Journal of Exposure Science and Environmental Epidemiology (2006) 16, 275–286
© 2006 Nature Publishing Group All rights reserved 1053-4245/06/$30.00
www.nature.com/jes

PM source apportionment and health effects: 1. Intercomparison of source apportionment results

PHILIP K. HOPKE,[a] KAZUHIKO ITO,[b] THERESE MAR,[c] WILLIAM F. CHRISTENSEN,[d] DELBERT J. EATOUGH,[e] RONALD C. HENRY,[f] EUGENE KIM,[a] FRANCINE LADEN,[g] RAMONA LALL,[b] TIMOTHY V. LARSON,[h] HAO LIU,[i] LUCAS NEAS,[j] JOSEPH PINTO,[k] MATTHIAS STÖLZEL,[l] HELEN SUH,[g] PENTTI PAATERO[m] AND GEORGE D. THURSTON[b]

[a]*Center for Air Resources Engineering and Science, Clarkson University, Potsdam, NY, USA*
[b]*Institute for Environmental Medicine, New York University, Tuxedo Park, NY, USA*
[c]*Department of Environmental Health, University of Washington, Seattle, WA, USA*
[d]*Department of Statistics, Brigham Young University, Provo, UT, USA*
[e]*Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT, USA*
[f]*Department of Civil and Environmental Engineering, Southern California University, Los Angeles, CA, USA*
[g]*Department of Environmental Health, Harvard School of Public Health, Boston, MA, USA*
[h]*Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA*
[i]*Department of Biostatistics, University of Washington, Seattle, WA, USA*
[j]*National Health & Environmental Effects Research Laboratory, US Environmental Protection Agency, Chapel Hill, NC, USA*
[k]*National Center for Environmental Assessment, US Environmental Protection Agency, Research Triangle Park, NC, USA*
[l]*Institute of Epidemiology, GSF Focus Network Aerosols and Health, GSF National Research Center for Environment and Health, Neuherberg, Germany*
[m]*Department of Physical Sciences, University of Helsinki, Helsinki, Finland*

During the past three decades, receptor models have been used to identify and apportion ambient concentrations to sources. A number of groups are employing these methods to provide input into air quality management planning. A workshop has explored the use of resolved source contributions in health effects models. Multiple groups have analyzed particulate composition data sets from Washington, DC and Phoenix, AZ. Similar source profiles were extracted from these data sets by the investigators using different factor analysis methods. There was good agreement among the major resolved source types. Crustal (soil), sulfate, oil, and salt were the sources that were most unambiguously identified (generally highest correlation across the sites). Traffic and vegetative burning showed considerable variability among the results with variability in the ability of the methods to partition the motor vehicle contributions between gasoline and diesel vehicles. However, if the total motor vehicle contributions are estimated, good correspondence was obtained among the results. The source impacts were especially similar across various analyses for the larger mass contributors (e.g., in Washington, secondary sulfate SE = 7% and 11% for traffic; in Phoenix, secondary sulfate SE = 17% and 7% for traffic). Especially important for time-series health effects assessment, the source-specific impacts were found to be highly correlated across analysis methods/researchers for the major components (e.g., mean analysis to analysis correlation, $r > 0.9$ for traffic and secondary sulfates in Phoenix and for traffic and secondary nitrates in Washington. The sulfate mean r value is >0.75 in Washington.). Overall, although these intercomparisons suggest areas where further research is needed (e.g., better division of traffic emissions between diesel and gasoline vehicles), they provide support the contention that $PM_{2.5}$ mass source apportionment results are consistent across users and methods, and that today's source apportionment methods are robust enough for application to $PM_{2.5}$ health effects assessments.
Journal of Exposure Science and Environmental Epidemiology (2006) **16,** 275–286. doi:10.1038/sj.jea.7500458; published online 12 October 2005

Keywords: *receptor modelling, source apportionment, $PM_{2.5}$, positive matrix factorization, PMF, Unmix.*

Introduction

Current National Ambient Air Quality Standards (NAAQS) for airborne particulate matter (PM) use airborne particle mass as the indicator for making air quality determinations. However, it seems highly likely that some types of particles are more toxic than others. Thus, the focus on all particles that contribute mass may lead to less efficient and less effective control strategies, relative to being able to focus directly on those particles that cause the adverse human health effects. Although current regulations only target total mass concentrations, future regulations could be focused onto the specific components that are related to inducing the adverse health effects. Even under current regulations, the implementation planning process in areas in nonattainment

1. Address all correspondence to: Professor P.K. Hopke, Center for Air Resources Engineering and Science, Clarkson University, Potsdam, NY 13699-5708, USA. Tel.: +1 315 268 3861; Fax: +1 315 268 4410; E-mail: hopkepk@clarkson.edu
Received 20 April 2005; accepted 29 July 2005; published online 12 October 2005

of the National Ambient Air Quality Standards could focus on the most toxic components of the ambient aerosol if they can be identified.

However, there are an enormous number of possible chemical species associated with the particles, and a potentially more effective approach would be to consider the airborne PM as a mixture of mass contributions by various source classes. These source types (e.g., spark-ignition vehicles, diesel powered vehicles, coal-, gas-, or oil-fired power plants, incinerators, etc.) have characteristic chemical and/or physical patterns (mixtures), and it may be useful to examine the relationship of specific source emissions on health effects since, if a limited number of sources contributed significantly to the effects, more targeted control strategies could then be devised to focus on those sources producing most of the health problems.

These source composition or physical properties profiles permit the contributions of these sources to the airborne PM mass to be apportioned. This area of research is called Receptor Modeling and has been an area of active research for over 30 years. There are a number of methods that have been used with the choice of technique being dependent on the amount of information that is available *a priori*. In general, these methods have been primarily employed as part of the development of air quality management plans. However, there have been very few published efforts to relate apportioned sources to human health effects (Ozkaynak and Thurston, 1987; Laden et al., 2000; Mar et al., 2000). As described in a companion paper by Thurston et al. (2005), a workshop was organized, under the auspices of the US PM Research Centers, at which a number of investigators presented their results on the analyses of two historical data sets: one from Washington, DC and the other from Phoenix, AZ. The apportioned $PM_{2.5}$ PM mass contributions were then included in health effects models. The results of the health effects analyses are presented in companion papers (Mar et al., 2005; Ito et al., 2005). This paper will describe the ambient $PM_{2.5}$ compositional data sets and the various methods used to identify the sources and perform the mass apportionment in each workshop city, for subsequent input to the health effects comparisons. It is the goal of this paper to evaluate the application of source apportionment methods and their application by multiple investigators and their associated uncertainties.

Data sets

Fine PM mass and speciation data sets from Washington, DC and Phoenix, AZ were selected for evaluation in the workshop because they were as consistent as possible with the newly instituted EPA Speciation Network, allowing these results to be indicative of what might be achieved when analyzing those data. These data sets are described in detail elsewhere in the literature (Ramadan et al., 2000, 2003; Kim and Hopke, 2004a), and briefly summarized below.

Washington DC

The $PM_{2.5}$ samples in this study were collected on Wednesdays and Saturdays at the IMPROVE monitoring site located in Washington, DC. This monitoring site is located near the Potomac River, 2 km southeast of Lincoln Memorial, 3 km northeast of Ronald Reagan Washington National Airport, and 30 m above sea level. Highways are closely situated to the north and west of the site. Integrated 24-h $PM_{2.5}$ samples were collected on Teflon, Nylon, and quartz filters. The Teflon filter was used for mass concentrations and analyzed via particle induced X-ray emission (PIXE) for Na to Mn, X-ray fluorescence (XRF) for Fe to Pb, proton elastic scattering analysis (PESA) for elemental hydrogen concentration (University of California at Davis, CA, USA). The Nylon filter was analyzed via ion chromatography (IC) for sulfate (SO_4^{2-}), nitrate (NO_3^-), and chloride (Cl^-) (Research Triangle Institute, NC, USA). The quartz filter was analyzed via IMPROVE/TOR protocol16 for eight temperature resolved carbon fractions (Desert Research Institute, NV, USA). This protocol volatilizes organic carbon (OC) by four temperature steps in a helium atmosphere: OC1 at 120°C, OC2 at 250°C, OC3 at 450°C, and OC4 at 550°C. After OC4 response returns to baseline or a constant value, the pyrolysed organic carbon (OP) is oxidized at 550°C in a mixture of 2% oxygen and 98% helium atmosphere prior to the return of reflectance to its original value. Then three element carbon (EC) fractions are measured in oxidizing atmosphere: EC1 at 550°C, EC2 at 700°C, and EC3 at 850°C.

Phoenix AZ

Daily, integrated 24-h samples were collected on 37 mm diameter Teflon and quartz filter media for fine particle mass and species measurements using a dual fine particle sequential sampler (DFPSS). The samples were collected during the time period from March 1995 through June 1998. A total of 981 samples was finally obtained. Two energy dispersive X-ray spectrometers were used to produce the chemical elemental concentration data; a custom-made machine from Lawrence Berkeley Laboratories (LBL) and a commercially available one from Kevex (KEV). Both XRF instruments employed multiple choices for secondary excitation and utilized a helium atmosphere rather than vacuum in order to preserve volatile species. The quartz filters collected with the DFPSS were analyzed by Sunset Laboratory, Forest Grove, OR, USA using the thermal optical transmission technique (Chow et al., 1993). This technique measured both OC and EC. Each sample was characterized by the measured concentrations of the following 46 chemical elements: Na, Mg, Al, Si, P, S, Cl, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Rb, Sr, Y, Zr, Mo, Rh, Pd, Ag,

Cd, Sn, Sb, Te, I, Cs, Ba, La, W, Au, Hg, Pb, OC, and EC. This 981×46 data matrix was used as the basis for the receptor modeling studies to infer possible aerosol sources. In this data matrix, there were some missing values (e.g., the LBL instrument could not quantify the elements Na and Mg due to the instrument design) and below-detection-limit values. The analytical uncertainty estimates associated with each measured concentration and the detection limits for both instruments were also reported for each analyte and sample.

Receptor model methods

The fundamental principle of receptor modeling is that mass conservation can be assumed and a mass balance analysis can be used to identify and apportion sources of airborne PM in the atmosphere. The approach to obtaining a data set for receptor modeling is to determine a large number of chemical constituents such as elemental concentrations in a number of samples. A mass balance equation can be written to account for all m chemical species in the n samples as contributions from p independent sources

$$x_{ij} = \sum_{p=1}^{P} g_{ip} f_{jp} \tag{1}$$

where x_{ij} is the measured concentration of the jth species in the ith sample, f_{jp} is the concentration of the jth species in material emitted by source p, g_{ip} is the contribution of the pth source to the ith sample, and e_{ij} is the portion of the measurement that cannot be fit by the model. There are a variety of ways to solve Eq. (1) depending on what information is available.

Source Profiles Known

If the number and nature of the sources in the region are known (i.e., P and f_{jp}'s), then the only unknown is the mass contribution of each source to each sample, g_{ip}. These values can be estimated using regression. This approach was first independently suggested by Winchester and Nifong (1971) and by Miller et al. (1972) and is now called the Chemical Mass Balance (CMB) model (Cooper et al., 1984). The problem is typically solved using effective-variance least-squares approach (Cooper et al., 1984) using software that is available from the US Environmental Protection Agency (Watson et al., 1990). Recent CMB studies have been reviewed by Chow and Watson (2002), who concluded that CMB effectively apportioned primary particle sources in a number of studies. It can be used to test emissions inventories. Overall, CMB models can perform well when the pollution sources and their source emission profiles are known with certainty, but this is not usually the case (e.g., due to lack of measured source profiles for particular local sources and changes in source profiles between emission and receptor locations from atmospheric processing). Therefore, other methods are usually needed to address the challenge of source apportionment when source information is not available.

Source Profiles Unknown

Due to the general lack of local source-specific emission information and changes in source emission constituent profiles between emission and impact site (e.g., due to the accretion of secondary aerosols, including sulfates), the area of the most active area of method development has been for the case when the source profiles are not known. These are forms of factor analysis, but are quite different, in practice, from traditional Principal Components Analysis (PCA). In factor analysis, the problem is expanded to the solution of the source profiles and contributions over a set of samples. Thus, the basic equation in matrix form is

$$X = GF' \tag{2}$$

where G is the matrix of source contributions and F' is the transpose of the matrix of source profiles.

Principal Component Analysis

PCAs has been applied to particle compositional data since the 1960s (Blifford and Meeker, 1967). Principal components and factor analysis are names given to several of the variety of forms of eigenvector analysis. PCA derives a limited set of components that explain as much of the *total* variance of all the observable variables (e.g., trace element concentrations) as possible. Other forms of factor analysis were originally developed and used in psychology to provide mathematical models of psychological theories of human ability and behavior (Harman, 1976). However, such eigenvector analyses have found wide application throughout the physical and life sciences. Unfortunately, a great deal of confusion exists in the literature in regard to the terminology of eigenvector analysis. Various changes in the way the method is applied has resulted in it being called factor analysis, PCA, principal components factor analysis, empirical orthogonal function analysis, Karhunen–Loeve transform, etc, depending on the way the data are scaled before analysis or how the resulting vectors are treated after the eigenvector analysis is completed. All of the methods have the same basic objectives; to compress the observable data variables into fewer, underlying dimensions, and to then identify the structure of interrelationships that exist between the variables measured or the cases studied.

Principal component and factor analyses attempt to simplify the description of a system by determining a minimum set of basis vectors that span the data space to be interpreted. In other words, a new set of variables are found as linear combinations of the measured variables so

that the observed variations in the system can be reproduced by a smaller number of these + factors. It has been widely used in studies of airborne PM composition data (e.g., Hopke et al., 1976; Roscoe et al., 1982; Gao et al., 1994). Classical PCA identifies source components and the amount of variance explained by each, but does not directly provide a quantitative source apportionment of the pollution mass in the form presented in Eq. (1). However, the component solutions can be manipulated to provide such a solution.

One approach is specific rotation factor analysis (Koutrakis and Spengler, 1987). It uses a targeted Procrustes rotation. Targets chosen were based on experience with the Six Cities Cohort, composition of sources, and knowledge of the sampling site. A few different targets were also considered.

An alternative approach called Absolute Principal Components Analysis (APCA) (Thurston and Spengler, 1985) has also been used to produce quantitative apportionments. The initial PCA results are rotated and then the component scores are uncentered (relative to a zero pollution reference case), allowing the $PM_{2.5}$ mass values to be regressed against the scores to provide scaling coefficients for both the component scores (related to source impacts) and component loadings (related to source profiles). The solutions are thus optimized to explain concentration data variance without restricting the acceptable mass concentrations. This solution minimizes subjective rotations since it uses a Varimax rotation. However, when source impacts are low, negative observable tracer concentration data (that can be reported at low ambient concentrations) can potentially result in negative source contributions with this method.

Confirmatory factor analysis (CFA) is similar to standard (exploratory) factor analysis, but is based on a hypothesized model and can yield a physically interpretable and uniquely estimable solution. The cost of CFA is that it requires that at least some of the rows of the source profile matrix be known. CFA has been used in the pollution source apportionment setting by CFA models (e.g., Yang, 1994; Gleser, 1997; Christensen and Sain, 2002). A recent modification of CFA is the iterated confirmatory factor analysis (ICFA) approach, which can take on aspects of chemical mass balance analysis, exploratory factor analysis, and CFA by assigning varying degrees of constraint to the elements of the source profile matrix when iteratively adapting the hypothesized profiles to conform to the data (Christensen et al., 2004). This approach attempts to maximally utilize *a priori* source profile information to reduce indeterminacy and enhance interpretability in a multivariate receptor modeling scenario.

Two newer factor-based approaches are UNMIX (Henry and Kim, 1990; Kim and Henry, 1999, 2000) and Positive Matrix Factorization (PMF) (Paatero, 1997, 1999; Paatero et al., 2002). As described below, these methods place restrictions on the possible source impact solutions to require them to meet certain physical constraints (e.g., non-negative source impacts).

Unmix

The fundamental philosophy behind the Unmix model is to impose as few assumptions as possible thereby letting the data speak for themselves. The Unmix model has been developed for the US EPA by Dr. Henry and has several unique features. The most recent version is EPA Unmix 3.0. Earlier versions do not have all the features mentioned below. Unmix has an advanced computationally intensive algorithm to estimate the number of sources that can be seen above the noise level in the data (Henry et al., 1999; Park et al., 2000). Given this estimated number of sources, Unmix uses PCA to reduce the dimensionality of the data space. Geometrical concepts of self-modeling curve resolution are used to ensure the results obey (to within error) nonnegativity constraints on source compositions and contributions. This, however, is not sufficient to uniquely determine the source compositions and contributions (Henry, 1987). Additional constraints determined from the data itself are needed. These are estimated by looking for edges in the data determined by points where one source is small compared to the other sources (Henry, 1997, 2003). Some special features of Unmix are the capability to replace missing data and the ability to estimate large numbers of sources (the current limit is 15) using duality concepts applied to receptor modeling (Henry, 2005). Unmix estimates uncertainties in the source compositions using a blocked bootstrap approach that takes into account serial correlation in the data. The latest version of Unmix is available from the US EPA, Dr. Gary Norris (Norris.Gary@epamail.epa.gov).

Recently, the model has also been applied to the Phoenix, Arizona data (Lewis et al., 2003). The analysis generated source profiles and overall average percentage source contribution estimates for five source categories: gasoline engines ($33 \pm 4\%$), diesel engines ($16 \pm 2\%$), secondary sulfate ($19 \pm 2\%$), crustal/soil ($22 \pm 2\%$), and vegetative burning ($10 \pm 2\%$). They were able to separate motor vehicle contributions into diesel and spark-ignition sources. Diesel emissions were identified by high elemental carbon relative to the OC whereas spark ignition vehicles had a profile with more organic than elemental carbon. The diesel source contributions were much larger on weekdays than weekends, as expected. These results are quite similar to those derived by Henry for this study. Lewis et al. compared their results to those of Ramadan et al. (2000) and found good agreement except for the traffic sources. Lewis et al. had more diesel and less gasoline emissions when compared to Ramadan et al., but the total traffic contributions were similar.

Positive Matrix Factorization

PMF takes a very different approach to the factor analysis problem. All of the other methods use an eigenvector analysis

based on a singular value decomposition (SVD). Any matrix, X, can be defined as

$$X = USV' = \overline{USV'} + E \tag{3}$$

where $\bar{U}$ and $\bar{V}$ are the first p columns of the U and V matrices. The U and V matrices are calculated from eigenvalue-eigenvector analyses of the XX' and $X'X$ matrices, respectively. It can be shown (Lawson and Hanson, 1974; Malinowski, 1991) that the second term on the right-hand side of Eq. (3) estimates X in the least-squares sense that it gives the lowest possible value for

$$\sum_{i=1}^{m}\sum_{j=1}^{n} e_{ij}^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left[x_{ij} - \sum_{p=1}^{P} g_{ip}f_{jp}\right]^2 \tag{4}$$

Thus, an eigenvector analysis is an *implicit* least-squares analysis in that it is minimizing the sum of squared residuals for the model. Paatero and Tapper (1993, 1994) show that in effect in PCA, there is scaling of the data by column or by row in order to normalize the data, and that this scaling will lead to distortions in the analysis. They further show that optimum scaling of the data would be to scale each data point individually so as to have the more precise data having more influence on the solution than points that have higher uncertainties. However, they show that point-by-point scaling results in a scaled data matrix that cannot be reproduced by a conventional factor analysis based on the singular value decomposition. Thus, PMF takes the approach of an *explicit* least-squares approach in which the method minimizes the object function:

$$Q = \mathop{\mathring{a}}_{j=1}^{n}\mathop{\mathring{a}}_{i=1}^{m}\left|\frac{x_{ij} - \mathring{a}_{p=1}^{P} g_{ip}f_{jp}}{s_{ij}}\right|^2 \tag{5}$$

where s_{ij} is an estimate of the "uncertainty" in the jth variable measured in the ith sample. The factor analysis problem is then to minimize $Q(E)$ with respect to G and F with the constraint that each of the elements of G and F is to be non-negative.

Over the past several years, several approaches to solving the PMF problem have been developed. Initially, a program called PMF2 utilizes a unique algorithm (Paatero, 1997) for solving the factor analytic task. PMF2 has been used in a number of recent factor analytic studies (Anttila et al., 1995; Huang et al., 1999; Polissar et al., 1996, 1998, 1999, 2001; Prendes et al., 1999; Xie et al., 1999a; Lee et al., 1999, 2002, 2003; Paterson et al., 1999; Chueinta et al., 2000; Ramadan et al., 2000; Anderson et al., 2001, 2002; Claiborn et al., 2002; Miller et al., 2002; Qin et al., 2002; Begum et al., 2004; Buzcu et al., 2003; Gao et al., 2003; Kim et al., 2003a, b, 2004; Larsen and Baker, 2003; Maykut et al., 2003; Qin and Oduyemi, 2003; Hien et al., 2004; Kim and Hopke, 2004a, b; Larson et al., 2004; Zhou et al., 2004).

Subsequently, an alternative approach that provides a flexible modeling system, the multilinear engine (ME), has been developed for solving the various PMF factor analysis least-squares problems (Paatero, 1999). ME has already been applied to a number of environmental problems (Xie et al., 1999b; Paatero and Hopke, 2002; Hopke et al., 2003; Kim et al., 2003c; Paatero et al., 2003; Ramadan et al., 2003; Chueinta et al., 2004). It is possible to expand the source contribution value to a set of products as follows:

$$x_{ij} = \sum_{p=1}^{P} m_{ip}f_{jp} + e_{ij} \tag{6}$$

where the source contribution, g_{ip}, has been replaced by the modeling term, m_{ip}. This modeling term describes the various physical factors that affect the observed concentrations

$$\begin{aligned} m_{ip} = & W(\omega_i, p)S(\sigma_i, p)I(\iota_i, p) \\ & P(\rho_i, p)C(\chi_i, p)\alpha_{ip}\sum_{h=1}^{24} D(\delta_{ih}, p)V(v_{ih}, p) \end{aligned} \tag{7}$$

where $D(\delta_{ih}, p)$ is the element of D with the index for the wind direction during hour h of day i for the pth source, $V(v_{ih}, p)$ is the element of V with the index for the wind speed during hour h of day i for the pth source, $W(\omega_i, p)$ is the element of W with the index corresponding to day i for the weekday/weekend factor for the pth source, $S(\sigma_i, p)$ is the

Table 1. Summary of the data analyses performed by the various participating groups.

Group	*Phoenix, AZ data*	*Washington, DC data*
BYU		
Eatough	UNMIX	UNMIX
Christensen		Iterated confirmatory FA
Clarkson (CU)		
Hopke/Kim	PMF2 ME Expanded Model (ME)	PMF2
GSF		
Stolzel	APCA	
Harvard (HU)		
Laden/Neas	Target rotated PCA	Target rotated PCA
NYU		
Thurston/Ito/Lall	PMF2 and APCA	PMF2, APCA, and single element multiple regression
USC		
Henry	UNMIX	UNMIX
U. Washington		
Larson	PMF2	

element of the seasonal matrix, S, with the index corresponding to the time-of-year classification of day i for the pth source, $I(\iota_i, p)$ is the inlet factor. $P(\rho_i, p)$ is a correction factor for days when there is precipitation while $C(\chi_i, p)$ is a factor for dry days.

Intercomparison results

The design of this intercomparison study was to provide the two data sets that are outlined above for which there existed both $PM_{2.5}$ compositional and mortality data. Each of the Source Apportionment Workshop participants provided source resolutions using their preferred method(s). Table 1 summarizes the data analyses that were performed on these data sets. These results were then combined with the mortality data in the analyses described in the companion papers by Ito et al. (2005) and Mar et al. (2005).

Table 2. Factor names as assigned by the participants for Washington, DC results.

Group/name analysis	n	Factor names
HU Target Rotated FA	3	Crustal, Coal/2ndary, Auto
BYU/DE Unmix	6	Wood Smoke, Unknown Pb, Mobile, Winter 2ndary, Summer 2ndary, Crustal
BYU/WC Confirmatory FA	8	Wood, Secondary Sulfate, Second_1, Soil, Second_2, Auto/Diesel, Salt, Pb Smelter
USC Unmix	8	Diesel Vehicle, Nitrate, Se, Soil, Sulfate, Gasoline Vehicle, V, As
CU/EK PMF2 (with C Fractions)	10	Sulf_1, Gasoline Vehicle, Sulf_2, Nitrate, Sulf_3, Incinerator, Aged Sea Salt, Soil, Diesel, Oil
CU/XS PMF2	8	Sec. Sulfate, Coal Combustion, Oil Combustion, Soil, Incinerator, Sea Salt, Nitrate, Motor Vehicles
NYU/A APCA	10	Sulfate, Soil, Traffic1, Oil, Traffic2, Industry, Incinerator, Wood, As, Salt
NYU/P PMF2	10	Oil, Soil, Wood, Sulfate1, Sulfate2, Nitrate, Traffic1, Traffic2, Se, Salt
NYU regression	7	Sulfate, Nitrate, Oil, Traffic, Wood, Soil, Primary Coal

Washington, DC

There was a wide range of results for the various analyses in terms of the number of source factors identified. Table 2 presents the names of the sources as provided by the participants that were identified from the Washington DC data. There are a number of similarities in the resolved profiles. The profiles for the secondary nitrate, oil-fired power plants, and soil show considerable similarities across the various Workshop research groups and analysis methods. One of the problems with factor analysis methods is that there is rotational ambiguity in the solutions (Henry, 1987). The problem is discussed in more detail by Paatero et al. (2002, 2005).

Many of the groups resolved two secondary sulfate factors, as had previously been reported by Polissar et al. (2001) and Song et al. (2001). These factors represent the differences in secondary photochemical transformation of SO_2 to SO_4^{2-} (that varies by season), relative to the primary emission of S and Se by coal-fired power plants (which is more consistent year-round). Thus, the total secondary sulfate contributions have been calculated by summing these multiple factors. The correlations between pairs of total sulfate contributions are given in Table 3. In general, there are strong correlations across these sulfate source contribution estimates. Figure 1 shows the distributions of the correlation coefficients obtained between the pairs of results from different investigators for the same source types.

For motor vehicles, there were some of the analyses that separated diesel emissions from spark-ignition vehicles, while other analyses only reported a total motor vehicle emissions source reported that diesels operating at very slow speed and in stop and go traffic produce OC/EC ratios that are like typical spark-ignition emissions (Shah et al., 2004). As much as five times more OC is generated than EC in the cold start/idle mode of heavy-duty diesel vehicles. Thus, the separation of these two sources based on OC and EC values is difficult.

Table 3. Correlation coefficients between the sum of sulfate factor contributions derived from the Washington, DC data.

	HU	BYU/WC	BYU/DE	USC	CU/EK	CU/XS	NYU/A	NYU/P
HU		0.690	0.530	0.413	0.483	0.544	0.655	0.537
BYU/WC	0.690		0.817	0.813	0.855	0.832	0.837	0.821
BYU/DE	0.530	0.817		0.927	0.912	0.953	0.787	0.917
USC	0.413	0.813	0.927		0.948	0.955	0.848	0.926
CU/EK	0.483	0.855	0.912	0.948		0.962	0.895	0.946
CU/XS	0.544	0.832	0.953	0.955	0.962		0.861	0.952
NYU/A	0.655	0.837	0.787	0.848	0.895	0.861		0.883
NYU/P	0.537	0.821	0.917	0.926	0.946	0.952	0.883	

In some cases, there was an identification of multiple "traffic" sources without specific attribution to the types of engine emissions involved. The multiple motor vehicle source contributions are summed and these values compared in Table 4. There is greater variability in these values with several of the time series clearly being quite different from the others.

An ANOVA analysis was performed on the mean source contributions to examine the between-source as compared to the between-group variance. Figure 2 presents the central estimate of the source contributions and the 95% confidence intervals. The results show that between-source variance is greater than between-group variance with $P<0.001$.

The mean and the standard deviations of the various source contributions are summarized in Table 5, where sources have been grouped based on their similarity in source composition profiles. With the exception of the BYU Confirmatory FA, there is good agreement in the secondary sulfate contributions. The soil contributions were very similar except for the Harvard PCA value. The total traffic contributions were similar with the exception of the BYU Unmix, the CU OC/EC PMF and the NYU PMF values that are lower than the others. Table 5 also provides the mean, standard deviations, standard error of the mean and percent standard error for each of the sources that were generally resolved from the Washington data. For the major contributing sources the agreement is quite reasonable as indicated by the percent standard error. For those sources for which their contributions were small, there tended to be more scatter in the results and a larger standard error.

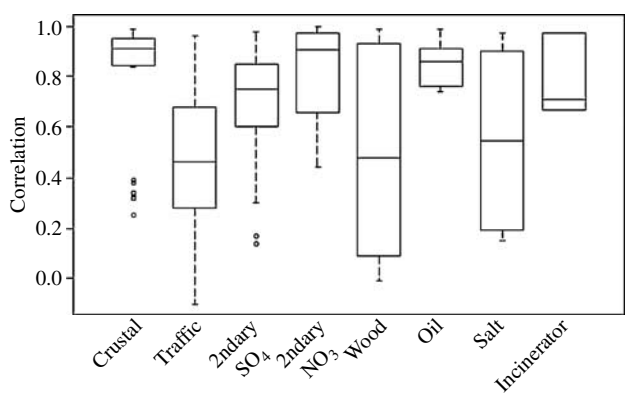

Figure 1. Box and whisker plots showing the distributions of correlation coefficients between the possible pairs of similar source contributions resolved from the Washington, DC data.

With respect to the other source types, similar contributions were obtained when similar profiles were extracted. However, there was not uniformity in the number and nature of the profiles. Some of this variability may arise from operator choices made in the use of these factor analysis methods and some may come from difference in data screening that led to differences in the chemical species and samples included in the analyses.

Phoenix, AZ

The names of the various sources that have been resolved from the analyses of the Phoenix data are presented in Table 6. The mean and standard deviations for the source contributions are summarized in Table 7. Again, as in the Washington, DC results, there is a general convergence in terms of the source profiles and contributions, particularly for the major mass contribution sources. Table 7 also provides mean contributions, standard deviations, standard error of the mean and percent standard error for the identified source categories. It is again necessary to combine source types like spark ignition and diesel to provide a total traffic estimate to obtain good concurrence among the various estimates. For the combined sources (traffic, metals, secondary), the agreement is quite reasonable given the different choices of elements, samples and model parameters to include in the analyses.

An analogous ANOVA analysis was performed for these results and the mean and confidence intervals are shown in Figure 3. Again the between-source variance is greater than between-group variance with $P<0.001$.

Figure 4 shows the distribution of correlation coefficients between the possible pairs of results for similar source types. Again, several of the groups have been able to resolve several traffic related sources that have been tentatively assigned to be diesel and spark-ignition emissions. In the case of "diesel"

Table 4. Correlation coefficients between the sum of motor vehicle factor contributions derived from the Washington, DC data.

	HU	BYU/WC	BYU/DE	USC	CU/EK	CU/XS	NYU/A	NYU/P
HU		0.354	0.351	0.084	0.290	0.458	0.248	0.419
BYU/WC	0.354		0.109	−0.002	0.193	−0.072	0.135	−0.049
BYU/DE	0.351	0.109		0.276	0.795	0.575	0.538	0.650
USC	0.084	−0.002	0.276		0.444	0.113	0.647	0.387
CU/EK	0.290	0.193	0.795	0.444		0.564	0.713	0.650
CU/XS	0.458	−0.072	0.575	0.113	0.564		0.463	0.898
NYU/A	0.248	0.135	0.538	0.647	0.713	0.463		0.606
NYU/P	0.419	−0.049	0.650	0.387	0.650	0.898	0.606	

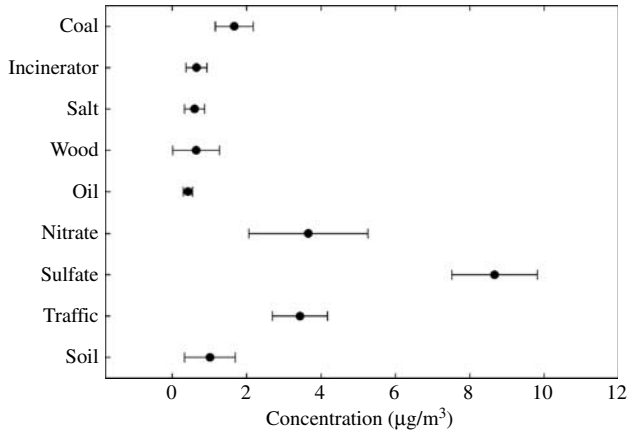

Figure 2. Plot of the ANOVA results for the Washington, DC data.

profiles, manganese appears as previously reported (Ramadan et al., 2000, 2003; Lewis et al., 2003). There were more unique sources identified by the various investigators that made comparisons more difficult than for Washington.

Conclusion

Many similar sources have been identified despite various investigators and different methods being employed. Crustal (soil), sulfate, oil, and salt were most unambiguously identified (generally highest correlation across the sites). Owing to the differences in the resolution of motor vehicle sources, the individual traffic sources are not as well correlated among the results. If the total motor vehicle contributions are estimated, better correspondence among the results is obtained. While these analyses indicate that it

Table 5. Estimated mean $PM_{2.5}$ source contributions (SD) in $\mu g/m^3$ for Washington, DC.

Group/name analysis	ID	Nfac	Soil/ crustal	Traffic & diesel	Secondary sulfate	Nitrate	Oil	Wood smoke	Salt	Incinerator	Others
Harvard Target FA	A	3	3.7 (2.0)	4.6 (2.0)	7.6 (4.0)						
BYU/WC Confirmatory FA	B	8	0.5 (1.0)	3.7 (2.5)	5.1 (4.3)	6.7 (3.5)		1.9 (1.9)	0.9 (0.9)		Pb smelter 0.5 (0.7)
BYU/DE Unmix	C	6	0.8 (0.6)	2.0 (1.6)	8.4 (7.3)	6.7 (5.6)		0.2 (0.5)		0.3 (0.4)	
USC/RH Unmix	D	8	0.7 (1.1)	4.7 (3.5) gas: 1.8 (1.8) dsl: 2.9 (2.6)	7.8 (6.8)	2.5 (2.6)	0.3 (0.3)				Primary coal: 1.2 (1.0) As: 0.7 (0.5)
CU PMF/C fractions	E	10	0.3 (0.5)	4.1 (3.1) gas: 3.8 (2.9) dsl: 0.3 (0.3)	10.6 (7.1) sulf1:7.7 (6.7) sulf2:1.9 (1.9) sulf3:1.1 (0.7)	1.6 (1.6)	0.3 (0.4)		0.4 (0.4)	0.7 (0.5)	
CU PMF/OC-EC	F	8	0.5 (0.7)	1.6 (1.2)	8.4 (6.5)	3.5 (3.0)	0.6 (0.8)		0.6 (0.5)	1.0 (0.6)	Primary coal: 1.7 (1.1)
NYU APCA	G	10	1.0 (1.5)	4.2 (2.8) trfc1: 1.8 (1.7) trfc2: 2.4 (2.1)	9.6 (7.1)		0.6 (1.2)	0.2 (0.4)	0.9 (1.4)	0.6 (0.7)	Industry1: 0.1 (.01) Industry2: 0.3 (0.7)
NYU PMF	H	10	0.5 (0.7)	2.5 (1.4) trfc1: 1.1 (0.7) trfc2: 1.4 (1.1)	10.7 (6.9) sulf1: 3.4 (3.5) sulf2: 7.3 (5.3)	2.5 (2.5)	0.3 (0.4)	0.3 (1.1)	0.2 (0.2)		Se: 1.5 (1.0)
NYU MLR	I	7	1.1 (1.3)	3.5 (1.3)	9.8 (6.3)	2.1 (1.9)	0.4 (0.4)	0.6 (0.7)			Primary coal: 2.1 (1.2)
Mean value (SD)			1.01 (1.04)	3.32 (1.18)	8.67 (1.76)	3.66 (2.16)	0.42 (0.15)	0.60 (0.31)	0.64 (0.72)	0.65 (0.29)	
SEM (%)			0.347 (34.3)	0.393 (11.8)	0.587 (6.77)	0.815 (22.3)	0.17 (40.8)	0.14 (23.0)	0.32 (50.5)	0.14 (22.2)	

Table 6. Factor names as assigned by the participants for Phoenix, AZ results.

Group	Analysis	
Eatough-Summer	Unmix	Primary, Secondary, Mobile, Crustal
Eatough-Winter	Unmix	Primary, Secondary, Mobile, Crustal
Ramadan et al. (2000)	PMF2	Biomass burning, diesel, soil, Cu smelter, motor vehicle, salt, secondary sulfate, other burning/fireworks
Ramadan et al. (2003)	ME	Diesel, biomass burning, soil, Cu smelter, coal fired power plant, salt, spark ignition
Hopke	Expanded Analysis	Soil, marine, fireworks, S + V, OCEC, Pb +, Cd +, Mn, wood, mass, metal, Cu
Henry	Unmix	Diesel, vegetative burning, sulfate, gasoline vehicles, soil, Pb, Ni, Zn, traffic (diesel + gasoline)
Stölzel	APCA	Traffic/oil, coal, wood, soil
Laden	PCA	Crustal material, mobile sources, local industry, fuel oil, woodsmoke
Larson	PMF2	Soil, secondary sulfates, metal, Cl rich, veg/motor, diesel, motor/veg
NYU	APCA	Crustal, traffic, smelter, primary coal burning, diesel, vegetative burning, secondary aerosol, sea salt, unknown (phosphorous)
NYU	PMF2	Diesel, secondary aerosol, vegetative burning, smelter, wood burning, primary coal burning, traffic, sea salt, crustal

Table 7. Estimated mean $PM_{2.5}$ source contributions (SD) in $\mu g/m^3$ for Phoenix, AZ.

	Motor vehicle	Diesel	Crustal	Sea salt	Biomass/wood combustion	Metals	Secondary	Other
Eatough-Summer	Mobile 2.95 (2.64)		1.17 (0.88)		Primary emissions: 0.57 (0.53)		3.83 (2.23)	
Eatough-Winter	Mobile 5.09 (4.82)		1.01 (0.87)		Primary emissions: 5.88 (5.73)		5.61 (3.08)	
Ramadan PMF	4.90 (4.06)	1.65 (1.40)	2.08 (1.30)	0.07 (0.16)	1.22 (0.92)	0.80 (1.16)	0.30 (0.67)	Wildfire/ fireworks: 0.76 (0.83)
Ramadan ME	3.58 (2.97)	0.44 (0.47)	0.84 (0.70)	0.05 (0.11)	1.00 (0.76)	0.18 (0.25)	1.30 (0.80)	
Hopke	4.35 (3.61)	0.60 (0.58)	2.58 (2.21)	0.12 (0.26)	0.46 (0.32)	Cu: 0.68 (0.66) Metals: 0.27 (0.37) Pb +: 0.21 (0.23)	1.72 (0.87)	Unmeas. mass: 2.35 (1.86) Firewks: 0.04 (0.08) Cd +: 0.16 (0.25)
Henry	3.94 (3.37)	1.14 (1.12)	2.53 (2.05)		0.93 (0.95)	Pb: 1.06 (1.38) Zn: 0.25 (0.33)	1.91 (1.21)	Ni: 0.58 (0.75)
Stölzel	Mobile 5.68 (3.33)		1.15 (0.65)		2.32 (1.26)		3.31 (1.68)	
Laden	Mobile: 7.66 (5.18)		1.82 (1.15)					
Larson	5.73 (4.43)	0.79 (0.78)	1.10 (0.81)	0.17 (0.31)	2.13 (1.58)	0.56 (0.74)	3.59 (2.19)	
NYU-A	6.45 (4.67)	0.39 (1.22)	2.08 (2.54)	0.08 (0.26)	1.22 (1.35)	0.71 (1.26)	1.79 (0.94)	Unknown (P): 0.41 (1.10)
NYU-P	4.82 (4.03)	1.61 (1.82)	1.17 (1.02)	0.08 (0.18)	1.72 (1.41)	0.16 (0.17)	2.86 (1.90)	Wood: 0.12 (0.27)
Mean value (SD)	5.66 (1.22)		1.64 (0.65)	0.095 (0.043)	1.44 (0.66)	1.34 (1.14)	2.39 (1.35)	
SEM (%)	0.384 (6.79)		0.205 (12.5)	0.017 (18.2)	0.248 (17.2)	0.548 (40.9)	0.426 (17.8)	

may be possible to separate diesel from gasoline vehicle impacts, further research is needed to ascertain the degree to which the various components of motor vehicular emissions (diesel and spark-ignition) can be separately identified and quantified. However, the overall consistency by source category provided impacts that were very similar for the larger mass contributors across various analyses (e.g., in Washington, secondary sulfate SE = 7% and 11% for traffic

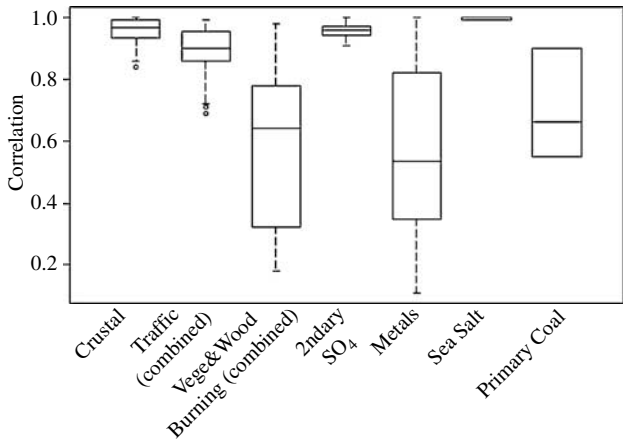

Figure 3. Box and whisker plots showing the distributions of correlation coefficients between the possible pairs of similar source contributions resolved from the Phoenix, AZ data.

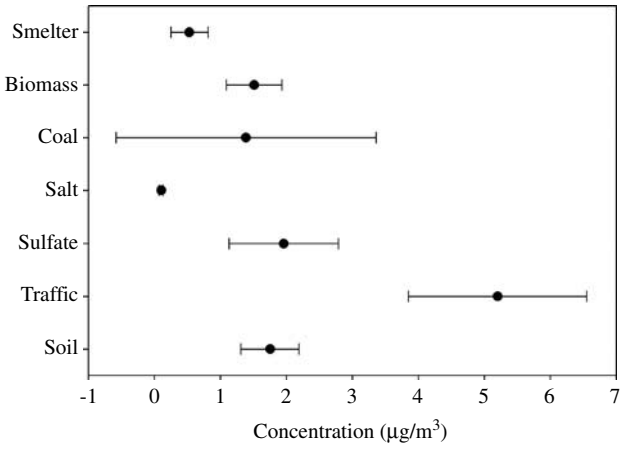

Figure 4. Plot of the ANOVA results for the Phoenix, AZ data.

impacts; in Phoenix, secondary sulfate SE = 17% and 7% for traffic).

Especially important for time-series health effects assessment, the source-specific impacts across analyses were also found to be highly correlated across analyses methods/researchers for the major components (e.g., mean analysis to analysis correlation, $r > 0.9$ for traffic and secondary sulfates in Phoenix and for traffic and secondary nitrates in Washington. The sulfate has mean $r > 0.75$ in Washington). Overall, although these intercomparisons suggest areas where further research is needed (e.g., into the better division of traffic emissions between diesel and gasoline fueled), they also provide considerable support to the contention that $PM_{2.5}$ mass source apportionment results are consistent across users and methods, and that today's mass apportionment methods are robust enough for reliable application to $PM_{2.5}$ health effects assessments.

Acknowledgements

The Workshop was organized under the auspices of the participating US EPA PM Health Effects Research Centers (Grant R827351 at NYU, R827351 at the University of Washington, R827353 at Harvard University, and R927354 at the University of Rochester). We thank the individual researchers who undertook participation in this workshop, often on their own time and resources. The information in this document has been subjected to review by EPA's National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. Support for the organization and administration of the workshop was also provided by the New York State Energy Research and Development Authority (NYSERDA Grant 375-34215). We also thank Columbia University's Arden House Conference Center in Harriman, NY for hosting the May, 2003 workshop that led to the writing of this manuscript.

References

Anderson M.J., et al. Source apportionment of exposure to toxic volatile organic compounds using Positive Matrix Factorization. *J Exp Anal Environ Epidemiol* 2001: 1: 295–307.

Anderson M.J., et al. Source apportionment of exposures to volatile organic compounds: II. Application of receptor models to TEAM study data. *Atmos Environ* 2002: 36: 3643–3658.

Anttila P., et al. Application of Positive Matrix Factorization to source apportionment: results of a study of bulk deposition chemistry in Finland. *Atmos Environ* 1995: 29: 1705–1718.

Begum B.A., et al. Investigation of sources of atmospheric aerosol at urban and semi-urban areas in Bangladesh. *Atmos Environ* 2004: 38: 3025–3038.

Blifford Jr. I.H., and Meaker G.O. A factor analysis model of large scale pollution. *Atmos Environ* 1967: 1: 147–157.

Buzcu B., et al. Source identification and apportionment of fine particulate matter in Houston, TX, using Positive Matrix Factorization. *Environ Eng Sci* 2003: 20: 533–545.

Chow J.C., et al. The DRI thermal/optical reflectance carbon analysis system: description, evaluation and applications in US air quality studies. *Atmos Environ* 1993: 27A(8): 1185–1201.

Chow J., and Watson J. Review of $PM_{2.5}$ and PM_{10} apportionment for fossil fuel combustion and other sources by the chemical mass balance receptor model. *Energy Fuels* 2002: 16: 222–260.

Christensen W.F., et al. Iterated confirmatory factor analysis for pollution source apportionment. Brigham Young University, Department of Statistics Technical Report Series #SD-101-R. *Environmetrics* 2004. Submitted.

Christensen W.F., and Sain S.R. Accounting for dependence in a flexible multivariate receptor model. *Technometrics* 2002: 44: 328–337.

Chueinta W., et al. Investigation of sources of atmospheric aerosol urban and suburban residential areas in Thailand by Positive Matrix Factorization. *Atmos Environ* 2000: 34: 3319–3329.

Chueinta W., et al. A multilinear model for spatial pattern analysis of the measurement of haze and visual effects (MOHAVE) project. *Environ Sci Technol* 2004: 38: 544–554.

Claiborn C.S., et al. Testing the metals hypothesis in Spokane, Washington. *Environ Health Persp* 2002: 110: 547–552.

Cooper J.A., et al. The effective variance weighting for least squares calculations applied to the mass balance receptor model. *Atmos Environ* 1984: 18: 1347–1355.

Gao N., et al. Receptor modeling for airborne ionic species collected in SCAQS, 1987. *Atmos Environ* 1994: 28: 1447–1470.

Gao S., et al. Aerosol chemistry, and light-scattering and hygroscopicity budgets during outflow from East Asia. *J Atmos Chem* 2003: 46: 55–88.

Gleser L.J. Some thoughts on chemical mass balance models. *Chemometrics Intell Lab Systems* 1997: 37: 15–22.

Harman H.H. *Modern Factor Analysis*, 3rd edn. Rev. University of Chicago Press: Chicago, 1976.

Henry R.C. Current factor analysis models are ill-posed. *Atmos Environ* 1987: 21: 1815–1820.

Henry R.C. History and fundamentals of multivariate air quality receptor models. *Chemom Intell Lab Systems* 1997: 37: 37–42.

Henry R.C. Multivariate receptor modeling by N-dimensional edge detection. *Chemom Intell Lab Systems* 2003: 65: 179–189.

Henry R.C. Duality in multivariate receptor models. *Chemom Intell Lab Systems* 2005: 77: 59–63.

Henry R.. et al. Comparing a new algorithm with the classic methods for estimating the number of factors. *Chemom Intell Lab Systems* 1999: 48: 91–97.

Henry R.C., and Kim B.M. Extension of self-modeling curve resolution to mixtures of more than three components. Part 1: finding the basic feasible region. *Chemom Intell Lab Systems* 1990: 8: 205–216.

Hien P.D., et al. PMF receptor modelling of fine and coarse PM_{10} in air masses governing monsoon conditions in Hanoi, Northern Vietnam. *Atmos Environ* 2004: 38: 189–201.

Hopke P.K., et al. The use of multivariate analysis to identify sources of selected elements in the Boston Urban Aerosol. *Atmos Environ* 1976: 10: 1015–1025.

Hopke P.K., et al. Receptor modeling of ambient and personal exposure samples: 1998 Baltimore particulate matter epidemiology-exposure study. *Atmos Environ* 2003: 37: 3289–3302.

Huang S., et al. Testing and optimizing two factor-analysis techniques on aerosol at Narragansett, Rhode Island. *Atmos Environ* 1999: 33: 2169–2185.

Ito K., Christensen W., Eatough D.J., Henry R.C., Kim E., Laden F., Lall R., Larson T.V., Neas L., Hopke P.K., and Thurston G. An investigation of inter-method variability in associations between source-apportioned fine particle mass and daily mortality in Washington, DC. *J Expos Anal Environ Epidemiol* 2005, submitted.

Kim B.M., and Henry R.C. Extension of self-modeling curve resolution to mixtures of more than three components. Part 2: finding the complete solution. *Chemom Intell Lab Syst* 1999: 49: 67–77.

Kim B.M., and Henry R.C. Extension of self-modeling curve resolution to mixtures of more than three components — Part 3. Atmospheric aerosol data simulation studies. *Chemom Intell Lab Systems* 2000: 52: 145–154.

Kim E., and Hopke P.K. Source apportionment of fine particles at Washington, DC utilizing temperature resolved carbon fractions. *J Air Waste Manage Assoc* 2004a: 54: 773–785.

Kim E., and Hopke P.K. Improving source identification of fine particles in a rural Northeastern US area utilizing temperature resolved carbon fractions. *J Geophys Res* 2004b: 109: D09204.

Kim E., et al. Source identification of atlanta aerosol by Positive Matrix Factorization. *J Air Waste Mange Assoc* 2003a: 53: 731–739.

Kim E., et al. Source identification of PM2.5 in an arid northwest U.S. city by Positive Matrix Factorization. *Atmospheric Res* 2003b: 66: 291–305.

Kim E., et al. Incorporation of parametric factors into multilinear receptor model studies of Atlanta Aerosol. *Atmospheric Environ* 2003c: 37: 5009–5021.

Kim E., et al. Improving source identification of Atlanta aerosol using temperature resolved carbon fractions in Positive Matrix Factorization. *Atmospheric Environ* 2004: 38: 3349–3362.

Koutrakis P., and Spengler J.D. Source apportionment of ambient particles in Steubenville, OH using specific rotation factor analysis. *Atmos Environ* 1987: 21: 1511–1519.

Laden F., et al. Association of fine particulate matter from different sources with daily mortality in six US cities. *Environ Health Perspect* 2000: 108: 941–947.

Larsen R.K., and Baker J.E. Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: a comparison of three methods. *Environ Sci Technol* 2003: 37: 1873–1881.

Larson T., et al. Source apportionment of indoor, outdoor and personal $PM_{2.5}$ in Seattle, WA using Positive Matrix Factorization. *J Air Waste Manage Assoc* 2004: 54: 1175–1187.

Lawson C.L., and Hanson R.J. *Solving Least-Squares Problems*. Prentice-Hall: Englewood Cliffs, NJ, 1974.

Lee E., et al. Application of Positive Matrix Factorization in source apportionment of particulate pollutants in Hong Kong. *Atmos Environ* 1999: 33: 3201–3212.

Lee J.H., et al. Identification of sources contributing to the Mid-Atlantic regional aerosol. *J Air Waste Manage* 2002: 52: 1186–1205.

Lee P.K.H., et al. Identification of the major sources contributing to PM2.5 observed in Toronto. *Environ Sci Technol* 2003: 37: 4831–4840.

Lewis C.W., et al. Source apportionment of phoenix $PM_{2.5}$ aerosol with the UNMIX receptor model. *J Air Waste Manage Assoc* 2003: 53: 325–338.

Malinowski E.R. *Factor Analysis in Chemistry*, 2nd edn. Wiley: New York, 1991.

Mar T.F., et al. Associations between air pollution and mortality in Phoenix, 1995–1997. *Environ Health Perspect* 2000: 108: 347–353.

Mar T.F., Koenig J.Q., Larson T.V., Christensen W., Eatough D.J., Henry R.C., Ito K., Kim E., Laden F., Lall R., Neas L., Hopke P.K., and Thurston G. An investigation of inter-method variability in associations between source-apportioned fine particle mass and daily mortality in Phoenix, AZ. *J Expo Anal Environ Epidemiol* 2005. submitted.

Maykut N.N., et al. Source apportionment of PM2.5 at an urban IMPROVE site in Seattle, Washington. *Environ Sci Technol* 2003: 37: 5135–5142.

Miller M.S., et al. A chemical element balance for the Pasadena aerosol. *J Colloid Interface Sci* 1972: 39: 165–176.

Miller S.L., et al. Source apportionment of exposures to volatile organic compounds. I. Evaluation of receptor models using simulated exposure data. *Atmos Environ* 2002: 36: 3629–3641.

Ozkaynak H., and Thurston G.D. Associations between 1980 US mortality rates and alternative measures of airborne particle concentration. *Risk Anal* 1987: 7: 449–460.

Paatero P. Least squares formulation of Robust, non-negative factor analysis. *Chemom Intell Lab Systems* 1997: 37: 23–35.

Paatero P. The multilinear engine — a table-driven least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J Comput Graph Stat* 1999: 8: 854–888.

Paatero P., and Hopke P.K. Utilizing wind direction and wind speed as independent variables in multilinear receptor modeling studies. *Chemom Intell Lab Systems* 2002: 60: 25–41.

Paatero P., and Tapper U. Analysis of different modes of factor analysis as least squares fit problems. *Chemometr Intell Lab Systems* 1993: 18: 183–194.

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

Paatero P., et al. Understanding and controlling rotations in factor analytic models. *Chemom Intell Lab Systems* 2002: 60: 253–264.

Paatero P., et al. Advanced factor analysis of spatial distributions of PM2.5 in the Eastern U.S. *Environ Sci Technol* 2003: 37: 2460–2476.

Paatero P., et al. A graphical diagnostic method for assessing the rotation in factor analytical models of atmospheric pollution. *Atmos Environ* 2005: 39: 193–201.

Park E.S., et al. Estimating the number of factors to include in a high-dimensional multivariate bilinear model. *Commun Stat Simul Comput* 2000: 29: 723–746.

Paterson K.G., et al. Analysis of air quality data using Positive Matrix Factorization. *Environ Sci Technol* 1999: 33: 635–641.

Polissar A.V., et al. The ratio of aerosol optical absorption coefficients to sulfur concentrations, as an indicator of smoke from forest fires when sampling in polar regions. *Atmos Environ* 1996: 30: 1147–1157.

Polissar A.V., et al. Atmospheric aerosol over Alaska: 2. Elemental composition and sources. *J Geophys Res* 1998: 103: 19045–19057.

Polissar A.V., et al. The aerosol at Barrow, Alaska: long-term trends and source locations. *Atmos Environ* 1999: 33: 2441–2458.

Polissar A.V., et al. Atmospheric aerosol over Vermont: chemical composition and sources. *Environ Sci Technol* 2001: 35: 4604–4621.

Prendes P., et al. Source apportionment of inorganic ions in airborne urban particles from Coruna city (NW of Spain) using Positive Matrix Factorization. *Tatlanta* 1999: 49: 165–178.

Qin Y., and Oduyemi K. Atmospheric aerosol source identification and estimates of source contributions to air pollution in Dundee, UK. *Atmos Environ* 2003: 37: 1799–1809.

Qin Y., et al. Comparative testing of PMF and CFA models. *Chemometr Intell Lab Systems* 2002: 61: 75–87.

Ramadan Z., et al. Identification of sources of Phoenix aerosol by Positive Matrix Factorization. *J Air Waste Manage Assoc* 2000: 50: 1308–1320.

Ramadan Z., et al. Comparison of Positive Matrix Factorization (PMF) and multilinear engine (ME-2) for the source apportionment of particulate pollutants. *Chemom Intell Lab Systems* 2003: 66: 15–28.

Roscoe B.A., et al. The use of principal components factor analysis to interpret particulate compositional data sets. *J Air Pollut Control Assoc* 1982: 32: 637–642.

Shah S.D., et al. Emission rates of particulate matter and elemental and organic carbon from in-use diesel engines. *Environ Sci Technol* 2004: 38: 2544–2550.

Song X.H., et al. Source of fine particle composition in the Northeastern US *Atmospheric Environ* 2001: 35: 5277–5286.

Thurston G.D., and Spengler J.D. A quantitative assessment of source contributions to inhalable particulate matter pollution in Metropolitan Boston. *Atmos Environ* 1985: 19: 9–26.

Thurston G.D., et al. The workshop on the source apportionment of PM health effects: inter-comparison of results and implications. *Environ Health Persp* 2005: in press.

Watson J.G., et al. The USEPA/DRI chemical mass balance receptor model, CMB 7.0. *Environ Software* 1990: 5: 38–49.

Winchester J.W., and Nifong G.D. Water pollution in Lake Michigan by trace elements from pollution aerosol fallout. *Water Air Soil Pollut* 1971: 1: 50–64.

Xie Y.L., et al. Identification of source nature and seasonal variations of Arctic aerosol by Positive Matrix Factorization. *J Atmos Sci* 1999a: 56: 249–260.

Xie Y.L., et al. Identification of source nature and seasonal variations of Arctic aerosol by the multilinear engine. *Atmos Environ* 1999b: 33: 2549–2562.

Yang H. Confirmatory factor analysis and its application to receptor modeling. Unpublished Ph.D. dissertation, University of Pittsburgh, Department of Mathematics and Statistics, 1994.

Zhou L., et al. Advanced factor analysis on Pittsburgh particle size distribution data. *Aerosol Sci Technol* 2004: 38(S1): 118–132.