

# The Application of Receptor Modeling to Air Quality Data

## Les modèles récepteur et leurs applications dans le domaine de la qualité de l'air

Philip K. HOPKE\*

### Abstract

Receptor modeling is the application of data analysis methods to elicit information on the sources of air pollutants. Typically, it employs methods of solving the mixture resolution problem using chemical composition data for airborne particulate matter samples. In such cases, the outcome is the identification of the pollution source types and estimates of the contribution of each source type to the observed concentrations. It can also involve efforts to identify the direction of local sources using wind directions or the locations of distant sources through the use of ensembles of air parcel back trajectories. In recent years, there have been improvements in the factor analysis methods that are applied in receptor modeling as well as easier application of trajectory methods. These methods are now in widespread use. The theoretical basics of the methods will be presented and the recent literature will be reviewed.

### Keywords

Receptor modeling. Chemical mass balance. Positive matrix factorization. Air pollution. Trajectory-ensemble methods.

### Résumé

La technique du modèle récepteur est basée sur l'application de méthodes d'analyse des données permettant d'extraire de l'information sur les sources de pollution de l'air. Typiquement, cette technique fait appel à des méthodes pour résoudre le problème du mélange des sources utilisant des données sur la composition chimique d'échantillons de poussières (prélevés dans leur environnement). Il est ainsi possible d'identifier les différents types de source et d'estimer leurs contributions relatives dans les concentrations observées. Cette technique inclut également la possibilité d'identifier la direction des sources en utilisant les directions du vent ou la localisation de sources éloignées en utilisant des rétro-trajectoires. Récemment, on a assisté à l'amélioration des méthodes d'analyse factorielle qui sont appliquées dans les modèles récepteur, et à celle des techniques de définition de trajectoires, qui sont devenues plus facilement applicables. Ces méthodes sont maintenant très largement utilisées. Les bases théoriques des méthodes seront présentées ainsi qu'une recherche bibliographique récente.

### Mots clés

Modèles récepteur. Bilan chimique de masse (CMB). Factorisation matricielle positive (PMF). Pollution atmosphérique. Méthodes de rétro-trajectoires.

\* Center for Air Resources Engineering and Science – Clarkson University – Box 5708 – Potsdam – New York 13699-5708 – USA.

## Introduction

The management of air quality is a difficult but an important problem. In general it involves the identification of the sources of materials emitted into the air, the quantitative estimation of the emission rates of the pollutants, the understanding of the transport of the substances from the sources to downwind locations and the knowledge of the physical and chemical transformation processes that can occur during that transport. All of those elements can then be put together into a mathematical model that can be used to estimate the changes in observable airborne concentrations that might be expected to occur if various actions are taken. Such actions could include the construction of new emission sources as new industries are built and begin to function or the imposition of emission controls on existing facilities in order to reduce the pollutant concentrations.

However, the atmosphere is a very complex system and it is necessary to simplify greatly the descriptions of reality in order to produce a mathematical model capable of being calculated on even the largest and fastest computers. Significant improvements have been made over the past 30 years in the application of chemical transport models as both computing power and the understanding of the atmosphere have improved. However, there are still many instances when the models are insufficient to permit the full development of effective and efficient air quality management strategies, particularly for airborne particulate matter. Hence it is necessary to have other methods available to assist in the identification of sources and the apportionment of the observed pollutant concentrations to those sources.

Such methods are receptor-oriented or receptor models since they are focused on the behavior of the ambient environment at the point of impact as opposed to the source-oriented dispersion models that focus on the transport, dilution and transformations that begin at the source and continue as the pollutants are transported to the sampling or receptor site. These methods have been applied primarily to airborne particulate matter. In the United States, there are two size ranges of particles that are regulated by the US Environmental Protection Agency. Particulate matter in the air with aerodynamic diameters less than 10  $\mu\text{m}$  is called  $\text{PM}_{10}$  whereas the mass concentration of particles less than 2.5  $\mu\text{m}$  is termed  $\text{PM}_{2.5}$ . A comprehensive view of receptor modeling can be found in Hopke [1991]. A review through 1996 was provided as part of a larger report on modeling by Seigneur *et al.* [1997]. An update by Hopke [2003] presented results from 1997 to 2003. Since that time, there has been a rapid expansion of applications of receptor models particularly various advanced forms of factor analysis. European applications of receptor modeling have been reviewed by Viana *et al.* [2008]. This present paper will focus on recent developments and applications primarily elsewhere in the world.

## Background

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 particulate matter in the atmosphere. This methodology has generally been referred to within the air pollution research community as receptor modeling [Hopke, 1985; 1991]. 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. Alternatively, automated electron microscopy can be used to characterize the composition and shape of particles in a series of particle samples. In either case, a mass balance equation can be written to account for all  $m$  chemical species in the  $n$  samples as contributions from  $p$  independent sources:

$$x_{ij} = \sum_{k=1}^p g_{ik} f_{kj} + e_{ij} \quad (1)$$

where  $x_{ij}$  is the measured concentration of the  $j^{\text{th}}$  species in the  $i^{\text{th}}$  sample,  $f_{kj}$  is the concentration of the  $j^{\text{th}}$  species in material emitted by source  $k$ ,  $g_{ik}$  is the contribution of the  $k^{\text{th}}$  source to the  $i^{\text{th}}$  sample and  $e_{ij}$  is the portion of the measurement that cannot be fit by the model.

There exist a set of natural physical constraints on the system that must be considered in developing any model for identifying and apportioning the sources of airborne particle mass [Henry, 1991]. The fundamental, natural physical constraints that must be obeyed are as follows:

- The original data must be reproduced by the model; the model must explain the observations.
- The predicted source compositions must be non-negative; a source cannot have a negative elemental concentration.
- The predicted source contributions to the aerosol must all be non-negative; a source cannot emit negative mass.
- The sum of the predicted elemental mass contributions for each source must be less than or equal to the total measured mass for each element; the whole is greater than or equal to the sum of its parts.

While developing and applying these models, it is necessary to keep these constraints in mind in order to be certain of obtaining physically realistic solutions.

Thus, receptor modeling is a variation on the "spectrochemical mixture resolution" problem in chemometrics. However, there are some additional complicating aspects since source profiles do not remain constant in contrast to molecular spectra and the environmental data tend to have much higher noise in the measurements. This comparison is discussed in more detail by Hopke [1995].

## Receptor models

### Sources Known

There are a variety of ways to solve Equation (1) depending on what information is available. If the number and nature of the sources in the region are known (i.e.  $p$  and  $f_{ik}$  values), then the only unknown is the mass contribution of each source to each sample,  $g_{ki}$ . This approach was first independently suggested by Winchester and Nifong [1971] and by Miller *et al.* [1972]. The problem is typically solved using an effective-variance least-squares approach [Watson *et al.*, 1990; US EPA, 2010a] that is now generally referred to as the chemical mass balance (CMB) model. Software [US EPA, 2010a] is available from the US Environmental Protection Agency. Solution methods using multivariate calibration methods have also been proposed, and were summarized in an earlier review [Seigneur *et al.*, 1997]. There have not been any new method developments in this area nor have there been many new source profiles developed. The US Environmental Protection Agency's library is contained in SPECIATE ([www.epa.gov/ttn/CHIEF](http://www.epa.gov/ttn/CHIEF)). There are some new profiles, particularly for spark-ignition and diesel vehicles, that have been measured and were recently added to the database. The most recent review of CMB studies was by Chow and Watson [2002].

### Sources Unknown

The area of active method development has been in the methods to be used when the source profiles are not known. These are forms of factor analysis, but completely different from traditional principal components analysis and related techniques. In factor analysis, the problem is expanded to the solution of the source profiles and contributions over a set of samples. Two new approaches that have actively used are UNMIX [Henry and Kim, 1989; Kim and Henry, 1999; 2000] and positive matrix factorization (PMF) [Paatero, 1997; 1999].

#### Unmix

Unmix is based on an eigenvector-based analysis to find "edges" in the data [Henry, 2003]. The model uses a transformation method based on the self-modeling curve resolution (SMCR) techniques. Since a unique solution is not possible [Henry, 1987], the SMCR technique restricts the feasible region of the real solution into a small region with explicit physical constraints, such as source compositions must be greater than or equal to zero. Explicit physical constraints form linear inequality constraints in the space spanned by the eigenvectors, and these constraints form the feasible region in eigenvectors' space. UNMIX is designed to resolve the most important sources contributing to the measured mass concentrations. Although the US Environmental Protection Agency has developed this model as a stand-alone program [USEPA, 2010b], it has not been widely used.

The model has been applied to data from Phoenix, AZ [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 ( $10 \pm 2\%$ ). In this study, the authors were able to separate motor vehicle contributions into diesel and spark-ignition sources. Diesel emissions were identified by high elemental carbon relative to the organic carbon whereas spark ignition vehicles had a profile with more organic than elemental carbon. They found a substantial difference in the contribution of diesel emissions between weekend and weekday samples.

#### Positive Matrix Factorization (PMF)

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). However, it can be shown [Lawson and Hanson, 1974; Malinowski, 1991] that a SVD is really solving the problem in a least-squares sense. 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] showed that in PCA, there is scaling of the data by column or by row and that this scaling will lead to distortions in the analysis. They further showed 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 showed 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:

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

where  $s_{ij}$  is an estimate of the "uncertainty" of the  $j$ th variable measured in the  $i$ th sample. The factor analysis problem is then to minimize  $Q(E)$  with respect to  $\mathbf{G}$  and  $\mathbf{F}$  with the constraint that each of the elements of  $\mathbf{G}$  and  $\mathbf{F}$  is to be nonnegative. As with any factor analysis model, rotational ambiguity can be a problem, but there is now a better understanding of the rotational problem [Paatero *et al.*, 2002; Paatero and Hopke, 2009] as well as a useful graphical diagnostic tool [Paatero *et al.*, 2005].

Two approaches have been developed to solve the PMF problem. Initially, a program called PMF2 utilized a unique algorithm [Paatero, 1997] for solving the factor analytic task. For small-and medium-sized problems, this algorithm was found to be more efficient than ALS methods [Hopke *et al.*, 1998]. Subsequently, an alternative approach that provides a flexible modeling system was developed for solving

the various PMF factor analysis least-squares problems [Paatero, 1999]. This approach, called the multilinear engine (ME),

PMF2 was initially applied to data sets of major ion compositions of daily precipitation samples collected over a number of sites in Finland [Juntto and Paatero, 1994] and samples of bulk precipitation [Anttila *et al.*, 1995] in which they are able to obtain considerable information on the sources of these ions. Polissar *et al.* [1996] applied the PMF2 program to Arctic data from seven National Park Service sites in Alaska as a method to resolve the major source contributions more quantitatively. Polissar *et al.* [1998] reanalyzed the Alaska data and proposed an approach to uncertainty estimation that has now been widely used in PMF applications.

PMF has now been used as a source apportionment tool in many airborne PM composition studies [Chueinta *et al.*, 2000; Begum *et al.*, 2004, 2005a, 2009; Kim and Hopke, 2004a, b; Kim *et al.*, 2003a, 2004a, b, 2007; Lee and Hopke, 2006; Lee *et al.*, 2006; Liu *et al.*, 2006; Hwang and Hopke, 2007; Pekney *et al.*, 2006; Sunder Raman and Hopke, 2007; Santoso *et al.*, 2008]. It has been applied to VOC data [Kim *et al.*, 2005], to Aerosol Mass Spectrometry (AMS) data [Lanz *et al.*, 2007; Ulbrich *et al.*, 2008] and to particle number size distributions [Kim *et al.*, 2004c; Zhou *et al.*, 2005a; Ogulei *et al.*, 2006; 2007 a,b; Kasumba *et al.*, 2009]. With the distribution of a version of PMF by the US EPA (2010c), it is now being routinely applied to many air pollution data sets.

### Advanced Model Applications

#### Extended Model

Based on the understanding of transport and dispersion as well as the likely variation in source emissions and atmospheric chemistry with day of week and season suggest that the incorporation of additional information such as wind speed and direction, weekday/weekend or season might permit improved resolution of sources. The availability of the multilinear engine [Paatero, 1999] permits building more complex models that can be fit to the data. Thus, a parametric approach to incorporating meteorological and temporal variables into the analysis has been developed and is term the expanded model.

To present the expanded factor analysis approach, the model is described from the viewpoint of one source, denoted by  $p$ . In reality, there are several sources and the observed concentrations are sums of contributions due to all sources,  $p = 1, \dots, P$ . In the customary bilinear analysis, the contribution  $r_{ijp}$  of source  $p$  on day  $i$  to concentration of chemical species  $j$  is represented by the product  $g_{ip}f_{jip}$ , where  $g_{ip}$  corresponds to the strength of source  $p$  on day  $i$ , and  $f_{jip}$  corresponds to the concentration of compound  $j$  in the emission signature of source  $p$ .

In the expanded PMF analysis, the bilinear equation (1) is augmented by another more complicated

set of equations that contain modeling information. In its most basic form, the contribution  $r_{ijp}$  of source  $p$  is represented by the following expression:

$$r_{ijp} = m_{ip}f_{jip} = \mathbf{D}(\delta_i, p) \mathbf{V}(v_i, p) f_{jip} \quad (3)$$

The known values  $\delta_i$  and  $v_i$  indicate wind direction and wind speed on day  $i$ . The symbols  $\mathbf{D}$  and  $\mathbf{V}$  represent matrices, consisting of unknown values to be estimated during the fitting process. Their columns numbered  $p$  correspond to source number  $p$ . Because of typographic reasons, their indices are shown in parentheses, not as subscripts. The index value  $\delta_i$  for day  $i$  is typically obtained by dividing the average wind direction of day  $i$  (in degrees) by ten and rounding to nearest integer. As an example, if source 2 comes strongly from the wind direction at  $90^\circ$ , then the element  $\mathbf{D}(9,2)$  is likely to become large. The values  $v_i$  are obtained from a chosen classification of wind speeds. The following classification was used in this work: 0 - 1.5 - 2.5 - 3.5 - 5.8 - ¥ m/s. Thus,  $v_i = 2$  for such days when the average wind speed is between 1.5 and 2.5 m/s.

In component form, the equations of the model are

$$x_{ij} = \sum_{p=1}^p g_{ip} f_{jip} + e_{ij} \quad (4)$$

$$x_{ij} = \sum_{p=1}^p m_{ip} f_{jip} + e'_{ij} = \sum_{p=1}^p \mathbf{D}(\delta_i, p) \mathbf{V}(v_i, p) f_{jip} + e'_{ij}$$

The notation  $m_{ip}$  does not indicate a factor element to be determined, such as  $g_{ip}$ , but the expression defined by the physical model in question. In different physical models,  $m_{ip}$  will correspond to different expressions. Because the variability of  $m_{ip}$  is restricted by the model, the second set of equations (4) will produce a significantly poorer fit to the data than the first set of equations (4). The physical model,  $m_{ip}$ , is one of multiple possible models depending on the understanding of the system under study while the mass balance in the first set of equations should be much more applicable. Thus, the error estimates connected with the second set of equations must be (much) larger than the error estimates connected with the first set of equations.

The task of solving this expanded PMF model means that values of the unknown factor matrices  $G$ ,  $F$ ,  $D$ , and  $V$  are to be determined so that the model fits the data as well as possible. In other words, the sum-of-squares value  $Q$ , defined by

$$Q = \sum_{i=1}^I \sum_{j=1}^J (e_{ij} / \sigma_{ij})^2 + \sum_{i=1}^I \sum_{j=1}^J (e'_{ij} / \sigma'_{ij})^2 \quad (5)$$

is minimized with respect to the matrices  $G$ ,  $F$ ,  $D$ , and  $V$ , while the residuals  $e_{ij}$  and  $e'_{ij}$  are determined by equations (4). The error estimates  $\sigma'_{ij}$  must be specified (much) larger than the corresponding error estimates  $\sigma_{ij}$ .

Since there are other sources of variation such as weekend/weekday source activity patterns or seasonal differences in emission rates or in atmospheric chemistry, additional factors are included in the model. In this case, wind direction, wind speed, time of year, and weekend/weekday will be used. In this case, 24 one-hour average values are available for wind speed and direction. Time of year will be aggregated into six two-month periods or *seasons*, indicated for each day  $i$  by the index variables  $\sigma_i$  (The Greek letter  $\sigma$  is used for two purposes:  $\sigma_{ij}$  indicates the error estimates of data values, while  $\sigma_i$  indicates the season number for day  $i$ ). For the values  $i = 1$  to  $i = 60$ ,  $\sigma_i = 1$ , meaning that January and February belong to the first season. For the values  $i = 61$  to  $i = 121$ ,  $\sigma_i = 2$ , and so on.

Instead of the basic equation (3), the non-linear dependencies are now defined by the following multilinear expression:

$$m_{ip} = \sum_{h=1}^{24} \mathbf{D}(\delta_{ih}, p) \mathbf{V}(v_{ih}, p) \mathbf{W}(\omega_p, p) \mathbf{S}(\sigma_p, p) \quad (6)$$

where  $\mathbf{D}(\delta_{ih}, p)$  is the element of  $\mathbf{D}$  with the index for the wind direction during hour  $h$  of day  $i$  for the  $p$ th source,  $\mathbf{V}(v_{ih}, p)$  is the element of  $\mathbf{V}$  with the index for the wind speed during hour  $h$  of day  $i$  for the  $p$ th source,  $\mathbf{W}(\omega_p, p)$  is the element of  $\mathbf{W}$  with the index corresponding to day  $i$  for the weekday/weekend factor for the  $p$ th source, and  $\mathbf{S}(\sigma_p, p)$  is the element of  $\mathbf{S}$  with the index corresponding to the time-of-year classification of day  $i$  for the  $p$ th source. Each of these matrices,  $\mathbf{D}$ ,  $\mathbf{V}$ ,  $\mathbf{W}$ , and  $\mathbf{S}$ , contain unknown values to be estimated in the analysis. The specific factor elements used to fit a particular data point are selected based on the hourly (D, V) or daily (W, S) values of the corresponding variables. Thus, these auxiliary variables are not fitted, but serve as indicators to the values to be fitted.

This model was first applied to a set of simulated data created by the EPA [Paatero and Hopke, 2002]. Sixteen distinct source profiles were used in Palookaville simulation: nine point sources, four industrial complexes, one area source and two highways. Hourly meteorological data including wind speed and direction were used in the ISC3 model to estimate the concentrations at the receptor site. The area profile was a mixture of dust and road profiles. All source profiles with the exception of the petroleum refinery were fixed. The latter profile had some built-in variability (coefficient of variation (CV) of approximately 25%). Temporal modulation of the source strengths (50% CV for most) was found to be essential in being able to resolve the sources by PMF or UNMIX. A total of 366 24-hsamples were generated at the receptor site.

Comparisons with known true data indicate that the analysis is successful. More factors could be determined than by the state-of-the-art bilinear technique PMF. Fifteen of the 16 sources could be resolved. Close inspection of the results reveals that

minor rotational problems still remain. They are mainly visible so that the strongest elements of the strongest factors tend to appear in the weaker factors. The directional information derived from the wind direction factors pointed to each of the point sources. The mass apportionment was much closer to the true values than could be obtained with the simple bilinear modeling. This analysis was based on 24h concentrations and 1-h weather data.

Subsequently it has been applied to several real world data sets [Kim *et al.*, 2003b; Begum *et al.*, 2005b; Zhou *et al.*, 2009] with varying degrees of success. In the case of Atlanta [Kim *et al.*, 2003b], additional sources could be resolved. For Washington, DC, the extended analysis did not resolve more sources, but provided somewhat different sources although there was generally reasonable correlation among the various solutions for this data set [Begum *et al.*, 2005b]. However, in Cleveland [Zhou *et al.*, 2009], they were unable to develop a source resolution that was significantly different than that obtained with the standard bilinear model alone (Equation 1). It is unclear why the expanded model appears to work for some data sets and not for others and further study will be needed.

#### Multiple Time Interval Data

One of the major developments of the past decade or so has been the deployment of continuous monitoring systems for a variety of chemical species associated with airborne particulate matter [Chow *et al.*, 2008]. These measurements are generally made over varying time intervals ranging from 5 minutes to 24 hours. Averaging data to the longest interval will reduce the information content of the data since short-term high events will be averaged out of the data. Thus, to avoid the need for averaging the data to longer intervals, a model was developed that permits utilizing each measurement time on its original time scale [Zhou *et al.*, 2004a; Ogulei *et al.*, 2005].

For each concentration value, there are contributions from several sources, and source contributions have to be averaged so that the concentration value and the source contribution are in the same sampling period. The main equation of the model is as below:

$$x_{sj} = 1/(t_{s2} - t_{s1} + 1) \sum_{p=1}^p (f_{jp} \sum_{i=t_{s1}}^{t_{s2}} g_{ip} \eta_{jm}) + e_{sj} \quad (7)$$

where  $s$  is the sample number,  $j$  stands for the species,  $t_{s2}$  is the end time and  $t_{s1}$  is the start time (described by the number of time units). The shortest sampling interval, 10 min for this data set, is chosen as the time unit. In Equation 1,  $x_{sj}$  is the concentration of  $j$ th species in  $s$ th sample,  $f_{jp}$  is the mass fraction of species  $j$  in particles from source  $p$ ,  $g_{ip}$  is the  $p$ th source mass contribution during the time units for the  $s$ th sample and  $e_{sj}$  is the residual. The source contributions are averaged over the sampling time of  $x_{sj}$ . If all species of all samples are measured within same durations,  $t_{s2} = t_{s1}$ , then Equation 7 becomes the conventional two way receptor model.

Replicated species were measured by more than one method with different time resolution. In Equation 1, different values of the subscript  $m$  correspond to different measurement methods (for one type of data,  $m = 1$ , for the next data type,  $m = 2$  and for a third different measurement,  $m = 3$ ). For each sample  $s$ , the  $m$  value corresponds to the method used for measuring the sample,  $m = m(s)$ . For a replicated species  $j$ , adjustment factors,  $\eta_{jm}$  in Equation 7, are used with the assumption that the concentration values measured by different methods are proportional. An adjustment factor close to 1 suggests a good agreement between the different measurements. Usually, the methods with longest periods provide better accuracy or are defined as reference methods. Thus, their adjustment factors are set to unity by default. For non-replicated species, no adjustment factors are needed, so that their  $\eta_{jm}$  are set to unity by default.

If a source contains no species measured with high temporal resolution, then there is no way to obtain a reliable high time resolution contribution series. To solve this problem, a regularization equation is used to smooth the time series of source contributions, as indicated by Equation 8, where  $g_{ip}$  is the source contribution from the  $p$ th source during the  $i$ th time unit.

$$g_{i+1,p} - g_{i,p} = 0 + \varepsilon_i \quad (8)$$

The total residual sum of squares is composed of residuals from both Equations 7 and 8. When a source includes few high time resolution species, there is limited influence from Equations 7 and 8 will eliminate the unreliable high time resolution details of the contribution series in order to minimize the residuals. When a source contains some high resolution species, reducing the residual in Equation 8 leads to an increase in the residuals of Equation 7 and the high temporal variations tend to be conserved. The balance between the two residuals can be controlled by multiplying the residual in Equation 8 by a small coefficient, such as  $\sim 0.1$ . This balancing was implemented by weighting the residuals with their uncertainties. In both of the applications of this model, good resolutions of the data in Pittsburgh [Zhou *et al.*, 2004a] and Baltimore [Ogulei *et al.*, 2005] were obtained.

#### Size-Composition Data

In some studies, particle samples are collected with cascade impactors such that the PM sample analyses provide size fractionated compositions over time. The DRUM impactor [Raabe *et al.*, 1988] coupled with synchrotron x-ray fluorescence [Bench *et al.*, 2002] can provide data resolved down to 1 hour time intervals for up to 8 particle size fractions. Again, these data cannot be fully exploited using the standard bilinear model because a give source does not necessary emit the same composition particles across the whole range of emitted particle sizes. Thus, a new conceptual model can be proposed in which each source profile is a matrix of particle composition as a function of size. The model to be fit

to the data is then the a sum of outer products of a matrix (source profile) times a vector of source contributions over all of the sources. This model has been applied to three size bin data collected in Detroit, MI [Pere-Trepat *et al.*, 2007].

Nine factors were identified: road salt, industrial (Fe + Zn), cloud processed sulfate, two types of metal works, road dust, local sulfate source, sulfur with dust, and homogeneously formed sulfate. Road salt had high concentrations of Na and Cl. Mixed industrial emissions are characterized by Fe and Zn. The cloud processed sulfate had a high concentration of S in the intermediate size mode. The first metal works represented by Fe in all three size modes and by Zn, Ti, Cu and Mn. The second included a high concentration of small size particle sulfur with intermediate size Fe, Zn, Al, Si, and Ca. Road dust contained Na, Al, Si, S, K and, Fe in the large size mode. The local and homogeneous sulfate factors show high concentrations of S in the smallest size mode, but different time series behavior in their contributions. Sulfur with dust is characterized by S and a mix of Na, Mg, Al, Si, K, Ca, Ti and Fe from the medium and large size modes. This study shows the utilization of time and size resolved DRUM data can assist in the identification of sources and atmospheric processes leading to the observed ambient concentrations.

#### Multiple Sample Type Data

In many panel studies of the effects of airborne particles on health, measurements are made in multiple environments. For example, Hopke *et al.* [2003] report on the analysis of elderly subjects living in a single multifamily residence. Measurements were made at a central outdoor site, an unoccupied room in the building and using personal samplers on specific individuals. Thus, different sources will affect different sample types. Only "external" sources of ambient particles will affect the outdoor samples. However, ambient particles will penetrate into indoor air and add to the exposure observed in the indoor and personal samples. Indoor sources such as cooking and the use of personal care products will not affect the outdoor samples.

The expanded receptor model for this study can be expressed as:

$$x_{ijdt} = \sum_{p=1}^N g_{ipdt} f_{jp} + \sum_{p=N+1}^{N+H} g_{ipdt} f_{jp} \quad (t = 1/2: \text{personal/indoor}) \quad (9)$$

$$x_{jdt} = \sum_{p=1}^N g_{pdt} f_{jp} \quad (t = 3: \text{outdoor}) \quad (10)$$

where  $i$  is the individual (subject or participant) index,  $j$  is the species index,  $d$  is the sampling date index,  $t$  is the type index,  $N$  is the number of external sources,  $H$  is the number of internal sources.  $x_{ijdt}$  denotes the concentration of species  $j$  in the sample of type  $t$  collected by subject  $i$  on date  $d$ ,  $g_{ipdt}$  denotes the contribution of source  $p$  to the sample of type  $t$

collected by subject  $i$  on date  $d$ ,  $f_{jp}$  denotes the relative concentration of species  $j$  in source  $p$ . Since the individuals did not have samples for every environment on every sampling date, the space of the four-dimensional matrixes,  $\mathbf{X}$  and  $\mathbf{G}$ , in equations 9 and 10 is not fully occupied.

In addition to the Baltimore elderly resident study, there have been analyses of data for cardiac patients in the Raleigh-Chapel Hill area of North Carolina [Zhao *et al.*, 2006] and of data for asthmatic children attending a special school for moderate to severe asthmatics in Denver [Zhao *et al.*, 2007a]. In the case of the Denver study, four external sources and three internal sources were resolved from the PM<sub>2.5</sub> data for the three different environments. Secondary nitrate and motor vehicle emissions were the two largest external sources in this study. Cooking was the largest internal source. A significant influence of indoor tobacco smoking on daily personal exposures to particles was observed for those houses in which smokers reside and the environmental tobacco smoke contribution correlated with urinary cotinine levels in these urban schoolchildren. The influence of the high traffic flow outside the school on the indoor air quality was also observed.

#### Mixed Known and Unknown Source Profiles

Recently, it has been shown that the PMF approach can be modified to utilize existing information on source profiles [Amato *et al.*, 2009a; Escrig *et al.*, 2009]. Amato *et al.* [2009a] applied the multilinear engine to data from an urban background site in Barcelona (Spain) to quantify the contribution of road dust resuspension to PM<sub>10</sub> and PM<sub>2.5</sub> concentrations. A recent emission profile of local resuspended road dust had been previously obtained [Amato *et al.*, 2009b]. This *a priori* information was introduced into the model as auxiliary terms in the object function to be minimized by the implementation of so-called "pulling equations" [Paatero and Hopke, 2009].

The multilinear engine permitted an enhanced solution when compared to the basic PMF2 results. The enhanced analysis identified road dust source which accounted for 6.9 µg/m<sup>3</sup> (17%) in PM<sub>10</sub>, 2.2 µg/m<sup>3</sup> (8%) of PM<sub>2.5</sub> and 0.3 µg/m<sup>3</sup> (2%) of PM<sub>1</sub> in addition to the sources identified by PMF2. These results reveal that resuspension was responsible of the 37%, 15% and 3% of total traffic emissions respectively in PM<sub>10</sub>, PM<sub>2.5</sub> and PM<sub>1</sub>. Therefore the overall traffic contribution resulted in 18 µg/m<sup>3</sup> (46%) in PM<sub>10</sub>, 14 µg/m<sup>3</sup> (51%) in PM<sub>2.5</sub> and 8 µg/m<sup>3</sup> (48%) in PM<sub>1</sub>. In PMF2 this mass explained by road dust resuspension was re-distributed among the rest of sources, increasing mostly the mineral, secondary nitrate and aged sea salt contributions.

Escrig *et al.* [2009] applied a similar approach to speciated PM<sub>10</sub> data obtained at three air quality monitoring sites between 2002 and 2007 in a highly industrialized area in Spain. The source apportionment of PM in this area is an especially difficult task. There are industrial mineral dust emissions that need

to be separately quantified from the natural sources of mineral PM. On the other hand, the diversity of industrial processes in the area results in a puzzling industrial emissions scenario. The availability of some specific source profiles for particular major industrial emissions permitted the resolution of the industrial emissions from other sources providing an opportunity to quantitatively evaluate the effectiveness of the abatement programs for air quality improvement in this region.

Thus, it is possible to develop specific conceptual models for a given study that makes physical and chemical sense and then apply those models to the data rather than trying to force all data sets to fit a bilinear model.

## Methods using local wind data

### Conditional Probability Function (CPF)

To analyze point source impacts from various wind directions, the conditional probability function (CPF) [Kim *et al.*, 2003a,b] as calculated using source contribution estimates coupled with wind direction values measured on site. To minimize the effect of atmospheric dilution, daily fractional mass contribution from each source relative to the total of all sources was used rather than using the absolute source contributions. The same daily fractional contribution was assigned to each 3 hours period of a given day to match to the 3 hours average wind direction. Specifically, the CPF is defined as

$$CPF_{\Delta\theta} = \frac{m_{\Delta\theta}}{n_{\Delta\theta}} \quad (11)$$

where  $m_{\Delta\theta}$  is the number of occurrences from wind sector  $\Delta\theta$  that exceeded the threshold criterion, and  $n_{\Delta\theta}$  is the total number of data from the same wind sector. The threshold is set to a relatively high percentile value in the distribution of fractional contributions from a given source. The sources are likely to be located to the directions that have high conditional probability values.

### Methods incorporating back trajectories

The dispersion models describe the transport of the particles from a source to the sampling location. However, using an analogous model of atmospheric transport, it is possible to calculate the position of the air being sampled backward in time from the receptor site from various starting times throughout the sampling interval. The trajectories are then used in residence time analysis (RTA), areas of influence analysis (AIA), quantitative bias trajectory analysis (QTBA), potential source contribution function (PSCF), and residence time weighted concentrations (RTWC). AIA, QTBA and RTWC have only been used

in a single publication for each method and those results are reviewed by Seigneur *et al.* [1997]. PSCF and Simplified QTBA have been primarily used in recent published studies.

### Potential Source Contribution Function (PSCF)

The potential source contribution function (PSCF) receptor model was originally developed by Ashbaugh *et al.* [1985] and Malm *et al.* [1986]. It has been applied in a series of studies over a variety of geographical scales [Cheng *et al.*, 1993; Gao *et al.*, 1993; 1994; 1996]. Air parcel back trajectories ending at a receptor site are represented by segment endpoints. Each endpoint has two coordinates (e.g. latitude, longitude) representing the central location of an air parcel at a particular time. To calculate the PSCF, the whole geographic region covered by the trajectories is divided into an array of grid cells whose size is dependent on the geographical scale of the problem so that the PSCF will be a function of locations as defined by the cell indices  $i$  and  $j$ .

Air parcel backward trajectories were related to the composition of collected material by matching the time of arrival of each trajectory at the receptor site. The movement of an air parcel is described as series of segment end points defined by their latitude and longitude. PSCF values for each grid cell were calculated by counting the trajectory segment endpoints that terminate within the grid cells. The number of endpoints that fall in the  $ij^{th}$  cell is  $n(i, j)$ . The number of endpoints for the same cell when the corresponding samples show concentrations higher than an arbitrarily criterion value is defined to be  $m(i, j)$ . The PSCF value for the  $ij^{th}$  cell is defined as

$$PSCF(i, j) = m(i, j) / n(i, j) \quad (12)$$

In the PSCF analysis, it is likely that the small values of  $n_{ij}$  produce high PSCF values with high uncertainties. In order to minimize this artifact, an empirical weight function  $W(n_{ij})$  proposed by Zeng and Hopke [1989] is commonly applied when the number of the end points per a particular cell was less than about three times the average values of the end points per cell.

Although the trajectory segment endpoints are subject to uncertainty, a sufficient number of endpoints should provide accurate estimates of the source locations if the location errors are random and not systematic. Cells containing emission sources would be identified with conditional probabilities close to one if trajectories that have crossed the cells effectively transport the emitted contaminant to the receptor site. The PSCF model thus provides a means to map the source potentials of geographical areas. It does not apportion the contribution of the identified source area to the measured receptor data.

Xie *et al.* [1999] used PSCF to examine the locations of the sources identified by the PMF analysis of the data from Alert. The results of these analyses

were in agreement with earlier efforts that examined the PSCF maps for the individual chemical constituents in the particle samples. Poissant [1999] used PSCF to examine the likely source locations for total gaseous mercury observed in the St Lawrence River valley. During the winter, fall and spring period the distribution of potential sources reasonably reproduces the North American Hg emission inventory. However, because a single fixed criterion was over the entire year and transport from many of the strong source areas was weak during the summer months, few source areas were observed during the summer data where the concentrations were the lowest. Polissar *et al.* [2001] examined the particle data (black carbon, light scattering and condensation nuclei counts) collected at Point Barrow, Alaska. They found that they could distinguish between biogenic sources of the small particles seen only with the condensation nuclei counter from anthropogenic larger particles that scatter and absorb light. The biogenic particles came primarily from the open areas of the North Pacific Ocean whereas most of the anthropogenic particles came from known industrialized areas of Russia. Begum *et al.* [2005c] have shown that PSCF could accurately identify the location of a large boreal wildfire that affected much of the northeastern United States in July 2002.

### Simplified Quantitative Trajectory Bias Analysis (SQTBA)

Quantitative trajectory bias analysis (QTBA) was developed by Keeler [1987] as a multiple site approach to be able to better identify source regions for measured downwind high concentrations. It was applied to data collected at a number of sites in northeastern United States [Keeler and Samson, 1989]. However, it is very difficult to implement and thus, the full approach has not been applied elsewhere. It is possible to use the basic framework, but simplify the analysis such that it becomes a practical approach to apply [Zhou *et al.*, 2004b; Brook *et al.*, 2004].

The probability of a tracer arriving at a point  $(x, y)$  at time  $t$  is given as:

$$A(x, t) = \int_{t-\tau}^t \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q(x, y, t | x', y', t') dx' dy' dt' \quad (13)$$

where  $Q(x, y, t | x', y', t')$  is the transition probability density function of an air parcel located at  $(x', y')$  and time  $t'$  arriving at the receptor site  $(x, y)$  at time  $t$ .

The transition probability  $Q$  is assumed to be approximately normally distributed about the trajectory with a standard deviation that increases linearly with time upwind:

$$Q(x, y, t | x', y', t') = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[ -\frac{1}{2} \left( \left( \frac{X-x'(t')}{\sigma_x} \right)^2 + \left( \frac{Y-y'(t')}{\sigma_y} \right)^2 \right) \right] \quad (14)$$

where  $(X, Y)$  is the coordinate of the grid center and  $x'(t')$  and  $y'(t')$  are the coordinates of the center line of the trajectory. The  $\sigma_x$  and  $\sigma_y$  are approximated by

$$\sigma_x(t') = \sigma_y(t') = at' \quad (15)$$

with a dispersion speed,  $a$ , equal to 5.4 km/h [Samson, 1980]. The potential mass-transfer field for a given trajectory,  $\bar{T}_k(x, y, t | x', y', t')$ , was integrated over the upwind period,  $\tau$ , of each trajectory to produce a two-dimensional probability of natural transport field.

$$\bar{T}_k(x, y, t | x', y') = \frac{\int_{t-\tau}^t Q(x, y, t | x', y', t') dt'}{\int_{t-\tau}^t dt'} \quad (16)$$

The resulting natural transport potential field,  $\bar{T}_k(x, y, t | x', y', t')$ , for trajectory  $k$ , was weighted by the corresponding concentration,  $\chi_k(x, y)$ , yielding a concentration-weighted mass transfer potential field:

$$\tilde{T}(x | x') = \sum_{k=1}^K T_k(x | x') \chi_k(x) \quad (17)$$

This definition of the weighted potential field is different from the definition given in Keeler [1987] since it is not divided by the sum of concentrations. Thus, the weighted potential field in Equation (17) has the dimension of concentrations.

In PSCF, when cells are crossed by small number of trajectories, false source areas may be found if some of the trajectories also pass real source areas. This problem was solved by Cheng *et al.* [1993] by down-weighting the PSCF values. This "tailing effect" [Cheng and Lin, 2001] problem also exists for SQTBA and RTWC. To solve this problem, the SQTBA field was down-weighted empirically by the following method.

A coefficient  $c_r$  is defined:

$$c_r = \frac{10K}{2\pi (at_0)^2} \quad (18)$$

where  $K$  is total number of trajectories and  $t_0$  is the length of the longest trajectory. The final SQTBA field is obtained by dividing the concentration-weighted field by unweighted field:

$$SQTBA(x, y | x', y') = \frac{\tilde{T}(x, y | x', y')}{\sum_{k=1}^K \bar{T}_k(x, y | x', y')} \left( 1 - \exp \left( - \frac{\sum_{k=1}^K \bar{T}_k(x, y | x', y')}{c_r} \right) \right) \quad (19)$$

The weighted field has the dimensions of concentrations and the unweighted natural field is dimensionless so the final SQTBA field has the dimensions of concentrations. This approach has been used in several studies [Zhou *et al.*, 2004b; Brook *et al.*, 2004; Zhao *et al.*, 2007b].

## Illustrative example

### St. Louis-Midwest Supersite

As an illustrative example, analysis of data from the St. Louis - Midwest Supersite will be presented. These data were analyzed by Lee *et al.* [2006]. St. Louis is a useful example since it is one of the few locations in the United States where there are still large industrial point sources. The metropolitan St. Louis area – which spans Illinois and Missouri – was selected as part of the Supersites program to provide comprehensive characterization of the ambient aerosol in a setting broadly representative of the urban Midwest. It is the latest in a series of studies which have classified sources of ambient fine particles in St. Louis. The Regional Air Pollution Study (RAPS) identified secondary sulfate, motor vehicles, and point sources as major contributors to the St. Louis aerosol [Alpert and Hopke, 1981; Liu *et al.*, 1982; Severin *et al.*, 1983; Chang *et al.*, 1988; Rheingrover and Gordon, 1988]. An analysis of the Harvard Six Cities studies data for St. Louis (1979–1988) identified mobile, crustal, metals, manganese and "coal" sources, the latter dominated by sulfate and presumably of regional origin [Laden *et al.*, 2000]. Both preliminary [Coutant and Swinton, 2002; Kenski and Koerber, 2002] and more-detailed [Lee and Hopke, 2005] source apportionment has been performed on Speciation Trends Network (STN) data for sites in the St. Louis area.

### Sample Collection and Chemical Analysis

Daily 24-hour integrated PM<sub>2.5</sub> samples were collected at the East St. Louis site (Latitude: 38.6122, Longitude: – 90.16028, Elevation: 184 m) from mid-April 2001 through September 2003. This analysis used data for the two-year period of June 1, 2001 to May 31, 2003. Samples were collected from midnight to midnight Central Standard Time using Harvard Impactors (HI) [Marple *et al.*, 1987] for PM<sub>2.5</sub> mass and elements, the Harvard-EPA Annular Denuder System (HEADS) [Koutrakis *et al.*, 1990] for major ions, and the University of Wisconsin low-volume sampling train [Bae *et al.*, 2004] for organic carbon (OC) and elemental carbon (EC). The monitoring location and selected major industrial facilities are shown in Figure 1.

Ambient particulate matter samples were collected onto 37 mm Teflon filters downstream of the HI sampling at 10 L/min. Gravimetric analysis was performed to determine PM<sub>2.5</sub> mass concentration (Harvard School of Public Health, Boston, MA) and the filters were subsequently analyzed by energy-dispersive X-ray fluorescence (XRF) for 40 elements (Desert Research Institute, Reno, Nevada).

HEADS sampled at 10 L/min and consisted of a glass inlet, an impactor, two glass annular denuders (sodium carbonate- and citric acid-coated denuders) in series, followed by a filter pack with a 47 mm Teflon filter followed by a 47 mm Nylon filter. Fine PM sulfate (SO<sub>4</sub><sup>2-</sup>) was determined directly from the Teflon filter, and fine PM nitrate (NO<sub>3</sub><sup>-</sup>) was determined by adding the Teflon filter nitrate and the Nylon filter nitrate. Fine

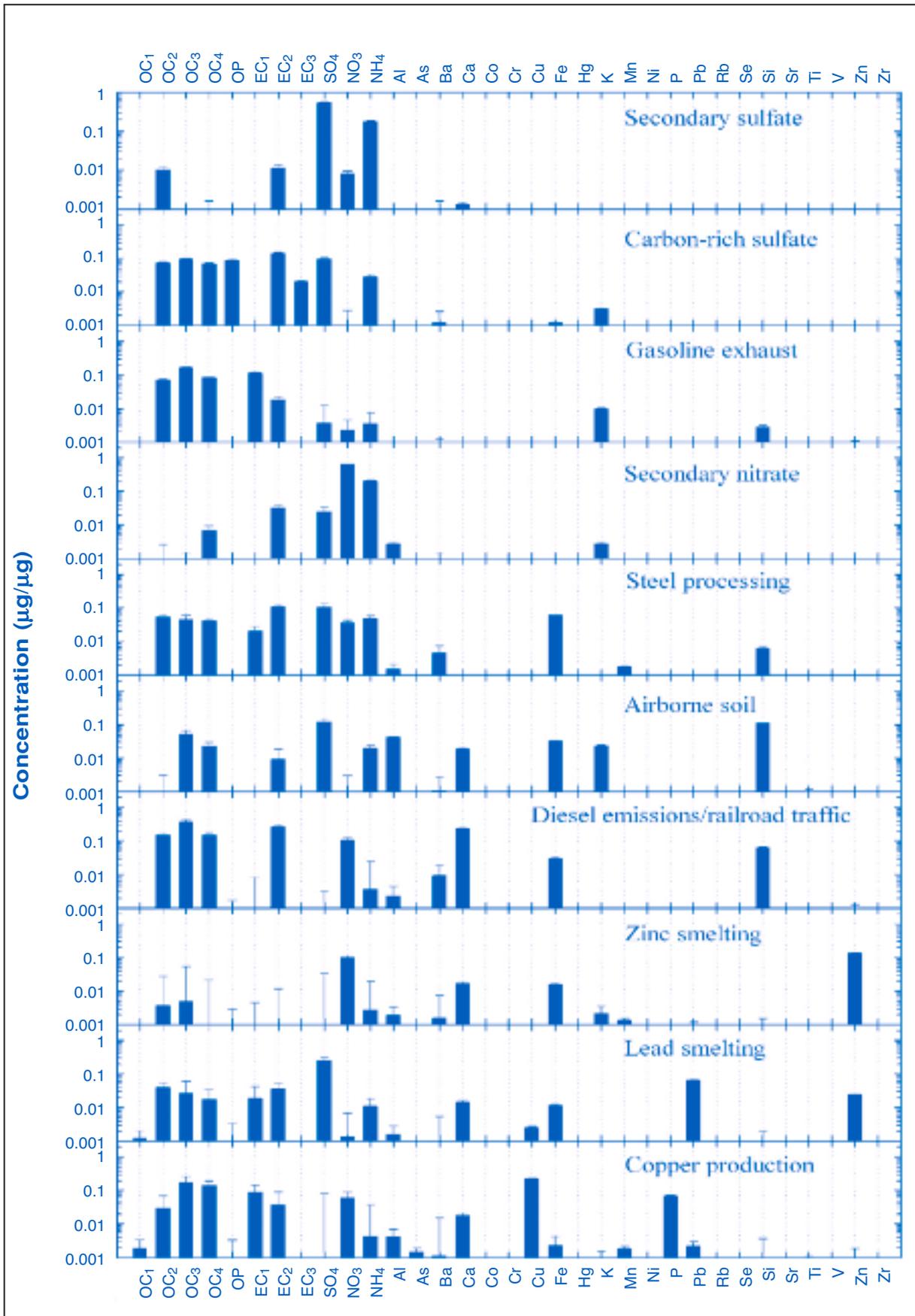


Figure 1.  
Predicted contributions (concentration ± standard deviation) of the PMF-identified sources impacting the East St. Louis site.

Tableau 1.  
Summary Statistics of PM<sub>2.5</sub> Samples Used in the East St. Louis Data Analysis.

	Concentration (ng/m <sup>3</sup> )				Number of BDL** Values (%)	Number of Missing Values (%)
	Geometric Mean*	Arithmetic Mean	Minimum	Maximum		
PM <sub>2.5</sub>	16 128	17 872	3 938	48 652	0	0
OC <sub>1</sub>	61.0	188	0.82	3 998	75 (10.6)	21 (3.0)
OC <sub>2</sub>	630	721	87.6	2 422	0	21 (3.0)
OC <sub>3</sub>	1 000	1 154	174	5 137	0	21 (3.0)
OC <sub>4</sub>	619	692	107	2 678	0	21 (3.0)
OP	249	395	0.0	3 072	14 (2.0)	21 (3.0)
EC <sub>1</sub>	230	515	2.66	3 477	35 (4.9)	21 (3.0)
EC <sub>2</sub>	973	1 038	278	3 422	0	21 (3.0)
EC <sub>3</sub>	61	92	1.02	345	30 (4.2)	21 (3.0)
SO <sub>4</sub> <sup>2-</sup>	3 109	4 028	444	20 762	0	0
NO <sub>3</sub> <sup>-</sup>	1 440	2 158	80	9 904	0	17 (2.4)
NH <sub>4</sub> <sup>+</sup>	1 566	1 912	80.5	6 760	0	33 (4.7)
Al	33.4	45.7	0.0	1 315	20 (2.8)	31 (4.4)
As	1.19	1.94	0.0	26.7	277 (39.1)	31 (4.4)
Ba	18.3	19.3	0.0	382	542 (76.4)	31 (4.4)
Ca	103	125	11.0	136	0	31 (4.4)
Co	0.57	0.77	0.0	7.06	281 (39.6)	31 (4.4)
Cr	0.528	0.31	0.0	8.36	614 (86.6)	31 (4.4)
Cu	8.32	23.7	0.0	746	14 (2.0)	31 (4.4)
Fe	102	125	15.0	1 145	0	31 (4.4)
Hg	0.86	0.94	0.0	62.7	607 (85.6)	31 (4.4)
K	60.7	70.2	12.4	408	0	31 (4.4)
Mn	2.69	4.0	0.0	84.8	69 (9.7)	31 (4.4)
Ni	0.29	0.25	0.0	30.5	564 (79.5)	31 (4.4)
P	4.47	11.9	0.0	305	360 (50.8)	31 (4.4)
Pb	10.9	18.5	0.58	593	6 (0.8)	31 (4.4)
Se	0.98	1.27	0.0	6.91	186 (26.2)	31 (4.4)
Si	94.3	127	9.82	3 489	0	31 (4.4)
Sr	0.62	0.88	0.0	8.98	274 (38.6)	31 (4.4)
Ti	1.45	2.11	0.0	10.3	446 (62.9)	31 (4.4)
V	0.88	0.66	0.0	15.8	565 (79.7)	31 (4.4)
Zn	29.9	44.5	2.19	1496	0	31 (4.4)
Zr	0.53	0.54	0.0	18.8	585 (82.5)	31 (4.4)

\* Data below the detection limit were replaced by the half of the detection limit values for the geometric mean calculations.

\*\* BDL, below the detection limit.

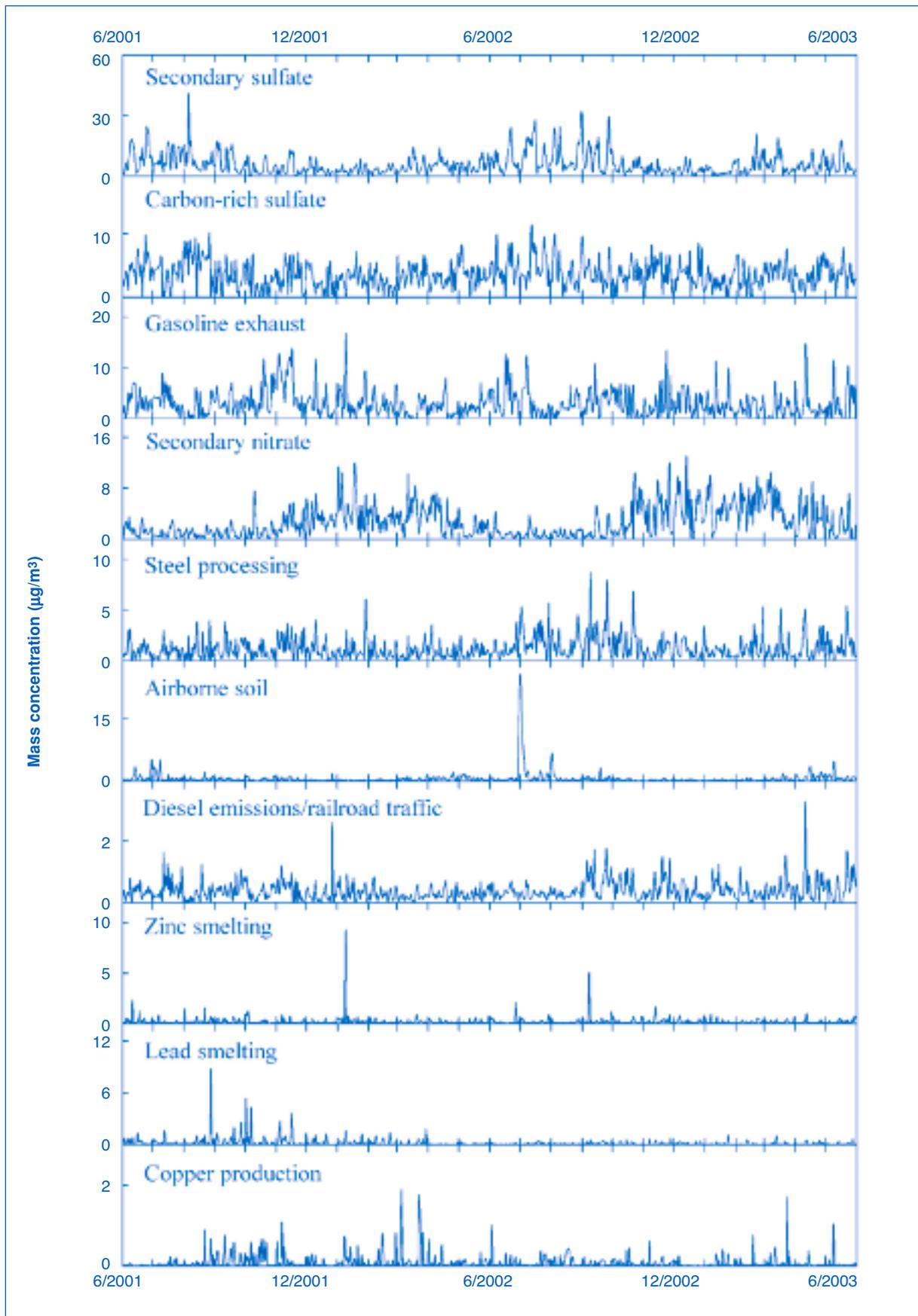


Figure 2.  
Temporal mass concentration contributions at the East St. Louis site as identified by the PMF analysis.

PM ammonium ( $\text{NH}_4^+$ ) was estimated by adding the Teflon filter ammonium and the stoichiometric ammonium associated with the Nylon filter nitrate with assumption that all nitrate on the Nylon filter was ammonium nitrate volatilized from the upstream Teflon filter. Ion chromatographic analyses were conducted by the Desert Research Institute and the Harvard School of Public Health. The OC/EC sampler consisted of a  $\text{PM}_{2.5}$  cyclone that operated at 24 L/min followed by two parallel OC/EC sampling channels. Each channel sampled at 12 L/min, and one channel consisted of an organics denuder followed by a 47 mm prebaked quartz fiber filter. The organics denuder was fabricated with replaceable parallel charcoal-impregnated filter strips to remove gas phase organic compounds prior to filter collection. Thermal-optical analyses were conducted on punches taken from the quartz filter, including 76 filter blanks and 20 laboratory blanks. The quartz filters were analyzed by the IMPROVE/Thermal Optical Reflectance (TOR) protocol [Chow *et al.*, 1993] (Desert Research Institute, Reno, Nevada). On-site hourly average wind direction and speed were measured at 10 m above ground level (AGL) at the East St. Louis site. Samples for which  $\text{PM}_{2.5}$  mass concentrations were missing or coded invalid were excluded from the analysis. Firework event day (July 4th and 5th) samples were also excluded.

The signal-to-noise (SN) ratio was used to screen PM species to be included in the analysis [Paatero and Hopke, 2003]. The SN ratio was calculated by dividing the sum of the values,  $x_{ij}$ , that are above the method detection limit (MDL) for a variable  $j$  in a sample  $i$  by the product of the number of values below the detection limit (BDL),  $m_{DLj}$ , and the MDL. Species with SN ratios between 0.2 and 2 (weak variable) or greater than 2 (good variable) were included in the analysis, while species with their SN ratios less than 0.2 (bad variable) were excluded. XRF sulfur showed good agreement with IC sulfate (estimated slope  $\pm$  standard error of  $2.3 \pm 0.01$ ,  $r^2 = 0.98$ ), and IC sulfate

was used in this source apportionment analysis. XRF sodium (Na) and magnesium (Mg) were excluded from the analysis because they had very high analytical uncertainties. EC1 concentration was corrected for pyrolytic carbon, OP, by subtracting OP from EC1. Species with  $\geq 90\%$  BDL were also excluded. Finally, a total of 709 samples and 33 species were used in the analysis. A summary of the  $\text{PM}_{2.5}$  speciation data used in this study is provided in Table 1.

### Source Identification

Ten sources were identified [Lee *et al.*, 2006] and the average mass apportionments for these sources are provided in Table 2. The identified sources included the four major industrial point sources (steel mill, lead smelter, zinc smelter, and copper products plant), soil, gasoline and diesel vehicle traffic, secondary sulfate and nitrate, and a carbon-rich sulfate factor. Thus, both local area and point sources and distant sources contribute to the fine particle mass concentrations. Figures 1 and 2 present the source profiles and source contribution plots, respectively.

To assist in the identification of the sources, CPF values were calculated for each source type. For the 4 major point sources, there is good agreement between the observed directions and the location of the specific facilities. For example, the CPF for the lead smelter is shown in Figure 3. Examination of the time series of resolved source contributions for the lead smelter (Figure 2) shows higher concentrations through the end of 2001, some higher concentrations during the initial months of 2002 and then relatively low concentrations until the end of the sampling program. It was determined that the smelter installed new control systems at the end of 2001, had some initial operating problems with these new control systems, and then they operated effectively for the rest of the period. Thus, the resolved contributions show a time pattern that corresponds well with the plant operations.

Table 2.  
Average Source Contributions from PMF  
to Measured  $\text{PM}_{2.5}$  Mass Concentration.

	Average source contribution (standard error), %
Secondary sulfate	32.6 (1.1)
Carbon-rich sulfate	19.6 (0.4)
Gasoline exhaust	16.4 (0.6)
Secondary nitrate	15.3 (0.5)
Steel processing	6.8 (0.2)
Airborne soil	4.2 (0.3)
Diesel emissions/railroad traffic	2.1 (0.1)
Zinc smelting	1.3 (0.1)
Lead smelting	1.3 (0.1)
Copper production	0.5 (0.04)

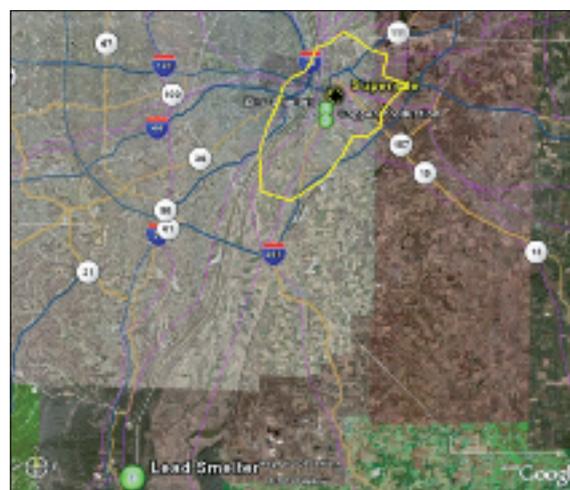


Figure 3.  
CPF plotted on a map of the region  
showing the influence of the primary lead smelter  
to the southwest of the sampling site.

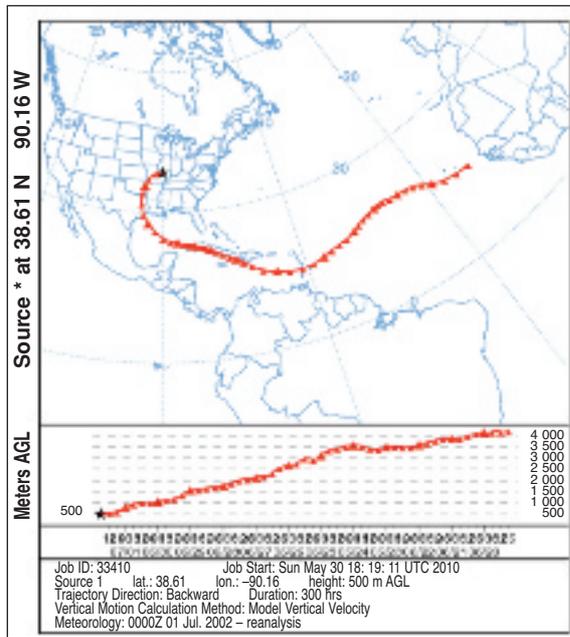


Figure 4.

Back trajectory plot showing the path of the air parcel arriving in St. Louis at 18:00 UTC on July 1, 2002.

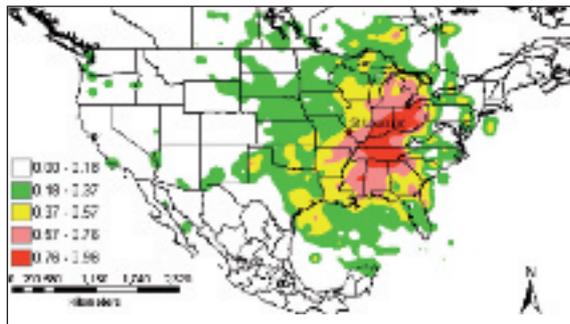


Figure 5.

PSCF map showing source regions for sulfate observed in St. Louis.

Soil is another primary source of airborne particles that normally does not contribute significantly to fine particle mass. It can be noted in Figure 2 that soil contributions are generally low, but there was a major spike early in July 2002. Examining the back trajectory calculated for that day (Figure 4) using the NOAA HYbrid Single-Particle Lagrangian Integrated Trajectory [Draxler and Rolph, 2010] shows that this high concentration event was the result of a Saharan Desert dust episode. This dust cloud was observed across the midwestern and eastern US in early July 2002. Thus, although the low common concentrations of soil arise from local sources primarily resuspended road dust, there are periodic episodes of long range transported dust that produce the highest observed concentrations.

Concentrations of secondary species such as sulfate and nitrate are generally driven by transport. However, Zhao *et al.* [2007b] have shown that there are higher concentrations of nitrate in urban areas of

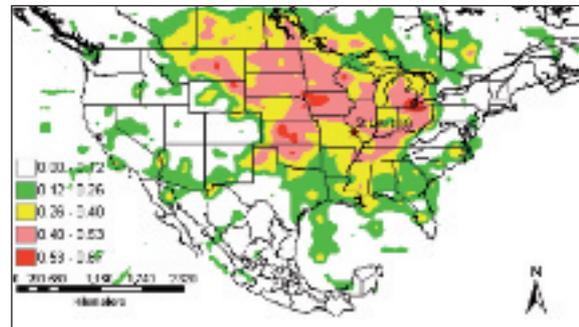


Figure 6.

PSCF map showing source regions for nitrate observed in St. Louis.

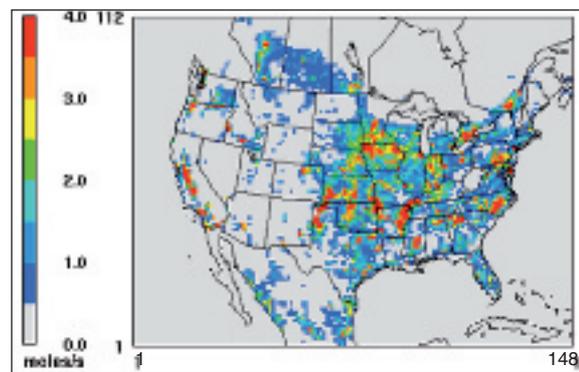


Figure 7.

Map showing the 2001 EPA estimates of ammonia emission rates across the United States.

the midwestern US whereas sulfate was higher in rural areas. They suggested that there was local production of nitric acid since the reaction of nitrogen dioxide with hydroxyl radical is an order of magnitude faster than the comparable reaction with sulfur dioxide. Examination of Figure 2 shows that sulfate is higher in the summer while nitrate is higher in the winter. Figure 5 shows the PSCF map for sulfate observed in St. Louis. Although prevailing winds are from the west, the areas of highest source emissions that influence this area are to the east where there are large numbers of coal-fired power plants along the Ohio and Tennessee Rivers.

There is an interesting contrast in the nitrate PSCF map (Figure 6) where there is some influence from the coal-fired power plants and possibly the Detroit area. However, there are areas of high probability in northwestern Iowa and Kansas where there are no known large scale emissions of  $\text{NO}_x$ . However, ammonium nitrate formation is an equilibrium process that depends on both nitric acid and ammonia concentrations. Figure 7 is an emissions inventory provided by the US EPA and shows high emissions areas for ammonia. The region in northwestern Iowa is dominated by large scale hog production whereas the emissions in Kansas are likely from extensive ammoniacal fertilizer use on large scale crop production.

Thus, commonly available methods provide good resolution of major source types. In the case of specific point sources that have easily defined emission profiles, the impacts of these specific facilities can be determined.

## Future directions

There is growing interest in the use of resolved source contributions in epidemiological studies of the relationships between airborne particle and adverse human health effects [Thurston *et al.*, 2005]. It is thought unlikely that all particles have equal toxicity, hence the problem then exists of how to organize data characterizing particle samples to enter appropriate statistical models. There are too many chemical components typically measured and there is often high correlation among them because they do come from a limited number of common sources. Hence it is anticipated that there will be an increased demand for the easy-to-use software that will permit even complex receptor models to be applied to a wider variety of available data.

Finally, as the EPA continues to reduce its air quality standard values and declares additional areas of the United States to be in non-attainment of the

PM<sub>2.5</sub> ambient air quality standard, there will be application of these receptor model methods to data such as those from the Speciation Network to provide information for state and local air quality management strategy development. A similar problem will arise in Europe as the new European PM<sub>10</sub> standards start to be enforced and areas are identified that have problems that require identification and quantitative apportionment of particle sources. Thus, receptor models continue to be developed and improved and there appears to be a substantial need for the application in the near future.

## Acknowledgment

This paper reflects the work of many students and post-doctoral associates in my laboratory. I would like to acknowledge their efforts in the development and implementation of these methods leading to the substantial literature in the field that is now available. The author also gratefully acknowledges the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT transport and dispersion model and/or READY website (<http://www.arl.noaa.gov/ready.php>) used in this publication.

## References

- Alpert DJ, Hopke PK. A determination of the sources of airborne particles collected during the regional air pollution study. *Atmos. Environ.* 1981 ; 15 : 675-87.
- Amato F, Pandolfi M, Escrig A, Querol X, Alastuey A, Pey J, Perez N, Hopke PK. Quantifying Road Dust Resuspension in Urban Environment by Multilinear Engine: a Comparison with PMF2. *Atmospheric Environ.* 2009a ; 43 : 2770-80.
- Amato F, Pandolfi M, Viana M, Querol X, Alastuey A, Moreno T. Spatial and chemical patterns of PM<sub>10</sub> in road dust deposited in urban environment. *Atmospheric Environment* 2009b ; 43 : 1650-9.
- Anttila P, Paatero P, Tapper U, Jaarvinen O. Application of positive matrix factorization to source apportionment: results of a study of bulk deposition chemistry in Finland. *Atmos. Environ.* 1995 ; 29 : 1705-18.
- Ashbaugh LL, Malm WC, Sadeh WZ. A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. *Atmos. Environ.* 1985 ; 19 : 1263-70.
- Bae MS, Schauer JJ, DeMinter JT, Turner JR, Smith D, Cary RA. Validation of a semi-continuous instrument for elemental carbon and organic carbon using a thermal-optical method. *Atmos. Environ.* 2004 ; 38 : 2885-93.
- Begum BA, Kim E, Biswas SK, Hopke PK. Investigation of sources of atmospheric aerosol at urban and semi-urban areas in Bangladesh. *Atmos. Environ.* 2004 ; 38 : 3025-38.
- Begum BA, Biswas SK, Kim E, Hopke PK, Khaliqzaman M. Investigation of sources of atmospheric aerosol at a hot spot area in Dhaka, Bangladesh. *J. Air & Waste Manage. Assoc.* 2005a ; 55 : 227-40.
- Begum BA, Hopke PK, Zhao W. Source Identification of Fine Particles in Washington DC by Expanded Factor Analysis Modeling. *Environ. Sci. Technol.* 2005b ; 55 : 227-40.
- Begum BA, Kim E, Jeong CH, Lee DW, Hopke PK. Evaluation of the potential source contribution function using the 2002 Quebec forest fire episode. *Atmospheric Environment* 2005 ; 39 : 3719-24.
- Begum BA, Biswas SK, Nasiruddin M, Hossain AMS, Hopke PK. Source Identification of Chittagong Aerosol by Receptor Modeling. *Environmental Engineering Science* 2009 ; 26 : 679-89.

- Bench G, Grant PG, Ueda D, Cliff SS, Perry KD, Cahill TA. The use of STIM and PESA to measure profiles of aerosol mass and hydrogen content, respectively, across Mylar rotating drum impactor samples. *Aerosol Science and Technology* 2002 ; 36 : 642-51.
- Brook JR, Johnson D, Mamedov A. Determination of the source areas contributing to regionally high warm season PM<sub>2.5</sub> in eastern North America. *J. Air Waste Manage. Assoc.* 2004 ; 54 : 1162-9.
- Chang SN, Hopke PK, Gordon GE, Rheingrover SW. Target transformation factor analysis of airborne particulate samples selected by wind-trajectory analysis. *Aerosol Sci. Technol.* 1988 ; 8 : 63-80.
- Cheng MD, Lin CJ. Receptor modeling for smoke of 1998 biomass burning in Central America. *Journal of Geophysical Research* 2001 ; 106 : 22871-86.
- Cheng MD, Hopke PK, Zeng Y. A receptor methodology for determining source regions of particle sulfate composition observed at Dorset, Ontario. *J. Geophys. Res.* 1993 ; 98 : 16839-49.
- Chow J, Watson J. Review of PM<sub>2.5</sub> and PM<sub>10</sub> apportionment for fossil fuel combustion and other sources by the chemical mass balance receptor model. *Energy & Fuels* 2002 ; 6 : 222-60.
- Chow JC, Watson JG, Pritchett LC, Pierson WR, Frazier CA, Purcell RG. The DRI thermal/optical reflectance carbon analysis system: Description, evaluation and applications in U.S. air quality studies. *Atmos. Environ.* 1993 ; 27A : 1185-201.
- Chow JC, Doraiswamy P, Watson JG, Antony-Chen LW, Ho SSH, Sodeman DA. Advances in integrated and continuous measurements for particle mass and chemical composition. *J. Air Waste Manage. Assoc.* 2008 ; 58 : 141-63.
- Chueinta W, Hopke PK, Paatero P. Investigation of sources of atmospheric aerosol in urban and suburban residential areas in Thailand by positive matrix factorization. *Atmos. Environ.* 2000 ; 34 : 3319-29.
- Coutant BW, Swinton KE. Preliminary source apportionment of speciated PM<sub>2.5</sub> measurements in St. Louis, paper #55 presented at the Symposium on Air Quality Measurement Methods and Technology – 2002, Air & Waste Management Association, San Francisco, CA, November 13–15.
- Draxler RR, Rolph GD. HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model access via NOAA ARL READY Website (<http://ready.arl.noaa.gov/HYSPLIT.php>). NOAA Air Resources Laboratory 2010, Silver Spring, MD.
- Escrig A, Monfort E, Celades I, Querol X, Amato F, Minguillón MC, Hopke PK. Application of optimally scaled target factor analysis for assessing source contribution of ambient PM<sub>10</sub>. *J. Air Waste Manage. Assoc.* 2009 ; 59 : 1296-307.
- Gao N, Cheng MD, Hopke PK. Potential source contribution function analysis and source apportionment of sulfur species measured at Rubidoux, CA during the Southern California Air Quality Study, 1987. *Anal. Chim. Acta* 1993 ; 277 : 369-80.
- Gao N, Cheng MD, Hopke PK. Receptor modeling for airborne ionic species collected in SCAQS, 1987. *Atmos. Environ.* 1994 ; 28 : 1447-70.
- Gao N, Hopke PK, Reid NW. Possible sources of some trace elements found in airborne particles and precipitation in Dorset, Ontario. *J. Air Waste Manage. Assoc.* 1996 ; 46 : 1035-47.
- Henry RC. Current factor analysis models are ill-posed. *Atmos. Environ.* 1987 ; 21 : 1815-20.
- Henry RC. Multivariate receptor models. In Receptor Modeling for Air Quality Management. Hopke PK (ed). Elsevier 1991, Amsterdam : 117-47.
- Henry RC. Multivariate receptor modeling by N-dimensional edge detection. *Chemom. Intell. Lab. Syst.* 2003 ; 65 : 179-89.
- Henry RC, Kim BM. Extension of self-modeling curve resolution to mixtures of more than three components. Part 1. Finding the basic feasible region. *Chemom. Intell. Lab. Syst.* 1989 ; 8 : 205-16.
- Hopke PK. Receptor Modeling in Environmental Chemistry. Wiley, New York 1985.
- Hopke PK (ed). Receptor Modeling for Air Quality Management. Elsevier, Amsterdam 1991.
- Hopke PK. The mixture resolution problem applied to airborne particle source apportionment. In Chemometrics in Environmental Chemistry, Springer 1995, Heidelberg : 47-86.
- Hopke PK. Recent developments in receptor modeling. *J. Chemometrics* 2003 ; 17 : 255-65
- Hopke PK, Paatero P, Jia H, Ross RT, Harshman RA. Three-way (PARAFAC) factor analysis: examination and comparison of alternative computational methods as applied to ill-conditioned data. *Chemom. Intell. Lab. Syst.* 1998 ; 43 : 25-42.
- Hopke PK, Ramadan Z, Paatero P, Norris G, Landis M, Williams R, Lewis CW. Receptor Modeling of Ambient and Personal Exposure Samples: 1998 Baltimore Particulate Matter Epidemiology-Exposure Study. *Atmospheric Environ.* 2003 ; 37 : 3289-302.
- Hwang IJ, Hopke PK. Estimation of source apportionment and potential source locations of PM<sub>2.5</sub> at a west coastal IMPROVE site. *Atmos. Environ.* 2007 ; 41 : 506-18.
- Junto S, Paatero P. Analysis of daily precipitation data by positive matrix factorization. *Environmetrics* 1994 ; 5 : 127-44.

- Kasumba J, Hopke PK, Chalupa DC, Utell MJ. Comparison of sources of submicron particle number concentrations measured at two sites in Rochester, NY. *Science of the Total Environment* 2009 ; 407 : 5071-84.
- Keeler GJ. (1987) A hybrid approach for source apportionment of atmospheric pollutants in the north-eastern United States, Ph.D. Dissertation, 1987, University of Michigan, Ann Arbor MI.
- Keeler GJ, Samson J. Spatial representativeness of trace element ratios. *Environmental Science and Technology* 1989 ; 23 : 1358-64.
- Kenski DM, Koerber M. CMB analysis of urban PM<sub>2.5</sub> data, paper presented at the Symposium on Air Quality Measurement Methods and Technology – 2002, Air & Waste Management Association, San Francisco, CA, November 13–15.
- Kim BM, Henry RC. 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 BM, Henry RC. Extension of self-modeling curve resolution to mixtures of more than three components. Part 3. Atmospheric aerosol data simulation studies. *Chemom. Intell. Lab. Syst.* 2000 ; 52 : 145-54.
- Kim E, Hopke PK, Edgerton ES. Source identification of Atlanta aerosol by positive matrix factorization. *J. Air & Waste Manage. Assoc.* 2003a ; 53 : 731-9.
- Kim E, Hopke PK, Paatero P, Edgerton ES. Incorporation of parametric factors into multilinear receptor models studies of Atlanta aerosol. *Atmospheric Environ.* 2003b ; 37 : 5009-21.
- Kim E, Hopke PK. Improving source identification of fine particles in a rural northeastern US area utilizing temperature-resolved carbon fractions. *J. Geophys. Res.* 2004a ; 109 : D09204.
- Kim E, Hopke PK. Source apportionment of fine particles at Washington, DC utilizing temperature resolved carbon fractions. *J. Air & Waste Manage. Assoc.* 2004b ; 54 (7) : 773-85.
- Kim E, Hopke PK, Edgerton ES. Improving source identification of Atlanta aerosol using temperature resolved carbon fractions in Positive Matrix Factorization. *Atmos. Environ.* 2004a ; 38 (20) : 3349-62.
- Kim E, Hopke PK, Larson TV, Mayjut NN, Lewtas J. Factor analysis of Seattle fine particles, *Aeros. Sci. And Techno.* 2004b ; 38 (7) : 724-38.
- Kim E, Hopke PK, Larson TV, Covert DS. Analysis of ambient particle size distributions using UNMIX and Positive Matrix Factorization. *Environmental Science and Technology* 2004c ; 38 : 202-9.
- Kim E, Brown SG, Hafner HR, Hopke PK. Characterization of Non-Methane Volatile Organic Compounds Sources in Houston during 2001 using Positive Matrix Factorization. *Atmospheric Environ.* 2005 ; 39 : 5934-46.
- Kim HS, Huh JB, Hopke PK, Holsen TM, Yi SM. Characteristics of the major chemical constituents of PM<sub>2.5</sub> and smog events in Seoul, Korea in 2003 and 2004. *Atmos. Environ.* 2007 ; 41 (32) : 6762-70.
- Koutrakis P, Wolfson JM, Brauer M, Spengler JD. Design of a glass impactor for an annular denuder/filter pack system. *Aerosol Sci. Technol.* 1990 ; 12 : 607-12.
- Laden F, Neas FM, Dockery DW, Schwartz J. Association of fine particulate matter from different sources with daily mortality in six U.S. cities. *Environ. Health Perspect.* 2000 ; 108 : 941-7.
- Lawson CL, Hanson RJ. Solving Least Squares Problems. Prentice-Hall: Englewood Cliffs, NJ, 1974.
- Lanz VA, Alfara MR, Baltensperger U, Buchmann B, Hueglin C, Prevot ASH. Source apportionment of submicron organic aerosol at an urban site by factor analytical modeling of aerosol mass spectra. *Atmos. Chem. Phys.* 2007 ; 7 (6) : 1503-22.
- Lee JH, Hopke PK. Apportioning sources of PM<sub>2.5</sub> in St. Louis, MO using speciation trends network data. *Atmos. Environ.* 2006 ; 40 : S360-77.
- Lee JH, Hopke PK, Turner JR. Source Identification of Airborne PM<sub>2.5</sub> at the St. Louis-Midwest Supersite. *J. Geophys. Res.* 2006 ; 111 : D10S10.
- Lewis CW, Norris GA, Henry RC, Conner TL. Source apportionment of Phoenix PM<sub>2.5</sub> aerosol with the UNMIX receptor model. *J. Air Waste Manage. Assoc.* 2003 ; 53 : 325-38.
- Liu CK, Roscoe BA, Severin KG, Hopke PK. The application of factor analysis to source apportionment of aerosol mass. *Am. Ind. Hyg. Assoc. J.* 1982 ; 43 : 314-8.
- Liu W, Wang Y, Russell AG, Edgerton ES. (2006) Enhanced source identification of Southeast aerosols using temperature resolved carbon fractions and gas phase components. *Atmos. Environ.* 2006 ; 40 : S445-66.
- Malinowski ER. Factor Analysis in Chemistry (2nd ed.). Wiley 1991, New York.
- Malm WC, Johnson CE, Bresch JF. Application of principal component analysis for purposes of identifying source-receptor relationships. In: Receptor Methods for Source Apportionment, Pace TG (ed) 1986. Air Pollution Control Association: Pittsburgh, PA, 127-48.
- Marple VA, Rubow KL, Turner W, Spengler JD. Low flow rate sharp cut impactors for indoor air sampling: design and calibration. *J. Air Poll. Control Assoc.* 1987 ; 37 : 1303-7.

- Miller MS, Friedlander SK, Hidy GM. A chemical element balance for the Pasadena aerosol. *J. Colloid Interface Sci.* 1972 ; 39 : 65-176.
- Ogulei D, Hopke PK, Zhou L, Pancras JP, Park SS, Ondov JM. Receptor Modeling for Multiple Time Resolved Species: The Baltimore Supersite; *Atmospheric Environ.* 2005 ; 39 : 3751-62.
- Ogulei D, Hopke PK, Wallace LA. Analysis of indoor particle size distributions from an occupied town-house using positive matrix factorization. *Indoor Air* 2006 ; 16 : 204-15.
- Ogulei D, Hopke PK, Ferro AR, Jaques PA. Factor analysis of submicron particle size distributions near a major United States-Canada trade bridge. *J Air Waste Manage Assoc* 2007a ; 57 : 190-203.
- Ogulei D, Hopke PK, Chalupa D, Utell M. Modeling source contributions to submicron particle number concentrations measured in Rochester, NY. *Aerosol Sci Tech* 2007b ; 41 : 179-201.
- Paatero P. Least squares formulation of robust, nonnegative factor analysis. *Chemom. Intell. Lab. Syst.* 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 : 854-88.
- Paatero P, Hopke PK. Utilizing wind direction and wind speed as independent variables in multilinear receptor modeling studies. *Chemom. Intell. Lab. Syst.* 2002; 60 : 25-41.
- Paatero P, Hopke PK. Discarding or downweighting high noise variables in factor analytic models. *Anal. Chim. Acta* 2003 ; 490 (1) : 277-89.
- Paatero P, Hopke PK. (2009) Rotational tools for factor analytic models. *J. Chemometrics* 2009 ; 23 : 91-100.
- Paatero P, Tapper U. Analysis of different modes of factor analysis as least squares fit problems. *Chemom. Intell. Lab. Syst.* 1993 ; 18 : 183-94.
- Paatero P, Hopke PK, Song XH, Ramadan Z. Understanding and controlling rotations in factor analytic models. *Chemom. Intell. Lab. Syst.* 2002 ; 60 : 253-64.
- Paatero P, Hopke PK, Begum BA, Biswas SK. A graphical diagnostic method for assessing the rotation in factor analytical models of atmospheric pollution. *Atmos. Environ.* 2005 ; 39 : 193-201.
- Pekney NJ, Davidson CI, Zhou L, Hopke PK. Application of PSCF and CPF to PMF-modeled sources of PM<sub>2.5</sub> in Pittsburgh. *Aeros. Sci. Technol.* 2006 ; 40 (10) : 952-61.
- Pere-Trepas E, Hopke PK, Paatero P. Source Apportionment of Time and Size Resolved Ambient Particulate Matter Measured with a Rotating DRUM Impactor. *Atmospheric Environ.* 2007 ; 41 : 5921-33.
- Poissant L. Potential sources of atmospheric total gaseous mercury in the St. Lawrence River valley. *Atmos. Environ.* 1999 ; 33 : 2537-47.
- Polissar AV, Hopke PK, Harris JM. Source regions for atmospheric aerosol measured at Barrow, Alaska. *Environ. Sci. Technol.* 2001 ; 35 : 4214-26.
- Polissar AV, Hopke PK, Malm WC, Sisler JF. 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 : 1147-57.
- Polissar AV, Hopke PK, Malm WC, Sisler JF. Atmospheric aerosol over Alaska: 2. Elemental composition and sources. *J. Geophys. Res.* 1998 ; 103 : 19045-57.
- Raabe OG, Braaten DA, Axelbaum RL, Teague SV, Cahill TA. (1988) Calibration Studies of the DRUM Impactor. *J. Aerosol Sci.* 1998 ; 19 : 183-95.
- Rheingrover SW, Gordon GE. Wind-trajectory method for determining compositions of particles from major air pollution sources. *Aerosol Sci. Technol.* 1988 ; 8 : 29-61.
- Samson PJ. Trajectory analysis of summer time sulfate concentrations in the northeastern United States. *Journal of Applied Meteorology* 1980 ; 19 : 1382-94.
- Santoso M, Hopke PK, Hidayat A, Dwiana LD. Sources identification of the atmospheric aerosol at urban and suburban sites in Indonesia by Positive Matrix Factorization. *Sci. Total Environ.* 2008 ; 397 : 229-37.
- Seigneur C, Pai P, Louis JF, Hopke PK, Grosjean D. Review of Air Quality Models for Particulate Matter, Report 4669. American Petroleum Institute 1997, Washington, DC.
- Thurston GD, Ito K, Mar T, Christensen WF, Eatough DJ, Henry RC, Kim E, Laden F, Lall R, Larson TV, Liu H, Neas L, Pinto J, Stolzel M, Suh H, Hopke PK. Workgroup report: Workshop on source apportionment of particulate matter health effects – Intercomparison of results and implications. *Environmental Health Perspectives* 2005 ; 113 : 1768-74.
- Ulbrich IM, Canagaratna MR, Zhang Q, Worsnop DR, Jimenez JL. Interpretation of organic components from positive matrix factorization of aerosol mass spectrometric data. *Atmos. Chem. Phys. Discuss.* 2008 ; 8 : 6729-91.
- U.S. Environmental Protection Agency (USEPA). Chemical Mass Balance (CMB) Model. 2010a. [www.epa.gov/ttn/scram/receptor\\_cmb.htm](http://www.epa.gov/ttn/scram/receptor_cmb.htm)
- U.S. Environmental Protection Agency (USEPA). EPA Unmix 6.0 Model. 2010b. [www.epa.gov/heasd/products/unmix/unmix.html](http://www.epa.gov/heasd/products/unmix/unmix.html)

- U.S. Environmental Protection Agency (USEPA). EPA Positive Matrix Factorization (PMF) 3.0 Model; 2010c. [www.epa.gov/heasd/products/pmf/pmf.html](http://www.epa.gov/heasd/products/pmf/pmf.html)
- Viana M, Kuhlbusch TAJ, Querol X, Alastuey A, Harrison RM, Hopke PK, Winiwarter W, Vallius A, Szidat S, Prevot ASH, Hueglin C, Bloemen H, Wahlin P, Vecchi R, Miranda AI, Kasper-Giebl A, Maenhaut W, Hitenberger R. Source apportionment of particulate matter in Europe: A review of methods and results. *J. Aerosol Sci.* 2008 ; 39 : 827-49.
- Watson JG, Robinson NF, Chow JC, Henry RC, Kim BM, Pace TG, Meyer EL, Nguyen Q. The USEPA/DRI chemical mass balance receptor model, CMB 7.0. *Environ. Software* 1990 ; 5 : 38-49.
- Winchester JW, Nifong GD. Water pollution in Lake Michigan by trace elements from pollution aerosol fallout. *Water Air Soil Pollut* 1971 ; 1 : 50-64.
- Xie YL, Hopke PK, Paatero P, Barrie LA, Li SM. Locations and preferred pathways of possible sources of Arctic aerosol. *Atmos. Environ.* 1999 ; 33 : 2229-39.
- Zeng Y, Hopke PK. A study on the sources of acid precipitation in Ontario, Canada. *Atmospheric Environment* 1989 ; 23 : 1499-509.
- Zhao W, Hopke PK, Norris G, Williams R, Paatero P. Source Apportionment and Analysis on Ambient and Personal Exposure Samples with a Combined Receptor Model and an Adaptive Blank Estimation Strategy. *Atmospheric Environ.* 2006 ; 40 : 3788-801.
- Zhao W, Hopke PK, Gelfand EW, Rabinovitch N. Use of an Expanded Receptor Model for Personal Exposure Analysis in Schoolchildren with Asthma. *Atmospheric Environ.* 2007a ; 41 : 4084-96.
- Zhao W, Hopke PK, Zhou L. Spatial Distribution of Source Locations for Particulate Nitrate and Sulfate in the Upper-Midwestern United States. *Atmospheric Environ.* 2007b ; 41 : 1831-47.
- Zhou L, Hopke PK, Paatero P, Ondov JM, Pancras JP, Penney NJ, Davidson CI. Advanced factor analysis for multiple time resolution aerosol composition data. *Atmospheric Environ.* 2004a ; 38 : 4909-20.
- Zhou L, Hopke PK, Liu W. Comparison of two trajectory-based models for locating particle sources for two rural New York sites. *Atmospheric Environment* 2004b ; 38 : 1955-63.
- Zhou L, Hopke PK, Stanier C, Pandis SN. Mining airborne particulate size distribution data by positive matrix factorization. *J Geophys Res* 2005a ; 110 : D07S19.
- Zhou L, Hopke PK, Stanier C, Pandis SN, Ondov JM, Pancras JP. Investigation of the relationship between chemical composition and size distribution of airborne particles by partial least squares (PLS) and positive matrix factorization (PMF). *J Geophys Res* 2005b : D07S18.

