EPA United States
Environmental Protection
Agency

EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide

R E S E A R C H   A N D   D E V E L O P M E N T

EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide

Gary Norris, Rachelle Duvall
U.S. Environmental Protection Agency
National Exposure Research Laboratory
Research Triangle Park, NC 27711

Steve Brown, Song Bai
Sonoma Technology, Inc.
Petaluma, CA 94954

U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

Disclaimer

EPA through its Office of Research and Development funded and managed the research and development described here under contract 68-W-04-005 to Lockheed Martin and EP-D-09-097 to Sonoma Technology, Inc. The User Guide has been subjected to Agency review and is cleared for official distribution by the EPA. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This User Guide is for the EPA PMF 5.0 program and the disclaimer for the software is shown below.

The United States Environmental Protection Agency through its Office of Research and Development funded and collaborated in the research described here under Contract Number EP-D-09-097 to Sonoma Technology, Inc.

Portions of the code are Copyright ©2005-2014 ExoAnalytics Inc. and Copyright ©2007-2014 Bytescout.

Acknowledgments

The Multilinear Engine is the underlying program used to solve the PMF problem in EPA PMF and version me2gfP4_1345c4 has been developed by Pentti Paatero at the University of Helsinki and Shelly Eberly at Geometric Tools (http://www.geometrictools.com/). Shelly Eberly, Pentti Paatero, Ram Vedantham, Jeff Prouty, Jay Turner, and Teri Conner have contributed to the development of this and prior versions of EPA PMF. EPA would like to thank EPA PMF Peer Reviewers for their comments on the software and user guide, and for providing an improved list of PMF references.

Table of Contents

List of Figures

List of Tables

Acronyms

Acronym	Definition
AMS	Aerosol mass spectrometer
BOD5	Biological oxygen demand
BS	Bootstrap
BS-DISP	Bootstrap-Displacement
CI	Confidence interval
CMB	Chemical mass balance
DDP	Discrete difference percentiles
DISP	Displacement
EC	Elemental carbon
EDXRF	Energy dispersive X-ray fluorescence
GUI	Graphical user interface
MDL	Method detection limit
ME	Multilinear Engine
ME-2	Multilinear Engine version 2
Obs/Pred	Observed/Predicted
OC	Organic carbon
PAMS	Photochemical assessment monitoring stations
PCA	Principal component analysis
PM	Particulate matter
PMF	Positive Matrix Factorization
S/N	Signal-to-noise ratio
TNMOC	Total non-methane organic carbon
TSS	Total suspended solids
VOC	Volatile organic compound

1.   Introduction

1.1   Model Overview

Receptor models are mathematical approaches for quantifying the contribution of sources to samples based on the composition or fingerprints of the sources. The composition or speciation is determined using analytical methods appropriate for the media, and key species or combinations of species are needed to separate impacts. A speciated data set can be viewed as a data matrix X of i by j dimensions, in which i number of samples and j chemical species were measured, with uncertainties u. The goal of receptor models is to solve the chemical mass balance (CMB) between measured species concentrations and source profiles, as shown in Equation 1-1, with number of factors p, the species profile f of each source, and the amount of mass g contributed by each factor to each individual sample (see Equation 1-1):

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij} \tag{1-1}$$

where e_{ij} is the residual for each sample/species. The CMB equation can be solved using multiple models including EPA CMB, EPA Unmix, and EPA Positive Matrix Factorization (PMF).

PMF is a multivariate factor analysis tool that decomposes a matrix of speciated sample data into two matrices: factor contributions (G) and factor profiles (F). These factor profiles need to be interpreted by the user to identify the source types that may be contributing to the sample using measured source profile information, and emissions or discharge inventories. The method is reviewed briefly here and described in greater detail elsewhere (Paatero and Tapper, 1994; Paatero, 1997).

Results are obtained using the constraint that no sample can have significantly negative source contributions. PMF uses both sample concentration and user-provided uncertainty associated with the sample data to weight individual points. This feature allows analysts to account for the confidence in the measurement. For example, data below detection can be retained for use in the model, with the associated uncertainty adjusted so these data points have less influence on the solution than measurements above the detection limit.

Factor contributions and profiles are derived by the PMF model minimizing the objective function Q (Equation 1-2):

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj}}{u_{ij}} \right]^2 \tag{1-2}$$

Q is a critical parameter for PMF and two versions of Q are displayed for the model runs.

- Q(true) is the goodness-of-fit parameter calculated including all points.

- Q(robust) is the goodness-of-fit parameter calculated excluding points not fit by the model, defined as samples for which the uncertainty-scaled residual is greater than 4.

The difference between Q(true) and Q(robust) is a measure of the impact of data points with high scaled residuals. These data points may be associated with peak impacts from sources that are not consistently present during the sampling period. In addition, the uncertainties may be too high, which result in similar Q(true) and Q(robust) values because the residuals are scaled by the uncertainty.

EPA PMF requires multiple iterations of the underlying Multilinear Engine (ME) to help identify the most optimal factor contributions and profiles. This is due to the nature of the ME algorithm that starts the search for the factor profiles with a randomly generated factor profile. This factor profile is systematically modified using the gradient approach to chart the optimal path to the best-fit solution. In spatial terms, the model constructs a multidimensional space using the observations and then traverses the space using the gradient approach to reach its final destination of the best solution along this path. The best solution is typically identified by the lowest Q(robust) value along the path (i.e., the minimum Q) and may be imagined as the bottom of a trough in the multidimensional space. Due to the random nature of the starting point, which is determined by the seed value and the path it dictates, there is no guarantee that the gradient approach will always lead to the deepest point in the multidimensional space (global minimum); it may instead find a local minimum. To maximize the chance of reaching the global minimum, the model should be run 20 times developing a solution and 100 times for a final solution, each time with a different starting point.

Because Q(robust) is not influenced by points that are not fit by PMF, it is used as a critical parameter for choosing the optimal run from the multiple runs. In addition, the variability of Q(robust) provides an indication of whether the initial base run results have significant variability because of the random seed used to start the gradient algorithm in different locations. If the data provide a stable path to the minimum, the Q(robust) values will have little variation between the runs. In other cases, the combination of the starting point and the space defined by the data will impact the path to the minimum, resulting in varying Q(robust) values; the lowest Q(robust) value is used by default since it represents the most optimal solution. It should be noted that a small variation in Q-values does not necessarily indicate that the different runs have low variability between source compositions.

Variability due to chemical transformations or process changes can cause significant differences in factor profiles among PMF runs. Two diagnostics are provided to evaluate the differences between runs: intra-run residual analysis and a factor summary of the species distribution compared to those of the lowest Q(robust) run. The user must evaluate all of the error estimates in PMF to understand the stability of the model results; the algorithms and ME output are described in Paatero et al. (2014). Variability in the PMF solution can be estimated using three methods:

1. Bootstrap (BS) analysis is used to identify whether there are a small set of observations that can disproportionately influence the solution. BS error intervals include effects from random errors and partially include effects of rotational ambiguity. Rotational ambiguity is caused by the existence of infinite solutions that are similar in many ways to the solution generated by PMF. That is, for any pair of matrices, infinite variations of the pair can be generated by a simple rotation. With only one constraint of non-negative source contributions, it is impossible to restrict this space of rotations. BS errors are generally robust and are not influenced by the user-specified sample uncertainties.

2. Displacement (DISP) is an analysis method that helps the user understand the selected solution in finer detail, including its sensitivity to small changes. DISP error intervals include effects of rotational ambiguity but do not include effects of random errors in the data. Data uncertainty can directly impact DISP error estimates. Hence, intervals for downweighted species are likely to be large.

3. BS-DISP (a hybrid approach) error intervals include effects of random errors and rotational ambiguity. BS-DISP results are more robust than DISP results since the DISP phase of BS-DISP does not displace as strongly as DISP by itself.

These methods are applied with three air pollution data sets in Brown et al. (2014). The paper provides an interpretation of the EPA error estimates based on the applications. Paatero et al. (2014) and Brown et al. (2014) are key references for EPA PMF and both provide details on the error estimates and their interpretation, which are only briefly covered in this guide.
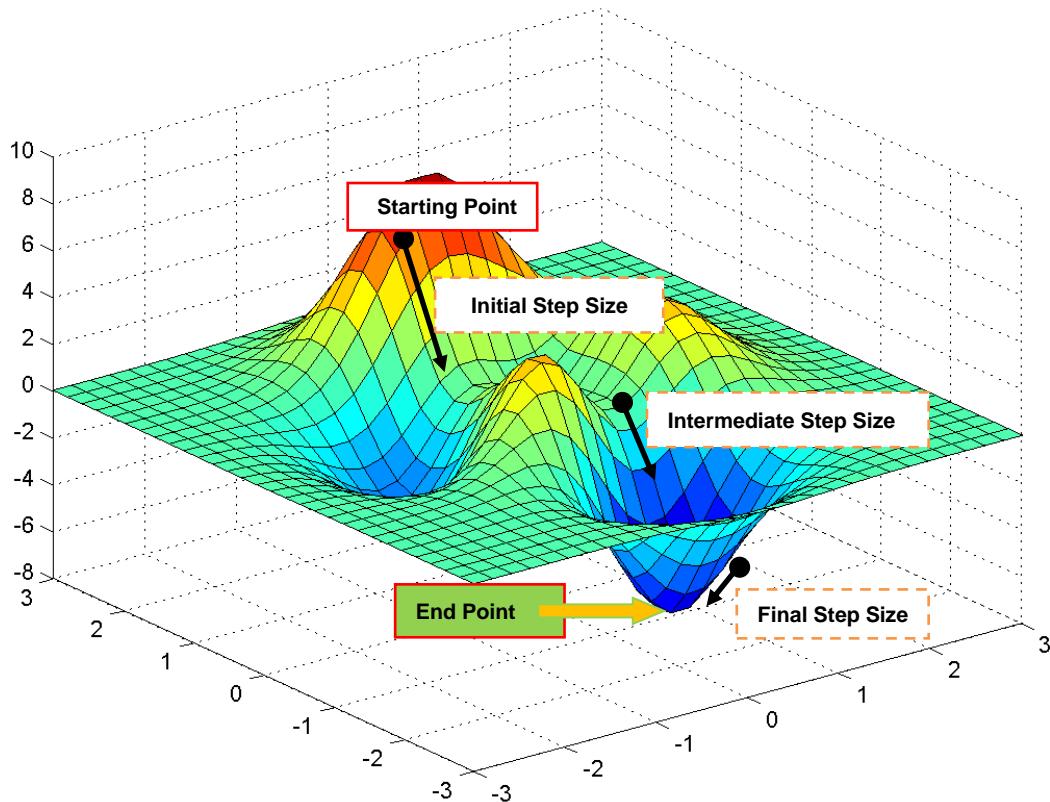
1.2   Multilinear Engine

Two common programs solve the PMF problem as described above. Originally, the program PMF2 (Paatero, 1997) was used. In PMF2, non-negativity constraints could be imposed on factor elements and measurements could be weighted individually based on uncertainties when determining the least squares fit. With these features, PMF2 was a significant improvement over previous principal component analysis (PCA) techniques for receptor modeling of environmental data. PMF2 was limited, however, in that it was designed to solve a very specific PMF problem. In the late 1990s, the ME, a more flexible program, was developed (Paatero, 1999). This program, currently in its second version and referred to as ME-2, includes many of the same features as PMF2 (for instance, the user is able to weight individual measurements and provide non-negativity constraints); however, unlike PMF2, ME-2 is structured so that it can be used to solve a variety of multilinear problems including bilinear, trilinear, and mixed models.

ME-2 was designed to solve the PMF problem by combining two separate steps. First, the user produces a table that defines the PMF model of interest. Then an automated secondary program reads the tabulated model parameters and computes the solution. When solving the PMF problem using EPA PMF, the first step is achieved via an input file that is produced by the EPA PMF user interface. Once the model has been specified, data and user specifications are fed into the secondary ME-2 program by EPA PMF. ME-2 solves the PMF equation iteratively, minimizing the sum-of-squares object function, Q, over a series of steps as shown in Figure 1. A stable solution has been reached when additional iterations to minimize Q provide diminishing returns. The search for the solution goes from coarser to a finer scale over three levels of iterations. The first level of iterations identifies the overall region of solution in space. In this

level, the change in Q (dQ) is required to be less than 0.1 over 20 consecutive steps in less than 800 steps. The second level identifies the neighborhood of the final solution. Here, dQ is required to be less than 0.005 over 50 consecutive steps in less than 2,000 total steps. The third level converges to the best possible Q-values (Paatero, 2000a) where dQ should be less than 0.0003 over 100 consecutive steps in less than 5,000 steps.

ME-2 typically requires a few hundred iterations for small data sets (less than 300 observations) and up to 2,000 for larger data sets (Paatero, 2000a). If a solution is not found that meets the requirements of any of the three levels, then a solution is non-convergent (Paatero, 2000a).

Figure 1. Conjugate Gradient Method – underpinnings of PMF solution search.

Output from ME-2 is read by EPA PMF and then formatted for the user to interpret. In addition, EPA PMF has three error estimate methods that are implemented through ME-2 and EPA PMF.

The differences between ME-2 and PMF2 model results have been examined in several studies through the application of each model to the same data set and comparison of the results. Overall, the studies showed similar results for the major components, but a greater uncertainty in the PMF2 solution (Ramadan et al., 2003) and better source separation using ME-2 (Kim et al., 2007). In two recent publications, the application of factor profile constraints by ME-2 resulted in a larger number of sources found (Amato et al., 2009; Amato and Hopke, 2012).

Version 5.0 of EPA PMF uses the most recent version of ME-2 and a PMF script file, which were developed by Pentti Paatero at the University of Helsinki and Shelly Eberly at Geometric Tools (March 3, 2014; me2gfP4_1345c4.exe and PMF_bs_6f8xx_sealed_GUI.ini).

1.3　Comparison to EPA PMF 3.0 and Other Methods

EPA PMF 5.0 has added two key components to EPA PMF 3.0: two additional error estimation methods and source contribution and profile constraints. Many other changes have been added to make the software easier to use, including the ability to read in multiple site data. The run time for the new error estimation methods can take from an hour to half a day depending on the number of factors and BS runs. The large amount of time is due to the high number of computations required for the robust error estimates. The PMF Model Development Quality Assurance Project Plan provides the details on the QA steps used to develop EPA PMF 5.0 and a number of interim versions between version 3.0 and 5.0. Version 4.2 was externally peer reviewed; the very useful comments were used to develop version 5.0 and improve the user guide.

Other comparable source apportionment models include Unmix and CMB. Although both models have aims similar to that of PMF, they have different mechanisms. Unmix identifies the "edges" in the data where the factor contribution from at least one factor is present only in negligible amounts. The edges are then used to determine the profile compositions and the number of sources in the data is provided. Unmix does not allow individual weighting of data points, as allowed by PMF. Although major factors resolved by PMF and Unmix are generally the same, Unmix does not always resolve as many factors as PMF (Pekney et al., 2006c; Poirot et al., 2001).

With CMB, the user must provide source profiles that the model uses to apportion mass. PMF and CMB have been compared in several studies. Rizzo and Scheff (2007a) compared the magnitude of source contributions resolved by each model and examined correlations between PMF- and CMB-resolved contributions. They found the major factors correlated well and were similar in magnitude; additionally, the PMF-resolved source profiles were generally similar to measured source profiles. In supplementary work, Rizzo and Scheff (2007b) used information from CMB PM source profiles to influence PMF results and used CMB results to help control rotations in PMF. Jaeckels et al. (2007) used organic molecular markers with elemental carbon (EC) and organic carbon (OC) in both CMB and PMF. Good correlations were found for most factors, with some biases present in a few of the factors. They also found an additional PMF factor that did not correspond to any CMB factors.

The models discussed above are complementary and, whenever possible, should be used along with PMF to make source apportionment results more robust. In addition, statistical receptor modeling methods have been developed by William F. Christensen at Brigham Young University and other researchers.

2.   Uses of PMF

PMF has been applied to a wide range of data, including 24-hr speciated $PM_{2.5}$, size-resolved aerosol, deposition, air toxics, high time resolution measurements such as those from aerosol mass spectrometers (AMS), and volatile organic compound (VOC) data. The References section (Section 9) provides numerous references where PMF has been applied. Additional discussion of uses of PMF is available in the Multivariate Receptor Modeling Workbook (Brown et al., 2007). Users are encouraged to read the papers that are relevant to their data as well as source profile measurement papers. The approaches used for PMF analyses have changed over the years as options such as constraints have been made available. Key references are summarized in Table 1.

Table 1. Summary of key references.

Reference	Key Points
Brinkman, G.; Vance, G.; Hannigan, M.P.; Milford, J.B. (2006). Use of synthetic data to evaluate positive matrix factorization as a source apportionment tool for $PM_{2.5}$ exposure data. Environ. Sci. Technol., 40(6): 1892-1901.	• Uses coefficient of determination (R^2) and normalized gross error (NGE) for the source contribution comparisons and the root mean squared error (RMSE) for source profile comparisons. • R^2 measures the fraction of the variance in the actual source contributions. • The NGE and RMSE are measures of the accuracy of the source contribution or profile estimate. • The RMSE was chosen for the profile comparisons to place the greatest weight on compounds present in the largest fractions, which are most important for source apportionment purposes, where total mass apportionment is the goal.
Chen, L.-W.A.; Lowenthal, D.H.; Watson, J.G.; Koracin, D.; Kumar, N.; Knipping, E.M.; Wheeler, N.; Craig, K.; Reid, S. (2010). Toward effective source apportionment using positive matrix factorization: Experiments with simulated $PM_{2.5}$ data. J. Air Waste Manage. Assoc., 60(1): 43-54.	• Uses a metric to measure the difference between known source profiles and PMF provided contributions. Uses a minimization technique to find the correct set of parameter values that helps closely match the true source profiles with predicted source profiles. • Not much on using the source profile uncertainties from the model output.

Reference	Key Points
Christensen, W.F.; Schauer, J.J. (2008). Impact of species uncertainty perturbation on the solution stability of positive matrix factorization of atmospheric particulate matter data. Environ. Sci. Technol., 42(16): 6015-6021.	• A perturbed uncertainty matrix is created by multiplying each original uncertainty value by a random multiplier generated from a log-normal distribution with a mean of 1 and a standard deviation (and CV) equal to 0.25, 0.50, or 0.75. The average values for the measure of relative error for the three scenarios are 8%, 14%, and 17%, respectively. • Relative errors associated with day-today estimates of source contributions can be more than double the size of the relative errors associated with estimates of average source contributions, with errors for four of 10 source contributions exceeding 30% for the largest-perturbation scenario. • The stability of source profile estimates in the simulation varies greatly between sources, with a mean correlation between perturbed gasoline exhaust profiles and the true profile equal to only 59% for the largest-perturbation scenario.
Hemann, J.G.; Brinkman, G.L.; Dutton, S.J.; Hannigan, M.P.; Milford, J.B.; Miller, S.L. (2009). Assessing positive matrix factorization model fit: a new method to estimate uncertainty and bias in factor contributions at the measurement time scale. Atmos. Chem. Phys., 9(2): 497-513.	• A novel method was developed to estimate model fit uncertainty and bias at the daily time scale, as related to factor contributions. A circular block BS is used to create replicate data sets, with the same receptor model then fit to the data. • Neural networks are trained to classify factors based upon chemical profiles, as opposed to correlating contribution time series, and this classification is used to align factor orderings across the model results associated with the replicate data sets. • The results indicate that variability in factor contribution estimates does not necessarily encompass model error: contribution estimates can have small associated variability across results yet also be very biased.
Henry, R.C.; Christensen, E.R. (2010). Selecting an appropriate multivariate source apportionment model result. Environ. Sci. Technol., 44(7): 2474-2481.	• Source apportionment results favor Unmix when edges in the data are well-defined and PMF when several zeros are present in the loading and score matrices. Because both models are seen to have potential weaknesses, both should be applied in all cases. • Recommend that the EPA approved versions of PMF and Unmix both be applied to environmental data sets. If the two produce very similar results, then one has added confidence based on the fact that two independent methods of analysis support each other. If the PMF and Unmix results are different, then examine the estimated source compositions: if these have many zeros the PMF result should be preferred, but only if the Unmix diagnostic edges plots show that one or more of the edges are not clearly defined by the data.

Reference	Key Points
Kim, E.; Hopke, P.K. (2007a). Comparison between sample-species specific uncertainties and estimated uncertainties for the source apportionment of the speciation trends network data. Atmos. Environ., 41(3): 567-575.	• The objective of this study is to compare the use of the estimated fractional uncertainties (EFU) for the source apportionment of $PM_{2.5}$ (particulate matter less than 2.5 μm in aerodynamic diameter) measured at the speciated trends network (STN) monitoring sites with the results obtained using SSU (standard STN uncertainties). Thus, the source apportionment of STN $PM_{2.5}$ data were performed and their contributions were estimated through the application of PMF for two selected STN sites, Elizabeth, NJ and Baltimore, MD with both SSU and EFU for the elements measured by X-ray fluorescence. The PMF resolved factor profiles and contributions using EFU were similar to those using SSU at both monitoring sites. The comparisons of normalized concentrations indicated that the STN SSU were not well estimated. This study supports the use of EFU for the STN samples to provide useful error structure for the source apportionment studies of the STN data. • Implies a flaw with uncertainties associated with STN data. Promotes EFU over SSN.
Latella, A.; Stani, G.; Cobelli, L.; Duane, M.; Junninen, H.; Astorga, C.; Larsen, B.R. (2005). Semicontinuous GC analysis and receptor modelling for source apportionment of ozone precursor hydrocarbons in Bresso, Milan, 2003. J. Chromatogr. A, 1071(1-2): 29-39.	• A new approach is presented, by which the input uncertainty is allowed to float as a function of the photochemical reactivity of the atmosphere and the stability of each individual compound.
Lowenthal, D.H.; Rahn, K.A. (1988). Tests of regional elemental tracers of pollution aerosols. 2. Sensitivity of signatures and apportionments to variations in operating parameters. Atmos. Environ., 22: 420-426.	• Straight forward use of PMF and Unmix along with HYSPLIT to confirm results using synthetic data.

Reference	Key Points
Miller, S.L.; Anderson, M.J.; Daly, E.P.; Milford, J.B. (2002). Source apportionment of exposures to volatile organic compounds. I. Evaluation of receptor models using simulated exposure data. Atmos. Environ., 36(22): 3629-3641.	• Four receptor-oriented source apportionment models were evaluated by applying them to simulated personal exposure data for select VOCs that were generated by Monte Carlo sampling from known source contributions and profiles. The exposure sources modeled are environmental tobacco smoke, paint emissions, cleaning and/or pesticide products, gasoline vapors, automobile exhaust, and wastewater treatment plant emissions. The receptor models analyzed are CMB, PCA/absolute principal component scores, PMF, and graphical ratio analysis for composition estimates/source apportionment by factors with explicit restriction, incorporated in the UNMIX model. • All models identified only the major contributors to total exposure concentrations. PMF extracted factor profiles that most closely represented the major sources used to generate the simulated data. • None of the models were able to distinguish between sources with similar chemical profiles. Sources that contributed 5% to the average total VOC exposure were not identified.
Reff, A.; Eberly, S.I.; Bhave, P.V. (2007). Receptor modeling of ambient particulate matter data using positive matrix factorization: Review of existing methods. J. Air Waste Manage. Assoc., 57(2): 146-154.	• Guidance for the application and use of PMF.
Shi, G.L.; Li, X.; Feng, Y.C.; Wang, Y.Q.; Wu, J.H.; Li, J.; Zhu, T. (2009). Combined source apportionment, using positive matrix factorization-chemical mass balance and principal component analysis/multiple linear regression-chemical mass balance models. Atmos. Environ., 43(18): 2929-2937.	• A straightforward application of PMF and PCA/MLR-CMB that deals with collinear sources and other real data issues.
Yuan, B., Min Shao, M.; Gouw, J.; David D. Parrish, D.; Lu, S.; Wang, M.; Zeng, L.; Zhang, Q.; Song, Y.; Zhang, J.; Hu, M, (2012). Volatile organic compounds (VOCs) in urban air: How chemistry affects the interpretation of positive matrix factorization (PMF) analysis, J. Geophys. Res., 117	• Impact of VOC atmospheric reactivity on PMF results. (VOCs) were measured online at an urban site in Beijing in August–September 2010.

Reference	Key Points
Zhang, Y.X.; Sheesley, R.J.; Bae, M.S.; Schauer, J.J. (2009). Sensitivity of a molecular marker based positive matrix factorization model to the number of receptor observations. Atmos. Environ., 43(32): 4951-4958.	• Impact of the number of observations on molecular marker-based positive matrix factorization (MM-PMF) source apportionment models, daily $PM_{2.5}$ samples were collected in East St. Louis, IL, from April 2002 through May 2003.

PMF requires a data set consisting of a suite of parameters measured across multiple samples. For example, PMF is often used on speciated $PM_{2.5}$ data sets with 10 to 20 species over 100 samples. An uncertainty data set, that assigns an uncertainty value to each species and sample, is also needed. The uncertainty data set is calculated using propagated uncertainties or other available information such as collocated sampling precision.

3.   Installing EPA PMF 5.0

EPA PMF 5.0 can be obtained from EPA by e-mailing *NERL_RM_Support@epa.gov*. To install the program, run EPA PMF 5.0 Setup.exe and follow the installation directions on the screen. The installation program creates an EPA PMF subfolder in the Program Files folder for the software and an EPA PMF subfolder in the Documents folder for data files. Installation problems and software error messages should be reported to Gary Norris at *RM_Support@epa.gov*.

EPA PMF 5.0 can be run on a personal computer using the Windows XP or Windows 7 operating system or higher. Users will need to have permission to write to the computer's C:\ drive in order to install and run EPA PMF; this may not be the default setting for some users. After installation, EPA PMF can be started by double clicking EPA PMF 5.0 icon on the desktop.

4.   Global Features

The user can access the following features throughout EPA PMF 5.0:

- **Sorting data.** Columns in tables can be sorted by left-clicking the mouse button on a column heading. Clicking once will sort the items in ascending order and clicking twice will sort the items in descending order. If a column has been sorted, an arrow will appear in the header indicating the direction in which it is sorted.

- **Saving graphics.** All graphical output can be saved in a variety of formats by right-clicking on an image. Available formats are .gif, .bmp, .png, and .tiff. In the same menu, the user can choose to copy or print a graphic. A stacked graph option is also available to combine profiles or time series on one page. When "copy" is selected, the graphic is copied to the clipboard. When "print" is selected, the graphic will automatically be sent to the local machine's default printer. When saving a graphic, a dialog box appears so that the user can change the file path and file name of the output file.

- **Undocking graphs.** Any graph can be opened in a new window by right-clicking on the graph and selecting Floating Window. The user can open as many windows as required. However, the graphs in the floating windows do not update when model parameters and output are changed.

- **Resizing sections within tabs.** Many tabs have multiple sections separated by a gray line (Figure 2; red arrows point to the gray bars that enable the user to adjust height and width). These sections can be resized by clicking on the gray line and dragging it to the desired location.

- **Indicating selected data points.** When the user moves the cursor over a point on a scatter plot or time series graph, the point is outlined with a dashed-line square, indicating the point to which the information in the status bar refers.

- **Using arrow keys on lists and tables.** After selecting (by clicking on or tabbing to) a list or table, the keyboard arrow keys can be used to change the selected row.

- **Accessing help files.** The left bottom corner of most screens has a "Help" shortcut that provides users access to a help file associated with the main functions in the current screen.

- **Using the status bar.** Most screens have a status bar across the bottom of the window that provides additional information to the user. This information changes based on the tab selected. Individual tab details are discussed in subsequent sections of this guide. An example of the status bar on the Concentration Scatter Plot screen is shown at the bottom of Figure 2.
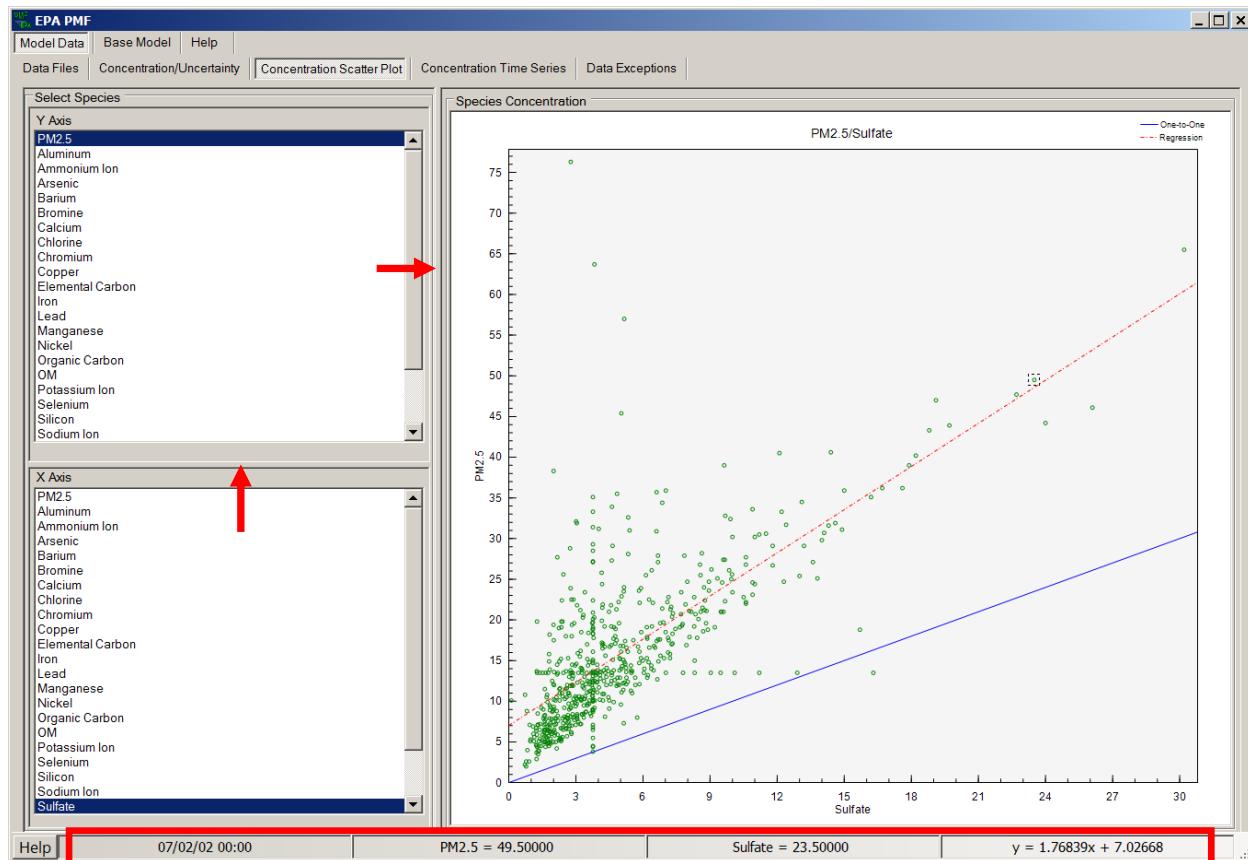
Figure 2. Example of resizable sections and status bar.

5.   Getting Started

Each time the EPA PMF 5.0 program is started, a splash screen with information about the development of the software and various copyrights is displayed. The user must click the OK button or press the spacebar or Enter key to continue.
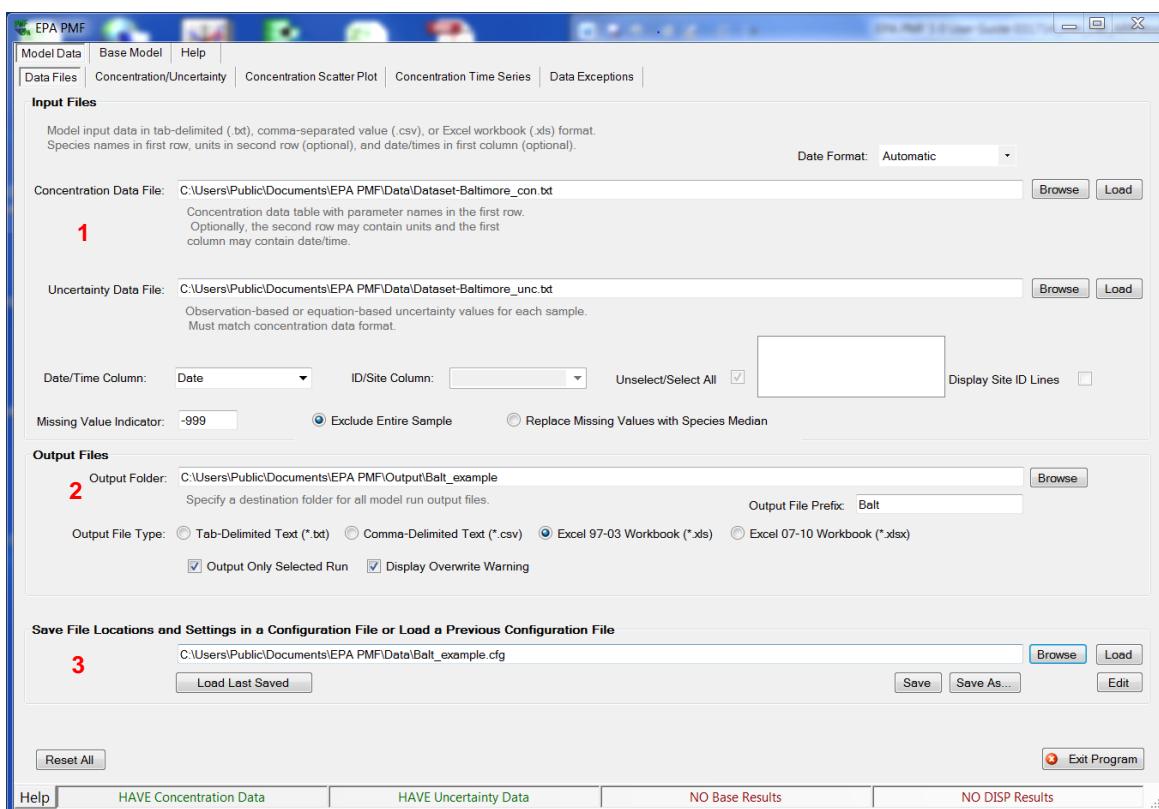
The first EPA PMF window is Data Files under the Model Data tab, as shown in Figure 3. On this screen, the user can provide file location information and make required choices that will be used in running the model. This screen has three sections: Input Files (Figure 3, 1), Output Files (Figure 3, 2), and Configuration File (Figure 3, 3), each of which is described in detail below. EPA PMF 5.0 can read multiple site data; time series plots of species concentrations or source contributions are displayed in the same order as the user provided data and PMF displays a vertical line separating the sites.

The status bar at the bottom of the Data Files screen indicates which section of the program has been completed. Prior to any user input on the Data Files screen, the status bar displays "NO Concentration Data, NO Uncertainty Data, NO Base Results, NO Bootstrap Results, NO BS-DISP Results, and NO DISP Results" in red. When a task is completed, "NO" is replaced with "Have" and the text color changes to green. In the Figure 3 example, concentration and uncertainty files have been provided to the program, so the first two items on the status bar are green. Base runs, BS runs, BS-DISP runs, and DISP runs have not been completed, so the last four items are red. The Baltimore PM files (Dataset_Baltimore_con.txt and Dataset_Baltimore_unc.txt) are part of the installation package and can be found in the "C:\Documents\EPA PMF\Data" folder, if the user installed the model using the default installation settings.

5.1   Input Files

Two input files are required by PMF: (1) sample species concentration values and (2) sample species uncertainty values or parameters for calculating uncertainty. EPA PMF accepts tab-delimited (.txt), comma-separated value (.csv), and Excel Workbook (.xls or .xlsx) files. Each file can be loaded either by typing the path into the "data file" input boxes or browsing to the appropriate file. If the file includes more than one worksheet or named range, the user will be asked to select the one they want to use. The concentration file has the species as columns and dates or sample numbers as rows, with headers for each (Figure 4). All standard date and time conventions are accepted and they are listed in the Date Format pull-down list. Four possible input options are accepted: (1) with sample ID only, (2) with Date/Time only, (3) with both Sample ID and Date/Time, (4) with no IDs or Date/Time. Units can be included as a second heading row in the concentration file, but are not required and units are not included in the uncertainty file. If units are supplied by the user, they will be used by the graphical user interface (GUI) for axis labels only and will not be used by the model. Blank cells are not accepted; the user will be prompted to examine the data and try again; species names cannot contain commas. If values less than -999 are found in the data set, the program will give a warning message but will continue. If these values are not real or are missing value indicators, the user should modify the data file outside the program and reload the data sets. Also, the names of each species must be unique. The user must specify the Date/Time and ID/Site

columns if they are included in the input data sets. The basic PMF functions are demonstrated using single site data and a multiple site example is shown in Section 8.1. Multiple site data should be sorted by Site and Date/Time before loading it into PMF. Lines deliminating Sample ID will not be displayed if a missing value is at the transition between Sample IDs and the option "exclude missing samples" is selected; missing transition samples should be removed or the option "replace missing samples with the species median" selected.

Figure 3. Example of the Input Files screen.

Sample species uncertainties should encompass errors such as sampling and analytical errors. For some data sets, the analytical laboratory or reporting agency provides an uncertainty estimate for each value. However, uncertainties are not always reported and, when they are not available, errors must be estimated by the user. A discussion of calculating uncertainties is provided in Reff et al. (2007).

EPA PMF 5.0 accepts two types of uncertainty files: observation-based and equation-based. The observation-based uncertainty file provides an estimate of the uncertainty for each species in a sample. It should have the same dimensions as the concentration file and the first column will still be a date, date time or sample number; however, the uncertainty file should not include units. If the concentration file contains a row of units, the uncertainty file will have one less row than the concentration file. The user will be notified if the column and row headers do not match, but the program will continue. In addition, the program will check to see if the dates or

sample numbers are the same between the concentration and uncertainty files and the program will not allow the data to be evaluated if there is a mismatch. If the headers are different due to naming conventions but actually have the same order, the user can proceed to the next step. If not, the user should correct the problem outside the GUI and reload the files. Negative values and zero are not permitted as uncertainties; EPA PMF will provide an error message and the user will have to remove these values outside EPA PMF and reload the uncertainty file.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Aluminum	Ammonium	Bromine	Calcium	Chlorine	Copper	EC	Iron	Lead	Manganes	Nickel	Nitrate	OC
2	DATE	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3
3	2/9/2000	0.0201	3.6020	0.0107	0.0676	0.0647	0.0059	3.1230	0.1497	0.0157	0.0043	0.0577	5.3700	7.3930
4	2/15/2000	0.0057	1.3740	0.0006	0.0325	0.0016	0.0019	1.0710	0.0673	0.0055	0.0004	0.0285	0.8785	3.3310
5	2/27/2000	0.0029	2.1860	0.0028	0.0422	0.0288	0.0028	0.6732	0.0727	0.0073	0.0002	0.0215	3.8820	5.2030
6	3/4/2000	0.0011	0.4501	0.0014	0.0329	0.0024	0.0010	0.5503	0.0483	0.0061	0.0004	0.0188	0.4562	3.6160
7	3/10/2000	0.0075	0.3099	0.0006	0.0247	0.0039	0.0003	0.2869	0.0565	0.0032	0.0016	0.0083	0.6763	2.8140
8	3/22/2000	0.0006	1.1570	0.0033	0.0265	0.0015	0.0029	0.9487	0.0821	0.0044	0.0012	0.0107	1.0670	2.4150
9	4/6/2000	0.0256	1.3520	0.0025	0.0863	0.0026	0.0041	2.1990	0.1492	0.0089	0.0034	0.0254	1.4660	4.7350
10	4/9/2000	0.0165	0.2800	0.0011	0.0263	0.0016	0.0003	0.8535	0.0396	0.0017	0.0019	0.0257	0.2515	1.6760
11	4/12/2000	0.0108	1.1290	0.0026	0.0304	0.0080	0.0046	0.9983	0.0959	0.0042	0.0001	0.0344	1.1900	2.6360
12	4/15/2000	0.0065	1.5640	0.0037	0.1075	0.0296	0.0059	3.1430	0.1976	0.0110	0.0026	0.0437	4.3040	6.9460
13	4/18/2000	0.0072	0.1983	0.0028	0.0351	0.0073	0.0017	0.6603	0.0539	0.0004	0.0027	0.0082	0.6816	1.9990
14	4/21/2000	0.0092	0.1432	0.0022	0.0250	0.0042	0.0023	0.7096	0.0765	0.0003	0.0009	0.0126	0.6017	1.7230
15	4/24/2000	0.0289	0.4066	0.0000	0.0337	0.0007	0.0006	1.1100	0.0830	0.0067	0.0005	0.0256	0.2174	2.4420
16	4/27/2000	0.0033	1.5030	0.0031	0.0329	0.0010	0.0024	1.4970	0.0840	0.0082	0.0013	0.0247	3.3670	3.5360
17	4/30/2000	0.0120	0.5734	0.0021	0.0442	0.0097	0.0022	0.6726	0.0741	0.0025	0.0041	0.0153	0.5117	3.3610
18	5/3/2000	0.0098	1.3200	0.0014	0.0365	0.0039	0.0015	1.1210	0.0735	0.0077	0.0000	0.0056	1.3380	4.2670
19	5/12/2000	0.0209	0.1049	0.0013	0.0394	0.0003	0.0033	1.2070	0.1108	0.0046	0.0000	0.0114	0.6438	3.8460
20	5/15/2000	0.0096	1.1600	0.0010	0.0337	0.0023	0.0002	0.8730	0.0902	0.0064	0.0004	0.0167	0.3547	3.1960
21	5/18/2000	0.0348	2.9630	0.0037	0.1088	0.0083	0.0066	1.9910	0.1519	0.0054	0.0031	0.0166	3.3450	6.1610
22	5/21/2000	0.0008	1.9910	0.0014	0.0409	0.0011	0.0025	0.4828	0.0449	0.0038	0.0018	0.0099	2.0890	2.5760
23	5/24/2000	0.0057	1.8440	0.0010	0.0386	0.0013	0.0005	1.4190	0.0957	0.0045	0.0007	0.0553	1.3390	4.2690

Figure 4. Example of formatting of the Input Concentration file.

The equation-based uncertainty file provides species-specific parameters that EPA PMF 5.0 uses to calculate uncertainties for each sample. This file should have one delimited row of species, with species names (Figure 5). The next row should be species-specific method detection limit (MDL) followed by the row of uncertainty (species-specific). Zeroes and negatives are not permitted for either the detection limit or the percent uncertainty. If the concentration is less than or equal to the MDL provided, the uncertainty (*Unc*) is calculated using a fixed fraction of the MDL (Equation 5-1; Polissar et al., 1998).

	A	B	C	D	E	F	G	H
1	unc	Aluminum	Ammoniur	Arsenic	Barium	Bromine	Calcium	Chlorine
2		2	0.00419	0.0125	0.00098	0.0068	0.0016	0.0038	0.002635
3		10	10	10	10	10	10	10
4								

Figure 5. Example of an equation-based uncertainty file.

$$Unc = \frac{5}{6} \times MDL \qquad\qquad (5\text{-}1)$$

If the concentration is greater than the MDL provided, the calculation is based on a user provided fraction of the concentration and MDL (Equation 5-2).

$$Unc = \sqrt{\left(Error\ Fraction \times concentration\right)^2 + \left(0.5 \times MDL\right)^2} \qquad \textbf{(5-2)}$$

A sample equation-based uncertainty file (Dataset-Baltimore_unc_eqn) has been provided in the C:\Documents\EPA PMF\Data folder. The equation-based uncertainty is useful if only the MDL and error percent are available; however, this approach will not capture errors associated with the specific samples. The uncertainties calculated by the equation-based method do not match the Dataset_Baltimore_unc.txt due to this simplification.

Users can specify a Missing Value Indicator (which can be any numeric value) in the Input Files box on the Data Files screen. The user should not choose a numeric indicator that could potentially be a real concentration. For example, if the user specifies "-999" as the missing value indicator, and chooses to replace the species with the median, the program will find all instances of "-999" in the data file and replace them with the species-specific median. The program will also replace all associated uncertainty values with a high uncertainty of four times the species-specific median. If all samples of a species are missing, that species is automatically categorized as "bad" and excluded from further analysis. The missing value indicator is used in the output files.

If a message is displayed that the dates/times do not match in the concentration and uncertainty files, the user needs to check the file dates/times and reload the data before being able to evaluate the data in PMF. If the dates/times in both files are the same, try saving both the concentration and data file in a different format, such as .csv or .txt.

5.2   Output Files

The user can specify the output directory ("Output Folder"), choose the EPA PMF output file types ("Output File Type" radio buttons) and define a prefix for output files ("Output File Prefix"). The prefix is added to the beginning of each file; for the example in Figure 3, the profiles will be saved as Balt_profile.xls. For the examples in the User Guide, the prefix is shown as an asterisk (*). The "Output File Type" includes tab-delimited text (.txt), comma-separated variable (.csv), or Excel Workbook (.xls). "Output File Prefix" is the prefix that will be used as the first part of any output file; this prefix can contain any letters and/or numbers (other characters such as "-" and "_" are not allowed). If this prefix is not changed when a new run is initiated, a warning will be displayed. If Excel Workbook output is selected, two output files are automatically created by EPA PMF during base runs and will be saved in the My Documents\EPA PMF\Output folder selected by the user: *_base.xls and *_diagnostics.xls. Each file has tabs with the PMF results.

- ***_base.xls** – Profiles, Contributions, Residual, Run Comparison
- ***_diagnostics.xls** – Summary, Input, Base Runs

If a delimited output is selected, the information in the Base Runs tab is provided as separate files and the diagnostics tab information is combined into one file. The following list provides the details on the data that are saved in the Excel output files.

Additional files are created and saved after conducting bootstrapping: (*_profile_boot), DISP (*_DISPres1, *_DISPres2, *_DISPres3, *_DISPres4), BS-DISP (*_BSDISP1, *_BSDISP2, *_BSDISP3, *_BSDISP4), Fpeak (*_fpeak), and/or constrained model runs (*_Constrained). The four files output for DISP and BS-DISP are for each dQmax; the runs using the lowest dQmax are used in the summary graphics and in the summary output file. The file *_ErrorEstimationSummary provides a summary of the base run and the error estimations that have been done using BS, DISP, and BS-DISP. The file *_profile_boot contains the number of BS runs mapped to each base run, each BS profile that was mapped to the base profile, and all bootstrapping statistics generated by the GUI. The file *_fpeak contains the profiles and contributions of each Fpeak run. When multiple base model runs are completed, by default, only the run with the lowest Q(robust) value is saved to the output, but the user may opt to include all runs in the output by unselecting "Output Only Selected Run."

5.3   Configuration Files
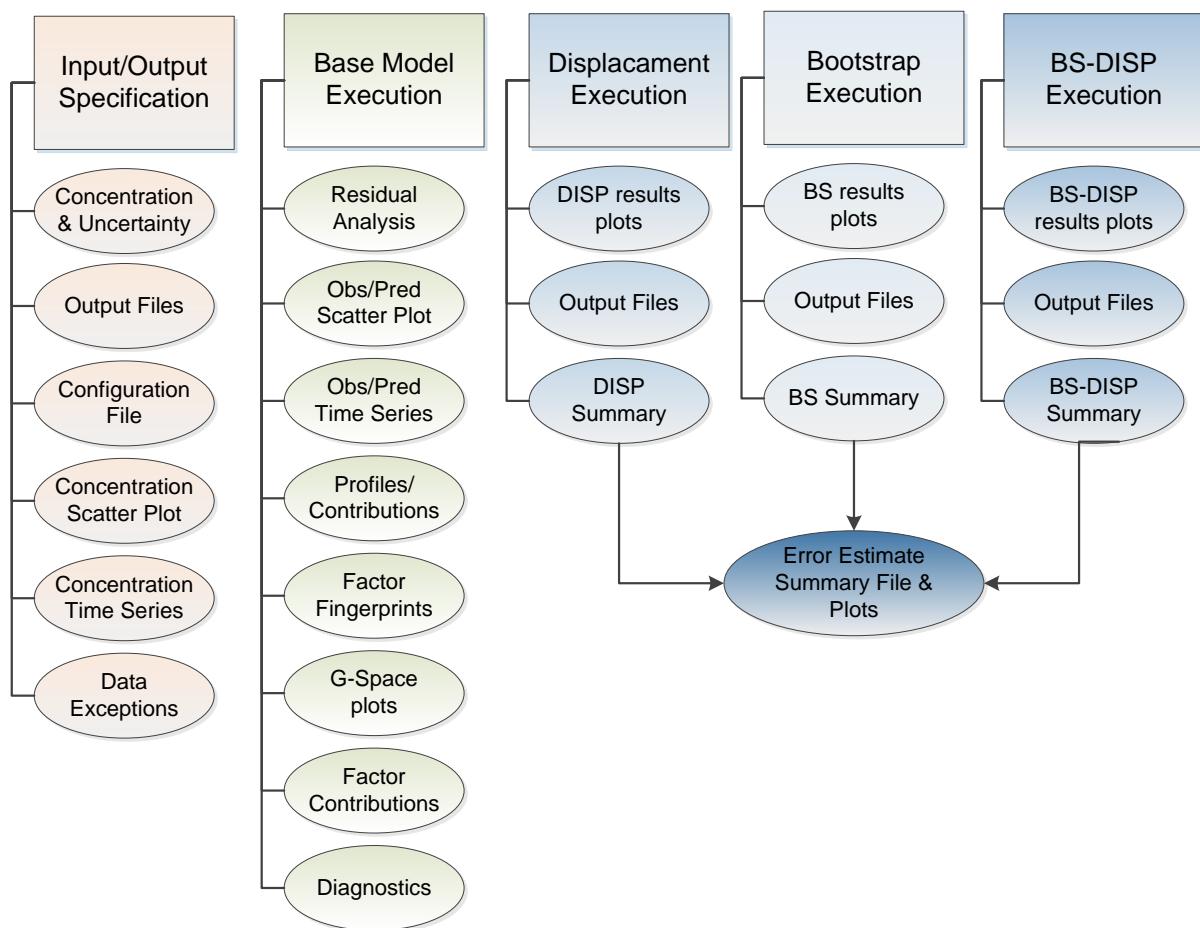
EPA PMF provides the option of saving run preferences and input parameters in a configuration file. The user must provide a name for a configuration file on the Input File Screen to create a configuration file. Information saved in the configuration file include specifications from the Data Files screen (e.g., input files, output file location, and output file type), species categorizations from the Concentration/Uncertainty screen, and all run specifications from the Base Model Runs screen, Fpeak Rotation screen, and Constrained Model Runs screen. Model output is not saved as part of the configuration file; however, the model random starting point or seed number is saved if the Random Start button is unchecked. To choose a configuration file, the user can click on "Browse" to browse to the correct path or type in a path and name. The user can also press the "Load Last" button or simply press "Enter" on the keyboard to load the most recently used configuration file. The "Save" and "Save As" buttons can be used to save the current settings to an existing or new configuration file.

Configuration files can be used on multiple computers or shared with collaborators, thereby avoiding a long list of preferences to replicate the results. Use the "Browse" button to locate and load the configuration file. The location of both the concentration and uncertainty files must be identified next. PMF does not store past run data; however, the results can be easily calculated by PMF as long as the same number of factors, runs, and a fixed seed is used (random start is not selected).

5.4   Suggested Order of Operations

The GUI is designed to give the user as much flexibility as possible when running the PMF model. However, certain steps must be completed to utilize the full potential of the provided tools. The order of operations is mainly based on how the tabs and functions are arranged (from left to right) in the program (Figure 6, Figure 7, and Figure 8); the sections in this user guide also follow this order. To begin using the program, the user must provide input files via

the Model Data - Data Files screen before other operations are available. The first time PMF is performed on the data set, the user should analyze the input data via the Concentration/Uncertainty, Concentration Scatter Plot, Concentration Time Series, and Data Exceptions screens. This step is usually followed by Base Model Runs and Base Model Results under the Base Model tab; these steps should be repeated as needed until the user reaches a reasonable solution. The solution is evaluated using the Error Estimation options starting with DISP and progressing to BS and BS-DISP; the output from the error estimation methods (DISP, BS, and BS-DISP) provides key information on the stability of the solution. All three error estimation methods are required to understand the uncertainty associated with the solution. Advanced users may wish to initiate Fpeak runs or constrained model runs based on a selected base run; both options are available under the Rotational Tools tab.
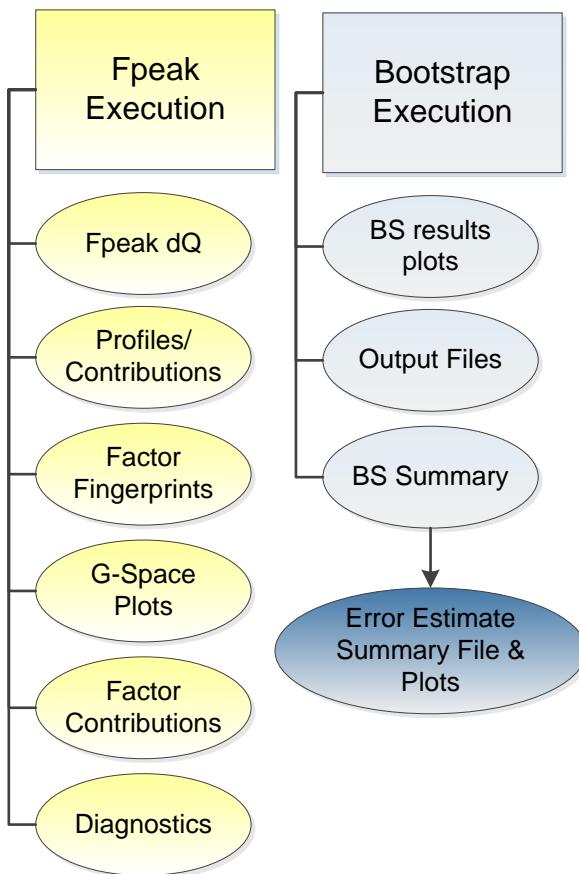
Figure 6. Flow chart of operations within EPA PMF – Base Model.

5.5　Analyze Input Data

Several tools are available to help the user analyze the concentration and uncertainty data before running the model. These tools help the user decide whether certain species should be excluded or downweighted (e.g., due to increased uncertainty or a low signal-to-noise ratio), or

if certain samples should be excluded (e.g., due to an outlier event). All changes and deletions should be reported with the final solution. The four screens for analyzing input data are described below.
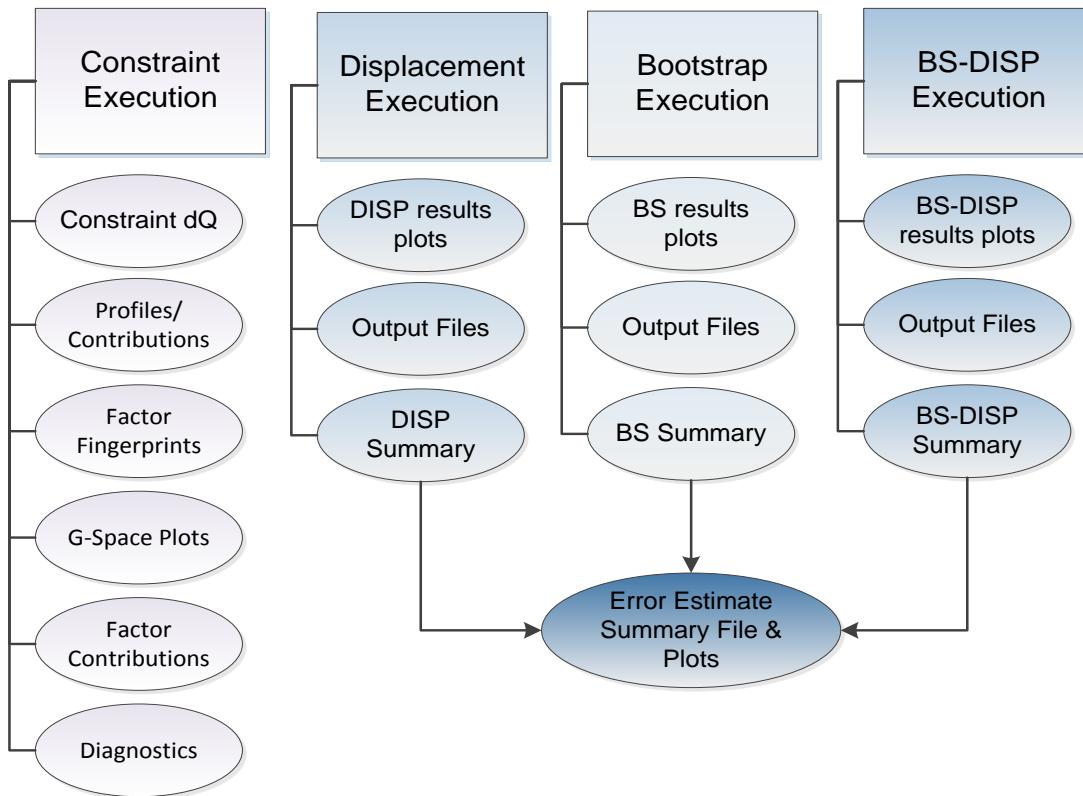
Figure 7. Flow chart of operations within EPA PMF – Fpeak.

5.5.1 Concentration/Uncertainty

Input data statistics and concentration/uncertainty scatter plots are presented in the Concentration/Uncertainty screen, as shown in Figure 9. The following statistics are calculated for each species and displayed in a table on the left of the screen (Figure 9, 1):

- Minimum (Min) – minimum concentration value
- 25th percentile (25th)
- Median – 50th percentile (50th)
- 75th percentile (75th)
- Maximum (Max) – maximum value reported
- Signal-to-noise ratio (S/N) – indicates whether the variability in the measurements is real or within the noise of the data

Figure 8. Flow chart of operations within EPA PMF – Constraints.
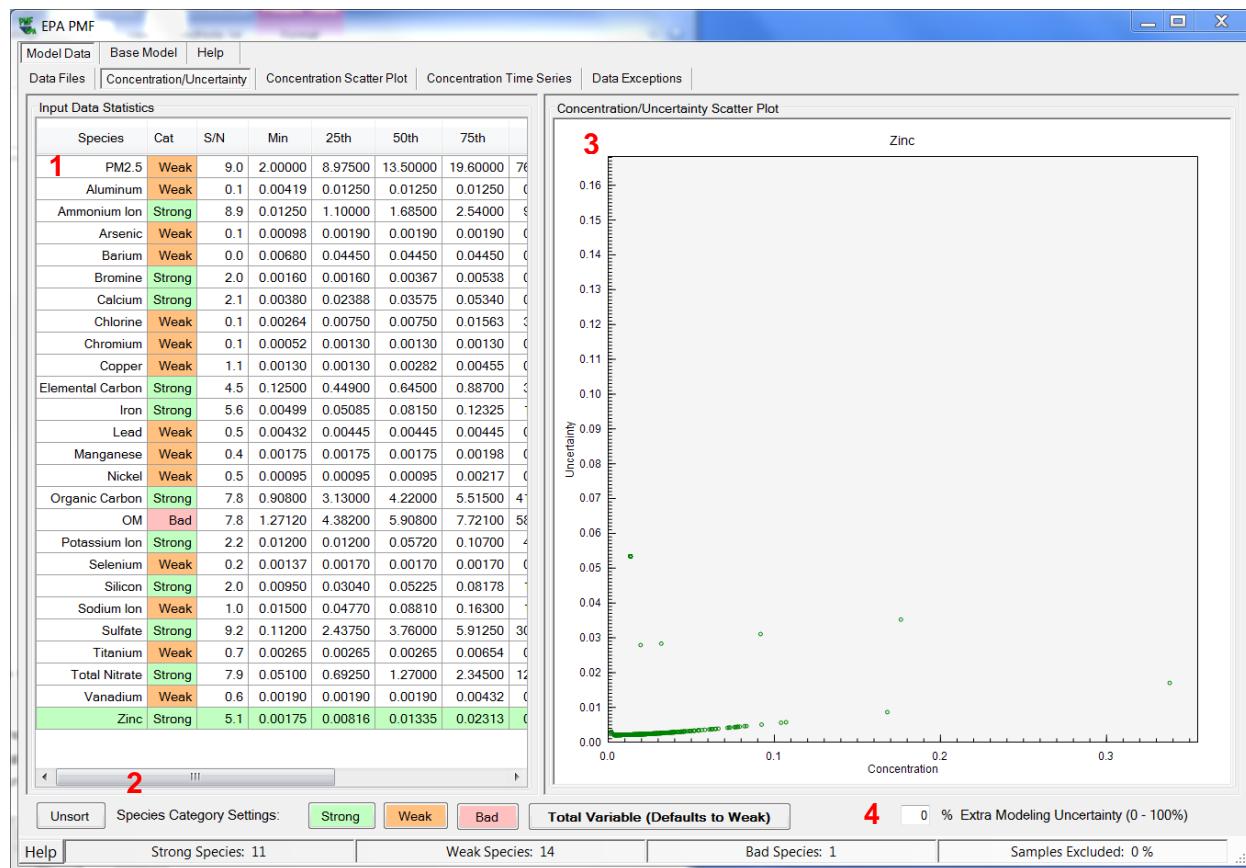
Percentiles are calculated using a weighted average approach (Equation 5-2):

$$L(n,p) = \frac{(n+1)p}{100}$$ **(5-2)**

$$L(n,p) = I + F$$

$$W_1 = 1 - F; \; W_2 = F; \; W_3 = 0$$

$$P = W_1 X_I + W_2 X_{(I+1)} + W_3 X_{(I+2)}$$

where *n* represents the number of non-missing values of the selected variable; *p* is the percentile of interest; *I* is the integer part of *L(n,p)*; *F* represents the fractional part of *L(n,p)*; W_1, W_2, and W_3 are weights; *P* is the pth percentile; and $X_1, X_2, \ldots, X_n$ represent the ordered values of the variable of interest.

The S/N calculation in EPA PMF has been revised in the new version. Previously, S/N of a given species was essentially the sum of the concentration values divided by the sum of uncertainty values. While reasonable, this could lead to different problems in certain specific situations. Artificially high S/N values would be obtained for species with a handful of high

concentration events, resulting in a S/N that may actually be higher than another species' S/N with more consistent signal. More seriously, artificially low S/N values could appear for species with a few missing values. Missing values are usually downweighted by very large uncertainty values, typically (much) larger than the largest concentration values in the species in question.

Figure 9. Example of the Concentration/Uncertainty screen.

If this process was done to the data prior to ingest into EPA PMF, such inflated uncertainty values will inflate the N in S/N calculations, resulting in a S/N that will be small enough to cause the classification of a perfectly strong variable as "weak." The latter problem has been repeatedly observed in practical work. In addition, the presence of slightly negative concentration values, not uncommon in environmental data, could artificially decrease S and hence the S/N of a species.

In the revised calculation, only concentration values that exceed the uncertainty contribute to the signal portion of the S/N calculation, because the concentration value is essentially equal to the sum of signal and noise, and therefore signal is the difference between concentration and uncertainty.

Two calculations are performed to determine S/N, where concentrations below uncertainty are determined to have no signal, and for concentrations above uncertainty, the difference between concentration (x_i) and uncertainty (s_i) is used as the signal (Equation 5-3):

$$d_{ij} = \left(\frac{x_{ij} - s_{ij}}{s_{ij}} \right) \ \text{if} \ x_{ij} > s_{ij}$$

$$d_{ij} = 0 \qquad \text{if} \ x_{ij} \leq s_{ij} \tag{5-3}$$

S/N is then calculated using Equation 5-4:

$$\left(\frac{S}{N} \right)_j = \frac{1}{n} \sum_{i=1}^{n} d_{ij} \tag{5-4}$$

The result with this new S/N calculation is that species with concentrations always below their uncertainty have a S/N of 0. Species with concentrations that are twice the uncertainty value have a S/N of 1. S/N greater than 1 may often indicate a species with "good" signal, though this depends on how uncertainties were determined. Negative concentration values do not contribute to the S/N, and species with a handful of high concentration events will not have artificially high S/N. While there are many methods to determine S/N, the one selected in the new version of EPA PMF may be more useful in environmental data analysis compared to the prior version, though with the caveat that the S/N is merely one of many analyses for screening data.

Based on these statistics and knowledge of analytical and sampling issues, the user can categorize a species as "Strong," "Weak," or "Bad" by selecting the species in the Input Data Statistics table (Figure 9, 1) and pressing the appropriate button under the table (Figure 9, 2). In addition, Alt+W, Alt+B, and Alt+G can be used to change a species category to Weak, Bad, or Good, respectively. The default value for all species is "Strong." A categorization of "Weak" triples the provided uncertainty, and a categorization of "Bad" excludes the species from the rest of the analysis. If a species is marked "Weak," the row is highlighted orange; if a species is marked "Bad," the row is highlighted pink. When choosing the category for each species, the user should consider the presence of sources that could be contributing to species based on measured profiles, tracer species for point sources that may have infrequent impacts, the number of samples that are missing or below the limit of detection, known problems with the collection or analysis of the species, and species reactivity.

A discussion of these considerations is provided in Reff et al. (2007). Detailed knowledge of the sources, sampling, and analytical uncertainties is the best way to decide on the species category. If detailed information about the data set is unavailable, the S/N ratios may be used to categorize one or more species. To conservatively use the S/N ratios to categorize species, categorize the species as "Bad" if the S/N ratio is less than 0.5 and "Weak" if the S/N ratio is greater than 0.5 but less than 1. For the sample Baltimore data set provided with the installation package (Dataset-Baltimore_con.txt and Dataset-Baltimore_unc.txt), these guidelines would result in aluminum, arsenic, barium, chlorine, chromium, manganese, and selenium categorized as "Bad" and lead, nickel, titanium, and vanadium as "Weak." Any changes made to the

user-provided uncertainty by making a species category "Weak" or by adding extra modeling uncertainty should be documented by the user and reported with the final solution.

For users familiar with EPA PMF, Table 2 shows a summary of the PMF input information for the Baltimore Example, which is used in Sections 5 and 6 to demonstrate PMF. This summary information will be presented for users who would like to run the software while learning about the new features and structure of EPA PMF 5.0.

A concentration/uncertainty scatter plot is displayed on the right of the screen (Figure 9, 3) and the plot shows the relationship between the concentration and the user provided or PMF calculated uncertainties. The species to be plotted is selected in the Input Data Statistics table either by clicking on the species row or scrolling up and down through the species and only one species can be displayed at a time. The statistics for each species are shown in the table: S/N; Minimum (Min), 25th, 50th, and 75th percentile; Maximum (Max), % Modeled Samples (number of samples with matched non-missing selected species divided by total number of input samples), and % Raw Samples (number of non-missing input samples divided by total number of input samples). For example, if four sites with equivalent number of data points and no missing data were ingested, and only one of the four sites was included for modeling, "% modeled samples"=25%, while "% raw samples"=100%, since there was no reduction of data directly upon ingest. If missing data were in the ingested data, and "exclude entire sample" for missing data was selected, both % modeled and % raw would be lower. The last two values are important because PMF requires that all good or weak category species be non-missing for the sample to be included in the PMF run. The % Modeled Samples and % Raw Samples can be used to identify the species that may be limiting the total number of samples used in a run.

Table 2. Baltimore example – summary of PMF input information.

**** Data Files ****			****** Base Run Summary ******		
Concentration file:	Dataset-Baltimore_con.txt		Number of base runs:	20	
Uncertainty file:	Dataset-Baltimore_unc.txt		Base random seed:	89	
			Number of factors:	7	
			Extra modeling uncertainty:	0	
Excluded Samples					
07/04/02					
07/07/02					
07/08/02					
12/31/02					
07/05/03					
01/01/05					
07/03/05					
07/01/06					
07/04/06					
**** Input Data Statistics ****					
Species	Category	S/N	Species	Category	S/N
PM2.5	Weak	9.0	Manganese	Weak	0.3
Aluminum	Weak	0.1	Nickel	Weak	0.5
Ammonium Ion	Strong	8.9	Organic Carbon	Strong	7.8
Arsenic	Weak	0.1	OM	Bad	7.8
Barium	Weak	0.0	Potassium Ion	Strong	2.1
Bromine	Strong	2.0	Selenium	Weak	0.2
Calcium	Strong	2.1	Silicon	Strong	2.0
Chlorine	Weak	0.1	Sodium Ion	Weak	1.0
Chromium	Weak	0.0	Sulfate	Strong	9.2
Copper	Weak	1.0	Titanium	Weak	0.7
Elemental Carbon	Strong	4.4	Total Nitrate	Strong	7.9
Iron	Strong	5.6	Vanadium	Weak	0.6
Lead	Weak	0.5	Zinc	Strong	5.1

The x-axis is the concentration, the y-axis is the uncertainty, and the graph title is the name of the species plotted. If users change a species categorization to "Weak," the concentration/uncertainty scatter plot for that species will be updated to three times the original uncertainty and the data points will be changed to orange squares. If users change a species categorization to "Bad," the graph for that species will not be displayed. A typical concentration and uncertainty relationship is a hockey stick shape where the MDL dominates the uncertainty at low concentrations and becomes linear as the percentage of the concentration dominates the uncertainty. Points with uncertainties that do not follow the general trend of the data should be further evaluated by reading available sampling and analytical reports.

The user can also add "Extra Modeling Uncertainty (0-100%)," which is applied to all species, by entering a value in the box in the lower right corner of the screen (Figure 9, 4). This value encompasses various errors that are not considered measurement or analytical errors and which are included in the user-provided uncertainty files. Issues that could cause modeling errors include variation of source profiles and chemical transformations in the atmosphere. The model uses the "Extra Modeling Uncertainty" variable to calculate "sigma," which corresponds to total uncertainty (modeling uncertainty plus species/sample-specific uncertainty). If the user specifies extra modeling uncertainty, all concentration/uncertainty graphs will be updated to reflect the increase in uncertainty. As shown in Equation 1-2, the uncertainty values are a critical input in the PMF model.

On this screen, the user can also specify a "Total Variable" (Figure 9, 2) that will be used by the program in the post-processing of results. For example, if the data used are $PM_{2.5}$ components, the total variable would be $PM_{2.5}$ mass. The user specifies the total variable by selecting the species and pressing the "Total Variable" button beneath the Input Data Statistics table. Because a total variable should not have a large influence on the solution, it should be given a high uncertainty. Therefore, when a species is selected as a total variable, its categorization is automatically set to "Weak." If the user has already adjusted the uncertainty of the total variable outside of PMF and wishes to categorize it as "Strong," the default characterization can be overridden by selecting "Strong" for the variable after selecting "Total Variable." A species designated "Bad" cannot be selected as a total variable, and a total variable cannot be made "Bad."

The status bar in the Concentration/Uncertainty screen displays the number of species of each category as well as the percentage of samples excluded by the user. Hot keys can be used to assign "Strong" (Alt-S), "Weak" (Alt-W), "Bad" (Alt-B), and "Total Variable" (Alt-T). The user can also sort the input data by clicking on the column headers. Clicking on the "Species" and "Cat" columns will sort the input data in alphabetical or reverse alphabetical order. Clicking on the remaining columns will sort the data in ascending or descending order. To return to the original species sort order (which corresponds to the order listed in the input concentration data file on the Data Files screen) the user can select "Unsort" (Figure 9, 2) or use a hot key (Alt-U).
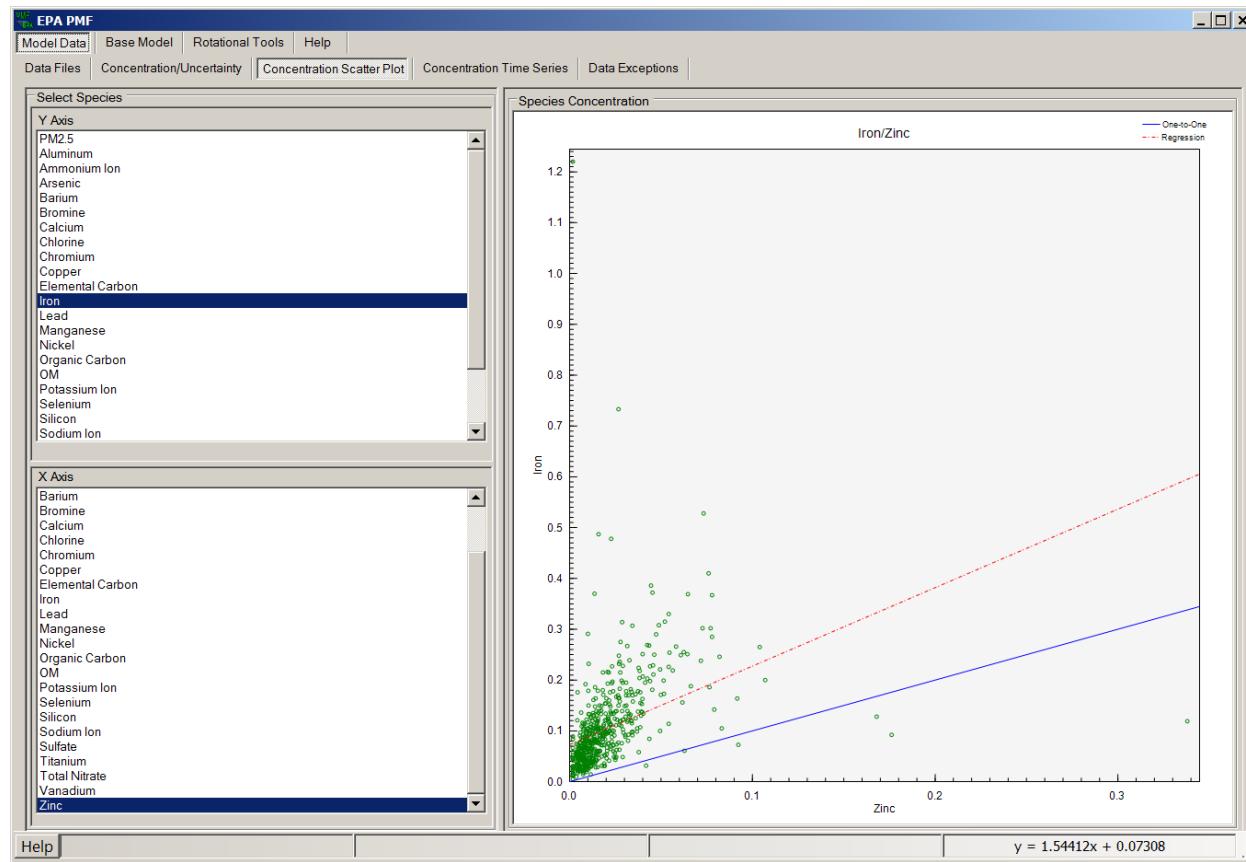
5.5.2   Concentration Scatter Plots

Scatter plots between species are a useful pre-PMF analysis tool; a correlation between species indicates a similar source type or source locations. The user should examine scatter plots to

look for expected relationships, as well as to look for other relationships that might indicate sources or source categories.

The Concentration Scatter Plot screen shows scatter plots between two user-specified species (Figure 10). The user selects the species for each axis in the appropriate "Y Axis" or "X Axis" list. Only one species can be selected for each axis. A one-to-one line (in blue) and linear regression line (in dashed red) are shown on the plot. Axis labels are the species names and units (if provided) and the plot title is "Y Axis Species/X Axis Species." Some examples of linear relationships between species indicate source impacts: iron and zinc for steel production and sulfate and ammonium ion for ammonium sulfate from coal-fired power plants.

As the user mouses over the points, the status bar at the bottom of the window shows the date, y-value, x-value, and the regression equation.

Figure 10. Example of a concentration scatter plot.

5.5.3　Concentration Time Series

Time series of species concentrations (Figure 11) are useful to determine whether expected temporal patterns are present in the data and whether there are any unusual events. By overlaying multiple species, the user can see if any unusual events are present across a group

of species that may indicate a shared source. The user should also examine time series for extreme events that should be excluded from modeling (for example, elevated potassium concentrations on the Fourth of July from fireworks). The firework impacts can show up both before and after the Fourth of July as well as on New Year's Eve (elevated concentrations on the January 1 sample).

The user can select up to 10 species in the Concentration Time Series list by checking the box next to each species name (Figure 11, 1). The selected species will be displayed in varying colors on the plot. To clear all species from the plot, the user should select "Clear Selections" below the list. Vertical orange lines denote January 1 of each year (if appropriate) for reference. A legend is provided at the top of the graph with species names and units (if available). Vertical lines separating points by SampleID can be toggled on the Data Files screen. A legend is provided at the top of the graph with species names and units (if available). The legend automatically updates with each selection. If data are not in order by date, e.g., if there are multiple SampleIDs for a given date, the x-axis will display "Sample Number", as the plot is simply a line plot, rather than a time series of sequential samples. The legend automatically updates with each selection. The status bar on this screen shows the selected sample date/time, the SampleID if provided, the number of samples included out of the total number of samples, and the percent of samples excluded by the user. The arrow buttons below the plot, or the right and left arrow keys on the keyboard, can be used to scroll through samples. If a group of samples is selected, the arrows will move the first selected sample forward/backward by one sample. Samples can be removed from analysis by selecting individual data points with a single mouse click or dragging the mouse over a range of dates. Pressing the "Exclude Samples" button below the plot will remove the samples and gray them out for all species (Figure 11, 2). Excluded samples can be included again by selecting the data point/range on any species time series graph and pressing "Restore Samples." If a sample is removed from analysis, it will not be included in the statistics or plots generated by EPA PMF or in any model output, but it is not removed from the original user input files. Hot keys can be used to exclude (Alt-E) or restore (Alt-R) selected samples. A number of samples impacted by fireworks were excluded: 07/04/02, 07/07/02, 07/08/02, 12/31/02, 07/05/03, 01/01/05, 07/03/05, 07/01/06, and 07/04/06. Impacts such as fireworks represent a challenge for PMF and multivariate models because they are infrequent short duration events with high concentrations.
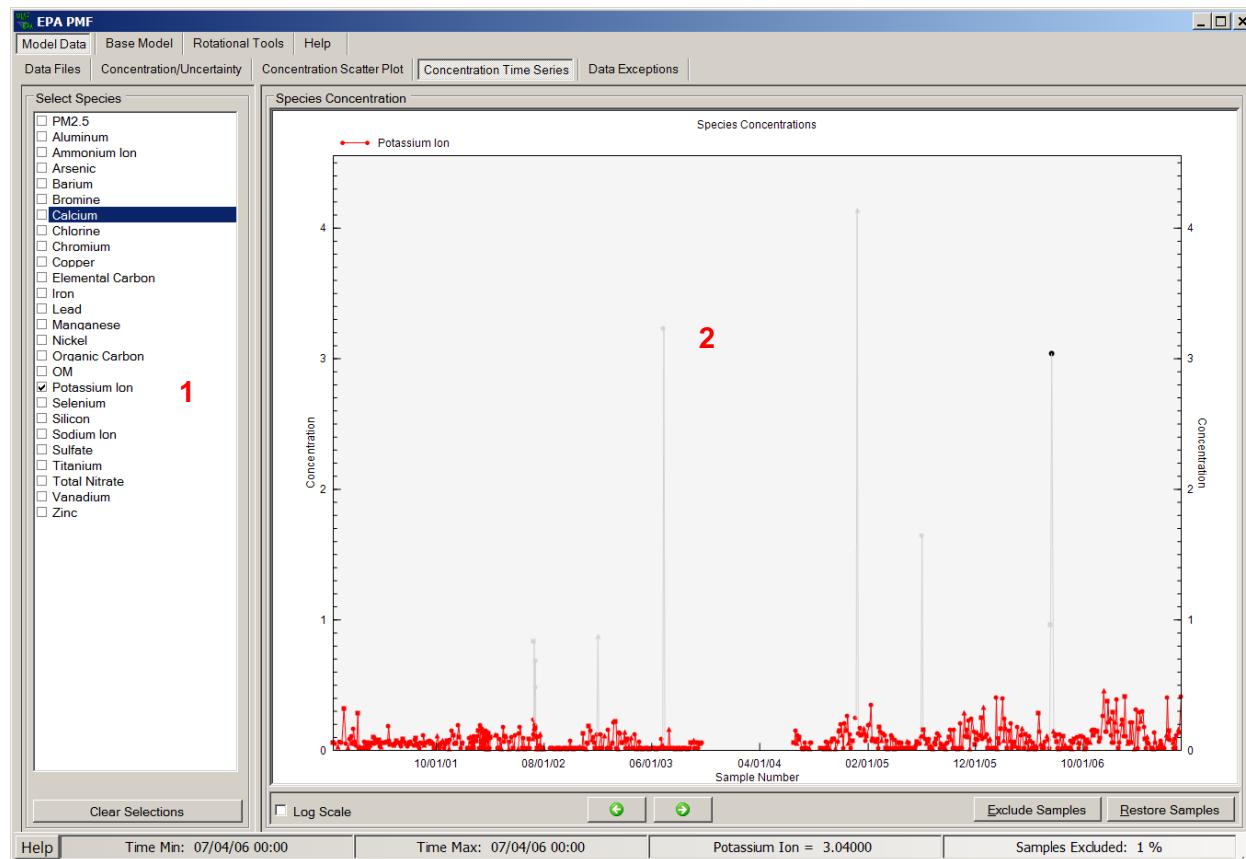
5.5.4   Data Exceptions

Changes made by the GUI to the input data are detailed in the Data Exceptions screen. These changes include designating a species "Weak" or "Bad," excluding a sample via the Concentration Time Series screen, or excluding a sample using "Missing Value Indicator" in the Data Files screen "Input Files" box. Click the right mouse button to save the data exceptions information.

5.6   Base Model Runs

Base Model Run produces the primary PMF output of profiles and contributions. The base model run uses a new random seed or starting point for iterations if the "Random Start" option is selected. A user can test whether the solution found is a local or global minimum by using many random seeds and examining whether the Q(robust) values are stable. A constant seed

can be set by unselecting the "Random Start" box. A constant seed with the same number of factors and runs will generate the same PMF result; the seed is also saved in the configuration file. The configuration file can be reloaded for additional evaluation of PMF solutions and can also be sent to collaborators for evaluation of a PMF solution.

Figure 11. Example of the Concentration Time Series screen with excluded and selected samples.

5.6.1   Initiating a Base Run
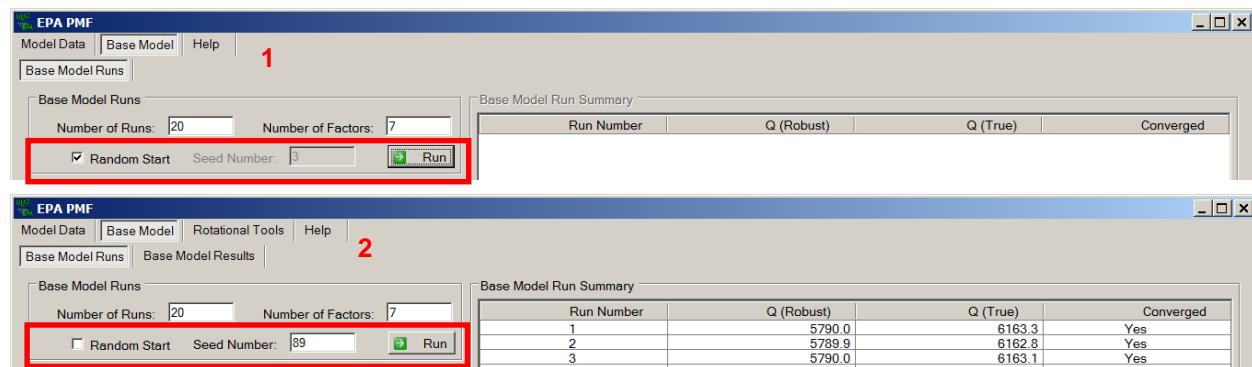
Base model runs are initiated on the Base Model Runs screen under the Base Model tab (Figure 12). The following parameters need to be specified:

- **"Number of Runs"** – the number of base runs to be performed; this number must be an integer between 1 and 999. The recommended number of runs is 20, which will allow for an evaluation of the variation in Q.

- **"Number of Factors"** – the number of factors the model should fit; this number must be an integer between 1 and 999. The number of factors to be chosen will depend on the user's understanding of the sources impacting samples, number of samples, sampling time resolution, and species characteristics.

- **"Seed"** – the starting point for each iteration in ME-2; the default is Random Start, which tells the GUI to randomly choose a starting point for each run. The random seed number is displayed in the "Seed Number" box (Figure 12, 1). To reproduce results, unselect the "Random Start" option, so that the seed number used will be saved as part of the .cfg file, and thus an identical solution can be recreated later using the same .cfg (Figure 12, 2).

After the aforementioned parameters are specified, the user should press the "Run" button in Base Model Runs to initiate the base runs. Once runs are initiated, the "Run Progress" box in the lower right corner of the screen activates. Base model runs can be terminated at any time by pressing the "Stop" button in the "Run Progress" box. The progress bar in this box also fills whenever runs are performed. No information about the runs will be saved or displayed if the runs are stopped.

The status bar on the Base Model Runs screen displays the same information as on the Data Files screen.

Figure 12. Example of the Base Model Runs screen showing Random Start (1) and Fixed Start (2).
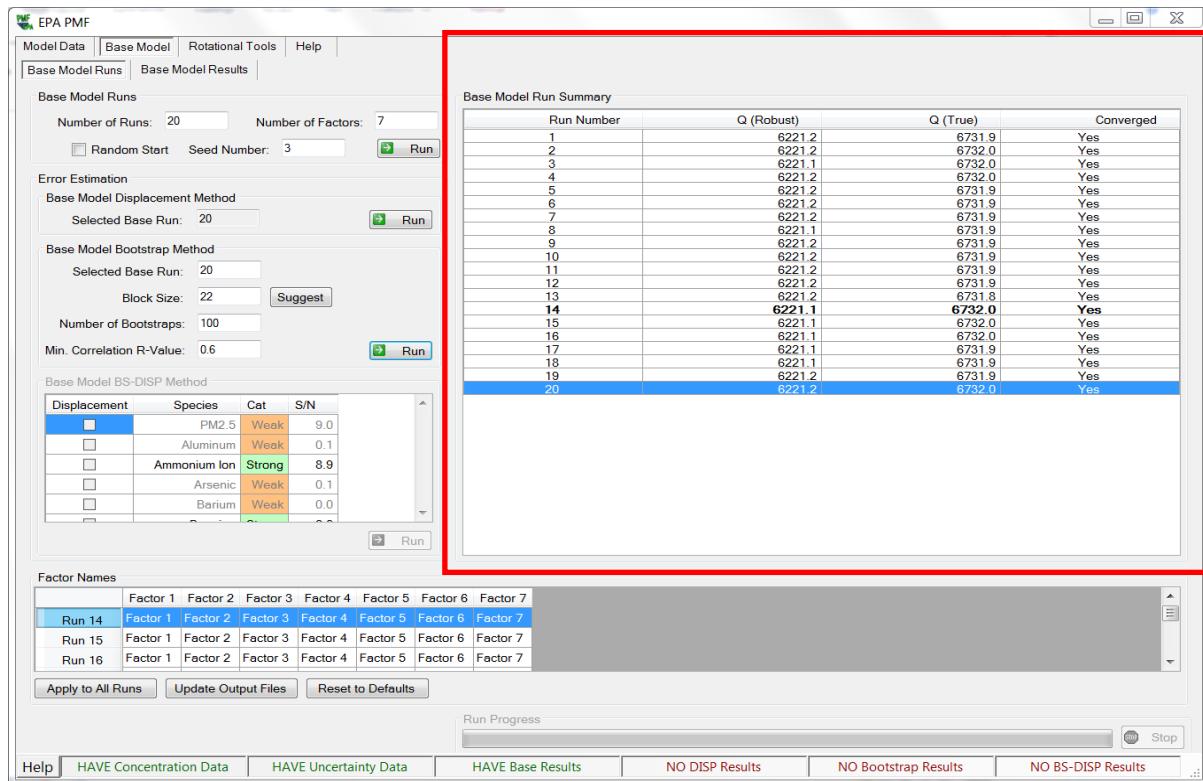
5.6.2 Base Model Run Summary

When the base runs are completed, a summary of each run appears on the right portion of the Base Model Runs screen in the Base Model Run Summary table (Figure 13, red box). The Q-values are goodness-of-fit parameters calculated using Equation 1-2 and are an assessment of how well the model fits the input data. The run with the lowest Q(robust) is highlighted and only the converged solutions should be investigated. Non-convergence implies that the model did not find any minima. Several things could cause the non-convergence, including uncertainties that are too low or specified incorrectly, or inappropriate input parameters.

The Q(robust) and Q(true) values provide a comparison of the fit of the runs; more detail is provided by comparing the residuals. The intra-run residual calculation compares the residuals between base runs by adding the squared difference between the uncertainty-scaled residuals for each pair of base runs (Equation 5-5):

$$d_{jkl} = \sum_i \left(r_{ijk} - r_{ijl} \right)^2 \qquad \textbf{(5-5)}$$

where *r* is the scaled residual, *i* is the sample, *j* is the variable, and *k* and *l* are two different runs. These results are shown in a matrix and can be used to identify runs with significantly different fits. Also, the paired species values for each run can be compared by adding the *d*-values (Equation 5-6).

Figure 13. Example of the Base Model Runs screen after base runs have been completed.

$$D_{kl} = \sum_j d_{jkl} \qquad \text{(5-6)}$$

The *D*-values are reported in a matrix of base run pairs. The user should examine this matrix for large variations, which indicate that two runs resulted in truly different solutions rather than merely being rotations of each other. If different solutions are seen, the user can then examine the *d*-values, which will indicate the individual species that are fitted differently across the runs.

The distribution of species concentration and percent of species sum results are also evaluated for each of these factors: Lowest Q, Minimum (Min), 25th percentile, 50th percentile, 75th percentile, Maximum (Max), Mean, Standard Deviation (SD), Relative Standard Deviation (SD*100/mean), and RSD % Lowest Q. Large variations in species distributions may indicate that the factor profile is changing due to process changes, reactivity, or measurement issues.

These intra-run variability results are recorded in the *_diag file and can be viewed through the GUI by selecting the Diagnostics tab and scrolling to "Scaled residual analysis." In addition, a factor summary of the species distribution compared to the lowest Q(robust) run is recorded in

the *_run_comparison file and can be viewed through the GUI by selecting the Diagnostics tab and the lower window "Run Comparison Statistics."

5.6.3   Base Model Results

Details of the base model run results are provided in the screens under the Base Model Results tab. The results for the run with the lowest Q(robust) value are automatically displayed. The user can change the run number either by highlighting it in the Base Model Run Summary table on the Base Model Runs screen, or by selecting the run number at the bottom of the Base Model Results screen.

Residual Analysis

The Residual Analysis screen (Figure 14) displays the uncertainty-scaled residuals in several formats for the selected run. At the left of the screen (Figure 14, 1), the user can select a species, which will be displayed in the histogram in the center of the screen (Figure 14, 2). The histogram shows the percent of all scaled residuals in a given bin (each bin is equal to 0.5). These plots are useful to determine how well the model fits each species. If a species has many large scaled residuals or displays a non-normal curve, it may be an indication of a poor fit.

The species in Figure 14 (sulfate) is well-modeled; all residuals are between +3 and -3 and they are normally distributed. Gray lines are provided for reference at +3 and -3. Selecting the "Autoscale Histogram" box will set the y-axis range maximum at +10% of the maximum bin count for each species. If the box is unchecked, the y-axis maximum is fixed at 100%. Species with residuals beyond +3 and -3 need to be evaluated in the Obs/Pred Scatter Plot and Time Series screens. Large positive scaled residuals may indicate that PMF is not fitting the species or the species is present in an infrequent source.

The screen also displays the samples with scaled residuals that are greater than a user-specified value (Figure 14, 3). The default value is 3.0. The residuals can be displayed as "Dates by Species" or "Species by Dates" by choosing the appropriate option above the table. When a species is selected in the list on the left (Figure 14, 1), the table on the right (Figure 14, 3) automatically scrolls to that species.

Observed/Predicted Scatter Plot

A comparison between observed (input data) values and predicted (modeled) values is useful to determine if the model fits the individual species well. Species that do not have a strong correlation between observed and predicted values should be evaluated by the user to determine whether they should be down-weighted or excluded from the model.

A table in the Obs/Pred Scatter Plot screen shows Base Run Statistics for each species (Figure 15, 1). These numbers are calculated using the observed and predicted concentrations to indicate how well each species is fit by the model. The statistics shown are the coefficient of determination (r^2), Intercept, Intercept SE (standard error), Slope, Slope SE, SE, and Normal Resid (normal residual). The table also indicates whether the residuals are normally distributed, as determined by a Kolmogorov-Smirnoff test. If the test indicates that the residuals are not

normally distributed, the user should visually inspect the histogram for outlying residuals. If not all statistics are visible, the user can use the scroll bars at the bottom and side of the table to display additional statistics. These statistics are also provided in the *_diag output file. The Obs/Pred Scatter Plot (Figure 15, 2) shows the observed (x-axis) and predicted (y-axis) concentrations for the selected species. A blue one-to-one line is provided on this plot for reference (a perfect fit would line up exactly on this line), and the regression line is shown as a dotted red line. The status bar on this screen (Figure 15) displays the date, x-value, y-value, and regression equation between predicted and observed data as data points are moused-over (Figure 15, 3).
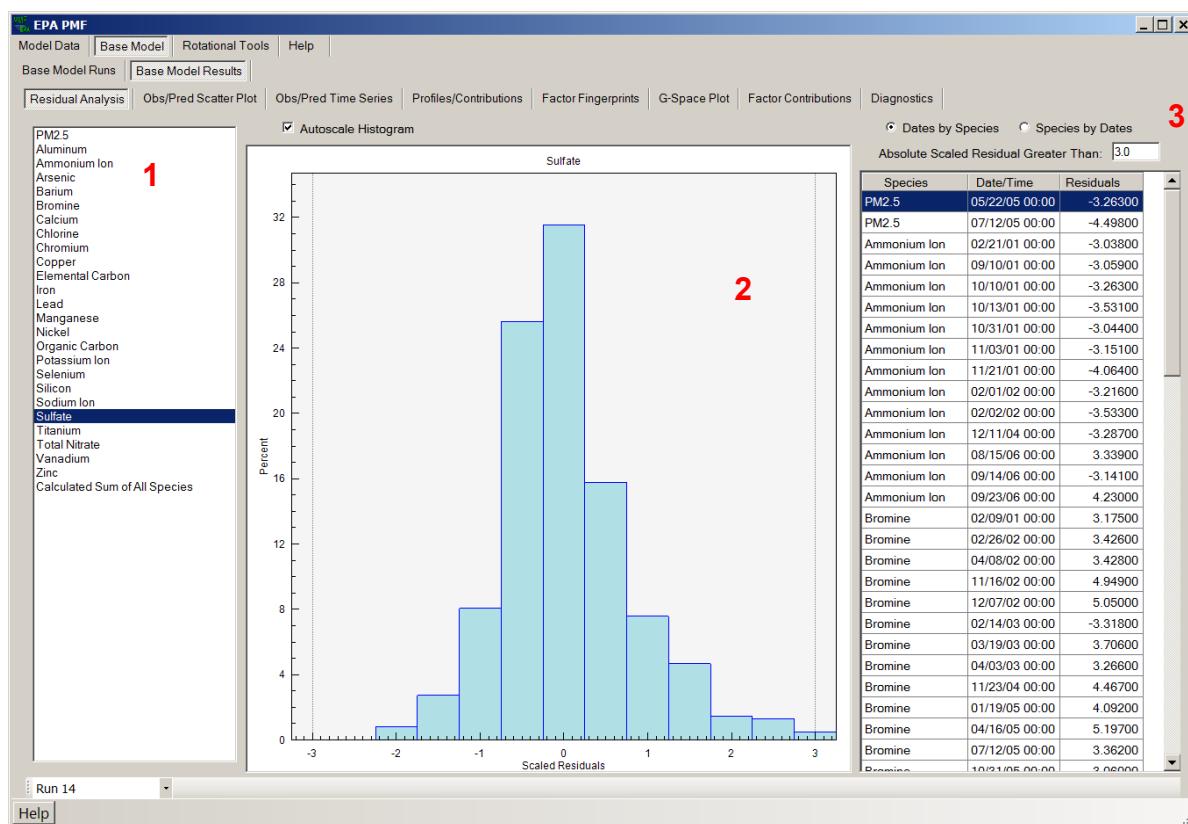
Figure 14. Example of the Residual Analysis screen.

Observed/Predicted Time Series

The data displayed on the Obs/Pred Scatter Plot screen are the same data displayed as a time series on the Obs/Pred Time Series screen (Figure 16). When a species is selected by the user, the observed (user-input) data for that species are displayed in blue and the predicted (modeled) data are displayed in red. The user can view this screen to determine when the model is fitting the observed data well. If the peak values of a species are not reproduced by the model, it may be advisable to exclude the species or change the species category to weak. The status bar on this screen displays the date, and the observed and predicted concentrations for the sample closest to the black vertical dotted reference line.
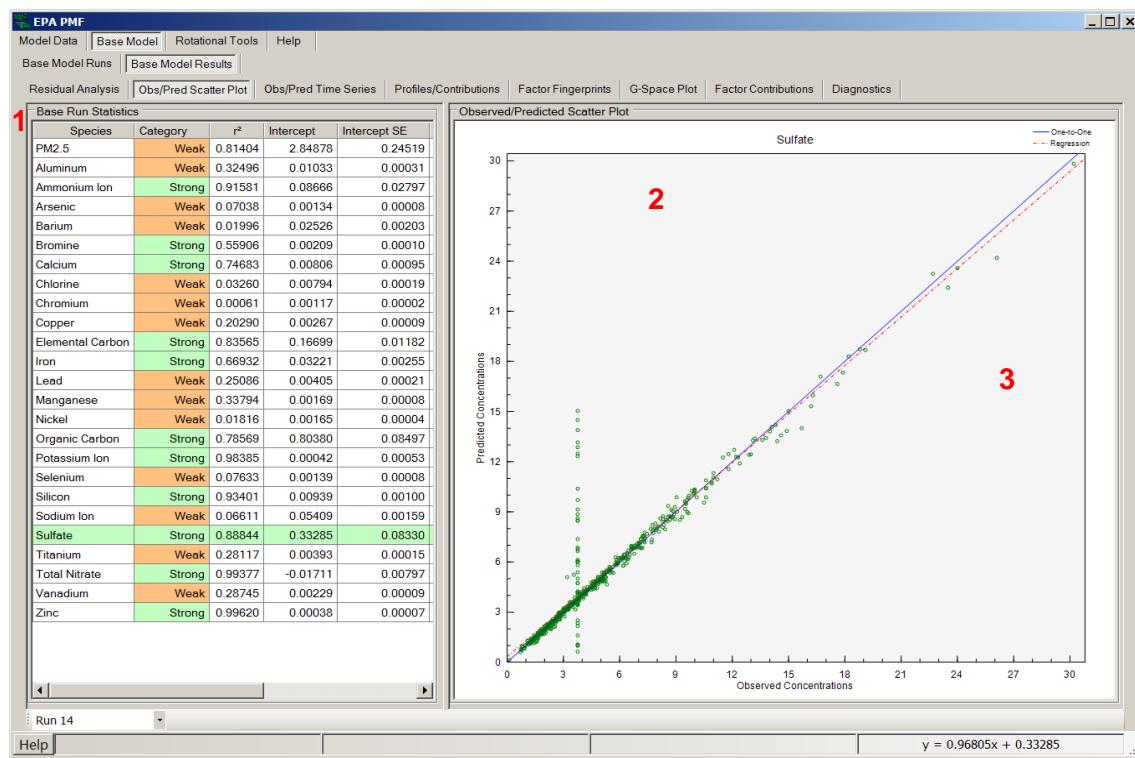
Figure 15. Example of the Obs/Pred Scatter Plot screen.
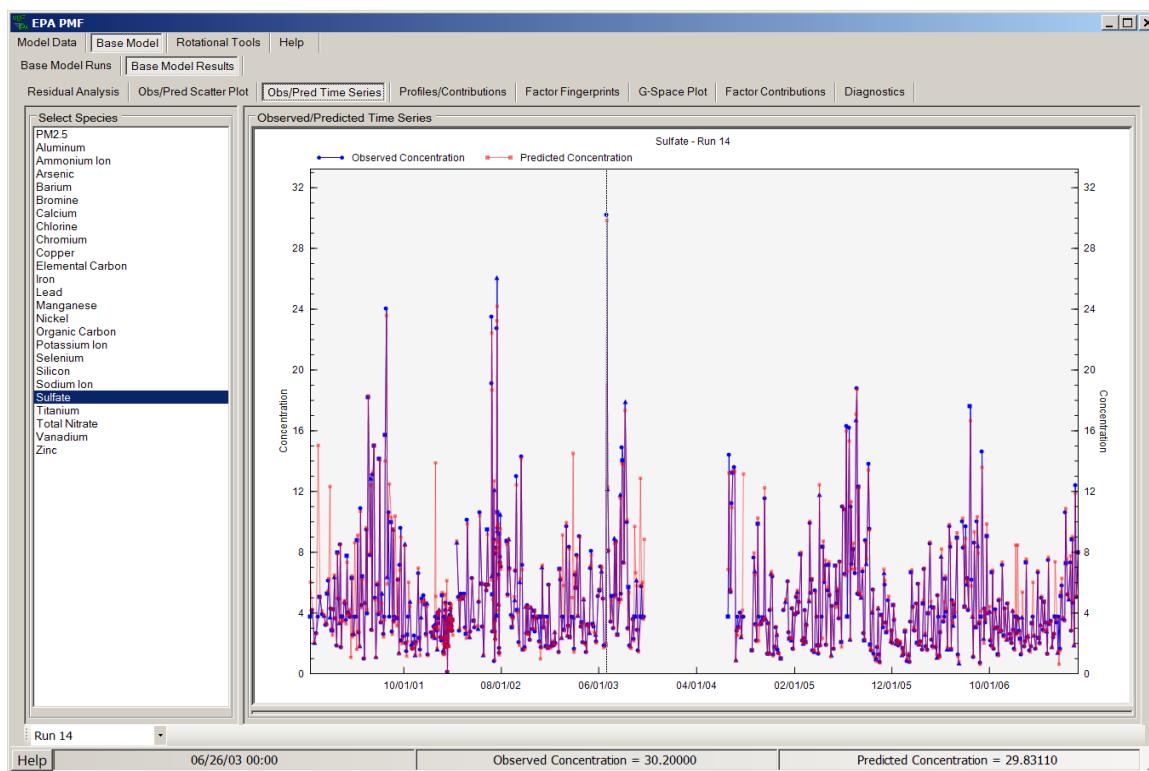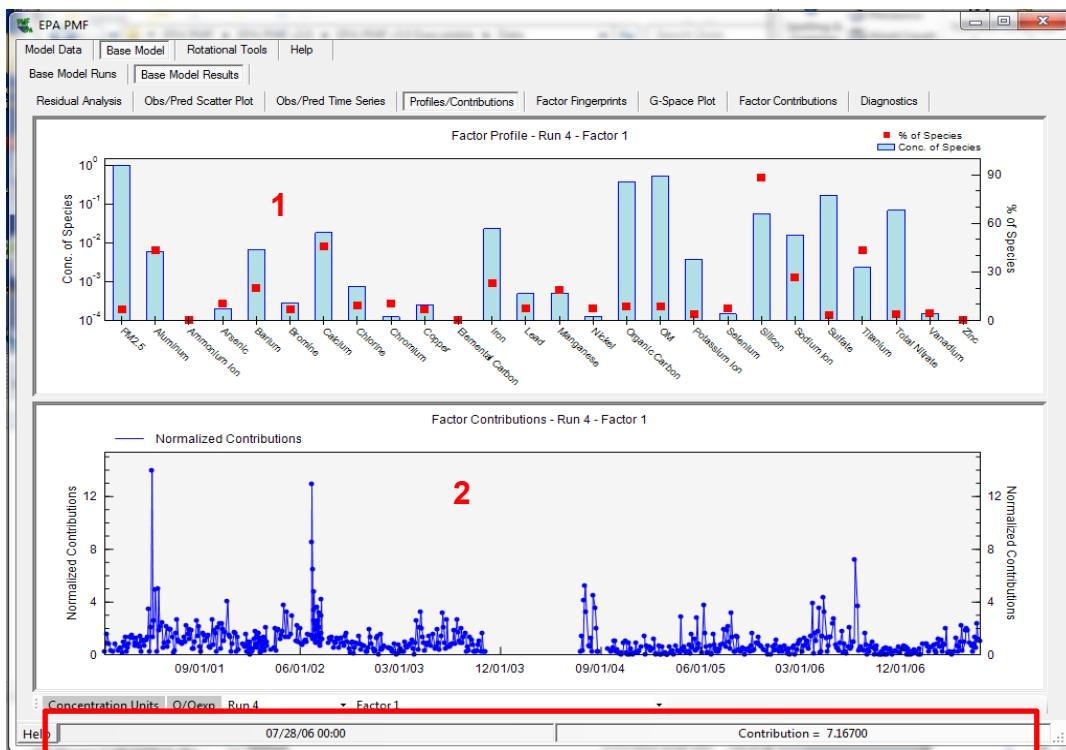
Figure 16. Example of the Obs/Pred Time Series screen.

Profiles/Contributions
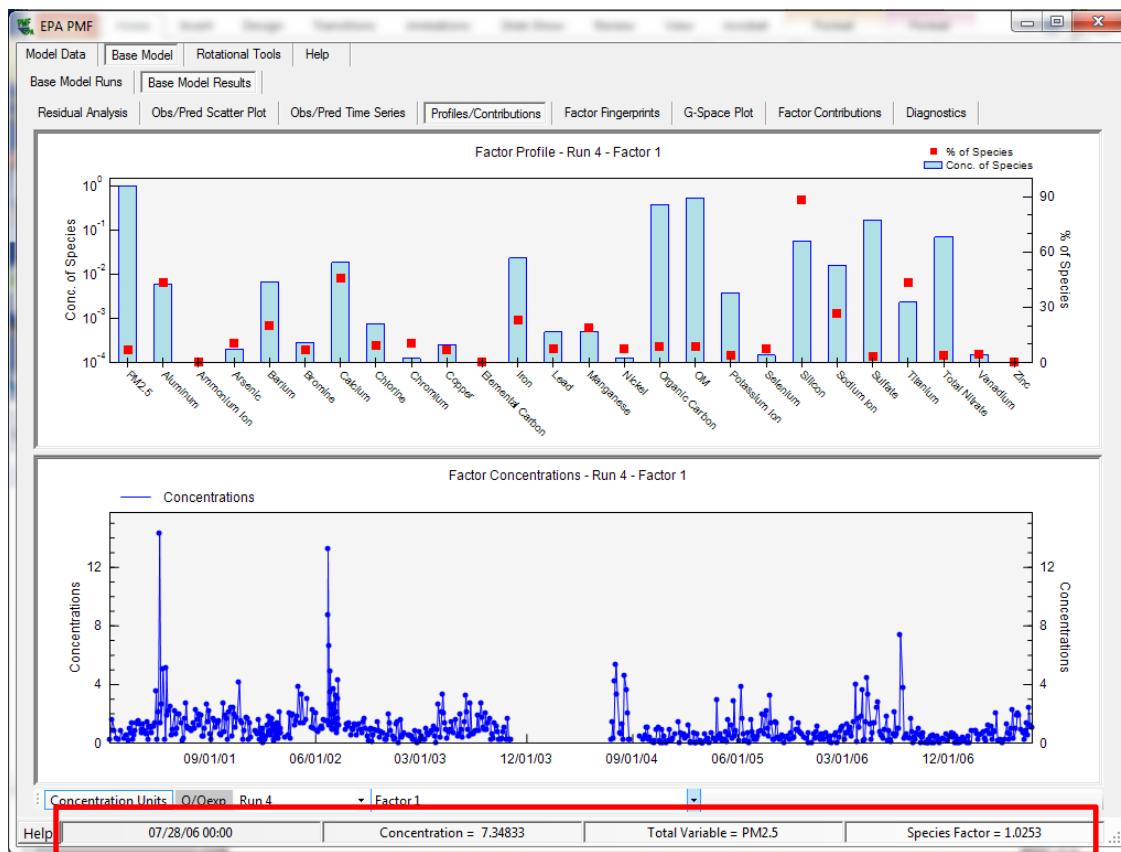
The factors resolved by PMF are displayed under the Profiles/Contributions screen. Two graphs are shown for each factor, one displaying the factor profile and the other displaying the contribution per sample of each factor (Figure 17). The profile graph, displayed on top (Figure 17, 1), shows the concentration of each species apportioned to the factor as a pale blue bar and the percent of each species apportioned to the factor as a red box. The concentration bar corresponds to the left y-axis, which is a logarithmic scale. The percent of species corresponds to the right y-axis. The bottom graph shows the contribution of each factor to the total mass by sample (Figure 17, 2). This graph is normalized so that the average of all contributions for each factor is 1. The status bar on this screen (Figure 17, red box) displays the date and contributions of data points as they are moused-over on the Factor Contributions plot.

Pull-down menus at the bottom of the Profiles/Contributions screen allow the user to easily compare runs and factors. Beginning in the bottom left corner, each run can be chosen by toggling to and clicking on the appropriate run number. The user can quickly compare runs to assess the stability of the solution or determine what, if any, individual species or factors are varying between runs. Users can switch between the factors resolved by PMF by using the pull-down menu second from the left. Factor 1 is currently selected. The user can create a stacked plot of the profiles or time series by first selecting either the factor profile plot or the factor concentration plot, right-clicking on the mouse to view the menu, and selecting "Stacked Graphs."

Figure 17. Example of the Profiles/Contributions screen.

If a total variable is selected, the user can select "Concentration Units" in the bottom left corner of the Profiles/Contributions screen to display the contributions in the same units as the total mass (Figure 18). If this option is selected, the GUI multiplies the contributions by the mass of the total variable in that factor. The status bar displays the date, factor contribution, total variable selected, and the species factor as they are moused-over on the Factor Contributions plot (Figure 18, red box). If no mass from the total variable is apportioned to the factor, the graph is not shown and the GUI instead displays "Total Variable mass is 0 for this run/factor."
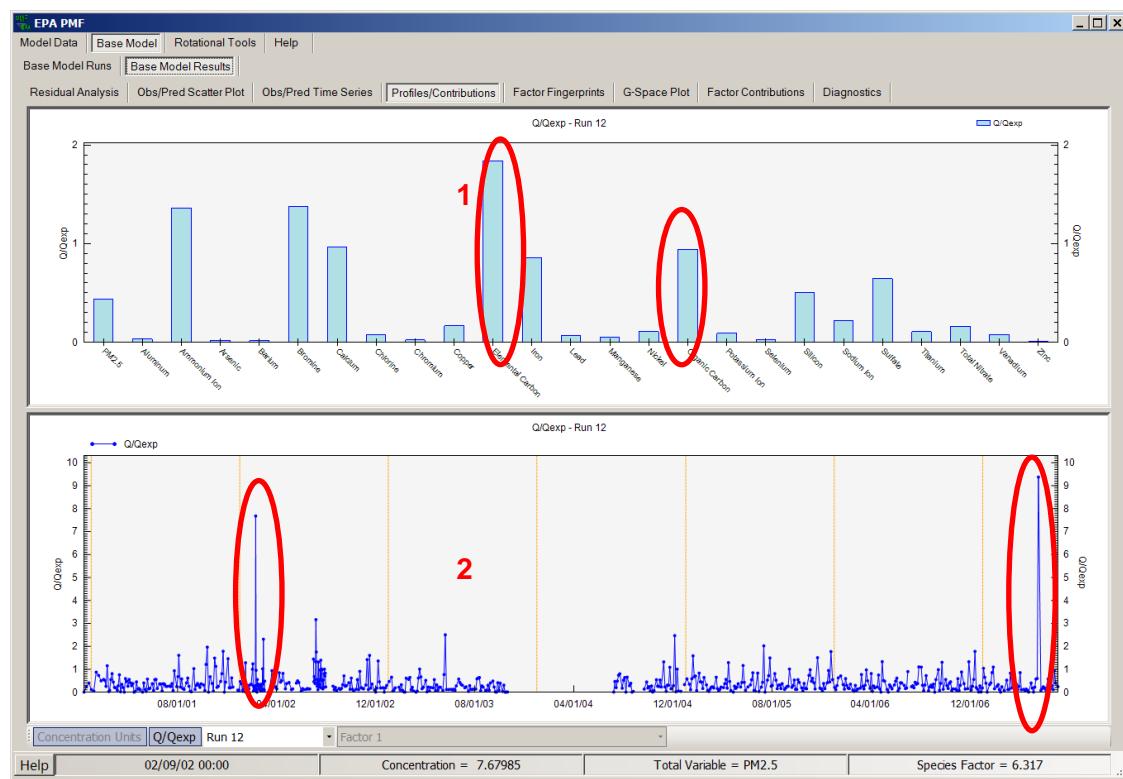
Figure 18. Example of the Profiles/Contributions screen with "Concentration Units" selected.

The user can give a factor a name in the Profiles/Contributions screen by right-clicking on the mouse to view the menu, selecting "factor name," typing in a unique name, and then pressing "Apply Factor Name." The new factor name(s) will appear on the Factor Fingerprints, G-Space Plot, Factor Contributions, and Diagnostics screens. Factor 1 has high concentrations of sulfate and ammonium ions and it represents secondary sulfate formation from the combustion of coal in power plants. The identification of factors from PMF requires review of measured species relationships. Some sources may be easily identified; an industrial source, for example, may be dominated by peaks in zinc concentrations. Other sources may be more difficult to identify.

The species Q/Qexpected (Q/Qexp) can be displayed by selecting the "Q/Qexp" toggle on the Profiles/Contributions tab (Figure 19). $Q_{expected}$ is equal to (number of non-weak data values in

X) - (numbers of elements in **G** and **F**, taken together). For example, for five factors, 642 samples, and 19 strong species, this equals (642*19) – ((5*642)+(5*19)), or 8893. For each species, the Q/Qexp for a species is the sum of the squares of the scaled residuals for that species, divided by the overall Qexpected divided by the number of strong species. For each sample, the Q/Qexp is the sum of the square of the scaled residuals over all species, divided by the number of species. Examining the Q/Qexp graphs is an efficient way to understand the residuals of the PMF solution, and in particular, what samples and/or species were not well modeled (i.e., have values greater than 2). A comparison of the species results shows that EC and OC have elevated Q/Qexp values, which might indicate that motor vehicle contribution could be better explained by adding another source (Figure 19, 1). Also, the time series of Q/Qexp values shows two days where the species concentrations were not fit as well compared to other days (Figure 19, 2). These days might have had unique source impacts and should be investigated further.
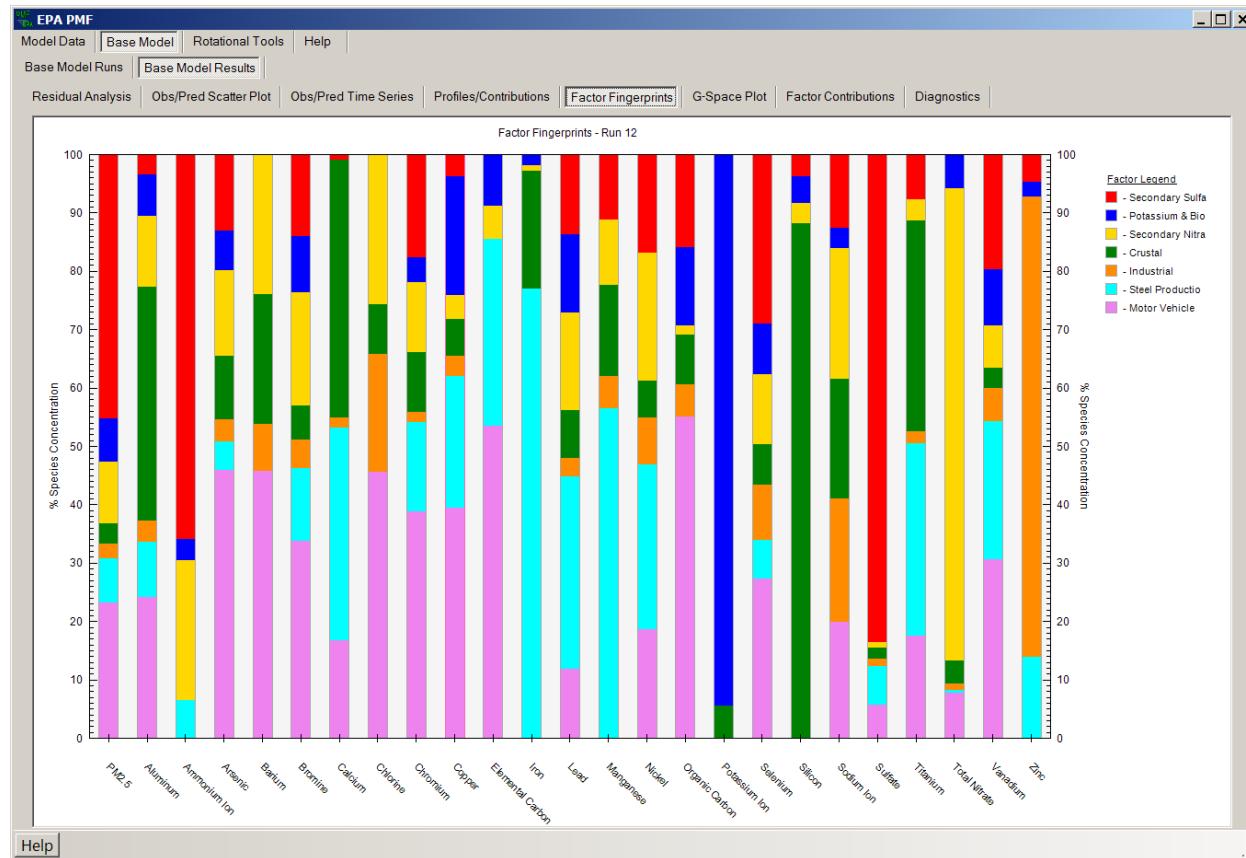
Figure 19. Example of the Profiles/Contributions screen with "Q/Qexp" selected.

Factor Fingerprints

The concentration (in percent) of each species contributing to each factor is displayed as a stacked bar chart in the Factor Fingerprints screen (Figure 20). This plot can be used to verify factor names and determine the distribution of the factors for individual species. The plot only displays the currently selected run. To change runs, the user can select a different run number
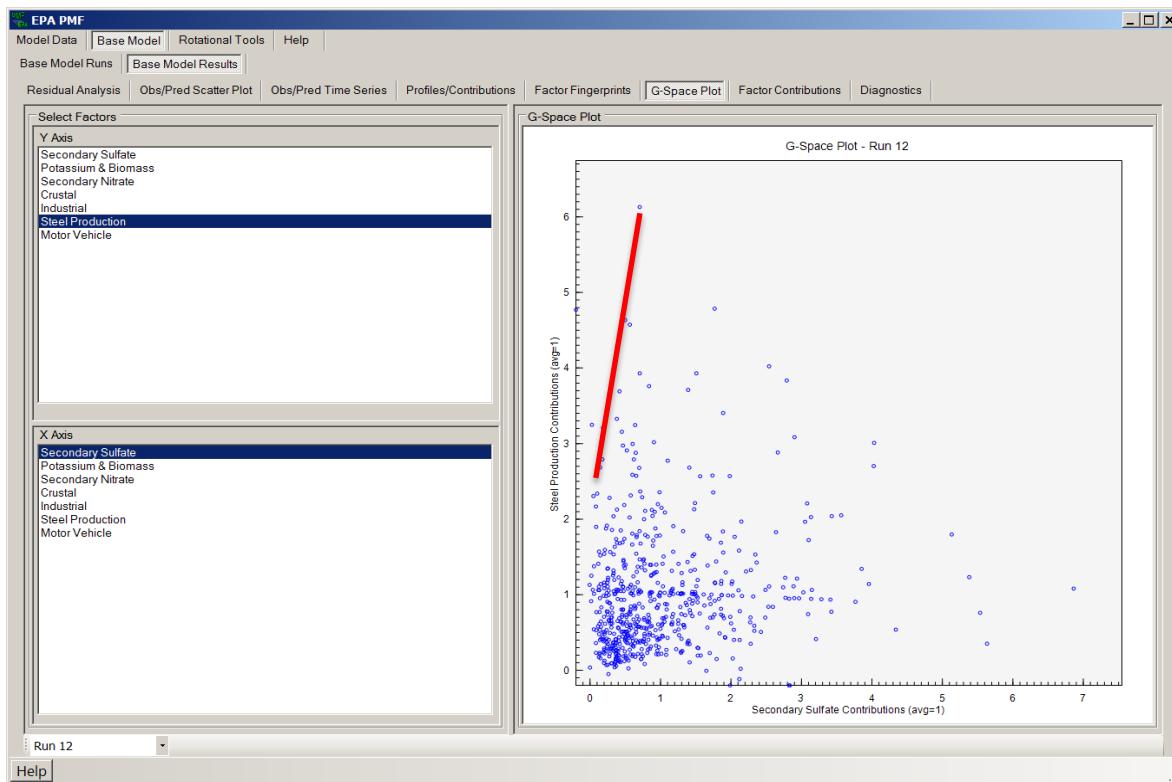
at the bottom left-hand corner of the Residual Analysis, Obs/Pred Scatter Plot, Obs/Pred Time Series, or Profiles/Contributions screens.

Figure 20. Example of the Factor Fingerpints screen.

G-Space Plot

The G-Space Plot screen (Figure 21) shows scatter plots of one factor versus another factor, which can be used to assess rotational abiguity as well as the relationship between source contributions. A more stable solution will have many samples with zero contributions on both axes, which provide greater stability in the PMF solution to less rotational ambiguity. A solution or combination of sources may also have no points on or near the axes, which results in greater rotational ambiguity. The user selects one factor for the y-axis and one factor for the x- axis from lists on the left of the screen. A scatter plot of these factors will be shown on the right of the screen. The plot in Figure 21 is an example of a non-optimal rotation of a factor, which has an upper edge that is not aligned with the axis in the G-Space plot (red line added for reference). In EPA PMF, the user can explore different rotations via the Fpeak option (Paatero et al., 2005), which is explained in detail in Section 6.1. The G-Space plots are also useful for understanding the relationship between the factor source contributions and the pattern in Figure 21 shows not relationship between regional secondary sulfate and local steel production.
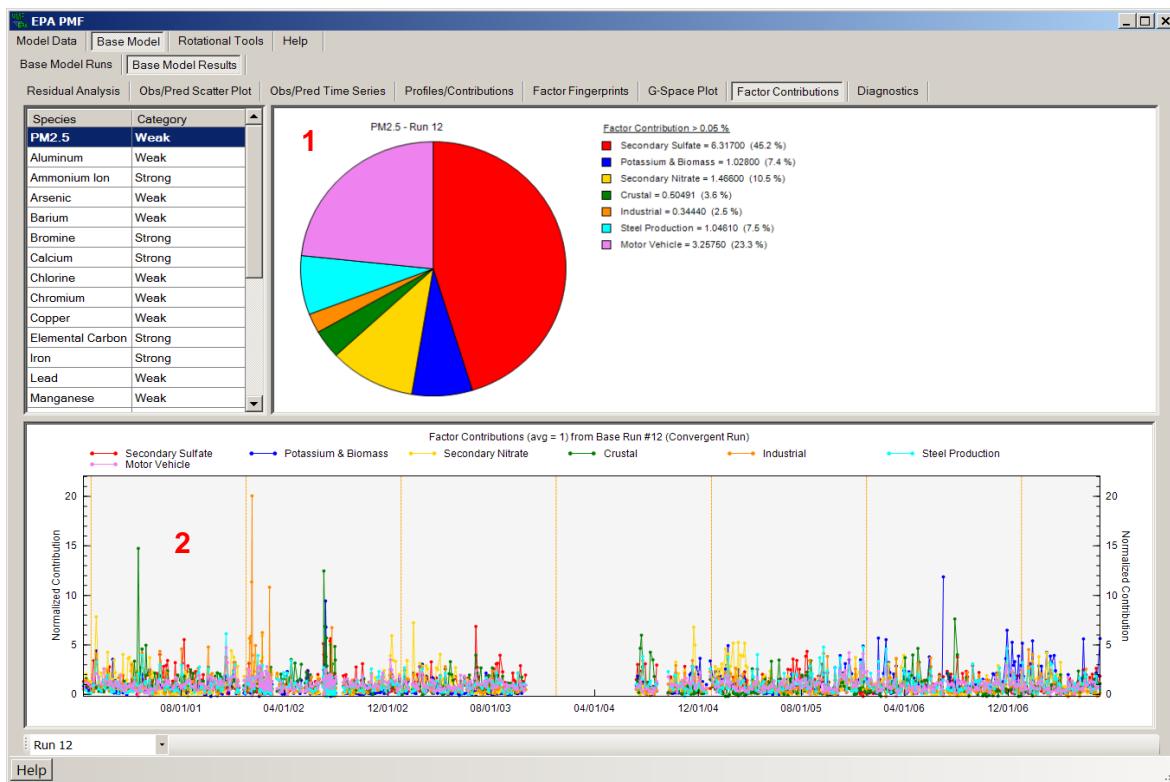
Figure 21. Example of the G-Space Plot screen with a red line indicating an edge.

Factor Contributions

The Factor Contributions screen (Figure 22) shows two graphs. The top graph is a pie chart which displays the distribution of each species among the factors resolved by PMF (Figure 22, 1). The species of interest is selected in the table on the left of the screen; the categorization of that species is also displayed for reference. If a total variable was chosen by the user under the Concentration/Uncertainty screen, that variable is boldfaced in the table. The pie chart for the selected species is on the right side of the screen. If the user has specified a total variable, the distribution of this variable across the factors will be of particular importance. The user may also want to examine the distribution of key source tracer species across factors. The bottom graph shows the contribution of all the factors to the total mass by sample (Figure 22, 2). The dotted orange lines denote January 1 of each year. The graph is normalized so that the average of all the contributions for each factor is 1, to allow for a comparison of the temporal pattern of source contributions.

Diagnostics

The Diagnostics screen displays two outputs, which are also saved in the output directory: *_diag and the *_run_comparison file.

Figure 22. Example of the Factor Contributions screen.

Output Files

After the base runs are completed, the GUI creates output files that contain all of the data used for the on-screen display of the results. The number of output files created depends on the type of output file selected: tab-delimited (*.txt) and comma-delimited (*.csv) create five output files – *_diag, *_contrib, *_profile, *_resid and *_runcomparison. Excel Workbook (*.xls) creates two output files – *_diag and *_base. The output files are saved to the directory specified in the "Output Folder" box in the Data Files screen, using the prefix specified in the "Output File Prefix" box.

- ***_diag** contains a record of the user inputs and model diagnostic information (identical to the Diagnostics screen).

- ***_contrib** contains the contributions for each base run used to generate the contribution graphs on the Profiles/Contributions tab. Contributions are sorted by run number. Normalized contributions are shown first, followed by contributions in mass units if a total variable is specified.

- ***_profile** contains the profiles for each base run used to generate the profile graphs on the Profiles/Contributions tab. Profiles are sorted by run number. Profiles in mass units are written first, followed by profiles in percent of species and concentration fraction of species total if a total mass variable is specified.

- ***_resid** contains the residuals (regular and scaled by the uncertainty) for each base run, used to generate the graphs and tables on the Residual Analysis screen.

- ***_run_comparison** contains a summary of the species distribution for each factor over all PMF runs and compared to the lowest Q(robust) run.

- ***_base** contains the *_contrib, *_profile, *_resid and *_run_comparison on separate worksheets in the same Excel Workbook. This output file only appears if the user selects "Excel Workbook" as the output file type.
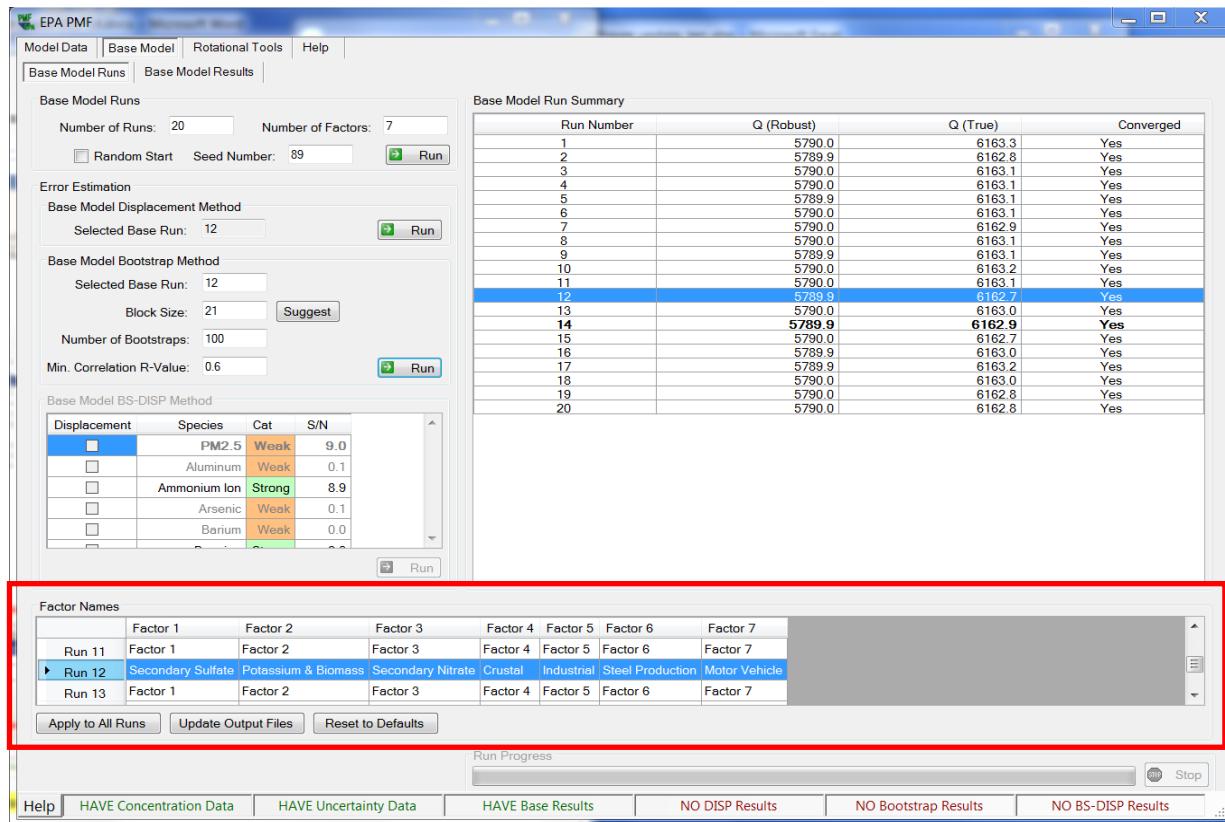
5.6.4   Factor Names on Base Model Runs Screen

The Factor Name can be entered or changed on the Profiles/Contributions screen or the Base Runs screen. After the base runs are completed, the "Factor Names" box located in the lower left portion of the Base Model Runs screen will be populated (Figure 23, red box). Each row in the matrix will be labeled by run number, in ascending order, and each column will be labeled by factor number, in ascending order. The table is then populated with the factor name associated with each column header.
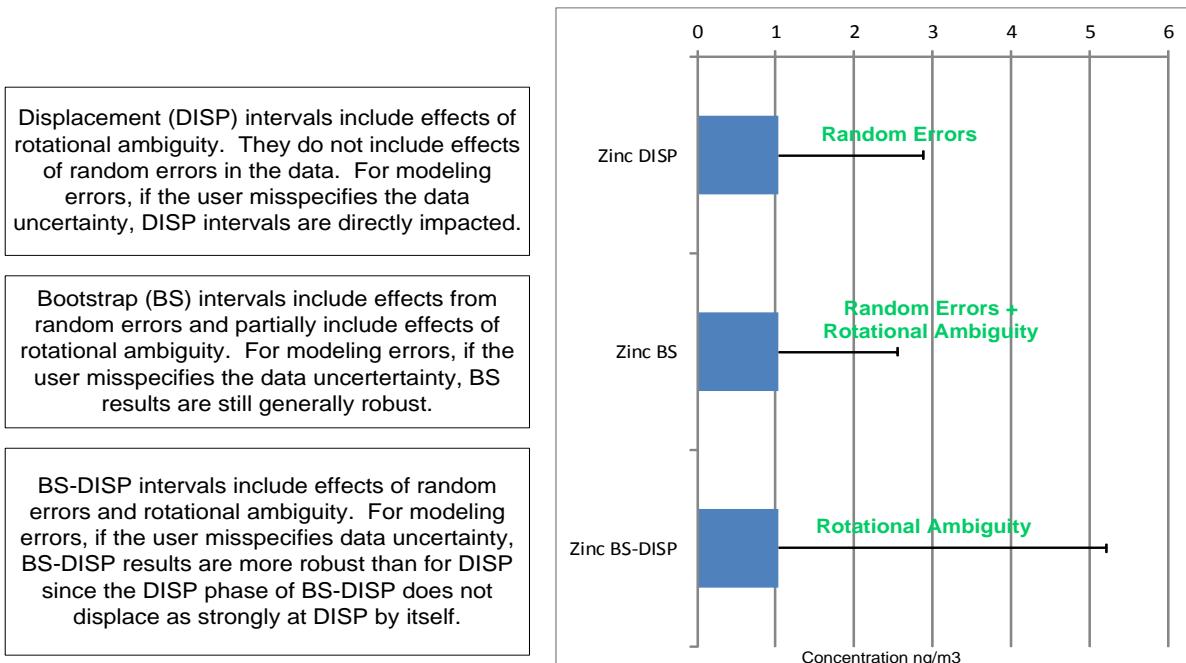
The factor names are used to indicate specific solutions in the tools for assessing model results. Users can input their own factor names, which will replace the defaults in the Factor Names table and be saved in the configuration file. The user can also set a unique factor name for all the base runs by inputting the name in one cell and then pressing the "Apply to All Runs" button; update factors names in the profile and contribution files by pressing the "Update Diag Files" button; or reload the default factor names into the Factor Names table by pressing "Reset to Defaults."

It should be noted that, if the user loads an existing configuration file with user-defined factor names and initiates base model runs with random seeds, the factor order in the run solutions may change. In this case, the GUI will generate a pop-up warning to remind the user to verify that previous factor names are appropriate.

Short descriptions of the error estimation methods available in PMF are shown in Figure 24 along with the example base factor concentration (blue) and upper error limits for the three methods. The upper error estimate for BS is the lowest for the zinc source and the estimates increase for the DISP and BS-DISP. Random errors are estimated with the BS method described in this section. Also, the Methods for Estimating Uncertainty in Factor Analytic Solutions paper (Paatero et al., 2014) provides a detailed description of the PMF error estimation methods.

Figure 23. Example of the Base Model Runs screen with default base model run factor names.

Displacement (DISP) intervals include effects of rotational ambiguity. They do not include effects of random errors in the data. For modeling errors, if the user misspecifies the data uncertainty, DISP intervals are directly impacted.

Bootstrap (BS) intervals include effects from random errors and partially include effects of rotational ambiguity. For modeling errors, if the user misspecifies the data uncertertainty, BS results are still generally robust.

BS-DISP intervals include effects of random errors and rotational ambiguity. For modeling errors, if the user misspecifies data uncertainty, BS-DISP results are more robust than for DISP since the DISP phase of BS-DISP does not displace as strongly at DISP by itself.

Figure 24. Comparison of upper error estimates for zinc source.

5.7   Base Model Displacement Error Estimation
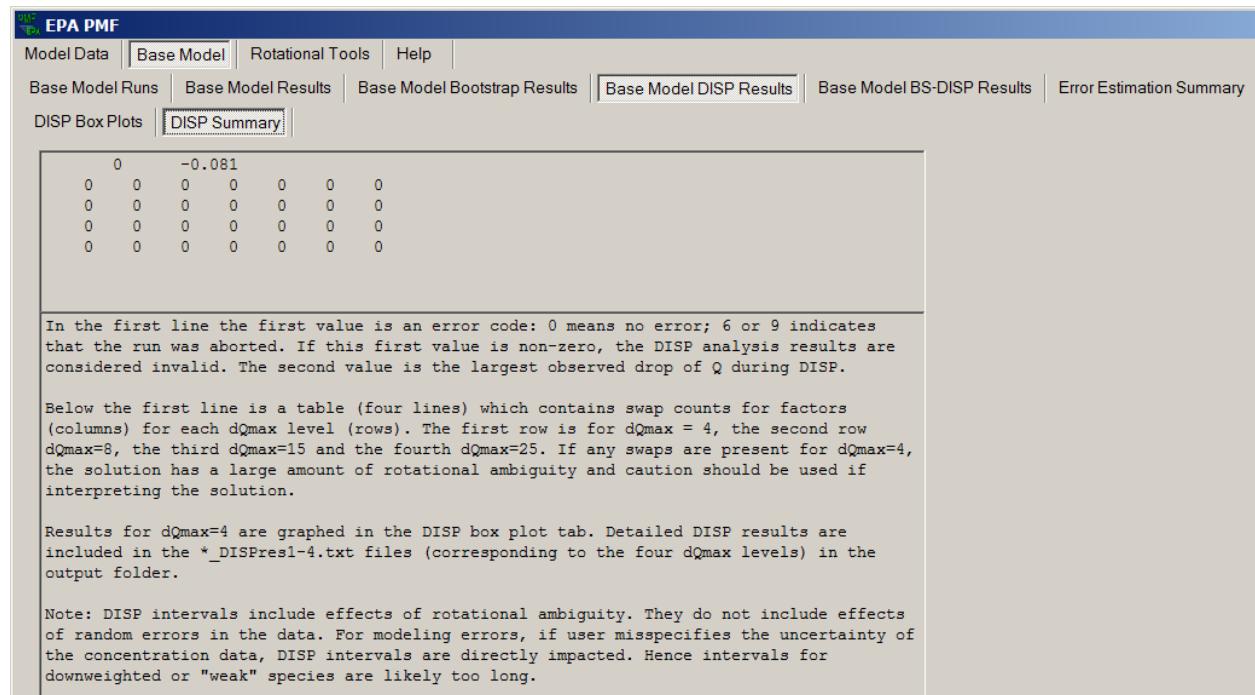
The DISP explicitly explores the rotational ambiguity in a PMF solution by assessing the largest range of source profile values without an appreciable increase in the Q-value. The DISP Error Estimation can be run without running BS or can be run after BS and BS-DISP (discussed in Sections 5.8 and 5.9, respectively). For the solution chosen by the user, each value in the factor profile is first adjusted up and down and then all other values are computed to achieve the associated PMF (convergence to a Q-minimum). It is important to note that the newly computed minimum Q-value (modified) may be different from the Q-value associated with the unadjusted solution (base). The adjustment in factor profile values (up and down) is always the maximum allowable, with the constraint that the difference (dQ = base - modified) because of this adjustment is no greater than the dQmax (dQ <= dQmax). The model generates results for the following dQMax values: 4, 8, 15, and 25. For each dQmax value, DISP is executed and intervals (minimum and maximum source profile values) are summarized for each element in each factor profile. For example, if 20 species are in a data set and a 7-factor model has been fitted, then the DISP method will estimate 20 x 7 = 140 intervals for each dQmax value.

Simulations indicate dQmax values of 4 and 8 provide the smallest error ranges with the least number of base factor values outside the range. EPA PMF provides results for all dQmax, but plots are only shown for dQmax of 4 because this should provide robust intervals for nearly all data sets. DISP intervals may be calculated for both the base model solutions and base model solutions with added constraints. Press the "Run" button in the Base Model Displacement Method box to start DISP.

The DISP output is shown in Figure 25, along with guidance on interpreting the output. When the DISP method is completed, two output files (*_DISPest.dat and *_DISP.txt) are saved in the directory specified in the Output Folder box in the Data Files screen. The .dat file is in a concise format most usable by software and is not intended for users to view; there are no labels in this file, only numbers. The .txt file is a very large text file with details about the models fitted and the resulting DISP intervals.

Four files are output from DISP, one for each dQmax used, and the user-provided output file prefix is placed at the start of the file name and is denoted in this user guide as an asterisk (*) (dQmax=4, 8, 16, 32; * _DISPres1, *_DISPres2, * _DISPres3, *_DISPres4). In each file, there is a line with two numbers, followed by four lines of data. In the first line, the first value is an error code: 0 means no error; 6 or 9 indicates that the run was aborted. If this first value is non-zero, the DISP analysis results are considered invalid. The second value is the largest observed drop of Q during DISP.

Below the first line is a four-line table that contains swap counts for factors (columns) for each dQmax level (rows). The first row is for dQmax = 4, the second row dQmax=8, the third dQmax=15 and the fourth dQmax=25. The swap counts are a key indicator of the stability of a PMF solution and swaps at dQmax = 4 or the first row in the table indicate that the solution should not be interpreted.

Figure 25. Example of the Base Model Displacement Summary screen.

If factor swaps occur for the smallest dQmax, it indicates that there is significant rotational ambiguity and that the solution is not sufficiently robust to be used. If the decrease in Q is greater than 1%, it likely is the case that no DISP results should be published unless DISP analysis is redone after finding the true global minimum of Q. To improve the solution, the number of factors could be reduced, marginal species could be excluded, or unusual events in time series plots could be excluded.

Below these diagnostics in the *_DISPresX data files are four blocks of data, where each column is a factor and each row a species: (1) the profile matrix upper bound, in concentration units; (2) the profile matrix lower bound, in concentration units; (3) the profile matrix upper bound, in % species units; (4) the profile matrix lower bound, in % species. The DISPPres files are output directly from ME and are for users who want to process the output. The DISP results for a dQmax of 4 are summarized in an easy-to-use file: *_ErrorEstimationSummary.

5.8   Base Model BS Error Estimation

BS is used to detect and estimate disproportionate effects of a small set of observations on the solution and also, to a lesser extent, effects of rotational ambiguity. BS data sets are constructed by randomly sampling blocks of observations from the original data set. The block length depends on the data set and is chosen so that each BS data set preserves the underlying serial correlation that may be present in the base data set. Blocks of observations are randomly selected until the BS data set is the same size as the original input data.

A number of BS data sets (e.g., 100) are then processed with PMF, and for each BS run, the BS factors are compared with the base run factors using the following method: the BS factor is mapped to the base factor with which the BS factor contribution has the highest correlation (and above a user-specified threshold). If no base factors correlate above the threshold for a given BS factor, that factor is considered "unmapped." This process is repeated for as many BS runs as the user specifies. There can be instances when multiple BS factors from the same run may be mapped to the same base factor.
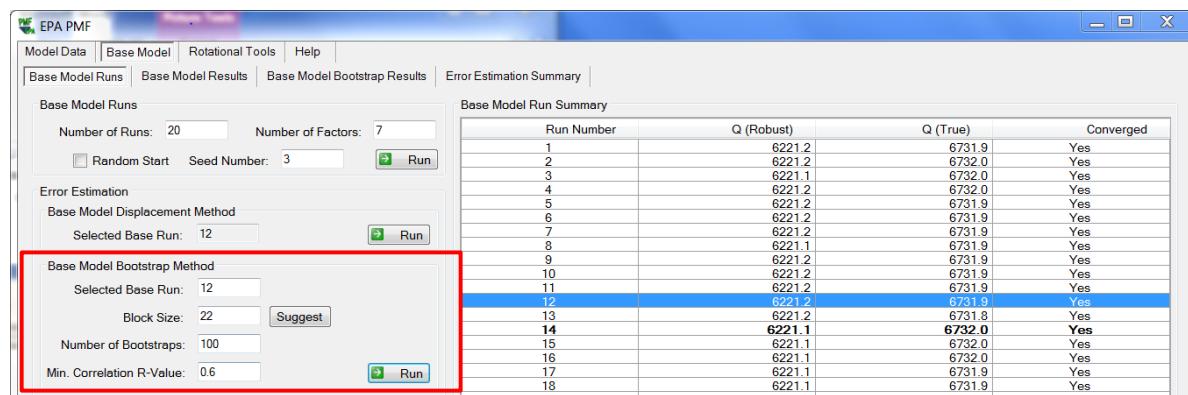
EPA PMF then summarizes all the bootstrapping runs. The user should examine the BS results to determine if the base run (blue square) is within the interquartile ranges (box) around the profiles. Species with their base run value outside of the interquartile range should be interpreted with caution because a small set of observations may have impacted the base run results or the species concentration in the factor could be insignificant. The mapping of BS factors to base factors will ideally be one-to-one. That is, factors from each BS run factor should match exactly one, and only one, base factor. However, it is likely that the presence (or absence) of a few critical observations can dramatically impact the BS factor profile. In such instances, the affected BS factors may closely match a particular base factor most of the times and some other base factor the rest of the time. In addition, specification of too many factors in the base model may also create a phantom factor. Any factor with approximately 80% or less mapping from the BS run should have the major contributing species in the profile investigated and further evaluation of the base model results should be done with the BS-DISP and DISP error estimation methods.

Initiating BSRuns

Bootstrapping captures the error associated with random errors and it is initiated under the Base Model tab, in the Base Model Runs screen (Figure 26, red box). As with the base runs, the user must make multiple choices prior to initiating the BS runs:

- **Base Run** – the base run to be used to map each BS run. The base run with the lowest Q(robust) is automatically provided; the user can enter another run number.

- **Block Size** – the number of samples that will be selected in each step of resampling. For example, a block size of three means that each BS block will comprise three samples from the input data set (i.e., samples 8-10 could be one block). The default block size is calculated according to Politis and White (2003), but can be overridden by the user. If the default has been overridden, the user can press the "Suggest" button to restore the default value.

- **Number of Bootstraps** – the number of BS runs to be performed. It is recommended that 100 BS runs be performed to ensure the robustness of the statistics; for preliminary analysis, 50 BS runs may be performed to quickly gauge the stability of a solution. A minimum of 20 BS runs are required.

- **Minimum Correlation R-Value** – the minimum Pearson correlation coefficient that will be used in the assignment of a BS run factor to a base run factor. The default value is 0.6. If a large number of factors are unmapped, the user may want to investigate the impact of lowering the R-value. This change should be reported with the final solution.

After all input parameters have been entered, the BS runs can be initiated by pressing the "Run" button inside the Base Model Bootstrap Method box. As with the base runs, the user can interrupt the runs by pressing the "Stop" button in the lower right corner of the Base Model Runs screen. No outputs will be saved or overwritten if the run is interrupted.

Figure 26. Example of the Base Model Runs screen highlighting the Base Model Bootstrap Method box.

5.8.1　Summary of BS Runs

A summary of base model BS runs is presented in the Base Bootstrap Summary screen under the Base Model Bootstrap Results tab (Figure 27), which appears only after the BS has been run. The first eight lines in this screen contain all the input parameters for bootstrapping, as specified by the user in the Base Model Runs screen. The summary screen also includes several tables that summarize the BS run results. The first table is a matrix of how many BS factors were matched to each base factor. The next table shows the minimum, maximum, median, and 25th and 75th percentiles of the Q(robust) values. The rest of the summary is the variability in each factor profile, also given as the mean, standard deviation, 5th percentile, 25th percentile, median, 75th percentile, and 95th percentile, using weighted average percentiles (see equation 5-2). The base run of each profile is included as the first column for reference, as is a column indicating if the base run profile is within the interquartile range of the BS run profiles.

EPA PMF also calculates the Discrete Difference Percentiles (DDP) associated with the BS runs and reports these values in the Base Bootstrap Summary screen. This method estimates the 90th and 95th percentile confidence intervals (CI) around the base run profile, reported as percentages. The DDP is calculated by taking the 90th and 95th percentiles of the absolute differences between the base run and the BS runs for each species in each profile and expressing it as a percentage of the base run value. If the DDP percent is greater than 999, a "+" is displayed on screen. The original value is saved in the output files (*_diag and *_boot). If the base run value for a species is zero, it is not possible to calculate the DDP; in these cases, an asterisk (*) is displayed. The DDP values can be used for reporting the BS error estimates.

For this example, the base and boot factors are matched except for three factors with three runs that were mapped to factor 7. The crustal (factor 4) and motor vehicle (factor 7) contain crustal

elements and the steel source also was mapped to three other sources, which could be due to BS not creating a data set with all of the samples with high steel production impacts. The total number of mapped factors may also not add up to the number of BS runs if the boot factor run did not converge. Mapping over 80% of the factors indicates that the BS uncertainties can be interpreted and the number of factors may be appropriate.
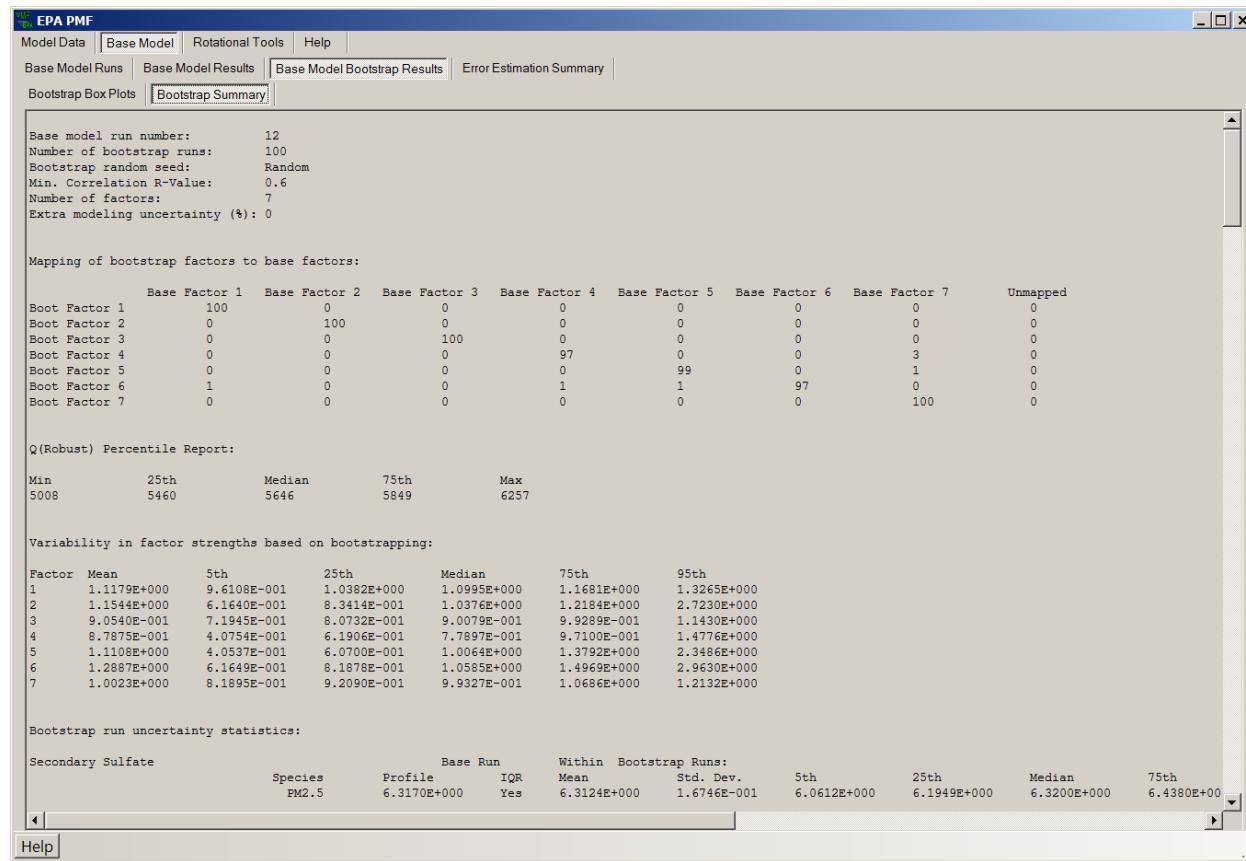
Figure 27. Example of the Base Bootstrap Summary screen.
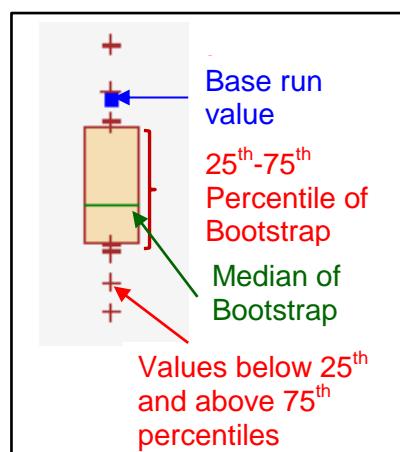
5.8.2   Base Bootstrap Box Plots

The variability in BS runs is shown graphically in the Base Bootstrap Box Plots screen (Figure 28). Two graphs are presented: the variability in the percentage of each species (Figure 28, 1) and the variability in the concentration of each species (Figure 28, 2), which corresponds to the Variability in Factor Profiles table in the Base Bootstrap Summary screen. In both box plots, the box (Figure 29) shows the interquartile range (25th–75th percentile) of the BS runs. The horizontal green line represents the median BS run and the red crosses represent values outside the interquartile range. The base run is shown as a blue box for reference. Values outside of the interquartile range are shown as red crosses. At the bottom of this screen, the base run numbers are grayed out and not selectable; however, the base run used for bootstrapping is highlighted in orange. The user can select the factor they want to view by clicking on the factor number across the bottom of the screen. The Variability in Concentration

of Species is shown in the bottom plot. Species with the base run profile value (blue box) outside interquartile range (tan box) should be interpreted only after evaluating the two additional error estimation results in PMF. These species have influential BS observations that biased either the base or BS runs; the DISP and BS-DISP will provide more reliable error estimates.

Figure 28. Example of the Base Bootstrap Box Plots screen.

Figure 29. Diagram of box plot.

5.9   Base Model BS-DISP Error Estimation

BS-DISP estimates the errors associated with both random and rotational ambiguity and it is run from the Error Estimation section of the Base Model Runs screen. BS-DISP may take many hours to run due to the number of combinations that are evaluated, so it is recommended that the user evaluate the BS-DISP results first with less than 100 BS runs (50 is recommended); for final BS-DISP results, use 100 BS runs.

BS-DISP is a combination of BS and the DISP method. The BS Error Estimation must be run before BS-DISP because each BS resample undergoes a DISP analysis so that error limits are found for all F (profile) factor elements. This process may be viewed as follows: each DISP defines the span of rotationally accessible space. Each BS resample moves this space around, randomly in different directions. Taken together, all the replications of the rotationally accessible space, in random locations, represent both the random uncertainty and the rotational uncertainty.
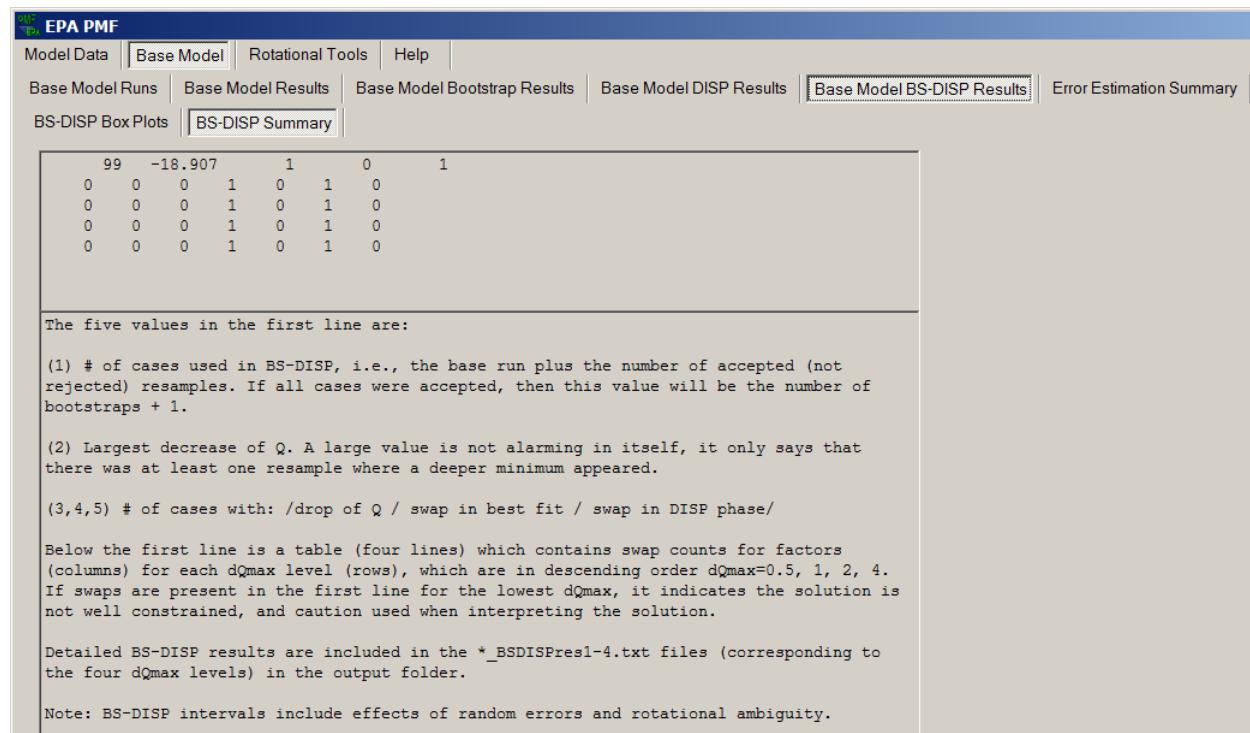
The limits obtained by displacing a factor element include both rotational ambiguity and variability due to input data uncertainty. To speed up computation of BS-DISP, it is suggested that only a small subset of all F factor elements are adjusted. Downweighted variables create a special problem in DISP computations. If such variables are adjusted, the error intervals can be very large (based on simulated data evaluations). The error estimates for downweighted species are best estimated from the results obtained from adjusting non-downweighted species.

BS-DISP provides the change in Q associated with the displacement. Occasionally, it is seen that displacements cause a significant decrease of Q, typically by tens or by hundreds of units. If such a decrease occurs in DISP or BS-DISP, it means that the base case solution was in fact not a global minimum, although it was assumed to be such. The value associated with a significant change in Q is still being evaluated, but the initial guidance is that a change in Q greater than 1% is significant. If the change in Q is greater than 0.5%, it is recommended to increase the number of Base Model runs to 40 to find a global minima.

A key output from DISP and BS-DISP analyses is the extent of factor swapping, usually resulting from a "not-well-defined" solution (i.e., a solution where factor identities are fluid). A sample BS-DISP output is shown in Figure 30 along with guidance on interpreting the output. Starting from the most plausible solution, it is possible to transform the solution gradually, without significant increase of Q, so that factor identities change. In the extreme case, factors may change so much that they exchange identities. This is called factor swap. Physically, a solution with swapped factors represents the same physical model as the original solution. However, the presence of factor swaps means that all those intermediate solutions also exist and must be considered as alternative solutions.

For a higher dQmax, a larger uncertainty interval or CI is usually obtained. The larger the interval, the higher the chance that it contains the true unknown value. CI is displayed along with the profile values in the BS-DISP Box Plots tab. The dQmax values are still being evaluated and a dQmax of 4 for DISP and 0.5 for BS-DISP provide lower bounds for the true uncertainty estimates if the input data uncertainties are reasonable. Smaller dQmax values are used in BS-DISP versus DISP because the combination of bootstrapping and DISP should

capture nearly all the uncertainty within the solution. All dQmax values should be evaluated to determine whether the solution is well-defined.

Figure 30. Example of the Base Model BS-DISP Summary screen.

Sample results from the BS-DISP Summary tab are shown in Figure 30 after using key species from each of the sources (sulfate, potassium ion, total nitrate, silicon, zinc, iron, and EC).

The BS-DISP results in Figure 30 show that the solution does not have significant rotational ambiguity and the base model and error estimates can be interpreted. Having no swaps at all, dQmax provides confidence that the solution is well constrained and the BS-DISP results can be reported.

If factor swaps are produced at dQmax = 0.5, then the number of factors in the solution and BS and DISP results need to be evaluated before reporting the BS-DISP results. Because the BS-DISP is a combination of BS and DISP, it is suggested that the results of each component be evaluated to understand what might be causing the swaps. Steps to reduce the number of swaps include reducing the number of factors and adding constraints.

Four files are output from BS-DISP, one for each dQmax used; the user-provided output file prefix is placed at the start of the file name and is denoted in this user guide as an asterisk (*) (dQmax=0.5, 1, 2, 4; * _BSDISPres1, *_BSDISPres2, * _BSDISPres3, *_BSDISPres4). These contain the same summary diagnostics that are provided in the BS-DISP Summary tab. The five values in the first line of diagnostics that are displayed within the EPA PMF program are:

1. *k,* the number of cases in the file. This includes both the full-data case and the accepted (not rejected) resamples; if all bootstrap cases were accepted, this value would be equal to one plus the number of bootstraps (the extra one run is an initialization run). If no cases were excluded, *k* should be equal to the number of bootstraps times the number of factors times the number of species selected for BS-DISP.

2. Largest decrease of Q. A large value is not necessarily alarming, but it indicates that there was at least one resample where a deeper minimum appeared. A large value for a decrease in Q is approximately 1% or more of Q(robust); more testing is required to provide better guidance on this value.

3. Number of cases with drop of Q.

4. Number of cases with swap in best fit.

5. Number of cases with swap in DISP.

Below the first line of diagnostics in the BS-DISP summary is a four-line table that contains swap counts for factors (columns) for each dQmax level (rows), which are in ascending order (dQmax=0.5, 1, 2, 4). In the best case, all of the swaps are zero; however, the probability of creating a BS data set that results in a swap is based on the data characteristics (i.e. peaks), the number of BS runs, and the number of factors. The profiles and DISP results should be evaluated to determine whether there is a reason for the swaps. A result with swaps between two factors is more reliable than swaps occurring across many factors. For this example, the swaps are occurring between the crustal (factor 4) and steel production (factor 6), which have many common elements. Also, the number of swaps is one for two factors, which indicates some ambiguity between the factors.

The output files from BS-DISP contain many blocks of data following the diagnostics shown in Figure 30. The first two blocks of data are the initial run data, with each row representing a species and each column a factor. The last line of each block is always a series of "1"s as a placeholder. There are four blocks of data for each BS resample: (1) profile matrix for BS resample #1 after displacing down, in concentration units; (2) profile matrix for BS resample #1 after displacing up, in concentration units; (3) profile matrix for BS resample #1 after displacing down, in % species; (4) profile matrix for BS resample #1 after displacing up, in % species. These four blocks are then repeated for each BS resample. The BSDISPPres files are output directly from ME and are for users who want to process the output. The BS-DISP results for a dQmax of 0.5 are summarized in an easy to use file: *_BaseErrorEstimationSummary.

5.10   Interpreting Error Estimate Results

A comprehensive set of error estimates are available and the results are added to the summary files for easy use after running each error estimation method (*_BaseErrorEstimationSummary, *_FpeakErrorEstimationSummary, *_ConstrainedErrorEstimationSummary). The summary files contain the species and diagnostics as well as the error estimates by factor for concentrations, percent of species sum, and percent of total variable if one is selected.

The error estimation information is summarized in the *_BaseErrorEstimationSummary file and the following figure after each error estimation method is run. The

*_BaseErrorEstimationSummary file has a useful summary of the factor error estimates: Base Value, BS 5th, BS Median, BS 95th, BS-DISP 5th, BS-DISP Average, BS-DISP 95th, DISP Min, DISP Average, and DISP Max. Figure 31 shows the error estimation summary plot for the three error estimates.
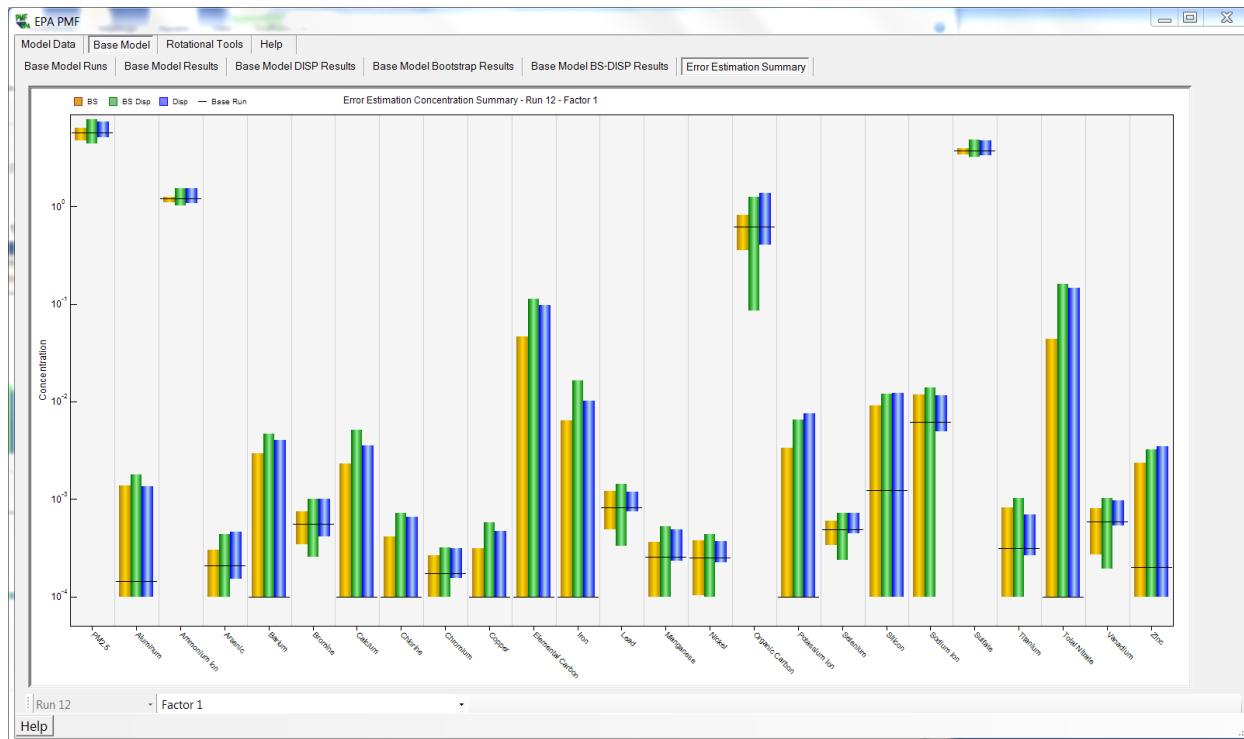
Figure 31. Error estimation summary plot.

6.   Rotational Tools

In general, the non-negativity constraint alone is not sufficient to produce a unique solution. An infinite number of plausible solutions may be generated and cannot be simply disqualified using mathematical algorithms. Rotating a given solution and evaluating how the rotated results fill the solution space is one approach to reduce the number of solutions. Additional information, such as known source contributions and/or source compositions, can also be used to reduce the number of solutions and to determine whether one solution is more physically realistic than other solutions.

Mathematically, a pair of factor matrices (*G* and *F*) that can be transformed to another pair of matrices (*G** and *F**) with the same Q-value is said to be "rotated." The transformation takes place as shown in Equation 6-1:

$$G^* = GT \ \text{ and } \ F^* = T^{-1}F \tag{6-1}$$

The T-matrix is a *p* x *p*, non-singular matrix, where *p* is the number of factors. In PMF, this is not strictly a rotation but rather a linear transformation of the *G* and *F* matrices. Due to the non-negativity constraints in PMF, a pure rotation (i.e., a specific T-matrix) is only possible if none of the elements of the new matrices are less than zero. If no rotation is possible, the solution is unique. Therefore, approximate rotations that allow some increase in the Q-value and prevent any elements in the solution from becoming negative are useful in PMF.
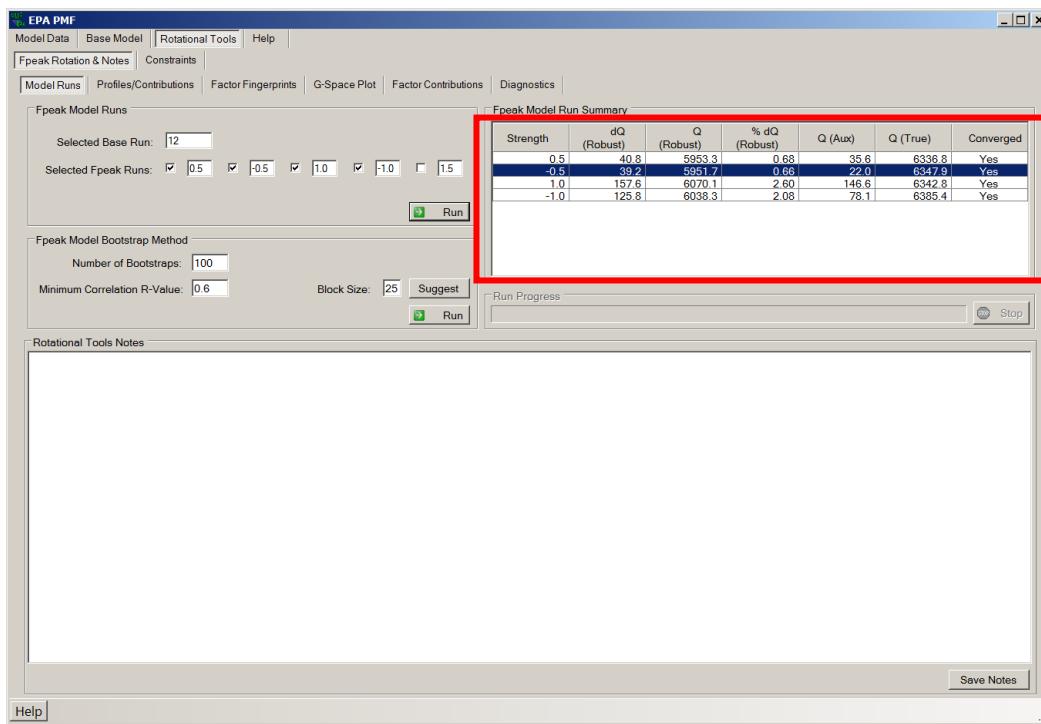
For some solutions, the non-negativity constraint is enough to ensure that there is little rotational ambiguity in a solution. If there are a sufficient number of zero values in the profiles (F-matrix) and contributions (G-matrix) of a solution, the solution will not rotate away from the "real" solution. However, in many cases, the non-negativity constraint is not sufficient to prevent rotation away from the "real" solution. To help determine whether an optimal solution has been found, the user should inspect the G-space plots for selected pairs of factors in the original solution. The current guidance is to select a regional source type such as coal-fired power plants (sulfate) and plot it against local industrial sources such as steel production (Fe).

6.1   Fpeak Model Run Specification

After evaluating the base run BS error estimates, the rotations should be explored. Fpeak runs are initiated by selecting "Rotational Tools," "Fpeak Rotation & Notes," and "Fpeak Model Runs." The base run with the lowest Q(robust) is automatically selected by the program as the run for Fpeak runs; this can be overridden by the user in the "Selected Base Run" box. The user can perform up to five Fpeak runs by checking the appropriate number of boxes and entering the desired strength of each Fpeak run. While there are no limits on the values that can be entered as Fpeak strengths (under "Selected Fpeak Runs"), generally values between -5 and 5 should be explored first. Positive Fpeak values sharpen the F-matrix and smear the G-matrix; negative Fpeak values smear the F-matrix and sharpen the G-matrix. More details on positive and negative Fpeak values can be found in Paatero (2000). The Fpeak strengths in ME-2 are not the same as those in PMF2; values of around five times the PMF2 values are

needed to produce comparable results in ME-2. Additionally, an Fpeak value of 0 is not allowed; EPA PMF will give the user an error message if 0 is entered in any Fpeak strength box.

Fpeak runs begin when the user presses the "Run" button on the Fpeak Model Runs screen. Base run and BS run results will not be lost when Fpeak is run. After the Fpeak runs are completed, a summary of the Fpeak results, with the same information contained in the Base Model Run Summary table, is shown in the Fpeak Model Run Summary table (Figure 32, red box). Additional results are displayed in: Fpeak Profiles/Contributions, Fpeak Factor Fingerprints, Fpeak G-Space Plot, Fpeak Factor Contributions, and Fpeak Diagnostics; these results should be used as a reference when evaluating the Fpeak runs. Fpeak is useful for examining the span of possible rotations, with an end result of more values at or near 0 in either the contributions or profiles, depending on whether a positive or negative Fpeak is used. Thus DISP and BS-DISP with Fpeak forcing will yield shorter EE intervals, potentially leading to incorrect interpretation of a solution.
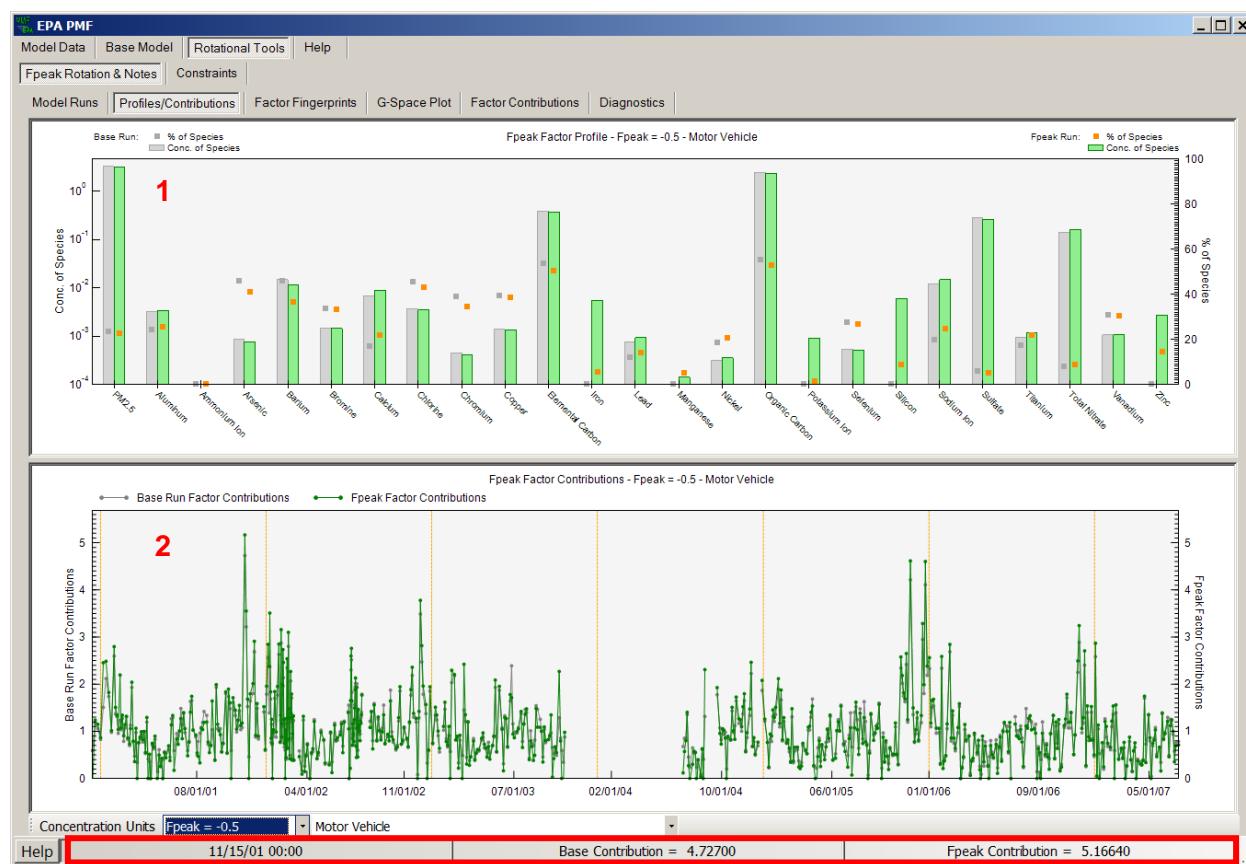
Figure 32. Example of the Fpeak Model Run Summary in the Fpeak Model Runs screen.

6.1.1   Fpeak Results

The Fpeak Profiles/Contributions screen presents profile (Figure 33, 1) and contribution (Figure 33, 2) plots for Fpeak runs (by Fpeak strength value and factor) and for the selected base run. In the profile graph, the concentration of species (left y-axis) is a green bar and the percent of species (right y-axis) is an orange box. For comparison, the original base run results are also displayed on the profile graph. The mass of the species (left y-axis) is a light gray bar and the percent of species (right y-axis) is a dark gray box. The contribution graph presents the time

series of factor contributions. Factor contributions for the base model results are also displayed (gray line). The Fpeak values are in the same order as entered on the Fpeak Model Runs screen; the factors are in the same order as those in Base Model Results. In these graphs, users should look for deviations (i.e., increases or decreases in a particular species in a factor) among Fpeak values and with the corresponding base run results. Users can select an Fpeak value and factor number by clicking on the desired number at the bottom of the screen. The status bar (Figure 33, red box) in the Fpeak Profiles/Contributions screen displays the date and contribution of data points closest to the mouse position on the contribution graph. The status bar displays the date, concentration, total variable selected, and the species factor as they are moused over on the Factor Contributions plot. If no mass from the total variable is apportioned to the factor, the graph is not shown and the GUI instead displays, "Total Variable mass is 0 for this run/factor."
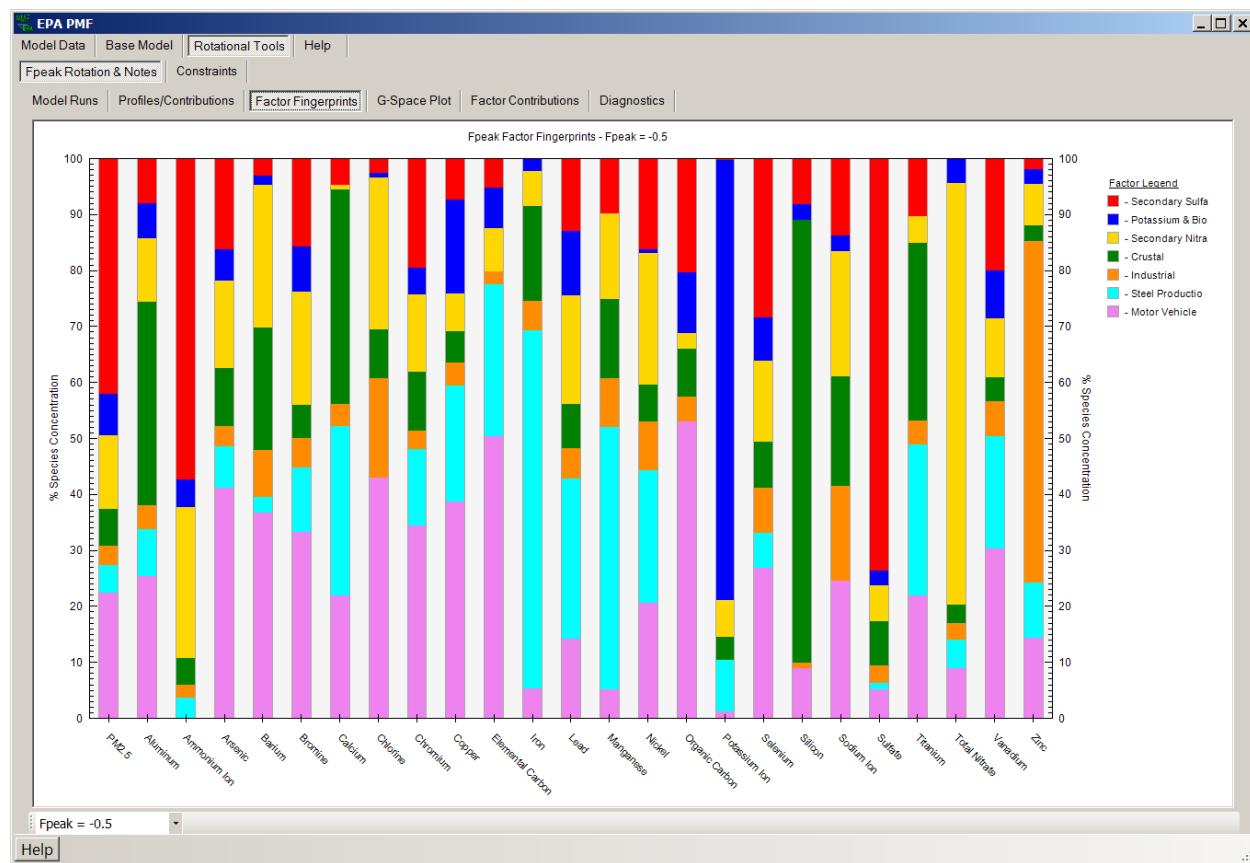
Figure 33. Example of the Fpeak Profiles/Contributions screen.

Fpeak Factor Fingerprints

The Fpeak Factor Fingerprints screen shows the concentration (in percent) of each species contributing to each factor as a stacked bar chart (Figure 34). This plot can be used to verify unique factor names and determine the distribution of the factors for individual species. Users should look for deviations (i.e., increases or decreases in a particular species in a factor) among
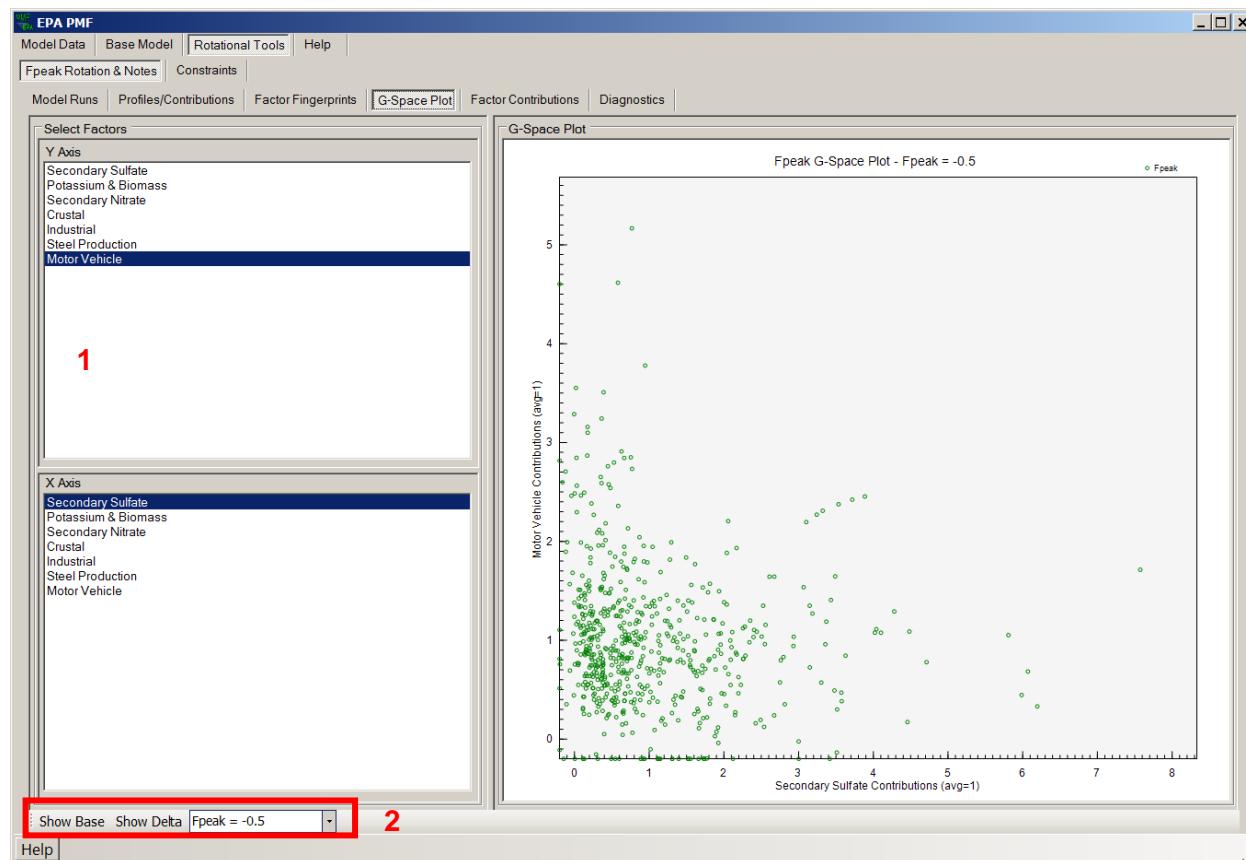
Fpeak values and the corresponding base run results. The user can select an Fpeak value by clicking on the desired number at the bottom of the screen.

Figure 34. Example of the Fpeak Factor Fingerprints screen.

Fpeak G-Space Plot

As in the Base Model Results screen, the Fpeak G-Space Plot screen shows a scatter plot of factors. The user assigns a factor to the x- and y-axes by selecting the desired factor from the lists on the left of the screen (Figure 35, 1). The Fpeak value to display, the base run G-space plot ("Show Base"), and the delta in G-space plots between the base run and an Fpeak run ("Show Delta") are selected at the bottom of the screen (Figure 35, 2). When an Fpeak value is selected in either the Fpeak Profiles/Contributions screen or the Fpeak G-Space Plot screen, it is automatically selected in the other screen. The user can also select a point in any Fpeak G-space plot by clicking on that point. The selected point will turn orange and the date and x-y values will be stored to the *_Fpeak_diag file. This feature helps the user identify and track rotations. For example, if a G-Space plot appears rotated, the user can mark the edge points. Using information such as meteorological conditions or emissions information, the user can determine whether these edge points are expected to have low contributions from the source.
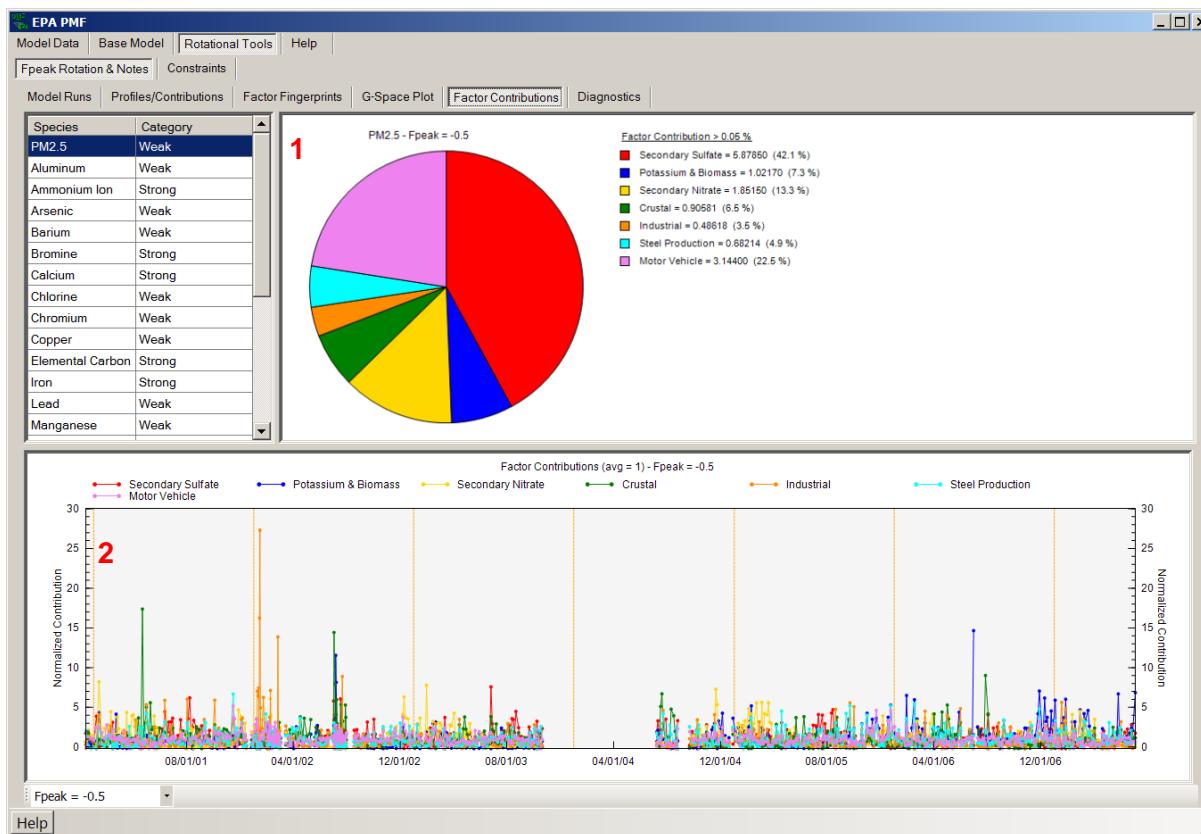
Figure 35. Example of the Fpeak G-Space Plot screen.

Fpeak Factor Contributions

The Fpeak Factor Contributions screen (Figure 36) shows two graphs. The top graph is a pie chart which displays the distribution of each species among the factors resolved by PMF (Figure 36, 1). The species of interest are selected from the table on the left of the screen; the categorization of that species is also displayed for reference. If a total variable was chosen by the user under the Concentration/Uncertainty screen, that variable is boldfaced in the table. The pie chart for the selected species appears on the right side of the screen. If the user has specified a total variable, the distribution of this variable across the factors will be of particular importance. The user may also want to examine the distribution of certain key species, such as toxic species, across factors. The bottom graph shows the contribution of all the factors to the total mass by sample (Figure 36, 2). The dotted orange reference lines denote January 1 of each year. The graph is normalized so that the average of all the contributions for each factor is 1.

Fpeak Diagnostics
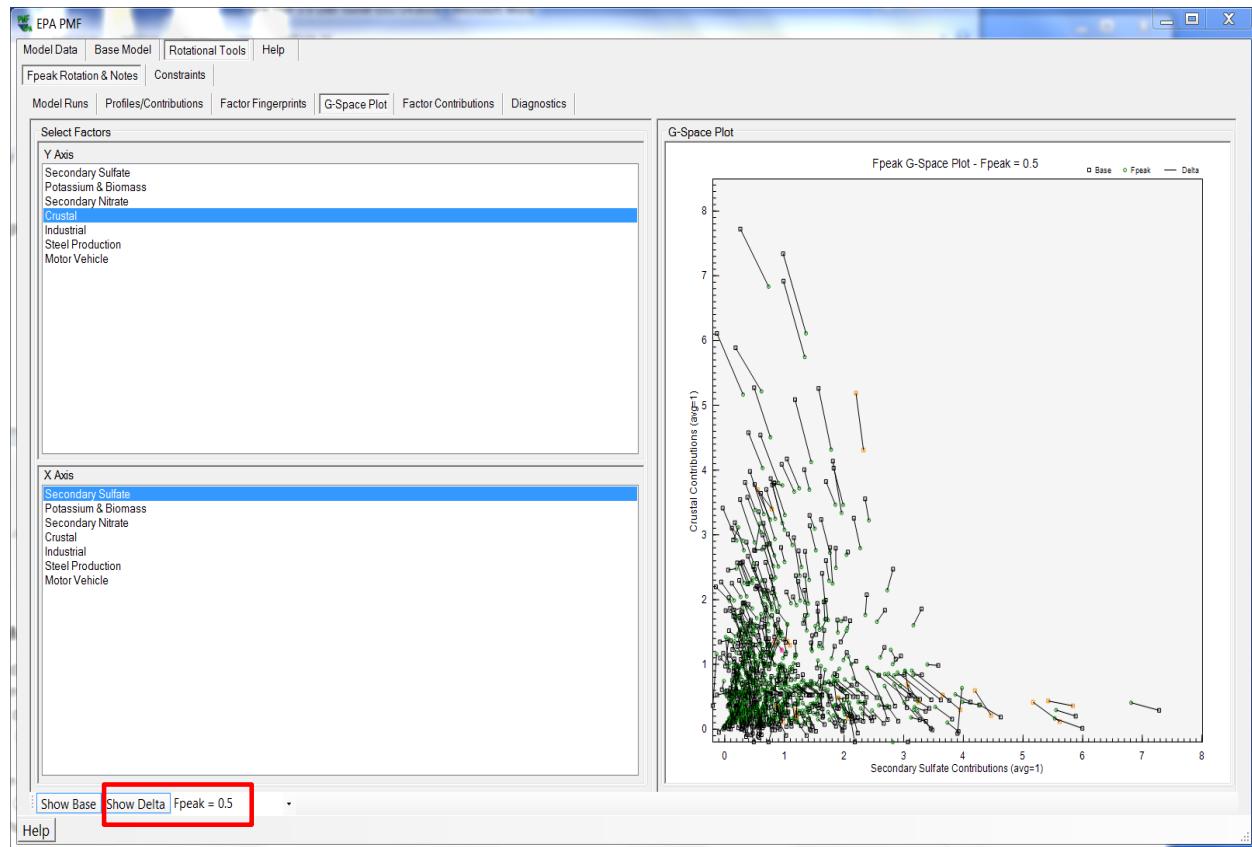
The Fpeak Diagnostics screen summarizes the Fpeak input parameters and output for reference (e.g., Fpeak run summary, factor profiles and contributions, and samples that are marked on the Fpeak G-space plot). All of the information on this screen is saved in *_Fpeak.

Figure 36. Example of the Fpeak Factor Contributions screen.

6.1.2   Evaluating Fpeak Results

Fpeak runs should be viewed by the user as a means of exploring the full space of the chosen PMF solution. Several aspects of the solution should be evaluated to understand how Fpeak changes the PMF solution. Users should first examine the *Q*-values of the Fpeak runs (available in the Fpeak Model Run Summary on the Fpeak Rotation & Notes → Fpeak Model Runs screen) to evaluate their increase from the base run *Q*-value. In a pure rotation, the *Q*-value would not change because the rotation is simply a linear transformation of the original solution. However, because of the non-negativity constraints of PMF, pure rotations are not usually possible and the rotations induced by Fpeak are approximate rotations, which change the *Q*-value. In general, an increase of the *Q*-value due to the Fpeak rotation with a dQ of less than 5% of the Base Run Q(robust) value is acceptable. Corresponding G-space plots of Fpeak solution factors should be examined to see if points move toward the axis or lower/zero contributions (Figure 37). Additionally, profiles and contributions should be examined to determine the impact of the rotation.

Figure 37. G-Space plot and delta between the base run contribution and Fpeak run contribution for each contribution point.

6.2   Constrained Model Operation

Source composition and contribution knowledge can be used to constrain a model run. For example, if a source is known to be inactive for a certain period, there should be no contributions from the factor that represents that source during the inactive time period. The contributions can be set to zero or pulled to zero and the penalty in Q is provided for moving the contribution from the optimal solution to one based on external knowledge. Another example is if a source profile from a nearby facility has been quantified, the user could constrain the profile in a factor that represents that facility type to match the measured profile. The amount of Q allowed for a constraint depends on the data set; however, 5% of Q(robust) is the current maximum that is recommended and PMF automatically calculates the amount of Q associated with a percent by entering a % dQ. Applications of using constraints are discussed in greater detail elsewhere (Norris et al., 2009; Paatero et al., 2002; Paatero and Hopke, 2008; Rizzo and Scheff, 2007).

6.2.1   Constrained Model Run Specification

The Constrained Model Runs screen is used to specify constraints associated with a variety of types of *a priori* information including: (a) creating constraints using the Expression Builder and

(b) specifying constraint points from the base model results and the constraints table. Starting with a selected base run, two types of constraints can be performed: (1) "hard pulling," which is imposed without regard to the change in the Q-value (e.g., a specific factor element in either the profile or the contribution matrix is set to zero, given a lower and upper limit, or fixed to its original value), or (2) "soft pulling," which has a limit of change allowed in the Q-value (e.g., an element or expression of elements is pulled up maximally, pulled down maximally, or pulled to a target value).

The Expression Builder has three radio buttons that users can select to define constraints as constant ratio (Figure 38), mass balance (Figure 39), or customized expression (Figure 40).

- **Ratio** (Figure 38) – Select a factor and two different species from the lists, and input the ratio in the "Value" text box.

- **Mass Balance** (Figure 39) – Select and add one or multiple factor-species into the text boxes on both sides of the equal sign under "Mass Balance" to set the balance equation. If needed, a number can be input into the "Coefficient" text box, which will be used as a coefficient for the species selected. Click the "Clear" buttons to remove the current specifications of the balance equation.

- **Custom** (Figure 40) – Specify a constraint by creating a customized equation. The customized equation can be based on either profiles (with species as element) or contributions (with sample as element). The custom equation must follow the same structure as the equations developed by the Expression Builder.

For each of the three Expression Builder functions, after the user defines a constraint and presses the "Add to Expressions" button, the corresponding equation in a standardized format will appear in the Expressions table (Figure 41, red box). Since the constraints defined using Expression Builder are "soft pulling," a limit of change in the Q-value must be specified. A default value (% dQ = 0.5) is provided in the Expressions table, which can be updated by users if needed. Users are also allowed to delete the selected constraints or all constraints by pressing the "Remove Selected Expressions" or "Remove All Expressions" buttons at the bottom of the Expressions table.

Source contributions can be constrained; the user can identify the points to be constrained in three graphs:

- On the Base Model → Base Model Results → Profiles/Contributions screen, left-click on the top graph to highlight a bar for the species to be constrained, then right-click the bar and select "Toggle Constraints" (Figure 42, 1).

- From the Base Model → Base Model Results → Profiles/Contributions screen, left-click on the bottom figure to select one data point or drag a square to select multiple data points, then right-click the data point and select "Toggle Constraints" (Figure 42, 2).

- From the Base Model → Base Model Results → Base G-Space Plot screen, left-click to select one data point or drag a square to select multiple data points, then right-click the data point(s) and select "Pull to X-Axis" or "Pull to Y-Axis" (Figure 43). The user can also select multiple data points pressing the CTRL button.
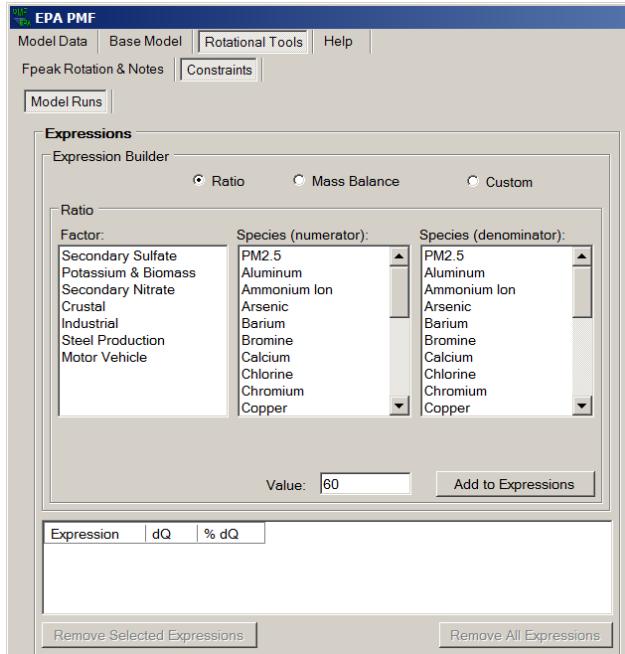
Figure 38. Expression Builder – Ratio.
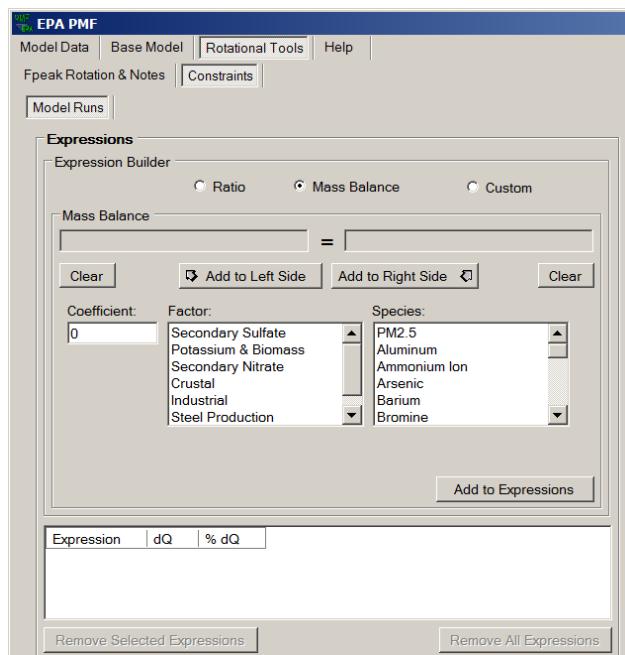
Figure 39. Expression Builder – Mass Balance.
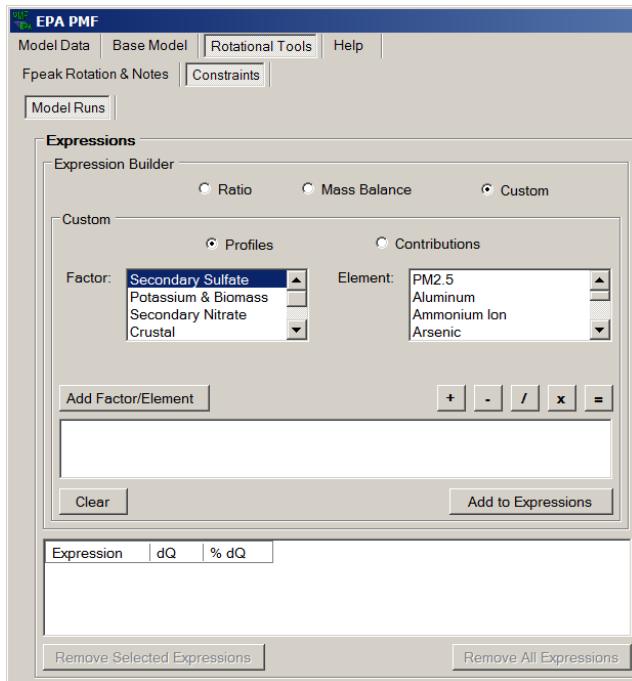
Figure 40. Expression Builder – Custom.
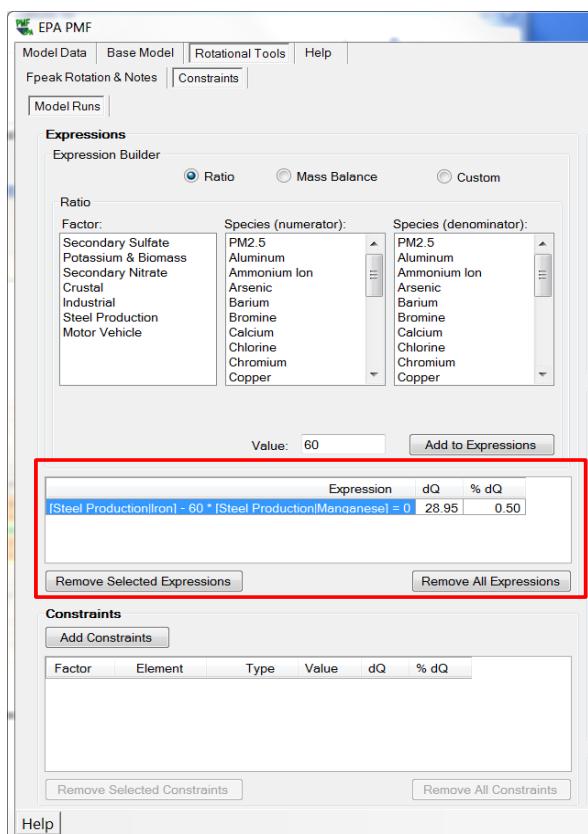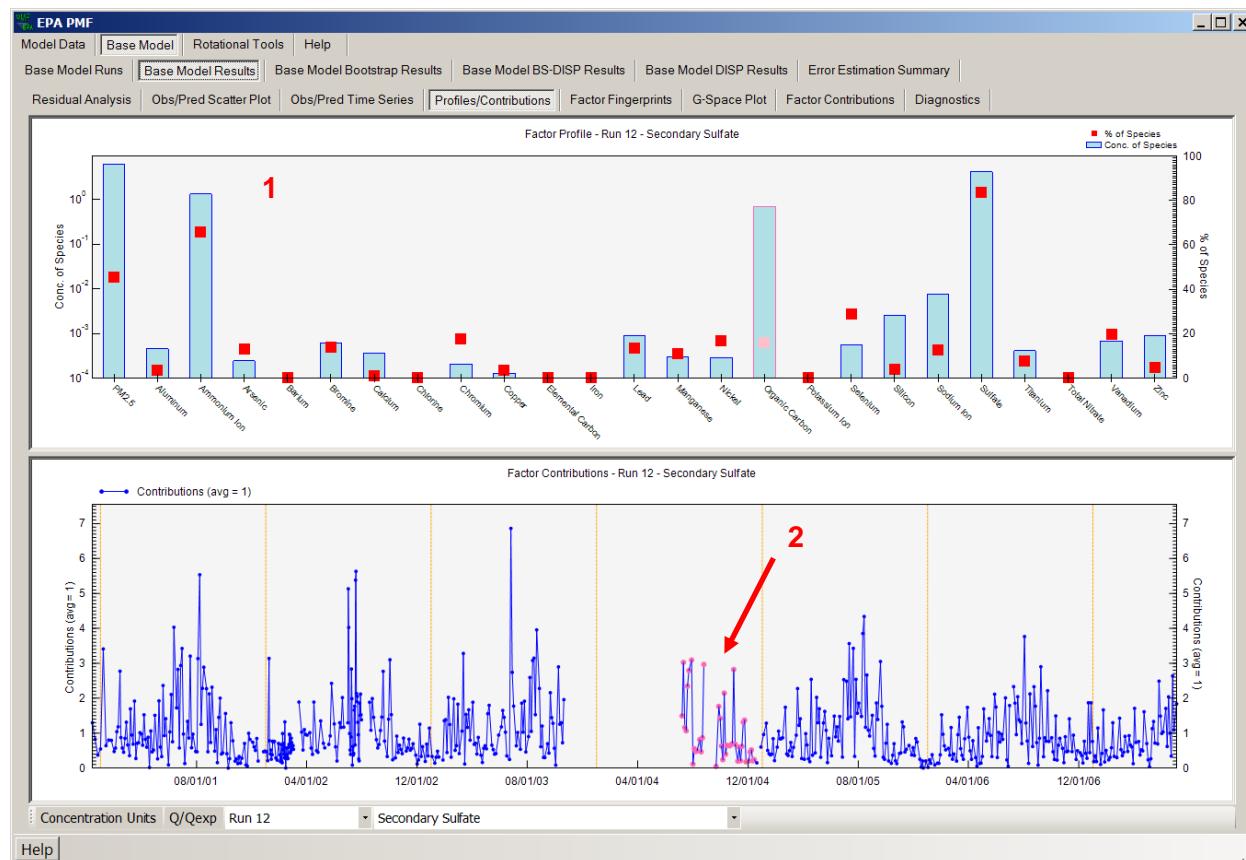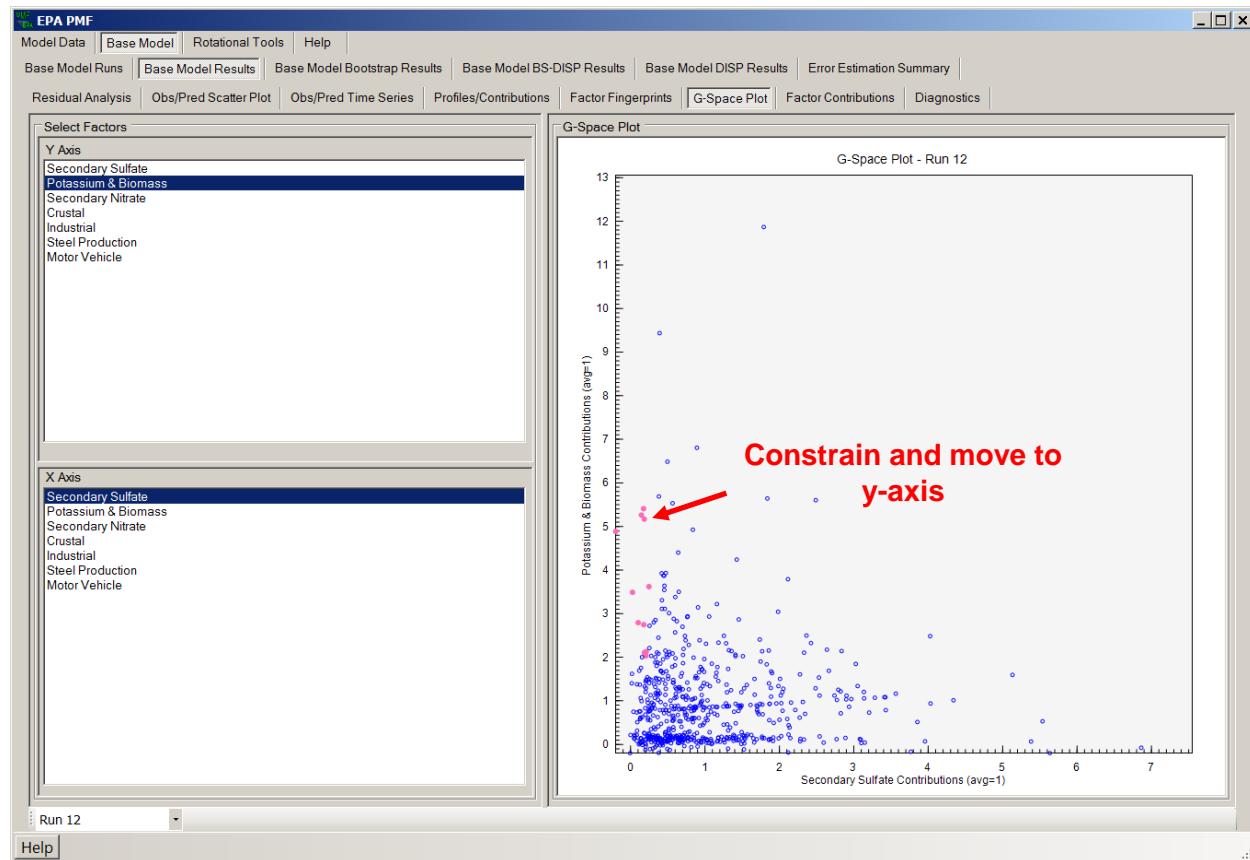
Figure 41. Example of expressions on the Constrained Model Runs screen.

Figure 42. Selecting constrained species and observations.
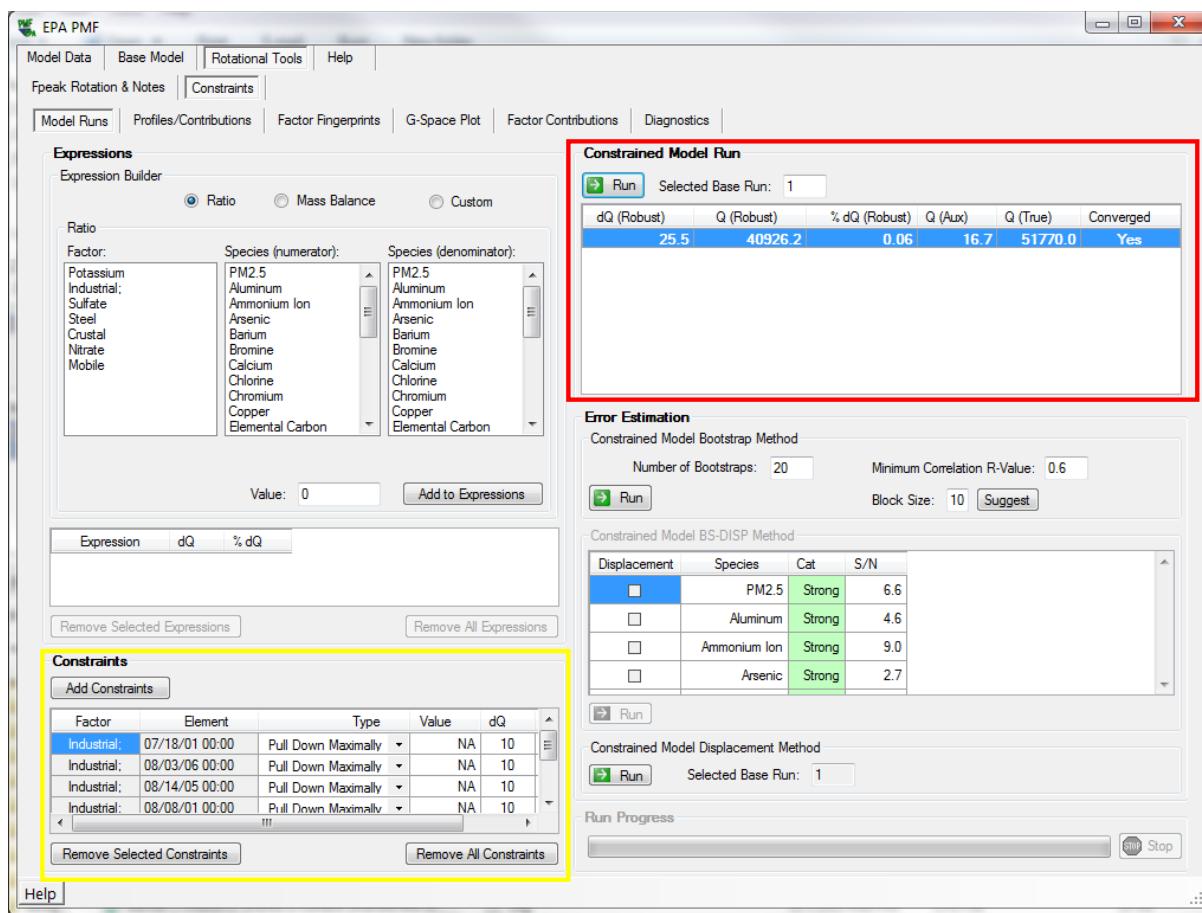
As discussed in Section 6.1.2, G-space plots in PMF solutions are evaluated to find edges that indicate rotational ambiguity and to determine if there are rotations in the solution. If users identify an edge in a G-space plot, constraints can be specified to pull the data points along the edge toward the axis (i.e., toward zero). The user should examine the points along the edge; if there is any *a priori* information that would indicate that a value should be zero (e.g., the source that the factor represents was inactive during a given time), the point should be pulled using the associated constraints. The strength of each pull is controlled by specifying a limit on the change in the Q-value. If the user wishes to perform a weak pull, a small limit on the change in the Q-value should be allowed. Conversely, if the user wishes to perform a strong pull, a large limit on the change in Q-value should be allowed. The strength of the pull should be based on *a priori* information about the pollutant sources that indicate that the contribution for the given sample should be zero. The user can select as many points in as many factors to pull as they wish.

Figure 43. Example of selecting points to pull to the y-axis in the G-space plot.

After the Constraint Points are defined in the previous three graphs, the Constraints table will appear on the Rotational Tools, Constraints screen, showing a constraint in each row (Figure 44, yellow box). Users then need to select one of the six constraint types included in the pull-down list (column "Type"):

- **Pull Down Maximally** – A factor element is pulled down maximally given a limit of change in the Q-value; users can update the default dQ-value.

- **Pull Up Maximally** – A factor element is pulled up maximally given a limit of change in the Q-value; users can update the default dQ-value.

- **Pull to Value** – A factor element is pulled to a target value given a limit of change in the Q-value (default % dQ = 0.5); users need to input the target value into the "Value" column.

- **Set to Zero** – A factor element is forced to equal zero, with no limit of change in the Q-value.

- **Set to Original Value** – A factor element is fixed to its original value, with no limit of change in the Q-value.

- **Define Limits** – A factor element is given a lower and upper limit; users need to input the "low/high" limit in the column "Value."

Figure 44. Example of the Constrained Model Run summary table.

It should be noted that the constraints defined through the Expression Builder or "Constrain Points" are specific for a selected base run. If users input another run number as the "Selected Base Run" under Constrained Model Run, all constraints associated with the previous base run will be removed from the Expressions and Constraints tables.

After the specification of all constrained model parameters, the user should press the "Run" button in the Constrained Model Run box to initiate the run for a constrained model. Once the run is initiated, the "Run Progress" box in the lower right corner of the screen activates and the constrained model run can be terminated at any time by pressing the "Stop" button. No information about the constrained model runs will be saved or displayed if the runs are stopped. When the constrained model run is completed, the summary table shows dQ, Q(robust), % dQ(robust), Q(Aux), Q(true), as well as whether the run converged (Figure 44, red box). Five new tabs with constrained model run results will appear, including Constrained Profiles/Contributions, Constrained Factor Fingerprints, Constrained G-Space Plot, Constrained Factor Contributions, and Constrained Diagnostics.
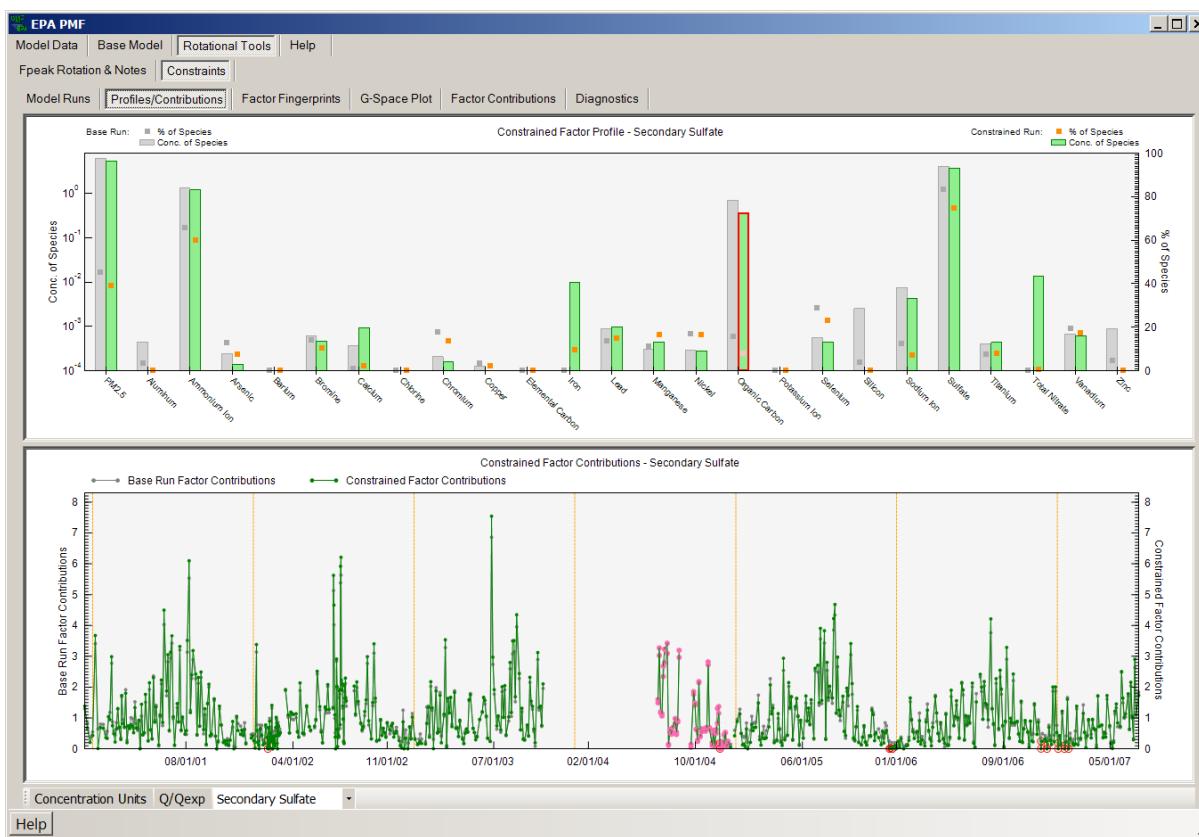
The % dQ (robust) value needs to be evaluated based on the amount of dQ that was used in the constraint(s). The % dQ(robust) shows the increase in Q due to the constraint(s). An increase of dQ of up to 1% for all of the constraints may be acceptable; however, the

interpretation of the factor profiles, contribution time series, and error estimation results are also critical. The Profiles/Contributions tab provides both the base and constrained factor profiles and well as the base and constrained factor time series. Evaluate all of the plots for all factors to understand the impact of the constraints and determine whether the constraint has provided a more interpretable solution.

Typically, species contributions to factors fall into two categories: (1) stiff, in that they will not significantly change or if they are constrained, unreasonable profiles are created; and (2) weak, in that they move easily and are typically not well modeled by PMF. The understanding of the stiff and weak key tracer species for sources allows for optimization of the solution using measured profile or other information. Weak species should be interpreted as easily moved between sources while stiff species are strongly associated with the factor and should be used in the interpretation of its source.

6.2.2   Constrained Profiles/Contribution Results
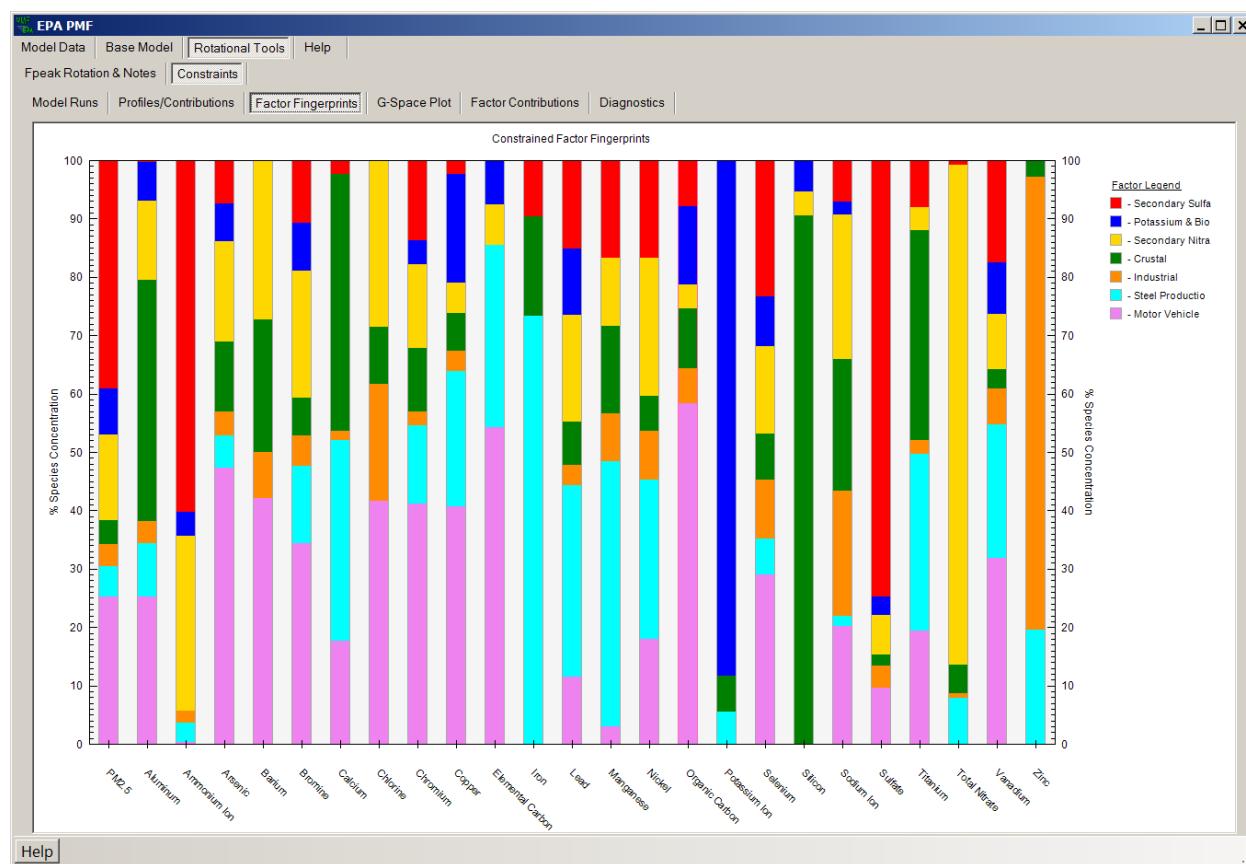
The Constrained Profiles/Contributions screen (Figure 45) shows factor profile and contributions graphs in the same format as those on the Fpeak Profiles/Contributions screen. The mass and percentage of species and the time series of factor contributions are presented for both the constrained model run and the selected base run. The user should look at the deviations in the results between the two model runs and examine the impact of constraints.

Figure 45. Example of the Constrained Profiles/Contributions screen.

Constrained Factor Fingerprints

The Constrained Factor Fingerprints screen shows the concentration (in percent) of each species contribution to each factor as a stacked bar chart (Figure 46). This plot can be used to verify unique factor names and determine the distribution of the factors for individual species. Users should look for deviations (i.e., increases or decreases in a particular species in a factor) with the specified constraint(s) and corresponding base run results.

Figure 46. Example of the Constrained Factor Fingerprints screen.
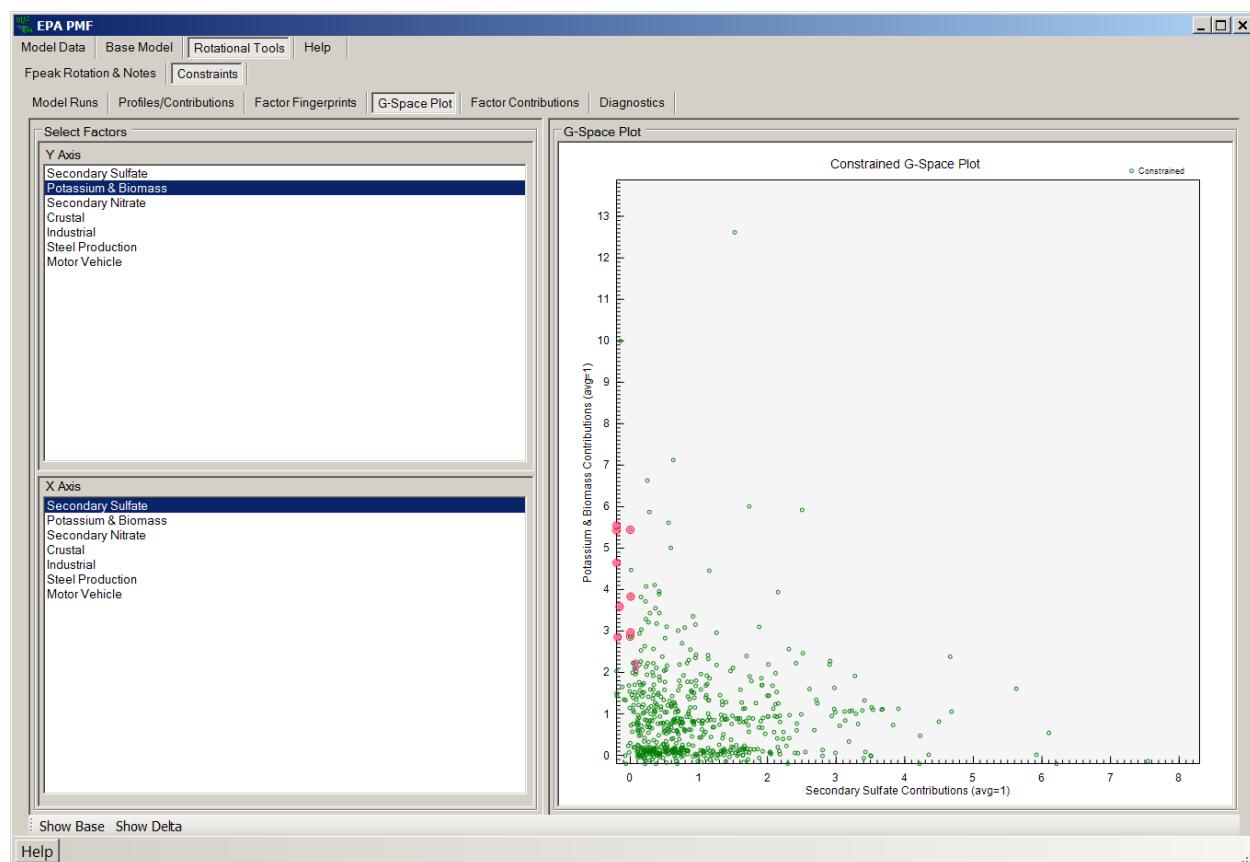
Constrained G-Space Plot

The Constrained G-Space Plot (Figure 47) presents the scatter plot of factor contributions for the constrained model run. Similar to the Fpeak G-Space Plot screen, the user can select "Show Base" to display the base run G-space plot and select "Show Delta" to display the difference in G-space plots between the constrained model run and the base run.

Constrained Factor Contributions

The Constrained Factor Contributions screen (Figure 48) shows two graphs. The top graph is a pie chart, which displays the distribution of each species among the factors resolved by PMF (Figure 48, 1). The species of interest is selected from the table on the left of the screen; the

categorization of that species is also displayed for reference. If a total variable was chosen by the user under the Concentration/Uncertainty screen, that variable is boldfaced in the table. The pie chart for the selected species appears on the right side of the screen. If the user has specified a total variable, the distribution of this variable across the factors will be of particular importance. The bottom graph shows the contribution of all the factors to the total mass by sample (Figure 48, 2). The dotted orange reference lines denote January 1 of each year. The graph is normalized so that the average of all the contributions for each factor is 1.

Figure 47. Example of the Constrained G-Space Plot screen.
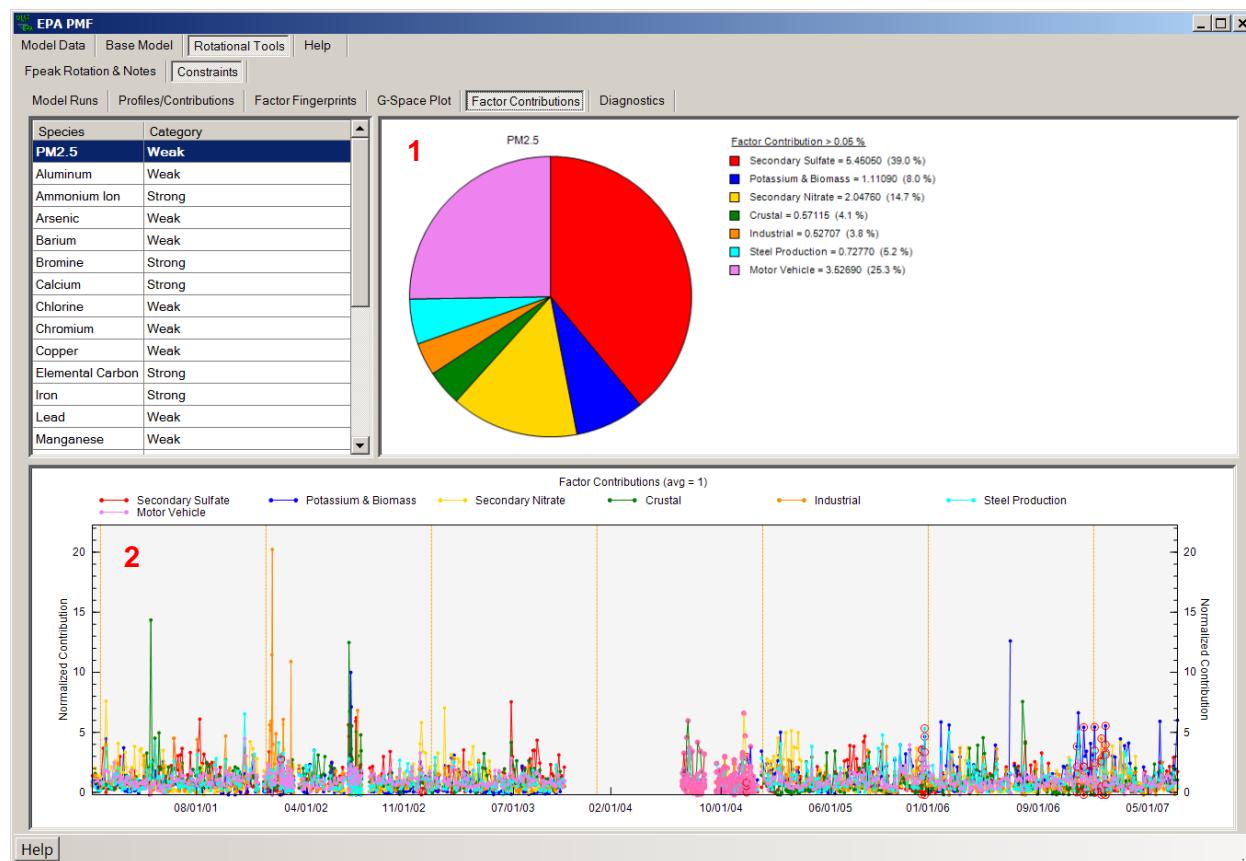
Constrained Diagnostics

The Constrained Diagnostics screen (Figure 49) includes a summary of the constrained model parameters and output for reference (e.g., constraint types, constrained model run summary table, factor profiles, and factor contributions). All of the information on this screen is saved in *_Constrained files.

Constrained BS-DISP and DISP Runs

The BS-DISP and DISP error estimation for the constrained model can be performed in the same manner as the error estimations for the base run. DISP run output files will be saved in

the directory specified in the Output Folder box in the Data Files screen. The DISP and BS-DISP files are saved as *_ConstraintedBSDISPres# and *ConstrainedDISPresd#.

Figure 48. Example of the Constrained Factor Contributions screen.

Constrained BS Runs and Results

A constrained model run can be bootstrapped in the same manner as base model runs. After a constrained model run is completed, the user can initiate a BS run for the constrained model in Constrained Model Bootstrapping. The constrained bootstrapping results are displayed in Constrained Bootstrap Box Plots and Constrained Bootstrap Summary in the same format as the Base Run bootstrapping output screens for easy comparison. The BS files are saved as *_Gcon_profile_boot.

6.2.3   Evaluating Constraints Results

Constraints can be used to reduce rotational ambiguity, to refine a solution, and to understand both stiff and weak factor species. All factors and source contribution time series must be evaluated to understand the impact of the constraint(s). In addition, the error estimation results need to be evaluated to determine if the constraint has changed the species factor contribution significantly. The guidance on constraints will continue to be developed as PMF is applied to

more data sets and the Training Exercises in Section 8 provide more examples on how to interpret the results.
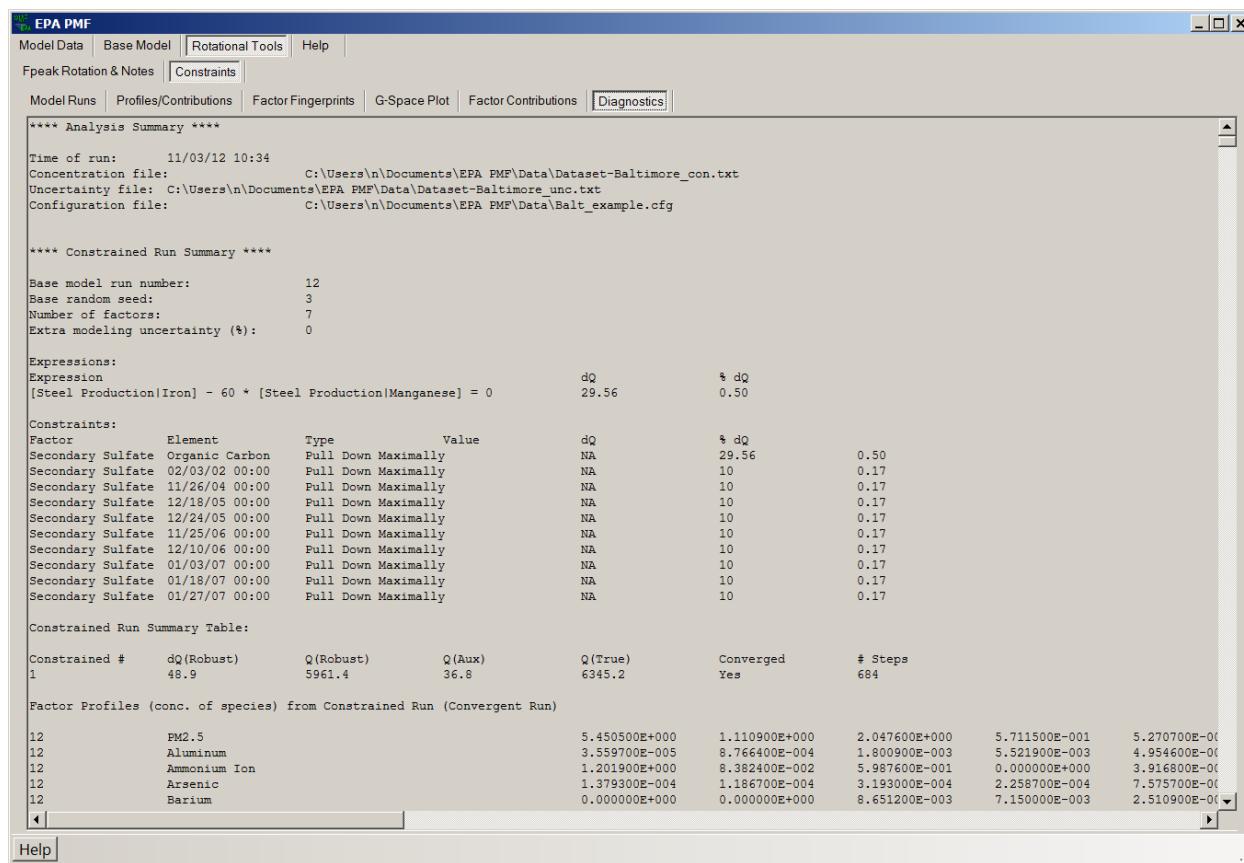
Figure 49. Example of the Constrained Diagnostics screen.

7. Troubleshooting

Common problems in EPA PMF 5.0, including the error messages generated by the GUI and the action the user should take to correct the problem, are detailed in Table 3. If a problem cannot be resolved using the following information, send an email to *NERL_RM_Support@epa.gov*.

Table 3. Common problems in EPA PMF 5.0.

Problem	Error Message	Action
Cannot run base runs	Access to the path 'C:\Program Files\EPA PMF 5.0\PMFData.txt' is denied. Please close all output files.	Turn off User Access Controls in Microsoft Vista
Column headers of concentration and uncertainty files do not match	Species names in uncertainty file do not match those in concentration file. Do you wish to continue?	If the names are correct, continue. If the columns are in a different order, correct and retry.
Number of columns in concentration file is not the same as in uncertainty file	Number of species in uncertainty file does not match the number of species in concentration file.	Select "OK" and examine input files. The same number of columns, in the same order, should be included in the concentration and uncertainty files. If named ranges are used, check that the ranges are defined correctly.
Number of rows in concentration file is not the same as in uncertainty file	Dates/times in uncertainty file do not match those in concentration file.	Select "OK" and examine input files. The same number of rows, sorted by the date/time, should be included in the concentration and uncertainty files. If named ranges are used, check that the ranges are defined correctly.
Blank cells are included in concentration file	Empty cells are not permitted in the concentration input file. Please check your data file.	Select "OK" and remove blank cells from input file before trying again.
Blank cells, zero values, or negative values are included in uncertainty file	Null, zero, and negative uncertainty values are not permitted. Please check your data file.	Select "OK" and remove inappropriate cells from input file before trying again.
Cannot save output files because one is open	The process cannot access the file 'file path and name' because it is being used by another process. Please close all output files.	Close file and select "Retry" or select "Cancel" to change the file path and name.

8.   Training Exercises

The following sections offer examples of PMF analyses of three types of data: (1) water samples collected at multiple locations during rainfall events; (2) hourly aerosol metals data from St Louis, Missouri; and (3) speciated VOC data from a Photochemical Assessment Monitoring Stations (PAMS) site in Baton Rouge, Louisiana. The data sets are installed in the EPA PMF/Data folder and are provided as examples for analyses. Users can follow the steps outlined in each example to better understand the PMF process and the interaction of the components described in this User Guide.

The examples all follow the flow shown in Figure 50, recommended for all PMF analyses. For some users, the Base Model may be sufficient. However, Fpeak can be used to optimize the solution and Constraints can be used to incorporate information on the source such as composition or emissions. Evaluating the error estimates is a critical component of a PMF analysis.
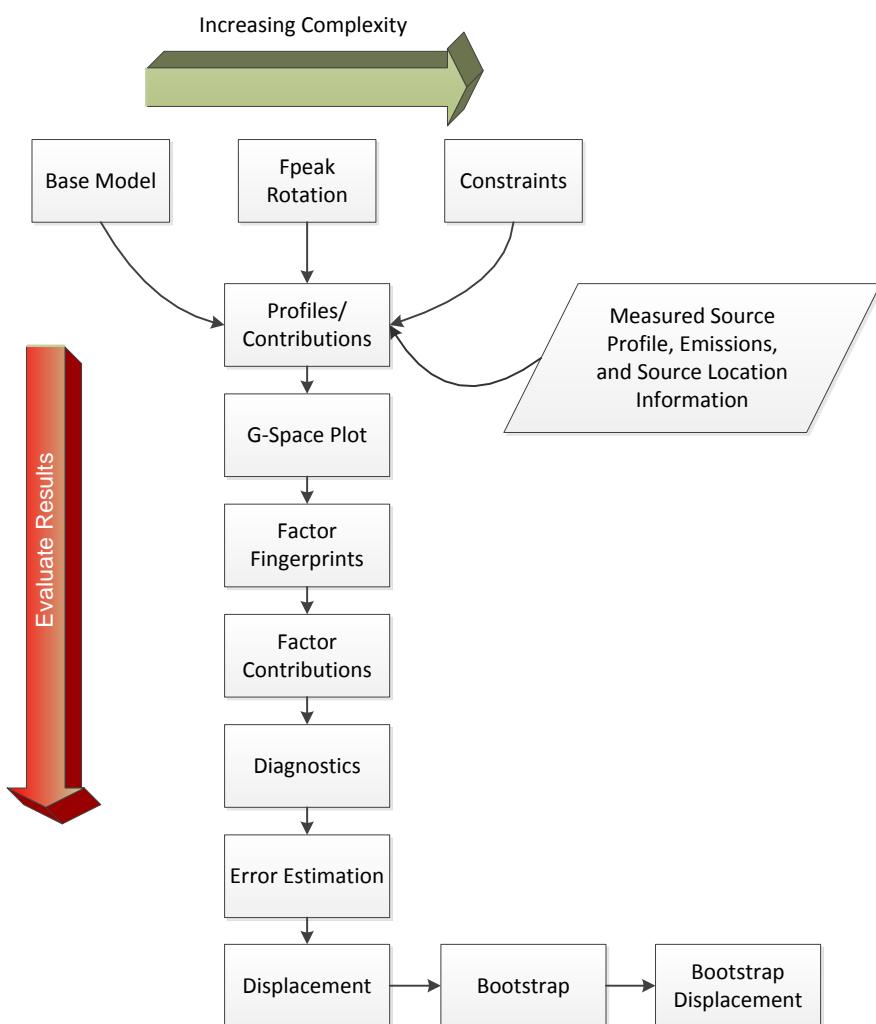
Figure 50. PMF results evaluation process.

8.1   Milwaukee Water Data

This exercise focuses on the data set provided in Mil_water_samples.xls. This exercise is intended to demonstrate the thought process as well as steps involved in evaluating a small data set with event sampling from multiple sites; it is not intended to be a complete source apportionment analysis. The PMF input parameters are summarized in Table 4 and all sites were used in the analysis.

Table 4. Milwaukee Example – Summary of PMF Input Information.
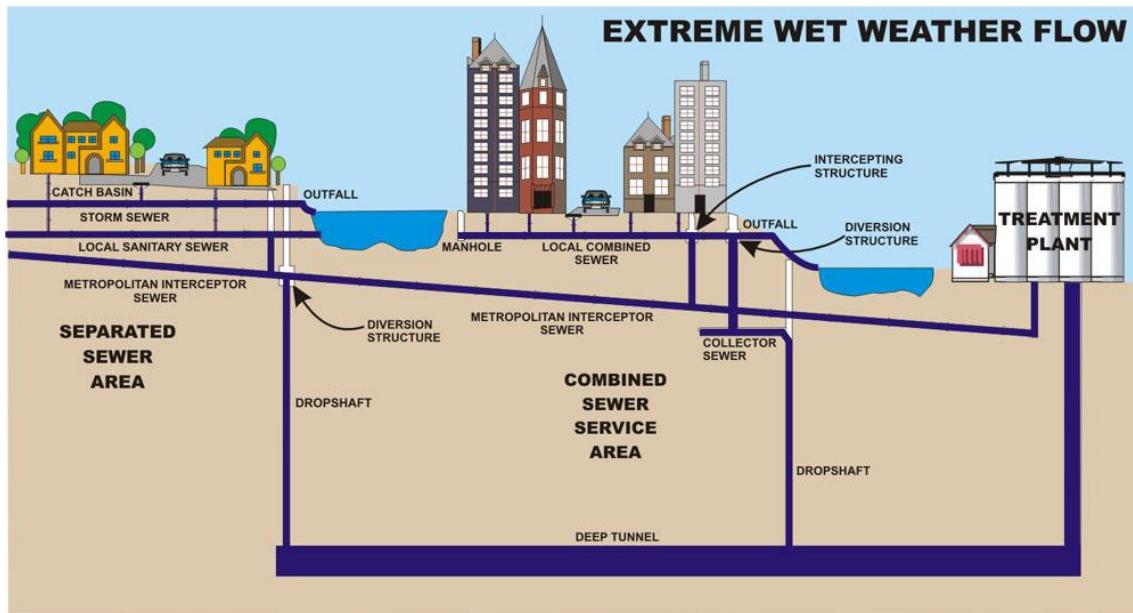
Data Files				****** Base Run Summary ******			
Concentration file:	Mil_water_samples.xlsx ("Conc" worksheet)			Number of base runs:		20	
Uncertainty file:	Mil_water_samples.xlsx ("Unc" worksheet)			Base random seed:		12	
				Number of factors:		3	
Excluded Samples				Extra modeling uncertainty (%):		0	
none							
**** Input Data Statistics ****							
Species	Category	S/N		Species		Category	S/N
BOD5	Strong	4		Cr		Strong	4
TSS	Strong	4		Cu		Strong	4
NH3	Strong	4		Pb		Strong	4
TP	Strong	4		Ni		Strong	4
Cd	Bad	4		Zn		Strong	4

8.1.1   Data Set Development

Soonthornnonda and Christensen (2008) conducted a source apportionment of pollutants contributing to combined sewer overflows (waste water + storm water) from the 19.5-mile (31.4 km) inline storage system in Milwaukee. A diagram of the deep tunnel system is shown in Figure 51 and more information can be found at http://v3.mmsd.com/DeepTunnel.aspx. Samples were collected from multiple sites on one day and the Mil_water_samples.xls file has three tabs: conc (concentration), unc (uncertainty), and site information. The paper reference is also included on the site tab.

Both CMB and a version of PMF that was developed by Bzdusek et al. (2006) were used for the data analysis and the data used for the PMF modeling was posted as supplemental information on the Environmental Science and Technology website[1]. In addition, the authors assumed 20% relative error of the elements of the data matrix. All of the species were initially used in the base model run, 3 factors, and 20 runs. A random seed was initially used to evaluate the variability in runs and the following results are based on a seed number of 12.

---

[1] http://www.researchgate.net/journal/0013-936X_Environmental_Science_and_Technology

Figure 51. Deep tunnel system.

8.1.2   Analyze Input Data

The species relationships were evaluated using the concentration scatter plots. The biological oxygen demand (BOD5) was not related to the total suspended solids (TSS) (Figure 52), indicating that they had separate sources. Also, the cadmium concentrations were only at two levels (Figure 53), potentially indicating an issue with using the species.

8.1.3   Base Model Runs

The obs/pred scatter plot was used to evaluate the base model results because the data were collected from multiple sites on the same date. All of the species have a linear relationship except for cadmium, as shown in Figure 53. Based on these results, cadmium was set to "bad" and the base model was re-run.

The stacked graph plot shown in Figure 54, which shows results similar to Bzdusek et al. (2006a), is created by selecting the top figure in the Profiles/Contributions screen, right-clicking, and selecting Stack Graphs. Select the new window and right-click for file saving options or use "Copy to Clipboard" to paste the figure into a document.

This data set poses some challenges for plotting since the samples were collected from multiple sites on the same day when it is was raining. Rather than on a fixed schedule, the sampling was event-based. The time-series plots have horizontal lines between the sites (Figure 55). Information on the site name and sampling time is displayed on the bottom bar after a point is selected on the figure. The user needs to evaluate whether combining the data in a PMF analysis is justified. The key receptor modeling assumption is the composition of the sources impacting the sites does not change between sites.
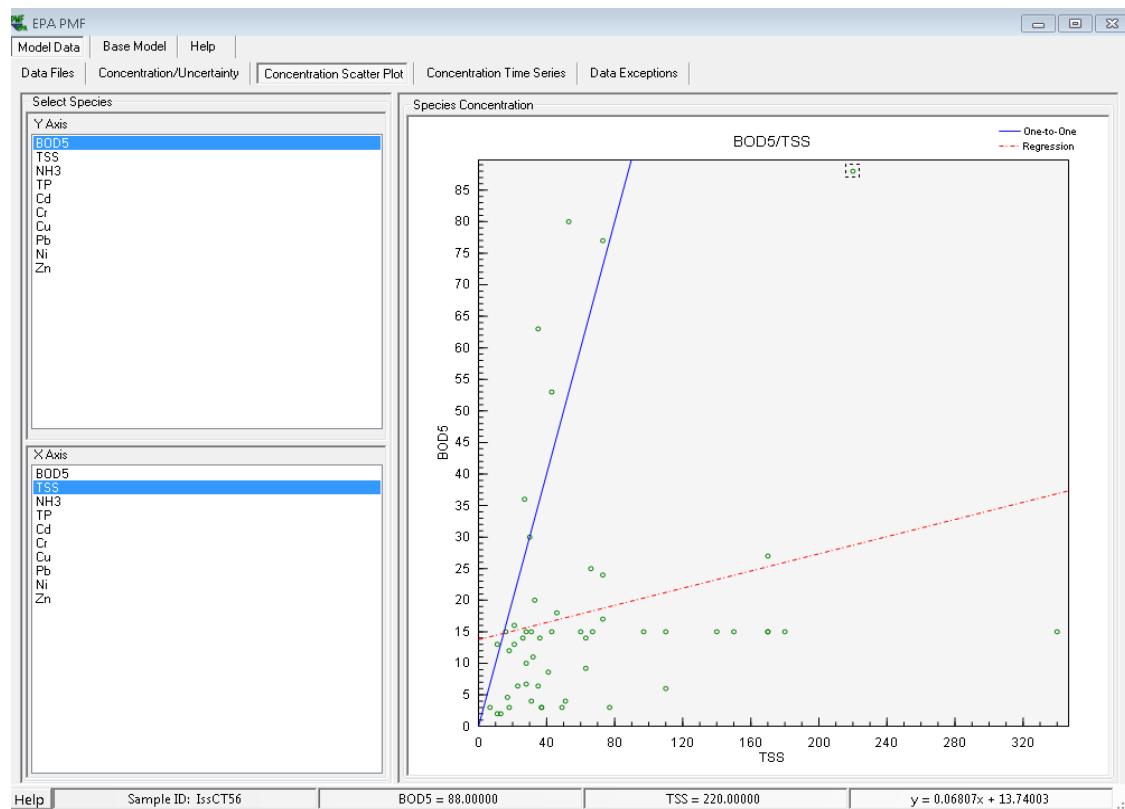
Figure 52. Scatter plot of BOD5 and TSS.
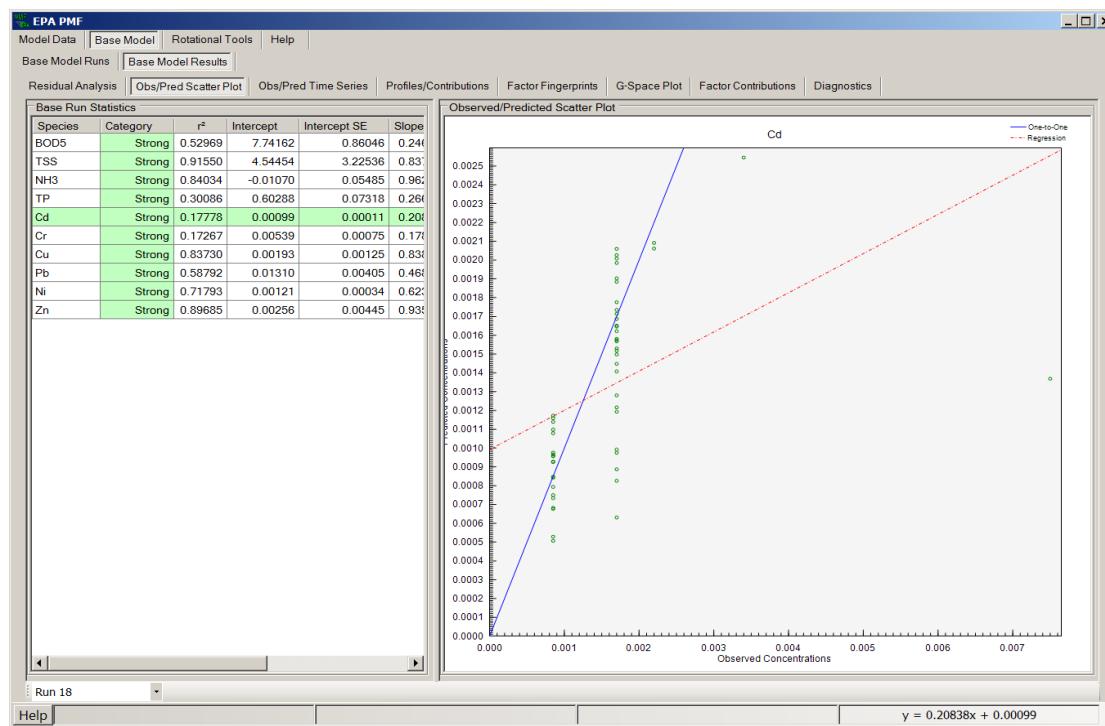
Figure 53. Example of observed/predicted results for cadmium.
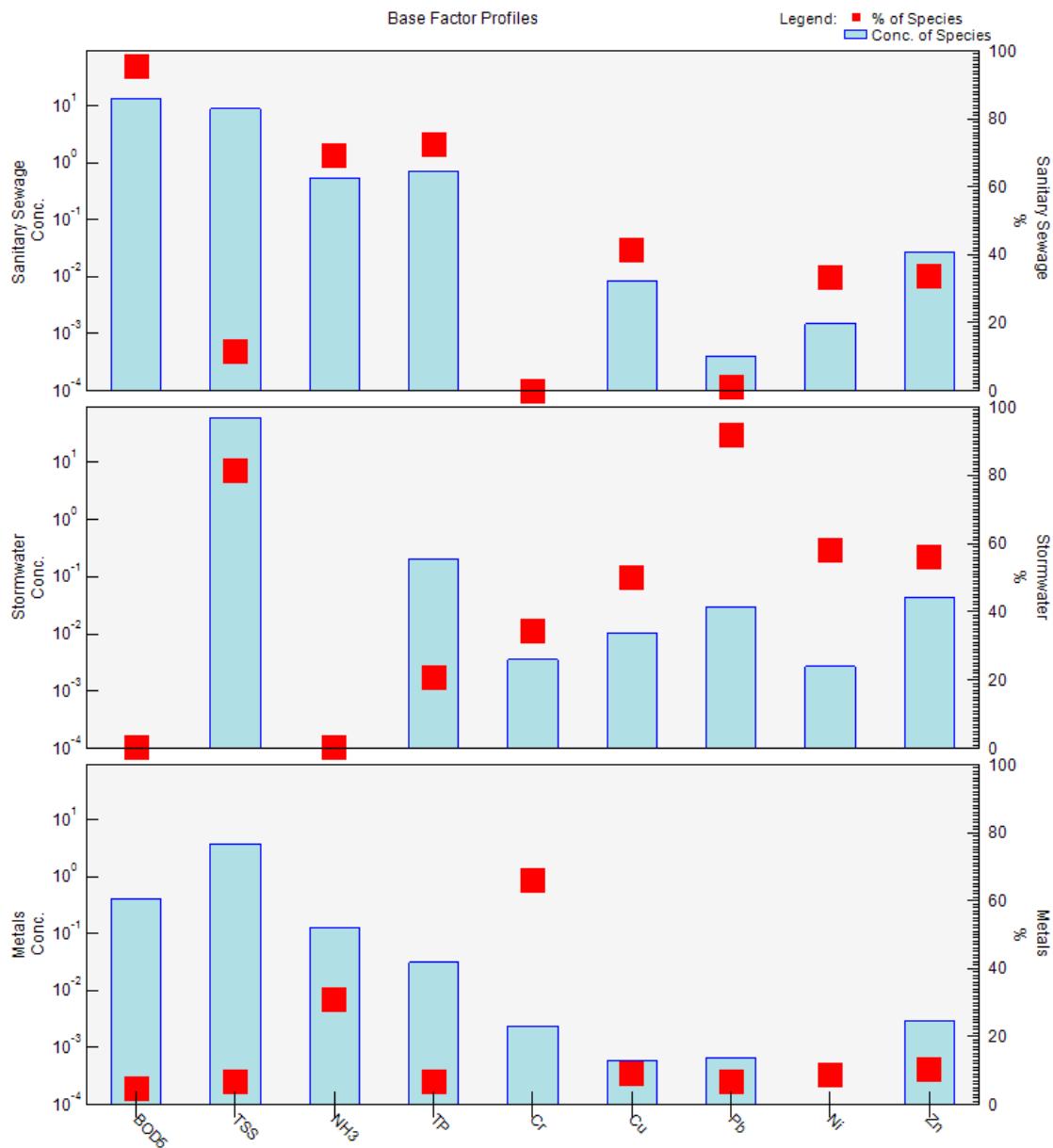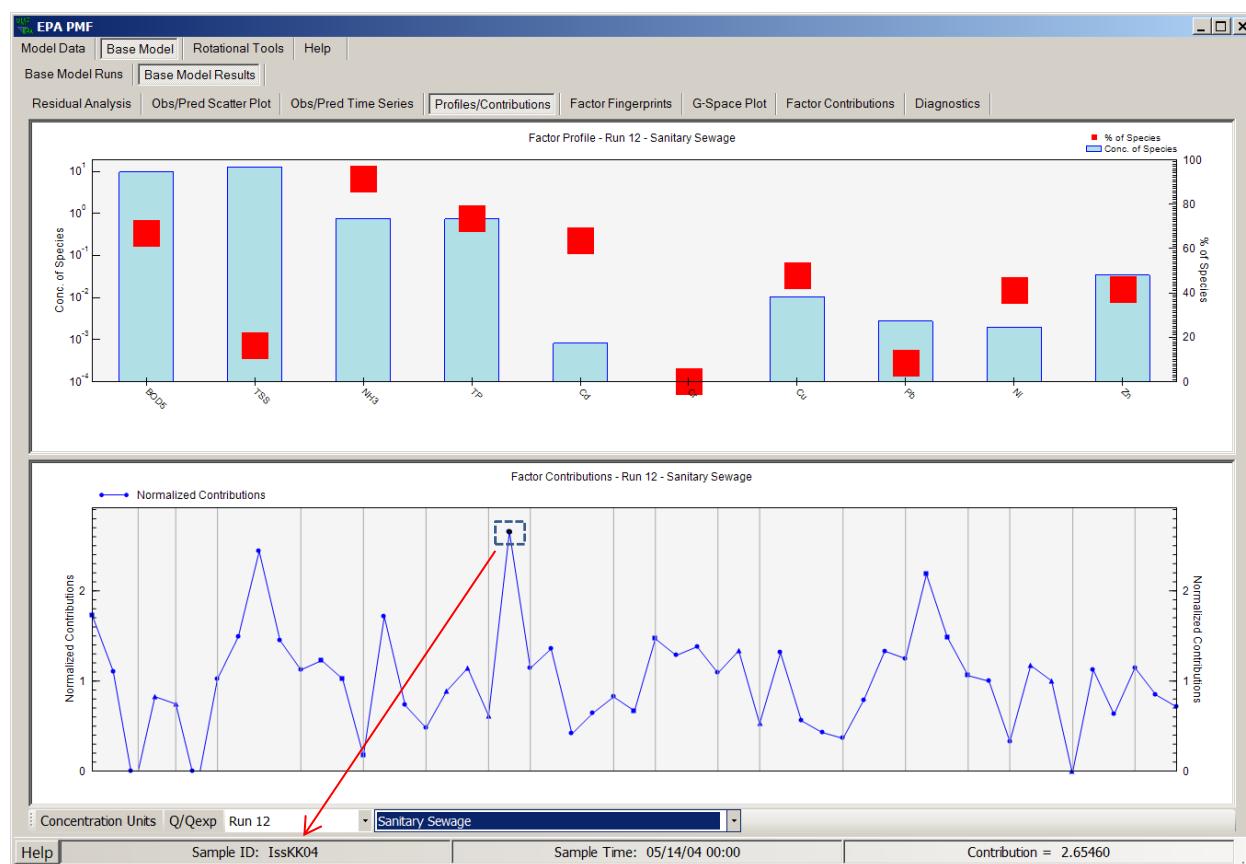
Figure 54. Stacked Graph plot.

The time series of source contributions and observed vs. predicted concentrations provide useful information. The time series of source contributions should show variability between sites and having one site that is impacted and the others with a negligible impact may indicate that the sources compositions are not uniform. The observed vs. predicted plot provide the most important information and sites that have large differences between the observed and predicted (residual) most likely are impacted by more unique sources and could be removed from the analysis. In both cases, a site or sites with significant differences in contributions or residuals need to be evaluated in more detail before keeping them in a multiple site PMF analysis. Time-series plots from the Milwaukee water data are used to demonstrate combining
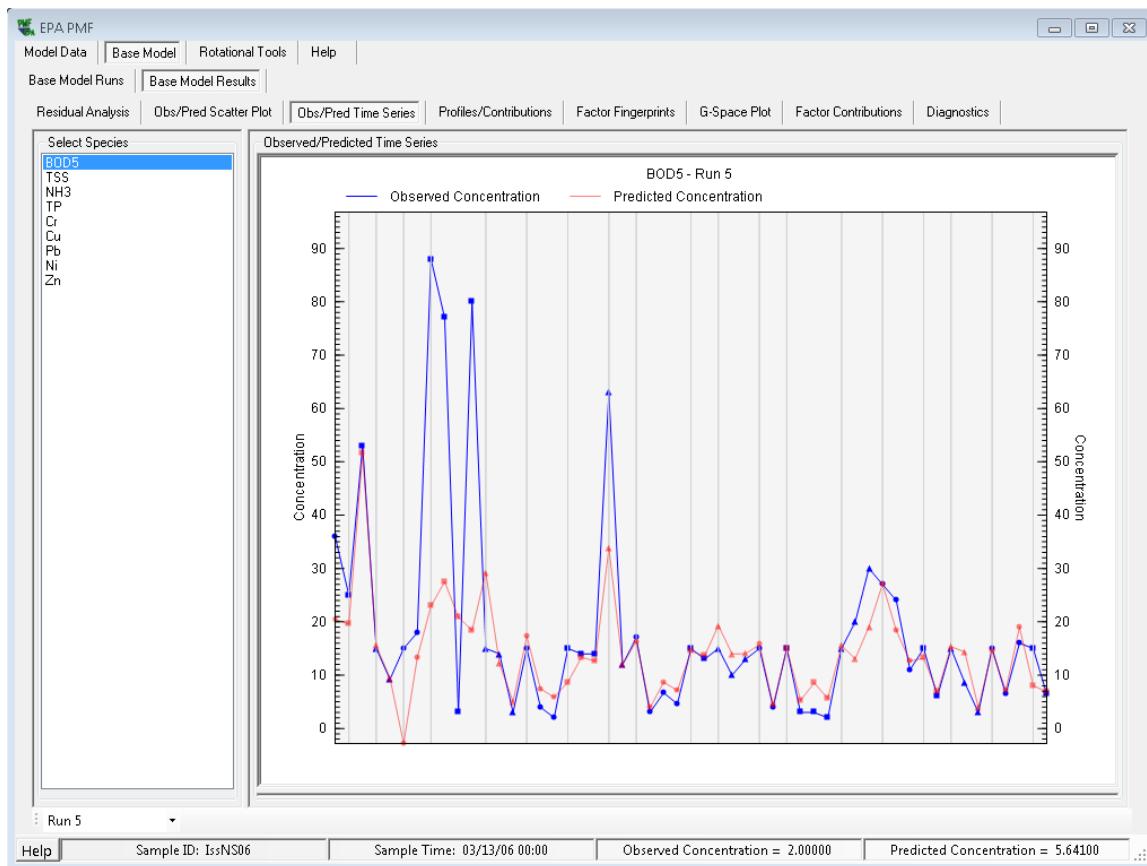
multiple sites in PMF (Figure 55, Figure 56) and the user is encouraged to run each site separately using the check box on the Data File screen and the combined analysis.

Figure 55. Profiles/Contributions Plot for mulitiple site data.

The relative magnitude of the source impacts varies across the sampling sites, however, the impacts are variable and multiple sites have both high and low source contributions. Combining the sites seems justified based on the variability between sites. The observed vs. predicted concentration time series also has lines between the sites (Figure 56). The time series shows that observed and predicted concentrations are large for a few sampling sites and low for others. The data from the sites with large differences should be evaluated in more detail to determine whether the samples should be combined in the PMF analysis.

The Q/Qexp plots should also be evaluated because it provides a complimentary time-series plot to the obs/pred species plots. Time series plots in the Rotational Tools also display the lines between the sites.
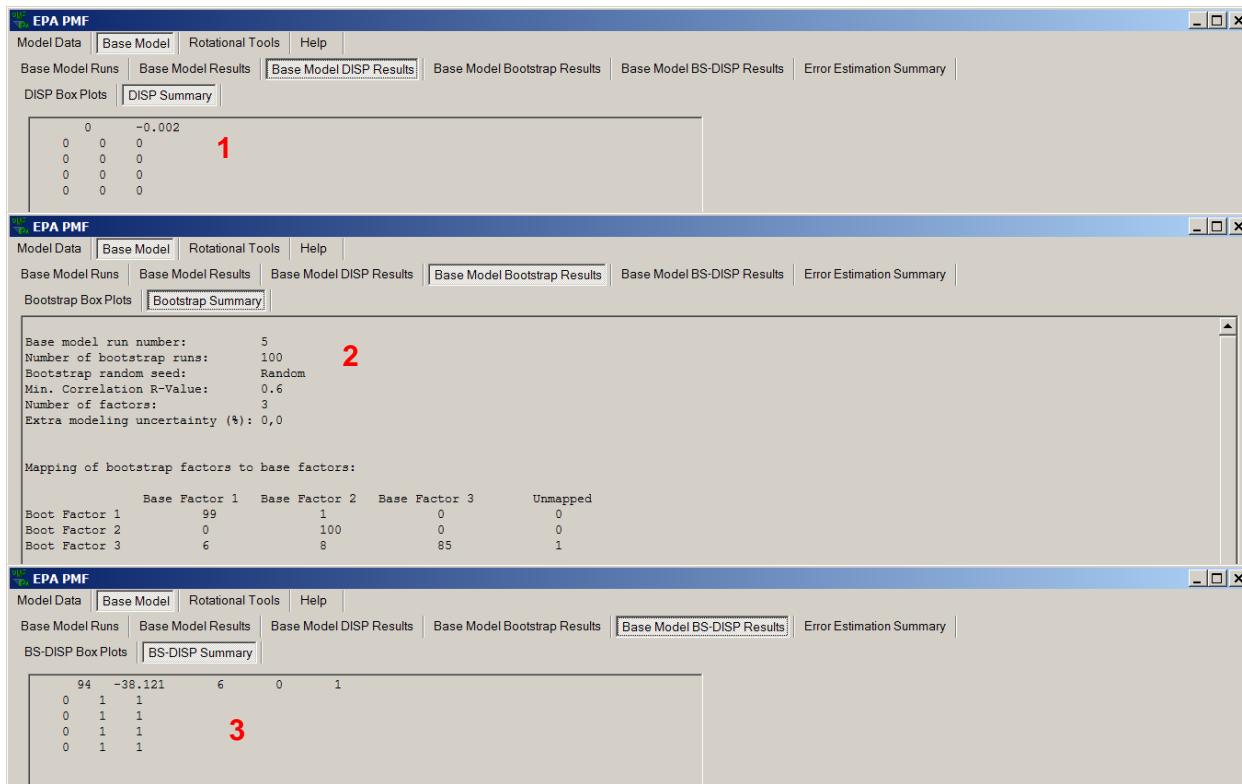
Figure 56. Observed/Predicted Time Series Plot for multiple site data.

8.1.4   Error Estimation

The BS, DISP, and BS-DISP results show some instability in the solution, which is due to the small size of the data set and limited number of factors. The error estimation results are shown in Figure 57.

- DISP results (Figure 57, 1) show that the solution is stable because no swaps are present.

- BS results (Figure 57, 2) for the metals source show that the source was mapped to the sanitary sewage and stormwater sources 6 and 8 times, respectively. This may be due to PMF not fitting this highly variable source and the BS data sets also might not have captured the variability in the metals.

- BS-DISP results (Figure 57, 3) highlight that the solution may not be reliable due to swaps across two factors. The number of swaps is low and the results may reflect the relatively small data set with variability introduced by many sampling sites.

Figure 57. Comparison of error estimation results.

It is recommended that all of the results be reported and explained, and that the *_ErrorEstimationSummary file should be provided as supplemental information for publications. The error estimation summary plot provides a summary of the error estimates. For this analysis, the BS-DISP errors, which capture both random errors and rotational ambiguity, have the largest range (Figure 58).
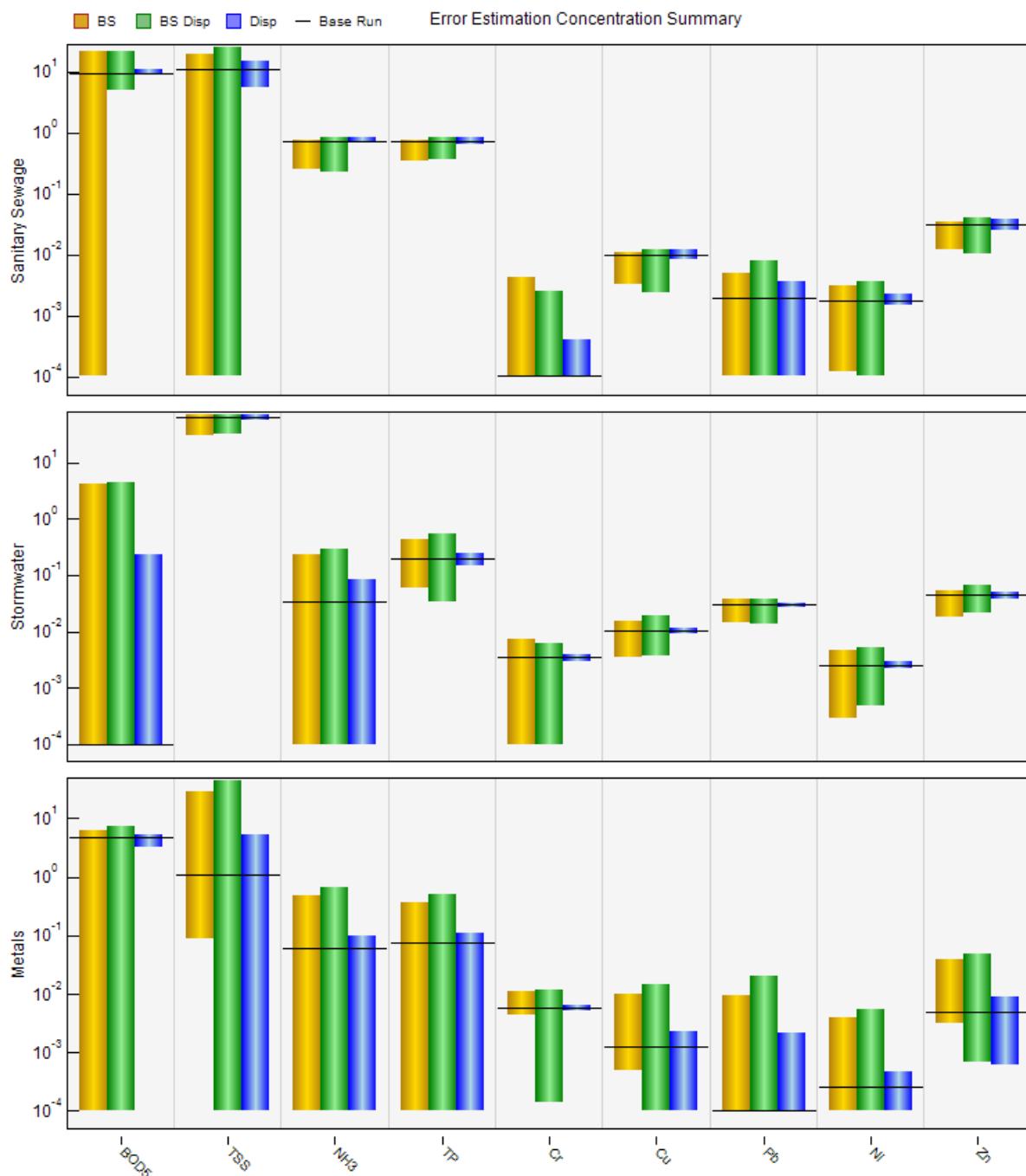
8.2   St. Louis Supersite PM$_{2.5}$ Data Set

This exercise focuses on the data set provided in Dataset-StLouis-con.csv and Dataset-StLouis-unc.csv. The exercise is intended to demonstrate the evaluation of base model results and addition of constraints using EPA PMF. A number of papers have been published on St. Louis particulate matter (PM) apportionment and Amato and Hopke (2012) have recently published an analysis of St. Louis data. The example given here is not a complete analysis; it illustrates how to analyze the data with PMF and the importance of evaluating the model results. The PMF input parameters are summarized in Table 5.

8.2.1   Data Set Development

The St. Louis PM data set includes 13 species and 420 hourly samples, taken during June 2001, November 2001, and March 2002 at the East St. Louis Supersite (Figure 59). The data were formatted in .csv files with each row representing one sample and each column one species. Uncertainty estimates by species and sample were provided by the analytical lab.
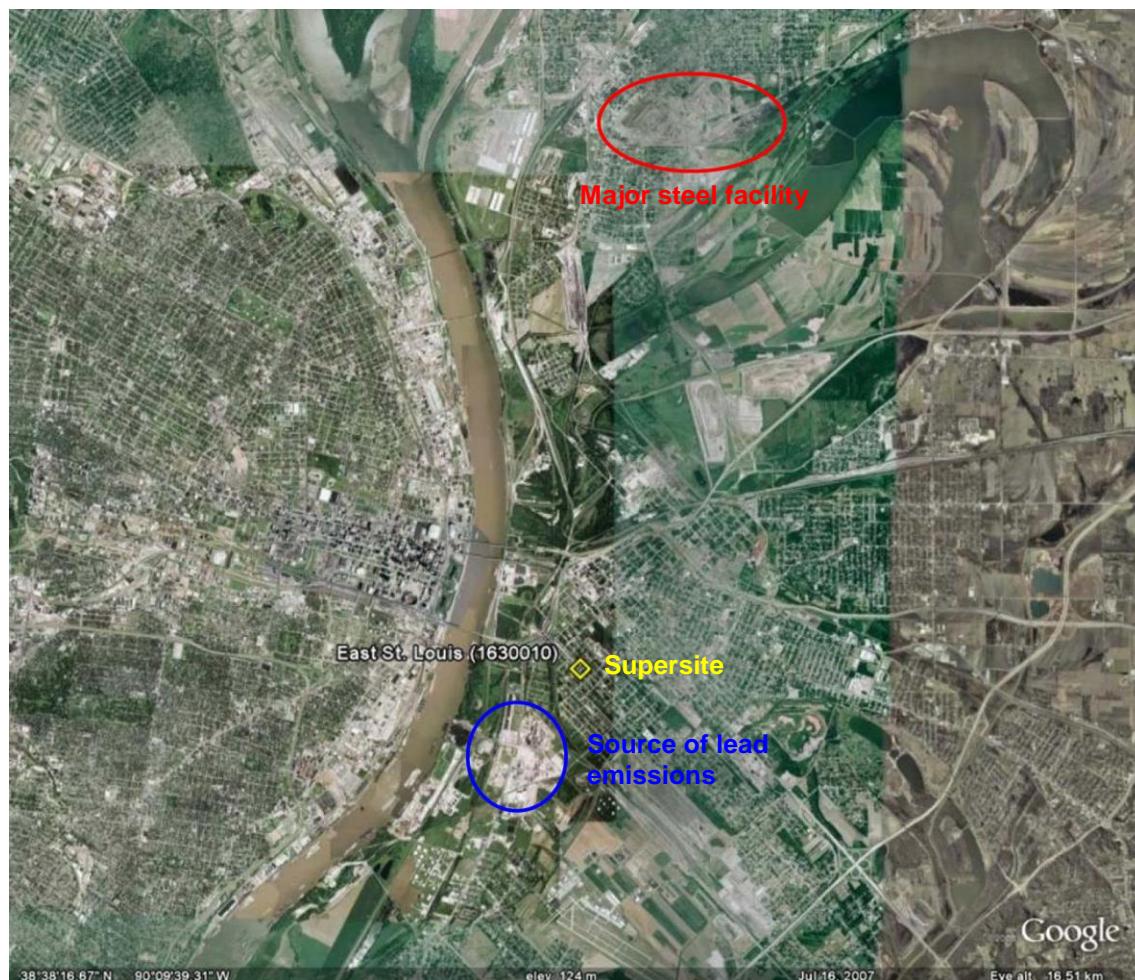
Samples below the detection limit were given an uncertainty of 5/6 the detection limit, missing samples were given an uncertainty of 4 times the median concentration, and samples above the detection limit were given an uncertainty of 1/3 the detection limit plus a sample-specific laboratory uncertainty. In particular, this data set was chosen to illustrate adding constraints to the PMF model based on known source profiles.

Figure 58. Error estimation summary plot of range of concentration by species in each factor.

Table 5. St. Louis Example – Summary of PMF input information.
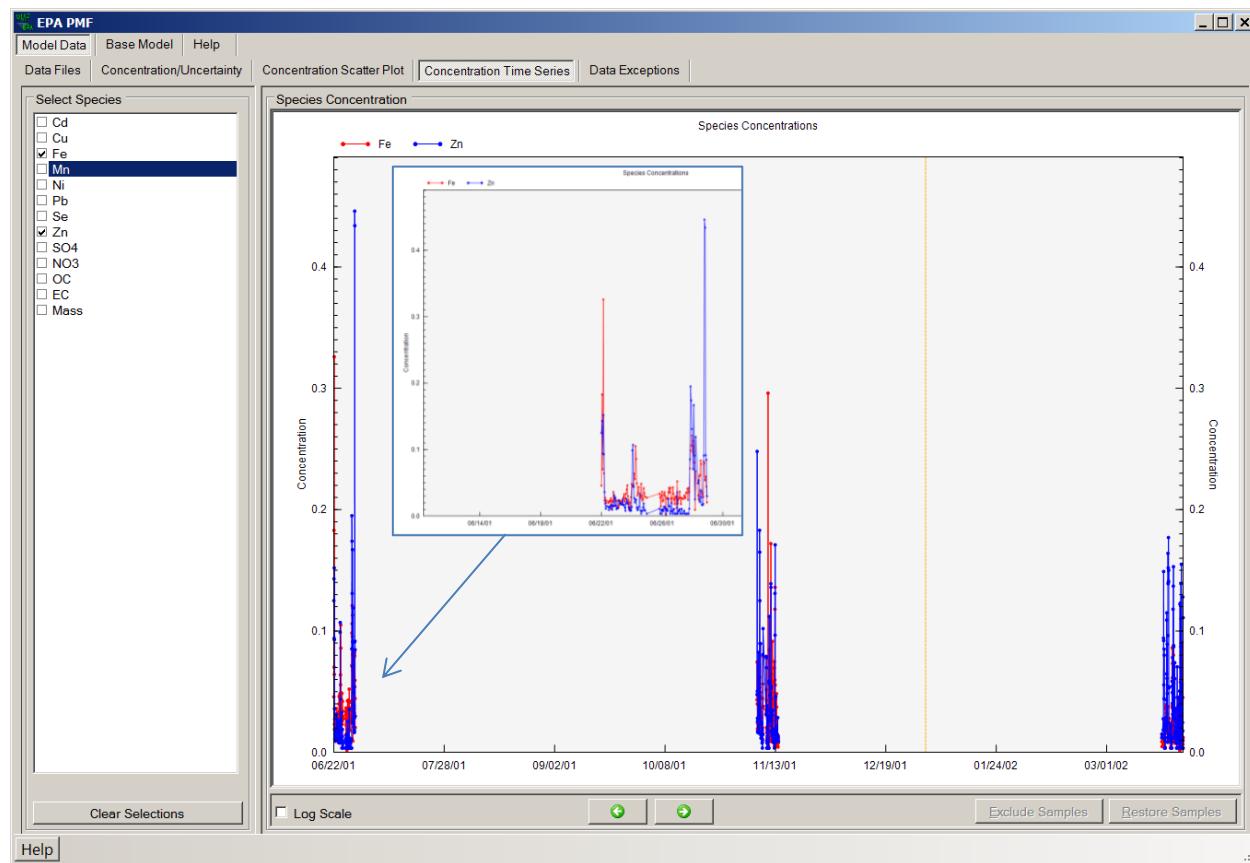
Data Files				****** Base Run Summary ******			
Concentration file:	Dataset-StLouis-con.csv			Number of base runs:		20	
Uncertainty file:	Dataset-StLouis-unc.csv			Base random seed:		30	
				Number of factors:		7	
Excluded Samples				Extra modeling uncertainty (%):		0	
none							
**** Input Data Statistics ****							
Species	Category	S/N		Species	Category	S/N	
Cd	Bad	0.80		Zn	Strong	5.05	
Cu	Strong	5.35		SO4	Strong	6.73	
Fe	Strong	2.30		NO3	Bad	5.31	
Mn	Strong	8.80		OC	Strong	3.59	
Ni	Weak	0.52		EC	Weak	0.67	
Pb	Strong	8.43		Mass	Weak	0.92	
Se	Weak	0.55					

Figure 59. Satellite image of St. Louis Supersite and major emissions sources.

8.2.2   Analyze Input Data

Characterizing Species (Concentration/Uncertainty and Concentration Time Series)

The species categories were set based on the guidance in Section 5.5.1. The user should first examine the input data to determine whether the species concentrations from expected sources are temporally related. For example, do iron and zinc concentrations vary together, indicating the presence of steel production or other sources? The time series of iron and zinc are shown in Figure 60. A zoomed-in graph of the time series is generated by both holding the "Alt" key, and the left mouse button while drawing a box around the period of interest. Select "Alt" and click the left mouse button to return to the original figure.
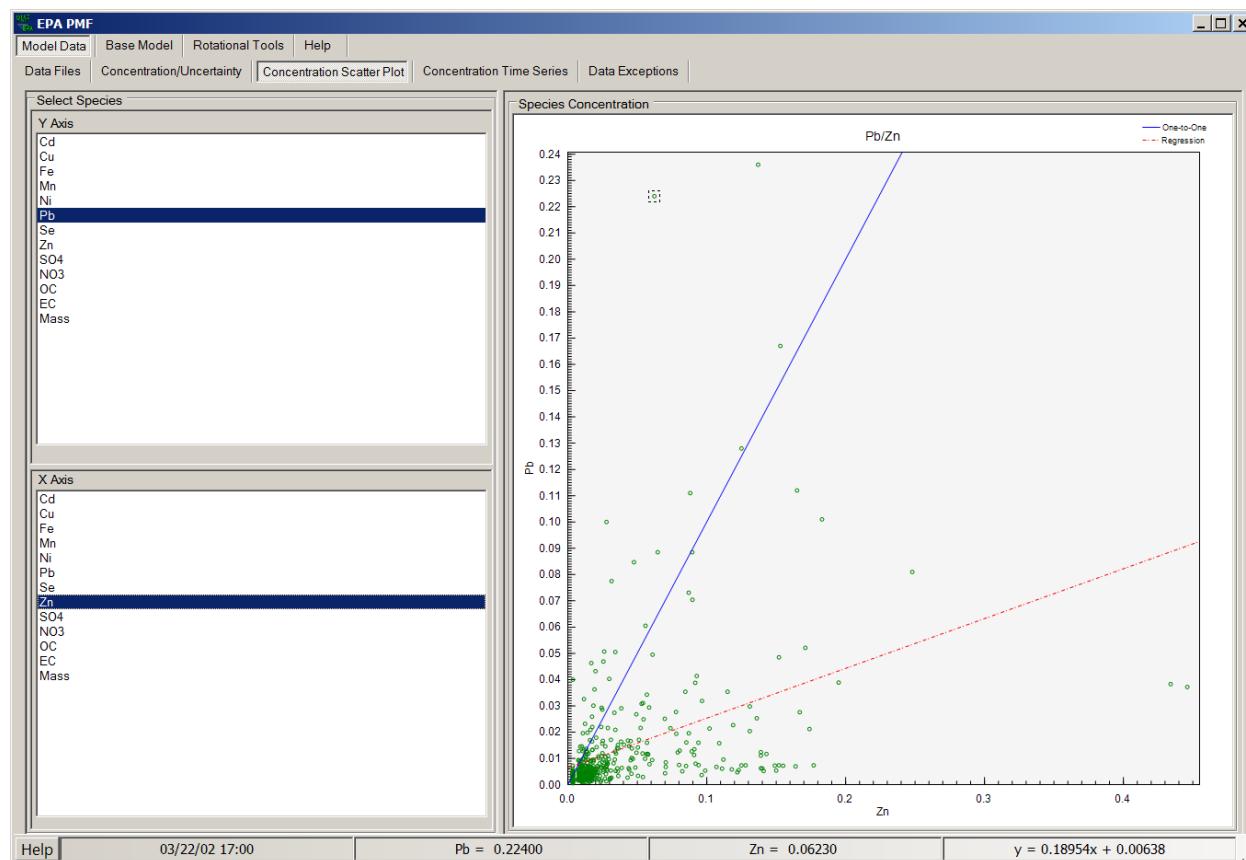
Figure 60. Concentration Time Series screen and zoomed-in diagram for the St. Louis data set.

The plot in Figure 60 shows a complex picture, because high zinc concentrations do not correspond to iron concentrations. This discrepancy may indicate a local source of zinc that does not include iron. In the case of this example in St. Louis, a zinc smelter was located near the monitoring site.

Relationships Between Species (Concentration Scatter Plot)

Scatter plots between species should be examined for relationships that indicate that a common source emitted both species (e.g., OC and EC are both emitted by mobile sources). In the St. Louis data set, lead and zinc are not related, which indicates two potential sources (Figure 61).

Figure 61. Concentration scatter plots for steel elements.

Excluding Samples (Concentration Time Series)
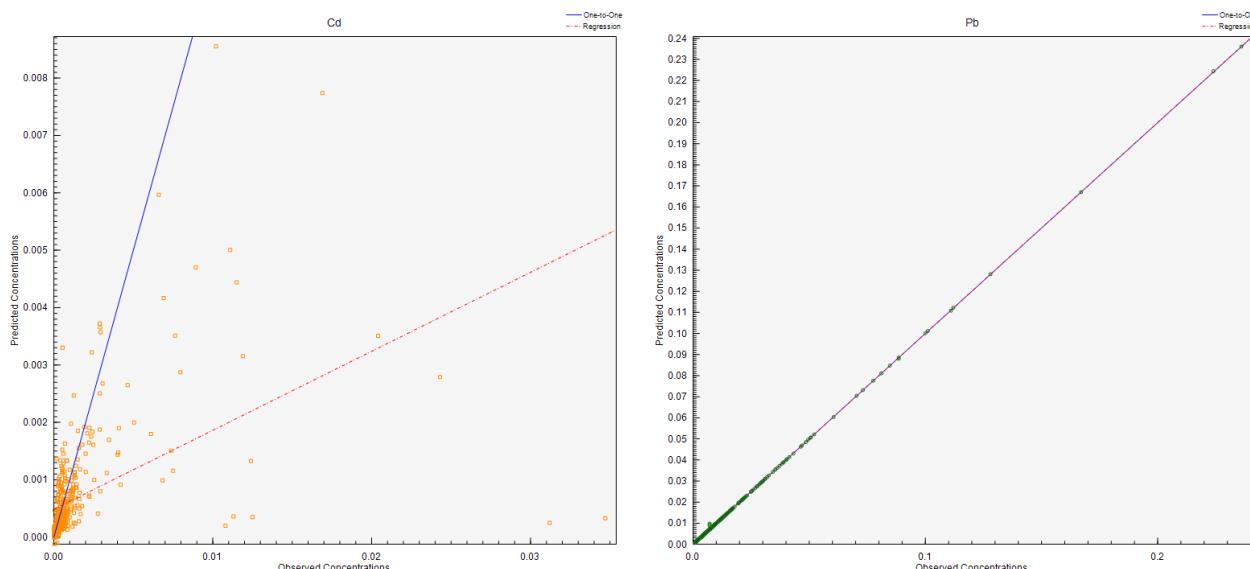
The user should examine the concentration time-series plots to verify that the species selected for PMF have expected seasonal patterns (e.g., high sulfate during the summer), as well as to identify unusual events (e.g., fireworks on the Fourth of July, which contribute to high levels of potassium, strontium, and other trace metals). Often, these events are easily identified. The samples taken during these identified events should be excluded because the overall profiles may not capture the unique composition of the source, or the profiles of non-event sources may be distorted. Exclude a sample by highlighting it and clicking "Exclude Samples" at the bottom right of the screen. All data exclusions must be well-justified and documented.

8.2.3   Base Model Runs

Initial Model Parameters (Base Model Runs)
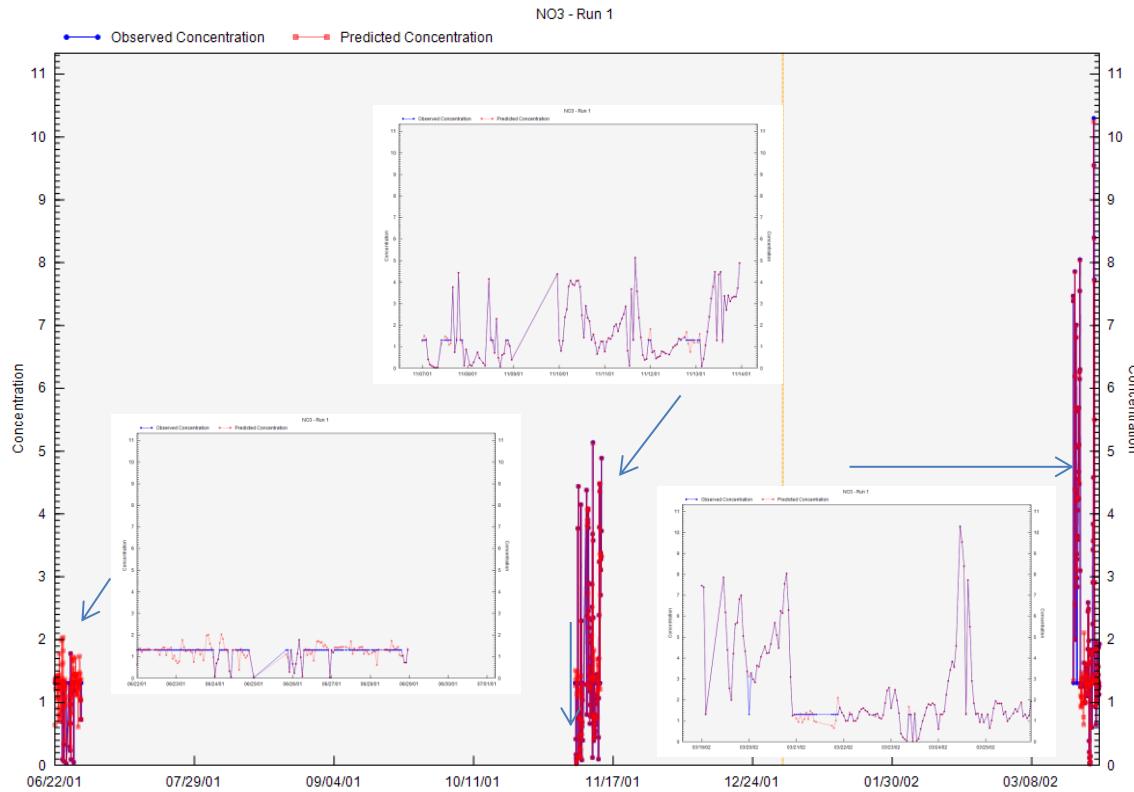
The model was run 20 times with 8 factors and a seed of 30. A constant seed was used to replicate results for training purposes and the runs converged and the Q values were very stable. The Q(robust) was about 10% lower than the Q(true), indicating some, but not heavy, impact of outliers on the Q-value.

Based on the observed-versus-predicted scatter plots and time series, some species, such as lead, were modeled well, and others, such as cadmium, were not well-modeled (Figure 62). This could be the result of incorrect uncertainties, improper categorization (e.g., as strong species), too few factors being modeled, not enough impacts from the source, or PMF incorrectly modeling the species variability. This lack of fitting trace species has been noticed for high-time-resolution sampling (one-hour frequency or less). A cadmium source such as an incinerator is most likely present near the monitoring site. However, the data does not have enough information for PMF to resolve it. The poorly modeled species (cadmium) should be categorized as "Bad."

Figure 62. Example of output graphs for cadmium (poorly modeled) and lead (well-modeled).

In addition, NO_3 (shown in the graphs in Figure 63) has many fixed values for the first intensive during the summer of 2001 that may be set at the MDL. This issue is not present for the next two time periods as shown in Figure 63 and NO_3 should be set as "Bad" if the entire data set is used and "Strong" if only the last two intensives are used.
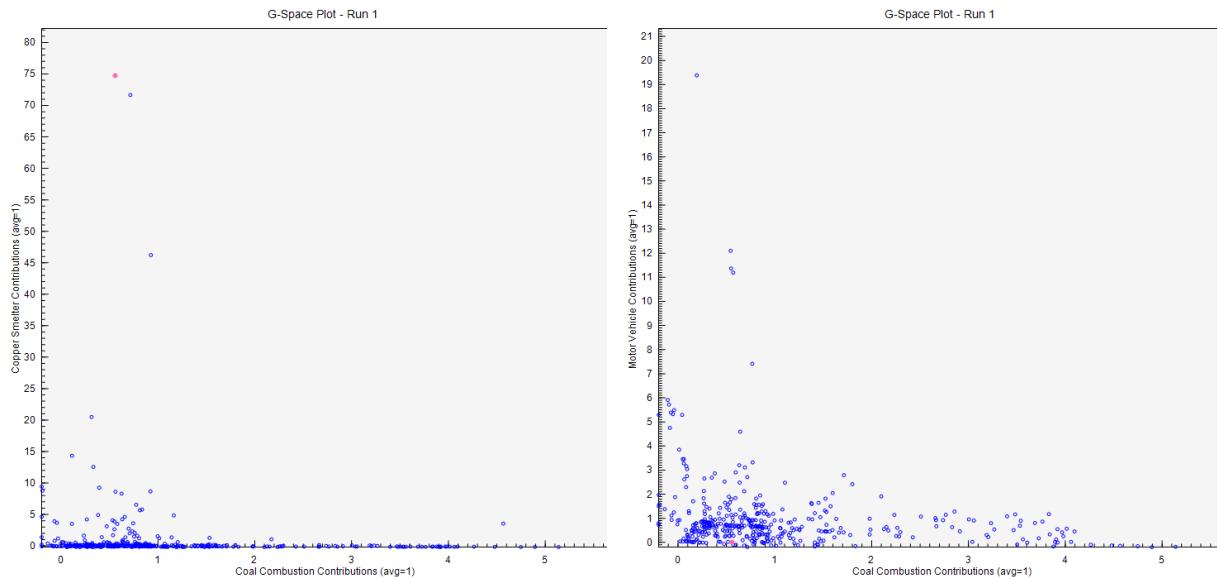
Figure 63. Example of inconsistencies in input data. The multiple points shown in blue in the lower left graphic are fixed values.

Rotations (G-Space Plots)

G-space plots of the solution should be examined to determine whether the contributions fill the solution space and there are edges or points with low or zero contributions. Selection of the species for these plots is important and species should be plotted against regional source indicators, such as coal-fired power plants. Figure 64 shows two examples, one with points near both axes and the other with points only on one axis. Fpeak should be evaluated to determine whether a more optimal solution can be found. If a point is selected in one figure, the same point will be highlighted in the other figures.

Factor Identification (Profiles/Contributions, Aggregate Contributions)

Factors may be identified using dominant species and temporal patterns. Nitrate was removed from the analysis and the number of factors was reduced to seven (since nitrate was one factor). The seven factors identified in the St. Louis data set represent a realistic solution based on known sources in the area, which are crustal (Mn), copper smelter (Cu), coal combustion (SO_4, Se), zinc smelter (Zn), iron and coal (Fe and EC), lead smelter (Pb), and motor vehicle (OC, EC). The iron and coal factor seems to be a mix of species and the factor is evaluated using the constraints later in this example. The factor profiles are shown in Figure 65.

Figure 64. Example of G-space plots for independent (left) and weakly dependent factors (right).

Mass Distribution (Factor Contributions)

Figure 66 shows the factor contributions as a pie chart for the total mass variable ($PM_{2.5}$). Evaluate the distribution of contributions to determine whether they are within the expected range for the samples. The major sources for this example are motor vehicles and coal combustion, with minor contributions from the crustal, zinc smelter, lead smelter, and copper smelter sources.

8.2.4 Error Estimation

A summary of the error estimate results from the *_ErrorEstimationSummary file are shown in Table 6 along with comments. The results are stable and no swaps were present. The *_ErrorEstimationSummary file should be reported with any publication and report.

This example demonstrates the iterative approach for evaluating a PMF solution: evaluate input data, calculate and evaluate base results, and evaluate error estimates. The Error Estimation Concentration Summary plot is shown in Figure 67.

8.2.5 Constrained Model Runs

Define Constrain Expressions (Expression Builder)

For the St. Louis data set, source profiles of local steel facilities were used to determine appropriate ratios of iron and manganese in the steel factor. Samples were analyzed as described in Pancras et al. (2005). This method provides total inorganic concentrations, which are comparable to the total inorganic concentrations from Energy Dispersive X-ray fluorescence (EDXRF). The profile of the Granite City Steelworks basic oxygen furnace was used as a representative sample, because it is believed to be impacting the site; the ratio of EDXRF iron to

manganese in the source profile was 60. The average ratio of iron to manganese in the St. Louis ambient air data was 10.8. However, the base model run results from PMF showed that the iron–to-manganese ratio of 51 was a little low based on the steel factor profiles. The ratio constraint was defined using the Expression Builder, which was interpreted as an autopull equation of iron minus 60 times the manganese in the steel factor, pulled to zero with a given dQ limit ([Steel|Fe] – 60 * [Steel|Mn] = 0). In addition, EC was selected in the iron and coal factor and the right mouse button was used to toggle EC as a constraint. This might allow EC to be better separated from the steel source. The % dQ was set at 5% for each constraint and the converged results used 2.1% dQ.
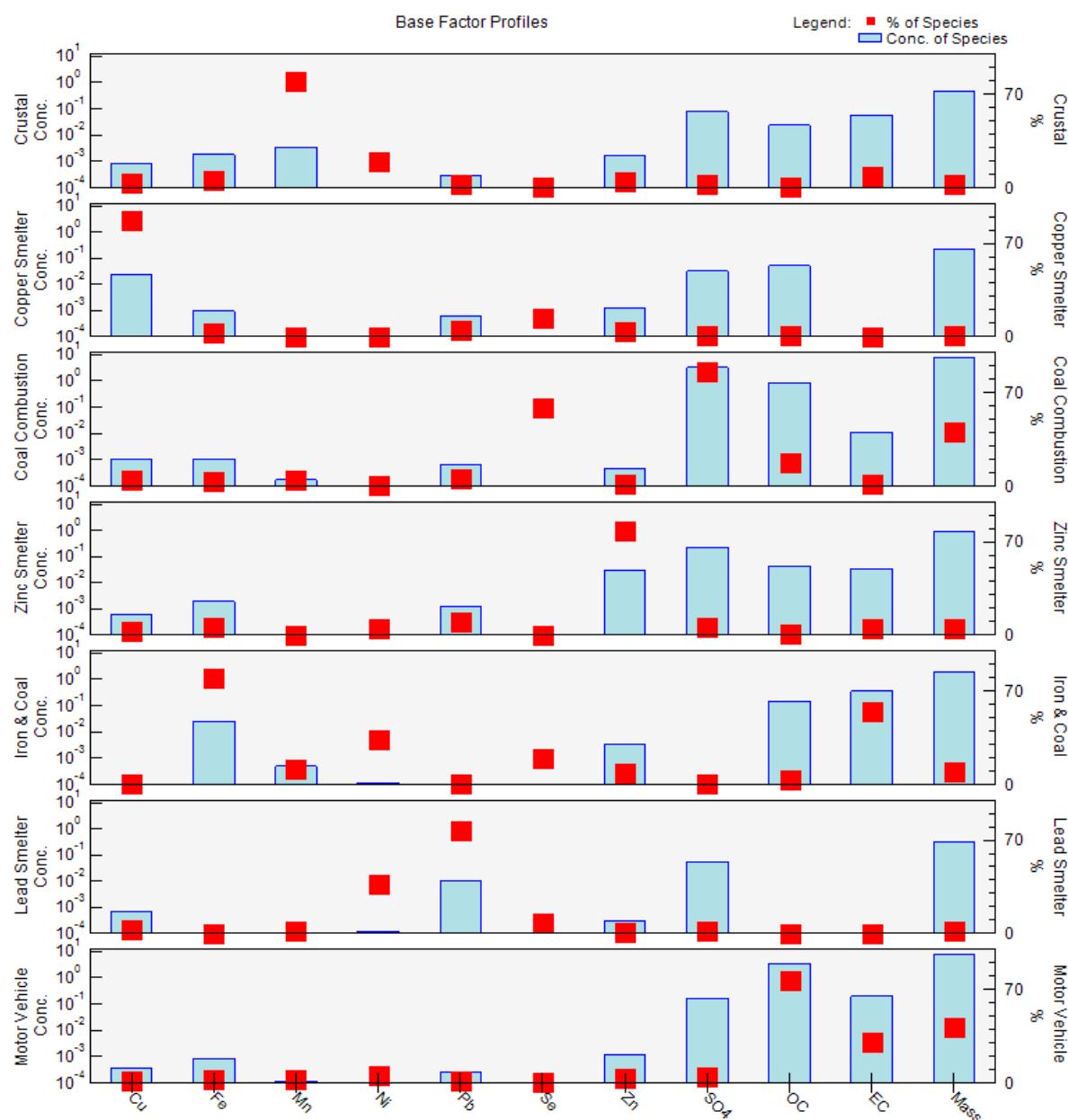
Figure 65. St. Louis stacked base factor profiles.
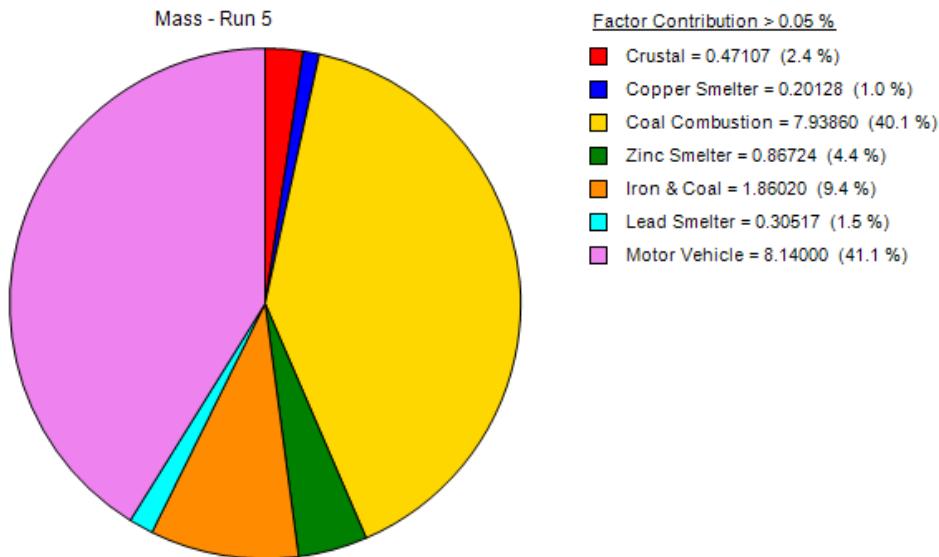
Figure 66. Distribution of mass for St. Louis PM$_{2.5}$.

Constrained Model Run Results (Constrained Profiles/Contributions and Diagnostics)

In the resulting constrained run, the ratio moved to 60 and the EC was also significantly reduced to around 40%, shown in Figure 68. It is important to remember that EC will be shifted to another factor. The largest change in profile was found for motor vehicles. This indicates that the constraints provide an improved result compared to the base run.

These changes did not have a large impact on the overall factor contributions to the mass (the iron and coal factor was reduced by 2.3% and the motor vehicle factor increased by 1.1%); however, it demonstrates the benefit of bringing in external information. After adding constraints, run all three error estimates and compare them to the base model results. The error estimate summary (Figure 69) does not show a significant change. In other data sets, the addition of constraints may reduce the size of error estimates by reducing rotational ambiguity.

8.3   Baton Rouge PAMS VOC Data Set

The following sections detail a PMF analysis of a Photochemical Air Monitoring Station (PAMS) VOC data set from Baton Rouge, Louisiana. The user should run EPA PMF 5.0 with the data sets provided in Dataset-BatonRouge-con.csv and Dataset-BatonRouge-unc.csv to follow the analyses described below. This exercise is intended to demonstrate the thought process and steps involved in reaching a solution using EPA PMF 5.0; it is not intended to be a complete source apportionment analysis. The PMF input parameters are summarized in Figure 69.
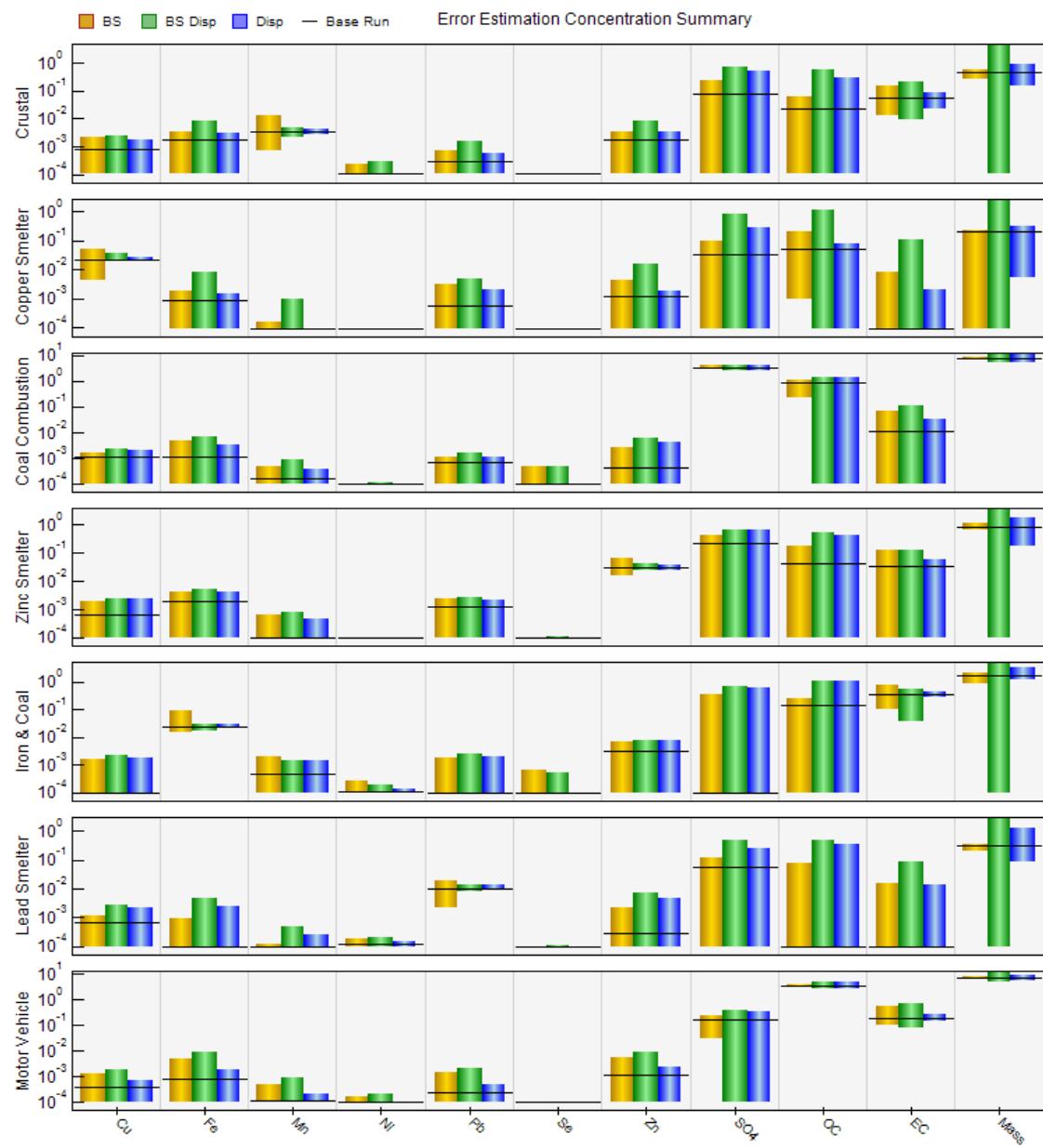
Figure 67. Summary of base run and error estimates.
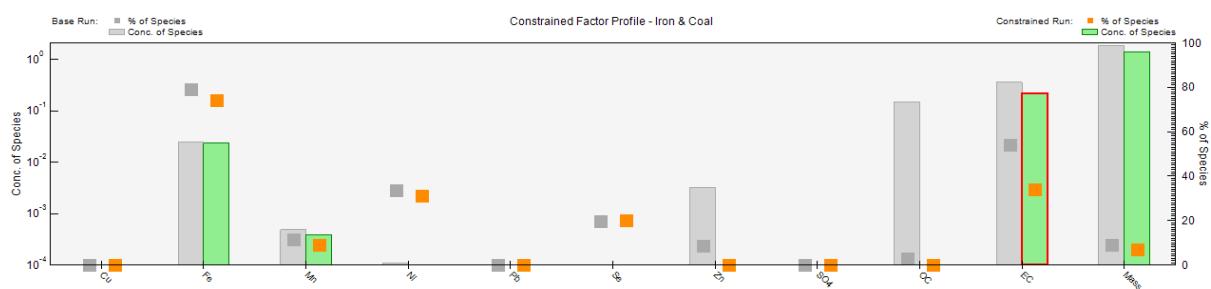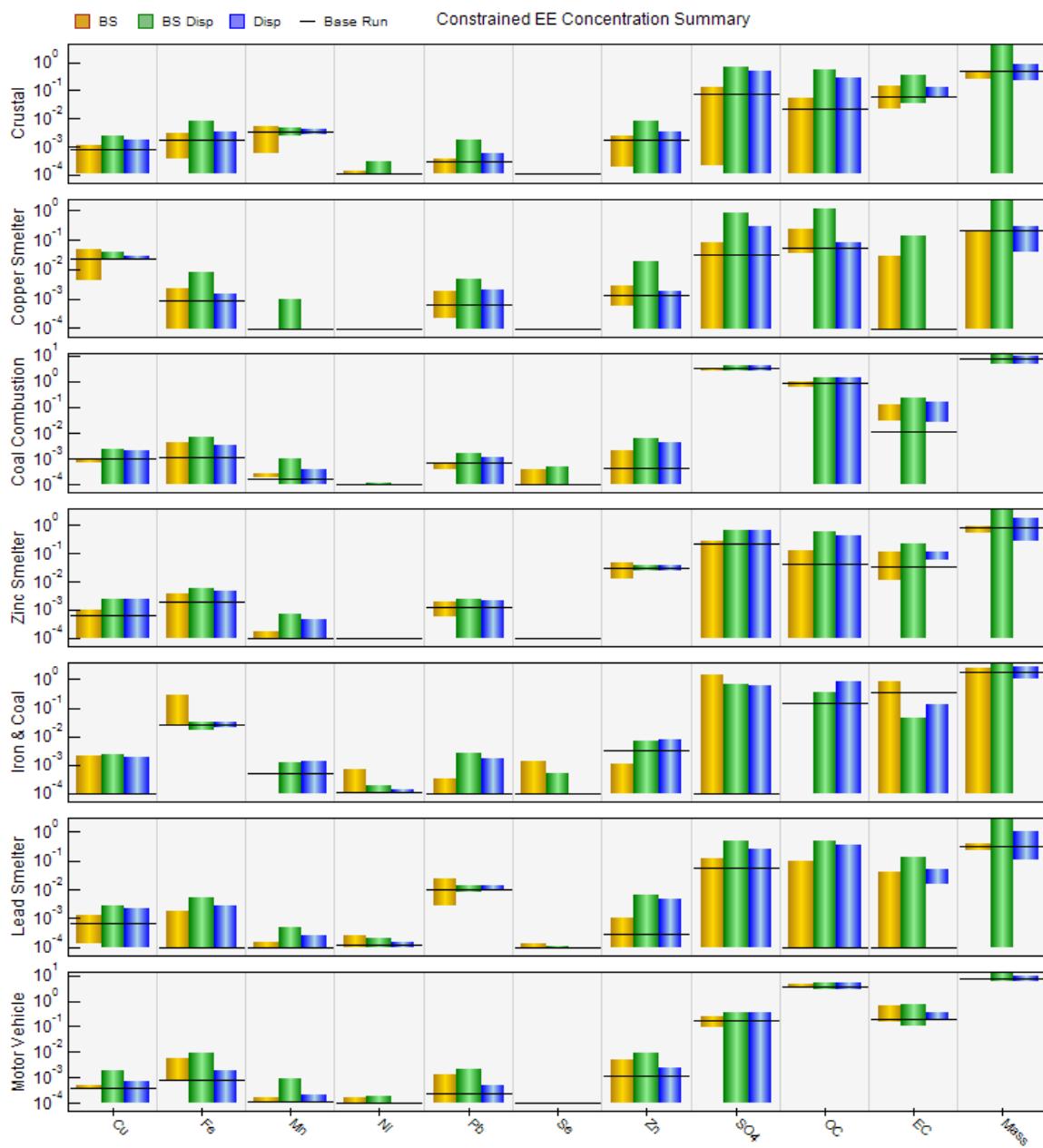
Figure 68. Comparison of base model and constrained model run profiles for the steel factor.

Table 6. Error Estimaton Summary results.

BS-DISP Diagnostics							
# of Cases:	101						
Largest Decrease in Q:	-0.382999986						
% dQ:	-0.370098358						
# of Decreases in Q:	0						
# of Swaps in Best Fit:	0						
# of Swaps in DISP:	0						
Swaps by Factor:	0	0	0	0	0	0	0

DISP Diagnostics							
Error Code:	0						
Largest Decrease in Q:	-0.035999998						
% dQ:	-0.034787313						
Swaps by Factor:	0	0	0	0	0	0	0

BS Mapping	Base Factor 1	Base Factor 2	Base Factor 3	Base Factor 4	Base Factor 5	Base Factor 6	Base Factor 7	Unmapped
Boot Factor 1	100	0	0	0	0	0	0	0
Boot Factor 2	0	100	0	0	0	0	0	0
Boot Factor 3	0	0	100	0	0	0	0	0
Boot Factor 4	0	0	0	100	0	0	0	0
Boot Factor 5	0	0	0	0	100	0	0	0
Boot Factor 6	0	0	0	0	0	100	0	0
Boot Factor 7	0	0	0	0	0	0	100	0

Figure 69. Summary of constrained run and error estimates.

8.3.1   Data Set Development

The concentration data for this analysis were downloaded from the EPA Air Quality System. Speciated volatile organic compound (VOC) data from 3-hr samples collected at the Baton Rouge PAMS site during June–August 2005 and June–September 2006 (307 samples) were downloaded for potential inclusion in PMF. Uncertainties are not regularly reported for PAMS VOC data. For this analysis, initial uncertainties were set for each species and sample at 15% of the concentration, unless the value was below detection, where the concentration was MDL/2 and uncertainty was (5/6)*MDL (Table 7).

Table 7. Baton Rouge Example – Summary of PMF input information.

Data Files				****** Base Run Summary ******		
Concentration file:	Dataset-BatonRouge-con.csv			Number of base runs:	20	
Uncertainty file:	Dataset-BatonRouge-unc.csv			Base random seed:	25	
				Number of factors:	4	
Excluded Samples				Extra modeling uncertainty (%):	0	
none						
**** Input Data Statistics ****						
Species	Category	S/N		Species	Category	S/N
124-Trimethylbenze	Bad	5.46		M-Ethyltoluene	Bad	5.53
224-Trimethylpenta	Strong	5.67		N-Butane	Strong	5.67
234-Trimethylpenta	Bad	5.55		N-Decane	Weak	5.20
23-Dimethylbutane	Bad	5.51		N-Heptane	Strong	5.67
23-Dimethylpentane	Bad	5.48		N-Hexane	Weak	5.62
2-Methylheptane	Weak	5.08		N-Nonane	Weak	5.43
3-Methylhexane	Bad	5.65		N-Octane	Weak	5.58
3-Methylpentane	Bad	5.62		N-Pentane	Weak	5.67
Acetylene	Strong	5.67		N-Propylbenzene	Bad	3.76
Benzene	Strong	5.67		N-Undecane	Bad	5.03
Cis-2-Butene	Bad	3.28		O-Ethyltoluene	Weak	5.00
Cis-2-Pentene	Bad	5.10		O-Xylene	Strong	5.67
Ethane	Bad	5.67		Propane	Strong	5.67
Ethylbenzene	Strong	5.67		Propylene	Weak	5.67
Ethylene	Weak	5.67		Styrene	Bad	4.95
Isobutane	Weak	5.67		Toluene	Strong	5.67
Isopentane	Weak	5.67		Trans-2-Butene	Bad	3.16
Isoprene	Bad	5.56		Trans-2-Pentene	Bad	5.43
Isopropylbenzene	Bad	2.32		Unidentified	Bad	1.00
M_P Xylene	Bad	5.67		TNMOC	Weak	0.75
M-Diethylbenzene	Bad	2.66				

8.3.2   Analyze Input Data
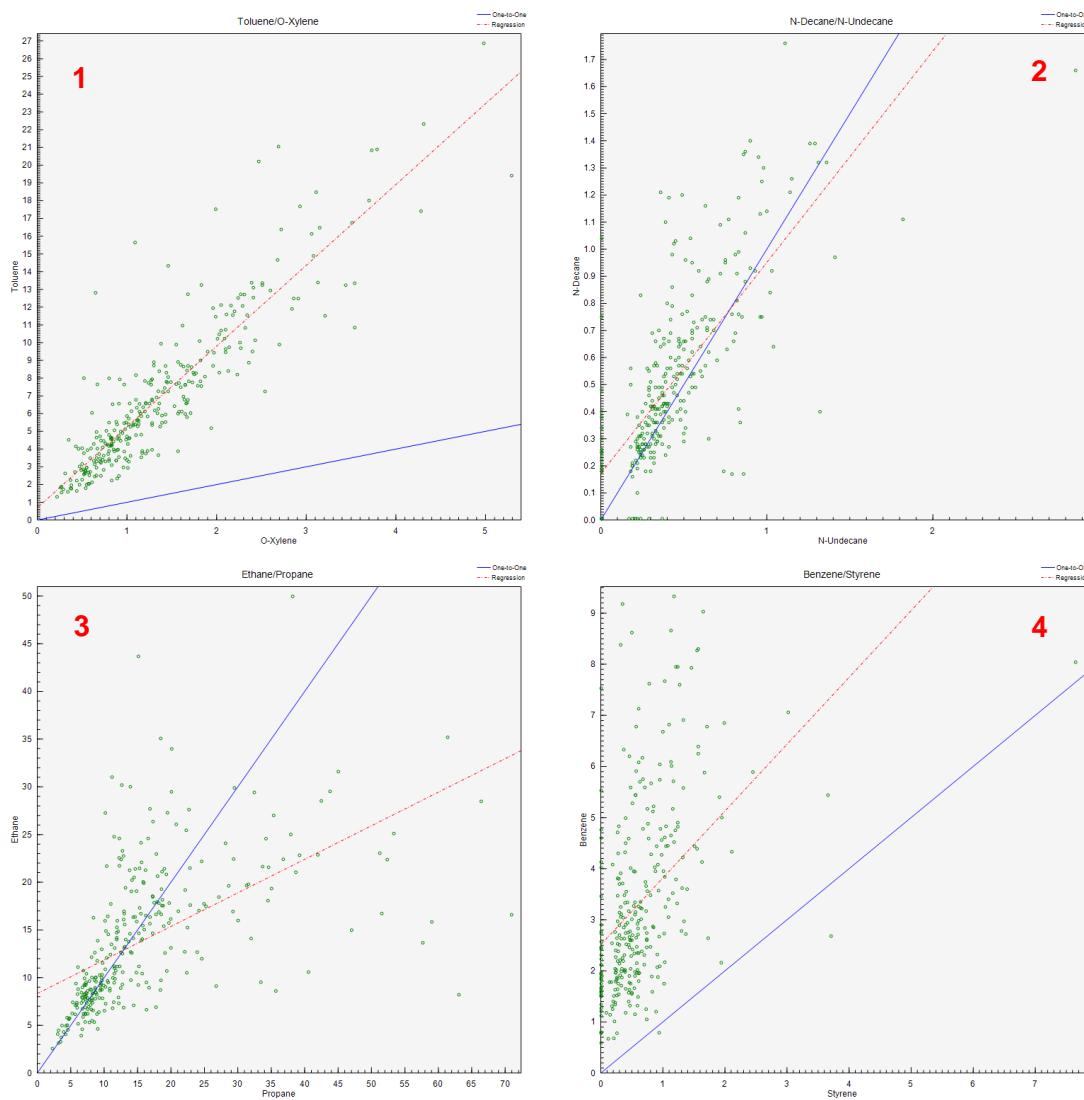
Characterizing Species (Concentration/Uncertainty and Concentration Time Series)

S/N ratios are not as useful in this analysis because all species were given a set uncertainty; therefore, species categorizations will be evaluated based on residuals and observed/predicted statistics after the initial base runs. Species with greater relative uncertainties were categorized as "Bad" and excluded from the analysis. For the initial run, all included species were categorized as "Strong" and all 21 species, including total non-methane organic compounds (TNMOC), were used.

Relationships Between Species (Concentration Scatter Plot)

Scatter plots between species are examined to evaluate relationships between the species that may indicate a common source. In the Baton Rouge data set, expected relationships between gasoline mobile source species, such as toluene and o-xylene (Figure 70, 1) and heavy-duty vehicle mobile source species, such as n-decane and n-undecane (Figure 70, 2) are indicated.

Ethane and propane (Figure 70, 3) show some evidence of two source influences that have different ethane and propane ratios, potentially indicating a mix of fresh sources from petrochemical processing/natural gas use and aged carryover from other areas. Benzene and styrene (Figure 70, 4), often mobile source-dominated species, were not well-correlated with other mobile source species; this lack of correlation is likely due to emissions of these species from the several large petrochemical sources in the area.

Figure 70. Relationships between ambient concentrations of various species.

Excluding Samples and Species (Concentration Time Series)

Time series of each pollutant were examined for extreme events and/or noticeable step changes in concentrations that should be removed from the analysis. Step changes (e.g., differences due to changes in laboratory analytical technique) may be mistakenly identified as separate sources of the species. If samples are removed due to unusual events in various

species, further data analysis outside EPA PMF could be used to confirm whether the data are real and informative.

8.3.3 Base Model Runs

Initial Model Parameters (Model Execution)

Initially, 20 base runs with 4 factors and a seed of 25 were explored. In this iteration, the Q-values varied by several hundred units, indicating the solution may not be stable. The species and categories are shown in Table 8. A number of the species categories were changed to "Weak" after the residuals and plots were evaluated as described below. Strong/Weak is shown in the Category column of Table 8 for species that were changed.

Table 8. VOC species categories.

Species	Category
1,2,4-Trimethylbenzene	Bad
2,2,4-Trimethylpentane	Strong
2,3,4-Trimethylpentane	Bad
2,3-Dimethylbutane	Bad
2,3-Dimethylpentane	Bad
2-Methylheptane	Strong/Weak
3-Methylhexane	Bad
3-Methylpentane	Bad
Acetylene	Strong
Benzene	Strong
Cis-2-Butene	Bad
Cis-2-Pentene	Bad
Ethane	Bad
Ethylbenzene	Strong
Ethylene	Strong/Weak
Isobutane	Strong/Weak
Isopentane	Strong/Weak
Isoprene	Bad
Isopropylbenzene	Bad
M_P Xylene	Bad
M-Diethylbenzene	Bad
M-Ethyltoluene	Bad
N-Butane	Strong
N-Decane	Strong/Weak
N-Heptane	Strong
N-Hexane	Strong/Weak
N-Nonane	Strong/Weak

Species	Category
N-Octane	Strong/Weak
N-Pentane	Strong/Weak
N-Propylbenzene	Bad
N-Undecane	Bad
O-Ethyltoluene	Strong/Weak
O-Xylene	Strong
Propane	Strong
Propylene	Strong/Weak
Styrene	Bad
Toluene	Strong
Trans-2-Butene	Bad
Trans-2-Pentene	Bad
Unidentified	Bad
TNMOC	Weak

8.3.4   Base Model Run Results

Model Reconstruction (Obs/Pred Scatter Plots, Obs/Pred Time Series)

Residuals of the species were analyzed and the histograms of scaled residuals (after selecting autoscale) are shown for benzene, which had a good fit, and poorly fit ethylene in Figure 71. In addition, the observed vs. predicted scatter plots and time series are shown in Figure 72 and Figure 73, respectively. Since PAMS data are only collected during the summer, the time-series plots have a missing time period during fall through spring. The scatter plots and the time series also show the difference between the observed and predicted concentrations. The poorly fit species have scaled residuals greater than 3.0 and the peak observations are not fit in the scatter or time-series plots. Species with a number of scaled residuals above 4 have peak concentrations that were not fit by PMF: 2-methylheptane, ethylene, isobutane, isopentane, n-decane, n-hexane, n-nonane, n-octane, n-pentane, o-ethyltoluene, and propylene. The category for these species was set to "Weak."

Factor Identification (Profiles/Contributions, Aggregate Contributions)

The base run was re-run and profiles and contributions were examined to identify factors. Measured profiles were used to support the identification of the factors and the factor names have been added to Figure 74 by right-clicking in Profiles/Contributions and naming the factors via the "Factor Name" option.
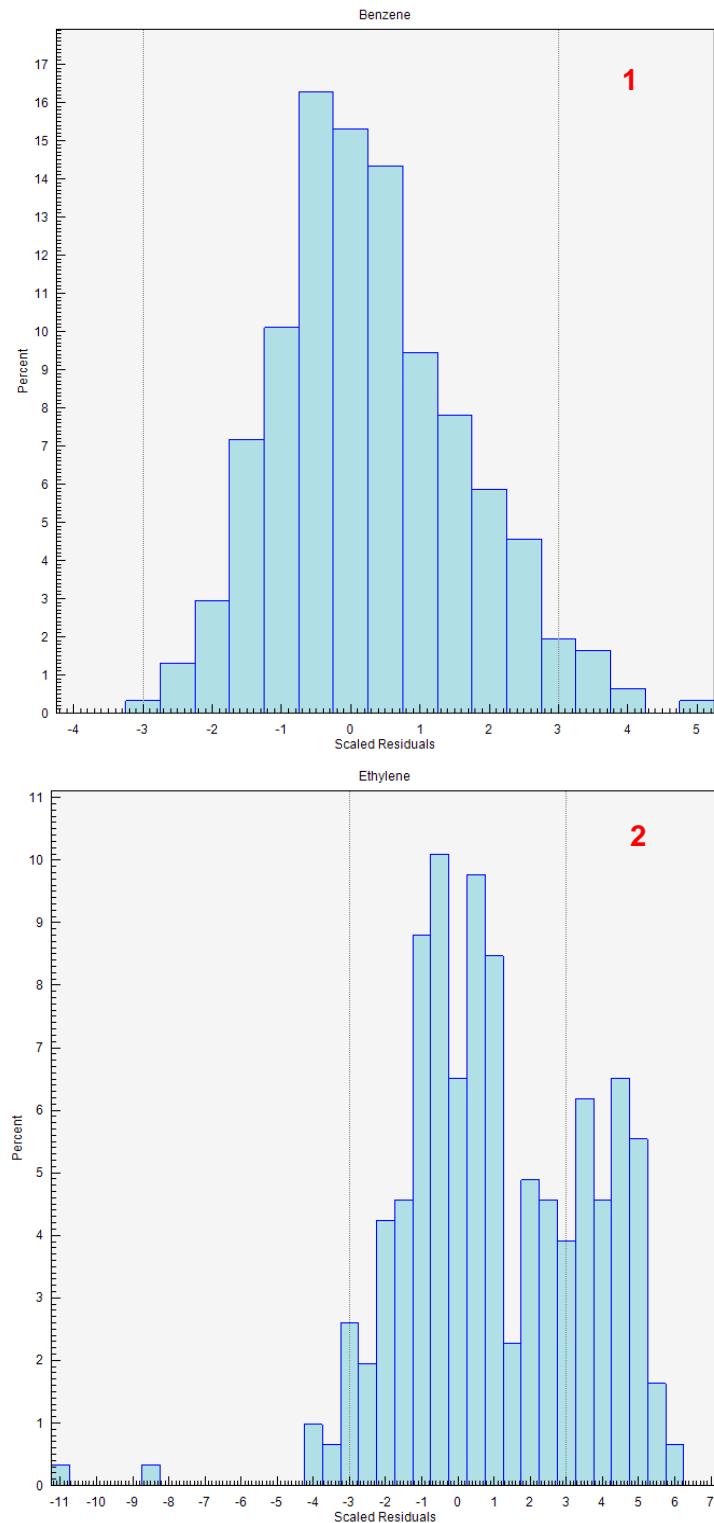
Figure 71. Histogram of scaled residuals for benzene (1) and ethylene (2).
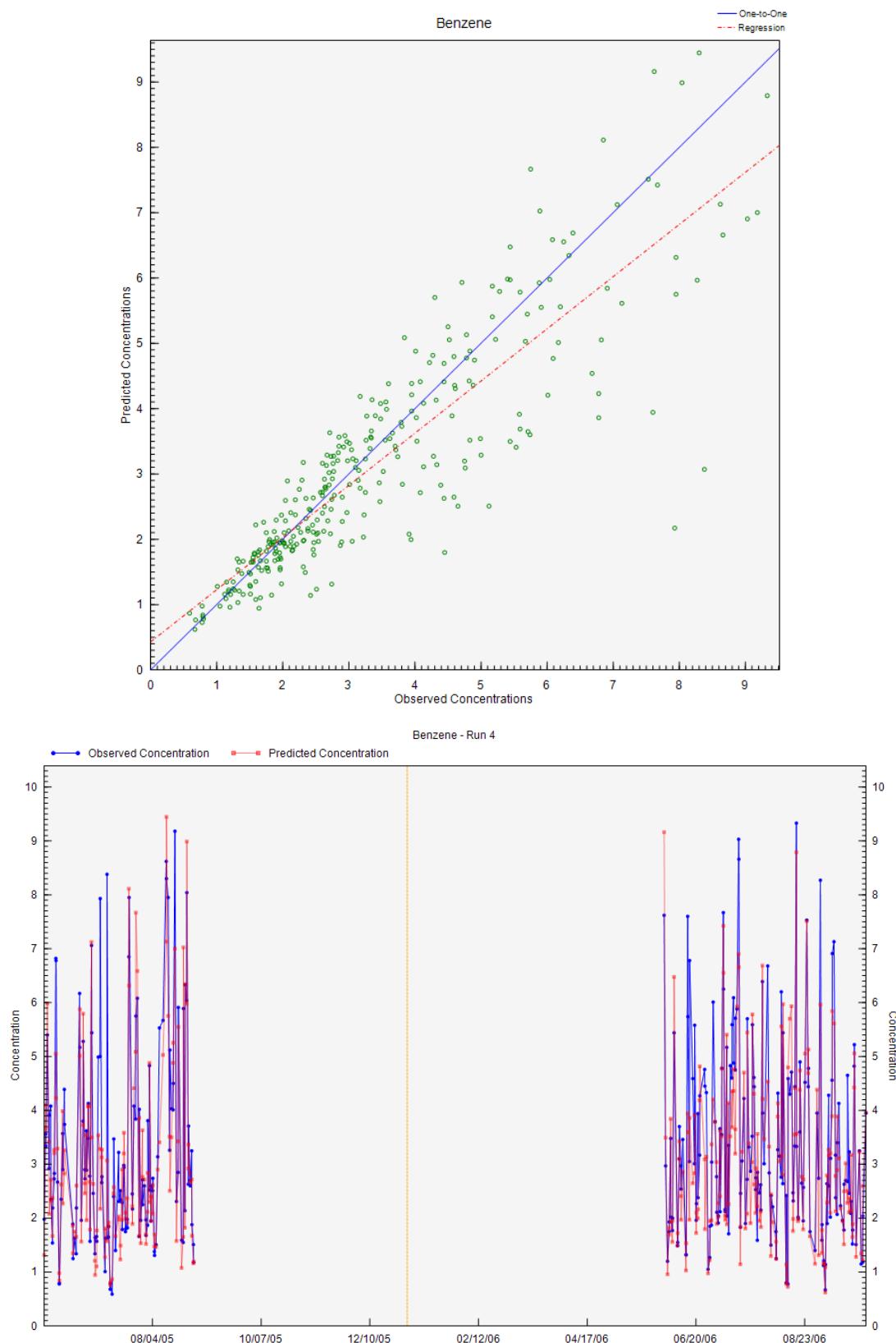
Figure 72. Observed/predicted plots for benzene.
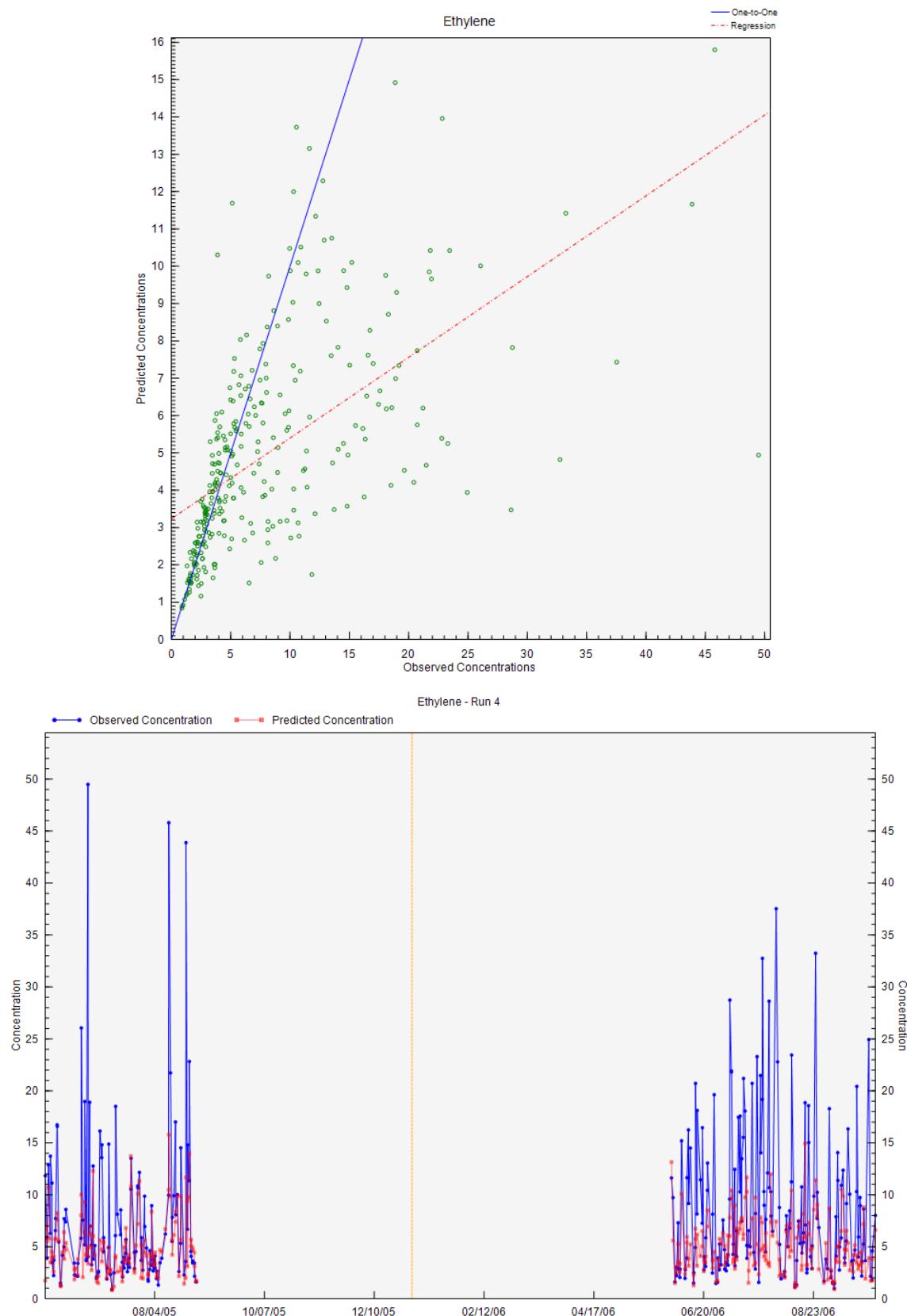
Figure 73. Observed/predicted plots for ethylene.
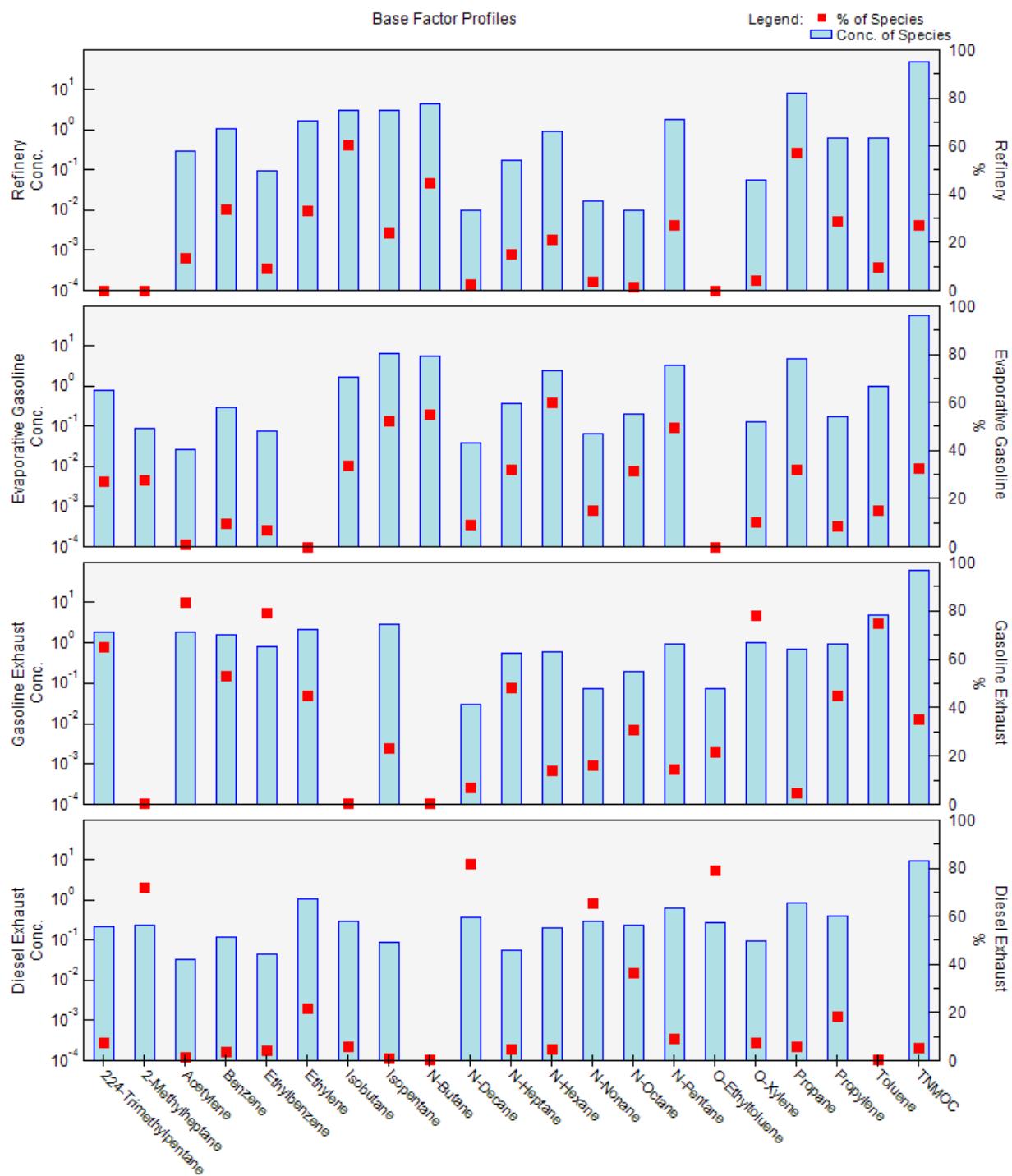
Figure 74. VOC factor profiles.

The PMF results were compared to measured profiles using the first and second columns from Fujita (2001), shown in Figure 75. The n-decane levels in the diesel exhaust profile (Tu_MchHD) are high compared to the vehicle emissions (Exh_J) and Figure 76 shows the factor fingerprint plot for which n-decane is predominately associated with the diesel factor. The acetylene contributions to sources will be discussed in later in this example. Acetylene is

predominately associated with vehicle emissions and has a small contribution to gasoline vapor. It is also present in the industrial source and diesel.

Table 1
Source profiles applied in the Paso del Norte Ozone Study (volume percent of sum of 55 PAMS target species)[a] MS target species)[a]

PROFILE		Juarez Vehicle Emissions (Exh_J)	Diesel Exhaust (Tu_MchHD)	Adjusted Propane Bus (Exh_Pba)	Liquid Gasoline (ME75R25P)	Gasoline Vapor (Maga_HS)	Liquefied Petroleum Gas (Prop_J)	Commercial Natural Gas (CNG_J)	Industrial (ChevFC)	Industrial (Zenco)	Surface Coating (COATcomp)
ethene		5.19 ± 0.78	9.87 ± 2.81	6.30 ± 0.95	0.00 ± 0.13	0.00 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	2.34 ± 0.35	2.65 ± 0.96	0.00 ± 0.44
ethane	*	0.95 ± 0.14	1.18 ± 0.68	2.08 ± 0.51	0.00 ± 0.13	0.00 ± 0.12	1.51 ± 0.30	78.06 ± 11.72	22.80 ± 14.90	9.98 ± 1.51	0.00 ± 0.44
acetylene	*	5.29 ± 1.47	2.55 ± 1.71	3.72 ± 0.65	0.00 ± 0.13	0.17 ± 0.13	0.00 ± 0.20	0.00 ± 0.20	0.41 ± 0.20	3.07 ± 1.38	0.00 ± 0.44
Propene		2.00 ± 0.73	3.99 ± 0.93	5.01 ± 2.39	0.00 ± 0.13	0.11 ± 0.12	0.13 ± 0.21	0.00 ± 0.20	3.30 ± 1.40	0.98 ± 0.23	0.00 ± 0.44
n-propane	*	3.06 ± 0.46	2.22 ± 1.05	57.90 ± 8.69	0.02 ± 0.01	1.44 ± 0.19	85.60 ± 12.86	15.75 ± 2.37	21.50 ± 3.24	21.49 ± 3.24	0.00 ± 0.44
isobutane	*	0.86 ± 0.13	0.27 ± 0.29	3.53 ± 1.74	0.53 ± 0.15	4.80 ± 0.50	2.47 ± 0.42	2.26 ± 0.39	4.27 ± 1.13	3.51 ± 2.20	0.00 ± 0.44
1-butene		0.00 ± 0.20	2.95 ± 0.55	0.00 ± 0.20	0.00 ± 0.01	0.06 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	1.05 ± 0.25	0.45 ± 0.13	0.00 ± 0.44
n-butane	*	3.72 ± 0.56	0.64 ± 1.58	17.77 ± 10.14	2.52 ± 0.60	19.82 ± 1.99	9.58 ± 4.71	3.76 ± 0.60	11.33 ± 6.42	3.54 ± 0.68	0.00 ± 0.44
t-2-Butene		0.96 ± 0.43	0.24 ± 0.44	1.65 ± 2.34	0.04 ± 0.01	2.19 ± 0.25	0.00 ± 0.20	0.00 ± 0.20	0.31 ± 0.12	0.09 ± 0.11	0.00 ± 0.44
c-2-butene		0.30 ± 0.05	0.29 ± 0.11	0.06 ± 0.01	0.01 ± 0.01	1.91 ± 0.23	0.00 ± 0.20	0.00 ± 0.20	0.28 ± 0.06	0.56 ± 0.79	0.00 ± 0.44
isopentane	*	8.80 ± 1.32	1.31 ± 3.43	0.57 ± 0.28	5.92 ± 1.57	36.40 ± 3.64	0.34 ± 0.21	1.30 ± 0.28	3.54 ± 1.14	3.41 ± 0.88	0.00 ± 0.00
1-pentene		0.42 ± 0.06	0.89 ± 0.21	0.05 ± 0.01	0.28 ± 0.07	0.98 ± 0.16	0.00 ± 0.20	0.00 ± 0.20	0.43 ± 0.29	0.18 ± 0.12	0.00 ± 0.44
n-pentane	*	4.64 ± 0.70	1.52 ± 1.27	0.00 ± 0.20	3.65 ± 0.69	11.91 ± 1.20	0.10 ± 0.20	1.20 ± 0.27	3.93 ± 0.61	3.77 ± 1.49	0.00 ± 0.44
isoprene	*	0.23 ± 0.04	0.00 ± 0.30	0.00 ± 0.20	0.00 ± 0.13	0.10 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.03 ± 0.05	0.18 ± 0.12	0.00 ± 0.44
t-2-Pentene		0.97 ± 0.15	0.36 ± 0.38	0.00 ± 0.20	0.75 ± 0.13	0.00 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.28 ± 0.20	0.25 ± 0.12	0.00 ± 0.10
c-2-pentene		0.49 ± 0.07	0.29 ± 0.19	0.00 ± 0.20	0.43 ± 0.08	0.93 ± 0.15	0.00 ± 0.20	0.00 ± 0.20	0.15 ± 0.10	0.13 ± 0.12	0.00 ± 0.44
2,2-dimethylbutane	*	0.37 ± 0.06	2.63 ± 0.98	0.00 ± 0.20	0.44 ± 0.07	1.16 ± 0.17	0.00 ± 0.20	0.00 ± 0.20	0.08 ± 0.08	0.16 ± 0.12	0.00 ± 0.44
cyclopentane	*	0.46 ± 0.07	0.32 ± 0.23	0.00 ± 0.20	0.46 ± 0.07	0.00 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.46 ± 0.08	0.18 ± 0.12	0.00 ± 0.44
2,3-dimethylbutane	*	1.19 ± 0.18	0.32 ± 0.62	0.00 ± 0.20	1.40 ± 0.09	0.00 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.37 ± 0.22	0.45 ± 0.13	0.00 ± 0.44
2-methylpentane	*	3.36 ± 0.50	1.97 ± 1.08	0.00 ± 0.20	3.97 ± 0.23	4.71 ± 0.49	0.00 ± 0.20	0.00 ± 0.20	1.39 ± 0.39	1.47 ± 0.25	0.01 ± 0.01
3-methylpentane	*	2.09 ± 0.31	0.91 ± 0.68	0.00 ± 0.20	2.61 ± 0.14	2.52 ± 0.28	0.00 ± 0.20	0.00 ± 0.20	0.81 ± 0.27	0.95 ± 0.18	0.01 ± 0.07
2-methyl-1-pentene		0.39 ± 0.06	0.22 ± 0.19	0.00 ± 0.20	0.00 ± 0.10	0.00 ± 0.10	0.00 ± 0.20	0.00 ± 0.20	0.10 ± 0.05	0.08 ± 0.11	0.00 ± 0.10
n-hexane	*	2.68 ± 0.40	0.96 ± 0.60	0.00 ± 0.20	3.46 ± 0.17	2.03 ± 0.24	0.00 ± 0.20	0.00 ± 0.20	1.95 ± 0.30	1.24 ± 0.22	0.00 ± 0.44
Methylcyclopentane	*	1.78 ± 0.27	0.62 ± 0.48	0.00 ± 0.20	2.27 ± 0.04	0.87 ± 0.15	0.00 ± 0.20	0.00 ± 0.20	1.13 ± 0.23	0.70 ± 0.17	0.04 ± 0.06
2,4-dimethylpentane	*	1.29 ± 0.19	0.36 ± 0.27	0.00 ± 0.20	1.58 ± 0.67	0.52 ± 0.14	0.00 ± 0.20	0.00 ± 0.20	0.65 ± 0.53	0.43 ± 0.13	0.01 ± 0.07
benzene	*	4.50 ± 0.68	3.19 ± 1.52	0.00 ± 0.20	2.64 ± 1.03	0.76 ± 0.14	0.00 ± 0.20	0.00 ± 0.20	1.23 ± 0.43	1.52 ± 0.38	0.00 ± 0.44
cyclohexane	*	0.90 ± 0.14	0.23 ± 0.09	0.00 ± 0.20	1.23 ± 0.02	0.20 ± 0.13	0.00 ± 0.20	0.00 ± 0.20	0.57 ± 0.33	0.45 ± 0.22	0.15 ± 0.23
2-methylhexane	*	1.51 ± 0.23	0.00 ± 0.11	0.00 ± 0.20	2.10 ± 0.15	0.48 ± 0.13	0.00 ± 0.20	0.00 ± 0.20	0.52 ± 0.09	0.87 ± 0.17	0.28 ± 1.22
2,3-dimethylpentane	*	1.98 ± 0.30	0.91 ± 0.37	0.00 ± 0.20	3.34 ± 0.22	0.31 ± 0.13	0.00 ± 0.20	0.00 ± 0.20	1.17 ± 0.92	0.75 ± 0.16	0.12 ± 0.21
3-methylhexane	*	1.67 ± 0.25	2.31 ± 1.29	0.00 ± 0.20	2.38 ± 0.12	0.48 ± 0.13	0.00 ± 0.20	0.00 ± 0.20	0.56 ± 0.18	0.59 ± 0.14	0.34 ± 0.45
2,2,4-trimethylpentane	*	3.04 ± 0.30	1.48 ± 1.26	0.00 ± 0.20	4.38 ± 0.69	2.13 ± 0.25	0.00 ± 0.20	0.00 ± 0.20	1.49 ± 1.60	3.92 ± 2.37	0.00 ± 0.44

[a] Profiles consisting of 100 percent isoprene (Biogenic) and 100 percent unidentified hydrocarbons (UNID) were also applied.

* Fitting species in both auto-GC and canister samples, c - fitting species in canister samples only.

Table 1 (*Continued*)
Source profiles applied in the Paso del Norte Ozone Study (volume percent of sum of 55 PAMS target species)[a] IS target species)[a]

PROFILE		Exh_J	Tu_MchHD	Exh_PBa	ME75R25P	Maga_HS	Prop_J	CNG_J	ChevFC	Zenco	COATcomp
n-heptane	*	1.62 ± 0.24	0.57 ± 0.30	0.00 ± 0.20	2.01 ± 0.10	0.37 ± 0.13	0.00 ± 0.20	0.00 ± 0.20	1.28 ± 0.55	0.61 ± 0.15	1.57 ± 4.82
methylcyclohexane	*	0.85 ± 0.13	0.44 ± 0.27	0.00 ± 0.20	0.84 ± 0.04	0.15 ± 0.13	0.00 ± 0.20	0.00 ± 0.20	0.74 ± 0.39	0.35 ± 0.15	2.61 ± 9.12
2,3,4-trimethylpentane	*	1.28 ± 0.19	0.32 ± 0.44	0.00 ± 0.20	1.87 ± 0.28	0.48 ± 0.13	0.00 ± 0.20	0.00 ± 0.20	0.71 ± 0.51	0.63 ± 0.15	0.05 ± 0.10
toluene	*	10.34 ± 1.55	4.50 ± 3.64	0.00 ± 0.20	15.04 ± 0.69	1.26 ± 0.18	0.00 ± 0.20	0.00 ± 0.20	2.23 ± 0.89	19.41 ± 2.92	6.91 ± 6.86
2-methylheptane	*	0.55 ± 0.08	0.00 ± 0.21	0.00 ± 0.20	0.80 ± 0.05	0.06 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.38 ± 0.07	0.24 ± 0.12	1.36 ± 2.34
3-methylheptane	*	0.70 ± 0.11	0.44 ± 0.22	0.00 ± 0.20	0.81 ± 0.05	0.05 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.23 ± 0.06	0.23 ± 0.12	0.96 ± 1.62
n-octane	*	0.53 ± 0.08	0.31 ± 0.17	0.00 ± 0.20	0.76 ± 0.04	0.05 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.95 ± 0.45	0.34 ± 0.21	3.44 ± 5.30
ethylbenzene		2.05 ± 0.31	2.85 ± 1.98	0.00 ± 0.20	3.35 ± 0.18	0.06 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.44 ± 0.08	0.90 ± 0.18	3.82 ± 2.85
mp-xylene		6.11 ± 0.92	10.99 ± 6.84	0.00 ± 0.20	10.70 ± 0.79	0.19 ± 0.13	0.00 ± 0.20	0.00 ± 0.20	0.88 ± 0.14	2.15 ± 0.51	15.40 ± 10.58
styrene		0.46 ± 0.07	1.85 ± 0.73	0.00 ± 0.20	0.28 ± 0.02	0.02 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.26 ± 0.10	1.80 ± 0.76	0.01 ± 0.03
o-xylene		2.46 ± 0.37	3.73 ± 2.27	0.00 ± 0.20	3.95 ± 0.28	0.06 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.54 ± 0.09	0.98 ± 0.20	7.18 ± 4.36
n-nonane	*	0.19 ± 0.03	1.12 ± 0.35	0.00 ± 0.20	0.30 ± 0.07	0.02 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.74 ± 0.31	0.40 ± 0.13	4.54 ± 2.30
isopropylbenzene		0.29 ± 0.04	0.33 ± 0.17	0.00 ± 0.20	0.25 ± 0.02	0.00 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.16 ± 0.07	0.11 ± 0.11	0.63 ± 0.57
n-propylbenzene		0.58 ± 0.09	1.06 ± 0.68	0.00 ± 0.20	0.88 ± 0.05	0.01 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.13 ± 0.05	0.27 ± 0.12	1.46 ± 1.46
m-ethyltoluene		1.71 ± 0.26	4.15 ± 2.88	0.00 ± 0.20	2.75 ± 0.18	0.05 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.26 ± 0.06	0.61 ± 0.15	0.00 ± 0.44
p-ethyltoluene		0.37 ± 0.52	1.40 ± 0.73	0.00 ± 0.20	1.26 ± 0.08	0.00 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.00 ± 0.05	0.00 ± 0.11	4.51 ± 1.66
1,3,5-trimethylbenzene		0.97 ± 0.15	2.09 ± 1.09	0.00 ± 0.20	1.43 ± 0.10	0.04 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.21 ± 0.06	0.57 ± 0.14	4.14 ± 2.64
o-ethyltoluene		0.66 ± 0.10	2.00 ± 1.12	0.00 ± 0.20	0.99 ± 0.06	0.00 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.12 ± 0.05	0.24 ± 0.12	0.00 ± 0.44
1,2,4-trimethylbenzene		2.74 ± 0.41	7.46 ± 4.64	0.00 ± 0.20	4.35 ± 0.27	0.07 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.37 ± 0.07	1.07 ± 0.21	11.13 ± 4.31
n-decane	c	0.27 ± 0.05	2.62 ± 0.64	0.00 ± 0.20	0.14 ± 0.02	0.02 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.55 ± 0.11	0.70 ± 0.30	18.12 ± 6.01
1,2,3-trimethylbenzene		0.00 ± 0.20	1.67 ± 1.04	0.00 ± 0.20	0.00 ± 0.10	0.00 ± 0.10	0.00 ± 0.20	0.00 ± 0.20	0.00 ± 0.20	0.00 ± 0.20	0.00 ± 0.10
m-diethylbenzene		0.00 ± 0.01	0.00 ± 0.11	0.00 ± 0.20	0.32 ± 0.01	0.00 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.00 ± 0.05	0.00 ± 0.11	0.39 ± 0.42
p-diethylbenzene		0.00 ± 0.20	0.00 ± 0.11	0.00 ± 0.20	0.00 ± 0.10	0.00 ± 0.10	0.00 ± 0.20	0.00 ± 0.20	0.00 ± 0.20	0.00 ± 0.20	0.00 ± 0.10
n-undecane	c	0.17 ± 0.03	5.32 ± 1.04	0.00 ± 0.20	0.49 ± 0.03	0.02 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	0.37 ± 0.07	0.42 ± 0.13	10.81 ± 8.14
Others		11.92 ± 1.19	8.51 ± 0.85	0.00 ± 0.20	27.01 ± 1.55	0.00 ± 0.12	0.00 ± 0.20	0.00 ± 0.20	7.25 ± 1.28	8.99 ± 0.91	0.00 ± 0.10
Unidentified	c	16.11 ± 2.43	0.40 ± 0.04	0.00 ± 0.20	0.00 ± 0.13	0.00 ± 0.12	0.27 ± 0.31	0.00 ± 0.20	13.37 ± 4.25	39.41 ± 5.93	186.46 ± 22.67
NMHC		128.03 ± 12.80	108.91 ± 10.89	100.00 ± 15.00	127.01 ± 12.70	100.00 ± 10.00	100.00 ± 15.00	100.00 ± 15.00	120.63 ± 12.06	148.40 ± 14.84	286.46 ± 28.65
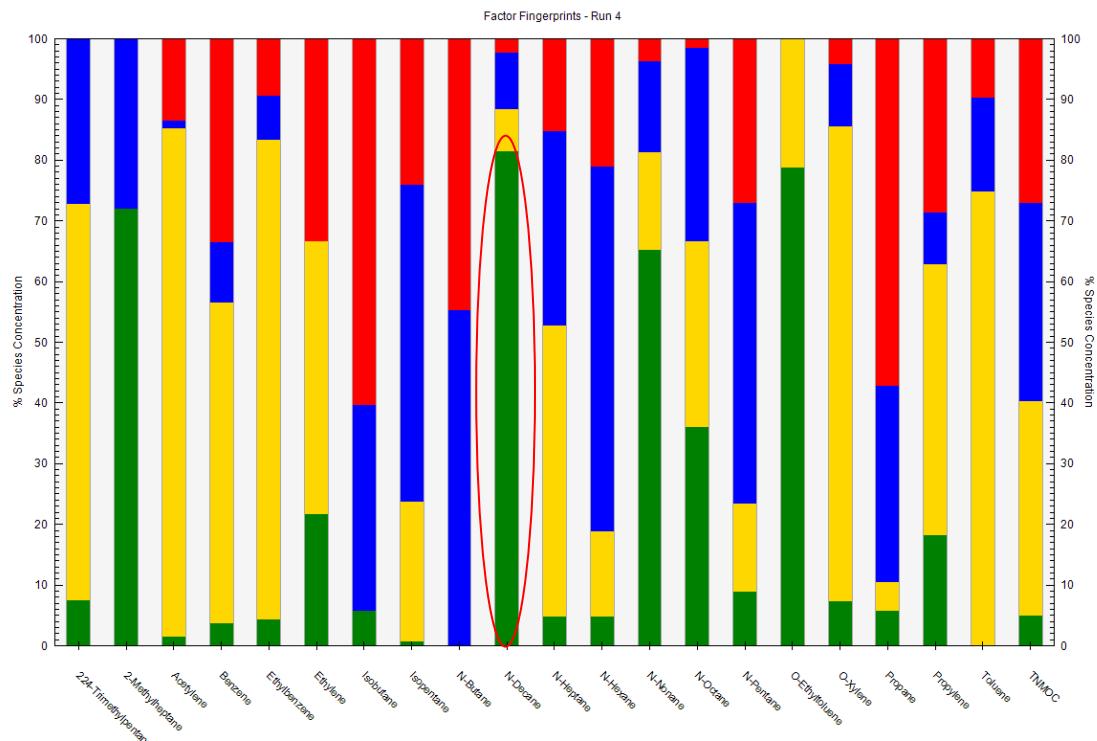
[a] Profiles consisting of 100 percent isoprene (Biogenic) and 100 percent unidentified hydrocarbons (UNID) were also applied.

* Fitting species in both auto-GC and canister samples, c - fitting species in canister samples only.

Figure 75. Measured VOC profile information. Source: Fujita (2001).

Figure 76. Factor fingerprint plot for VOCs.

Rotations (G-Space Plots)
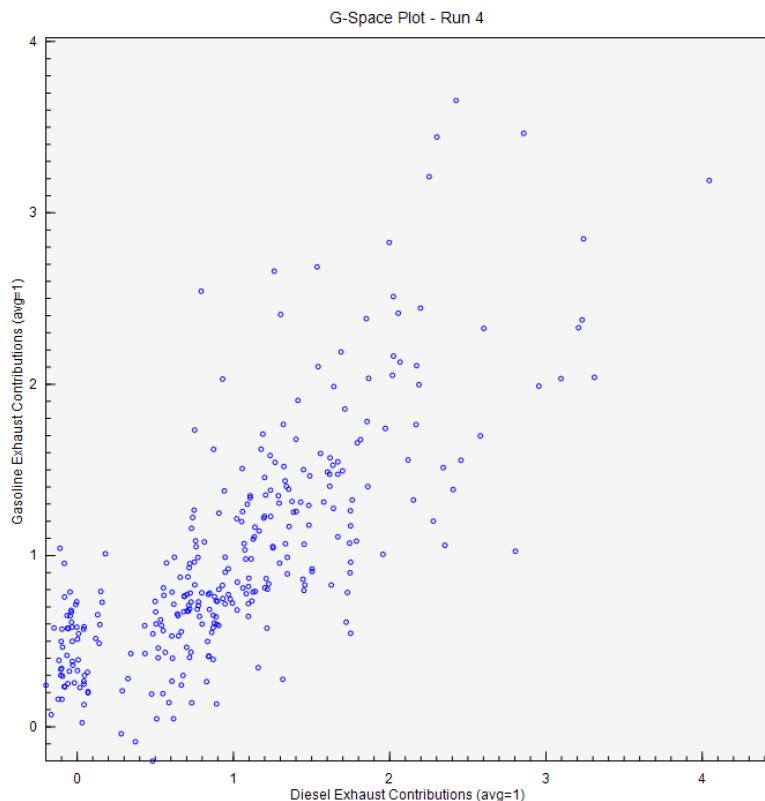
The G-space plot of the motor vehicle and the diesel exhaust source contributions had a weak linear relationship (Figure 77). This may indicate that the diesel motor vehicle source may be mixed with the motor vehicle source, or another source of diesel combustion may be present. The other G-space plot pairings showed the points were distributed across the solution space between the axes. Fpeak should be investigated to determine whether a rotation moves points to the axes.

Species Distribution (Factor Pie Chart)
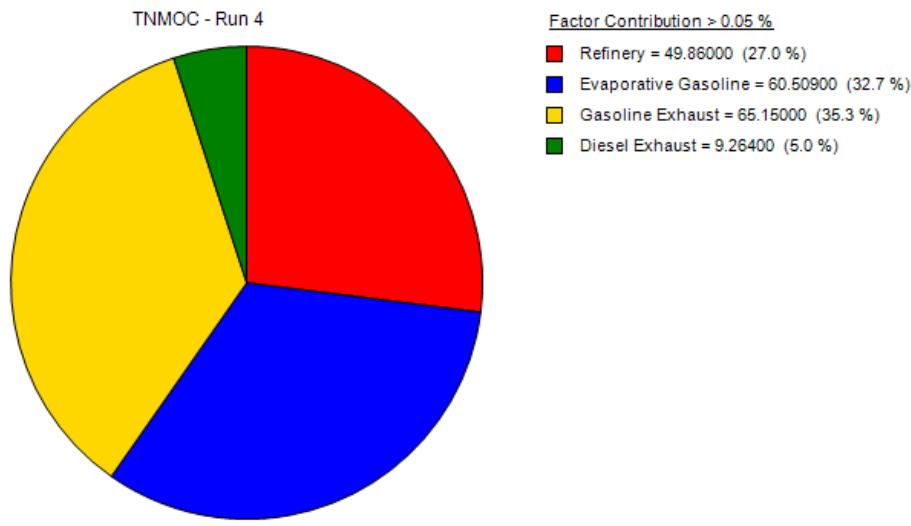
The total variable (TNMOC) was mainly contributed to by motor vehicle exhaust and gasoline vapor. The industrial component was also a major contributor, as shown in Figure 78.

8.3.5 Fpeak

Examination of the Fpeak G-space plots of motor vehicle exhaust vs. gasoline vapor showed that some optimization might be gained using an Fpeak of -1.0. The focus of this example is to demonstrate source profile constraints, so the Fpeak result will not be discussed further. The base, Fpeak, and constrained model results should be compared to determine whether the rotational tools and constraints provide a different interpretation of the factors and contributions.

Figure 77. G-Space plot of motor vehicle and diesel exhaust.

Figure 78. Apportionment of TNMOC to factors resolved in the initial 4-factor base run.

Error Estimate Summary
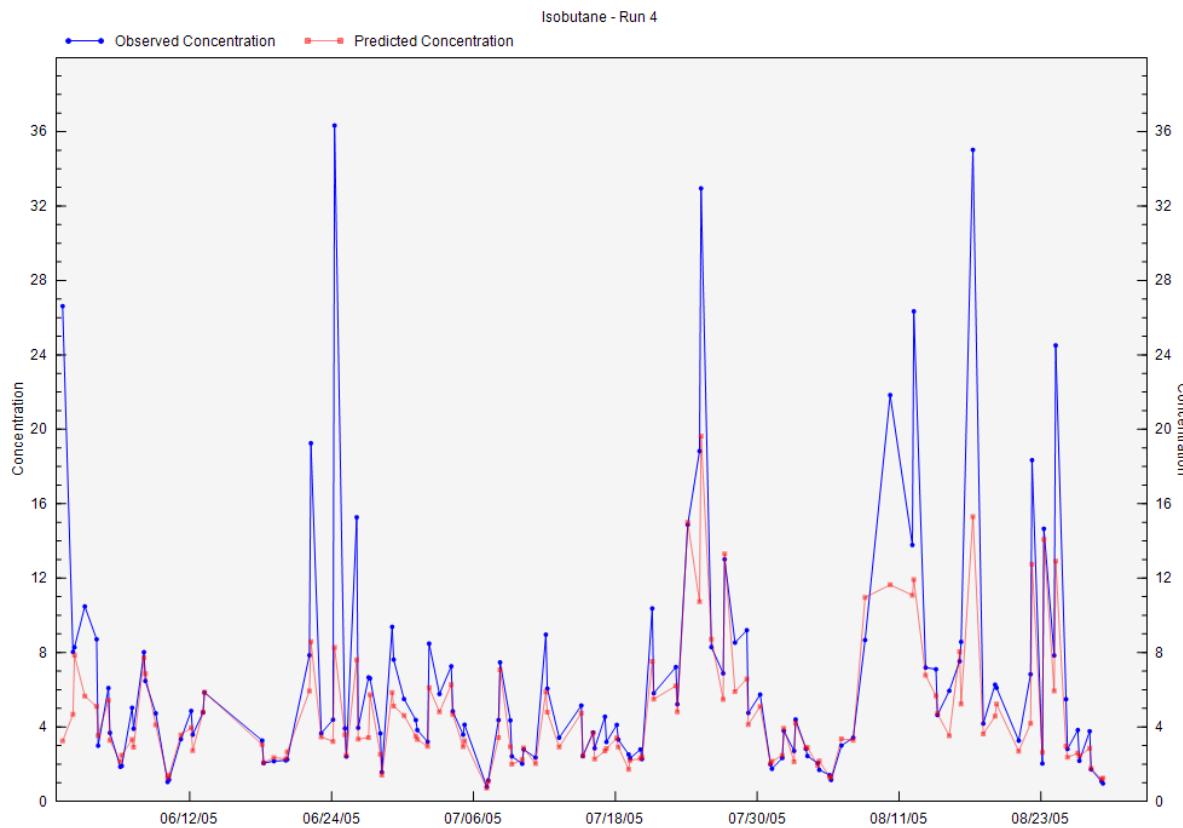
As shown in Table 9, not all of the base factors were mapped to the boot factors and the number of factors that were not correctly mapped is approximately 80%, which is relatively stable. The unmapped factors are due to the combination of the high variability in the data and PMF not fitting all of the spikes in the data (Figure 79). All of the "Strong" species were selected for the BS-DISP error estimation. The number of DISP swaps is zero and the BS-DISP swaps are distributed across three factors. The number of swaps in BS-DISP is relatively high and the BS results and model fit statistics need to be evaluated before reporting results.

Table 9. Base run boostrap mapping.

BS-DISP Diagnostics				
# of Cases:	87			
Largest Decrease in Q:	-6.846000195			
% dQ:	-0.138746462			
# of Decreases in Q:	0			
# of Swaps in Best Fit:	1			
# of Swaps in DISP:	13			
Swaps by Factor:	1	3	4	0

DISP Diagnostics				
Error Code:	0			
Largest Decrease in Q:	0			
% dQ:	0			
Swaps by Factor:	0	0	0	0

BS Mapping					
	Base Factor 1	Base Factor 2	Base Factor 3	Base Factor 4	Unmapped
Boot Factor 1	80	8	8	4	0
Boot Factor 2	0	92	6	2	0
Boot Factor 3	0	0	100	0	0
Boot Factor 4	0	0	13	87	0
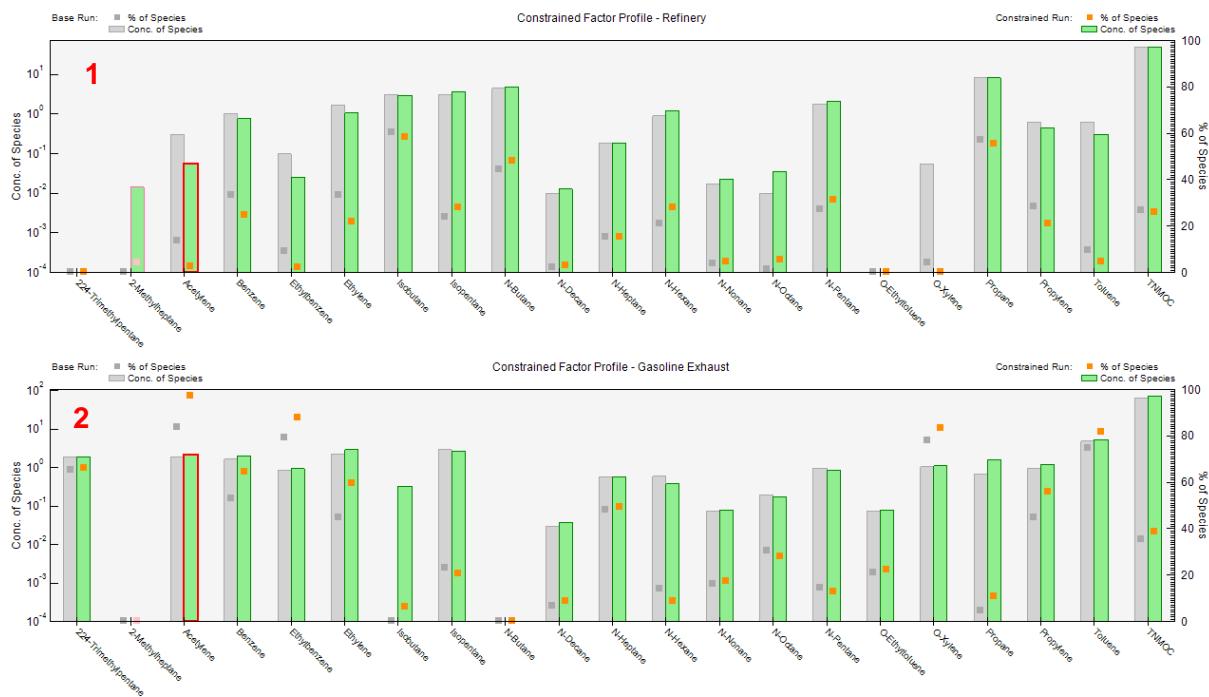
Figure 79. Observed vs. Predicted Time Series for refinery species.

8.3.6 Constrained Model Runs

Constraints were used to determine if the acetylene is strongly associated with the industrial source because acetylene is a key tracer for motor vehicle exhaust. In the base run, 84 and 14 percent of the acetylene was associated with the gasoline exhaust and refinery factors, respectively. Acetylene was selected in the refinery factor using toggle constraints and it was constrained using "Pull Down Maximally" with a 1% dQ and acetylene was also constrained in the gasoline exhaust factor using "Pull Up Maximally" with a 1% dQ.

The base run and constrained run results are shown in Figure 80. The constraint used 0.84% dQ and acetylene was pulled to zero in the refinery factor (Figure 80, 1) and increased to almost 100% in the gasoline exhaust factor (Figure 80, 2). The low amount of dQ needed to move acetylene indicates that it is not a firm feature of the refinery factor and that acetylene can be used as a tracer for gasoline motor vehicle exhaust.

Figure 80. Percent of species associated with a source (1) and Toggle Species Constraint (2).

9.   PMF & Application References

Adhikary, B.; Kulkarni, S.; Dallura, A.; Tang, Y.; Chai, T.; Leung, L.R.; Qian, Y.; Chung, C.E.; Ramanathan, V.; Carmichael, G.R. (2008). A regional scale chemical transport modeling of Asian aerosols with data assimilation of AOD observations using optimal interpolation technique. *Atmos. Environ.*, 42(37): 8600-8615.

Aiken, A.C.; DeCarlo, P.F.; Kroll, J.H.; Worsnop, D.R.; Huffman, J.A.; Docherty, K.S.; Ulbrich, I.M.; Mohr, C.; Kimmel, J.R.; Sueper, D.; Sun, Y.; Zhang, Q.; Trimborn, A.; Northway, M.; Ziemann, P.J.; Canagaratna, M.R.; Onasch, T.B.; Alfarra, M.R.; Prevot, A.S.H.; Dommen, J.; Duplissy, J.; Metzger, A.; Baltensperger, U.; Jimenez, J.L. (2008). O/C and OM/OC ratios of primary, secondary, and ambient organic aerosols with high-resolution time-of-flight aerosol mass spectrometry. *Environ. Sci. Technol.*, 42(12): 4478-4485.

Aiken, A.C.; Salcedo, D.; Cubison, M.J.; Huffman, J.A.; DeCarlo, P.F.; Ulbrich, I.M.; Docherty, K.S.; Sueper, D.; Kimmel, J.R.; Worsnop, D.R.; Trimborn, A.; Northway, M.; Stone, E.A.; Schauer, J.J.; Volkamer, R.M.; Fortner, E.; de Foy, B.; Wang, J.; Laskin, A.; Shutthanandan, V.; Zheng, J.; Zhang, R.; Gaffney, J.; Marley, N.A.; Paredes-Miranda, G.; Arnott, W.P.; Molina, L.T.; Sosa, G.; Jimenez, J.L. (2009). Mexico City aerosol analysis during MILAGRO using high resolution aerosol mass spectrometry at the urban supersite (T0) - Part 1: Fine particle composition and organic source apportionment. *Atmos. Chem. Phys.*, 9(17): 6633-6653.

Aiken, A.C.; de Foy, B.; Wiedinmyer, C.; DeCarlo, P.F.; Ulbrich, I.M.; Wehrli, M.N.; Szidat, S.; Prevot, A.S.H.; Noda, J.; Wacker, L.; Volkamer, R.; Fortner, E.; Wang, J.; Laskin, A.; Shutthanandan, V.; Zheng, J.; Zhang, R.; Paredes-Miranda, G.; Arnott, W.P.; Molina, L.T.; Sosa, G.; Querol, X.; Jimenez, J.L. (2010). Mexico City aerosol analysis during MILAGRO using high resolution aerosol mass spectrometry at the urban supersite (T0) - Part 2: Analysis of the biomass burning contribution and the non-fossil carbon fraction. *Atmos. Chem. Phys.*, 10(12): 5315-5341.

Allan, J.D.; Williams, P.I.; Morgan, W.T.; Martin, C.L.; Flynn, M.J.; Lee, J.; Nemitz, E.; Phillips, G.J.; Gallagher, M.W.; Coe, H. (2010). Contributions from transport, solid fuel burning and cooking to primary organic aerosols in two UK cities. *Atmos. Chem. Phys.*, 10(2): 647-668.

Amato, F.; Pandolfi, M.; Escrig, A.; Querol, X.; Alastuey, A.; Pey, J.; Perez, N.; Hopke, P.K. (2009). Quantifying road dust resuspension in urban environment by Multilinear Engine: A comparison with PMF2. *Atmos. Environ.*, 43(17): 2770-2780.

Amato, F. and Hopke, P.K. (2012) Source apportionment of the ambient PM$_{2.5}$ across St. Louis using constrained positive matrix factorization. *Atmos. Environ.*, 46(2012): 329-337

Anderson, M.J.; Miller, S.L.; Milford, J.B. (2001). Source apportionment of exposure to toxic volatile organic compounds using positive matrix factorization. *J. Expo. Anal. Environ. Epidemiol.*, 11(4): 295-307.

Anderson, M.J.; Daly, E.P.; Miller, S.L.; Milford, J.B. (2002). Source apportionment of exposures to volatile organic compounds II. Application of receptor models to TEAM study data. *Atmos. Environ.*, 36(22): 3643-3658.

Anttila, P.; Paatero, P.; Tapper, U.; Järvinen, O. (1994). Application of positive matrix factorization to source apportionment: Results of a study of bulk deposition chemistry in Finland. *Atmos. Environ.*, 29: 1705-1718.

Banta, J.R.; McConnell, J.R.; Edwards, R.; Engelbrecht, J.P. (2008). Delineation of carbonate dust, aluminous dust, and sea salt deposition in a Greenland glaciochemical array using positive matrix factorization. *Geochemistry Geophysics Geosystems*, 9

Bari, M.A.; Baumbach, G.; Kuch, B.; Scheffknecht, G. (2009). Wood smoke as a source of particle-phase organic compounds in residential areas. *Atmos. Environ.*, 43(31): 4722-4732.

Baumann, K.; Jayanty, R.K.M.; Flanagan, J.B. (2008). Fine particulate matter source apportionment for the Chemical Speciation Trends Network site at Birmingham, Alabama, using Positive Matrix Factorization. *J. Air Waste Manage. Assoc.*, 58: 27-44.

Begum, B.A.; Kim, E.; Biswas, S.K.; Hopke, P.K. (2004). Investigation of sources of atmospheric aerosol at urban and semi-urban areas in Bangladesh. *Atmos. Environ.*, 38(19): 3025-3038.

Begum, B.A.; Biswas, S.K.; Kim, E.; Hopke, P.K.; Khaliquzzaman, M. (2005). Investigation of sources of atmospheric aerosol at a hot spot area in Dhaka, Bangladesh. *J. Air Waste Manage. Assoc.*, 55(2): 227-240.

Begum, B.A.; Hopke, P.K.; Zhao, W.X. (2005). Source identification of fine particles in Washington, DC, by expanded factor analysis modeling. *Environ. Sci. Technol.*, 39(4): 1129-1137.

Begum, B.A.; Biswas, S.K.; Hopke, P.K.; Cohen, D.D. (2006). Multi-element analysis and characterization of atmospheric particulate pollution in Dhaka. *AAQR*, 6(4): 334-359. aaqr.org.

Begum, B.A.; Biswas, S.K.; Nasiruddin, M.; Hossain, A.M.S.; Hopke, P.K. (2009). Source identification of Chittagong aerosol by receptor modeling. *Environmental Engineering Science*, 26(3): 679-689.

Begum, B.A.; Biswas, S.K.; Markwitz, A.; Hopke, P.K. (2010). Identification of Sources of Fine and Coarse Particulate Matter in Dhaka, Bangladesh. *AAQR*, 10(4): 345-U1514.

Bhanuprasad, S.G.; Venkataraman, C.; Bhushan, M. (2008). Positive matrix factorization and trajectory modelling for source identification: A new look at Indian Ocean Experiment ship observations. *Atmos. Environ.*, 42(20): 4836-4852.

Bon, D.M.; Ulbrich, I.M.; de Gouw, J.A.; Warneke, C.; Kuster, W.C.; Alexander, M.L.; Baker, A.; Beyersdorf, A.J.; Blake, D.; Fall, R.; Jimenez, J.L.; Herndon, S.C.; Huey, L.G.; Knighton, W.B.; Ortega, J.; Springston, S.; Vargas, O. (2011). Measurements of volatile organic compounds at a suburban ground site (T1) in Mexico City during the MILAGRO 2006 campaign: measurement comparison, emission ratios, and source attribution. *Atmos. Chem. Phys.*, 11(6): 2399-2421.

Brinkman, G.; Vance, G.; Hannigan, M.P.; Milford, J.B. (2006). Use of synthetic data to evaluate positive matrix factorization as a source apportionment tool for $PM_{2.5}$ exposure data. *Environ. Sci. Technol.*, 40(6): 1892-1901.

Brown, S.G.; Frankel, A.; Raffuse, S.M.; Roberts, P.T.; Hafner, H.R.; Anderson, D.J. (2007). Source apportionment of fine particulate matter in Phoenix, AZ, using positive matrix factorization. *J. Air Waste Manage. Assoc.*, 57(6): 741-752.

Brown, S.G.; Frankel, A.; Hafner, H.R. (2007). Source apportionment of VOCs in the Los Angeles area using positive matrix factorization. *Atmos. Environ.*, 41(2): 227-237.

Brown S.G., Wade K.S., and Hafner H.R. (2007) Multivariate receptor modeling workbook. Prepared for the U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC, by Sonoma Technology, Inc., Petaluma, CA, STI-906207.01-3216, August.

Brown, S.G .; Eberly, S.;. Pentti, P.; Norris, G.A. (2014) Methods for Estimating Uncertainty in PMF Solutions: Examples with Ambient Data. *submitted*.

Bullock, K.R.; Duvall, R.M.; Norris, G.A.; McDow, S.R.; Hays, M.D. (2008). Evaluation of the CMB and PMF models using organic molecular markers in fine particulate matter collected during the Pittsburgh Air Quality Study. *Atmos. Environ.*, 42(29): 6897-6904.

Buset, K.C.; Evans, G.J.; Leaitch, W.R.; Brook, J.R.; Toom-Sauntry, D. (2006). Use of advanced receptor modelling for analysis of an intensive 5-week aerosol sampling campaign. *Atmos. Environ.*, 40(Suppl. 2): S482-S499.

Buzcu-Guven, B.; Brown, S.G.; Frankel, A.; Hafner, H.R.; Roberts, P.T. (2007). Analysis and apportionment of organic carbon and fine particulate matter sources at multiple sites in the Midwestern United States. *J. Air Waste Manage. Assoc.*, 57(5): 606-619.

Buzcu-Guven, B.; Fraser, M.P. (2008). Comparison of VOC emissions inventory data with source apportionment results for Houston, TX. *Atmos. Environ.*, 42(20): 5032-5043.

Buzcu, B.; Fraser, M.P. (2006). Source identification and apportionment of volatile organic compounds in Houston, TX. *Atmos. Environ.*, 40(13): 2385-2400. ISI:000236773000014.

Bzdusek, P.A.; Lu, J.; Christensen, E.R. (2006) PCB congeners and dechlorination in sediments of Sheboygan River, Wisconsin, determined by matrix factorization. *Environ. Sci. Technol.*, 40(1), 120-129. Available at http://dx.doi.org/10.1021/es050083p.

Chan, Y.C.; Cohen, D.D.; Hawas, O.; Stelcer, E.; Simpson, R.; Denison, L.; Wong, N.; Hodge, M.; Comino, E.; Carswell, S. (2008). Apportionment of sources of fine and coarse particles in four major Australian cities by positive matrix factorisation. *Atmos. Environ.*, 42(2): 374-389.

Chan, Y.C.; Hawas, O.; Hawker, D.; Vowles, P.; Cohen, D.D.; Stelcer, E.; Simpson, R.; Golding, G.; Christensen, E. (2011). Using multiple type composition data and wind data in PMF analysis to apportion and locate sources of air pollutants. *Atmos. Environ.*, 45(2): 439-449.

Chand, D.; Hegg, D.A.; Wood, R.; Shaw, G.E.; Wallace, D.; Covert, D.S. (2010). Source attribution of climatically important aerosol properties measured at Paposo (Chile) during VOCALS. *Atmos. Chem. Phys.*, 10(22): 10789-10801.

Chen, L.-W.A.; Watson, J.G.; Chow, J.C.; Magliano, K.L. (2007). Quantifying $PM_{2.5}$ source contributions for the San Joaquin Valley with multivariate receptor models. *Environ. Sci. Technol.*, 41(8): 2818-2826.

Chen, L.-W.A.; Lowenthal, D.H.; Watson, J.G.; Koracin, D.; Kumar, N.; Knipping, E.M.; Wheeler, N.; Craig, K.; Reid, S. (2010). Toward effective source apportionment using positive matrix factorization: Experiments with simulated $PM_{2.5}$ data. *J. Air Waste Manage. Assoc.*, 60(1): 43-54. http://pubs.awma.org/gsearch/journal/2010/1/10.3155-1047-3289.60.1.43.pdf.

Chen, L.-W.A.; Watson, J.G.; Chow, J.C.; DuBois, D.W.; Herschberger, L. (2011). $PM_{2.5}$ source apportionment: Reconciling receptor models for U.S. non-urban and urban long-term networks. *J. Air Waste Manage. Assoc.*, 61(11): 1204-1217.

Cheng, I.; Lu, J.; Song, X.J. (2009). Studies of potential sources that contributed to atmospheric mercury in Toronto, Canada. *Atmos. Environ.*, 43(39): 6145-6158.

Cherian, R.; Venkataraman, C.; Kumar, A.; Sarin, M.M.; Sudheer, A.K.; Ramachandran, S. (2010). Source identification of aerosols influencing atmospheric extinction: Integrating PMF and PSCF with emission inventories and satellite observations. *Journal of Geophysical Research-Atmospheres*, 115

Chiou, P.; Tang, W.; Lin, C.J.; Chu, H.W.; Tadmor, R.; Ho, T.C. (2008). Atmospheric aerosols over two sites in a southeastern region of Texas. *Canadian Journal of Chemical Engineering*, 86(3): 421-435.

Chiou, P.; Tang, W.; Lin, C.J.; Chu, H.W.; Ho, T.C. (2009). Atmospheric aerosol over a southeastern region of Texas: Chemical composition and possible sources. *Environ. Mon. Assess.*, 14(3): 333-350.

Chiou, P.; Tang, W.; Lin, C.J.; Chu, H.W.; Ho, T.C. (2009). Comparison of atmospheric aerosols between two sites over Golden Triangle of Texas. *International Journal of Environmental Research*, 3(2): 253-270.

Choi, E.; Heo, J.B.; Hopke, P.K.; Jin, B.B.; Yi, S.M. (2011). Identification, apportionment, and photochemical reactivity of non-methane hydrocarbon sources in Busan, Korea. *Water Air and Soil Pollution*, 215(1-4): 67-82.

Choi, H.W.; Hwang, I.J.; Kim, S.D.; Kim, D.S. (2004). Determination of source contribution based on aerosol number and mass concentration in the Seoul subway stations. *J. Korean Society for Atmos. Environ.*, 20(1): 17-31.

Christensen, W.F.; Schauer, J.J. (2008). Impact of species uncertainty perturbation on the solution stability of positive matrix factorization of atmospheric particulate matter data. *Environ. Sci. Technol.*, 42(16): 6015-6021.

Chueinta, W.; Hopke, P.K.; Paatero, P. (2000). Investigation of sources of atmospheric aerosol at urban and suburban residential areas in Thailand by positive matrix factorization. *Atmos. Environ.*, 34(20): 3319-3329.

Chueinta, W.; Hopke, P.K.; Paatero, P. (2004). Multilinear model for spatial pattern analysis of the measurement of haze and visual effects project. *Environ. Sci. Technol.*, 38(2): 544-554.

Cohen, D.D.; Crawford, J.; Stelcer, E.; Bac, V.T. (2010). Characterisation and source apportionment of fine particulate sources at Hanoi from 2001 to 2008. *Atmos. Environ.*, 44(3): 320-328.

Coutant, B.W.; Kelly, T.; Ma, J.; Scott, B.; Wood, B.; Main, H.H. (2002). Source Apportionment Analysis of Air Quality Data: Phase 1 - Final Report. prepared by Mid-Atlantic Regional Air Management Assoc., Baltimore, MD, http://www.marama.org/visibility/SA_report/

Cuccia, E.; Bernardoni, V.; Massabo, D.; Prati, P.; Valli, G.; Vecchi, R. (2010). An alternative way to determine the size distribution of airborne particulate matter. *Atmos. Environ.*, 44(27): 3304-3313.

DeCarlo, P.F.; Ulbrich, I.M.; Crounse, J.; de Foy, B.; Dunlea, E.J.; Aiken, A.C.; Knapp, D.; Weinheimer, A.J.; Campos, T.; Wennberg, P.O.; Jimenez, J.L. (2010). Investigation of the sources and processing of organic aerosol over the Central Mexican Plateau from aircraft measurements during MILAGRO. *Atmos. Chem. Phys.*, 10(12): 5257-5280.

Dogan, G.; Gullu, G.; Tuncel, G. (2008). Sources and source regions effecting the aerosol composition of the Eastern Mediterranean. *Microchemical Journal*, 88(2): 142-149.

Dreyfus, M.A.; Adou, K.; Zucker, S.M.; Johnston, M.V. (2009). Organic aerosol source apportionment from highly time-resolved molecular composition measurements. *Atmos. Environ.*, 43(18): 2901-2910.

Du, S.; Belton, T.J.; Rodenburg, L.A. (2008). Source apportionment of polychlorinated biphenyls in the tidal Delaware River. *Environ. Sci. Technol.*, 42(11): 4044-4051.

Du, S.; Wall, S.J.; Cacia, D.; Rodenburg, L.A. (2009). Passive air sampling for polychlorinated biphenyls in the Philadelphia metropolitan area. *Environ. Sci. Technol.*, 43(5): 1287-1292.

Du, S.Y.; Rodenburg, L.A. (2007). Source identification of atmospheric PCBs in Philadelphia/Camden using positive matrix factorization followed by the potential source contribution function. *Atmos. Environ.*, 41: 8596-8608.

Dutton, S.J.; Vedal, S.; Piedrahita, R.; Milford, J.B.; Miller, S.L.; Hannigan, M.P. (2010). Source apportionment using positive matrix factorization on daily measurements of inorganic and organic speciated $PM_{2.5}$. *Atmos. Environ.*, 44(23): 2731-2741.

Eatough, D.J.; Anderson, R.R.; Martello, D.V.; Modey, W.K.; Mangelson, N.E. (2006). Apportionment of ambient primary and secondary $PM_{2.5}$ during a 2001 summer intensive study at the NETL Pittsburgh site using PMF2 and EPA UNMIX. *Aerosol Sci. Technol.*, 40 (10): 925-940.

Eatough, D.J.; Mangelson, N.F.; Anderson, R.R.; Martello, D.V.; Pekney, N.J.; Davidson, C.I.; Modey, W.K. (2007). Apportionment of ambient primary and secondary fine particulate matter during a 2001 summer intensive study at the CMU supersite and NETL Pittsburgh site. *J. Air Waste Manage. Assoc.*, 57(10): 1251-1267.

Eatough, D.J.; Grover, B.D.; Woolwine, W.R.; Eatough, N.L.; Long, R.; Farber, R. (2008). Source apportionment of 1 h semi-continuous data during the 2005 Study of Organic Aerosols in Riverside (SOAR) using positive matrix factorization. *Atmos. Environ.*, 42(11): 2706-2719.

Eatough, D.J.; Farber, R. (2009). Apportioning visibility degradation to sources of $PM_{2.5}$ using positive matrix factorization. *J. Air Waste Manage. Assoc.*, 59(9): 1092-1110.

Eberly, S.I. (2005). EPA PMF 1.1 User's Guide. prepared by U.S. Environmental Protection Agency, Research Triangle Park, NC,

Engel-Cox, J.A.; Weber, S.A. (2007). Compilation and assessment of recent positive matrix factorization and UNMIX receptor model studies on fine particulate matter source apportionment for the eastern United States. *J. Air Waste Manage. Assoc.*, 57(11): 1307-1316.

Escrig, A.; Monfort, E.; Celades, I.; Querol, X.; Amato, F.; Minguillon, M.C.; Hopke, P.K. (2009). Application of optimally scaled target factor analysis for assessing source contribution of ambient PM_{10}. *J. Air Waste Manage. Assoc.*, 59(11): 1296-1307.

Favez, O.; El Haddad, I.; Piot, C.; Boreave, A.; Abidi, E.; Marchand, N.; Jaffrezo, J.L.; Besombes, J.L.; Personnaz, M.B.; Sciare, J.; Wortham, H.; George, C.; D'Anna, B. (2010). Intercomparison of source apportionment models for the estimation of wood burning aerosols during wintertime in an Alpine city (Grenoble, France). *Atmos. Chem. Phys.*, 10(12): 5295-5314.

Friend, A.J.; Ayoko, G.A. (2009). Multi-criteria ranking and source apportionment of fine particulate matter in Brisbane, Australia. *Environmental Chemistry*, 6(5): 398-406.

Friend, A.J.; Ayoko, G.A.; Elbagir, S.G. (2011). Source apportionment of fine particles at a suburban site in Queensland, Australia. *Environmental Chemistry*, 8(2): 163-173.

Fry, J.L.; Kiendler-Scharr, A.; Rollins, A.W.; Brauers, T.; Brown, S.S.; Dorn, H.P.; Dube, W.P.; Fuchs, H.; Mensah, A.; Rohrer, F.; Tillmann, R.; Wahner, A.; Wooldridge, P.J.; Cohen, R.C. (2011). SOA from limonene: Role of NO3 in its generation and degradation. *Atmos. Chem. Phys.*, 11(8): 3879-3894.

Fujita E.M. (2001) Hydrocarbon source apportionment for the 1996 Paso del Norte Ozone Study, The Science of the Total Environment 276: 171-184.

Furusjo, E.; Sternbeck, J.; Cousins, A.P. (2007). PM_{10} source characterization at urban and highway roadside locations. *Sci. Total Environ.*, 387: 206-219.

Gaimoz, C.; Sauvage, S.; Gros, V.; Herrmann, F.; Williams, J.; Locoge, N.; Perrussel, O.; Bonsang, B.; d'Argouges, O.; Sarda-Esteve, R.; Sciare, J. (2011). Volatile organic compounds sources in Paris in spring 2007. Part II: source apportionment using positive matrix factorisation. *Environmental Chemistry*, 8(1): 91-103.

Gao, N.; Gildemeister, A.E.; Krumhansl, K.; Lafferty, K.; Hopke, P.K.; Kim, E.; Poirot, R.L. (2006). Sources of fine particulate species in ambient air over Lake Champlain Basin, VT. *J. Air Waste Manage. Assoc.*, 56(11): 1607-1620.

Gietl, J.K.; Klemm, O. (2009). Source identification of size-segregated aerosol in Munster, Germany, by factor analysis. *Aerosol Sci. Technol.*, 43(8): 828-837.

Gilardoni, S.; Vignati, E.; Marmer, E.; Cavalli, F.; Belis, C.; Gianelle, V.; Loureiro, A.; Artaxo, P. (2011). Sources of carbonaceous aerosol in the Amazon basin. *Atmos. Chem. Phys.*, 11(6): 2747-2764.

Gildemeister, A.E.; Hopke, P.K.; Kim, E. (2007). Sources of fine urban particulate matter in Detroit, MI. *Chemosphere*, 69: 1064-1074.

Gong, F.; Wang, B.T.; Fung, Y.S.; Chau, F.T. (2005). Chemometric characterization of the quality of the atmospheric environment in Hong Kong. *Atmos. Environ.*, 39(34): 6388-6397.

Grahame, T.; Hidy, G.M. (2007). Pinnacles and pitfalls for source apportionment of potential health effects from airborne particle exposure. *Inhal. Toxicol.*, 19(9): 727-744.

Gratz, L.E.; Keeler, G.J. (2011). Sources of mercury in precipitation to Underhill, VT. *Atmos. Environ.*, 45(31): 5440-5449.

Green, M.C.; Xu, J. (2007). Causes of haze in the Columbia River Gorge. *J. Air Waste Manage. Assoc.*, 57(8): 947-958.

Grover, B.D.; Eatough, D.J. (2008). Source apportionment of one-hour semi-continuous data using positive matrix factorization with total mass (nonvolatile plus semi-volatile) measured by the R&P FDMS monitor. *Aerosol Sci. Technol.*, 42(1): 28-39.

Gu, J.W.; Pitz, M.; Schnelle-Kreis, J.; Diemer, J.; Reller, A.; Zimmermann, R.; Soentgen, J.; Stoelzel, M.; Wichmann, H.E.; Peters, A.; Cyrys, J. (2011). Source apportionment of ambient particles: Comparison of positive matrix factorization analysis applied to particle size distribution and chemical composition data. *Atmos. Environ.*, 45(10): 1849-1857.

Hagler, G.S.W.; Bergin, M.H.; Salmon, L.G.; Yu, J.Z.; Wan, E.C.H.; Zheng, M.; Zeng, L.M.; Kiang, C.S.; Zhang, Y.H.; Schauer, J.J. (2007). Local and regional anthropogenic influence on $PM_{2.5}$ elements in Hong Kong. *Atmos. Environ.*, 41(28): 5994-6004.

Hammond, D.M.; Dvonch, J.T.; Keeler, G.J.; Parker, E.A.; Kamal, A.S.; Barres, J.A.; Yip, F.Y.; Brakefield-Caldwell, W. (2008). Sources of ambient fine particulate matter at two community sites in Detroit, Michigan. *Atmos. Environ.*, 42(4): 720-732.

Han, J.S.; Moon, K.J.; Kim, Y.J. (2006). Identification of potential sources and source regions of fine ambient particles measured at Gosan background site in Korea using advanced hybrid receptor model combined with positive matrix factorization. *Journal of Geophysical Research-Atmospheres*, 111(D22)ISI:000242740700001.

Han, J.S.; Moon, K.J.; Lee, S.J.; Kim, Y.J.; Ryu, S.Y.; Cliff, S.S.; Yi, S.M. (2006). Size-resolved source apportionment of ambient particles by positive matrix factorization at Gosan background site in East Asia. *Atmos. Chem. Phys.*, 6: 211-223.

Harrison, R.M.; Beddows, D.C.S.; Dall'Osto, M. (2011). PMF analysis of wide-range particle size spectra collected on a major highway. *Environ. Sci. Technol.*, 45(13): 5522-5528.

Hawkins, L.N.; Russell, L.M.; Covert, D.S.; Quinn, P.K.; Bates, T.S. (2010). Carboxylic acids, sulfates, and organosulfates in processed continental organic aerosol over the southeast Pacific Ocean during VOCALS-REx 2008. *Journal of Geophysical Research-Atmospheres*, 115

Healy, R.M.; Hellebust, S.; Kourtchev, I.; Allanic, A.; O'Connor, I.P.; Bell, J.M.; Healy, D.A.; Sodeau, J.R.; Wenger, J.C. (2010). Source apportionment of $PM_{2.5}$ in Cork Harbour, Ireland using a combination of single particle mass spectrometry and quantitative semi-continuous measurements. *Atmos. Chem. Phys.*, 10(19): 9593-9613.

Hedberg, E.; Gidhagen, L.; Johansson, C. (2005). Source contributions to PM_{10} and arsenic concentrations in Central Chile using positive matrix factorization. *Atmos. Environ.*, 39(3): 549-561.

Hegg, D.A.; Warren, S.G.; Grenfell, T.C.; Doherty, S.J.; Larson, T.V.; Clarke, A.D. (2009). Source attribution of black carbon in Arctic snow. *Environ. Sci. Technol.*, 43(11): 4016-4021.

Hegg, D.A.; Warren, S.G.; Grenfell, T.C.; Doherty, S.J.; Clarke, A.D. (2010). Sources of light-absorbing aerosol in Arctic snow and their seasonal variation. *Atmos. Chem. Phys.*, 10(22): 10923-10938.

Hellebust, S.; Allanic, A.; O'Connor, I.P.; Wenger, J.C.; Sodeau, J.R. (2010). The use of real-time monitoring data to evaluate major sources of airborne particulate matter. *Atmos. Environ.*, 44(8): 1116-1125.

Hemann, J.G.; Brinkman, G.L.; Dutton, S.J.; Hannigan, M.P.; Milford, J.B.; Miller, S.L. (2009). Assessing positive matrix factorization model fit: a new method to estimate uncertainty and bias in factor contributions at the measurement time scale. *Atmos. Chem. Phys.*, 9(2): 497-513.

Henry, R.C. (2002). Multivariate receptor models - Current practice and future trends. *Chemom. Intell. Lab. Sys.*, 60(1-2): 43-48. doi:10.1016/S0169-7439(01)00184-8.

Henry, R.C.; Christensen, E.R. (2010). Selecting an appropriate multivariate source apportionment model result. *Environ. Sci. Technol.*, 44(7): 2474-2481.

Heo, J.B.; Hopke, P.K.; Yi, S.M. (2009). Source apportionment of $PM_{2.5}$ in Seoul, Korea. *Atmos. Chem. Phys.*, 9(14): 4957-4971.

Hersey, S.P.; Craven, J.S.; Schilling, K.A.; Metcalf, A.R.; Sorooshian, A.; Chan, M.N.; Flagan, R.C.; Seinfeld, J.H. (2011). The Pasadena Aerosol Characterization Observatory (PACO): chemical and physical analysis of the western Los Angeles basin aerosol. *Atmos. Chem. Phys.*, 11(15): 7417-7443.

Hien, P.D.; Bac, V.T.; Thinh, N.T.H. (2004). PMF receptor modelling of fine and coarse PM_{10} in air masses governing monsoon conditions in Hanoi, northern Vietnam. *Atmos. Environ.*, 38(2): 189-201. ISI:000188210700003.

Hien, P.D.; Bac, V.T.; Thinh, N.T.H. (2005). Investigation of sulfate and nitrate formation on mineral dust particles by receptor modeling. *Atmos. Environ.*, 39(38): 7231-7239. ISI:000233671700003.

Hodzic, A.; Jimenez, J.L.; Madronich, S.; Canagaratna, M.R.; DeCarlo, P.F.; Kleinman, L.; Fast, J. (2010). Modeling organic aerosols in a megacity: potential contribution of semi-volatile and intermediate volatility primary organic compounds to secondary organic aerosol formation. *Atmos. Chem. Phys.*, 10(12): 5491-5514.

Hopke, P.K.; Xie, Y.L.; Paatero, P. (1999). Mixed multiway analysis of airborne particle composition data. *J. Chemometrics*, 13: 343-352.

Hopke, P.K. (2000). A Guide to Positive Matrix Factorization. prepared by Clarkson University, Clarkson University-Department of Chemistry,

Hopke, P.K.; Ramadan, Z.; Paatero, P.; Norris, G.A.; Landis, M.S.; Williams, R.W.; Lewis, C.W. (2003). Receptor modeling of ambient and personal exposure samples: 1998 Baltimore Particulate Matter Epidemiology-Exposure Study. *Atmos. Environ.*, 37(23): 3289-3302. doi: 10.1016/S1352-2310(03)00331-5.

Hopke, P.K.; Ito, K.; Mar, T.; Christensen, W.F.; Eatough, D.J.; Henry, R.C.; Kim, E.; Laden, F.; Lall, R.; Larson, T.V.; Liu, H.; Neas, L.; Pinto, J.; Stolzel, M.; Suh, H.; Paatero, P.; Thurston, G.D. (2006). PM source apportionment and health effects: 1. Intercomparison of source apportionment results. *J. Expo. Anal. Environ. Epidemiol.*, 16: 275-286. doi:10.1038/sj.jea.7500458.

Hopke, P.K. (2010). Discussion of "Sensitivity of a molecular marker based positive matrix factorization model to the number of receptor observations" by YuanXun Zhang, Rebecca J. Sheesley, Min-Suk Bae and James J. Schauer. *Atmos. Environ.*, 44(8): 1138.

Hu, D.; Bian, Q.J.; Lau, A.K.H.; Yu, J.Z. (2010). Source apportioning of primary and secondary organic carbon in summer $PM_{2.5}$ in Hong Kong using positive matrix factorization of secondary and primary organic tracer data. *Journal of Geophysical Research-Atmospheres*, 115

Hu, S.H.; McDonald, R.; Martuzevicius, D.; Biswas, P.; Grinshpun, S.A.; Kelley, A.; Reponen, T.; Lockey, J.; LeMasters, G. (2006). UNMIX modeling of ambient $PM_{2.5}$ near an interstate highway in Cincinnati, OH, USA. *Atmos. Environ.*, 40(Suppl. 2): S378-S395.

Huang, S.L.; Arimoto, R.; Rahn, K.A. (2001). Sources and source variations for aerosol at Mace Head, Ireland. *Atmos. Environ.*, 35(8): 1421-1437.

Huang, X.F.; Yu, J.Z.; He, L.Y.; Yuan, Z.B. (2006). Water-soluble organic carbon and oxalate in aerosols at a coastal urban site in China: Size distribution characteristics, sources, and formation mechanisms. *Journal of Geophysical Research-Atmospheres*, 111(D22)

Huang, X.F.; Yu, J.Z.; Yuan, Z.B.; Lau, A.K.H.; Louie, P.K.K. (2009). Source analysis of high particulate matter days in Hong Kong. *Atmos. Environ.*, 43(6): 1196-1203.

Huang, X.F.; Zhao, Q.B.; He, L.Y.; Hu, M.; Bian, Q.J.; Xue, L.A.; Zhang, Y.H. (2010). Identification of secondary organic aerosols based on aerosol mass spectrometry. *Science China-Chemistry*, 53(12): 2593-2599.

Huang, X.F.; He, L.Y.; Hu, M.; Canagaratna, M.R.; Kroll, J.H.; Ng, N.L.; Zhang, Y.H.; Lin, Y.; Xue, L.; Sun, T.L.; Liu, X.G.; Shao, M.; Jayne, J.T.; Worsnop, D.R. (2011). Characterization of submicron aerosols at a rural site in Pearl River Delta of China using an Aerodyne High-Resolution Aerosol Mass Spectrometer. *Atmos. Chem. Phys.*, 11(5): 1865-1877.

Hubble, M. (2000). Phoenix Source Apportionment Studies: Positive Matrix Factorization (PMF) and Unmix Applications for $PM_{2.5}$ Source Apportionment. prepared by Arizona Department of Environmental Quality, Arizona Department of Environmental Quality-Phoenix, AZ,

Huffman, J.A.; Docherty, K.S.; Aiken, A.C.; Cubison, M.J.; Ulbrich, I.M.; DeCarlo, P.F.; Sueper, D.; Jayne, J.T.; Worsnop, D.R.; Ziemann, P.J.; Jimenez, J.L. (2009). Chemically-resolved aerosol volatility measurements from two megacity field studies. *Atmos. Chem. Phys.*, 9(18): 7161-7182.

Hwang, I.; Hopke, P.K. (2006). Comparison of source apportionments of fine particulate matter at two San Jose Speciation Trends Network sites. *J. Air Waste Manage. Assoc.*, 56(9): 1287-1300.

Hwang, I.; Hopke, P.K. (2007). Estimation of source apportionment and potential source locations Of $PM_{2.5}$ at a west coastal IMPROVE site. *Atmos. Environ.*, 41(3): 506-518.

Hwang, I.; Hopke, P.K.; Pinto, J.P. (2008). Source apportionment and spatial distributions of coarse particles during the Regional Air Pollution Study. *Environ. Sci. Technol.*, 42(10): 3524-3530.

Hwang, I.J.; Bong, C.K.; Lee, T.J.; Kim, D.S. (2002). Source identification and quantification of coarse and fine particles by TTFA and PMF. *J. Korean Society for Atmos. Environ.*, 18(E4): 203-213.

Hwang, I.J.; Kim, D.S. (2003). Estimation of quantitative source contribution of ambient PM_{10} using the PMF model. *J. Korean Society for Atmos. Environ.*, 19(6): 719-731.

Iijima, A.; Tago, H.; Kumagai, K.; Kato, M.; Kozawa, K.; Sato, K.; Furuta, N. (2008). Regional and seasonal characteristics of emission sources of fine airborne particulate matter collected in the center and suburbs of Tokyo, Japan as determined by multielement analysis and source receptor models. *J. Environ. Monit.*, 10(9): 1025-1032.

Ito, K.; Xue, N.; Thurston, G. (2004). Spatial variation of $PM_{2.5}$ chemical species and source-apportioned mass concentrations in New York City. *Atmos. Environ.*, 38(31): 5269-5282.

Jacobson, M.Z.; Kaufman, Y.J. (2006). Wind reduction by aerosol particles. *Geophys. Res. Lett.*, 33(24)

Jaeckels, J.M.; Bae, M.S.; Schauer, J.J. (2007). Positive matrix factorization (PMF) analysis of molecular marker measurements to quantify the sources of organic aerosols. *Environ. Sci. Technol.*, 41(16): 5763-5769.

Jagoda, C.A.; Charnbers, S.; David, D.C.; Dyer, L.; Wang, T.; Zahorowski, W. (2007). Receptor modelling using positive matrix factorisation, back trajectories and radon-222. *Atmos. Environ.*, 41(32): 6823-6837.

Jeong, C.H.; Evans, G.J.; Dann, T.; Graham, M.; Herod, D.; bek-Zlotorzynska, E.; Mathieu, D.; Ding, L.; Wang, D. (2008). Influence of biomass burning on wintertime fine particulate matter: Source contribution at a valley site in rural British Columbia. *Atmos. Environ.*, 42(16): 3684-3699.

Jia, Y.L.; Clements, A.L.; Fraser, M.P. (2010). Saccharide composition in atmospheric particulate matter in the southwest US and estimates of source contributions. *J. Aerosol Sci.*, 41(1): 62-73.

Jia, Y.L.; Fraser, M. (2011). Characterization of saccharides in size-fractionated ambient particulate matter and aerosol sources: The contribution of Primary Biological Aerosol Particles (PBAPs) and soil to ambient particulate matter. *Environ. Sci. Technol.*, 45(3): 930-936.

Jimenez, J.; Wu, C.F.; Claiborn, C.; Gould, T.; Simpson, C.D.; Larson, T.; Liu, L.J.S. (2006). Agricultural burning smoke in eastern Washington - part 1: Atmospheric characterization. *Atmos. Environ.*, 40(4): 639-650.

Johnson, K.S.; de Foy, B.; Zuberi, B.; Molina, L.T.; Molina, M.J.; Xie, Y.; Laskin, A.; Shutthanandan, V. (2006). Aerosol composition and source apportionment in the Mexico City Metropolitan Area with PIXE/PESA/STIM and multivariate analysis. *Atmos. Chem. Phys.*, 6(12): 4591-4600.

Jorquera, H.; Rappengluck, B. (2004). Receptor modeling of ambient VOC at Santiago, Chile. *Atmos. Environ.*, 38(25): 4243-4263.

Junninen, H.; Monster, J.; Rey, M.; Cancelinha, J.; Douglas, K.; Duane, M.; Forcina, V.; Muller, A.; Lagler, F.; Marelli, L.; Borowiak, A.; Niedzialek, J.; Paradiz, B.; Mira-Salama, D.; Jimenez, J.; Hansen, U.; Astorga, C.; Stanczyk, K.; Viana, M.; Querol, X.; Duvall, R.M.; Norris, G.A.; Tsakovski, S.; Wahlin, P.; Horak, J.; Larsen, B.R. (2009). Quantifying the impact of residential heating on the urban air quality in a typical European coal combustion region. *Environ. Sci. Technol.*, 43(20): 7964-7970.

Juntto, S.; Paatero, P. (1994). Analysis of daily precipitation data by positive matrix factoriztion. *Environmetrics*, 5: 127-144.

Juvela, M.; Lehtinen, K.; Paatero, P. (1996). The use of positive matrix factorization in the analysis of molecular line spectra. *ROYAL ASTR. SOC.*, 280(2)

Karanasiou, A.; Moreno, T.; Amato, F.; Lumbreras, J.; Narros, A.; Borge, R.; Tobias, A.; Boldo, E.; Linares, C.; Pey, J.; Reche, C.; Alastuey, A.; Querol, X. (2011). Road dust contribution to PM levels - Evaluation of the effectiveness of street washing activities by means of Positive Matrix Factorization. *Atmos. Environ.*, 45(13): 2193-2201.

Karanasiou, A.A.; Siskos, P.A.; Eleftheriadis, K. (2009). Assessment of source apportionment by Positive Matrix Factorization analysis on fine and coarse urban aerosol size fractions. *Atmos. Environ.*, 43(21): 3385-3395.

Karnae, S.; Kuruvilla, J. (2011). Source apportionment of fine particulate matter measured in an industrialized coastal urban area of South Texas. *Atmos. Environ.*, 45(23): 3769-3776.

Kasumba, J.; Hopke, P.K.; Chalupa, D.C.; Utell, M.J. (2009). Comparison of sources of submicron particle number concentrations measured at two sites in Rochester, NY. *Sci. Total Environ.*, 407(18): 5071-5084.

Ke, L.; Liu, W.; Wang, Y.; Russell, A.G.; Edgerton, E.S.; Zheng, M. (2008). Comparison of $PM_{2.5}$ source apportionment using positive matrix factorization and molecular marker-based chemical mass balance. *Sci. Total Environ.*, 394(2-3): 290-302.

Keeler, G.J.; Landis, M.S.; Norris, G.A.; Christianson, E.M.; Dvonch, J.T. (2006). Sources of mercury wet deposition in Eastern Ohio, USA. *Environ. Sci. Technol.*, 40(19): 5874-5881. ISI:000240826000015.

Kertesz, Z.; Szoboszlai, Z.; Angyal, A.; Dobos, E.; Borbely-Kiss, I. (2010). Identification and characterization of fine and coarse particulate matter sources in a middle-European urban environment. *Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms*, 268(11-12): 1924-1928.

Kim, E.; Hopke, P.K.; Paatero, P.; Edgerton, E.S. (2003). Incorporation of parametric factors into multilinear receptor model studies of Atlanta aerosol. *Atmos. Environ.*, 37 (36): 5009-5021.

Kim, E.; Hopke, P.K.; Edgerton, E.S. (2003). Source identification of Atlanta aerosol by positive matrix factorization. *J. Air Waste Manage. Assoc.*, 53(6): 731-739.

Kim, E.; Larson, T.V.; Hopke, P.K.; Slaughter, C.; Sheppard, L.E.; Claiborn, C. (2003). Source identification of PM$_{2.5}$ in an arid northwest U.S. city by positive matrix factorization. *Atmos. Res.*, 66: 291-305.

Kim, E.; Hopke, P.K.; Larson, T.V.; Covert, D.S. (2004). Analysis of ambient particle size distributions using UNMIX and positive matrix factorization. *Environ. Sci. Technol.*, 38(1): 202-209.

Kim, E.; Hopke, P.K. (2004). Comparison between conditional probability function and nonparametric regression for fine particle source directions. *Atmos. Environ.*, 38(28): 4667-4673.

Kim, E.; Hopke, P.K.; Larson, T.V.; Maykut, N.N.; Lewtas, J. (2004). Factor analysis of Seattle fine particles. *Aerosol Sci. Technol.*, 38(7): 724-738.

Kim, E.; Hopke, P.K.; Edgerton, E.S. (2004). Improving source identification of Atlanta aerosol using temperature resolved carbon fractions in positive matrix factorization. *Atmos. Environ.*, 38(20): 3349-3362.

Kim, E.; Hopke, P.K. (2004). Improving source identification of fine particles in a rural northeastern US area utilizing temperature-resolved carbon fractions. *Journal of Geophysical Research-Atmospheres*, 109(D09204): 1-13. doi:2003JD004199.

Kim, E.; Hopke, P.K. (2004). Source apportionment of fine particles at Washington, DC, utilizing temperature-resolved carbon fractions. *J. Air Waste Manage. Assoc.*, 54(7): 773-785.

Kim, E.; Brown, S.G.; Hafner, H.R.; Hopke, P.K. (2005). Characterization of non-methane volatile organic compounds sources in Houston during 2001 using positive matrix factorization. *Atmos. Environ.*, 39(32): 5934-5946.

Kim, E.; Hopke, P.K. (2005). Identification of fine particle sources in mid-Atlantic US area. *Water Air and Soil Pollution*, 168(1-4): 391-421.

Kim, E.; Hopke, P.K. (2005). Improving source apportionment of fine particles in the eastern United States utilizing temperature-resolved carbon fractions. *J. Air Waste Manage. Assoc.*, 55(10): 1456-1463.

Kim, E.; Hopke, P.K.; Kenski, D.M.; Koerber, M. (2005). Sources of fine particles in a rural Midwestern US area. *Environ. Sci. Technol.*, 39(13): 4953-4960.

Kim, E.; Hopke, P.K.; Pinto, J.P.; Wilson, W.E. (2005). Spatial variability of fine particle mass, components, and source contributions during the Regional Air Pollution Study in St. Louis. *Environ. Sci. Technol.*, 39(11): 4172-4179.

Kim, E.; Hopke, P.K. (2006). Characterization of fine particle sources in the Great Smoky Mountains area. *Sci. Total Environ.*, 368(2-3): 781-794.

Kim, E.; Hopke, P.K. (2007). Comparison between sample-species specific uncertainties and estimated uncertainties for the source apportionment of the speciation trends network data. *Atmos. Environ.*, 41(3): 567-575.

Kim, E.; Hopke, P.K. (2007). Source identifications of airborne fine particles using positive matrix factorization and US environmental protection agency positive matrix factorization. *J. Air Waste Manage. Assoc.*, 57(7): 811-819.

Kim, E.; Hopke, P.K. (2008). Source characterization of ambient fine particles at multiple sites in the Seattle area. *Atmos. Environ.*, 42(24): 6047-6056.

Kim, E.; Turkiewicz, K.; Zulawnick, S.A.; Magliano, K.L. (2010). Sources of fine particles in the South Coast area, California. *Atmos. Environ.*, 44(26): 3095-3100.

Kim, M.; Deshpande, S.R.; Crist, K.C. (2007). Source apportionment of fine particulate matter (PM$_{2.5}$) at a rural Ohio River Valley site. *Atmos. Environ.*, 41: 9231-9243.

Lambe, A.T.; Logue, J.M.; Kreisberg, N.M.; Hering, S.V.; Worton, D.R.; Goldstein, A.H.; Donahue, N.M.; Robinson, A.L. (2009). Apportioning black carbon to sources using highly time-resolved ambient measurements of organic molecular markers in Pittsburgh. *Atmos. Environ.*, 43(25): 3941-3950.

Lan, Z.J.; Chen, D.L.; Li, X.A.; Huang, X.F.; He, L.Y.; Deng, Y.G.; Feng, N.; Hu, M. (2011). Modal characteristics of carbonaceous aerosol size distribution in an urban atmosphere of South China. *Atmos. Res.*, 100(1): 51-60.

Lanz, V.A.; Alfarra, M.R.; Baltensperger, U.; Buchmann, B.; Hueglin, C.; Prevot, A.S.H. (2007). Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra. *Atmos. Chem. Phys.*, 7(6): 1503-1522.

Lanz, V.A.; Hueglin, C.; Buchmann, B.; Hill, M.; Locher, R.; Staehelin, J.; Reimann, S. (2008). Receptor modeling of C-2-C-7 hydrocarbon sources at an urban background site in Zurich, Switzerland: changes between 1993-1994 and 2005-2006. *Atmos. Chem. Phys.*, 8(9): 2313-2332.

Lanz, V.A.; Henne, S.; Staehelin, J.; Hueglin, C.; Vollmer, M.K.; Steinbacher, M.; Buchmann, B.; Reimann, S. (2009). Statistical analysis of anthropogenic non-methane VOC variability at a European background location (Jungfraujoch, Switzerland). *Atmos. Chem. Phys.*, 9(10): 3445-3459.

Lanz, V.A.; Prevot, A.S.H.; Alfarra, M.R.; Weimer, S.; Mohr, C.; DeCarlo, P.F.; Gianini, M.F.D.; Hueglin, C.; Schneider, J.; Favez, O.; D'Anna, B.; George, C.; Baltensperger, U. (2010). Characterization of aerosol chemical composition with aerosol mass spectrometry in Central Europe: An overview. *Atmos. Chem. Phys.*, 10(21): 10453-10471.

Lapina, K.; Paterson, K.G. (2004). Assessing source characteristics of $PM_{2.5}$ in the eastern United States using positive matrix factorization. *J. Air Waste Manage. Assoc.*, 54(9): 1170-1174.

Larsen, R.K., III; Baker, J.E. (2003). Source apportionment of polycyclic aromatic hydrocarbons in the urban atmosphere: A comparison of three methods. *Environ. Sci. Technol.*, 37: 1873-1881.

Larson, T.; Gould, T.; Simpson, C.; Liu, L.J.S.; Claiborn, C.; Lewtas, J. (2004). Source apportionment of indoor, outdoor, and personal $PM_{2.5}$ in Seattle, WA, using positive matrix factorization. *J. Air Waste Manage. Assoc.*, 54(9): 1175-1187.

Larson, T.V.; Covert, D.S.; Kim, E.; Elleman, R.; Schreuder, A.B.; Lumley, T. (2006). Combining size distribution and chemical species measurements into a multivariate receptor model of $PM_{2.5}$. *Journal of Geophysical Research-Atmospheres*, 111(D10): D10S09. doi:10.1029/2005JD006285.

Latella, A.; Stani, G.; Cobelli, L.; Duane, M.; Junninen, H.; Astorga, C.; Larsen, B.R. (2005). Semicontinuous GC analysis and receptor modelling for source apportionment of ozone precursor hydrocarbons in Bresso, Milan, 2003. *J. Chromatogr. A*, 1071(1-2): 29-39.

Laupsa, H.; Denby, B.; Larssen, S.; Schaug, J. (2009). Source apportionment of particulate matter ($PM_{2.5}$) in an urban area using dispersion, receptor and inverse modelling. *Atmos. Environ.*, 43(31): 4733-4744.

Lee, E.; Chan, C.K.; Paatero, P. (1999). Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong. *Atmos. Environ.*, 33(19): 3201-3212.

Lee, J.H.; Yoshida, Y.; Turpin, B.J.; Hopke, P.K.; Poirot, R.L.; Lioy, P.J.; Oxley, J.C. (2002). Identification of sources contributing to mid-Atlantic regional aerosol. *J. Air Waste Manage. Assoc.*, 52(10): 1186-1205.

Lee, J.H.; Gigliotti, C.L.; Offenberg, J.H.; Eisenreich, S.J.; Turpin, B.J. (2004). Sources of polycyclic aromatic hydrocarbons to the Hudson River Airshed. *Atmos. Environ.*, 38(35): 5971-5981.

Lee, J.H.; Hopke, P.K. (2006). Apportioning sources of $PM_{2.5}$ in St. Louis, MO using speciation trends network data. *Atmos. Environ.*, 40(Suppl. 2): S360-S377.

Lee, J.H.; Hopke, P.K.; Turner, J.R. (2006). Source identification of airborne $PM_{2.5}$ at the St. Louis-Midwest Supersite. *Journal of Geophysical Research-Atmospheres*, 111(D10S10): 1-12. doi:10.1029/2005JD006329.

Lee, P.K.H.; Brook, J.R.; Dabek-Zlotorzynska, E.; Mabury, S.A. (2003). Identification of the major sources contributing to $PM_{2.5}$ observed in Toronto. *Environ. Sci. Technol.*, 37(21): 4831-4840.

Lee, S.; Liu, W.; Wang, Y.H.; Russell, A.G.; Edgerton, E.S. (2008). Source apportionment of $PM_{2.5}$: Comparing PMF and CMB results for four ambient monitoriniz sites in the southeastern United States. *Atmos. Environ.*, 42(18): 4126-4137.

Lei, C.; Landsberger, S.; Basunia, S.; Tao, Y. (2004). Study of $PM_{2.5}$ in Beijing suburban site by neutron activation analysis and source apportionment. *Journal of Radioanalytical and Nuclear Chemistry*, 261(1): 87-94. ISI:000221903800011.

Lestari, P.; Mauliadi, Y.D. (2009). Source apportionment of particulate matter at urban mixed site in Indonesia using PMF. *Atmos. Environ.*, 43(10): 1760-1770.

Leuchner, M.; Rappengluck, B. (2010). VOC source-receptor relationships in Houston during TexAQS-II. *Atmos. Environ.*, 44(33): 4056-4067.

Li, Z.; Hopke, P.K.; Husain, L.; Qureshi, S.; Dutkiewicz, V.A.; Schwab, J.J.; Drewnick, F.; Demerjian, K.L. (2004). Sources of fine particle composition in New York city. *Atmos. Environ.*, 38(38): 6521-6529.

Liang, J.Y.; Kaduwela, A.; Jackson, B.; Gurer, K.; Allen, P. (2006). Off-line diagnostic analyses of a three-dimensional PM model using two matrix factorization methods. *Atmos. Environ.*, 40(30): 5759-5767. ISI:000241217500003.

Liang, J.Y.; Fairley, D. (2006). Validation of an efficient non-negative matrix factorization method and its preliminary application in Central California. *Atmos. Environ.*, 40(11): 1991-2001.

Liggio, J.; Li, S.M.; Vlasenko, A.; Sjostedt, S.; Chang, R.; Shantz, N.; Abbatt, J.; Slowik, J.G.; Bottenheim, J.W.; Brickell, P.C.; Stroud, C.; Leaitch, W.R. (2010). Primary and secondary organic aerosols in urban air masses intercepted at a rural site. *Journal of Geophysical Research-Atmospheres*, 115

Lingwall, J.W.; Christensen, W.F. (2007). Pollution source apportionment using a priori information and positive matrix factorization. *Chemom. Intell. Lab. Sys.*, 87(2): 281-294.

Liu, S.; Takahama, S.; Russell, L.M.; Gilardoni, S.; Baumgardner, D. (2009). Oxygenated organic functional groups and their sources in single and submicron organic particles in MILAGRO 2006 campaign. *Atmos. Chem. Phys.*, 9(18): 6849-6863.

Liu, W.; Hopke, P.K.; Han, Y.J.; Yi, S.M.; Holsen, T.M.; Cybart, S.; Kozlowski, K.; Milligan, M. (2003). Application of receptor modeling to atmospheric constituents at Potsdam and Stockton, NY. *Atmos. Environ.*, 37(36): 4997-5007.

Liu, W.; Hopke, P.K.; VanCuren, R.A. (2003). Origins of fine aerosol mass in the western United States using positive matrix factorization. *Journal of Geophysical Research-Atmospheres*, 108(D23)doi:10.1029/2006JD007978.

Liu, W.; Wang, Y.H.; Russell, A.; Edgerton, E.S. (2005). Atmospheric aerosol over two urban-rural pairs in the southeastern United States: Chemical composition and possible sources. *Atmos. Environ.*, 39(25): 4453-4470.

Liu, W.; Wang, Y.H.; Russell, A.; Edgerton, E.S. (2006). Enhanced source identification of southeast aerosols using temperature-resolved carbon fractions and gas phase components. *Atmos. Environ.*, 40(Suppl. 2): S445-S466.

Logue, J.M.; Small, M.J.; Robinson, A.L. (2009). Identifying priority pollutant sources: Apportioning air toxics risks using positive matrix factorization. *Environ. Sci. Technol.*, 43(24): 9439-9444.

Lonati, G.; Ozgen, S.; Giugliano, M. (2007). Primary and secondary carbonaceous species in $PM_{2.5}$ samples in Milan (Italy). *Atmos. Environ.*, 41(22): 4599-4610.

Lopez, M.L.; Ceppi, S.; Palancar, G.G.; Olcese, L.E.; Tirao, G.; Toselli, B.M. (2011). Elemental concentration and source identification of PM_{10} and $PM_{2.5}$ by SR-XRF in Cordoba City, Argentina. *Atmos. Environ.*, 45(31): 5450-5457.

Lowenthal, D.H.; Watson, J.G.; Koracin, D.; Chen, L.-W.A.; DuBois, D.; Vellore, R.; Kumar, N.; Knipping, E.M.; Wheeler, N.; Craig, K.; Reid, S. (2010). Evaluation of regional scale receptor modeling. *J. Air Waste Manage. Assoc.*, 60(1): 26-42. http://pubs.awma.org/gsearch/journal/2010/1/10.3155-1047-3289.60.1.26.pdf.

Lowenthal, D.H.; Rahn, K.A. (1988). Tests of regional elemental tracers of pollution aerosols. 2. Sensitivity of signatures and apportionments to variations in operating parameters. *Atmos. Environ.*, 22: 420-426.

Lu, J.H.; Wu, L.S. (2004). Technical details and programming guide for a general two-way positive matrix factorization algorithm. *Journal of Chemometrics*, 18(12): 519-525. ISI:000229692100001.

Markus, A.; Matsaev, V. (1994). The failure of factorization of positive matrix functions on noncircular contours. *LINEAR ALGEBRA & APPL*, 208/209: 231.

Marmur, A.; Mulholland, J.A.; Russell, A.G. (2007). Optimized variable source-profile approach for source apportionment. *Atmos. Environ.*, 41(3): 493-505.

Marmur, A.; Liu, W.; Wang, Y.; Russell, A.G.; Edgerton, E.S. (2009). Evaluation of model simulated atmospheric constituents with observations in the factor projected space: CMAQ simulations of SEARCH measurements. *Atmos. Environ.*, 43(11): 1839-1849.

Martello, D.V.; Pekney, N.J.; Anderson, R.R.; Davidson, C.I.; Hopke, P.K.; Kim, E.; Christensen, W.F.; Mangelson, N.F.; Eatough, D.J. (2008). Apportionment of ambient primary and secondary fine particulate matter at the Pittsburgh National Energy Laboratory particulate matter characterization site using positive matrix factorization and a potential source contributions function analysis. *J. Air Waste Manage. Assoc.*, 58(3): 357-368.

Mazzei, F.; Lucarelli, F.; Nava, S.; Prati, P.; Valli, G.; Vecchi, R. (2007). A new methodological approach: The combined use of two-stage streaker samplers and optical particle counters for the characterization of airborne particulate matter. *Atmos. Environ.*, 41(26): 5525-5535.

Mazzei, F.; D'Alessandro, A.; Lucarelli, F.; Nava, S.; Prati, P.; Valli, G.; Vecchi, R. (2008). Characterization of particulate matter sources in an urban environment. *Sci. Total Environ.*, 401(1-3): 81-89.

Mazzei, F.; Prati, P. (2009). Coarse particulate matter apportionment around a steel smelter plant. *J. Air Waste Manage. Assoc.*, 59(5): 514-519.

McGuire, M.L.; Jeong, C.H.; Slowik, J.G.; Chang, R.Y.W.; Corbin, J.C.; Lu, G.; Mihele, C.; Rehbein, P.J.G.; Sills, D.M.L.; Abbatt, J.P.D.; Brook, J.R.; Evans, G.J. (2011). Elucidating determinants of aerosol composition through particle-type-based receptor modeling. *Atmos. Chem. Phys.*, 11(15): 8133-8155.

McMeeking, G.R.; Morgan, W.T.; Flynn, M.; Highwood, E.J.; Turnbull, K.; Haywood, J.; Coe, H. (2011). Black carbon aerosol mixing state, organic aerosols and aerosol optical properties over the United Kingdom. *Atmos. Chem. Phys.*, 11(17): 9037-9052.

Mehta, B.; Venkataraman, C.; Bhushan, M.; Tripathi, S.N. (2009). Identification of sources affecting fog formation using receptor modeling approaches and inventory estimates of sectoral emissions. *Atmos. Environ.*, 43(6): 1288-1295.

Miller, S.L.; Anderson, M.J.; Daly, E.P.; Milford, J.B. (2002). Source apportionment of exposures to volatile organic compounds. I. Evaluation of receptor models using simulated exposure data. *Atmos. Environ.*, 36(22): 3629-3641.

Mohr, C.; Richter, R.; DeCarlo, P.F.; Prevot, A.S.H.; Baltensperger, U. (2011). Spatial variation of chemical composition and sources of submicron aerosol in Zurich during wintertime using mobile aerosol mass spectrometer data. *Atmos. Chem. Phys.*, 11(15): 7465-7482.

Mooibroek, D.; Schaap, M.; Weijers, E.P.; Hoogerbrugge, R. (2011). Source apportionment and spatial variability of PM(2.5) using measurements at five sites in the Netherlands. *Atmos. Environ.*, 45(25): 4180-4191.

Moon, K.J.; Han, J.S.; Ghim, Y.S.; Kim, Y.J. (2008). Source apportionment of fine carbonaceous particles by positive matrix factorization at Gosan background site in East Asia. *Environ. Int.*, 34(5): 654-664.

Moreno, T.; Perez, N.; Querol, X.; Amato, F.; Alastuey, A.; Bhatia, R.; Spiro, B.; Hanvey, M.; Gibbons, W. (2010). Physicochemical variations in atmospheric aerosols recorded at sea onboard the Atlantic-Mediterranean 2008 Scholar Ship cruise (Part II): Natural versus anthropogenic influences revealed by PM_{10} trace element geochemistry. *Atmos. Environ.*, 44(21-22): 2563-2576.

Morino, Y.; Ohara, T.; Yokouchi, Y.; Ooki, A. (2011). Comprehensive source apportionment of volatile organic compounds using observational data, two receptor models, and an emission inventory in Tokyo metropolitan area. *Journal of Geophysical Research-Atmospheres*, 116

Morishita, M.; Keeler, G.J.; Wagner, J.G.; Harkema, J.R. (2006). Source identification of ambient $PM_{2.5}$ during summer inhalation exposure studies in Detroit, MI. *Atmos. Environ.*, 40(21): 3823-3834. ISI:000238827200001.

Morishita, M.; Keeler, G.J.; Kamal, A.S.; Wagner, J.G.; Harkema, J.R.; Rohr, A.C. (2011). Identification of ambient $PM_{2.5}$ sources and analysis of pollution episodes in Detroit, Michigan using highly time-resolved measurements. *Atmos. Environ.*, 45(8): 1627-1637.

Ng, N.L.; Herndon, S.C.; Trimborn, A.; Canagaratna, M.R.; Croteau, P.L.; Onasch, T.B.; Sueper, D.; Worsnop, D.R.; Zhang, Q.; Sun, Y.L.; Jayne, J.T. (2011). An Aerosol Chemical Speciation Monitor (ACSM) for routine monitoring of the composition and mass concentrations of ambient aerosol. *Aerosol Sci. Technol.*, 45(7): 770-784.

Ng, N.L.; Canagaratna, M.R.; Jimenez, J.L.; Zhang, Q.; Ulbrich, I.M.; Worsnop, D.R. (2011). Real-time methods for estimating organic component mass concentrations from Aerosol Mass Spectrometer data. *Environ. Sci. Technol.*, 45(3): 910-916.

Nicolas, J.; Chiari, M.; Crespo, J.; Orellana, I.G.; Lucarelli, F.; Nava, S.; Pastor, C.; Yubero, E. (2008). Quantification of Saharan and local dust impact in an arid Mediterranean area by the positive matrix factorization (PMF) technique. *Atmos. Environ.*, 42(39): 8872-8882.

Nicolas, J.; Chiari, M.; Crespo, J.; Galindo, N.; Lucarelli, F.; Nava, S.; Yubero, E. (2011). Assessment of potential source regions of PM$_{2.5}$ components at a southwestern Mediterranean site. *Tellus Series B-Chemical and Physical Meteorology*, 63(1): 96-106.

Norman, A.L.; Barrie, L.A.; Toom-Sauntry, D.; Sirois, A.; Krouse, H.R.; Li, S.M.; Sharma, S. (1999). Sources of aerosol sulphate at Alert: Apportionment using stable isotopes. *J. Geophys. Res.*, 104(D9): 11619-11631.

Norris, G., Vedantham, R., Wade, K., Zahn, P., Brown, S., Paatero, P., Eberly, S., and Foley, C. (2009) Guidance document for PMF applications with the Multilinear Engine. EPA 600/R-09/032, Prepared for the U.S. Environmental Protection Agency, Research Triangle Park, NC, April.

Ogulei, D.; Hopke, P.K.; Wallace, L.A. (2006). Analysis of indoor particle size distributions in an occupied townhouse using positive matrix factorization. *Indoor Air*, 16(3): 204-215.

Ogulei, D.; Hopke, P.K.; Zhou, L.M.; Pancras, J.P.; Nair, N.; Ondov, J.M. (2006). Source apportionment of Baltimore aerosol from combined size distribution and chemical composition data. *Atmos. Environ.*, 40(Suppl. 2): S396-S410.

Ogulei, D.; Hopke, P.K.; Ferro, A.R.; Jaques, P.A. (2007). Factor analysis of submicron particle size distributions near a major United States-Canada trade bridge. *J. Air Waste Manage. Assoc.*, 57(2): 190-203.

Oh, M.S.; Lee, T.J.; Kim, D.S. (2011). Quantitative source apportionment of size-segregated particulate matter at urbanized local site in Korea. *AAQR*, 11(3): 247-264.

Owega, S.; Khan, B.U.Z.; D'Souza, R.; Evans, G.J.; Fila, M.; Jervis, R.E. (2004). Receptor modeling of Toronto PM$_{2.5}$ characterized by aerosol laser ablation mass spectrometry. *Environ. Sci. Technol.*, 38(21): 5712-5720.

Paatero, J.; Hopke, P.K.; Song, X.H.; Ramadan, Z. (2002). Understanding and controlling rotations in factor analytic models. *Chemom. Intell. Lab. Sys.*, 60(1-2): 253-264. doi:10.1016/S0169-7439(01)00200-3.

Paatero, P.; Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5: 111-126.

Paatero, P. (1997). Least squares formulation of robust non-negative factor analysis. *Chemom. Intell. Lab. Sys.*, 37: 23-35.

Paatero, P. (1998). User's guide for positive matrix factorization programs PMF2 and PMF3 Part 1: Tutorial. prepared by University of Helsinki, Helsinki, Finland,

Paatero, P. (1999). The multilinear engine-A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *Journal of Computational and Graphical Statistics*, 8: 854-888.

Paatero, P. (2000). User's guide for positive matrix factorization programs PMF2 and PMF3 Part 2: Reference. prepared by University of Helsinki, Helsinki, Finland,

Paatero, P.; Hopke, P.K.; Song, X.H.; Ramadan, Z. (2002). Understanding and controlling rotations in factor analytical models. *Chemom. Intell. Lab. Sys.*, 60: 253-264.

Paatero, P.; Hopke, P.K.; Hoppenstock, J.; Eberly, S.I. (2003). Advanced factor analysis of spatial distributions of PM$_{2.5}$ in the eastern United States. *Environ. Sci. Technol.*, 37(11): 2460-2476.

Paatero, P.; Hopke, P.K.; Begum, B.A.; Biswas, S.K. (2005). A graphical diagnostic method for assessing the rotation in factor analytical models of atmospheric pollution. *Atmos. Environ.*, 39(1): 193-201.

Paatero P. and Hopke P.K. (2008) Rotational tools for factor analytic models implemented by using the multilinear engine. *Chemometrics*. 23 (2): 91-100

Paatero, P., Eberly, S., Brown, S. G., and Norris, G. A.(2014) "Methods for estimating uncertainty in factor analytic solutions", Atmos. Meas. Tech., 7, 781-797, doi:10.5194/amt-7-781-2014.

Pancras, J.P.; Ondov, J.M.; Poor, N.; Landis, M.S.; Stevens, R.K. (2006). Identification of sources and estimation of emission profiles from highly time-resolved pollutant measurements in Tampa, FL. *Atmos. Environ.*, 40(Suppl. 2): S467-S481.

Pancras J.P., Ondov J.M., Zeisler R. (2005) Multi-element electrothermal AAS determination of 11 marker elements in fine ambient aerosol slurry samples collected with SEAS-II. Analytica Chimica Acta 538: 303–312.

Pandolfi, M.; Viana, M.; Minguillon, M.C.; Querol, X.; Alastuey, A.; Amato, F.; Celades, I.; Escrig, A.; Monfort, E. (2008). Receptor models application to multi-year ambient PM_{10} measurements in an industrialized ceramic area: Comparison of source apportionment results. *Atmos. Environ.*, 42(40): 9007-9017.

Paterson, K.G.; Sagady, J.L.; Hooper, D.L. (1999). Analysis of air quality data using positive matrix factorization. *Environ. Sci. Technol.*, 33(4): 635-641.

Pekney, N.J.; Davidson, C.I.; Zhou, L.M.; Hopke, P.K. (2006a). Application of PSCF and CPF to PMF-modeled sources of $PM_{2.5}$ in Pittsburgh. *Aerosol Sci. Technol.*, 40(10): 952-961.

Pekney, N.J.; Davidson, C.I.; Bein, K.J.; Wexler, A.S.; Johnston, M.V. (2006b). Identification of sources of atmospheric PM at the Pittsburgh Supersite, Part I: Single particle analysis and filter-based positive matrix factorization. *Atmos. Environ.*, 40(Suppl. 2): S411-S423.

Pekney, N.J.; Davidson, C.I.; Robinson, A.; Zhou, L.M.; Hopke, P.K.; Eatough, D.J.; Rogge, W.F. (2006c). Major source categories for $PM_{2.5}$ in Pittsburgh using PMF and UNMIX. *Aerosol Sci. Technol.*, 40(10): 910-924.

Pitz, M.; Gu, J.; Soentgen, J.; Peters, A.; Cyrys, J. (2011). Particle size distribution factor as an indicator for the impact of the Eyjafjallajokull ash plume at ground level in Augsburg, Germany. *Atmos. Chem. Phys.*, 11(17): 9367-9374.

Poirot, R.L.; Wishinski, P.R.; Hopke, P.K.; Polissar, A.V. (2001). Comparitive application of multiple receptor methods to identify aerosol sources in northern Vermont. *Environ. Sci. Technol.*, 35(23): 4622-4636.

Poirot, R.L.; Wishinski, P.R.; Hopke, P.K.; Polissar, A.V. (2002). Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont (vol 35, pg 4622, 2001). *Environ. Sci. Technol.*, 36(4): 820.

Polissar, A.V.; Hopke, P.K.; Paatero, P.; Malm, W.C.; Sisler, J.F. (1998). Atmospheric aerosol over Alaska 2. Elemental composition and sources. *J. Geophys. Res.*, 103(D15): 19045-19057.

Polissar, A.V.; Hopke, P.K.; Paatero, P.; Kaufmann, Y.J.; Hall, D.K.; Bodhaine, B.A.; Dutton, E.G.; Harris, J.M. (1999). The aerosol at Barrow, Alaska: Long-term trends and source locations. *Atmos. Environ.*, 33(16): 2441-2458.

Polissar, A.V.; Hopke, P.K.; Poirot, R.L. (2001). Atmospheric aerosol over Vermont: Chemical composition and sources. *Environ. Sci. Technol.*, 35(23): 4604-4621.

Polissar, A.V.; Hopke, P.K.; Harris, J.M. (2001). Source regions for atmospheric aerosol measured at Barrow, Alaska. *Environ. Sci. Technol.*, 35(21): 4214-4226.

Politis D.N. and White H. (2003) Automatic block-length selection for the dependent bootstrap. Prepared by the University of California at San Diego, La Jolla, CA, February.

Prendes, P.; Andrade, J.M.; Lopez-Maha, P. (1999). Source apportionment of inorganic ions in airborne urban particles from Coruna City using positive matrix factorization. *Talanta*, 49(1): 165.

Qi, L.; Nakao, S.; Malloy, Q.; Warren, B.; Cocker, D.R. (2010). Can secondary organic aerosol formed in an atmospheric simulation chamber continuously age? *Atmos. Environ.*, 44(25): 2990-2996.

Qin, Y.; Oduyemi, K.; Chan, L.Y. (2002). Comparative testing of PMF and CFA models. *Chemom. Intell. Lab. Sys.*, 61(1-2): 75-87. doi:10.1016/S0169-7439(01)00175-7.

Qin, Y.; Oduyemi, K. (2003). Atmospheric aerosol source identification and estimates of source contributions to air pollution in Dundee, UK. *Atmos. Environ.*, 37(13): 1799-1809.

Qin, Y.J.; Kim, E.; Hopke, P.K. (2006). The concentrations and sources of $PM_{2.5}$ in metropolitan New York city. *Atmos. Environ.*, 40(Suppl.2): S312-S332.

Raatikainen, T.; Vaattovaara, P.; Tiitta, P.; Miettinen, P.; Rautiainen, J.; Ehn, M.; Kulmala, M.; Laaksonen, A.; Worsnop, D.R. (2010). Physicochemical properties and origin of organic groups detected in boreal forest using an aerosol mass spectrometer. *Atmos. Chem. Phys.*, 10(4): 2063-2077.

Raja, S.; Biswas, K.F.; Husain, L.; Hopke, P.K. (2010). Source apportionment of the atmospheric aerosol in Lahore, Pakistan. *Water Air and Soil Pollution*, 208(1-4): 43-57.

Ramadan, Z.; Song, X.H.; Hopke, P.K. (2000). Identification of sources of Phoenix aerosol by positive matrix factorization. *J. Air Waste Manage. Assoc.*, 50(8): 1308-1320.

Ramadan, Z.; Eickhout, B.; Song, X.H.; Buydens, L.M.C.; Hopke, P.K. (2003). Comparison of positive matrix factorization and multilinear engine for the source apportionment of particulate pollutants. *Chemom. Intell. Lab. Sys.*, 66(1): 15-28. doi:10.1016/S0169-7439(02)00160-0.

Raman, R.S.; Hopke, P.K. (2007). Source apportionment of fine particles utilizing partially speciated carbonaceous aerosol data at two rural locations in New York State. *Atmos. Environ.*, 41: 7923-7939.

Raman, R.S.; Ramachandran, S. (2010). Annual and seasonal variability of ambient aerosols over an urban region in western India. *Atmos. Environ.*, 44(9): 1200-1208.

Raman, R.S.; Ramachandran, S.; Kedia, S. (2011). A methodology to estimate source-specific aerosol radiative forcing. *J. Aerosol Sci.*, 42(5): 305-320.

Raman, R.S.; Ramachandran, S. (2011). Source apportionment of the ionic components in precipitation over an urban region in Western India. *Environmental Science and Pollution Research*, 18(2): 212-225.

Reff, A.; Eberly, S.I.; Bhave, P.V. (2007). Receptor modeling of ambient particulate matter data using positive matrix factorization: Review of existing methods. *J. Air Waste Manage. Assoc.*, 57(2): 146-154.

Richard, A.; Gianini, M.F.D.; Mohr, C.; Furger, M.; Bukowiecki, N.; Minguillon, M.C.; Lienemann, P.; Flechsig, U.; Appel, K.; DeCarlo, P.F.; Heringa, M.F.; Chirico, R.; Baltensperger, U.; Prevot, A.S.H. (2011). Source apportionment of size and time resolved trace elements and organic aerosols from an urban courtyard site in Switzerland. *Atmos. Chem. Phys.*, 11(17): 8945-8963.

Rizzo, M.J.; Scheff, P.A. (2004). Assessing ozone networks using positive matrix factorization. *Environmental Progress*, 23(2): 110-119.

Rizzo, M.J.; Scheff, P.A. (2007). Fine particulate source apportionment using data from the USEPA speciation trends network in Chicago, Illinois: Comparison of two source apportionment models. *Atmos. Environ.*, 41(29): 6276-6288.

Rizzo, M.J.; Scheff, P.A. (2007). Utilizing the Chemical Mass Balance and Positive Matrix Factorization models to determine influential species and examine possible rotations in receptor modeling results. *Atmos. Environ.*, 41(33): 6986-6998.

Robinson, N.H.; Hamilton, J.F.; Allan, J.D.; Langford, B.; Oram, D.E.; Chen, Q.; Docherty, K.; Farmer, D.K.; Jimenez, J.L.; Ward, M.W.; Hewitt, C.N.; Barley, M.H.; Jenkin, M.E.; Rickard, A.R.; Martin, S.T.; McFiggans, G.; Coe, H. (2011). Evidence for a significant proportion of Secondary Organic Aerosol from isoprene above a maritime tropical forest. *Atmos. Chem. Phys.*, 11(3): 1039-1050.

Rodriguez, S.; Alastuey, A.; Alonso-Perez, S.; Querol, X.; Cuevas, E.; Abreu-Afonso, J.; Viana, M.; Perez, N.; Pandolfi, M.; de la Rosa, J. (2011). Transport of desert dust mixed with North African industrial pollutants in the subtropical Saharan Air Layer. *Atmos. Chem. Phys.*, 11(13): 6663-6685.

Santoso, M.; Hopke, P.K.; Hidayat, A.; Diah, D.L. (2008). Source identification of the atmospheric aerosol at urban and suburban sites in Indonesia by positive matrix factorization. *Sci. Total Environ.*, 397(1-3): 229-237.

Sarnat, J.A.; Marmur, A.; Klein, M.; Kim, E.; Russell, A.G.; Sarnat, S.E.; Mulholland, J.A.; Hopke, P.K.; Tolbert, P.E. (2008). Fine particle sources and cardiorespiratory morbidity: An application of chemical mass balance and factor analytical source-apportionment methods. *Environ. Health Perspect.*, 116(4): 459-466.

Sauvage, S.; Plaisance, H.; Locoge, N.; Wroblewski, A.; Coddeville, P.; Galloo, J.C. (2009). Long term measurement and source apportionment of non-methane hydrocarbons in three French rural areas. *Atmos. Environ.*, 43(15): 2430-2441.

Schnelle-Kreis, J.; Sklorz, M.; Orasche, J.; Stolzel, M.; Peters, A.; Zimmermann, R. (2007). Semi volatile organic compounds in ambient PM$_{2.5}$. Seasonal trends and daily resolved source contributions. *Environ. Sci. Technol.*, 41(11): 3821-3828.

Shi, G.L.; Li, X.; Feng, Y.C.; Wang, Y.Q.; Wu, J.H.; Li, J.; Zhu, T. (2009). Combined source apportionment, using positive matrix factorization-chemical mass balance and principal component analysis/multiple linear regression-chemical mass balance models. *Atmos. Environ.*, 43(18): 2929-2937.

Shim, C.; Wang, Y.; Yoshida, Y. (2008). Evaluation of model-simulated source contributions to tropospheric ozone with aircraft observations in the factor-projected space. *Atmos. Chem. Phys.*, 8(6): 1751-1761.

Shrivastava, M.K.; Subramanian, R.; Rogge, W.F.; Robinson, A.L. (2007). Sources of organic aerosol: Positive matrix factorization of molecular marker data and comparison of results from different source apportionment models. *Atmos. Environ.*, 41(40): 9353-9369.

Slowik, J.G.; Vlasenko, A.; McGuire, M.; Evans, G.J.; Abbatt, J.P.D. (2010). Simultaneous factor analysis of organic particle and gas mass spectra: AMS and PTR-MS measurements at an urban site. *Atmos. Chem. Phys.*, 10(4): 1969-1988.

Slowik, J.G.; Brook, J.; Chang, R.Y.W.; Evans, G.J.; Hayden, K.; Jeong, C.H.; Li, S.M.; Liggio, J.; Liu, P.S.K.; McGuire, M.; Mihele, C.; Sjostedt, S.; Vlasenko, A.; Abbatt, J.P.D. (2011). Photochemical processing of organic aerosol at nearby continental sites: contrast between urban plumes and regional aerosol. *Atmos. Chem. Phys.*, 11(6): 2991-3006.

Sofowote, U.M.; McCarry, B.E.; Marvin, C.H. (2008). Source apportionment of PAH in Hamilton Harbour suspended sediments: Comparison of two factor analysis methods. *Environ. Sci. Technol.*, 42(16): 6007-6014.

Sofowote, U.M.; Hung, H.; Rastogi, A.K.; Westgate, J.N.; Deluca, P.F.; Su, Y.S.; McCarry, B.E. (2011). Assessing the long-range transport of PAH to a sub-Arctic site using positive matrix factorization and potential source contribution function. *Atmos. Environ.*, 45(4): 967-976.

Song, X.H.; Polissar, A.V.; Hopke, P.K. (2001). Sources of fine particle composition in the northeastern US. *Atmos. Environ.*, 35(31): 5277-5286.

Song, Y.; Zhang, Y.H.; Xie, S.D.; Zeng, L.M.; Zheng, M.; Salmon, L.G.; Shao, M.; Slanina, S. (2006). Source apportionment of PM$_{2.5}$ in Beijing by positive matrix factorization. *Atmos. Environ.*, 40(8): 1526-1537. ISI:000236306800012.

Song, Y.; Zhang, Y.H.; Xie, S.D.; Zeng, L.M.; Zheng, M.; Salmon, L.G.; Shao, M.; Slanina, S. (2006). Source apportionment of PM$_{2.5}$ in Beijing by positive matrix factorization (vol 40, pg 1526, 2006). *Atmos. Environ.*, 40(39): 7661-7662. ISI:000242289800018.

Song, Y.; Xie, S.D.; Zhang, Y.H.; Zeng, L.M.; Salmon, L.G.; Zheng, M. (2006). Source apportionment of PM$_{2.5}$ in Beijing using principal component analysis/absolute principal component scores and UNMIX. *Sci. Total Environ.*, 372(1): 278-286.

Song, Y.; Shao, M.; Liu, Y.; Lu, S.H.; Kuster, W.; Goldan, P.; Xie, S.D. (2007). Source apportionment of ambient volatile organic compounds in Beijing. *Environ. Sci. Technol.*, 41(12): 4348-4353.

Song, Y.; Tang, X.Y.; Xie, S.D.; Zhang, Y.H.; Wei, Y.J.; Zhang, M.S.; Zeng, L.M.; Lu, S.H. (2007). Source apportionment of PM$_{2.5}$ in Beijing in 2004. *J. Hazard. Mat.*, 146(1-2): 124-130.

Song, Y.; Dai, W.; Shao, M.; Liu, Y.; Lu, S.H.; Kuster, W.; Goldan, P. (2008). Comparison of receptor models for source apportionment of volatile organic compounds in Beijing, China. *Environ. Poll.*, 156(1): 174-183.

Song, Y.; Dai, W.; Wang, X.S.; Cui, M.M.; Su, H.; Xie, S.D.; Zhang, Y.H. (2008). Identifying dominant sources of respirable suspended particulates in Guangzhou, China. *Environmental Engineering Science* , 25(7): 959-968.

Soonthornnonda, P.; Christensen, E.R. (2008). Source apportionment of pollutants and flows of combined sewer wastewater. *Water Research*, 42(8-9): 1989-1998.

Sun, Y.L.; Zhang, Q.; Zheng, M.; Ding, X.; Edgerton, E.S.; Wang, X.M. (2011). Characterization and source apportionment of water-soluble organic matter in atmospheric fine particles (PM(2.5)) with high-resolution aerosol mass spectrometry and GC-MS. *Environ. Sci. Technol.*, 45(11): 4854-4861.

Sundqvist, K.L.; Tysklind, M.; Geladi, P.; Hopke, P.K.; Wiberg, K. (2010). PCDD/F source apportionment in the Baltic Sea using positive matrix factorization. *Environ. Sci. Technol.*, 44(5): 1690-1697.

Tandon, A.; Yadav, S.; Attri, A.K. (2010). Coupling between meteorological factors and ambient aerosol load. *Atmos. Environ.*, 44(9): 1237-1243.

Tauler, R.; Viana, M.; Querol, X.; Alastuey, A.; Flight, R.M.; Wentzell, P.D.; Hopke, P.K. (2009). Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies. *Atmos. Environ.*, 43(26): 3989-3997.

Thimmaiah, D.; Hovorka, J.; Hopke, P.K. (2009). Source apportionment of winter submicron Prague aerosols from combined particle number size distribution and gaseous composition data. *AAQR*, 9(2): 209-236.

Thornhill, D.A.; Williams, A.E.; Onasch, T.B.; Wood, E.; Herndon, S.C.; Kolb, C.E.; Knighton, W.B.; Zavala, M.; Molina, L.T.; Marr, L.C. (2010). Application of positive matrix factorization to on-road measurements for source apportionment of diesel- and gasoline-powered vehicle emissions in Mexico City. *Atmos. Chem. Phys.*, 10(8): 3629-3644.

Thurston, G.D.; Ito, K.; Mar, T.; Christensen, W.F.; Eatough, D.J.; Henry, R.C.; Kim, E.; Laden, F.; Lall, R.; Larson, T.V.; Liu, H.; Neas, L.; Pinto, J.; Stolzel, M.; Suh, H.; Hopke, P.K. (2005). Workgroup report: Workshop on source apportionment of particulate matter health effects - Intercomparison of results and implications. *Environ. Health Perspect.*, 113(12): 1768-1774.

Tian, F.L.; Chen, J.W.; Qiao, X.L.; Cai, X.Y.; Yang, P.; Wang, Z.; Wang, D.G. (2008). Source identification of PCDD/Fs and PCBs in pine (Cedrus deodara) needles: A case study in Dalian, China. *Atmos. Environ.*, 42(19): 4769-4777.

Tsai, J.; Owega, S.; Evans, G.; Jervis, R.; Fila, M.; Tan, P.; Malpica, O. (2004). Chemical composition and source apportionment of Toronto summertime urban fine aerosol (PM$_{2.5}$). *Journal of Radioanalytical and Nuclear Chemistry*, 259(1): 193-197.

Tsimpidi, A.P.; Karydis, V.A.; Zavala, M.; Lei, W.; Molina, L.; Ulbrich, I.M.; Jimenez, J.L.; Pandis, S.N. (2010). Evaluation of the volatility basis-set approach for the simulation of organic aerosol formation in the Mexico City metropolitan area. *Atmos. Chem. Phys.*, 10(2): 525-546.

Tsimpidi, A.P.; Karydis, V.A.; Zavala, M.; Lei, W.; Bei, N.; Molina, L.; Pandis, S.N. (2011). Sources and production of organic aerosol in Mexico City: insights from the combination of a chemical transport model (PMCAMx-2008) and measurements during MILAGRO. *Atmos. Chem. Phys.*, 11(11): 5153-5168.

U.S.EPA (2010). EPA Positive Matrix Factorization (PMF) 3.0 model. prepared by U.S. Environmental Protection Agency, Research Triangle Park, NC, http://www.epa.gov/heasd/products/pmf/pmf.html

Uchimiya, M.; Arai, M.; Masunaga, S. (2007). Fingerprinting localized dioxin contamination: Ichihara anchorage case. *Environ. Sci. Technol.*, 41(11): 3864-3870.

Ulbrich, I.M.; Canagaratna, M.R.; Zhang, Q.; Worsnop, D.R.; Jimenez, J.L. (2009). Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data. *Atmos. Chem. Phys.*, 9(9): 2891-2918.

Vaccaro, S.; Sobiecka, E.; Contini, S.; Locoro, G.; Free, G.; Gawlik, B.M. (2007). The application of positive matrix factorization in the analysis, characterisation and detection of contaminated soils. *Chemosphere*, 69: 1055-1063.

Vecchi, R.; Chiari, M.; D'Alessandro, A.; Fermo, P.; Lucarelli, F.; Mazzei, F.; Nava, S.; Piazzalunga, A.; Prati, P.; Silvani, F.; Valli, G. (2008). A mass closure and PMF source apportionment study on the sub-micron sized aerosol fraction at urban sites in Italy. *Atmos. Environ.*, 42(9): 2240-2253.

Vecchi, R.; Bernardoni, V.; Cricchio, D.; D'Alessandro, A.; Fermo, P.; Lucarelli, F.; Nava, S.; Plazzalunga, A.; Valli, G. (2008). The impact of fireworks on airborne particles. *Atmos. Environ.*, 42(6): 1121-1132.

Vedal, S.; Hannigan, M.P.; Dutton, S.J.; Miller, S.L.; Milford, J.B.; Rabinovitch, N.; Kim, S.Y.; Sheppard, L. (2009). The Denver Aerosol Sources and Health (DASH) study: Overview and early findings. *Atmos. Environ.*, 43(9): 1666-1673.

Vestenius, M.; Leppanen, S.; Anttila, P.; Kyllonen, K.; Hatakka, J.; Hellen, H.; Hyvarinen, A.P.; Hakola, H. (2011). Background concentrations and source apportionment of polycyclic aromatic hydrocarbons in south-eastern Finland. *Atmos. Environ.*, 45(20): 3391-3399.

Viana, M.; Pandolfi, M.; Minguillon, M.C.; Querol, X.; Alastuey, A.; Monfort, E.; Celades, I. (2008). Inter-comparison of receptor models for PM source apportionment: Case study in an industrial area. *Atmos. Environ.*, 42(16): 3820-3832.

Viana, M.; Amato, F.; Alastuey, A.; Querol, X.; Moreno, T.; Dos Santos, S.G.; Herce, M.D.; Fernandez-Patier, R. (2009). Chemical tracers of particulate emissions from commercial shipping. *Environ. Sci. Technol.*, 43(19): 7472-7477.

Viana, M.; Salvador, P.; Artinano, B.; Querol, X.; Alastuey, A.; Pey, J.; Latz, A.J.; Cabanas, M.; Moreno, T.; Dos Santos, S.G.; Herce, M.D.; Hernandez, P.D.; Garcia, D.R.; Fernandez-Patier, R. (2010). Assessing the performance of methods to detect and quantify African dust in airborne particulates. *Environ. Sci. Technol.*, 44(23): 8814-8820.

Vlasenko, A.; Slowik, J.G.; Bottenheim, J.W.; Brickell, P.C.; Chang, R.Y.W.; Macdonald, A.M.; Shantz, N.C.; Sjostedt, S.J.; Wiebe, H.A.; Leaitch, W.R.; Abbatt, J.P.D. (2009). Measurements of VOCs by proton transfer reaction mass spectrometry at a rural Ontario site: Sources and correlation to aerosol composition. *Journal of Geophysical Research-Atmospheres*, 114

Wang, D.G.; Tian, F.L.; Yang, M.; Liu, C.L.; Li, Y.F. (2009). Application of positive matrix factorization to identify potential sources of PAHs in soil of Dalian, China. *Environ. Poll.*, 157(5): 1559-1564.

Wang, H.B.; Shooter, D. (2005). Source apportionment of fine and coarse atmospheric particles in Auckland, New Zealand. *Sci. Total Environ.*, 340(1-3): 189-198.

Wang, Y.; Zhuang, G.S.; Tang, A.H.; Zhang, W.J.; Sun, Y.L.; Wang, Z.F.; An, Z.S. (2007). The evolution of chemical components of aerosols at five monitoring sites of China during dust storms. *Atmos. Environ.*, 41(5): 1091-1106.

Wang, Y.G.; Hopke, P.K.; Chalupa, D.C.; Utell, M.J. (2011). Effect of the shutdown of a coal-fired power plant on urban ultrafine particles and other pollutants. *Aerosol Sci. Technol.*, 45(10): 1245-1249.

Watson, J.G.; Chow, J.C. (2004). Receptor models for air quality management. *EM*, 10(Oct.): 27-36.

Watson, J.G.; Chen, L.-W.A.; Chow, J.C.; Lowenthal, D.H.; Doraiswamy, P. (2008). Source apportionment: Findings from the U.S. Supersite Program. *J. Air Waste Manage. Assoc.*, 58(2): 265-288. http://pubs.awma.org/gsearch/journal/2008/2/10.3155-1047-3289.58.2.265.pdf.

Willis, R.D. (2000). Workshop on UNMIX and PMF as applied to PM$_{2.5}$. Report Number EPA/600/A-00/048; prepared by U.S. Environmental Protection Agency, Research Triangle Park, NC, for US EPA,

Wingfors, H.; Hagglund, L.; Magnusson, R. (2011). Characterization of the size-distribution of aerosols and particle-bound content of oxygenated PAHs, PAHs, and n-alkanes in urban environments in Afghanistan. *Atmos. Environ.*, 45(26): 4360-4369.

Wu, C.F.; Larson, T.V.; Wu, S.Y.; Williamson, J.; Westberg, H.H.; Liu, L.J.S. (2007). Source apportionment of PM$_{2.5}$ and selected hazardous air pollutants in Seattle. *Sci. Total Environ.*, 386: 42-52.

Xiao, R.; Takegawa, N.; Zheng, M.; Kondo, Y.; Miyazaki, Y.; Miyakawa, T.; Hu, M.; Shao, M.; Zeng, L.; Gong, Y.; Lu, K.; Deng, Z.; Zhao, Y.; Zhang, Y.H. (2011). Characterization and source apportionment of submicron aerosol with aerosol mass spectrometer during the PRIDE-PRD 2006 campaign. *Atmos. Chem. Phys.*, 11(14): 6911-6929.

Xie, Y.L.; Hopke, P.K.; Paatero, P. (1998). Positive matrix factorizaiton applied to a curved resolution problem. *J. Chemometrics*, 12(6): 357-364.

Xie, Y.L.; Hopke, P.K.; Paatero, P.; Barrie, L.A.; Li, S.M. (1999). Identification of source nature and seasonal variations of Arctic aerosol by positive matrix factorization. *J. Atmos. Sci.*, 56(2): 249-260.

Xie, Y.L.; Hopke, P.K.; Paatero, P.; Barrie, L.A.; Li, S. (1999). Identification of source nature and seasonal variations of Arctic aerosol by the multilinear engine. *Atmos. Environ.*, 33(16): 2549-2562.

Xie, Y.L.; Berkowitz, C.M. (2006). The use of positive matrix factorization with conditional probability functions in air quality studies: An application to hydrocarbon emissions in Houston, Texas. *Atmos. Environ.*, 40(17): 3070-3091.

Yakovleva, E.; Hopke, P.K.; Wallace, L. (1999). Receptor modeling assessment of particle total exposure assessment methodology data. *Environ. Sci. Technol.*, 33(20): 3645-3652.

Yatkin, S.; Bayram, A. (2008). Source apportionment of PM_{10} and $PM_{2.5}$ using positive matrix factorization and chemical mass balance in Izmir, Turkey. *Sci. Total Environ.*, 390(1): 109-123.

Yli-Tuomi, T.; Paatero, P.; Raunemaa, T. (1996). The soil factor in Rautavaara aerosol in positive matrix factorization solutions with 2 to 8 factors. *J. Aerosol Sci.*, 27(supplement 1): S671-S672. doi:10.1016/0021-8502(96)00408-9.

Yli-Tuomi, T.; Hopke, P.K.; Paatero, P.; Basunia, M.S.; Landsberger, S.; Viisanen, Y.; Paatero, J. (2003). Atmospheric aerosol over Finnish Arctic: Source analysis by the multilinear engine and the potential source contribution function. *Atmos. Environ.*, 37(31): 4381-4392. doi: 10.1016/S1352-2310(03)00569-7.

Yu, J.Z.; Yang, H.; Zhang, H.Y.; Lau, A.K.H. (2004). Size distributions of water-soluble organic carbon in ambient aerosols and its size-resolved thermal characteristics. *Atmos. Environ.*, 38(7): 1061-1071.

Yuan, H.; Zhuang, G.S.; Li, J.; Wang, Z.F.; Li, J. (2008). Mixing of mineral with pollution aerosols in dust season in Beijing: Revealed by source apportionment study. *Atmos. Environ.*, 42(9): 2141-2157.

Yuan, Z.B.; Yu, J.Z.; Lau, A.K.H.; Louie, P.K.K.; Fung, J.C.H. (2006). Application of positive matrix factorization in estimating aerosol secondary organic carbon in Hong Kong and its relationship with secondary sulfate. *Atmos. Chem. Phys.*, 6(1): 25-34.

Yuan, Z.B.; Lau, A.K.H.; Zhang, H.Y.; Yu, J.Z.; Louie, P.K.K.; Fung, J.C.H. (2006). Identification and spatiotemporal variations of dominant PM_{10} sources over Hong Kong. *Atmos. Environ.*, 40(10): 1803-1815.

Yuan, Z.B.; Lau, A.K.H.; Shao, M.; Louie, P.K.K.; Liu, S.C.; Zhu, T. (2009). Source analysis of volatile organic compounds by positive matrix factorization in urban and rural environments in Beijing. *Journal of Geophysical Research-Atmospheres*, 114

Yuan, B., Min Shao, M.; Gouw, J.; David D. Parrish, D.; Lu, S.; Wang, M.; Zeng, L.; Zhang, Q.; Song, Y.;

Zhang, J.;Hu, M, (2012), Volatile organic compounds (VOCs) in urban air: How chemistry affects the interpretation of positive matrix factorization (PMF) analysis, J. Geophys. Res., 117

Yue, W.; Stolzel, M.; Cyrys, J.; Pitz, M.; Heinrich, J.; Kreyling, W.G.; Wichmann, H.E.; Peters, A.; Wang, S.; Hopke, P.K. (2008). Source apportionment of ambient fine particle size distribution using positive matrix factorization in Erfurt, Germany. *Sci. Total Environ.*, 398(1-3): 133-144.

Zhang, Q.; Alfarra, M.R.; Worsnop, D.R.; Allan, J.D.; Coe, H.; Canagaratna, M.R.; Jimenez, J.L. (2005). Deconvolution and quantification of hydrocarbon-like and oxygenated organic aerosols based on aerosol mass spectrometry. *Environ. Sci. Technol.*, 39(13): 4938-4952.

Zhang, W.; Guo, J.H.; Sun, Y.L.; Yuan, H.; Zhuang, G.S.; Zhuang, Y.H.; Hao, Z.P. (2007). Source apportionment for,urban PM_{10} and $PM_{2.5}$ in the Beijing area. *Chinese Science Bulletin*, 52(5): 608-615.

Zhang, Y.; Sheesley, R.J.; Schauer, J.J.; Lewandowski, M.; Jaoui, M.; Offenberg, J.H.; Kleindienst, T.E.; Edney, E.O. (2009). Source apportionment of primary and secondary organic aerosols using positive matrix factorization (PMF) of molecular markers. *Atmos. Environ.*, 43(34): 5567-5574.

Zhang, Y.X.; Schauer, J.J.; Shafer, M.M.; Hannigan, M.P.; Dutton, S.J. (2008). Source apportionment of in vitro reactive oxygen species bioassay activity from atmospheric particulate matter. *Environ. Sci. Technol.*, 42(19): 7502-7509.

Zhang, Y.X.; Sheesley, R.J.; Bae, M.S.; Schauer, J.J. (2009). Sensitivity of a molecular marker based positive matrix factorization model to the number of receptor observations. *Atmos. Environ.*, 43(32): 4951-4958.

Zhao, W.; Hopke, P.K.; Karl, T. (2004). Source identification of volatile organic compounds in Houston, Texas. *Environ. Sci. Technol.*, 38(5): 1338-1347.

Zhao, W.X.; Hopke, P.K. (2004). Source apportionment for ambient particles in the San Gorgonio wilderness. *Atmos. Environ.*, 38(35): 5901-5910.

Zhao, W.X.; Hopke, P.K. (2006). Source identification for fine aerosols in Mammoth Cave National Park. *Atmos. Res.*, 80(4): 309-322.

Zhao, W.X.; Hopke, P.K. (2006). Source investigation for ambient $PM_{2.5}$ in Indianapolis, IN. *Aerosol Sci. Technol.*, 40(10): 898-909.

Zhou, L.; Hopke, P.K.; Zhao, W.X. (2009). Source apportionment of airborne particulate matter for the Speciation Trends Network site in Cleveland, OH. *J. Air Waste Manage. Assoc.*, 59(3): 321-331.

Zhou, L.M.; Kim, E.; Hopke, P.K.; Stanier, C.O.; Pandis, S.N. (2004). Advanced factor analysis on Pittsburgh particle size-distribution data. *Aerosol Sci. Technol.*, 38(Suppl. 1): 118-132.

Zhou, L.M.; Hopke, P.K.; Liu, W. (2004). Comparison of two trajectory based models for locating particle sources for two rural New York sites. *Atmos. Environ.*, 38(13): 1955-1963.

Zhou, L.M.; Hopke, P.K.; Stanier, C.O.; Pandis, S.N.; Ondov, J.M.; Pancras, J.P. (2005). Investigation of the relationship between chemical composition and size distribution of airborne particles by partial least squares and positive matrix factorization. *Journal of Geophysical Research-Atmospheres*, 110(D7)

Zhou, L.M.; Kim, E.; Hopke, P.K.; Stanier, C.; Pandis, S.N. (2005). Mining airborne particulate size distribution data by positive matrix factorization. *Journal of Geophysical Research-Atmospheres*, 110(D7): D07S19. doi:10.1029/2004JD004707.

Zota, A.R.; Willis, R.; Jim, R.; Norris, G.A.; Shine, J.P.; Duvall, R.M.; Schaider, L.A.; Spengler, J.D. (2009). Impact of mine waste on airborne respirable particulates in northeastern Oklahoma, United States. *J. Air Waste Manage. Assoc.*, 59(11): 1347-1357.