns in the original data matrix, each m/z , have a one-to-one mapping to the new data matrix.
- Bootstrapping results provide a box plot graph of each m/z, to indicate correlation with the original data matrix.
- A stable solution would match all the profiles to a factor in the bootstrapped case, with a high percentage within the new sets interquartile range (IQR).

Bootstrapping
(Continued)



Graph provided by EPA PMF of Concentration vs. profile for factor two of 03/19/2006 bootstrapped results indicate robustness of the statistics.

Horizontal line represents median bootstrap run, red crosses represent values outside of IQR.

Results

- Four OA factors were found

    - More oxidized (OOA-I)

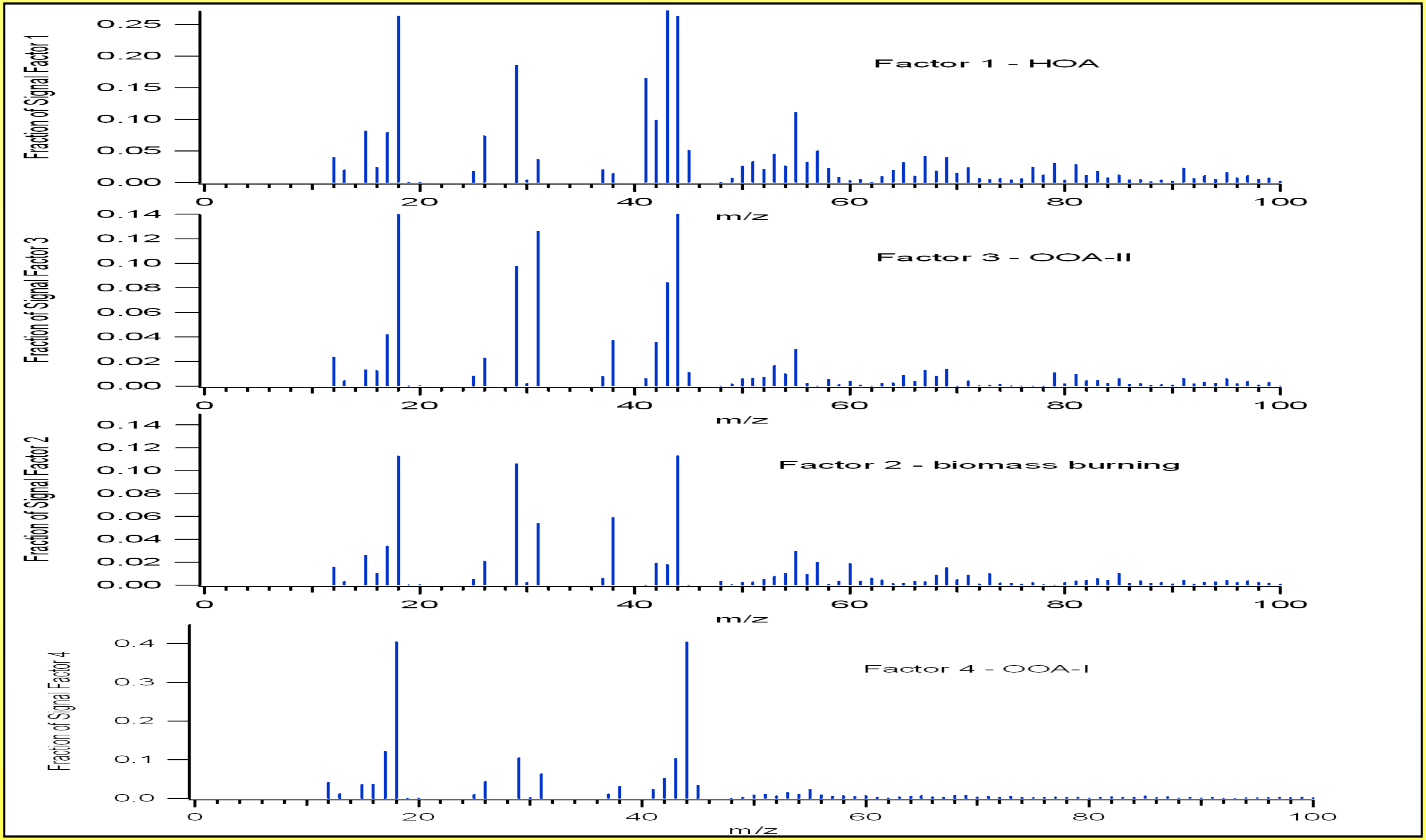
    - Less oxidized (OOA-II)

    - Hydrocarbon-like (HOA)

    - Factor related to biomass burning, which was expected because of the high level of forest fires around time of study

Results

(Continued)

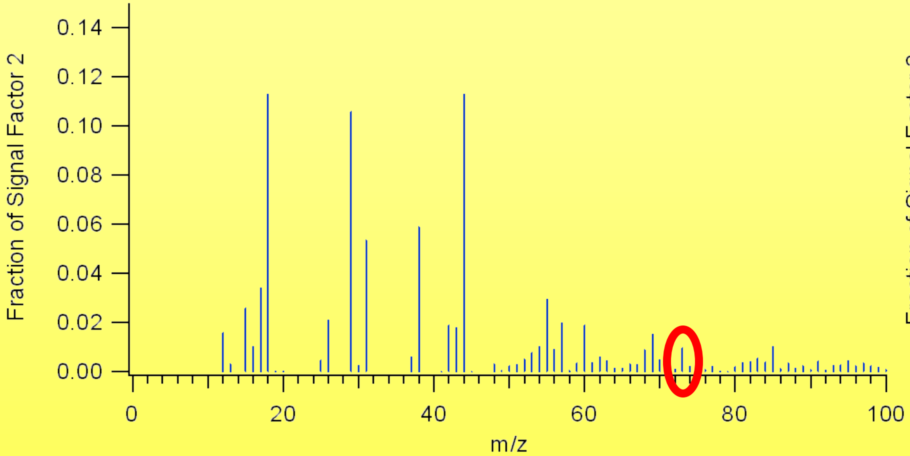


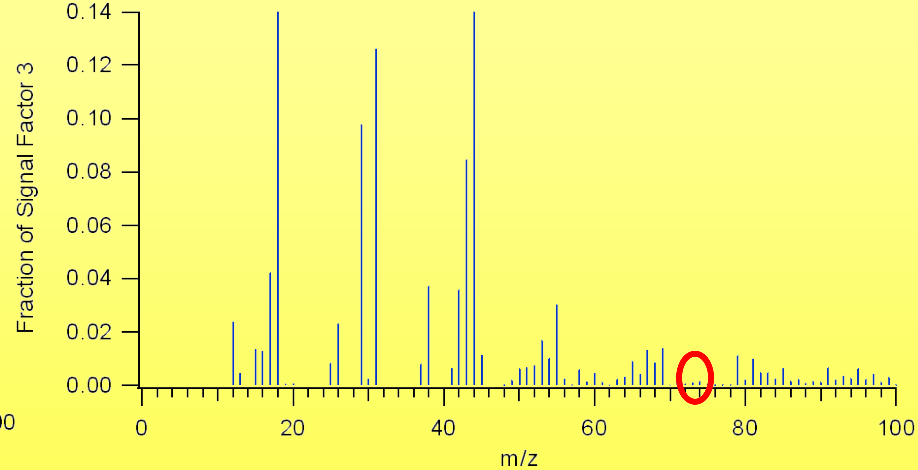
Factor profiles for 3/18/06

Results

(Continued)

- OOA-II and the biomass burning factor were difficult to distinguish, the main difference was a peak at m/z 73 that is characteristic of burning biomass.





Factor 2 was found to be biomass burning, while factor 3 was found to be OOA-II by the lack of m/z 73

Discussion

- EPA PMF showed effective results when processing Mexico City ToF-AMS data
- It should be known that there are many decisions left to the user; therefore, there may be discrepancies within the results.
- Some quantification issues for m/z 27, and 31.
  - They were removed from the analysis so that they would not compromise the results.
- $Q_{robust} << Q_{true} * 1.5$, which was a good indication that outliers were not affecting the factors.

- $Q_{robust} / Q_{expected}$ values were less than 1 indicating that the uncertainty of the data may have been over-estimated*. Additional runs may be needed to explore this.

*I. M. Ulbrich, M.R. Canagaratna, Q. Zhang, D. R. Worsnop, J. L. Jimenez. "Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data". *Atmospheric Chemistry and Physics*, vol. 9, 2891-2918, 2009.

Discussion

(Continued)

- Factors may have had a correlation coefficient < 0.8 with either tracers or known OA spectra.
  - factors produced from the 3 days correlated well with one another indicating that they were all real solutions.
  - The difficulty with correlating the factors produced to tracers is that mixing patterns in the air are complex, and each factor or tracer may be affected differently.
  - Comparing factors to known OA spectra is difficult because OA spectra are either produced in a lab atmosphere which excludes many factors that may affect the results, or constructed by PMF which adds uncertainty to their validity.
- Rotation of solutions was not explored extensively in this study, but it is another aspect of PMF that should be analyzed in order to validate the chosen solution.
- Bootstrapping results for 16, 17, 18 mapped 82.1%, 77.3%, and 78.9% respectively, of data within the IQR, suggesting robustness of the analysis.

Acknowledgements

I would like to thank….

- The GCEP for sponsoring my research.
- Brookhaven National Laboratory for providing me with the facilities to work.
- My mentor Dr. Yin-Nan Lee for all of his help.