

# **SOURCE ALLOCATION OF COLUMBIA GORGE IMPROVE DATA WITH POSITIVE MATRIX FACTORIZATION**

Keith Rose, Air Monitoring Specialist  
U.S. Environmental Protection Agency, Region 10  
November 15, 2004

## **Introduction**

Positive Matrix Factorization (PMF) was used to model IMPROVE monitoring data collected at the Mt. Zion and Wishram monitoring sites in the Columbia River Gorge National Scenic Area. PMF was able to identify source categories that contribute to the fine particulates (PM<sub>2.5</sub>) measured at each site, the average percent PM<sub>2.5</sub> mass allocation from each source category, and the time-dependent allocation from each source category. PMF identified 9 source categories for the Mt. Zion site based on 1996-98 data, and, 9 source categories for the Wishram site based on 1993-95 data. The results of this study was compared to the PMF analysis of the 1996-98 Columbia Gorge IMPROVE data presented in "Chemical Mass Balance Source Apportionment of PM<sub>2.5</sub> Aerosol in the Columbia River Gorge" (Kuhns et. al.).

## **Methods**

### **PMF Model**

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-dimensional version of PMF (PMF2) was used to analyze the Columbia Gorge data. PMF2 solves the equation:

$$\mathbf{X} = \mathbf{GF} + \mathbf{E}$$

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 the PMF model, 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.

### Model Operating Parameters

For analysis of the Columbia Gorge IMPROVE data, PMF was run in the robust mode suggested for analyzing environmental data (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 disproportionately affecting the results.

PMF provides several 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 the PMF analysis of particulate data collected in Hong Kong, Lee et. al. achieved good results with the heuristically-computed error model. In this study the heuristically-computed model was chosen for analysis of the Columbia Gorge data.

### Source Allocation

In this study the PMF calculated masses for each source were adjusted through a linear regression of the measured mass and calculated mass similar to that used by Maykut et. al. Percent source mass allocation was determined by dividing the adjusted mass of each source by the adjusted total mass of all sources. The linear regression was accomplished by using the “LINEST” function in Excel. This function provides three parameters that indicate the “goodness of fit” of the regression. These parameters are “ $r^2$ ”, the slope of the regression line, and the uncertainty in each source regression 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.

### Determining the Number of PMF Factors

One of the most important model parameters to determine when using PMF is the number of factors (sources) which PMF will use to model the data. The Chemical Mass Balance (CMB) analysis of Columbia Gorge data performed by Kuhns et. al. identified seven source categories contributing to the fine particulates measured at the Mt. Zion and Wishram sites. The source categories were ammonium sulfate (secondary sulfate), ammonium nitrate (secondary nitrate), soil, marine aerosols, vegetative burning, aluminum smelters and motor vehicles. An emission inventory of sources impacting air quality in the Columbia Gorge compiled by the states of Washington and Oregon showed that pulp mill emissions are a significant source of  $PM_{2.5}$  in the Columbia Gorge. CMB was not able to resolve a pulp mill source in the Columbia Gorge but PMF identified a pulp mill source at Mt. Zion (Kuhns et. al.). In the same report, PMF was also able to resolve mobile combustion sources into gasoline-powered and diesel-powered mobile sources at the Mt. Zion site. If a pulp mill source is added to the seven Columbia Gorge sources identified by CMB, and mobile sources could be resolved into gasoline-powered

and diesel-powered mobile sources, this would result in a total of nine possible source categories that could be identified by PMF. Therefore, for this analysis, the number of factors (sources) used for the PMF model was varied between eight and ten.

To determine the number of sources that gave the best solution for the Mt. Zion and Wishram data, the following method was used. First, source profiles (F matrices) generated by each run were compared to CMB profiles for the Columbia Gorge (Kuhns et. al.), and to PMF profiles generated for Seattle IMPROVE data (Maykut et. al.) to identify the most physically reasonable source profiles. Second, the PMF “Q” function of each solution was checked to see if its value approximately equaled the number of elements in the “X” matrix minus the number of elements in the “G” matrix (Paatero, 1996). Third, the “goodness of fit” regression parameters for each resolution were examined to see which solution achieved the best “fit”.

### Data Selection

Data used for this analysis were the IMPROVE PM<sub>2.5</sub> speciated data from the Mt. Zion and Wishram monitoring sites in the Columbia Gorge National Scenic Area. Data and data uncertainties reported as “zero” were replaced with a value of ½ Minimum Detection Limit (MDL). Dates that had missing data, and species that had a substantial number of values below the laboratory MDL, were eliminated from this analysis. Species retained for this analysis included Al, Br, Ca, Cu, EC1, EC2, Fe, K, H, Pb, Na, OC1, OC2, OC3, OC4, nitrate, sulfate, S, Si, and Zn. The 1996-98 data set for Mt. Zion consisted of 198 sampling days and the 1996-98 data set for Wishram included 303 sampling days. The 1993-95 Wishram data set, which was analyzed, consisted of 261 sampling days.

## **Results and Discussion**

### Identification of Source Profiles

PMF solutions for eight, nine, and ten sources were generated for Mt. Zion and Wishram for the 1996-98 data. The eight-source solution for Mt. Zion was rejected because it generated a soil profile that contained 12% aluminum, which is too high for soil in western Washington. The eight-source Wishram solution was rejected because it generated mobile source profiles that did not match those identified by Maykut et. al. The 10-source solutions for Mt. Zion and Wishram were rejected because they each generated a combustion-like profile that could not be identified.

The difference between the value of the “Q” function and the corresponding number of elements for the X-G matrices for each Mt. Zion and Wishram solution are shown in Table 1. The lower values for the 9 and 10 source solutions indicate that the 8-source solutions are not optimal. The goodness of fit for each solution is shown in Table 2. The  $r^2$  and slope parameters for the 9-source and 10-source Mt. Zion solutions were equal. Base on goodness of fit, minimization of the “Q” function, and similarities to source profiles in the literature, the best solution for Mt. Zion was a 9-source solution.

**Table 1.** Q Function Value minus X-G Elements for Each Solution

Number of Factors	Mt. Zion	Wishram
8	804	626
9	517	386
10	360	317

**Table 2.** Linear Regression “Goodness of Fit”

Number of Factors	Mt. Zion 1996-98 Data (r <sup>2</sup> /slope)	Wishram 1993-95 Data (r <sup>2</sup> /slope)	Wishram 1996-98 (r <sup>2</sup> /slope)
8	Rejected*	0.94/0.97	Rejected*
9	0.91/1.0	0.94/0.99	Rejected*
10	0.91/1.0	0.94/0.98	Rejected*

\* Solution contained a regression uncertainty greater than the regression value for one source.

All the PMF solutions for the Wishram 1996-98 data were rejected because they contained a regression uncertainty that was greater than the corresponding regression value for one source. As a substitute for the 1996-98 Wishram data, the 1993-95 Wishram data were analyzed. PMF analysis of the 1993-95 Wishram data generated viable solutions for the 8, 9, and 10 sources. Based on the characteristics of the source profiles and the “goodness of fit”, the best solution for Wishram was a 9-source solution. The nine source profiles generated by PMF for Mt. Zion and Wishram are shown in Appendix A.

For both sites, PMF generated four source profiles that had relatively small amounts of organic or elemental carbon and contained significant amounts of one or more inorganic species. These source profiles were similar to non-combustion source profiles used in the CMB analysis of Columbia Gorge data. The inorganic species in each of these profiles, and their associated sources, are shown in Table 3.

**Table 3.** Inorganic PMF Source Profiles and Associated Sources

Profile inorganic species	Source
Sulfate and sulfur	Secondary sulfate
Nitrate	Secondary nitrate
Silicon, iron, potassium and calcium	Soil
Sodium and bromine	Marine aerosols

For both sites, PMF generated five profiles associated with combustion sources that had relatively large amounts of organic and elemental carbon, and contained significant amounts of one or more inorganic species. The organic and inorganic composition of each profile and its potential source are shown in Table 4.

**Table 4.** PMF Combustion Source Profiles and Potential Sources

<b>Profile Composition</b>	<b>Potential Source</b>
OC, EC, K	Vegetative Burning
OC, EC, lead, zinc	Gasoline-powered sources
OC, EC, iron	Diesel-powered sources
OC, EC, Al, Fe	Aluminum Smelters
OC, EC, K, Na, Fe	Pulp Mills

The vegetative burning profiles could be readily identified because they contained the highest amount of organic carbon, a large OC3 fraction, small amounts of EC1 and EC2, and potassium. Of all the profiles, the vegetative burning profiles had the highest OC/EC ratios. These profiles were also similar to the composite burn profile used in the Columbia Gorge CMB analysis.

The diesel-powered source profiles could be identified because they contained the highest amount of EC (36-43%), moderate amount of OC (25-30%), an EC/OC mass ratio of about 1.4, and a trace amount of iron. The gasoline-powered source profiles were identified by a high amount of OC (35-43%), smaller amount of EC (16-17%), an EC/OC mass ration of about 0.4, and trace amounts of lead and zinc.

The aluminum smelter source profiles were identified by a species composition indicative of smelters including a high amount of aluminum (17-20%), high amount of OC (32-33%), lower amount of EC (10-20%), and a trace amount of iron (2.0-2.7%). The pulp mill source profiles were identified by a species composition indicative of pulp mills including a large amount of OC (35-39%), and small amounts of potassium (3.2-4.4%), Fe (1.4-1.6%), and sodium (1.7-3.2%). The pulp mill profile for Mt. Zion also had a high amount of sulfur (17%), which is indicative of pulp mills.

#### Source Allocation

PMF percent mass allocations to each source for the 1996-98 Mt. Zion data and for the 1993-95 Wishram data are shown in Table 5. The vegetative burning and secondary sulfate sources were the largest two sources at the Mt. Zion and Wishram sites and had almost equal allocations at each site. Equal allocations of secondary sulfate (ammonium sulfate) between the two sites is reasonable since the conversion of SO<sub>2</sub> to sulfate is dependent on ambient levels of moisture and oxidants, and typically occurs on a time scale of days to weeks. Over this time period, plumes from point sources disperse over large regions resulting in a uniform blanket of sulfate aerosols.

The secondary nitrate (ammonium nitrate) allocation was about 90% larger at Wishram than at Mt. Zion. Higher levels of ammonium nitrate at Wishram could be due to NO<sub>2</sub> emitted by the Portland Gas and Electric power plant located at Boardman, Oregon. During the period of 1993-95, the Boardman power plant emitted an average of 7,000 tons of NO<sub>2</sub> per year. The NO<sub>2</sub> from the Boardman power plant, reacting with ammonium released from surrounding animal feed lots, would have created ammonium nitrate that would have been detected at the Wishram site when winds were blowing from the east.

**Table 5.** Source Allocation (Percent Total PM<sub>2.5</sub> Mass)

<b>Source</b>	<b>Mt. Zion</b>	<b>Wishram</b>
Vegetative Burning	24.1	24.0
Ammonium Sulfate	24.2	24.4
Ammonium Nitrate	7.8	14.7
Marine Aerosols	11.9	5.1
Soil	8.9	12.2
Diesel Vehicles	11.1	7.7
Gasoline Vehicles	2.6	3.8
Pulp Mills	6.1	2.5
Aluminum Reduction	3.3	5.6

Marine aerosols contributed about 130% more mass at Mt. Zion than at Wishram, which is reasonable since Mt. Zion is on the west side of the Cascade Mountains and is more frequently exposed to storms coming off the Pacific Ocean. Soil contributed about 40% more mass at Wishram, which would be expected since Wishram is in the drier, more arid climate found on the east side of the Cascades where the soil is more susceptible to wind-blown erosion. Pulp mill contribution at Mt. Zion was about 140% larger than at Wishram. This is reasonable since the Mt. Zion site is located 15 km east of a large pulp mill in Camas, and 40-50 km west of two forest product facilities that may have emissions similar to pulp and paper mills. The contribution of aluminum smelters to the Wishram site was about 70% larger than at the Mt. Zion site. This is a feasible result since the Wishram site is located between the Northwest Aluminum Smelter located at the Dalles (16 km to the SW), and the Goldendale Aluminum Smelter located at John Day Dam (24 km to the east). The Reynold Metals facility, the only aluminum source close to Mt. Zion, had very low particulate emissions during the 1992-97 period.

Based on the diesel and gasoline mass allocations, the diesel-to-gasoline mass ratio for Mt. Zion was 4.2, which is reasonable for a site close to a large urban area such as Portland. The ratio for Mt. Zion is larger than the urban diesel-to-gasoline ratio of 3.2 for the city of Pasadena, California derived by CMB modeling using organic source markers (Schauer et. al.). The ratio for Wishram was 2.0, which is slightly below the national ratio of 2.3 based on the national 1997 PM<sub>2.5</sub> emission inventory (National Academy of Sciences).

### Time-Dependent Source Allocations

Time-dependent PMF source allocations for the Mt. Zion and Wishram sources are shown in Appendix B. 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

<b>Source</b>	<b>Mt. Zion</b>	<b>Wishram</b>
Vegetative Burning	August-October	September-February
Ammonium Sulfate	May-August	May-October
Ammonium Nitrate	October-January	November-March
Marine Aerosols	September-June	October-March
Soil	July-September	June-October
Diesel-powered Sources	No pattern	No pattern
Gasoline-powered Sources	August-October	November-January
Pulp Mills	March-August	No pattern
Aluminum Reduction	No pattern	No pattern

Both Mt. Zion and Wishram sites are impacted by vegetative burning during the late summer and early fall. Vegetative burning during this time of year is due to forest fires and agricultural burning. However, residential wood combustion may account for a large fraction of fine particulate emissions during the winter months, which may be the reason why vegetative burning remains high at Wishram through the winter. Ammonium sulfate is high at both sites during the summer and early when solar intensity and air temperatures are high. Conversion of SO<sub>2</sub> to sulfate is by an oxidation reaction that involves the OH<sup>-</sup> radical, which is present during ozone formation in the presence of sunlight and high temperatures. Ammonium nitrate is highest at both sites during the winter months, which would be expected since ammonium nitrate has a very low vapor pressure and its formation is more stable during wintertime conditions.

The most significant marine influence at Mt. Zion (September-June) is longer than at Wishram (October-March) as would be expected. Soil makes its largest contribution at Mt. Zion during the months of July-September, and over a longer period (June-October) at Wishram, as would be expected. The highest level of soil was recorded at Mt. Zion on April 29, 1998, which coincides with the arrival of an Asian dust storm in the Pacific Northwest. Gasoline-powered sources show the highest contribution during August-October at Mt. Zion, and November-January at Wishram. These periods may coincide with periods of high vehicle traffic or by periods of air stagnation that trap vehicle exhaust close to the ground. The pulp mill contribution is the largest at Mt. Zion during the period of March-August, but does not seem to have a seasonal pattern at the Wishram site. Diesel-powered sources and aluminum smelters do not appear to have a seasonal pattern at either site.

## Conclusions

This study generated feasible 9-source solutions for the 1996-98 Mt. Zion data and the 1993-95 Wishram data. Diesel and gasoline-powered source profiles were resolved at both sites, and the profiles for each of these sources were more consistent with those in the literature. The aluminum smelter and pulp mill source profiles were more consistent between the Mt. Zion and Wishram sites, and the allocations from these sources were also more realistic based on the distances from these sources and the emission inventories of these sources.

The PMF results in this paper differ from the previous PMF results in Kuhns et. al. In this study 9 sources were identified for each site, instead of 8 sources as in the previous study. All 9 sources had characteristics of sources impacting the Columbia Gorge and had time-dependent allocations characteristic of these sources. Modifications made in this analysis over the prior analysis were: 1) diesel and gasoline-powered source profiles were identified using markers identified by Maykut et. al., 2) data and data uncertainties reported as “zero” were replaced with  $\frac{1}{2}$  MDL, and 3) OC1 data were included in the analysis to help distinguish between diesel and gasoline-powered sources. Diesel and gasoline-powered source profiles generated by PMF in this study, which were based on monitoring data collected at ambient air monitoring sites distant from the sources, differ from mobile source profiles constructed from monitoring data collected near these sources. These differences in source profiles may be due either to sampling artifacts, laboratory analysis artifacts, or actual chemical changes that occur when organic compounds react in the atmosphere as they travel from the sources to the monitoring sites (Maykut et. al.).

The PMF results indicate that vegetative burning and secondary sulfate, each at about 24% percent mass allocation, are the major sources of fine particulates at both the Mt. Zion and Wishram sites. PMF results also show that combined diesel and gasoline-powered sources are the third highest contributors at Mt. Zion (13.7%), and are the fourth highest contributors at Wishram (11.5%). These results are in contrast to the CMB modeling of Columbia Gorge IMPROVE data in Kuhns et. al. The CMB results indicated that the major source of fine particulates for 1996-98 at both sites was motor vehicles (Mt. Zion - 43%, Wishram - 41%), and that the contribution of vegetative burning at each site was relatively small (Mt. Zion - 9%, Wishram - 7%). PMF also predicts a higher contribution of secondary nitrate at Mt. Zion and Wishram than CMB. PMF source allocations for the secondary sulfate, aluminum smelters, soil, and marine sources were similar to the CMB results.



## **References**

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

Kuhns et. al, Chemical Mass Balance Source Apportionment of PM2.5 Aerosol in the Columbia River Gorge, submitted to the Oregon Department of Environmental Quality, March 31, 2003.

Lee et al., 1999. Application of Positive Matrix Factorization in Source Apportionment of Particle Pollutants in Hong Kong, *Atmospheric Environment*, 33, 3201-3212.

Maykut et. al., Source Apportionment of PM2.5 at an Urban IMPROVE Site in Seattle Washington, *Environmental Science and Technology*, 2003, 37, 5135-5142.

National Academy of Sciences, Modeling Mobile Source Emissions, National Academy Press: Washington, DC, 2000, pp. 20-28.

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

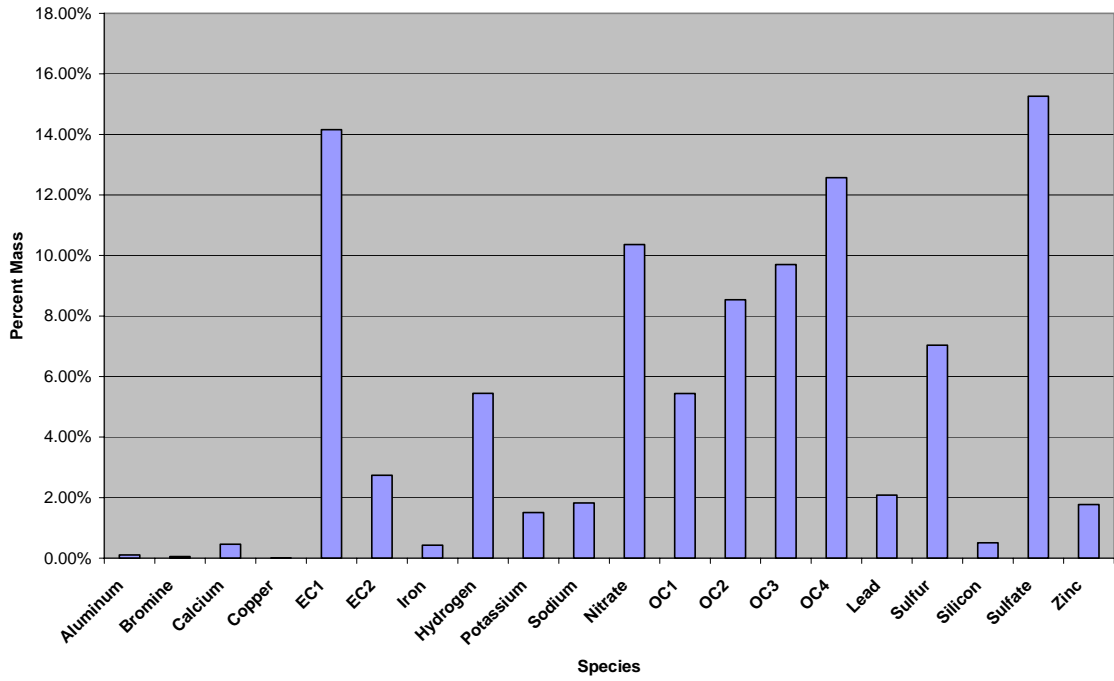
Paatero, P., 1996. User's Guide for Positive Matrix Factorization Programs PMF2.EXE and PMF3.EXE, University of Helsinki, Helsinki.

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

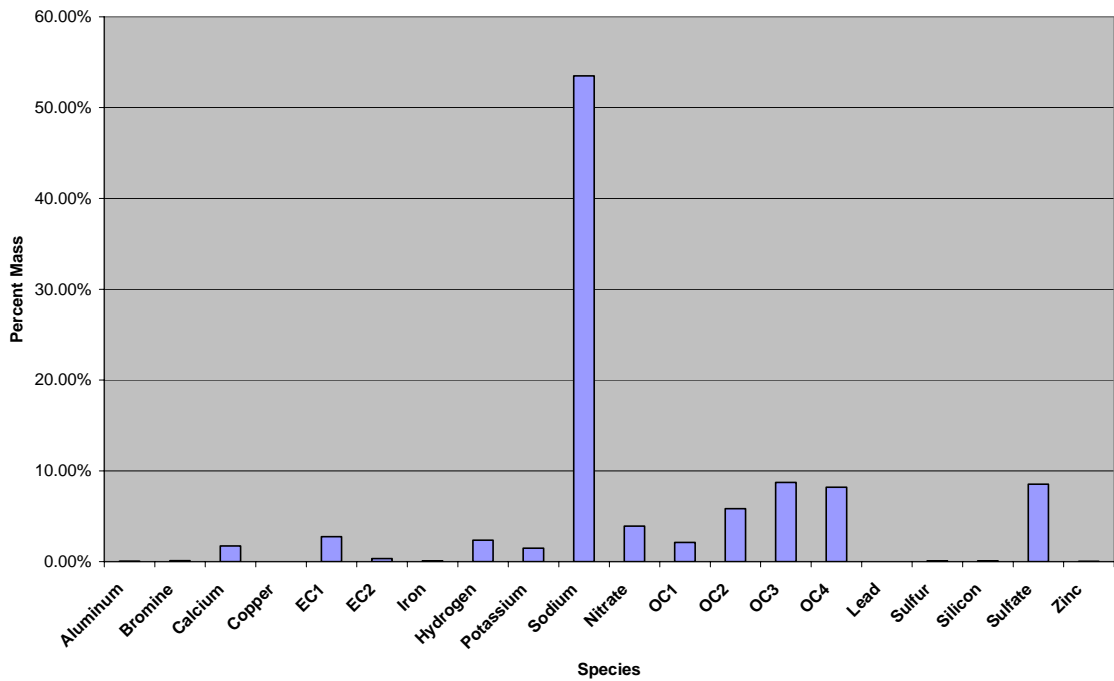
Schauer et. al., *Atmos. Environ.* 1996, 30 (22), 3837-3855.

**APPENDIX A**  
**PMF SOURCE PROFILES**

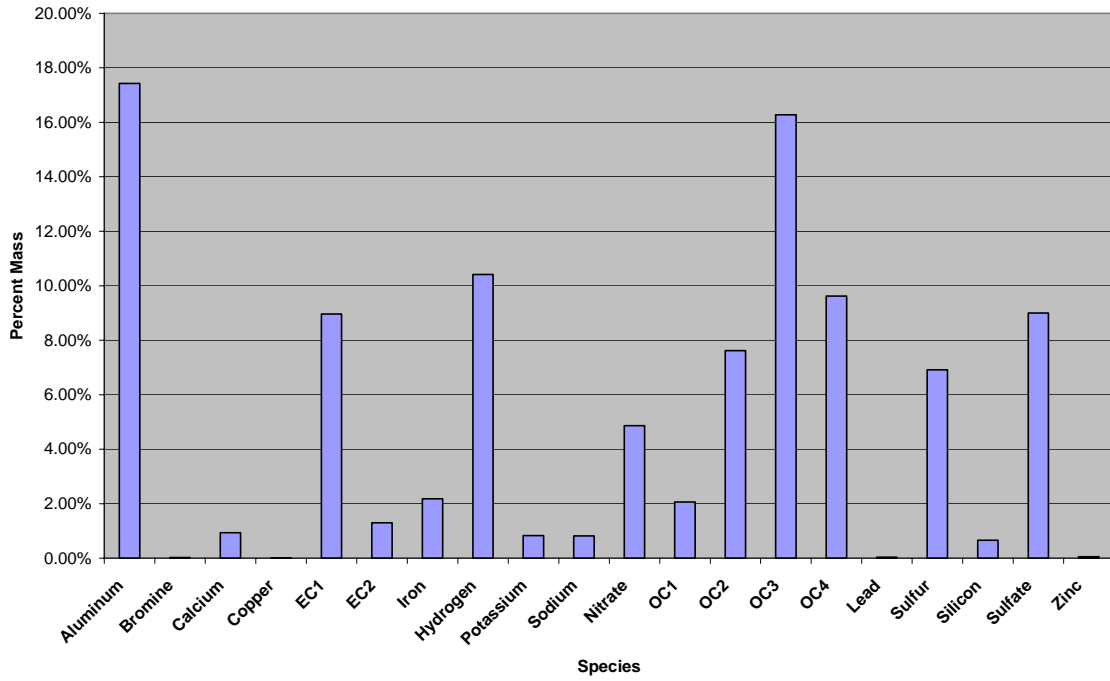
Mt. Zion - Gasoline Profile



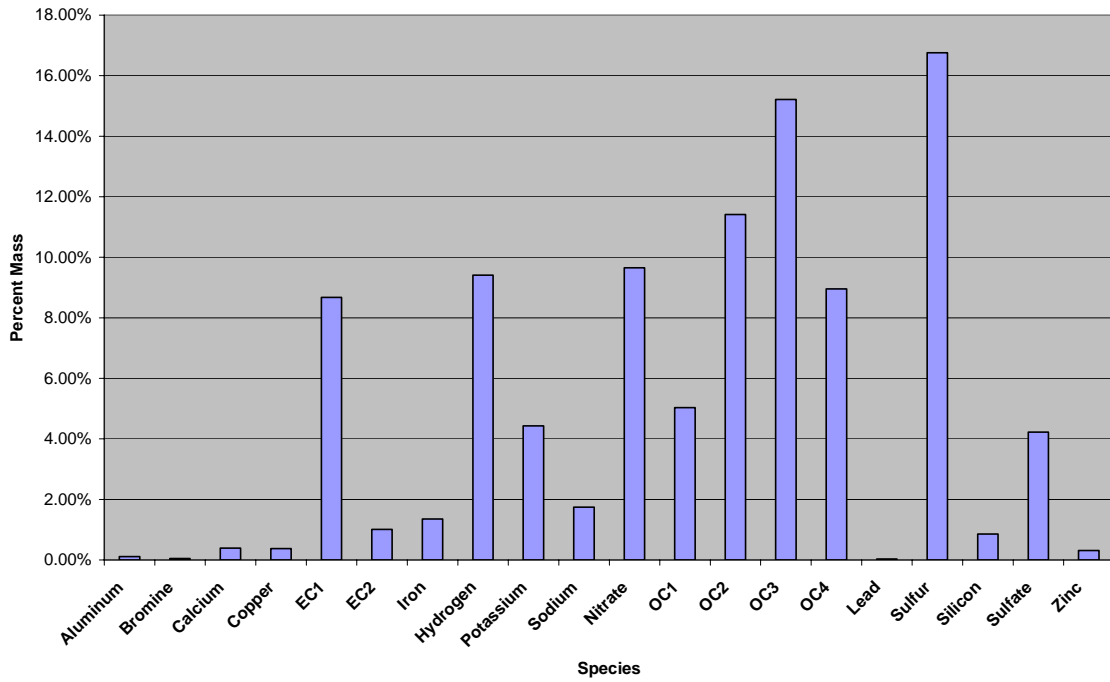
Mt. Zion - Marine Profile



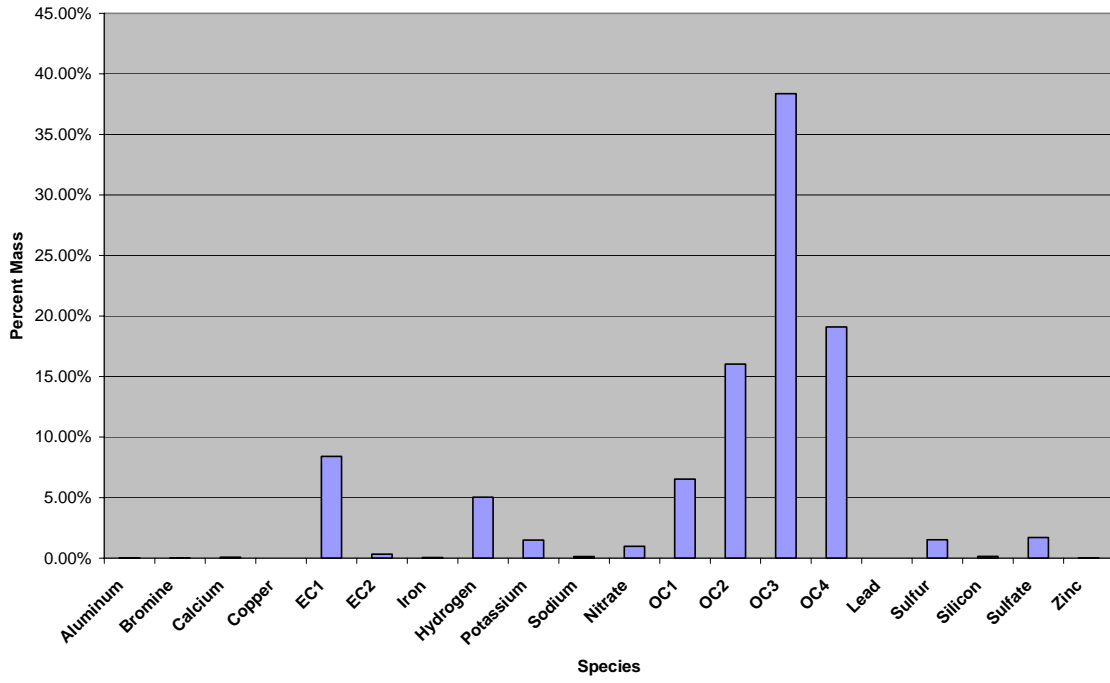
Mt. Zion - Aluminum Smelter Profile



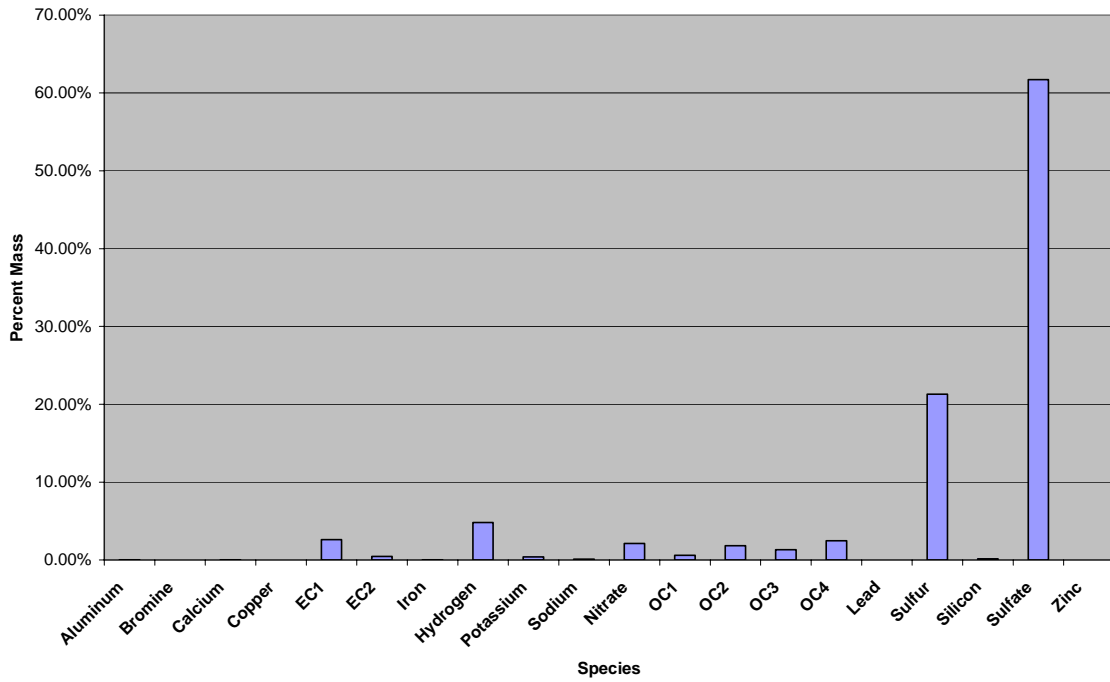
Mt. Zion - Pulp Mill Profile



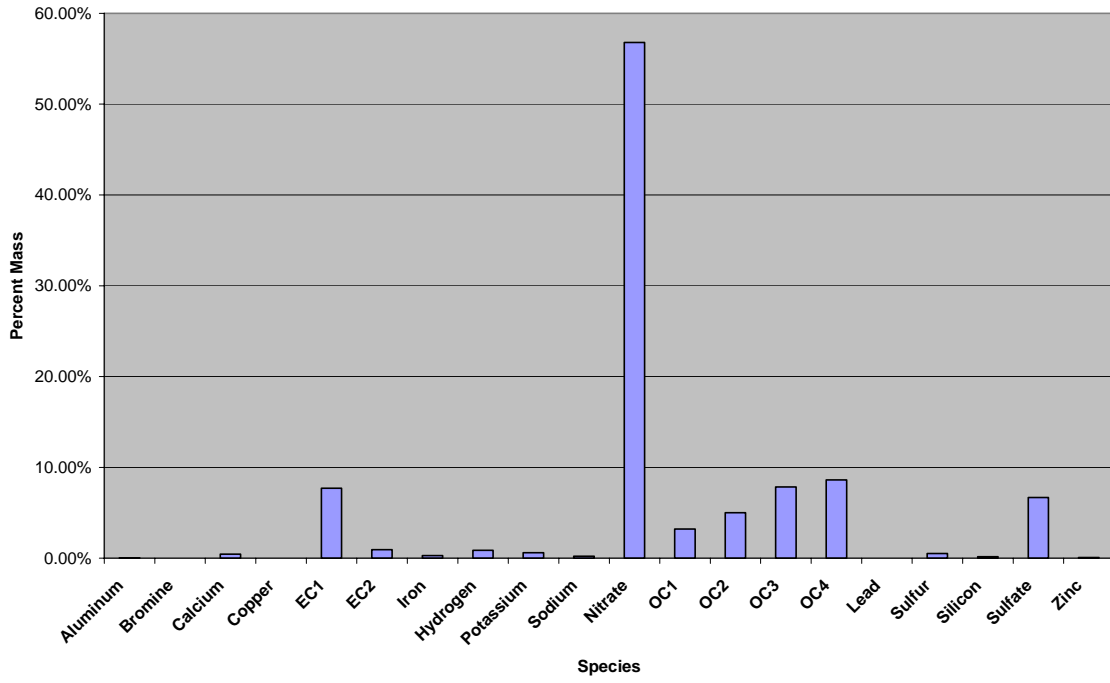
Mt. Zion - Vegetative Burning Profile



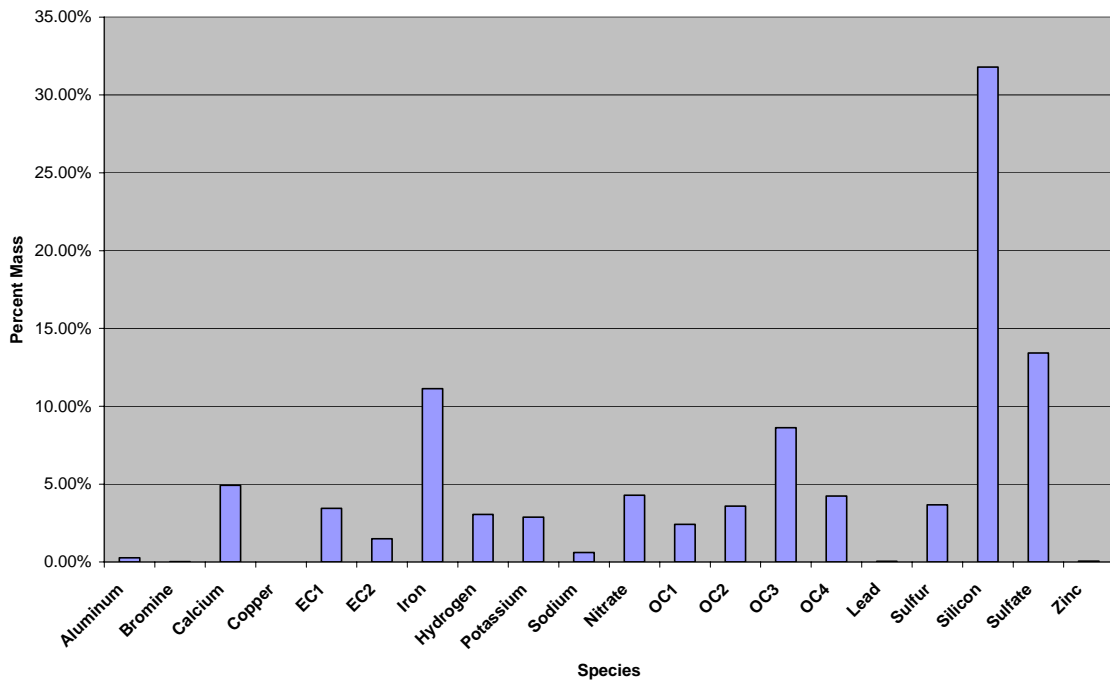
Mt. Zion - Secondary Sulfate Profile



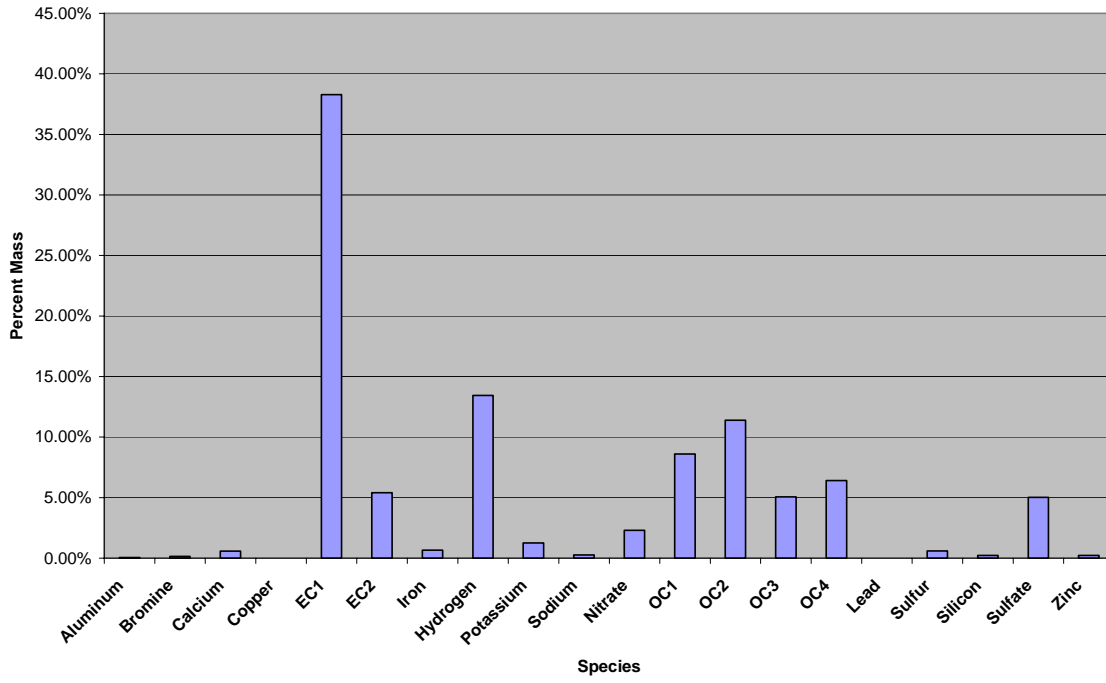
Mt. Zion - Secondary Nitrate Profile



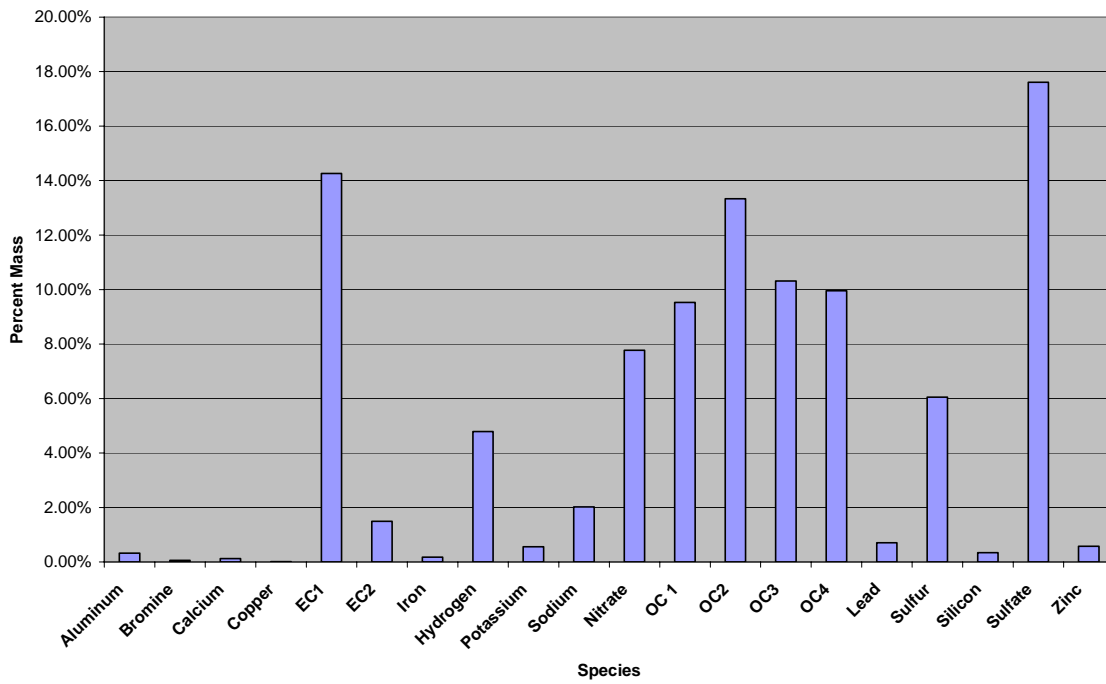
Mt. Zion - Soil Profile



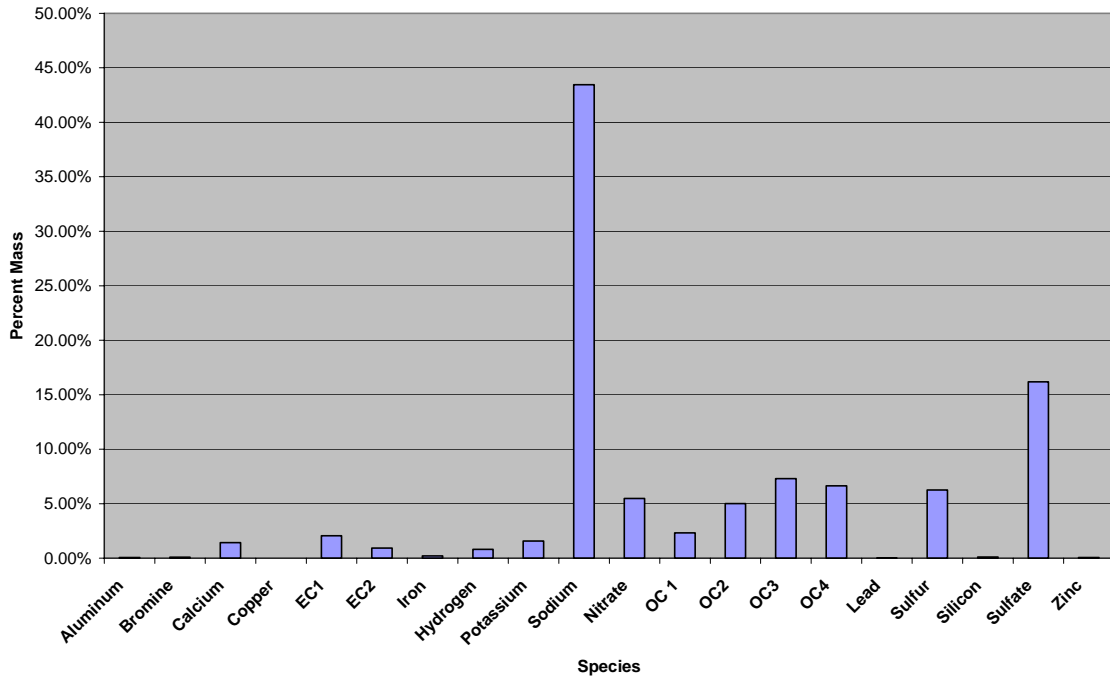
**Mt. Zion - Diesel Profile**



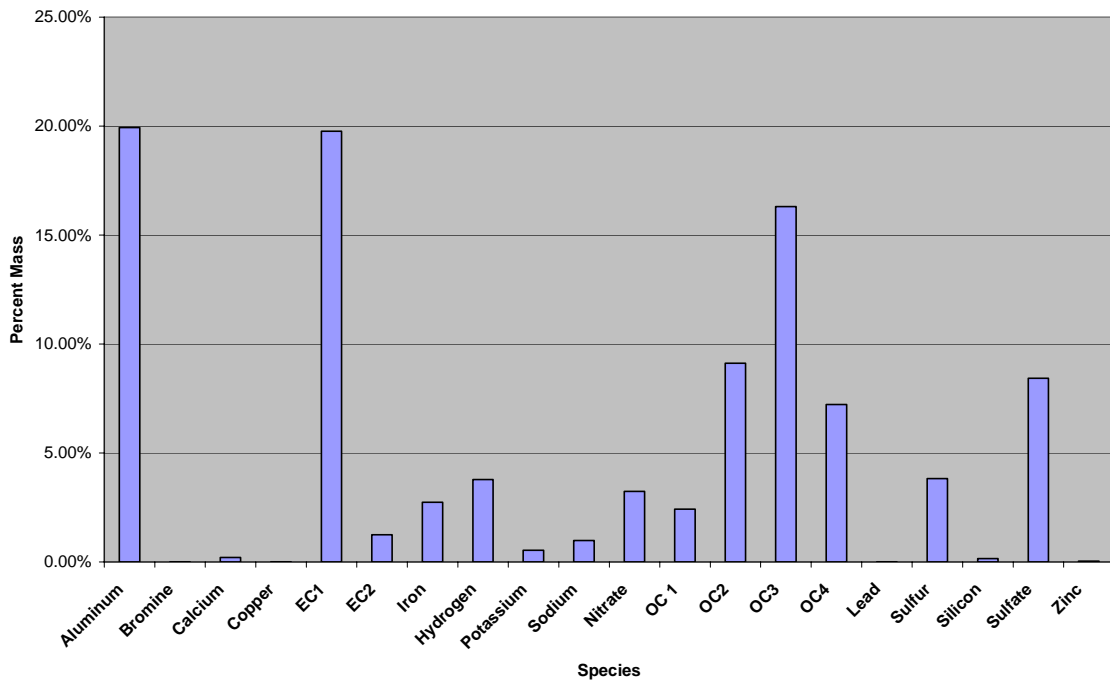
**Wishram - Gasoline Profile**



Wishram - Marine Profile

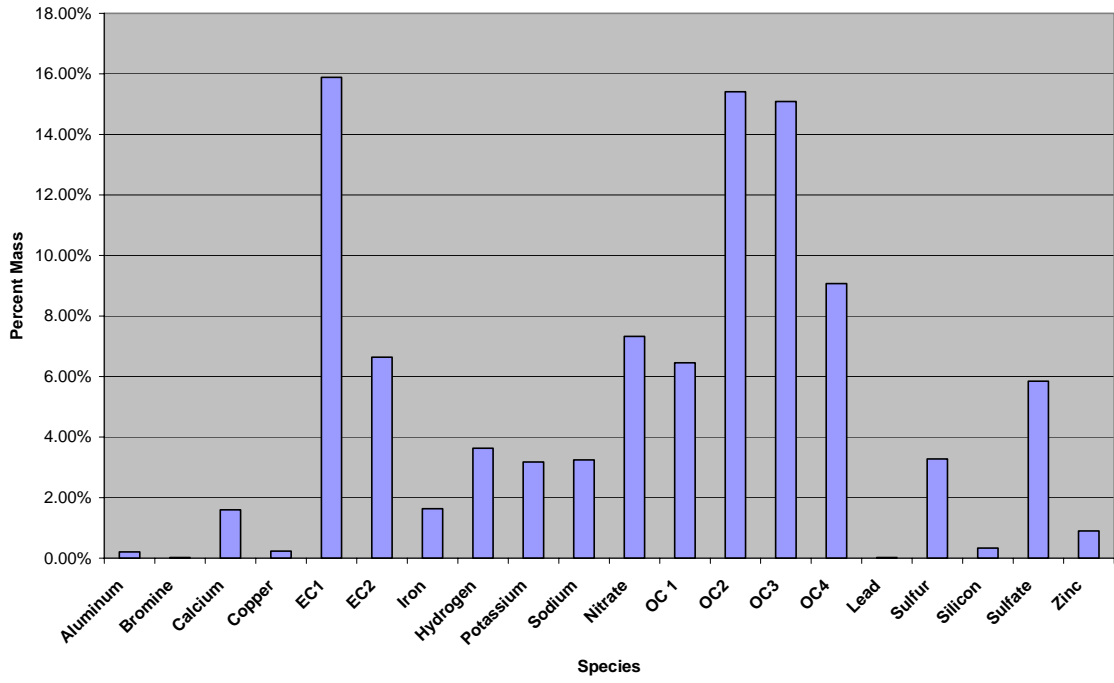


Wishram - Aluminum Smelter Profile

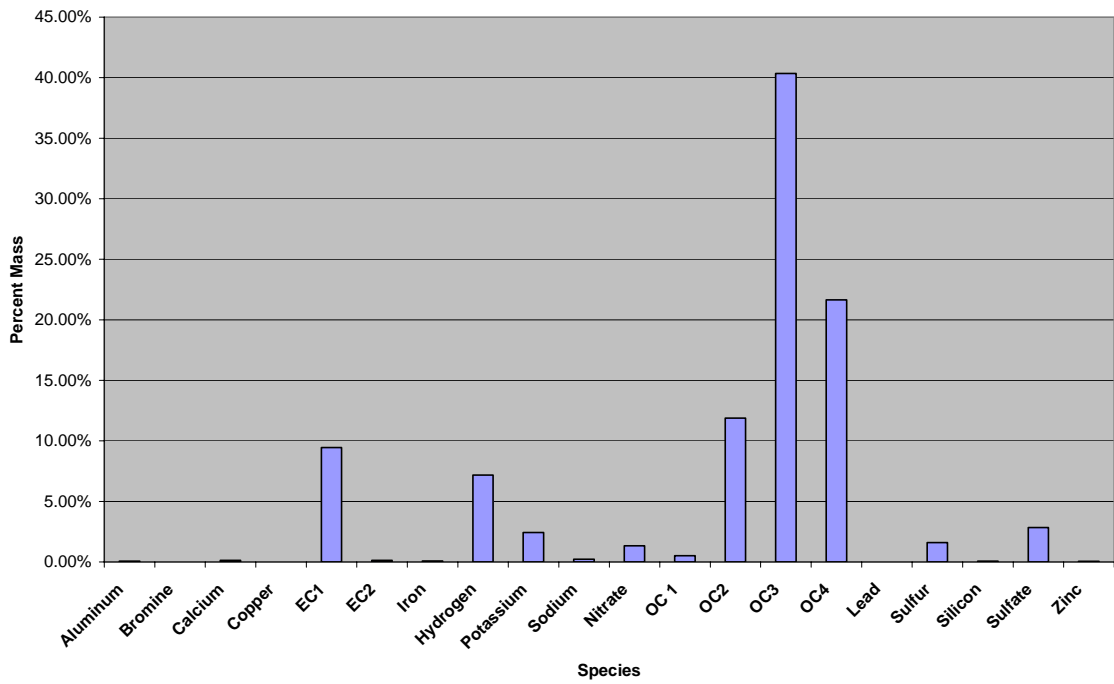




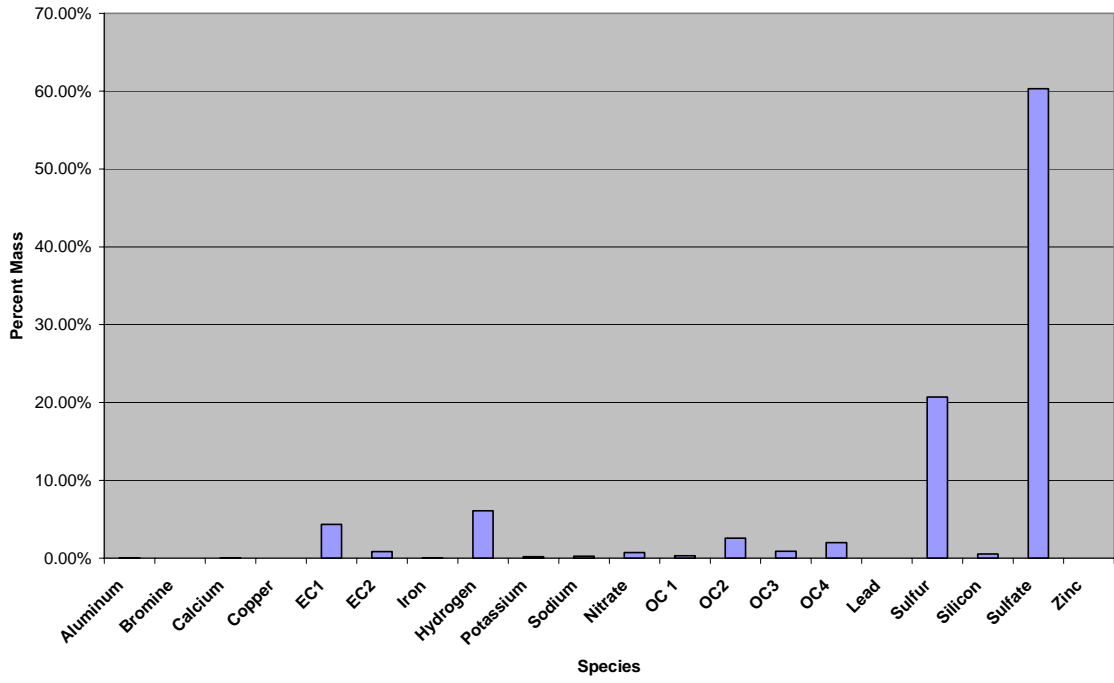
Wishram - Pulp Mill Profile



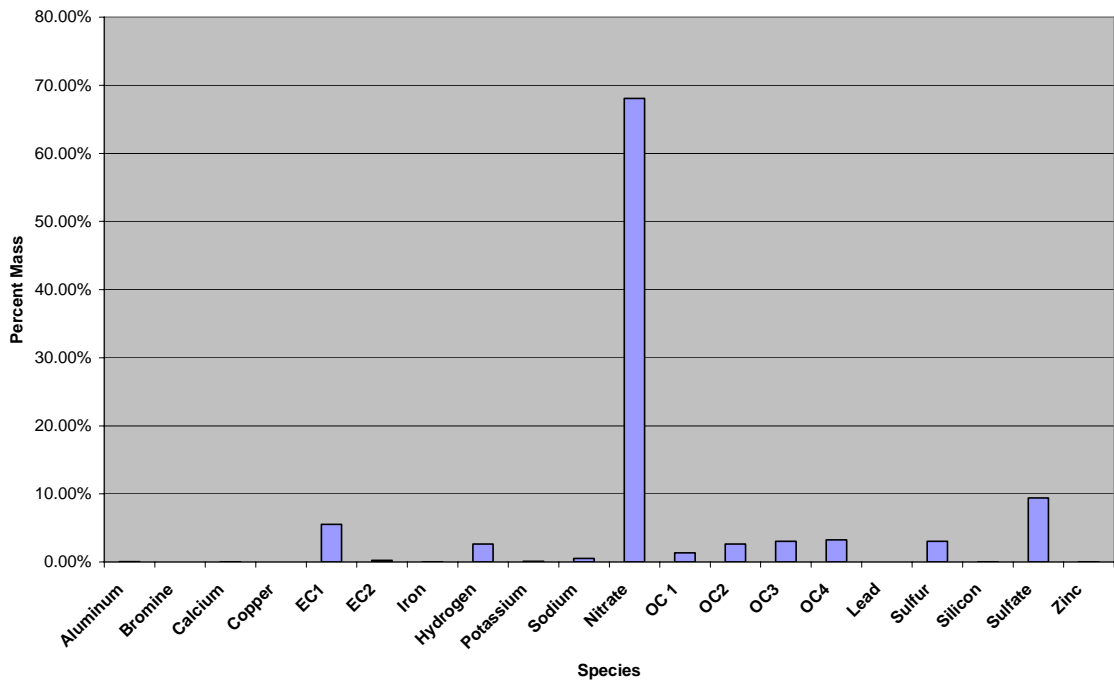
Wishram - Vegetative Burning Profile



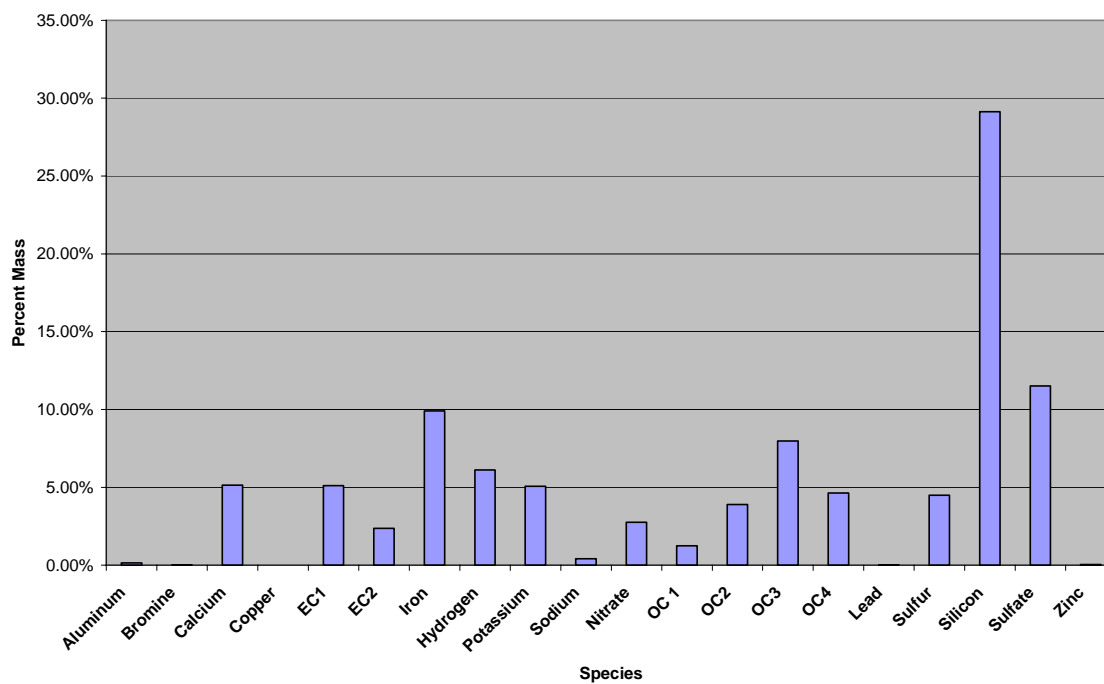
Wishram - Secondary Sulfate Profile



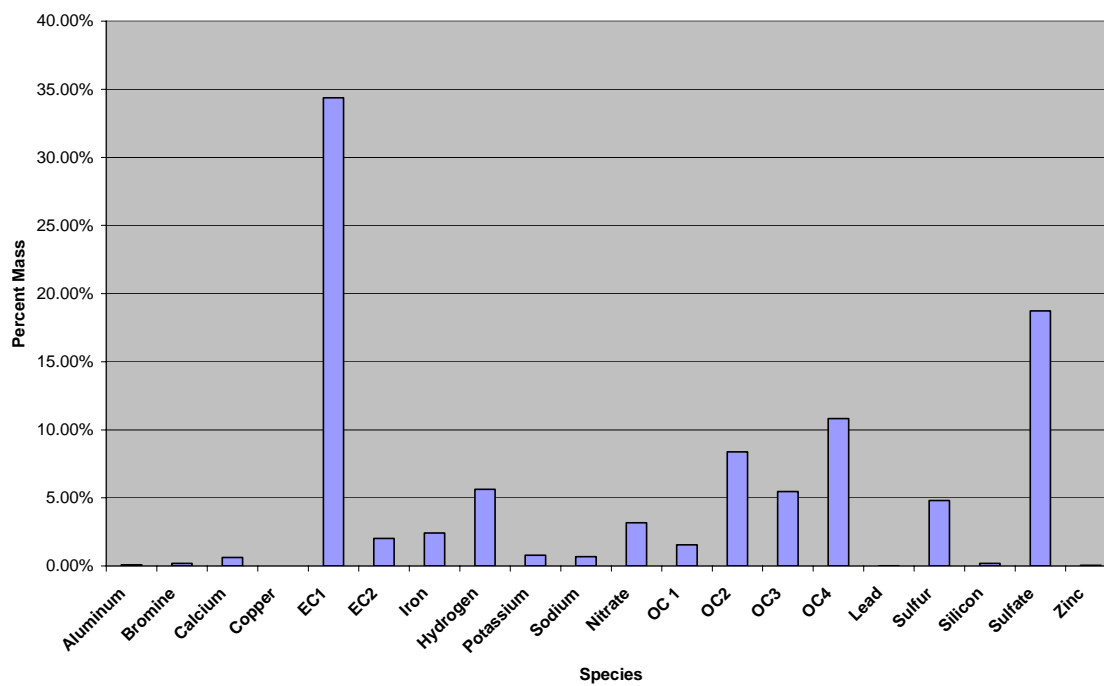
Wishram - Secondary Nitrate Profile



Wishram - Soil Profile

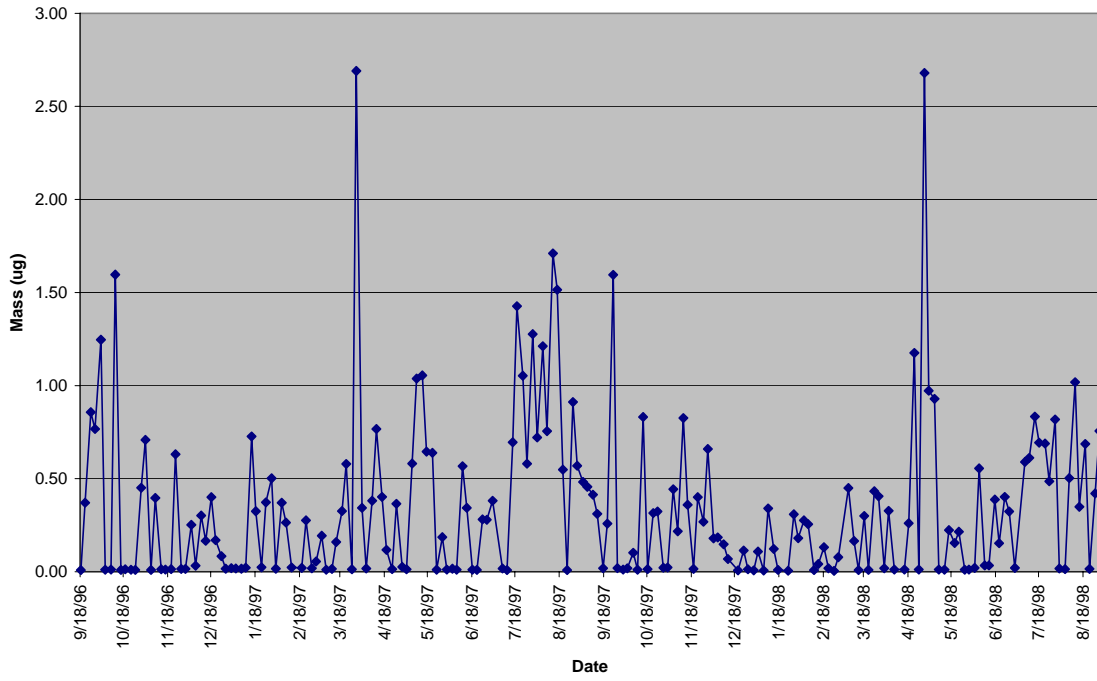


Wishram - Diesel Profile

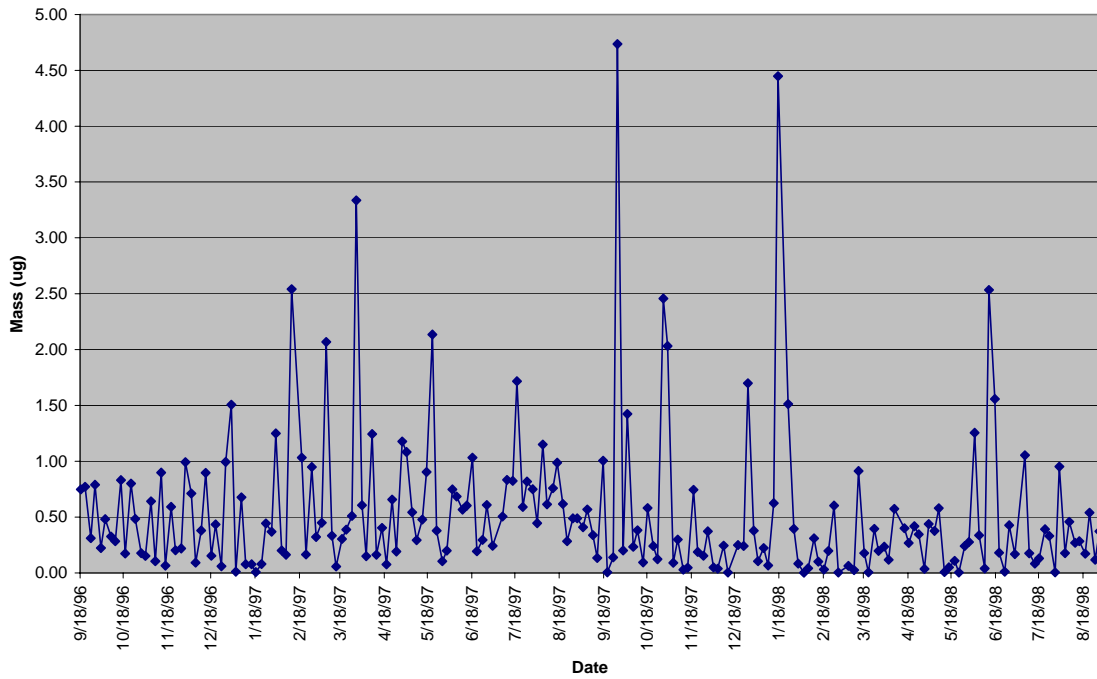


**APPENDIX B**  
**PMF TIME SERIES SOURCE ALLOCATIONS**

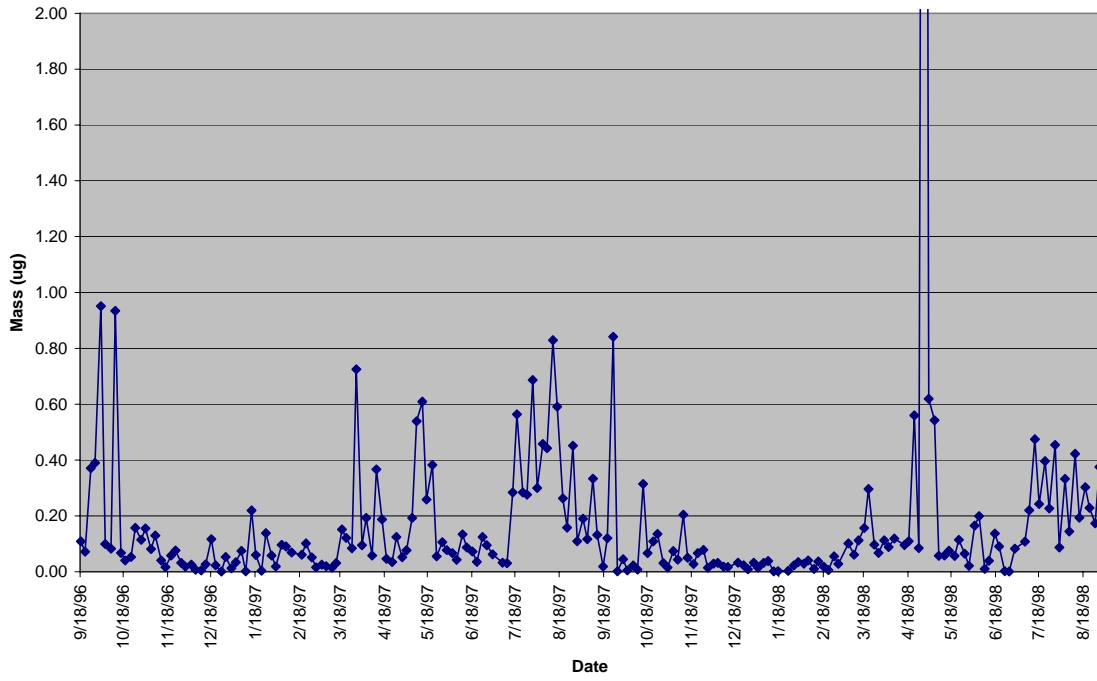
Mt. Zion - Aluminum Smelter Allocation



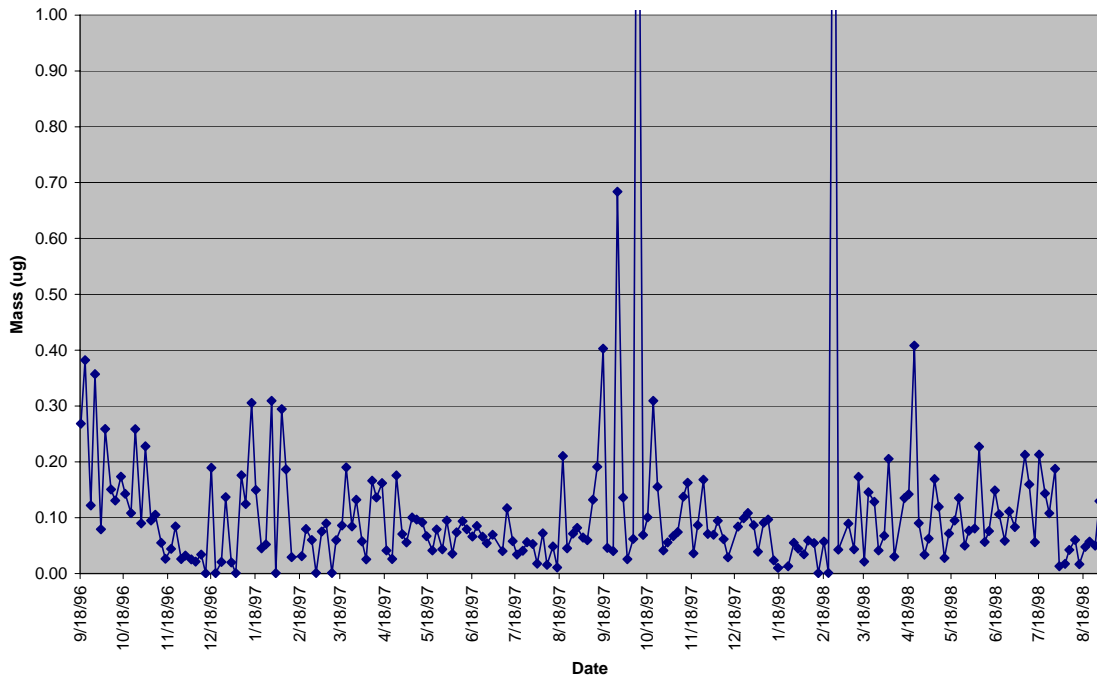
Mt. Zion - Marine Allocation



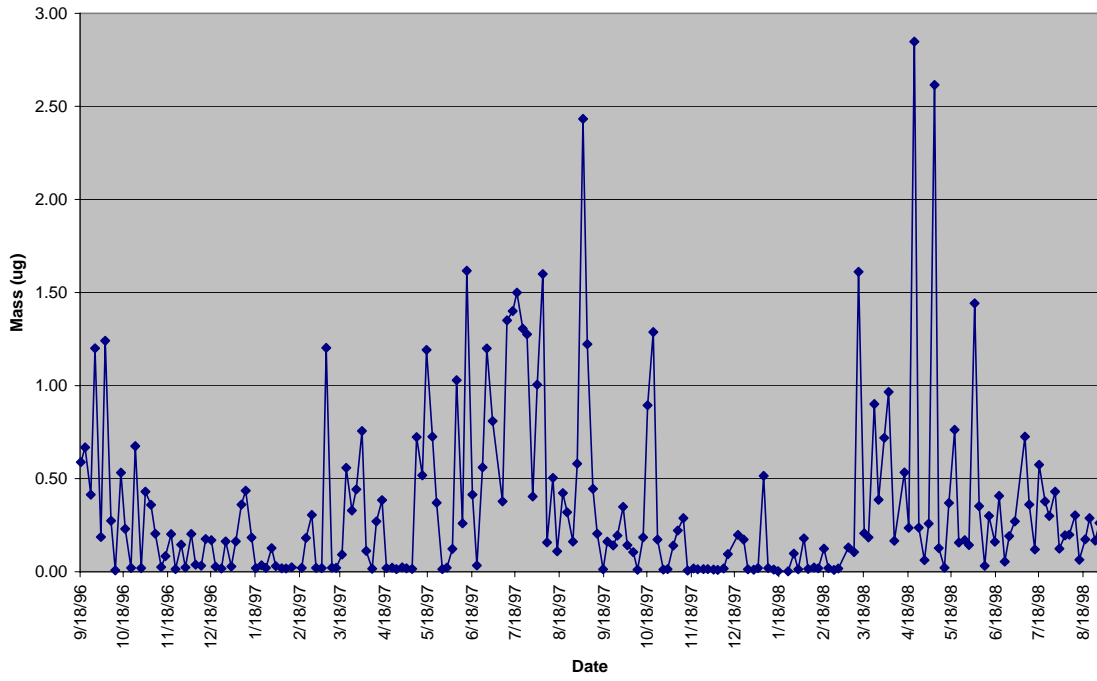
### Mt. Zion - Soil Allocation



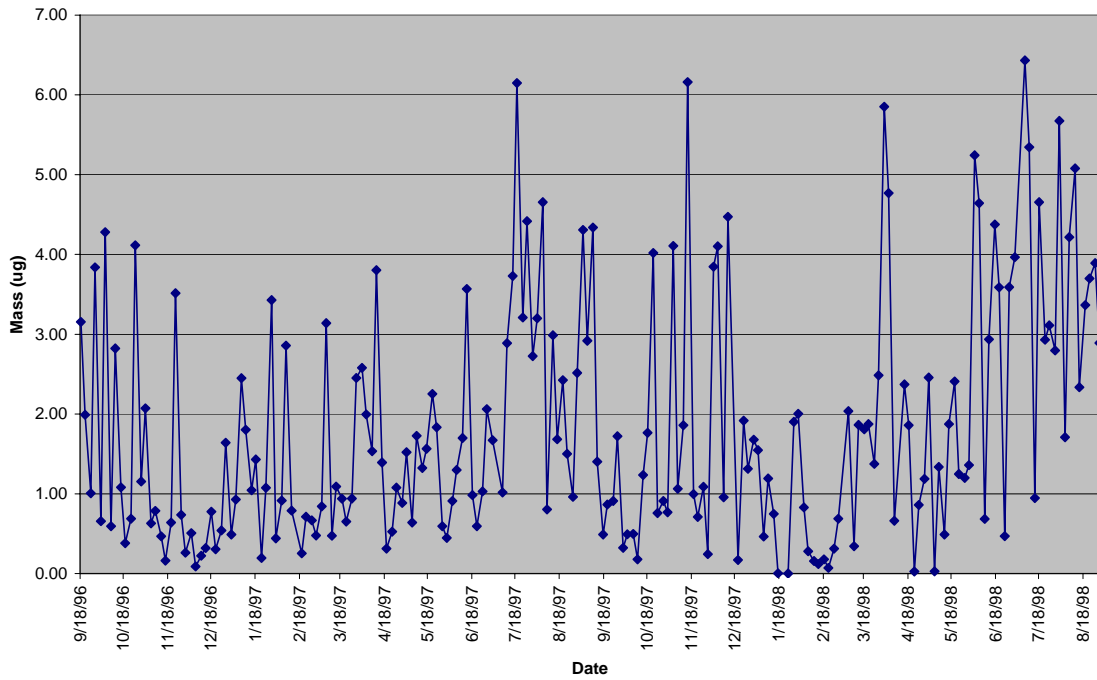
### Mt. Zion - Gasoline Allocation



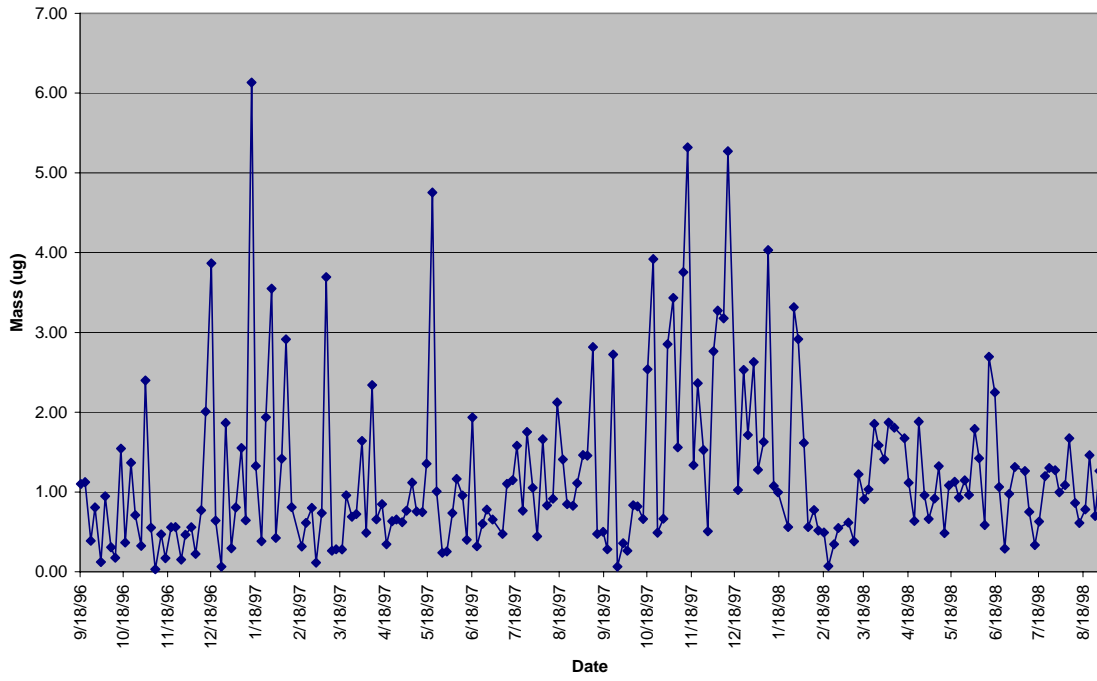
Mt. Zion - Pulp Mill Allocation



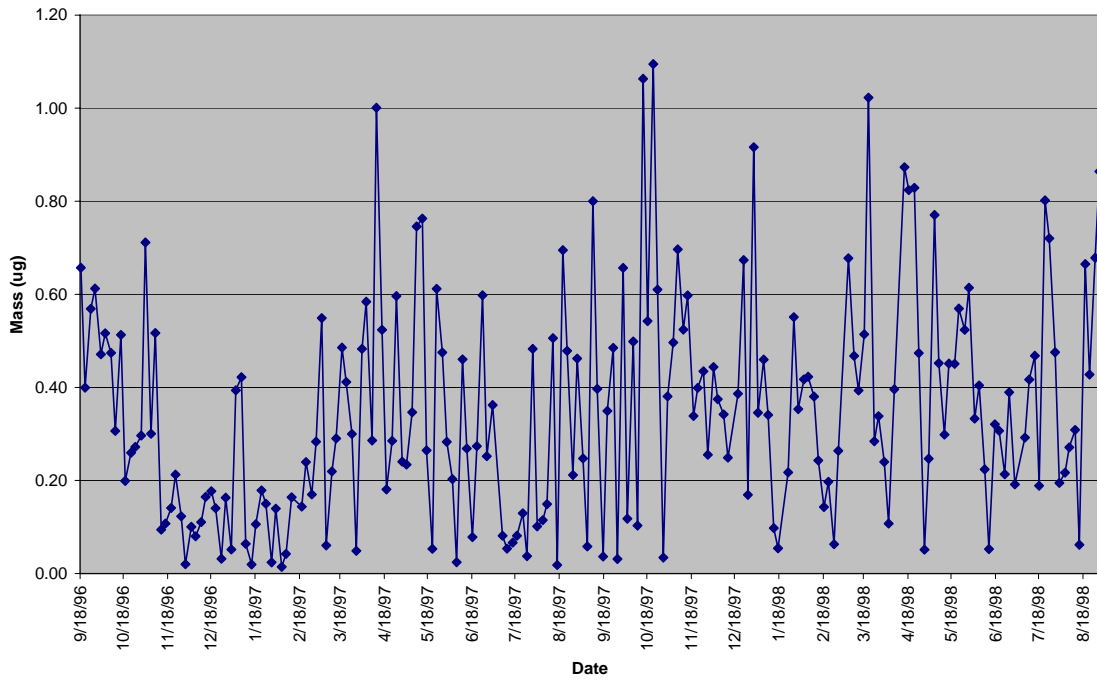
Mt. Zion - Secondary Sulfate Allocation



Mt. Zion - Secondary Nitrate Allocation

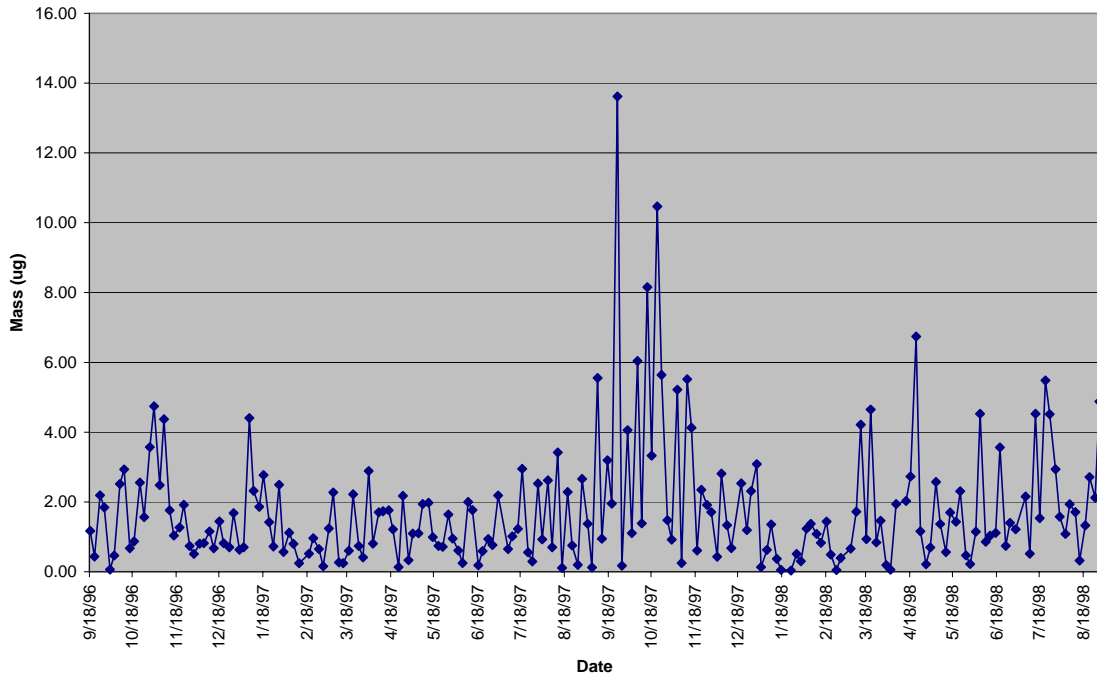


Mt. Zion - Diesel Allocation

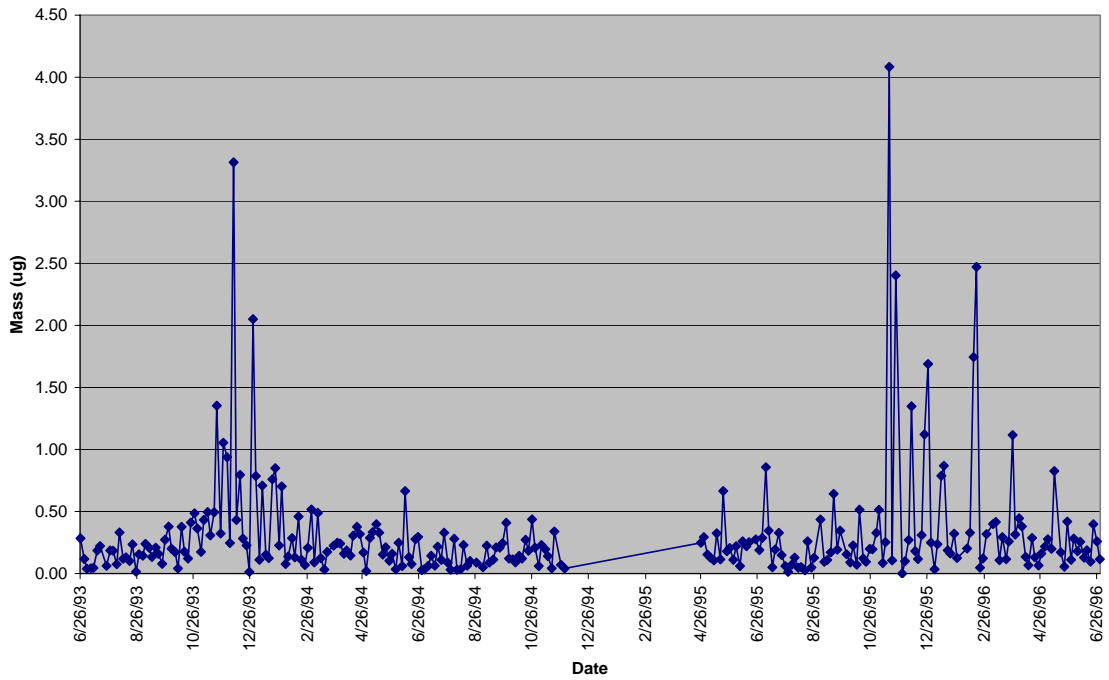




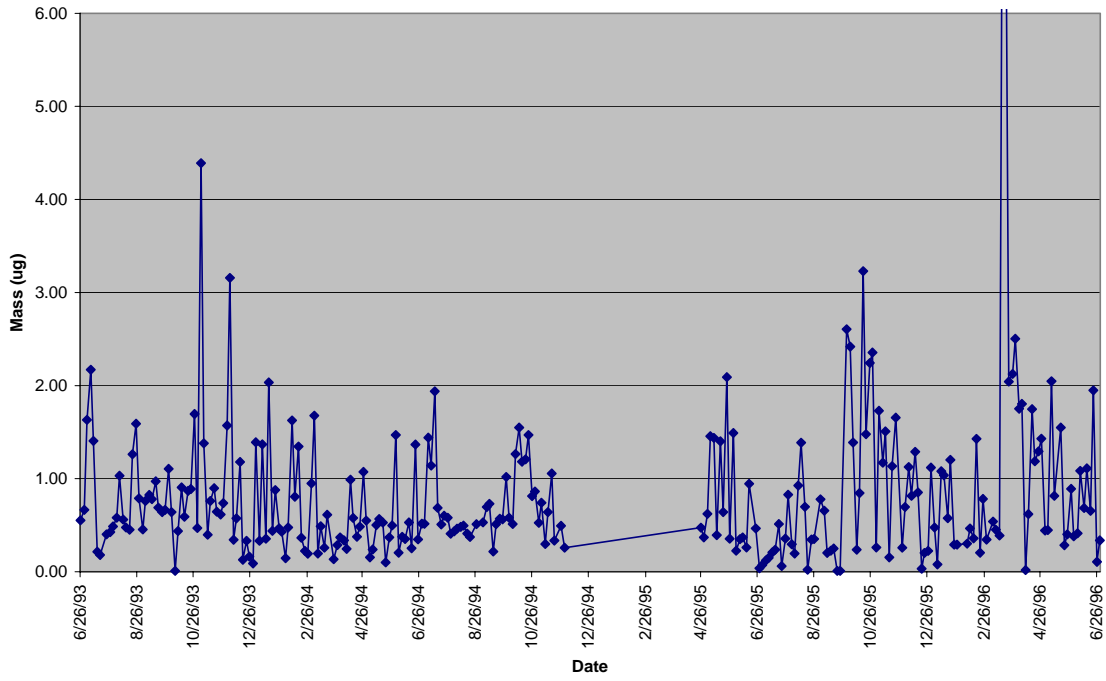
### Mt. Zion - Vegetative Burning Allocation



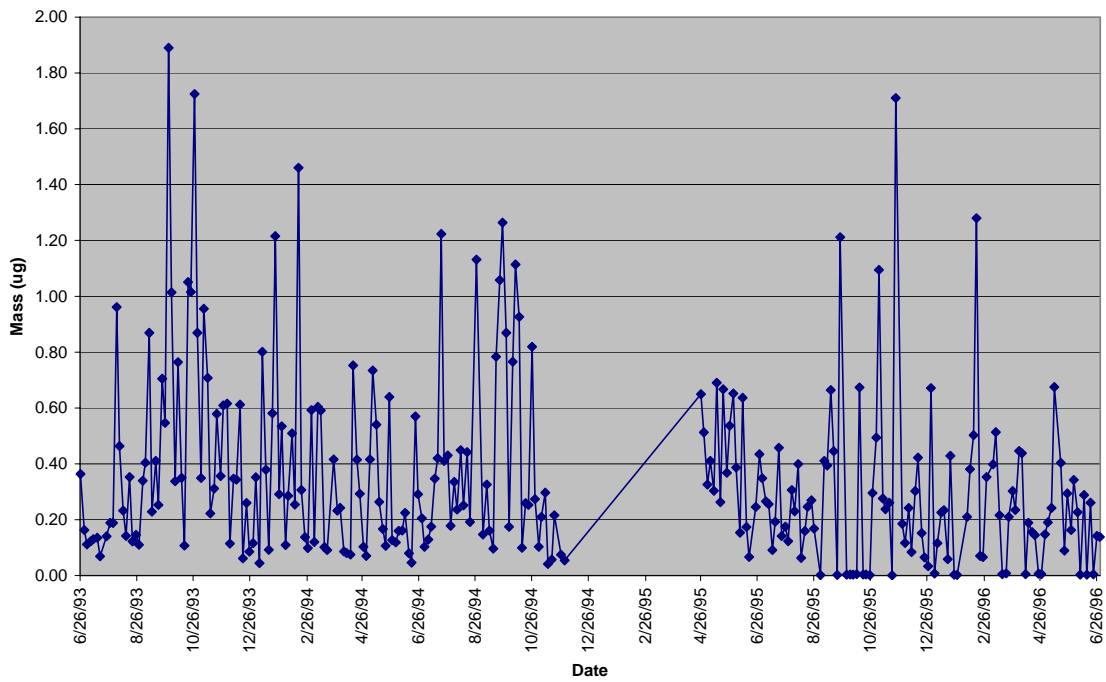
### Wishram - Gasoline Allocation



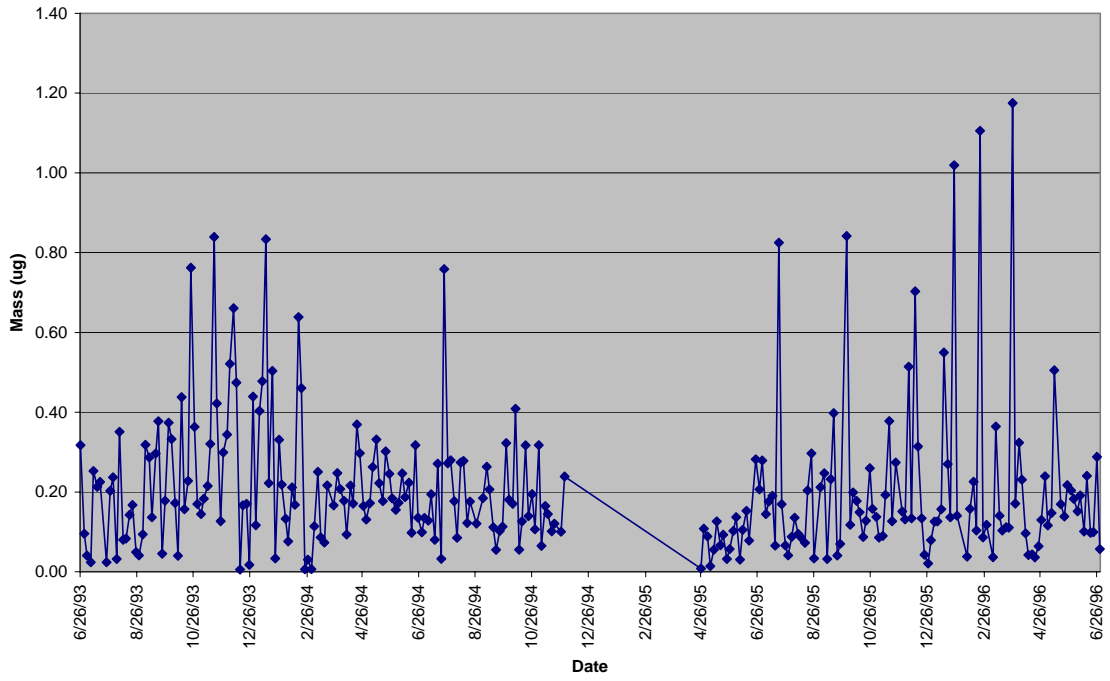
### Wishram - Marine Allocation



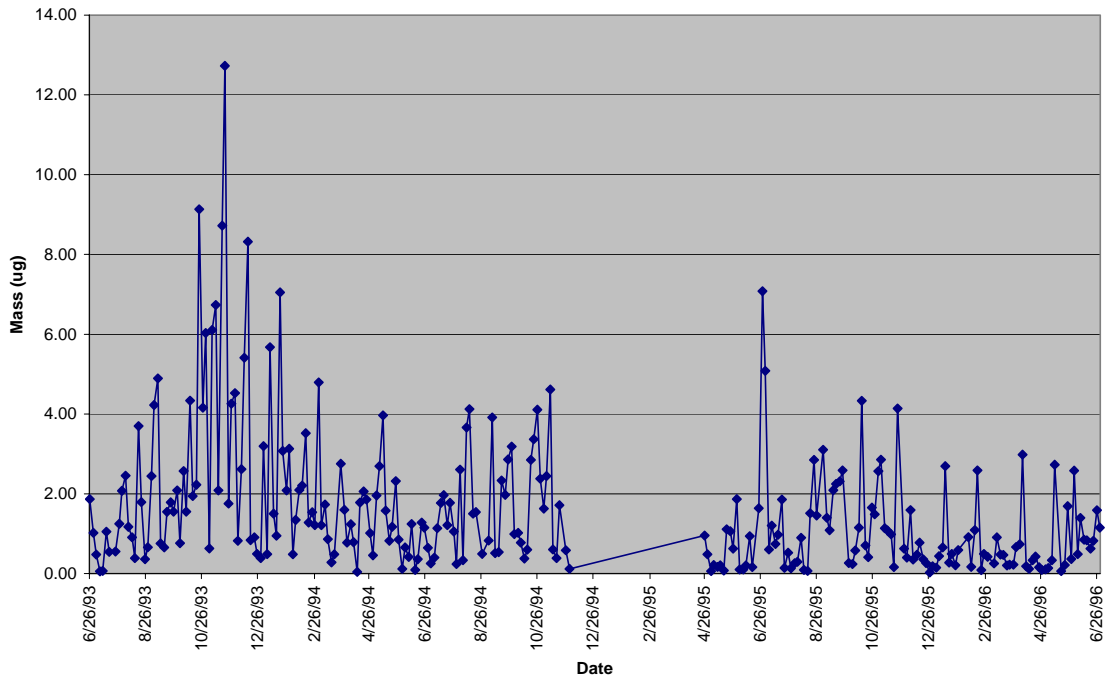
### Wishram - Aluminum Smelter Allocation



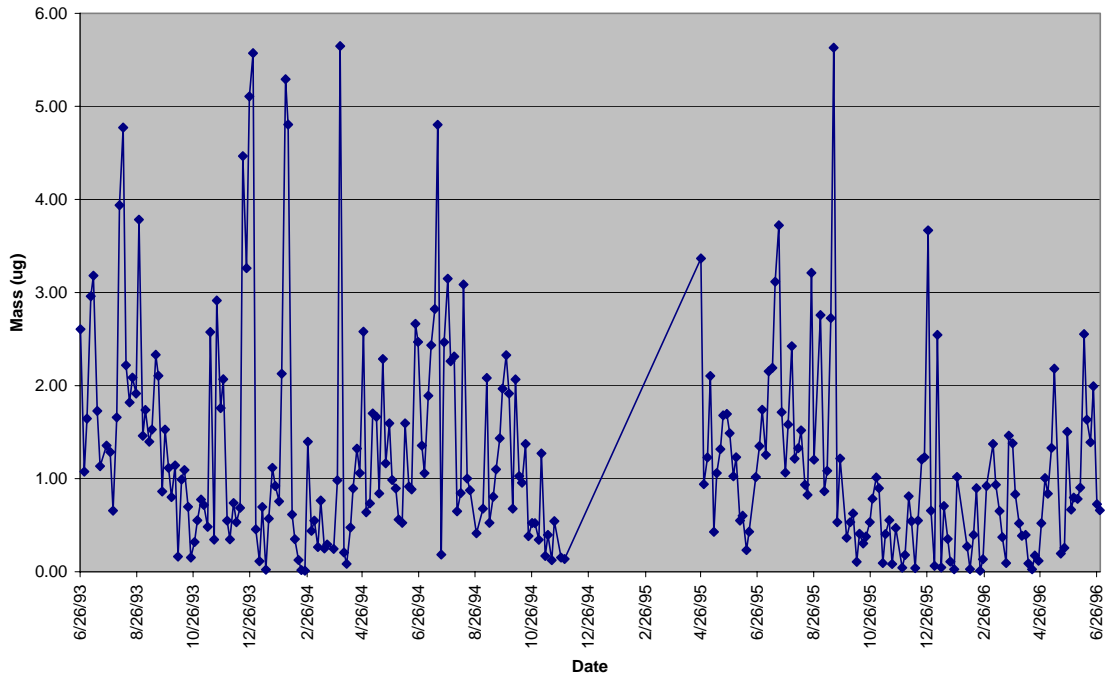
### Wishram - Pulp Mill Allocation



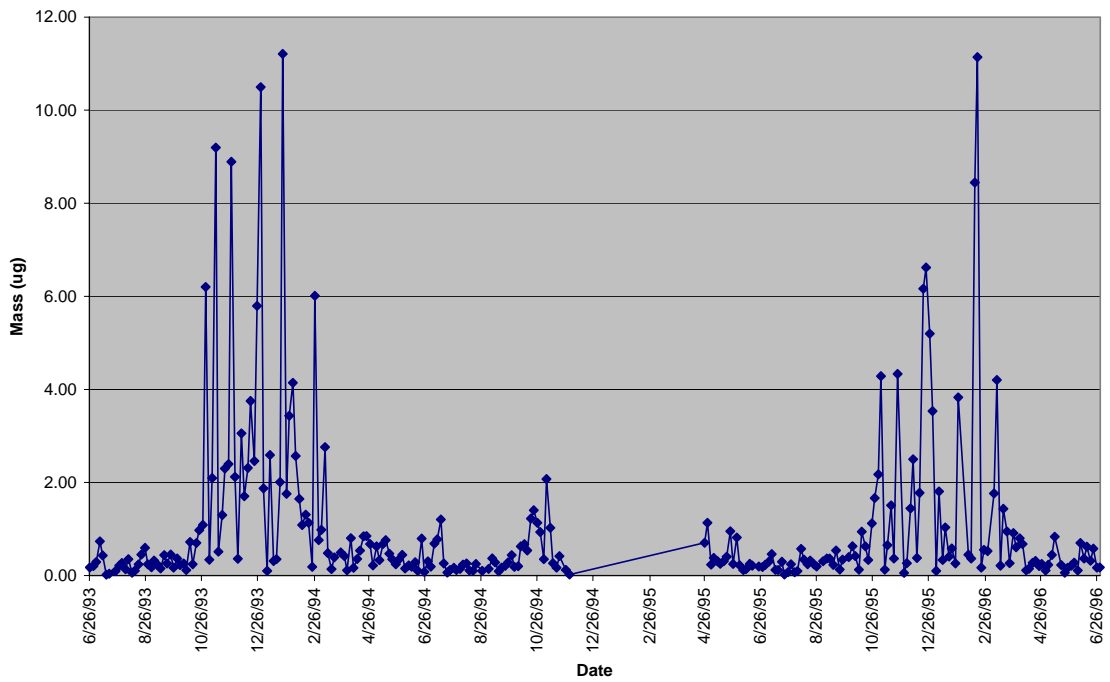
### Wishram - Vegetative Burning Allocation



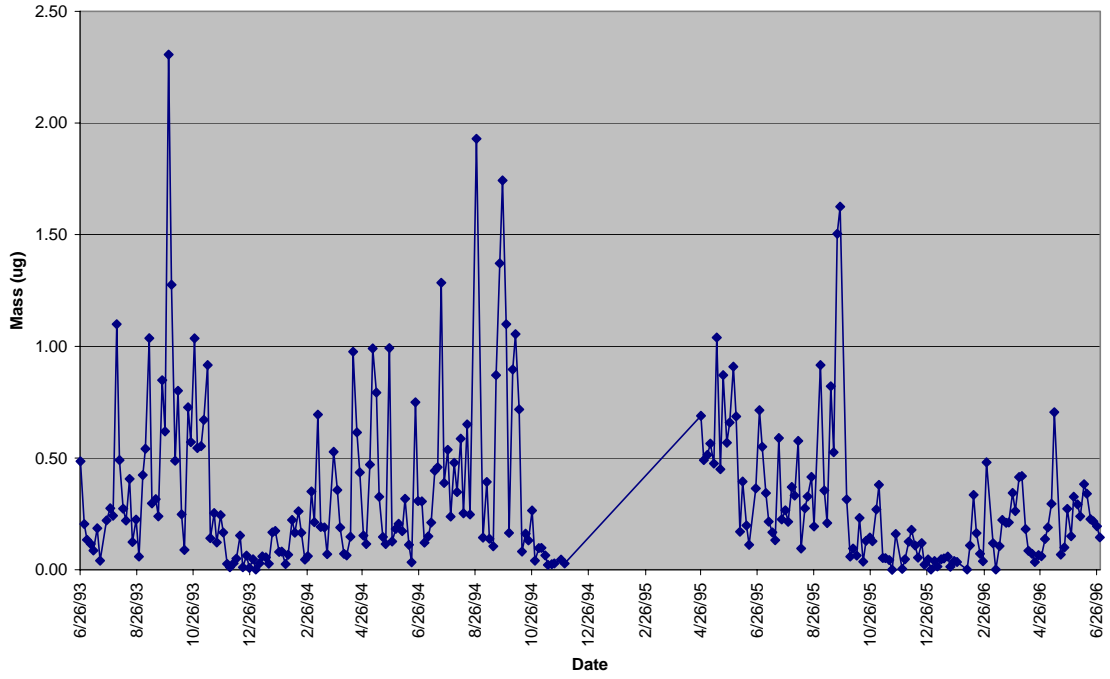
Wishram - Secondary Sulfate Allocation



Wishram - Secondary Nitrate Allocation



Wishram - Soil Allocation



Wishram - Diesel Allocation

