Dissertations, Master's Theses and Master's Reports

2019

# MOLECULAR CHARACTERIZATION OF FREE TROPOSPHERIC ORGANIC AEROSOL AND THE DEVELOPMENT OF COMPUTATIONAL TOOLS FOR MOLECULAR FORMULA ASSIGNMENT

Simeon Schum
*Michigan Technological University*, skschum@mtu.edu

MOLECULAR CHARACTERIZATION OF FREE TROPOSPHERIC ORGANIC
AEROSOL AND THE DEVELOPMENT OF COMPUTATIONAL TOOLS FOR
MOLECULAR FORMULA ASSIGNMENT

By

Simeon K. Schum

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Chemistry

MICHIGAN TECHNOLOGICAL UNIVERSITY

2019

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Chemistry.

Department of Chemistry

Dissertation Advisor: *Lynn R. Mazzoleni*

Committee Member: *Sarah A. Green*

Committee Member: *Kathryn A. Perrine*

Committee Member: *Evan S. Kane*

Committee Member: *David C. Podgorski*

Department Chair: *John A. Jaszczak*

# Table of Contents

# Preface

Chapter 3, major portions of Chapter 2, and Section 6.2 were previously published in "Molecular and physical characteristics of aerosol at a remote free troposphere site: implications for atmospheric aging" by Schum et al. as published in *Atmospheric Chemistry and Physics* (2018). The author contributions to Schum et al. (2018) include: Simeon K. Schum, Katja Džepina, Paulo Fialho, Claudio Mazzoleni, and Lynn R. Mazzoleni performed the sample collection, measurements, and the data analysis. FT-ICR MS analysis was done by Simeon K. Schum and Lynn R. Mazzoleni. Bo Zhang performed the FLEXPART model simulations and interpretation. All authors contributed to the discussion and interpretation of the results and to the writing of the manuscript.

# Acknowledgements

First, I need to thank my parents for all their hard work and support, which has afforded me the ability to spend such a long time working on my education, and for instilling in me the desire to pursue education to this extent. I also must thank my advisor Dr. Lynn Mazzoleni, without her advice and support I would not have been able to accomplish what I have over the course of my studies, nor have the opportunities to contribute to the variety of projects that I participated in. Also, thanks to my group members who helped a lot with testing the code for MFAssignR, helping to improve it immensely along with being around to talk with and help to solve other problems. Additional thanks to the Michigan Tech Chemistry Department, Graduate School, and Earth, Planetary, and Space Sciences Institute for funding during my studies.

# Abstract

Organic aerosol affects human health and climate. These effects are largely determined by the composition of the organic aerosol, which is a complex mixture of species. Understanding the complexity of organic aerosol is critical to determining its effect on human health and climate. In this study, long range transported organic aerosol collected at the Pico Mountain Observatory was analyzed using ultrahigh resolution mass spectrometry. Organic aerosol transported in the free troposphere had an overall lower extent of oxidation than aerosol transported in the boundary layer. It was hypothesized that the lower oxidation was related to a more viscous phase state of the aerosol during transport. The results suggest that biomass burning organic aerosol injected into the free troposphere are more persistent than organic aerosol in the boundary layer. A sample was also analyzed using tandem FT-ICR MS/MS fragmentation, providing information about the functional group composition in the aerosol sample. This was done using a segmented scan approach, which revealed an unprecedented molecular complexity of unfragmented precursor ions. In addition to the expected $CO_2$ and $H_2O$ neutral losses, neutral losses corresponding to carbonyl functional groups ($C_2H_4O$, CO) were observed. The abundance of carbonyl functional groups suggests a slower rate of aging in the atmosphere. Analysis of nitrogen and sulfur containing neutral losses highlighted a surprising abundance of reduced nitrogen and sulfur loss ($NH_3$ and $SH_2$). This further supports the hypothesis of slower aging in the free troposphere. Additional research was done to develop an R software package (MFAssignR) to perform molecular formula assignment with improved decision-making transparency, noise estimation, isotope identification, and mass

recalibration. MFAssignR was found to assign the same molecular formula as other molecular formula assignment methods for the majority (97-99%) of mass peaks that were assigned a molecular formula by the compared methods. Additionally, MFAssignR was more effective at assigning molecular formulas to low intensity peaks relative to the other methods tested, leading to more overall molecular formula assignments. MFAssignR is available via GitHub and is the first open source package to contain a full pipeline of functions for data preparation and analysis for ultrahigh resolution mass spectrometry.

# 1 Introduction

## 1.1 Overview of Aerosol

Atmospheric aerosol are condensed particles suspended in a gas (Poschl, 2005). These particles range from liquid to solid, meaning that anything from inorganic dust to a cloud droplet can technically be considered an aerosol particle, although traditionally cloud droplets are not considered to be aerosol particles (Poschl, 2005). The diameter of aerosol particles is typically in the range of $10^{-9}$ m to $10^{-4}$ m (Poschl, 2005). In many cases, aerosol is categorized by its aerodynamic diameter as $PM_1$, $PM_{2.5}$, or $PM_{10}$ (Shiraiwa et al., 2017b). $PM_1$ has an aerodynamic diameter of $\leq 1$ μm and is classified as ultrafine aerosol, $PM_{2.5}$ is classified as fine aerosol with an aerodynamic diameter $\leq 2.5$ μm, while $PM_{10}$ includes particles with an aerodynamic diameter $\leq 10$ μm and is considered to be coarse aerosol (Shiraiwa et al., 2017b).

Atmospheric aerosol are complex mixtures with inorganic or organic contributions. Typical sources of inorganic aerosol are both natural and anthropogenic, such as wind-blown crustal materials (sand and soil) (D'Almeida and Schutz, 1983;Wetzel et al., 2003; Englebrecht et al., 2016), ash emissions from volcanic eruption (Schumann et al., 2013), sea spray (O'Dowd et al., 2004; Prather et al., 2013), road dust from roadways and vehicle wear (Chiou et al., 2007; Cote et al., 2018), concrete production (Chiou et al., 2007), and gas emissions from industrial and volcanic sources such as $SO_2$ (Schumann et al., 2011). Organic aerosol can come from biological sources (pollen, spores) (Burge, 2002; Taylor et al., 2002), biomass burning (wildfire) (Capes et al., 2008; Bougiatioti et al., 2014), biogenic emissions (Virtanen et al., 2010), sea spray (O'Dowd et al., 2004;

Prather et al., 2013), and anthropogenic sources such as fossil fuel combustion, cooking, heating, and crop burning (Fine et al., 2004; Volkamer et al., 2006; Buonanno et al., 2009; Lin et al., 2010). A schematic of the selected emission sources that most relevant for the production of organic aerosol is shown in Figure 1.1. We will focus on organic aerosol for the remainder of this introduction because it is the focus of the studies presented in this dissertation.
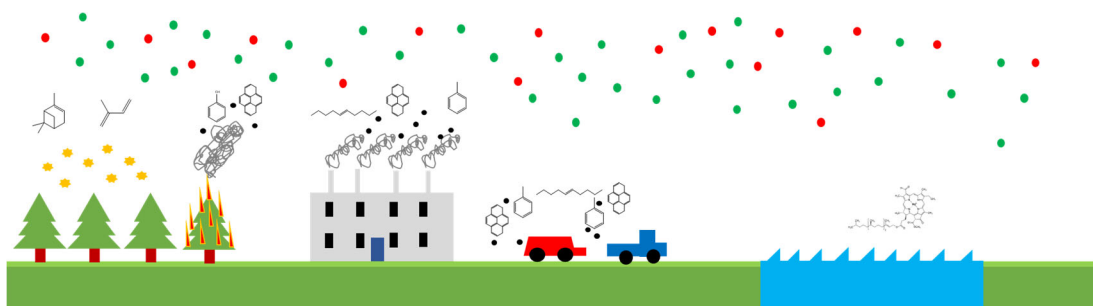


**Figure 1.1**. Schematic of selected emission sources of organic aerosol. The green and red dots represent fresh (green), and aged (red) aerosol particles. The yellow dots represent pollen or fungal spores. The black dots represent soot particles. The molecular structures represent potential gas phase pollutants being emitted from the various sources.

## 1.2 Gas Phase SOA Production

Organic aerosol is a complex mixture made up of thousands of different molecules and isomers (Walser et al., 2007; Mazzoleni et al., 2012; O'Brien et al., 2013; Wozniak et al., 2014; Džepina et al., 2015). There are two major types of organic aerosol, primary and secondary. Primary organic aerosol (POA) is condensed material directly emitted to the atmosphere as a condensed particle (Poschl, 2005). Examples include fungal spores or smoke from wildfires. Secondary organic aerosol (SOA) is formed through chemical reactions in the atmosphere that produce condensed species (Poschl, 2005; Volkamer et

al., 2006). SOA production can occur in the gas phase through processes referred to as auto-oxidation (Kundu et al., 2012; Ehn et al., 2012; Jokinen et al., 2014; Mutzel et al., 2015), via condensed phase pathways within aerosol water or droplets (Ervens et al., 2008; Lim et al., 2010; Ervens et al., 2011), or through heterogeneous pathways (Jang et al., 2002). Biogenic SOA is often formed from trace gases released from trees such as terpenes (Ferman et al., 1981; Griffin et al., 1999). Precursors of SOA from anthropogenic sources are typically aromatic or aliphatic hydrocarbons such as toluene, benzene, heptanal, and 1-decene (Gelenscer et al., 2004). Once emitted, these gases can be oxidized in the atmosphere by oxidants such as OH radicals, $O_3$, and $NO_x$ (Kanakidou et al., 2005). Oxidation of the trace gases eventually leads to gas-to-particle partitioning into a liquid or solid state as the volatility of the compound decreases and other molecules bind and coagulate together (Kroll and Seinfeld, 2008). This oxidation can proceed very rapidly leading to highly oxidized molecular species within a few hours downwind of their emission source (Jimenez et al., 2009; Vakkari et al., 2014). Oxidation of biogenic terpene precursor molecules can also result in new particle formation as observed in forested environments such as Hyytiala (Virtanen et al., 2010). Emissions from biogenic SOA can contribute 12-70 Tg/yr of organic aerosol to the atmosphere (Hallquist et al., 2009). In contrast, the production of SOA from anthropogenic precursors is on the order of 2-12 Tg/yr (Hallquist et al., 2009).

## 1.3   Aqueous Phase SOA Production

SOA can also be produced by aqueous phase processes within cloud or fog droplets (Blando and Turpin, 2000; Lim et al. 2010; Ervens et al., 2011). Oxidized species are

produced in the aqueous phase become aerosol particles when the water evaporates (Carlton et al., 2007). Aqueous phase processing has been shown to lead to an overall higher extent of oxidation than is typically observed from in gas phase experiments (Lim et al., 2010; Ervens et al., 2011). Although, some recent studies have suggested that the oxidation from both methods can be comparable (Ehn et al., 2012; Jokinen et al., 2014). Laboratory studies have shown rapid processing of glyoxal in aqueous solutions produces organic material that is similar to organic aerosol present in the atmosphere (McNeil, 2015; Hawkins et al., 2016). Predicted pathways for the production of SOA from aqueous processing include Maillard reactions (Hawkins et al., 2016), Fenton reactions (Moonshine et al., 2008), radical reactions, non-radical reactions and oligomerization reactions (McNeil, 2015). Common species that undergo these reactions are water soluble aldehydes, epoxides, glyoxal, organic acids, amines, and phenols (McNeil, 2015). Some studies have shown that aqueous processing of glyoxal and pinene SOA together closely match observations of aged aerosol in the atmosphere, especially with respect to the production of low volatility oxygenated organic aerosol (LV-OOA) (Lee et al., 2012). Aqueous processing of methylglyoxal has been found to lead to the production of brown carbon (BrC), which has an uncertain effect on radiative forcing (De Haan et al., 2017). The primary source of BrC is biomass burning, which is discussed in Section 1.4. The heterogeneous pathway for SOA production occurs as an interaction between typically gas phase compounds and solid phase particles, where the solid phase particles provide a surface for the reaction to take place (Jang et al., 2002), or directly reacting with the gases. Heterogeneous reactions are often acid-catalyzed (Jang et al., 2002) and have been shown to increase organic aerosol formation by a factor of 5 (Jang et al., 2002). The

HNO$_3$ and H$_2$SO$_4$ produced from anthropogenic emissions of SO$_2$ and NO$_x$ are significant contributors to the acid-catalyzed reactions common in heterogeneous pathways.

## 1.4   Biomass Burning Aerosol

Biomass burning contributes to both POA and SOA (Simoneit, 2002). Biomass burning emissions undergoes atmospheric processing according to all of the pathways described above. Biomass burning aerosol is generally less oxidized than aerosol from biogenic or anthropogenic sources (Aiken et al., 2008; Bougiatioti et al., 2014). Biomass burning produces black and brown carbon which contributes to its ability to directly absorb light, and affect the Earth's radiative balance (Desyaterik et al., 2013; Lin et al., 2015). The production of light absorbing aerosol from biomass burning depends on the intensity of the fire. Fires that burn vigorously lead to more complete combustion which typically produce more black carbon (Hopkins et al., 2007), while a smoldering fire tends to produce a larger amount of light-absorbing brown carbon (BrC) and tar balls (Chakrabarty et al., 2010). The fuel type also plays a significant role in the composition and absorption of the biomass burning aerosol (Levin et al., 2010). Some of the aromatic compounds that are produced via incomplete combustion are polycyclic aromatic hydrocarbons (PAH), many of which are known to be carcinogenic and mutagenic (Perraudin et al., 2006; Bignal et al., 2008). Biomass burning aerosol contributes a large fraction of the total organic aerosol loading in the atmosphere (Pratt et al., 2010) and as the frequency and size of wildfires increases (Turetsky et al., 2011), the amount of aerosol due to biomass burning is expected to increase (Spracklen et al., 2009). Recently,

the absorbance of BrC aerosol produced from biomass burning has been gaining attention because it has less constrained absorbance characteristics relative to black carbon (Andrae and Gelencser, 2006; Chakrabarty et al., 2010; Desyaterik et al., 2013; Lin et al., 2015).

BrC is organic aerosol that absorbs light with a wavelength dependence. BrC does not absorb light at all wavelengths as efficiently as black carbon, but it does have increased absorption in the visible wavelengths relative to typical SOA (Andrae and Gelencser, 2006; Chakrabarty et al., 2010). Thus, BrC has an influence on the radiative forcing of the planet, and this means that biomass burning aerosol has a more complicated effect than was previously considered (Andrae and Gelencser, 2006). Studies of BrC during transport have suggested that the light-absorbing BrC components are mostly removed within ~1 day in the boundary layer (Forrister et al., 2015; Laing et al., 2016), which would limit its long-range effects. However, based on studies of PAH transport in the free troposphere (Shrivastava et al., 2017), there is precedence for longer lifetime of aerosol species transported in the free troposphere, which may also be applicable to BrC. This could imply greater long-range effects of BrC if it can be transported to the free troposphere.

## 1.5   Aerosol Lifetime and Transport

The lifetime of aerosol in the atmosphere can extend from hours to weeks (Poschl, 2005) and is dependent on the composition of the aerosol and the meteorological conditions it encounters (Poschl, 2005; Schum et al., 2018). The typical removal mechanisms for

organic aerosol are wet deposition, dry deposition, and oxidative degradation in the

atmosphere (Poschl, 2005). Wet deposition can occur when an aerosol particle acts as a

cloud condensation nucleus (CCN), forming a cloud or rain droplet or when an aerosol

particle is scavenged by a falling water droplet (Poschl, 2005). Dry deposition occurs

either when a particle is too heavy and falls from the sky due to gravity or when it

impacts a surface while aloft, such as a building. (Stefanis et al., 2009). Oxidation of

aerosol through interactions with gas phase oxidants such as (OH, $NO_x$, $O_3$) and aqueous

phase processing can also lead to the removal of aerosol from the atmosphere. Eventually

as the aerosol components are oxidized, they become fragmented into lower molecular

weight volatile compounds, which can evaporate from the aerosol particle or be dissolved

in a water droplet (Kessler et al., 2010; Kroll et al., 2011) and are removed from the

aerosol particle. The lifetime of aerosol in the atmosphere also impacts how far it can be

transported from its source. Several studies have shown transport of several thousand km

from the source region (Damoah et al., 2004; Dirksen et al., 2009). Biomass burning

events are a common source of these long range transported aerosol. This is largely due

to the pyro-convection that can occur over fires, causing aerosol to be injected directly

into the free troposphere (Val Martin et al., 2008a, Kahn et al., 2008). Aerosol in the free

troposphere typically have a longer lifetime than aerosol in the boundary layer due to

fewer oxidants and the cool, dry conditions can cause the aerosol to be in a solid state,

increasing its resistance to oxidative processes (Koop et al., 2011; Lignell et al., 2014;

Hinks et al., 2016). This may also extend to the BrC species, which typically have

lifetimes on the order of 1 day when in the boundary layer (Laing et al., 2015; Forrister et

al. 2016). The transport of these aerosol particles can cause problems downwind of the

emission sources due to the effect aerosol has on climate and human health (Shiraiwa et al., 2017b). A schematic of aerosol production, pyro-convection and transport of aerosol can be seen in Figure 1.2.
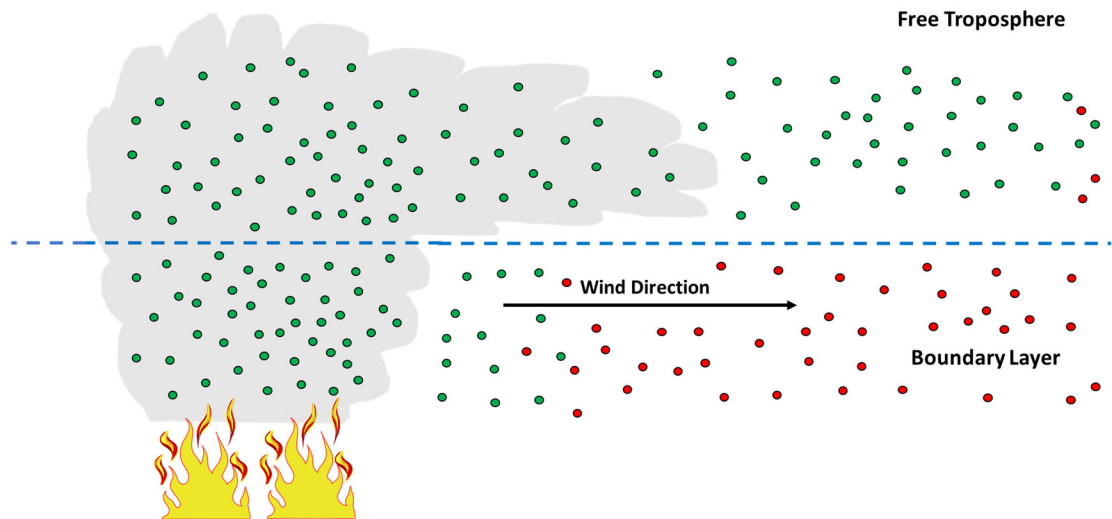


**Figure 1.2.** Schematic depicting wildfire pyro-convection of aerosol to the free troposphere. Green dots indicate fresh aerosol, while red dots indicate increased oxidation of the aerosol particles. This is done to suggest the difference in rate of oxidation in the free troposphere relative to the boundary layer. Fire image source: www.kisspng.com/png-cartoon-fire-flame-clip-art-bbq-fire-cliparts-183620/download-png.html

## 1.6 Climate and Health Effects of Aerosol

Relative to atmospheric gases such as $CO_2$ and $CH_4$, the effect of aerosol on the climate is much less certain (IPCC 2013). Depending on its source and composition aerosol can have either a warming or cooling effect. This effect can be direct or indirect, which makes it more difficult to constrain than if aerosol could only affect the climate in one way. The direct effect of aerosol on the climate is through scattering or absorption of incoming radiation. Most SOA and sulfate aerosol are considered to be dominated by

scattering (Lin et al., 2014), which has a cooling effect on the atmosphere. In contrast, soot and BrC species that are emitted from biomass burning strongly absorb light and can contribute to warming (Lin et al., 2014). Since biomass burning is a significant contributor to the global aerosol loading (Spracklen et al., 2009; Pratt et al., 2010), its light absorbing components may have a significant impact on global radiative forcing. Aerosol also has an indirect effect on the climate through its ability to act as a cloud condensation nuclei (CCN). Clouds also have a scattering effect on incoming radiation, which leads to a cooling effect. If aerosol is not hygroscopic, it is less likely to act as a CCN (Massoli et al., 2010).

The hygroscopicity and number of aerosol also have an effect on visibility in urban and rural environments (Cheng and Tsai, 2000). Visibility reduction from aerosol happens in several national parks, such as the Blue Ridge Mountains, where the decrease in visibility is due to biogenic emissions that readily form aerosol, which can directly or indirectly obscure views (Ferman et al., 1981). It can also happen in urban environments, where the aerosol fog can be more hazardous because it contains products from fossil fuel combustion, which may be toxic.

More importantly urban aerosol and subsequent smog can contain toxic chemicals that are hazardous for human, animal, and plant health (Shiraiwa et al., 2017b). Urban aerosol has been shown to contain PAH and other toxic molecules (Ollivon et al., 2002) that are known to be carcinogenic. The size of aerosol allows them to be directly inhaled, where some are even small enough ($\leq 0.1$ μm) to enter the bloodstream, allowing them to be transported to other organs in the body (Shiraiwa et al., 2017b). In general aerosol have

been linked to an average decrease in life expectancy of 1.2 years across the globe, with certain regions, such as Eastern Europe (Russia) having a decrease of 2.5 years (Shiraiwa et al., 2017b). Pulmonary, cardiovascular, and asthma problems have been linked to aerosol exposure in several studies (Dockery et al., 1996; Taylor et al., 2002; Franck et al., 2011). Aerosol particles can have many detrimental health and climate effects and the major drivers of those impacts are the physical and chemical characteristics of the aerosol particles.

## 1.7   Oxidation, Hygroscopicity, Volatility, and Viscosity of Aerosol

### 1.7.1   Accretion and Oxidation of Aerosol

The major methods of aerosol transformation in the atmosphere are accretion and oxidation. Accretion can lead to larger molecules as multiple smaller molecules bind together through hemiacetal formation, aldol condensation, ester formation, organosulfate formation, and Criegee reactions with alcohol, water, or acids (Kroll and Seinfeld, 2008). These accretion reactions occur in the particle phase and do not increase the oxidation of the aerosol species (Kroll and Seinfeld, 2008). The interaction with gaseous and particle phase species with reactive species such as $O_3$ and OH radicals leads to oxidative transformations of the molecules (Rudich et al., 2007). The reaction pathways are similar to those that lead to the production of aerosol described previously (Rudich et al., 2007; Kroll and Seinfeld, 2008). Rapid atmospheric oxidation has been observed in the boundary layer for aerosol from anthropogenic (Jimenez et al., 2009), biogenic (Corrigan et al., 2013; Ehn et al., 2014; Jokinen et al., 2014), and biomass burning emission sources (Capes et al., 2008; Bougiatioti et al., 2014; Vakkari et al., 2015). In contrast, some

studies have reported lower relative oxidation for aerosol that was transported in the free troposphere (Džepina et al., 2015; Schum et al., 2018). Increased oxidation leads to increased water solubility and hygroscopicity (Rudich et al., 2007; Kroll and Seinfeld, 2008; Massoli et al., 2010), and decreased volatility (Jimenez et al., 2009; Li et al., 2016) in condensed molecules. Continued oxidation can also lead to fragmentation and re-volatilization of the small fragments (Kroll et al., 2011).

### 1.7.2   Aerosol Hygroscopicity

Hygroscopicity refers to the ability of aerosol to take up water from its surroundings. Some types of inorganic aerosol are very hygroscopic, such as sulfate and nitrate aerosol (Cruz and Pandis, 2000; Lightstone et al., 2000). For this reason, these types of inorganic aerosol have been included in models in order to predict the hygroscopicity of aerosol (Petters and Kreidenweis, 2007) and to predict the aerosol water content (Nenes et al., 1998). Inorganic components of aerosol are a significant portion of the overall aerosol mass, so considering them is a good way to predict hygroscopicty and aerosol water content. However, organic material typically makes up 20-90 % of the total aerosol mass fraction (Jimenez et al., 2009), so understanding its contribution to the aerosol hygroscopicity is very important. Hygroscopicity of an aerosol particle relates to its ability to act as a cloud condensation or ice nuclei (CCN and IN respectively) (Massoli et al., 2010; China et al., 2017). As the hygroscopicity increases it becomes a better nucleus (Massoli et al., 2010). The formation of these cloud droplets leads to wet deposition or additional oxidation through aqueous oxidation pathways (McNeil et al., 2015).

### 1.7.3   Aerosol Volatility

In addition to the hygroscopicity, the oxidation of organic species affects the volatility of the molecules in organic aerosol. Volatility determines the equilibrium of molecules between the gas and condensed phases. Five categories of volatility are defined for atmospherically relevant species; volatile organic compounds (VOC), semi-volatile organic compounds (SVOC), intermediate volatility organic compounds (IVOC), low volatility organic compounds (LVOC), and extremely low oxidation organic compounds (ELVOC). The volatility of a compound affects its ability to nucleate new aerosol particles (Kanakidou et al., 2005) and its transformation in the atmosphere. Several methods have been developed to predict the volatility of organic molecules in aerosol (Donahue et al., 2011; Li et al., 2016). These methods rely on the molecular composition of the species present in the organic aerosol to do these predictions. The method described in Donahue et al., (2011) was designed for compounds that contain carbon, hydrogen, and oxygen, while the method from Li et al. (2016) can be applied to molecular formulas with sulfur and nitrogen. These methods make assumptions about the functional groups that are present in the species because the functional groups have an impact on the volatility of the molecule (Kanakidou et al., 2005). For example, a carboxyl group causes a greater decrease in the volatility than a hydroxyl group (Kanakidou et al., 2005; Rothfuss and Petters, 2016). In general, the more oxygen functional groups that are present on a molecule, the less volatile it is (Kanakidou et al., 2005). The volatility is also related to the viscosity and phase state of aerosol particles (Li et al., 2016; Rothfuss and Petters, 2016).

### 1.7.4 Aerosol Viscosity

Viscosity is the measure of the flow rate of a substance. As viscosity increases aerosol particles are considered to become marble-like amorphous spheres. Until recently it was expected that the majority of organic aerosol was liquid or perhaps semi-solid, but in a study of newly formed particles over Hyytiala Finland (Virtanen et al., 2010) showed that some particles "bounced" when drawn into a multi-stage aerosol impactor designed to collect particles with different aerodynamic diameters. The "bouncing" suggested that the aerosol was in a solid state. This study provided field confirmation for some laboratory studies that had suggested aerosol could exist in a highly viscous solid-state (Zobrist et al., 2008; Koop et al., 2011). The viscosity of aerosol is very important because it can affect the degree of aerosol oxidation, potentially contributing to greater transport distances for molecular species than are predicted with the assumption of liquid aerosol (Shrivastava et al., 2017). Aerosol in the liquid state is more susceptible to chemical and photo oxidation because the oxidants can more easily mix with the organic species within an aerosol particle (Berkemeier et al., 2014; Lignell et al., 2014; Hinks et al., 2016). When aerosol is in the solid phase it can have a greater resistance to these environmental conditions, which may lead to lower oxidation than would be observed for a similar particle in the liquid state (Berkemeier et al., 2014; Lignell et al., 2014; Hinks et al., 2016). Contributing factors to the phase state of organic aerosol can be its composition (Rothfuss and Petters, 2016; Zelenyuk et al., 2017) and its environmental conditions (Koop et al., 2011; Shiraiwa et al., 2017a; DeRieux et al., 2018). A recent study by Zelenyuk et al. (2017) has shown that the addition of PAH to pinene SOA can lead to the formation of a highly viscous "shell" that greatly decreases the diffusivity of oxidants

through the particle. Since PAH are a known product of incomplete combustion from both wildfires and anthropogenic sources, this may suggest a greater propensity for these aerosol to be viscous in the atmosphere. However, the biggest driver of viscosity, is the meteorological conditions around the aerosol, particularly the ambient temperature and relative humidity (Schum et al., 2018). In cold, dry conditions the aerosol are much more likely to be in a solid state than if it is warm and wet (Koop et al., 2011; Shiraiwa et al., 2017a). The major reason for this is that water is an excellent plasticizer for typical aerosol species (Koop et al., 2011; Shiraiwa et al., 2017a). A plasticizer is a compound that can cause a compound to soften. A plasticizer works by lowering the overall glass transition temperature ($T_g$) of the mixture (Koop et al., 2011) and since the $T_g$ of water is 136 K compared to a typical aerosol compound with $T_g$ of ~300 K (Schum et al., 2018), it is a very effective plasticizer. $T_g$ is not the same as the melting temperature ($T_m$), but is related. The relationship between $T_g$ and the melting temperature has been determined to be $T_g \approx 0.6 * T_m$ (Koop et al., 2011; Rothfuss and Petters, 2016). $T_g$ is appropriate for determining the phase state of aerosol particles because it is an un-ordered mixture of many different organic and inorganic molecules, which makes it more glass-like than crystalline. The glass transition temperature can be related to viscosity by considering the ambient temperature and relative humidity. If the ambient temperature is higher than $T_g$, then the compound will be either liquid or semi-solid, and if $T_g$ is greater than the ambient temperature, the compound will be solid (Shiraiwa et al., 2017a). The relative humidity plays a significant role because it can lower the $T_g$ of a compound based on the plasticizing effect of water (Koop et al., 2011; Shiraiwa et al., 2017a). For example, if a compound has a dry $T_g$ well above the ambient temperature, it can still be liquid or semi-

solid after the relative humidity adjustment to a wet $T_g$ value. This effect can be estimated using the Gordon-Taylor equation, which was developed for polymers, but works for organic aerosol as well (Gordon and Taylor, 1952; Koop et al., 2011; Shiraiwa et al., 2017a). The potential importance of viscosity on aerosol lifetime and transport was highlighted in several recent studies (Shiraiwa et al., 2017a; Shrivastava et al., 2017), where it was suggested that the solid phase state of aerosol could lead to an increased lifetime and transport distance of aerosol species expected to have much shorter lifetimes. The study by Shrivastava et al. (2017) found that by considering phase state in their modeled transport of PAH they better reproduced field observations of PAH concentrations in regions far from their sources. The transport of carcinogenic PAH has a potential important impact on human health. Our recent study of organic aerosol collected at the Pico Mountain Observatory (PMO) observed a lower than expected oxidation of aerosol species after long range transport, which was attributed, in part, to the aerosol phase state during transport (Schum et al., 2018; see also Section 3.6).

### 1.7.5 Predicting Aerosol Viscosity

Recent studies by Shiraiwa et al. (2017a) and DeRieux et al. (2018) presented methods to estimate the $T_g$ of molecular compounds with masses from 100 -1000 Da that contained carbon, hydrogen, and oxygen. In Shiraiwa et al. (2017a) the estimation covered masses from 100-400 Da and was used to model the phase state of SOA at different levels of the atmosphere, finding that at higher altitudes SOA was predicted to be more solid. This observation is consistent with the expectation of cold, dry conditions leading to solid aerosol particles. The equation from DeRieux et al. (2018) was optimized for masses

from 100-1000 Da and was used in a recent study by Schum et al. (2018; see also Section 3.6) to predict the glass transition temperature of organic aerosol collected at PMO, more discussion about this is in Chapter 3. The viscosity, volatility, and hygroscopicity of aerosol are all affected by the functional groups that are present on the molecules (Rothfuss and Petters, 2016), a discussion of functional groups and their effect on these characteristics is provided in the following paragraph.

## 1.8   Functional Groups

### 1.8.1   Functional Group Overview

The functional groups of organic aerosol species are important to their chemical, physical, and optical properties (Kanakidou et al., 2005; Rothfuss and Petters, 2016). This has motivated multiple studies to determine which functional groups are present in aerosol. Most studies focused on the bulk composition of functional groups by FT-IR or NMR analysis (Decesari et al., 2000; 2007; Hawkins and Russell, 2010; Takahama et al., 2013). These studies have reported that oxidized functional groups such as carboxylic acids, alcohols, and ketones are common components of aerosol species, in addition to the carbon backbone functional groups such as alkanes, alkenes, and aromatics. The prevalence of certain functional groups has been shown to increase or decrease with atmospheric aging of the aerosol. For example, a study by Hawkins and Russell (2010) showed that fresh biomass burning aerosol had a large fraction of carbonyl groups, but as the sample aged, the fraction of carboxyl increased as carbonyl groups decreased. The increase in carboxyl functional groups with aging has been demonstrated in several studies (Hawkins and Russell, 2010; Takahama et al., 2013). Functional group analysis

has also shown the presence of sulfur and nitrogen containing functional groups such as sulfate, nitrate, and nitro groups (Clegg et al., 2008).

### 1.8.2 Effects of Functional Groups

The presence of the above-mentioned functional groups can affect the hygroscopicity, volatility, viscosity, and light absorption of the aerosol species in the atmosphere (Kanakidou et al., 2005; Rothfuss and Petters, 2016; Li et al., 2016). In recent studies the effect of multiple hydroxyl groups on viscosity was investigated (Rothfuss and Petters, 2016). They found that with each additional hydroxyl group that was added, the viscosity of the particles increased by approximately 0.8 orders of magnitude. They observed a similar trend for the addition of carboxyl and nitrate groups (Rothfuss and Petters, 2016). Other functional groups such as carbonyls and esters did not have an impact as significant (Rothfuss and Petters, 2016). Similar studies have been performed with regard to the hygroscopicity of species with additional oxygenated functional groups which were found to lead to increased hygroscopicity (Massoli et al., 2010). It has also been reported in the literature that the addition of oxygen to a molecule can lead to a decrease in the volatility (Li et al., 2016). Functional groups on the molecular species have also been suggested to have some influence on the light absorbing ability of organic aerosol particles (Gelenscer et al., 2003; Kanakidou et al., 2005; Phillips and Smith, 2014; 2015). The studies by Phillips and Smith (2014; 2015) suggested that some of the absorbance observed for aerosol species, especially at higher wavelengths, was due to charge transfer between adjacent functional groups such as carboxylic acids and alcohols. Measured and predicted functional groups are often used in modeling studies in order to predict the

17

interactions of aerosol in the atmosphere (Kanakidou et al., 2005; Petters and

Kreidenweis, 2007; Rothfuss and Petters, 2016; Clegg et al., 2019 in review) therefore

more information is valuable to better constrain these models. Bulk functional groups do

not provide mass information, or any specifics about molecular structure. A study of

select species in atmospheric organic matter by LeClair et al. (2012), investigated

functional groups using tandem ultrahigh resolution MS/MS analysis (LeClair et al.,

2012). Lower resolution tandem MS/MS has been used to investigate select species in

dissolved organic matter (Leenheer et al., 2001). Such studies have seen limited use in

analyzing complex mixtures. A major reason for this is the mass spectral complexity

makes it difficult to perform comprehensive analysis due to the large number of peaks to

be interpreted. Chapter 4 of this dissertation provides a discussion of the composition and

functional groups identified through MS/MS for a long range transported organic aerosol

sample.

## 1.9    Instrumentation

### 1.9.1    Example of Ultrahigh Resolution MS Applications

Organic aerosol is analyzed using many offline analytical methods such as NMR, FT-IR,

UV-Vis spectroscopy, fluorescence spectroscopy, ion chromatography, and mass

spectrometry (Duarte et al., 2005; Decessari et al., 2007; Agarwal et al., 2010; Hawkins

and Russell et al., 2010). Here we focus on ultrahigh resolution mass spectrometry.

Ultrahigh resolution MS has been used to analyze a variety of natural organic matter such

as rain water (Altieri et al., 2009), fog water (Mazzoleni et al., 2010), cloud water (Zhao

et al., 2013; Cook et al., 2017), aerosol (Schmitt-Kopplin et al., 2010; Wozniak et al.,

2014; Willoughby et al., 2016; Brege et al., 2018), dissolved organic matter (Hertkorn et al., 2008; Kujawinski et al., 2009; Hawkes et al., 2016), and even the organic matter in a meteorite (Schmitt-Kopplin et al., 2010). There are two instruments capable of ultrahigh resolving power, one is the Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS), and the other is the compact high field Orbitrap Elite MS (ThermoScientific). The FT-ICR MS is the first instrument capable of ultrahigh resolution and was developed by Comissarow and Marshall (1974). Certain versions of this instrument are capable of having a resolving power greater than 2,700,000 at mass to charge (m/z) of 400 (Smith et al., 2018), which is superior to the Orbitrap. The Orbitrap was developed by Makarov (2000). The compact high field Orbitrap can reach a resolving power of 240,000 at m/z 400, and with a developer add-on it can reach a resolving power of 480,000 at m/z 400. This resolution is sufficient to separate peaks that differ by mDa mass differences such as $SH_4$ vs. $C_3$, which has a mass difference of 3.4 mDa. Separation of these peaks is necessary to identify sulfur containing compounds, which are an important component of many aerosol samples (Schmitt-Kopplin et al., 2010). The ultrahigh resolution offered by these instruments is critical to obtaining the molecular level composition of complex mixtures such as organic aerosol (Schmitt-Kopplin et al., 2010).

### 1.9.2   Brief Theory of FT-ICR MS

FT-ICR instruments use a super conducting magnet to obtain extreme mass resolving (Marshall and Hendrickson, 2008). Generated ions are transported through ion focusing and mass filtering components and then accumulated using an ion trap, before ICR

analysis. When the ions enter the ICR cell they encounter a magnetic field which bends

the ions path into a circle by the Lorentz force. This path is the cyclotron resonance

motion of the ion. The initial cyclotron radii are small and unable to be measured, so the

ions are excited by an RF electric field pulse which excites the ions and causes their

cyclotron radii to expand to move according to their cyclotron frequency while

continuing their original circular orbit (Marshall and Hendrickson, 2008). Ions with the

same mass form coherent packages, which induce an image current on the detector plates.

This time-domain signal is then converted to frequency-domain spectrum via Fourier

transformation which can be converted to mass measurements (Marshall and

Hendrickson, 2008). A schematic of an ICR cell, the movement of ions within it, and the

transformation of the signal to a mass spectra can be seen in Figure 1.3.
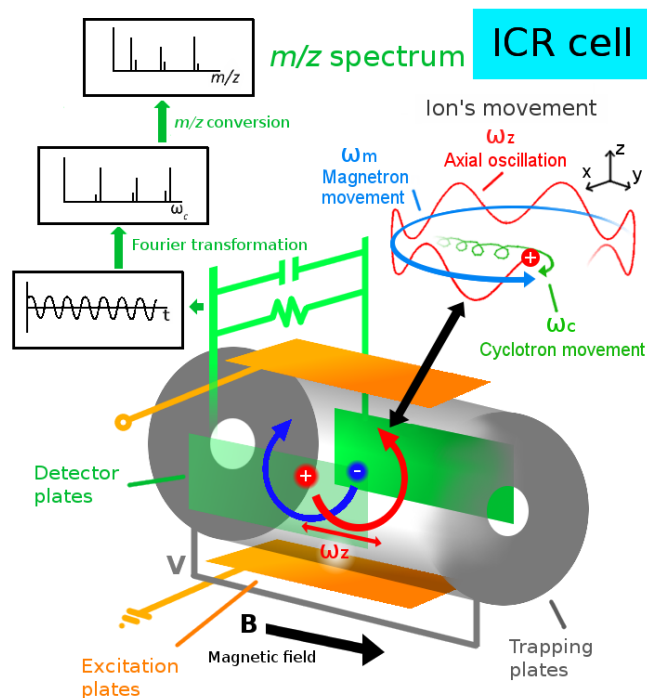


**Figure 1.3**. Schematic of an ICR cell with an illustration of the ion movement. Image source: https://commons.wikimedia.org/wiki/File:FTICR_cell.png

### 1.9.3 Brief Theory of Orbitrap MS

For the Orbitrap MS the initial stages are similar to the FT-ICR, but before the Orbitrap is a bent ion trap called a C-trap. The C-trap is used to compact the ion beam so that they can be injected simultaneously into the Orbitrap at the correct velocity and angle. The velocity and angle are very important to ensure that the ions properly orbit the inner spindle. The inner spindle is electrically charged to attract the ions, but when the ions are injected at the proper velocity and angle, they start to orbit the spindle instead of crashing into it (Makarov, 2000). The voltage on the inner spindle electrode creates an electric potential with the virtual ground outer electrode (Michalski et al., 2012). This potential increases at the end of the spindle and pushes the ions back toward the center of the trap. This back and forth movement, while orbiting the center spindle is the cause of the increased resolving power because it separates the ions by mass (Makarov, 2000). Then the ion image current is detected and Fourier transformed into mass measurements, like with the FT-ICR MS (Makarov, 2000). A schematic of an Orbitrap can be seen in Figure 1.4.
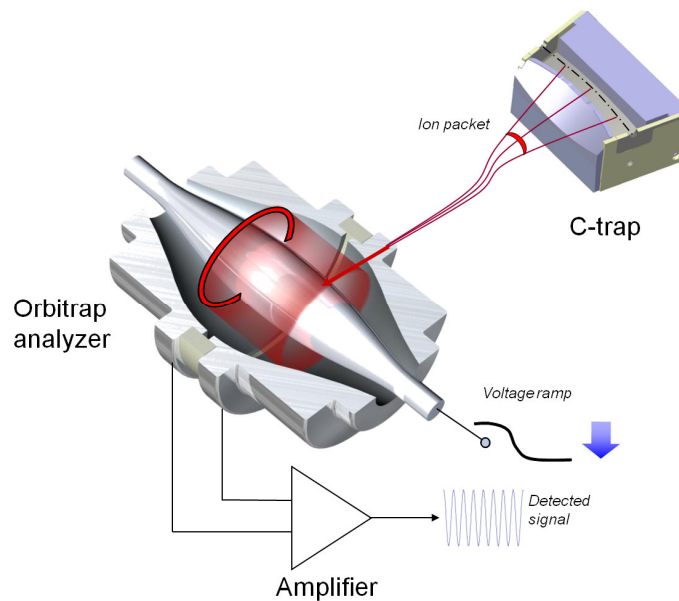
**Figure 1.4**. Schematic of an Orbitrap analyzer demonstrating the spindle shape and ion injection from the C-trap. Image source: https://en.wikipedia.org/wiki/Orbitrap.

## 1.10  Methods of Ionization

Typically, ultrahigh resolution MS is coupled with a soft ionization technique, to allow

for the detection and measurement of molecular ions. The most common method of soft

ionization is electrospray ionization (ESI). ESI works by applying a high voltage to a

stream of liquid coming from a small inner diameter capillary, this field causes excess

charge accumulation leading to the formation of a Taylor cone (Cech and Enke, 2001).

When the Rayleigh limit is reached, a spray of charged droplets are released (Cech and

Enke, 2001). These charged droplets then travel towards the inlet of the instrument,

during which time the solvent evaporates, causing the most surface-active ions to be

enhanced (Cech and Enke, 2001; de Hoffman and Stroobant, 2007). This ionization

process can be used to produce negative or positive ions depending on the voltage

polarity. Negative mode ESI ionizes polar, acidic molecules most efficiently (Cech and Enke, 2001) and so it is commonly used as the main ionization source for organic aerosol, since the species are expected to be relatively polar and acidic (Bougiatioti et al., 2016). Positive ESI is more effective for relatively reduced and basic molecules (Cech and Enke, 2001). It has begun to be used more commonly in recent studies of natural organic matter as analysts attempt to get a more comprehensive view of the molecular composition of their samples (Hertkorn et al., 2008). Other ionization methods that are becoming more common are atmospheric pressure chemical ionization (APCI) and atmospheric pressure photoionization (APPI). Compared to ESI these methods are more efficient in ionizing less polar, and higher aromaticity compounds. APPI, in particular, is better for aromatic species (Hertkorn et al., 2008). APPI works by vaporizing a sample stream using a heated chamber to rapidly dry the droplets. Photons from a lamp (typically krypton) are used to ionize molecules, which then are transported into the instrument (de Hoffmann and Stroobant, 2007). Sometimes a dopant such as toluene is added to the solvent to facilitate ionization of the analyte through the transfer of a proton, or electron. (de Hoffmann and Stroobant, 2007). Similar to APPI, APCI occurs in the gas phase, so after vaporization molecules are ionized using a corona discharge needle (de Hoffmann and Stroobant, 2007), which causes proton transfer (positive mode), abstraction (negative mode), or adduct formation (both), causing the ionization of the molecules (de Hoffman and Stroobant, 2007). For the studies presented in this dissertation, negative ESI was primarily used.

## 1.11 Competitive Ionization and Segmented Scanning

ESI is a competitive process, meaning that if a particular molecule is very easily ionized it will be detected more easily than other ions in the mixture. This means that the observed ion intensity is not directly related to the analytes concentration in the sample (Cech and Enke, 2001) thus the ion intensity is not quantitative. Additionally, most ion trap type instruments have a maximum limit on the number of ions. This is done to avoid space-charge effects, which can significantly impact the mass measurement accuracy of the measurements (Belov et al., 2003; Kozhinov et al., 2013). The combination of the ESI competitive process and the limited number of ions can lead to suppression of the lower abundance or lower ionization efficiency species in a sample. This happens because the species that are easily ionized/more abundant will fill up the ion trap quickly, making it much more difficult to detect the low efficiency/abundance ions. A schematic of this process is given in Figure 1.5. One way to increase the overall sensitivity to less efficient/abundant ions is to use a segmented scanning technique (Southam et al., 2007). Using shorter ion segments, there is a smaller mass range of ions attempting to enter the trap, so the detection of less abundant ions is improved. This method has been used previously for crude oil samples (Gaspar and Schrader, 2012), but has not been reported for organic aerosol prior to this dissertation (Chapter 4).
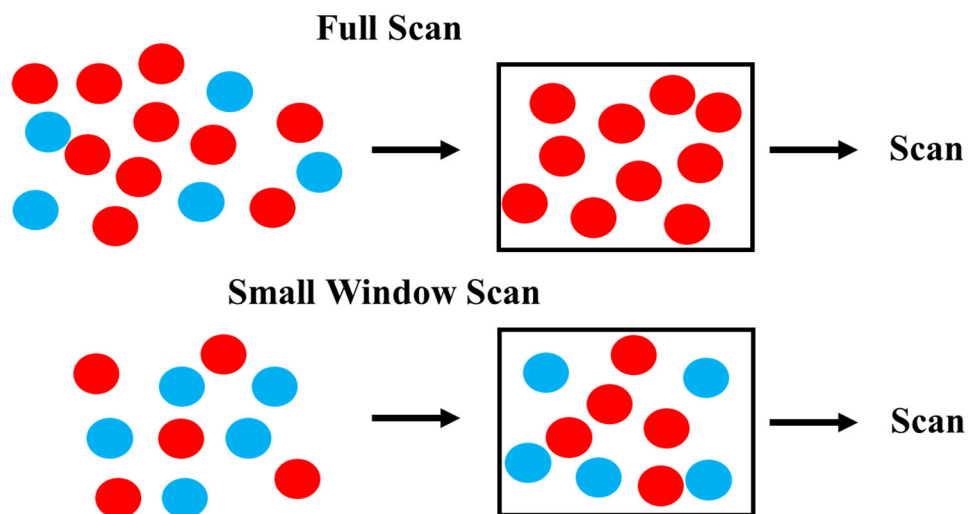
**Figure 1.5.** Schematic of full scan analysis vs. small window analysis. Red circles represent high concentration/ionization efficiency species, blue represent low concentration/ionization efficiency species.

## 1.12  MS/MS Fragmentation

### 1.12.1  Fragmentation Methods

While the ability to detect unfragmented molecular ions is very helpful for determining

the molecular mass and formula for molecules it provides very little information about

the structure of the molecule that was detected. To get structural information requires the

use of tandem MS/MS to fragment the ions and provide functional group and other

structural information (Gross, 2017). There are a variety of fragmentation methods that

are used in ultrahigh resolution MS instrumentation such as collision induced dissociation

(CID), infrared multiphoton dissociation (IRMPD), electron capture dissociation (ECD),

and electron transfer dissociation (ETD). ECD and ETD are methods that are well suited

for the fragmentation of large multiply charged ions such as proteins (Gross, 2017). They

work by sending a beam of low energy electrons at the ion (ECD) or transferring an

electron from a singly charged anthracene or fluoranthene ion to the multiply charged ion

(ETD) (de Hoffman and Stroobant, 2007). IRMPD fragments ions by directing IR laser

pulses at the ions, this increases the energy in the bonds until they fragment (de Hoffman

and Stroobant, 2007; Gross, 2017). The most commonly used fragmentation method is

CID, which fragments ions by having them collide with non-reactive gas molecules such

as He or $N_2$ (de Hoffmann and Stroobant, 2007; Gross, 2017). This fragmentation takes

place in the linear ion trap (Michalski et al., 2012).

### 1.12.2  Fragmentation Mechanism Overview

The following is a brief overview of some of the more common and well-described

fragmentation mechanisms in MS/MS. Simple cleavages are either radical site initiated

(α-cleavage) or charge site initiated (inductive cleavage) (McLafferty & Turecek, 1993;

Gross, 2017; de Hoffman and Stroobant, 2007). Radical site-initiated fragmentation

occurs in odd electron (OE) ions, where an electron is either removed (positive mode) or

added (negative mode) to ionize the molecule. In an alpha cleavage, the odd electron is

donated to form a new bond with an adjacent atom. This causes the cleavage of the other

bond to this atom, leading to a neutral loss. Inductive cleavage happens at the site of the

charge and can apply to even electron ions. Unlike the alpha cleavage, inductive cleavage

attracts an electron pair from a bond to the charge site, leading to the cleavage of that

bond. The ability of a charge site to do this is influenced by the electronegativity of the

atom, halogens have the strongest tendency to do this, followed by oxygen and sulfur,

with nitrogen and carbon being the least likely (McLafferty & Turecek, 1993). There are

also rearrangement fragmentation pathways such as the McLafferty rearrangement where an unpaired electron can be donated through space, forming a new bond on an adjacent atom. The McLafferty rearrangement is a radical site-initiated process, although there are similar processes for charge site-initiated rearrangements.

Negative ions are often even electron ions and contain less energy than positive ions, causing them to produce fewer fragments. A review by Bowie et al. (1990) highlighted four major fragmentation pathways for even electron negative ions. One was a simple homolytic cleavage with the loss of a radical causing the formation of a stable radical anion. The second was the formation of an anion complex, which can undergo several reactions such as elimination processes and direct displacement of the anion. The third was a reaction where the initial deprotonated ion doesn't really fragment, but after a proton transfer a new anion is formed that can fragment. There is also the possibility of rearrangement reactions. Other identified pathways for negative ions are charge site driven fragmentation and charge remote fragmentation (Stroobant et al., 1995). The charge site driven fragmentation is analogous to the charge site driven fragmentation of a positive ion and the charge remote fragmentation (CRF) occurs when the charge is unable to migrate to the point of fragmentation, which is more prevalent when an ion has increased unsaturation (Stroobant et al., 1995). CRF can also occur for positive ions (de Hoffman and Stroobant, 2007).

### 1.12.3 Interpretation of MS/MS Fragmentation

The fragments generated by these mechanisms can then be used to interpret the original structure. For complex mixtures such as organic aerosol, it is difficult to isolate peaks for

fragmentation. So, many peaks are fragmented simultaneously. To complicate matters even further, each measured precursor mass is likely a mixture of isomers (Zark et al., 2017). So, the fragment ions may originate from many different structures. Nonetheless, analysis of the neutral losses can be useful because it can provide information about the functional groups. Some neutral losses, such as $H_2O$, can be lost from either a hydroxyl or carboxyl group, which makes it harder to determine a particular structure (Kerwin et al., 1996; Leenheer et al., 2001). The water loss from a carboxylic acid can occur by removal of the hydroxyl group (Jensen et al., 1985; Bowie et al., 1990; Kerwin et al., 1996) or by having two carboxyl groups interact with one another to form a five or six membered ring that leads to the elimination of $H_2O$ (Leenheer et al., 2001; Witt et al., 2009). For both pathways, the loss of $H_2O$ from a carboxylic acid is expected to be favored for aliphatic molecules relative to more olefinic or aromatic molecules (Kerwin et al., 1996; Leenheer et al., 2001). The loss of $CO_2$ is more specific to carboxylic acids (Leenheer et al., 2001; LeClair et al., 2012). Other commonly observed neutral losses are $C_2H_4O$, CO, $CH_4O$, $C_2O_4$, and $H_4O_2$. The $C_2H_4O$ neutral loss is likely related to a ketone carbonyl functional group, $CH_4O$ is likely a methoxy group, while $C_2O_4$ and $H_4O_2$ represent double losses of $CO_2$ and $H_2O$ respectively. While the CO loss is often related to carbonyl functional groups, it can also be produced from phenols via multistep rearrangements (Gross, 2017). The large number of ways that CO can be produced makes it difficult to use for structure identification (Gross, 2017). These functional groups can be used to improve model estimates of aerosol hygroscopicity (Clegg et al., 2019 *in review*).

## 1.13  Data Analysis, Preparation, and Molecular Formula Assignment

### 1.13.1  Mass Calibration

Before the raw data can be used to draw conclusions, it must be analyzed. The most

powerful aspect of ultrahigh resolution MS is that it can obtain exact mass measurements

that allow the analyst to determine molecular formulas. There are many considerations

that must be taken before assigning a molecular formula to a mass. One of the key

considerations is the mass accuracy of the measurement. While ultrahigh resolution MS

instruments are capable of measuring exact masses, if they are not carefully calibrated the

masses may not be within the error tolerance of the theoretical mass of the compounds.

The first step of ensuring mass accuracy is performing an external calibration of the

instrument measurements prior to analysis. This is typically done with a mix of

compounds from the instrument manufacturer or a custom calibration solution, such as

arginine clusters (Schmitt-Kopplin et al., 2010). The initial external calibration is critical

to obtaining good mass accuracy in measurements, but it is not typically sufficient to get

mass accuracy down to the parts-per-billion level, which is preferred for highly complex

mixtures such as petroleum and aerosol. Mass error is typically reported in ppm for

measurements with ultrahigh resolution MS and is calculated using Equation 1:

$$ppm\ error = \frac{Th\_mass - Exp\_mass}{Th\_mass} * 10^6 \qquad \textbf{Eq.1}$$

where *ppm error* is the mass error, *Th_mass* is the theoretical exact mass of the molecular

formula, and *Exp_mass* is the experimentally measured mass of the peak assigned with

the formula. In order to get the required mass accuracy additional internal recalibration

techniques must be applied. One method uses an added compound of a known mass and identity and then is used to correct the mass spectrum based on the mass shift observed for that peak. This method is not the best option for complex mixtures because the spiked compound may distort the abundances of the analyte peaks due to ionization competition (Zhang et al., 2011) and also because the mass shifts are not the same across the entire spectrum (Savory et al., 2011), which makes it unlikely that a range of masses will be recalibrated correctly. A common method of doing recalibration in complex mixtures is to choose a series of peaks that are commonly present in a sample and use those to recalibrate the spectrum (Savory et al., 2011). These recalibrant peaks can be contaminants in the sample or any other known series that can easily be identified. There are a variety of methods that have been developed to recalibrate mass spectral data from ultrahigh resolution MS instrumentation. Many of these methods were developed for FT-ICR MS (Wong et al., 2006; Savory et al., 2011) largely because it has been around longer than the Orbitrap, but some more recent studies have developed recalibration methods that were specifically designed for Orbitrap (Gorshkov et al., 2010), while others are applicable to both Orbitrap and FT-ICR (Kozhinov et al., 2013).

### 1.13.2 Noise Estimation

Another key aspect of data preparation for analysis is to determine the noise level of the signal. If the noise is over or underestimated there is the potential for analyte peaks to be removed (over estimation) or for noise peaks to be included as analytes (underestimation) (Riedel and Dittmar, 2014). There have been many methods reported for evaluating the noise level of a sample (Riedel and Dittmar, 2014; Zhurov et al., 2014; Kilgour et al.,

2017; Zielinski et al., 2018). Some simple methods to estimate the noise involve finding

an extended range of the mass spectrum that does not show any analyte signal and then

using that range to calculate the average noise (Kew et al., 2017). This method is easy to

implement, but it can lead to incorrect assessments of the noise level because the noise is

not necessarily the same across the entire spectrum (Hawkes et al., 2016), in fact a study

has found that the region that is most often free of analytes also has the highest noise

level for FT-ICR (Hawkes et al., 2016). Some other methods use histograms of the

intensity values of the raw mass spectrum, which can show a clear separation between the

low intensity "noise" peaks and the high intensity analyte peaks (Zhurov et al., 2014;

Zielinski et al., 2017). Another reported method uses the average of the noise for multiple

blank samples run under the same conditions as the test sample to estimate the noise level

(Riedel and Dittmar, 2014). Another method for noise estimation uses isotopic patterns to

identify the noise level in a segmented way across the entire mass spectrum (Kilgour et

al., 2017). Still other methods use the mass defect of the masses in order to pull out the

peaks that are most likely to be noise peaks and then averages them in order to provide an

estimate of the noise (Riedel and Dittmar, 2014; see also Section 5.3). All of these

methods have advantages and disadvantages in terms of implication and effectiveness,

but one of them should be used to remove the noise in a mass spectrum before attempting

to assign molecular formulas. More discussion of these noise estimation methods is in

Chapter 5 of this dissertation.

### 1.13.3  Molecular Formula Assignment

After the noise estimation and recalibration of the mass spectrum has been performed molecular formulas can be more confidently assigned to the measured masses. There are a multitude of commercial, open source, and other reported methods for molecular formula assignment (Kujawinski and Behn, 2006; Tolic et al., 2017; Leefmann et al., 2019; Green and Perdue, 2015; Stranz, 2015; Kunenkov et al., 2009; Tziotis et al., 2011; PetroOrg, 2014; Schum et al., 2019). The most simple method of molecular formula assignment is to take the experimental mass and try every combination of atoms within limits set by the user to determine a molecular formula that matches the mass. This brute force method will eventually get the correct answer, but it will also generate a large number of molecular formulas that do not make any chemical sense.

*1.13.3.1 Quality Assurance Parameters*

The most egregious incorrect formulas can be removed by a variety of tests and rules such as the Senior valence rules (Senior, 1951; Kind and Fiehn, 2007; Green and Perdue, 2015), the nitrogen rule (Kind and Fiehn, 2007), the DBE-oxygen rule (Herzsprung et al., 2014), and atomic ratios (oxygen-to-carbon, hydrogen-to-carbon, etc.). Three Senior rules are used for determining chemical feasibility of a molecular formula, one is that the sum of valences or total number of odd valence atoms is even, the second is that the sum of valences must be greater than or equal to the maximum valence times two, and the third is that the sum of valences is greater than or equal to the number of atoms times 2 minus 1 (Kind and Fiehn, 2007). The valences refer to the number of open holes in the outermost layer of electrons, so for example carbon typically has a valence value of 4.

Some atoms such as sulfur can often have valences anywhere from 2 to 6 depending on the oxidation of the sulfur atom, so previous studies have used the largest valence values to try to avoid removing formulas that are otherwise chemically feasible (Kind and Fiehn, 2007). The nitrogen rule states that any molecule with an odd number of nitrogen atoms must have an odd neutral mass (Kind and Fiehn, 2007). The DBE minus oxygen rule (DBE-O) was used to separate formulas for data collected in the negative mode into possible, unlikely, and unthinkable molecular formulas (Herzsprung et al., 2014). The typical range of possible molecular formulas was from $-10 \leq DBE\text{-}O \leq 13$. Using these rules and others a large portion of the chemically non-feasible assignments can be removed. More description of these rules and others is provided in Chapter 5.

*1.13.3.2 Formula Extension*

Mathematical calculations are at the core of all the methods that are used to assignment molecular formulas, but many use formula extensions in order to lessen the computational burden and to lower the chances of incorrect assignments. A Formula extension uses a seed formula with a known molecular formula to assign other masses that are related to it through chemical series. These chemical series can be identified by using a Kendrick mass defect (KMD) analysis with a specific base unit. A common base unit is $CH_2$, the exact mass of which can be used to renormalize the measured masses so that when masses vary only by the number of $CH_2$ (14.01565 Da) they have the same KMD This is analogous to how masses are typically normalized to $^{12}C$ having the exact mass of 12. To support the KMD value, a z* value can be calculated (Hsu et al., 1994; Stenson et al., 2003). This value is the nominal mass series value for the mass. If z* and

KMD are used in conjunction with each other, any masses that vary by $CH_2$ must have the same value of both. This relationship means that if any mass within the series is assigned a molecular formula, the other masses can be assigned a molecular formula by adding or subtracting $CH_2$ units from the assigned formula. Any repeating mass unit can be used as a formula extension, some that have been used in molecular formula assignment packages include O, $CH_2O$, $H_2$, $H_2O$, and $CO_2$ (Stranz, 2015; Tolic et al., 2017; Schum et al., 2019) Typically, the masses at the low end of the series are directly assigned molecular formulas and then the assignment is extended to the rest of the masses in the series. The primary reason for assigning the low masses is that as the mass increases, the number of possible molecular formula assignments increases exponentially, even when chemical feasibility tests are applied (Kind and Fiehn, 2007, Koch et al., 2007).

*1.13.3.3 Formula Assignment Methods*

While many formula assignment methods use a mathematical calculation to get preliminary assignments, some also use database matching as a method to assign molecular formulas (Kujawinski and Behn, 2006; Tolic et al., 2017; Leefmann et al., 2018). Database matching means that the assignment method has a list of exact masses with their formulas and it will check the experimental masses against the masses within its database and if there is a match, it will assign that as the molecular formula for the mass. This can be useful for increasing the speed of assignment because it negates a large portion of the time-consuming calculation, but it is limited to molecular formulas in the database. Large databases can limit the impact of this, because as more atoms and larger

masses are put into a database, it can become unmanageably large and impractical to use (Leefmann et al., 2018). Instead of brute force methods, some studies have developed algorithms that perform calculations, but are able to avoid the majority of the loops required to try all the difference combinations of molecules by assigning the formulas using low mass moieties such as $CH_4O_{-1}$ and $C_4O_{-3}$ (Green and Perdue, 2015; Perdue and Green, 2015). A detailed discussion of this method is provided in Chapter 5.

## 1.14 Organization of the Dissertation

The first chapter of this dissertation is the introduction. The second chapter contains the instrumental methods, parameters, and equations used for Chapters 3 and 4. Chapters 3-5 describes the results of three projects that comprised the majority of my doctoral work which was focused on the analyzing the molecular composition of aerosol samples collected at the Pico Mountain Observatory, and the development of software tools for data preparation and molecular formula assignment for ultrahigh resolution mass spectrometry data.

The first project (described in Chapter 3) is the analysis of three organic aerosol samples (PMO-1, PMO-2, PMO-3) that were collected at the Pico Mountain Observatory, this study has already been published (Schum et al., 2018). The samples in this study were analyzed using FT-ICR MS, ion chromatography (IC), and an organic carbon/elemental carbon analyzer (OC/EC). FT-ICR MS was used to determine the molecular composition of the samples, while IC and OC/EC were used to determine bulk anion/cation and organic carbon concentrations respectively. The key observation obtained from FT-ICR

MS analysis was that the overall extent of oxidation of PMO-1 and PMO-3 aerosol was lower than expected based on their atmospheric transport time. The study hypothesized that the transport of the sample through the free troposphere, which is generally cool and dry, contributed to an increased viscosity of the organic aerosol and subsequently increased its resistance to oxidation, as has been suggested by other studies (Koop et al., 2011; Lignell et al., 2014; Hinks et al., 2016). In contrast, PMO-2 was transported through the boundary layer and was evaluated to determine if aqueous processing was a contributing factor to its higher observed oxidation despite its shorter transport time. The ion concentrations obtained from ion chromatography were used as supplemental data for identifying the sources of the samples and markers of potential aqueous processing. Additionally, FLEXible PARTicle (FLEXPART) model analysis was performed by a collaborator for these samples. The model provided the transport times, possible emission sources, and the ambient temperature and RH, which was used to predict the glass transition temperature using the equations from DeRieux et al., (2018).

The second project (described in Chapter 4) is a deeper analysis of PMO-1 using a segmented scanning approach with tandem MS/MS using an FT-ICR MS. The segmented scanning approach was used to improve the instrument sensitivity and provide functional group information molecular species present in organic aerosol. This study highlights an unprecedented mass spectral complexity resulting from the segmented scanning technique. Some implications about the composition of biomass burning organic aerosol are presented in this work, including the importance of non-oxygen heteroatoms and highly aromatic molecular species. The other major aspect of this chapter is the

identification of functional groups using MS/MS fragmentation. The type and number of functional groups is discussed, but the main focus is relating the observations of functional groups to the molecular aspects of the base molecule (e.g. aromaticity, oxidation, size). This analysis was largely done using principle components analysis (PCA) with the functional group loss types as the variables. The observations from this analysis were used to interpret relationships between certain types of neutral losses and also some potential environmental interactions.

The third project (described in Chapter 5) is the development of a software package for ultrahigh resolution MS data preparation and molecular formula assignment. The package is called MFAssignR and is written in R and has been released on GitHub (Schum et al., 2019). The package contains functions for noise estimation, isotope identification, recalibration, and molecular formula assignment. This package represents one of a few open source packages that contain all these functions necessary for the analysis of ultrahigh resolution MS data, typically this pipeline is only available in commercial software (Leefmann et al., 2018). Chapter 5 will discuss the theory behind the functions, how they work, and the literature that led to their development, along with demonstrations of the function outputs. Chapter 6 provides the main conclusions and implications for the studies aerosol and ultrahigh resolution MS data analysis.

# 2 Methods

## 2.1 Pico Mountain Observatory

The Pico Mountain Observatory (PMO) (also denoted as PICO-NARE, OMP, and PIC in various studies) is located on Pico Island in the Azores archipelago. The observatory itself is in the summit caldera of Pico Mountain 2225 m above sea level and was established by the late Dr. Richard Honrath and colleagues in 2001. The boundary layer height in this region is typically 500-2000m during the summer (Kleissl et al., 2007; Rémillard et al., 2012; Zhang et al., 2017) which places the observatory within the free troposphere for the majority of the summer (Zhang et al., 2017). The ability to do long term sampling within the free troposphere is nearly unique to this site, in fact a recent study suggested that PMO is one of the best free tropospheric sampling mountain sites in the world (Collaud Coen et al., 2018). The ability to collect aerosol in the free troposphere is important because the characteristics of the free troposphere are different than the boundary layer in terms of temperature, relative humidity, and oxidant concentrations, which all have an impact on aerosol aging and transport. The Azores-Bermuda anticyclone causes persistent downward mixing from the upper free troposphere and stratosphere, which allows the collection of aerosol at those levels of the atmosphere on occasion (Zhang et al., 2017). A study using the FLEXible PARTicle dispersion model (FLEXPART) by Zhang et al., (2017) has shown that this site is most often influenced by North American outflow (30-40%). More specifically, in the summer months 15% of the air masses that intercept PMO are North American anthropogenic outflows, while 7.3% are influenced by wildfires (Zhang et al., 2017). The relative

prevalence of these types of outflows make PMO an excellent site to study the effects of long-range transport on North American gas and aerosol emissions (Val Martin et al., 2008a). Several studies at PMO have investigate outflows of $NO_x$, $NO_y$, $CH_4$, non-methane hydrocarbons, and $O_3$ gases (Val Martin et al., 2006; Pfister et al., 2006; Val Martin et al., 2008a; Val Martin et al., 2008b; Helmig et al., 2015). Fewer have investigated aerosol at the site (Fialho et al., 2005; China et al., 2015; Džepina et al., 2015; China et al., 2017), with only one (Džepina et al., 2015) looking into the molecular and chemical composition of aerosol, prior to the study described here (Schum et al., 2018). An image of the mountain and the station itself can be seen in Figure 2.1.
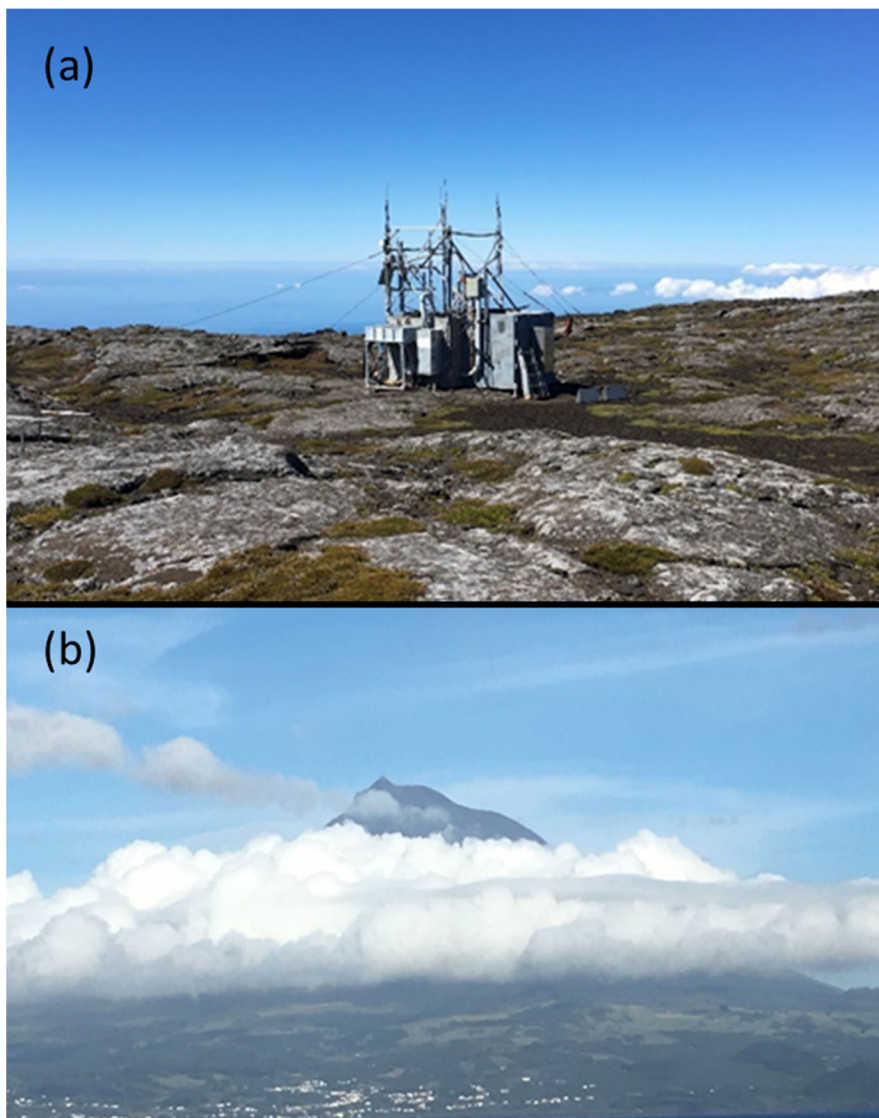
**Figure 2.1.** Photos of the research station on the summit caldera of Pico Mountain (a) and Pico Mountain from neighboring Faial Island with the mountain summit above the cloud layer (b).

## 2.2 Sample Collection

PM$_{2.5}$ samples were collected at PMO on 8.5 x 10 in. quartz fiber filter using high volume air samplers (EcoTech HiVol 3000, Warren, RI, USA) operated at an average volumetric flow rate of 84 m$^3$ hr$^{-1}$ for 24 h. Prior to sampling, the filters were wrapped in clean,

heavy-duty aluminum foil and baked at 500 $^0$C for ~ 8 hours to remove organic artifacts associated with the filters. Afterward, they were placed in antistatic sealable bags until deployment. We deployed four air samplers at the site, each was set up with a filter simultaneously and programmed to start one day after another, allowing for continuous sample collection for up to four consecutive days. This procedure was used to maximize the number of filters collected. Daily visits and maintenance were prohibited by the time consuming and strenuous hike necessary to reach the site. The sampled filters were removed and returned to the same aluminum wrapper and bag. The samples were then brought down the mountain and stored in a freezer until cold transport back to Michigan Tech where they were stored in a freezer until analysis. Three samples, collected in consecutive years at PMO, on 27-28 June 2013, 5-6 July 2014, and 20-21 June 2015 were analyzed in this study. The sampling time for all samples was 24 hours; on 27-28 June the sampling began at 19:00, on 5-6 July and on 20-21 June the sampling began at 15:00, all local times.

## 2.3   Sample Preparation

### 2.3.1   FT-ICR MS

The samples for FT-ICR-MS analysis were selected based on the organic carbon concentration. Selected samples typically had more than 1000 μg of organic carbon per quartz filter. Sample preparation was described in detail in previous studies from our group (Mazzoleni et al., 2010, 2012; Zhao et al., 2013; Džepina et al., 2015). Briefly, one quarter of the quartz filter was cut into strips, placed in a pre-washed and baked 40 mL glass vial, and then extracted using ultrasonic agitation in Optima LC/MS grade

deionized water (Fisher Scientific, Waltham, MA, USA) for 30 minutes. The extract was then filtered using a pre-baked quartz filter syringe to remove undissolved material and quartz filter fragments. The sample filter was then sonicated again in 10 mL of Optima LC/MS grade deionized water for 30 minutes, filtered, and then added to the original 30 mL of filtrate yielding a total of 40 mL. Ice packs were used during the sonication to ensure the water temperature stayed below 25 °C. The water-soluble organic carbon (WSOC) compounds were then isolated using Strata-X (Phenomenex, Torrance, CA, USA) reversed phase solid phase extraction (SPE) cartridges to remove inorganic salts that can adduct with organic compounds during electrospray ionization. During the reversed phase SPE, losses of highly water soluble, low molecular weight (MW) and hydrophobic, high MW organic compounds are expected. Thus, the resulting WSOC is the SPE-recovered fraction. The cartridges were pre-conditioned with acetonitrile and LC/MS grade water before the 40 mL filtrate was applied to the cartridges at a rate of ~ 1 mL/min. The cartridges were eluted with 2 mL of an aqueous acetonitrile solution (90/10 acetonitrile/water by volume) and stored in the freezer until analysis. The procedural loss of ionic low MW compounds such as oxalate can lead to an underprediction of the organic aerosol O/C and overprediction of the average glass transition temperatures ($T_g$). To investigate this, we used the concentrations of the prominent organic anions measured with ion chromatography to estimate the abundance of these compound relative to the compounds detected by FT-ICR MS. The low MW organic anion corrected average O/C values correlated with the trends of the original O/C values, however the significance of impacts varies with the measured analyte concentrations and the assumptions associated with the uncertain mass fraction of the molecular formula composition (Table SM4).

When low MW organic anions were included in the estimated average dry $T_g$ values, they dropped by $\leq 2.5$ %, which was deemed relatively insignificant. More discussion of this can be seen in section 2.7

### 2.3.2 OC/EC

Sample preparation consisted of taking 16 mm diameter circular punches from the sample filters and putting them in the OC/EC instrument. This was done at least three times for each sample.

### 2.3.3 Ion Chromatography

The samples were prepared using the California Air Resource Board method (California Environmental Protection Agency, 2011). Briefly, five square punches of 3.98 cm$^2$ each were taken from each filter and placed into a pre-cleaned 15 mL disposable centrifuge tube, to which 100 µL of isopropanol was added to help dissolve the less soluble organic species. Finally, 12 mL of 18.2 MOhm deionized water from an Easy Pure water system (Barnstead, ThermoFisher Waltham, MA, USA) were added to each centrifuge tube. These samples were then sonicated for 60 minutes with blue ice added to the sonication bath to keep the temperature below 25 $^0$C. Once sonicated, the samples were stored in the refrigerator overnight and transferred with 0.45 µm nylon syringe filters (Fisher Brand, Waltham, MA, USA) and sterile 3 mL syringes (BD, Franklin Lakes, NJ, USA) to 5 mL IC vials (Thermo Scientific, Waltham, MA, USA) the following day. The samples were then run on the IC system.

## 2.4  Instrument and Quality Assurance Parameters

### 2.4.1  FT-ICR General Parameters and Quality Assurance

Ultrahigh resolution mass spectrometric analysis was done using FT-ICR MS with ESI at the Woods Hole Oceanographic Institution (Thermo Scientific LTQ Ultra). The samples were analyzed using direct infusion ESI in the negative ion mode. Negative polarity is effective for the deprotonation of polar organic molecules (Mazzoleni et al., 2010), which are expected to dominate the organic aerosol mass fraction and were the focus of this study. The spray voltage ranged from 3.15 to 3.40 kV depending on the ionization stability with a sample flow rate of 4 to 5 µL/min. We used a scan range of $m/z$ 100 – 1000 with a mass resolving power of 400,000 (defined at $m/z$ 400) for all samples. The samples were run in duplicate and 200 transient scans were collected. The transients were co-added for each replicate run using the Midas Co-Add tool and molecular formula assignments were made using Composer software (Stranz, 2015), as described in previous studies (Mazzoleni et al., 2012; Džepina et al., 2015).

The resulting molecular formula assignments underwent additional quality assurance (QA) data filtering to remove chemically unreasonable formulas with respect to O/C, hydrogen to carbon ratio (H/C), double bond equivalent (DBE), and absolute PPM error as described in the Supplemental Information of Putman et al. (2012). Molecular formulas in common with the instrument blanks with signal intensity ratios < 3 were removed; meanwhile analytes in common with the field blanks with signal intensity ratios < 3 were flagged. Specifically, two formulas ($C_{17}H_{34}O_4$ and $C_{19}H_{38}O_4$) observed in PMO-1 could not be classified as pertaining only to the field blank and so they were not

removed. The quality assurance (QA) was performed to eliminate improbable molecular formulas and possible contaminants for each of the ultrahigh resolution FT-ICR mass spectra consistent with previous studies from our group (Putman et al. 2012; Mazzoleni et al. 2012; Džepina et al. 2015). First, molecular formulas with extremely high or low O/C (>2, <0.1), H/C (>2.2, <0.3), and DBE (>20) are removed. All of the known solvent contaminant peaks and isolated assignments that are not associated with a $CH_2$ homologous series are removed. Blank subtraction was done by finding the ratio of intensities between a sample and a blank and then formulas with intensities < 3x larger than in the blank are removed or flagged. After this is done, replicate analyses of the samples are evaluated and only the formulas that are present in both replicates are retained. To produce the final data set for each sample, the replicates were aligned and only the molecular formulas found in both replicates after QA were retained.

### 2.4.2 FT-ICR Tandem MS/MS

Scan windows were set every 5 m/z from m/z 165 to 465, and m/z 520 with a width of 6 m/z, allowing 0.5 m/z overlap between each scan to ensure complete coverage and address edge effect to some extent as describe in Kiddo-Soule et al. (2010). Overall, a total range of ~300 m/z was collected. Scan windows in different mass ranges were collected using different resolving powers, m/z 165 to 295 was collected in 100K (@ m/z 400) resolving power, m/z 300 to 415 was collected in 200K, and m/z 420 to 465 (and 520) was collected in 400K. This was done to increase the scan rate at the lower mass ranges where fewer masses were expected, and then to have higher resolving power at the higher mass ranges where the complexity of the unfragmented and fragmented scans was

expected to increase. The fragmentation method used was collision induced dissociation (CID) with fragmentation energies of 25-35 depending on how the scan range fragmented. For each mass window, 55 scans were collected and then co-added using the MIDAS Co-Add tool before being assigned formulas using Composer. The assignment parameters were 1-100 carbon, 4-200 hydrogen, 1-60 oxygen, 0-3 nitrogen, and 0-1 sulfur. The abundance cut off was set at 10 times the signal to noise. After formula assignment quality assurance steps to remove the unlikely formulas were done according to the parameters previously used in other studies (Putman et al., 2012; Džepina et al., 2015). Additionally, the DBE minus oxygen cut, described by Herzsprung et al. (2014) was employed to remove as many questionable formulas as possible.

### 2.4.3 Organic and Elemental Carbon Analysis

For each sample collected, a minimum of three circular 16 mm diameter filter punches were analyzed. If all three punches had consistent organic carbon concentrations (relative standard deviation, RSD < 15 %), the average value was used to determine the total loading of OC on the filter and in the air during the sampling period. If the replicates were inconsistent, more replicates were analyzed until at least three were consistent. Elemental carbon measurements were also obtained with this instrument, but in nearly all cases they were below the detection limit (~0.1 μgC/filter punch), so those values are not reported.

### 2.4.4   Ion Chromatography

PMO samples were also analyzed for major anions and cations using ion chromatography (IC). Anion analysis was performed using a Dionex ICS-2100 instrument (Thermo Scientific) with an AS-17-C analytical and guard column set (Thermo Scientific) using a KOH generator for gradient elution.  The gradient elution had the following steps: -5 – 0 min., Equilibrate, 1 mM KOH, 0 – 15 min., Isocratic, 1 mM KOH, 15 – 20 min., Ramp, 1 – 10 mM KOH, 20 – 30 min., Isocratic, 10 mM KOH, 30 – 40 min., Ramp, 10 – 20 mM KOH, 40 – 45 min., Isocratic, 20 mM KOH, 45 – 55 min., and Ramp, 20 – 40 mM KOH. Cation analysis was performed using a Dionex ICS-1100 instrument with CS-12A analytical and guard column set (Thermo Scientific) and an isocratic 20 mM methanesulfonic acid eluent. The instruments were operated in parallel using split flow from autosampler. After the ion concentrations were determined, they were background subtracted using field blanks from PMO.

### 2.5   Data Analysis

All analysis of data was done using custom scripts written in the R programming language (R Core Team, 2017). Packages used to perform data analysis and visualization are dplyr, (Wickam et al., 2018), ggplot2 (Wickam, 2016), tidyr (Wickam and Henry, 2018), colorRamps (Keitt, 2012), lemon (Edwards, 2018), cowplot (Wilke, 2017) ggbiplot (Vu, 2011), and base R functions. The PCA analysis described in Chapter 4 was done with the base R function prcomp() and the visualization was done with the ggbiplot and ggplot2 packages.

### 2.5.1 PCA Analysis

A long-standing method for investigating the correlation between different data sets is

principal components analysis (PCA), which has been used in the past for looking into

the correlation between different samples (Zhao, 2014, Wozniak et al., 2014). PCA

reduces the dimensions in a multivariate data using a linear combination of variables to

create new variables, called principal components (PC). These PC each represent a

certain amount of the variance in the sample with PC1 accounting for the most variance,

PC2 for the second most, etc. There are the same number of PC as there are variables in

the data set, but typically the first 2 or 3 PC are enough to account for the vast majority of

the variance in the data set can be accounted for. The amount of variance each PC

accounts for is related to their eigenvalue which is determined based on the sum of

squared distance from the origin for each of the observations on the line of best fit for the

data set. The eigenvalues for each of the PC (which are eigenvectors) are added up and

the percentage of the total each eigenvalue makes up determines the percentage of

variance each PC accounts for. A visualization of the variance from each PC is the Scree

plot, which shows the amount of variance each PC makes up, an example Scree plot can

be seen in Figure 2.2. Determining how many PC are important for representing the

sample as accurately as possible is facilitated with this plot because the decrease in

variance is observed visually. In general, "elbow" of the plot is where the number of

important PC values can be determined. With the PC determined, the point can be plotted

in a 2-D plot with PC1 and PC2 (PC3, PC4, etc.) on each axis and the correlations

between the different data points can become apparent. The loadings of each variable can

be added to this plot to make a so-called PCA biplot. The loadings represent the

importance of each variable in the variance accounted for by each PC, the greater the

distance for the origin the greater the effect. An example biplot can be seen in Figure 2.3,

for this plot the loadings are represented by arrows, so the longer the arrow, the more

effect on variance. The angle between points is also important, if two arrows are in the

same direction that means that they are correlated, if the arrows are at 90° from each

other, they are not correlated, and if the arrows are at 180° they are negatively correlated.
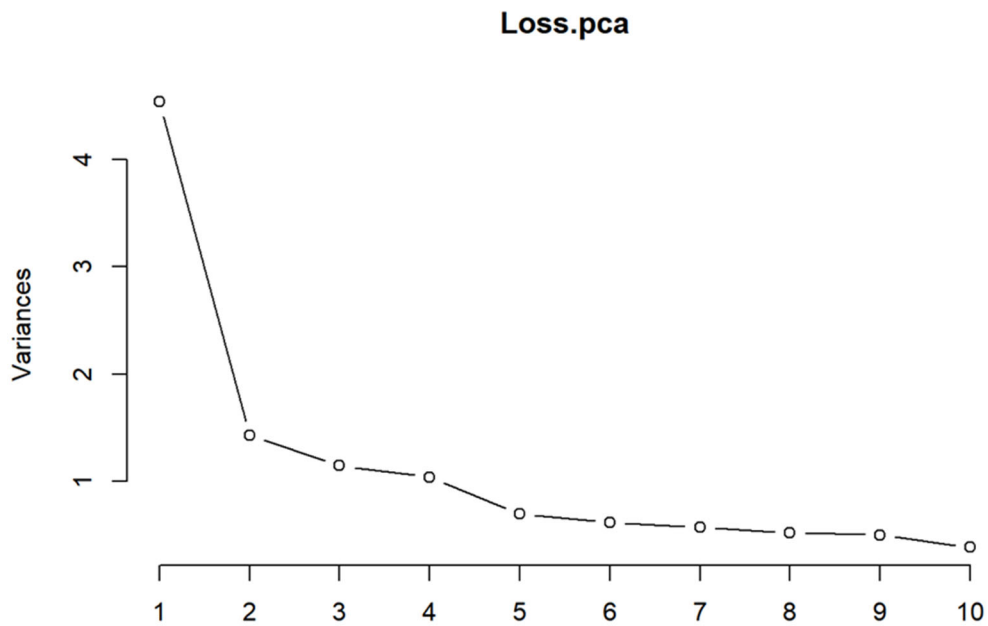


**Figure 2.2.** Example Scree plot for 10 PC values. The "elbow" of this plot can be observed at PC2.
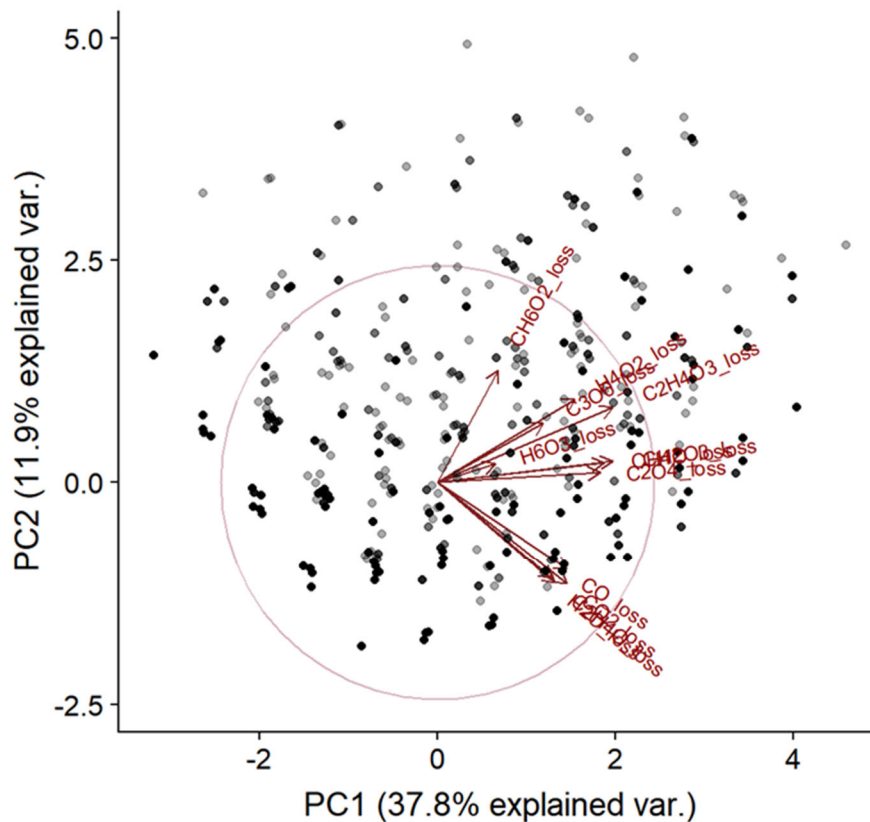
Figure 2.3. Example PCA biplot using PC1 and PC2.

## 2.6 FLEXPART Numerical Simulations

Retroplume analysis was conducted using the Lagrangian particle dispersion model

FLEXPART. (Seibert and Frank, 2004; Stohl et al., 2005; Owen and Honrath, 2009). In

this study we report three specific events, one that took place on June 27 (19:00) -28

(19:00), 2013 (PMO-1), one on July 05 (15:00) - 06 (15:00), 2014 (PMO-2), and one on

June 20 (15:00) - 21 (15:00), 2015 (PMO-3). FLEXPART was used to determine the

sources, ages, and transport pathways of the aerosol samples collected at PMO.

FLEXPART backward simulations (also called retroplumes) were driven by meteorology

fields from the Global Forecast System (GFS) and its Final Analysis with 3-hour

temporal resolution, 1° horizontal resolution, and 26 vertical levels. The output was saved

in a grid with a horizontal resolution of 1° latitude by 1° longitude, and eleven vertical

levels from the surface to 15,000 m a.s.l. For each simulation, 80 thousand air parcels

were released from the receptor and transported backwards for 20 days to calculate a

source-receptor relationship (in units of s kg-1, Seibert and Frank, 2004). FLEXPART

retroplumes were then multiplied with CO emission inventories (kg s-1) from the

Emissions Database for Global Atmospheric Research (EDGAR version 3.2 (Olivier and

Berdowski, 2001)) and the Global Fire Assimilation System (Kaiser et al., 2012) to

estimate the influence from anthropogenic and wildfire sources, respectively. The

FLEXPART CO tracer calculated with this approach indicates the relative contributions

from anthropogenic and biomass burning emissions. Since CO chemistry and dry

deposition are not considered in the FLEXPART setup, the absolute FLEXPART CO

value does not reproduce the actual CO concentrations at Pico. FLEXPART does not

consider the background CO accumulated in the atmosphere. The difference between

FLEXPART CO and the actual CO largely depends on these factors. In previous

applications of this approach, FLEXPART CO was able to estimate the episodes of CO

enhancement due to transport of emissions (e.g., Brown et al., 2009; Stohl et al., 2007;

Warneke et al., 2009). This approach has been used in several PMO studies and

successfully captured elevated CO periods (e.g., Džepina et al., 2015; Zhang et al., 2014;

2017) and it is used here to assist in the interpretation of the chemical composition in this

work. In addition to the typical FLEXPART simulations done for the site (e.g.,

retroplume, CO source apportionment), we extracted the ambient temperature and

relative humidity (RH) from the GFS analysis data for model grids along the FLEXPART

simulated transport pathways. These parameters were then used to estimate the glass

transition temperatures (Tg) of the organic aerosol components during transport, based on

its molecular composition from ultrahigh resolution MS, using estimation methods

recently developed by Shiraiwa et al. (2017a) and extended to higher masses by DeRieux

et al. (2018).


## 2.7   Error Sensitivity Tests

### 2.7.1   Estimation of the SPE Sample Preparation Effect on the Total WSOC Properties

As mentioned in the main paper, the solid phase extraction of WSOC results in the loss of

some low molecular weight (MW) polar organic species and some high MW nonpolar

species. Since the low molecular weight species are studied using ion chromatography,

the major ions and their concentrations are known. Specifically, oxalate, formate, and

acetate which are the most abundant can cause an underprediction of the average O/C

value and an overprediction of average $T_g$ values for a sample. In an effort to constrain

the potential impact of this effect on our results, we used the concentrations of 5 organic

acids detected by ion chromatography, along with their O/C values, and Boyer-Kauzmann

rule ($T_g = g*T_m$, $g = 0.7$) (Shiraiwa et al., 2017a; DeRieux et al., 2018) estimated $T_g$

values to estimate the weighted and unweighted average O/C and $T_g$ values for the three

samples if the organic acids were included.

To estimate the contribution of each organic anion to the overall organic mass, the mass

concentrations were normalized by the organic mass (OM) concentration. The organic

mass concentration was estimated by multiplying the measured OC concentration by 2,

consistent with El-Zanan et al. (2005). The mass fractions were then multiplied by the

sum of the total ion abundance in the mass spectrometry data (using assumption that they

made up 50, 70, or 100% of organic mass) in order to roughly determine the potential ion

abundances relative to those that were studied after SPE. While the results of the negative

mode ESI do not likely represent the entirety of all organic species in the aerosol extracts,

the polar compounds expected in water extracts are best ionized by this method and are

expected to make up the majority of species (< 70%; Samburova et al., 2013). The

estimated relative abundance of the low MW anions was used in conjunction with their

O/C and $T_g$ values to calculate a new weighted average for all of the compounds available

for each sample. Using this approximation, we found that the addition of these small

compounds had a very minor impact on the $T_g$ values because they decreased by < 2.5 %

at most. The percentage increase of the O/C values for the three samples varied from 20

% for PMO-1 to 42% for PMO-2. The ion weighted O/C value for PMO-1 is 0.58, for

PMO-2, 0.81, and for PMO-3 is 0.57. The results confirm our observation that PMO-2

has much higher average O/C than PMO-1 or PMO-3, and that PMO-1 and PMO-3 have

low O/C, considering their transport time in comparison to other studies (Bougiatioti et

al., 2014). Based on these results, we conclude that although the loss of some low MW

compounds using SPE can cause an under or an over prediction of some values, the effect

in the case of the samples in this study is minimal and does not change the implications

of our observations. Tables 2.1-2.5 contain the values described in this discussion.

**Table 2.1.** The concentrations of the ions used for the estimation and the organic mass (OM) concentration. The values are in $\mu g/m^3$ air.

| Ion | PMO-1 | PMO-2 | PMO-3 |
|---|---|---|---|
| **Formate** | $0.0289 \pm 0.0003$ | $0.00438 \pm 0.00007$ | $0.0119 \pm 0.0001$ |
| **Acetate** | $0.0519 \pm 0.0001$ | $0.004587 \pm 0.000005$ | $0.0071 \pm 0.0002$ |
| **Oxalate** | $0.0938 \pm 0.00070$ | $0.0897 \pm 0.00181$ | $0.0522 \pm 0.00002$ |
| **Malonate** | $0.00605 \pm 0.0003$ | $0.00548 \pm 0.0007$ | $0.0045 \pm 0.0003$ |
| **Lactate** | $0.0292 \pm 0.0004$ | $0.0019 \pm 0.0001$ | $0.00467 \pm 0.0001$ |
| **OM** | $4.14 \pm 0.04$ | $0.956 \pm 0.052$ | $1.74 \pm 0.20$ |

**Table 2.2.** The percent mass fraction of each ion.

| Ion | PMO-1 | PMO-2 | PMO-3 |
|---|---|---|---|
| **Formate** | 0.698 | 0.458 | 0.684 |
| **Acetate** | 1.25 | 0.479 | 0.409 |
| **Oxalate** | 2.27 | 9.38 | 3.00 |
| **Malonate** | 0.146 | 0.573 | 0.259 |
| **Lactate** | 0.705 | 0.199 | 0.268 |

**Table 2.3.** The estimated $T_g$ values for the acid form of each ion as estimated using their melting points and the Boyer-Kauzmann rule with g = 0.7.

| Ion | $T_g$ |
|---|---|
| **Formate** | 197.1 |
| **Acetate** | 202.83 |
| **Oxalate** | 324.21 |
| **Malonate** | 285.6 |
| **Lactate** | 202.79 |

**Table 2.4.** Estimated average O/C values when the ions are considered. The table contains the results for 3 assumptions of the organic mass fraction represented by the FT-ICR MS identified species (100%, 70%, 50%). The numbers in parentheses show the percent change in average O/C from the O/C without ions considered.

| Sample | RA Weighted O/C without Ions (100%) | Ions and RA Weighted O/C (100%) | Ions and RA Weighted O/C (70%) | Ions and RA Weighted O/C (50%) |
|--------|-------------------------------------|---------------------------------|--------------------------------|--------------------------------|
| **PMO-1** | 0.48 | 0.53 (10.42%) | 0.55 (14.58%) | 0.58 (20.83%) |
| **PMO-2** | 0.57 | 0.70 (22.81%) | 0.75 (31.58%) | 0.81 (42.11%) |
| **PMO-3** | 0.45 | 0.52 (15.56%) | 0.54 (20.00%) | 0.57 (26.67%) |

**Table 2.5.** Estimated average $T_g$ values when the ions are considered. The table contains the results for 3 assumptions of the organic mass fraction represented by the FT-ICR MS identified species (100%, 70%, 50%). The numbers in parentheses show the percent change in average $T_g$ from the $T_g$ without ions considered. All $T_g$ values are in K.

| Sample | RA Weighted $T_g$ without Ions (100%) | Ions and RA Weighted $T_g$ (100%) | Ions and RA Weighted $T_g$ (70%) | Ions and RA Weighted $T_g$ (50%) |
|--------|---------------------------------------|-----------------------------------|----------------------------------|----------------------------------|
| **PMO-1** | 328.75 | 324.38 (1.33%) | 322.67 (1.85%) | 320.51 (2.51%) |
| **PMO-2** | 326.45 | 324.43 (0.619%) | 323.71 (0.839%) | 322.85 (1.10%) |
| **PMO-3** | 326.88 | 324.41 (0.756%) | 323.44 (1.05%) | 322.22 (1.43%) |

## 2.8 Equations

**Equation 2. 1.** The average oxidation state of carbon ($OS_C$) from Kroll et al. (2011). Sulfur and nitrogen can change the oxidation of the molecular species, which varies based on their oxidation state. In this case, sulfur and nitrogen are assumed to be fully oxidized.

Eq. 2. 1
$$OS_C \approx 2 * \frac{\#O}{\#C} - \frac{\#H}{\#C} - 5 * \frac{\#N}{\#C} - 6 * \frac{\#S}{\#C}$$

**Equation 2. 2.** The aromaticity index (AI) from Koch and Dittmar (2006; 2016). This is the most conservative method for calculating aromaticity as it assumes that all oxygen is in carbonyl groups. The threshold for olefinic species is $0 < AI \le 0.5$, for aromatic it is $0.5 < AI \le 0.67$, and for condensed aromatic it is $0.67 < AI \approx 1$. All other species are defined as AI = 0 making them aliphatic.

Eq. 2. 2
$$AI = \frac{1 + C - O - S - 0.5 * H - 0.5 * N}{C - O - S - N}$$

**Equation 2.3.** The modified aromaticity index ($AI_{mod}$) from Koch and Dittmar (2006; 2016). Here it is assumed that half of the oxygen is in carbonyl groups. The threshold for olefinic species is $0 < AI_{mod} \le 0.5$, for aromatic it is $0.5 < AI_{mod} \le 0.67$, and for condensed aromatic it is $0.67 < AI_{mod} \approx 1$. All other species are defined as $AI_{mod} = 0$ making them aliphatic.

Eq. 2. 3
$$AI_{mod} = \frac{1 + C - 0.5 * O - S - 0.5 * H - 0.5 * N}{C - 0.5 * O - S - N}$$

**Equation 2. 4.** The number of double bond (and rings) equivalents (DBE). C represents carbon, H represents hydrogen, X represents halogens, and N represents nitrogen. Elements with 2 covalent bonds (oxygen and sulfur) are cancelled out in this equation.

Eq. 2.4
$$DBE = \#C + 1 - \frac{\#H}{2} - \frac{\#X}{2} + \frac{\#N}{2}$$

**Equation 2. 5.** Estimation of the glass transition temperature from DeRieux et al., 2018. $n_C^0$ is the carbon reference number (12.13 ± 2.66), $b_C$, $b_H$, and $b_O$ are the contributions of each atom to $T_g$, and $b_{CH}$ and $b_{CO}$ represent the contribution of carbon-hydrogen and carbon-oxygen bonds respectively. For CHO molecular formulas the values for these terms are as follows: $b_C$ = 10.95 ± 13.60, $b_H$ = -41.82 ± 14.78, $b_{CH}$ = 21.61 ± 5.30, $b_O$ = 118.96 ± 9.72, and $b_{CO}$ = -24.38 ± 4.21. This equation determines the dry glass transition temperature. The Gordon-Taylor Equation (Eq. 2-6) is required to convert the glass transition temperature for non-dry conditions.

**Eq. 2. 5**

$$T_g = (n_C^0 + ln(n_C))\, b_C + ln(n_H)\, b_H + ln(n_C)\, ln(n_H)\, b_{CH} + ln(n_O)\, b_O + ln(n_C)\, ln(n_O)\, b_{CO}$$

**Equation 2. 6.** Gordon-Taylor Equation from DeRieux et al. (2018). This is used to calculate the glass transition temperature in humid conditions. $w_{org}$ is the mass fraction of organics, $T_{g,w}$ is the glass transition temperature for water (136 K), $k_{GT}$ is the Gordan-Taylor constant (assumed to be 2.5, consistent with DeRieux et al., 2018 and Shiraiwa et al., 2017a), and $T_{g,org}$ is the dry glass transition temperature calculated by Eq. S5. For more detail see DeRieux et al., 2018 and Shiraiwa et al., 2017a.

**Eq. 2. 6**

$$T_g(w_{org}) = \frac{(1 - w_{org})T_{g,w} + \dfrac{1}{k_{GT}}w_{org}T_{g,org}}{(1 - w_{org}) + \dfrac{1}{k_{GT}}w_{org}}$$

**Equation 2.7.** Adapted Gordon-Taylor equation with inputs for relative humidity (RH), dry glass transition temperature, and ambient temperature. This generates the phase state ratio (PSR), which predicts the phase the molecular species is likely in, PSR >= 1 is solid, PSR >= 0.8 & PSR < 1 is semi-solid, and PSR < 0.8 is liquid. This equation converts $w_{org}$ to a relative humidity dependent term as described in DeRieux et al. (2018) and Shiraiwa et al. (2017a), and converts $1/k_{GT}$ to 0.4, which is its value using the assumption of $k_{GT}$ equals 2.5. $T_{amb}$ is the ambient temperature. The relative humidity dependent $T_g$ is calculated using the calculation in the numerator.

**Eq. 2.7**

$$
\text{Phase State Ratio} = \frac{\left(\left(1 - \dfrac{1.4 - \frac{1.4*RH}{100}}{1.4 - \frac{1.28*RH}{100}}\right) * 136 + 0.4 * \left(\dfrac{1.4 - \frac{1.4*RH}{100}}{1.4 - \frac{1.28*RH}{100}}\right) * T_{g\,org}\right)}{\left(\left(1 - \dfrac{1.4 - \frac{1.4*RH}{100}}{1.4 - \frac{1.28*RH}{100}}\right) + 0.4 * \left(\dfrac{1.4 - \frac{1.4*RH}{100}}{1.4 - \frac{1.28*RH}{100}}\right)\right)}{T_{amb}}
$$

# 3 Physical and Molecular Composition of Long-Range Transported Free Tropospheric Organic Aerosol

## 3.1 Brief Introduction to Study

The first study to be described in this dissertation is the analysis and interpretation of 3 long range transported organic aerosol samples that were collected at the Pico Mountain Observatory (PMO) in the summers of 2013, 2014, and 2015. This study was published as Schum et al., (2018) and this chapter borrows heavily from that paper.

For a brief overview for the following chapter, three organic aerosol samples were collected at the Pico Mountain Observatory (PMO) in the summers of 2013, 2014, and 2015. The samples were collected on quartz filter paper for 24 hours each on June 27-28, 2013 (PMO-1), July 5-6, 2014 (PMO-2), and June 20-21, 2015 (PMO-3). The sampling time began at 19:00 for PMO-1, and 15:00 for PMO-2 and PMO-3, all times local. The samples were then stored in a freezer before being transported back to Michigan Tech, where they were again stored in a freezer until analysis. For this study the samples were analyzed with three different instruments, OC/EC analyzer (Model 4, Sunset Laboratory Inc. Tigard, OR, USA) for bulk organic carbon concentration, ion chromatography (IC) for bulk anion and cations concentration, and Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) to investigate the molecular composition of the samples. The sample preparation and parameters used for these instruments can be seen in Chapter 2. To get back trajectory and source region information we used FLEXPART modeling, which was performed by our collaborator Dr. Bo Zhang.

## 3.2   FLEXPART Retroplume Simulation Results

Representative FLEXPART retroplumes for the three samples are shown in Figures 3.1-3.3. PMO-1 (Figure 1) was largely influenced by North American outflow transported relatively high ($\geq 5$ km) in the free troposphere. Based on the FLEXPART carbon monoxide (CO) modeling (Figure 3.4), PMO-1 was impacted by wildfire emissions from Canada. The transport time for PMO-1 air masses from North America to PMO was about 7 days. The free tropospheric transport is likely due to the high injection heights (Val Martin et al., 2008a; 2010) of organic aerosol from wildfire events in northwestern Quebec (See Figures 3.4, 3.5). Similar events at PMO have been identified previously by (Val Martin et al., 2006; 2008a). The air masses intercepted during PMO-3 were North American outflows that traveled in the lower free troposphere across the Northern Atlantic Ocean to Western Europe before circling back to PMO. The transport time for the PMO-3 air masses from North America to PMO was roughly 10 days (Figure 3.2). After a northward transport to Western Europe in the jet stream during the first 4-5 days, the simulated plume turned to the south and west, arriving at PMO from Europe in about 2-4 days. This air mass was most likely influenced by wildfire emissions in western and central Canada (U.S. Air Quality, Smog Blog. alg.umbc.edu). Similar to PMO-1, FLEXPART CO source apportionment (Figure 3.4) suggests this sample was influenced by fire, although considering the OC concentration and transport time, it was much more diluted than what was observed in PMO-1. In contrast, the PMO-2 air masses traveled relatively low ($\leq 2$ km) over the "Rust Belt" (Illinois, Indiana, Michigan, Ohio, Pennsylvania, and New York) of the United States and stayed at approximately the same altitude until it reached the observatory 2-4 days later (Figure 3.3). This transport pattern

suggests that the aerosol was predominantly transported in the boundary layer on its way to the PMO and was primarily influenced by a mixture of continental U.S. anthropogenic and biogenic emissions. This was supported by the FLEXPART CO simulations as well (Figure 3.4). The height of the boundary layer over the continent generally ranges from 500-2500 m and is strongly affected by diurnal cycles, seasonal effects, and topography (Liu and Liang, 2010); overall, the continental boundary layer height generally increasing during the day and during the summer months. This suggests that PMO-2 was within the boundary layer over the United States.
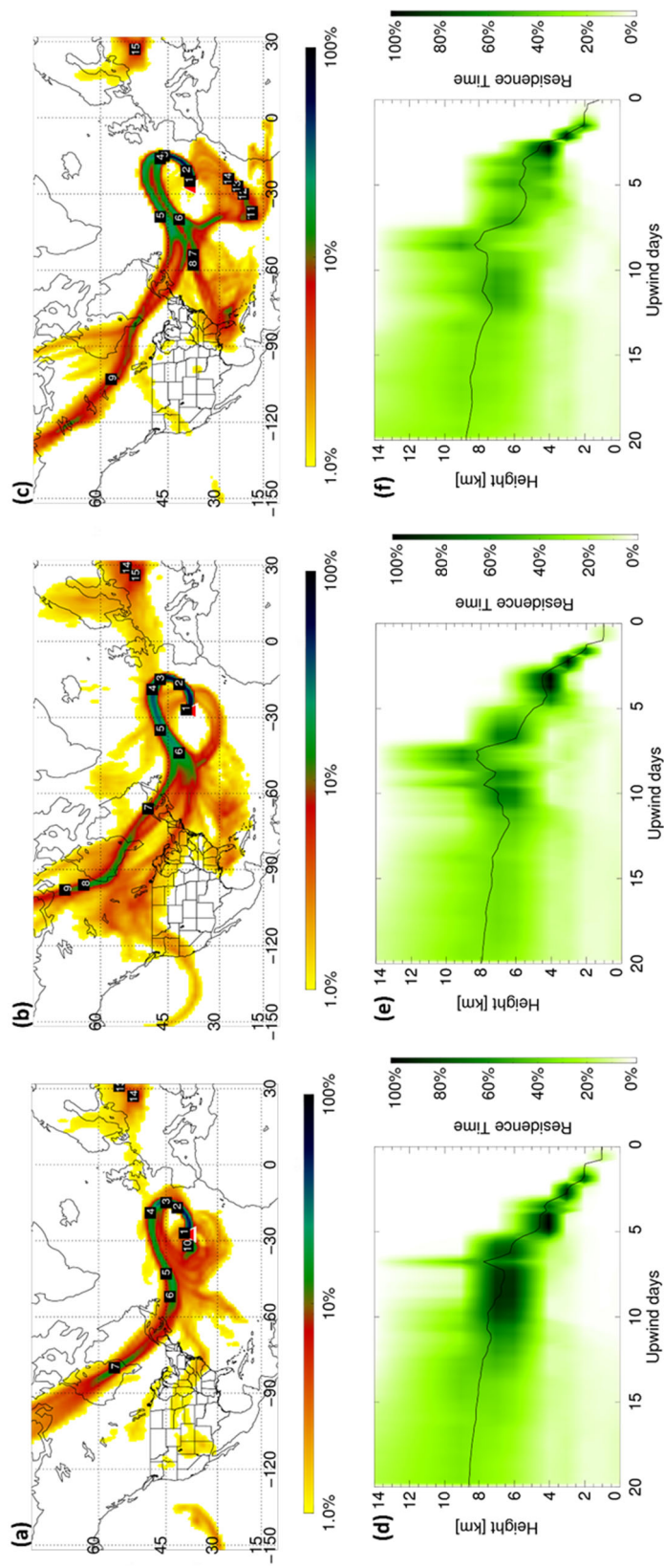
**Figure 3.1.** FLEXPART retroplumes the sampling period for PMO-1. June 27, 2013, 18:00 (a, d), June 28, 2013, 06:00 (b, e), and June 28, 2013, 18:00 (c, f).

62
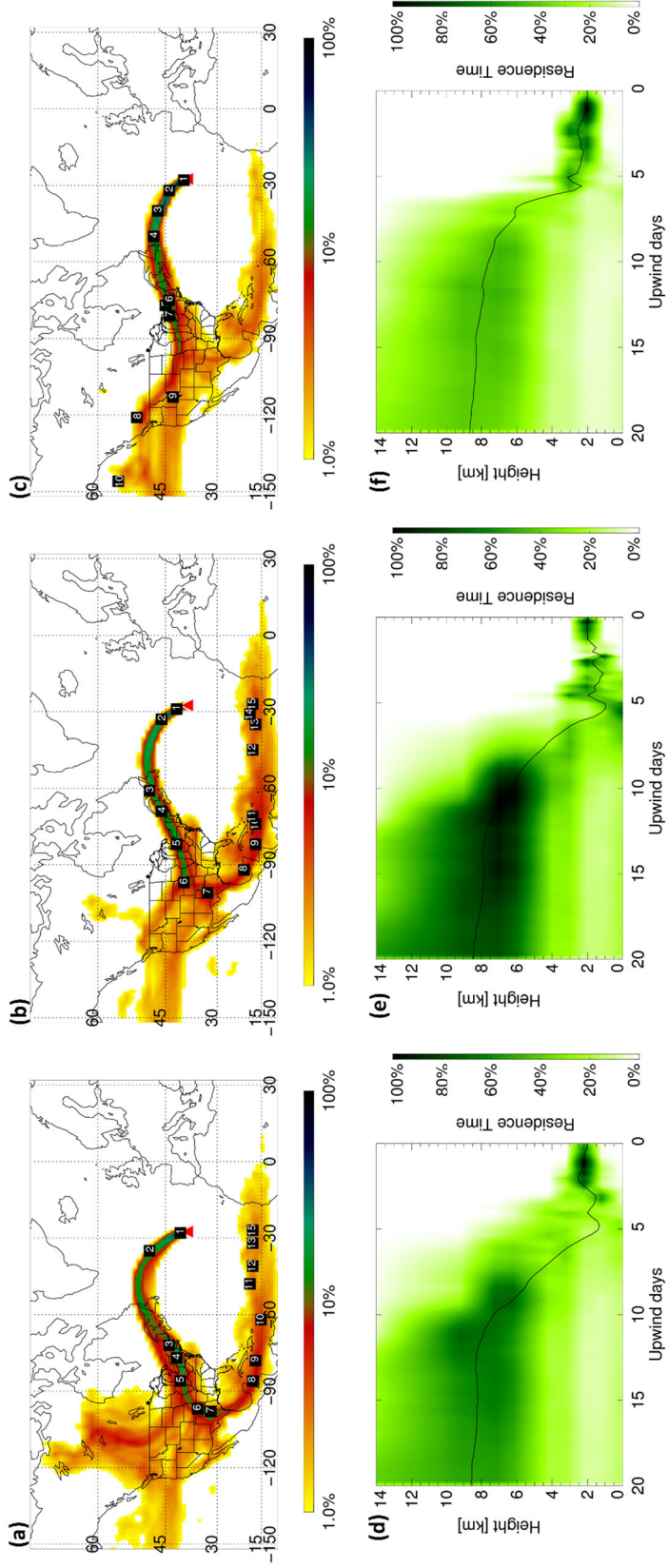
**Figure 3.2.** FLEXPART retroplumes the sampling period for PMO-2. July 05, 2014, 15:00 (a, d), July 06, 2014, 03:00 (b, e), and

July 06, 2014, 15:00 (c, f).

63
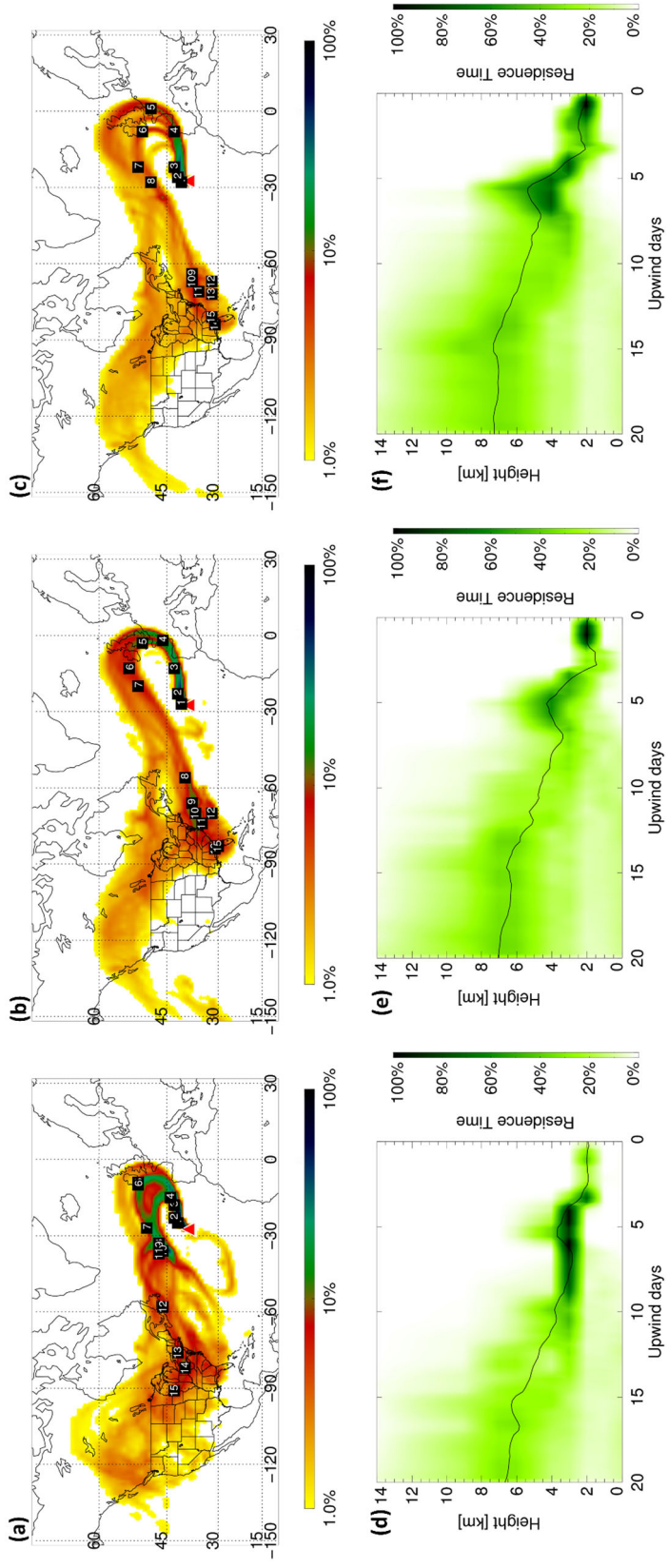
**Figure 3.3.** FLEXPART retroplumes the sampling period for PMO-3. June 20, 2015, 15:00 (a, d), June 21, 2015, 03:00 (b, e), and June 21, 2015, 15:00 (c, f).
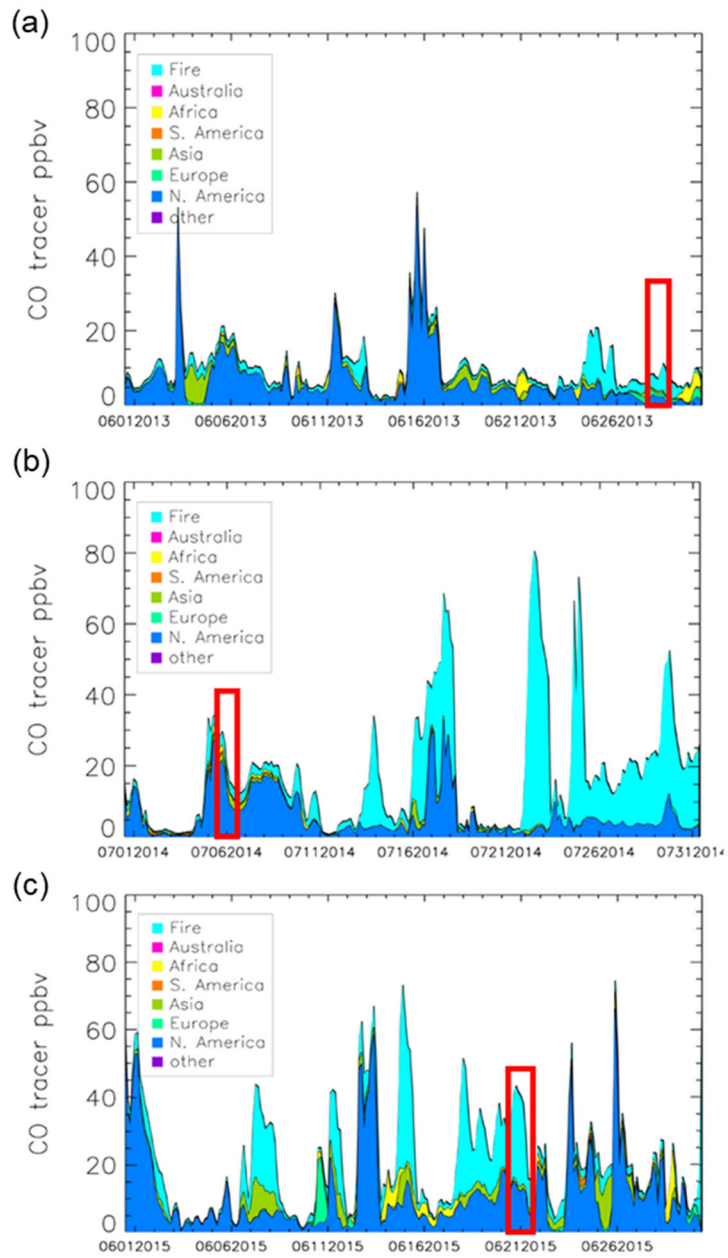
**Figure 3.4.** FLEXPART carbon monoxide source apportionment plot. PMO-1 (a), PMO-2 (b), PMO-3 (c). The red rectangle highlights the sampling period for each sample.
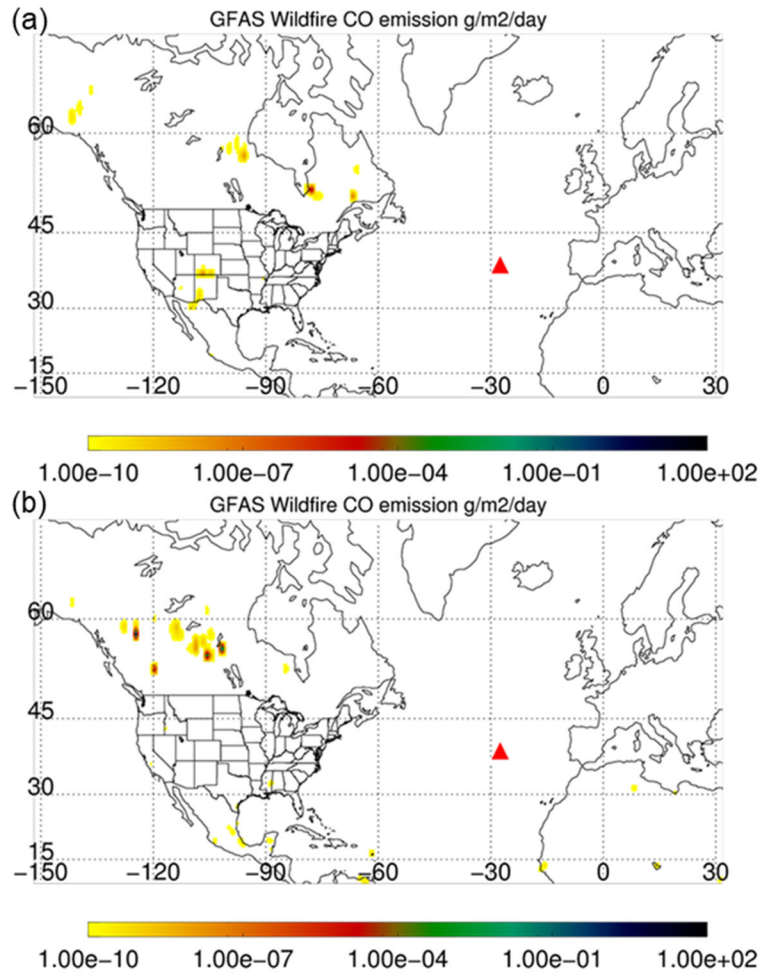
**Figure 3.5.** Wildfire emissions from GFAS dataset for the week corresponding to the PMO-1 event (a) and the PMO-3 event (b). Note the strong fire in western Quebec, which spatially coincides with the most likely path in the PMO-1 retroplume. Multiple fires in central and western Canada may have impacted PMO-3, although they are not spatially proximate to the most likely path in the PMO-3 retroplume.

## 3.3 Overview of the Aerosol Chemistry: OC/EC and IC

In this study, we present the detailed composition of three individual samples collected for 24 hours on 27-28 June 2013, 5-6 July 2014, and 20-21 June 2015 at the PMO. These samples, referred to as PMO-1, PMO-2 and PMO-3 hereafter, were selected after analysis of organic and elemental carbon (OC/EC) were performed for all 127 aerosol samples collected in this study. The three selected samples all had elevated organic carbon (OC) concentrations (Table 3.1) representing the capture of a pollution plume. After blank subtraction, the median OC of the samples collected over the summers of 2013-2015 was $0.16 \pm 0.018$ µg/m$^3$. The minimum OC level measured was lower than the average blank concentration and the maximum was $2.07 \pm 0.017$ µg/m$^3$ (PMO-1). The most abundant anions and cations in these samples are also shown in Table 3.1. The presence of these ions is consistent with the results of other studies (Yu et al., 2005; Aggarwal and Kawamura, 2009).

The concentrations of common anions and cations can offer important insight regarding cloud processing and emission sources (Table 3.1). Specifically, the elevated level of sulfate observed in the PMO-2 sample can be an indicator of anthropogenic influence, cloud processing, or marine influence (Yu et al., 2005). We also observed elevated oxalate concentrations in PMO-1 and PMO-2. Oxalate is known to co-vary with sulfate concentrations in the atmosphere when they are formed by aerosol-cloud processing (Yu et al., 2005; Sorooshian et al., 2007). Thus, the oxalate to sulfate ratio can be an indication of cloud processing (Sorooshian et al., 2007); in general, a higher ratio is the result of increased cloud processing. As described in Sorooshian et al. (2007), the oxalate

concentrations increase with cloud processing because there is more time for it to be produced, leading to an increased ratio. PMO-1 had the highest oxalate to sulfate ratio (0.278), followed by PMO-3 (0.124), and PMO-2 (0.084). The observed oxalate to sulfate ratios for these samples are all much higher than what was reported in Sorooshian et al. (2007) suggesting other factors may have impacted the ion concentrations. Specifically, an enrichment of oxalate from biomass combustion plumes (Cao et al., 2017) likely contributed to the observed concentrations of these ions in PMO-1 and PMO-3. The bulk concentration of oxalate in PMO-2 is similar to PMO-1, but the sulfate in PMO-2 is much higher, leading to a low oxalate to sulfate ratio. Based on FLEXPART simulations it is likely that PMO-2 underwent aqueous phase processing (see Sect. 3.5), but the high concentration of sulfate from possible anthropogenic and marine sources appears to have obscured the oxalate-sulfate relationship (Yu et al., 2005; Sorooshian et al., 2007).

Despite inconsistencies in the replicate potassium measurements for PMO-1, elevated potassium levels were observed, indicating contributions from biomass combustion (Duan et al., 2004). PMO-3 had slightly elevated potassium relative to PMO-2, but not as high as PMO-1. Chloride was also present in PMO-1 and PMO-3, which has been shown in some studies to be a minor product of biomass burning, depending on the fuel burned (Levin et al., 2010; Liu et al., 2017).

The nitrate levels in PMO-2 were significantly lower than what was observed in PMO-1 and PMO-3, which is consistent with the observation that the marine boundary layer promotes the rapid removal of $HNO_3$ (Val Martin et al., 2008b). This is also consistent with removal due to cloud scavenging (Dunlea et al., 2009). The elevated nitrate in PMO-

1 and PMO-3 is consistent with the observation of elevated $NO_y$ and $NO_x$ in the plumes of wildfire emissions made in previous studies at PMO (Val Martin et al., 2008a) and a lack of recent cloud scavenging (Dunlea et al., 2009).

Despite the low altitude transport, the major ion concentrations in PMO-2 do not strongly support a major influence from marine sources (Quinn et al., 2015; Kirpes et al., 2017). However, the increased concentration of methane sulfonic acid (MSA) in PMO-2 relative to PMO-1 and PMO-3 suggests some degree of marine influence. To estimate this, we used the non-background subtracted sodium concentration as an upper limit to estimate sea salt sulfate according the method described in Chow et al. (2015), this led to a maximum sea salt sulfate contribution of 25 %. The equation for this estimation can be seen in Equation 1:

$$nssSO_4^{2-} = [SO_4^{2-}] - 0.252 * [Na^+] \qquad \textbf{Eq. 1}$$

where $nssSO_4^{2-}$ is the estimated non-sea salt sulfate, $SO_4^{2-}$ is the measured sulfate concentration, and $Na^+$ is the measured sodium concentration. The influence of marine sources supports boundary layer transport. However, the results indicate that marine aerosol is not likely a major component of PMO-2, perhaps because the rate of PMO-2 transport was very fast.

**Table 3.1.** Average concentrations (μg/m³) of major ions and organic carbon.

| Component | PMO-1 | PMO-2 | PMO-3 |
|---|---|---|---|
| **Acetate** | $0.0519 \pm 0.0001$ | $0.004587 \pm 0.000005$ | $0.0071 \pm 0.0002$ |
| **Formate** | $0.0289 \pm 0.0003$ | $0.00438 \pm 0.00007$ | $0.0119 \pm 0.0001$ |
| **MSA** | $0$ | $0.00439 \pm 0.00006$ | $0.00232 \pm 0.00002$ |
| **Chloride** | $0.0104 \pm 0.0003$ | $0$ | $0.0310 \pm 0.0001$ |
| **Nitrate** | $0.189 \pm 0.002$ | $0.0173 \pm 0.0004$ | $0.3010 \pm 0.00028$ |
| **Sulfate** | $0.338 \pm 0.004$ | $1.07 \pm 0.01$ | $0.421 \pm 0.003$ |
| **Oxalate** | $0.0938 \pm 0.00070$ | $0.0897 \pm 0.00181$ | $0.05222 \pm 0.00002$ |
| **Sodium** | $0.2101 \pm 0.0004*$ | $0.2560 \pm 0.0001*$ | $0.548 \pm 0.005*$ |
| **Ammonium** | $0.1364 \pm 0.0004$ | $0.2394 \pm 0.00001$ | $0.1193 \pm 0.00062$ |
| **Potassium** | $0.0791 \pm 0.0020**$ | $0.0126 \pm 0.0002$ | $0.0197 \pm 0.0003$ |
| **OC** | $2.07 \pm 0.02$ | $0.478 \pm 0.026$ | $0.87 \pm 0.10$ |

*Sodium concentrations were not background subtracted due to inconsistent and high blank levels, they are included to provide an upper limit on the approximate sodium concentrations.

**Replicate measurements of potassium were inconsistent. The sample could not be re-analyzed because there was no remaining sample. Standard deviation was determined by looking at the average standard deviation of 36 potassium measurements in other samples. This sample should only be considered as elevated potassium and only minimal importance placed on the actual measured value.

## 3.4 Molecular Formula Oxidation Metrics: O/C and $OS_C$

In this section, we describe the detailed molecular formula composition of the three individual samples PMO-1, PMO-2, and PMO-3. Overall, nearly 80% of the observed mass spectral peaks in the ultrahigh resolution mass spectra were assigned molecular formulas, which is comparable to previous studies (Zhao et al., 2013; Džepina et al., 2015). After removing the duplicate molecular formulas containing $^{13}C$ or $^{34}S$, a total of 3168 (PMO-1), 2121 (PMO-2), and 1820 (PMO-3) monoisotopic molecular formulas remained. Groups of molecular formulas were assigned based on their elemental composition $C_cH_hN_nO_oS_s$, including: carbon, hydrogen, and oxygen (CHO); carbon, hydrogen, nitrogen, and oxygen (CHNO); and carbon, hydrogen, oxygen, and sulfur (CHOS). The most frequently observed compositions were CHO and CHNO. The reconstructed negative ion mass spectra of the monoisotopic molecular formulas for each of the samples are provided in Figure 3.6. Visual comparisons of the mass spectra indicate that PMO-2, which was likely transported through the North American continental and North Atlantic marine boundary layer, has an increased prevalence of higher O/C ratio formulas compared to the two samples transported through the free troposphere. Considering the ion distribution and normalized relative abundances, PMO-1 and PMO-3 mass spectra look quite similar with a high frequency of individual O/C values < 0.5. This may suggest similar emission sources or aerosol processing. In contrast, PMO-2 has a stronger relative influence of molecular compositions with higher O/C ratios. The O/C histogram plots in Figure 3.6 provide additional evidence for the O/C differences between the two types of samples (free troposphere and boundary layer)

71

due to the difference in the O/C distribution. A version of the mass spectrum with the full

y axis can be seen in Figure 3.7.
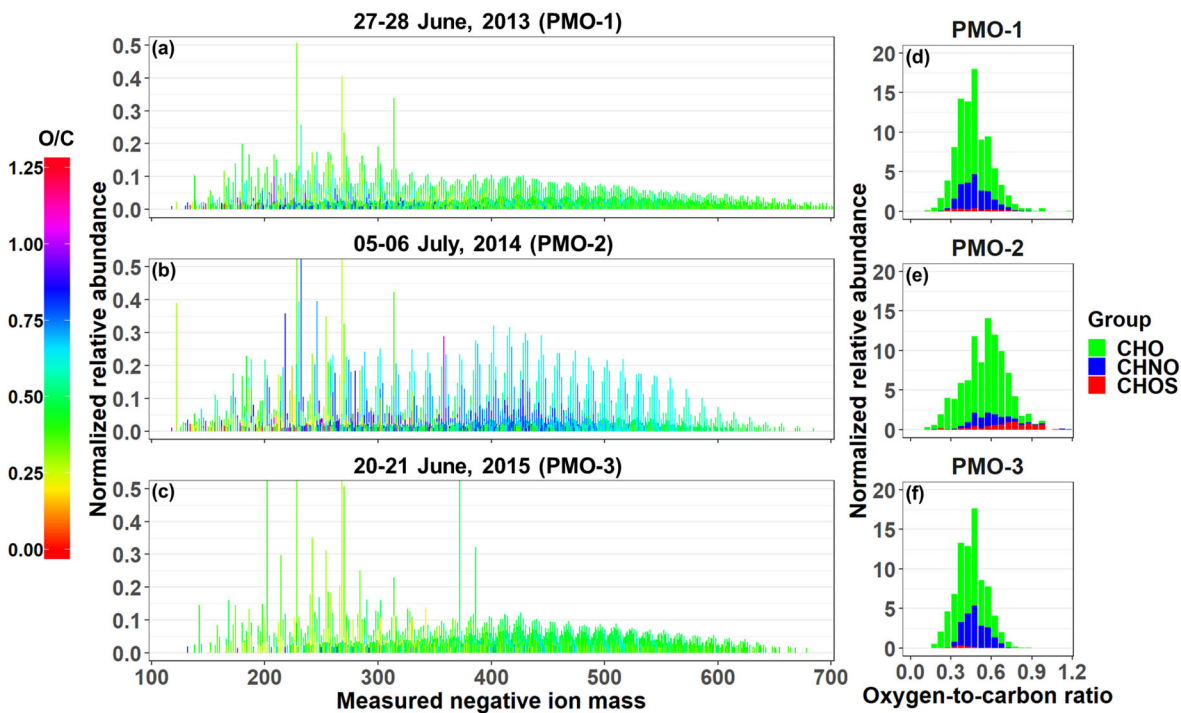


**Figure 3.6.** Reconstructed negative ion mass spectra (a-c) and O/C histograms (d-f) for the three PMO samples. The color in the mass spectra indicates the O/C value for the molecular formula it represents. The tallest peaks in the mass spectra exceed the range, this was done to improve the visibility of the lower abundance species (see also Figure 3.7).
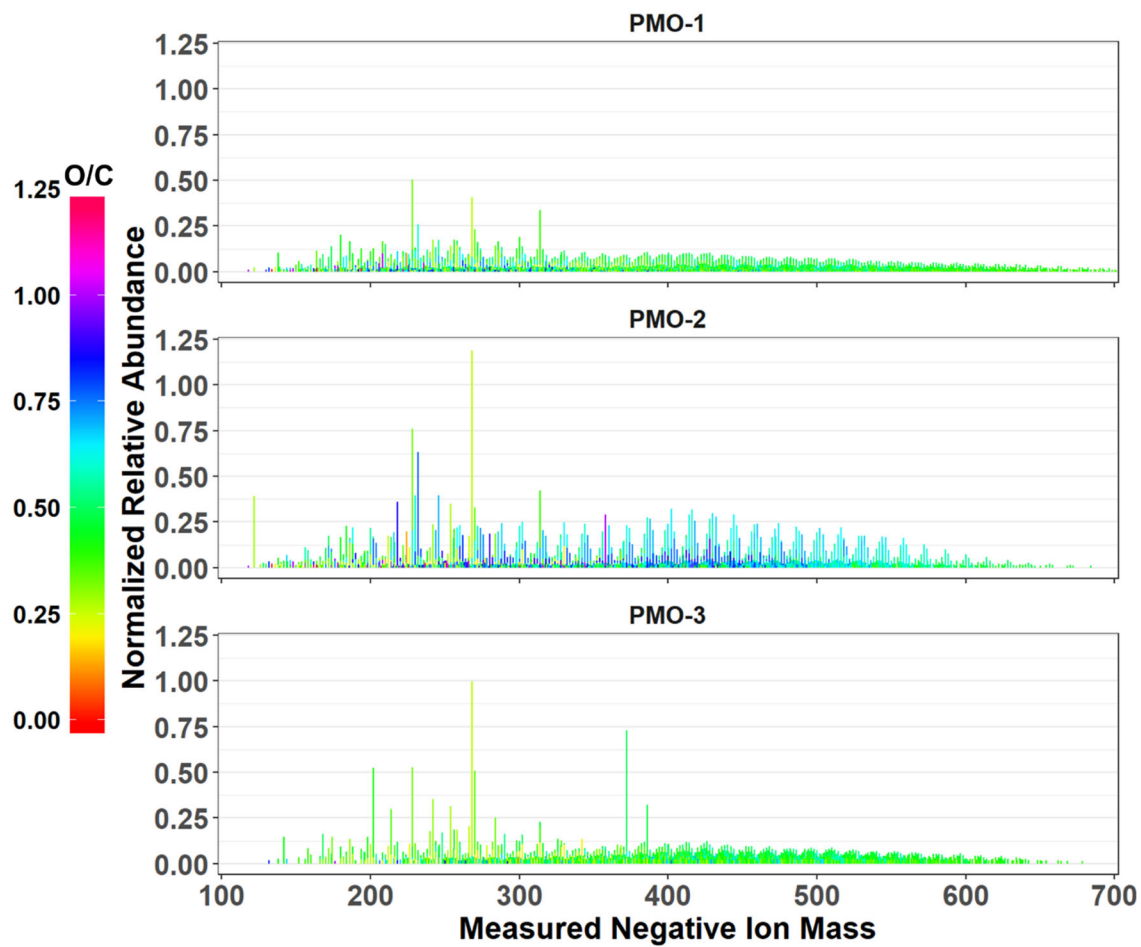
**Figure 3.7.** Reconstructed mass spectra showing the full abundance of all peaks. This is the same plot as Figure 3.6, just with fully expanded y axis.

The North American boundary layer outflow of organic aerosol captured in PMO-2 was likely influenced by SOA (Zhang et al., 2007) and thus is expected to have a higher initial O/C value compared to pyro-convected wildfire emissions of organic aerosol (e.g., Aiken et al., 2008; Jimenez et al., 2009; Bougiatioti et al., 2014). Although the initial compositions are unknown, we anticipated that the samples with longer transport times (~ 1 week for PMO-1 and PMO-3) would be at least similar or perhaps more oxidized than PMO-2 which had a much shorter transport time (~ 3 days). This expectation was based on literature describing secondary organic aerosol formation and aging (Volkamer et al., 2006; Jimenez et al., 2009) and the reported molecular composition of continental boundary layer aerosol (Mazzoleni et al., 2012; Huang et al., 2014). The lower oxidation observed in the free tropospheric samples transported for 7-10 days is consistent with our previous study at this site reported in Džepina et al. (2015). In fact, when we compared the molecular formula composition of the free tropospheric aerosol sample "9/24" from Džepina et al. (2015) to the free tropospheric samples in this study (PMO-1 and PMO-3), we observed that 86% and 91% of the molecular formulas are in common. FLEXPART simulations from both studies indicated these samples were all affected by wildfire emissions, contributing to their similarity. In contrast, only 75% of the formulas found in the boundary layer sample (PMO-2) were common with those in Džepina et al. (2015). These comparisons are provided in Table 3.2.

**Table 3.2.** Number of common species between this study and previous studies at PMO (Džepina et al., 2015) and SPL (Mazzoleni et al., 2012). The percentages indicate the percent of common species for the sample indicated by the row names.

| | PMO-1 | PMO-2 | PMO-3 | Džepina 2015 | Mazzoleni 2012 |
|---|---|---|---|---|---|
| **PMO-1** | X | 1697 (53.6%) | 1633 (51.5%) | 2730 (86.2%) | 1951 (61.6%) |
| **PMO-2** | 1697 (80.0%) | X | 1253 (59.1%) | 1585 (74.7%) | 1661 (78.3%) |
| **PMO-3** | 1633 (89.7%) | 1253 (68.8%) | X | 1704 (90.6%) | 1429 (76.0%) |

As observed in the mass spectra and histograms presented in Figure 3.6, the samples have noticeable differences in the distribution of O/C values. This is also reflected in the abundance weighted mean O/C values for the samples: $0.48 \pm 0.13$ (PMO-1), $0.57 \pm 0.17$ (PMO-2), and $0.45 \pm 0.11$ (PMO-3). Note that these O/C values are averages of thousands of individual measurements, as such the standard deviation represents the range of values and not uncertainties. A visual representation of the distribution of the O/C and OSc values is presented in Figures 3.8 and 3.9. These plots show the box plot distribution for each oxidation metric with a white box, the external "violin" component show the number density of formulas with those O/C or OSc values. The box plots show the median value with a horizontal black line within the "box" portion of the plot, the top of the box represents the 1st quartile, and the bottom of the box represents the 3rd quartile. The solid vertical line "whisker" represents Q3-1.5*interquartile range (IQR, Q3–Q1, maximum), and Q1-1.5*(IQR, minimum). We note that the relative abundance of compounds in ESI mass spectra is not directly proportional to their solution concentration, other factors including surface activity and polarity impact the ionization efficiency (Cech & Enke, 2001). Nonetheless, the abundance does differentiate trends

between the samples and the assigned molecular formulas represent a collection of multifunctional isomers (e.g., LeClair et al., 2012). For completeness, both the abundance weighted average values for various metrics of aerosol oxidation and saturation (Table 3.3) and the unweighted average values (Table 3.4) are reported.
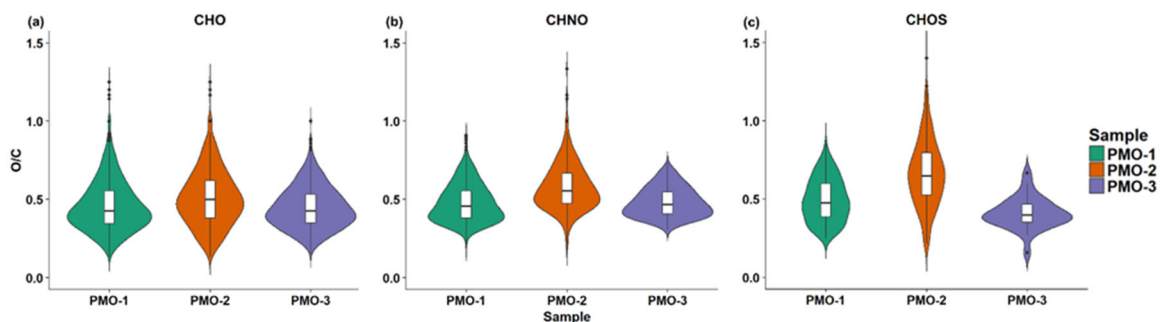


**Figure 3.8.** Violin plots showing the number distribution of species according to their O/C values, separated by molecular groups.
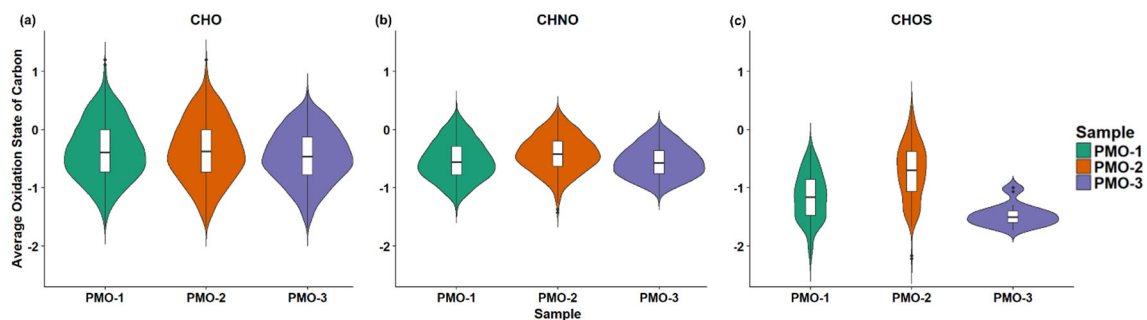


**Figure 3.9.** Violin plots showing the number distribution of species according to their $OS_C$ values, separated by molecular groups.

**Table 3.3.** Molecular formula composition with abundance weighted average values and the numbers of formulas for each elemental group.

| Sample | Group | $O/C_w$ | $H/C_w$ | $DBE_w$ | $OS_{Cw}$ | Number |
|--------|-------|---------|---------|---------|-----------|--------|
| **PMO-1** | All | $0.48 \pm 0.13$ | $1.30 \pm 0.28$ | $7.74 \pm 3.38$ | $-0.42 \pm 0.43$ | 3168 |
| **PMO-2** | All | $0.57 \pm 0.17$ | $1.36 \pm 0.22$ | $6.42 \pm 2.54$ | $-0.30 \pm 0.46$ | 2121 |
| **PMO-3** | All | $0.45 \pm 0.11$ | $1.34 \pm 0.41$ | $7.45 \pm 3.15$ | $-0.50 \pm 0.41$ | 1820 |
| **PMO-1** | CHO | $0.47 \pm 0.14$ | $1.31 \pm 0.29$ | $7.43 \pm 3.68$ | $-0.37 \pm 0.44$ | 1848 |
| **PMO-2** | CHO | $0.55 \pm 0.17$ | $1.35 \pm 0.25$ | $6.43 \pm 3.66$ | $-0.26 \pm 0.45$ | 1281 |
| **PMO-3** | CHO | $0.44 \pm 0.14$ | $1.37 \pm 0.31$ | $6.93 \pm 3.82$ | $-0.48 \pm 0.48$ | 1183 |
| **PMO-1** | CHNO | $0.49 \pm 0.15$ | $1.2 \pm 0.26$ | $9.44 \pm 3.09$ | $-0.50 \pm 0.3$ | 1120 |
| **PMO-2** | CHNO | $0.59 \pm 0.14$ | $1.25 \pm 0.19$ | $8.20 \pm 2.19$ | $-0.38 \pm 0.29$ | 561 |
| **PMO-3** | CHNO | $0.49 \pm 0.14$ | $1.23 \pm 0.21$ | $9.25 \pm 2.41$ | $-0.52 \pm 0.25$ | 608 |
| **PMO-1** | CHOS | $0.48 \pm 0.14$ | $1.78 \pm 0.35$ | $2.87 \pm 3.28$ | $-1.20 \pm 0.42$ | 200 |
| **PMO-2** | CHOS | $0.74 \pm 0.34$ | $1.57 \pm 0.23$ | $4.05 \pm 2.45$ | $-0.54 \pm 0.51$ | 274 |
| **PMO-3** | CHOS | $0.40 \pm 14$ | $1.90 \pm 0.47$ | $1.60 \pm 4.29$ | $-1.50 \pm 0.20$ | 29 |

77

**Table 3.4.** Unweighted average values with standard deviation and the numbers for each elemental group.

| Sample | Group | O/C | H/C | DBE | $OS_C$ | Number |
|---|---|---|---|---|---|---|
| **PMO-1** | All | $0.47 \pm 0.15$ | $1.29 \pm 0.32$ | $8.28 \pm 3.69$ | $-0.47 \pm 0.49$ | 3168 |
| **PMO-2** | All | $0.55 \pm 0.19$ | $1.37 \pm 0.27$ | $6.61 \pm 2.85$ | $-0.42 \pm 0.49$ | 2121 |
| **PMO-3** | All | $0.46 \pm 0.12$ | $1.33 \pm 0.27$ | $7.83 \pm 3.07$ | $-0.51 \pm 0.42$ | 1820 |
| **PMO-1** | CHO | $0.46 \pm 0.16$ | $1.28 \pm 0.32$ | $8.16 \pm 3.68$ | $-0.36 \pm 0.50$ | 1848 |
| **PMO-2** | CHO | $0.51 \pm 0.18$ | $1.37 \pm 0.28$ | $6.59 \pm 2.79$ | $-0.36 \pm 0.53$ | 1281 |
| **PMO-3** | CHO | $0.45 \pm 0.13$ | $1.36 \pm 0.28$ | $7.34 \pm 3.11$ | $-0.46 \pm 0.45$ | 1183 |
| **PMO-1** | CHNO | $0.48 \pm 0.12$ | $1.21 \pm 0.24$ | $9.40 \pm 3.08$ | $-0.53 \pm 0.34$ | 1120 |
| **PMO-2** | CHNO | $0.58 \pm 0.15$ | $1.26 \pm 0.18$ | $8.03 \pm 2.27$ | $-0.42 \pm 0.32$ | 561 |
| **PMO-3** | CHNO | $0.48 \pm 0.09$ | $1.24 \pm 0.19$ | $9.08 \pm 2.33$ | $-0.55 \pm 0.27$ | 608 |
| **PMO-1** | CHOS | $0.50 \pm 0.14$ | $1.77 \pm 0.19$ | $3.04 \pm 1.61$ | $-1.16 \pm 0.41$ | 200 |
| **PMO-2** | CHOS | $0.67 \pm 0.26$ | $1.61 \pm 0.25$ | $3.89 \pm 2.12$ | $-0.72 \pm 0.49$ | 274 |
| **PMO-3** | CHOS | $0.41 \pm 0.10$ | $1.90 \pm 0.21$ | $1.72 \pm 1.22$ | $-1.47 \pm 0.19$ | 29 |

Another way to compare the oxidation differences between the samples is with a difference mass spectrum (Figure 3.10). A difference mass spectrum compares two samples by subtracting the abundance of a molecular formula in one sample from the abundance in another, if a particular formula is more abundant in one sample versus the other the peak will show up in that half of the plot. Figure 3.10 shows that even if some of the higher O/C molecular formulas are found in PMO-1 and PMO-3, they are all more abundant in PMO-2, supporting the conclusion that PMO-2 is likely more oxidized than the other samples. Additionally, the difference mass spectrum demonstrates that PMO-1 and PMO-3 have a lot of similarity, with PMO-1 having slightly more of the high O/C species.
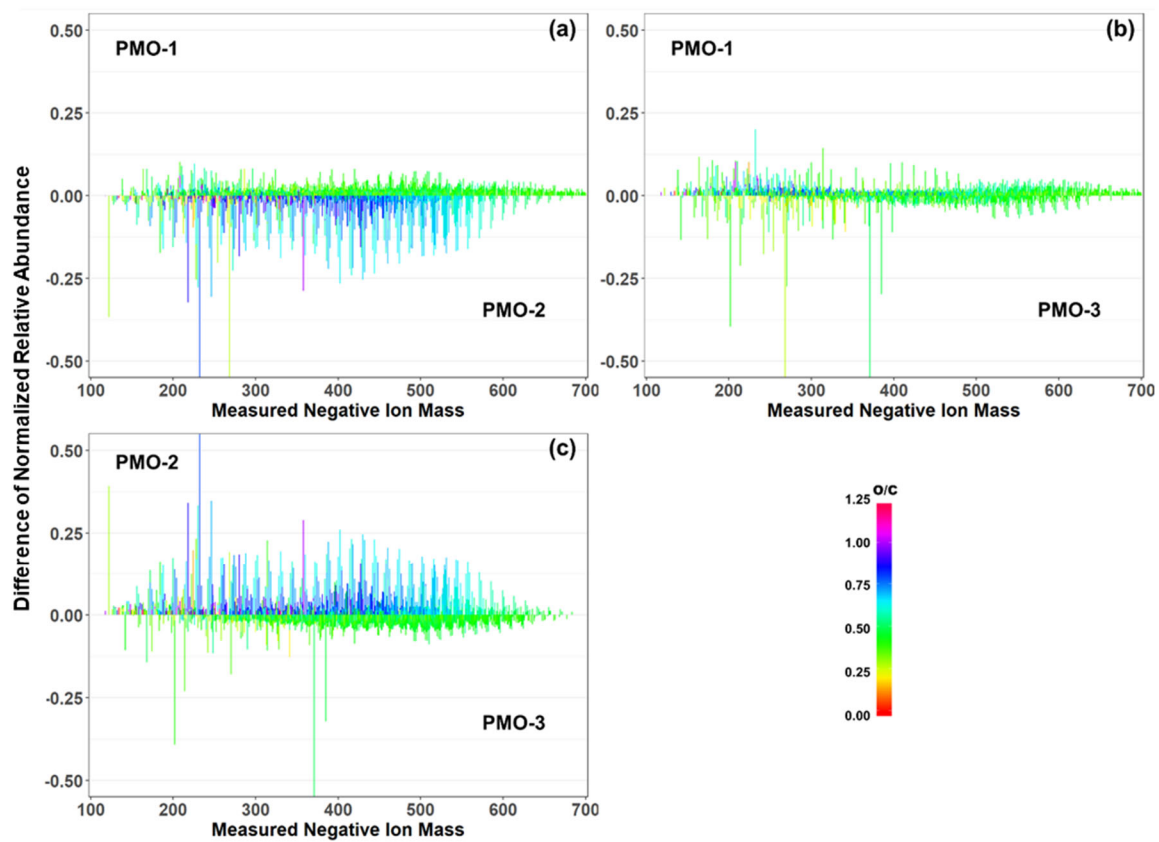
**Figure 3.10.** Difference mass spectra comparing the three PMO samples. The species more abundant in one sample or another are elevated in the correspondingly labeled half of the plot. PMO-1 vs. PMO-2 (a), PMO-1 vs. PMO-3 (b), and PMO-2 vs. PMO-3 (c).

Additional O/C distribution insight was derived from separating the species into CHO, CHNO, CHOS elemental groups. For example, the comparison of the species with CHO formulas in each sample indicates a smaller relative difference between PMO-2 aerosol compared to PMO-1 and PMO-3, with the PMO-2 aerosol having a higher average O/C value ($0.55 \pm 0.17$ (PMO-2) compared to $0.47 \pm 0.14$ (PMO-1) and $0.44 \pm 0.14$ (PMO-3)). Meanwhile 85 - 98% of the CHO species in each sample are present in at least one other sample, with 848 (42 - 78%) of the formulas being found in all three samples, as shown in Figure 3.11. Panel a shows the number of molecular formulas that are common and unique between the various samples and Panel b shows the same data, normalized to 100% to make some of the groups more clear than they are in Panel a. These results suggest that the CHO composition may be fairly uniform throughout the atmosphere, without a significant abundance of clear marker species after long-range transport, regardless of the source region and transport time. This observation is consistent with other studies which have observed the decay of marker species after ~ 24 hours (Bougiatioti et al., 2014; Forrister et al., 2015).

**Figure 3.11.** Molecular formulas common to all three samples and those unique to each sample presented as the total number of formulas (a) and as a percent of total number of formulas (b).
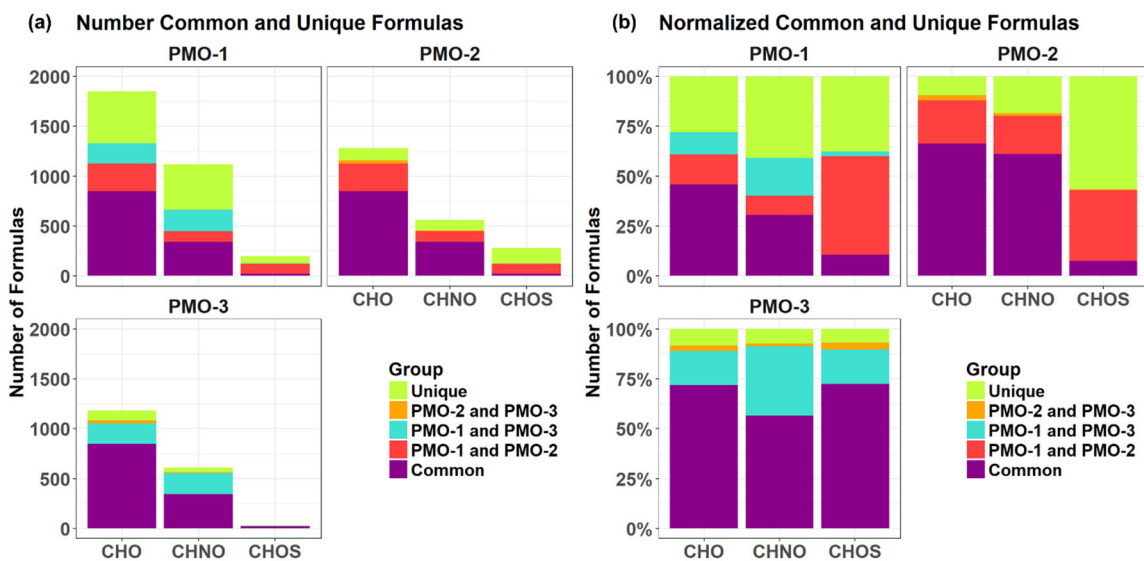
In contrast, the CHNO molecular formulas demonstrate stronger differences that correlate with the overall O/C ratio. The average O/C value for the CHNO formulas in PMO-2 was $0.59 \pm 0.14$ compared to $0.49 \pm 0.15$ in PMO-1 and $0.49 \pm 0.14$ PMO-3 (Table 3.3). Differences in the elemental ratios are often visualized using the van Krevelen plot, which shows the correlation of H/C *vs.* O/C. The van Krevelen plots for the three samples with the unique CHNO formulas present in each sample are shown in Figure 3.12. Most of the unique CHNO species in PMO-2 (68%) fall in the more oxidized region of the plot (Tu et al., 2016) with high overall O/C values. This differs from the PMO-1 unique species that are predominantly on the less oxidized, low O/C side of the plot, or the oxidized aromatic region. Another observation from the CHNO species is more identified species in both PMO-1 (1120) and PMO-3 (608) than in PMO-2 (561), despite the higher total number of molecular species in PMO-2 compared to PMO-3. This is potentially due to the enrichment of $NO_x$ and $NO_y$ species as previously observed in wildfire pollution events (Val Martin et al., 2008a), which may in turn lead to an increased nitrogen content in the organic aerosol species. The nitrogen containing species show a distinct difference in terms of the total oxidation between the two sets of samples, more so than the CHO compounds. This implies that much of the distinction between aerosol sources may come from heteroatom containing species.

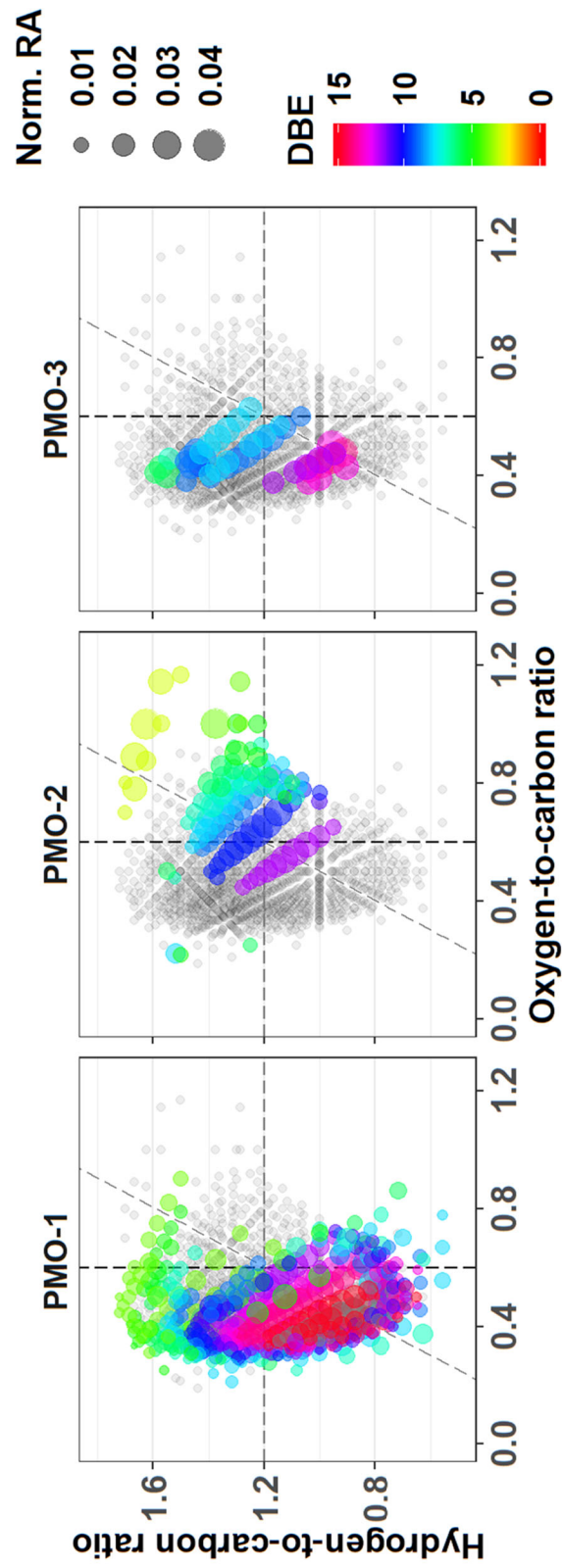**Figure 3.12.** Van Krevelen plots for the CHNO species with all identified CHNO species (grey) and unique species (colored). The color represents the number of double bond equivalents in the corresponding molecular formula. The diagonal line is an oxidation line ($OS_C = 0$ for C, H, O elements; Tu et al., 2016), where the more oxidized formulas are on the right side and less oxidized are on the left.

84

The difference in O/C is even more evident in the sulfur containing formulas (CHOS). The PMO-2 CHOS species have a much higher average O/C ratio ($0.74 \pm 0.34$) than what is observed in PMO-1 ($0.48 \pm 0.14$). Consistent with the CHNO formulas, the PMO-2 unique CHOS formulas (55% of unique formulas) are present in the oxidized region of the plot, whereas those in PMO-1 are nearly completely in the less oxidized region of the van Krevelen plot (Figure 3.13). The Kendrick mass defect plot (Figure 3.13c) also demonstrates a clear difference between the two samples. Most of the unique CHOS compounds in PMO-2 are located on the lower mass, higher defect side of the plot, while the PMO-1 formulas are on the higher mass, lower defect side. This difference is due to the larger amount of oxygen present in the PMO-2 formulas, which would lead to a greater Kendrick mass defect (KMD) than the more reduced CHOS formulas present in PMO-1. The higher oxygen content of PMO-2 aerosol is supported by its higher O/C ratio when compared to PMO-1 as shown in box plots (Figure 3.13d). Very few CHOS molecular formulas ($N = 29$) were identified in PMO-3 and most of them ($N = 26$ of 29 total) were also present in PMO-1. Due to the small number of identified CHOS formulas in PMO-3, we did not consider it in the comparison between CHOS formulas in the samples. The increased number of sulfur species observed in PMO-2 are likely associated with the anthropogenic emission sources in the North American boundary layer. Overall, the observed differences in the O/C ratios between the boundary layer transported aerosol (PMO-2) compared to the free troposphere transported aerosol (PMO-1 and PMO-3) highlight differences in the aging and lifetime of aerosol relative to its transport pathway and emission source.

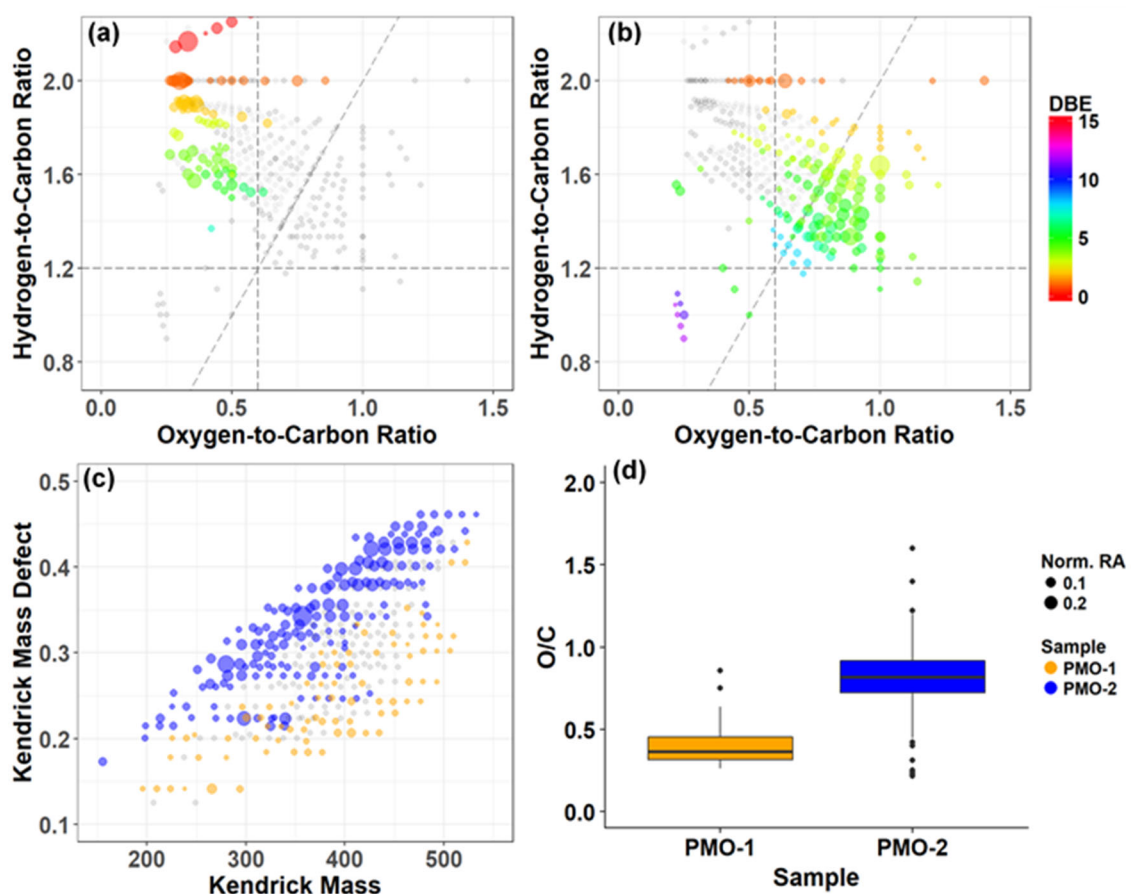**Figure 3.13.** A comparison of PMO-1 (a) and PMO-2 (b) CHOS molecular formulas using van Krevelen, Kendrick plots (c), and O/C box plots (d). Common CHOS molecular formulas (grey) and unique CHOS molecular formulas (colored) are indicated in a-c.

Another commonly used metric of aerosol oxidation is the average oxidation state of carbon ($OS_C$) described by Kroll et al. (2011). The average $OS_C$ includes both hydrogen and oxygen for the average oxidation of carbon in each molecular formula. Additionally, we assumed all nitrogen and sulfur were present as nitrate and sulfate functional groups and calculated the $OS_C$ with the appropriate corrections (Chapter 2, Equation S1). The average $OS_C$ values (Table 3.3) for the three samples show again that PMO-2 is more oxidized than the other two samples. The average $OS_C$ values for the CHO formulas in PMO-1 and PMO-2 are very similar (Table 3.3), but as shown in the histograms in Figure 3.14, their relative abundance distributions are quite different. The $OS_C$ *vs.* carbon number plots in Figure 3.14 show slight differences between PMO-1 and PMO-2, mostly in the distribution of the sulfur containing formulas. However, the similarity of the PMO-1 and PMO-3 samples and their difference from the PMO-2 sample is quite clear in the visual comparisons of the histograms of the $OS_C$ values with their normalized relative abundances. The observation of an overall lower oxidation in PMO-1 and PMO-3 may support the findings of Aiken et al. (2008) and Bougiatioti et al. (2014) who reported that biomass burning aerosol are less oxidized than other types of aerosol, even after some aging. Conversely, the overall higher oxidation of PMO-2 implies that the sampled aerosol was likely more hygroscopic, included more efficient cloud condensation nuclei (Massoli et al., 2010), or had components of a less volatile nature (Ng et al., 2011) than PMO-1 and PMO-3.
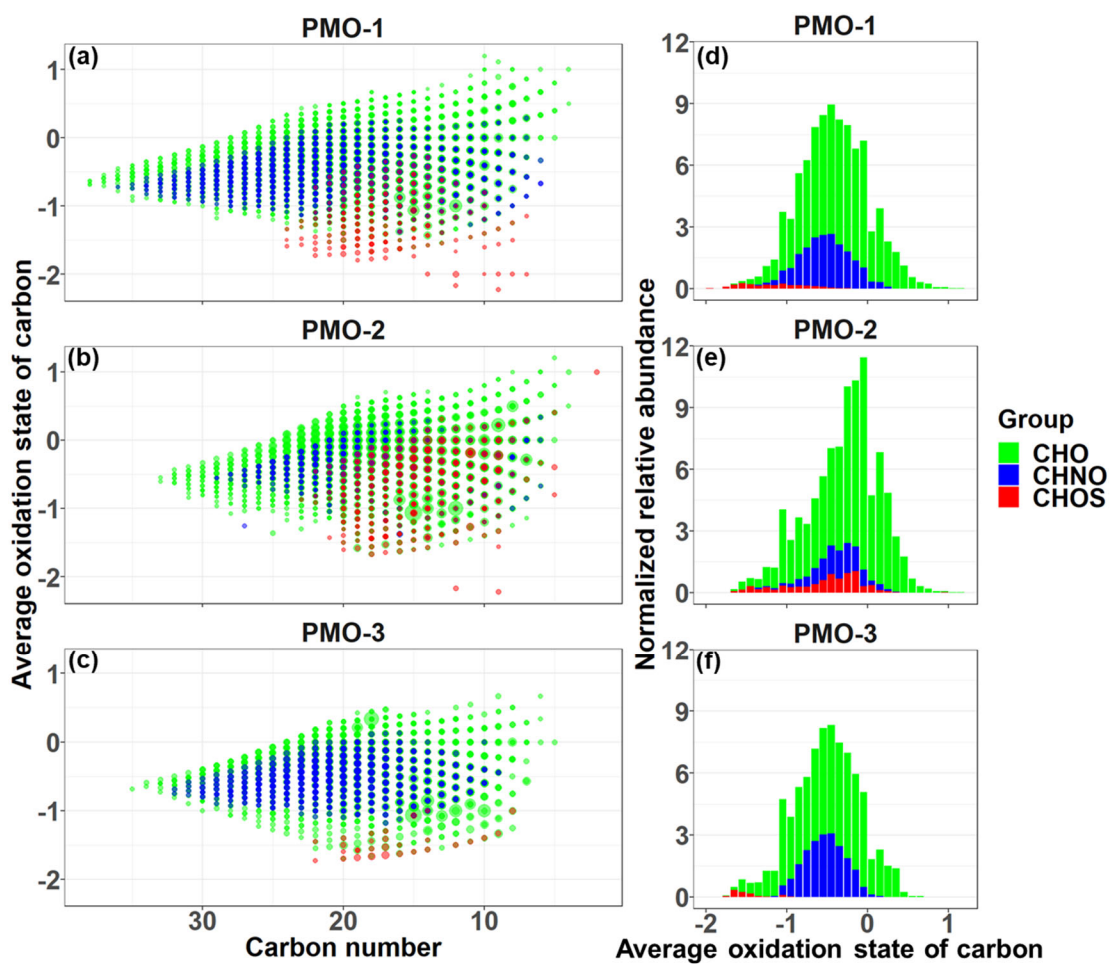
**Figure 3.14.** Average OS$_C$ vs. carbon number plots for molecular formula identified in each of PMO samples (a-c). The size of the symbols is scaled to the analyte relative abundance and the color represents the elemental group: CHO (green), CHNO (blue), and CHOS (red). The right panel (d-f) contains average OS$_C$ histograms based on the sum of normalized abundance.

## 3.5   Molecular Formula Aromaticity and Brown Carbon

The aromaticity of the samples is also different between the two groups of aerosol samples. Based on the aromaticity index (AI, Eq. 2.2; $AI_{mod}$, Eq. 2.3; Koch and Dittmar, 2006; 2016, Chapter 2), the free tropospheric aerosol samples (PMO-1 and PMO-3) are more aromatic than the convected boundary layer aerosol (PMO-2; Figure 3.15). The presence of more aromatic species in the long-range transported wildfire-influenced aerosol may lead to increased light absorption (Bao et al., 2017) and perhaps an increased resistance to oxidation (Perraudin et al., 2006). Aromatic species can also be associated with the presence of brown carbon (BrC; Desyaterik et al., 2013). Aromaticity is heavily dependent on the H/C ratio and the DBE (Chapter 2, Eq. 2.4), where low H/C and high DBE indicate aromatic structure.
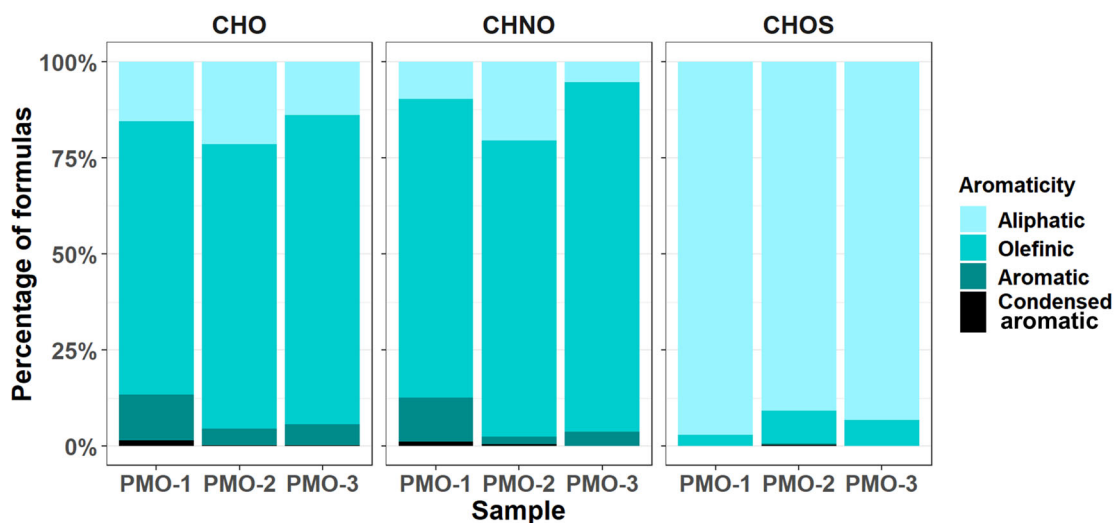


**Figure 3.15.** Normalized bar charts for the aromaticity of the three PMO samples, calculated using the Koch and Dittmar (2006; 2016) modified aromaticity index ($AI_{mod}$).

Histograms depicting the distribution of H/C and DBE values for the three samples are

shown in Figure 3.16. As observed previously, PMO-1 and PMO-3 are more similar to

each other than compared to PMO-2. Likewise, PMO-1 and PMO-3 exhibit an increase in

the number frequency of higher DBE species, which is not observed in PMO-2,

supporting the observation of an increased overall aromaticity for these free tropospheric

aerosol samples. The difference is also demonstrated by the H/C histograms because

PMO-1 has a higher proportion of its formulas below H/C = 1 relative to PMO-2, which

indicates the potential for higher aromaticity. Many aromatic compounds, such as PAH

are known to be carcinogens, and are a product of incomplete combustion biomass

burning and anthropogenic emissions (Perraudin et al., 2006; Bignal et al., 2008).
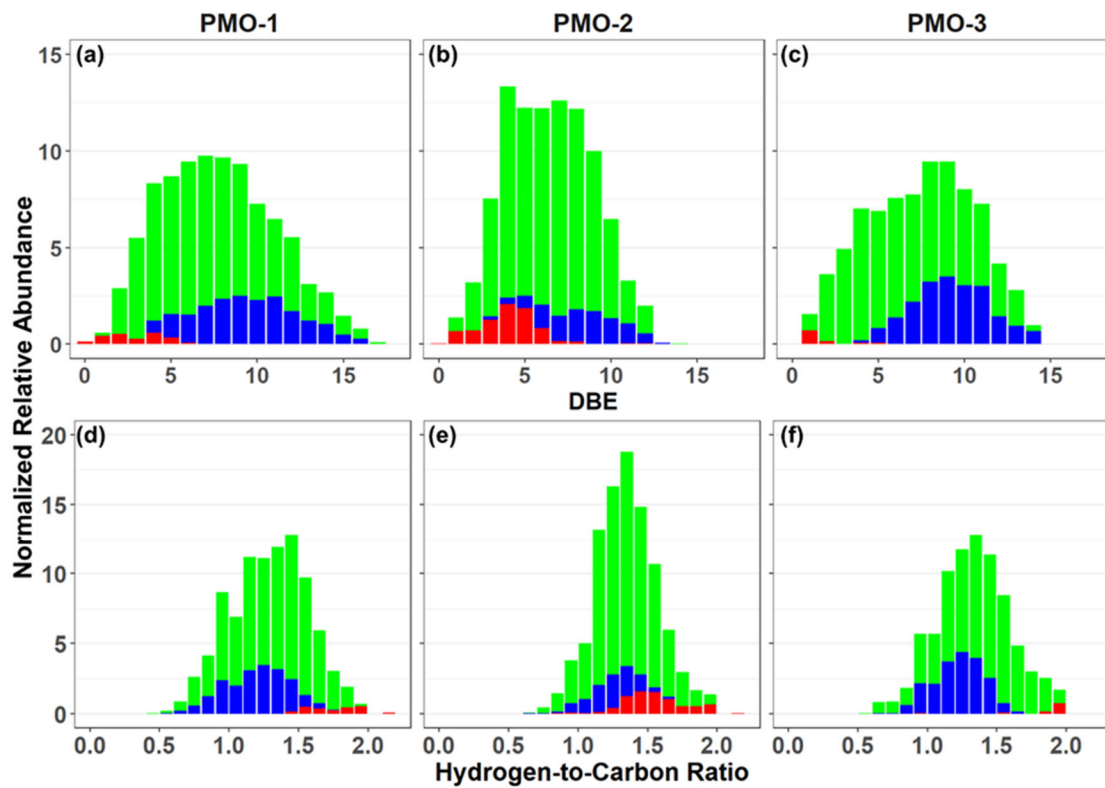
**Figure 3.16.** Histograms of the molecular formula DBE (a-c) and H/C (d-f).

Generally, BrC is considered to be aromatic or olefinic in nature (Bao et al., 2017). In our observations, the two samples influenced by wildfire show the greatest amount of olefinic and aromatic species, which is likely associated with the presence of BrC compounds. Additional evidence for the presence of BrC in PMO-1 comes from aethalometer measurements using the 7 wavelength aethalometer (Magee Scientific Company, Berkeley, California, USA) located at the site, which detected a wavelength-dependent peak with an Ångström exponent of 1.3 during the sampling period. Ångström exponents above 1 suggest the presence of BrC or iron oxides. Based on the retroplume analysis, Ångström exponent, wavelength dependent absorbance, and comparison to similar samples (Džepina et al., 2015), the detected peak is most likely the result of BrC. Figure 3.17 contains the aethalometer observations for this event and for PMO-2. In contrast with PMO-1, the results for PMO-2 show very little increase in absorbance and no wavelength dependence during the sampling period. This means the aerosol in this period was not strongly absorbing, which is consistent with the hypothesis that it is anthropogenic SOA, which is not generally strongly absorbing in visible wavelengths. Difficulties with the instrument prevented similar data from being collected for PMO-3, although based on the retroplumes, ambient conditions, and molecular characteristics similar results seem likely. In addition to the aethalometer response, PMO-1 contained species that were related to BrC in studies by Iinuma et al. (2010) and Lin et al. (2016) (Table 3.5). This observation provides evidence for the persistence of BrC species, which is contrary to the observations by Forrister et al. (2015) who concluded that BrC is mostly removed within 24 hours. Additionally, the high concentration of OC for this sample

makes it seem unlikely that we observed just a minor residual fraction. Perhaps, the

lifetime of BrC is dependent on additional ambient conditions that influence aerosol

oxidation and phase state.

**Table 3.5**. Molecular formulas identified in brown carbon by Iinuma et al. (2010) and Lin et al. (2016), where "Yes" means the formula was observed in our sample and, "No" means it was not.

| Formula | Observed | Citation |
|---------|----------|----------|
| $C_7H_7NO_4$ | Yes | Iinuma et al. 2010; Lin et al. 2016 |
| $C_6H_5NO_3$ | Yes | Lin et al. 2016 |
| $C_6H_5NO_4$ | Yes | Lin et al. 2016 |
| $C_6H_6N_2O_6$ | No | Lin et al. 2016 |
| $C_6H_4NO_4$ | No | Lin et al. 2016 |
| $C_{10}H_9NO_3$ | No | Lin et al. 2016 |
| $C_8H_7NO_4$ | Yes | Lin et al. 2016 |
| $C_8H_7NO_3$ | Yes | Lin et al. 2016 |
| $C_9H_7NO_4$ | Yes | Lin et al. 2016 |
| $C_{10}H_7NO_4$ | Yes | Lin et al. 2016 |
| $C_8H_8O_3$ | Yes | Lin et al. 2016 |
| $C_9H_6O_3$ | Yes | Lin et al. 2016 |
| $C_{10}H_8O_4$ | Yes | Lin et al. 2016 |
| $C_{13}H_8O_5$ | Yes | Lin et al. 2016 |
| $C_{13}H_8O_6$ | Yes | Lin et al. 2016 |
| $C_{15}H_{10}O_6$ | Yes | Lin et al. 2016 |
| $C_{16}H_{12}O_6$ | Yes | Lin et al. 2016 |
| $C_{16}H_{12}O_7$ | Yes | Lin et al. 2016 |
| $C_{17}H_{14}O_8$ | Yes | Lin et al. 2016 |
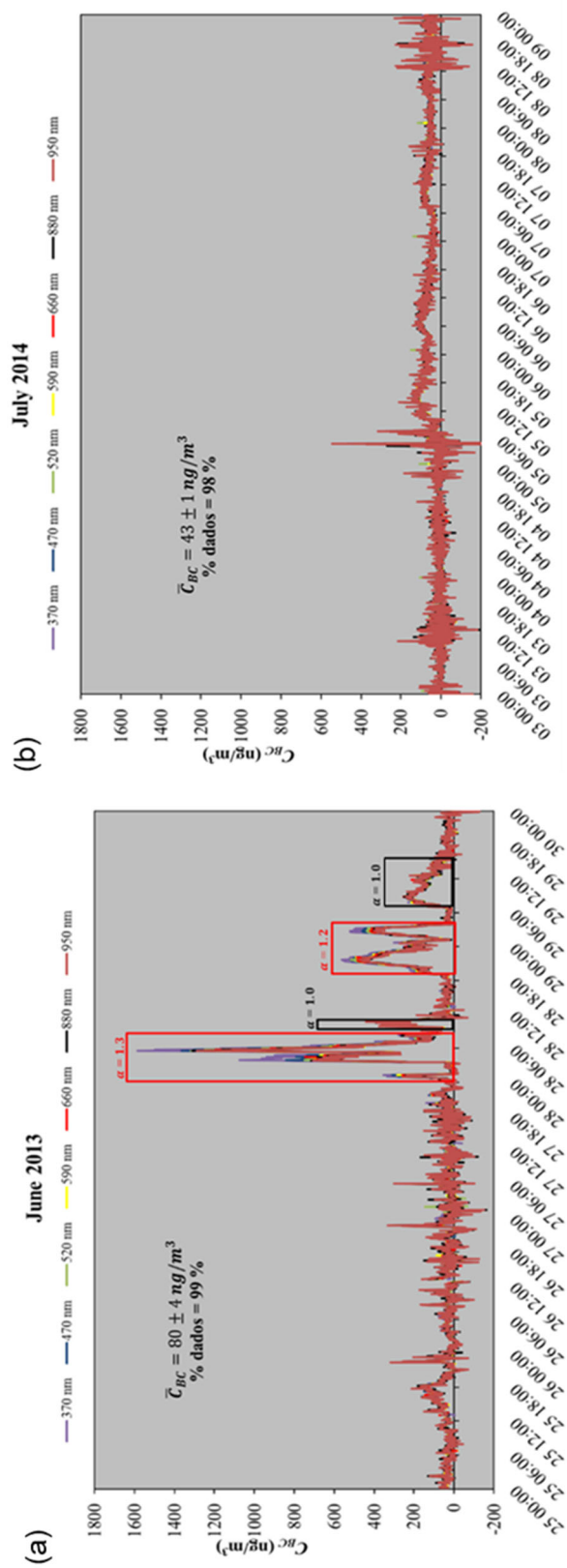
**Figure 3.17.** Aethalometer results for the PMO-1 (a) and PMO-2 (b) sampling periods. A few days before and after each sampling period are included for reference. Note the tallest wavelength dependent peak present from approximately 28 June at 00:00 to 28 June at 09:00 local time (first red box), which corresponds to the sampling period of PMO-1. When the absorption angstrom exponent ($\alpha$) equals 1, it suggests black carbon, when it is $\geq$ 1.2 it suggests presence of brown carbon as well. Due to instrument maintenance/repair no data were available for the time period associated with PMO-3. The sampling period for PMO-2 did not show wavelength dependence.

## 3.6 Phase State, Volatility, and Cloud Processing: Implications for the Observed Aerosol Oxidation

Atmospheric aging processes are influenced by ambient conditions, such as temperature and water vapor, and the concentrations of reactive species. Recently, Shrivastava et al. (2017) reported observations of long-range transported PAH from Asia to North America and suggested an enhanced lifetime due to a probable glassy aerosol phase state during transport. Additionally, model simulations reported by Shiraiwa et al. (2017a) indicated that model SOA is predicted to be semi-solid or glassy at altitudes above 2000 m in the northern hemisphere. Since the PMO aerosol was sampled at 2225 m above sea level, we examined the estimated glass transition temperature ($T_g$) of the studied WSOC species in addition to the markers of aqueous phase processing for the three PMO samples. Increased aerosol viscosity has been shown to decrease the rate of photodegradation (Lignell et al., 2014; Hinks et al., 2016) and water diffusivity (Berkemeier et al., 2014). Both photodegradation and water diffusion are expected to strongly affect the oxidation and aging of aerosol species during transport.

In general, lower volatility typically inversely correlates with $T_g$ (Shiraiwa et al., 2017a) and viscosity. As such, a better understanding of the potential volatility can be important for evaluating the viscosity of the PMO aerosol. Using the parameters reported by Donahue et al. (2011) and Li et al. (2016), we estimated the volatility of the FT-ICR MS identified organic aerosol molecular compositions (Figs. 3.18 and 3.19, respectively). These parameters are based on the molecular formulae of the aerosol species present in the samples. The Donahue et al. method is only applicable to CH and CHO formulas,

while the Li et al. method is also applicable to CHNO, CHNOS, and CHOS molecular formulas. In both cases compounds with known volatilities were used to generate a plot that could be used to find a relationship between the volatility and the molecular composition. The volatilities calculated by these methods are not equivalent, in general, the volatility calculated with the Li method is equal to the Donahue volatility multiplied by ~2. Figure 3.20 demonstrates this comparison. As expected, based on the length of transport for the samples, the majority of formulas show extremely low volatility. Interestingly, PMO-2 has a larger number of higher abundance molecular formulas with extremely low volatility and elevated oxidation relative to PMO-1 and PMO-3 (Figure 3.21). This highlights the relationship between O/C and volatility, where volatility is expected to decrease as O/C increases when the mass range is constant (Ng et al., 2011); the relationship between oxygen and carbon and its effect on volatility is used by both Donahue et al. (2011) and Li et al. (2016) to estimate volatility. Similarly, lower volatility is expected to lead to lower diffusivity in aerosol even at elevated RH as demonstrated by Ye et al. (2016).
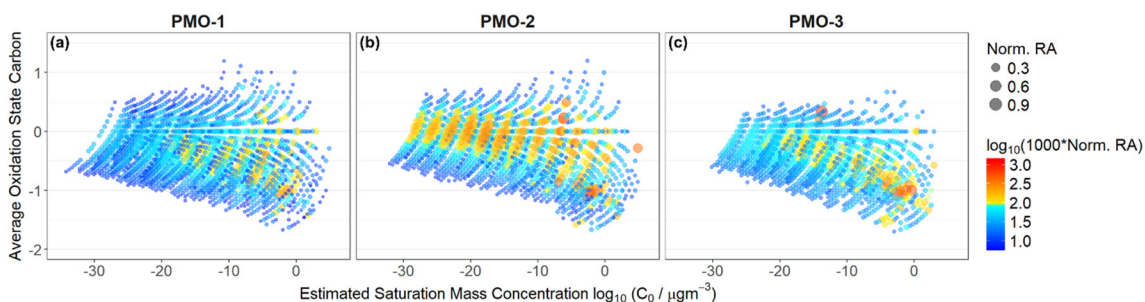


**Figure 3.18.** OS$_C$ *vs.* volatility plots for the three samples. Volatility estimates were made using the Donahue et al. (2011) method. Only volatility for CHO species can be estimated with this method. Color is the logarithm of the normalized relative abundance

multiplied by 1000. Of interest is the increased abundance of low volatility, higher oxidation species in PMO-2 relative to PMO-1 and PMO-3, indicating the importance of these species to this sample and highlighting a difference between these samples.
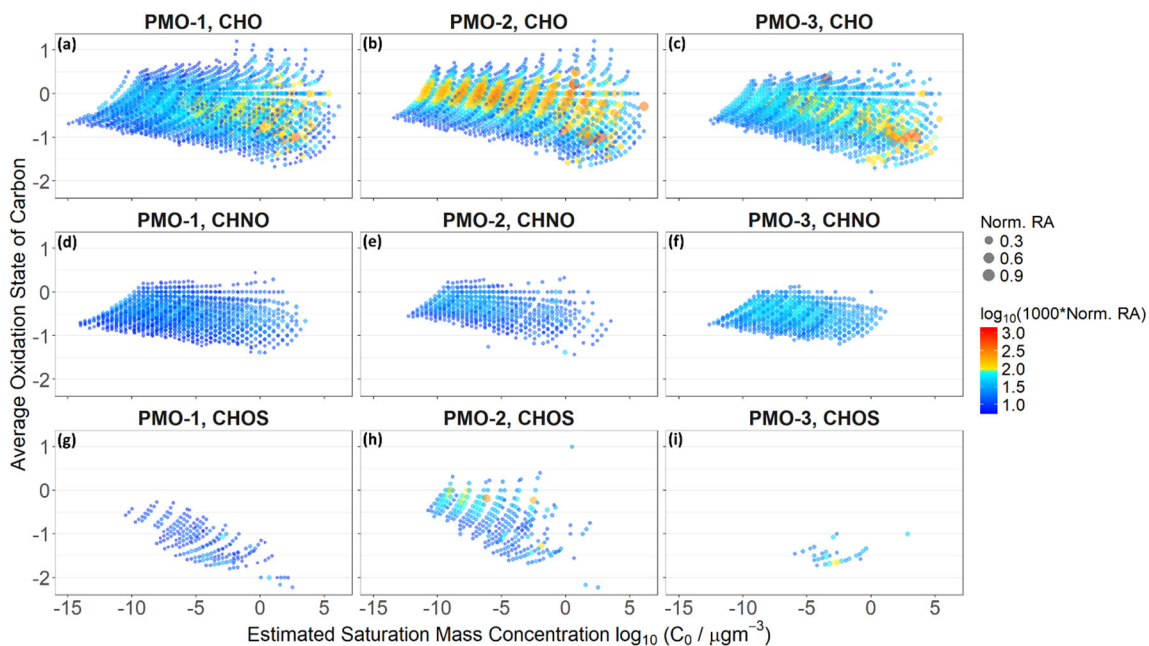


**Figure 3.19.** Group separated $OS_C$ *vs.* volatility plots for the three samples. Volatility estimated using the Li et al. (2016) method. Color is the logarithm of the normalized RA multiplied by 1000. The same increase in abundance for low volatility, higher oxidation species is observed in this figure as in Figure 3.18.
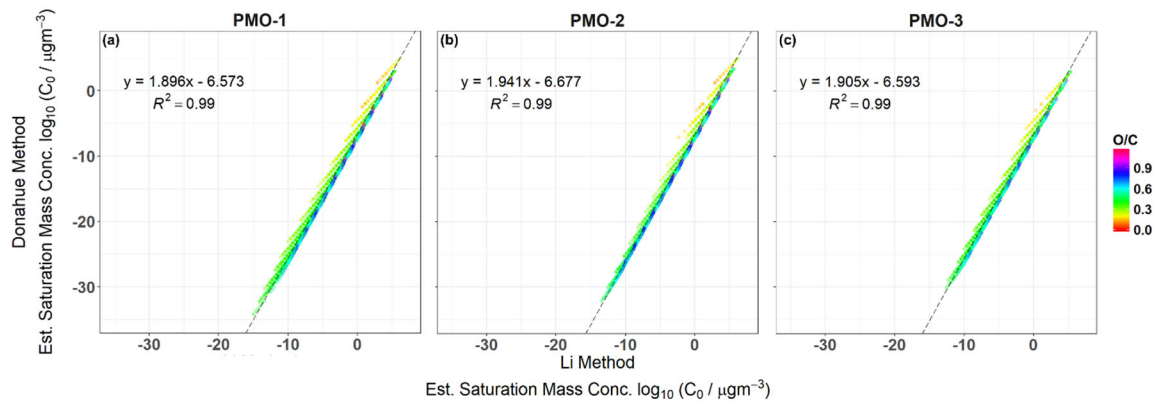
**Figure 3.20.** Correlation plot of volatility calculated by the Li et al. (2016) method and the Donahue et al. (2011) method. The relation is linear although the slope of the line is nearly two, indicating that the Donahue method predicts values that are roughly two times what the Li method predicts, at least for the low and extremely low volatility species presented here.
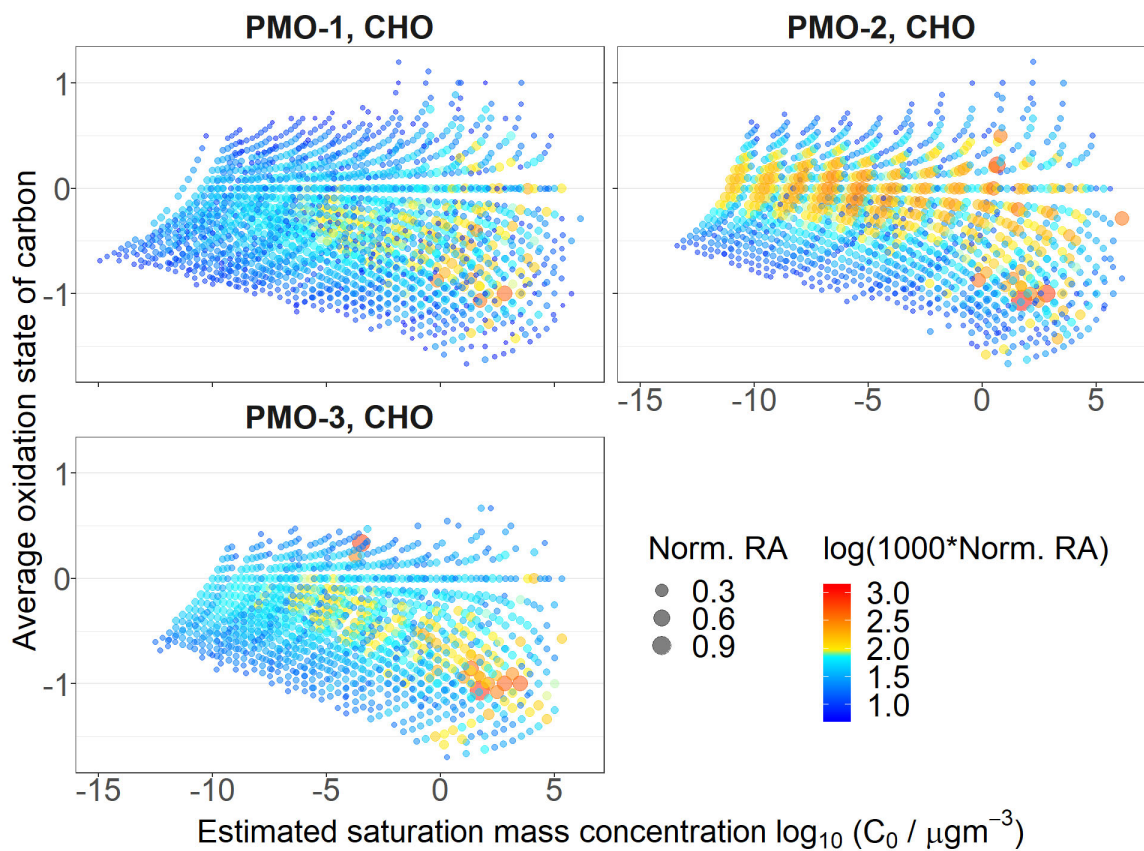
**Figure 3.21.** OS$_C$ *vs.* volatility estimated using the Li et al. (2016) method for the CHO species in the three samples. The size is determined by the normalized relative abundance and the color is determined by the logarithm of the normalized relative abundance multiplied by 1000.

As predicted in earlier studies (Shrivastava et al., 2017; Shiraiwa et al., 2017a), particles transported in the free troposphere are likely semi-solid to solid, where the actual particle viscosity depends on the ambient conditions and the composition of the particles. Thus, to better understand the potential phase state associated with the PMO organic aerosol, we first estimated the dry $T_g$ for the identified CHO molecular formulas in each of the PMO aerosol using the estimation method by DeRieux et al. (2018; Chapter 2, Eq. 2.5). We then converted the dry $T_g$ to the RH dependent $T_g$ (below). Currently $T_g$ can only be estimated for CHO species, however the CHO species were the most frequently observed and constituted a major fraction of the total relative abundance in the PMO negative ion mass spectra. Assuming the identified CHO compositions are fairly representative of the total organic aerosol composition, a comparison of the $T_g$ values to the ambient temperature ($T_{amb}$) provides an indication of the likely phase state of the organic aerosol particles. Generally, if $T_g$ exceeds $T_{amb}$, a glassy solid state is predicted, likewise, if $T_g$ is less than $T_{amb}$ then either a semi-solid or liquid state is predicted depending on the ratio magnitude (Shiraiwa et al., 2017a; DeRieux et al., 2018). Although the exact composition of the total organic aerosol is yet unknown, the identified water-soluble organic compounds provide a reasonable upper limit for the estimated $T_g$ values. Under this assumption, the CHO molecular formulas in PMO-1 and PMO-3 had higher average dry $T_g$ values than PMO-2 (Table 3.6, Figure 3.22), which implies that they would be more viscous than PMO-2, given similar atmospheric conditions.

**Table 3.6.** Estimated average dry $T_g$ for the three samples. All values are in K.

|  | PMO-1 | PMO-2 | PMO-3 |
|---|---|---|---|
| **Unweighted** | 333.76 | 328.94 | 333.56 |
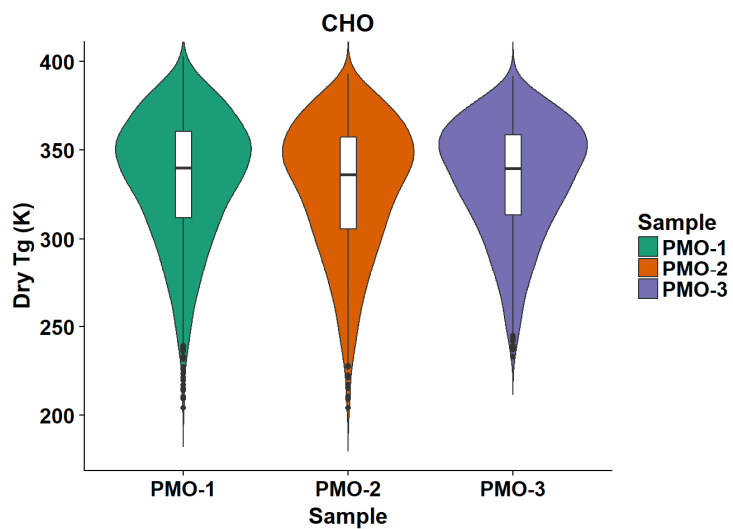


**Figure 3.22.** Violin plots showing the distribution of dry $T_g$ values for each sample.

Water is known to be a strong plasticizer relative to typical aerosol species (Koop et al., 2011; Shiraiwa et al., 2017a; Reid et al., 2018), thus it can decrease $T_g$ and the overall aerosol viscosity. Therefore, it's important to consider the ambient relative humidity when estimating the $T_g$. Using the extracted ambient temperature and RH from the GFS along the FLEXPART retroplumes and the Gordon-Taylor equation (Chapter 2, Eqs. 2.6 – 2.7), the calculated dry $T_g$ were modified to RH-dependent $T_g$ for the CHO molecular species. The distributions of the $T_g$ values for the three PMO samples based on one standard deviation of the ambient conditions are shown as boxplots in Figure 3.23. The range of ambient temperature and RH extracted from the GFS along the FLEXPART simulated path yields a wide range of $T_g$ values (Figs. 3.23). The estimates were taken back only 5 days due to the increasing range of possible meteorological conditions associated with the spread in the air masses as shown in Figures 3.1-3.3. Overall, the distributions of $T_g$ values in PMO-1 and PMO-3 generally exceed the ambient temperature (Figure 3.23), implying that particles mostly containing these compounds would likely be solid. To account for the low molecular weight organic anions not observed in the FT-ICR mass spectra, their mass concentrations and $T_g$ values (estimated using the Boyer-Kauzmann rule (Koop et al., 2011; Shiraiwa et al., 2017a; DeRieux et al., 2018)) are also shown in Figure 3.23. The three most prevalent low molecular weight organic acids indicate the potential impact of those compounds on the overall $T_g$ value of a particle that contains them. Oxalic acid was estimated to have a similar $T_g$ value to a majority of the higher MW species identified in PMO-2, but it is slightly lower than the

majority of species in PMO-1 and PMO-3. However, the mass fraction of oxalate is 3 times lower in PMO-1 and PMO-3 (2.3 and 3.0 %) compared to PMO-2 (9.4 %).

The results suggest that aerosol in PMO-1 and PMO-3 was overall less susceptible to atmospheric oxidation due to the aerosol phase state during free tropospheric long-range transport than it may have been in the boundary layer with higher ambient RH and temperature. A more viscous phase state during transport may also explain the presence of persistent BrC species in PMO-1, where the BrC species are protected from oxidation similarly to the long-lived PAHs observed by Shrivastava et al. (2017). In contrast to the observations from PMO-1 and PMO-3, much of the PMO-2 $T_g$ distribution falls below the ambient temperature implying a semi-solid or liquid state during the final 5 days of transport. This indicates an increased susceptibility to oxidation processes in the atmosphere (Shiraiwa et al., 2011), such as aqueous phase processing. The possibility of aqueous phase processing is also supported by the extracted GFS RH in Figure 3.23, which is above 50% for the last 5 days of PMO-2 transport. The potential for liquid/semi-solid aerosol in the boundary layer is consistent with other studies (Shiraiwa et al., 2017a; Maclean et al., 2017) due to the increased RH in the boundary layer and the plasticizing effect of water. Although, we note the PMO-2 average dry $T_g$ values were 4-5° lower than those of PMO-1 and PMO-3. Overall, the estimates of dry $T_g$ and RH-dependent $T_g$ provide an otherwise unattainable upper limit estimate of the aerosol phase state of the sampled free tropospheric aerosol in this study.
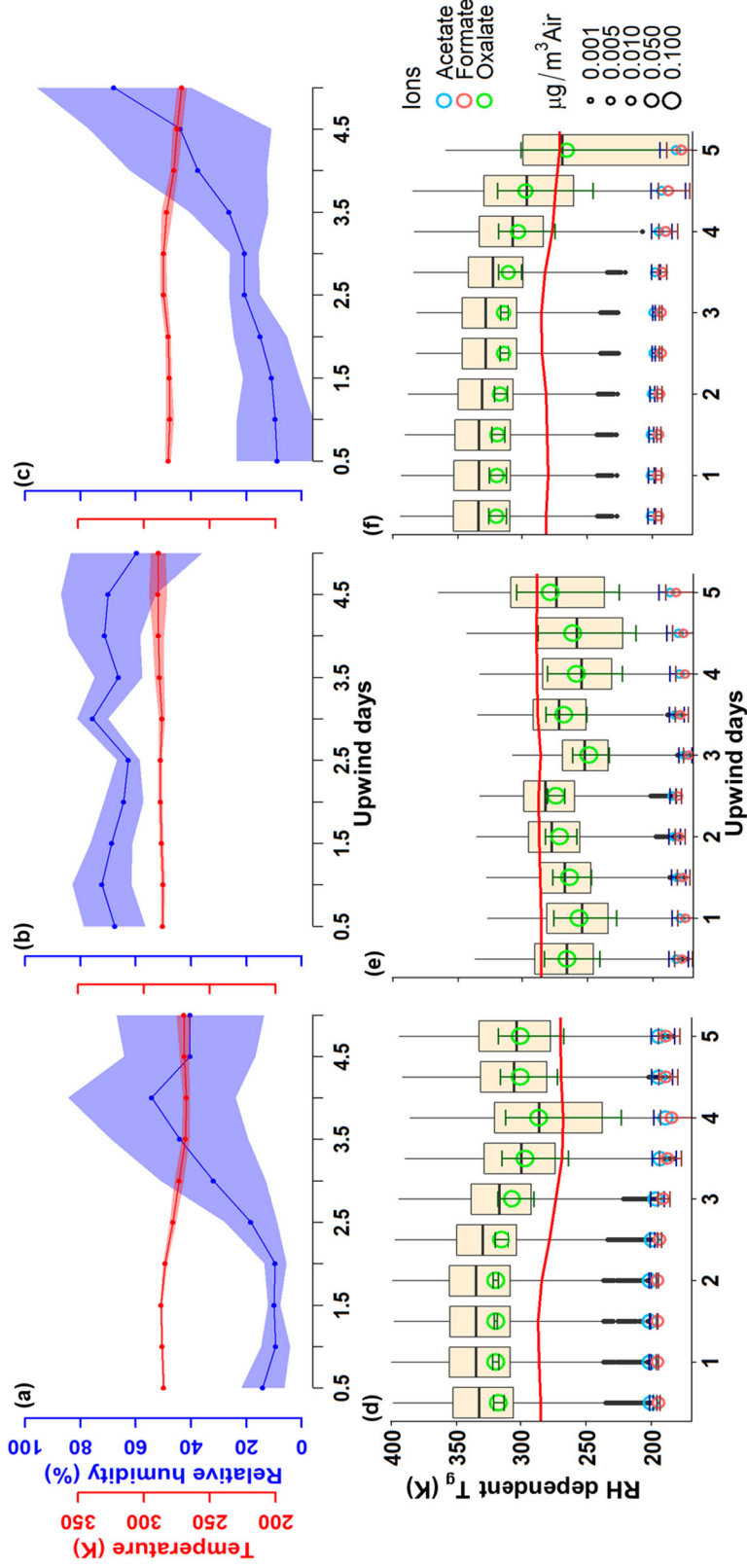
**Figure 3.23.** Panels a-c contain the ambient conditions extracted from the GFS analysis along the FLEXPART modeled path weighted by the residence time for PMO-1, PMO-2, and PMO-3, respectively. The line represents the mean value and the shading represents one standard deviation of values. Panels d-f contain the boxplot distributions of the relative humidity dependent $T_g$ values for molecular formulas using the maximum, mean, and minimum RH for PMO-1, PMO-2, and PMO-3, respectively. The $T_g$ values for the full composition of each sample were calculated using the maximum, mean, and minimum RH and then all three sets of data are combined and plotted as a single distribution for each time period. The open circles represent the abundance and Boyer-Kauzmann estimated $T_g$ for the acid forms of the three most abundant low MW organic ions, the bars around the circles represent the range of possible $T_g$ values for those compounds when the range of RH is considered. The red line demonstrates the ambient temperature at each time point, as extracted from GFS. The centerline of the boxplot represents the median, the top and bottom of the "box" represent the third and first quartiles, respectively. The "whiskers" represent Q3 + 1.5* interquartile range (IQR, Q3-Q1) (maximum), and Q1 − 1.5*(IQR) (minimum).

As described above, the most obvious difference in the molecular composition of PMO-2 vs. PMO-1 and PMO-3 is the increased extent of oxidation. In fact, most of the unique species observed in PMO-2 are in the highly oxidized region of the van Krevelen plot (Figure 3.24). However, the exact oxidation pathways that led to the increased oxidation observed for PMO-2 and its initial composition are unclear. Both gas phase and aqueous phase reactions lead to SOA, where aqueous SOA components can have higher O/C values than gas phase SOA components (Lim et al., 2010; Ervens et al., 2011). The high numbers of CHNO and CHOS molecular formulas observed here are consistent with secondary components associated with an emission plume likely enriched in $SO_2$, NOx, and $O_3$ pertaining to its expected anthropogenic influence. All three of these reactive species have been shown to lead to production and oxidation of SOA in the atmosphere (Hoyle et al., 2016; Bertrand et al., 2018).
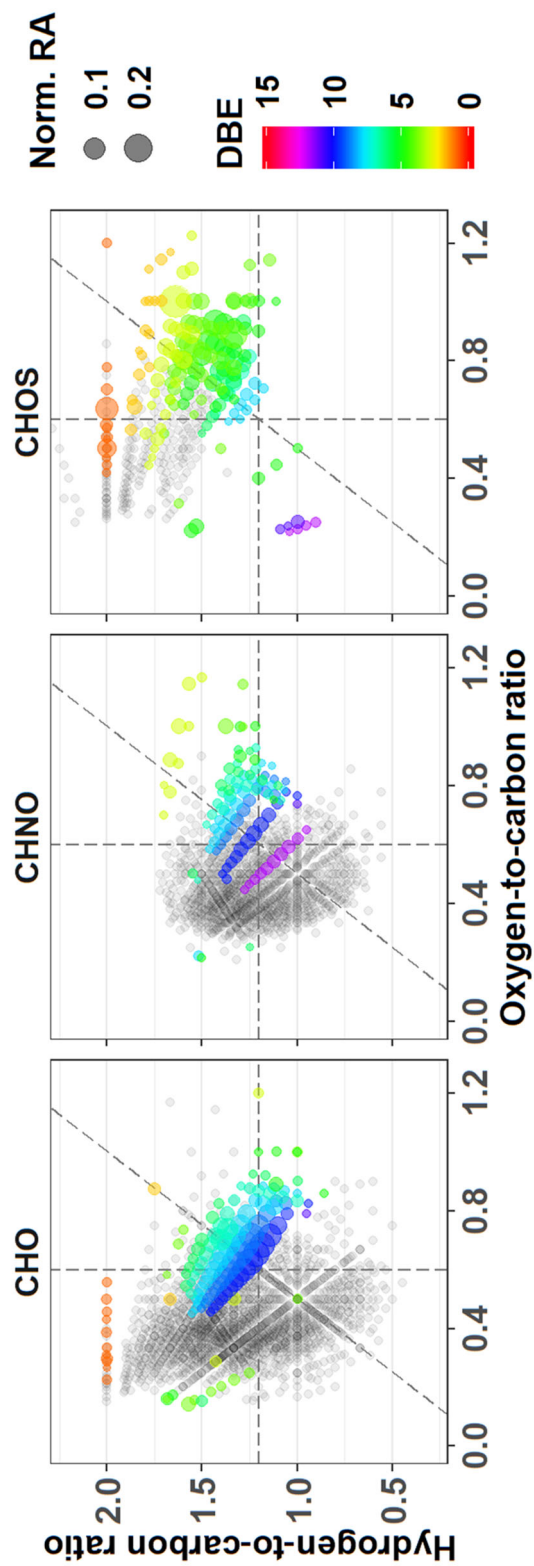
**Figure 3.24.** PMO-2 van Krevelen plots for unique molecular formulas separated by group. Symbols are scaled to indicate the normalized relative abundance. The DBE is indicated for each of unique molecular formulas using colored symbols. Formulas common with other samples are provided in grey for context.
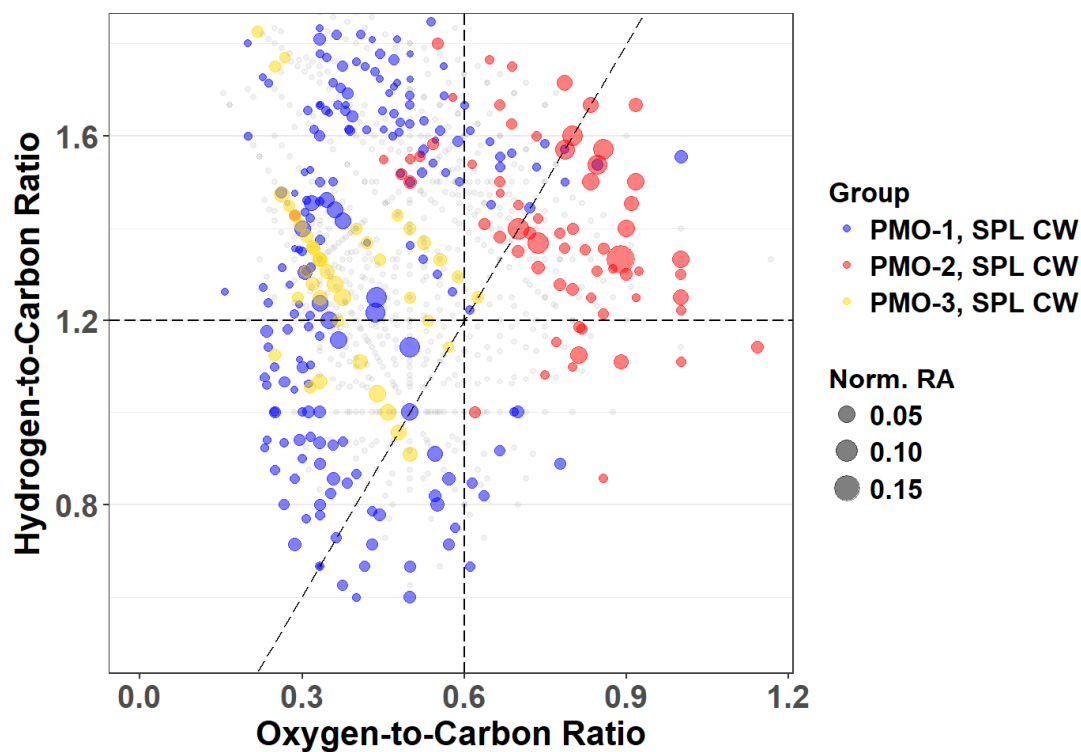
**Figure 3.25.** Van Krevelen plot showing the molecular formulas that are common to only one PMO sample and the cloud water samples from SPL (Zhao et al., 2013). PMO-2 (red) molecular formulas located nearly exclusively in highly oxidized region of plot, may indicate cloud processing. Common molecular formulas from either PMO-1 (blue) and PMO-3 (gold) may be related to the biomass combustion that influenced the supercooled cloud water collected in the winter at SPL. Formulas that are common to two or more PMO samples and CW are in grey.

**Table 3.7.** Number of molecular formulas and their average O/C values (unweighted O/C and RA weighted O/C (O/C$_w$)) uniquely common between this study and ambient aqueous organic matter (Mazzoleni et al., 2010; Zhao et al., 2013; Cook et al., 2017). Uniquely common means that the formula is common between only one of the PMO samples and the aqueous organic matter sample. CW indicates cloud water, the numbers in parentheses are the percentage of total formulas.

| Sample | # Common Formula | O/C | O/C$_w$ |
|---|---|---|---|
| **PMO and Fog (Mazzoleni et al., 2010)** | | | |
| PMO-1 | 202 (6.4%) | 0.38 | 0.39 |
| PMO-2 | 48 (2.3%) | 0.5 | 0.55 |
| PMO-3 | 11 (0.60%) | 0.29 | 0.29 |
| **PMO and CW (Cook et al., 2017)** | | | |
| PMO-1 | 2 (0.063%) | 0.82 | 0.82 |
| PMO-2 | 23 (1.1%) | 0.8 | 0.81 |
| PMO-3 | 1 (0.055%) | 0.36 | 0.36 |
| **PMO and CW (Zhao et al., 2013)** | | | |
| PMO-1 | 197 (6.2%) | 0.42 | 0.42 |
| PMO-2 | 70 (3.3%) | 0.76 | 0.8 |
| PMO-3 | 42 (2.3%) | 0.38 | 0.38 |

Cloud and aqueous processing have also been shown to increase the oxidation of atmospheric organic matter (e.g., Ervens et al., 2008; Zhao et al., 2013; Cook et al., 2017; Brege et al., 2018). Comparisons of the detailed molecular composition of the PMO samples with studies of cloud (Zhao et al., 2013; Cook et al., 2017) and fog (Mazzoleni et al., 2010) organic matter indicate that the formulas uniquely common to only PMO-2 have higher O/C, which supports aqueous phase processing during transport. These results are provided in Figure 3.25 and Table 3.7. Studies have shown that the reactive species emitted from anthropogenic plumes ($SO_2$, $NO_x$, $O_3$) can play a role in the oxidation of the organic species that are dissolved in water (Blando and Turpin, 2000; Chen et al., 2008; Ervens et al., 2011); furthermore, studies have shown aerosol liquid water content contributes to aqueous production of SOA (Volkamer et al., 2006; Lim et al., 2010). The elevated RH extracted from the GFS for this plume (Figure 3.23) indicates the presence of aerosol liquid water and is consistent with its ubiquitous nature (Nguyen et al., 2016). Additionally, PMO-2 had a strongly elevated non-sea salt sulfate concentration relative to PMO-1 and PMO-3, which also indicates aqueous phase processing (Crahan et al., 2004; Yu et al., 2005; Sorooshian et al., 2007; Hoyle et al., 2016). Oxalate, another well-known marker of potential aqueous phase processing (Warneck 2003; Crahan et al., 2004; Yu et al., 2005; Sorooshian et al., 2007; Carlton et al., 2007), was also elevated in PMO-2. The organic mass fraction of oxalate was 9.4 % in PMO-2 compared to 2.3 % and 3.0 % in PMO-1 and PMO-3. The nitrate concentration in PMO-2 was very low compared to PMO-1 or PMO-3 (Table 3.1), supporting aqueous phase processed aerosol in PMO-2. While clearly gas phase SOA cannot be excluded,

110

several lines of evidence suggest that aqueous phase oxidation likely influenced the
chemical and physical characteristics of the PMO-2 aerosol to a larger extent than those
of PMO-1 and PMO-3 based on the observed molecular characteristics, major ion
concentrations (Figure 3.26), and the model simulated transport pathways and GFS
meteorology.
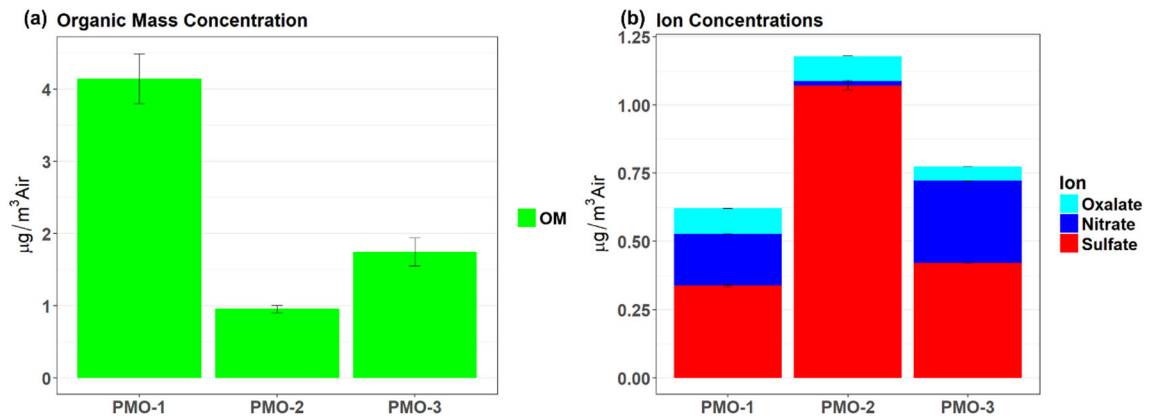


**Figure 3.26.** Organic mass concentrations (a) and sulfate, nitrate and oxalate
concentrations (b).

# 4 Tandem MS/MS Fragmentation of PMO-1

## 4.1 Precursor Molecular Complexity and Composition

### 4.1.1 Sample Overview

The sample described here is a long range transported biomass combustion aerosol sample from North America collected at the PMO on June 27, 2013. The air mass history was determined using the FLEXPART model as described in Schum et al. (2018). FLEXPART is a Lagrangian model that releases thousands of massless particles and follows their path back in time using information from the Global Forecast Service (GFS). This modeling provides information about the transport path, time, and emission sources, all of which are important when interpreting the results from an analysis of organic aerosol. More detailed discussion of FLEXPART can be seen in Chapter 2.

### 4.1.2 Full Scan vs. Segmented Scanning

When analyzed using full scan (m/z 100-1000), 3168 identified monoisotopic masses were identified (Schum et al. 2018). In this work a segmented scanning method was applied to the range of m/z 162-468 and m/z 518-523, using 6 m/z segments centered every 5 m/z units. Doing this over 9000 monoisotopic molecular formulas were assigned. In contrast, only 2051 monoisotopic molecular formulas were assigned in the same range for the full scan sample. Segmented scanning has been shown to increase the sensitivity of FT-ICR MS measurements (Southam et al., 2007), allowing better detection of low concentration or low ionization efficiency molecular species, though it has not been used previously for atmospheric aerosol samples. The reasons for this were discussed in

Chapter 2. This increase in the number of identified molecular formulas demonstrates and increased molecular complexity that has been overlooked in typical ultrahigh resolution MS analyses of organic aerosol.

Over 90% of the molecular formulas from the full scan within the same mass range were also identified using the segmented scanning approach. An example of the increased mass spectral complexity is shown in Figure 4.1 using the reconstructed mass spectra for each sample at nominal m/z 445. Over 50 molecular formulas were assigned to the isobaric peaks at this nominal mass with the spectral stitching approach, while only 11 CHO molecular formulas were assigned using the full scan approach. In addition to an increase in the absolute number of identified molecular formulas this figure highlights an increase in the detection of heteroatom containing molecular formulas. The heteroatom containing molecular formulas are a significant component of most environmental (Willoughby et al., 2016, Wozniak et al., 2014), but the frequently do not ionize efficiently. The lower intensity peaks can be overlooked in full scan analyses.
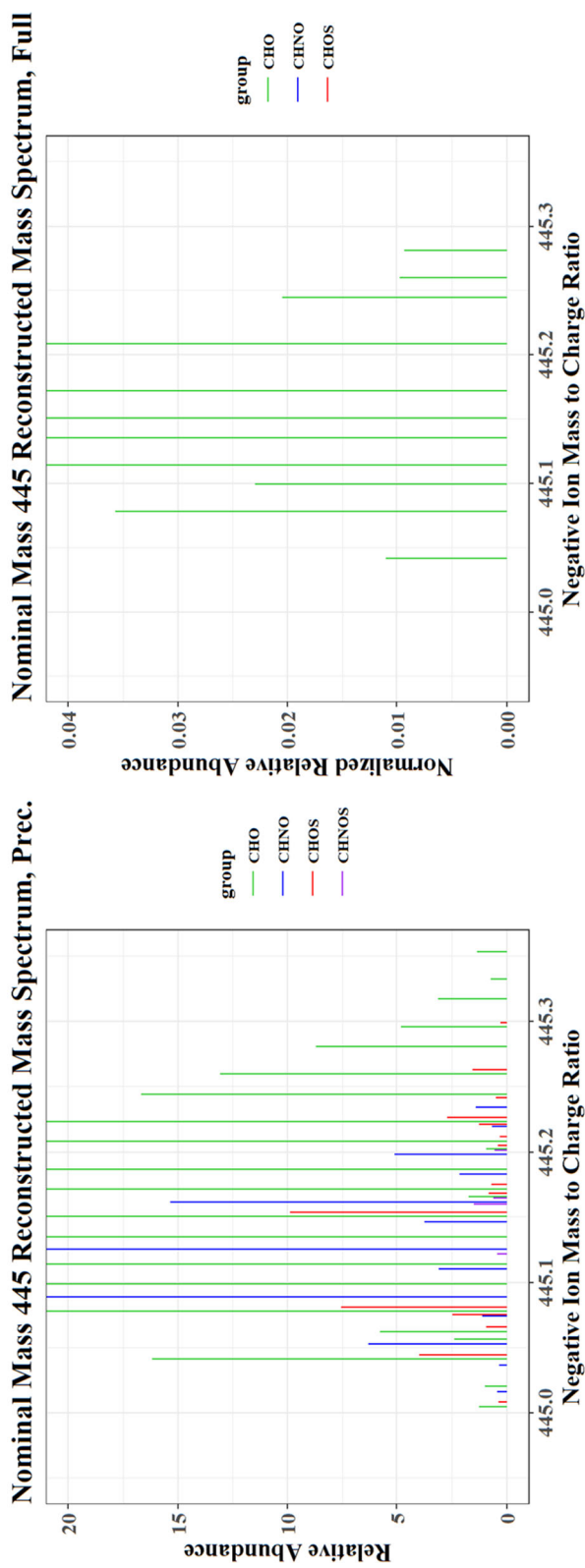
**Figure 4.1.** Reconstructed mass spectra of nominal mass 445 for the segmented scan data (left) and the full scan (right). A total of 57 monoisotopic molecular formulas are observed in the segmented scan version, compared to 11 monoisotopic molecular formulas observed in the full scan version. Also note the presence of nitrogen and sulfur containing species in the segmented scan version.

The elemental group analysis provides an interesting comparison of the two scan methods. In the full scan, CHO molecular formulas make up roughly 60% of the total number of molecular formulas, followed by 30% CHNO, and 10% CHOS. In contrast, the window scan, the CHNO molecular formulas make up 42 %, followed by 26%, then CHOS and CHNOS with 16% and 15.6% respectively. This shift in the elemental groups suggests that CHNO molecular formulas may be more prevalent in organic aerosols than previously reported a full scan of the sample. There is also an increase in the number of observed CHOS and CHNOS molecular formulas in the results from the spectral stitching approach compared to full scan analysis. Combined these sulfur containing classes make up roughly 30% of the identified molecular formulas. Similarly, the nitrogen containing formulas make up 57% of the total formulas. Overall the N or S containing molecular formulas represent a greater fraction than the CHO formulas.

### 4.1.3   Heteroatoms

The reason for the increase in the N and S containing molecular formulas is due to their low intensity, compared to CHO. Perhaps the CHO molecular formulas ionize more efficiently with negative ESI than the CHNO, CHNOS, and CHOS molecular formulas. Since the instrument limits the total number of ions, those with a higher efficiency may out compete those with a lower efficiency even if they have the same mass concentration. Many studies have shown the importance of nitrogen containing species in biomass combustion organic aerosol (Willoughby et al., 2016, Desyaterik et al., 2013; Lin et al., 2015), so the presence of a large number of such species in this sample is consistent with those observations. In fact, when compared to the elemental group breakdown reported

115

by Willoughby et al. (2016) the results here closely match, despite the difference in the age of the ambient biomass combustion. Since the spectral stitching method was required to observe these nitrogen containing compounds, it may also suggest that they are less abundant. Many of the observed nitrogen species are defined as aromatic based on the modified aromaticity index ($AI_{mod}$) (Koch and Dittmar, 2006;2015). Multiple studies have suggested that aromatic nitrogen species are an important component of brown carbon (BrC) (Desyaterik et al., 2013; Lin et al., 2015). The presence of these compounds in long-range transported aerosol suggests a longer lifetime than previously estimated (Forrister et al., 2015). The presence of absorbing aerosol species further downwind than expected based on the typical degradation rates may be due to the aerosol phase state during transport (Shrivastava et al., 2017; Chapter 3). Viscous aerosol has an increased resistance to chemical degradation via oxidative processes (Koop et al., 2011; Hinks et al., 2016; Lignell et al., 2014; Shiraiwa et al., 2017a), and an aerosol solid phase is more likely in cold, dry atmospheres, such as that of the free troposphere.

### 4.1.4   Aromaticity

One of the most interesting differences between the two versions of this sample is the increase in the number of detected aromatic compounds. Combustion is known to have more aromatic compounds than other types of organic aerosol (Willoughby et al., 2016), in part due to the presence of lignin pyrolysis products (Simoneit et al., 1993), so the presence of aromatic species is not a surprise. In fact, in previous full scan analysis this sample was found to be the most aromatic of the three samples analyzed (Schum et al., 2018). The increase of aromatic species can be seen in the van Krevelen plot in Figure

116

4.2 which shows the identified molecular species separated by elemental groups. The molecular formulas in the low H/C and low O/C region of the plot are considered to be aromatic and a large number of these aromatic species contain nitrogen. This suggests an importance of aromatic nitrogen in biomass combustion aerosol consistent with observations from other studies (Willoughby et al., 2016, Desyaterik et al., 2013; Lin et al., 2015). Using the:$AI_{mod}$ calculation developed by Koch and Dittmar (2006; 2015) the extent of unsaturations can be estimated. The results are shown using a histogram containing the molecular formulas classified by $AI_{mod}$ and by elemental group (Figure 4.3). While overall the aromatic and condensed aromatic molecular formulas are still the least common groups, they represent a much larger fraction of the total than was previously reported in Schum et al. (2018). The improved observation of aromatic compounds is important because they may have absorbing characteristics, in turn leading to increased aerosol absorption (Chakrabarti et al., 2010; Desyaterik et al., 2013). Aromatic compounds are also a significant component of brown carbon (BrC), which is a topic of great interest (Chakrabarti et al., 2010; Desyaterik et al., 2013) due to its absorption.
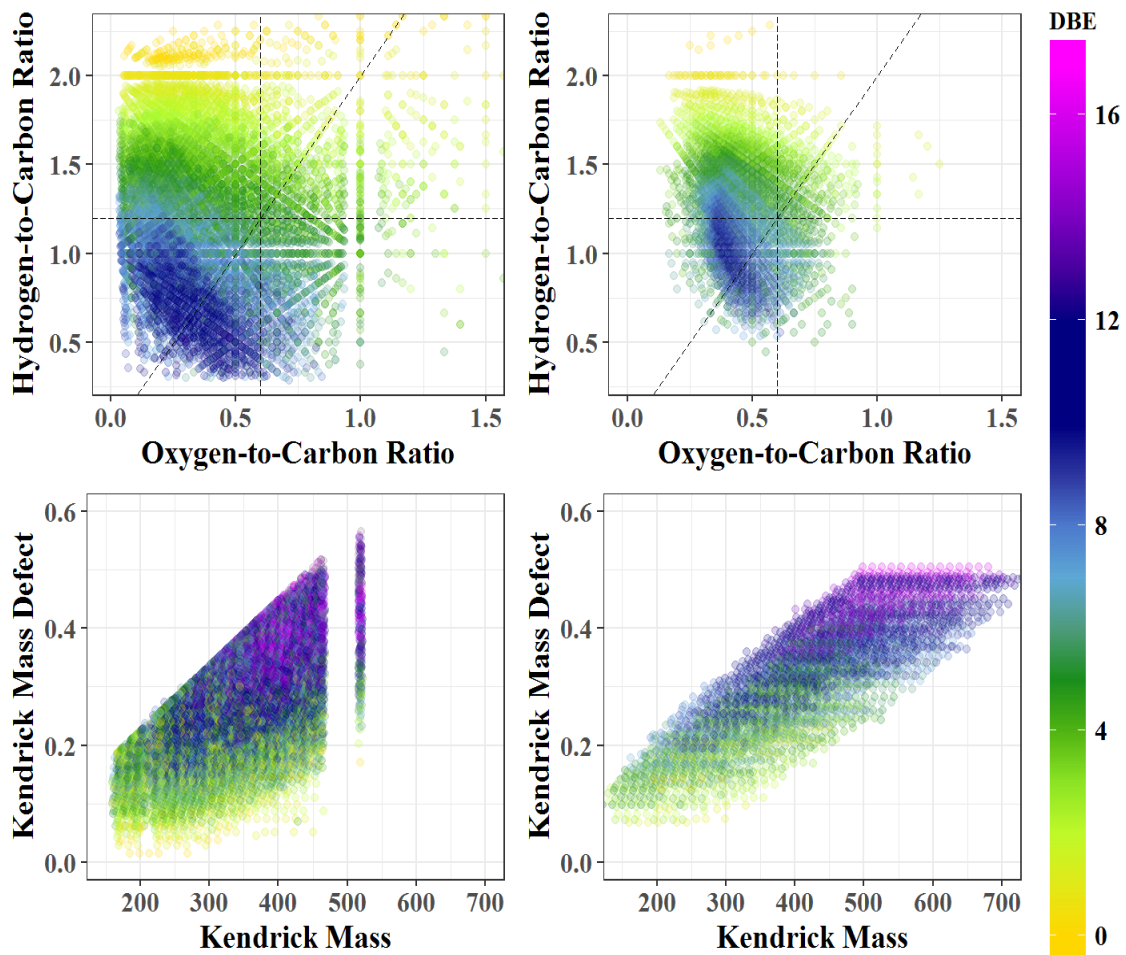
**Figure 4.2.** VK and KMD plots for the Precursor data (Left) and the full scan data (Right). The increased number of species in the lower left-hand corner of the segmented scan VK indicates an increased number of aromatic compounds. Color is determined by the double bond equivalents (DBE).
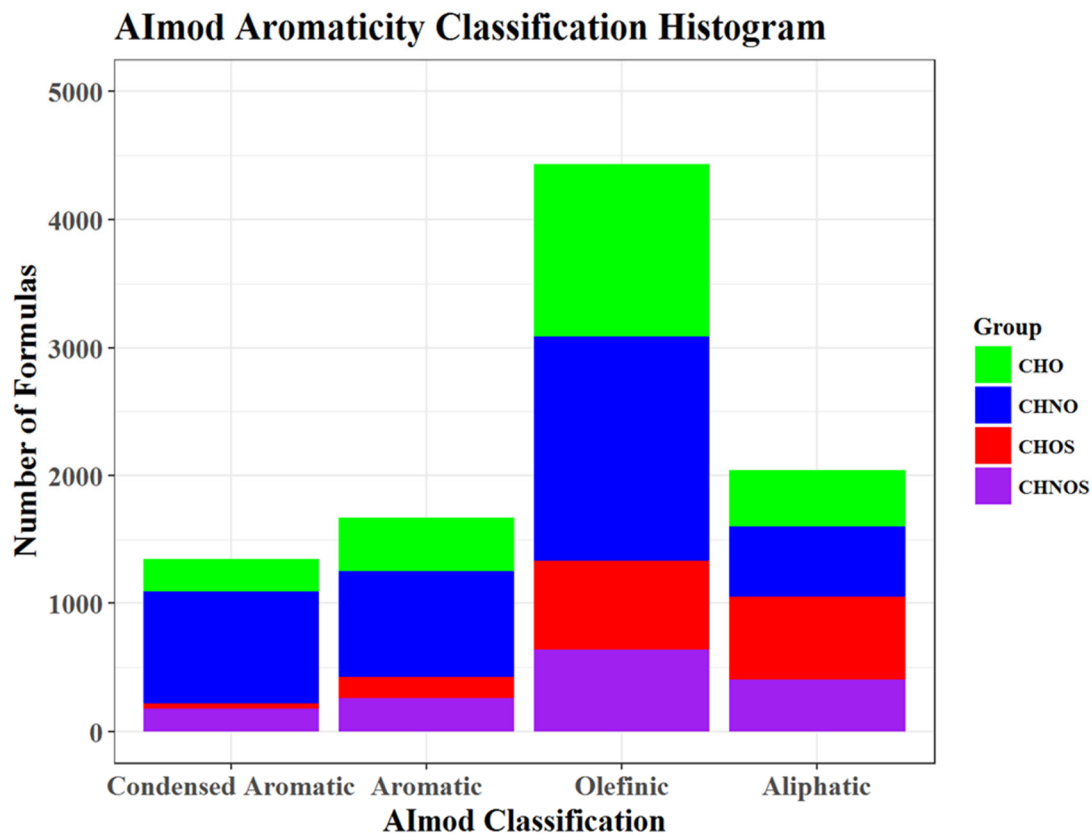
**Figure 4.3.** Histogram for number of detected formulas separated by aromaticity and group. Aromaticity was calculated using the $AI_{mod}$ calculation. Note the large number of condensed aromatic and aromatic species present.

### 4.1.5 Oxidation

Another interesting property of the aromatic species is their molecular average oxidation state of carbon as estimated by the OSc equation in Kroll et al. (2011) (Eq. 4.1).

$$OS_C \approx 2 * \frac{\#O}{\#C} - \frac{\#H}{\#C} - 5 * \frac{\#N}{\#C} - 6 * \frac{\#S}{\#C} \quad \textbf{Eq. 4.1}$$

When the OSc is plotted as shown in Figure 4.4, it is clear that on average the aromatic and condensed aromatic species in CHO and CHNO species are more oxidized than the olefinic or aliphatic species. This level of oxidation may be an example of water-soluble soot. Decessari et al. (2002) studied soot oxidation using ozone and found that it became water soluble with multiple acidic functional groups (carboxylic acids). The MS/MS analysis performed on this sample provides some additional insight into the applicability of this concept to this atmospheric aerosol sample.
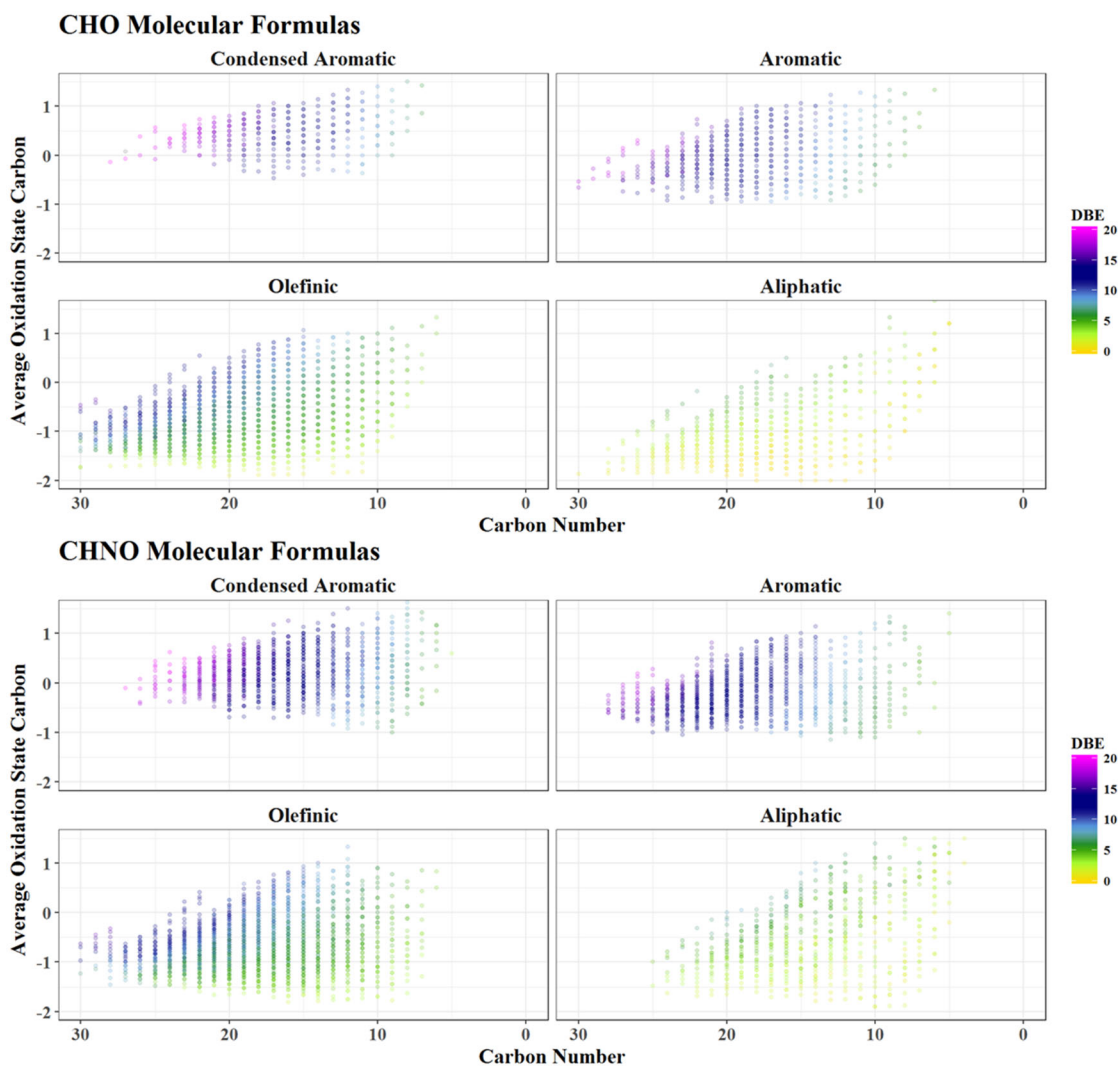
**Figure 4.4.** Aromaticity separated Average Oxidation State of Carbon vs. Carbon Number plot. The top four plots correspond to the CHO molecular formula, while the bottom four correspond to CHNO molecular formulas. In both cases, the aromatic compounds are generally more oxidized than the species that are classified as olefinic or aliphatic.

## 4.2 MS/MS Fragmentation Results

### 4.2.1 Background and Reasons for Functional Group Analysis

MS/MS analysis provides functional group information for common neutral losses. Structural information is important for model predictions of organic aerosol chemistry and their physico-chemical properties, such as the hygroscopicity (Clegg et al., 2019 in review; Reid et al., 2018; Petters et al., 2017), viscosity (Rothfuss and Petters, 2016; Reid et al., 2018; Song et al., 2016; Grayson et al., 2017), and light absorbance (Phillips and Smith, 2014; 2015). Despite the value of this information, relatively few studies have done detailed MS/MS analysis of the functional groups present in atmospheric organic matter (LeClair et al., 2012). Most studies focused on functional group information use bulk analytical methods such as FT-IR or NMR spectroscopy (Takahama et al., 2013, Hawkins and Russell, 2010, Decesari et al., 2000; 2007). Analysis of these complex natural mixtures with MS/MS fragmentation is challenging because of the spectral complexity at each nominal mass. Thus, it is impossible to get a clean fragmentation pattern of any single ion of interest. Additionally, each ion likely represents a mixture of several different isomers (Zark et al., 2017). This means that each molecular formula can show the loss of more functional groups than are chemically feasible for a given structure, thus it is difficult to determine a structure for the exact masses. However, the predictive power associated with functional groups led us to analyze PMO-1 using collision induced dissociation (CID) for the same windows described in Section 4.1. To find matching precursor and fragment ions the precursor formulas were used to create the expected fragment ion formula and then this expected formula was checked against the

assigned fragment ion formulas. If there was a match, then the precursor formula was considered to have demonstrated that loss.

## 4.2.2   Observed Neutral Losses and Corresponding Functional Groups

There are many pathways that lead to neutral losses during ion fragmentation. Most of the described pathways are for positive ions (de Hoffman and Stroobant, 2007). While these paths may not be directly applicable to negative ions, the general principles are similar (Gross, 2017).

Of the over 9000 molecular formulas identified for PMO-1 using this segmented scan method, 7181 distinct molecular formulas were identified with at least one detectable neutral loss. The majority of the formulas that did not have a neutral loss had low relative abundances. Therefore, the fragment ions were likely below the signal-to-noise threshold. For reference, the average intensity of a precursor mass with at least one neutral loss is 158,660, while the average intensity for a precursor mass without any neutral losses is 41,945. A histogram showing the studied neutral losses is shown in Figure 4.5. The color represents the elemental group (Panel a) and aromaticity (Panel b). The most common neutral losses were $CO_2$, $C_2H_4O$, and $H_2O$. The loss of $CO_2$ and $H_2O$ is consistent with a previous study of atmospheric organic matter by LeClair et al. (2012), but the $C_2H_4O$ is different. The identity and reason for the loss of $C_2H_4O$ will be discussed in Section 4.2.3. The loss of $CO_2$ represents of a carboxyl group, and the loss of $H_2O$ represents either a hydroxyl (-OH) group, or a carboxyl group (-COOH) that loses its hydroxyl group. The water loss from a carboxylic acid can either occur by simple cleavage of the hydroxyl group on the carboxyl group (Jensen et al., 1985; Bowie et al., 1990; Kerwin et

al., 1996) or by rearrangement involving two carboxyl groups. In the formation of a five or six membered ring $H_2O$ is eliminated (Leenheer et al., 2001; Witt et al., 2009). In both cases, the neutral loss of $H_2O$ is expected to be favored for more aliphatic molecules (Kerwin et al., 1996; Leenheer et al., 2001). $C_2H_4O$ is likely a neutral loss of $-C(O)CH_3$. Other commonly observed neutral losses were CO, $CH_4O$, $C_2O_4$, and $H_4O_2$. The CO neutral loss is often related to the carbonyl functional group (C=O), $CH_4O$ is likely a methoxy group ($-OCH_3$), while $C_2O_4$ and $H_4O_2$ represent double losses of $CO_2$ and $H_2O$ respectively. While CO loss is often related to carbonyl functional groups, it can also be eliminated from phenols via multistep rearrangements (Gross, 2017). The large number of ways that CO can be eliminated makes it difficult to interpret the molecular structure (Gross, 2017).
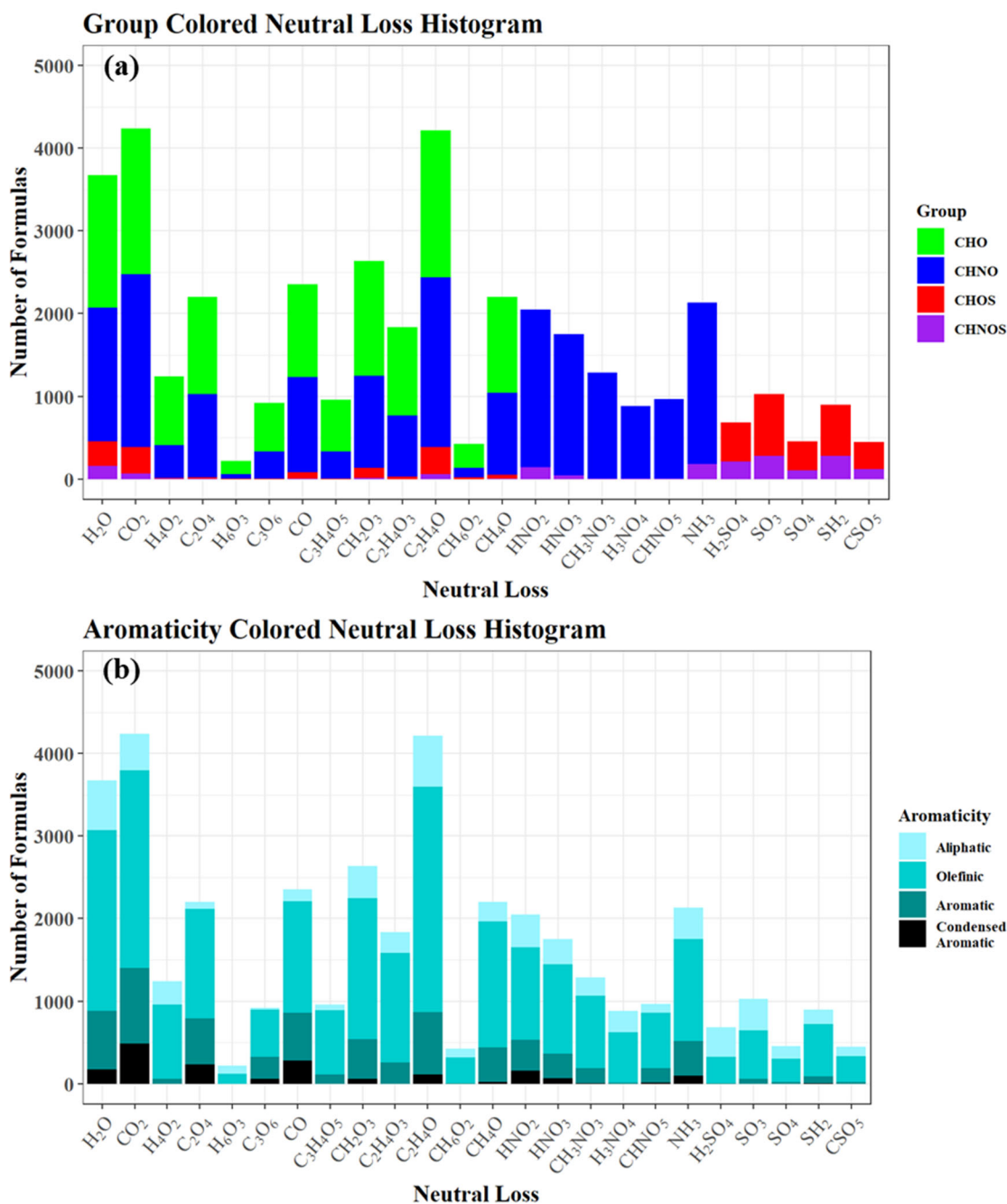
**Figure 4.5**. Bar plot showing the observed neutral losses. Colored by elemental group (a), and aromaticity index(b)

### 4.2.3   Carbonyl Functional Groups and Relationship to Aging

The carbonyl groups are of particular interest because they are known to be a major functional group in fresh biomass burning aerosol (Hawkins et al., 2010; Takahama et al., 2013). The abundance of carbonyl neutral losses provides additional confidence in our assessment of PMO-1 as a biomass burning sample, and may relate to the lower than expected extent of oxidative aging that was the focal point of Chapter 3. As the aerosol ages, according to Hawkins and Russell (2010), the number of ketone groups decreases and the number of carboxyl groups increases. In their study of biomass burning aerosol the carboxylic acid to ketone ratio increased from 0.35 to 1.3 over 4 days of transport (Hawkins and Russell, 2010), Similarly, another study was unable to detect ketone groups after 5 days of transport (Shaw et al., 2010). In this study, a rough comparison based on the number of molecular formulas with a $CO_2$ loss or a $C_2H_4O$ loss yields a ratio of approximately 1. This falls within the reported range for 4 days of transport (Hawkins and Russell, 2010), but this sample was transported for approximately 7 days from its major emission source before collection at the PMO, potentially indicating decreased rate of aerosol aging.  This estimation may be limited because only carbonyls that are adjacent to the terminal carbon are considered and the $CO_2$ losses may not represent all carboxyl groups present in the structures. However, the observed decreased extent of aerosol aging may support the hypothesis presented in Schum et al. (2018) and Chapter 3 that the samples had lower oxidation which was attributed to a solid-state during transport in the free troposphere.

### 4.2.4 Functional Group Analysis

#### *4.2.4.1 Data Visualization with Principal Components Analysis*

Due to the extent of available data from the large range of precursor and fragment ions

that were analyzed for this sample it became necessary to use different data visualization

methods than typically used. A major goal of this analysis was to provide insight on the

general molecular characteristics (oxidation and aromaticity) of the molecular formulas

that have certain neutral losses. Also of interest are the relationships between different

functional groups. A long-standing statistical method for investigating the correlations

between different variables is principal components analysis (PCA). It has been used in

the past to investigate the correlation between different ambient samples (Zhao, 2014;

Wozniak et al., 2014). Due to the large number of data points available within this data

set we chose to use PCA to find correlations between the different types of neutral losses

observed for the molecular formulas.

#### *4.2.4.2 Carbon, Hydrogen, and Oxygen Containing Neutral Losses*

4.2.4.2.1 PC1 vs. PC2

The first set of neutral losses investigated were the C, H, and O based neutral losses,

including; $CO_2$, $C_2O_4$, $C_3O_6$, $H_2O$, $H_4O_2$, $H_6O_3$, $CH_2O_3$, CO, $C_2H_4O$, $CH_4O$, $C_2H_4O_3$, and

$CH_6O_2$. These losses represent several of the most commonly observed neutral losses.

The scree plot shows the amount of variance explained by each principal component (PC)

and is shown in Figure 4.6. PC1 always accounts for the most variance, followed by PC2,

etc. This plot indicates how many of the PCs are necessary to account for a majority of

the variance in a sample. In this case, PC1 and PC2 would be enough to account for

49.7% of the variance, although the inclusion of PC3 and PC4 provides an additional 9.6 and 8.7%, respectively.
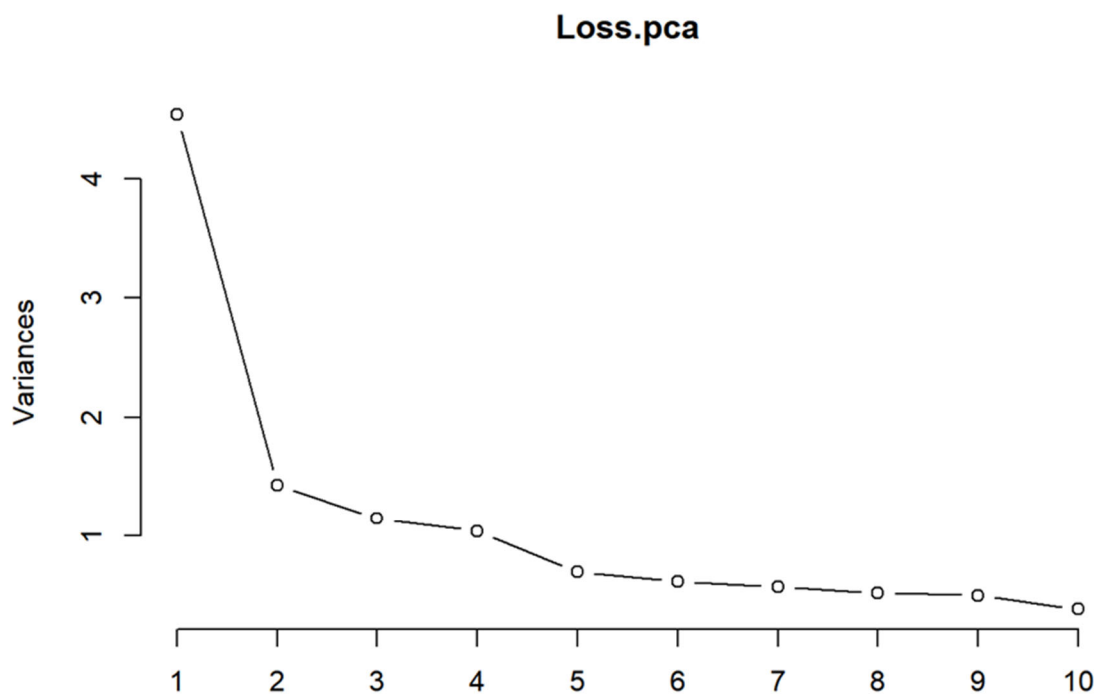


**Figure 4.6**. Scree plot demonstrating the variances accounted for by each PC for the PCA analysis of molecular species showing neutral losses of $CO_2$, $C_2O_4$, $C_3O_6$, $H_2O$, $H_4O_2$, $H_6O_3$, $CH_2O_3$, $CO$, $C_2H_4O$, $CH_4O$, $C_2H_4O_3$, or $CH_6O_2$. The x axis is the PC number (PC1, PC2, etc.).

First, we look at the PCA biplot for PC1 and PC2 (Figure 4.7). The points on the plot

represent a scatter plot of the PC values for the different molecular formulas in the

sample, some are darker than others because there are many overlapped data points in the

exact same location of the plot. The arrows coming from the center of the plot show the

specific loadings of the variables used for the PCA analysis, in this case they are the

observations of neutral losses. In general, the longer an arrow is, the greater its

contribution to the variance of the sample. The direction of the arrow indicates which PC

the variable is most correlated with. The angle between points is also important, if two

arrows are in the same direction that means that they are correlated, if the arrows are at

90˚ from each other, they are not correlated, and if the arrows are at 180˚ they are

negatively correlated. The arrows in Figure 4.7 indicate that CO, $CO_2$, and $H_2O$ are

correlated with each other, but they are not correlated with $CH_6O_2$ using PC1 and PC2. A

general correlation also exists between $C_3O_6$, $H_4O_2$, $C_2H_4O_3$, and $H_6O_3$. In general, most

of the losses are somewhat correlated using these two PCs, with the exception of the

$CH_6O_2$ loss and the losses 90˚ from that loss. To interpret and provide molecular

composition context to these results, the scatter plots of the PC values are shown with an

indication of whether a molecular formula showed a particular neutral loss. In addition,

some metrics of the chemical compositions, including: DBE, the number of oxygen

atoms, and the number of carbon atoms are shown (Figure 4.8). The scatter plot points

are jittered to avoid the over-plotting that was observed in Figure 4.7.
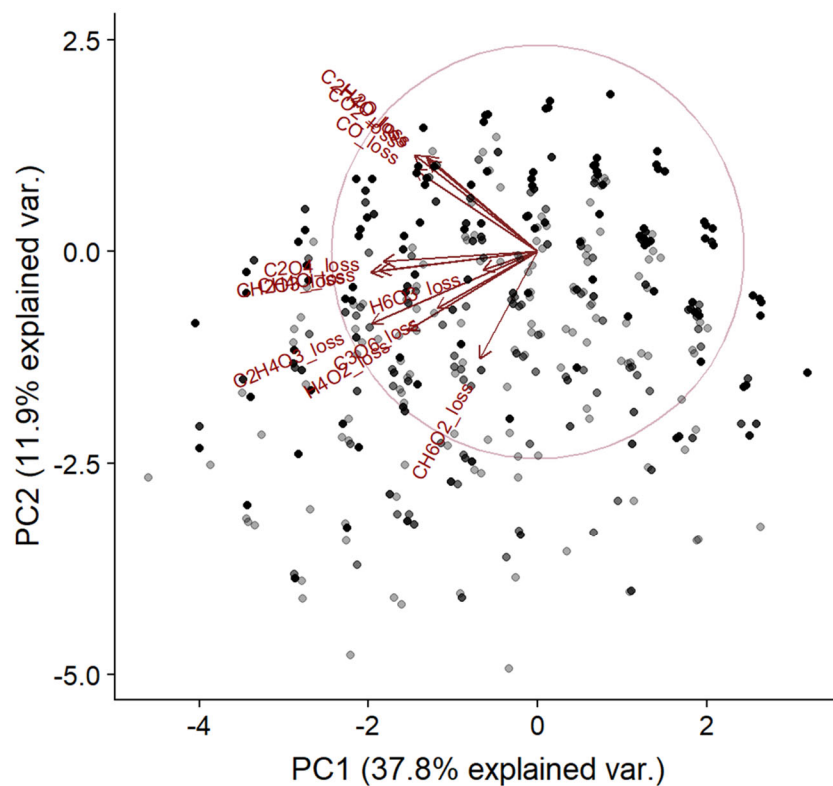
Figure 4.7. PC1 vs. PC2 PCA biplot for molecular species with a neutral loss of $CO_2$, $C_2O_4$, $C_3O_6$, $H_2O$, $H_4O_2$, $H_6O_3$, $CH_2O_3$, CO, $C_2H_4O$, $CH_4O$, $C_2H_4O_3$, or $CH_6O_2$. Most are correlated somewhat, except $CH_6O_2$ with the neutral losses of CO, $CO_2$, $H_2O$, and $C_2H_4O$.
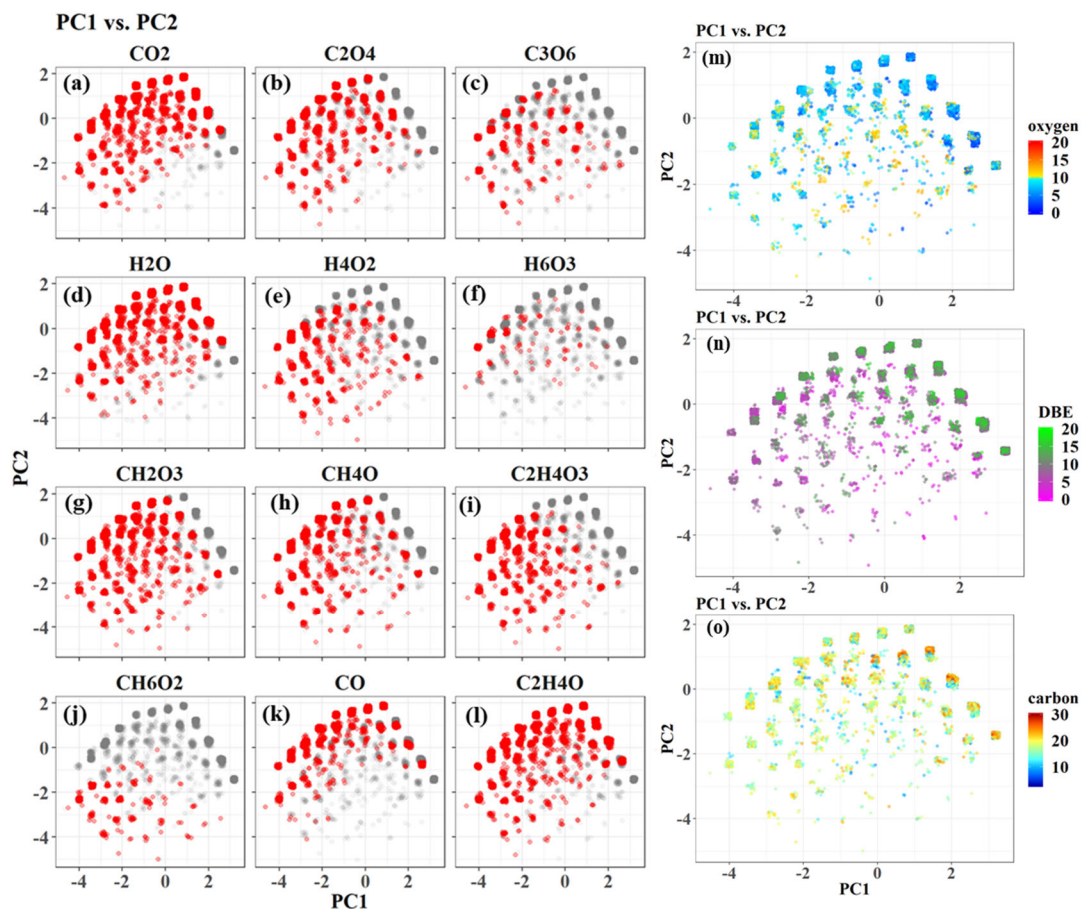
**Figure 4.8.** Colored PC plots PC1 vs. PC2. Panels a-l show the molecular species with the neutral loss (red) and without the neutral loss (gray). For example, Panel a shows the molecular species with a $CO_2$ loss in red, while the species that did not have a $CO_2$ loss, but did have one of the other losses are shown in gray. Panels m-o show the same plots, only colored with oxygen number (m), DBE (n), and carbon number (o). All of these plots use the same data as Figure 4.7, but to avoid the overlapping that occurred in that plot, the data points have been jittered, which allows the points that were overlapped to be seen.

Thus, the variation within a point can be observed. From the composition focused plots on the right side, a separation between the species with high and low oxygen, DBE, and carbon is observed. Matching these points to the regions where a specific type of loss is observed (red color on the plots) allows us to get a general idea of the characteristics associated with these neutral losses. The upper edge of the PC scatter plot contains the highest DBE values, only the molecular formulas with $CO_2$, $H_2O$, CO, and $C_2H_4O$ losses are present on this lower edge, suggesting these functional groups are more prevalent on aromatic formulas. The loss of CO is particularly frequent in the low oxygen and high DBE region of the plot. This suggests that the carbonyl losses are more prevalent from more aromatic and less oxidized species. Additionally, the similarity in PCA distribution between the CO and $C_2H_4O$ losses suggests that they are both representative of a similar functional group, likely containing a carbonyl. The distribution of CO losses differs from the $CH_4O$ (methoxy) based losses. The losses of $CH_4O$ are shifted to the lower DBE and higher oxygen regions of the plot. This is especially true when the $CH_4O$ loss happens in conjunction with a $CO_2$ or $H_2O$ neutral loss ($C_2H_4O_3$ and $CH_6O_2$ respectively). This suggests that the methoxy groups are more likely to occur from species that are more aliphatic and oxidized, relative to those showing a CO loss. This information is valuable because it provides a basis for assumptions about the most likely functional groups in an organic aerosol sample. For example, a highly oxidized and somewhat aliphatic sample, could be assumed to contain a higher number of methoxy functional groups rather than carbonyl functional groups. There are some additional observed differences for the $CO_2$

and $H_2O$ based losses, but they are more clear in the plots with PC1 and PC3, which are described in the following paragraph.

4.2.4.2.2   PC1 vs. PC3

The PCA biplot with PC1 and PC3 is shown in Figure 4.9. It does not show much separation between the majority of the loss types, but the multiple $CO_2$ and $H_2O$ losses do get separated to some extent. The multiple $CO_2$ losses are on the top half of the plot space while the multiple $H_2O$ losses are on the bottom. As in the previous biplot, the molecular formula with single $CO_2$ and single $H_2O$ losses are well correlated with each other, which indicates the molecular formulas with these losses are similar. This supports the idea that most of the $H_2O$ losses observed here are likely due to losses from carboxylic acids. This is consistent with the increased presences of carboxyl groups with increasing oxidation during transport (Hawkins and Russell et al., 2010). Even though this sample was not as strongly oxidized as may be expected based on its transport time (Schum et al., 2018; Chapter 3), it is likely that some aging took place. To get a better idea of the differences between the molecular formulas showing multiple $CO_2$ vs. multiple $H_2O$ we can look at Figure 4.10, which applies the same concept as described for Figure 4.8 with PC1 and PC3. Here a much clearer distinction between molecular formulas with higher DBE and oxygen is shown. Additionally, some separation between molecular formulas with higher and lower carbon numbers are observed. The lower half of the plot contains the molecular formulas with the highest oxygen, DBE, and carbon (Panels m-o). This is the same region where the $CO_2$ losses are most common, indicating that $CO_2$ losses are favorable for molecular formulas with an increased aromaticity and extent of oxidation.

133

In contrast, most of the H$_2$O losses are in the lower half of the plot, indicating the

molecular formulas with lower DBE and oxidation. These observations are consistent

with the two different ways that H$_2$O can be eliminated from a carboxylic acid, and

suggest that a more aliphatic molecule can more easily lose the hydroxyl from a

carboxylic acid. The loss of CO$_2$ is preferential as the molecule becomes more aromatic.

The likely reason for this is that the bond from an aromatic ring to the carboxyl group is

somewhat easy to break.
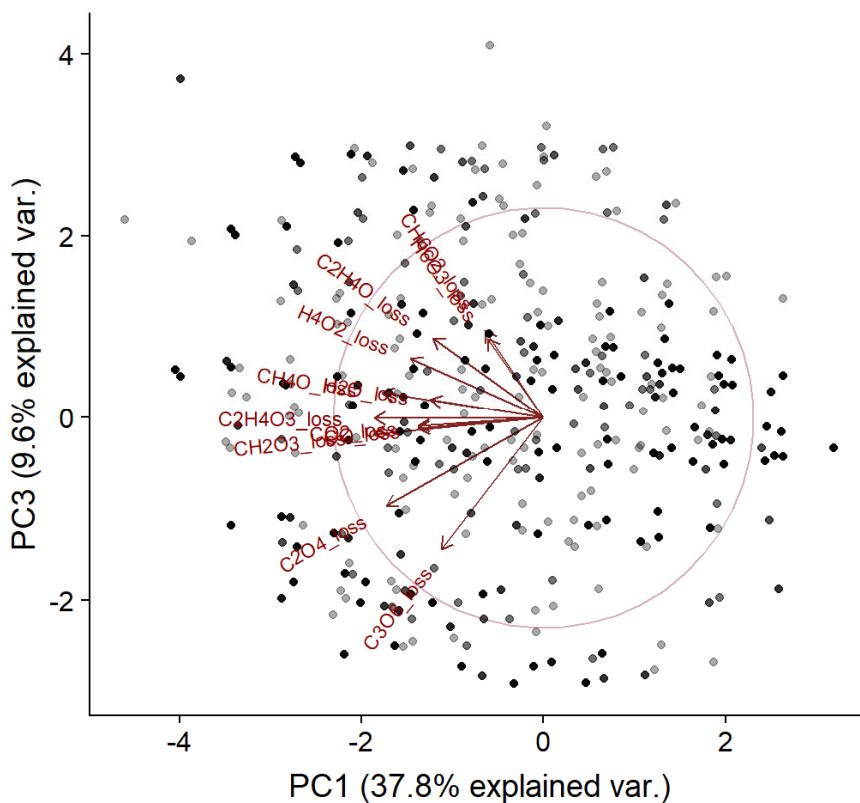


**Figure 4.9.** PC1 vs. PC3 PCA biplot for molecular species with a neutral loss of CO$_2$, C$_2$O$_4$, C$_3$O$_6$, H$_2$O, H$_4$O$_2$, H$_6$O$_3$, CH$_2$O$_3$, CO, C$_2$H$_4$O, CH$_4$O, C$_2$H$_4$O$_3$, or CH$_6$O$_2$.
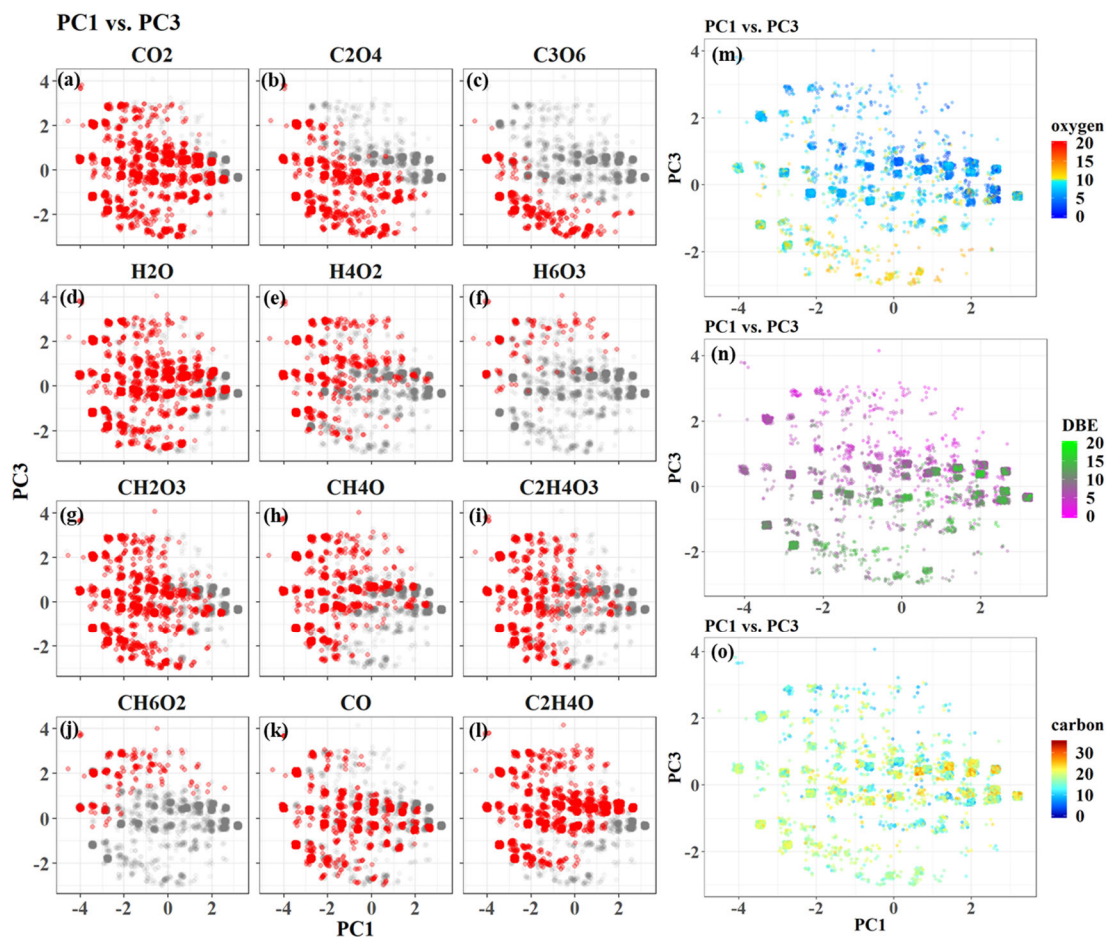
**Figure 4.10**. Colored PC plots PC1 vs. PC3. Panels a-l show the molecular species with the neutral loss (red) and without the neutral loss (gray). For example, Panel a shows the molecular species with a $CO_2$ loss in red, while the species that did not have a $CO_2$ loss, but did have one of the other losses are shown in gray. Panels m-o show the same plots, only colored with oxygen number (m), DBE (n), and carbon number (o). All of these plots use the same data as Figure 4.9, but to avoid the overlapping that occurred in that plot, the data points have been jittered, which allows the points that were overlapped to be seen.

### 4.2.4.2.3  Average Relative Abundance of Fragment Ion Comparison

This trend is illustrated in Figure 4.11 with a comparison of the relative abundance (RA) of the fragment ions that are due to the $H_2O$ and $CO_2$ losses. Only the precursor ions with both losses were considered in this comparison. It should be noted that the results from this figure are not quantitative as there are many aspects that govern the RA, such as the number of molecular formulas containing the functional group, the strength of the bond that was broken, and the stability of the fragment ion (McLafferty and Turecek, 1993). So, this comparison of RA is only intended to provide general themes about the conditions leading to a preference for a certain kind of loss. In the top panel, the average RA for the fragment ions for precursor molecular formulas within each of the aromaticity classification are shown. As described above, in aliphatic species the $H_2O$ loss fragment ion has a higher RA than the $CO_2$ loss fragment ion, but as the aromaticity increase, the $CO_2$ loss fragment ion become more abundant. This increasing importance of the $CO_2$ loss is shown more clearly using the ratio of the $H_2O$ loss fragment ion RA divided by the $CO_2$ loss fragment ion RA (bottom panel). While it is likely that when multiple $H_2O$ losses were observed ($H_4O_2$, $H_3O_6$), some of them are from actual alcohols, the similarity between the precursor peaks showing single $H_2O$ and $CO_2$ loss suggests that the majority of the $H_2O$ losses observed in this sample were due to a loss of a hydroxyl from a carboxyl group.
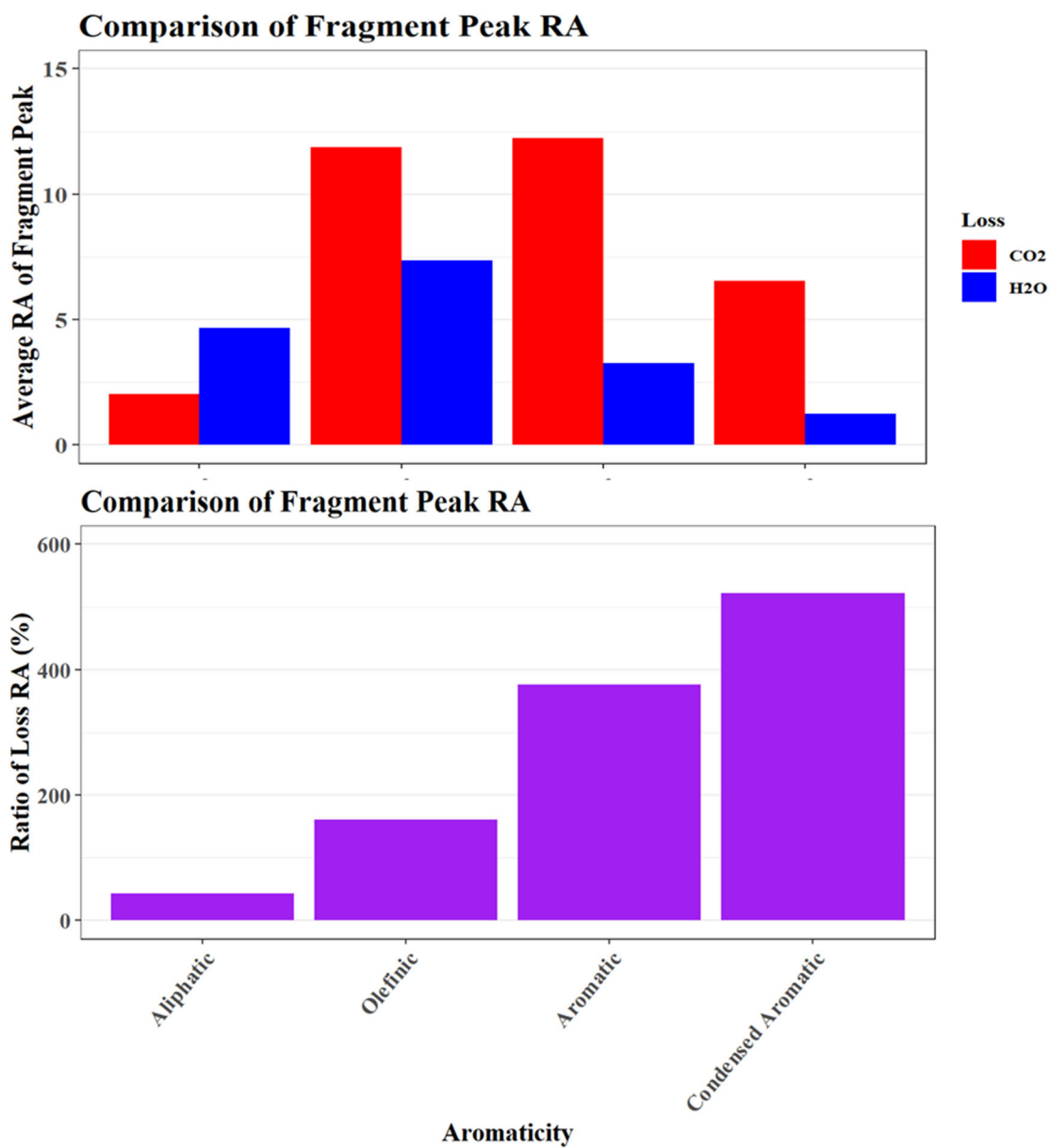
**Figure 4.11.** Comparing average RA trends for molecular species of different aromaticity with $CO_2$ or $H_2O$ losses. The top panel shows the average RA of the $CO_2$ or $H_2O$ loss fragment ion for precursor ions within each aromaticity classification. The bottom panel shows the ratio of the average RA shown in the top panel. The ratio is calculated by dividing the RA of the $CO_2$ fragment by the RA of the $H_2O$ fragment. The tallest bar in the bottom panel demonstrates that the $CO_2$ fragment ion is on average ~ 520% larger than the $H_2O$ fragment ion.

### 4.2.5 Comparison to Storm Peak Lab Aerosol

The results for PMO-1 were then compared to the MS/MS results for a sample collected at the Storm Peak Lab (SPL). In this case there was much less correlation between the molecular formulas that show $CO_2$ and $H_2O$ losses (Figure 4.12). This suggests that the molecules with these losses were somewhat different from each other, which may be due to the SPL aerosol having more hydroxyl functional groups. The aerosol collected at SPL likely had very little aging and different sources, and transport conditions. The SPL aerosol represents a relatively typical boundary layer aerosol with an influence of biogenic SOA collected at a rural site (Mazzoleni et al., 2012) and thus it doesn't have much influence from biomass burning. The observed differences between these two samples suggests that different ratios of functional groups may be present in different types of aerosol, highlighting the need for detailed molecular characterization of representative aerosol from many locations and transport paths.
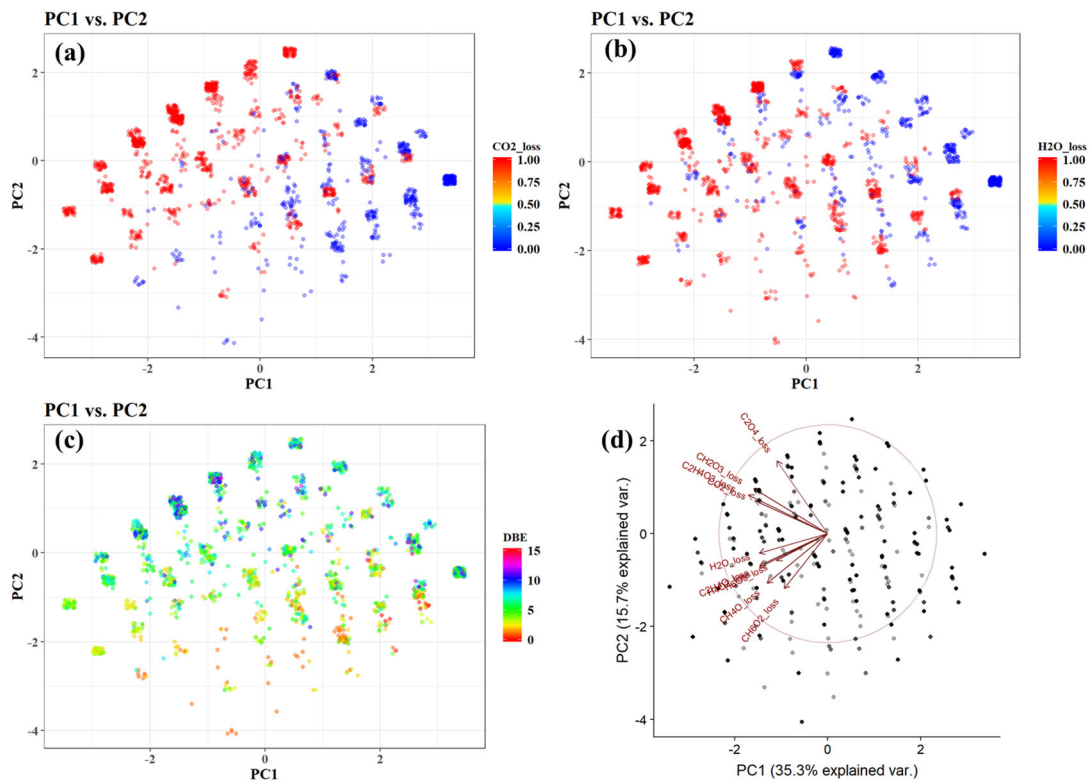
**Figure 4.12.** PC1 vs. PC2 PCA plots for molecular species showing the neutral losses of $CO_2$, $C_2O_4$, $C_3O_6$, $H_2O$, $H_4O_2$, $H_6O_3$, $CH_2O_3$, $CO$, $C_2H_4O$, $CH_4O$, $C_2H_4O_3$, or $CH_6O_2$ in an aerosol sample collected at the Storm Peak Laboratory. Panels a and b show the molecular species with a $CO_2$ (a) or $H_2O$ (b) loss in red, while molecular species that don't show that loss are in blue. Panel c shows the plot colored by DBE. Panel d shows the PCA biplot for the species showing the selected losses in the sample. Panels a-c are jittered to better demonstrate the complexity within each point in Panel d.

### 4.2.6 Analysis of Species with Multiple $CO_2$ and $H_2O$ Neutral Losses

#### 4.2.6.1 PC1 vs. PC2

To do some additional analysis of the $CO_2$ and $H_2O$ losses, another PCA analysis was done with only the single, double, and triple $CO_2$ and $H_2O$ losses. The PCA biplot with PC1 and PC2 is shown in Figure 4.13. All three of the $CO_2$ losses are well correlated with each other, but the $H_2O$ losses are somewhat less correlated. The single $H_2O$ loss is correlated with the $CO_2$ losses, while the $H_6O_3$ loss is well separated from the other losses and it is only somewhat correlated with the $H_4O_2$ loss. This suggests that the molecular formulas with multiple $H_2O$ losses are not very similar to the molecular formulas with $CO_2$ losses. This may imply that the molecular formulas with multiple $H_2O$ losses are alcohols. The loss type PCA plots for these six loses are shown in Figure 4.14. As observed in Figure 4.13, the $H_6O_3$ losses are separated from the other observed losses, although there are a few molecular formulas that show other losses in addition to the triple water loss. For this PCA plot, the $CO_2$ class losses are generally shifted towards the left side of the plot, while the $H_2O$ class losses tended to shift down, especially as more $H_2O$ losses are observed. The observed shift towards the left for the $CO_2$ class losses correlates with an increase in the number of oxygen and DBE, which is consistent with the previous observations. Meanwhile, the $H_2O$ class losses correlate with a decreased DBE and number of oxygen atoms.
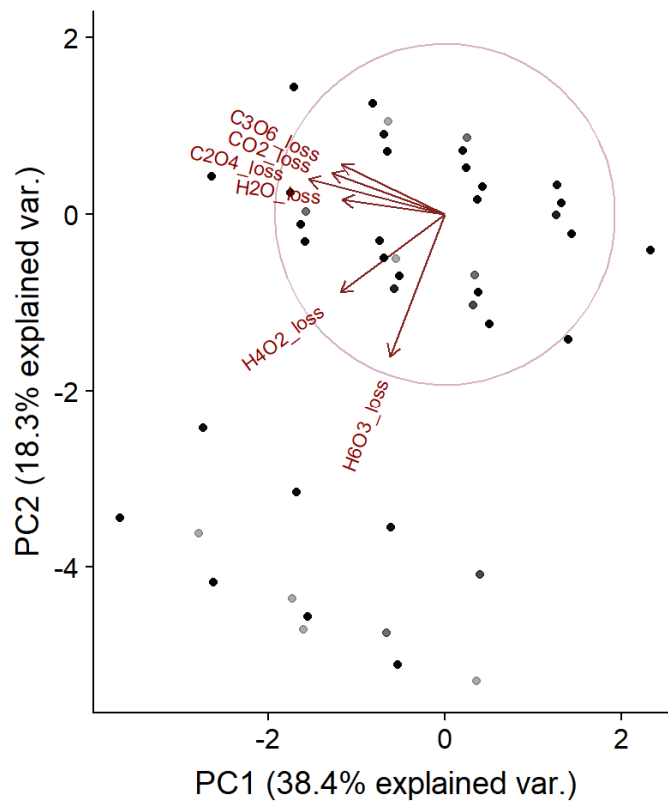
**Figure 4.13.** PC1 vs. PC2 PCA biplot for molecular species with a neutral loss of $CO_2$, $C_2O_4$, $C_3O_6$, $H_2O$, $H_4O_2$, or $H_6O_3$.
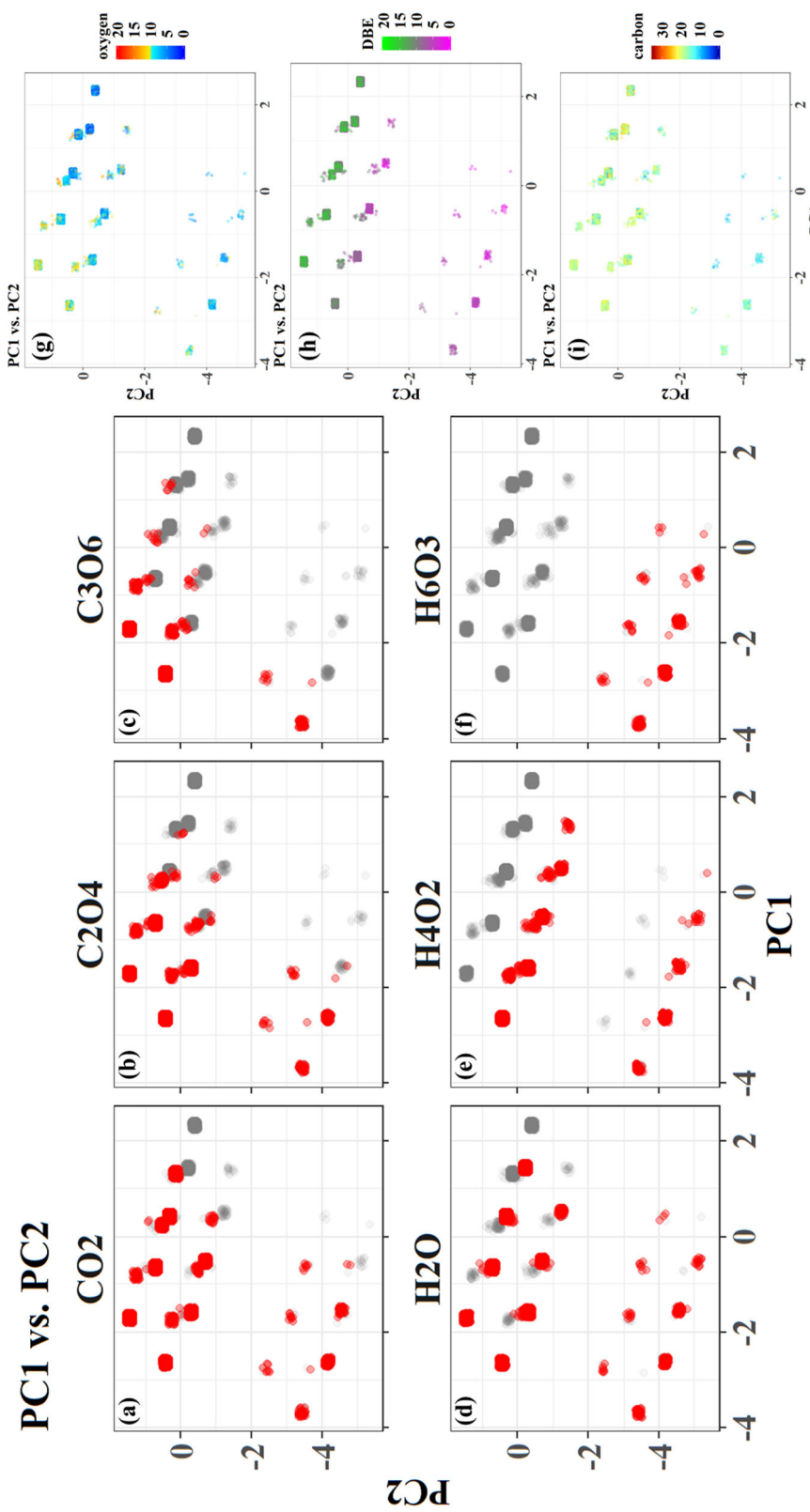
**Figure 4.14.** Colored PC plots PC1 vs. PC2. Panels a-f show the molecular species with the neutral loss (red) and without the neutral loss (gray). For example, Panel a shows the molecular species with a $CO_2$ loss in red, while the species that did not have a $CO_2$ loss, but did have one of the other losses are shown in gray. Panels g-i show the same plots, only colored with oxygen number (g), DBE (h), and carbon number (i). All of these plots use the same data as Figure 4.13, but to avoid the overlapping that occurred in that plot, the data points have been jittered, which reveals the complexity within each point in Figure 4.13.

142

One key point of interest in these plots is the properties of the molecular formulas with multiple $CO_2$ losses and triple $H_2O$ loss. These losses from the same molecular formula provide evidence for an isomeric mixture as suggested by Zark et al. (2017). They may also highlight the ability of aliphatic molecular formulas to lose $H_2O$ via the cleavage of the hydroxyl group from a carboxyl group (Jensen et al., 1985; Bowie et al., 1990; Kerwin et al., 1996). These observations may suggest a direct bond cleavage (Jensen et al., 1985; Bowie et al., 1990; Kerwin et al., 1996) instead of the indirect ring forming bond cleavage (Leenheer et al., 2001; Witt et al., 2009). This is further supported by the fact that the amount of oxygen present on these molecular formulas is on average insufficient to contain the six carboxyl groups required for the ring forming pathway. To further investigate the compositional properties of those molecular formulas with both a triple $H_2O$ loss and one of the $CO_2$ based losses, a van Krevelen (VK) plot showing was made to show the common molecular formulas (Figure 4.15a). The average O/C and H/C values are represented by triangles in the plot (Figure 4.15a). Interestingly, the shift in composition between molecular formulas with a single $CO_2$ loss vs. those that showed three $CO_2$ losses was almost entirely dependent on the H/C value, with little influence from the O/C value. For comparison, the same type of VK plot with the molecular formulas with $CO_2$ losses without a loss of $H_6O_3$ is shown in Figure 4.15b. There a strong dependence on both H/C and O/C for the molecular formulas with multiple $CO_2$ losses. The observations from these plots suggest that as the molecular formula composition shifts toward the more oxidized and aromatic species, the likelihood of observing multiple $CO_2$ losses increases.
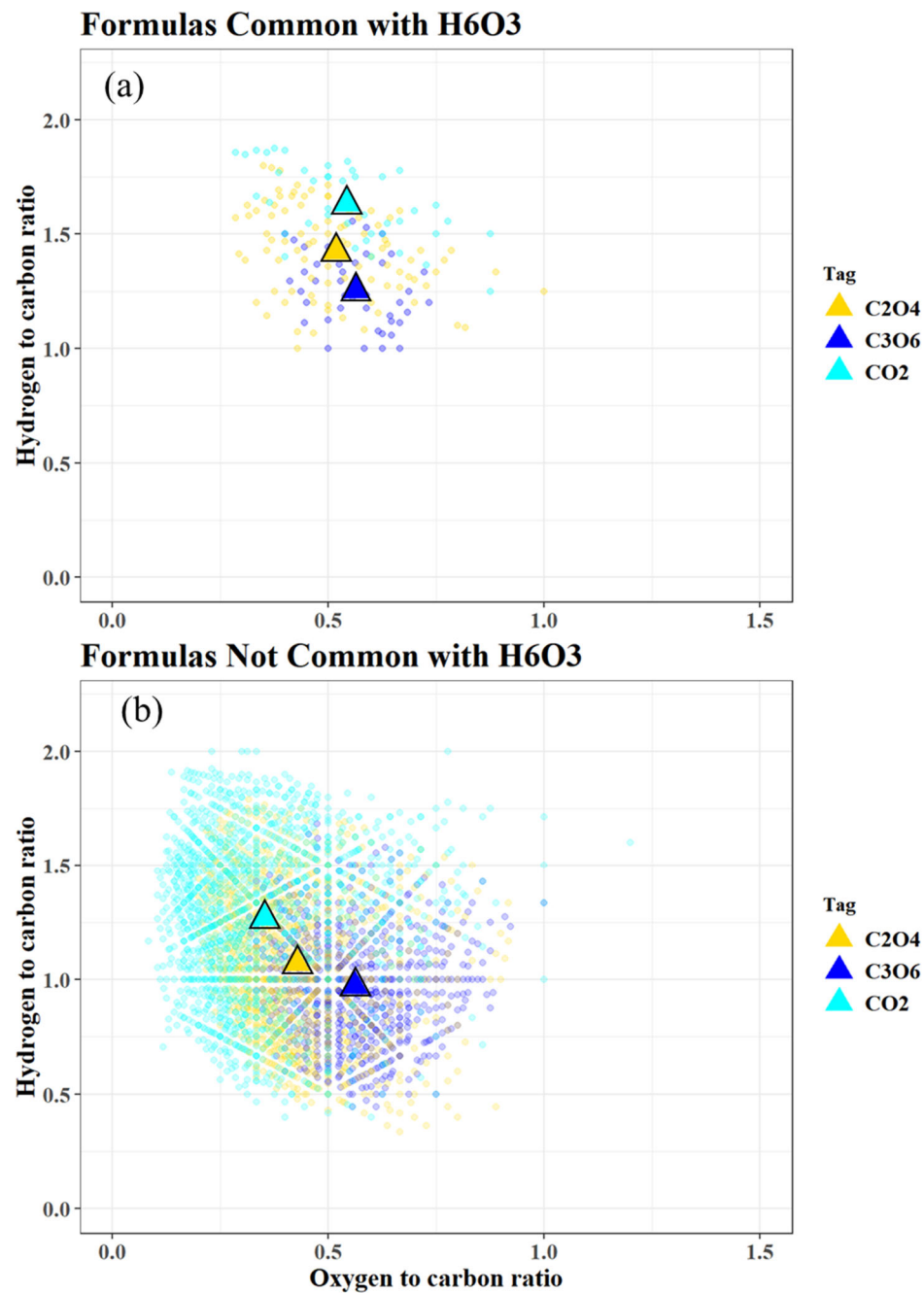
**Figure 4.15**. VK plots demonstrating formulas that exhibit $H_6O_3$ loss and one of the $CO_2$ based loses (a), and all other $CO_2$ based losses (b). The large triangles represent the average H/C and O/C ratios for the molecular formulas demonstrating $CO_2$ (cyan), $C_2O_4$ (gold), or $C_3O_6$ (blue) neutral losses.

This is highlighted in Figure 4.16 which shows the comparison of $CO_2$ and $C_3O_6$ loss fragment peak intensities analogous to Figure 4.11. The $CO_2$ losses are always the more abundant fragment ion but the relative importance of the $C_3O_6$ loss begins to increase as the aromaticity increases. This supports the observations from the VK plots in Figure 4.15. A similar plot, with $CO_2$ and $C_2O_4$ is shown in Figure 4.17. Interestingly, the pattern of an increasing RA with increasing aromaticity is not observed, the reason for this is unclear at this time.
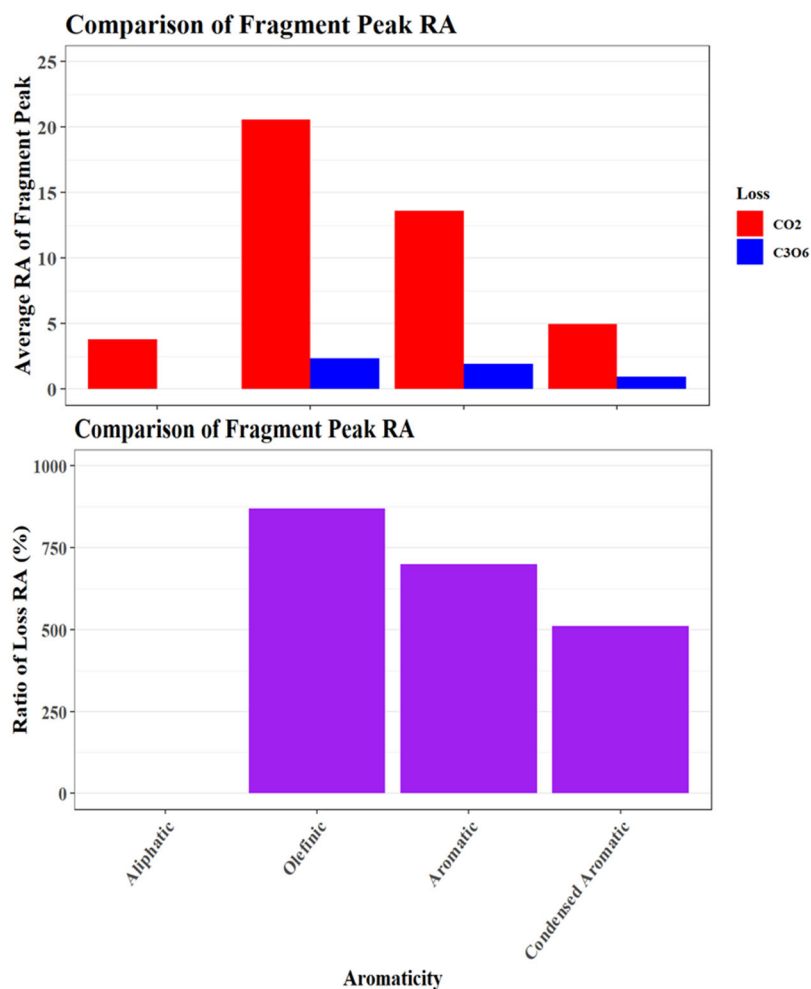
**Figure 4.16.** Comparing average RA trends for molecular species of different aromaticity with $CO_2$ or $C_3O_6$ losses. The top panel shows the average RA of the $CO_2$ (red) or $C_3O_6$ (blue) loss fragment ion for precursor ions within each aromaticity classification. The bottom panel shows the ratio of the average RA shown in the top panel. The ratio is calculated by dividing the RA of the $CO_2$ fragment by the RA of the $C_3O_6$ fragment. The taller the bar, the larger the RA of $CO_2$ fragment ion is relative to the $C_3O_6$ fragment ion.
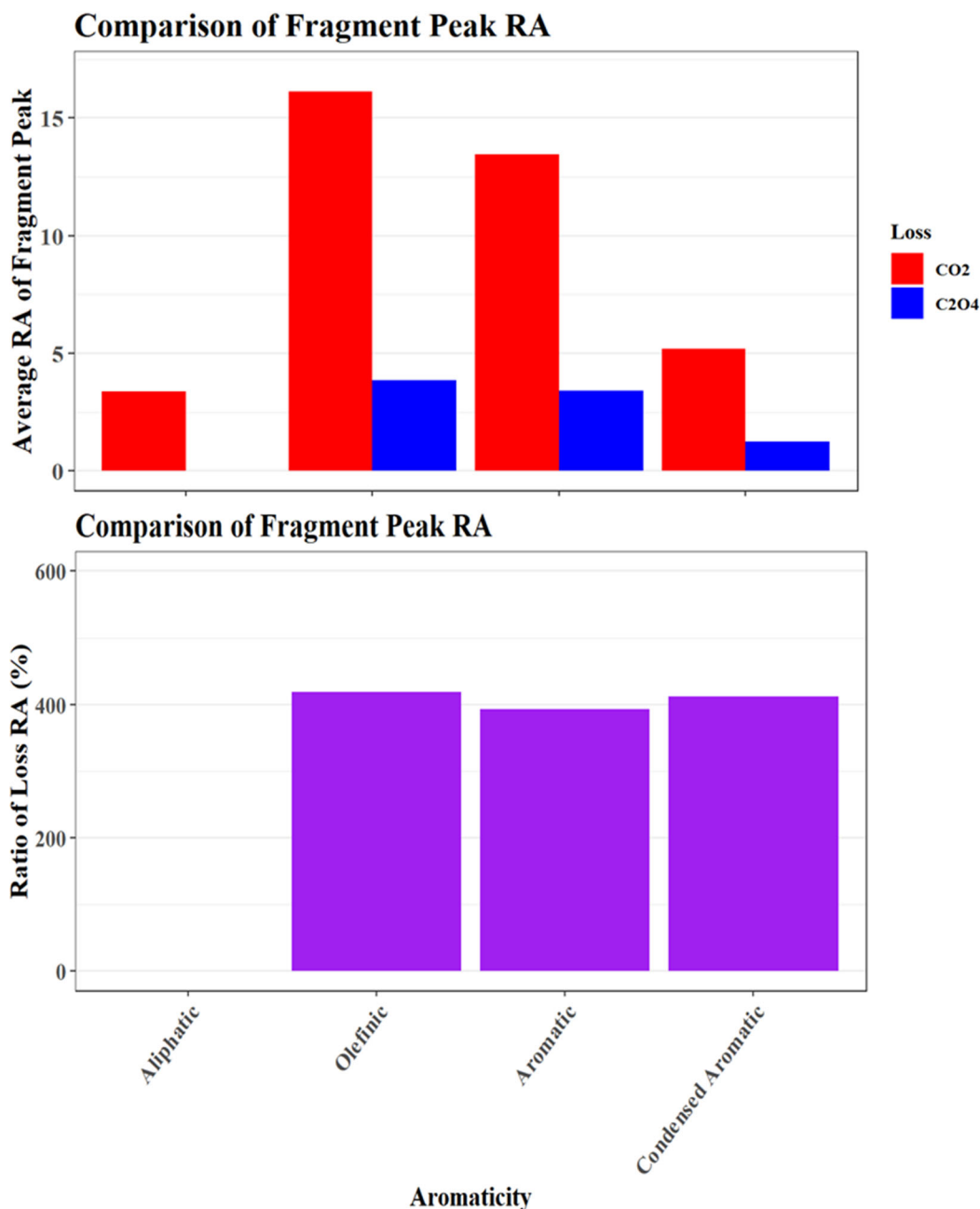
**Figure 4.17**. Comparing average RA trends for molecular species of different aromaticity with $CO_2$ or $C_2O_4$ losses. The top panel shows the average RA of the $CO_2$ (red) or $C_2O_4$ (blue) loss fragment ion for precursor ions within each aromaticity classification. The bottom panel shows the ratio of the average RA shown in the top panel. The ratio is calculated by dividing the RA of the $CO_2$ fragment by the RA of the $C_2O_4$ fragment. The taller the bar, the larger the RA of $CO_2$ fragment ion is relative to the $C_2O_4$ fragment ion.

*4.2.6.2   Van Krevelen Analysis of Species with $CO_2$ and $H_2O$ Loss*

A VK plot for all of the molecular formulas with at least one of the six $CO_2$ or $H_2O$ based losses are shown in Figure 4.18. Since the plots show all of the molecular formulas with each of the losses, many of the molecular formulas show up in several of the plots (e.g. $H_2O$ and $CO_2$). The molecular formulas with the loss are in color (scaled by DBE) and the other molecular formulas with at least one of the other losses are in gray. Here a clear difference between the molecular formulas with multiple $CO_2$ losses and those with multiple $H_2O$ losses is observed. As more $CO_2$ losses are observed, the molecular formulas shift down and to the right, indicating an increased aromaticity and oxidation. In contrast, the $H_2O$ losses shift up and to the right, indicating a similar increase in oxidation, but a decrease in the aromaticity. These observations are consistent with what has been previously described in Section 4.2.4.2.3.
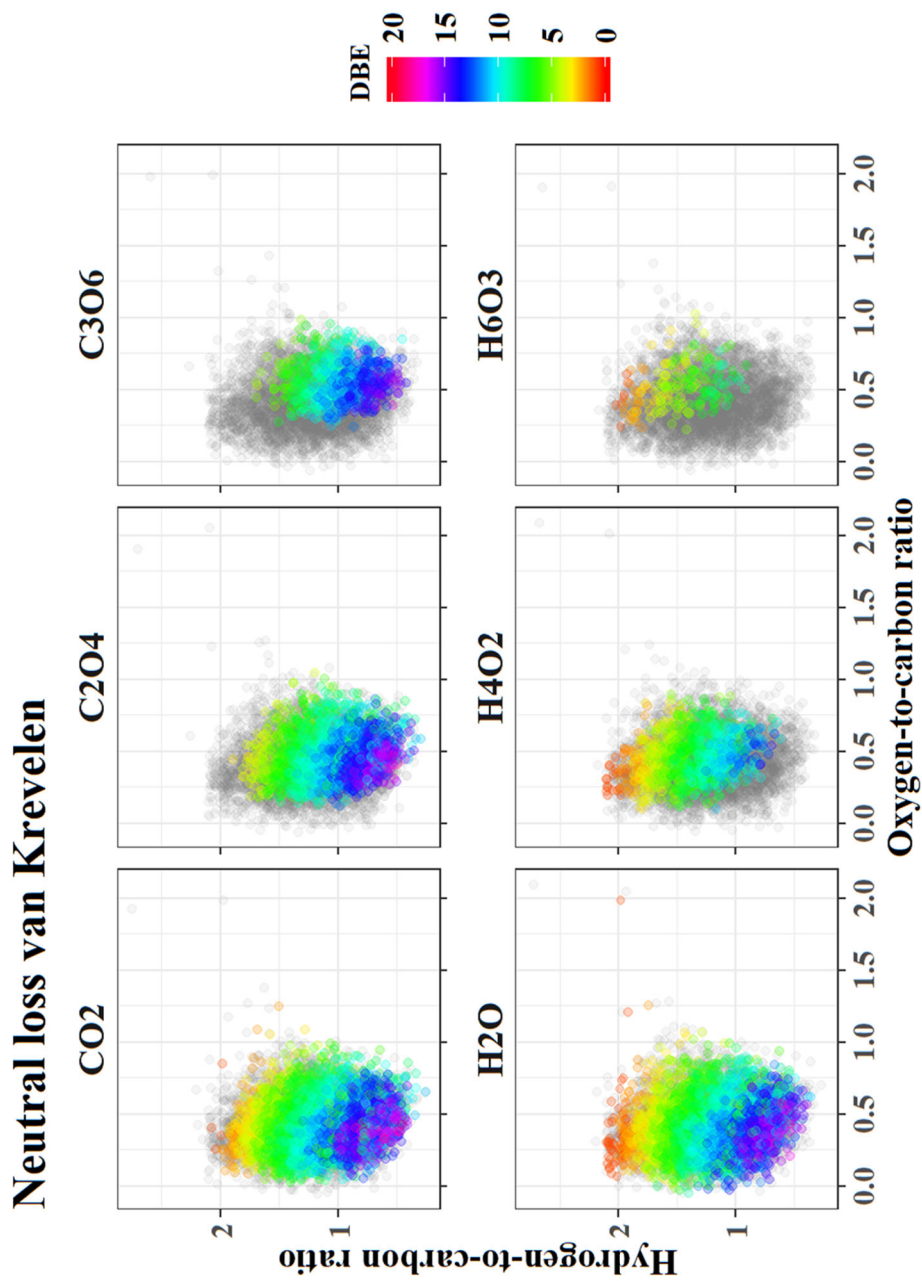
**Figure 4.18.** VK plots showing the molecular formulas that demonstrate each type of neutral loss (color) specified by the plot facet title ($CO_2$, etc.) and all formulas that show at least one of these loss types (gray), but not the one the plot is labeled for. The color is determined by the DBE value for the molecular formula.

### 4.2.7   Nitrogen Containing Neutral Losses

In addition to the C, H, and O based losses, the neutral losses with nitrogen and sulfur were also investigated. Since the majority of the identified molecular formulas in this sample are nitrogen containing, it is expected that many of the observed nitrogen species contain oxygenated functional groups such as nitrate or nitro. Furthermore, oxygenated nitrogen groups are expected because negative ESI ionizes acidic compounds more efficiently (Cech and Enke, 2001). The nitrogen containing neutral losses investigated were: $HNO_2$ (nitro functional group), $HNO_3$ (nitrate functional group), $CH_2NO_4$ (nitro + $H_2O$), $CHNO_5$ (nitro + $CO_2$), and $NH_3$ (amine). Despite the expectation of low $NH_3$, it was the second most common nitrogen-based loss (Figure 4.5), the reason for this is explored in the following paragraphs.

#### 4.2.7.1   PC1 vs. PC2

To interpret the observations PCA was performed on the molecular formulas with one of the five nitrogen containing neutral losses (Figure 4.19). From this plot of PC1 and PC2 it is clear that the molecular formulas with an $NH_3$ losses are not correlated with any of the nitrate losses and are almost negatively correlated with the loss of $HNO_2$. The loss and composition colored plot of PC1 and PC2 (Figure 4.20) does not provide much additional information about the characteristics of the molecular formulas with the loss of $NH_3$ relative to the nitrate losses. No clear trend or difference in the number of oxygen, carbon, or DBE was observed between the molecular formulas with the different losses. Molecular formulas with amine groups should not ionize well in the negative ion mode, so it is expected that some other acidic functional group is also present for ionization.
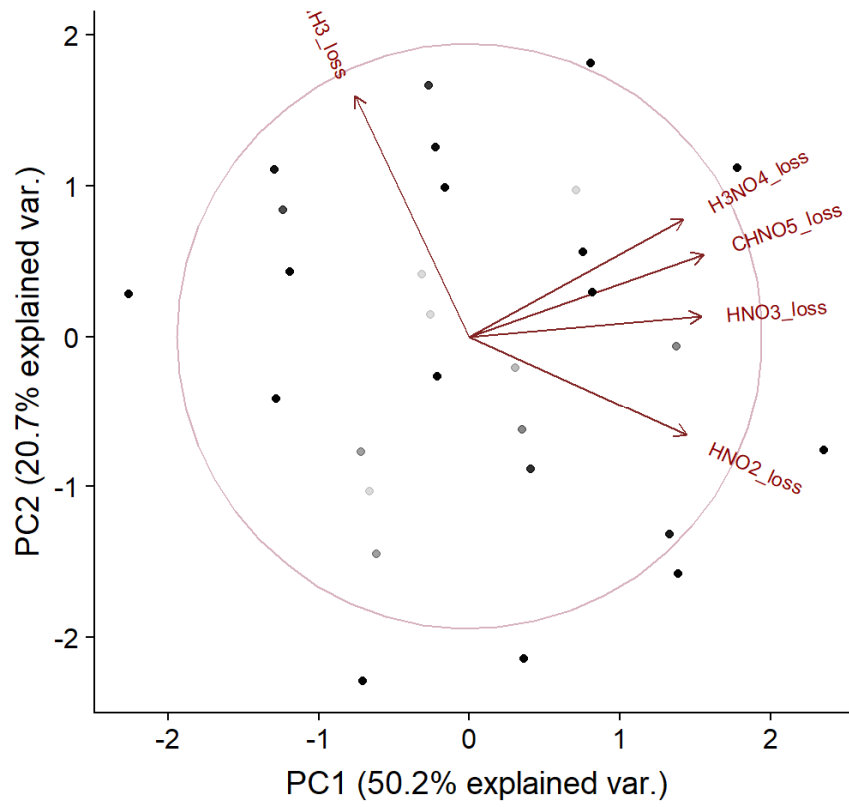
**Figure 4.19.** PC1 vs. PC2 PCA biplot for molecular species with a neutral loss of $HNO_2$, $HNO_3$, $CHNO_5$, $H_3NO_4$, or $NH_3$.

**Figure 4.20**. Colored PC plots PC1 vs. PC2. Panels a-e show the molecular species with the neutral loss (red) and without the neutral loss (gray). For example, Panel a shows the molecular species with a $CO_2$ loss in red, while the species that did not have a $CO_2$ loss, but did have one of the other losses are shown in gray. Panels f-h show the same plots, only colored with oxygen number (g), DBE (h), and carbon number (i). All of these plots use the same data as Figure 4.19, but to avoid the overlapping that occurred in that plot, the data points have been jittered, which reveals the complexity within each point in Figure 4.19.

### 4.2.8  Nitrogen Containing Neutral Losses with C, H, O Neutral Losses

#### 4.2.8.1  PC1 vs. PC2: Amine Investigation

To investigate this, another PCA with the nitrogen losses was performed including the losses described in Section 4.2.4.2. The analysis was restricted to the CHNO group of molecular formulas because if all groups are included, the major separation is due to the heteroatoms and not the neutral losses (Figure 4.21). This is also why CHNOS was not included in this portion of the analysis, because as shown in Figure 4.21, they correlate much more strongly with the CHOS group than with the CHNO group.
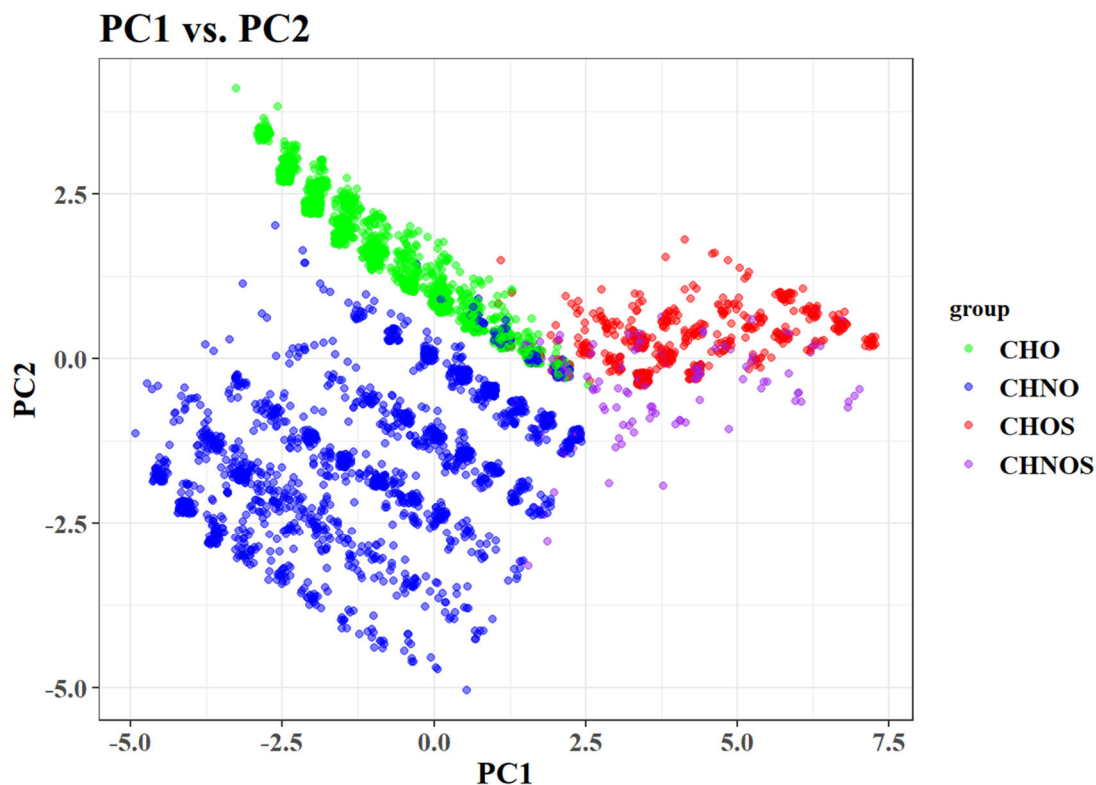


**Figure 4.21**. Jittered PCA plot for PC1 vs. PC2 for molecular species with a neutral loss of $HNO_2$, $HNO_3$, $CHNO_5$, $H_3NO_4$, $NH_3$, $CO_2$, $C_2O_4$, $C_3O_6$, $H_2O$, $H_4O_2$, $H_6O_3$, $CH_2O_3$, $CO$, $C_2H_4O$, $CH_4O$, $C_2H_4O_3$, or $CH_6O_2$. Color is set by elemental group CHO (green), CHNO (blue), CHOS (red), and CHNOS (purple).

153

The PCA biplot for the CHNO nitrogen-containing losses is given in Figure 4.22. The plot indicates that although the loss of $NH_3$ is not tightly correlated with other nitrogen losses, it is somewhat correlated with $H_2O$, $CO_2$, $CO$, and $C_2H_4O$ losses. This suggests that if a molecular formula has amine group, it likely also has a carboxyl, hydroxyl, or carbonyl group if it is detected as a negative ion. This likely counteracts the basicity of the amine group and allows the molecule to be ionized.
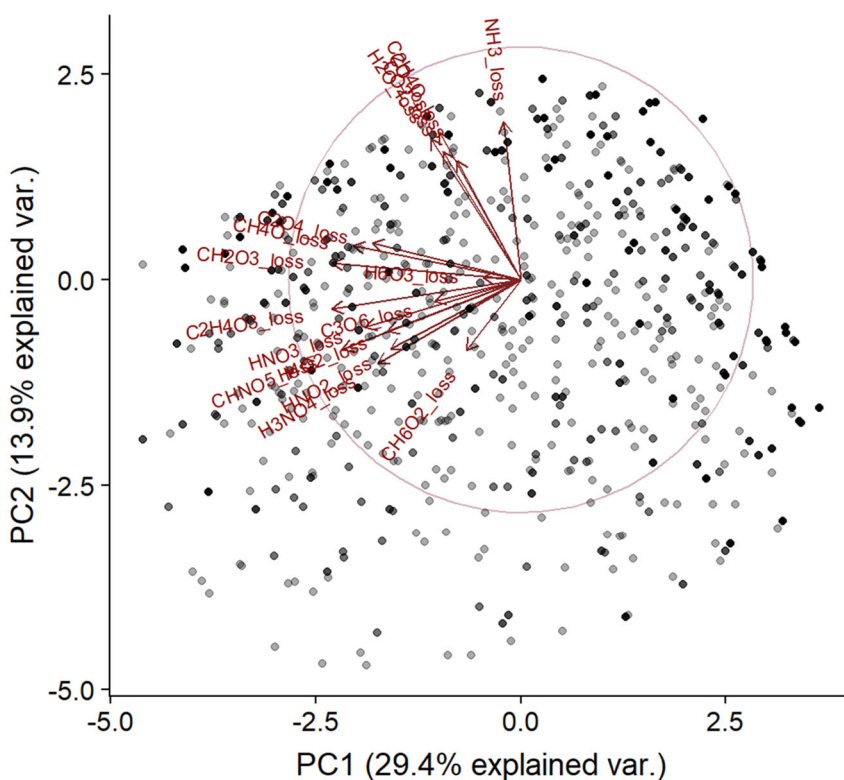


**Figure 4.22.** PC1 vs. PC2 PCA biplot for only CHNO molecular species with a neutral loss of $HNO_2$, $HNO_3$, $CHNO_5$, $H_3NO_4$, $NH_3$, $CO_2$, $C_2O_4$, $C_3O_6$, $H_2O$, $H_4O_2$, $H_6O_3$, $CH_2O_3$, $CO$, $C_2H_4O$, $CH_4O$, $C_2H_4O_3$, or $CH_6O_2$.

Of the 2136 molecular formulas that show a loss of $NH_3$, 1948 of them have at least one additional loss, with $CO_2$, $H_2O$, CO, and $C_2H_4O$ making up 1805 of these. There are 187 molecular formulas with only a loss of $NH_3$, but they are all low intensity peaks, suggesting that the fragments for the other losses were likely below the detection limit. The favorability of fragmentation at a primary amine (Gross, 2017) may lead to the $NH_3$ fragment ion being above the detection limit. The common combination of $NH_3$ and $CO_2$ losses suggests the presence of amino acids in the organic aerosol which is consistent with previous studies (Ge et al., 2011) including aerosol from biomass burning (Mace et al., 2003; Ge et al., 2011). This PCA does an improved job of showing the types of molecular formulas with the loss of $NH_3$ relative to those with an oxygenated nitrogen loss (Figure 4.23). The molecular formulas with an $NH_3$ loss are shifted toward the upper edge of the plot area, where the $H_2O$, $CO_2$, CO, and $C_2H_4O$ losses are predominately observed. This provides more support for the correlations indicated in Figure 4.22. This upper edge tends to have lower oxygen and higher DBE than the regions below it, which suggests the molecular formulas that are characteristic of $NH_3$ losses are more aromatic and less oxidized than molecular formulas with the loss of a nitro or nitrate group.
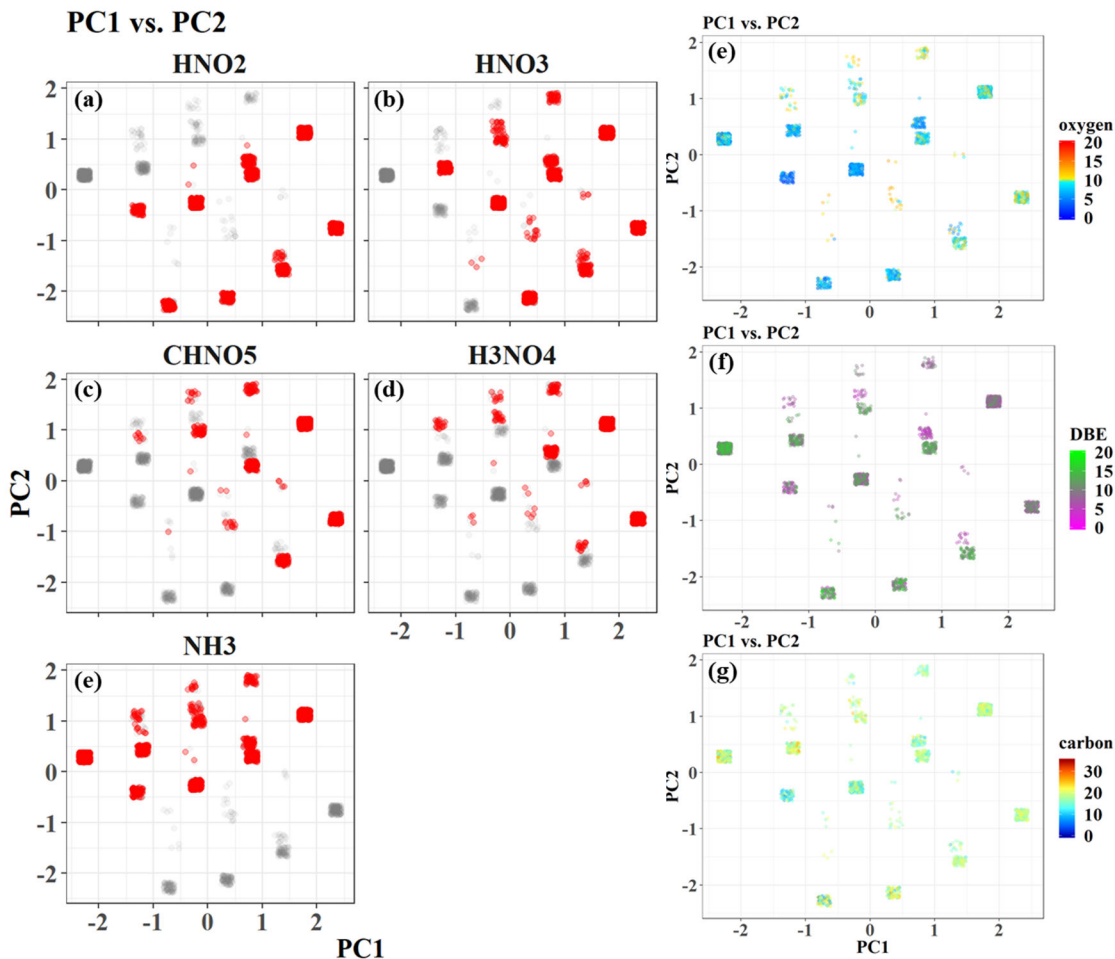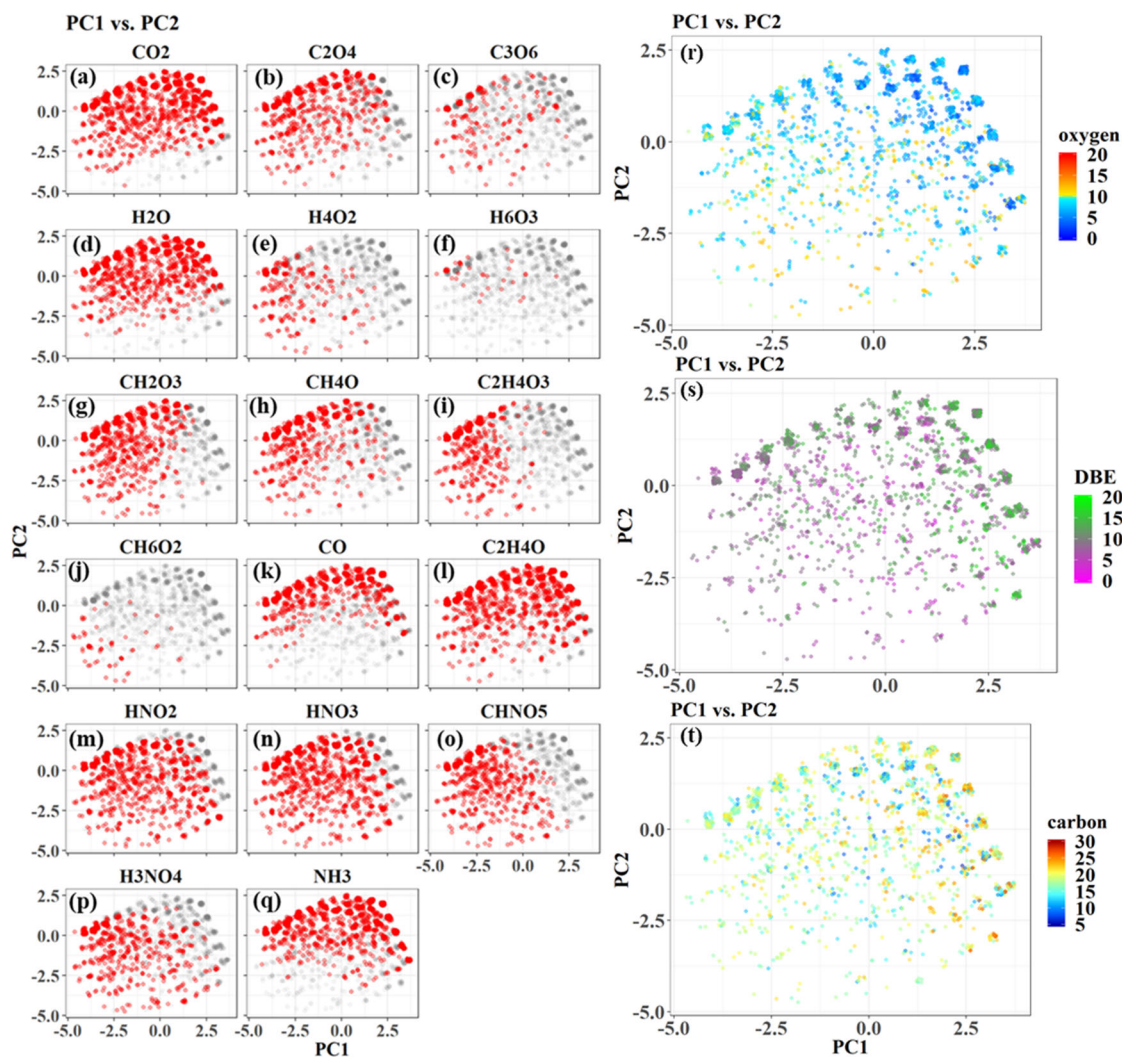
**Figure 4.23**. Colored PC plots PC1 vs. PC2. Panels a-q show the molecular species with the neutral loss (red) and without the neutral loss (gray). For example, Panel a shows the molecular species with a $CO_2$ loss in red, while the species that did not have a $CO_2$ loss, but did have one of the other losses are shown in gray. Panels r-t show the same plots, only colored with oxygen number (r), DBE (s), and carbon number (t). All of these plots use the same data as Figure 4.22, but to avoid the overlapping that occurred in that plot, the data points have been jittered, which reveals the complexity within each point in Figure 4.22.

*4.2.8.2  PC1 vs. PC2 Oxygenated Nitrogen Losses*

The oxygenated nitrogen losses tended to shift towards the bottom left of the plot which generally contains molecular formulas with higher oxygen and lower DBE. When the RA of the $NH_3$ fragment ions is compared to the RA of the $HNO_2$ fragment ions (Figure 4.24), the $NH_3$ losses are more abundant in formulas that are more aliphatic. This suggests that for more aliphatic molecules the presence of an amine group is preferable to a nitro group, but as the aromaticity increases, the presence of a nitro group becomes more favorable. Nitro groups are known to be present in aromatic aerosol produced in the presence of $NO_x$ so this relationship is not unexpected (Grosjean, 1992). Furthermore, studies of the bond between aromatic ring and the nitro group have suggested it is relatively weak, making it a likely fragmentation point (Rice et al., 2002).
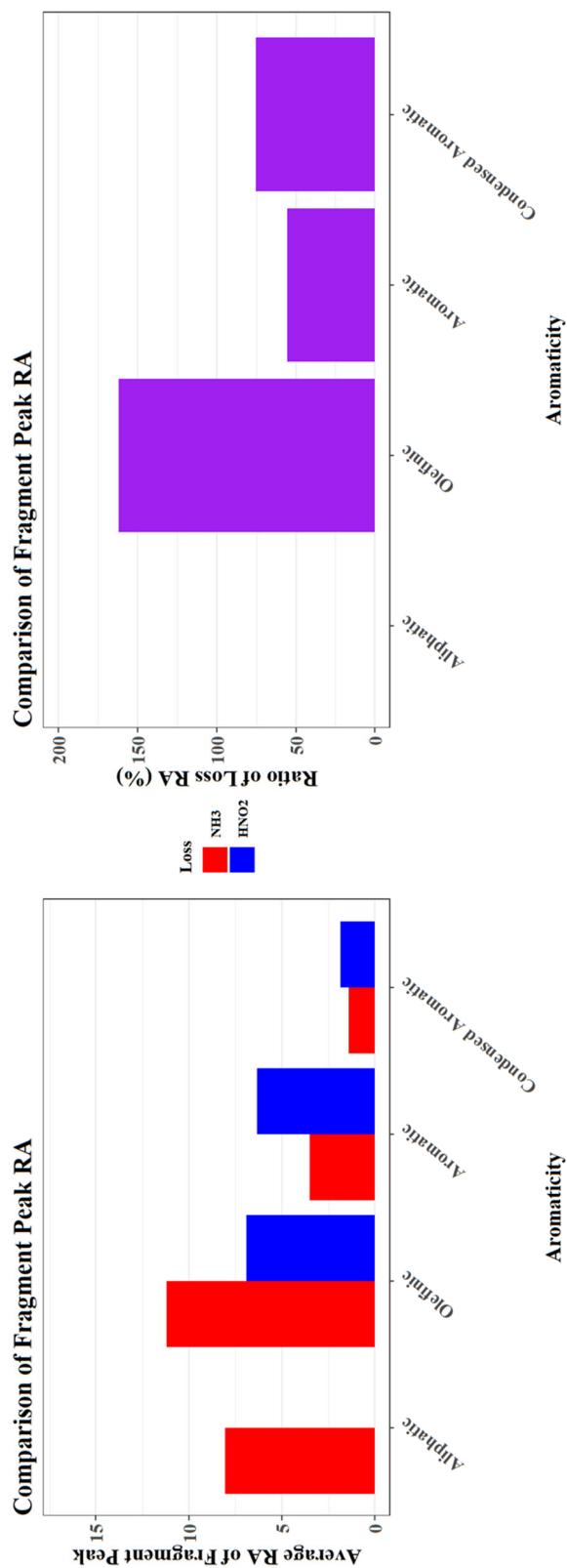
**Figure 4.24.** Comparing average RA trends for molecular species of different aromaticity with $NH_3$ or $HNO_2$ losses. The top panel shows the average RA of the $NH_3$ (red) or $HNO_2$ (blue) loss fragment ion for precursor ions within each aromaticity classification. The bottom panel shows the ratio of the average RA shown in the top panel. The ratio is calculated by dividing the RA of the $NH_3$ fragment by the RA of the $HNO_2$ fragment. The taller the bar, the larger the RA of $NH_3$ fragment ion is relative to the $HNO_2$ fragment ion.

The PCA plots in Figure 4.23 do not show much of a difference between the molecular formulas with an $HNO_2$ or $HNO_3$ loss. However, the RA trends plot (Figure 4.25) indicates that the loss of $HNO_2$ becomes more important as the aromaticity of the base molecule increases. The results suggest that either a nitro group is more common in more aromatic molecular formulas, or that a nitro group is preferentially cleaved relative to a nitrate group in more aromatic molecular formulas. The first option is more likely because nitroaromatic molecules are well known components of biomass burning (Iinuma et al., 2010; Kahnt et al., 2013). Additionally, nitrate groups are expected to be present in aliphatic compounds, while nitro groups are present in aromatic groups.
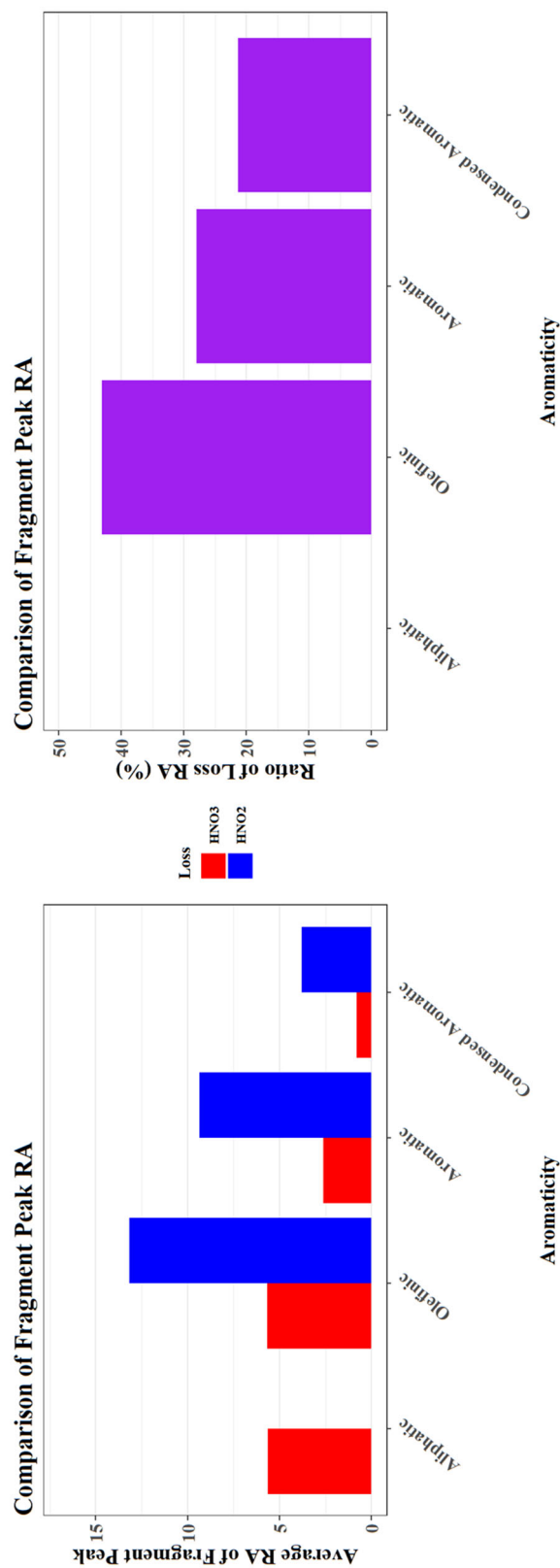
**Figure 4.25.** Comparing average RA trends for molecular species of different aromaticity with $HNO_3$ or $HNO_2$ losses. The top panel shows the average RA of the $HNO_3$ (red) or $HNO_2$ (blue) loss fragment ion for precursor ions within each aromaticity classification. The bottom panel shows the ratio of the average RA shown in the top panel. The ratio is calculated by dividing the RA of the $HNO_3$ fragment by the RA of the $HNO_2$ fragment. The taller the bar, the larger the RA of $HNO_3$ fragment ion is relative to the $HNO_2$ fragment ion.

The prevalence of the nitro groups supports a biomass burning influenced aerosol

(Iinuma et al. 2010; Kahnt et al., 2013). The losses of $H_3NO_4$ ($HNO_3 + H_2O$) and $CHNO_5$

($HNO_3 + CO_2$) highlight the previously observed difference between molecular formulas

with a $CO_2$ loss and an $H_2O$ loss. The VK plot with these two types of losses (Figure

4.26) demonstrates that molecular formulas showing the loss of $CHNO_5$ are distributed in

the high O/C and lower H/C region of the plot, whereas the $H_3NO_4$ losses are

concentrated in the high O/C and higher H/C region. The difference between these

regions can be characterized as more aromatic for the $CHNO_5$ losses and less aromatic

for the $H_3NO_4$ losses, which is consistent with what was previously observed from

comparing the multi-$H_2O$ losses to multi-$CO_2$ losses. This suggests that muli-functional

compounds showing concurrent functional group losses are more aromatic when $CO_2$ is

involved and more aliphatic when $H_2O$ is.
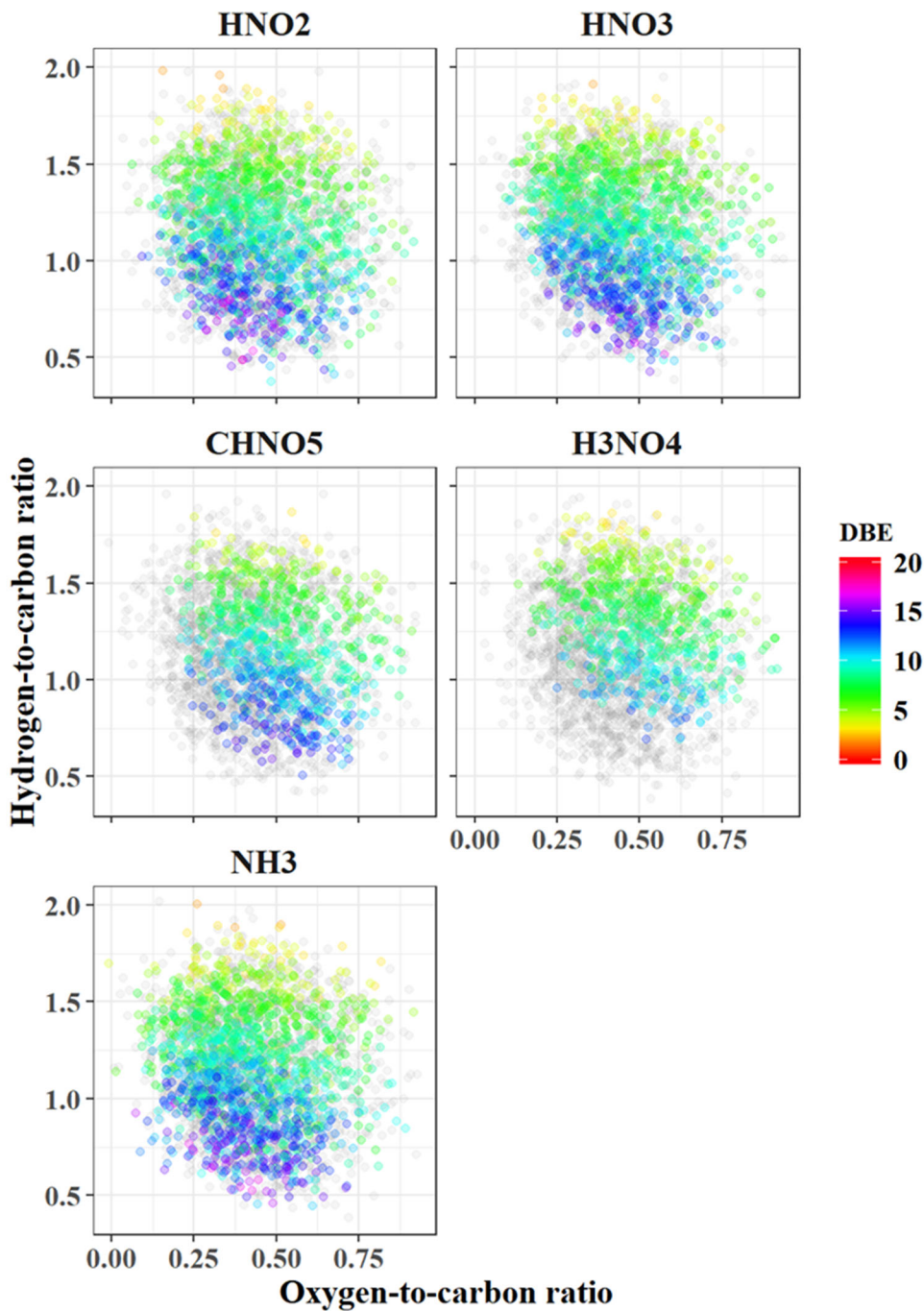
**Figure 4.26.** VK plots showing the molecular formulas that demonstrate each type of neutral loss (color) specified by the plot facet title (HNO2, etc.) and all formulas that show at least one of these loss types (gray), but not the one the plot is labeled for. The color is determined by the DBE value for the molecular formula.

### 4.2.8.3 Interpretation of Nitro Groups

Overall, the nitro functional groups are the most abundant type of nitrogen containing loss. This is consistent with the composition of biomass burning aerosol (Iinuma et al., 2010; Kahnt et al., 2013). Many of the nitro containing molecular formulas are aromatic, indicating that they are likely light absorbing (Desyaterik et al., 2013; Lin et al., 2015). Some of the possible structures of the selected set of molecular formulas with nitro losses are shown in Figure 4.27. The molecular formulas are $C_7H_7NO_4$, $C_6H_5NO_5$, and $C_7H_6N_2O_4$. The predicted structures are supported by the secondary losses observed in addition to the nitro loss. For example, $C_7H_7NO_4$ has a methoxy group in its proposed structure and in addition a methoxy loss is observed for that molecular formula. In contrast, the proposed structure of $C_6H_5NO_5$ does not contain a methoxy group, and no methoxy group is observed, supporting this hypothesized structure. Meanwhile, the proposed structure for $C_7H_6N_2O_4$ does not include any OH groups, which is consistent with the lack of an observed water loss for this molecular formula. All of these structures contain aromatic rings, making it likely that they are absorbing species (Desyaterik et al., 2013; Lin et al., 2015). It is possible that the nitro groups are also contributing to the absorbance of these species as well. Recent studies have also suggested that charge transfer may be responsible for some of the absorbance observed in brown carbon aerosol (Philips and Smith, 2014; 2015), due to the ability of aldehyde, ketone, alcohol, and carboxylic acid groups to undergo charge transfer interactions, increasing the absorbance at higher wavelengths in particular. While those studies do not mention nitro groups, a study by Nagakura (1955) demonstrated intramural charge transfer in molecules with

163

nitro groups, and suggested that it may be possible for intermural molecular interactions as well. Other studies have highlighted that nitro-phenols in particular are important for the observed absorbance in brown carbon aerosol (Desyaterik et al., 2013; Lin et al., 2015), but the impact of the nitro group itself was not discussed. While this idea is speculative at this point, an increased absorbance due to the nitro groups in organic aerosol components could be an important factor to consider when evaluating the absorption potential of brown carbon. The results from this MS/MS work indicate a high frequency of nitro groups which may contribute to an enhanced absorption which warrants further investigation, including theoretical studies focused on their specific impact.
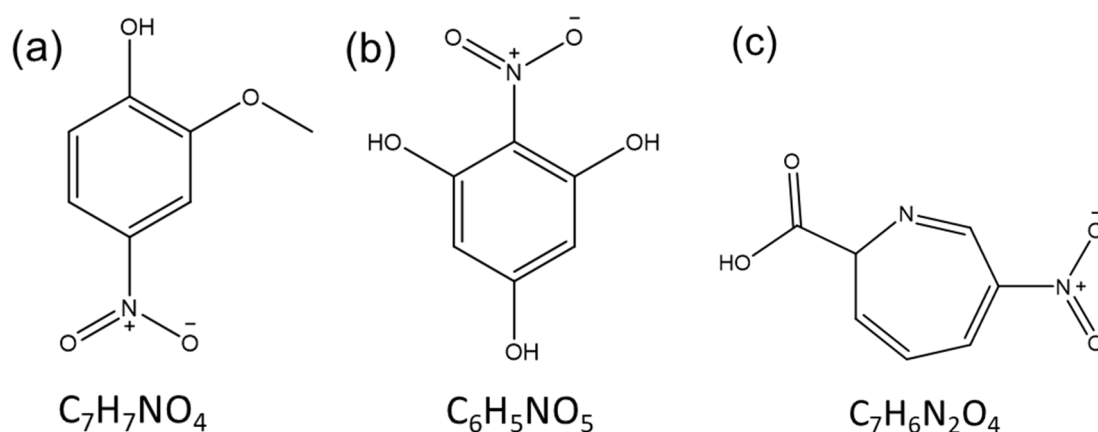


**Figure 4.27.** Proposed structures for some nitro aromatic formulas observed in PMO-1. Panel a shows $C_7H_7NO_4$, Panel b shows $C_6H_5NO_5$, and Panel c shows $C_7H_6N_2O_4$

### 4.2.9  Sulfur Containing Neutral Losses

Neutral losses of $SO_3$, $SO_4$, $H_2SO_4$, $CSO_5$, and $SH_2$ were also examined. The loss of $SO_3$ was the most commonly observed, followed by $SO_4$ and $SH_2$. Similar to the $NH_3$ loss, the observation of a reduced sulfur functional group is somewhat surprising. When the molecular formulas with sulfur losses were compared using PCA, the sulfate neutral losses were all somewhat correlated, but the $SH_2$ loss was not in the PC1 and PC2 biplot (Figure 4.28). The PCA plots provide some insight into the general characteristics of the specific molecular formulas with the loss of $SH_2$ relative to the ones with sulfate losses. Specifically, the $SH_2$ loses are all located in the lower half of the PCA plots, which is a region that contains relatively low numbers of oxygen and an increased DBE value (Figure 4.29), suggesting that thiol groups are more likely to be present on less oxygenated and aromatic molecules.
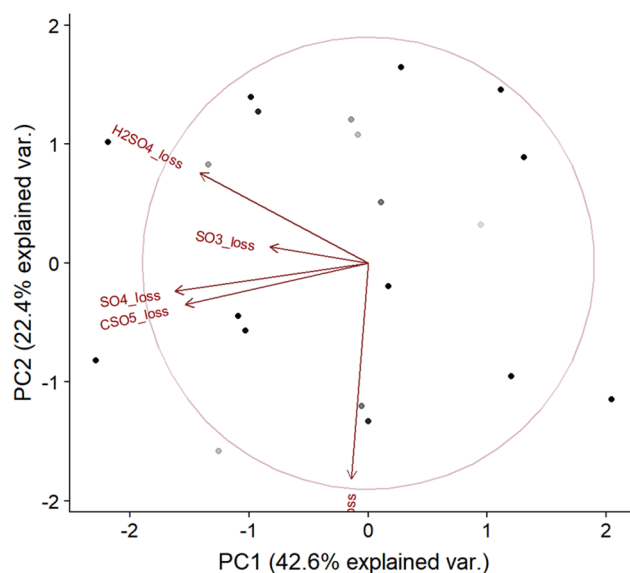


**Figure 4.28.** PC1 vs. PC2 PCA biplot for molecular species with a neutral loss of $SO_3$, $SO_4$, $H_2SO_4$, $CSO_5$, or $SH_2$.
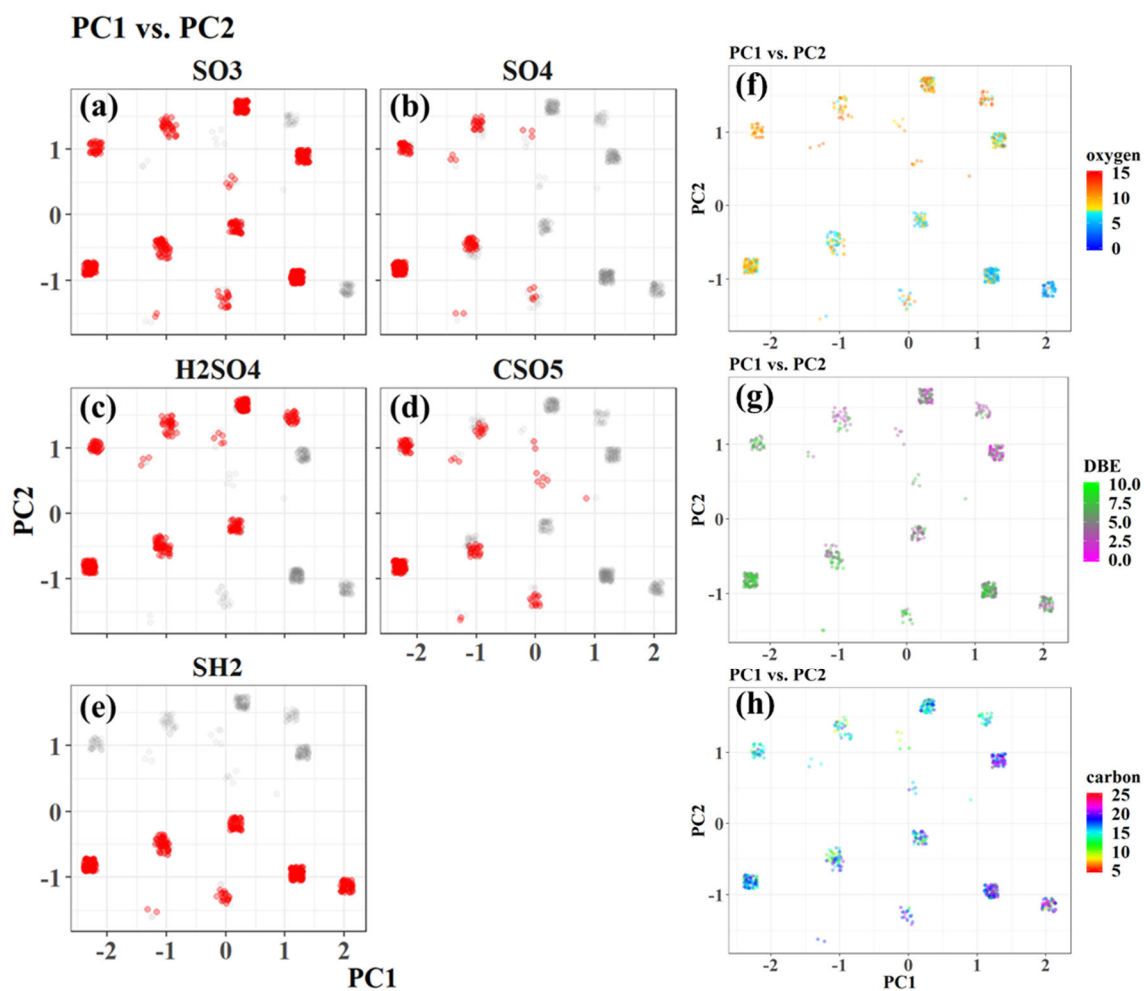
**Figure 4.29.** Colored PC plots PC1 vs. PC2. Panels a-e show the molecular species with the neutral loss (red) and without the neutral loss (gray). For example, Panel a shows the molecular species with a $CO_2$ loss in red, while the species that did not have a $CO_2$ loss, but did have one of the other losses are shown in gray. Panels f-h show the same plots, only colored with oxygen number (r), DBE (s), and carbon number (t). All of these plots use the same data as Figure 4.28, but to avoid the overlapping that occurred in that plot, the data points have been jittered, which reveals the complexity within each point in Figure 4.28.

### 4.2.10 Sulfur Containing Neutral Losses with C, H, O Neutral Losses

However, there is still the question of how the $SH_2$ containing molecular formulas were ionized, so the sulfur-based losses were added to the C, H, and O based losses and PCA was done again for the CHOS and CHNOS molecular formulas (Figure 4.30). In this case the functional group correlations were less clear, but the $SH_2$ loss was most correlated with $C_2H_4O$, CO, $CO_2$, and $SO_3$ losses. Only one sulfur atom was allowed when the molecular formula assignment was done, so any overlap between an $SO_3$ and $SH_2$ loss would suggest isomerization. The correlation with $C_2H_4O$, CO, and $CO_2$ suggests that $SH_2$ is related to carbonyl and carboxyl groups to some extent, which may explain how they are ionized. The relative low number of detected sulfur-containing species and subsequently, their fragments is likely due to the low abundance of these molecular formulas relative to the CHO and CHNO molecular formulas obscures the conclusions.
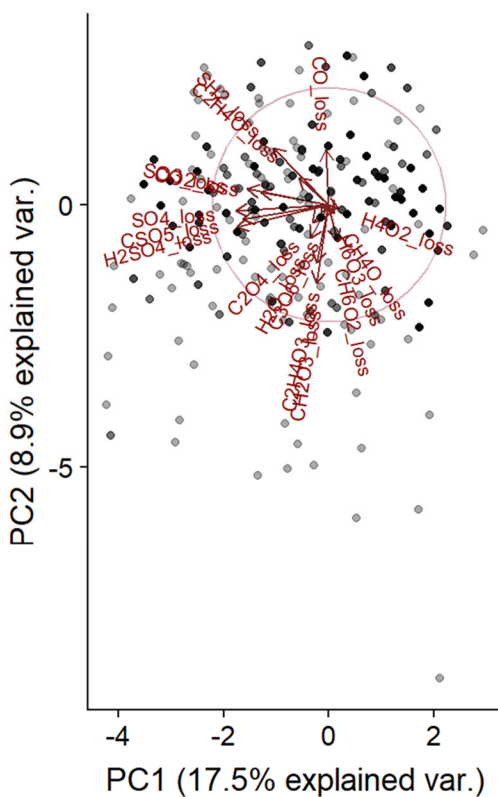
**Figure 4.30**. PC1 vs. PC2 PCA biplot for only CHOS and CHNOS molecular species with a neutral loss of $SO_3$, $SO_4$, $H_2SO_4$, $CSO_5$, $SH_2$, $CO_2$, $C_2O_4$, $C_3O_6$, $H_2O$, $H_4O_2$, $H_6O_3$, $CH_2O_3$, $CO$, $C_2H_4O$, $CH_4O$, $C_2H_4O_3$, or $CH_6O_2$.

**4.2.11 Van Krevelen Analysis of Species with Sulfur Containing Neutral Losses**

The VK plot of the sulfur losses (Figure 4.31) provides some additional information about the general composition of the molecular formulas with the sulfur containing losses. $SO_3$ is common across the range of sulfur species, but $SO_4$ is shifted down and to the right, as is $CSO_5$ ($SO_3 + CO_2$). This suggests a preference for more oxygenated and aromatic molecules. $H_2SO_4$ ($SO_3 + H_2O$) shifts toward the upper right, suggesting a similar dependence on oxygenation, but also aliphatic species. In contrast, the molecular formulas with $SH_2$ loss are shifted slightly to the left, indicating that it is more prevalent on less oxygenated species, consistent with previous observations. The comparison of the RA for the fragment ions for the $SH_2$ and $SO_3$ losses suggests that $SH_2$ is more favored as the aromaticity increases. Since the molecular formulas are expected to be isomers of each other due to the sulfur limitation, this may suggest that the reduced thiol group becomes more common for more aromatic compounds. This could be a valuable piece of information for predictions involving sulfur containing species in aerosol because the hygroscopicity of the thiol group will be different than a sulfate group.
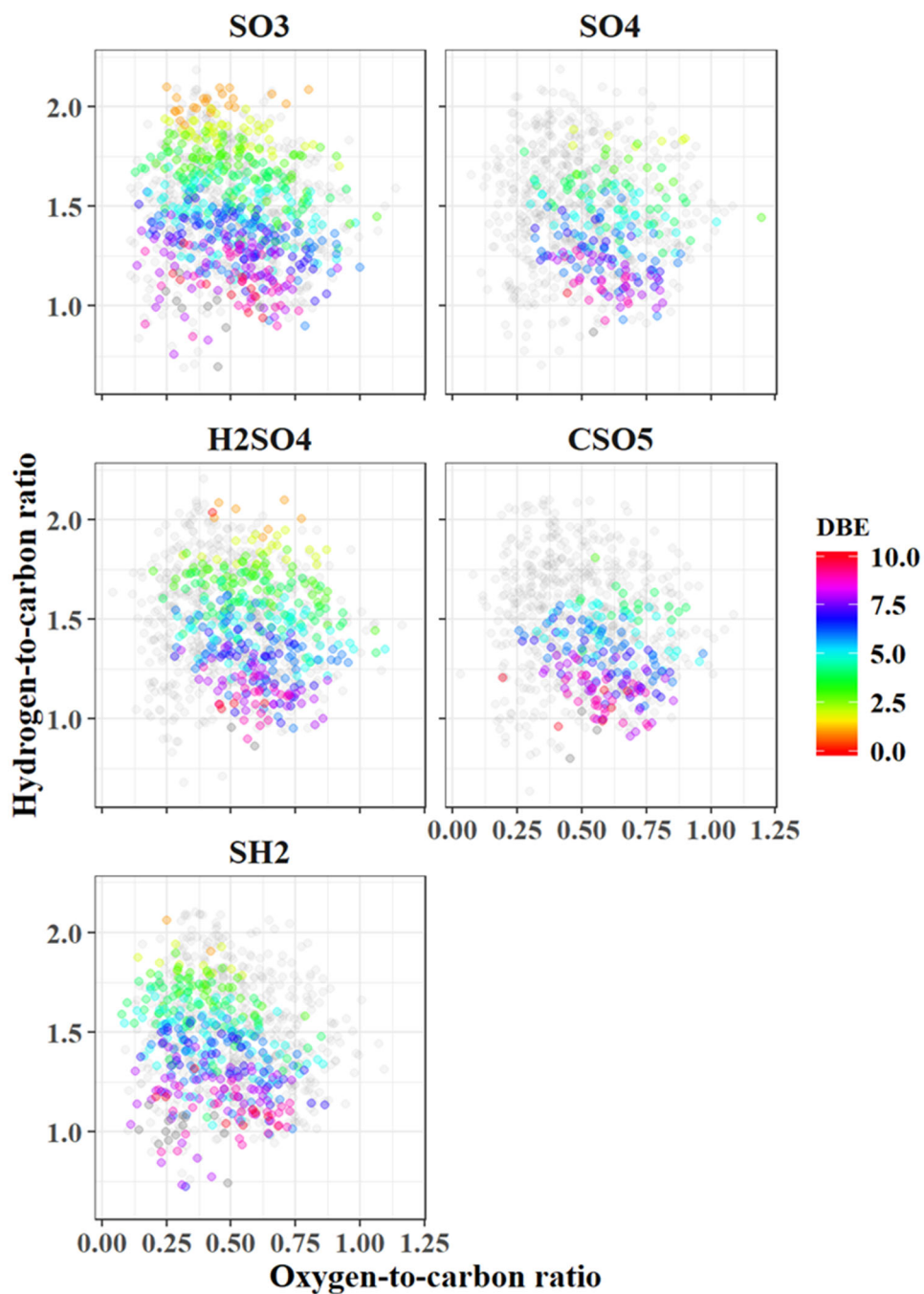
**Figure 4.31**. VK plots showing the molecular formulas that demonstrate each type of neutral loss (color) specified by the plot facet title (SO₃, etc.) and all formulas that show at least one of these loss types (gray), but not the one the plot is labeled for. The color is determined by the DBE value for the molecular formula.

### 4.2.12  Proposed Structures for Selected Molecular Formulas

One of the major goals of MS/MS analysis is to determine the molecular structures for the species of interest. In this case, it is difficult to determine any unequivocal molecular structures because each mass is representative of multiple isomers. However, some potential structures can be suggested for molecules that have specific loss types, such as $H_6O_3$ and $C_3O_6$, because they require the loss of three individual $H_2O$ or $CO_2$, making their structures somewhat easier to predict because more of the functional groups are constrained. To demonstrate some potential molecular structures and their general fragmentation pattern we have provided potential structures of two molecular formulas that show $C_3O_6$ loss are provided (Figure 4.32-4.33). $C_{13}H_{16}O_7$, is one of the most abundant molecular formulas with 3 $CO_2$ losses in it and it has a DBE value of 6. Due to the 3 DBE for the 3 carboxylic acids, an aromatic ring is unlikely because it would require 4 DBE. The hypothetical structure has 3 carboxylic acids that can be lost as shown in the series of structures (Figure 4.32). Another molecular formula with the 3 $CO_2$ losses is $C_{23}H_{14}O_{11}$, with 17 DBE. This example has one of the highest DBE values. A proposed structure for this compound is presented in Figure 4.33. This structure has four conjugated benzene rings making it a condensed aromatic molecule. The conjugated rings are similar to what would be expected for soot, but the oxygenated functional groups allow the molecule to be water soluble and provide the polarity necessary to be ionized by negative mode electrospray ionization. This type of structure is consistent with what was theorized by Decesari et al. (2002) for the oxidation of soot via atmospheric processes until it was water soluble. This type of structure would contribute to brown

carbon (BrC) due to its probable light absorption. Since BrC is expected to be a major component of this sample due to its wildfire influence the presence of such a molecular structure seems reasonable. The aldehydes and carboxylic acids could also potentially contribute to charge transfer based absorbance as was described by Phillips and Smith (2014; 2015).
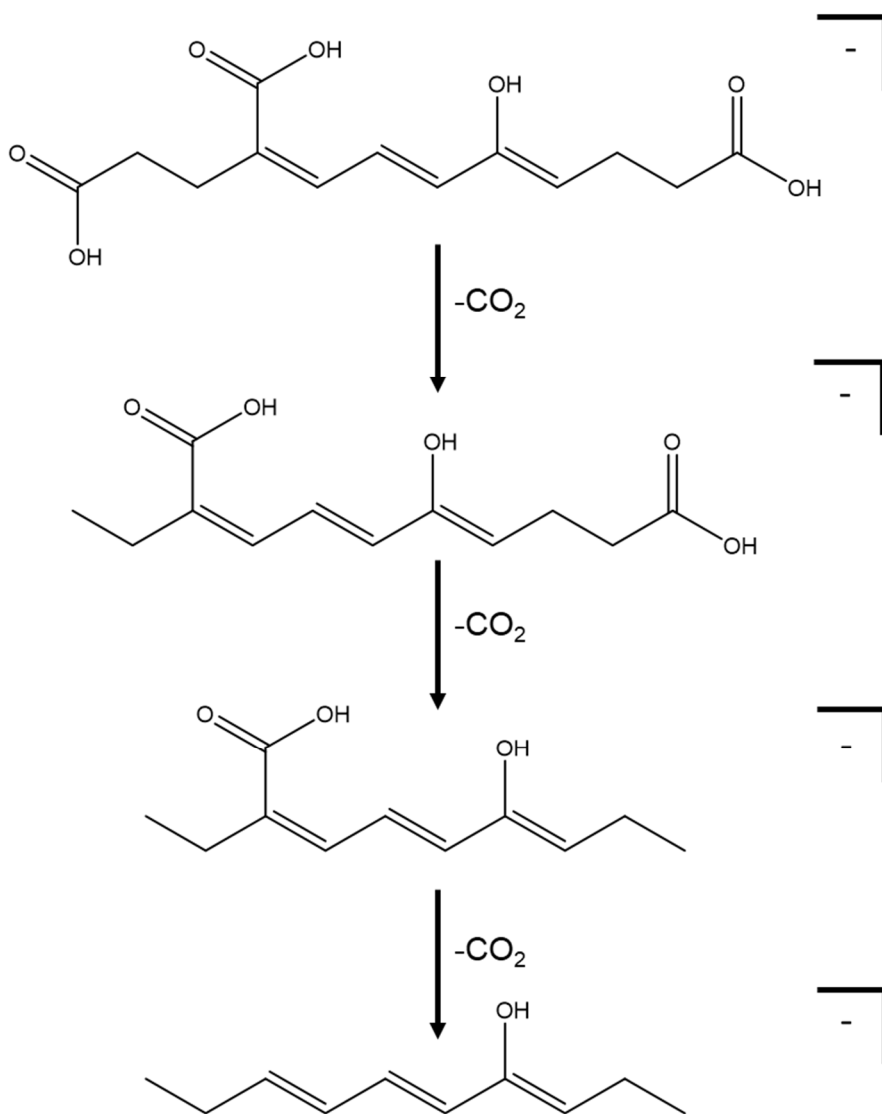


**Figure 4.32**. Proposed structure and fragmentation of $C_{13}H_{16}O_7$
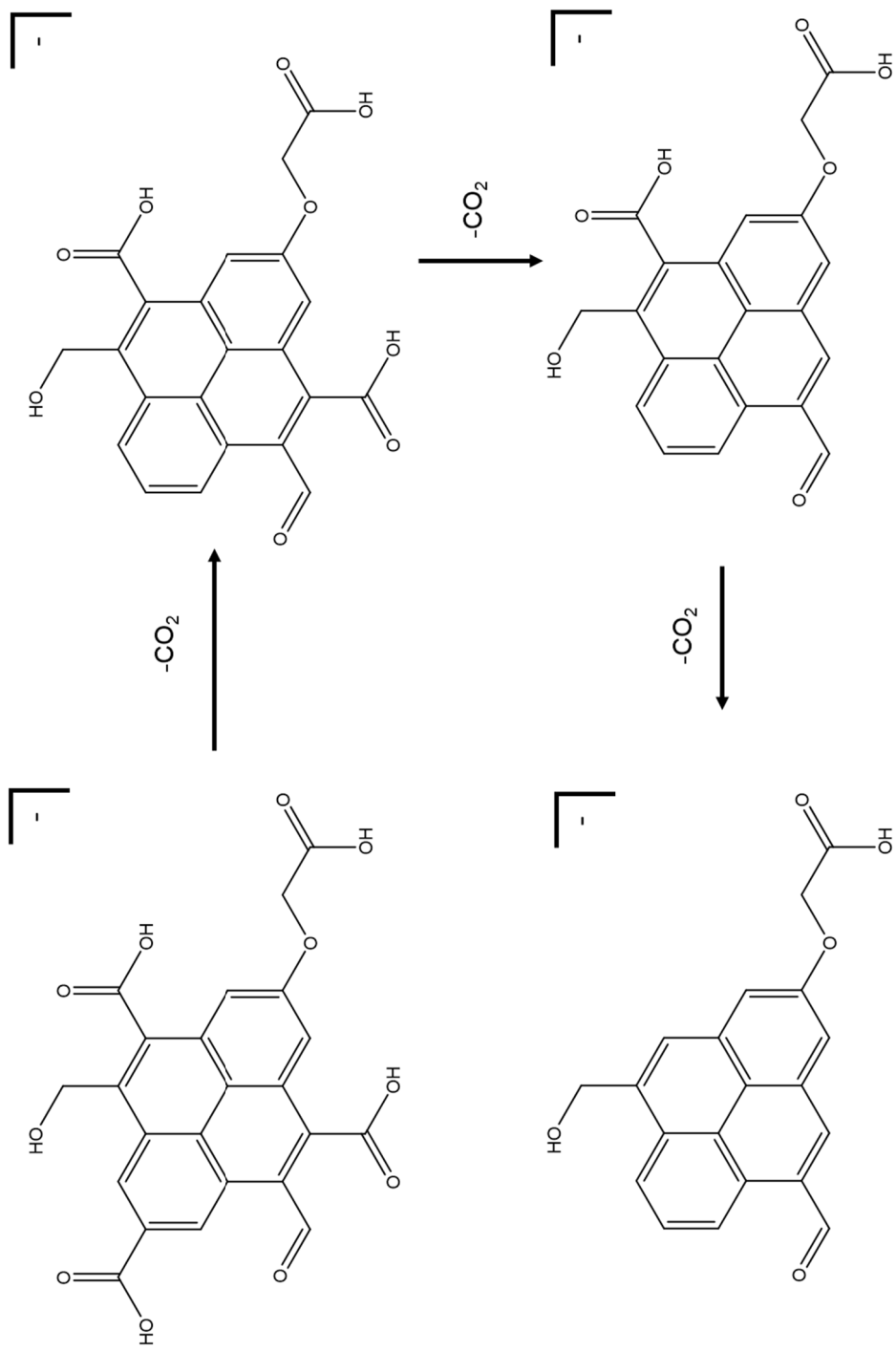
**Figure 4.33**. Proposed structure and fragmentation of $C_{23}H_{14}O_{11}$

173

# 5  MFAssignR

## 5.1  Background

The ultimate goal ultrahigh resolution mass spectrometry is to obtain exact mass measurements for accurate molecular formula assignment. This type of identification is very powerful for an improved understanding of the composition of natural organic matter. However, there are several factors that can make it difficult to obtain an accurate molecular formula assignment. One of the biggest factors is that multiple molecular formulas can be assigned to the same measured mass. This necessitates the use of quality assurance parameters that can choose the correct molecular formula out of many potential options. There are two main ways that molecular formulas can be assigned, database matching or calculation. Database matching is when a measured mass it matched against a database containing many molecular formulas, and whichever ones match the mass within an error tolerance are accepted pending additional QA. This method is generally fast, but is limited to only the molecular formulas in the database and no database can realistically be completely comprehensive. The calculation methods are based on using the exact masses of atoms to calculate a molecular formula that has a theoretical mass within the error tolerance for the measured mass. This brute force method is more flexible than database matching, but can be slower to run. There are several methods available to do molecular formula assignment, some of which are open source (UltraMassExplorer, Formularity) and others that are commercial (Composer, PetroOrg). Some of the programs assign molecular formulas using database searches (UME, Formularity), while others calculate the molecular formulas and use formula extension to

assign other molecular formulas (Composer). In general, most of the methods have a lack

of transparency as to how they assign molecular formulas, and what parameters are taken

into consideration when determining whether the molecular formulas are correct. This

uncertainty in how molecular formula assignment methods handle ambiguous

assignments, led us to develop our own method with more transparency as to how this is

handled. MFAssignR, written in the R programming language can be used on any

computer. This ensured easy access to the code because R is free to download through the

R website (www.r-project.org). The preeminent function in MFAssignR is the function

for molecular formula assignment, MFAssign(). The core of this function is the CHOFIT

algorithm which was developed by Green and Perdue (2015) in the Pascal programming

language. We adapted the CHOFIT algorithm to do molecular formula assignment and

added many additional parameters to improve its ability to assign a variety of

heteroatoms and quality assurance for the molecular formulas that get assigned. While

developing the MFAssign functions to assign molecular formulas, we decided to expand

the R package so that it also included noise estimation, isotope filtering, and

recalibration. These are all very important components of getting the best data possible

from ultrahigh resolution mass spectrometry. This means the final package contains a

complete pipeline for the assignment and analysis of ultrahigh resolution mass

spectrometry data, which is something that until recently was only available in

commercial software (Leefmann et al., 2018). As such, this software package represents

an important contribution to studies of natural organic matter (NOM) such as aerosol,

soil, and aquatic organic species using ultrahigh resolution mass spectrometry. The

following chapter describes each of the functions and includes some background

information regarding other methods that have been developed to do the same thing, or that our functions were based on.

## 5.2 Molecular Formula Assignment

### 5.2.1 Methods of Formula Assignment

MFAssignR has two functions (MFAssign() and MFAssignCHO()) for molecular formula assignment which are used for different aspects of formula assignment. MFAssign() is a multi-element function that includes non-oxygen heteroatoms and provides low ambiguity in the assignments. MFAssignCHO() is identical to MFAssign(), except that it only can assign molecular formulas with C, H, and O. Thus, it run faster, and is more useful for preliminary molecular formula assignment used to determine possible recalibrant ions. At the core of each function is the CHOFIT algorithm developed by Perdue and Green (2015). CHOFIT makes use of low mass moieties to assign C, H, and O containing molecular formulas and is faster relative to the traditional methods of using brute force looping and Diophantine equations to assign all of the molecular formulas (Meija, 2006; Kunenekov et al., 2009). CHOFIT was originally written in Pascal, so we adapted it to the R programming language. MFAssign() is meant to provide users with a more transparent method for formula assignment, especially for possible ambiguous molecular formula assignments compared to previous methods that have been developed. Ambiguity of molecular formula assignment increases with increasing molecular weight or the number of possible heteroatoms because the number of mathematically possible molecular formulas increases exponentially (Koch et al., 2007; Kind and Fiehn, 2007). Even when chemical feasibility is used to constrain the

assignments, there are many ambiguous molecular formulas. Ambiguity is often removed with other software tools using pre-defined rules and it is possible that one of the removed ambiguous formulas is actually the correct formula. Many molecular formula assignment software tools have quality assurance (QA) parameters and rules to reduce ambiguous assignments to only one molecular formula per mass (Stranz et al., 2015; Tolic et al., 2017). These restrictions lead to incorrect assignments, especially when sample types are inconsistent with those used to design the software (e.g. DOM for Formularity and petroleum extracts for Composer. A recent software package, called UltraMassExplorer (UME) provides ambiguous assignments without QA parameters (Leefman et al., 2018). While this is very useful for evaluating the molecular formula assignment options for a particular mass, UME is limited to the m/z range of 100-700, assigns molecular formulas using a database matching approach, and is restricted to C, H, O, N, P, S, $^{13}$C, $^{34}$S, $^{15}$N (Leefman et al., 2018).

### 5.2.2 Methods to Limit Ambiguity

When it comes to data analysis it is often helpful to have as few ambiguous assignments as possible because they make it difficult to know exactly the composition of a sample. For this reason, MFAssignR has options to limit the ambiguous molecular formula assignments to those that are the most likely. To do this, we incorporated many tools for QA remove less probable molecular formula assignments. We first review the QA tools and a priori assumptions used in existing formula assignment software. Formularity (Tolic et al., 2017), applies the assumption that the molecular formula with the fewest number of non-oxygen heteroatoms is the best molecular formula. This assumption has

been made in several NOM studies (Kujawinski and Behn, 2006; Ohno and Ohno, 2013; Tolic et al., 2017) and it is a fair assumption to make in general. However, since more heteroatoms are expected for molecular formula assignment (e.g. soil or wastewater NOM), it becomes possible that a formula with more heteroatoms is more probable. An example of this is the comparison of $C_{17}H_8O_5$ and $C_9H_{12}N_2O_7S$ which were both assigned for m/z 293.0436. If the formula with the maximum number of heteroatoms is removed, the $N_2S$ formula will not be assigned which may not always be the correct option. To separate the two formulas into separate peaks would require a resolving power > 600,000 at m/z 400 (Table 5.1).

**Table 5.1.** Mass differences and required resolving power for selected formula transitions.

| Formula Difference | Δmass | Resolving Power at m/z 400 |
|---|---|---|
| $C_xH_yN_3O_zS$ vs. $C_{x1}H_{y1}O_{z1}$ w/ $^{13}C$ | 0.63 mDa | 635K |
| $C_xH_yN_3O_z$ vs. $C_{x1}H_{y1}O_{z1}S$ w/ $^{13}C$ | 0.244 mDa | 1.64M |
| $CH_4$ vs. O | 3.639 mDa | 110K |
| $C_4$ vs. $O_3$ | 1.525 mDa | 262K |
| $C_xH_yN_2O_zS$ vs. $C_{x1}H_{y1}O_{z1}$ | 0.652 mDa | 613K |
| $SH_4$ vs. $C_3$ | 3.372 mDa | 119K |

In contrast, Composer makes the assumption that the more hydrocarbon-like molecular formula is the correct one (Stranz, 2015). Although it is reasonable to make this a priori decision for petroleum extracts, it may not be correct for more oxidized samples, such atmospheric organic aerosol. To avoid making a priori assumptions about the molecular formula composition, we developed a data dependent method to decrease a majority of the ambiguity using molecular formula extensions. Earlier versions of molecular formula

extensions were used in molecular formula assignment (Kujawinksi and Behn, 2006; Kunenkov et al., 2009; Stranz, 2015; Tolic et al., 2017). The molecular formula extensions provide confidence in the assignments at higher masses, but have not previously been used to decrease their ambiguity. Another algorithm that has been reported is the CHOFIT algorithm from Green and Perdue (2015) which was used as the core of our molecular formula assignment code. A description of the core CHOFIT algorithm, and the methods employed to limit ambiguity is below.

### 5.2.3 Theory of CHOFIT Algorithm

The core of the formula assignment functions in MFAssignR is the CHOFIT algorithm, which was developed by Green and Perdue (2015). The CHOFIT algorithm itself only assigns CH or CHO containing molecular formulas, but with the addition of some nested loops it is possible to include any heteroatom that is desired. The CHOFIT algorithm assigns molecular formulas using a low mass moieties (LMM) approach with $CH_4O_{-1}$ and $C_4O_{-3}$ which represent elemental exchanges. These LMMs form non-parallel lines within a single nominal mass window in van Krevelen (VK) space. Each line represents a different homologous series where the difference is $CH_4O_{-1}$ (blue lines) or $C_4O_{-3}$ (gold lines). Each series represents a common exchange series that is present within each nominal mass window. The lines for each moiety intercept at a location in the negative quadrants of the VK plot, thus the formulas that are within a straight line are related to each other by one of the LMMs. To use this method for formula assignment, the exact mass (EM) of an unidentified molecule is converted to its nominal mass (NM) and then the hydrocarbon that contains the highest possible carbon number is calculated. Then the

molecular formula is checked to see if it is on the $CH_4O_{-1}$ "mixing line". The mixing line is the line of molecular formulas that differ in formula by the low mass moiety within a single nominal mass. It is analogous to a $CH_2$ series across a spectrum. A representation of these mixing lines can be seen in Figure 5.1. If the difference between the measured EM and the EM of the initial formula is an integer value, then the same number of $CH_4O_{-1}$ LMMs are added to the initial formula to get the final formula. If the difference it is not an integer, then the LMM $C_4O_{-3}$ is subtracted from the initial formula to move the formula to the next mixing line, where the $CH_4O_{-1}$ test is performed again. This is done until a final formula is assigned, or the valid compositional space is used up. The VK plot for the CHO formulas within a single nominal mass is shown in Figure 5.1. This figure, which is adapted from Perdue and Green (2015), helps to highlight the mixing lines. This method greatly increases the speed of the function relative to other molecular formula calculators by decreasing the number of loops required to assign valid molecular formulas. An example of the calculations that contribute to assigning a molecular formula with the CHOFIT algorithm can be seen in Table 5.2.

**Table 5.2**. General steps of molecular formula assignment using the CHOFIT algorithm. First, the maximum number of carbon for the nominal mass (NM) of the neutral molecular mass is calculated and, the remaining mass is made up with hydrogen. Then, the mass difference between the measured mass (143.0349 Da in this case) and the exact mass (EM) of the trial formula is calculated and divided by the exact mass of $CH_4O_{-1}$ (0.03639 Da). If this value is not an integer, the LMM $C_4O_{-3}$ is subtracted from the first trial formula and the process repeats. This continues until an integer value is found. At this point, the requisite number of $CH_4O_{-1}$ LMM are added/subtracted from the trial formula in order to determine the final molecular formula. In this case 2 $CH_4O_{-1}$ LMM are added to the 3$^{rd}$ trial formula to obtain the final formula of $C_6H_8O_4$.

| Step | NM | Carbon | Hydrogen | Oxygen | EM | # of $CH_4O_{-1}$ moieties |
|---|---|---|---|---|---|---|
| **First trial formula** | 144 | 12 | 0 | 0 | 144 | 1.16 |
| **2nd after subtracting $C_4O_{-3}$** | 144 | 8 | 0 | 3 | 143.9847 | 1.58 |
| **3rd after subtracting $C_4O_{-3}$ again** | 144 | 4 | 0 | 6 | 143.9695 | 1.99 |
| **Final Formula, after adding 2 $CH_4O_{-1}$** | 144 | 6 | 8 | 4 | **144.0423** | |

While the CHOFIT core is unable to directly assign heteroatoms, combinatorial formulas and nested loops can be used to assign heteroatoms. The mass of a heteroatom combination is removed from the overall measured EM to obtain the CHO core. The core is then be assigned using the CHOFIT algorithm and then the heteroatom combination is added back to the final formula. This method generates all possible molecular formulas within the error tolerance, leading to a large number of ambiguous assignments. While generation of the maximum number of formulae affords ultimate method transparency, it can add additional complexity for the analysis and interpretation of the data. To address this, we introduced a number of optional QA steps and parameters to reduce the ambiguity in a data-dependent way.
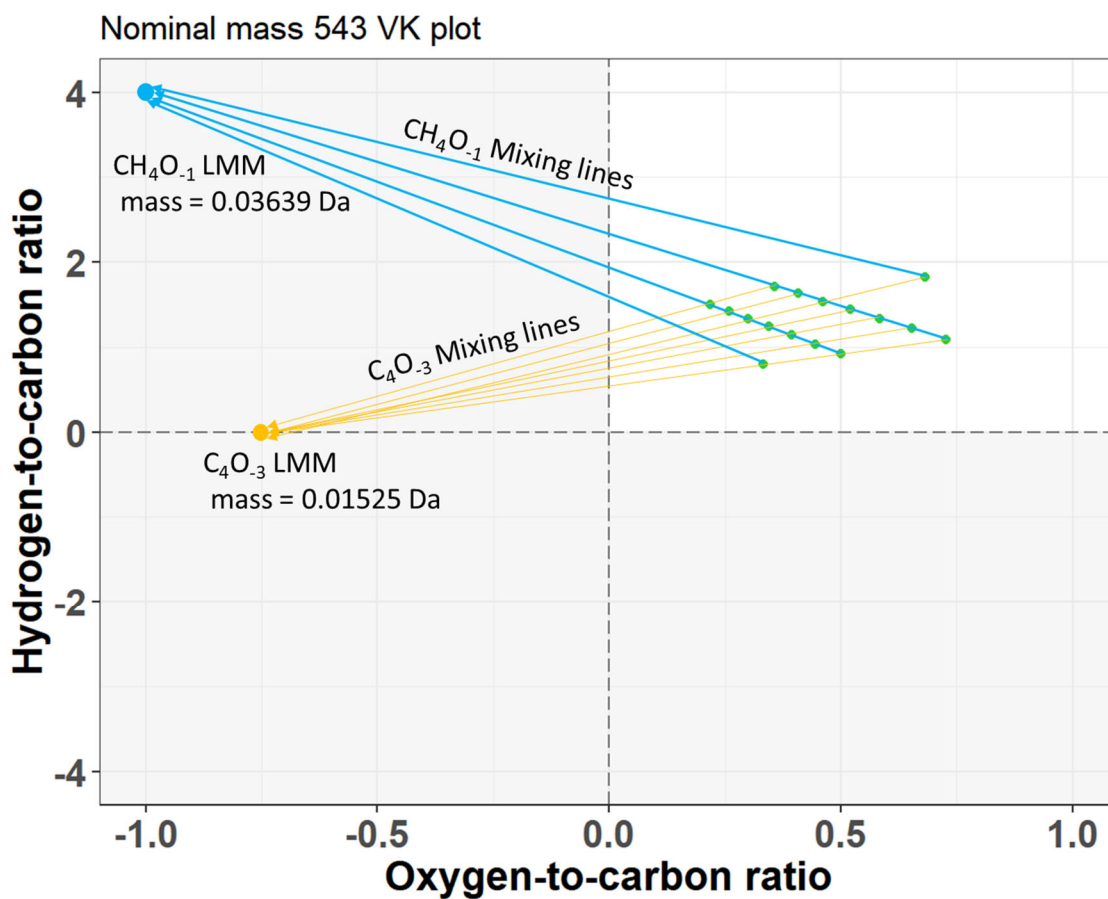
**Figure 5.1**. Adaptation of figure from Perdue and Green to help explain how molecular formulas are assigned in CHOFIT. Blue color is related to the $CH_4O_{-1}$ low mass moiety, meaning that the dots connected by blue lines vary by $CH_4O_{-1}$, while the gold color corresponds to the $C_4O_{-3}$ low mass moiety, dots connected by gold lines vary by $C_4O_{-3}$. The green points represent molecular formulas assigned at nominal mass m/z 543 in PMO-2. The unshaded quadrant represents the positive space where real molecular formulas exist, the shaded regions represent the negative space of a van Krevelen plot.

### 5.2.4 Formula Extension Background and as Way to Limit Ambiguity

The most significant method for reducing the ambiguity in MFAssignR comes from the use of formula extensions. A building block approach using formula extensions has been used to assign molecular formulas in several software packages (Kujawinski and Behn, 2006; Kunenkov et al., 2009; Tziotis et al., 2011; Tolic et al., 2017;). The Compound Identification Algorithm (CIA) described by Kujawinski and Behn, (2006) used a molecular formulas seed approach where all of the masses related to the seed by $CH_2$, $H_2$, or O mass differences are assigned by adding or subtracting the requisite number of those building blocks. Formularity extended this approach with the addition of $CH_4O_{-1}$, $CO_2$, $C_2H_4O$, and $C_2H_2O$ as possible building blocks. The molecular formula extension relationships of measured masses are determined using their KMD values. When the KMD values are the same, the masses are considered to be in the same homologous series. Thus, the molecular formulas are assigned by adding or subtracting the appropriate number of building blocks equal to the mass difference. The formula extensions provide confidence in the molecular formula assignments, especially at higher masses because they are related to peaks at lower masses within the mass spectrum, providing an overall lower number of molecular formulas. If the molecular formulas are calculated for higher mass peaks without restricting them to lower masses, the number of possible molecular formulas increases exponentially (Kind and Fiehn, 2007; Koch et al., 2007). In MFAssignR, the formula extension is used to improve formula assignment, and decrease the ambiguity of formula assignments without making a priori decisions about the composition of the sample, as described previously (Ohno and Ohno, 2013; Kujawiniski

and Behn, 2006; Stranz, 2015). Five formula extensions are used in MFAssign() and MFAssignCHO(). The extensions used are $CH_2$, $H_2O$, O, $H_2$, and $CH_2O$. Some of these can be redundant ($H_2$ and O vs. $H_2O$ for example), but the redundancy provides increased confidence because if a particular molecular formula has multiple relationships it is more likely to be a correct assignment. This assumption is the basis of the approach used in MFAssign and MFAssignCHO to decrease the molecular formula ambiguity with formula extensions. Formula extensions, require a good "seed" formula (Kujawinski and Behn, 2006; Koch et al., 2007). In MFAssign and MFAssignCHO, the preliminary seed formulas are generated by the CHOFIT algorithm where only the unambiguous molecular formula assignments are used as seed formulas. The remaining monoisotopic masses are added to the unassigned mass list which are assigned using the formula extension approach. In MFAssignR, a combination of the Kendrick mass defect, and the z* (Hsu et al., 1994; Stenson et al., 2003) for each of the bases is used. The masses are then matched using KMD and z* in several steps as shown in Figure 5.2. The molecular formula extension process begins with masses below the user defined de novo threshold. It then performs a user defined number of loops to assign molecular formulas using the 5 extensions. At the end of each loop, the molecular formulas that are unambiguously assigned act as seeds for the next loop. To determine the unambiguous assignments, at this point, a basic elemental ratio test is applied to the assigned molecular formulas to remove those that have an O/C or H/C ratio that exceeds the user defined limits. After this step, the function determines how many seed formulas are related to the newly assigned formulas, if there are two assignments for the same mass, the function chooses the formula that has more "paths" to it. The paths indicate the number of seed formulas

185

related to the new assignment. The selected formula is then treated as an unambiguous assignment for the next loop. After the series of loops are finished for the first segment of the mass spectrum, formula extensions are performed on the next segment, which uses seeds up to the de novo threshold + 200. This process is repeated until the entire mass spectrum has been covered. Formula extensions significantly reduce the number of ambiguous assignments, but some ambiguous assignments still make it through. These ambiguous molecular formulas are identified in the output.
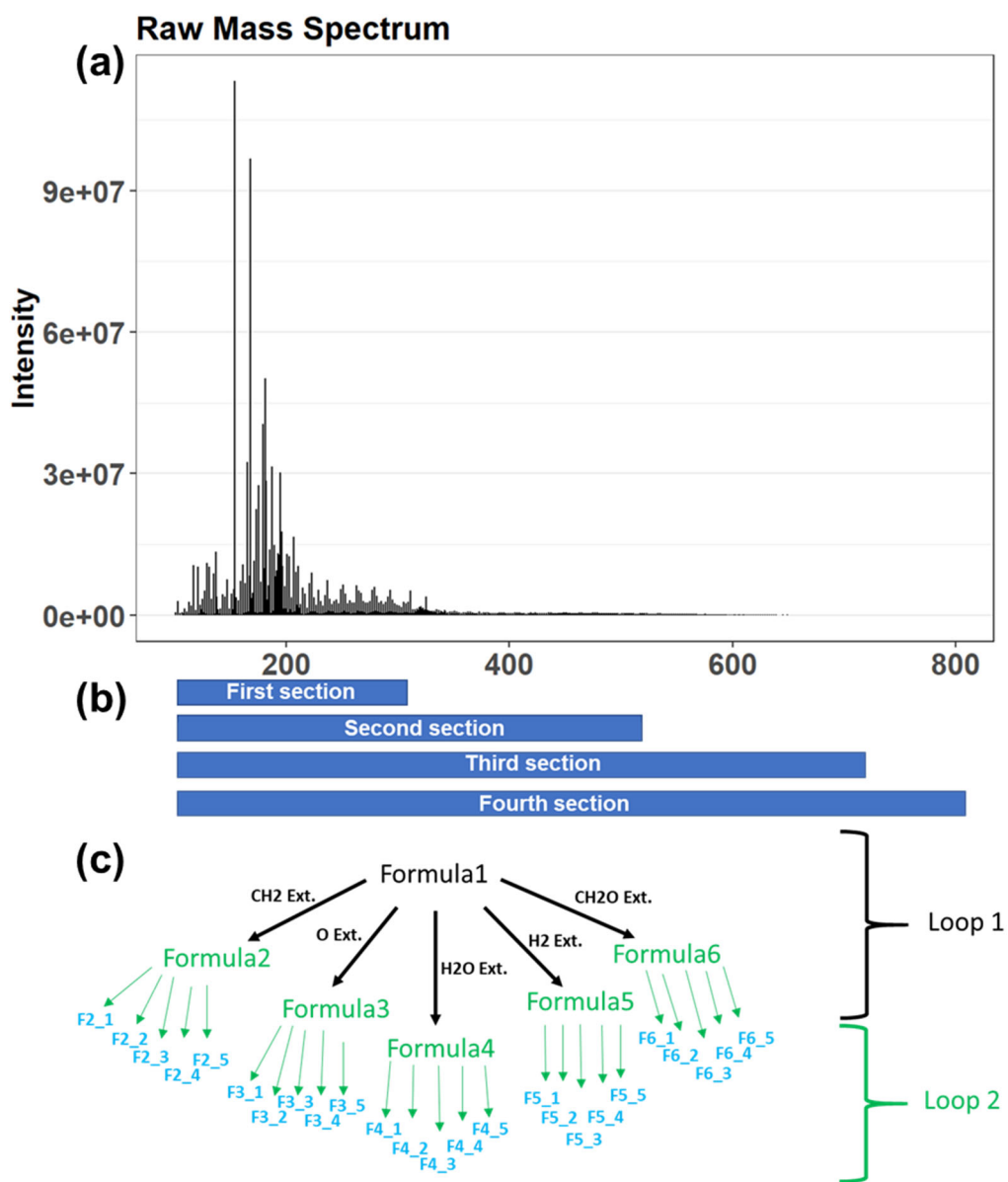
**Figure 5.2.** Schematic for formula extension. Panel a shows an example mass spectrum, Panel b in conjunction with the mass range in Panel a demonstrates the how the segments of formula extension cover the full mass range over several steps, Panel c shows how a single "seed" formula can be used to assign many molecular formulas with the 5 formula extension bases.

187

### 5.2.5   Isotope Molecular Formula Correction

Part of the molecular formula assignment is matching the masses flagged as potential

polyisotopic molecular formulas to the assigned monoisotopic masses. The initial

separation of monoisotopic and polyisotopic is done with the IsoFiltR function which is

described more thoroughly in Section 5.4. IsoFiltR does a good job of identifying

monoisotopic and polyisotopic masses, but some are still incorrectly identified as

polyisotopic. To address this issue, a second round of formula extensions is done for the

polyisotope masses that were not, or were incorrectly, matched to a monoisotopic mass.

The masses that are incorrectly matched to a monoisotopic mass are typically those that

IsoFiltR flags as a $^{34}$S mass, but shouldn't be flagged as such. These masses can be

matched to a sulfur containing molecular formula, but if the sulfur containing formula

doesn't have a corresponding $^{13}$C mass it is not likely correct. The logic is that if the

sulfur containing formula can have a matching $^{34}$S mass, it should also have a matching

$^{13}$C mass because the $^{13}$C mass should be more abundant and easier to detect. So, if a

sulfur containing molecular formula has a $^{34}$S isotope, but no $^{13}$C isotope, then it is

considered to be incorrect for the purposes of this test, and the "$^{34}$S" mass is added to the

list for secondary assignment via formula extension. This secondary step ensures that all

isotopic peaks are given the opportunity to be assigned a molecular formula. Formula

extensions are always used to assign molecular formulas in MFAssign and

MFAssignCHO, but if the user wants to have more ambiguous assignments the path

counting aspect of formula extension can be turned off. This allows all ambiguous

assignments that fulfill the other QA parameters to be reported in the output. The

effectiveness of the formula extensions is shown in Figure 5.3, which shows the number

of paths that are used to select molecular formulas. This figure illustrates how the

formula extensions are used to decide between "correct" and "incorrect" assignments

without additional *a priori* decisions such as choosing the formula with the lowest

number of heteroatoms.
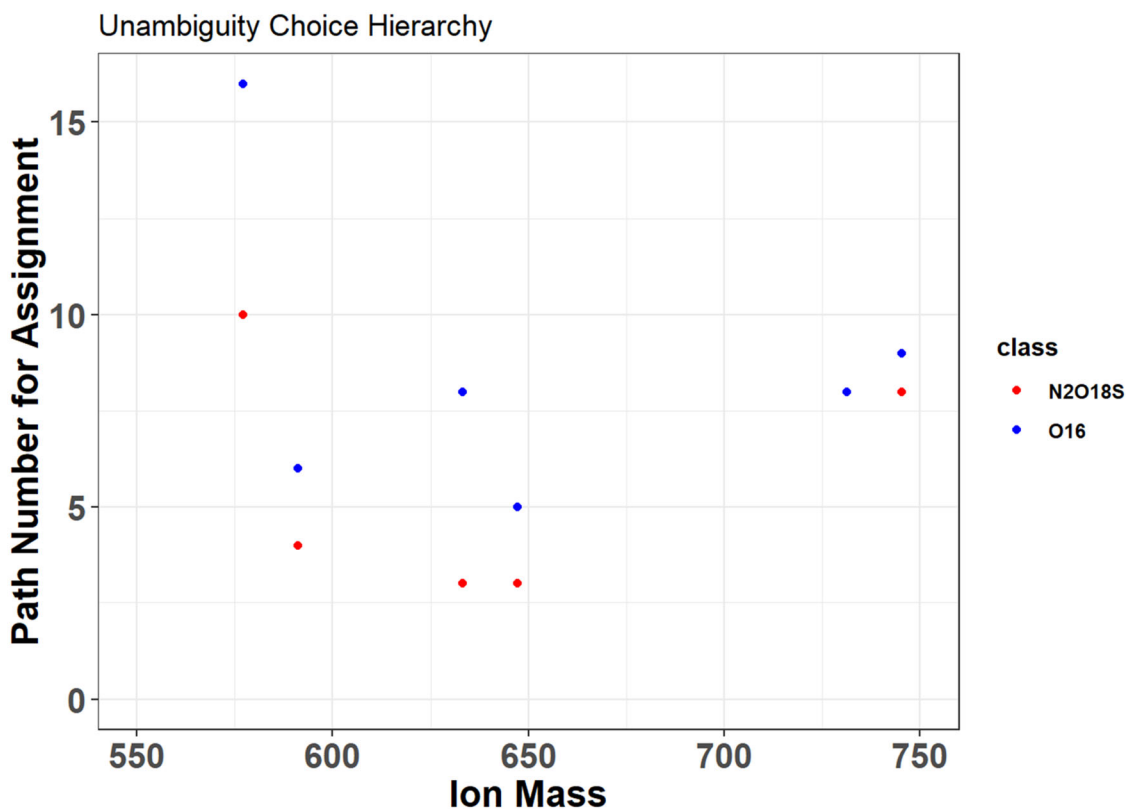


**Figure 5.3.** Demonstrating the number of paths for each formula assignment. This is how the formula extension decides on the correct formula, whichever one has the higher number of paths ($O_{16}$ in this case) is chosen as the correct molecular formula. The molecular classes being compared here are $N_2O_{18}S$ (red) and $O_{16}$ (blue). Each set of points represents an ion mass that has two molecular formulas preliminarily assigned.

## 5.2.6  Quality Assurance Parameters

### 5.2.6.1  Sulfur Isotope Check

In addition to the formula extensions, there are several other QA parameters that can help decrease ambiguity and improve the quality of formula assignments. One of the methods is the sulfur isotope check, which checks all sulfur containing assignments that come out of the CHOFIT core to see if they have a matching sulfur isotope mass. If they do, the assignment can be used as a seed to assign other molecular formulas, if not, the assignment is discarded and the mass will be assigned during formula extension. This ensures that all sulfur containing molecular formulas are related in some way to a molecular formula with a sulfur isotope to help confirm its identity.

### 5.2.6.2  Nominal Mass Series Check

Additionally, a nominal mass series check, ensures that molecular formulas within a nominal mass window with 36.4 mDa mass differences all vary by the exchange of O with $CH_4$ (Koch et al. 2007). If a molecular formula within the series doesn't have enough oxygen to account for the rest of the masses in the series, it would be deemed incorrect. For example, if a molecular formula with one oxygen is assigned, but is also related to 3 higher masses in series, it is likely incorrect because the subsequent molecular formula would have a negative number of oxygen to follow the trend (Koch et al., 2007). The concept of $CH_4$ vs. O is implicit in the CHOFIT core algorithm (Perdue and Green, 2015; Green and Perdue, 2015), but this additional step can decrease ambiguity in the formulas that are assigned via formula extensions.

### 5.2.6.3  High Heteroatom Check

The final optional QA parameter is the high heteroatom check. This check chooses the

molecular formula with the least number of non-oxygen heteroatoms consistent with

previous studies (Kujawinski and Behn, et al., 2006; Ohno and Ohno, 2013; Tolic et al.,

2017). This parameter is generally applicable to NOM, but because it has the potential to

limit atmospherically relevant assignments, we developed the other QA parameters to

reduce ambiguity, as described previously.

### 5.2.6.4  User Controlled QA Parameters

In addition to the optional QA parameters, several others can be set by the user,

including: the min/max oxygen-to-carbon ratio (O/C), hydrogen-to-carbon ratio (H/C),

and double-bond-equivalent minus oxygen (DBE-O) parameters. The O/C and H/C

parameters can be set according to the expected composition of the sample; typical

boundaries are 0.1 to 2 for O/C and 0.3 to 2.5 for H/C. The DBE-O parameter was

developed by Herzsprung et al. (2014) as a way to remove unlikely molecular formula,

and the default boundaries are -13 to 13. The minimum number of allowed oxygen can

also be set by the user.

### 5.2.6.5  Default QA Parameters

Several QA parameters are not subject to user inputs because they are related to basic

chemical formula feasibility instead of the more qualitative assessment parameters

discussed previously. These include the Senior rules (Senior, 1952; Kind and Fiehn,

2007; Green and Perdue, 2015), nitrogen rule, large atom rule, maximum hydrogen rule,

and maximum double bond equivalent rule (Lobodin et al., 2012). The Senior rules were developed by Senior (1952) and can be used to limit molecular formulas to only those that are chemically feasible with regard to bonding and valence levels. The valence level of the atoms represents the number of open spots in their outer electron shell. For example, carbon typically has a valence level of 4 because that is the number of open spots in the outer octect of electrons. Rule 1 states that the sum of valences or the total number of atoms having odd valences is even (Kind and Fiehn, 2007). Rule 2 states that the sum of valences is greater than or equal to twice the maximum valence (Kind and Fiehn, 2007). Rule 3 states the sum of valences is greater than or equal to 2n-1, where n is the number of atoms (Kind and Fiehn, 2007). These rules are used in several molecular formula assignment software tools including: Seven Golden Rules by Kind and Fiehn, (2007) and the original CHOFIT (Green and Perdue, 2015). While the Senior rules themselves cannot be changed, the valence level of S, N, and P can be set by the user, which can be useful if oxidized sulfur with a valence level of 6 is expected. The nitrogen rule states that a molecular formula with an odd number of nitrogen must have an odd neutral mass. The maximum hydrogen rule states that the number of hydrogen cannot exceed 2 * #carbon + 2. The maximum DBE rule as described by Lobodin et al. (2012) states that the DBE cannot exceed 90% of the total number of carbon and nitrogen atoms in a molecular formula. The large atom rule states that the number of atoms with a mass greater than $^{12}C$ must be less than the value of the exact mass of the molecular formula divided by 13.

### 5.2.7 Advantages of MFAssignR for Heteroatoms

Most molecular formula assignment software tools assign molecular formulas containing C, H, N, O, S, and P, potentially with the isotope masses of $^{13}$C and $^{34}$S (Kujawinski and Behn, 2006, Green and Perdue, 2015, Leefmann et al., 2018, Stranz, 2015). Recently, Tolic et al. (2017) developed a method that can also include select halogens, but it is based on a database search, and thus is less flexible than methods that can directly calculate molecular formulas. Using the CHOFIT algorithm, it is theoretically possible to add as many heteroatoms as desired. Therefore, in addition to $^{12}$C, $^{1}$H, $^{14}$N, $^{16}$O, $^{32}$S, and $^{31}$P, we added $^{2}$H, $^{15}$N, $^{35}$Cl, $^{37}$Cl, and $^{19}$F. The addition of these heteroatoms is very helpful in performing experiments with isotopically labeled species (Leverton, 2019), and in investigating halogenated wastewater species, which may be found, for example, in outflow from wastewater treatment facilities. Following this overview of MFAssign, we now turn to the functions that are necessary to ensuring that the results from MFAssign are as robust as possible.

## 5.3   Noise Estimation

### 5.3.1   Importance of Noise Estimation and Methods for its Estimation

First, we discuss the instrument noise estimation function in MFAssignR, KMDNoise, and why noise estimation is important to ensure good data quality (Riedel and Dittmar, 2014, Kilgour et al., 2017). The noise that is being considered here is thermal and electrical noise from the instrument itself. This noise is inherent in any measurement due to the movements of electrons and imperfections in the instrumentation. Many methods exist for estimating the noise level. The most simple approach evaluates the peak

intensity in regions without analytes (Kew et al., 2017). However, the noise level is not necessarily consistent throughout the spectrum, especially for FT-ICR MS (Hawkes et al., 2016), so these small regions may not provide the most accurate assessment of the noise level. Slightly more advanced methods rely on the assumption that all peaks with a mass defect of 0.3 to 0.9 are noise peaks (Riedel and Dittmar, 2014). This assumption ignores the possibility of higher intensity multiply charged peaks or harmonic signals in this range for NOM. This assumption may lead to higher estimations of noise than is appropriate due to the inclusion of the higher intensity multiply charged or harmonic peaks. Harmonic signals can be produced by a variety of factors such as signal saturation of the detector, the inherent finiteness of an ICR cell, or issues with the Fourier transform of the data (Mathur et al., 2009). They can also be produced by physical problems with the instrument itself such as bad connections in the wires (Mathur et al., 2009). In the Orbitrap, harmonics are also possible, but are more limited relative to ICR instruments (Makarov, 2000; Zubarev and Makarov, 2013). More advanced noise estimation methods such as those described by Zhurov et al. (2014); Zielinski et al. (2018), Kilgour et al. (2017), and Riedel and Dittmar (2014) can be effective, but are often more difficult to implement. For example, the main method described in Riedel and Dittmar (2014) requires the use of multiple blanks to estimate the noise for samples that are run using the same parameters. AutoPiquer (Kilgour et al., 2017) uses isotopic fine structures to pull out peaks from the noise thereby separating the analyte signal from surrounding noise peaks. The methods described by Zhurov et al. (2014) and Zielinski et al. (2018) are very similar to each other, both use a histogram distribution of the peak intensities in the raw spectrum to separate the noise peaks from the analyte signal, the only difference is that

the Zhurov et al. (2014) method uses the log10 of intensity, while Zielinski et al. (2018) uses the native intensity. These methods can work well when the noise and intensity are well separated (Figure 5.4a), but oftentimes they are not (Figure 5.4b).
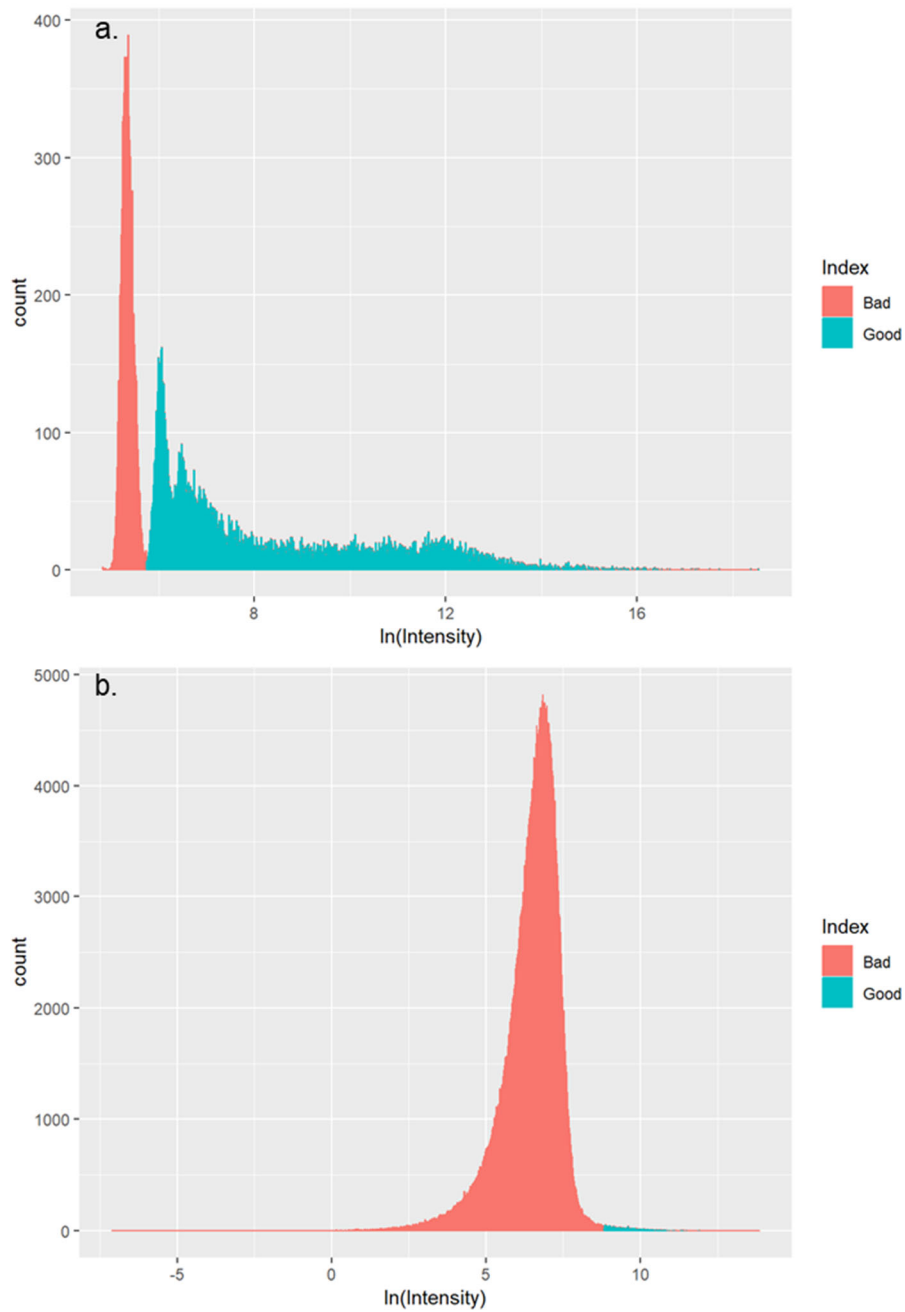
**Figure 5.4**. Histograms demonstrating the distributions of raw intensity values in two different aerosol samples used for the Zhurov et al., (2014) intensity histogram noise estimation method. Panel a is an example of the method working well, while Panel b is an example of it not working well.

### 5.3.2 Theory of KMDNoise

Due to the limitations described above, we developed a new method for estimating the noise in a mass spectrum, using the Kendrick mass defect (KMD) values from a raw mass spectrum called the KMD slice method. Traditionally, a KMD is calculated using a $CH_2$ base, where masses that differ by the exact mass of $CH_2$ have the same KMD. This improves the identification of molecular formulas because in theory, if you know the identity of one of the peaks in a $CH_2$ homologous series you can identify the rest by adding or subtracting the requisite number of $CH_2$ groups. Typically, KMD analysis is performed on data that has already had the noise removed, but an interesting pattern of analyte and noise peaks can be observed from the raw mass spectral peaks. Since the mass of noise peaks is random, the KMD values for the noise peaks are also random. This leads to "islands" of analyte peaks surrounded by a "sea" of low intensity noise peaks. This is shown in Figure 5.5. Some of the high intensity peaks form smaller "islands" above and below the largest "island"; these peaks represent either ions that are multiply charged or harmonics. It is very difficult to tell whether the peaks are multiply charged or harmonic, but the intensity of these peaks would bias the estimated noise level when using a noise estimation based on normal mass defect (Riedel and Dittmar, 2014). This highlights the improved flexibility of the KMD slice method because it can avoid those regions more easily. The inclusion of the noise peaks also causes the entire theoretical range of the KMD plot to be filled, which allows the identification of the overall slope of a KMD plot. The overall equation for a KMD plot is $y = 0.1132x + b$, with y being the KMD value, x being the measured ion mass, and b being the y-intercept. Changing b

allows different segments of the spectrum to be isolated, in most mass spectra, good

values for b are 0.2 and 0.05 to select a region of the plot with as few high intensity peaks

as possible (Figure 5.5b). The peaks within this "slice" are then averaged with that value

representing the average noise level for the overall spectrum. This value can used in

conjunction with a user-defined signal to noise ratio (typically 3 – 10) to remove peaks

that have an intensity that is too low.
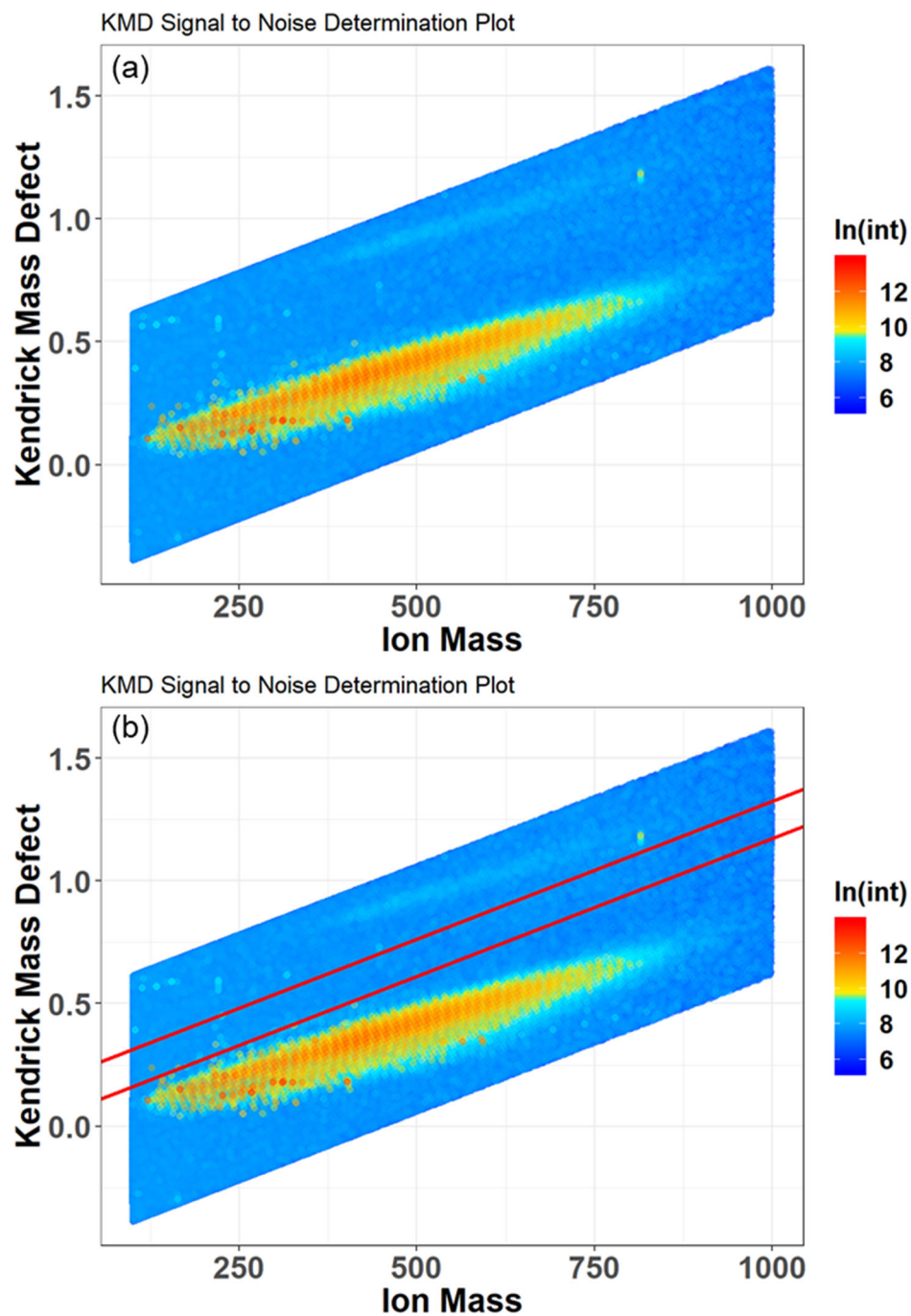
**Figure 5.5**. Showing the KMD plot for the raw mass spectrum of PMO-2. Panel a contains the plot without additional information. Note the light blue "island" near the top of the plot; these represent doubly charged or harmonic plots with a mass defect of ~0.5. Panel b shows the same plot, but with the KMDNoise function boundaries indicating where the noise is estimated with the default settings of that function.

### 5.3.3 Noise Estimation Method Comparison

A comparison of some of the noise estimation methods described above is shown in Figure 5.6. Here the noise estimations obtained from the "simple", Riedel and Dittmar (2014) mass defect, Zhurov et al. (2013) histogram, and the KMD slice estimation methods for the raw mass spectrum of PMO-2 collected with an FT-ICR MS with 400K resolving power at m/z 400 are presented. The "simple" method uses an analyte free region of the spectrum to estimate the noise. The results of this comparison (Figure 5.6) show that the estimation with KMDNoise produces the lowest signal-to-noise threshold, while still being comfortably above the noise level. The Riedel and Dittmar (2014) method has a noticeably higher signal-to-noise relative to the KMD slice method. This is likely due to the inclusion of more intense peaks in the mass defect ~0.5 range (Figure 5.5). The "simple" method produced the highest signal to noise ratio, which is likely due to the increased noise in the range of m/z 950-1000 where it was estimated. This is a limitation of the "simple" method as the noise increases with increasing mass and has been described elsewhere (Hawkes et al. 2016). The Zhurov et al. (2013) method had the second lowest estimated noise, but it did not work correctly because the intensity histogram was unable to separate the analyte and noise peaks (Figure 5.4b). In this case, the estimated noise was based on the mode of the distribution, which is mentioned as a secondary option as described in Zhurov et al. (2013). So, although the estimated noise with this method is similar to what we see for the KMD slice method, the KMD slice method is much easier to use, and provides more reproducible results.
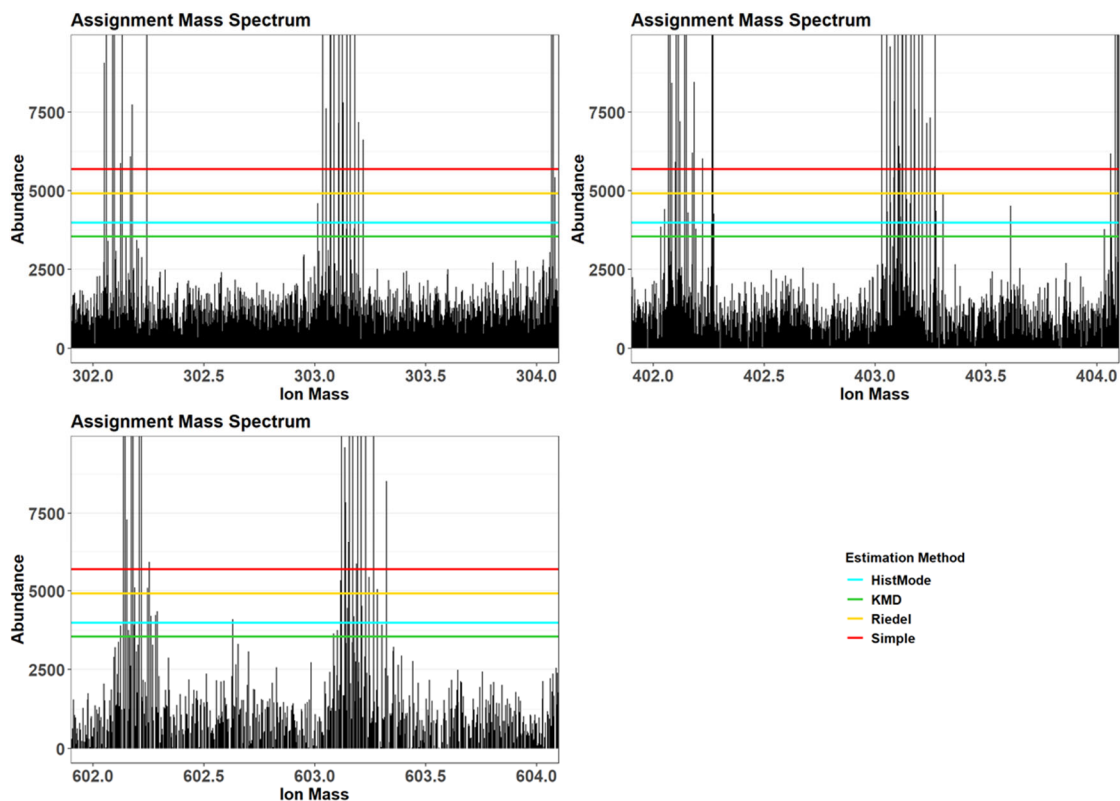
**Figure 5.6**. Showing the noise levels estimated by 4 different methods of noise estimation. The noise estimation with the Riedel and Dittmar method (Riedel) is in gold, the "Simple" method estimate is in red, the estimation based on the mode of the histogram distribution (HistMode) is in cyan, and the noise estimation from the KMD Slice method (KMD) is in green. Each panel represents a different mass range in the spectrum of PMO-2.

### 5.3.4 Similarities and Differences to Other Methods of Noise Estimation

This method is quite simple, but has not been reported previously. It uses the idea of estimating the noise based on the intensity of peaks in a region without analytes, consistent with the less advanced of the two methods described by Riedel and Dittmar (2014). Essentially what the KMD slice method does is pull out the noise peaks from regions that do not have analyte peaks, but instead of needing a range over several consecutive m/z as with the original, it can get them from every range that has noise in it across the spectrum, as opposed to a relatively small region at the high end of the mass range. It is somewhat similar to the mass defect method described by Riedel and Dittmar (2014) in that it uses assumptions about mass defects to estimate the noise over the entire spectrum. The primary advantage of this new method though is that it is more precise about which peaks are used for the noise estimation. The most obvious advantage is due to the observation of doubly charged or harmonic signals that are sometimes present in mass spectra. These peaks have a natural mass defect of ~0.5, placing them in the range of mass defects used to estimate the noise with the method described by Riedel and Dittmar (2014). Their inclusion may lead to an over estimation of the noise, which in turn may cause low intensity analyte peaks to be removed from the mass spectrum. In the KMD slice method however, these doubly charged or harmonic peaks form a visible secondary cluster above the cluster of singly charged peaks. This cluster can then be avoided by using the appropriate boundaries, providing a more accurate assessment of the actual noise level.

## 5.4   Isotope Filtering

### 5.4.1   Importance of Identifying Polyisotope Masses

A function to identify likely polyisotopic masses was developed for MFAssignR called

IsoFiltR(). Polyisotopic masses are masses where the molecular formula responsible for

them contain more than one type of isotope for one of the types of atoms in the molecule.

For example, a molecular formula with a $^{13}C$ atom present would be responsible for a

polyisotopic mass. Identification of polyisotopic masses before or during formula

assignment is key to ensuring good quality data. If these masses are not correctly

identified, they can be incorrectly assigned as monoisotopic masses, and lead to an

incorrect interpretation of the molecular composition. An example of a common incorrect

formula assignment is when a mass that is representative of a molecular formula that

contains a $^{13}C$ is assigned a monoisotopic molecular formula of the basic form

$C_xH_yN_3O_zS$. This elemental exchange has a mass difference of $\Delta mass = 0.63$ mDa, which

makes it difficult to resolve for many ultrahigh resolution mass spectrometers (Table 5.1).

An example of the formulas that result from such an exchange is for m/z 350.0816 where

the correct molecular formula is $C_{12}{}^{13}C_1H_{18}O_{11}$, but a molecular formula of $C_{15}H_{17}N_3O_5S$

can also be assigned. The mass measurement error cannot even be used to differentiate

these assignments because the error was lower for the incorrect assignment. This can be

seen visually in Figure 5.7. The resolving power needed to separate these peaks and

others at m/z 400 can be seen in Table 5.1. A van Krevelen (VK) plot with the ambiguous

MFs is shown in Figure 5.8. In this case it would be relatively easy to recognize that the

$N_3O_xS$ assignments are less likely because they have a large number of heteroatoms for

natural organic matter (NOM) (Kujawinski and Behn, 2006; Ohno and Ohno, 2013). The distribution of the species in a triangular shape at low O/C ratios may also raise concern for someone analyzing the data. In other cases, it is not always possible to know whether an assignment should be a monoisotopic or polyisotopic molecular formula assignment, especially when additional heteroatoms are expected. For this reason, many formula assignment software packages offer some sort of polyisotope identification during formula assignment (Tolic et al., 2017; Leefmann et al., 2018; Stranz, 2015), which are either based on mass differences or database matching. In NOM the most important isotopes that are $^{13}$C and $^{34}$S, so they are most often considered.
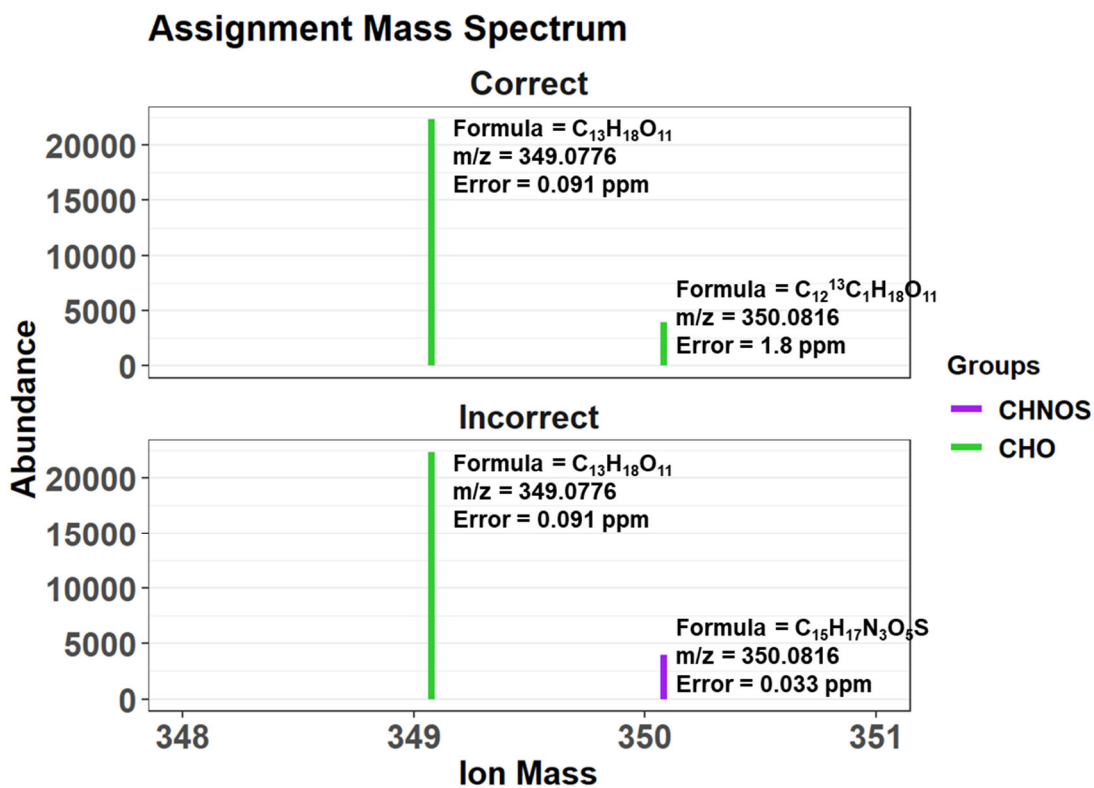
**Figure 5.7**. Reconstructed mass spectra showing how a mass can be incorrectly assigned as a monoisotopic mass. The figure includes the corresponding monoisotopic mass and formula, along with the mass measurement error for all the assigned molecular formulas.
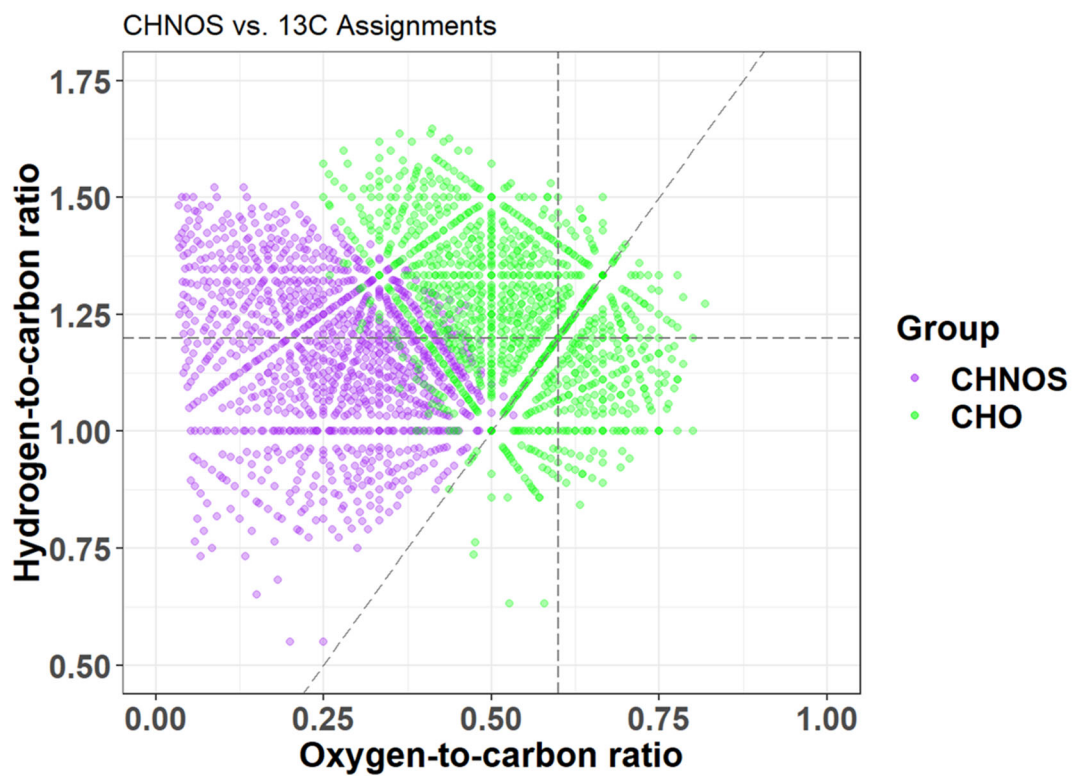
**Figure 5.8**. $^{13}C$ isotopic CHO formulas assigned with $^{13}C$ and when $^{13}C$ is not allowed. Each ion mass is assigned both a $^{13}C$ CHO formula (green) and a CHNOS (purple) formula.

### 5.4.2 Considerations for Polyisotope Identification

Every isotope has a specific abundance in nature. $^{13}C$ has a natural abundance of 1.10 % meaning that any particular carbon atom has a 1.10 % chance of being $^{13}C$ (de Hoffmann and Stroobant, 2007). Therefore, the abundance of the single $^{13}C$ mass peak in the mass spectra scales with the number of carbon atoms. The theoretical intensity of a $^{13}C$ isotope peak is calculated using Equation 5.1:

$$Int.Ratio = \#carbon \times 0.9890^{\#carbon-1} \times 0.011 \qquad \textbf{Eq. 5.1}$$

where *Int.Ratio* is equal to the ratio of the mono/poly intensities, 0.9890 is equal to the natural abundance of the $^{12}C$ atom, 0.011 is equal to the natural abundance of the $^{13}C$ atom, and #carbon is equal to the number of carbon atoms in the molecular formula. A molecule with 20 carbon for example, would have an associated $^{13}C$ peak intensity that is 22.2% of the intensity of the monoisotopic peak. Due to the large numbers of carbon present in most organic matter, $^{13}C$ is an important isotope for identification and confirmation of molecular species. The theoretical intensity of the $^{34}S$ peak is calculate using Equation 5.2:

$$Int.Ratio = \#sulfur \times 0.958^{\#sulfur-1} \times 0.042 \qquad \textbf{Eq. 5.2}$$

where Int. Ratio is equal to the ratio of the mono/poly intensities, 0.958 is the natural abundance of the $^{32}S$ atoms, 0.042 is the natural abundance of the $^{34}S$ atom, and #sulfur is equal to the number of sulfur atoms in the molecular formula. Due to the limitations of ultrahigh resolution MS, the theoretical abundance of isotopes is not as reliable as it is

with other types of mass spectrometry such as isotope ratio mass spectrometry (Muccio and Jackson, 2008; Weber et al., 2011). For analytes with a monoisotopic peak intensity only slightly above the noise level, the associated polyisotopic peak can fall below the S/N threshold. Furthermore, low abundances deviate from the expected isotope abundances according to the effect of isotope dilution (Weber et al., 2011). Figure 5.9 provides a demonstration of isotope dilution using the abundances of the polyisotopic masses with $^{34}$S peaks. The principle is the same for $^{13}$C (not shown). In Figure 5.9, one sees that when a set of peaks is well above the S/N level, the isotopic ratio is nearly equivalent to the theoretical value, but as the monoisotopic peak gets closer to the signal to noise ratio, the isotopic ratio deviates significantly. Due to this deviation from the theoretical values, it is difficult to identify all of the potential polyisotope peaks using isotope. Many formula assignment software tools that consider polyisotopic masses such as $^{13}$C, use expected isotopic patterns and check them against a database containing those patterns or by simple mass difference (Stranz, 2015; Tolic et al., 2017; Leefmann et al., 2018). MFAssignR does not use a database to assignment molecular formula, and a simple mass difference is not solely sufficient for identifying polyisotopic masses.
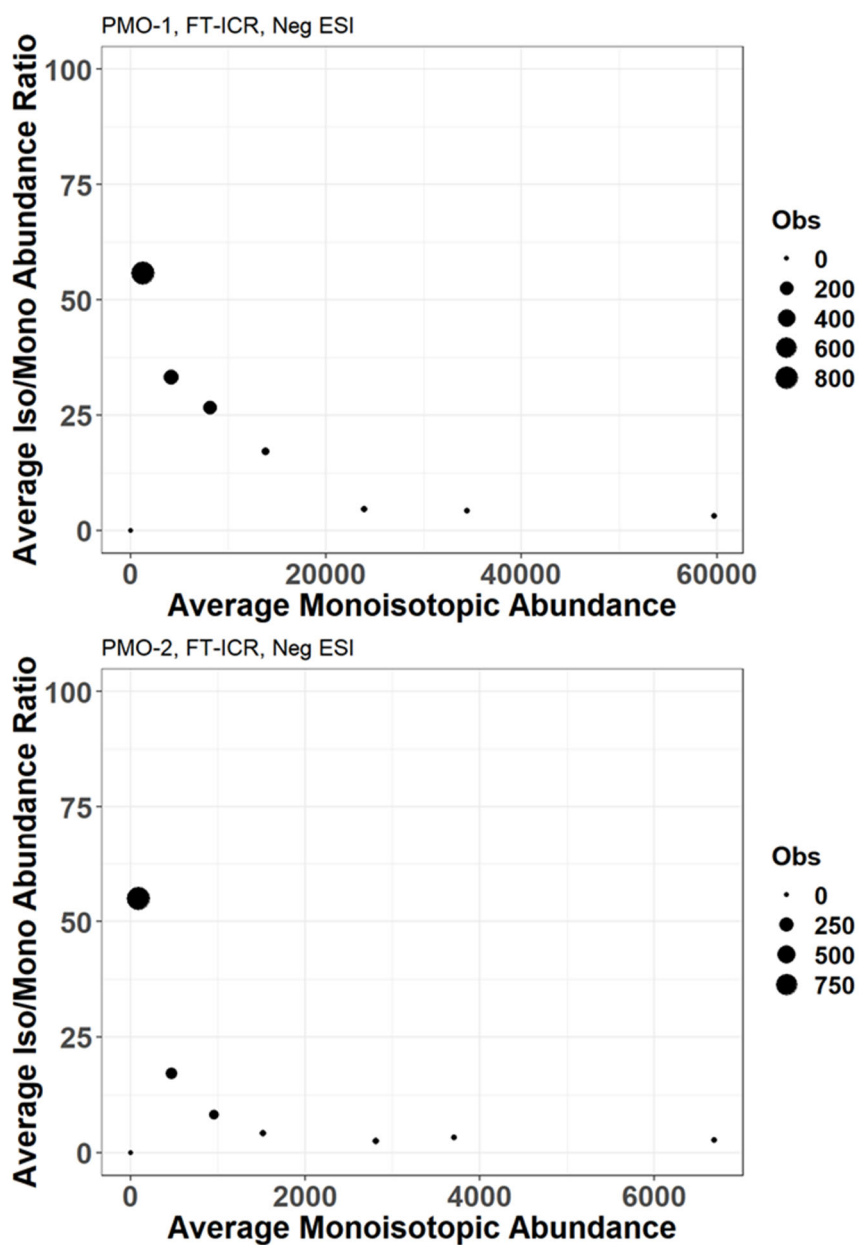
**Figure 5.9.** Demonstration of $^{34}$S isotope dilution for sulfur species in in PMO-1 (top) and PMO-2 (bottom). Lower Average Monoisotopic Abundance (AMA) indicates that the molecular formula lower intensity and is approaching the noise. The larger AMA indicate that the formula is getting further above the noise level. As the AMA increases, the points plateau at around 5%, which is what would be expected for the natural abundance of $^{34}$S (4.2%).

### 5.4.3   Theory of IsoFiltR Function

The IsoFiltR function in MFAssignR identifies probable polyisotopic masses containing 1 or 2 $^{13}$C or 1 $^{34}$S before molecular formula assignment by using four quality assurance (QA) steps. This represents a unique attempt to identify polyisotopic masses prior to formula assignment for UHR MS data, as previous literature methods were either based on database patterns (Yang et al., 2015; Tolic et al., 2017), mass differences after assignment (Leefman et al., 2018), or were made for lower resolution instruments (Zheng et al., 2018). The steps for identifying polyisotopic masses are described below.

#### 5.4.3.1   Step One: Mass Matching

IsoFiltR first matches every mass in the spectrum with every other mass, making a data frame with $N^N$ observations, where N is the number of observed peaks in the mass spectrum. This can become unmanageable for a standard computer, so the mass list is broken into 10 overlapping sections and the difference in mass between the matched observations is calculated. The observation pairs that match within $\pm$ 5 ppm of the theoretical mass difference of the isotope of interest (1.003355 Da for $^{13}$C and 1.995797 Da for $^{34}$S) are retained and all other pairs are removed from further consideration. After this is done for each of the sections, they are re-combined into a single dataset and all duplicate mass pairs are removed before the next step of filtering.

### 5.4.3.2   Step Two: Isotopic KMD Series

IsoFiltR then makes use of the mass difference between $^{12}C$ and $^{13}C$ (or $^{32}S$ and $^{34}S$) with

a KMD. The equation to calculate KMD requires the calculation of the Kendrick mass

(KM) the equation for KM and KMD are below (Equations 5.3 and 5.4):

$$KM = mass * 14.01565/14 \qquad \textbf{Eq. 5.3}$$

$$KMD = KM - NM \qquad \textbf{Eq. 5.4}$$

Typically, KMD is used with a $CH_2$ base, such that the masses are re-normalized by the

exact mass of $CH_2$ (14.01565 Da) instead of carbon. This causes the masses that differ by

the exact mass difference of 14.01565 to have the same KMD value. Since the $CH_2$ based

KMD is not helpful for identifying isotope pairs, the mass difference for $^{12}C$ and $^{13}C$

(1.003355 Da) or $^{32}S$ and $^{34}S$ (1.995797 Da) is used instead. Similar to the $CH_2$ KMD, if

the mass difference of $^{12}C$ and $^{13}C$ is used as the KMD base, all of the masses that differ

by that mass will have the same KMD value, making them easier to identify. After the

$^{13}C$ or $^{34}S$ KMD values are calculated for the preliminary monoisotopic and polyisotopic

masses, those values are subtracted from one another. Theoretically, the number should

be exactly 0, but because there is some inaccuracy in the measurements, some degree of

error is allowed. The remaining quality assurance (QA) steps will be enough to limit the

number of false positives in the identification of polyisotopic masses. Thus, if the

absolute value of the subtracted number is $< 0.00149$ the pair of peaks are considered to

be in a series and are moved to the next step.

### 5.4.3.3 Step Three: Resolution Enhanced Kendrick Mass Defect

IsoFiltR then uses the so-called resolution enhanced KMD (Zheng et al., 2018; Fouquet et al., 2017). Essentially, the resolution enhanced KMD (KMDr) divides a repeating mass unit (for example 14.0565 for $CH_2$) by an experimentally derived integer. This separates the isotope mass pairs by a consistent value. To obtain the separation desired, we used the $CH_2$ base (14.01565 Da) and two integers for dividing it. For $^{34}S$ the integer is 12 and for $^{13}C$ the integer is 21. The integer for $^{13}C$ was obtained from Zheng et al. (2018), while the integer for $^{34}S$ was determined by trial and error. Using the calculation and the integers mentioned previously we observed a consistent difference in the KMDr values for $^{13}C$ and $^{34}S$ peaks relative to their $^{12}C$ and $^{32}S$ counterparts. For the separation of $^{12}C$ and $^{13}C$, the KMDr difference was either -0.496 or 0.503, and for the separation of $^{32}S$ and $^{34}S$ the difference was either -0.291 or 0.709. A visual representation of this can be seen in Figure 5.10. The difference was calculated by subtracting the KMDr value for the suspected polyisotopic mass from the monoisotopic mass. As in the second step, there is some allowance of measurement error in these values to account for potential inaccuracies in the measurements. For the purpose of filtering the isotope pairs, the limits are $-0.4975 <$ KMDrDiff $< -0.494501$ and $0.501501 <$ KMDrDiff $< 0.5045$ for $^{13}C$, and $-0.29349 <$ KMDrDiff $< -0.29051$ and $0.7075 <$ KMDrDiff $< 0.70949$ for $^{34}S$. Only the mass pairs that differed by the allowed KMDr difference are passed on to the fourth step of the isotope filtering.
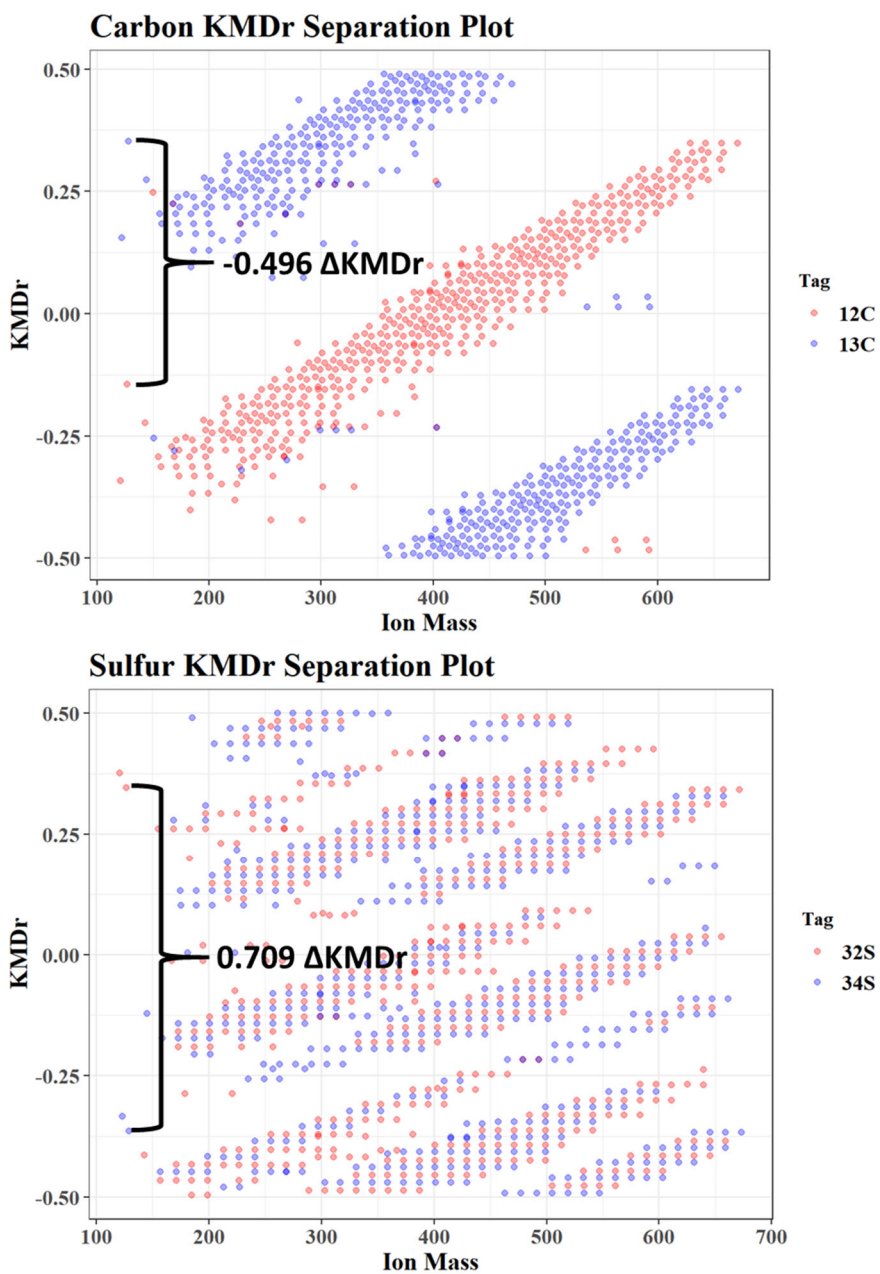
**Figure 5.10**. Demonstration of the effect of the resolution enhance KMD analysis used for identifying polyisotopic pairs. The top panel shows $^{12}C/^{13}C$, while the bottom panel shows $^{32}S/^{34}S$. The brackets represent one of the $\Delta$KMDr between a matched pair of isotopic masses. The other $\Delta$KMDr comes from the wrap-around of the points in the plot when the absolute value of KMDr exceeds 0.5. This wrapping is best visualized for the blue species in the $^{12}C/^{13}C$ panel.

*5.4.3.4 Step Four: Natural Isotopic Abundances*

For the reasons noted previously, the isotopic abundances are not always completely accurate for FT MS data so the ratio limits are somewhat relaxed. The theoretical $^{34}S/^{32}S$ intensity ratio is 4.2%, this ratio is kept the same across the entire spectrum because it is unlikely that there are > 1 sulfur atom present in the molecular formulas. However, since the abundance of sulfur containing molecular formulas is low, the sulfur ratio is set to 30% in IsoFiltR by default. Therefore, the $^{34}S$ peak abundance must not exceed 30% of the potential $^{32}S$ peak that it is matched to by default. It is set to this level because of the isotope dilution as shown in Figure 5.8, although the value can be changed by the user. $^{13}C$ on the other hand, cannot be set to a single ratio across the entire spectrum because the number of expected carbon atoms is too variable. Since the number is expected to increase with mass, the minimum and maximum $^{13}C/^{12}C$ ratio increases with mass. Specifically, a theoretical upper and lower limit of the $^{13}C/^{12}C$ ratio was calculated for 100 Da bins across the entire spectrum (100-200 Da, 200-300 Da, etc.). The upper limit was estimated by taking the mass at the upper edge of the bin and determining the maximum number of carbon atoms, without consideration to chemical feasibility. The isotopic ratio was determined using the sisweb.com isotope pattern generator. The abundance of the single and double $^{13}C$ peaks were used to set the limit on isotope ratio. The same concept was applied for the lower limit, using the maximum number of $CH_4O$ units at the lower limit of each mass bin. To further increase the flexibility of the abundance ranges, another parameter was included in the IsoFiltR function to loosen (lower value) or tighten (larger value) the carbon abundance windows; the default value

is 0.1. The pairs of masses that make it through this step of the QA are then finalized as monoisotopic and polyisotopic masses and are exported from the function as two separate lists. The monoisotopic masses are directly assigned molecular formulas using either MFAssign or MFAssignCHO and the polyisotopic masses are assigned molecular formulas by default when they are matched to the assigned monoisotopic formula. Occasionally, the IsoFiltR function incorrectly defines peaks as polyisotopic. For this reason, there is a secondary step within the MFAssign and MFAssignCHO functions to assign molecular formulas to the incorrect polyisotopic masses whenever possible.

### 5.4.4   IsoFiltR Test

To test the capability of IsoFiltR to accurately separate polyisotopic peaks and monoisotopic peaks the masses that had been defined as polyisotopic were assigned molecular formulas with MFAssign. Of the 1576 peaks flagged as potential $^{13}$C masses, 1269 were assigned a molecular formula, of which 1259 were assigned a $^{13}$C containing molecular formula. This suggests that the method does a reasonable job of isolating $^{13}$C polyisotopic peaks. Similarly, of the 549 masses that are flagged as $^{34}$S peaks, 459 are assigned a molecular formula, with only 11 being assigned as $^{34}$S molecular formulas. The reason why so few peaks are identified as $^{34}$S is that there is a relatively limited number of sulfur-containing species in the studied sample (PMO-2) and most of the sulfur peaks were present with low intensities, meaning that the isotope masses may be below the signal-to-noise threshold. The most abundant sulfur containing peaks from the molecular formulas $C_{16}H_{26}O_3S$ and $C_{17}H_{28}O_3S$ both have $^{34}$S peaks assigned, suggesting that the primary limitation is the low abundance of sulfur containing peaks. The peaks

215

that were assigned as $^{34}$S incorrectly can still be assigned a monoisotopic molecular formula due to a secondary formula extension in the MFAssign function which was described in Section 5.2.

## 5.5   Recalibration

### 5.5.1   Reason for Recalibration

In MFAssignR two functions were developed to do mass recalibration Recal() and Recal_2(). The goal of FT-MS is to obtain exact mass measurements. Despite the ultrahigh resolving power of the Orbitrap and FT-ICR mass spectrometers, the mass measurements can have mass shifts which can cause increased mass error when formula assignment is done. Some causes of this are random error from thermal and instrument noise (Kozhinov et al., 2013), space charge effects (Easterling et al., 1999; Wenger et al., 2011; Kozhinov et al., 2013), imperfections in the instrument hardware and electronics (Mathur et al., 2009), and issues with data conversion from the time to frequency domain (Gross et al., 2017). To address this issue the instrument and data must be calibrated. The first step is calibrating the instrument itself. This is done using a solution of compounds that are distinct from the analytes and is called external calibration. There are calibration solutions that can be purchased from the instrument manufacturers or other calibrant solutions such as arginine clusters can be used (Schmitt-Kopplin et al., 2010).  External calibration of the instrument before analysis is important to decreasing the mass error, but it is not always sufficient to mass error in the sub 1 ppm range, which is necessary for complex mixtures such as aerosol or (Schmitt-Kopplin et al., 2010; Wenger et al., 2011; Smith et al., 2018). To further improve the measurement accuracy an internal calibration

is performed on the acquired data after it has been collected. Classically, internal calibration involves spiking a sample with a known quantity of a known compound and then correcting the rest of the data based on the instrument response (Zhang et al., 2011). This is not always practical when analyzing complex mixtures such as environmental NOM due to possible interferences between the sample and the spiked calibrant (Zhang et al., 2011, Wenger et al., 2011). Spiking a solution can work, but it assumes that the mass error of the spiked calibrant is representative of the rest of the mass spectrum, which may not always be appropriate.

### 5.5.2   Methods for Mass Recalibration from Literature

Recently an iterative method was developed (Kozhinov et al., 2013) using a binomial coefficient weighted average to estimate the mass error correction term first across the whole spectrum. The coefficient is based on relatively few peaks and then within many small segments of the scan range as peaks are assigned molecular formulas in parallel. This eventually leads to a majority of the measured masses serving as recalibrants and eliminates the systematic error, leaving only the random error which cannot be corrected for. The systematic error comes from the instrument itself and is related to the electrical connections, physical imperfections in the instrument itself, and space charge effects. Since the recalibration uses the measured masses it is independent of the instrument platform (Kozhinov et al., 2013). This provides a good basis for the recalibration method in MFAssignR package. The Kozhinov method uses the concept of segmented recalibration previously described by Savory et al. (2011) and by Wong et al. (2006). This segmented recalibration approach works well because the measurement error with

respect to mass can change over the mass spectrum. This means that within relatively small windows the mass errors are more similar than the errors in other parts of the mass spectrum.

### 5.5.3 MFAssignR Mass Recalibration

In MFAssignR, we have implemented a recalibration function that is based on the methods described by Kozhinov et al. (2013) and Savory et al. (2011). The Kozhinov method is used to calculate the mass error function, which is used to correct the masses, and also the concept of segmented recalibration from Savory et al. (2011) is used to remove systematic bias. There are two versions of the recalibration function with a slight difference in how the recalibrants are chosen. In the first version, Recal(), users select an initial series of recalibrants series that are extended using molecular formula extensions based on the $H_2$ and O homologous. A user defined number of recalibrant peaks are selected based on relative abundance or intensity within the user defined mass range bins. Only the tallest peaks within each mass bin are used and then checked to see if there are any $^{13}C$ isotope peaks matched to them. If so, those $^{13}C$ peaks are also added to the recalibrants list. In the second version, Recal_2(), only the user defined recalibrants are used for the recalibration of the mass spectrum.

### *5.5.3.1 Recalibration Equations*

The following equations represent our implementation of the recalibration equation described in Kozhinov et al. (2013). The full recalibration method included more iterations of recalibration, and a secondary abundance-based term, which are not used in

218

MFAssignR. The following steps are required to calculate the mass error function. First, the binomial coefficient average is calculated to weight the mass error function. It is determined using the standard equation for a binomial coefficient (Kozhinov et al., 2013), shown in Equation 5.4:

$$C_j^k = \frac{(k)!}{(j)! \times (k-j)!} \quad \text{Eq. 5.4}$$

where $C_j^k$ is the binomial coefficient, $k$ is n minus 1, and $j$ is the row number from 0 to $k$, n is the total number of recalibrants. After the weights have been calculated for each of the recalibrants, the weighted mass error function is determined for each recalibrant using Equation 5.5:

$$\varepsilon_j = C_j^k \times \frac{E_{mass,j} - Th_{mass,j}}{Th_{mass,j}} \quad \text{Eq. 5.5}$$

where $\varepsilon_j$ is the error for a mass j, $E_{mass,j}$ is the experimental mass j, $Th_{mass,j}$, is the theoretical mass corresponding to the formula assigned to $E_{mass,j}$, and $C_j^k$ is the weight calculated in Equation 2. To determine the mass error function for the segment of the mass spectrum, use Equation 5.6:

$$\varepsilon = \frac{\sum \varepsilon_j}{2^k} \quad \text{Eq. 5.6}$$

where $\varepsilon$ is the mass error function used to recalibrate the spectrum. Mass recalibration of the measured masses is done using Equation 5.7:

$$mass_{recal} = \frac{mass}{1 + \varepsilon} \quad \text{Eq. 5.7}$$

where *mass* is the original measured mass, *mass$_{recal}$* is the recalibrated mass, and $\varepsilon$ is the mass error function.

### 5.5.3.2   Implementation of Recalibration

As opposed to recalibrating the entire spectrum at one time using a single mass error function a segmented approach is used adapted from Savory et al. (2011). To do this, the mass spectrum is divided into user defined segments and the mass error functions for each segment are calculated as described above and used to recalibrate the masses within the segment. This greatly increases the mass accuracy of the assignments by removing systematic bias. The recalibration requires at least three recalibrants in each segment, so any segment that does not have three recalibrants is recalibrated using the mass error function from the prior segment with three recalibrants. This commonly occurs at the high mass range as recalibrant peaks are more difficult to identify. This can lead to an increase in the recalibrated mass error because the systematic bias is not fully removed.

### 5.5.4   Recalibration Test

The overall effectiveness of this recalibration procedure is demonstrated by Figure 5.11, which compares the error plots for masses before and after recalibration. The error without recalibration follows trends that increase and decrease systematically. In the recalibrated plot, most of the points are in a flat line near 0-1 ppm mass error. While additional improvements are planned for this, it currently provides a noticeable improvement in the overall mass accuracy.
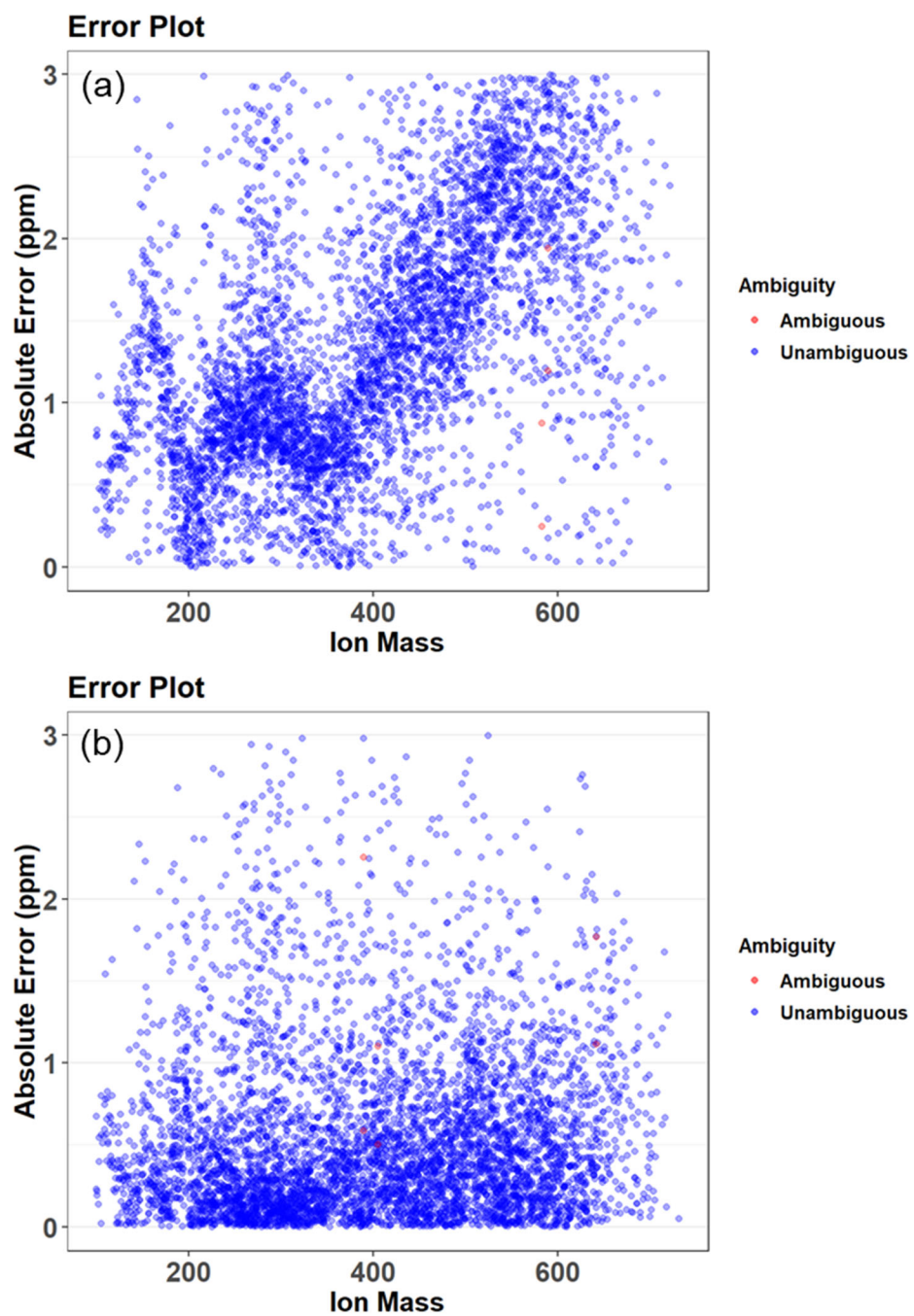
**Figure 5.11**. Showing the mass error for a sample of biomass burning aerosol before (Panel a) and after (Panel b) recalibration with the Recal() function.

## 5.6 Comparison of MFAssignR to Other Molecular Formula Assignment Methods

To demonstrate the effectiveness of MFAssignR as a method for molecular formula assignment relative to other available opensource (Formularity, UME) and commercial (Composer) software tools, the four methods were used to assign molecular formulas to an identical mass list. The mass list was the raw mass spectrum for the PMO-2 aerosol sample, which was discussed in previous chapters. Prior to molecular formula assignment with any of the methods, the noise level and mass recalibration were done using functions in the MFAssignR package. This ensured that all of the available masses for assignment were identical. It was necessary to use the functions from MFAssignR, because the other software packages do not have the capability to do noise estimation and mass recalibration conveniently. This is one of the major advantages of the MFAssignR package relative to the other available molecular formula assignment software packages. In each methods the assignments were limited to 3 nitrogen and 1 sulfur with a maximum absolute error of 1 ppm. The de novo cutoff was set to m/z 300 when possible (i.e. MFAssignR, Formularity, Composer). In UME the data were assigned molecular formulas using the NOM 4 CHNOSP database. The mass range limit of UME was m/z 100-700 (the entire range available), while the other methods had a range of m/z 100-1000.

The results of this comparison are in Table 5.3 and Figure 5.12. In general, the majority of masses assigned by MFAssignR and another method are assigned to the same molecular formula (Common). Only in a few cases are the molecular formulas different

between the methods (Different). MFAssignR generally assigns more molecular formulas to the spectrum than the other methods, meaning it can provide a more comprehensive picture of the molecular formula composition. UME is comparable within the available mass range, but because it is limited to m/z $\leq$ 700, MFAssignR assigns many more molecular formulas (n = 698 for this sample). Of particular interest is the much larger number of assignments from MFAssignR relative to Composer. Composer was used to do the molecular formula assignment for the samples discussed in the previous chapters, and after conservative QA steps, there were 2121 monoisotopic molecular formula assignments for PMO-2. In this case, with relatively limited QA, Composer was able to assign 4904 monoisotopic molecular formulas, which is still markedly fewer than that of MFAssignR (6489 assignments). MFAssignR does a better job of assigning molecular formulas to mass with low abundance, relative to the other methods, which accounts for most of the difference, as shown in Figure 5.12. Figure 5.13 shows the same type of plot as in Figure 5.12, but for the PMO-1 sample instead, demonstrating that this result is consistent.

The results highlight the reasons for developing our own method for formula assignment, because without MFAssignR we could not have known about the many additional molecular formulas that could be assigned. Overall, it seems that MFAssignR is very comparable to other established methods of formula assignment. This in itself is a valuable contribution, but when the other functions regarding noise estimation, isotope filtering, recalibration, and ability to assign a variety of heteroatoms and isotopes are considered, this package represents a uniquely powerful opensource tool for the data

preparation and molecular formula assignment for ultrahigh resolution mass spectrometry data.

**Table 5.3**. Number of formulas identified with each molecular formula assignment method (Total Assigned), number common with MFAssignR (Common), number of peaks that were assigned different molecular formula than MFAssignR (Different), number of masses that were only assigned with each non-MFAssignR method relative to MFAssignR (Unique), and the number of formulas assigned only by MFAssignR relative to the non-MFAssignR methods (Unique MFAssignR).

|  | UME | Formularity | Composer | MFAssignR |
|---|---|---|---|---|
| **Total Assigned** | 8955* | 5510 | 4904 | 6974 |
| **Common** | 5522 | 5247 | 4804 | X |
| **Different** | 187 | 137 | 19 | X |
| **Unique** | 538 | 70 | 81 | X |
| **Unique MFAssignR** | 1272 | 1590 | 2151 | X |

* The number of formulas assigned by UME includes masses that were assigned multiple molecular formulas.

**(a)** UME vs. MFAssignR, PMO-2

**(b)** Formularity vs. MFAssignR, PMO-2

**(c)** Composer vs. MFAssignR, PMO-2
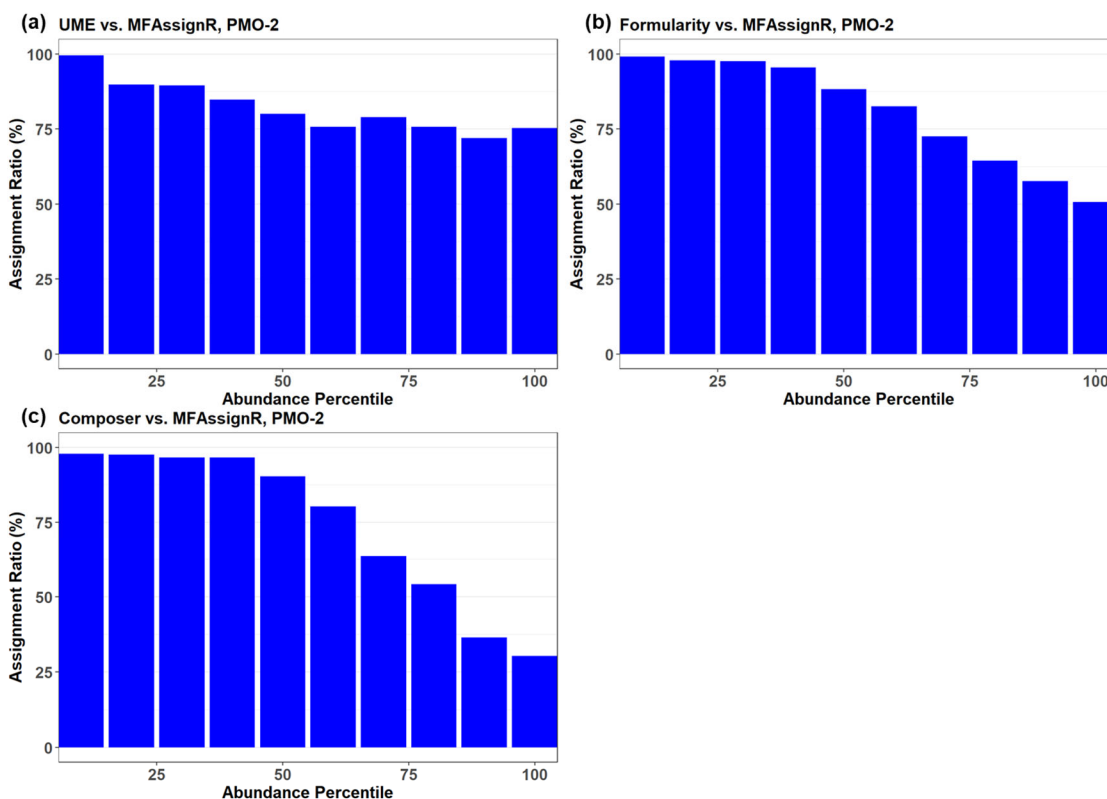
**Figure 5.12**. Comparison of the number of molecular formulas assigned based on the abundance of the peak using sample PMO-2. Panel a shows the comparison between UME and MFAssignR, Panel b shows the comparison between Formularity and MFAssignR, and Panel c shows the comparison between Composer and MFAssignR. Abundance Percentile breaks the abundance into a series of bins, the top 10% most abundant peaks are in Pecentile 10, the next 10% are Percentile 20, etc. If the assignment ratio is near 100 that means that the non-MFAssignR method assigned the same number of molecular formulas as MFAssignR within that abundance percentile. For example, in Class 1 all three methods are essentially equivalent to MFAssignR, however, for the least abundant peaks (Pecentile 100) Composer only assigns about 30% as many molecular formulas. This figure does not compare the molecular formula assignments to see if they are the same between each method, it only compares the number of assignments that are made within each abundance class.
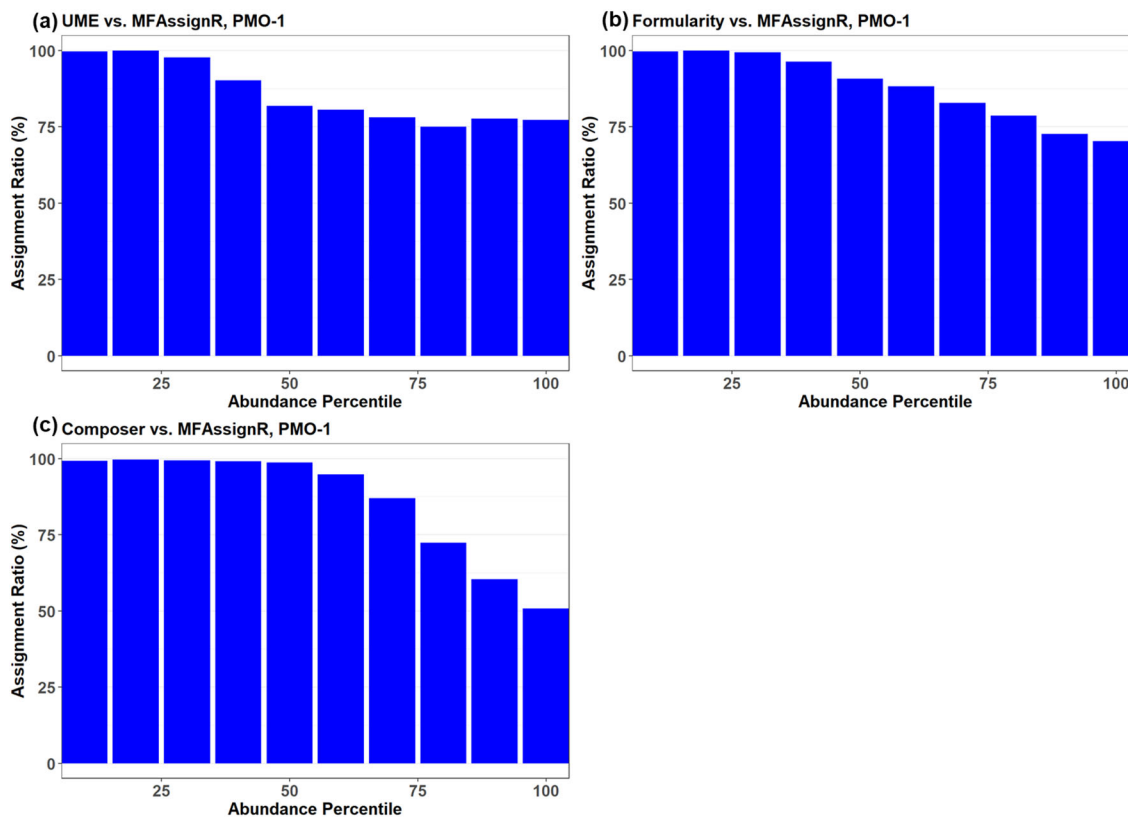
**Figure 5.13**. Comparison of the number of molecular formulas assigned based on the abundance of the peak using sample PMO-1. Panel a shows the comparison between UME and MFAssignR, Panel b shows the comparison between Formularity and MFAssignR, and Panel c shows the comparison between Composer and MFAssignR. Abundance Percentile breaks the abundance into a series of bins, the top 10% most abundant peaks are in Pecentile 10, the next 10% are Percentile 20, etc. If the assignment ratio is near 100 that means that the non-MFAssignR method assigned the same number of molecular formulas as MFAssignR within that abundance percentile. For example, in Class 1 all three methods are essentially equivalent to MFAssignR, however, for the least abundant peaks (Pecentile 100) Composer only assigns about 50% as many molecular formulas. This figure does not compare the molecular formula assignments to see if they are the same between each method, it only compares the number of assignments that are made within each abundance class.

# 6  Conclusion

## 6.1  Overview

The major focus of this dissertation was to provide a deeper understanding of the molecular formula composition of free tropospheric long range transported organic aerosol. This aerosol is of interest because it can provide insight into the multiphase processes that aerosol undergoes while in the atmosphere, especially with regards to the extent of oxidation that occurs in the atmosphere. The molecular formula composition aids physical property predictions including: viscosity, volatility, hygroscopicity, and light absorption. Each of these are important factors pertaining to the effect of aerosol on the climate system. In particular, we focused on samples that were influenced by wildfire events, which are a significant contributor to global aerosol loading. These large -scale biomass burning events also produce aerosol that is light absorbing. Thus, information about the aging of light absorbing species is important for improving predictions of their aerosol radiative effect on the climate.

Samples of organic aerosol were collected at the Pico Mountain Observatory in the Azores archipelago during the summers of 2013, 2014, and 2015. OC/EC and IC analyses were conducted to determine the bulk organic carbon and ion concentrations and ultrahigh resolution analysis FT-ICR MS was done to determine the molecular level composition of the samples. Two of the three samples were likely influenced by biomass burning, while one was likely anthropogenic in origin. The molecular formula analysis of these samples provided unique insight into their potential viscosity during transport. Highly viscous or solid aerosol transported in the free troposphere would likely have a

slower rate of oxidation and consequently more persistent light absorbing brown carbon

species from large-scale biomass burning aerosol. One of these samples was further

analyzed using tandem ultrahigh resolution FT-ICR MS/MS, which provided a more

detailed look at the molecular complexity of the aerosol. In addition; information about

the prevalence of specific functional groups for the molecular species was obtained. The

detailed analysis of functional groups is valuable for better model predictions of the

viscosity, volatility, hygroscopicity, and light absorption of organic aerosol.

Another focus of this dissertation was the development of software tools to streamline

and improve molecular formula assignment for ultrahigh resolution MS data collected

using either Orbitrap MS or FT-ICR MS. The software package containing these tools is

called MFAssignR. The code was written in the R programming language and was

released on GitHub. The functions apply methods drawn from the literature regarding the

best strategies for data preparation and molecular formula assignment. The goal of this

project was to produce a method for molecular formula assignment that was operationally

transparent and flexible with regard to multielement molecular formula assignments. As

the project progressed, it shifted towards a developing a full pipeline of functions for data

preparation and molecular formula assignment.

## 6.2  Long-Range Transported Aerosol Collected at the Pico Mountain Observatory

Aerosol samples collected on 27-28 June 2013 (PMO-1), 5-6 July 2014 (PMO-2), and

20-21 June 2015 (PMO-3) at the Pico Mountain Observatory were analyzed using

ultrahigh resolution FT-ICR mass spectrometry for molecular formula composition

228

determination. FLEXPART retroplumes for the sampled air masses indicated that: (a) PMO-1 and PMO-3 aerosol were transported predominantly through the free troposphere and were primarily influenced by wildfire emissions; and (b) PMO-2 aerosol were transported primarily through the boundary layer over the Northeast continental U.S. and the North Atlantic Ocean and was largely influenced by anthropogenic and biogenic sources. Although elevated levels of organic carbon, sulfate, and oxalate were found in all three samples, PMO-2 had the overall highest mass fractions of oxalate and sulfate indicating a clear influence of aqueous phase processing. The molecular formula assignments indicated differences in the aerosol oxidation rates between aerosol transported in the free troposphere (PMO-1 and PMO-3) and the boundary layer transported aerosol (PMO-2). These observations suggest that the transport pathways, in addition to the emission sources, contributed to the observed differences in the organic aerosol oxidation. The ambient temperature and RH at upwind times were extracted from the GFS analysis in FLEXPART and were used to estimate the glass transition temperatures of the aerosol species during transport. The results suggest that the organic aerosol components extracted from PMO-1 and PMO-3 were considerably more viscous due to lower RH than those from PMO-2 and thus were less susceptible to oxidation. The relationship between aerosol viscosity and its susceptibility to oxidation in the free troposphere is well supported (e.g., Koop et al., 2011; Berkemeier et al., 2014; Lignell et al., 2014; Shiraiwa et al., 2017a). These results suggest that biomass burning emissions and brown carbon injected into the free troposphere are more resistant to removal than aerosol transported in the boundary layer, due largely to the ambient temperature and relative humidity in the free troposphere. Although more work is needed to better

constrain the molecular composition of long-range transported aerosol and the processes that affect it during transport, the presented results have broader implications for the aging of long-range transported aerosol that is rapidly convected to the free troposphere.

## 6.3   Ultrahigh Resolution FT-ICR MS/MS Analysis of Free Tropospheric Organic Aerosol

The ultrahigh resolution FT-ICR MS/MS analysis of PMO-1 has provided novel functional group information using exact mass pairing of ions for free tropospheric organic aerosol. The most common losses were $CO_2$, $C_2H_4O$, and $H_2O$. The prevalence of $CO_2$ and $H_2O$ losses are consistent with a similar study of atmospheric organic matter in fog previously reported by LeClair et al. (2012). However, the loss of $C_2H_4O$ was not previously reported and likely represents a methyl carbonyl from a ketone near the end of a carbon chain. Ketones have been reported to be significant components of biomass burning aerosol according to the results of bulk methods (Hawkins and Russel, 2010), although they decrease with increased aging. Considering the transport time of PMO-1, the fraction of ketone functional groups are somewhat high, suggesting that the aging of aerosol proceeded more slowly than expected, potentially due to free tropospheric transport as hypothesized in Schum et al. (2018). The analysis of these functional groups is particularly important for modeling studies because different functional groups have different interactions with the environment (Clegg et al., 2019 in review). For example, hydroxyl and carboxyl groups are known to greatly increase the hygroscopicity of organic aerosol (Petters et al., 2017; Reid et al., 2018; Clegg et al., 2019 *in review*), which plays a significant role in the ability of an aerosol particle to act as a cloud

condensation nucleus (Massoli et al., 2010). Being able to constrain the fractions of these functional groups present in aerosol is important for better predictions of aerosol-water interactions in the atmosphere (Clegg et al., 2019, *in review*). The viscosity of organic aerosol can also be affected by the functional groups present in its component species. For example, carboxylic acids can lead to increased viscosity (Rothfuss and Petters, 2016; Song et al., 2016; Grayson et al., 2017; Reid et al., 2018). This makes the aerosol particles more solid and thus more resistant to oxidative processes in the atmosphere (Lignell et al., 2014; Hinks et al., 2016; Shrivastava et al., 2017; Schum et al., 2018). The carboxyl, hydroxyl, and ketone groups may also increase molecular absorption characteristics via change transfer reactions (Phillips and Smith, 2014; 2015). A comparison to aerosol collected at the Storm Peak Lab showed less correlation between the molecular formulas showing $H_2O$ and $CO_2$ losses than is observed in PMO-1. This suggests that the structural characteristics of those molecules are different. More hydroxyl groups are likely present in the regional biogenic SOA compared to PMO-1 where $H_2O$ losses were strongly correlated to the carboxyl losses. The investigation of nitrogen-containing neutral losses demonstrated the prevalence of nitro, nitrate, and amine groups in the aerosol molecular formulas. The amine groups were almost always observed in conjunction with a more polar/acidic functional group, highlighting their multifunctional nature. The nitro groups are consistent with the influence of biomass burning aerosol and particularly nitroaromatic compounds. Due to the ability of nitro groups to participate in charge transfer, it was hypothesized that these groups could contribute to longer wavelength light absorption of the aerosol analogous to the carboxyl, hydroxyl, and ketone groups. The neutral losses with sulfur demonstrated a surprising

prevalence of thiol groups in addition to the expected sulfate groups. The thiol groups were found to correlate well with more polar/acidic functional groups, which would allow them to be observed in negative ESI. The thiol groups were found to be more prevalent on aromatic molecules, suggesting that aromatic compounds with sulfur may be more likely reduced than oxidized. These results highlight the complexity of organic aerosol and the functional groups that are present in it. These results can be used to improve model predictions of aerosol and its interactions in the atmosphere.

## 6.4   MFAssignR

We decided to develop and in-house method for molecular formula assignment, called MFAssignR due to a lack of transparency and flexibility in the assignment of molecular formulas using a commercial software package. The core of the molecular formula assignment uses the CHOFIT algorithm developed by Green and Perdue (2015). The code was rewritten and expanded in the R programming language for our purposes. In order to improve its functionality, we added many quality assurance (QA) parameters to remove formulas that are likely incorrect and to decrease the ambiguity of formula assignments for a single mass. This is accomplished using formula extensions in a data dependent way to decrease the ambiguity without making *a priori* decisions about which molecular formulas be assigned. Despite adding parameters to decrease ambiguity in a data dependent way, we also leave the ambiguity that cannot be resolved by our methods so that the user can make the decision instead of the program, thereby providing more transparency. The ability to assign a variety of heteroatoms was incorporated, allowing novel analysis of isotopically labeled samples and chlorinated wastewater. In addition to

the molecular formula assignment, new methods for the estimation of the noise level in ultrahigh resolution MS and identification of isotope masses prior to molecular formula assignment were developed. The noise estimation method simplifies the estimation of noise, providing an improvement in accuracy over other noise estimation methods. This makes it a valuable contribution to the analysis of ultrahigh resolution MS data. An isotope identification function can be used to identify $^{13}$C and $^{34}$S isotopes and is important for decreasing the number of incorrect formula assignments. It was developed using well known mass relationships in a unique way providing a reasonably robust method to identify isotopic masses not based on molecular formula assignment, which is necessary for identification in other molecular formula assignment methods. The final component of MFAssignR is the mass recalibration functions. They were developed by expanding on previously reported methods (Savory et al., 2011; Kozhinov et al., 2014). The recalibration functions remove systematic bias to improve the mass accuracy to < 1 ppm. A comparison of the new MFAssignR package to pre-existing methods of molecular formula assignment found that for samples PMO-1 and PMO-2 MFAssignR generally assigns molecular formulas to ~1000 masses that are not assigned a molecular formula by the methods we compared it to (Formularity, UME, Composer). This is likely due to the extensive formula extension in MFAssignR, and the fact that two of the three methods tested used database searches, which limits the potential formulas to be assigned to some extent. In contrast, the other methods generally only assigned molecular formulas to ~100 masses that were not also assigned by MFAssignR. The exception was UME which had ~500 such assignments. Also, relatively few masses assigned a molecular formula by MFAssignR and one of the other methods were assigned different molecular

formulas. These numbers were presented in Table 5.3. Of particular interest is the comparison between MFAssignR and Composer, because the previous work (including Chapter 3 and 4 of this dissertation) was done with Composer. This indicates more information may be present in the mass spectra. Furthermore, the ability of MFAssignR functions to do noise estimation and recalibration is not available in UME and limited in Formularity (only recalibration) demonstrating improved functionality of MFAssignR. The transparency and comparability of MFAssignR to these other recently established methods of molecular formula assignment suggests that MFAssignR will be a valuable contribution to the field of ultrahigh resolution MS analysis.

## 6.5 Future Work

The work presented in this dissertation represents an important first step in providing information about the molecular formula composition of long range transported organic aerosol and molecular formula assignment methods. There are a few topics that deserve further study.

1. Further analysis of organic aerosol transported with the boundary layer and free troposphere is needed to investigate the oxidation and potential phase state of those samples. This would provide additional support for the hypothesis that free tropospheric transport can lead to lower oxidation due to the increased potential for solid state aerosol. Further analysis could confirm that aerosol transported in the free troposphere is markedly different than boundary layer aerosol, which is important for understanding the effect of aerosol on the climate.

2. More ultrahigh resolution MS/MS of organic aerosol is another important avenue of future research. As highlighted by comparison to biogenic SOA, further analysis of a variety of samples may provide valuable information regarding the molecular properties of organic aerosol. The use of PCA to do this analysis was valuable in highlighting the correlations, and should be used for future analysis as well. This information has implications for the aging, volatility, viscosity, and hygroscopicity of aerosol from different sources.

3. Further study regarding the potential for the nitro groups to contribute to light absorption via charge transfer is recommended. Few studies have directly studied the contribution of the nitro group to light absorption, so additional research, including computational predictions would be valuable to see how charge transfer works between the nitro, other functional groups, and the aromatic ring and if it could be related to increased light absorption. This would have an impact on radiative forcing predictions for biomass burning aerosol.

4. One additional avenue of work highlighted by this dissertation is continued improvements to the MFAssignR molecular formula assignment package. Further improvements for the isotope filtering and recalibration functions are recommended, in addition to continued improvements to the formula assignment itself, including the addition of more allowed elements for formula assignment.

# 7 References

Agarwal, S., Aggarwal, S. G., Okuzawa, K., and Kawamura, K.: Size distributions of dicarboxylic acids, ketoacids, α-dicarbonyls, sugars, WSOC, OC, EC and inorganic ions in atmospheric particles over Northern Japan: implication for long-range transport of Siberian biomass burning and East Asian polluted aerosols, Atmospheric Chemistry and Physics, 10, 5839-5858, 10.5194/acp-10-5839-2010, 2010.

Aggarwal, S. G., and Kawamura, K.: Carbonaceous and inorganic composition in long-range transported aerosols over northern Japan: Implication for aging of water-soluble organic fraction, Atmospheric Environment, 43, 2532-2540, 10.1016/j.atmosenv.2009.02.032, 2009.

Aiken, A. C., Decarlo, P. F., Kroll, J. H., Worsnop, D. R., Huffman, J. A., Docherty, K. S., Ulbrich, I. M., Mohr, C., Kimmel, J. R., Sueper, D., Sun, Y., Zhang, Q., Trimborn, A., Northway, M., Ziemann, P. J., Canagaratna, M. R., Onasch, T. B., Alfarra, M. R., Prevot, A. S. H., Dommen, J., Duplissy, J., Metzger, A., Baltensperger, U., and Jimenez, J. L.: O/C and OM/OC Ratios of Primary, Secondary, and Ambient Organic Aerosols with High-Resolution Time-of-Flight Aerosol Mass Spectrometry, Environmental Science & Technology, 42, 8, 10.1021/es703009q, 2008.

Altieri, K. E., Turpin, B. J., and Seitzinger, S. P.: Oligomers, organosulfates, and nitrooxy organosulfates in rainwater identified by ultra-high resolution electrospray ionization FT-ICR mass spectrometry, Atmospheric Chemistry and Physics, 9, 10, www.atmos-chem-phys.net/9/2533/2009/, 2009.

Andreae, M. O., and Gelencs´er, A.: Black carbon or brown carbon? The nature of light-absorbing carbonaceous aerosols, Atmospheric Chemistry and Physics, 6, 18, www.atmos-chem-phys.net/6/3131/2006/, 2006.

Bao, H., Niggemann, J., Luo, L., Dittmar, T., and Kao, S. J.: Aerosols as a source of dissolved black carbon to the ocean, Nat Commun, 8, 510, 10.1038/s41467-017-00437-3, 2017.

Belov, M. E., Zhang, R., Strittmatter, E. F., Prior, D. C., Tang, K., and Smith, R. D.: Automated Gain Control and Internal Calibration with External Ion Accumulation Capillary Liquid Chromatography-Electrospray Ionization-Fourier Transform Ion Cyclotron Resonance, Analytical Chemistry, 75, 11, 10.1021/ac0206770, 2003.

Berkemeier, T., Shiraiwa, M., Pöschl, U., and Koop, T.: Competition between water uptake and ice nucleation by glassy organic aerosol particles, Atmospheric Chemistry and Physics, 14, 12513-12531, 10.5194/acp-14-12513-2014, 2014.

Bertrand, A., Stefenelli, G., Jen, C. N., Pieber, S. M., Bruns, E. A., Ni, H., Temime-Roussel, B., Slowik, J. G., Goldstein, A. H., El Haddad, I., Baltensperger, U., Prévôt, A. S. H., Wortham, H., and Marchand, N.: Evolution of the chemical fingerprint of biomass burning organic aerosol during aging, Atmospheric Chemistry and Physics, 18, 7607-7624, 10.5194/acp-18-7607-2018, 2018.

Bignal, K. L., Langridge, S., and Zhou, J. L.: Release of polycyclic aromatic hydrocarbons, carbon monoxide and particulate matter from biomass combustion in a wood-fired boiler under varying boiler conditions, Atmospheric Environment, 42, 8863-8871, 10.1016/j.atmosenv.2008.09.013, 2008.

Blando, J. D., and Turpin, B. J.: Secondary organic aerosol formation in cloud and fog droplets: a literature evaluation of plausibility, Atmospheric Environment, 34, 10, 1 3 5 2 - 2 3 1 0 ( 9 9 ) 0 0 3 9 2 - 1, 2000.

Bougiatioti, A., Stavroulas, I., Kostenidou, E., Zarmpas, P., Theodosi, C., Kouvarakis, G., Canonaco, F., Prévôt, A. S. H., Nenes, A., Pandis, S. N., and Mihalopoulos, N.: Processing of biomass-burning aerosol in the eastern Mediterranean during summertime, Atmospheric Chemistry and Physics, 14, 4793-4807, 10.5194/acp-14-4793-2014, 2014.

Bougiatioti, A., Nikolaou, P., Stavroulas, I., Kouvarakis, G., Weber, R., Nenes, A., Kanakidou, M., and Mihalopoulos, N.: Particle water and pH in the eastern Mediterranean: source variability and implications for nutrient availability, Atmospheric Chemistry and Physics, 16, 4579-4591, 10.5194/acp-16-4579-2016, 2016.

Bowie, J. H.: The fragmentations of even-electron organic negative ions, Mass Spectometry Reviews, 9, 31, ccc 0277-7037/90/030349-3, 1990.

Brege, M., Paglione, M., Gilardoni, S., Decesari, S., Facchini, M. C., and Mazzoleni, L. R.: Molecular insights on aging and aqueous-phase processing from ambient biomass burning emissions-influenced Po Valley fog and aerosol, Atmospheric Chemistry and Physics, 18, 13197-13214, 10.5194/acp-18-13197-2018, 2018.

Brown, S. S., deGouw, J. A., Warneke, C., Ryerson, T. B., Dub´e, W. P., Atlas, E., Weber, R. J., Peltier, R. E., Neuman, J. A., Roberts, J. M., Swanson, A., Flocke, F., McKeen, S. A., Brioude, J., Sommariva, R., Trainer, M., Fehsenfeld, F. C., and Ravishankara, A. R.: Nocturnal isoprene oxidation over the Northeast United States in summer and its impact on reactive nitrogen partitioning and secondary organic aerosol, Atmospheric Chemistry and Physics, 9, 16, www.atmos-chem-phys.net/9/3027/2009/, 2009.

Buonanno, G., Morawska, L., and Stabile, L.: Particle emission factors during cooking activities, Atmospheric Environment, 43, 3235-3242, 10.1016/j.atmosenv.2009.03.044, 2009.

Burge, H. A.: An update on pollen and fungal spore aerobiology, Journal of Allergy and Clinical Immunology, 110, 544-552, 10.1067/mai.2002.128674, 2002.

Cao, F., Zhang, S. C., Kawamura, K., Liu, X., Yang, C., Xu, Z., Fan, M., Zhang, W., Bao, M., Chang, Y., Song, W., Liu, S., Lee, X., Li, J., Zhang, G., and Zhang, Y. L.: Chemical characteristics of dicarboxylic acids and related organic compounds in PM2.5 during biomass-burning and non-biomass-burning seasons at a rural site of Northeast China, Environ Pollut, 231, 654-662, 10.1016/j.envpol.2017.08.045, 2017.

Capes, G., Johnson, B., McFiggans, G., Williams, P. I., Haywood, J., and Coe, H.: Aging of biomass burning aerosols over West Africa: Aircraft measurements of chemical composition, microphysical properties, and emission ratios, Journal of Geophysical Research, 113, 10.1029/2008jd009845, 2008.

Carlton, A. G., Turpin, B. J., Altieri, K. E., Seitzinger, S., Reff, A., Lim, H.-J., and Ervens, B.: Atmospheric oxalic acid and SOA production from glyoxal: Results of aqueous photooxidation experiments, Atmospheric Environment, 41, 7588-7602, 10.1016/j.atmosenv.2007.05.035, 2007.

Cech, N. B., and Enke, C. G.: Practical implications of some recent studies in electrospray ionization fundamentals, Mass Spectrom Rev, 20, 362-387, 10.1002/mas.10008, 2001.

Chakrabarty, R. K., Moosmüller, H., Chen, L. W. A., Lewis, K., Arnott, W. P., Mazzoleni, C., Dubey, M. K., Wold, C. E., Hao, W. M., and Kreidenweis, S. M.: Brown carbon in tar balls from smoldering biomass combustion, Atmospheric Chemistry and Physics, 10, 6363-6370, 10.5194/acp-10-6363-2010, 2010.

Chang, O., Procedure for the analysis of particulate anions and cations in motor vehicle exhaust by ion chromatography, California EPA Air Resources Board, 2011.

Chen, Z. M., Wang, H. L., Zhu, L. H., Wang, C. X., Jie, C. Y., and Hua, W.: Aqueous-phase ozonolysis of methacrolein and methyl vinyl ketone: a potentially important source of atmospheric aqueous oxidants, Atmospheric Chemistry and Physics, 8, 11, www.atmos-chem-phys.net/8/2255/2008/, 2008.

Cheng, M. T., and Tsai, Y. I.: Characterization of visibility and atmospheric aerosols in urban, suburban, and remote areas, The Science of the Total Environment, 263, 14, S 0 0 4 8 - 9 6 9 7 0 0 00670-7, 2000.

China, S., Scarnato, B., Owen, R. C., Zhang, B., Ampadu, M. T., Kumar, S., Džepina, K., Dziobak, M. P., Fialho, P., Perlinger, J. A., Hueber, J., Helmig, D., Mazzoleni, L. R., and Mazzoleni, C.: Morphology and mixing state of aged soot particles at a remote marine free troposphere site: Implications for optical properties, Geophysical Research Letters, 42, 1243-1250, 10.1002/2014gl062404, 2015.

China, S., Alpert, P. A., Zhang, B., Schum, S., Džepina, K., Wright, K., Owen, R. C., Fialho, P., Mazzoleni, L. R., Mazzoleni, C., and Knopf, D. A.: Ice cloud formation potential by free tropospheric particles from long-range transport over the Northern Atlantic Ocean, Journal of Geophysical Research: Atmospheres, 122, 3065-3079, 10.1002/2016jd025817, 2017.

Chiou, P., Tang, W., Lin, C.-J., Chu, H.-W., and Ho, T. C.: Atmospheric Aerosol over a Southeastern Region of Texas: Chemical Composition and Possible Sources, Environmental Modeling & Assessment, 14, 333-350, 10.1007/s10666-007-9120-8, 2007.

Chow, J. C., Lowenthal, D. H., Chen, L. W., Wang, X., and Watson, J. G.: Mass reconstruction methods for PM2.5: a review, Air Qual Atmos Health, 8, 243-263, 10.1007/s11869-015-0338-3, 2015.

Clegg, S. L., Kleeman, M. J., Griffin, R. J., and Seinfeld, J. H.: Effects of uncertainties in the thermodynamic properties of aerosol components in an air quality model – Part 2: Predictions of the vapour pressures of organic compounds, Atmospheric Chemistry and Physics, 8, 17, www.atmos-chem-phys.net/8/1087/2008/, 2008.

Clegg, S. L., Mazzoleni, L. R., Samburova, V., Taylor, N. F., Collins, D. R., Schum, S. K., and Hallar, A. G.: Modelling the hygroscopic growth factors of aerosol material containing a large water-soluble organic fraction, collected at the Storm Peak Laboratory, Atmospheric Environment, 2019.

Collaud Coen, M., Andrews, E., Aliaga, D., Andrade, M., Angelov, H., Bukowiecki, N., Ealo, M., Fialho, P., Flentje, H., Hallar, A. G., Hooda, R., Kalapov, I., Krejci, R., Lin, N.-H., Marinoni, A., Ming, J., Nguyen, N. A., Pandolfi, M., Pont, V., Ries, L., Rodríguez, S., Schauer, G., Sellegri, K., Sharma, S., Sun, J., Tunved, P., Velasquez, P., and Ruffieux, D.: The topography contribution to the influence of the atmospheric boundary layer at high altitude stations, Atmospheric Chemistry and Physics Discussions, 1-44, 10.5194/acp-2017-692, 2017.

Collaud Coen, M., Andrews, E., Aliaga, D., Andrade, M., Angelov, H., Bukowiecki, N., Ealo, M., Fialho, P., Flentje, H., Hallar, A. G., Hooda, R., Kalapov, I., Krejci, R., Lin, N.-H., Marinoni, A., Ming, J., Nguyen, N. A., Pandolfi, M., Pont, V., Ries, L., Rodríguez, S., Schauer, G., Sellegri, K., Sharma, S., Sun, J., Tunved, P., Velasquez, P., and Ruffieux, D.: Identification of topographic features influencing aerosol observations at high altitude stations, Atmospheric Chemistry and Physics, 18, 12289-12313, 10.5194/acp-18-12289-2018, 2018.

Comisarow, M. B., and Marshall, A. G.: Fourier transform ion cyclotron resonance spectroscopy, Chemical Physics Letters, 25, 2, https://doi.org/10.1016/0009-2614(74)89137-2, 1974.

Cook, R. D., Lin, Y.-H., Peng, Z., Boone, E., Chu, R. K., Dukett, J. E., Gunsch, M. J., Zhang, W., Tolic, N., Laskin, A., and Pratt, K. A.: Biogenic, urban, and wildfire influences on the molecular composition of dissolved organic compounds in cloud water, Atmospheric Chemistry and Physics, 17, 15167-15180, 10.5194/acp-17-15167-2017, 2017.

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

Cote, C. D., Schneider, S. R., Lyu, M., Gao, S., Gan, L., Holod, A. J., Chou, T. H. H., and Styler, S. A.: Photochemical Production of Singlet Oxygen by Urban Road Dust, Environmental Science & Technology Letters, 5, 92-97, 10.1021/acs.estlett.7b00533, 2018.

Crahan, K. K., Hegg, D., Covert, D. S., and Jonsson, H.: An exploration of aqueous oxalic acid production in the coastal marine atmosphere, Atmospheric Environment, 38, 3757-3764, 10.1016/j.atmosenv.2004.04.009, 2004.

Cruz, C. N., and Pandis, S. N.: Deliquescence and Hygroscopic Growth of Mixed Inorganic-Organic Atmospheric Aerosol, Environmental Science & Technology, 34, 7, 2000.

D'Almeida, G. A., and Schutz, L.: Number Mass and Volume Distribution of Mineral Aerosol and Soils of the Sahara, Journal of Climate and Applied Meteorology, 22, 11, 1983.

Damoah, R., Spichtinger, N., Forster, C., James, P., Mattis, I., Wandinger, U., Beirle, S., Wagner, T., and Stohl, A.: Around the world in 17 days – hemispheric-scale transport of forest fire smoke from Russia in May 2003, Atmospheric Chemistry and Physics, 4, 11, 1680-7324/acp/2004-4-1311, 2004.

De Haan, D. O., Hawkins, L. N., Welsh, H. G., Pednekar, R., Casar, J. R., Pennington, E. A., de Loera, A., Jimenez, N. G., Symons, M. A., Zauscher, M., Pajunoja, A., Caponi, L., Cazaunau, M., Formenti, P., Gratien, A., Pangui, E., and Doussin, J. F.: Brown Carbon Production in Ammonium- or Amine-Containing Aerosol Particles by Reactive Uptake of Methylglyoxal and Photolytic Cloud Cycling, Environ Sci Technol, 51, 7458-7466, 10.1021/acs.est.7b00159, 2017.

de Hoffman, E., and Stroobant, V.: Mass Spectrometry Principles and Applications 3rd Edition, Wiley, 2007.

Decesari, S., Facchini, M. C., Fuzzi, S., and Tagliavini, E.: Characterization of water-soluble organic compounds in atmospheric aerosol: A new approach, Journal of Geophysical Research: Atmospheres, 105, 1481-1489, 10.1029/1999jd900950, 2000.

Decesari, S., Facchini, M. C., Matta, E., Mircea, M., Fuzzi, S., Chughtai, A. R., and Smith, D. M.: Water soluble organic compounds formed by oxidation of soot, Atmospheric Environment, 36, 6, S 1352 -2310(02 )0 0141-3, 2002.

Decesari, S., Mircea, M., Cavalli, F., Fuzzi, S., Moretti, F., and Tagliavini, E.: Source Attribution of Water-Soluble Organic Aerosol by Nuclear Magnetic Resonance Spectroscopy, Environmental Science & Technology, 41, 6, 10.1021/es0617111 2007.

DeRieux, W.-S. W., Li, Y., Lin, P., Laskin, J., Laskin, A., Bertram, A. K., Nizkorodov, S. A., and Shiraiwa, M.: Predicting the glass transition temperature and viscosity of secondary organic material using molecular composition, Atmospheric Chemistry and Physics, 18, 6331-6351, 10.5194/acp-18-6331-2018, 2018.

Desyaterik, Y., Sun, Y., Shen, X., Lee, T., Wang, X., Wang, T., and Collett, J. L.: Speciation of "brown" carbon in cloud water impacted by agricultural biomass burning in eastern China, Journal of Geophysical Research: Atmospheres, 118, 7389-7399, 10.1002/jgrd.50561, 2013.

Dirksen, R. J., Folkert Boersma, K., de Laat, J., Stammes, P., van der Werf, G. R., Val Martin, M., and Kelder, H. M.: An aerosol boomerang: Rapid around-the-world transport of smoke from the December 2006 Australian forest fires observed from space, Journal of Geophysical Research, 114, 10.1029/2009jd012360, 2009.

Dockery, D. W., Cunningham, J., Damokosh, A. L., Neas, L. M., Spengler, J. D., Koutrakis, P., Ware, J. H., Raizenne, M., and Speizer, F. E.: Health Effects of Acid Aerosols on North American Children: Respiratory Symptoms, Environmental Health Perspectives, 104, 6, 1996.

Donahue, N. M., Epstein, S. A., Pandis, S. N., and Robinson, A. L.: A two-dimensional volatility basis set: 1. organic-aerosol mixing thermodynamics, Atmospheric Chemistry and Physics, 11, 3303-3318, 10.5194/acp-11-3303-2011, 2011.

Duan, F., Liu, X., Yu, T., and Cachier, H.: Identification and estimate of biomass burning contribution to the urban aerosol organic carbon concentrations in Beijing, Atmospheric Environment, 38, 1275-1282, 10.1016/j.atmosenv.2003.11.037, 2004.

Duarte, R. M. B. O., Pio, C. A., and Duarte, A. C.: Spectroscopic study of the water-soluble organic matter isolated from atmospheric aerosols collected under

different atmospheric conditions, Analytica Chimica Acta, 530, 7-14, 10.1016/j.aca.2004.08.049, 2005.

Džepina, K., Mazzoleni, C., Fialho, P., China, S., Zhang, B., Owen, R. C., Helmig, D., Hueber, J., Kumar, S., Perlinger, J. A., Kramer, L. J., Dziobak, M. P., Ampadu, M. T., Olsen, S., Wuebbles, D. J., and Mazzoleni, L. R.: Molecular characterization of free tropospheric aerosol collected at the Pico Mountain Observatory: a case study with a long-range transported biomass burning plume, Atmospheric Chemistry and Physics, 15, 5047-5068, 10.5194/acp-15-5047-2015, 2015.

Edwards, S. M.: lemon: Freshing Up your 'ggplot2' Plots. R package version 0.4.1. https://CRAN.R-project.org/package=lemon, 2018.

Ehn, M., Thornton, J. A., Kleist, E., Sipila, M., Junninen, H., Pullinen, I., Springer, M., Rubach, F., Tillmann, R., Lee, B., Lopez-Hilfiker, F., Andres, S., Acir, I. H., Rissanen, M., Jokinen, T., Schobesberger, S., Kangasluoma, J., Kontkanen, J., Nieminen, T., Kurten, T., Nielsen, L. B., Jorgensen, S., Kjaergaard, H. G., Canagaratna, M., Maso, M. D., Berndt, T., Petaja, T., Wahner, A., Kerminen, V. M., Kulmala, M., Worsnop, D. R., Wildt, J., and Mentel, T. F.: A large source of low-volatility secondary organic aerosol, Nature, 506, 476-479, 10.1038/nature13032, 2014.

El-Zanan, H. S., Lowenthal, D. H., Zielinska, B., Chow, J. C. and Kumar, N.: Determination of the organic aerosol mass to organic carbon ratio in IMPROVE samples, Chemosphere, 60(4), 485–496, doi:10.1016/j.chemosphere.2005.01.005, 2005.

Engelbrecht, J. P., Moosmüller, H., Pincock, S., Jayanty, R. K. M., Lersch, T., and Casuccio, G.: Technical note: Mineralogical, chemical, morphological, and optical interrelationships of mineral dust re-suspensions, Atmospheric Chemistry and Physics, 16, 10809-10830, 10.5194/acp-16-10809-2016, 2016.

Ervens, B., Carlton, A. G., Turpin, B. J., Altieri, K. E., Kreidenweis, S. M., and Feingold, G.: Secondary organic aerosol yields from cloud-processing of isoprene oxidation products, Geophysical Research Letters, 35, 10.1029/2007gl031828, 2008.

Ervens, B., Turpin, B. J., and Weber, R. J.: Secondary organic aerosol formation in cloud droplets and aqueous particles (aqSOA): a review of laboratory, field and model studies, Atmospheric Chemistry and Physics, 11, 11069-11102, 10.5194/acp-11-11069-2011, 2011.

Ferman, M. A., Wolff, G. T., and Kelly, N. A.: The Nature and Sources of Haze in the Shenandoah Valley/Blue Ridge Mountains Area, Journal of the Air Pollution Control Association, 31, 1074-1082, 10.1080/00022470.1981.10465329, 1981.

Fialho, P., Hansen, A. D. A., and Honrath, R. E.: Absorption coefficients by aerosols in remote areas: a new approach to decouple dust and black carbon absorption coefficients using seven-wavelength Aethalometer data, Journal of Aerosol Science, 36, 267-282, 10.1016/j.jaerosci.2004.09.004, 2005.

Fine, P. M., Cass, G. R., and Simoneit, B. R. T.: Chemical Characterization of Fine Particle Emissions from the Wood Stove Combustion of Prevalent United States Tree Species, Environmental Engineering Science, 21, 17, 2004.

Forrister, H., Liu, J., Scheuer, E., Dibb, J., Ziemba, L., Thornhill, K. L., Anderson, B., Diskin, G., Perring, A. E., Schwarz, J. P., Campuzano-Jost, P., Day, D. A., Palm, B. B., Jimenez, J. L., Nenes, A., and Weber, R. J.: Evolution of brown carbon in wildfire plumes, Geophysical Research Letters, 42, 4623-4630, 10.1002/2015gl063897, 2015.

Fouquet, T., and Sato, H.: Extension of the Kendrick Mass Defect Analysis of Homopolymers to Low Resolution and High Mass Range Mass Spectra Using Fractional Base Units, Anal Chem, 89, 2682-2686, 10.1021/acs.analchem.6b05136, 2017.

Franck, U., Odeh, S., Wiedensohler, A., Wehner, B., and Herbarth, O.: The effect of particle size on cardiovascular disorders--the smaller the worse, Sci Total Environ, 409, 4217-4221, 10.1016/j.scitotenv.2011.05.049, 2011.

Gaspar, A., and Schrader, W.: Expanding the data depth for the analysis of complex crude oil samples by Fourier transform ion cyclotron resonance mass spectrometry using the spectral stitching method, Rapid Commun Mass Spectrom, 26, 1047-1052, 10.1002/rcm.6200, 2012.

Ge, X., Wexler, A. S., and Clegg, S. L.: Atmospheric amines – Part I. A review, Atmospheric Environment, 45, 524-546, 10.1016/j.atmosenv.2010.10.012, 2011.

Gelencser, A., Hoffer, A., Kiss, G., Tombacz, E., Kurdi, R., and Bencze, L.: In-situ Formation of Light-Absorbing Organic Matter in Cloud Water, Journal of Atmospheric Chemistry, 45, 9, 2003.

Gordon, M., and Taylor, J. S.: Ideal copolymers and the second-order transitions of synthetic rubber. i. non-crystalline copolymers, Journal of chemical technology & biotechnology, 2, 8, 1952.

Gorshkov, M. V., Good, D. M., Lyutvinskiy, Y., Yang, H., and Zubarev, R. A.: Calibration function for the Orbitrap FTMS accounting for the space charge effect, J Am Soc Mass Spectrom, 21, 1846-1851, 10.1016/j.jasms.2010.06.021, 2010.

Grayson, J. W., Evoy, E., Song, M., Chu, Y., Maclean, A., Nguyen, A., Upshur, M. A., Ebrahimi, M., Chan, C. K., Geiger, F. M., Thomson, R. J., and Bertram, A. K.: The effect of hydroxyl functional groups and molar mass on the viscosity of non-crystalline organic and organic–water particles, Atmospheric Chemistry and Physics, 17, 8509-8524, 10.5194/acp-17-8509-2017, 2017.

Green, N. W., and Perdue, E. M.: Fast graphically inspired algorithm for assignment of molecular formulae in ultrahigh resolution mass spectrometry, Anal Chem, 87, 5086-5094, 10.1021/ac504166t, 2015.

Griffin, R. J., Cocker, D. R., Flagan, R. C., and Seinfeld, J. H.: Organic aerosol formation from the oxidation of biogenic hydrocarbons, Journal of Geophysical Research: Atmospheres, 104, 3555-3567, 10.1029/1998jd100049, 1999.

Grosjean, D.: In Situ Organic Aerosol Formation During a Smog Episode: Estimated Production and Chemical Functionality, Atmospheric Environment, 26, 11, 1992.

Gross, J. H.: Mass Spectrometry A Textbook, 10.1007/978-3-319-54398-7, 2017.

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, T. F., Monod, A., Pr´ev^ot, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, Atmospheric Chemistry and Physics, 9, 82, www.atmos-chem-phys.net/9/5155/2009/, 2009.

Hawkes, J. A., Dittmar, T., Patriarca, C., Tranvik, L., and Bergquist, J.: Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter, Anal Chem, 88, 7698-7704, 10.1021/acs.analchem.6b01624, 2016.

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

Hawkins, L. N., Lemire, A. N., Galloway, M. M., Corrigan, A. L., Turley, J. J., Espelien, B. M., and De Haan, D. O.: Maillard Chemistry in Clouds and Aqueous Aerosol As a Source of Atmospheric Humic-Like Substances, Environ Sci Technol, 50, 7443-7452, 10.1021/acs.est.6b00909, 2016.

Helmig, D., Muñoz, M., Hueber, J., Mazzoleni, C., Mazzoleni, L., Owen, R. C., Val-Martin, M., Fialho, P., Plass-Duelmer, C., Palmer, P. I., Lewis, A. C., and Pfister, G.: Climatology and atmospheric chemistry of the non-methane hydrocarbons

ethane and propane over the North Atlantic, Elementa: Science of the Anthropocene, 3, 10.12952/journal.elementa.000054, 2015.

Hertkorn, N., Frommberger, M., Witt, M., Koch, B. P., Schmitt-Kopplin, P., and Perdue, E. M.: Natural Organic Matter and the Event Horizon of Mass Spectrometry, Analytical Chemistry, 80, 12, 10.1021/ac800464g, 2008.

Herzsprung, P., Hertkorn, N., von Tumpling, W., Harir, M., Friese, K., and Schmitt-Kopplin, P.: Understanding molecular formula assignment of Fourier transform ion cyclotron resonance mass spectrometry data of natural organic matter from a chemical point of view, Anal Bioanal Chem, 406, 7977-7987, 10.1007/s00216-014-8249-y, 2014.

Hinks, M. L., Brady, M. V., Lignell, H., Song, M., Grayson, J. W., Bertram, A. K., Lin, P., Laskin, A., Laskin, J., and Nizkorodov, S. A.: Effect of viscosity on photodegradation rates in complex secondary organic aerosol materials, Phys Chem Chem Phys, 18, 8785-8793, 10.1039/c5cp05226b, 2016.

Hopkins, R. J., Lewis, K., Desyaterik, Y., Wang, Z., Tivanski, A. V., Arnott, W. P., Laskin, A., and Gilles, M. K.: Correlations between optical, chemical and physical properties of biomass burn aerosols, Geophysical Research Letters, 34, 10.1029/2007gl030502, 2007.

Hoyle, C. R., Fuchs, C., Järvinen, E., Saathoff, H., Dias, A., El Haddad, I., Gysel, M., Coburn, S. C., Tröstl, J., Bernhammer, A. K., Bianchi, F., Breitenlechner, M., Corbin, J. C., Craven, J., Donahue, N. M., Duplissy, J., Ehrhart, S., Frege, C., Gordon, H., Höppel, N., Heinritzi, M., Kristensen, T. B., Molteni, U., Nichman, L., Pinterich, T., Prévôt, A. S. H., Simon, M., Slowik, J. G., Steiner, G., Tomé, A., Vogel, A. L., Volkamer, R., Wagner, A. C., Wagner, R., Wexler, A. S., Williamson, C., Winkler, P. M., Yan, C., Amorim, A., Dommen, J., Curtius, J., Gallagher, M. W., Flagan, R. C., Hansel, A., Kirkby, J., Kulmala, M., Möhler, O., Stratmann, F., Worsnop, D. R., and Baltensperger, U.: Aqueous phase oxidation of sulphur dioxide by ozone in cloud droplets, Atmospheric Chemistry and Physics, 16, 1693-1712, 10.5194/acp-16-1693-2016, 2016.

Hsu, C. S., Liang, Z., and Campana, J. E.: Hydrocarbon Characterization by Ultrahigh Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*, Analytical Chemistry, 66, 6, 0003-2700/94/0366-0850, 1994.

Huang, R. J., Zhang, Y., Bozzetti, C., Ho, K. F., Cao, J. J., Han, Y., Daellenbach, K. R., Slowik, J. G., Platt, S. M., Canonaco, F., Zotter, P., Wolf, R., Pieber, S. M., Bruns, E. A., Crippa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., An, Z., Szidat, S., Baltensperger, U., El Haddad, I., and Prevot, A. S.: High secondary aerosol contribution to particulate pollution during haze events in China, Nature, 514, 218-222, 10.1038/nature13774, 2014.

Iinuma, Y., Boge, O., Grafe, R., and Herrmann, H.: Methyl-Nitrocatechols: Atmospheric Tracer Compounds for Biomass Burning Secondary Organic Aerosols, Environmental Science & Technology, 44, 7, 10.1021/es102938a 2010.

IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jang, M., Czoschke, N. M., Lee, S., and Kamens, R. M.: Heterogeneous Atmospheric Aerosol Production by AcidCatalyzed Particle-Phase Reactions, Science, 298, 4, 2002.

Jensen, N. J., Tomer, K. B., and Gross, M. L.: Gas-phase ion decomposition occurring remote to a charge site, Journal of the American Chemical Society, 107, 1863-1868, 10.1021/ja00293a010, 1985.

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn , M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, E. J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Džepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of Organic Aerosols in the Atmosphere, Science, 326, 5, 10.1126/science.1180353, 2009.

Jokinen, T., Sipila, M., Richters, S., Kerminen, V. M., Paasonen, P., Stratmann, F., Worsnop, D., Kulmala, M., Ehn, M., Herrmann, H., and Berndt, T.: Rapid autoxidation forms highly oxidized RO2 radicals in the atmosphere, Angew Chem Int Ed Engl, 53, 14596-14600, 10.1002/anie.201408566, 2014.

Kahn, R. A., Chen, Y., Nelson, D. L., Leung, F.-Y., Li, Q., Diner, D. J., and Logan, J. A.: Wildfire smoke injection heights: Two perspectives from space, Geophysical Research Letters, 35, 10.1029/2007gl032165, 2008.

Kahnt, A., Behrouzi, S., Vermeylen, R., Safi Shalamzari, M., Vercauteren, J., Roekens, E., Claeys, M., and Maenhaut, W.: One-year study of nitro-organic compounds and their relation to wood burning in PM10 aerosol from a rural site in Belgium, Atmospheric Environment, 81, 561-568, 10.1016/j.atmosenv.2013.09.041, 2013.

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J. J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527-554, 10.5194/bg-9-527-2012, 2012.

Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. E. L., Tsigaridis, K., Vignati, E., Stephanou, E. G., and Wilson, J.: Organic aerosol and global climate modelling: a review, Atmospheric Chemistry and Physics, 5, 71, 1680-7324/acp/2005-5-1053, 2005.

Keitt, T.: colorRamps: Builds color tables. R package version 2.3. https://CRAN.R-project.org/package=colorRamps, 2012

Kerwin, J. L., Wiens, A. M., and Ericsson, L. H.: Identification of Fatty Acids by Electrospray Mass Spectrometry and Tandem Mass Spectrometry, Journal of Mass Spectrometry, 31, 9, CCC 1076-5 174/96/020184-09, 1996.

Kessler, S. H., Smith, J. D., Che, D. L., Worsnop, D. R., Wilson, K. R., and Kroll, J. H.: Chemical Sinks of Organic Aerosol: Kinetics and Products of the Heterogeneous Oxidation of Erythritol and Levoglucosan, Environmental Science & Technology, 44, 6, 10.1021/es101465m 2010.

Kew, W., Goodall, I., Clarke, D., and Uhrin, D.: Chemical Diversity and Complexity of Scotch Whisky as Revealed by High-Resolution Mass Spectrometry, J Am Soc Mass Spectrom, 28, 200-213, 10.1007/s13361-016-1513-y, 2017.

Kilgour, D. P. A., Hughes, S., Kilgour, S. L., Mackay, C. L., Palmblad, M., Tran, B. Q., Goo, Y. A., Ernst, R. K., Clarke, D. J., and Goodlett, D. R.: Autopiquer - a Robust and Reliable Peak Detection Algorithm for Mass Spectrometry, J Am Soc Mass Spectrom, 28, 253-262, 10.1007/s13361-016-1549-z, 2017.

Kind, T., and Fiehn, O.: Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC Bioinformatics, 8, 105, 10.1186/1471-2105-8-105, 2007.

Kirpes, R. M., Bondy, A. L., Bonanno, D., Moffet, R. C., Wang, B., Laskin, A., Ault, A. P., and Pratt, K. A.: Secondary Sulfate is Internally Mixed with Sea Spray Aerosol and Organic Aerosol in the Winter-Spring Arctic, Atmospheric Chemistry and Physics Discussions, 1-29, 10.5194/acp-2017-998, 2017.

Kleissl, J., Honrath, R. E., Dziobak, M. P., Tanner, D., Val Martín, M., Owen, R. C., and Helmig, D.: Occurrence of upslope flows at the Pico mountaintop observatory: A

case study of orographic flows on a small, volcanic island, Journal of Geophysical Research: Atmospheres, 112, 10.1029/2006jd007565, 2007.

Koch, B. P., and Dittmar, T.: From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter, Rapid Communications in Mass Spectrometry, 20, 926-932, 10.1002/rcm.2386, 2006.

Koop, T., Bookhold, J., Shiraiwa, M., and Poschl, U.: Glass transition and phase state of organic compounds: dependency on molecular properties and implications for secondary organic aerosols in the atmosphere, Phys Chem Chem Phys, 13, 19238-19255, 10.1039/c1cp22617g, 2011.

Kozhinov, A. N., Zhurov, K. O., and Tsybin, Y. O.: Iterative method for mass spectra recalibration via empirical estimation of the mass calibration function for Fourier transform mass spectrometry-based petroleomics, Anal Chem, 85, 6437-6445, 10.1021/ac400972y, 2013.

Kroll, J. H., and Seinfeld, J. H.: Chemistry of secondary organic aerosol: Formation and evolution of low-volatility organics in the atmosphere, Atmospheric Environment, 42, 3593-3624, 10.1016/j.atmosenv.2008.01.003, 2008.

Kroll, J. H., Donahue, N. M., Jimenez, J. L., Kessler, S. H., Canagaratna, M. R., Wilson, K. R., Altieri, K. E., Mazzoleni, L. R., Wozniak, A. S., Bluhm, H., Mysak, E. R., Smith, J. D., Kolb, C. E., and Worsnop, D. R.: Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol, Nat Chem, 3, 133-139, 10.1038/nchem.948, 2011.

Kujawinski, E. B., and Behn, M. D.: Automated Analysis of Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra of Natural Organic Matter, Analytical Chemistry, 78, 11, 10.1021/ac0600306, 2006.

Kujawinski, E. B., Longnecker, K., Blough, N. V., Vecchio, R. D., Finlay, L., Kitner, J. B., and Giovannoni, S. J.: Identification of possible source markers in marine dissolved organic matter using ultrahigh resolution mass spectrometry, Geochimica et Cosmochimica Acta, 73, 4384-4399, 10.1016/j.gca.2009.04.033, 2009.

Kundu, S., Fisseha, R., Putman, A. L., Rahn, T. A., and Mazzoleni, L. R.: High molecular weight SOA formation during limonene ozonolysis: insights from ultrahigh-resolution FT-ICR mass spectrometry characterization, Atmospheric Chemistry and Physics, 12, 5523-5536, 10.5194/acp-12-5523-2012, 2012.

Kunenkov, E. V., Kononikhin, A. S., Perminova, I. V., Hertkorn, N., Gaspar, A., Schmitt-Kopplin, P., Popov, I. A., Garmash, A. V., and Nikolaev, E. N.: Total Mass Difference Statistics Algorithm: A New Approach to Identification of High-Mass Building Blocks in Electrospray Ionization Fourier Transform Ion

Cyclotron Mass Spectrometry Data of Natural Organic Matter, Analytical Chemistry, 81, 10, 10.1021/ac901476u, 2009.

Laing, J. R., Jaffe, D. A., and Hee, J. R.: Physical and optical properties of aged biomass burning aerosol from wildfires in Siberia and the Western USA at the Mt. Bachelor Observatory, Atmospheric Chemistry and Physics, 16, 15185-15197, 10.5194/acp-16-15185-2016, 2016.

Leclair, J. P., Collett, J. L., and Mazzoleni, L. R.: Fragmentation analysis of water-soluble atmospheric organic matter using ultrahigh-resolution FT-ICR mass spectrometry, Environ Sci Technol, 46, 4312-4322, 10.1021/es203509b, 2012.

Lee, A. K. Y., Hayden, K. L., Herckes, P., Leaitch, W. R., Liggio, J., Macdonald, A. M., and Abbatt, J. P. D.: Characterization of aerosol and cloud water at a mountain site during WACS 2010: secondary organic aerosol formation through oxidative cloud processing, Atmospheric Chemistry and Physics, 12, 7103-7116, 10.5194/acp-12-7103-2012, 2012.

Leefmann, T., Frickenhaus, S., and Koch, B. P.: UltraMassExplorer: a browser-based application for the evaluation of high-resolution mass spectrometric data, Rapid Commun Mass Spectrom, 33, 193-202, 10.1002/rcm.8315, 2019.

Leenheer, J. A., Rostad, C. E., Gates, P. M., Furlong, E. T., and Ferrer, I.: Molecular Resolution and Fragmentation of Fulvic Acid by Electrospray Ionization/Multistage Tandem Mass Spectrometry, Analytical Chemistry, 73, 11, 10.1021/ac0012593, 2001.

Levin, E. J. T., McMeeking, G. R., Carrico, C. M., Mack, L. E., Kreidenweis, S. M., Wold, C. E., Moosmüller, H., Arnott, W. P., Hao, W. M., Collett, J. L., and Malm, W. C.: Biomass burning smoke aerosol properties measured during Fire Laboratory at Missoula Experiments (FLAME), Journal of Geophysical Research, 115, 10.1029/2009jd013601, 2010.

Li, Y., Pöschl, U., and Shiraiwa, M.: Molecular corridors and parameterizations of volatility in the chemical evolution of organic aerosols, Atmospheric Chemistry and Physics, 16, 3327-3344, 10.5194/acp-16-3327-2016, 2016.

Lightstone, J. M., Onasch, T. B., Imre, D., and Oatis, S.: Deliquescence, Efflorescence, and Water Activity in Ammonium Nitrate and Mixed Ammonium Nitrate/Succinic Acid Microparticles, Journal of Physical Chemistry A, 104, 10, 10.1021/jp002137h, 2000.

Lignell, H., Hinks, M. L., and Nizkorodov, S. A.: Exploring matrix effects on photochemistry of organic aerosols, Proc Natl Acad Sci U S A, 111, 13780-13785, 10.1073/pnas.1322106111, 2014.

Lim, Y. B., Tan, Y., Perri, M. J., Seitzinger, S. P., and Turpin, B. J.: Aqueous chemistry and its role in secondary organic aerosol (SOA) formation, Atmospheric Chemistry and Physics, 10, 10521-10539, 10.5194/acp-10-10521-2010, 2010.

Lin, G., Penner, J. E., Flanner, M. G., Sillman, S., Xu, L., and Zhou, C.: Radiative forcing of organic aerosol in the atmosphere and on snow: Effects of SOA and brown carbon, Journal of Geophysical Research: Atmospheres, 119, 7453-7476, 10.1002/2013jd021186, 2014.

Lin, P., Engling, G., and Yu, J. Z.: Humic-like substances in fresh emissions of rice straw burning and in ambient aerosols in the Pearl River Delta Region, China, Atmospheric Chemistry and Physics, 10, 6487-6500, 10.5194/acp-10-6487-2010, 2010.

Lin, P., Liu, J., Shilling, J. E., Kathmann, S. M., Laskin, J., and Laskin, A.: Molecular characterization of brown carbon (BrC) chromophores in secondary organic aerosol generated from photo-oxidation of toluene, Phys Chem Chem Phys, 17, 23312-23325, 10.1039/c5cp02563j, 2015.

Lin, P., Aiona, P. K., Li, Y., Shiraiwa, M., Laskin, J., Nizkorodov, S. A., and Laskin, A.: Molecular Characterization of Brown Carbon in Biomass Burning Aerosol Particles, Environ Sci Technol, 50, 11815-11824, 10.1021/acs.est.6b03024, 2016.

Liu, S., and Liang, X.-Z.: Observed Diurnal Cycle Climatology of Planetary Boundary Layer Height, Journal of Climate, 23, 5790-5809, 10.1175/2010jcli3552.1, 2010.

Liu, X., Huey, L. G., Yokelson, R. J., Selimovic, V., Simpson, I. J., Müller, M., Jimenez, J. L., Campuzano-Jost, P., Beyersdorf, A. J., Blake, D. R., Butterfield, Z., Choi, Y., Crounse, J. D., Day, D. A., Diskin, G. S., Dubey, M. K., Fortner, E., Hanisco, T. F., Hu, W., King, L. E., Kleinman, L., Meinardi, S., Mikoviny, T., Onasch, T. B., Palm, B. B., Peischl, J., Pollack, I. B., Ryerson, T. B., Sachse, G. W., Sedlacek, A. J., Shilling, J. E., Springston, S., St. Clair, J. M., Tanner, D. J., Teng, A. P., Wennberg, P. O., Wisthaler, A., and Wolfe, G. M.: Airborne measurements of western U.S. wildfire emissions: Comparison with prescribed burning and air quality implications, Journal of Geophysical Research: Atmospheres, 122, 6108-6129, 10.1002/2016jd026315, 2017.

Mace, K. A.: Water-soluble organic nitrogen in Amazon Basin aerosols during the dry (biomass burning) and wet seasons, Journal of Geophysical Research, 108, 10.1029/2003jd003557, 2003.

Maclean, A. M., Butenhoff, C. L., Grayson, J. W., Barsanti, K., Jimenez, J. L., and Bertram, A. K.: Mixing times of organic molecules within secondary organic aerosol particles: a global planetary boundary layer perspective, Atmospheric Chemistry and Physics, 17, 13037-13048, 10.5194/acp-17-13037-2017, 2017.

Makarov, A.: Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis, Analytical Chemistry, 72, 7, 10.1021/ac991131p, 2000.

Marshall, A. G., and Hendrickson, C. L.: High-resolution mass spectrometers, Annu Rev Anal Chem (Palo Alto Calif), 1, 579-599, 10.1146/annurev.anchem.1.031207.112945, 2008.

Massoli, P., Lambe, A. T., Ahern, A. T., Williams, L. R., Ehn, M., Mikkilä, J., Canagaratna, M. R., Brune, W. H., Onasch, T. B., Jayne, J. T., Petäjä, T., Kulmala, M., Laaksonen, A., Kolb, C. E., Davidovits, P., and Worsnop, D. R.: Relationship between aerosol oxidation level and hygroscopic properties of laboratory generated secondary organic aerosol (SOA) particles, Geophysical Research Letters, 37, n/a-n/a, 10.1029/2010gl045258, 2010.

Mathur, R., and O'Connor, P. B.: Artifacts in Fourier transform mass spectrometry, Rapid Commun Mass Spectrom, 23, 523-529, 10.1002/rcm.3904, 2009.

Mazzoleni, L. R., Ehrmann, B. M., Shen, X., Marshall, A. G., and Collett, J. L.: Water-Soluble Atmospheric Organic Matter in Fog: Exact Masses and Chemical Formula Identification by Ultrahigh-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, 44, 8, 10.1021/es903409k, 2010.

Mazzoleni, L. R., Saranjampour, P., Dalbec, M. M., Samburova, V., Hallar, A. G., Zielinska, B., Lowenthal, D. H., and Kohl, S.: Identification of water-soluble organic carbon in non-urban aerosols using ultrahigh-resolution FT-ICR mass spectrometry: organic anions, Environmental Chemistry, 9, 10.1071/en11167, 2012.

McLafferty, F. W., and Turecek, F.: Interpretation of Mass Spectra, University Science Books, 1993.

McNeill, V. F.: Aqueous organic chemistry in the atmosphere: sources and chemical processing of organic aerosols, Environ Sci Technol, 49, 1237-1244, 10.1021/es5043707, 2015.

Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Mü̈ller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J.-J., Cox, J., Horning, S., Mann, M., and Makarov, A.: Ultra High Resolution Linear Ion Trap Orbitrap Mass Spectrometer (Orbitrap Elite) Facilitates Top Down LC MS/MS and Versatile Peptide Fragmentation Modes, Molecular & Cellular Proteomics, 11, 12, 10.1074/mcp.O111.013698, 2012.

Moonshine, M., Rudich, Y., Katsman, S., and Graber, E. R.: Atmospheric HULIS enhance pollutant degradation by promoting the dark Fenton reaction, Geophysical Research Letters, 35, 10.1029/2008gl035285, 2008.

Muccio, Z., and Jackson, G. P.: Isotope Ratio Mass Spectrometry, Analyst, 134, 213-222, 10.1039/b808232d, 2009.

Mutzel, A., Poulain, L., Berndt, T., Iinuma, Y., Rodigast, M., Boge, O., Richters, S., Spindler, G., Sipila, M., Jokinen, T., Kulmala, M., and Herrmann, H.: Highly Oxidized Multifunctional Organic Compounds Observed in Tropospheric Particles: A Field and Laboratory Study, Environ Sci Technol, 49, 7754-7761, 10.1021/acs.est.5b00885, 2015.

Nagakura, S.: Intramolecular Charge Transfer Spectra Observed with Some Compounds Containing the Nitro or the Carbonyl Group, The Journal of Chemical Physics, 23, 1441-1445, 10.1063/1.1742324, 1955.

Nenes, A., Pandis, S. N., and PILINIS, C.: ISORROPIA: A New Thermodynamic Equilibrium Model for Multiphase Multicomponent Inorganic Aerosols, Aquatic Geochemistry, 4, 30, 1998.

Ng, N. L., Canagaratna, M. R., Jimenez, J. L., Chhabra, P. S., Seinfeld, J. H., and Worsnop, D. R.: Changes in organic aerosol composition with aging inferred from aerosol mass spectra, Atmospheric Chemistry and Physics, 11, 6465-6474, 10.5194/acp-11-6465-2011, 2011.

Nguyen, T. K. V., Zhang, Q., Jimenez, J. L., Pike, M., and Carlton, A. G.: Liquid Water: Ubiquitous Contributor to Aerosol Mass, Environmental Science & Technology Letters, 3, 257-263, 10.1021/acs.estlett.6b00167, 2016.

O'Dowd, C. D., Facchini, M. C., Cavalli, F., Ceburnis, D., Mircea, M., Decesari, S., Fuzzi, S., Jun Yoon, Y., and Putaud, J.-P.: Biogenically driven organic contribution to marine aeroso, Nature, 431, 5, 0.1038/nature02959, 2004.

Ohno, T., and Ohno, P. E.: Influence of heteroatom pre-selection on the molecular formula assignment of soil organic matter components determined by ultrahigh resolution mass spectrometry, Anal Bioanal Chem, 405, 3299-3306, 10.1007/s00216-013-6734-3, 2013.

Olivier, J.G.J. and Berdowski, J.J.M.: Global emissions sources and sinks, in: The Climate System, Berdowski, J., Guicherit, R. and Heij, B.J., A.A. Balkema Publishers/Swets & Zeitlinger Publishers, Lisse, The Netherlands. 33-78, 2001.

Ollivon, D., Blanchoud, H., Motelay-Massei, A., and Garban, B.: Atmospheric deposition of PAHs to an urban site, Paris,France, Atmospheric Environment, 36, 10, S 1352 -2310(02 )0 0089-4, 2002.

Owen, R. C., and Honrath, R. E.: Technical note: a new method for the Lagrangian tracking of pollution plumes from source to receptor using gridded model output,

Atmospheric Chemistry and Physics, 9, 19, www.atmos-chem-phys.net/9/2577/2009/, 2009.

Perdue, E. M., and Green, N. W.: Isobaric molecular formulae of C, H, and O: a view from the negative quadrants of van Krevelen space, Anal Chem, 87, 5079-5085, 10.1021/ac504165k, 2015.

Perraudin, E., Budzinski, H., and Villenave, E.: Kinetic Study of the Reactions of Ozone with Polycyclic Aromatic Hydrocarbons Adsorbed on Atmospheric Model Particles, Journal of Atmospheric Chemistry, 56, 57-82, 10.1007/s10874-006-9042-x, 2006.

PetroOrg, http://software.petroorg.com/, 2014.

Petters, M. D., and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmospheric Chemistry and Physics, 7, 11, www.atmos-chem-phys.net/7/1961/2007/, 2007.

Petters, S. S., Pagonis, D., Claflin, M. S., Levin, E. J. T., Petters, M. D., Ziemann, P. J., and Kreidenweis, S. M.: Hygroscopicity of Organic Compounds as a Function of Carbon Chain Length and Carboxyl, Hydroperoxy, and Carbonyl Functional Groups, J Phys Chem A, 121, 5164-5174, 10.1021/acs.jpca.7b04114, 2017.

Pfister, G. G., Emmons, L. K., Hess, P. G., Honrath, R., Lamarque, J. F., Val Martin, M., Owen, R. C., Avery, M. A., Browell, E. V., Holloway, J. S., Nedelec, P., Purvis, R., Ryerson, T. B., Sachse, G. W., and Schlager, H.: Ozone production from the 2004 North American boreal fires, Journal of Geophysical Research, 111, 10.1029/2006jd007695, 2006.

Phillips, S. M., and Smith, G. D.: Light Absorption by Charge Transfer Complexes in Brown Carbon Aerosols, Environmental Science & Technology Letters, 1, 382-386, 10.1021/ez500263j, 2014.

Phillips, S. M., and Smith, G. D.: Further evidence for charge transfer complexes in brown carbon aerosols from excitation-emission matrix fluorescence spectroscopy, J Phys Chem A, 119, 4545-4551, 10.1021/jp510709e, 2015.

Poschl, U.: Atmospheric aerosols: composition, transformation, climate and health effects, Angew Chem Int Ed Engl, 44, 7520-7540, 10.1002/anie.200501122, 2005.

Prather, K. A., Bertram, T. H., Grassian, V. H., Deane, G. B., Stokes, M. D., Demott, P. J., Aluwihare, L. I., Palenik, B. P., Azam, F., Seinfeld, J. H., Moffet, R. C., Molina, M. J., Cappa, C. D., Geiger, F. M., Roberts, G. C., Russell, L. M., Ault, A. P., Baltrusaitis, J., Collins, D. B., Corrigan, C. E., Cuadra-Rodriguez, L. A., Ebben, C. J., Forestieri, S. D., Guasco, T. L., Hersey, S. P., Kim, M. J., Lambert,

W. F., Modini, R. L., Mui, W., Pedler, B. E., Ruppel, M. J., Ryder, O. S., Schoepp, N. G., Sullivan, R. C., and Zhao, D.: Bringing the ocean into the laboratory to probe the chemical complexity of sea spray aerosol, Proc Natl Acad Sci U S A, 110, 7550-7555, 10.1073/pnas.1300262110, 2013.

Pratt, K. A., Heymsfield, A. J., Twohy, C. H., Murphy, S. M., DeMott, P. J., Hudson, J. G., Subramanian, R., Wang, Z., Seinfeld, J. H., and Prather, K. A.: In Situ Chemical Characterization of Aged Biomass-Burning Aerosols Impacting Cold Wave Clouds, Journal of the Atmospheric Sciences, 67, 2451-2468, 10.1175/2010jas3330.1, 2010.

Putman, A. L., Offenberg, J. H., Fisseha, R., Kundu, S., Rahn, T. A., and Mazzoleni, L. R.: Ultrahigh-resolution FT-ICR mass spectrometry characterization of α-pinene ozonolysis SOA, Atmospheric Environment, 46, 164-172, 10.1016/j.atmosenv.2011.10.003, 2012.

Quinn, P. K., Collins, D. B., Grassian, V. H., Prather, K. A., and Bates, T. S.: Chemistry and related properties of freshly emitted sea spray aerosol, Chem Rev, 115, 4383-4399, 10.1021/cr500713g, 2015.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reid, J. P., Bertram, A. K., Topping, D. O., Laskin, A., Martin, S. T., Petters, M. D., Pope, F. D., and Rovelli, G.: The viscosity of atmospherically relevant organic particles, Nat Commun, 9, 956, 10.1038/s41467-018-03027-z, 2018.

Rémillard, J., Kollias, P., Luke, E., and Wood, R.: Marine Boundary Layer Cloud Observations in the Azores, Journal of Climate, 25, 7381-7398, 10.1175/jcli-d-11-00610.1, 2012.

Rice, B. M., Sahu, S., and Owens, F. J.: Density functional calculations of bond dissociation energies for NO2 scission in some nitroaromatic molecules, Journal of Molecular Structure (Theochem), 583, 4, S0166-1280(01)00782-5, 2002.

Riedel, T., and Dittmar, T.: A method detection limit for the analysis of natural organic matter via Fourier transform ion cyclotron resonance mass spectrometry, Anal Chem, 86, 8376-8382, 10.1021/ac501946m, 2014.

Rothfuss, N. E., and Petters, M. D.: Influence of Functional Groups on the Viscosity of Organic Aerosol, Environ Sci Technol, 51, 271-279, 10.1021/acs.est.6b04478, 2017.

Rudich, Y., Donahue, N. M., and Mentel, T. F.: Aging of organic aerosol: bridging the gap between laboratory and field studies, Annu Rev Phys Chem, 58, 321-352, 10.1146/annurev.physchem.58.032806.104432, 2007.

Samburova, B., Hallar, A. G., Mazzoleni, L. R., Saranjampour, P., Lowenthal, D., Kohl, S. D., Zielinska, B.: Composition of water-soluble organic carbon in non-urban atmospheric aerosol collected at the Storm Peak Laboratory. Environ. Chem., 10(5), 370-380, https://doi.org/10.1071/EN13079, 2013.

Savory, J. J., Kaiser, N. K., McKenna, A. M., Xian, F., Blakney, G. T., Rodgers, R. P., Hendrickson, C. L., and Marshall, A. G.: Parts-per-billion Fourier transform ion cyclotron resonance mass measurement accuracy with a "walking" calibration equation, Anal Chem, 83, 1732-1736, 10.1021/ac102943z, 2011.

Schmitt-Kopplin, P., Gabelica, Z., Gougeon, R. D., Fekete, A., Kanawati, B., Harir, M., Gebefuegi, I., Eckel, G., and Hertkorn, N.: High molecular diversity of extraterrestrial organic matter in Murchison meteorite revealed 40 years after its fall, Proc Natl Acad Sci U S A, 107, 2763-2768, 10.1073/pnas.0912157107, 2010.

Schum, S. K., Zhang, B., Džepina, K., Fialho, P., Mazzoleni, C., and Mazzoleni, L. R.: Molecular and physical characteristics of aerosol at a remote free troposphere site: implications for atmospheric aging, Atmospheric Chemistry and Physics, 18, 14017-14036, 10.5194/acp-18-14017-2018, 2018.

Schum, S.K., L.R. Mazzoleni, L.E. Brown, MFAssignR: An R package for Data Preparation and Molecular Formula Assignment, R package version 0.0.2,http://github.com/skschum/MFAssignR, doi: 10.5281/zenodo.1471471, 2019.

Schumann, U., Weinzierl, B., Reitebuch, O., Schlager, H., Minikin, A., Forster, C., Baumann, R., Sailer, T., Graf, K., Mannstein, H., Voigt, C., Rahm, S., Simmet, R., Scheibe, M., Lichtenstern, M., Stock, P., Rüba, H., Schäuble, D., Tafferner, A., Rautenhaus, M., Gerz, T., Ziereis, H., Krautstrunk, M., Mallaun, C., Gayet, J. F., Lieke, K., Kandler, K., Ebert, M., Weinbruch, S., Stohl, A., Gasteiger, J., Groß, S., Freudenthaler, V., Wiegner, M., Ansmann, A., Tesche, M., Olafsson, H., and Sturm, K.: Airborne observations of the Eyjafjalla volcano ash cloud over Europe during air space closure in April and May 2010, Atmospheric Chemistry and Physics, 11, 2245-2279, 10.5194/acp-11-2245-2011, 2011.

Seibert, P., and Frank, A.: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, Atmospheric Chemistry and Physics, 4, 13, 1680-7324/acp/2004-4-51, 2004.

Shiraiwa, M., Li, Y., Tsimpidi, A. P., Karydis, V. A., Berkemeier, T., Pandis, S. N., Lelieveld, J., Koop, T., and Poschl, U.: Global distribution of particle phase state

in atmospheric secondary organic aerosols, Nat Commun, 8, 15002, 10.1038/ncomms15002, 2017a.

Shiraiwa, M., Ueda, K., Pozzer, A., Lammel, G., Kampf, C. J., Fushimi, A., Enami, S., Arangio, A. M., Frohlich-Nowoisky, J., Fujitani, Y., Furuyama, A., Lakey, P. S. J., Lelieveld, J., Lucas, K., Morino, Y., Poschl, U., Takahama, S., Takami, A., Tong, H., Weber, B., Yoshino, A., and Sato, K.: Aerosol Health Effects from Molecular to Global Scales, Environ Sci Technol, 51, 13545-13567, 10.1021/acs.est.7b04417, 2017b.

Shrivastava, M., Cappa, C. D., Fan, J., Goldstein, A. H., Guenther, A. B., Jimenez, J. L., Kuang, C., Laskin, A., Martin, S. T., Ng, N. L., Petaja, T., Pierce, J. R., Rasch, P. J., Roldin, P., Seinfeld, J. H., Shilling, J., Smith, J. N., Thornton, J. A., Volkamer, R., Wang, J., Worsnop, D. R., Zaveri, R. A., Zelenyuk, A., and Zhang, Q.: Recent advances in understanding secondary organic aerosol: Implications for global climate forcing, Reviews of Geophysics, 55, 509-559, 10.1002/2016rg000540, 2017.

Simoneit, B. R. T.: Biomass burning — a review of organic tracers for smoke from incomplete combustion, Applied Geochemistry, 17, 34, S0883-2927(01)00061-0, 2002.

Simonelt, B. R. T., Rogge, W. F., Mazurek, M. A., Standley, L. J., Hildemann, L. M., and Cass, G. R.: Lignin Pyrolysis Products, Lignans, and Resin Acids as Specific Tracers of Plant Classes in Emissions from Biomass Combustion, Environmental Science & Technology, 27, 9, 0013-936X/93/0927-2533, 1993.

Slade, J. H. and Knopf, D. A.: Multiphase OH oxidation kinetics of organic aerosol: The role of particle phase state and relative humidity, Geophys. Res. Lett., 41(14), 5297–5306, doi:10.1002/2014gl060582 , 2014.

Smith, D. F., Podgorski, D. C., Rodgers, R. P., Blakney, G. T., and Hendrickson, C. L.: 21 Tesla FT-ICR Mass Spectrometer for Ultrahigh-Resolution Analysis of Complex Organic Mixtures, Anal Chem, 90, 2041-2047, 10.1021/acs.analchem.7b04159, 2018.

Song, Y. C., Haddrell, A. E., Bzdek, B. R., Reid, J. P., Bannan, T., Topping, D. O., Percival, C., and Cai, C.: Measurements and Predictions of Binary Component Aerosol Particle Viscosity, J Phys Chem A, 120, 8123-8137, 10.1021/acs.jpca.6b07835, 2016.

Sorooshian, A., Lu, M.-L., Brechtel, F., Jonsson, H., Feingold, G., Flagan, R. and Seinfeld, J.: On the Source of Organic Acid Aerosol Layers above Clouds, Environ. Sci. Technol., 41(13), 4647–4654, doi:10.1021/es0630442, 2007.

Southam, A. D., Payne, T. G., Cooper, H. J., Arvanitis, T. N., and Viant, M. R.: Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method, Analytical Chemistry, 79, 8, 10.1021/ac062446p, 2007.

Spracklen, D. V., Mickley, L. J., Logan, J. A., Hudman, R. C., Yevich, R., Flannigan, M. D., and Westerling, A. L.: Impacts of climate change from 2000 to 2050 on wildfire activity and carbonaceous aerosol concentrations in the western United States, Journal of Geophysical Research, 114, 10.1029/2008jd010966, 2009.

Stranz, D.: Composer64 v1.5.0, Sierra Analytics Inc. 2015.

Stefanis, N.-A., Theoulakis, P., and Pilinis, C.: Dry deposition effect of marine aerosol to the building stone of the medieval city of Rhodes, Greece, Building and Environment, 44, 260-270, 10.1016/j.buildenv.2008.03.001, 2009.

Stenson, A. C., Marshall, A. G., and Cooper, W. T.: Exact Masses and Chemical Formulas of Individual Suwannee River Fulvic Acids from Ultrahigh Resolution Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra, Analytical Chemistry, 75, 10, 10.1021/ac026106p, 2003.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmospheric Chemistry and Physics, 5, 14, 1680-7324/acp/2005-5-2461, 2005.

Stroobant, V., de Hoffman, E., Libert, R., and Van Hoof, F.: Fast atom bombardment mass spectrometry and low energy collision induced tandem mass spectrometry of tauroconjugated bile acid anions, Journal of the American Society for Mass Spectrometry, 6, 9, 1044-0305(95)00203-P, 1995.

Takahama, S., Johnson, A., and Russell, L. M.: Quantification of Carboxylic and Carbonyl Functional Groups in Organic Aerosol Infrared Absorbance Spectra, Aerosol Science and Technology, 47, 310-325, 10.1080/02786826.2012.752065, 2013.

Taylor, P. E., Flagan, R. C., Valenta, R., and Glovsky, M. M.: Release of allergens as respirable aerosols: A link between grass pollen and asthma, Journal of Allergy and Clinical Immunology, 109, 51-56, 10.1067/mai.2002.120759, 2002.

Tolic, N., Liu, Y., Liyu, A., Shen, Y., Tfaily, M. M., Kujawinski, E. B., Longnecker, K., Kuo, L. J., Robinson, E. W., Pasa-Tolic, L., and Hess, N. J.: Formularity: Software for Automated Formula Assignment of Natural and Other Organic Matter from Ultrahigh-Resolution Mass Spectra, Anal Chem, 89, 12659-12665, 10.1021/acs.analchem.7b03318, 2017.

Tu, P., Hall, W. A. t., and Johnston, M. V.: Characterization of Highly Oxidized Molecules in Fresh and Aged Biogenic Secondary Organic Aerosol, Anal Chem, 88, 4495-4501, 10.1021/acs.analchem.6b00378, 2016.

Turetsky, M. R., Kane, E. S., Harden, J. W., Ottmar, R. D., Manies, K. L., Hoy, E., and Kasischke, E. S.: Recent acceleration of biomass burning and carbon losses in Alaskan forests and peatlands, Nature Geoscience, 4, 27-31, 10.1038/ngeo1027, 2010.

Tziotis, D., Hertkorn, N., and Schmitt-Kopplin, P.: Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity, Eur J Mass Spectrom (Chichester), 17, 415-421, 10.1255/ejms.1135, 2011.

U.S. Air Quality, Smog Blog. alg.umbc.edu, last access: 9 January 2018.

Vakkari, V., Kerminen, V.-M., Beukes, J. P., Tiitta, P., van Zyl, P. G., Josipovic, M., Venter, A. D., Jaars, K., Worsnop, D. R., Kulmala, M., and Laakso, L.: Rapid changes in biomass burning aerosols by atmospheric oxidation, Geophysical Research Letters, 41, 2644-2651, 10.1002/2014gl059396, 2014.

Val Martin, M., Honrath, R. E., Owen, R. C., and Lapina, K.: Large-scale impacts of anthropogenic pollution and boreal wildfires on the nitrogen oxides over the central North Atlantic region, Journal of Geophysical Research, 113, 10.1029/2007jd009689, 2008a.

Val Martin, M., Honrath, R. E., Owen, R. C., and Li, Q. B.: Seasonal variation of nitrogen oxides in the central North Atlantic lower free troposphere, Journal of Geophysical Research, 113, 10.1029/2007jd009688, 2008b.

Val Martin, M., Logan, J. A., Kahn, R. A., Leung, F.-Y., Nelson, D. L., and Diner, D. J.: Smoke injection heights from fires in North America: analysis of 5 years of satellite observations, Atmospheric Chemistry and Physics, 10, 20, www.atmos-chem-phys.net/10/1491/2010/, 2010.

Val Martín, M., Honrath, R. E., Owen, R. C., Pfister, G., Fialho, P., and Barata, F.: Significant enhancements of nitrogen oxides, black carbon, and ozone in the North Atlantic lower free troposphere resulting from North American boreal wildfires, Journal of Geophysical Research: Atmospheres, 111, n/a-n/a, 10.1029/2006jd007530, 2006.

Virtanen, A., Joutsensaari, J., Koop, T., Kannosto, J., Yli-Pirila, P., Leskinen, J., Makela, J. M., Holopainen, J. K., Poschl, U., Kulmala, M., Worsnop, D. R., and Laaksonen, A.: An amorphous solid state of biogenic secondary organic aerosol particles, Nature, 467, 824-827, 10.1038/nature09455, 2010.

Volkamer, R., Jimenez, J. L., San Martini, F., Džepina, K., Zhang, Q., Salcedo, D., Molina, L. T., Worsnop, D. R., and Molina, M. J.: Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected, Geophysical Research Letters, 33, 10.1029/2006gl026899, 2006.

Vu, V. Q.: ggbiplot: A ggplot2 based biplot. R package version 0.55. http://github.com/vqv/ggbiplot, 2011.

Walser, M. L., Desyaterik, Y., Laskin, J., Laskin, A., and Nizkorodov, S. A.: High-resolution mass spectrometric analysis of secondary organic aerosol produced by ozonation of limonene, Phys Chem Chem Phys, 10, 1009-1022, 10.1039/b712620d, 2008.

Warneck, P.: In-cloud chemistry opens pathway to the formation of oxalic acid in the marine atmosphere, Atmospheric Environment, 37, 2423-2427, 10.1016/s1352-2310(03)00136-5, 2003.

Warneke, C., Bahreini, R., Brioude, J., Brock, C. A., de Gouw, J. A., Fahey, D. W., Froyd, K. D., Holloway, J. S., Middlebrook, A., Miller, L., Montzka, S., Murphy, D. M., Peischl, J., Ryerson, T. B., Schwarz, J. P., Spackman, J. R., and Veres, P.: Biomass burning in Siberia and Kazakhstan as an important source for haze over the Alaskan Arctic in April 2008, Geophysical Research Letters, 36, n/a-n/a, 10.1029/2008gl036194, 2009.

Weber, R. J., Southam, A. D., Sommer, U., and Viant, M. R.: Characterization of isotopic abundance measurements in high resolution FT-ICR and Orbitrap mass spectra for improved confidence of metabolite identification, Anal Chem, 83, 3737-3743, 10.1021/ac2001803, 2011.

Wetzel, M. A.: Physical, chemical, and ultraviolet radiative characteristics of aerosol in central Alaska, Journal of Geophysical Research, 108, 10.1029/2002jd003208, 2003.

Wickham, W.: François, R., Henry, L., and Müller, K. dplyr: A Grammar of Data Manipulation. R package version 0.7.7. https://CRAN.R-project.org/package=dplyr, 2018.

Wickham, H.: ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham, W., Henry, L.: tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions. R package version 0.8.1. https://CRAN.R-project.org/package=tidyr, 2018.

Willoughby, A., Wozniak, A., and Hatcher, P.: Detailed Source-Specific Molecular Composition of Ambient Aerosol Organic Matter Using Ultrahigh Resolution Mass Spectrometry and 1H NMR, Atmosphere, 7, 10.3390/atmos7060079, 2016.

Wilke, C. O.: cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 0.8.0. https://CRAN.R-project.org/package=cowplot, 2017.

Witt, M., Fuchser, J., and Koch, B. P.: Fragmentation Studies of Fulvic Acids Using Collision Induced Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, Analytical Chemistry, 81, 7, 10.1021/ac802624s 2009.

Wong, J. P. S., Nenes, A., and Weber, R. J.: Changes in Light Absorptivity of Molecular Weight Separated Brown Carbon Due to Photolytic Aging, Environ Sci Technol, 51, 8414-8421, 10.1021/acs.est.7b01739, 2017.

Wozniak, A. S., Willoughby, A. S., Gurganus, S. C., and Hatcher, P. G.: Distinguishing molecular characteristics of aerosol water soluble organic matter from the 2011 trans-North Atlantic US GEOTRACES cruise, Atmospheric Chemistry and Physics, 14, 8419-8434, 10.5194/acp-14-8419-2014, 2014.

Yang, M., Zhou, Z., and Guo, D. A.: A strategy for fast screening and identification of sulfur derivatives in medicinal Pueraria species based on the fine isotopic pattern filtering method using ultra-high-resolution mass spectrometry, Anal Chim Acta, 894, 44-53, 10.1016/j.aca.2015.07.050, 2015.

Ye, Q., Robinson, E. S., Ding, X., Ye, P., Sullivan, R. C., and Donahue, N. M.: Mixing of secondary organic aerosols versus relative humidity, Proc Natl Acad Sci U S A, 113, 12649-12654, 10.1073/pnas.1604536113, 2016.

Zark, M., Christoffers, J., and Dittmar, T.: Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS, Marine Chemistry, 191, 9-15, 10.1016/j.marchem.2017.02.005, 2017.

Zelenyuk, A., Imre, D. G., Wilson, J., Bell, D. M., Suski, K. J., Shrivastava, M., Beranek, J., Alexander, M. L., Kramer, A. L., and Massey Simonich, S. L.: The effect of gas-phase polycyclic aromatic hydrocarbons on the formation and properties of biogenic secondary organic aerosol particles, Faraday Discuss, 200, 143-164, 10.1039/c7fd00032d, 2017.

Zhang, B., Owen, R. C., Perlinger, J. A., Kumar, A., Wu, S., Val Martin, M., Kramer, L., Helmig, D., and Honrath, R. E.: A semi-Lagrangian view of ozone production tendency in North American outflow in the summers of 2009 and 2010, Atmospheric Chemistry and Physics, 14, 2267-2287, 10.5194/acp-14-2267-2014, 2014.

Zhang, B., Owen, R. C., Perlinger, J. A., Helmig, D., Val Martín, M., Kramer, L., Mazzoleni, L. R., and Mazzoleni, C.: Ten-year chemical signatures associated with long-range transport observed in the free troposphere over the central North

Atlantic, Elementa: Science of the Anthropocene, 5, 20, 10.1525/journal.elementa.194, 2017.

Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Allan, J. D., Coe, H., Ulbrich, I., Alfarra, M. R., Takami, A., Middlebrook, A. M., Sun, Y. L., Džepina, K., Dunlea, E., Docherty, K., DeCarlo, P. F., Salcedo, D., Onasch, T., Jayne, J. T., Miyoshi, T., Shimono, A., Hatakeyama, S., Takegawa, N., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Williams, P., Bower, K., Bahreini, R., Cottrell, L., Griffin, R. J., Rautiainen, J., Sun, J. Y., Zhang, Y. M., and Worsnop, D. R.: Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenically-influenced Northern Hemisphere midlatitudes, Geophysical Research Letters, 34, n/a-n/a, 10.1029/2007gl029979, 2007.

Zhang, Y., Wen, Z., Washburn, M. P., and Florens, L.: Improving proteomics mass accuracy by dynamic offline lock mass, Anal Chem, 83, 9344-9351, 10.1021/ac201867h, 2011.

Zhao, Y., Hallar, A. G., and Mazzoleni, L. R.: Atmospheric organic matter in clouds: exact masses and molecular formula identification using ultrahigh-resolution FT-ICR mass spectrometry, Atmospheric Chemistry and Physics, 13, 12343-12362, 10.5194/acp-13-12343-2013, 2013.

Zhao, Yunzhu, "MOLECULAR CHARACTERIZATION OF ATMOSPHERIC ORGANIC MATTER IN BIOGENIC SECONDARY ORGANIC AEROSOL, AMBIENT AEROSOL AND CLOUDS", Dissertation, Michigan Technological University, 2014.

Zheng, Q., Morimoto, M., Sato, H., and Fouquet, T.: Resolution-enhanced Kendrick mass defect plots for the data processing of mass spectra from wood and coal hydrothermal extracts, Fuel, 235, 944-953, 10.1016/j.fuel.2018.08.085, 2019.

Zhurov, K. O., Kozhinov, A. N., Fornelli, L., and Tsybin, Y. O.: Distinguishing analyte from noise components in mass spectra of complex samples: where to cut the noise?, Anal Chem, 86, 3308-3316, 10.1021/ac403278t, 2014.

Zielinski, A. T., Kourtchev, I., Bortolini, C., Fuller, S. J., Giorio, C., Popoola, O. A. M., Bogialli, S., Tapparo, A., Jones, R. L., and Kalberer, M.: A new processing scheme for ultra-high resolution direct infusion mass spectrometry data, Atmospheric Environment, 178, 129-139, 10.1016/j.atmosenv.2018.01.034, 2018.

Zobrist, B., Marcolli, C., Pedernera, D. A., and Koop, T.: Do atmospheric aerosols form glasses?, Atmospheric Chemistry and Physics, 8, 24, www.atmos-chem-phys.net/8/5221/2008/, 2008.

Zubarev, R. A., and Makarov, A.: Orbitrap mass spectrometry, Anal Chem, 85, 5288-5296, 10.1021/ac4001223, 2013.

# A    Copyright Documentation

Images in Figure 1.3 and Figure 1.4 are from Wikimedia Commons. They are licensed for reuse under the Creative Commons Attribution-Share Alike 4.0 International license and the Creative Commons Attribution-Share Alike 3.0 Unported license, respectively. The fire images in Figure 1.2 are from kisspng.com. They are licensed under the personal use license. Please see below for full citation and attribution information.

Figure 1.3: "ICR cell schematic of a FTICR mass spectrometer" by Keministi at the English language Wikipedia. Licensed under CC BY-SA 4.0 via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:FTICR_cell.png.  Accessed April 2019.

Figure 1.4: "Ion injection into an Orbitrap mass analyzer" by Thermo Fisher Scientific at the English language Wikipedia. Licensed under CC BY-SA 3.0 via Wikimedia Commons - https://commons.wikimedia.org/wiki/File:OrbitrapMA%26Injector.png. Accessed April 2019.

Figure 1.2: "Cartoon Fire" by kisspng.com. Licensed for Personal Use via kisspng.com - https://www.kisspng.com/png-cartoon-fire-flame-clip-art-bbq-fire-cliparts-183620/download-png.html Accessed April 2019