Supporting information

Quantification of major particulate matter species from a single filter type using infrared spectroscopy – Application to a large-scale monitoring network

Bruno Debus[1], Andrew T. Weakley[1], Satoshi Takahama[2], Kathryn M. George[1,3], Bret Schichtel[4], Scott Copeland[5], Anthony S. Wexler[1,6], Ann M. Dillner[1]*

[1] Air Quality Research Center, University of California, Davis, California, 95616, USA

[2] ENAC/IIE, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

[3] Monitoring and Laboratory Division, California Air Resources Board, Sacramento, CA 95811, USA

[4] National Park Service, Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO 80523, USA

[5] Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, 80523, USA

[6] Departments of Mechanical and Aerospace Engineering, Civil and Environmental Engineering, and Land, Air and Water Resources, University of California, Davis, California, 85616, USA

*Correspondence to: Ann M. Dillner (amdillner@ucdavis.edu)

Table of Contents

Section S1: Global Model for OC, EC and TC and Bias Investigation

Global Model for OC, EC and TC

A simple calibration design relying on a subset of filters from all sites (Figure S1-1) and seasons was created using 2015 data. Spatiotemporal variability in composition is approximated by retaining every fifth sample to build the model. The Global model, summarized in Table S1-1, shows satisfactory agreement between TOR and FT-IR predicted concentrations with R^2 equal to 0.977, 0.889 and 0.984 for OC, EC, and TC, respectively. Overall, FT-IR OC, EC and TC residuals show a near zero bias with relative values not exceeding 2 – 3 % while the relative error have a moderate spread ranging from 12% (TC) to 27 % (EC). Notably, the inclusion of additional samples into the calibration did not provide better model performance, which indicates that the size of the calibration set is sufficiently large to capture the variability in atmospheric composition across the network.

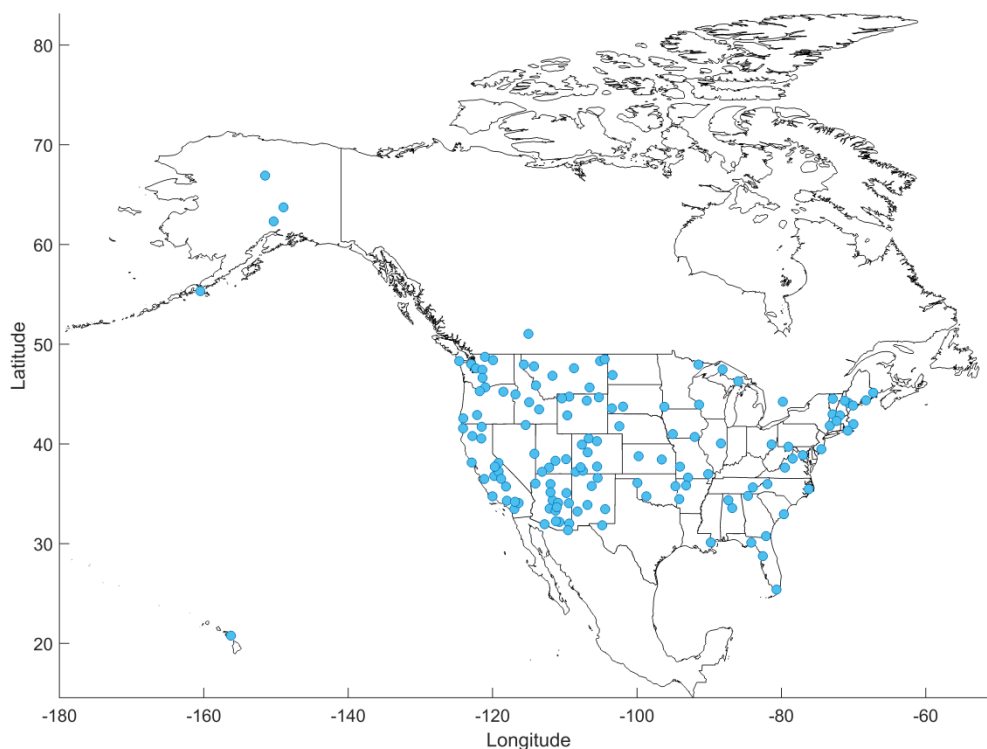



Figure S1 – Spatial distribution of the 161 IMPROVE sites from the year 2015.

Table S1-1 – *Global FT-IR OC, EC and TC model performance for the years 2015, for 19,608 PTFE filters collected at all 161 North American sites, compared with the Global models for OC and EC proposed by Dillner & Takahama featuring 832 filters sampled at 7 sites in 2011.*

	Model	R^2	Bias (μg/m3)	Bias (%)	Error (μg/m3)	Error (%)	MDL (μg/m3)	< MDL (%)
OC	*Global (2015)*	0.977	0.01	1.6	0.08	13.3	0.07	0.5
OC	*Global (2011)*	0.98	0	0	0.08	10.7	0.07	1.1
EC	*Global (2015)*	0.889	0	2.8	0.02	26.8	0.03	13.6
EC	*Global (2011)*	0.98	0	0.3	0.02	17.8	0.03	13.6
TC	*Global (2015)*	0.984	0.01	1.2	0.09	12.0	0.07	0.4

A proof of concept analysis was demonstrated using a similar calibration design at seven IMPROVE sites in 2011. For the sake of comparison, figures of merit associated with the 2011 Global model are reported in Table S1-1 for OC and EC (TC not reported in 2011 paper). Despite the obvious differences in sample population and site number, both 2011 and 2015 models present nearly identical performance with the exception of relative bias and error. The small drop in model performance observed for the 2015 Global model are ascribed to the much broader range of aerosol composition and concentration resulting from expanding the calibration from 7 to 161 sites. For instance, the maximum TOR OC concentration in 2015 (44.6 μg/m^3) is about six times larger than in 2011 (7.3 μg/m^3). The dramatic expansion in carbon content in 2015 is likely connected with biomass burning emissions from wildfires and anthropogenic emissions from the additional three urban sites. To further establish the equivalence between the two models, residuals and their distributions are compared in Figure S1-2, in which FT-IR residuals are compared to the reproducibility of TOR measurements obtained from collocated quartz filter sampling. Noticeably, the interquartile range associated with FT-IR OC and EC residuals for both Global 2011 and 2015 models is consistent with the variability in collocated TOR. More detailed examinations show positive residuals for low predicted concentrations and random residuals less than TOR uncertainty for moderate to high concentrations (below). The combination of these metrics supports the conclusion that generalizing the calibration to the entire network has a negligible impact on prediction accuracy, as long as the calibration contains a sufficient collection of samples representative of the aerosol composition.

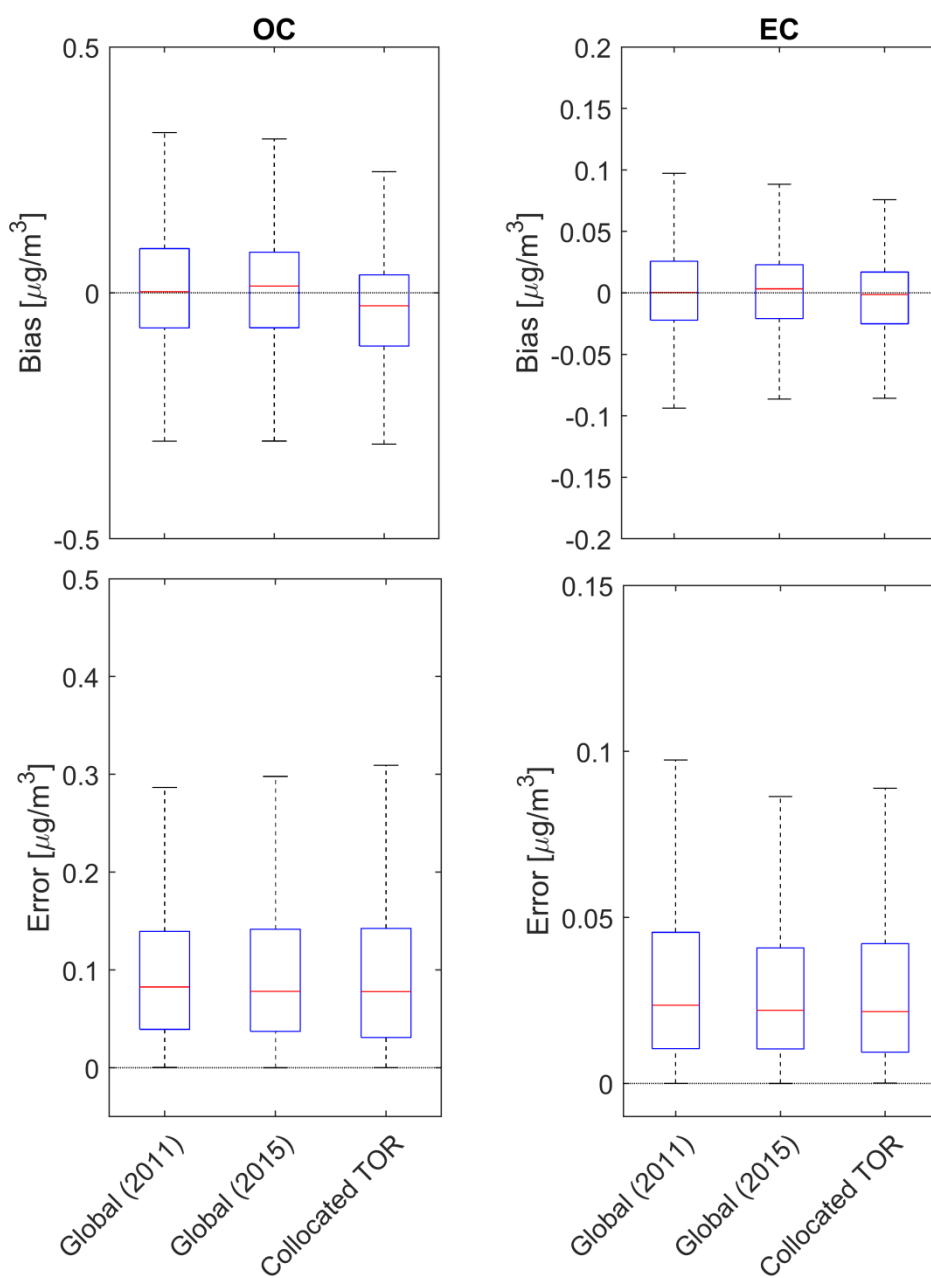



Figure S1-2 – *Bias (top) and error (bottom) for OC (left) and EC (right) for Global model (Dillner et al., 2015 a, b), Global model with all sites in 2015 and collocated TOR. To enhance visualization, samples extending outside ± 1.5 times the interquartile range are not displayed. TC 2015 residuals for FT-IR and Collocated TOR are similar to OC.*

Section S2: TOR EC outlier screening

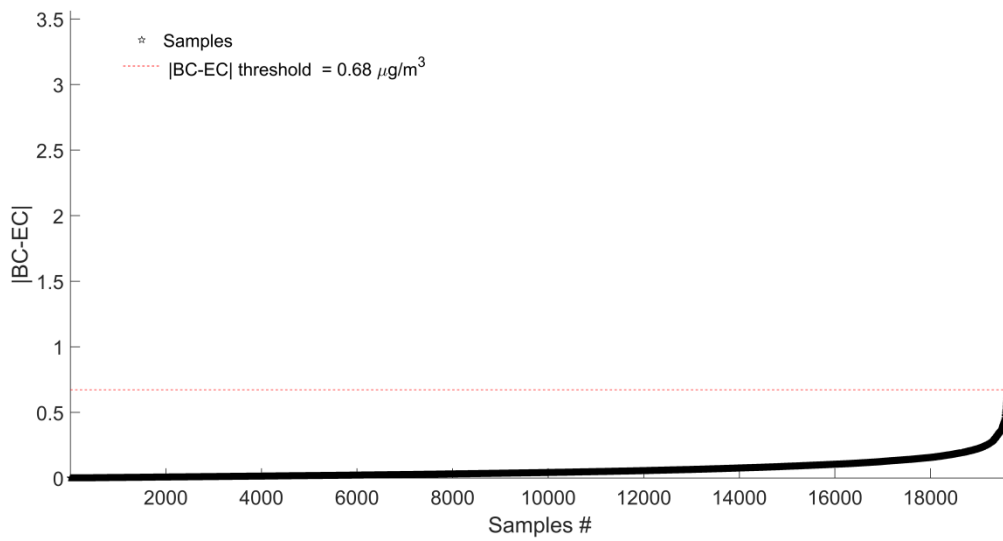



Figure S2-1 – Establishment of the |BC-EC| threshold for the identification of EC outlier samples from the year 2015.

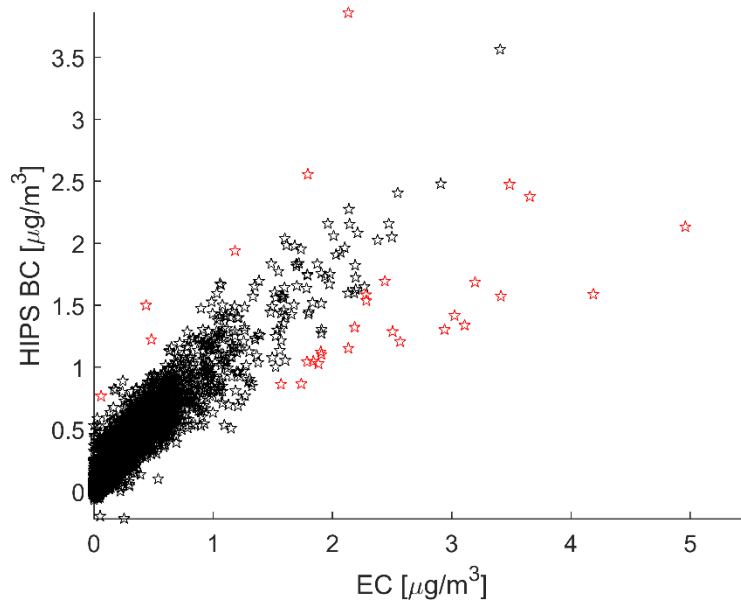



Figure S2-2 – Comparison of HIPS BC and TOR EC concentrations from 2015 samples. The red markers indicate the 33 samples with |BC-EC| > 0.68 µg/m³ significantly deviating from linearity.

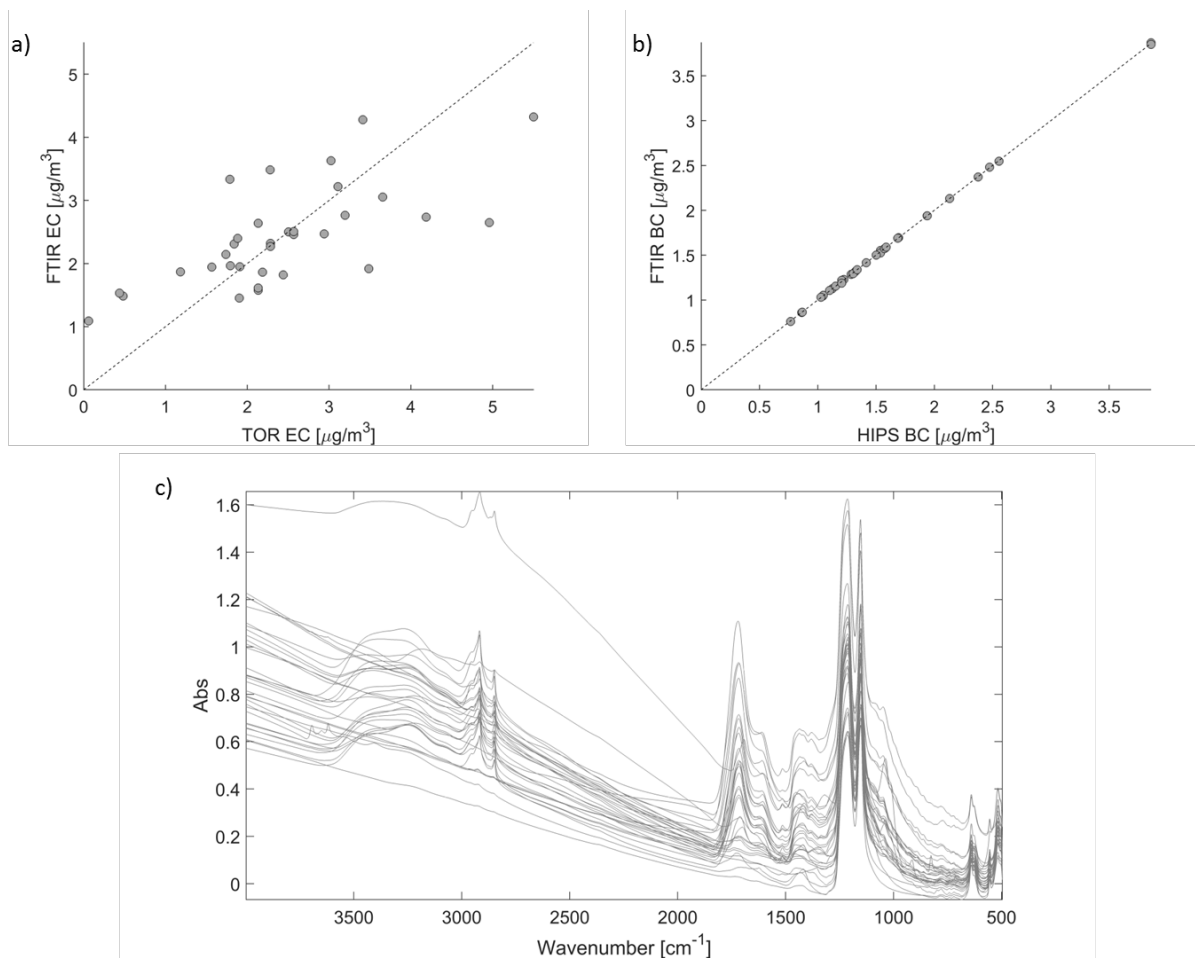



Figure S2-3 – a) FT-IR EC and b) BC predicted concentrations based on the set of 33 sample with |BC-EC| > 0.68 μg/m³ for the year 2015. Although the set of spectra c) is the same for both models, a calibration based on TOR EC values yield poorly predicted concentrations compared to a HIPS BC calibration. FT-IR EC predictions are characterized by an R^2 of 0.459 and a relative bias and error of 1.6 % and 23.7 %, respectively. On the contrary, FT-IR BC predictions feature a R^2 of nearly one, a relative bias of 0.03 % and a relative error of 0.2 %.

Section S3: FT-IR based detection of biomass burning samples

Prior to quantifying the area under hydroxyl, aliphatic and carbonyl functional groups, all characteristic of biomass burning emissions (Popovicheva et al., 2020; Takahama et al., 2011), effects of baseline and offset fluctuations from sample to sample were eliminated. For this purpose, spectral regions featuring hydroxyl – aliphatic (3545 – 2400 cm[-1]) and carbonyl (1820 – 1530 cm[-1]) signatures are baselined separately using an adaptive iteratively reweighted penalized least squares procedure (Eilers, 2003). The area under each baselined spectrum is then integrated in the 1820 – 1660 cm[-1] and the 3445 – 2400 cm[-1] windows. Due to the significant overlap between hydroxyl and aliphatic IR bands (Figure S3-1), an additional baselining step is required to differentiate the areas of the two functional groups. A linear interpolation between 2990 and 2725 cm[-1], corresponding to the edges of the aliphatic peaks, is performed before subtracting the fitted line within the same interval and compute the area under the resulting curve. The area of the hydroxyl group is adjusted accordingly by subtracting the aliphatic area. Finally, the area of each functional group relative to the cumulative hydroxyl – aliphatic – carbonyl area is computed.

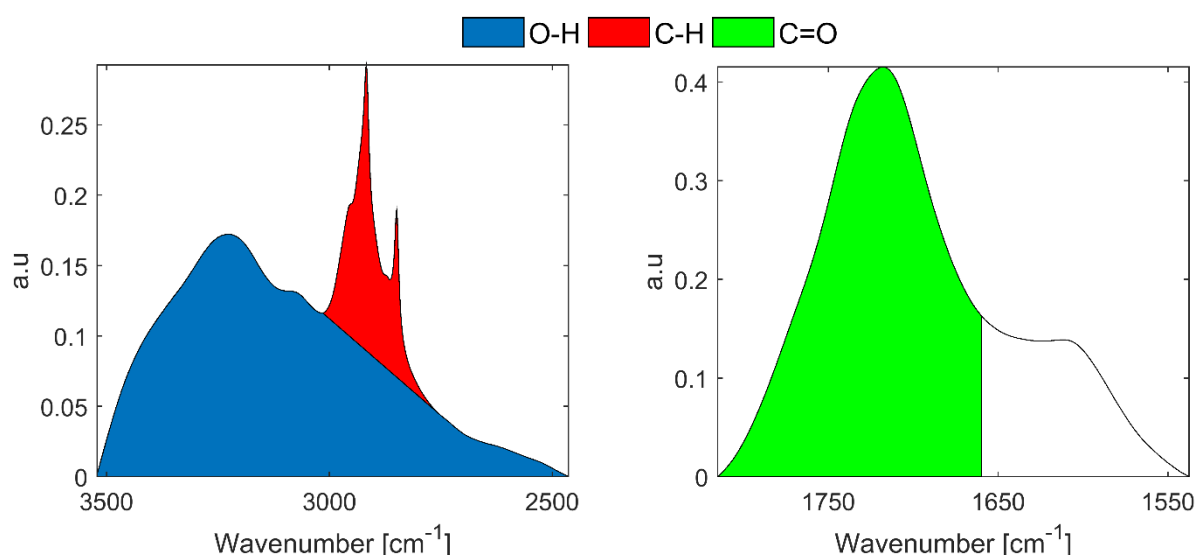



Figure S3-1 – Integrated areas retained for hydroxyl, aliphatic and carbonyl functional groups. Areas are computed following an adaptive iteratively reweighted penalized least squares baselining procedure. The sample associated with the above baselined spectrum was collected at the Barrier Lake site (Canada) on August 28th 2015.

Baselined spectra in the hydroxyl – aliphatic and carbonyl regions were subjected to principal component analysis (Wold et al., 1987) and their respective set of scores was used to compute the squared Mahalanobis distance (D_i^2) per sample for both spectral windows. By performing this procedure for each site individually, D_i^2 gives us an indication of how similar or dissimilar spectra are compared to the mean spectrum.

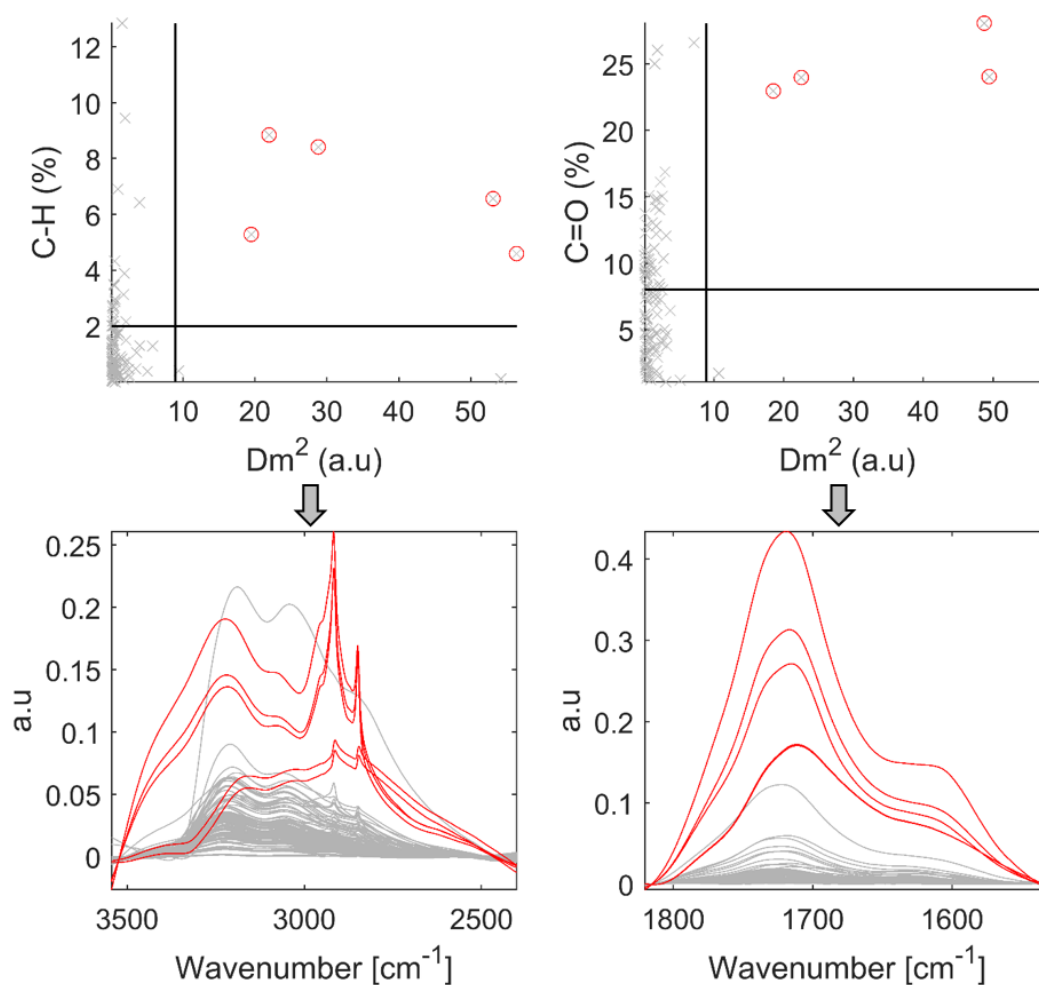



Figure S3-2 – Top plots show squared Mahalanobis distance vs % CH (left) and % C=O (right) for one year of PTFE filters collected at the Fort Peck Indian Reservation (MT) site in 2015. Samples that meet both dissimilarity and FG area criteria are circled in red. Using the same color code, the bottom plots show the corresponding FT-IR spectrum associated with each sample.

The figures below (Figure S3-3 through S3-5) provide information about the samples that were selected as biomass burning for each sample year. The high carbon content, OC/EC ratio, location, season and spectra all confirm that the selected spectra are likely contain biomass burning aerosol.

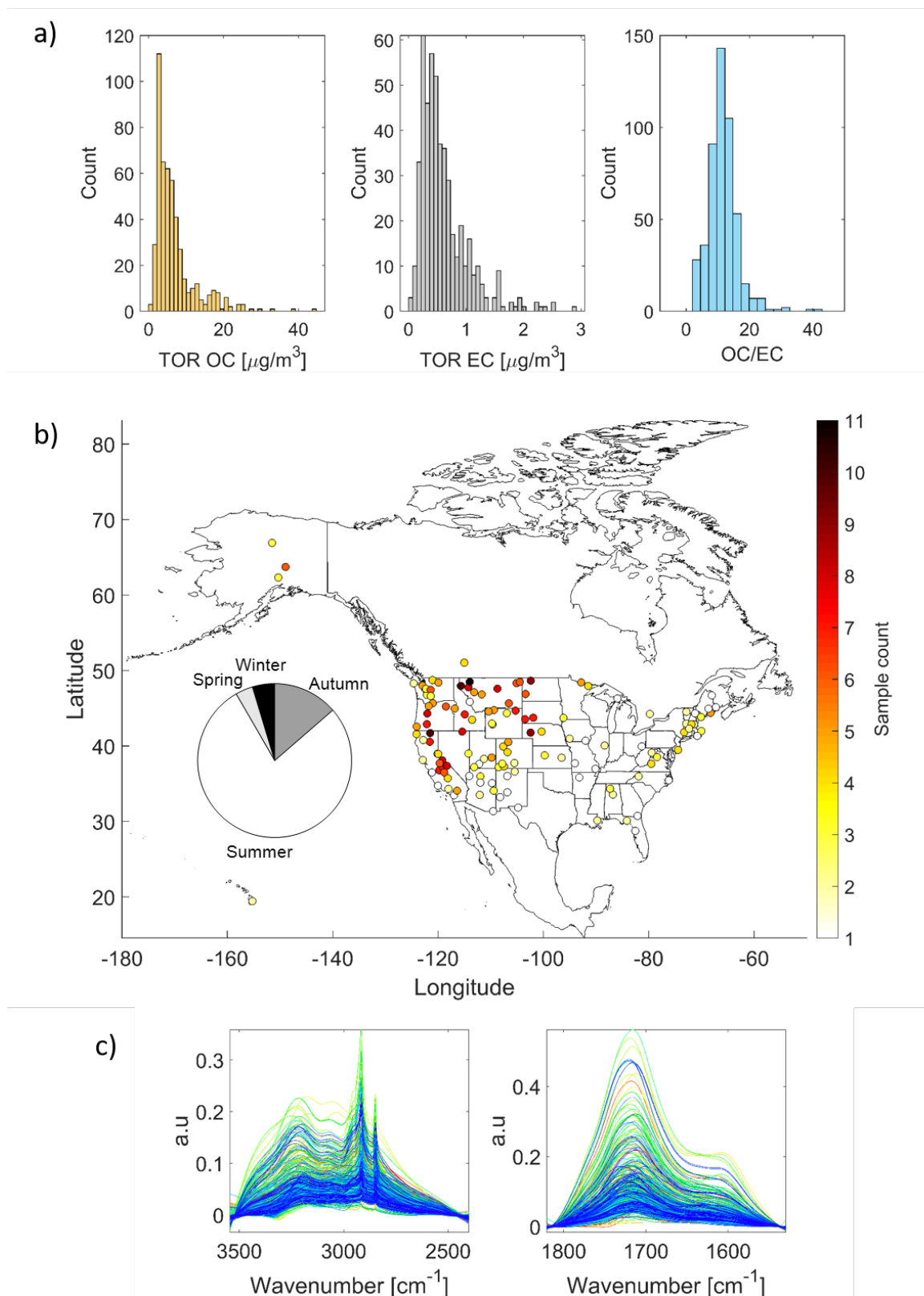



Figure S3-3 – a) Distribution in carbon concentrations and OC/EC ratios for the 492 samples classified as biomass burning in 2015, b) their corresponding spatial and seasonal distribution and c) associated baselined spectra in the hydroxyl – aliphatic (left) and carbonyl (right) functional group regions.

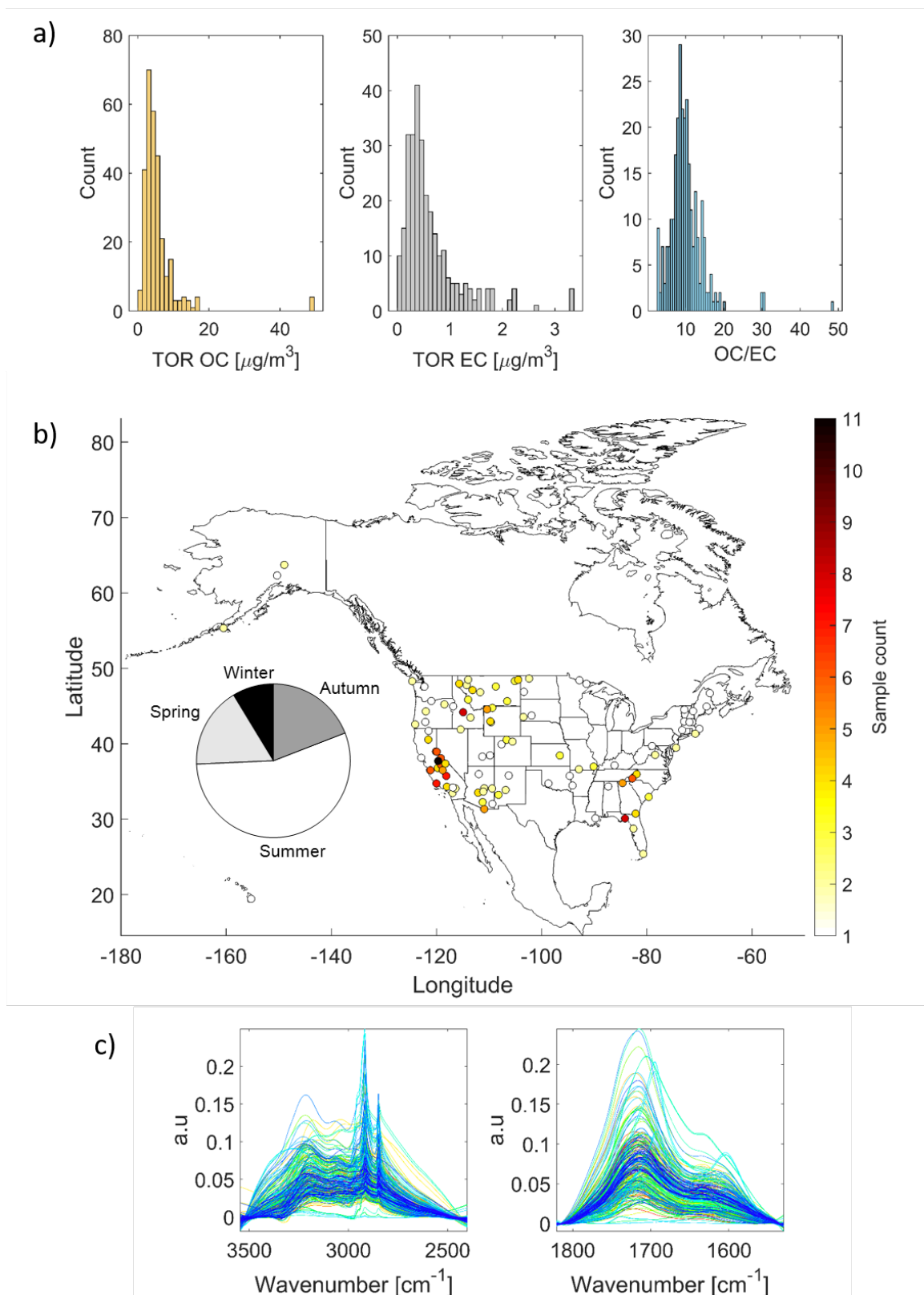



Figure S3-4 – a) Distribution in carbon concentrations and OC/EC ratios for the 288 samples classified as biomass burning in 2016, b) their corresponding spatial and seasonal distribution and c) associated baselined spectra in the hydroxyl – aliphatic (left) and carbonyl (right) functional group regions.

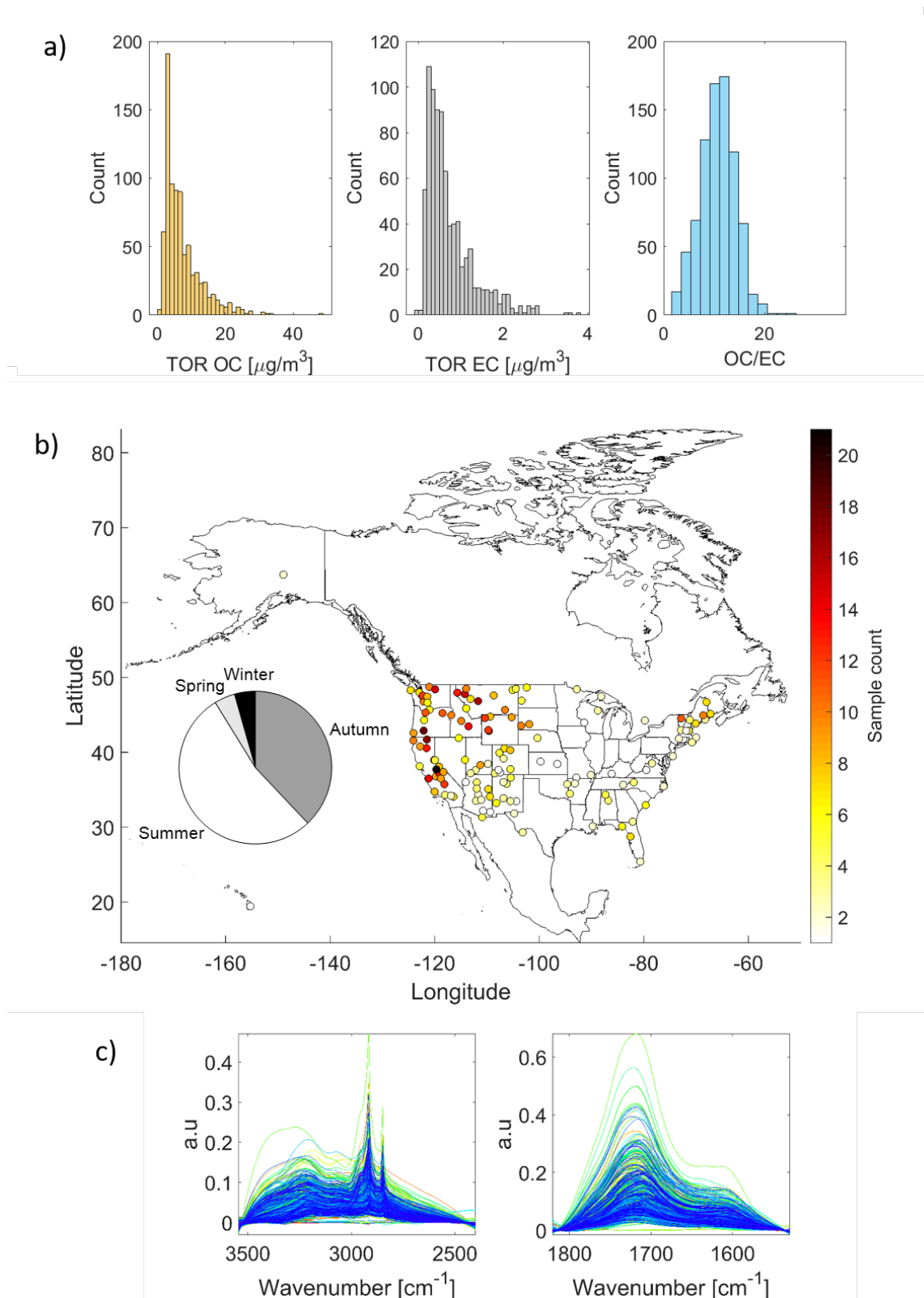



Figure S3-5 – a) Distribution in carbon concentrations and OC/EC ratios for the 817 samples classified as biomass burning in 2017, b) their corresponding spatial and seasonal distribution and c) associated baselined spectra in the hydroxyl – aliphatic (left) and carbonyl (right) functional group regions.

Section S4: Gaussian Mixture Model (GMM) and Cluster Interpretation

GMM Methodology

A GMM was developed to cluster FT-IR spectra based on PM$_{2.5}$ composition. A GMM assumes an FT-IR spectrum's probability density function may be approximated as a weighted sum of multivariate normal distributions (Bilmes, J. A., 1998; Hastie, T et al., 2009). Specifically, an individual spectrum (x_i) is assigned to the Gaussian-cluster (j) that maximizes the following conditional distribution

$$p(j|x_i, \hat{\mu}_j, \hat{\Sigma}_j) = \frac{\hat{\pi}_j N(x_i|\hat{\mu}_j, \hat{\Sigma}_j)}{\sum_{k=1}^{K} \hat{\pi}_k N(x_i|\hat{\mu}_k, \hat{\Sigma}_k)} \tag{1}$$

where $p(j|x_i, \hat{\mu}_j, \hat{\Sigma}_j)$ is the conditional probability that spectrum i belongs to cluster j, π_k is the mixture component (or weight) associated with the kth Gaussian ($\sum \pi_k = 1$), and $N(x|\mu_k, \Sigma_k)$ the kth Gaussian distribution with mean μ_k and variance-covariance matrix Σ_k.

To minimize the number of parameters estimated during GMM fitting, dimensionality reduction the FT-IR spectra was performed. First, raw spectra were transformed to second derivative spectra using a 2nd order, 21 point, Savitzky-Golay filter (Savitzky and Golay, 1964) and differenced with filter blank spectra. Each blank-corrected spectrum was divided by its Euclidean norm to minimize the effect of total aerosol mass concentration and emphasize aerosol composition (Bro and Smilde, 2003). The following ranges in the FT-IR spectra were included as they contain the predominant organic, inorganic, and mineral aerosol absorption bands: 3500 – 2650 cm^{-1}, 1800 – 1350 cm^{-1}, and 1050 – 800 cm^{-1} (Allen et al., 1994). Finally, a principal component analysis (PCA) was applied to the spectra with the optimal number of components estimated using five-fold cross validation (Abdi and Williams, 2010). Nine principal components were retained for analysis with the principal component scores used as inputs to the GMM.

The number of Gaussians (clusters) K was determined empirically as follows. First, K is set equal to one and the Expectation-Maximization (EM) algorithm is initialized with random parameter values (Dempster et al., 1977; Fraley and Raftery, 2002). Next, parameters are estimated using the EM algorithm and log-likelihoods recorded. GMM fitting is repeated 100 times for distinct randomizations to prevent the EM algorithm from converging to local maxima. The preceding steps are repeated for 2 through 40 Gaussians yielding a total of 4000 models. The optimum GMM model is selected according to a minimizing Bayesian Information Criterion (Schwarz, 1978). For the 2015 dataset, the optimal number of clusters is met for K = 21.

GMM dossiers for cluster interpretation

The prototypical spectra associated with each of the 21 GMM clusters were first evaluated by considering the dominant absorption bands in Field Blank Corrected (FBC) spectra. As the interpretation of derivative spectra is difficult for the infrared spectra of filter-bound PM, summary measures associated with each cluster were calculated from trace elements and ions, geographical dispersion plotted and assessed, and seasonal information aggregated to provide a complete picture of the aerosol selected by the GMM. This information was collected together into "dossiers" for each of the 21 clusters. To only extract the main composition related information required for cluster interpretation, a restricted number of tracers is considered (n = 15). Titles were given by the authors to reflect their judgment as to overall sources attributable (on average) to a given cluster. The first dossier is explained in detail in the figure caption below. Others are considered self-explanatory given a perusal of the dossier subplots and title.

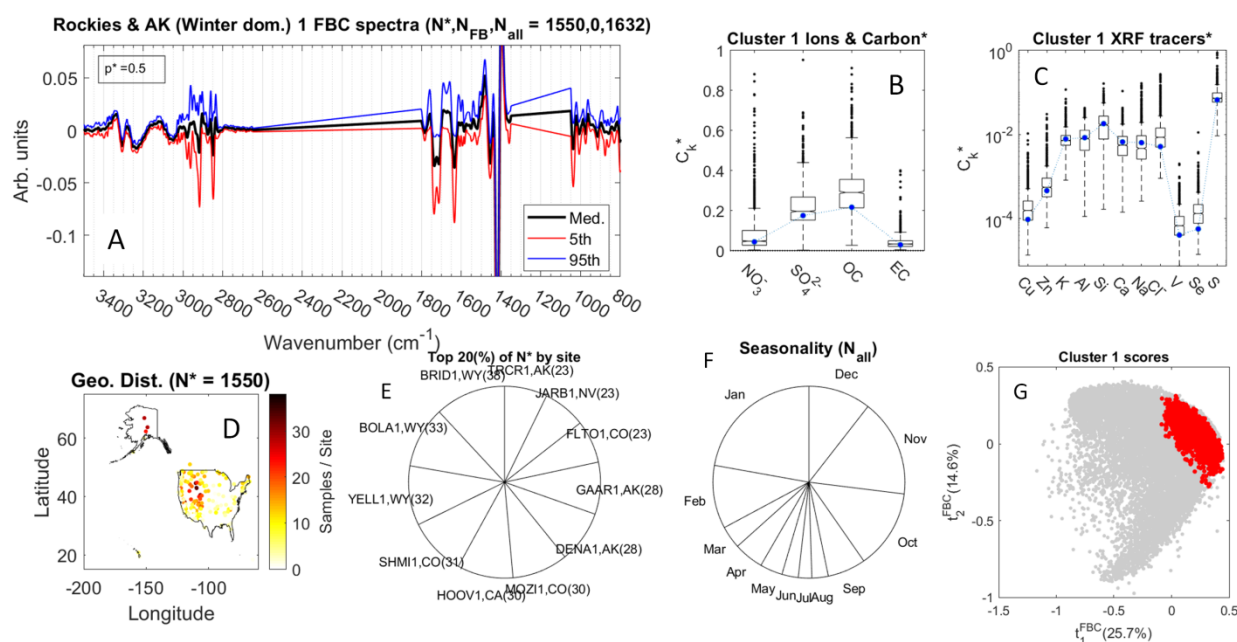



Figure S4-1a – GMM dossier associated with cluster n°1.

Description:

- The title, *i.e.*, "Rockies & AK (winter dom.) 1 FBC spectra (N^*,N_{FB},N_{all}=1550,0,1632)" relays the interpretation of the cluster, any seasonal information, and spectra preprocessing. The title also indicates in brackets the number of IMPROVE spectra above the conditional probability criterion (N^*; p^*=0.5), the number of field blanks in the cluster (N_{FB}), and the total number of spectra cluster by the GMM (N_{all}). In this cluster, most spectra (1550/1632) are considered to have "prototypical" composition.

- Panel A shows the median, 5th and 95th percentile FBC spectra.

- Panel B summarizes the distribution of nitrate, sulfate, OC, and EC in these samples, normalized to total PM$_{2.5}$ (C_k^*). Blue dots connected by dashed lines show median normalized concentrations for the entire IMPROVE network. Minimum detection limits (MDLs) for each aerosol species were downloaded from the FED database on November 21, 2018. Measured

concentrations falling below the MDL were imputed as the MDL/2 in order to generate meaningful boxplots and subsequent pie charts.

- Panel C follows similar scaling rules for trace elements associated with traffic emissions (Cu, Zn), biomass burning and soil (K, Al, Si, Ca), marine aerosol (Na, Cl⁻), shipping (V), and coal combustion (Se, S).

- Panel D shows the geographic distribution of samples per site in cluster n°1. The figure is false colored according to the concentration of samples in a site. Here, samples are clustered primarily in the Rocky Mountains and Alaska.

- Panel E presents a pie chart ranking each site according to the number of samples that fell into this cluster. Samples must have fallen above the $p* = 0.5$ criterion level implying that they should best exemplify (on average) the sources attributable to this cluster. In the interest of space and readability, on the sites containing approximately 20% of the data (0.2*1550) were plotted. Here, we see that the top 3 sites were all located in Wyoming with 38, 33, and 32 while the other two are located in Colorado (31) and California (30).

- Panel F shows that samples in this cluster were typically collected in late fall to early spring (~75%). About half of all samples were collected in the winter (Nov, Dec, Jan).

- Panel G shows the location of the cluster on the first two principal component scores derived from normalized FBC spectra (t_1^{FBC}, t_2^{FBC}). Note that this only represents the first (dominant) two-dimensional principal subspace as nine components were used for GMM clustering.

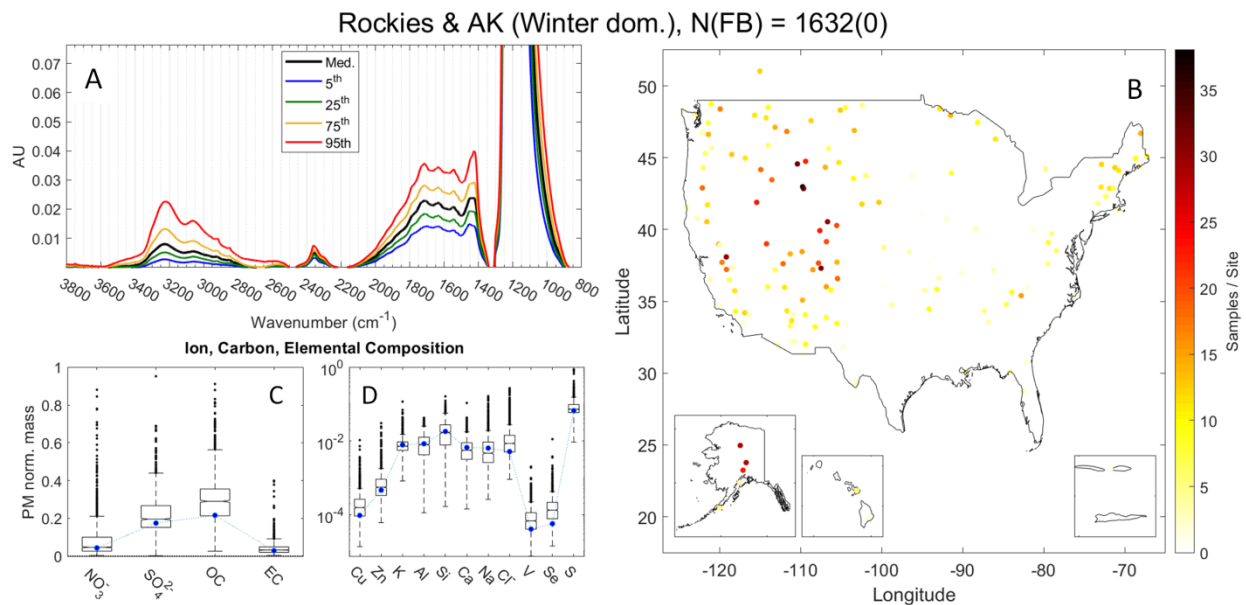



Figure S4-1b – Auxiliary GMM dossier associated with cluster n°1. Here, baseline corrected absorption spectra (A), expanded map (B) and tracer information are plotted (C, D).

Cluster nº2:

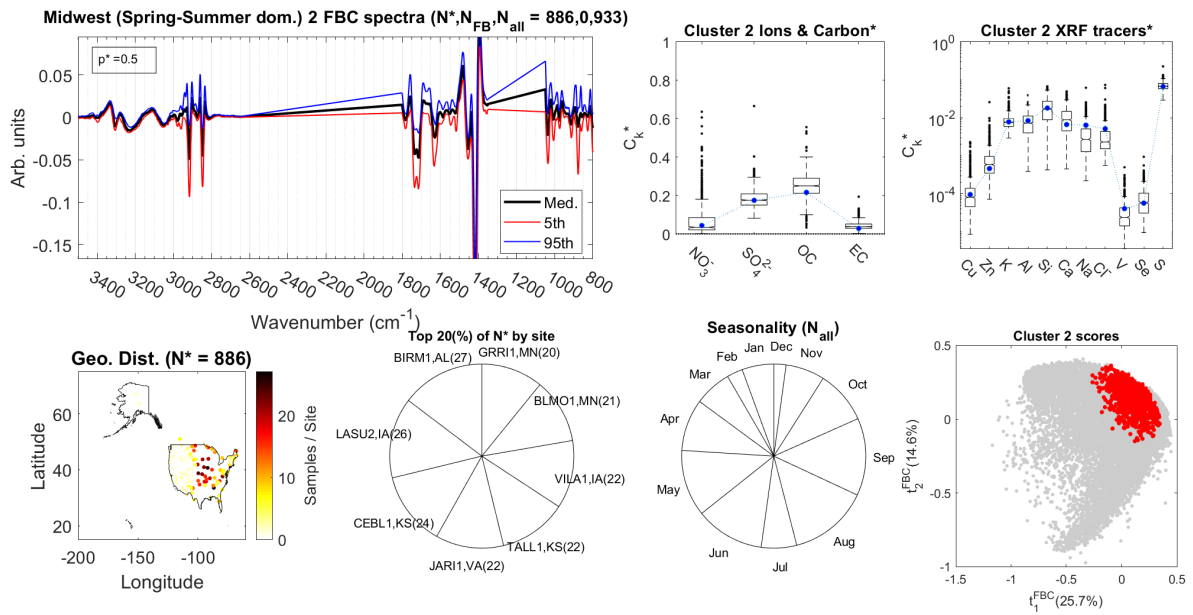



Figure S4-2a – GMM dossier associated with cluster nº2.

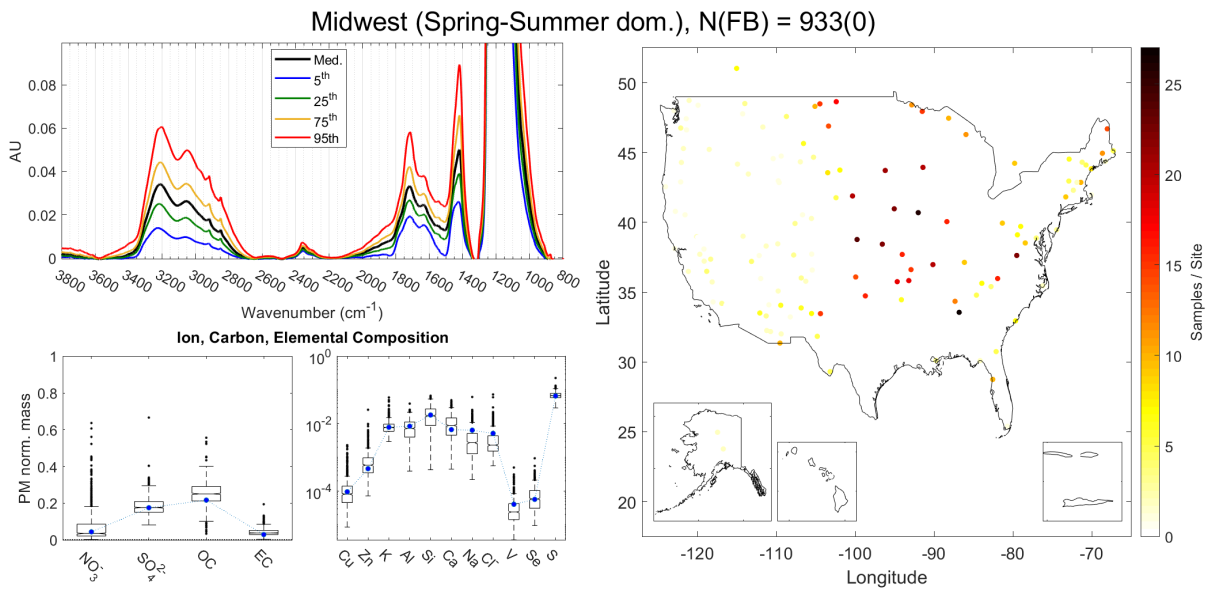



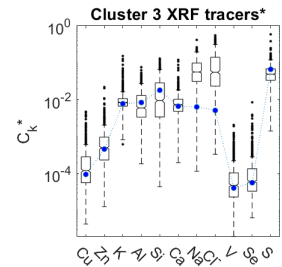
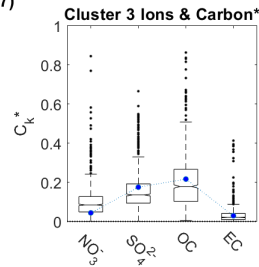
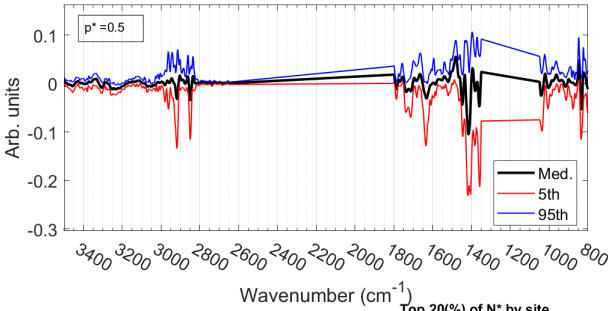
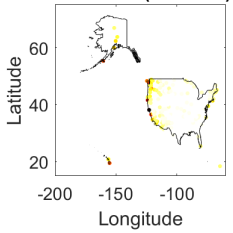
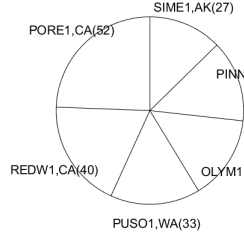
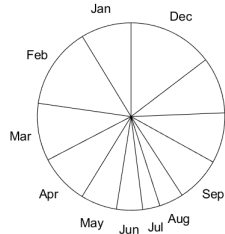
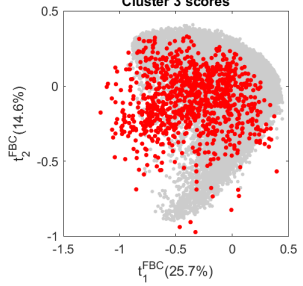
Figure S4-2b – Auxiliary GMM dossier associated with cluster nº2.

Cluster nº3:

Figure S4-3a – GMM dossier associated with cluster nº3.

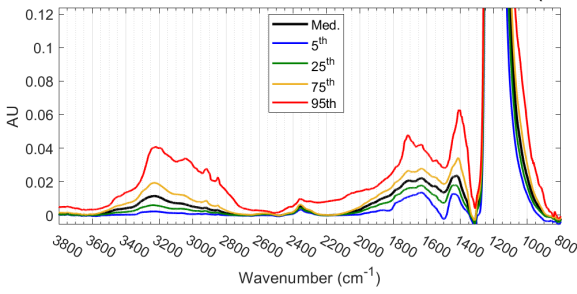
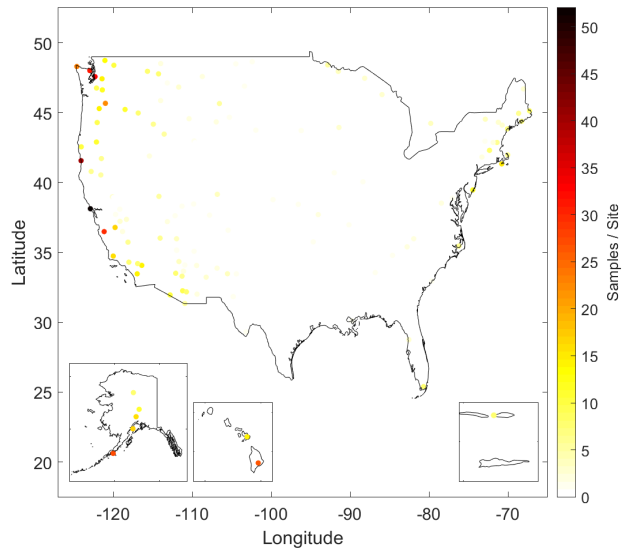
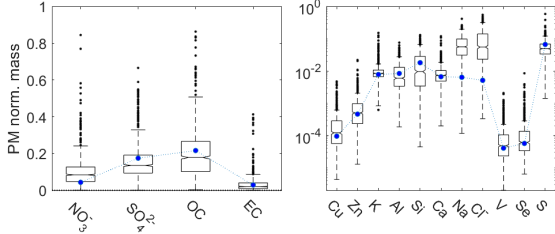



Figure S4-3b – Auxiliary GMM dossier associated with cluster nº3.

Cluster nº4:

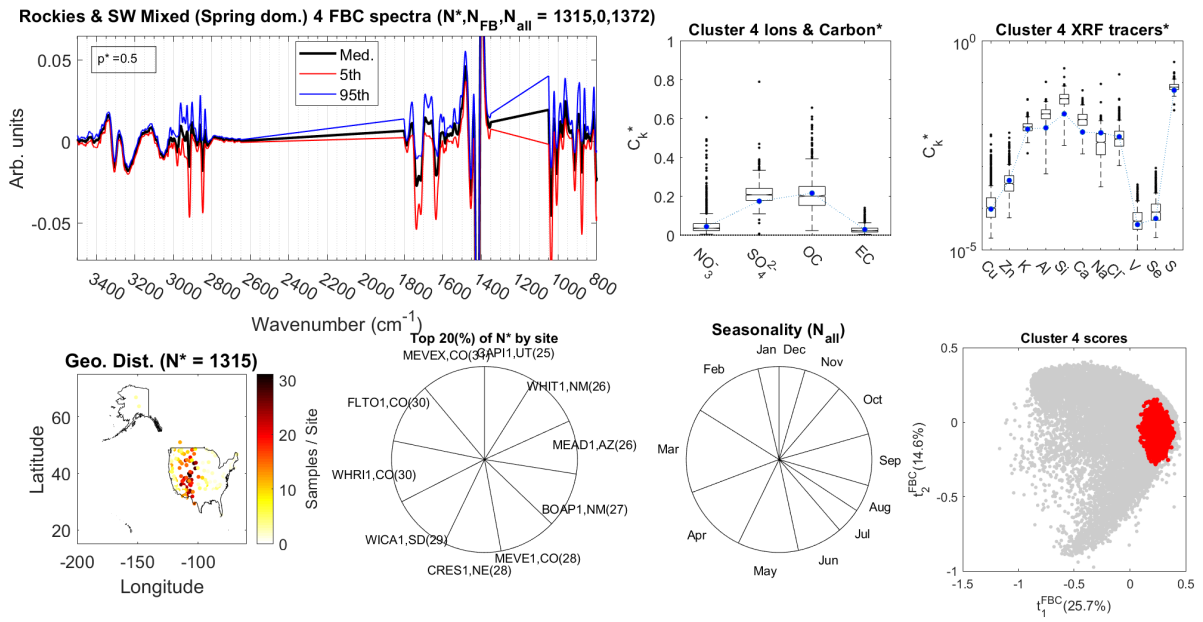



Figure S4-4a – GMM dossier associated with cluster nº4.

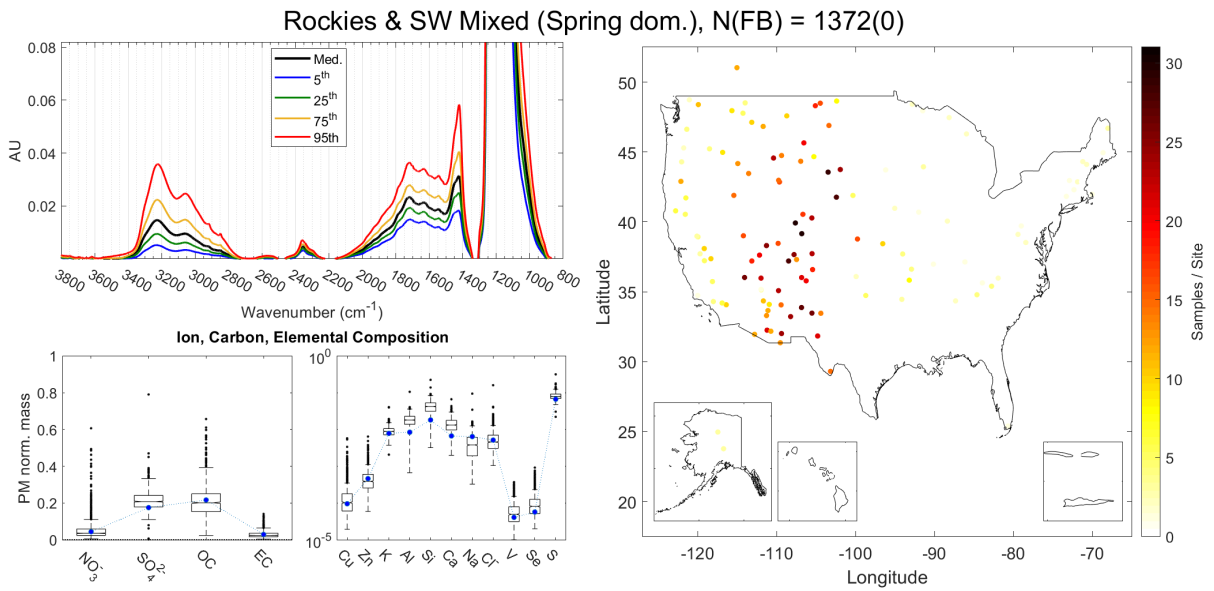



Figure S4-4b – Auxiliary GMM dossier associated with cluster nº4.

Cluster nº5:

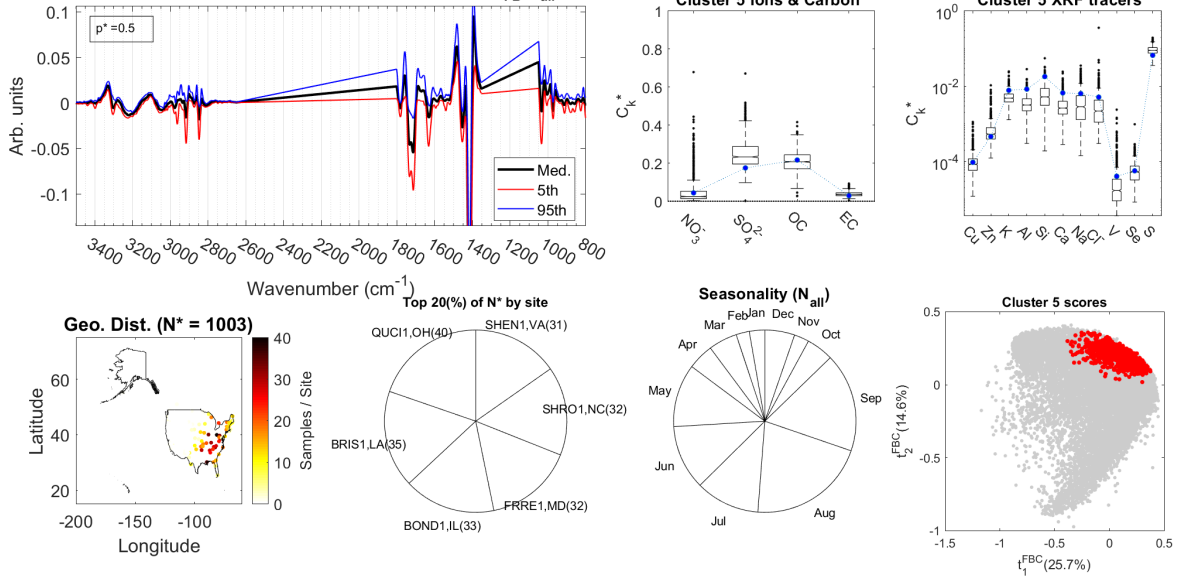



Figure S4-5a – GMM dossier associated with cluster nº5.

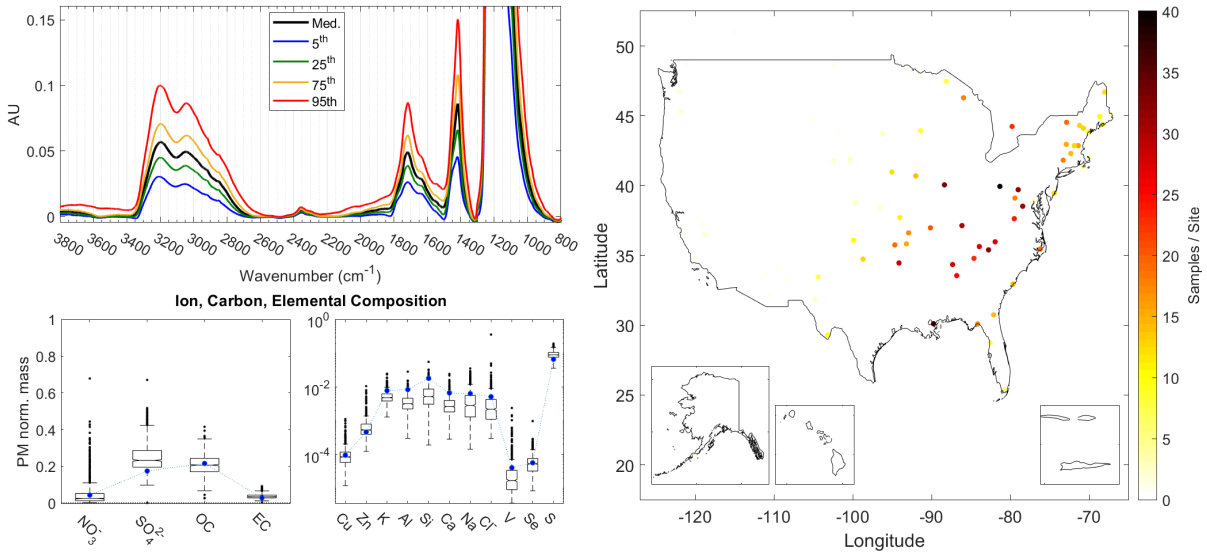



Figure S4-5b – Auxiliary GMM dossier associated with cluster nº5.

Cluster nº6:

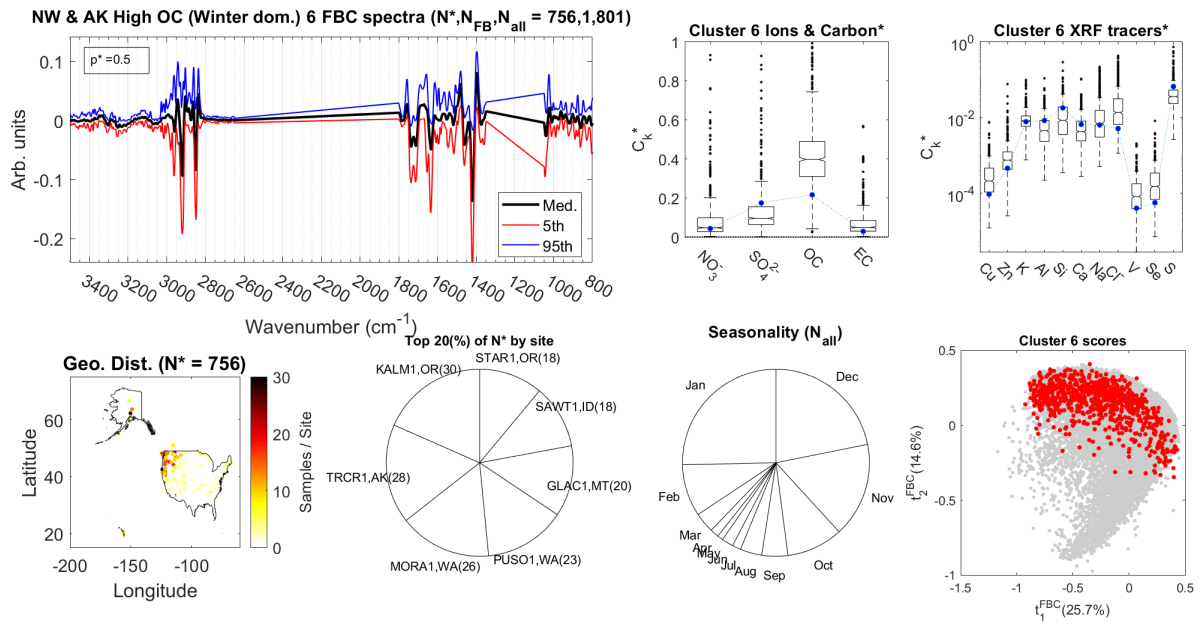



Figure S4-6a – GMM dossier associated with cluster nº6.

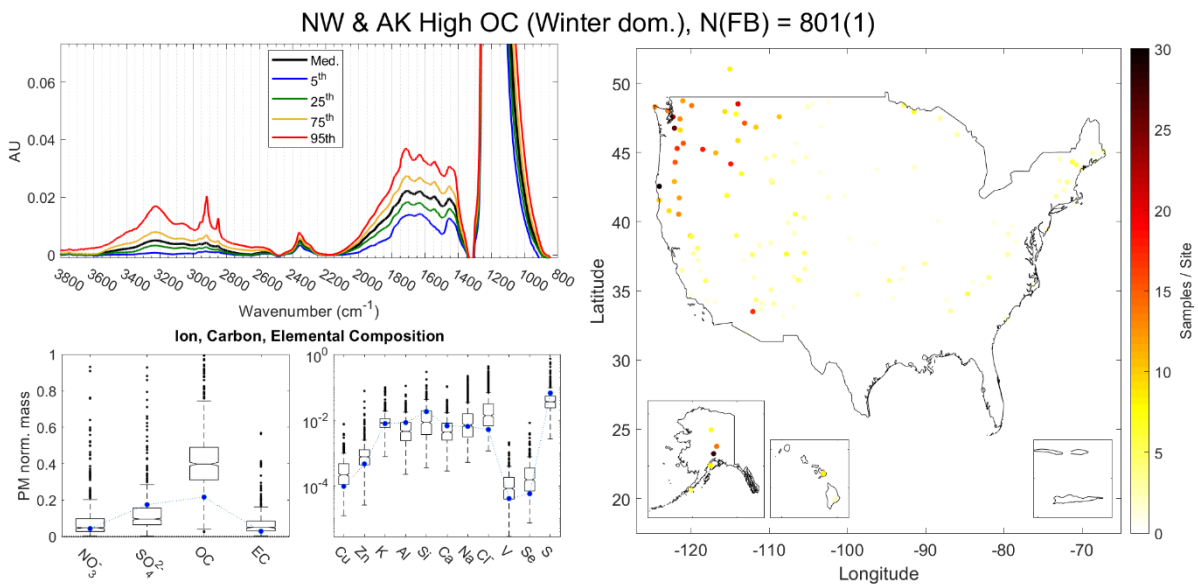



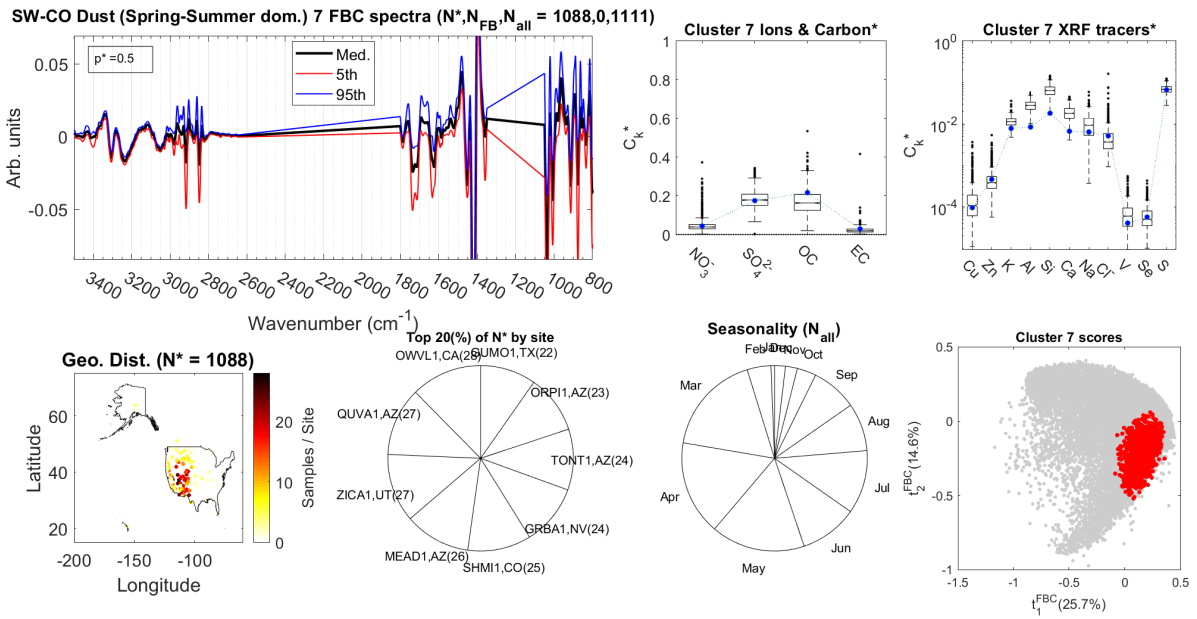
Figure S4-6b – Auxiliary GMM dossier associated with cluster nº6.

Cluster nº7:

Figure S4-7a – GMM dossier associated with cluster nº7.

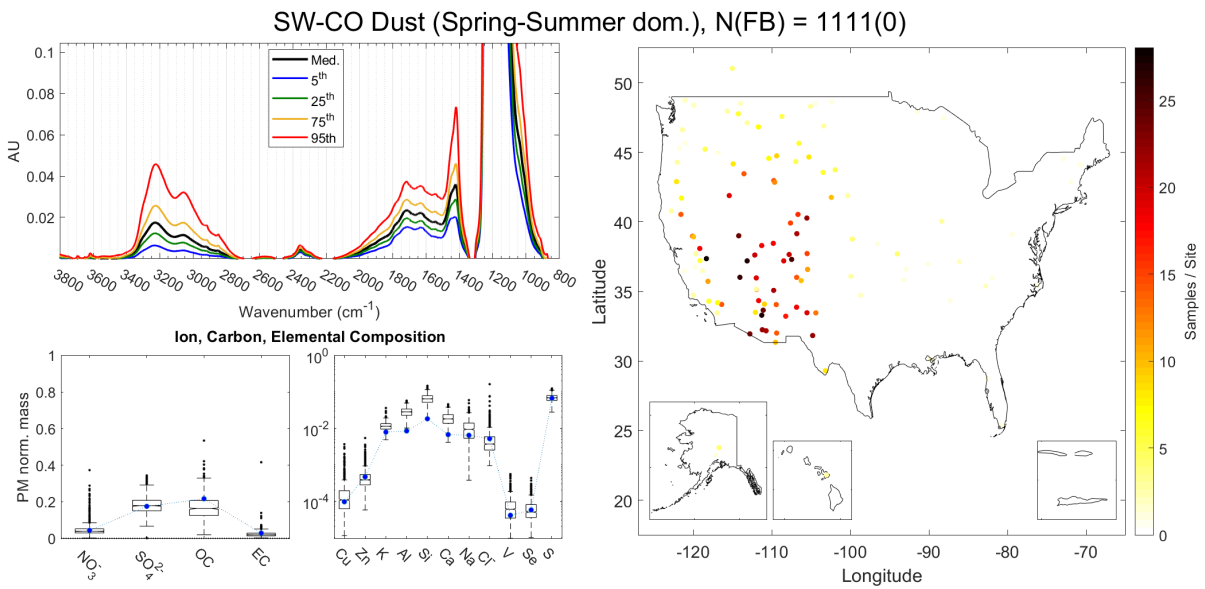



Figure S4-7b – Auxiliary GMM dossier associated with cluster nº7.

Cluster nº8:

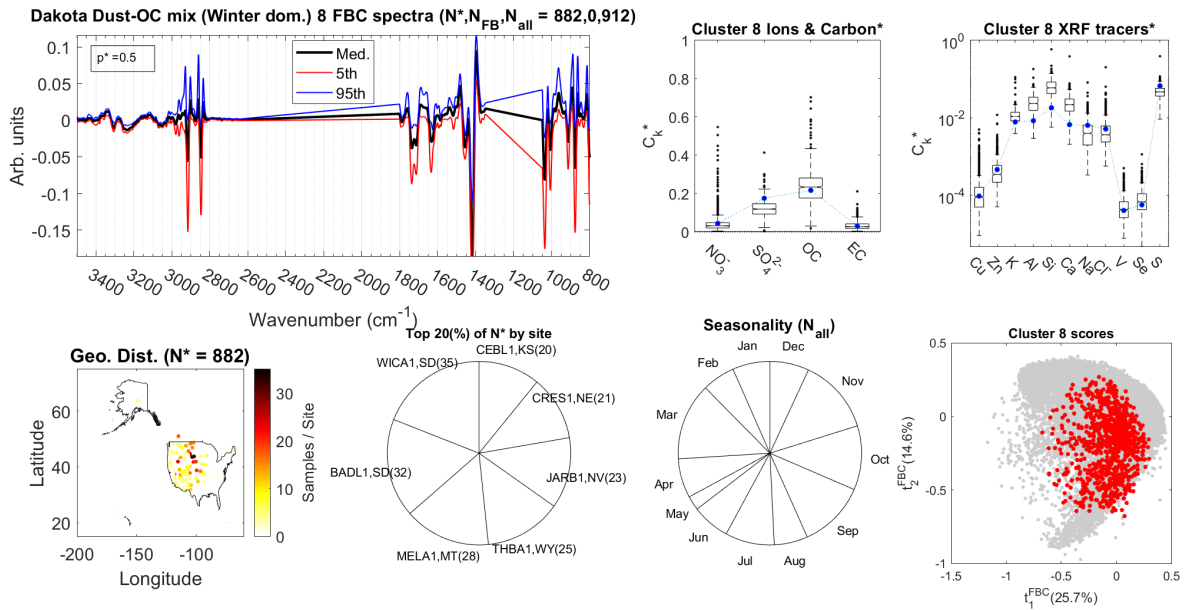



Figure S4-8a – GMM dossier associated with cluster nº8.

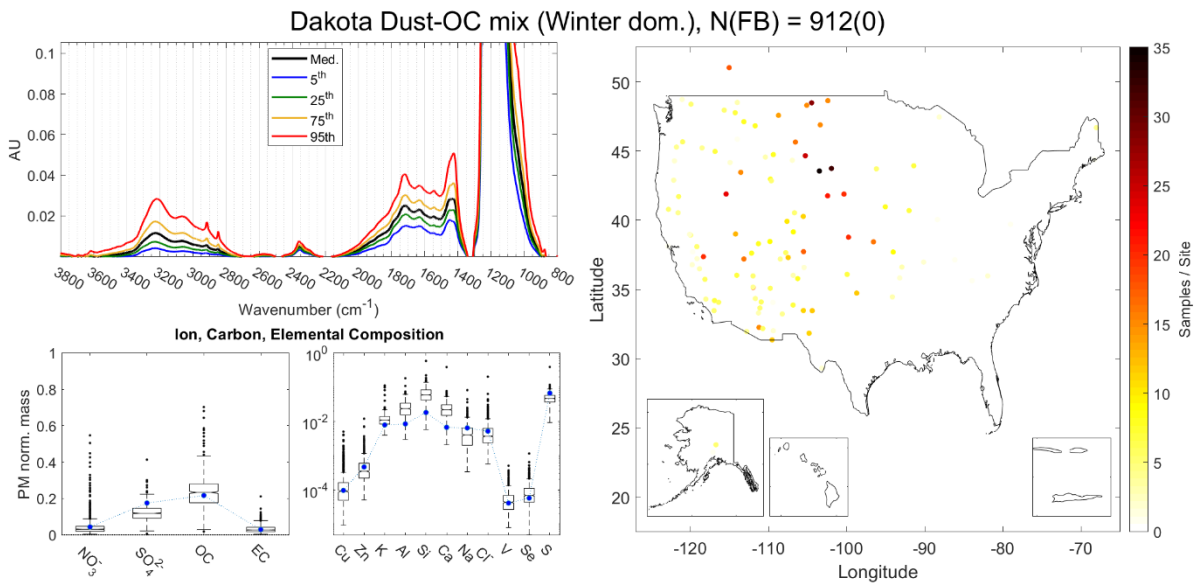



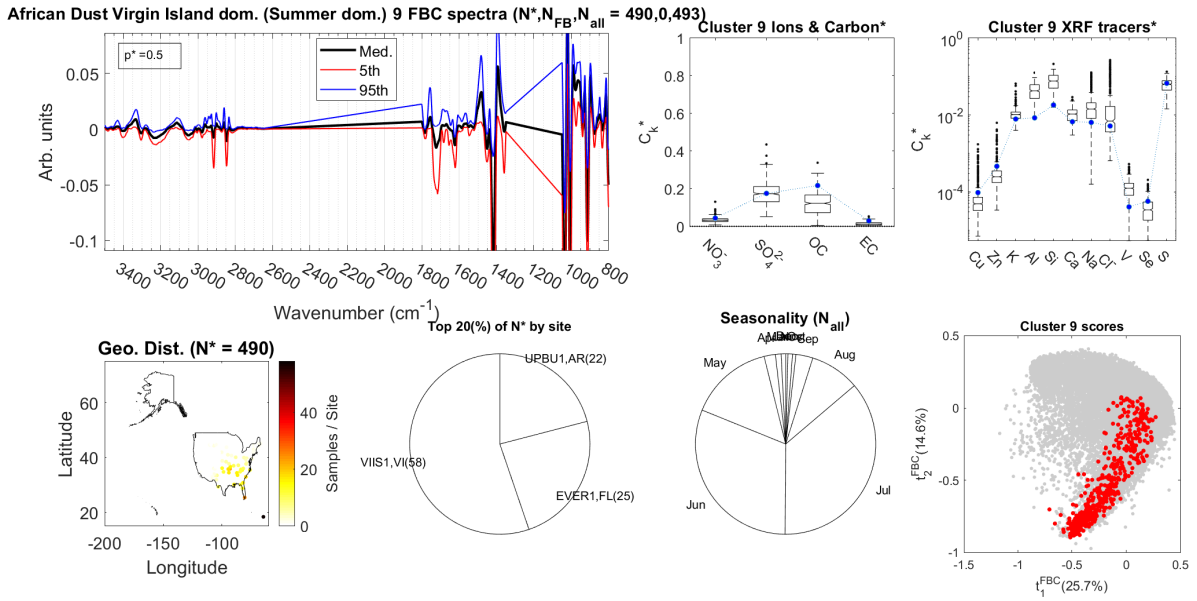
Figure S4-8b – Auxiliary GMM dossier associated with cluster nº8.

Cluster nº9:

Figure S4-9a – GMM dossier associated with cluster nº9.

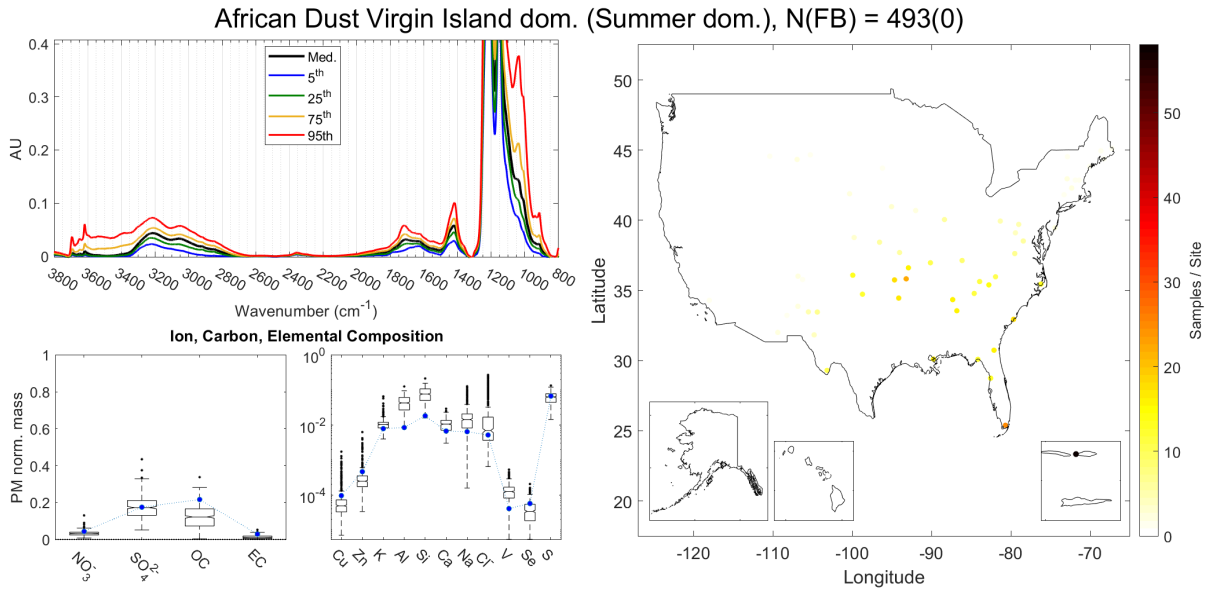



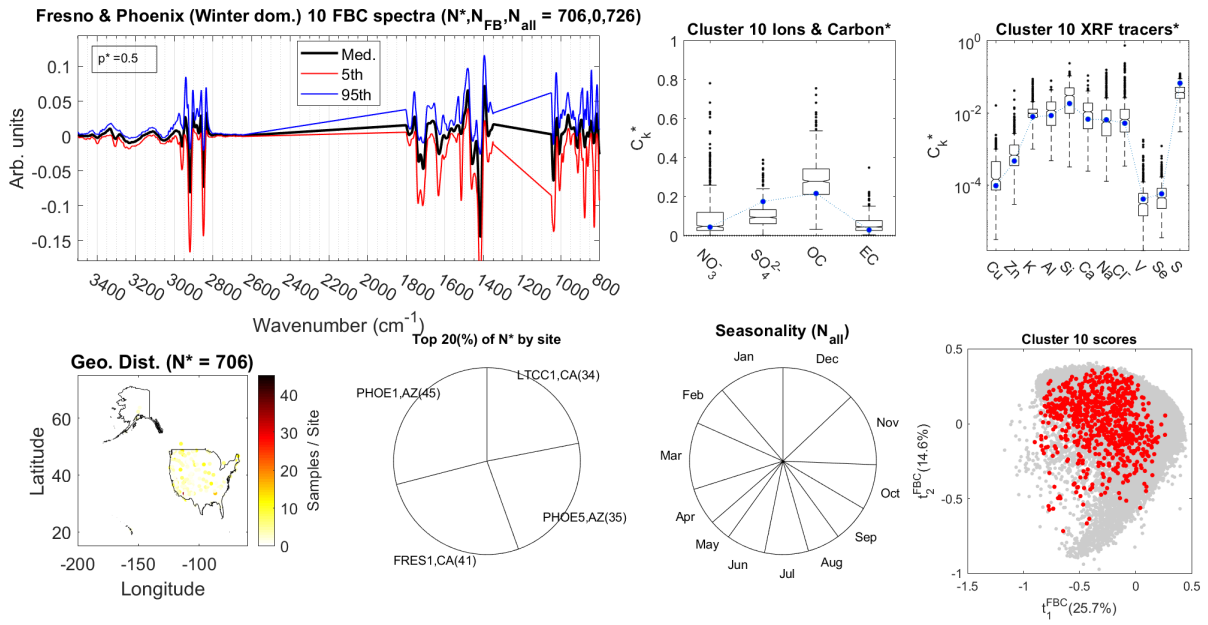
Figure S4-9b – Auxiliary GMM dossier associated with cluster nº9.

Cluster nº10:

Figure S4-10a – GMM dossier associated with cluster nº10.

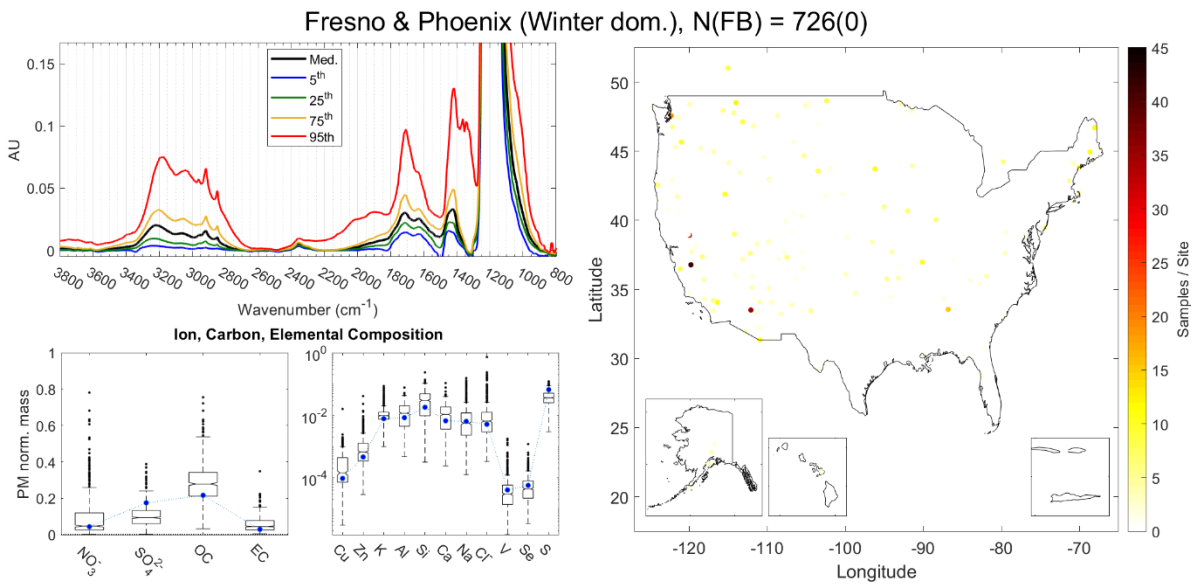



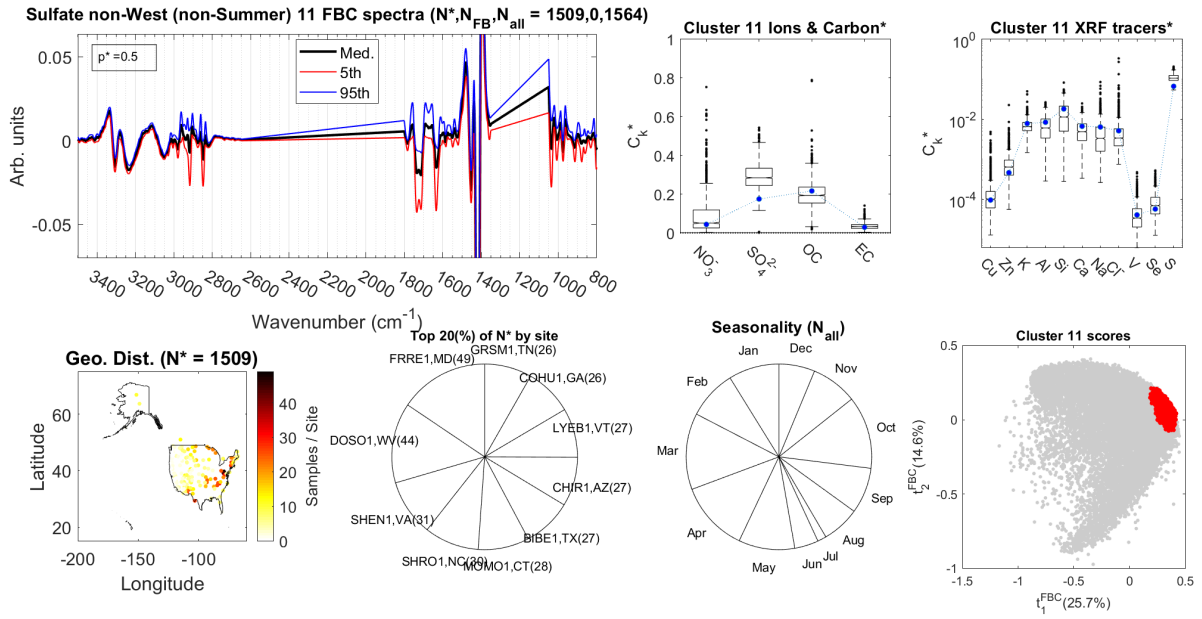
Figure S4-10b – Auxiliary GMM dossier associated with cluster nº10.

Cluster nº11:

Figure S4-11a – GMM dossier associated with cluster nº11.

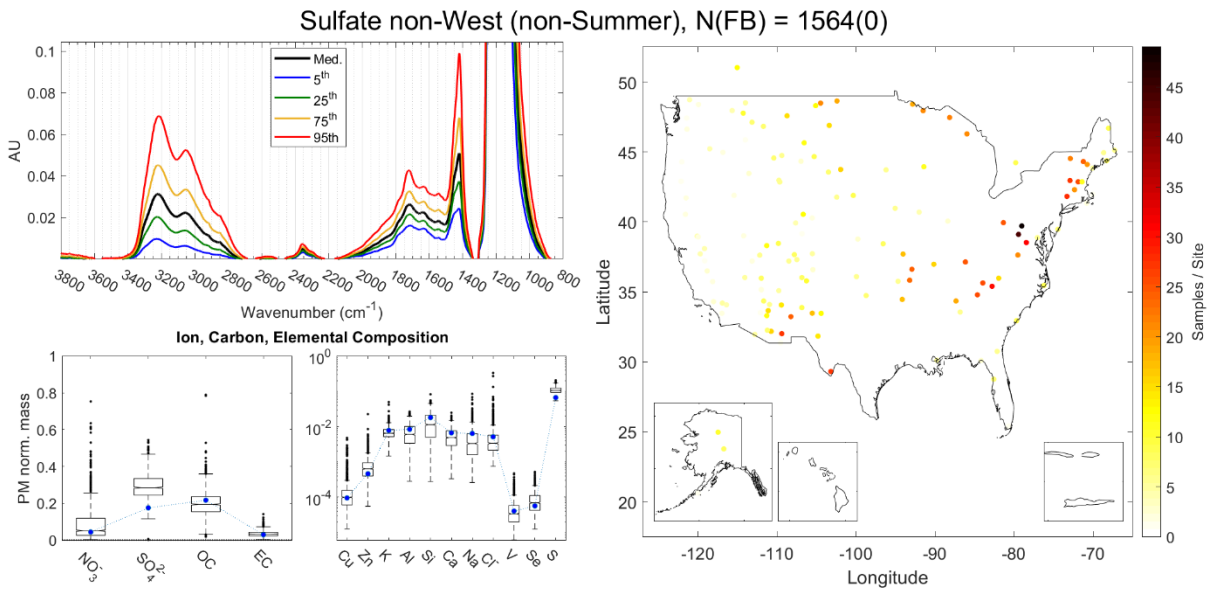



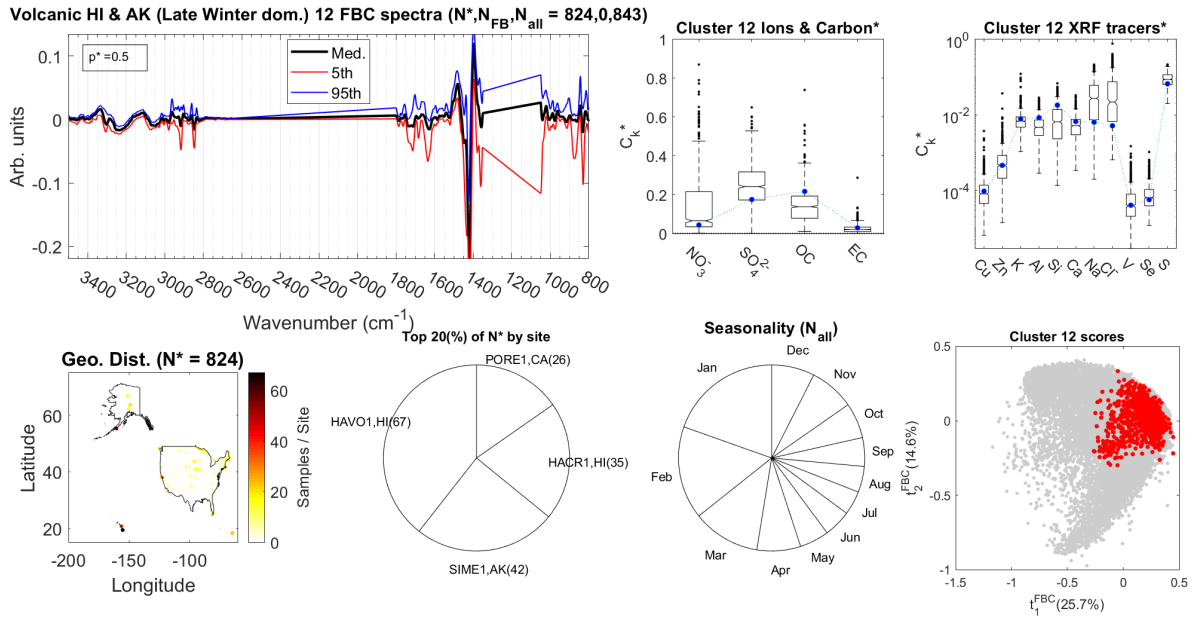
Figure S4-11b – Auxiliary GMM dossier associated with cluster nº11.

Cluster nᵒ12:

Figure S4-12a – GMM dossier associated with cluster nᵒ12.

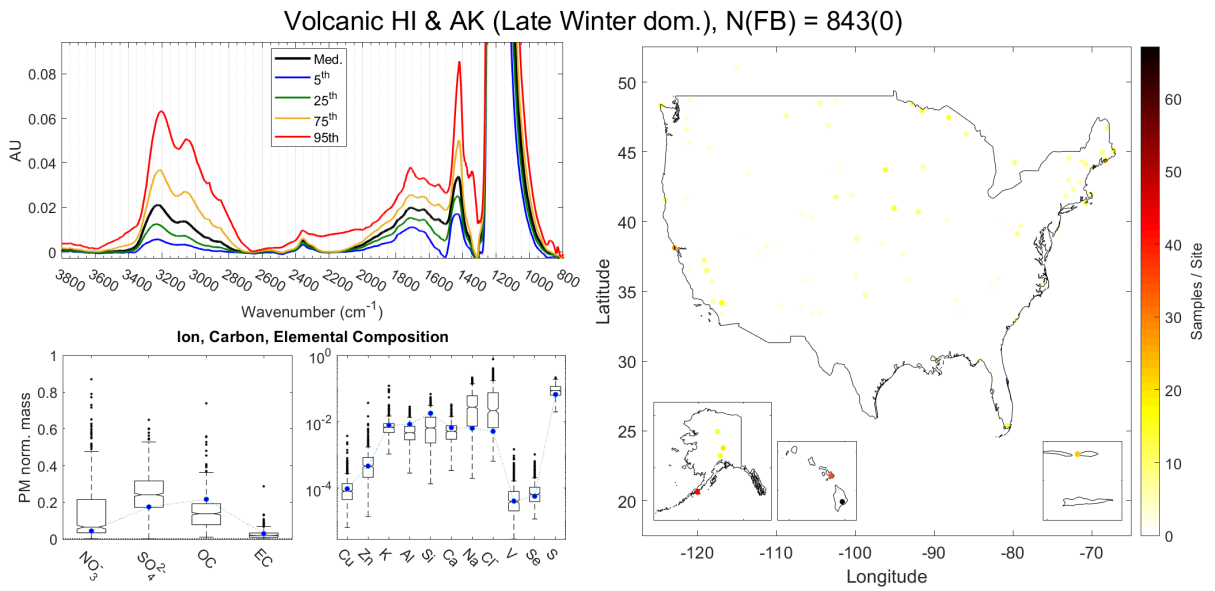



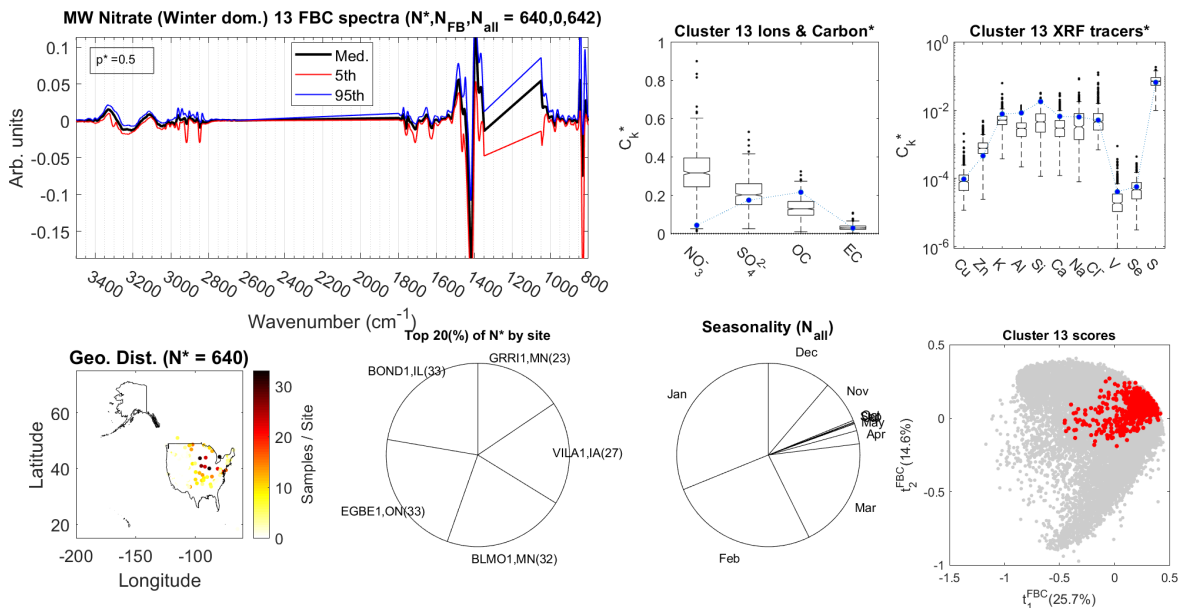
Figure S4-12b – Auxiliary GMM dossier associated with cluster nᵒ12.

Cluster n°13:

Figure S4-13a – GMM dossier associated with cluster n°13.

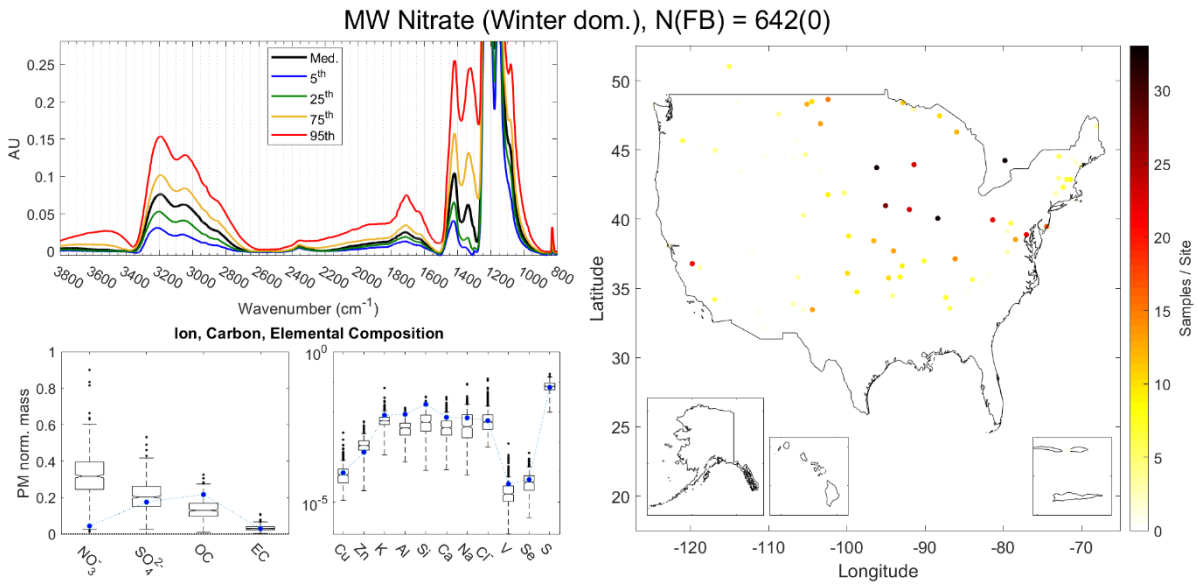



Figure S4-13b – Auxiliary GMM dossier associated with cluster n°13.

Cluster nº14:

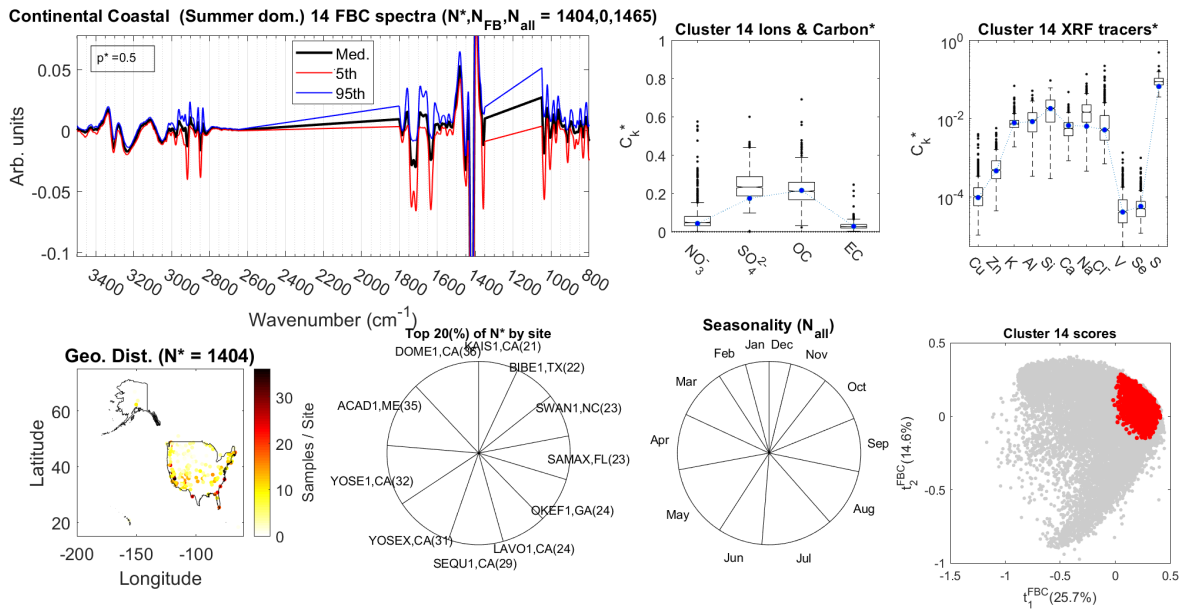



Figure S4-14a – GMM dossier associated with cluster nº14.

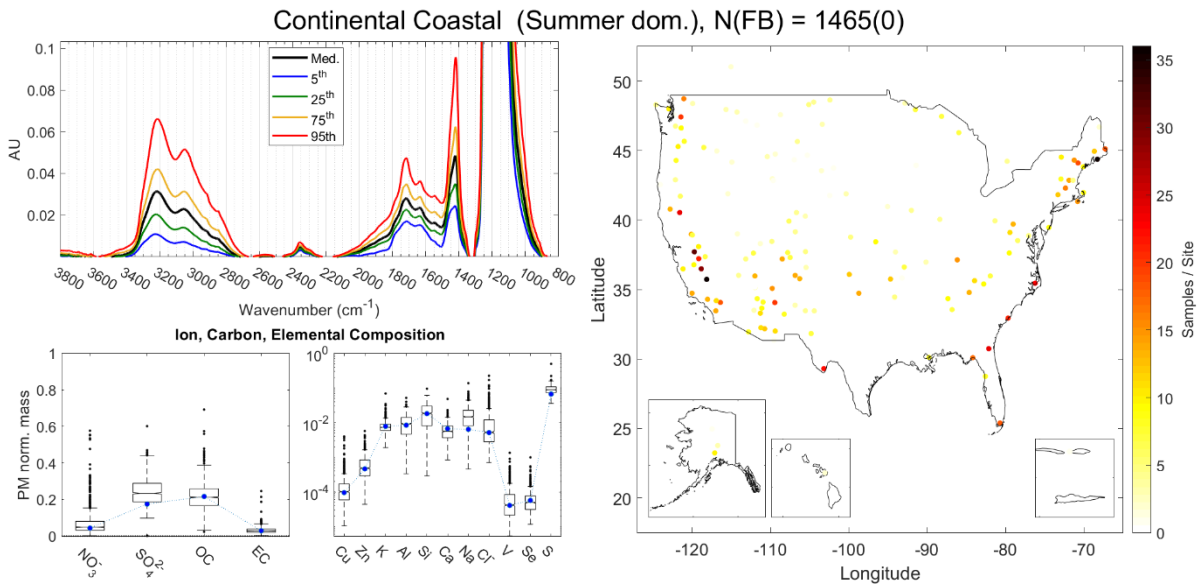



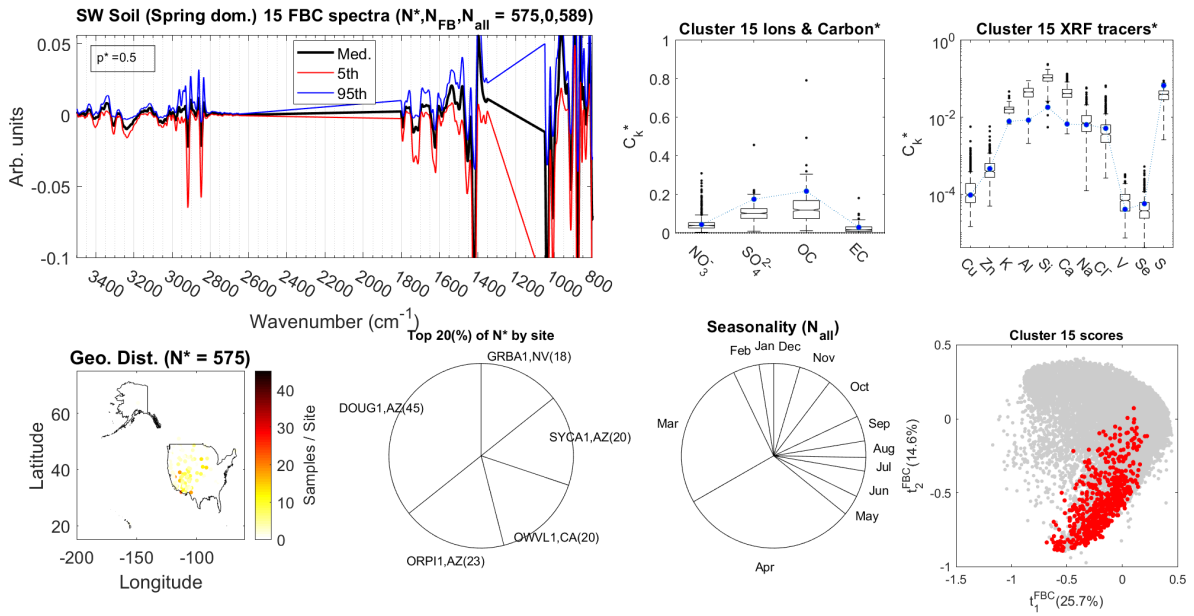
Figure S4-14b – Auxiliary GMM dossier associated with cluster nº14.

Cluster nº15:

Figure S4-15a – GMM dossier associated with cluster nº15.

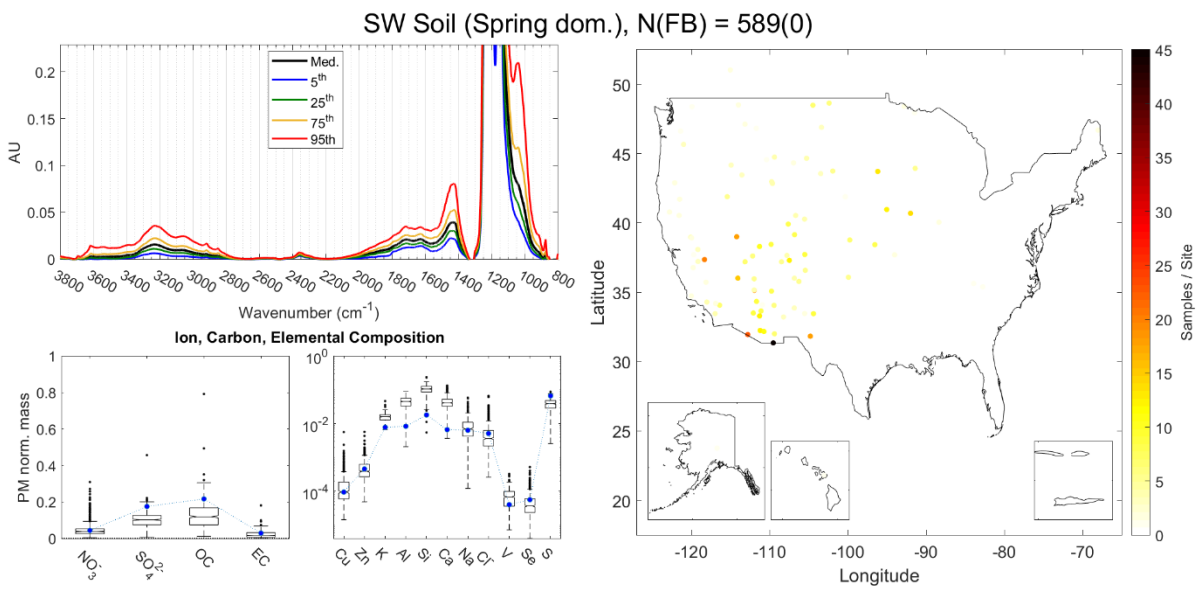



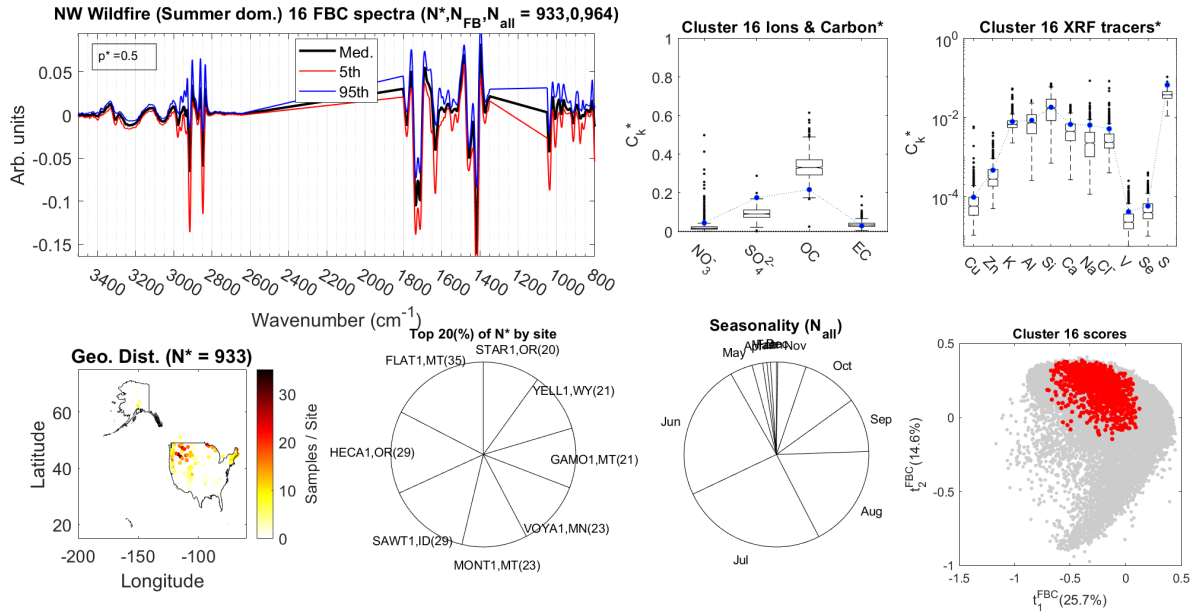
Figure S4-15b – Auxiliary GMM dossier associated with cluster nº15.

Cluster nº16:

Figure S4-16a – GMM dossier associated with cluster nº16.

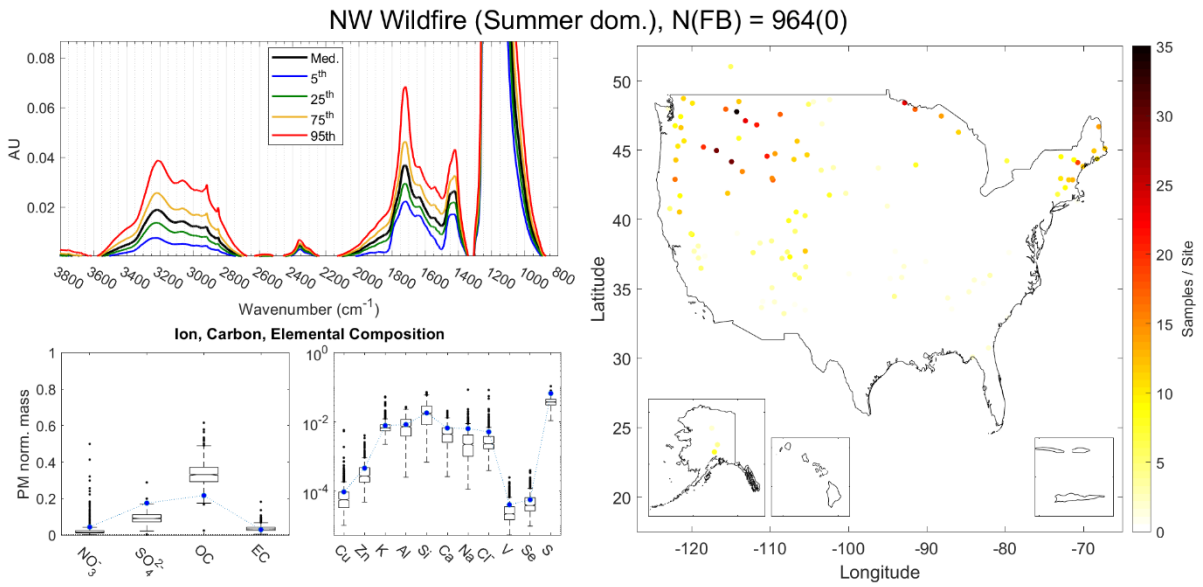



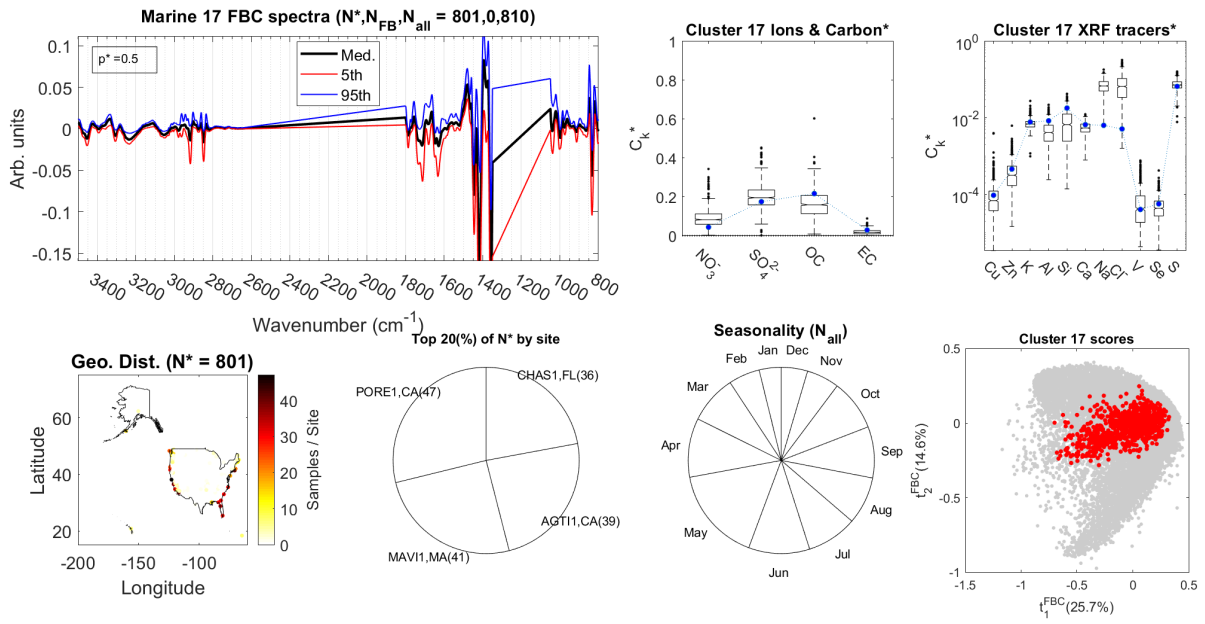
Figure S4-16b – Auxiliary GMM dossier associated with cluster nº16.

Cluster nº17:

Figure S4-17a – GMM dossier associated with cluster nº17.

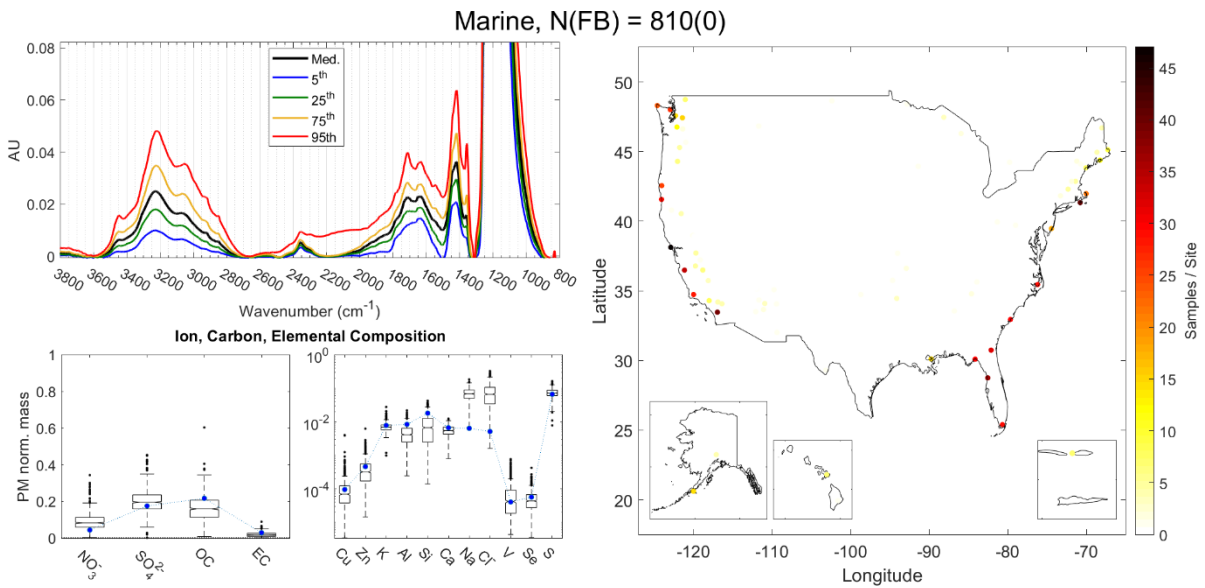



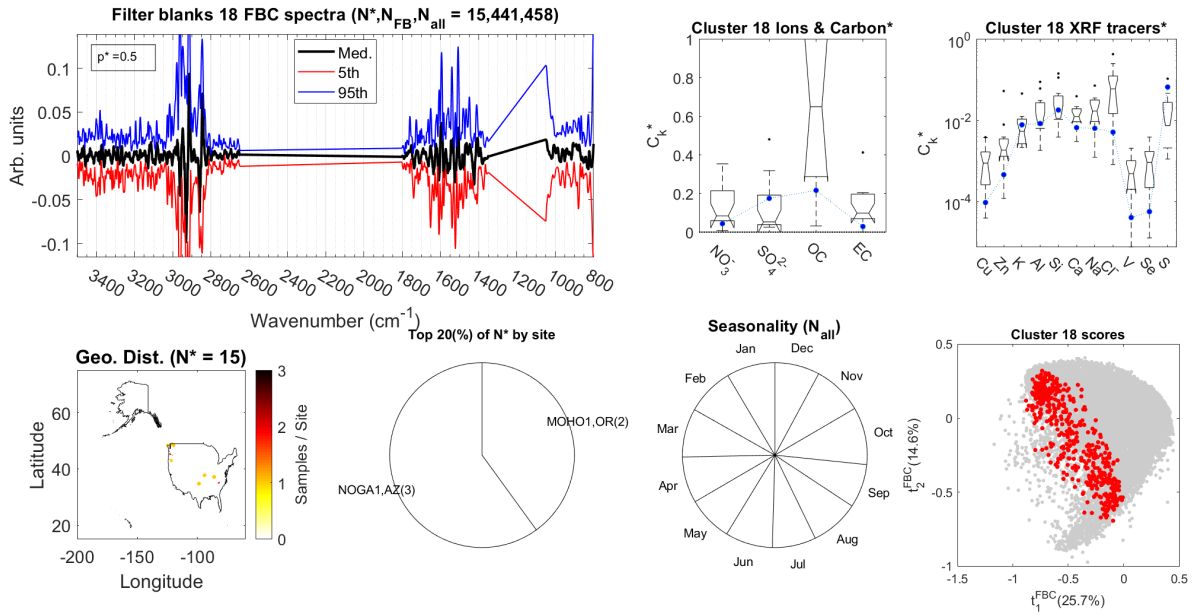
Figure S4-17b – Auxiliary GMM dossier associated with cluster nº17.

Cluster nᵒ18:

Figure S4-18a – GMM dossier associated with cluster nᵒ18.

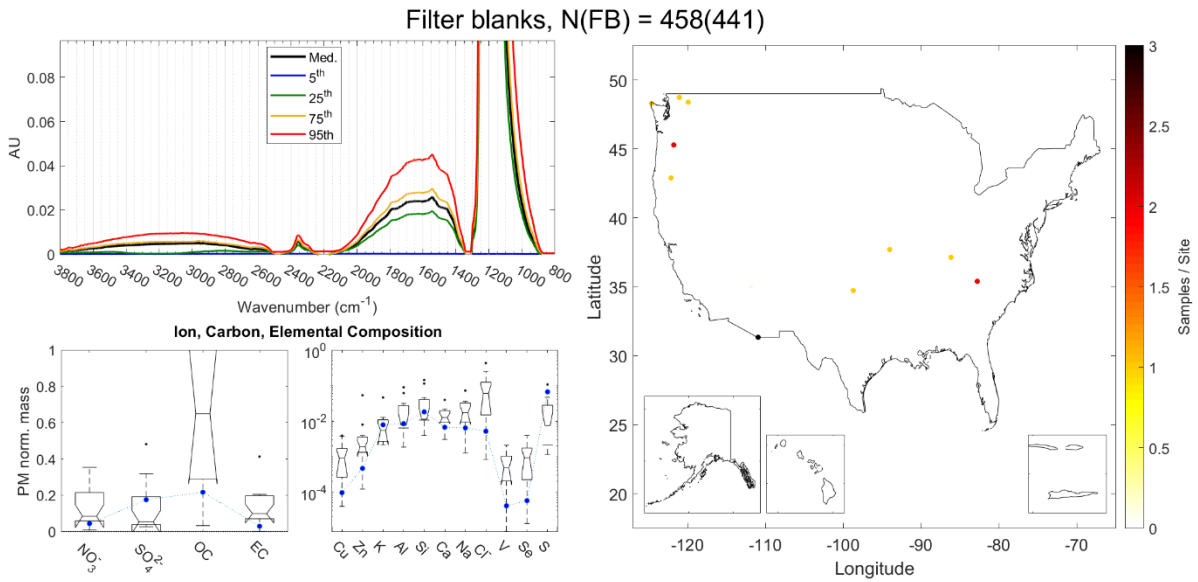



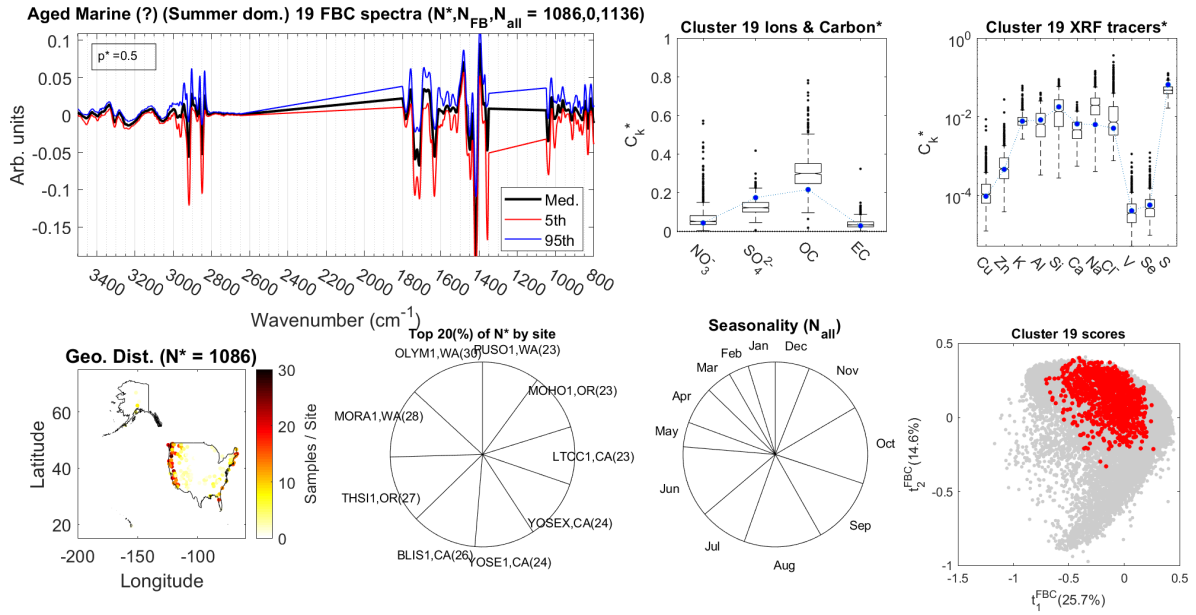
Figure S4-18b – Auxiliary GMM dossier associated with cluster nᵒ18.

Cluster nº19:

Figure S4-19a – GMM dossier associated with cluster nº19.

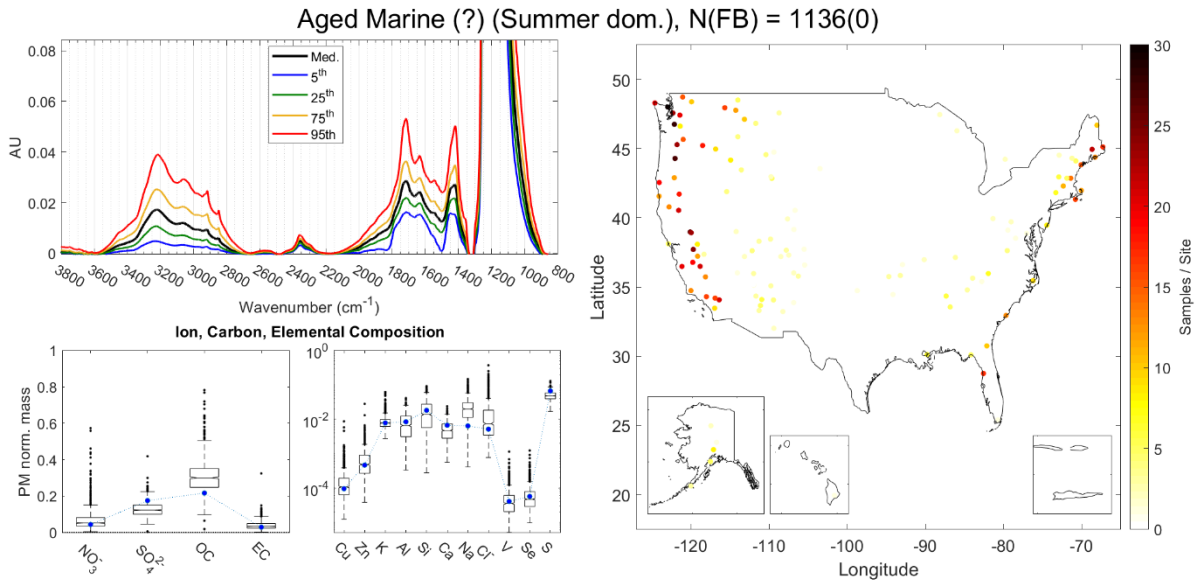



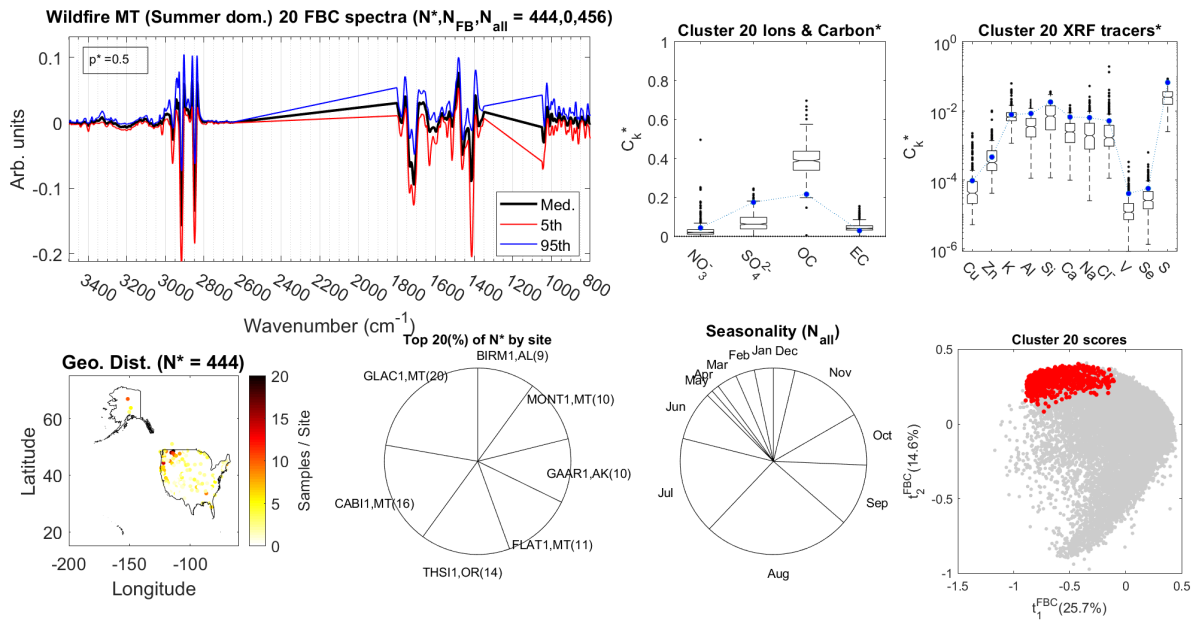
Figure S4-19b – Auxiliary GMM dossier associated with cluster nº19.

Cluster nº20:

Figure S4-20a – GMM dossier associated with cluster nº20.

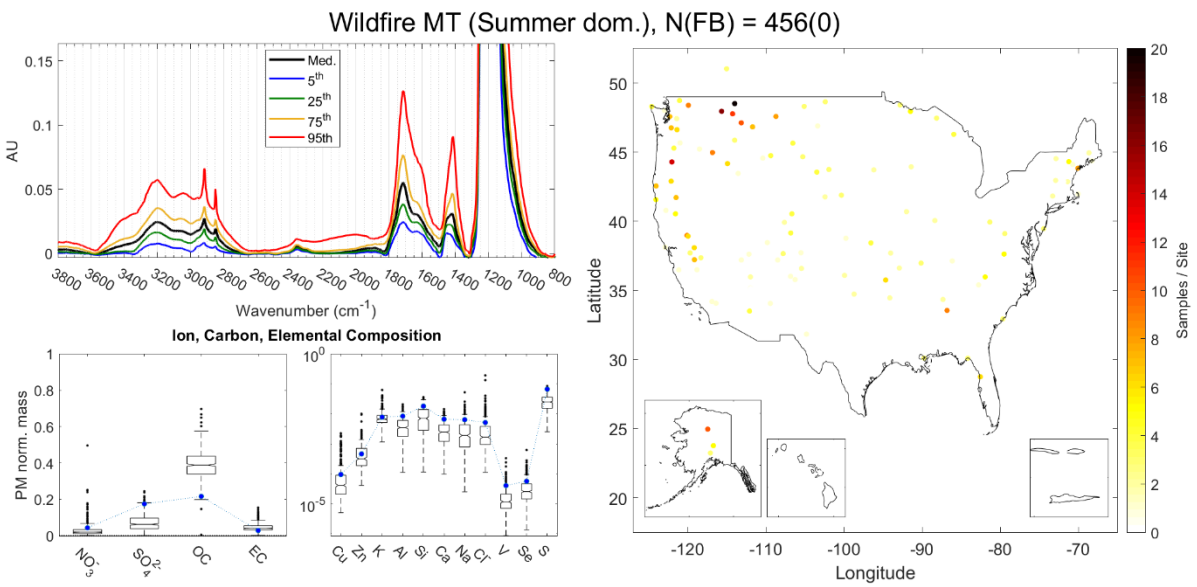



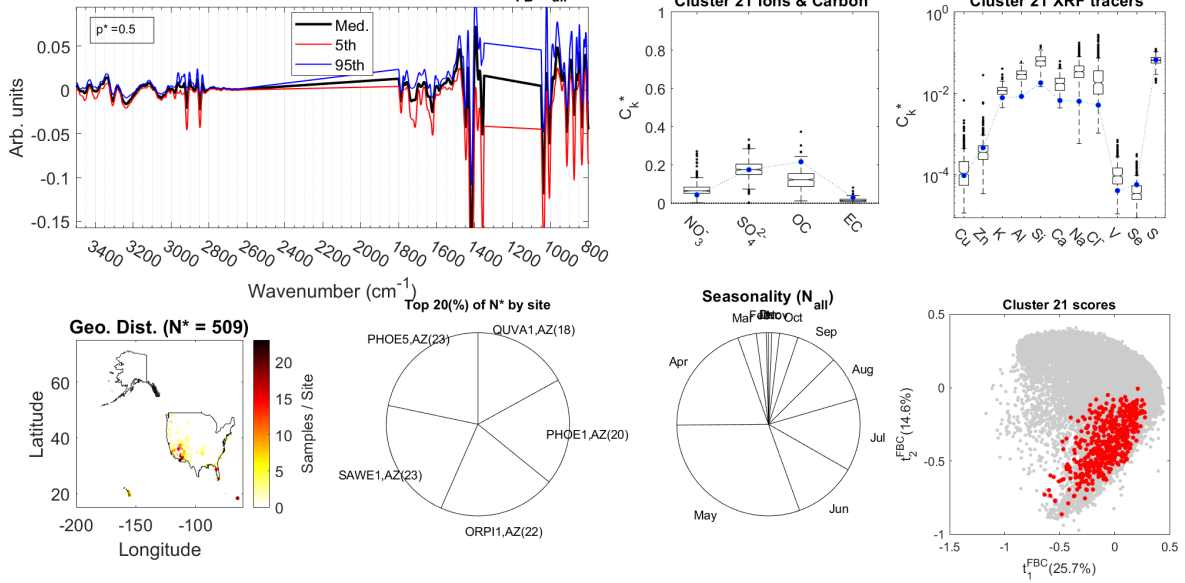
Figure S4-20b – Auxiliary GMM dossier associated with cluster nº20.

Cluster nᵒ21:

Figure S4-21a – GMM dossier associated with cluster nᵒ21.

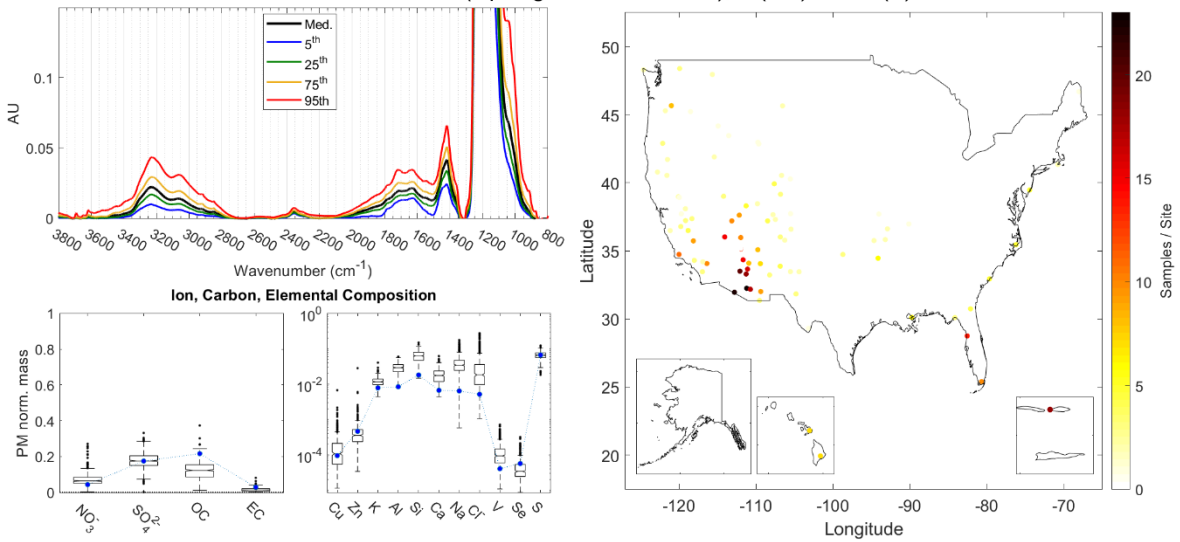



Figure S4-21b – Auxiliary GMM dossier associated with cluster nᵒ21.

Section S5: Biomass Burning - Calibration design

Due to the broad variability in the raw spectral response, the Biomass Burning model is sensitive to baseline fluctuations and offsets. In practice, samples whose spectral class is underrepresented in the calibration are not well accounted for by the model and tend to yield unsatisfactory predicted concentrations. To prevent such an issue and ensure a homogeneous representation of all spectral shapes and patterns in the calibration, the following procedure has been implemented. First, biomass burning spectra are arranged in a matrix *[X]* containing on each row the raw spectrum of a given sample and on each column the corresponding spectral intensity at each wavenumber. Starting from *[X]*, the Hat matrix is computed whose diagonal gives an estimate of sample leverage according to Eq. (1):

$$[H] = [X]([X]^T [X])^{-1}[X]^T \tag{1}$$

In addition to sample leverage, providing an indication of extreme or most dissimilar spectra, the OC/EC ratio is also considered to ensure a representative distribution of carbonaceous aerosol concentrations is taken into account as part of the sample selection procedure. By subjecting both quantities to the Kennard-Stone algorithm (Kennard and Stone, 1969), a uniform subset of biomass burning samples balancing spectral pattern and aerosol composition can be identified. The sample size to be retained for calibration is set to two-third of the total number of biomass burning samples. This procedure is repeated for each sampling year. Figure S5-1 shows the data partitioning achieved for the year 2015.

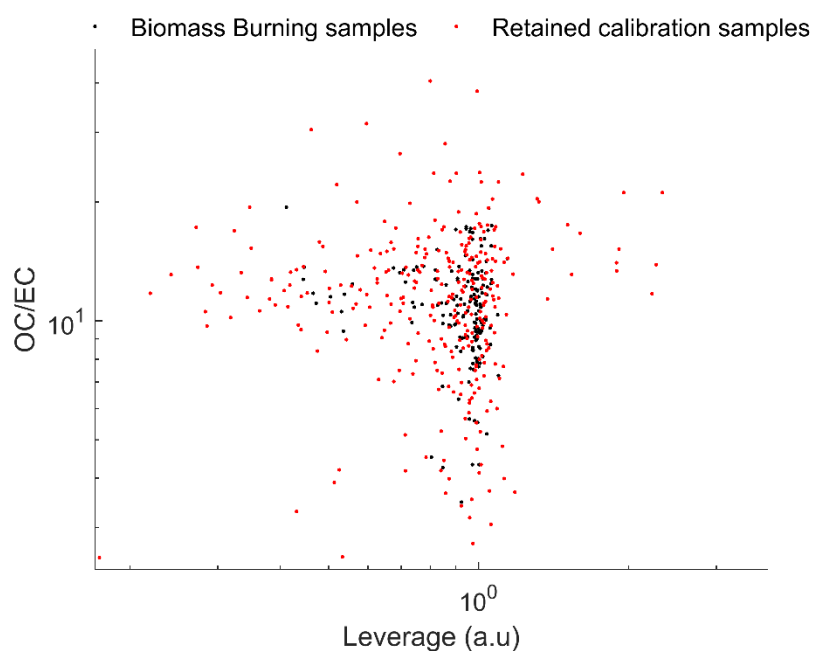



Figure S5-1 – Kennard-Stone partitioning for the selection of biomass burning samples to be retained for calibration from the year 2015. To enhance visualization, results are shown in logarithmic scale. The uniform spread of calibration samples across the Leverage – OC/EC subspace suggests the retained samples are suitable for developing a robust Biomass Burning model.

Section S6: Gaussian Mixture Model & site selection

Using the minimum Bayesian Information Criterion (BIC), an optimum GMM solution with 21 clusters was retained among all attempted models. Out of the 21 clusters, 20 are characteristic of atmospheric aerosols while the remaining cluster (#18) is specific to field blank spectra. With the exception of cluster #18, uninformative in term of $PM_{2.5}$ composition, a single site per cluster was selected to represent the variability in major aerosols constituents experienced throughout the year 2015. The decision regarding which site to retain is inferred from the sample probability to belong to a particular cluster and the number of occurrence of each site within that cluster. In practice, the site with the largest number of classified spectra that maximize the probability density function (PDF) is selected by default. Additional rules, as described in the paper, are also implemented to prevent situations where duplicate sites are selected. To have a better perspective of the spatial distribution of the 20 retained sites, all locations have been marked on the map shown in Figure S6-1.

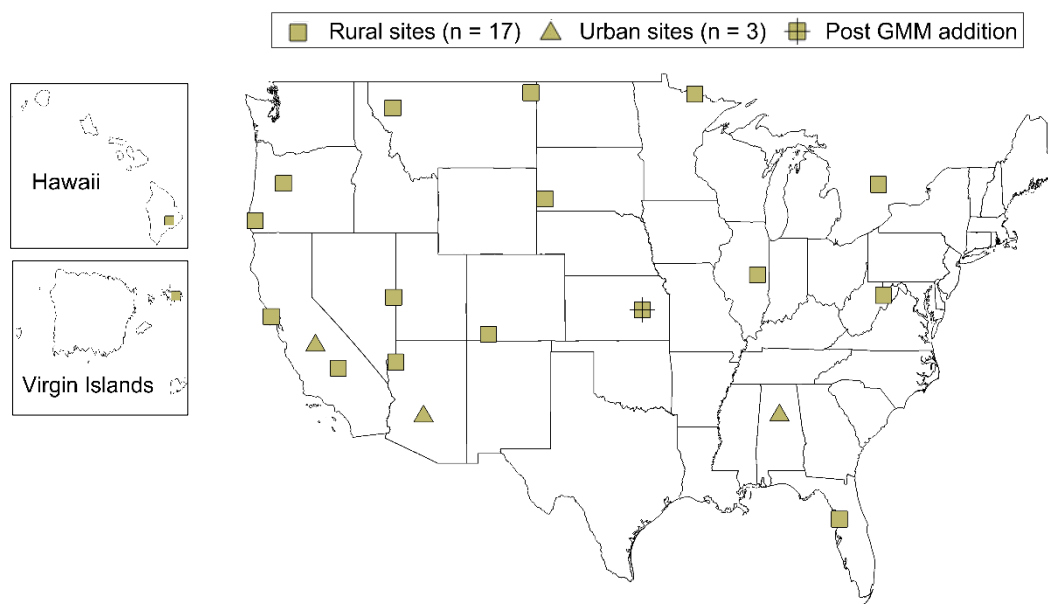



Figure S6-1 – Spatial coverage of the 20 sites representative of $PM_{2.5}$ composition retained following the GMM analysis and the one additional site added to represent cluster #2. Cluster #18 was discarded as it contains only field blanks spectra.

In Figure S6-1, the spatial distribution of the 20 sites presents a reasonable balance between eastern and western locations and covers both continental and overseas regions within the network. Noticeably, no representative site have been identified in the Midwest where characteristic $PM_{2.5}$ sources related with agricultural activities (resuspended road dust, fuel combustion) have been previously reported (Kundu and Stone, 2014). Initially, the absence of Midwest sites is considered surprising as cluster #2 was found to contain a large collection of sites in this region with high sample occurrence and probability (Figure S7-2a). A closer inspection at cluster #2 reveals a broad distribution of sites encompassing multiple states, including Alabama where Birmingham was selected as the most representative site with 27 classified spectra. Due to the large spread of this cluster, decision was made

to extract a second site to enhance the spatial coverage in the Midwest. Based on the GMM dossier (Figure S7-2a), the Tallgrass Prairie National Preserve site (KS) with 22 classified spectra is an adequate choice. With this new addition, 21 sites will now be considered to develop the Limited model, which corresponds to about 13 % of the network capacity. A brief description of the retained sites is available in Table S6.   Figure S6-3 shows the seasonal distribution of the each cluster.   Figure S6-4 shows the distribution of species in the selected sites compared to the rest of the network indicating that the selected sites reasonably approximate the range of concentrations of species in the network.

Table S6 – Summary of the GMM clusters and their 21 most representative sites retained for Limited modeling. The second column indicates the number of classified spectra (p > 0.5) from each selected site. No site has been drawn from the uninformative cluster #18 featuring field blank spectra only.

Cluster #	# of samples in cluster	Site name	Site code	State	Type
1	31	Shamrock Mines	SHMI1	CO	Rural
2	27	Birmingham	BIRM1	AL	Urban
2	22	Tallgrass Prairie	TALL1	KS	Rural
3	52	Point Reyes	PORE1	CA	Rural
4	29	Wind Cave	WICA1	SD	Rural
5	33	Bondville	BOND1	IL	Rural
6	30	Kalmiopsis	KALM1	OR	Rural
7	26	Meadview	MEAD1	AZ	Rural
8	28	Medicine Lake	MELA1	MT	Rural
9	58	Virgin Islands	VIIS1	VI	Rural
10	41	Fresno	FRES1	CA	Urban
11	44	Dolly Sods	DOSO1	WV	Rural
12	67	Hawaii Volcanoes	HAVO1	HI	Rural
13	33	Egbert (Canada)	EGBE1	ON	Rural
14	36	Dome Land	DOME1	CA	Rural
15	18	Great Basin	GRBA1	NV	Rural
16	23	Voyageurs	VOYA2	MN	Rural
17	36	Chassahowitzka	CHAS1	FL	Rural
19	27	Three Sisters	THSI1	OR	Rural
20	11	Flathead	FLAT1	MT	Rural
21	20	Phoenix	PHOE1	AZ	Urban

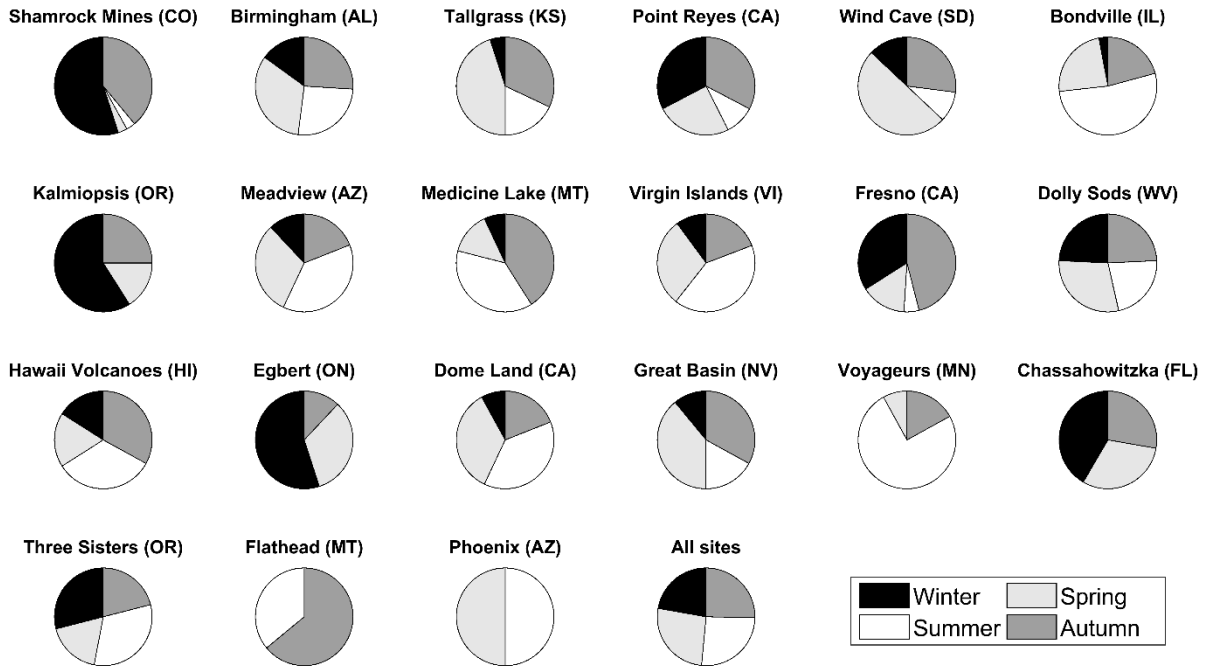



Figure S6-2 – Seasonal distribution of the classified spectra (p > 0.5) from each representative site and for each GMM cluster.

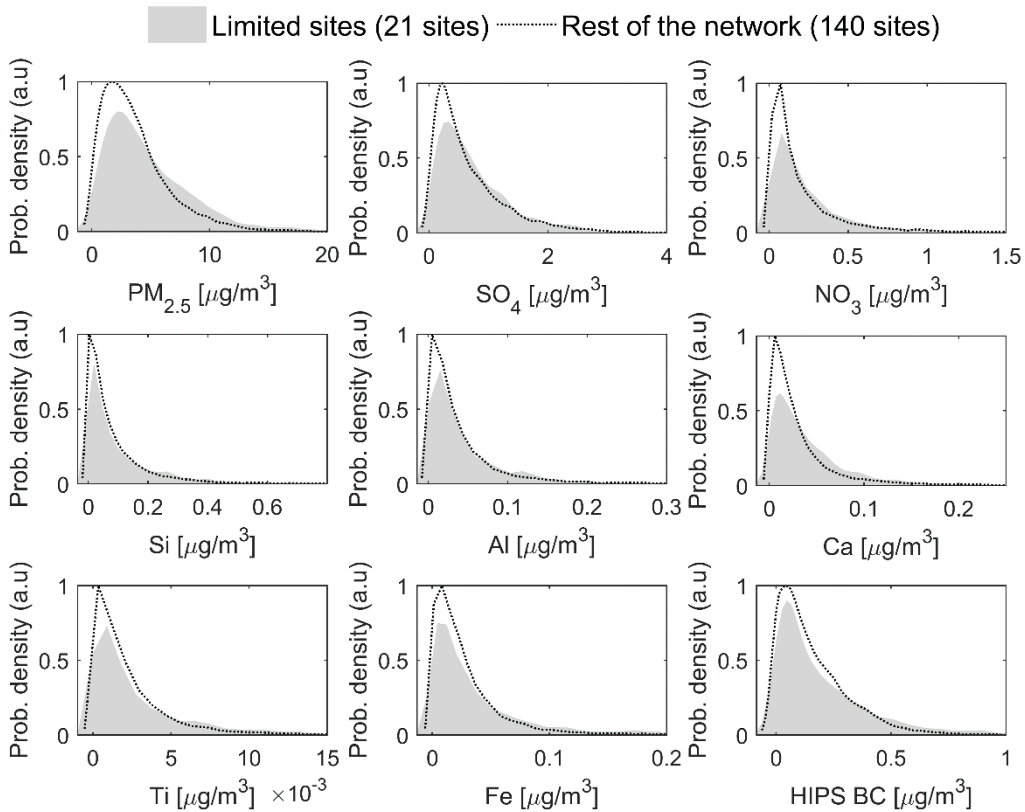



Figure S6-3 Distributions of mass, ions, elements and HIPS BC for the sites used in the limited model compared to the rest of the sites.

Section S7: Biomass Burning model performance

Table S7-1 – *Inter-model performance comparison between Global and Biomass Burning predicted concentrations for the 492 samples classified as biomass burning in 2015. Values confirm Biomass Burning modeling is not necessary required for OC and TC but brings substantial improvement for EC. In the absence of field blank samples, MDL is not reported.*

	Model	R^2	Bias (μg/m3)	Bias (%)	Error (μg/m3)	Error (%)	MDL (μg/m3)	< MDL (%)
OC	Global	0.973	0.05	1.2	0.32	6.1	-	-
	Biomass Burning	0.984	0.02	0.7	0.33	6.2	-	-
EC	Global	0.747	0.01	2.6	0.07	15	-	-
	Biomass Burning	0.902	0.01	2.2	0.07	13	-	-
TC	Global	0.978	0.10	2.1	0.30	5.6	-	-
	Biomass Burning	0.984	0.05	0.9	0.35	6.2	-	-

Section S8: Bias in least-squares regression.

The factors that influence FT-IR method accuracy are readily assessed by plotting residual-error against any independent variable of interest (e.g., sampling date). Here we define residual-error as the difference between the FT-IR prediction ($\hat{y}_i$) and reference measure (y_i) or

$$\hat{e}_i = \hat{y}_i - y_i. \qquad\qquad [1]$$

Takahama et al., 2019 illustrates such a procedure for several sites in the IMPROVE network where OC residual-errors are plotted against sampling date (*t*) following a transformation of the errors to standard normal covariates. These figures allow one to ascertain short- and long-run fluctuations in OC predictions, with the former corresponding to random error and later (perhaps) to time-dependent biases. Method error may also be assessed by plotting residual errors against prediction ($\hat{y}_i$) to gain insight into method bias as a function of sample mass loading. Residual-error and reference measurements (y_i) are not independent for least-squares estimators (Besalú et al., 2006; Draper and Smith, 1998, pp. 63-64,173,638). This well-understood, but often overlooked, aspect of least-squares estimators has the consequence of introducing an *apparent* linear bias into residual plots dependent on the strength of correlation (R^2) between the predictor variables (i.e., infrared absorption) and dependent variable (e.g., TOR OC). In fact, for our definition of residual error in [1] the apparent trend will have a negative slope with magnitude of $(1 - R^2)$ and intercept $(1 - R^2)\bar{y}$. Therefore, our biases are evaluated against the predicted concentrations below (Figure S8-1 and S8-2) to assess biases that are not introduced by the least squares estimation method.

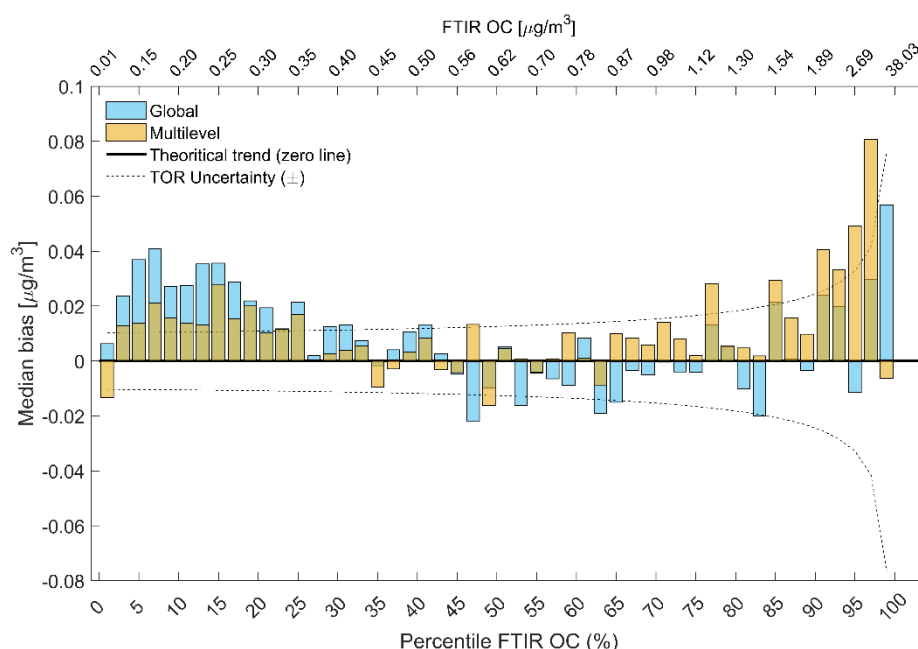



Figure S8-1 – Median bias of Global and Multilevel models for FT-IR OC concentrations bins. Each bin contains 2.5% of the data. Values on the top x-axis correspond to the bounds of the concentration per bin. While the Multilevel model leads to a significant drop in bias at concentrations below 0.35 µg/m³, a moderate increase for samples predicted above 1.54 µg/m³ is also observed. The magnitude of the

reported Multilevel bias is comparable with TOR uncertainty boundaries at low levels and above at high concentrations except for the highest bin. TC performs similar to OC.

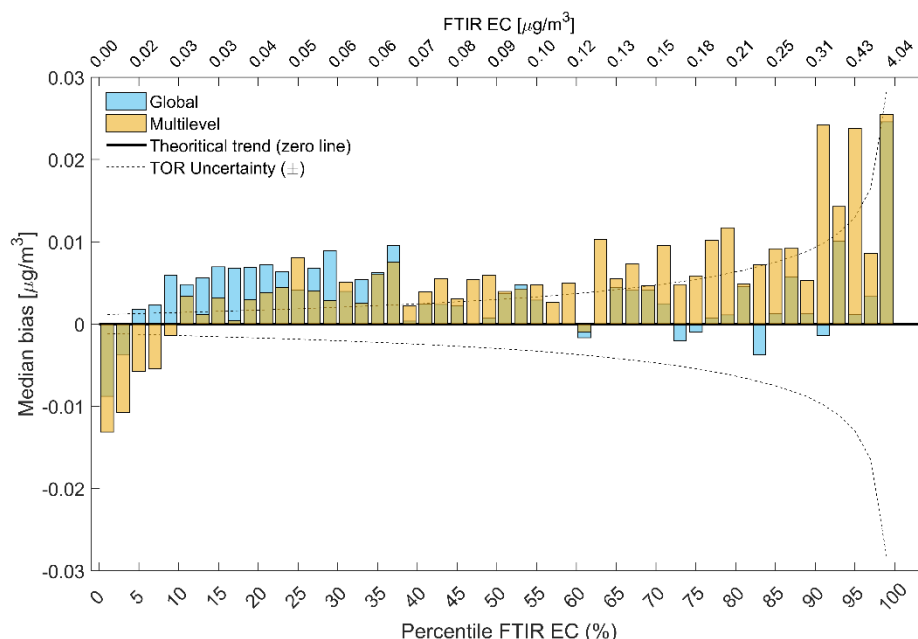



Figure S8-2 – Median bias of Global and Multilevel models for FT-IR EC concentrations bins. Each bin contains 2.5% of the data. Values on the top x-axis correspond to the bounds of the concentration per bin. While the Multilevel model leads to a significant drop in bias at concentrations for the lower 40th percentile, a moderate increase for samples predicted above 90th percentile. In either cases, the magnitude of the reported Multilevel bias is comparable with TOR uncertainty boundaries.

Section S9: Multilevel models for non-carbonaceous IR active atmospheric species in 2015 samples

To assess the absorption bands most strongly associated with prediction of each species, the Variable Importance in Project (VIP) scores from the PLS models are examined. This metric characterizes the direction vector of the PLS model weighted by the explained sum-of-squares, with higher values indicating variables (wavenumbers) important for explaining the variation in the target variable (analyte) (Wold et al., 1983). In principle, the average value of this metric should be one, so values greater than one indicate noteworthy variables. However, the actual threshold varies by the number of variables and signal strength (Chong and Jun, 2005). The interpretation must further consider that the PLS model does not necessarily need to use all absorbance bands from a substance for quantification, and absorption bands of interferents used by the model can also have high VIP scores. In our application, the PTFE absorption bands often exhibit high VIP values presumably because these bands are required by the model for correction of the absorption and scattering by the filter media (Takahama et al., 2016). Examination of VIP scores has the potential to elucidate the major forms (including specific polymorphs) of mineral oxides associated with each element by identifying the most important absorption bands used for prediction; alternatively, insight into elements that are predicted through its correlation with (combination of) other IR-active substances can presumably be garnered with this approach. While an exhaustive treatment is beyond the scope of this paper and should be the topic of another dedicated study, we include a cursory examination below.

According to the VIP score analysis, the SO_4 model appears to place importance S-O stretch between 1130-1080 cm^{-1} and SO_4 deformation between 670-580 cm^{-1} (Mayo, 2004), whereas the NO_3 model appears to place importance on the N-O stretch at 1410-1340 cm^{-1} (Mayo, 2004) (Figure S9-1a). N-H stretch of ammonia near 3200 cm^{-1} also appears to be relevant for the SO_4 model. IR spectra of oxides of Si and Al have distinct, identifiable features – for instance, silica exhibits peak near 1100 cm^{-1} and 500 cm^{-1} due to Si-O stretching and bending, respectively; aluminum oxide exhibits a broad peak near 600 comprised of various Al-O vibrational modes (e.g., (Barker, 1963; Farmer, 1968; Kieffer, 1979; Saniger, 1995) (Mayo, 2004). However, the Si and Al PLS models appear to be using the same features as the VIP scores are nearly identical. The region near 600 cm^{-1} exhibits a relatively high VIP score in both models; the importance of the Si-O stretch is unclear as the PTFE absorbance is strong there also, and this band is used by models for other metals. We note however that the predictions are not identical as the regression coefficients differ (in part due to the difference in number of latent variables used by each PLS model). Spectra of TiO_2 (e.g., rutile and anatase) have absorption between 900-700 and below 600 cm^{-1} (with the band near 500 cm^{-1} providing strongest discrimination among polymorphs) (Busani and Devine, 2005)(Mayo, 2004), but the VIP score profile of the Ti model is similar to that of Si and Al. However, its features are weighted differently (based on the magnitude of the VIP scores) than Si and Al models and some differences can be observed near 500 cm^{-1} (Figure S9-1b). VIP scores of Si and Fe also differ near 500 cm^{-1} but also near 829 cm^{-1}; yellow ferric oxide ($Fe_2O_3 \cdot H_2O$) and goethite (FeOOH) (Figure S9-1c) is reported to have absorbances in this region due to Fe-O bonds (e.g., (Weckler and Lutz, 1998; Namduri and Nasrazadani, 2008; Margenot et al., 2017) (Mayo, 2004). The VIP score profile for the Ca model is strikingly different and suggests that C-O (from carbonate) and Ca-O absorption bands near 1450 and 875 cm^{-1} are used for prediction (Figure S9-1d) (e.g., (Tsyganenko and Filimonov, 1973)Ramdas 1953, Gunasekaran et al. 2006, Galvan-Ruiz et al., 2009). The VIP scores for the Ti, Fe, and Ca models all exhibit a sharp peak near 3615 cm^{-1}, which is likely related to surface hydroxyl groups of oxide species (e.g., (Tsyganenko and Filimonov, 1973).

Figure S9-1. Variable Importance in Projection (VIP) scores for PLS models from 2015. VIP scores for Si model is shown in all plots to establish basis for comparison.

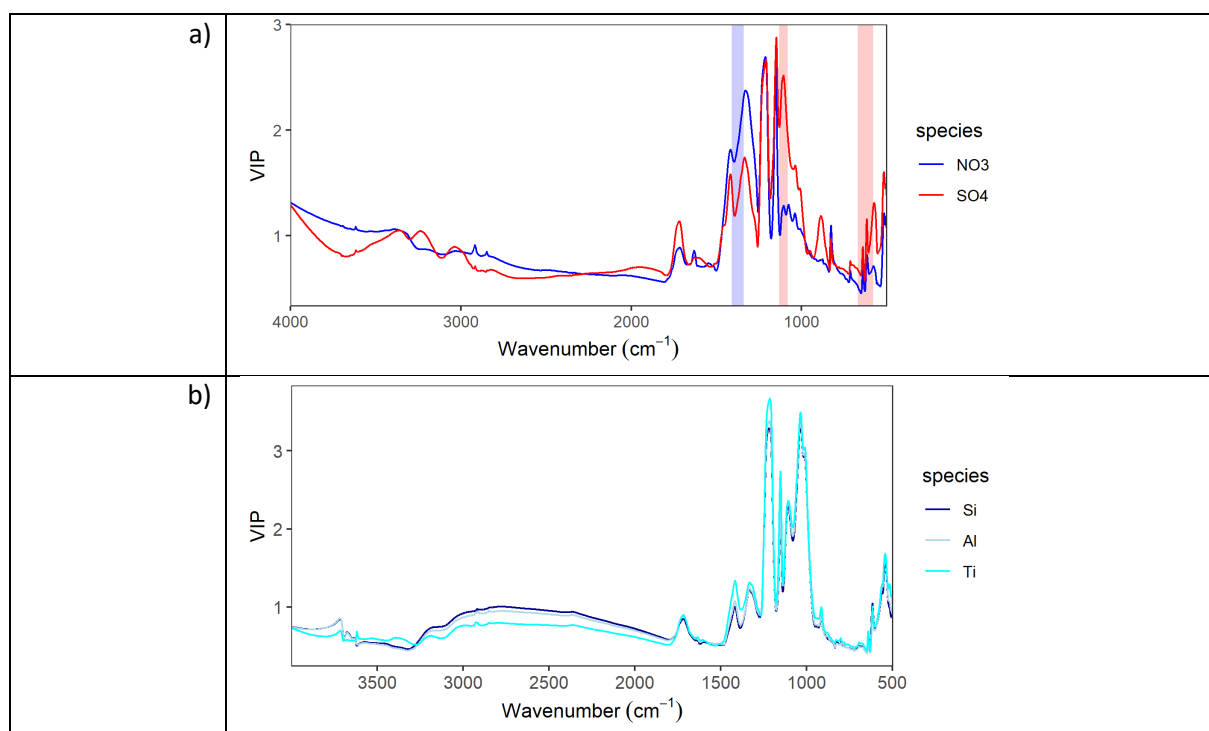
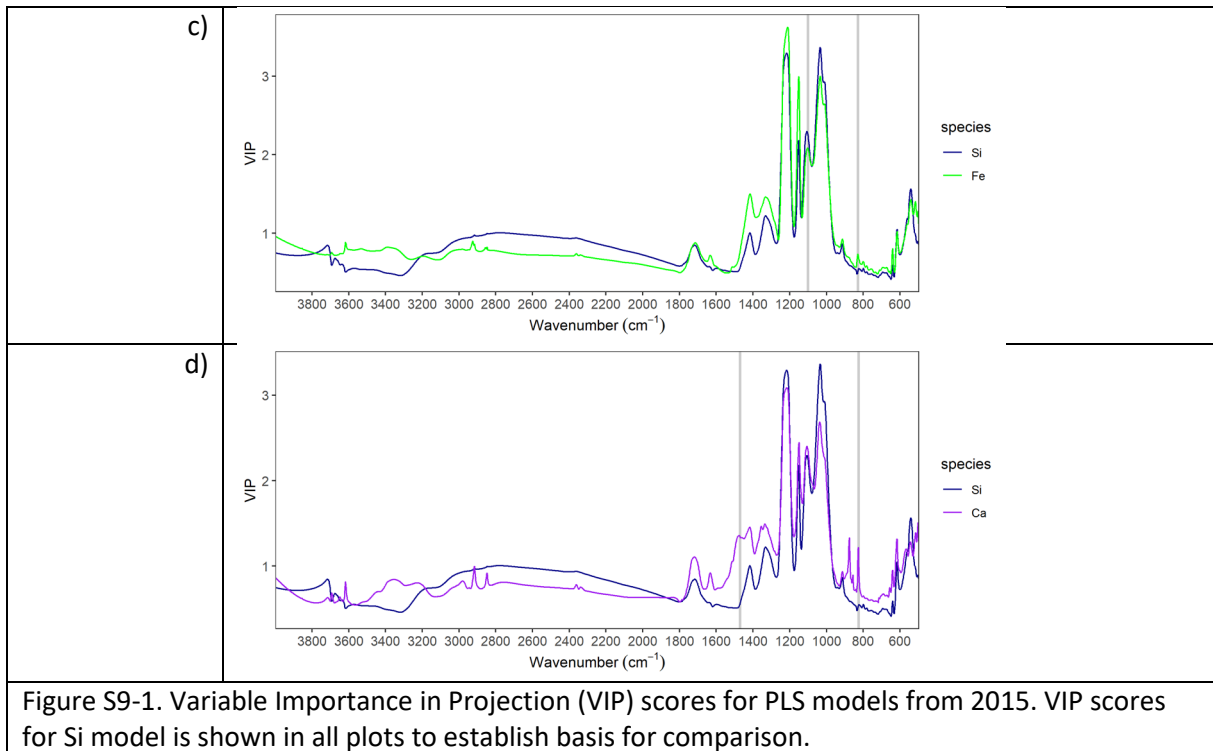



Figure S9-2 shows the cross plot for all predicted species in 2015 and Figure S9-3 shows the residuals for FTIR compared to collocated data for most predicted species in 2015. Unlike any other crustal elements, titanium concentrations presents a curious pattern. In addition to the expected sample distribution (grey markers), where most samples scatter around the 1:1 line, some data points are characterized by a large negative bias (purple markers) but nearly fall on a straight line. Since those samples all originate from the Sycamore Canyon (AZ) site, unique sources of titanium are likely the reason behind the observed discrepancy. Because the same pattern has been identified in all subsequent years (Section S10), the hypothesis of a specific soil and dust composition in the area is strongly favored. Likely, the specificities of local crustal composition are not well accounted for in the Multilevel model and as such a specific titanium model might be required at this location.

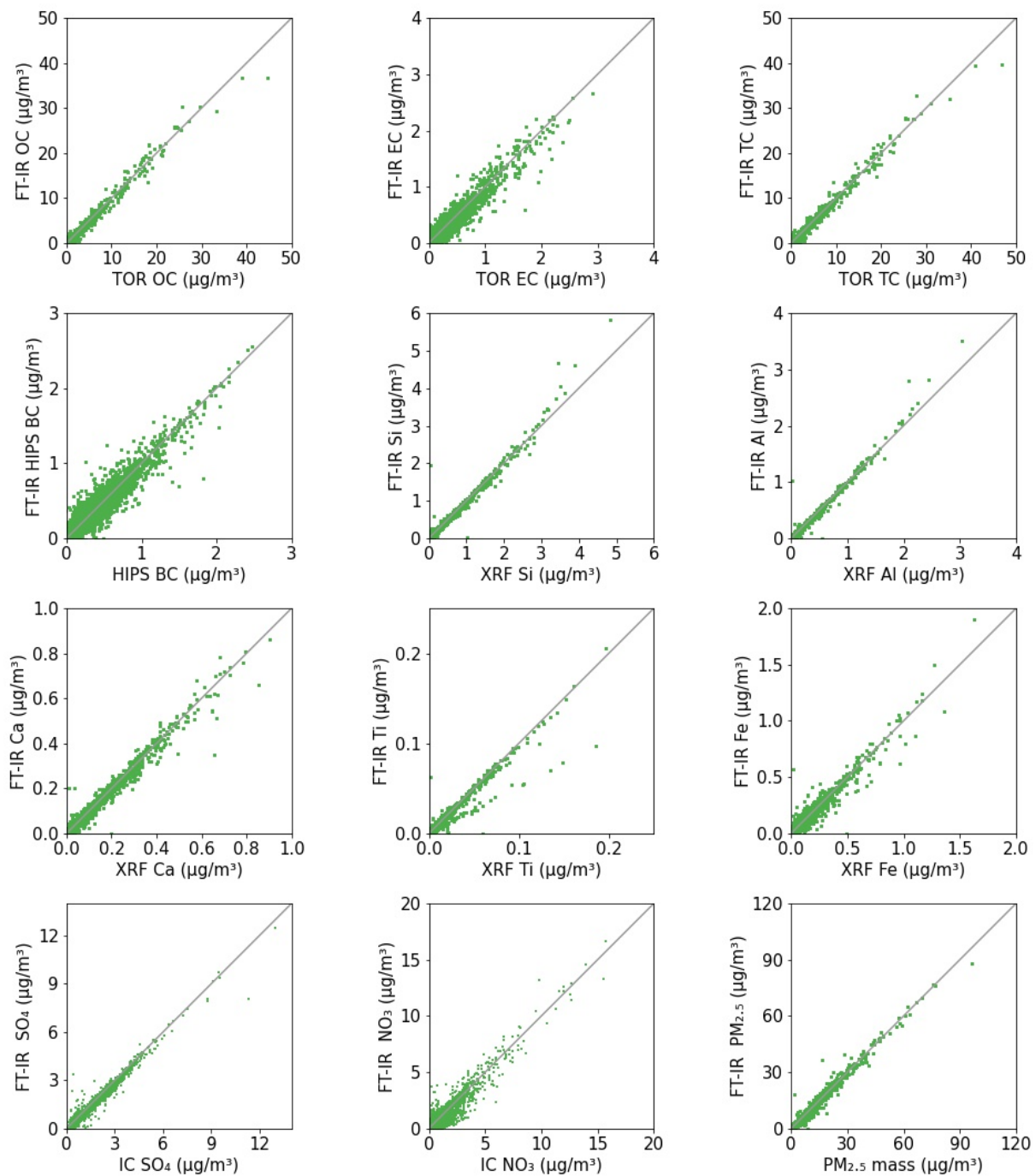



Figure S9-2 – Multilevel FT-IR concentrations for all predicted atmospheric species regressed against their reference measurement. Each subplot contains 19,608 data points from the year 2015.

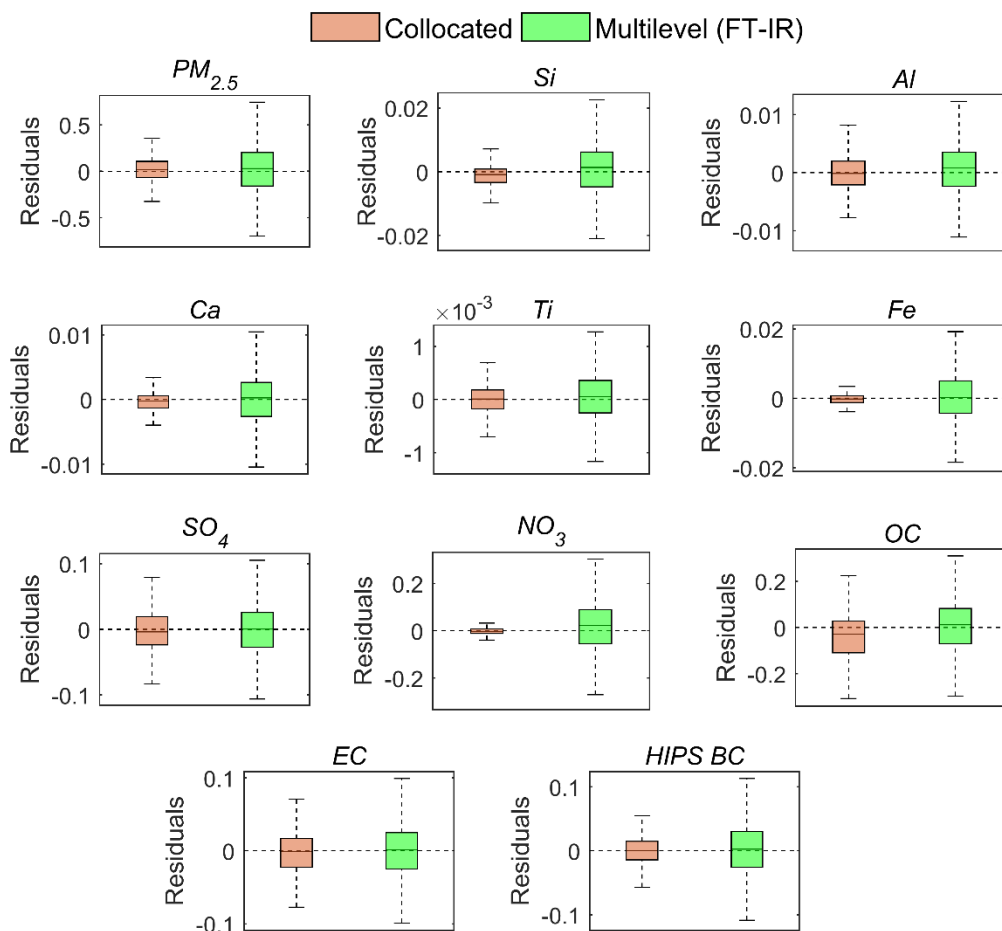



Figure S9-3 – Comparison between collocated samples (orange) and Multilevel (green) residuals for various atmospheric species evaluated from samples collected in 2015. Residuals are reported in µg/m³. For the sake of clarity, samples with residuals exceeding 1.5 times the interquartile range away from the top or bottom whiskers are not shown.

Section S10: Multilevel models calibrations – Long-term stability assessment (2016 – 2017)

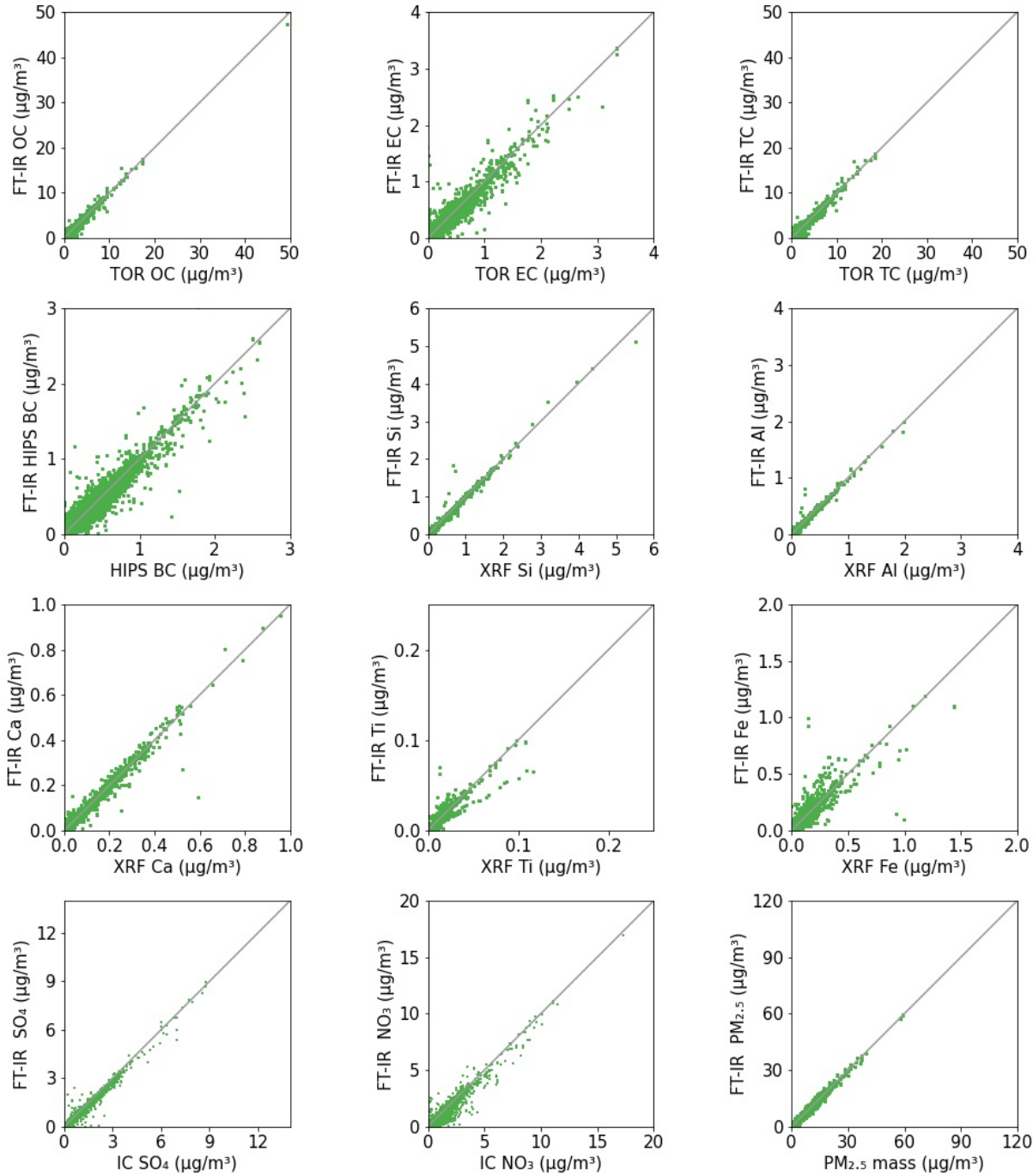



Figure S10-1 – Multilevel FT-IR concentrations for all predicted atmospheric species regressed against their reference measurement. Each subplot contains 19,849 data points from the year 2016.

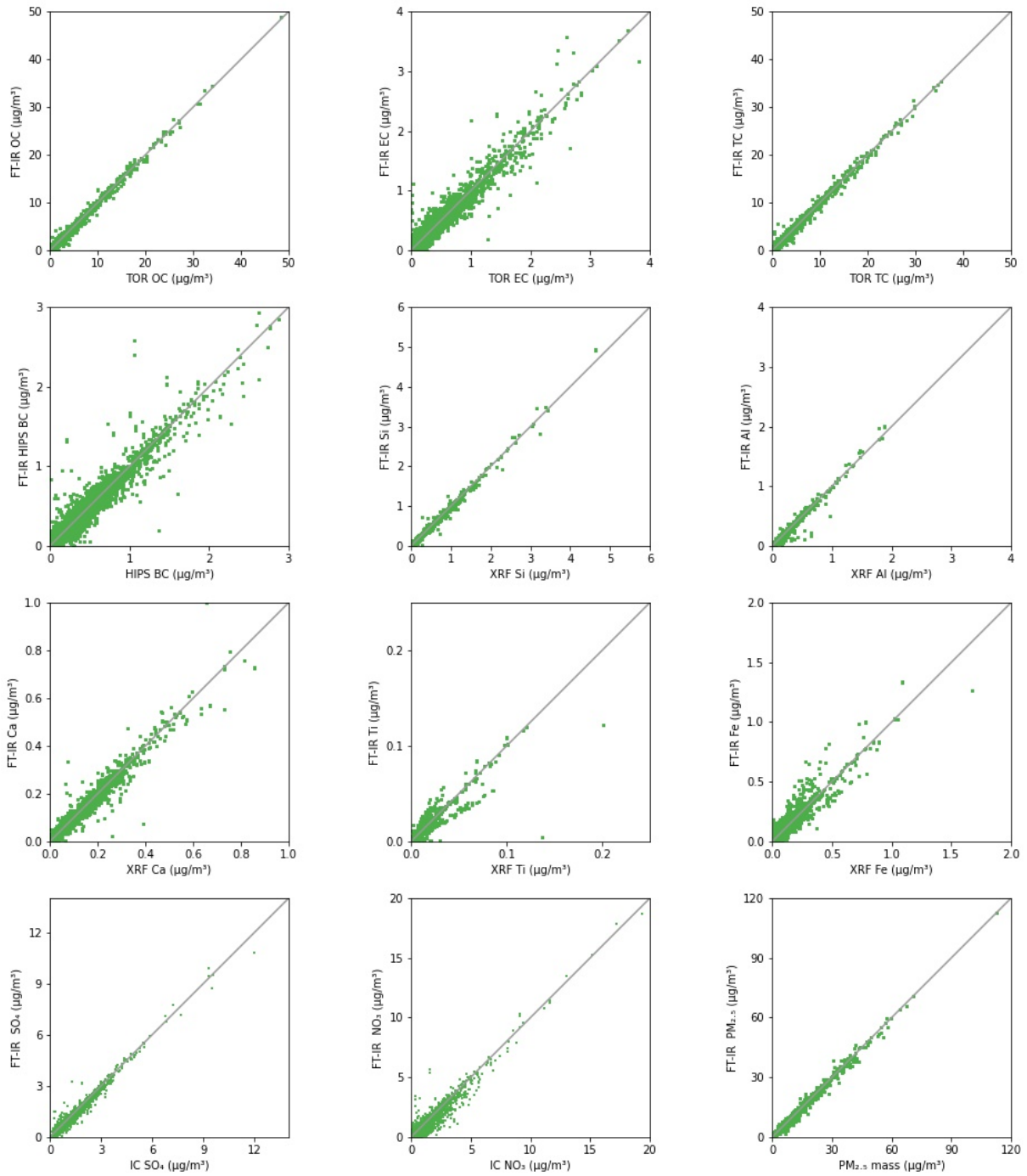



Figure S10-2 – Multilevel FT-IR concentrations for all predicted atmospheric species regressed against their reference measurement. Each subplot contains 22,005 data points from the year 2017.

Table S10-1– *Combined set of Multilevel model performance covering the 2016 sampling period for several IR-active atmospheric constituents. All 19,849 spectra analyzed by FT-IR during that interval, including duplicate and replicate measurements, were considered for model evaluation.*

Species	R^2	Bias [μg/m³]	Bias (%)	Error [μg/m³]	Error (%)	MDL [μg/m³]	< MDL (%)
OC	0.974	0.01	0.7	0.08	14.9	0.05	1
EC	0.905	0	5.2	0.02	31.4	0.04	22.7
TC	0.978	0.01	0.7	0.09	14.3	0.06	1.2
BC	0.906	0	-4.7	0.03	27.1	0.04	23
Si	0.986	0	0.1	0	9.3	0.01	18.5
Al	0.983	0	0.6	0	11.8	0	13.3
Ca	0.975	0	0.5	0	11.5	0	12.2
Ti	0.901	0	0.7	0	21.5	0	19.7
Fe	0.814	0	0.5	0	27	0.01	21.8
SO$_4$	0.984	0	0.8	0.02	5.5	0.03	0.4
NO$_3$	0.9	0.02	12.8	0.07	53.8	0.07	25.8
PM$_{2.5}$ Mass	0.986	0.03	1.2	0.16	6.4	0.23	0.9

Table S10-2– *Combined set of Multilevel model performance covering the 2017 sampling period for several IR-active atmospheric constituents. All 22,005 spectra analyzed by FT-IR during that interval, including duplicate and replicate measurements, were considered for model evaluation.*

Species	R^2	Bias [μg/m³]	Bias (%)	Error [μg/m³]	Error (%)	MDL [μg/m³]	< MDL (%)
OC	0.987	0.01	1.5	0.08	11.9	0.07	1.3
EC	0.913	0	1.1	0.02	25.5	0.02	10.8
TC	0.988	0.02	1.7	0.09	11.5	0.08	1.2
BC	0.922	0	-0.2	0.02	19.5	0.03	11.5
Si	0.993	0	0.9	0	9.7	0.01	11.5
Al	0.982	0	0.7	0	12.6	0	11.2
Ca	0.974	0	2.6	0	14.3	0	8.7
Ti	0.879	0	0.4	0	22.5	0	19.1
Fe	0.905	0	2.6	0	26.8	0	15.4
SO$_4$	0.979	0	0.3	0.03	6.4	0.03	1
NO$_3$	0.906	0.02	14.4	0.07	49.1	0.06	17.5
PM$_{2.5}$ Mass	0.989	0	0	0.17	6.1	0.18	0.6

References:

Abdi, H. and Williams, L. J.: Principal component analysis, WIREs Computational Statistics, 2, 433–459, https://doi.org/10.1002/wics.101, 2010.

Allen, D. T., Palen, E. J., Haimov, M. I., Hering, S. V., and Young, J. R.: Fourier Transform Infrared Spectroscopy of Aerosol Collected in a Low Pressure Impactor (LPI/FTIR): Method Development and Field Calibration, Aerosol Science and Technology, 21, 325–342, https://doi.org/10.1080/02786829408959719, 1994.

Barker, A. S.: Infrared Lattice Vibrations and Dielectric Dispersion in Corundum, Phys. Rev., 132, 1474–1481, https://doi.org/10.1103/PhysRev.132.1474, 1963.

Besalú, E., Julián-Ortiz, J. V. de., Iglesias, M., and Pogliani, L.: An overlooked property of plot methods, Journal of Mathematical Chemistry, 39, 475–484, https://doi.org/10.1007/s10910-005-9035-z, 2006.

Bilmes, J. A.: A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, 4:126, 1998.

Bro, R. and Smilde, A. K.: Centering and scaling in component analysis, Journal of Chemometrics, 17, 16–33, https://doi.org/10.1002/cem.773, 2003.

Busani, T. and Devine, R. A. B.: Dielectric and infrared properties of TiO 2 films containing anatase and rutile, Semicond. Sci. Technol., 20, 870–875, https://doi.org/10.1088/0268-1242/20/8/043, 2005.

Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, 39, 1–38, 1977.

Draper, N. R. and Smith, H.: Applied Regression Analysis, 1998.

Eilers, P. H. C.: A Perfect Smoother, Anal. Chem., 75, 3631–3636, https://doi.org/10.1021/ac034173t, 2003.

Farmer, V. C.: Infrared spectroscopy in clay mineral studies, Clay miner., 7, 373–387, https://doi.org/10.1180/claymin.1968.007.4.01, 1968.

Fraley, C. and Raftery, A. E.: Model-Based Clustering, Discriminant Analysis, and Density Estimation, Journal of the American Statistical Association, 97, 611–631, https://doi.org/10.1198/016214502760047131, 2002.

Hastie, T, Tibshirani, R, and Friedman, J: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer New York, New York, NY, 2009.

Kennard, R. W. and Stone, L. A.: Computer aided design of experiments, 11, 137–148, https://doi.org/10.1080/00401706.1969.10490666, 1969.

Kieffer, S. W.: Thermodynamics and lattice vibrations of minerals: 2. Vibrational characteristics of silicates, Rev. Geophys., 17, 20, https://doi.org/10.1029/RG017i001p00020, 1979.

Kundu, S. and Stone, E. A.: Composition and sources of fine particulate matter across urban and rural sites in the Midwestern United States, 16, 1360–1370, https://doi.org/10.1039/C3EM00719G, 2014.

Margenot, A. J., Calderón, F. J., Goyne, K. W., Mukome, F. N. D., and Parikh, S. J.: IR Spectroscopy, Soil Analysis Applications, in: Encyclopedia of Spectroscopy and Spectrometry (Third Edition), edited by: Lindon, J. C., Tranter, G. E., and Koppenaal, D. W., Academic Press, Oxford, 448–454, https://doi.org/10.1016/B978-0-12-409547-2.12170-5, 2017.

Namduri, H. and Nasrazadani, S.: Quantitative analysis of iron oxides using Fourier transform infrared spectrophotometry, Corrosion Science, 50, 2493–2497, https://doi.org/10.1016/j.corsci.2008.06.034, 2008.

Popovicheva, O., Ivanov, A., and Vojtisek, M.: Functional Factors of Biomass Burning Contribution to Spring Aerosol Composition in a Megacity: Combined FTIR-PCA Analyses, 11, https://doi.org/10.3390/atmos11040319, 2020.

Saniger, JoséM.: Al-O infrared vibrational frequencies of γ-alumina, Materials Letters, 22, 109–113, https://doi.org/10.1016/0167-577X(94)00234-7, 1995.

Savitzky, A. and Golay, M. J. E.: Smoothing and differentiation of data by simplified least squares procedures, 36, 1627–1639, https://doi.org/10.1021/ac60214a047, 1964.

Schwarz, G.: Estimating the Dimension of a Model, 6, 461–464, 1978.

Takahama, S., Schwartz, R. E., Russell, L. M., Macdonald, A. M., Sharma, S., and Leaitch, W. R.: Organic functional groups in aerosol particles from burning and non-burning forest emissions at a high-elevation mountain site, Atmos. Chem. Phys., 11, 6367–6386, https://doi.org/10.5194/acp-11-6367-2011, 2011.

Takahama, S., Ruggeri, G., and Dillner, A. M.: Analysis of functional groups in atmospheric aerosols by infrared spectroscopy: sparse methods for statistical selection of relevant absorption bands, 9, 3429–3454, https://doi.org/10.5194/amt-9-3429-2016, 2016.

Takahama, S., Dillner, A. M., Weakley, A. T., Reggente, M., Bürki, C., Lbadaoui-Darvas, M., Debus, B., Kuzmiakova, A., and Wexler, A. S.: Atmospheric particulate matter characterization by Fourier transform infrared spectroscopy: a review of statistical calibration strategies for carbonaceous aerosol quantification in US measurement networks, Atmos. Meas. Tech., 12, 525–567, https://doi.org/10.5194/amt-12-525-2019, 2019.

Tsyganenko, A. A. and Filimonov, V. N.: Infrared spectra of surface hydroxyl groups and crystalline structure of oxides, Journal of Molecular Structure, 19, 579–589, https://doi.org/10.1016/0022-2860(73)85136-1, 1973.

Weckler, B. and Lutz, H. D.: Lattice vibration spectra. Part XCV. Infrared spectroscopic studies on the iron oxide hydroxides goethite (β), akaganéite (β), lepidocrocite (γ), and feroxyhite (δ), European Journal of Solid State and Inorganic Chemistry, 35, 531–544, https://doi.org/10.1016/S0992-4361(99)80017-4, 1998.

Wold, S., Martens, H., and Wold, H.: The Multivariate Calibration-problem In Chemistry Solved By the PLS Method, 973, 286–293, 1983.

Wold, S., Esbensen, K., and Geladi, P.: Principal component analysis, Chemometrics and Intelligent Laboratory Systems, 2, 37–52, https://doi.org/10.1016/0169-7439(87)80084-9, 1987.