

For Presentation at: International Symposium on Measurement of Toxic and Related Air Pollutants, Research Triangle Park, NC, September 12-14, 2000

Sources of Fine Particle Concentration and Composition in Northern Vermont

Richard L. Poirot

Paul R. Wishinski

Vermont Department of Environmental Conservation
Building 3 South, 103 South Main Street
Waterbury, VT 05671-0402

Philip K. Hopke

Alexander V. Polissar

Department of Chemical Engineering
Clarkson University
Box 5705
Potsdam, NY 13699-5705

ABSTRACT

This study applies and compares results of four receptor modeling techniques to a common set of IMPROVE-like, speciated fine particle measurement data collected at remote site in northwestern Vermont between 1988 and 1995. Two multivariate mathematical models - Positive Matrix Factorization and UNMIX - were applied to the measurement data, and identified seven "common" sources, which have similar compositions and similar fine mass contributions in both models. Two ensemble backward trajectory techniques - Potential Source Contribution Function and Residence-Time Analysis - are also applied to evaluate and interpret the mathematical model results. The trajectory techniques indicate a strong regional character to the upwind locations associated with high contributions identified independently by the mathematical techniques. These convergent results among the multiple methods provide a degree of confidence that each of the receptor methods can provide useful insights for air quality management, when applied to speciated fine particle data of this nature. Midwestern fuel combustion (with separate seasonal source characteristics), local Woodsmoke, and East Coast urban source influences collectively accounted for nearly 90% of the average fine mass concentrations at the receptor. Smaller contributions were also identified from Canadian Smelters, Canadian motor vehicles (and/or other Mn sources), and windblown soil.

INTRODUCTION

Concentrations of fine particles (< 2.5 microns) in the ambient air are typically composed of complex mixtures of chemical species, originating from a wide range of natural sources and human activities, distributed over transport distances of tens to thousands of kilometers from their origins. Recent US EPA regulatory programs for fine particles (PM_{2.5}) and regional haze have led to a dramatic expansion of ambient measurements of PM_{2.5} mass and chemical composition. Within the next few years, vast amounts of complex new data will become available, with associated mandates to use these data to develop efficient emissions control strategies to attain national health standards and to reduce regional haze in Class I federal lands.

Receptor models, which attribute pollution to sources through mathematical and/or meteorological interpretation of ambient measurement data, should prove to be useful PM_{2.5} air quality management tools, especially as the quantity and quality of speciated fine particle data increases, and as improved methods are developed for treating secondary aerosol formation in these models.¹ The Positive Matrix Factorization (PMF) model^{2,3} and UNMIX^{4,5} are two state-of-the-art multivariate mathematical models which have potential applicability to analysis of speciated fine particle data. These two models were recently applied to common data sets of artificial data (with known sources) and ambient measurement data (with unknown sources from the Phoenix urban area), with results compared and discussed at a recent EPA workshop.^{6,7}

In the current study, the PMF and UNMIX models are applied to a common set of IMPROVE-like speciated PM_{2.5} data from a remote background site in northwestern VT. The mathematical model results are compared and further evaluated by two ensemble backward air trajectory techniques: the Potential Source Contribution Function⁸ and Residence-Time Analysis.⁹ Through this “4-way” comparison, some degree of confidence in the usefulness of the different receptor modeling tools applied to this kind of measurement data can be gained in areas where model results are convergent. Conversely, areas of divergent model results identify topics for future research.

MEASUREMENT DATA

Fine particle measurements were conducted at a remote background site in Underhill, VT (44.53 N, -72.86 W, 400 m. elev.) between September, 1988 and June, 1995, as part of the NESCAUM Regional Particle Monitoring Network.^{10,11} Samples were collected on Teflon filters using one of the four modular samplers (Module A) routinely employed in the Interagency Monitoring of Protected Visual Environments (IMPROVE) network. Filters were subsequently analyzed at the Crocker Nuclear Laboratory, U. California, Davis (IMPROVE analytical laboratory) for mass (gravimetric), light absorption (B_{abs} - by laser integrating plate - LIPM), elemental hydrogen (by proton elastic scattering analysis - PESA), and multiple elements with molecular weights ranging from Na through Pb (initially by proton induced x-ray emission - PIXE, and starting June, 1992, by a combination of PIXE and x-ray fluorescence - XRF).

Samplers were run 24 hours, midnight to midnight, every Wednesday and Saturday (IMPROVE sampling schedule) and also every sixth day (routine EPA PM-10 sampling schedule), with a

resultant data set of 854 samples with 28 measurement variables (see Figure 1). These data include reported concentrations (in ng/m^3 for all variables except B_{abs} , which is in 10^{-8}m^{-1}) and analytical uncertainties for all concentrations above minimum detection limits (MDL). For attempted measurements below (and above) MDL, a varying MDL is quantified separately for each species and each sample. There are occasional “missing data” (no reported measurement), for one or more species (most often mass or B_{abs}) in 16 of the 854 daily samples. For use as model input, B_{abs} was reduced by a factor of 0.4 to convert from light absorption to an estimated mass concentration (of elemental carbon). This adjustment is based on the relationship between B_{abs} and EC at two nearby sites in the NESCAUM network where direct measurements of EC and B_{abs} were conducted concurrently with the Underhill, VT measurements. B_{abs} analytical uncertainty values were doubled to reflect the additional uncertainties associated with this mass conversion.

MATHEMATICAL MODELS AND METHODS

“Factor analysis”, employing a traditional mass balance approach to analysis of multivariate air pollution data, has provided useful insights on sources of gaseous hydrocarbons and speciated aerosols over the past several decades. No assumed knowledge of meteorology or emissions inventories is required, and the model identifies the chemical composition of the identified sources. However, as observed by Henry,¹² the traditional factor analysis approach is ill-posed, and can produce an large number of equally “correct” (mathematically) but different answers, many of which are physically implausible. PMF and UNMIX are refinements of the traditional factor analysis method, which employ different approaches to constrain results to a “best fit” solution, which is both mathematically and physically feasible.

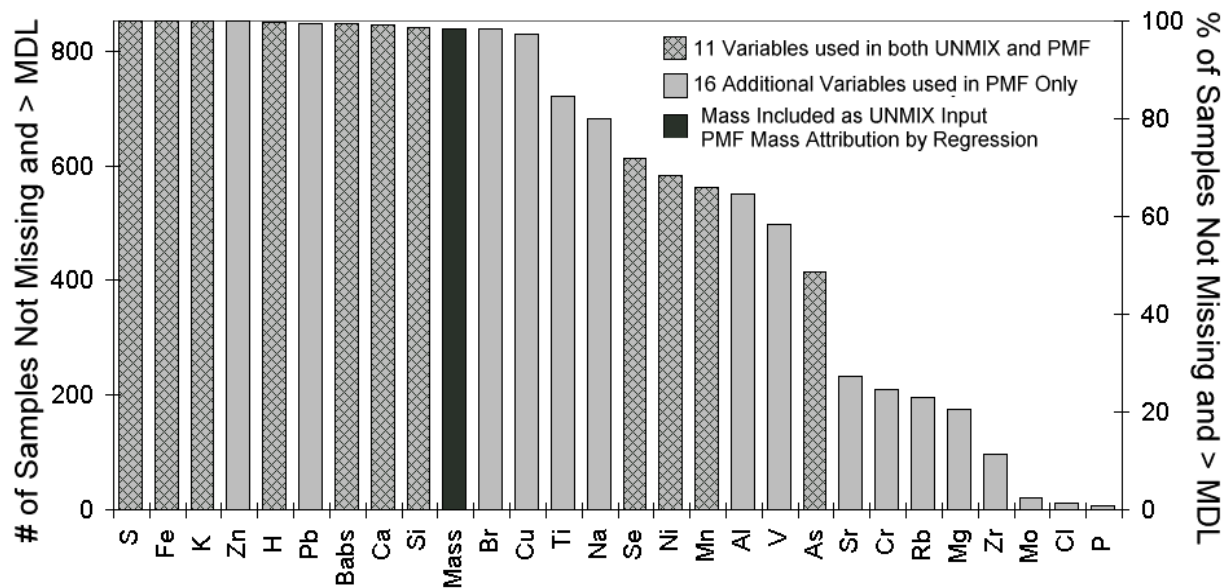
In the current study the PMF and UNMIX models were run independently - PMF by the Clarkson team, and UNMIX by the VT team - both starting with an identical raw data set composed of 854 samples, 28 variables, with associated MDLs and analytical uncertainties. PMF and UNMIX are described in detail elsewhere²⁻⁶ and are summarized in only general terms here, with emphasis on those differences between the models which resulted in use of different input data by the modelers. Both models, as applied here, are intended to identify the number of discernable sources of influence on the data, the source compositions, the daily contributions from each source to concentrations of fine mass (and of other species used as input to the models), and the associated uncertainties.

One key difference between the models is that PMF allows for weighting of individual data points, while UNMIX does not. This weighting option allows inclusion of measurement uncertainties, and also provides for innovative treatments of data which are missing or below MDL. In the current PMF application, missing data points were “filled” by using mean values as concentrations, with associated uncertainties of ten times the mean. For data points below MDL, half the (varying) MDL was used for concentration, with an uncertainty equal to the standard deviation of the MDL. Using this approach, all 27 variables (excluding mass) and all 854 samples were employed as input data for PMF. Mass attribution for the resulting PMF sources was determined by a weighted least squares regression of the daily source scores vs. fine mass.

In UNMIX, which does not include options for individual data point weighting, missing or below MDL data can be treated in one of two general ways: (a) by substituting specific values for missing or below MDL data points, or (b) by removing observations or variables for which missing or below MDL data are encountered. In the current study, a combination of these data censoring techniques was employed. Samples with missing data for one or more species were eliminated, reducing the sample size from 854 to 838 observations. An average MDL was calculated for each variable, and values below MDL were replaced with one half this (constant) average MDL for each species. The decision to employ a constant, rather than varying MDL resulted from a range of sensitivity tests using alternative “hole filling” procedures, and is explained in more detail by Poirot and Hopke.¹³ The general concern was that the variance in daily (analytical) MDLs appears to be subject to analytical influences, unrelated to air pollution sources (and often leading to “no feasible solution” in UNMIX sensitivity runs). The number of species was also reduced to 12 UNMIX input variables (including fine mass). The selected species were chosen using a combination of trial and error and the “UNMIX overnight” option, which allows consideration of multiple numbers and combinations of input variables. The general objective in choosing UNMIX input variables was to maximize the number of input species and resultant sources, while producing physically realistic and interpretable results.

The species included as input to PMF and UNMIX and the percentage of observations with reported concentrations above MDL are summarized in Figure 1. Whereas 854 observations x 27 variables were employed as PMF input, less than half as many data points (838 observations of 12 variables) were employed as UNMIX input. Different approaches were also employed in the treatment of analytical uncertainties, MDLs and mass attribution in the two models.

Figure 1. Underhill, VT Fine Particle Variables used as Input for UNMIX and PMF Model Runs



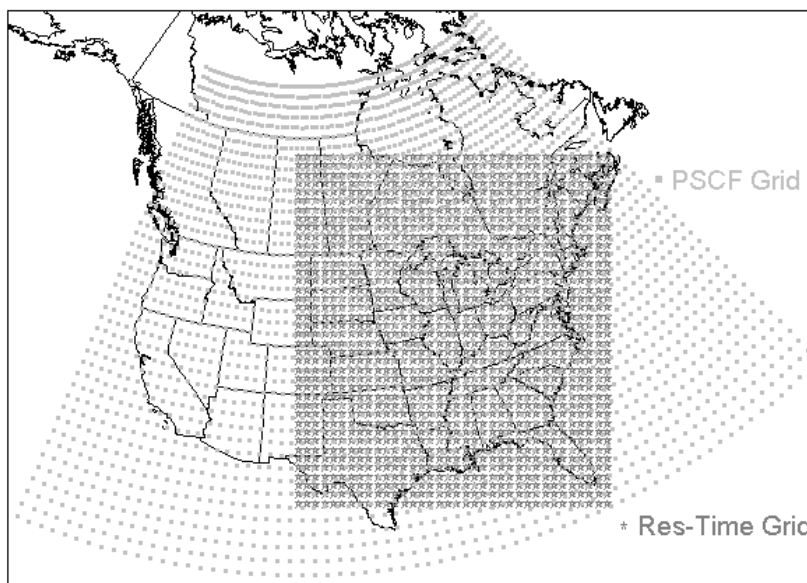
ENSEMBLE BACKWARD TRAJECTORY TECHNIQUES

A set of detailed backward air mass history calculations had been previously calculated for the Underhill, VT site for the years 1989 through 1996 - covering about 95% of the dates for which fine particle data and mathematical model results were available. These air mass histories were calculated using the CAPITA Monte Carlo model,¹⁴ with an NGM¹⁵ meteorological driver, and include backward trajectory (horizontal and vertical location) positions for each of 10 hypothetical particles, released every 2 hours from the receptor location and tracked backward in time for 5 days. Details of these air mass history calculations are described in detail in Poirot *et al.*¹⁶ and associated references.

In the current study, these CAPITA Monte Carlo trajectories were employed in two ensemble backward trajectory techniques, the Potential Source Contribution Function (PSCF) and Residence-Time Analysis (RTA), to evaluate and interpret the results from the multivariate mathematical models. As with the mathematical models, the trajectory techniques were applied independently by the Clarkson team (PSCF) and the VT team (RTA). The trajectory techniques are similar in that both apply spatial grids to track and sort the trajectories as a function of metrics derived from the ambient measurement data - in this case the daily source contributions identified by the mathematical models. The trajectory techniques differ in: the domains, orientations and size of the trajectory tracking grid cells, the metrics employed to attribute trajectories to grid cells, and the metrics calculated from the gridded results.

The PSCF tracking grid covers most of North America with 2,966 grid cells of 1 degree by 1 degree latitude and longitude. The RTA grid is composed of 1,440 squares of 80 by 80 km each covering the eastern United States and Southeastern Canada. It may be noted that the RTA squares are equal-area (on this stereographic map projection) and are smaller than the PSCF grid cells south of about 60 degrees North latitude.

Figure 2. PSCF and Residence-Time Analysis Trajectory Tracking Grids



The individual CAPITA Monte Carlo Trajectories were expressed as a series of latitude-longitude coordinates with an initial location 1 hour upwind of the receptor location, followed by additional coordinates every 3 hours backward in time for 5 days. The trajectory locations are

only tracked within the respective domains, such that the average trajectory duration within the RTA domain, for example, is slightly less than 72 hours. In the PSCF approach, trajectories are assigned to and temporally disaggregated to the grid cells in their paths by a simple count of trajectory endpoints within each 1 degree cell. In the RTA approach, a more computationally intensive technique is employed to disaggregate trajectory sub-segments to each grid square and to determine the time (in hours) that each sub-segment spends over each grid square.

In the PSCF approach as applied here, a single metric is employed to define the “potential source contribution”. A count of all trajectory endpoints in each grid cell for all sampling days at the receptor defines an “all day count”. A second count of trajectory endpoints is determined for a “high day” subset of trajectories (in this case, the highest 40% of daily source contributions for each of the sources identified by the mathematical models). The PSCF is defined as the ratio of high day endpoints to all day endpoints in each grid cell. A PSCF ratio of <0.4 would indicate that a given grid cell is less likely to be up wind if the source contribution at the receptor was high than it is on an everyday basis, while a ratio of 0.8 would indicate a cell is twice as likely to be upwind on days when the source contribution is high as it is on an everyday basis. Given the large PSCF grid domain and resultant sparse trajectory coverage of the more distant grid cells, an inverse distance weighting function is applied to the PSCF ratios to minimize spurious results that often result from large ratios between very small numbers in the most sparsely covered squares.

In the RTA approach, a variety of different metrics can be applied to the resultant counts of hours in the equal-area grid squares.^{17, 18} One set of RTA metrics, referred to as “concentration-based sorting” begins with the conversion of the gridded trajectory hours to “probability fields” in which, for a given scenario of dates, the “upwind probability” of trajectory location in a given grid square is defined as the fraction of hours in that square compared to the total hours in all 1440 squares. An “everyday probability field” is calculated for a scenario of all sample days at the receptor, and provides an indication of areas most likely to be upwind of the receptor on a long-term or climatological basis. A “high day probability field” can be calculated for various definitions of “high” contributions at the receptor, for example upper 50th, 75th, or 90th percentile days, etc. The “incremental probability” for a given high day scenario is defined as the difference between the high day and everyday probability fields. An incremental probability field for or an upper 60th percentile definition of high day would be a direct analog to the PSCF, except that the RTA metric is determined by subtraction (how much greater is the high day probability than everyday?), while the PSCF metric is determined by division (what fraction of total trajectories passing over a cell occur on high concentration days?). A second series of RTA metrics, referred to as “location-based sorting” calculates a summary statistic (mean, median, percentile, etc.) from concentrations (or in this case source contributions) at the receptor for all days with trajectories residing over a each grid square. The summary statistic is weighted by the hours over square of the individual trajectories. As with the PSCF metric, the results from location-based sorting are sensitive to the sparse trajectory coverage of distant grid squares, and a censoring function is applied to exclude calculations in squares with sparse coverage.

MATHEMATICAL MODEL RESULTS

The UNMIX model (run with 838 observations of 12 input variables) identified 7 sources. The PMF model (run with 854 observations of 27 input variables) identified 11 sources. The average fine mass source contributions are summarized in Table 1 and source composition profiles are presented in Table 2. A comparison of the daily reconstructed fine mass contributions from all the model sources with measured fine mass concentrations is displayed in Figure 3.

Table 1. Average Underhill, VT PM_{2.5} Mass Contributions from PMF and UNMIX Models

Sources	PMF Average Mass (ng/m ³)	PMF % Reconstructed Mass	UNMIX Average Mass (ng/m ³)	UNMIX % Reconstructed Mass
MW Summer Coal	4200	53.13	4643	55.22
Woodsmoke	1205	15.24	1314	15.63
MW Winter Coal	593	7.50	1189	14.14
East Coast Oil	545	6.90	643	7.65
Can. Mn Sources	173	2.19	323	3.84
Soil	321	4.06	208	2.47
Canadian Smelter	98	1.24	89	1.05
PMF-Zn-Pb	581	7.35		
PMF-Cu	122	1.54		
PMF-Na-S	52	0.65		
PMF-Salt	15	0.19		
Reconstructed Mass	7904	100	8408	100

It may be noted that both models reproduce the daily mass reasonably well. UNMIX, which included mass as an input variable, reproduced the average and daily mass somewhat more closely (essentially attributing all the mass among the 7 identified sources). PMF, which apportioned mass by regression from the daily mass and source scores, leaves some of the mass unexplained by the 11 sources it identified from the 27 non-mass input variables.

The source names indicated in Tables 1 and 2 are not produced by the models, but rather reflect the best judgement of the modelers.

Figure 3. Comparison of Measured and Modeled PM_{2.5} Mass

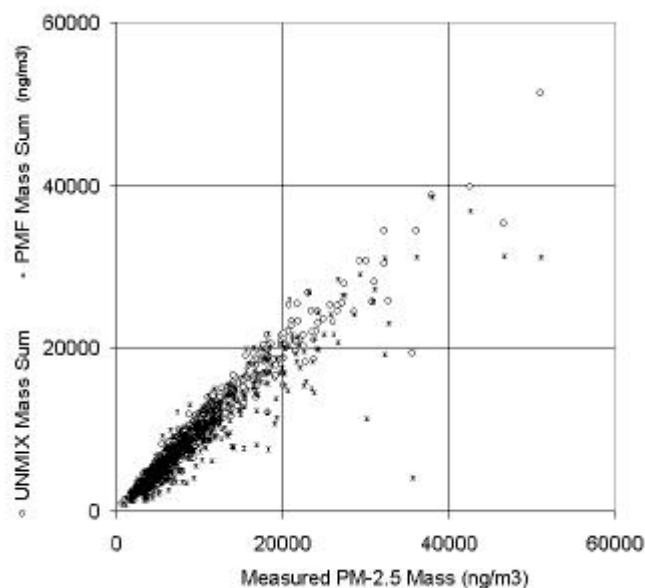


Table 2. PMF and UNMIX Source Compositions for Underhill, VT, 9/88 - 5/95

	MW Summer Coal	MW Winter Coal	Woodsmoke	East Coast Oil	Soil	Canadian Mn Sources	Canadian Smelter	Zn-Pb	Salt	Cu	Na-S
PMF											
S	0.15630	0.12151	0.01569	0.14066	0.03035	0.04105	0.08584	0.02899	0.05451	0.10574	0.66484
H	0.04987	0.04262	0.03491	0.04688	0.05098	0.03222	0.05608	0.04911	0.02495	0.05759	0.03667
Babs*	0.03138	0.10482	0.09199	0.04440	0.04977	0.05405	0.06347	0.06602	0.05003	0.04004	0.06775
K	0.00018	0.00444	0.01907	0.00090	0.01080	0.01487	0.00638	0.00505	0.00937	0.00324	0.01267
Se	0.00003	0.00057	0.00000	0.00000	0.00001	0.00002	0.00001	0.00000	0.00005	0.00001	0.00002
Ca	0.00035	0.00462	0.00024	0.00194	0.00958	0.00961	0.01103	0.00081	0.01877	0.00281	0.02623
Fe	0.00084	0.00499	0.00075	0.00073	0.02116	0.00642	0.00150	0.00336	0.00121	0.00065	0.01060
Si	0.00007	0.00430	0.00050	0.00092	0.10586	0.00925	0.01179	0.00143	0.01104	0.00386	0.00873
Mn	0.00000	0.00004	0.00001	0.00001	0.00016	0.00663	0.00013	0.00002	0.00021	0.00008	0.00017
Ni	0.00001	0.00001	0.00000	0.00127	0.00001	0.00001	0.00003	0.00001	0.00004	0.00002	0.00006
As	0.00000	0.00001	0.00000	0.00001	0.00001	0.00002	0.00346	0.00001	0.00005	0.00003	0.00007
Al	0.00002	0.00028	0.00008	0.00021	0.06596	0.00066	0.00085	0.00019	0.00152	0.00093	0.00218
Br	0.00000	0.00086	0.00031	0.00012	0.00025	0.00002	0.00112	0.00003	0.00051	0.00004	0.01214
Cl	0.00001	0.00006	0.00003	0.00007	0.00011	0.00019	0.00031	0.00006	0.20578	0.00025	0.00062
Cr	0.00002	0.00006	0.00004	0.00002	0.00006	0.00010	0.00025	0.00006	0.00014	0.00007	0.00121
Cu	0.00001	0.00003	0.00001	0.00003	0.00003	0.00010	0.00011	0.00002	0.00019	0.01022	0.00020
Mg	0.00017	0.00111	0.00110	0.00131	0.00248	0.00406	0.00603	0.00110	0.01074	0.00378	0.03209
Mo	0.00002	0.00009	0.00009	0.00005	0.00009	0.00011	0.00032	0.00009	0.00031	0.00016	0.00105
Na	0.00147	0.00050	0.00019	0.00059	0.00067	0.00361	0.00306	0.00061	0.17715	0.00383	1.09613
P	0.00005	0.00015	0.00011	0.00010	0.00021	0.00025	0.00040	0.00016	0.00056	0.00024	0.00238
Pb	0.00006	0.00055	0.00022	0.00045	0.00005	0.00397	0.00075	0.00154	0.00021	0.00039	0.00172
Rb	0.00000	0.00000	0.00005	0.00000	0.00002	0.00020	0.00001	0.00000	0.00001	0.00003	0.00002
Sr	0.00000	0.00000	0.00002	0.00001	0.00006	0.00021	0.00003	0.00001	0.00008	0.00003	0.00037
Ti	0.00021	0.00002	0.00001	0.00003	0.00306	0.00059	0.00010	0.00009	0.00047	0.00012	0.00158
V	0.00000	0.00004	0.00013	0.00132	0.00002	0.00057	0.00024	0.00002	0.00017	0.00027	0.00235
Zn	0.00004	0.00031	0.00008	0.00030	0.00067	0.00059	0.00122	0.00949	0.00134	0.00720	0.00173
Zr	0.00000	0.00000	0.00001	0.00002	0.00001	0.00034	0.00004	0.00001	0.00002	0.00006	0.00020
UNMIX											
S	0.13128	0.12863	0.03890	0.14023	0.13821	0.00183	0.09384				
H	0.04456	0.04445	0.04060	0.05051	0.09158	0.05260	0.05203				
Babs*	0.02883	0.08350	0.08583	0.05042	0.03090	0.04609	0.09071				
K	0.00033	0.00512	0.01587	0.00326	0.02144	0.00848	0.00945				
Se	0.00000	0.00034	0.00001	0.00000	0.00003	0.00000	-0.00001				
Ca	-0.00030	0.00297	0.00433	-0.00102	0.05567	0.00569	0.00347				
Fe	0.00040	0.00487	0.00068	0.00134	0.06548	0.00446	0.00248				
Si	0.00264	0.00482	-0.00381	0.00271	0.23386	0.01298	0.00518				
Mn	0.00001	0.00002	0.00005	-0.00001	0.00008	0.00381	0.00010				
Ni	0.00000	0.00000	0.00002	0.00115	-0.00012	-0.00004	-0.00002				
As	0.00000	0.00001	0.00002	0.00001	-0.00003	0.00001	0.00366				
Mass**	4642.92	1188.79	1314.19	643.38	207.78	322.66	88.64				

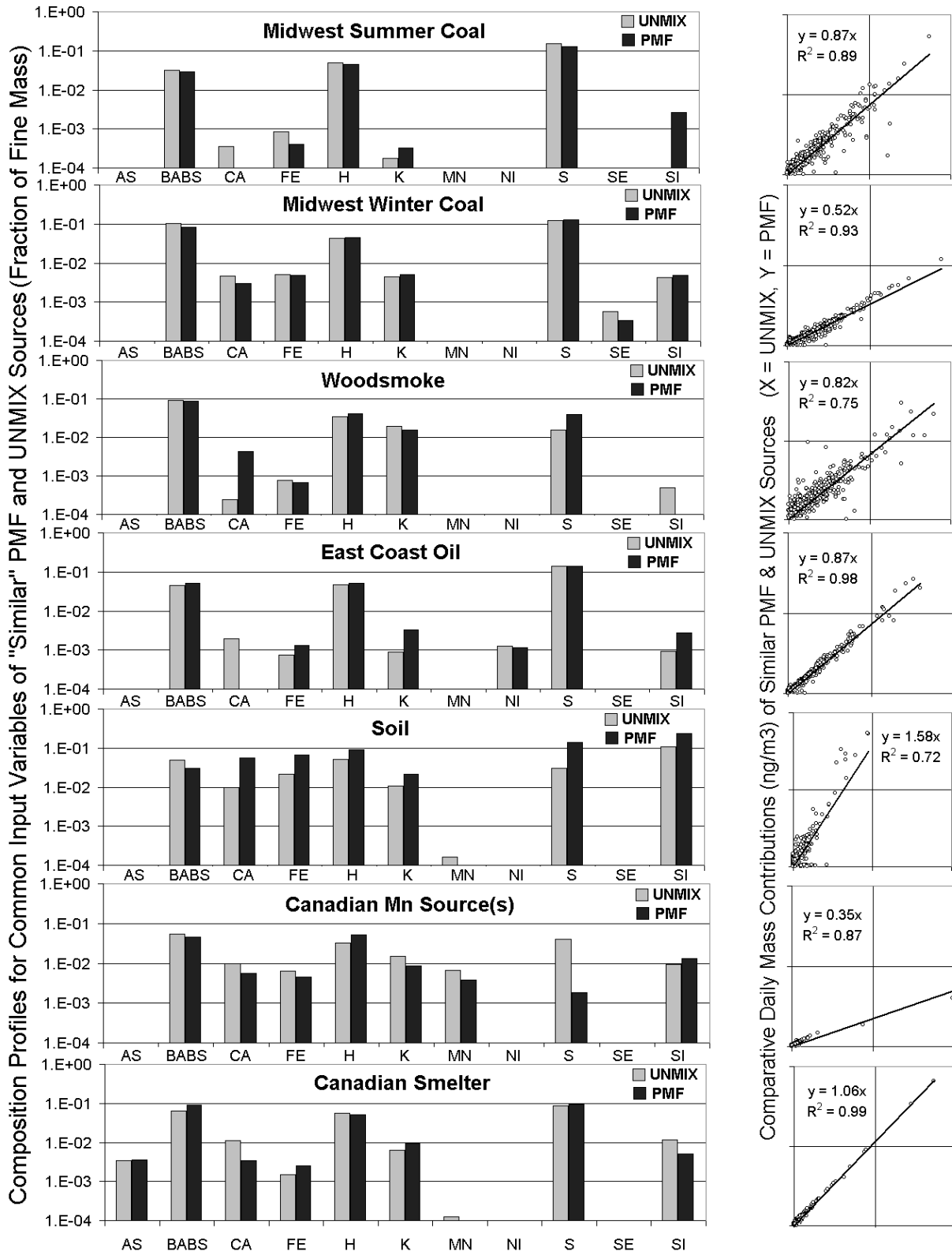
Notes. Units are fractions of fine mass, except:

* Babs, originally in inverse megameters was multiplied by 0.4 to convert to estimated EC mass

** Mass, included as an input variable in UNMIX, is expressed in ng/m³ in this table

These interpretations are based on the source compositions, time series of the daily source contributions and subsequent PSCF and Residence-Time analysis results to follow. Note also that common source names have been applied to “similar” sources that were identified by both the PMF and UNMIX models. Again, this judgement of similarity is derived solely from interpretations by the modelers, and is based in turn upon comparisons of the source profiles and daily source contributions resulting from the models. These source profiles (for those elements used as common input for both models) and daily source contributions are displayed in Figure 4.

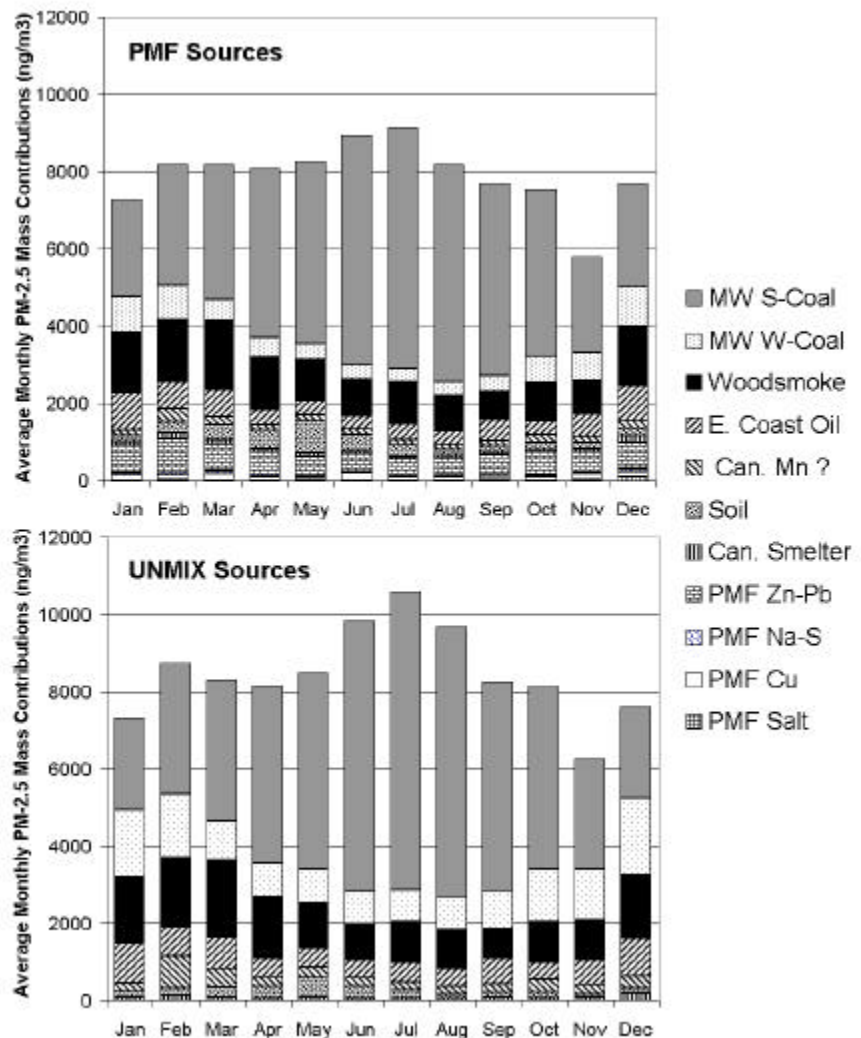
Figure 4. Comparison of “Similar” UNMIX and PMF Source Compositions and Daily PM_{2.5} Contributions



Note that in most cases the source profiles are similar, the daily source contributions are well correlated ($R^2 > 0.75$), and the slopes of the daily mass comparisons are generally within 25% of 1 to 1. Notable exceptions occur for the MW Winter Coal and the Canadian Mn sources, which have similar composition profiles and highly correlated daily contributions, but show substantially higher mass contributions from UNMIX than for PMF - hence the sources are “similar”, but their mass attribution differs. The Soil sources have similar profiles, but show the poorest daily correlation ($R^2 = 0.73$), with a daily mass contribution from the PMF Soil source which is about 50% higher than from the UNMIX Soil source. Despite these discrepancies, the authors feel its reasonable to conclude that for the 7 sources identified by the UNMIX model, there were “similar” counterpart sources also identified by PMF.

Average monthly source contributions are displayed in Figure 5, and provide some insights into the selected source names for the similar PMF and UNMIX sources. The largest source - named “MW Summer Coal” accounts for about half the average mass and has a strong summer maxima in both models. The source named “MW Winter Coal” displays an opposite seasonality, with a strong winter maxima. These two Midwestern sources together account for a majority of the Sulfur (80%) in both models, but the “Summer” source has a much higher S:Se ratio, indicative of more efficient sulfur gas to particle conversion chemistry. The sources named “East Coast Oil” and “Woodsmoke” also exhibit winter maxima, consistent with seasonal increases in wood and oil fuel combustion. The “Soil” source peaks in Spring, while the “Canadian Mn” and “Canadian Smelter” sources do not exhibit strong seasonal patterns.

Figure 5. Seasonal PM_{2.5} Contributions from PMF and UNMIX Sources



ENSEMBLE TRAJECTORY RESULTS

Additional insights into the nature of the identified common PMF and UNMIX sources are provided through a trajectory-based evaluation of the upwind locations associated with “high” concentrations of these sources. In Figure 6, an attempt is made to compare, as directly as possible, the difference between the PSCF and RTA trajectory tracking methods.

Figure 6 employs the same source (the PMF “East Coast Oil” source), the same definition of “high contribution” (the 60th percentile, or highest 40% of the source contribution days), the same metric (ratio of high day to every day), the PSCF inverse distance weighting function, and the same common grid domain (the RTA domain) to the plots on both sides of the figure. On the left side, the plotted ratio of “high day” to everyday is based on trajectory endpoints in 1 degree by 1 degree grid cells, and on the right side, the plotted ratio is based on the ratio of “high day” to everyday hours in 80x80 km squares. In both cases, the areas with the highest PSCF ratios are clearly centered on the east coast urban corridor. The similarity between the plots indicates that for this particular metric (ratio of high day to everyday) there is no substantial difference between the PSCF and RTA trajectory tracking and gridding techniques.

Figure 6. Comparison of PSCF and RTA Trajectory Tracking Methods, illustrated by Ratio of 60th % to Everyday for PMF East Coast Oil Source

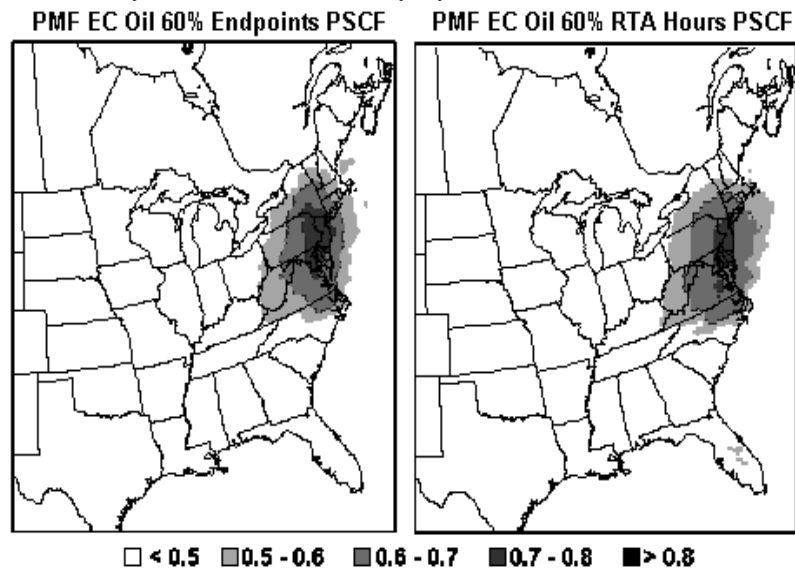
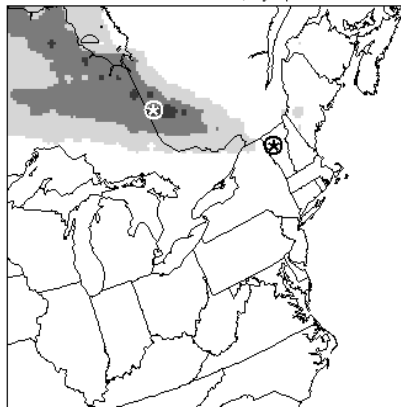


Figure 7. Comparison of RTA Metrics for UNMIX Canadian Smelter Source

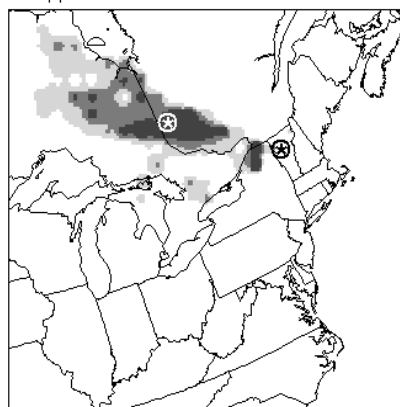
Average VT Fine Mass Contribution (ng/m³) from UNMIX Can. Smelter source, by upwind location



100-125 125-150 > 150 ng/m³

⊗ Noranda Smelter

Incremental Probability (x1000) for Upper 10% UNMIX Can. Smelter source



0.8-1.0 1.0-1.2 > 1.2

⊗ Underhill Receptor

Figure 7 illustrates two different Residence-Time metrics as applied to the UNMIX “Canadian Smelter” source. The left plot is based on calculations of an average value of the daily source contribution for all trajectories passing through each grid square. The right plot shows the incremental probability (high day minus everyday) for upper 10% of the daily source contributions.

Figure 8 displays the Residence-Time Analysis upwind incremental probability fields for the highest 10 percent of daily source contributions for the 7 similar sources identified independently by the UNMIX and PMF mathematical models

Figure 8. RTA Incremental Probability Fields for Top 10% of Daily UNMIX and PMF Source Contributions

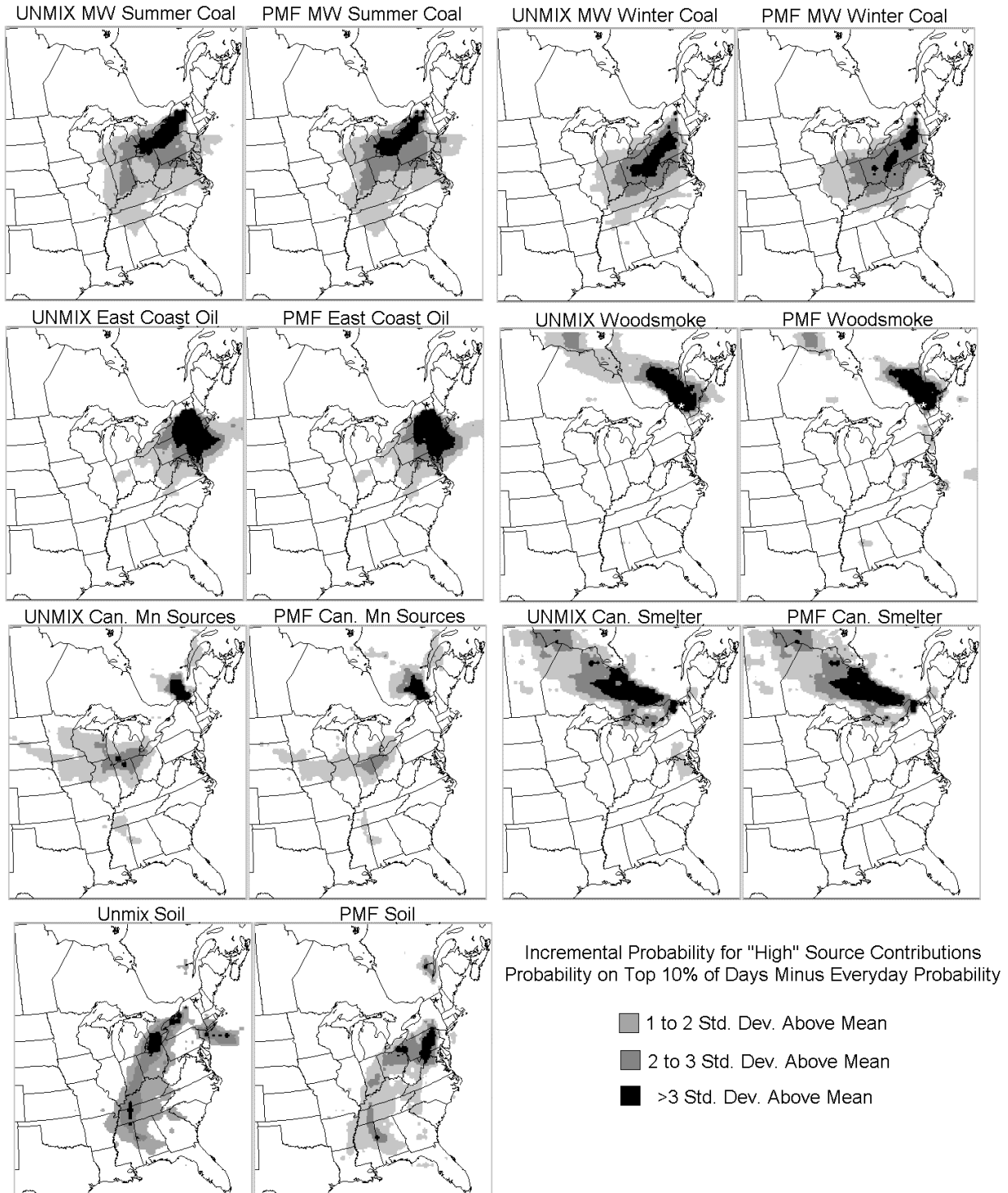
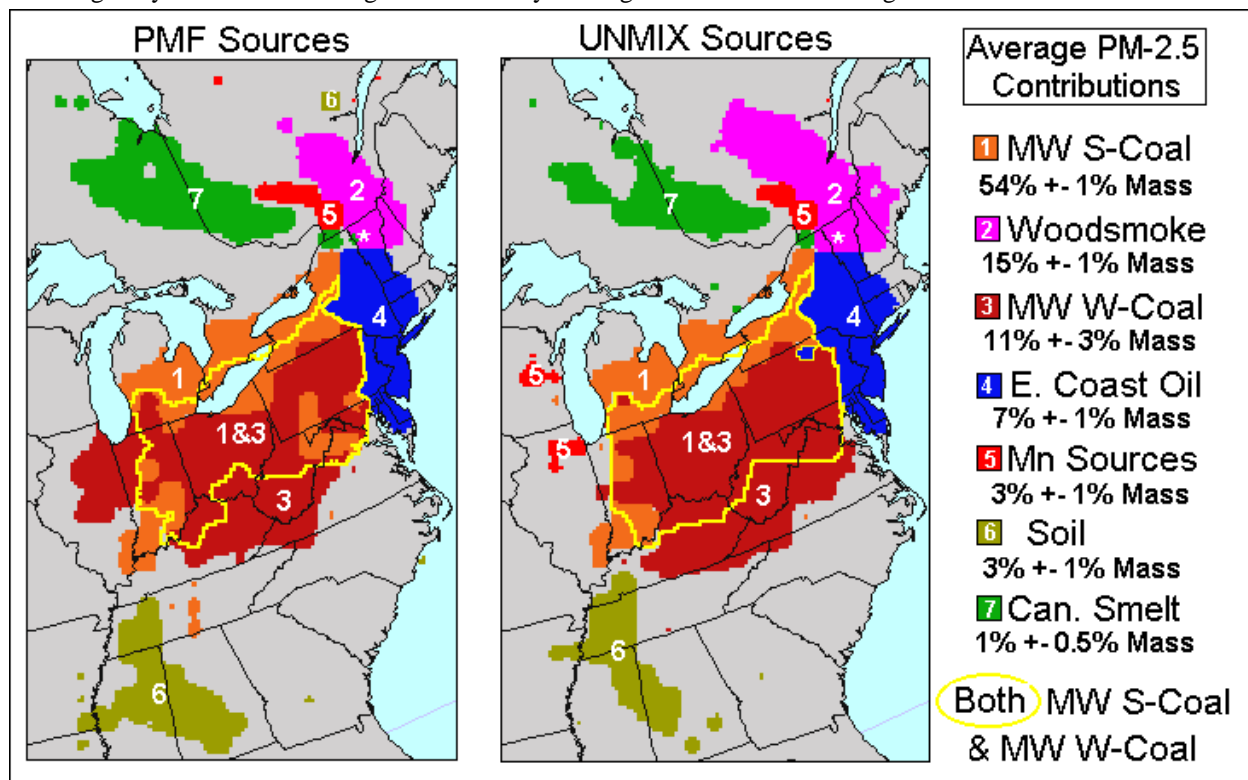


Figure 9 attempts to summarize the comparative results of the mathematical and trajectory techniques in a single image. Like Figure 8, it is based on RTA incremental probability fields for the upper 10 % of daily source contributions for the 7 common sources identified independently by the PMF and UNMIX models. In this case, locations are only shaded if the incremental probabilities exceed a “high” threshold (of 0.1%), and only for the source which had the highest upwind probability among all the sources. For clarity a “water mask” is applied to show only locations over land areas. The indicated contributions to fine mass at the receptor (*) are based on an average of the PMF and UNMIX results. While there is some degree of overlap between areas of high probability among adjacent source regions, this overlap is especially strong between the sources identified as “Midwestern Summer and Winter Coal”, and so an additional isopleth is added to identify the common area where the probability for both these sources was especially high (> 0.12%). Generally these sources are associated with the same upwind source region.

Figure 9. Incremental Probability Fields for Top 10% Contributions of Common PMF & UNMIX Sources, Showing Only Locations with Highest Probability Among All Sources and Average PM-2.5 Mass Contributions



DISCUSSION

The source identified as “Canadian Smelter” accounts for a majority of the elemental arsenic apportioned by PMF and UNMIX, but only accounts for about 1% of the average fine mass concentration at the receptor. Its trajectory-based assessment suggests that this influence may come from a single large smelter located near the town of Noranda near the Quebec/Ontario border (see Figure 7). The time series for this source indicates a substantial reduction occurring in early 1990, coincident with major revisions to emissions control processes at Noranda.

The source identified as “Canadian Mn Sources”, which accounts for a majority of elemental Mn but only about 3 % of the average fine mass, also appears to have a predominantly Canadian origin. The upwind probability field in Figure 9 suggests a uniquely high probability over the relatively nearby Montreal urban area, which is consistent with the influence of Canadian motor vehicles, burning gasoline with a uniquely Canadian Mn additive (methylcyclopentadienyl manganese tricarbonyl or MMT). However, a large Mn alloy production plant just southwest of Montreal in Beauharnois, Quebec may have also been an important Mn source during the first several years of this sampling period.¹⁶ Also, as may be noted in Figure 8, the Mn source has a second area of relatively high incremental probability along the Great Lakes, possibly associated with industrial Mn alloy use, MMT gasoline additive or both.

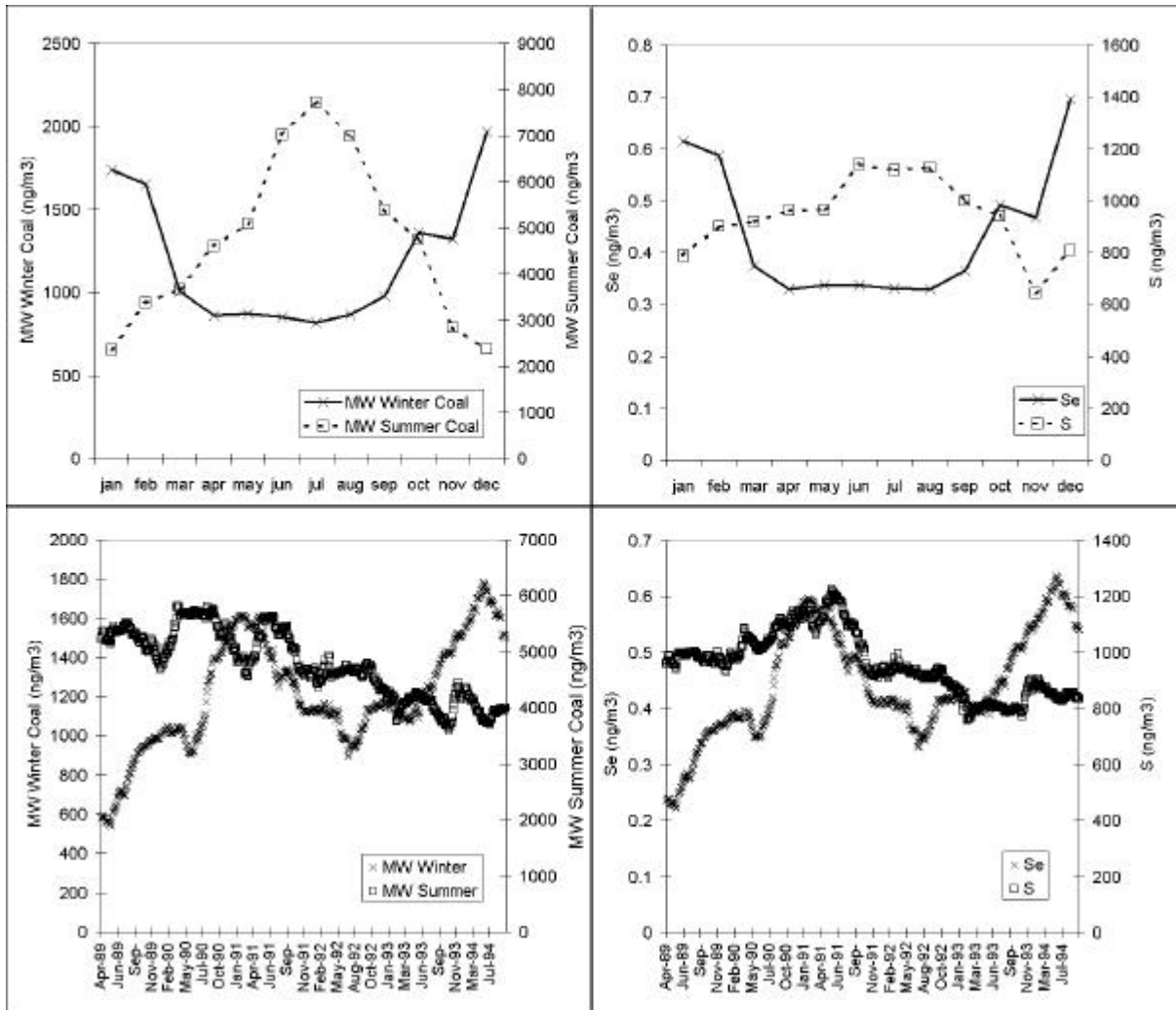
The “Soil” source accounts for a large fraction of the crustal elements (Si, Fe, Ca, Al) and about 3% of the fine mass. Its regional location is more ambiguous than suggested in Figure 9 (see Figure 8), as its areas of highest upwind probability overlap with, but are less distinct than, those of several other sources. The RTA location-based sorting for the soil source(s) indicates highest average soil associated with locations in the extreme southwest of the RTA grid. It seems likely that windblown dust emissions from relatively distant arid regions to the southwest are significant contributors, but that there are also fine soil contributions associated with dry conditions and high wind speeds from a wide range of locations.

The source identified as “East Coast Oil” clearly has a strong upwind orientation along the North East Coast urban corridor (Figures 6, 8, 9), where residual and distillate oil are predominant “fuels of choice” for utility, industrial and residential heating sources. This source accounts for a majority of the nickel (UNMIX and PMF) and vanadium (PMF), and about 7 % of the average fine mass at Underhill. It is likely that this Northeast Coast source is a regional influence, rather than specific to the (strongly “flavored”) oil combustion emissions which provide relatively unique tracers. That is, when VT is influenced by the oil burning sources to the South, it is likely to be concurrently influenced by all the emissions (including motor vehicles, for example) in the densely populated Northeast urban corridor.

The “Woodsmoke” source accounts for a high fraction of the elemental potassium and B_{abs} and about 15% of the fine mass. While there were no direct measurements of carbon in this sampling program, the high B_{abs} fraction suggests a significant elemental carbon content, and an indirect estimate of organic matter by the “non-sulfate hydrogen” method^{10, 11} ($13.75*(H-S/4)$) suggests that organics may account for about half the mass associated with this source. The woodsmoke time series indicates a general winter maxima (consistent with residential wood burning) but also shows occasional summer spikes, several of which coincide with periods of known forest fire impacts (which were coincidentally located in western Quebec). Its incremental probability field is the only one among all the sources which shows high incremental probability in the area immediately surrounding the receptor site, indicating a strong local source influence from residential wood combustion in northern New England and southwestern Quebec.

The sources identified as “Midwestern Summer Coal” and “Midwestern Winter Coal” together account for about 80 % of the sulfur and 65% of the fine mass at the receptor. Their incremental probability fields indicate a large common source region extending from the lower Great Lakes to south of the Ohio River Valley, encompassing the locations of many large, sulfur emitting utility and industrial sources. Despite the common regional origin indicated by the trajectory techniques, the mathematical models both separate this Midwestern influence into two separate sources, which are uncorrelated in their daily contributions at the receptor. One partial explanation for this mathematical separation may be in the seasonal changes in sulfur conversion chemistry, which would typically result in more efficient sulfate aerosol formation in the summer (in relation to other primary emission elements like Se). The opposite seasonality of these Midwestern sources is clearly evident in Figure 5, although note that even in the winter, the mass contribution from the “MW Summer” source is often as high as or higher than the “MW Winter” source, and so this choice of names is not ideal.

Figure 10. Comparison of Seasonal (top) and Long-Term Trends (1 year moving average - bottom) in UNMIX Midwestern Sources (left) and S and Se Concentrations (right) at Underhill, VT, 1988-95



Receptor models like PMF and UNMIX require an assumption of constant source composition, and can be useful for apportionment of secondary species like sulfate only to the extent that the secondary species is sampled at the receptor in a relatively constant ratio to other primary emission variables from the source. Figure 10 shows (upper left) the strong seasonal differences in the average monthly contributions for the two Midwestern sources resulting from the UNMIX model, and (upper right) similar seasonal differences in the raw sulfur and selenium concentrations used as input for both the PMF and UNMIX models. It can be noted that not only does sulfur increase in summer, but selenium concentrations (relatively constant between April and September) increase dramatically in the winter. Thus it appears to be not just seasonal changes in aerosol sulfur, but even larger changes in the seasonal S:Se ratios that helps break these sources apart in the PMF and UNMIX results.

The lower half of Figure 10 shows (left) the (smoothed) long-term moving average daily contributions from the UNMIX Midwestern sources. The moving average is based on 120 samples, an averaging period of roughly one year, with the data points plotted at a date in the middle of the 120 day averaging period. The lower right plot in Figure 10 shows a (very) similar plot of the time trend in raw S and Se data used as input for the models. The decline in sulfur concentrations is consistent with documented reductions in Eastern US sulfur emissions. The reasons for the concurrent increase in Se concentrations are unclear, but it appears that differences in both the seasonal and long-term trends in Midwestern sulfur and selenium emissions may be contributing to the different Midwestern sources identified by the models, and that these different sources may be reflective of changing fuel characteristics on a seasonal and long-term basis.

CONCLUSIONS

Despite differences in the fundamental mechanisms employed by the PMF and UNMIX models and substantial differences in the manipulated raw data used as input for these models, both models identified seven sources with very similar chemical compositions and daily fine mass contributions at the Underhill, VT receptor site. This similarity between the mathematical model results is further illustrated by the common upwind probability fields identified by the trajectory-based techniques. The trajectory results, in turn, add some degree of confidence that the mathematical models have identified sources that are physically meaningful, and also provide additional insights into the regional nature of these sources. Conversely, the convergence of results among the different techniques also provides additional confidence that the trajectory techniques can provide useful information on source locations, at least for a remote receptor site like Underhill, which is likely to be influenced by a range of relatively distant source regions.

REFERENCES

1. Hopke, P.K. (1997) Receptor modeling for air quality management, in Proceedings: AWMA / AGU Specialty Conference on Visual Air Quality: Aerosols and Global Radiation Balance, Bartlett, NH.
2. Paatero, P. and U. Tapper (1994) *Environmetrics* **5**:111-126.
3. Paatero, P. (1998) User's Guide for Positive Matrix Factorization programs PMF2 and PMF3.

4. Henry, R.C., C.W. Lewis, and J.F. Collins (1994), *Environ. Sci. Technol.* **28**:823-832.
5. Henry, R.C. (2000) UNMIX Version 2 Manual, Prepared for the U. S. Environmental Protection Agency.
6. Willis, R.D. (2000) Workshop on UNMIX and PMF as Applied to PM_{2.5}, EPA/600/A-00/48.
7. Eberly, S.I., B.W. Coutant and C.W. Lewis (2000) EPA workshop on source apportionment Tools UNMIX and PMF as applied to PM_{2.5} synthetic data set generation, this conference.
8. Ashbaugh, L.L., W.C. Malm and W.Z.Sadeh (1983) A methodology for establishing the probability of the origin of air masses containing high pollutant concentrations, *76th Annual APCA Meetings*, Atlanta, GA.
9. Poirot, R. L. and P.R. Wishinski (1986) *Atmos. Environ.* **20**: 1457-1469.
10. Flocchini, R.G., T.A. Cahill, R.A. Eldred and P.J. Feeney (1990) Particulate sampling in the Northeast: a description of the NESCAUM network, Transactions: *A&WMA/EPA Spec. Conf. on Visibility and Fine Particles*, C.V. Mathai, Ed., Estes Park, CO.
11. Poirot, R.L., P.J. Galvin, N. Gordon, S. Quan, A. Van Arsdale, and R.G. Flocchini (1991) Annual and seasonal fine particle composition in the Northeast: Second year results from the NESCAUM monitoring network, 91-49.1, *84th Annual A&WMA Meeting*, Vancouver, Canada..
12. Henry R. C. (1987) *Atmospheric Environment*, **21**:1815-1820.
13. Poirot, R.L. and P.K. Hopke (2000) The Potential Influence of Data Artifacts on Receptor Model Results, this conference.
14. Schichtel, B.A.; Husar, R.B.(1996) *J. of Air & Waste Manage. Assoc.*, **47**: 331-343.
15. Rolph, G D. (1996) NGM Archive TD-6140, January 1991 - June 1996, Prepared for National Climatic Data Center (NCDC).
16. Poirot, R., P. Wishinski, B. Schichtel , P. Girton (1999) Air Trajectory Pollution Climatology for the Lake Champlain Basin, AGU Water Monograph on Lake Champlain Research and Implementation, T. and P. Manley, Eds., American Geophysical Union, Washington, DC.
17. Wishinski, P. R. and R. L. Poirot, Long-Term Ozone Trajectory Climatology for the Eastern US, Part I: Methods, 98-A613, *91st Annual A&WMA Meeting*, San Diego, CA, 1998.
18. Poirot, R. L. and P. R. Wishinski, Long-Term Ozone Trajectory Climatology for the Eastern US, Part II: Results, 98-A615, *91st Annual A&WMA Meeting*, San Diego, CA, 1998.

KEYWORDS

Positive Matrix Factorization, UNMIX, receptor models, trajectory models, Potential Source Contribution Function, Residence-Time Analysis, fine particles, speciation, transport.