1 **Quantification of major particulate matter species from a single filter type using**

2 **infrared spectroscopy – Application to a large-scale monitoring network**

3 Bruno Debus[1], Andrew T. Weakley[1], Satoshi Takahama[2], Kathryn M. George[1,3], Anahita Amiri

4 Farahani[1], Bret Schichtel[4], Scott Copeland[5], Anthony S. Wexler[1,6], Ann M. Dillner[1]*

5 [1] Air Quality Research Center, University of California, Davis, California, 95616, USA

6 [2] ENAC/IIE, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

7 [3]Monitoring and Laboratory Division, California Air Resources Board, Sacramento, CA 95811,

8 USA

9 [4]National Park Service, Cooperative Institute for Research in the Atmosphere, Colorado State

10 University, Fort Collins, CO 80523, USA

11 [5] Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins,

12 CO, 80523, USA

13 [6]Departments of Mechanical and Aerospace Engineering, Civil and Environmental Engineering,

14 and Land, Air and Water Resources, University of California, Davis, California, 85616, USA

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

16 **Abstract**

17 To enable chemical speciation, monitoring networks collect particulate matter (PM) on different

18 filter media, each subjected to one or more analytical techniques to quantify PM composition

19 present in the atmosphere. In this work, we propose an alternate approach that uses one filter

20 type (teflon or polytetrafluoroethylene, PTFE, commonly used for aerosol sampling) and one

21 analytical method, Fourier Transform Infrared (FT-IR) spectroscopy to measure almost all of the

22 major constituents in the aerosol. In the proposed method, measurements using the typical

23 multi-filter, multi-analytical techniques are retained at a limited number of sites and used as

24 calibration standards. At all remaining sites, only sampling on PTFE and analysis by FT-IR is

25 performed. This method takes advantage of the sensitivity of the mid-IR domain to various

26 organic and inorganic functional groups and offers a fast and inexpensive way of exploring sample

27 composition. As a proof of concept, multiple years of samples collected within the Interagency

28 Monitoring of PROtected Visual Environment network (IMPROVE) are explored with the aim of

29 retaining high quality predictions for a broad range of atmospheric compounds including total

30 mass, organic (OC), elemental (EC) and total (TC) carbon, sulfate, nitrate and crustal elements.

Findings suggest that models based on only 21 sites, covering spatial and seasonal trends in atmospheric composition, are stable over a three year period within the IMPROVE network with acceptable prediction accuracy ($R^2 > 0.9$, median bias less than 3%) for most constituents. The major limitation is measuring nitrate as it is known to volatilize off of PTFE filters. Incorporating additional sites at low cost, partially replacing existing, more time and cost intensive techniques or using the FT-IR data for quality control, are among the potential benefits of one-filter, one-method approach.

1 Introduction

In the United States, filter-based chemical speciation of ambient aerosols has been in operation for decades to quantify trends, assess transport and atmospheric transformation, identify sources of air pollution, evaluate impacts of pollution regulations, assess impacts on visibility, radiative forcing, human and ecosystem health and evaluate atmospheric and climatological models. The two federally funded speciation networks, the Interagency Monitoring of PROtected Visual Environments (IMPROVE) and the Chemical Speciation Network (CSN) collect 24-hour filter samples using three filter media: polytetrafluoroethylene for analysis by gravimetry, hybrid integrating plate and sphere (HIPS), and x-ray fluorescence (XRF), quartz for thermal optical reflectance (TOR) and nylon for ion chromatography. Over the decades of operation, the analytical methods have evolved with efforts to maintain consistency in trends while also adopting improved methodology and retiring obsolete equipment. Impacts of many of these changes have been addressed in the literature (Hyslop et al., 2015, 2012; White et al., 2016; Zhang et al., 2021; Chow et al., 2007a, 2015) and in data advisories posted on the IMPROVE website (http://vista.cira.colostate.edu/Improve/data-advisories/).

In this paper, we explore the use of Fourier transform-infrared spectroscopy (FT-IR) to reproduce most of the existing speciation data based on the optical activity of the components in the mid-IR. The number and bands of organic compounds are numerous, but generally group frequencies can be found above 1500 cm^{-1} and compound-specific spectral patterns ("fingerprint region") below this frequency; down to approximately 700 cm^{-1} (for example, Weakley et al., 2016; Mayo et al., 2004). Graphitic carbon displays peaks near 1600 cm^{-1} due to lattice defects (Tuinstra and Koenig, 1970; Friedel and Carlson, 1971), displacement vibrations near 868 cm^{-1} (Nemanich et al., 1977), and a broad, sloping absorbance between 4000 and 1500 cm^{-1} due to the tail of the electronic transition more strongly observed in the UV (Parks et al., 2021). Inorganic substances containing polyatomic ions such as sulfate, nitrate, and ammonium have strong vibrational modes above 600 cm^{-1} (Mayo, 2004). Crystalline and amorphous geological minerals in the form of oxides (which include hydroxides and oxyhydroxides) have distinct internal vibrational modes influenced by the electronegativity of the metal to which the oxygen is bonded (Busca and Resini, 2006; Chukanov and Chervonnyi, 2016; Margenot et al., 2017).

FT-IR spectra with partial least squares (PLS) calibrations have been shown to reproduce OC and EC concentrations using organic and graphitic carbon absorption bands, respectively, at a limited number of sites in the IMPROVE network (Dillner and Takahama, 2015a, b; Reggente et al., 2016), CSN (Weakley et al., 2016, 2018a) and FRM (Weakley et al., 2018b). Takahama et al. (2019) reviews these findings and the overall framework to be used for the two phases of such statistical calibrations: model building (sample selection, spectral preparation, model generation, model selection, model evaluation, and model understanding) and operation (error anticipation and model updating). Inorganic ions and geological mineral absorption bands have been used for

75   chemical speciation of these substances using FT-IR in prior studies (e.g., Cunningham et al.,
76   1974; McClenny et al., 1985; Pollard et al., 1990; Bogard et al., 1982; Foster and Walker, 1984).

77   Organic absorption bands are useful for measuring OC but also provide spectral information
78   needed to add detailed knowledge of composition not currently measured in air quality
79   monitoring networks – such as organic matter (OM) and organic functional group composition –
80   which is the subject of other work (Reggente et al., 2019; Boris et al., 2019, 2021; Burki et al.,
81   2020). Such calibrations, also combined with factor analytic approaches, can provide source
82   characterization on par with more costly mass spectrometric techniques (Boris et al., 2021;
83   Yazdani et al., 2021a; Hawkins et al., 2010; Takahama et al., 2011; Liu et al., 2012; Corrigan et al.,
84   2013).

85   Although FT-IR shows promise for measuring many constituents in aerosol, it is not without its
86   challenges. One limitation is that not all PM constituents can be measured, or measured with
87   high sensitivity, from the FT-IR spectrum. For instance, NaCl and $MgCl_2$ do not have IR-active
88   substituents. While a multitude of spectral signatures associated with mineral dust arise from
89   their constituent bonds – e.g., the metal-oxygen bonds in oxides (the oxide form is explicitly
90   assumed in estimating dust mass concentrations from elemental composition for the IMPROVE
91   network), some must be predicted from correlation with other constituents (e.g., some forms of
92   iron) if IR-activity is lacking. Other substances are IR-active but have weak responses, such as
93   graphitic carbon (Niyogi et al., 2006; Parks et al., 2021).   The absorption and scattering by the
94   PTFE filter also pose challenges for quantitative analysis. The PTFE-based material changes over
95   time due to change in manufacturer or manufacturing process, and is difficult to fully characterize
96   a priori or treat with simple blank subtraction techniques. PTFE absorption limits full access to
97   the range of spectroscopic information in the mid-IR, for instance in the region of carbon-oxygen
98   bonds that can lead to less than full recovery of OM mass. Additionally, scattering leads to
99   broadly-varying slope in the group frequency region. This scattering artifact is minimized by
100   baselining (Kuzmiakova et al., 2016) and using many standards that have a range of scattering
101   and absorption observed in the network (Debus et al., 2019), yet these techniques can still lead
102   to errors in quantification. Weakley et al., (2018b) demonstrated that calibrations built using one
103   brand of filter can be accurately extended to another brand of PTFE filter with numerically
104   marginal but statistically significant increase in method error (e.g., +2% error for α=0.05).
105   However, these studies are insufficient to generalize findings to all types of sampling filters.

106   The goal of this work is to assess the capability of using FT-IR to measure the aerosol chemical
107   composition at all IMPROVE sites. FT-IR quickly and non-destructively collects information-rich
108   spectra from routinely collected PTFE filter samples. Ambient samples from strategically-
109   selected IMPROVE sites are used for calibration and reasonably mimic the composition, matrix
110   effects and substrates of the unknowns, all of which theoretically lead to accurate estimations of
111   concentrations. Using all samples from selected sites reduces maintenance, shipping, processing

112 and coordinating required to maintain intermittent quartz and nylon filter sampling at every site.
113 Sites are selected using data from 2015 and are used for calibrating samples from 2015-2017.
114 Samples from all other (non-calibration) IMPROVE sites are predicted and compared to routine
115 IMPROVE data to assess the quality of prediction. Aerosol components to be measured include
116 TC, OC, EC, inorganic ions, soil elements, particulate mass, and light absorption.

2   Methods

2.1   Network data

119 IMPROVE samples collected every third day at all North American sites (Section S1) from January
120 2015 through December 2017 are included in this study. Fine particulate matter (aerodynamic
121 diameter less than 2.5 micrometers) is deposited on 25 mm diameter filters
122 polytetrafluoroethylene (PTFE, Teflo, Pall Gelman) and quartz filters by sampling air at a nominal
123 flowrate of 22.8 liters per minute from midnight to midnight local time. Parallel 37 mm nylon
124 filters are collected at the same flow flowrate. PTFE filters are analyzed by multiple instruments
125 and archived for future analysis. Nylon filters and a portion of each quartz filter undergoes
126 destructive analysis and a remaining part of the quartz filters are retained for archive.

127 Over the period covered in this study, two different TOR instruments were employed to measure
128 OC, EC and TC. Quartz filters sampled prior to 2016 where analyzed on a DRI Model 2001 thermal
129 optical carbon analyzers (Chow et al., 1993) while filters collected beginning in January of 2016
130 were analyzed on a DRI Model 2015 multi-wavelength thermal optical carbon instrument (Magee
131 Scientific – Berkley, USA)(Chow et al., 2015). Both instruments use the IMPROVE_A protocol
132 (Chow et al., 2007b), which outlines the temperature step, gaseous environment in the
133 instrument and that reflectance is used to define the split point between OC and EC. To correct
134 for gas phase adsorption onto the quartz filter, the monthly median field blank OC concentration
135 is subtracted from each OC measurement during that sample month. Carbon concentrations are
136 reported in $\mu g/m^3$.

137 An in-house Hybrid Integrating Plate and Sphere (HIPS) system evaluates light absorption from
138 the PTFE filters in the IMPROVE network (White et al., 2016). In this work, the measured
139 absorption coefficient (*Fabs*) is converted into a TOR EC equivalent concentration assuming a
140 Fabs / EC ratio of 10 $m^2 g^{-1}$ (Malm et al., 1994). The resulting value, referred to as HIPS Black
141 Carbon (HIPS BC), is used as part of a quality control procedure to evaluate potential outliers in
142 TOR EC measurements.

143 Data from gravimetry and X-ray fluorescence (XRF) analysis obtained from PTFE filters and ion
144 chromatography from the nylon filters are also used in this study. Additional information on
145 routine   IMPROVE   methods   can   be   found   on   the   IMPROVE   website
146 (http://vista.cira.colostate.edu/Improve/).   IMPROVE   data   are   available   online   at
147 (http://views.cira.colostate.edu/fed).

2.2 Outlier removal

Data were screened for outliers to eliminate their influence on the results. Out of the ~61,500 total number of samples in the three-year period, fewer than 800 were excluded from the analysis due sampling issues or missing TOR, XRF or IC data. In addition, 65 samples collected at the Wheeler Peak Wilderness (New Mexico) site between November 2015 and April 2016 were excluded due to an EC contamination caused by a diesel-powered ski lift.

Potential outliers in TOR measurements were investigated by regressing TOR EC against HIPS BC concentrations. Samples with differences exceeding a predefined threshold value (0.68 µg/m^3) were flagged as potential outliers (section S2). The status of these samples was confirmed by building separate TOR EC and HIPS BC calibrations. The poor agreement between TOR EC and FT-IR EC concentrations contrasts with the nearly 1:1 relationship HIPS BC and FT-IR BC predicted values indicating that TOR EC concentrations were likely compromised (Section S2). For the period considered in this study, 112 samples with faulty TOR EC values were identified and excluded from further analysis. The number of valid sample spectra retained in this study is 61,462.

2.3 Fourier-transform infrared (FT-IR) analyses

Since 2015, all PTFE in the IMPROVE network have been analyzed by infrared spectroscopy for research and evaluation purposes. FT-IR measurement occurs after gravimetric analysis and prior to XRF and HIPS to prevent possible loss of volatile species under the mild vacuum in XRF. Three FT-IR spectrometers including one Tensor 27 and two Tensor 2 instruments (Bruker Optics, Billerica, MA) equipped with a pre-aligned mid-IR source and a liquid nitrogen-cooled wide-band mercury cadmium telluride (MCT) detector were used for spectra acquisition in the range 4000 - 420 cm^{-1} by averaging 512 scans at a nominal resolution of 4 cm^{-1}. The single beam signal associated with each PTFE filter was converted to an absorbance spectrum using the most recent zero reference signal, updated hourly.

Previously, it was determined that calibration transfer between multiple FT-IR instruments is not required as long as their spectral response is carefully matched by controlling a set of key environmental and instrumental parameters (Debus et al., 2019). Briefly, each mercury cadmium telluride (MCT) detector is connected to an automatic liquid nitrogen micro dosing system (NORHOF, Ede, Netherlands) designed to improve signal stability and maintain a high signal to noise ratio. The repeatability and reproducibility of the filter position relative to the IR beam is controlled via a house-built sample chamber (4.0 × 5.1 × 4.5 cm) mounted inside the instrument sample compartment. Details regarding the chamber design have been published elsewhere (Debus et al., 2019). Finally, the contribution of water vapor and carbon dioxide to the signal was minimized by continuously purging both the sample chamber and the optical bench with a VCD Series CO_2 adsorber / dryer system (PureGas LLC, Broomfield, CO). For each sample, the acquisition procedure involves a 4 minutes purge period followed by a spectrum collection lasting

185 about 1 minute. An in-house macro interfaced to the OPUS software (Bruker Optics, Billerica,
186 MA) controls each step. PTFE filters were measured in transmission mode without sample
187 preparation. No interpolated data (from zero-filling) are included in the final raw spectra.
188 Collected spectra are subjected to weekly quality control procedures detailed in (Debus et al.,
189 2019). Duplicate and replicate measurements were also performed to evaluate instrument
190 stability and found to be within +/- 10%.

191 ## 2.4   Multivariate Calibration Model - Partial Least Squares (PLS) Regression
192 While the presence of certain category of atmospheric compounds can be identified qualitatively
193 from an FT-IR spectrum, an accurate quantification of their concentration requires calibration.
194 PLS is a commonly used algorithm to relate a multi-wavenumber measurement to any particular
195 sample properties such as concentration (Wold et al., 2001). In brief, PLS maximizes the co-
196 variance between a set of response variables (species measurements) and a reference
197 measurement (FT-IR spectra) from which equivalent predicted values are desired. In so doing,
198 the optimal combination of response variables best describing the reference measurement is
199 identified and the selected features are used to build a multivariate calibration. With all least-
200 squares calibration methodologies, concentration-dependent biases in residuals that are
201 determined by the quality of fit (R^2) and dynamic range of the data are expected due to the nature
202 of least-squares estimation (Besalú et al., 2006; Draper and Smith, 1998, pp. 63-64,173,638). For
203 further discussion of these biases, see Section S1.

204 The applicability of PLS to quantify carbonaceous aerosol species (Reggente et al., 2016; Weakley
205 et al., 2016, 2018a) or functional groups (Boris et al., 2019; Ruthenburg et al., 2014) collected on
206 PTFE filters in various monitoring networks and field campaigns has been successfully
207 demonstrated. A complete review of the implementation of PLSR calibration in the framework
208 of atmospheric particulate matter characterization has been recently published (Takahama et al.,
209 2019).

210 To evaluate model performance, FT-IR predicted concentrations were regressed against their
211 reference measurement to quantify residuals and a series of metrics. Reported figures of merit
212 include the coefficient of determination (R^2), bias, error and the method detection limit (MDL).
213 Residuals are defined as the difference between predicted and reference concentrations, bias
214 corresponds to the median residual while error is the median absolute residual. To facilitate inter-
215 model comparison, relative performance metrics were calculated by normalizing the values by
216 their reference value. FT-IR PLSR calibration MDL was estimated from field blank predicted
217 concentrations as the 95th percentile minus the median residuals, as is done for other species in
218 the   IMPROVE   network   http://vista.cira.colostate.edu/improve/wp-
219 content/uploads/2021/07/IMPROVE-SOP-351_Data-Processing-and-Validation_2021_final.pdf.
220 Performance is reported for all samples together regardless if the samples were included in the
221 calibration. This enables comparison between models with different samples used for calibration.

222 For further insight into model prediction accuracy, the distribution in FT-IR residuals is
223 qualitatively compared with residuals from collocated measurements. Collocated quartz filters
224 are collected at the Everglades (FL), Hercules-Glades (MO), Medicine Lake (MT) and Phoenix (AZ)
225 sites. Similarly, collocated Teflon filters are sampled at Mesa Verde (CO), Proctor Maple Research
226 Facility (VT), Saint Marks National Wildlife Refuge (FL), Yosemite (CA) and Phoenix (AZ) sites while
227 collocated nylon filters are featured at the Phoenix (AZ), Frostburg Reservoir (MD), Mammoth
228 Cave (KY) and San Gabriel (CA) sites.

229 Data handling and analysis was performed in Matlab R2018a (The MatWorks, Inc, Natick, MA,
230 United States) using the statistics and signal processing toolboxes. PLS was computed via the
231 libPLS Matlab package (v1.9) (Li et al., 2018).

2.5   FT-IR Calibrations for Predicting PM Composition

233 This section presents the design of calibrations for quantifying the concentration of major
234 atmospheric species by taking advantage of the composition-based information embedded
235 within an FT-IR spectrum. In practice, spectra are calibrated against reference measurements
236 from TOR, XRF, IC, HIPS and gravimetric analysis with the aim of predicting concentrations of
237 atmospheric constituents using only spectra of PTFE filters as input.

238 A multilevel model (Snijders and Bosker, 2011; Takahama et al., 2019) is proposed in which
239 dedicated calibration models for subgroups of samples are constructed, and applied according
240 to a predetermined selection criterion for each sample. This model considers two subgroups: i)
241 samples determined to be dominated by biomass burning, which are calibrated with similar
242 samples, and ii) the remaining samples, which are calibrated with samples from a limited number
243 of sites.

244 To establish baseline performance metrics for comparison, a "Global model" in which a single
245 calibration (for each species) is constructed from all samples considered together is described in
246 Section S1 (Supplement).

247 The first step in the development of the Multilevel model consists of screening for biomass
248 burning samples. These samples are removed from consideration during the site selection
249 process. A simple detection method combining estimates of key functional group spectral peak
250 areas and spectral dissimilarity metrics were used to segregate biomass burning samples from all
251 other samples. Next, a Gaussian Mixture Model (GMM) was applied to the spectra of all non-
252 biomass burning samples. The GMM exploits the specificity of the infrared signal for organic and
253 inorganic species. The GMM was implemented with the aim of clustering the non-biomass
254 burning FT-IR spectra into groups sharing similar spectral features (Section 2.5.2). Those groups
255 were later used as part of the methodology for selecting sites with representative atmospheric
256 composition. Spectra from the year 2015 were used as a benchmark to validate the biomass
257 burning detection strategy, build the GMM and establish the list of representative sites where

258 multi-filter collection/multi-analyses should be retained (section 2.5.2). The identified biomass
259 burning samples are used to build a calibration for biomass burning samples (Section 2.5.1).

260 ## 2.5.1   Biomass burning model
261 FT-IR spectra were used to estimate functional group areas and calculate spectral dissimilarities
262 metrics to segregate biomass burning samples from all other samples. Although this paper
263 focuses on using FT-IR to measure the major aerosol components in routine speciated aerosol
264 monitoring networks, FT-IR is more frequently used to measure organic functional groups (e.g.
265 (Russell et al., 2011; Ruthenburg et al., 2014; Boris et al., 2019). Specific regions in the IR spectra
266 correspond to specific functional groups. Peak areas, calculated from baseline corrected spectra
267 (see Section S3 for baseline procedure), for carbonyl, OH and CH were used rather than functional
268 group calibrations for simplicity. Because the relative functional group peak area tends to
269 increase significantly as the cumulative peak area decreases, typically for low mass deposition
270 samples, an estimate of spectral dissimilarities, the squared Mahalanobis distance (D_i^2), for each
271 site is also considered to minimize false detection. The Mahalanobis distance (Mahalanobis,
272 1936; Cios et al., 1998) is a measure of the spectral dissimilarity between a given spectrum at a
273 site and the mean spectrum at the site. Taking advantage of D_i^2 and relative functional group
274 areas, a set of criteria were established from observations at known wildfire sites during wildfire
275 season (O'Dell et al., 2019). First, samples collected under heavy smoke conditions whose spectra
276 fulfill C–H ≥ 2 %, C=O ≥ 15 % and $D_i^2 \geq 3\,\overline{D^2}$ were detected (Section S3). This group of spectra
277 tend to have large D_i^2 values and, consequently, the $3\,\overline{D^2}$ threshold often excludes samples with
278 low to moderate biomass burning contributions. For a more inclusive detection, spectra from the
279 first group were removed from consideration, the D_i^2 values are updated for each sample and
280 the plots were regenerated. The cut-off value for the relative carbonyl functional group area was
281 lowered to 8 % while other parameters were not changed. Spectra identified by the first and
282 second rounds are considered biomass burning samples. This procedure is performed for each
283 site and for each year of sample collection (Section S.3)

284 Recent work has shown that smoke samples may be identified using techniques such as cluster
285 analysis and labeling (Burki et al., 2020) similar to the GMM used here and through detection of
286 molecular markers – levoglucosan and lignin – or peak profiles in FT-IR spectra (Yazdani et al.,
287 2021a, b). For the large data set in this work (~20,000 samples in 2015), cluster analysis
288 resulted in multiple clusters that could be associated with smoke-impacted samples likely due
289 to the variations in fuel, oxidation conditions, and contributions from other sources. Therefore,
290 for this work we selected a single group of smoke-impacted samples based on specific organic
291 features known to be present in FT-IR spectra. While the criteria for smoke-impact labeling can
292 be defined differently according to each intended purpose, the method presented here is
293 demonstrated to sufficiently partition the samples for building accurate submodels to predict
294 concentrations of PM constituents.

295 While ions and crustal species are not necessary correlated with wildfire emissions, the Biomass
296 Burning sub-model accounts for interferences that are necessary to track in order to maintain
297 high prediction accuracy for samples collected on smoky days.

2.5.2 Limited Sites Model

299 To assess major PM$_{2.5}$ composition regimes in the network and to identify representative sites to
300 use as calibration standards in the Limited Sites model, screening of all FT-IR spectra (except
301 samples identified as biomass burning samples) across all locations and seasons was performed
302 by building a Gaussian Mixture Model (GMM) (Bilmes, J. A., 1998; Hastie, T et al., 2009). The basic
303 idea behind GMM is to group FT-IR spectra into clusters of similar spectral shape using a
304 probabilistic approach describing the likelihood that any given spectrum belongs to a particular
305 class. To minimize the concentration dependence and emphasize composition, raw spectra were
306 transformed to second derivative spectra using a 2nd order, 21 point, Savitzky-Golay filter
307 (Savitzky and Golay, 1964), differenced with filter blank spectra and divided by their respective
308 Euclidean norm (Bro and Smilde, 2003). Additional details about the GMM pre-processing and
309 implementation as well as cluster interpretation are provided in Section S4.

310 After classification, a single site per cluster was selected to represent the atmospheric
311 composition captured in that cluster. For any given cluster, the retained location was defined as
312 the site with the largest number of classified spectra with the highest probabilities of belonging
313 to that cluster. To prevent misleading site selection and enhance spatial coverage, the following
314 set of decision rules were used: *i)* if the same site is representative of two clusters, it is ascribed
315 to the cluster with the largest number of classified spectra from that site, *ii)* if none of the
316 retained sites accounts for a given spatial region or known source type in the network, an
317 additional site with the highest number of classified spectra is selected from a nearby cluster,
318 and *iii)* only sites under continuous operation between 2015 and 2017 are eligible for selection.
319 Criteria *ii)* was invoked once to add a site in the Midwest to improve spatial coverage in that
320 region and to capture prescribed fire emissions in Kansas. All non-biomass burning samples from
321 selected sites were used as FT-IR calibration standards for all species and all non-biomass burning
322 samples are predicted with these models. Once established, the selected sites are not re-
323 evaluated but instead were used in all subsequent years as would occur in practice.

2.5.3 Application of Multilevel Model

325 The Multilevel model is the combined FT-IR predicted concentrations from the Limited sites and
326 Biomass burning models. Multilevel modeling will be discussed in the context of carbonaceous
327 aerosols before extending the modeling to other atmospheric constituents with detectable
328 infrared signatures. In addition to OC and EC, species evaluated for FT-IR prediction include PM$_{2.5}$
329 mass, soil elements (silicon, aluminum, calcium, titanium, iron), anions (sulfate, nitrate) and HIPS
330 BC. As mentioned previously, NaCl is not IR active and so there is no direct measure of seasalt
331 from FT-IR. Next, the years 2016 and 2017 will be examined to assess the long-term stability of

332 the proposed Multilevel strategy by screening for smoke samples and re-calibrating each year
333 using the sites selected using 2015 data.

3 Results and discussion

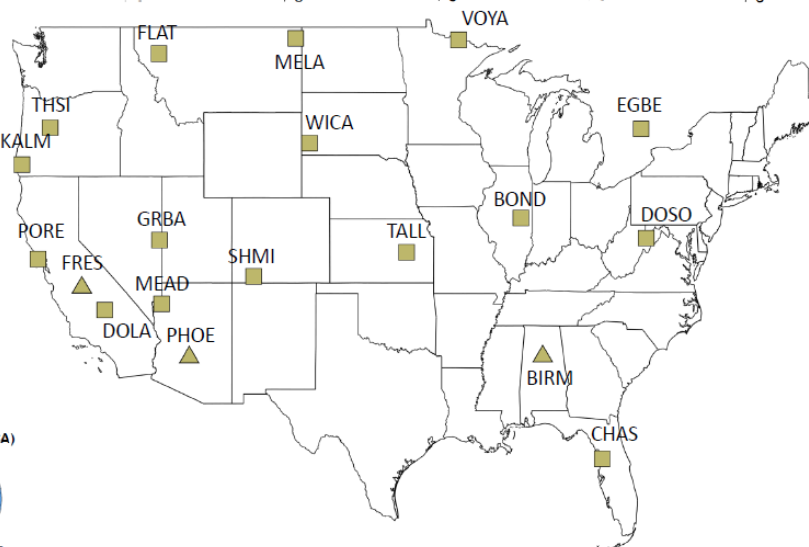
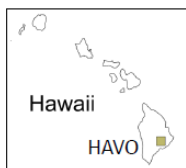
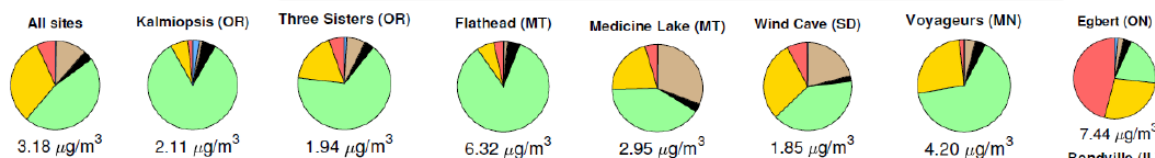
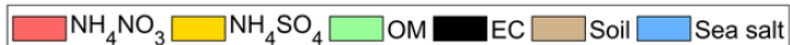
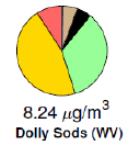
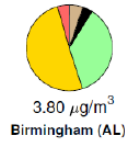
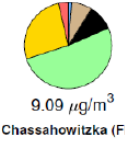
335 In the following sections, the quality of FT-IR based calibrations for quantifying aerosol
336 composition across continental US and their long-term applicability to large speciation
337 monitoring networks will be assessed. Section 3.1 describes the selected calibration samples for
338 the Biomass Burning and Limited Sites models. In Section 3.2, Biomass Burning and Limited
339 model performance will be briefly reviewed before exploring the Multilevel FT-IR predictions for
340 all samples. Initially focused on carbonaceous species on PTFE samples collected in 2015, FT-IR
341 predictions will be extended to other atmospheric constituents and years.

3.1 Multilevel modeling – Calibration sample selection

3.1.1 Biomass burning sample selection

344 Using the methods described above, 492 samples impacted by biomass burning emissions were
345 identified in 2015 (2.5 % of the network), 288 samples in 2016 (1.5 %), and 817 samples in 2017
346 (3.7 %). The mean OC concentration of the biomass burning samples range was 5.6 – 8.3 $\mu g/m^3$
347 with maximum concentrations extending from 44 to 97 $\mu g/m^3$ over the three year period.
348 Similarly, per year, the mean EC concentration varies between 0.6 – 0.9 $\mu g/m^3$ with maximums
349 up to 2.9 – 3.9 $\mu g/m^3$. Mean OC/EC ratios are larger than 7, in agreement with past literature
350 (Schichtel et al., 2008; Sorooshian et al., 2011). Analysis of the detected samples shows reliable
351 spatial and seasonal distributions, consistent with biomass burning emissions predominantly in
352 summer and fall across the Pacific North West and Northwestern US (Section S3). Two-thirds of
353 the identified samples were selected (Section S5) as calibration standards for the calibration and
354 resulting model was applied to the remaining third of the smoke impacted samples.

3.1.2 Limited Sites model – clusters and retained sites

356 Figure 1 shows the spatial distribution and annual average composition (from routine IMPROVE
357 data) of the 21 sites selected for the Limited sites model calibratoin. From a spatial standpoint,
358 retained sites appear reasonably scattered across the network including Hawaii and the Virgin
359 Islands. Clusters are represented by a distribution of urban and rural sites. One urban cluster is
360 represented by Fresno and contains mostly urban samples from Fresno and Phoenix. All other
361 clusters contain mostly rural and pristine sites. However, two other urban sites were retained,
362 Phoenix and Birmingham. The Phoenix cluster contains samples from the southwest in the
363 spring. The Birmingham site along with the Tallgrass site represent a non-western cluster in the
364 spring and summer.

367

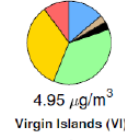
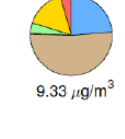
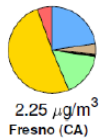
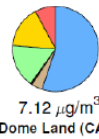
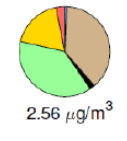
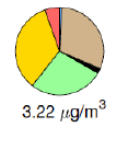
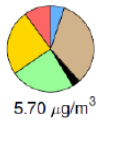
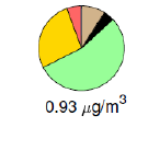
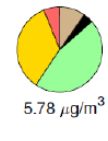
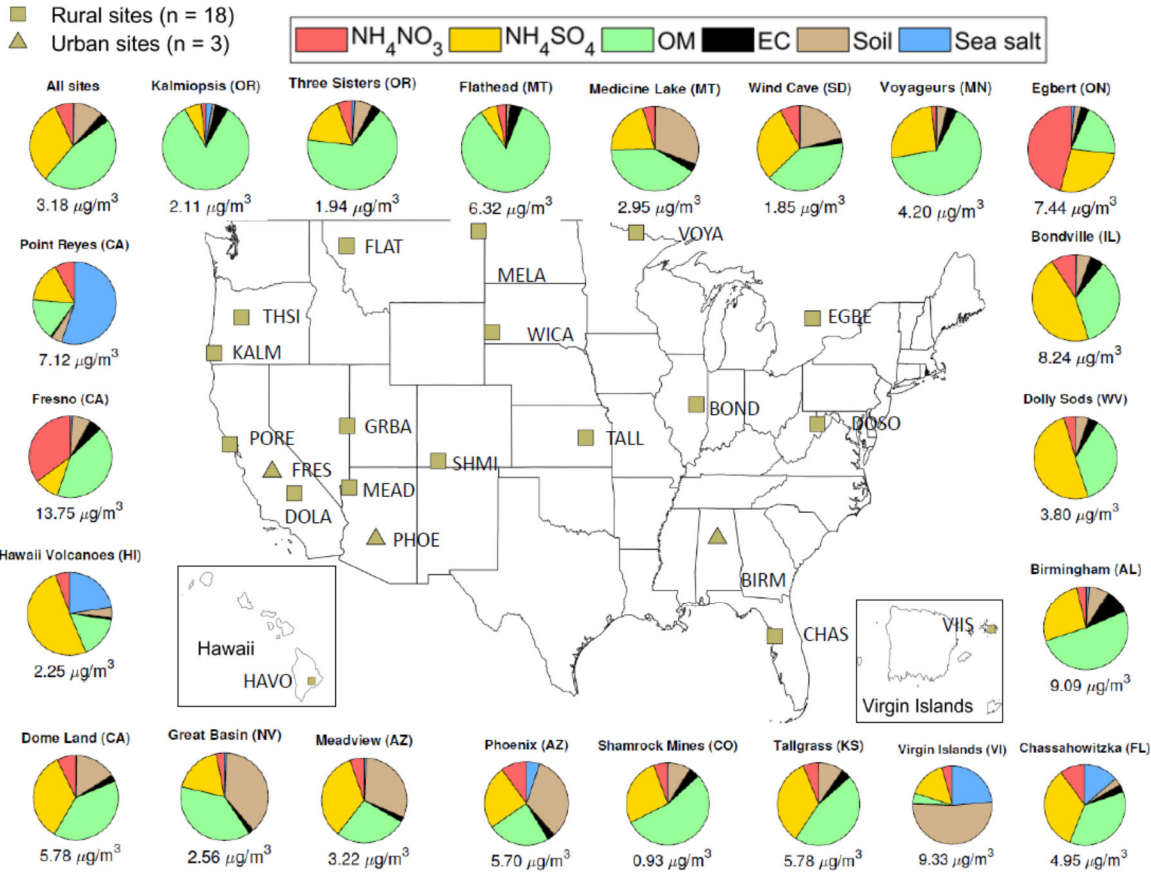
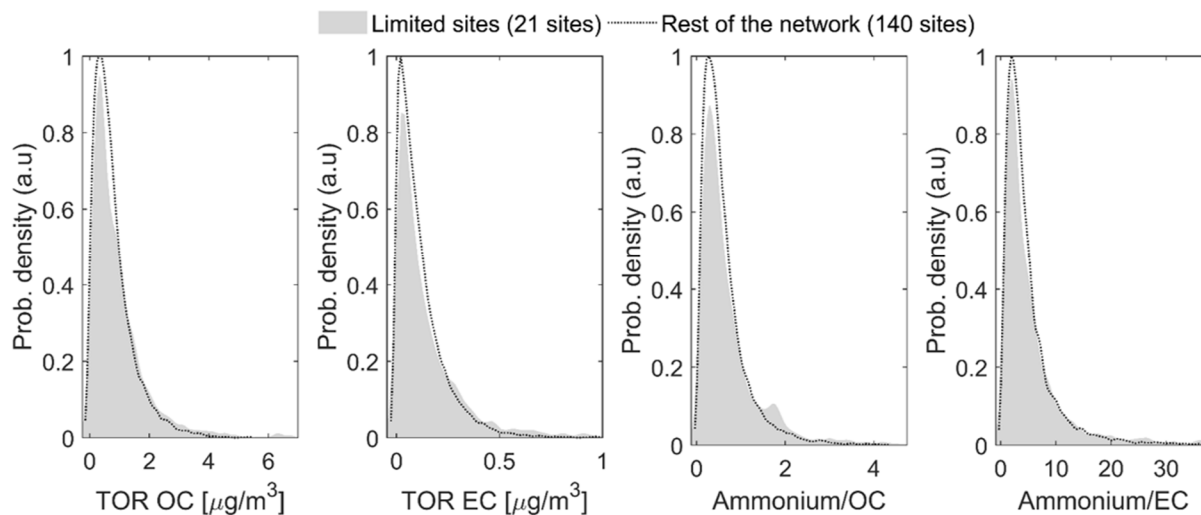



Figure 1: Spatial distribution, median PM$_{2.5}$ concentration and composition of the 21 representative sites. The composition is obtained from routine IMPROVE (non-FT-IR) measurements and the IMPROVE reconstructed fine mass equation (http://vista.cira.colostate.edu/Improve/reconstructed-fine-mass/) to calculate soil and seasalt. Sites are identified by the four letter site code which is the first four letters of a single word site name (Fresno = FRES) or the first two letters of the first and second word for two word site names (Dome Land = DOLA). ~~The composition is obtained from IMPROVE measurements and the IMPROVE reconstructed fine mass equation.~~ The top left pie chart representing the median PM$_{2.5}$ composition across all sites and samples is given for comparison.

The clusters are also seasonally distributed (Section S6): five clusters are dominated by fall - winter samples, ten clusters containing summer samples (along with varying number of spring and fall samples), two clusters are predominately spring and one is spring - fall. Three clusters have little seasonality.

381 Because FT-IR spectra are clustered based on composition, the first step in assessing the
382 representativeness of the 21 sites is to compare the concentration ranges. For this purpose,
383 distributions in TOR OC and EC concentrations excluding biomass burning samples are compared
384 for the 21 sites used for calibration and the 140 remaining sites. In Fig. 2, the two probability
385 density functions are very similar for both OC and EC despite large differences in sample
386 populations (2572 and 16,543, respectively). In addition to matching the range of carbonaceous
387 concentrations observed in the rest of the network, the presence of species interfering with
388 organic functional groups should also be accounted for by the calibration. Because ammonium
389 absorptions overlap with carbonaceous absorptions, situations where ammonium to OC and
390 ammonium to EC ratios are different between calibration samples and samples to be predicted
391 were associated with additional sources of bias and error (Dillner and Takahama, 2015a, b).
392 Although not measured in IMPROVE, ammonium concentration is approximated from nitrate and
393 sulfate assuming both species are fully neutralized. The corresponding probability distribution in
394 Fig. 2 confirms the equivalence between the ranges of ammonium/OC and ammonium/EC
395 concentrations spanned by the Limited sites samples and the overall network. In section S6, Fig.
396 S6-3 shows reasonable agreement between the selected sites and the rest of the network for
397 PM2.5 mass, ions, elements and HIPS BC. Together, these results suggest the list of 21 sites is a
398 suitable representation of network variations in OC and EC and their relative proportion to
399 ammonium, and for all other predicted constituents.

400

401 **Figure 2**: Comparison of probability density function for OC, EC and ammonium concentrations
402 in 2015 between the 21 sites retained for Limited calibration and the rest of the network.

403 The spatial and seasonal as well as the urban and rural diversity supports the compositional
404 diversity of the selected sites as shown in Fig. 1.   The three urban sites have distinct
405 characteristics. At the Fresno, CA site, the composition is dominated by nitrate (35 %) and organic

406 matter (42 %) with an autumn – winter pattern consistent with agriculture and residential wood
407 burning activities (Ngo et al., 2010) as well as with the formation of secondary aerosols during
408 stagnation events and a low inversion layer (Watson and Chow, 2002). Phoenix, AZ site features
409 a strong soil component (33%) associated with spring dust storms and windblown dust and equal
410 proportions of ammonium sulfate (25 %) and OM (24 %) occurring mostly in spring and summer.
411 The ammonium sulfate and organic matter has been attributed to regional power production and
412 traffic (Brown et al. 2007). In contrast, Birmingham samples show little seasonal trend with
413 elevated OM (52 %) and EC (10 %) fractions originating from various combustion processes
414 including vehicle exhaust, biomass burning and biogenic secondary organic aerosols (Blanchard
415 et al. 2016). The other dominant species at this site is ammonium sulfate (26 %), characteristic of
416 coal burning and industrial activities in the East (Watson et al. 2015).

417 Among rural sites, four noticeable patterns in $PM_{2.5}$ composition are distinguishable. The first
418 corresponds to OM fractions accounting for more than two-thirds of the filter mass. High OM
419 samples are encountered at four locations in Northwestern US, namely the Kalmiopsis (OR),
420 Three Sisters (OR), Flathead (MT) and Voyageurs (MT) sites. Samples from Voyageurs (MN) and
421 Flathead (MT) sites are from Summer-Fall and present elevated median $PM_{2.5}$ concentrations
422 (4.20 $\mu g/m^3$ and 6.32 $\mu g/m^3$, respectively) and very large percentage of OM consistent with
423 biomass burning emissions. Kalmiopsis (OR) and Three Sisters (OR) samples have a lower and
424 nearly identical median $PM_{2.5}$ concentration (≈ 2 $\mu g/m^3$) but differ in their monthly distribution
425 with the former displaying more winter samples than any other season whereas the later shows
426 little seasonality.

427 The second type of sites have high OM and sulfate concentrations. Both Shamrock Mines (CO)
428 and Tallgrass Prairie (KS) sites have larger OM than sulfate content. However, the Colorado site
429 has more autumn – winter samples, represents samples in the Rockies and Alaska and an overall
430 small median $PM_{2.5}$ concentration (< 1 $\mu g/m^3$). The Kansas site has a majority of spring samples,
431 representing non-western samples and has a significantly larger $PM_{2.5}$ concentration (≈ 6 $\mu g/m^3$)
432 that is attributed to prescribed burning near the Tallgrass site (Whitehill et al. 2019).   Other sites
433 have higher median sulfate concentrations (~50%) than OM concentrations (~40%) such as Dolly
434 Sods (WV) and Bondville (IL). The monthly sample distribution indicates seasonal influences:
435 Bondville (IL) samples are mostly from the summer and the concentrations are relatively high
436 while the Dolly Sods (WV) site samples are mostly not in the summer with lower concentrations.
437 Because the spectra were normalized to minimize influence of concentration, these two clusters
438 likely have different organic composition even though the relative amount of OM is the same.
439 Finally, situations where sulfate and OM are present in equal proportions (≈ 36 %) are reported
440 at the Dome Land (CA) and Chassahowitzka (FL) pristine sites mainly featuring spring – summer
441 and winter samples, respectively.

442 A third group of noteworthy PM$_{2.5}$ compositions at rural sites contain a large fraction of (> 20 %
443 of the total mass). The Virgin Islands (VI) site presents the highest soil fraction across the network
444 52 % of the total PM$_{2.5}$ mass, mostly originating from long-range Sahara soil dust transport
445 (Holmes and Miller 2004). In addition to sulfate and OM, elevated soil contributions are observed
446 for the Wind Cave (SD), Meadview (AZ), Medicine Lake (MT), and Great Basin (NV) sites with soil
447 content between 20 and 40%. Although the seasonality is somewhat different between these
448 sites, they all have many samples from the spring suggesting the dust is due at least in part to
449 spring dust storms and may also contain resuspended road dust and more localized dust sources.

450 A fourth and final distinct category of PM$_{2.5}$ composition includes a collection of sites with unique
451 local atmospheric pollution sources, specific to those locations. The Hawaii Volcanoes (HI) site
452 where sulfur emitted as part as the volcanic activity, contains 51% sulfate along with sea salt (23
453 %). Another location with unique composition is the Point Reyes (CA) site where the sea salt
454 contribution reaches 55% of the median filter mass for the clustered samples, larger than any
455 other marine site in the network. Finally, the Egbert (ON) Canadian site, representing the upper
456 Midwest in winter is dominated by nitrate (46 %), sulfate (27 %) and OM (20 %).

457 As described above, the 21 sites retained for the Limited sites sub-calibration present seasonal,
458 regional and compositional features consistent with known or expected trends in PM$_{2.5}$ across
459 the network. The median PM$_{2.5}$ mass at those locations covers a broad range of concentrations
460 ranging from 0.93 µg/m^3 to 13.75 µg/m^3 and includes both urban and rural sites. Capturing the
461 large variability in PM$_{2.5}$ composition and concentration is essential to ensure the proposed site
462 list is a representative subset of the parent network. However, it should be mentioned that the
463 proposed site list is not unique but is one of the many feasible solutions since sites whose samples
464 clustered together in the GMM are likely inter-exchangeable.

3.2 Evaluation of Biomass Burning Model

466 Prior to describing the overall results from the Multilevel model, the Biomass Burning model is
467 evaluated to determine if the biomass burning model improves predictions for those samples.
468 To evaluate the quality of the biomass burning model, the predictions are compared to a global
469 model (section S1) which contains a few samples from all 160 sites which are mostly non-smoke
470 samples but does contain a few smoke samples. Visual inspection of Fig. 3 suggests the
471 equivalence of the biomass burning models to the global model at the lower end of the
472 concentration range. However, improvement in prediction accuracy can be claimed at high
473 concentrations for the Biomass Burning model. The gain in model performance is subtle for OC
474 and TC; however, for EC, predictions benefit from having a dedicated calibration for samples
475 impacted by wildfire emissions, with an increase in R^2 from 0.747 to 0.902 (Section S7).

476

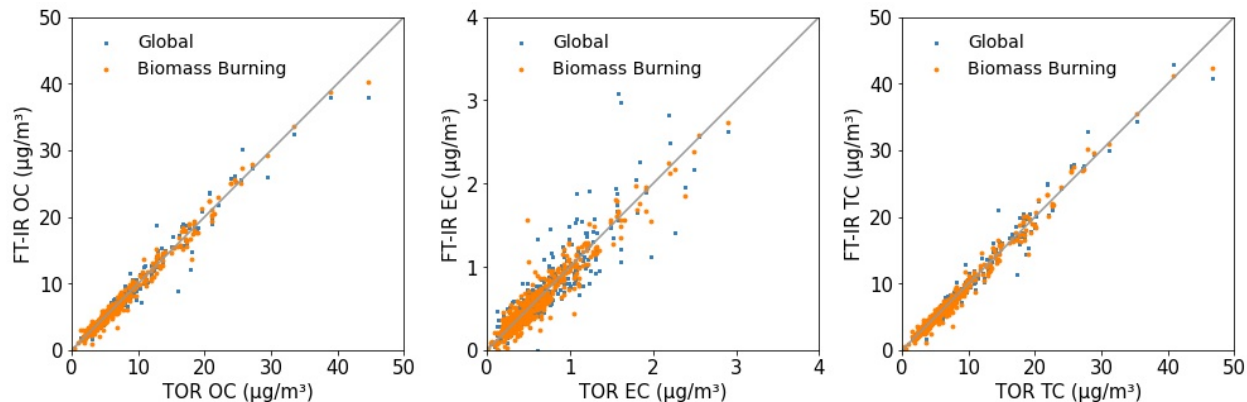



Figure 3: Inter-model OC (left), EC (middle), TC (right) comparison between global (section S1) and Biomass Burning predicted concentrations for the 492 samples classified as biomass burning in 2015. EC prediction, in particular, benefit from having a dedicated Biomass Burning calibration model.

Therefore, we retain the biomass burning model as part of the multilevel model and present the results for the Multilevel model below.

3.3   Multilevel modeling – Performance evaluation

3.3.1 Carbonaceous aerosol predictions

Figure 4 shows the correspondence between FT-IR Multilevel concentrations for OC and EC and TOR measurements for 2015 (plot for TC can be found in Section S9) and Table 1 lists the prediction metrics for all 3 carbonaceous components. The visual agreement between FT-IR and the reference measurements of OC and EC is high but EC shows higher scatter than the other measurements. Table 1 indicates that FT-IR OC and TC has higher prediction quality than EC but both perform satisfactorily. FT-IR OC and TC error is on par with TOR precisions (Table 1) indicating that on average FT-IR does not add significant additional error to the measurement. FT-IR EC predictions, however, have higher error than TOR precision. With respect to reference (TOR) measurements, concentration-dependent biases in residuals that are determined by the quality of fit (R^2) and dynamic range of the data are expected due to the nature of least-squares estimation (Besalú et al., 2006; Draper and Smith, 1998). For bias defined as FT-IR predictions minus the reference (TOR) measurement, least-squares estimator causes an apparent linear bias which is positive at the low end of the concentration range and negative at the high end of the concentration range. (see Section S8 for further discussion). The satisfactory agreement between FT-IR and TOR concentrations as well as the equivalence agreements using the global model (Section S1) support the validity of the proposed Multilevel modeling in the context of carbonaceous aerosols prediction from PTFE filters in large speciation networks.

502

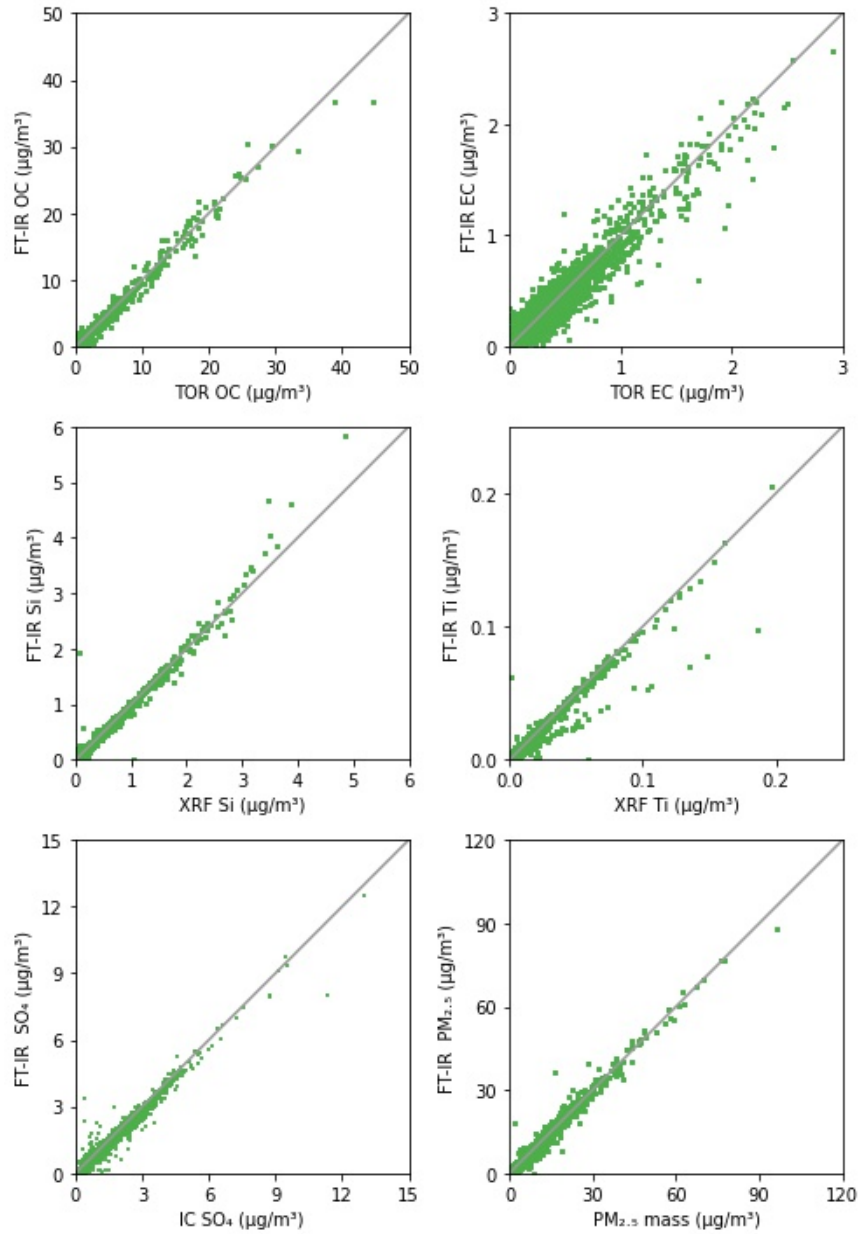



Figure 4: Comparison of predicted FT-IR OC, EC, Si, Ti, SO₄ and mass concentrations using the Multilevel model against their reference measurements. Each subplot contains all 19,608 samples collected in the year 2015.

506

507 **Table 1**: Summary of Multilevel model performance for IR-active atmospheric constituents for
508 19,608 spectra analyzed by FT-IR in the year 2015.

Species	R^2	Bias [μg/m³]	Bias (%)	Error [μg/m³]	Error (%)	Reference Data Error[1] (%)	MDL [μg/m³]	< MDL (%)
OC	0.983	0.01	1.6	0.08	12	10	0.06	0.9
EC	0.912	0	1.7	0.02	30	15	0.04	20.7
TC	0.984	0.01	1.2	0.08	12	11	0.07	1.3
BC	0.92	0	-0.3	0.03	23	---	0.04	19.3
Si	0.983	0	2.2	0.01	11	13	0.01	9.7
Al	0.985	0	2.2	0	12	10	0	4.7
Ca	0.979	0	1.1	0	13	9	0	6.9
Ti	0.941	0	2.7	0	21	16	0	14.9
Fe	0.95	0	1.1	0	25	8	0.01	19
SO₄	0.983	0	0.1	0.03	6	2	0.03	0.9
NO₃	0.927	0.02	15.3	0.07	54	8	0.07	21.8
PM₂.₅ Mass	0.985	0.03	1	0.18	6	6	0.25	1.1

509 [1]Median relative error for TOR, XRF, IC and gravimetric analysis. OC, EC and TC median relative error estimated
510 from collocated sampling as measurement error/uncertainty is not reported by IMPROVE for this components.
511 For all other components, the normalized error was calculated as the uncertainty divided by the concentration
512 prior to selecting the median. BC is not reported by IMPROVE so measurement error is not estimated.

513 In addition to OC, EC and TC, light absorption which is predominantly due to black carbon, is also
514 a measure of one fraction of the carbonaceous aerosol. FT-IR calibrations are found to be
515 adequate for replicating HIPS BC measurements (Section S9). As expected, the corresponding
516 model is similar in performance to ~~its~~ EC with R^2 and relative error of 0.920 and 23.3 %,
517 respectively (Table 1). FT-IR BC residuals have a broader interquartile range than in the HIPS BC
518 collocated data (Section S9). We attribute this effect to a difference in signal to noise ratio and
519 sensitivity to chemical interferents between the two analytical techniques. While HIPS exploits
520 the strong absorption properties of refractory carbon in the visible domain, the weak absorptivity
521 of EC in the mid-infrared domain (Niyogi et al., 2006) and the presence of overlapping species
522 makes the quantification less accurate.

523 Further exploration of the regional and site by site quality of data is made via contiguous United
524 States (CONUS) maps of annual median reference method concentrations (left), difference in %
525 below MDL between FTIR and reference method (middle) and relative error (right) are shown in
526 Figure 5 for OC and EC. TC is very similar to OC and BC is very similar to EC.

527 Maps for all constituents with annual median FTIR and reference method concentrations, as well
528 as annual median prediction metrics are shown in Section S10.

529

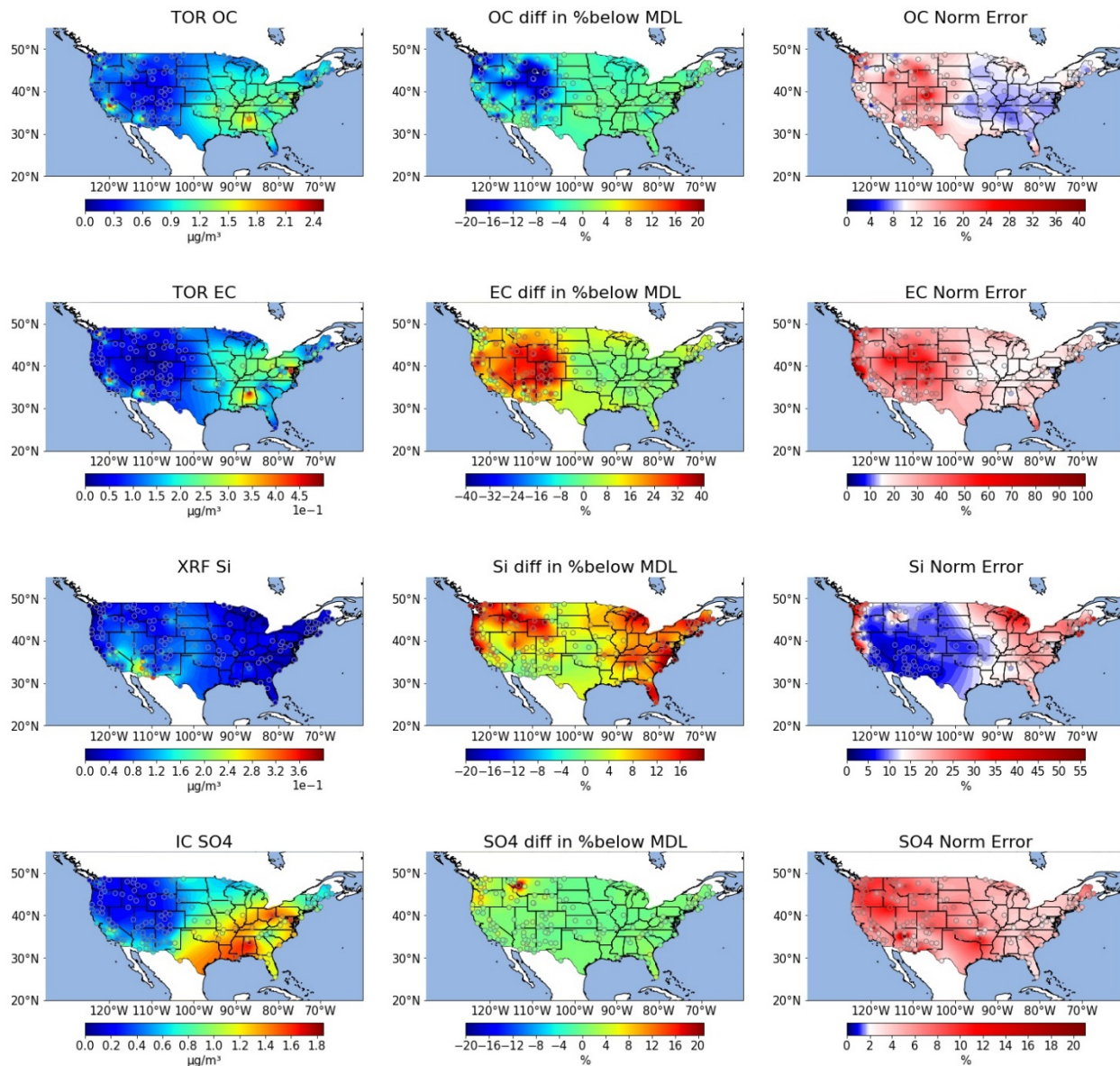



Figure 5. Annual median reference method concentrations (left), difference in % below MDL (middle) and normalized relative error (right) per site for OC, EC, silicon, and sulfate for CONUS for 2015. For the MDL plot, sites in green and blue indicate that the FTIR has the same of fewer samples below MDL than the reference method. Sites in yellow and red have more samples below MDL for FTIR than for the reference method. For the relative error maps, the median relative error of the reference method estimated using methods described in Table 1 is white. For sites in blue, FTIR has lower relative error than the reference method and sites in red are higher.

Annual median maps of FT-IR OC and TOR OC as well as maps of FT-IR EC and TOR EC (Supplemental material S10) are nearly identical. As shown in Figure 5, annual median OC and EC concentrations are highest at the four urban IMPROVE sites of Seattle, WA, Fresno, CA, Phoenix, AZ and Birmingham, AL than the rural sites and are higher in the east than in the west. For OC, the relative error is lower than the TOR relative error in the east (where concentrations are higher) and higher than TOR relative error in the west. OC has an equal or lower number of samples below MDL than TOR at all sites. For EC, FTIR relative error is higher than TOR relative error at almost all sites. The percentage of samples of EC that are below MDL for FTIR is similar to are slightly higher than TOR in the eastern US where EC concentrations are higher and are significantly higher than TOR in the western US where concentrations are lower. These patterns indicate that FTIR does not add error to OC measurements when concentrations are above 0.75 µg/m^3 but does add some error at lower OC concentrations and for EC measurements.

3.3.2   Elemental oxide predictions

Taking advantage of known mineral absorbance bands in the mid-infrared range (Hahn et al., 2018; Madejová and Komadel, 2001; Senthil Kumar and Rajkumar, 2013) (Section S9), Multilevel calibrations for soil elements were evaluated for the five crustal elements commonly used to estimate soil: silicon, aluminum, calcium, titanium, and iron (Table 1 and Fig. 4 for Si and Ti). All models present a satisfactory agreement between XRF and FT-IR predicted concentrations (R^2 > 0.94). The quality of prediction of the elemental oxides falls into two groups. The first group includes silicon, aluminum and calcium and is characterized by moderate relative errors (11 – 13 %), similar in magnitude to the FT-IR OC model (12 %) and have similar errors to XRF measurements. This indicates that like OC and TC, on-average FT-IR does not add additional uncertainty. The second group includes titanium and iron which have larger relative errors (20.9 – 24.8 %), analogous to HIPS BC and EC models (23.3 – 30 %). Comparing residuals to collocated XRF measurements (Section S9) shows that the FT-IR based models have a larger interquartile range. For Fe, XRF uncertainty is quite low and FT-IR adds additional uncertainty to the measurement. XRF Ti measurements have higher error than the other elements but there is an incremental increase in error due to FT-IR. In addition, cross plots of titanium concentrations show a bifurcation (Fig. 4). While most samples fall near their expected titanium concentration, samples collected at the Sycamore Canyon (AZ) site present a systematic negative bias, consistent across years, tentatively attributed to a site-specific soil composition not accounted for by the Limited calibration. Takahama et al. (2019) demonstrated several methods to identify the possible occurrence of anomalous predictions in OC and EC based on comparison of new sample spectra to calibration spectra based on projected PLS scores and regression residual vectors.

575 These samples with systematic negative bias in titanium predictions can presumably be identified
576 using such an approach, provided that compositional differences are detected in the IR spectrum.
577 Although distinct IR fingerprints exist, FT-IR calibrations for quantifying mineral contents should
578 be interpreted with care as specific elements may be indirectly quantified through their
579 correlation with another element even in the absence of clear IR signature (Hahn et al., 2018).
580 For instance, the variable importance in projection (VIP) scores for the Si, Al, and Ti calibrations
581 suggests use of similar spectral variables, with small differences, for prediction of these species
582 (Section S9). However, the 21 GMM sites coverage still meet the necessary requirements for
583 providing a reliable insight into soil composition in the IMPROVE network.

584 Figure 5 shows the distribution of concentrations of XRF Si across CONUS. The highest annual
585 median concentrations are in the southwest. Similar patterns are found for Al, Ca, Ti and Fe
586 except that high Fe concentrations are also observed at the urban sites, particularly Fresno, CA
587 and Birmingham, AL (Figures S10-6 through S10-9). For Si, FTIR normalized error is lower than
588 XRF in the west where concentrations are higher. For Ca, Ti and Al, FT-IR normalized error is
589 lower only in the southwest. For Fe, FT-IR is above XRF normalized error. The percentage of
590 samples below MDL are similar to XRF (0-10% different) in the southwest and central US and
591 modestly higher (15-20%) in the northwest and eastern US for Si. For Fe, the spatial pattern is
592 similar but the FTIR % below MDL is up to 50 % higher than XRF. However, for Al, Ca, and Ti, FTIR
593 percent below MDL is approximately the same or lower than XRF at all sites.

3.3.3   Inorganic ions

595 The two most abundant inorganic anions quantified in the network: nitrate and sulfate can also
596 be measured by FT-IR (absorption bands used for prediction are discussed in Section S9). FT-IR
597 sulfate concentrations display a satisfactory agreement with the reference IR measurements
598 (Fig. 4). Model performance metrics include R^2 above 0.98 and relative error of 6 %. The relative
599 error is the same as for FT-IR $PM_{2.5}$ and lower than OC, TC, and Si (Table 1). However, IC
600 measurements have even lower error than FT-IR sulfate. Compared to sulfate, FT-IR nitrate
601 concentrations (Section S9) are characterized by a moderate drop in the overall model
602 performance (R^2 = 0.927) while relative bias and error exceed 15 % and 50 %, respectively and
603 the error far exceeds reference IC nitrate measurement error. A direct comparison against
604 differential nitrate concentrations at collocated sites highlights the broad uncertainty in
605 determining nitrate content from PTFE filters (Section S9, Figure S9-3). Unlike nylon filters for
606 which nitrate is trapped on the surface, nitrate is known to evaporate from PTFE filters and the
607 extent of volatilization is dependent on temperature and relative humidity during and after
608 sampling. This causes a discrepancy between the mass of nitrate deposited onto the nylon filter
609 and the mass of nitrate on the PTFE filter (Eldred and Ashbaugh, 2004), making therefore FT-IR
610 calibrations with the nitrate measurements by IC from nylon filters as the reference method
611 error proneshould be used with caution. Although there are physical limitations to measuring

612 ambient nitrate on PTFE filters, a measure of nitrate on PTFE filters which corresponds to its
613 contribution to the gravimetric mass is useful for mass closure and data validation. FT-IR has
614 been shown to be useful for measuring and evaluating nitrate under controlled laboratory
615 conditions (ex. Wu et al., 2007). For network samples, ~~FT-IR based nitrate concentrations,~~
616 ~~measured in this way, should be considered with caution. A possible alternative is to develop a~~
617 ~~set of~~ nitrate could be measured using laboratory calibration standards and this effort will be
618 addressed in future work ~~of ammonium nitrate for FT-IR calibration. The nitrate mass on the~~
619 ~~PTFE would be useful for mass closure exercises on the PTFE filter but would not adequately~~
620 ~~assess particulate nitrate in the atmosphere~~.

621 The annual median sulfate concentration by IC is shown in Figure 5. Annual median
622 concentrations are highest in the southeast and eastern US with a gradient in concentrations
623 observed across the midwest. The median relative error for sulfate by IC is only 2% and all FTIR
624 sulfate has higher relative error. There are however, spatial differences. In the eastern US
625 sulfate relative error is less than 15% but in the west, it is considerably higher, peaking in
626 Wyoming where concentrations are very low.   The % below MDL is very similar for FTIR and IC
627 across the continent. Due to volatility of nitrate, the nitrate metrics for FTIR are poor compared
628 to sulfate (Figure S10-11).

3.3.4   PM$_{2.5}$ mass predictions

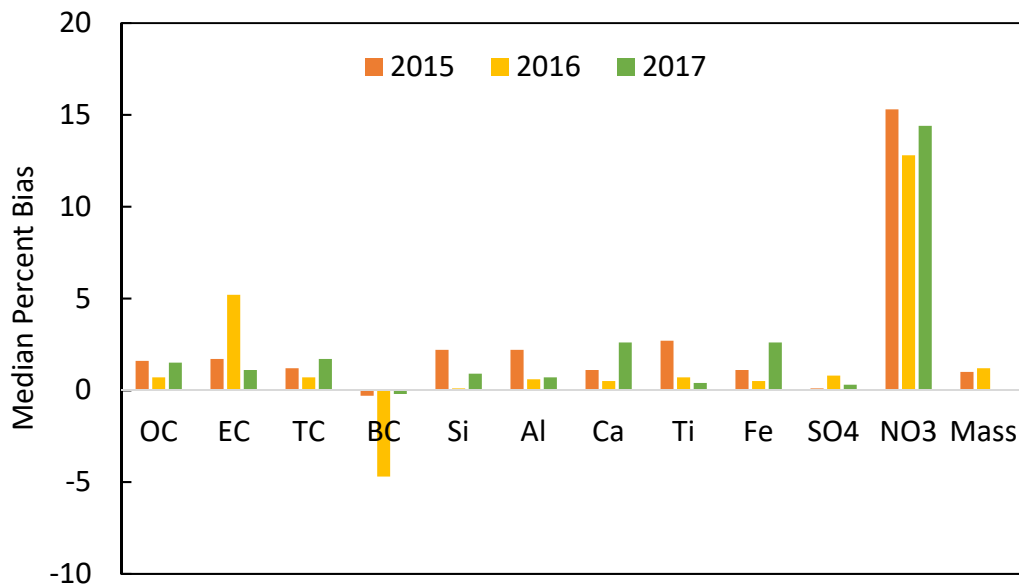
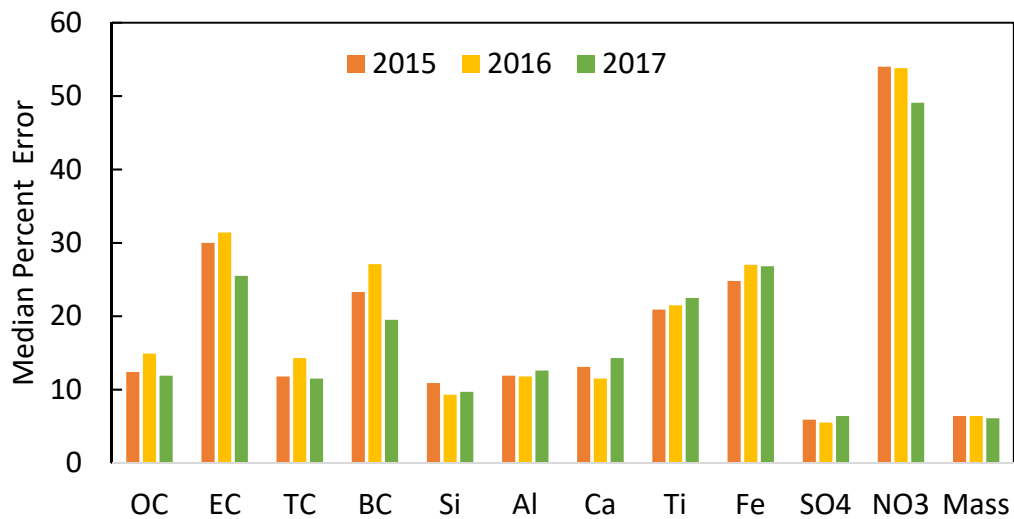
630 Since the major aerosol species are shown to be reasonably well measured by FT-IR, it was
631 anticipated that PM$_{2.5}$ mass calibration would perform well. The PM$_{2.5}$ model presents reliable
632 filter mass predictions (R^2 = 0.985) characterized by relative bias and error that are 1/3 to 1/2 of
633 those for OC and on par with gravimetric error (Table 1). The cross plot of gravimetric mass and
634 FT-IR predictions (Fig. 4) and maps of predictions metrics (Figure S10-12) show~~s~~ that PM$_{2.5}$ mass
635 can be accurately predicted across a broad concentration range indicating that FT-IR spectra of
636 PTFE filters do not contain interferents or other limitations that make PM mass predictions error
637 prone.

3.4   Long term stability

639 Finally, Multilevel calibrations are extended to 2016 and 2017 to evaluate the inter-year
640 consistency and determine if the assumptions behind Limited Sites and Biomass Burning models
641 remain valid over time. For each sampling year, new calibrations were developed following the
642 framework established for 2015. Models were recalibrated with data from the 21 sites and
643 biomass burning samples were detected by the functional group screening procedure. Fig. 6~~5~~
644 shows the median relative bias (top) and error (bottom) for the three years of data (cross plots
645 and prediction metrics shown for all predicted species for 2016 and 2017 in Section S11~~10~~).

646 These results indicate that the modeling methodology provided reasonably consistent results
647 across all three years.

648 Normalized bias for most species is below 3% and normalized error is consistent for all species
649 across all three years. The relative bias for EC and BC are similar to other species in 2015 and
650 2017 but in 2016 they are larger in magnitude than the other two years and different in sign.
651 2016 is the first year of TOR data from the multiwavelength TOR instruments (Chow et al., 2015)
652 so higher bias could be potentially be related to the new instruments. However, the HIPS
653 instrument was overhauled beginning in 2017 which provides no explanation for high bias in 2016
654 (http://vista.cira.colostate.edu/improve/Data/QA_QC/Advisory/da0041/da0041_HIPSmodificat
655 ions.pdf).   Further, the EC and BC calibrations are independent of each other except for using
656 the same filters for calibrations (as all species do) so the fact that the median bias is roughly equal
657 but opposite in sign is not due to codependence of the models.

658 In future work, calibrations models will be updated more frequently than annually with the most
659 recent year of ambient samples which may smooth biases and errors due to changes to
660 atmospheric condition and instrument drift.

661

662

663   Figure 65. (top) Median Percent Bias and (bottom) Median Percent Error for each constituent
664   measured for each year.

665

4   Conclusion

667   In this paper, we investigate the feasibility of an FT-IR method that uses ambient samples as
668   calibration standards for use by a large monitoring network. In this method, all sites in the

669   network collect PTFE filters for FT-IR analysis. A few select sites, used for calibration, would retain
670   all sampling and analyses of current IMPROVE sites to enable re-calibration of the FTIR method
671   on a routine basis. Re-calibration is especially important as atmospheric changes occur and as
672   conditions in the network evolve over time, for example new reference instruments, new or
673   significantly modified FT-IR instruments, changes to sampling protocol or possibly change in filter
674   material. The validity of such a design was evaluated with all PTFE filters collecting $PM_{2.5}$ aerosols
675   at 161 IMPROVE sites in 2015 and then tested for all filters in 2016 and 2017.

676   A multi-level modeling algorithm was used whereby smoke impacted samples are identified and
677   predicted by one model and the rest of the samples are predicted by another model developed
678   from 21 selected IMPROVE sites. The data from the two models are combined to evaluate
679   performance of the FT-IR method. The selection of sites was performed such that if one of the
680   21 sites ceases to operate, another site, selected from the same compositional cluster can be
681   used for calibration.

682   The cross-plots and prediction metrics indicate that the Multilevel model is equivalent to
683   conventional calibrations built from samples from every available site. Reliable performance in
684   predicted concentrations were reported for a broad range of atmospheric constituents with
685   detectable infrared signatures such as OC, EC, TC, sulfate, soil elements (Si, Al, Ca, Ti, Fe), light
686   absorption, and $PM_{2.5}$ mass. Due to volatilization off the PTFE filter, nitrate measurements were
687   found to be unsatisfactory. The calibration method was develop using data from 2015, and the
688   same methodology was applied to 2016 and 2017. The model performance metrics in all three
689   years were similar. Results across ~61,500 FT-IR spectra highlight the suitability of the Multilevel
690   calibration design to quantify multiple atmospheric $PM_{2.5}$ species in large monitoring networks.

691   This work presents an alternative, lower cost, filter analysis method to measure speciated aerosol
692   in a routine monitoring network. This could be a valuable addition to routine speciated aerosol
693   monitoring networks, such as IMPROVE, by incorporating monitoring sites that collect samples
694   on only a PTFE filter for subsequent analysis. This would provide the opportunity to have a subset
695   of less expensive monitoring site, which could be used for scoping studies to understand the
696   aerosol composition in unmonitored locations. It could also serve as a network cost savings
697   method by having a subset of network sites collect aerosol samples on only a Teflon filter.
698   However, the inability to measure particulate nitrate is a significant deficiency for using this
699   method to replace existing monitoring sites. The FTIR derived aerosol concentrations are also a
700   semi-independent measurement from the routine speciated aerosol measurements. Therefore,
701   routine FTIR measurements would provide valuable QA/QC information for any speciated
702   monitoring network. In addition, FTIR derived concentrations could be used to substitute for
703   missing concentrations in the case where the Teflon sample is valid, but filter samples or analyses
704   on the nylon or quartz fiber filters are not.

705 For IMPROVE's urban counterpart, the CSN network, after evaluation of the quality of predictions
706 in CSN, this framework could be used to accomplish goals similar to those of IMPROVE.
707 Additionally, this method could be used to predict samples collected in the Federal Reference
708 Method (FRM) network which is a PM mass only network. In addition~~Finally~~, this method, with
709 appropriate ambient standards, could be applied at other regional or international monitoring
710 networks or sites to provide low-cost comprehensive composition data. As shown in our previous
711 work, additional data, including an estimate of organic matter and it's functional group
712 composition, can also be obtained from FT-IR spectra of PTFE filters, further increasing the utility
713 of infrared spectroscopy of aerosol samples.

714

5 Data availability

Data is available at https://doi.org/10.25338/B8TP8V.

6 Author contribution

BD developed the software, performed the formal analysis and visualization for much of the manuscript and wrote the original draft of the manuscript, ATW developed software, performed formal analysis and visualization of the GMM work, ST participated in conceptualization, methodology software, visualization and reviewing and editing the manuscript. KMG developed parts of the biomass burning identification methodology. AAF performed the final visualizations for parts of the paper and was responsible for additional visualizations made to the paper based on reviewer comments. BS, SC and ASW, provided input throughout the project and reviewed and edited the manuscript, AMD conceptualized of and acquired funding for this project, developed methodology, performed project administration and supervision and reviewed, edited and finalized the manuscript.

7 Competing interests

The authors declare that they have no conflict of interest.

8 Acknowledgments

The authors acknowledge funding from the National Park Service in cooperation with the Environmental Protection Agency (P18AC01222). Thanks to Anahita Amiri Farahani for assisting with figures. We are particularly grateful to Kelsey Seibert for overseeing daily FT-IR operations at the University of California Davis and to the numerous undergraduate students who performed spectra collection from 2015 to 2017.

9 References

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.

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.

746 Boris, A. J., Takahama, S., Weakley, A. T., Debus, B. M., Fredrickson, C. D., Esparza-Sanchez, M., Burki, C.,
747 Reggente, M., Shaw, S. L., Edgerton, E. S., and Dillner, A. M.: Quantifying organic matter and functional
748 groups in particulate matter filter samples from the southeastern United States – Part 1: Methods,
749 Atmos. Meas. Tech., 12, 5391–5415, https://doi.org/10.5194/amt-12-5391-2019, 2019.

750 Boris, A. J., Takahama, S., Weakley, A. T., Debus, B. M., Shaw, S. L., Edgerton, E. S., Joo, T., Ng, N. L., and
751 Dillner, A. M.: Quantifying organic matter and functional groups in particulate matter filter samples from
752 the southeastern United States - Part 2: Spatiotemporal trends, 14, 4355–4374,
753 https://doi.org/10.5194/amt-14-4355-2021, 2021.

754 Bro, R. and Smilde, A. K.: Centering and scaling in component analysis, Journal of Chemometrics, 17, 16–
755 33, https://doi.org/10.1002/cem.773, 2003.

756 Burki, C., Reggente, M., Dillner, A. M., Hand, J. L., Shaw, S. L., and Takahama, S.: Analysis of functional
757 groups in atmospheric aerosols by infrared spectroscopy: method development for probabilistic
758 modeling of organic carbon and organic matter concentrations, 13, 1517–1538,
759 https://doi.org/10.5194/amt-13-1517-2020, 2020.

760 Busca, G. and Resini, C.: Vibrational Spectroscopy for the Analysis of Geological and Inorganic Materials,
761 in: Encyclopedia of Analytical Chemistry, John Wiley & Sons, Ltd, 2006.

762 Chow, J. C., Watson, J. G., Pritchett, L. C., Pierson, W. R., Frazier, C. A., and Purcell, R. G.: The dri
763 thermal/optical reflectance carbon analysis system: description, evaluation and applications in U.S. Air
764 quality studies, 27, 1185–1201, https://doi.org/10.1016/0960-1686(93)90245-T, 1993.

765 Chow, J. C., Watson, J. G., Chen, L. W. A., Chang, M. C. O., Robinson, N. F., Trimble, D., and Kohl, S.: The
766 IMPROVE_A Temperature Protocol for Thermal/Optical Carbon Analysis: Maintaining Consistency with a
767 Long-Term Database, 57, 1014–1023, https://doi.org/10.3155/1047-3289.57.9.1014, 2007a.

768 Chow, J. C., Watson, J. G., Chen, L. W. A., Chang, M. C. O., Robinson, N. F., Trimble, D., and Kohl, S.: The
769 IMPROVE_A Temperature Protocol for Thermal/Optical Carbon Analysis: Maintaining Consistency with a
770 Long-Term Database, 57, 1014–1023, https://doi.org/10.3155/1047-3289.57.9.1014, 2007b.

771 Chow, J. C., Wang, X., Sumlin, B. J., Gronstal, S. B., Chen, L.-W. A., Trimble, D. L., Watson, J. G., Kohl, S. D.,
772 Mayorga, S. R., Riggio, G., Hurbain, P. R., Johnson, M., and Zimmermann, R.: Optical Calibration and
773 Equivalence of a Multiwavelength Thermal/Optical Carbon Analyzer, 15, 1145–1159,
774 https://doi.org/10.4209/aaqr.2015.02.0106, 2015.

775 Chukanov, N. V. and Chervonnyi, A. D.: Infrared Spectroscopy of Minerals and Related Compounds,
776 Springer International Publishing, 2016.

777 Cios, K., Pedrycz, W., and Swiniarski, R. W.: Data Mining Methods for Knowledge Discovery, Kluwer
778 Academic Publishers, Norwell, MA, USA, 495 pp., 1998.

779 Corrigan, A. L., Russell, L. M., Takahama, S., Äijälä, M., Ehn, M., Junninen, H., Rinne, J., Petäjä, T.,
780 Kulmala, M., Vogel, A. L., Hoffmann, T., Ebben, C. J., Geiger, F. M., Chhabra, P., Seinfeld, J. H., Worsnop,
781 D. R., Song, W., Auld, J., and Williams, J.: Biogenic and biomass burning organic aerosol in a boreal forest
782 at Hyytiälä, Finland, during HUMPPA-COPEC 2010, 13, 12233–12256, https://doi.org/10.5194/acp-13-
783 12233-2013, 2013.

784 Debus, B., Takahama, S., Weakley, A. T., Seibert, K., and Dillner, A. M.: Long-Term Strategy for Assessing
785 Carbonaceous Particulate Matter Concentrations from Multiple Fourier Transform Infrared (FT-IR)
786 Instruments: Influence of Spectral Dissimilarities on Multivariate Calibration Performance, 73, 271–283,
787 https://doi.org/10.1177/0003702818804574, 2019.

788 Dillner, A. M. and Takahama, S.: Predicting ambient aerosol thermal-optical reflectance measurements
789 from infrared spectra: Elemental carbon, 8, 4013–4023, https://doi.org/10.5194/amt-8-4013-2015,
790 2015a.

791 Dillner, A. M. and Takahama, S.: Predicting ambient aerosol thermal-optical reflectance (TOR)
792 measurements from infrared spectra: Organic carbon, 8, 1097–1109, https://doi.org/10.5194/amt-8-
793 1097-2015, 2015b.

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

795 Eldred, Robert. and Ashbaugh, L. L.: Loss of particle nitrate from Teflon sampling filters: Effects on
796 measured gravimetric mass in California and in the IMPROVE network, 54, 93–104,
797 https://doi.org/10.1080/10473289.2004.10470878, 2004.

798 Friedel, R. A. and Carlson, G. L.: Infrared spectra of ground graphite, 75, 1149–1151,
799 https://doi.org/10.1021/j100678a021, 1971.

800 Hahn, A., Vogel, H., Andó, S., Garzanti, E., Kuhn, G., Lantzsch, H., Schüürman, J., Vogt, C., and Zabel, M.:
801 Using Fourier transform infrared spectroscopy to determine mineral phases in sediments, Sedimentary
802 Geology, 375, 27–35, https://doi.org/10.1016/j.sedgeo.2018.03.010, 2018.

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

805 Hawkins, L. N., Russell, L. M., Covert, D. S., Quinn, P. K., and Bates, T. S.: Carboxylic acids, sulfates, and
806 organosulfates in processed continental organic aerosol over the southeast Pacific Ocean during
807 VOCALS-REx 2008, 115, https://doi.org/10.1029/2009jd013276, 2010.

808 Hyslop, N. P., Trzepla, K., and White, W. H.: Reanalysis of Archived IMPROVE PM2.5 Samples Previously
809 Analyzed over a 15-Year Period, Environ. Sci. Technol., 46, 10106, 2012.

810   Hyslop, N. P., Trzepla, K., and White, W. H.: Assessing the Suitability of Historical PM2.5 Element
811   Measurements for Trend Analysis, Environ. Sci. Technol., 49, 9247–9255,
812   https://doi.org/10.1021/acs.est.5b01572, 2015.

813   Kuzmiakova, A., Dillner, A. M., and Takahama, S.: An automated baseline correction protocol for infrared
814   spectra of atmospheric aerosols collected on polytetrafluoroethylene (Teflon) filters, 9, 2615–2631,
815   https://doi.org/10.5194/amt-9-2615-2016, 2016.

816   Li, H.-D., Xu, Q.-S., and Liang, Y.-Z.: libPLS: An integrated library for partial least squares regression and
817   linear discriminant analysis, 176, 34–43, https://doi.org/10.1016/j.chemolab.2018.03.003, 2018.

818   Liu, S., Ahlm, L., Day, D. A., Russell, L. M., Zhao, Y., Gentner, D. R., Weber, R. J., Goldstein, A. H., Jaoui,
819   M., Offenberg, J. H., Kleindienst, T. E., Rubitschun, C., Surratt, J. D., Sheesley, R. J., and Scheller, S.:
820   Secondary organic aerosol formation from fossil fuel sources contribute majority of summertime organic
821   mass at Bakersfield, 117, https://doi.org/10.1029/2012JD018170, 2012.

822   Madejová, J. and Komadel, P.: Baseline studies of the clay minerals society source clays: Infrared
823   methods, Clays and Clay Minerals, 49, 410–432, https://doi.org/10.1346/CCMN.2001.0490508, 2001.

824   Mahalanobis, P., C,: On the generalized distance in statistics, 2, 49–55, 1936.

825   Malm, W. C., Sisler, J. F., Huffman, D., Eldred, R. A., and Cahill, T. A.: Spatial and seasonal trends in
826   particle concentration and optical extinction in the United States, 99, 1347–1370,
827   https://doi.org/10.1029/93jd02916, 1994.

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

832   Mayo, D. W., Miller, F. A., and Hannah, R. W.: Course Notes on the Interpretation of Infrared and Raman
833   Spectra, John Wiley & Sons, Hoboken, NJ, 2004.

834   Nemanich, R. J., Lucovsky, G., and Solin, S. A.: Infrared active optical vibrations of graphite, Solid State
835   Communications, 23, 117–120, https://doi.org/10.1016/0038-1098(77)90663-9, 1977.

836   Ngo, M. A., Pinkerton, K. E., Freeland, S., Geller, M., Ham, W., Cliff, S., Hopkins, L. E., Kleeman, M. J.,
837   Kodavanti, U. P., Meharg, E., Plummer, L., Recendez, J. J., Schenker, M. B., Sioutas, C., Smiley-Jewell, S.,
838   Haas, C., Gutstein, J., and Wexler, A. S.: Airborne particles in the San Joaquin Valley may affect human
839   health, 64, 12–16, https://doi.org/10.3733/ca.v064n01p12, 2010.

840   Niyogi, S., Bekyarova, E., Itkis, M. E., McWilliams, J. L., Hamon, M. A., and Haddon, R. C.: Solution
841   Properties of Graphite and Graphene, 128, 7720–7721, https://doi.org/10.1021/ja060680r, 2006.

842 O'Dell, K., Ford, B., Fischer, E. V., and Pierce, J. R.: Contribution of wildland-fire smoke to US PM2.5 and
843 its influence on recent trends, 53, 1797–1804, https://doi.org/10.1021/acs.est.8b05430, 2019.

844 Parks, D. A., Griffiths, P. R., Weakley, A. T., and Miller, A. L.: Quantifying elemental and organic carbon in
845 diesel particulate matter by mid-infrared spectrometry, null, 55, 1014–1027,
846 https://doi.org/10.1080/02786826.2021.1917764, 2021.

847 Reggente, M., Dillner, A. M., and Takahama, S.: Predicting ambient aerosol thermal-optical reflectance
848 (TOR) measurements from infrared spectra: Extending the predictions to different years and different
849 sites, 9, 441–454, https://doi.org/10.5194/amt-9-441-2016, 2016.

850 Reggente, M., Dillner, A. M., and Takahama, S.: Analysis of functional groups in atmospheric aerosols by
851 infrared spectroscopy: systematic intercomparison of calibration methods for US measurement network
852 samples, Atmos. Meas. Tech., 12, 2287–2312, https://doi.org/10.5194/amt-12-2287-2019, 2019.

853 Russell, L. M., Bahadur, R., and Ziemann, P. J.: Identifying organic aerosol sources by comparing
854 functional group composition in chamber and atmospheric particles, 108, 3516–3521,
855 https://doi.org/10.1073/pnas.1006461108, 2011.

856 Ruthenburg, T. C., Perlin, P. C., Liu, V., McDade, C. E., and Dillner, A. M.: Determination of organic matter
857 and organic matter to organic carbon ratios by infrared spectroscopy with application to selected sites in
858 the IMPROVE network, 86, 47–57, https://doi.org/10.1016/j.atmosenv.2013.12.034, 2014.

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

861 Schichtel, B. A., Malm, W. C., Bench, G., Fallon, S., McDade, C. E., Chow, J. C., and Watson, J. G.: Fossil
862 and contemporary fine particulate carbon fractions at 12 rural and urban sites in the United States, 113,
863 https://doi.org/10.1029/2007jd008605, 2008.

864 Senthil Kumar, R. and Rajkumar, P.: Characterization of minerals in air dust particles in the state of
865 Tamilnadu, India through ftir spectroscopy, 2013, 22221–22248, https://doi.org/10.5194/acpd-13-
866 22221-2013, 2013.

867 Snijders, T. A. B. and Bosker, R. J.: Multilevel Analysis: An Introduction to Basic and Advanced Multilevel
868 Modeling, SAGE, 370 pp., 2011.

869 Sorooshian, A., Wonaschütz, A., Jarjour, E. G., Hashimoto, B. I., Schichtel, B. A., and Betterton, E. A.: An
870 aerosol climatology for a rapidly growing arid region (southern Arizona): Major aerosol species and
871 remotely sensed aerosol properties, 116, https://doi.org/10.1029/2011jd016197, 2011.

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

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

880 Tuinstra, F. and Koenig, J. L.: Raman Spectrum of Graphite, 53, 1126–1130,
881 https://doi.org/10.1063/1.1674108, 1970.

882 Watson, J. G. and Chow, J. C.: A wintertime PM2.5 episode at the Fresno, CA, supersite, Atmospheric
883 Environment, 36, 465–475, https://doi.org/10.1016/S1352-2310(01)00309-0, 2002.

884 Weakley, A. T., Takahama, S., and Dillner, A. M.: Ambient aerosol composition by infrared spectroscopy
885 and partial least-squares in the chemical speciation network: Organic carbon with functional group
886 identification, 50, 1096–1114, https://doi.org/10.1080/02786826.2016.1217389, 2016.

887 Weakley, A. T., Takahama, S., Wexler, A. S., and Dillner, A. M.: Ambient aerosol composition by infrared
888 spectroscopy and partial least squares in the chemical speciation network: Multilevel modeling for
889 elemental carbon, 52, 642–654, https://doi.org/10.1080/02786826.2018.1439571, 2018a.

890 Weakley, A. T., Takahama, S., and Dillner, A. M.: Thermal/optical reflectance equivalent organic and
891 elemental carbon determined from federal reference and equivalent method fine particulate matter
892 samples using Fourier transform infrared spectrometry, 52, 1048–1058,
893 https://doi.org/10.1080/02786826.2018.1504161, 2018b.

894 White, W. H., Trzepla, K., Hyslop, N. P., and Schichtel, B. A.: A critical review of filter transmittance
895 measurements for aerosol light absorption, and de novo calibration for a decade of monitoring on PTFE
896 membranes, 50, 984–1002, https://doi.org/10.1080/02786826.2016.1211615, 2016.

897 Wold, S., Sjöström, M., and Eriksson, L.: PLS-regression: A basic tool of chemometrics, 58, 109–130,
898 https://doi.org/10.1016/S0169-7439(01)00155-1, 2001.

899 Yazdani, A., Dudani, N., Takahama, S., Bertrand, A., Prevot, A. S. H., El Haddad, I., and Dillner, A. M.:
900 Characterization of primary and aged wood burning and coal combustion organic aerosols in an
901 environmental chamber and its implications for atmospheric aerosols, 21, 10273–10293,
902 https://doi.org/10.5194/acp-21-10273-2021, 2021a.

903 Yazdani, A., Dillner, A. M., and Takahama, S.: Estimating mean molecular weight, carbon number, and
904 OM/OC with mid-infrared spectroscopy in organic particulate matter samples from a monitoring
905 network, 14, 4805–4827, https://doi.org/10.5194/amt-14-4805-2021, 2021b.

906 Zhang, X. L., Trzepla, K., White, W., Raffuse, S., and Hyslop, N. P.: Intercomparison of thermal-optical
907 carbon measurements by Sunset and Desert Research Institute (DRI) analyzers using the IMPROVE_A
908 protocol, 14, 3217–3231, https://doi.org/10.5194/amt-14-3217-2021, 2021.

909