2014

MOLECULAR CHARACTERIZATION OF ATMOSPHERIC ORGANIC MATTER IN BIOGENIC SECONDARY ORGANIC AEROSOL, AMBIENT AEROSOL AND CLOUDS

Yunzhu Zhao
Michigan Technological University

Recommended Citation

MOLECULAR CHARACTERIZATION OF ATMOSPHERIC ORGANIC MATTER IN BIOGENIC SECONDARY ORGANIC AEROSOL, AMBIENT AEROSOL AND CLOUDS

By

Yunzhu Zhao

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Environmental Engineering

MICHIGAN TECHNOLOGICAL UNIVERSITY

2014

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

Department of Chemistry

Dissertation Advisor: *Lynn R Mazzoleni*

Committee Member: *Sarah A Green*

Committee Member: *Judith A Perlinger*

Committee Member: *Raymond A Shaw*

Committee Member: *Shiliang Wu*

Department Chair: *Cary F Chabalowski*

Table of Content

List of Figures

List of Tables

Abstract

Atmospheric aerosol affects the Earth's energy budget, reduces visibility and influences human health. The organic composition of aerosol is quite complex and continuously evolves through various atmospheric processes. To gain a deeper understanding of the molecular composition of atmospheric organic matter (AOM), chamber-generated biogenic secondary organic aerosol (SOA), ambient aerosol and cloud water samples were studied. Ultrahigh resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry was used to provide detailed molecular characterization of the atmospheric samples. Due to the extremely high mass resolution and mass accuracy, thousands of individual molecular formulas were identified in all of the samples studied. Multivariate statistical analysis methods were evaluated to compare the similarities and differences of the sample compositions. The biogenic SOA from three individual monoterpene precursors and a sesquiterpene precursor have clusters of peaks in their mass spectra, indicating that high molecular weight oligomers are a major component of the SOA. The monoterpene SOA have similar molecular compositions, which are different from the sesquiterpene SOA composition. The indicator species of SOA were identified using multivariate statistical analysis. Daily 24-hour water-soluble organic carbon samples from ambient aerosol collected at the Storm Peak Laboratory (SPL) show similar bulk chemical properties regarding their average elemental ratios and double bond equivalents. Using multivariate statistical analysis, the site meteorological conditions were found to affect the aerosol molecular composition. Days with strong UV radiation and high temperature were found to contain large numbers of biogenic SOA molecular formulas. Days with high relative humidity and high sulfate ion concentrations were found to contain many sulfur-containing compounds, suggesting their aqueous phase formation. The collection of cloud samples at the SPL provided an opportunity to study aqueous processing of AOM. The cloud composition was affected by biomass burning and SOA. Comparisons of the sample compositions indicate biogenic SOA components are commonly observed in ambient aerosol and cloud samples collected at the SPL. Thus, the ambient samples were used to confirm the biogenic SOA indicator species identified in chamber-generated SOA. This study of the three types of atmospheric samples helps to understand the composition of AOM with respect to atmospheric processes.

Chapter 1 Introduction

1.1 Introduction to atmospheric aerosol

1.1.1 Definitions, sources and sinks of atmospheric aerosol

Atmospheric aerosol exists as a fine suspension of solid and liquid particles surrounded by gases in the atmosphere. Often the term aerosol refers to the condensed phase components associated with fine particulate matter (Seinfeld and Pandis, 2006). Though the atmospheric particulate matter (PM) range in size from 10^{-9} m (molecules and molecular clusters) to 10^{-4} m (small raindrops) in diameter (Poschl, 2005), fine PM usually refers to the aerosols with aerodynamic diameters ≤ 2.5 µm (denoted as PM2.5).

Aerosols emitted from various emission sources to the atmosphere are referred to as primary and those formed in the atmosphere via chemical reactions are referred to as secondary. Sources of primary aerosols include wind-driven or mechanical suspensions of soil and rock debris, volcanic eruptions, sea sprays, biomass burning, incomplete combustion of fuels, and biological debris like pollens or fungi. Secondary aerosols are formed through gas-to-particle conversion of condensable species or formed through condensed phase reactions. Inorganic gases like sulfur dioxide, nitrogen dioxide and ammonia can convert into particulate phase sulfate, nitrate and ammonium by gaseous phase oxidation or heterogeneous reactions. Volatile organic compounds (VOCs) can be oxidized in the atmosphere and the less volatile oxidation products condense to form secondary organic aerosols (SOA). Sometimes this is referred to as gaseous SOA (Ervens et al., 2011). SOA can also arise from reactions in the condensed phase, which is sometimes referred to as aqueous SOA. Section 1.2 will describe gaseous SOA and aqueous SOA in further detail. Thus, aerosols are generated through natural and anthropogenic processes. Resident times in the atmosphere vary from hours to weeks (Raes et al., 2000; Williams et al., 2002). Aerosol properties evolve during their residence, referred to as aerosol aging. Aerosols are removed from the atmosphere through dry and wet deposition. Dry deposition results from gravitational settling of the particles to the Earth's surface, which is more efficient for large particles than small particles. Wet deposition occurs with precipitation, where aerosol particles serve as nuclei or are scavenged by collisions with a cloud or rain droplet.

1.1.2 Significance of atmospheric aerosol

Atmospheric aerosols are important in many aspects. They affect the Earth's energy budget, which is relevant to global warming and climate change. They also have effects on human health because they can be inhaled deep into the lungs or make contact with people's skin. Aerosols reduce visibility of elevated concentrations, which have been reported often thought out the world, especially in the megacities of the developing world.

Aerosols affect the Earth's radiation budget. The radiative forcing of several atmospheric components has been documented in the Intergovernmental Panel on Climate Change report (IPCC, 2013). Despite the well-known warming effects exerted by the long-lived greenhouse gases like CO_2, CH_4, halocarbons and N_2O, many aerosol components are considered to have cooling effects on the Earth's temperature, by directly and indirectly reflecting incoming solar radiation. The absorption or reflection of solar radiation by aerosols is known as the aerosol direct effect. Aerosols also serve as cloud condensation nuclei (CCN) to form clouds, which interact with solar radiation; this is known as the aerosol indirect effect. Both direct and indirect effects play an important role in the Earth's energy budget (IPCC, 2013). The light absorbing species in aerosols include black carbon (IPCC 2013) and brown carbon (Desyaterik et al., 2013) and they have a positive radiative forcing. The aerosol components with cooling effects or a negative radiative forcing include nitrate, sulfate, organic carbon and so on.

Atmospheric aerosols also have adverse impacts on human health. It has been well established that exposure to aerosols causes or enhances stress on the respiratory system, cardiovascular system and allergic diseases, though the mechanism of action affecting health is poorly understood (Bernstein et al., 2004; Englert, 2004; Pope and Dockery, 2006). Cohen et al. (2005) reported that "PM2.5 causes about 3 % of mortality from cardiopulmonary disease, about 5% of mortality from cancer of the trachea, bronchus, and lung, and about 1 % of mortality from acute respiratory infections in children under 5 years, worldwide." Studies have shown that despite the wide variety of aerosol properties like size or composition, the observed relative health risk estimates per unit of particulate matter rarely depend on the particle characterization (Finlayson-Pitts and Pitts, 1997; Davidson et al., 2005). In addition, not a single or a few chemical components of aerosols were found to be dominantly responsible for the adverse health effects (Harrison and Yin, 2000; Davidson et al., 2005), suggesting that either the target components are

not detected by available routine methods, or that there are common reactive intermediates found in most particles (Kao and Friedlander, 1995; Finlayson-Pitts and Pitts, 1997). In October 2013, the World Health Organization announced outdoor air pollution, as a whole, is classified as carcinogenic to humans; the particulate matter, is evaluated separately and also identified as carcinogenic to humans (http://www.iarc.fr/en/media-centre/iarcnews/pdf/pr221_E.pdf). The detailed molecular composition of aerosols may help to understand the health effects of aerosols.

An elevated concentration of aerosol in the atmosphere also reduces visibility, giving a "hazy" appearance in the atmosphere (Finlayson-Pitts and Pitts, 2000). The particles themselves may absorb visible light, leading to a shorter visible distance. Besides, they can form fog or haze, which indirectly reduce the visibility. The number of particles in the atmosphere affects the distribution of water vapor. For example, a large number of particles provides more cloud condensation nuclei. If a higher number of particles are activated, the water vapor associated with each particle is less and thus smaller droplets are formed. It is difficult for the small droplets to grow big enough to form rain droplets. Therefore the formed fog or haze has a longer lifetime, which influences the visibility for a longer time. Besides the visibility effect of the fogs, they indirectly influence the climate, as described in Section 1.1.2. The composition of the particles plays a role in this droplet-formation process and the particles with more hygroscopic (water-affinity) components will be more easily activated. Water-soluble organic carbon is an important fraction of the hygroscopic components in aerosols and will be introduced in Section 1.4.

Other roles of atmospheric aerosol include affecting the distribution and abundance of atmospheric trace gases and influencing the biogeochemical circulation of the organic species.

1.1.3 Atmospheric aerosol composition and organic aerosols

Regarding the chemical composition of atmospheric aerosols, inorganic ions have been commonly observed. They are major components of the aerosols with relatively high concentrations quantified using ion chromatography. The use of inorganic ions to serve as markers for aerosol source apportionment motivates the study of inorganic ions. Frequently identified anions include sulfate (SO_4^{2-}), nitrate (NO_3^-), phosphate (PO_4^{3-}) and chloride (Cl^-). Cations like ammonium (NH_4^+), sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), magnesium (Mg^{2+}),

aluminum (Al^{3+}), cadmium (Cd^{3+}), iron (Fe^{2+} and Fe^{3+}), lead (Pd^{2+}) and silicon (Si^{2+}) have been identified in aerosols, though usually only ammonium shows high concentration. Black or elemental carbon and organic carbon are also significant aerosol components with quantities comparable to inorganic ions. But unlike the inorganic species, the speciation of organics is far from being well-understood due to its high complexity; thus usually organic carbon is considered as a whole group in aerosol composition. Organic carbon comprises 20-50% of the aerosol mass at continental mid-latitudes and up to 90% in tropical forested areas (Kanakidou et al., 2005). An example of the mass faction of aerosols with inorganics and organics is shown in Figure 1.1.
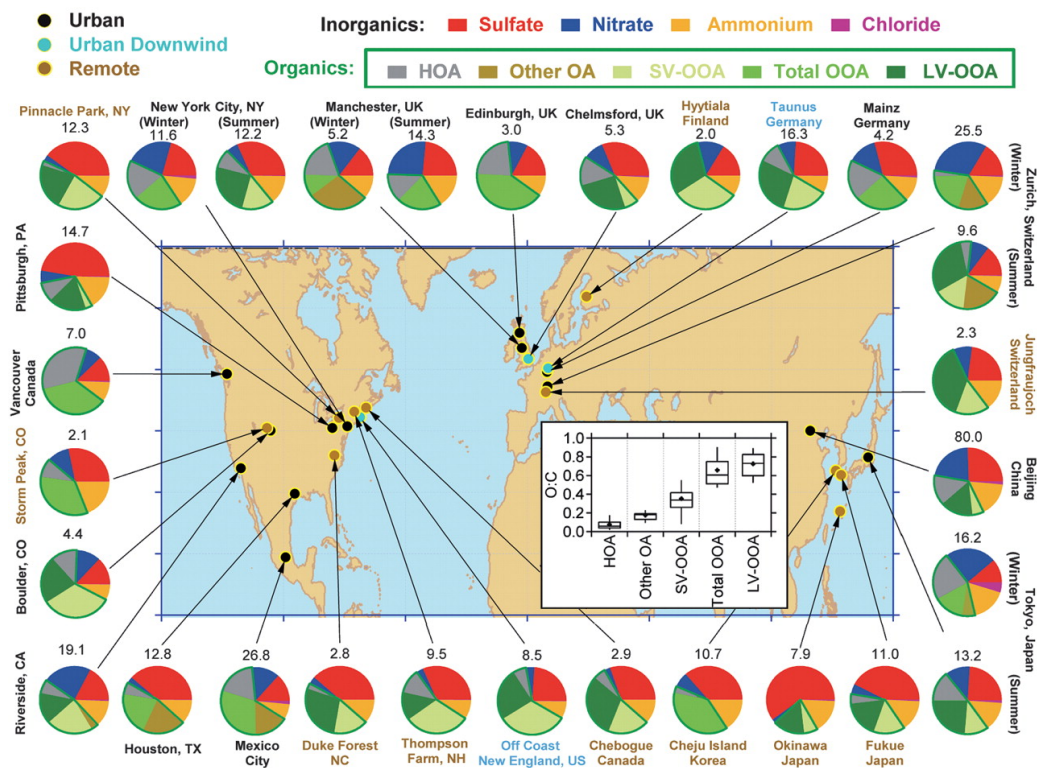
Figure 1.1 Total mass concentration (in micrograms per cubic meter) and mass fractions of nonrefractory inorganic and organic components in submicrometer aerosols measured with the aerosol mass spectrometer (AMS) at multiple surface locations in the Northern Hemisphere. This figure is from Jimenez et al. (2009), reprinted with permission from AAAS.

1.2 Biogenic secondary organic aerosol

Formation of gaseous SOA starts with gaseous oxidation of VOCs. The most common atmospheric oxidants are hydroxyl radical (OH·), nitrate radical (NO$_3$·) and ozone (O$_3$). A radical (OH·, NO$_3$·, etc.) initiates oxidation either by abstraction of a hydrogen atom or addition to a C-C double bond (Kroll and Seinfeld, 2008), to form an alkyl radical (R·). In ozonolysis, an alkene is cleaved by ozone to form a carbonyl and an energetically excited carbonyl oxide ("Criegee intermediate"). The excited intermediate proceeds to form an alkyl radical through the "hydroperoxide channel", or become a stabilized Criegee intermediate, which then reacts with water or oxygenated organics. The alkyl radicals formed from the initial oxidation, combine with oxygen (O$_2$) readily in the atmosphere to form alkylperoxy radicals (RO$_2$·). RO$_2$· react with various atmospheric gases or radicals like NO$_2$, NO, HO$_2$·, RO$_2$·, NO$_3$· and so on, to generate peroxynitrates, organic nitrates, hydroperoxides, alcohols, carbonyls, etc. The addition of oxygen-containing functional groups (ketone, hydroxyl, hydroperoxyl, nitrate, carboxylic acid and so on) in the oxidation products lower the compound vapor pressure, compared to that of the original VOC (Kroll and Seinfeld, 2008). Thus the products are condensable and can become particulate matter.

SOA yield depends on the partitioning of the semi-volatile VOC oxidation products, between the gas and condensed phases. Pankow (1994a, b) proposed a partitioning model with the following equation (equation 1-1):

$$Kp = \frac{F/TSP}{A} \qquad\qquad 1\text{-}1$$

where Kp (m^3 µg^{-1}) is a temperature-dependent partitioning constant, TSP (µg m^{-3}) is the concentration of total suspended particulate material, and F (µg m^{-3}) and A (µg m^{-3}) are the particulate- and gaseous concentrations of the compound of interest, respectively. Thus a semi-volatile compound with a lower saturation vapor pressure would favor the condensed phase; the high concentration of total organic mass loading (TSP) helps the semi-volatile compounds to partition to the aerosol phase.

Once the organic compounds are in the condensed phase, they undergo particle phase reactions (Kroll and Seinfeld, 2008). Non-oxidative processes, or accretion reactions, are evidenced by the observation of high molecular weight species in SOA, including the formation of

peroxyhemiacetals, hemiacetals, aldol addition products, esters or acid anhydrides, sulfate esters and Criegee intermediate adducts. Also, the particle phase organic species can be oxidized by $OH\cdot$, $NO_3\cdot$, O_3, etc., with mechanisms similar to the gaseous phase reactions. The formed SOA evolve continuously in the atmosphere until removed by dry or wet deposition to the ground or surface water.

What kinds of VOCs in the atmosphere contribute significantly to SOA formation? VOCs are emitted to the atmosphere naturally or biogenically (e.g., from vegetation) and anthropogenically (e.g., from the production and usage of industrial solvents). The global flux of biogenic VOC was estimated to be 1150 TgC for the year of 1990 with the largest mass fraction attributed to isoprene (503 TgC) (Gelencser, 2004). Monoterpenes contributed 127 TgC to the flux (Gelencser, 2004). In contrast, the total anthropogenic VOC flux was only 110 Tg yr^{-1} (Gelencser, 2004). Regarding the chemical class of VOCs, the standard view is that only a few classes of VOCs dominate SOA formation (Kroll and Seinfeld, 2008), in spite of a large number of VOCs are emitted. The anthropogenic VOCs that generate significantly to aerosol mass are aromatic and large aliphatic species. Examples of the aromatic species include toluene, ethylbenzene, ethyltoluene, *n*-propylbenzene, xylene, trimethylbenzene, tetramethylbenzene and so on; examples of the large aliphatic species are like 1-octene, 1-decene, heptanal and nonanal (Gelencser, 2004). In terms of the hundreds of biogenic VOCs, only monoterpenes (with molecular formula: $C_{10}H_{16}$) and sesquiterpenes (molecular formula: $C_{15}H_{24}$) are thought to significantly contribute to SOA formation (Gelencser, 2004). Terpenes are biogenic hydrocarbons with building blocks of C_5H_8, which are mainly emitted by conifers and broad-leaved trees. The structures of commonly studied biogenic VOCs with effective SOA formation are shown in Figure 1.2. Recently, isoprene (2-methyl-1,3-butadiene, C_5H_8) was found to be capable of producing SOA too (Claeys et al., 2004a; Claeys et al., 2004b; Edney et al., 2005; Surratt et al., 2006b; Ng et al., 2008; Carlton et al., 2009). Even though the yield might be minor, the overall contribution of isoprene to SOA mass could be large due to the large amount isoprene in the global emissions.

Consistent with the biogenic and anthropogenic VOC inventories, the global estimated biogenic SOA fluxes (12-70 Tg/yr) are larger than the anthropogenic SOA fluxes (2-12 Tg/yr) (Hallquist et al., 2009). Biogenic SOA has been proven to contribute significantly to ambient aerosols in rural and forested areas with large amount of trees from direct field measurements using different

methods. For example, Chan et al. (2010) observed and quantified isoprene-derived epoxydiols (species identified in chamber studies) in ambient aerosols in the southeastern United States. Surratt et al. (2006a; 2008) attributed organosulfates found in aerosol collected at Alabama and Georgia to be from isoprene, α-pinene, β-pinene and limonene-like SOA. Sun et al (2011) used positive matrix factorization (PMF) of the high-resolution mass spectra to classify the biogenic SOA fraction of the ambient aerosols. Mazzoleni et al. (2012) compared the components identified using ultrahigh resolution mass spectrometry of ambient aerosols to those of chamber generated SOA and reported observations of a large number of biogenic SOA components in the ambient aerosols.
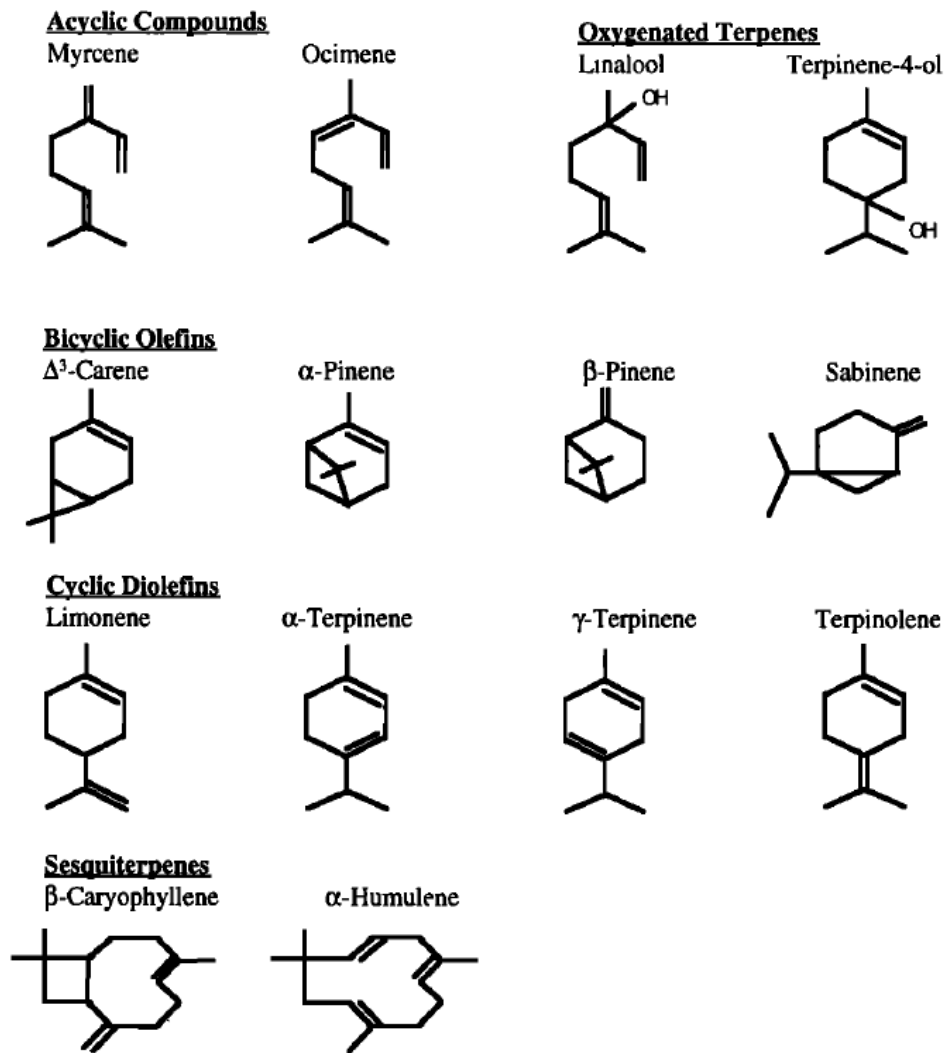
Figure 1.2 Chemical structures of the biogenic VOCs relevant for SOA formation. This figure is from Griffin et al. (1999), reprinted with permission from John Wiley and Sons.

Chamber studies enable researchers to focus on biogenic SOA characterization. In the ambient environment, the aerosol is composed of inorganic ions, primary organic aerosol components, anthropogenic SOA and biogenic SOA components. The composition is affected by a great number of environmental parameters like temperature, wind, solar radiation, trace gas concentration and so on. However, in the atmospheric aerosol chambers the environmental variables can be controlled. For instance, a specific biogenic VOC and oxidant can be selected and their concentrations can be controlled, reactions can proceed under different relative humidity conditions and the reaction time can be adjusted.

Besides the studies of mass yield and the reaction kinetics of SOA formation (e.g., Griffin et al., 1999; Jonsson et al., 2005; Jonsson et al., 2008), the chemical composition of biogenic SOA attracted much recent attention. Gas chromatography (GC) was widely used to detect and quantify the organic species in the particle phase, often with chemical derivatizations. From these studies, several organic compounds were identified in monoterpene and sesquiterpene SOA experiments and some of them are referred as tracer compounds for biogenic SOA or SOA from specific VOC precursors. For example, pinonaldehyde, pinonic acid and cis-pinic acid were commonly identified in α-pinene SOA experiments (Christoffersen et al., 1998; Hoffmann et al., 1998; Atkinson and Arey, 2003; Jaoui et al., 2005); nopinone was observed in β-pinene SOA (Hatakeyama et al., 1991; Wisthaler et al., 2001); β-caryophyllinic acid was identified as a tracer for β-caryophyllene SOA (Jaoui et al., 2007). Almost all of these identified species are 1st generation oxidation products with molecular weights less than 300 Da. Since the first observation of a stable dimer of cis-pinic and norpinic acid in the α-pinene/ozone system by Hoffmann et al (1998), high molecular weight oligomers were observed in several recent studies (Heaton et al., 2007; Iinuma et al., 2007; Reinhardt et al., 2007), although only a few of them were recently identified with known molecular formulas or chemical structures (Kristensen et al., 2013). They are mostly formed in the particle phase under multiple generations. Ultrahigh resolution mass spectrometry analysis enables the full scan of the particulate phase species and their molecular weight identification (Bateman et al., 2009; Kundu et al., 2012; Putman et al., 2012). Usually the mass spectra show several groups of peaks, attributed to monomers, dimers and so on.

 Previous studies of molecular characterization of the SOA composition typically only characterized the SOA from one kind of VOC precursor. In this study, we generated SOA from

individual α-pinene, β-pinene, D-limonene or β-caryophyllene ozonolysis experiments under different conditions. In this study, the molecular composition of several SOA samples will be compared. α-pinene, β-pinene and D-limonene are the three most commonly studied monoterpenes with the highest global emissions (Griffin et al., 1999); β-caryophyllene is the most frequently studied and abundant sesquiterpene (Helmig et al., 2007).

Biogenic SOA tracers help to identify the sources of ambient aerosols. They can also be incorporated into atmospheric models for various purposes. Traditionally, biogenic SOA samples were analyzed using GC and the highest abundance peaks were identified as tracer compounds. However, the GC methods usually detect a limited number of tracer compounds at one time and most of them are low molecular weight species (e.g., monomers). In this study, the identification of biogenic SOA tracers from different precursors using ultrahigh resolution mass spectrometry analysis methods will be presented. The full scan mass spectra of the organic species in SOA facilitate the identification of dimer and even trimer tracers.

1.3 Water-soluble organic carbon in aerosols

A major component of the organic carbon in ambient aerosols is water-soluble and hence is referred to as water-soluble organic carbon (WSOC). Saxena and Hildemann (1996) proposed the threshold of 1 gram of solute dissolving into 100 grams of water to distinguish water-soluble and water-insoluble atmospheric organics. It should be noted that an aerosol particle is a highly complex mixture of chemical species. The "matrix" of a specific compound in the particle is composed of many other species. Thus the solubility of the specific compound in the unknown matrix could be different from its solubility in pure water. For example, neutralization by alkaline ions can easily render a water-insoluble compound water-soluble (Gelencser, 2004). In most cases, the term WSOC from the practical view, represents the aerosol constituents that can be extracted by water. In the early studies (around 1950s) of organic composition, aerosols were primarily extracted with non-polar solvents like benzene, cyclohexane or dichloromethane for GC analysis. The identified compounds only comprised a small fraction of the total organic mass (usually less than 20%). Thus, the introduction of polar solvents such as water and methanol for aerosol extractions helped to recover a more substantial fraction of the aerosol organic carbon. Note that besides the WSOC in aerosols, there is also WSOC in cloud droplets

and fog droplets. In those cases, the droplet water acts as the natural solvent. The WSOC in ambient clouds will be discussed in Section 1.4.

The importance of water-soluble organic compounds is related to their affinity for water, or rather their hygroscopicity. The presence of WSOC facilitates the particles to absorb water vapor from the gas phase, which alters many physiochemical characteristics of the particles. For instance, the size and the phase of the particles change if water is absorbed, and thus the optical properties and the lifetime of the particles are influenced correspondingly. The species in the particles undergo aqueous phase reactions once water partitions to the particle, which significantly alters the aerosol chemical composition (Ervens et al., 2011). More importantly, water-soluble organic compounds are often associated with the CCN activities of the particles due to their water-absorbing ability. Inorganic aerosols, mostly sulfate aerosols, were thought to be major cloud nuclei. In 1993, Novakov and Penner (1993) reported evidence suggesting that organic aerosol also plays a key role in cloud nucleation. Formation of fog or cloud droplets, strengthened by the WSOC in aerosol changes the solar radiation budget of the Earth (the indirect effect of aerosols on climate).

It is not surprising that water-soluble organic compounds are those with hydrophilic functional groups, such as hydroxyl, carbonyl or carboxylic groups. The following are some examples of WSOC species previously identified in atmospheric particles. Kawamura et al. (2013) and Kundu et al. (2010) extracted atmospheric aerosols using organic-free pure water and used GC-mass spectrometry (GC-MS) for analysis after chemical derivatization to identify dicarboxylic acids, ketocarboxylic acids and α-dicarbonyls in biomass burning aerosols. More specifically, they identified saturated n-dicarboxylic acids (C_2 to C_{11} with chemical formulas: $HOOC\text{-}(CH_2)_{0-9}\text{-}COOH$), branched dicarboxylic acids (methylmalonic acid: $HOOC\text{-}CH(CH_3)\text{-}COOH$, methylsuccinic acid: $HOOC\text{-}CH(CH_3)\text{-}CH_2\text{-}COOH$ and methylglutaric acid: $HOOC\text{-}CH(CH_3)\text{-}(CH_2)_2\text{-}COOH$), unsaturated carboxylic acids (C_4 and C_5 with carbon-carbon double bonds, and diacids with phenyl groups like phthalic acid: $HOOC\text{-}(C_6H_4)\text{-}COOH$, isophthalic acid and terephthalic acid), multifunctional dicarboxylic acids (hydroxysuccinic acid: $HOOC\text{-}CH(OH)\text{-}CH_2\text{-}COOH$, ketomalonic acid: $HOOC\text{-}HC(O)\text{-}COOH$ and ketopimelic acid: $HOOC\text{-}CH_2\text{-}CH_2\text{-}HC(O)(CH_2)_2\text{-}COOH$), ketocarboxylic acids (pyruvic acid: $CH_3\text{-}(O)C\text{-}CHO$, glyoxylic acid or 2-oxoethanoic acid: $OHC\text{-}COOH$, 3-oxopropanoic acid, 4-oxobutanoic acid, 5-oxopentanoic acid, 7-oxoheptanoic acid, 8-oxooctanoic acid, 9-oxononanoic acid) and α-dicarbonyls (glyoxal: $OHC\text{-}CHO$ and methylglyoxal:

CH$_3$-O(C)-CHO). Samy et al. (2010) used a similar water extraction and derivatization procedure for GC-MS analysis of 84 water-soluble analytes in aerosols and clouds collected at the top of a mountain in Colorado, US. Besides the low molecular weight acids mentioned above (maximum carbon number: 11), they identified relatively larger water-soluble compounds with up to 24 carbon atoms. In addition to the carbonyls and acids, they also observed methoxyphenols and other biomass burning related compounds like levoglucosan (C$_6$H$_{10}$O$_5$), cholesterol (C$_{27}$H$_{46}$O) and β-sitosterol (C$_{29}$H$_{50}$O). Samburova et al. (2013) quantified over 45 water-soluble organics and found fairly high concentrations of sugar alcohols, sugars and sugar anhydrates in aqueous extracts of aerosols.

Despite the hundreds of water-soluble organic species identified, they are still an incomplete coverage of the WSOC in atmospheric aerosol composition in terms of the carbon mass. It has been suggested that the major missing fraction is high molecular weight compounds or macromolecular compounds (typically > 300 Da). In 1980, Simoneit (1980) first separated particulate organics into fulvic acid, humic acid and residual carbon fractions. Later in 1983, Likens et al. (1983) observed macromolecular organics in precipitation samples. Mukai and Ambe (1986) found and compared the brown substances from airborne particulate matter to the extracts from the local soil, dead leaves, smoke from burning plant matter, and soot from automotive exhaust. They suggested agricultural burning as the primary source of this brown, high molecular weight acidic material. Later in 1998, Havers et al. (1998) studied the spectroscopic characterization of "humic-like substances" in airborne particulate matter. They introduced the term "HULIS", which has been commonly used in the recent literature to represent the water-soluble high molecular weight species in aerosol and clouds. The observed spectroscopic properties of these compounds closely resemble those of natural humic and fulvic acids. The first efforts were dedicated to the determination of the molecular size of HULIS. It is commonly agreed that they have molecular weight up to 1000 Da as evidenced by ultrafiltration analysis (Havers et al., 1998), laser desorption/ionization mass spectrometry (Samburova et al., 2005) and later ultrahigh resolution MS analysis (e.g., Kiss et al., 2001; Wozniak et al., 2008). Determination of the chemical functional groups has been facilitated by means of UV/VIS, Fourier-transformed infrared (FT-IR), nuclear magnetic resonance (NMR), MS analysis and so on (Havers et al., 1998; Decesari et al., 2005; Sullivan and Weber, 2006; Herckes et al., 2007; Dron et al., 2010; Fan et al., 2013). Commonly observed functional groups include hydroxyl groups of

alcohols and phenols, carboxylic groups, carbonyl groups and aromatic rings. Organosulfates (Gómez-González et al., 2008) and organic nitrogen (appearing as amines, nitrate esters, peptides and nitrooxy-organosulfate compounds) (Herckes et al., 2007; Dron et al., 2010) were also identified as HULIS components. Graber and Rudich (2006) reviewed the studies of HULIS characterization and stated that compared to aquatic and terrestrial humic substances, HULIS usually has a smaller average molecular weight, lower aromatic moiety content, greater surface activity and better droplet activation ability.

Where does HULIS, a major component of WSOC, come from? It is a complex mixture of components from primary organic aerosol and SOA with reaction products from both gas phase and condensed phase reaction pathways. Fossil fuel combustion produces HULIS (Baduel et al., 2009; El Haddad et al., 2009). Biomass burning also contributes to HULIS composition (Hoffer et al., 2004; Schmitt-Kopplin et al., 2010). A major component of HULIS is formed by atmospheric chemical reactions that take place either in the particulate phase or aqueous phase. SOA was found to be an important constituent of HULIS or WSOC, especially in the summer when photochemical reactions are favored (Wozniak et al., 2008; Mazzoleni et al., 2012). Aqueous phase reaction products are important constituents of HULIS, too. It has been proven that aqueous reaction of methylglyoxal with inorganic salts produce light absorbing high molecular weight compounds (Sareen et al., 2010).

Despite the widely studied properties of HULIS and WSOC, they are still regarded as an unresolved complex mixture since little is known about their molecular composition. A new type of molecular characterization of airborne WSOC began in the late 2000s with the introduction of electrospray ionization Fourier transform-ion cyclotron mass spectrometry (ESI FT-ICR MS) for atmospheric sample analysis. Ultrahigh resolution FT-ICR MS enables the assignment of molecular formulas to all of the ionized species observed in samples. A large amount of chemical information can be inferred from the chemical formulas of the individual species. Thus, it significantly improved our understanding of atmospheric processes. For example, the degree of oxygenation of an atmospheric sample can be represented by the average oxygen-to-carbon ratio (O/C), which is calculated from the atomic O/C ratios of the identified species in the sample. Organic compounds with heteroatoms (for example, nitrogen and sulfur) were commonly observed and their compound classes can be inferred from the molecular formulas or from tandem MS/MS analysis (LeClair et al., 2012).

Due to the complexity of the data, most studies characterize only a few samples of atmospheric organic matter on a molecular level. The organic composition of the aerosols, for example, collected at a certain location was studied at a single time point. In this study, we collected daily 24-hour aerosol samples continuously over 1 month. The molecular characterization of the WSOC is presented on a daily basis for the month and can be compared. In addition, the correlation of the observed organic species with the inorganic species will be presented. The effects of the meteorological parameters and ozone concentrations on WSOC composition will also be discussed.

1.4 Water-soluble organic carbon in clouds

(Section 1.4 is from the introduction of a publication by Zhao, Y., Hallar, A. G., and Mazzoleni, L. R. titled "Atmospheric organic matter in clouds: exact masses and molecular formula identification using ultrahigh-resolution FT-ICR mass spectrometry" published in Atmospheric Chemistry and Physics, volume 13, 12343-12362 in 2013. The articles published by Copernicus Publications (the publisher of Atmospheric Chemistry and Physics) are licensed under the Creative Commons Attribution 3.0 License together with an author copyright.)

Cloud and fog droplets play a major role in the processing of atmospheric organic matter (AOM) (Blando and Turpin, 2000; Collett et al., 2008; Mazzoleni et al., 2010; Ervens et al., 2011; Herckes et al., 2013). Under conditions of high relative humidity, aerosol particles may act as cloud condensation nuclei where water vapor condenses onto them to form droplets. Suspended droplets are subject to gas-liquid equilibrium of volatile compounds and solid-liquid equilibrium of non-volatile compounds (Graedel and Weschler, 1981), thus new transport and transformation pathways affecting the composition of AOM may occur (Collett et al., 2008; Lim et al., 2010; Ervens et al., 2011). Volatile water-soluble compounds like glyoxal and methylglyoxal can readily transfer into the aqueous phase of droplets due to their high effective Henry's law constant (e.g., H_{eff} = 3.710 x 10^3 M atm^{-1} at 25°C for methylglyoxal (Betterton and Hoffmann, 1988)). Similarly, organic compounds in CCN aerosol can partition into the aqueous phase. Since water-soluble organic carbon may comprise up to 90% of the total organic carbon mass in aerosols (Samburova et al., 2013) and the particle activation process usually favors particles with large fraction of water-soluble components (Facchini et al., 1999) a large fraction

of the WSOC may partition to the aqueous phase. Thus a considerable number of species, either from the gas phase or the particle phase, can be found in the liquid phase of the droplets.

The multiphase environment of the droplets facilitates aqueous phase reactions of the water-soluble species. A number of laboratory experiments mimicking wet aerosols showed precursors like glyoxal or methylglyoxal react with inorganic species like ammonium nitrate or ammonium sulfate to produce a complex mixture of light absorbing organic compounds (Carlton et al., 2007; De Haan et al., 2009; Galloway et al., 2009; Perri et al., 2009; Shapiro et al., 2009; Volkamer et al., 2009; Sareen et al., 2010; Sun et al., 2010). Likewise, a number of laboratory experiments have demonstrated the production of secondary organic compounds under cloud-relevant conditions (Tan et al., 2010; Yasmeen et al., 2010; Bateman et al., 2011), indicating cloud processing of AOM. In fact, in-cloud production of oxalic acid (Crahan et al., 2004), organic acids (Sorooshian et al., 2010), organosulfates (Pratt et al., 2013) and secondary organic aerosol (SOA; Wonaschuetz et al., 2012) has been observed in ambient measurements. Aerosol high molecular weight compounds (also known as "oligomers") may also originate from aqueous phase reactions (Altieri et al., 2008; De Haan et al., 2010; Lim et al., 2010; Yasmeen et al., 2010). Upon droplet evaporation (> 90% of cloud droplets evaporate, (Pruppacher and Klett, 1996)), aerosols with new composition are released. Volkamer and colleagues (Volkamer et al., 2007) suggested missing aerosol aqueous phase reactions or cloud processing may be the reason for the under predictions of organic aerosol mass using atmospheric models (Heald et al., 2005; Yu et al., 2008). The aerosol mass burden and the oxygen-to-carbon ratio of aged ambient organic aerosol are larger than that of SOA from dry smog chamber (Aiken et al., 2008; Ng et al., 2010). Aqueous processing of the SOA could explain their high O/C ratios (Ervens et al., 2011; Waxman et al., 2013). Therefore, it is crucial to study the composition of ambient cloud samples for a better understanding of SOA formation.

Clouds and fog water composition has been studied for decades, however very little study of the high molecular weight compounds in ambient cloud water has been done. Inorganic ions including Na^+, NH_4^+, Mg^{2+}, Ca^{2+}, K^+, Cl^-, NO_3^-, SO_4^{2-}, Br^-, PO_4^{3-} were observed commonly in clouds at various places and seasons (Weathers et al., 1988; Hindman et al., 1994; Deininger and Saxena, 1997; Collett et al., 2002; Simeonov et al., 2003; Lowenthal et al., 2004; Decesari et al., 2005; Gioda et al., 2011; Gioda et al., 2013). Several low molecular weight organic species have been studied and are believed to be transferred from the gas phase (Marinoni et al., 2004),

including formic acid, acetic acid (Keene et al., 1995; Laj et al., 1997), formaldehyde (Keene et al., 1995; Munger et al., 1995; Laj et al., 1997), glyoxal, methylglyoxal (Munger et al., 1995), phenol and nitrophenol (Lüttke and Levsen, 1997a; Lüttke et al., 1997b). Dozens of other compounds were observed in clouds, mostly monocarboxylic acids, dicarboxylic acids, alcohols and aldehydes (Limbeck and Puxbaum, 2000; Marinoni et al., 2004; van Pinxteren et al., 2005; Samy et al., 2010; Charbouillot et al., 2012). Usually the identified low molecular weight organics comprise a small fraction of the total organic mass in cloud water samples. For example, van Pinxteren et al (2005) identified ~20% of the dissolved organic carbon by measuring the individual organic compounds. Similarly, Facchini et al (1999) measured ~120 individual organic compounds in fog water which only accounted on average less than 5% of total WSOC. Compounds similar to the high molecular weight organic compounds or HULIS found in aerosols (Havers et al., 1998; Graber and Rudich, 2006) were also recently observed in clouds (Feng and Möller, 2004; Decesari et al., 2005; Reyes-Rodríguez et al., 2009). For example, Reyes-Rodriguez et al. (2009) observed the high molecular weight organic compounds in cloud samples to be mostly aliphatic and oxygenated, with a small amount of aromatic compounds based on study of the functional groups. Overall, the previous studies either observed several individual compounds or observed functional groups rather than the complex mixture of the organic compounds on a molecular level. In this study, we will use ultrahigh resolution mass spectrometry to full scan the complex organics in clouds.

1.5 Ultrahigh resolution mass spectrometry analysis of atmospheric samples

Electrospray ionization (ESI) coupled with Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS) provides detailed molecular characterization of organic matter due to its extreme high resolution and mass accuracy (Marshall et al., 1998; Kujawinski, 2002; Sleighter and Hatcher, 2007). ESI is a soft ionization method which minimizes the fragmentation of analytes, allowing detection of intact molecules, while the ultrahigh resolution and high accuracy of FT-ICR allows assignment of a unique chemical formula to each peak detected. FT-ICR MS has been successfully applied to ambient fog water samples (Mazzoleni et al., 2010; LeClair et al., 2012), rainwater samples (Altieri et al., 2009a; Altieri et al., 2009b; Altieri et al., 2012) and aerosols (Reemtsma et al., 2006; Wozniak et al., 2008; Laskin et al., 2009; Schmitt-

Kopplin et al., 2010; Lin et al., 2012b; Mazzoleni et al., 2012). This powerful analytical approach was applied to characterize the complex mixture of chamber-generated biogenic SOAs, ambient aerosol-derived water-soluble organic carbon and water-soluble organic carbon in clouds. The comprehensive characterization of the AOM composition after negative ion ESI FT-ICR MS will be discussed.

1.6 Application of statistical analysis in atmospheric samples

In the study of chamber-generated biogenic SOA, approximately 20 SOA samples were generated and analyzed using ultrahigh resolution ESI FT-ICR MS; similarly in the study of aerosol-derived WSOC approximately 30 daily WSOC samples were collected and analyzed using ESI-FT_ICR MS. In both cases, I will compare the molecular compositions of the SOA samples and the compositions of the WSOC samples to understand the sample-composition relationships within the two datasets. As will be described in Chapters 2 – 5, the atmospheric samples are quite complex with thousands of monoisotopic molecular formulas identified in each of the samples. To deal with the large datasets from ultrahigh resolution MS, I will apply multivariate statistical analysis in the data analysis. Hierarchical cluster analysis and principal component analysis have been applied in a few studies with ultrahigh resolution MS data as a sample analysis tool for marine dissolved organic matter (Kujawinski et al., 2009; Sleighter et al., 2010), petroleum samples (Hur et al., 2010), marine rainwater (Altieri et al, 2012) and atmospheric aerosols (Kourtchev et al., 2014; Wozniak et al., 2014). With the help of multivariate statistical analysis, the similarities and differences of molecular compositions of the atmospheric samples will be explored.

1.7 Overview of the organization of the dissertation

This dissertation began with an introduction of the key aspects of aerosol and clouds relevant to the research presented in the later chapters. My doctoral research was composed from three projects, including: molecular characterization of chamber-generated biogenic SOA, ambient WSOC in aerosol and ambient WSOC in clouds. The three projects all included detailed characterization of the atmospheric organic composition and are related to each other by their goal to understand the composition of atmospheric aerosol and clouds. The biogenic SOA samples were generated in the laboratory with controlled experimental parameters to focus on the composition of biogenic SOA, thus they are expected to have a less complex molecular

composition than the ambient atmospheric samples. The ambient WSOC collected at the Storm Peak Laboratory was highly influenced by local biogenic emissions and thus is expected to contain biogenic SOA components. To evaluate this, the composition was compared to the chamber-generated SOA. Various data preparation methods for multivariate statistical were tested on the controlled biogenic SOA samples, and subsequently applied to the analysis of the ambient WSOC samples. The study of WSOC in clouds provides insight on the AOM composition in the atmospheric aqueous phase. Both biogenic SOA and aerosol-derived WSOC components are expected to appear in cloud WSOC and undergo atmospheric aqueous processes.

Chapter 2 will describe the research methods for each of the projects. This includes the atmospheric sample collection, preparation, instrumental analysis and data processing methods. Due to the large dataset generated from the ultrahigh resolution MS analysis and the good of comparing the samples, multivariate statistical analysis methods were explored. The fundamental mathematical principles for these methods are also introduced in Chapter 2. The results and discussion for each of the three projects will be presented in Chapters 3 – 5, each chapter focusing on one project. In Chapter 3, the molecular composition of 20 biogenic SOA samples generated in the laboratory is presented. The molecular compositions of different SOA samples were compared, and molecular markers of biogenic SOA were identified and evaluated using ambient atmospheric samples. The controlled laboratory experiments provide an ideal test case for the development of the multivariate statistical analysis methods to be applied to the ambient aerosol-derived WSOC analysis. In Chapter 4 the results of more than 30 daily ambient aerosol WSOC samples is discussed. Multivariate statistical analyses were used to find similarities and differences among the samples and the relationships between the other field observations (inorganic ion concentrations, meteorological data and so on) and the WSOC molecular composition were also explored. Chapter 5 focuses on the composition of two ambient cloud samples. The conclusions from these studies are presented in Chapter 6. The aim of these projects was to improve our understanding of atmospheric chemical processes involving organic compounds, which is crucial for climate change, human health and so on.

Chapter 2 Methods

In this chapter, the methods for conducting research are described. This includes methods for the following three research topics: biogenic secondary organic aerosol, ambient aerosol-derived water-soluble organic carbon and ambient supercooled cloud composition. They will be presented in separate sections in this chapter. Each section describes the sample generation or collection, sample preparation, instrumental analysis, data processing and statistical analysis methods will be described.

2.1 Chamber-generated biogenic secondary organic aerosol

2.1.1 Biogenic secondary organic aerosol generation and collection

A total of 19 biogenic secondary organic aerosol (SOA) samples were generated in an aerosol chamber at the Los Alamos National Laboratory (LANL), which are referred to as LANLSOA samples. They are from the ozonolysis of individual biogenic precursors including α-pinene, β-pinene, D-limonene and β-caryophyllene. The experiments were conducted with a variety of relative humidity (RH) conditions with or without cyclohexane as an OH radical scavenger. α-pinene, β-pinene and D-limonene are biogenic monoterpenes with the molecular formula $C_{10}H_{16}$ (molecular weight 136 Da). A-pinene and β-pinene are isomers with an endocyclic or an exocyclic double bond, respectively. D-limonene on the other hand has both an endocyclic and an exocyclic double bond. B-caryophyllene is a sesquiterpene with the molecular formula $C_{15}H_{24}$ (molecular weight 204 Da) with an endocyclic and an exocyclic double bond. The structures of the four biogenic precursors of this study are shown in Figure 2.1.

α-pinene   β-pinene   D-limonene   β-caryophyllene
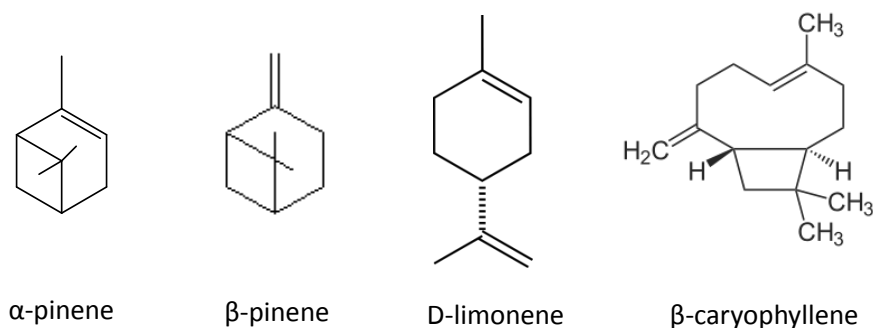
Figure 2.1 The structure of the biogenic precursors used to generate the LANLSOA samples. The aerosol chamber is a 1.5 m^3 flexible Teflon bag (thickness: 5mm) suspended on a bench top. It is covered with a black fabric to simulate dark conditions. Before each experiment, the chamber was flushed with ultrahigh purity compressed air until the particle concentration was less than 2 particles cm^{-3}, which was monitored by a condensation particle counter (CPC, TSI 3025A). Prior to introducing the reactants to the chamber, the RH was adjusted and recorded using a humidity sensor (Vaisala PTU300). The RH was adjusted to 0, 4% or 30% in different experiments (Table 2.1). A heated port inlet was used to introduce the terpene vapors into the chamber for 15 min before ozone was added. A 4 μL aliquot of α-pinene, β-pinene or D-limonene was injected to produce the target concentration of 500 ppb; likewise 4 μL of β-caryophyllene was injected to produce a terpene concentration of 290 ppb, due to its difference in molecular weight and density compared to the monoterpenes. Then ozone was introduced until a concentration of 250 ppb was reached, according to an ozone monitor (2B Technologies, Model 205). Seed aerosols were not added to any of the experiments. In some of the experiments (Table 2.1), 25 μL of cyclohexane was introduced into the chamber though a heated port (concentration of cyclohexane in chamber was ~3770 ppb). Cyclohexane was used as an OH radical scavenger to react with the OH radicals generated during the ozonolysis of terpenes (Jonsson et al., 2008). A heat mat located below the chamber was maintained at 25 – 28 °C to mix the reactants homogeneously throughout the chamber. Aerosol particles were observed within seconds after ozone was introduced to the chamber. The particle number concentration reached ~ 20000 particles cm^{-3}. The reactants were allowed to react for an hour before sampling began.

Table 2.1 LANLSOA experimental conditions.

Terpene	Sample Name	Experiment Date	Terpene Concenration, ppb (Volume, μL)	Ozone, min (conc, ppb)	Target RH %	cyclohexane, uL (concentration, ppb)
α-Pinene	AP2_0	6/15/2010	500 (4)	5 (250)	0	0 (0)
	AP3_4	7/26/2010	500 (4)	5 (250)	4	0 (0)
	AP4_30	7/27/2010	500 (4)	5 (250)	30	0 (0)
β-Pinene	BP8_0	6/23/2010	500 (4)	5 (250)	0	0 (0)
	BP9_0	6/24/2010	500 (4)	5 (250)	0	0 (0)
	BP10_4	7/22/2010	500 (4)	5 (250)	4	0 (0)
	BP11_30	7/23/2010	500 (4)	5 (250)	30	0 (0)
Limonene	LM1_0	7/8/2010	500 (4)	5 (250)	0	0 (0)
	LM2_4	7/28/2010	500 (4)	5 (250)	4	0 (0)
	LM3_30	5/24/2011	500 (4)	5 (250)	30	0 (0)
	LM4_0_CH	5/25/2011	500 (4)	5 (250)	0	25 (3770)
	LM5_4_CH	5/26/2011	500 (4)	5 (250)	4	25 (3770)
	LM6_30_CH	5/27/2011	500 (4)	5 (250)	30	25 (3770)
β-Caryophyllene	CP1_0	6/1/2011	290*(4)	5 (250)	0	0 (0)
	CP2_4	6/3/2011	290*(4)	5 (250)	4	0 (0)
	CP3_30	6/6/2011	290*(4)	5 (250)	30	0 (0)
	CP4_0_CH	6/8/2011	290*(4)	5 (250)	0	25 (3770)
	CP5_4_CH	6/13/2011	290*(4)	5 (250)	4	25 (3770)
	CP6_30_CH	6/14/2011	290*(4)	5 (250)	30	25 (3770)

*Differences in ppb are mainly attributed to molecular weight and density change between the monoterpenes and the sesquiterpene.

The resulting SOA from each experiment was collected onto a prebaked quartz fiber filter (600°C for > 8 hours; 47 mm diameter) with a flow rate of 6 LPM for 2.5 – 5 hours without a denuder. The sampled filters were stored in petri dished (Pall Corporation) wrapped with aluminum foil in a freezer at -20 °C. The samples were shipped cold overnight with blue ice packs to Michigan Tech. Half of each filter was extracted with 5 mL of a 1:1 mixture of acetonitrile (Chromasolv, Sigma Aldrich) and water (ACS, Sigma Aldrich). Acetonitrile is not expected to react with the SOA components (Bateman et al., 2008); water may hydrate the analyte molecules potentially altering the functional groups. The extraction was carried out in a sonicating bath for 30 min. The extracts was decanted and kept at -20 °C until analysis. The chamber blanks were treated following the same procedures, except the terpenes and ozone were not added to the chamber.

Each sample was labeled to indicate the precursor, sequence number, relative humidity and presence of cyclohexane (Table 2.1). The precursor is identified with one of the following abbreviations: α-pinene (AP), β-pinene (BP), D-limonene (LM), or β-caryophyllene (CP). The number following the precursor indicates the sequence number of the experiments. The percent relative humidity is placed after an underscore. If cyclohexane was present in the chamber, CH is placed after another underscore at the end (e.g., AP3_4_CH indicates an ozonolysis experiment with α-pinene at 4% RH with cyclohexane).

2.1.2 Secondary organic aerosol sample analysis

Ultrahigh resolution mass spectrometry analysis of the aerosol extracts was performed using a hybrid linear ion trap Fourier-transform ion cyclotron resonance mass spectrometer (LTQ FT Ultra, Thermo Scientific) equipped with an electrospray ionization (ESI) source. The diluted solution was infused directly into the ESI interface, the ESI parameters were adjusted to obtain a stable ion current with minimum ion injection time into the mass analyzer. After optimization the infusion flow rate was 4 μl/min, the ESI needle voltage was between -3.7kV and -4.0kV, and the capillary temperature was 265 °C. No sheath gas was used. The sample delivery apparatus was flushed in between each analysis with a minimum of 500 μL of acetonitrile and water (50:50) and 500 μL of 100% acetonitrile until background noise levels were reached. Besides the instrumental parameters, the electrospray efficiencies of the analytes depend on their chemical structures. Analytes with hydroxyl or carboxyl functional groups favor ionization in the negative ion electrospray mode. Many SOA species contain these hydrophilic functional groups, thus

most of the SOA components are expected to be observed in the negative ion mass spectra; though not all the organic carbon species are ionized. The relative abundances of the peaks in the mass spectra are not only related to the compound concentrations in the mixture, but also influenced by their electrospray efficiencies. Thus the relative abundance does not exactly reflect the compound concentration. The highly complex mixture and the lack of the organic standard hinder the quantification of the organic compounds. Quantification of the SOA components was not conducted in this study. Negative ion mass spectra were collected using full scan analysis for the range of 100 < m/z < 1000. Mass resolving power, $m/\Delta m_{50\%}$, in which $\Delta m_{50\%}$ is peak full width at half-maximum peak height, was set at 400,000 (at m/z 400) for all spectra. Automatic gain control was used to consistently fill the LTQ with the same number of ions ($n = 10^6$) for each acquisition and to avoid space charge effects from over-filling the mass analyzer. The instrument was externally calibrated in the negative ion mode with a standard solution of sodium dodecyl sulfate and taurocholic acid. The resulting mass accuracy was better than 2 ppm. Over 200 individual mass spectra recorded in the time domain were collected and stored as transients for each sample using Thermo Xcalibur software.

2.1.3 Molecular formula assignment

Ultrahigh resolution FT-ICR mass spectra were processed using Composer software (Sierra Analytics, Version 1.0.5) as previously described in Putman et al. (2012) and Mazzoleni et al. (2012). Each mass spectrum was processed individually with the following procedures. Approximately 200 time domain transients were co-added to improve the signal-to-noise ratio (Kujawinski et al., 2002) and the signal reproducibility (Kido Soule et al., 2010). Internal recalibration of the mass spectra (Sleighter and Hatcher, 2008) was performed to improve the molecular formula assignment accuracy. Several high relative abundance SOA components were used as internal recalibrants. They were selected due to their high relative abundances, long CH_2 homologous series (species with same molecular composition only deferring in CH_2); and several CH_2 homologous series of formulas were combined as the internal recalibrant list to try to span the whole m/z range. The peaks in the spectrum are internally recalibrated by reference to the expected exact m/z of the internal recalibrants. The lists of internal recalibrants for formula assignment of LANLSOA mass spectra from different precursors are given in Table 2.2. The match tolerances for the internal recalibrants were set at 3 ppm with a minimum relative abundance (RA) of 18 times the root mean square (RMS) of the signal-to-noise ratio between

the range m/z 900 - 1000. The molecular formula calculator was set to allow up to 100 carbon (C), 200 hydrogen (H) and 20 oxygen (O) atoms per molecular formula composition. The calculator is based on the PREDATOR algorithm (Blakney et al., 2011) and uses a Kendrick mass defect (KMD, see Section 2.1.4) analysis (Hughey et al., 2001) to sort ions into CH_2 homologous series and then assigns the de novo molecular formulas ≤ 500 Da. All the ions > 500 Da either belong to a CH_2 homologous series with de novo formulas of ≤ 500 Da or were not assigned. An assignment threshold of 6 times the RMS of the signal-to-noise ratio was applied to the detected peaks. The molecular formulas resulting in measurement errors > 3 ppm were discarded. Additional data filtering of the assigned formulas was done by applying the rules and assumptions as described by Koch et al. (2005) and Putman et al. (2012).

Table 2.2 Internal recalibrants used for LANLSOA formula assignment.

alpha-pinene		beta-pinene		limonene		beta-caryophyllene	
m/z	Formula	m/z	Formula	m/z	Formula	m/z	Formula
129.019332	C5H6O4	129.019332	C5H6O4	189.040462	C7H10O6	173.009161	C6H6O6
143.034982	C6H8O4	143.034982	C6H8O4	203.056112	C8H12O6	187.024812	C7H8O6
157.050632	C7H10O4	157.050632	C7H10O4	217.071762	C9H14O6	201.040462	C8H10O6
171.066282	C8H12O4	171.066282	C8H12O4	231.087412	C10H16O6	215.056112	C9H12O6
185.081932	C9H14O4	185.081932	C9H14O4	245.103062	C11H18O6	229.071762	C10H14O6
199.097583	C10H16O4	199.097583	C10H16O4	259.118712	C12H20O6	243.087412	C11H16O6
213.113233	C11H18O4	213.113233	C11H18O4	273.134362	C13H22O6	257.103062	C12H18O6
227.128883	C12H20O4	227.128883	C12H20O4	287.150012	C14H24O6	271.118712	C13H20O6
241.144533	C13H22O4	241.144533	C13H22O4	301.165662	C15H26O6	285.134362	C14H22O6
255.160183	C14H24O4	255.160183	C14H24O4	315.181312	C16H28O6	299.150012	C15H24O6
269.175833	C15H26O4	269.175833	C15H26O4	329.196962	C17H30O6	313.165662	C16H26O6
283.191483	C16H28O4	283.191483	C16H28O4	343.212612	C18H32O6	327.181312	C17H28O6
297.207133	C17H30O4	297.207133	C17H30O4	343.212612	C19H34O6	341.196962	C18H30O6
311.222783	C18H32O4	311.222783	C18H32O4	347.09837	C12H16O10	355.212612	C19H32O6
259.045941	C10H12O8	259.045941	C10H12O8	347.09837	C13H18O10	369.228262	C20H34O6
273.061591	C11H14O8	273.061591	C11H14O8	347.09837	C14H20O10	383.243912	C21H36O6
287.077241	C12H16O8	287.077241	C12H16O8	361.11402	C15H22O10	397.259563	C22H38O6
301.092891	C13H18O8	301.092891	C13H18O8	375.129671	C16H24O10	411.275213	C23H40O6
315.108541	C14H20O8	315.108541	C14H20O8	389.145321	C17H26O10	425.290863	C24H42O6
329.124191	C15H22O8	329.124191	C15H22O8	403.160971	C18H28O10	439.306513	C25H44O6
343.139841	C16H24O8	343.139841	C16H24O8	417.176621	C19H30O10	453.322163	C26H46O6
357.155491	C17H26O8	357.155491	C17H26O8	431.192271	C20H32O10	467.337813	C27H48O6
371.171141	C18H28O8	371.171141	C18H28O8	445.207921	C21H34O10	481.353463	C28H50O6
385.186791	C19H30O8	385.186791	C19H30O8	459.223571	C22H36O10	495.369113	C29H52O6
399.202442	C20H32O8	399.202442	C20H32O8	473.239221	C23H38O10	509.384763	C30H54O6
413.218092	C21H34O8	413.218092	C21H34O8	487.254871	C24H40O10	523.400413	C31H56O6
427.233742	C22H36O8	427.233742	C22H36O8	501.270521	C25H42O10	537.416063	C32H58O6
441.249392	C23H38O8	441.249392	C23H38O8	515.286171	C26H44O10	551.431713	C33H60O6
455.265042	C24H40O8	455.265042	C24H40O8	529.301821	C27H46O10	401.108935	C17H22O11
469.280692	C25H42O8	469.280692	C25H42O8	543.317471	C28H48O10	415.124585	C18H24O11
483.296342	C26H44O8	483.296342	C26H44O8	557.333121	C29H50O10	429.140235	C19H26O11
417.10385	C17H22O12	417.10385	C17H22O12	491.140629	C19H26O14	443.155885	C20H28O11
431.1195	C18H24O12	431.1195	C18H24O12	491.140629	C20H28O14	457.171535	C21H30O11
445.13515	C19H26O12	445.13515	C19H26O12	505.156279	C21H30O14	471.187185	C22H32O11
459.1508	C20H28O12	459.1508	C20H28O12	519.171929	C22H32O14	485.202835	C23H34O11
473.16645	C21H30O12	473.16645	C21H30O12	533.187579	C23H34O14	499.218486	C24H36O11
487.1821	C22H32O12	487.1821	C22H32O12	547.203229	C24H36O14	513.234136	C25H38O11
501.19775	C23H34O12	501.19775	C23H34O12	561.218879	C25H38O14	527.249786	C26H40O11
515.2134	C24H36O12	515.2134	C24H36O12	575.23453	C26H40O14	541.265436	C27H42O11
529.22905	C25H38O12	529.22905	C25H38O12	589.25018	C27H42O14	555.281086	C28H44O11
543.2447	C26H40O12	543.2447	C26H40O12	603.26583	C28H44O14	569.296736	C29H46O11
557.26035	C27H42O12	557.26035	C27H42O12	617.28148	C29H46O14	583.312386	C30H48O11
571.276	C28H44O12	571.276	C28H44O12	631.29713	C30H48O14	597.328036	C31H50O11
585.29165	C29H46O12	585.29165	C29H46O12	645.31278	C31H50O14	611.343686	C32H52O11
599.307301	C30H48O12	599.307301	C30H48O12	659.32843	C32H52O14	625.359336	C33H54O11
613.322951	C31H50O12	613.322951	C31H50O12	673.34408	C33H54O14	639.374986	C34H56O11
627.338601	C32H52O12	627.338601	C32H52O12	687.35973	C34H56O14	653.390636	C35H58O11
641.354251	C33H54O12	641.354251	C33H54O12	701.37538	C35H58O14	667.406286	C36H60O11
655.369901	C34H56O12	655.369901	C34H56O12	715.39103	C36H60O14	681.421936	C37H62O11
603.193059	C26H36O16	603.193059	C26H36O16	729.40668	C37H62O14	695.437586	C38H64O11
617.208709	C27H38O16	617.208709	C27H38O16	677.229838	C27H38O18	683.292044	C33H48O15
631.224359	C28H40O16	631.224359	C28H40O16	677.229838	C28H40O18	697.307694	C34H50O15
645.240009	C29H42O16	645.240009	C29H42O16	677.229838	C29H42O18	711.323345	C35H52O15
659.255659	C30H44O16	659.255659	C30H44O16	691.245488	C30H44O18	725.338995	C36H54O15

673.271309	C31H46O16	673.271309	C31H46O16	705.261138	C31H46O18	739.354645	C37H56O15
687.286959	C32H48O16	687.286959	C32H48O16	719.276788	C32H48O18	753.370295	C38H58O15
701.302609	C33H50O16	701.302609	C33H50O16	733.292438	C33H50O18	767.385945	C39H60O15
715.318259	C34H52O16	715.318259	C34H52O16	747.308088	C34H52O18	781.401595	C40H62O15
729.333909	C35H54O16	729.333909	C35H54O16	761.323738	C35H54O18	795.417245	C41H64O15
743.349559	C36H56O16	743.349559	C36H56O16	775.339389	C36H56O18	809.432895	C42H66O15
757.365209	C37H58O16	757.365209	C37H58O16	789.355039	C37H58O18	823.448545	C43H68O15
771.380859	C38H60O16	771.380859	C38H60O16	803.370689	C38H60O18	837.464195	C44H70O15
785.396509	C39H62O16	785.396509	C39H62O16	817.386339	C39H62O18	851.479845	C45H72O15
				831.401989	C40H64O18		
				845.417639	C41H66O18		
				859.433289	C42H68O18		

28

2.1.4 Sample composition descriptive terms

Double bond equivalents (DBE) describe the saturation degree of a molecular formula. It counts the number of double bonds and rings in the formula. DBE is determined by equation 2-1 (McLafferty and Tureek, 1993):

$$DBE = c - h/2 + n/2 + 1 \qquad\qquad\qquad 2\text{-}1$$

for each elemental composition $C_cH_hN_nO_oS_s$. A formula with higher DBE value is more unsaturated.

Kendrick mass analysis uses Kendrick mass defect (KMD) to sort the molecular formulas into homologous series. The most commonly used homologous series is CH_2 homologous series, which is an intrinsic character of natural organic matters (Reemtsma et al., 2006; Koch et al., 2007; Sleighter and Hatcher, 2008). The KMD is determined from equations 2-2 and 2-3 (Hughey et al., 2001):

$$\text{Kendrick mass (KM)} = \text{(measured mass)} \times (14.00000)/(14.01565) \qquad 2\text{-}2$$

$$\text{KMD} = \text{nominal Kendrick mass (NKM)} - \text{KM} \qquad\qquad 2\text{-}3$$

Using these equations, compounds of the same CH_2 homologous series will have the same KMD. When KMD is plotted against Kendrick mass, formulas of the same CH_2 homologous series form a horizontal line.

Van Krevelen analysis is used to understand the composition characteristics of complex formula dataset. All the individual formulas in the dataset are plotted on the van Krevelen diagram with the elemental oxygen-to-carbon ratio (O/C) as x-axis and hydrogen-to-carbon ratio (H/C) as y-axis. The degree of alkylation, hydrogenation, hydration and oxidation of complex samples can be evaluated with the van Krevelen diagram (Kim et al., 2003; Sleighter and Hatcher, 2007). Along the x-axis, compounds with a higher oxidation degree lie farther to the right. Along the y-axis, compounds with a higher saturation degree are located closer to the top of the diagram. Aliphatic compounds have high H/C ratios (≥ 1.5) and low O/C ratios (≤ 0.5); aromatic hydrocarbons have low H/C ratios (≤ 1.0) and low O/C ratios (≤ 0.5) (Mazzoleni et al., 2012).

2.1.5 Multivariate statistical analysis

2.1.5.1 Hierarchical cluster analysis

Hierarchical cluster analysis (HCA) finds the samples with similar molecular composition and clusters them closely together. It starts with individual samples. Thus there are initially as many clusters as samples. The most similar samples (samples with the smallest distance or largest similarity) are first grouped, and these initial groups are merged according to their similarities or distances. Eventually, all of the subgroups are fused into a single cluster (Johnson and Wichern, 2007). HCA was conducted using R software. The function *hcluster* in the package of "amap" (Lucas, 2014) was used for hierarchical cluster analysis in this study.

The basic tool for fusing the samples or sample groups is the similarity or distance measurement. In statistics, different distance measurements are applied to different situations to find the "natural" grouping of the observations. Around 10 distance measurement methods are available in the function *hcluster* in R, including *euclidean*, *maximum*, *manhattan*, *canberra*, *binary*, *pearson*, *correlation*, *spearman* and *kendall*. They were all tested to examine the similarities of the LANLSOA samples. The description of the distance methods can be found in the R help files at http://cran.r-project.org/web/packages/amap/amap.pdf.

When fusing individual samples and sample groups, linkage methods are applied to determine the distance between them. Commonly used linkage methods include *single*, *average* and *complete* linkages. The linkage criteria are illustrated schematically in Figure 2.2. Single linkage uses the smallest distance between the nearest members of two groups as the distance of the groups. Complete linkage uses the largest distance between the furthest members of the two groups as the distance of the groups. Average linkage uses the average distances between the members in two groups.

It is recommended to try different distance functions and linkage methods on a given problem. If the results are similar, then the result is considered to be "natural" (Johnson and Wichern, 2007).

Cluster distance

$$d_{24}$$

$$d_{15}$$

$$\frac{d_{13} + d_{14} + d_{15} + d_{23} + d_{24} + d_{25}}{6}$$
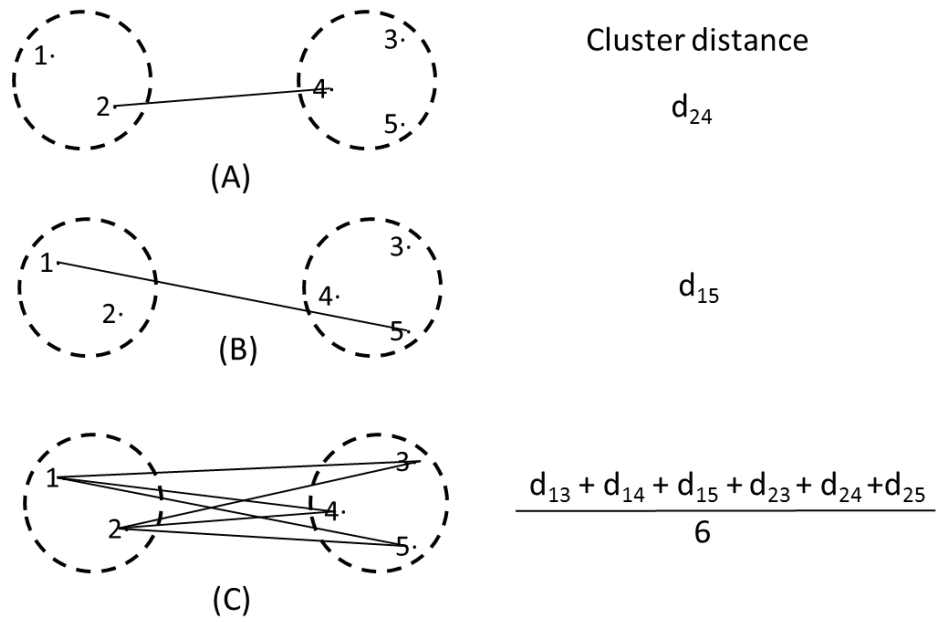
Figure 2.2 Intercluster distances for single linkage (A), complete linkage (B) and average linkage (C). Adapted from Johnson and Wichern, 2007.

2.1.5.2 Principal component analysis

Principal component analysis (PCA) is a dimension reduction tool. It uses a linear combination of the variables to generate new variables, which are called principal components (PCs). The PCs are ordered by the variance they carry, with PC1 representing the largest variance, PC2 representing second largest variance and so on. The number of the new variables (the PCs) is the same as the number of original variables. But usually only the first few PCs are selected for further statistical analysis. In this way the dimension of the dataset is reduced. In practice, a correlation matrix of the dataset is first generated. Then the eigenvalues and the eigenvectors of the correlation matrix are calculated. It turns out the eigenvectors are the new PCs and the corresponding eigenvalues are relevant to the variance the PCs carry. The eigenvector corresponding to the largest eigenvalue is PC1, the eigenvector corresponding to the second largest eigenvalue is PC2 and so on. When PCA is followed by further statistics, there is always the question of how many components to retain. There is no definitive answer to this question (Johnson and Wichern, 2007). A useful visual aid is the Scree plot. With the eigenvalues ordered form largest to smallest, a Scree plot is a plot of the magnitude of the eigenvalues (in some cases PCs are used) versus their number. An example of a Scree plot with 6 eigenvalues is shown in Figure 2.3. To determine the appropriate number of components, we look for an elbow (bend) in the Scree plot. The number of components is taken to be at the point where the remaining eigenvalues are relatively small and all about the same size (Johnson and Wichern, 2007). In Figure 2.3, the elbow occurs at approximately i = 3. Thus it appears that two (or perhaps three) sample principal components effectively summarize the total sample variance. There is a large number of calculations involved in the principal component analysis of the LANLSOA samples, due to the large LANLSOA dataset with thousands of formulas as variables. PCA was conducted using R software with the function of *prcomp* in the basic package.

In this study, PCA is mainly used to explore the relationships between the LANLSOA samples and the molecular formulas. Usually the first two or three PCs were selected to visually show the analysis results. The variable loadings and the sample scores for every PC are given in the PCA results. If PC1 and PC2 are selected, the formulas can be plotted on the PC1-PC2 space with their loadings; this is called a PCA loading plot. Similarly, the samples can also be plotted on the PC1-PC2 space with their scores; this is called a PCA score plot. When the variables and the samples are plotted together, the plot is a PCA biplot. The relationships between the samples and the

molecular formulas can be inferred from the PCA biplot based on their relative locations. Generally the variables at a similar location as the samples are the variables rich in those samples. In addition, the further away a variable is form the origin of the plot (in this case the variable has large absolute PC loading), the more contribution that variable has to the variation of the samples.
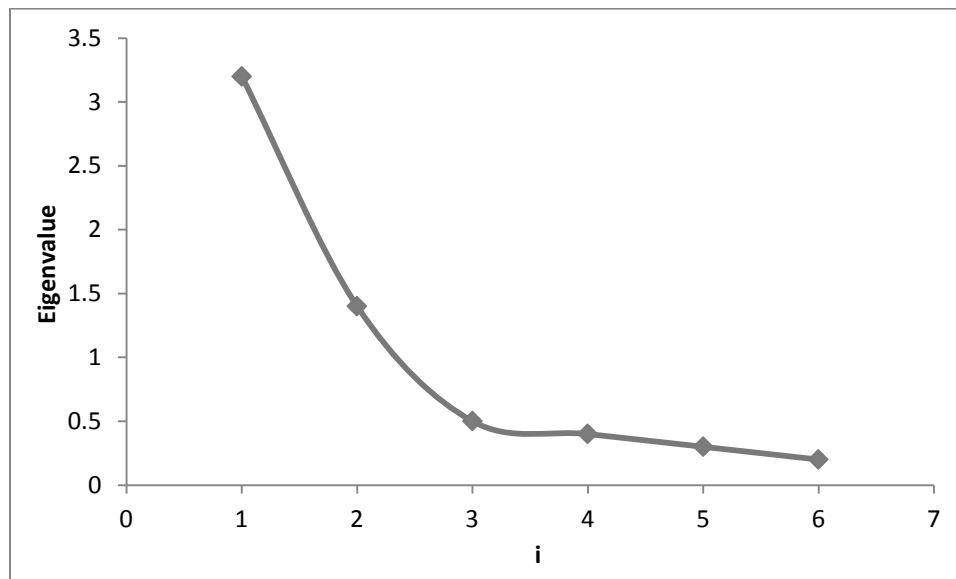
Figure 2.3 An example of Scree plot from principal component analysis with 6 eigenvalues or principal components. Adapted from Johnson and Wichern, 2007.

2.2 Ambient aerosol-derived water-soluble organic carbon

2.2.1 Ambient aerosol collection

The aerosol samples were collected at the Storm Peak Laboratory (SPL). This remote continental site is located on the western summit of Mt. Werner (3220 m a.s.l.) in the northwestern Colorado Park Range near Steamboat Springs. The mountain-top laboratory facilitates observations of free tropospheric air and in-cloud conditions. SPL is situated at the tree line on a 70 km ridge oriented perpendicular to the prevailing westerly winds (Hallar et al., 2013). SPL experiences transport from distant sources including urban areas, power plants, and wildfires (Obrist et al., 2008). There are also abundant biogenic emissions from the pine trees at SPL during summer months (Amin et al., 2012).

Thirty-four daily aerosol samples with aerodynamic diameter ≤ 2.5 μm ($PM_{2.5}$) were collected from June 25th, 2010 to July 28th, 2010. Four high-volume sampler (TE-5170-D, Tisch Environmental Inc., Cleves, OH) located on the roof of the SPL were operated at 1.13 m^3 min^{-1} flow rate for daily sampling with pre-cleaned 8 x 10 inches (20.3 x 25.4 cm) Teflon-impregnated glass fiber (TIGF) filters (Fibrefilm T 60A20, Pall Life Sciences, Port Washington, NY). Filters were changed daily at approximately 11 am Mountain Daylight Time (MDT). The loaded filters were packed with aluminum foil individually and stored at – 6 °C. The names of the samples and the exact sampling time are listed in Table 2.3.

Table 2.3 Daily aerosol sample names and the sampling time.

Sample name	Sampling Date	Start time	Stop time (next day)	Sampling period
0625	25-Jun-10	11:15	10:50	23 h 35 m
0626	26-Jun-10	11:04	10:40	23 h 35 m
0627	27-Jun-10	10:58	10:40	23 h 42 m
0628	28-Jun-10	10:55	13:14	26 h 19 m
0629	29-Jun-10	13:29	11:23	21 h 54 m
0630	30-Jun-10	11:46	10:45	22 h 59 m
0701	1-Jul-10	11:00	10:40	23 h 40 m
0702	2-Jul-10	11:00	10:48	23 h 48 m
0703	3-Jul-10	11:03	13:13	26 h 10 m
0704	4-Jul-10	13:34	10:43	21 h 09 m
0705	5-Jul-10	11:01	10:40	23 h 39 m
0706	6-Jul-10	10:55	10:24	23 h 29 m
0707	7-Jul-10	10:37	12:11	25 h 34 m
0708	8-Jul-10	12:31	11:20	22 h 49 m
0709	9-Jul-10	11:36	11:15	23 h 39 m
0710	10-Jul-10	11:35	10:43	23 h 08 m
0711	11-Jul-10	11:03	12:04	25 h 01m
0712	12-Jul-10	12:26	11:28	23 h 02m
0713	13-Jul-10	11:47	10:45	22 h 58 m
0714	14-Jul-10	11:02	10:37	23 h 35 m
0715	15-Jul-10	11:53	10:43	22 h 50 m
0716	16-Jul-10	10:56	10:41	23 h 45 m
0717	17-Jul-10	10:54	10:42	23 h 48 m
0718	18-Jul-10	10:55	10:41	23 h 46 m
0719	19-Jul-10	10:56	10:44	23 h 48 m
0720	20-Jul-10	10:57	10:57	24 h 00 m
0721	21-Jul-10	11:20	10:52	23 h 32 m
0722	22-Jul-10	11:11	10:42	23 h 31 m
0723	23-Jul-10	10:58	10:44	23 h 46 m
0724	24-Jul-10	11:02	10:42	23 h 40 m
0725	25-Jul-10	10:59	10:41	23 h 42 m
0726	26-Jul-10	10:55	10:40	23 h 45 m
0727	27-Jul-10	10:53	10:41	23 h 48 m
0728	28-Jul-10	10:58	10:41	23 h 43 m

2.2.2 Water-soluble organic carbon sample extraction and preparation

The water-soluble organic carbon (WSOC) in aerosols was obtained by extracting the aerosol-loaded filters with high purity water and isolated it from the inorganic water-soluble components using a reversed-phase solid phase extraction technique. No pH adjustment was applied to the extraction and isolation procedures. Studies showed that the addition of acid to atmospheric samples catalyzes the carbonyl chemistry (Gelencser, 2004) and accretion reactions to form oligomers (Iinuma et al., 2007; Altieri et al., 2008). Some studies used 0.1 M NaOH to extract aerosol organics, which could recover a larger mass fraction than extractions with pure water (Feczko et al., 2007; Baduel et al., 2009). In this study, we focus on the water-soluble organic carbon in aerosols, which is relevant to the hygroscopicity and indirect climate effect of aerosols. To keep the WSOC component intact and preserve its original composition as in the atmosphere, no acid or base was introduced in the sample preparation. This is consistent with other molecular composition studies of WSOC (Wozniak et al., 2008; Schmitt-Kopplin et al., 2010; Mazzoleni et al., 2012; Rincon et al., 2012).

A punch of 25mm diameter was taken from every TIGF filter. The four punches from the four filters collected on the same day (with four high-volume samplers) were put together in a pre-cleaned glass vial with tweezers. The tweezers were rinsed with isopropyl alcohol between uses. 10 mL of high purity water was added to the vial for extraction in an ultrasonic bath for 30 min. The extracts were filtered with 0.2 µm PTFE membrane filters (Whatman, Inc.). The filtered water-soluble aerosol extracts were placed in pre-cleaned amber bottles and stored at – 20 °C for further analysis steps. Besides the analysis of daily WSOC samples, aliquots of the extracts from different days were combined into several "composite samples" based on back trajectory analysis (Hallar et al., 2013). One of the composite samples (Sample #4) was discussed in Mazzoleni et al. (2012). In this study, we focus on the molecular composition of daily WSOC.

The daily extract of each sample was prepared for ESI FT-ICR MS analysis using solid phase extraction to concentrate the analytes and remove the inorganic ions, which compromise the analyte electrospray ionization efficiencies and lead to analyte artifacts (e.g., salt adducts). The Strata-X (Phenonmenex, Torrance, CA) cartridges were used for daily WSOC preparation. They contain both hydrophilic and hydrophobic functional groups, which were found to retain a diverse spectrum of analytes and. Each Strata-X cartridge was pre-conditioned with 2 mL of each

of the following solvents in series: methanol, acetonitrile and water. Then the filtered extract (~ 10 mL due to the extraction and filtration loss) was applied to the cartridge. The cartridges were rinsed with 2 mL of high purity water and dried. Then the WSOC was washed out with 1.5 mL of acetonitrile (Chromsolv for HPLC, Sigma-Aldrich). Extraction recoveries were not determined in this study, however a large fraction of the high molecular weight WSOC is expected to be retained due to the both hydrophilic and hydrophobic characteristics of the Strata-X cartridges. The blank sample was prepared following the same procedure. Samples were stored at – 5 °C until further analysis.

2.2.3 Water-soluble organic carbon sample analysis and molecular formula assignment

The daily WSOC samples were analyzed using an ultrahigh resolution FT-ICR MS. The instrument was described in Section 2.1.2. Samples were diluted with acetonitrile (50 sample: 50 acetonitrile, v/v) before direct injection into the ESI source to be better electrospray ionization. Negative ion mass spectra were recorded with a mass range of m/z 100 to 1000 and a resolving power of 400,000 (defined at m/z 400). The ESI conditions for each of the daily samples are listed in Table 2.4. The other instrument operation procedures were the same as given in Section 2.1.2. As previously mentioned, quantification of the WSOC components was not conducted in this study. Usually a sample was analyzed more than one time and the repeated runs were distinguished in the sample ID with "r" (sometimes also a number after the "r") after the sample date. For example, 0627 is the first run of the WSOC sample collected on June 27th, 2010; 0726r is the second run of the same sample; 0627r2 is the third run of the sample. Only the best of multiple runs was selected for further discussion in Chapter 4 for each sample. The selection criteria is based on the ionization consistency in the total ion current of the ~ 200 time domain transients, which implies the sample components electrosprayed well during the whole run. 8 of the 34 daily WSOC samples were analyzed twice (two "good" runs) and serve as analytical replicates. They include: 0628, 0707, 0710, 0713, 0717, 0722, 0724 and 0727. The sample replicates are indicated with "rep" after the date in the sample names. For example, 0707rep represents the analytical replicate for the sample 0707.

Table 2.4 The electrospray parameters for daily water-soluble organic carbon analysis.

Sample ID	Sample name	ESI needle voltage (kv)	Sample flow rate (ul/min)	Position of ESI probe
FilterBlank1	FilterBlank1	4.10	5	B1/2
FilterBlank2r	FilterBlank2	4.20	5	B1/2
FilterBlank3	FilterBlank3	3.80	5	B1/2
F062510r3	0625	3.90	4	B1/2
F062610r	0626	3.45	3	B1/2
F062710r3	0627	3.45	3	B1/2
F062810	0628	3.00	3	B1/2
F062810rep	0628rep	3.45	3	B1/2
F062910	0629	3.80	3	B1/2
F063010r2	0630	3.45	3	B1/2
F070110r	0701	3.80	4	C
F070210r2	0702	3.45	3	B1/2
F070310	0703	3.25	4	B
F070410	0704	3.35	4	C
F070510	0705	3.35	4	C
F070610	0706	3.45	4	C
F070710	0707	3.35	4	B1/2
F070710repr2	0707rep	3.55	3	B1/2
F070810r	0708	3.30	4	B1/2
F070910	0709	3.70	4	B1/2
F071010	0710	3.35	3	B1/2
F071010rep	0710rep	3.25	3	B1/2
F071110	0711	3.30	3	B1/2
F071210	0712	3.30	3	B1/2
F071310r	0713	4.00	3	B1/2
F071310rep	0713rep	3.35	3	B1/2
F071410	0714	3.10	3	B1/2
F071510	0715	3.10	3	B1/2
F071610	0716	3.10	3	B1/2
F071710	0717	3.10	3	B1/2
F071710rep	0717rep	3.10	3	B1/2
F071810	0718	3.10	3	B1/2
F071910r	0719	3.10	3	B
F072010r	0720	3.13	3	B
F072110	0721	3.10	3	B
F072210	0722	3.00	3	B1/2
F072210rep	0722rep	3.45	3	B1/2
F072310	0723	3.25	3	B
F072410	0724	3.50	3	B1/2
F072410rep	0724rep	3.50	3	B1/2
F072510	0725	3.40	3	B1/2
F072610	0726	3.50	3	B1/2
F072710	0727	3.05	3	B1/2
F072710rep	0727rep	3.45	3	B1/2
F072810	0728	3.45	3	B1/2

The molecular formula assignment was done with the same procedure as discussed in Section 2.1.3 for LANLSOA. Internal recalibration was applied using a list of the high relative abundance formulas in the daily WSOC samples. The internal recalibrants are listed in Table 2.5. The only difference between the formula assignment for LANLSOA and for daily WSOC is the type and number of elemental atoms allowed. Heteroatoms including nitrogen (N) and sulfur (S) were allowed in the formula assignment of the ambient WSOC samples. The molecular formula calculator was set to allow up to 70 carbon (C), 140 hydrogen (H) and 25 oxygen (O) 3 nitrogen (N) and 1 sulfur (S) atoms per molecular formula.

Table 2.5 Internal recalibrants used for daily WSOC formula assignment.

m/z	Formula	m/z	Formula
129.019332	C5H6O4	129.019332	C5H6O4
171.066282	C8H12O4	171.066282	C8H12O4
185.081932	C9H14O4	185.081932	C9H14O4
199.097583	C10H16O4	199.097583	C10H16O4
213.113233	C11H18O4	213.113233	C11H18O4
227.128883	C12H20O4	227.128883	C12H20O4
231.051026	C9H12O7	231.051026	C9H12O7
241.144533	C13H22O4	241.144533	C13H22O4
255.160183	C14H24O4	255.160183	C14H24O4
269.066676	C12H14O7	269.066676	C12H14O7
283.082326	C13H16O7	283.082326	C13H16O7
315.144927	C15H24O7	315.144927	C15H24O7
353.160577	C18H26O7	353.160577	C18H26O7
361.077635	C14H18O11	361.077635	C14H18O11
367.176227	C19H28O7	367.176227	C19H28O7
381.191877	C20H30O7	381.191877	C20H30O7
387.093285	C16H20O11	387.093285	C16H20O11
395.207527	C21H32O7	395.207527	C21H32O7
409.223177	C22H34O7	409.223177	C22H34O7
415.161008	C19H28O10	415.161008	C19H28O10
427.270127	C23H40O7	427.270127	C23H40O7
443.19226	C21H32O10	443.19226	C21H32O10
445.171535	C20H30O11	445.171535	C20H30O11
471.187185	C22H32O11	471.187185	C22H32O11
489.233038	C23H36O11	489.233038	C23H36O11
499.109329	C21H24O14	499.109329	C21H24O14
527.249786	C26H40O11	527.249786	C26H40O11
529.265436	C26H42O11	529.265436	C26H42O11
541.305192	C28H46O10		
555.171929	C25H32O14		
557.295639	C28H44O11		
559.311289	C28H46O11		
569.296736	C29H46O11		
569.339412	C30H50O10		
571.312386	C29H48O11		
585.328036	C30H50O11		
597.328036	C31H50O11		
611.23453	C29H40O14		

2.2.4 Sample composition descriptive terms

The sample composition descriptive terms introduced in Section 2.1.4 were also used here to describe the daily WSOC sample composition. Besides those, relative abundance weighted properties including relative abundance weighted oxygen-to-carbon ratios (O/C$_w$), relative abundance weighted hydrogen-to-carbon ratios (H/C$_w$), relative abundance weighted double bond equivalents (DBE$_w$) and relative abundance weighted organic mass to organic carbon ratios (OM/OC$_w$). They were determined using the following equations (Sleighter and Hatcher, 2008; Bateman et al., 2009):

O/C$_w$= Σ($w_i o_i$)/ Σ($w_i c_i$) 2-4

H/C$_w$= Σ($w_i h_i$)/ Σ($w_i c_i$) 2-5

DBE$_w$= Σ($w_i DBE_i$)/ Σw_i 2-6

OM/OC$_w$ = Σ($w_i OM/OC_i$)/ Σw_i 2-7

for elemental compositions of $C_c H_h N_n O_o S_s$, where w_i is the relative abundance for each individual molecular formula, i.

2.2.5 Multivariate statistical analysis

Multivariate statistical analysis including hierarchical cluster analysis and principal component analysis were applied to the daily WSOC. The methods of conducting the statistical analysis are the same as the analysis of the LANLSOA dataset and described in detail in Section 2.1.5.

2.3 Cloud water

2.3.1 Cloud water sample collection

To investigate the composition of cloud water AOM, samples of ambient clouds consisting of supercooled droplets were collected at the SPL. Clouds that surround the SPL with supercooled droplets form in the wintertime (Lowenthal et al., 2002). A custom cloud sieve with stainless steel strands of 0.5 mm diameters was used to collect supercooled droplets by impaction (Figure 2.4). As described by Hindman et al. (1992), cloud droplets with diameter > 8µm will be collected with a 50% collection efficiency on the 0.5 mm strand diameter with an average wind speed of 2 m/s (average wind speed during sample collection time periods). Cloud sieves were mounted on the SPL rooftop deck railing on the west side during cloud events. After ~20

minutes, the cloud sieves were carried into a cold laboratory space to remove the frozen cloud water (collected as rime) from the strands using a stainless steel blade and collection tray. The collected rime was stored in a freezer at -5°C. Four cloud water samples were collected during short separate cloud events. The first sample (CW1) is a composite of two samples collected between 8:30 and 9:00 pm on February 24th, 2010 and 7:00 and 8:00 am on February 25th. The second composite sample (CW2) is a composite of two samples, collected between 8 and 11 am on February 26th. The 24-hour air back trajectory analyses using NOAA HYSPLIT model (Draxler and Rolph) showed the air coming from west and northwest of SPL. The particle concentrations (with diameter > 3 nm) were approximately 1000/cm3 during the sampling time as determined by a Condensation Particle Counter (TSI, model #3025). One field blank sample was prepared by leaving the cloud sieve in the air in the absence of a cloud event and then rinsed with high purity water.
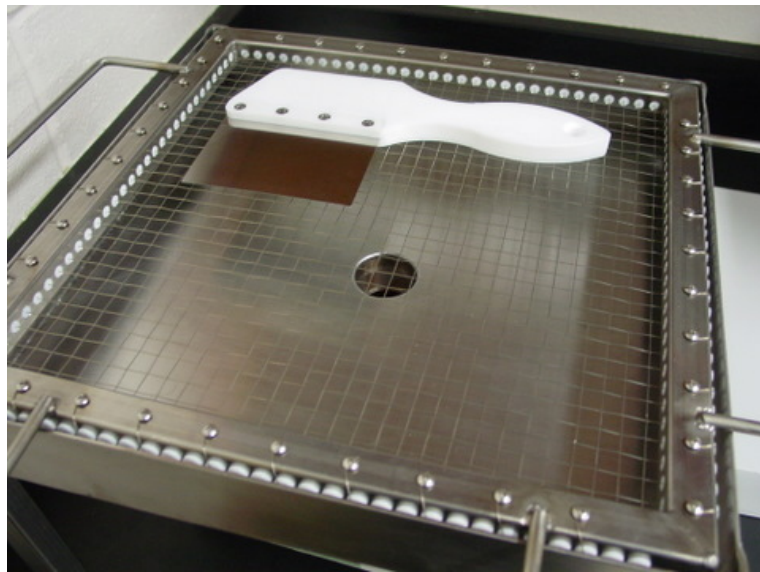
Figure 2.4 Photo of the custom stainless steel and teflon cloud sieve and rime collection tray.

2.3.2 Cloud sample preparation

Cloud water samples were prepared for analysis using a reversed phase solid phase extraction method to concentrate the analytes and remove inorganic ions (Mazzoleni et al., 2010). Strata-X SPE cartridges were conditioned with consecutive applications of the following: 1 mL of isopropyl alcohol, 2 mL of acetonitrile, 2 mL of acidified methanol containing 0.1% formic acid, and 2 mL of aqueous 0.1% formic acid. Then 100 mL of cloud water without pH adjustment was applied at a rate of 1-2 mL min^{-1} to the cartridge to allow the AOM components to partition to the SPE stationary phase. pH values of the cloud water samples were not determined in this study. Hindman et al. (2006) measured the pH values of wintertime clouds collected at Storm Peak Laboratory from 1983 to 2004 and reported a pH range of 3.7 to 5.7, thus the cloud samples in this study are expected to be acidic. Then the cartridges were rinsed with high purity acidified water to remove salts and dried. Some of the low molecular weight analytes (< 100 Da) are expected to be lost in this step (Samburova et al., 2013). Analytes were eluted with 1.5 mL of acetonitrile. Extraction recoveries were not determined in this study, however a large fraction of the high molecular weight cloud water organic carbon is expected to be retained due to the both hydrophilic and hydrophobic characteristics of the Strata-X cartridges. The field blank and lab blank were prepared using the same method except that 100 mL of high purity water without pH adjustment was applied to the SPE cartridge for the lab blank. All solvents were HPLC grade or higher. Prepared samples were stored at -5°C until further analysis.

2.3.3 Cloud sample analysis

Ultrahigh resolution mass spectrometry analysis was performed on the cloud samples. Cloud water samples were infused directly into the ESI interface, the ESI parameters were adjusted to obtain a stable ion current with minimum ion injection time into the mass analyzer. After optimization the infusion flow rate was 4 µl/min, the ESI needle voltage was -3.7kV, and the capillary temperature was 265 °C. Negative ion mass spectra were collected using full scan analysis for the range of 100 < m/z < 1000. Mass resolving power was set at 200,000 (at m/z 400) for all spectra. Other procedures are similar in Section 2.1.2. Quantification of the cloud water components was not conducted in this study due to uncertainties associated with electrospray efficiencies and an overall lack of suitable analytes for quantitation.

2.3.4 Data processing and assignment of molecular formula composition

Ultrahigh resolution FT-ICR mass spectra were processed using Composer software (Sierra Analytics, Version 1.0.5) as discussed in Section 2.1.4 and 2.2.4. A set of naturally occurring cloud water analytes from the two cloud water samples was selected to represent internal recalibrants over the mass range of 100-650 Da. The internal recalibrants are listed in Table 2.6. The match tolerances for the internal recalibrants were set at 3 ppm with a minimum relative abundance (RA) of 18 times the root mean square (RMS) of the signal-to-noise ratio between $900 < m/z < 1000$ (RMS of signal-to-noise ratio was 0.018 % for CW1 and 0.038 % for CW2). The molecular formula calculator was set to allow up to 70 carbon (C), 140 hydrogen (H), 25 oxygen (O), 4 nitrogen (N) and 1 sulfur (S) atoms per molecular formula composition. The other formula assignment procedures are the same as described in Section 2.1.4.

Table 2.6 Internal recalibrants used for cloud sample formula assignment.

m/z	Formula
129.019357	C5H6O4
143.035007	C6H8O4
157.050657	C7H10O4
171.066307	C8H12O4
185.081958	C9H14O4
199.097608	C10H16O4
213.113258	C11H18O4
227.128908	C12H20O4
241.144558	C13H22O4
255.160208	C14H24O4
269.175858	C15H26O4
283.191508	C16H28O4
297.207158	C17H30O4
311.222808	C18H32O4
325.238458	C19H34O4
367.285408	C22H40O4
259.045966	C10H12O8
273.061616	C11H14O8
287.077266	C12H16O8
301.092916	C13H18O8
315.108566	C14H20O8
329.124216	C15H22O8
343.139866	C16H24O8
357.155516	C17H26O8
371.171166	C18H28O8
385.186817	C19H30O8
399.202467	C20H32O8
413.218117	C21H34O8
427.233767	C22H36O8
441.249417	C23H38O8
455.265067	C24H40O8
469.280717	C25H42O8
483.296367	C26H44O8
497.312017	C27H46O8
511.327667	C28H48O8

Molecular formulas were assigned to the field and lab blank samples using the same method. The blank mass spectra contained several low molecular weight negative ions (m/z < 400), most of which were assigned unambiguous molecular formulas as described above. Approximately 800 of the assigned monoisotopic formulas were in common with the cloud water samples CW1 and CW2. The common molecular formulas represent approximately 20% of the assigned molecular formulas in cloud water samples. Although some carryover of the sample components to blanks may have occurred (Mazzoleni et al., 2012), other sources of contamination could not be ruled out. Thus, all of the common molecular formulas assigned to negative ions in either of the blank samples were removed from the presented cloud water composition.

Chapter 3 Molecular characterization of biogenic secondary organic aerosol

A total of 19 biogenic secondary organic aerosol samples were generated in an aerosol chamber at the Los Alamos National Laboratory (LANLSOA samples). The samples resulted from the ozonolysis of individual biogenic SOA precursors including α-pinene, β-pinene, D-limonene and β-caryophyllene with varied relative humidity conditions (0%, 4% and 30%) and with or without cyclohexane as a hydroxyl radical (OH) scavenger. Ultrahigh resolution FT-ICR mass spectrometry was used to analyze the molecular composition of the biogenic SOAs and so far two manuscripts have resulted from this dataset. Kundu et al. (2012) reported the composition of D-limonene SOA generated with low relative humidity (RH= 0 and 4%) and without cyclohexane (sample LM1_0 and LM2_4). The paper focused on the formation of high molecular weight SOA components. Another paper by Kundu et al. (In preparation) reports the molecular composition of β-caryophyllene SOA. In this dissertation, the molecular composition of all 19 LANLSOA samples is reported. In Section 3.1, I will discuss the chemical characterization of the biogenic SOA samples by the means of the mass spectra, Kendrick mass defect analysis, van Krevelen plots, carbon number trends analysis and so on. In Section 3.2, I will compare the molecular composition of the SOA samples using multivariate statistical analysis. Various data preparation methods were tested, including RA normalization and RA standardization. The formulas characteristic to the SOA samples from different precursors were explored. The potential biogenic SOA indicator species were verified by comparison with the ambient atmospheric samples of several other studies.

3.1 Composition of biogenic secondary organic aerosol

3.1.1 Molecular formula assignments

After the molecular formula assignment and quality control procedures were applied as described in Section 2.1.3, approximately 1000 monoisotopic molecular formulas with molecular weights in the range of 100 - 1000 Da were assigned to each of the LANLSOA samples. They contain the elements of carbon (C), hydrogen (H) and oxygen (O). Molecular formulas containing ^{13}C were also identified and corresponded to the monoisotopic formulas, which provides an intrinsic validation of the formula assignments.
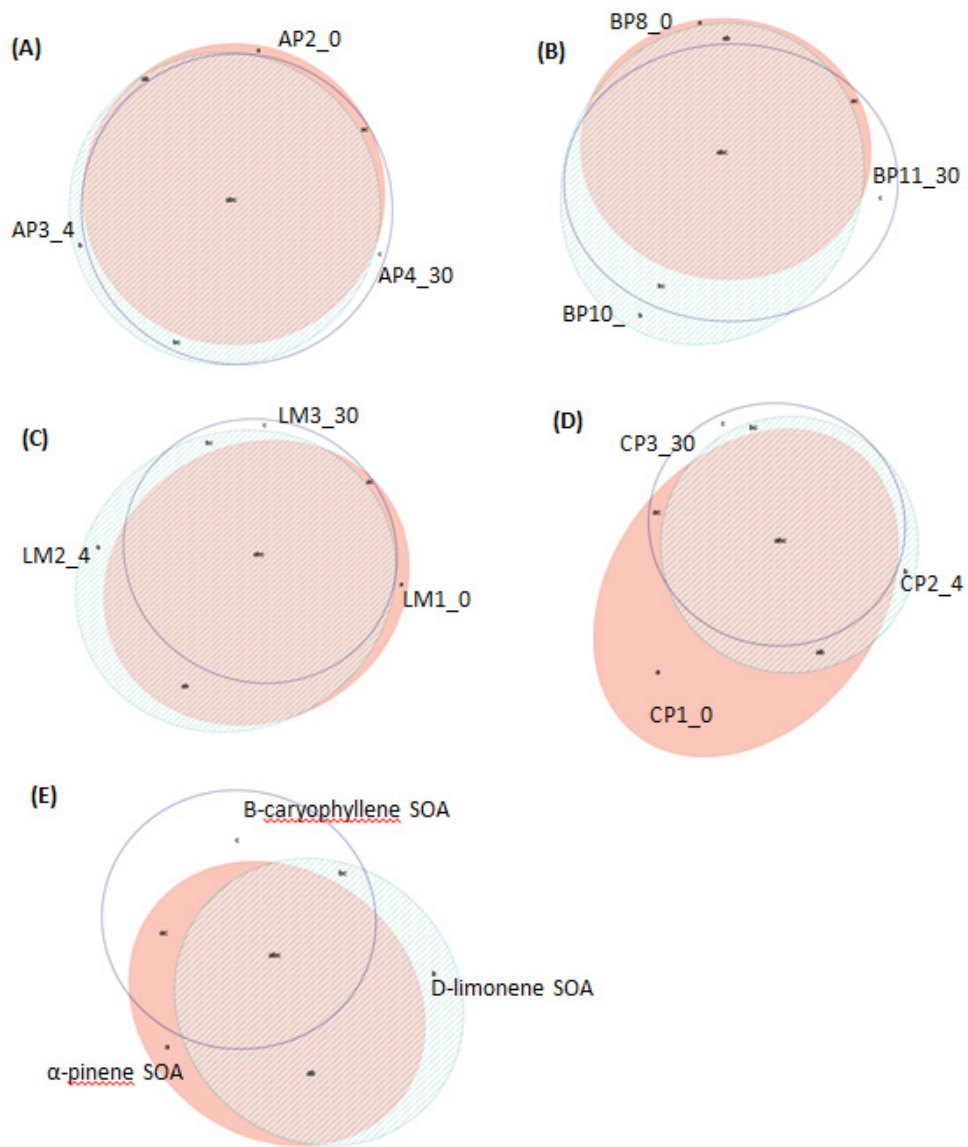
Figure 3.1 The Venn diagrams for the α-pinene SOA (A), β-pinene SOA (B), D-limonene SOA (C), β-caryophyllene SOA (D) and LANLSOA (E) samples.

A large number of common molecular formulas were observed among the SOA samples, especially those from the same precursor. Area proportional Venn diagrams were created to show the number of common and unique molecular formulas in the SOA samples (Figure 3.1). The three α-pinene SOA samples (AP2_0, AP3_4 and AP4_30) share 1107 common molecular formulas (the common area of the three ellipses in Figure 3.1 A), which is much more than the number of unique formulas for each of the α-pinene SOA samples. There are around 800 formulas common in the three β-pinene SOA samples: BP8_0, BP10_4 and BP11_30 (BP9_0 is an experimental replicate of BP8_0 and is not included in the Venn diagram) (Figure 3.1 B). Three D-limonene SOA samples (LM1_0, LM2_4 and LM3_30) without cyclohexane were used to create the Venn diagram in Figure 3.1 C. Again the largest fraction of the diagram is the common area of the three ellipses, which corresponds to the common formulas in the three D-limonene samples. Similarly, three β-caryophyllene SOA samples (CP1_0, CP2_4 and CP3_30) without cyclohexane were used to create the Venn diagram for β-caryophyllene SOA samples in Figure 3.1 D. Though the common area of the three ellipses is still the largest fraction in the diagram as in Figure 3.1 A to C, there are many unique formulas in the sample CP1_0. This observation reflects the effect of RH on the β-caryophyllene SOA composition.

The molecular composition of the SOA samples from different precursors is also compared using the Venn diagram (Figure 3.1 E). Due the high similarity between α-pinene and β-pinene SOA sample composition, the Venn diagram was created with α-pinene, D-limonene and β-caryophyllene SOA samples. Only the common formulas in all the three α-pinene SOA samples in Figure 3.1 A, the common formulas in all the three D-limonene SOA samples in Figure 3.1 C and the common formulas in all the three β-caryophyllene SOA samples in Figure 3.1 D were used in Venn diagram (Figure 3.1 E) generation. A large fraction of the α-pinene SOA and D-limonene SOA components are in common. B-caryophyllene SOA shows similarities to α-pinene and D-limonene SOA, but approximately 1/3 of the β-caryophyllene SOA formulas are unique. Overall, the SOA samples from different precursors show similar molecular composition, but differences are observed between β-caryophyllene SOA and the other types of SOA in terms of precursors.

It should be noted that we use the molecular formulas in SOA composition comparison. A molecular formula might represent various organic compounds with different chemical structures. Due to the high complexity of atmospheric sample composition, it is almost

impossible with the currently available analytical techniques to achieve complete speciation of the atmospheric molecular species. Even though the molecular formulas do not reflect chemical structures, they still provide much insight toward the molecular composition of atmospheric samples.

3.1.2 Characteristics of the biogenic SOA components

The reconstructed mass spectra of the monoisotopic molecular formulas for 18 LANLSOA samples are shown in Figure 3.2. Samples BP9_0 and BP8_0 were generated with the same experimental conditions, so the reconstructed mass spectrum of BP9_0 is not shown here. Consistent with the high number of monoisotopic molecular formulas identified in all of the samples, the mass spectra are quite complex with a large number of isobaric ions.
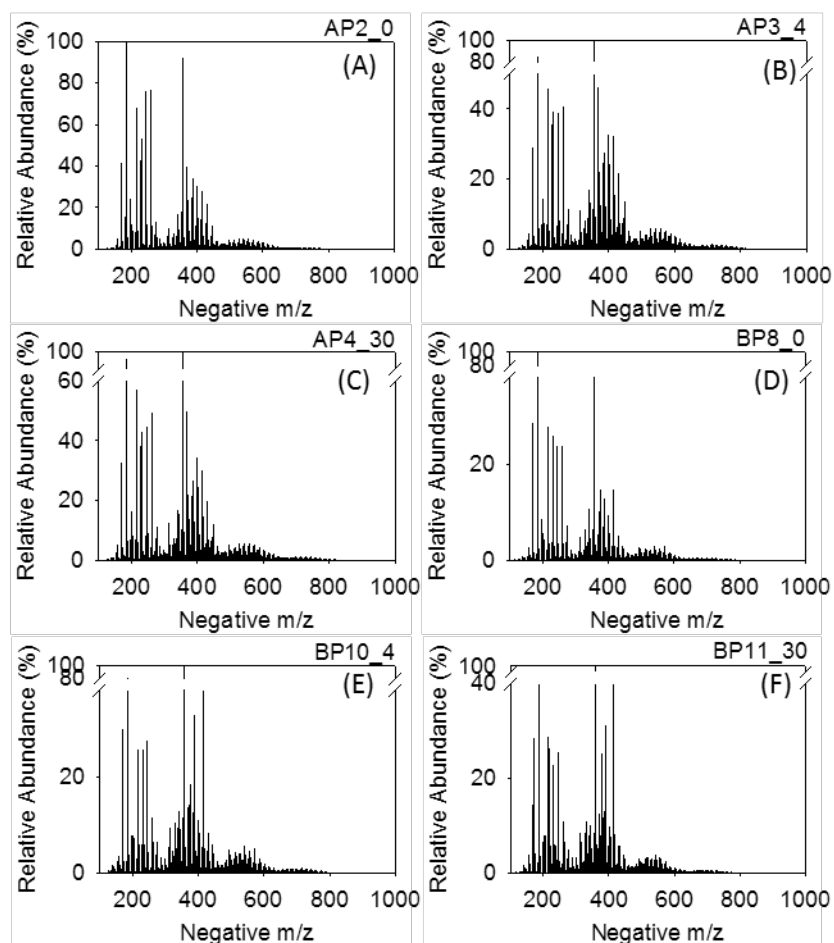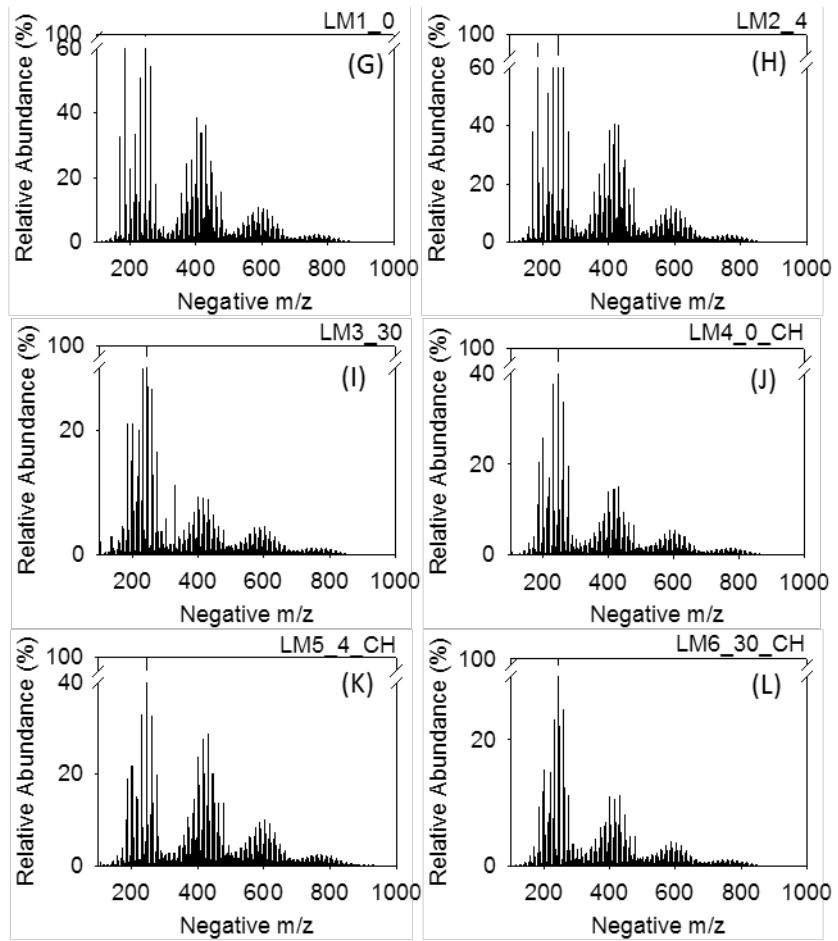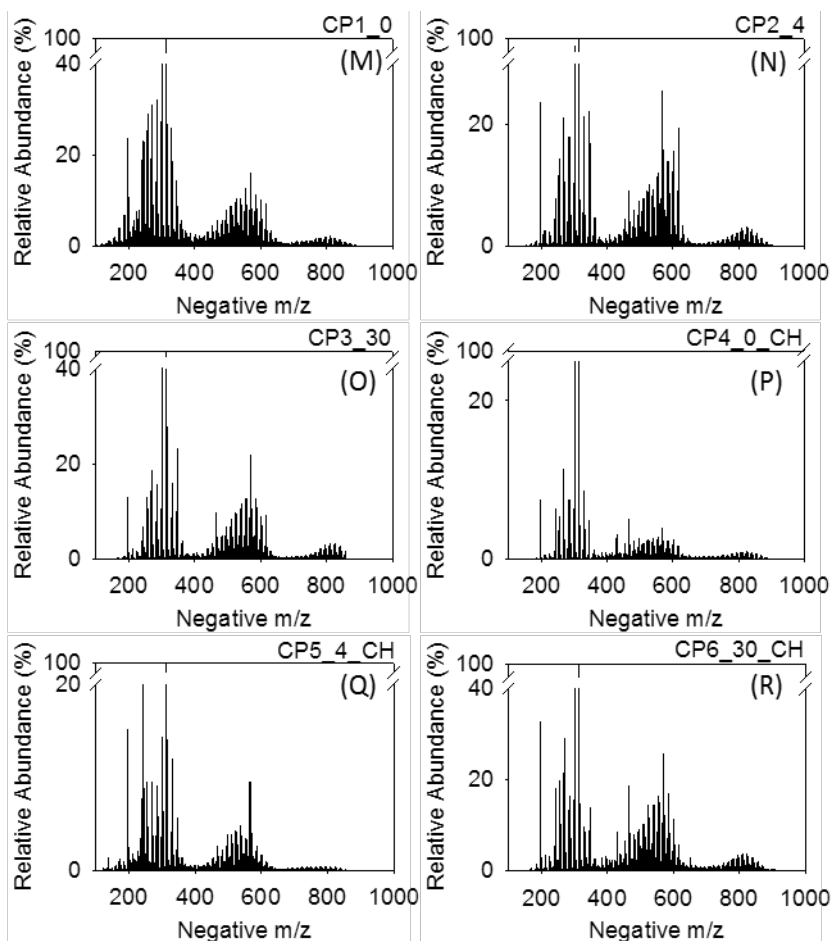
Figure 3.2 Reconstructed mass spectra of LANLSOA sample AP2_0 (A), AP3_4 (B), AP4_30 (C), BP8_0 (D), BP10_4 (E), BP11_30 (F), LM1_0 (G), LM2_4 (H), LM3_30 (I), LM4_0_CH (J), LM5_4_CH (K), LM6_30_CH (L), CP1_0 (M), CP2_4 (N), CP3_30 (O), CP4_0_CH (P), CP5_4_CH (Q) and CP6_30_CH (R).

A clear characteristic observed in all of the mass spectra is the clustering of the analyte peaks into groups, regardless of the relative humidity and presence or absence of the OH radical scavenger used to generate the SOA samples. This clustering of peaks has been previously observed in studies of α-pinene SOA (Tolocka et al., 2004; Reinhardt et al., 2007; Gao et al., 2010; Putman et al., 2012; Kourtchev et al., 2014), β-pinene SOA (Gao et al., 2010) and D-limonene SOA (Walser et al., 2008; Bateman et al., 2009; Kundu et al., 2012) using ultrahigh and high resolution MS techniques. This is especially interesting because the studied SOA were generated under various experimental conditions, some even with acidic seed aerosols. The clustering characteristic is also observed in the β-caryophyllene SOA mass spectra included in this study. The clustering of the peaks in the mass spectra is due to accretion reactions of

functionalized products that lead to high molecular weight SOA formation (also referred to as "oligomer formation"). Tolocka et al. (2004) referred to the clusters as monomer, dimer, trimer and so forth with the widely-held assumption that the accretions occur predominantly between first generation SOA products thus a "terpenoid" structure is preserved. In Putman et al. (2012) the analyte clusters are referred to as Group I, Group II, Group III and so on with the consideration that multiple accretions involving the Criegee intermediates, hydroperoxides and terpenoid degradation products and they do not necessary retain the terpenoid structure. In addition to the expected "terpenoid" structure of the high molecular weight species, many compounds without the terpenoid structures are also expected in the highly complex mixture. Thus in this dissertation, I will use this terminology to be consistent with Putman et al., (2012) and Kundu et al. (2012; 2014 In preparation).

Four groups of analyte clusters in the m/z range from 100 to 1000 are observed in the monoterpene (α-pinene, β-pinene and D-limonene) SOA mass spectra. They are defined as follows: Group I: 100 < m/z < 300; Group II: 300 < m/z < 475; Group III: 475 < m/z < 650 and Group IV: 650 < m/z < 825. Three groups of peaks are observed in the β-caryophyllene SOA mass spectra, which are defined as: Group I: 100 < m/z < 400; Group II: 400 < m/z < 700 and Group III: 700 < m/z < 1000. A higher molecular weight range is observed in the sesquiterpene SOA mass spectra because of the higher molecular weight of sesquiterpene precursor ($C_{15}H_{24}$ with molecular weight 204 Da) compared to the monoterpene precursor ($C_{10}H_{16}$ with molecular weight 136 Da). Generally, the relative abundances (RAs) of the tallest species in each group decrease from Group I to Group IV. An exception is observed in the α- and β-pinene SOA mass spectra where the tall peaks in Groups I and II have comparable RAs.

The analyte cluster characteristic of the biogenic SOA components is also observed in the Kendrick mass defect analysis shown in Figure 3.3. These Kendrick mass defect plots show CH_2-homologous series of the identified formulas. The CH_2-homologous series are a series of molecular formulas with the same elemental composition except they differ by the number of CH_2 units. Thus, those with the same Kendrick mass defect compose a straight horizontal line in the diagram. The CH_2-homologous series observed in all of the SOA samples provide an intrinsic confirmation of the molecular formula assignments. To examine this further, SOA samples from each precursor were selected with 4% and 30% relative humidity and without OH scavenger (Figure 3.3). The size of the symbols (bubbles) represents the molecular formula RAs for each of

the specific samples. The Groups I to IV of the monoterpene SOA samples and Groups I to III in the sesquiterpene SOA samples are clearly observed in the diagrams. Some of the homologous series span up to 3 of the Groups. The Kendrick mass defect of the SOA components increases with the molecular weight. Since the molecular formulas contain C, H and O and C does not contribute to the Kendrick mass defect (exact mass of ^{12}C=12.0000 Da), H and O contribute to a negative and positive Kendrick mass defect. This is because the exact mass of ^{1}H and ^{16}O are 1.0078 Da and 15.9949 Da, which are more or less than their nominal masses of 1 and 16 Da. Thus, as the Kendrick mass defect increases, so does the number of oxygen atoms and vice versa. This is confirmed by examining the average numbers of O in each of the Group (Table 3.1). The average O number increases from ~6 to ~16 from Group I to IV for the monoterpene SOA (or I to III for the sesquiterpene SOA), indicating a higher number of oxygen-containing functional groups in the oligomers than in the monomers.
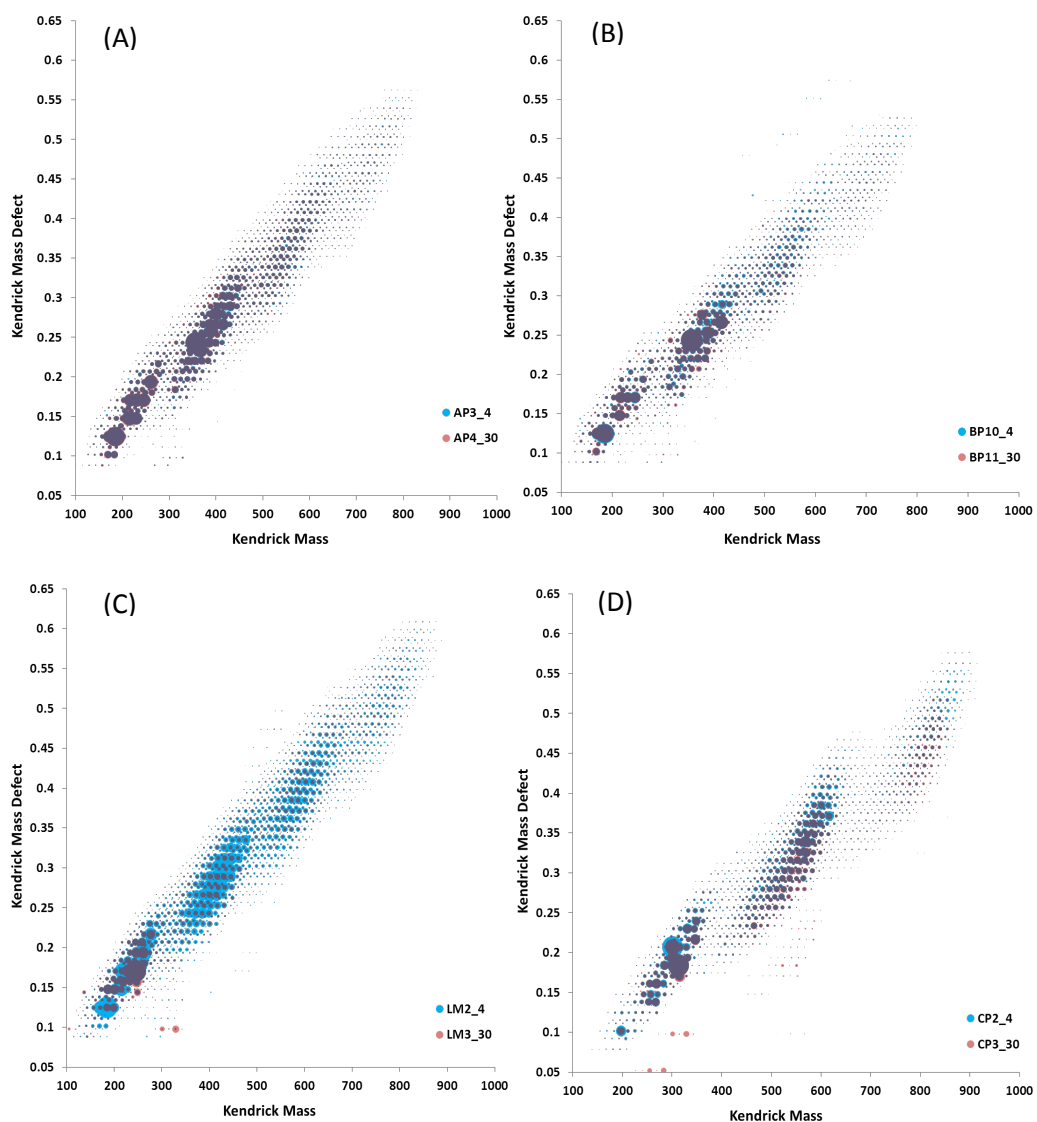
Figure 3.3 Kendrick mass defect diagrams for selected α-pinene SOA samples (A), β-pinene SOA samples (B), limonene SOA samples (C) and β-caryophyllene SOA samples (D). The symbol sizes are scaled on a sample basis based on the molecular formula RAs.

Table 3.1 The range and average values with their standard deviations of the DBE, oxygen number, oxygen-to-carbon ratio and hydrogen-to-carbon ratio for each of the LANLSOA samples generated under 4% and 30% relative humidity without OH scavenger.

	DBE		#O		O/C		H/C	
	Range	Avg.±Std	Range	Avg.±Std	Range	Avg.±Std	Range	Avg.±Std
α-Pinene								
AP3_4								
Group I	1 to 7	3.44±1.30	3 to 10	5.53±1.67	0.17 to 1.25	0.57±0.23	0.86 to 2	1.52±0.24
Group II	1 to 9	5.15±1.67	4 to 14	8.68±2.09	0.21 to 0.91	0.48±0.15	1.2 to 2	1.55±0.17
Group III	4 to 12	7.08±1.52	7 to 17	12.25±2.23	0.24 to 0.75	0.47±0.11	1.27 to 1.76	1.55±0.10
Group IV	6 to 11	8.81±1.23	11 to 19	15.42±2.08	0.29 to 0.64	0.45±0.08	1.43 to 1.68	1.55±0.06
Overall	1 to 11	6.18±2.53	3 to 19	10.59±4.30	0.17 to 1.25	0.49±0.17	0.86 to 2	1.55±0.17
AP4_30								
Group I	1 to 7	3.45±1.26	3 to 10	5.65±1.67	0.17 to 1.25	0.58±0.24	0.86 to 2	1.52±0.23
Group II	1 to 9	5.15±1.69	4 to 14	8.76±2.06	0.21 to 0.91	0.49±0.15	1.2 to 2	1.55±0.17
Group III	4 to 11	7.02±1.48	7 to 17	12.30±2.20	0.25 to 0.75	0.47±0.11	1.27 to 1.76	1.55±0.10
Group IV	6 to 11	8.74±1.30	11 to 19	15.56±1.99	0.3 to 0.67	0.45±0.08	1.41 to 1.7	1.55±0.07
Overall	1 to 11	6.18±2.31	3 to 19	10.72±3.900	0.17 to 1.25	0.49±0.15	0.86 to 2	1.54±0.15
β-Pinene								
BP10_4								
Group I	1 to 7	3.47±1.41	3 to 9	5.40±1.59	0.17 to 2	0.56±0.25	0.86 to 2	1.52±0.26
Group II	1 to 8	5.19±1.59	4 to 13	8.40±1.96	0.18 to 0.77	0.46±0.13	1.22 to 2	1.55±0.16
Group III	3 to 12	6.95±1.50	7 to 23	12.31±2.44	0.25 to 1.33	0.48±0.16	1 to 1.77	1.55±0.11
Group IV	4 to 11	8.50±1.16	11 to 23	15.10±1.90	0.3 to 1.05	0.44±0.08	1.4 to 1.73	1.56±0.07
Overall	1 to 11	5.91±2.25	3 to 23	10.06±3.95	0.17 to 2	0.48±0.17	0.86 to 2	1.54±0.16
BP11_30								
Group I	1 to 6	3.22±1.33	3 to 9	5.48±1.62	0.17 to 1.25	0.58±0.24	0.86 to 2	1.56±0.26
Group II	1 to 8	4.88±160	3 to 20	8.53±2.18	0.13 to 2	0.47±0.18	1.2 to 2	1.58±0.16
Group III	4 to 11	6.77±1.37	7 to 17	12.08±1.98	0.25 to 0.7	0.46±0.10	1.27 to 1.77	1.57±0.10
Group IV	6 to 12	8.31±1.07	11 to 18	14.70±1.75	0.3 to 0.59	0.43±0.07	1.31 to 1.7	1.58±0.06
Overall	1 to 12	5.58±2.23	3 to 18	9.81±3.71	0.17 to 2	0.49±0.17	0.86 to 2	1.57±0.17
D-Limonene								
LM2_4								
Group I	1 to 9	3.71±1.60	2 to 10	5.47±1.71	0.17 to 2	0.57±0.26	0.75 to 2	1.48±0.28
Group II	1 to 10	5.26±1.96	4 to 14	8.84±2.12	0.21 to 0.92	0.50±0.15	1.07 to 2	1.53±0.20
Group III	3 to 12	7.07±1.92	8 to 18	12.71±2.12	0.27 to 0.75	0.49±0.11	1 to 1.82	1.53±0.15
Group IV	1 to 12	8.92±1.77	12 to 22	16.90±2.25	0.32 to 0.96	0.48±0.09	1.27 to 2	1.57±0.09
Overall	1 to 12	6.42±2.54	2 to 22	11.56±4.57	0.17 to 2	0.50±0.15	0.75 to 2	1.53±0.19
LM3_30								
Group I	1 to 6	3.27±1.45	3 to 10	5.78±1.60	0.25 to 2	0.63±0.25	0.86 to 2	1.55±0.27
Group II	1 to 8	5.08±1.61	4 to 14	9.19±2.00	0.2 to 0.91	0.52±0.14	1.2 to 2	1.54±0.17
Group III	3 to 10	6.73±1.70	6 to 17	12.96±2.00	0.21 to 0.74	0.51±0.10	1.28 to 1.86	1.56±0.13
Group IV	5 to 11	8.57±1.50	13 to 20	16.92±1.72	0.37 to 0.67	0.49±0.06	1.38 to 1.74	1.56±0.08
Overall	1 to 11	6.11±2.56	3 to 20	11.61±4.61	0.20 to 2	0.53±0.16	0.86 to 2	1.55±0.17
β-caryophyllene								
CP2_4								
Group I	2 to 7	4.05±1.24	2 to 11	6.33±2.07	0.14 to 0.78	0.44±0.14	1.29 to 1.88	1.59±0.14
Group II	3 to 10	6.74±1.60	5 to 17	10.70±2.46	0.18 to 0.65	0.39±0.10	1.36 to 1.88	1.59±0.10
Group III	7 to 12	9.26±1.30	10 to 19	14.86±2.29	0.24 to 0.49	0.36±0.06	1.46 to 1.71	1.60±0.06
Overall	2 to 12	6.89±2.32	2 to 19	10.96±3.78	0.14 to 0.78	0.39±0.10	1.29 to 1.88	1.59±0.10
CP3_30								
Group I	1 to 8	3.96±1.34	2 to 11	5.99±2.04	0.11 to 0.73	0.41±0.14	1.29 to 2	1.60±0.16
Group II	1 to 10	6.44±1.72	4 to 16	10.21±2.49	0.11 to 0.65	0.37±0.10	1.36 to 2	1.61±0.11
Group III	6 to 12	9.03±1.32	10 to 18	13.92±1.93	0.23 to 0.46	0.34±0.06	1.46 to 1.79	1.61±0.06
Overall	1 to 12	6.50±2.32	2 to 18	10.14±3.52	0.11 to 0.73	0.37±0.10	1.29 to 2	1.61±0.12

In addition to the differences in the O numbers among the SOA Groups, other chemical property differences between the Groups can be used to infer the accretion reaction (aka "oligomerization") pathways. The elemental ratios of oxygen-to-carbon (O/C) and hydrogen-to-carbon (H/C) indicate the degree of oxidation and saturation of the species. O/C and H/C for each of the sample Groups for selected LANLSOA samples are shown as box plots in Figure 3.4 and the averages and ranges are listed in Table 3.1. Generally, the average O/C ratio decreases from ~0.60 in Group I to ~0.45 in Group IV of the monoterpene SOA and from ~0.45 in Group I to ~0.35 in Group III of β-caryophyllene SOA. The H/C ratios remain steady or in some cases slightly increase to 1.50-1.60 for the monoterpene SOA and ~1.60 for the sesquiterpene SOA. Reinhardt et al. (2007) and Putman et al. (2012) observed the same O/C and H/C trends in α-pinene SOA samples from different chamber experiments. Likewise, Kundu et al. (2012) observed the same trends in the limonene SOA of the LANLSOA dataset. Here in this study, the trends remain the same for all of the SOA samples regardless of the experimental conditions, reflecting similar accretion reactions (or oligomer formation pathways) in different biogenic SOA experiments in terms of the precursors and the chamber conditions. The molecular formula DBE values reflect the number of double bonds and rings in the potential structures, which also indicates the degree of saturation. Putman et al. (2012) reported an increase in the DBE values from Group I to IV in α-pinene SOA; similarly both Bateman et al. (2009) and Kundu et al. (2012) observed increases in the DBE values from Group I to IV in D-limonene SOA. An increase in the DBE values with respect to the Group numbers or molecular weight is observed in all of the SOA samples from this study (Table 3.1). The average DBE values ranged from ~3 to ~9. Overall from Group I to Group IV for the monoterpene SOA and Group I to Group III for sesquiterpene SOA, the organic species contain more O atoms and more DBE thus more O-containing functional groups are expected, including hydroxyl, carbonyl (ketone and aldehyde), and carboxyl functional groups.
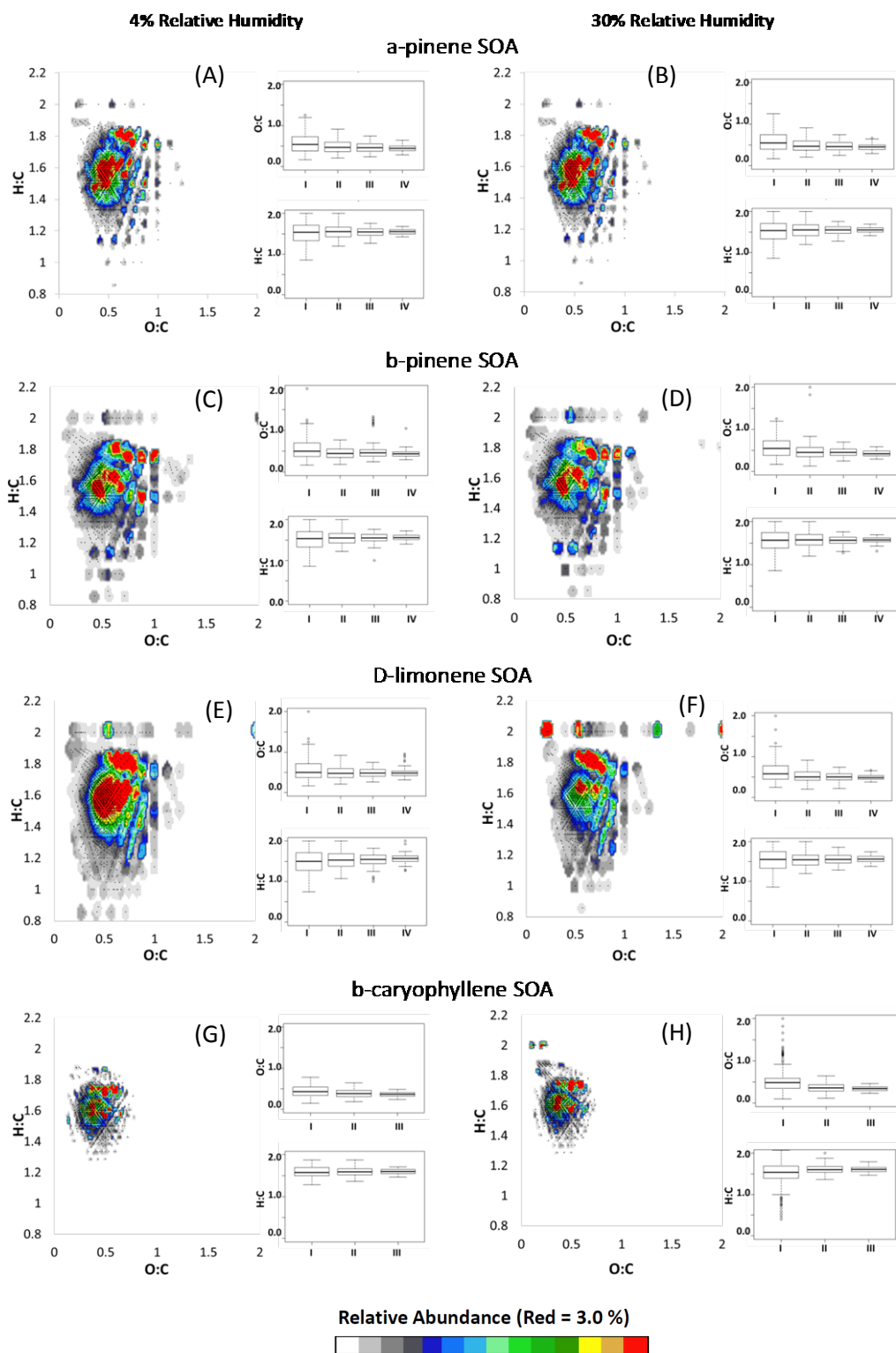
Figure 3.4 The isoabundance van Krevelen diagrams with box plots showing O/C and H/C ratios of the SOA Groups for SOA sample AP4_4 (A), AP4_30 (B), BP10_4 (C), BP11_30 (D), LM2_4 (E), LM3_30 (F), CP2_4 (G) and CP3_30 (H).

58

Studies showed that the high molecular weight products (aka oligomers) are formed quickly (within seconds) once the biogenic precursor reacts with the oxidant. There are no kinetic barriers for oligomer formation and it is unlikely that a large number of steps are involved in their formation (Heaton et al., 2007; Bateman et al., 2009). Tolocka et al. (2004) proposed aldol condensation and gem-diol reactions as the primary pathways of oligomer formation in α-pinene SOA. Bateman et al. (2009) suggested reactions between the Criegee intermediates and stable first-generation products of limonene ozonolysis as the dominant mechanism of oligomerization. Kundu et al. (2012) examined the formation of high molecular weight (>300 Da) limonene SOA components from the building blocks of low molecular weight SOA components, hydroperoxides and Criegee radicals. Based on the number of possible pathways for the observed species, they reported hemi-acetal reactions to be the most likely pathway for high molecular weight limonene SOA formation followed by the hydroperoxide and Criegee reaction channels. Aldol and ester condensation reactions were reported to be the least dominant pathway because no reduction in the H/C with respect to molecular weight was observed. No reports were found on sesquiterpene SOA oligomerization pathways, although the principles of reactions are expected to be similar to those okf the monoterpenes. The oligomers from various building blocks with different reaction pathways have diverse molecular formulas and structures. Due to the observed complex array of species, the SOA is expected to result from a complex array of radical and non-radical accretion reaction channels. However, there might be different dominating pathways in different SOA experimental conditions.

3.1.3 Van Krevelen analysis and the carbon number trends of the LANLSOA samples

The van Krevelen diagrams in Figure 3.4 show the O/C and H/C for the individual species of selected LANLSOA samples (the SOA samples generated under 4% and 30% relative humidity without OH scavenger). The color in the plots represents the cumulative relative abundances of the species at a given point. Overall, a majority of the molecular formulas identified in the monoterpene SOA (α-pinene, β-pinene and D-limonene SOA) have O/C values ranging from 0.1 to 1.0 and H/C from 1.0 to 2.0. The highest relative abundance species have O/C values of ~0.5 and H/C values of ~ 1.6. All of the monoterpene SOA has quite similar elemental ratio distributions. Further, the elemental ratio distributions of SOA generated from the different RH are also quite similar. The β-caryophyllene SOA components are less various in terms of their O/C and H/C ratios; the O/C range from 0.2 to 0.8 and H/C range from 1.4 to 1.8 for most

compounds. The average O/C and H/C ratios of β-caryophyllene SOA are around 0.38 and 1.60, respectively. Due to the higher number of carbon atoms in the sesquiterpene, the β-caryophyllene SOA samples are less oxidized and more saturated than the monoterpene SOA samples.

The trends in the number of O and DBE values with respect to the number of C for the LANLSOA samples were examined. The average O number and the DBE values of the formulas within bins at intervals of every 5 carbon atoms were calculated for the selected LANLSOA samples (Figure 3.5). The general trends for all of the SOA samples studied are that both O number and DBE values increase with C numbers. This indicates that the larger SOA compounds contain more O functional groups and more double bonds or rings in the potential structures for the molecular formulas. Again the SOA samples generated from the same precursor with different RH conditions show almost the same values, reflecting their similar molecular composition.

(B)

(C)

61
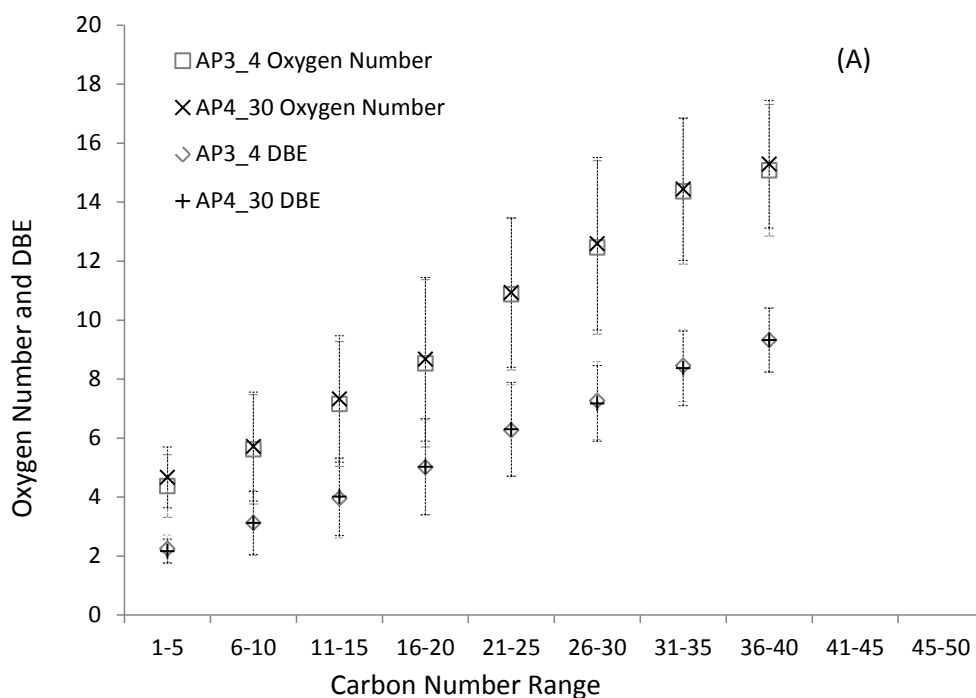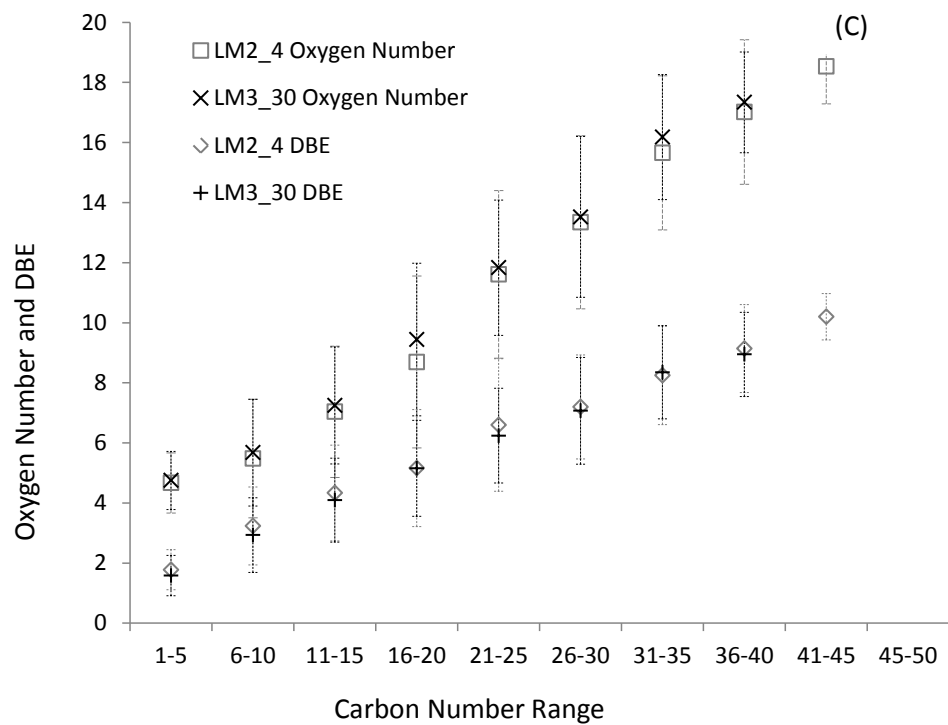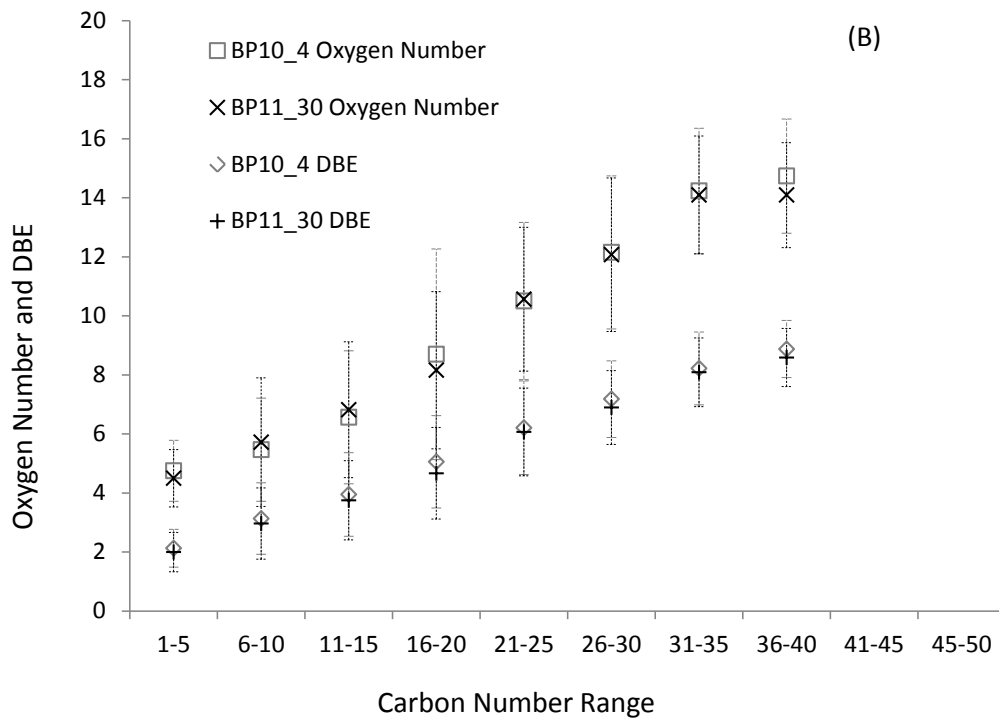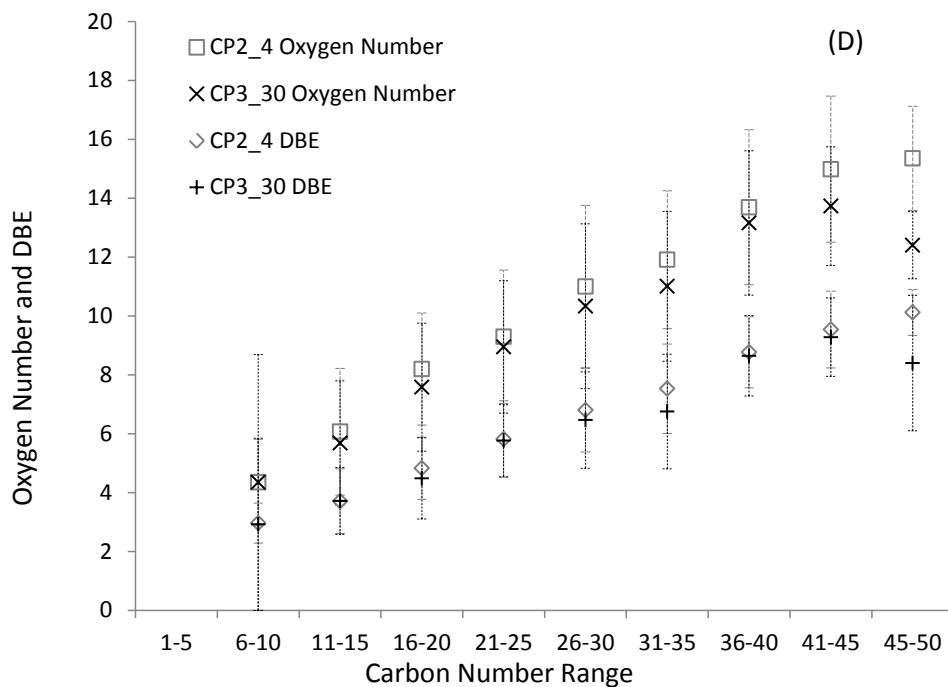
Figure 3.5 Carbon number trends of the oxygen number and DBE values for selected α-pinene SOA samples (A), β-pinene SOA samples (B), limonene SOA samples (C) and β-caryophyllene SOA samples (D).

3.2 Statistical analysis of the LANLSOA molecular composition and biogenic SOA indicator species

From the discussion in Section 3.1, it is clear the biogenic SOA from different precursors generated under different conditions share a large number of common molecular formulas and many similar chemical properties. Yet still, subtle differences are also observed among the LANLSOA samples. In this section, the use of multivariate statistical analysis methods to comprehensively compare the molecular composition of the SOA is presented. Biogenic SOA indicator species will be explored to improve the source apportionment of ambient atmospheric samples.

The assigned and quality assured monoisotopic molecular formulas in all of the LANLSOA samples were combined together into a dataset for multivariate statistical analysis. A total of 2653 distinct molecular formulas were retained after removing the duplicates. The molecular formulas and their relative abundances in each of the LANLSOA samples were used as the statistical analysis input. The combined dataset contains 19 observations (19 LANLSOA samples) with 2653 variables (the molecular formulas). Several variations of data preparation were tested and finally a scaled normalized RA was used in the multivariate statistical analysis to determine the biogenic SOA indicator species (Section 3.2.5).

3.2.1 Comparison of the LANLSOA samples using a heat map

A summary of the molecular composition of the biogenic SOA is shown graphically using a heat map in Figure 3.6. Since there are a total of 2653 distinct molecular formulas identified in the LANLSOA samples, the resulting heat map using all of the formulas would likely result in an extremely complex heat map and it would be difficult to extract useful information from it. For this reason, I selected 50 molecular formulas with the highest relative abundances from α-pinene SOA, β-pinene SOA, D-limonene SOA and β-caryophyllene SOA. The formulas with their relative abundances in the specific samples were combined and the duplicate formulas were removed, leaving a list of 96 distinct molecular formulas. As shown in Figure 3.6, the highest RA formulas typically contain 8 – 30 C atoms, 3 – 13 O atoms with a range of DBE values from 2 to 7.

The LANLSOA samples are shown according to the output from the hierarchical cluster analysis (Section 2.1.5) with the sample labels at the bottom of the heat map. The hierarchical cluster

analysis was done using the default heat map function parameters in R, including the "Euclidean" distance method and "complete" linkage. The selection of the hierarchical clustering parameters will be discussed in Section 3.2.2. In the heat map, the β-caryophyllene SOA samples have a clearly different molecular composition from the others (monoterpene SOA). For the monoterpenes SOA samples, α-pinene SOA and β-pinene SOA have very similar molecular compositions, which are slightly different from the D-limonene SOA. The high RA molecular formulas in monoterpene SOA samples contain C numbers 8 – 11 and 16 – 21, while the high RA formulas in the sesquiterpene SOA contain C numbers 11 – 16 and 27 – 30, reflecting the terpenoid characteristics of the biogenic SOA (aka monomers and dimers).

(A)

(B)
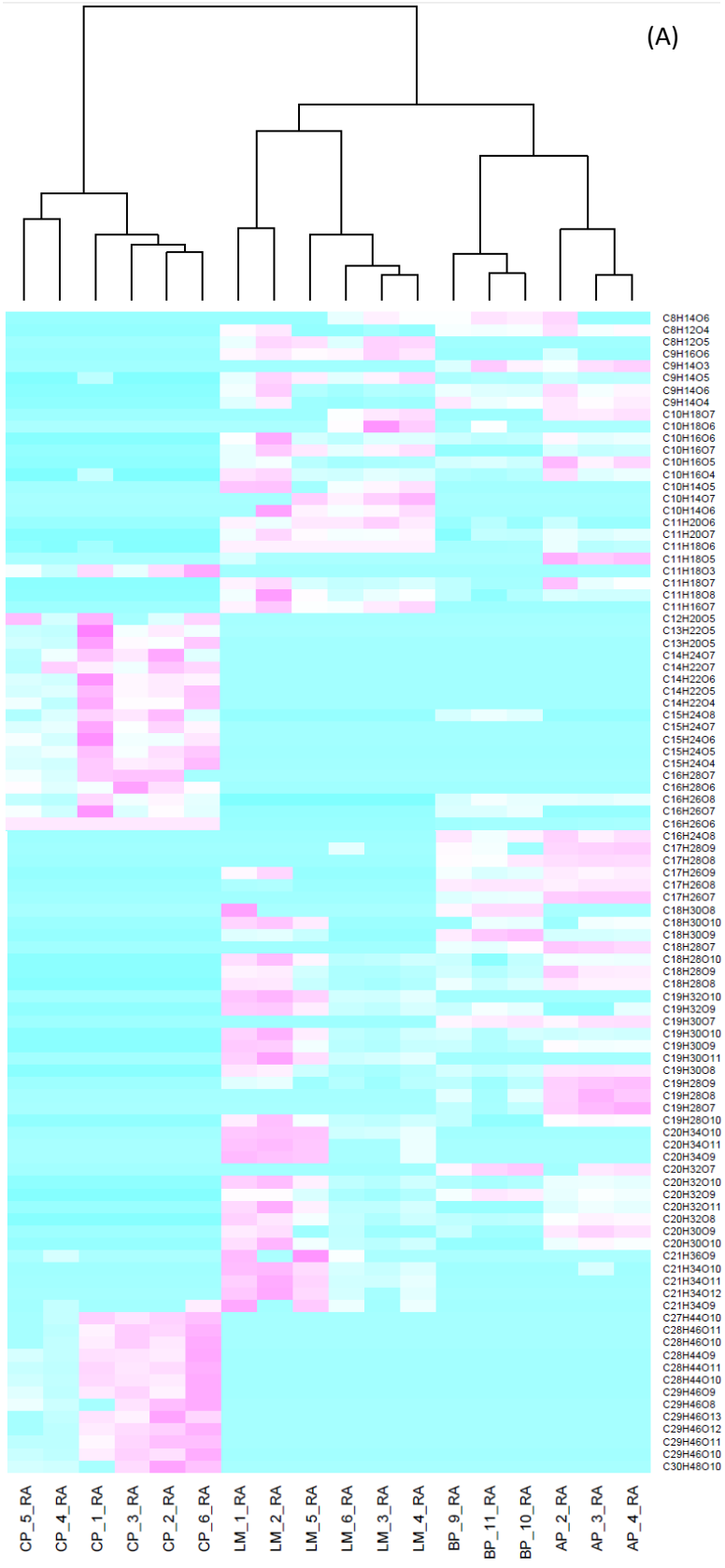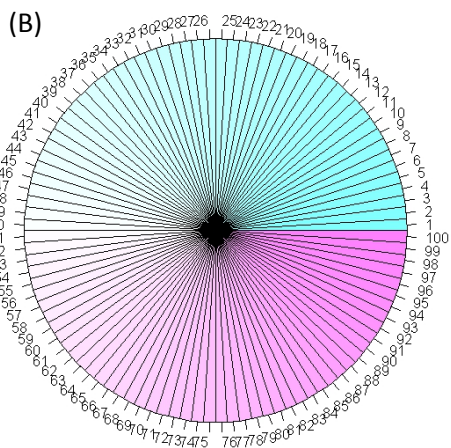
Figure 3.6 LANLSOA sample heat map using 96 molecular formulas with the 50 highest relative abundances in each of the samples (A). The clustering of the samples uses the "Euclidean" distance method and "complete" linkage. The molecular formula relative abundances are represented with the colors shown in the pie chart (B).

3.2.2 Multivariate statistical analysis using molecular formula relative abundances and normalized relative abundances

All of the 2653 variables (or molecular formulas) were used in the hierarchical cluster analysis (HCA) and principal component analysis (PCA) multivariate statistical analysis methods of the LANLSOA composition. Though the analyte relative abundances in the mass spectra are the most common way to represent the relative magnitude of the compounds, they are not directly comparable unless several strict criteria are met (to be discussed in more detail in Section 4.2). The molecular formula RA is relative to the base peak of the mass spectra and depends on the electrospray and sample matrix conditions. A way to address this problem is to normalize the RA on a sample basis. To do this, each of the analyte RAs are divided by the sum of all RAs for the sample and this is done consistently to each of the samples. In this way, the normalized RAs of the same formula in different spectra are dimensionless and comparable to each other. The modified combined dataset still contains all 19 samples and 2653 variables but contains normalized RAs instead of measured RAs.

3.2.2.1 Hierarchical cluster analysis and selection of the hierarchical cluster analysis parameters

Hierarchical cluster analysis sorts the LANLSOA samples based on their molecular composition similarities and differences. The samples with similar compositions are clustered together in the dendrogram, according to the "distance" of the sample similarities. Some HCA dendrograms generated with various HCA parameters are shown in Figure 3.7. Overall, the LANLSOA samples are separated into two big clusters: one for the monoterpene SOA (α-pinene, β-pinene and D-limonene SOA) and the other for the sesquiterpene SOA (β-caryophyllene SOA). Within the monoterpene SOA cluster, D-limonene SOA is separated from the α- and β-pinene SOA. This indicates that regardless of the SOA generation conditions (RH and OH scavenger), the SOA samples from the same precursor have very similar molecular compositions compared to the SOA from other precursors. In addition, the sesquiterpene SOA molecular composition is substantially different than the monoterpene SOA. The SOA from α-pinene and β-pinene, two structural isomers with one double bond and two rings in their structures, are very similar to each other. However they are substantially different from the D-limonene SOA, which has two double bonds and one ring in its structure. Although not shown here, the samples group

consistently regardless of the HCA parameters. This observation confirms the sorting of the molecular compositions and indicates some inherent molecular characteristics of the LANLSOA samples.
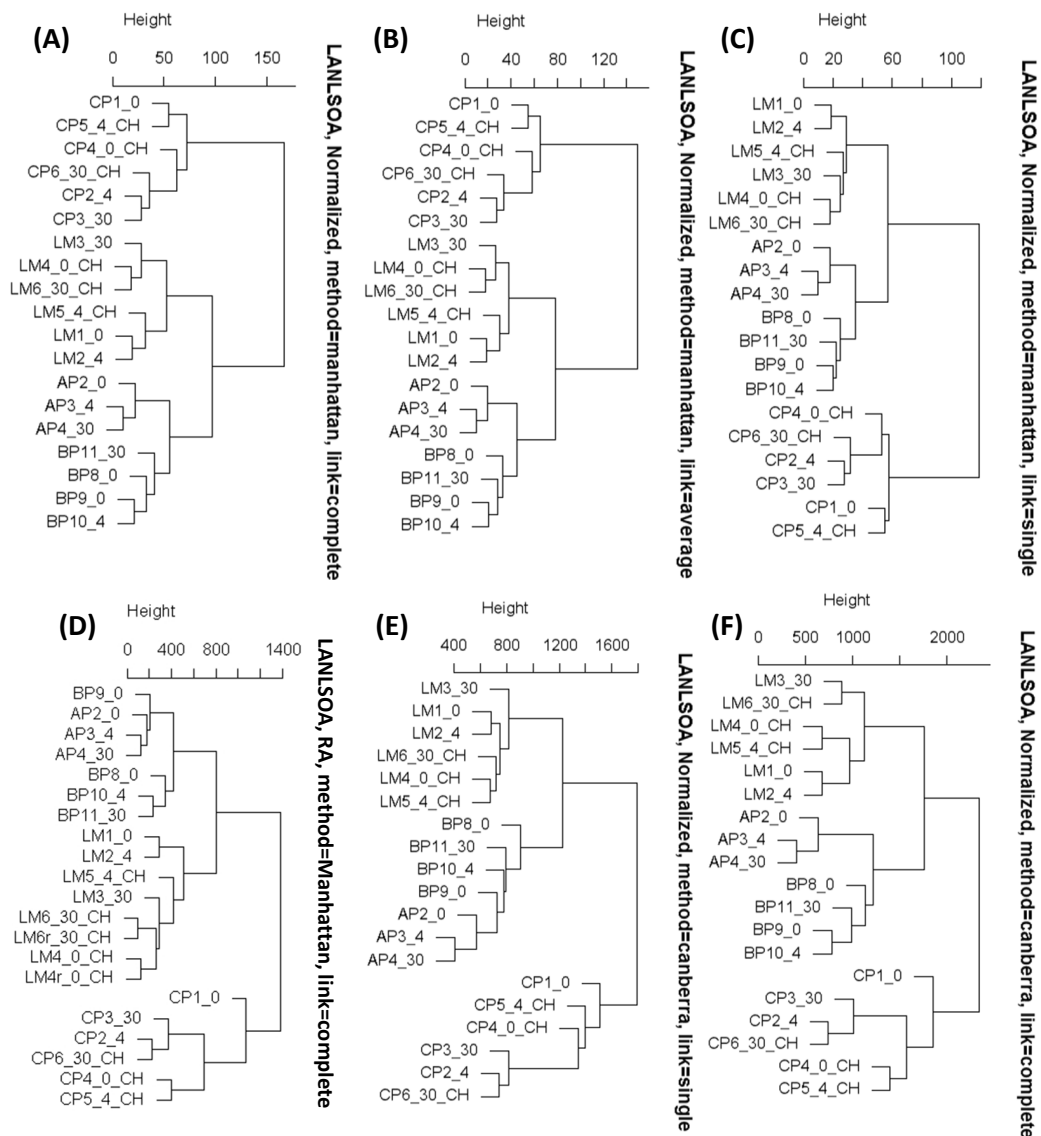
Figure 3.7 The HCA dendrograms for the LANLSOA. The dendrograms were generated using the Manhattan distance method with complete (A), average (B) and single (C) linkages, with normalized RA. (D) dendrogram was generated using the same HCA parameters as (A) dendrogram but without normalization of the RA. The dengrograms in (E) and (F) result from the Canberra distance method with single and complete linkage, respectively, with normalized RA.

Minor differences in the sample groupings within the precursor types are observed in the dendrograms. For example, different grouping patterns of the 6 D-limonene SOA samples were observed using different ways to measure the distances between a pair of samples in HCA. There is not a single perfect measure to reflect the "natural" grouping of the observations in all of the cases. Thus, I tested all of the available distance measurements in the *hcluster* function for R, including "Euclidean", "Maximum", "Manhattan", "Canberra", "Binary", "Pearson", "Correlation", "Spearman" and "Kendall" distances. HCA was conducted using each of distance measurement methods with complete linkage, average linkage and single linkage. Those linkages are the commonly used linkage methods (Johnson and Wichern, 2007) and will be discussed in more detail in next paragraph. Though the general grouping patterns are consistent with separation of α- and β-pinene SOA, D-limonene SOA and β-caryophyllene SOA using all of the distance methods (which again confirms the grouping), in some cases CP4_0_CH and CP5_4_CH stood out from the other LANLSOA samples (e.g., using the Maximum distance method with all three linkages and Euclidean distance method with average linkage); in other cases α-pinene SOA and β-pinene SOA are not separated from each other within the monoterpene SOA cluster. The dendrograms generated using the Manhattan (shown in Figure 3.7 A to C), Correlation and Pearson distance methods, regardless of the linkage requirements used, provided a clean separation of α-pinene SOA, β-pinene SOA, D-limonene SOA and β-caryophyllene SOA samples. Thus we selected the Manhattan distance method in the following analysis using HCA.

The linkage method sets the criteria for determining the distance between groups of observations instead of individual observations. Single linkage uses the minimum distance between two groups or in other words it uses the distance between the nearest members in the two groups as the distance between groups. Complete linkage uses the maximum distance between the observations in the two groups or in other words it uses the distance between the farthest members in the two groups as the distance between groups. Average linkage uses the average distance between pairs of members in respective sets as the group distance (see more detail in Section 2.1.5). As mentioned in the previous paragraph, all three linkages were tested using every available distance method. For a given distance method, complete linkage tends to separate the SOA samples from a given precursor clearly while the single and average linkages do not. An example was made using the Canberra distance method with both the single linkage

and complete linkage (Figures 3.7 E and F). Here the α-pinene SOA and β-pinene SOA samples are mixed with the single linkage in Figure 3.7 E, but they are clearly separated with the complete linkage requirement in Figure 3.7 F. α-pinene SOA (1107 common formulas in AP2_0, AP3_4 and AP4_30) and β-pinene SOA (759 common formulas in BP8_0, BP10_4 and BP11_30) share 752 common molecular formulas, but there are 355 formulas only observed in α-pinene SOA. This indicates some minor composition differences are observed for the two types of SOA. Thus, α -pinene SOA and β-pinene SOA are expected to be separated in the analysis. Thus the complete linkage is used in the following HCA analysis.

The effects of relative humidity and the presence or absence of an OH scavenger on biogenic SOA composition is complex. The effects are different for each of the precursors; the influence of the RH depends on the OH scavenger and influence of the OH scavenger also depends on the RH conditions. For example in Figure 3.7 A, for the D-limonene SOA, two samples without OH scavenger LM1_0 and LM2_4 show similar molecular composition; while with OH scavenger, LM4_0_CH and LM6_30_CH group closely reflecting their similar composition. Further investigation with additional SOA experiments is necessary to understand the coupled effects of the OH scavenger and the RH on the biogenic SOA composition.

HCA dendrograms from analyses conducted using the molecular formula measured RAs without any normalization is shown in Figure 3.7 D. For comparison, Figure 3.7 D was generated with the same HCA parameters as Figure 3.7 A, including the Manhattan distance method and complete linkage. The only difference here is how the RA was used (measured RAs vs. normalized RAs). The dendrogram with the measured RAs shows separated monoterpene and sesquiterpene SOA and D-limonene SOA is separated from the α- and β-pinene SOA. Overall, the chamber-generated SOA samples are much less various in terms of their molecular composition compared to ambient samples. Even when the RA were not normalized, the sample grouping pattern is quite similar to that of the normalized RAs. Regardless, there are still differences between Figures 3.7 A and D. The most significant difference is the grouping of sample BP9_0 with the α-pinene SOA in Figure 3.9D, reflecting the necessity of the RA normalization in preparation for the statistical analysis.

3.2.2.2 Principal component analysis

Principal component analysis (PCA; Section 2.1.5) combines the original variables linearly to generate new variables, which are called principal components (PCs). Principal components are ordered by the variance they carry. PC1 represents the highest variance among all the PCs and PC2 represents the second highest variance and so on. PCA groups the observations based on the variation of their variables, like the HCA. But more importantly, PCA allows us to explore the relationships between the observations (the LANLSOA samples) and their variables (the molecular formulas). Using this analysis, an attempt to identify the marker species or indicator compounds for specific LANLSOA samples was made.

The PC1 and PC2 generated with the normalized RA carry more than 80% of the original variance; specifically PC1 carries 63% and PC2 carries 21%. The first two PCs are used to show the PCA results. The PC loadings of both the LANLSOA samples and the molecular formulas are shown in the PCA biplot (Figure 3.8). Consistent with the HCA dendrograms, all of the SOA samples from the same precursor group together reflecting their molecular composition similarities. In the PCA biplot, we observe that PC1 separates the monoterpene and sesquiterpene SOA. All of the SOA samples from the sesquiterpene (β-caryophyllene) have positive PC1 scores; all of the SOA samples from the monoterpenes (α-pinene, β-pinene and D-limonene) have negative PC1 scores. In addition, PC2 separates the α- and β-pinene SOA (which are isomers) from the D-limonene SOA.
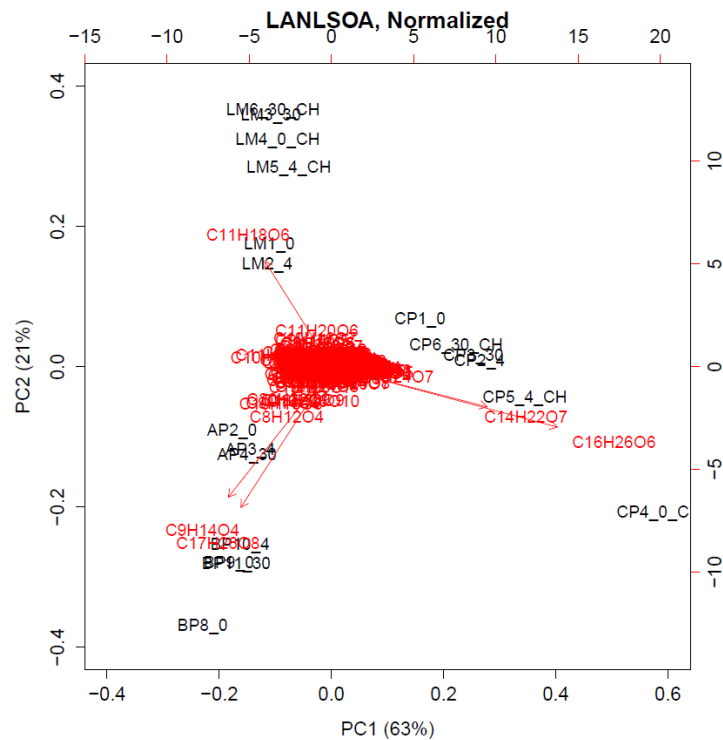
Figure 3.8 PCA biplot conducted with normalized RA.

According to the principal component analysis, the variables in the same area as the observations are the characteristic variables for those observations. In this study, the molecular formulas with positive PC1 loadings are the characteristic formulas of the β-caryophyllene SOA; the molecular formulas with negative PC1 and positive PC2 loadings are characteristic of the D-limonene SOA and the formulas with both negative PC1 and negative PC2 loadings are characteristic formulas of the α- and β-pinene SOA.
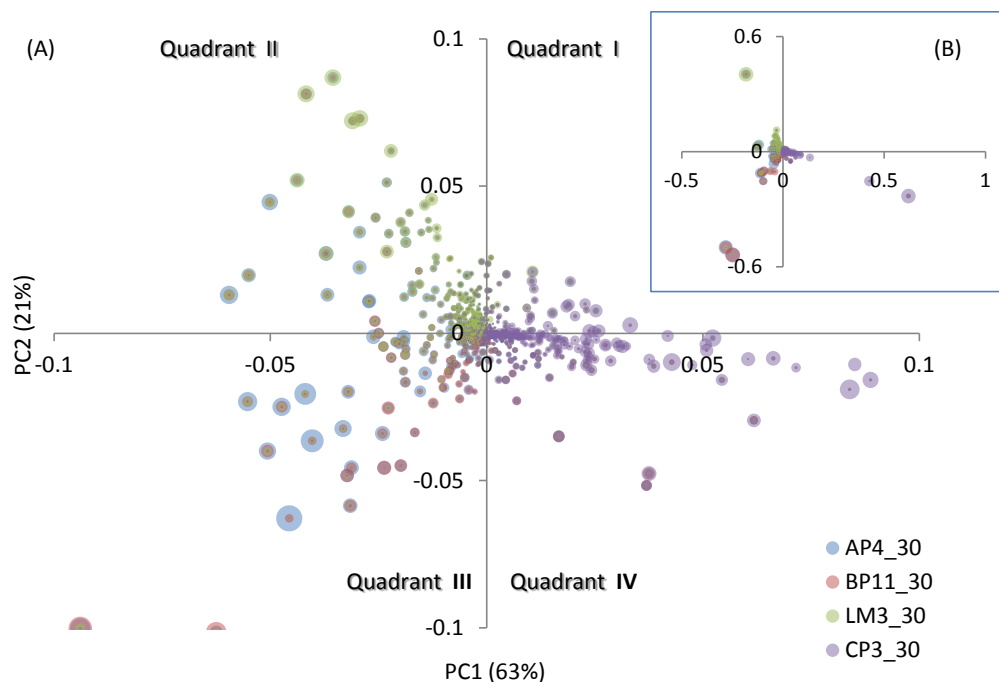
Figure 3.9 The PCA loading bubble plot with bubble size representing the relative abundances of molecular formulas from specific samples (AP4_30, BP11_30, LM3_30 and CP3_30); Plot (A) is an excerpt of the full plot (zoomed in around the origin); and plot (B) is the fulll plot. PCA was conducted with normalized RAs.

This is confirmed with the PCA loading bubble plot shown in Figure 3.9. Here representative SOA samples generated with 30% RH and without OH scavenger from each precursor were selected, including: AP4_30, BP11_30, LM3_30 and CP3_30. The molecular formulas of these samples are plotted on PC1-PC2 shown with different colors and with the bubble size representing the RAs of the molecular formulas for each of the specific samples. As shown in Figure 3.9, even though the molecular formulas in a specific sample spread over the quadrants of the PCA loading plot, the formulas with high RAs (large bubble size) of a specific sample are generally at the same locations as the sample. For example, the high RA formulas of the β-caryophyllene SOA (CP3_30) are in quadrants I and IV where the β-caryophyllene SOA samples are located. In the PCA loading plot, the variables with the greater distances from the origin have higher contributions to the variation among the observations. In this case, the formulas far away from the origin are more responsible for the location of the LANLSOA samples in the biplot. In the PCA loading plot showing the results from the analysis using the normalized RAs, the bubble size becomes larger and larger further away from the origin, reflecting the importance of the high RA formulas on the variation of the LANLSOA molecular composition. Notice there are several molecular

formulas standing out of the center region in Figure 3.9 B. The two in the β-caryophyllene SOA region (in quadrant IV) are $C_{16}H_{26}O_6$ and $C_{14}H_{22}O_7$; the one in D-limonene SOA sample region is $C_{11}H_{18}O_6$ and the two in α- and β-pinene SOA region are $C_9H_{14}O_4$ and $C_{17}H_{28}O_8$. They correspond to the tallest peaks in the mass spectra and are the species expected to have the highest concentrations in LANLSOA samples.

Several characteristic molecular formulas for the α- and β-pinene SOA, D-limonene SOA and β-caryophyllene SOA based on the principal component analysis with normalized RAs are listed in Table 3.2. They are the molecular formulas with the largest distances from the origin in the regions corresponding to the LANLSOA samples in PCA biplot. The α- and β-pinene SOA characteristic molecular formulas mostly contain ~10 or 20 C atoms and probably represent first generation products and dimers of the first generation products of α- and β-pinene ozonolysis. The first generation products (or monomers) keep the precursor DBE of 3, while the dimers have DBE values of 5 or 6. Similarly, D-limonene SOA characteristic formulas contain ~10 or 20 C atoms likely corresponding to the monomers and dimers, respectively. The monomers have DBE values of 2 – 4 and dimers have DBE values of 4 or 5 compared to a DBE of 3 for the precursor. B-caryophyllene SOA characteristic molecular formulas have ~15 or 30 C atoms in their molecular formulas, which again might represent monomers and dimers of first generation products with DBE values 3 or 4 and 7, respectively (compared to a DBE value of 4 in the precursor). There are several more dimer molecular formulas with DBE values of 7, which appear to be characteristic of the β-caryophyllene SOA. But, they have slightly smaller distances and so they are not shown in the table. The O/C and H/C ratios of the characteristic molecular formulas are consistent with the elemental ratios of the high RA formulas in the corresponding SOA samples as shown in the van Krevelen plot (Figure 3.4). Thus, these characteristic molecular formulas are potential indicator species of biogenic SOA samples studied here.

Table 3.2 Characteristic molecular formulas of α- and β-pinene SOA, D-limonene SOA and β-caryophyllene SOA based on the principal component analysis with normalized RAs.

α- and β-pinene SOA						D-limonene SOA						β-caryophyllene SOA					
Distance	Formula	Theor. m/z	DBE	O/C	H/C	Distance	Formula	Theor. m/z	DBE	O/C	H/C	Distance	Formula	Theor. m/z	DBE	O/C	H/C
0.5921	C17H26O8	357.1555	5	0.47	1.53	0.4429	C11H18O6	245.1031	3	0.55	1.64	0.6610	C16H26O6	313.1657	4	0.38	1.63
0.5727	C9H14O4	185.0819	3	0.44	1.56	0.1323	C10H16O6	231.0874	3	0.60	1.60	0.4533	C14H22O7	301.1293	4	0.50	1.57
0.1816	C8H12O4	171.0663	3	0.50	1.50	0.1242	C11H18O7	261.0980	3	0.64	1.64	0.1364	C14H24O7	303.1449	3	0.50	1.71
0.1572	C10H16O5	215.0925	3	0.50	1.60	0.1161	C11H20O6	247.1187	2	0.55	1.82	0.0901	C14H22O5	269.1394	4	0.36	1.57
0.1525	C9H14O6	217.0718	3	0.67	1.56	0.0938	C10H16O7	247.0823	3	0.70	1.60	0.0860	C16H28O6	315.1813	3	0.38	1.75
0.1373	C20H32O9	415.1974	5	0.45	1.60	0.0914	C9H14O5	201.0768	3	0.56	1.56	0.0857	C11H18O3	197.1183	3	0.27	1.64
0.1192	C18H30O9	389.1817	4	0.50	1.67	0.0787	C9H16O6	219.0874	2	0.67	1.78	0.0724	C12H20O5	243.1238	3	0.42	1.67
0.1113	C16H26O10	377.1453	4	0.63	1.63	0.0786	C8H12O5	187.0612	3	0.63	1.50	0.0685	C15H24O7	315.1449	4	0.47	1.60
0.0776	C19H28O7	367.1762	6	0.37	1.47	0.0680	C11H18O8	277.0929	3	0.73	1.64	0.0668	C14H22O6	285.1344	4	0.43	1.57
0.0665	C16H24O8	343.1398	5	0.50	1.50	0.0671	C20H32O10	431.1923	5	0.50	1.60	0.0635	C16H26O7	329.1606	4	0.44	1.63
0.0645	C18H28O8	371.1711	5	0.44	1.56	0.0658	C11H16O7	259.0823	4	0.64	1.45	0.0610	C15H24O4	267.1602	4	0.27	1.60
0.0599	C18H28O9	387.1661	5	0.50	1.56	0.0609	C19H30O9	401.1817	5	0.47	1.58	0.0606	C15H24O8	331.1398	4	0.53	1.60
0.0580	C19H30O7	369.1919	5	0.37	1.58	0.0584	C19H30O10	417.1766	5	0.53	1.58	0.0566	C15H24O6	299.1500	4	0.40	1.60
0.0553	C18H28O7	355.1762	5	0.39	1.56	0.0562	C21H34O10	445.2079	5	0.48	1.62	0.0523	C29H46O11	569.2967	7	0.38	1.59
0.0544	C11H18O5	229.1081	3	0.45	1.64	0.0522	C11H20O7	263.1136	2	0.64	1.82	0.0512	C13H20O5	255.1238	4	0.38	1.54
0.0535	C19H30O8	385.1868	5	0.42	1.58	0.0473	C11H22O6	249.1344	1	0.55	2.00	0.0508	C14H22O4	253.1445	4	0.29	1.57
0.0514	C9H14O3	169.0870	3	0.33	1.56	0.0469	C19H32O10	419.1923	4	0.53	1.68	0.0487	C15H24O5	283.1551	4	0.33	1.60
0.0491	C20H32O7	383.2075	5	0.35	1.60	0.0460	C10H16O4	199.0976	3	0.40	1.60	0.0439	C16H28O8	347.1711	3	0.50	1.75
0.0467	C19H28O9	399.1661	6	0.47	1.47	0.0459	C10H14O7	245.0667	4	0.70	1.40	0.0418	C13H22O5	257.1394	3	0.38	1.69
0.0463	C19H28O8	383.1711	6	0.42	1.47	0.0452	C20H32O11	447.1872	5	0.55	1.60	0.0402	C16H28O7	331.1762	3	0.44	1.75

*The formulas are ordered by their distances from the origin in the PCA loading plot. The distance is calculated using the following formula: distance = ((PC1 loading)^2 + (PC2 loading)^2)$^{1/2}$.

3.2.3 Multivariate statistical analysis using a logarithm conversion of the molecular formula relative abundances

The multivariate statistical analysis appears to evaluate the molecular composition of the LANLSOA samples quite well. The samples with similar molecular composition (the SOA from the same precursors) are grouped consistently in the HCA and PCA and indicator species of the SOA samples were explored. However, the presented results are driven by just a few molecular formulas with much higher RAs than the others. Thus, the roles of the low RA formulas are potentially masked. To lower the significance of the high RA formulas and enhance the importance of the low RA formulas in the multivariate statistical analysis, we re-scaled the RAs by converting them to a log scale with base 10. To avoid negative values when taking the logarithm on the RAs directly (if RA is < 1%), the measured RAs were normalized to the smallest RA in a sample by dividing all of the measured RA values in the sample by the smallest RA. Then the logarithm was taken on the ratios. HCA and PCA were conducted with the logarithm of the RA ratios.
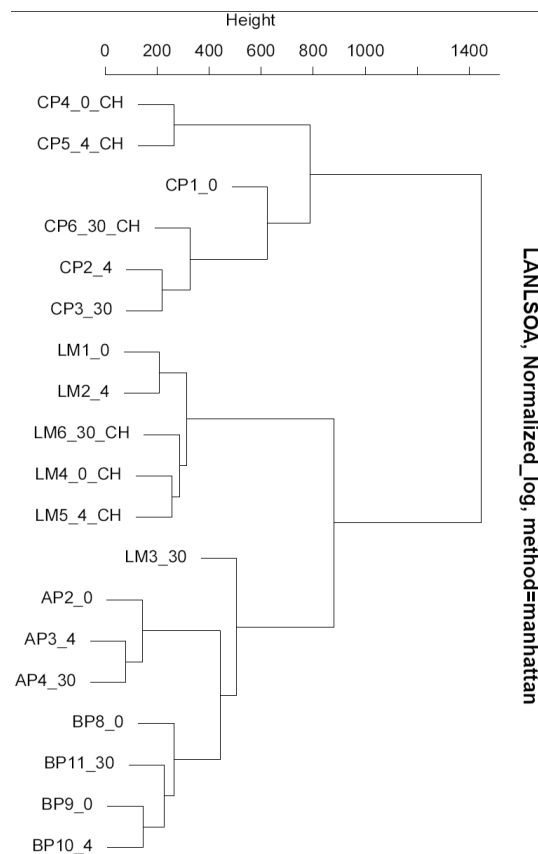
Figure 3.10 HCA dendrogram with logarithm of relative abundance ratios. Manhattan distance method and complete linkage were used.

Similar to the clustering analysis with the normalized RAs, the HCA dendrogram with the log scaled RA ratios shows a clean separation of the monoterpene and sesquiterpene SOA samples and the D-limonene SOA is separated from the α- and β-pinene SOA (Figure 3.10). But one of the D-limonene SOA samples (LM3_30) is unexpectedly grouped with the α- and β-pinene SOA. This might be caused by the un-normalization of the relative abundances. Normalization makes the RAs of the same formula from different spectra comparable, but when the RAs are divided by the smallest RA in a sample, the normalization effect is canceled out.

The PCA results conducted with the log scaled RA is shown with a biplot in Figure 3.11. PC1 and PC2 represent 64% and 16% of the original variance, which is somewhat lower than the PCA results using the normalized RA dataset. The location and grouping of the LANLSOA samples on the biplot are quite similar to the previous results (Figure 3.8), but the distribution of the molecular formulas is very different. As expected, the formulas spread over the biplot evenly without any standing out and far away from the others; this is because the RAs were re-scaled

77

by taking the logarithm of the RA ratios. Again, the molecular formulas with positive PC1 loadings are characteristic of the β-caryophyllene SOA; the formulas with negative PC1 and positive PC2 loadings are characteristic of the D-limonene SOA and the molecular formulas with both negative PC1 and negative PC2 loadings are characteristic of the α- and β-pinene SOA. This is confirmed with the PCA loading bubble plot in Figure 3.12. Still the higher RA molecular formulas are more distant from the origin. But unlike the biplot with normalized RA (Figure 3.8), the majority of formulas are not compressed around the origin, reflecting their increased significance in the analysis as a result of the log rescaling. Several characteristic formulas of the LANLSOA samples are listed in Table 3.3 with the same selection rules as described for Table 3.2. Similar to the previous list of characteristic formulas identified by the PCA using the normalized RA, the molecular formulas are highly likely to be monomers and dimers of the first generation ozonolysis products. In fact, several of the characteristic formulas are common between the two lists in Table 3.2 and Table 3.3. However in the log scaled analysis, there are more dimers identified as characteristic molecular formulas in Table 3.3. This is because typically the dimers have lower RAs than the monomers in mass spectra and taking the logarithm enhances the roles of lower RAs in the multivariate statistical analysis.
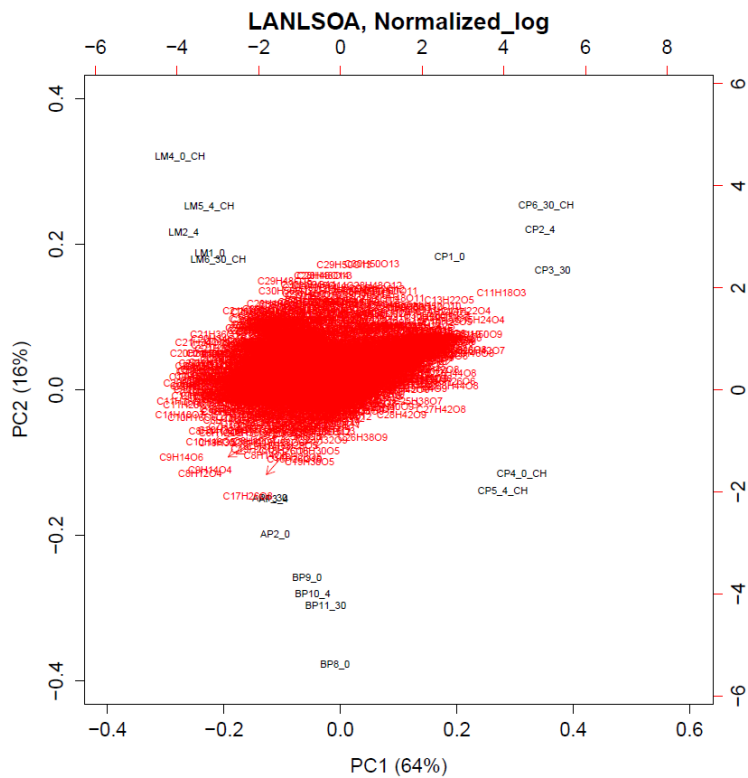
Figure 3.11 PCA biplot conducted with the logarithm scaled RA ratios.
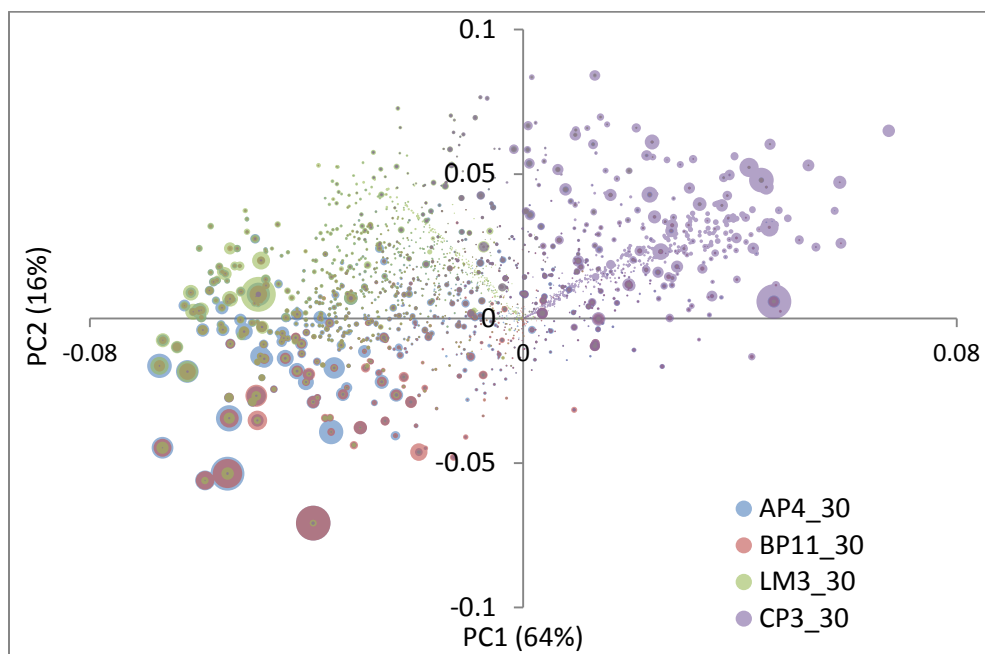
Figure 3.12 The PCA loading bubble plot with bubble size representing the relative abundances of formulas in specific samples. PCA was conducted with the logarithm scaled RA ratios.

Table 3.3 Characteristic molecular formulas of α- and β-pinene SOA, D-limonene SOA and β-caryophyllene SOA based on principal component analysis with the logarithm scaled RA ratios.

| α- and β-pinene SOA						D-limonene SOA						β-caryophyllene SOA					
Distance	Formula	Theor. m/z	DBE	O/C	H/C	Distance	Formula	Theor. m/z	DBE	O/C	H/C	Distance	Formula	Theor. m/z	DBE	O/C	H/C
0.0812	C8H12O4	171.0663	3	0.5000	1.5000	0.0770	C28H48O14	607.2971	5	0.5000	1.7143	0.0936	C11H18O3	197.1183	3	0.2727	1.6364
0.0807	C17H26O8	357.1555	5	0.4706	1.5294	0.0765	C28H48O13	591.3022	5	0.4643	1.7143	0.0852	C30H50O13	617.3179	6	0.4333	1.6667
0.0802	C9H14O6	217.0718	3	0.6667	1.5556	0.0763	C29H48O15	635.2920	6	0.5172	1.6552	0.0835	C29H50O13	605.3179	5	0.4483	1.7241
0.0765	C9H14O4	185.0819	3	0.4444	1.5556	0.0715	C30H50O14	633.3128	6	0.4667	1.6667	0.0756	C13H22O5	257.1394	3	0.3846	1.6923
0.0691	C11H18O7	261.0980	3	0.6364	1.6364	0.0701	C29H48O14	619.2971	6	0.4828	1.6552	0.0751	C15H24O4	267.1602	4	0.2667	1.6000
0.0670	C11H18O8	277.0929	3	0.7273	1.6364	0.0699	C30H50O15	649.3077	6	0.5000	1.6667	0.0747	C14H22O4	253.1445	4	0.2857	1.5714
0.0647	C11H20O7	263.1136	2	0.6364	1.8182	0.0698	C29H50O14	621.3128	5	0.4828	1.7241	0.0711	C28H48O12	575.3073	5	0.4286	1.7143
0.0647	C10H16O6	231.0874	3	0.6000	1.6000	0.0660	C21H34O12	477.1978	5	0.5714	1.6190	0.0692	C30H50O11	585.3280	6	0.3667	1.6667
0.0643	C10H16O5	215.0925	3	0.5000	1.6000	0.0657	C27H46O14	593.2815	5	0.5185	1.7037	0.0689	C31H52O12	615.3386	6	0.3871	1.6774
0.0608	C8H14O6	205.0718	2	0.7500	1.7500	0.0651	C21H38O10	449.2392	3	0.4762	1.8095	0.0685	C31H50O9	565.3382	7	0.2903	1.6129
0.0604	C18H30O9	389.1817	4	0.5000	1.6667	0.0647	C27H46O13	577.2866	5	0.4815	1.7037	0.0683	C31H52O10	583.3488	6	0.3226	1.6774
0.0593	C19H30O10	417.1766	5	0.5263	1.5789	0.0646	C20H34O11	449.2028	4	0.5500	1.7000	0.0678	C31H52O13	631.3335	6	0.4194	1.6774
0.0578	C8H12O6	203.0561	3	0.7500	1.5000	0.0640	C39H62O17	801.3914	9	0.4359	1.5897	0.0669	C29H50O12	589.3230	5	0.4138	1.7241
0.0561	C20H32O9	415.1974	5	0.4500	1.6000	0.0638	C28H46O15	621.2764	6	0.5357	1.6429	0.0669	C14H24O7	303.1449	3	0.5000	1.7143
0.0558	C18H28O11	419.1559	5	0.6111	1.5556	0.0634	C21H36O12	479.2134	4	0.5714	1.7143	0.0667	C29H48O13	603.3022	6	0.4483	1.6552
0.0556	C18H28O10	403.1610	5	0.5556	1.5556	0.0628	C20H36O11	451.2185	3	0.5500	1.8000	0.0659	C30H50O12	601.3230	6	0.4000	1.6667
0.0548	C18H30O10	405.1766	4	0.5556	1.6667	0.0627	C20H32O11	447.1872	5	0.5500	1.6000	0.0655	C29H48O11	571.3124	6	0.3793	1.6552
0.0538	C8H14O8	237.0616	2	1.0000	1.7500	0.0626	C29H48O16	651.2870	6	0.5517	1.6552	0.0650	C14H22O7	301.1293	4	0.5000	1.5714
0.0529	C19H28O7	367.1762	6	0.3684	1.4737	0.0620	C10H16O7	247.0823	3	0.7000	1.6000	0.0643	C29H48O12	587.3073	6	0.4138	1.6552
0.0524	C8H14O5	189.0768	2	0.6250	1.7500	0.0616	C38H62O17	789.3914	8	0.4474	1.6316	0.0642	C26H42O7	465.2858	6	0.2692	1.6154

3.2.4 Multivariate statistical analysis using only the presence and absence of molecular formulas

Although the analyte relative abundances in the mass spectra indicate the relative magnitude of their concentrations, they do not exactly correspond to sample concentrations. The relative abundance in a mass spectrum depends on various factors besides its amount in the sample, including: the ionization parameters, the functional groups and their ionization efficiency, the sample matrix and so on. Kujawinski et al. (2009) and Altieri et al. (2012) used the presence and absence of molecular formulas in samples for their multivariate statistical analysis. The advantages and limitations of using presence/absence of formulas and using RA of formulas are discussed in more detail in Section 4.2.1. Using presence and absence, the RA of the molecular formulas is not directly considered. As long as the analyte appears in the mass spectrum above some defined signal-to-noise threshold, it is defined as "present", otherwise it is "absent". I conducted a similar set of statistical analysis after converting all of the measured RA values above the signal-to-noise threshold to 1 and those below to 0 in the combined LANLSOA dataset.

The HCA results shown in the dendrogram (Figure 3.13) are quite similar to those conducted with the previously described data preparation methods in Section 3.2.2 and Section 3.2.3. This indicates the binary presence/absence method compares the LANLSOA molecular composition fairly well, although one of the β-pinene SOA samples (BP9_0) is grouped with the α-pinene SOA. The first two PCs in the PCA results carry 54% of the total variance (40% on PC1 and 14% on PC2) (Figure 3.14). The LANLSOA samples are located in similar regions of the PCA biplots as shown previously (Figure 3.8 and 3.11). The primary difference between these analyses is the locations of the molecular formulas in the biplots. As show in Figure 3.15 of the bubble plot, the formulas spread evenly over the four quadrants, however the large bubbles which represent the formulas with large RAs are mostly centered around the origin. In fact, there are hundreds of molecular formulas overlapping on the origin and they are the formulas appearing in all of the LANLSOA samples, including several with large RA. The presence of these molecular formulas in all of the samples does not contribute to the sample variation, so they have zero PC loadings.
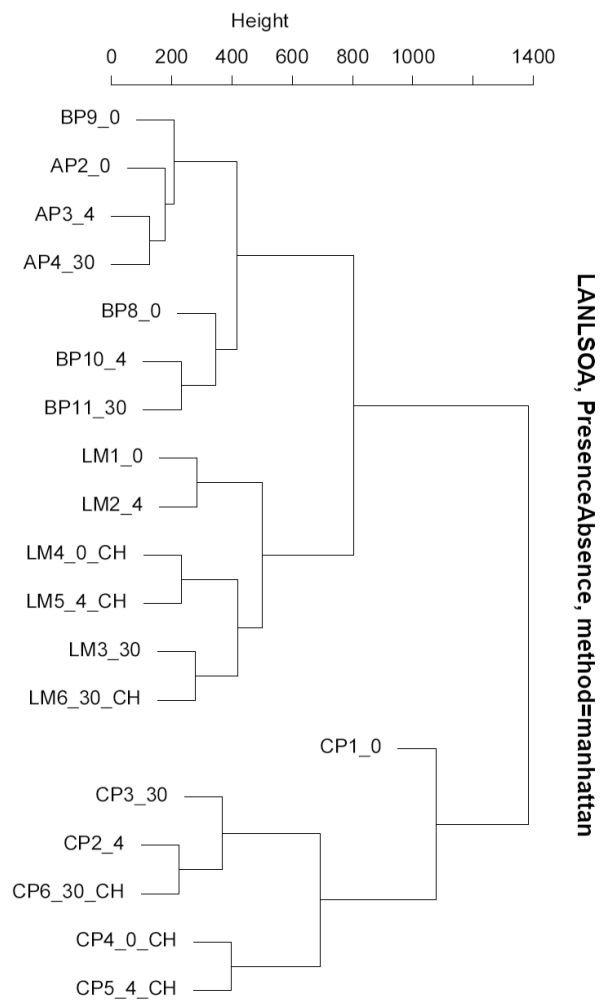
Figure 3.13 HCA dendrogram with only the presence and absence of the molecular formulas. The Manhattan distance method and complete linkage parameters were used.

Figure 3.14 PCA biplot conducted with only the formula presence and absence.

Figure 3.15 The PCA loading bubble plot with bubble size representing the relative abundances of the molecular formulas in specific samples. PCA was conducted with only the presence and absence of formulas.

The characteristic molecular formulas for the LANLSOA samples were selected using the same criteria as discussed in Section 3.2.2 and Section 3.2.3 and are listed in Table 3.4. In contrast to the characteristics observed previously, the molecular formulas in Table 3.4 are mostly compounds with molecular weights greater than 400 Da with high numbers of C atoms. Typically, the ambient organic species with relatively high concentrations have molecular weights less than 400 Da as observed using either traditional gas chromatography and mass spectrometry analysis (e.g., Samburova et al., 2013) or ultrahigh resolution analysis (Mazzoleni et al., 2012; Zhao et al., 2013). Although the characteristic formulas identified with the presence/absence data preparation might be trimers or tetramers of the first generation ozonolysis products, their abundances in atmospheric samples would be likely too low to detect well with the available analytical techniques. Thus they are not likely to be very helpful for biogenic SOA identification in ambient atmospheric samples.

Table 3.4 Characteristic molecular formulas of α- and β-pinene SOA, D-limonene SOA and β-caryophyllene SOA based on principal component analysis with only the presence and absence of molecular formulas.

α- and β-pinene SOA						D-limonene SOA						β-caryophyllene SOA					
Distance	Formula	Theor. m/z	DBE	O/C	H/C	Distance	Formula	Theor. m/z	DBE	O/C	H/C	Distance	Formula	Theor. m/z	DBE	O/C	H/C
0.0522	C17H28O4	295.1915	4	0.24	1.65	0.0659	C40H66O16	801.4278	8	0.40	1.65	0.0658	C40H66O17	817.4227	8	0.43	1.65
0.0510	C5H8O6	163.0248	2	1.20	1.60	0.0659	C29H50O14	621.3128	5	0.48	1.72	0.0658	C39H64O16	787.4122	8	0.41	1.64
0.0483	C32H46O12	621.2917	10	0.38	1.44	0.0659	C30H52O13	619.3335	5	0.43	1.73	0.0658	C39H64O17	803.4071	8	0.44	1.64
0.0446	C32H46O13	637.2866	10	0.41	1.44	0.0659	C30H52O14	635.3284	5	0.47	1.73	0.0625	C39H64O15	771.4172	8	0.38	1.64
0.0418	C18H30O4	309.2071	4	0.22	1.67	0.0659	C30H52O12	603.3386	5	0.40	1.73	0.0624	C41H66O17	829.4227	9	0.41	1.61
0.0411	C21H36O12	479.2134	4	0.57	1.71	0.0659	C37H62O16	761.3965	7	0.43	1.68	0.0624	C41H64O17	827.4071	10	0.41	1.56
0.0411	C20H34O12	465.1978	4	0.60	1.70	0.0659	C38H64O16	775.4122	7	0.42	1.68	0.0624	C40H66O15	785.4329	8	0.38	1.65
0.0411	C19H32O12	451.1821	4	0.63	1.68	0.0659	C38H64O15	759.4172	7	0.39	1.68	0.0624	C41H64O18	843.4020	10	0.44	1.56
0.0411	C21H34O13	493.1927	5	0.62	1.62	0.0659	C40H60O18	827.3707	11	0.45	1.50	0.0620	C41H62O18	841.3863	11	0.44	1.51
0.0411	C19H30O13	465.1614	5	0.68	1.58	0.0623	C39H66O16	789.4278	7	0.41	1.69	0.0598	C42H66O18	857.4176	10	0.43	1.57
0.0411	C19H34O10	421.2079	3	0.53	1.79	0.0621	C27H48O12	563.3073	4	0.44	1.78	0.0598	C30H52O11	587.3437	5	0.37	1.73
0.0411	C18H30O12	437.1665	4	0.67	1.67	0.0621	C39H62O19	833.3813	9	0.49	1.59	0.0598	C32H54O12	629.3543	6	0.38	1.69
0.0411	C20H30O13	477.1614	6	0.65	1.50	0.0621	C39H64O18	819.4020	8	0.46	1.64	0.0596	C39H64O14	755.4223	8	0.36	1.64
0.0411	C19H28O13	463.1457	6	0.68	1.47	0.0621	C40H64O19	847.3969	9	0.48	1.60	0.0591	C39H66O15	773.4329	7	0.38	1.69
0.0411	C18H28O13	451.1457	5	0.72	1.56	0.0621	C40H62O19	845.3813	10	0.48	1.55	0.0585	C40H64O17	815.4071	9	0.43	1.60
0.0411	C34H54O17	733.3288	8	0.50	1.59	0.0621	C40H66O18	833.4176	8	0.45	1.65	0.0585	C40H62O18	829.3863	10	0.45	1.55
0.0411	C17H28O12	423.1508	4	0.71	1.65	0.0621	C26H46O12	549.2917	4	0.46	1.77	0.0572	C41H66O18	845.4176	9	0.44	1.61
0.0411	C26H38O16	605.2087	8	0.62	1.46	0.0612	C40H64O18	831.4020	9	0.45	1.60	0.0548	C25H44O10	503.2862	4	0.40	1.76
0.0411	C33H52O17	719.3132	8	0.52	1.58	0.0576	C42H66O19	873.4126	10	0.45	1.57	0.0522	C40H64O15	783.4172	9	0.38	1.60
0.0411	C21H32O14	507.1719	6	0.67	1.52	0.0560	C26H46O11	533.2967	4	0.42	1.77	0.0522	C41H64O16	811.4122	10	0.39	1.56

3.2.5 Multivariate statistical analysis using scaled normalized relative abundances

It was demonstrated in Section 3.2.2 that the multivariate statistical analysis with normalized relative abundances is driven by a few of the molecular formulas with very high relative abundances. To enhance the roles of the lower RA molecular formulas, the relative abundances were re-scaled by taking the logarithm of the RA ratios (measured RA divided by the smallest measured RA in the sample) on a sample basis, in Section 3.2.3. But by taking the ratio, the effect of normalization to total RA in the sample is cancelled out. In this section, we standardize the variables (relative abundances of the formulas) on a variable basis. There are various standardization methods. Commonly used standardization techniques include auto-scaled, mean normalization, maximum normalization, range normalization, minimum-maximum transformation, half-range and central value transformation and so on (Moreda-Pineiro et al., 2001; Xue et al., 2011). We applied the most traditional standardization method, auto-scale, as a data preparation step for the LANLSOA dataset before multivariate statistical analysis. Using the converted combined dataset with normalized RAs, mean normalized RA values of a formula in all the samples were subtracted from the normalized RA, and then divided by the standard deviation of the normalized RA values of a formulas in all of the samples, using the following equation:

$$nRA_{ij}^{\ scaled} = (nRA_{ij} - \overline{x_j}) / S(x_j)$$

where $nRA_{ij}^{\ scaled}$ is the scaled normalized RA of formula j (from 1 to 2653) in sample i (from 1 to 19); nRA_{ij} is normalized RA of formula j in sample i; $\overline{x_j}$ is the mean normalized RA of formula j in the 19 samples and $S(x_j)$ is the standard deviation of normalized RA of formula j in the 19 samples. With the auto-scaled standardization, all of the variables have a zero mean (also called "centered") and unity standard deviation (also called "scaled"). All of the variables are regarded with the same significance in the multivariate statistical analysis, avoiding the large variance of the large RA formulas that drive the statistical analysis results. The "center" and "scale" conversion of the normalized RAs treats every formula with the same significance in PCA, no matter if the formula has ~100% RA or only ~1% RA in the samples, but the variation (the RA) of the a formula in all the samples are stilled considered in the multivariate statistical analysis. For example, formula 1 has 90% to 100% RA in the 19 samples, formula 2 has 0% to 10% RA in the

86

19 samples. The roles of the two formulas play in the statistics are treated as the same after

centering and scaling while the variation of the formula RA are still used in the statistics. But

both the formulas have the same range, for example, from -1 to 1, instead formula 1 is ~10

times magnitude of formula 2. This mathematical data preparation method is more general than

the methods attempted in Section 3.2.2 to 3.2.4 It is also more practical for ambient

atmospheric sample applications, since this data preparation treats every variable with the same

significance so other variables in addition to the molecular formulas from the ultrahigh

resolution analysis can be added to the dataset for statistical analysis to reveal their

relationships. An important application of this will be discussed in Chapter 4.

The results of HCA conducted with the scaled normalized RA of the formulas are shown in Figure

3.16. Again the results are similar to the dendrograms from analyses with other data

preparation methods in Section 3.2.2 to 3.2.4; the SOA samples group together with respect to

the SOA precursors, indicating their molecular composition similarities. α- and β-pinene SOA

have a more similar composition compared to D-limonene SOA, and all of the monoterpene SOA

is clustered separately from the β-caryophyllene SOA.
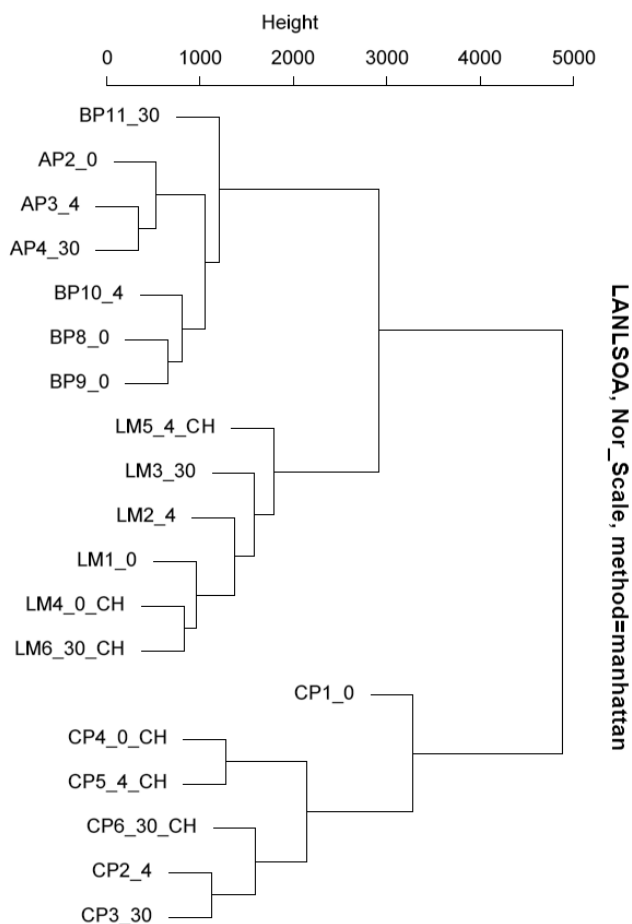
Figure 3.16 HCA dendrogram with the scaled normalized RA of the formulas. The Manhattan distance method and complete linkage were used.

In the principal component analysis with the scaled normalized RA of the molecular formulas, PC1 carries 38% of the variance and PC2 carries 16% of the total variance. The PCA biplot and score plot are shown in Figure 3.17; the score plot more clearly shows the locations of the LANLSOA samples. Overall, the samples have quite similar locations as shown in the previous PCA biplots with other data preparation methods. The molecular formulas are spread over the four quadrants and compose a large circle on the PC1-PC2 space, which can also be between seen in the PCA loading bubble plot in Figure 3.18. Most of the formulas are located at the edge of the circle, reflecting their significance in the principle component analysis.
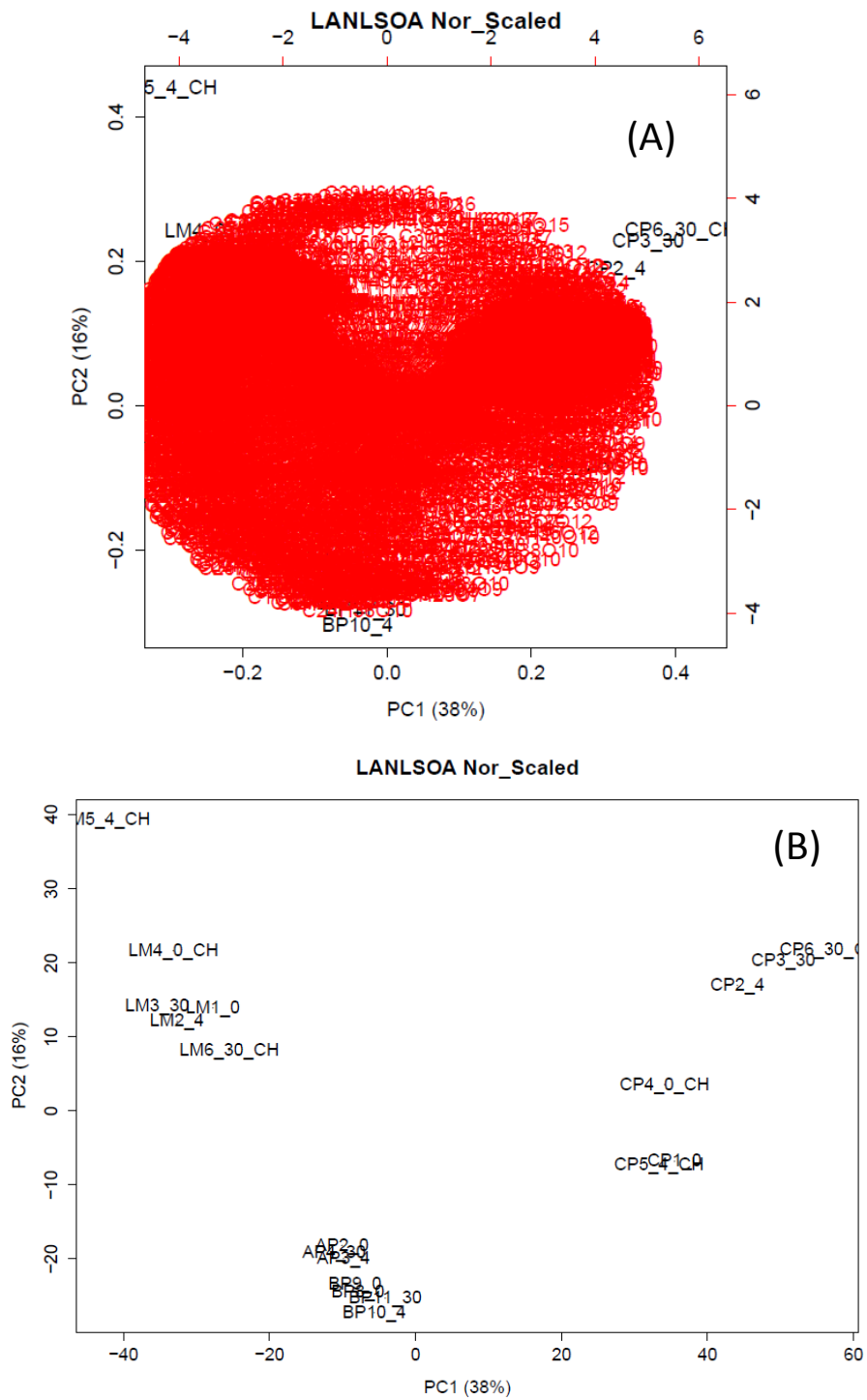
Figure 3.17 PCA biplot (A) and score plot (B) conducted with scaled normalized RA of the formulas.
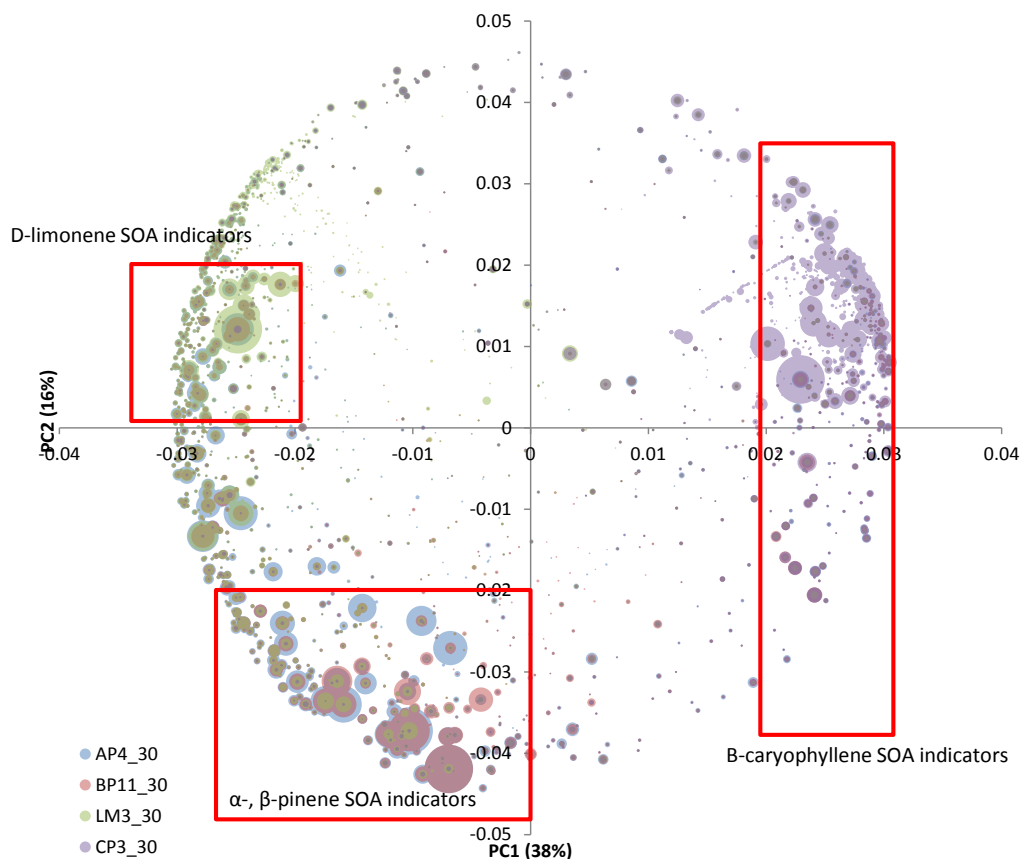
Figure 3.18 The PCA loading bubble plot with bubble size representing the relative abundances of formulas in specific samples. PCA was conducted with the scaled normalized RA of formulas. The boxes indicate where the characteristic formulas are located.

In the PCA biplot, the formulas farthest away from the origin have more contributions to the variation of the samples than the formulas around the origin. Thus, this is considered for the selection of the characteristic molecular formulas or the indicator species for specific samples in previous sections (3.2.2 to 3.2.5). However in the biplot in Figure 3.18, a majority of the formulas are on the edge of the circle and thus they have similar distances from the origin. This is because the molecular formulas are scaled and treated with equivalent significance in the statistical analysis in this case. To identify the SOA indicator species, we selected the formulas located in the similar regions as the samples they are to indicate. For example, the characteristic formulas of β-caryophyllene SOAs (located at the far right of the biplot) are the formulas also at far right of the biplot (Figure 3.18). The "similar region" rule is accompanied by the size of the bubbles. Usually the compounds with the larger amounts in the samples are better candidates for indicator species since their detection in samples will be enhanced. Thus for β-caryophyllene

90

SOA indicators, a region with PC1 > 0.02 was selected. Similarly, the region with PC1 < 0 and PC2 < -0.02 was selected for α- and β-pinene SOA and the region with PC1 < -0.02 and 0 < PC2 < 0.02 was selected for D-limonene SOAs. There are hundreds of molecular formulas in each box illustrated in Figure 3.18 to potentially represent the specific LANLSOA samples. 20 molecular formulas with the highest RA in AP4_30, BP11_30, LM3_30 and CP3_30 (the representative samples for SOA from each precursor) are listed in Tables 3.5, 3.6 and 3.7 for α- and β-pinene SOA, D-limonene SOA and β-caryophyllene SOA. They are the molecular formulas characteristic to the specific SOA samples, in other words they are the indicator species for SOA from each of those precursors with 30% RH.

Table 3.5 Characteristic molecular formulas of α- and β-pinene SOA based on principal component analysis with the scaled normalized RA, and their detection in ambient samples. See the text a description of the ambient samples. The values in the ambient sample columns are the formulas relative abundances. "y" and "n" represent the presence and absence of the formula in the ambient sample when RA is not available.

Formula	Theor.m/z	DBE	O/C	H/C	RA in AP4_30	RA in BP11_30	Mazzoleni et al., 2012 (aerosol)	Wozniak et al., 2008 (aerosol at NY)	Wozniak et al., 2008 (aerosol at VA)	Schmitt-Kopplin et al., 2010 (aerosol)	Zhao et al., 2013 (cloud 1)	Zhao et al., 2013 (cloud 2)	Mazzoleni et al., 2010 (fog)
C17H26O8	357.1555	5	0.47	1.53	100.00	100.00	49.33	y	y	4.65	n	n	3.03
C9H14O4	185.0819	3	0.44	1.56	95.38	71.21	54.62	n	n	2.90	n	n	35.98
C10H16O5	215.0925	3	0.50	1.60	57.06	28.80	83.14	n	n	6.48	n	n	26.23
C19H28O7	367.1762	6	0.37	1.47	49.68	5.08	12.38	y	y	7.16	n	n	4.53
C9H14O6	217.0718	3	0.67	1.56	38.48	26.36	17.46	n	n	4.69	n	n	7.25
C11H18O5	229.1081	3	0.45	1.64	38.40	5.48	45.85	y	y	3.97	82.56	58.82	21.35
C19H28O9	399.1661	6	0.47	1.47	34.26	4.33	44.34	y	y	4.68	n	n	n
C8H12O4	171.0663	3	0.50	1.50	32.63	28.39	23.79	n	n	1.16	n	n	16.32
C20H32O9	415.1974	5	0.45	1.60	30.21	40.24	35.26	y	y	3.16	n	n	n
C18H28O9	387.1661	5	0.50	1.56	26.59	8.31	52.8	y	y	3.72	n	n	n
C19H30O8	385.1868	5	0.42	1.58	24.47	9.95	28.66	y	y	5.53	n	n	1.53
C18H28O8	371.1711	5	0.44	1.56	21.83	11.14	39.27	y	y	5.51	n	n	2.51
C19H28O8	383.1711	6	0.42	1.47	21.37	5.19	22.39	y	y	5.84	n	n	1.30
C17H26O7	341.1606	5	0.41	1.53	16.64	8.03	26.71	y	y	7.59	24.17	n	5.08
C18H28O7	355.1762	5	0.39	1.56	15.77	8.42	21.71	y	y	7.01	24.07	n	5.64
C16H24O8	343.1398	5	0.50	1.50	15.26	10.09	60.47	y	y	5.32	n	n	3.04
C19H30O7	369.1919	5	0.37	1.58	14.19	12.58	14.16	y	y	11.95	15.69	n	4.01
C9H14O3	169.0870	3	0.33	1.56	14.08	14.61	8.7	n	n	0.68	n	n	13.39
C17H26O9	373.1504	5	0.53	1.53	13.59	7.52	50.91	y	y	4.18	n	n	n
C10H16O3	183.1027	3	0.30	1.60	13.46	2.30	10.42	n	n	2.36	n	n	53.96

C18H30O9	389.1817	4	0.50	1.67	12.71	31.09	24.86	y	y	5.33	n	n	n
C20H32O7	383.2075	5	0.35	1.60	10.55	11.69	8.98	y	y	7.17	n	14.36	2.37
C19H32O8	387.2024	4	0.42	1.68	8.99	13.13	17.43	y	y	3.47	n	n	4.82
C18H30O8	373.1868	4	0.44	1.67	7.31	7.90	26.08	y	y	3.58	n	n	n
C18H30O7	357.1919	4	0.39	1.67	5.30	8.60	19.86	y	y	6.34	22.56	19.01	2.56
C16H26O10	377.1453	4	0.63	1.63	5.08	25.27	9.17	n	n	n	n	10.01	1.18

Table 3.6 Characteristic molecular formulas of D-limonene SOA based on principal component analysis with the scaled normalized RA, and their detection in ambient samples. See the text for a description of the ambient samples. The values in the ambient sample columns are the formulas relative abundances. "y" and "n" represent the presence and absence of the formula in the ambient sample when RA is not available.

Formula	Theor. m/z	DBE	O/C	H/C	RA in LM3_30	Mazzoleni et al., 2012 (aerosol)	Wozniak et al., 2008 (aerosol at NY)	Wozniak et al., 2008 (aerosol at VA)	Schmitt-Kopplin et al., 2010 (aerosol)	Zhao et al., 2013 (cloud 1)	Zhao et al., 2013 (cloud 2)	Mazzoleni et al., 2010 (fog)
C11H18O6	245.1031	3	0.55	1.64	100.00	30.72	y	y	2.89	n	n	15.92
C11H20O6	247.1187	2	0.55	1.82	27.05	5.98	y	y	1.41	n	n	5.69
C8H12O5	187.0612	3	0.63	1.50	21.05	18.55	n	n	3.06	n	n	8.92
C9H14O5	201.0768	3	0.56	1.56	21.02	34.99	n	n	5.20	n	n	14.70
C10H16O7	247.0823	3	0.70	1.60	20.23	12.11	n	n	2.04	n	n	3.90
C9H16O6	219.0874	2	0.67	1.78	19.97	3.41	n	n	1.64	n	n	3.94
C11H18O8	277.0929	3	0.73	1.64	16.58	4.81	n	n	0.71	n	n	n
C10H18O6	233.1031	2	0.60	1.80	14.12	7.34	n	n	1.17	n	n	4.74
C11H16O7	259.0823	4	0.64	1.45	13.80	22.44	y	y	2.28	89.41	66.28	2.67
C11H20O7	263.1136	2	0.64	1.82	12.68	3.77	n	n	n	n	n	1.66
C11H22O6	249.1344	1	0.55	2.00	12.26	1.29	n	n	n	n	n	1.40
C10H14O7	245.0667	4	0.70	1.40	9.40	19.22	n	n	3.36	64.66	41.80	1.82
C20H32O10	431.1923	5	0.50	1.60	8.90	32.98	y	y	1.75	20.63	16.93	n
C10H14O6	229.0718	4	0.60	1.40	8.73	46.65	n	n	5.32	n	n	6.55
C10H14O5	213.0768	4	0.50	1.40	8.44	33.37	n	n	4.99	n	n	17.51
C10H18O7	249.0980	2	0.70	1.80	8.29	2.01	n	n	n	n	n	1.05
C9H14O7	233.0667	3	0.78	1.56	7.43	5.28	n	n	2.19	n	n	1.47
C9H12O5	199.0612	4	0.56	1.33	7.15	15.88	n	n	3.14	n	n	7.76
C11H16O8	275.0772	4	0.73	1.45	6.56	8.87	n	n	2.21	n	n	n
C19H32O10	419.1923	4	0.53	1.68	6.39	13.77	y	y	n	10.89	10.78	n

Table 3.7 Characteristic molecular formulas of β-caryophyllene SOA based on principal component analysis with the scaled normalized RA, and their detection in ambient samples. See the text for a description of the ambient samples. The values in the ambient sample columns are the formulas relative abundances. "y" and "n" represent the presence and absence of the formula in the ambient sample when RA is not available.

Formula	Theor.m/z	DBE	O/C	H/C	RA in CP3_30	Mazzoleni et al., 2012 (aerosol)	Wozniak et al., 2008 (aerosol at NY)	Wozniak et al., 2008 (aerosol at VA)	Schmitt-Kopplin et al., 2010 (aerosol)	Zhao et al., 2013 (cloud 1)	Zhao et al., 2013 (cloud 2)	Mazzoleni et al., 2010 (fog)
C16H26O6	313.1657	4	0.38	1.63	100.00	32.7	y	y	5.40	n	n	6.14
C14H22O7	301.1293	4	0.50	1.57	50.38	92.21	y	y	4.84	49.88	38.42	3.86
C14H24O7	303.1449	3	0.50	1.71	30.70	34.01	y	y	2.13	n	n	3.15
C16H28O6	315.1813	3	0.38	1.75	27.95	20.33	y	y	11.96	n	n	5.23
C16H28O8	347.1711	3	0.50	1.75	23.10	14.13	y	y	1.33	n	n	n
C29H46O11	569.2967	7	0.38	1.59	21.71	2.93	n	n	n	n	n	n
C14H22O5	269.1394	4	0.36	1.57	18.71	66.35	y	y	5.10	27.30	21.40	13.49
C29H48O11	571.3124	6	0.38	1.66	18.28	2.16	y	n	n	12.50	n	n
C15H24O8	331.1398	4	0.53	1.60	15.98	55.67	y	y	3.75	n	n	2.67
C14H22O6	285.1344	4	0.43	1.57	15.74	83.91	y	y	4.95	n	n	7.92
C15H26O9	349.1504	3	0.60	1.73	14.46	n	n	n	n	10.09	8.21	n
C15H24O4	267.1602	4	0.27	1.60	14.42	49.67	y	y	3.40	n	n	26.35
C15H24O7	315.1449	4	0.47	1.60	13.43	78.48	y	y	4.77	45.47	35.92	5.47
C11H18O3	197.1183	3	0.27	1.64	13.05	5.4	n	n	0.97	n	n	10.20
C13H20O5	255.1238	4	0.38	1.54	12.85	65.51	y	y	5.06	26.85	21.66	11.84
C28H46O11	557.2967	6	0.39	1.64	12.78	3.1	n	n	0.94	14.02	n	n
C29H46O10	553.3018	7	0.34	1.59	12.77	1.51	n	n	1.21	n	1.94	n
C29H46O12	585.2917	7	0.41	1.59	12.68	4.91	n	n	n	4.96	n	n
C16H28O7	331.1762	3	0.44	1.75	12.41	21.47	y	y	3.03	n	n	3.75
C14H22O4	253.1445	4	0.29	1.57	11.67	20.58	y	y	4.45	n	n	16.00

Similar to the characteristic molecular formulas identified using the normalized RAs, the indicator species show clear terpenoid characteristics. They contain C numbers of ~ 10, 15, 20 or 30 representing first generation products and their dimers. The selected monomers retain the precursor DBE values and the dimers have DBE values that are 2 or 3 higher than the precursor DBE consistent with the reported accretion reactions (e.g., Kundu et al., 2012). Some of the identified indicator species were previously proposed by other researchers as marker species. To confirm these indicator species, they were examined in selected ambient atmospheric samples that were influenced by biogenic emissions and are expected to contain ambient biogenic SOA. The ambient atmospheric samples used for this comparison are from published studies with ultrahigh and high resolution MS analysis methods for the identification of molecular formulas. They include: two aerosol-derived water-soluble organic carbon (WSOC) samples collected at rural sites from Wozniak et al. (2008): winter sample "NY" (sample site: Millbrook, NY) and summer sample "VA" (Sample site: Gloucester County, VA); one biomass burning affected aerosol-derived WSOC sample from Schmitt-Kopplin et al. (2010); one aerosol-derived WSOC sample from Mazzoleni et al. (2012) collected at the Strom Peak Laboratory (SPL) during the summer; two winter cloud samples from SPL reported in Zhao et al. (2013) and one winter fog sample from Mazzoleni et al. (2010). Overall, the majority of the identified biogenic SOA indicator species were detected in the ambient samples.

The aerosol-derived WSOC from Mazzoleni et al (2012) is rich with SOA indicators from all of the precursors in this study. Samples from Wozniak et al (2008) contain many α- and β-pinene SOA indicator species, followed by several β-caryophyllene SOA indicators and some D-limonene SOA indicator compounds. Almost all the α- and β-pinene SOA indicator species were detected in the samples reported by Schimitt-Kopplin et al. (2010). Around half of the indicator species appeared in the fog water reported by Mazzoleni et al. (2010), confirming the biogenic SOA influence on the sample component as reported by the authors. The cloud water samples in Zhao et al. (2013) do not contain many of the identified biogenic SOA indicator species, probably due to the lower emissions of biogenic precursors at the mountain site and slow atmospheric processing with the lower winter temperature compared to other ambient samples. But as will be seen in Chapter 5, though not many biogenic SOA indicator species were observed in the cloud samples, the studied clouds contain a high number of biogenic SOA components when

compared to the LANLSOA samples. Probably the high RA indicator species were converted to other compounds by aqueous processing.

In summary, the biogenic indicator species identified using PCA with scaled normalized RA were confirmed with the atmospheric observations. These species could be used in atmospheric source apportionment and perhaps be accommodated in to atmospheric models to improve the tracking of biogenic SOA.

Chapter 4 Aerosol-derived water-soluble organic carbon

Water-soluble organic carbon (WSOC) comprises a large fraction of aerosol organic mass. It is relevant to aerosol hygroscopicity and thus it plays an important role toward the aerosol indirect effect on the climate system. In this chapter, the molecular characterization of 34 aerosol-derived WSOC samples is presented. The organic compositions of the WSOC samples were compared using multivariate statistical analysis methods. Other aerosol measurements and meteorological parameters were also examined together with the WSOC molecular composition statistically. The relationships among the meteorological parameters and the aerosol components (inorganics and organics) are discussed.

4.1 Molecular characterization of daily water-soluble organic carbon

4.1.1 Molecular formula assignments

The molecular compositions of 24-hour samples of aerosol-derived WSOC were analyzed using negative ion mode ESI ultrahigh resolution FT-ICR mass spectrometry. Generally, $78-88\%$ of the total ion current for each WSOC sample mass spectra was assigned to unambiguous molecular formulas. Approximately 3000 monoisotopic molecular formulas, with the form of $C_cH_hN_nO_oS_s$, were identified in each sample. Consistent with previous observations of biogenic SOA (Putman et al., 2012), aerosol WSOC (Mazzoleni et al., 2012), fog water (Mazzoleni et al., 2010) and cloud water (Zhao et al., 2013), a large percentage of monoisotopic formulas were identified with their corresponding polyisotopic molecular formulas. This observation provides confirmation of the molecular formula assignments. In general, corresponding polyisotopic molecular formulas with ^{13}C were observed for over 80% of the monoisotopic formulas in each sample and those with ^{34}S were observed for ~70% of the assigned S-containing monoisotopic formulas in each sample. The molecular weights of the molecular formulas containing ^{13}C are 1.0034 Da higher than the corresponding monoisotopic molecular formulas, indicating the anions were singly charged. The fraction of the peaks assigned in the spectrum, the number of monoisotopic formulas and the percentages of the corresponding polyisotopic formulas identified in each sample are listed in Table 4.1. Several common molecular formulas were found in the sample blanks. Since the blanks were analyzed in between the samples, it is possible to observe some common species as carryover from the samples in the ionization source. The trace carryover

residues were likely resuspended and ionized by the clean solvent when the blanks were analyzed. In addition, the intensities of the common molecular formulas in the blanks are much lower than in the samples. As discussed by Mazzoleni et al. (2012), it takes a much longer time for the ions in blank samples to fill in the ion trap using the auto gain control ($n = 1x10^6$) than the sample ions. Thus, the amount of time require for analysis further reflects the low intensities of the blank ions. Due to the potential for carryover in the electrospray ion source and the low intensities of common formulas and the amount of time required for analysis, the common molecular formulas were not removed from the samples.

Table 4.1 Summary of the daily WSOC molecular formula assignment and the number of common monoisotopic formulas with the chamber generated biogenic SOA (LANLSOA samples discussed in Chapter 3).

Sample ID	Sample name	Assignable peaks (%)	RMS of signal-to-noise ratio	Assignment threshold (%)*	Number of monoisotopic molecular formulas	Number of CHO formulas	Number of CHNO formulas	Number of CHOS formulas	Number of CHNOS formulas	13C identified in all the monoisotopic formulas (%)	34S identified in sulfur-containing monoisotopic formulas (%)	Number of common formulas with LANLSOA
F062510r3	0625	87.01	0.090	0.5	4035	1779	1298	724	234	83.39	68.16	1114
F062610	-	85.71	0.042	0.3	2954	1205	919	588	242	85.51	67.00	-
F062610r	0626	84.43	0.048	0.3	2797	1169	903	483	242	86.70	67.17	965
F062710r2	-	85.61	0.088	0.5	3115	1263	1020	602	229	84.69	70.52	-
F062710r3	0627	84.63	0.061	0.4	2586	1116	849	447	174	86.72	68.72	969
F062810	0628	85.23	0.019	0.1	4147	1498	1481	743	425	83.41	69.01	1113
F062810-rep	0628	85.68	0.033	0.2	3092	1249	1077	531	235	86.38	71.54	1026
F062910	0629	86.42	0.034	0.2	3566	1425	1150	656	335	84.89	67.71	1013
F063010r	-	85.95	0.058	0.3	2862	1177	871	565	249	84.00	71.01	-
F063010r2	0630	85.61	0.046	0.3	2528	1127	834	414	153	86.43	69.14	939
F070110r	0701	84.64	0.032	0.2	3629	1295	1215	667	452	85.01	67.20	1014
F070210	-	84.00	0.117	0.7	2351	1102	717	445	87	86.52	71.62	-
F070210r2	0702	78.91	0.071	0.4	2309	1127	683	401	98	86.92	68.54	898
F070310-25pct	-	84.71	0.031	0.2	3207	1256	1050	569	332	86.31	70.48	-
F070310-50pct	0703	85.11	0.036	0.2	3130	1266	1031	557	276	85.30	70.59	997
F070410	0704	84.29	0.066	0.4	3791	1394	1158	767	472	83.22	64.25	964
F070510	0705	86.56	0.093	0.6	2951	1162	1037	576	176	85.60	68.35	966
F070610	0706	85.76	0.021	0.1	3702	1328	1192	752	430	83.93	66.50	1064
F070710	0707	86.58	0.053	0.3	3644	1446	1161	682	355	84.71	67.12	1050
F070710-rep-r2	0707	86.40	0.084	0.5	3180	1344	972	587	277	85.09	68.98	1039
F070810r	0708	83.20	0.045	0.3	2682	1209	768	472	233	86.17	66.67	921
F070910	0709	86.70	0.081	0.5	3751	1586	1190	641	334	82.59	62.46	993
F071010	0710	84.63	0.036	0.2	2851	1283	866	484	218	85.90	68.23	918
F071010-rep	0711	86.96	0.061	0.4	2816	1338	897	402	179	85.83	70.91	924
F071110	0711	83.01	0.048	0.3	2764	1095	962	485	222	85.60	71.15	953
F071210	0712	85.95	0.051	0.3	3341	1224	1200	657	260	84.74	69.47	1006
F071310r	0713	85.77	0.113	0.7	2706	1133	844	570	159	86.73	69.82	975
F071310-rep	0713	84.62	0.078	0.5	2294	1045	682	458	109	87.75	71.60	910

F071410	0714	82.67	0.064	0.4	3091	1159	1137	498	297	87.19	69.31	989
F071510	0715	84.38	0.096	0.6	3092	1218	1151	502	221	85.96	64.87	1022
F071610	0716	84.50	0.082	0.5	3023	1231	1110	511	171	85.01	67.16	1032
F071710	0717	79.60	0.066	0.4	3174	1224	1267	504	178	85.32	69.11	1017
F071710-rep	0717	79.39	0.069	0.4	3204	1242	1223	504	233	85.04	69.34	1025
F071810	0718	83.29	0.077	0.5	2755	1147	950	506	152	84.46	72.64	940
F071910r	0719	85.42	0.080	0.5	2766	1229	938	513	86	85.94	65.94	1010
F072010r	0720	85.55	0.068	0.4	3274	1597	1118	428	131	82.93	72.81	995
F072110	0721	82.45	0.056	0.3	2609	1167	873	402	167	85.36	69.42	904
F072210	0722	83.98	0.057	0.3	2890	1139	1104	423	224	84.74	70.79	897
F072210-rep	0722	86.49	0.076	0.5	2881	1201	1052	446	182	86.40	70.38	961
F072310r	0723	84.44	0.108	0.6	2532	1099	806	506	124	86.49	70.18	951
F072410	0724	84.31	0.089	0.5	3291	1415	1075	598	203	84.44	69.79	1118
F072410-rep	0724	84.96	0.095	0.6	3131	1357	992	601	181	85.40	70.59	1078
F072510	0725	82.40	0.049	0.3	3193	1332	1087	503	271	84.81	71.32	1075
F072610	0726	83.63	0.086	0.5	3287	1378	1108	539	262	84.76	69.54	1107
F072710	-	82.36	0.077	0.5	2848	1193	987	479	189	86.55	73.20	-
F072710r	0727	82.67	0.083	0.5	3228	1314	1076	584	254	85.66	69.57	1050
F072710-rep	0727	81.84	0.082	0.5	3187	1301	1035	599	252	86.29	69.10	1060
F072810	0728	85.17	0.051	0.3	3095	1245	1036	587	227	85.43	70.88	1037

* Six times the root-mean-square of the signal-to-noise ratio of the relative abundance was applied as an assignment threshold.

Similar patterns among the 34 WSOC samples can be observed in the reconstructed mass spectra containing only the identified monoisotopic molecular formulas. The reconstructed mass spectra of selected samples are shown in Figure 4.1. Consistent with previously reported observations from ultrahigh resolution mass spectrometry of atmospheric samples (*e.g.*, aerosols (Schmitt-Kopplin et al., 2010), clouds (Zhao et al., 2013)), the mass spectra are highly complex and contain a high density of analyte peaks. The negative ions have mass ranges from m/z 100 – 700 and the high relative abundance peaks are in the range of m/z 200 – 400. Interestingly in all the mass spectra, two tall peaks in the range of m/z 500 - 600 stand out from the others due to their high relative abundances. They are N- and S-containing compounds with the identified molecular formulas of $C_{24}H_{43}NO_{10}S$ (m/z 536.2535) and $C_{26}H_{47}NO_{10}S$ (m/z 564.2848). In some of the mass spectra (*e.g.*, 0626, 0710, 0720 in Figure 4.1), there are other tall peaks between m/z 500 and m/z 600. Many of them are molecular formulas in the same class, including the molecular formulas $C_{22}H_{39}NO_{10}S(CH_2)_{0-6}$ and $C_{24}H_{41}NO_{10}S(CH_2)_{0-4}$. The DBE values of these two CH_2-homologous series are 4 and 5, respectively. Considering the high number of oxygen atoms and low DBE values, these high relative abundance species are likely nitrooxy organosulfates. Organosulfates and nitrooxy organosulfates have been observed previously in ambient biogenic aerosols (Surratt et al., 2006; Surratt et al., 2008).
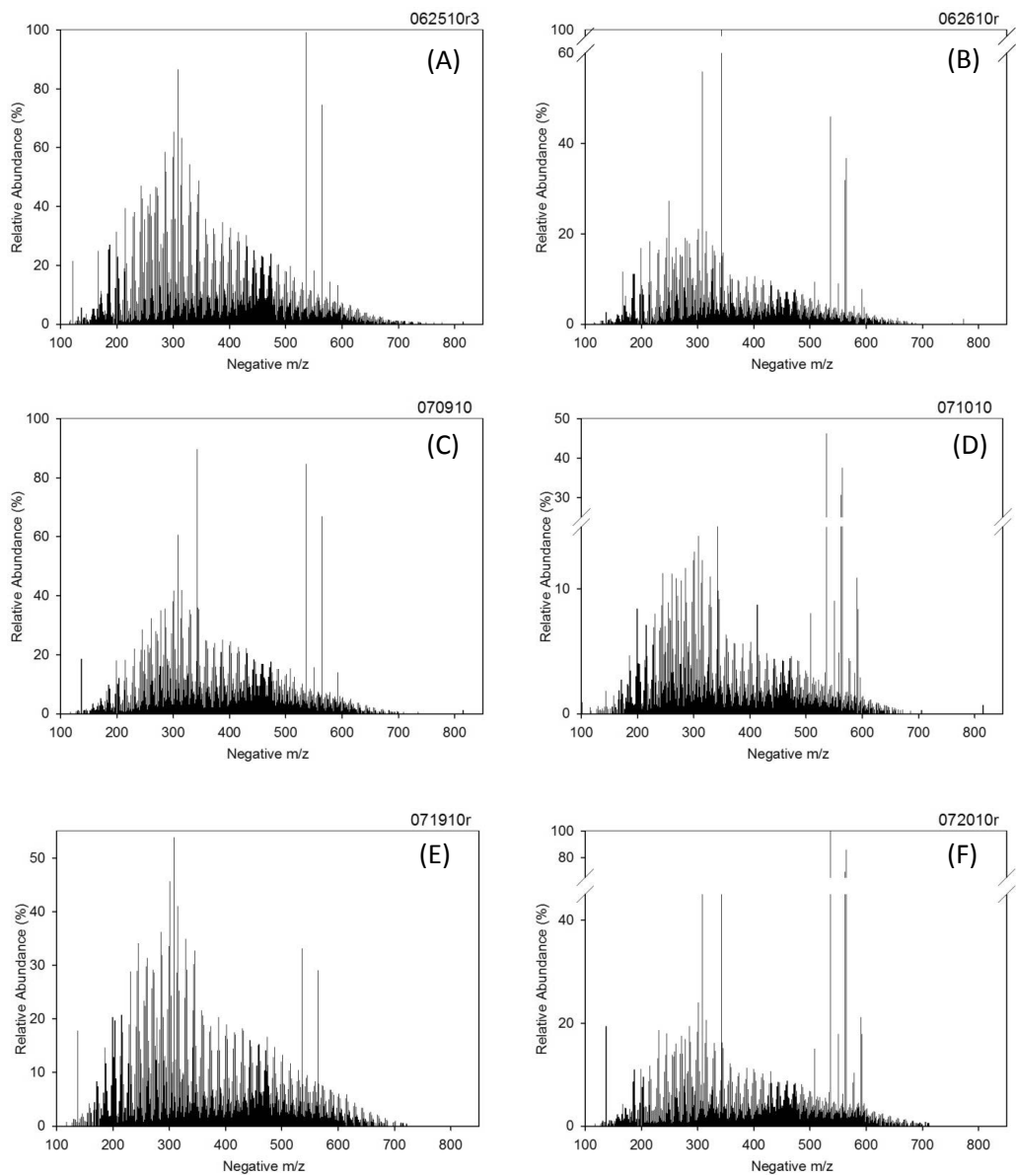
Figure 4.1 Reconstructed mass spectra of selected daily WSOC sample 0625 (A), 0626 (B), 0709 (C), 0710 (D), 0719 (E) and 0720 (F).

Atmospheric samples analyzed using ultrahigh resolution MS usually contain several different compound groups based on the elemental composition (Wozniak et al., 2008; Altieri et al., 2009; Mazzoleni et al., 2010). The 4 elemental groups include: CHO compounds (molecular formulas with C, H and O), CHNO compounds (formulas with C, H, O and N), CHOS compounds (formulas with C, H, O and S) and CHNOS compounds (formulas with C, H, N, O and S) and are commonly observed in atmospheric samples. In all of the daily WSOC samples, the CHO formulas represent the largest number fraction of the formulas identified, followed by CHNO, CHOS and CHNOS molecular formulas. This is consistent with other studies (*e.g.*, Schmitt-Kopplin et al., 2010; Mazzoleni et al., 2012). The numbers of identified molecular formulas for each of the elemental compositions of each sample are listed in Table 4.1.

4.1.2 CH_2-homologous series analysis

CH_2-homologous series of molecular formulas were clearly observed in all of the WSOC samples and identified using the Kendrick mass defect analysis. As shown in the Kendrick mass plots in Figure 4.2, the dots composing a horizontal straight line represent the molecular formulas differing by $-CH_2-$. This phenomenon has been commonly observed in natural organic matter (Kujawinski et al., 2002; Sleighter and Hatcher, 2007) and atmospheric organic matters (*e.g.,* Zhao et al., 2013). It should be noted that the addition of $-CH_2-$ can occur at any position in the molecular structure and does not necessarily indicate simple lengthening of the carbon backbone (*e.g.*, alkanes).
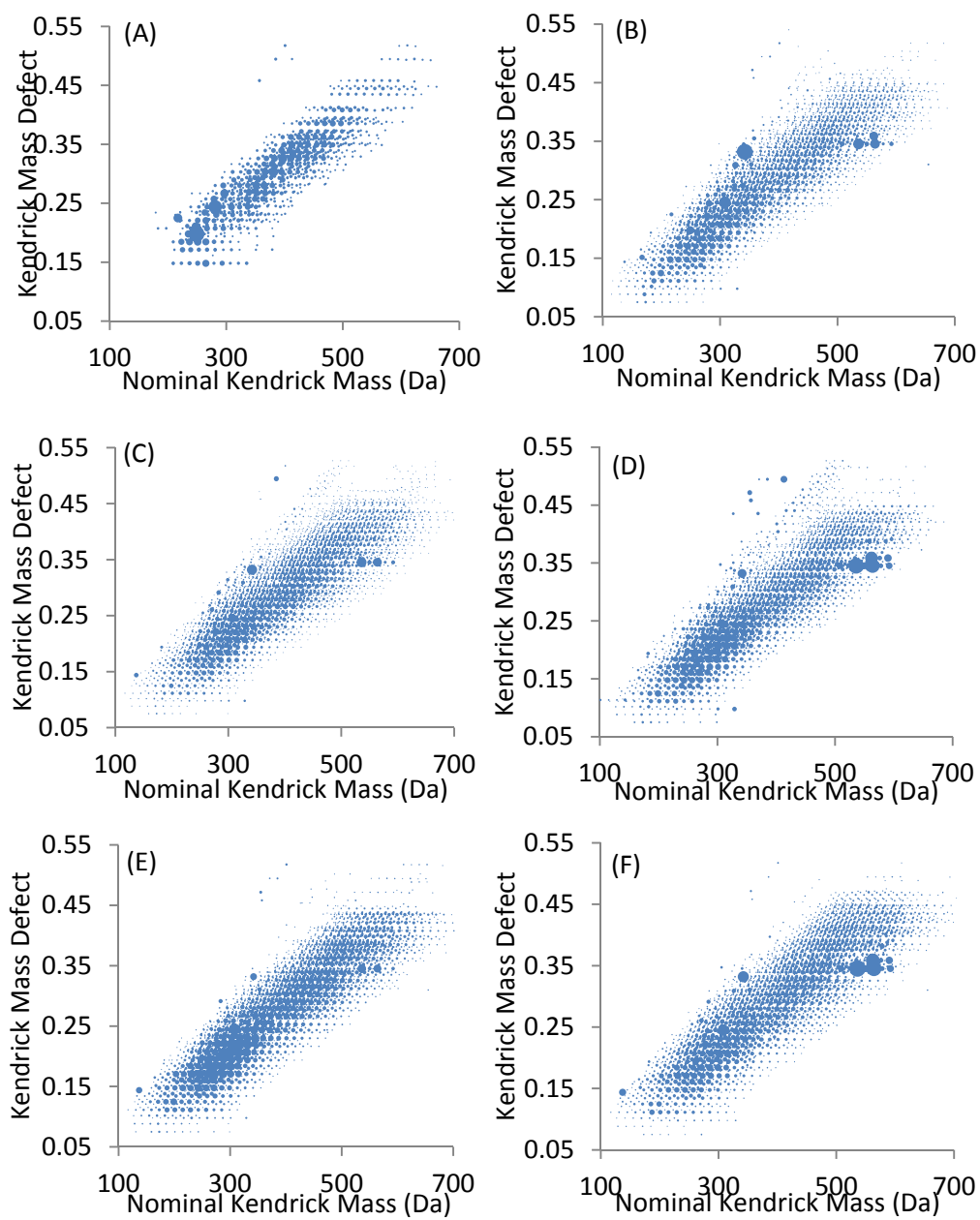
Figure 4.2 Kendrick mass plots for selected WSOC samples 0625 (A), 0626 (B), 0709 (C), 0710 (D), 0719 (E) and 0720 (F). The sizes of the dots represent the relative abundance of the molecular formulas in the samples.

4.1.3 Elemental ratios and double bond equivalents

The elemental ratios of the WSOC samples, including the oxygen-to-carbon ratio (O/C) and the hydrogen-to-carbon ratio (H/C), reflect the chemical properties of the aerosol-derived WSOC. The O/C and H/C ratios of the compounds within each of the elemental groups (CHO, CHNO, CHOS and CHNOS) of selected samples are shown in the van Krevelen diagrams of Figure 4.3. Overall, the compounds of the same elemental group in different WSOC samples have quite similar distributions in the van Krevelen diagrams. Almost all the CHO compounds have O/C values < 1.0. The high relative abundance CHO compounds are typically within the H/C range of 1.2 - 2.0. Biogenic SOA compounds fall into the space defined by the following ranges of elemental ratios: O/C of 0.3 - 1.0 and H/C of 1.2 - 2.0 according to previous studies (Reinhardt et al., 2007; Putman et al., 2012). Mazzoleni et al. (2012) studied the WSOC molecular composition of composited aerosol WSOC samples collected at the SPL from July 13th 2010 to July 20th 2010. They observed a significant number of WSOC molecular formulas common to those in monoterpene and sesquiterpene ozonolysis SOA from chamber experiments.

109
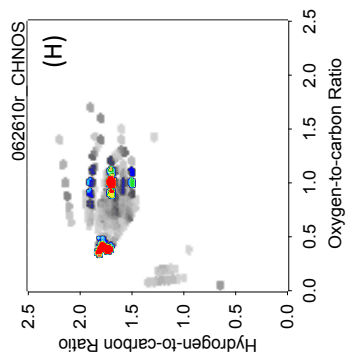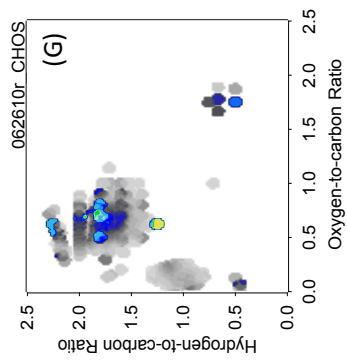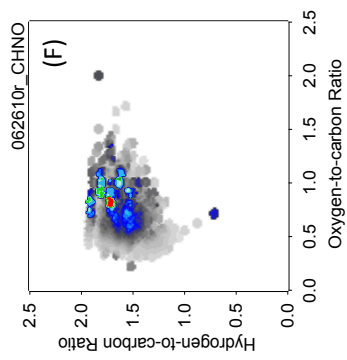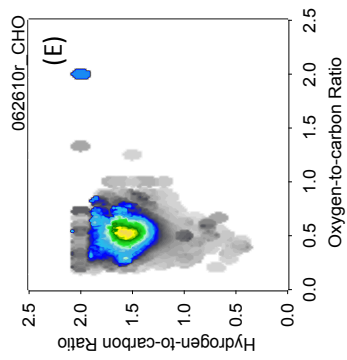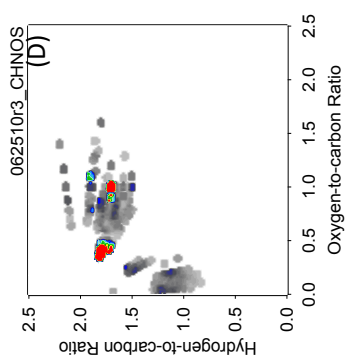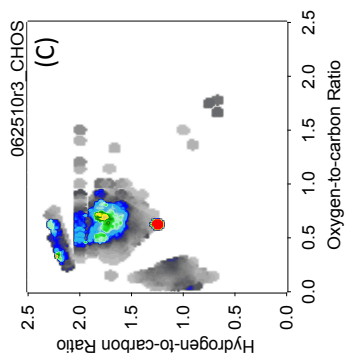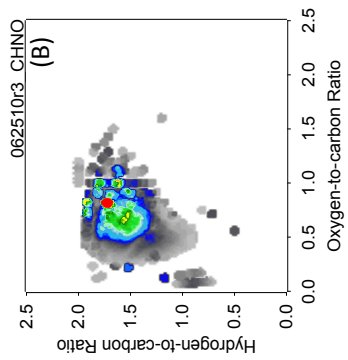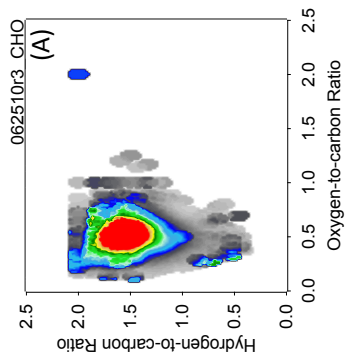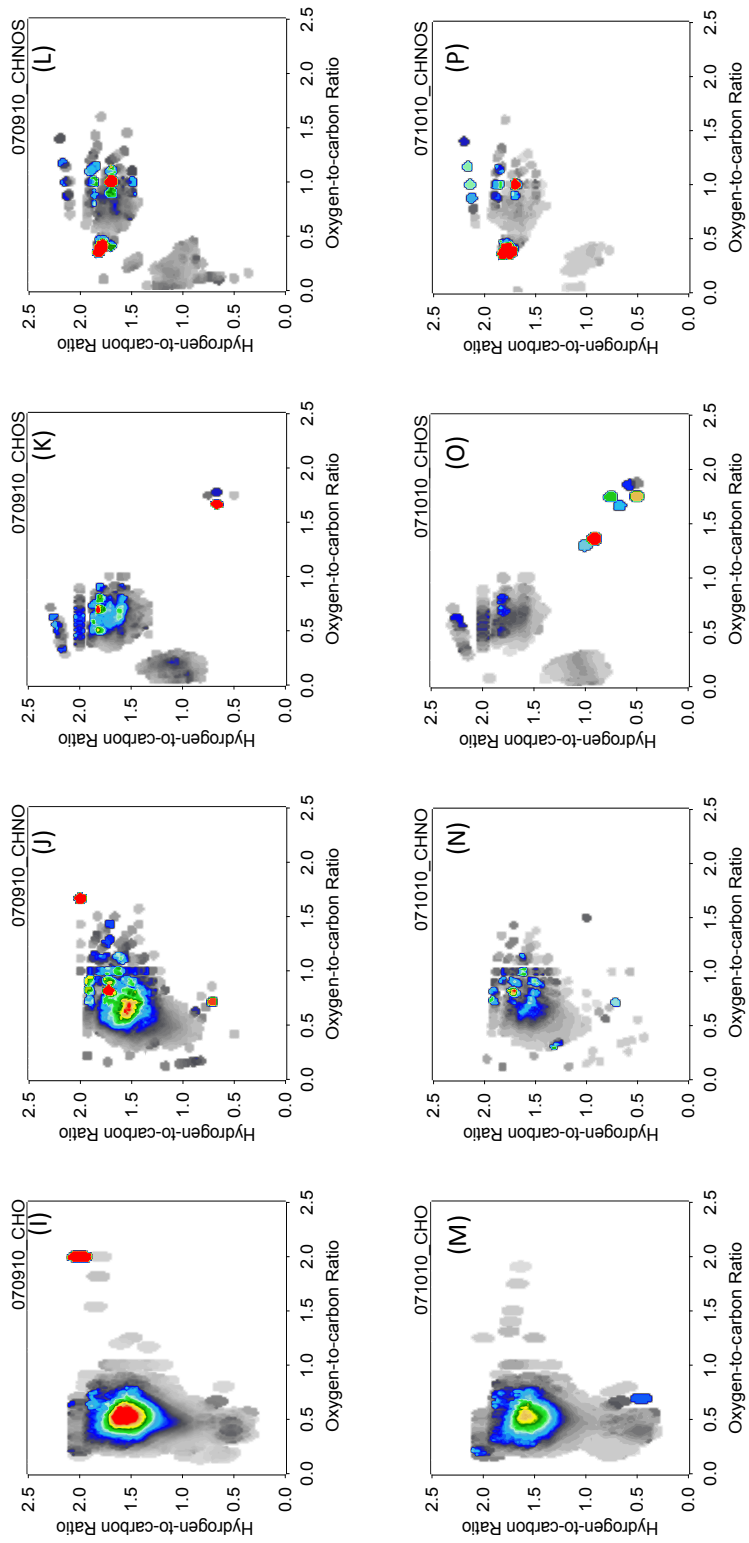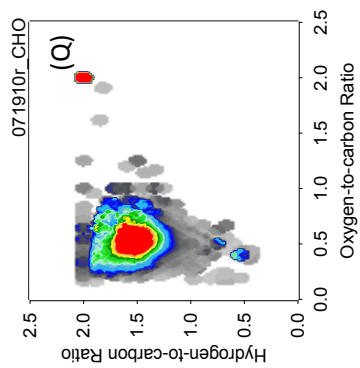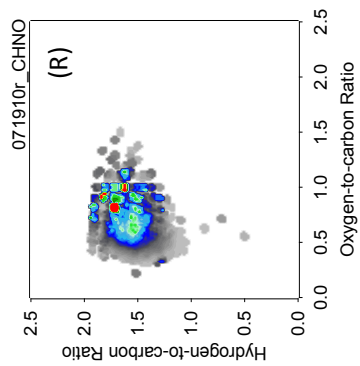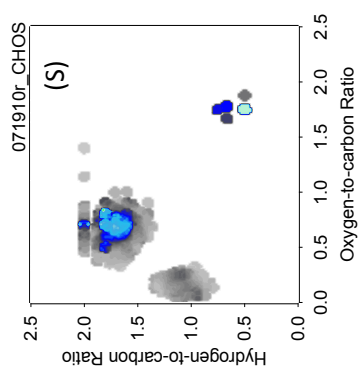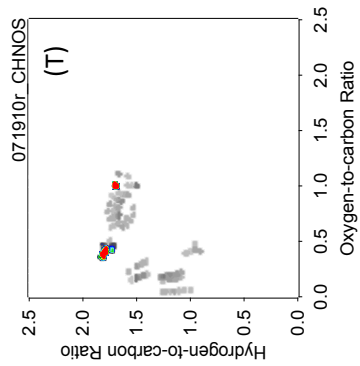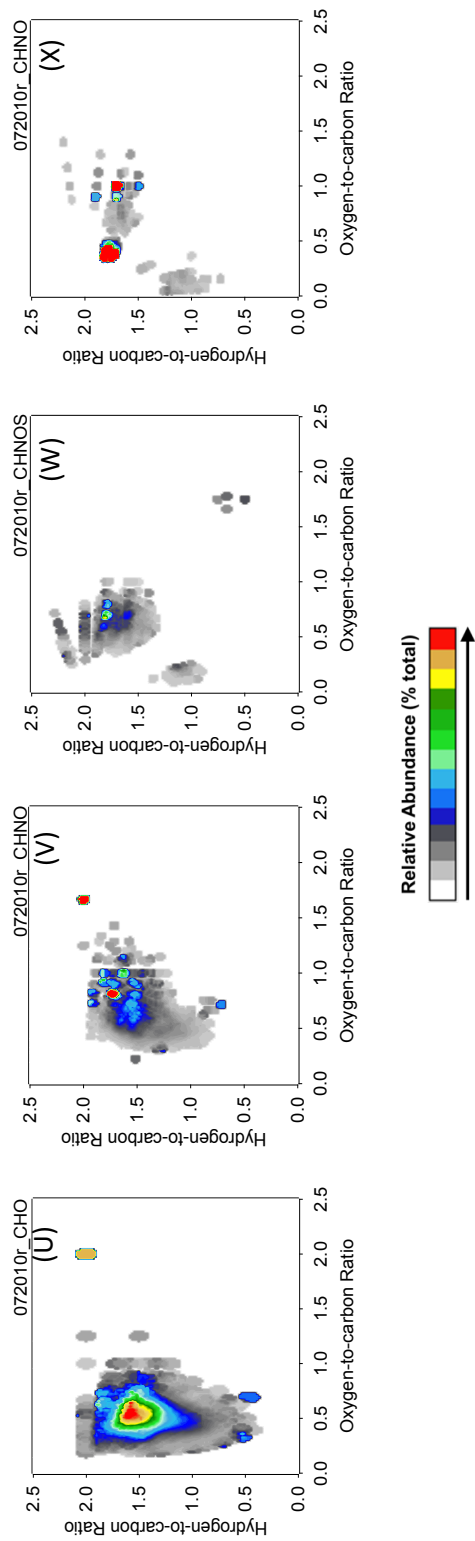
Figure 4.3 Isoabundance van Krevelen plots for sample 0625 CHO compounds (A), CHNO compounds (B), CHOS compounds (C) and CHNOS compounds (D); sample 0626 CHO compounds (E), CHNO compounds (F), CHOS compounds (G) and CHNOS compounds (H); sample 0709 CHO compounds (I), CHNO compounds (J), CHOS compounds (K) and CHNOS compounds (L); sample 0710 CHO compounds (M), CHNO compounds (N), CHOS compounds (O) and CHNOS compounds (P); sample 0719 CHO compounds (Q), CHNO compounds (R), CHOS compounds (S) and CHNOS compounds (T); sample 0720 CHO compounds (U), CHNO compounds (V), CHOS compounds (W) and CHNOS compounds (X).

111

Similar to the CHO compounds, the CHNO compounds of different samples are distributed similarly in the van Krevelen plots (Figure 4.3). They form a triangle shape in the plot, towards the point of O/C = 0 and H/C = 2.0, with all the CHNO compounds having H/C ≤ 2.0. Unlike the CHO compounds in which the high relative abundance formulas are centered around an O/C value of 0.5, the high relative abundance CHNO formulas are in the O/C range of 0.5 - 1.0. Thus, the CHNO compounds are more oxidized than the CHO compounds.

S-containing compounds (CHOS and CHNOS) are distributed clearly into separate regions in the van Krevelen diagrams (Figure 4.3), reflecting the presence of different functional groups in their molecular structures. One subclass of the S-containing compounds form a triangle shape in the plot towards the point of O/C = 0 and H/C = 2.0 similar to the CHNO compounds, but with higher H/C values ranging up to H/C = 2.5. Another subclass is the aromatic S-containing compounds with low O/C (< 0.5) and low H/C (< 1.5) ratios. They may contain reduced S in the molecular structures (*e.g.,* sulfide, disulfide). Reduced S-containing molecular formulas have been observed previously in aerosols (Mazzoleni et al., 2012; Kundu et al., 2013) and in clouds (Zhao et al., 2013). Currently, the structure and sources of these compounds are unknown, however they likely originate from anthropogenic emission sources. A small number of CHOS compounds comprise the third subclass. They are highly oxidized and unsaturated compounds with O/C 1.5 - 2.0 and H/C around 0.5. Example molecular formulas, include: $C_8H_4O_{16}S$, $C_9H_6O_{17}S$, $C_8H_4O_{15}S$ and so on. They likely contain highly oxidized S functional groups such as sulfate or sulfite.

(A)

(B)

113
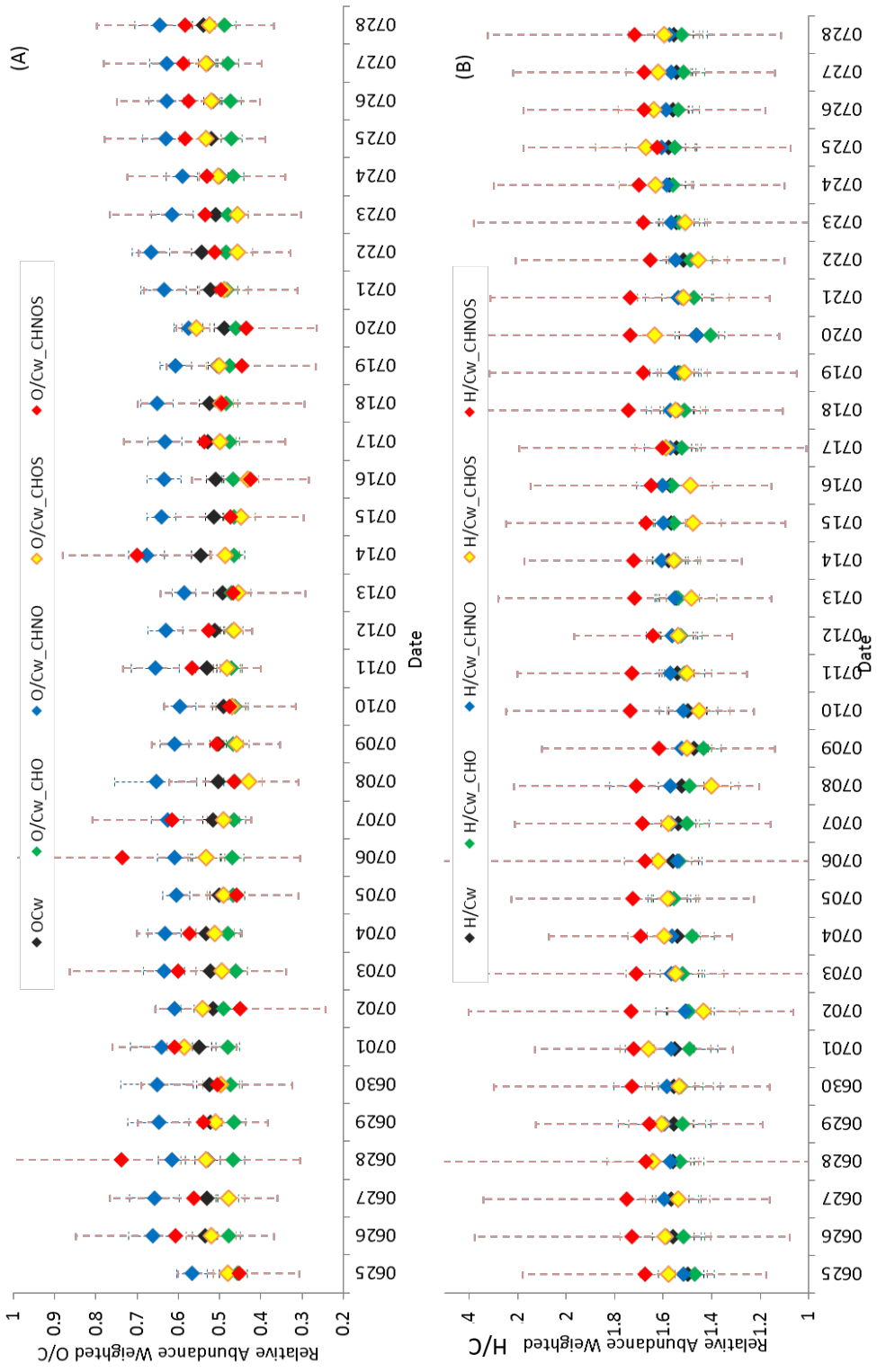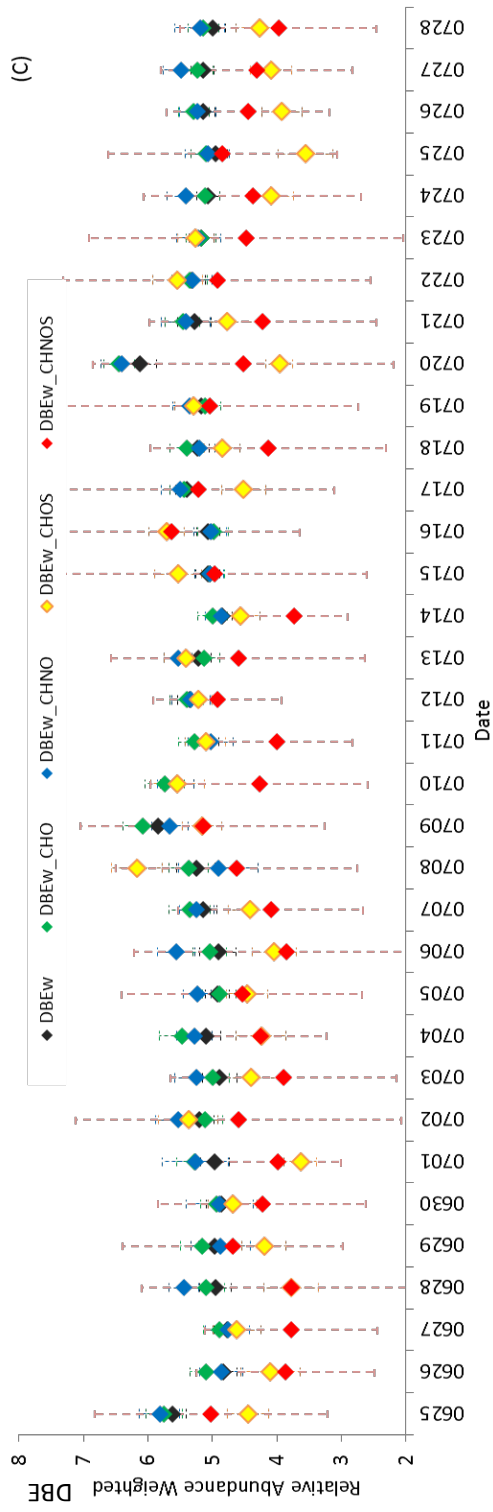
Figure 4.4 Relative abundance weighted trends of oxygen-to-carbon ratios (A), hydrogen-to-carbon ratio (B) and DBE values (C) across the study period. The error bars show the propagated standard error from the standard deviations.

The elemental ratios and DBE values for all of the WSOC samples with respect to their elemental groups are shown in Figure 4.4. Shown in the plots are the relative abundance (RA) weighted elemental ratios and DBE (O/C$_w$, H/C$_w$ and DBE$_w$). These values represent the relative amounts of the individual molecular formulas. The O/C$_w$ ratios determined for the CHO compounds of each sample are quite similar across the samples (~0.47 ± 0.07). The CHNO compounds have consistent and relatively high O/C$_w$, indicating they are more oxidized. In most WSOC samples O/C$_w$ ratios of CHNO compounds are > 0.6. Zhao et al. (2013) also reported high O/C ratios (O/C$_w$ > 0.7) for the CHNO compounds in cloud samples collected at the same site. The O/C$_w$ ratios of the S-containing compounds are highly variable across the samples with a range of 0.4 - 0.7. Most of the H/C$_w$ ratios of the WSOC samples and subgroups are in the range of 1.5 to 1.6 except for those of the CHNOS compounds, which have consistently higher H/C$_w$ ratios (~ 1.7) for the study period. The H/C$_w$ ratios of CHNO species are always higher than those of CHO species, reflecting a higher degree of saturation for the CHNO compounds compared to the CHO compounds. The DBE$_w$ values of the WSOC samples and the CHO and CHNO subgroups are consistently 5 - 6, but the values of the S-containing groups are relatively low indicating saturated structures. Some exceptions to this are observed with the high DBE$_w$ values on several days, including: 0625, 0709, 0710 and 0720. As will be shown in Section 4.2 and 4.3, the WSOC samples collected on these days have different molecular composition than others.

Overall, the O/C$_w$, H/C$_w$ and DBE$_w$ values of the samples are quite consistent and within the expected ranges for elemental ratios (*e.g.,* Aiken et al., 2008). The elemental ratios and DBE values of the CHO groups and the CHNO groups vary together, indicating a similar origin or formation mechanisms of the two groups of components. A likely source is biogenic SOA. Due to the large number of CHO and CHNO compounds and their high relative abundances compared to the S-containing compounds, the overall elemental ratios and DBE values vary with the CHO and CHNO groups. The elemental ratios and DBE values of the S-containing compounds vary largely within a sample and across the samples, indicating the various S-containing functional groups and compound structures, and thus various compound sources and atmospheric processes. N-containing compounds (CHNO and CHNOS) have high O/C ratios and high H/C ratios compared to the other subgroups in the samples, indicating that N-containing compounds are more oxidized and saturated. It is highly possible that they contain oxidized N functional groups like nitrates and nitrites.

4.1.4 Carbon number trends

The aerosol-derived WSOC molecular composition shows interesting trends with respect to the number of carbon atoms ($C_{\#}$) in the molecular formulas. The average elemental ratios and DBE values of the formulas with the same number of carbon atoms were determined for each carbon number, plotted for selected samples shown in Figure 4.5. O/C ratios decrease gradually with increasing carbon numbers. For example, a change from O/C > 1.0 at C_5 to O/C < 0.5 at C_{35} is observed in Figure 4.5. At the same time, the H/C ratios decrease with increasing carbon numbers. A subtle change from greater than 1.7 at C_5 to ~1.4, at C_{20} was observed in the H/C ratios. In contrast, the DBE values rise rapidly from C_2 to over C_{12}. This indicates the molecular formulas with higher carbon numbers, or the larger WSOC molecules, are less oxidized and less saturated than the formulas with lower numbers of carbon atoms. These compounds might be formed through the accretion of small molecules. Many of the double bonds and rings from all of the building blocks remain during the accretion reactions and thus the product molecules have higher DBE values.

Generally, for each of the compound groups (CHO, CHNO, CHOS and CHNOS) the total relative abundance of the formulas increases with the carbon number at the low end, reaches a peak between $C_{15} - C_{25}$ and then decreases (Figure 4.6). A few total relative abundance spikes were observed at C_{10}, C_{15} and C_{20}. CHO compounds have high total RAs at ~C_{15}; CHNO compounds have high total RAs ~C_{10}, and sometimes also C_{20}. The S-containing compounds show quite high total RAs at C_{10} compared to the adjacent peaks. Since the building block of the biogenic terpenes is C_5H_8, biogenic SOA appears to strongly impact the aerosol-derived WSOC composition. The tall total RA peaks of CHNOS compounds at C_{24}, C_{26} and C_{28} correspond to the peaks in the range of m/z 500-600 of the mass spectra in Figure 4.1, which might be the nitrooxy organosulfates as discussed in Section 4.1.1.

116
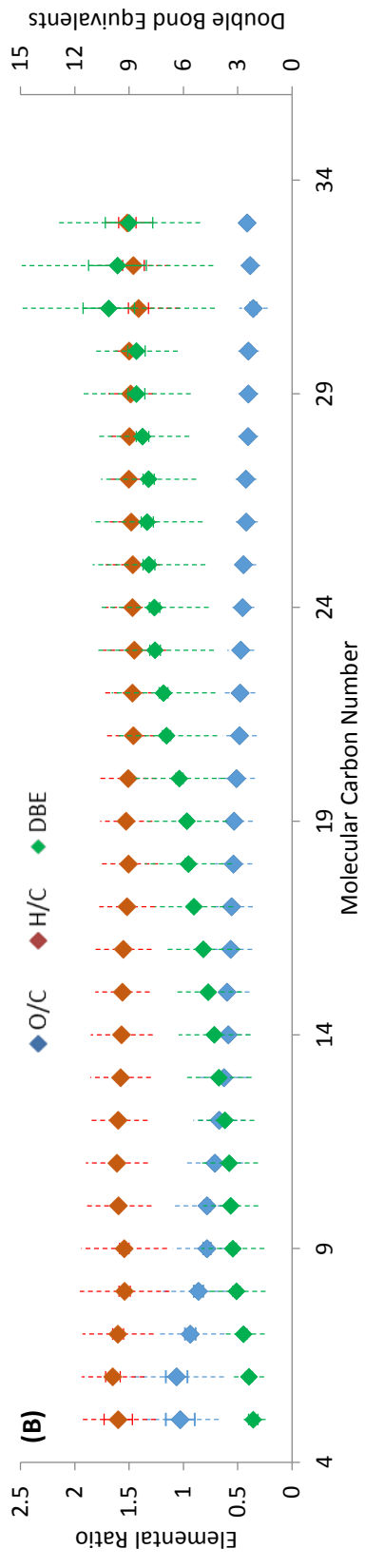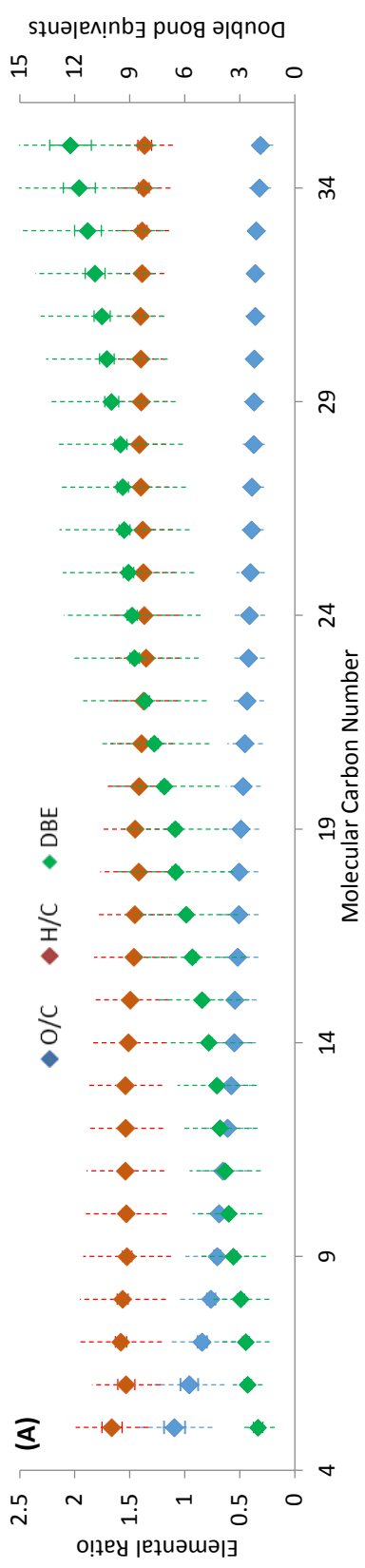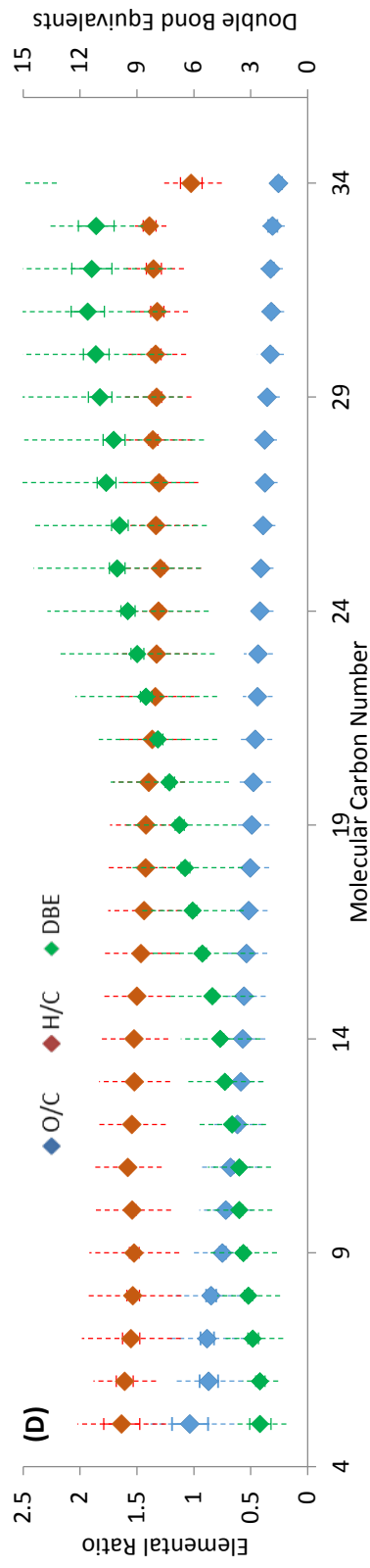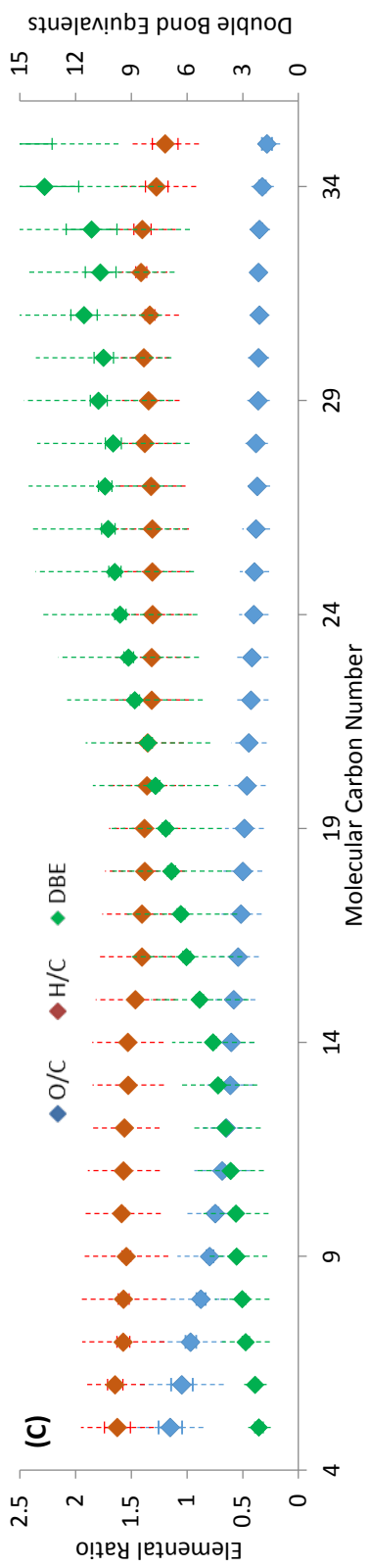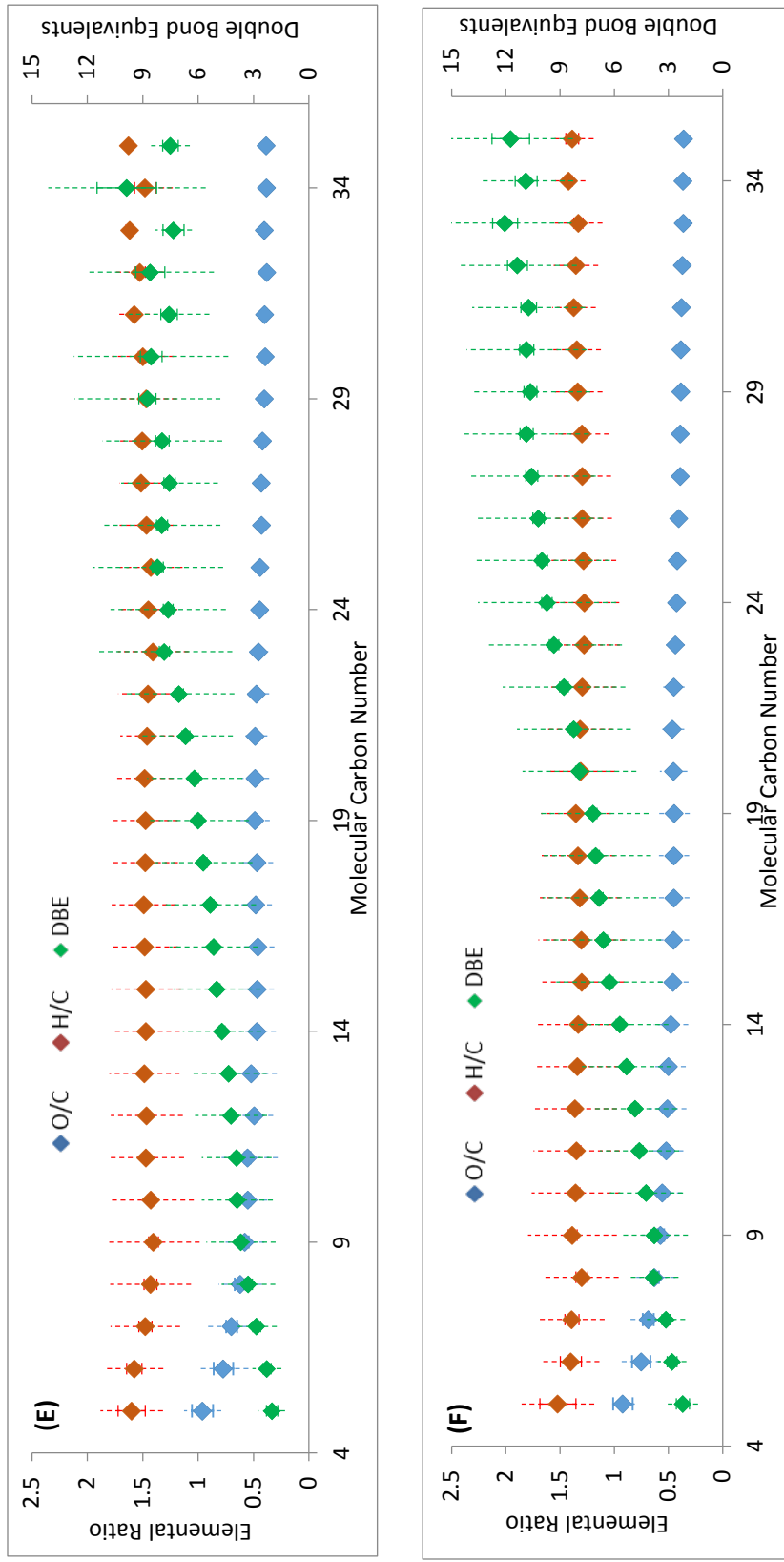
Figure 4.5 Carbon number trends of the elemental ratios and DBE values determined at each carbon number for daily sample 0625 (A), 0626 (B), 0709 (C), 0710 (D), 0719 (E) and 0720 (F). Uncapped dashed line bars equal 1 standard deviation and capped solid line bars equal 1 standard error.
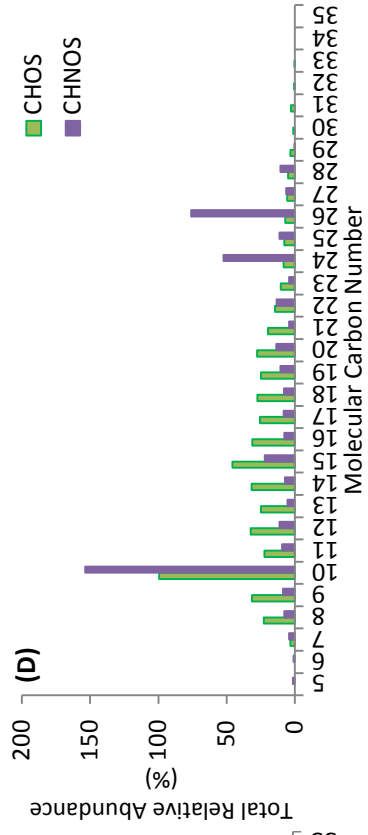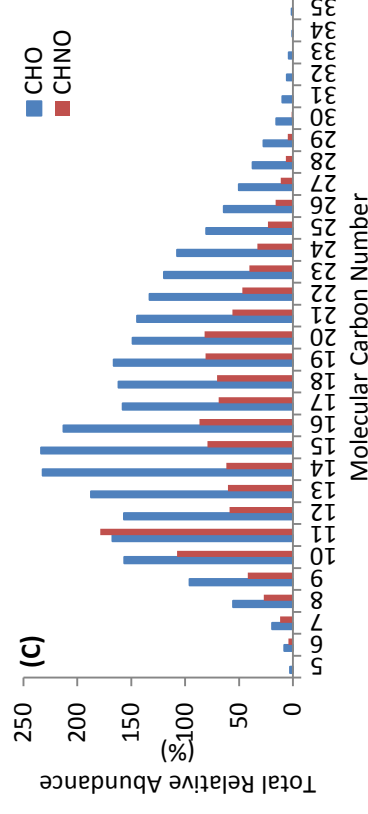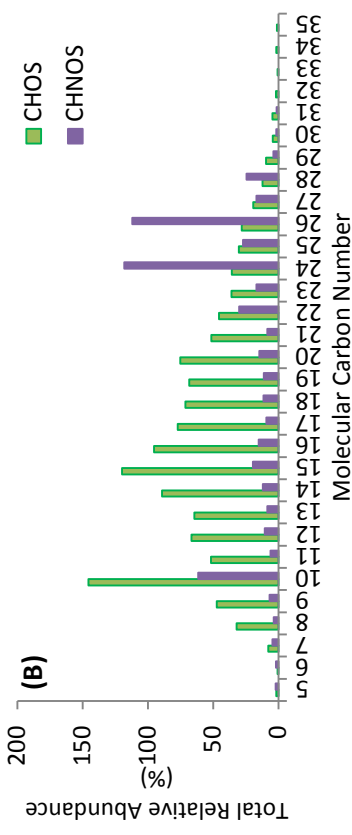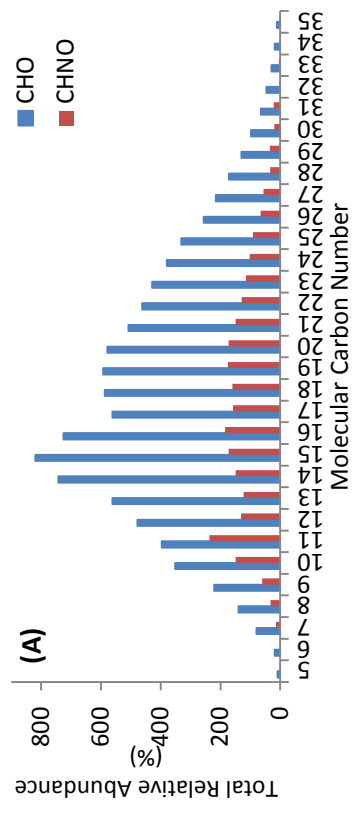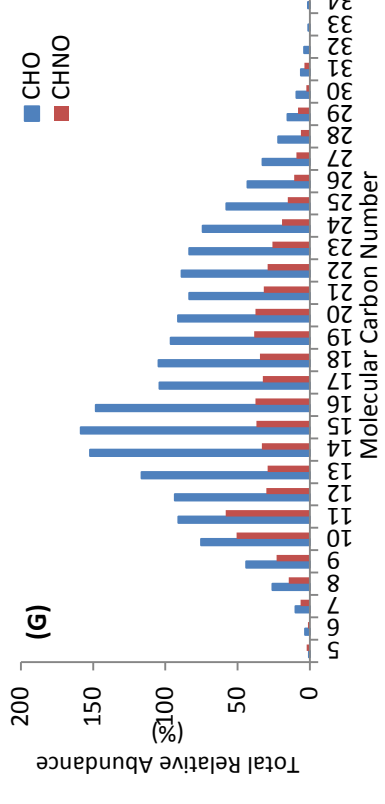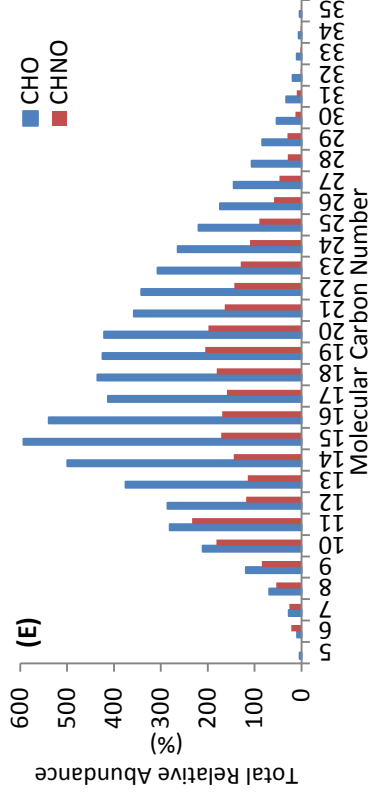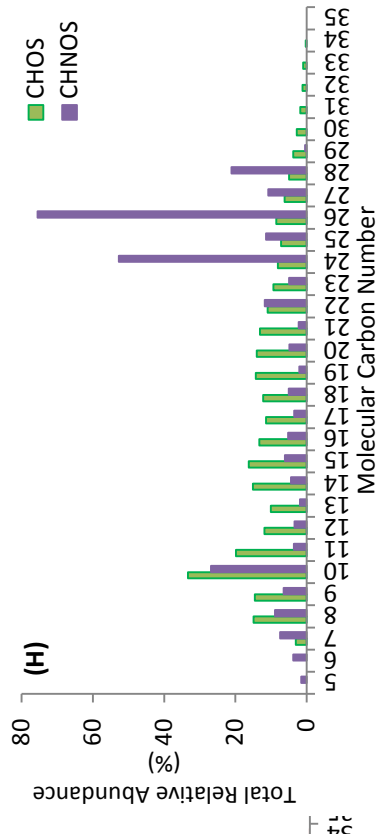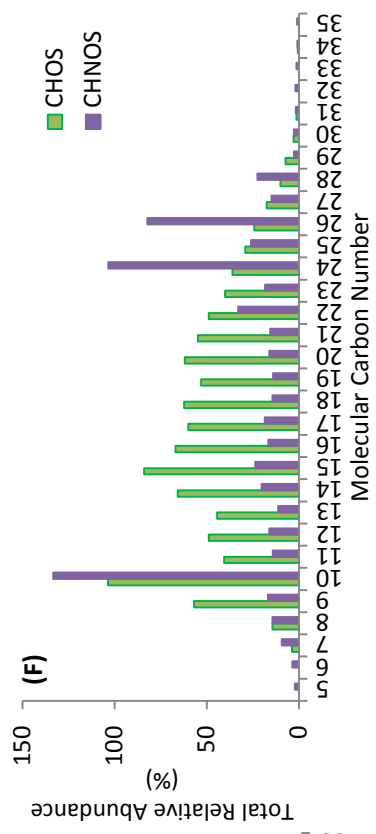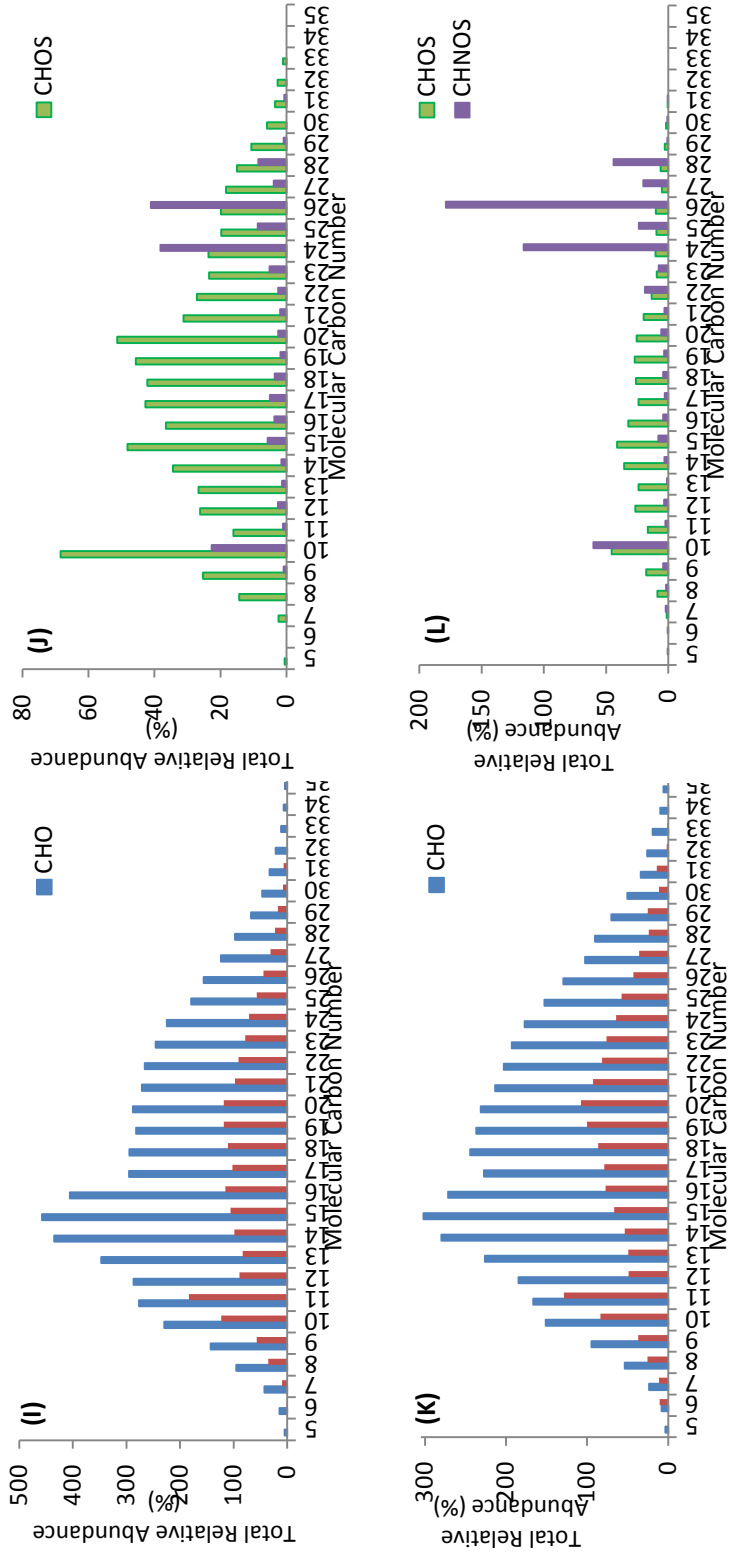
Figure 4.6 Carbon number trends of total relative abundance at each carbon number for WSOC sample 0625 CHO and CHNO compounds (A); sample 0625 CHOS and CHNOS compounds (B); sample 0626 CHO and CHNO compounds (C); sample 0626 CHOS and CHNOS compounds (D); sample 0709 CHO and CHNO compounds (E); sample 0709 CHOS and CHNOS compounds (F); sample 0710 CHO and CHNO compounds (G); sample 0710 CHOS and CHNOS compounds (H); sample 0719 CHO and CHNO compounds (I); sample 0719 CHOS and CHNOS compounds (J); sample 0720 CHO and CHNO compounds (K); sample 0720 CHOS and CHNOS compounds (L).

4.1.5 Numerical comparison of the WSOC samples

Since both the van Krevelen plots (Figure 4.3) and the carbon number trends plot (Figure 4.6) indicate biogenic SOA influences on the aerosol-derived WSOC composition, I compared the molecular formulas in each WSOC sample with the biogenic SOA composition discussed in Chapter 3. Only the CHO compounds were compared because the chamber generated SOA does not include N- or S-containing species. The numbers of common molecular formulas found in both WSOC and biogenic SOA for each WSOC sample are shown in Table 4.1. Generally, more than 80% of the CHO compounds in the WSOC samples were found in the biogenic SOA dataset, which verifies the major biogenic SOA component in aerosol-derived WSOC in this study.

As shown in Figure 4.4, the elemental ratios and DBE values of each WSOC sample are quite similar across the study period. To examine the similarities of their molecular composition, the appearance frequencies of the molecular formulas in the 42 datasets (34 WSOC samples plus 8 replicates) were calculated. Shown in Figure 4.7, among the 8322 unique monoisotopic molecular formulas (with 6 times the root-mean-square of the signal-to-noise ratio) from all of the WSOC samples, 941 (11%) of them appeared in all of the 42 datasets. At the other end of the plot, there are 1907 unique molecular formulas that are only present in one of the datasets. Most of these unique formulas have quite low relative abundances in the samples. If a threshold of 10 times the root-mean-square of the signal-to-noise ratio is applied to the samples for molecular formula assignment, 5960 unique formulas were identified among all the samples. Their appearance frequencies in the 42 datasets are also shown in Figure 4.7. The numbers of molecular formulas in only a few of the samples (in 1 to 5 of the samples) decreased significantly compared to the plot for 8322 formulas. The biogenic SOA components might partially explain the high number frequency of molecular formulas in most of the samples (*e.g.*, 941 formulas in all the samples). The samples were collected continuously over about one month in the summer of 2010. At the Storm Peak Lab, biogenic SOA formation is expected to consistently contribute to the WSOC composition. Hallar et al. (2011) have found persistent daily new particle formation in summer at this area, which makes the daily WSOC samples consisting biogenic SOA components. The environmental conditions did not change dramatically over the study period (Section 4.3), thus the WSOC composition is not expected to significantly change. However, there are still formulas appearing in only a few of the samples. Even for the formulas with large appearance frequencies, they may have variable RAs values among the samples. Thus the

comparison of the molecular composition of the WSOC samples is necessary and multivariate statistical analysis methods are needed to deal with the large datasets and their complexity.
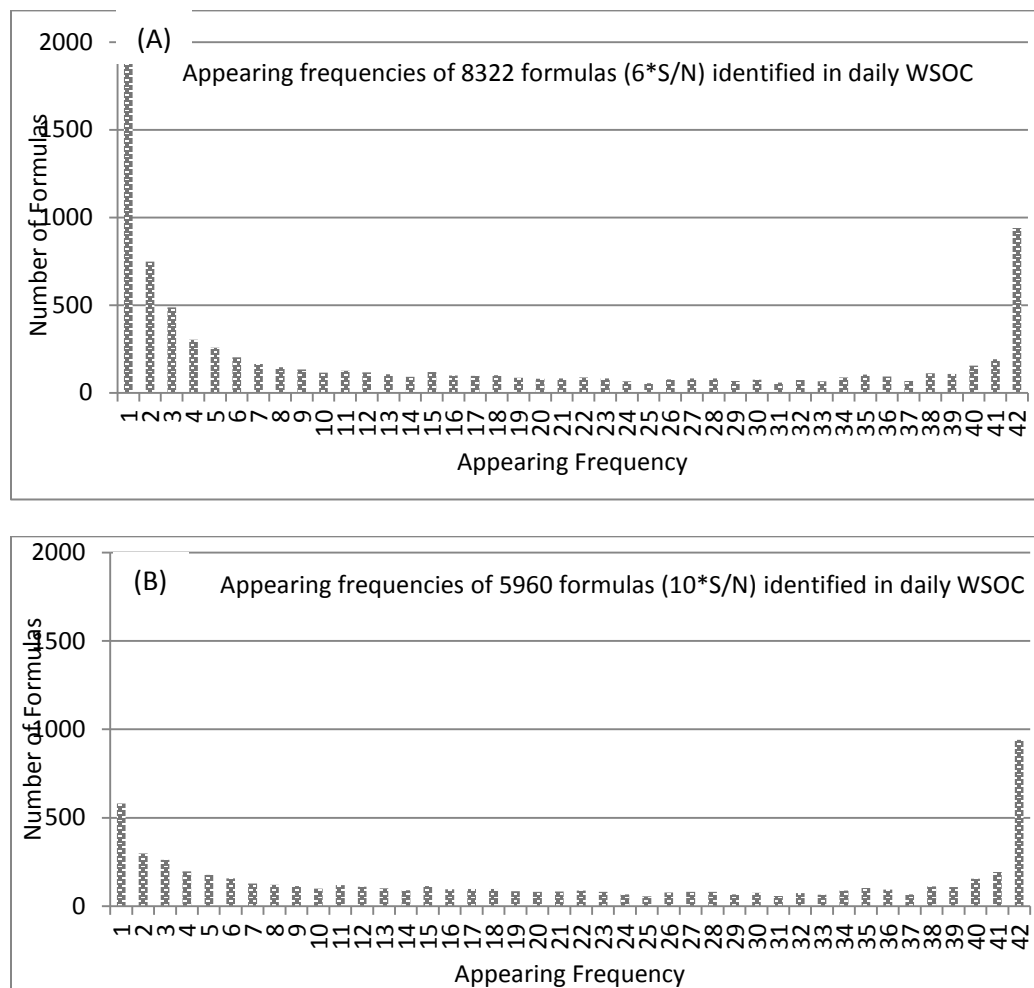
Figure 4.7 The appearing frequencies of the formulas in the 42 WSOC datasets (34 samples plus 8 replicates) using 6 times signal-to-noise ratio (A) and 10 times signal-to-noise ratio (B).

4.2 Multivariate statistical analysis of the WSOC composition

The daily aerosol-derived WSOC samples share a large number of common molecular formulas; they also contain unique molecular formulas, which appear only in a few of the specific samples. Comprehensive comparison of the composition of the daily WSOC samples helps to find the WSOC samples with similar composition. It also is used to examine the characteristic molecular formulas for those similar samples and to explore the atmospheric processes that are responsible for their composition. Since each of the 34 daily WSOC samples and the 8 sample replicates contain ~3000 molecular formulas, multivariate statistical analysis methods are required to conduct the comparison. As discussed in Chapter 3, we used multivariate statistical analysis methods, including: hierarchical cluster analysis (HCA) and principal component analysis (PCA).

To prepare the dataset for multivariate statistical analysis, the processed ultrahigh resolution MS data from each WSOC sample was combined. Duplicate molecular formulas were removed, leaving 8322 distinct formulas, which are regarded as the variables in the statistical analysis. The RAs (or converted values from the RAs) of the formulas in the samples comprise the input of the analysis. If a formula is not present in a sample, its RA is 0 in that sample. The original dataset is composed of 42 observations (34 daily WSOC samples and 8 sample replicate) with 8322 variables (the formulas).

Various data preparation methods were evaluated to best reflect the inherent variation of the WSOC molecular composition across the month and to examine how the environmental conditions affect the composition variation. They are presented in each subsection of Section 4.2. Since usually a few PCs are selected to represent the PCA results which carry a variable portion of the total variance and HCA results carry 100% of the original variance, we used the HCA dendrogram to inspect the data preparation methods. The "Manhattan" distance method is a commonly used distance measurement method in statistics; it was used in the HCA unless another is noted. The HCA tests conducted for the LANLSOA discussed in Chapter 3 show that the complete linkage is the best among the commonly used single, average and complete linkages so it was selected in this study unless another is noted. The groupings of the analysis replicates in HCA dendrograms are the most important criteria to examine the data preparation methods. The analytical replicate pairs are expected to have the most similar molecular

composition, so they should cluster most closely in the dendrograms. In addition, the samples with similar bulk chemical properties (calculated from the ultrahigh resolution MS data) potentially have similar molecular composition. This was used to assist examining the HCA results. For example, as discussed in Section 4.2.5, several WSOC samples with high DBE values usually group closely in the dendrograms and thus show similar molecular composition. Also, the measurements other than ultrahigh resolution MS were used to help examine the multivariate statistical results. For example, the back trajectory analysis showed over several days the air was consistently coming from the same direction. Thus the WSOC samples collected on those days are expected to have similar molecular composition.

Sections 4.2.1 through 4.2.7 will present and discuss the multivariate statistical analysis results with different data preparation methods. The best method found in this study uses scaled (also called "standardized") normalized RA (consistent with the discussion in Chapter 3 for LANLSOA), will be presented in Section 4.2.7. Following that in Section 4.3, the addition of meteorological parameters and other aerosol chemistry are examined together with the ultrahigh resolution MS dataset using the refined multivariate statistical analysis methods.

4.2.1 Use formula presence and absence in the WSOC samples

Of the 8322 distinct molecular formulas identified in the aerosol-derived WSOC samples, around 10% of them appeared in all the samples, while the others were identified in one or more of the samples. In previous studies (Kujawinski et al. (2009) and Altieri et al. (2012)), a presence-absence matrix of variables was used for the multivariate statistical analysis of ultrahigh resolution MS data. I tested the HCA and PCA analyses using only the presence and absence of the formulas in the samples. To conduct the analysis, the original dataset was modified to the following: all the non-zero RAs were converted to 1 and the zero RAs were kept as 0 in the dataset.

The HCA dendrogram using the presence-absence of formulas in the daily WSOC samples is shown in Figure 4.8. The samples 0709, 0625 and 0720, which are the samples with the high DBE values, group together and are separated from the other samples. Sample 0704, which was influenced by a local fireworks event, stands out by itself. However, half of the replicate pairs do not appear to be closely related in the dendrogram, including the pairs 0628, 0707, 0722 and 0727. Multivariate statistical analyses using the presence-absence matrix were also conducted

for the elemental groups, including: CHO species only, CHNO species only, CHOS species only and CHNOS species only. The results were not helpful and so they are not shown here.
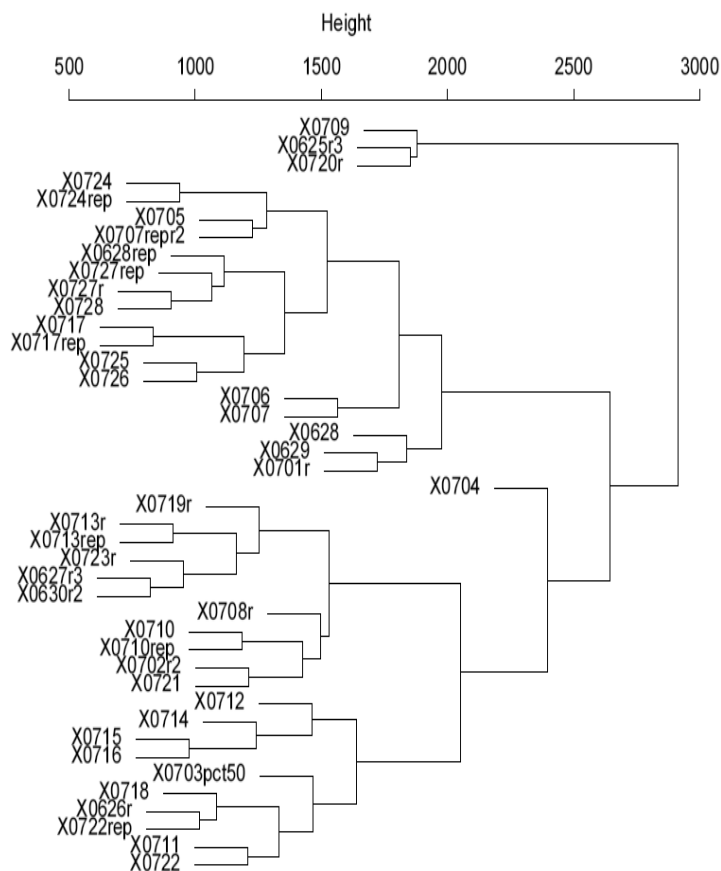
Figure 4.8 The HCA dendrogram using the presence-absence of molecular formulas in the daily WSOC samples, using the Manhattan distance method and complete linkage.

The non-paired sample replicates in the HCA dendrogram using the presence-absence matrix indicate that a fraction of the useful information is lost with the binary variables representing the presence or absence of the species. In this approach, all of the "presence" is regarded as 1 in the matrix, so the relative magnitude of the formulas is not fully considered. All the present formulas were considered equally and all the absent formulas were considered equally in the analysis. It is true that the relative abundances of the formulas do not exactly reflect their amount in the samples, since the complex organic species in the WSOC samples have different electrospray efficiencies in the ion source. But the RAs still represent the relative amount of the compound in a sample to a large extent. Another inconvenience of using the presence-absence method is that the presence or absence of a formula in a sample depends on the RA threshold

selected. In this study, a molecular formula is assigned only if its RA is higher than 6 times the root-mean-square of the signal-to-noise ratio. It is highly possible that a formula assigned in some samples also appears in other samples but its RA is lower than the threshold but it's labeled as 0 for the sample in the presence-absence matrix. Similarly if a threshold of 10 times the root-mean-square of the signal-to-noise ratio is used, the formulas between 6 and 10 times will become "absent". Thus for the molecular formulas near the threshold, subtle changes in the RA can have significant impacts on the inclusion of a variable. However if RAs were used, it can be seen that even though these formulas appear in some samples with 6 times the root-mean-square of the signal-to-noise ratio threshold, their RAs are quite small. This is in contrast to the assignment of the "1"s with equal importance to all observed molecular formulas in the presence-absence matrix. Thus to reduce the effect from the assignment threshold a better option would be to use the RA values in the multivariate statistical analysis so that the formula relative magnitude is considered.

4.2.2 Multivariate statistical analysis with relative abundances and normalized relative abundances

Due to the drawbacks of using only the presence-absence of the formulas in WSOC samples for multivariate statistical analysis, I conducted analyses with molecular formulas relative abundances using the original combined dataset. The HCA dendrogram using the relative abundances of the molecular formulas in the daily WSOC samples with the Manhattan distance method and complete linkage is shown in Figure 4.9. The sample replicates are expected to have quite similar molecular composition, thus closely paired replicates is used as a criterion to examine the data preparation methods for multivariate statistics. Again this method resulted in only 4 of the 8 replicate pairs grouped closely together in the dendrogram. The multivariate statistical analyses using formula relative abundances were also conducted for the CHO species only, CHNO species only, CHOS species only and CHNOS species only. The elemental group results did not provide insight toward atmospheric implications and thus are not shown here.

128
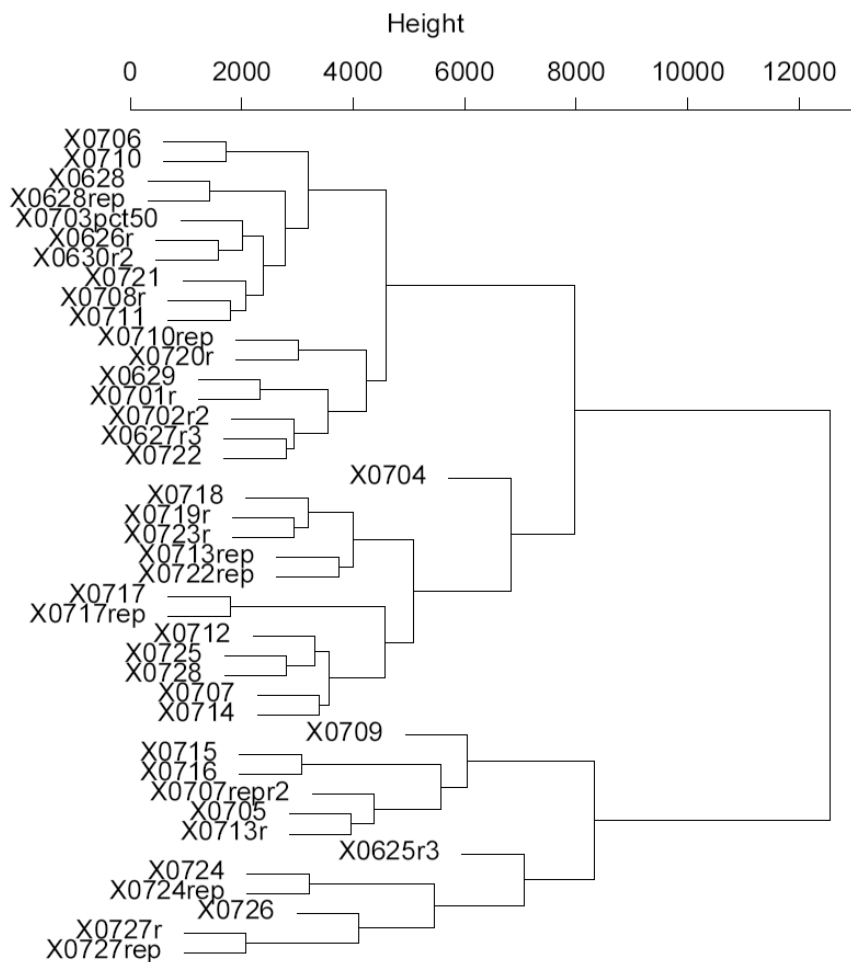
Figure 4.9 The HCA dendrogram using the relative abundances of formulas in daily WSOC samples with the Manhattan distance method and complete linkage.

Why does the HCA using the molecular formula relative abundances not reflect the good reproducibility of the replicates? First, the RAs of the formulas show intensities of the analytes in the mass spectrum relative to the intensity of the base peak (the tallest peak) of that spectrum. The base peak of the mass spectra can vary, leaving the other peaks not comparable between multiple spectra using the measured RA values. Second, even when the base peak is the same for multiple mass spectra, the base peak compound could have different mass concentrations in the samples thus it has different intensities in the spectra. Third, even when the base peak in multiple spectra is the same and has the same intensity, the heights of the other peaks in the spectra depend on their amount in the samples and on their electrospray efficiencies. Fourth, as shown in the reconstructed mass spectra, the WSOC samples are complex mixtures with thousands of formulas detected in every sample. The electrospray

efficiency of a compound is affected by all of the other species in the sample. Therefore, the RA values of an analyte in different samples cannot be compared directly. Some data preparation procedures are needed before the multivariate statistical analysis.

Hur et al. (2009) and later on Wozniak et al. (2014) normalized the relative abundance of individual peaks to the summed relative abundance of each peak list. In this way, the analyte magnitude in a sample spectrum is not normalized to the base peak, but to the total identified ion current. Thus the analyte magnitudes in different samples are comparable. Similarly, I normalized each of the analyte intensities to the sum of intensity for all of the identified analytes within the sample to create a new matrix with the "normalized RAs". The RAs in the original database were replaced with normalized RAs and multivariate statistical analyses were conducted. The HCA dendrogram using normalized relative abundances is shown in Figure 4.10. All the replicate pairs group closely, which partially confirms the data preparation method of normalization. The HCA and PCA analyses with normalized RAs show much improved characterization of the daily WSOC composition, which will be discussed in detail in Section 4.2.5.
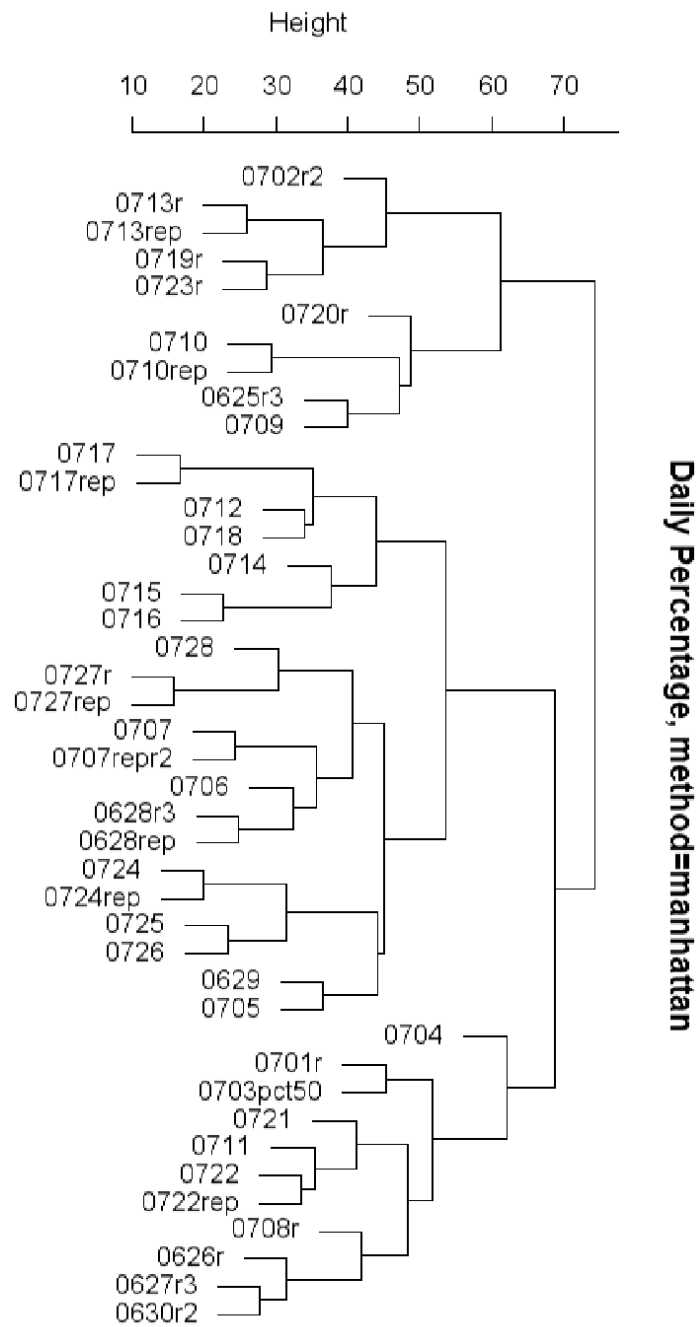
Figure 4.10 HCA dendrogram using normalized relative abundances and the Manhattan distance method and complete linkage.

4.2.3 Multivariate statistical analysis on the sample bulk properties

The sample bulk properties were calculated for each of the sample based on their molecular formulas, such as the relative abundance weighted O/C, H/C and DBE values as discussed in Section 4.1. Sleighter et al. (2010) conducted HCA and PCA analyses using 6 "magnitude-weighted" bulk parameters of dissolved organic matter samples analyzed by ultrahigh resolution MS, including: the magnitude-weighted DBE, carbon number, H/C, O/C, DBE/C and DBE/O ratios. Thus, I tested the multivariate statistical analysis on the daily WSOC samples in this study using the same sample bulk properties. Similar to the calculations for H/C_w, O/C_w and DBE_w described in Chapter 2, the relative abundance weighted DBE/C ratio, DBE/O ratio and carbon number of a WSOC sample are determined using the following equations:

$$DBE/C_w = \Sigma(w_i DBE_i) / \Sigma(w_i c_i) \qquad \text{4-1}$$

$$DBE/O_w = \Sigma(w_i DBE_i) / \Sigma(w_i o_i) \qquad \text{4-2}$$

$$C\#_w = \Sigma(w_i c_i) / \Sigma w_i \qquad \text{4-3}$$

for elemental compositions of $C_c H_h N_n O_o S_s$, where w_i is the relative abundance of each individual molecular formula, i. The relative abundance weighted bulk properties of the daily WSOC samples used in HCA and PCA analyses are shown in Table 4.2.

Table 4.2 Relative abundance weighted bulk properties of the daily WSOC samples.

Sample	H/C$_w$	O/C$_w$	DBE$_w$	DBE/C$_w$	DBE/O$_w$	C#$_w$
0625	1.4987	0.4764	5.6085	0.3139	0.6591	17.8642
0626	1.5627	0.5230	4.8096	0.2903	0.5551	16.5658
0627	1.5714	0.5187	4.7533	0.2850	0.5496	16.6762
0628	1.5614	0.5284	4.9479	0.2886	0.5461	17.1473
0628rep	1.5607	0.5264	4.9039	0.2884	0.5479	17.0018
0629	1.5576	0.5194	4.9594	0.2918	0.5618	16.9978
0630	1.5600	0.5130	4.8352	0.2915	0.5683	16.5850
0701	1.5544	0.5460	4.9564	0.2956	0.5414	16.7683
0702	1.5057	0.5156	5.2068	0.3149	0.6108	16.5337
0703	1.5572	0.5183	4.8797	0.2939	0.5669	16.6061
0704	1.5429	0.5322	5.1005	0.3004	0.5645	16.9770
0705	1.5714	0.5018	4.9249	0.2802	0.5584	17.5772
0706	1.5585	0.5302	4.9143	0.2901	0.5471	16.9406
0707	1.5368	0.5169	5.1512	0.3005	0.5813	17.1438
0707rep	1.5424	0.5087	5.1521	0.2953	0.5805	17.4462
0708	1.5255	0.5032	5.2569	0.3068	0.6097	17.1341
0709	1.4739	0.5036	5.8408	0.3295	0.6544	17.7237
0710	1.4977	0.4912	5.5489	0.3178	0.6471	17.4590
0710rep	1.5277	0.4847	5.3491	0.3011	0.6213	17.7629
0711	1.5431	0.5315	5.1050	0.3001	0.5647	17.0093
0712	1.5420	0.5107	5.3501	0.2941	0.5759	18.1921
0713	1.5458	0.4928	5.2163	0.2901	0.5887	17.9802
0713rep	1.5488	0.4798	5.0019	0.2897	0.6038	17.2680
0714	1.5764	0.5454	4.8541	0.2857	0.5238	16.9920
0715	1.5684	0.5131	5.0763	0.2832	0.5518	17.9277
0716	1.5724	0.5107	5.0666	0.2797	0.5477	18.1131
0717	1.5454	0.5281	5.3945	0.2936	0.5559	18.3752
0717rep	1.5433	0.5246	5.4009	0.2941	0.5607	18.3624
0718	1.5417	0.5247	5.2437	0.2953	0.5627	17.7594
0719	1.5391	0.5060	5.1816	0.2952	0.5833	17.5553
0720	1.4629	0.4891	6.1251	0.3317	0.6783	18.4636
0721	1.5223	0.5231	5.2808	0.3091	0.5909	17.0854
0722	1.5158	0.5437	5.3265	0.3160	0.5813	16.8535
0722rep	1.5291	0.5173	5.2158	0.3049	0.5895	17.1048
0723	1.5448	0.5095	5.1711	0.2922	0.5735	17.6971
0724	1.5743	0.4987	5.0662	0.2746	0.5506	18.4498
0724rep	1.5698	0.4922	5.0541	0.2774	0.5636	18.2189
0725	1.5788	0.5205	4.9561	0.2762	0.5307	17.9428
0726	1.5604	0.5160	5.1425	0.2843	0.5510	18.0860
0727	1.5467	0.5278	5.1546	0.2932	0.5555	17.5831
0727rep	1.5475	0.5272	5.1348	0.2928	0.5555	17.5356
0728	1.5569	0.5380	4.9997	0.2904	0.5398	17.2162

The hierarchical cluster analysis (Figure 4.11) shows the separation of the WSOC samples with high DBE values (0625, 0709 and 0720) from the other samples. But the sample replicate pairs still do not group closely in the dendrogram using the sample bulk properties. Although the molecular formulas in a sample have a wide range of chemical properties (*e.g.*, O/C of formulas in a sample might range from 0.3 to 2.0), these chemical properties are averaged when the bulk properties are calculated. Overall, the calculated bulk properties are quite similar from sample to sample. Consider the O/C ratio again, the O/C_w of the samples are in among the small range of 1.5 to 1.6. Thus, the bulk properties of the samples such as O/C_w are so close to each other that in some cases the differences are even smaller than the error of the replicates. So, in the dendrogram some samples are grouped closely with other samples instead of their replicates.

(A)

Height

135
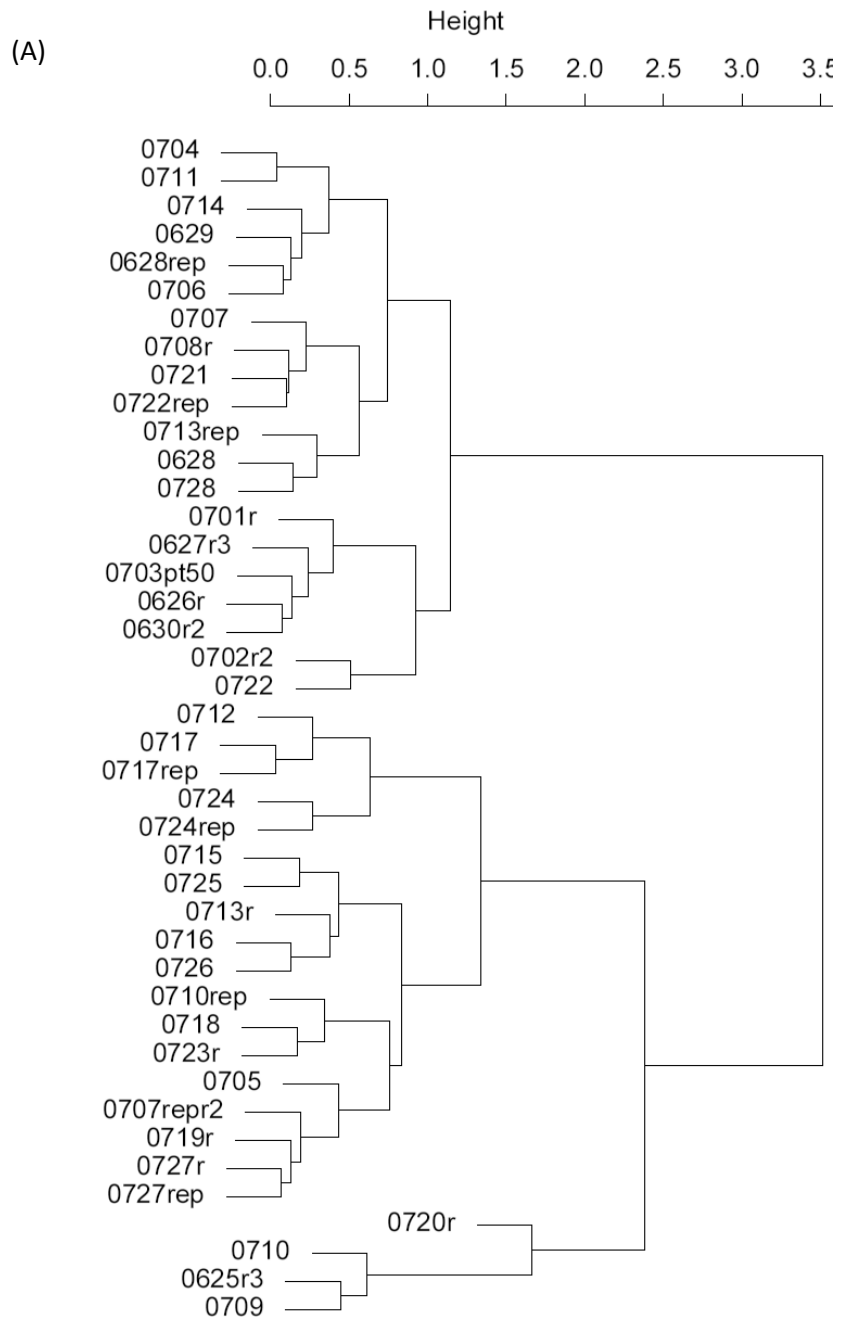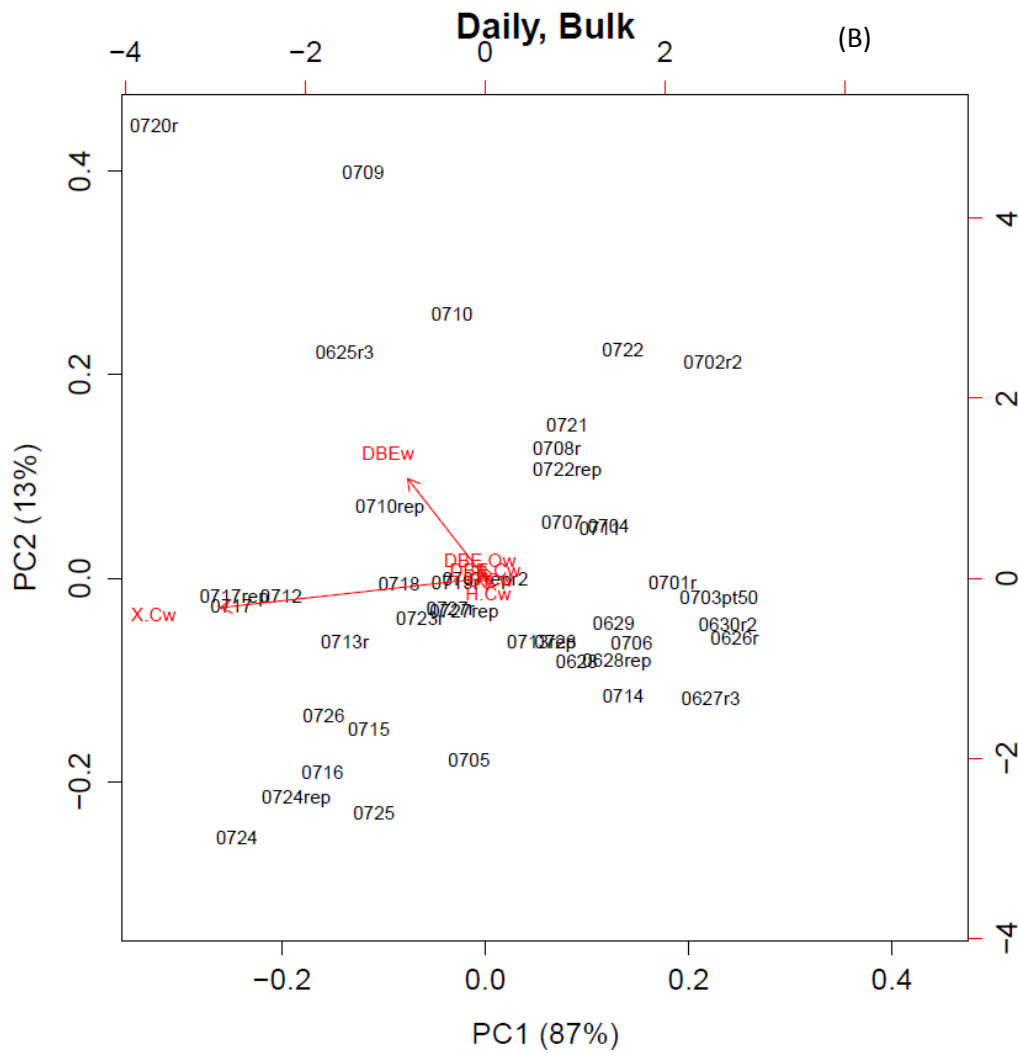
Figure 4.11 HCA dendrogram with Manhattan distance method and complete linkage (A) and PCA biplot (B) using the bulk properties of the daily WSOC samples.

The results of the principal components analysis are shown in Figure 4.11. Consistent with the primary goal of PCA, the first two PCs represent most of the variables (PC1 = 87% and PC2 = 13%). As shown in Figure 4.11, the DBE_w and $C\#_w$ drive the PCA result. This is because the values of DBE_w and $C\#_w$ are much larger than the other four bulk properties and their variances are larger too. They play a much more important role in the multivariate statistical analysis than the other variables, although the 6 bulk properties should show comparable significance in the analysis. Thus data preparation or transformation is needed when the variables are not comparable in magnitude or in different units. This will be discussed in Section 4.2.7.

4.2.4 Dimension reduction for the combined WSOC composition dataset

4.2.4.1 Dimension reduction by elevating the molecular formula assignment threshold

The combined WSOC composition dataset contains 42 observations with 8322 variables from the distinct molecular formulas. The high number of variables presents a statistical challenge, because not all of the measured variables are "important" for the analysis. The multivariate statistical analysis results will show clearer distinctions of the WSOC composition if the variables that contribute little to the analysis can be eliminated. In an attempt to reduce the variables, the compounds with low RA and likely low concentrations in the samples are considered to be less important than the higher RA compounds. Thus, I increased the formula assignment threshold. Instead of using a threshold equal to 6 times the root-mean-square of the signal-to-noise ratio (RMS of S/N), I applied factors of 10, 15 and 20 times the RMS of S/N and the new thresholds were calculated for every sample. The molecular formulas with RAs less than 10 times the RMS of S/N were removed from the samples, leaving 5960 distinct molecular formulas in the combined dataset. Similarly when 15 and 20 times the RMS of S/N were applied, the retained numbers of distinct formulas were 4437 and 3502, respectively. Normalization of the molecular formula RAs is necessary for multivariate statistical analysis as discussed in Section 4.2.2. So for the analysis with 10 times the RMS of S/N, I re-normalized the RAs in the dataset with the new sample base; similarly for 15 and 20 times the RMS of S/N datasets.

The HCA dendrograms resulting from using the new sample matrixes after the thresholds of 6, 10, 15 and 20 times the RMS of S/N ratios are shown in Figure 4.12. The Manhattan distance method and complete linkage were applied in each case. The HCA results with 6 times the RMS of S/N uses the most formulas, which could carry the most useful information and show the

137

inherent characterizations of WSOC composition, so it is used in comparisons to select the threshold. The HCA dendrograms using 6 and 10 times the RMS of S/N are almost the same, except for the little grouping difference of samples 0629 and 0705. The dendrogram from the HCA using 15 times RMS of S/N is much different from the one with 6 times RMS of S/N. Thus, the formula assignment threshold of 10 times the RMS of S/N is used in the subsequent multivariate statistical analysis.

(A)

Daily Percentage, method=manhattan

(B)

DailyAerosol10StoN, method=manhattan

(C)

Height

DailyAerosol15StoN, method=manhattan
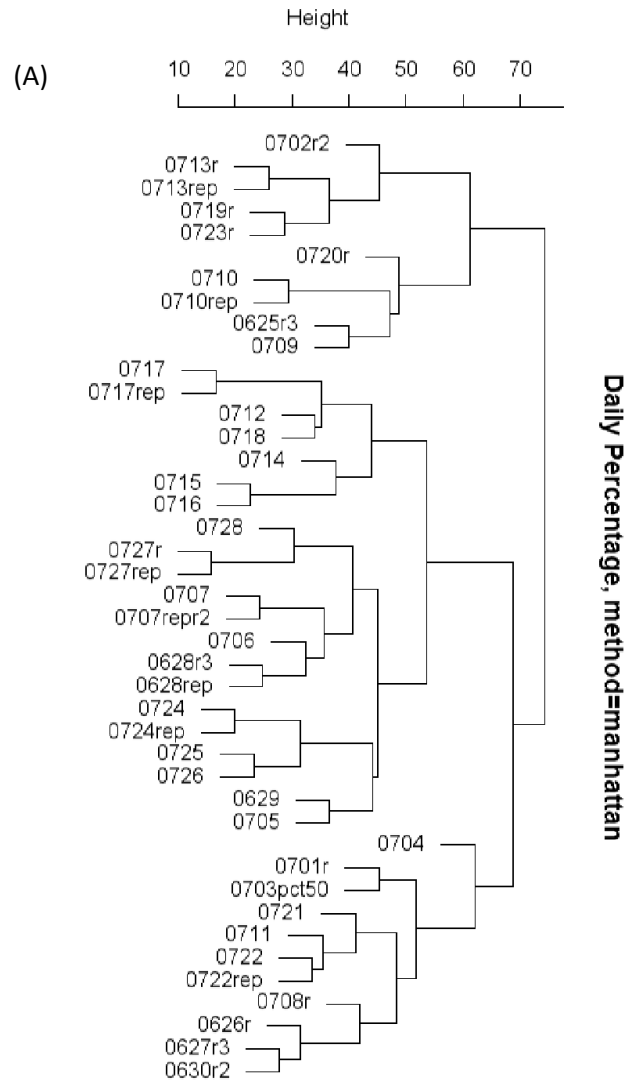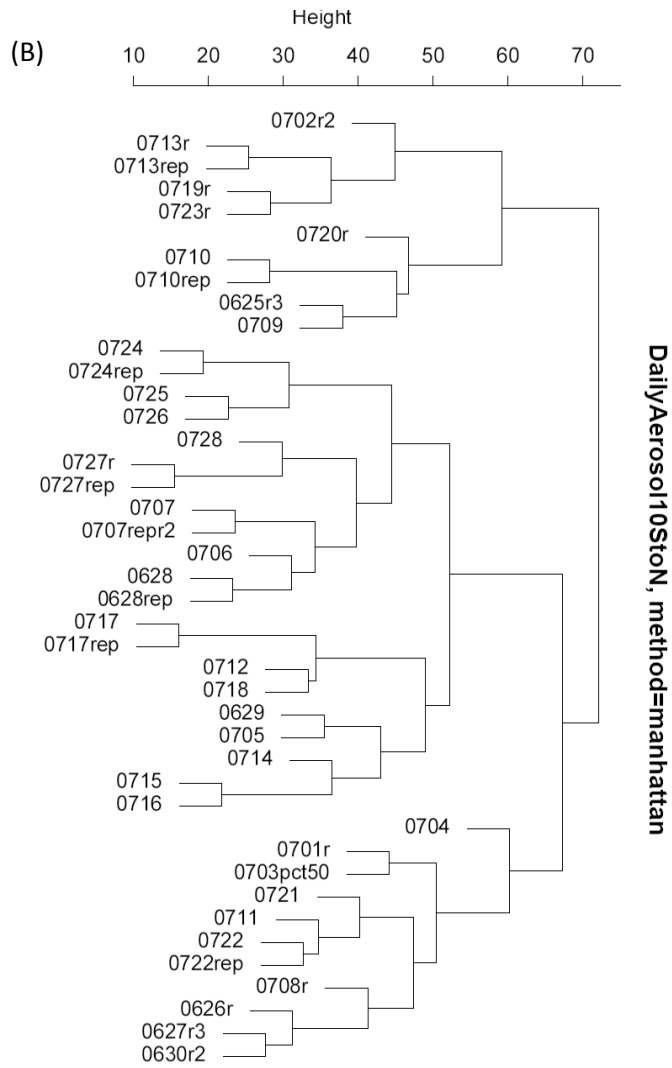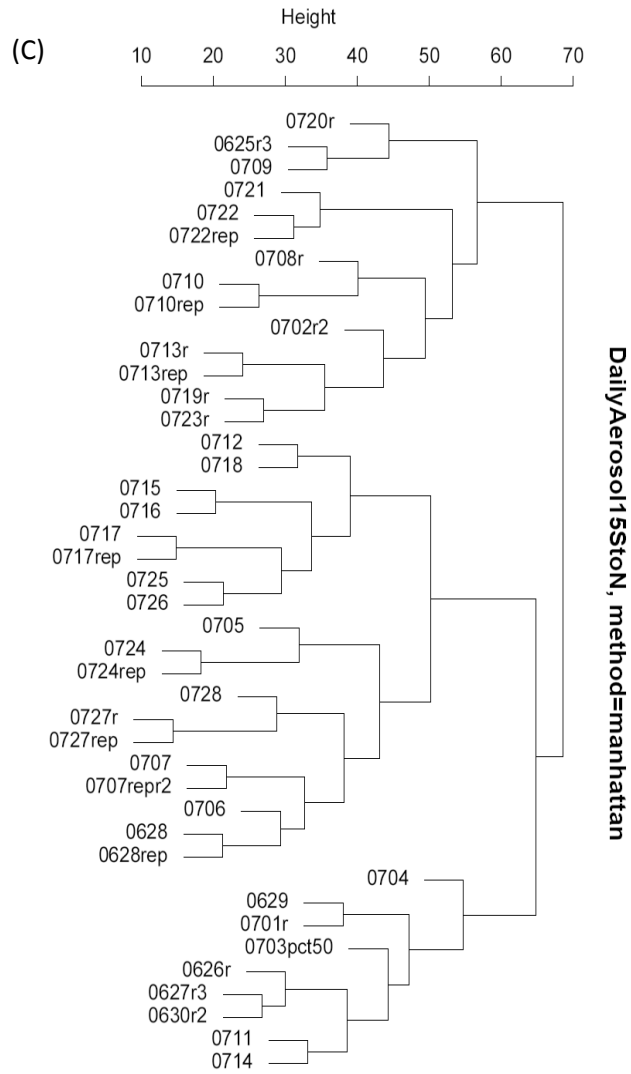
Figure 4.12 The HCA dendrograms of the daily WSOC samples with normalized RAs and using the Manhattan distance method and complete linkage. The molecular formulas were assigned with thresholds at 6 (A), 10 (B), 15 (C) and 20 (D) times the root-mean-square of signal-to-noise ratio, respectively.

Using 6 times the RMS of S/N, the 1st PC in the PCA represents 28% of the variance and the 2nd PC represents 24% of the original variance, as shown in Figure 4.14. The variances represented by the first two PCs with 10, 15 and 20 times RMS of S/N are quite similar to that. The scree plot shown in Figure 4.13 shows the variances represented by the PCs (up to 42) when different thresholds were applied for the formulas assignment. With the increase of the molecular formula assignment threshold the dataset dimensions were reduced and the first two PCs carry more and more of the dataset's original variance. But the differences are minor and are hardly noticeable in Figure 4.13. Thus, the thresholds greater than 10 do not assign much more variance to PC1 and PC2. The PCA biplots using 6 and 10 times the RMS of S/N are almost the same as shown in Figure 4.14. Considering the results of both HCA and PCA using different formulas assignment thresholds for the combined dataset dimension reduction, 10 times the RMS of S/N was selected for the subsequent analysis. This retains the characteristics of using 6 times the RMS of S/N and reduces the number of variables (from 8322 to 5960).
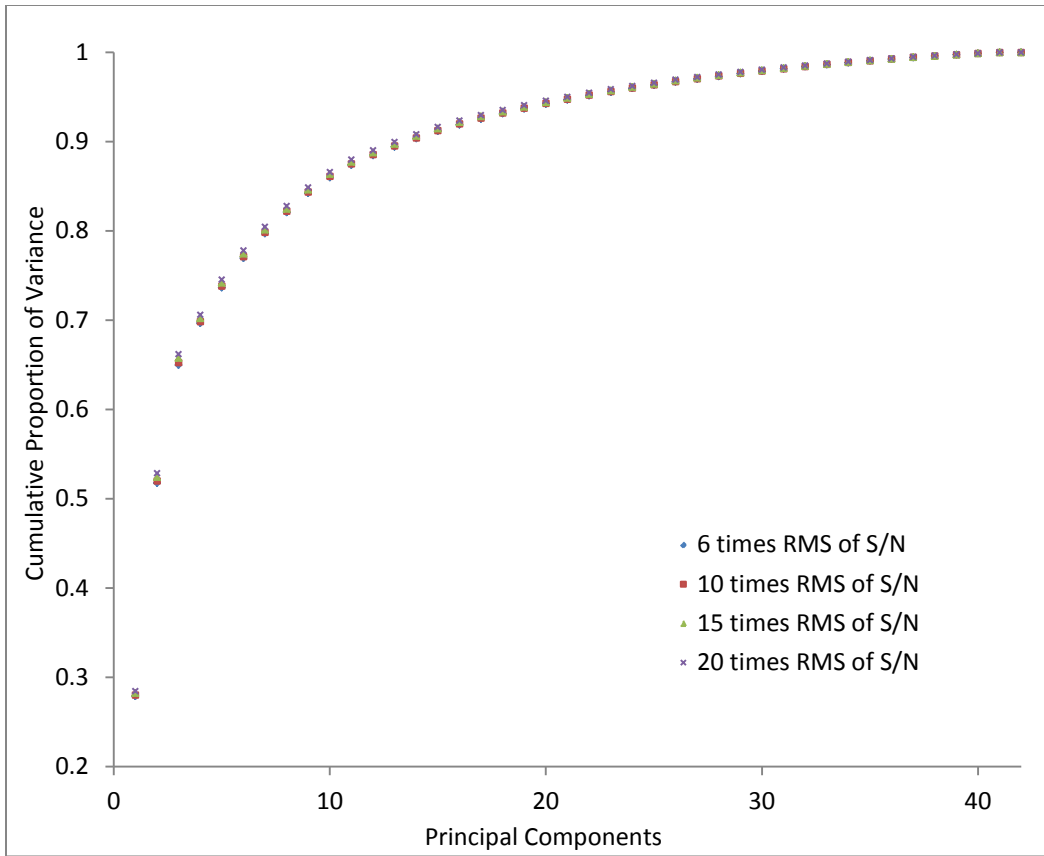
Figure 4.13 The cumulative proportion of the variance carried by the PCs in PCA using different formula assignment thresholds.

Figure 4.14 The PCA biplots of daily WSOC samples with normalized RAs. The molecular formulas were assigned with 6 (A) and 10 (B) times root-mean-square of signal-to-noise ratio as threshold.

4.2.4.2 Dimension reduction using other methods

In addition to increasing the molecular formula assignment threshold, other dataset dimension reduction methods were attempted, as discussed in next three paragraphs. Since the dendrogram conducted with 6 times RMS of S/N uses the most variables and carries the most molecular composition information, it was used for evaluating the other dimension reduction methods. Some of the statistical analysis results were less promising, as evaluated by: the pairing/nonpairing of replicates in the HCA dendrogram; or the similarities of the WSOC composition reflected by the dendrogram compared to those from the non-reduced dataset results. Some of the multivariate statistical analysis results were similar to those from using 10 times the RMS of S/N, which confirmed the formula assignment threshold method. In future research, other dimension reduction methods might be adapted to improve the representation of the molecular composition variation.

The variables far away from the origin in the PCA loading plot contribute more than the variables near the origin to PCA. Thus, as an attempt to reduce the dataset dimensions, some of the variables near the origin were removed based on the PCA loading plot using 6 times the RMS of S/N with normalized RAs. First, the molecular formulas with PCA loadings |PC1|<0.001 and |PC2|<0.001 and |PC3|<0.001 were removed, leaving 4788 distinct formulas. The HCA dendrogram is quite different from the one with 6 times the RMS of S/N. Then, a less strict boundary was applied and the molecular formulas with loadings |PC1|<0.0005 and |PC2|<0.0005 and |PC3|<0.0005 were removed, leaving 6263 distinct formulas. The HCA dendrogram is exactly the same as the results using 10 times the RMS of S/N. Since the molecular formulas near the origin are usually those with low RAs, this dimension reduction method is similar to increasing the formula assignment threshold. But the removal of formulas in this case is not on an individual sample basis.

Inspired by Sleighter et al. (2010), 500 molecular formulas with the highest RAs were selected from each sample representing approximately 15% to 25% of the total number of formulas in the WSOC samples. The combined dataset yields a total of 1031 distinct molecular formulas. The RAs were renormalized on an individual sample basis for HCA and PCA. The results have the replicates grouped closely in the HCA dendrogram but the groupings of the samples are quite

different from those using 6 times the RMS of S/N. This indicates that the dataset with 1031 molecular formulas are too few to accurately characterize the WSOC composition.

Since there are 941 molecular formulas that appear in all the samples, they were eliminated from the dataset with normalized RAs. This was done to test the reduction of sample similarities. However, the HCA and PCA results were not as expected regarding the grouping of replicates and so on. This implies that although the 941 show up in all the samples, their magnitudes, or normalized RAs are varied. And these differences play a role in differentiating the WSOC sample compositions.

4.2.5 HCA and PCA with dimension reduced normalized RAs

From the discussion in 4.2.1 to 4.2.4, characterization of the molecular composition of the daily-WSOC molecular composition requires individual molecular formulas and their normalized relative abundances and improvements were made when the dataset dimension was reduced to 5960 variables using 10 times the RMS of S/N. A more in-depth evaluation of the multivariate statistical analysis results is discussed below.

In the HCA dendrogram in Figure 4.15, 4 groups of samples based on the big clusters in the dendogram are highlighted with colors. The group of samples highlighted in red is separated from the other 3 groups of samples. The samples highlighted in green are separated from the samples highlighted with blue and yellow, which are most closely related. The groups are labeled as Group 1 - 4 samples highlighted in red, blue, yellow and green, respectively,

In the dendrogram, all of the 8 replicate pairs are grouped closely together as expected (in green boxes in the dendrogram), which partially confirms the statistical analysis and data preparation methods. The WSOC samples with the high DBE values as discussed in Section 4.1 (0625, 0709, 0710 and 0720) are shown to have quite similar molecular compositions in the dendrogram (the samples in the black box). In addition, some of the WSOC samples collected on the days with a consistent air mass origin also cluster together as shown in the red boxes in the dendrogram. Hallar et al. (2013) examined the air mass origins during the study period using NOAA Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT) back trajectory analysis. They found during the days from 0624 to 0628, the air masses comes from the west 72-hour prior to the SPL; from 0713 to 0720, air arrived at the SPL from the west with origins in California and Nevada 72 hours previously; from 0727 to 0728, the wind direction was from the south or

southwest. The origin of air masses is reflected in the HCA dendrogram (Figure 4.15). Although not all the samples defined to be from consistent air mass trajectories are clustered together, the origin did affect the WSOC molecular composition as shown in the red boxes in the dendrogram.

The PCA score plot is shown in Figure 4.16 with the samples highlighted with the corresponding colors as in Figure 4.15. The first two PCs in the PCA represent more than 50% of the variance. Notice the consistency between HCA and PCA. The samples of the same Group in the dendrogram (Figure 4.15) are also located near each other on the PCA plot (Figure 4.16). Consider the samples of Groups 2, 3 and 4, the PC2 separates Group 4 samples from Groups 2 and 3, while PC1 further separates the samples of Groups 2 and 3. The Group 1 samples have negative PC1 loadings but are not separated clearly from the other samples. This is probably caused by the partial variance (52%) carried by PC1 and PC2. I tried to include PC3 in the PCA plots, which did not provide much improvement. Thus more PCs might be necessary though are hard to be shown visually in plots. Group 1 samples might have different chemical properties, for examples, 0710 and 0720 are two of the samples with high DBE. Thus the Group 1 samples based on HCA did not group closely in the two dimensional PCA plot which only partial variance (52% instead of 100%) were used. Though the Group 1 samples are not separated clearly from the other samples, HCA and PCA still show substantial consistency regarding the grouping of the samples, which confirms the multivariate statistical analyses on the aerosol-derived WSOC molecular composition.
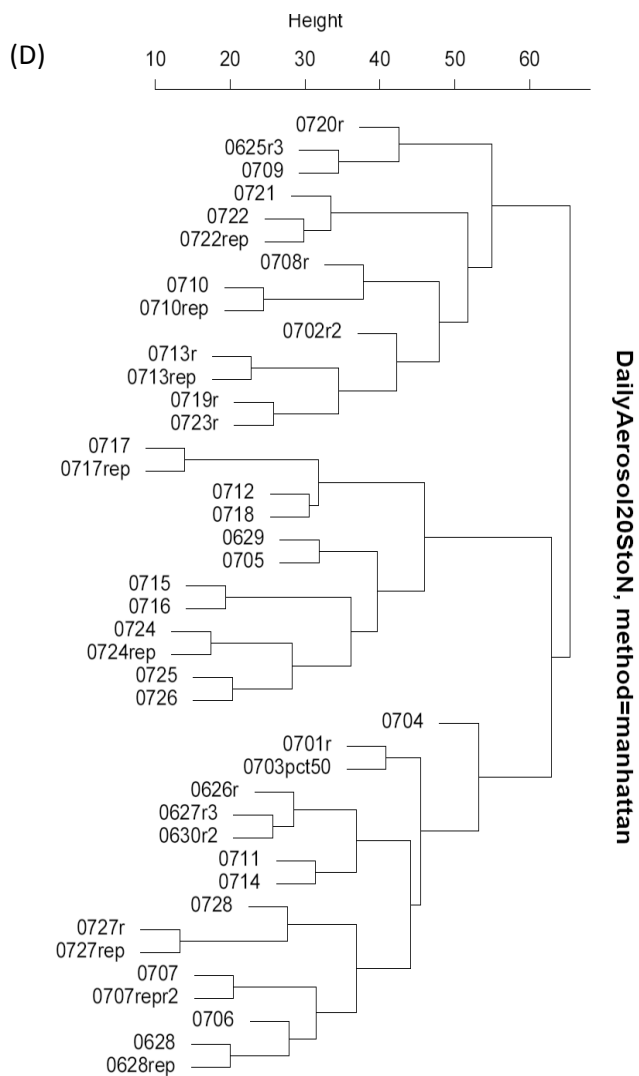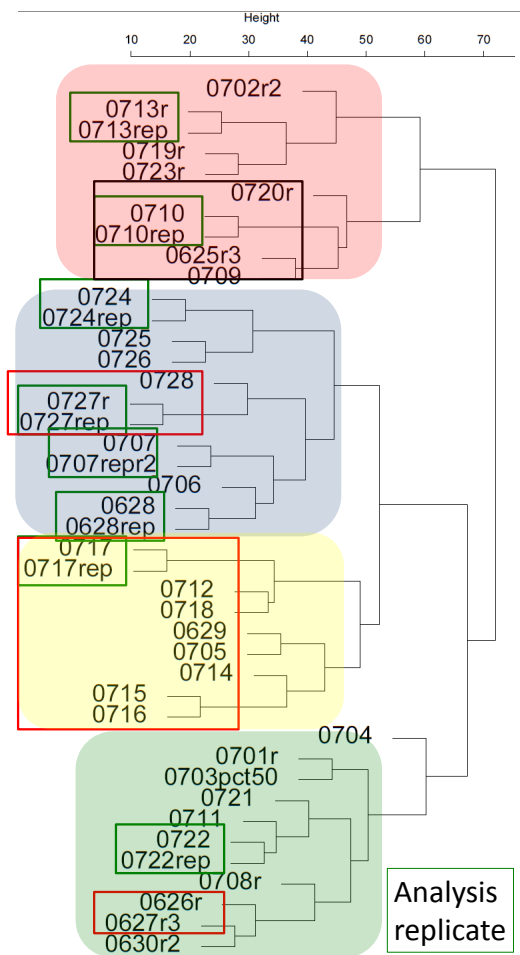
Figure 4.15 The HCA dendrogram of the daily WSOC samples with normalized RAs using the Manhattan distance method and complete linkage. The molecular formulas were assigned with a threshold of 10 times the root-mean-square of the signal-to-noise ratio. The green boxes represent samples with their analytical replicates; the samples in the black box have high DBE values; the dates in red boxes have a similar origin. The samples highlighted in red, blue, yellow and green are defined as Group 1, 2, 3 and 4 samples, respectively.
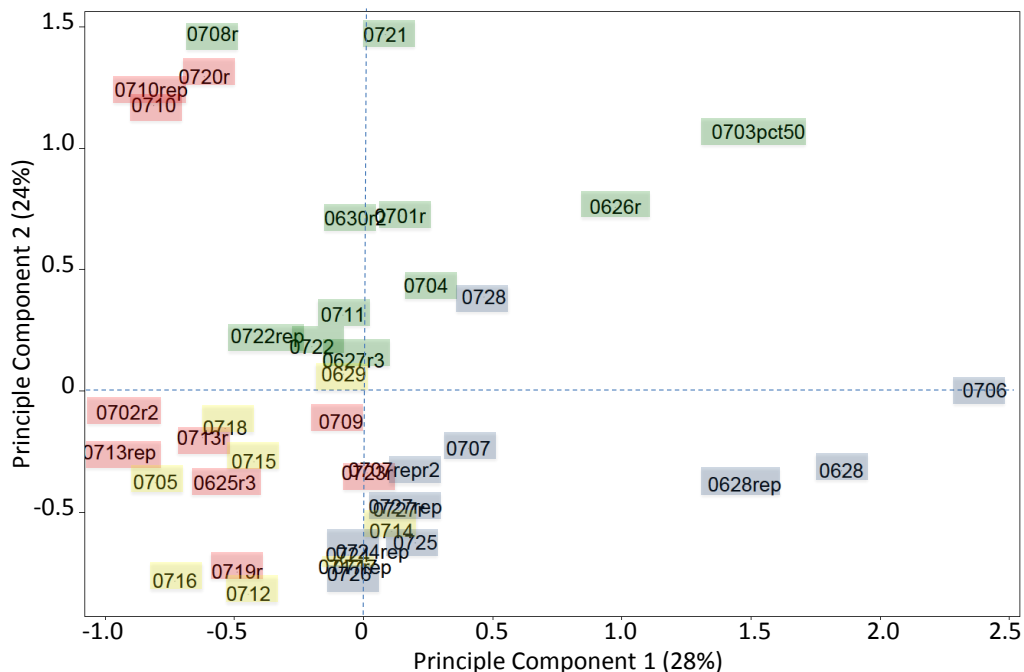
Figure 4.16 The PCA score plot of daily WSOC samples with normalized RAs. The molecular formulas were assigned with a threshold of 10 times the root-mean-square of the signal-to-noise ratio. The colors correspond to the colors in the dendrogram shown in Figure 4.15. The samples highlighted in red, blue, yellow and green are defined as Group 1, 2, 3 and 4 samples, respectively.

The distribution of molecular formulas is shown in Figure 4.17 with their PC1 and PC2 loadings. The molecular formulas characteristic to the WSOC samples located in each quadrant were examined. Since the variables farther away from the origin have larger PC absolute values, they contribute more to the analysis than the ones closer to the origin. The top 10 molecular formulas with the largest distances from the origin in each quadrant are listed in Table 4.3. Consider the samples from Groups 2 - 4, the Group 4 samples (green) are rich in formulas with N and S. Shown in Table 4.3, many of the quadrant 1 molecular formulas have large carbon atom numbers, while most of the quadrant 2 formulas have ~ 10 carbon atoms. The quadrant 3 formulas are mostly CHO compounds with ~15 or 25 carbon atoms. The molecular formulas in quadrants 2 and 3 are likely SOA derivatives, where those of quadrant 2 are monoterpene monomers and those of quadrant 3 are sesquiterpene SOA derivatives or dimers of isoprene

and monoterpene SOA compounds. Quadrant 4 formulas are mostly CHO compounds too, but with more carbon atoms (more than 20) than quadrant 3 formulas.

Figure 4.17 The PCA biplot (A) and zoomed-in loading plot (B) of daily WSOC samples with normalized RAs. The molecular formulas were assigned with a threshold of 10 times the root-mean-square of the signal-to-noise ratio.

Table 4.3 The 10 molecular formulas with the largest distances from the origin in each quadrant of the PCA loading plot as shown in Figure 4.17.

I) + PC1, + PC2	II) - PC1, + PC2	III) - PC1, - PC2	IV) + PC1, - PC2
$C_{10}H_{17}NO_{10}S$	$C_{24}H_{43}NO_{10}S$	$C_{23}H_{34}O_{12}$	$C_{24}H_{38}O_{13}$
$C_{11}H_{19}NO_9$	$C_{26}H_{45}NO_{10}S$	$C_{24}H_{36}O_{12}$	$C_{23}H_{36}O_{13}$
$C_{10}H_{17}NO_7S$	$C_{26}H_{47}NO_{10}S$	$C_3H_6O_6$	$C_{22}H_{32}O_{12}$
$C_{10}H_{17}NO_8$	$C_{28}H_{49}NO_{10}S$	$C_{14}H_{24}O_6$	$C_{12}H_{22}O_7S$
$C_{10}H_{18}O_7S$	$C_{25}H_{45}NO_{10}S$	$C_{14}H_{22}O_7$	$C_{25}H_{40}O_{13}$
$C_{10}H_{17}NO_9S$	$C_{28}H_{51}NO_{10}S$	$C_{25}H_{38}O_{12}$	$C_{23}H_{38}O_{12}$
$C_{10}H_{18}O_5S$	$C_{22}H_{39}NO_{10}S$	$C_6H_{12}N_2O_{10}$	$C_{22}H_{34}O_{13}$
$C_{10}H_{17}NO_{11}S$	$C_{11}H_{18}O_7$	$C_{26}H_{40}O_{12}$	$C_{12}H_{20}O_5$
$C_{10}H_{18}O_8S$	$C_{27}H_{47}NO_{10}S$	$C_{16}H_{26}O_7$	$C_{26}H_{42}O_{13}$
$C_{11}H_{19}NO_{10}$	$C_8H_4O_{14}S$	$C_{15}H_{24}O_7$	$C_{24}H_{40}O_{12}$

4.2.6 Other multivariate statistical analyses using the dimension reduced normalized RAs

I also conducted other statistical analyses using the normalized RAs for the molecular formulas assigned with a threshold of 10 times the RMS of the S/N. But the analyses results did not provide much valuable characterization of the WSOC composition. Though the results are not shown here, they are documented for future research considerations. Different distance measurement methods other than Manhattan method were tested using the HCA. The dendrograms were not consistent with each other, indicating that the grouping of the WSOC samples is subject to the distance methods selected. The HCA and PCA were tested using only the molecular formulas in common with LANLSOA and also using the dataset without any LANLSOA common formulas.

Also since the relative magnitude of the SOA molecular formulas plays a role in determining the WSOC composition, they should be included in the statistical analysis. The RAs were converted to a logarithm ratio to the smallest RA in a sample to conduct statistical analysis. The sample replicates do not group together in the dendrogram, reflecting the importance of the normalization of the RAs in the analysis.

4.2.7 Multivariate statistical analysis using scaled normalized RA

In the PCA biplot (Figure 4.17) with normalized RAs, a few of the formulas stand out of the region with a large number of formulas and drive the PCA result. They are the formulas with high RAs. To lower the effects of the high RA so that the importance of the lower RA formulas are not masked, the normalized RAs were centered to 0 and scaled to have a standard deviation of 1 done on every samples. In this way, all of the molecular formulas are regarded more equally in the statistical analysis. The scaling was done on every variable, as described in more detail in Section 3.2.5.

Considering the Euclidean, Manhattan and Pearson distance methods with complete linkage (the methods works well on the LANSOA), the major groupings are different with different distance measurements methods (Figure 4.18). However, several samples always show strong similarities regardless of the distance measurement methods used (Figure 4.8). First, the sample 0704 is usually separated from the other samples. There are a large number of unique molecular formulas only observed in 0704 compared to other samples. On the evening of July 4th, a

fireworks event took place near the mountain and is likely responsible for the composition differences. Second, samples 0720, 0625 and 0709 with high DBE values are always grouped closely in the dendrograms. Third, samples 0712, 0714, 0715, 0716, 0717 and 0718 show similar organic composition. Back trajectory analysis for these days indicates a consistent flow of air coming to the SPL from the west. Samples 0724, 0725 and 0726 with similar back trajectories usually group together; samples 0727 and 0728 with similar back trajectories group together, too. In general the three sets of samples above are clustered in one large group, indicating major similarities in their composition. Generally, the samples collected on adjacent days tend to have similar composition. The atmospheric aerosols have lifetime of days to weeks, thus a fraction of the daily aerosols collected might have been formed and processed in previous days so composition similarities are observed in the aerosol samples collected the day before.
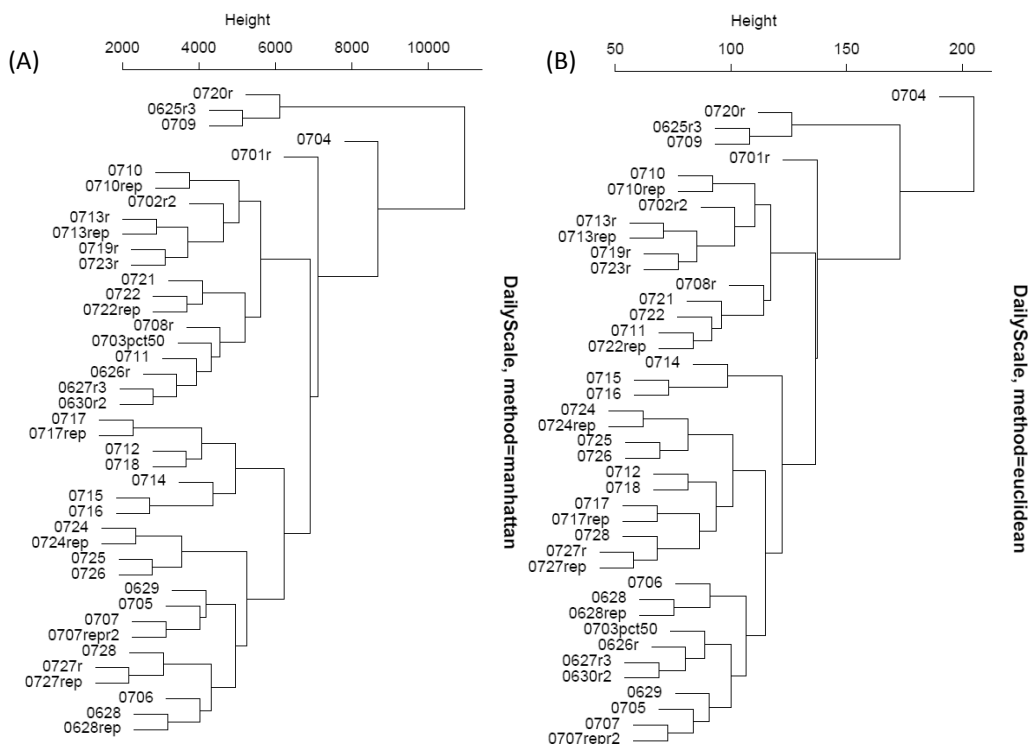
Figure 4.18 Cluster dendrograms using the scaled normalized RA and Manhattan (A) and Euclidean (B) distance methods.

Principal component analysis (PCA) was conducted on the WSOC samples to examine their molecular composition. The PCs extracted from the aerosol-derived WSOC sample dataset using scaled normalized RA are shown in Figure 4.19 (the Scree plot). PC1 represents 14.1% of the variance; PC2 represents 10.1% of the variance, followed by PC3 (7.6%). Usually the number of PCs selected for further analysis is based on the "elbow" of the Scree plot. In this case 6 PCs can be selected. However, the main purpose of the application of PCA in this study is to examine the groupings of the WSOC samples and find the formulas associated with the sample groups, instead of conducting further analysis with the PCs. To show the PCA results visually, I selected the first two PCs, PC1 and PC2, for further discussion. The WSOC samples and their molecular formulas can be shown in the PC1-PC2 space, respectively. They are compared to find the relationships between the samples and the formulas. It can be seen from the following discussion that though PC1 plus PC2 only carry ~ 23% of the variance, the characterizations of the WSOC composition are represented by the PC1 and PC2 result. The 3-dimensional plots with PC1, PC2 and PC3 as the axes were also examined. But, the addition of PC3 did not improve the groupings of the samples and it is more difficult to see the results visually than the 2-

dimensional plots with only PC1 and PC2. Thus we will use the first two PCs in the discussion of the PCA results.
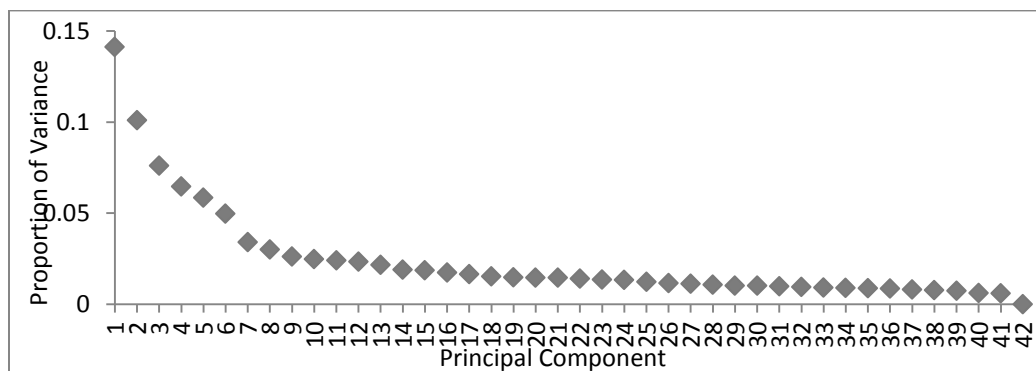
Figure 4.19 Scree plot of the PCA analysis of the WSOC samples.

The aerosol-derived WSOC samples are plotted on the PC1-PC2 space based on the PCA analysis (Figure 4.20). Consistent with the HCA dendrogram, the sample 0704 is clearly separated from the other samples with a very small PC2 value (negative value). To see the characterizations of the other samples, I removed 0704 from the dataset and conducted the PCA analysis again. Later on in Section 4.3, the meteorological parameters and other aerosol components will be used in the multivariate statistical analysis with the ultrahigh resolution MS data using the scaled normalized RA. Since organic carbon (OC) and elemental carbon (EC) data are not available for 0708, it is not included in the analysis with the meteorological parameters. To make the multivariate statistical analysis results with only ultrahigh resolution MS data comparable to those with meteorological parameter, 0708 is also not included here.

DailyWSOC

DailyWSOC

Figure 4.20 PCA score (A) and loading plot (B) for the WSOC samples using the scaled normalized RA.

Figure 4.21 PCA sample plot without the 0704 and 0708 samples.

The distribution of the sample days in the PCA results without 0704 and 0708 is shown in Figure 4.21. Consistent with the HCA results, the samples 0720, 0625 and 0709 with high DBE_w values are grouped together in the lower left corner of the PCA plot with negative PC1 and negative PC2 coordinates. Also consistent with the HCA results, most of the samples 0712 - 0728 are closely associated in the lower right corner (with positive PC1 and negative PC2) of the plot. The samples 0720, 0721 and 0722 are less closely associated in this 0712 0728 group, indicating differences in the molecular composition. Previously, it was shown that sample 0720 has high DBE_w value among the studied samples. Regarding 0721 and 0722, it was found that the organic carbon contents in the aerosol samples collected on July 21st and July 22nd are substantially lower than the samples collected before and after (Figure 4.22). Less organic carbon was observed in the aerosols on those two days and the molecular composition is also different from the adjacent samples. The potential reasons were examined and are discussed in Section 4.3. The characteristic formulas for the samples are also discussed in Section 4.3.

(A)
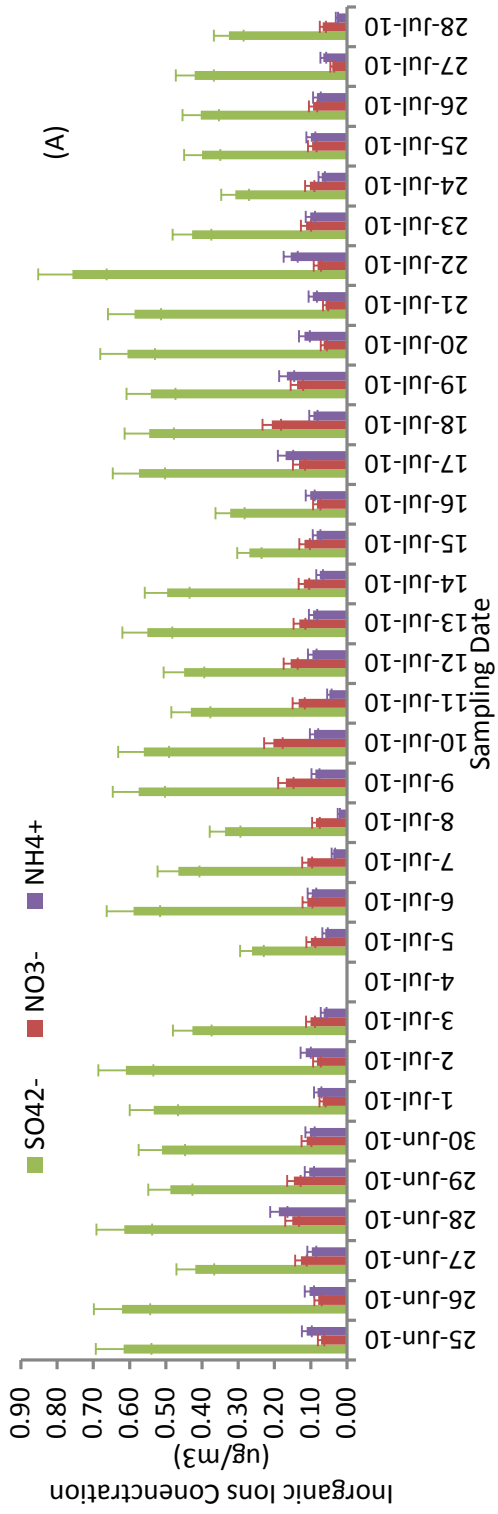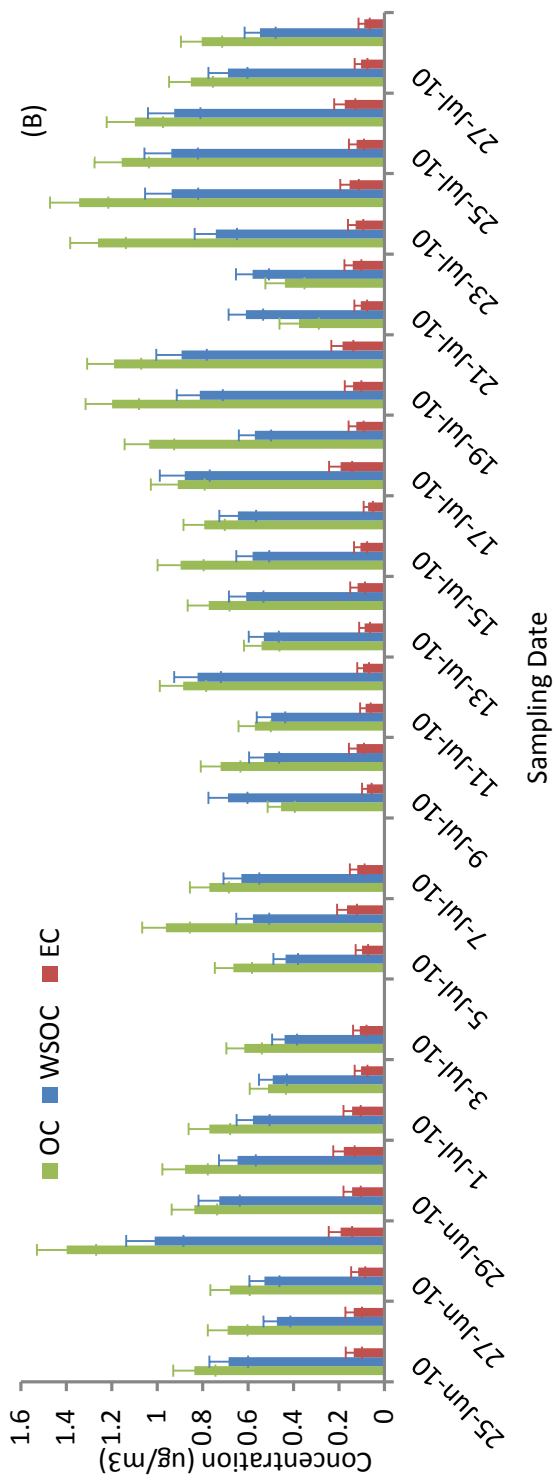
Figure 4.22 The inorganic ion concentrations (A), organic carbon, elemental carbon and WSOC concentrations (B) of the aerosol samples.

163

4.2.8 Molecular formula relationships

An evaluation of the molecular formulas was done to examine their relationships. This was done using HCA to cluster the molecular formulas, instead of the samples. To accomplish this, the 20 molecular formulas with the highest RAs were selected from each of the samples and combined and then the duplicate formulas were removed to form a new dataset. This dataset contains 60 distinct molecular formulas (observations) and 42 samples (variables) including replicates. The RAs were normalized on an individual sample basis and then evaluated using the HCA. In this way, the molecular formulas were clustered instead of the samples. The HCA dendrogram of the molecular formulas with the Manhattan distance method and complete linkage is shown in Figure 4.23. Generally the species with the same elemental compositions are grouped closely, indicating they co-vary in the samples. Many of the closely grouped formulas are in the same CH_2-homologous series. Although, some clusters include formulas with different elemental composition, such as $C_{10}H_{17}NO_8$, $C_{11}H_{18}O_7$ and $C_{11}H_{18}O_6$.
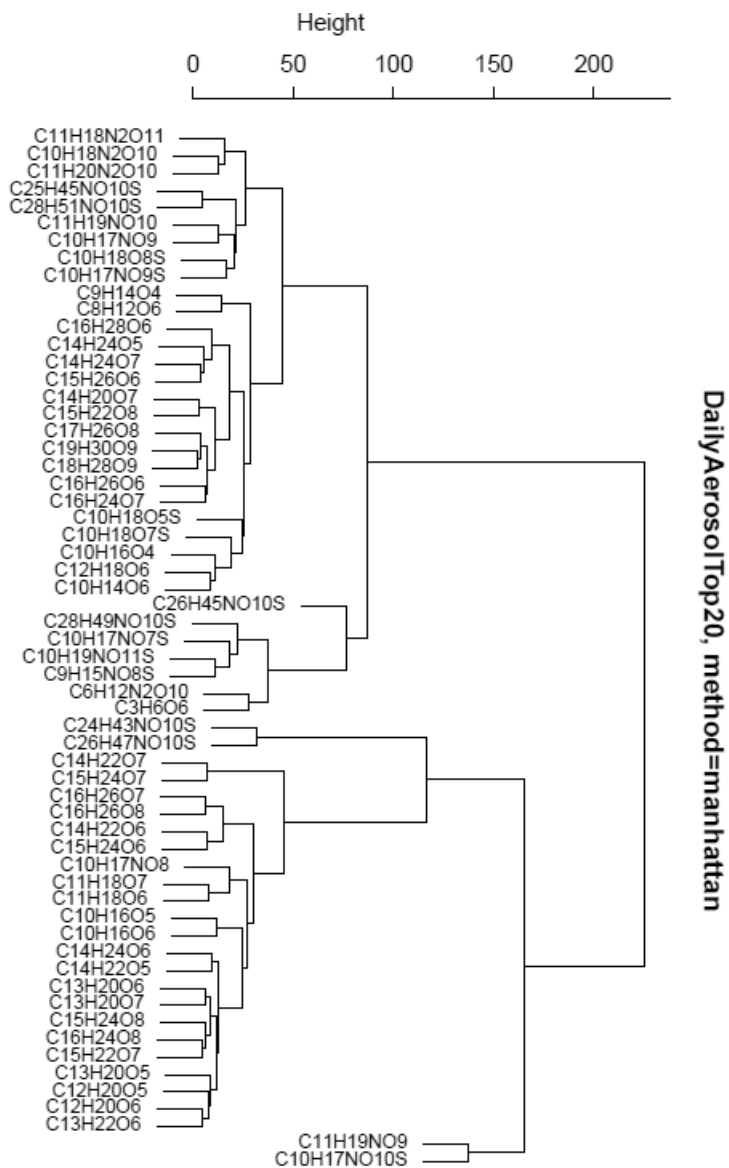
Figure 4.23 The HCA dendrogram of the top 60 high abundance molecular formulas in the daily WSOC samples using the Manhattan distance method and complete linkage.

4.3 Multivariate statistical analysis with additional aerosol and meteorological measurements

4.3.1 Aerosol and meteorological measurements

To better understand the WSOC composition differences and the environmental parameters affecting the molecular composition, additional aerosol measurements and meteorology parameters (referred to "the other measurements") were examined together with the ultrahigh resolution MS data using HCA and PCA. The other aerosol measurements include: mass concentrations of OC, EC, WSOC, NH_4^+, NO_3^- and SO_4^{2-} as reported by Hallar et al., (2013). The meteorological measurements include air temperature, pressure, wind speed, wind direction, relative humidity (RH), and solar radiation as total, direct and diffuse radiation. The meteorological data were obtained from http://www.wrcc.dri.edu/weather/strm.html; the radiation data were obtained from http://uvb.nrel.colostate.edu/UVB/index.jsf; and the ozone data was acquired from the SPL. The data were averaged to correspond to the filter sampling start and stop times. The averaged values are shown in Table 4.4. All of the ultrahigh resolution MS normalized RA data and the other measurements were rescaled before the multivariate statistical analyses were conducted.

Table 4.4 The additional measurements used in the multivariate statistical analysis with the average and standard deviation of each measurement.

Date	NO$_3^-$ (µg/m³)	SO$_4^{2-}$ (µg/m³)	NH$_4^+$ (µg/m³)	WSOC (µg/m³)	OC (µg/m³)	EC (µg/m³)	Temperature (°C)	Relative Humidity (%)	Wind Speed (m/s)	Wind Direction (degree)	Pressure (milli bar)	Ozone Conc. (ppbv)	Total Horizontal UV irradiance (Watts/m²)	Diffuse UV irradiance (Watts/m²)	Direct UV irradiance (Watts/m²)
25-Jun-10	0.07	0.62	0.11	0.69	0.836	0.135	12.26	52.8	5.0	204	695.0	45.99	110.46	36.21	79.63
26-Jun-10	0.08	0.62	0.10	0.47	0.690	0.135	10.31	56.7	3.4	252	694.2	49.58	38.18	30.85	7.68
27-Jun-10	0.13	0.42	0.10	0.53	0.679	0.115	12.75	36.6	3.2	273	695.1	59.46	110.51	34.22	81.84
28-Jun-10	0.15	0.61	0.19	1.01	1.398	0.193	14.35	44.1	4.8	183	697.7	59.14	111.76	33.34	84.15
29-Jun-10	0.15	0.49	0.10	0.73	0.836	0.142	12.61	48.3	4.7	173	697.4	52.40	94.41	49.99	47.44
30-Jun-10	0.11	0.51	0.10	0.65	0.876	0.178	12.58	52.1	4.8	198	697.0	56.27	70.64	46.73	25.47
1-Jul-10	0.07	0.53	0.08	0.58	0.770	0.142	11.06	67.8	4.7	198	695.2	53.09	43.57	36.56	7.35
2-Jul-10	0.08	0.61	0.11	0.49	0.512	0.103	11.78	61.7	6.0	180	691.0	51.18	106.73	47.50	63.92
3-Jul-10	0.10	0.43	0.06	0.44	0.617	0.108	11.47	39.3	5.1	199	690.2	47.12	72.82	39.40	35.25
4-Jul-10	NA	NA	NA	NA	0.472	0.107	3.27	86.0	5.4	258	692.6	50.82	98.35	49.50	52.79
5-Jul-10	0.10	0.26	0.06	0.43	0.665	0.099	10.49	42.2	5.1	218	693.8	41.60	111.66	36.82	80.49
6-Jul-10	0.11	0.59	0.10	0.58	0.961	0.164	11.09	58.8	5.1	178	693.5	45.56	91.15	46.46	47.79
7-Jul-10	0.11	0.46	0.04	0.63	0.770	0.119	8.28	77.9	4.8	173	697.2	46.04	91.91	50.47	44.31
8-Jul-10	0.09	0.34	0.02	0.42	NA	NA	8.76	64.9	3.1	167	698.8	52.01	112.17	43.35	73.96
9-Jul-10	0.17	0.57	0.09	0.69	0.454	0.077	11.45	50.6	4.6	258	697.9	56.95	113.18	36.81	82.31
10-Jul-10	0.20	0.56	0.09	0.53	0.721	0.123	10.93	52.5	5.3	255	695.7	54.09	76.14	51.37	26.61
11-Jul-10	0.13	0.43	0.05	0.50	0.570	0.083	9.17	70.8	5.7	244	693.7	51.05	103.09	46.68	60.80
12-Jul-10	0.16	0.45	0.10	0.82	0.886	0.094	14.15	42.7	8.1	251	693.4	51.53	89.72	49.65	43.14
13-Jul-10	0.13	0.55	0.09	0.53	0.540	0.087	13.70	54.6	6.8	248	695.0	46.00	93.02	56.50	39.37
14-Jul-10	0.12	0.50	0.08	0.61	0.773	0.118	13.17	49.9	3.9	259	699.2	49.08	113.40	42.50	76.69
15-Jul-10	0.12	0.27	0.08	0.58	0.896	0.105	14.04	27.3	5.1	275	700.3	52.24	112.21	33.77	84.93
16-Jul-10	0.08	0.32	0.10	0.65	0.793	0.071	17.45	22.1	3.3	207	699.8	52.22	112.22	33.95	84.84
17-Jul-10	0.13	0.57	0.17	0.88	0.910	0.193	16.63	38.0	3.0	272	698.7	50.66	93.51	49.90	46.65
18-Jul-10	0.21	0.55	0.09	0.57	1.034	0.124	15.69	40.7	4.7	236	696.8	50.40	112.02	36.91	81.61
19-Jul-10	0.14	0.54	0.17	0.81	1.197	0.138	16.32	34.2	3.8	257	695.9	48.79	100.44	47.27	58.08
20-Jul-10	0.06	0.61	0.12	0.89	1.190	0.185	13.07	67.0	4.6	188	696.3	51.95	106.52	51.83	59.01
21-Jul-10	0.06	0.59	0.09	0.61	0.375	0.104	11.96	82.2	4.4	200	695.2	46.26	43.97	39.53	5.05
22-Jul-10	0.08	0.76	0.16	0.58	0.438	0.139	12.80	61.1	4.9	227	695.3	51.91	73.33	35.68	40.62

Date														
23-Jul-10	0.11	0.43	0.10	0.74	1.260	0.126	15.00	37.3	3.7	259	697.4	45.92	100.84	46.20	59.57
24-Jul-10	0.10	0.31	0.07	0.94	1.343	0.153	15.96	37.2	2.5	189	697.9	42.20	109.79	33.65	83.33
25-Jul-10	0.10	0.40	0.10	0.94	1.156	0.122	16.32	37.3	3.6	226	696.8	43.78	95.74	62.98	35.98
26-Jul-10	0.09	0.40	0.08	0.93	1.098	0.175	15.63	43.5	3.8	278	696.1	52.08	110.19	37.55	79.72
27-Jul-10	0.04	0.42	0.07	0.69	0.851	0.102	13.22	69.1	5.3	168	698.2	47.08	109.59	46.04	69.99
28-Jul-10	0.07	0.33	0.03	0.55	0.804	0.089	12.70	75.7	4.6	183	699.6	44.57	45.01	38.39	7.13
Average	0.11	0.49	0.09	0.66	0.83	0.13	12.66	52.44	4.62	221.62	696.1	49.97	93.48	42.90	54.63
Standard	0.04	0.12	0.04	0.16	0.27	0.03	2.84	15.92	1.13	36.59	2.43	4.41	22.58	7.72	25.41

*The inorganic ion concentrations and WSOC, OC and EC concentrations were determined from filters.

**Temperature, relative humidity, wind speed, wind direction, pressure and ozone concentration are averaged values from the start sampling time to the end sampling time (~24 hours) of each date.

***The UV radiation data is the sum of radiation measured at 7 wavelengths: 300, 305, 311, 317, 325, 332 and 368 nm, from 12pm to 2pm local time on every sampling day.

4.3.2 Hierarchical cluster analysis

Hierarchical cluster analysis was conducted to provide insight on the WSOC molecular composition with respect to the other measured parameters such as the averaged meteorological conditions of the sample days. Thus the dendrogram reflects clustering of the comprehensive conditions of the samples including the aerosol components and the meteorological parameters, instead of clustering only WSOC molecular composition. The dendrogram of the results were generated using the Manhattan distance and complete linkage methods (Figure 4.24). The only difference between this analysis and that presented in Figure 4.18A is the addition of the aerosol and meteorological measurements data. The results of the two datasets are quite similar to each other. For example, 0720, 0625 and 0709 are grouped and separated from the others; and 0712 to 0718 except 0713 are grouped together in both dendrograms. Overall, the addition of the aerosol and meteorological measurements does not significantly affect the clusters. This is at least partially due to the preparation of the data before the multivariate statistical analysis. Since all of the variables, including the molecular formulas and the aerosol and meteorological measurements, were "centered" to have mean values of 0 and "scaled" to have standard deviations of 1, all of the variables contribute the same to the hierarchical cluster analysis. The WSOC molecular formulas comprise a much larger fraction of the total number of variables than the other measurements (5710 molecular formulas without sample 0704 and 0708, compared to 15 other measurements), thus the groupings are dominated by the WSOC molecular formulas. The variables can be adjusted to have different center values and scales so that they are assigned different contributions to the cluster analysis. For example, the concentration of aerosol organic species might be lower than the SO_4^{2-} concentrations, thus SO_4^{2-} could be regarded more significantly than the individual organic species when clustering aerosol composition. However in this study, we focused on the molecular composition of aerosol-derived WSOC. Also the meteorological parameters with different units were included in the statistics. To make all the variables (the formulas, aerosol and meteorological measurements) comparable with each, the aerosol and meteorological measurements were centered and scaled like those have been done on the ultrahigh resolution MS data (Section 4.2.7).
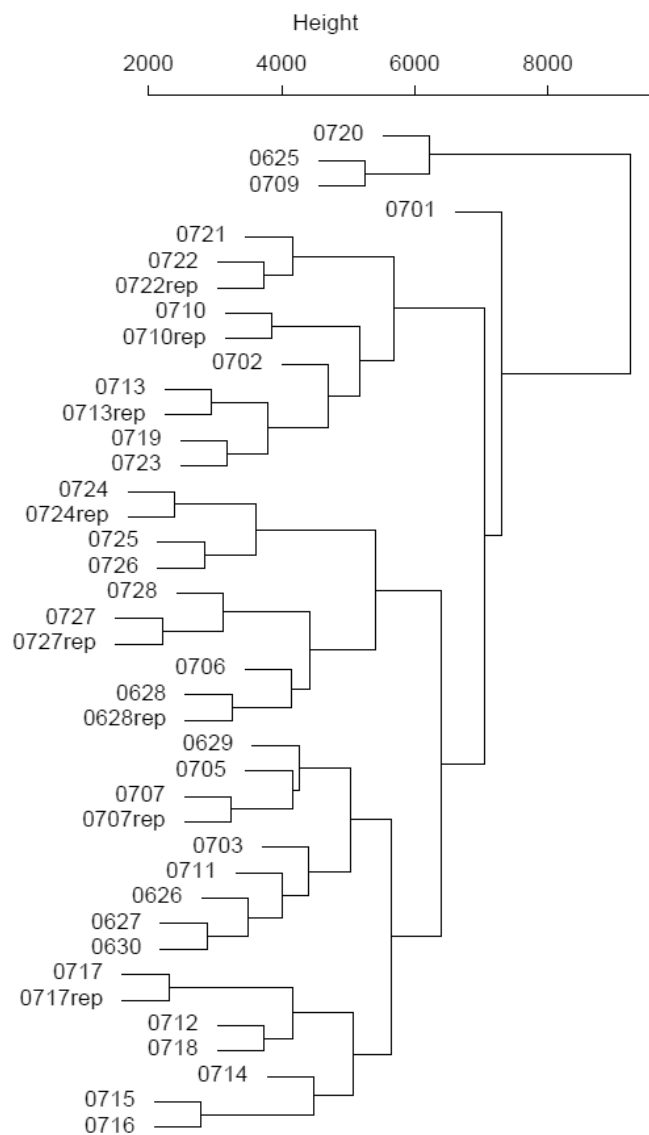
Figure 4.24 Dendrogram of the hierarchical cluster analysis of all the measurements (the WSOC molecular formulas, the aerosol component concentrations, meteorological parameters and ozone concentrations). The Manhattan distance measurement and complete linkage methods were applied.

4.3.3 Principal component analysis

4.3.3.1 General characterization of PCA

Principal component analysis was conducted on the WSOC composition together with the aerosol and meteorological measurements to examine the effects of the meteorological conditions on WSOC composition. Similar to the results of the PCA with only the WSOC composition (Figure 4.20 and 4.21), the first two PCs carry ~24% of the original variance (Figure 4.25 and 4.26). The first principal component (PC1) represents 15.4% of total variance and the second one (PC2) represents 8.2% of the variance. The addition of PC3 does not add much to the total variance, but sacrifices the clarity of the visual plots and thus is not included. Although a small fraction of the total variance observed in 5725 variables (5960 formulas from 10 times the RMS of S/N minus the unique formulas appearing in 0704 and 0708) is represented by the first two principal components, a great amount of valuable information about the variable relationships is indicated by the results. In addition, the locations of the samples and variables shown in the 2-dimensional PCA plots here are largely consistent with the non-centered and non-scaled PC1-PC2 PCA plots in which the first two PCs represent more than 90% of the variance (Figure 4.27). Thus we will use the first two PCs for further discussion.

The sample dates are plotted on the PC1-PC2 space in Figure 4.25 based on their PCA scores. The distributions of samples are almost the same as shown in Figure 4.21 (PCA score plot of WSOC molecular composition only). Though the dates here represent the comprehensive conditions of the sampling dates including the aerosol components and the meteorological conditions, their distribution on the PC1-PC2 plot is driven by the WSOC molecular formulas due to their high number of variables. The effect of the addition of the aerosol and meteorological measurements (addition of the 15 variables to the 5710 variables of formulas) is minor and even rarely seen on the PC1-PC2 plot with ~24% of the original variance. The similar distribution of the dates on the PCA score plots with and without the other measurements (Figure 4.25 and Figure 4.21) favors the data analysis; the dates shown on Figure 4.25 could be regarded as WSOC samples and the variables (the formulas and other measurements) could be related to the WSOC samples.
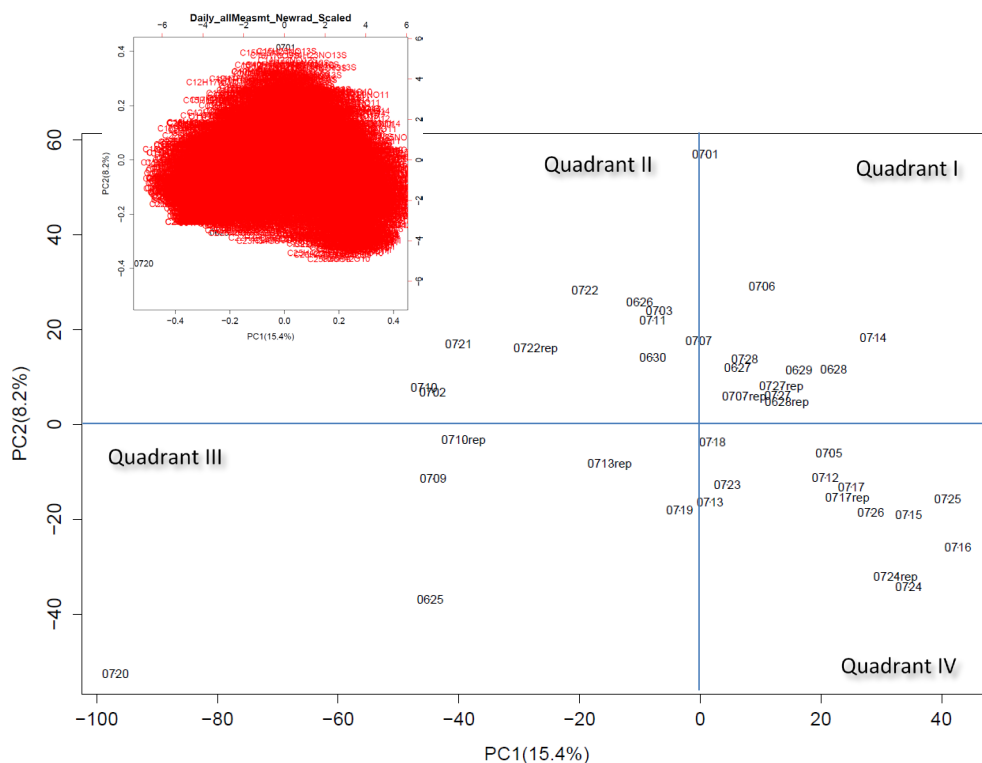
Figure 4.25 PCA score plot with the additional measurements. The PCA biplot is shown in the insert.

The results of the PCA with the WSOC molecular composition and the additional aerosol and meteorological measurements provide insights to the relationships among these variables. The molecular formulas, the aerosol chemistry (concentrations of NO_3^-, SO_4^{2-} and NH_4^+, organic carbon content, elemental carbon content, WSOC content), the meteorological parameters (relative humidity, wind direction, wind speed, atmospheric pressure, temperature, total horizontal UV irradiance, direct UV irradiance, diffuse UV irradiance) and the concentrations of ozone are plotted on the PC1-PC2 space based on their PCA loadings (Figure 4.26). In PCA, the score of an observation (or a sample in this case) on a dimension (*e.g.*, score of PC1) is the linear combination of the variable loadings on that dimension for that sample. The variables with large absolute loading values (either positive or negative) have a more significant impact on the variation of the samples than the variables with smaller absolute loading values. In other words as shown visually on the PCA loading plot, the more distant a variable is from the origin, the more impact it has on the variations of the WSOC samples.
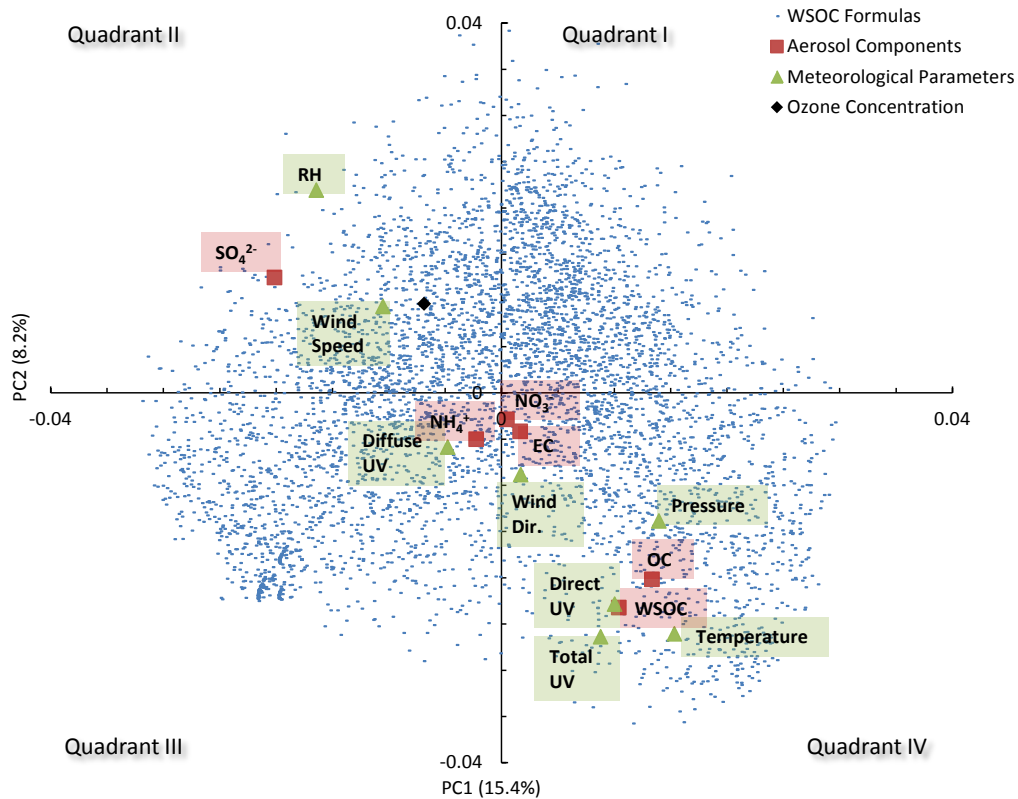
Figure 4.26 The PCA loading plot of the WSOC molecular formulas and the additional aerosol and meteorological measurements. The aerosol and meteorological measurements are labeled in the plot.

The molecular formulas of the WSOC are spread over the four quadrants of the PCA loading plot (Figure 4.26). Most of the aerosol and meteorological measurements are located in quadrants II and IV. For example, SO_4^{2-}, RH and wind speed are in quadrant II and temperature, pressure, direct UV irradiance, total horizontal UV irradiance, OC and WSOC are in quadrant IV. Others are around the origin. This distribution of variables is consistent with the PCA loading plot containing the non-centered and non-scaled variables (including both the WSOC molecular formulas and the other measurements). In the loading plot without scaling shown in Figure 4.27, the first two PCs represent 91.3% of the total variance. However, the absolute values of some of "the other measurements" are several orders of magnitude larger than the aerosol components causing them to stretch out in the PC1-PC2 space. For example, the values of the total horizontal UV irradiance are around 100 W m^{-2} and the temperatures are around 10 ˚C, while most of the normalized RAs of the formulas are around 0.01. The "stretched out" variables represent more of the variation, which results in the first two PCs carrying much more of the total variance

(Figure 4.27). The centering and scaling of the variable magnitudes eliminates the "stand out" variable effect on the principal component analysis. This yields a lower representation of the total variance by the first two PCs (Figure 4.26), but the atmospheric implications are the same. Though the centering and scaling treatment regards all of the variables as having the same significance in the principal component analysis, the non-centering and non-scaling method is not more realistic. For example, the values of the total horizontal UV irradiance are generally 10 times of the temperature, but the importance of the UV radiation in the multivariate statistical analysis is not necessary 10 times higher than the temperature. Thus due to the high number and complexity of variables, the variables were centered and scaled before the PCA.
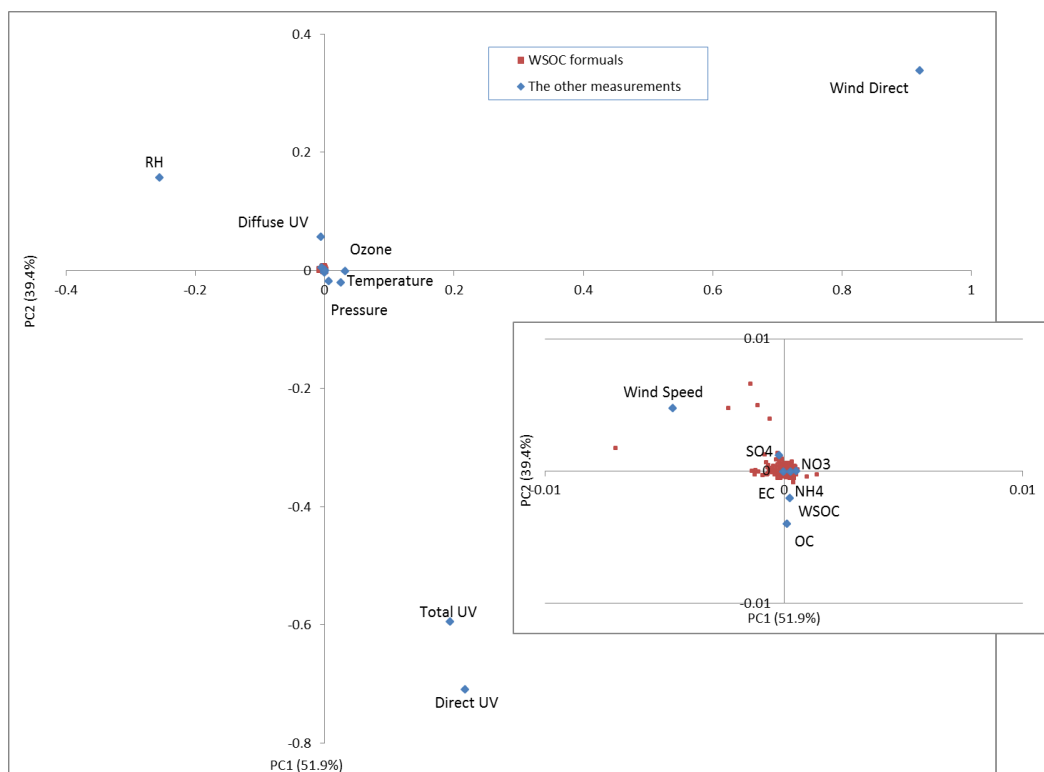
Figure 4.27 PCA loading plot with non-centered, non-scaled WSOC molecular formulas and the other measurements, for comparison with the centered and scaled PCA loading plot in Figure 4.26. The "other measurements" shown with blue diamonds are labeled. The zoomed in plot around the origin is shown in the excerpt.

4.3.3.2 Molecular formula compositions on the PCA plot

A major purpose of conducting PCA is to explore the molecular formulas and other key variables for compositional signatures of atmospheric aerosol processes. Thus those molecular formulas that are rich in some of the samples in terms of both appearance and relative abundance and are correlated to key variables are of interest for further study. The relative abundances of the WSOC components for select samples can be shown in the "PCA bubble loading plots" (Figure 4.16). In Figure 4.28 A, the WSOC molecular formulas of sample 0706 are shown in the PC1-PC2 space with the symbol (or bubble) size representing the relative abundances of the formulas in that sample. The aerosol and meteorological measurements are also shown for reference in the plot with a uniform symbol size. The sample 0706 was selected to represent the composition of the samples in Quadrant I on the PCA score plot in Figure 4.25. Similarly, the samples 0721

175

(Figure 4.28 B), 0720 (Figure 4.28 C) and 0716 (Figure 4.28 D) were selected to represent samples in Quadrants II, III and IV as defined in Figure 4.25, respectively.

0721 (B)

PC2 (8.2%)

PC1 (15.4%)

- Formulas
- Aerosol Components
- Meteorological Parameters
- Ozone

0720 (C)

PC2 (8.2%)

PC1 (15.4%)

- Formulas
- Aerosol Components
- Meteorological Parameters
- Ozone

177
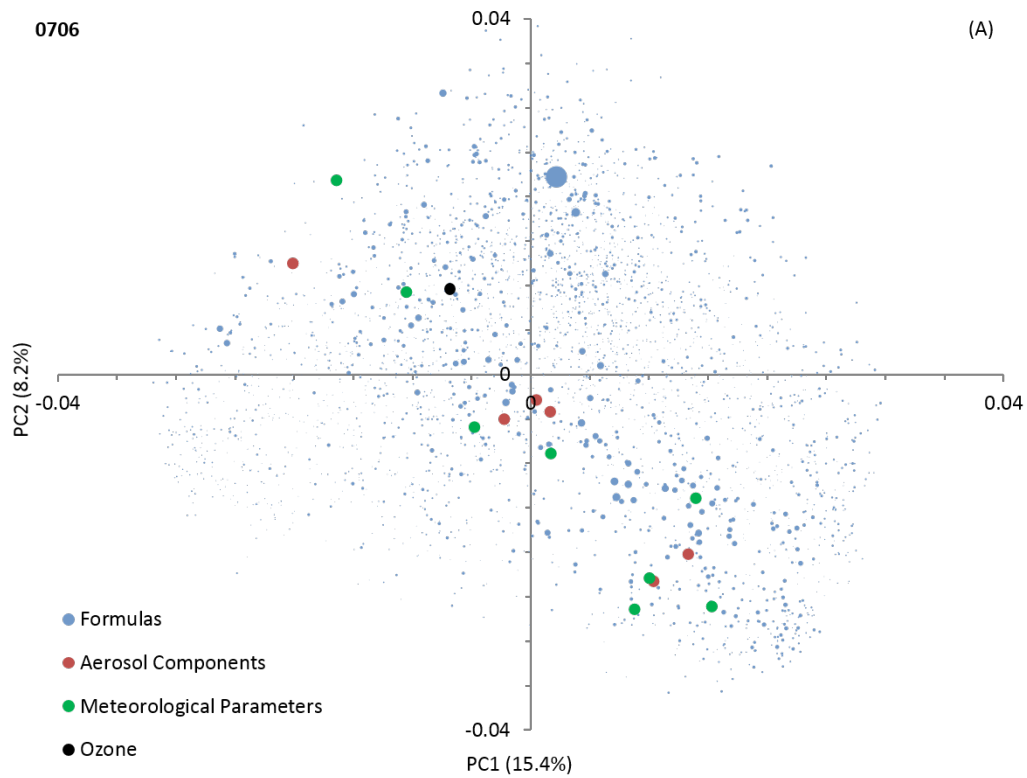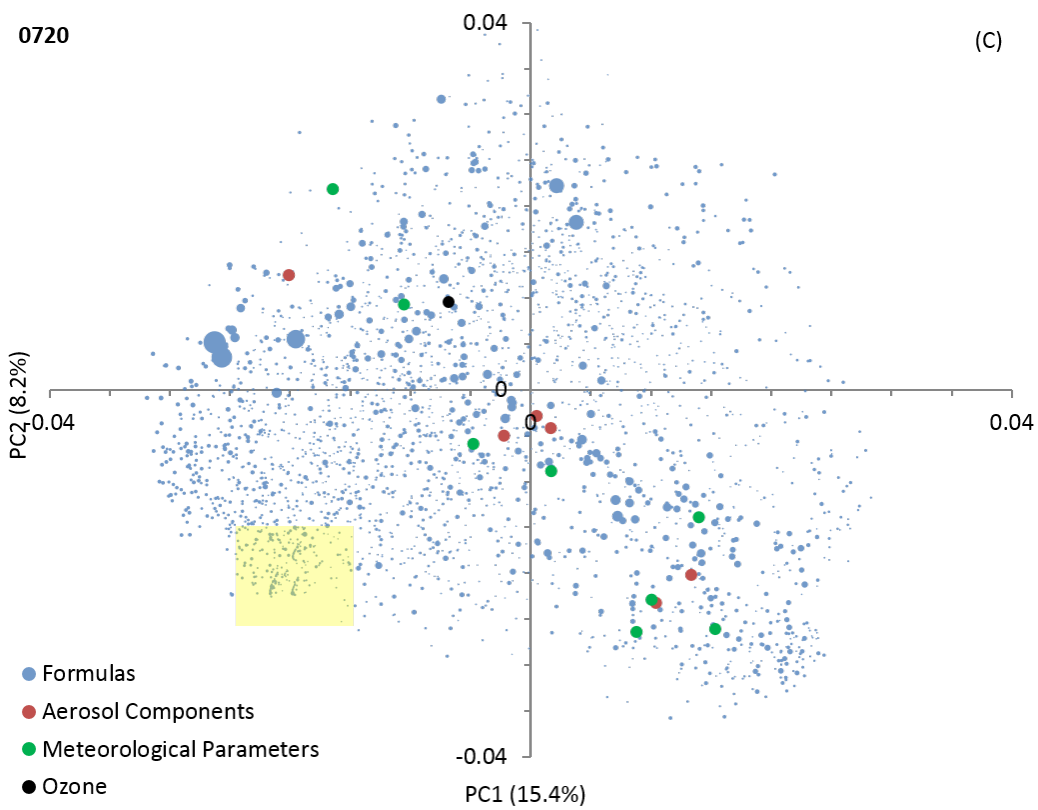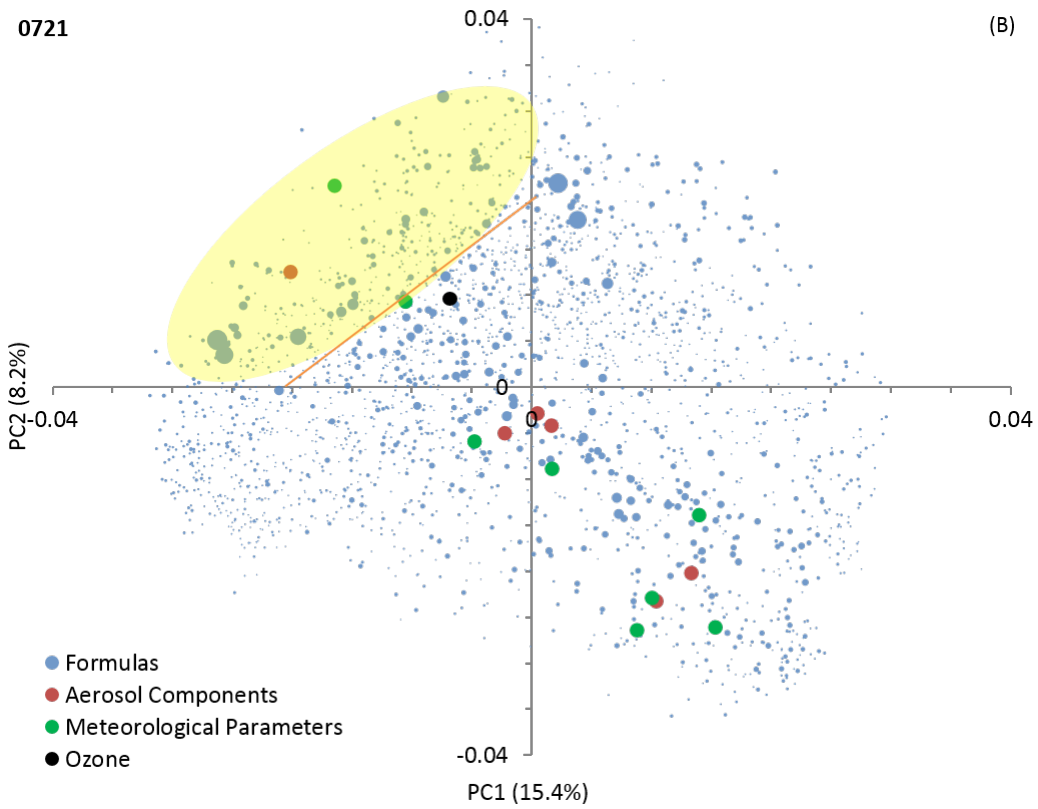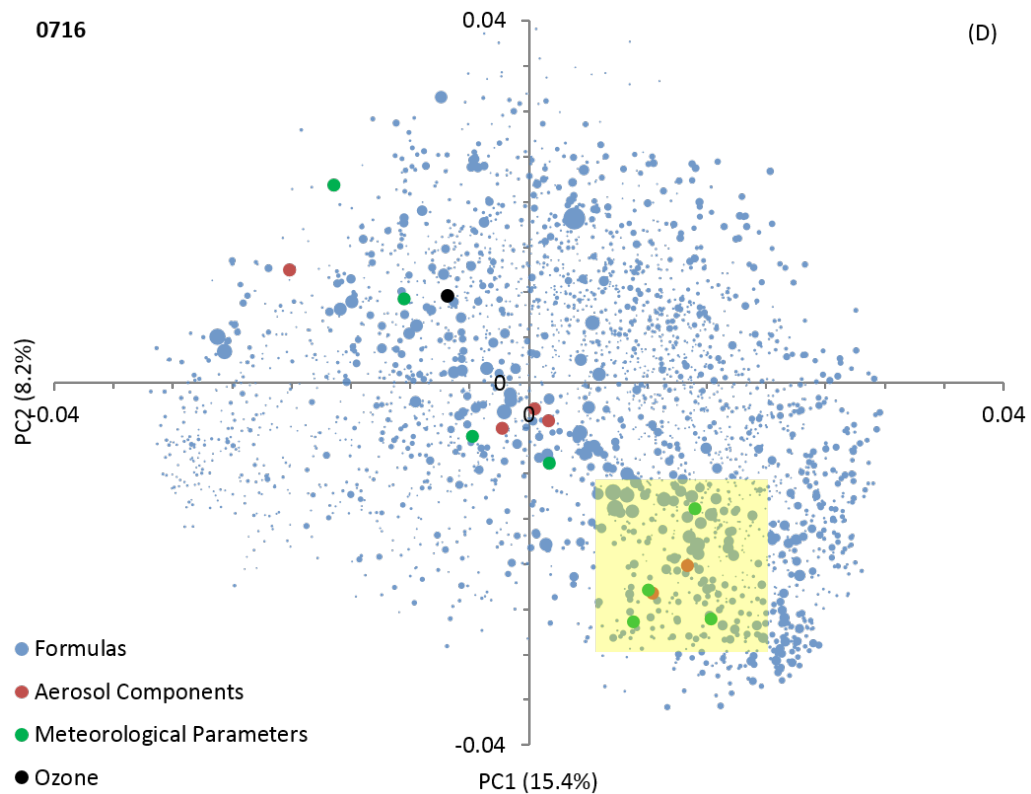
Figure 4.28 PCA loading plots with symbol size representing the relative abundances of the WSOC molecular formulas in samples 0706 (A), 0721 (B), 0720 (C) and 0716 (D). The other measurements are also shown for reference. The samples in A, B, C and D represent the samples co-located in quadrants I, II, III and IV of the PCA score plot as shown in Figure 4.25, respectively.

4.3.3.3 Aerosol measurements and meteorological parameters in PCA plot

Regarding the aerosol chemistry, SO_4^{2-} is located in quadrant II of the PCA loading plot with negative PC1 and positive PC2 loadings; NH_4^+ is located in quadrant III with negative PC1 and PC2 loadings; NO_3^-, elemental carbon, organic carbon and WSOC content reside in quadrant IV with positive PC1 and negative PC2 loadings (Figure 4.26). Among the major aerosol components measured in this study (not including the WSOC molecular composition) SO_4^{2-}, organic carbon and WSOC content are further away from the origin than NO_3^-, NH_4^+ and elemental carbon in the PCA loading plot. Thus the concentrations of SO_4^{2-}, organic carbon and WSOC are more significant in terms of explaining the overall variation in the WSOC composition than NO_3^-, NH_4^+ and elemental carbon.

Regarding the meteorological parameters, RH and wind direction are located in quadrant II; diffuse UV irradiance is located in quadrant III; and the other parameters are in quadrant IV (Figure 4.26). It is shown in Figure 4.26 that RH, direct UV irradiance and temperature are the variables farthest away from the origin, so they have the most significant impact among the measured meteorological conditions on the aerosol-derived WSOC composition. Their effect on the WSOC molecular composition will be discussed in detail later. The variable of wind direction is located in quadrant IV of the PCA loading plot (Figure 4.26). Wind direction indicates the recent history of the air masses intercepted at the sampling site. The air mass history and upwind emission sources are an important factor influencing the aerosol composition, particularly when aerosols arrive at the site from clearly different wind directions. In this study, the daily aerosol samples were collected continuously over about a month at the site. The wind direction was predominantly from the west and did not significantly change during the study period. Almost all of the daily average wind directions were between 180 and 270 degree, indicating the wind is generally from the southwest (Table 4.4). Thus the impact of wind direction on the aerosol composition is not as strong as RH or temperature.

Back trajectory analysis results provide further information about the air mass history. Consistent with the wind direction measurements, the air was generally from the west (or northwest or southwest) over the 72-hour time period prior to reaching the SPL. The results from the NOAA HYSPLIT back trajectory analysis for selected dates (0625, 0626, 0709, 0710, 0719 and 0720) are shown in Figure 4.29. Wozniak et al. (2014) reported that the air mass

origins were largely responsible for the variation in the aerosol-derived WSOC molecular

composition of samples collected in the North Atlantic. They distinguished North American-

influenced, North African-influenced and marine-influenced aerosol samples collected during

the 2011 trans-North Atlantic US GEOTRACES cruise. However, the aerosol samples studied here

have very similar air back trajectories and are influenced by local SOA production. Thus, the

composition of the samples is not clearly grouped based on the back trajectory analysis alone.

Instead, as discussed later, the local meteorological conditions are shown to affect the aerosol

composition.

(A)

NOAA HYSPLIT MODEL
Backward trajectories ending at 1400 UTC 25 Jun 10
EDAS Meteorological Data
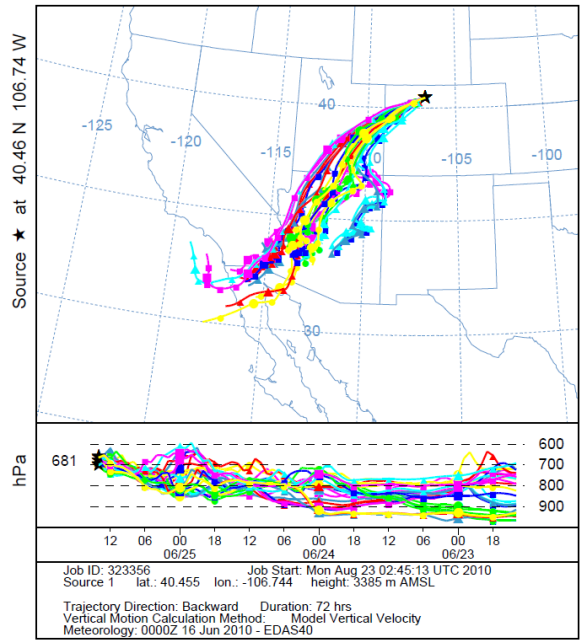
(B)

NOAA HYSPLIT MODEL
Backward trajectories ending at 1400 UTC 26 Jun 10
EDAS Meteorological Data
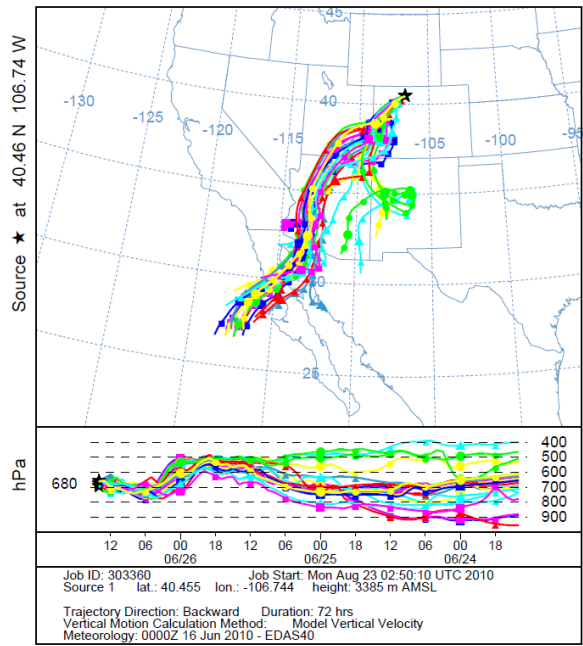
(C)

NOAA HYSPLIT MODEL
Backward trajectories ending at 1400 UTC 09 Jul 10
EDAS Meteorological Data

Source ★ at 40.46 N 106.74 W

hPa

684

500
550
600
650
700
750
800
850

12 06 00 18 12 06 00 18 12 06 00 18
07/09   07/08   07/07

Job ID: 396996   Job Start: Thu Aug 26 04:38:44 UTC 2010
Source 1   lat.: 40.455   lon.: -106.744   height: 3390 m AMSL

Trajectory Direction: Backward   Duration: 72 hrs
Vertical Motion Calculation Method: Model Vertical Velocity
Meteorology: 0000Z 01 Jul 2010 - EDAS40
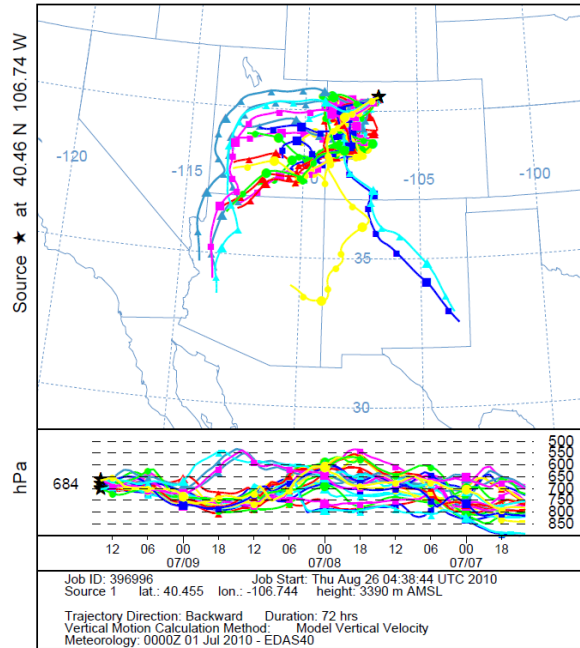
(D)

NOAA HYSPLIT MODEL
Backward trajectories ending at 1400 UTC 10 Jul 10
EDAS Meteorological Data
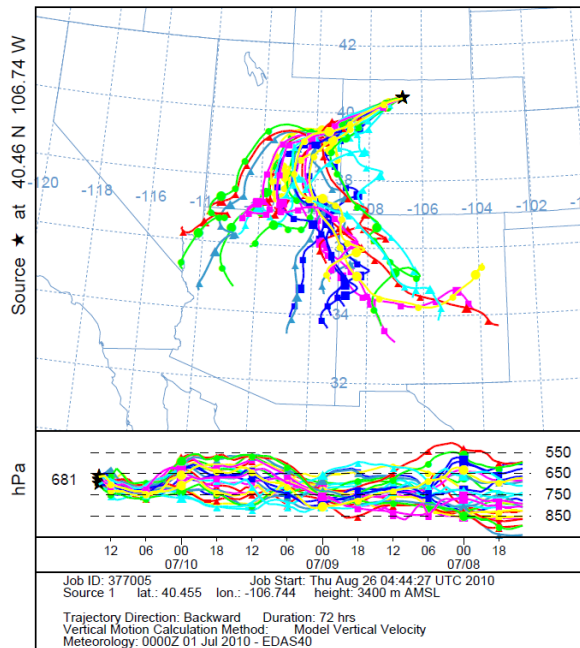
Source ★ at 40.46 N 106.74 W

hPa

681

550
650
750
850

12 06 00 18 12 06 00 18 12 06 00 18
07/10   07/09   07/08

Job ID: 377005   Job Start: Thu Aug 26 04:44:27 UTC 2010
Source 1   lat.: 40.455   lon.: -106.744   height: 3400 m AMSL

Trajectory Direction: Backward   Duration: 72 hrs
Vertical Motion Calculation Method: Model Vertical Velocity
Meteorology: 0000Z 01 Jul 2010 - EDAS40

182
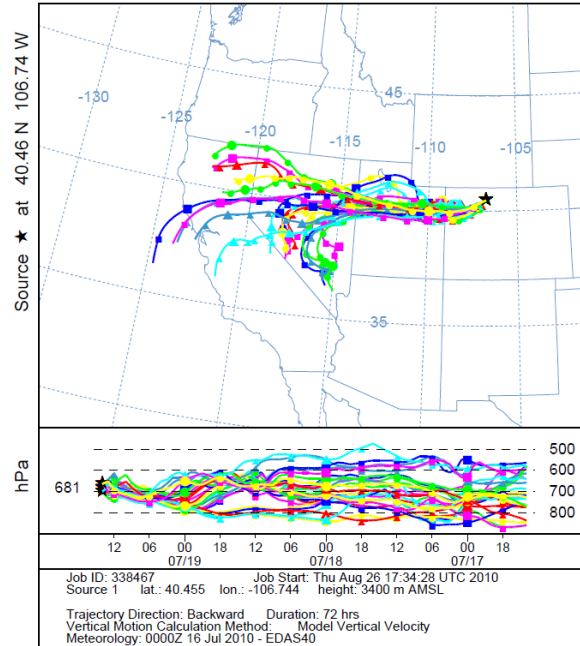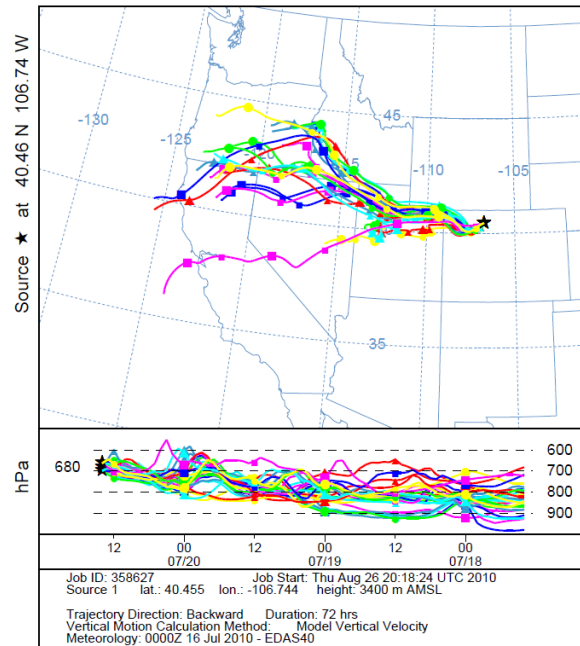
Figure 4.29 Back trajectory analysis results using the NOAA HYSPLIT model for sample 0625 (A), 0726 (B), 0709 (C), 0710 (D), 0719 (E) and 0720 (F).

4.3.3.4 Correlation of the variables

In the PCA loading plots with two dimensions of PC1 and PC2, the variables at similar regions have similar PC1 and PC2 loadings. Thus they have similar contributions to the variations of the samples or in other words, they co-vary in the samples. With the distribution of the variables on the PCA loading plot, the relationships among the PCA variables can be explored.

4.3.3.4.1 OC, WSOC, temperature and UV radiation

In this analysis, a strong correlation between the organic carbon and water-soluble organic carbon concentrations in the aerosol samples is shown in Figure 4.26. This positive relationship between OC and WSOC was confirmed by their concentrations shown in Figure 4.22. As introduced in Chapter 1, WSOC is usually a large fraction of the total organic mass as observed in various studies and this is especially true for the rural areas. In this study, WSOC generally represents ~80% of OC. The environmental parameters that mainly affect the WSOC concentration in aerosols at the SPL can be determined from the PCA results (Figure 4.26). As shown, the temperature, total horizontal UV irradiance and direct UV irradiance are the three meteorological parameters mostly closely associated with the observed aerosol WSOC concentrations. Of these, the direct UV irradiance is especially significant as indicated by the near overlap with WSOC in the PCA loadings plot. In fact, the three parameters are expected to positively correlate with each other. Direct UV irradiance is a major part of the total horizontal UV irradiance. It is the solar radiation reaching the Earth's surface without being reflected. The solar radiation heats up the Earth surface air, thus strong UV radiation results in high air temperature. The combined influence of the strong UV radiation and high temperature leads to high WSOC concentrations in aerosols.

Several potential reasons might explain these correlations between OC, WSOC, UV irradiance and temperature. First, UV radiation produces atmospheric oxidants, which generate biogenic SOA. The most important atmospheric oxidant in the daytime is the hydroxyl radical (OH). The major source of OH is the UV photolysis of ozone to generate electronically excited $O(^1D)$, followed by its reaction with water vapor (Finlayson-Pitts and Pitts, 1997):

$$O_3 + h\nu \ (\lambda < 336 \ nm) \rightarrow O(^1D) + O_2$$

$$O(^1D) + H_2O \rightarrow 2OH$$
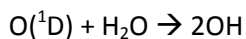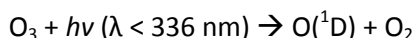
Another important oxidizing agent in the atmosphere, O_3, is also mainly formed through photochemical reactions involving NO_2 (Finlayson-Pitts and Pitts, 1997):

$$NO_2 + h\nu \ (\lambda < 420 \ nm) \rightarrow NO + O(^3P)$$

$$O(^3P) + O_2 + M \rightarrow O_3$$
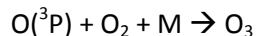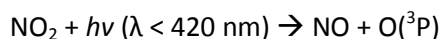
These reactions indicate that stronger UV radiation generates more oxidizing agent in the atmosphere. Oxidants react readily with the biogenic VOCs emitted by the vegetation in the vicinity of the SPL to generate biogenic secondary organic aerosols. In fact, as discussed in Section 4.1, a substantial fraction of the WSOC molecular formulas was observed to be in common with the LANLSOA samples. Dzepina et al. (2009) also observed increases in the oxygenated organic aerosol (a surrogate of SOA) concentrations with the OH concentrations and these corresponded with an increase in the UV irradiance from 9 am to 3 pm in Mexico City. At SPL, strong UV radiation results in more biogenic SOA and because most of the biogenic SOA components are water-soluble this contributes to WSOC. Second, strong UV radiation corresponds to a high degree of photochemical aging of aerosol organics. The photochemically formed oxidants continuously react with either the gas phase organics or the species of the condensed phase to produce oxidized products. Aging processes transform organic compounds to products with more hydrophilic functional groups and they make the organic compounds more hygroscopic and water-soluble (Jimenez et al., 2009). In addition to these oxidation reactions, it has been proposed that photochemical or radical reactions are one of the primary mechanisms for the condensed phase formation of organosulfates, which are water-soluble (McNeill et al., 2012). Thus the sampling dates with higher UV irradiance are expected to have higher WSOC concentrations. Third, it is well established that vegetation emits more biogenic VOCs at higher temperatures (*e.g.*, Kesselmeier and Staudt, 1999). On the days with high average temperatures more SOA is formed due to the higher concentrations of VOCs, thus high WSOC content in aerosols is expected. Hennigan et al. (2009) also observed a positive relationship between the particle phase WSOC and temperature during the summer time in Atlanta at a site largely affected by biogenic SOA. In addition, the high temperatures also accelerate the photochemical aging reactions mentioned above to produce more WSOC species.

Comparing the PCA loading plot (Figure 4.26) and the PCA score plot (Figure 4.25), the samples in the same region as the variables of WSOC, OC, temperature and UV radiation are the samples

characterized by these variables. These samples include 0712, 0715 - 0717 and 0724 - 0726. The variables of temperature, total horizontal UV irradiance, direct UV irradiance are high on these days and the concentrations of OC and WSOC are high in the aerosol samples. Further, the molecular formulas in the same region are closely associated with these variables. There are hundreds of molecular formulas in this region (far from the origin in quadrant IV). Though the RAs of the formulas in the ESI-MS analysis do not reflect their exact concentrations in the samples, they still represent their relative amounts to a large extent. So some high RA formulas were selected to represent this region. Since the formulas at similar locations on the PCA loading plot usually co-vary in the samples, the sample of 0716 can be used to represent the high samples in quadrant IV and the formula RAs in this sample can be used to represent the relative amount of the molecular species in these samples. The WSOC molecular formulas in sample 0716 with their RAs are shown in the PCA loading bubble plot of Figure 4.28 D. The formulas in the area defined by PC1 > 0.005 and PC2 < -0.01 were selected to represent the WSOC formulas affected by high temperature and strong UV irradiance (the highlighted region in Figure 4.28 D). The 20 molecular formulas in the highlighted region in Figure 4.28 D with the highest RAs in sample 0716 are shown in Table 4.5. The RAs of these formulas are expected to be slightly different in other samples also affected by the temperature and UV radiation (samples of 0712, 0715, 0717 and 0724 to 0726), but they show high RAs in all the samples at this region (0712, 0715, 0716, 0717, 0724, 0725 and 0726).

Table 4.5 The 20 molecular formulas in the highlighted region in Figure 4.28 D with the highest RAs in sample 0716.

Molecular Formula	Theoretical m/z	RA in 0716 (%)	DBE	O/C	H/C
$C_{14}H_{22}O_7$	301.1293	48.24	4	0.50	1.57
$C_{15}H_{24}O_7$	315.1449	46.29	4	0.47	1.60
$C_{14}H_{24}O_6$	287.1500	39.75	3	0.43	1.71
$C_{16}H_{26}O_7$	329.1606	39.71	4	0.44	1.63
$C_{16}H_{26}O_8$	345.1555	36.84	4	0.50	1.63
$C_{15}H_{24}O_8$	331.1398	33.09	4	0.53	1.60
$C_{13}H_{22}O_6$	273.1344	31.71	3	0.46	1.69
$C_{15}H_{26}O_6$	301.1657	30.98	3	0.40	1.73
$C_{14}H_{24}O_7$	303.1449	29.51	3	0.50	1.71
$C_{14}H_{24}O_5$	271.1551	28.45	3	0.36	1.71
$C_{16}H_{28}O_6$	315.1813	27.06	3	0.38	1.75
$C_{16}H_{28}O_7$	331.1762	25.67	3	0.44	1.75
$C_{13}H_{22}O_5$	257.1394	25.17	3	0.38	1.69
$C_3H_6O_6$	137.0092	25.06	1	2.00	2.00
$C_{15}H_{26}O_7$	317.1606	24.31	3	0.47	1.73
$C_{17}H_{28}O_8$	359.1711	23.00	4	0.47	1.65
$C_{20}H_{32}O_{10}$	431.1923	22.99	5	0.50	1.60
$C_{20}H_{32}O_9$	415.1974	22.99	5	0.45	1.60
$C_{18}H_{28}O_9$	387.1661	22.76	5	0.50	1.56
$C_{19}H_{30}O_9$	401.1817	22.71	5	0.47	1.58

The highest RA formulas of 0716 listed in Table 4.5 are all CHO species. They show clear biogenic

SOA characteristics with carbon numbers of ~15 or 20, O/C around 0.5 and H/C around 1.6 (see

also the discussion in Chapter 3). The highlighted formulas in quadrant IV in Figure 4.28 D were

compared with the LANLSOA formulas presented in Chapter 3. There are a total of 670

molecular formulas in the highlighted area and 483 of these are CHO species. Approximately

83% of these CHO compounds were also observed in the biogenic LANLSOA, confirming the

strong biogenic SOA characteristics of the molecular formulas associated with temperature and

UV radiation. Moreover, many of these highlighted molecular formulas in the WSOC samples are

the taller peaks in the LANLSOA spectra. Shown in Figure 4.30 are the representative LANLSOA

mass spectra of AP4 (α-pinene SOA), BP11 (β-pinene SOA), LM3 (limonene SOA) and CP3 (β-

caryophyllene SOA). The red and blue peaks are the common species between the highlighted

WSOC formulas and the LANLSOA, while the black peaks represent the unique formulas to

LANLSOA. The formulas listed in Table 4.5 are shown in blue in the reconstructed mass spectra.

The molecular formulas observed to be associated with temperature and UV radiation in the

aerosol-derived WSOC are in the mass range of m/z 300 – 600. This indicates they are higher

molecular weight biogenic SOA species defined as monoterpene SOA Groups II and III (Putman et al., 2012; Chapter 3) or sesquiterpene SOA Groups I and II (Kundu et al., 2014 In preparation; Chapter 3) with high relative abundances. The high degree of similarities between the highlighted molecular formulas and the chamber biogenic SOA confirms the relationship between the OC and WSOC concentrations in aerosols and the ambient temperature and UV irradiance.

(A)

AP4

Formuals highlighted in Qudrant IV
Formuals unique to AP4
20 top RA formulas in Qudrant IV

RA (%)

Negative m/z

(B)

BP11

Formuals highlighted in Qudrant IV
Formuals unique to BP11
20 top RA formulas in Qudrant IV

RA (%)

Negative m/z
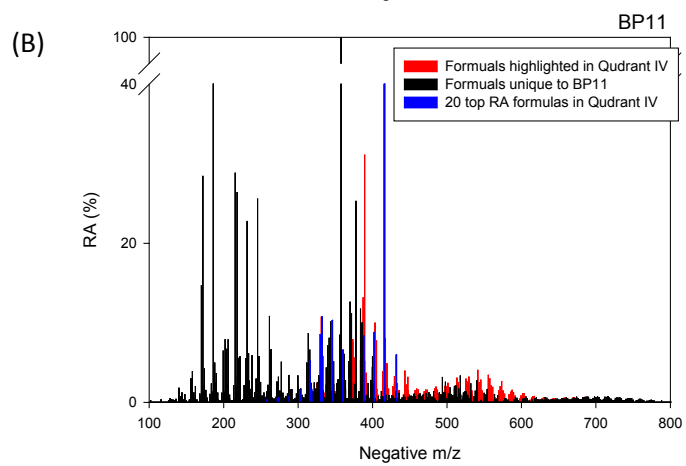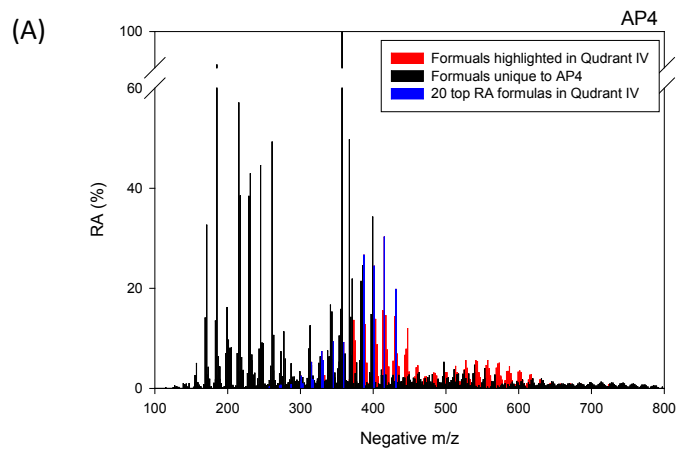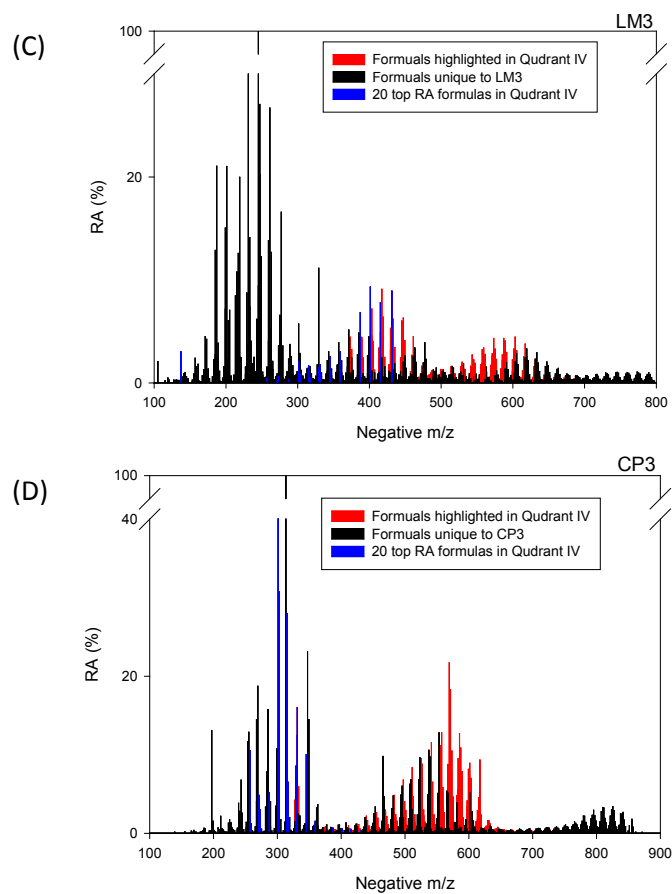
Figure 4.30 Reconstructed mass spectra of the representative LANLSOA sample AP_4 (A), BP_11 (B), LM_3 (C) and CP_3 (D). The red and blues peaks represent the molecular formulas also observed in the highlighted regions shown in Figure 4.28D. The molecular formulas shown in blue in the mass spectra are listed in Table 4.5.

4.3.3.4.2 Relative humidity and the sulfate ion concentration

The relative humidity (RH) and the sulfate ion concentration (SO_4^{2-}) are also strongly correlated in the PCA results. These two variables are both located in quadrant II of the PCA loading plot in Figure 4.26. A positive correlation between RH and SO_4^{2-} is expected. More than two thirds of the S(IV) or sulfur dioxide (SO_2) is oxidized to S(VI) or sulfate in the atmospheric aqueous phase (Harris et al., 2013). Many pathways have been proposed for the oxidation of sulfur in the aqueous phase. The well-studied reactions with best reaction rates include oxidation by dissolved ozone, hydrogen peroxide, organic peroxides, and molecular oxygen with iron or manganese catalysis (Seinfeld and Pandis, 2006). On the sampling dates when the RH is high, there is also a high tendency to form atmospheric water either in the form of aerosol water or cloud droplets. Thus the correlation of the sulfate concentrations and RH as observed in this study indicates the transformation of S(IV) to S(VI) occurs in the aqueous phase.

Shown in the PCA score plot in Figure 4.25, the samples correlated with RH and sulfate concentrations include 0721, 0722, 0702 and 0710. Thus, the WSOC composition of these days is likely highly affected by the RH and sulfate. The RAs of the molecular formulas of these days are expected to co-vary, so 0721 can be used to represent these samples. The RA of the formulas from the ultrahigh resolution MS analysis of 0721 are shown in the PCA loading bubble plot in Figure 4.28B. The molecular formulas near RH and SO_4^{2-} in the plot are most associated with the RH and SO_4^{2-} concentrations. The formulas far away from the origin in quadrant II and above the line connecting the points (-0.02, 0) and (0, 0.02) on PC1 and PC2 were selected for further analysis. There are a total of 265 molecular formulas in this region. The 20 formulas with highest RAs in 0721 are shown in Table 4.6. A large fraction of them are heteroatom-containing molecular formulas with nitrogen and sulfur. They might be nitrooxy organosulfates formed in atmospheric aqueous phase. Many of these top 20 formulas also show biogenic SOA characteristics with carbon number of ~10 and elemental ratios and DBE values similar to the biogenic SOA components. Both laboratory and field studies have confirmed the formation of organosulfates in biogenic SOA (Surratt et al., 2006; Surratt et al., 2008). Since these characteristic molecular formulas are positively correlated with SO_4^{2-}, they might be the biogenic precursors of organosulfates.

Table 4.6 The 20 molecular formulas in the highlighted region shown in Figure 4.28 B with the highest RAs in sample 0721.

Molecular Formula	Theoretical m/z	RA in 0721 (%)	DBE	O/C	H/C
$C_{24}H_{43}NO_{10}S$	536.2535	58.16	4	0.42	1.79
$C_{26}H_{47}NO_{10}S$	564.2848	46.45	4	0.38	1.81
$C_{26}H_{45}NO_{10}S$	562.2691	39.76	5	0.38	1.73
$C_{10}H_{17}NO_8$	278.0881	21.90	3	0.80	1.70
$C_{11}H_{18}O_7$	261.0980	19.11	3	0.64	1.64
$C_{11}H_{18}O_6$	245.1031	16.40	3	0.55	1.64
$C_{10}H_{16}O_5$	215.0925	12.37	3	0.50	1.60
$C_{10}H_{14}O_6$	229.0718	11.91	4	0.60	1.40
$C_{22}H_{39}NO_{10}S$	508.2222	11.54	4	0.45	1.77
$C_{25}H_{45}NO_{10}S$	550.2691	11.39	4	0.40	1.80
$C_{10}H_{16}O_4$	199.0976	11.08	3	0.40	1.60
$C_{10}H_{15}NO_8$	276.0725	9.72	4	0.80	1.50
$C_{28}H_{51}NO_{10}S$	592.3161	9.64	4	0.36	1.82
$C_{11}H_{16}O_7$	259.0823	9.26	4	0.64	1.45
$C_{11}H_{17}NO_9$	306.0831	8.87	4	0.82	1.55
$C_{11}H_{16}O_6$	243.0874	8.43	4	0.55	1.45
$C_{10}H_{14}O_5$	213.0768	8.26	4	0.50	1.40
$C_{10}H_{15}NO_9$	292.0674	8.09	4	0.90	1.50
$C_{10}H_{17}NO_9S$	326.0551	8.05	3	0.90	1.70
$C_{11}H_{17}NO_8$	290.0881	7.91	4	0.73	1.55

4.3.3.4.3 Samples and formulas with high DBE

A group of molecular formulas located in quadrant III were observed to be unique in sample 0720 (Figure 4.28 C); some of the molecular formulas are present in other plots but have quite small symbol sizes. This is also true for the samples 0625 and 0709 in addition to 0720. These are the three samples with the highest DBE_w values (Figure 4.4) among all of the WSOC samples. The dates located far away from the origin in quadrant III of the PCA loading plot are those with relatively low temperature, weak UV radiation, and low relative humidity. Thus the samples are expected to have a low degree of photochemical aging and weak aqueous phase processing relative to the samples in quadrants II and IV. Thus the molecular formulas of these three samples are expected to be less oxidized and may contain a high number of double bonds or aromatic rings.

The group of molecular formulas unique to quadrant III of Figure 4.28 C occupies the space defined by the PC1 values of approximately -0.025 to -0.015 and PC2 values of approximately -0.025 to -0.016 (highlighted area in Figure 4.28 C). There are 276 molecular formulas in this region. The molecular formulas in this region have quite low RAs. The highlighted formulas with the 20 highest RAs in 0720 are shown in Table 4.6. The RAs range from 2 - 5%. Almost all of the molecular formulas contain more than 20 carbon atoms with molecular weights greater than 400 Da. The DBE values of the 20 formulas are quite high and almost all of them are greater than 10. In fact, all of the 276 formulas in the highlighted region have very high DBE values. The average DBE for these molecular formulas is 12.2. The average O/C and H/C ratios of the 276 formulas are 0.39 and 1.08, respectively. The low O/C and H/C ratios of the formulas in this group compared to the average values of all of the formulas in the samples (Figure 4.4) indicate that these organic species are considerably less oxidized or less processed in the atmosphere without strong UV radiation or high water vapor content.

Table 4.7 The 20 molecular formulas in the highlighted region shown in Figure 4.28 C with the highest RAs in sample 0720.

Molecular Formula	Theoretical m/z	RA in 0720 (%)	DBE	O/C	H/C
$C_{31}H_{39}NO_{10}$	584.2501	4.27	13	0.32	1.26
$C_{26}H_{34}O_{12}$	537.1978	3.35	10	0.46	1.31
$C_{25}H_{30}O_{11}$	505.1715	2.78	11	0.44	1.20
$C_{26}H_{32}O_{11}$	519.1872	2.76	11	0.42	1.23
$C_{24}H_{28}O_{10}$	475.1610	2.71	11	0.42	1.17
$C_{25}H_{30}O_{10}$	489.1766	2.53	11	0.40	1.20
$C_{24}H_{26}O_{11}$	489.1402	2.48	12	0.46	1.08
$C_{28}H_{28}O_{12}$	555.1508	2.45	15	0.43	1.00
$C_{21}H_{22}O_9$	417.1191	2.41	11	0.43	1.05
$C_{24}H_{24}O_{11}$	487.1246	2.37	13	0.46	1.00
$C_{22}H_{24}O_9$	431.1348	2.37	11	0.41	1.09
$C_{25}H_{30}O_{12}$	521.1665	2.34	11	0.48	1.20
$C_{23}H_{26}O_{12}$	493.1352	2.34	11	0.52	1.13
$C_{24}H_{28}O_{12}$	507.1508	2.31	11	0.50	1.17
$C_{23}H_{24}O_{10}$	459.1297	2.29	12	0.43	1.04
$C_{19}H_{26}O_5$	333.1707	2.18	7	0.26	1.37
$C_{23}H_{24}O_9$	443.1348	2.12	12	0.39	1.04
$C_{20}H_{20}O_9$	403.1035	2.04	11	0.45	1.00
$C_{22}H_{24}O_{12}$	479.1195	2.02	11	0.55	1.09

4.3.3.4.4 Other correlated variables

Hennigan et al. (2008c) observed a strong correlation between WSOC and NO_3^- in Mexico City and explained that both were secondary products produced in the atmosphere. In the gas phase, OH radicals oxidize NO_2 to form NO_3^-. In this study, WSOC and NO_3^- are also positively correlated. Shown in the PCA loading plot in Figure 4.26, both WSOC and NO_3^- locate in quadrant IV, though the loadings of NO_3^- are quite small and did not contribute much to the variation of the WSOC samples. Several studies by Hennigan and colleagues (Hennigan et al., 2008a; Hennigan et al., 2008b; Hennigan et al., 2009) show the positive effect of water vapor or RH on the partitioning of WSOC to the particle phase when the RH is higher than 70%. In this study, we characterized the particle phase concentration of WSOC instead of the fraction of WSOC in particle phase; also the RH during the sampling dates was rarely higher than 70%. Thus a correlation between RH and WSOC is not observed. Wind speed, diffuse UV radiation, pressure and ozone concentration is also included in the PCA analysis. Since these variables are all 24 hour averaged values, they did not show much variation among the dates or much contribution to the WSOC molecular composition. So they are not discussed in detail. Further studies are necessary to examine their effects on aerosol compositions.

In summary, multiple environmental factors affect the aerosol-derived WSOC molecular composition and these factors can occur simultaneously. As discussed above, no single parameter could effectively explain the grouping of the WSOC samples. The composition of the samples is affected by the measured parameters and even additional ones not measured in this study. In the PCA score plot (Figure 4.25), the molecular composition of 0716 is highly affected by the temperature and UV irradiance. However other environmental parameters also affect the molecular composition of WSOC sample 0716, though their contribution might be smaller than temperature and UV radiation. Similarly the RH and sulfate do not solely influence the composition 0721, the other ambient variables and air mass histories are expected to also influence the molecular composition, but to a lesser extent. For the samples near the origin such as 0718 (Figure 4.25), all of the environmental variables play important roles in their composition, although their roles are comparable with each other instead of any one apparently higher than others. Atmospheric processing of the WSOC molecular composition is quite complex and multiple parameters are involved. The multivariate statistical analyses reveal the

important factors affecting the WSOC organic composition and the corresponding formulas were explored.

Chapter 5 Cloud water molecular composition

(This chapter is from the results and discussion of a publication by Zhao, Y., Hallar, A. G., and Mazzoleni, L. R. titled "Atmospheric organic matter in clouds: exact masses and molecular formula identification using ultrahigh-resolution FT-ICR mass spectrometry" published in Atmospheric Chemistry and Physics, volume 13, 12343-12362 in 2013. The articles published by Copernicus Publications (the publisher of Atmospheric Chemistry and Physics) are licensed under the Creative Commons Attribution 3.0 License together with an author copyright.)

5.1 Mass spectra and molecular formula assignments

Molecular formulas of the form, $C_cH_hN_nO_oS_s$, were assigned to the negative ion ultrahigh resolution FT-ICR mass spectra of AOM extracted from cloud water samples. Overall, 82 - 83% of the total ion current of the CW1 and CW2 sample mass spectra were assigned. After blank subtraction and quality assurance filtering approximately 3000 monoisotopic molecular formulas were unambiguously identified for CW1 and CW2 (Table 5.1). Assignments of corresponding polyisotopic molecular formulas containing ^{13}C were observed for 71% and 69% of the formulas assigned to CW1 and CW2. Likewise, polyisotopic molecular formulas with ^{34}S were observed for 54% and 50% of the S-containing molecular formulas assigned to CW1 and CW2. Detection of polyisotopic anions is more limited than detection of monoisotopic anions due to the lower RA which reflects the natural abundances of ^{13}C at 1% and ^{34}S at 4%. The high number frequency of corresponding polyisotopic molecular formulas with ^{13}C at exactly 1.0034 Da higher than the monoisotopic molecular formulas confirms the molecular formula assignment and the singly charged state of the detected anions. No evidence of multiply charged species was observed.

Reconstructed mass spectra of the monoisotopic molecular formulas assigned to the cloud water samples after blank subtraction are shown in Figure 5.1. In both of the cloud water mass spectra, the anions occupy a mass range of 100 – 700 Da, covering both the range of previously identified cloud water carbonyl species (Limbeck and Puxbaum, 2000; Marinoni et al., 2004; van Pinxteren et al., 2005; Samy et al., 2010; Charbouillot et al., 2012) and the range of water-soluble macromolecular compounds also known as "HULIS" (Feng and Möller, 2004). Macromolecular compounds were also observed in fog water by several studies (Krivacsy et al.,

2000; Herckes et al., 2002a; Herckes et al., 2002b; Cappiello et al., 2003). Similar to the ambient

aerosol collected at this site, the highest RA anions were observed with m/z from 200 to 400

(Mazzoleni et al., 2012). Consistent with the ultrahigh resolution mass spectra of other

atmospheric samples like aerosols (Wozniak et al., 2008; Schmitt-Kopplin et al., 2010), fog water

(Mazzoleni et al., 2010; LeClair et al., 2012) and rainwater (Altieri et al., 2009b; Altieri et al.,

2012), the mass spectra for cloud water samples are quite complex with a high number of

isobaric anions. An example of the isobaric complexity with 14 molecular formulas within 0.25

Da is shown in an excerpt of the reconstructed CW1 mass spectrum (Figure 5.1B).

Table 5.1 Cloud water composition characteristics by elemental groups

		All	CHO	CHNO	CHOS	CHNOS
CW1	Number frequency	3140	840	1638	281	381
	O/C	0.62 ± 0.37	0.54 ± 0.35	0.72 ± 0.34	0.45 ± 0.38	0.45 ± 0.34
	O/C$_w$	0.63 ± 0.03	0.47 ± 0.04	0.77 ± 0.04	0.43 ± 0.09	0.44 ± 0.04
	H/C	1.46 ± 0.36	1.42 ± 0.36	1.56 ± 0.29	1.37 ± 0.45	1.21 ± 0.41
	H/C$_w$	1.53 ± 0.07	1.48 ± 0.11	1.61 ±0.10	1.41 ± 0.27	1.17 ± 0.10
	DBE	6.30 ± 3.37	6.36 ± 3.56	5.58 ± 2.62	6.76 ± 4.21	8.98 ± 3.72
	DBE$_w$	5.19 ± 0.19	5.50 ± 0.35	4.63 ± 0.22	5.56 ± 0.92	8.64 ± 0.53
	OM/OC	2.08 ± 0.56	1.83 ± 0.47	2.25 ± 0.54	1.89 ± 0.56	2.02 ± 0.53
	OM/OC$_w$	2.20 ± 0.06	1.79 ± 0.09	2.44 ± 0.08	1.98 ± 0.23	2.14 ± 0.11
CW2	Number frequency	2724	712	1431	221	360
	O/C	0.61 ± 0.34	0.51 ± 0.28	0.73 ± 0.31	0.43 ± 0.37	0.43 ± 0.35
	O/C$_w$	0.63 ± 0.02	0.47 ± 0.03	0.74 ± 0.04	0.42 ± 0.05	0.43 ± 0.04
	H/C	1.46 ± 0.35	1.41 ± 0.34	1.56 ± 0.30	1.40 ± 0.44	1.21 ± 0.37
	H/C$_w$	1.53 ± 0.06	1.46 ± 0.11	1.61 ± 0.09	1.41 ± 0.17	1.19 ± 0.11
	DBE	6.29 ± 3.29	6.35 ± 3.18	5.60 ± 2.84	6.35 ± 3.89	8.89 ± 3.42
	DBE$_w$	5.25 ± 0.18	5.64 ± 0.32	4.75 ± 0.22	5.52 ± 0.55	8.00 ± 0.52
	OM/OC	2.06 ± 0.52	1.79 ± 0.37	2.25 ± 0.50	1.86 ± 0.54	1.99 ± 0.55
	OM/OC$_w$	2.16 ± 0.05	1.78 ± 0.08	2.35 ± 0.08	1.96 ± 0.14	2.18 ± 0.11

Note: Average values (O/C, H/C, DBE and OM/OC) are the mean and standard deviation of each data subset. Relative abundance weighted values (O/C$_w$, H/C$_w$, DBE$_w$ and OM/OC$_w$) were determined using equations 2-4 to 2-7 and have propagated standard error using the standard deviations for each subset.
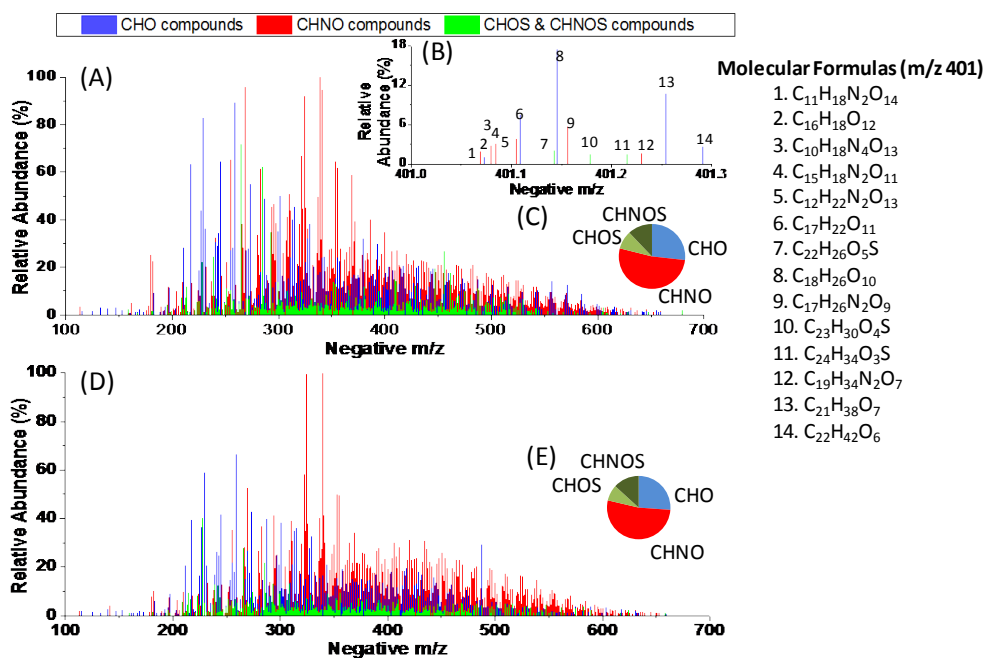
Figure 5.1. Reconstructed mass spectra of the identified monoisotopic molecular formulas for CW1 (A and B) and CW2 (D): an excerpt of the CW1 mass spectrum from m/z 401.00 to 401.30 is shown in (B) with a list of the numbered analytes and their identified molecular formulas. Pie charts indicate the number fraction of the identified monoisotopic molecular formula groups for CW1 (C) and CW2 (E).

200

Four subgroups of molecular formulas were determined based on their elemental composition, they include: compounds containing only C, H and O (CHO), compounds containing C, H, N and O (CHNO), compounds containing C, H, O and S (CHOS), and compounds containing C, H, N, O and S (CHNOS). The CHO compounds (N = 840 and 712) represented ~ 26% of the total number of assigned molecular formulas from the CW1 and CW2 sample mass spectra (Figure 5.1C and 5.1E). In general, a clear pattern of mass differences of the CHO compounds can be seen in both of the mass spectra (blue peaks in Figure 5.1) with spacing of the high intense anions of 14 Da. This has been commonly observed in natural organic matter samples (Stenson et al., 2003; Koch et al., 2007; Mazzoleni et al., 2010). This pattern is a likely consequence of the naturally occurring CH_2 "homologous series" or formula extensions in organic matter (Hughey et al., 2001). All of the identified CHO molecular formulas belong to a CH_2 homologous series (Figure 5.2). The series range from 2 to 26 molecular formulas in length with an average length of ~10. However, it should be noted that the addition of the $-CH_2$ to a molecular formula does not necessary lengthen the compounds' carbon backbone, but may appear in any aspect of the compounds' molecular structure. Moreover, each molecular formula identified may represent several organic compounds with different chemical structures (Hertkorn et al., 2008). Therefore, the clouds have extreme complex organic composition considering the ~3000 molecular formulas identified in each sample. In previous studies, the CHO compounds typically represented the highest number fraction of the overall compounds (Wozniak et al., 2008; Schmitt-Kopplin et al., 2010; Altieri et al., 2012; Mazzoleni et al., 2012). However in this study, CHNO compounds (N = 1638 and 1431) are the most frequently observed compounds in both cloud water samples and represented > 50% of the total number of molecular formulas (Figure 5.1C and 5.1E). Large amount of organic nitrogen was also observed in fog sample by previous study (Herckes et al., 2007).From the mass spectra, the CHNO compounds (red peaks in Figure 5.1)with the highest relative abundances are in the mass range of m/z 200 – 400 and the highest density is in the mass range of 300< m/z < 600. Consistent with the CHO compounds, a wide number of CH_2 homologous series were observed in the cloud water CHNO compounds (Figure 5.2). The S-containing compounds (CHOS and CHNOS) represented ~ 21% of the total number of molecular formulas (Figure 5.1C and 5.1E). 281 CHOS and 381 CHNOS molecular formulas were identified in CW1 and 221 CHOS and 360 CHNOS formulas were identified in CW2. Typically the S-containing compounds (green peaks in Figure 5.1) are in the mass range of 200< m/z < 600.

Often the relative abundance of these compounds are lower than the compounds in other groups. The CH$_2$ homologous series of the S-containing compounds are less clear in the reconstructed mass spectra. However, they are clearly shown in the Kendrick mass defect plots (Figure 5.2). The composition and molecular properties for each of the subgroups will be discussed further in the following sections.
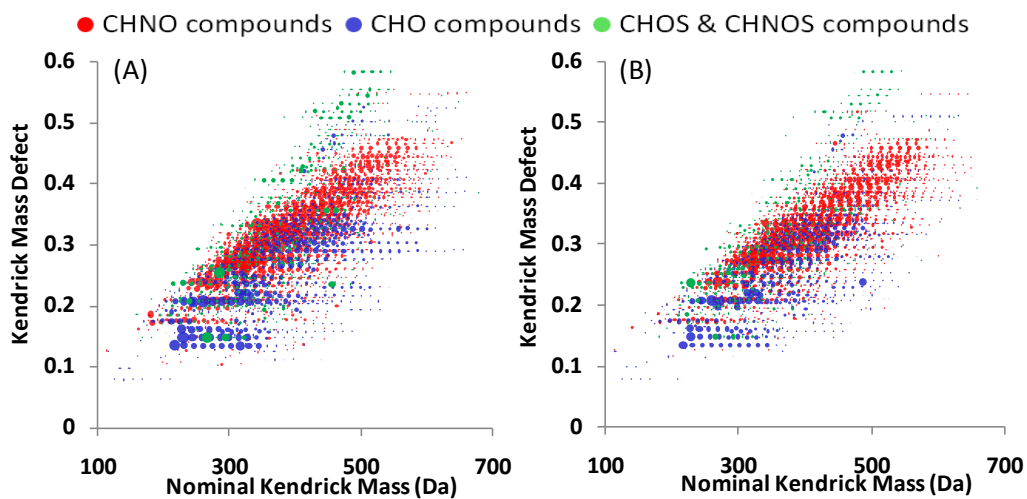
Figure 5.2. The Kendrick mass defect vs. nominal Kendrick mass for the monoisotopic molecular formulas identified in CW1 (A) and CW2 (B) with scaled symbol sizes representing the relative abundance of the compounds.

5.2 Compounds containing only C, H and O

Oxygen-to-carbon ratios (O/C) and hydrogen-to-carbon ratios (H/C) are commonly used to describe the chemical properties of aerosol (Aiken et al., 2007; Aiken et al., 2008; Jimenez et al., 2009; Bateman et al., 2010; Kroll et al., 2011). In general, O/C represents the degree of oxidation and H/C reflects the degree of saturation of the composition studied. The O/C and H/C elemental ratios of the individual CHO compounds of the cloud water samples are plotted in Figure 5.3A (for CW1) and Figure 5.4A (for CW2). The van Krevelen diagram (O/C vs. H/C) indicates the oxidation, hydration, hydrogenation and alkylation relationships between the observed molecular formulas (Kim et al., 2003; Wu et al., 2004; Heald et al., 2010). In Figure 5.3A and 5.4A, we plotted simultaneously the analyte isoabundance and individual points for each of the CHO molecular formulas. This permits visualization of the CH_2 homologous series of molecular formulas in lines pointing to the upper left corner (H/C = 2.0) of the plot and their significance in terms of isoabundance. In both samples, a majority of the CHO compounds have O/C ratios < 1.0 and H/C ratio from 0.5 to 2.0. The highest intensity molecular formulas are in the vicinity of O/C = 0.5 and H/C = 1.5. Interestingly, there are several compounds with O/C ratios > 1.0 in both samples and many of these have 18-22 oxygen atoms and molecular weights of 400-500 Da. These highly oxidized high molecular weight compounds may be the products of aqueous phase reactions (Lee et al., 2011). In addition, a few CHO compounds are located in the lower left corner of the van Krevelen diagram (O/C < 0.3, H/C < 1). These appear to be aromatic compounds with high DBE values. Examples of the molecular formulas in this region include the homologous series of $C_{20}H_8O_4(CH_2)_{1-7}$ and $C_{19}H_6O_5(CH_2)_{1-6}$. The average O/C ratios for the CW1 and CW2 CHO compounds are 0.54 (± 0.35) and 0.51 (± 0.28) (Table 5.1). Likewise, the average H/C ratios are 1.42 (± 0.36) and 1.41 (± 0.34). Similarly, Altieri et al. (2009b) observed an average O/C ratio of 0.7 (± 0.5) and an average H/C ratio of 1.5 (± 0.4) for the CHO compounds in ambient rainwater. Thus, the CHO compounds reported here are less oxidized and somewhat less saturated, likely reflecting seasonal differences and cloud processing time differences. Other reported elemental ratios for CHO compounds in ambient aerosol water-soluble organic compounds, include: 0.47 (± 0.2) and 1.42 (± 0.3) (Mazzoleni et al., 2012); 0.6 and 1.69 (Rincon et al., 2012); and 0.40 (± 0.21) and 1.29 (± 0.35) (Lin et al., 2012a). Note, the solid phase extraction step used in this study may not retain well some of the low molecular weight CHO compounds like formic acid, acetic acid or oxalic acid, which usually have high O/C ratios (>1).

Thus, the O/C ratios of the clouds may be somewhat higher than reported here. Overall, the O/C and H/C ratios of CHO compounds in the cloud water samples are within the wide range of the previously reported values, but are closer to those of aerosol water-soluble organic carbon than rainwater. This is probably related to the different extents of aqueous processing and temperature in supercooled and precipitating clouds.
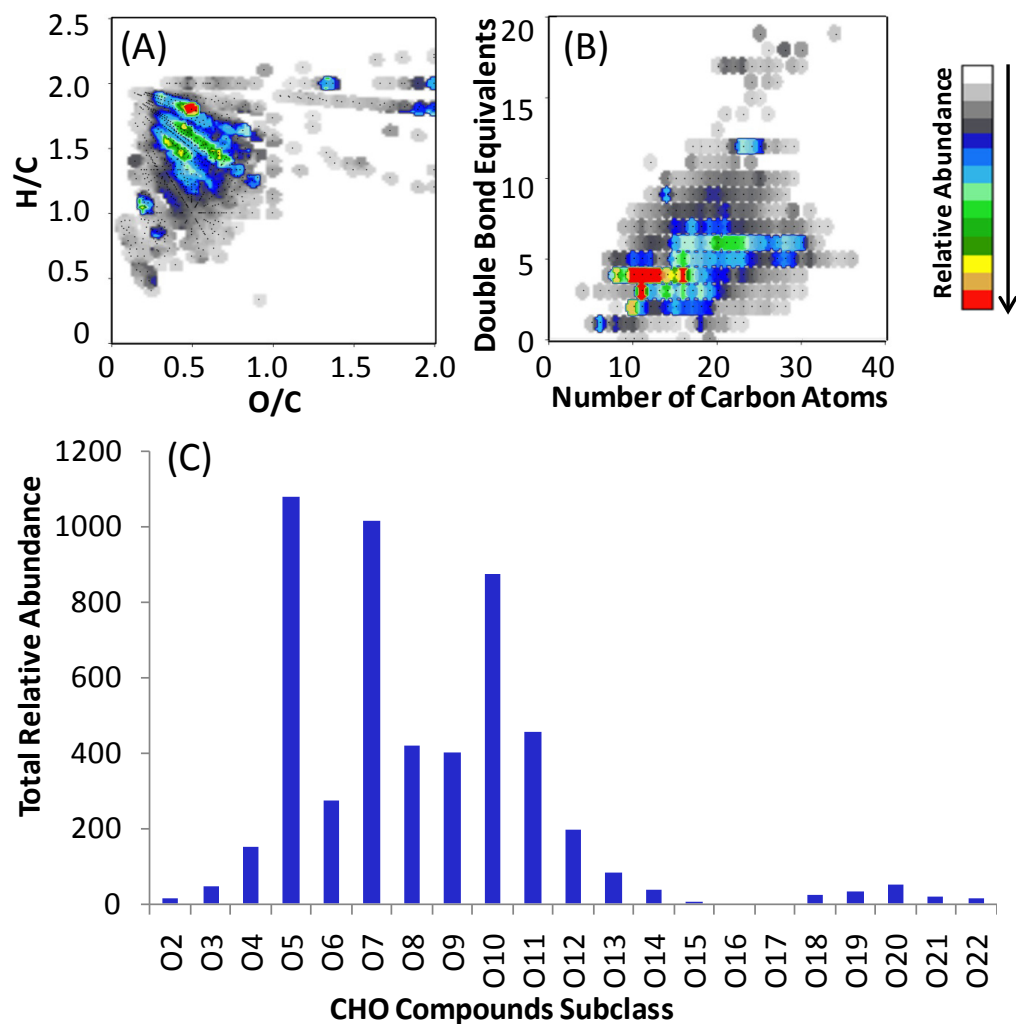
Figure 5.3. Isoabundance van Krevelen diagram for CW1 CHO compounds (A). Isoabundanceplot of double bond equivalents (DBE) vs. number of carbon atoms in molecular formulas for CW1 CHO compounds (B). Total relative abundance (RA) of each subclass for CW1 CHO compounds(C).
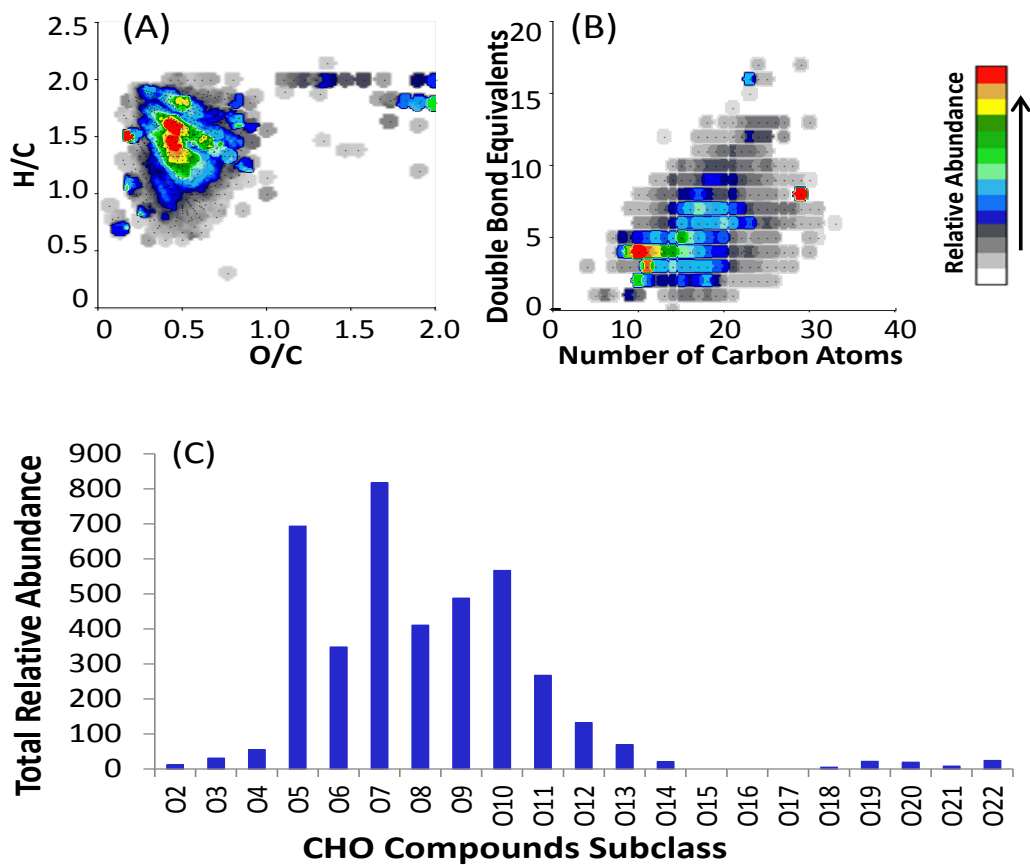
Figure 5.4. Isoabundance van Krevelen diagram for CW2 CHO compounds (A). Isoabndance plot of double bond equivalents (DBE) vs. number of carbon atoms in molecular formulas for CW2 CHO compounds (B). Total relative abundance (RA) of each subclass for CW2 CHO compounds (C).

Another measure of molecular unsaturation is the DBE, which indicates the number of double bonds and rings in a molecular structure. A wide range of DBE values (0-19) were observed among the cloud water CHO compounds with a clear trend of increasing DBE values with increasing carbon content (C_3 - C_{36}) (Figure 5.3B). In both samples, the majority (> 90%) of the compounds has DBE values ≤ 10 and the high relative abundance compounds have DBE values within the range of 2-7. Overall, the average DBE values for CW1 and CW2 CHO compounds are 6.4 (± 3.6) and 6.3 (± 3.2), similar to that of CHO compounds from other studies. Average DBE values for CHO compounds of 5.6 and 6.4 were reported by Lin et al. (2012a) and Mazzoleni et al. (2012) for ambient aerosol; Mazzoleni et al. (2010) reported an average DBE value for CHO compounds of 5.3 for low molecular weight fog water AOM; and Putman et al. (2012) reported an average DBE value for CHO compounds of 7.0 for chamber generated α-pinene ozonolysis SOA. Interestingly, several CHO compounds with high DBE values (i.e., 14-19) were observed and are located in the aromatic region (O/C < 0.3 and H/C < 1.0) of the van Krevelen diagram (Figure 5.3A). Generally, these aromatic compounds are within the mass range of 300-400 Da and have < 6 O atoms. The relative abundances of the compounds with respect to carbon are also illustrated in Figure 5.3B. The higher intensity compounds are compounds with C_5 - C_{25}. Among these compounds, the highest relative abundances are ~ C_{10}, followed by ~ C_{15}. There is also a relatively high intensity spot ~ C_{20} in the CW1 sample. These high relative abundance compounds at C_{10}, C_{15}, and C_{20} indicate terpene characteristics in the cloud water CHO compounds (Claeys et al., 2013).

To evaluate the oxygen content, the cloud water CHO compounds were separated into subclasses based on the number of oxygen atoms ($O_\#$) in the molecular formulas, labeled as O_1, O_2 and so on. The total relative abundance for each of the subclasses are shown in Figure 5.3C. Overall, cloud water CHO compounds range from O_2 to O_{22}, with the majority of compounds in the range of O_2 to O_{15}. These subclasses with variable oxygen numbers may reflect multiple and or various functional groups like hydroxyl, ether, peroxyl, carbonyl, carboxyl or ester. Interestingly, the highest relative abundances were observed for the O_5, O_7 and O_{10} subclasses. The O_{10} subclass contains the highest number of CHO compounds (~ 130 molecular formulas) while the highest relative abundance CHO compounds were observed in the O_5 and O_7 subclasses. Examples of the high relative abundance CHO compounds detected in both samples include: $C_{10}H_{18}O_5$, $C_{11}H_{18}O_5$, $C_{10}H_{14}O_7$, $C_{11}H_{16}O_7$, $C_{12}H_{18}O_7$, $C_{13}H_{20}O_7$, and $C_{14}H_{22}O_7$. These high

relative abundance compounds have DBE values of 2, 3 and 4. Interestingly, no formulas in the O_{16} and O_{17} classes were observed, but there are several compounds in the higher subclasses (O_{18} - O_{22}). These high oxygen numbers correspond to compounds with O/C ratios > 1.0 in the van Krevelen diagram.

The reconstructed mass spectra of the cloud water CHO compounds have similar characteristics to laboratory generated SOA samples. Three groups of high abundance anions were observed, they include: m/z 200-350; m/z 350-500; and m/z 500-650 (Figure 5.5).Similar clusters of high relative abundance anions (sometimes referred to as "oligomers" or monoterpene accretion products in SOA samples) have been observed in SOA samples and have decreasing ion intensities with increasing molecular weight (Reinhardt et al., 2007; Bateman et al., 2011; Kundu et al., 2012; Putman et al., 2012). Also the average O/C and H/C ratios of the cloud water CHO compounds are consistent with those determined for biogenic SOA samples (Bateman et al., 2009, 2010; Putman et al., 2012). In addition, there are several "hot spots" at C_{10}, C_{15} and C_{20} in the isoabundance DBE plots (Figure 5.3B) representing the high relative abundance compounds. The DBE values and the carbon numbers of the high relative abundance compounds reflect monoterpene and sesquiterpene characteristics (Putman et al., 2012). To further explore the SOA composition similarities, the cloud water CHO molecular formulas were compared to the CHO molecular formulas assigned to several laboratory generated biogenic SOA mass spectra. The biogenic SOA samples used in the comparison include: α-pinene ozonolysis SOA, β-pinene ozonolysis SOA, d-limonene ozonolysis SOA, and β-caryophyllene ozonolysis SOA, all of which were generated under dark conditions without an OH scavenger (the LANLSOA samples discussed in Chapter 3). Of the CW1 and CW2 CHO molecular formulas, 58 and 64%, respectively were identical to those in the biogenic SOA samples (Figure 5.5). Furthermore, most of these common molecular formulas are high intensity anions in the cloud water samples. Overall, the three groups of anions assigned as CHO compounds in the reconstructed mass spectra are less distinct than those in the SOA mass spectra. This reflects the more complex conditions for ambient aerosols and clouds than laboratory experimental conditions. Some of the differences include: the presence of biogenic SOA from unevaluated precursors; aqueous phase reactions which alter the composition; and compounds from other sources, etc. Despite the differences between ambient conditions and those of laboratory experiments, we observed highly similar characteristics in the CHO molecular formulas of cloud water and biogenic SOA.

This suggests that a high percentage of the cloud condensation nuclei in the Rocky Mountain region during the cloud sampling study contain biogenic SOA components. Similarly, Sun et al. (2011) using aerosol mass spectrometry found ~90 % of the aerosol water-soluble organic carbon at a rural site (Centreville, Alabama) in the summer are biogenic related. The fraction decreased in the winter, but still comprised ~50% of the total water-soluble organic carbon.

In addition to biogenic SOA, the CHO composition of the clouds appears to be affected by residential wood combustion. Guaiacols and syringols (aka methoxyphenols) are commonly found in biomass combustion emissions and are from the pyrolysis of wood lignin (Hawthorne et al., 1992; Mazzoleni et al., 2007). Methoxyphenols have been shown by Sun et al. (2010) to contribute to aqueous SOA and have been observed in ambient foggy atmospheres (Sagebiel and Seiber, 1993). Two of the molecular formulas considered to be markers of syringol and guaiacol aqueous phase reactions in the Sun et al. (2010) experiments were observed in the cloud water samples. They are: $C_{16}H_{18}O_6$ (RA = 7% in both samples) and $C_{14}H_{14}O_4$ (RA = 2.7% in CW1 and 4% in CW2). As described, molecular formulas assigned to exact mass measurements using ultrahigh resolution FT-ICR MS provide valuable insights regarding the chemical structure (e.g., elemental ratios and DBE). A large fraction of the high relative abundance O_5 (50 out of 79 formulas in CW1, 39 out of 63 formulas in CW2) and O_7 (68 out of 74 formulas in CW1, 74 out of 77 formulas in CW2) CHO compounds have DBE values ≥ 4. This combination is consistent with methoxyphenol derived structures since there is both enough oxygen for the functional groups of $-OCH_3$ and $-OH$ and enough DBE for the benzene ring. Samy et al. (2010) analyzed aerosol and supercooled cloud water samples collected at the SPL. They found abundant concentrations of methoxyphenols in both the aerosols and the cloud samples, which suggested a strong biomass burning chemical signature. The emissions likely arise from nearby residential wood combustion in the Yampa valley to the west of the SPL. In this study, the back trajectory analysis indicates the air was coming from west and northwest during the sampling periods (Figure 2.6 and 2.7) thus it was likely influenced by residential activities in the Yampa valley. Similarly Collett et al. (2008) also reported fog processing of carbonaceous particles from wood smoke. Residential wood combustion may contribute to the observed biogenic SOA in the cloud samples. It is known that terpenoid emissions from living plants are highly correlated to ambient temperature (Kesselmeier and Staudt, 1999), thus biogenic SOA is generally not expected to be significant during the winter time periods. However, trees have a large storage reservoir of the

monoterpenes compared to their emission rates (Lerdau et al., 1994; Amin et al., 2012). Thus, the stored monoterpenes may be released during residential wood combustion for SOA formation.

5.3 Compounds containing only C, H, N and O

A wide range of elemental ratios was observed for the CHNO compounds (Figure 5.6A). In contrast to the elemental ratios of the CHO compounds where a majority of the compounds have O/C < 1.0, many of the CHNO compounds have O/C > 1.0 and they have some of the highest relative abundances of the CHNO group. Most of these CHNO compounds with high O/C ratios also have high H/C ratios (>1.2), suggesting the compounds are both highly saturated and oxygenated. For example, the molecular formulas of the homologous series $C_5H_{10}N_2O_9(CH_2)_{1-3}$ with high relative abundances show these characteristics. Also, there are a large number (~ 50%) of CHNO molecular formulas with O/C < 0.7. In general, these compounds have a relatively low relative abundance and have a large range of H/C ratios (0.3 to 2.2). Overall, the average O/C ratios of the CHNO compounds in CW1 and CW2 are 0.72 (± 0.34) and 0.73 (± 0.31), respectively. The average H/C ratios for the CHNO compounds in CW1 and CW2 are 1.56 (± 0.29) and 1.56 (± 0.30) (Table 5.1). Previous studies of aerosol WSOC and fog AOM observed overall less oxygenated CHNO compounds, with average O/C ratios typically < 0.5 (Wozniak et al., 2008; Mazzoleni et al., 2010; Schmitt-Kopplin et al., 2010). However in the study of rainwater AOM by Altieri et al. (2009b), several CHNO compounds were identified with high O/C ratios (some even higher than 2.0). They reported an average O/C ratio of 1.6 (± 1.4) for the CHNO compounds. The presence of highly oxygenated CHNO compounds (O/C > 1.0) appears to be related to aqueous phase reactions in cloud and rainwater droplets.

Consistent with the H/C elemental ratios, the majority of the CHNO compounds have DBE values ≤ 10 (Figure 5.6B). Similar to the CHO compounds, the number of carbon atoms in the formulas range from $C_2 – C_{35}$ and the DBE values increase with the number of carbon atoms. The high intensity CHNO compounds with $C_2 – C_{20}$ have DBE values ≤ 6. The most abundant compounds are those with C_{11} and low DBE values (2 - 4). These values indicate that most of the CHNO compounds are relatively saturated, especially those with higher relative abundances. Only a small fraction (~3%) of the CHNO compounds has high DBE values from 11 to 19. As illustrated in Figure 5.6B, the high DBE CHNO compounds have low relative abundances. Overall, the average DBE values are 5.6 (± 2.6) and 5.6 (± 2.8) for CW1 and CW2 CHNO compounds. Therefore the cloud water CHNO compounds are mostly saturated with a few exceptions.

Cloud water CHNO compounds contain 1- 4 nitrogen atoms (N_1-N_4) per molecular formula. The compounds with one nitrogen (N_1) represented ~ 43% of the CHNO compounds in both samples, followed by compounds containing N_2 (~ 29%), N_3 (~ 21%) and N_4 (~ 7 %) in both samples. To further examine the N in the molecular formulas, CHNO subclasses were defined by the numbers of nitrogen and oxygen in the molecular formulas, they include: NO_{3-23}, N_2O_{3-18}, N_3O_{2-16} and N_4O_{1-19}. The total relative abundance for the compounds contained within the subclasses is shown in Figure 5.6C. Interestingly, the plot clearly indicates the high relative abundance CHNO compounds are in the subclasses of NO_{6-14}, N_2O_{8-16} and N_3O_{11-16}. The compounds in high relative abundance subclasses correspond to the hot spots in the upper right region of the isoabundance van Krevelen diagram (O/C > 0.7, H/C > 1.5; Figure 5.6A). All the high relative abundance CHNO subclasses are rich in oxygen with respect to nitrogen. Almost all of these classes show oxygen-to-nitrogen ratio (O/N) ≥ 3. Thus, the CHNO compounds appear to contain a large number of organonitrates (with the functional group of $-NO_3$), however CHNO compounds are not exclusively organonitrates since multifunctional compounds are anticipated. For example a study using an ultrahigh resolution tandem mass spectrometry technique, reported only 63% HNO_3 losses from CHNO and CHNOS compounds in fog water samples (LeClair et al., 2012).
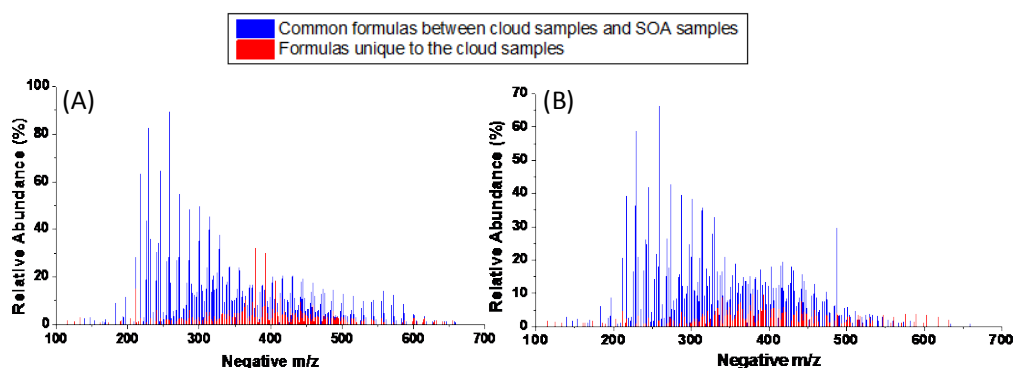
Figure 5.5. Reconstructed mass spectra of the CHO compounds in CW1 (A) and CW2 (B). Blue peaks show the common formulas between the cloud water samples and biogenic SOA samples (LANLSOS samples discussed in Chapter 3: combined data of ozonolysis of α-pinene, β-pinene, d-limonene and β-caryophyllene); pink peaks show the formulas unique to cloud water samples.

Organonitrates or nitrooxy-organosulfates have been observed in several AOM samples (Allen et al., 1994; Nielsen et al., 1998; Zhang and Anastasio, 2001; Garnes and Allen, 2002; Herckes et al., 2007; Reemtsma et al., 2006; Surratt et al., 2006; Altieri et al., 2008; Bruns et al., 2010; Fry et al., 2013). The organonitrates or nitrooxy-organosulfates observed in the clouds may be from dissolution of CCN aerosol, gas organic compounds containing nitro- or nitrate groups partitioned to the droplets (Lüttke and Levsen, 1997; Lüttke et al., 1997), or aqueous formation within the cloud droplets. Alkyl nitrates have relatively high vapor pressures, so they are not expected to readily partition to the particle phase. In a study by Nielson et al. (1998), only alkyl nitrates with > 18 carbon atoms can contribute to the particulate organic nitrogen mass at 20°C. However hydroxyl or carboxylic acid functional groups lower the vapor pressure (Arp and Goss, 2009), allowing the multifunctional organonitrates with less carbon atoms to partition to the particle phase. This coupled with the low temperatures during the sampling period (~ -10°C) enhances the significance of organonitrates in the condensed phase. Therefore it is highly possible that a considerable number of organonitrates are present in the supercooled cloud droplet samples. Similarly, nitrate esters were also observed in fog water (Herckes et al., 2007).

As mentioned ~50% of the CHNO compounds have O/C ratios < 0.7 (Figure 5.6A). These low O/C ratios are similar to those reported by Schmitt-Kopplin et al. (2010) for a biomass burning aerosol sample. In that study, almost all the identified nitrogen containing species had O/C ratios less than 0.7. The lower oxygen content suggests the presence of reduced nitrogen compounds. Wood combustion emissions are known to emit reduced nitrogen compounds. In a

214

study by Laskin et al. (2009) a substantial fraction of N-heterocyclic alkaloid compounds were observed in biomass burning aerosols. In another study by Bateman et al. (2010), an O/C ratio of 0.19 was reported for combustion of dried pine needles and sticks. Thus, the residential wood combustion in the vicinity of the SPL is a likely contributor to the reduced nitrogen containing species observed in the cloud water samples. However this does not exclude the possibility of aqueous phase reactions resulting in nitrogen containing species like those formed by aqueous phase reactions of glyoxal (Galloway et al., 2009; Shapiro et al., 2009) or methylglyoxal (De Haan et al., 2010; Sareen et al., 2010) with $(NH_4)_2SO_4/NH_4NO_3$. Reaction products, such as amine, imine or imidazole, provide another likely source of the reduced nitrogen compounds in the cloud water samples. Alkyl amines have also been detected in ambient aerosols (Gilardoni et al., 2009) and fog water (Zhang and Anastasio, 2001; Herckes et al., 2007). Regardless of their origin, reduced nitrogen functional groups can be oxidized in the cloud droplets to form oxidized functional groups as indicated by the high O/C ratios of products from aqueous phase reactions (Lim et al., 2010; Ervens et al., 2011).

Figure 5.6. Isoabundance van Krevelen diagram for CW1 CHNO compounds (A). Isoabundance plot of (DBE) vs. number of carbon atoms in molecular formulas for CW1 CHNO compounds (B). Total relative abundance (RA) of each subclass for CW1 CHNO compounds, the dashed lines separate the subclasses by number of nitrogen atoms contained in the formulas (C).

Figure 5.7. Isoabundance van Krevelen diagram for CW2 CHNO compounds (A). Isoabndance plot of double bond equivalents (DBE) vs. number of carbon atoms in molecular formulas for CW2 CHNO compounds (B). Total relative abundance (RA) of each subclass for CW2 CHNO compounds, the dashed lines separate the subclasses by number of nitrogen atoms contained in the molecular formulas (C).

5.4 Compounds containing only C, H, O and S and C, H, N, O and S

A wide range of elemental ratios is shown in Figure 5.8A for the CHOS compounds and Figure 5.10A for the CHNOS compounds. Similar to the CHO (Figure 5.3A) and CHNO compounds (Figure 5.6A), the elemental ratios align into several straight lines indicating CH2homologous series. However instead of one large group in the isoabundance van Krevelen diagrams, subgroups of the CHOS and CHNOS compounds in the isoabundance van Krevelen diagrams are depicted. Some obvious differences in the isoabundance van Krevelen diagrams of the CHOS and CHNOS compounds compared to those of the CHO and CHNO compounds were observed (Figure 5.8A and 5.10A). First, the presence of a high number of highly unsaturated CHOS and CHNOS compounds was observed. Second, we observed a wider overall range of elemental ratios with variable relative abundances. These compound properties are explored in more detail in the following paragraphs.

Four subgroups of CHOS molecular formulas with varied unsaturations or DBE values were observed (Figure 5.8). The CHOS compounds in subgroup 1 with H/C > 2 are highly saturated with DBE values of 0 or 1 (Figure 5.8B). These compounds belong to the subclasses of O_4S, O_5S and O_6S (Figure. 6C). Subgroup 2 CHOS compounds are less saturated and contain the subclasses of $O_{7-12}S$ with relatively low DBE values (mostly 3 or 4). The CHOS compounds in subgroup 3 are highly oxygenated with the subclasses of $O_{15-18}S$. They are less saturated than those in subgroup 2 with DBE values ranging from 7 to 9. Most of the subgroup 3 molecular formulas have relatively high molecular weight, typically > 400 Da. Thus from subgroups 1 to 3, the DBE values increase (from 0 to 9) with the increase of oxygen content in the molecular formulas (O_4 to O_{18}). The CHOS compounds in subgroup 4 are the most unsaturated of the CHOS compounds. The DBE values of these molecular formulas range from 8 to 19, suggesting aromatic moieties. Unlike the compounds in the previous subgroups, these have low O/C ratios (< 0.4). The subclasses of $O_{1-3}S$ in this group contain too little oxygen to contain organic sulfate functional groups. Overall around 1/3 of the detected CHOS compounds are in subgroup 4 which appears to be reduced sulfur species.
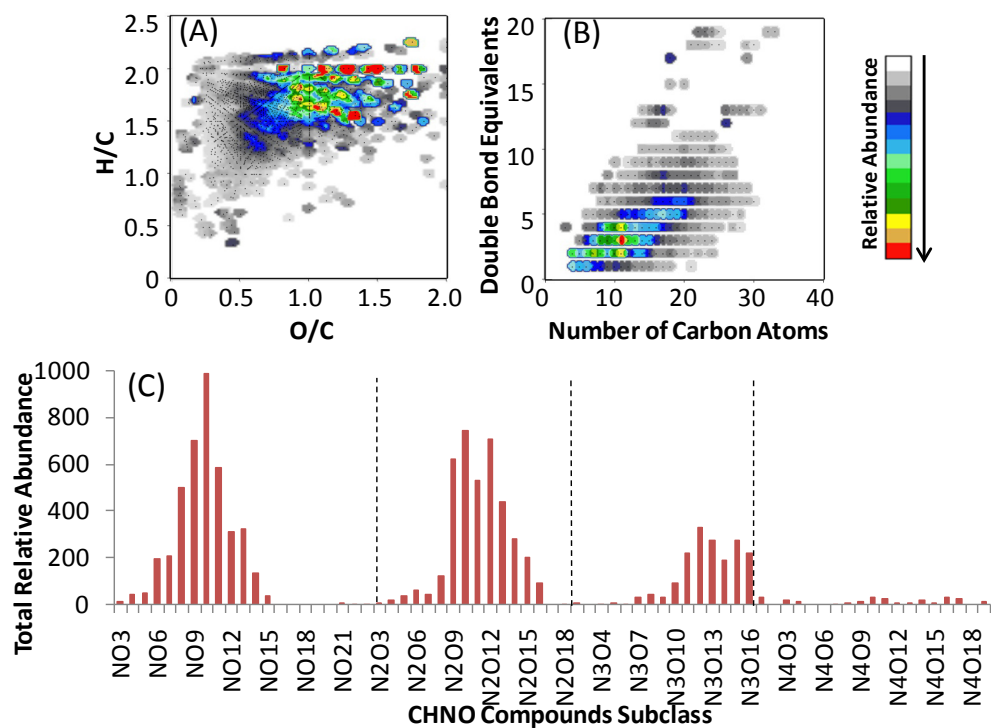
Figure 5.8. Isoabundance van Krevelen diagram for CW1 CHOS compounds (A). Isoabundance plot of (DBE) vs. number of carbon atoms in molecular formulas for CW1 CHOS compounds (B). Number of formulas identified in each of CW1 CHOS subclasses (C). The red circles and boxes indicate the approximate groupings (1 to 4) of the CW1 CHOS formulas.
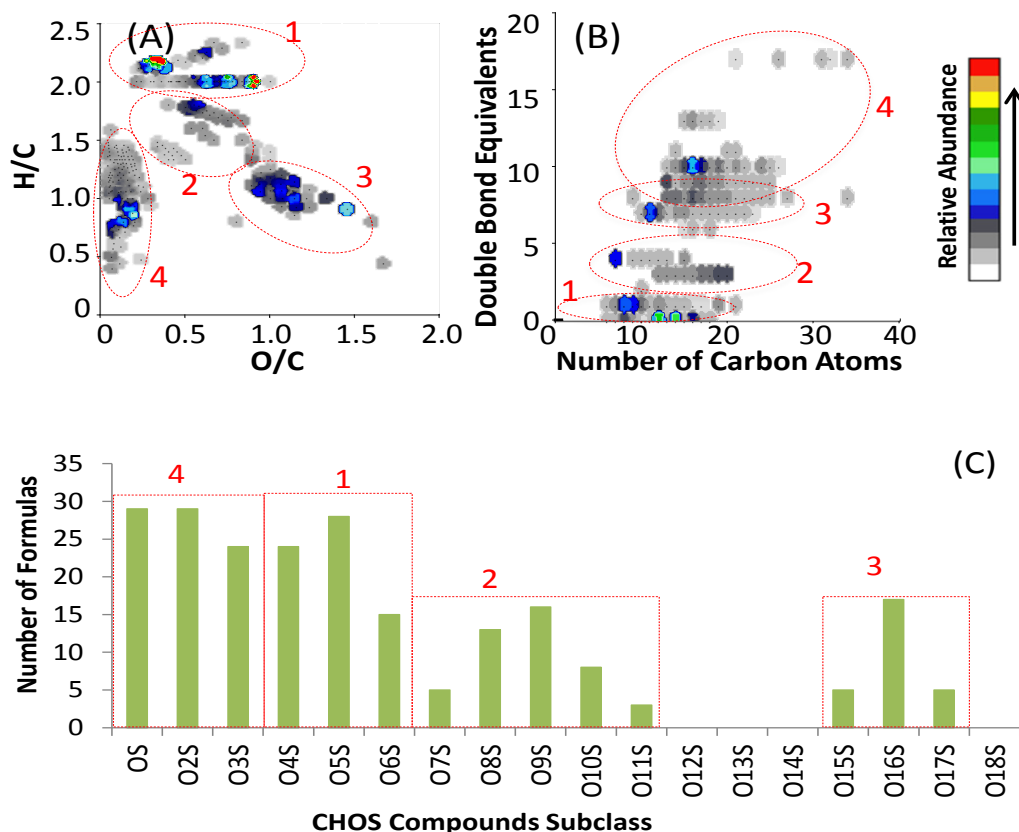
Figure 5.9. Isoabundance van Krevelen diagram for CW2 CHOS compounds (A). Isoabndance plot of double bond equivalents (DBE) vs. number of carbon atoms in molecular formulas for CW2 CHOS compounds (B). Number of molecular formulas identified in each of CW2 CHOS subclasses (C). The red circles and boxes indicate the approximate groupings (1 to 4) of the CW2 CHOS formulas.

Similarly, the CHNOS compounds have similar elemental ranges (Figure 5.10A). The number of CHNOS compounds in subgroups 1-3 are less than those of the CHOS compounds and the subgroup 4 CHNOS compounds are clearly dominant among the CHNOS compounds. Overall, a total of 75% of the CHNOS compounds are in subgroup 4 and 25% in subgroups 1-3. The CHNOS compounds in subgroups 1-3 have DBE values ranging from 1 to 8 (Figure 5.10B). These compounds contain 1 or 2 nitrogen atoms in the molecular formulas with a high number of oxygen atoms (> 7; subclasses $NO_{7-12,18}S$ and $N_2O_{12-16}S$; Figure 5.10C). Based on the oxygen numbers and DBE values, some of the molecular formulas may represent organonitrates, organosulfates or nitrooxy-organosulfates. In contrast, the subgroup 4 CHNOS compounds are highly aromatic with DBE values from 6 to 19. Most of the CHNOS compounds with 3 or 4 nitrogen atoms in their molecular formulas are in this group, including: $N_3O_{1-11}S$ and $N_4O_{1-5}S$.

Further, CHNOS compounds with 1 or 2 nitrogen atoms and a low number of oxygen atoms are in this group, including: $NO_{2-8}S$ and $N_2O_{2-9}S$. Therefore consistent with the cloud water CHOS compounds, the CHNOS compounds also include characteristics that are either highly oxygenated or highly aromatic. The aromatic S-containing compounds are dominant in number and thus greatly influence the average DBE value for these elemental classes.

Considering the adequate oxygen content and low DBE values (DBE for a sulfate functional group is 0), the S-containing compounds detected in subgroups 1-3 can be organosulfates or nitrooxy-organosulfates. Examples are like the several high relative abundance low MW CHOS compounds detected in both samples (200-300 Da): $C_{12}H_{26}O_4S$, $C_9H_{18}O_8S$, $C_{14}H_{30}O_4S$ and $C_8H_{16}O_6S$. LeClair et al. (2012) also detected these molecular formulas in fog water and used MS/MS to confirm the sulfate group in the structures Organosulfates have been detected commonly in atmospheric samples (Surratt et al., 2006; Gómez-González et al., 2008; Hatch et al., 2011). In fact, most of the atmospheric S-containing organic compounds reported in the literature are organosulfates or nitrooxy-organosulfates. Observations from both chamber simulations (Iinuma et al., 2007; Liggio et al., 2007; Surratt et al., 2007) and field measurements (Surratt et al., 2008; Stone et al., 2012; Pratt et al., 2013) indicate that organosulfates and nitrooxy-organosulfates can form in biogenic secondary organic aerosols. The identified organosulfates in biogenic SOA usually have carbon atom numbers similar to their biogenic precursors, e.g., ~5 or 10 (Surratt et al., 2008). In the cloud water samples, most of the CHOS and CHNOS compounds with less than 12 carbon atoms contain more than 4 oxygen atoms. Considering the cloud water biogenic SOA characterization identified from the cloud water CHO compounds, these organosulfates and nitrooxy-organosulfates compounds (S-containing compounds in subgroups 1-3) may have originated from biogenic precursors. Several high molecular weight organosulfate candidates (> 300 Da) with more carbon atoms (16 to 33) were also detected in the cloud water, including: $C_{17}H_{18}O_{16}S$, $C_{18}H_{38}O_6S$ and $C_{17}H_{20}O_{16}S$. The high molecular weight organosulfates have been observed previously in atmospheric aerosols (Romero and Oehme, 2005; Reemtsma et al., 2006). Mostly AOM organosulfates are secondary reaction products (Hatch et al., 2011). Two formation mechanisms were proposed, including: esterification of hydroxyl groups with sulfuric acid (Liggio et al., 2005; Surratt et al., 2007) and acid catalyzed reactions of epoxides (Minerath and Elrod, 2009). Both of the pathways require sulfuric acid anions (Surratt et al., 2007; Minerath and Elrod, 2009). Considering the coal fired

power plants located to the west of the SPL area which are emission sources of SO_2 and the aqueous phase processing that may take place within cloud droplets, formation of organosulfates is likely.
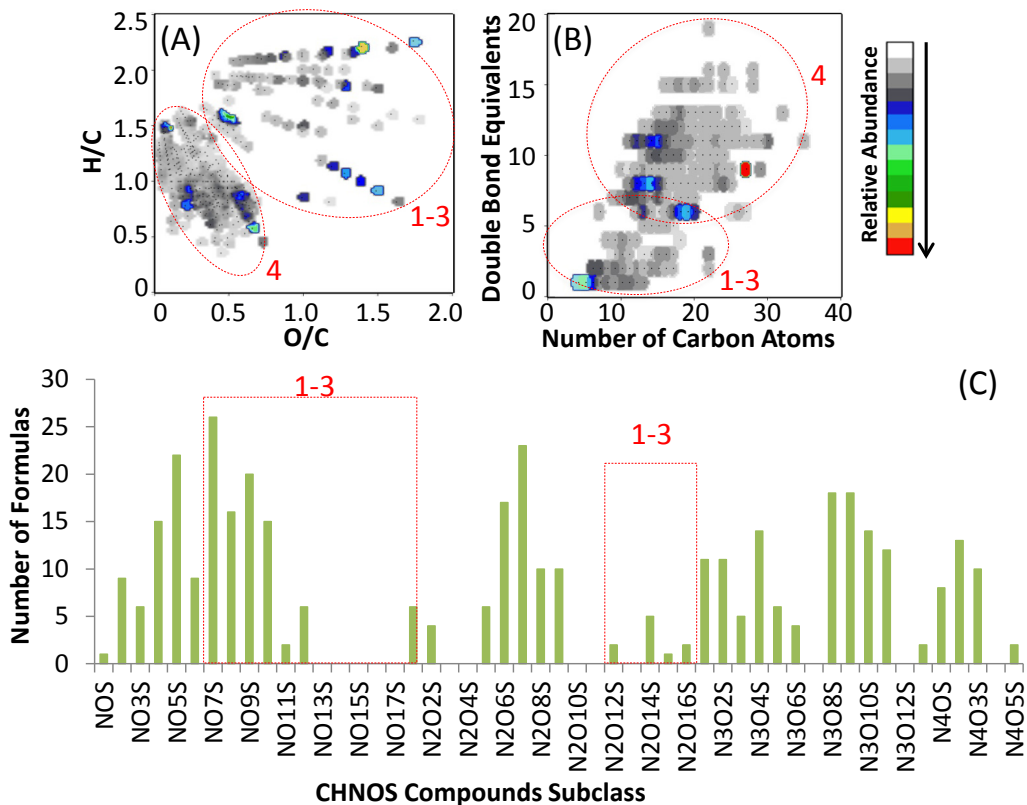
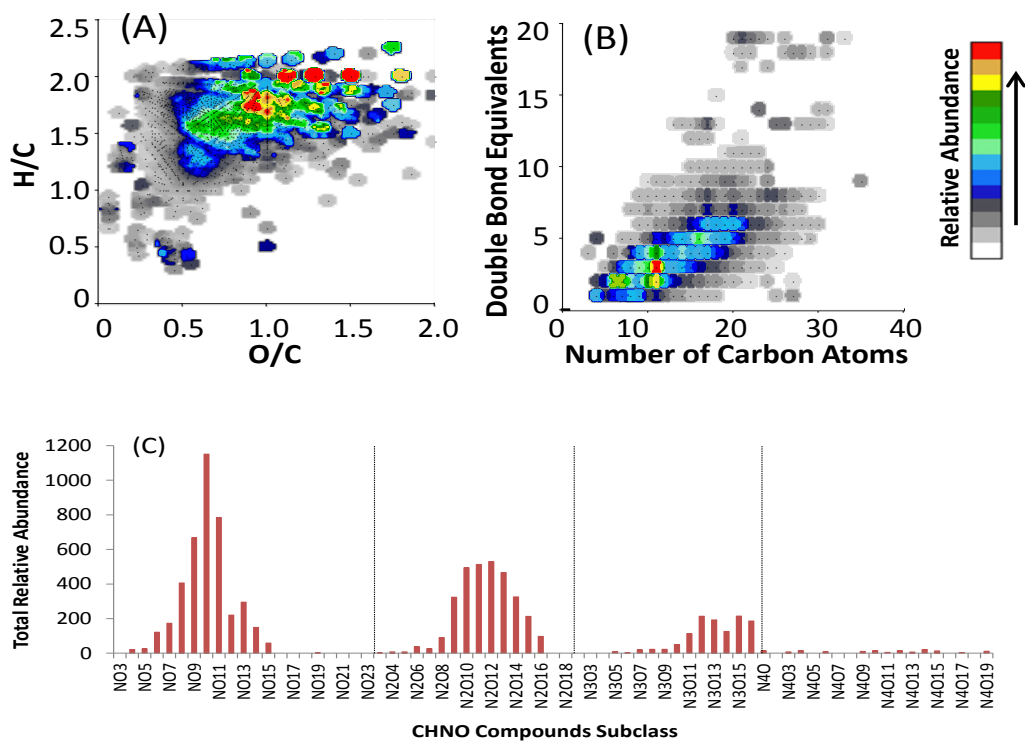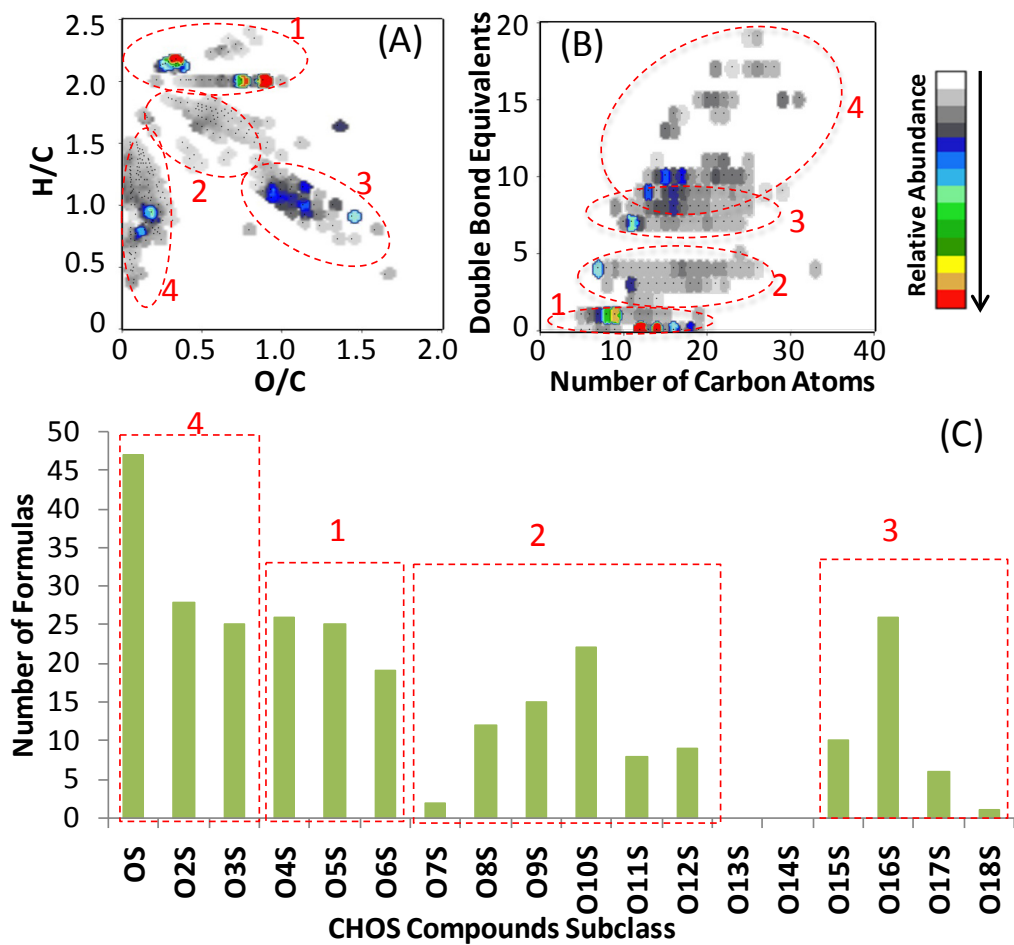**Figure 5.10.** Isoabundance van Krevelen diagram for CW1 CHNOS compounds (A). Isoabundance plot of (DBE) vs. number of carbon atoms in molecular formulas for CW1 CHNOS compounds (B). Number of formulas identified in each of CW1 CHNOS subclasses (C). The red circles indicate the approximate groupings (1 to 4) of the CW1 CHNOS formulas. The bars not boxed in (C) belong to group 4.
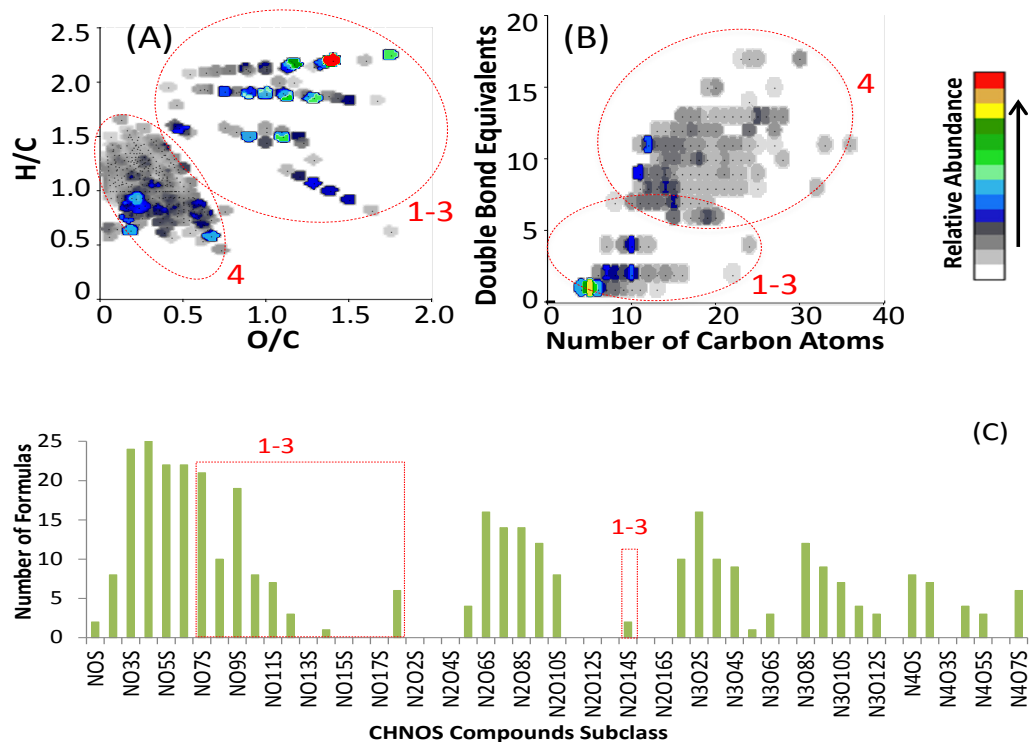
Figure 5.11. Isoabundance van Krevelen diagram for CW2 CHNOS compounds (A). Isoabndance plot of double bond equivalents (DBE) vs. number of carbon atoms in molecular formulas for CW2 CHNOS compounds (B). Number of molecular formulas identified in each of CW2 CHNOS subclasses (C). The red circles and boxes indicate the approximate groupings (1 to 4) of the CW2 CHNOS formulas. The bars not boxed in (C) belong to group 4.

Although organosulfates and nitrooxy-organosulfates were observed, a dominant number of S-containing compounds contain reduced S (e.g., aromatic S) due to insufficient oxygen content in the molecular formulas (subgroup 4 in Figure 5.8 and 5.10). AOM with reduced S is rarely reported (Saranjampour, 2012; Kundu et al., 2013). Most of the high DBE S-containing compounds observed in cloud water have high carbon numbers (> 10; Figure 5.8B and 5.11B). Instead of being secondary products like the organosulfates, they may be from primary emissions. The diesel fuel used to power snow cats used to groom the ski area at night near the SPL may be a source of reduced S compounds. Further investigation is needed to determine the origin of the unambiguously identified reduced S compounds.

5.5 Cloud water AOM properties

Overall, the bulk properties of the two cloud water samples are very similar (Table 5.1). The average O/C ratios of the CW1 and CW2 samples are 0.62 (± 0.37) and 0.61 (± 0.34) and the

average H/C ratios are 1.46 (± 0.36) and 1.46 (± 0.35). The average elemental ratios of each compound group with and without relative abundance weighting (equation 2-4 and 2-5) are summarized in Table 5.1. The relative abundance weighted elemental ratios are very similar in most cases to the unweighted values discussed here. As shown in Table 5.1, the O/C ratios of the CHO, CHOS, and CHNOS compounds are all approximately 0.5, but the O/C of the CHNO compounds is ~0.7. Thus the overall elemental ratios are highly influenced by the CHNO compounds, given their high number frequency and high relative abundance. Low molecular weight organic compounds like formate, acetate and formaldehyde observed previously in ambient samples (van Pinxteren et al., 2005; Herckes et al., 2007; Collett et al., 2008; Samy et al., 2010) but not here have relatively high O/C ratios (>1). In general the overall O/C ratios observed here are consistent with the O/C ratios observed for other aerosol water-soluble organic carbon and AOM samples, but they are at the higher end of the reported range. For example, Aiken et al. (2008) reported O/C ratios for various ambient aerosols from 0.16 to 0.76 using aerosol mass spectrometry. Similarly, Lee et al. (2012) estimated O/C ratios of low-volatility cloud water organics to be between 0.52 and 0.59. The observed cloud water O/C values are higher than Mazzoleni et al. (2010) observed in fog water collected in Fresno, California (0.43), but lower than Altieri et al. (2009b) observed in rainwater collected in New Jersey (0.96). Similar to the O/C ratios, the H/C ratios of CHNO compounds are higher than those of the CHO, CHOS and CHNOS compounds, so the overall H/C values are mainly driven by the high number and relative abundances of CHNO compounds. The observed H/C ratios for the cloud water are also within the range of reported values for aerosol water-soluble organic carbon and AOM samples. Aiken et al. (2008) reported H/C ratios for ambient aerosols from 1.41 to 1.89. A similar range of H/C values were observed in fog and rainwater samples.

Average DBE values and relative abundance weighted average DBE values of each compound group are listed in Table 5.1. The average DBE values of the CW1 and CW2 samples are 6.30 (± 3.37) and 6.29 (± 3.29). Again despite the aromatic compounds identified in the CHO, CHOS and CHNOS groups, the high number of saturated CHNO compounds lowers the overall average DBE values. The relative abundance weighted average values of the samples are even lower than the unweighted average DBE values due to the influence of high relative abundance saturated CHNO compounds. Similarly, the relative abundance weighted DBE value for the summer aerosol water-soluble organic carbon at the SPL was observed to be 5.35 (± 0.05) (Mazzoleni et

al., 2012). In general, the DBE values for the cloud water AOM are consistent with a strong biogenic influence (e.g. oxidation products of terpenes and methoxyphenols) and are comparable to other aerosol water-soluble organic carbon and AOM samples.

The average and relative abundance weighted organic mass-to-organic carbon ratios (OM/OC) for the samples and each of the compound groups were also evaluated (Table 5.1). Consistent with previous studies (Altieri et al., 2009b; Mazzoleni et al., 2012), we define OM/OC as the measured mass divided by the calculated mass of C in the assigned formulas. Overall, the OM/OC ratios for the clouds were ~ 2.1. As expected, the compound groups with heteroatoms (N and S) have higher OM/OC ratios than the CHO compound group. The observed OM/OC ratios for the clouds are slightly higher than those for the ambient aerosols (Aiken et al., 2008; Mazzoleni et al., 2012), reflecting a higher oxidized characteristic and the importance of N- and S-containing compounds identified in clouds.

Although the overall bulk properties for the two cloud water samples are quite similar, detectable compositional differences between the two samples were observed within the N-containing compounds. The reconstructed difference mass spectrum between the CW1 and CW2 CHNO compounds (Figure 5.12) shows several unique peaks with high relative abundance in CW1 in the range of m/z 250 - 350, while those in CW2 are within the range of m/z 350 - 450. The unique high relative abundance CHNO compounds in CW1 are mostly saturated compounds with low DBE values (2 and 4). Some examples of these compounds are the homologs of N_2O_{10} with DBE values of 2 ($C_4H_8(CH_2)_{1-5}N_2O_{10}$), homologs of N_2O_9 with DBE values of 4 ($C_8H_{12}(CH_2)_{1-3}N_2O_9$). However, the unique high relative abundance CHNO compounds in CW2 contain one nitrogen atom instead of two. Some example compounds are homologs of NO_{10} with DBE values of 3, 5 or 7; homologs of NO_{11} with DBE values of 4, 5, 6 or 8. The average N/C ratios for the unique CHNO compounds are 0.15 in CW1 and 0.13 in CW2. Similarly, the average N/C for the unique CHNOS compounds in CW1 is higher than in CW2 (0.12 versus 0.10). These observations may be the result of nighttime NO_x chemistry (Iinuma et al., 2007; Fry et al., 2013). At night, nitrate radical (NO_3) is produced from the reaction of O_3 and NO_2, which can react with organics (RO_2) to produce organonitrates ($RONO_2$, (Kroll and Seinfeld, 2008)). As discussed in Section 3.3, organonitrates observed in the clouds may be from the CCN particulate phase, gas phase or formed within the aqueous phase. The NO_x chemistry can be involved in all these pathways at night. For example, more organonitrates would be produced in gas phase, partitioned to the

aerosols. Also, it has been shown that nitrate radical can transfer from atmospheric gas phase to aqueous phase and is the main source of nitrate radical in aqueous phase (Herrmann et al., 2010). Oxidation of reduced N-containing compounds or nitrate radical oxidation of CHO or CHOS compounds in the aqueous phase may occur (Perraud et al., 2010). Since part of the CW1 composite was collected at night, it is reasonable to observe more N-containing compounds with higher nitrogen content in the molecular formulas from nighttime NO_x chemical pathways in CW1 than in CW2.

Top 10 Positive RA Formulas
1. $C_9H_{14}N_2O_{12}$
2. $C_7H_{14}N_2O_9$
3. $C_{10}H_{16}N_2O_{12}$
4. $C_{10}H_{20}N_2O_{10}$
5. $C_{10}H_{16}N_2O_9$
6. $C_6H_{12}N_2O_9$
7. $C_8H_{16}N_2O_{10}$
8. $C_{11}H_{22}N_2O_9$
9. $C_7H_{14}N_2O_{10}$
10. $C_{11}H_{18}N_2O_{12}$

Top 10 Negative RA Formulas
1. $C_{16}H_{25}NO_{10}$
2. $C_{15}H_{23}NO_{10}$
3. $C_{16}H_{27}NO_{11}$
4. $C_{17}H_{25}NO_{11}$
5. $C_{17}H_{29}NO_{11}$
6. $C_{14}H_{21}NO_{10}$
7. $C_{18}H_{27}NO_{11}$
8. $C_{19}H_{29}NO_{11}$
9. $C_{16}H_{25}NO_9$
10. $C_{15}H_{23}NO_9$
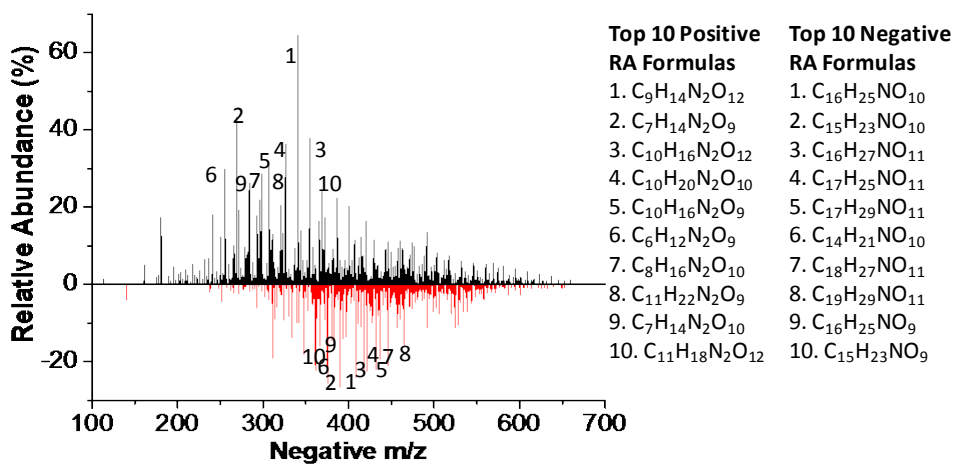
Figure 5.12. Difference mass spectra of CHNO compounds between CW1 and CW2. Positive relative abundance (black peaks) means the compound has higher relative abundance in CW1, negative relative abundance (red peaks) means the compound has higher relative abundance in CW2. The molecular formulas with top 10 positive and top 10 negative relative abundances are also shown in the plot.

Chapter 6 Conclusions

This dissertation aims to provide a deeper understanding of the molecular composition characteristics of atmospheric organic matter. The molecular composition of samples from three projects were studied to better understand the composition of atmospheric aerosol with respect to atmospheric processes, including biogenic secondary organic aerosol (SOA), ambient meteorological conditions and cloud processes. Biogenic SOA is an important component of ambient aerosol, especially in remote areas with large amounts of vegetation. Biogenic SOA samples were generated in the laboratory from 3 monoterpene precursors and a sesquiterpene precursor to allow the study of their molecular compositions. In addition, ambient aerosols were collected at the Storm Peak Laboratory (SPL), a location with substantial biogenic VOC emissions and daily new particle formation events. The molecular compositions of the aerosol water-soluble organic carbon (WSOC) and the biogenic SOA were evaluated using multivariate statistical analysis methods. Finally, cloud samples were collected at the SPL to study the aqueous processing of atmospheric organic matter. The cloud composition was influenced by biomass burning and subsequent SOA. A unique molecular formula composition was observed in the winter cloud samples probably due to the aqueous phase reactions in clouds.

6.1 Biogenic secondary organic aerosol

To study the characterization of biogenic secondary organic aerosol molecular composition and explore the "tracer compounds" of biogenic secondary organic aerosols, 19 biogenic SOA samples were generated in an aerosol chamber with varied experimental conditions (the LANLSOA samples). They are from the ozonolysis of individual monoterpene biogenic precursors: α-pinene, β-pinene, D-limonene or a sesquiterpene biogenic precursor: β-caryophyllene. The ultrahigh resolution ESI FT-ICR MS analysis showed that all of the SOA samples in this study have quite complex molecular composition with approximately 1000 monoisotopic molecular formulas identified in every SOA sample. The SOA samples have clusters of peaks in their mass spectra indicating that oligomers are a major component of biogenic SOA. From m/z 100 to 1000, there are four clusters of peaks referred to as Group I to IV compounds for monoterpene SOAs and three clusters of peaks referred as Group I to III compounds for sesquiterpene SOA. All of the SOA samples in this study have decreasing oxygen-to-carbon (O/C) ratios and constant or slightly increasing hydrogen-to-carbon (H/C) ratios with

increasing molecular weight or Group number. Also, the number of double bonds and/or rings in the compounds increases with molecular weight or Group number. Overall, monoterpene SOA samples have average O/C ratios of 0.50 and H/C ratios of 1.55. Sesquiterpene SOA samples are less oxidized than monoterpene SOA and has average O/C ratios of 0.38 and are more saturated with H/C ratios around 1.60.

The SOA samples share a large number of common molecular formulas. The molecular composition of the 19 samples was compared using multivariate statistical analysis methods due to their large datasets. The SOA samples generated with controlled parameters in the chamber have less complex composition than ambient atmospheric samples, thus they provide an ideal case for examination of the application of multivariate statistical analysis using ultrahigh resolution MS data. In this work, several data preparation methods were tested. The relative abundances (RA) of the molecular formulas used in the statistical analysis provide substantial information on the sample composition compared to only using presence and absence of the formulas. This is especially true in cases like these where a high number of common formulas are observed. Normalization of the RA on a sample basis is necessary in statistical analysis. Standardization or scaling of the normalized RAs on a variable (or molecular formula) basis ensures all the molecular formulas contribute more equally in the statistical analysis, so that the low RA formulas are not masked by the high RA formulas as was shown when only the normalized RAs were used. The various methods tested in this study show quite similar multivariate statistical results, i.e., the SOA samples from the same precursor have similar molecular composition, and this indicates the results and conclusions drawn from the multivariate statistical analysis are robust.

The multivariate statistical analysis including hierarchical cluster analysis (HCA) and principal component analysis (PCA) showed that the SOA samples from the same precursor have quite similar molecular compositions, regardless of the experimental conditions. The monoterpene SOA samples and sesquiterpene SOA samples are separated in both HCA dendrogrms and PCA biplots as expected, reflecting larger differences in their molecular composition. Although the monoterpenes and the sesquiterpene follow the same ozonolysis reaction principles as described in Section 1.2, the difference in the precursors' molecular formula and structure ($C_{10}H_{16}$ for monoterpene and $C_{15}H_{24}$ for sesquiterpene) lead to the different molecular compositions observed for monoterpene SOA and sesquiterpene SOA. Of the monoterpene

SOA, the samples from α-pinene and β-pinene have more similar composition than the samples from D-limonene. Thus, tracer compounds for α- and β-pinene SOA, D-limonene SOA and β-caryophyllene SOA were determined. Beside the first generation ozonolysis products (monomers) reported in previous studies, tracer compounds of dimers, trimers and tetramers were also identified. Of them, the dimers are more applicable in practice as biogenic SOA indicator species due to their higher amount in the atmospheric samples compared to trimers and tetramers. Example dimer indicator species include $C_{19}H_{28}O_7$ and $C_{19}H_{28}O_9$ for α- and β-pinene SOA, $C_{20}H_{32}O_{10}$ and $C_{19}H_{32}O_{10}$ for D-limonene SOA and $C_{29}H_{46}O_{11}$ and $C_{29}H_{48}O_{11}$ for β-caryophyllene SOA. The comprehensive study of biogenic SOA composition improves our understanding on molecular characterization of biogenic SOA composition and the indicator species identified through multivariate statistical analysis may help in source apportionment of ambient atmospheric samples.

6.2 Ambient aerosol water-soluble organic carbon

Thirty-four daily water-soluble organic carbon (WSOC) samples collected consecutively at the Storm Peak Laboratory were studied. The molecular composition of the WSOC from ambient aerosols was analyzed using ultrahigh resolution FT-ICR MS. All of the ambient WSOC samples have highly complex organic composition with approximately 3000 monoisotopic molecular formulas with molecular weights ranging from 100 to 800 Da identified in every sample. They are labeled as CHO, CHNO, CHOS and CHNOS compounds with respect to their elemental composition. In all of the WSOC samples, CHO and CHNO compounds comprised the largest number fractions of the identified formulas and generally have higher relative abundances than the S-containing (CHOS and CHNOS) compounds. The sample average O/C ratios, H/C ratios and DBE values are mainly influenced by their CHO and CHNO compounds. The WSOC samples over the sampling period show similar chemical properties with regard to the average elemental ratios and DBE values. They have relative abundance weighted O/C ratios around 0.47, H/C ratios around 1.55 and DBE values around 5. The examination of the sample carbon number trends and comparison of the WSOC samples with LANLSOA samples indicated that the compositions of these WSOC samples are highly affected by biogenic secondary organic aerosols.

Multivariate statistical analysis was applied to the WSOC composition to examine the similarities and differences of the sample molecular compositions. Similar to the LANLSOA samples, the methods of data preparation were also tested on the WSOC samples to find the best methods and accommodate additional variables. In the hierarchical cluster analysis, the sample analysis replicates show the most similar molecular composition when the normalized RA was used but not the presence/absence of the formulas or the un-normalized RA, indicating the valuable information provided by the intensities of the formulas, and the necessity of normalizing the RA in every sample. Based on the test results, the scaled normalized RAs were used in the multivariate statistical analysis. The ambient samples showed different groupings when different parameters were selected in the hierarchical cluster analysis, indicating that there is not a "natural" grouping characteristic for the ambient aerosol WSOC samples with respect to their molecular composition alone. However, some WSOC samples always grouped closely in HCA dendrogram regardless of the parameters used, indicating their similar molecular composition. The ambient WSOC composition was affected by various meteorological parameters. None of the meteorological parameters played a clearly dominating role in determining the WSOC molecular composition. But through principal component analysis, some of the samples were found to have been strongly affected by a few meteorological parameters and share similar components. Samples collected on July 16th, July 24th, etc. were strongly affected by strong UV radiation and high temperature and these samples contain a large number of molecular formulas consistent with biogenic SOA components. Samples collected on July 21st, July 22nd, etc. were influenced by high relative humidity and these samples contain many CHOS and CHNOS molecular formulas which are expected to be formed in atmospheric aqueous phases. The study of the molecular compositions of the daily aerosol-derived WSOC not only provides the characterization of ambient WSOC composition, but also, for the first time, reveals their molecular level composition changes across one month time period. In addition, the multivariate statistical analysis of the WSOC molecular composition along with meteorological parameters helps to understand the relationships between the meteorological parameters and ambient WSOC molecular composition.

6.3 Ambient cloud water composition

Approximately 3000 monoisotopic molecular formulas were assigned to two supercooled cloud water samples collected at the Storm Peak Laboratory. An overall complex organic composition with the elements of C, H, N, O, and S was observed. As well as a wide range of DBE values similar to other ambient aerosol water-soluble organic carbon and aqueous AOM samples was observed. CHO, CHNO, CHOS and CHNOS compounds with molecular weights up to 700 Da were observed in the cloud samples. Approximately 70% of the identified molecular formulas were observed in the mass range of 300 to 500 Da. Unique to this observation, the CHNO compounds had the highest number frequency and relative abundances. The average elemental ratios of O/C = 0.6 and H/C = 1.5 and DBE values of 6.3 were largely influenced by the highly oxygenated CHNO compounds in the clouds. The numbers of CHNO compounds were slightly higher than the aerosol WSOC collected in the summer at this site. It should be noted that the composition differences between the ambient daily aerosol WSOC and the clouds not only stemmed from the aqueous processing in the cloud droplets, but also from the different sampling seasons. Though the ambient aerosol and cloud samples were both collected at the SPL, the summer time aerosols experienced different environmental conditions than the winter time clouds, such as higher temperature, stronger biogenic VOC emissions and so on. Consistent with the summer aerosol WSOC composition, the majority of the cloud water CHO compounds matched CHO compounds assigned to chamber generated biogenic SOA. Furthermore, the S-containing species observed in the clouds suggest biogenic contributions to the cloud water AOM. Biomass stored terpenes are likely released during residential wood combustion, yielding significant biogenic SOA-like composition. Biomass combustion products also influence the cloud water AOM composition as evidenced by the high number (~1500 unique formulas) of nitrogen containing compounds. Approximately one-third of the CHOS and approximately three-quarters of the CHNOS compounds were identified as reduced S-containing compounds. They might be from diesel fuel combustion or power plant emissions, but further investigation is needed to determine their origin. More nitrogen atoms were associated with the high RA molecular formulas assigned to the cloud sample with a nighttime composite than the daytime sample (2 versus 1 nitrogen atom), likely reflecting the nighttime NO_x chemistry. The molecular composition determined from ultrahigh resolution FT-ICR mass spectrometry provides insights into the organic composition of cloud water AOM in the Rocky Mountain area and indicates

significant biogenic SOA and residential wood combustion contributions to the overall composition.

Biogenic SOA
- Molecular composition characterization of SOA from different precursors
- Indicator species for α- and β-pinene SOA, D-limonene SOA and β-caryophyllene SOA

A major component

Statistics test for ambient samples

Confirms the SOA indicator species

Aerosol-derived WSOC
- Molecular composition characterization of ~ 30 daily WSOC samples
- Effect of meteorological conditions on WSOC molecular composition
- Molecular formulas characteristic of different meteorological conditions

Confirms the SOA indicator species

A major component

A major component

Clouds
- Overall chemical properties of cloud samples including elemental ratios and DBE
- Effect of biogenic SOA on cloud molecular composition
- Effect of biomass combustion on cloud composition
- Effect of nighttime NO_x chemistry on aloud composition

Figure 6.1 The achievements of the dissertation and their relationships.

Overall the achievements of this dissertation from the three projects and their relationships are summarized in Figure 6.1. Biogenic SOA components are commonly observed in the two types of ambient atmospheric samples (aerosols and clouds), which were collected at a location (SPL) with significant biogenic emissions. The controlled biogenic SOA samples were also used for multivariate statistical analysis tests, which were applied to the ambient samples. Conversely, the study of aerosols and clouds provides real atmospheric samples, which were used to confirm the biogenic SOA indicator species identified in chamber-generated SOA.

Chapter 7 Recommendations for future work

From the discussions of the findings in this dissertation, several subtopics are worthy of further study.

1. Apply other multivariate statistical analysis methods to the ultrahigh resolution MS data of atmospheric samples. Two types of multivariate statistical analysis including the hierarchical cluster analysis and principal component analysis were applied to the ultrahigh resolution ESI FT-ICR MS data. Many studies using Aerosol Mass Spectrometry (AMS) have analyzed the composition of atmospheric organic matter using Positive Matrix Factorization (PMF) (e.g., Sun et al., 2011). It would be very interesting to try the PMF on these ultrahigh resolution ESI FT-ICR MS datasets. The data preparation might be easier than it was for HCA and PCA.

2. Test other RA standardization methods before the HCA and PCA analysis. When standardizing the normalized RA for statistical analysis, the traditional standardization method, auto-scaling, was applied. This method treats all the formulas with the same role in statistics. However some compounds have higher concentrations than others in the samples, which might be a more important aspect to represent the sample molecular composition. Other standardization methods might be applied to reflect these aspects. For example, the formulas can be ranked from 1 to 10. The formulas with RA 90-100 % could have rank 10, while the formulas with RA 0-10% could have rank 1. In this way the variables are re-scaled to show their differences.

3. Further study is necessary to reveal the effects of RH and the OH scavenger on the biogenic SOA composition. These effects are very complex, because they are different for different precursors. More biogenic SOA samples generated with various RH and OH scavenger conditions are necessary to further study the effects. For example, the SOA samples can be generated under RH of 10%, 30%, 50%, 70% and so on. Other OH scavengers like 2-butanol or hexane could also be tested

References

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, edited by: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Aiken, A. C., DeCarlo, P. F., and Jimenez, J. L.: Elemental Analysis of Organic Species with Electron Ionization High-Resolution Mass Spectrometry, Analytical Chemistry, 79, 8350-8358, 2007.

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, 4478-4485, 2008.

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

Altieri, K. E., Seitzinger, S. P., Carlton, A. G., Turpin, B. J., Klein, G. C., and Marshall, A. G.: Oligomers formed through in-cloud methylglyoxal reactions: Chemical composition, properties, and mechanisms investigated by ultra-high resolution FT-ICR mass spectrometry, Atmospheric Environment, 42, 1476-1490, 2008.

Altieri, K. E., Turpin, B. J., and Seitzinger, S. P.: Composition of Dissolved Organic Nitrogen in Continental Precipitation Investigated by Ultra-High Resolution FT-ICR Mass Spectrometry, Environmental Science & Technology, 43, 6950-6955, 2009a.

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, 2533-2542, 2009b.

Altieri, K. E., Hastings, M. G., Peters, A. J., and Sigman, D. M.: Molecular characterization of water soluble organic nitrogen in marine rainwater by ultra-high resolution electrospray ionization mass spectrometry, Atmospheric Chemistry and Physics, 12, 3557-3571, 2012.

Amin, H., Atkins, P. T., Russo, R. S., Brown, A. W., Sive, B., Hallar, A. G., and Huff Hartz, K. E.: Effect of Bark Beetle Infestation on Secondary Organic Aerosol Precursor Emissions, Environmental Science & Technology, 46, 5696-5703, 2012.

Arp, H. P. H., and Goss, K.-U.: Ambient Gas/Particle Partitioning. 3. Estimating Partition Coefficients of Apolar, Polar, and Ionizable Organic Compounds by Their Molecular Structure, Environmental Science & Technology, 43, 1923-1929, 2009.

Atkinson, R., and Arey, J.: Gas-phase tropospheric chemistry of biogenic volatile organic compounds: a review, Atmospheric Environment, 37, Supplement 2, 197-219, 2003.

Baduel, C., Voisin, D., and Jaffrezo, J. L.: Comparison of analytical methods for Humic Like Substances (HULIS) measurements in atmospheric particles, Atmospheric Chemistry and Physics, 9, 5949-5962, 2009.

Bateman, A. P., Walser, M. L., Desyaterik, Y., Laskin, J., Laskin, A., and Nizkorodov, S. A.: The Effect of Solvent on the Analysis of Secondary Organic Aerosol Using Electrospray

Ionization Mass Spectrometry, Environmental Science & Technology, 42, 7341-7346, 2008.

Bateman, A. P., Nizkorodov, S. A., Laskin, J., and Laskin, A.: Time-resolved molecular characterization of limonene/ozone aerosol using high-resolution electrospray ionization mass spectrometry, Physical Chemistry Chemical Physics, 11, 7931-7942, 2009.

Bateman, A. P., Nizkorodov, S. A., Laskin, J., and Laskin, A.: High-Resolution Electrospray Ionization Mass Spectrometry Analysis of Water-Soluble Organic Aerosols Collected with a Particle into Liquid Sampler, Analytical Chemistry, 82, 8010-8016, 2010.

Bateman, A. P., Nizkorodov, S. A., Laskin, J., and Laskin, A.: Photolytic processing of secondary organic aerosols dissolved in cloud droplets, Physical Chemistry Chemical Physics, 13, 2011.

Bernstein, J. A., Alexis, N., Barnes, C., Bernstein, I. L., Nel, A., Peden, D., Diaz-Sanchez, D., Tarlo, S. M., Williams, P. B., and Bernstein, J. A.: Health effects of air pollution, Journal of Allergy and Clinical Immunology, 114, 1116-1123, 2004.

Betterton, E. A., and Hoffmann, M. R.: Henry's law constants of some environmentally important aldehydes, Environmental Science & Technology, 22, 1415-1418, 1988.

Blakney, G., Robinson, D., Ly, N., Kelleher, N., Hendrickson, C., Marshall, A., and editors: Predator: Pci data station for FT-ICR mass spectrometry, 53rd American Society of Mass Spectrometry Annual Conference on Mass Spectrometry & Allied Topics; 2005 June 4-9, San Antonio, TX, 2005.

Blakney, G. T., Hendrickson, C. L., and Marshall, A. G.: Predator data station: A fast data acquisition system for advanced FT-ICR MS experiments, International Journal of Mass Spectrometry, 306, 246-252, 2011.

Blando, J. D., and Turpin, B. J.: Secondary organic aerosol formation in cloud and fog droplets: a literature evaluation of plausibility, Atmospheric Environment, 34, 1623-1632, 2000.

Bruns, E. A., Perraud, V., Zelenyuk, A., Ezell, M. J., Johnson, S. N., Yu, Y., Imre, D., Finlayson-Pitts, B. J., and Alexander, M. L.: Comparison of FTIR and Particle Mass Spectrometry for the Measurement of Particulate Organic Nitrates, Environmental Science & Technology, 44, 1056-1061, 2010.

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, 2007.

Carlton, A. G., Wiedinmyer, C., and Kroll, J. H.: A review of Secondary Organic Aerosol (SOA) formation from isoprene, Atmospheric Chemistry and Physics, 9, 4987-5005, 2009.

Chan, M. N., Surratt, J. D., Claeys, M., Edgerton, E. S., Tanner, R. L., Shaw, S. L., Zheng, M., Knipping, E. M., Eddingsaas, N. C., Wennberg, P. O., and Seinfeld, J. H.: Characterization and Quantification of Isoprene-Derived Epoxydiols in Ambient Aerosol in the Southeastern United States, Environmental Science & Technology, 44, 4590-4596, 2010.

Charbouillot, T., Gorini, S., Voyard, G., Parazols, M., Brigante, M., Deguillaume, L., Delort, A.-M., and Mailhot, G.: Mechanism of carboxylic acid photooxidation in atmospheric aqueous phase: Formation, fate and reactivity, Atmospheric Environment, 56, 1-8, 2012.

Christoffersen, T. S., Hjorth, J., Horie, O., Jensen, N. R., Kotzias, D., Molander, L. L., Neeb, P., Ruppert, L., Winterhalter, R., Virkkula, A., Wirtz, K., and Larsen, B. R.: cis-pinic acid, a possible precursor for organic aerosol formation from ozonolysis of α-pinene, Atmospheric Environment, 32, 1657-1661, 1998.

Claeys, M., Graham, B., Vas, G., Wang, W., Vermeylen, R., Pashynska, V., Cafmeyer, J., Guyon, P., Andreae, M. O., Artaxo, P., and Maenhaut, W.: Formation of Secondary Organic Aerosols Through Photooxidation of Isoprene, Science, 303, 1173-1176, 2004a.

Claeys, M., Wang, W., Ion, A. C., Kourtchev, I., Gelencsér, A., and Maenhaut, W.: Formation of secondary organic aerosols from isoprene and its gas-phase oxidation products through reaction with hydrogen peroxide, Atmospheric Environment, 38, 4093-4098, 2004b.

Claeys, M., Szmigielski, R., Vermeylen, R., Wang, W., Shalamzari, M., and Maenhaut, W.: Tracers for Biogenic Secondary Organic Aerosol from α-Pinene and Related Monoterpenes: An Overview, in: Disposal of Dangerous Chemicals in Urban Areas and Mega Cities, edited by: Barnes, I., and Rudziński, K. J., NATO Science for Peace and Security Series C: Environmental Security, Springer Netherlands, 227-238, 2013.

Cohen, A. J., Ross Anderson, H., Ostro, B., Pandey, K. D., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet, J. M., and Smith, K.: The Global Burden of Disease Due to Outdoor Air Pollution, Journal of Toxicology and Environmental Health, Part A, 68, 1301-1307, 2005.

Collett, J. L., Bator, A., Sherman, D. E., Moore, K. F., Hoag, K. J., Demoz, B. B., Rao, X., and Reilly, J. E.: The chemical composition of fogs and intercepted clouds in the United States, Atmospheric Research, 64, 29-40, 2002.

Collett, J. L., Herckes, P., Youngster, S., and Lee, T.: Processing of atmospheric organic matter by California radiation fogs, Atmospheric Research, 87, 232-241, 2008.

Dalbec, M. M., Zhao, Y., Fisseha, R., Putman, A. L., Rahn, T. A., and Mazzoleni, L. R.: Molecular Formula Characterization of Biogenic Secondary Organic Aerosol: Descriptive Statistical Evaluation (In preparation), 2013.

Davidson, C. I., Phalen, R. F., and Solomon, P. A.: Airborne Particulate Matter and Human Health: A Review, Aerosol Science and Technology, 39, 737-749, 2005.

De Haan, D. O., Tolbert, M. A., and Jimenez, J. L.: Atmospheric condensed-phase reactions of glyoxal with methylamine, Geophysical Research Letters, 36, L11819, 2009.

De Haan, D. O., Hawkins, L. N., Kononenko, J. A., Turley, J. J., Corrigan, A. L., Tolbert, M. A., and Jimenez, J. L.: Formation of Nitrogen-Containing Oligomers by Methylglyoxal and Amines in Simulated Evaporating Cloud Droplets, Environmental Science & Technology, 45, 984-991, 2010.

Decesari, S., Facchini, M. C., Fuzzi, S., McFiggans, G. B., Coe, H., and Bower, K. N.: The water-soluble organic component of size-segregated aerosol, cloud water and wet depositions from Jeju Island during ACE-Asia, Atmospheric Environment, 39, 211-222, 2005.

Deininger, C. K., and Saxena, V. K.: A validation of back trajectories of air masses by principal component analysis of ion concentrations in cloud water, Atmospheric Environment, 31, 295-300, 1997.

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, 2013.

Draxler, R.R. and Rolph, G.D: HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model access via NOAA ARL READY Website (http://www.arl.noaa.gov/HYSPLIT.php). NOAA Air Resources Laboratory, College Park, MD.

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

Dzepina, K., Volkamer, R. M., Madronich, S., Tulet, P., Ulbrich, I. M., Zhang, Q., Cappa, C. D., Ziemann, P. J., and Jimenez, J. L.: Evaluation of recently-proposed secondary organic aerosol models for a case study in Mexico City, Atmos. Chem. Phys., 9, 5681-5709, 2009.

Edney, E. O., Kleindienst, T. E., Jaoui, M., Lewandowski, M., Offenberg, J. H., Wang, W., and Claeys, M.: Formation of 2-methyl tetrols and 2-methylglyceric acid in secondary organic aerosol from laboratory irradiated isoprene/NOX/SO2/air mixtures and their detection in ambient PM2.5 samples collected in the eastern United States, Atmospheric Environment, 39, 5281-5289, 2005.

El Haddad, I., Yao, L., Nieto-Gligorovski, L., Michaud, V., Temime-Roussel, B., Quivet, E., Marchand, N., Sellegri, K., and Monod, A.: In-cloud processes of methacrolein under simulated conditions – Part 2: Formation of secondary organic aerosol, Atmos. Chem. Phys., 9, 5107-5117, 2009.

Englert, N.: Fine particles and human health—a review of epidemiological studies, Toxicology Letters, 149, 235-242, 2004.

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, 2011.

Facchini, M. C., Fuzzi, S., Zappoli, S., Andracchio, A., Gelencsér, A., Kiss, G., Krivácsy, Z., Mészáros, E., Hansson, H.-C., Alsberg, T., and Zebühr, Y.: Partitioning of the organic aerosol component between fog droplets and interstitial air, Journal of Geophysical Research: Atmospheres, 104, 26821-26832, 1999.

Fan, X., Song, J., and Peng, P. a.: Comparative study for separation of atmospheric humic-like substance (HULIS) by ENVI-18, HLB, XAD-8 and DEAE sorbents: Elemental composition, FT-IR, 1H NMR and off-line thermochemolysis with tetramethylammonium hydroxide (TMAH), Chemosphere, 93, 1710-1719, 2013.

Feczko, T., Puxbaum, H., Kasper-Giebl, A., Handler, M., Limbeck, A., Gelencser, A., Pio, C., Preunkert, S., and Legrand, M.: Determination of water and alkaline extractable atmospheric humic-like substances with the TU Vienna HULIS analyzer in samples from six background sites in Europe, J. Geophys. Res.-Atmos., 112, 9, 2007.

Feng, J., and Möller, D.: Characterization of Water-Soluble Macromolecular Substances in Cloud Water, Journal of Atmospheric Chemistry, 48, 217-233, 2004.

Finlayson-Pitts, B. J., and Pitts, J. N.: Tropospheric Air Pollution: Ozone, Airborne Toxics, Polycyclic Aromatic Hydrocarbons, and Particles, Science, 276, 1045-1051, 1997.

Finlayson-Pitts, B. J., and Pitts, J. N.: Chemistry of the Upper and Lower Atmosphere, Academic Press, San Diego, 2000.

Fry, J. L., Draper, D. C., Zarzana, K. J., Campuzano-Jost, P., Day, D. A., Jimenez, J. L., Brown, S. S., Cohen, R. C., Kaser, L., Hansel, A., Cappellin, L., Karl, T., Hodzic Roux, A., Turnipseed, A., Cantrell, C., Lefer, B. L., and Grossberg, N.: Observations of gas- and aerosol-phase organic nitrates at BEACHON-RoMBAS 2011, Atmos. Chem. Phys. Discuss., 13, 1979-2034, 2013.

Galloway, M. M., Chhabra, P. S., Chan, A. W. H., Surratt, J. D., Flagan, R. C., Seinfeld, J. H., and Keutsch, F. N.: Glyoxal uptake on ammonium sulphate seed aerosol: reaction products and reversibility of uptake under dark and irradiated conditions, Atmospheric Chemistry and Physics, 9, 3331-3345, 2009.

Gao, Y., Hall, W. A., and Johnston, M. V.: Molecular Composition of Monoterpene Secondary Organic Aerosol at Low Mass Loading, Environmental Science & Technology, 44, 7897-7902, 2010.

Garnes, L. A., and Allen, D. T.: Size Distributions of Organonitrates in Ambient Aerosol Collected in Houston, Texas, Aerosol Science and Technology, 36, 983-992, 2002.

Gelencser, A.: Carbonaceous Aerosol, Springer, The Netherlands, 2004.

Gilardoni, S., Liu, S., Takahama, S., Russell, L. M., Allan, J. D., Steinbrecher, R., Jimenez, J. L., De Carlo, P. F., Dunlea, E. J., and Baumgardner, D.: Characterization of organic ambient aerosol during MIRAGE 2006 on three platforms, Atmos. Chem. Phys., 9, 5417-5432, 2009.

Gioda, A., Reyes-Rodríguez, G. J., Santos-Figueroa, G., Collett, J. L., Jr., Decesari, S., Ramos, M. d. C. K. V., Bezerra Netto, H. J. C., de Aquino Neto, F. R., and Mayol-Bracero, O. L.: Speciation of water-soluble inorganic, organic, and total nitrogen in a background marine environment: Cloud water, rainwater, and aerosol particles, J. Geophys. Res., 116, D05203, 2011.

Gioda, A., Mayol-Bracero, O. L., Scatena, F. N., Weathers, K. C., Mateus, V. L., and McDowell, W. H.: Chemical constituents in clouds and rainwater in the Puerto Rican rainforest: Potential sources and seasonal drivers, Atmospheric Environment, 68, 208-220, 2013.

Gómez-González, Y., Surratt, J. D., Cuyckens, F., Szmigielski, R., Vermeylen, R., Jaoui, M., Lewandowski, M., Offenberg, J. H., Kleindienst, T. E., Edney, E. O., Blockhuys, F., Van Alsenoy, C., Maenhaut, W., and Claeys, M.: Characterization of organosulfates from the photooxidation of isoprene and unsaturated fatty acids in ambient aerosol using liquid chromatography/(−) electrospray ionization mass spectrometry, Journal of Mass Spectrometry, 43, 371-382, 2008.

Graber, E. R., and Rudich, Y.: Atmospheric HULIS: How humic-like are they? A comprehensive and critical review, Atmospheric Chemistry and Physics, 6, 729-753, 2006.

Graedel, T. E., and Weschler, C. J.: Chemistry within aqueous atmospheric aerosols and raindrops, Reviews of Geophysics, 19, 505-539, 1981.

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, 1999.

Hallar, A. G., Lowenthal, D. H., Chirokova, G., Borys, R. D., and Wiedinmyer, C.: Persistent daily new particle formation at a mountain-top location, Atmospheric Environment, 45, 4111-4115, 2011.

Hallar, A. G., Lowenthal, D. H., Clegg, S. L., Samburova, V., Taylor, N., Mazzoleni, L. R., Zielinska, B. K., Kristensen, T. B., Chirokova, G., McCubbin, I. B., Dodson, C., and Collins, D.: Chemical and hygroscopic properties of aerosol organics at Storm Peak Laboratory, Journal of Geophysical Research: Atmospheres, 118, 4767-4779, 2013.

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., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, T. F., Monod, A., Prevot, 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, Atmos. Chem. Phys. Discuss., 9, 3555-3762, 2009.

Harris, E., Sinha, B., van Pinxteren, D., Tilgner, A., Fomba, K. W., Schneider, J., Roth, A., Gnauk, T., Fahlbusch, B., Mertes, S., Lee, T., Collett, J., Foley, S., Borrmann, S., Hoppe, P., and Herrmann, H.: Enhanced Role of Transition Metal Ion Catalysis During In-Cloud Oxidation of SO2, Science, 340, 727-730, 2013.

Harrison, R. M., and Yin, J.: Particulate matter in the atmosphere: which particle properties are important for its effects on health?, Science of The Total Environment, 249, 85-101, 2000.

Hatakeyama, S., Izumi, K., Fukuyama, T., Akimoto, H., and Washida, N.: Reactions of OH with α-pinene and β-pinene in air: Estimate of global CO production from the atmospheric oxidation of terpenes, Journal of Geophysical Research: Atmospheres, 96, 947-958, 1991.

Hatch, L. E., Creamean, J. M., Ault, A. P., Surratt, J. D., Chan, M. N., Seinfeld, J. H., Edgerton, E. S., Su, Y., and Prather, K. A.: Measurements of Isoprene-Derived Organosulfates in Ambient Aerosols by Aerosol Time-of-Flight Mass Spectrometry - Part 1: Single Particle Atmospheric Observations in Atlanta, Environmental Science & Technology, 45, 5105-5111, 2011.

Havers, N., Burba, P., Lambert, J., and Klockow, D.: Spectroscopic Characterization of Humic-Like Substances in Airborne Particulate Matter, Journal of Atmospheric Chemistry, 29, 45-54, 1998.

Hawthorne, S. B., Miller, D. J., Langenfeld, J. J., and Krieger, M. S.: PM-10 high-volume collection and quantitation of semi- and nonvolatile phenols, methoxylated phenols, alkanes, and polycyclic aromatic hydrocarbons from winter urban air and their relationship to wood smoke emissions, Environmental Science & Technology, 26, 2251-2262, 1992.

Heald, C. L., Jacob, D. J., Park, R. J., Russell, L. M., Huebert, B. J., Seinfeld, J. H., Liao, H., and Weber, R. J.: A large organic aerosol source in the free troposphere missing from current models, Geophys. Res. Lett., 32, L18809, 2005.

Heald, C. L., Kroll, J. H., Jimenez, J. L., Docherty, K. S., DeCarlo, P. F., Aiken, A. C., Chen, Q., Martin, S. T., Farmer, D. K., and Artaxo, P.: A simplified description of the evolution of organic aerosol composition in the atmosphere, Geophys. Res. Lett., 37, L08803, 2010.

Heaton, K. J., Dreyfus, M. A., Wang, S., and Johnston, M. V.: Oligomers in the Early Stage of Biogenic Secondary Organic Aerosol Formation and Growth, Environmental Science & Technology, 41, 6129-6136, 2007.

Helmig, D., Ortega, J., Duhl, T., Tanner, D., Guenther, A., Harley, P., Wiedinmyer, C., Milford, J., and Sakulyanontvittaya, T.: Sesquiterpene Emissions from Pine Trees – Identifications, Emission Rates and Flux Estimates for the Contiguous United States, Environmental Science & Technology, 41, 1545-1553, 2007.

Hennigan, C. J., Bergin, M. H., Dibb, J. E., and Weber, R. J.: Enhanced secondary organic aerosol formation due to water uptake by fine particles, Geophys. Res. Lett., 35, L18801, 2008a.

Hennigan, C. J., Bergin, M. H., and Weber, R. J.: Correlations between Water-Soluble Organic Aerosol and Water Vapor: A Synergistic Effect from Biogenic Emissions?, Environmental Science & Technology, 42, 9079-9085, 2008b.

Hennigan, C. J., Sullivan, A. P., Fountoukis, C. I., Nenes, A., Hecobian, A., Vargas, O., Peltier, R. E., Case Hanks, A. T., Huey, L. G., Lefer, B. L., Russell, A. G., and Weber, R. J.: On the volatility and production mechanisms of newly formed nitrate and water soluble organic aerosol in Mexico City, Atmospheric Chemistry and Physics, 8, 3761-3768, 2008c.

Hennigan, C. J., Bergin, M. H., Russell, A. G., Nenes, A., and Weber, R. J.: Gas/particle partitioning of water-soluble organic aerosol in Atlanta, Atmospheric Chemistry and Physics, 9, 3613-3628, 2009.

Herckes, P., Leenheer, J. A., and Collett, J. L.: Comprehensive Characterization of Atmospheric Organic Matter in Fresno, California Fog Water, Environmental Science & Technology, 41, 393-399, 2007.

Herrmann, H., Hoffmann, D., Schaefer, T., Bräuer, P., and Tilgner, A.: Tropospheric Aqueous-Phase Free-Radical Chemistry: Radical Sources, Spectra, Reaction Kinetics and Prediction Tools, ChemPhysChem, 11, 3796-3822, 2010.

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, 8908-8919, 2008.

Hindman, E., Campbell, M. A., and Borys, R. D.: A 10-year record of cloud-droplet physical and chemical properties at a mountaintop site in Colorado, J. Appl. Meteor, 33, 797-807, 1994.

Hindman, E. E., Carter, E. J., Borys, R. D., and Mitchell, D. L.: Collecting Supercooled Cloud Droplets as a Function of Droplet Size, Journal of Atmospheric and Oceanic Technology, 9, 337-353, 1992.

Hoffer, A., Kiss, G., Blazsó, M., and Gelencsér, A.: Chemical characterization of humic-like substances (HULIS) formed from a lignin-type precursor in model cloud water, Geophysical Research Letters, 31, L06115, 2004.

Hoffmann, T., Bandur, R., Marggraf, U., and Linscheid, M.: Molecular composition of organic aerosols formed in the α-pinene/O3 reaction: Implications for new particle formation processes, Journal of Geophysical Research: Atmospheres, 103, 25569-25578, 1998.

Hughey, C. A., Hendrickson, C. L., Rodgers, R. P., Marshall, A. G., and Qian, K.: Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra, Analytical Chemistry, 73, 4676-4681, 2001.

Hur, M., Yeo, I., Park, E., Kim, Y. H., Yoo, J., Kim, E., No, M.-h., Koh, J., and Kim, S.: Combination of Statistical Methods and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for More Comprehensive, Molecular-Level Interpretations of Petroleum Samples, Analytical Chemistry, 82, 211-218, 2009.

Hutchings, J., Robinson, M., McIlwraith, H., Triplett Kingston, J., and Herckes, P.: The Chemistry of Intercepted Clouds in Northern Arizona during the North American Monsoon Season, Water, Air, & Soil Pollution, 199, 191-202, 2009.

Iinuma, Y., Muller, C., Berndt, T., Boge, O., Claeys, M., and Herrmann, H.: Evidence for the Existence of Organosulfates from a-Pinene Ozonolysis in Ambient Secondary Organic Aerosol, Environmental Science & Technology, 41, 6678-6683, 2007a.

Iinuma, Y., Muller, C., Boge, O., Gnauk, T., and Herrmann, H.: The formation of organic sulfate esters in the limonene ozonolysis secondary organic aerosol (SOA) under acidic conditions, Atmospheric Environment, 41, 5571-5583, 2007b.

Jacobson, M. C., Hansson, H. C., Noone, K. J., and Charlson, R. J.: Organic atmospheric aerosols: Review and state of the science, Reviews of Geophysics, 38, 267-294, 2000.

Jaoui, M., Kleindienst, T. E., Lewandowski, M., Offenberg, J. H., and Edney, E. O.: Identification and Quantification of Aerosol Polar Oxygenated Compounds Bearing Carboxylic or Hydroxyl Groups. 2. Organic Tracer Compounds from Monoterpenes, Environmental Science & Technology, 39, 5661-5673, 2005.

Jaoui, M., Lewandowski, M., Kleindienst, T. E., Offenberg, J. H., and Edney, E. O.: β-caryophyllinic acid: An atmospheric tracer for β-caryophyllene secondary organic aerosol, Geophys. Res. Lett., 34, L05816, 2007.

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., E., Dunlea, 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., Dzepina, 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, 1525-1529, 2009.

Johnson, R. A., and Wichern, D. W.: Applied Multivariate Statistical Analysis, 6th ed., Pearson Prentice Hall, New Jersey, 2007.

Jonsson, Å. M., Hallquist, M., and Ljungström, E.: Impact of Humidity on the Ozone Initiated Oxidation of Limonene, Δ3-Carene, and α-Pinene, Environmental Science & Technology, 40, 188-194, 2005.

Jonsson, Å. M., Hallquist, M., and Ljungström, E.: The effect of temperature and water on secondary organic aerosol formation from ozonolysis of limonene, delta-carene and α-pinene, Atmos. Chem. Phys., 8, 6541-6549, 2008a.

Jonsson, Å. M., Hallquist, M., and Ljungström, E.: Influence of OH Scavenger on the Water Effect on Secondary Organic Aerosol Formation from Ozonolysis of Limonene, delta-Carene, and α-Pinene, Environmental Science & Technology, 42, 5938-5944, 2008b.

Kampf, C., Waxman, E., Slowik, J., Dommen, J., Pfaffenberger, L., Praplan, A., Prevot, A., Baltensperger, U., Hoffmann, T., and Volkamer, R.: Effective Henry's law partitioning and the salting constant of glyoxal in aerosols containing sulfate, Environmental Science & Technology, 2013.

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, 1053-1123, 2005.

Kao, A. S., and Friedlander, S. K.: Temporal Variations of Particulate Air Pollution: A Marker for Free Radical Dosage and Adverse Health Effects?, Inhalation Toxicology, 7, 149-156, 1995.

Kawamura, K., Tachibana, E., Okuzawa, K., Aggarwal, S. G., Kanaya, Y., and Wang, Z. F.: High abundances of water-soluble dicarboxylic acids, ketocarboxylic acids and α-dicarbonyls in the mountaintop aerosols over the North China Plain during wheat burning season, Atmos. Chem. Phys., 13, 8285-8302, 2013.

Keene, W. C., Mosher, B. W., Jacob, D. J., Munger, J. W., Talbot, R. W., Artz, R. S., Maben, J. R., Daube, B. C., and Galloway, J. N.: Carboxylic acids in clouds at a high-elevation forested site in central Virginia, Journal of Geophysical Research: Atmospheres, 100, 9345-9357, 1995.

Kesselmeier, J., and Staudt, M.: Biogenic Volatile Organic Compounds (VOC): An Overview on Emission, Physiology and Ecology, Journal of Atmospheric Chemistry, 33, 23-88, 1999.

Kido Soule, M. C., Longnecker, K., Giovannoni, S. J., and Kujawinski, E. B.: Impact of instrument and experiment parameters on reproducibility of ultrahigh resolution ESI FT-ICR mass spectra of natural organic matter, Organic Geochemistry, 41, 725-733, 2010.

Kim, S., Kramer, R. W., and Hatcher, P. G.: Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram, Analytical Chemistry, 75, 5336-5344, 2003.

Kiss, G., Varga, B., Gelencsér, A., Krivácsy, Z., Molnár, Á., Alsberg, T., Persson, L., Hansson, H.-C., and Cristina Facchini, M.: Characterisation of polar organic compounds in fog water, Atmospheric Environment, 35, 2193-2200, 2001.

Koch, B. P., Witt, M., Engbrodt, R., Dittmar, T., and Kattner, G.: Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Geochimica et Cosmochimica Acta, 69, 3299-3308, 2005.

Koch, B. P., Dittmar, T., Witt, M., and Kattner, G.: Fundamentals of Molecular Formula Assignment to Ultrahigh Resolution Mass Data of Natural Organic Matter, Analytical Chemistry, 79, 1758-1763, 2007.

Kourtchev, I., Fuller, S. J., Giorio, C., Healy, R. M., Wilson, E., O'Connor, I., Wenger, J. C., McLeod, M., Aalto, J., Ruuskanen, T. M., Maenhaut, W., Jones, R., Venables, D. S., Sodeau, J. R., Kulmala, M., and Kalberer, M.: Molecular composition of biogenic secondary organic aerosols using ultrahigh-resolution mass spectrometry: comparing laboratory and field studies, Atmos. Chem. Phys., 14, 2155-2167, 2014.

Kristensen, K., Enggrob, K. L., King, S. M., Worton, D. R., Platt, S. M., Mortensen, R., Rosenoern, T., Surratt, J. D., Bilde, M., Goldstein, A. H., and Glasius, M.: Formation and occurrence of dimer esters of pinene oxidation products in atmospheric aerosols, Atmos. Chem. Phys., 13, 3763-3776, 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, 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, advance online publication, 2011.

Kujawinski, E. B.: Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (ESI FT-ICR MS): Characterization of Complex Environmental Mixtures, Environmental Forensics, 3, 207-216, 2002.

Kujawinski, E. B., Freitas, M. A., Zang, X., Hatcher, P. G., Green-Church, K. B., and Jones, R. B.: The application of electrospray ionization mass spectrometry (ESI MS) to the structural characterization of natural organic matter, Organic Geochemistry, 33, 171-180, 2002a.

Kujawinski, E. B., Hatcher, P. G., and Freitas, M. A.: High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Humic and Fulvic Acids: Improvements and Comparisons, Analytical Chemistry, 74, 413-419, 2002b.

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, 2009.

Kundu, S., Kawamura, K., Andreae, T. W., Hoffer, A., and Andreae, M. O.: Molecular distributions of dicarboxylic acids, ketocarboxylic acids and alpha-dicarbonyls in biomass burning aerosols: implications for photochemical production and degradation in smoke layers, Atmospheric Chemistry and Physics, 10, 2209-2225, 2010.

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, 2012.

Kundu, S., Quraishi, T. A., Yu, G., Suarez, C., Keutsch, F. N., and Stone, E. A.: Evidence and quantitation of aromatic organosulfates in ambient aerosols in Lahore, Pakistan, Atmos. Chem. Phys., 13, 4865-4875, 2013.

Kundu, S., Mazzoleni, L. R., Fisseha, R., Putman, A. L., and Rahn, T. A.: Molecular composition of β-caryophyllene ozonolysis SOA formed in humid and dry conditions, In preparation.

Laj, P., Fuzzi, S., Facchini, M. C., Lind, J. A., Orsi, G., Preiss, M., Maser, R., Jaeschke, W., Seyffer, E., Helas, G., Acker, K., Wieprecht, W., Möller, D., Arends, B. G., Mols, J. J., Colvile, R. N., Gallagher, M. W., Beswick, K. M., Hargreaves, K. J., Storeton-West, R. L., and Sutton, M. A.: Cloud processing of soluble gases, Atmospheric Environment, 31, 2589-2598, 1997.

Laskin, A., Smith, J. S., and Laskin, J.: Molecular Characterization of Nitrogen-Containing Organic Compounds in Biomass Burning Aerosols Using High-Resolution Mass Spectrometry, Environmental Science & Technology, 43, 3764-3771, 2009.

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, Environmental Science & Technology, 46, 4312-4322, 2012.

Lee, A. K. Y., Herckes, P., Leaitch, W. R., Macdonald, A. M., and Abbatt, J. P. D.: Aqueous OH oxidation of ambient organic aerosol and cloud water organics: Formation of highly oxidized products, Geophys. Res. Lett., 38, L11805, 2011.

Lerdau, M., Dilts, S. B., Westberg, H., Lamb, B. K., and Allwine, E. J.: Monoterpene emission from ponderosa pine, Journal of Geophysical Research: Atmospheres, 99, 16609-16615, 1994.

Liggio, J., Li, S.-M., and McLaren, R.: Reactive uptake of glyoxal by particulate matter, Journal of Geophysical Research: Atmospheres, 110, D10304, 2005.

Liggio, J., Li, S.-M., Brook, J. R., and Mihele, C.: Direct polymerization of isoprene and α-pinene on acidic aerosols, Geophysical Research Letters, 34, L05814, 2007.

Likens, G. E., Edgerton, E. S., and Galloway, J. N.: The composition and deposition of organic carbon in precipitation, Blackwell Publishing Ltd, 16-24, 1983.

Lim, H.-J., Carlton, A. G., and Turpin, B. J.: Isoprene Forms Secondary Organic Aerosol through Cloud Processing:â€‰ Model Simulations, Environmental Science & Technology, 39, 4441-4446, 2005.

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, 2010.

Limbeck, A., and Puxbaum, H.: Dependence of in-cloud scavenging of polar organic aerosol compounds on the water solubility, J. Geophys. Res., 105, 19857-19867, 2000.

Lin, P., Rincon, A. G., Kalberer, M., and Yu, J. Z.: Elemental Composition of HULIS in the Pearl River Delta Region, China: Results Inferred from Positive and Negative Electrospray High Resolution Mass Spectrometric Data, Environmental Science & Technology, 46, 7454-7462, 2012a.

Lin, P., Yu, J. Z., Engling, G., and Kalberer, M.: Organosulfates in Humic-like Substance Fraction Isolated from Aerosols at Seven Locations in East Asia: A Study by Ultra-High-Resolution Mass Spectrometry, Environmental Science & Technology, 46, 13118-13127, 2012b.

Loeffler, K. W., Koehler, C. A., Paul, N. M., and De Haan, D. O.: Oligomer Formation in Evaporating Aqueous Glyoxal and Methyl Glyoxal Solutions, Environmental Science & Technology, 40, 6318-6323, 2006.

Lowenthal, D. H., Borys, R. D., and Wetzel, M. A.: Aerosol distributions and cloud interactions at a mountaintop laboratory, Journal of Geophysical Research: Atmospheres, 107, 4345, 2002.

Lowenthal, D. H., Borys, R. D., Choularton, T. W., Bower, K. N., Flynn, M. J., and Gallagher, M. W.: Parameterization of the cloud droplet–sulfate relationship, Atmospheric Environment, 38, 287-292, 2004.

Lucas, Antoine: amap: Another Multidimensional Analysis Package. R package version 0.8-12, 2014. http://CRAN.R-project.org/package=amap

Lüttke, J., and Levsen, K.: Phase partitioning of phenol and nitrophenols in clouds, Atmospheric Environment, 31, 2649-2655, 1997.

Lüttke, J., Scheer, V., Levsen, K., Wünsch, G., Neil Cape, J., Hargreaves, K. J., Storeton-West, R. L., Acker, K., Wieprecht, W., and Jones, B.: Occurrence and formation of nitrated phenols in and out of cloud, Atmospheric Environment, 31, 2637-2648, 1997.

Marinoni, A., Laj, P., Sellegri, K., and Mailhot, G.: Cloud chemistry at the Puy de DÃ´me: variability and relationships with environmental factors, Atmospheric Chemistry and Physics, 4, 715-728, 2004.

Marshall, A. G., Hendrickson, C. L., and Jackson, G. S.: Fourier transform ion cyclotron resonance mass spectrometry: A primer, Wiley Subscription Services, Inc., A Wiley Company, 1-35, 1998.

Mazzoleni, L. R., Zielinska, B., and Moosmüller, H.: Emissions of Levoglucosan, Methoxy Phenols, and Organic Acids from Prescribed Burns, Laboratory Combustion of Wildland Fuels, and Residential Wood Combustion, Environmental Science & Technology, 41, 2115-2122, 2007.

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, Environmental Science & Technology, 44, 3690-3697, 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, 285-297, 2012.

McLafferty, F. W., and Tureek, F.: Interpretation of mass spectra, 4th ed., University Science Books, Sausalito, CA, 1993.

McLafferty FW, T. F.: Interpretation of mass spectra, 4th ed., Sausalito, CA, 1993.

McNeill, V. F., Woo, J. L., Kim, D. D., Schwier, A. N., Wannell, N. J., Sumner, A. J., and Barakat, J. M.: Aqueous-Phase Secondary Organic Aerosol and Organosulfate Formation in Atmospheric Aerosols: A Modeling Study, Environmental Science & Technology, 46, 8075-8081, 2012.

Minerath, E. C., and Elrod, M. J.: Assessing the Potential for Diol and Hydroxy Sulfate Ester Formation from the Reaction of Epoxides in Tropospheric Aerosols, Environmental Science & Technology, 43, 1386-1392, 2009.

Moreda-Pineiro, A., Marcos, A., Fisher, A., and Hill, S. J.: Evaluation of the effect of data pre-treatment procedures on classical pattern recognition and principal components analysis: a case study for the geographical classification of tea, Journal of Environmental Monitoring, 3, 352-360, 2001.

Mukai, H., and Ambe, Y.: Characterization of a humic acid-like brown substance in airborne particulate matter and tentative identification of its origin, Atmospheric Environment (1967), 20, 813-819, 1986.

Munger, J. W., Jacob, D. J., Daube, B. C., Horowitz, L. W., Keene, W. C., and Heikes, B. G.: Formaldehyde, glyoxal, and methylglyoxal in air and cloudwater at a rural mountain site

in central Virginia, Journal of Geophysical Research: Atmospheres, 100, 9325-9333, 1995.

Ng, N. L., Kwan, A. J., Surratt, J. D., Chan, A. W. H., Chhabra, P. S., Sorooshian, A., Pye, H. O. T., Crounse, J. D., Wennberg, P. O., Flagan, R. C., and Seinfeld, J. H.: Secondary organic aerosol (SOA) formation from reaction of isoprene with nitrate radicals (NO3), Atmospheric Chemistry and Physics, 8, 4117-4140, 2008.

Ng, N. L., Canagaratna, M. R., Zhang, Q., Jimenez, J. L., Tian, J., Ulbrich, I. M., Kroll, J. H., Docherty, K. S., Chhabra, P. S., Bahreini, R., Murphy, S. M., Seinfeld, J. H., Hildebrandt, L., Donahue, N. M., DeCarlo, P. F., Lanz, V. A., Prevot, A. S. H., Dinar, E., Rudich, Y., and Worsnop, D. R.: Organic aerosol components observed in Northern Hemispheric datasets from Aerosol Mass Spectrometry, Atmospheric Chemistry and Physics, 10, 4625-4641, 2010.

Nielsen, T., Platz, J., Granby, K., Hansen, A. B., Skov, H., and Egeløv, A. H.: Particulate organic nitrates: sampling and night/day variation, Atmospheric Environment, 32, 2601-2608, 1998.

Novakov, T., and Penner, J. E.: Large contribution of organic aerosols to cloud-condensation-nuclei concentrations, Nature, 365, 823-826, 1993.

Obrist, D., Hallar, A. G., McCubbin, I., Stephens, B. B., and Rahn, T.: Atmospheric mercury concentrations at Storm Peak Laboratory in the Rocky Mountains: Evidence for long-range transport from Asia, boundary layer contributions, and plant mercury uptake, Atmospheric Environment, 42, 7579-7589, 2008.

Pankow, J. F.: An absorption model of the gas/aerosol partitioning involved in the formation of secondary organic aerosol, Atmospheric Environment, 28, 189-193, 1994a.

Pankow, J. F.: An absorption model of gas/particle partitioning of organic compounds in the atmosphere, Atmospheric Environment, 28, 185-188, 1994b.

Parshintsev, J., Hyötyläinen, T., Hartonen, K., Kulmala, M., and Riekkola, M.-L.: Solid-phase extraction of organic compounds in atmospheric aerosol particles collected with the particle-into-liquid sampler and analysis by liquid chromatography-mass spectrometry, Talanta, 80, 1170-1176, 2010.

Perraud, V., Bruns, E. A., Ezell, M. J., Johnson, S. N., Greaves, J., and Finlayson-Pitts, B. J.: Identification of Organic Nitrates in the NO3 Radical Initiated Oxidation of α-Pinene by Atmospheric Pressure Chemical Ionization Mass Spectrometry, Environmental Science & Technology, 44, 5887-5893, 2010.

Perri, M. J., Seitzinger, S., and Turpin, B. J.: Secondary organic aerosol production from aqueous photooxidation of glycolaldehyde: Laboratory experiments, Atmospheric Environment, 43, 1487-1497, 2009.

Pope, C. A., and Dockery, D. W.: Health Effects of Fine Particulate Air Pollution: Lines that Connect, Journal of the Air & Waste Management Association, 56, 709-742, 2006.

Poschl, U.: Atmospheric aerosols: Composition, transformation, climate and health effects, Angewandte Chemie-International Edition, 44, 7520-7540, 2005.

Pruppacher, H. R. a. K., J. D.: Microphysics of clouds and precipitation, 1st ed., D. Reidel Publishing Company, Dordrecht, 1978.

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 a-pinene ozonolysis SOA, Atmospheric Environment, 46, 164-172, 2012.

Raes, F., Dingenen, R. V., Vignati, E., Wilson, J., Putaud, J.-P., Seinfeld, J. H., and Adams, P.: Formation and cycling of aerosols in the global troposphere, Atmospheric Environment, 34, 4215-4240, 2000.

Reemtsma, T., These, A., Venkatachari, P., Xia, X., Hopke, P. K., Springer, A., and Linscheid, M.: Identification of Fulvic Acids and Sulfated and Nitrated Analogues in Atmospheric Aerosol by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, Analytical Chemistry, 78, 8299-8304, 2006.

Reinhardt, A., Emmenegger, C., Gerrits, B., Panse, C., Dommen, J., Baltensperger, U., Zenobi, R., and Kalberer, M.: Ultrahigh Mass Resolution and Accurate Mass Measurements as a Tool To Characterize Oligomers in Secondary Organic Aerosols, Analytical Chemistry, 79, 4074-4082, 2007.

Reyes-Rodríguez, G. J., Gioda, A., Mayol-Bracero, O. L., and Collett Jr, J.: Organic carbon, total nitrogen, and water-soluble ions in clouds from a tropical montane cloud forest in Puerto Rico, Atmospheric Environment, 43, 4171-4177, 2009.

Rincon, A. G., Calvo, A. I., Dietzel, M., and Kalberer, M.: Seasonal differences of urban organic aerosol composition - an ultra-high resolution mass spectrometry study, Environmental Chemistry, 9, 298-319, 2012.

Romero, F., and Oehme, M.: Organosulfates – A New Component of Humic-Like Substances in Atmospheric Aerosols?, Journal of Atmospheric Chemistry, 52, 283-294, 2005.

Samburova, V., Zenobi, R., and Kalberer, M.: Characterization of high molecular weight compounds in urban atmospheric particles, Atmospheric Chemistry and Physics, 5, 2163-2170, 2005.

Samburova, V., Hallar, A. G., Mazzoleni, L. R., Saranjampour, P., Lowenthal, D., Kohl, S., and Zielinska, B.: Composition of water-soluble organic carbon in nonurban atmospheric aerosol collected at the Storm Peak Laboratory, Environmental Chemistry: Accepted, 2013a.

Samburova, V., Hallar, A. G., Mazzoleni, L. R., Saranjampour, P., Lowenthal, D., Kohl, S. D., and Zielinska, B.: Composition of water-soluble organic carbon in non-urban atmospheric aerosol collected at the Storm Peak Laboratory, Environmental Chemistry, -, 2013b.

Samy, S., Mazzoleni, L. R., Mishra, S., Zielinska, B., and Hallar, A. G.: Water-soluble organic compounds at a mountain-top site in Colorado, USA, Atmospheric Environment, 44, 1663-1671, 2010.

Saranjampour, P.: Characterization of water-soluble organic compounds in ambient aerosol using ultrahigh-resolution electrospray ionization fourer transform ion cyclotron resonance mass spectrometry, MS, Department of Chemistry, Michigan Technological University, 2012.

Sareen, N., Schwier, A. N., Shapiro, E. L., Mitroo, D., and McNeill, V. F.: Secondary organic material formed by methylglyoxal in aqueous aerosol mimics, Atmospheric Chemistry and Physics, 10, 997-1016, 2010.

Saxena, P., and Hildemann, L. M.: Water-soluble organics in atmospheric particles: A critical review of the literature and application of thermodynamics to identify candidate compounds, Journal of Atmospheric Chemistry, 24, 57-109, 1996.

Schmitt-Kopplin, P., Gelencsér, A., Dabek-Zlotorzynska, E., Kiss, G., Hertkorn, N., Harir, M., Hong, Y., and Gebefügi, I.: Analysis of the Unresolved Organic Fraction in Atmospheric Aerosols with Ultrahigh-Resolution Mass Spectrometry and Nuclear Magnetic Resonance Spectroscopy: Organosulfates As Photochemical Smog Constituents†, Analytical Chemistry, 82, 8017-8026, 2010.

Seinfeld, J. H., and Pandis, S. N.: Atmospheric Chemistry and Physics, 2nd ed., John Wiley & Sons, New Jersey, 2006.

Shapiro, E. L., Szprengiel, J., Sareen, N., Jen, C. N., Giordano, M. R., and McNeill, V. F.: Light-absorbing secondary organic material formed by glyoxal in aqueous aerosol mimics, Atmospheric Chemistry and Physics, 9, 2289-2300, 2009.

Simeonov, V., Kalina, M., Tsakovski, S., and Puxbaum, H.: Multivariate statistical study of simultaneously monitored cloud water, aerosol and rainwater data from different elevation levels in an alpine valley (Achenkirch, Tyrol, Austria), Talanta, 61, 519-528, 2003.

Simoneit, B. R. T.: Eolian particulates from oceanic and rural areas—their lipids fulvic and humic acids and residual carbon, Physics and Chemistry of the Earth, 12, 343-352, 1980.

Sleighter, R. L., and Hatcher, P. G.: The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter, Journal of Mass Spectrometry, 42, 559-574, 2007.

Sleighter, R. L., and Hatcher, P. G.: Molecular characterization of dissolved organic matter (DOM) along a river to ocean transect of the lower Chesapeake Bay by ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Marine Chemistry, 110, 140-152, 2008.

Sleighter, R. L., Liu, Z., Xue, J., and Hatcher, P. G.: Multivariate Statistical Approaches for the Characterization of Dissolved Organic Matter Analyzed by Ultrahigh Resolution Mass Spectrometry, Environmental Science & Technology, 44, 7576-7582, 2010.

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, 1275-1284, 2003.

Stone, E. A., Yang, L., Yu, L. E., and Rupakheti, M.: Characterization of organosulfates in atmospheric aerosols at Four Asian locations, Atmospheric Environment, 47, 323-329, 2012.

Sullivan, A. P., and Weber, R. J.: Chemical characterization of the ambient organic aerosol soluble in water: 2. Isolation of acid, neutral, and basic fractions by modified size-exclusion chromatography, J. Geophys. Res., 111, D05315, 2006.

Sun, Y., Zhang, Q., Zheng, M., Ding, X., Edgerton, E. S., and Wang, X.: Characterization and Source Apportionment of Water-Soluble Organic Matter in Atmospheric Fine Particles (PM2.5) with High-Resolution Aerosol Mass Spectrometry and GC–MS, Environmental Science & Technology, 45, 4854-4861, 2011.

Sun, Y. L., Zhang, Q., Anastasio, C., and Sun, J.: Insights into secondary organic aerosol formed via aqueous-phase reactions of phenolic compounds based on high resolution mass spectrometry, Atmospheric Chemistry and Physics, 10, 4809-4822, 2010.

Surratt, J. D., Kroll, J. H., Kleindienst, T. E., Edney, E. O., Claeys, M., Sorooshian, A., Ng, N. L., Offenberg, J. H., Lewandowski, M., Jaoui, M., Flagan, R. C., and Seinfeld, J. H.: Evidence for Organosulfates in Secondary Organic Aerosol, Environmental Science & Technology, 41, 517-527, 2006a.

Surratt, J. D., Murphy, S. M., Kroll, J. H., Ng, N. L., Hildebrandt, L., Sorooshian, A., Szmigielski, R., Vermeylen, R., Maenhaut, W., Claeys, M., Flagan, R. C., and Seinfeld, J. H.: Chemical Composition of Secondary Organic Aerosol Formed from the Photooxidation of Isoprene, The Journal of Physical Chemistry A, 110, 9665-9690, 2006b.

Surratt, J. D., Lewandowski, M., Offenberg, J. H., Jaoui, M., Kleindienst, T. E., Edney, E. O., and Seinfeld, J. H.: Effect of Acidity on Secondary Organic Aerosol Formation from Isoprene, Environmental Science & Technology, 41, 5363-5369, 2007.

Surratt, J. D., Goìmez-Gonzaì lez, Y., Chan, A. W. H., Vermeylen, R., Shahgholi, M., Kleindienst, T. E., Edney, E. O., Offenberg, J. H., Lewandowski, M., Jaoui, M., Maenhaut, W., Claeys, M., Flagan, R. C., and Seinfeld, J. H.: Organosulfate Formation in Biogenic Secondary Organic Aerosol, The Journal of Physical Chemistry A, 112, 8345-8378, 2008.

Tan, Y., Carlton, A. G., Seitzinger, S. P., and Turpin, B. J.: SOA from methylglyoxal in clouds and wet aerosols: Measurement and prediction of key products, Atmospheric Environment, 44, 5218-5226, 2010.

Tolocka, M. P., Jang, M., Ginter, J. M., Cox, F. J., Kamens, R. M., and Johnston, M. V.: Formation of Oligomers in Secondary Organic Aerosol, Environmental Science & Technology, 38, 1428-1434, 2004.

van Pinxteren, D., Plewka, A., Hofmann, D., Müller, K., Kramberger, H., Svrcina, B., Bächmann, K., Jaeschke, W., Mertes, S., Collett Jr, J. L., and Herrmann, H.: Schmücke hill cap cloud and valley stations aerosol characterisation during FEBUKO (II): Organic compounds, Atmospheric Environment, 39, 4305-4320, 2005.

Volkamer, R., Jimenez, J. L., San Martini, F., Dzepina, 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, Geophys. Res. Lett., 33, L17811, 2006.

Volkamer, R., San Martini, F., Molina, L. T., Salcedo, D., Jimenez, J. L., and Molina, M. J.: A missing sink for gas-phase glyoxal in Mexico City: Formation of secondary organic aerosol, Geophysical Research Letters, 34, L19807, 2007.

Volkamer, R., Ziemann, P. J., and Molina, M. J.: Secondary Organic Aerosol Formation from Acetylene (C2H2): seed effect on SOA yields due to organic photochemistry in the aerosol aqueous phase, Atmospheric Chemistry and Physics, 9, 1907-1928, 2009.

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, Physical Chemistry Chemical Physics, 10, 1009-1022, 2008.

Waxman, E. M., Dzepina, K., Ervens, B., Lee-Taylor, J., Aumont, B., Jimenez, J. L., Madronich, S., and Volkamer, R.: Secondary organic aerosol formation from semi- and intermediate-volatility organic compounds and glyoxal: Relevance of O/C as a tracer for aqueous multiphase chemistry, Geophysical Research Letters, 40, 978-982, 2013.

Weathers, K. C., Likens, G. E., Bormann, F. H., Bicknell, S. H., Bormann, B. T., Daube, B. C., Eaton, J. S., Galloway, J. N., Keene, W. C., and et al.: Cloudwater chemistry from ten sites in North America, Environmental Science & Technology, 22, 1018-1026, 1988.

Williams, J., de Reus, M., Krejci, R., Fischer, H., and Ström, J.: Application of the variability-size relationship to atmospheric aerosol studies: estimating aerosol lifetimes and ages, Atmos. Chem. Phys., 2, 133-145, 2002.

Wisthaler, A., Jensen, N. R., Winterhalter, R., Lindinger, W., and Hjorth, J.: Measurements of acetone and other gas phase product yields from the OH-initiated oxidation of terpenes by proton-transfer-reaction mass spectrometry (PTR-MS), Atmospheric Environment, 35, 6181-6191, 2001.

Wozniak, A. S., Bauer, J. E., Sleighter, R. L., Dickhut, R. M., and Hatcher, P. G.: Technical Note: Molecular characterization of aerosol-derived water soluble organic carbon using

ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Atmospheric Chemistry and Physics, 8, 5099-5111, 2008.

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, Atmos. Chem. Phys. Discuss., 14, 6427-6470, 2014.

Wu, Z., Rodgers, R. P., and Marshall, A. G.: Two- and Three-Dimensional van Krevelen Diagrams:A Graphical Analysis Complementary to the Kendrick Mass Plot for Sorting Elemental Compositions of Complex Organic Mixtures Based on Ultrahigh-Resolution Broadband Fourier Transform Ion Cyclotron Resonance Mass Measurements, Analytical Chemistry, 76, 2511-2516, 2004.

Xue, J., Lee, C., Wakeham, S. G., and Armstrong, R. A.: Using principal components analysis (PCA) with cluster analysis to study the organic geochemistry of sinking particles in the ocean, Organic Geochemistry, 42, 356-367, 2011.

Yasmeen, F., Sauret, N., Gal, J. F., Maria, P. C., Massi, L., Maenhaut, W., and Claeys, M.: Characterization of oligomers from methylglyoxal under dark conditions: a pathway to produce secondary organic aerosol through cloud processing during nighttime, Atmospheric Chemistry and Physics, 10, 3803-3812, 2010.

Yu, S., Mathur, R., Schere, K., Kang, D., Pleim, J., Young, J., Tong, D., Pouliot, G., McKeen, S. A., and Rao, S. T.: Evaluation of real-time PM2.5 forecasts and process analysis for PM2.5 formation over the eastern United States using the Eta-CMAQ forecast model during the 2004 ICARTT study, J. Geophys. Res., 113, D06204, 2008.

Zhang, Q., and Anastasio, C.: Chemistry of fog waters in California's Central Valley--Part 3: concentrations and speciation of organic and inorganic nitrogen, Atmospheric Environment, 35, 5629-5643, 2001.

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., Dzepina, 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, Geophys. Res. Lett., 34, L13801, 2007.

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, Atmos. Chem. Phys., 13, 12343-12362, 2013.

Appendix: Permission to use copyrighted materials

For Figure 1.1:

This is a License Agreement between Yunzhu Zhao ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by The American Association for the Advancement of Science, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3506391365729
License date	Nov 12, 2014
Licensed content publisher	The American Association for the Advancement of Science
Licensed content publication	Science
Licensed content title	Evolution of Organic Aerosols in the Atmosphere
Licensed content author	J. L. Jimenez, M. R. Canagaratna, N. M. Donahue, A. S. H. Prevot, Q. Zhang, J. H. Kroll, P. F. DeCarlo, J. D. Allan, H. Coe, N. L. Ng, A. C. Aiken, K. S. Docherty, I. M. Ulbrich, A. P. Grieshop, A. L. Robinson, J. Duplissy, J. D. Smith, K. R. Wilson, V. A. Lanz, C. Hueglin, Y. L. Sun, J. Tian, A. Laaksonen, T. Raatikainen, J. Rautiainen, P. Vaattovaara, M. Ehn, M. Kulmala, J. M. Tomlinson, D. R. Collins, M. J. Cubison, E., J. Dunlea, J. A. Huffman, T. B. Onasch, M. R. Alfarra, P. I. Williams, K. et al.
Licensed content date	Dec 11, 2009
Volume number	326
Issue number	5959
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print and electronic
Portion	Figure
Number of figures/tables	1
Order reference number	None
Title of your thesis / dissertation	Molecular characterization of atmospheric organic matter in biogenic secondary organic aerosol, ambient aerosol and clouds
Expected completion date	Dec 2014
Estimated size(pages)	250
Total	0.00 USD

Terms and Conditions

American Association for the Advancement of Science TERMS AND CONDITIONS

Regarding your request, we are pleased to grant you non-exclusive, non-transferable permission, to republish the AAAS material identified above in your work identified above, subject to the terms and conditions herein. We must be contacted for permission for any uses other than those specifically identified in your request above.

The following credit line must be printed along with the AAAS material: "From [Full Reference Citation]. Reprinted with permission from AAAS."

For Figure 1.2:

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

This is a License Agreement between Yunzhu Zhao ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3506401066937
License date	Nov 12, 2014
Licensed content publisher	John Wiley and Sons
Licensed content publication	Journal of Geophysical Research: Atmospheres
Licensed content title	Organic aerosol formation from the oxidation of biogenic hydrocarbons
Licensed copyright line	Copyright 1999 by the American Geophysical Union.
Licensed content author	Robert J. Griffin,David R. Cocker,Richard C. Flagan,John H. Seinfeld
Licensed content date	Sep 21, 2012
Start page	3555
End page	3567
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1
Will you be translating?	No
Title of your thesis / dissertation	Molecular characterization of atmospheric organic matter in biogenic secondary organic aerosol, ambient aerosol and clouds
Expected completion date	Dec 2014
Expected size (number	250

of pages)

Total   0.00 USD