Atmospheric Environment 42 (2008) 6897–6904

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Evaluation of the CMB and PMF models using organic molecular markers in fine particulate matter collected during the Pittsburgh Air Quality Study

Kerry R. Bullock [a], Rachelle M. Duvall [b], Gary A. Norris [b], Stephen R. McDow [b], Michael D. Hays [a,*]

[a] National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA
[b] National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

ARTICLE INFO

Article history:
Received 13 September 2007
Received in revised form 2 May 2008
Accepted 6 May 2008

Keywords:
Chemical mass balance model
Positive matrix factorization
Source apportionment
Organic tracers
Pittsburgh Air Quality Study

ABSTRACT

This analysis investigated different possible strategies for source apportionment of airborne fine particulate matter ($PM_{2.5}$) using data collected as part of the Pittsburgh Air Quality Study (PAQS). More specifically, we apportioned the organic fraction of the winter and summer season $PM_{2.5}$ using two source–receptor models – the EPA Chemical Mass Balance 8.2 (CMB) and EPA Positive Matrix Factorization 1.1 (PMF) models – and tested several case scenarios with each model by varying either the chemical species or source profiles used as model input. Moreover, we added the constraint of selecting only individual molecular marker species with concentrations above their minimum quantitative limits. Model results suggest that the molecular marker and source profile selection can strongly affect the model, as reflected in the source contribution estimates determined by both CMB and PMF. Biomass burning and mobile emissions sources were identified by both models as being major source contributors in Pittsburgh. A third source was consistent with a meat cooking profile but was more likely a combination of cooking and secondary organic aerosol.

As expected, the relative proportion of each source's contribution depended on both the season and on whether the CMB or PMF model was applied. Selecting fewer species in CMB resulted in less mass being apportioned, and an unrealistically large wood burning contribution estimate. Swapping a wildfire profile for one of the two wood burning profiles also resulted in less mass being apportioned in the winter. The results suggest that CMB can distinguish between fireplace burning and wildfire contributions when appropriate species are included. The gasoline/diesel split also varied by up to an order of magnitude, depending on which model was applied and which species were fit.

Published by Elsevier Ltd.

1. Introduction

Particulate matter has been implicated in a number of health effects, including respiratory symptoms, bronchitis, heart attack and premature deaths (Koike and Kobayashi, 2006; Dockery et al., 1993). Because PM is regulated at the ambient level receptor models, such as the Chemical Mass Balance (CMB) and Positive Matrix Factorization (PMF) models, are commonly used tools in regional and local air quality modeling. Source–receptor models use chemical signatures (e.g. trace elements, elemental carbon, organic markers, etc.) of source emissions and ambient samples both to identify the sources and to quantify source contributions to a sampling site. Advances in source sampling and analytical techniques have allowed for

* Corresponding author.
E-mail address: hays.michael@epa.gov (M.D. Hays).

1352-2310/$ – see front matter Published by Elsevier Ltd.
doi:10.1016/j.atmosenv.2008.05.011

greater identification of organic species in fine particulate matter ($PM_{2.5}$) which in turn have increased the pool of chemical species available for use in receptor models. Earlier studies focusing on U.S. air quality have primarily used the CMB model to apportion the organic fraction of PM (Schauer et al., 1996; Schauer and Cass, 2000; Zheng et al., 2002; Fraser et al., 2003; Fujita et al., 2003; Sheesley et al., 2004; Subramanian et al., 2006; Robinson et al., 2006a,b,c). Apportioning the PM organic fraction using PMF is more limited (Larsen and Baker, 2003; Lee et al., 2004; Buzcu and Fraser, 2006; Shrivastava et al., 2007; Jaeckels et al., 2007).

Using individual organic marker concentrations in source–receptor models presents a number of challenges. It is not always clear which organic markers should be selected as fitting species in the models because many markers undergo photochemical reaction (Robinson et al., 2006d) or fall below detection limits in the atmosphere (e.g. PAHs). Selecting source profiles for CMB also leads to uncertainties in model performance. Source profiles within the same source category or class (e.g. biomass burning) can have varying chemical compositions that influence apportionment results. Furthermore, source profiles that represent actual sources in every region are not always available, in which case the "best existing" profile must be selected as a surrogate even though it may not accurately represent the source emissions in the sampling area of interest.

The analysis presented in this paper investigates different possible strategies for obtaining optimal source apportionment results using $PM_{2.5}$ organic markers. Both the CMB and PMF models were used to apportion the regional and local sources impacting Pittsburgh from June 2001 to July 2002. In this study, we developed a method for selecting the organic species included in source–receptor models and compared the apportionment results obtained from CMB and PMF. For the CMB analysis, we selected a number of biomass combustion source profiles to evaluate the CMB model performance. We also discuss differences between CMB and PMF apportionments.

2. Experimental methods

2.1. Sample collection and analysis

Daily 24-h quartz-PUF ($d_p \leq 2.5\ \mu m$) samples were collected in Pittsburgh, PA from June 2001 to July 2002 (96 samples total) as part of the Pittsburgh Air Quality Study (PAQS) (Wittig et al., 2004). Overall, sampling was conducted every sixth day, with two intensive sampling periods (July 5 to August 3, 2001 and January 3 to January 12, 2002) containing daily measurements. Organic and elemental carbon (OC and EC) were measured by thermal evolution and combustion (Birch and Cary, 1996). A total of 48 organic marker species were measured by gas chromatography/mass spectrometry (GC–MS) (Subramanian et al., 2006).

2.2. Source–receptor modeling

Source apportionment was conducted using the EPA CMB 8.2 Model (Coulter, 2004) and the EPA PMF 1.1 Model (Eberly, 2005). The CMB 8.2 model uses an effective variance least-squares algorithm to apportion the ambient data to selected source profiles. The PMF model uses a constrained, weighted, least-squares algorithm to generate source profiles and source contributions from ambient data. For the CMB analysis, the ambient samples were averaged into two seasonal profiles based on the average monthly low temperatures in Pittsburgh. Samples collected from November to March were labeled as the "winter" season, and samples collected from April to October were labeled as the "summer" season.

Prior to running the PMF and CMB models, a method was developed for selecting species to include as input for each model. Because the PMF model generates factors and source contributions based on data from an entire sampling period, it is important to select species that are stable and conserved in the atmosphere and have at least half of the data above detection limits. Species selection is equally critical in CMB since species conservation is one of the assumptions of the CMB model. To determine which species to select, the method detection limit (MDL) was calculated for each organics species measured. The MDL was then multiplied by a factor of 5 to obtain the minimum quantitative limit (MQL). The MQL indicates the level at which the GC–MS instrument is able to reliably quantify an organic specie. Species concentrations that were above the MQL in less than 25% of the samples were eliminated. The remaining organic species were included in the PMF and CMB model runs. For the PMF model, depending on the number of times the measured species was above the MQL during the sampling period, the species was assigned to a category (either strong, moderately strong, or weak). Species that were above the MQL at least 75% of the time were assigned as strong, those that were above the MQL between 50 and 75% of the time were designated moderately strong, and those that were above the MQL between 25 and 50% of the time were designated weak. Because they aid in the interpretation of the apportionment solution, organic markers measured below MQLs are typically included in the apportionment model calculation. However, the concentrations of organic markers detected below their MQLs are more uncertain. Our method of organic species selection was developed in an effort to reduce the uncertainty in the source apportionment that might otherwise result from including organic markers detected below their MQLs. Despite the constraints the species selection process imposed, numerous organic markers (at least 24 of 48 possible species) were retained for modeling. Species that were eliminated based on MQLs included *n*-Tetratriacontane, iso-Hentriacontane, anteiso-Hentriacontane, Pimaric Acid, Dehydroabietic Acid, and all but four of the hopanes and cholestanes.

Using the CMB and PMF models and the species selection technique, a number of case scenarios were investigated. Table 1 displays the scenarios and corresponding species selected for this study. Case A included 33 fitting species (32 organic marker species and elemental carbon). Case B was formed as a subset of the Case A species list to examine the effect of fitting fewer species to the models. Case B included 24 fitting species (23 organic species and elemental carbon). There is evidence that these species