Aerosol Science and Technology, 47:310–325, 2013
Copyright © American Association for Aerosol Research
ISSN: 0278-6826 print / 1521-7388 online
DOI: 10.1080/02786826.2012.752065

Quantification of Carboxylic and Carbonyl Functional Groups in Organic Aerosol Infrared Absorbance Spectra

Satoshi Takahama,[1,2] Anita Johnson,[1,3] and Lynn M. Russell[1]
[1]*Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA*
[2]*Atmospheric Particle Research Laboratory, APRL/IEE/ENAC, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*
[3]*Department of Civil, Architectural, and Environmental Engineering, Drexel University, Philadelphia, Pennsylvania, USA*

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

Atmospheric aerosols are one of the least understood components of the climate system and incur adverse health effects on susceptible populations. Organic aerosols can make up as much as 80% of atmospheric aerosols (Lim and Turpin 2002), and so its quantification and characterization plays an important role in reducing our uncertainty with regards to aerosol impacts on health and climate. As the number of organic molecules in the atmosphere are diverse in number (Hamilton et al. 2004), we advance a functional group representation of organic molecules as measured by Fourier transform infrared spectroscopy (FTIR) to characterize the chemical composition of particle samples. This study describes and evaluates the algorithm introduced by Russell et al. (2009) for apportionment and quantification of oxygenated (carbonyl and hydroxyl) functional groups from infrared absorption spectra. Molar absorptivities for carbonyl and hydroxyl bonds in carboxylic groups are obtained for several dicarboxylic compounds, and applied to a multifunctional compound and mixture to demonstrate the applicability of this method for more complex samples. Furthermore, functional group abundances of two aldehydic compounds, 2-deoxy-d-ribose and glyceraldehyde, atomized from aqueous solution are in quantitative agreement with number of bonds predicted after transformation of these compounds into diols. The procedure for spectra interpretation and quantitative analysis is described through the context of an algorithm in which contributions of background and analyte absorption to the infrared spectrum are apportioned by the superposition of lineshapes constrained by laboratory measurements.

Received 6 June 2012; accepted 4 November 2012.
The authors would like to acknowledge W. R. Leaitch, T. S. Bates, P. K. Quinn, and D. Baumgardner, and the science and field teams of GoMACCS, MILAGRO, VOCALS-REx, ICEALOT, Barrow, Scripps Pier, and Whistler campaigns for assistance with sample collection and operation; L. Ahlm, R. Bahadur, A. Corrigan, D. Day, A. Frossard, S. Gilardoni, L. Hawkins, S. Liu, R. Modini, R. Schwartz, and P. Shaw for feedback on the algorithm; S. Gilardoni, L. Hawkins, and S. Liu for helpful discussions during development; and article reviewers for helpful comments. The authors acknowledge NSF ATM-0904203 for funding.
Address correspondence to Lynn M. Russell, Scripps Institute of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0221, USA. E-mail: lmrussell@ucsd.edu

[Supplementary materials are available for this article. Go to the publisher's online edition of *Aerosol Science and Technology* to view the free supplementary files.]

1. INTRODUCTION

Atmospheric aerosols contain some mixture of nitrate, sulfate, ammonium, crustal elements, sea salt, black carbon, and organic molecules (Seinfeld and Pandis 2006). The number of molecules comprising the organic aerosol fraction is numerous (Hamilton et al. 2004) and remains undercharacterized with respect to their quantity and potential impacts on health and climate (Kanakidou et al. 2005). Measurement of organic aerosol composition and abundance in the atmosphere provide a basis from which we can infer sources, predict potential impacts on health and climate, and provide constraints on representations for mathematical and computational modeling of the atmospheric system for simulating alternative scenarios relevant to adaptation or policy decision-making.

Given the dimensionality and domain size of organic aerosol characteristics (comprising size, shape, quantity, and composition), its representation must be constructed through the combination of analytical techniques that provide complementary information. Each method varies in the amount of organic content (mass concentration) it can quantify, and to what extent individual species can be resolved (chemical specificity). Composition is typically quantified in units of organic carbon (OC) through evolved gas analysis (EGA) (Subramanian et al. 2004), mass fragments through mass spectrometry (Schauer et al. 1999; Allan et al. 2003; Williams et al. 2006; Goldstein et al. 2008), atomic composition also through aerosol mass spectrometry (time-of-flight aerosol mass spectrometry [TOF-AMS]; Chhabra et al. 2010), or functional bonds through spectroscopy (Fourier transform infrared spectroscopy [FTIR], Maria et al. 2003; nuclear magnetic resonance [NMR], Decesari et al. 2007; near-edge X-ray absorption fine structure [NEXAFS], Russell 2003; Raman spectroscopy, Chan et al. 2006).

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

Single-particle methods (aerosol time-of-flight mass spectrometry [ATOFMS], Noble and Prather 2000; scanning electron microscopy with energy dispersive X-ray [SEM-EDX], Laskin et al. 2006; transmission electron microscopy with electron energy-loss spectroscopy [TEM-EELS], Adachi and Buseck 2008; scanning transmission X-ray microscopy with NEXAFS [STXM-NEXAFS], Russell 2003) can provide metrics linked to number, mass composition, and morphology, but often not all together, and representativeness is dependent on sampling statistics. Some techniques target a particular size fraction (cluster chemical ionization mass spectrometry [cluster-CIMS], Zhao et al. 2010; thermal desorption chemical ionization mass spectrometry [TD-CIMS], Smith et al. 2004), or probe particular sections of individual particles (time-of-flight secondary ion mass spectrometry [TOF-SIMS], Tervahattu et al. 2002). Few methods measure organic aerosols online or *in-situ* (aerodyne aerosol mass spectrometer [AMS], Allan et al. 2003; aerosol chemical speciation monitor [ACSM], Ng et al. 2010; thermal desoprtion aerosol gas chromatography [TAG], Williams et al. 2006), while a large number of methods characterize particles offline. In the latter case, samples are collected on a substrate and subsequently analyzed by one or more methods. For these analytical techniques, some fraction of particle ensembles collected on a substrate through filtration or impaction is quantified with optional sample preparation a priori (e.g., extraction or derivitization). Each instrument serves to provide enlightenment on some aspect particle composition or morphology relevant for source apportionment or property prediction.

Infrared absorbance spectra of condensed-phase matter from FTIR provides a rich sequence of signals (Allen et al. 1994; Sax et al. 2005; Ofner et al. 2011) from which we can extract information regarding organic aerosol composition and mass. Each spectrum is interpreted as a combination of absorption lineshapes (characteristic absorption profiles) which can represent either molecular mixtures (Russell et al. 2009) or organic functional groups (Maria et al. 2003). Application of statistical learning methods which retain the full dimensionality of the measured signal has provided discrimination of constituent components or mixture classes (Takahama et al. 2011). Factor-analytic decomposition of ambient aerosol spectra have revealed component profiles which retain distinguishing spectral features that relate to original sources of organic aerosols (Russell et al. 2011). These feature vectors have also been used for unsupervised learning (cluster) analysis and regression analysis (in a chemical-balance-type approach) to group ambient aerosol spectra by source influence or similar mixture composition (Takahama et al. 2011) that agree with interpretations from the factor analysis methods.

A reduced representation of the FTIR spectrum in molar quantities of functional groups has been used to describe organic functional group abundance in atmospheric, indoor, and laboratory-generated organic aerosols (Allen et al. 1994; Russell 2003; Sax et al. 2005; Reff et al. 2007; Coury and Dillner 2008; Polidori et al. 2008). A functional group representation of the complex mixture of organic aerosols is physically meaningful to the extent that it has been linked to predictive models of molecular interactions in the condensed phase within an established thermodynamic framework (i.e., activity coefficients from group-contribution methods, Fredenslund et al. 1975), from which hygroscopic properties and volatilities can be estimated (Asher et al. 2002; Ming and Russell 2002; Topping et al. 2005; Chang and Pankow 2006; Erdakos et al. 2006; Pankow and Chang 2008; Quinn et al. 2006; Zuend et al. 2008, 2011). Relative functional group mass or molar compositions have also been linked with the identification of samples heavily influenced by marine, combustion, biogenic, or biomass burning sources (Liu et al. 2009; Hawkins and Russell 2010; Takahama et al. 2011).

Previous studies have detailed the quantification of CH groups from saturated, unsaturated, and ring structures (Maria et al. 2003; Gilardoni et al. 2007; Russell et al. 2009); amine CNH_2 (Liu et al. 2009; Russell et al. 2009); alcohol and phenol COH (Bahadur et al. 2010 and Russell et al. 2010, respectively); organosulfate $COSO_3$ by solvent-rinsing (Hawkins et al. 2010); and organonitrate $CONO_2$ groups (Day et al. 2010). This work describes and evaluates the carboxylic COH, alcohol COH, and non-acid carbonyl C=O separation and quantification initially introduced by Russell et al. (2009). As the the quantity and type of oxygenated functional groups is linked with atmospheric aging (Rudich et al. 2007; Jimenez et al. 2009), polarity and hygroscopicity (Pankow and Chang 2008; Jimenez et al. 2009; Pankow and Barsanti 2009; Zuend et al. 2010; Donahue et al. 2011), solution-phase miscibility (Song et al. 2012; Zuend and Seinfeld 2012), and modifications to vapor–pressures relative to unoxygenated compounds (Chattopadhyay and Ziemann 2005), reducing the uncertainty in this fraction provides a valuable contribution to organic aerosol analysis. This manuscript describes an algorithmic method of quantification in which commonly employed geometric curves—parameterized from laboratory and reference spectra—are superposed to reconstruct measured spectra. This approach for resolving the abundance of organic molecular bonds leads to a consistent interpretation of FTIR spectra across different users, and provides an explicit framework through which additional information can be introduced to reduce analytical uncertainty.

2. METHODS

2.1. Algorithm Description

2.1.1. Background Correction

To quantify the amount of absorption due to analyte on our substrate, we first apply several background corrections. One correction is applied to remove the additional extinction contributed from scattering and absorption by the Teflon (PTFE) filter on which samples are collected and analyzed. We first subtract the spectrum obtained before ambient air sampling (presample) from that obtained after ambient air sampling (postsample). For this, we adopt a procedure described by Maria

TABLE 1
Components of spectrum absorbance represented as scaled shapes, lines, and Gaussian peaks as a function of wavenumber $\tilde{\nu}$

Representation	Expression		Parameters
Fixed shape	$\beta a(\tilde{\nu})$	(T1-1)	β: Scaling factor
Line/polynomial	$\sum_{i=0}^{p} \beta_i \tilde{\nu}^i$	(T1-2)	β: Polynomial coefficients p: Degree of polynomial
Gaussian lineshape	$\beta \exp[-\frac{(\tilde{\nu}-\mu)^2}{2\sigma^2}]$	(T1-3)	β: Amplitude μ: Peak location σ: Width parameter

et al. (2003), whereby we scale each presample spectrum to the postsample spectrum in the neighborhood of 1200 cm^{-1} prior to subtraction (Equation (T1-1) in Table 1), to account for the possible stretching of the Teflon due to the sampling procedure, which would (linearly) affect its optical thickness. The remaining variations which are not due to analyte absorption are accounted for with polynomial or linear baselines, depending on the region. The simple baselining methods used here are unable to account for the full curvature on the absorbance imposed by the substrate; presample spectrum subtraction is therefore applied. The choice of baselining the presample subtracted spectrum as opposed to taking the difference of baselined presample and postsample spectra follows from the reasoning that an error is introduced each time the baseline is approximated; we baseline once after presample spectrum subtraction to minimize the accumulation of this error. In practice, however, the order of operations does not affect the baselined spectra significantly (except in samples close to the limit of detection).

For the region between 4000 and 2000 cm^{-1}, we find the parameters for a polynomial (Equation (T1-2)) for wavelengths, $\tilde{\nu}$, regressed to a set of wavelengths in the background region, $\mathcal{W}_m$. $\mathcal{W}_m$ is parameterized by two baselining points (b_l, b_u) and region widths on either side of absorption region (Δ_l, Δ_u):

$$\mathcal{W}_m = \{\nu : b_l - \Delta_l \leq \nu < b_u + \Delta_u\} \ . \qquad [1]$$

A 3rd-order polynomial ($p = 3$) is empirically determined to baseline blank sample spectra better than a linear model ($p = 1$) (Figure 1). As a blank sample is not expected to contain any analytes, a background correction should in principle "zero-out" the spectrum; the remaining contributions to the absorbance after baselining provides an estimate in the error of the method. Specifically, the deviation in the corrected baseline about the mean value for a large number of samples is the notable metric, as a nonzero but consistent bias can simply be removed by subtraction. From this perspective, the $p = 3$ polynomial reduces the deviation in a baselined spectrum by an average of 70% across relevant wavenumbers over the linear case. Furthermore, additional subtraction of the mean absorbance of blank

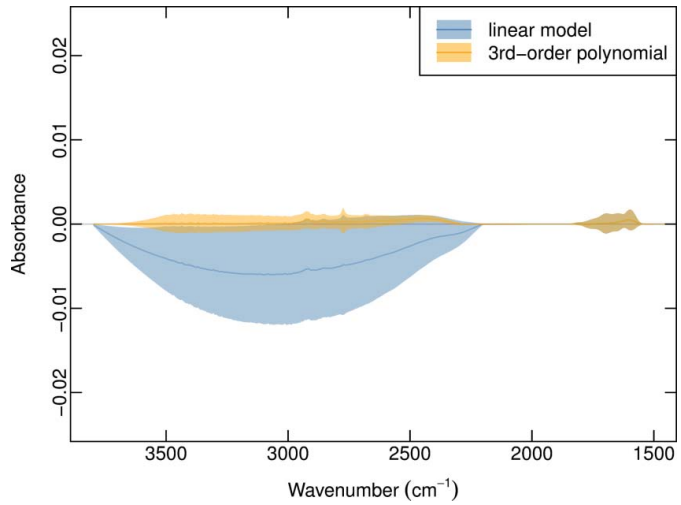

FIG. 1. Comparison of linear and polynomial baselines estimated for blank sample spectra from the CalMex campaign (185 filters used for this analysis; Takahama et al. 2012). Lines indicate mean values for each wavenumber; shaded areas encompass ±1 standard deviation about the mean. (Color figure available online.)

samples after baselining is not necessary, as it is statistically indistinguishable from the zero value.

For the region between 2000 and 1400 cm^{-1}, we model the baseline with a line defined by endpoints corresponding to local minima around the absorbing regions (as commonly employed by spectroscopists). Let $s(\tilde{\nu})$ = detrended spectrum. We model the baseline as a line (Equation (T1-2)) defined by two baselining points, $\tilde{\nu}_u$ and $\tilde{\nu}_l$:

$$\tilde{\nu}_i = \arg\min_{\tilde{\nu}} \{s(\tilde{\nu}) : b_{1,i} \leq \tilde{\nu} < b_{2,i}\} \ \forall i \in \{l, u\} \ . \qquad [2]$$

The default parameters for fitting in both regions are shown in Table 2. Particularly for the region between 4000 and 2000 cm^{-1}, these values were chosen from observations of the detrended spectra (Figure 2) but can be modified according to requirements of individual samples (for instance, higher loadings may require applying the baseline over a wider region). Detrending by differencing is a method commonly used in time series analysis (Box et al. 2008) and spectroscopic analysis (Griffiths and Haseth 2007); a gradual variation is removed from sequential observations (over successive wavenumbers in this case). The positions of onset of perturbations in the signal due to analyte absorbance can then be located in the differenced spectrum.

TABLE 2
Baselining parameters defined in Equations (1) and (2); examples of b_x and Δ_x for $x = \{u, l\}$ are illustrated in Figure 2

Parameter	$x = (u)pper$	$(l)ower$
b_x	3625	2300
Δ_x	200	200
$b_{1,x}$ (lower)	1805	1500
$b_{2,x}$ (upper)	1845	1610

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

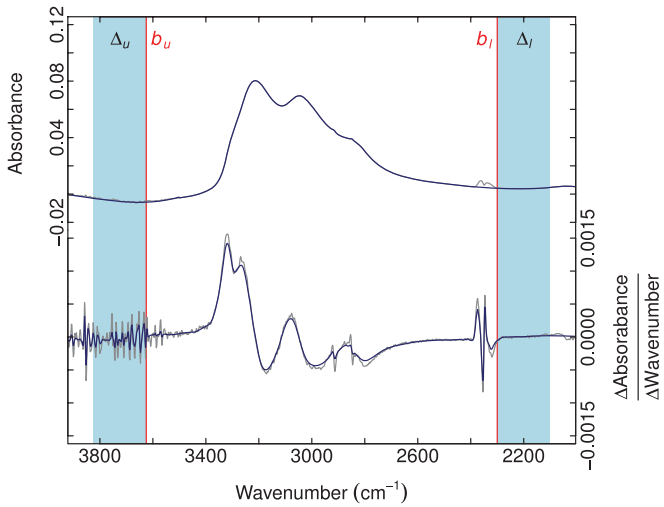

FIG. 2. Example of detrended spectrum (Section 2.1.1) used to determine background regions (vertical lines and shaded rectangles); parameter names correspond to those used in Equation (1) and Table 2 for the baseline regression model. (Color figure available online.)

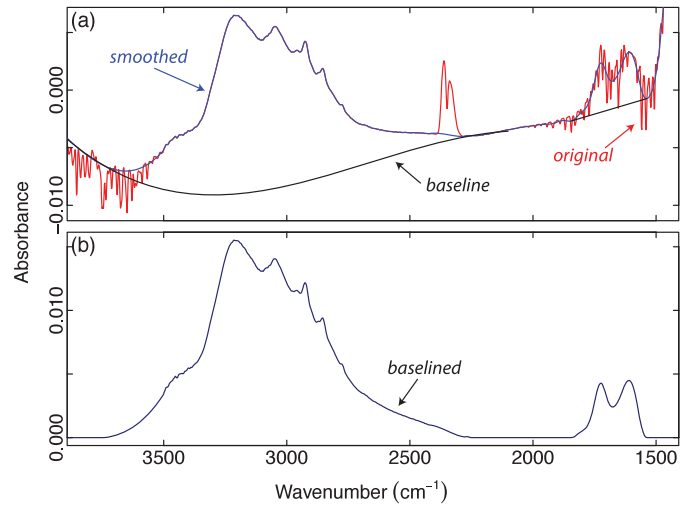

FIG. 3. Example of original and smoothed spectra (a) and baselined spectra (b). Negative absorbance (a) can arise from variations in Teflon thicknesses before and after sampling, as deformations can occur through stretching and imprinting of sample grid located behind the filter. (Color figure available online.)

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

The background regions used for regression (Equation (1) are defined by either side of these wavenumbers (Figure 2).

In a few cases, we apply a non-negativity constraint on the baseline fitting (Section S1). A few additional techniques are introduced to aid in the fitting or subsequent quantification, which includes the biased and wavelet smoothing over the region of water vapor interference above 3600 cm^{-1} and between 2000 and 1500 cm^{-1}, respectively, and spline interpolation to remove the CO_2 absorption at 2300 cm^{-1} (Section S1). An example of a modeled baseline and the baselined spectrum is shown in Figure 3.

2.1.2. *Band-Fitting*

We apportion the absorption by scaling fixed absorption profiles to the spectrum, or fitting parameterized distributions (in this case, Gaussian) to particular regions of the spectrum. Performing a least-squares minimization for all fitting parameters (including scaling factors and peak amplitudes, locations, and widths) simultaneously poses a challenge as the dimensionality and search space becomes large. As some absorbance profiles can be fitted independently of another (as determined by fitting region, Table 3) or scaled to uniquely absorbing regions, we instead segment the parameter space and apply a sequential fitting process. The order of fitting is chosen such that the components for which fitting parameters can be determined least ambiguously is selected first, or fitted together when this order is not resolvable (and the parameter space to be searched is manageable). Based on this heuristic, the order of fitting and subtraction chosen is as follows (the chosen representation of lineshape is given in parentheses): carboxylic COH (Equation (T1-1)), ammonium NH (Equation (T1-1)), alcohol COH (Equation (T1-3)), alkane CH (Equation (T1-3)), and carboxylic C=O (Equation (T1-3)) and amine CNH_2 (Equation (T1-3)) together (further details provided in Section S2). As the absorbance regions of the latter two are separate from the rest, they can be fitted independently of the other groups. While the presence of additional functional groups may affect the fitting of any single functional group (Section S5), the estimation errors of those fitted later in this sequence do not affect estimates of those fitted prior. After peaks for these groups are fitted, a series of additional peaks are fitted for quantifying the alkene and aromatic functional groups (Section S2). Absorbances by organonitrate $CONO_2$ and organosulfate $COSO_3$ used for quantification (at <1000 cm^{-1}) are mutually exclusive to these regions, and are described by Day et al. (2010) and Hawkins and Russell (2010), respectively.

For carboxylic COH and ammonium NH stretching vibrations, we scale and subtract a fixed absorbance profile (Equation (T1-1)). β_k is defined as the scaling factor for component k. Let $s(\tilde{\nu})$ be either the spectrum/residual after baseline correction, detrending, and/or other fitting, and $a_k(\tilde{\nu})$ the absorption profile for component k:

$$\beta_k = \min\{s(\tilde{\nu})/a_k(\tilde{\nu}) : \nu \in \mathcal{W}_k\}$$

where $\mathcal{W}_k$ is the set of wavenumbers which defines the scaling region for component k. The final contribution is reconstructed as $\beta_k a_k(\tilde{\nu})$. As few atmospherically-relevant compounds exhibit broad absorption in the region below 2600 cm^{-1}, we use this knowledge to subtract the carboxylic COH lineshape first. We then subtract ammonium by scaling in the region primarily around 3100 cm^{-1} as established by Maria et al. (2003) and Gilardoni et al. (2007). For this task, we require lineshapes. For ammonium, we use an absorption profile from a scan corresponding to a filter loaded in the laboratory with 6 μg of

TABLE 3
Initial values and constraints used for Gaussian peak-fitting. The Gaussian form and parameters are from Equation (T1-3)

		Initial values		Constraints*		
Functional group	Peak no.	μ (cm^{-1})	σ (cm^{-1})	μ (cm^{-1})	σ (cm^{-1})	Region (cm^{-1})
Alcohol	1	3500–3400	30–40	3500–3290	10–100	3670–3150
Alcohol	2	3350–3205	30–40	3450–3200	10–75	3670–3150
Unidentified	1	3139	19	3142–3136	9–19.5	–
Unidentified	2	3073	20	3076–3070	10–20	–
Aromatic	1	3050	3.5	–	–	–
Unidentified	3	3011	23.5	3014–3008	13.5–24	–
Alkene	1	2980	3.5	–	–	–
Unidentified	4	2955	12	2958–2952	2–12.5	–
Alkane	1	2925	10–25	2855–2849	10–42.5	2930–2790
Alkane	2	2882	10–25	2886–2876	10–37.5	2930–2790
Alkane	3	2852	10–25	2932–2921	10–25	2930–2790
Alkane	4	2800	10	2815–2790	10–22.5	2930–2790
Carbonyl	1	1717	10–20	1720–1714	7.5–30	1850–1500
Amine	1	1625	10–20	1630–1620	7.5–30	1850–1500

*Peaks for which constraints are not specified indicate that peak parameters are fixed to "initial" values.

$(NH_4)_2SO_4$. The exact loading does not have significance except that the absorption is not saturated and the signal-to-noise ratio is sufficiently above the limit of detection. The shape for carboxylic COH is derived from a novel method as described in Section 2.1.3.

After each of these fixed profiles are fitted, Gaussian peaks for alkane, alcohol COH, carbonyl, and amine are fitted to the residual remaining after subtraction. For this task, we search for the appropriate parameters, $(\beta_k, \mu_k, \sigma_k)$ for a Gaussian function (g_k; Equation (T1-3)) and union of absorbing regions of each peak ($\mathcal{W}_k$), box constraints (B_k), for set of peaks ($\mathcal{P}$):

$$\left\{\boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}\right\} = \left\{\underset{\beta_k,\mu_k,\sigma_k}{\arg\min}\left[s(\tilde{\nu}) - \sum_k g_k(\tilde{\nu})\right]^2 : \tilde{\nu} \in \bigcup_k \mathcal{W}_k \vee B_k(\beta_k, \mu_k, \sigma_k) \vee k \in \mathcal{P}\right\}.$$

As absorbing regions of alcohol COH and alkane CH are mutually exclusive to a first order (Table 3), they are fitted separately (with alcohol COH first). Carbonyl C=O and amine CNH_2 are fitted together.

The initial values and constraints for Gaussian peaks are derived from analysis of laboratory standards, and manually fitted peaks of Gilardoni et al. (2007) (Table 3). We use one of two convex, constrained minimization algorithms (nl2sol or L-BFGS-B implemented in the R statistical package; R Development Core Team 2012); a set of 35 initial values and constraints are used to seed the minimization, and the one with lowest deviance or residual is selected. In the case of the alkane CH group, we initially fit a single Gaussian peak and use its width to revise the fitting region for the four peaks.

2.1.3. *Separation of Oxygenated Fraction*

Carboxylic COH has a very broad absorption between the regions 3400 and 2400 cm^{-1}, and overlaps with the absorption region for alkane CH. Therefore, it is difficult to independently parameterize its contribution to the absorption in this region. For this purpose we apply a statistical separation technique, whereby combined absorptions of CH and COH over a spectrum of compounds are decomposed into relative contributions from CH and COH components using the Positive Matrix Factorization (PMF) algorithm (Paatero and Tapper 1994). The solutions generated by PMF are further modified to impose additional constraints of smoothness and quantitive apportionment for our aerosol sample spectra.

We selected a number of carboxylic acids with varying number contributions of carboxylic COOH and CH bonds, and obtained their spectra from the NIST Chemistry WebBook (http://webbook.nist.gov/chemistry/http://webbook.nist.gov/chemistry/) (Figure 4a). Using the WebBook archive permitted rapid consolidation of infrared spectra of compounds spanning a wide range of carboxylic COOH to alkane CH ratios; considerations for their variable preparation and data treatment are described in Section S3. We subject this set of spectra to a factor analytic decomposition (described further in Section S3), in which the set is represented by a linear combination of a smaller number of components. We specify a two-factor solution, as the variation among spectra is expected to arise from the variable contributions of alkane CH and carboxylic COH in the selected compounds. Among the factor analytic rotations (Paatero et al.

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

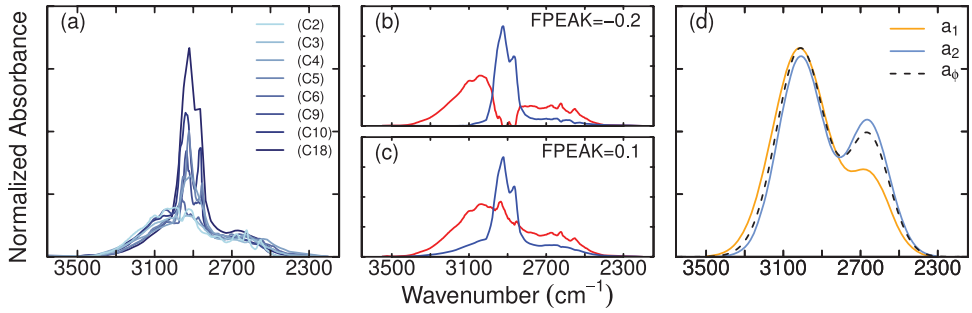

FIG. 4. (a) Normalized spectra from NIST Chemistry WebBook, (b) and (c) possible CH (blue/dark grey) and COH (red/light grey) components from PMF analysis, and (d) derived lineshapes for carboxylic COH and final lineshape of carboxylic COH from calibration (Equation (3)). (Color figure available online.)

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

2002) explored, we find two reoccurring solutions which are shown in Figures 4b and c. One component is identified as the absorption arising from CH bonds from comparisons of reference standards of aliphatic compounds with strong absorbances appearing between 3000 and 2800 cm^{-1} (blue/dark grey profiles in Figures 4b and c). The other profile (red/light grey lines in Figures 4b and c) is attributed to the COH absorption (the broad range is consistent with the expected absorption band of carboxylic COH; Lambert et al. 1998), and the illustration reflects ways in which the COH absorption approximately varies with wavenumber in these two cases. Our goal is to further constrain this wavenumber-dependence in a way that it is in agreement with our method of alkane CH quantification, so we find a combination of the two profiles that satisfies this criteria. We introduce a degree of freedom in combining these solutions for the carboxylic COH component, represented as a fractional contribution (ϕ) of each profile to the final lineshape $a(\tilde{\nu})$ (Equation (3)). Carboxylic acid reference spectra $s(\tilde{\nu})$ consisting of only carboxylic COH and CH bonds are represented by a superposition of $a(\tilde{\nu})$ scaled by a factor β, and a sum of Gaussian peaks representing the absorbance due to CH bonds (Equation (4)):

$$a_\phi(\tilde{\nu}) = \phi a_1(\tilde{\nu}) + (1-\phi)a_2(\tilde{\nu}) \quad [3]$$

$$s(\tilde{\nu}) = \beta a_\phi(\tilde{\nu}) + \sum_k g_k(\tilde{\nu}) \,. \quad [4]$$

The appropriate value of ϕ and the effective absorptivity of the resulting carboxylic COH lineshape is determined by fitting Equation [4] to adipic acid reference spectra for which carbonyl absorptivities were determined (Gilardoni et al. 2007). The criterion for ϕ is that the alkane CH absorptivity, α_{aCH}, obtained from fitting the Gaussian peaks to the residual of $s(\tilde{\nu}) - \beta a_\alpha(\tilde{\nu})$ as described in Section 2.1.2 approximately matches our known value. ϕ was varied between 0.5 and 1.0 which resulted in α_{aCH} between 1.1 and 0.5; a value of $\phi = 0.77$ was selected. This fitting procedure resulted in an estimated value for the carboxylic COH absorptivity of 65.0 $cm^{-1}\,\mu mol^{-1}$. The final shape is shown in Figure 4d. The final lineshape of carboxylic COH absorption derived here is qualitatively consistent with those illustrated by Holes et al. (1997) and Sax et al. (2005), but is fixed by calibration to quantitatively separate the respective fractions of carboxylic COH and alkane CH.

As implied in this successive fitting method, the two bonds comprising carboxylic COOH functional groups (hydroxyl COH and carbonyl C=O) are treated separately. The moles of carbonyl functional groups quantified by our Gaussian peak-fitting is referred to as "total" carbonyl, and we apportion this quantity to carboxylic acid groups and nonacid, or "excess," carbonyl (i.e., ketones and aldehydes), and in turn refine our estimate of the amount of carboxylic COOH functional groups in our sample. Let n_{cCOH} = moles of carboxylic COH and $n_{tC=O}$ = moles of total carbonyl. We map these quantities to carboxylic COOH and excess carbonyl C=O, $(n_{COOH}, n_{C=O})$, using the following approach:

$$(n_{COOH}, n_{C=O}) = \begin{cases} (n_{cCOH} + n_{tC=O}, 0) & \text{if } n_{tC=O} \approx n_{cCOH} \\ (n_{cCOH}, n_{tC=O} - n_{cCOH}) & \text{if } n_{tC=O} \gg n_{cCOH} \\ (n_{cCOH}, 0) & \text{otherwise .} \end{cases} \quad [5]$$

There are some observations for which $n_{tC=O} \ll n_{cCOH}$, denoted by the "otherwise" case, which is a physically implausible scenario (given our explicit assumptions in Table 3). It is likely that the quantity of carbonyl C=O is underestimated for these samples, as the stretching frequency of carbonyl can be shifted to lower frequencies in the presence of conjugated (C=C or C=O) bonds nearby (Blout et al. 1948; Coates 2006). Based on this understanding, the moles of carboxylic COH are used to estimate the moles of carboxylic COOH for these ($n_{tC=O} \ll n_{cCOH}$) cases.

2.2. Experimental Method

Several organic compounds were selected for either calibration or evaluation of the algorithm described in Section 2.1. Compounds chosen for calibration are those for which band-fitting and quantification follows from the least amount of uncertainty (i.e., least amount of interference). Some of the

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

chosen compounds have been reported to be present in the atmosphere (e.g., pinonic acid (Aiken et al. 2008); 2-deoxy-d-ribose and glyceraldehyde (Russell et al. 2010)). Others have been used in determining the shape of the carboxylic COH lineshape (as described in Section 2.1.3), but previously from the NIST Chemistry WebBook database; e.g., succinic and azelaic acids) or a previous measurement (adipic acid, by Gilardoni et al. 2007). Pinonic acid and a 50/50 molar % mixture of 2-deoxy-d-ribose and succinic acid were analyzed to evaluate the algorithm for characterization of multifunctional samples (with regards to co-occurrence of carboxylic and alcohol COH, especially). Collectively, these reference compounds delineate the axes of the composition domain defined by Russell et al. (2011).

Solutions of 1 mM concentrations were prepared by dissolving or mixing selected compounds (azelic acid: Aldrich, 246379-25G, 98%, solid; succinic acid: Fluka, 14078, 99.5%, solid; 2-deoxy-d-ribose: Sigma, 31170-5G-F, 99%, solid; cis-pinonic acid: Aldrich, 110108-5G, 98%, solid) in distilled, deionized water (ACS ASTM Type I, Ricca chemical CAS: 7732-18-5), or ethanol (Sigma-Aldrich 200 proof HPLC/Spectrophotometric grade 459828) for one solution of 2-deoxy-d-ribose. Atomized aerosol samples were generated using a TSI Model 3076 Aerosol Generator using these solutions. Particles were dried online using a diffusion dryer filled with $CaSO_4$ and collected on 37 mm Teflon (PTFE) filters (Pall Corporation). The sample collection area is restricted to a circular area with a diameter of 10 mm in the center of the filter by Teflon masking elements placed above and below each filter. These masking elements are effectively thin Teflon washers with inner and outer diameters of 10 and 37 mm, respectively. Filters were allowed to equilibriate in <55% RH conditions for 24 h before analyzing with a microbalance (Sartorius) with 0.1 μg resolution for gravimetric analysis, and a Bruker Tensor 27 spectrometer with DTGS detector for measurement of infrared transmission using 2 cm^{-1} resolution. Consideration of aerosol water retention and effects on functional group quantification is provided in Section S4. A background spectrum was acquired for every five sample spectra, and each spectrum is the average of 128 scans. Filters were scanned with the spectrometer before and after sample collection, and the FTIR sample chamber was purged with pure nitrogen for three minutes prior to and during scanning for each filter.

Mass loadings on filters are determined through difference of presample and postsample weights of the substrate by gravimetric analysis (in the single-solute case, there is no assumption of dissolution estimates or solution concentrations required). The measured gravimetric mass is divided by the compound molecular weight and multiplied by functional group abundance on a per molecule basis to estimate the total moles of functional groups in each sample. The quantity referred to as "absorptivity" in this manuscript is adopted from previous use in analysis of organic functional group abundance in atmospheric aerosols (Maria et al. 2002, 2003), and is defined as the integrated absorbance (cm^{-1}, as absorbance is unitless) per micromole of functional group; collectively expressed in units of $cm^{-1}\,\mu mol^{-1}$. This absorptivity we report is the integrated molar absorption coefficient divided by the the sample collection area (a circle with a diameter of 10 mm for our configuration), and its derivation from the Beer–Lambert Law is shown in Section S4.

3. PARAMETER QUANTIFICATION AND EVALUATION

3.1. Absorptivities

Samples of reference standards are collected by the method described in Section 2.2, and the algorithmic decomposition described in Section 2.1 is applied to the infrared spectra of these samples. The integrated absorbance is obtained by integrating the lineshapes of apportioned components with respect to wavenumber (Section 2.1.2).

Values of integrated absorbances for each functional group are regressed against measured quantity of functional groups (expressed in moles) with a one-parameter linear model (containing a slope with no intercept, derived from Beer–Lambert Law as described in Section S4) for individual compounds (Figures 5a–c). The absorptivities calculated from regression slopes for each compound are shown in Figure 6. The 10 saccharide compounds (from the work by Russell et al. 2010) are regressed as a single group for alcohol COH absorptivity, as absorbance at a single concentration was measured for each compound (Figures 5c and 6a). Carboxylic acid standard spectra exhibited a range of absorbance shapes, particularly in the region between 3400 and 2400 cm^{-1} (COH-stretching region). Therefore, the residual area after alkane CH lineshape subtraction, $s(\tilde{\nu}) - \sum_k g_k(\tilde{\nu})$ from Equation (4), is considered to be the best estimate of integrated absorbance for carboxylic COH as the compounds used for calibration contain no other functional group. The areas and absorpitivities calculated by this approach are shown in Figures 5a and 6c.

Absorptivities are generally consistent for each functional group (Maria et al. 2003; Gilardoni et al. 2007; Russell et al. 2009). However, small variations in the absorbance band and absorptivity occur for different conditions in which each chemical bond exists, and it is therefore reasonable to expect some variation in absorptivities for pure compound standards. When the absorptivities among compounds are not statistically significantly different, the points can be regressed together to get a single absorptivity value without any loss of information. For cases in which any of the regression slopes for each compound are significantly different, we can aggregate them according to our interpretation of mixture composition. This aggregation of varying absorptivities is expressed as a weighted sum. Let $w_i =$ weights, $\alpha_i =$ absorptivity, and $\bar{\alpha} =$ weighted absorptivity:

$$\bar{\alpha} = \sum_{i=1}^{n} w_i \alpha_i \ . \qquad [6]$$

Weights can reflect prior beliefs about the proportion of each of the n bond types in each sample. In the absence of prior

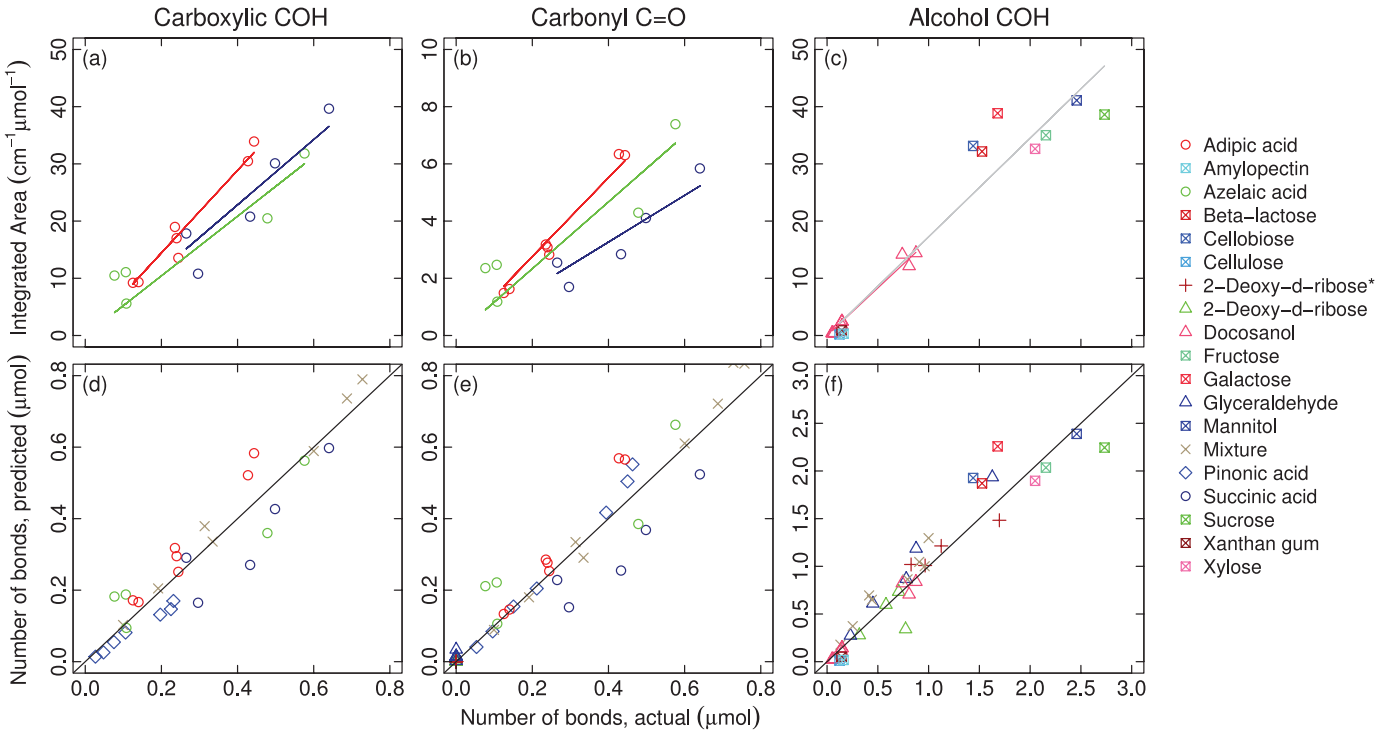

FIG. 5. Molar absorptivity calibration for each functional group and subsequent evaluation by functional group (columnwise organization): integrated absorbance (top row), predicted number of functional group bonds (lower panel). (*) indicates dissolution of 2-deoxy-d-ribose in ethanol rather than water for atomization (Section 2.2). The diagonal lines spanning lower left to upper right corners of plot boxes in lower panels indicate the $x = y$ relationship. Saccharide compounds measurements are reported by Russell et al. (2010) (Table S2). (Color figure available online.)

information of sample composition, we use a uniform ("uninformed") distribution, where $w_i = w = 1/n$. A uniform weighting is also used for multicomponent mixtures (e.g., ambient samples) where these individual variations are expected to be diminishingly small.

Absorptivities for each functional group calculated by this method are shown in Table 4. All alcohol COH bond measurements for docosanol and saccharides in aqueous solution (measured and reported by Gilardoni et al. [2007] and Russell et al. [2010], respectively) are regressed together, as there is not a significant variation among these compounds. In this table, we report a carboxylic COH absorptivity slightly lower than that illustrated in Figures 5a and 6c, as we use a scaled area to represent an approximate carboxylic COH absorbance. This scaled area is calculated as the integral of $\beta a_\phi(\tilde{\nu})$ from Equation (4), where $a_\phi(\tilde{\nu})$ is the fixed carboxylic COH profile

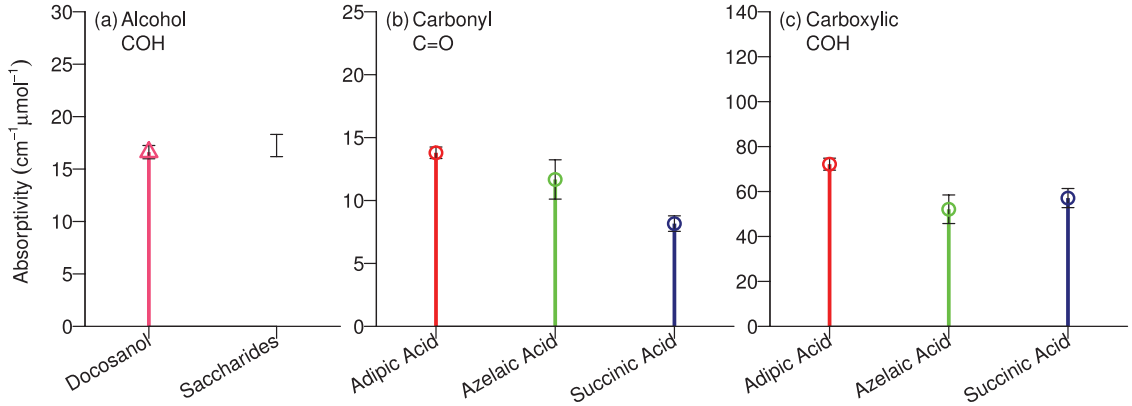

FIG. 6. Molar absorptivities for each functional group estimated from Figure 5. Black error bars represent ± standard error of regression slope, or propagated standard errors in case of aggregation of multiple regression slopes (Section 3.1). Saccharide compounds measurements are reported by Russell et al. (2010) (Table S2). (Color figure available online.)

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

TABLE 4
Absorptivities for oxygenated functional groups

Functional group	No. of compounds	No. of samples	Absorptivity ± std. error (cm^{-1} μmol)	Individual values ± std. error (cm^{-1} μmol)
Carboxylic COH	3	17	47.7 ± 6.6	Adipic acid (59.5 ± 1.8), azelaic acid (43.7 ± 5.2), succinic acid (39.8 ± 3.7)
Carbonyl C=O	3	17	11.2 ± 1.7	Adipic acid (13.8 ± 0.5), azelaic acid (11.7 ± 1.6), succinic acid (8.2 ± 0.6)
Alcohol COH	11	19	17.2 ± 0.7	(Gilardoni et al. 2007; Russell et al. 2010)

derived in Section 2.1.3. The reduced absorptivity is calculated in the same manner with the exception that the scaled area, rather than residual area, is used as the regressor. Using a fixed carboxylic COH profile paired with its corresponding absorptivity is more suitable for ambient sample spectra, as the residual area estimated by the subtraction method described above is likely to contain absorption due to other bonds present in atmospheric mixtures. The purpose of using the residual area in Figure 5a is to illustrate the applicability of the Beer–Lambert Law for the unapproximated area (i.e., linearity with respect to molar abundance), but we are also able to reproduce the molar quantities of carboxylic COH present in our reference samples using the scaled area approach (as evaluated in Figure 5d, and also in Table 5 described below). The variability in carboxylic COH absorptivities among compounds are approximately similar regardless of which metric is used. Carboxyic COH and carbonyl C=O absorptivities vary by 21% and 25%, respectively (as measured by relative standard deviation, or %RSD), among the dicarboxylic acids analyzed in this study. The prediction errors for bond abundance quantification (in n μmoles) are within 35% on average, and are also reported as molar values in Table 5. In this Table, mean bias and mean error between moles of predicted ($\hat{n}$) and actual (n) number of bonds for N measurements are defined as $\frac{1}{N}\sum_{i=1}^{N}(\hat{n}_i - n_i)$ and $\frac{1}{N}\sum_{i=1}^{N}|\hat{n}_i - n_i|$, respectively, and their relative values are defined by normalizing by n_i prior to averaging.

Allen et al. (1994) suggested that acid C=O has a stronger molar absorptivity than ketonic C=O by a factor of 2 or 3. According to this assertion, we should underpredict the quantity of carbonyl C=O for compounds containing ketonic C=O if we apply the molar absorptivity for carbonyl C=O derived from carboxylic acids. For instance, we should predict molar quantities 1.3 to 1.5 times higher for pinonic acid, which has a 1:1 molar ratio of ketonic to carboxylic C=O. We find that this is not the case, as the aggregated absorptivity derived from adipic, azelaic, and succinic acids (which only contain carboxylic C=O) applied to pinonic acid (containing both carboxylic and ketonic C=O) shows that the predicted moles of carbonyl C=O is within 5% of the actual values (Figure 5e).

Mean absorptivities reported in Table 4 are used in Section 3.2 for evaluation of the algorithm described previously (Section 2.1), as these uniformly-weighted absorptivities are representative of the subset of compounds analyzed for this study. The value to be used for ambient samples merits further consideration. Baselining parameters were slightly adjusted from the work of Russell et al. (2009) to be more consistent across reference compounds; this adjustment results in absorptivities for adipic acid that are 8% lower for both carboxylic COH and carbonyl C=O bonds. This difference is within analytical uncertainty

TABLE 5
Errors in predicted molar quantities of functional groups, corresponding to Figure 5. Mean bias and mean error are defined in Section 3.1

Functional group	Standards	Mean bias (μmol)	Mean error (μmol)	Mean relative bias (%)	Mean relative error (%)	Range (μmol)
Alcohol COH	Saccharides	0.04	0.16	−12	27	0.05–2.73
Alcohol COH	2-Deoxy-d-ribose, Mixture†	0.06	0.14	14	22	0.13–1.70
Carboxylic COH	Acids	−0.01	0.06	−1	21	0.03–0.64
Carboxylic COH	Mixture†	0.04	0.04	8	9	0.10–0.76
Carbonyl C=O	Acids	−0.01	0.06	−2	17	0.05–0.64
Carbonyl C=O	Mixture†	0.02	0.04	1	8	0.10–0.76
Carboxylic COOH	Acids	0.00	0.06	0	21	0.03–0.76
Carbonyl C=O	Ketone	0.06	0.06	33	33	0.03–0.23

†Equimolar mixture of 2-deoxy-d-ribose and succinic acid.

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

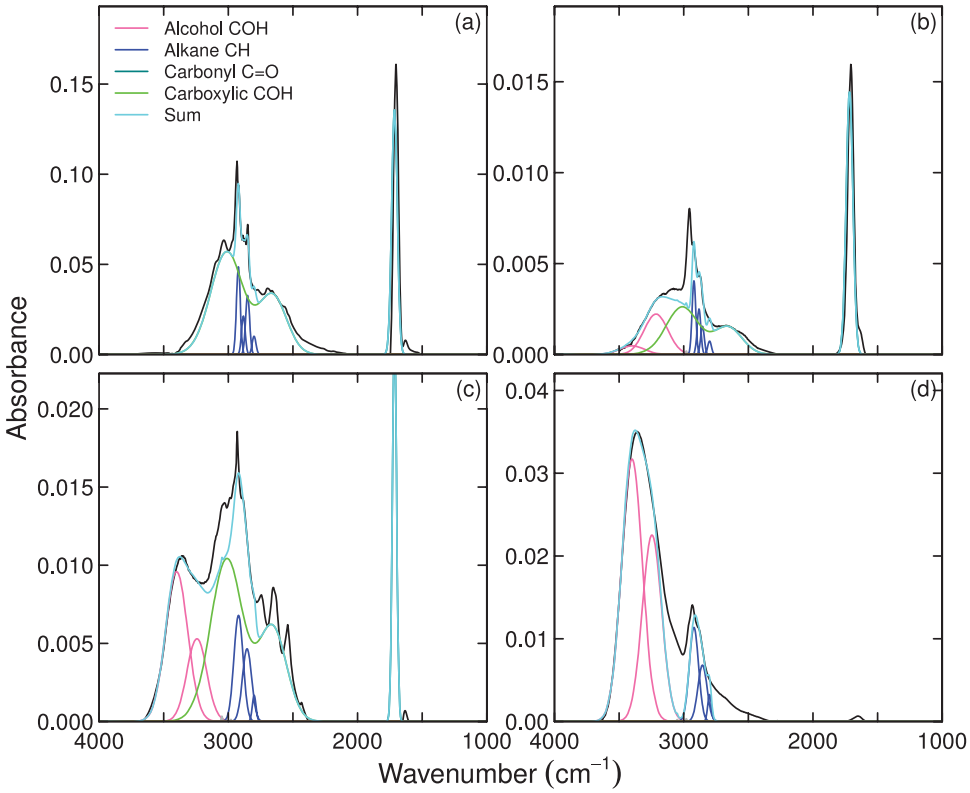

FIG. 7. Example of apportioned lineshapes for selected laboratory standard spectra of aerosols generated from aqueous solutions of (a) azelaic acid, (b) pinonic acid, (c) equimolar mixture of succinic and 2-deoxy-d-ribose, and (d) 2-deoxy-d-ribose. Curves include measured spectrum and fitted lineshapes. (Color figure available online.)

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

and can statistically and effectively be considered insignificant. While it remains that adipic acid—used previously for calibration by Gilardoni et al. (2007) and Russell et al. (2009)—has the highest absorptivity of the three dicarboxylic acids compounds analyzed here, the full range of possible absorptivities for monocarboxylic, alkanoic acids, and hydroxyacid among other plausible carboxylic group configurations remain uncharacterized. Therefore, we recommend the continued use of the values cited by Russell et al. (2009) (effectively corresponding to a weight of $\alpha = 1$ for adipic acid and zero for all others; Equation (6)) in our studies for the merit of consistency and comparability, until further experiments determine a value representative for a larger class of organic acids.

3.2. Evaluation: Laboratory-Generated Aerosols

Examples of reference sample spectra apportioned by contributions from constituent functional groups as described in Section 2.1 are shown in Figure 7. Compounds in neat preparation, especially small, highly-oxygenated compounds, are subject to accumulation of specific intermolecular forces—e.g., hydrogen bonding between protonated and deprotonated groups (Hay and Myneni 2007). Repeatable deviations from average or expected lineshapes (as expressed by aforementioned parameterizations) are therefore observed for a selected subset of samples. However, these deviations are expected to become less severe in mixtures (a relevant consideration for analysis of atmospheric samples), and is confirmed by reasonably accurate predictions for moles of carboxylic COH, alcohol COH, and carboxylic C=O (Figures 5d–f and Table 5).

A comparison of quantified moles of total carbonyl C=O and carboxylic COH for reference samples are shown in Figure 8. The calibrated absorptivities for carboxylic COH and carbonyl C=O absorption yield approximately equivalent number of moles for each where expected (all carbonyl C=O is associated with the carboxylic acid group), and departs from the $n_{\mathrm{tC{=}O}} = n_{\mathrm{cCOH}}$ relationship when ketonic C=O is present. As seen in Figure 8, the number of moles of carbonyl C=O in pinonic acid exceeds the equimolar relationship, which is the expected result and illustrates the way in which the measured carbonyl C=O can be apportioned between carboxylic acid and other carbonyl. The anticipated value of the slope is two, according to our understanding that pinonic acid contains one mole of ketonic C=O for every mole of carboxylic group. The mole ratio of carbonyl C=O to carboxylic COH as quantified by regression indicates that we overestimate the slope by 52% in this case. This discrepancy is attributed to the underprediction of quantified carboxylic COH (Figure 5d), possibly due to variations in molar absorptivity or departure of absorption lineshape

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

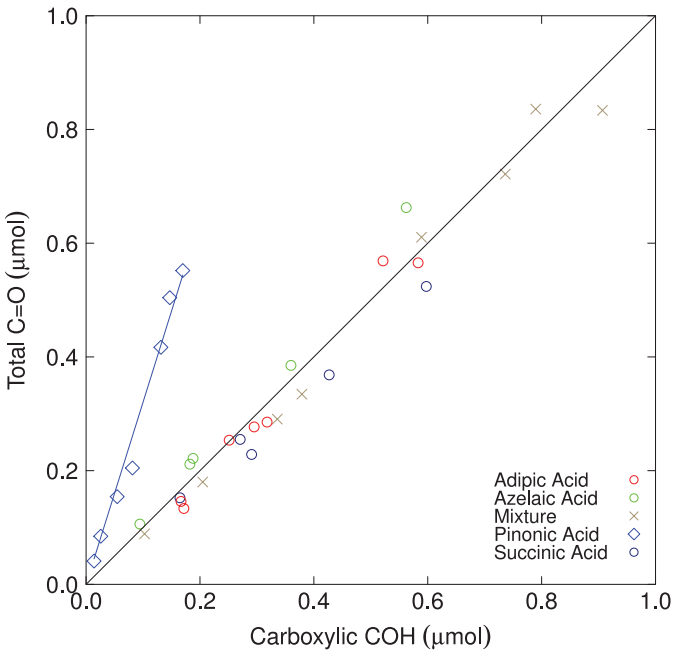

FIG. 8. Comparison of quantified amount of total carbonyl C=O and carboxylic COH. Glyph symbols and colors are the same as for Figure 5. The diagonal line spanning lower left to upper right corner of plot box indicates the $x = y$ relationship. (Color figure available online.)

from the profile used for our analysis. Applying our algorithm (Equation (5)), this leads to a 34% underprediction and 33% overprediction in the number of moles of carboxylic COH (and hence carboxylic COOH) and ketonic C=O, respectively.

Aldehydic C=O has not been reported separately from ketonic C=O in previous studies using this algorithm. The aldehyde C=O absorption peak appears at approximately 1700 cm^{-1}, depending on surrounding oxygenation or saturation, and is possibly separable from acidic or ketonic C=O when differences in peak location are detectable. For aldehyde-containing compounds analyzed in this study, no carbonyl C=O peak was detected, and Fermi doublet peaks (from the CH stretching mode) expected between 2860 and 2710 cm^{-1} (Lambert et al. 1998) indicating the presence of aldehydic C=O groups were not observed in our absorbance spectra. Many aldehydic compounds are reactive and can hydrate to form diols containing alcohol COH groups in aqueous solution (Schwarzenbach et al. 2002). In agreement with our understanding that 2-deoxy-d-ribose in aqueous solvent is converted into a mixture of deoxyribopyranose (75%) and deoxyribofuranose (25%) which contain no carbonyl C=O bonds, we do not detect carbonyl C=O (nor the aldehydic doublet) and observe the predicted amount of alcohol COH for 2-deoxy-d-ribose aerosols generated from aqueous solution. We make a similar observation for aerosols generated from an equimolar mixture of 2-deoxy-d-ribose and succinic acid in aqueous solvent (Figures 5f and 7c). 2-Deoxy-d-ribose in an ethanol solution undergoes a similar conversion; no aldehydic doublet or quantifiable carbonyl C=O is observed in these aerosol samples. Glyceraldehyde is predicted to form glyceraldehyde dimers in aqueous solution, which is also supported by our predictions of functional group abundance (Figure 5f).

3.3. Evaluation: Ambient Samples

We apply the same algorithm described in Sections 2.1.2 and 2.1.3 to infrared spectra of ambient aerosols (Figure 9). We find that the >90% of the integrated absorbance is accounted for when the total area is >50 cm^{-1} by the algorithm described above (Figure 10). The range of fitted peak parameters are shown in Table 6. The amount of absorption by functional group and their absorptivities (Section 3.1 and Gilardoni et al. 2007) are used to estimate the OM mass concentrations for each sample using the algorithm described in Sections 4 and S6.

A comparison of quantified total C=O and carboxylic COH for ambient samples shows a wider deviation from equimolar quantities at lower concentrations than observed for laboratory standards. In addition, there are several campaigns for which the quantified moles of carboxylic COH exceeds total carbonyl C=O ($n_{cCOH} < n_{tC=O}$; Equation (5)) beyond normal variations about the $n_{tC=O} = n_{cCOH}$ line (Figure 11). These fall into the "otherwise" case (Equation (5)) discussed in Section 2.1.3, and we rely on carboxylic COH to determine the amount of carboxylic COOH in these cases. Other explanations consistent with observation may be the overestimation of carboxylic COH area, underestimation of the true carboxylic COH absorptivity, or overestimation of carboxylic C=O absorptivity for the given samples, but are considered less likely given the potential range of biases characterized in Section 3.2. However, further studies are necessary to strengthen this argument.

TABLE 6
Range (5–95th percentiles) of Gaussian peak-fitting parameter values for field campaigns since 2006 (residuals displayed in Figure 10). The Gaussian form and parameters are from Equation (T1-3)

Functional group	Peak no.	μ (cm^{-1})	σ (cm^{-1})
Alcohol	1	3400–3600	30–100
Alcohol	2	3200–3355	30–75
Alkane	1	2921–2931	10–25
Alkane	2	2876–2886	10–25.0
Alkane	3	2849–2855	10–25.5
Alkane	4	2790–2814	10–22.5
Unidentified	1	3136–3142	18.0–19.7
Unidentified	2	3070–3076	10.8–20.2
Aromatic	1	3050–3050	3.5–3.5
Unidentified	3	3008–3014	13.4–23.9
Alkene	1	2980–2980	3.5–3.5
Unidentified	4	2953–2958	8.0–12.3
Carbonyl	1	1714–1720	10–29.9
Amine	1	1620–1630	10–30

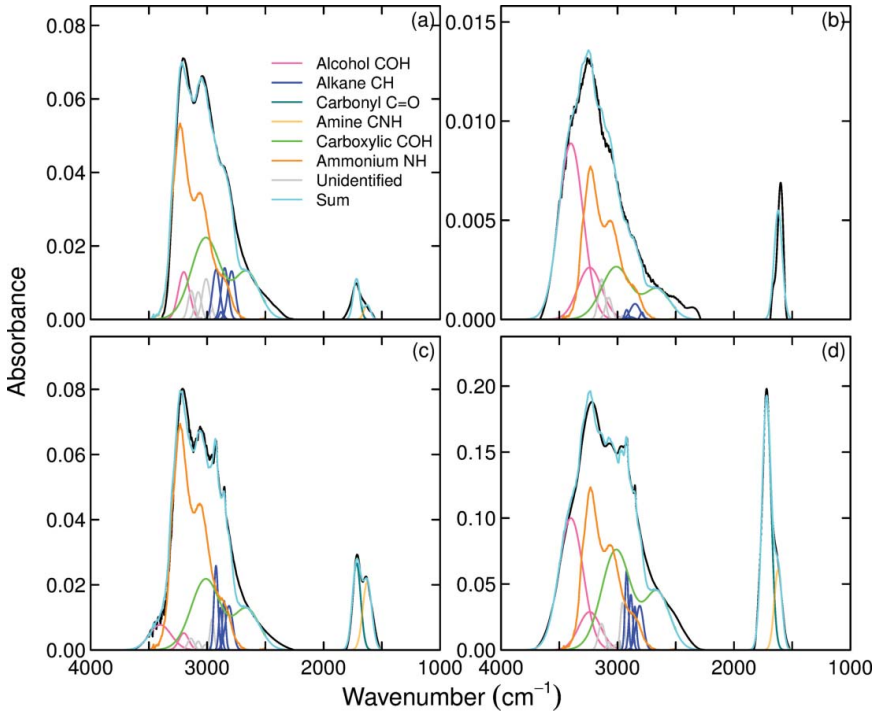

FIG. 9. Examples of apportioned area for selected ambient sample spectra: from (a) Whistler Peak, 2009; (b) N. Atlantic Ocean, 2009 (ICEALOT); (c) Mexico City, 2006 (MILAGRO); and (d) Tijuana, 2010 (CalMex). Curves are colored as in Figure 7, with the addition of ammonium NH, alkene CH, aromatic CH, and unidentified (Section S6) groups. (Color figure available online.)

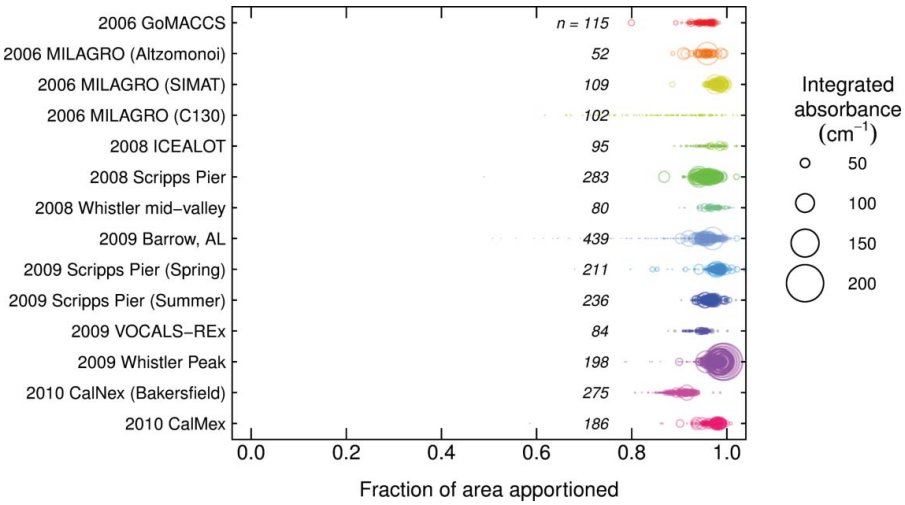

FIG. 10. Summary of spectra area of ambient aerosols apportioned according to the lineshape fitting described in Section 2.1.2. Field campaigns since 2006 are included: 2006 GoMACCS (Russell et al. 2009); 2006 MILAGRO (Gilardoni et al. 2009; Liu et al. 2009); 2008 Whistler mid-valley (Schwartz et al. 2010); 2008 ICEALOT (Frossard et al. 2011); 2008 Scripps Pier (Hawkins and Russell 2010); 2009 Barrow, AL (Shaw et al. 2010); 2009 Scripps Pier Summer (Day et al. 2010) and Spring (Liu et al. 2011); 2009 VOCALS-REx (Hawkins et al. 2010); 2009 Whistler Peak (Takahama et al. 2011); 2010 CalNex Bakersfield (Alhm2012); 2010 CalMex (Takahama et al. 2012). (Color figure available online.)

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

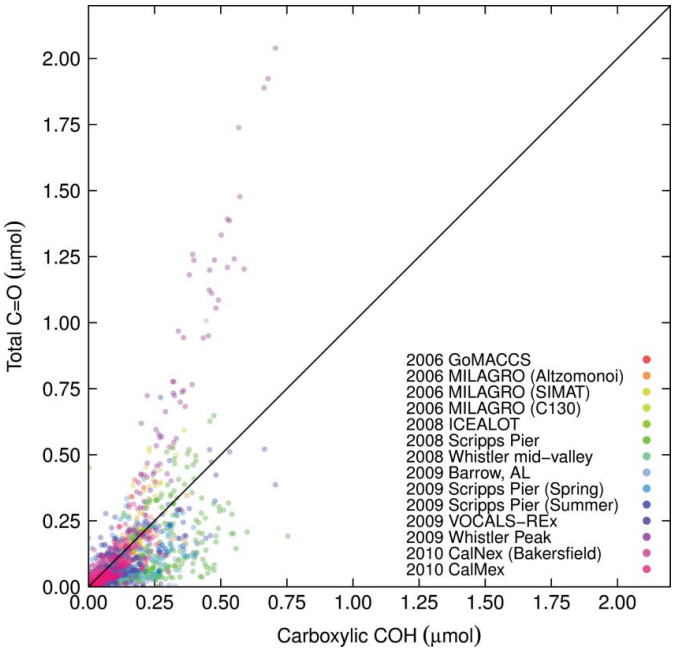

FIG. 11. Comparison of quantified amount of total carbonyl C=O and carboxylic COH. Color scheme indicates field campaign and is consistent with those used for Figure 10; the samples included for each campaign is also identical. The diagonal line indicates the $x = y$ relationship. (Color figure available online.)

In Figure 11, the presence of nonacid C=O is apparent in several campaigns (from the observation that $n_{\mathrm{tC=O}} \gg n_{\mathrm{cCOH}}$; Equation (5)), and it is desirable to know whether this is due to contributions from aldehydic or ketonic C=O. Given the ubiquity of aerosol-water interactions in the atmosphere (Rood et al. 1989; Martin 2000), we expect that a subset of aldhehydes ambient aerosol samples will be hydrated and carbonyl C=O will be undetected by FTIR analysis (as reported in Section 3.2). In the broader scope of chemical constituents present in the atmospheric particles, however, hydration is not expected for all aldehydic compounds. Compounds such as glyceraldehyde (included in this study) have strongly electron withdrawing substituents with low steric hindrance, leading to high equilibrium constants for hydration ($K_{Hydr} \gg 1$; Wasa and Musha 1970; Lewis and Wolfenden 1973; Creighton et al. 1988). Propionaldehyde, acetaldehyde, and possibly other aldehydic compounds expected to be present in atmospheric particles (De Andrade et al. 1993; Yu et al. 1999) have lower hydration equilibrium constants ($K_{Hydr} \lesssim 1$; Lewis and Wolfenden 1973; Guthrie 1975), which can be due to stabilizing electronic, symmetry, or steric effects of the unhydrated molecule (Guthrie 1975; Buschmann et al. 1980; Wiberg et al. 1994). Therefore, caution is warranted when interpreting spectra from ambient aerosol (and also smog chamber) samples. However, in the sample spectra associated with excess carbonyl C=O (Figure 11), the characteristic aldehydic doublet (described in Section 3.2) was not detected in these ambient sample spectra. Thus, the excess carbonyl C=O was attributed to ketonic C=O. The presence and quantity of ketonic C=O has been correlated with biogenic SOA and biomass burning aerosol in various campaigns (Liu et al. 2009; Hawkins and Russell 2010; Schwartz et al. 2010; Liu et al. 2011; Takahama et al. 2011). A higher OM fraction of carboxylic COOH, on the other hand, has been associated with anthropogenic combustion (Liu et al. 2009; Russell et al. 2009); Hawkins and Russell 2010; (Schwartz et al. 2010; Liu et al. 2011; Takahama et al. 2011, 2012), and alcohol COH with marine aerosol (Hawkins and Russell 2010; Hawkins et al. 2010; Russell et al. 2010; Takahama et al. 2012), highlighting the relevance of functional group characterization in source identification and apportionment (Russell et al. 2011).

4. MOLECULAR BONDS TO ATOMIC COMPOSITION AND MASS CONCENTRATIONS

Having a set of measurements which provides the total number of moles of functional group j, we can estimate the associated mass and other chemical metrics from inferred atomic composition. We relate the total number of moles of functional group j, n_j, to the number of moles of atom type i, n_i, from the following equation:

$$n_i = \sum_{j \in \mathcal{G}} \sum_{k \in \mathcal{D}} \lambda_{ijk} f_{jk} n_j \ . \qquad [7]$$

f_{jk} is the fraction of moles of j in configuration k, defined such that $\sum_{k \in \mathcal{D}} f_{jk} = 1$ for all j in $\mathcal{G}$; λ_{ijk} is the element of the apportionment array which maps moles of functional group j in configuration k to the moles of i. $\mathcal{G}$ is the set of all functional groups, and $\mathcal{D}$ is the set of possible configurations (degree of saturation and functionalization). If we let $\bar{\mathbf{\Lambda}} = [\bar{\lambda}_{ij}]_{i \in \mathcal{A}, j \in \mathcal{G}}$ where $\bar{\lambda}_{ij} = \sum_{k \in \mathcal{D}} \lambda_{ijk} f_{jk}$, we can rewrite Equation [7] as a matrix product,

$$\mathbf{n}_{\mathcal{A}} = \bar{\mathbf{\Lambda}}\, \mathbf{n}_{\mathcal{G}} \ , \qquad [8]$$

where $\mathcal{A}$ is the set of atoms. From $\mathbf{n}_{\mathcal{A}}$, we can estimate the organic mass (OM), H/C ratio, O/C ratio, and functionalization rate (Dron et al. 2010) (Section S6). The choice of $\bar{\mathbf{\Lambda}}$ is based on the analysis by Russell (2003), where values of f_{jk} are inferred from field measurements (Section S6).

Various field studies indicate that the OM reported by this method is generally within ±30% of other independent estimates (AMS: Gilardoni et al. 2009; Russell et al. 2009; Frossard et al. 2011; Liu et al. 2011). Estimated O/C atomic ratios reported by the algorithm are generally lower and exhibit a smaller dynamic range than O/C values measured by AMS and ACSM in the field (even after time-averaging of AMS and ACSM measurements to the time resolution of FTIR filter samples) (Russell et al. 2009; Takahama et al. 2012). For secondary organic aerosol (SOA) produced in an environmental chamber, O/C and H/C ratios estimated with canonical and adjusted $\bar{\mathbf{\Lambda}}$s (Section

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

S6) are bounded by AMS values reported at higher time resolution (Chhabra et al. 2011).

5. CONCLUSIONS

This work describes the algorithm and parameters by which infrared absorption regions relevant for quantification of carboxylic COOH and carbonyl C=O are apportioned into their constituent components using simple lineshapes and a sequential fitting process. Using a statistical decomposition method, we extract an average lineshape for carboxylic COH from profiles of a fatty acid series and assign an absorptivity from a laboratory calibration from carboxylic acid compounds and mixtures. In addition to adipic acid which was used previously (Gilardoni et al. 2007), we use azelaic and succinic acids to obtain molar absorptivities for carboxylic C=O and COH bonds. Our algorithm for quantifying carboxylic COOH is applied to a multifunctional compound (pinonic acid), and to a multifunctional mixture consisting of succinic acid and 2-deoxy-d-ribose (50/50 molar %) with 21% average prediction accuracy. The ketonic C=O in pinonic acid is detected by our method by the amount of carbonyl in molar excess of carboxylic COH, though overestimated by 33% due to a bias in calibration. Functional group abundances of aldehydic compounds (glyceraldehyde and 2-deoxy-d-ribose) were in quantitative agreement (within 30%) with the presence of diols formed from hydration reactions in aqueous solution. Given the prevalence of metastable states of aerosols in the atmosphere (Rood et al. 1989), many such aldehydic compounds can be hydrated in aqueous particles. In ambient sample spectra from field campaigns to date, doublet peak absorption characteristic of aldehydic compounds have not been observed, leading to attribution of excess carbonyl C=O to ketonic rather than aldehydic C=O. However, as there are aldehydic compounds expected in aqueous particles which are not expected to hydrate completely (e.g., acetaldehyde, pinonaldehyde), caution is still warranted when interpreting ambient and smog chamber aerosol spectra. The software programmed from this algorithm has been applied to the analysis of particles collected on Teflon substrates during various campaigns (Gilardoni et al. 2009; Liu et al. 2009; Russell et al. 2009; Bahadur et al. 2010; Hawkins and Russell 2010; Hawkins et al. 2010; Shaw et al. 2010; Day et al. 2010; Schwartz et al. 2010; Chhabra et al. 2011; Frossard et al. 2011; Liu et al. 2011; Takahama et al. 2011, 2012). The articulation of this algorithm and its implementation improves reproducibility of FTIR spectra interpretation across users, and provides an explicit framework by which additional constraints can be incorporated to reduce the uncertainty in the analysis.

REFERENCES

Adachi, K., and Buseck, P. R. (2008). Internally Mixed Soot, Sulfates, and Organic Matter in Aerosol Particles from Mexico City. *Atmos. Chem. Phys.,* 8:6469–6481.

Aiken, A. C., Decarlo, P. F., Kroll, J. H., Worsnop, D. R., Huffman, J. A., Docherty, K. S., et al. (2008). O/C and OM/OC Ratios of Primary, Secondary, and Ambient Organic Aerosols with High-Resolution Time-of-Flight Aerosol Mass Spectrometry. *Environ. Sci. Technol.,* 42:4478–4485, doi:10.1021/es703009q.

Allen, D. T., Palen, E. J., Haimov, M. I., Hering, S. V., and Young, J. R. (1994). Fourier-Transform Infrared-Spectroscopy of Aerosol Collected In A Low-pressure Impactor (lpi/ftir)—Method Development and Field Calibration. *Aerosol Sci. Technol.,* 21:325–342, doi:10.1080/02786829408959719.

Allan, J. D., Jimenez, J. L., Williams, P. I., Alfarra, M. R., Bower, K. N., Jayne, J. T., et al. (2003). Quantitative Sampling using an Aerodyne Aerosol Mass Spectrometer—1. Techniques of Data Interpretation and Error Analysis. *J. Geophys. Res.-Atmos.,* 108:4090, doi:10.1029/2002JD002358.

Asher, W. E., Pankow, J. F., Erdakos, G. B., and Seinfeld, J. H. (2002). Estimating the Vapor Pressures of Multi-Functional Oxygen-Containing Organic Compounds using Group Contribution Methods. *Atmos. Environ.,* 36: PII S1352–2310(01)00 564–7, doi:10.1016/S1352-2310(01)00564-7.

Bahadur, R., Uplinger, T., Russell, L. M., Sive, B. C., Cliff, S. S., Millet, D. B., et al. (2010). Phenol Groups in Northeastern US Submicrometer Aerosol Particles Produced from Seawater Sources. *Environ. Sci. Technol.,* 44:2542–2548, doi:10.1021/es9032277.

Blout, E. R., Fields, M., and Karplus, R. (1948). Absorption Spectra .6. the Infrared Spectra of Certain Compounds Containing Conjugated Double Bonds. *J. Am. Chem. Soc.,* 70:194–198, doi:10.1021/ja01181a056.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control* (4th ed.), John Wiley & Sons, Inc., Hoboken, NJ.

Buschmann, H.-J., Ffldner, H.-H., and Knoche, W. (1980). The Reversible Hydration of Carbonyl Compounds in Aqueous Solution. Part I, The Keto/Gem-diol Equilibrium. *Berichte der Bunsengesellschaft fflr physikalische Chemie,* 84:41–44, doi:10.1002/bbpc.19800840109.

Chan, M. N., Lee, A. K. Y., and Chan, C. K. (2006). Responses of Ammonium Sulfate Particles Coated with Glutaric Acid to Cyclic Changes in Relative Humidity: Hygroscopicity and Raman Characterization. *Environ. Sci. Technol.,* 40:6983–6989, doi:10.1021/es060928c.

Chang, E. I., and Pankow, J. F. (2006). Prediction of Activity Coefficients in Liquid Aerosol Particles Containing Organic Compounds, Dissolved Inorganic Salts, and Water – Part 2: Consideration of Phase Separation Effects by an X-UNIFAC Model. *Atmos. Environ.,* 40:6422–6436, doi:10.1016/j.atmosenv.2006.04.031.

Chattopadhyay, S., and Ziemann, P. J. (2005). Vapor Pressures of Substituted and Unsubstituted Monocarboxylic and Dicarboxylic Acids Measured using an Improved Thermal Desorption Particle Beam Mass Spectrometry Method. *Aerosol Sci. Technol.,* 39:1085–1100, doi:10.1080/02786820500421547.

Chhabra, P. S., Flagan, R. C., and Seinfeld, J. H. (2010). Elemental Analysis of Chamber Organic Aerosol using an Aerodyne High-Resolution Aerosol Mass Spectrometer. *Atmos. Chem. Phys.,* 10:4111–4131, doi:10.5194/acp-10-4111-2010.

Chhabra, P. S., Ng, N. L., Canagaratna, M. R., Corrigan, A. L., Russell, L. M., Worsnop, D. R., et al. (2011). Elemental Composition and Oxidation of Chamber Organic Aerosol. *Atmos. Chem. Phys.,* 11:8827–8845, doi:10.5194/acp-11-8827-2011.

Coates, J. (2006). Interpretation of Infrared Spectra, A Practical Approach, in: *Encyclopedia of Analytical Chemistry*, Meyers, R. A., (ed), John Wiley & Sons, Ltd., Chichester, p. 10815–10837.

Coury, C., and Dillner, A. M. (2008). A Method to Quantify Organic Functional Groups and Inorganic Compounds in Ambient Aerosols using Attenuated Total Reflectance FTIR Spectroscopy and Multivariate Chemometric Techniques. *Atmos. Environ.,* 42:5923–5932, doi:10.1016/j.atmosenv.2008.03.026.

Creighton, D. J., Migliorini, M., Pourmotabbed, T., and Guha, M. K. (1988). Optimization of Efficieny in the Glyoxylase Pathway. *Biochemistry,* 27:7376—7384.

Day, D. A., Liu, S., Russell, L. M., and Ziemann, P. J. (2010). Organonitrate Group Concentrations in Submicron Particles with High Nitrate and Organic Fractions in Coastal Southern California. *Atmos. Environ.*, 44:1970–1979, doi:10.1016/j.atmosenv.2010.02.045.

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

De Andrade, J. B., Pinheiro, H. L. C., and Andrade, M. V. (1993). Determination Of Formaldehyde and Acetaldehyde Associated to Atmospheric Aerosols by HPLC. *Int. J. Environ. Anal. Chem.,* 52:49–56, doi: 10.1080/03067319308042847.

Decesari, S., Mircea, M., Cavalli, F., Fuzzi, S., Moretti, F., Tagliavini, E., et al. (2007). Source Attribution of Water-Soluble Organic Aerosol by Nuclear Magnetic Resonance Spectroscopy. *Environ. Sci. Technol.,* 41:2479–2484, doi:10.1021/es061711l.

Donahue, N. M., Epstein, S. A., Pandis, S. N., and Robinson, A. L. (2011). A Two-Dimensional Volatility Basis Set: 1. Organic-Aerosol Mixing Thermodynamics. *Atmos. Chem. Phys.,* 11:3303–3318, doi:10.5194/acp-11-3303-2011.

Dron, J., El Haddad, I., Temime-Roussel, B., Jaffrezo, J.-L., Wortham, H., and Marchand, N. (2010). Functional Group Composition of Ambient and Source Organic Aerosols Determined by Tandem Mass Spectrometry. *Atmos. Chem. Phys.,* 10:7041–7055, doi:10.5194/acp-10-7041-2010.

Erdakos, G. B., Chang, E. I., Pankow, J. F., and Seinfeld, J. H. (2006). Prediction of Activity Coefficients in Liquid Aerosol Particles Containing Organic Compounds, Dissolved Inorganic Salts, and Water - Part 3: Organic Compounds, Water, and Ionic Constituents by Consideration of Short-, mid-, and Long-Range Effects using X-UNIFAC.3. *Atmos. Environ.,* 40:6437–6452, doi:10.1016/j.atmosenv.2006.04.001.

Fredenslund, A., Jones, R., and Prausnitz, J. (1975). Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. *AIChE J.,* 21:1086–1099.

Frossard, A. A., Shaw, P. M., Russell, L. M., Kroll, J. H., Canagaratna, M. R., Worsnop, D. R., et al. (2011). Springtime Arctic Haze Contributions of Submicron Organic Particles from European and Asian Combustion Sources. *J. Geophys. Res.-Atmos.,* 116:D05205, doi:10.1029/2010JD015178.

Gilardoni, S., Liu, S., Takahama, S., Russell, L. M., Allan, J. D., Steinbrecher, R., et al. (2009). Characterization of Organic Ambient Aerosol During MIRAGE 2006 on Three Platforms. *Atmos. Chem. Phys.,* 9:5417–5432.

Gilardoni, S., Russell, L. M., Sorooshian, A., Flagan, R. C., Seinfeld, J. H., Bates, T. S., et al. (2007). Regional Variation of Organic Functional Groups in Aerosol Particles on Four US East Coast Platforms During the International Consortium for Atmospheric Research on Transport and Transformation 2004 Campaign. *J. Geophys. Res.-Atmos.,* 112:D10S27, doi: 10.1029/2006JD007737

Goldstein, A. H., Worton, D. R., Williams, B. J., Hering, S. V., Kreisberg, N. M., Panic, O., et al. (2008). Thermal Desorption Comprehensive Two-Dimensional Gas Chromatography for in-situ Measurements of Organic Aerosols. *J. Chromatogr. A,* 1186:340–347, doi:10.1016/j.chroma.2007.09.094.

Griffiths, P., and Haseth, J. A. D. (2007). *Fourier Transform Infrared Spectrometry* (2nd ed.), John Wiley & Sons, Hoboken, NJ.

Guthrie, J. P. (1975). Carbonyl Addition-Reactions – Factors Affecting Hydrate-Hemiacetal and Hemiacetal-Acetal Equilibrium-Constants. *Canadian J. Chem.-Rev. Canadienne De Chimie,* 53:898–906, doi:10.1139/v75-125.

Hamilton, J. F., Webb, P. J., Lewis, A. C., Hopkins, J. R., Smith, S., and Davy, P. (2004). Partially Oxidised Organic Components in Urban Aerosol using GCXGC-TOF/MS. *Atmos. Chem. Phys.,* 4:1279–1290.

Hawkins, L. N., and Russell, L. M. (2010). Oxidation of Ketone Groups in Transported Biomass Burning Aerosol from the 2008 Northern California Lightning Series Fires. *Atmos. Environ.,* 44:4142–4154, doi: 10.1016/j.atmosenv.2010.07.036.

Hawkins, L. N., Russell, L. M., Covert, D. S., Quinn, P. K., and Bates, T. S. (2010). Carboxylic Acids, Sulfates, and Organosulfates in Processed Continental Organic Aerosol Over the Southeast Pacific Ocean During VOCALS-REx 2008. *J. Geophys. Res.-Atmos.,* 115:D13201, doi:10.1029/2009JD013276.

Hay, M. B., and Myneni, S. C. B. (2007). Structural Environments of Carboxyl Groups in Natural Organic Molecules from Terrestrial Systems. Part 1: Infrared Spectroscopy. *Geochimica Et Cosmochimica Acta,* 71:3518–3532, doi:10.1016/j.gca.2007.03.038.

Holes, A., Eusebi, A., Grosjean, D., and Allen, D. T. (1997). FTIR Analysis of Aerosol Formed in the Photooxidation of 1,3,5-Trimethylbenzene. *Aerosol Sci. Technol.,* 26:516–526: doi:10.1080/02786829708965450.

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J. H., et al. (2009). Evolution of Organic Aerosols in the Atmosphere. *Science,* 326:1525–1529, doi:10.1126/science.1180353.

Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., et al. (2005). Organic Aerosol and Global Climate Modelling: A Review. *Atmos. Chem. Phys.,* 5:1053–1123.

Lambert, J. B., Shurvell, H. F., Lightner, D., and Cooks, R. G. (1998). *Organic Structural Spectroscopy*, Prentice Hall, Upper Saddle River, NJ.

Laskin, A., Cowin, J. P., and Iedema, M. J. (2006). Analysis of Individual Environmental Particles using Modern Methods of Electron Microscopy and X-Ray Microanalysis. *J. Electron Spectrosc. Relat. Phenom.,* 150:260–274, doi:10.1016/j.elspec.2005.06.008.

Lewis, C. A., and Wolfenden, R. (1973). Influence of Pressure on the Equilibrium of Hydration of Aliphatic Aldehydes. *J. Am. Chem. Soc.,* 95:6685–6688, doi: 10.1021/ja00801a026.

Lim, H. J., and Turpin, B. J. (2002). Origins of Primary and Secondary Organic Aerosol in Atlanta: Results' of Time-Resolved Measurements During the Atlanta supersite experiment. *Environ. Sci. Technol.,* 36:4489–4496, doi: 10.1021/es0206487.

Liu, S., Day, D. A., Shields, J. E., and Russell, L. M. (2011). Ozone-Driven Daytime Formation of Secondary Organic Aerosol Containing Carboxylic Acid Groups and Alkane Groups. *Atmos. Chem. Phys.,* 11:8321–8341, doi: 10.5194/acp-11-8321-2011.

Liu, S., Takahama, S., Russell, L. M., Gilardoni, S., and Baumgardner, D. (2009). Oxygenated Organic Functional Groups and Their Sources in Single and Submicron Organic Particles in MILAGRO 2006 Campaign. *Atmos. Chem. Phys.,* 9:6849–6863.

Maria, S. F., Russell, L. M., Turpin, B. J., and Porcja, R. J. (2002). FTIR Measurements of Functional Groups and Organic Mass in Aerosol Samples Over the Caribbean. *Atmos. Environ.,* 36:5185–5196, doi:10.1016/S1352-2310(02)00654-4.

Maria, S. F., Russell, L. M., Turpin, B. J., Porcja, R. J., Campos, T. L., Weber, R. J., et al. (2003). Source Signatures of Carbon Monoxide and Organic Functional Groups in Asian Pacific Regional Aerosol Characterization Experiment (ACE-Asia) Submicron Aerosol Types. *J. Geophys. Res.-Atmos.,* 108(D23):8637, doi:10.1029/2003JD003703.

Martin, S. T. (2000). Phase Transitions of Aqueous Atmospheric Particles. *Chem. Rev.,* 100:3403–3453, doi:10.1021/cr990034t.

Ming, Y., and Russell, L. M. (2002). Thermodynamic Equilibrium of Organic-Electrolyte Mixtures in Aerosol Particles. *Aiche J.,* 48:1331–1348.

Ng, N. L., Canagaratna, M. R., Zhang, Q., Jimenez, J. L., Tian, J., Ulbrich, I. M., et al. (2010). Organic Aerosol Components Observed in Northern Hemispheric Datasets from Aerosol Mass Spectrometry. *Atmos. Chem. Phys.,* 10:4625–4641, doi:10.5194/acp-10-4625-2010.

Noble, C. A., and Prather, K. A. (2000). Real-Time Single Particle Mass Spectrometry: A Historical Review of a Quarter Century of the Chemical Analysis of Aerosols. *Mass Spectrom. Rev.,* 19:248–274, doi:10.1002/1098-2787(200007)19:4¡248.

Ofner, J., Krueger, H. . U., Grothe, H., Schmitt-Kopplin, P., Whitmore, K., and Zetzsch, C. (2011). Physico-Chemical Characterization of SOA Derived from Catechol and Guaiacol—a Model Substance for the Aromatic Fraction of Atmospheric HULIS. *Atmos. Chem. Phys.,* 11:1–15, doi:10.5194/acp-11-1-2011.

Paatero, P., Hopke, P. K., Song, X. H., and Ramadan, Z. (2002). Understanding and Controlling Rotations in Factor Analytic Models. *Chemometr. Intell. Lab.*, 60:253–264.

Paatero, P., and Tapper, U. (1994). Positive Matrix Factorization – A Nonnegative Factor Model with Optimal Utilization of Error-estimates of Data Values. *Environmetrics,* 5:111–126.

Pankow, J. F., and Barsanti, K. C. (2009). The Carbon Number-Polarity Grid: A Means to Manage the Complexity of the Mix of Organic Compounds

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

When Modeling Atmospheric Organic Particulate Matter. *Atmos. Environ.,* 43:2829–2835, doi:10.1016/j.atmosenv.2008.12.050.

Pankow, J. F., and Chang, E. I. (2008). Variation in the Sensitivity of Predicted Levels of Atmospheric Organic Particulate Matter (OPM). *Environ. Sci. Technol.,* 42:7321–7329, doi:10.1021/es8003377.

Polidori, A., Turpin, B. J., Davidson, C. I., Rodenburg, L. A., and Maimone, F. (2008). Organic PM2.5: Fractionation by Polarity, FTIR Spectroscopy, and OM/OC Ratio for the Pittsburgh Aerosol. *Aerosol Sci. Technol.,* 42:233–246, doi:10.1080/02786820801958767.

Quinn, P. K., Bates, T. S., Coffman, D., Onasch, T. B., Worsnop, D., Baynard, T., et al. (2006). Impacts of Sources and Aging on Submicrometer Aerosol Properties in the Marine Boundary Layer Across the Gulf of Maine. *J. Geophys. Res.-Atmos.,* 111:D23S36, doi:10.1029/2006JD007582.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, ISBN 3-900051-07-0.

Reff, A., Turpin, B. J., Offenberg, J. H., Weisel, C. P., Zhang, J., Morandi, M., et al. (2007). A Functional Group Characterization of Organic PM2.5 Exposure: Results from the RIOPA Study RID C-3787-2009. *Atmos. Environ.,* 41:4585–4598, doi:10.1016/j.atmosenv.2007.03.054.

Rood, M. J., Shaw, M. A., Larson, T. V., and Covert, D. S. (1989). Ubiquitous Nature of Ambient Metastable Aerosol. *Nature,* 337:537–539, doi:10.1038/337537a0.

Rudich, Y., Donahue, N. M., and Mentel, T. F. (2007). Aging of Organic Aerosol: Bridging the Gap Between Laboratory and Field Studies. *Ann. Rev. Phys. Chem.,* 58:321–352, doi:10.1146/annurev.physchem.58.032806.104432.

Russell, L. M. (2003). Aerosol Organic-Mass-to-Organic-Carbon Ratio Measurements. *Environ. Sci. Technol.,* 37:2982–2987, doi:10.1021/es026123w.

Russell, L. M., Bahadur, R., and Ziemann, P. J. (2011). Identifying Organic Aerosol Sources by Comparing Functional Group Composition in Chamber and Atmospheric Particles. *Proc. Natl. Acad. Sci. USA,* 108:3516–3521, doi:10.1073/pnas.1006461108.

Russell, L. M., Hawkins, L. N., Frossard, A. A., Quinn, P. K., and Bates, T. S. (2010). Carbohydrate-Like Composition of Submicron Atmospheric Particles and Their Production from Ocean Bubble Bursting. *Proc. Natl Acad. Sci. USA,* 107:6652–6657, doi:10.1073/pnas.0908905107.

Russell, L. M., Takahama, S., Liu, S., Hawkins, L. N., Covert, D. S., Quinn, P. K., et al. (2009). Oxygenated Fraction and Mass of Organic Aerosol from Direct Emission and Atmospheric Processing Measured on the R/V Ronald Brown During TEXAQS/GoMACCS 2006. *J. Geophys. Res.-Atmos.,* 114: D00F05, doi:10.1029/2008JD011275.

Sax, M., Zenobi, R., Baltensperger, U., and Kalberer, M. (2005). Time Resolved Infrared Spectroscopic Analysis of Aerosol Formed by Photo-Oxidation of 1,3,5-Trimethylbenzene and Alpha-Pinene. *Aerosol Sci. Technol.,* 39:822–830, doi:10.1080/02786820500257859, rID F-1113-2010.

Schauer, J. J., Kleeman, M. J., Cass, G. R., and Simoneit, B. R. T. (1999). Measurement of Emissions from Air Pollution Sources. 1. C-1 Through C-29 Organic Compounds from Meat Charbroiling. *Environ. Sci. Technol.,* 33:1566–1577.

Schwartz, R. E., Russell, L. M., Sjostedt, S. J., Vlasenko, A., Slowik, J. G., Abbatt, J. P. D., et al. (2010). Biogenic Oxidized Organic Functional Groups in Aerosol Particles from a Mountain Forest Site and Their Similarities to Laboratory Chamber Products. *Atmos. Chem. Phys.,* 10:5075–5088, doi:10.5194/acp-10-5075-2010.

Schwarzenbach, R. P., Gschwend, P. M., and Imboden, D. M. (2002). *Environmental Organic Chemistry* (2nd ed.), John Wiley & Sons, Hoboken, NJ.

Seinfeld, J. H., and Pandis, S. N. (2006). *Atmospheric Chemistry and Physics* (2nd ed.), John Wiley & Sons, New York.

Shaw, P. M., Russell, L. M., Jefferson, A., and Quinn, P. K. (2010). Arctic Organic Aerosol Measurements Show Particles from Mixed Combustion in Spring Haze and from Frost Flowers in Winter. *Geophys. Res. Lett.,* 37:L10803, doi:10.1029/2010GL042831.

Smith, J. N., Moore, K. F., McMurry, P. H., and Eisele, F. L. (2004). Atmospheric Measurements of Sub-20 nm Diameter Particle Chemical Composition by Thermal Desorption Chemical Ionization Mass Spectrometry. *Aerosol Sci. Technol.,* 38:100–110, doi:10.1080/02786820490249036.

Song, M., Marcolli, C., Krieger, U. K., Zuend, A., and Peter, T. (2012). Liquid-Liquid Phase Separation and Morphology of Internally Mixed Dicarboxylic Acids/Ammonium Sulfate/Water Particles. *Atmos. Chem. Phys.,* 12:2691–2712, doi:10.5194/acp-12-2691-2012.

Subramanian, R., Khlystov, A. Y., Cabada, J. C., and Robinson, A. L. (2004). Positive and Negative Artifacts in Particulate Organic Carbon Measurements with Denuded and Undenuded Sampler Configurations. *Aerosol Sci. Technol.,* 38:27–48, doi:10.1080/02786820390229354.

Takahama, S., Johnson, A., Morales, J. G., Russell, L., Duran, R., Rodriguez, G., et al. (2012). Submicron Organic Aerosol in Tijuana, Mexico, from Local and Southern California Sources During the CalMex Campaign. *Atmos. Environ.* , doi:10.1016/j.atmosenv.2012.07.057.

Takahama, S., Schwartz, R. E., Russell, L. M., Macdonald, A. M., Sharma, S., and Leaitch, W. R. (2011). 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, doi:10.5194/acp-11-6367-2011.

Tervahattu, H., Juhanoja, J., and Kupiainen, K. (2002). Identification of an Organic Coating on Marine Aerosol Particles by TOF-SIMS. *J. Geophys. Res.-Atmos.,* 107(D19):1319, doi:10.1029/2001JD001403.

Topping, D. O., McFiggans, G. B., and Coe, H. (2005). A Curved Multi-Component Aerosol Hygroscopicity Model Framework: Part 2—Including Organic Compounds. *Atmos. Chem. Phys.,* 5:1223–1242.

Wasa, T., and Musha, S. (1970). Polarographic Behavior of Glyoxal And Its Related Compounds. *Bull. Univ. Osaka Prefecture. Ser. A, Eng. Nat. Sci.,* 19:169–180.

Wiberg, K. B., Morgan, K. M., and Maltz, H. (1994). Thermochemistry of Carbonyl Reactions. 6. A Study of Hydration Equilibria. *J. Am. Chem. Soc.,* 116:11 067–11 077, doi:10.1021/ja00103a024.

Williams, B. J., Goldstein, A. H., Kreisberg, N. M., and Hering, S. V. (2006). An In-Situ Instrument for Speciated Organic Composition of Atmospheric Aerosols: Thermal Desorption Aerosol GC/MS-FID (TAG). *Aerosol Sci. Technol.,* 40:627–638, doi:10.1080/02786820600754631.

Yu, J. Z., Griffin, R. J., Cocker, D. R., Flagan, R. C., Seinfeld, J. H., and Blanchard, P. (1999). Observation of Gaseous and Particulate Products of Monoterpene Oxidation in Forest Atmospheres. *Geophys. Res. Lett.,* 26:1145–1148, doi:10.1029/1999GL900169.

Zhao, J., Eisele, F. L., Titcombe, M., Kuang, C., and McMurry, P. H. (2010). Chemical Ionization Mass Spectrometric Measurements of Atmospheric Neutral Clusters using the Cluster-CIMS. *J. Geophys. Res.– Atmos.,* 115: doi:10.1029/2009JD012606.

Zuend, A., Marcolli, C., Booth, A. M., Lienhard, D. M., Soonsin, V., Krieger, U. K., et al. (2011). New and Extended Parameterization of the Thermodynamic Model AIOMFAC: Calculation of Activity Coefficients for organic-inorganic mixtures containing carboxyl, hydroxyl, carbonyl, ether, ester, alkenyl, alkyl, and Aromatic Functional Groups. *Atmos. Chem. Phys.,* 11:9155–9206, doi:10.5194/acp-11-9155-2011.

Zuend, A., Marcolli, C., Luo, B. P., and Peter, T. (2008). A Thermodynamic Model of Mixed Organic-Inorganic Aerosols to Predict Activity Coefficients. *Atmos. Chem. Phys.,* 8:4559–4593.

Zuend, A., Marcolli, C., Peter, T., and Seinfeld, J. H. (2010). Computation of Liquid-Liquid Equilibria and Phase Stabilities: Implications for RH-Dependent Gas/Particle Partitioning of Organic-Inorganic Aerosols. *Atmos. Chem. Phys.,* 10:7795–7820, doi:10.5194/acp-10-7795-2010.

Zuend, A., and Seinfeld, J. H. (2012). Modeling the Gas-Particle Partitioning of Secondary Organic Aerosol: The Importance of Liquid-Liquid Phase Separation. *Atmos. Chem. Phys.,* 12:3857–3882, doi:10.5194/acp-12-3857-2012.

Downloaded by [EPFL Bibliothèque] at 02:08 04 February 2013

Supplemental Information for "Quantification of carboxylic acid and carbonyl functional groups in organic aerosol infrared absorbance spectra"

S. Takahama, A. Johnson, L. M. Russell

S1 Background corrections

In a few cases, we apply a non-negativity constraint on the baseline fitting. For a baseline modeled as $y = f_\beta(x) + \epsilon$ with the base-case assumption of centered disturbances ($E[\epsilon] = 0$), we find an unconstrained least-squares solution over the region, $\mathcal{W}_n$. Let $e_j|_\beta = s(\tilde{\nu}_j) - f_\beta(\tilde{\nu}_j)$, and β is a vector of regression coefficients:

$$\beta = \arg\min_\beta \sum_{j \in \mathcal{W}_m} (e_j|_\beta)^2 \ .$$

When $E[\epsilon] \geq 0$, non-negativity constraints in the form of a logarithmic barrier (Arora, 2004) are applied:

$$\beta = \arg\min_\beta \textstyle\sum_{j \in \mathcal{W}_m} (e_j|_\beta)^2 - \alpha \sum_{j \in \mathcal{W}_n} P(e_j|_\beta) \ ,$$

where

$$P(e_j) = \begin{cases} log(e_j) & \text{if } e_j > \phi \\ log(\phi) & \text{otherwise} \end{cases} \ .$$

P effectively defines the penalty function; ϕ is chosen to be the machine precision value (smallest value for which $1+\phi \neq 1$). α is a strength parameter. In the implementation of this algorithm, we initialize parameter estimates from ordinary (nonlinear) least squares (a self-starting approach), and iteratively increase α until the non-negativity constraint is met.

Additionally, several other methods are used to remove undesired disturbances from the signal which interfere with baselining or band-fitting (as illustrated in Figure 3). Frequently, there is a high-frequency signal from water vapor between 3800 and 3600 cm^{-1} remaining after nitrogen purging, and this component is highly asymmetric (about the "true" baseline) depending on whether the water vapor was greater during scanning prior to or after sample collection. As this component can impact the fitting of the polynomial baseline (Section 2.1.1), this interference is removed by replacing the actual absorbance greater than 3500 cm^{-1} with a 3rd-order polynomial fitted to the 50 highest [if $s(\tilde{\nu} = 3748\,\mathrm{cm}^{-1}) < 0$] or lowest (otherwise case) absorbances over the region 3550 and 3900 cm^{-1} (approximately comprising 175 points) prior to the baselining. Contributions to absorbance from high-frequency water vapor disturbances which can disrupt baselining of the carbonyl region (between 1800 and 1600 cm^{-1}) are decreased by adjusting wavelet coefficients obtained by Daubechies 8-tap filter decomposition, using hard thresholds of 10^2 for first 4 levels and 10^{-5} for the last 4 (this functionality is provided by the `wavShrink` function of the `wmtsa` library, implemented in the R statistical package; R Development Core Team, 2012). Absorption of CO_2 occurs at approximately 2300 cm^{-1}, and this peak is removed by interpolating between 2420 and 2280 cm^{-1} with a cubic spline.

The different methods are used in accordance with the structure of the interference. Wavelet decomposition is used for the 1800-1600 cm^{-1} region as the low-frequency component varies less predictably than for the region between 3800 and 3600 cm^{-1}; the former region is less amenable for the same biased smoothing approach. If the biased smoothing is not applied for the region between 3800 and 3600 cm^{-1}, the regression

would be influenced by the presence of the water vapor and the line would be fitted through all of the high-frequency values, rather than over or under in accordance with the smoothed baseline. Since the baseline at 1800-1600 cm^{-1} is defined by a line fitted through two points, we apply the wavelet decomposition to ensure that the two points are not extreme with respect to the low-frequency signal. Smoothing for CO_2 seldom affects the quantification and is largely cosmetic, but can affect the baselining if b_r (Equation 1) is moved toward 2300 cm^{-1}.

S2 Band-fitting

In this section, we briefly expand on the explanation of the combination and sequence of fitted lineshapes for the region approximately between 3600 and 2400 cm^{-1} as described in Section 2.1.2. As the carboxylic COH absorption band is quite broad (primarily between 3400 and 2400 cm^{-1}) and overlaps with the other lineshapes, it is difficult to determine a priori how much of the absorption should be apportioned to alcohol COH, ammonium NH, and alkane CH in these regions. Based on evaluation of absorbing species below 2600 cm^{-1} (Lambert et al., 1998), we conclude that carboxylic COH is an approximately uniquely absorbing group in this region for atmospheric aerosol samples. Therefore, we use this region to scale up a fixed carboxylic COH lineshape (Section 2.1.3). Next, the absorption profile of ammonium NH is scaled and subtracted following the protocol of Gilardoni et al. (2007). Gaussian peaks of alcohol COH (two peaks), alkane CH (four peaks), and the rest (alkene CH, aromatic CH, and four unidentified peaks) are then fitted, in this sequence (the order of some do not matter as they do not affect the peak-fitting of other bonds). Gaussian peaks for carbonyl C=O and amine CNH_2 are also fitted together but not in sequence with the listed peaks above, as their absorbance regions are isolated from the other peaks. As reported by Russell et al. (2009); Hawkins and Russell (2010); Schwartz et al. (2010); and Takahama et al. (2011), alkene and aromatic bonds, which occur as sharp absorption bands or peaks, are seldom observed in ambient samples and are estimated to be below detection limit. While these peaks preclude geometrical detection, in practice, there are four additional ("Unidentified") peaks (Table 3 and 6) which are fitted simultaneously with the alkene and aromatic peaks, as it is evident that there are co-absorption by other analytes in the region between 3050 and 2950 cm^{-1} which account for the broad absorption bands in the region unapportioned by our algorithm. In a similar method to the carboxylic COH shape derivation (Section 2.1.3), the additional components were derived from a PMF analysis on the residuals in this region after the fitting in Section 2.1.2. The total area of these identified peaks are generally $\ll$10% of the total integrated area of the baselined region between 3800 and 1500 cm^{-1}, but prevents overestimation of alkene and aromatic compounds in the absence of sharp, characteristic peaks. Primary amine CNH_2 groups are also reported to exhibit stretching-mode absorbances between 3550 and 3250 cm^{-1} (Lambert et al., 1998), but are not considered in our algorithm at this time. This may lead to an overestimation of alcohol COH functional groups, but would be dependent on the absorptivity and absorption profiles of amine CNH_2 for various compounds. For instance, the absorption profile for L- and D-alanine show very little absorbance in this region, whereas tridecylamine and tetradecylamine show sharp lineshapes [e.g., as can be viewed in Spectral Database for Organic Compounds (SDBS); http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi] in contrast to the broad features of the alcohol COH band in the region. As shown by examples of ambient spectra in Figure 9, absorbances between 3550 and 3250 cm^{-1} do not contain sharp features indicative of high molar absorptivities that are interfering with quantification of alcohol COH (this observation also applies for other samples not included in the Figure), but this may be an issue to be considered in future studies. As the abundance of amine CNH_2 is quantified by absorption in the region between 1850 and 1500 cm^{-1} in our algorithm, amine CNH_2 absorbances at higher wavenumbers will not affect our quantification, though considering them in future analyses may increase the robustness of our estimates.

S3 Separation of oxygenated fraction

The infrared spectra published by the Coblentz Society (from NIST Chemistry WebBook spectra are taken) vary in preparation, concentration, phase state, spectral resolution, and transmission levels (http://webbook.nist.gov/chemistry/coblentz/). The assumption invoked is that these differences across samples and spectra are smaller than the overwhelming variations imposed by differences in the COOH to CH functional group ratios. The spectra and descriptions regarding their sample collection is listed in Table S1. Transmission spectra are converted to absorbance by taking the negative logarithm; the incident radiation (I_0) is effectively taken as the maximum value of transmission over the spectra window of 3800 and 2200 cm^{-1}. The spectra were scaled such that the mean of absorbances at 3100 and 2600 cm^{-1} are approximately unity for Figure 4a, though the Positive Matrix Factorization (PMF; Paatero and Tapper, 1994) algorithm for solving the factor analytic problem in chemometrics does not actually require scaling of samples or spectra a priori (Paatero and Tapper, 1994), as the the scaling is reflected in the factor strengths.

PMF is one method which prescribes the solution according to non-negativity constraints in the component spectra (factors) and their contributions, subject to weighting of sample and variable by uncertainties (in this analysis, we assume uniform uncertainties). PMF has been previously used for the interpretation of atmospheric particulate matter concentrations (e.g., Lee et al., 1999), precipitation composition (Juntto and Paatero, 1994), meteorological variables (e.g., Paatero and Hopke, 2002), size distributions (e.g., Zhou et al., 2005), mass fragment spectra (e.g., Lanz et al., 2007; Ulbrich et al., 2009), X-ray absorption spectra (e.g., Liu et al., 2009; Takahama et al., 2010), and FTIR spectra (e.g., Russell et al., 2009, 2011). There are several degrees of freedom which must be specified by the user in application of PMF: the number of factors, the rotational parameter (FPEAK), and seed value. Using various seed values does not appear to yield additional unique solutions for infrared spectra, so we focus on the selection of the first two parameters. As stated above, the variation among spectra in the selected molecules presumably arise from the different proportions of CH and COH bonds present in each molecule (Section S3). A two-factor solution is thus selected, with the expectation that each factor represents one of these two components. The rotational parameter, FPEAK, is a single parameter used to control the linear combinations which can be constructed to represent the possible solutions (Paatero et al., 2002). While there is a potentially infinite combination of profile and strength matrices which can satisfy the factor analytic expression, we find two reocurring solutions which are interpreted as described in Section 2.1.3.

From examination of our raw PMF solutions, we observe that the ratio of factor strengths of components identified as COH to CH does not monotonically increase with respect to the true molar ratio of COH to CH in the molecules selected. There is a clear bifurcation in the relationship; for instance, for the FPEAK=-0.2 solution shown in Figure 4b, the overall correlation between COH/CH factor and bond ratios is $r = 0.57$, while the subset consisting of glutaric, succinic, and malonic acids have a correlation of $r = 0.99$ and the complimentary set consisting of stearic, decanoic, azelaic, adipic, and acetic acids have a correlation of $r = 0.98$. While the two sets are highly linear, the slopes of factor ratios regressed against bond ratios vary by a factor of 3.5. This may reflect the variations from interferences or concentrations in the Coblentz Society spectra, but may also be attributed to the combination of true variations in absorption profiles or absorptivities among these compounds (the sample preparation methods listed in Table S1 do not indicate that systematic differences should necessarily be observed according to this grouping). We have included both sets of spectra in our PMF decomposition in the hopes of obtaining a solution which corresponds to the average absorption profile of carboxylic COH across these molecules, though there is room for further evaluation and refinement.

The solutions generated by PMF along a continuum of FPEAK values correspond to a least squares solution of scaled distances (between the actual and reconstructed spectra) which satisfy the standard non-negativity constraints. For our task, we require two additional constraints that the COH profile be smooth (i.e., not discontinuous over the range of absorption), and that the profile is relevant for functional group quantification in our aerosol samples. While an algorithm can be constructed to impose all of these constraints during the iterative search of the component strength and profiles (Jiang and Ozaki, 2002; de Juan and Tauler, 2006), we instead use PMF as a readily-available tool to generate solutions from which we obtain a carboxylic COH profile and matching absorptivity which satisfies our additional constraints. Therefore, our

selected PMF factors are re-expressed as a pair of Gaussian peaks to ensure smoothness, and solutions are superposed through an adjustable degree of freedom to find an absorption profile and absorptivity which provides suitable quantification of molar abundances in our samples (Section 2.1.3).

The Coblentz Society spectra archive was selected on the merit that a large number of spectra could be consolidated with minimal material and labor investment. Addressing the issue of purity, sample prepration, and concentration range of samples for the initial PMF analysis would presumably make the range of solutions more directly suitable for our aerosol sample spectra with minimal modification. But in the processes of imposing additional constraints as discussed above, the systematic biases or errors that may have existed in the selection of the COH profile and absorptivity do not appear to be strongly retained. As reported by our manuscript, the final carboxylic COH profile (with calibrated absorptivities) can reproduce actual molar abundance of carboxylic COH in our samples within 21%, and with less than 5% bias on average (Section 3.2 and Table 5).

S4 Experimental methods

For FTIR analysis, interference of liquid water on organic functional group quantification is considered to be negligible in our analysis. Liquid-phase water is in principle observable by FTIR (e.g., Bertram et al., 1996; Martin et al., 1997; Liu et al., 2008); the OH stretching (with overtones of bending) band for water is centered at $3404\,\mathrm{cm}^{-1}$ contains multiple components with the highest mode appearing at $3600\,\mathrm{cm}^{-1}$ at 25°C (Venyaminov and Prendergast, 1997). In aerosol spectra, an effective absorbing peak from liquid water is reported at $3450\,\mathrm{cm}^{-1}$ (Cziczo and Abbatt, 2000) when present. This band is not observed in our sample spectra because the molar absorptiviy of water is low in comparison to organic analytes (Venyaminov and Prendergast, 1997), but also and primarily for the reason that the water is presumably evaporated during the nitrogen purge period of the sample compartment headspace prior to scan acquisition. An extensive study of hydrates in laboratory and ambient spectra (Frossard and Russell, 2012) finds that this form of bound water is also not likely to interfere in organic functional group quantification in most spectra of dried ambient submicron particles.

For gravimetric analysis, aerosol water is absent in organic acids standards after drying with the diffusion dryer, as the efflorescence RH of organic acids studied in our laboratory experiments are between 34-85% RH (Chan et al., 2008), with deliquescence RHs either not observed or greater than 90% RH. The hygroscopicity of aerosols containing alcohols are generally less than that for organic acids (Hemming and Seinfeld, 2001), but their efflorescence and deliquescence RHs are less well known. The lack of evidence of diol dehydration into carbonyls (Sections 3.2 and 3.3) suggests that some water may be retained by these aerosols, but agreement of measured and predicted moles of functional groups for alcohol compounds (Section 3.2) suggests that the quantity is small.

Our reported form of functional group absorptivity is derived from the Beer-Lambert Law, and its origins are described here. A standard expression for the Beer-Lambert Law (Atkins, 1998) for an absorbing group i at wavenumber $\tilde{\nu}$ in terms of absorbance, $A_i'(\tilde{\nu})$, is

$$A_i'(\tilde{\nu}) = \epsilon_i'(\tilde{\nu}) c_i l \,,$$

where $\epsilon_i'(\tilde{\nu})$ is the wavenumber-dependent molar absorption coefficient, c_i is the molar concentration, and l is the path length through the sample. As the absorption bands of groups considered in our study are broad, the integral over each band rather than the absorbance at a single wavenumber is used for molar quantification of our functional groups (Atkins, 1998). The relationship with integrated absorbance, $A_i = \int A_i'(\tilde{\nu}) d\tilde{\nu}$, and the integrated absorption coefficient, $\epsilon_i = \int \epsilon_i'(\tilde{\nu}) d\tilde{\nu}$, is $A_i = \epsilon_i c_i l$. For n_i moles of the absorbing group accumulated uniformly (Maria et al., 2003) over area a_m and thickness l on our substrate, $c_i = n_i/(a_m l)$. If we let $\epsilon_i^* = \epsilon_i / a_m$, we arrive at the expression,

$$A_i = \epsilon_i^* n_i \,.$$

The parameter ϵ_i^* is the slope obtained from our linear regressions (with no intercept) in Section 3.1 and is referred to as the absorptivity, following the convention of Maria et al. (2003). The standard integrated

absorption coefficient, ϵ_i, can be calculated by multiplying our reported absorptivities by the sample collection area. This area is $a_m = \pi/4 \cdot (10\,\mathrm{mm})^2$ for our method, as described in Section 2.2.

S5 Limits of Detection

Detection limits reported by Russell et al. (2009) were determined by considering the (two-sigma) variability in absorbance for each functional group region after baselining of blank sample spectra. As our apportionment of absorbance is rather conservative (we are generally underbiased in apportioning area for each functional group), the abundance of absent functional groups can be over-reported if spectroscopically-adjacent functional groups are present in the sample. The presence of all functional groups are often assumed for ambient samples and fitted according to Table 3, though we apply a priori knowledge of the components in the system to guide the fitting in our laboratory sample spectra.

The influence of neighboring absorption bands can be illustrated by considering that Type I errors (detecting analyte when there is none) can increase for amine CNH_2 if carbonyl C=O is present. None of our samples in this work contained amine CNH_2, but several reference compounds studied contained carbonyl C=O bonds (succinic acid, pinonic acid, aqueous mixture of succinic acid and 2-deoxy-d-ribose, azelaic acid); carbonyl C=O and amine CNH_2 are spectroscopically adjacent and have overlapping absorption bands. The 95th percentile of quantified amine CNH_2 using the fitting specified by Table 3 for these spectra is 0.22 μmole. In contrast, the 95th percentile value for compounds that do not contain interfering carbonyl C=O bonds (docosanol) or aldehydes expected to hydrate and form alcohols (2-deoxy-d-ribose in aqueous solution, 2-deoxy-d-ribose in ethanol, glyceraldehyde in aqueous solution) is 0.01 μmole using the same algorithm, in accordance with the reported value by Russell et al. (2009). This effect can also be expected for functional groups with broad, overlapping absorbance profiles, such as alcohol and carboxylic COH. Ninety-fifth percentile thresholds for quantified abundances of absent functional groups in the presence of functional groups with overlapping bands are reported in Table S2. The reason for higher limits of detection (relative to Russell et al., 2009) discussed above is supported by correlations of quantified abundance of absent functional groups with actual moles of spectroscopically-adjacent functional groups actually present in the samples (Table S2).

The values in Table S2 do not necessarily determine limits for quantification, however, and reflect the extreme cases for which these threshold were calculated. For example, the area near $2600\,\mathrm{cm}^{-1}$ in broadly-absorbing samples containing alcohol COH groups (e.g., Figure 7d) is integrated as carboxylic COH if this group is assumed to be present and scaled according to our algorithm (Section 2.1.3). This type of misapportionment of absorbance leads to a quantified abundance (upper-limit) of 0.18 μmol for carboxylic COH (in samples lacking carboxylic groups) reported in Table S2. However, if we examine our capability for quantification of carboxylic COH in pinonic acid — which contains carboxylic groups — the linearity in response is apparent (r=1.00) even below the $0.18\,\mu$mol threshold (Figure 5d). There is a -35% systematic bias in the prediction of carboxylic COH for reasons suggested in Section 3.2, but this error is not due to random variability limiting the quantification below this threshold. As such, the thresholds in Table S2 differ in interpretation from conventional analytical detection limits. As quantification of functional group abundance is still possible below these values, some consideration must precede their application to various sample types.

S6 Molecular bonds to atomic composition and mass concentrations

While the main body of the manuscript focuses on the apportionment and quantification of carboxylic COOH and carbonyl C=O, the framework used for mass quantification is presented here for completeness. If we let $\mathbf{n}_{\mathcal{A}} = [n_i]_{i\in\mathcal{A}}$, a vector of moles of atom i in set of atoms $\mathcal{A} = \{\mathrm{C, O, H, N, S}\}$, and $\mathbf{n}_{\mathcal{G}} = [n_j]_{j\in\mathcal{G}}$, a vector of moles of functional group (FG) j in set of groups $\mathcal{G} = \{$aCOH, rCH, eCH, aCH, COOH, C=O, CNH_2, $CONO_2$, $COSO_3\}$, we can relate the two quantities by a weighted apportionment matrix, $\bar{\mathbf{\Lambda}} = \left[\bar{\lambda}_{ij}\right]_{i\in\mathcal{A},j\in\mathcal{G}}$ (Equation 8). The

convention of using lowercase letters used to prefix bond names for elements of $\mathcal{G}$ is adopted from Russell (2003) and is interpreted as follows: for COH, a = alcohol; for CH, r = aromatic, e = alkene, and a = alkane. Russell (2003) defined values for $\bar{\lambda}_{ij}$ (shown in Equation S2) based on an analysis of the functional group index (*FGI*; Table S3), which is a carbon-chain analogue of the bulk functionalization rate (Dron et al., 2010) and represents the ratio of functional groups to carbon atoms in the chain on a mole-basis. We also explicitly include the degree of saturation by alkane CH in enumerating the possibilities for the functional group configurations. The canonical apportionment matrix $\bar{\mathbf{\Lambda}}$ for ambient samples is shown in Equation S2. Given $\mathbf{n}_{\mathcal{A}} = \bar{\mathbf{\Lambda}}\mathbf{n}_{\mathcal{G}}$ (Equation 8), each array is defined as

$$\mathbf{n}_{\mathcal{A}} = \begin{pmatrix} n_{\mathsf{C}} \\ n_{\mathsf{O}} \\ n_{\mathsf{H}} \\ n_{\mathsf{N}} \\ n_{\mathsf{S}} \end{pmatrix}, \ \mathbf{n}_{\mathcal{G}} = \begin{pmatrix} n_{\mathsf{aCOH}} \\ n_{\mathsf{rCH}} \\ n_{\mathsf{eCH}} \\ n_{\mathsf{aCH}} \\ n_{\mathsf{COOH}} \\ n_{\mathsf{C{=}O}} \\ n_{\mathsf{CNH2}} \\ n_{\mathsf{CONO2}} \\ n_{\mathsf{COSO3}} \end{pmatrix}, \text{ and } \bar{\mathbf{\Lambda}} = \begin{pmatrix} 0.5 & 1 & 1 & 0.5 & 1 & 1 & 0.25 & 0.5 & 0.5 \\ 1 & 0 & 0 & 0 & 2 & 1 & 0 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{S1}$$

As summarized by Russell et al. (2011), alkene and aromatic compounds are estimated to contribute 5-10% of OM mass (the method for estimating their contribution to the absorption spectrum is described in Section S2); given their high limit of detection these components are often removed from $\mathbf{n}_{\mathcal{G}}$ and $\bar{\mathbf{\Lambda}}$ for the calculation of total OM in these cases. Elements of $\bar{\mathbf{\Lambda}}$ can be adjusted to account for knowledge about the saturation and functionality for known compounds in the sample, or by measurements provided by other instrumentation. For instance, a global parameter was introduced into this apportionment matrix by Chhabra et al. (2011) to constrain variations in *FGI* values between the canonical case (Equation S2) with those derived from the molecular structure of the parent, precursor compound for interpretation of spectra of secondary organic aerosol produced in the Caltech environmental chambers.

As discussed by Russell (2003), the aforementioned FTIR method misses and underestimates carbon content in highly-branched compounds. However, the magnitude of this underestimation may not be significant for many ambient measurements, as the OM reported by this approach is often in reasonable agreement with other analytical methods (each with their own set of biases), such as AMS (e.g., Russell et al., 2009; Gilardoni et al., 2009; Frossard et al., 2011; Liu et al., 2011). However, when additional information is provided or assumed, these can be augmented by adding unknown quantities on a proportional basis — e.g., $n_{\mathsf{O,revised}} = n_{\mathsf{O}} + (n_{\mathsf{COC}}/n_{\mathsf{C}})\, n_{\mathsf{C}}$ where n_{COC} is the assumed number of ether bonds (Chhabra et al., 2011). Such augmentations can be performed by prepending an operator to $\bar{\mathbf{\Lambda}}$; these operators will in most cases be an altered identity matrix. For example, letting $\mathbf{P}$ denote a prefix operator which assumes a fixed number of ether groups per carbon atom,

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ n_{\mathsf{COC}}/n_{\mathsf{C}} & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The full product would then be written as $\mathbf{n}_{\mathcal{A}} = \mathbf{P}\bar{\mathbf{\Lambda}}\mathbf{n}_{\mathcal{G}}$. Finally, we can use $\mathbf{n}_{\mathcal{G}}$ and $\mathbf{n}_{\mathcal{A}}$ to calculate various molar ratios and mass quantities, including mass, H/C ratio, O/C ratio, and functionalization rate (Dron et al., 2010). Let $\mathbf{m}_{\mathcal{A}} = [m_i]_{i\in\mathcal{A}}$, a vector of atomic mass units for atom i, then

$$\begin{aligned} \text{O/C atomic ratio} &= n_{\mathsf{O}}/n_{\mathsf{C}}, \\ \text{H/C atomic ratio} &= n_{\mathsf{H}}/n_{\mathsf{C}}, \\ \text{FG}_j \text{ functionalization rate} &= n_j/n_{\mathsf{C}} \ \forall\, j \in \mathcal{G}, \\ \text{FG}_j \text{ mass} &= \sum_{i\in\mathcal{A}} m_i \bar{\lambda}_{ij} n_j \ \forall\, j \in \mathcal{G}, \\ \text{OC mass} &= m_{\mathsf{C}} n_{\mathsf{C}}, \text{ and} \\ \text{OM mass} &= \mathbf{m}_{\mathcal{A}}^T \mathbf{n}_{\mathcal{A}} \ . \end{aligned}$$

While the main body of the manuscript focuses on the apportionment and quantification of carboxylic COOH and carbonyl C=O, the framework used for mass quantification is presented here for completeness. If we let $\mathbf{n}_{\mathcal{A}} = [n_i]_{i\in\mathcal{A}}$,

a vector of moles of atom i in set of atoms $\mathcal{A} = \{\text{C, O, H, N, S}\}$, and $\mathbf{n}_\mathcal{G} = [n_j]_{j\in\mathcal{G}}$, a vector of moles of functional group (FG) j in set of groups $\mathcal{G} = \{\text{aCOH, rCH, eCH, aCH, COOH, C=O, CNH}_2\text{, CONO}_2\text{, COSO}_3\}$, we can relate the two quantities by a weighted apportionment matrix, $\bar{\mathbf{\Lambda}} = \left[\bar{\lambda}_{ij}\right]_{i\in\mathcal{A},j\in\mathcal{G}}$ (Equation 8). The convention of using lowercase letters used to prefix bond names for elements of $\mathcal{G}$ is adopted from Russell (2003) and is interpreted as follows: for COH, a = alcohol; for CH, r = aromatic, e = alkene, and a = alkane. Russell (2003) defined values for $\bar{\lambda}_{ij}$ (shown in Equation S2) based on an analysis of the functional group index (*FGI*; Table S3), which is a carbon-chain analogue of the bulk functionalization rate (Dron et al., 2010) and represents the ratio of functional groups to carbon atoms in the chain on a mole-basis. We also explicitly include the degree of saturation by alkane CH in enumerating the possibilities for the functional group configurations. The canonical apportionment matrix $\bar{\mathbf{\Lambda}}$ for ambient samples is shown in Equation S2. Given $\mathbf{n}_\mathcal{A} = \bar{\mathbf{\Lambda}}\mathbf{n}_\mathcal{G}$ (Equation 8), each array is defined as

$$\mathbf{n}_\mathcal{A} = \begin{pmatrix} n_\mathsf{C} \\ n_\mathsf{O} \\ n_\mathsf{H} \\ n_\mathsf{N} \\ n_\mathsf{S} \end{pmatrix}, \ \mathbf{n}_\mathcal{G} = \begin{pmatrix} n_\mathsf{aCOH} \\ n_\mathsf{rCH} \\ n_\mathsf{eCH} \\ n_\mathsf{aCH} \\ n_\mathsf{COOH} \\ n_\mathsf{C{=}O} \\ n_\mathsf{CNH2} \\ n_\mathsf{CONO2} \\ n_\mathsf{COSO3} \end{pmatrix}, \text{ and } \bar{\mathbf{\Lambda}} = \begin{pmatrix} 0.5 & 1 & 1 & 0.5 & 1 & 1 & 0.25 & 0.5 & 0.5 \\ 1 & 0 & 0 & 0 & 2 & 1 & 0 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{S2}$$

As summarized by Russell et al. (2011), alkene and aromatic compounds are estimated to contribute 5-10% of OM mass (the method for estimating their contribution to the absorption spectrum is described in Section S2); given their high limit of detection these components are often removed from $\mathbf{n}_\mathcal{G}$ and $\bar{\mathbf{\Lambda}}$ for the calculation of total OM in these cases. Elements of $\bar{\mathbf{\Lambda}}$ can be adjusted to account for knowledge about the saturation and functionality for known compounds in the sample, or by measurements provided by other instrumentation. For instance, a global parameter was introduced into this apportionment matrix by Chhabra et al. (2011) to constrain variations in *FGI* values between the canonical case (Equation S2) with those derived from the molecular structure of the parent, precursor compound for interpretation of spectra of secondary organic aerosol produced in the Caltech environmental chambers.

As discussed by Russell (2003), the aforementioned FTIR method misses and underestimates carbon content in highly-branched compounds. However, the magnitude of this underestimation may not be significant for many ambient measurements, as the OM reported by this approach is often in reasonable agreement with other analytical methods (each with their own set of biases), such as AMS (e.g., Russell et al., 2009; Gilardoni et al., 2009; Frossard et al., 2011; Liu et al., 2011). However, when additional information is provided or assumed, these can be augmented by adding unknown quantities on a proportional basis — e.g., $n_{\mathsf{O,revised}} = n_\mathsf{O} + (n_\mathsf{COC}/n_\mathsf{C})\, n_\mathsf{C}$ where n_COC is the assumed number of ether bonds (Chhabra et al., 2011). Such augmentations can be performed by prepending an operator to $\bar{\mathbf{\Lambda}}$; these operators will in most cases be an altered identity matrix. For example, letting $\mathbf{P}$ denote a prefix operator which assumes a fixed number of ether groups per carbon atom,

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ n_\mathsf{COC}/n_\mathsf{C} & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The full product would then be written as $\mathbf{n}_\mathcal{A} = \mathbf{P}\bar{\mathbf{\Lambda}}\mathbf{n}_\mathcal{G}$. Finally, we can use $\mathbf{n}_\mathcal{G}$ and $\mathbf{n}_\mathcal{A}$ to calculate various molar ratios and mass quantities, including mass, H/C ratio, O/C ratio, and functionalization rate (Dron et al., 2010). Let $\mathbf{m}_\mathcal{A} = [m_i]_{i\in\mathcal{A}}$, a vector of atomic mass units for atom i, then

$$\begin{aligned} \text{O/C atomic ratio} &= n_\mathsf{O}/n_\mathsf{C}, \\ \text{H/C atomic ratio} &= n_\mathsf{H}/n_\mathsf{C}, \\ \text{FG}_j \text{ functionalization rate} &= n_j/n_\mathsf{C} \ \forall\, j \in \mathcal{G}, \\ \text{FG}_j \text{ mass} &= \sum_{i\in\mathcal{A}} m_i \bar{\lambda}_{ij} n_j \ \forall\, j \in \mathcal{G}, \\ \text{OC mass} &= m_\mathsf{C} n_\mathsf{C}, \text{ and} \\ \text{OM mass} &= \mathbf{m}_\mathcal{A}^T \mathbf{n}_\mathcal{A}\ . \end{aligned}$$

References

Arora, J. (2004). *Introduction to Optimum Design*, Elsevier Academic Press, 2nd edn.

Atkins, P. (1998). *Physical Chemistry*, W. H. Freeman and Company, New York, 6th edn.

Bertram, A. K., Patterson, D. D., and Sloan, J. J. (1996). Mechanisms and temperatures for the freezing of sulfuric acid aerosols measured by FTIR extinction spectroscopy, *Journal of Physical Chemistry*, 100:2376–2383, doi: 10.1021/jp952551v.

Chan, M. N., Kreidenweis, S. M., and Chan, C. K. (2008). Measurements of the hygroscopic and deliquescence properties of organic compounds of different solubilities in water and their relationship with cloud condensation nuclei activities, *Environmental Science & Technology*, 42:3602–3608, doi:10.1021/es7023252.

Chhabra, P. S., Ng, N. L., Canagaratna, M. R., Corrigan, A. L., Russell, L. M., Worsnop, D. R., Flagan, R. C., and Seinfeld, J. H. (2011). Elemental composition and oxidation of chamber organic aerosol, *Atmospheric Chemistry and Physics*, 11:8827–8845, doi:10.5194/acp-11-8827-2011.

Cziczo, D. J. and Abbatt, J. P. D. (2000). Infrared observations of the response of NaCl, MgCl2, NH4HSO4, and NH4NO3 aerosols to changes in relative humidity from 298 to 238 K, *Journal of Physical Chemistry A*, 104:2038–2047, doi:10.1021/jp9931408.

de Juan, A. and Tauler, R. (2006). Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications, *Critical Reviews in Analytical Chemistry*, 36:163–176, doi:10.1080/10408340600970005.

Dron, J., El Haddad, I., Temime-Roussel, B., Jaffrezo, J.-L., Wortham, H., and Marchand, N. (2010). Functional group composition of ambient and source organic aerosols determined by tandem mass spectrometry, *Atmospheric Chemistry and Physics*, 10:7041–7055, doi:10.5194/acp-10-7041-2010.

Frossard, A. A. and Russell, L. M. (2012). Removal of Sea Salt Hydrate Water from Seawater- Derived Samples by Dehydration, *submitted*.

Frossard, A. A., Shaw, P. M., Russell, L. M., Kroll, J. H., Canagaratna, M. R., Worsnop, D. R., Quinn, P. K., and Bates, T. S. (2011). Springtime Arctic haze contributions of submicron organic particles from European and Asian combustion sources, *Journal of Geophysical Research-Atmospheres*, 116, doi:10.1029/2010JD015178.

Gilardoni, S., Russell, L. M., Sorooshian, A., Flagan, R. C., Seinfeld, J. H., Bates, T. S., Quinn, P. K., Allan, J. D., Williams, B., Goldstein, A. H., Onasch, T. B., and Worsnop, D. R. (2007). Regional variation of organic functional groups in aerosol particles on four US east coast platforms during the International Consortium for Atmospheric Research on Transport and Transformation 2004 campaign, *Journal of Geophysical Research-Atmospheres*, 112, doi:10.1029/2006JD007737.

Gilardoni, S., Liu, S., Takahama, S., Russell, L. M., Allan, J. D., Steinbrecher, R., Jimenez, J. L., De Carlo, P. F., Dunlea, E. J., and Baumgardner, D. (2009). Characterization of organic ambient aerosol during MIRAGE 2006 on three platforms, *Atmospheric Chemistry and Physics*, 9:5417–5432.

Hawkins, L. N. and Russell, L. M. (2010). Oxidation of ketone groups in transported biomass burning aerosol from the 2008 Northern California Lightning Series fires, *Atmospheric Environment*, 44:4142–4154, doi: 10.1016/j.atmosenv.2010.07.036.

Hemming, B. L. and Seinfeld, J. H. (2001). On the Hygroscopic Behavior of Atmospheric Organic Aerosols, *Industrial & Engineering Chemistry Research*, 40:4162–4171, doi:10.1021/ie000790l.

Jiang, J. H. and Ozaki, Y. (2002). Self-modeling curve resolution (SMCR): Principles, techniques, and applications, *Applied Spectroscopy Reviews*, 37:321–345, doi:10.1081/ASR-120014359.

Juntto, S. and Paatero, P. (1994). Analysis of Daily Precipitation Data By Positive Matrix Factorization, *Environmetrics*, 5:127–144, doi:10.1002/env.3170050204.

Lambert, J. B., Shurvell, H. F., Lightner, D., and Cooks, R. G. (1998). *Organic Structural Spectroscopy*, Prentice Hall.

Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prevot, A. S. H. (2007). Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra, *Atmospheric Chemistry and Physics*, 7:1503–1522.

Lee, E., Chan, C. K., and Paatero, P. (1999). Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong, *Atmospheric Environment*, 33:3201–3212, doi:10.1016/S1352-2310(99)00113-2.

Liu, L., Mishchenko, M. I., and Arnott, W. P. (2008). A study of radiative properties of fractal soot aggregates using the superposition T-matrix method, *Journal of Quantitative Spectroscopy & Radiative Transfer*, 109:2656–2663, doi:10.1016/j.jqsrt.2008.05.001.

Liu, S., Takahama, S., Russell, L. M., Gilardoni, S., and Baumgardner, D. (2009). Oxygenated organic functional groups and their sources in single and submicron organic particles in MILAGRO 2006 campaign, *Atmospheric Chemistry and Physics*, 9:6849–6863.

Liu, S., Day, D. A., Shields, J. E., and Russell, L. M. (2011). Ozone-driven daytime formation of secondary organic aerosol containing carboxylic acid groups and alkane groups, *Atmospheric Chemistry and Physics*, 11:8321–8341, doi:10.5194/acp-11-8321-2011.

Maria, S. F., Russell, L. M., Turpin, B. J., Porcja, R. J., Campos, T. L., Weber, R. J., and Huebert, B. J. (2003). Source signatures of carbon monoxide and organic functional groups in Asian Pacific Regional Aerosol Characterization Experiment (ACE-Asia) submicron aerosol types, *Journal of Geophysical Research-Atmospheres*, 108, doi:10.1029/2003JD003703.

Martin, S. T., Salcedo, D., Molina, L. T., and Molina, M. J. (1997). Phase transformations of micron-sized H2SO4/H2O particles studied by infrared spectroscopy, *Journal of Physical Chemistry B*, 101:5307–5313, doi: 10.1021/jp970607a.

Paatero, P. and Hopke, P. K. (2002). Utilizing wind direction and wind speed as independent variables in multilinear receptor modeling studies, *Chemometrics and Intelligent Laboratory Systems*, 60:25–41, doi:10.1016/S0169-7439(01)00183-6.

Paatero, P. and Tapper, U. (1994). Positive Matrix Factorization - A Nonnegative Factor Model With Optimal Utilization of Error-estimates of Data Values, *Environmetrics*, 5:111–126.

Paatero, P., Hopke, P. K., Song, X. H., and Ramadan, Z. (2002). Understanding and controlling rotations in factor analytic models, *Chemometrics and Intelligent Laboratory Systems*, 60:253–264.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org/`, ISBN 3-900051-07-0.

Russell, L. M. (2003). Aerosol organic-mass-to-organic-carbon ratio measurements, *Environmental Science & Technology*, 37:2982–2987, doi:10.1021/es026123w.

Russell, L. M., Takahama, S., Liu, S., Hawkins, L. N., Covert, D. S., Quinn, P. K., and Bates, T. S. (2009). Oxygenated fraction and mass of organic aerosol from direct emission and atmospheric processing measured on the R/V Ronald Brown during TEXAQS/GoMACCS 2006, *Journal of Geophysical Research-Atmospheres*, 114, D00F05, doi:10.1029/2008JD011275.

Russell, L. M., Bahadur, R., and Ziemann, P. J. (2011). Identifying organic aerosol sources by comparing functional group composition in chamber and atmospheric particles, *Proceedings of the National Academy of Sciences of the United States of America*, 108:3516–3521, doi:10.1073/pnas.1006461108.

Schwartz, R. E., Russell, L. M., Sjostedt, S. J., Vlasenko, A., Slowik, J. G., Abbatt, J. P. D., Macdonald, A. M., Li, S. M., Liggio, J., Toom-Sauntry, D., and Leaitch, W. R. (2010). Biogenic oxidized organic functional groups in aerosol particles from a mountain forest site and their similarities to laboratory chamber products, *Atmospheric Chemistry and Physics*, 10:5075–5088, doi:10.5194/acp-10-5075-2010.

Takahama, S., Liu, S., and Russell, L. M. (2010). Coatings and clusters of carboxylic acids in carbon-containing atmospheric particles from spectromicroscopy and their implications for cloud-nucleating and optical properties, *Journal of Geophysical Research-Atmospheres*, 115, D01 202, doi:10.1029/2009JD012622.

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

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L. (2009). Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, *Atmospheric Chemistry and Physics*, 9:2891–2918.

Venyaminov, S. Y. and Prendergast, F. G. (1997). Water (H2O and D2O) Molar Absorptivity in the 10004000 cm1Range and Quantitative Infrared Spectroscopy of Aqueous Solutions, *Analytical Biochemistry*, 248:234 – 245, doi:10.1006/abio.1997.2136.

Zhou, L. M., Hopke, P. K., Stanier, C. O., Pandis, S. N., Ondov, J. M., and Pancras, J. P. (2005). Investigation of the relationship between chemical composition and size distribution of airborne particles by partial least squares and positive matrix factorization, *Journal of Geophysical Research-Atmospheres*, 110, D07S18, doi: 10.1029/2004JD005050.

Tables

Table S1: List of acid spectra used for PMF (from NIST Chemistry Webbook).

CAS	Compound	n_{C}	n_{cCOH}	n_{aCH}	Preparation
64-19-7	Acetic acid	2	1	3	solution (10% in CCl4)
141-82-2	Malonic acid	3	2	2	solid (Fluorolube mull)
110-15-6	Succinic acid	4	2	4	solid (Nujol mull)
110-94-1	Glutaric acid	5	2	6	solid (Nujol mull)
24-04-9	Adipic acid	6	2	8	aerosol[†]
123-99-9	Azelaic acid	9	2	24	solid (Fluorolube mull)
334-48-5	Decanoic acid	10	1	19	solution (10% in CCl4)
57-11-4	Stearic acid	18	1	35	solid

[†]Measured by atomization and filter collection in the laboratory by Gilardoni et al. (2007).

Table S2: Ninety-fifth percentile values of quantified moles of absent functional group when spectroscopically-adjacent functional groups are present. Correlation coefficients indicate the correlation between the quantified moles of absent functional groups (first column) and number of moles actually present for the functional group in the adjacent band (third column). Juxtaposed are detection limits reported by Russell et al. (2009), which are derived from two-sigma variability of absorbances in spectroscopic region of each functional group after background correction (no analytes present).

Functional group	95% percentile (μmol)	Adjacent band	Correlation coefficient	Standards (no. of samples)	Russell et al. (2009) (μmol)
Alcohol COH	0.18	Carboxylic COH	0.56	adipic acid (7), azelaic acid (5), pinonic acid (7), succinic acid (5)	0.02
Amine CNH_2	0.22	Carbonyl C=O	0.73	azelaic acid (5), mixture (8), pinonic acid (7), succinic acid (5)	0.01
Carboxylic COH	0.18	Alcohol COH	0.82	2-deoxy-d-ribose (4), 2-deoxy-d-ribose* (4), docosanol (6), glyceraldehyde (5)	0.008

Table S3: Elements of the molar apportionment array, λ_{ijk}, from Equation (8). The index, k, representing the degree of saturation by CH (as number of moles, n_{aCH}) and functionalization by FG j (characterized by functional group index, FGI_j; Russell, 2003) is defined by enumeration of the pairs, $(n_{\text{aCH}}, FGI_j) \in \mathcal{D}$. By convention, n_{aCH} does not include the number of alkane CH bonds in j (i.e., $n_{\text{aCH}} = 0$ when j corresponds to the alkane CH group as in rows 2 and 12).

				moles of atom i (per mole of FG j)				
k	(n_{aCH}, FGI_j)	FG (j)	Configuration	C	O	H	N	S
1	(0, 1)	Alcohol COH	R_3COH	1	1	1	0	0
1	(0, 1)	Alkane CH	R_3CH	1	0	1	0	0
1	(0, 1)	Amine CNH_2	R_3CNH_22	1	0	2	1	0
1	(0, 1)	Organonitrate $CONO_2$	R_3CONO_2	1	3	0	1	0
1	(0, 1)	Organosulfate $COSO_3$	R_3COSO_3	1	4	0	0	1
2	(0, 2)	Alcohol COH	$R_2C(OH)_2$	0.50	2	2	0	0
2	(0, 2)	Alkane CH	R_2CH_2	0.50	0	2	0	0
2	(0, 2)	Amine CNH_2	$R_2C(NH_2)_2$	0.50	0	4	2	0
2	(0, 2)	Organonitrate $CONO_2$	$R_2C(ONO_2)_2$	0.50	6	0	2	0
2	(0, 2)	Organosulfate $COSO_3$	$R_2C(OSO_3)_2$	0.50	8	0	0	2
3	(0, 3)	Alcohol COH	$RC(OH)_3$	0.33	3	3	0	0
3	(0, 3)	Alkane CH	RCH_3	0.33	0	3	0	0
3	(0, 3)	Amine CNH_2	$RC(NH_2)_3$	0.33	0	6	3	0
3	(0, 3)	Organonitrate $CONO_2$	$RC(ONO_2)_3$	0.33	9	0	3	0
3	(0, 3)	Organosulfate $COSO_3$	$RC(OSO_3)_3$	0.33	12	0	0	3
4	(1, 1)	Alcohol COH	R_2CHOH	0.50	1	1	0	0
4	(1, 1)	Amine CNH_2	R_2CHNH_2	0.50	0	2	1	0
4	(1, 1)	Carbonyl C=O	RCOH	0.50	1	1	0	0
4	(1, 1)	Carbonyl C=O	RCOR	1	1	0	0	0
4	(1, 1)	Carbonyl C=O	RCOOH	1	2	1	0	0
4	(1, 1)	Organonitrate $CONO_2$	R_2CHONO_2	0.50	3	1	1	0
4	(1, 1)	Organosulfate $COSO_3$	R_2CHOSO_3	0.50	4	1	0	1
5	(1, 2)	Alcohol COH	$RCH(OH)_2$	0.33	2	2	0	0
5	(1, 2)	Amine CNH_2	$RCH(NH_2)_2$	0.33	0	4	2	0
5	(1, 2)	Organonitrate $CONO_2$	$RCH(ONO_2)_2$	0.33	6	1	2	0
5	(1, 2)	Organosulfate $COSO_3$	$RCH(OSO_3)_2$	0.33	8	1	0	2
6	(2, 1)	Alcohol COH	RCH_2OH	0.33	1	1	0	0
6	(2, 1)	Amine CNH_2	RCH_2NH_2	0.33	0	2	1	0
6	(2, 1)	Organonitrate $CONO_2$	RCH_2ONO_2	0.33	3	2	1	0
6	(2, 1)	Organosulfate $COSO_3$	RCH_2OSO_3	0.33	4	2	0	1