Available online at www.sciencedirect.com
ScienceDirect
Atmospheric Environment 41 (2007) 6986–6998
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

Utilizing the Chemical Mass Balance and Positive Matrix Factorization models to determine influential species and examine possible rotations in receptor modeling results

Michael J. Rizzo[a,*], Peter A. Scheff[b]

[a]USEPA, Office of Air Quality Planning and Standards, Air Quality Analysis Division, Air Quality Analysis Group, Research Triangle Park, NC 27711, USA
[b]University of Illinois—Chicago, School of Public Health, 2121 W. Taylor, Chicago, IL 60612, USA

Received 10 October 2006; received in revised form 7 May 2007; accepted 9 May 2007

Abstract

Data from two of the United States Environmental Protection Agency's Speciation Trends Network fine particulate matter sites within Chicago, Illinois were used to examine the influence that the results and profiles of the Chemical Mass Balance (CMB) receptor model have on the source contributions and profiles of the Positive Matrix Factorization (PMF) model. This was accomplished using the target shape technique, which utilizes a priori information from the CMB source profiles inputted into the PMF model. The target shape methodology involves inputting specific information for the source profiles into the PMF model as it is resolving source profile and contribution matrices. The target shape results demonstrated it is possible to determine in both the CMB and PMF source profiles those species, which do not influence the solutions of either model.

A second method utilizing information from the CMB results was used to impose a condition where the Motor Vehicles source never had a zero contribution as was applied to the CMB model. This involved utilizing an edge rotation to rotate the PMF results to yield a different solution without worsening the fit of the original results. The purpose of this work is to achieve a rotation, which produced a PMF solution where all of the Motor Vehicles contributions were greater than zero. Comparing the rotated Motor Vehicles and Sulfates source contributions in PMF to those obtained from CMB showed a better correlation between the PMF Motor Vehicles contributions to the original CMB results than those prior to rotation.

Keywords: Source apportionment; Positive Matrix Factorization; Chemical Mass Balance; $PM_{2.5}$; Speciation Trends Network (STN); Rotations; Influential species; Receptor modeling

1. Introduction

A previous work discussed a comparison between the results for the Positive Matrix Factorization (PMF) and Chemical Mass Balance (CMB) receptor models (Rizzo and Scheff, in press). The two models used data collected from the United States Environmental Protection Agency's (USEPA) Fine Particulate Speciation Trends Network for two sites in Chicago, Illinois. The results showed good consistency between the 10 source PMF and 9 source CMB solutions. However, issues of rotational

*Corresponding author. Tel.: +1 919 541 3604; fax: +1 919 541 3613.
E-mail address: rizzo.michael@epa.gov (M.J. Rizzo).

1352-2310/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2007.05.008

ambiguity in the PMF solution need to be further investigated to see if rotating the PMF solution is able to improve the comparability between the two models. Rotational ambiguity refers to the uncertainty in the PMF results which could allow for the minimization or maximization of species loadings in the source profile matrix along with the complementary change in the source contribution matrix to yield a better overall and more understandable solution without greatly changing the predicted values of the model. The changes made to the source profile matrix through the rotation are also capable of identifying the magnitude of a species' influence on the solution.

Techniques have been suggested as to how to investigate rotational ambiguity in factor analytic models (Paatero et al., 2002, 2005; Henry, 1997a–c). One such method is the use of target shapes or a priori information regarding the structure of the source profile matrix and can include information from measured source profiles used for the CMB model. Previous studies have used a priori data to pull certain species in the source profile to zero using the PMF2 software through the FKEY matrix (Lee et al., 1999; Kim et al., 2004; Zhao and Hopke, 2004). However, these studies did not determine the magnitude of the influence of pulling selected species to zero in the final result. If a species can be pulled to zero in PMF without altering the model's prediction, it would further indicate the possible influence that species would have in a CMB solution. Thus one could use a series of sensitivity analyses to see how those results change for each model. In this case, information from the CMB profiles used in the companion piece to this work will be utilized to see what effects they have on the PMF profiles. By examining any changes, one can further critique the importance of certain species in a specific source profile in PMF as well as determine the overall quality of established source profiles used in the CMB model utilizing the fine particulate speciation data collected in Chicago, Illinois.

In addition to quantifying species' influences, an important aspect of utilizing PMF in receptor modeling is to assess the possible rotational ambiguity, which may yield the same fit of the data matrix. Rotational ambiguity refers to the flexibility in the original solution to allow the "re-scaling" of the results. A valid rotation of factor analytic results "re-scales" the source profiles and contributions to yield a better overall and more understandable solution without worsening the fit of the original results. As noted previously there have been techniques suggested as to how to investigate rotational ambiguity in factor analytic models. One such technique is the examination of edges, which are linear boundary patterns observed in plots of intersource relationships. Usually these edges are rotated so that members in either the F (source profile) or G (source contribution) matrices are maximized where there are "peak" and zero values but very few intermediate values. This type of rotation is done to better achieve a more "simple structure" in the factor analytic solution.

Because factor analytic models assume that there are sufficient number of days where every source in the model has a zero contribution, it can lead to results where sources with a greater probability of always contributing have instances where they do not contribute anything. A prime example would be Motor Vehicles where in large urban areas, it is highly improbable that there will ever be a time when vehicles are not contributing something to the overall ambient fine particulate concentration especially during a 24 h average composite air sample. Therefore, it would be important to determine if the PMF solution can be rotated without greatly increasing the model's prediction error so that the Motor Vehicle contribution across the entire time span examined is never zero. This work will examine the use of the edge technique in assessing the rotational ambiguity in the PMF model and comparing those results with their corresponding values from the CMB model for the fine particulate speciation data collected in Chicago, Illinois.

2. Methodology

The Chicago data from the USEPA Speciation Trends Network used for this work including procedures for data preparation and modeling techniques have been previously described (Rizzo and Scheff, in press). The Chicago data set resulted in a PMF solution having a total of 10 factors or sources: Vegetative Burning, Motor Vehicles, Fe/Mn, a Copper source, Road Salt, Steel/Metals Processing, Industrial, Secondary Sulfates, Secondary Nitrates and Soil (Rizzo and Scheff, in press). The source contributions and profiles from PMF were in turn compared with the resulting 9 source solution from the CMB analysis which included contributions for Vegetative Burning, Motor Vehicles, Road Salt, Steel Mills, Electric Generation