Evaluation of PM2.5 in Chicago by Chemical Mass Balance and Positive Matrix Factorization models

P. Scheff[1] & M. Rizzo[2]
[1]University of Illinois at Chicago
[2]U.S. Environmental Protection Agency Region 5

Abstract

Chemical Mass Balance (CMB) and Positive Matrix Factorization (PMF) models were used to analyze fine particulate matter data from two sites within the city of Chicago. Measurements of metals, organic and elemental carbon, sulfate, nitrate, and gaseous criteria pollutants from the PM2.5 speciation network were evaluated. CMB and PMF results were both strongly influenced by the measurement uncertainty. Variables with a high percentage of measurements below the detection limit were heavily down-weighted so that the models would not be overly influenced by low or unknown concentrations. Variables that were usually above the detection limit were weighted by the root mean square average of 10 % of the measured concentration and the corresponding detection limit.

The analysis yielded a nine source CMB and a 10 factor PMF solution for the Chicago sites. Sources represented by the factors were identified using established source profiles from literature and mass to mass ratios of species. The sources identified included secondary sulfate and nitrate, motor vehicles, coal-fired utilities, vegetative burning, wind blown dust, salt used to de-ice roadways and steel production. CMB and PMF predictions for source contribution and composition were compared and contrasted. The two models provided remarkably consistent results. The estimated daily contributions from each source revealed seasonal patterns which also aided in source identification.
Keywords: Chemical Mass Balance, Positive Matrix Factorization, Particulate Speciation Trends Network, receptor models, source composition.

1   Introduction

In 2000, the United States Environmental Protection Agency established the Fine Particulate Speciation Trends Network to expand on its existing PM2.5 monitoring activities. The purpose of the network is to characterize individual species which compose the total fine particulate measured at the Agency's Federal Reference Method (FRM) PM2.5 monitoring sites. The data from the speciation network serves an important role in aiding the Agency in determining which species are the most prevalent in areas of the nation thus allowing for the formulation of control strategies. For example, studies have already shown that secondary sulfates comprise large part of the fine particulate in the Eastern part of the United States while secondary nitrates dominate the total PM2.5 in the Western United States [1].

Another use planned for the data collected through the speciation network is to support fine particulate matter source apportionment studies. Traditional source apportionment techniques have centered around the use of the Chemical Mass Balance (CMB) model which uses source composition (independent variables) and speciated air quality data (dependent variable) to determine source contributions for either gases, particles or a combination of both [2,3]. One limitation of this approach is that source profiles for many of the important PM2.5 sources are not well defined. A complementary receptor modeling technique which does not require the specification of source profiles to provide an indication of possible source impacts is Positive Matrix Factorization (PMF). This technique is a form of factor analysis where the underlying co-variability of many variables (e.g. sample to sample variation in PM species) is described by a smaller set factors (e.g. PM sources) to which the original variables are related. PMF has been used in a variety of source apportionment and spatial analysis studies [4, 5]. One important advantage of PMF is that it does not require source profile information to determine the possible source contributions as with the CMB model. Furthermore, CMB assumes that the fitting species used in the analysis are non-reactive, that is they do not change significantly in the atmosphere between the time of emission and collection at the receptor location. PMF does not impose this restriction. However, it can be difficult to identify potential sources defined by the PMF model without source profile information to compare the predicted composition with. This work will compare and contrast CMB and PMF solutions for the same data set in Chicago and examine the strengths and weaknesses of the models.

2   Methods

Both models solve the following equation:

$$X = G\,F + E$$

where X (n x Sp)= matrix of Sp speciated concentrations on n days;
G (n x f) = matrix of f source contributions on n days;

F (f x Sp) = matrix of f source profiles for Sp species; and

E (n x Sp) = matrix of errors for Sp species on n days.

For the CMB model, X and F are measured and G and E are solved by least squares statistics. For the PMF model, the number of sources is specified and the G and F matrices are estimated with a non-negative constraint (concentrations and compositions can not be less than zero). CMB calculations were performed with the SAS programming language using the effective variance weighted least squares procedure [6]. Details of the PMF analysis have been presented [7]. PMF calculations were performed for a 10-factor model with the Multilinear Engine 2 (ME2).

Source fingerprints were derived from published studies. Sources modeled included soil, biomass burning, motor vehicles (gas and diesel composite), steel industry, coal combustion, road salt, petroleum refineries, and secondary sulfate and nitrate. The vehicle profile was compiled from a series of composites obtained from the Northern Front Range Air Quality Study (NFRAQS) conducted by Desert Research Institute [8]. The secondary sulfate and nitrate profiles were calculated using the mass ratios of the individual constituents for each of the two salts. Wood burning profiles from NFRAQS and a study conducted by Jamie Schauer were used to construct the vegetative burning profile [8, 11]. The steel profile was constructed from a series of profiles for a different steel processes used in a South Africa steel mill for a study conducted by Desert Research Institute [12]. The petroleum refinery profile is an average of a composite petroleum refinery profile from NFRAQS and a separate refinery profile collected during a study in Robbins, Illinois examining the emissions from a local waste incinerator [8, 13]. The rock salt profile represents the analysis of the bulk road salt used for snow removal in the Chicago metropolitan area [13]. The soil profile is a composite of several profiles created during the Robbins incinerator study by Desert Research Institute and represents the local soil component for the Chicago area [13]. The coal plant profile represents the average of the profiles from NFRAQS and two other studies which examined the chemical composition of power plants using high sulfur coal which is common in the Midwestern United States [8, 9, 10].

Data were combined from the two speciation-trends sites within Chicago for 2001 and 2002 to create a data set of 241 observations. Each observation included 4 gases (SO_2, NO_x, NH_4, and CO), 4 ions ($SO_4^=$, NO_3^-, Na^+ and K^+), 2 carbon fractions (organic and elemental carbon) and 48 elements. Multivariate factor analytic techniques have been shown to be sensitive to variables with a high proportion of data less than the minimum detectable limit (MDL). Thus, the uncertainty for each value was based on a signal to noise analysis of the importance of individual species given the number of samples above the method detection limit [14].

One of the difficult judgments in PMF modeling is determining the number of sources to include in the model. The number of factors to be included in the final solution was determined from a preliminary factor analysis where the eigen value corresponding to the inclusion of each successive variable was plotted against the number of variables included in the analysis. This is commonly

known as a Scree plot and gives an indication of the number of factors appropriate for a solution when the line approaches horizontal. For the Chicago example, solutions of five to eleven factors were calculated in ME2 and compared to the results obtained by the PMF analytical tool. This analysis showed consistent results between the two methods for a 10 factor solution.

3 Results

Table 1 shows descriptive statistics for the CMB and PMF solutions. The table includes tentative identifications of the PMF sources. Figure 1 is a scatter plot matrix of CMB and PMF daily source concentrations and show similarity and differences between model predictions.

Table 1:   Average CMB and PMF source contributions.

Model	Source	Minimum	Maximum	Median	Mean	Std Dev
CMB	Coal	0.000	3.43	0.134	0.206	0.311
CMB	Soil	0.000	10.87	0.288	0.489	0.896
CMB	Steel	0.000	4.26	0.193	0.340	0.541
CMB	Burning	0.000	11.57	1.460	1.781	1.447
CMB	Vehicle	0.642	12.56	4.44	4.71	2.240
CMB	Salt	0.000	3.34	0.00	0.135	0.445
CMB	Refinery	0.000	3.54	0.049	0.088	0.268
CMB	Sulfate	0.479	30.2	3.48	4.99	4.55
CMB	Nitrate	0.143	18.66	2.19	3.07	2.91
PMF	Utility	0.000	1.181	0.120	0.188	0.197
PMF	Soil	0.000	6.33	0.578	0.830	0.832
PMF	Steel	0.000	1.695	0.090	0.164	0.204
PMF	Fe Mn	0.000	2.84	0.249	0.398	0.436
PMF	Copper	0.000	4.372	0.184	0.296	0.458
PMF	Burning	0.000	6.63	0.439	0.662	0.816
PMF	Vehicle	0.000	9.04	2.55	2.65	1.377
PMF	Salt	0.000	4.59	0.060	0.228	0.575
PMF	Sulfate	0.000	39.4	3.79	5.94	5.91
PMF	Nitrate	0.000	20.2	2.36	3.44	3.42

   Table 1 and figure 1 show strong agreement between these two models. For example, predictions for road salt, sulfate and nitrate are strongly correlated with very similar mean concentrations. Other sources with moderate correlations include biomass burning, soil and vehicles. CMB vehicle and biomass burning predictions are higher than PMF while PMF soil concentrations are higher than CMB solutions. Except for burning, mean differences are all less than a factor of 2. Figure 1 also shows that the CMB steel source is moderately correlated to the PMF iron and manganese source. The mean concentrations of these sources are also similar.
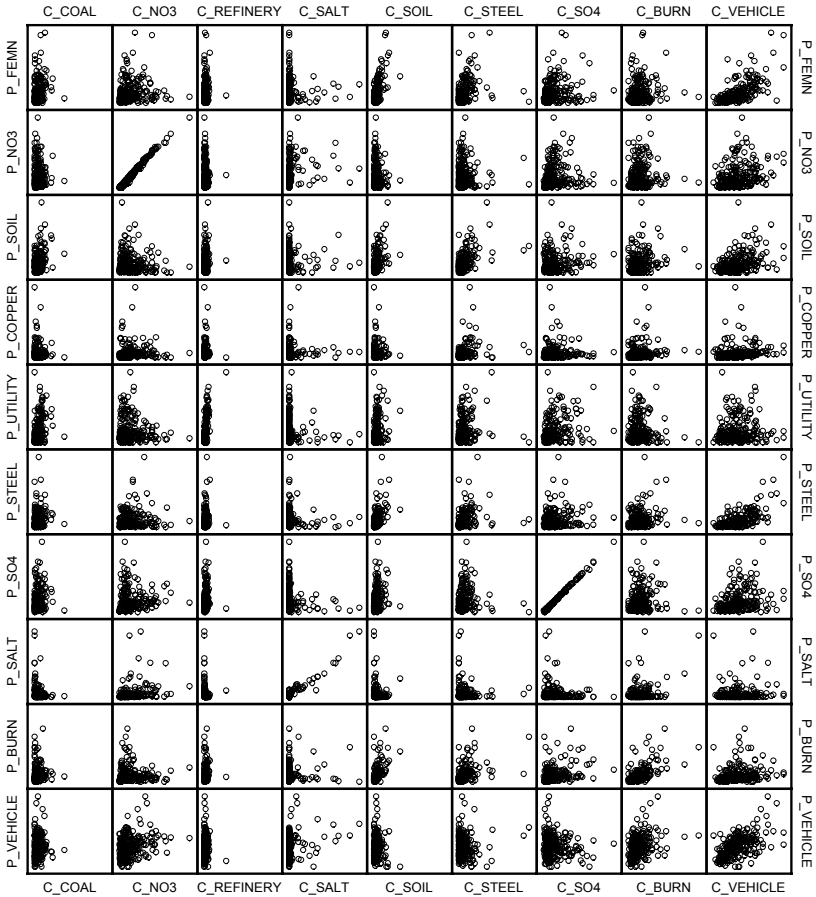
Figure 1: Scatter plot matrix of CMB versus PMF solutions.

4   Discussion

One of the challenges of PMF modeling is the naming of the predicted sources. This is done by comparing the predicted source compositions with known source compositions (F matrix) and evaluating the variability of the sample to sample

predictions (G matrix). For example, sources with seasonal patterns can be identified when plotting the G matrix (predicted source concentrations) as a time series. In this study, the preliminary identifications in table 1 were determined by time series plots as well as comparison with CMB predicted source contributions and measured source profiles.
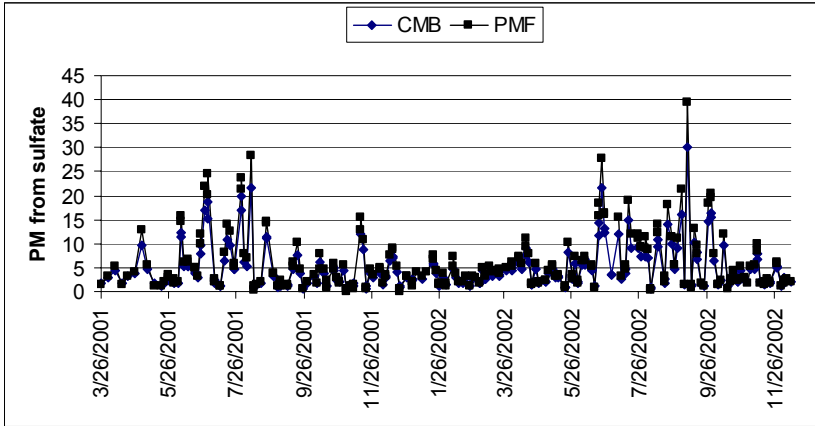

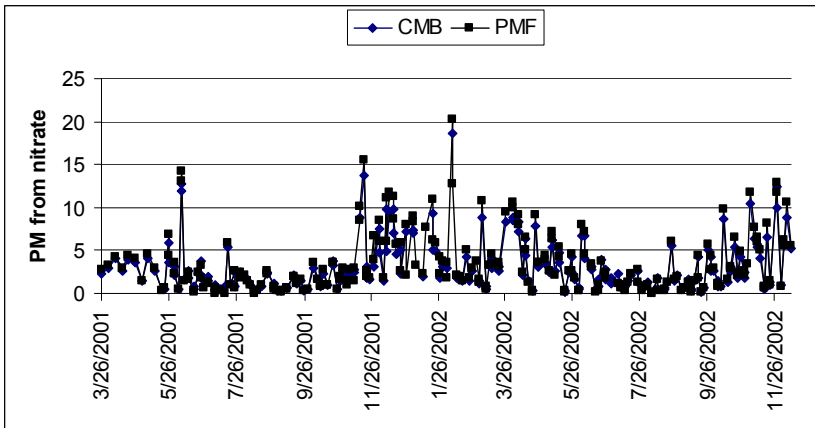Figure 2: Time series plot of CMB and PMF predicted sulfate.

Figure 3: Time series plot of CMB and PMF predicted nitrate.

Figure 2 shows the time series plot of CMB and PMF predictions for sulfate and shows increased sulfate concentrations during the summer. In contrast, figure 3 shows the time series plot for nitrate and shows higher levels during the winter.

Figures 4, 5, 6 and 7 compare PMF predicted source compositions for the vehicles, road salt, soil and biomass burning sources with the measured source profiles used in the CMB analysis. The agreement between the measured vehicle profile and the PMF estimated profile (figure 4) is remarkable. The predicted profile captures the major components perfectly including the gases (CO, NO$_x$

and SO$_2$) and the major particle components (elemental and organic carbon). While the PMF prediction overestimates the vehicle sulfate component, it does an excellent job with many metals including Ca, Cr, Fe, Mg, P, Pb, and V.
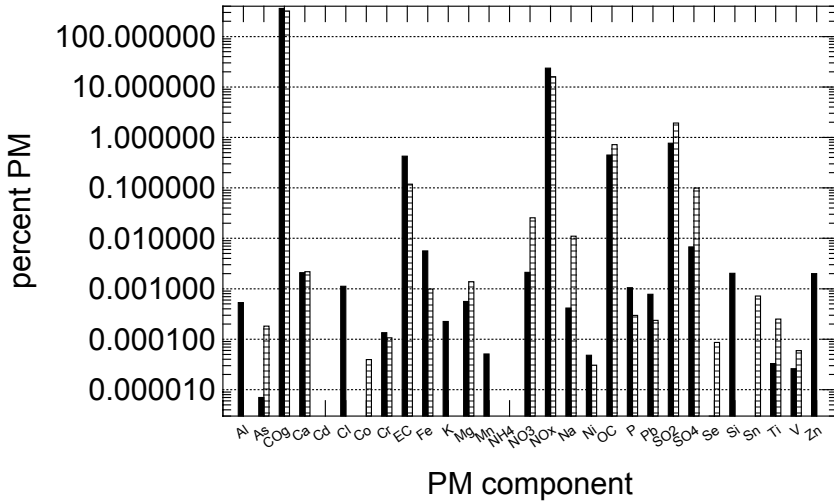

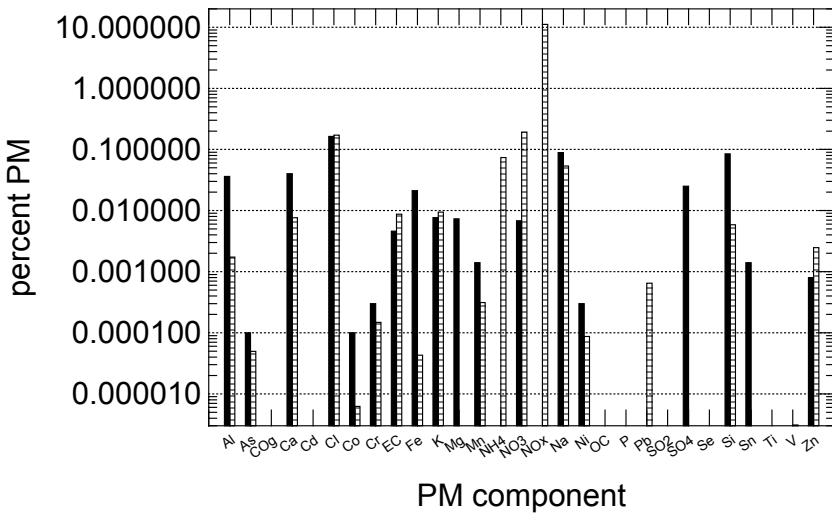Figure 4: Comparison of CMB ■ and PMF □ vehicle fingerprints.

Figure 5: Comparison of CMB ■ and PMF □ road salt fingerprints.

The estimated profile for road salt aerosol (figure 5) also has a number of excellent predictions. For example, the sodium and chlorine predictions are almost exactly equal to the measured composition. Other reasonable good predictions include As, Ca, Cr, Ec, K, Mn, Ni, and Zn. It is interesting that the

PMF fingerprint includes predictions for gases associated with motor vehicle emission (NO_x, and NH_4^+) as well as Pb which was not in the measured source composition. This could be an indication of problems with the measured composition. This excellent agreement is also seen in the strong correlation between the CMB and PMF predictions in figure 1.
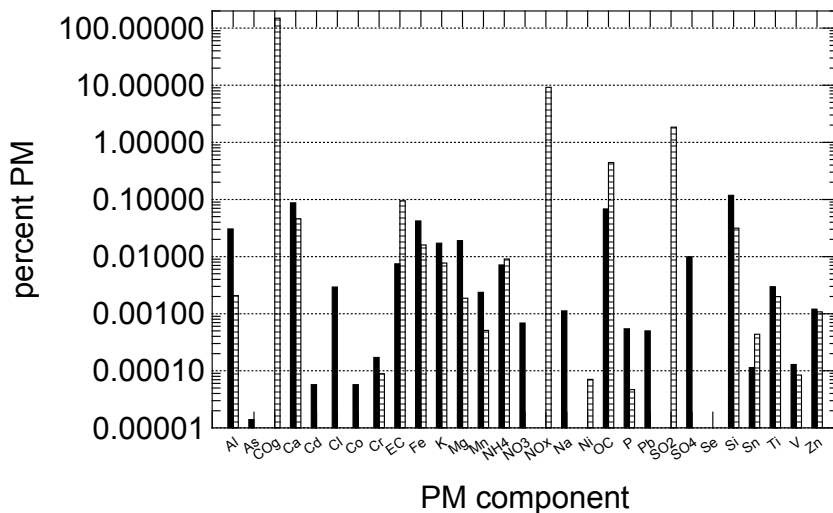

Figure 6: Comparison of CMB ■ and PMF □ soil fingerprints.

The comparison of measured to predicted soil composition (figure 6) is also interesting. The figure shows excellent agreement for many of the major soil elements including Ca, Fe, K, Si, Ti, and Zn. There is also good agreement for other major components including NH_4^+ and organic carbon. Many of the minor elements also compare very well including Cr, Sn and V. Possible problems with the predicted composition include fairly high predictions for gases (CO, NO_x, and SO_2) which are clearly not associated with soil particles. The PMF prediction also did not find the sulfate and Cl in the soil particles and underestimated the Al mass fraction. One strength of the ME2 for performing PMF calculations is that it allows the partial specification of the F matrix (source composition). Future runs with this data set will explore the effect of removing the gases from the predicted soil composition.

Figure 7 compares the measured and predicted compositions for the bio-mass burning source. The predicted composition for major gases including CO and SO_2 is excellent along with major PM components including elemental and organic carbon and K, a frequently used tracer for bio-mass burning. The PMF model did not pick up the Al, Cl, NH_4^+, NO_x, P and Zn in the measured composition while over predicting the As, Fe, NO_3^-, $SO_4^=$, Si, Sn, Ti and V.

5   Conclusions

The 10-factor PMF and 9-source CMB models produced remarkably similar predictions for PM2.5 aerosol in Chicago. Predictions for road salt, sulfate and

nitrate are strongly correlated with very similar mean concentrations. Other sources with moderate correlations include biomass burning, soil and vehicles. Average CMB vehicle and biomass burning predictions are higher than PMF while PMF soil concentrations are higher than CMB solutions. Except for burning, mean differences are all less than a factor of 2.
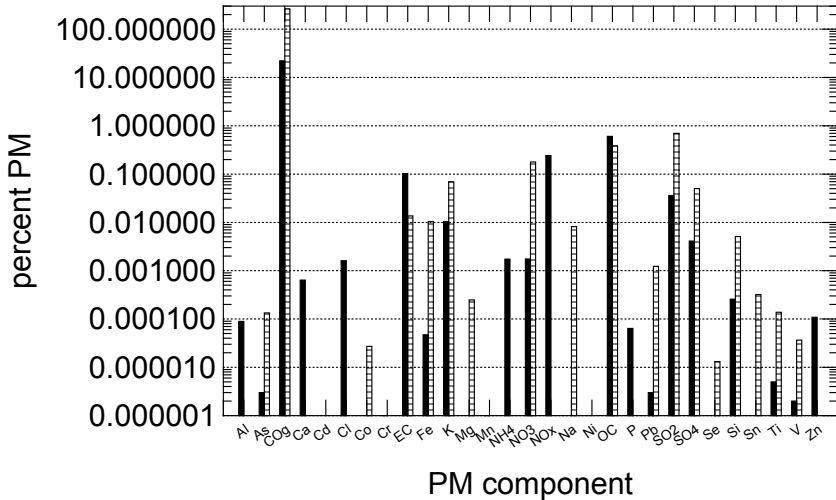

Figure 7: Comparison of CMB ■ and PMF □ biomass burning fingerprints.

It was not always easy to name the PMF factors. For example, figure 1 shows that the CMB steel source is moderately correlated to the PMF iron and manganese source. The mean concentrations of these sources are also similar. It is not clear if our identification of the PMF steel source is correct. We were also able to include a source for PM emissions from petroleum refineries in the CMB analysis. However, the PMF model was not able to find a factor with similar composition. The only source that the CMB refinery prediction correlated with was the source identified as PMF utility suggesting that our identification of that source may also not be correct.

Despite these differences, the ME2 solution of the PMF model was able to reproduce many of the major features of a number of important PM2.5 sources in Chicago including motor vehicles, soil, road salt, biomass burning, sulfate and nitrate. The next step in this analysis is to use the ME2's ability to partially specify the F matrix and begin to force the two models toward a more common set of source compositions.

6   Disclaimer

The views expressed in this paper are those of the authors and not necessarily those of the U.S. Environmental Protection Agency.

References

[1]   U.S. Environmental Protection Agency National *Air Quality and Emissions Trends Report* http://www.epa.gov/airtrends/

[2]   Watson, J.G.; Zhu, T.; Chow, J.C.; Engelbrecht, J.; Fujita, E.M. and Wilson, W.E., Receptor modelling application framework for particle source apportionment. *Chemosphere* **49(9)**, pp. 1093-1136, December, 2002.

[3]   Schauer, J. J.; Fraser, M. P.; Cass, G. R. and Simoneit, B. R. T., Source Reconciliation of Atmospheric Gas-Phase and Particle-Phase Pollutants during a Severe Photochemical Smog Episode, *Environmental Science and Technology* **36(17)**, pp. 3806-3814, 2002.

[4]   Liu, W.; Hopke, P.K.; Han, Y.;Yi, S.; Holsen, T.M.;Cybart, S.; Kozlowski, K. and Milligan, M., Application of Receptor Modelling to Atmospheric Constituents at Potsdam and Stockton, NY, *Atmospheric Environment* **37**, pp 4997-5007, 2003.

[5]   Kim, E.; Larson, T.V.; Hopke, P.K.; Slaughter, C.; Sheppard, L.E. and Claiborn, C., Source identification of $PM_{2.5}$ in an arid Northwest U.S. City by positive matrix factorization, *Atmospheric Research* **66(4)**, pp. 291-305, May-June, 2003.

[6]   Watson, J.G., Cooper, J.A., Huntzicker, J.J., The effective variance weighting for least squares calculations applied to the mass balance receptor model, *Atmospheric Environment* **18**, pp. 1347-55, 1984.

[7]   Rizzo, M. and Scheff, P., Determining Potential Sources of Fine Particulate Matter using Positive Matrix Factorization for Chicago, IL. For presented at the Annual Conference of the Air and Waste Management Association, Indianapolis, IN, June 20-24, 2004.

[8]   http://www.nfraqs.colostate.edu/nfraqs/index2.html

[9]   Lee, S Win., Source Profiles of Particulate Matter Emissions from a Pilot-Scale Boiler Burning North American Coal Blends. *Journal of the Air and Waste Management Association.* **51**, pp. 1568-1578. 2001.

[10]   Watson et al., Simulating Changes in Source Profiles from Coal-Fired Power Stations: Use in Chemical Mass Balance of PM2.5 in the Mount Zirkel Wilderness, *Energy and Fuels*, **16(2)**, 2002.

[11]   Schauer, J. et al., Measurement of Emissions from Air Pollution Sources. 3. C1-C29 Organic Compounds from Fireplace Combustion of Wood, *Environmental Science and Technology,* **35(9)**, pp. 1716-1728, 2001.

[12]   Engelbrecht, J. P., et al., The Comparison of Source Contributions from Residential Coal and Low-Smoke Fuels, Using CMB Modeling, in South Africa, *Environmental Science & Policy*. **5(2)**, pp 157-167, April 2002.

[13]   Watson, J.G. et al., Annual Report for the Robbins Particulate Study. Prepared for VERSAR, Inc. by Desert Research Institute. Document No. 7100.4F3. October 1999.

[14]   Paatero, P. and Hopke, P.K., Discarding or Down weighting High-Noise Variables in Factor Analytic Models, *Analytica Chimica Acta* **490**. pp 277-289, 2003.