



Source apportionment of PM_{2.5} inside two diesel school buses using partial least squares discriminant analysis with chemical mass balance

Timothy Larson^{1,2}, Barbara Zielinska³, Rob Ireson⁴, L.J. Sally Liu^{2,5,6}

¹ Department of Civil and Environmental Engineering, Box 352700 University of Washington, Seattle, WA 98195

² Department of Environmental and Occupational Health Sciences, University of Washington

³ Atmospheric Sciences Division, Desert Research Institute, Reno, NV.

⁴ Air Quality Management Consulting, Greenbrae, CA.

⁵ Swiss Tropical and Public Health Institute

⁶ University of Basel, Basel, Switzerland

ABSTRACT

Uncertainty-weighted partial least squares discriminant analysis was used to identify key species that were subsequently included in the EPA CMB8.2 chemical mass balance model to assess PM_{2.5} source contributions from a previously published data set on school bus self-pollution. Estimates from this two-step modeling approach, herein referred to as effective variance discriminant analysis chemical mass balance (EVDA–CMB) were compared for eight separate runs with independent estimates from a synthetic tracer method. EVDA–CMB model predictions agreed favorably with those from the tracer method ($R^2 = 0.83, 0.96$ and 0.48 , for contributions from the bus tailpipe, the engine crankcase and from other sources, respectively). Predictions from the traditional CMB model (without prior species selection), did not agree as well with the tracer method estimates of the bus tailpipe and engine crankcase contributions ($R^2 = 0.18, 0.69$, respectively), but did agree as well with the contributions from other sources ($R^2 = 0.60$). Although this study required discrimination of only a few sources, the same approach could be applied to the more general receptor modeling problem as an initial screening procedure, including approaches that optimize the choice of variables based on ambient data. This is important given that the number of species available for use in receptor modeling is rapidly expanding.

Keywords:

Chemical mass balance
Partial least squares discriminant analysis
Source apportionment

Article History:

Received: 18 June 2010

Revised: 23 December 2010

Accepted: 19 January 2011

Corresponding Author:

Timothy Larson

Tel: +1-206-543-6815

Fax: +1-206-685-3836

E-mail: tlarson@uw.edu

© Author(s) 2011. This work is distributed under the Creative Commons Attribution 3.0 License.

doi: 10.5094/APR.2011.019

1. Introduction

Receptor modeling has been used to estimate the contribution of various sources to measured airborne particulate matter concentrations (Henry, 1997; Hopke and Song, 1997; Seigneur et al., 1997). Traditionally, the U.S. EPA has recommended using the effective variance weighted chemical mass balance (CMB) receptor model (Miller et al., 1972; Watson et al., 1984), although less constrained multi-variate approaches have recently been widely used (Paatero, 1997; Henry et al., 1999; Paatero, 1999; Henry, 2003). More recent applications of the CMB method have explored the use of unique particulate organic tracers (Schauer et al., 1996; Zheng et al., 2002) as well as combined particulate and gaseous tracers (Schauer and Cass, 2000; Schauer et al., 2002).

One less well known alternative to CMB is partial least squares regression (PLS). PLS was originally developed by Herman Wold (Wold, 1966; Wold, 1981) and took his name when it was applied to the over-determined regression problem (Wold et al., 1983; Geladi and Kowalski, 1986). It was first applied to the aerosol source apportionment problem by Frank and Kowalski (1985). Vong et al. (1988) showed how PLS could solve this apportionment as a discriminant analysis problem. This latter approach has since been used in a limited number of similar studies (Larson and Vong, 1989; Vong, 1993; Wang and Larson, 1993; Norris, 1998). Similar to the effective variance-weighting scheme used in the EPA's CMB model (Watson et al., 1984), Norris (1998) introduced the idea of

uncertainty weighted PLS, thereby accounting for individual species measurement uncertainties.

Here we apply uncertainty weighted PLS in order to determine key tracer species for subsequent use in a traditional chemical mass balance (CMB) model in order to estimate the source contributions to PM_{2.5} inside a school bus. This two-step CMB model incorporating prior PLS discriminant analysis is one realization of what we refer to here as an effective variance discriminant analysis chemical mass balance model (EVDA–CMB).

The PLS algorithm provides an automated way to identify and highlight those species that differentiate the proposed sources, down-weighting the other species (Vong et al., 1988; Larson and Vong, 1989; Norris, 1998). These species are then used in CMB. The subsequent source contribution estimates are then compared with independent estimates of the relative contributions from each source that have been established by the use of unique, synthetic source tracers (Ireson et al., 2004; Zielinska et al., 2008; Liu et al., 2010).

Our data set is described in more detail elsewhere (Zielinska et al., 2008; Liu et al., 2010) and consists of ambient filter samples taken inside two diesel school buses and source samples taken from the tailpipe, from the crankcase road draft tube, and from the roadway traversed by each bus ("other sources"). Our initial attempts at CMB were only moderately successful in deducing the

relative source contributions to in-bus concentrations as judged by comparison with results obtained from the tracer-based method. This discrepancy was due in part to the relatively large number of measured species in this data set and the accompanying difficulty in selecting the appropriate species for use with CMB. We therefore decided to explore the use of an alternate species selection method for use with CMB, as described below.

2. Methods

The sampling and analysis methods are described in detail elsewhere (Zielinska et al., 2008; Liu et al., 2010) and briefly here. Unique, synthetic tracers were added to both the fuel supply and the lubricating oil. Tris(norbornadiene)iridium(III)acetylacetonate, an organometallic iridium complex was dissolved in toluene (1 g:225 mL) and added to each bus' fuel tank to track tailpipe exhaust particulate. Fully deuterated normal hexatriacontane ($n\text{-C}_{36}\text{D}_{74}$ or $d\text{-alkane}$) was dissolved in the bus' lubricant oil (100 g:18.9 L) to track crankcase emissions.

Source sampling involved using an on-board dilution tunnel (Weaver and Petty, 2004) to collect $\text{PM}_{2.5}$ samples from the tailpipe and the crankcase, respectively, of each bus. A lead vehicle drove the same route as the bus, ahead of the bus by approximately 5 minutes. A set of source profiles to represent other sources was developed based upon computed mass fractions of the species measured on the lead vehicle samples. In addition, a total of eight in-bus/lead vehicle sample pairs (Teflon and quartz filters) were taken using identical UMD impactors at 120 L/min. The windows in the lead vehicle were wide open during all sampling runs. The concentrations of particulate organic compounds are described in detail by Zielinska and co-workers (Zielinska et al., 2008).

The uncertainties for the in-bus samples were taken directly from the reported analytical uncertainties. The measurement uncertainties for the XRF and OC/EC fractions were reported using standard EPA protocols. The analytical uncertainties for the organic species were based on known deuterated internal standards. Compounds for which authentic standards were not available were quantified based on the response factor of standards most closely matched in structure and retention characteristics (Zielinska et al., 2008). There were three sets of source samples taken for each of the eight runs (with one sample excluded due to sampling issues). The average analytical uncertainties of the three samples taken during each run were used as the uncertainties in this analysis.

2.1. CMB diagnostics

The standard EPA model, CMB8.2, was used in this analysis. It employs a weighted ordinary least squares solution to the following mass balance equation

$$C = FS + \varepsilon \quad (1)$$

where C ($n \times 1$) is the vector of observed concentrations of n species ($\mu\text{g}/\text{m}^3$), F ($n \times p$) is a source profile matrix of n species from p sources ($\mu\text{g}/\mu\text{g mass}$), S ($p \times 1$) is the source contribution vector ($\mu\text{g mass}/\text{m}^3$), and ε ($n \times 1$) is the vector of random measurement errors. The species are weighted by their respective measurement uncertainties involving an iterative procedure that includes the one standard deviation measurement uncertainties for the i^{th} species in both the source and ambient samples, σ_{source} ($\mu\text{g}_i/\mu\text{g mass}$) and σ_{amb} ($\mu\text{g mass}/\text{m}^3$) respectively (Watson et al., 1984). Specifically, the weighted equation that is actually solved is:

$$C_W = F_W S + \varepsilon' \quad (2)$$

where

$$C_W = (V_e)^{-0.5} C \quad (3)$$

and

$$F_W = (V_e)^{-0.5} F \quad (4)$$

V_e ($n \times n$) is the diagonal effective variance matrix whose off-diagonal elements are zero and whose diagonal elements are:

$$v_{e_{ii}} = (\sigma_{\text{amb}})_i^2 + \sum_{j=1}^p (\sigma_{\text{source}})_{i,j}^2 (s_j)^2 \quad (5)$$

where s_j is the contribution from the j^{th} source. The fact that the s_j are computed from Equation (2) means that the CMB8.2 algorithm is implicit and thus iterative. The first iteration initially assumes all the s_j are zero in Equation (5) and then computes the s_j from Equations (2)–(4) for use in subsequent iterations (Watson et al., 1984). The iteration procedure is stopped when the current and prior value of s_j are within one percent of each other. The final source contribution estimates in the original mass concentration units are then computed as:

$$S = F_W^+ C_W = \left(F^t (V_e)^{-1} F \right)^{-1} F^t (V_e)^{-1} C \quad (6)$$

By definition, the modified pseudo-inverse matrix (MPIN) is given as:

$$\text{MPIN} = F_W^+ = \left(F^t (V_e)^{-1} F \right)^{-1} F^t (V_e)^{-0.5} \quad (7)$$

Guidance is provided within CMB8.2 on those species that are influential and thus should be included in the model. Specifically, the elements of the normalized MPIN matrix, whose values range from -1 to 1 , should be greater than 0.5 for species that are to be retained in the model (Kim and Henry, 1999; Watson, 2004).

Additional run diagnostics in CMB8.2 provide measures of the collinearity of the given set of weighted source profiles, including Henry's (1992) eligible space based on the singular value decomposition of the weighted F matrix as follows:

$$(V_e)^{-0.5} F = A D V^t \quad (8)$$

where A ($n \times n$) and V ($p \times p$) are orthogonal matrices and D is a diagonal matrix with p nonzero and positive elements called the singular values of the decomposition. V is the matrix of eigenvectors of the decomposition. The eligible space is that spanned by these eigenvectors with inverse singular values less than or equal to the maximum score uncertainty. The estimable sources are those with a user defined minimum source projection within the estimable space, set at a default value of 0.95 . CMB8.2 provides suggestions for combining highly collinear profiles (Henry, 1992), but provides no additional guidance on species selection so as to minimize collinearity of existing source profiles. Several authors have suggested alternative methods to minimize the collinearity problem, including ridge regression (Hopke, 1985) and non-negative principal component regression (Shi et al., 2009).

2.2. Species selection based on effective variance weighted discriminant analysis

As an alternate species selection strategy, we present here an effective variance weighted, partial least-squares discriminant analysis algorithm to select influential species for inclusion in the CMB model (EVDA-CMB) while minimizing collinearity. Source contributions from the crankcase, the tailpipe and other sources

Download English Version:

<https://daneshyari.com/en/article/4434953>

Download Persian Version:

<https://daneshyari.com/article/4434953>

[Daneshyari.com](https://daneshyari.com)