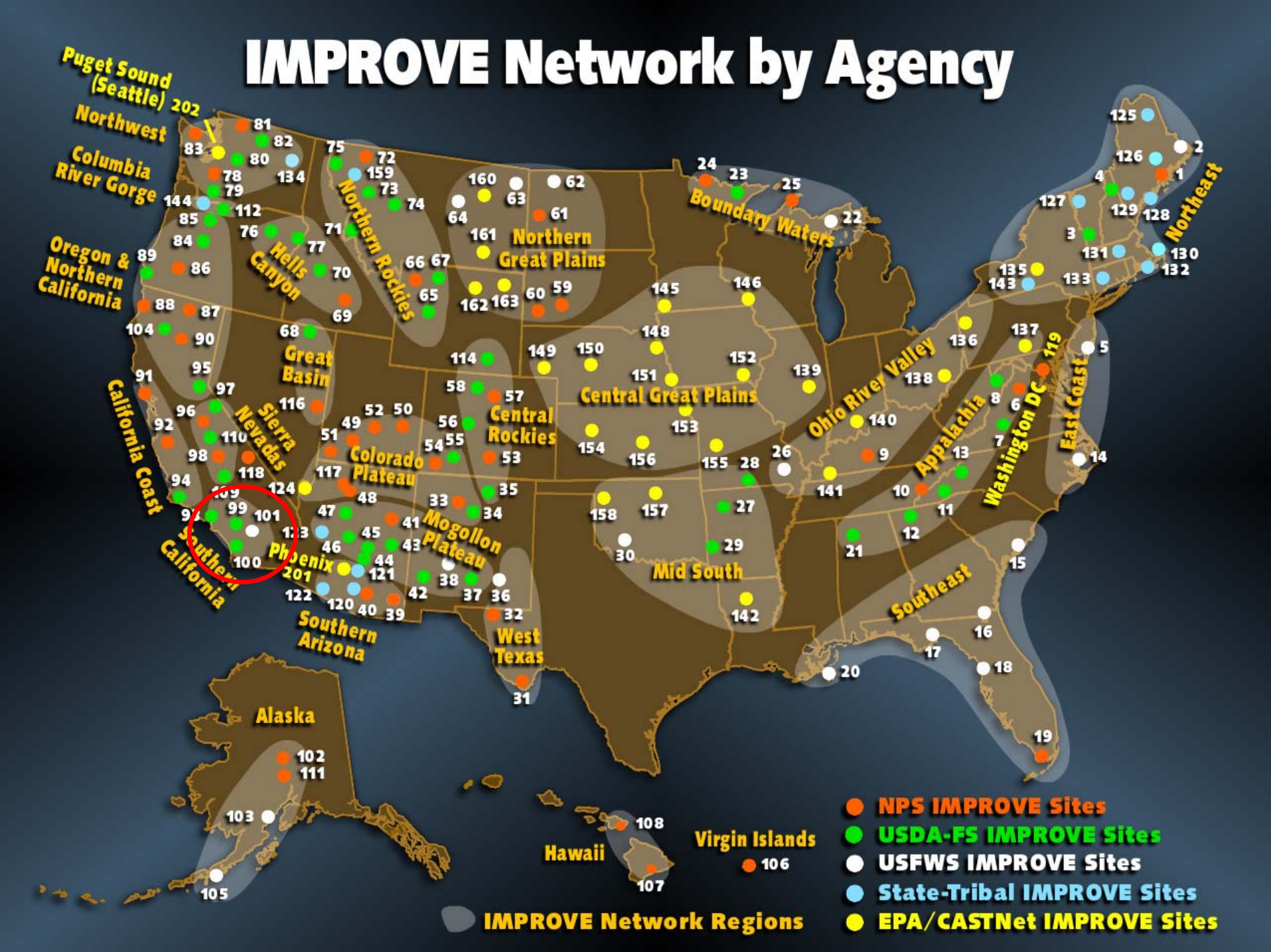
COHA Update

Jin Xu

Update

- 2003 and 2004 back-trajectories – done
- PMF modeling by groups using 2000 to 2004 IMPROVE data – done
- Analysis of PMF results – ongoing
  - General analysis and discussion – decide how many factors are reasonable for each group, will finish soon. More modeling calculation will be done if necessary.
  - Trajectory analysis – ongoing
  - Spatial and temporal analysis – ongoing, may result in regrouping of the sites and more PMF modeling
  - Episode analysis
  - Other analysis – carbon-based factors – ongoing
    - 2002 fire database from WRAP, other years from Dr. Tim Brown's group in DRI. Satellite data and images archived.
    - Case study
    - Similar trajectory analysis as for the causes of dust resultant haze


IMPROVE Network by Agency
Puget Sound (Seattle) 202
Northwest
Columbia River Gorge
Oregon & Northern California
Hells Canyon
Northern Rockies
Northern Great Plains
Boundary Waters
Northeast
Great Basin
Sierra Nevadas
California Coast
Southern California
Colorado Plateau
Central Rockies
Central Great Plains
Ohio River Valley
Appalachia
Washington DC
East Coast
Mogollon Plateau
Phoenix
Southern Arizona
West Texas
Mid South
Southeast
Alaska
Hawaii
Virgin Islands
106
IMPROVE Network Regions
NPS IMPROVE Sites
USDA-FS IMPROVE Sites
USFWS IMPROVE Sites
State-Tribal IMPROVE Sites
EPA/CASTNet IMPROVE Sites

Receptor Modeling - Positive Matrix Factorization (PMF) and Chemical Mass Balance (CMB)

- Mathematical technique for determining the contributions of various sources to a given sample of air

$$\begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_i \end{bmatrix} = \begin{bmatrix} SP_{11} & SP_{12} & \cdots & SP_{1j} \\ SP_{21} & SP_{22} & \cdots & SP_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ SP_{i1} & SP_{i2} & \cdots & SP_{ij} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_j \end{bmatrix}$$

SP_{ij} – Source Profile: Emissions of compound i from source j (100%).

I_j – Contribution of source j ($\mu g/m^3$).

C_i – Concentration of compound i ($\mu g/m^3$).

	CMB	PMF
Input	Both C and SP	Only C
Output	Only I	Both SP and I

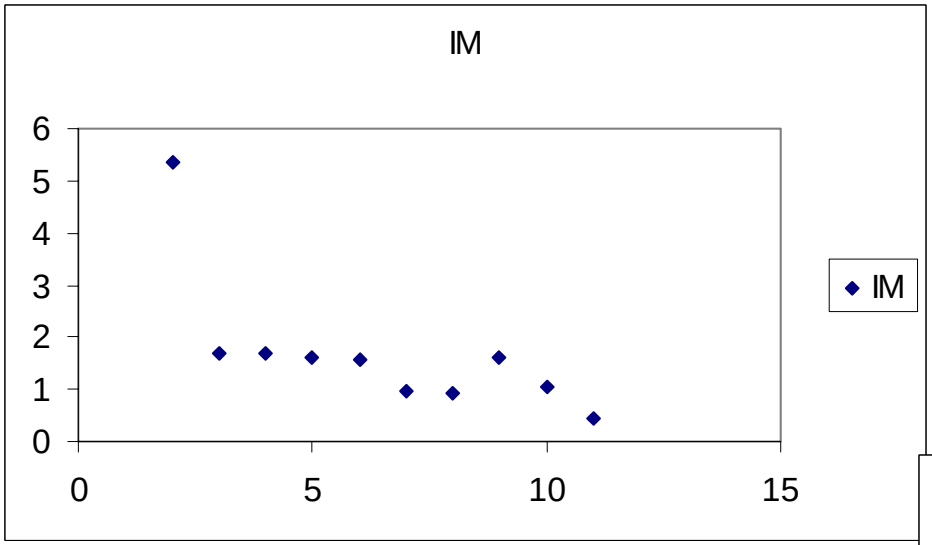
Receptor Modeling - Positive Matrix Factorization (PMF) and Chemical Mass Balance (CMB) (Cont.)

	CMB	PMF
Assumptions	Composition of source emissions is relatively constant Emissions do not react or selectively deposit between source and receptor (mass is conserved) Source profiles are linearly independent For CMB, all major sources should be included in the model	
Limitations	Reactive compounds Only identifies categories of sources, not individual sources Identifies only relative contributions, not mass emission rates	
Limitations	Must know source profiles High sensitivity to uncertainty / error in source profiles Omission of a source can lead to large errors	Pure statistical model large number of samples (100+) are needed Need to make arbitrary decision of the number of sources (factors)

Number of Factors – Southern CA

- No negative regression coefficient(s) between $PM_{2.5}$ mass and G factors.
- The sum of the scaled profiles should be less than unity (well, < 2).
- Other PMF output parameters:

Number of Factors – Southern CA



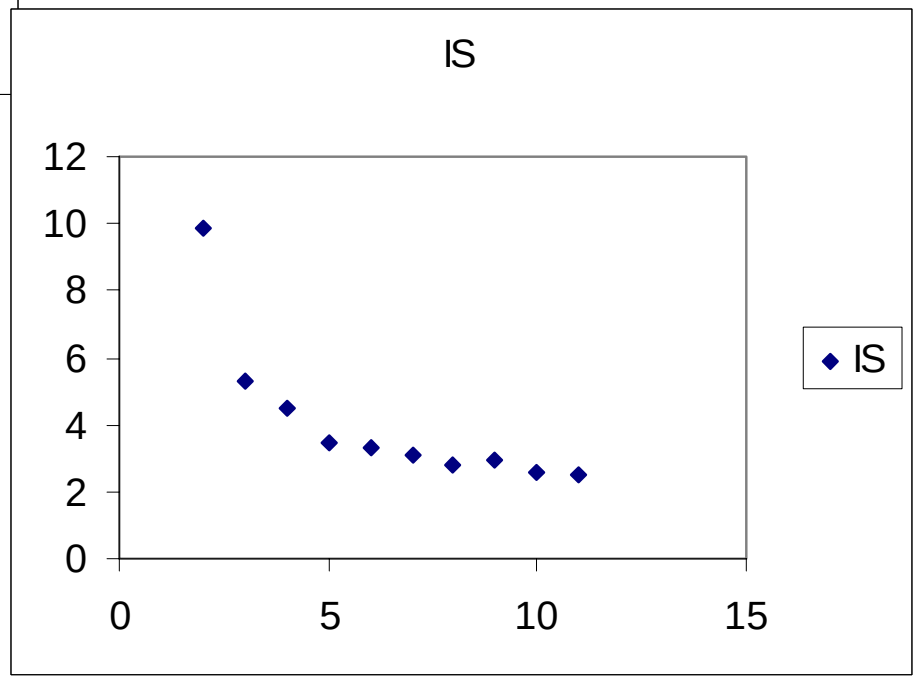
$r_{ij} = e_{ij} / S_{ij}$, while $e_{ij} = x_{ij} - GF$

$$IM = \max_{j=1..m}\left(\frac{1}{n}\sum_{i=1}^{n} r_{ij}\right)$$

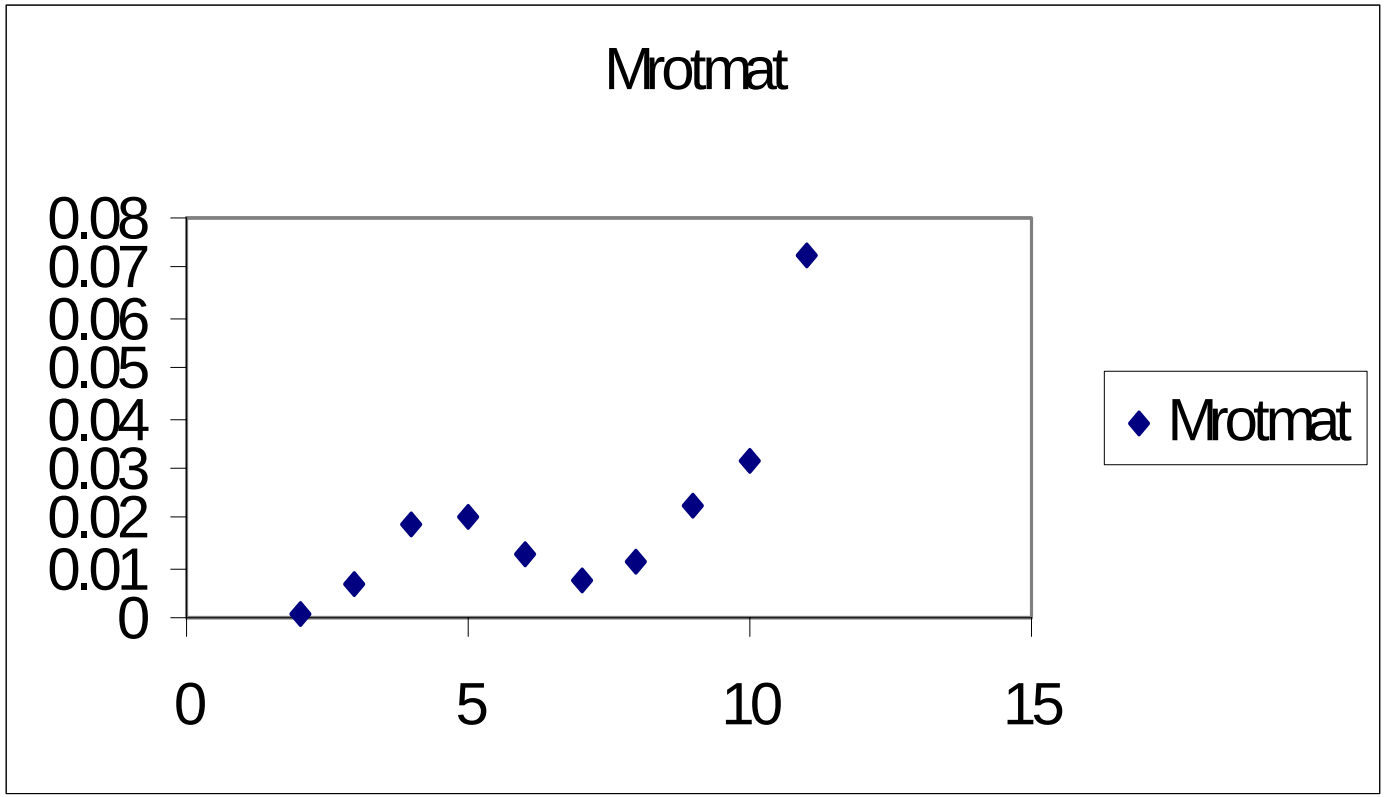
Species having the least fit

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

Species having the most imprecise fit



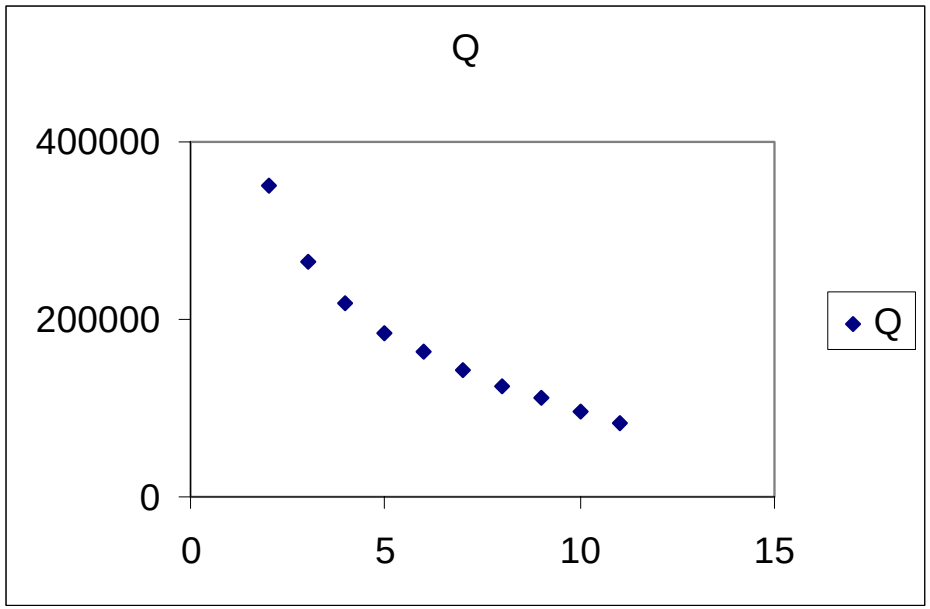
Number of Factors – Southern CA



Rotmat – a matrix resulting from each PMF computation, is used for detecting the degree of rotational freedom of the factors

Largest element in rotational matrix (Mrotmat) is used to show worst case in rotational freedom

Number of Factors – Southern CA



where

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

$$h_{ij}^2 = 1, if \ \left|e_{ij}/s_{ij}\right| \le \alpha; h_{ij}^2 = \left|e_{ij}/s_{ij}\right|/\alpha.$$

If there are no outliers, Q should be approximately equal to the number of entries in the data array. Possible reasons for an excessively large Q:

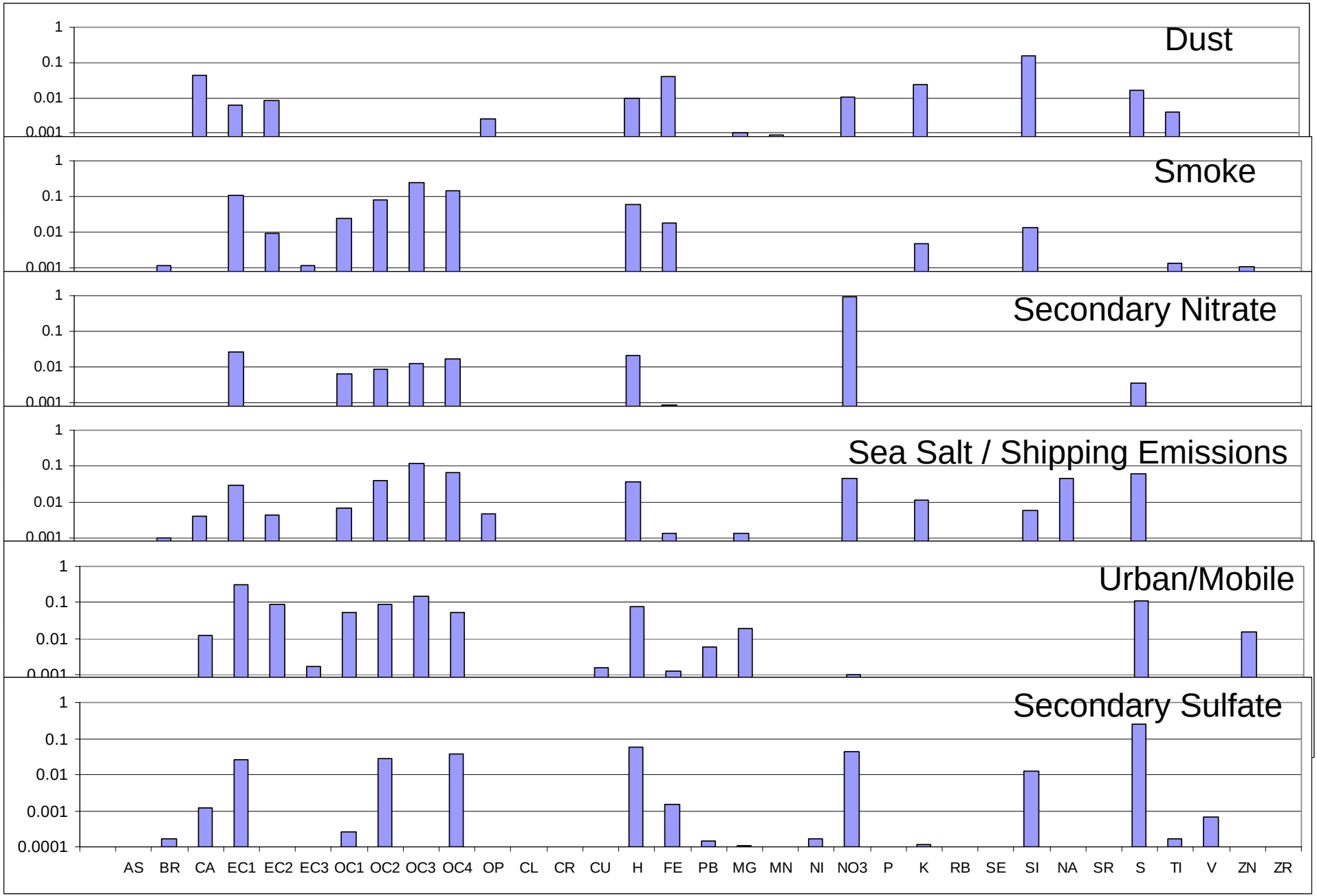
1. The original std-dev are too small
2. There are many outliers
3. More factors are needed
4. The data do not obey a bi-linear model, i.e. PMF is not a suitable model
5. The iteration did not converge or converged to a local minimum.

For a 6 factor modeling, 3.6% of the data entries are outliers, i.e. $e_{ij}/s_{ij} > 4$.

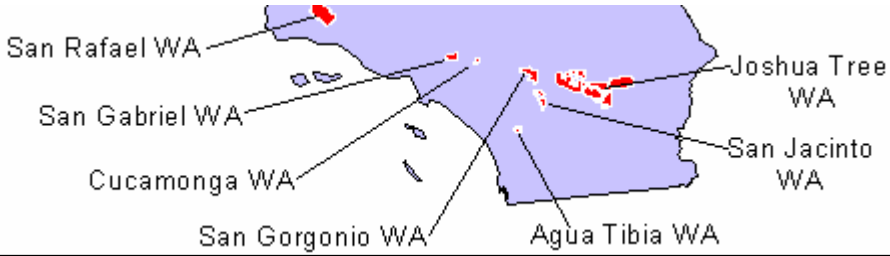
Number of Factors – Southern CA

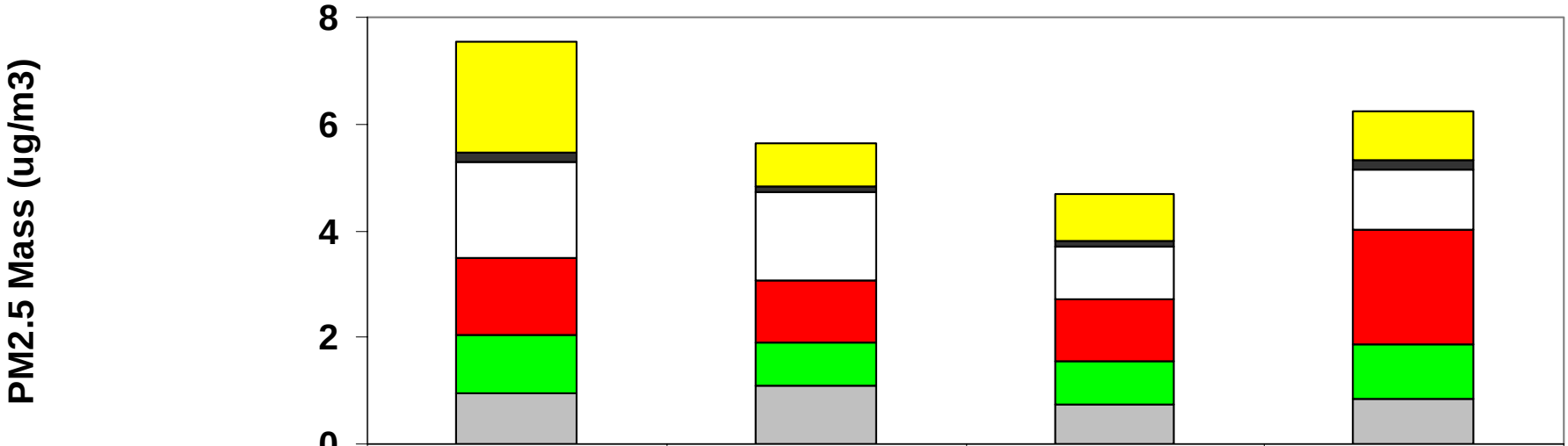
- Factor 5 – 7 should be used.
- How many factors between 5 and 7 should we choose? – Judgement based on literature, known source profiles, and experience

PMF for Southern CA – 6 factors



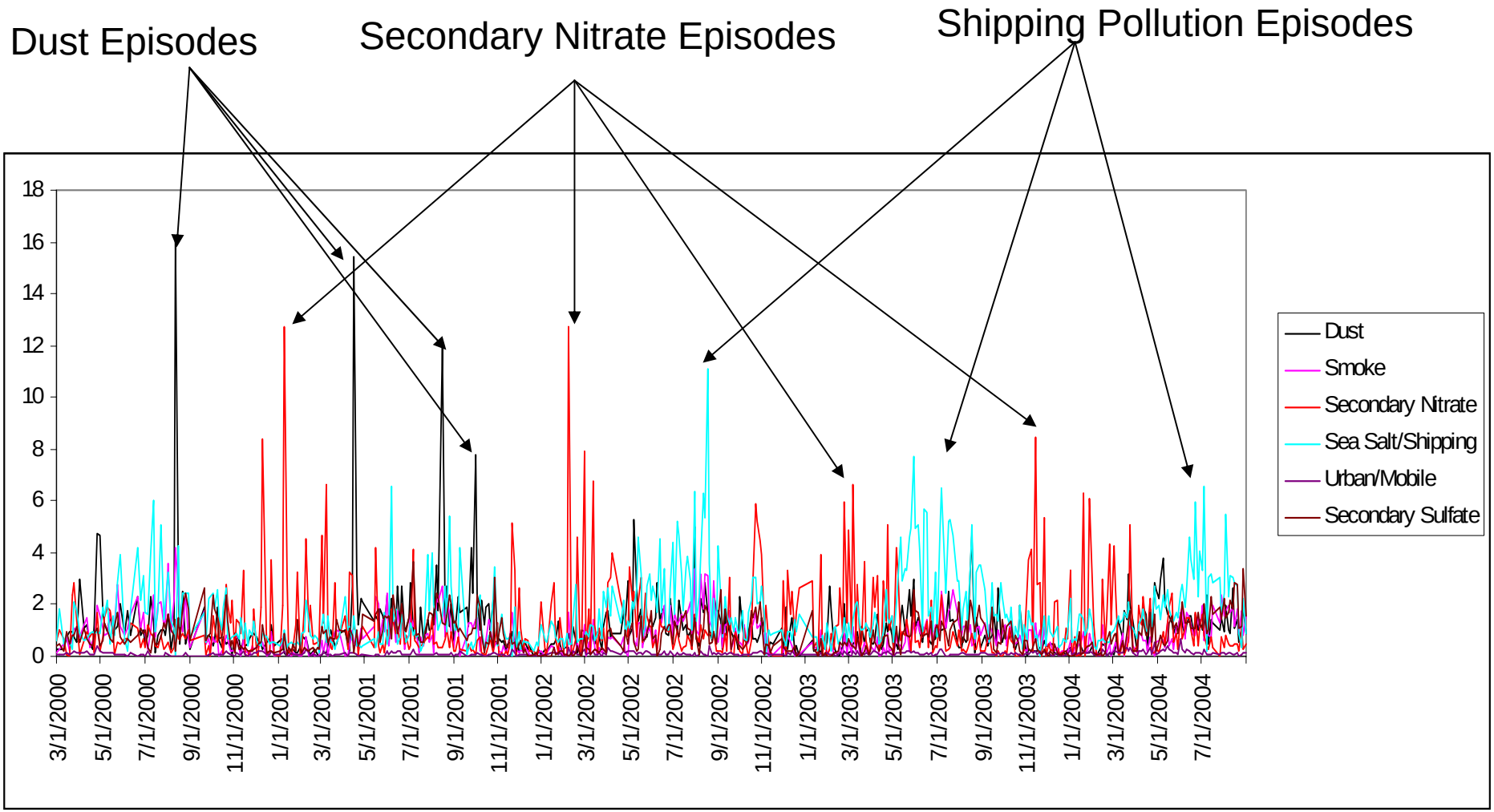
Average Contribution of Each Factor to $PM_{2.5}$



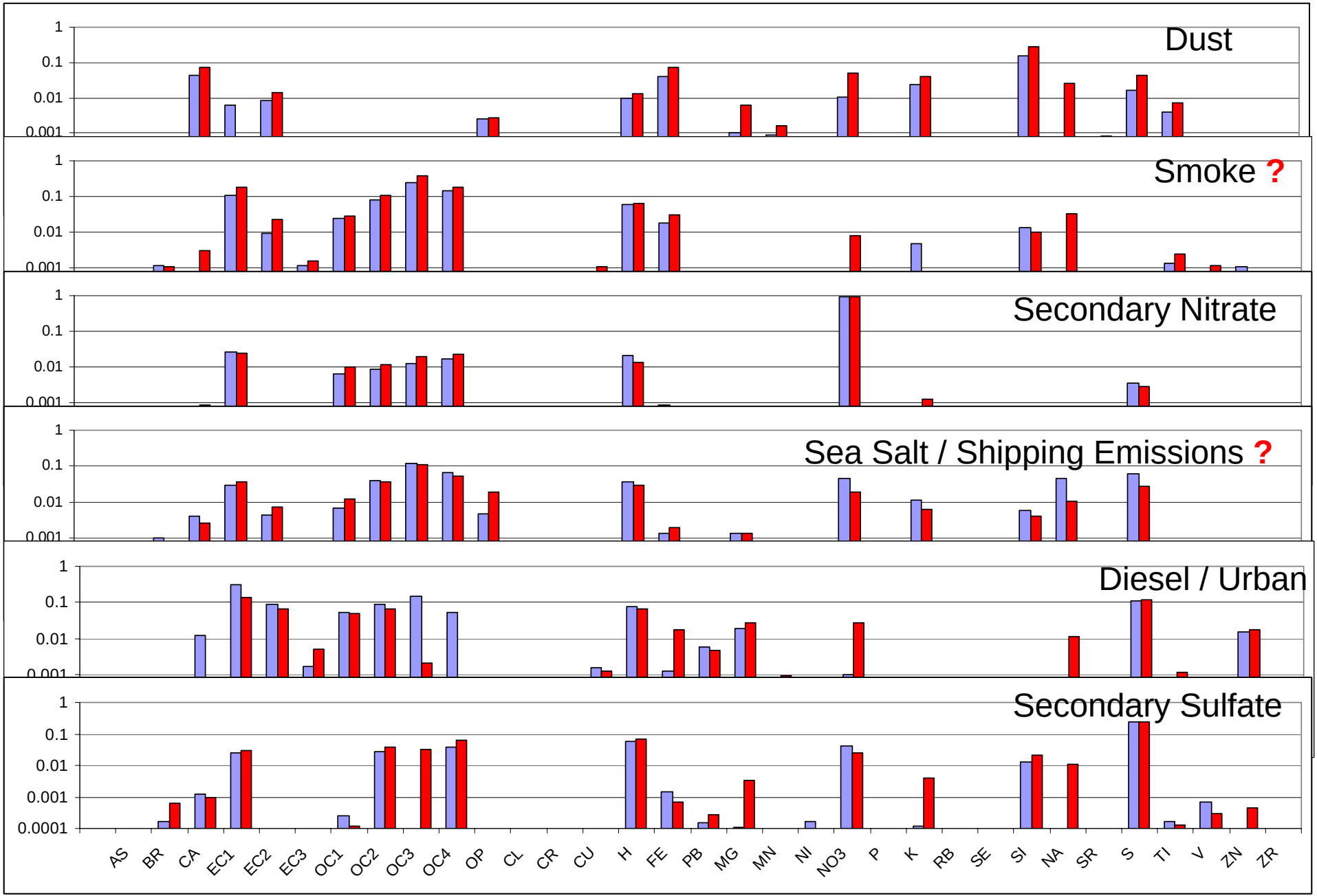


	AGTI1	JOSH1	SAGA1	SAGO1
Secondary Sulfate	2.09	0.80	0.87	0.93
Urban/Mobile	0.15	0.10	0.14	0.16
Sea Salt/Shipping	1.80	1.65	0.98	1.13
Secondary Nitrate	1.46	1.18	1.15	2.15
Smoke	1.10	0.79	0.84	1.04
Dust	0.94	1.11	0.72	0.83

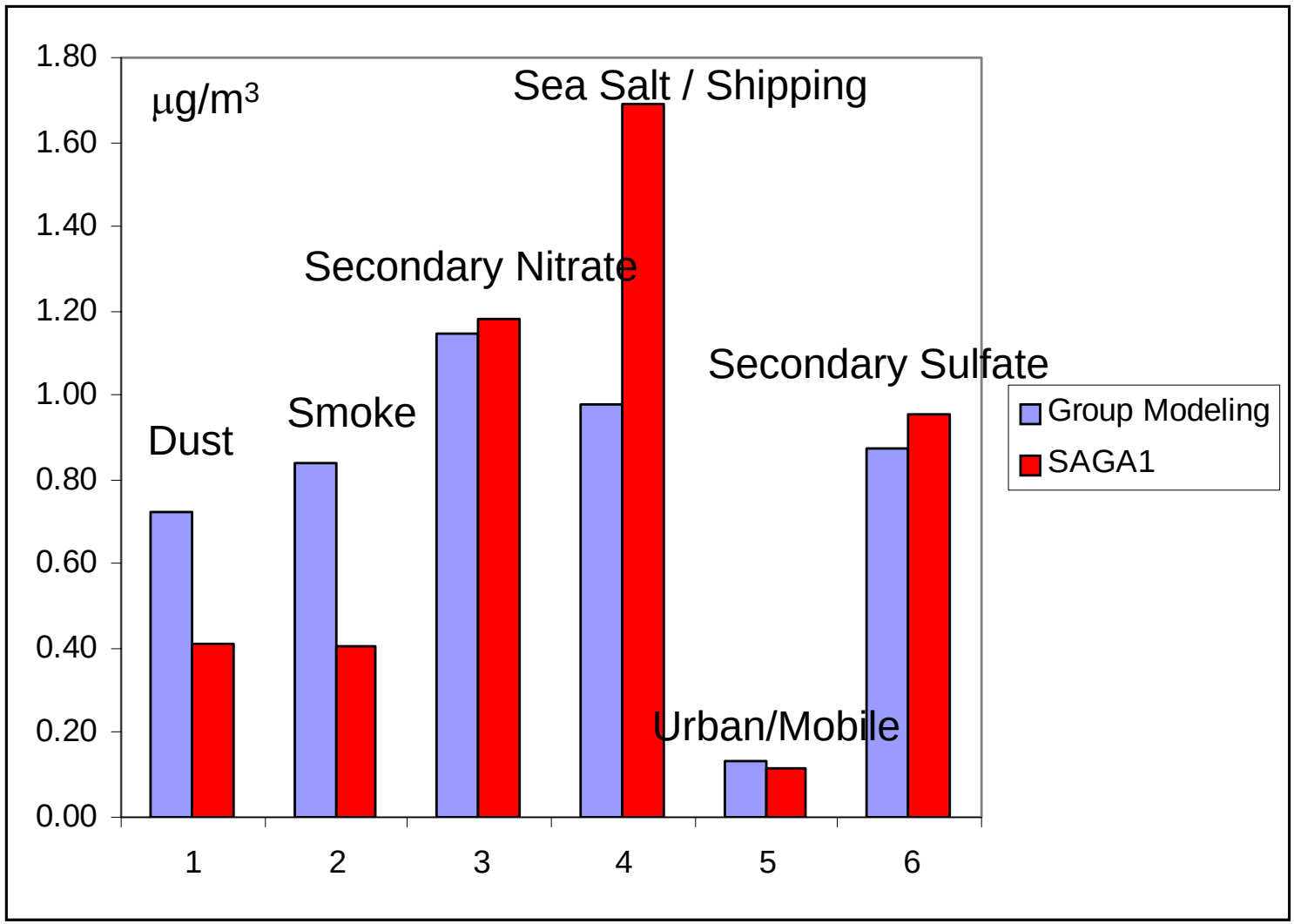
Time Series of Factor Contributions at JOSH1



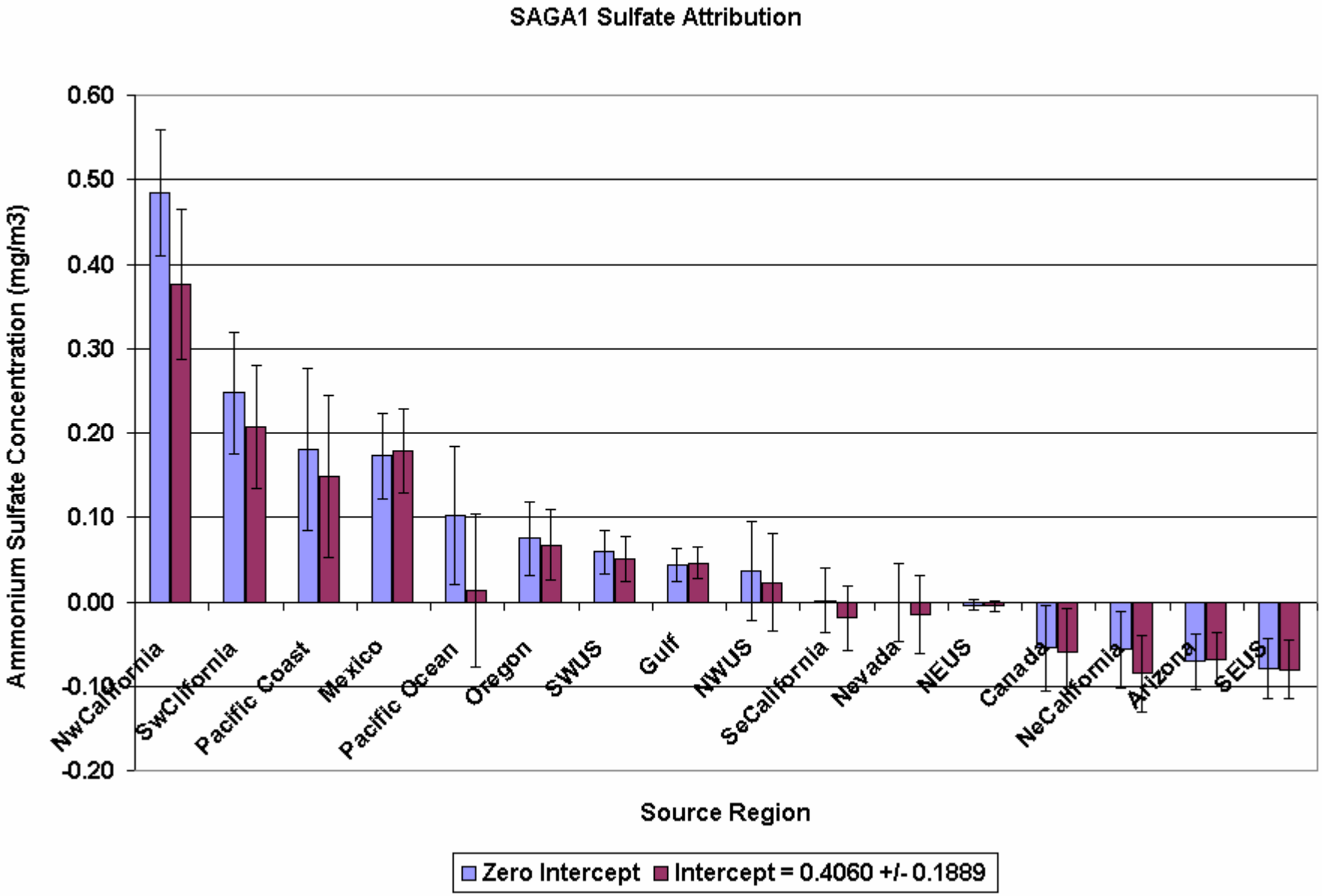
Comparison of Group Modeling (Blue) and SAGA1 Individual Modeling (Red)



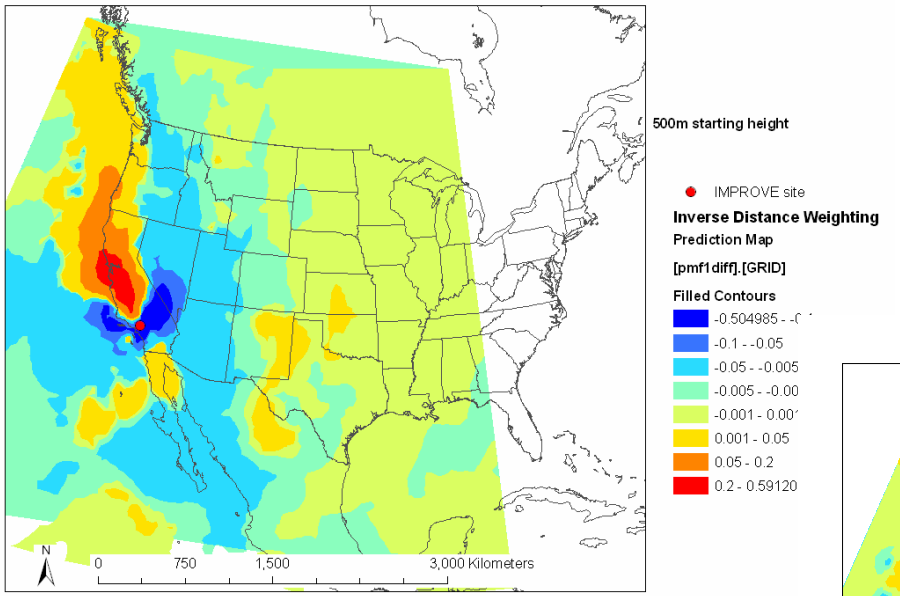
Contribution of Each Factor to PM2.5 Mass in SAGA1



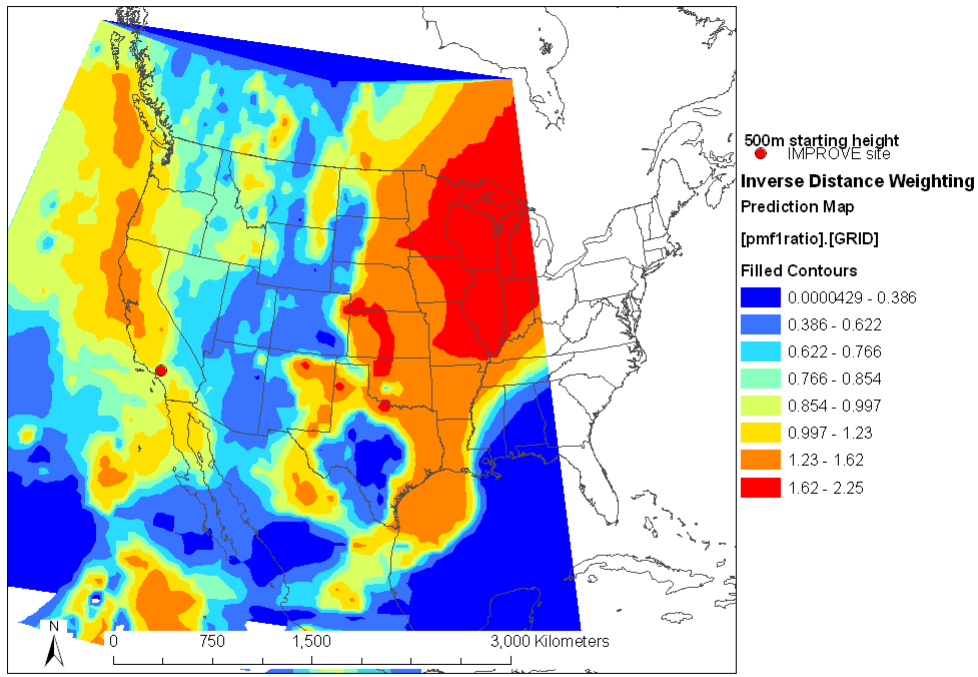
SAGA1 Trajectory Regression Analysis Results



Trajectory Analysis of SAGA1 Individual PMF Modeling Results

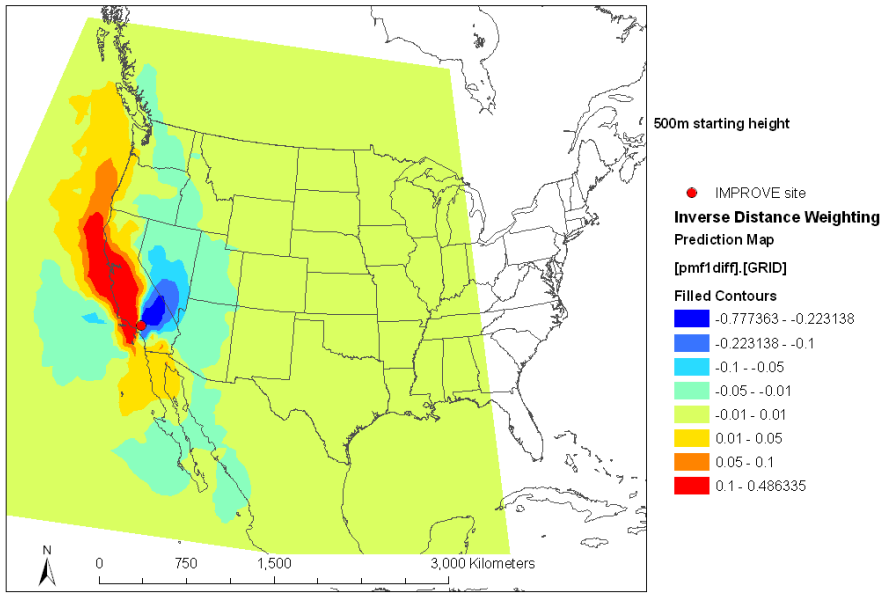


PMF Weighted - Unweighted

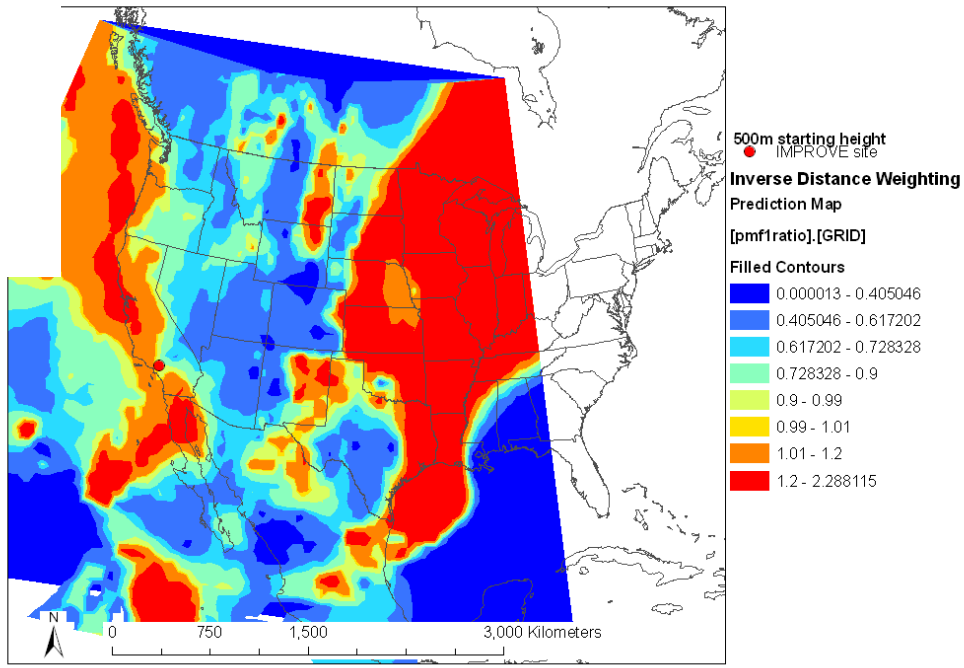


PMF Weighted / Unweighted

Trajectory Analysis of SAGA1 Group PMF Modeling Results

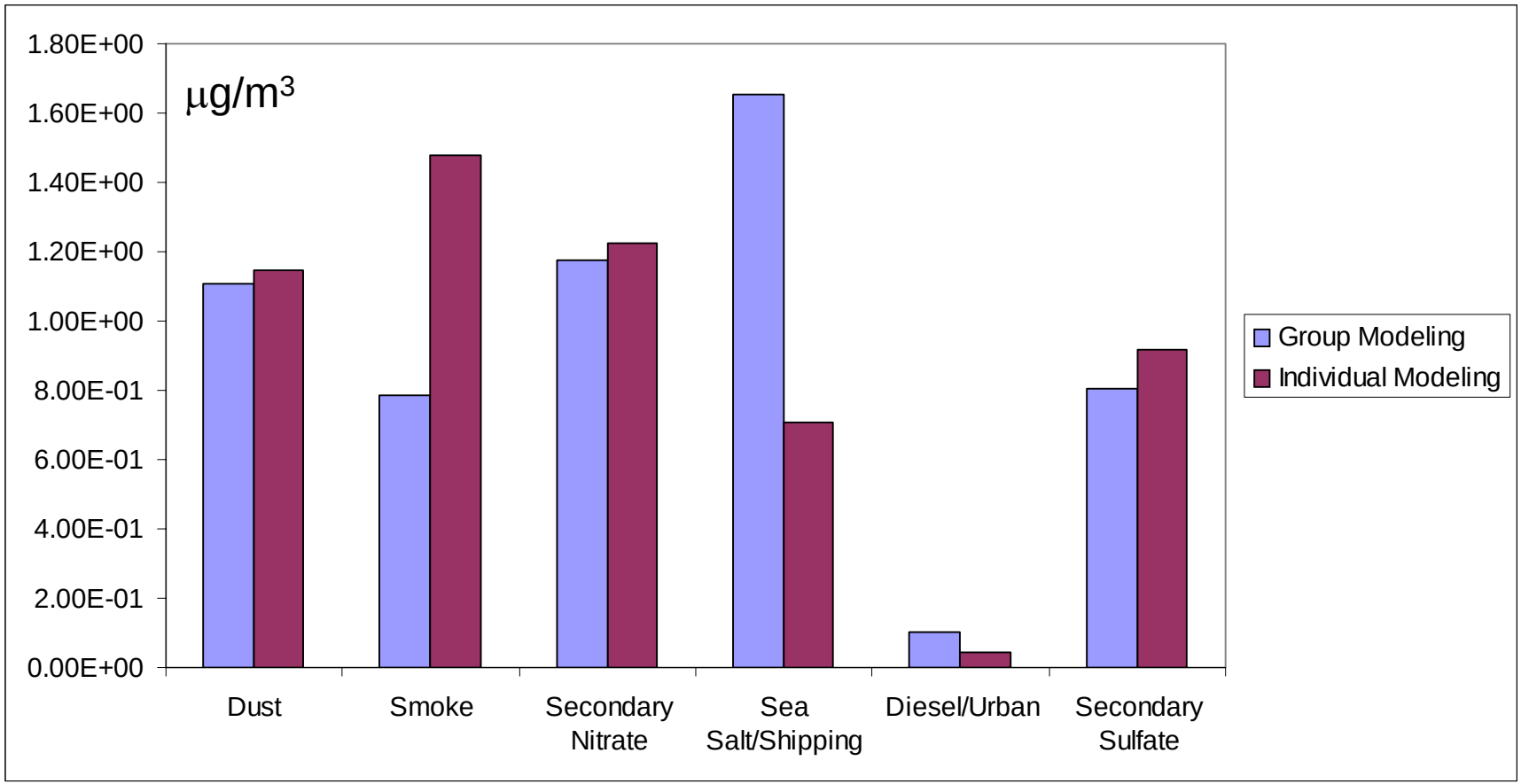


PMF Weighted - Unweighted

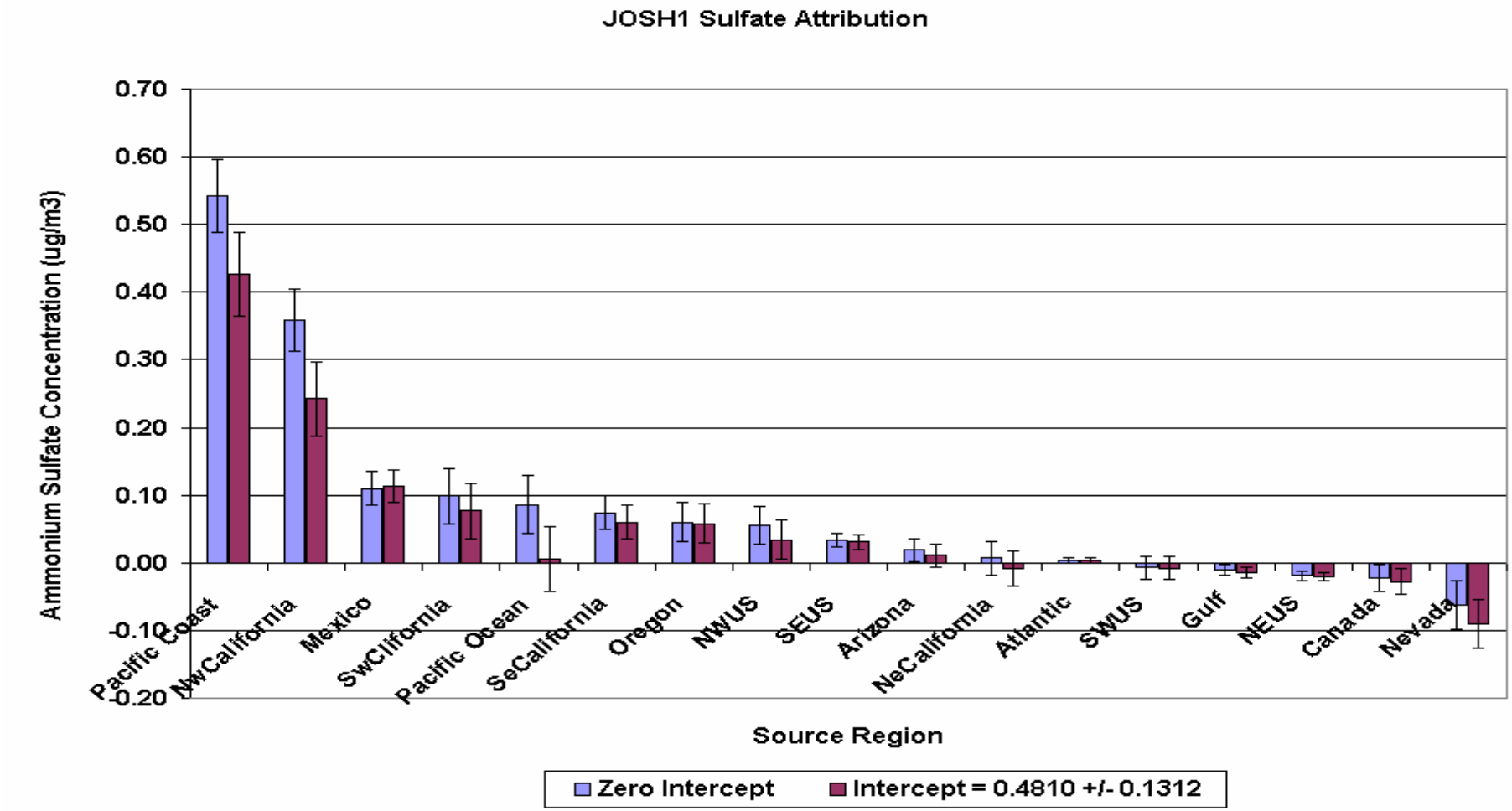


PMF Weighted / Unweighted

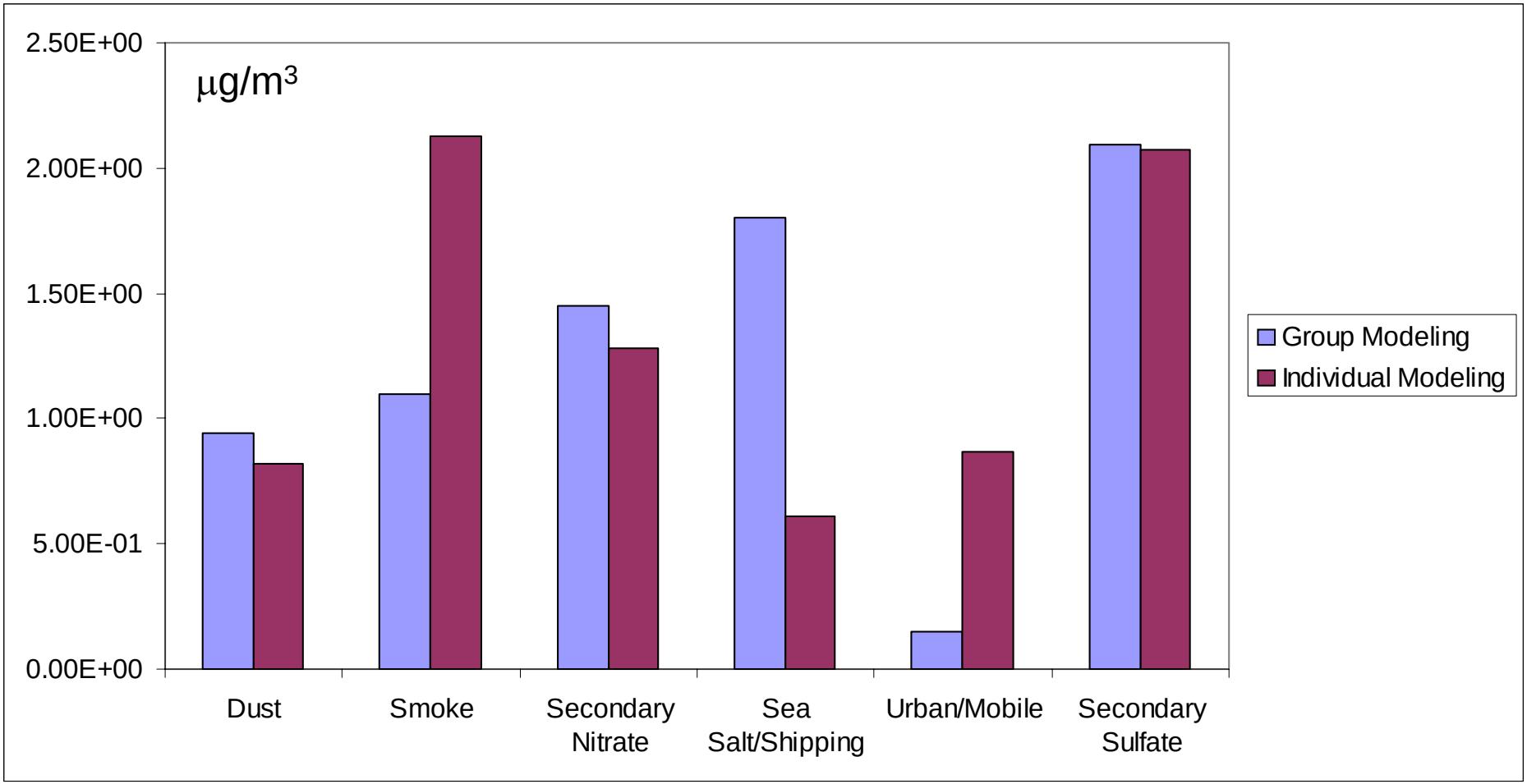
Contribution of Each Factor to PM2.5 Mass in JOSH1



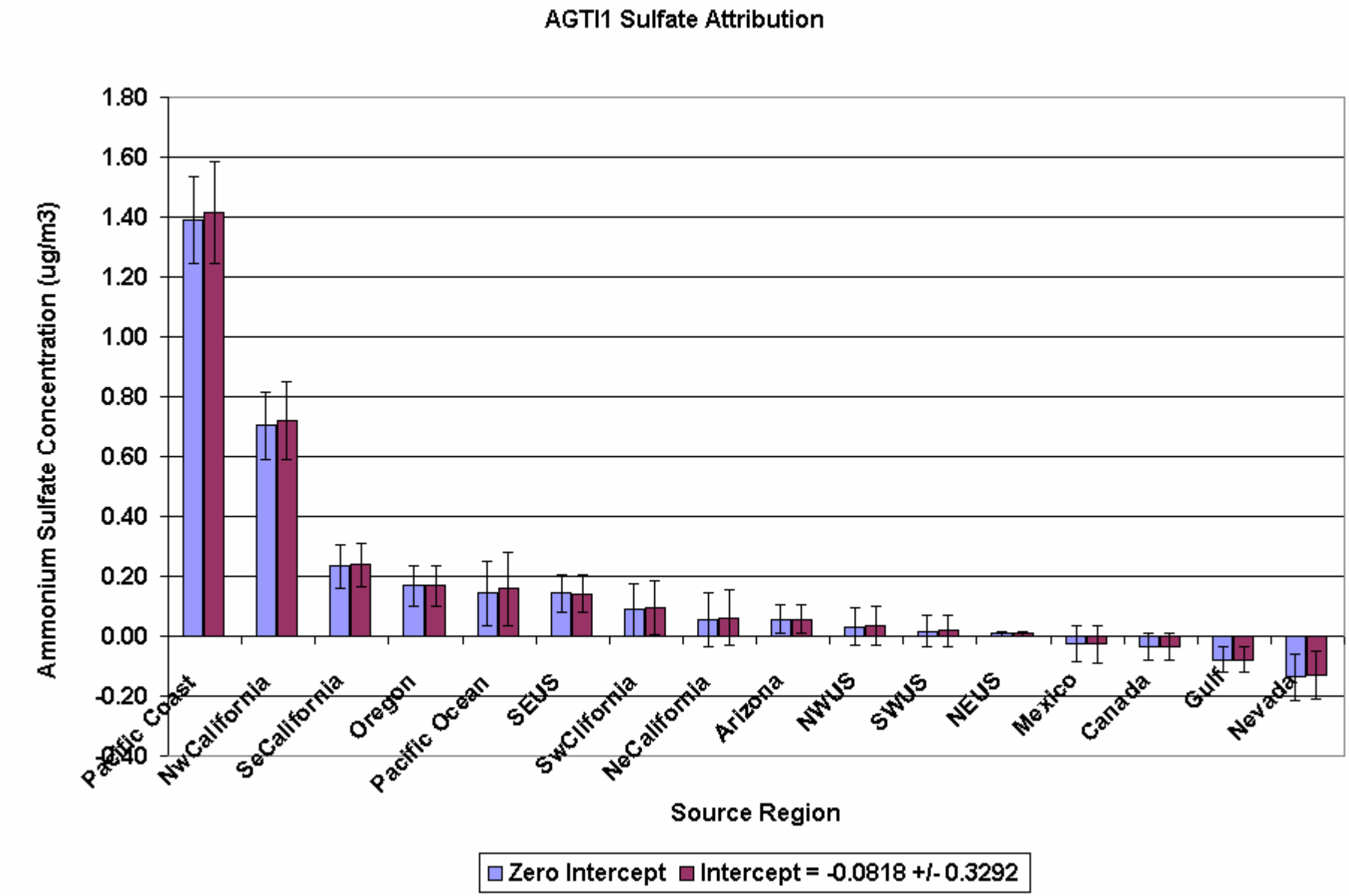
JOSH1 Trajectory Regression Analysis Results



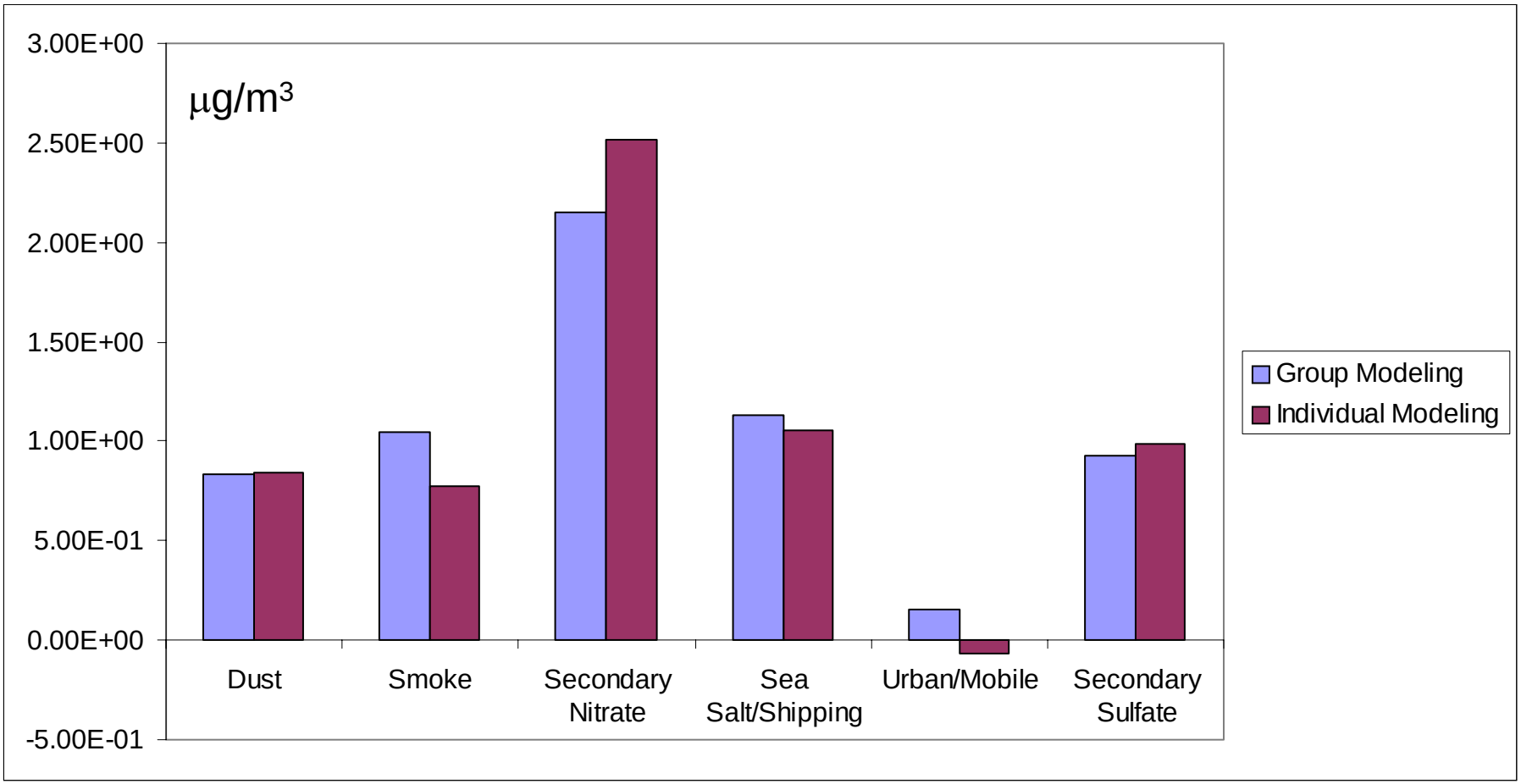
Contribution of Each Factor to PM2.5 Mass in AGTI1



AGTI1 Trajectory Regression Analysis Results



Contribution of Each Factor to PM2.5 Mass in SAGO1



Group modeling is doing a better job for the Southern CA group?

- Clearer factors due to strong source signatures in certain sites in the group (especially true for IMPROVE sites because they are in remote areas with well mixed pollutions)
- Partially solved the collinearity problem of some sources
- More data

Trajectory Analysis of PMF Results