

# PMF modeling for WRAP COHA

## 1 Introduction

In order to identify the sources of aerosols in the western United States, Positive Matrix Factorization (PMF) receptor model is applied to the 24-hr integrated aerosol chemical composition data obtained at the Class I areas of the Western Regional Air Partnership (WRAP) region through the Interagency Monitoring of Protected Visual Environments (IMPROVE) program. The IMPROVE sites in the WRAP region are grouped into 18 sub-regions (including West Texas region) as shown in Figure 1.

PMF is applied to each group to generate profiles of source factors. Normalized source profiles and the quantitative source contributions for each resolved factor are calculated. The major sources that contribute to the aerosol loadings in the western United States are identified. The similarities and differences of chemical source profiles, and the major aerosol source contributors in different regions of the western United States are investigated. Based on the profile and the daily contribution to aerosol concentration of each source factor, the contributions of source factors to the aerosol light extinction coefficients are estimated using the IMPROVE method. The importance of the major aerosol sources to regional haze and visibility in the Class I areas of the western United States is discussed.

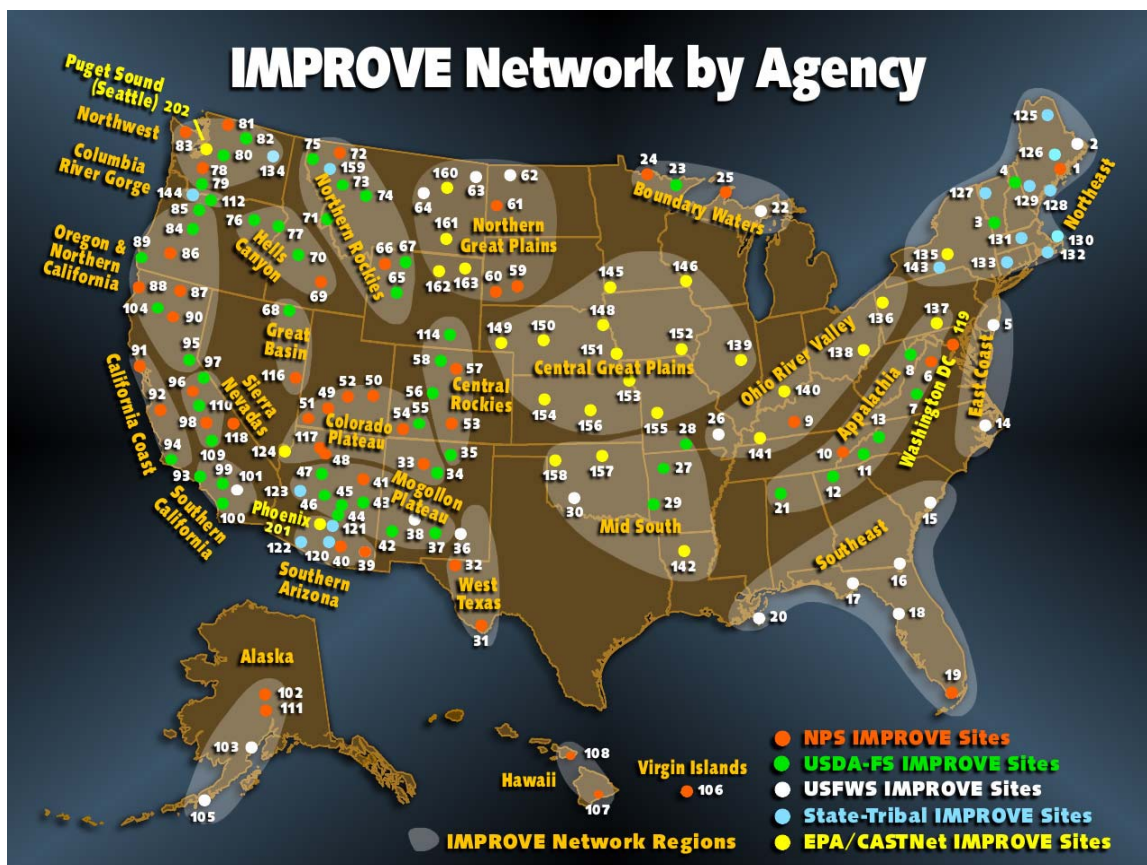


Figure 1. Grouping of Class I areas.

## 2 Model description

PMF is a statistical method that identifies a user specified number of source profiles (i.e. relative composition particle species for each source) and source strengths for each sample period that reduce the difference between measured and PMF fitted mass concentration. Equations 1 – 4 show the major steps of the model calculation.

$$X = GF + E \quad (1)$$

where:  $\mathbf{X}$  ( $n * Sp$ ) = a matrix of observed fine particulate species concentrations with the dimensions of number of observations by the number of species

$\mathbf{G}$  ( $n * f$ ) = a matrix of source contributions by observation day with the dimensions of number of observations by the number of factors

$\mathbf{F}$  ( $f * Sp$ ) = a matrix of source profiles with the dimensions of number of factors by the number of species

$\mathbf{E}$  ( $n * Sp$ ) = a matrix of random errors with the dimensions of number of observations by number of species

The F and G matrices of the final solution are then normalized according to the following equations to determine the quantitative source contributions ( $C_i$ ,  $\mu\text{g}/\text{m}^3$ ) and profiles for each source ( $S_i$ ,  $\mu\text{g}/\mu\text{g}$ ).

$$S_i = \frac{F_{ij}}{FM_i} \quad (2)$$

where:  $S_i$  = the row of the source profile matrix for source i

$F_{ij}$  = the source profile value for specie j of source i

$FM_i$  = the calculated average total fine mass contribution for source i

$$C_i = G_{ki} * FM_i \quad (3)$$

where:  $C_i$  = the column of the source contribution matrix for source i

$G_{ki}$  = the source contribution on day k for source i

$FM_i$  = the calculated average total fine mass contribution for source  $i$

The  $FM_i$  is determined by regression total  $PM_{2.5}$  mass concentrations in the  $k$ th sample ( $m_k$ ) against estimated source contribution values.

$$m_k = \sum_{i=1}^f G_{ki} * FM_i \quad (4)$$

## 2.1 Input data

1. Download aerosol  $PM_{10}$  and  $PM_{2.5}$  mass and chemical speciation data including analytical uncertainty and minimum detection limit from the VIEWS web site.
2. Data are screened to remove the days when either  $PM_{10}$  or  $PM_{2.5}$  mass concentration is missing.
3. Data value and associated uncertainty are calculated as:

### **If data is missing Then**

data value = geometric mean of the measured values

uncertainty = 4 \* geometric mean of the measured values

### **Else if data bellows detection limit**

data value = 1/2 \* detection limit

uncertainty = 5/6 \* detection limit

### **Else**

data value = measured data

uncertainty = analytical uncertainty + 1/3 \* detection limit

## 2.2 Major Control parameters

1. **Running Mode:** Robust Mode, the value of outlier threshold distance  $\alpha = 4.0$  (i.e. if the residue exceeds 4 times of the standard deviation, a measured value is considered outlier). The least squares formulation thus becomes:

$$Q = \sum_{i=1}^n \sum_{j=1}^m (e_{ij}/h_{ij}s_{ij})^2,$$

where

$$h_{ij}^2 = 1, \text{ if } |e_{ij}/s_{ij}| \leq \alpha; h_{ij}^2 = |e_{ij}/s_{ij}|/\alpha.$$

2. **Error Mode** (decides the standard deviation of the data  $S_{ij}$ ):

EM = -12 (based on observed value)  $S_{ij} = T_{ij}$

3. **FPEAK and FKEY Matrix (controls the rotation)** – default: 0 (central), may try different numbers

### 2.3 Output data

1. Source (factor) profiles for each group of sites.

Source (factor) profiles (see example as shown in Figure 2) which represent the fraction of each aerosol component to total  $PM_{2.5}$  mass for PMF defined sources (factors) are calculated for each group of sites.

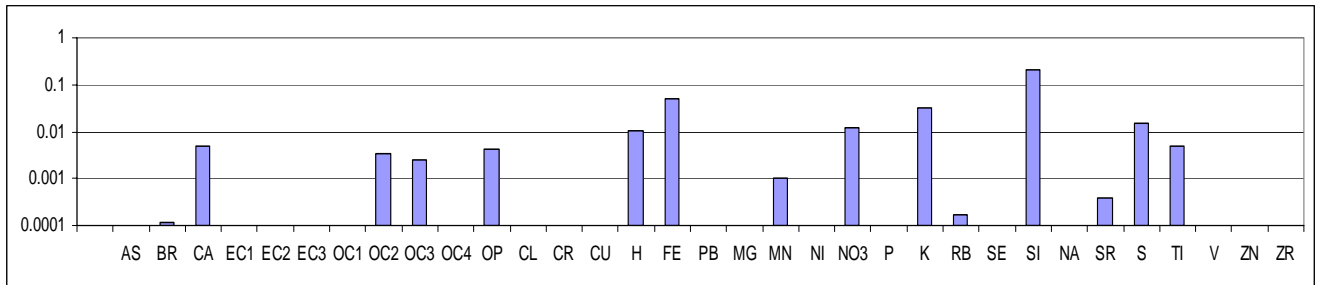


Figure 2. Example source (factor) profile (unit:  $\mu\text{g}/\mu\text{g}$ ) based on PMF modeling (dust factor at GRCA2)

2. Contribution of each source (factor) to aerosol mass and light extinction for each sampling day.

For each IMPROVE site in the Class I areas of western United States, contributions of each source (factor) to aerosol mass and light extinction for each sampling day are calculated (see example as shown in Figure 3).

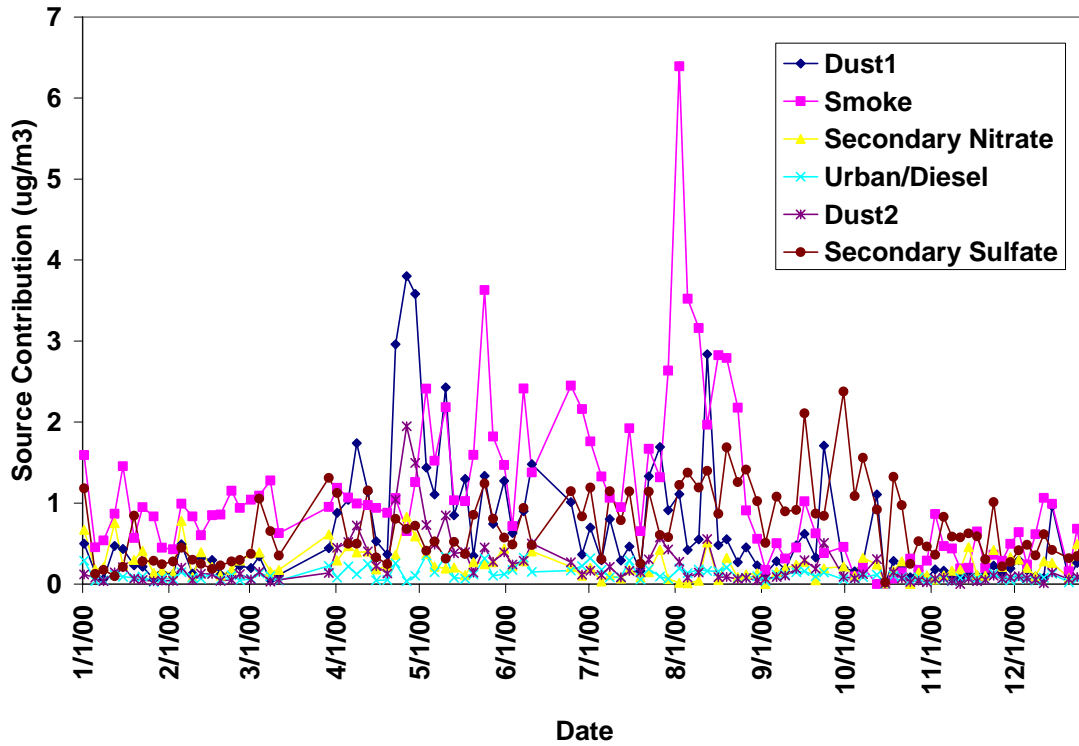


Figure 3. Contributions of major aerosol sources to  $PM_{2.5}$  mass in GRCA2 during the sampling days of 2000.

#### 2.4 Determination of the number of sources (factors)

For each group of sites, PMF model is run with 5 to 10 factors. Then the following methods are used to decide how many factors (sources) should be chosen for each area.

1. The regression coefficients  $FM_i$  should be positive. If the regression produces a negative coefficient, it suggests that maybe too many source factors have been used.
2. The sum of the scaled profiles should be less than unity. If any of them is much higher than unity (i.e.  $>2$ ) it suggests that too few sources may have been chosen.
3. Other PMF output parameters:

a. Q value

b. 
$$IM = \max_{j=1..m} \left( \frac{1}{n} \sum_{i=1}^n r_{ij} \right)$$
 where  $r_{ji} = e_{ij} / S_{ij}$

c. 
$$IS = \max_{j=1..m} \left( \sqrt{\frac{1}{n-1} \sum_{i=1}^n (r_{ij} - \bar{r}_j)^2} \right)$$

d. Largest element in Rotmat matrix

4. Judgement based on literature, known CMB source profiles, and experience

## 2.5 Coarse particles

Try to assign coarse mass to source factors using multi-linear regression. If no reasonable results can be achieved, coarse particles may be considered individually when calculating reconstructed extinction.

## 3 Analysis of PMF modeling results

### 3.1 General discussion

Run PMF and interpret the results for groups of sites in the WRAP. For selected regions, run on individual sites and groups to evaluate the effects of grouping (sensitivity analysis). Discuss the major sources of aerosols in each Class I area, and their contributions to aerosol mass and light extinction based on PMF modeling.

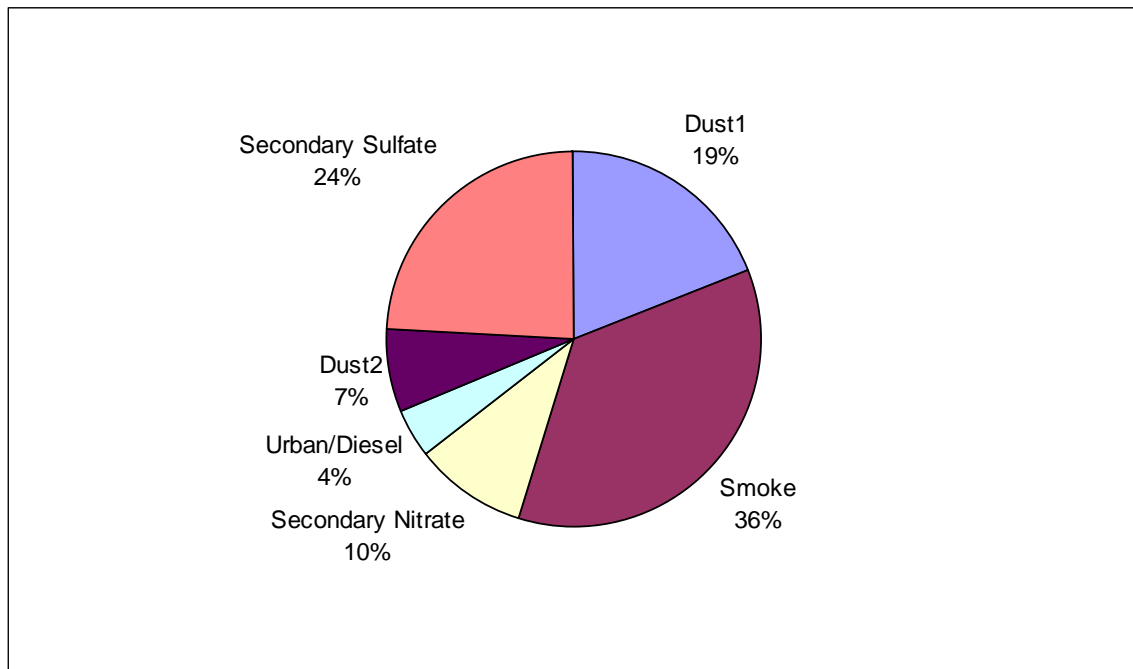


Figure 4. Average Contribution of Each Factor to GRCA2 PM<sub>2.5</sub> Mass during 2000-2004.

### 3.2 Trajectory analysis

Generate difference (see example in Figure 5) and ration maps of PMF factor weighted and un-weighted residence time to show the major source regions of sources (factors) defined by PMF. Compare the results with the emissions inventories data to evaluate level of confidence in results.

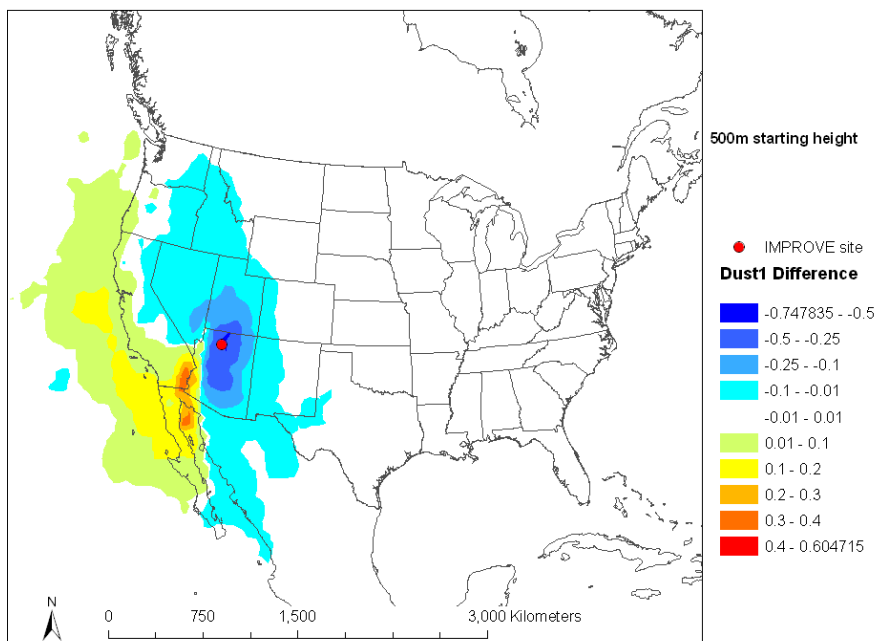


Figure 5. Difference map of PMF dust factor weighted and un-weighted residence time at GRCA2

### 3.3 Spatial and temporal analysis

Carry out spatial and temporal analysis of PMF results. Regroup the sites using aerosol data and PMF modeling results if it is found to be necessary.

### 3.4 Episode analysis

Investigate the robustness of the PMF results through case study (episode) analysis (e.g. for days with high factor loadings for specific factors, such as carbon based factors).

### 3.5 Other analysis

Organics is one of the major aerosol chemical components in the western United States. Preliminary analysis suggests that a large portion of the organics is from forest fires (smoke). But other sources such as mobile emissions and secondary organics may also have significant contributions to organic aerosols in the atmosphere. For carbon-based factors, more detailed analysis will be done than for the other factors, including:

1. Run PMF for a couple of selected sites: some of them with known big contributions from smoke and the others without. Compare the PMF results.
2. Calculate the percentage contribution of each factor to OC day by day. Compare the average percentage contributions of “smoke” and other factors to OC in the worst, typical and best days.
3. Investigate the relationship of OC / EC and the loading of the “smoke” factor.
4. Look at the time series of the factor contributions, especially those of the “smoke” factor. Pick out the highest 20% “smoky” days, and days when OC is the highest 20% while “smoke” factor loading is relatively small (e.g. in the low 50%) of 2002 for the selected sites and compare with the fire database (i.e. WRAP 2002 fire emissions inventory).
5. Check out back trajectory maps of those days picked out in #4 to see if air is from the fire area.