Atmospheric
Measurement
Techniques

Open Access

Methods for estimating uncertainty in factor analytic solutions

P. Paatero[1], **S. Eberly**[2], **S. G. Brown**[3], and **G. A. Norris**[4]

[1]University of Helsinki, Dept. of Physics, Helsinki, Finland
[2]Geometric Tools, LLC, Redmond, Washington, USA
[3]Sonoma Technology, Inc., Petaluma, California, USA
[4]US EPA, Office of Research and Development, Research Triangle Park, North Carolina, USA

Correspondence to: G. A. Norris (norris.gary@epa.gov)

Abstract. The EPA PMF (Environmental Protection Agency positive matrix factorization) version 5.0 and the underlying multilinear engine-executable ME-2 contain three methods for estimating uncertainty in factor analytic models: classical bootstrap (BS), displacement of factor elements (DISP), and bootstrap enhanced by displacement of factor elements (BS-DISP). The goal of these methods is to capture the uncertainty of PMF analyses due to random errors and rotational ambiguity. It is shown that the three methods complement each other: depending on characteristics of the data set, one method may provide better results than the other two. Results are presented using synthetic data sets, including interpretation of diagnostics, and recommendations are given for parameters to report when documenting uncertainty estimates from EPA PMF or ME-2 applications.

1 Introduction

1.1 EPA PMF and ME-2

The multivariate factor analysis tools PMF2, ME-2, and EPA PMF (Environmental Protection Agency positive matrix factorization, which is built on ME-2) are widely used for numerous applications, particularly for analyses of ambient air quality data (Poirot et al., 2001; Reff et al., 2007; Kim and Hopke, 2007; Engel-Cox and Weber, 2007; Norris et al., 2008; Ke et al., 2008; Ulbrich et al., 2009; Brown et al., 2012). Each tool performs a positive matrix factorization (PMF) that decomposes a matrix of speciated sample data into two matrices – factor contributions and factor profiles. A speciated data set may be viewed as a data matrix $\mathbf{X}$ of

dimensions n by m, in which n samples and m chemical species were measured. Rows and columns of $\mathbf{X}$ and of related matrices are indexed by i and j, respectively. The goal of modeling with PMF is to identify the number of factors p, the species profile $\boldsymbol{f}_k$ of each factor k, and the amount of mass $\boldsymbol{g}_k$ contributed by each factor k to each individual sample (Eq. 1):

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

where e_{ij} is the residual and c_{ij} denotes the modeled part for each sample/species. The method is described in greater detail elsewhere (Paatero and Tapper, 1994; Paatero, 1997). Regarding notation, capital bold-face letters denote entire matrices, $\boldsymbol{g}_k$ denotes columns of the factor contribution matrix $\mathbf{G}$, and $\boldsymbol{f}_k$ denotes rows of factor profile matrix $\mathbf{F}$.

Original versions of PMF2, ME-2, and EPA PMF provided uncertainty estimates for $\mathbf{F}$ and sometimes $\mathbf{G}$. However, these estimates did not explicitly include rotational uncertainty of the results. The present work corrects this deficiency for ME-2 and EPA PMF, presents three methods for estimating uncertainty, and discusses each method's strengths and weaknesses. The error estimation methods described in this work have been implemented in version 5.0 of EPA PMF, to be released in 2013. See http://www.epa.gov/heasd/research/pmf.html and Norris et al. (2008) for details.

1.2 Two interpretations of Eq. (1)

Equation (1) may be employed in two ways. One is when $\mathbf{F}$ and $\mathbf{G}$ contain known values. This approach is used when generating simulated data that mimic real measurements. In

this case, the *data errors* e_{ij} are pseudorandom values, often generated from normal distributions with mean zero and standard deviation equal to s_{ij}. These *data uncertainties* s_{ij} are known values specified in a simulation. Multiplying $\mathbf{F}$ and $\mathbf{G}$ and adding $\mathbf{E}$ produces $\mathbf{X}$, the simulated matrix of measurements to be modeled by PMF. Fitted values for $\mathbf{F}$ and $\mathbf{G}$ can then be compared to the true values that were used to simulate $\mathbf{X}$.

Alternatively, Eq. (1) may be employed when the measured (or simulated) matrix $\mathbf{X}$ is known and the matrix of *estimated data uncertainties* u_{ij} has been estimated. This approach is used to determine the values of unknown matrices $\mathbf{F}$ and $\mathbf{G}$. In simulations, data uncertainties u_{ij} may be set equal to uncertainties s_{ij}. When analyzing real data, data uncertainties u_{ij} are estimated by the users so that u_{ij} approximate the unknown true uncertainties s_{ij}. In some situations, *adjusted data uncertainties* are used. For example, to downweight species j, one may set $u_{ij} \approx 3\,s_{ij}$ for a chosen species j.

1.3   Details of the PMF model

In PMF, factor elements are constrained so that no sample can have a significantly negative factor contribution. Also, PMF allows each data value to be individually weighted. This feature allows analysts to adjust the influence of each data point, depending on the confidence in the measurement. For example, data below detection limit can be retained for use in the model with the associated uncertainty adjusted to give these data points less influence on the solution than data above the detection limit. The PMF solution minimizes the object function Q (Eq. 2) based upon the estimated data uncertainties (or adjusted data uncertainties) u_{ij} and with factor matrix elements g_{ik} and f_{kj} subject to non-negativity constraints.

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

ME-2 performs iterations via the conjugate gradient algorithm until convergence to a minimum Q value.

1.4   Origins of uncertainty in PMF analyses

$\mathbf{F}^1$ and $\mathbf{G}^1$ are used to denote a solution of Eq. (1) obtained by solving Eq. (2). Uncertainty analysis of PMF modeling attempts to estimate a range or interval of plausible values around each element of matrix $\mathbf{F}^1$. This interval is estimated so that with a high probability it will include the true value of $\mathbf{F}$. The ends of the range will be called *upper and lower interval estimates* of $\mathbf{F}$ or simply *upper and lower estimates* of $\mathbf{F}$. The uncertainty analysis must take into account all aspects of solving Eq. (2), such as non-negativity constraints.

Uncertainty in PMF analyses arises from three main causes, as described below: (1) random errors in data values; (2) rotational ambiguity; and (3) modeling errors.

Random errors in data values are those that arise from the measurement process, even if measurement systems have been properly calibrated so that no systematic bias is present. All measured data contain random errors – measure something twice and two different values will be obtained.

Uncertainty caused by rotational ambiguity is specific to factor analytic models. Rotational ambiguity arises because bilinear factor analytic models are ill-posed, meaning there are multiple solutions ($\mathbf{G}$, $\mathbf{F}$) with the same value of Q (Henry, 1987). In some special cases, a rotationally unique solution is possible if there are a sufficient number of zero values in true matrices $\mathbf{G}$ and $\mathbf{F}$ (Anderson, 1984). In analytical chemistry (AC), presence of zero values is often known a priori. For example in chromatograms, each component ("peak") shape begins with a number of zero values. Non-zero values are not possible at a large distance before the peak proper. In comparison, presence of zero values in true factors is less predictable in environmental measurements. In most cases, some rotational ambiguity remains in environmental modeling.

The problem of rotations has been discussed in Paatero et al. (2002) and recently in more detail in Paatero and Hopke (2009). The present work offers numerical methods for estimating the rotational non-uniqueness for any given data set where unknown numbers of zero values may be present in true factors. However, the present work does not explicitly consider the effect of inserting additional numerical constraints on factors, as suggested by Paatero and Hopke (2009) and demonstrated by Amato et al. (2009) and by Amato and Hopke (2012).

The extent of possible rotations is limited by non-negativity constraints imposed on the solution and by the number of zero values present in the fitted $\mathbf{G}$ and $\mathbf{F}$ (Paatero et al., 2002). With a small number of zero values, this uncertainty may dominate other types of uncertainty. In extreme cases, presence of large rotations may prevent well-defined modeling altogether, as discussed later in detail. With larger numbers of zero values, rotations may be present in small amounts, so that a useful albeit non-unique solution is possible. Sometimes, rotational uniqueness may be observed, especially if only a small number of factors is fitted. It is seen that measurements should be arranged so that variation between individual samples is as large as possible, e.g., making sure that measurements are performed during different weather patterns. Note that the terms "rotational ambiguity" and "rotational uncertainty" are used here to represent slightly different ideas. "Rotational ambiguity" denotes the concept that multiple mathematical solutions can yield the same or practically the same fit (one with almost identical Q values). "Rotational uncertainty" is used when discussing the amount of rotational ambiguity in a more quantitative sense.

Modeling errors are those caused by using a model that is a simplification of the true physical–chemical phenomena. The PMF model describes what is believed to happen in nature. However, modeling errors can arise if the real process in nature is different from what is captured in Eq (1). Some examples include variation of source profiles with time (e.g., because of chemical transformations during transit or chemical variations in the source itself), incorrectly specified number of factors p, incorrectly estimated data uncertainties u_{ij}, contamination of samples, correlated (i.e., non-random or systematic) errors in data values, and weak or sporadic sources that cannot be represented by dedicated factors. Adjustments to measured data may also introduce modeling error. For example, if data below detection level are censored, then the resulting matrix $\mathbf{X}$ will not be in relationship to matrices $\mathbf{G}$ and $\mathbf{F}$ as stipulated by Eq. (1). Wrong decisions about outlier status may also introduce modeling error, as pointed out by an anonymous reviewer. For example, in difficult snow conditions, highway traffic would be non-existent and hence, traffic emissions unusually low. Such high-snow samples would be valuable for determining correct rotations, especially for the traffic-emissions factor. However, such high-snow samples may appear to be outliers. If they are downweighted as outliers, then a serious modeling error is made, leading to loss of critical information.

Effects of modeling errors are difficult to estimate because the causes of such errors generally cannot be explicitly formulated. As emphasized by an anonymous reviewer, it is expected that in environmental data, modeling errors are much more significant than in AC measurements. In a follow-up paper, to be submitted soon, we apply the error estimation methods described in this work to three real data sets where modeling errors may be present.

It is noted that other definitions of modeling error have been used in literature. For example, Tauler (2001) includes rotational ambiguity with modeling error.

The relative importance of the three causes of uncertainty depends on the size of the data set being modeled. As the size of the data set increases, the significance of random errors decreases, due to the law of large numbers; the significance of rotational uncertainty also decreases because the number of zero entries in true $\mathbf{G}$ factors often increases. On the other hand, effects of non-random modeling errors are not likely to decrease with increasing size because the law of large numbers does not apply to non-random disturbances. Thus, the relative significance of modeling errors may be assumed to be highest in the largest data sets. Large data sets may, however, contain enough information so that their models may be enhanced to include the real data's problematic features, which cannot be modeled with a small data set.

1.5 Significance of different Q values when comparing alternative models of a data set

The difference of Q values obtained from alternative PMF models is often used as a criterion for rejecting models with "too high" Q values. Examples include comparison of models with different numbers of factors and rejection of competing solutions obtained from repeated random starts of PMF modeling. In error estimation, Q values are used for similar purposes: acceptable solutions must not have "too high" Q values. The obvious question is: how high is "too high"? Unfortunately, this question does not have a clear answer. If there were absolutely no modeling errors, then a change in Q (dQ) by, for example, 20 units might be considered "too high", and this limit would be independent of the number of data values in the data set. In real life, modeling errors complicate the situation because modeling errors differ in different types of measurements. Even in one type of measurement, modeling errors may depend on details of individual experimental situations. It may be assumed that the effect of modeling errors is dependent on the number of data values. It appears likely that the variation in Q values caused by modeling errors is proportional to the number of data values and hence also proportional to Q values.

It follows that no a priori percentage value may be given for assessing variations of Q values. In some data, a significant variation of Q might be 1 %. In other data, it might be 5 % or 0.5 %. It seems that an understanding of the significance of Q variations must be based on empirical evidence. It is crucial that such evidence be relevant to the case at hand. Thus, for instance, observations of Q variations in speciated aerosol measurements may not be applicable to analysis of aerosol mass spectra, water quality data, or other data sets. For example, observed variations in Q from displacement of factor elements (DISP) for reasonable models of the simulated data presented later in this paper appear not to be significant for percentages less than 0.1 %. This percentage may or may not be appropriate when analyzing actual ambient measurements, such as speciated aerosol data, aerosol mass spectra, water quality data, human exposure data, or other types. Experience applying the error methods described in this work to various types of data and numerous data sets is required before it will be known if a fixed percentage is realistic for multiple or all types of data.

1.6 Overview of uncertainty estimation methods

Many uncertainty estimation methods base their estimates on analyses of a number of *perturbed versions* of the original data set. Each perturbed data set is analyzed in a similar way as the original data. The collection of all perturbed results is then used to derive uncertainty estimates for the original unperturbed results. Using a collection of results allows analysts to review a distribution for each factor element to

evaluate the stability of solutions instead of having to rely on a single point estimate.

Pseudorandom (or random) numbers are needed for generation of perturbed versions of the data set. For this reason, the generic term "Monte Carlo methods" is sometimes used for the methods that generate perturbed versions of the data set. In particular, noise insertion (see below) might be called "Monte Carlo".

One of the classical methods for estimating uncertainty is *error propagation*, which originates in astronomy. For this method, data uncertainties (i.e., standard deviations of observations) are assumed known. Then the covariance matrix of computed results is obtained by applying the well-known error propagation formula that is based on a linear approximation around the measured values. No perturbed versions are generated in classical error propagation. *Noise insertion* is a computation-intensive variation of the classical error propagation method. In this method, a number (b_r) of perturbed versions of the original data set are generated in the following way: each perturbed version is of the same dimensions as the original data set. In each version, each original data value is perturbed by a pseudorandom artificial additive noise value whose standard deviation equals the estimated uncertainty of the data value to be perturbed. Each perturbed version is modeled similarly as the original data set, creating a collection of b_r perturbed solutions. The variances and covariances of the distribution of perturbed results are then used as the uncertainty estimates of original unperturbed results. In comparison to original error propagation, noise insertion has the advantage that no linearization is needed and nonnegativity constraints and other imposed constraints are correctly handled. Error propagation and noise insertion account for uncertainty caused by random errors in the data but not for uncertainty caused by rotational ambiguity or modeling errors.

Bootstrap analysis (BS) perturbs the original data set by *resampling*. In each perturbed or *resampled* version, some randomly chosen rows of the original matrix occur multiple times, while other rows do not occur at all. Each resampled data set is decomposed into profile and contribution matrices using PMF (Norris et al., 2008). BS has an advantage of not depending on the average level of error estimates of data values: if all data error estimates are scaled by an arbitrary coefficient r, BS results will stay the same, provided that outlier reweighting does not induce a change. Uncertainties estimated by BS may be too small or too large if significant correlation of data errors is not properly handled by techniques such as blocked resampling. BS is not specifically designed to explore rotational ambiguity, although some rotational uncertainty is captured in the analysis of the resamples. Since rotational uncertainty is limited by the number of zero values in **G** and **F**, and since the resampling for BS may omit some or all of the **G** zero values, BS may estimate a large variation in a PMF solution, especially in small data sets. Whether this large variation is appropriate depends on the reliability of the zero values. If the zero values are erroneous or are not expected to recur, then the large variation is correct. If the zero values are reliable, then the large variation is not correct. With regard to modeling errors, it is not known how well BS captures the uncertainty from this cause.

Displacement analysis (DISP) obtains uncertainty estimates for individual variables in fitted **F** by repeatedly fitting the model such that each variable in turn is perturbed (displaced) from its fitted value. Each displacement is extended until the object function Q increases by a maximum allowed change in Q ($\mathrm{d}Q^{\max}$). Each such extended displacement is interpreted as the upper or lower interval estimate of the perturbed variable. DISP captures the uncertainty caused by data errors, provided that the user-provided data uncertainties are correct for the data and they obey the assumptions of the PMF model. DISP uncertainty estimates underestimate real uncertainties if data errors are correlated, modeling errors are present, or actual data errors exceed assumed data uncertainties. On the other hand, DISP uncertainty estimates overestimate real uncertainties if actual data errors are smaller than those assumed. By design, DISP captures the uncertainty from rotational ambiguity. As with other methods, it is not known how well DISP captures uncertainty from modeling errors.

2 Previous work

2.1 Uncertainty of factor analytic results in analytical chemistry

Most prior work in assessing uncertainties of factor analytic results has been carried out with methods applied in AC (analytical chemistry). Unfortunately, most of these methods are not applicable for use in environmental source apportionment (ESA). One reason for this is that data uncertainties play a lesser role in AC because chromatogram data are usually more precise than ESA data. A second reason is that AC data are more structured than ESA data. For example, in chromatograms, if the data have been corrected to baseline, then each true component may be assumed to have a number of consecutive zero values preceding the peak. The first AC results that are applicable are due to Gemperline (1999). In this work, structural features typical of AC are not utilized. Instead, rotations of the computed **G** and **F** factors are considered under feasibility constraints, typically under nonnegativity of **G** and **F**. By using non-linear optimization algorithms, two "extreme" rotation matrices $\mathbf{T}_k$ are determined for each factor k of the model. For each factor k, those matrices minimize and maximize the fraction $\mathbf{X}_k$ of matrix $\mathbf{X}$ that is explained by factor k.

In order to discuss the method of Gemperline, Eq. (1) is written in the following form (Eq. 3):

$$\mathbf{X} = \mathbf{GF} + \mathbf{E} = \sum_{k=1}^{p} g_k f_k + \mathbf{E} = \sum_{k=1}^{p} \mathbf{X}_k + \mathbf{E}. \qquad (3)$$

Here, g_k denotes column k of $\mathbf{G}$, f_k denotes row k of $\mathbf{F}$, and $\mathbf{X}_k = g_k f_k$ is the part of data matrix $\mathbf{X}$ that is explained by factor k. The non-linear optimization problem for factor k is the following (Eq. 4):

Given $\mathbf{G}^0$ and $\mathbf{F}^0$,

define $\mathbf{G} = \mathbf{G}^0 \mathbf{T}_k$ and $\mathbf{F} = \mathbf{T}_k^{-1} \mathbf{F}^0$,

determine $\mathbf{T}_k$ such that $\mathbf{G} \geq 0$, $\mathbf{F} \geq 0$, and $\qquad$ (4)

norm$(\mathbf{X}_k) = $ norm$(g_k f_k)$ is maximized (or minimized).

The vectors g_k and f_k obtained by maximization constitute the upper interval estimates for factor k. Similarly, minimization produces lower interval estimates.

Tauler and co-workers have continued to develop the method originated by Gemperline (Tauler, 2001; Abdollahi et al., 2009; Jaumot and Tauler, 2010). The last two references contain useful literature references to other work in this field. In the 2009 paper, an illustrative example of the optimization task is presented for the two-factor case ($p = 2$). In the original Gemperline paper, *sum of elements* was used as the *norm* in Eq. (4). In later papers, other norms have also been used, such as the Frobenius norm (square root of sum of squares). It appears that slightly different results may be obtained with different norms. Also, scaling of rows and columns of matrix $\mathbf{X}$ may influence the obtained uncertainty limits.

There is a fundamental difference between the present work and the works of Gemperline and Tauler (G–T). The G–T limits for factor k represent values that might be obtained by factor k in one particular solution of the factor analytic problem. Our limits, on the other hand, represent limits of values of individual factor elements – these limits are determined individually, without regard to each other. Thus, a collection of upper-interval estimate values of factor k computed by one of our methods produces a hyperbox that may contain points that are not feasible solutions of the problem. It follows that our limits are expected to be wider than the G–T limits. Although the two methods produce different results, neither of them is wrong because they solve different mathematical problems.

2.2 Uncertainty of factor analytic results in environmental research

The earliest contribution towards understanding rotational ambiguity in factor analysis is probably by Henry (1987). In this work, the importance of rotational uncertainty is emphasized, while no methods are presented for deriving uncertainty limits. Later, Henry (1997) developed Unmix, a model for solving Eq. (1) subject to non-negativity constraints. Included with the Unmix model are estimates of uncertainty in factor profiles, estimates derived using block bootstrapping.

Hedberg et al. (2005) tested the robustness of the PMF model with a cross-validation method. They analyzed randomly reduced data sets that included 85 %, 70 %, 50 %, and 33 % of the original samples. In this way they tested the ability of the model to reconstruct the factors initially found when modeling the original data set. On average, for all factors, the relative standard deviation increased from 7 % to 25 % for the variables identifying the factors, when decreasing the data set from 85 % to 33 % of the samples.

The cross-validation method of Hedberg et al. (2005) is conceptually similar to the bootstrap method used in present work. However, they used cross-validation only for qualitative confirmation of PMF modeling, not for determining uncertainty limits.

In literature, atmospheric scientists have used the Fpeak rotational tool of program PMF2 to understand rotational uncertainty of the solution. This practice provides only a lower limit for rotational uncertainty. Specifically, varying the Fpeak parameter traces a one-dimensional path through the rotationally accessible domain. In most cases, though, the rotationally accessible domain is many-dimensional; for these cases, Fpeak will demonstrate only a lower limit for rotational uncertainty (Paatero et al., 2002). Rotational error analysis requires an upper limit, and this is not achievable by the Fpeak of program PMF-2 nor by the simpler one-parameter Fpeak of program ME-2. DISP and bootstrapping enhanced with DISP (BS-DISP) provide such upper limits.

3 Methodology

3.1 Overview of uncertainty estimation methods in ME-2 and EPA PMF

Three uncertainty estimation methods are now available in ME-2 and EPA PMF: bootstrapping (BS), dQ-controlled displacement of factor elements (DISP), and bootstrapping enhanced with DISP (BS-DISP). BS is a typical statistical method for estimating uncertainty. As implemented, BS involves resampling the input data set, fitting PMF model parameters for this resampled data set, and then using the variations among these resampled or "bootstrapped" fitted profiles to estimate the uncertainty of the initial PMF solution. BS has been available in EPA PMF v1.1 and all subsequent versions, and many publications have reported uncertainty estimates from EPA PMF.

Since BS does not explicitly include rotational ambiguity, DISP was developed. DISP intervals, however, are directly impacted by inaccuracies in data uncertainties. Thus, a method combining BS's strength with data errors and DISP's strength with rotational uncertainty was developed into the method BS-DISP. Details of the DISP and BS-DISP methods are presented below. Since BS is a standard statistical method, descriptions of its theoretical foundations are left to textbooks (e.g., Efron and Tibshirani, 1993).

The goal of DISP is to provide uncertainty estimates in such cases where data errors obey the assumptions of the PMF model (i.e., uncorrelated data errors with known data uncertainties) and there are no modeling errors. DISP

uncertainty estimates contain good estimates of rotational uncertainty as demonstrated with synthetic data sets (discussed in Sect. 4). However, DISP uncertainty estimates underestimate real uncertainties if data errors are correlated, modeling errors are present, or actual data errors exceed assumed data uncertainties. In order to obtain more reliable estimation of uncertainty due to data errors, a BS or BS-DISP analysis may additionally be performed and results compared to those from DISP. BS or BS-DISP are also necessary techniques for estimating uncertainty for species that are downweighted in the PMF analysis (i.e., species for which the adjusted data uncertainty values have intentionally been increased to reduce their influence in the minimization of Q). For such species, uncertainties estimated by DISP are known to be too large. BS-DISP is a combination of bootstrap and displacement methods in which each resampled data set is decomposed into profile and contribution matrices and then fitted elements in $\mathbf{F}$ are displaced. The collection of all results from the process of resampling, decomposing, and displacing is then summarized to derive uncertainty estimates. Intuitively, this process may be viewed as follows: each BS resample results in one solution that is randomly located within the rotationally accessible space. Then, the DISP analysis determines an approximation for the rotationally accessible space around that solution. Taken together, all the approximations of rotationally accessible spaces for randomly located solutions represent both the random uncertainty and the rotational uncertainty for the modeled solution to the complete data set. Since both the BS and DISP phases explore the rotationally accessible space, the DISP phase may be executed with weaker displacements than when only DISP is used to estimate uncertainties. As a result, BS-DISP is less sensitive to inaccuracies in data uncertainties.

In principle, BS-DISP should determine the rotational uncertainty well. However, data sets with a scarcity of *rotation-blocking zero values* in $\mathbf{G}$ factors pose the same problem for BS-DISP as with classical BS. Specifically for resamples omitting some or all of the zero values, large rotations are possible. To reduce the impact of these large rotations, the 5th percentile of minimum interval estimates and 95th percentile of maximum interval estimates may be used. There is insufficient practical experience with varied data sets to know whether using these, or any, percentiles adequately addresses this issue.

3.2 Mathematical approach in DISP

This section describes the computations for DISP, whether DISP is performed alone or as the second phase of BS-DISP. Computations are first described for well-defined cases, those for which factors do not change so much after displacement that they exchange identities ("factor swapping"). Later, computations are presented for the case complicated by factor swapping.

Superscripts are used for denoting different variants of a matrix. As an example, $(\mathbf{G}^0, \mathbf{F}^0)$ and $(\mathbf{G}^1, \mathbf{F}^1)$ may denote two different solutions of a PMF problem. Usually, $(\mathbf{G}^0, \mathbf{F}^0)$ denotes the solution obtained by PMF when no displacements are applied. Individual factor elements are then denoted by using both subscripts and superscripts; for example, g_{ik}^0 and f_{kj}^0 may denote the elements of matrices $\mathbf{G}^0$ and $\mathbf{F}^0$.

For DISP analyses, $\mathbf{F}$ factor elements are chosen, one by one, to be displaced. The chosen element is denoted by f_{kj}, so that k denotes the factor and j denotes the variable. Usually, only a subset of all $\mathbf{F}$ elements is chosen to be displaced. Details of why and how to choose are discussed later.

The DISP approach is based on the increase of the PMF sum-of-squares function Q. The function may be the basic Q defined as follows by Eq. (5):

$$Q = \min_{\mathbf{F},\mathbf{G}} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\left(x_{ij} - \sum_{k=1}^{p} g_{ik} \, f_{kj} \right) \Big/ u_{ij} \right)^2 , \qquad (5)$$

where all elements of $\mathbf{G}$ and $\mathbf{F}$ have been determined so as to achieve best possible fit (i.e., lowest possible value of sum of squares). However, the function Q may also be any enhanced form of the object function, such as a robust sum obtained by reweighting of outlying data values or a sum enhanced by penalty terms like those used for pulling chosen factor elements towards preferred values. (In special cases, some elements of $\mathbf{G}$ and/or $\mathbf{F}$ may be constrained by the user so that these elements are not variable at all. Such elements are not considered variable in the minimization.)

The notation Q^{opt} denotes the value of Q function for the PMF model that is about to be processed by DISP ("DISPed"). For pure DISP, Q^{opt} is thus the Q value obtained in the base case PMF run. For BS-DISP, Q^{opt} is the Q value obtained in PMF modeling of the current resampled data set. In both cases, Q^{opt} represents the solution of Eq. (5), i.e., a minimum with respect to all elements of factor matrices $\mathbf{G}$ and $\mathbf{F}$. The numerical values of Q^{opt} from base case and Q^{opt} from any of the resampled cases have no obvious relationship, usually they are different and either one may be larger. The notation $Q\left(f_{kj} = d\right)$ denotes the smallest sum-of-squares value obtained when constraining the indicated factor f_{kj} to a fixed feasible value d and minimizing over all other $\mathbf{G}$ and $\mathbf{F}$ factor elements. Finally, the increase of Q is denoted by Eq. (6):

$$\mathrm{d}Q\left(f_{kj} = d\right) \quad = \quad Q\left(f_{kj} = d\right) - Q^{\text{opt}}. \qquad (6)$$

The essence of DISP is to find the largest and smallest feasible values $d^{\max}$ and $d^{\min}$ such that

$$\begin{aligned} \mathrm{d}Q\left(f_{kj} = d^{\max}\right) &\leq \mathrm{d}Q^{\max} \\ \mathrm{d}Q\left(f_{kj} = d^{\min}\right) &\leq \mathrm{d}Q^{\max}, \end{aligned} \qquad (7)$$

where $\mathrm{d}Q^{\max}$ is a predetermined maximum allowable change in Q (Eq. 7). The values $d^{\max}$ and $d^{\min}$ can be determined by using any available non-linear optimization algorithm. In

this work, the ME-2 program is used under control of an enhanced script.

The obtained values $d^{\max}$ and $d^{\min}$ respectively represent upper and lower interval estimates for factor element f_{kj}. The limit value $dQ^{\max}$ is chosen by the user. In practice, the DISP approach is implemented so that estimation is performed using a set of four $dQ^{\max}$ values chosen by the user. Thus four pairs of upper and lower interval estimates are obtained for each displaced factor element. A typical set of $dQ^{\max}$ values would be {4, 8, 16, 32} for DISP and {0.5, 1, 2, 4} for BS-DISP. Larger $dQ^{\max}$ values usually produce wider uncertainty intervals which in turn usually have higher probabilities of including true unknown values. However, wider intervals may be so wide that they cannot support meaningful conclusions. For DISP, analogy with customary linear least squares models suggests that executing with $dQ^{\max} = 4$ results in interval estimates that are minima for the true uncertainty estimates, provided the user-specified data uncertainties are reasonable for the data (see the Supplement for additional discussion). If a minimum interval estimate is sufficient to support or refute a postulated hypothesis, then no additional uncertainty analysis is warranted.

The choice of $dQ^{\max}$ values will depend on assumed magnitudes of modeling errors, as discussed in Sect. 1.5. Reliable estimates of modeling errors are usually not available. It follows that $dQ^{\max}$ values cannot be deduced from statistical theory. Experimental evidence must be used.

3.3 Implementation of DISP in ME-2 and EPA PMF

Equations (5) to (7) would lead to a straightforward and reasonably efficient algorithm. However, they cannot be applied as such because of the automatic dynamic reweighting that is used for several purposes, most importantly for robust estimation, in PMF. With such reweighting, the numerical value of Q changes whenever the weights are recomputed. Such changes of Q are not directly related to changes in the fit. Hence, such changes cannot be used as a basis of uncertainty estimation.

As a substitute, the DISP approach estimates dQ values using a partial derivative (or *gradient*) of Q with respect to the displaced variable (Eq. 8):

$$dQ\left(f_{kj} = d\right) =$$
$$0.5 \sum_{v=1}^{z} (d_v - d_{v-1}) \left(\left.\frac{\partial Q}{\partial f_{kj}}\right|_{f_{kj}=d_v} + \left.\frac{\partial Q}{\partial f_{kj}}\right|_{f_{kj}=d_{v-1}} \right)$$
$$= 0.5 \sum_{v=1}^{z} (d_v - d_{v-1}) \left(\mathrm{grad}\left(f_{kj} = d_v\right) + \mathrm{grad}\left(f_{kj} = d_{v-1}\right) \right)$$
$$\text{where } d_0 = f_{kj}^0 \quad \text{and} \quad d_z = d. \tag{8}$$

This definition is based on a sequence of z displaced values d_v, generated automatically by the algorithm. The model is fitted using each displaced value in turn, and the corresponding gradient values are saved. The proxy dQ value is

obtained using displacement step lengths and gradients at each displaced point. This method is approximate and becomes more accurate if a larger number of intermediate displacements are used for reaching the final displacement d. The quality of approximation has been observed in cases where no dynamic reweighting is present so that actual Q values may be used for computing non-approximate dQ values. The sequences created automatically by the current implementation of DISP appear to be a satisfactory compromise between computational efficiency and accuracy of approximation. Determination of the sequence of displaced values d_v is based on various heuristic principles designed to balance between too-long displacements (indicated by sudden increase of gradient and dQ or by reversal of gradient) and too-short – and hence inefficient – displacements. If a displacement is found to be too long, it is rejected and a shorter displacement is attempted instead.

The sequence of displaced values does not usually hit the desired value for dQ, namely $dQ = dQ^{\max}$, as required by the definition of the uncertainty interval in Eq. (7). As shown in Eq. (9), the sequence generally ends so that

$$dQ\left(f_{kj} = d_{z-1}\right) < dQ^{\max} \tag{9}$$
$$dQ\left(f_{kj} = d_z\right) > dQ^{\max}. \tag{10}$$

In order to obtain the desired critical value ($d^{\max}$ or $d^{\min}$), an interpolation is performed. It is assumed that the gradient changes linearly in the interval ($d_{z-1} < d < d_z$). With this assumption, the value $d^{\max}$ for displacing up may be computed (Eq. 10) so that

$$dQ\left(f_{kj} = d^{\max}\right) \approx dQ^{\max}. \tag{11}$$

Similarly, when displacing down, the value $d^{\min}$ is obtained so that (Eq. 11)

$$dQ\left(f_{kj} = d^{\min}\right) \approx dQ^{\max}. \tag{12}$$

These interpolations are computed separately for each of the four $dQ^{\max}$ values. Using the interpolated displacement values and factor matrices computed at each displacement, it is possible to also interpolate the values of factor matrices $\mathbf{G}$ and $\mathbf{F}$ so that the interpolated values correspond to the solution of Eq. (7). In current implementation, only elements of factor matrix $\mathbf{F}$ are interpolated, however.

It is to be noted that displacements do not proceed past lower or upper constraints for each displaced factor element. Whenever the constraint is violated, the last displacement is truncated so that it exactly corresponds to the constraint for the variable. If the dQ at the constraint value does not exceed the chosen $dQ^{\max}$, then the constraint value is used as the interval estimate of the variable. For this reason, lower interval estimates of $\mathbf{F}$ factor elements may appear as exactly zero.

3.4 Active and passive estimation with DISP and BS-DISP

The intervals obtained by displacing a factor element f_{kj} include both rotational ambiguity and uncertainty due to assumed data uncertainties. In order to speed up computation of BS-DISP, it is preferable to displace a small subset of all **F** factor elements, the *active elements* of **F**. Usually, one would displace those variables important for factor identification or variables key to a particular question.

It is possible to estimate uncertainty intervals for those factor elements that are not displaced. Intervals for such *passive* factor elements are obtained as a by-product during displacements of active elements. As described, all elements of **F** are obtained for each (interpolated) displacement that solves Eq. (7). The DISP algorithm finds largest and smallest values $f_{kj}^{\max}$ and $f_{kj}^{\min}$ of each passive element f_{kj} among *all* interpolated **F** matrices that occur while displacing all active elements. These extreme values constitute *passive interval estimates* for the passive (non-displaced) **F** factor elements.

If a sufficient number of **F** elements are displaced actively, then passive interval estimates reflect rotational ambiguity well for the remaining passive elements. In contrast, passive interval estimates do not contain uncertainty due to assumed data uncertainties of the passive factor elements. In BS-DISP, assumed data uncertainties play a minor role because uncertainty caused by data noise is mainly estimated by resampling. Thus passive estimation is useful in BS-DISP, provided that the number of active elements is large enough that rotationally accessible space is exhaustively visited. In DISP, however, passive interval estimates are less useful because they ignore data uncertainties of passive factor elements. For this reason, in pure DISP computations one would prefer to displace all factor elements.

Downweighted variables create a special problem in DISP computations. If such variables are displaced, their obtained active interval estimates will be much too long; because the assumed data uncertainties are much too large, using the default $dQ^{\max}$ limits will result in very large residuals for the downweighted variables. The best compromise seems to be that downweighted variables are never chosen for active estimation in DISP or in BS-DISP. If not active, downweighted variables will obtain passive interval estimates, intervals that may be too short from DISP but satisfactory from BS-DISP.

3.5 Factor swaps in DISP from not-well-defined solutions

Starting from one good solution, it may be 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." A solution with swapped factors represents the same physical model as the original solution. However, the presence of factor swaps means that all intermediate solutions must be considered as alternative solutions. In such a case, the modeling supports a many-dimensional infinite population of solutions where it is not possible to single out one of these solutions as "the solution," hence the terminology "not-well-defined (NWD) solution." Often, factor swaps occur only within a subset of all factors. Then the modeling may provide useful information about those factors that do not participate in swaps. DISP and BS-DISP analyses provide diagnostic output to aid in the identification of factors involved with swapping.

The significance of factor swaps from NWD solutions came as a surprise. There is little practical knowledge about these situations, and therefore conclusions in this section are of preliminary nature.

To detect factor swaps, consider two solutions: the original solution ($\mathbf{G}^0$, $\mathbf{F}^0$) and the transformed solution ($\mathbf{G}^1$, $\mathbf{F}^1$). Testing for swaps may be based on **G** matrices or on **F** matrices. In the case of complete swaps, testing using either matrix produces identical conclusions. In borderline cases where factors change significantly but a complete swap does not occur, the **G** and **F** tests are not fully equivalent. Equations (12) through (15) are given for testing **G** matrices. **F** tests are obtained by replacing **G** with **F** in the equations. Two methods are available for detecting factor swaps: one based on cross correlations and the other based on regression.

For cross correlations, "uncentered" correlation coefficient r between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined by

$$r = \mathrm{corr}(\boldsymbol{u}, \boldsymbol{v}) = \frac{\boldsymbol{u}'\boldsymbol{v}}{\sqrt{\boldsymbol{u}'\boldsymbol{u}\ \boldsymbol{v}'\boldsymbol{v}}}. \tag{13}$$

This differs from Pearson correlation, which is centered and is commonly used both in social and biological sciences and also in chemistry and engineering.

$$\text{Define centered variables}: \quad \tilde{\boldsymbol{u}} = \boldsymbol{u} - \bar{\boldsymbol{u}} \quad \tilde{\boldsymbol{v}} = \boldsymbol{v} - \bar{\boldsymbol{v}}$$
$$r = \mathrm{corr}_{\text{Pearson}}(\boldsymbol{u}, \boldsymbol{v}) = \frac{\tilde{\boldsymbol{u}}'\tilde{\boldsymbol{v}}}{\sqrt{\tilde{\boldsymbol{u}}'\tilde{\boldsymbol{u}}\ \tilde{\boldsymbol{v}}'\tilde{\boldsymbol{v}}}}. \tag{14}$$

Because Pearson correlations can be meaningless if some factors are nearly constant, uncentered correlations are used to detect factor swaps. Specifically, a matrix of correlation coefficients is computed, so that each matrix element is the correlation coefficient between one column of $\mathbf{G}^0$ and one column of $\mathbf{G}^1$. A factor swap is seen in this correlation matrix so that two or more diagonal entries are small, while corresponding off-diagonal entries are ≈ 1.

In the regression approach, a transformation matrix (or *regression matrix*) **T** is computed for approximating $\mathbf{G}^1$ by a transformed $\mathbf{G}^0$. The approximation is defined by

$$\mathbf{G}^1 = \mathbf{G}^0\mathbf{T} + \mathbf{E} \approx \mathbf{G}^0\mathbf{T}, \tag{15}$$

where matrix **T** is obtained from

$$\mathbf{T} = \left(\mathbf{G}^{0\prime}\mathbf{G}^0\right)^{-1}\mathbf{G}^{0\prime}\mathbf{G}^1. \tag{16}$$

It is assumed that $\mathbf{G}^0$ is of full column rank. If there are no factor swaps, $\mathbf{T}$ is approximately diagonal, so that off-diagonal elements are small positive and negative values. With a factor swap, the rows of $\mathbf{T}$ become permuted so that at least two diagonal elements change positions with smaller off-diagonal elements.

3.6 Decrease in Q with DISP

Occasionally displacements cause a significant decrease of Q, typically by tens or even hundreds of units. If such a decrease occurs in DISP analysis or when analyzing the complete (not resampled) data in BS-DISP, it means that the base case solution was in fact not a global minimum, although it was assumed to be such. This is a fatal error and invalidates the DISP analysis. It is necessary to go back to solving the original PMF model again, perhaps using many more random starts, to find the global minimum. Then the DISP analysis may be continued.

Decrease of Q may also occur when performing displacements in the analysis of BS-DISP resamples. Such a decrease indicates that resampling created a new minimum, different from the original base case solution. In one case, the initial not-displaced fit of this BS resample did not succeed in finding the new global minimum, while the displacement "nudged" the solution towards the global minimum. In such a case, it is best to reject the resample because no meaningful error limits can be obtained. The overall BS-DISP analysis remains valid, even if a few resamples get rejected, though currently there is no way to quantify the number of rejections that will yield meaningful results.

3.7 Development and modeling of synthetic data sets

Simulated data were designed to demonstrate the three uncertainty estimation methods. The data were generated using partial results from a PMF application to $PM_{2.5}$ speciated data collected in Phoenix (Eberly and Reff, 2007). Fitted g_k and f_k for four of seven factors from the previous PMF analysis were selected to represent the true matrices $\mathbf{G}$ and $\mathbf{F}$. Four factors – representing copper smelting, coal combustion, aged sea salt, and soil – were used to simplify the simulation and modeling. Some factors are small contributors on average and others are large, a desired characteristic for the simulated data. Specifically, average contributions are 49 % for coal combustion, 2 % for aged sea salt, 9 % for copper smelting, and 40 % for soil. Sixteen species are included: $PM_{2.5}$, elemental carbon (EC), organic carbon (OC), Si, S, Cl, K, Ca, Ti, Mn, Fe, Ni, Cu, Zn, Se, and Pb. Profiles for the four factors are included in the Supplemental, Table S1.

To generate the simulated data, $\mathbf{G}$ and $\mathbf{F}$ based on the four previously modeled factors were multiplied to form $\mathbf{C}$, per notation described in Eq. (1). Error-containing values $\mathbf{X}$ were obtained from pseudorandom distributions of log-normal variates with mean $\mathbf{C}$ and standard deviation $\mathbf{S}$, where

$\mathbf{S}$ was specified by two equations to evaluate impacts of standard deviations on uncertainty estimates. Case 1 assumed small errors such that $s_{ij} = 0.05\,c_{ij}$. Case 2 assumed realistic errors such that $s_{ij} = z_j c_{ij}$, where z_j varied from a small value of 0.05 for well-measured species to a value of 1.2 for species with large measurement errors. Specifically, values for z_j are Ca 0.2; Cl 0.5; Cu 0.2; EC 0.12; Fe 0.1; K 0.1; Mn 0.15; Ni 1.25; OC 0.1; Pb 0.5; $PM_{2.5}$ 0.08; S 0.05; Se 0.4; Si 0.35; Ti 0.9; and Zn 0.13. For this work, a simplifying assumption was made that detection limits are approximately zero.

The object function Q in Eq. (2) requires user-provided data uncertainties u_{ij}. These were set equal to the data uncertainties used in deriving the simulated values, namely $u_{ij} = s_{ij}$. In reality, the user rarely knows the exact amount of uncertainty in the actual data. To simulate this discrepancy, one additional case was modeled. For Case 3, the data were generated using the small errors of $s_{ij} = 0.05\,c_{ij}$, but the uncertainties given for Equation 2 were derived by $u_{ij} = 0.001 + 0.03\,x_{ij}$. Case 3 contained another intentional inconsistency: a total of 5 factors were fitted, one more than were used to generate the data.

Data sets comprised either 50 or 261 samples. Modeling was done through direct interaction with ME-2 via PMF_bs_6f1.ini and me2gfP4_1345c4.exe, rather than EPA PMF. The lower limit allowed for fitted $\mathbf{G}$ factor elements was -0.10, error model -12 was used, and the block size for bootstrapping was 1. For each data set analyzed, 15 base case runs were executed to determine a solution presumably associated with the global minimum for Q.

3.8 Computational workload in different methods

Rough estimates of computational workload (and hence, of computing times) are given in this section. The computing load ($=$ time) of one PMF modeling, using random starting values, is denoted by one time unit (t). Thus a typical initial modeling run will amount to $20\,t$. Denote by a_d and a_b the numbers of actively displaced $\mathbf{F}$ elements in DISP and BS-DISP, respectively. Assume for this estimation a large data set, having $m = 30$ and $p = 10$. In this example, $a_d = mp = 300$ if all $\mathbf{F}$ elements are selected to be active.

The number of bootstrap resamples (same for BS-only and BS-DISP) is denoted by b_r. Assume $b_r = 50$. A BS-only run consists usually of b_r instances of PMF modeling, each about one unit. Thus BS-only amounts to b_r units. It is seen that a BS-only run, with $50\,t$, is not much slower than the initial run with $20\,t$.

In a DISP process acting on one active $\mathbf{F}$ element, a varying number of PMF-like models are fitted starting from non-random initial values. In easy cases, with well-defined solutions and no rotational ambiguity, the total load from one active $\mathbf{F}$ element may be just a few units. With lots of rotational ambiguity and maybe NWD solutions, the load may be tens of units. As an example, it is assumed in this estimation that

an average DISP process for one active $\mathbf{F}$ element amounts to $10\,t$. With the assumed dimensions, the load of one complete DISP run, with all $\mathbf{F}$ elements active, would amount to $10\,mp = 3000\,t$. This is about 150 times longer than the initial run. Note that the actual results may vary by a large amount, depending on the rotational ambiguity and NWD character of the model.

In a BS-DISP run, the number of DISP processes is $b_r a_b$. If all $\mathbf{F}$ are active, this amounts to $b_r a_b = b_r\,mp = 15\,000$ DISP processes and an estimated workload of $10\,b_r a_b = 10\,b_r\,mp = 150\,000\,t$. Again, the actual workload may vary by a large amount, at least by a factor of 3. In the other extreme, it might be possible to run with only $a_b = 10$, i.e., only one $\mathbf{F}$ element active in each $\mathbf{F}$ row. Then one would have 500 DISP processes and a workload of $5000\,t$. This is comparable to a complete DISP run with workload of $3000\,t$.

It is seen that computing times may easily grow impossibly long. Hence, even with DISP, it is useful to omit less important species from active status. Also, one might keep many elements of a weak factor inactive even if all elements of a strong factor are made active. With BS-DISP, it is necessary to only have a small number of active elements. For this reason, BS-DISP will in most cases need support from a separate DISP run (with an increased number of active elements) in order to get realistic estimates for those elements that cannot be active in BS-DISP.

The authors have identified a method to improve the convergence rate of ME calculations which will help with the computational time in future. It is hoped that an order-of-magnitude gain might be obtained. If this succeeds, it will make more complete BS-DISP runs possible.

3.9   Estimation of errors of factor matrix G

This work only derives uncertainty estimates of $\mathbf{F}$ factor elements. These uncertainty estimates apply also to estimates of average pollution contribution from each factor because all modeling is performed under the constraint that average $\mathbf{G}$ values must be in unity for each factor. However, it would also be important to obtain uncertainties of specific $\mathbf{G}$ values or functions of $\mathbf{G}$ values such as the largest 10 %, weekday/weekend ratios, or seasonal contributions to aid in development of air quality management strategies. Also, individual $\mathbf{G}$ matrix errors would be very useful for future model development since hybrid approaches that combine meteorology and source contributions need to account for the uncertainties. Unfortunately, estimation of $\mathbf{G}$ errors could not be included in this work plan for several reasons. First, $\mathbf{F}$ uncertainties were of higher priority because they are needed in order that factors may be more reliably identified with sources by showing which components were fitted confidently and which components were too uncertain to aid with identification. Second, it has not been possible to devise a straightforward method for estimating $\mathbf{G}$ uncertainties. This is so because the dimensional situation with $\mathbf{G}$ and $\mathbf{F}$ is not symmetric and displacement of $\mathbf{G}$ values may not be a reliable method. Rotational uncertainty of $\mathbf{G}$ values may perhaps be obtained as an extension of the current work in future. However, so far it is unclear how to combine rotational $\mathbf{G}$ uncertainty with the $\mathbf{G}$ uncertainty caused by random errors of individual data values.

4   Results and discussion

DISP, BS, and BS-DISP were run for each of the three synthetic cases. For Cases 1 and 2, the correct number of factors, four, was fitted. For Case 3, five factors were fitted, one more than needed. Modeling resulted in fitted factors for Cases 1 and 2 of soil, salt, copper, and coal. For Case 3, with the $n = 50$ data set, the factors are soil, salt, copper, coal, and an extra factor composed of some EC, OC, Ni, S, and $PM_{2.5}$; in the $n = 261$ data set the factors are soil, salt, copper, and coal split into two pieces. No species were downweighted, so all species were active in DISP. DISP results were generated with $dQ^{\max}$ values of 4, 8, 15, 25, the values used in EPA PMF. For BS, factors were assigned to base case factors based on uncentered correlations of contributions (i.e., time series). A correlation of 0.80 or larger was required for the assignment to be valid. Three hundred bootstraps were used for this demonstration. For BS-DISP, only those species key in factor identification were active in the DISP phase: Ca, Cl, Cu, Fe, $PM_{2.5}$, S, and Ti. BS-DISP was executed using 50 BS runs and $dQ^{\max}$ values of 0.5, 1, 2 and 4, the values used in EPA PMF.

4.1   Analysis of synthetic data sets – diagnostics

Table 1 summarizes the diagnostics reported by ME-2 for data sets with 50 or 261 samples. For brevity, detailed discussion of these diagnostics is confined to the data sets with 50 samples. Diagnostic results were similar for the 261-sample data set. To put decreases of Q into perspective, robust Q values for the data set with 50 samples were 500–600 for Cases 1 and 2 and 340 for Case 3. For the data set with 261 samples, robust Q values were approximately 3000 for Cases 1 and 2 and 1800 for Case 3.

Decrease in Q for DISP. Small decreases in Q (less than 0.2) were reported for Cases 1 and 2, indicating that these solutions were global minima. A large value (greater than 2.5) was reported for Case 3, providing the first indication that there is something problematic with the modeling.

Swapped Factors for DISP. For Cases 1 and 2, no factors swapped for any values of $dQ^{\max}$, indicating that these were all well-defined PMF solutions. For Case 3, the copper factor was not involved in swaps for the smallest $dQ^{\max}$ value, so DISP interval estimates for this factor were reliable and realistic for the smallest displacement. All other factors of Case 3 were involved with swaps for each $dQ^{\max}$ value, and therefore DISP cannot provide error estimates for these factors.

Table 1. Summary of error estimation diagnostics by data set and case study.

Data set	Case study	DISP		BS		BS-DISP	
		Decrease in Q	Factors involved in swaps	BS factor assignment to base case factors	# Rejected BS resamples	Decrease in Q	Factors involved in swaps
$N = 50$	Case 1: small errors	0.1	No swaps.	Copper: 99 % Others: 100 %	0	0.2	No swaps.
	Case 2: realistic errors	0.0	No swaps.	Salt: 67 % Others: 99 %–100 %	12 out of 50, some due to swapping of factors, some due to decreases in Q.	21	No swaps for dQ^{max} = 0.5, minimal swaps (1–4) for each factor for larger dQ^{max}.
	Case 3: small errors; too many factors	2.6	Factors 1–4 swap 1–8 times, numerous swaps for Factor 5 (extra).	Extra factor 5: 20 % Others: 100 %	0 out of 10*	0.5	All factors involved in swaps.
$N = 261$	Case 1: small errors	0.2	No swaps.	All: 100 %	0	0.4	No swaps.
	Case 2: realistic errors	0.0	No swaps.	Salt: 92 % Others: 100 %	2 out of 50 due to decreases in Q.	75	No swaps.
	Case 3: small errors; too many factors	28	No swaps for soil, copper, salt, many swaps for coal and extra factor.	Coal and extra Factor 5: 80 % Others: 100 %	0 out of 10*	0.1	All factors involved in numerous swaps, usually hundreds of swaps.

* Used 10 bootstrap resamples because of the large number of factor swaps.

The extra factor (Factor 5) was involved in numerous swaps compared to the other factors, confirming that one too many factors was modeled. When only four factors, the true number, were modeled for Case 3, the DISP diagnostics indicated no factor swaps.

Assigning BS Factors to Base Case Factors. All bootstrap factors were assigned to base case factors in 99–100 % of every bootstrap resample for Case 1. For Case 2, the salt factor was not consistently identified in 33 % of the resamples. This lack of reproducibility was likely caused by two compounding issues. One was that the factor was composed of just one species, Cl, with a small amount of EC. The other was that the factor's contributions were defined by a few large values that could be excluded in BS resamples. For such resamples, this factor could be incorporated into other factors. For Case 3, all factors were reproduced in every bootstrap, except that Factor 5 (the extra factor that is comprised of small pieces of several species) was rarely found, confirming that one too many factors was modeled.

Decrease in Q and swapped factors in BS-DISP. In Case 1, no swaps occurred in the initial refitting of the full data set and no BS resamples were rejected because of swaps or large decreases in Q. This indicates that Case 1 was a well-defined PMF solution. For Case 2, diagnostics showed that 16 % of the resamples exhibited large decreases in Q and 8 % contained swapped factors. The large decrease in Q compared to Case 1 is likely due to the larger data uncertainties used in Case 2. This indicates that Case 2 was not as well defined as Case 1, but there were few enough rejected resamples that error estimates summarized for the accepted resamples were likely reliable and robust. For Case 3, all factors

were involved in numerous swaps, indicating serious problems with the modeling and warning that interval estimates should not be interpreted.

4.2 Analysis of synthetic data sets – interval estimate examples

Output from DISP, BS, and BS-DISP includes interval estimates for each element for each factor and diagnostics for evaluating the trustworthiness of the interval estimates. As discussed in Sect. 3.2, estimates of intervals are calculated as follows: for DISP, endpoints of the uncertainty interval for a specific **F** factor element are the minimum value for that factor element observed in all displacements and the maximum value for that factor element observed in all displacements. For BS, the endpoints of the uncertainty interval for a factor element are the 5th and 95th percentile values for that factor element from all bootstrap resamples. For BS-DISP, each bootstrap resample is displaced and minimum and maximum values are calculated for each factor element as described for DISP. Then percentiles are taken across the resamples, the 5th percentile of the minima and the 95th percentile of the maxima, to create the final interval estimate.

Many intervals were estimated: one for each factor element for each error method for each data set studied. Table 2 contains upper and lower interval estimates for all error methods for a selected case, Case 2, for two selected species: PM$_{2.5}$, a species of interest across all factors (Table 2a), and Cu (Table 2b), a typical example of a key species for identifying one of the factors. For the sake of brevity, only Case 2 is presented, since the data uncertainties for this case are more typical for ambient measurements.

Table 2a. Lower and upper interval estimates of PM$_{2.5}$ (µg m^{-3}) by factor for Case 2 (realistic errors) for data sets with 50 or 261 samples.

	Salt factor True PM$_{2.5}$ = 0.10	Copper factor True PM$_{2.5}$ = 0.42	Soil factor True PM$_{2.5}$ = 1.82	Coal factor True PM$_{2.5}$ = 2.24
	Data set with 50 samples			
DISP	(0.00, 0.69)	(0.12, 0.62)	(1.23, 1.97)	(2.08, 2.49)
BS	(0.06, 0.75)	(0.15, 0.69)	(1.16, 1.90)	(1.52, 2.38)
BS-DISP	(0.00, 0.85)	(0.12, 0.93)	(1.17, 2.48)	(1.54, 2.64)
	Data set with 261 samples			
DISP	(0.06, 0.18)	(0.33, 0.59)	(1.59, 1.92)	(2.08, 2.36)
BS	(0.10, 0.28)	(0.36, 0.54)	(1.56, 1.82)	(1.98, 2.27)
BS-DISP	(0.07, 0.32)	(0.33, 0.63)	(1.52, 1.99)	(2.00, 2.37)

Table 2b. Lower and upper interval estimates of Cu (µg m^{-3}) by factor for Case 2 (realistic errors) for data sets with 50 or 261 samples.

	Salt factor True Cu = 0.0000	Copper factor True Cu = 0.0025	Soil factor True Cu = 0.0007	Coal factor True Cu = 0.0001
	Data set with 50 samples			
DISP	(0.0000, 0.0001)	(0.0017, 0.0022)	(0.0003, 0.0009)	(0.0000, 0.0003)
BS	(0.0000, 0.0005)	(0.0015, 0.0021)	(0.0003, 0.0007)	(0.0000, 0.0003)
BS-DISP	(0.0000, 0.0011)	(0.0012, 0.0023)	(0.0001, 0.0008)	(0.0000, 0.0004)
	Data set with 261 samples			
DISP	(0.0000, 0.0001)	(0.0021, 0.0025)	(0.0005, 0.0008)	(0.0001, 0.0003)
BS	(0.0000, 0.0001)	(0.0022, 0.0024)	(0.0005, 0.0007)	(0.0001, 0.0002)
BS-DISP	(0.0000, 0.0001)	(0.0022, 0.0025)	(0.0005, 0.0008)	(0.0001, 0.0003)

For PM$_{2.5}$ for the data set with 50 samples, the salt factor's overall contribution is uncertain, with possible values ranging up to 7 times the true amount. Comparatively, the soil and coal factors' PM$_{2.5}$ mass estimates are more robust, with estimates ranging from about half of the true amount to just 10 % more for DISP and BS and 20–30 % more for BS-DISP. The copper factor is in between, with PM$_{2.5}$ estimates ranging from a third of the true value to 1.5 to 2 times the true amount. The size of these intervals may seem large, but this data set contains just 50 samples. For comparison, intervals for the data set with 261 samples are included in the lower halves of Tables 2a and 2b. The markedly shorter intervals for the larger data set show the power of having more data. Intervals estimated from the smaller data set support the idea presented in Sect. 1.6 about the sensitivity of BS to zero values in **G**, as evidenced by the long BS (and therefore BS-DISP) intervals compared to DISP. This difference nearly disappears for the larger data set, supporting the idea presented in Sect. 1.4 that rotational uncertainty plays a lesser role in larger data sets.

For Cu, again the intervals for the larger data set are markedly shorter than those for the smaller data set. Another note is that many of the intervals do not contain the true amount of Cu for the copper factor. That is, these error methods do not always produce intervals that contain the true value.

4.3 Analysis of synthetic data sets – summary of comparisons

As seen in Table 2, different error methods can produce the shortest interval depending on the data set. Sometimes an error method's interval includes the true value and sometimes it does not. Given the large number of intervals estimated, it is challenging to determine which error method is consistently producing shorter intervals or intervals that include true values. To aid in the comparison of one error method to another, summary statistics that aggregated over all factor elements were calculated (see Table 3). Three summaries were calculated. One was *percent coverage*, the number of intervals containing true **F** factor elements divided by total number of **F** factor elements. The second and third were *median* and *average ratios* for intervals. These were calculated as follows: length and midpoint of each interval for each **F** factor element were computed. Then length was divided by midpoint to create a unitless quantity that can be compared across factor elements of differing magnitude. Median and average ratios were calculated across all **F** factor elements.

Table 3a. Summaries of **F** interval estimates for data sets with 50 observations.

Method for estimating intervals (number of bootstraps, dQ^{max})	First row: summary over all **F** factors Second row: summary over all **F** factors excluding salt (% coverage, median and avg ratios of length to middle of interval)	
Case 1. Small errors	Subset 1	Subset 2
DISP (n/a, 4)	98 %, 0.82, 1.04 98 %, 0.51, 0.86	100 %, 0.74, 1.00 100 %, 0.54, 0.85
BS (300, n/a)	77 %, 0.91, 1.05 73 %, 0.57, 0.88	73 %, 0.93, 1.00 71 %, 0.62, 0.87
BS–DISP (50, 0.5)	100 %, 1.28, 1.25 100 %, 1.01, 1.08	100 %, 1.25, 1.19 100 %, 0.93, 1.05
Case 2. Realistic errors		
DISP (n/a, 4)	95 %, 1.49, 1.31 96 %, 1.03, 1.15	100 %, 1.47, 1.32 100 %, 0.93, 1.16
BS (300, n/a)	78 %, 1.53, 1.36 81 %, 1.16, 1.23	81 %, 1.39, 1.24 79 %, 0.82, 1.06
BS–DISP (50, 0.5)	97 %, 2.00, 1.54 96 %, 1.59, 1.39	98 %, 1.74, 1.36 98 %, 0.98, 1.21

Table 3b. Summaries of **F** interval estimates for data sets with 261 observations.

Method for estimating intervals (number of bootstraps, dQ^{max})	First row: summary over all **F** factors Second row: summary over all **F** factors excluding salt (% coverage, median and avg ratios of length to middle of interval)	
Case 1. Small errors	Subset 1	Subset 2
DISP (n/a, 4)	95 %, 0.45, 0.77 94 %, 0.32, 0.69	94 %, 0.45, 0.84 92 %, 0.36, 0.71
BS (300, n/a)	75 %, 0.79, 0.99 69 %, 0.40, 0.80	56 %, 0.39, 0.78 60 %, 0.27, 0.66
BS–DISP (50, 0.5)	98 %, 0.70, 0.96 98 %, 0.53, 0.81	97 %, 0.57, 0.92 96 %, 0.45, 0.76
Case 2. Realistic errors		
DISP (n/a, 4)	92 %, 0.77, 1.01 90 %, 0.49, 0.86	89 %, 0.80, 1.05 85 %, 0.47, 0.86
BS (300, n/a)	75 %, 0.44, 0.82 71 %, 0.31, 0.69	59 %, 0.58, 0.86 63 %, 0.42, 0.80
BS–DISP (50, 0.5)	91 %, 0.85, 1.05 88 %, 0.60, 0.90	83 %, 0.82, 1.05 77 %, 0.53, 0.86

To test repeatability of results, two replicates of each data set were generated and modeled. The original data set contained 783 observations. For the 261-day replicates, every third sample was retained, starting with the first sample for Subset 1 and the second sample for Subset 2. For the 50-day replicates, every 15th sample was retained, starting with the first sample for Subset 1 and the second sample for Subset 2. DISP results are presented only if no swaps occurred and if Q decreased minimally (less than 0.5). BS was run with 300 resamples and results are presented only for assignments of BS factors to base case factors with uncentered correlations of 0.80 or higher and for which only one bootstrap factor is allowed to be assigned to each base case factor. BS-DISP was run for 50 of the BS resamples. Summaries are formulated only of such BS resamples in which no swaps occurred. Interval estimates were summarized over all factors (upper row

in table cells) and also over all factors excluding the sea salt factor (lower row) since modeling of bootstrapped resamples did not always fit a factor highly correlated with the sea salt factor (as described in Table 1). Case 3, the case in which modeling error was introduced, was excluded from this summary analysis since diagnostics for this case indicated problems, as discussed in Sect. 4.1. Results are presented in Tables 3a and 3b.

These summaries show that percent coverage is generally high, greater than 90 %, except for BS and for BS-DISP with the larger data set for Case 2. Also, DISP generally provides the shortest intervals, except for Case 2 with the larger data set, where BS provides the shortest intervals.

Results from Subset 1 and Subset 2 are similar for DISP and BS-DISP. Unexpectedly, BS results vary by subset. The reason is unclear at this time, but it may have to do with the number of zeros in $\mathbf{G}$ for the two subsets. For the data set with 50 samples, Subset 1 has 3, 3, 7, and 1 zeros and Subset 2 has 1, 6, 12, and 1 zeros for coal, salt, copper, and soil factors, respectively. For the data set with 261 samples, Subset 1 has 6, 17, 27, and 8 zeros and Subset 2 has 5, 26, 44, and 7 zeros for coal, salt, copper, and soil factors, respectively. It does not take many zeros to reduce rotational uncertainty; thus, the larger number of zeros for Subset 2 of the smaller data set could explain the shorter intervals. The cause for lower percentage coverages for BS for Subset 2 is unknown.

As expected and seen with the examples presented in the previous section, it is noted that intervals are shorter for the larger data set. This is true for all methods and both case studies. What is not expected is that percentage coverage is lower for the larger data set. The cause is unclear; however, a proposed explanation is that the likelihood of excessively long intervals is higher for smaller data sets because there are fewer zeros in $\mathbf{G}$. These excessively long intervals will in turn result in unnaturally high coverage.

The conclusion from the analysis of these synthetic data sets is that DISP consistently provides intervals that have high coverage (> 90 %) and that are shorter than those provided by BS or BS-DISP. BS-DISP sometimes provides intervals with higher coverage than DISP, but these intervals are generally longer. The performance of error estimation techniques will depend on the details of each individual data set. Here, the differences seen for supposedly similar case studies 1 and 2 illustrate the variability found between data sets.

Although patterns in relative merits of the three uncertainty estimation techniques are developing, applying these inferences to all PMF analyses is premature. Variation in characteristics of data sets (e.g., number of samples, number of zeroes in $\mathbf{G}$) and modeling errors (e.g., inappropriate number of factors, discrepancies between s_{ij} and u_{ij}, handling of values below method detection limit) may lead to different relative merits. In order to achieve the best possible uncertainty estimations, the evaluation approach of this paper should preferably be repeated whenever PMF error

estimation is applied to new kinds of data sets: simulations with realistic true data patterns should be performed and merits of uncertainty estimates should be evaluated. A forthcoming manuscript (Brown et al., 2014) will present case studies of ambient data and interpretation of results from the three error estimation techniques.

5 Reporting recommendations for PMF analyses

Reff et al. (2007) performed a literature review of publications of PMF applications. The purpose of the review was to document the numerous decisions that users of PMF must make to perform such applications and to encourage that future publications of PMF applications include enough details for readers to evaluate, reproduce, or compare results between different studies. In a continuing effort to help make the reporting of results from EPA PMF and ME-2 more systematic among researchers, we have summarized recommendations on what to report while documenting uncertainty estimates from PMF analyses. This is not an exhaustive list, and every data set may require that additional information be reported. To increase the understanding of the behavior of these uncertainty estimates with different types of data, it is recommended that all three techniques be applied and specific details about and estimated intervals from each method be reported. For cases where this is not possible or reasonable, it is recommended that such reasoning be included in the publication.

BS. Report the number of resamples analyzed and the size of percentiles of the obtained distribution of results chosen for error limits, e.g., 5th and 95th percentiles. Also report the percentage of BS factors assigned to each base case factor and the number of BS factors not assigned to any base case factor.

DISP. Report species not displaced such as those downweighted, the absolute and relative decrease in Q, and the number of factor swaps. If factor swaps occur for the smallest dQ^{max}, 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.

BS-DISP. As with BS and DISP, report the number of BS resamples analyzed, the size of percentiles chosen for error limits, the species actively displaced, the decrease in Q, and the number of factor swaps.

6 Conclusions

Exercises presented with synthetic data suggest that error intervals estimated by DISP, BS and BS-DISP capture with high probability profile values that truly underlie the modeled observations. Numerous simulations were performed in

addition to those reported in this work. All indicate that if data uncertainties are known and there are no modeling errors, then the DISP method consistently produces good coverage of true values using the shortest possible uncertainty intervals. In the more difficult cases where data uncertainties are not well known, the bootstrap-based methods BS and BS-DISP seem to work satisfactorily, provided that there are no modeling errors. A solution's stability can also be evaluated via the fraction of times each factor is mapped in BS and if any swaps occur in DISP. These results provide critical information on whether a solution should be interpreted.

The uncertainty estimation with DISP depends on the user's defined maximum allowed change in Q ($\mathrm{d}Q^{\max}$). For simulated data, this dependence was illustrated in this work. For real data, mathematical derivation is impossible because of the presence of modeling errors. Practical experience is needed in order to understand the dependence on $\mathrm{d}Q^{\max}$. Such understanding might be attempted by partitioning a real data set in various ways and comparing the partition-partition variation of profiles against their DISP uncertainty estimates. In a companion paper, to be submitted soon, several real-data analyses will be reported. It should be noted that the dependence of uncertainty intervals on $\mathrm{d}Q^{\max}$ depends on the amount of rotational ambiguity. If the model has no rotational ambiguity, then uncertainties computed by DISP are expected to be proportional to square root of $\mathrm{d}Q^{\max}$. At the other extreme, if the rotational uncertainty is dominant, then the computed uncertainties are expected to be almost independent of $\mathrm{d}Q^{\max}$. In EPA PMF, the DISP method is implemented so that uncertainties are always computed for four different $\mathrm{d}Q^{\max}$ values. In this way, the influence of $\mathrm{d}Q^{\max}$ values on uncertainty estimates is easy to see for each specific data set.

In order to speed up computations, some factor elements may be defined as passive in DISP and BS-DISP processes. Defining some elements as passive has no influence on the uncertainty intervals obtained for active (actively displaced) factor elements. Uncertainty intervals for active factor elements are reliable regardless of how many and which elements are defined as passive, provided the user-provided data uncertainties and $\mathrm{d}Q^{\max}$ are correct. Thus it is safe to define uninteresting factor elements as passive in order to speed up computations. Note though that defining a factor element as passive will usually underestimate its computed uncertainty. Specifically, the uncertainty for a factor element defined as passive will be less than or equal to the uncertainty computed for that factor element if it were defined as active. Thus factor elements critical for associating a factor with a source should always be defined as active.

The present work offers no quantitative results for the situations where significant modeling errors exist. It was seen that one type of modeling error, specifying more factors than the data support, leads to diagnostics that suggest to an attentive PMF user that there are too many factors. However, it is not currently known whether diagnostics will be as clear if multiple modeling errors are present. For example, censoring a large number of values below detection limit, another type of modeling error, may invalidate uncertainty analysis by BS, DISP, and BS-DISP.

It was seen that some data sets produce large rotational uncertainties for some or all factors so that interval estimates may extend down to zero even for some of the defining "key" species. In such cases, factor identities may become fluid, often indicated by factor swaps. The obtained uncertainty intervals are then imprecise because of the difficulty of defining the borderline between rotations and swaps. Although the methods will correctly indicate that uncertainties are large, they may not produce quantitative results for these large intervals. On the other hand, this "weakness" caused by factor swapping may not be important in practical work. Simply put, it does not matter whether uncertainty is rather large or very large.

When interpreting large uncertainties, there is a conceptual issue that warrants highlighting. Suppose a factor is associated with a known source or sources based on the initial computed composition. For example, suppose factor F1 is identified as "Diesel vehicles" based on a high value of EC. Now suppose that the estimated uncertainty for EC for factor F1 shows that there may be low or no EC apportioned to the factor. This would then call into question the association of this factor with the postulated source. Therefore, when discussing uncertainties, they should be called uncertainties in factor F1, not uncertainties in the diesel factor. If the uncertainties are small enough that the source or sources associated with a factor are not called into question, then it is reasonable to refer to the uncertainties as uncertainties in the source profile. When reporting results, it is important to document each factor for which the size of the uncertainties calls into question the source or sources initially associated with that factor.

If large uncertainties are obtained for a PMF solution, the next step is for the analyst to determine whether physical-chemical arguments can be applied to reduce the variability of the results. Different constraints can be defined, for example, by constraining certain **G** or **F** factor elements to be zero (Paatero et al., 2002). Narrower uncertainty intervals will be obtained. However, no results from such experiments are included in this work.

It has been customary to report uncertainties in the symmetric form, as "best fit $\pm$ uncertainty". In the present case, such a formulation is not adequate since uncertainty intervals need not be symmetric. Uncertainties should be reported in an unsymmetrical formulation, for example, as "best fit $+ u - d$" where u and d represent the width of interval up and down from best fit, respectively. It should be noted that these intervals are not standard deviations of "errors". Rather, their nature is that of "Confidence Intervals", meaning that with a high (albeit often unknown) probability, the intervals contain the unknown true values.

The addition of DISP and BS-DISP capabilities in EPA PMF and ME-2 will help users better understand sources of variability in their PMF results. Such understanding may include identifying samples that are highly influential in the error estimation, identifying species for which user-provided data errors are too low or too high, or determining that too many factors have been modeled. Using DISP, BS, and BS-DISP as a suite of techniques for estimating uncertainty in PMF solutions can be more illuminating than using just one technique, much as using multiple receptor models to analyze a data set can provide more insight into the solution than using just one.

Comparing merits of different estimation principles is not straightforward, because widely varying characteristics are inherent in data sets and numerous types of modeling errors may occur. For the synthetic data developed for this work, it was seen that BS had longer uncertainty intervals and lower coverage, DISP had shorter uncertainty intervals and higher coverage, and BS-DISP had high coverage with uncertainty interval lengths between those of BS and DISP. This suggests that DISP and BS-DISP are better at assessing uncertainty than BS.

Supplementary material related to this article is available online at http://www.atmos-meas-tech.net/7/781/2014/amt-7-781-2014-supplement.pdf.

References

Abdollahi, H., Maeder, M., and Tauler, R.: Calculation and meaning of feasible band boundaries in multivariate curve resolution of a two-component system, Anal. Chem., 81, 2115–2122, 2009.

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

Amato, F., Pandolfi, M., Escrig, A., Querol, X., Alastuey, A., Pey, J., Perez, N., and Hopke, P. K.: Quantifying road dust resuspension in urban environment by multilinear engine: a comparison with PMF2, Atmos. Environ., 43, 2770–2780, 2009.

Anderson, T. W.: An Introduction to Multivariate Statistical Analysis, 2nd Edn., Wiley, New York, 1984.

Brown, S. G., Lee, T., Norris, G. A., Roberts, P. T., J. L. Collett, Jr., Paatero, P., and Worsnop, D. R.: Receptor modeling of near-roadway aerosol mass spectrometer data in Las Vegas, Nevada, with EPA PMF, Atmos. Chem. Phys., 12, 309–325, doi:10.5194/acp-12-309-2012, 2012.

Brown, S. G., Paatero, P., Eberly, S., and Norris, G.: Methods for estimating uncertainty in PMF solutions: examples using ambient air and water quality data, in preparation, 2014.

Eberly, S. I. and Reff, A.: Evaluation of rotational control in multilinear engine based receptor modeling approaches. Poster presented at the American Association for Aerosol Research Conference, Reno, NV, 24–28 September, 2007.

Efron, B. and Tibshirani, R. J.: An introduction to the bootstrap, Chapman & Hall, New York, 1993.

Engel-Cox, J. A. and Weber S. A.: 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, 1307–1316, 2007.

Gemperline, P. J.: Computation of the range of feasible solutions in self-modeling curve resolution algorithms, Anal. Chem., 71, 5398–5404, 1999.

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

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

Henry, R. C.: History and fundamentals of multivariate air quality receptor models, Chemometrics and Intelligent Laboratory Systems, 37, 525–530, 1997.

Jaumot, J. and Tauler, R.: MCR-BANDS: a user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution. Chemometrics and Intelligent Laboratory Systems, 103, 96–107, 2010.

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

Kim, E. and Hopke, P. K.: Source identifications of airborne fine particles using positive matrix factorization and U.S. Environmental Protection Agency positive matrix factorization, J. Air Waste Manage. Assoc., 57, 811–819, 2007.

Norris, G., Vedantham, R., Wade, K. S., Brown, S. G., Prouty, J. D., and Foley, C.: EPA positive matrix factorization (PMF) 3.0 fundamentals and user guide. Prepared for the US. Environmental Protection Agency, Washington, DC, by the National Exposure Research Laboratory, Research Triangle Park; Sonoma Technology, Inc., Petaluma, CA; and Lockheed Martin Systems Engineering Center, Arlington, VA, EP-D-05-004; STI-907045.05-3347-UG, October, 2008.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, Chemometrics and Intelligent Laboratory Systems, 37, 23–35, 1997.

Paatero, P. and Hopke, P. K.: Rotational Tools for Factor Analytic Models, J. Chemometrics, 23, 91–100, 2009.

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

Paatero, P., Hopke, P. K., Song, X. H., and Ramadan, Z.: Understanding and controlling rotations in factor analytic models, Chemometrics an Intelligent Laboratory Systems, 60, 253–264, 2002.

Poirot, R. L., Wishinski, P. R., Hopke, P. K., and Polissar, A. V.: Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont, Environ. Sci. Technol., 35, 4622–4636, 2001.

Reff, A., Eberly, S. I., and Bhave, P. V.: Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods, J. Air Waste Manage. Assoc., 57, 146–154, 2007.

Tauler, R.: Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution, J. Chemometrics, 15, 627–646, 2001.

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, Atmos. Chem. Phys., 9, 2891–2918, doi:10.5194/acp-9-2891-2009, 2009.