Environmental Science & Technology

Article

pubs.acs.org/est

Coupling Chemical Transport Model Source Attributions with Positive Matrix Factorization: Application to Two IMPROVE Sites Impacted by Wildfires

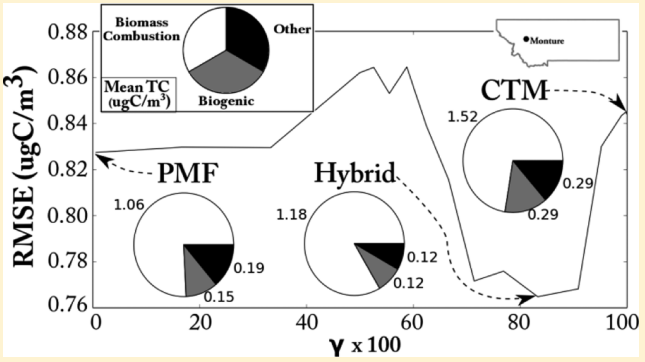
Timothy M. Sturtz,[†] Bret A. Schichtel,[§] and Timothy V. Larson*[,†,‡]

[†]Department of Civil and Environmental Engineering and [‡]Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, Washington 98195, United States

[§]Cooperative Institute for Research in the Atmosphere/NPS, Colorado State University, Fort Collins, Colorado 80523, United States

Supporting Information

ABSTRACT: Source contributions to total fine particle carbon predicted by a chemical transport model (CTM) were incorporated into the positive matrix factorization (PMF) receptor model to form a receptor-oriented hybrid model. The level of influence of the CTM versus traditional PMF was varied using a weighting parameter applied to an object function as implemented in the Multilinear Engine (ME-2). The methodology provides the ability to separate features that would not be identified using PMF alone, without sacrificing fit to observations. The hybrid model was applied to IMPROVE data taken from 2006 through 2008 at Monture and Sula Peak, Montana. It was able to separately identify major contributions of total carbon (TC) from wildfires and minor contributions from biogenic sources. The predictions of TC had a lower cross-validated RMSE than those from either PMF or CTM alone. Two unconstrained, minor features were identified at each site, a soil derived feature with elevated summer impacts and a feature enriched in sulfate and nitrate with significant, but sporadic contributions across the sampling period. The respective mean TC contributions from wildfires, biogenic emissions, and other sources were 1.18, 0.12, and 0.12 ugC/m^3 at Monture and 1.60, 0.44, and 0.06 ugC/m^3 at Sula Peak.



1. INTRODUCTION

Models that accurately describe source contributions to ambient fine particle mass and composition are an important air quality management tool. These models span a spectrum from purely deterministic models based on a priori knowledge of emissions, meteorology, and chemistry, to multivariate receptor models based on ambient pollutant measurements at a given receptor site.[1−11] Frequently, one of these two modeling approaches acts in an independent, supporting role to the other.[12,13]

Some investigators have focused on combining deterministic models with receptor-based particle measurement approaches to form a single "hybrid" model. Different approaches to combining these models include the use of genetic algorithms,[14−17] ensemble methods,[18−24] multiplicative bias correction,[25,26] and nonlinear optimization.[27] A subset of these hybrid modeling approaches include a deterministic chemical transport model (CTM) that includes secondary formation of particle mass. Of these, even fewer also include receptor information on particle composition in addition to particle mass.[22,25,26,28,29]

Here we present a hybrid model that explicitly combines predictions from a CTM with those from the Positive Matrix Factorization receptor model within the framework of the Multilinear Engine.[30] Our model is an extension of one initially proposed by Schichtel et.al.[28] that was developed using a synthetic data set. In this case, we apply an extension of this latter hybrid model to actual data at two rural IMPROVE monitoring sites in Montana with the goal of distinguishing contributions to total fine particle carbon from biogenic sources versus those from biomass combustion due to wildfires. On the basis of correlations between soluble potassium and particulate organic carbon, the impact of wildfires at Western U.S. IMPROVE sites is known to be significant.[31−33] Recent CTM modeling supports this conclusion, although correlations between predicted and observed particulate carbon values at these same Western U.S. sites are lower than in other regions,[34] providing additional motivation for the development of our hybrid model.

Received: June 5, 2014
Revised: September 1, 2014
Accepted: September 2, 2014
Published: September 2, 2014

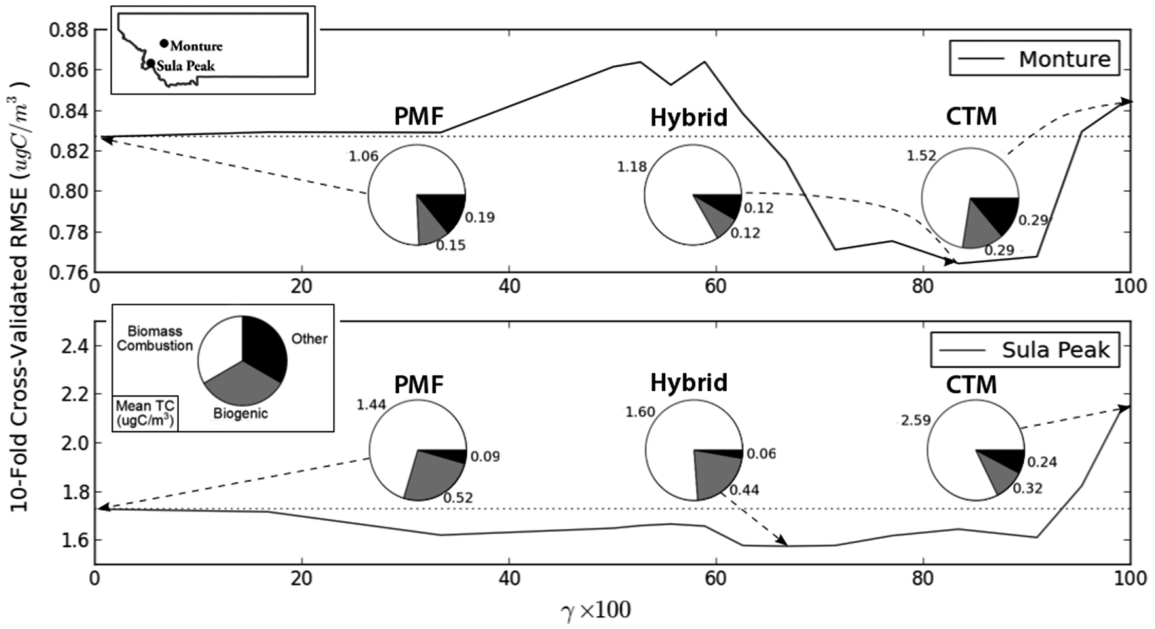


Figure 1. Cross-Validated Root Mean Square Error (RMSE) of Total Carbon versus weighting parameter (see eq 5) for Monture and Sula Peak. The black line represents the RMSE and the dotted line represents the PMF RMSE value. In addition, the pie charts for the PMF, CTM, and hybrid (minimum γ) models are provided for each site, annotated with the mean concentrations (ug/m^3) of biomass combustion, biogenic, and all other sources. A general map of Montana with the site locations is provided as an inset.

Table 1. RMSE Values (Both Cross-Validated and Not Cross-Validated) and R^2 at Monture and Sula Peak for the γ_{min} Hybrid Model and Constrained PMF against Observed Concentrations for Each Element and Total Carbon

	Monture			Sula Peak		
		RMSE/R^2 (cross-validated RMSE/R^2)			RMSE/R^2 (cross-validated RMSE/R^2)	
element	S/N	hybrid	const. PMF	S/N	hybrid	const. PMF
AL	2.85	0.037/0.93 (0.062/0.77)	0.039/0.92 (0.068/0.73)	2.42	0.012/0.95 (0.015/0.93)	0.012/0.94 (0.014/0.94)
BR	9.80	0.000/0.73 (0.000/0.56)	0.000/0.73 (0.001/0.54)	11.45	0.000/0.77 (0.001/0.52)	0.000/0.77 (0.001/0.53)
CA	18.42	0.028/0.60 (0.031/0.57)	0.028/0.60 (0.032/0.51)	18.25	0.020/0.61 (0.023/0.51)	0.020/0.61 (0.023/0.51)
EC1	5.48	0.198/0.96 (0.368/0.88)	0.281/0.96 (0.434/0.84)	5.19	0.353/0.99 (0.657/0.86)	0.250/0.99 (0.751/0.85)
EC2	1.55	0.042/0.47 (0.030/0.74)	0.041/0.55 (0.045/0.57)	1.43	0.027/0.80 (0.037/0.69)	0.027/0.80 (0.040/0.69)
OC1	1.12	0.297/0.81 (0.333/0.67)	0.312/0.78 (0.341/0.65)	1.25	0.464/0.89 (0.702/0.69)	0.470/0.87 (0.739/0.63)
OC2	2.57	0.410/0.93 (0.546/0.79)	0.454/0.91 (0.564/0.77)	2.56	0.610/0.96 (1.174/0.81)	0.694/0.96 (1.279/0.70)
OC3	2.39	0.141/0.96 (0.173/0.94)	0.104/0.98 (0.218/0.91)	2.21	0.254/0.98 (0.408/0.88)	0.222/0.99 (0.471/0.85)
OC4	2.37	0.110/0.77 (0.097/0.82)	0.102/0.81 (0.108/0.78)	1.95	0.106/0.82 (0.138/0.75)	0.101/0.84 (0.146/0.72)
OP	2.48	0.211/0.91 (0.281/0.87)	0.218/0.93 (0.309/0.85)	2.43	0.537/0.82 (0.229/0.96)	0.457/0.83 (0.261/0.95)
CU	1.90	0.000/0.14 (0.000/0.19)	0.000/0.14 (0.000/0.18)	1.86	0.000/0.06 (0.000/0.10)	0.000/0.06 (0.000/0.10)
H	13.62	0.039/0.98 (0.097/0.91)	0.046/0.98 (0.121/0.86)	15.52	0.066/0.99 (0.257/0.86)	0.078/0.99 (0.299/0.80)
FE	18.75	0.005/0.99 (0.016/0.90)	0.005/0.99 (0.020/0.86)	18.71	0.005/0.98 (0.012/0.91)	0.005/0.98 (0.011/0.91)
PB	3.47	0.000/0.55 (0.000/0.45)	0.000/0.55 (0.000/0.46)	3.11	0.000/0.36 (0.000/0.35)	0.000/0.34 (0.000/0.35)
MN	11.32	0.002/0.42 (0.003/0.37)	0.002/0.43 (0.003/0.32)	7.44	0.001/0.69 (0.001/0.39)	0.001/0.68 (0.001/0.31)
NO_3	4.21	0.082/0.35 (0.098/0.15)	0.079/0.40 (0.097/0.16)	6.40	0.106/0.79 (0.166/0.39)	0.100/0.80 (0.159/0.50)
K	15.92	0.018/0.93 (0.031/0.82)	0.010/0.98 (0.047/0.50)	18.65	0.016/0.96 (0.038/0.80)	0.018/0.97 (0.042/0.82)
SI	11.00	0.070/0.95 (0.136/0.80)	0.075/0.95 (0.151/0.75)	8.52	0.021/0.97 (0.040/0.91)	0.021/0.97 (0.040/0.91)
SR	1.81	0.000/0.62 (0.000/0.63)	0.000/0.62 (0.000/0.57)	1.85	0.000/0.65 (0.000/0.49)	0.000/0.65 (0.000/0.48)
SO_4	10.67	0.177/0.70 (0.264/0.51)	0.177/0.70 (0.234/0.57)	8.66	0.118/0.77 (0.233/0.45)	0.120/0.76 (0.226/0.48)
TI	5.25	0.001/0.98 (0.001/0.90)	0.001/0.98 (0.002/0.86)	6.15	0.000/0.98 (0.001/0.91)	0.000/0.98 (0.001/0.92)
ZN	16.06	0.002/0.34 (0.002/0.20)	0.002/0.31 (0.002/0.20)	9.28	0.002/0.54 (0.002/0.30)	0.002/0.52 (0.002/0.25)
TC		0.574/0.98 (0.765/0.91)	0.636/0.97 (0.827/0.88)		0.888/0.99 (1.577/0.86)	0.908/0.99 (1.729/0.82)

2. METHODS

2.1. Monitoring Data. We used three years (2006−2008) of speciated $PM_{2.5}$ data from two IMPROVE sites located in western Montana—Monture and Sula Peak (vista.cira.colostate.edu/improve) whose locations are shown in the inset of Figure 1. The temporal extent was selected based on available CTM results that were used as model constraints (see Section 2.2). The methodology described in Polissar et al.[35] was implemented to determine measurement uncertainties. We assessed the average concentration to measurement uncertainty ratio (signal-to-noise, S/N) using the methodology of Norris and Vedantham[36] and removed any species with average S/N <

0.2, and down weighted by a factor of 3 species with $0.2 \leq S/N < 2.0$.[37] In addition, species with reported concentrations below their detection limit or missing in over half the samples were removed from the analysis. The remaining species we considered are summarized in Table 1. Finally, an additional 8% and 25% of the respective samples from Monture and Sula Peak were removed if the mass reconstruction was outside IMPROVE limits.[13,38]

2.2. Chemical Transport Model. At the two monitoring sites, we used predictions of fine particle carbon based on the CAPITA Monte Carlo Lagrangian chemical transport model (CTM) available for the 2006−2008 calendar years.[39] The model considered 5 day upwind trajectories with accompanying emissions and atmospheric reactions along each trajectory. It was recently implemented at each IMPROVE site in the United States and shown to have similar performance metrics to CMAQ in predicting fine particle carbon across the U.S..[34] The model provided source contribution estimates of primary and/or secondary carbonaceous fine particles from the following source categories: biomass combustion, biogenic, mobile, area, oil, point, and other. Uncertainty estimates for each source were taken from the work previously conducted by Schichtel et al.[28]

2.3. Combined CTM/PMF Model. *2.3.1. Main Equations.* We implemented a modified version of the Positive Matrix Factorization (PMF) receptor model using the Multilinear Engine version 2.[30] The standard PMF model solves the basic mass balance equation (eq 1) for source contributions, g_{ik}, source profiles, f_{kj}, and model error, e_{ij}, for $i = 1,n$ samples, $j = 1,m$ species, and $k = 1,p$ sources. Species concentrations, x_{ij}, corresponding uncertainties, σ_{ij}, and the user-defined number of sources, p, serve as model inputs.

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + \varepsilon_{ij}, \quad \text{where } g_{ik},\ f_{kj} > 0 \tag{1}$$

To add prior source contribution constraints from the CTM, an additional set of equations were solved simultaneously with eq 1. Specifically, we have added equations representing each of the $t = 1,v$ contributions, g'_{it}, to total fine particulate carbon predicted by the CTM model, in this case, for $v = 2$ sources: total biomass combustion ($k = 1$) and biogenic emissions ($k = 2$). In general, the t^{th} CTM constraint is given by eq 2.

$$g'_{it} = g_{it} I_t + \varepsilon'_{it}, \quad \text{where } t \subseteq k \tag{2}$$

where I_t represents a diagonal matrix which is solved by the model to account for potential multiplicative bias in the CTM predictions.

2.3.2. Profile Constraints. The thermal fractions of carbon for each source are normalized as follows,

$$\sum_{s=1}^{w} f_{ks} = 1 \pm u_s, \quad \text{where } s \subseteq j \tag{3}$$

where $s = 1,w$ carbon fractions. In this case, we set $u_s = 0.001$. Equation 3 rescales f such that g_{ik} in eqs 1 and 2 represent total fine particle carbon and f_{kj} represents mass fraction of species j in source k relative to total carbon.

On the basis of prior knowledge of fire emissions in this region, we constrained the biomass combustion profile such that the value of f for potassium is[33] > 0.01 and that the corresponding values for NO_3 and SO_4 are <0.05.[40]

Additionally, we constrain each of the $l = 1,b$ secondary feature profiles, in this case the biogenic source, so that each of the $r = 1,c$ noncarbonaceous species are near zero. This assumes a primary biogenic source consisting only of VOCs. While organic sulfate and organic nitrate are found in some environments, we assume they are very low and set to near 0 because of the remote environments of Sula Peak and Monture and low SO_2 and NO_x concentrations.

$$f_{lr} = 0 \pm u_{lr}, \quad \text{where } l \subseteq k \text{ and } r \subseteq j \tag{4}$$

In this case, we set $u_{lr} = 1 \times 10^{-5}$.

For the biomass combustion source we have limited the species to contain the carbon thermal fractions, potassium, nitrate, sulfate, and hydrogen. The decision to limit these species was based on the biomass combustion-like source profile resolved for these sites by the pure PMF model and the EPA SPECIATE database. To impose this constraint we applied eq 4 with the above set of $r = 11$ species. Going forward, PMF will refer to the $\gamma = 0$ scenario.

2.3.3. Penalized Object Function. In ME-2, eqs 1 through 4 are solved by minimizing an object function, Q, through the use of a preconditioned conjugate gradient algorithm. The object function (eq 5) includes a weighting parameter associated with each of the applied constraints.

$$Q = (1-\gamma) \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{\varepsilon_{ij}}{\sigma_{ij}}\right]^2 + (\gamma) \sum_{i=1}^{n} \sum_{t=1}^{v} \left[\frac{\varepsilon'_{it}}{\omega_{it}}\right]^2 \left[\frac{\varepsilon'_{it}}{\omega_{it}}\right]^2 + \sum_{s=1}^{w} u_s + \sum_{l=1}^{b} \sum_{r=1}^{c} u_{lr} \tag{5}$$

The first two terms in eq 5 represent the residuals from eqs 1−3, where σ_{ij} is the species measurement uncertainty, ω_{it} is the CTM uncertainty, and γ is a user-defined weighting parameter that allows one to weigh the CTM predictions (eq 2) relative to those from the bivariate mass balance model (eqs 1, 3, and 4). We implement this in ME-2 by specifying auxiliary formulas using error mode −14 (error modes defined in Supporting Information, SI, Table S2) for the first two terms. The second two terms in eq 5 represent the prior source profile constraints (eqs 3 and 4) and are implemented with auxiliary formulas using error mode −12.

The uncertainties associated with the CTM results (eq 6) were defined as a function of estimated fractional error, λ_t, and a fixed minimum CTM error, β, for the CTM model.

$$\omega_{it} = \sqrt{(\lambda_t * g'_{it})^2 + \beta^2} \tag{6}$$

2.3.4. Model Implementation. For an initial user-specified number of sources, p, and a given value of the weighting parameter, γ, multiple model runs were conducted at 40 different starting points chosen randomly, and the chosen p-source baseline model was selected based on the minimum value of Q (= Q_{min}). The model was executed using a standard and a 10-fold cross-validation approach (as described in the next paragraph). These profile constrained model runs were conducted assuming a range of values for p with γ set equal to zero, representing the basic PMF solution (eqs 1, 3, and 4) without additional information from the CTM (eq 2). The final value of p was chosen based on the following criteria: the smallest value of p where a change in the ratio of cross-validated Q_{min} to $Q_{theoretical}$ approaches zero and user judgment based upon qualitative agreement between each f_{kj} and prior knowledge of source profiles from known source types within the modeled region. $Q_{theoretical}$, the χ^2 value, represents an

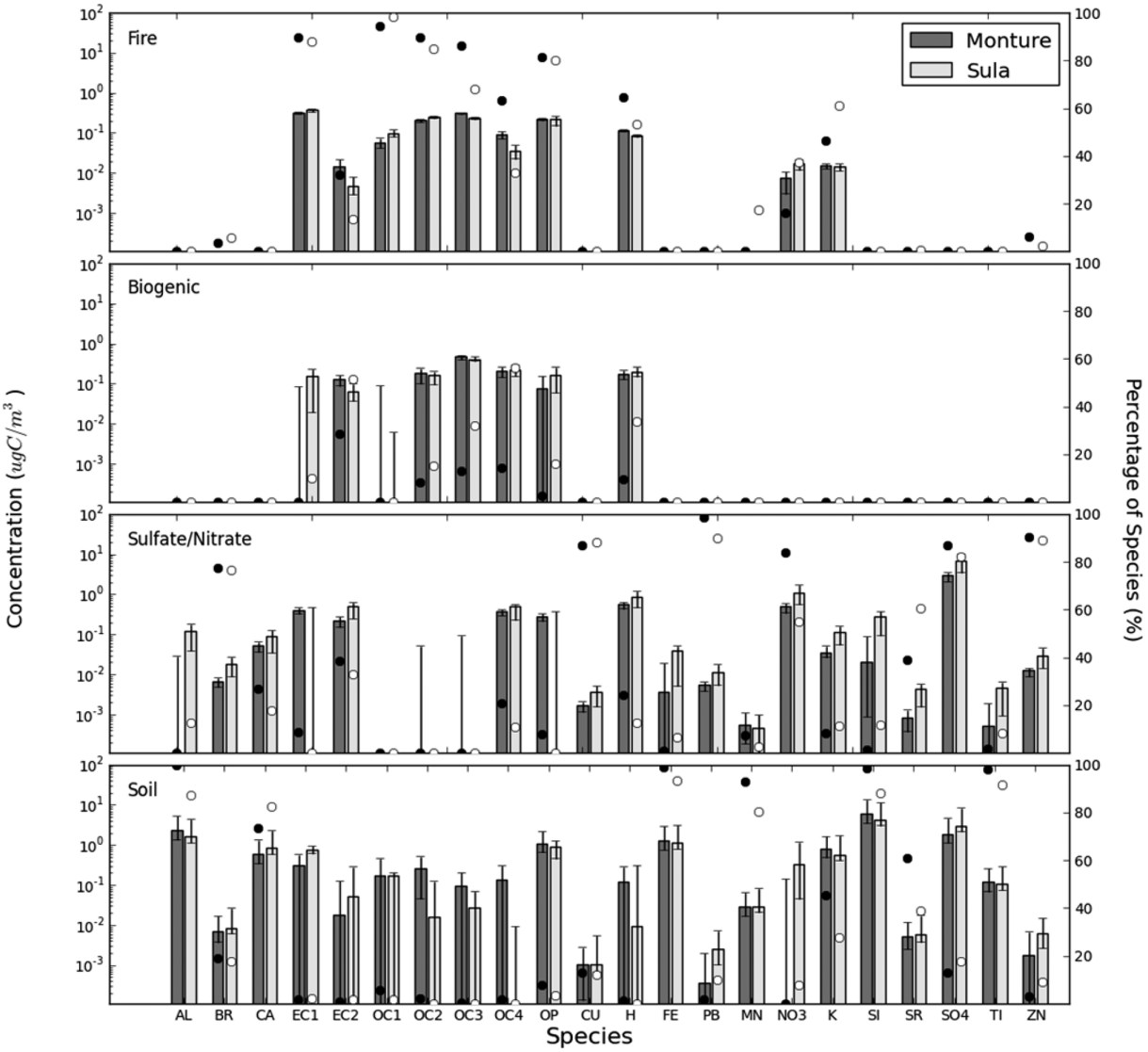


Figure 2. Hybrid model derived source profiles for Monture and Sula Peak. The dark bars represent the average species contributions to total carbon for Monture and the light bars represent Sula Peak. The bootstrapped 95% confidence intervals are also shown. The circles represent the percent of overall predicted concentrations for a given species associated with a feature at Monture (black circles) and Sula Peak (white circles).

approximation of the degrees of freedom for a given value of p, where base run $Q_{\text{theoretical}} = (n - p) * (m - p)$, 10-fold cross-validated $Q_{\text{theoretical}} = (n - p) * (m * 0.9 * 10 - p) + (m * 0.1 * 10 - p)$, m is number of samples, and n is the number of species.[41,42] Plots of the ratio between cross-validated Q_{min} and $Q_{\text{theortical}}$ as a function of p are provided in the SI (Figure S3).

A final value of γ was chosen based on a minimum in the estimated root-mean-square error (RMSE) of predicted versus measured total carbon. For a given value of γ, the RMSE was computed via a 10-fold cross validation procedure to ensure a robust result. The measurements and CTM results were initially assigned to one of 10 separate groups. The model was then run 10 times, leaving one group out each time. The contributions of total carbon were predicted for the 10% missing values using ME-2 with the derived profiles from a given run. The resulting 10 groups of predictions were then combined into a single data set and the RMSE of total carbon was determined for each value of γ.

After selecting the γ associated with the minimum RMSE, blocked bootstrapping[36] was then used to estimate the uncertainties of the f_{kj} and the average values of g_{ik}. A sample block size of 4 was defined and profile matching was conducted on the predicted contributions of each species with an acceptable match defined as an $R^2 > 0.6$ across all contributions for a given model run.

Here, the hybrid approach was applied using 17 different values of γ and solved for 4 features. The Monture and Sula Peak IMPROVE monitor data were modeled using CTM predicted contributions from biomass combustion and biogenic sources as constraints (see eq 2). Total biomass combustion was used instead of specifying a priori separate secondary and primary biomass combustion features because of the high Pearson correlations between the CTM predictions of these two source contributions to total carbon, (0.97 and 0.98 at Monture and Sula Peak respectively). This high correlation would likely cause difficulty in effectively separating the two sources, thus the sum of the CTM predicted contributions from these two sources was used to constrain the single biomass combustion PMF feature as described earlier.

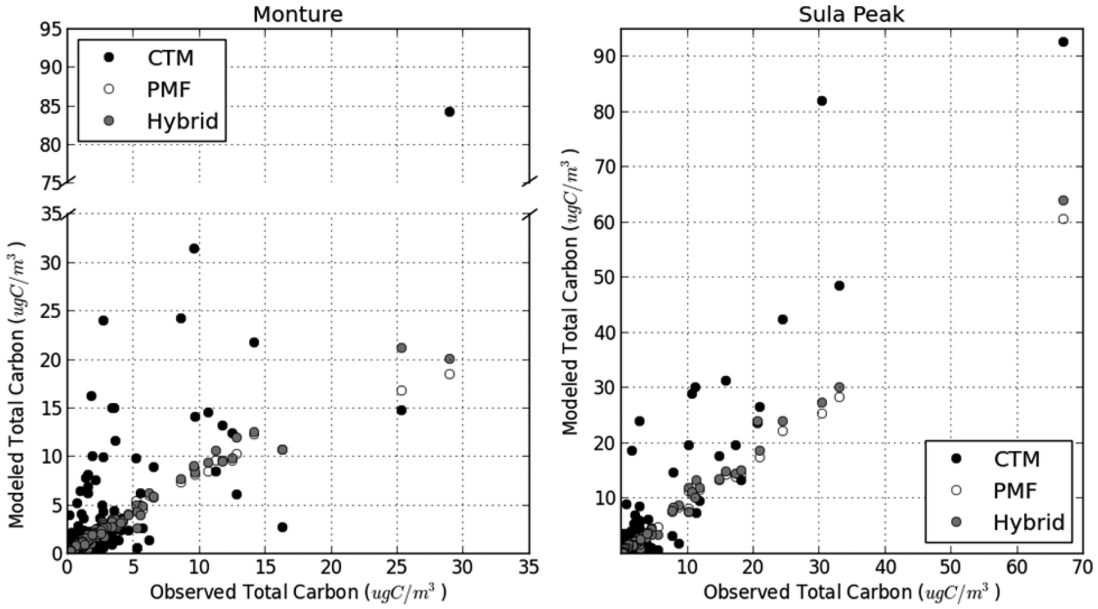


Figure 3. Predicted Versus Observed Total Carbon for the CTM, PMF, and hybrid models at Monture and Sula Peak. The black dots represent the CTM, the gray dots represent the hybrid model, and the white dots represent PMF. The white dots are somewhat obscured due to similarities in the fit of PMF and the hybrid solution.

To determine the CTM uncertainty, the fractional error was set to λ = 1.18 and 0.66 for the biomass combustion and biogenic features,[28] respectively, and a minimum CTM error of β = 0.1 was used. A sensitivity analysis of these assumed values was also conducted.

3. RESULTS

At both sites, we chose a four feature model based on the criteria described earlier (see SI Figure S3 for plots of cross-validated $Q/Q_{theoretical}$ versus p). At each site, the hybrid model was defined by a minimum value of the cross-validated total carbon RMSE as a function of γ (Figure 1). The γ associated with the minimum cross-validated total carbon RMSE at Monture and Sula Peak was 0.83 and 0.67, respectively. Additional error and bias metrics were calculated (Table S4 in the SI) for the cross-validated model and provide additional confidence in the minimum γ selected by the RMSE metric. The $Q/Q_{theoretical}$ without cross-validation for Monture and Sula Peak at the minimum γ was 2.19 and 1.94, respectively. The value of γ_{min} and the associated total carbon RMSE were relatively insensitive to changes in the values of λ and β in eq 6 as summarized in Figures S7 and S8 and Table S3 in the SI. In both the PMF and hybrid models, two features were identified in addition to the constrained biomass combustion and biogenic source features: a sulfate/nitrate-rich feature and a soil feature. The profiles for all four of these features at each site for the final hybrid model are shown in Figure 2. Overall, the derived profiles at the two sites are similar and, where differences occur, the associated bootstrapped uncertainties span these differences with the exception of Al in the sulfate and nitrate rich feature. Profiles of the four features derived from the PMF model are provided in Figure 2.

Performance statistics of both the hybrid and PMF model across all modeled species are provided in Table 1. In general, the hybrid model performs slightly better than the PMF model across all species. In addition, the predicted average contributions to total carbon from biomass combustion, biogenic sources, and "other" sources (sulfate/nitrate-rich and soil) are shown in Figure 1 as pie charts for γ values associated with PMF (γ = 0), CTM (γ = 1) and the hybrid model. Figure 3 shows predicted versus observed total carbon for each sample at each site for all three models. The hybrid model displays the best agreement, given that it by definition has the lowest RMSE. For the PMF and hybrid models, a time series of the predicted contributions (Figures S2 and S3) and the source profiles (Figure S4) at each site can be found in the SI. Seasonal performance statistics of modeled total carbon are provided in SI Table S1 for the hybrid and CTM model; associated scatter plots are provided in SI Figure S6.

4. DISCUSSION

We present a methodology that allows CTM source attribution results to be applied as constraints within the PMF model via ME-2. This methodology provides the ability to separate features that would not be identified using PMF alone, while maintaining the fit to observations that is common to PMF analyses. Additionally, interpretation of constrained hybrid factors is simplified when correlation is maintained with the CTM constraints. Here, we applied the hybrid methodology using a rudimentary CTM and IMPROVE speciated measurements to distinguish biomass combustion and secondary biogenic impacts from total carbon. However, the general model is useful for assessment of CTM predicted particle mass or other species of interest.

Our models did not attempt to distinguish between primary and secondary biomass because the CTM model predictions of these two features were highly correlated. While the scaling factor in eq 2 accounts for multiplicative bias in the CTM, it causes ambiguity between CTM impact constraints when they are near perfectly correlated and lacking tracer-like species. Here, the results of the model represent combined primary, aged primary, and secondary biomass combustion impacts within a single source profile. These CTM results were therefore combined due to their high correlation and the potential of the constrained ME-2 features to swap between the two separate biomass features. However, the separation of these

two features using the hybrid model would, in principle, be possible using other CTM models, data from other sites further from the fire origins, or additional marker species specific to primary and/or secondary particles.

The biomass combustion feature therefore represents both primary and secondary particles and the contributions are dominated by summer and autumn fire events in all three years (SI Figures S1 and S2). Confidence in the biomass burning hybrid modeling derives from the fact that the hybrid model biomass combustion profiles, shown in Figure 2, are largely composed of organic carbon thermal fractions, and the wildfire CTM constraint on this feature is consistent with the high percentage of pyrolytic carbon relative to the other thermal fractions. Additionally, the biomass combustion contributions predicted by the hybrid model are reasonably well correlated with the CTM predictions of total wildfire (Pearson correlation coefficients of 0.72 and 0.91 for Monture and Sula Peak, respectively) even though the CTM predicts consistently larger contributions than the hybrid model (see Figure 3).

The average ratio of K to OC from the biomass combustion profiles at Monture and Sula were 0.036 (s.d. = 0.017) and 0.026 (s.d. = 0.016), respectively. Assuming that the K in this feature is water-soluble, these values are consistent with contributions from primary fire emissions, specifically with the fact that (1) the range of values of primary emissions of woody debris sampled near these sites (Northern Rockies region) as reported by Munchack et.al.;[33] (2) the fires occurred throughout this region in all three years and many of the largest CTM-modeled smoke impacts were from these more local fires, indicating less aged smoke; and (3) the largest impacts from this feature to total carbon did not occur at the same time (see SI Figure S5) at both sites, implying the lack of a broad regional haze from long-range transport.

The benefit of the hybrid model over the pure CTM results is evident in Figure 3 and SI Figure S6, where it is seen that high measured TC concentrations were often significantly over or underestimated by the CTM. For example, at Monture, MT during the autumn, the CTM model regularly over and underestimated the measured TC by 2−22 $\mu g/m^3$. This is likely due to errors in the biomass burning contribution, since all large overestimations were associated with biomass burning impacts contributing 80% or more to the simulated TC. As discussed by Schichtel et al.,[34] these biases are at least partly due to errors in the biomass burning emissions and transport and dilution of the smoke plumes. These biases are not seen in the hybrid model results and the hybrid model fit the measured total carbon better than the CTM during the year (Figure 3) and each season, particularly during times of high total carbon contributions (see SI Figure S6 and Table S1). In addition, changes of 2 $\mu g/m^3$ or more between the hybrid and CTM model occurred only for the biomass burning contributions. Consequently, in the hybrid model, the biogenic and other nonfire contributions were not affected by biases in the biomass burning contributions.

The biogenic feature was constrained to have a purely carbon and hydrogen based profile, shown in Figure 2, and the resulting contributions have strong seasonality peaking in the summer months as expected (see SI Figures S1 and S2). The ability to distinguish the biogenic feature from the biomass combustion in the hybrid model is a result of both the profile and CTM constraints. When the PMF model was run without these additional constraints, these sources were lumped into a single factor. In addition to improving the separation of correlated sources, the incorporation of CTM helped to interpret the PMF factors. For example, the biogenic feature was highly correlated with the CTM-predicted biogenic contributions.

The soil profiles are provided in Figure 2. The feature identified as soil is the majority contributor to Al, Ca, Fe, Mn, Si, and Ti. In addition to elements commonly found in soils, we identified the presence of K and Sr, a result in agreement with the surface geochemistry identified by Shacklette and Boerngen.[44] A comparison of the relative composition of K, Fe, and Ca reported by Shacklette is consistent with our profile. The contributions from the soil feature have a high level of seasonality with the peaks in the Spring and Summer, due to a combination of dry soils and long-range transport[45] (see SI Figures S1 and S2). There are modest correlations between the biomass combustion feature and the soil feature at both sites (r = 0.52 and 0.56 at Monture and Sula Peak, respectively), and it is plausible that some of these soil particles were generated via updrafts during fire events and directly emitted along with the biomass smoke.

The nitrate and sulfate dominated profiles, shown in Figure 2, are enriched with Cu, Pb, and Zn and present as misaligned, sharp peaks, potentially implicating nearby industrial and area sources. Generally, the composition of the profiles at these two sites are similar with the possible exception of Sr, Zn, and Al. In any case, this feature was found to have a low correlation with the biomass combustion source.

To examine the possible source(s) associated with the sulfate/nitrate-rich feature, we used the NOAA hysplit model to assess 48 h back trajectories leading to the three large total carbon spikes observed in Monture and 2 large spikes at Sula Peak (see SI Figures S1 and S2). At Monture, two of the spikes were associated with winds coming from the West and most recently passing over the city of Missoula. During the sampling period a large pulp and paper mill was active in Missoula that may have contributed to these spikes. The third spike at Monture was associated with winds from the Northwest near Libby where current mining operations[46] may have also contributed. The uncertainty associated with Sr, Zn, and Al in the Monture profile may be due to the sporadic impacts from both paper milling and mining facilities. At Sula Peak, the winds arrived from the West during these spikes, passing over active gold, cobalt, and molybdenum mines and processing facilities.[47] A comparison of this feature (Figure 2) to the EPA SPECIATE database found that the "Regional smelter background" SPECIATE profile is in agreement with the enriched values of Cu, Pb, Sr, Zn, NO_3, and SO_4 in this feature.

The sensitivity of the hybrid model to variations in the CTM uncertainty equation were explored by modifying the relative error fraction between biomass combustion and biogenic and by modifying the minimum model error values (SI Figures S7 and S8). Changes in the relative error fractions caused only small variations in the RMSE curve and identified similar minimum γ values. The feature profiles and compositions at the minimum in each case was identical or near identical, indicating that the model is robust to the error fraction estimates for dominant features. Minimum model error values of β = 0.01, 0.1, and 1.0 demonstrated consistent minimum γ identification. With β values of 10 or 100 the RMSE minimum was no longer identifiable. The lack of a minimum in the cross-validated RMSE indicates no improvement over PMF, likely due to the large assumed error associated with the CTM and thus the lack of control by the constraining CTM features.

Comparisons to source apportionment studies at Monture and Sula Peak were sought to better understand the performance of the hybrid model. To the authors' knowledge, only one publication[43] contains comparable results of carbonaceous aerosols at these two sites. The study applied carbon 14 (^{14}C) dating to particulate samples for calibration of EC/TC ratios and to separate TC into fossil and contemporary fractions. In the case of rural sites, the contemporary fraction is largely TC derived from biomass combustion and biogenic secondary formation. For the Sula Peak and Monture regions, the ^{14}C study showed that the contemporary fraction is expected to be between 0.90 ± 0.24 during the summer and 0.90 ± 0.15 during the winter. Our results (SI Table S4) demonstrate that the hybrid model provides results similar to that of the ^{14}C study and more near the average than the PMF or CTM models alone. The PMF results were within the error bounds of the ^{14}C study for both seasons, and the CTM results were within the error bounds during the summer but were well outside the bounds during the winter season.

Here we provided a framework for direct coupling of predictions from a CTM model with those from PMF via ME-2. The CTM modeling results used here provided estimates of total carbon.

However our modeling framework provides the flexibility to assess particle mass or other species of interest as predicted by a CTM. The use of CTM constraints provides guidance and increased confidence in identifying PMF derived features. The differing γ_{min} at the two sites modeled here emphasizes the need to assess a range of γ values for model applications at other sites or with other CTMs. This is likely due to differing errors in the PMF and CTM model at the two locations, and the γ_{min} should weight the model with the least error. An analysis of additional sites and the overall variability in γ_{min} in relation to the PMF and CTM modeling errors is worthy of future investigation.

Other investigators have coupled CTM results with receptor-based models. The approach most similar to ours is that of Maier and colleagues.[22] Their ensemble model uses CTMs and PMF to derive average contributions from sources, uses these contributions to derive source profiles using an inverse chemical mass balance (CMB) approach, and finally uses the new source profiles in a common CMB approach to estimate contributions. Our model differs by directly applying the CTM predictions with uncertainty to the PMF model. This allows us to resolve sources that are not necessarily identified by PMF alone.

■ ASSOCIATED CONTENT

Supporting Information

Figure S1, Hybrid model derived feature contributions for the Monture site; Figure S2, hybrid model derived feature contributions for the Sula Peak site; Figure S3, cross-validated $Q/Q_{theoretical}$ versus number of factors; Figure S4, hybrid model derived biomass combustion source profiles of the hybrid and PMF solutions at Monture and Sula peaks; Figure S5, side-by-side comparison of the biomass combustion feature contributions from Monture and Sula peaks; Figure S6, comparison of CTM and hybrid modeled total carbon versus observed total carbon for Monture and Sula peaks across seasons; Figure S7, impact on the cross-validated total carbon RMSE curve due to variations in the relative error fraction (biomass combustion: biogenic) of the CTM uncertainty at the Monture site; Figure S8, impact on the cross-validated total carbon RMSE curve due to variations in the minimum model error (ugC/m^3) of the CTM uncertainty at the Monture site; Table S1, modeled total carbon seasonal performance statistics from the CTM and the hybrid model (γ_{min}) at Monture and Sula Peak; Table S2, ME-2 error mode equations; Table S3, hybrid model sensitivity analysis; Table S4, extended performance metrics of the cross-validated hybrid model; and Table S5, comparison of modeled TC contemporary fractions to contemporary fractions determined by carbon dating. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +1 206 543 6815; fax: 206 685 3836; e-mail: tlarson@uw.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Joint Fire Science Program (09-1-03-1). The assumptions, findings, conclusions, judgments, and views presented herein are those of the authors and should not be interpreted as representing the National Park Service polices.

■ REFERENCES

(1) Seigneur, C.; Pai, P.; Hopke, P. K.; Grosjean, D. Modeling atmospheric particulate matter. *Environ. Sci. Technol.* **1999**, *33*, 80A−86A.

(2) Watson, J. G. Overview of receptor model principles. *J. Air Pollut. Control Assoc.* **1984**, *34*, 619−623.

(3) Schauer, J. J.; Lough, G. C.; Shafer, M. M. Characterization of metals emitted from motor vehicles. *Health Effects Instit.* **2006**, *133*.

(4) Reff, A.; Eberly, S. I.; Bhave, P. V. Receptor modeling of ambient particulate matter data using positive matrix factorization: Review of existing methods. *J. Air Waste Manage. Assoc. (1995)* **2007**, *57*, 146−54.

(5) 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.; Hueglin, C.; Bloemen, H.; Wåhlin, P.; Vecchi, R.; Miranda, a. I.; Kasper-Giebl, a.; Maenhaut, W.; Hitzenberger, R. Source apportionment of particulate matter in Europe: A review of methods and results. *J. Aerosol Sci.* **2008**, *39*, 827−849.

(6) Pant, P.; Harrison, R. M. Critical review of receptor modelling for particulate matter: A case study of India. *Atmos. Environ.* **2012**, *49*, 1−12.

(7) Zhang, Y.; Bocquet, M.; Mallet, V.; Seigneur, C.; Baklanov, A. Real-time air quality forecasting, Part I: History, techniques, and current status. *Atmos. Environ.* **2012**, *60*, 632−655.

(8) Zhang, Y.; Bocquet, M.; Mallet, V.; Seigneur, C.; Baklanov, A. Real-time air quality forecasting, Part II: State of the science, current research needs, and future prospects. *Atmos. Environ.* **2012**, *60*, 656−676.

(9) Belis, C. a.; Karagulian, F.; Larsen, B. R.; Hopke, P. K. Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe. *Atmos. Environ.* **2013**, *69*, 94−108.

(10) El-harbawi, M. Air quality modelling, simulation and computational methods: A review. *Environ. Rev.* **2013**, *24*, 149−179.

(11) Lingwall, J. W.; Christensen, W. F. Pollution source apportionment using a priori information and positive matrix factorization. *Chemom. Intel. Lab. Syst.* **2007**, *87*, 281−294.

(12) Marmur, A.; Park, S.-K.; Mulholland, J. a.; Tolbert, P. E.; Russell, A. G. Source apportionment of PM2.5 in the southeastern United States using receptor and emissions-based models: Conceptual differences and implications for time-series health studies. *Atmos. Environ.* **2006**, *40*, 2533−2551.

(13) Watson, J. G. Visibility: Science and regulation. *J. Air Waste Manage. Assoc.* **2002**, *52*, 628−713.
(14) Haupt, S. E.; Young, G. S.; Allen, C. T. Validation of a receptor−dispersion model coupled with a genetic algorithm using synthetic data. *J. Appl. Meteorol. Climatol.* **2006**, *45*, 476−490.
(15) Allen, C.; Young, G.; Haupt, S. Improving pollutant source characterization by better estimating wind direction with a genetic algorithm. *Atmos. Environ.* **2007**, *41*, 2283−2289.
(16) Allen, C. T.; Haupt, S. E.; Young, G. S. Source characterization with a genetic algorithm−coupled dispersion−backward model incorporating SCIPUFF. *J. Appl. Meteorol. Climatol.* **2007**, *46*, 273−287.
(17) Khlaifi, A.; Ionescu, A.; Candau, Y. Pollution source identification using a coupled diffusion model with a genetic algorithm. *Math. Comput. Simul.* **2009**, *79*, 3500−3510.
(18) McKeen, S.; Chung, S. H.; Wilczak, J.; Grell, G.; Djalalova, I.; Peckham, S.; Gong, W.; Bouchet, V.; Moffet, R.; Tang, Y.; Carmichael, G. R.; Mathur, R.; Yu, S. Evaluation of several PM 2.5 forecast models using data collected during the ICARTT/NEAQS 2004 field study. *J. Geophys. Res.* **2007**, *112*, D10S20.
(19) McKeen, S.; Grell, G.; Peckham, S.; Wilczak, J.; Djalalova, I.; Hsie, E.-Y.; Frost, G.; Peischl, J.; Schwarz, J.; Spackman, R.; Holloway, J.; De Gouw, J.; Warneke, C.; Gong, W.; Bouchet, V.; Gaudreault, S.; Racine, J.; McHenry, J.; McQueen, J.; Lee, P.; Tang, Y.; Carmichael, G. R.; Mathur, R. An evaluation of real-time air quality forecasts and their urban emissions over eastern texas during the summer of 2006 second Texas air quality study field study. *J. Geophys. Res.* **2009**, *114*, D00F11.
(20) Chuang, M.-T.; Zhang, Y.; Kang, D. Application of WRF/Chem-MADRID for real-time air quality forecasting over the southeastern United States. *Atmos. Environ.* **2011**, *45*, 6241−6250.
(21) Gebhart, K. a.; Schichtel, B. a.; Malm, W. C.; Barna, M. G.; Rodriguez, M. a.; Collett, J. L. Back-trajectory-based source apportionment of airborne sulfur and nitrogen concentrations at Rocky Mountain National Park, Colorado, USA. *Atmos. Environ.* **2011**, *45*, 621−633.
(22) Maier, M. L.; Balachandran, S.; Sarnat, S. E.; Turner, J. R.; Mulholland, J. a; Russell, A. G. Application of an ensemble-trained source apportionment approach at a site impacted by multiple point sources. *Environ. Sci. Technol.* **2013**, *47*, 3743−51.
(23) Lee, D.; Balachandran, S.; Pachon, J.; Shankaran, R.; Lee, S.; Mulholland, J. A.; Russell, A. G. Ensemble-trained PM2.5 source apportionment approach for health studies. *Environ. Sci. Technol.* **2009**, *43*, 7023−7031.
(24) Balachandran, S.; Chang, H. H.; Pachon, J. E.; Holmes, H. A.; Mulholland, J. A.; Russell, A. G. Bayesian-based ensemble source apportionment of PM 2.5. *Environ. Sci. Technol.* **2013**, *47* (23), 13511−13518.
(25) Schichtel, B. a.; Malm, W. C.; Gebhart, K. a.; Barna, M. G.; Knipping, E. M. A hybrid source apportionment model integrating measured data and air quality model results. *J. Geophys. Res.* **2006**, *111*, 1−19.
(26) Stroud, C. a.; Moran, M. D.; Makar, P. a.; Gong, S.; Gong, W.; Zhang, J.; Slowik, J. G.; Abbatt, J. P. D.; Lu, G.; Brook, J. R.; Mihele, C.; Li, Q.; Sills, D.; Strawbridge, K. B.; McGuire, M. L.; Evans, G. J. Evaluation of chemical transport model predictions of primary organic aerosol for air masses classified by particle component-based factor analysis. *Atmos. Chem. Phys.* **2012**, *12*, 8297−8321.
(27) Yuval, S. B.; Broday, D. M. Data-driven nonlinear optimisation of a simple air pollution dispersion model generating high resolution spatiotemporal exposure. *Atmos. Environ.* **2013**, *79*, 261−270.
(28) Schichtel, B. A.; Malm, W. G.; Collett, J. L.; Sullivan, A. P.; Holden, A. S.; Patterson, L. A.; Rodriguez, M. A.; Barna, M. G.Estimating the contribution of smoke to fine particulate matter using a hybrid-receptor model. *Air Waste Manage. Aerosol Atmos. Optics*, 2008.
(29) Hu, Y.; Balachandran, S.; Pachon, J. E.; Baek, J.; Ivey, C.; Holmes, H.; Odman, M. T.; Mulholland, J. a.; Russell, a. G. Fine particulate matter source apportionment using a hybrid chemical transport and receptor model approach. *Atmos. Chem. Phys.* **2014**, *14*, 5415−5431.
(30) 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*, 854−888.
(31) Park, R. J.; Jacob, D. J.; Logan, J. a. Fire and biofuel contributions to annual mean aerosol mass concentrations in the United States. *Atmos. Environ.* **2007**, *41*, 7389−7400.
(32) Jaffe, D.; Hafner, W.; Chand, D.; Westerling, A.; Spracklen, D. Interannual variations in PM2.5 due to wildfires in the western United States. *Environ. Sci. Technol.* **2008**, *42*, 2812−2818.
(33) Munchak, L. a.; Schichtel, B. a.; Sullivan, A. P.; Holden, A. S.; Kreidenweis, S. M.; Malm, W. C.; Collett, J. L. Development of wildland fire particulate smoke marker to organic carbon emission ratios for the conterminous United States. *Atmos. Environ.* **2011**, *45*, 395−403.
(34) Schichtel, B. a.; Rodriguez, M. a.; Barna, M. G.; Gebhart, K. a.; Pitchford, M. L.; Malm, W. C. A semi-empirical, receptor-oriented Lagrangian model for simulating fine particulate carbon at rural sites. *Atmos. Environ.* **2012**, *61*, 361−370.
(35) Polissar, A. V.; Hopke, P. K.; Paatero, P.; Malm, W. C.; Sisler, J. F. Atmospheric aerosol over Alaska 2. Elemental composition and sources. *J. Geophys. Res.* **1998**, *103*.
(36) Norris, G.; Vedantham, R. *EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide*; 2008.
(37) Paatero, P.; Hopke, P. K. Discarding or downweighting high-noise variables in factor analytic models. *Anal. Chim. Acta* **2003**, *490*, 277−289.
(38) Frank, N. H. Retained nitrate, hydrated sulfates, and carbonaceous mass in federal reference method fine particulate matter for six eastern U.S. cities. *J. Air & Waste Manage. Assoc. (1995)* **2006**, *56*, 500−511.
(39) Schichtel, B. a.; Husar, R. B. Regional simulation of atmospheric pollutants with the CAPITA Monte Carlo model. *J. Air Waste Manage. Assoc.* **1997**, *47*, 301−333.
(40) Reid, J. S.; Koppmann, R.; Eck, T. F.; Eleuterio, D. P. A review of biomass burning emissions Part II: Intensive physical properties of biomass burning particles. *Atmos. Chem. Phys.* **2005**, *5*, 799−825.
(41) Comero, S.; Capitani, L.; Gawlik, B. M. Positive Matrix Factorisation (PMF) An introduction to the chemometric evaluation of environmental monitoring data using PMF. *EUR − Scientific and Technical Research Series*; EU Joint Center for Research, 2009; 58 pp. DOI: 10.2788/2497.
(42) Paatero, P.; Tapper, U. Analysis of different modes of factor analysis as least squares fit problems. *Chemom. Intel. Lab. Syst.* **1993**, *18*, 183−194.
(43) Schichtel, B. a.; Malm, W. C.; Bench, G.; Fallon, S.; McDade, C. E.; Chow, J. C.; Watson, J. G. Fossil and contemporary fine particulate carbon fractions at 12 rural and urban sites in the United States. *J. Geophys. Res.* **2008**, *113*, D02311.
(44) Shacklette, H. T.; Boerngen, J. G. *Element Concentrations in Soils and Other Surfici 1 Materials of the Conterminous United Stat Element Concentrations in Soils and Other Surficial Materials of the Conterminous United States*; Washington, D.C., 1984; pp 1−63.
(45) Hand, J. L.; Schichtel, B. a.; Pitchford, M.; Malm, W. C.; Frank, N. H. Seasonal composition of remote and urban fine particulate matter in the United States. *J. Geophys. Res.* **2012**, *117*, D05209.
(46) McCulloch, R. Montana Mines and Exploration−2012 http://www.mbmg.mtech.edu/gmr/gmr-mines_exploration.asp (accessed Nov 22, 2013).
(47) Idaho Geological Survey Map of 2006 mining-Exploration Activity http://www.mineidaho.com/about/mines/ (accessed Nov 22, 2013).