Development of Advanced Factor Analysis Methods for Carbonaceous PM Source Identification and Apportionment

Philip K. Hopke

Center for Air Resources Engineering and Science

Clarkson University

Investigators

- This project represents a collaboration among
  - Pentti Paatero, University of Helsinki
  - Ron Henry, University of Southern California
  - Cliff Spiegelman, Texas A&M University
- Work was conducted by
  - Eugene Kim, Weixiang Zhao, Jong Hoon Lee, David Ogulei, and Ramya Sunder Raman

Objectives

- The objective of this project was to combine the best features of the two advanced factor analysis models, UNMIX and Positive Matrix Factorization (PMF), and to test the effectiveness of this improved factor analysis methodology by analysis of the data developed in the various supersites with an emphasis on data from the New York City Supersite and other data from New York State.

Methodological Research

- Part of the effort in the project was methodological studies.
    - Duality of Solutions
        - Singular value decomposition of the data leads to two sets of eigenvectors. One set of eigenvectors spans a space in which source compositions are points and source contributions are hyperplanes. This space is shown to be dual to the space spanned by the second set of eigenvectors of the data in which source compositions are hyperplanes and source contributions are points. The duality principle has been applied to greatly increase the computational speed of the Unmix multivariate receptor model.

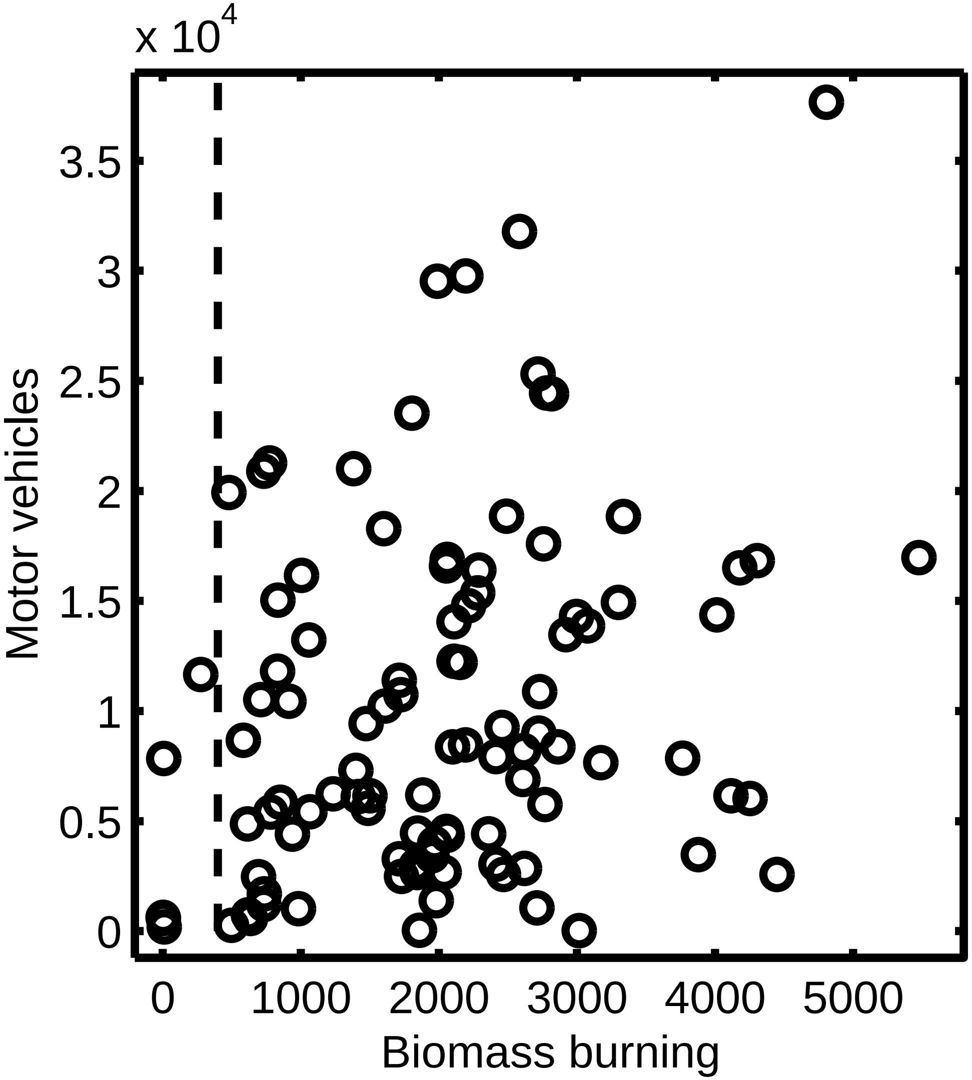
Methodological Research

- G-Space Edges
  - Scatter plots are created of pairs of source contribution factors. When factors are plotted in this way, unrealistic rotations appear as oblique edges that define the distribution of points away from one (or both) of the coordinate axes. With a correct rotation, the limiting edges usually coincide with the axes or lay parallel with them. Inspection of the plots helps one in choosing a realistic rotation.

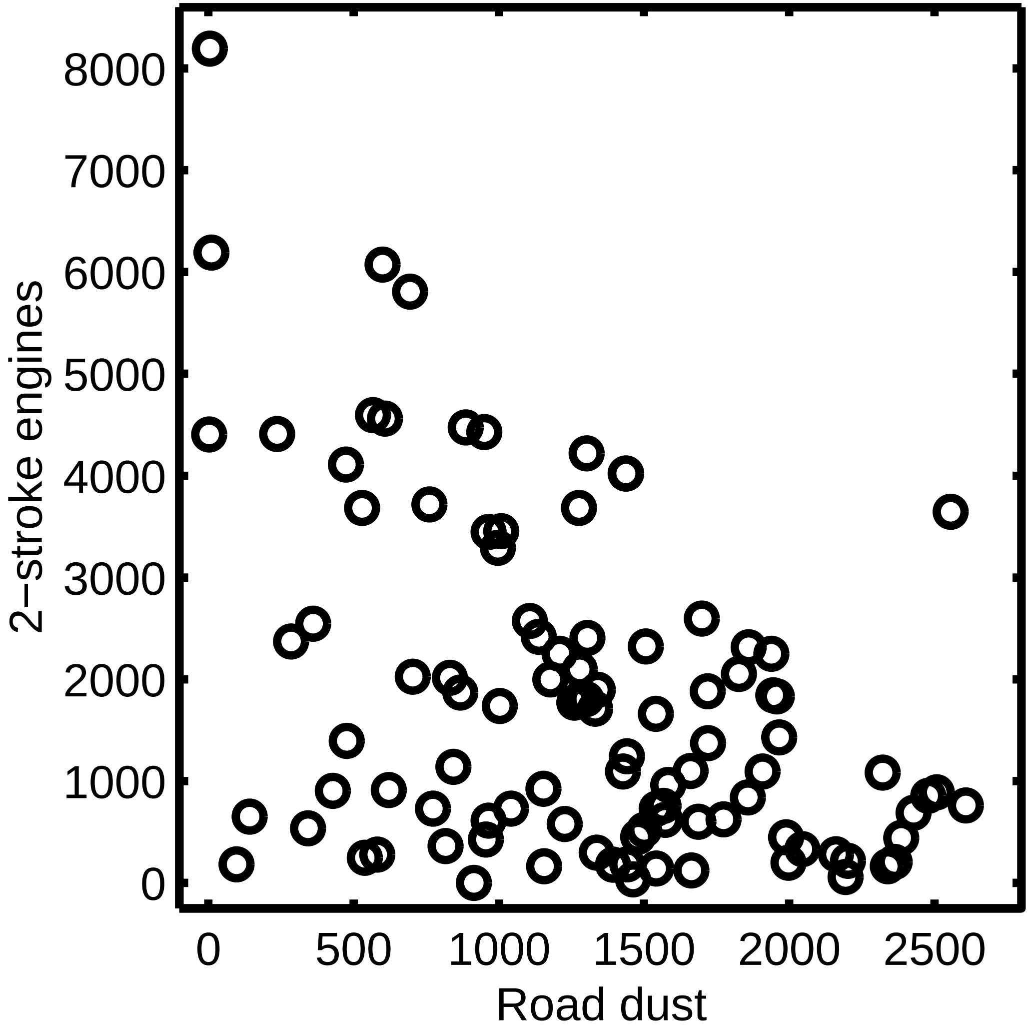
G-Space Edges

- If the two factors are independent of one another, then the resulting contribution values should completely fill the scatter plot and there should be no correlation between them.

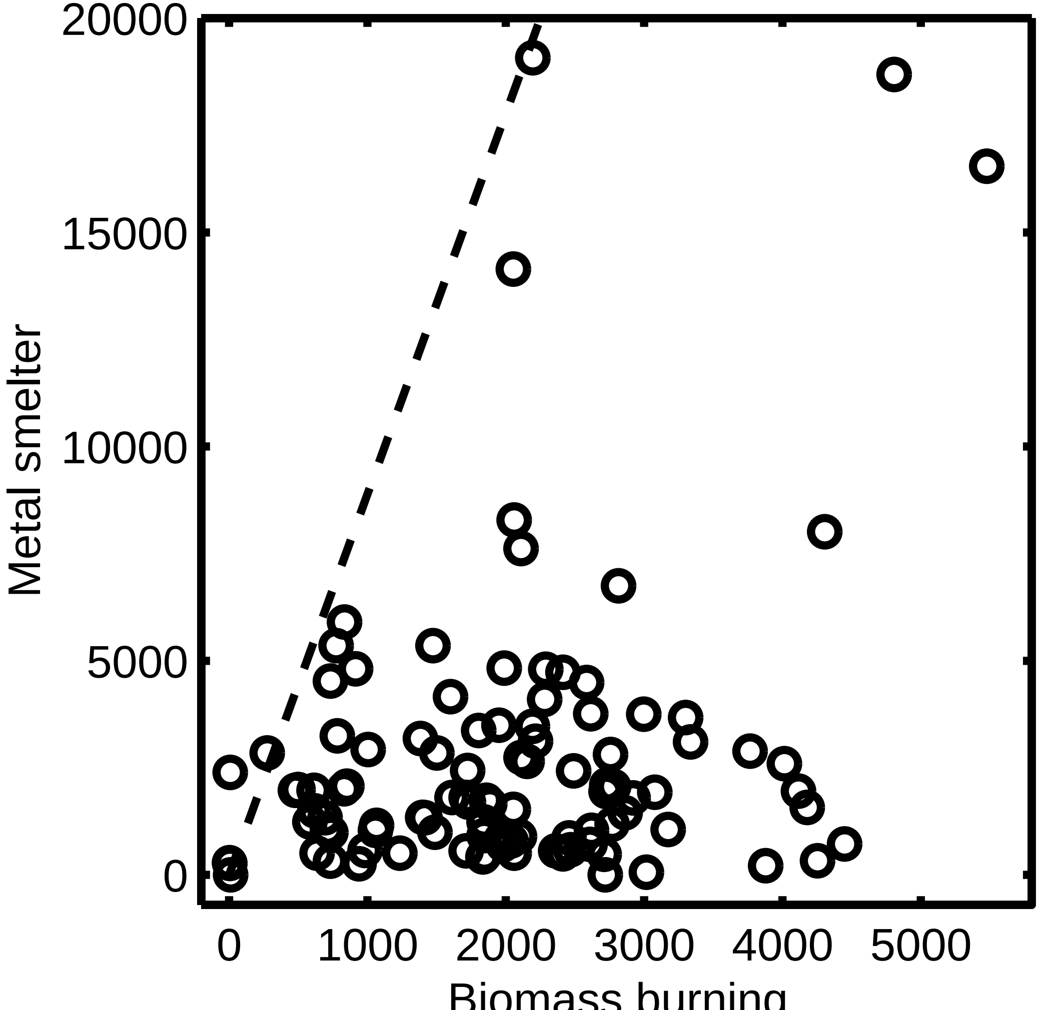
G-Space Edges



G-Space Edges



G-Space Edges



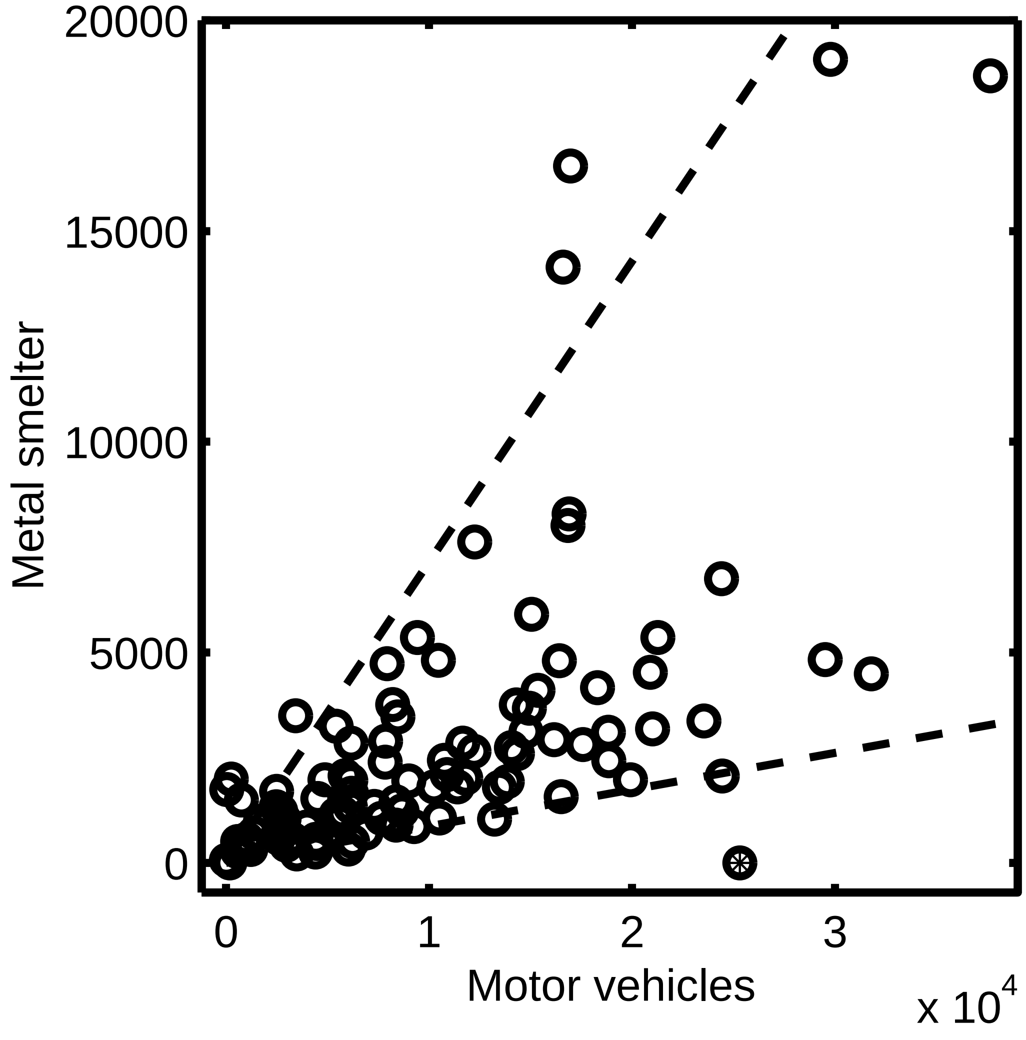
G-Space Plots

- Obviously there is an edge in this plot.

- Does it make sense that these two factors are correlated?

- If not, it suggests the need for a rotation.

G-Space Edges

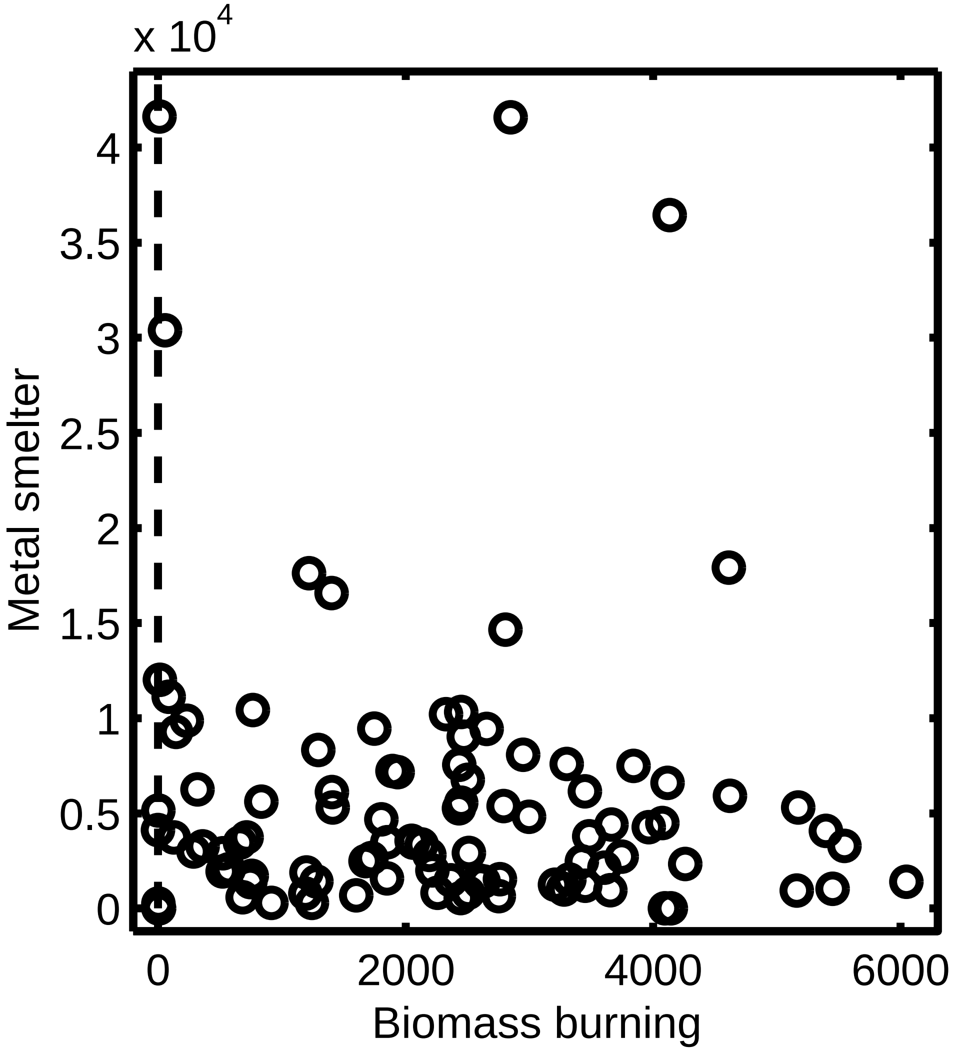


G-Space Edges

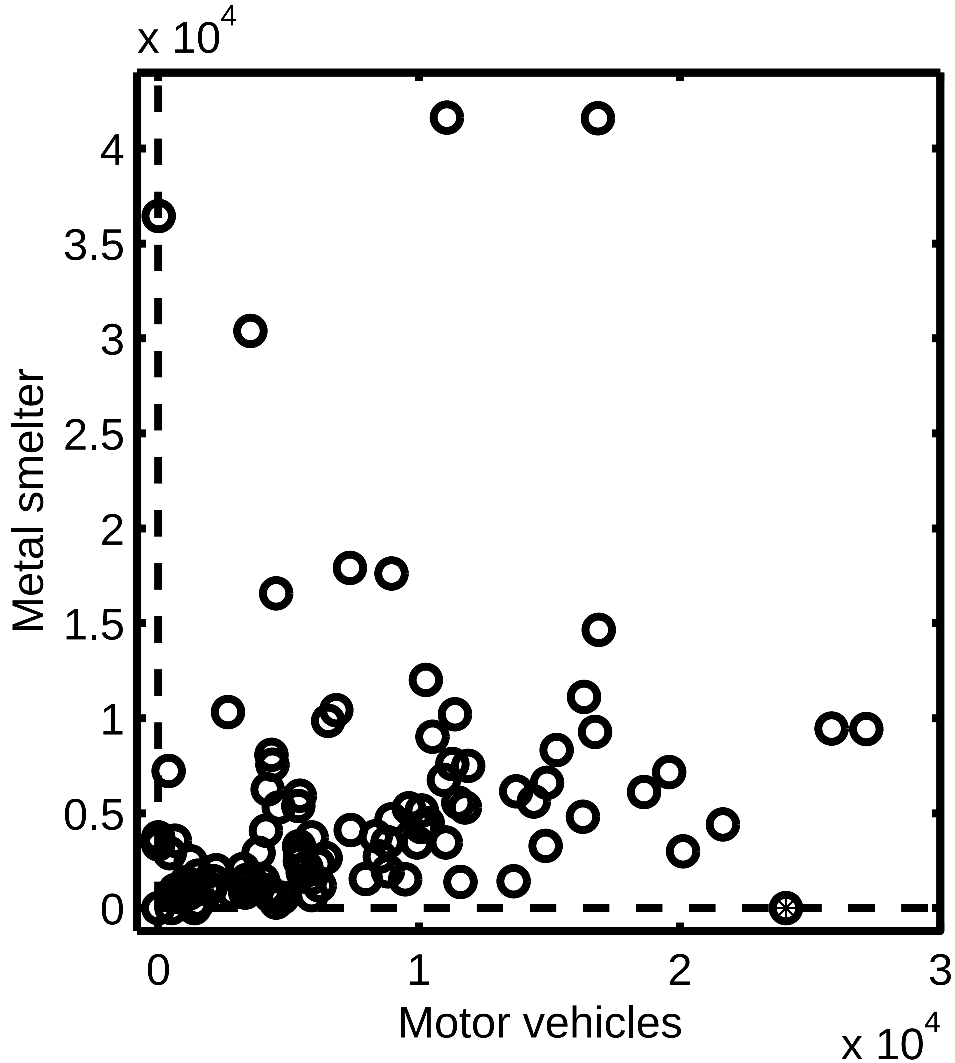
- Note that there can be points outside the apparent edge.

- These points should be checked to be sure they belong. However, it may be necessary to ignore them.

G-Space Edges



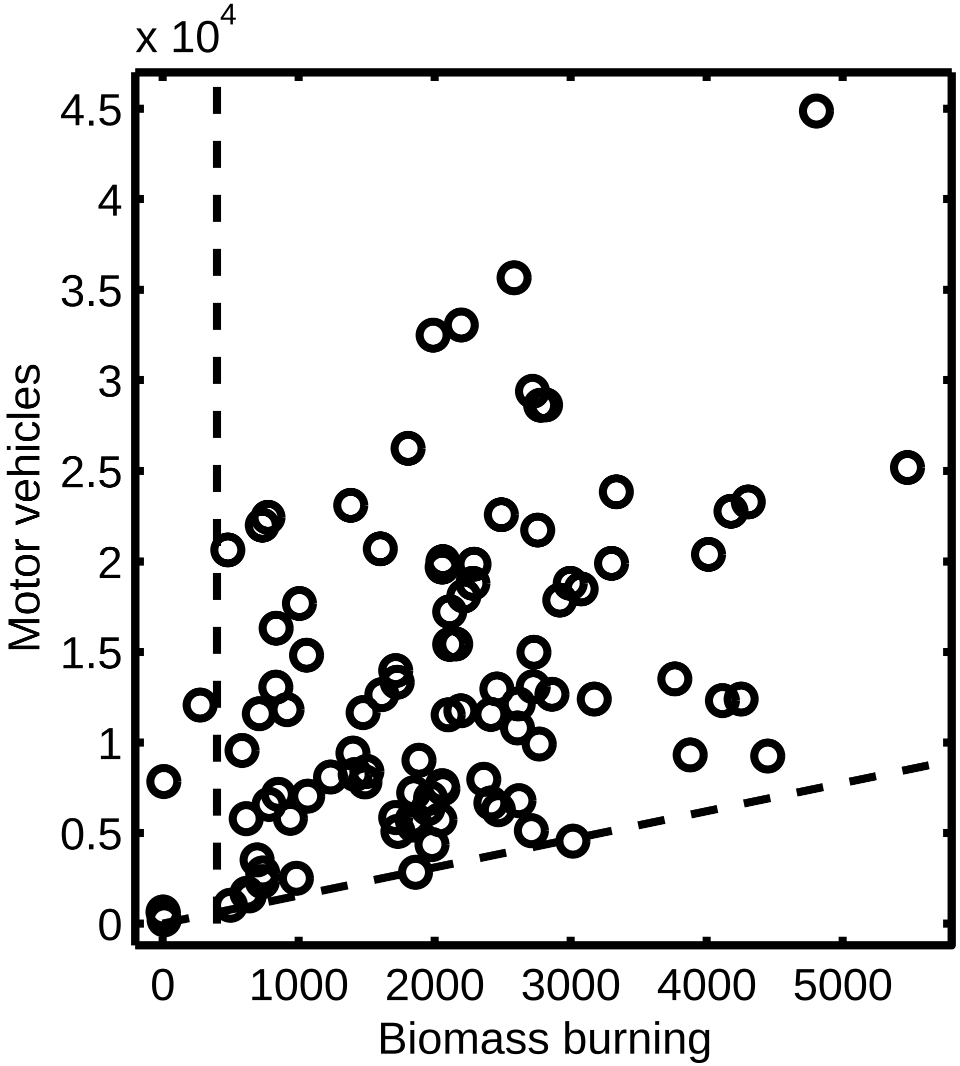
G-Space Edges



G-Space Edges

- Even after rotation, edges can persist.

G-Space Edges

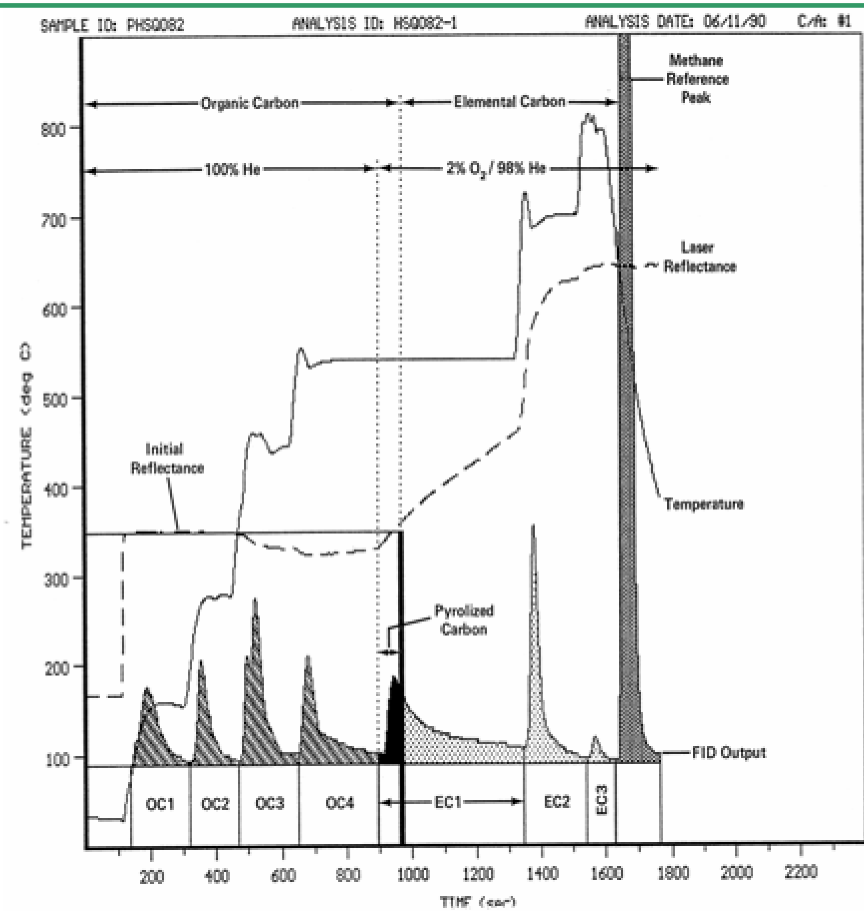


Applications

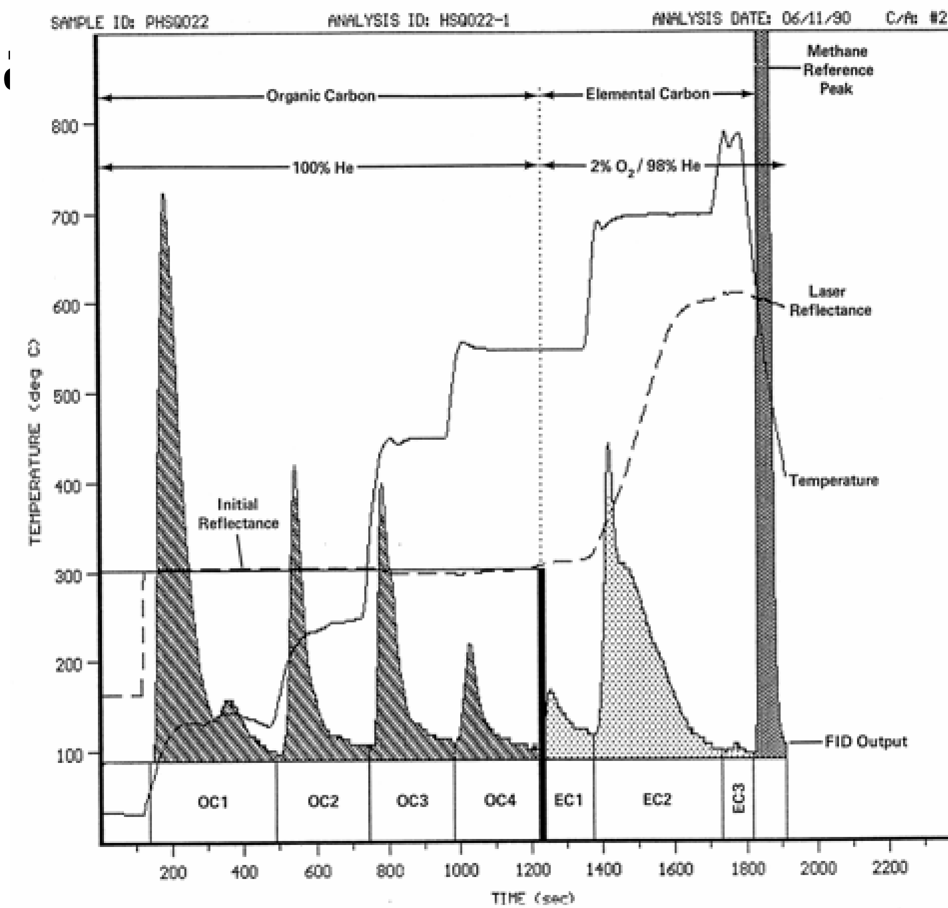
- Particle Composition Data
  - Use of IMPROVE carbon fractions
  - Use of STN data

IMPROVE

Gasoline-fueled vehicles



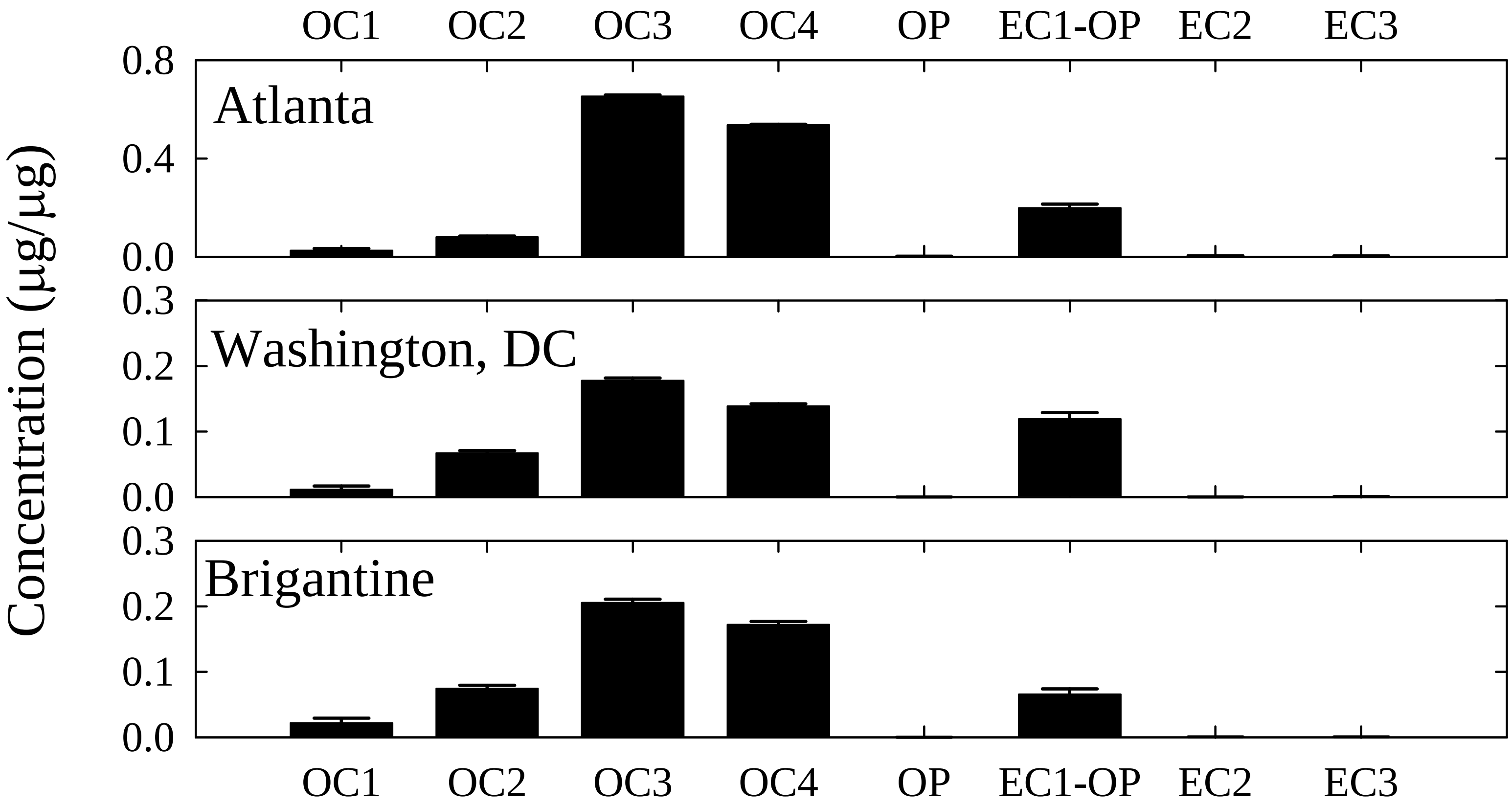
Diesel-fueled vehicles



Watson, J.G.; Chow, J.C.; Lowenthal, D.H.; Pritchett, L.C.; Frazier, C.A.; Neuroth, G.R.; and Robbins, R. (1994). Differences in the carbon composition of source profiles for diesel- and gasoline-powered vehicles. *Atmos. Environ.*, **28**(15):2493-2505.

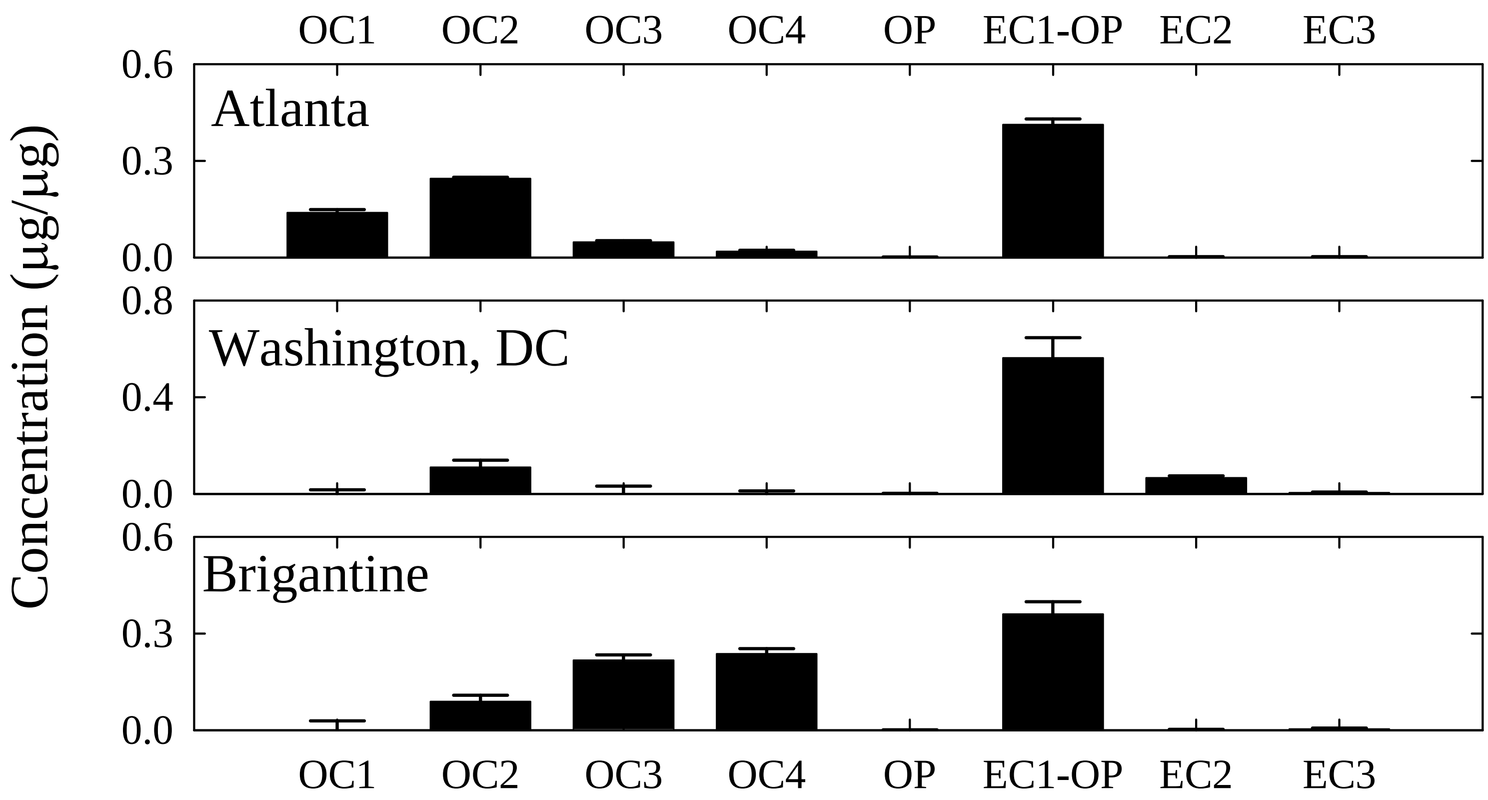
IMPROVE

Gasoline Vehicles



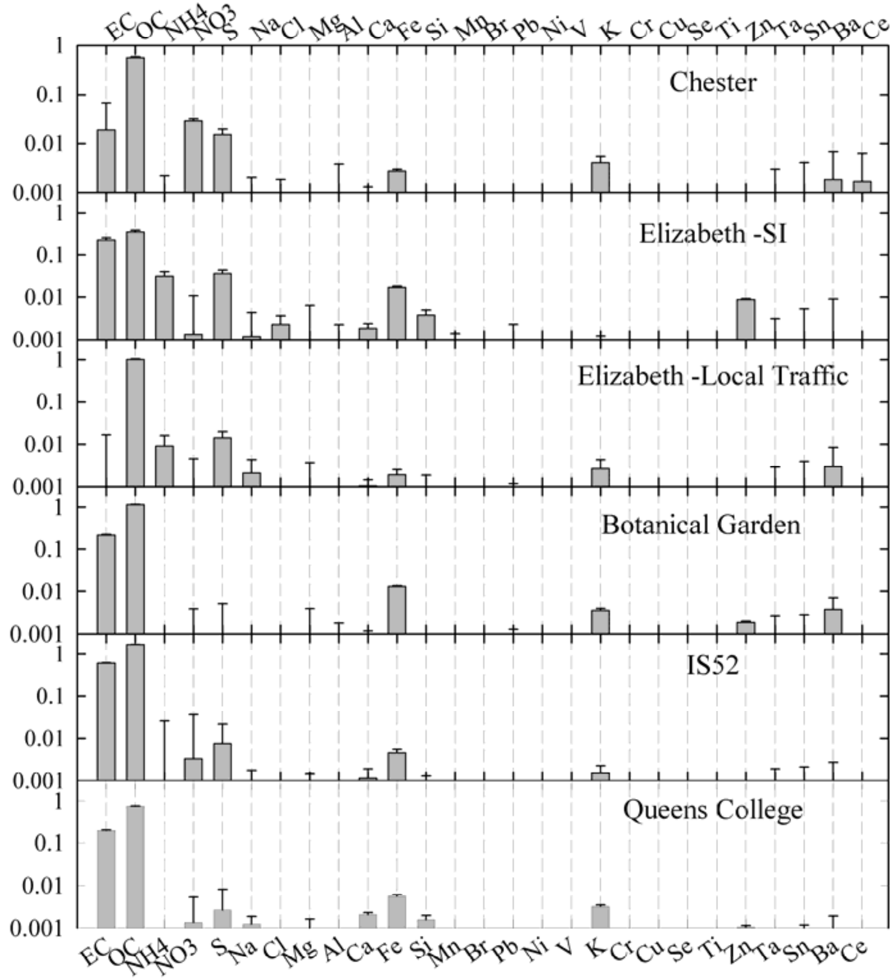
IMPROVE

Diesel Vehicles



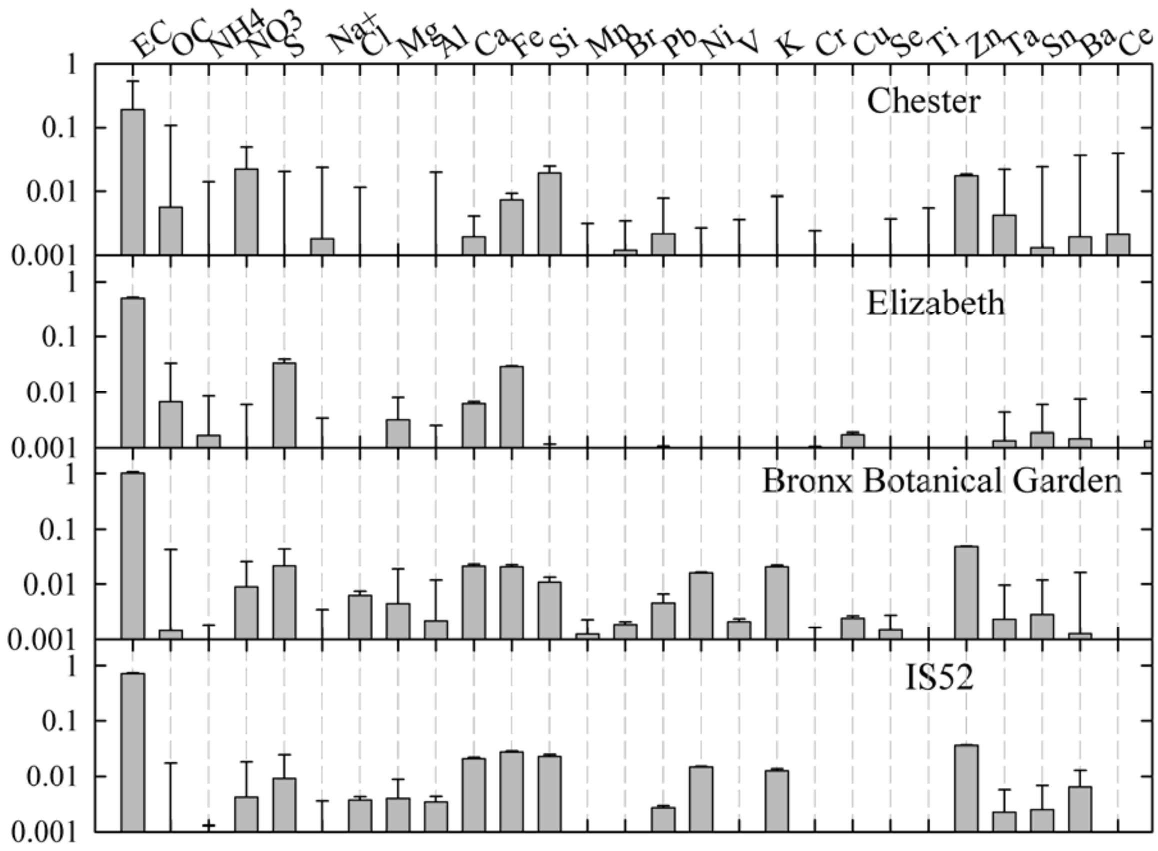
STN-NYC

Gasoline Vehicles



STN-NYC

Diesel Vehicles



Gasoline – Diesel Split

- Can gasoline vehicular emissions be separated from diesel emissions?
  - Shah et al. (*Environ. Sci. Technol.* 38 (9), 2544-2550, 2004) show that stop and go and creeping diesel vehicles emit roughly 50:50 OC/EC as measured with the NIOSH protocol.
  - Problems of “smokers” looking like “diesel” emissions

Gasoline – Diesel Split

- Thus, the “diesel” profile tends to reflect the emissions from vehicles moving at highway speed (i.e., min OC/EC ratio)
- “Gasoline” reflects the maximum OC/EC ratio
- However, the oil additive trace elements tend to go into the “diesel” profile.

Gasoline – Diesel Split

- Does the choice of IMPROVE or NIOSH protocols affect the apportionment and the assignment of mass to “diesel,” “gasoline,” and other major carbonaceous aerosol sources like biomass burning.
- We have an opportunity to make such a comparison using data from the St.Louis-Midwest Supersite.

Comparison of PMF using either IMPROVE or NIOSH Data

- Using daily integrated PM2.5 samples obtained at the St. Louis-Midwest Supersite, OC/EC analyses were performed by the two protocols:
  - OC-EC were originally analyzed at UW-Madison with the ACE-Asia variant of the NIOSH protocol
  - Subsequently, the same samples were analyzed at DRI using the IMPROVE protocol

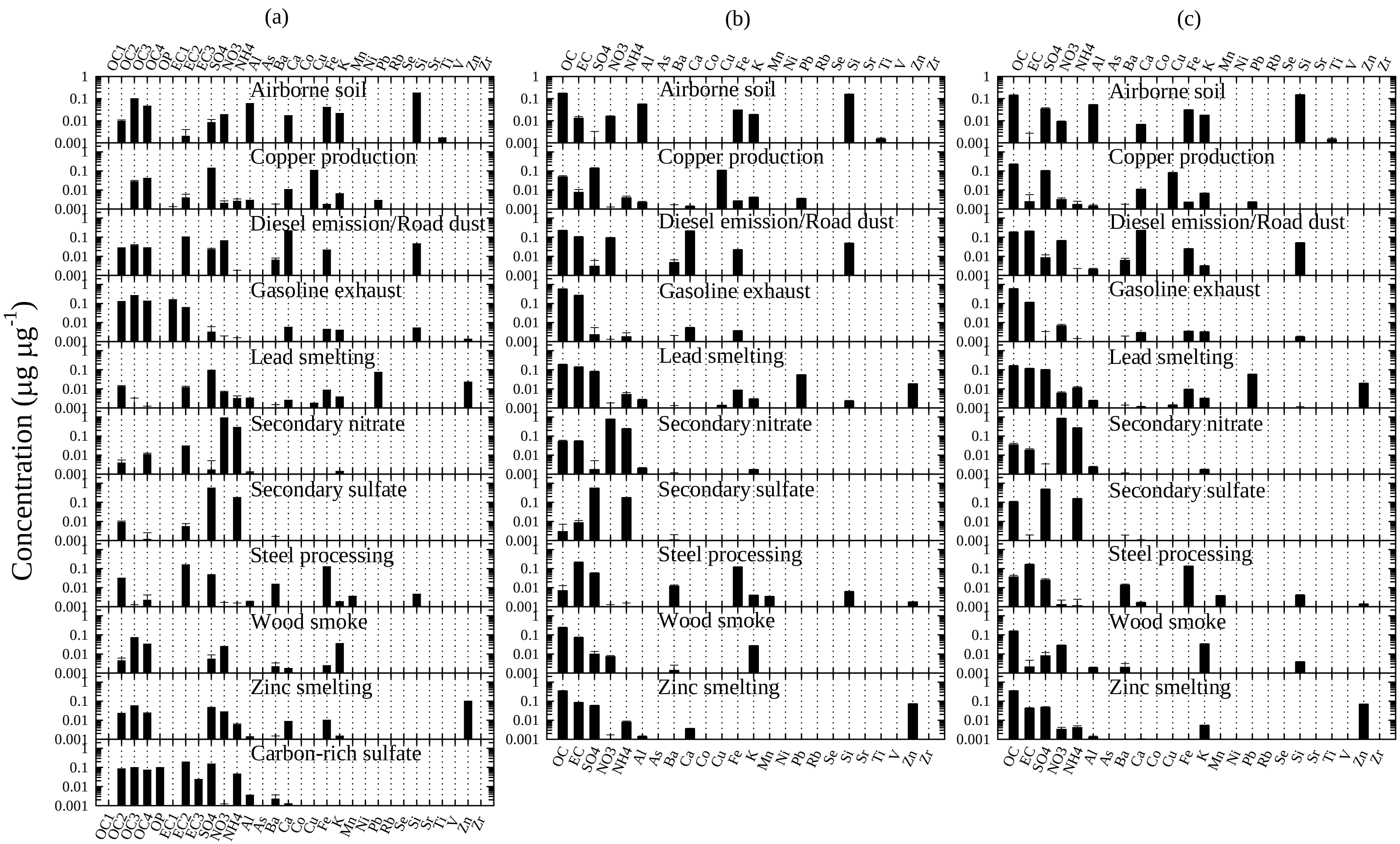
Comparison of PMF using either IMPROVE or NIOSH Data

- Analysis was undertaken for three sets of PM2.5 speciation data at the St. Louis-Midwest Supersite in which each set differs only in the carbon concentrations.
  - The first set (679 samples for 31 species) has eight carbon fractions (OC1 to OC4, OP, and EC1 to EC3) from the IMPROVE protocol.
  - The second set (679 samples for 25 species) included only the total IMPROVE OC (TOC = OC1 + OC2 + OC3 + OC4 + OP) and EC fractions (TEC = EC1 - OP + EC2 + EC3), respectively.
  - The last set (679 samples for 25 species) contains OC and EC concentration obtained by the NIOSH analysis.

Comparison of PMF using either IMPROVE or NIOSH Data

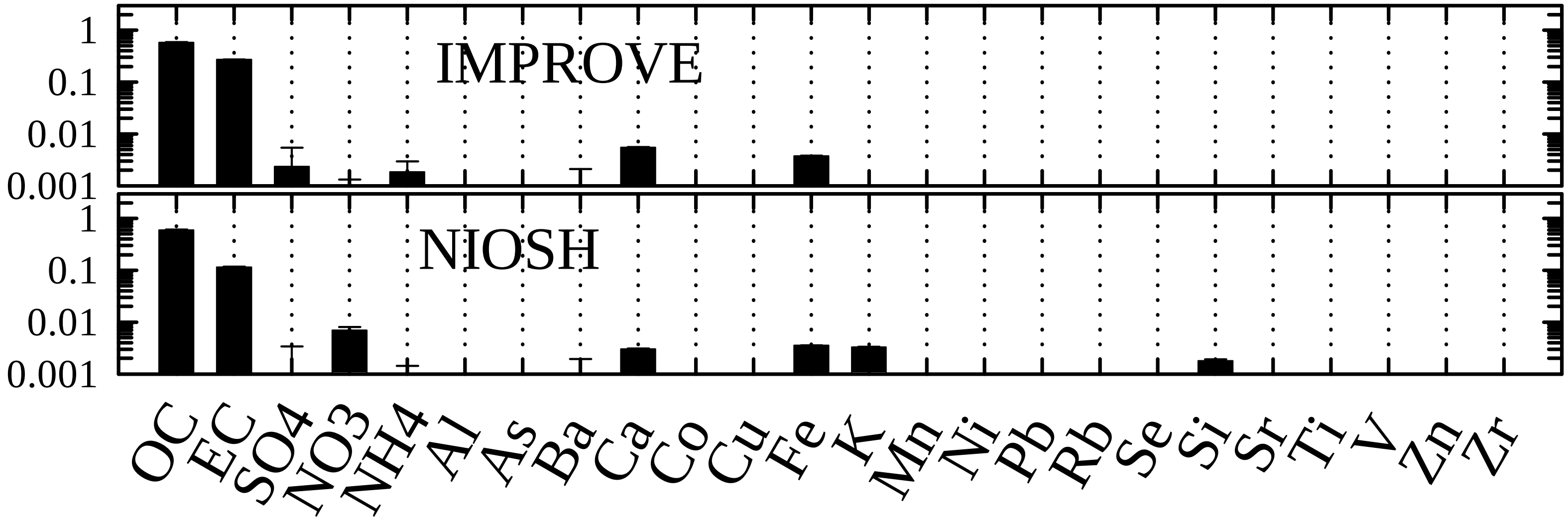
- Solutions with 11 factors, 10 factors, and 10 factors were obtained by IMPROVE carbon fractions, IMPROVE TOC-TEC values, and NIOSH OC-EC values, respectively, for the St. Louis Supersite PM2.5.

Comparison of PMF using either IMPROVE or NIOSH Data



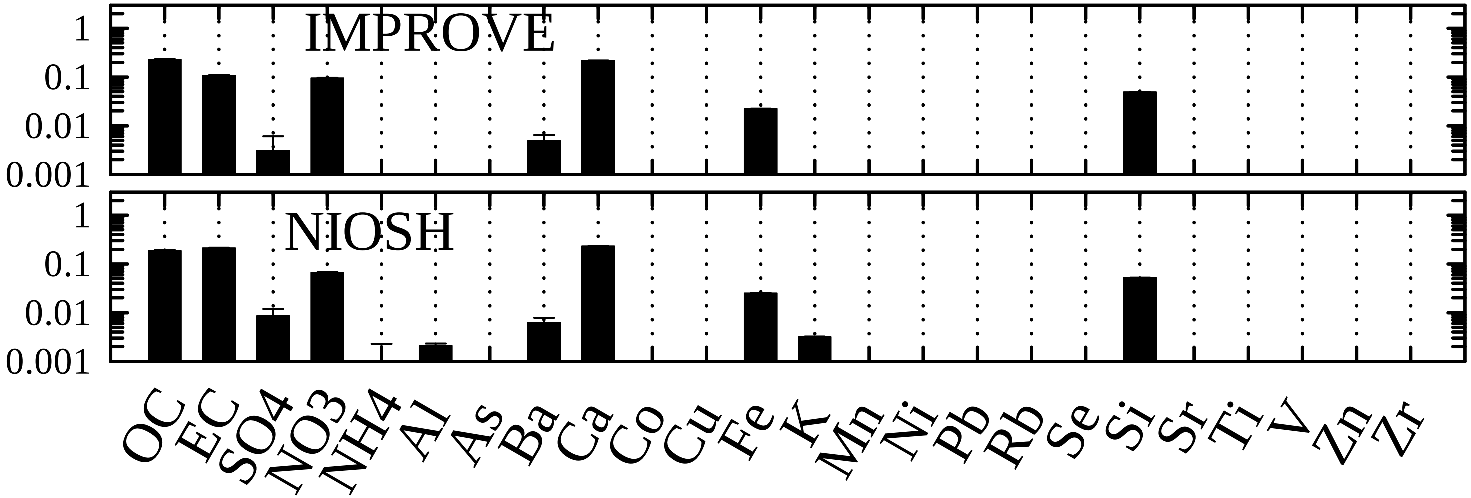
Comparison of PMF using either IMPROVE or NIOSH Data

Gasoline



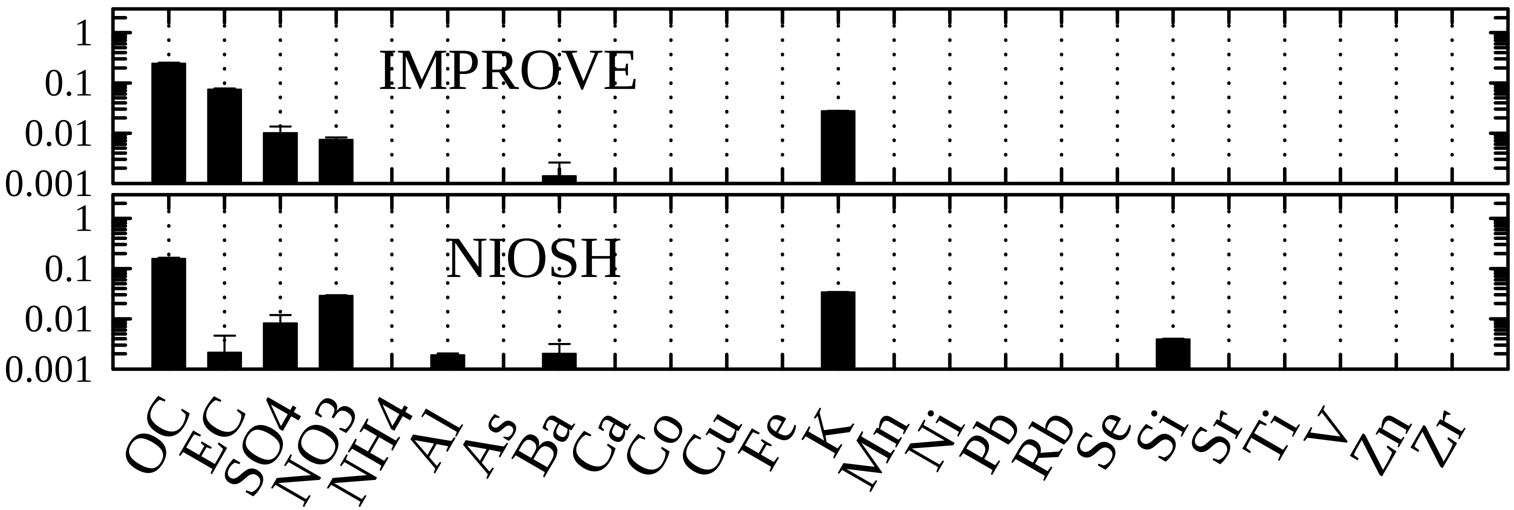
Comparison of PMF using either IMPROVE or NIOSH Data

Diesel



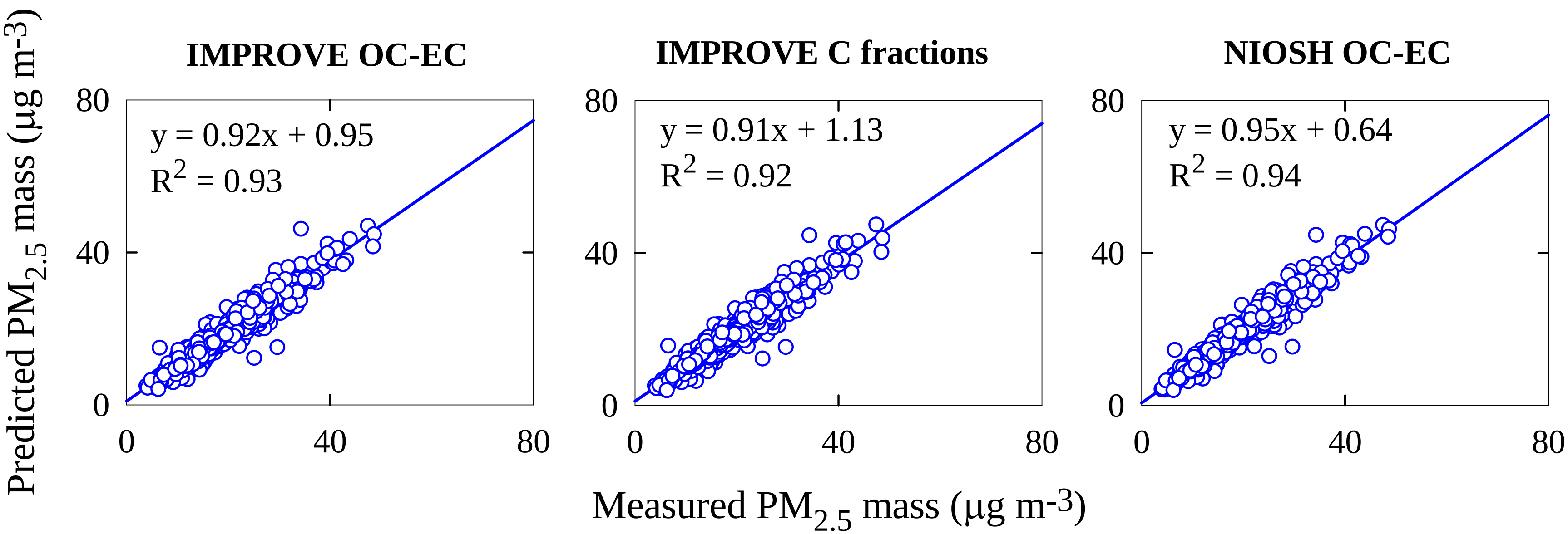
Comparison of PMF using either IMPROVE or NIOSH Data

Biomass Burning



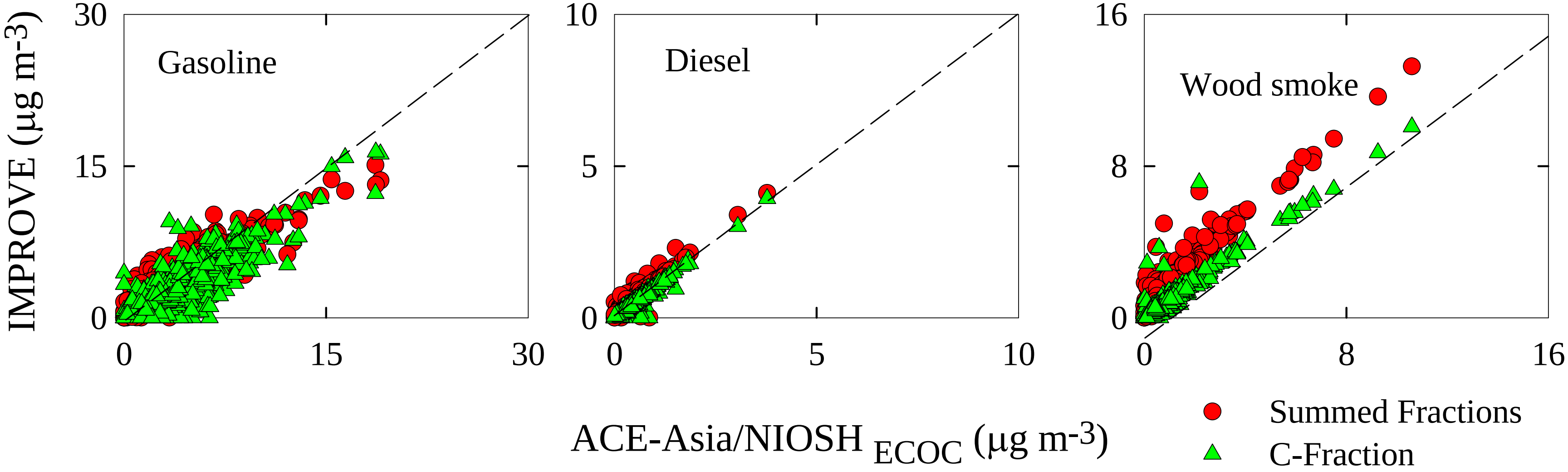
Comparison of PMF using either IMPROVE or NIOSH Data

Mass Reconstruction



Comparison of PMF using either IMPROVE or NIOSH Data

Contribution Comparisons



Publications

Publications

- Comparison between Conditional Probability Function and Nonparametric Regression for Fine Particle Source Directions, E. Kim, and P.K. Hopke, *Atmospheric Environ.* 38: 4667–4673 (2004).
- A Graphical Diagnostic Method For Assessing the Rotation In Factor Analytical Models Of Atmospheric Pollution, P. Paatero, P.K. Hopke, B.A. Begum, S.K. Biswas, *Atmospheric Environ.* 39: 193-201 (2005).
- Duality in Multivariate Receptor Model, Ronald C. Henry, *Chemom. Intell. Lab. Syst.* 77: 59-63 (2005).
- Improving Source Apportionment of Fine Particles in the Eastern United States Utilizing Temperature-Resolved Carbon Fractions. E. Kim and P.K. Hopke, *J. Air & Waste Manage. Assoc.* 55:1456–1463 (2005).
- Estimation of Organic Carbon Blank Values and Error Structures of the Speciation Trends Network Data for Source Apportionment, E. Kim, P.K. Hopke, and Y. Qin, *J. Air Waste Manage. Assoc.* 55: 1190 - 1199 (2005).
- Identification of Fine Particle Sources in Mid-Atlantic US Area, E. Kim and P.K. Hopke, *Water, Soil, and Air Pollution* 168: 391-421 (2005).
- Analysis of Indoor Particle Size Distributions from an Occupied Townhouse using PMF, D. Ogulei, P.K. Hopke, L. Wallace, *Indoor Air* 16:204-215 (2006).
- Source Apportionment of Baltimore Aerosol from Combined Size Distribution and Chemical Composition Data, D. Ogulei, P.K. Hopke, L. Zhou, J.P. Pancras, N. Nair, J.M. Ondov, *Atmospheric Environ.* 40: S396-S410 (2006).
- Comparison between Sample-Species Specific Uncertainties and Estimated Uncertainties for the Source Apportionment of the Speciation Trends Network Data, E. Kim and P.K. Hopke, *Atmospheric Environ.* 41: 567-575 (2007).
- Factor Analysis of Submicron Particle Size Distributions Near a Major United States-Canada Trade Bridge, D. Ogulei, P.K. Hopke, A.R. Ferro, P.A. Jaques, *J. Air Waste Manage. Assoc.* 57: 190-203(2007).
- Modeling Source Contributions to Submicron Particle Number Concentrations Measured in Rochester, NY, D. Ogulei, P.K. Hopke, D.C. Chalupa, and M.J. Utell, *Aerosol Sci. Technol.* 41: 179-201 (2007).
- Spatial Distribution of Source Locations for Particulate Nitrate and Sulfate in the Upper-Midwestern United States, W. Zhao, P.K. Hopke, and L. Zhou, *Atmospheric Environ.* 41:1831-1847 (2007)..
- A Computation Saving Jackknife Approach to Receptor Model Uncertainty Statements for Serially Correlated Data, C.H. Spiegelman and E.S. Park, Chemometrics and Intelligent Laboratory Systems (in press, 2007).

QUESTIONS?