## SOURCE ALLOCATION AND VISIBILITY IMPAIRMENT IN TWO CLASS I AREAS WITH POSITIVE MATRIX FACTORIZATION

Keith A. Rose Senior Environmental Scientist U.S. Environmental Protection Agency, Region 10 Rose.Keith@epa.gov January 18, 2006

### **Introduction**

In the 1977 amendments to the Clean Air Act (CAA), Congress set a national goal of improving visibility in mandatory Class I Federal areas by controlling sources of visibility- impairing pollutants. In 1988 the States, Federal Land Managers, and Environmental Protection Agency (EPA) initiated the IMPROVE monitoring program to measure speciated fine particulate (PM2.5) concentrations in Class I areas, primarily in national parks and wilderness areas. The purpose of this monitoring program was to identify which pollutants are causing impairment of visibility in Class I areas, and to identify the sources responsible for these pollutants. In 1999 EPA issued the Regional Haze Rule to identify the requirements that States must meet for developing State Implementation Plans to control sources in each State that contribute to visibility impairment in any Class I area.

In this study, a statistical method known as Positive Matrix Factorization (PMF) was used to analyze IMPROVE monitoring data collected at two west coast Class I areas over two time periods, 1991-1995 and 2000-2003. These Class I areas were Mt. Rainier National Park in Washington, and Yosemite National Park in California. PMF generated source profiles associated with each source of fine particulates, and generated a timedependent series of fine particulate concentrations from each source in these two Class I areas. The light extinction (a measurement of visibility impairment) of each source was determined by summing the light extinction of all the light absorbing chemical species in each source. The average light extinction of each source for 1991-1995 and 2000-2003 were summed to determine the total light extinction in each Class I area for these time periods. The total light extinction in each Class I area in 2002 was also determined for the worst 20% visibility days, which is one of the parameters identified in the Regional Haze Rule to determine progress towards improving visibility in Class I areas. The composition of the biomass source in Yosemite was also examined to determine the relative contribution of fine particulates from biomass combustion and biogenic emissions to visibility impairment on the 20% worst visibility days in 2002. The results show that PMF can be used as a tool to help determine which sources have the most significant impact on visibility in Class I areas, and how the visibility impairment from each source varies between time periods.

### **Methods**

PMF is a variant of Factor Analysis with non-negative factor elements. It is a factor analysis method with individual weighting of matrix elements first described by Paatero and Tapper, and Paatero (1997). The PMF approach can be used to analyze 2-dimensional and 3-dimensional matrices. The 2-demensional version of PMF was used to analyze the Class I area data in this study. PMF solves the equation:

## $\mathbf{X} = \mathbf{G}\mathbf{F} + \mathbf{E}$

In this equation, "X" is the matrix of measured values, "G" and "F" are the factor matrices to be determined, and "E" is the matrix of residuals, the unexplained part of "X". In PMF, the solution is a weighted Least Squares fit, where the known standard deviations for each value of "X" are used for determining the weights of the residuals in matrix "E". The objective of PMF is to minimize the sum of the weighted residuals. PMF uses information from all samples by weighting the squares of the residuals with the reciprocals of the squares of the standard deviations of the data values.

In environmental pollution problems, one row of "X" would consist of the concentrations of all chemical species in one sample, and one column of "X" would be the concentration of one species for each of the samples. One row of the computed "F" matrix would be the source profile for one source, and the corresponding column of "G" would be the amount of this source in each individual sample. Required input matrices for PMF are "X", the measured values, and "X<sub>std-dev</sub>", the standard deviations (uncertainties) of the measured values. PMF requires that all values and uncertainties are positive values, therefore missing data and zero values must be omitted or replaced with appropriate substitute values.

#### **PMF Operating Parameters**

For analysis of the IMPROVE data, PMF was run in the robust mode suggested for analyzing environmental data by Paatero (1996). In the robust mode, the standard deviations used for weighting the residuals are dynamically readjusted through an iterative process. This process prevents excessively large values in the data set from disproportionally affecting the results. PMF provides error models to calculate the standard deviations of the data values. According to Paatero (1996), recommended error models for environmental data include the lognormal distribution model and the heuristically-computed model. The lognormal model works well if the data have a lognormal distribution, but that is not always the case for environmental data. In this study the heuristically-computed model was chosen for analysis of IMPROVE data.

#### Adjustment of PMF Source Concentrations

In this study, the daily PMF calculated concentrations for each source (G matrix) were adjusted through a linear regression with the measured total concentrations. The linear regression was accomplished by using the "LINEST" function in Excel. This function provides three parameters that indicate the "goodness of fit" between the measured concentration and the sum of the calculated concentrations. These parameters

are " $r^2$ ", the slope of the regression line, and the uncertainty in each source regression adjustment factor. The best fit is achieved when the regression parameters " $r^2$ " and "slope" each equal 1.0, and the uncertainty in each regression factor is smaller than the value of the corresponding regression factor.

## Data Selection

Data used for each Class I area in this analysis were from the years 1991-95 and 2000-03. Dates that had missing data, and species that had substantial values below the laboratory minimum detection limit (MDL), were eliminated from this analysis. Species used in this analysis included: calcium, copper, elemental carbon fractions (EC1 and EC2), iron, potassium, hydrogen, sodium, lead, organic carbon fractions (OC2, OC3 and OC4), nitrate, sulfate, sulfur, silicon, and zinc. Data and data uncertainties reported as "zero" by the laboratory were replaced with a value of ½ the MDL.

## **Results and Discussion**

## Determining the Number of Sources

The most difficult challenge in using PMF to evaluate environmental data is determining the number of sources that are contributing to the contaminants collected at the monitor. In this study, five, six and seven-source solutions were generated for both Class I areas. A two-step process was used to determine which solutions generated by PMF provided the most feasible number of sources for each Class I area. First, the generated source profiles were compared to source profiles identified in previous published PMF studies. Specifically, the source profiles for each solution (F matrices) were compared to the Columbia Gorge PMF source profiles (Rose) and to those identified by the PMF analysis of Seattle IMPROVE data (Maykut et. al.). Second, the "goodness of fit" for each solution was examined to see which solutions had the best linear regression between measured and calculated source concentrations. The results of this two-step process, to identify source profiles and determine the "goodness of fit", are shown in Table 1.

Site	Time	# Sources	$r^2$	Slope	Source Profiles
	Period				
Mt. Rainier	1991-95	6	0.95	0.83	All Identified
Mt. Rainier	1991-95	7	0.95	0.81	All Identified
Mt. Rainier	2000-03	6	0.94	0.60	All Identified
Mt. Rainier	2000-03	7			Unidentified Profiles
Yosemite	1991-95	6	0.96	0.68	All Identified
Yosemite	1991-95	7	0.96	0.68	All Identified
Yosemite	2000-03	5	0.97	0.67	All Identified
Yosemite	2000-03	6			Unidentified Profiles
Yosemite	2000-03	7			Unidentified Profiles

## Table 1. Evaluation of PMF Solutions for Mt. Rainier and Yosemite National Parks

These results show that the six-source solutions for Mt. Rainier for 1991-1995 and 2000-2003 generated acceptable results. The seven-source solution for Mt. Rainier for 1991-1995 was also acceptable, while the seven-source solution for 2000-2003 contained unidentified source profiles. For Yosemite, the six and seven-source solutions for 1991-1995 were acceptable, while only the five-source solution was acceptable for 2000-2003. In all cases where there were acceptable results from multiple solutions, the solution with the higher number of sources always contained a diesel-powered mobile source profile and a gasoline-powered mobile source profile. In order to directly compare the total mobile (combined diesel and gasoline) source contributions between the two time periods, only those solutions which contained a combined mobile source profile were used for further analysis in this study.

## Identification of Source Profiles

PMF source profiles for each Class I area for are shown in Appendix A. PMF generated four source profiles for each Class I area that had relatively small amounts of organic or elemental carbon and contained significant amounts one or more inorganic species. These source profiles were similar to non-combustion source profiles generated by PMF analysis of Columbia Gorge and Seattle IMPROVE data. The inorganic species in each profile and its associated source are shown in Table 2.

Profile species	Source		
Sulfate	Secondary sulfate		
Nitrate	Secondary nitrate		
Silicon, Fe, K and Ca	Soil		
Sodium	Marine aerosols		

#### Table 2. PMF Inorganic Profiles and Associated Sources

For each Class I area, PMF also generated two profiles with high amounts of organic and elemental carbon (Table 3) similar to the biomass and mobile source profiles identified by PMF analysis of Columbia Gorge and Seattle IMPROVE data. The mobile (gasoline and diesel) source profiles contain the highest amounts of EC, moderate amounts of OC, and trace amounts of iron, lead and zinc. The biomass profiles contain the highest amounts of organic carbon, a large OC3 fraction, relatively smaller amounts of EC1 and EC2, and trace amounts of potassium.

## Table 3. PMF High-Carbon Profiles and Associated Sources

<b>Profile Species</b>	Source
OC, EC, K	Biomass
OC, EC, Pb, Zn, and Fe	Mobile sources

## Source Contributions to PM2.5 Concentrations

The average and 90 percentile daily PM2.5 concentrations from each source in each Class I area, for both the 1991-1995 and 2000-2003 time periods, are shown in Tables 4 and 5. Biomass contributed the largest amount of fine particulates in both Class

I areas for both time periods. However, between the two time periods, average biomass concentrations at Mt. Rainier decreased by 42%, while biomass concentrations at Yosemite increased 27%. Secondary sulfate contributed the second highest amount of fine particulates at both Class I areas. Between 1991-1995 and 2000-2003, average concentrations of the secondary sulfate source decreased by 37% at Mt. Rainier, and by 23% at Yosemite.

	1991-95	1991-95	2000-03	2000-03
Source	Average	90 Percentile	Average	90 Percentile
Biomass	2.31	5.14	1.34	3.0
Secondary Sulfate	1.62	4.0	1.02	2.36
Secondary Nitrate	0.42	1.01	0.28	0.63
Mobile Sources	0.51	1.06	0.34	0.82
Soil	0.46	1.21	0.44	1.07
Marine	0.35	0.79	0.27	0.6

## Table 5. PM2.5 Concentration by Source at Yosemite (ug/m<sup>3</sup>)

Source	1991-95 Average	1991-95 90 Percentile	2000-03 Average	2000-03 90 Percentile
Biomass	1.70	3.17	2.16	5.41
Secondary Sulfate	1.10	2.29	0.85	1.96
Secondary Nitrate	0.62	1.46	0.61	1.39
Mobile Sources	0.30	0.57	0.36	0.64
Soil	0.65	1.35	0.64	1.26
Marine	0.27	0.61	n/a*	n/a*

\* The Yosemite 2000-03 solution did not include a marine source.

## Time-Dependent Source Concentrations

Trends in source concentrations for both Class I areas for the 1991-1995 time period are shown in Appendix B. At Mt. Rainier, secondary sulfate, secondary nitrate, soil and marine sources showed seasonal trends. At Yosemite, biomass, secondary sulfate, secondary nitrate, soil and marine sources showed seasonal trends. Months of the year during which each source made its highest contribution at each site are shown in Table 6.

## Table 6. Months of Highest Source Contribution

Source	Mt. Rainier NP	Yosemite NP	
Biomass	October-March	May-October	
Secondary Sulfate	April-October	April-October	
Secondary Nitrate	April-October	March-November	
Soil	March-October	April-October	
Marine	March-October	May-September	
Mobile Sources	No pattern	No pattern	

#### Source Visibility Impairment

Visibility impairment caused by fine particles, expressed in terms of the light extinction coefficient Bext (units of inverse megameters, 1/Mm), is given in equation 3.8 in Chapter 3 of the report titled "Spatial and Seasonal Patterns and Temporal Variability of Haze and its Constituents in the United States: Report III" (Malm):

 $Bext = (3 m^{2}/g) Ft(RH)[sulfate] + (3 m^{2}/g)Ft(RH)[nitrate] + (4 m^{2}/g)[OC] + (10 m^{2}/g)[EC] + (1 m^{2}/g)[soil]$ 

Where: Ft(RH) = annual average relative humidity factor

The Bext for the secondary sulfate and secondary nitrate sources were determined by assuming that these sources consisted only of ammonium sulfate and ammonium nitrate, respectively. The Bext for the biomass and mobile sources were determined by assuming that the only visibility impairing components in these sources were OC and EC. Ft(RH) for Mt. Rainier was set at a value of 4.5, and the Ft(RH) for Yosemite was set at a value of 2.1. Using this approach, the average and 90 percentile Bext due to each source, based on average and 90 percentile concentrations of each source (tables 4 and 5), are shown in Tables 7 and 8.

#### Table 7. Average Source Bext (1/Mm)

	Mt. Rainier	Mt. Rainier	Yosemite	Yosemite
Source	1991-95	2000-03	1991-95	2000-03
Biomass	12.7	7.4	8.5	10.8
Secondary Sulfate	15.9	10.0	6.9	3.9
Secondary Nitrate	4.4	2.9	3.9	3.0
Mobile Sources	3.6	2.4	2.2	2.6
Soil	0.5	0.5	0.6	0.6

### Table 8. 90 Percentile Source Bext (1/Mm)

	Mt. Rainier	Mt. Rainier	Yosemite	Yosemite
Source	1991-95	2000-03	1991-95	2000-03
Biomass	28.26	16.57	15.85	27.05
Secondary Sulfate	39.26	23.14	14.36	8.99
Secondary Nitrate	10.58	6.53	9.18	6.84
Mobile Sources	7.48	5.79	4.18	4.62
Soil	1.32	1.22	1.25	1.18

Figures 1 through 4 show the average percent of visibility impairment due to each source, relative to the total visibility impairment for all sources, in each Class I area for the 1991-1995 and 2000-2003 time periods.

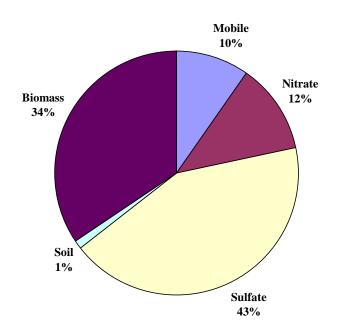
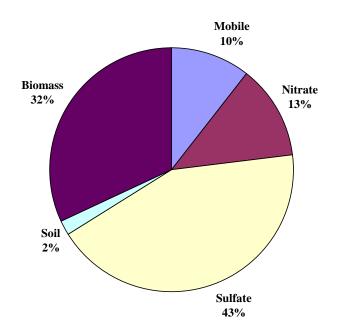


Figure 1. Percent Source Visibility Impairment at Mt. Rainier for 1991-1995

Figure 2. Percent Source Visibility Impairment at Mt. Rainier for 2000-2003



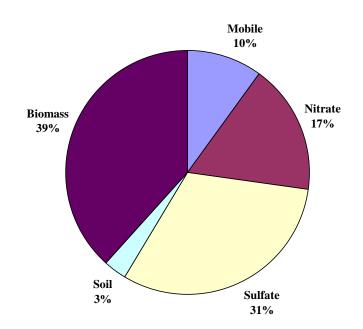
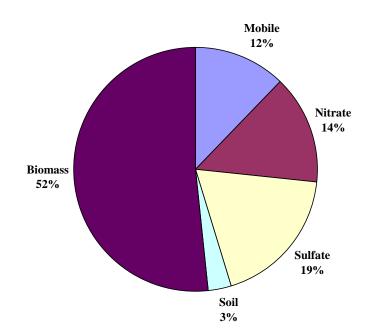


Figure 3. Percent Source Visibility Impairment at Yosemite for 1991-1995

Figure 4. Percent Source Visibility Impairment at Yosemite for 2000-2003



Figures 1 and 2 show that that the most significant source of visibility impairment at Mt. Rainier is secondary sulfate, and that the second most significant source is biomass. These figures also show that the relative percents of visibility impairment from all sources at Mt. Rainier remained about the same for both 1991-1995 and 2000-2003. Figures 3 and 4 show that the most significant source of visibility impairment at Yosemite was biomass, and that the second largest source was secondary sulfate. Figures 3 and 4 also show that the relative percent of visibility impairment due to biomass substantially increased at Yosemite between 1991-1995 and 2000-2003, while the relative percent due to secondary sulfate substantially decreased between these two time periods. Visibility impairment due to secondary nitrate was the third largest of all sources at both Class I areas, and visibility impairment due to mobile sources was the fourth largest.

Figures 5 and 6 show the average visibility impairment of all sources at Mt. Rainier and Yosemite for the 20% best and worst visibility days in 2002. Figure 5 shows that an average of 44% of the visibility impairment on the 20% worst days at Mt. Rainier was due to fine particulates from secondary sulfate sources, and 27% of the visibility impairment was due to fine particulates from biomass sources. Figure 6 shows that an average of 66% of the visibility impairment on the 20% worst days at Yosemite was due to fine particulates from biomass sources.

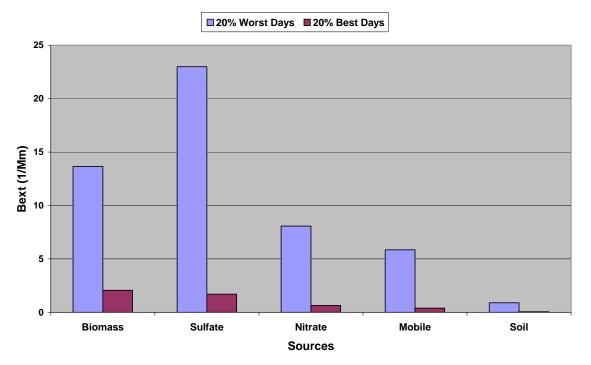


Figure 5. Bext on 20% Worst and 20% Best Days - Mt. Rainier 2002

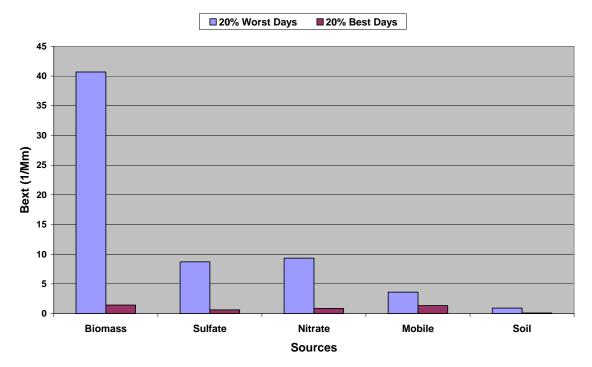


Figure 6. Bext on 20% Worst and 20% Best Days -Yosemite 2002

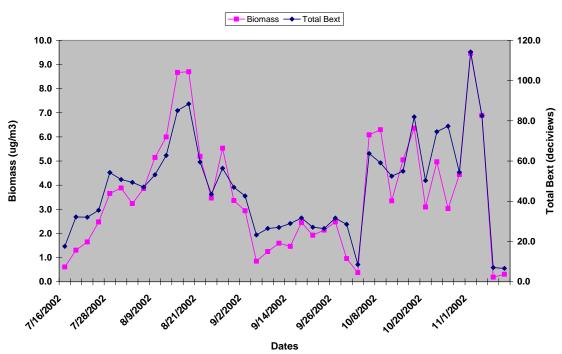
## Composition of the Yosemite Biomass Source

For the 2002 fire season in Yosemite (July-October), fine particulates from biomass sources were responsible for most of the visibility impairment on the 20% worst visibility days. This is shown in Figure 7 where total daily Bext is closely correlated with the daily PMF biomass concentrations. The question remains whether visibility impairment caused by biomass particulates during this period was caused primarily by biomass smoke or by secondary organic aerosols (SOAs) produced by oxidation of biogenic emissions of volatile organic compounds (VOCs).

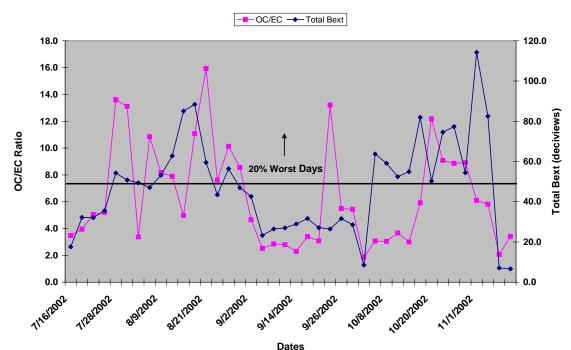
Alessio et. al. demonstrated that aged smoke plumes that travel long distances contain significant amounts of SOAs associated with oxidation of biogenic VOCs associated with fires in the Mediterranean. This was also demonstrated in a study conducted in Yosemite National Park by Engling et. al. In this study, called the Yosemite Aerosol Characterization Study (YACS), organic tracers for biomass combustion (smoke) and tracers for SOAs were collected in Yosemite during the period of July 14, 2002, to September 5, 2002. Biomass smoke tracers, including anhydrosugars, methoxyphenols, and resin acids, were used to determine contributions of primary biomass smoke to PM2.5. To determine the contribution of SOAs to organic carbon aerosol, monoterpene oxidation products and other organic compounds of secondary origin, such as dicarboxylic acids, were also measured. The results of this study showed that in addition to several local wildfires and prescribed burns, two regional haze episodes at Yosemite were strongly influenced by smoke from biomass that was subject to long-range transport. These long-range transport plumes contained a combination of primary

biomass combustion products and SOAs. SOAs contributed about 80% of the total fine particulate OC concentration during these long-range transport events.

During the periods of time when the organic aerosols measured in the YACS were attributed to long-range transport and were mostly composed of SOAs, the average OC/EC ratio of fine particulates measured was about 9.9. During the period when smoke from local wildfires was the primary source of organic aerosols, 65% of organic aerosols consisted of biomass smoke products and the average OC/EC ratio was 3.6. This indicates that fine particulates mostly composed of SOAs have relatively high OC/EC ratios, whereas fine particulates composed mostly of biomass smoke products have low OC/EC ratios. In order to estimate the relative contribution of SOAs and biomass smoke to visibility impairment on the 20% worst visibility days during the Yosemite 2002 fire season, the daily OC/EC concentration ratio for each day during this period was plotted against total Bext for each day (Figure 8). Figure 8 shows that on about half of the 20% worst visibility days the OC/EC ratio ranged from 7 to 16, and on the other half of the days the OC/EC ratio ranged from 3 to 6. This indicates that there was approximately an equal contribution of SOAs and biomass smoke to fine particulates on the 20% worst days.







#### Figure 8 - Yosemite OC/EC Ratios vs Total Bext

### **Conclusions**

PMF generate source profiles for biomass, secondary sulfate, secondary nitrate, mobile sources, soil and marine aerosols that contribute to fine particulate concentrations measured at Mt. Rainier and Yosemite National Parks. The trends in these source concentrations were also identified. At Mt. Rainier, secondary sulfate, secondary nitrate, soil and marine aerosols showed seasonal trends. At Yosemite, biomass, secondary sulfate, secondary nitrate, soil and marine aerosols showed seasonal trends. At Yosemite, biomass, secondary sulfate, secondary nitrate, soil and marine aerosols showed seasonal trends. Biomass was responsible for the highest average concentrations of fine particulates in both Class I areas for the 1991-1995 and 2000-2003 time periods, and the second highest concentrations were due to secondary sulfate. At Mt. Rainier, average concentrations of particulates due to biomass and secondary sulfate decreased between 1991-1995 and 2000-2003. At Yosemite, average concentrations due to biomass increased between these two time periods, while concentrations of secondary sulfate decreased.

Average and 90 percentile source concentrations were used to determine the average and 90 percentile visibility impairment due to each source for both the 1991-1995 and 2000-2003 time periods. At Mt. Rainier, the largest source of visibility impairment was secondary sulfate, and the second largest source was biomass. At Yosemite, the largest source of visibility impairment was biomass, and the second largest source was secondary sulfate. At Mt. Rainier, between the periods of 1991-1995 and 2000-2003, average visibility impairment due to secondary sulfate decreased from a Bext of 15.9 1/Mm to 10.0 1/Mm, and average visibility impairment due to biomass decreased from 12.7 1/Mm to 7.4 1/Mm. At Yosemite, between the periods of 1991-1995 and

2000-2003, average visibility impairment due to biomass increased from a Bext of 8.5 1/Mm to 10.8 1/Mm, and average visibility impairment due to secondary sulfate decreased from 6.9 1/Mm to 3.9 1/Mm. Visibility impairment due to secondary nitrate was the third largest of all sources at both Class I areas, and visibility impairment due to mobile sources was the fourth largest. On the 20% worst visibility days in 2002, 44% of the visibility impairment on at Mt. Rainier was due to fine particulates from secondary sulfate sources, and 27% of the visibility impairment was due to biomass sources. For Yosemite, 66% of the visibility impairment on the 20% worst days was due to fine particulates from biomass sources. Analysis of the OC/EC ratios on the 20% worst visibility days in Yosemite in 2002, in which visibility impairment was mostly attributed to biomass particulates, indicates that there was approximately an equal contribution from SOAs and smoke to these biomass particulates. Most likely, the SOAs measured during this period were associated with long-range transport of smoke plumes from wildfires.

## **References**

Alessio, G. A., Lillis, M. d., Fanelli, M., Pinelli, P. and Loreto, F. Direct and Indirect Impacts of Fire on Isoprenoid Emissions from Mediterranean Vegetation. Functional Ecology 18, 357-364, 2004.

Chow, J., and Watson, J. Western Washington 1996-97 PM2.5 Source Apportionment Study, 1998.

Engling, G., Herckes, P., Kreidenweis, S. M., Malm, W. C., and Collett, J. L. Composition of Fine Organic Aerosol in Yosemite National Park During the 2002 Yosemite Aerosol Characterization Study. Atmospheric Environment, submitted.

Kim E., Hopke, P. K., Larson, T. V., Maykut, N. N., and Lewtas, J. Factor Analysis of Seattle Fine Particles. Aerosol Sci. Technol. 38 (7): 724-738, 2004.

Malm, W. C. Spatial and Seasonal Patterns and Temporal Variability of Haze and its Constituents in the United States: Report III. CIRA, ISSN: 0737-5352-47, May 2000.

Maykut, N. N., Lewtas, J., Kim, E., and Larson, T. V. Source Apportionment of PM2.5 at an Urban IMPROVE Site in Seattle Washington. Environmental Science and Technology, 2003, 37, 5135-5142.

Paatero, P., and Tapper, U. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5, 111-126, 1994.

Paatero, P. User's Guide for Positive Matrix Factorization Programs PMF2.EXE and PMF3.EXE, University of Helsinki, Helsinki, 1996.

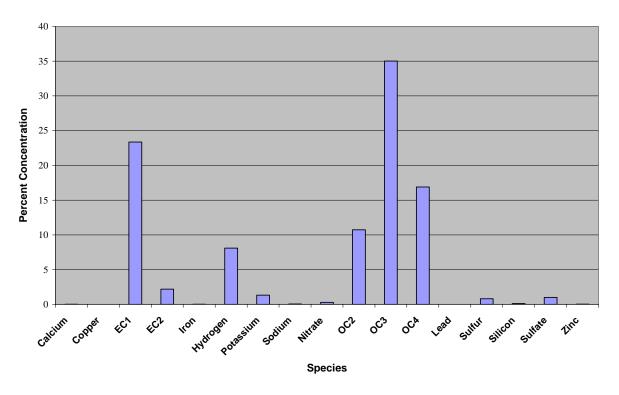
Paatero, P. Least squares formulation of robust non-negative factor analysis. Chemometrics and Intelligent Laboratory Systems, 37, 23-35, 1997.

Rose, K. Source Allocation of Columbia Gorge IMPROVE Data with Positive Matrix Factorization, Appendix E of "Chemical Concentration Balance Source Apportionment of PM2.5 Aerosol in the Columbia River Gorge". Oregon Department of Environmental Quality Report, March 31, 2003.

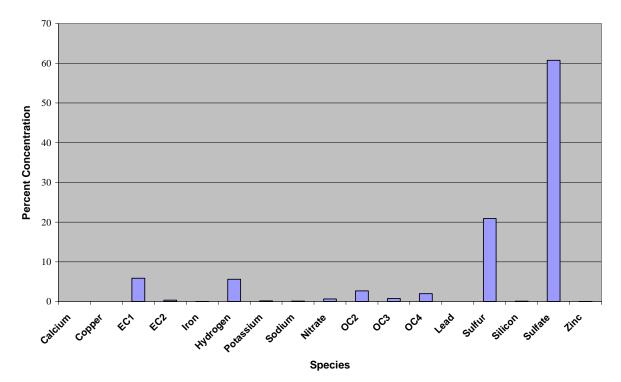
# APPENDIX A

## SOURCE PROFILES

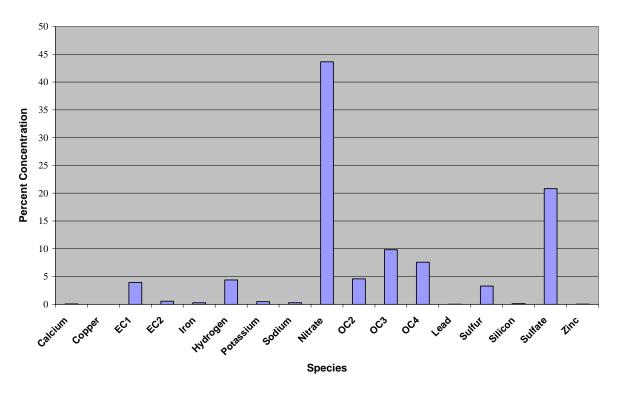
Biomass Burning Profile - Mt. Rainier 1991-95

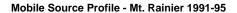


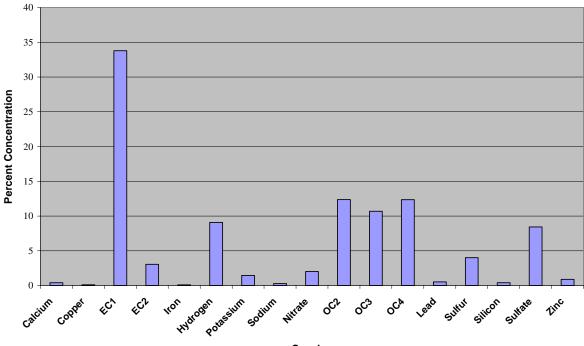




Secondary Nitrate Profile - Mt. Rainier 1991-95

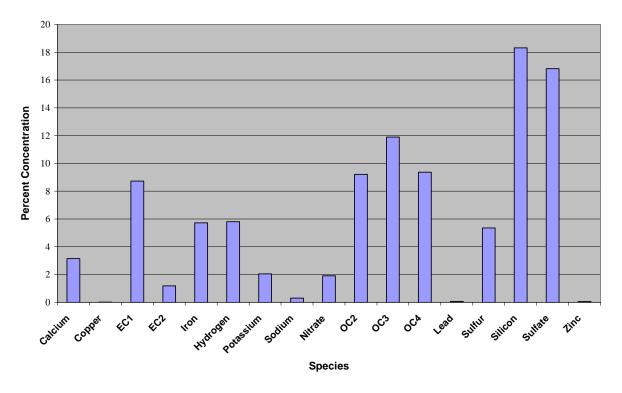




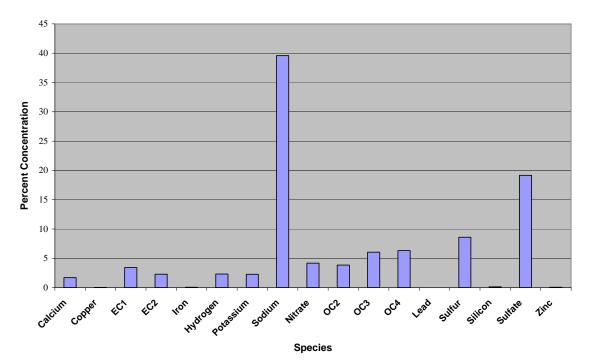


Species

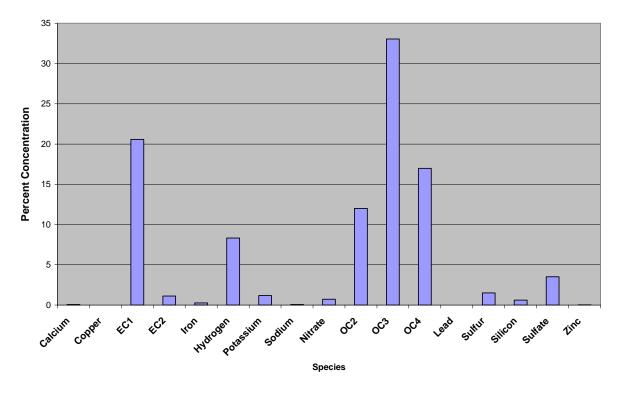
#### Soil Profile - Mt. Rainier 1991-95



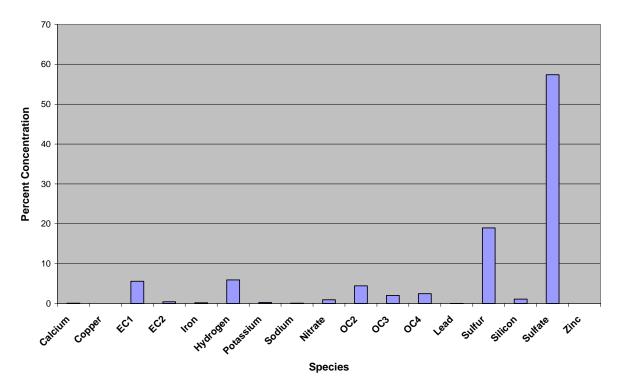




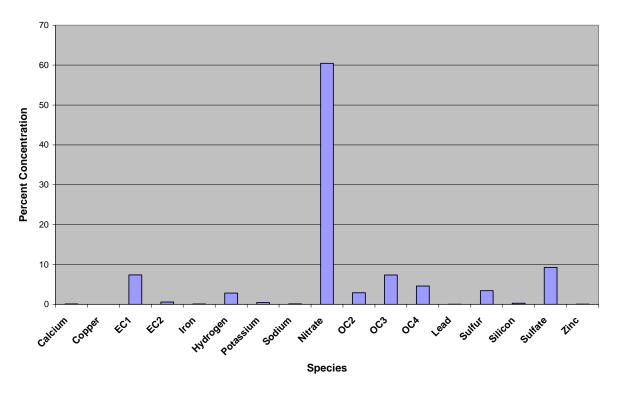
#### **Biomass Burning Profile - Yosemite 1991-95**



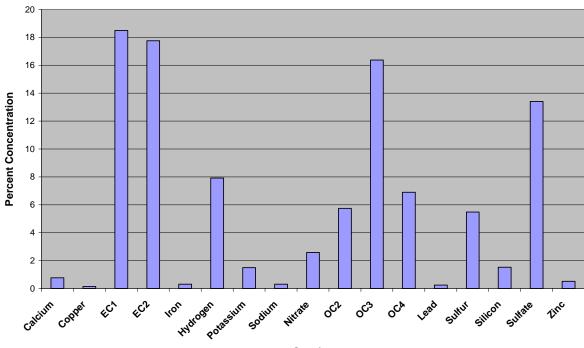




Secondary Nitrate Profile - Yosemite 1991-95

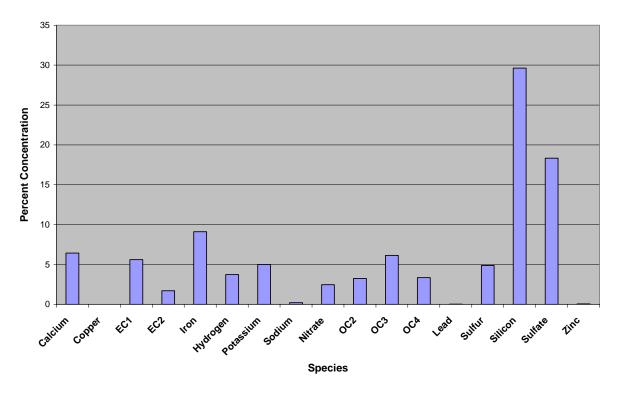


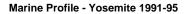


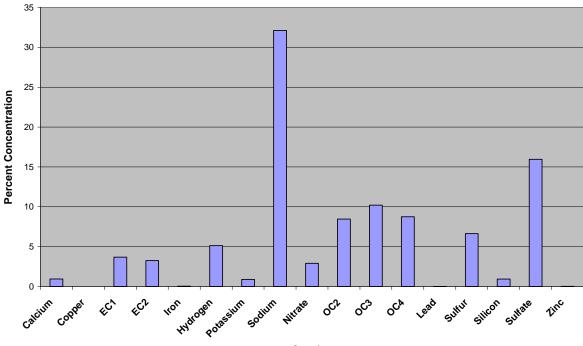


Species

#### Soil Profile - Yosemite 1991-95





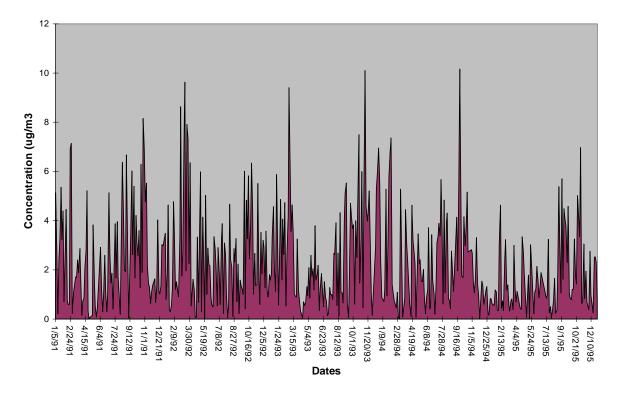


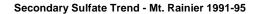
Species

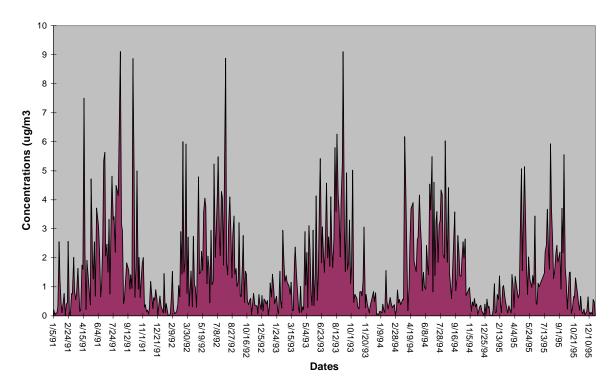
# APPENDIX B

# TIME DEPENDENT SOURCE CONCENTRATIONS

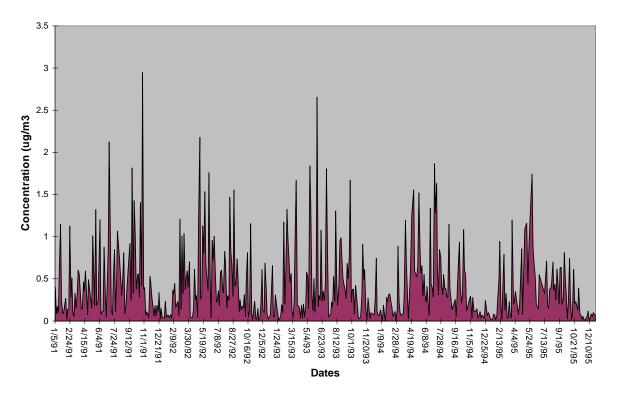
Biomass Source Trend - Mt. Rainier 1991-95

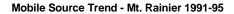


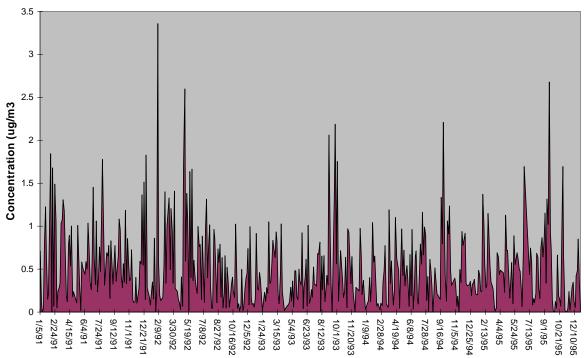




Secondary Nitrate Trend - Mt. Rainier 1991-95

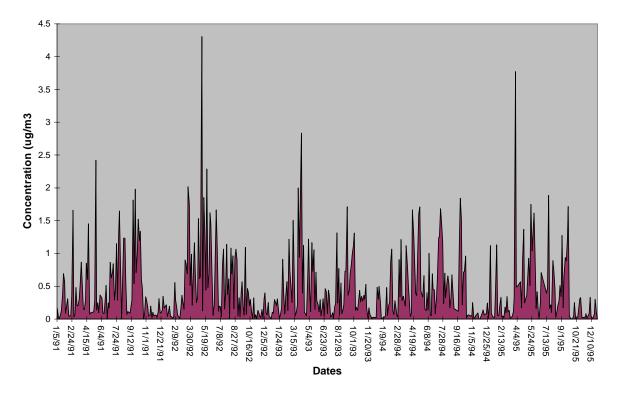


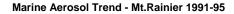


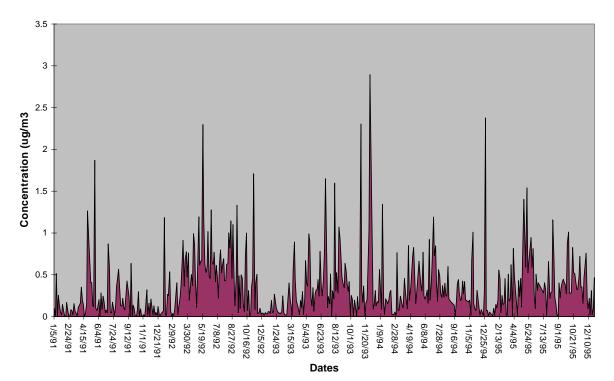


Dates

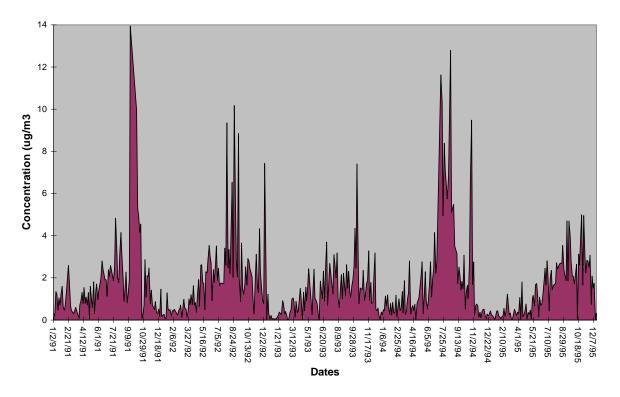
Soil Trend - Mt. Rainier 1991-95



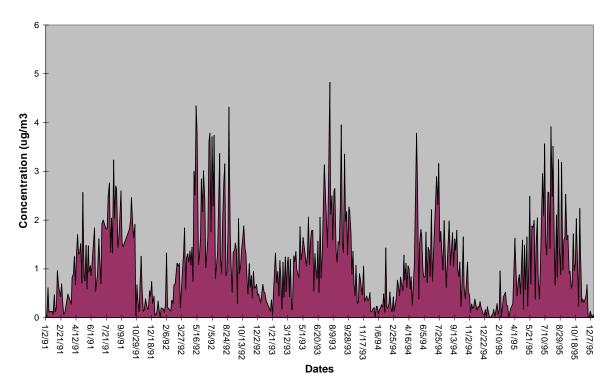




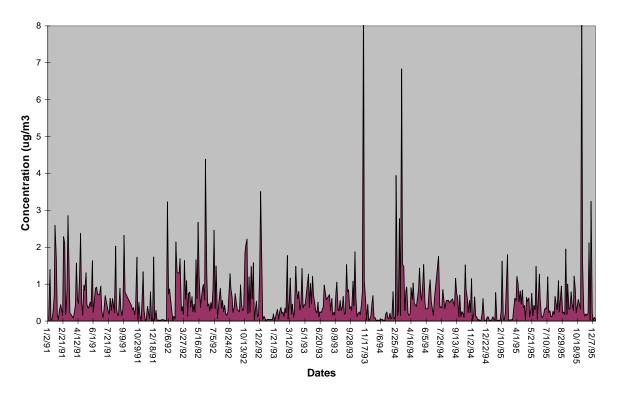
Biomass Source Trend - Yosemite 1991-95

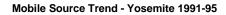


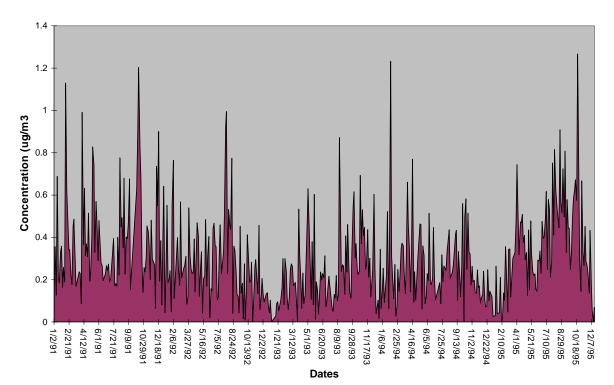




Secondary Nitrate Trend - Yosemite 1991-95







Soil Trend - Yosemite 1991-95

