Environmental Modeling

PM2.5 Source Apportionment Using a Hybrid Environmental Receptor Model

L.-W. Antony Chen, and Jun-ji Cao

Environ. Sci. Technol., **Just Accepted Manuscript** • Publication Date (Web): 03 May 2018

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

$PM_{2.5}$ Source Apportionment Using a Hybrid Environmental Receptor Model

L.-W. Antony Chen[†,*] and Junji Cao[‡]

[†]Department of Environmental and Occupational Health, School of Community Health Sciences, University of Nevada, Las Vegas, NV, 89154, USA

[‡]Key Laboratory of Aerosol Chemistry & Physics (KLACP), Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

[*]Corresponding author: antony.chen@unlv.edu, 1(702)895-1420

Abstract

A Hybrid Environmental Receptor Model (HERM) that unifies the theory of effective-variance chemical mass balance (EV-CMB) and positive matrix factorization (PMF) models was developed to support the weight-of-evidence approach of air pollution source apportionment. The HERM software is capable of 1) conducting EV-CMB analysis for multiple samples in a single iteration; 2) calculating EV-CMB and PMF source contributions as well as middle grounds (hybrid mode) between the two using partial source information available for the study region; 3) reporting source contribution uncertainties and sample-/species-specific fitting performance measures; 4) interfacing with MS Excel® for convenient data inputs/outputs and analysis. Initial testing with simulated and real-world $PM_{2.5}$ (fine particulate air pollutants with aerodynamic diameter < 2.5 μm) datasets show that HERM reproduces EV-CMB results from existing software but with more tolerance to collinearity and better uncertainty estimates. It also shows that partial source information helps reduce rotational ambiguity in PMF, thus producing more accurate partitioning between highly correlated sources. Moreover, source profiles generated from the hybrid mode can be more representative of the study region than those acquired from other studies or calculated by PMF with no source information. Strategies to use HERM for source apportionment are recommended in the paper.

Keywords

Receptor model, chemical mass balance, PMF, $PM_{2.5}$ source apportionment

INTRODUCTION

Receptor models have been widely used for source apportionment of particulate and gaseous air pollutants, allowing control efforts to be focused on sources that contribute most to the environmental and health effects.[1-5] In principle the speciation of pollutants at a receptor site reflects the emissions of individual sources and their chemical compositions, also known as source profiles. The most general form of chemical mass balance (CMB) model that links source profiles to ambient chemical composition considers the atmospheric transport and transformation,[6-7] thus:

$$C_{ik} = \sum_{j} (T_{ijk} F_{ij})(D_{jk} Q_{jk}) \tag{1}$$

where

C_{jk} : the measured concentration of a pollutant *i* at sample *k*

Q_{jk}: the total emission from source *j* corresponding to the sample *k*

D_{jk}: the fraction of emissions arriving at the receptor site due to atmospheric transport

F_{ij} : the source profile, i.e., fractional quantity of pollutant *i* in source *j* emission

T_{ijk}: describe how the source profiles evolve/transformation during the transport

In an ideal situation where F_{ij} are measured accurately and comprehensively for the region of interest and where atmospheric transformation is negligible ($T_{ijk} \sim 1$) or can be simulated adequately, Eq. (1) is simplified to:

$$C_{ik} = \sum_{j=1}^{J} F_{ij} S_{jk} \tag{2}$$

where *J* indicates the number of sources that impact the receptor site and the source contribution S_{jk} (equal to $D_{jk}Q_{jk}$) can be quantified from measured C_{ik} and F_{ij} by non-weighted linear regression, providing that number of species is more than the number of sources in the model.

The effective variance (EV) regression[8] takes into account uncertainties in both C_{ik} and F_{ij} resulting from either measurement or variability in source emissions. EV-CMB solves for S_{jk} (j = 1 to J for sample k) that minimize the reduced chi-square:

$$\chi_k^2 = \frac{1}{I-J}\sum_{i=1}^{I}\frac{\left(C_{ik}-\sum_{j=1}^{J}F_{ij}S_{jk}\right)^2}{\sigma_{C_{ik}}^2+\sum_{j=1}^{J}\sigma_{F_{ij}}^2 S_{jk}^2} \quad (3)$$

where $\sigma_{C_{ik}}$ and $\sigma_{F_{ij}}$ are uncertainties of the measured concentrations and profile abundances, respectively. I and J are the number of species and sources, respectively; I - J that precedes the summation accounts for the degree of freedom (DF) in the model. EV refers to the denominator in Eq. (3), thus:

$$EV_{ik} = \sigma_{C_{ik}}^2 + \sum_{j=1}^{J}\sigma_{F_{ij}}^2 S_{jk}^2 \quad (4)$$

Watson et al.[8] developed an iterative algorithm, later adopted by the EPA CMB software,[9-11] to solve Eq. (3). This algorithm works on one sample at a time, starting with the solution of ordinary weighted linear regression (in that case $EV_{ik} = \sigma_{C_{ik}}^2$ only) for initial S_{jk}, updating EV at each iteration based on new S_{jk}, and continuing until S_{jk} is converged. The final χ_k^2 suggests the goodness of fit. There is no non-negative constraint in the algorithm, though the EPA CMB software enables a "source elimination mode" that automatically removes sources with negative contribution and recalculates S_{jk}. In addition, convergence may not be achieved if highly collinear source profiles are included in the model.

The development of Multi-Linear Engine (ME-2)[12] offers an alternative to solve Eq. (3). ME-2 uses an iterative conjugate gradient algorithm to approach a local and/or global minimum for any defined multilinear problems such as CMB. It can handle multiple samples by expanding the definition of reduced chi-square in Eq. (3) to:

$$\chi^2 = \frac{1}{K(I-J)} \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{\left(C_{ik} - \sum_{j=1}^{J} F_{ij} S_{jk}\right)^2}{EV_{ik}} \quad (5)$$

ME-2 solves S_{jk} for all sources (j = 1 to J) and all samples (k = 1 to K, where K is the number of samples) simultaneously. Note DF in the model increases to $K(I - J)$. Theoretically Eq. (5) is equivalent to Eq. (3) since S_{jk} that minimize every χ_k^2 defined in Eq. (3) must also minimize the overall χ^2 in Eq. (5). Nonnegativity constraints have been implemented in ME-2 and so source contributions can only be zero or above. As we will show, the conjugate gradient algorithm tolerates collinearity better than the conventional EV regression in EPA CMB software. It produces solutions even when EPA CMB fails to converge.

Assuming no uncertainty associated with any F_{ij} (i.e., $\sigma_{F_{ij}}$ = 0), Eq. (5) would be reduced to that implemented by the positive matrix factorization (PMF) model, thus:

$$\chi^2 = \frac{1}{K(I-J)-IJ} \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{\left(C_{ik} - \sum_{j=1}^{J} F_{ij} S_{jk}\right)^2}{\sigma_{C_{ik}}^2} = \frac{1}{KI-J(K+I)} \sum_{k=1}^{K} \sum_{i=1}^{I} \frac{\left(C_{ik} - \sum_{j=1}^{J} F_{ij} S_{jk}\right)^2}{\sigma_{C_{ik}}^2} \quad (6)$$

PMF, a factor analysis model, gains popularity in the last two decades for PM and volatile organic compounds (VOCs) source apportionment.[13] It is typically applied to CMB problems where appropriate source profiles are not available, let alone source profile uncertainties, due to the lack of source testing data and/or substantial atmospheric modification of primary emissions. The model seeks F_{ij} and S_{jk} that minimize χ^2 in Eq. (6) simultaneously. Since all F_{ij} are unspecified, DF in the model is reduced by $I \times J$ from Eq. (5) to Eq. (6). PMF relies on variability in chemical composition across ambient samples and therefore work best for a large dataset (i.e., many C_{ik})

with highly variable source contributions. The popular EPA PMF 5.0 software employs ME-2 to solve Eq. (6).[14] The main issue with PMF is the rotational ambiguity, i.e., F_{ij} and S_{jk} matrixes can be rotated in opposite direction to yield new solutions. This often leads to non-unique solutions despite the nonnegativity constrains on both F_{ij} and S_{jk}, and some of the solutions may not even be physically possible. Although PMF calculations do not involve source profiles explicitly, the resulting "factors" are often interpreted based on how they compare with known source profiles.[15-17]

Source apportionment by EV-CMB and PMF has been compared in recent studies for rural[17-19], urban[20,21], and industrial[22-24] environments. While they both quantify major source contributions, biases between the two are often attributed to CMB profiles being representative of "fresh" source emissions ignoring transformation or "aging" between the source and receptor. Although PMF factors better capture the aging process, they inevitably mix sources together. Moreover, EV-CMB more likely resolves minor sources[17,18,23], and its performance is best with locally-measured source profiles[22,24]. One major shortage of these studies is the lack of using simulated datasets to evaluate the absolute accuracy of the models. On the other hand, Shi et al.[25] used simulated data to evaluate the EV-CMB performance under serious collinearity conditions.

This paper describes the development and evaluation of a Hybrid Environmental Receptor Model (HERM), which is built upon the ME-2 solution to EV-CMB problems (Eq. [5]). HERM differs from the current CMB software (i.e., EPA CMB v8.2) in the ability to analyze one or multiple samples in a single iteration, inherent non-negativity constraints, and better tolerance to collinearity. Most important of all, HERM bridges EV-CMB to PMF by allowing the use of incomplete or partial source profiles. In many situations, the lack of high-quality source profile(s) for every known source hinders successful CMB source apportionment. A few studies attempted to incorporate source information into PMF or ME-2 by constraining ratios of marker species in the factors.[26,27] HERM can take all reliable source profile information while estimating unknown sources and/or missing species in the source profiles. This feature also helps characterize "aged" source profiles when they vary substantially from source testing results (i.e., the "fresh" source profiles). When no source profiles are used, HERM would return to the PMF configuration (Eq. [6]) to calculate factor profiles and contributions. Virtually the model is capable of reporting both EV-CMB and PMF source apportionment, as well as any middle ground between the two. The current Chinese Academy of Sciences (CAS) HERM software comes with a Microsoft Excel® user interface to facilitate data input, output, and analysis. Simulated particulate matter (PM) data were generated to evaluate the HERM performance with different degrees of source information. Moreover, the model was applied to a real-world PM dataset previously analyzed by EV-CMB to offer additional insights into the receptor modeling process.

TECHNICAL APPROACHES

Algorithms

The ME-2 Basic_2way (B2W) script was modified to accommodate HERM requirements. B2W solves the PMF problem assuming all F_{ij} and S_{jk} are unknown and to be solved. The model inputs include ambient measurements C_{ik}, uncertainty $\sigma_{C_{ik}}$, and the number of factors J. In addition to the CMB equation (Eq. [2]), B2W implements a normalization scheme that constrains the average source contribution, $\sum_{k=1}^{K} S_{jk} / K$, to 1 for each factor j, thus limiting the number of possible solutions. The modifications to B2W include the following:

- Select the non-robust mode to calculate χ^2, as robust mode automatically downweight apparent outliers[12] and so would not be consistent with EV-CMB calculations. The HERM software allows easy switch between the robust and non-robust mode.

- Read source profiles into the model, with the number of profiles no more than J. Lock F_{ij} that correspond to the profiles (i.e., fix them to the initial values throughout iteration). Assign a priori (or random) values to non-locked F and all S elements to begin the first iteration.
- Read profile uncertainties ($\sigma_{F_{ij}}$) into the model for calculating EV. Assume zero $\sigma_{F_{ij}}$ for any non-specified or non-locked F_{ij}.
- Remove the auxiliary equations that normalize the average of S_{jk} (over all samples) to unity, considering that F_{ij} are locked.
- Replace error $\sigma^2_{C_{ik}}$ with EV_{ik} (Eq. [4]) and update it at every iteration of conjugate gradient calculation using S_{jk} from the previous iteration until the convergence is reached. Final values of S_{jk} is reported as source contribution estimates.

In the case of conventional EV-CMB problem where each factor is assigned a full source profile (i.e., all F_{ij} are locked), HERM reports χ^2_k and χ^2 as defined in Eqs. (3)-(5), along with source contribution S_{jk}. Uncertainty (i.e., standard deviation $\sigma_{S_{jk}}$) of S_{jk} is then estimated by:

$$\sigma^2_{Sjk} = (F'(dEV_k)^{-1}F)^{-1}_{jj} \times \chi^2_k \qquad (7)$$

where F is the $I\times J$ profile matrix and dEV_k is an $I\times I$ diagonal matrix with diagonal elements $(dEV_k)_{ii} = EV_{ik}$. Eq. (7) takes into account both the EV and goodness of fit,[28] though EPA CMB ignores the latter (χ^2_k)[10,11] A larger χ^2_k indicates worse fit and certainly larger uncertainty in the source contribution estimate. The sample-specific correlation of fitting (r^2_k) is also calculated:[11]

$$r_k^2 = 1 - \frac{(I-J)\chi_k^2}{\sum_{i=1}^{I} \frac{C_{ik}{}^2}{EV_{ik}}} \tag{8}$$

Higher r_k^2 and lower χ_k^2 generally suggest the particular sample is fitted better by the model. In addition, HERM calculates species-specific χ_i^2 and r_i^2, where:

$$\chi_i^2 = \frac{I}{K(I-J)} \sum_{k=1}^{K} \frac{\left(C_{ik} - \sum_{j=1}^{J} F_{ij} S_{jk}\right)^2}{EV_{ik}} \tag{9}$$

$$r_i^2 = 1 - \frac{K(I-J)\chi_i^2}{I \times \left(\sum_{k=1}^{K} \frac{C_{ik}{}^2}{EV_{ik}}\right)} \tag{10}$$

χ_i^2 and r_i^2 help diagnosis of the results, e.g., identifying species that are not fitted as well (high χ_i^2 and low r_i^2) across all samples. They are not reported by the current EPA CMB software.

If HERM needs to solve profiles that are not assigned a priori and/or some species that are missing in the profiles (i.e., the “hybrid” or PMF mode), EV_{ik} is generalized to:

$$EV_{ik}^{*} = \sigma_{C_{ik}}^2 + \sum_{j=1}^{J} (\sigma_{F_{ij}}^2 S_{jk}^2 + \beta \delta_{ij} \sigma_{C_{ik}}^2) \tag{11}$$

Here $\delta_{ij} = 0$ if source profile element F_{ij} is specified and $\delta_{ij} = 1$ when F_{ij} is unknown or missing in the profiles, thus setting $\sigma_{F_{ij}}$ to zero. β is an adjustable factor with a default value of 1. The last term in Eq. (11) avoids the model to overweight unspecified profile species in the fitting process due to a zero uncertainty. Missing (unlocked) F_{ij} also decrease DF in the model, and therefore definitions of χ^2, χ_k^2, and χ_i^2 should be modified accordingly. For the hybrid mode,

$$\chi^{*2} = \frac{1}{K(I-J) - \sum_{i=1}^{I}\sum_{j=1}^{J}\delta_{ij}} \sum_{k=1}^{K}\sum_{i=1}^{I} \frac{\left(C_{ik} - \sum_{j=1}^{J} F_{ij} S_{jk}\right)^2}{EV_{ik}^*} \qquad (12)$$

is used in the calculation, instead of Eq. (5) for the EV-CMB mode. Eq. (12) returns to Eq. (5) when all F_{ij} are locked ($\delta_{ij} = 0$), and it becomes the PMF formulation when no profile information is used ($\sigma_{F_{ij}} = 0$, $\delta_{ij} = 1$); in that case,

$$\chi^{*2} = \frac{1}{1+\beta J}\chi^2 \qquad (13)$$

where χ^2 is that defined in Eq. (6). Other generalized formulas are listed in the supporting information Table S1.

User Interface

The current CAS HERM v1.8 software takes inputs in Microsoft Excel® format. Each input file should contain 6 tabs: 1) speciated ambient measurements (C_{ik}); 2) speciated measurement uncertainties ($\sigma_{C_{ik}}$); 3) source profiles (F_{ij}); 4) source profile uncertainties ($\sigma_{F_{ij}}$); 5) source profile specifications (keys); and 6) other model parameters, all of which are organized in matrix form (see supporting information Figure S1 for an example). $\sigma_{C_{ik}}$ is determined from the measurement precision (%) and minimal detection limit (MDL) of each species[29], with examples shown in Table S2, while $\sigma_{F_{ij}}$ also takes into account the standard deviation of the averaged abundances from multiple source testings[30,31]. Typically the first species in ambient measurements and in source profiles is the normalization (total) species, such as PM mass or total VOCs concentration. The software allows users to specify profile keys corresponding to specific F_{ij} to be either "locked" (EV-CMB mode) or "non-locked" (hybrid/PMF mode). Species will be fitted whether they are locked or not. To exclude a species from fitting, one can remove the species from the ambient measurements or assign it a relatively large uncertainty so it contributes little to χ^2. It should be noted that typically EV-CMB does not fit the total species, and instead compares it with that reconstructed from the solution to inform the model performance (e.g., "%mass" in EPA CMB v8.2, see Coulter[11]). Other parameters in CAS HERM include the number of species (I), samples (K), and sources (J, specified plus unspecified), as well as a seed for the random numbers and the number of repeated runs with different seeds.

CAS HERM passes the input to ME-2, which starts iteration with initial profiles, if specified, or random values. Upon convergence, ME-2 passes the final source profiles and contributions to CAS HERM, along with χ^2 for each run. Further CAS HERM calculates sample-specific χ_k^2, r_k^2, and $\sigma_{S_{jk}}$, as well as species-specific χ_i^2 and r_i^2. A scatter plot of measured versus calculated concentrations for each species is presented, along with the breakdown of source contributions to that species. Due to the numerical nature of ME-2, repeated runs can yield different results, and the users can select to report one (e.g, with the lowest χ^2) or multiple run results for further analysis. All CAS HERM outputs are also in MS Excel® format with different information displayed in different tabs. Input information is included in the output file to facilitate data management, comparison, and interpretation.

The current CAS HERM does not contain error estimation tools such as bootstrapping (BS) or displacement of factor elements (DISP) that are implemented in the EPA PMF 5.0[14] software. These tools aim at quantifying uncertainties in factor profiles resolved by PMF due partly to noises and rotational ambiguity and could help evaluate the robustness of model solutions, especially if HERM is required to address unknown and/or incomplete source profiles. They will be integrated into future versions. Meanwhile, repeated runs (typically 10-20, with random initial values) in CAS HERM provide a clue for the model robustness and the solution with the lowest χ^2 is used in the following discussions.

Simulated and Ambient Test Datasets

Simulated $PM_{2.5}$ (fine PM with aerodynamic diameter < 2.5 μm) data were generated from 5 real-world source profiles, including a secondary ammonium sulfate (AMSUL), a secondary ammonium nitrate (AMNIT), a biomass burning (BB), a motor vehicle exhaust (MV), and an urban dust (U-Dust) profiles, used in the Reno $PM_{2.5}$ source apportionment study.[32] Each profile consists of water-soluble ions (NO_3^-, $SO_4^=$, NH_4^+, Na^+, K^+), organic carbon (OC), elemental carbon (EC), and thermal/optical carbon fractions as quantified by the IMPROVE_A protocol,[33] elements (Al to Pb), levoglucosan, as well as selected polycyclic aromatic hydrocarbons (PAHs), hopanes, and alkanes, for a total of 44 species that are normalized to the $PM_{2.5}$ mass (see supporting information Table S2). For each sample, source profiles were perturbed stochastically from the defined means (F_{ij}) and standard deviations ($\sigma_{F_{ij}}$) of the 5 sources. They were then multiplied by pre-specified S_{jk} (0 – 10 μg/m^3 of $PM_{2.5}$ for AMNIT and BB, and 0 – 5 μg/m^3 of $PM_{2.5}$ for AMSUL, MV and U-Dust) to determine the speciated $PM_{2.5}$ concentrations at the receptor site (Eq. [2]), which were finally perturbed to simulate "would-be" measured values, C_{ik}, according to the defined measurement uncertainties ($\sigma_{C_{ik}}$). Therefore, the simulated C_{ik} reflect both the source variability and measurement errors.

Two sets of simulated data, each of which contained 50 samples, were developed to challenge the receptor models. The first set (Scenario A) assumed correlations between the AMNIT and BB contributions (r^2 = 0.5) and between the MV and U-Dust contributions (r^2 = 0.8), a real-world situation as found by Chen et al.[32]. There were no correlations (r^2 < 0.1) between any other pairs of sources. For the other set of data (Scenario B), AMNIT was replaced by a road dust source (R-Dust) that had a varying degree of collinearity with U-Dust. The R-Dust source profile, i.e., $F_{i,R\text{-}Dust}$, simulates U-Dust being contaminated by brake wear, thus:

$$F_{i,R-Dust} = \alpha \times F_{i,U-Dust} + (1-\alpha) \times F_{i,Brake} \tag{14}$$

where $F_{i,Brake}$ is the source profile of brake wear[34] with high iron (Fe) and manganese (Mn) contents, and α determines the degree of collinearity ranging from 0 (no collinearity) to 1 (full collinearity). It should be noted that collinearity also depends on C_{ik}, the receptor data to be fitted.[35] The R-Dust profile uncertainty, $\sigma_{F_{i,R-Dust}}$, was calculated following the rule of error propagation. No correlations were assumed for any pairs of source contributions in this scenario. Scenario B therefore was based on AMSUL, BB, MV, U-Dust, and a range of R-Dust, while the other principles for constructing ambient C_{ik} remained the same as Scenario A.

Ambient $PM_{2.5}$ data acquired from the Bliss State Park (BSP), California, and previously analyzed for source apportionment[18] served to further test the receptor models. BSP, located in the scenic Lake Tahoe Basin, is part of the Interagency Monitoring of PROtected Visual Environments (IMPROVE) network designed to track the long-term trends of visibility in U.S. national parks and wildlife reserves.[36,37] The site is impacted by local sources, particularly wood burning in nearby communities and wildlands and traffic from tourists, as well as long-range transport of natural and anthropogenic pollutants. The IMPROVE network quantifies only inorganic species, including mass, NO_3^-, $SO_4^=$, H^+, OC, EC, and 21 elements, on an every 3rd day basis. Based on EPA PMF and EV-CMB models, Green et al.[18] attributed $PM_{2.5}$ during 2005–2009 to 9 sources, i.e., AMSUL, AMNIT, wood burning with both high and low combustion efficiencies (BBh and BBl), motor vehicles (MV), two road dusts (RDust1 and RDust2), Asian dust (ADust), and miscellaneous coal combustion (Coal), with the wood burning emissions dominating throughout the year. Source profiles used for EV-CMB (Table S3) differed appreciably from those resolved by PMF.[18] Although these source apportionment results satisfied general receptor modeling guidelines,[38,39] there was a discrepancy between the measured and EV-CMB-calculated $PM_{2.5}$ mass. This discrepancy might result from some source profiles being unrepresentative. Particularly, the wood burning profiles that were acquired near the burns represented fresh smoke better than aged smoke that actually impacted the BSP site.[40]

RESULTS

Consistency of HERM with EPA CMB

HERM and EPA CMB was first applied to the simulated "Scenario A" dataset using known source profiles (i.e., all profiles are "locked"). Both models calculated S_{jk} and $\sigma_{S_{jk}}$ for the 50 samples based on EV-CMB, and they are compared with actual source contributions in Table 1. All 50 HERM and EPA CMB iterations converged and no sources were eliminated due to negative contribution. HERM reproduced the exact EPA CMB results with respect to source contribution S_{jk} ($r^2 = 1$, with the same means for corresponding sources). The minor differences, much smaller than the calculated source contribution uncertainty $\sigma_{S_{jk}}$, are attributed to the numerical precision of calculations, resulting in residual-to-uncertainty ratios (R/U ratios, model-versus-model) that are << 1 (Table 1). Source apportionment by HERM (or EPA CMB) captures the variations of actual source contributions well ($r^2 > 0.96$) and on average deviates from the true breakdowns by <2%. R/U ratios calculated from the difference in actual and modeled S_{jk} as well as modeled $\sigma_{S_{jk}}$ for individual samples are distributed roughly around unity, suggesting a reasonable estimate of source contribution uncertainties. However, the median R/U ratio (actual-versus-model) is 0.92 and 0.71 for HERM and EPA CMB, respectively, compared to the expected value of 1 (see supporting information Figure S2).

The eligible space dimension, i.e., the maximum number of sources that are estimable in the EV-CMB model, according to Henry[41] and calculated by EPA CMB[11] is always 5 (Table S4). Estimable sources have a contribution uncertainty <20% of $PM_{2.5}$ concentration (a predefined threshold), and when all the sources are estimable, as in this case, it corroborates no collinearity among the source profiles.

Table 2 shows the comparison for "Scenario B" with a varying degree of collinearity between U-Dust and R-Dust. For median-to-high collinearity, the eligible space dimension is reduced from 5 to 4 (Table S4), confirming similarity between at least two source profiles in the model. U-Dust and R-Dust are classified as inestimable (collinear) sources as they have small projections (<0.95) within the eligible space.[41] This means uncertainties associated with the U-Dust and R-Dust contributions would be above the threshold.

HERM reproduced EPA CMB results in the cases of low and median collinearity, though for some samples (3 in the low collinearity and 26 in the median collinearity case) U-Dust or R-Dust was eliminated by EPA CMB due to negative contributions. HERM attributed zero contributions to all the sources eliminated by EPA CMB and provided uncertainty estimates. For the three non-collinear sources, AMSUL, BB, and MV, both HERM and EPA CMB yielded expected source contributions. EPA CMB, however, appears to overestimate the source contribution uncertainty, as most of the actual-versus-model R/U ratios it reports are less than 0.5. HERM reports smaller, and more reasonable, uncertainties. Source apportionment between the two collinear sources, U-Dust and R-Dust, are not as accurate, as r^2 decreases to 0.7 – 0.9 and 0.2 – 0.3 in the low and median collinearity case, respectively, when compared with the actual source contributions (Table 2). The discrepancy is also reflected in the relatively large source contribution uncertainties from

HERM. Even in the median collinearity case, the median R/U ratio (actual-versus-model) for the two collinear sources remains at 0.83 from HERM, much closer to 1 in comparison with 0.40 from EPA CMB.

When collinearity is even higher, HERM starts to report source contributions that deviate from those of EPA CMB, and EPA CMB starts to report non-convergence in which no source contribution would be determined (see the high collinearity case in Table 2). Both HERM and EPA CMB fail to partition contributions from collinear sources, though HERM continues to report source contributions and uncertainties for all the samples, yielding a median actual-versus-model R/U ratio of 0.33 (or 0.24 for the two collinear sources). In practice, large uncertainties (i.e., $\sigma_{S_{jk}}$) alert users the potential collinearity in the model. The R/U ratio distributions in this case show that EPA CMB overestimates source contribution uncertainties more than HERM for the 3 non-collinear sources but underestimates source contribution uncertainties severely for the two collinear sources causing most R/U ratios > 2.5.

HERM was applied to 226 BSP samples acquired 2008-2009, using the same 9 source profiles combination as prior EPA CMB analysis (Table 3). This leads to an overall χ^2 of 1.8 (χ_k^2 : 0.46 – 20; r_k^2 : 0.43 – 0.98). The 9 sources explained 87% of measured $PM_{2.5}$. EPA CMB reported 12 non-convergent samples and eliminated a number of sources due to negative contributions. The eligible space dimension ranges from 6 to 9 (Table S4), and so collinearity does occur in some of the samples. Specifically, BBh and ADust have the most small projections in the eligible space, likely due to their collinearity with BBl and RDust2, respectively.

Other than the non-convergent samples and a few exceptions (with R/U ratio > 0.5, model-versus-model), HERM reproduced the EPA CMB source apportionment for the BSP dataset (Table 3). The exceptions for AMSUL, MV, and Coal are attributed to a single outlier

(5/16/2009), which also explains the low correlation ($r^2 = 0.43$) between the HERM- and EPA CMB-calculated Coal combustion contributions. Removing the outlier improves r^2 to 1.0 (see Figure S2). The 5/16/2009 sample features the highest calcium (Ca) concentration in the dataset that may introduce collinearity between the Coal and Asian dust source profiles, both of which contains an elevated Ca fraction (6.5% for Coal and 4.0% for Asian dust). In fact, collinearity resulted in one of the three dust sources being eliminated by EPA CMB for many samples. HERM avoided non-convergence and reported source contributions for every sample. It also shows relatively large uncertainties associated with the road dust contributions (Figure S4).

A scatter plot of χ_i^2 versus r_i^2 is used to evaluate HERM's fitting performance (Figure 1). Most of the species in the simulated Scenario A dataset are fitted well with $r_i^2 > 0.95$ taking into account the effective variance (Eq. [10]). Exceptions include 10 elements and 2 organic markers (Figure 1a). However, none of the species show $\chi_i^2 > 1$, suggesting that they contribute little to the overall χ^2 due to relatively large uncertainty (i.e., low signal-to-noise ratio) of the species in the source profiles, ambient measurements, or both. In the case of real-world BSP dataset, however, a few species that are not fitted well by the current HERM 9-source model, such as Zn, Ni, Pb and Br show $r_i^2 < 0.8$ and $\chi_i^2 >> 1$ (Figure 1b). There are therefore "real" discrepancies between the measured and modeled concentrations. This alerts users that different source profiles and/or additional sources may be needed in the model to explain variations of these species.

Application of HERM for unknown sources

In real-world applications, representative source profiles may not be available for all the sources that contribute to ambient $PM_{2.5}$, and HERM is better run in the hybrid mode. For our

Scenario A, AMSUL and AMNIT are hypothetical profiles for secondary ammonium salts formed in the atmosphere and U-Dust can be acquired for regions of interest at a relatively low cost through resuspension.[30,42] On the other hand, MV and BB source profiles likely result from other studies and deviate from the actual emissions that impact the receptor site. It is logical to specify only AMSUL, AMNIT, and U-Dust in the source apportionment by HERM, and let the model calculate other source profiles. The first trials include the three specified source profiles (and their uncertainties) as well as 0 to 4 unspecified source profiles, for a total of 3 – 7 sources in the HERM analysis. Figure 2 shows that χ^2 decreases substantially from 3 to 5 sources and levels off thereafter. This indicates that 5 sources sufficiently explain the variability in the dataset, as expected. In practice, such tests alerts users to focus on a 5-source model.

Four different conditions were examined under a 5-source model: 1) 3 sources specified (AMSUL, AMNIT, and U-Dust); 2) 4 sources specified (AMSUL, AMNIT, MV, and U-Dust); 3) no sources specified; and 4) no sources specified by EPA PMF 5.0 (Table 4). HERM was used for the first 3 conditions. When missing only the BB profile, HERM was able to report source contribution estimates as accurate as HERM or EPA CMB using all 5 source profiles ($r^2 > 0.97$, with χ^2 of 0.12 and a median actual-versus-model R/U ratio of 1.1). When the MV profile was also removed, the model still predicted BB well but underestimated the U-Dust contribution significantly ($r^2 = 0.54$). The R/U ratios, particularly for U-Dust, increased substantially leading to a median value of 2.2 (9.2 for U-Dust). Therefore, the discrepancy, resulted from the strong correlation between the MV and U-Dust contributions, is not captured in the source contribution uncertainty estimates. A few crustal elements (e.g., Al, Si, Ca, and Fe) are mixed into the calculated MV source profile (Figure S5); this confirms the challenge for receptor model to separate correlated sources without specific source profiles.

The two conditions without any source profile inputs generally failed to yield accurate source contribution estimates (Table 4). EPA PMF underestimated BB and U-Dust contributions while overestimating the others for which the actual and modeled source contributions remain highly correlated ($r^2 > 0.93$). The HERM source apportionment differ from that of EPA PMF, likely due to different ME-2 settings (e.g., nonrobust versus robust). Other causes of the difference are explained in Kim and Hopke[43]. All corresponding source contributions between the two models show strong correlations ($r^2 > 0.91$), and the median model-versus-model R/U ratio is 2.0, lower than their median actual-versus-model R/U ratios (HERM: 6.6; EPA PMF: 6.0). Generally, they agree with each other better than with the actual source contributions.

Improvement of source apportionment with HERM

Source apportionment results can usually be improved with additional information that serve as constraints to a receptor model. Even if the full source profile is unavailable, it is possible to introduce to the prior knowledge that MV (tailpipe) emissions contain little crustal elements, such as silicon (Si) and Ca, into the HERM modeling. This was done by specifying an incomplete source profile with only two zero elements (Si and Ca), along with three full source profiles (AMSUL, AMNIT, and U-Dust), in the HERM input file to establish a 5-source model for the Scenario A dataset (Table 5). The resulting MV and U-Dust contributions agree with actual values better ($r^2 > 0.98$) than those acquired previously using only the three full source profiles. The median actual-versus-model R/U ratio drops from 2.2 to 1.3 while the overall χ^2 increases little from 0.093 to 0.12. HERM also closely reproduces the expected MV source profile (Figure S5). This example illustrates how additional source information help separate correlated sources.

In the previous BSP $PM_{2.5}$ source apportionment, the road and Asian dust source profiles were developed locally[18] and, along with AMSUL and AMNIT, can be representative of corresponding sources or atmospheric processes. The MV profile that is a composite from dynamometer testing[44] should represent tailpipe emissions of a modern fleet (low-emitting gasoline vehicles). On the other hand, the BB and Coal profiles are more uncertain. Wildfire smoke impacts BSP from time to time, for which source profile may substantially differ from BBh and BBl acquired from a much smaller scale laboratory combustion.[45] Since there are not industrial sources in the Lake Tahoe Basin, the "Coal" contributions must originate from long-range transport and chemically resemble mixed industrial emissions. Figure 3 shows the dependence of χ^2 on the number of sources when the first 4 sources (AMSUL, AMNIT, RDust2, ADust) are specified in HERM. Though it is not as obvious as Figure 2, the trend suggests 6 or 7 sources to be the most appropriate. Thus the three least contributing sources in Table 3, i.e., BBh, RDust1, and/or Coal, may be merged with other sources.

The 6- and 7-source models were constructed by HERM (Table 6), and these models all appeared robust as χ^2 varied little in repeated runs. Based on correlations with the prior model results, the two additional sources in the 6-source model were identified as BB ($r^2 = 0.97$) and MV ($r^2 = 080$). However, industrial markers such as As, Br, Pb, Se, Zn, and S show higher than expected fractions in the derived "MV" profile, suggesting its coupling with mixed industrial emissions (noted by "MV + Ind." in Table 6). A 7-source model with 3 unspecified sources could not separate them, possibly due to some correlation and/or collinearity between the two. When adding the default MV profile in the model input (i.e., 5 specified plus 2 unspecified sources), however, HERM was able to separate motor vehicle and industrial contributions. Table 6 compares source apportionment by the HERM 6-source (4+2), HERM 7-source (5+2), and EV-

CMB 9-source (from Table 3) models. For the 4 pre-specified sources and calculated BB, the HERM 6- and 7-source models estimate essentially the same contributions considering the reported uncertainty (median R/U ratio < 0.2). With the input of MV source profile, the 7-source model distinguishes the MV contribution while achieving a better fit (i.e., lower χ^2). Unlike EV-CMB which underestimates $PM_{2.5}$ mass, both HERM models explain $PM_{2.5}$ mass within 2% by allowing part of the profiles to vary. The hybrid models attribute more mass to BB and transported industrial emissions but less mass to AMSUL and MV. Particularly, MV fraction in $PM_{2.5}$ is >11% by EV-CMB and only 2% by the HERM hybrid 7-source model. A concurrent emission inventory[46] supports the latter as basinwide onroad vehicles and recreational boats account for <2% the primary $PM_{2.5}$ emission. Unrepresentative biomass burning and industrial source profiles may have caused EV-CMB to overestimate the MV contribution.

The derived BB source profile is similar to BBl where OC, EC, and K dominate (Figure S6) but with higher EC/OC (0.12 vs. 0.047) and lower K/OC ratios (0.011 versus 0.014). Sulfur is the most enriched species in both the derived industrial and Coal source profiles (Figure S6), though the Se/S ratio differs significantly between the two (0.00052 vs. 0.016). A low ratio typically means substantial aging, and one should note that the ambient Se/S ratio never exceeded 0.001 and averaged only 0.00014 over the entire period. Moreover, industrial elements including Br, Zn, and Pb are more enriched in the derived industrial than in the measured Coal source profile; this results in them being fitted better (higher r_i^2 and lower χ_i^2) by the hybrid 7-source model (Figure 4) than by EV-CMB with 9 sources (Figure 1), at a small cost to the K and Se fittings. The fitting for Ni and EC also improves. In general, the hybrid model explains well the variations of species in the BSP dataset.

DISCUSSION AND RECOMMENDATION

Receptor model is an important tool for air quality management. Since none of the modeling approaches is without biases or uncertainties, a weight-of-evidence (WOE) approach that takes into account multiple model results is strongly recommended in practice.[19,23,24,39] This paper introduces the hybrid environmental receptor model (HERM) that can perform EV-CMB and PMF, two most popular receptor models for $PM_{2.5}$ source apportionment, using a unified algorithm and evaluates it with simulated and real-world datasets. In the EV-CMB mode, where all source profiles/uncertainties are specified, HERM is shown to yield source attributions nearly identical to EPA CMB v8.2 but with 1) more tolerance to collinearity and 2) better estimate of source contribution uncertainty even when collinearity occurs. In the PMF mode where no source information is used, HERM and EPA PMF 5.0 source contributions are highly correlated but not the same due to different modeling preferences (e.g., non-robust versus robust).

HERM allows a hybrid mode that takes partial source information such as incomplete source profiles to pursue a middle ground between EV-CMB and PMF. This is particularly useful since the inclusion of only reliable source profiles in the model avoids poor fitting in EV-CMB while reducing the rotational degree of freedom in PMF analysis. HERM implements the constraints differently from EPA PMF in that it uses source profile uncertainties explicitly in the effective variance fitting. Preliminary tests show that partial information improves source apportionment. It could help separate sources of which contributions are highly correlated thus presenting a major challenge to PMF. It also calculates source profiles that are more representative of the study region than profiles acquired from somewhere else.

More tests are warranted to determine how the best performance of HERM may be achieved with different datasets and also how the robust mode, if implemented, will alter the source apportionment in the EV-CMB or hybrid mode. The convenience of the model's user interface will facilitate the investigation, as it allows all input and output parameters in a single MS Excel® file for easier data processing and comparison. In addition to source contribution and uncertainty values, HERM calculates reduced chi^2 (χ^2) to inform users the overall goodness of fit, χ_k^2 and r_k^2 to assess sample-specific fits, and χ_i^2 and r_i^2 to assess species-specific fits. This helps identify outliers for potential removal from the model. When practicing receptor modeling, users are recommended to first determine the possible number(s) of sources (J) by examining the dependence of χ^2 on J. HERM in different modes (EV-CMB, hybrid, and PMF) using non-robust and robust calculations should be carried out with their results compared and reconciled to support the WOE approach of source apportionment.

ACKNOWLEDGEMENT

This project was supported by a joint venture (#2015003P) between Institute of Earth Environment, Chinese Academy of Sciences and University of Nevada, Las Vegas. Additional support was provided by the National Research Program for Key Issues in Air Pollution Control. The authors thank Dr. Shally Pang at Xi'an University of Posts & Telecommunications for assisting the software development. Comments from three anonymous reviewers are highly appreciated.

Supporting Information

Figures showing the model interface and various performance measures for source contribution/profile estimates, and tables documenting model formulation, source profiles used in this study as well as a collinearity diagnosis for these source profiles.

REFERENCES

1. Watson, J.G., Chow, J.C. and Fujita, E.M., Review of volatile organic compound source apportionment by chemical mass balance. *Atmos. Environ.* **2001**, *35*(9), 1567-1584.

2. Cao, J.J., Wu, F., Chow, J.C., Lee, S.C., Li, Y., Chen, S.W., An, Z.S., Fung, K.K., Watson, J.G., Zhu, C.S. and Liu, S.X., Characterization and source apportionment of atmospheric organic and elemental carbon during fall and winter of 2003 in Xi'an, China. *Atmos. Chem. Phys.* **2005**, *5*(11), 3127-3137.

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

4. Viana, M., Kuhlbusch, T.A.J., Querol, X., Alastuey, A., Harrison, R.M., Hopke, P.K., Winiwarter, W., Vallius, M., Szidat, S., Prévôt, A.S.H. and Hueglin, C., Source apportionment of particulate matter in Europe: A review of methods and results. *J. Aerosol Sci.* **2008**, *39*(10), 827-849.

5. Huang, R.J., Zhang, Y., Bozzetti, C., Ho, K.F., Cao, J.J., Han, Y., Daellenbach, K.R., Slowik, J.G., Platt, S.M., Canonaco, F. and Zotter, P., High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **2014**, *514*(7521), 218-222.

6. Henry, R.C., Lewis, C.W., Hopke, P.K. and Williamson, H.J., Review of receptor model fundamentals. *Atmos. Environ. (1967)* **1984**, *18*(8), 1507-1515.

7. Watson, J.G., Chen, L.-W.A., Chow, J.C., Doraiswamy, P. and Lowenthal, D.H., Source apportionment: findings from the US supersites program. *J. Air Waste Manage. Assoc.* **2008**, *58*(2), 265-288.

8. Watson, J.G., Cooper, J.A. and Huntzicker, J.J., The effective variance weighting for least squares calculations applied to the mass balance receptor model. *Atmos. Environ. (1967)* **1984**, *18*(7), 1347-1355.

9. Watson, J.G., Robinson, N.F., Chow, J.C., Henry, R.C., Kim, B.M., Pace, T.G., Meyer, E.L. and Nguyen, Q., The USEPA/DRI chemical mass balance receptor model, CMB 7.0. *Environ. Softw.* **1990**, *5*(1), 38-49.

10. Watson, J.G., Robinson, N.F., Lewis, C., Coulter, T., Chow, J.C., Fujita, E.M.,

Lowenthal, D.H., Conner, T.L., Henry, R.C. and Willis, R.D., *Chemical mass balance receptor model version 8 (CMB8) user's manual*. Prepared for US Environmental Protection Agency, Research Triangle Park, NC, by Desert Research Institute, Reno, NV, **1997**.

11. Coulter C.T., *EPA-CMB8.2 Users Manual*. Prepared for the Office of Air Quality Planning & Standards (OAQPS), U.S. Environmental Protection Agency, Research Triangle Park, NC. EPA-452/R-04-011, **2004**.
12. Paatero, P., The multilinear engine—a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J. Comput. Graph. Stat.* **1999**, *8*(4), 854-888.
13. Hopke, P.K., Review of receptor modeling methods for source apportionment. *J. Air Waste Manage. Assoc.* **2016**, *66*(3), 237-259.
14. Norris, G.A., Duvall, R., Brown, S.G. and Bai, S., *EPA Positive Matrix Factorization (PMF) 5.0 fundamentals and User Guide.* Prepared for the US Environmental Protection Agency Office of Research and Development, Washington, DC. EPA/600/R-14/108, **2014**.
15. 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.* **2007**, *57*(2), 146-154.
16. Chen, L.-W.A., Watson, J.G., Chow, J.C. and Magliano, K.L., Quantifying $PM_{2.5}$ source contributions for the San Joaquin Valley with multivariate receptor models. *Environ. Sci. Tech.* **2007**, *41*(8), 2818-2826.
17. Chen, L.-W.A., Watson, J.G., Chow, J.C., DuBois, D.W. and Herschberger, L., $PM_{2.5}$ source apportionment: reconciling receptor models for US nonurban and urban long-term networks. *J. Air Waste Manage. Assoc.* **2011**, *61*(11), 1204-1217.
18. Green, M.C., Chen, L.-W.A., DuBois, D.W. and Molenar, J.V., Fine particulate matter and visibility in the Lake Tahoe Basin: Chemical characterization, trends, and source apportionment. *J. Air Waste Manage. Assoc.* **2012**, *62*(8), 953-965.
19. Lee, S., Liu, W., Wang, Y., Russell, A.G. and Edgerton, E.S., Source apportionment of $PM_{2.5}$: Comparing PMF and CMB results for four ambient monitoring sites in the southeastern United States. *Atmos. Environ.* **2008**, *42*(18), 4126-4137.

20. Favez, O., Haddad, I.E., Piot, C., Boréave, A., Abidi, E., Marchand, N., Jaffrezo, J.L., Besombes, J.L., Personnaz, M.B., Sciare, J. and Wortham, H., Inter-comparison of source apportionment models for the estimation of wood burning aerosols during wintertime in an Alpine city (Grenoble, France). *Atmos. Chem. Phys.* **2010**, *10*(12), 5295-5314.

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

22. Cesari, D., Donateo, A., Conte, M. and Contini, D., Inter-comparison of source apportionment of PM_{10} using PMF and CMB in three sites nearby an industrial area in central Italy. *Atmos. Res.* **2016**, *182*, 282-293.

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

24. Taiwo, A.M., Harrison, R.M. and Shi, Z., A review of receptor modelling of industrially emitted particulate matter. *Atmos. Environ.* **2014**, *97*, 109-120.

25. Shi, G.L., Feng, Y.C., Zeng, F., Li, X., Zhang, Y.F., Wang, Y.Q. and Zhu, T., Use of a nonnegative constrained principal component regression chemical mass balance model to study the contributions of nearly collinear sources. *Environ. Sci. Tech.* **2009**, *43*(23), 8867-8873.

26. Liu, G.R., Shi, G.L., Tian, Y.Z., Wang, Y.N., Zhang, C.Y. and Feng, Y.C., Physically constrained source apportionment (PCSA) for polycyclic aromatic hydrocarbon using the Multilinear Engine 2-species ratios (ME2-SR) method. *Sci. Total Environ.* **2015**, *502*, 16-21.

27. Sofowote, U.M., Su, Y., Dabek-Zlotorzynska, E., Rastogi, A.K., Brook, J. and Hopke, P.K., Sources and temporal variations of constrained PMF factors obtained from multiple-year receptor modeling of ambient $PM_{2.5}$ data from five speciation sites in Ontario, Canada. *Atmos. Environ.* **2015**, *108*, 140-150.

28. Shi, G.L., Zhou, X.Y., Feng, Y.C., Tian, Y.Z., Liu, G.R., Zheng, M., Zhou, Y. and Zhang, Y.H., An improved estimate of uncertainty for source contribution from effective variance Chemical Mass Balance (EV-CMB) analysis. *Atmos. Environ.* **2015**, *100*, 154-158.

29. Chen, L.-W.A., Watson, J.G., Chow, J.C., DuBois, D.W. and Herschberger, L., Chemical mass balance source apportionment for combined $PM_{2.5}$ measurements from US non-urban and urban long-term networks. *Atmos. Environ.* **2010**, *44*(38), 4908-4918.
30. Chow, J.C., Watson, J.G., Ashbaugh, L.L. and Magliano, K.L., Similarities and differences in PM_{10} chemical source profiles for geological dust from the San Joaquin Valley, California. *Atmos. Environ.* **2003**, *37*(9), 1317-1340.
31. Chow, J.C., Watson, J.G., Kuhns, H., Etyemezian, V., Lowenthal, D.H., Crow, D., Kohl, S.D., Engelbrecht, J.P. and Green, M.C., Source profiles for industrial, mobile, and area sources in the Big Bend Regional Aerosol Visibility and Observational study. *Chemosphere* **2004**, *54*(2), 185-208.
32. Chen, L.-W.A., Watson, J.G., Chow, J.C., Green, M.C., Inouye, D. and Dick, K., Wintertime particulate pollution episodes in an urban valley of the Western US: a case study. *Atmos. Chem. Phys.* **2012**, *12*(21), 10051-10064.
33. Chow, J.C., Watson, J.G., Chen, L.-W.A., Chang, M.O., Robinson, N.F., Trimble, D. and Kohl, S., The IMPROVE_A temperature protocol for thermal/optical carbon analysis: maintaining consistency with a long-term database. *J. Air Waste Manage. Assoc.* **2007**, *57*(9), 1014-1023.
34. Chow, J.C., Watson, J.G., Chen, L.-W.A. and Divita, F., *Contemporary inorganic and organic speciated particulate matter source profiles for geological material, motor vehicles, vegetative burning, industrial boilers, and residential cooking*. Prepared for Pechan and Associates, Inc., Springfield, VA, by Desert Research Institute, Reno, NV, **2006**.
35. Lowenthal, D.H., Chow, J.C., Watson, J.G., Neuroth, G.R., Robbins, R.B., Shafritz, B.P. and Countess, R.J., The effects of collinearity on the ability to determine aerosol contributions from diesel-and gasoline-powered vehicles using the chemical mass balance model. *Atmos. Environ. A-Gen.* **1992**, *26*(13), 2341-2351.
36. Murphy, D.M., Chow, J.C., Leibensperger, E.M., Malm, W.C., Pitchford, M., Schichtel, B.A., Watson, J.G. and White, W.H., Decreases in elemental carbon and fine particle mass in the United States. *Atmos. Chem. Phys.* **2011**, *11*(10), 4679-4686.
37. Chen, L.-W.A., Chow, J.C., Watson, J.G. and Schichtel, B.A., Consistency of long-term elemental carbon trends from thermal and optical measurements in the IMPROVE

network. *Atmos. Meas. Tech.* **2012**, *5*(10), 2329-2338.

38. Watson, J.G., Zhu, T., Chow, J.C., Engelbrecht, J., Fujita, E.M. and Wilson, W.E., Receptor modeling application framework for particle source apportionment. *Chemosphere* **2002**, *49*(9), 093-1136.

39. Watson, J.G., Chow, J.C., Chen, L.-W.A., Engling, G. and Wang, X.L., Source apportionment: principles and methods. In *Airborne Particulate Matter* (Ed. Roy Harrison, pp. 72-125). Royal Society of Chemistry, **2016**.

40. Malamakal, T., Chen, L.-W.A., Wang, X., Green, M.C., Gronstal, S., Chow, J.C. and Watson, J.G., Prescribed burn smoke impact in the Lake Tahoe Basin: model simulation and field verification. *Int. J. Environ. Pollut.* **2013**, *52*(3-4), 225-243.

41. Henry, R.C., Dealing with near collinearity in chemical mass balance receptor models. *Atmos. Environ. A-Gen.* **1992**, *26*(5), 933-938.

42. Chow, J.C., Watson, J.G., Houck, J.E., Pritchett, L.C., Rogers, C.F., Frazier, C.A., Egami, R.T. and Ball, B.M., A laboratory resuspension chamber to measure fugitive dust size distributions and chemical compositions. *Atmos. Environ.* **1994**, *28*(21), 3463-3481.

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

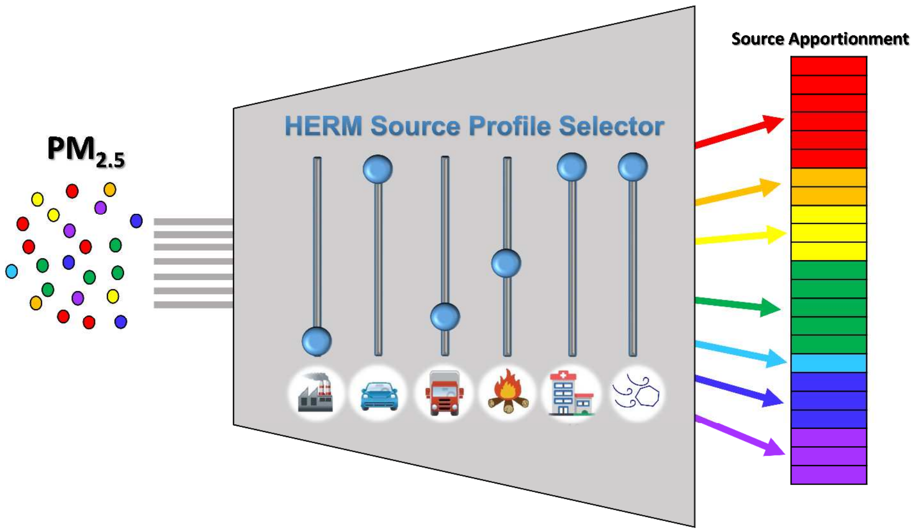
44. Fujita, E.M., Zielinska, B., Campbell, D.E., Arnott, W.P., Sagebiel, J.C., Mazzoleni, L., Chow, J.C., Gabele, P.A., Crews, W., Snow, R. and Clark, N.N., Variations in speciated emissions from spark-ignition and compression-ignition motor vehicles in California's south coast air basin. *J. Air Waste Manage. Assoc.* **2007**, *57*(6), 705-720.

45. Chen, L.W., Verburg, P., Shackelford, A., Zhu, D., Susfalk, R., Chow, J.C. and Watson, J.G., Moisture effects on carbon and nitrogen emission from burning of wildland biomass. *Atmos. Chem. Phys.* **2010**, *10*(14), 6617-6625.

46. Kuhns, H.D., Chang, M.C.O., Chow, J.C., Etyemezian, V., Chen, L.W.A., Nussbaum, N.J., Nathagoundenpalayam, S.K., Trimble, T.C., Kohl, S.D., MacLaren, M. and Abu-Allaban, M., *DRI Lake Tahoe source characterization study*. Prepared for California Air Resources Board, Sacramento, CA, by Desert Research Institute, Reno, NV, **2004**.

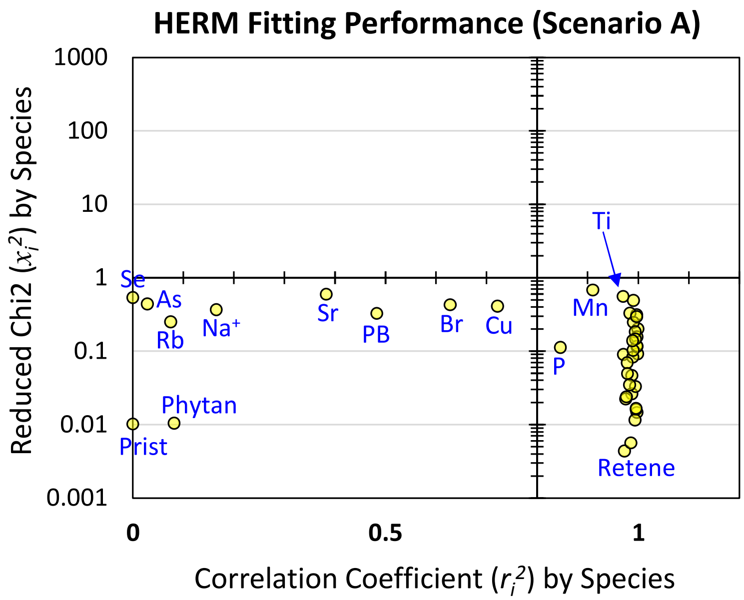
645 **Table of content (TOC) graph**

646

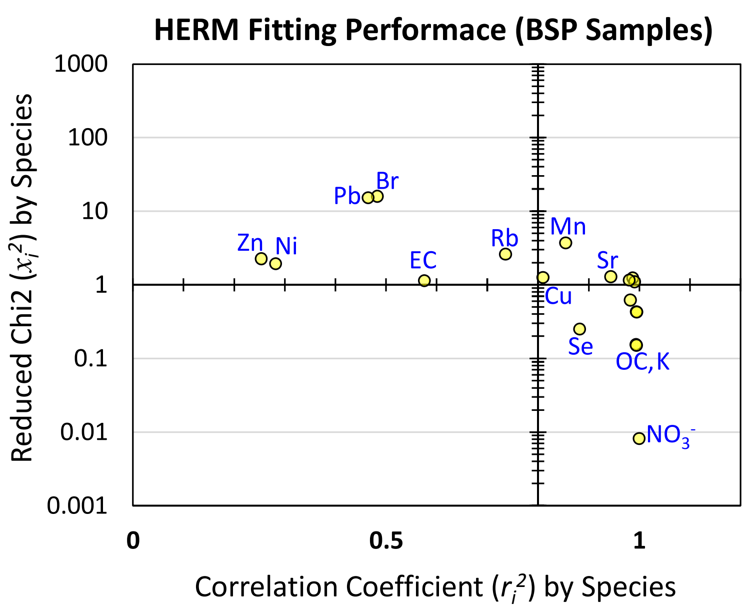


647

648



(a)



(b)

Figure 1. HERM fitting performance examined by species-specific residual (χ_i^2) and correlation coefficient (r_i^2) for the (a) simulated Scenario A (b) BSP dataset (EV-CMB mode, see Table 1 and 3). Species noted in blue show relatively extreme χ_i^2 and/or r_i^2.

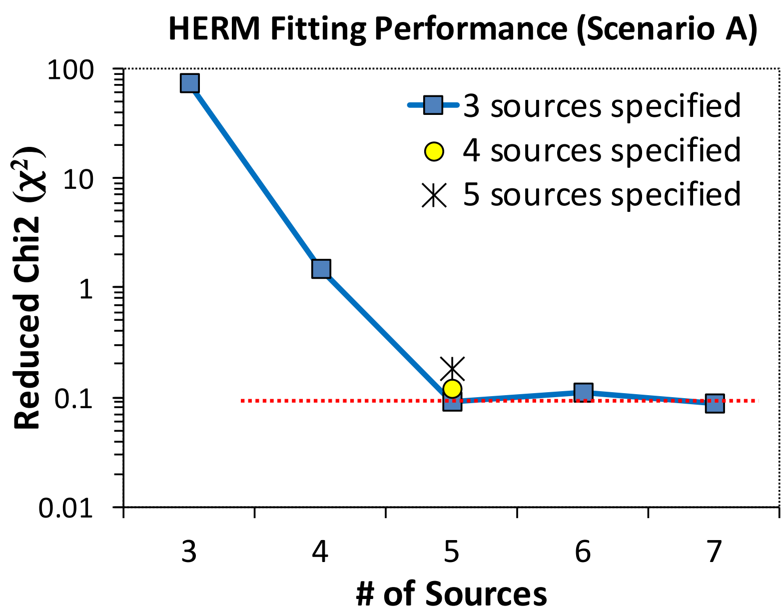


Figure 2. HERM fitting performance for the Scenario A dataset examined by the overall residual (χ^2) as a function of the total number of sources and number of sources specified in the model. AMSUL, AMNIT, and U-Dust are among the 3 sources specified. Additionally, MV is included in the "4 or 5 sources specified" and BB is included in the "5 sources specified".

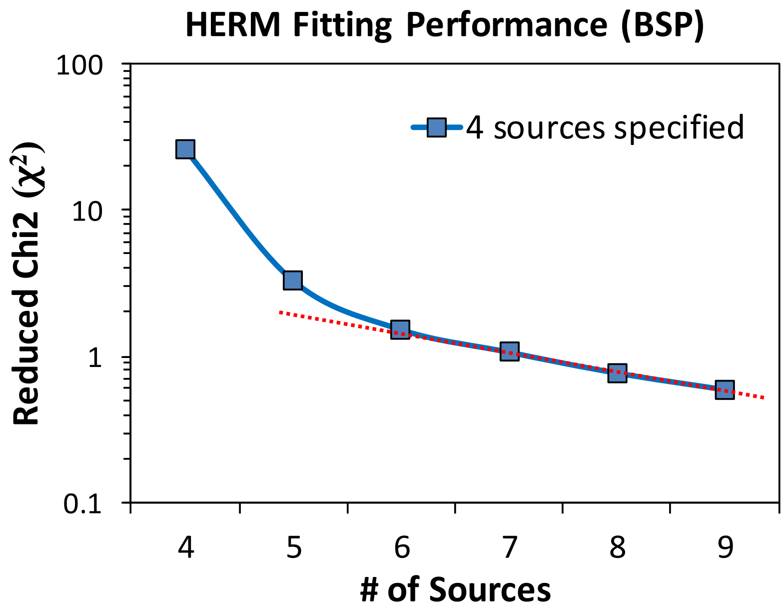
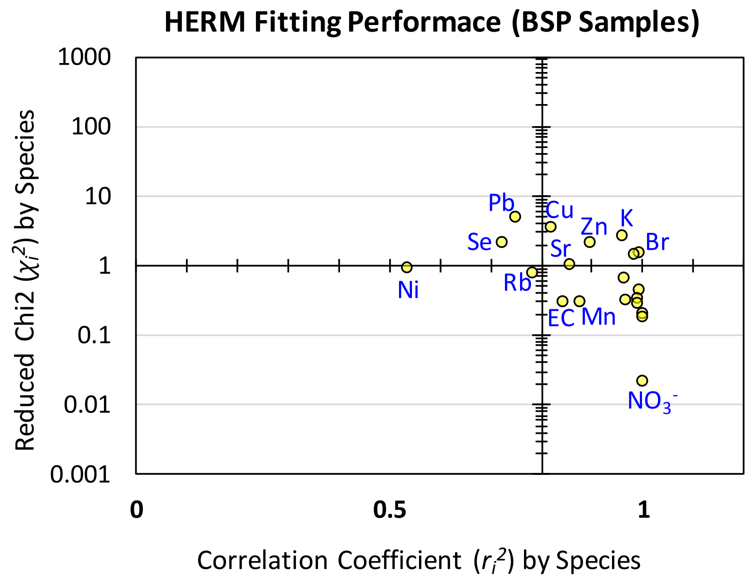


Figure 3. HERM fitting performance for the BSP (2008-2009) dataset examined by the overall residual (χ^2) as a function of the total number of sources when 4 sources, AMSUL, AMNIT, RDust2, and ADust, have been specified.



664

665 **Figure 4.** HERM fitting performance examined by species-specific residual (χ_i^2) and correlation

666 coefficient (r_i^2) for the BSP 2008-2009 dataset (hybrid 7-source model, see Table 6). Species

667 noted in blue show relatively extreme χ_i^2 and/or r_i^2 .

Table 1. Source apportionment of simulated $PM_{2.5}$ speciation dataset (Scenario A) by CAS HERM and EPA CMB, compared with the actual source contributions.

	Samples	Mean Contribution[*] (μg m^{-3})			Correlation (r^2)			R/U Ratio[†] (<0.5/0.5-1.5/1.5-2.5/>2.5)												Source Eliminated[‡]		Noncon-vergence[□]	
Source(s)	#	Actual (x)	HERM (y)	CMB (z)	x vs y	x vs z	y vs z	x vs y				x vs z				y vs z				HERM	CMB	HERM	CMB
AMSUL	50	2.591	2.599	2.599	0.983	0.983	1.000	19	**21**	7	3	24	**20**	4	2	50	**0**	0	0	0	0	0	0
AMNIT	50	4.817	4.757	4.757	0.988	0.988	1.000	12	**20**	11	7	19	**23**	7	1	50	**0**	0	0	0	0	0	0
BB	50	4.866	4.777	4.777	0.963	0.963	1.000	17	**19**	12	2	16	**16**	10	8	50	**0**	0	0	0	0	0	0
MV	50	2.423	2.459	2.460	0.979	0.980	1.000	23	**22**	4	1	17	**18**	6	9	50	**0**	0	0	0	0	0	0
U-Dust	50	2.330	2.313	2.314	0.983	0.983	1.000	12	**17**	11	10	21	**25**	3	1	50	**0**	0	0	0	0	0	0
Sum		17.026	16.906	16.906	HERM: x^2 = 0.182																		

[*]Actual source contribution (S_{jk}) and those derived by HERM and EPA CMB models are noted as x, y, and z, respectively. Mean values take into account all available data.

[†]Residue-Uncertainty (R/U) ratio of x and y is calculated by $|y-x|/\sigma_y$ where σ_y is the source contribution uncertainty estimated by HERM. The ratios are then categorized into 4 ranges: <0.5, 0.5-1.5, 1.5-2.5, and >2.5 with numbers in each range shown in the table. Similarly, R/U ratio of x and z is calculated by $|z-x|/\sigma_z$ where σ_z is the source contribution uncertainty estimated by EPA CMB. R/U ratio of y and z is calculated by $|y-z|/(\sigma_y^2 + \sigma_z^2)^{½}$.

[‡]Number of source eliminated due to negative source contribution. When occurring, no uncertainty estimate is provided by EPA CMB.

[□]Number of non-convergence due to collinearity. When occurring, no uncertainty estimate is provided by HERM or EPA CMB.

Table 2. Source apportionment of simulated $PM_{2.5}$ speciation dataset (Scenario B) by CAS HERM and EPA CMB, compared with the actual source contributions.

	Samples	Mean Contribution* ($\mu g\ m^{-3}$)			Correlation (r^2)			R/U Ratio† (<0.5/0.5-1.5/1.5-2.5/>2.5)												Source Eliminated‡		Noncon-vergence□	
Source(s)	#	Actual (x)	HERM (y)	CMB (z)	x vs y	x vs z	y vs z	x vs y				x vs z				y vs z				HERM	CMB	HERM	CMB
Low collinearity between U-Dust and R-Dust (α = 0.9)																							
AMSUL	50	2.404	2.418	2.418	0.992	0.992	1.000	24	**19**	7	0	44	**6**	0	0	50	**0**	0	0	0	0	0	0
BB	50	4.758	4.831	4.831	0.985	0.985	1.000	26	**21**	3	0	45	**5**	0	0	50	**0**	0	0	0	0	0	0
MV	50	4.899	4.887	4.887	0.993	0.993	1.000	25	**19**	6	0	44	**6**	0	0	50	**0**	0	0	0	0	0	0
U-Dust	50	2.479	2.256	2.257	0.738	0.738	1.000	8	**14**	15	13	19	**26**	2	0	47	**0**	0	0	0	3	0	0
R-Dust	50	2.488	2.597	2.597	0.892	0.892	1.000	11	**18**	12	9	26	**21**	2	1	50	**0**	0	0	0	0	0	0
Sum		17.029	16.990	16.990	HERM: x^2 = 0.151																		
Median collinearity between U-Dust and R-Dust (α = 0.99)																							
AMSUL	50	2.404	2.418	2.418	0.990	0.990	1.000	16	**25**	9	0	42	**8**	0	0	50	**0**	0	0	0	0	0	0
BB	50	4.758	4.900	4.900	0.979	0.979	1.000	19	**27**	4	0	48	**1**	1	0	50	**0**	0	0	0	0	0	0
MV	50	4.899	4.820	4.820	0.989	0.989	1.000	26	**18**	6	0	45	**4**	1	0	50	**0**	0	0	0	0	0	0
U-Dust	50	2.479	2.589	2.588	0.179	0.179	1.000	15	**26**	8	1	22	**4**	4	8	38	**0**	0	0	0	12	0	0
R-Dust	50	2.488	2.337	2.338	0.276	0.277	1.000	13	**28**	8	1	23	**3**	1	9	36	**0**	0	0	0	14	0	0
Sum		17.029	17.063	17.063	HERM: x^2 = 0.171																		
High collinearity between U-Dust and R-Dust (α = 0.998)																							
AMSUL	50	2.404	2.401	2.444	0.985	0.985	1.000	24	**18**	5	3	39	**9**	1	0	49	**0**	0	0	0	0	0	1
BB	50	4.758	4.769	4.699	0.987	0.987	1.000	26	**21**	3	0	45	**4**	0	0	49	**0**	0	0	0	0	0	1
MV	50	4.899	4.896	4.864	0.996	0.996	1.000	30	**18**	2	0	46	**3**	0	0	49	**0**	0	0	0	0	0	1
U-Dust	50	2.479	1.739	1.656	0.029	0.017	0.997	47	**3**	0	0	7	**1**	2	10	20	**0**	0	0	0	29	0	1
R-Dust	50	2.488	3.165	3.208	0.051	0.055	0.997	48	**2**	0	0	8	**3**	1	24	36	**0**	0	0	0	13	0	1
Sum		17.029	16.970	16.871	HERM: x^2 = 0.158																		

*†‡□ See footnotes in Table 1.

Table 3. Source apportionment of ambient $PM_{2.5}$ speciation dataset (BSP 2008-2009) by CAS HERM and EPA CMB.

	Samples	Mean Contribution* ($\mu g\ m^{-3}$)			Correlation (r^2)	R/U Ratio† (<0.5/0.5-1.5/1.5-2.5/>2.5)				Source Eliminated‡		Noncon-vergence□	
Source(s)	#	Actual	HERM (y)	CMB (z)	y vs z	y vs z				HERM	CMB	HERM	CMB
AMSUL	226		0.555	0.553	1.000	213	**1**	0	0	0	0	0	12
AMNIT	226		0.161	0.158	1.000	210	**0**	0	0	0	4	0	12
RDust1	226		0.005	0.005	0.998	109	**0**	0	0	0	105	0	12
RDust2	226		0.123	0.122	0.991	134	**3**	0	0	0	77	0	12
ADust	226		0.506	0.514	0.999	195	**0**	0	0	0	19	0	12
BBh	226		0.105	0.107	1.000	174	**0**	0	0	0	40	0	12
BBl	226		1.358	1.363	1.000	188	**0**	0	0	0	26	0	12
MV	226		0.419	0.422	0.996	212	**1**	0	0	0	1	0	12
Coal	226		0.029	0.025	0.427	195	**1**	0	0	0	18	0	12
Sum		3.760	3.261	3.269	HERM: x^2 = 1.81								

*Source contribution (S_{jk}) derived by HERM and EPA CMB models are noted as y and z, respectively. Mean values take into account all available data. The 9 sources include ammonium sulfate (AMSUL), ammonium nitrate (AMNIT), two road dusts (RDust1, RDust2), Asian dust (ADust), wood burning with both low and high combustion efficiencies (BBh and BBl), traffic (MV), and miscellaneous coal combustion (Coal).

†Residue-Uncertainty (R/U) ratio of y and z is calculated by $|y-z|/(\sigma_y^2 + \sigma_z^2)^{1/2}$, where σ_y and σ_z is the source contribution uncertainty estimated by HERM and EPA CMB, respectively. The ratios are then categorized into 4 ranges: <0.5, 0.5-1.5, 1.5-2.5, and >2.5 with numbers in each range shown in the table.

‡Number of source eliminated due to negative source contribution. When occurring, no uncertainty estimate is provided by EPA CMB.

□Number of non-convergence due to collinearity. When occurring, no uncertainty estimate is provided by HERM or EPA CMB.

Table 4. Source apportionment of simulated $PM_{2.5}$ speciation dataset (Scenario A) by CAS HERM and EPA PMF 5.0, compared with the actual source contributions.

	Samples	**Mean Contribution[*] (μg m^{-3})**			**Correlation (r^2)**			**R/U Ratio[†] (<0.5/0.5-1.5/1.5-2.5/>2.5)**												**Noncon-vergence[‡]**	
Source(s)	#	Actual (x)	HERM^{4+1} (y)	HERM^{3+2} (z)	x vs y	x vs z	y vs z	x vs y				x vs z				y vs z				HERM^{4+1}	HERM^{3+2}
AMSUL	50	2.591	2.610	2.808	0.982	0.975	0.993	19	**22**	7	2	5	**14**	9	22	10	**23**	14	3	0	0
AMNIT	50	4.817	4.748	4.828	0.986	0.986	1.000	13	**24**	10	3	5	**18**	17	10	33	**11**	4	2	0	0
BB	50	4.866	4.750	4.701	0.987	0.994	0.992	20	**15**	6	9	19	**25**	5	1	33	**17**	0	0	0	0
MV	50	2.423	2.513	3.393	0.977	0.996	0.968	8	**17**	10	15	3	**4**	12	31	4	**8**	10	28	0	0
U-Dust	50	2.330	2.213	1.124	0.982	0.543	0.552	11	**24**	10	5	0	**0**	2	48	3	**2**	6	39	0	0
Sum		17.026	16.834	16.854	HERM^{4+1}: χ^2 = 0.121; HERM^{3+2}: χ^2 = 0.093																
Source(s)	#	Actual (x)	HERM^{0+5} (y)	PMF (z)	x vs y	x vs z	y vs z	x vs y				x vs z				y vs z				HERM^{0+5}	PMF
AMSUL	50	2.591	2.474	3.368	0.976	0.937	0.976	10	**18**	7	15	2	**3**	5	40	1	**2**	6	41	0	0
AMNIT	50	4.817	6.765	5.989	0.976	0.955	0.973	4	**5**	2	39	3	**3**	4	40	7	**16**	4	23	0	0
BB	50	4.866	2.366	2.705	0.570	0.557	0.981	1	**1**	3	45	3	**3**	4	40	10	**19**	15	6	0	0
MV	50	2.423	3.963	3.186	0.995	0.958	0.962	4	**5**	3	38	3	**4**	2	41	7	**10**	5	28	0	0
U-Dust	50	2.330	1.310	1.594	0.098	0.207	0.915	0	**4**	2	44	0	**6**	6	38	13	**13**	15	9	0	0
Sum		17.026	16.878	16.842	HERM^{0+5}: χ^2 = 0.152; PMF: χ^2 = 0.161																

[*]Actual source contribution (S_{jk}) and those derived by HERM or EPA PMF models are noted as x, y, or z, respectively. Mean values take into account all available data. HERM^{4+1} specifies 4 source profiles (AMSUL, AMNIT, MV, and U-Dust) while calculating 1 source profile (BB). HERM^{3+2} specifies 3 source profiles (AMSUL, AMNIT, and U-Dust) while calculating 2 source profiles (BB and MV). HERM^{0+5} and PMF calculate all 5 profiles (non-specified). Calculated source profiles are matched to the known sources by ranking the correlation coefficients across source contributions.

[†][‡] See footnotes in Table 1.

Table 5. Source apportionment of simulated $PM_{2.5}$ speciation dataset (Scenario A) by CAS HERM, compared with the actual source contributions.

	Samples	**Mean Contribution*** **(μg m^{-3})**		**Correlation** **(r^2)**	**R/U Ratio†** **(<0.5/0.5-1.5/1.5-2.5/>2.5)**			
Source(s)	#	Actual (x)	$HERM^{3+2'}$ (y)	x vs y	x vs y			
AMSUL	50	2.591	2.806	0.975	5	**14**	10	21
AMNIT	50	4.817	4.825	0.986	8	**20**	13	9
BB	50	4.866	4.460	0.989	7	**25**	12	6
MV	50	2.423	2.294	0.996	16	**33**	1	0
U-Dust	50	2.330	2.470	0.987	7	**9**	14	20
Sum		17.026	16.855	$HERM^{3+2'}$: x^2 = 0.124				

*Actual source contribution (S_{jk}) and those derived by HERM are noted as x and y, respectively. Mean values take into account all available data. $HERM^{3+2'}$ specifies 3 source profiles (AMSUL, AMNIT, and U-Dust) while also specifying the silicon (Si) and calcium (Ca) contents in one of the two unknown source profiles to be zero. Other profile elements are calculated by the model. Derived source profiles are matched to BB or MV according to correlation coefficients across source contributions.

†See footnotes in Table 1.

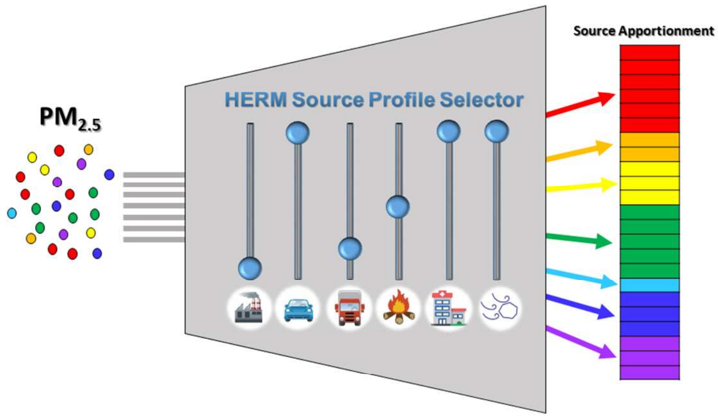
Table 6. Source apportionment of ambient $PM_{2.5}$ speciation dataset (BSP 2008-2009) by CAS HERM models.

Samples	Mean Source Contribution* ($\mu g\ m^{-3}$)						Correlation (r^2)			R/U Ratio† (<0.5/0.5-1.5/1.5-2.5/>2.5)											
#	EV-CMB Sources (x)		HERM^{4+2} Sources (y)		HERM^{5+2} Sources (z)		x vs y	x vs z	y vs z	x vs y				x vs z				y vs z			
226	**AMSUL**	0.555	**AMSUL**	0.488	**AMSUL**	0.481	0.986	0.987	0.999	56	**160**	10	0	27	**98**	58	43	221	**5**	0	0
226	**AMNIT**	0.161	**AMNIT**	0.174	**AMNIT**	0.172	0.991	0.992	1.000	204	**22**	0	0	177	**49**	0	0	226	**0**	0	0
226	**RDust1**	0.005																			
226	**RDust2**	0.123	**RDust2**	0.096	**RDust2**	0.092	0.254	0.435	0.844	158	**62**	6	0	121	**86**	19	0	219	**7**	0	0
226	**ADust**	0.506	**ADust**	0.515	**ADust**	0.510	0.966	0.973	0.995	169	**54**	3	0	123	**88**	14	1	218	**8**	0	0
226	**BBh**	0.105																			
226	**BBl**	1.358	**BB**~	2.217	**BB**~	2.256	0.966	0.972	0.999	38	**85**	79	24	11	**34**	39	142	204	**22**	0	0
226	**MV**	0.419	**MV + Ind.**~	0.204	**Ind.**~	0.124	0.803	0.759	0.936	11	**146**	62	7	0	**1**	3	222	7	**40**	67	112
226	**Coal**	0.029			**MV**	0.076		0.017						88	**107**	27	4				
	Sum	3.261		3.694		3.712	χ^2 = 1.81 (EV-CMB), 1.53 (HERM^{4+2}), and 1.23 (HERM^{5+2})														

*Source contribution (S_{jk}) derived by three HERM models are noted as x, y and z, respectively. EV-CMB is accomplished by HERM using 9 full source profiles (same as Table 3), HERM^{4+2} specifies 4 source profiles while calculating 2 source profiles, and HERM^{5+2} specifies 5 source profiles while calculating 2 source profiles. The last two use the HERM hybrid mode. Mean values take into account all available data.

~Source (profiles) calculated by HERM. "Ind." stands for mixed industrial emissions.

†Residue-Uncertainty (R/U) ratio of x and y is calculated by $|y-x|/(\sigma_x^2 + \sigma_y^2)^{½}$ where σ_x and σ_y are the source contribution uncertainty estimated by HERM. The ratios are then categorized into 4 ranges: <0.5, 0.5-1.5, 1.5-2.5, and >2.5 with numbers in each range shown in the table. Similarly, R/U ratio of x and z is calculated by $|z-x|/(\sigma_x^2 + \sigma_z^2)^{½}$ and R/U ratio of y and z is calculated by $|y-z|/(\sigma_y^2 + \sigma_z^2)^{½}$.



TOC Image

338x190mm (96 x 96 DPI)