AEROSOL LIQUID WATER: MEASUREMENT, TRENDS, AND IMPLICATIONS FOR ATMOSPHERE-BIOSPHERE INTERACTIONS

By

THIEN KHOI V. NGUYEN

A Dissertation submitted to the Graduate School-New Brunswick Rutgers, The State University of New Jersey In partial fulfillment of the requirements For the degree of Doctor of Philosophy Graduate Program in Environmental Sciences

Written under the direction of Ann Marie G. Carlton

And approved by

________________________
________________________
________________________
________________________

New Brunswick, New Jersey

May, 2016
ABSTRACT OF THE DISSERTATION

Aerosol Liquid Water: Measurement, Trends, And Implications For Atmosphere-Biosphere Interactions

By THIEN KHOI V. NGUYEN

Dissertation Director:
Ann Marie G. Carlton

Particle-phase liquid water is a ubiquitous and abundant component of atmospheric aerosols and influences hydrological cycling, fate and transport of trace species, visibility, human health, and regional climate. Aerosol water mass concentrations are functions of relative humidity ($RH$), temperature, and aerosol concentration and composition. Previous studies focused on inorganic species, and their relationships to particle hygroscopicity are included in atmospheric models. But there is a dearth of field studies and related ambient information pertaining to water modulation by organic compounds and vice versa. The extent to which water facilitates transfer of biogenically derived carbon from the gas to particle phase to form secondary organic aerosol (SOA), which contributes to the fine particle matter ($PM_{2.5}$) burden, is poorly constrained. My dissertation aims to 1) better understand the chemical and thermodynamic controls on aerosol water in the Southeast U.S., an area known to contain high concentrations of SOA, through instrument development and field measurements, 2) quantify the temporal, spatial, and aloft trends in aerosol water through observational estimates, and 3) reconcile
discrepancies between existing surface and satellite measurements of PM across the continental U.S. using aerosol water. My field measurements of aerosol water found a minimum of 7% water in particles by volume and a diel cycle in water content controlled, in part, by aerosol hygroscopicity, that ranged from 1-5 µg m⁻³. Next, semi-observational estimates of aerosol water by an inorganic thermodynamic model discovered that aerosol water mass has decreased in the Southeast by 79% between 2001-2012. Similar reductions of biogenically derived SOA mass in the area are consistent with an aerosol water-mediated mechanism and the correlation in water and SOA mass is statistically robust. Finally, estimates of aerosol water across the continental U.S. yield seasonal differences that were highest in the Southeast and modulated by aerosol chemistry and not RH. Near surface trends in aerosol water are qualitatively similar to space-based aerosol optical thickness measurements and potentially resolve surface and satellite PM measurements. Ultimately, this work contributes to improving our understanding of the connections among aerosol water, biogenic and anthropogenic emissions, and biogenic SOA formed in the presence of anthropogenic perturbations.
This work would also not have been possible without those who have supported me during my time as a graduate student at Rutgers University. First, I would like to thank my advisor, mentor, and friend, Dr. Ann Marie Carlton, who has never failed to provide me with amazing advice, constructive criticism, and support. Naturally, when she suggested we spend a whole summer in rural Alabama, I followed suit. And I’m also glad she never gave up on me (even during that one time I missed our destination while driving in Alabama) and has always continued to push me towards excellence. I am also very grateful to the other members of my dissertation committee – Dr. Barbara Turpin, Dr. Gediminas Mainelis, and Dr. Markus Petters. Barb, I am so honored to have you as a “grandmother” in my academic genealogy. Gedi, I learned so much about air pollution and instrumentation from you. And Markus, it was a pleasure working on Chapter 2 with you, and I have greatly appreciated for your friendship over these years. The co-authors of my work also deserve recognition: Sarah Suda, Hongyu Guo, Rodney Weber, Shannon Capps, Virendra Ghate, Qi Zhang, Jose Jimenez, and Max Pike. I am also grateful for my current and past lab mates: Neha, Caroline, Mike, Amy, and Bryan. You have all made working here an enjoyable experience. Thank you to Dr. Alan Robock for useful conversations about meteorology and pressure levels. And thank you to my fellow graduate students and colleagues at Rutgers for your humor, support, and friendship, as well as staff members of the Department of Environmental Sciences for your administrative assistance. Additionally, my friends and families have also provided me with much support throughout my life and time at Rutgers, and I thank them for their
unwavering love and for filling my life with happy memories. This dissertation is dedicated to my parents.

I also thank the Atmospheric Administration, Atmospheric Research & Analysis Inc., for providing meteorological and speciated ion data for the SEARCH sites, as well as Eric Edgerton, who has been running the SEARCH network since 1998. The IMPROVE database used in this work can be found at http://vista.cira.colostate.edu/improve/, while the NCEP Reanalysis data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA from their Web site at http://www.esrl.noaa.gov/psd/. I am also grateful towards the Environmental Protection Agency, the National Oceanic and National Center for Atmospheric Research, the Electric Power Research Institute, and all of the SOAS participants for their help and support during the SOAS campaign. Finally, much gratitude is also given towards Athanasios Nenes for developing ISORROPIA and providing it to the community.

Finally, I would like to acknowledge my funding sources. My graduate education and research were funded by the Department of Education Graduate Assistance in Areas of National Need (GAANN) Fellowship, the Environmental Protection Agency (EPA) STAR grant R835041, and the National Science Foundation grant AGS-1242155.

Parts of this dissertation have been previously published as follows:

1) Chapter 2 has been published in:

2) Chapter 3 has been published in:

# TABLE OF CONTENTS

Abstract of the Dissertation ........................................................................................................... ii
Acknowledgments ........................................................................................................................ iv
Table of Contents ......................................................................................................................... vii
List of Tables ............................................................................................................................... xi
List of Illustrations ....................................................................................................................... xii

## CHAPTER 1. INTRODUCTION ............................................................................. 1

1.1 Motivation ......................................................................................................................... 1
1.2 Hypothesis and Specific Aims ......................................................................................... 2
1.3 Background: Atmospheric Aerosols .............................................................................. 3
1.4 Background: Aerosol Water ........................................................................................... 4
1.5 Aerosol Water Ubiquity and Abundance ....................................................................... 9
1.6 References ....................................................................................................................... 18

## CHAPTER 2. TRENDS IN PARTICLE-PHASE LIQUID WATER DURING THE SOUTHERN OXIDANT AND AEROSOL STUDY ......................................................... 34

2.1 Abstract ......................................................................................................................... 34
2.2 Introduction ...................................................................................................................... 35
2.3 Experimental Section ..................................................................................................... 38
   2.3.1 Instrument Design ............................................................................................... 38
   2.3.2 Instrument Performance ..................................................................................... 42
   2.3.3 Data Reduction .................................................................................................... 44
   2.3.4 Sources of Uncertainty ....................................................................................... 49
   2.3.5 Chemical Composition ....................................................................................... 53

vii
CHAPTER 5. SUMMARY, IMPLICATIONS, AND FUTURE DIRECTIONS ....... 137

5.1 Summary and Implications ........................................................................................................... 137

5.1.1 Trends in particle-phase liquid water during the Southern Oxidant and Aerosol Study ................................................................. 137

5.1.2 Decreasing aerosol water is consistent with organic carbon trends in the southeast United States ...................................................................................... 138

5.1.3 Reconciling satellite aerosol optical thickness and surface fine particle mass through aerosol liquid water................................................................. 139

5.2 Future Directions and Implications .......................................................................................... 140

5.2.1 Hygroscopicity Intercomparison ...................................................................................... 140

5.2.1 Model Validation .............................................................................................................. 143

5.2.3 Global Aerosol Water and Temperature Trends ............................................................. 145

5.2.3 Epidemiological Study ................................................................................................. 146

5.3 References .................................................................................................................................. 147

Appendix A: Supporting Information for Chapter 1 ............................................................................ 155

Appendix B: Supporting Information for Chapter 2 ............................................................................ 156

Appendix B1. Meteorological Differences in Diel Cycle of Aerosol Water .... 156

Appendix B2. SMPS Volume Validation .................................................................................... 156

Appendix B3. Analysis Without Modal Fit .................................................................................... 157

Appendix B4. Relative Humidity Dependence of Kappa ............................................................. 158

Appendix B5. References ................................................................................................................. 159

Appendix C: Supporting Information for Chapter 3 ............................................................................ 167

Appendix D: Supporting Information for Chapter 4 ............................................................................ 177
Appendix E: R Codes for Chapter 1 ................................................................. 178
Appendix F: R Codes for Chapter 2 ............................................................... 186
Appendix G: R Codes for Chapter 3 .............................................................. 209
Appendix H: R Codes for Chapter 4 .............................................................. 232
LIST OF TABLES

Table 2-1. Mean and quartiles of number concentration $N_{conc}$, relative humidity $RH$, temperature ($T$), particle phase water volume $V_w$, semi-volatile volume $V_{sv}$ lost during drying, volume of solutes $V_s$, $\kappa_{vol}$, $\kappa_D1$, $\kappa_D2$, and $gf_{vol}$ ........................................................ 70

Table 2-2. Parameters describing two-mode distributions of number concentration, surface area, and volume for the unperturbed ($C_{unperturbed,i}$), dry-humidified ($C_{dry-humidified,i}$), and dry states ($C_{dry,i}$) ........................................................................................................ 71

Table 2-3. Table of sensitivity ratios ........................................................................................................ 72

Table 3-1. Linear trend analysis (lm function in R) for a) mean annual averages of parameters with time and b) mean annual averages of water ........................................................................................................... 107

Table A-1. Average values of parameters at locations used to estimate aerosol water for AMS studies .......................................................... 155

Table B-1. Mean and quartiles of parameters for unfitted data ........................................ 161

Table B-2. Solute concentrations for E-AIM model initialization ........................................ 161

Table C-1. Annual means of data. .................................................................................. 167
LIST OF ILLUSTRATIONS

Figure 1-1. Overview of thesis proposal ................................................................. 31

Figure 1-2. Map of aerosol species for urban, urban downwind, and rural sites .......... 32

Figure 1-3. Bar chart of aerosol species mass concentrations for urban, urban downwind, and rural sites ................................................................. 33

Figure 2-1. Instrument setup schematic ................................................................. 73

Figure 2-2. Example time series for 8 June and 14 July 2013 .............................. 74

Figure 2-3. Example data from 5 July 2013 ......................................................... 75

Figure 2-4. Time series of 1) temperature ($T$, °C) in brown, 2) dew point temperature ($T_{dew}$, °C), 3) relative humidity ($RH$, %), 4) aerosol dry volume ($V_d$, µm$^3$ cm$^-3$), 5) volumetric growth factor ($gf_{vol}$, dimensionless), 6) volumetric $\kappa$ ($\kappa_{vol}$, dimensionless), and 7) particle phase liquid water ($V_w$, µm$^3$ cm$^-3$) ............................................................. 76

Figure 2-5. Summary number concentration, surface area, and volume distributions for 3 June to 15 July 2013 ................................................................. 77

Figure 2-6. Campaign-averaged diel profiles of total particle phase liquid water mass concentrations and $RH$ binned into hourly intervals ................................. 78

Figure 2-7. Diel profile of volumetric growth factor binned into hourly intervals ...... 79

Figure 2-8. Diurnal trends of smaller diameter mode, larger diameter mode, and volumetric $\kappa$ ........................................................................................................ 80

Figure 2-9. Diurnal trend of semi-volatile volumes lost during drying binned into hourly intervals ..................................................................................... 81

Figure 2-10. Diurnal trend of select aerosol chemical species measured by PILS-IC and PILS-WSOM ................................................................................. 82
Figure 2-11. Diurnal trend of dry aerosol volumes binned into hourly intervals .......... 83

Figure 3-1. Local regression (LOESS) analysis of aerosol liquid water estimates and organic carbon measurements for all SEARCH sites from 2001-2012 ......................... 108

Figure 3-2. LOESS trends in a) temperature, b) relative humidity, c) sulfate, d) nitrate, e) ammonium, and f) organic carbon for urban and rural sites ................................. 109

Figure 3-3. LOESS monthly trends in aerosol liquid water in the southeast U.S. from 2001-2012 for $80 \leq RH < 85$, $85 \leq RH < 90$, and $90 \leq RH < 95$ ................................. 110

Figure 3-4. Mean annual PM$_{2.5}$ aerosol compositional mass concentrations for SEARCH sites from 2001-2012 using estimated inorganic ion derived water content .................. 111

Figure 3-5. Mean annual averages of estimates of liquid water content from organic fraction for overall, urban, and rural data ................................................................. 112

Figure 4-1. Seasonal differences in aerosol water estimated using temperature and RH data for surface and surface to boundary layer pressure ............................................. 130

Figure 4-2. Time series of quartiles and medians for aerosol water estimated using surface temperature and RH reanalysis data for 1988 – 2013 ....................................... 131

Figure 4-3. Spatial map of seasonal differences in surface organic matter for (a) 2000-2007 and (b) 1988 – 2013 .............................................................................................. 132

Figure 4-4. Time series of quartiles and medians for surface measurements of organic matter for 1988 – 2013 ...................................................................................... 133

Figure 4-5. Time series of quartiles and medians for surface measurements of inorganics for 1988 – 2013 .............................................................................................. 134

Figure 4-6. Time series of quartiles and medians for surface relative humidity for 1988 – 2013 .............................................................................................. 135
Figure 4-7. Time series of quartiles and medians for estimated extinction coefficient for 1988 – 2013 ................................................................. 136

Figure 5-1. Intercomparison results of $\kappa$ for amongst optical by GA Tech, ARA, and NOAA/CIRES and particle sizing methods by Rutgers and EFPL ........................................... 150

Figure 5-2. Diurnal profile of $\kappa$ during SOAS ..................................................... 151

Figure 5-3. Diurnal profile of $\kappa$, June 7-13, 2013 ............................................... 152

Figure 5-4. Boxplots of diurnal trends in aerosol water for ISORROPIA semi-observational estimates and SOAS hygroscopicity data from Nguyen et al., 2014 adjusted to SEARCH $RH$ ............................................................................................................. 153

Figure 5-5. Global changes in surface temperature using NASA GISS data for the month of September from 1900-2015 ...................................................................................... 154

Figure B-1. Diurnal profile of $RH$ and aerosol water for 21 June 2013, a day with no rainfall ........................................................................................................................................ 162

Figure B-2. Comparison of particle-phase liquid water for a 5 day period with heavy rainfall versus no rainfall .................................................................................................................. 163

Figure B-3. Time series of TEOM mass and SMPS dry volume ............................... 164

Figure B-4. Scatterplot of TEOM particulate matter mass and SMPS dry volume ............................................................................................................................ 165

Figure B-5: Variation of the E-AIM Model III derived osmotic coefficient with water activity .................................................................................................................. 166

Figure C-1. Locations of SEARCH monitoring sites ............................................. 168

Figure C-2. LOESS estimates of trends for water and organic carbon ...................... 169

Figure C-3. LOESS estimated monthly trends in SO$_4$ and NH$_4$ .......................... 170
Figure C-4. GGplot2 trends in ambient isoprene .......................................................... 171

Figure C-5. LOESS trends of ozone concentrations ..................................................... 172

Figure C-6. GGplot2 trends in acidity index ............................................................... 173

Figure C-7. Ggplot2 trends for organic carbon to inorganic ratios .............................. 174

Figure C-8. LOESS trends of OC and SOC ............................................................... 175

Figure C-9. LOESS estimated monthly trends in water, OC, and SOC ....................... 176

Figure D-1. Map of IMPROVE sites divided up into six regions .............................. 177
CHAPTER 1. INTRODUCTION

1.1 Motivation

Atmospheric aerosols affect human health and welfare, global climate, visibility, and ecosystems. Aerosols are directly emitted and form through a myriad of chemical reactions involving a variety of anthropogenic and biogenic precursors (Kanakidou et al., 2005; Hallquist et al., 2009). Consequently, atmospheric aerosols consist of a mix of chemically diverse compounds ranging a spectrum of volatility. Semi-volatile compounds in the atmosphere, including water, undergo equilibrium partitioning between the condensed and gaseous phases (Murphy et al., 1998; Pöschl, 2005; Robinson et al., 2007). Particle phase liquid water and its interactions with aerosol chemical components are associated with many environmental effects: acid deposition (Calvert et al., 1985), impaired visibility through light scattering (e.g. Malm et al., 1994; Park et al., 2004a; Pitchford et al., 2007), regional climate through effects on aerosol optical depth (Pilinis et al., 1995, Leibensperger et al., 2012), and climate through effects on cloud condensation and ice nuclei (e.g. Cruz and Pandis, 1997; Pöschl, 2005). Quantitative characterization of ambient aerosol size, mass, and chemical composition, including water content, is essential to understand the fate and transport of chemicals in the Earth’s atmosphere, and to develop effective strategies that mitigate aerosol-related problems. Furthermore, anthropogenic-induced climate change has brought about a need to better understand the processes by which human activity influences biosphere-atmosphere interactions. In particular, the impact of anthropogenic pollutants on the fate of biogenic volatile organic compounds is poorly constrained. It has been well established that anthropogenic pollution facilitates biogenic secondary organic aerosol (SOA) formation (Kroll et al.,
2005a; Lane et al., 2008; De Gouw and Jimenez, 2009; Hoyle et al., 2009; Carlton et al., 2010; Spracklen et al., 2011; Shilling et al., 2013). Biogenic volatile hydrocarbon emissions can react with emissions from human activity to alter the oxidative capacity of the atmosphere, impact the global carbon cycle, and form more functionalized semi-volatile hydrocarbon products that can absorb into the organic or aqueous aerosol phase to form SOA (Pankow, 1994; Odum et al., 1996; Seinfeld and Pankow, 2003; Donahue et al., 2006).

1.2 Hypothesis and Specific Aims

To improve the understanding of the effects of anthropogenic perturbations on aerosol liquid water and biogenic secondary organic aerosol formation, I 1) measure, co-located with a variety of chemical gas phase and aerosol phase measurements, in situ aerosol water content to determine the chemical and thermodynamic controls on particle phase liquid water content 2) conduct a multivariate analysis of temporal and spatial trends of aerosol water to quantify anthropogenic effects on aerosol liquid water in the Southeast and relate to organic particulate matter (PM) mass trends in the region, and 3) analyze aerosol water at the surface and aloft and qualitatively reconcile differences between water and satellite data with spatial trends to better understand potential impacts on air quality and climate across the continental U.S. (Figure 1-1).

I hypothesize that particle phase liquid water is regulated by anthropogenic emissions and facilitates biogenic secondary organic aerosol (SOA) formation. To support this hypothesis, I 1) directly measure particle phase liquid water and assess whether there is a positive relationship with aerosol constituents from anthropogenic sources (e.g., sulfate and nitrate), 2) evaluate historical trends in particle water relative to
changes in anthropogenic particle mass concentrations for hygroscopic species and demonstrate a relationship with organic aerosol in the context of a plausible biogenic SOA formation mechanism, 3) investigate geospatial and seasonal trends in particle phase water near the Earth’s surface qualitatively estimated by satellite aerosol optical thickness (AOT) measurements across the United States, with focus on areas where biogenic SOA mass concentrations are highest.

1.3 Background: Atmospheric Aerosols

Atmospheric particulate matter is either directly emitted (primary) or form in the atmosphere through chemical reactions involving a variety of anthropogenic and biogenic precursors (secondary) (Seinfeld and Pandis, 2006). Increased mortality and morbidity in communities is associated with elevated atmospheric particulate matter concentrations and there is no consensus regarding a safe level (Dockery and Pope, 1994; Samet et al., 2000; Smith et al., 2000). Historical episodes of extreme air pollution have resulted in many mortalities and illnesses. An episode in Meuse Valley Belgium in 1930 led to 60 deaths (Nemery et al., 2001) and the Great London Smog of 1952, which led to estimated deaths ranging from 4000 (Logan, 1953) to 12,000 (Bell et al., 2004) deaths. Adverse effects from breathing airborne PM include respiratory symptoms of asthmatics and reduced lung function (Michaels and Kleinman, 2000). Thus, a more detailed understanding of the sources and chemical composition of PM will help illuminate the necessary controls needed to reduce mortality associated with air pollution.

Fine particulate matter (PM$_{2.5}$) that contributes to atmospheric pollution is composed of a mixture of different compounds that come from a variety of sources. These speciated components include water-soluble inorganic species such as sulfate
(SO$_4^{2-}$), nitrate (NO$_3^-$), and ammonium (NH$_4^+$). The main sources of sulfate and nitrate aerosols, for example, are atmospheric oxidation of SO$_2$ to H$_2$SO$_4$ and nitrogen oxides (NO$_x$) and HNO$_3$ respectively (Seinfeld and Pandis, 2006). The gaseous precursors originate primarily from combustion (Liao et al., 2003; Liao et al., 2004). Other sources of SO$_2$ include volcanic emissions (Kiehl and Rodhe, 1995) and the oxidation of dimethyl sulfide (DMS) from oceans (Liao et al., 2003) while other sources of NO$_x$ include lightning, soils, and wildfires. Ammonium can form from the neutralization of sulfate to ammonium sulfate (NH$_4$)$_2$SO$_4$ and the neutralization of nitrate to ammonium nitrate (NH$_4$)NO$_3$ if excess ammonia (NH$_3$), which can originate from fertilizer and biological sources, is available. Other aerosol components include organic condensates from oxidation of gaseous emissions and lower volatility hydrocarbons from primary combustion sources, elemental or black carbon from fuel combustion. Though widely acknowledged to occur, the degree to which man-made pollution alters biogenic emissions, fluxes and their ultimate fate remains poorly understood.

1.4 Background: Aerosol Water

Conceptually, the atmospheric aerosol particle consists of three constituents: the dry component, which consists of a mix of low-volatility sulfates, nitrates, elemental carbon, and organic compounds, and semi-volatile compounds that partition between the condensed and gaseous phases, and water, which is also semi-volatile. Water is an abundant atmospheric constituent that is present in the condensed phase as a function of relative humidity (RH), temperature, aerosol concentration and chemical composition (Zhou et al., 2011). Aerosol water also impairs visibility by affecting aerosol light scattering, extinction coefficients, and aerosol optical depths. Poor summertime visibility
in the eastern U.S. is primarily due to high sulfate exposed to high $RH$ (Malm et al., 1994; Park et al., 2004b; Pitchford et al., 2007).

The aqueous phase also provides a medium for the partitioning of polar, water-soluble gas phase species (Asa-Awuku et al., 2010), thus potentially facilitating SOA formation (Blando and Turpin, 2000; Carlton et al., 2009; Ervens et al., 2011). In the traditional SOA formation pathway (Pankow, 1994; Odum et al., 1996), gas-phase oxidation yields semi-volatile compounds that can partition onto the dry, organic aerosol matrix. More recently, it was discovered that small, water-soluble gases can undergo aqueous-mediated chemistry in clouds, fogs, and wet aerosols and form low volatility compounds that remain in the particle phase and contribute to the atmospheric PM burden (e.g. Blando and Turpin, 2000; Ervens et al., 2008; Lim et al., 2010; Ervens et al., 2011). It is important to more fully comprehend water-mediated processes because the majority of organic compounds in the atmosphere are small, water-soluble gases (Grosjean et al., 1996; Nolte et al., 2001). Cloud processing has been hypothesized to be an important source of SOA (Blando and Turpin, 2000; Warneck, 2003; Ervens et al., 2004; Lim et al., 2005), in which reactive organics oxidize in the interstitial spaces of clouds to form highly water-soluble compounds (e.g., aldehydes) that readily partition into cloud droplets, which then oxidize further to form less volatile organics (e.g. carboxylic acids and oligomers) (Altieri et al., 2006; Carlton et al., 2006; Carlton et al., 2007; Altieri et al., 2008). In general, the relative contributions of cloud and aerosol aqueous SOA to total SOA mass depend on the time scales air parcels are exposed to supersaturated (clouds) and subsaturated (aerosols) conditions. The processing time in aerosol water is typically longer (timescale of hours to days) than processing in cloud
water (timescale of minutes), suggesting that the vast majority of aqueous SOA may form in aerosol water and be comprised of products with higher carbon number than their precursors (Ervens et al., 2011). In the Southeast U.S. in particular, isoprene epoxydiols (IEPOX)-derived SOA (Chan et al., 2010; Froyd et al., 2010; Surratt et al., 2010; McNeill et al., 2012; Pye et al., 2013) has been observed to account for about 28% (Budisulistiorini et al., 2015) to 33% (Budisulistiorini et al., 2013) of the total fine organic aerosol mass and this constituent can only form if water is present in the condensed phase (Nguyen et al., 2014a). IEPOX-derived SOA is formed from the ring-opening of the epoxide group, which can be activated by a proton transfer from a strong acid such as sulfuric acid, and then followed by nucleophilic addition of available nucleophiles in the condensed phase (Minerath et al., 2008; Eddingsaas et al., 2010; Surratt et al., 2010). Thus IEPOX SOA chemistry is enhanced in the presence of acidified sulfate seed aerosol and condensed-phase nucleophiles (Surratt et al., 2010; Nguyen et al., 2014a). As formation of IEPOX SOA, a major component of organic aerosol in the Southeast, requires water and aerosol water further provides the medium for aqueous partitioning of organic gases to form SOA, we hypothesize that in the Southeast U.S., particle-phase liquid water is anthropogenic and facilitates biogenic SOA formation.

The co-emission of biogenic volatile organic compounds (BVOCs) and water vapor, which are functions of temperature, multiphase SOA formation as a source of aerosol mass, and the influence of aerosols on surface temperatures, are all linked (Hennigan et al., 2008b). In regions of BVOCs emissions, water-soluble compounds (e.g., formic acid, bulk water soluble organic compounds) readily partition to particle water, whereas in regions dominated by anthropogenic volatile organic compounds
(VOCs), the partitioning preference is to the organic aerosol matrix (Hennigan et al., 2009; Liu et al., 2012; Zhang et al., 2012). Since aerosol water serves as a medium to partition polar, water-soluble gas phase organic species (Asa-Awuku et al., 2010; Prisle et al., 2010), there is a critical need to understand the relationship between anthropogenic influences on liquid water and subsequent biogenic SOA formation. Conversely, relationships between aerosol water and organic species are poorly understood, and there are few field measurements linking the two.

Liquid water is estimated to represent a substantial fraction of total tropospheric aerosol volume at $RH > 85\%$ (Kreidenweis et al., 2008). Water is predicted to exceed total aerosol dry mass by 2 to 3 times globally (global average 15 mg m$^{-2}$ liquid water vs. 6.5 mg m$^{-2}$ dry aerosol mass) (Liao and Seinfeld, 2005; Lee and Adams, 2010). This suggests that atmospheric water is more accessible than condensed organic matter for partitioning of gas phase organic species. Despite the abundance and importance of aerosol water, it is not routinely measured, actual mass concentrations are not well known, and model predictions are poorly constrained. One form of measurement, the hygroscopic tandem differential mobility analyzer (HTDMA) (Liu et al., 1978; Rader and McMurry, 1986), is often used in laboratory experiments to analyze the hygroscopicity of compounds. In the field, most studies also use HTDMAs to establish links between inorganic species, particle hygroscopicity, ambient relative humidity, and condensed phase liquid water (e.g. McMurry and Stolzenburg, 1989; Berg et al., 1998). Swietlicki et al. (2008) synthesized a review of these HTDMA measurements. These relationships are included in thermodynamic modules of atmospheric chemistry models, but there are few field measurements focused on the organic constituent and aqueous partitioning (Zhang
et al., 1993; Dick et al., 2000). Another recent method for characterizing water is the Dry-Ambient Aerosol Size Spectrometer (DAASS) (Stanier et al., 2004), which was employed for a year in the Pittsburgh Air Quality Study (PAQS) (Khlystov et al., 2005). The DAASS measures water content and volumetric growth factor, but does not account for aqueous partitioning and effects on biogenic SOA formation. In Chapter 2, I use a DMA system to quantify ambient mass concentrations of particle-phase liquid water in the Southeast U.S.

The question arises as to how anthropogenic activity influences aerosol water, and how these interactions affect or have historically affected the partitioning of organic gases to the condensed phase and subsequent biogenic SOA formation, which is what this work addresses. Effective evaluation of particle chemical composition and size in relation to water uptake will therefore provide critical insight into the effects of biogenic and anthropogenic emissions on multiphase organic chemistry, as explored in Chapter 2. Chapter 3 then addresses how aerosol water may be affected by anthropogenic emissions. To do this, I will, for the first time in this field, construct a historical trends analysis of aerosol liquid water mass concentrations, compare them with anthropogenic ion mass concentrations, and then provide a plausible explanation for organic aerosol trends in the Southeast U.S. Chapter 4 will then provide a response to the ongoing debate in the literature about why seasonal differences in aerosol optical thickness (AOT) is strongest over the Southeast U.S. I will analyze trends in aerosol water mass concentrations across the continental U.S. to explore geospatial and temporal trends in regional seasonality of aerosol water mass concentrations. All of this work will allow for more accurate physicochemical modules in model predictions and enable more effective strategies of air
quality and climate mitigation.

1.5 Aerosol Water Ubiquity and Abundance

Aerosol liquid water is a ubiquitous aerosol constituent. Field measurements confirm the ubiquity of the metastable state (Martin et al., 2008; Nguyen et al., 2014b), in which water is always present. In certain photochemically active, humid locations influenced by anthropogenic emissions (e.g., the eastern U.S. (Carlton and Turpin, 2013)), aerosol water is an abundant medium available to partition polar, water-soluble gas phase organic gases to the condensed phase (Asa-Awuku et al., 2010; Parikh et al., 2011) to facilitate SOA formation (Carlton et al., 2009; Ervens et al., 2011; Carlton and Turpin, 2013; Hodas et al., 2014). Recently noted improved visibility (Attwood et al., 2014) and reductions in organic carbon mass in the southeast U.S. can be mechanistically linked to decreasing trends in aerosol water mass concentrations (Nguyen et al., 2015; Chapter 3).

Despite the importance of aerosol water, actual geospatial and temporal trends in mass concentrations are not well known. Routine surface mass networks (e.g., Interagency Monitoring of Protected Visual Environments (IMPROVE)) and most particle measurement techniques (e.g., the aerosol mass spectrometer (AMS)) make measurements at a prescribed RH to standardize the contribution of aerosol water. Historically, laboratory and field investigations of aerosol water were typically motivated by visibility concerns and focused on characterizing particle growth factors as a function of RH for single particles and the bulk aerosol population (Zhang et al., 1993; Pitchford and Mcmurry, 1994; Malm et al., 1996; Dick et al., 2000; Kroll et al., 2005b; Hennigan et al., 2008a; Sorooshian et al., 2008). The goals of these projects were to predict visibility
and to develop control strategies for visibility improvement in Class I areas. The projects were performed either by calculations of scattering efficiencies using Mie theory (e.g. Ouimette et al., 1981; Hasan and Dzubay, 1983; Sloane, 1983; Sloane and Wolff, 1985; Zhang et al., 1994) or by regression-derived scattering efficiencies for each particle species (e.g. Malm et al., 1996; Sisler and Malm, 2000; Malm et al., 2003; Park et al., 2006; Pitchford et al., 2007). Water uptake by organics has also been studied (e.g. Dick et al., 2000; Kreidenweis et al., 2006) but this process is neglected in the inorganic thermodynamics model ISORROPIA (Fountoukis and Nenes, 2007) used to estimate aerosol water mass concentrations because organic contributions are complex and difficult to quantify.

Typically, techniques to measure aerosol hygroscopic properties examine how particle growth responds to changes in RH relative to a dry reference state (e.g., ~5% RH), and does not explicitly measure water content of the unperturbed aerosol at ambient RH (Dick et al., 2000). Field studies of aerosol hygroscopicity, from which water can be inferred, are conducted and provide valuable knowledge. For example, the Southeastern Aerosol and Visibility Study (SEAVS) in the Great Smoky Mountains measured aerosol growth as a function of RH and found water uptake was more than predicted due to inorganic species alone and the excess was positively associated with particle organic content (Dick et al., 2000). However, these field studies are not routine, have primarily been short term and regional in scope (Dick et al., 2000; Khlystov et al., 2005; Levin et al., 2013; Mei et al., 2013; Levin et al., 2014; Nguyen et al., 2014b). Insufficient knowledge of aerosol liquid water can lead to a misunderstanding regarding atmosphere-biosphere interactions, the fate and transport of trace species in the atmosphere, and can
hinder development of effective control strategies to mitigate water-related PM impacts on air quality, climate, visibility, and human health and welfare.

Chemically characterized particle measurements with the AMS highlight the ubiquity and dominance of oxygenated organic species in aerosols in the anthropogenically-influenced Northern Hemisphere mid-latitudes (Zhang et al., 2007), and in Southern Hemisphere locations (Chen et al., 2009; Martin et al., 2010; Poschl et al., 2010; Carbone et al., 2013; Tiitta et al., 2014). In this work, we estimate aerosol water mass concentrations from speciated AMS measurements at field campaign locations for which we can readily estimate campaign average meteorological conditions. Explicit understanding and constraint in atmospheric models of aerosol water are essential. Uncertainties in the magnitude and direction of the direct and indirect effect of aerosols remain an obstacle for reliable climate change prediction, especially when aerosol water remains largely unmeasured and model predictions poorly constrained. A global perspective of the presence of aerosol water represents a key knowledge gap.

To explore geospatial patterns in aerosol water, we estimate mass concentrations at locations with consistent submicron aerosol composition measurements worldwide. We estimate and compare, in a relative sense, average aerosol water concentrations for 21 field campaigns with the thermodynamic model ISORROPIAv2.1 (http://isorropia.eas.gatech.edu) using AMS particle mass concentrations of inorganic species (NH$_4^+$, SO$_4^{2-}$, NO$_3^-$) from Zhang et al. (2007) and other datasets (Chen et al., 2009; Martin et al., 2010; Poschl et al., 2010; Carbone et al., 2013; Tiitta et al., 2014) in the AMS Global Database (https://sites.google.com/site/amsglobaldatabase/), and RH and temperature data from the Climate Forecast System Reanalysis (CFSR;
ISORROPIA assumes thermodynamic equilibrium between the gaseous and condensed phases of inorganic species and water, and a \( \text{NH}_4^+-\text{SO}_4^{2-}-\text{NO}_3^- \) metastable system. The model has been shown to be valid within the errors of field measurements (Moya et al., 2001; Zhang et al., 2002; Yu et al., 2005). The AMS study locations take place primarily during the summer months, which are June, July, and/or August for Northern Hemisphere (NH) and December, January, February for Southern Hemisphere (SH) locations. The exceptions are Pittsburgh (September; NH), Manaus (February and March; SH), Mexico City (April and May; NH), Edinburgh (November; NH), Weybourne (April and May; NH), Duke Forest (September; NH), Hyytiala (March and April, NH), Hohenpeißenberg (May, NH), Santiago (August through November; SH), and Welgegund (September 2010 through August 2011; SH).

Uncertainties in retrieved meteorology data for the study locations and time periods will affect the magnitudes of estimated aerosol water mass concentrations. CFSR 1000 mb temperature and 2 m \( RH \) at 0.5 x 0.5 degree spatial and 6-hour temporal resolutions are averaged and paired in space and time for each of the respective AMS studies. As surface and 2 m temperature is unavailable in CFSRv.1, this may create an isobaric error in temperature that will slightly reduce the estimated magnitude of aerosol water mass concentrations at mountainous sites such as Storm Peak and Jungfraujoch. Furthermore, a slight inconsistency arises for Welgegund where observations run from September 2010 to August 2011. CFSRv.1, which ran from 1979 through the end of 2010, was replaced by CFSRv.2 beginning in 2011, and 2 m \( RH \) became unavailable in
Thus beginning in 2011, $RH$ for Welgegund was retrieved at 1000 mb level instead, and the difference in average $RH$ between the two time periods is minimal (3%).

Another limitation is this application of ISORROPIA does not consider water uptake by organic compounds, and laboratory studies demonstrate that organic constituents can modulate particle hygroscopicity depending on the chemical composition (Saxena et al., 1995; Cruz and Pandis, 2000; Dick et al., 2000; Kreidenweis et al., 2008; Suda et al., 2012; Sareen et al., 2013). The effects of organic compounds on aerosol water are complex (Decesari et al., 2000; Shimmo et al., 2004) and water uptake is typically assumed to be additive. Thus these semi-observational estimates by ISORROPIA likely provide a lower bound for aerosol water mass concentrations, though there is uncertainty. The Zdanovskii-Stokes-Robinson (ZSR) mixing rule, however, can be used to describe hygroscopic growth of most aerosol mixtures, including for organic compounds (Choi and Chan, 2002; Svenningsson et al., 2006; Petters and Kreidenweis, 2007). We estimate potential addition of water from the organic aerosol components according to $\kappa$-Kohler theory and the ZSR mixing rule, as described in Petters and Kreidenweis (2007) and Kreidenweis et al. (2008):

$$V_{w,o} = V_o \kappa_{org} \frac{a_w}{1-a_w}, \tag{1}$$

where $a_w$ is water activity (dimensionless), $V_{w,o}$ and $V_o$ are the volumes of water and organic matter ($\mu$m$^3$ cm$^{-3}$), and $\kappa_{org}$ is the hygroscopicity parameter of the organic component (dimensionless). $V_o$ is calculated by dividing the mass of organic matter by a typical organic density of 1.4 g cm$^{-3}$ (Turpin and Lim, 2001). For simplicity and due to a lack of data on particle diameter, we assume $a_w$ is equivalent to $RH$. The expected difference between $RH$ and water activity is between 0.01 and 0.02 in absolute units.
(Tang, 1996), and this may result in an RH dependent overestimate in κ ranging from 4 to 11% (Nguyen et al., 2014b).

Hygroscopicity of the organic fraction ($\kappa_{org}$) depends on particle chemical composition and can vary with location. Historical $\kappa_{org}$ data for the AMS study site locations are not available. Previous studies suggest that typical $\kappa_{org}$ values are ~0.1 (Prenni et al., 2007; Engelhart et al., 2008; Wex et al., 2009; Engelhart et al., 2011; Pierce et al., 2012) although individual species show a range when separated by polarity (Suda et al., 2012). With increasing distance from pollution sources, the initially hydrophobic particles become more hygroscopic due to oxidative processes in the atmosphere (Moffet and Prather, 2009; Ervens et al., 2010). Thus we estimate water contribution by organic compounds using $\kappa_{org}$ values of 0.08, 0.11, and 0.13 for urban, urban downwind, and rural locations, respectively, to correspond to increasing distance from pollution sources. The value of $\kappa_{org} = 0.13$ for rural sites is consistent with $\kappa_{org}$ estimated at a rural, forested, and mountainous site in Colorado (Levin et al., 2014).

Globally, semi-empirical observational estimates of aerosol water mass concentrations vary by amount and fractional contribution to total aerosol mass (Figures 1-2 and 1-3, Table A-1). As the AMS studies take place during different years and different seasons, it is important to note that the presented semi-observational estimates of aerosol water for each region are for a particular time period and are meant to present the ubiquity and dominance of aerosol water rather than a direct comparisons of actual concentrations among different locations. ISORROPIA estimates aerosol water mass concentration is highest in humid urban areas with high concentrations of hygroscopic aerosol constituents such as sulfate. Sulfate can form in urban and urban downwind
locations from oxidation of sulfur dioxide emissions from anthropogenic energy sectors such as fuel combustion. The Beijing study has the highest estimated water concentrations at 78 \( \mu \text{g m}^{-3} \), while the Chebogue study in rural Nova Scotia has the lowest estimated water concentrations at 0.35 \( \mu \text{g m}^{-3} \).

The regional contrast in aerosol water mass concentrations demonstrates how complex interactions among meteorology, aerosol mass concentration, and chemical composition, in particular the presence of anthropogenic hygroscopic constituents, control aerosol water. The average water concentrations for urban, urban downwind, and rural areas are estimated to be 12, 11, and 3 \( \mu \text{g m}^{-3} \) respectively. The water to dry mass ratios demonstrate a near opposite pattern, at 0.46, 1.14, and 1.07 respectively. Aerosol water fraction is highest during sampling in Hyytiala, a rural area in Finland with the highest RH (97%) of all the sites, at 79% of total aerosol mass. Water fraction is lowest for Riverside, an urban area in California, at 3.2% of total aerosol mass. One interesting note is that Edinburgh is an urban area and yet its water mass concentration during the study period is relatively low at 1.2 \( \mu \text{g m}^{-3} \) (or 2.9% of total mass). This is due primarily to low concentrations of sulfate (0.52 \( \mu \text{g m}^{-3} \), a highly hygroscopic aerosol constituent despite relatively high RH (78%). In contrast, New York City and Pittsburgh, both urban areas with relatively hygroscopic aerosol (3.9 and 7.0 \( \mu \text{g m}^{-3} \) sulfate respectively), contain low aerosol water concentrations (0.8 and 78.1 \( \mu \text{g m}^{-3} \) respectively) due to low average RH during the time period of study (21% and 42% respectively). Additionally, the AMS study at Duke Forest, a rural site located in the southeast U.S. with an average RH of 53%, contains 0.7 \( \mu \text{g m}^{-3} \) aerosol water, nearly double the amount of water (0.4 \( \mu \text{g m}^{-3} \)) at Manaus, a rural site in Brazil with an average RH of 76%. Although the Pittsburgh
study has lower $RH$ than the Edinburgh study and the Duke Forest study has lower $RH$ than the Manaus study, Pittsburgh and Duke Forest have more water due to more hygroscopic particle constituents, a consequence of a higher sulfate mass concentration arising from fossil fuel combustion.

Organic aerosol constituents vary in hygroscopicity and can add water to atmospheric aerosol depending on organic mass concentrations, types of organic compounds, and meteorology (Figure 1-3, Table A-1). When aerosol water sensitivity to organic compounds is considered, AMS-measured aerosol in the Beijing study maintains the highest estimated water concentrations of all sampled locations at 86 $\mu$g m$^{-3}$, and the Chebogue study remains the lowest at 0.42 $\mu$g m$^{-3}$. When organic compound hygroscopicity is considered, the average water concentrations for urban, urban downwind, and rural areas are estimated to be 14, 15, and 3.2 $\mu$g m$^{-3}$ respectively. The water to dry mass ratios are 0.53, 1.3, and 1.4 respectively. Aerosol water fraction is once again highest for the Hyytiala study period, at 85% of total aerosol mass, and again lowest for Riverside at 3.7% of total aerosol mass. Water mass concentrations increase due to consideration of organic water at 21%, 29%, and 31% for urban, urban downwind, and rural areas, in order of increasing $\kappa_{\text{org}}$ values for the location categories. Average water concentrations when $\kappa_{\text{org}}$ is considered, however, are highest in urban downwind areas, contrasting with the maximum average in urban areas when only inorganic water is calculated. This arises from a combination of higher $\kappa_{\text{org}}$ at urban downwind areas than at urban areas, and of higher organic compound mass concentration at urban downwind areas than at rural areas (9.8, 5.8, and 1.4 $\mu$g m$^{-3}$ for urban, urban downwind, and rural sites, respectively). Organic water absolute mass concentration is highest in Beijing at 8.4
µg m⁻³, where organic compound concentration is highest at 25 µg m⁻³. Organic water fractional contribution to total mass is the highest in Hyytiala at 26.3%. Aerosol water mass concentrations attributed to inorganic compounds (29% of total aerosol mass on average) far exceed water mass concentrations attributed to organic compounds (7.4% of total aerosol mass on average) for all study times and locations. Thus geospatial patterns in aerosol water do not change due the addition of estimated organic water.

Aerosol concentration and chemical speciation play critical roles that vary regionally and control water mass concentrations and its impacts. Consideration of local meteorology alone is insufficient to properly characterize aerosol water. This work suggests the need for more detailed studies of water and how it affects and is affected by aerosol chemistry. Understanding the role of aerosol in the atmosphere and how it varies in response to emissions from human activity is a critical concern in the consideration of future air quality in a changing global climate, and within the context of energy needs and choices. Water vapor in the atmosphere is predicted to increase due to enhanced rates of evapotranspiration in a warmer world (Boucher et al., 2004) and global energy needs will continue to increase with the growing population and increased energy use in developing areas. These factors may increase the amount of aerosol water that modulates the fate, transport, and transformation of trace species, in particular polar, water-soluble organic compounds that represent a major fraction of the global atmospheric organic gas burden.(Nolte et al., 2001) The resulting pollutant precursor emissions arising from energy decisions and environmental policies represent key uncertainties that complicate future prediction of aerosol hygroscopicity and subsequently aerosol water impacts.
1.6 References


Figure 1-1. Overview of thesis proposal
**Figure 1-2.** Aerosol species for urban (blue), urban downwind (black), and rural (pink) sites. Fractional species are sulfate (red), nitrate (dark blue), ammonium (orange), organic matter (green), and water (light blue). Absolute water mass concentrations by inorganic compounds as estimated by ISORROPIA are indicated in blue text.
Figure 1-3. Aerosol species mass concentrations for urban (blue text locations), urban downwind (black text locations), and rural (pink text locations) sites. Fractional species are sulfate (red), nitrate (dark blue), ammonium (orange), organic matter (green), inorganic water (light blue), and organic water (darker light blue). Organic water is calculated using an assumed hygroscopicity of 0.08, 0.11, and 0.13 for urban, urban downwind, and rural locations, respectively.
CHAPTER 2. TRENDS IN PARTICLE-PHASE LIQUID WATER DURING THE SOUTHERN OXIDANT AND AEROSOL STUDY

2.1 Abstract

We present in situ measurements of particle phase liquid water. Measurements were conducted 3 June to 15 July 2013 during the Southern Oxidant and Aerosol Study in the southeast U.S. The region is dominated by biogenic emissions, impacted by anthropogenic pollution, photochemically active, humid, and known to contain high concentrations of organic aerosol mass. Measurements characterized mobility number size distributions of ambient atmospheric aerosols in three states: unperturbed, dry, and dry-humidified. Unperturbed measurements describe the aerosol distribution at ambient temperature and relative humidity. For the dry state, the sample was routed through a cold trap upstream of the inlet then re-heated, while for the dry-humidified state the sample was re-humidified after drying. The total volume of water and semi-volatile compounds lost during drying was quantified by differencing dry and unperturbed volumes from the integrated size spectra, while semi-volatile volumes lost during drying were quantified differencing unperturbed and dry-humidified volumes. Results indicate that particle phase liquid water was always present. Throughout the SOAS campaign, median water mass concentrations typically ranged from 1 to 5 µg m⁻³, but were as high as 73 µg m⁻³. On non-raining days, morning time (6am-9am) median mass concentrations exceeded 15 µg m⁻³. Hygroscopic growth factors followed a diel cycle and exceeded two from 7 to 9 a.m. local time. The hygroscopicity parameter kappa ranged from 0.14 to 0.46 and hygroscopicity increased with increasing particle size. An observed diel cycle in kappa is consistent with changes in aerosol inorganic content. Unperturbed and dry-
humidified aerosol volumes did not result in statistically discernible differences, demonstrating that drying did not lead to large losses in dry particle volume. We anticipate that our results will help improve the representation of aerosol water content and aqueous phase mediated partitioning in photochemical models.

2.2 Introduction

Anthropogenic-induced climate change has brought about a need to better understand the processes by which human activity influences biosphere-atmosphere interactions in order to effectively predict future impacts on air quality and the environment. In particular, the effect of anthropogenic pollutants on the fate of biogenic volatile organic compounds has not been well constrained. Biogenic emissions can interact with pollutants to alter the oxidative capacity of the atmosphere (Lelieveld et al., 2008), impact the global carbon cycle (Guenther, 2002), and also to form SOA (Pöschl, 2005; Lelieveld et al., 2008). Since liquid water serves as a medium to partition polar, water-soluble gas phase organic species, there is a critical need to understand the relationship between anthropogenic influences on liquid water and subsequent biogenic SOA formation. Effective evaluation of particle chemical composition and size in relation to water uptake will provide critical insight into the effects of biogenic and anthropogenic emissions on multiphase organic chemistry.

The role of water in SOA formation in a particular environment is complex. In addition to the availability and relative abundance of particle phase liquid water, SOA enhancement through water-mediated pathways depends on the amount of water-soluble gas-phase material (Carlton and Turpin, 2013). For absorptive partitioning, SOA formation is dependent on the activity coefficient of the organics in the solution (Pankow
et al., 2001, Barley et al., 2009). The presence of inorganic material mixed with hydrophobic organic compounds may lead to liquid-liquid phase separation (Bertram et al., 2011) and the presence of liquid-liquid equilibria may influence partitioning (Prisle et al., 2010). Partitioning of water-soluble material is also influenced by the presence of inorganic material through salting effects (Lim et al., 2010; Knote et al., 2014).

Several techniques exist to measure aerosol hygroscopic properties. In general, hygroscopicity measurements characterize the change in aerosol properties in response to perturbations in $RH$. Sorooshian et al. (2008) provide a detailed overview of the different available techniques. Broadly, the approaches can be classified into methods that probe single sizes and methods that probe the entire aerosol. Popular single size methods include the hygroscopicity tandem differential mobility analyzer (HTDMA) technique (Liu et al., 1978; Rader and McMurry, 1986; Suda and Petters, 2013), the laminar flow tube approach (Stratmann et al., 2004; Wex et al., 2005), and the single particle levitation approaches (Tang, 1996; Mitchem and Reid, 2008). The HTDMA technique has been routinely deployed in field experiments, providing rich datasets for particle hygroscopic growth factors in a wide range of environments (e.g. McMurry and Stolzenburg, 1989; Berg et al., 1998; Dick et al., 2000; Swietlicki et al., 2008 (and references therein)). Popular methods that characterize the change in the entire aerosol with humidity include humidified nephelometry (Rood et al., 1985), aerosol hydration spectrometry (Stanier et al., 2004; Hegg et al., 2008; Snider and Petters, 2008; Engelhart et al., 2011), and gravimetric methods (Mikhailov et al., 2013). Single size methods are generally more precise and less ambiguous to interpret relative to bulk techniques. Bulk techniques,
however, are useful because they characterize the entire aerosol rather than a subset and thus are needed to directly measure total water volume.

Here we report measurements obtained with an aerosol preconditioning system coupled with a scanning mobility particle sizer (SMPS) to explicitly measure total aerosol liquid water and to estimate the loss of semi-volatile compound volume during particle drying. Our study is motivated by the need to explore the hypothesis that particle water enhances biogenic SOA volume through aqueous-mediated partitioning of biogenically-derived organic species as the result of anthropogenic perturbations (Carlton and Turpin, 2013). The humid, photochemically active summer of the southeast U.S is ideal for conducting this study. Biogenic SOA mass concentrations are typically large (Lewis et al., 2004; Kleindienst et al., 2007; Ding et al., 2008) demonstrate positive relationships to $RH$ (Hatch et al., 2011), and have been shown to be enhanced by the presence of sulfates and nitrates that affect particle water uptake (Surratt et al., 2010). Further, model predictions suggest liquid water mass concentrations are high (Carlton and Turpin, 2013), and that aqueous phase water attributed to anthropogenic sulfate may influence biogenic SOA mass (Carlton et al., 2010; Hoyle et al., 2011) in the area.

The instrument was deployed as a part of the Southern Oxidant and Aerosol Study (SOAS), a collaborative field campaign during the summer season near Talladega National Forest in Brent, Alabama. The site is situated at ($32.903^\circ N$, $87.250^\circ W$), and at an elevation of 126 m. Data were collected from 3 June to 15 July 2013. Our principle objectives were to measure continuous in situ aerosol volume distributions of the dry, and particle phase liquid water constituents over the six-week time period of SOAS, and to identify chemical and thermodynamic controls on particle phase liquid water content.
This work provides a description and analysis of the instrument design, data reduction methods, and field measurement results, and also explores influences on ambient water content.

2.3 Experimental Section

2.3.1 Instrument Design

A schematic of the instrument and experimental setup is presented in Figure 2-1. Ambient air entered the instrument via a preconditioning inlet assembly consisting of a copper tube (9.5 mm ID, ~60 cm length) embedded in an aluminum block that was located ~1 m above ground level under the roof of an open shed. The temperature of the copper tube could be stably controlled between $T_{\text{inlet}} - 30\pm0.19^\circ\text{C}$ and $T_{\text{inlet}} + 50^\circ\text{C}$ (TE Technology LC-061). An optional Nafion membrane water-to-gas humidifier (PermaPure MH 110-48) was used to condition the aerosol. Subsequently the aerosol passed through a charge neutralizer (Aerosol Dynamics Inc. Model 100; Russell et al., 1996) holding four fresh $^{210}\text{Po}$ charge strips (NRD StaticMaster 2U500) with a nominal total activity of 2 mCi to confer a well-defined Boltzmann-type charge distribution to the particles. The aerosol was routed through an equilibration section (not pictured) before entering a high-flow differential mobility analyzer (DMA; Stolzenburg et al., 1998). The DMA sheath flow was controlled by a critical orifice (O’Keefe Controls Co., 9 L min$^{-1}$) and configured in recirculation mode. Monodisperse aerosol exiting the DMA was counted by a condensation particle counter (CPC; TSI 3772) operated at 1 L min$^{-1}$. Aerosol transit times were 10 s through the inlet loop, 6 s through the charge neutralizer, 5 s in the equilibration section between the Nafion humidifier and the DMA entrance, and 10.4 s through the DMA column. The residence of sample at the measurement RH is
comparable to those used in other studies of hygroscopicity (cf. Table 1, Duplissy et al., 2009).

The DMA was operated in scanning mobility particle sizer (SMPS) mode (Wang and Flagan, 1990). Negative voltage applied to the inner rod was held steady at 7 kV for 60 to 190 s, depending on the holding time of the present system state, and was then followed by an exponential decrease to 5 V over 300 s. Mapping between the time-varying electric field and selected particle mobility was achieved using the method of Wang and Flagan (1990). Conversion between particle mobility and mobility diameter followed standard DMA theory (Knutson and Whitby, 1975). The diameter range was determined by the sheath flow rate and DMA dimensions (60 cm tall, 11.6 cm ID outer cylinder, 10 cm OD collection rod) and spanned from 13 nm to ~1.1 \( \mu \)m. The number size distribution was found via a standard inversion that accounts for the transmission of multiply charged particles using the method described in Petters et al. (2007) with empirically determined adjustments accounting for particle transmission efficiencies that are described in detail in Section 3.2.2.

Relative humidity and temperature control of the DMA column were unchanged from previous versions of the instrument (Suda and Petters, 2013). Since the instrument was placed inside a well-ventilated shed, and since the objective was to track ambient temperatures, the neoprene insulation used by Suda and Petters (2013) was removed. Despite best efforts the temperature inside the shed was slightly higher than the outside, resulting in lower relative humidities inside the instrument relative to the values reported by the meteorology station. Both sheath and sample flows were optionally passed through Nafion humidifiers connected to a recirculating water bath. The water temperature
determined the dew point temperature of the sheath and sample stream and was actively controlled by Laboratory Virtual Instrument Engineering Workbench (LabVIEW) to match the ambient dew point temperature measured at the inlet. Temperature of the DMA column was measured using thermistors at the entrance, middle and bottom of the instrument. Two aluminum sleeves with PID controlled thermoelectric heat exchangers that were mounted on the outside were used to reduce the standard deviation of the three thermistor temperatures to 0.21±0.16°C. The nominal RH inside the instrument was computed from the average of the three thermistor temperatures and the flow rate weighted average of the measured sheath and sample dew point temperatures from the RH sensors embedded in the flows (HC2 Rotronics, Hygroclip, RH = ±0.8% accuracy). This method of humidity control and measurement is sufficient for measuring aerosol hygroscopic growth and activity coefficients in tandem DMA studies at RH < 90% (Suda and Petters, 2013).

The panels in Figure 2-1 show three instrument states. In the unperturbed state, no humidity conditioning occurred to the sample prior to particle sizing. The temperature of the preconditioning copper tube and the DMA column temperatures equaled the temperature measured at the inlet. The Nafion humidifier was bypassed. The sheath flow was actively humidified to match the relative humidity of the sample stream. This configuration measured the particle size distribution with minimal perturbation to sample temperature and RH.

In the dry instrument state, the temperature of the copper tube was chilled to 30 °C below the inlet temperature (Figure 2-1, panel 2). The temperature drop causes substances with dew point temperatures higher than the tubing temperature to condense
onto the wall. Measurements of relative humidity downstream of the copper tube confirmed that the dew point of the sample was less than or equal to the temperature of the aluminum block the copper tube was embedded in. Upon exiting the copper tube the sample flow warmed, resulting in a sharp drop in the water vapor saturation ratio. The sheath flow was conditioned in the same manner as the sample flow. This configuration measured the dry particle size distribution at $RH \sim 10\%$.

The dry-humidified state is similar to the unperturbed state (Figure 2-1, panel 3). The difference is that the sample aerosol was dried using the same method as the dry instrument state. A temperature reduction of 30°C is expected to drop the saturation vapor pressure of typical semi-volatile organic compounds by 1-2 orders of magnitude (Booth et al., 2010). Compounds that become supersaturated (i.e. solution contains more dissolved solute than is predicted by the solubility limit) will condense onto the wall or on the particles. Warming of the sample flow to the original temperature results in lower saturation ratio of gas-phase organic species. Subsequent re-equilibration of the semi-volatile compounds may result in net loss of organic mass from the particle. We note that compounds having high vapor pressure and low saturation ratios (e.g. glyoxal) will not be affected by this mechanism and may re-equilibrate with the aqueous phase. This configuration was designed to test if the drying procedure resulted in net removal of particle volume.

The full duty cycle of the instrument was unperturbed, dry-humidified, and dry, followed by an automated cleaning cycle (not pictured in Figure 2-1). During the cleaning cycle the temperature of the copper tube was warmed to 30 °C above outside temperature and back flushed with ambient air to remove water and other condensed
substances from the tube. Complete cleaning was ensured by verifying that the dew-point temperature measured before and after the copper tube were indistinguishable within experimental uncertainty. Since the temperature of the copper tube had to be adjusted between the instrument states, there was a 160 s delay between the cleaning cycle and unperturbed state, 190 s between the unperturbed and dry-humidified states, 60 s between the dry-humidified and dry states. A total of 600 s was allotted for the cleaning cycle. Approximately two unperturbed, dry-humidified, and dry size distributions were acquired per hour.

### 2.3.2 Instrument Performance

Particle sizing accuracy was verified in the laboratory using polystyrene latex spheres (PSL; 102±3 nm; Thermo Scientific, lot #36489). The resulting measured size distribution was in agreement with the PSL size. Particle transmission efficiencies were quantified in the lab and the field using effloresced ammonium sulfate particles. For these tests, particles were atomized from a stock solution (99.9% pure; Sigma-Aldrich; deionized water, ~ 18.2 MΩ cm) dried in silica gel diffusion dryer (TSI 3062), charge neutralized, and size selected by a second DMA (same origin and dimensions as the primary DMA, operated at 9:2 sheath-to-monodisperse flow ratio). Monodisperse aerosol was then routed to the SMPS and a second CPC (TSI 3771). Prior to these tests the two CPCs were intercompared sampling lab air side-by-side. Concentrations correlated well and the systematic offset between the instruments was 7.6%, which is within the manufacturer’s tolerance (±10%). Using this setup, the kernel function accounting for DMA transfer entering the inversion algorithm (Petters et al., 2007) for the selected bin resolution (60 bins) was determined for a series of mobility sizes between 20 and 600 nm.
Reduced transmission efficiencies were encountered for \( D < 50 \) nm, presumably due to diffusion losses. To account for reduced transmission a diameter dependent loss correction was included into the inversion matrix. Overall performance of the instrument was confirmed by comparing DMA integrated number concentration with CPC measured number concentration obtained either in simultaneously or sequentially with the size distribution scan. Concentrations from the integrated size distribution and the CPC agreed within 10%.

The aforementioned procedure does not capture losses that may occur in the preconditioning inlet assembly. These losses were different for the unperturbed, dry-humidified, and dry state. To account for differential transmission between the different states, the unperturbed, dry-humidified, and dry states are multiplied by a constant factor 1, 1.2, and 1.12 respectively to match the concentrations of a co-located CPC that was available prior to the start of the campaign (27–29 May) as well as 3–15 July. Between 3 June and 3 July instrument performance was validated by disconnecting the CPC from the DMA during several cleaning cycles. After 3 July the CPC sample flow was reduced to 0.5 L min\(^{-1}\) to accommodate 0.5 L min\(^{-1}\) flow for a cloud condensation nuclei counter added to the setup. Thus the total flow through the DMA was unchanged. Number distribution readings from the CPC were adjusted accordingly to account for the reduced number of particles counted. The regular system performance checks were also carried out after the change. No degradation of instrument performance was observed. Figure 2-2 shows example time series of particle number concentration observed with the SMPS and condensation particle counters. The figure shows that there is generally satisfactory agreement between SMPS and CPC derived concentrations, although there are some time
periods when the SMPS undercounted relative to the CPC (e.g. Figure 2-2 bottom panel 15–20 hr). The reasons for this are not entirely clear but undercounting occurs preferentially when fine mode aerosol are present. This indicates that transmission correction for particles with $D < 50$ nm may not fully account for the combined losses in the inlet assembly and the DMA column. We do not believe that these deviations are of concern since aerosol volume is most sensitive to accumulation mode sizes, and thus the error in volume introduced by these deviations is small.

2.3.3 Data Reduction

The volume of water associated with the aerosol is described using the hygroscopicity model of Petters and Kreidenweis (2007) and Kreidenweis et al. (2008):

$$V_w = V_d \kappa \frac{a_w}{1-a_w}, \quad (2-1)$$

where $a_w$ is the water activity, $V_w$ and $V_d$ are the volume of water and dry aerosol components, and $\kappa$ is the hygroscopicity parameter. In general $\kappa$ may vary with water activity and depends on particle chemical composition. For a particle composed of multiple components the particle’s $\kappa$ can be computed from the volume-weighted mixture of its dry components

$$\kappa = \sum \varepsilon_i \kappa_i, \quad (2-2)$$

where $\varepsilon_i$ and $\kappa_i$ are the volume fraction and hygroscopicity parameter of the $i^{th}$ component comprising the particle. Water activity and RH are related via

$$a_w = RH exp \left( \frac{4 \sigma_w M_w}{\rho_w RT D} \right)^{-1}, \quad (2-3)$$

where, $\sigma_w$, $M_w$ and $\rho_w$ represent the surface tension, molecular weight, and density of water, respectively, $T$ is the absolute temperature, $R$ is the ideal gas constant, $D$ is the
humidified particle diameter, and RH is the fractional relative humidity. Two hygroscopic growth factors can be defined:

\[ g_{vol} = \frac{V_{w} + V_{dry}}{V_{dry}} \quad \text{and} \quad g_{D} = \frac{D}{D_{d}}, \]  \hspace{0.5cm} (2-4)

where \( g_{vol} \) and \( g_{D} \) are the volume and diameter based growth factors, and \( D_{d} \) is the dry particle diameter. The dry particle composition can be conceptually divided into a low volatile (lv) and a semi-volatile (sv) fraction. The component parameter sets are \( \{ \varepsilon_{lv}, \kappa_{lv}, V_{w,lv} \) \) and \( \{ \varepsilon_{sv}, \kappa_{sv}, V_{w,sv} \) \) \( \text{and} \) \( V_{d,sv} \) \}, respectively. Assuming that volumes are additive (see Section 2.4 for further discussion), the total volume of water and dry components are \( V_{w} = V_{w,lv} + V_{w,sv} \) \) \( \text{and} \) \( V_{d} = V_{d,lv} + V_{d,sv} \).

Aerosol water and aerosol semi-volatile volumes are quantified using the measured aerosol volume of three consecutively measured instrument states: unperturbed, dry-humidified, and dry. The measured volumes of the three states correspond to

\[ V_{unperturbed} = V_{w,lv} + V_{w,sv} + V_{d,lv} + V_{d,sv} \]

\[ V_{dry} = V_{d,lv} \]

\[ V_{dry-humidified} = V_{w,lv} + V_{d,lv}, \]  \hspace{0.5cm} (2-5)

where \( V_{unperturbed} \), \( V_{dry} \), and \( V_{dry-humidified} \) are the total particle volumes measured for each instrument state, again assuming that volumes are additive. The semi-volatile volumes in Eq. (2-5) are operationally defined to correspond to the amount that was lost during the drying procedure. Consequently aerosol water volume associated with the low volatile material is

\[ V_{w,lv} = V_{dry-humidified} - V_{dry}. \]  \hspace{0.5cm} (2-6)

As will be discussed in Section 2.4, the contribution of \( V_{d,sv} \) to \( V_{w} \) was not discernible. The semi-volatile volume lost during drying is
\[ V_{sv,\text{instrument}} = V_{\text{unperturbed}} - V_{\text{dry-humidified}}. \]  

(2-7)

We note that the instrument defined loss of \( V_{sv,\text{instrument}} \) implies a loss of water that was hygroscopically bound in the particle. The amount of water associated with the semi-volatile fraction can be determined combining Eq. (2-7) with Eqs. (2-1) and (2-2) and solving for the amount of dry semi-volatile volume lost

\[ V_{d,sv} = V_{sv,\text{instrument}} \frac{1-a_w}{1-a_w+\kappa_{sv}a_w}. \]  

(2-8)

Eq. (2-8) demonstrates that if \( \kappa_{sv} = 0 \) (corresponding to a hypothetical non-hygroscopic material that is lost during drying) \( V_{sv,\text{instrument}} \) is an accurate measure of \( V_{d,sv} \). For cases where \( \kappa_{sv} > 0 \), \( V_{sv,\text{instrument}} \) will overestimate the amount of semi-volatile material lost during drying. Eq. (2-8) can be used to estimate the magnitude of this effect.

Volumes of water and semi-volatiles entering Eqs. (2-6) and (2-7), respectively, are derived from the size distribution data. Volume distributions were calculated from the number size distributions assuming particle sphericity following Seinfeld and Pandis (2006). Figure 2-3 shows example histograms of the inverted 60-bin representation of the number and volume size distribution. The data show that the size distribution is bimodal with mode diameters \( D_{p1} \sim 0.06 \) μm and \( D_{p2} \sim 0.2 \) μm. The relative contributions of number and volume to the ratio of mode #1 to #2 are \( \sim10:1 \) and \( \sim1:2 \), respectively. Low number concentrations result in poor counting statistics leading to the apparent noisiness of the volume distribution histograms in the larger size mode. A single false count in the largest diameter bins can significantly bias the total volume derived from the spectra. False counts sometimes arise due to arcing in the DMA column. Arcing occurs at high relative humidity and high electric potential leading to a breakdown of the electric field in
the column. This enables transmission of smaller particles that are falsely sized in the large bins. These particles would appear as a peak at $D \sim 1 \mu\text{m}$ (not seen in Figure 2-3).

To filter out possible contributions from arcing, reduce the noisiness from low counting statistics, and identify the mode diameters in each scan, log-normal distribution functions (Hatch and Choate, 1929) are fitted to the distribution data. A non-linear least square fitting routine is used to minimize the residual between the data and the distribution function. Artifacts from rogue counts at high diameters are filtered since a 3rd mode is not allowed. Distribution functions are fit to both the number and volume distributions separately. Example fits to the histograms are superimposed in Figure 2-3. To test the efficacy of the fits to represent the data, the fit and data integrated moments were compared. For each distribution the relative difference between integrated number concentration from the data and the fit were $1.4\pm5\%$, where $1.4\%$ denotes the average error and $\pm5\%$ the fit-to-fit variability of the error. Similarly, the relative difference between the integrated volume from the data and the fit were $1.6\%\pm5.3\%$. The variability in the data greatly exceeds any artificial error introduced by the two-mode parameterization, indicating that the two-mode parameterization is appropriate to model the size distribution. Measured volume concentrations were correlated with mass-concentrations from a co-located tapered element oscillating microbalance (Patashnick and Rupprecht, 1991, see supplementary information). The offset between these measurements is broadly consistent with an aerosol density of $1.5 \text{ g cm}^{-3}$, a value that is reasonable for the measured aerosol chemical composition. Note that the average number and volume fraction of particles with $D < 50 \text{ nm}$ to smaller mode is $30.6\%$ and $19\%$, respectively. For those sizes instrument performance is degraded due to diffusional
losses. However, since the contribution of $D < 50$ nm particles is less than 50% and since the potential error is likely serious only for $D < 30$ nm we do not expect that diffusional losses affect the fit-determined mode diameter.

The fitted volumes $V_{\text{unperturbed}}$ and $V_{\text{dry}}$ enter Eq. (2-6) to find $V_w$. Furthermore, $V_{\text{dry}}$ defines $V_d$ in Eq. (2-1). Calculations were also performed using the data distributions directly and summary statistics are provided in the online supplement. For these calculations, we assume $a_w$ is equivalent to $RH$, and the implications of this assumption are discussed later on in Section 2.2.4.

Alternatively, growth factors are defined by calculating the ratio of the shift in the mode diameters from the log-normal fit between $V_{\text{unperturbed}}$ and $V_{\text{dry}}$:

$$g_{f_i} = \frac{D_{pgi,\text{unperturbed}}}{D_{pgi,\text{dry}}},$$

(2-9)

where $g_{f_i}$ is the diameter growth factor of the $i^{th}$ log-normal mode and $D_{pgi,\text{unperturbed}}$ and $D_{pgi,\text{dry}}$ are the fit-returned mode diameters of the $i^{th}$ log-normal mode for the sequential unperturbed and dry measurement cycles. Conversion from $RH$ to water activity is performed using Eq. (2-3) and with $D_{pgi,\text{unperturbed}}$ as the particle diameter. The resulting diameter growth factors can be used to define $\kappa$ values that characterize the hygroscopicity of the individual modes:

$$\kappa_{Di} = (g_{f_i}^3 - 1)(1 - a_w)a_w^{-1},$$

(2-10)

where $\kappa_{Di}$ is the diameter based hygroscopicity parameter for the $i^{th}$ log-normal mode.

The fitted volumes $V_{\text{unperturbed}}$ and $V_{\text{dry-humidified}}$ enter Eq. (7) to find $V_{SV, \text{instrument}}$. Since the scans were sequential, relative humidity was not necessarily the same between the two scans. To account for $RH$ variations from the target value the $V_{\text{dry-humidified}}$ volume was empirically corrected to the $RH$ corresponding the unperturbed cycle as
recommended by Gysel et al. (2009). In our case, the correction is performed as follows. First, $\kappa_{\text{vol, dry-humidified}}$ is determined from $V_{\text{dry-humidified}}$ and $V_{\text{dry}}$ using Eq. (1) and $RH_{\text{dry-humidified}}$. Second, we obtain the corrected state using:

$$V_{\text{dry-humidified, corrected}} = V_{\text{dry}} \kappa_{\text{vol}} \frac{RH_{\text{unperturbed}}}{1 - RH_{\text{unperturbed}}} + V_{\text{dry}},$$

(2-11)

where $V_{\text{dry-humidified, corrected}}$ is the corrected volume for the dry-humidified state, $\kappa_{\text{vol}}$ is the size distribution integrated volume based hygroscopicity parameter computed from Eq. (2-1), and $RH_{\text{unperturbed}}$ is the average $RH$ measured during the unperturbed state.

### 2.3.4 Sources of Uncertainty

Since $RH_{\text{unperturbed}}$ was slightly lower than that measured at the meteorological station due to slight heating of the shed, the reported water contents represent a lower estimate. The relative difference between the two $RH$ sensors, computed as $(RH_{\text{met}} - RH_{\text{shed}})/RH_{\text{met}}$ was 15%. Eq. (2-11) can be used to estimate the corrected water content at conditions deviating from instrument conditions.

The DMA measures mobility diameter, which for spherical particles is equal to the volume equivalent diameter. During the unperturbed and dry-humidified state, particles that contain water will likely have a nearly spherical shape because water preferentially adsorbs at the corners and edges of the particle (Mikhailov et al., 2004). However, some particles may have a crystalline structure when dried. Particle shape irregularities increase the drag inside the DMA and result in a larger electric mobility diameter measured. Compounds with a cubic structure, for example, have a shape factor of 1.08 (e.g. Kelly and McMurry, 1992). Gysel et al. (2002) calculated that the relation between the volume equivalent diameter and the mobility diameter of a cubic particle is 0.96, which indicates a 4% overestimate of particle diameter by the DMA for cubic
particles. As a result, the volume distribution calculated from the mobility diameters of irregularly shaped particles during the dry state may be overestimated, which would result in an underestimate of water content when we difference the dry-humidified and dry states.

The calculations also assume volume additivity, i.e. that aerosol species volumes are independent of solution concentration and of species mixing fractions. Detailed studies of water activity for inorganic compounds show that excess volume of mixing can be large for some systems, particularly at low water content (Clegg et al., 1998; Wexler and Clegg, 2002). However, other studies find that volume additivity holds, although this may be due to the cancellation of excess volumes (Stokes and Robinson, 1966). Mikhailov et al. (2004) found volume additivity to be a safe assumption for aerosols composed of proteins and salts for which the individual densities and behavior in solution were known. In the absence of detailed knowledge of composition and the component behavior in solution, as is the case with ambient aerosol, volume additivity has typically been assumed (Dick et al., 2000; Speer et al., 2003). Dick et al. (2000) constrained their results with the calculation that for simple aqueous inorganic solutions volume additivity mis-estimates particle phase liquid water by -7% for H$_2$SO$_4$ at $RH = 90\%$ and by < 5% for deliquesced NH$_4$HSO$_4$, (NH$_4$)$_3$H(SO$_4$)$_2$, and (NH$_4$)$_2$SO$_4$, thus providing an estimate of error for different ammonium-to-sulfate ratios for the inorganic fraction of ambient aerosols. In sum we conclude that the assumption of volume additivity may introduce a small error in our calculations when molecular interactions in the solution deviate from the pure components.
Since multiple sizes contribute to the fitted volumes, the conversion from instrument $RH$ (Section 2.2.1) to $a_w$ via Eq. (2-3) is not straightforward. We therefore assume $a_w$ is equivalent to $RH$ averaged over the scan to compute $\kappa_{\text{vol}}$. The expected difference between $RH$ and water activity are between 0.01 and 0.02 in absolute units (e.g. Tang, 1996) and will result an $RH$ dependent overestimate in $\kappa_{\text{vol}}$ ranging from 4 to 11%.

Changes in aerosol solution from stable to metastable conditions during the dry-humidified state can also introduce errors into the calculations of water volume. Inside the Nafion tube during the dry-humidified state, the $RH$ approaches ~100% at the temperature that approximates the dew point temperature of the outside air. Thus most particles that can deliquesce at subsaturated relative humidity will contain water exiting the Nafion. Particles may or may not effloresce prior to entering the DMA. If $V_{\text{dry-humidified}}$ exceeds $V_{\text{unperturbed}}$, it may be possible that a sufficiently large volume of water was added by the deliquescence of water on particles that were effloresced under unperturbed conditions.

Some compounds do not effloresce and thus may retain water at low $RH$. In that case the measured $V_{\text{dry}}$ will be overestimated and the instrumented-inferred $gf_{\text{vol}}$ and $\kappa_{\text{vol}}$ will be biased low. The dry instruments state corresponds to $RH = 12.5\% \pm 2.3\%$, where 2.3% denotes the variability of the RH in absolute units at the dry state throughout the campaign. Eq. 1 can be used to estimate the magnitude of particle-bound water at $a_w = 0.12$, assuming the $\kappa$ is invariant with $a_w$. The relative error in the inferred $\kappa_{\text{vol}}$ is
e = \left( \frac{\kappa_{\text{vol,dry}} - \kappa_{\text{vol,0.12}}}{\kappa_{\text{vol,dry}}} \right),
where $\kappa_{\text{vol,dry}}$ is the true hygroscopicity, and $\kappa_{\text{vol,0.12}}$ is the hygroscopicity one would calculate if particle-bound water was present at $a_w=0.12$. 

The estimated error is independent of $a_w$ and scales with $\kappa_{\text{vol,dry}}$. For $\kappa_{\text{vol,dry}} = 0.01$ and $\kappa_{\text{vol,dry}} = 0.6$ the error is $e=-0.1\%$ and $e=-7.6\%$, respectively.

Although losses of semi-volatile compounds during drying are expected, the total amount lost unlikely corresponds to what would be expected from thermodynamic equilibrium. Several reasons are listed here. First, fast kinetics for aerosol dissolution are required. However, this may not be the case if the initial $RH$ is low, as the aerosol becomes viscous and formation of a homogenous liquid phase may be slow (Vaden et al., 2011; Renbaum-Wolff et al., 2013). Second, evaporation rates must be fast enough to remove semi-volatiles in the inlet even though evaporation rates may be slow (Bilde et al., 2003). Succinic acid, for example, has been measured to have an evaporation rate of $2.76 \text{ nm s}^{-1}$ at 299.8 K and 64.1% $RH$ (Koponen et al., 2007). The residence time of the preconditioning inlet may not be adequate to capture full evaporation of succinic acid. Third, the phase state of the particles after drying is uncertain. Sub-cooled liquid vapor pressure may be up to three orders of magnitude higher than solid vapor pressure (Booth et al., 2010). Since partitioning between gas and particle phase is dependent on the saturation vapor pressure of the organic compound in the liquid state (Pankow, 1994), semi-volatile content may be underestimated if particles are in the solid state after drying. However, water in the organic phase could reduce organic equilibrium partial pressure according to Raoult’s law (Pankow and Chang, 2008), though this effect did not play a strong role on $\alpha$-pinene SOA formation (Prisle et al., 2010). Detailed studies with proxy systems are needed to quantify the amount removed as a function of compound volatility in the context of particle phase/viscosity/diffusion and non-ideal solution effects. Therefore, the tests with the preconditioning inlet only test whether the selected drying
procedure lead to loss of material during SOAS and is not intended to be used as a measure of the semi-volatile fraction.

2.3.5 Chemical Composition

PM$_{2.5}$ or PM$_{1}$ water soluble ions were measured by a Particle-Into-Liquid-Sampler coupled with an Ion Chromatograph (PILS-IC; Metrohm 761 Compact IC. Similar setups are described in previous field studies (Orsini et al., 2003; Liu et al., 2012). Metrosep A Supp 5 anion column and C 4 150 cation column (Metrohm USA, Riverside FL) were used to separate PILS liquid sample anions sulfate, nitrate, chloride, oxalate, acetate, formate and cations ammonium, sodium, potassium, calcium, magnesium with 20 min duty cycle. Water soluble organic carbon (WSOC) was measured by coupling the PILS with a total carbon analyzer ) or a total carbon analyzer (TOC, Model 900, GE Analytical Instruments; Boulder, CO). The TOC was operated in normal mode with a 6-min duty cycle. The PILS sample air flow rate was 16.8 ± 0.35 L/min and URG cyclones inlets provided PM cut sizes of PM$_{2.5}$ for the 1st half of field study (June 1 to June 22) and PM$_{1}$ for the latter half (June 23 to July 15). Honeycomb acid (phosphoric acid) and base (sodium carbonate) coated denuders remove interfering gases. Water-soluble organic matter (WSOM) was then estimated from WSOC measurements using a conversion factor of 2.1 (Turpin and Lim, 2001). The sample inlet was ~7m above ground level and ~4m long. The inlet line was insulated inside a trailer (typical indoor T was 25°C) and less than 1m in length to minimize possible changes in aerosol composition prior to measurement. The trailer was located adjacent to the shed housing the SMPS system. Periodic 1 hour blank measurements were made every day by placing a HEPA filter (Pall corp.) on the cyclone inlet. All data were blank corrected.
2.4 Results

Figure 2-4 provides a time-series of the measured data for the six weeks of data collection. A campaign-based statistical summary of selected quantities is summarized in Tables 2-1 and 2-2. Temperatures typically fluctuated between ~24 °C during nighttime and ~30 °C during daytime. Minor fluctuations in the dew point temperature ($T_{dew}$) indicate that the diel cycle of RH was driven mostly by temperature variations – highest at low temperatures and lowest at high temperatures. Daily fluctuations for RH were typically within 60% to 77%. Periods of heavy rainfall included 5–7 June, 18 June, and 3–6 July. The highest amount of precipitation for a single day was 45.8 mm on 6 June. Periods with little to no rainfall included 8 June, 11–12 June, 14–16 June, 19–22 June, 25–26 June, and 9–12 July. Periods with noticeably high amounts of rainfall, such as 5 and 6 June, resulted in low levels of dry ($V_d$) and water aerosol ($V_w$) volume. During periods with no rainfall, such as the days between 19 and 22 June, dry particle volume steadily increased, presumably due to the lack of an efficient aerosol sink.

Statistics of the campaign average bimodal distributions of number concentration, surface area, and volume for the three-instrument inlet states are summarized in Table 2-2. A graphical interpretation of Table 2-2 is provided in Figure 2-5. Average number concentrations were ~2300 cm$^{-3}$ and 500 cm$^{-3}$ for the smaller and larger diameter mode, respectively. Surface area distributions are higher for the unperturbed and dry-humidified states than for the dry state due to the additional water uptake. For the smaller diameter mode, surface areas average at 54.0 $\mu$m$^2$ cm$^{-3}$ for both the unperturbed and dry-humidified states, and 54.4 $\mu$m$^2$ cm$^{-3}$ for the dry state, while for the larger mode, the averages are 130, 127, and 95 $\mu$m$^2$ cm$^{-3}$ for the unperturbed, dry-humidified, and dry...
states respectively. This indicates that water uptake increases the aerosol surface area by 33.7% for the large diameter mode, while the effect for the small diameter mode is not observed, suggesting that the smaller mode is less hygroscopic. (Note that the surface area of the small mode nominally shrinks by 0.07% upon humidification. Possible reasons are noise in the measurement and that some fraction of the particles is picked up in mode #1 for the “dry” fit and mode #2 for the humidified fit.). Similarly, the volume distributions for the unperturbed and dry-humidified state are both larger than the distribution for the dry state. Volume averages are 2.20, 2.18, and 1.49 µm$^3$ cm$^{-3}$ for the smaller diameter mode and 6.50, 6.93, and 3.87 µm$^3$ cm$^{-3}$ for the larger diameter mode, for the unperturbed, dry-humidified, and dry states, respectively. The observed size distribution statistics are consistent with previous measurements of accumulation mode remote continental aerosol size distributions (Jaenicke, 1993; Martin et al., 2010; Levin et al., 2012). Notably, small particle events indicative of nucleation (Levin et al., 2012) were not observed during the campaign. Number and volume concentrations were significantly higher during SOAS than during the Amazonian Aerosol Characterization Experiment (Martin et al., 2010), suggesting anthropogenic influence, likely from sulfates, on the SOAS site (Carlton and Turpin, 2013). One would expect that the presence of sulfates would result in more hygroscopic aerosol in Alabama compared to the pristine Amazon rainforest.

Figure 2-4 summarizes trends of particle phase liquid water ($V_w$) throughout the campaign. Measured water concentrations always exceeded 0 µg m$^{-3}$, typically varied between 1 and 5 µg m$^{-3}$, and peaked at 73 µg m$^{-3}$ on 17 June. The campaign time-series data indicate a strong diel cycle for $V_w$. A campaign-average representation of the diel
cycle is presented in Figure 2-6. We note that this average diel cycle was suppressed during periods with heavy washout. Analysis contrasting selected time periods is provided in Appendix B1. Figure 2-6 shows that water volume peaked during the early morning hours ~7 a.m. to 9 a.m. local time (LT). On average, the median value for water concentrations was 2.9 µg m\(^{-3}\), and the interquartile range was between 1.3 and 4.9 µg m\(^{-3}\) (Table 2-1). For 7 a.m. to 9 a.m. LT, however, the median value for water mass concentrations was 5.8 µg m\(^{-3}\). The start of the decreasing transition for RH occurred in the morning at approximately 6 a.m. LT. The peak in water content occurred shortly after RH decreased and RH continued its decline until it increased again at 4 p.m. LT. Water mass concentrations leveled off after the morning peak, and despite the RH increase after 4 p.m. LT, water mass concentrations did not experience a noticeable increase in response during that time period. The partial decorrelation of RH and water volume suggests that either aerosol dry volume or aerosol hygroscopicity evolved during the day.

The evolution of composition is demonstrated further by the volumetric growth factor data. Figure 2-7 shows the diurnal profile of \(g_{\text{vol}}\) overlaid with expected \(g_{\text{vol}}\) assuming an invariant \(\kappa\) throughout the day. Analysis in Appendix B1 is performed for a volumetric growth factor profile for a single day as an illustrative example of a day with no rainfall. The data show that \(g_{\text{vol}}\) is always higher than 1, and exceeds 2 between 7 and 9 a.m. LT. Overall, the \(g_{\text{vol}}\) median is 1.61 (Table 2-1). The data also show that the observed interquartile range for \(g_{\text{vol}}\) is between 1.37 and 2.01. These values are similar to those found during the Pittsburgh Air Quality Study (Khlystov et al., 2005). The overlaid expected growth factor for different values of \(\kappa\) in the plot suggest that the aerosol hygroscopicity during the campaign changed systematically throughout the day, which is
further explored in the diel cycle of $\kappa_{\text{vol}}$ shown in Figure 2-8. Overall, the median observed value for $\kappa_{\text{vol}}$ is 0.27 and the interquartile range is from 0.14 to 0.46. At night, $\kappa_{\text{vol}}$ is steady at median values of $\sim$0.18 and quartiles of $\sim$ 0.08 to 0.3. After 7 a.m. LT, $\kappa_{\text{vol}}$ increases to about a median value of $\sim$0.48 and quartiles of $\sim$ 0.2 to 0.7 followed by a gradual decrease throughout the rest of the day. The hygroscopicity statistics for the two size modes are summarized in Table 2-1. Medians of $\kappa$ for the smaller and larger diameter mode average to $\kappa_{D1} = 0.13$ and $\kappa_{D2} = 0.29$. In comparison, medians for $\kappa_{\text{vol}}$ average to 0.27, which more closely matches $\kappa_{D2}$ than $\kappa_{D1}$. $\kappa_{D1}$ also exhibits greater variability than either $\kappa_{\text{vol}}$ or $\kappa_{D2}$, which is most likely attributable to noisiness in the fittings on a scan by scan basis that increases variability in the diameter growth factors for the mode.

The measured $\kappa_{\text{vol}}$ are consistent with a mix of hygroscopic organic and inorganic compounds (Petters and Kreidenweis, 2007; Kreidenweis et al., 2008; Suda et al., 2012). The observed $\kappa_{\text{vol}}$ are significantly larger than those in organic dominated environments encountered in remote continental forests (Gunthe et al., 2009; Sihto et al., 2011; Levin et al., 2012; Mikhailov et al., 2013), which vary between $0.1 < \kappa < 0.2$. In these studies the aerosol were dried prior to measurement. We therefore also tested whether drying removed semi-volatile compounds and thus bias the observed $\kappa_{\text{vol}}$.

Results for the diel cycle of the semi-volatile fraction computed from Eqs. (2-7) and (2-11) are summarized in Figure 2-9. No clear diel trend is noticeable. At times the average semi-volatile volume concentration is negative. However, the 95% confidence interval of the mean consistently envelops zero. This is consistent with the null hypothesis that no semi-volatiles were removed in our preconditioning system and that
the measurement has random error. We note that the confidence limits are $\pm 0.5 \mu g m^{-3}$, which implies that point-to-point variability in the semi-volatile retrieval led to significant uncertainty. Losses of semi-volatile material less than $0.5 \mu g m^{-3}$ may have occurred but were undetectable with our method.

The diel evolution of the hygroscopic species sulfate, nitrate, ammonia, and WSOM is shown in Figure 2-10. Ammonia and WSOM are approximately constant while sulfate and nitrate show a diel cycle with a daytime maximum. We note that these speciated compounds only represent a subsection of the total aerosol since dust and elemental carbon are not included. However, the sum of the median mass concentrations for sulfate (1.7 $\mu g m^{-3}$), nitrate (0.07 $\mu g m^{-3}$), ammonia (0.47 $\mu g m^{-3}$), and WSOM (3.15 $\mu g m^{-3}$) is 5.39 $\mu g m^{-3}$ and these species likely played an important role in governing aerosol water uptake during SOAS.

2.5 Discussion

The dynamics governing the total water volume that is available for water-mediated partitioning and condensed aqueous phase reactions are complex. First, our results show that at minimum 7% of the aerosol volume is composed of water. Day-to-day variations in aerosol hygroscopicity are relatively minor and total water volumes scale with dry aerosol volume and $RH$. The fractional water content is highest during the early morning hours when temperatures are lowest and $RH$ is at a maximum. During those times volume growth factors exceed two, implying that the aerosol composition is dominated by water. Second, on timescales longer than the diel cycle, dry aerosol volume is the dominant determinant of absolute water content. For example, between 3 and 9 June (Figure 2-4), dry aerosol volume increased from 1 to $10 \mu m^3 cm^{-3}$, with concomitant
increases in water volumes. Dry aerosol volume concentrations result from a balance in production (emissions and secondary production via chemical reactions) and removal (dry deposition, washout, and venting to the free troposphere) processes. The buildup of dry aerosol volume during stagnation events seems to be tied to the meteorological conditions. Warm temperatures, high actinic flux, and absence of deep convection and precipitation lead to accumulation of aerosol volume. The availability of a significant aqueous phase (by total volume) is thus correlated with, and contingent on, suitable weather conditions. Finally, the diel cycle of aerosol hygroscopicity suggests that significant changes in aerosol composition may occur on the time scale of a few hours. This is supported by changes in the inorganic aerosol fraction derived from the PILS measurements shown in Figure 10e. Variation range from 20% to 80% for the inorganic fraction and the diel cycle in organic fraction mimics that of $\kappa$ (Figure 2-10f). However, the actual relationship between chemical composition and $\kappa$ is complex (Kreidenweis et al., 2008). Assuming that the PILS measured chemical composition captures most of the species responsible for water uptake and assuming a representative composition for the WSOM fraction we estimated the $RH$ dependence using the E-AIM model (Clegg et al., 1998). These calculations show that the reduction of $RH$ from 90% to 40% may cause a 60% increase in $\kappa$ (see supplemental information). The diel fluctuation in $\kappa_{vol}$ is from 0.14 to 0.46. Notably, the aerosol is most hygroscopic when the fractional water content is larger, which is consistent with the modeled $RH$ dependency of $\kappa$. The magnitude of the RH dependence, however, is insufficient to explain the full diel $\kappa$-cycle which is driven by a combination of the two effects. We note that deconvolution of the two effects will require closure calculations between time-resolved chemical composition and $\kappa$ that
are contingent upon complete speciation of the aerosol (including black carbon and dust),
detailed information about the composition of WSOM, and estimation of aerosol
solubility and density (see supplemental information).

Figure 2-11 shows that there was not strong diel profile in dry aerosol volume. Thus diel changes in aerosol hygroscopicity are indicative of dynamic aerosol chemistry
that does not involve major volume changes. Another possible explanation for the
observed $\kappa_{vol}$ changes are chemical transformations of the organic fraction itself that lead
to corresponding changes in $\kappa_{org}$. Specifically, $\kappa_{org}$ has been shown to increase with higher
oxygen-to-carbon ratio (Jimenez et al., 2009), smaller molecular size (Petters et al.,
2009), and higher solubility in water (Petters and Kreidenweis, 2008). The presence of an
aqueous phase combined with the morning actinic flux could drive aqueous phase
oxidation reactions or lead to hydrolysis reactions resulting in smaller, more oxidized,
and water-soluble organic molecules or ions.

Perhaps consistent with aqueous phase processes, the hygroscopicity parameter $\kappa$
increases with particle size, similar to trends of $\kappa$ with particle size were observed in
previous field studies (Gunthe et al., 2009; Wiedensohler et al., 2009; Levin et al., 2012).
The higher $\kappa$ values for the larger mode suggest the presence of more hygroscopic solutes
such as more oxidized or hydrolyzed organic species, sulfates or nitrates. A direct
consequence of the size dependence is that the aqueous phase volume is predominately
located with the larger particles sizes both due to the larger dry aerosol volume and the
larger hygroscopicity. This initiates the question: of the three parameters influencing
particle phase liquid water (dry aerosol volume, hygroscopicity, and $RH$), which are the
most important under ambient conditions?
We address this question using sensitivity analysis following the approach of (Feingold, 2003; McFiggans et al., 2006). The sensitivity is defined as $S(X_i) = \frac{\partial \ln V_w}{\partial \ln X_i}$, where $X_i$ is one of $V_d$, RH, or $\kappa_{vol}$. Comparing values of $S(X_i)$ describes their relative importance given equal perturbations $\delta V_d$, $\delta RH$, or $\delta \kappa_{vol}$. Based on Eq. (2-1), sensitivities $S(\kappa_{vol})$ and $S(V_d) = 1$, while $S(RH) = 1/(1 - RH)$. These sensitivities, combined with values for $\delta X_i$ are summarized in Table 2-3. The interpretation of $S(X_i) = 1$ is that e.g. a 10% relative change in parameter $X_i$ will lead to a 10% relative change in $V_w$. Changes in hygroscopicity, dry aerosol volume, and $RH$ are potentially cross-correlated. For example, the condensation of sulfuric acid onto organic aerosol will lead to both increases in $\kappa$ and $V_d$, thereby influencing $V_w$ via both pathways. In contrast, the condensation of nonhygroscopic organic material (i.e., $\kappa = 0$) will lead to no increase in $V_w$ as the effects of increased volume and decreased hygroscopicity cancel after application of the ZSR mixing rule for $\kappa$ (Petters and Kreidenweis, 2007). From observations during SOAS, the perturbations $\delta \kappa$ and $\delta V_d$ are comparable, resulting in 0–200% variability of $V_w$ depending on the hygroscopicity of the compound that controls the variability of the dry aerosol volume. Although the relative variability in $RH$ is typically lower than $\kappa$ or $V_d$, the relative sensitivity of water volume to changes in $RH$ is much larger observed at the site. At $RH = 40\%$, 70\%, and 90\%, $S(RH) = 1.66$, 3.33, and 10, respectively. Thus, at $RH = 90\%$, a 1% relative fluctuation in $RH$ will be equivalent to controls on water content by a 10% change in dry aerosol volume, which is likely driven by changes in anthropogenic ions of sulfate, nitrate, and ammonium. As a result, the diel trends in water are primarily controlled by changes in relative humidity. The strong dependence of $V_w$ on relative humidity is similar to that reported in previous
studies (i.e., Khlystov et al., 2005). Since dew points are relatively constant during the day at the site location (see Figure 2-4), we believe that the diel $V_w$ cycle is indirectly controlled by temperature. Correlations between ambient temperature, relative humidity, and particle phase liquid water can potentially obfuscate the attribution of semi-volatile partitioning into the condensed phase to either the presence of water due to increased $RH$ or reduced vapor pressure due to reduced temperature. Although this conceptual distinction is important, it is irrelevant for this study because loss of semi-volatile volumes $V_{sv}$ during our drying procedure was not statistically discernible (Figure 2-9). The absence of this effect needs to be interpreted in light of the experimental limitations discussed in Section 2.2.4 (particle phase/viscosity/diffusion and non-ideal solution effects) as well as in context of the organic mass concentration, whether partitioning should follow Raoult’s law, the mass of pre-existing non-volatile and dissolved material, and not the least the expected amounts of semi-volatile compounds (Barley et al., 2009). For heavily aged SOA, loss of semi-volatile compounds may be minimal. Small differences in mass may be more difficult to detect if overall aerosol mass loadings are large. Improved methodology will be needed to further reduce the uncertainty to resolve removal of less than 0.5 $\mu$g m$^{-3}$ amounts, establish volatility limits that can be detected using proxy system, lengthen residence time between the exit of the cold trap and SMPS to ensure full equilibration, and test whether the conclusion is valid when single particle sizes are considered.

2.6 Conclusion

We present measurements of dry aerosol volume, particle phase liquid water, and dry aerosol volume changes that may occur during particle drying. The measurements
were obtained during the SOAS campaign at Talladega National Forest in Brent, AL from 3 June to 15 July 2013 to collect six weeks of continuous in situ measurements of ambient aerosols. Aqueous phase liquid water was always present, comprising at minimum 7% of aerosol volume. Water mass concentrations typically ranged from 1 to 5 µg m⁻³. Liquid water was the dominant aerosol constituent from ~7 a.m. to 9 a.m. LT. The diel water-content cycle was dominated by changes in RH, which mostly fluctuated due to changes in temperature. A strong diel cycle in aerosol hygroscopicity is consistent with observed systematic changes in aerosol composition and a water-content dependency of κ. Aerosol hygroscopicity increased with particle size. The method of particle drying used in this study did not lead to removal of significant material that was associated with the aqueous phase.

2.7 References


Table 2-1. Mean and quartiles of number concentration $N_{\text{conc}}$, relative humidity $RH$, temperature ($T$), particle phase water volume $V_w$, semi-volatile volume $V_{sv}$ lost during drying, volume of solutes $V_d$, $\kappa_{\text{vol}}$, $\kappa_{D1}$, $\kappa_{D2}$, and $gf_{\text{vol}}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>Mean</th>
<th>25% Quartile</th>
<th>50% Quartile</th>
<th>75% Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{conc}}$</td>
<td>cm$^{-3}$</td>
<td>2629</td>
<td>1531</td>
<td>2155</td>
<td>3136</td>
</tr>
<tr>
<td>$RH$</td>
<td>%</td>
<td>68.94</td>
<td>59.95</td>
<td>70.85</td>
<td>77.10</td>
</tr>
<tr>
<td>$T$</td>
<td>°C</td>
<td>27.48</td>
<td>24.85</td>
<td>26.73</td>
<td>30.06</td>
</tr>
<tr>
<td>$V_w$</td>
<td>μm$^3$ cm$^{-3}$</td>
<td>4.00</td>
<td>1.34</td>
<td>2.88</td>
<td>4.86</td>
</tr>
<tr>
<td>$V_{sv}$</td>
<td>μm$^3$ cm$^{-3}$</td>
<td>0.18</td>
<td>-0.65</td>
<td>0.22</td>
<td>1.05</td>
</tr>
<tr>
<td>$V_d$</td>
<td>μm$^3$ cm$^{-3}$</td>
<td>5.29</td>
<td>2.95</td>
<td>4.65</td>
<td>7.17</td>
</tr>
<tr>
<td>$\kappa_{\text{vol}}$</td>
<td>dimensionless</td>
<td>0.33</td>
<td>0.14</td>
<td>0.27</td>
<td>0.46</td>
</tr>
<tr>
<td>$\kappa_{D1}$</td>
<td>dimensionless</td>
<td>0.47</td>
<td>-0.11</td>
<td>0.13</td>
<td>0.78</td>
</tr>
<tr>
<td>$\kappa_{D2}$</td>
<td>dimensionless</td>
<td>0.38</td>
<td>0.12</td>
<td>0.29</td>
<td>0.56</td>
</tr>
<tr>
<td>$gf_{\text{vol}}$</td>
<td>dimensionless</td>
<td>1.78</td>
<td>1.37</td>
<td>1.61</td>
<td>2.01</td>
</tr>
</tbody>
</table>
Table 2-2. Parameters describing two-mode distributions of number concentration, surface area, and volume for the unperturbed \((C_{\text{unperturbed},i})\), dry-humidified \((C_{\text{dry-humidified},i})\), and dry states \((C_{\text{dry},i})\).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>Number N</th>
<th>Surface S</th>
<th>Volume V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(C_{\text{unperturbed},i})</td>
<td>cm(^{-3})</td>
<td>2254</td>
<td>507</td>
<td>54.0</td>
</tr>
<tr>
<td>(gmd_{\text{unperturbed},i})</td>
<td>µm</td>
<td>0.099</td>
<td>0.24</td>
<td>0.11</td>
</tr>
<tr>
<td>(gsd_{\text{unperturbed},i})</td>
<td></td>
<td>1.54</td>
<td>1.43</td>
<td>1.54</td>
</tr>
<tr>
<td>Mode</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(C_{\text{dry-humidified},i})</td>
<td>cm(^{-3})</td>
<td>2238</td>
<td>511</td>
<td>54.0</td>
</tr>
<tr>
<td>(gmd_{\text{dry-humidified},i})</td>
<td>µm</td>
<td>0.11</td>
<td>0.23</td>
<td>0.12</td>
</tr>
<tr>
<td>(gsd_{\text{dry-humidified},i})</td>
<td>1.49</td>
<td>1.49</td>
<td>1.5</td>
<td>1.49</td>
</tr>
<tr>
<td>Mode</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(C_{\text{dry},i})</td>
<td>cm(^{-3})</td>
<td>2355</td>
<td>403</td>
<td>54.4</td>
</tr>
<tr>
<td>(gmd_{\text{dry},i})</td>
<td>µm</td>
<td>0.096</td>
<td>0.2</td>
<td>0.10</td>
</tr>
<tr>
<td>(gsd_{\text{dry},i})</td>
<td></td>
<td>1.47</td>
<td>1.43</td>
<td>1.47</td>
</tr>
</tbody>
</table>
Table 2-3. Table of sensitivity ratios $S(X_i)$ based on Eq. (2-1) for $X_i$ equal to $\kappa_{\text{vol}}, V_d, \text{and } RH$. The table also reports the typical range of these parameters during the SOAS campaign.

<table>
<thead>
<tr>
<th>$X_i$</th>
<th>$S(X_i)$</th>
<th>Typical range during SOAS</th>
<th>Average of range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_{\text{vol}}$</td>
<td>1</td>
<td>$0.1 &lt; \kappa_{\text{vol}} &lt; 0.6$</td>
<td>0.3</td>
</tr>
<tr>
<td>$V_d$</td>
<td>1</td>
<td>$1 &lt; V_d &lt; 10 \ \mu g \ \text{m}^{-3}$</td>
<td>5 $\mu g \ \text{m}^{-3}$</td>
</tr>
<tr>
<td>$RH \frac{1}{1 - RH}$</td>
<td>$0.4 &lt; RH &lt; 1$</td>
<td>$0.7$</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2-1. Instrument setup schematic. Red lines indicate the sample flow, black lines indicate the sheath flow, orange lines indicate the cleaning cycle flow, and grey lines indicate no flow. The column central to the figure is the differential mobility analyzer (DMA). Sheath and sample flows in the diagram are downward in the center and upward at the sides. The cleaning cycle is similar to the unperturbed cycle but with solenoid valve A toggled such that flushing air flows toward the actuated ball valve G and is split between the sample path to the CPC and the inlet, reversing the flow through the temperature controlled preconditioner.
Figure 2-2. Top panel: Example time series from 8 June. Symbols correspond to article number concentration obtained from integration over the size distribution (red = unperturbed, blue = dry-humidified, black = dry). Grey line corresponds to 1 Hz particle concentration from the CPC upstream of the SMPS with the inlet opened to ambient air during cleaning cycles. The plot shown here is a sample plot of the comparisons for 8 June. Bottom panel: Example time series from 14 July. Symbols are the same as in the top panel. Grey line corresponds to 1 Hz particle concentration from a second CPC sampling side-by-side with the SMPS.
Figure 2-3. Example data from 5 July 2013. Histograms represent measured 60-bin representation of the number size distribution (top row) and derived volume size distributions (bottom row) for the unperturbed (left panels) dry-humidified (middle panels) and dry (right panels) instrument states. Red lines correspond to a bimodal log-normal distribution fitted to the data. Data-derived and fit-derived number and volume concentrations are provided in the legend. Blue numbers indicate the mode diameters obtained from the fit to the distributions.
Figure 2-4. From top to bottom in forefront of figure: Time series of 1) temperature ($T$, °C) in brown, 2) dew point temperature ($T_{dew}$, °C) in dark grey, 3) relative humidity ($RH$, %) in green, 4) aerosol dry volume ($V_d$, µm$^3$ cm$^{-3}$) in pink, 5) volumetric growth factor ($g_{v}_{ol}$, dimensionless) in light blue, 6) volumetric $k$ ($k_{v}_{ol}$, dimensionless) in orange, and 7) particle phase liquid water ($V_w$, µm$^3$ cm$^{-3}$) in dark blue. Precipitation (Precip, mm day$^{-1}$) is represented in grey bars at the bottom of the plot. All of the data except for precipitation were recorded by the SMPS. Gaps in the data indicate periods when the instrument was not operating. Precipitation data are from the ARA SEARCH network.
Figure 2-5. Summary number concentration (top row), surface area (middle row), and volume (bottom row) distributions for 3 June to 15 July 2013. The left panels indicate the unperturbed state, the middle panels the dry-humidified state, and the right panels the dry state. The central black lines show the median distribution, and the shaded gray regions represent the interquartile range.
Figure 2-6. Campaign-averaged diel profiles of total particle phase liquid water mass concentrations and RH binned into hourly intervals. The central lines indicate the median and the shaded regions indicate the interquartile range.
Figure 2-7. Diel profile of volumetric growth factor, $gf_{vol}$, binned into hourly intervals. The central lines indicate the median and the shaded area represents interquartile range. Grey dotted lines indicating median growth factors that were calculated using the same $RH$ and dry volume, and assumed $\kappa$ values ranging from 0 to 0.6.
Figure 2-8. The top, middle, and bottom panels depict diurnal trends of smaller diameter mode, larger diameter mode, and volumetric $\kappa$, respectively. The trends are campaign-averaged and binned into hourly intervals. The central lines indicate the median values and the shaded area represents interquartile range.
Figure 2-9. Diurnal trend of semi-volatile volumes lost during drying binned into hourly intervals. The central white line shows the mean of the measurements. The shaded blue area indicates a 95% confidence interval for the mean.
Figure 2-10. Diurnal trend of select aerosol chemical species measured by PILS-IC and PILS-WSOM. The central black line shows the median of the measurements. The shaded area indicates the interquartile range. Panel (e) shows the ratio of $\sum (b) + (c) + (d)$ over $\sum (a) + (b) + (c) + (d)$. Panel (f) is identical to Figure 8.
Figure 2-11. Diurnal trend of dry aerosol volumes binned into hourly intervals. The central black line shows the median of the measurements. The shaded red area indicates the interquartile range.
CHAPTER 3. DECREASING AEROSOL WATER IS CONSISTENT WITH ORGANIC CARBON TRENDS IN THE SOUTHEAST UNITED STATES

3.1 Abstract

Water is a ubiquitous and abundant component of atmospheric aerosols. It influences light scattering, the hydrological cycle, atmospheric chemistry, and secondary particulate matter (PM) formation. Despite the critical importance of aerosol liquid water, mass concentrations are not well known. Using speciated ion and meteorological data from the Southeastern Aerosol Research and Characterization network, we employ the thermodynamic model ISORROPIAv2.1 to estimate water mass concentrations and evaluate trends from 2001 to 2012 in urban and rural locations. The purpose of this study is to better understand the historical trends of aerosol liquid water in the southeast U.S. in the context of improved air quality and recently noted reductions in particulate organic carbon (OC). Aerosol water mass concentrations decreased by ~79% from 2001 to 2012 in the region, consistent with decreases in anthropogenic ion precursors. Decreases are more prominent in rural than in urban areas. Fractional contribution of water to PM also decreases during the same time period, and this is consistent with recently noted improvements in visibility. Decreases in mass concentrations of aerosol water are also consistent with decreases in mass concentrations of OC. These findings support the hypothesis that aerosol liquid water facilitates formation of biogenic secondary organic aerosol (SOA) and that biogenically derived SOA is modulated in the presence of anthropogenic perturbations.

3.2 Introduction

The Clean Air Act of 1970 and the Amendments of 1990 authorized national and
regional emission control programs enacted by the U.S. Environmental Protection Agency from 1990 to the present day that greatly reduced emissions of sulfur dioxide, nitrogen oxides, volatile organic compounds, carbon monoxide, and particulate matter (Hand et al., 2012). Implementation of performance standard rules provided decreases in acid deposition and large improvements to local and regional ambient air quality, in particular in the southeast U.S. (Butler et al., 2001; Likens et al., 2001; Blanchard et al., 2013; Hidy et al., 2014). Decreases in sulfate and nitrate PM$_{2.5}$ constituents during this time period are consistent with environmental rules specifically targeted at their gas phase precursors: SO$_x$ and NO$_x$. For example, in the U.S., anthropogenic SO$_2$ emissions declined 60% between 1990 and 2010 and ambient particulate sulfate mass concentrations decreased as a consequence (Hand et al., 2012). Title IV and Section 110 of the 1990 Clean Air Act Amendments contributed to the 50% reduction from 1999 to 2010 in emissions of anthropogenic NO$_x$ (Blanchard et al., 2013), a chemical species that impacts particulate nitrate mass concentrations (Mysliwiec and Kleeman, 2002).

In the southeast U.S., mean annual fine particulate organic carbon (OC) mass concentrations have decreased at monitoring locations that are part of the Southeastern Aerosol Research and Characterization (SEARCH) network over the past decade (Blanchard et al., 2013; Attwood et al., 2014; Hidy et al., 2014). This is intriguing because environmental rules have not been specifically designed to reduce ambient particulate OC mass concentrations, though recent fuel economy standards (United States Environmental Protection Agency, 2012) could reduce the fossil fraction of ambient particulate OC mass. Plausible explanations for the recent decreases in OC mass at the SEARCH sites are complicated by the fact that OC is complex, involving both primary
and secondary material. The carbon in the southeast U.S. derives predominantly from the biosphere (Kleindienst et al., 2007; Weber et al., 2007) and is therefore not mainly derived from petrochemical emissions of volatile organic compounds (VOCs) that have a fossil carbon signature. Secondary organic aerosol (SOA) forms from interactions of anthropogenic and biogenic precursors (Hallquist et al., 2009; Carlton et al., 2010b; de Gouw et al., 2011; Hoyle et al., 2011) through semi-volatile partitioning (Pankow, 1994) and water-mediated chemistry (Volkamer et al., 2009).

Modeling experiments suggest that the dominant SOA formation pathway in the southeast is through reactions in aerosol liquid water (Carlton and Turpin, 2013). Ambient concentrations of aerosol liquid water are controlled by meteorology and chemistry, as aerosol liquid water is present in the condensed phase as a function of relative humidity ($RH$), temperature, aerosol concentration and chemical composition (Zhou et al., 2011). A prior modeling study estimated aerosol liquid water was high in the eastern U.S. because of high $RH$ and high concentrations of sulfate in the region in 2003, the period corresponding to the start of this trends analysis (Carlton and Turpin, 2013). In the eastern U.S., field studies demonstrate that sulfate increases the amount of aerosol liquid water concentrations at a given relative humidity (Malm et al., 1994; Sisler and Malm, 1994). Such anthropogenic influences on particle hygroscopicity and the amount of aerosol water are well-established in theory (Nenes et al., 1998) and in field studies (Zhang et al., 1993; Malm et al., 1994; Pitchford and Mcmurry, 1994; Sisler and Malm, 1994; Khlystov et al., 2005; Hodas et al., 2014). During the Southern Oxidant and Aerosol Study in the southeast U.S., aerosol water mass was maximum when the aerosol hygroscopicity (i.e. $\kappa$) was highest, corresponding to when the inorganic fraction was
Aerosol water can serve as a medium to partition polar, water-soluble gas phase organic gases to the condensed phase (Asa-Awuku et al., 2010; Parikh et al., 2011), facilitating SOA formation (Carlton et al., 2009; Ervens et al., 2011; Carlton and Turpin, 2013; Hodas et al., 2014). Laboratory studies indicate organic material can also modulate particle hygroscopicity and water uptake in both directions, depending on the nature of the organics (Sareen et al., 2013). Field studies in the Great Smoky Mountains demonstrate a positive relationship between aerosol water and particle organic content (Dick et al., 2000). The fraction of organic material in the particle phase often increases with increasing RH (Ervens et al., 2011), in particular in the southeast U.S. (Hennigan et al., 2008; Hennigan et al., 2009; Zhang et al., 2012).

Trends in aerosol liquid water mass concentrations are not well known because historical measurements of the water component of ambient aerosols have been limited and conventional PM mass measurement techniques remove water from the aerosol. There is a critical need to understand the relationship between anthropogenic influences on aerosol liquid water and subsequent biosphere-atmosphere interactions that form particulate OC. Regional studies of trends in aerosol constituents are also needed to evaluate the effectiveness of existing rules in the context of ongoing improvements in air quality because meteorological conditions and atmospheric chemical regimes vary from region to region and in a changing climate. Future energy choices and environmental regulation are key uncertainties that make future prediction of pollutants difficult (Jacob and Winner, 2009), but in the southeast U.S., future ozone concentrations are predicted to increase due to climate change (Hogrefe et al., 2004). Global weather patterns are predicted to be more stagnant, which is typically associated with poor air quality (Jacob
and Winner, 2009) and water vapor is expected to increase (Boucher et al., 2004). It is plausible that water-mediated SOA formation will be more prevalent with the changing climate.

In this work, we test the hypothesis that observed decreases in organic aerosol mass concentration are partially attributable to changes in the liquid water partitioning medium. Specifically, for the southeast United States from 2001 to 2012, we evaluate the correlation of measured temporal trends in OC with trends in semi-empirical observational estimates of ambient aerosol liquid water concentrations, which are estimated from the thermodynamic model ISORROPIA (v2.1) based on speciated inorganic aerosol composition and meteorological measurements from SEARCH sites. We also compare differences in aerosol liquid water concentrations between urban and rural sites. We focus on the southeast U.S. because of the recent attention drawn to the region due to the cooling haze and decreasing organic aerosol concentrations (Goldstein et al., 2009; Blanchard et al., 2013; Attwood et al., 2014) and because of the recent field campaigns in the area such as the Southern Oxidant and Aerosol Study (SOAS; http://soas2013.rutgers.edu).

3.3 Experimental Section

3.3.1 Measurements of Surface Air Quality and Meteorology

The SEARCH program is a surface network of eight routinely monitored sites across four states in both urban and rural locations (Hansen et al., 2003; Edgerton et al., 2005, 2006) (Figure C-1). Birmingham (BHM), Pensacola (PNS), Jefferson Street (JST), and Gulfport (GFP) are classified as urban sites in Alabama, Florida, Georgia, and Mississippi, respectively. Centreville, Pensacola-outlying aircraft landing field (OLF),

### 3.3.2 Water Associated With Inorganics

Mass concentrations of particle phase liquid water for all sites are estimated using the inorganic aerosol thermodynamic equilibrium model ISORROPIA2.1 (Nenes et al., 1998; Fountoukis and Nenes, 2007). Water mass is calculated based on inorganic species ($NH_4^+$, $SO_4^{2-}$, $NO_3^{-}$) mass concentrations, temperature and $RH$. ISORROPIA assumes a $NH_4^+$-$SO_4^{2-}$-$NO_3^{-}$ metastable system, which is consistent with the recent finding that aerosols in the southeast always contain liquid water (minimum 7% water by mass) (Nguyen et al., 2014b) and these ions (Guo et al., 2014). The assumption of
thermodynamic equilibrium between the gaseous and condensed phases of inorganic aerosol has been shown to be valid within the errors of field measurements via comparisons of field data with gas/particle partitioning data calculated by ISORROPIA (Moya et al., 2001; Zhang et al., 2002; Yu et al., 2005). ISORROPIA is run in reverse mode because gaseous concentrations of NH$_3$ for the SEARCH sites are not available for the entire time period.

### 3.3.3 Water from organic fraction

ISORROPIAV2.1 does not consider organic compounds. Laboratory studies have shown that the organic fraction of atmospheric aerosols can change particle hygroscopicity depending on the chemical composition (Saxena et al., 1995; Cruz and Pandis, 2000; Dick et al., 2000; Suda et al., 2012; Sareen et al., 2013). The organic aerosol fraction is complex (Decesari et al., 2000; Shimmo et al., 2004) and not fully defined qualitatively or quantitatively. However, the Zdanovskii-Stokes-Robinson (ZSR) mixing rule could be used to describe hygroscopic growth of most aerosol mixtures (Choi and Chan, 2002; Svenningsson et al., 2006; Petters and Kreidenweis, 2007). Thus, we estimate potential excess water from the organic components according to $\kappa$-Kohler theory and the ZSR mixing rule, as described in Petters and Kreidenweis (2007) and Kreidenweis et al. (2008) (Eq. 3-1):

$$V_{w,o} = V_o \kappa_{org} \frac{a_w}{1-a_w},$$  \hspace{1cm} (3-1)

where $a_w$ is water activity (dimensionless) as defined in Eq. 3-1, $V_{w,o}$ and $V_o$ are the volumes of water and organic matter ($\mu$m$^3$ cm$^{-3}$), and $\kappa_{org}$ is the hygroscopicity parameter of the organic component (dimensionless). $V_o$ is calculated by dividing the mass of OM by a typical organic density of 1.4 g cm$^{-3}$ (Turpin and Lim, 2001). $\kappa_{org}$ depends on
particle chemical composition. Since historical $\kappa_{\text{org}}$ data for the SEARCH locations are not well known, we explore sensitivity to this parameter by employing a range in $\kappa_{\text{org}}$ between 0.01 and 0.2, which is a typical range for organic species (Petters and Kreidenweis, 2007), in our calculations to demonstrate how organic aerosol liquid water mass concentrations may influence the decadal trend as a function of $\kappa_{\text{org}}$. For simplicity and due to lack of data on particle diameter, we assume $a_w$ is equivalent to $RH$. The expected difference between $RH$ and water activity are between 0.01 and 0.02 in absolute units (Tang, 1996), and previous studies indicate that this may result in an $RH$ dependent overestimate in $\kappa_{\text{vol}}$ ranging from 4 to 11% (Nguyen et al., 2014b).

### 3.3.4 Statistical Analysis

Both linear and non-linear trends in the data are analyzed. We use the locally weighted scatter plot smoothing (LOESS) function in the R statistical software package (www.r-project.org/) to analyze non-linear trends in the observationally estimated data (R Core Team, 2013). The LOESS function utilizes a nearest neighbors algorithm to determine a least squares fit for localized subsets of data. For linear trends, we use the lm function in R to compute regression statistics. We also use the ggplot2 package (Wickham, 2009) to fit a generalized additive model with integrated smoothness estimation to some of the data, as indicated in the relevant figures. We also test whether removing forest fire time periods from the data affect the water and OC trends by screening concentrations of carbon monoxide $> 700$ ppb. For this analysis, water mass estimates for $RH$ levels between 80% and 95%, where aerosols are deliquesced, are
separated by month, into three \( RH \) bins: 80-85\%, 85-90\%, and 90-95\%, and screened for potential effects of forest fires.

### 3.4 Results and Discussion

Across the SEARCH network, surface mass concentrations of aerosol liquid water associated with inorganic species as estimated by ISORROPIA decreased by 79\% from 2001 to 2012 (Figure 3-1 and C-2). The decrease is driven largely by trends at the rural sites, which exhibit an 82\% decrease in aerosol liquid water compared to a 51\% decrease over the decade at urban monitoring locations. Linear trend analysis indicates the overall aerosol liquid water mass decreases for all SEARCH sites with time are robust (\( R^2 = 0.77 \)) and statistically discernible (\( p = 0.0002 \)) (Table 3-1). Urban water decrease (\( R^2 = 0.52 \)) is not as robust as rural water decrease (\( R^2 = 0.69 \)), possibly due to rising urban \( \text{NO}_3^- \) and \( \text{NH}_4^+ \) concentrations from 2001-2005 (Figure 3-2). In contrast, rural \( \text{NO}_3^- \) and \( \text{NH}_4^+ \) concentrations decreased from 2001-2005. Both urban and rural \( \text{SO}_4^{2-} \) increased during this time period, suggesting that the differences in water trends at urban and rural areas are not fully represented by \( \text{SO}_4^{2-} \). Uncertainty in the estimates of aerosol water is difficult to quantify because this usage of ISORROPIA neglects water uptake by organic compounds and other ions. Discrepancies in PM mass concentrations are estimated to vary between 3–7\% (Edgerton et al., 2005). Adjusting the speciated PM data used in this study in ISORROPIA to account for this range of error results in a median ~5\% error in aerosol water estimates (Figure 3-1), which varies according to temperature and RH. Further uncertainties due to fires and organic compounds are discussed later on in this chapter. Overall, across the SEARCH network decreases in mass concentrations of the inorganic compounds from 2001-2012 are robust and statistically discernible.
Semi-empirical observational estimates of aerosol liquid water mass concentrations exhibit interannual variability consistent with local meteorology (e.g. temperature and $RH$; Figures 3-2a and 3-2b; Figure 3-3), but decadal trends are influenced by the decreasing ion concentrations of $SO_4^{2-}$, $NO_3^-$, and $NH_4^+$ (Figures 3-2c-f) that affect particle hygroscopicity. Higher temperatures in urban areas (Figure 2a) suggest an urban heat island effect (Oke, 1973; Kim, 1992) that reduces local $RH$ and impedes aerosol liquid water uptake. Despite the higher concentrations of hygroscopic inorganic compounds such as sulfate, nitrate, and ammonium in urban areas (Figure 3-2c-e), ISSORROPIA-estimated aerosol liquid water is less. Rural areas exhibit higher $RH$ levels than urban areas (Figure 3-2b), and atmospheric aerosols grow with increasing $RH$ (Tang and Munkelwitz, 1993; Seinfeld and Pandis, 2006). The network average of 52 $\mu$g m$^{-3}$ for aerosol liquid water in 2001 decreased to 11 $\mu$g m$^{-3}$ in 2012 (Table C-1). Recent direct measurements of aerosol liquid water during the SOAS campaign at the SEARCH CTR site averaged 1-5 $\mu$g m$^{-3}$ (Nguyen et al., 2014b). These values are lower than the SEARCH 2012 average and suggest the overall decreasing trend in aerosol liquid water in the southeast U.S. at SEARCH monitoring locations continues until at least the summer of 2013.

Monthly analysis of aerosol liquid water data reveals seasonal differences that drive the yearly trend (Figure 3-3 and 3-4). No noticeable offset in water trends is detected due to fires as screened (concentration values on days and locations when CO > 700 ppb) in this work. Other fire screening methods may not influence aerosol water trends either because they are also unlikely to perturb decadal trends in meteorology and anthropogenic inorganic chemical species that impact particle hygroscopicity and drive
the decrease in aerosol water. Seasonal differences in the decadal water trends in the southeast region, however, are noticeable, and these trends may be explained by both local meteorology and particle hygroscopicity. The most noticeable decrease in water concentrations occurs May to October with less interannual variability from January to April and November to December. In the colder months, mass concentrations of \( \text{SO}_4^{2-} \) and \( \text{NH}_4^+ \) are low due to emissions and formation chemistry (Seinfeld and Pandis, 2006) and control the relatively low aerosol water concentrations (Figure C-3). A peak in water content occurs before 2006 and is detectable from February to October (Figure 3-3). Figure 3-2b suggests that these years have high \( RH \), which may have resulted in environments that facilitate aerosol liquid water formation and phase transition of gas phase water. Additionally, Figure 3-2c shows that sulfate mass concentrations decreased after 2006, possibly due to increased gas use (switch away from coal) (Fichman, 2011; de Gouw et al., 2014) and diesel fuel reformation (United States Environmental Protection Agency, 2000). We note that between 2010 and 2012, measured urban temperature increased by over 2 °C, but measured urban \( RH \) levels do not decrease proportionally. The interactions between meteorology and aerosol chemistry suggest complexities in the aerosol water trends.

Robust and statistically discernible particle OC reductions (Table 3-1) are consistent with aerosol liquid water reductions (Carlton and Turpin, 2013). Decreases in anthropogenic emissions from combustion sources, including diesel and gasoline engine emissions, and biomass combustion (Blanchard et al., 2012) could reduce organic aerosol concentrations, but previous summertime studies in Texas (Lemire et al., 2002), Georgia (Weber et al., 2007), Tennessee (Lewis et al., 2004), and Alabama (Xu et al., 2015)
demonstrate that particulate OC in the southeast is primarily secondary biogenic, even in urban areas. In the wintertime, fossil contribution to OC can be large though the fractional contribution is uncertain (Ding et al., 2008). Decreases in primary non-fossil OC emissions could also potentially reduce ambient OC concentrations. Ambient isoprene mixing ratios increase from 2001 and 2012 (Figure C-4) either because biogenic emissions increased or loss processes decreased (i.e., less OH oxidation). Although the hydroxyl radical is the main chemical daytime sink for isoprene, measurements are not routine and it is difficult to robustly evaluate trends. Photolysis of ozone in the presence of water vapor is the main source of OH (Monks, 2005) and median ozone concentrations from 2001-2012 are relatively stable (Figure C-5). Stable ozone concentrations may be a useful indicator for southeast oxidative capacity over the decade. However, decreases in NO\textsubscript{x} could result in changes to gas phase OH oxidation of isoprene (Kroll et al., 2005; Carlton et al., 2009). If isoprene emissions are increasing, this is not suggestive that other biogenic emissions such as primary OC emissions are decreasing, but the cause of rising isoprene mixing ratios is an open question.

Trends in anthropogenic NO\textsubscript{x} mixing ratios during this time modulate biogenic SOA mass concentrations through complex non-linear mechanisms. NO\textsubscript{x} can hinder biogenic SOA formation during gas-to-particle partitioning in dry (low RH) environments for isoprene (Kleindienst et al., 2007; Carlton et al., 2009; Chan et al., 2010) and enhance SOA yields from sesquiterpenes (Ng et al., 2007). NO\textsubscript{x} facilitates formation of the nitrate radical that can form SOA from monoterpenes and sesquiterpenes during nighttime chemistry (Fry et al., 2014). Process-level modeling suggests high NO\textsubscript{x} enhances SOA formation from isoprene in clouds (Ervens et al., 2008) and possibly in wet aerosols. In
the southeast U.S., photochemically active periods tend to be dominated by high $RH$ and high biogenic emissions since temperature is a major driver of emission and evapotranspiration (Hennigan et al., 2008) and water-mediated chemistry is predicted to dominate SOA formation in the southeast (Carlton and Turpin, 2013). While NO$_x$ mixing ratios have been declining the eastern U.S., the impact on SOA mass is unclear and difficult to detangle without similar trend analysis in specific tracer species, which are not available.

Organic compounds may vaporize out of particle phase to the gas phase in a warming climate, but temperatures have not increased in the southeast (Figure 3-2a). Additionally, isoprene epoxydiol (IEPOX) SOA chemistry is enhanced in the presence of acidified sulfate seed aerosol and condensed-phase nucleophiles (Surratt et al., 2010; Nguyen et al., 2014a). Calculations of an acidity index using ion charge balance – Eq. (6) from Guo et al. (2014) and an assumed $\kappa_{org}$ of 0.13, consistent with the $\kappa_{org}$ mean of 0.126 that the same authors measured at CTR – suggest increasing aerosol alkalinity, but the data are not robust ($R^2 = 0.35$) (Figure C-6). Recent literature also indicate high aerosol acidity in the Southeast despite declining sulfate concentrations (Weber et al., 2016). Thus, potential decadal trends in aerosol acidity and effects on organic aerosol mass concentrations remain unclear.

Aerosol liquid water mass fraction has decreased between 2001-2012 ($R^2 = 0.77$, p-value < 0.05) for the southeast U.S. region. The $R^2$ value between mean annual water and mean annual OC is 0.45 (p-value < 0.05), which does not suggest a strong relationship between the two parameters, but the $R^2$ value between the LOESS water trends and the LOESS OC trends are 0.81 (p-value < 0.05), which does indicate a strong
relationship. The statistically discernible correlation in decadal scale reductions between these LOESS trends indicates that aerosol liquid water may play a role in OC reductions.

We also investigate how aerosol composition in the southeast U.S. changed from 2001-2012. Figure 4 indicates the mean annual PM$_{2.5}$ aerosol compositional mass concentrations for SEARCH sites from 2001-2012, where SEARCH estimates organic matter (OM) to be $1.4 \times OC$, although this number has been shown to vary (Turpin and Lim, 2001). The higher aerosol water mass concentrations in August compared to January suggest a conducive summertime environment for aqueous chemistry. Previous studies conducted in the southeast U.S. in Atlanta show positive correlations between aerosol water and partitioning of water soluble organic compounds, particularly at $RH >70\%$ (Hennigan et al., 2009).

The decrease in aerosol water fractional contribution to total aerosol mass by 23% between 2001 and 2012 may limit the formation of organic material through aqueous partitioning, which may contribute to the decrease in observed OC concentrations. The decrease in the fractional contribution of water to PM is consistent with recently improved visibility in the southeast U.S. (Attwood et al., 2014). That study also found increasing organic aerosol fraction (Attwood et al., 2014), but we find no statistically discernible change in the OC to SO$_4^{2-}$ ratio even when fires are screened (Figure C-7). We also observe that the OC to (SO$_4^{2-}$ + NO$_3^{-}$) ratios do not indicate any clear trend. Secondary organic carbon (SOC) was estimated for the SEARCH sites using an EC-tracer method that employs monthly average primary ratios of transported primary OC to elemental carbon (EC) ratios from Carlton et al. (2010a). The data indicate an overall decrease in SOC until 2009 (Figure C-8), although monthly trends in SOC are complex
(Figure C-9), possibly because the monthly average ratios are not sufficient. These estimates may be improved through future work on data gained from SOAS. Alternatively, water might not be the limiting factor for IEPOX SOA (Nguyen et al., 2014a) when concentrations are high, such as during this time period. It is also possible that concentrations of OC from fires are not adequately screened.

The data further show that the excess water due to organics may be generally higher at rural sites (Dick et al., 2000; Hennigan et al., 2009; Ervens et al., 2011) than at urban sites, possibly due to higher RH levels at rural sites, more hygroscopic aerosol, or hydrophobic organic compounds and soot that limit water uptake at urban sites. To explore uncertainty in aerosol liquid water trends induced by not including organic material effects on particle hygroscopicity by ISSOROPIA, we conduct seven sensitivities in the trend for a range of κ values reported in the literature. We estimate the 24-hour average aerosol water content if measured OC contributed to total aerosol water uptake (Figure 3-5). The average percent difference between inorganic water and total inorganic and organic water for all years and hygroscopicity levels is 16%, 28%, and 8% for overall, at the rural sites, and at the urban sites, respectively. However, the OC contribution to aerosol water is highly variable, largely uncertain, contingent upon RH levels. The uncertainty (and neglect in ISORROPIA) of organic contribution to aerosol water does not change the conclusion that aerosol water mass has been decreasing.

These results are consistent with the hypothesis that anthropogenic inorganic species modulate biogenic SOA through aerosol liquid water effects. We investigated five possible reasons as to why organic aerosol concentrations in the Southeast U.S. have decreased over the past decade. These are: 1) decreases in anthropogenic VOC
concentrations, 2) decreases in primary non-fossil organic carbon concentrations, 3) increases in temperature, 4) increases in pH, and 5) decreases in aerosol water mass concentrations. Reasons 1-3 are ruled out while uncertainties remain with reason 4. Exploration of reason 5 demonstrates consistencies between decreases in aerosol water mass concentrations and decreases in organic aerosol mass concentrations. Continued field evaluation with aerosol water measurements may provide useful constraints to improve and develop thermodynamic models such as ISORROPIA and others (Clegg et al., 2001; Zuend et al., 2008) and ultimately to connect measurements of aerosol water in space and time. More research into how aerosol composition and water trends in the southeast U.S. compare with different regions in the U.S. and world are also needed to further test this hypothesis.

3.5 References


Table 3-1. Linear trend analysis (lm function in R) for a) mean annual averages of parameters with time and b) mean annual averages of parameters with mean annual averages of water.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$R^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Parameter vs. time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall water</td>
<td>0.772</td>
<td>0.0002</td>
</tr>
<tr>
<td>Rural water</td>
<td>0.692</td>
<td>0.0008</td>
</tr>
<tr>
<td>Urban water</td>
<td>0.518</td>
<td>0.0083</td>
</tr>
<tr>
<td>Water fraction</td>
<td>0.771</td>
<td>0.0002</td>
</tr>
<tr>
<td>OC</td>
<td>0.634</td>
<td>0.0019</td>
</tr>
<tr>
<td>$SO_4$</td>
<td>0.758</td>
<td>0.0002</td>
</tr>
<tr>
<td>$NO_3$</td>
<td>0.826</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>$NH_4$</td>
<td>0.708</td>
<td>0.0006</td>
</tr>
<tr>
<td>(b) Parameter vs. water</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OC</td>
<td>0.451</td>
<td>0.0168</td>
</tr>
<tr>
<td>Rural OC vs. rural water</td>
<td>0.286</td>
<td>0.0735</td>
</tr>
<tr>
<td>Urban OC vs. urban water</td>
<td>0.341</td>
<td>0.0461</td>
</tr>
<tr>
<td>OC loess (vs. water loess)</td>
<td>0.807</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>$SO_4$</td>
<td>0.634</td>
<td>0.0019</td>
</tr>
<tr>
<td>$NO_3$</td>
<td>0.724</td>
<td>0.0004</td>
</tr>
<tr>
<td>$NH_4$</td>
<td>0.684</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
Figure 3-1. Local regression (LOESS) analysis of aerosol liquid water estimates and organic carbon measurements for all SEARCH sites from 2001-2012. Trends for the entire SEARCH network for water and OC are depicted in black and green, respectively. The data for both water and OC have also been separated by urban (grey, dashed) and rural (orange, dashed with circles) monitoring sites. The red error bars on the overall water plot indicate the error introduced to the water measurements due to measurement uncertainty.
Figure 3-2. LOESS trends in a) temperature, b) relative humidity, c) sulfate, d) nitrate, e) ammonium, and f) organic carbon for urban (grey) and rural (orange) sites.
Figure 3-3. LOESS monthly trends in aerosol liquid water in the southeast U.S. from 2001-2012 for $80 \leq RH < 85$ (red), $85 \leq RH < 90$ (green), and $90 \leq RH < 95$ (blue). Dotted lines indicate screening out fires ([CO] $>$ 700 ppb are excluded).
Figure 3-4. Mean annual PM$_{2.5}$ aerosol compositional mass concentrations for SEARCH sites from 2001-2012 using estimated inorganic ion derived water content. Note the change in y-axis scale for water mass concentrations for August (bottom panel).
Figure 3-5. Mean annual averages of estimates of liquid water content from organic fraction for overall (black), urban (grey), and rural (orange) data. The dashed lines indicate ranges of organic water for $\kappa$ ranging from 0.01 (lowest dashed line) to 0.2 (highest dashed line). Bolded lines indicate water content from the inorganic component.
CHAPTER 4. RECONCILING SATELLITE AEROSOL OPTICAL THICKNESS AND SURFACE FINE PARTICLE MASS THROUGH AEROSOL LIQUID WATER

4.1 Abstract

Summertime aerosol optical thickness (AOT) in the Southeast U.S. is sharply enhanced compared to wintertime AOT. This seasonal pattern is unique to the Southeast and is of particular interest because temperatures have not warmed over the past 100 years, contrasting with trends in other U.S. regions. In this work, we attempt to reconcile the spatial and temporal distribution of AOT over the U.S. with surface mass measurements by examining trends in aerosol liquid water, an aerosol constituent that scatters radiation and is detected by AOT measurements but removed in mass measurements at routine surface monitoring sites. We employ the thermodynamic model ISORROPIAv2.1 to estimate aerosol water mass concentrations at IMRPOVE sites using measured ion mass concentrations and NARR meteorological data. Our findings suggest aerosol water provides a plausible explanation for the geographical and seasonal patterns in AOT and can reconcile previously noted discrepancies with surface mass measurements.

4.2 Introduction

Accurate understanding of the fate and transport of trace species in the atmosphere and their impacts on air quality and climate relies on consistent and long-term measurements from a variety of platforms. As these platforms often observe different aspects of the same phenomena, have different strengths and shortcomings, the collected observations can lead to different scientific conclusion, and hence need to be reconciled. A recently noted example of such discrepancy is between space-based aerosol
optical thickness (AOT) measurements and surface-based measurements of aerosol mass (Ford and Heald, 2013). Goldstein et al. (2009) previously demonstrate that summertime AOT in the Southeast U.S., location of the “warming hole” (Meehl et al., 2012), is enhanced over wintertime values by a factor of 2-3 (Goldstein et al., 2009). The high AOT and strong seasonality evident in data from the MultiAngle Imaging Spectro-Radiometer (MISR) (Kahn et al., 2005) and the Moderate Resolution Imaging Radiometer (MODIS) (Kaufman et al., 1997) is not characteristic of other regions of the U.S. They hypothesize that anthropogenic and biogenic emissions combine to form secondary organic aerosol (SOA), a near-term climate forcer that results in a cooling haze over the Southeast U.S. This conclusion was formed from consistencies between the spatial distribution, seasonal distribution, and temperature dependence of natural biogenic volatile organic compound (BVOC) emissions with AOT as well as observations of modern $^{14}$C and BVOC oxidation products with summer maxima. However, Ford and Heald (2013) note that surface mass measurements of organic aerosols in the Southeast U.S. do not exhibit the same seasonality. These authors conclude that surface measurements do not support the biogenic SOA and cooling haze hypothesis because seasonality in surface mass is insufficient to explain AOT observations. They suggest the possibility of an aloft source of organic aerosol below 700hPa, possibly fire-related, to explain the seasonal AOT differences. Fire counts in the region nearly double from winter to summer in 2007-2009 (Ford and Heald, 2013), but biomass burning in the Southeast generally peaks in the late winter and early spring (Zeng et al., 2008; Tian et al., 2009; Zhang et al., 2010). Both studies rule out aerosol liquid water uptake as an explanation for the geospatial pattern and seasonality in Southeast AOT through analysis
of only relative humidity (RH) seasonality. This methodology is insufficient because aerosol water is driven by not only changes in RH, but also by changes in temperature, particle concentration, and chemical composition (Petters and Kreidenweis, 2007; Nguyen et al., 2014).

In this work, we explore the hypothesis that aerosol water, removed in mass measurements but observed by radiation measurements, reconciles satellite AOT observations and surface mass measurements. Aerosol water is an abundant and ubiquitous component of tropospheric aerosol volume (Martin et al., 2008; Lee and Adams, 2010), with the highest concentrations over the continental U.S. predicted to be in the eastern part of the country (Liao and Seinfeld, 2005; Carlton and Turpin, 2013) as a consequence of high RH and aerosol chemical properties. A recent modeling study estimated aerosol liquid is high in the eastern U.S. because of high RH and high concentrations of sulfate in the region in 2003 (Carlton and Turpin, 2013). Additionally in the eastern U.S., field studies demonstrate that sulfate increases the amount of aerosol water concentrations at a given relative humidity (Malm et al., 1994; Sisler and Malm, 1994). Trends in aerosol water mass concentrations are not well known because historical measurements of the water component of ambient aerosols have been limited and conventional particulate matter (PM) mass measurement techniques remove water from the aerosol. Historically, field studies predicting aerosol water focused on visibility and characterizing particle growth factors as a function of RH for single particles and the bulk aerosol population (Zhang et al., 1993; Pitchford and Mcmurry, 1994; Malm et al., 1996; Dick et al., 2000; Kroll et al., 2005; Hennigan et al., 2008; Sorooshian et al., 2008). In the Southeast U.S., recent quasi-direct measurements during the Summer Oxidant and
Aerosol Study (SOAS) indicate that aerosol water is present, abundant, and ubiquitous during the summer (Nguyen et al., 2014; Guo et al., 2015).

Satellites measure radiation and it is well established that aerosol water effectively scatters radiation in the visible part of the spectrum (e.g. Malm et al., 1994; Park et al., 2004; Pitchford et al., 2007). On the contrary, routine surface networks and most particle measurement techniques remove aerosol water during sampling or filter equilibration. Numerous studies have included water when reconstructing extinction to predict and/or monitor visibility. Historically, calculations of scattering efficiencies were done by applying Mie theory to measured size-resolved chemically speciated PM species (e.g. Ouimette et al., 1981; Hasan and Dzubay, 1983; Sloane, 1983; Sloane and Wolff, 1985; Zhang et al., 1994) and by regression-derived scattering efficiencies (e.g. Malm et al., 1996; Sisler and Malm, 2000; Malm et al., 2003; Park et al., 2006; Pitchford et al., 2007). Discrepancies between measured extinction of unperturbed aerosol when compared to Mie theory calculations applied to measured (and dry) aerosol size distributions are reduced with the acknowledgment of the size distribution of unperturbed (and wet) aerosols. In regression-derived scattering efficiencies, water uptake is accounted for as a function of $RH$ on a sampling-period-by-sampling-period basis using Mie theory and an assumed size distribution and laboratory measured aerosol growth curves. For the Southeast U.S., contributions by aerosol water were not adequately resolved in the previous literature. In determining if AOT could be affected by hygroscopic growth of aerosols, Goldstein et al. (2009) did not find a significant relationship between $RH$ and AOT at the Walker Branch AERONET site, while Ford and Heald (2013) converted the differences in $RH$ to differences in aerosol mass extinction
efficiency and did not find large seasonal differences. Our work is novel because for the first time in this field, we estimate regional and spatial trends in aerosol water mass concentrations across the continental U.S. and reconcile these trends with trends in AOT.

Here we apply the inorganic thermodynamic model ISORROPIA (v2.1) to simulate aerosol water at the surface and throughout the planetary boundary layer (PBL) from chemically-speciated particle mass measurements at Interagency Monitoring of PROtected Visual Environments (IMPROVE) sites spanning across the U.S. from 1988–2013. The IMPROVE program (http://vista.cira.colostate.edu/improve/) was initiated in 1987 for the implementation of extensive long term monitoring of visibility impairment and aerosol chemistry in national parks and remote wilderness areas across the U.S. at the surface (Malm et al., 1994; Malm et al., 2004). In the IMPROVE algorithm for visibility, the sulfate and nitrate mass extinction efficiency terms include a water growth factor that is a function of $RH$ (displayed as $f(RH)$ multiplied by a constant dry extinction efficiency) (Pitchford et al., 2007). We test the hypothesis that liquid water, removed in surface measurements, reconciles satellite AOT observations and surface mass measurements. To do so we compare regional and temporal trends in aerosol water across the continental U.S. to provide plausible explanations for the current debate in the literature.

4.3 Experimental Section

IMPROVE data were downloaded from public archives (http://views.cira.colostate.edu/fed/DataWizard/Default.aspx) on July 13, 2015 for 210 unique locations (Figure D-1) spanning the continental U.S. from 1988-2013. We separate the IMPROVE sites into six regions as defined in Figure D-1: Northeast, Southeast, Midwest, Central, Mountain West, and West Pacific. We use IMPROVE
surface data for $\text{SO}_4^{2-}$, $\text{NO}_3^-$, and organic matter (OM). IMPROVE defines OM as 1.8 times measured organic carbon (OC) mass, which is within the range of 1.58 to 2.07 OM/OC ratios estimated experimentally at national parks in the IMPROVE network (El-Zanan et al., 2005), and the range suggested by Turpin and Lim (2001). In general, the speciated data are available every three days. IMPROVE assumes all sulfate is in the form of ammonium sulfate and all nitrate is in the form of ammonium nitrate (Malm et al., 2004), though some studies have shown these assumptions to be inaccurate at some locations for sulfate (e.g. Gebhart et al., 1994; Malm et al., 2000; Tolocka and Turpin, 2012) and nitrate (e.g. Malm et al., 2003). Inaccuracies of this assumption would affect aerosol hygroscopicity and thus introduce uncertainty into the magnitudes of the aerosol water calculations, but should not affect the aggregated spatial and seasonal trends in aerosol water.

Mass concentrations of particle phase liquid water for all sites are estimated using the inorganic aerosol thermodynamic equilibrium model ISORROPIA v2.1 (Nenes et al., 1998; Fountoukis and Nenes, 2007). The ISORROPIA model used in this study assumes $\text{SO}_4^{2-}$-$\text{NO}_3^-$ metastable particles, which is applicable in the Southeast, where measurements during SOAS found that aerosols contained these ions (Guo et al., 2015) and at minimum 7% water by mass (Nguyen et al., 2014). $\text{NH}_4^+$ is not considered in the model runs due to lack of full data records for all sites, which will result in lower-bound estimates of absolute aerosol water mass concentrations because water uptake is typically assumed to be additive (Stokes and Robinson, 1966). Regional trends in aerosol water are valid assuming no systematic differences in ammonia emissions over the considered time period. However, lower-bound estimates of aerosol water are expected in agricultural
areas or locations with large numbers of Concentrated Animal Feeding Operations (CAFOs) and associated ammonia emissions (e.g. Midwest and Southeast U.S. (Goebes et al., 2003; Wilson and Serre, 2007; Saylor et al., 2010)). Another source of uncertainty is that ISORROPIAv2.1 does not consider water uptake by organic compounds, though laboratory studies demonstrate that organic constituents can modulate particle hygroscopicity depending on chemical composition (Saxena et al., 1995; Cruz and Pandis, 2000; Dick et al., 2000; Suda et al., 2012; Sareen et al., 2013). Yet this is difficult to quantify because the effects of organic compounds on aerosol water are complex (Decesari et al., 2000; Shimmo et al., 2004). While OM introduces more aerosol water, inorganic water dominates over organic contributions (Choi and Chan, 2002). These semi-observational estimates by ISORROPIA likely provide a lower bound for aerosol water mass concentrations, and large-scale aerosol water trends in space and time are valid in a relative sense (Nguyen et al., 2014). Measured speciated chemical data are used with ISORROPIA to calculate historical trends in semi-empirical observational estimates of aerosol water for the time period of 1988-2013 and to evaluate seasonal trends to reconcile the results of Goldstein et al. (2009) and Ford and Heald (2013). Vertical profiles of chemical composition are unavailable for the aloft analysis, so we assume a well-mixed planetary boundary layer, consistent with recent findings in the Southeast that suggest that vertical profiles of chemically specific constituents are similar within the PBL (Wagner et al., 2015). Seasonal differences in the data are calculated using the same summer and winter definitions as in Goldstein et al. (2009) and Ford and Heald (2013). Summer is representative of the months of June, July, and August while winter is representative of the months of December, January, and February.
To ensure consistency among the meteorological data across the IMPROVE network and as co-located RH and temperature data was unavailable at many sites, we used output from the North American Regional Reanalysis (NARR) model [Mesinger et al. 2006]. The profiles of specific humidity and temperature together with values of PBL depth were extracted from the grid boxes closest to the locations of the monitoring stations. The NARR data is available 8-times daily at 29 pressure levels and 0.3° (~32 km) resolution for the North American region. Although the NARR model simulates the real atmosphere by assimilating available observations, previous studies have shown that the model fails to accurately simulate the real atmosphere under some atmospheric conditions (e.g. Berg et al., 2015; Santanello et al., 2015). We assume no systematic bias perturbs wide scale geographic or seasonal patterns and this introduces uncertainty. Specific humidity is converted to RH using Equation 4-1 below adapted from Bolton (1980), where \( q \) is specific humidity and dimensionless (e.g. kg/kg), \( p \) is pressure in mb, and \( T \) is temperature in °C. The meteorology data is then averaged over the same time period as IMPROVE speciated measurements.

\[
RH = \frac{q p}{6.112 \left(0.378 q + 0.622\right) \exp\left(\frac{17.67 T}{T + 243.5}\right)}
\]

(4-1)

We also estimate light extinction from IMPROVE particle monitoring data using the original IMPROVE algorithm (Eq. 1 in Pitchford et al., 2007). For this calculation, we use NARR RH and a suggested cutoff of 98% RH (Malm et al., 1994). Seasonality in the extinction coefficient is calculated using the same definition for seasonal differences described earlier in this section.
4.4 Results and Discussion

The spatial patterns in aerosol water seasonality are consistent with AOT seasonality measured by MODIS and MISR from 2001-2007 as presented by Goldstein et al. (2009) (Figures 4-1 and 4-2). This pattern is observed in aerosol water estimated at the surface (Figures 4-1a,b), and averaged from the surface through the PBL (Figures 4-1c,d) for 2000-2007 (Figures 4-1a,c) and for all years 1988-2013 (Figures 4-1b,d). Aerosol water seasonal differences are highest in the eastern U.S. and centered around the Ohio Valley and Southeast, and lowest in the western U.S., with notable exceptions near Los Angeles and San Francisco. Seasonal differences in AOT spatial data from MODIS and MISR (Goldstein et al., 2009) are also high in the eastern U.S. and low in the western U.S. and qualitatively similar to geospatial patterns of aerosol water in Figure 4-1. Mean aerosol water concentrations during the observed time period from highest to lowest are 8.9, 7.7, 5.1, 3.8, 1.3, and 0.9 µg m\(^{-3}\) for the Southeast, Northeast, Midwest, Central, West Pacific, and Mountain West regions respectively (Figure 4-2). The highest concentrations of aerosol water are therefore found in the eastern U.S. in the Northeast and Southeast, while the lowest concentrations are found in the western U.S. in the West Pacific and Mountain West. The geospatial patterns of aerosol water concentrations are consistent with previous modeling work by Carlton and Turpin (2013). The time series also suggest seasonal shifts in aerosol water concentrations that are most prominent in the eastern U.S. and that resemble AOT seasonality in the Southeast U.S. represented in Figure 1d from Goldstein et al. (2009).

Geospatial patterns and the seasonality of surface organic aerosol mass concentrations are *not* qualitatively similar to MODIS and MISR AOT seasonality from
Seasonality in surface OM mass concentrations is observed across the U.S. (Figure 4-4). Mean OM concentrations are 3.2, 2.2, 2.1, 2.0, 1.9, and 1.5 \( \mu g \text{ m}^{-3} \) for the Southeast, Northeast, West Pacific, Central, Midwest, and Mountain West regions respectively (Figure 4-4). While OM concentrations are highest in the Southeast on average, the highest seasonal differences (summer – winter) in OM are found in the western U.S., and not in the eastern U.S. (Figure 4-3). There are also no stark contrasts between east and west OM mass concentrations and seasonality. This is in sharp contrast to the spatial and regionally-specific seasonality in aerosol water (Figures 4-1 and 4-2) and AOT.

Seasonality in aerosol water in the Southeast is modulated by aerosol chemistry and not \( RH \). There is strong seasonality in boundary layer \( RH \) in the western U.S. but not the eastern U.S. (Figure 4-5). Mean mass concentrations of inorganic species that are known to impact particle hygroscopicity (e.g., sulfate) show an east to west gradient, and are higher in the eastern U.S. (Figure 4-6d,e). Mean total inorganic compound concentrations from highest to lowest are 3.9, 2.6, 2.6, 2.6, 1.2, and 0.8, \( \mu g \text{ m}^{-3} \) for the Southeast, Northeast, Central, Midwest, West Pacific, and Mountain West regions respectively (Figure 4-6). Additionally, there is a stronger seasonal dependence in aerosol inorganic mass concentrations in the eastern vs. western U.S. (Figure 4-6). Total inorganic compound concentrations in the Northeast and Southeast regions of the U.S. have been decreasing since 2005 (Figure 4-6d,e), (Blanchard et al., 2013; Nguyen et al., 2015), as have seasonal differences in aerosol water (Figure 4-2d,e), in particular since 2005 (Figure 4-2d,e). As aerosol water mass concentrations decrease, the seasonal contrast has also decreased and was maximum in the eastern U.S. in this time series for
the period of 2000-2007. Similar trends are found for regional seasonality in the extinction coefficient, in which extinction seasonal differences are highest in the Southeast and lowest in the western U.S. (Figure 4-7). The geospatial and seasonal trends in aerosol inorganic mass concentrations and extinction are similar to geospatial trends in aerosol water concentrations and AOT. This is in sharp contrast to spatial patterns of \( RH \) seasonality.

It has been suggested that poor summertime visibility in the eastern U.S. is primarily due to high concentrations of sulfate aerosols exposed to high \( RH \) (Malm et al., 1994; Park et al., 2004; Pitchford et al., 2007). Decreases in sulfate in the Southeast have been linked with recently noted improvements in visibility and changes in surface radiation measurements in the Southeast U.S. (Attwood et al., 2014), location of the warming hole that is recently noted to be disappearing due to rising temperatures from 2000-2011 (Yu et al., 2014; Meehl et al., 2015). Temporal trends of aerosol water seasonality in the southeast U.S. are not inconsistent with the presence and decreasing magnitude of the warming hole.

### 4.5 Conclusions

Semi-observational estimates of aerosol water across the continental U.S. from 2000 to 2007 reveal qualitative similarities between aerosol water mass concentration and AOT. First, we find that geospatial patterns and regional seasonality of aerosol water mass at the surface and through the PBL over the continental U.S. is qualitatively similar to MODIS and MISR AOT from 2000-2007. On the contrary, geospatial patterns in the seasonality of organic aerosol mass at the surface in the continental U.S. are qualitatively dissimilar to MODIS and MISR AOT from 2000-2007. Additional evaluation of
precursors to aerosol water demonstrates that seasonality in aerosol water in the Southeast is modulated by aerosol chemistry and not RH.

4.6 References


Figure 4-1. Seasonal differences in aerosol water estimated using chemically characterized particle data from IMPROVE and temperature and RH data from the NARR at (a) surface for 2000-2007, (b) surface for 1988-2013, (c) surface to PBL for 2000-2007, and (d) surface to PBL for 1988-2013.
Figure 4-2. Time series of quartiles (blue shade) and medians (black) for aerosol water estimated in the a) West Pacific, b) Mountain West, c) Central, d) Northeast, e) Southeast, and f) Midwest using surface temperature and RH reanalysis data for 1988-2013. Dashed lines in panel e) indicate the time period analyzed in Goldstein et al. (2009).
Figure 4-3. Spatial map of seasonal differences in surface organic matter for (a) 2000-2007 and (b) 1988–2013.
Figure 4-4. Time series of quartiles (grey shading) and medians (black line) for surface measurements of OM for 1988–2013.
Figure 4-5. Time series of quartiles (grey shading) and medians (black line) for surface RH for 1988–2013.
Figure 4-6. Time series of quartiles (grey shading) and medians (black line) for surface measurements of inorganics (defined as the sum of sulfate and nitrate mass concentrations) for 1988–2013.
Figure 4-7. Time series of quartiles (grey shading) and medians (black line) for estimated extinction coefficients at IMPROVE sites for 1988–2013.
CHAPTER 5. SUMMARY, IMPLICATIONS, AND FUTURE DIRECTIONS

5.1 Summary and Implications

5.1.1 Trends in particle-phase liquid water during the Southern Oxidant and Aerosol Study

Particle-phase liquid water contributes to total aerosol mass concentrations. We presented measurements of dry aerosol volume, particle phase liquid water, and dry aerosol volume changes that occur during particle drying. The measurements were obtained during the SOAS campaign at Talladega National Forest in Brent, AL from 3 June to 15 July 2013 to collect six weeks of continuous in situ measurements of ambient aerosols. The particle-sizing instrument used in this study measures size distributions of ambient atmospheric aerosols in three states: unperturbed, dried, and dried then re-humidified. Unperturbed measurements characterize the aerosol distribution at ambient conditions. For dry spectra, the sample is routed through a cold trap (ΔT = -30K) upstream of the DMA inlet. The total volume of water and semi-volatile compounds (SV) lost during drying is quantified by differencing dry and unperturbed volumes from the integrated size spectra, while SV volumes are quantified by re-humidifying the sample and referencing to the unperturbed state. Results indicate that aqueous phase liquid water is an important contributor to ambient aerosol volume in the southeast U.S. during early morning times when the relative humidity differential is largest. Water is always present, comprising at minimum 7% of aerosol volume, and water mass concentrations ranged from 1 to 5 µg m\(^{-3}\) overall and exceeded 15 µg m\(^{-3}\) during non-raining days. Liquid water was the dominant aerosol constituent from ~7 a.m. to 9 a.m. LT. The diel water-content cycle was dominated by changes in \(RH\), which fluctuated primarily due to changes in
temperature. A strong diel cycle in aerosol hygroscopicity is consistent with observed systematic changes in aerosol composition and a water-content dependency of $\kappa$. Aerosol hygroscopicity increased with particle size. The method of particle drying used in this study did not lead to removal of significant material that was associated with the aqueous phase. Measurements were conducted June 1 – July 15, 2013, during the Southern Oxidant and Aerosol Study (SOAS) in the southeast U.S., a biogenically dominated and photochemically active environment impacted by anthropogenic pollution and known to contain high concentrations of organic aerosol mass. These results will be useful in future improvements of chemical mechanisms in atmospheric photochemical models, enabling more accurate model predictions for climate and air quality.

5.1.2 Decreasing aerosol water is consistent with organic carbon trends in the southeast United States

Using speciated ion and meteorological data from the Southeastern Aerosol Research and Characterization network, we employ the thermodynamic model ISORROPIAv2.1 to estimate water mass concentrations and evaluate trends from 2001 to 2012 in urban and rural locations. The purpose of this study is to better understand the historical trends of aerosol liquid water in the southeast U.S. in the context of improved air quality and recently noted reductions in particulate organic carbon (OC). Aerosol water mass concentrations decrease by $\sim$79% from 2001 to 2012 in the region. The decrease is driven largely by trends at the rural sites, which exhibit an 82% decrease in aerosol liquid water compared to a 51% decrease over the decade at urban monitoring locations. Fractional contribution of water to particulate matter (PM) also decreases during the same time period, and this is consistent with recently noted improvements in
visibility. Semi-empirical observational estimates of aerosol liquid water mass concentrations exhibit interannual variability consistent with local meteorology, but decadal trends are influenced by the decreasing ion concentrations of $\text{SO}_4^{2-}$, $\text{NO}_3^-$, and $\text{NH}_4^+$ that affect particle hygroscopicity. Furthermore, monthly analysis of aerosol liquid water data reveals seasonal differences that drive the yearly trend. The most noticeable decrease in water concentrations occurs May to October with less interannual variability from January to April and November to December. In the colder months, mass concentrations of $\text{SO}_4^{2-}$ and $\text{NH}_4^+$ are low due to lower emissions and less photochemistry, which control the relatively low wintertime aerosol water concentrations. Robust and statistically discernible particle OC reductions are consistent with aerosol liquid water reductions while acidity effects remain unclear. We rule out causes of OC reductions that pertain to decreases in anthropogenic volatile organic compounds (VOCs) mass concentrations, decreases in primary non-fossil OC mass concentrations, decreases in temperature, and increases in pH. These findings are consistent with the hypotheses that aerosol liquid water facilitates formation of biogenic secondary organic aerosol (SOA) and that biogenically derived SOA is modulated in the presence of anthropogenic perturbations.

5.1.3 **Reconciling satellite aerosol optical thickness and surface fine particle mass through aerosol liquid water**

Summertime aerosol optical thickness (AOT) in the Southeast U.S. is sharply enhanced compared to wintertime AOT. This seasonal pattern is unique to the Southeast and is of particular interest because temperatures have not warmed over the past 100 years, contrasting with trends in other U.S. regions. In this work, we attempt to reconcile
the spatial and temporal distribution of AOT over the U.S. with surface mass measurements by examining trends in aerosol water, an aerosol constituent that scatters radiation and is removed from aerosols in mass measurements at routine surface monitoring sites. We employ the thermodynamic model ISORROPIAv2.1 to estimate aerosol water mass concentrations at IMRPOVE sites using measured ion mass concentrations and NARR meteorological data. Our findings suggest aerosol water provides a plausible explanation for the geographical and seasonal patterns in AOT and can reconcile previously noted discrepancies with surface mass measurements. Semi-observational estimates of aerosol water across the continental U.S. reveal qualitative similarities between aerosol water mass concentration and aerosol optical thickness. First, we find that geospatial patterns and regional seasonality of aerosol water mass through the boundary layer over the continental U.S. is qualitatively similar to MODIS and MISR AOT from 2000-2007. On the contrary, geospatial patterns in the seasonality of organic aerosol mass at the surface in the continental U.S. are qualitatively dissimilar to MODIS and MISR AOT from 2000-2007. Additional evaluation of precursors to aerosol water demonstrate that seasonality in aerosol water in the Southeast is modulated by aerosol chemistry and not RH. Further analysis beyond the scope of this study will be needed to determine the presence or disappearance of the Southeast warming hole and whether aerosol water provides a plausible explanation.

5.2 Future Directions and Implications

5.2.1 Hygroscopicity Intercomparison

Particle phase liquid water and its influence on chemical composition is associated with many health and welfare effects: acid deposition (Calvert et al., 1985),
impaired visibility through light scattering (Malm et al., 1994), and climate through
effects on cloud condensation and ice nuclei (Cruz and Pandis, 1997; Pöschl, 2005).
Quantitative characterization of ambient aerosol liquid water content is essential to
adequately understand the fate and transport of chemicals in the Earth’s atmosphere, and
to develop effective strategies that mitigate aerosol-related problems. Comparisons of
different methods that measure aerosol liquid water and aerosol hygroscopicity are
necessary to ensure accurate data and improve measurement techniques.

In our preliminary analysis, we compare more than 5000 data points from the
field of aerosol hygroscopicity using five different instruments and different techniques.
The measurements were conducted 1 June to 15 July 2013 during SOAS in the
southeastern U.S. Two different hygroscopicity techniques – optical and particle sizing
methods – are compared for $\kappa$ at RH < 90%. The optical methods include one
nephelometer (optical scattering at 530 nm) from Georgia Institute of Technology (GA
Tech), one nephelometer from Atmospheric Research & Analysis (ARA), and an optical
detector (optical extinction, 405 nm) from the Cooperative Institute for Research in
Environmental Sciences (CIRES) and the National Oceanic and Atmospheric
Administration (NOAA). The particle sizing methods include our differential mobility
analysis (see Chapter 2) and the Dry-Ambient Aerosol Size Spectrometer (DAASS) from
École polytechnique fédérale de Lausanne (EPFL).

The volume of water associated with the aerosol is described using the
hygroscopicity model of Petters and Kreidenweis (2007). To calculate the hygroscopicity
parameter $\kappa$ for the optical methods, we compute the equation depicted below,

$$\kappa = \frac{f(RH) - 1}{\frac{RH_{amb}}{1-RH_{amb}} - f(RH)\left(\frac{RH_{dry}}{1-RH_{dry}}\right)}$$

(5-1)
where \( RH_{\text{amb}} \) and \( RH_{\text{dry}} \) are the RH of the wet and dry scans and

\[
f(RH) = \frac{B_{\text{scat, wet}}}{B_{\text{scat, dry}}}. \tag{5-2}
\]

\( B_{\text{scat, wet}} \) is the scattering coefficient of the wet scan while \( B_{\text{scat, dry}} \) is the scattering coefficient of the dry scan. To calculate \( \kappa \) for the particle sizing methods, we compute the equation,

\[
\kappa = \frac{V_{\text{amb}}}{V_{\text{dry}}} (RH^{-1} - 1) \tag{5-3}
\]

where \( V_{\text{amb}} \) and \( V_{\text{dry}} \) are the integrated volume of ambient (i.e., unperturbed) and dry aerosol scans.

Additionally, for a particle composed of multiple components the particle’s \( \kappa \) can be computed from the volume-weighted mixture of its dry components, \( \kappa = \sum \epsilon_i \kappa_i \), where \( \epsilon_i \) and \( \kappa_i \) are the volume fraction and hygroscopicity parameter of the \( i^{\text{th}} \) component comprising the particle. Assuming an approximate effective two component mixture of organic and inorganic species, and assigning \( \kappa_{\text{org}} \sim 0.1 \) (Prenni et al., 2007; Engelhart et al., 2008; Wex et al., 2009; Engelhart et al., 2011; Pierce et al., 2012) and \( \kappa_{\text{inorg}} \sim 0.6 \) (Petters and Kreidenweis, 2007) one can estimate the relative organic volume fractions (\( \epsilon_{\text{org}} \)) in the mixture.

The hygroscopicities calculated during the SOAS study period indicate aerosol components containing varying mixtures of inorganic and organic compounds. On average for all instrumentation methods, diel fluctuations in overall \( \kappa \) range from 0.13 to 0.29, which corresponds to \( \epsilon_{\text{org}} = 0.94 \) and \( \epsilon_{\text{org}} = 0.62 \), respectively (Figure 5-1). The particle sizing methods measurements calculate higher \( \kappa \) than the optical methods and the DAASS has the highest measurements of all methods, especially after 1 July. The diurnal profile of median \( \kappa \) of all instruments for SOAS reveals similar trends in hygroscopicity.
amongst the instruments but different magnitudes of $\kappa$ (Figure 5-2). For 7 June to 13 June, a period when all instruments were operating, the diurnal trend in $\kappa$ is less prominent for the optical techniques than for the particle sizing techniques (Figure 5-3), though we note that the NOAA/CiRES instrument does not humidify to ambient $RH$.

Diel fluctuations in $\kappa$ for this time period range from $\sim 0.2$ to $0.5$, which corresponds to $\epsilon_{org} = 0.8$ and $\epsilon_{org} = 0.2$, respectively. $\kappa$ exhibits a diel cycle in both analyses, suggesting that aerosol composition changes on time scales of a few hours. These patterns demonstrate that $\kappa$ may vary with water activity and depends on particle chemical composition.

Further research will be needed to determine reconciliation procedures between different instrument methods and improve aerosol water and hygroscopicity measurement accuracy. As particle-sizing methods measure higher $\kappa$ than the optical methods, a factor can be developed to resolve the two methods. This analysis will help further the representation of aerosol water content and aqueous phase mediated partitioning of atmospheric photochemical models.

### 5.2.2 Model Validation

Model validation of aerosol water predictions is needed to effectively predict future air quality and climate change. Semi-observational estimates of water presented in Chapters 1, 3, and 4 of this work have been calculated using ISORROPIAv.2.1, which is used as a module in the Community Multi-scale Air Quality (CMAQ) model by government agencies such as the Environmental Protection Agency (EPA) to predict air quality. However, no work has yet been published to validate model predictions of aerosol water with measurements. A better understanding of the trends in aerosol water and how it is represented in the model is therefore needed to improve development of
models such as ISORROPIA, Extended AIM Aerosol Thermodynamics Model (E-AIM, Clegg et al., 2001), and Aerosol Inorganic–Organic Mixtures Functional groups Activity Coefficients (AIOMFAC, Zuend et al., 2008).

Initial comparisons of ISORROPIA with SOAS measurements from Chapter 2 can provide the necessary basis for this future work. SOAS hygroscopicity data $\kappa$ from Nguyen et al. (2014) and SEARCH $RH$ is used to calculate aerosol water using Eq. 5-3. ISORROPIA is then run using SEARCH meteorology data to predict aerosol water for time periods corresponding to time periods from Nguyen et al. (2014) with available data. For this work, we again assume a $\text{NH}_4^+\text{-SO}_4^{2-}\text{-NO}_3^-$ metastable system and we run the model in reverse mode.

Diurnal trends in aerosol water are lower than aerosol water calculated from measured hygroscopicity data (Figure 5-4). ISORROPIA does not consider water uptake by organic compounds, and laboratory studies demonstrate that organic constituents can modulate particle hygroscopicity depending on the chemical composition (Saxena et al., 1995; Cruz and Pandis, 2000; Dick et al., 2000; Suda et al., 2012; Sareen et al., 2013). However, the effects of organic compounds on aerosol water are complex (Decesari et al., 2000; Shimmo et al., 2004) and data on all aerosol chemical constituents during SOAS is not readily available. Since water uptake is typically assumed to be additive, this may explain the underestimates by ISORROPIA. The difference between measured and predicted aerosol water is also highest in the early morning hours between 7 to 9 a.m. where aerosol water peaks in mass concentrations. The highest difference is about 25 $\mu$g m$^{-3}$ at 7 a.m. Further research is needed to determine if organic compound concentrations peak during this same morning period. It is also possible that the underestimates by
ISORROPIA may be additionally due to not considering water uptake by other speciated ions such as sodium, chloride, magnesium, calcium, and potassium.

**5.2.3 Global Aerosol Water and Temperature Trends**

Global climate change threatens the planet’s ecosystems. Successful strategies to alleviate the most severe environmental outcomes are mired in uncertainty, in part, because the atmospheric chemistry of the Earth system that regulates and responds to pollution is complex and interconnected. The Southeast has not warmed like other regions of the U.S. in response to global climate change (Goldstein et al., 2009; Portmann et al., 2009). In Chapter 4, we explored the possible link between aerosol water and the seasonal differences in aerosol optical thickness in the Southeast. Further analysis of the relationship between aerosol water and the temperature trends in the Southeast and the rest of the world will be useful to help better understand past, current, and future regional and global climate change.

To better understand global temperature trends in the past century, we construct an analysis of global surface temperature trends from 1900 to 2015 for the month of September using data from the Global Historical Climatology Network (GHCN) v3 (http://data.giss.nasa.gov/gistemp/maps/). Land data are selected from the National Aeronautics and Space Administration's Goddard Institute for Space Studies (NASA GISS; Hansen et al., 2010), ocean data from the Extended Reconstructed Sea Surface Temperature (ERSST) v4, and a base period of 1951 to 1980.

Results indicate a cooling trend in the Southeast U.S. of -0.2 °C in the time period of analysis (Figure 5-5). On the contrary, most parts of the world have experienced a warming trend, in particular Brazil and central Australia. Regional areas of cooling also
include western Africa, eastern China, and parts of Russia. Further analysis into aerosol water concentrations in these regions and their historical trends could lead to key understanding behind their impacts on regional and climate.

5.2.4 Epidemiological Study

The National Morbidity, Mortality, and Air Pollution Study (NMMAPS) is one of the largest multi-city daily time series studies that addressed mortality related to airborne particles less than 10 μm in aerodynamic diameter (PM$_{10}$) in the United States (Samet et al., 2000a). NMMAPS was designed to address concerns about city selection bias, publication bias, and influence of co-pollutants. The study included two analyses: 20 and 90 large cities (Dominici et al., 2003b; Dominici et al., 2003a; Peng et al., 2005) in its mortality analysis. Although the PM-mortality effect estimates were somewhat sensitive to various modeling and city selection choices, there was “consistent evidence that the levels of fine particulate matter in the air are associated with the risk of death from all causes and from cardiovascular and respiratory illnesses” (Samet et al., 2000b). Overall, the study found a 0.21% increase in total mortality per 10 μg m$^{-3}$ elevation in PM$_{10}$ concentrations (Samet et al., 2000a).

Future studies building upon the work completed by NMMAPS can interpret the relative risks of PM$_{10}$ across 20 of the most populated areas in the U.S. (Table A2, Samet et al., 2000a) with aerosol water. Mass concentrations of particle-phase liquid water can be calculated by information on speciated PM components in these cities during the study period of 1987-1994. It is hypothesized that chemistry in atmospheric waters (i.e., in clouds and particle phase water) is a substantial source of SOA (Blando and Turpin, 2000; Gelencser and Varga, 2005), and thus aerosol water may be linked to the PM$_{10}$
mortality risks that NMMAPS analyzed. A better understanding of aerosol water mortality will help improve public health policies pertaining to precursor aerosol water pollutant emissions in order to highlight all of the major possibilities for control.

5.3 References


Figure 5-1. Intercomparison results of $\kappa$ for amongst optical methods by GA Tech (purple), ARA (red), and NOAA/CIFRES (orange) and particle sizing methods by Rutgers (black) and EFPL’s DAASS (blue). The plot also shows precipitation in mm (grey bars, bottom), temperature in °C (brown line, top), and $RH$ (green line, second from top). The yellow bars on the plot indicate intensive periods during the SOAS field campaign.
Figure 5-2. Diurnal profile of $\kappa$ during SOAS

June 1 - July 15, 2013

Hours past midnight (local time)

$K$

RUTGERS DAASS ARA GA TECH NOAA/CIRES
Figure 5-3. Diurnal profile of $\kappa$, June 7-13, 2013
Figure 5-4. Boxplots of diurnal trends in aerosol water for ISORROPIA semi-observational estimates (black) and SOAS hygroscopicity data from Nguyen et al., 2014 adjusted to SEARCH $RH$ (blue).
Figure 5-5. Global changes in surface temperature using NASA GISS data for the month of September from 1900-2015.
**APPENDIX A: SUPPORTING INFORMATION FOR CHAPTER 1**

**Table A-1.** Average values of parameters at locations used to estimate aerosol water for AMS studies

<table>
<thead>
<tr>
<th>Location</th>
<th>Units</th>
<th>Study Period</th>
<th>Temp</th>
<th>RH</th>
<th>SO$_4^{2-}$</th>
<th>NO$_3^-$</th>
<th>NH$_4^+$</th>
<th>Organics</th>
<th>Inorganic water</th>
<th>Organic water</th>
<th>Total water</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>°C</td>
<td>%</td>
<td>µg m$^{-3}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>40</td>
<td>Jul 2006</td>
<td>23.2</td>
<td>85.4</td>
<td>19</td>
<td>15</td>
<td>12</td>
<td>25</td>
<td>78.1</td>
<td>8.4</td>
<td>86.4</td>
</tr>
<tr>
<td>Chebogue</td>
<td>43.8</td>
<td>Jul-Aug 2004</td>
<td>31.7</td>
<td>33</td>
<td>0.89</td>
<td>0.08</td>
<td>0.34</td>
<td>1.6</td>
<td>0.3</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Duke Forest</td>
<td>36</td>
<td>Sep 2004</td>
<td>29.7</td>
<td>53</td>
<td>1.1</td>
<td>0.1</td>
<td>0.32</td>
<td>1.3</td>
<td>0.7</td>
<td>0.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>56</td>
<td>Nov 2000</td>
<td>5</td>
<td>78.1</td>
<td>0.52</td>
<td>0.28</td>
<td>0.3</td>
<td>1.9</td>
<td>1.2</td>
<td>0.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Hohenpeissenberg</td>
<td>50.2</td>
<td>May 2002</td>
<td>6.2</td>
<td>89</td>
<td>3.9</td>
<td>1.4</td>
<td>1.4</td>
<td>9.5</td>
<td>12.5</td>
<td>6</td>
<td>18.6</td>
</tr>
<tr>
<td>Houston</td>
<td>29.8</td>
<td>Aug-Sep 2000</td>
<td>36.9</td>
<td>64.7</td>
<td>4.9</td>
<td>0.38</td>
<td>1.5</td>
<td>6</td>
<td>4.6</td>
<td>0.6</td>
<td>5.2</td>
</tr>
<tr>
<td>Hyttiala</td>
<td>61.9</td>
<td>Mar-Apr 2003</td>
<td>-3.6</td>
<td>97</td>
<td>0.31</td>
<td>0.24</td>
<td>0.28</td>
<td>1.1</td>
<td>7.3</td>
<td>3.3</td>
<td>10.6</td>
</tr>
<tr>
<td>Jungfraujoch</td>
<td>46.3</td>
<td>Jun-Jul 2002</td>
<td>10.7</td>
<td>93</td>
<td>0.51</td>
<td>0.2</td>
<td>0.33</td>
<td>1.2</td>
<td>3.6</td>
<td>1.5</td>
<td>5.1</td>
</tr>
<tr>
<td>Mace Head</td>
<td>53.3</td>
<td>Jul-Sep 2002</td>
<td>9.1</td>
<td>89</td>
<td>0.75</td>
<td>0.05</td>
<td>0.23</td>
<td>0.5</td>
<td>2</td>
<td>0.4</td>
<td>2.4</td>
</tr>
<tr>
<td>Manaus</td>
<td>-2.58</td>
<td>Feb-Mar 2008</td>
<td>27.1</td>
<td>75.6</td>
<td>0.23</td>
<td>0.011</td>
<td>0.032</td>
<td>0.77</td>
<td>0.4</td>
<td>0.2</td>
<td>0.6</td>
</tr>
<tr>
<td>Manchester</td>
<td>53.5</td>
<td>Jun 2001</td>
<td>6.4</td>
<td>90.4</td>
<td>3.4</td>
<td>3.7</td>
<td>2.3</td>
<td>4.9</td>
<td>25.7</td>
<td>2.6</td>
<td>28.4</td>
</tr>
<tr>
<td>Mexico City</td>
<td>19.4</td>
<td>Apr-May 2003</td>
<td>35.6</td>
<td>32</td>
<td>3.1</td>
<td>3.7</td>
<td>2.1</td>
<td>22</td>
<td>1.5</td>
<td>0.6</td>
<td>2.1</td>
</tr>
<tr>
<td>NYC</td>
<td>40.8</td>
<td>Jul-Aug 2001</td>
<td>34.9</td>
<td>20.6</td>
<td>3.9</td>
<td>0.68</td>
<td>1.7</td>
<td>5.9</td>
<td>0.8</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>40.4</td>
<td>Sep 2002</td>
<td>22.6</td>
<td>41.7</td>
<td>7</td>
<td>0.87</td>
<td>2.4</td>
<td>4.4</td>
<td>3.6</td>
<td>0.2</td>
<td>3.8</td>
</tr>
<tr>
<td>Riverside</td>
<td>33.9</td>
<td>Jul-Aug 2005</td>
<td>35</td>
<td>16.9</td>
<td>3.3</td>
<td>4.3</td>
<td>2.4</td>
<td>8.9</td>
<td>0.6</td>
<td>0.1</td>
<td>0.7</td>
</tr>
<tr>
<td>Santiago</td>
<td>-33.45</td>
<td>Aug-Nov 2011</td>
<td>17.9</td>
<td>78</td>
<td>2.19</td>
<td>4.11</td>
<td>2.74</td>
<td>17.3</td>
<td>9.2</td>
<td>3.5</td>
<td>12.7</td>
</tr>
<tr>
<td>Storm Peak</td>
<td>40.5</td>
<td>Apr 2004</td>
<td>20.2</td>
<td>60</td>
<td>0.59</td>
<td>0.22</td>
<td>0.4</td>
<td>0.9</td>
<td>0.6</td>
<td>0.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Tokyo</td>
<td>35.7</td>
<td>Jul-Aug 2003</td>
<td>23.6</td>
<td>77.7</td>
<td>3.2</td>
<td>1.5</td>
<td>1.8</td>
<td>6.5</td>
<td>7.2</td>
<td>1.3</td>
<td>8.5</td>
</tr>
<tr>
<td>Vancouver</td>
<td>49.3</td>
<td>Aug 2001</td>
<td>24.7</td>
<td>36.3</td>
<td>0.86</td>
<td>0.55</td>
<td>0.68</td>
<td>4.9</td>
<td>0.4</td>
<td>0.2</td>
<td>0.6</td>
</tr>
<tr>
<td>Welgegund</td>
<td>-26.57</td>
<td>Sep 2010 – Aug 2011</td>
<td>20.1</td>
<td>76.5</td>
<td>2.32</td>
<td>0.44</td>
<td>0.95</td>
<td>3.49</td>
<td>4.2</td>
<td>1.1</td>
<td>5.2</td>
</tr>
<tr>
<td>Weybourne</td>
<td>53</td>
<td>Apr-May 2004</td>
<td>4</td>
<td>85.7</td>
<td>2</td>
<td>2.1</td>
<td>1.3</td>
<td>2.1</td>
<td>10.2</td>
<td>1</td>
<td>11.2</td>
</tr>
</tbody>
</table>
APPENDIX B: SUPPORTING INFORMATION FOR CHAPTER 2

Appendix B1. Meteorological Differences in Diel Cycle of Aerosol Water

Conclusions based on campaign-based average diel cycles may be influenced by different regimes. To explore the possible influence of averaging we repeat the analysis performed in Figure 2-7 for a single day in Figure B-1. Comparing Figures 2-2 to 2-7 and B-1 suggests that the reported diel trend accurately captures the behavior for a single day. An unusual cold and rainy period occurred between 4 and 8 July 2013. During that time, aerosol volume was low and diel temperature and RH fluctuations were minimal. Figure B-2 contrasts the diel cycle for particle phase liquid water for a stagnation event with a pronounced diel cycle and the cold period. The data show that no diel cycle was observed during this atypical period, suggesting that conclusions reached about typical regional southeast U.S. aerosol must be interpreted in the context of the meteorological setting.

Appendix B2. SMPS Volume Validation

The SMPS was co-located with PM 2.5 aerosol loadings determined by a tapered element oscillating microbalance (TEOM) (Patashnick and Rupprecht, 1991). The TEOM instrument was operated by Atmospheric Administration, Atmospheric Research & Analysis, Inc. as a part of the SEARCH network. SEARCH operations, sampling, and measurement procedures can be found in Hansen et al. (2003), Edgerton et al. (2005) and Edgerton et al. (2006). The TEOM (R&P Model 1400 a/b) uses oscillating microbalance to provide continuous measurements of PM$_{2.5}$ dry mass concentration at 5-minute intervals. Briefly, particles with aerodynamic particle diameter $D > 2.5$ µm are removed using a cyclone impactor and particles are dried to RH < 20% using a Nafion drier. To
compare the TEOM mass with the SMPS volume concentrations SMPS dry volume
timestamps were consequently rounded to the nearest fifth minute.

Figure B-3 shows the SMPS volume concentration and TEOM mass
concentration time series. Note that a mass concentration of 1 $\mu g\ m^{-3}$ equals 1 $\mu m^3\ cm^{-3}$
assuming spherical particles with unit density of 1 g cm$^{-3}$. The time-series shows that the
TEOM and SMPS fluctuate in tandem. The scatterplot of the two time-series is shown in
Figure B-4, which shows that the average offset between the two measurements is $\sim 1.5$,
corresponding to an effective aerosol density of 1.5 g cm$^{-3}$. Sulfate salts and organics
composed $\sim 50\%$ of particles during SOAS. The density of ammonium sulfate and
organics are $\rho = 1.77\ g\ cm^{-3}$ and $0.8 < \rho < 1.5$, respectively. Notably, the density of the
organic fraction is most uncertain since it depends on the number and quality of
functional groups associated with the carbon chain (e.g. Girolami, 1994). We point out
that this closure between TEOM mass and SMPS volume is approximate since the two
techniques have different size cutoffs (2.5 $\mu m$ vs. 1 $\mu m$), the SMPS volume may be
biased due to particle shape effects, and residual water present at RH < 20% and the
presence of volatile compounds may differentially affect the two measurements (Moya et
al., 2011). Nevertheless, the correlation between TEOM mass and SMPS volume,
combined with the observed effective density of 1.5 g cm$^{-3}$ is reasonable to suggest that
the SMPS volume measurements were not significantly biased.

**Appendix B3. Analysis Without Modal Fit**

The mean and quartiles of parameters without resorting to lognormal parameter
fitting are reported in Table B-1. Particles with $D > 700\ nm$ were excluded to screen for
arcing artifacts. Water volume, semi-volatile volume, \( k_{\text{vol}} \), and \( g_{f_{\text{vol}}} \) were calculated using the unfitted data and Eqs. 2-6, 2-7, 2-1, and 2-4 respectively.

**Appendix B4. Relative Humidity Dependence of Kappa**

Thermodynamic model calculations were performed to test the potential influence of relative humidity on the diel cycle of kappa. Model III of the online version of the Extended Aerosol Inorganic Model (http://www.aim.env.uea.ac.uk/aim/model3/mod3rhw.php) (Clegg et al., 1998; Clegg et al., 2001) was initialized with fixed amounts of \( \text{NH}_4^+ \), \( \text{H}^+ \), \( \text{SO}_4^{2-} \), \( \text{NO}_3^- \), and a single organic species defined in Table B-2. Molar amounts were estimated to approximately represent the PILS measured mass concentrations of sulfate, nitrate, ammonia, and WSOM. \( \text{H}^+ \) concentrations were obtained by forcing charge balance. Metastable solutions were expected to be present during the dry-humidified cycle. Therefore the formation of solids was suppressed in the model. Since speciation of the organic fraction was not available, we assumed that the average carbon chain length was \( C_{10} \) and that the functional group composition consists of one acid, three hydroxyls, and one carbonyl group. The molecular mass of the organic compound corresponds to 232 g mol\(^{-1}\). The assumed ratio of hydroxyl to carboxyl groups is 3:1 and similar to measurements of ambient monoterpene SOA (~2.7:1) (Russell et al., 2011). Activity coefficients were estimated using the semi-empirical UNIFAC group contribution method (Fredenslund et al., 1975) and tabulated interaction parameters of Hansen et al. (1991). The E-AIM model computes the water content \([\text{H}_2\text{O}]_{\text{aq}}\) of the solution and the molarity of dissolved species: \([\text{NH}_4^+]_{\text{aq}}\), \([\text{HSO}_4^-]_{\text{aq}}\), \([\text{NO}_3^-]_{\text{aq}}\), \([\text{OH}^-]_{\text{aq}}\), \([\text{NH}_3]_{\text{aq}}\), and \([\text{WSOM}]_{\text{aq}}\), for a specified water activity. The sum of these parameters corresponds to the number of moles
in solution. The osmotic coefficient is obtained from the well-known relationship

\[ a_w^{-1} = 1 + \Phi \frac{n_s}{n_w} \]  \hspace{1cm} (B-1)

where \( n_s \) is the number of moles of dissolved species in solution, \( n_w \) is the number of moles of water, \( a_w \) is the water activity, and \( \Phi \) is the apparent osmotic coefficient. Variations in the osmotic coefficient with relative humidity are responsible for the humidity dependence of \( \kappa \) (Petters and Kreidenweis, 2007). Note that the dissociation of molecules and potential removal of molecules from solution via solubility would be captured in this definition of \( \Phi \) since the number of moles of solution, \( n_s \), is directly obtained from the output of the thermodynamic model. Figure B-5 shows the variation of \( \Phi \) with \( a_w \) for the illustrative test solution. The model \( \Phi \) increases with decreasing \( a_w \). A change in \( a_w \) from 0.9 to 0.4 results in a ~60% increase in \( \Phi \) and a concomitant increase in \( \kappa \). Thus the increase in \( \kappa_{vol} \) during daytime (Figures 2-8 and 2-10) is partly driven by changes in aerosol hygroscopicity with RH. Precise deconvolution of the relative importance of changing RH and chemical composition in determining the aerosol hygroscopicity will require a better description of the organic fraction, a description of the aerosol mixing state, estimation of organic aerosol molecular weight and density, and inclusion of unresolved non-hygrosopic species (dust, black carbon) to relate \( \Phi \) to \( \kappa \).

**Appendix B5. References**


Table B-1. Mean and quartiles of number concentration $N_{\text{conc}}$, particle phase water volume $V_w$, semi-volatile volume $V_{sv}$, volume of solutes $V_d$, and $gf_{\text{vol}}$ for unfitted data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>Mean</th>
<th>25% Quartile</th>
<th>50% Quartile</th>
<th>75% Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{conc}}$</td>
<td>cm$^{-3}$</td>
<td>2732</td>
<td>1541</td>
<td>2185</td>
<td>3249</td>
</tr>
<tr>
<td>$V_w$</td>
<td>µm$^3$ cm$^{-3}$</td>
<td>3.91</td>
<td>1.14</td>
<td>2.7</td>
<td>4.97</td>
</tr>
<tr>
<td>$V_{sv}$</td>
<td>µm$^3$ cm$^{-3}$</td>
<td>0.0099</td>
<td>-0.85</td>
<td>0.19</td>
<td>1.31</td>
</tr>
<tr>
<td>$V_d$</td>
<td>µm$^3$ cm$^{-3}$</td>
<td>5.748</td>
<td>3.3</td>
<td>5.0</td>
<td>7.793</td>
</tr>
<tr>
<td>$k_{\text{vol}}$</td>
<td>dimensionless</td>
<td>0.31</td>
<td>0.12</td>
<td>0.24</td>
<td>0.41</td>
</tr>
<tr>
<td>$gf_{\text{vol}}$</td>
<td>dimensionless</td>
<td>1.72</td>
<td>1.31</td>
<td>1.57</td>
<td>1.95</td>
</tr>
</tbody>
</table>

Table B-2. Solute concentrations for E-AIM model initialization. The UNIFAC functional groups representation of the organic molecule is 1*COOH 3*OH 1*CH$_2$CO 3*CH$_2$ 3*CH 1*CH$_3$.

<table>
<thead>
<tr>
<th></th>
<th>Cations</th>
<th>Anions</th>
<th>Organic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NH$_4^+$</td>
<td>H$^+$</td>
<td>SO$_4^{2-}$</td>
</tr>
<tr>
<td>Molecular Mass (g mol$^{-1}$)</td>
<td>18.03851</td>
<td>1.008</td>
<td>96.07</td>
</tr>
<tr>
<td>Mass concentration (µg m$^{-3}$)</td>
<td>0.46</td>
<td>0.01</td>
<td>1.71</td>
</tr>
<tr>
<td>Molar concentration (nmol m$^{-3}$)</td>
<td>25.4</td>
<td>11.3</td>
<td>17.8</td>
</tr>
</tbody>
</table>
Figure B-1. Same as Figure 2-7 for 21 June 2013, a day with no rainfall. Grey dotted lines indicating median growth factors that were calculated using the same $RH$ and dry volume, and assumed $\kappa$ values ranging from 0 to 0.6. The black dotted line indicates the $RH$. 
Figure B-2. Comparison of particle-phase liquid water for a 5 day period with heavy rainfall (4–8 July, blue) versus a 5 day period with no rainfall (18–22 June, brown). The central lines indicate the median and the shaded regions indicate the interquartile range.
Figure B-3. Time series of TEOM mass (blue) and SMPS dry volume (black).
Figure B-4. Scatterplot of TEOM particulate matter mass ($\mu g$ m$^{-3}$) and SMPS dry volume ($\mu m^3$ cm$^{-3}$) with 1:1 and 1:2 line. The 1:1.5 line indicates the best-fit line calculated from the data.
Figure B-5: Variation of the E-AIM Model III derived osmotic coefficient with water activity.
APPENDIX C: SUPPORTING INFORMATION FOR CHAPTER 3

Table C-1. Annual means of data for (a) overall, (b) urban, and (c) rural sites.

<table>
<thead>
<tr>
<th>Year</th>
<th>Water µg m(^{-3})</th>
<th>RH dimensionless</th>
<th>Temperature °C</th>
<th>SO(_4) µg m(^{-3})</th>
<th>NO(_3) µg</th>
<th>NH(_4) µg</th>
<th>OC µg-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Overall sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>51.9</td>
<td>0.727</td>
<td>19.1</td>
<td>4.21</td>
<td>0.809</td>
<td>1.58</td>
<td>3.39</td>
</tr>
<tr>
<td>2002</td>
<td>28.2</td>
<td>0.751</td>
<td>18.6</td>
<td>3.78</td>
<td>0.663</td>
<td>1.24</td>
<td>3.02</td>
</tr>
<tr>
<td>2003</td>
<td>30.4</td>
<td>0.730</td>
<td>15.7</td>
<td>4.11</td>
<td>0.768</td>
<td>1.34</td>
<td>3.27</td>
</tr>
<tr>
<td>2004</td>
<td>42.3</td>
<td>0.737</td>
<td>16.9</td>
<td>4.56</td>
<td>0.691</td>
<td>1.45</td>
<td>3.15</td>
</tr>
<tr>
<td>2005</td>
<td>28.2</td>
<td>0.729</td>
<td>17.6</td>
<td>4.46</td>
<td>0.533</td>
<td>1.50</td>
<td>3.21</td>
</tr>
<tr>
<td>2006</td>
<td>26.4</td>
<td>0.682</td>
<td>17.8</td>
<td>3.96</td>
<td>0.626</td>
<td>1.49</td>
<td>3.28</td>
</tr>
<tr>
<td>2007</td>
<td>17.7</td>
<td>0.683</td>
<td>18.9</td>
<td>3.55</td>
<td>0.515</td>
<td>1.21</td>
<td>3.20</td>
</tr>
<tr>
<td>2008</td>
<td>20.5</td>
<td>0.696</td>
<td>17.8</td>
<td>3.29</td>
<td>0.478</td>
<td>1.13</td>
<td>2.54</td>
</tr>
<tr>
<td>2009</td>
<td>21.0</td>
<td>0.724</td>
<td>16.7</td>
<td>2.34</td>
<td>0.460</td>
<td>0.95</td>
<td>2.23</td>
</tr>
<tr>
<td>2010</td>
<td>12.7</td>
<td>0.695</td>
<td>17.2</td>
<td>2.36</td>
<td>0.508</td>
<td>0.97</td>
<td>2.82</td>
</tr>
<tr>
<td>2011</td>
<td>10.9</td>
<td>0.684</td>
<td>17.5</td>
<td>2.62</td>
<td>0.485</td>
<td>1.04</td>
<td>2.55</td>
</tr>
<tr>
<td>2012</td>
<td>10.7</td>
<td>0.711</td>
<td>18.9</td>
<td>1.97</td>
<td>0.312</td>
<td>0.77</td>
<td>2.33</td>
</tr>
<tr>
<td>(b) Urban sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>18.8</td>
<td>0.700</td>
<td>20.4</td>
<td>4.99</td>
<td>0.710</td>
<td>1.43</td>
<td>3.77</td>
</tr>
<tr>
<td>2002</td>
<td>27.0</td>
<td>0.756</td>
<td>20.3</td>
<td>4.08</td>
<td>0.796</td>
<td>1.29</td>
<td>3.56</td>
</tr>
<tr>
<td>2003</td>
<td>21.2</td>
<td>0.715</td>
<td>16.2</td>
<td>4.40</td>
<td>0.935</td>
<td>1.43</td>
<td>3.70</td>
</tr>
<tr>
<td>2004</td>
<td>32.3</td>
<td>0.719</td>
<td>16.1</td>
<td>4.69</td>
<td>0.854</td>
<td>1.53</td>
<td>3.55</td>
</tr>
<tr>
<td>2005</td>
<td>27.6</td>
<td>0.727</td>
<td>18.2</td>
<td>4.88</td>
<td>0.687</td>
<td>1.69</td>
<td>3.42</td>
</tr>
<tr>
<td>2006</td>
<td>24.7</td>
<td>0.657</td>
<td>17.5</td>
<td>4.33</td>
<td>0.878</td>
<td>1.79</td>
<td>3.57</td>
</tr>
<tr>
<td>2007</td>
<td>15.4</td>
<td>0.679</td>
<td>19.3</td>
<td>3.65</td>
<td>0.607</td>
<td>1.24</td>
<td>3.48</td>
</tr>
<tr>
<td>2008</td>
<td>20.7</td>
<td>0.688</td>
<td>18.3</td>
<td>3.45</td>
<td>0.578</td>
<td>1.21</td>
<td>2.80</td>
</tr>
<tr>
<td>2009</td>
<td>19.7</td>
<td>0.703</td>
<td>16.1</td>
<td>2.45</td>
<td>0.546</td>
<td>1.03</td>
<td>2.35</td>
</tr>
<tr>
<td>2010</td>
<td>11.8</td>
<td>0.683</td>
<td>17.4</td>
<td>2.52</td>
<td>0.617</td>
<td>1.05</td>
<td>3.04</td>
</tr>
<tr>
<td>2011</td>
<td>10.0</td>
<td>0.667</td>
<td>18.1</td>
<td>2.78</td>
<td>0.577</td>
<td>1.14</td>
<td>2.55</td>
</tr>
<tr>
<td>2012</td>
<td>9.25</td>
<td>0.713</td>
<td>19.9</td>
<td>2.03</td>
<td>0.336</td>
<td>0.80</td>
<td>2.52</td>
</tr>
<tr>
<td>(c) Rural sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>68.5</td>
<td>0.740</td>
<td>18.5</td>
<td>3.81</td>
<td>0.859</td>
<td>1.66</td>
<td>2.91</td>
</tr>
<tr>
<td>2002</td>
<td>29.1</td>
<td>0.747</td>
<td>17.3</td>
<td>3.56</td>
<td>0.563</td>
<td>1.21</td>
<td>2.40</td>
</tr>
<tr>
<td>2003</td>
<td>38.0</td>
<td>0.742</td>
<td>15.3</td>
<td>3.87</td>
<td>0.629</td>
<td>1.26</td>
<td>2.64</td>
</tr>
<tr>
<td>2004</td>
<td>52.6</td>
<td>0.755</td>
<td>17.7</td>
<td>4.43</td>
<td>0.523</td>
<td>1.38</td>
<td>2.52</td>
</tr>
<tr>
<td>2005</td>
<td>28.6</td>
<td>0.731</td>
<td>17.2</td>
<td>4.16</td>
<td>0.423</td>
<td>1.37</td>
<td>2.90</td>
</tr>
<tr>
<td>2006</td>
<td>27.5</td>
<td>0.699</td>
<td>18.1</td>
<td>3.72</td>
<td>0.459</td>
<td>1.29</td>
<td>2.73</td>
</tr>
<tr>
<td>2007</td>
<td>19.9</td>
<td>0.687</td>
<td>18.6</td>
<td>3.46</td>
<td>0.424</td>
<td>1.18</td>
<td>2.62</td>
</tr>
<tr>
<td>2008</td>
<td>20.3</td>
<td>0.704</td>
<td>17.3</td>
<td>3.12</td>
<td>0.380</td>
<td>1.06</td>
<td>2.00</td>
</tr>
<tr>
<td>2009</td>
<td>22.4</td>
<td>0.743</td>
<td>17.3</td>
<td>2.23</td>
<td>0.377</td>
<td>0.87</td>
<td>1.92</td>
</tr>
<tr>
<td>2010</td>
<td>13.4</td>
<td>0.704</td>
<td>17.0</td>
<td>2.24</td>
<td>0.425</td>
<td>0.91</td>
<td>2.27</td>
</tr>
<tr>
<td>2011</td>
<td>11.7</td>
<td>0.700</td>
<td>16.9</td>
<td>2.48</td>
<td>0.403</td>
<td>0.96</td>
<td>2.56</td>
</tr>
<tr>
<td>2012</td>
<td>12.6</td>
<td>0.707</td>
<td>17.6</td>
<td>1.97</td>
<td>0.312</td>
<td>0.77</td>
<td>2.16</td>
</tr>
</tbody>
</table>
Figure C-1. Locations of SEARCH monitoring sites.
Figure C-2. LOESS estimates of trends for water (black) and organic carbon (green) for rural (orange, left) and urban (grey, right) SEARCH sites.
Figure C-3. LOESS estimated monthly trends in SO\(_4\) (solid lines) and NH\(_4\) (dashed lines) in the SE U.S. from 2001-2012 for 80 \(\leq RH < 85\) (red), 85 \(\leq RH < 90\) (green), and 90 \(\leq RH < 95\) (blue).
Figure C-4. Ggplot2 smoothness estimation (solid line) and standard error (shaded area) for ambient isoprene from the PAMS sites extracted for the SEARCH states.
Figure C-5. Local regression (LOESS) trends of ozone concentrations for rural (orange) and urban (grey) sites for SEARCH states.
Figure C-6. Ggplot2 smoothness estimation (solid line) and standard error (shaded area) for index for acidity calculated using ion charge balance.
Figure C-7. Ggplot2 smoothness estimation (solid lines) and standard error (shaded regions) for OC to SO$_4$ (left panels: a and c) and OC to SO$_4$ + NO$_3$ (right panels: b and d). The top panels (a and b) have been screened for fires and the bottom panels (c and d) represent the raw data.
Figure C-8. Localized regression (LOESS) trends of OC (green) and SOC estimates (red) for all SEARCH sites from 2001-2012.
Figure C-9. Localized regression (LOESS) trends for water (dashed, blue), OC (red), and SOC estimates (green) by month.
Figure D-1. Map of IMPROVE sites divided up into six regions
APPENDIX E: R CODES FOR CHAPTER 1

## This recreates Zhang pie chart map
(http://www.asrc.cestm.albany.edu/qz/AMS_Global_Database/img_global.jpg), but with water data
## Use Zhang AMS data and RH and T data from June 21, 2014 at 7am.
##

```r
# library(rworldmap)
# dF <- getMap()@data
# mapPies( dF,nameX="LON", nameY="LAT",
nameZs=c('POP_EST','POP_EST','POP_EST','POP_EST') )
#
# # add.pie(z=c(east, west), x=lon, y=lat, radius=sqrt(tot),
# col=c(alpha("orange", 0.6), alpha("blue", 0.6)),
# labels="",À,À)
#
# # map("worldHires","Canada,À, xlim=c(-141,-53), ylim=c(40,85),
# col="gray90,À, fill=TRUE)
#
# urban areas (blue)
# <100 miles downwind of major cites (black)
# rural/remote areas (>100 miles downwind; pink).

con <- read.table("/Users/Khoi/Dropbox/Rworldmap/Dataset -
withRH.csv",header=T,sep="",colClasses=c("character",
rep("numeric", 12),
"character","character","character"),na.strings="NA")
coniso = data.frame(Location=con$Location,
Na=c(0),SO4=con$Sulfate,NH3=con$Ammonium, NO3=
con$Nitrate,Cl=c(0),Ca=c(0),K=c(0),Mg=c(0),RH=as.numeric(con$RH),TEMP=as.numeric(con$Temp))

##write to ISORROPIA
header <- c(
  "Input units (0=umol/m3, 1=ug/m3) ; sample input file
  1
Problem type (0=forward, 1=reverse); Phase state (0=solid+liquid,
1=metastable)
1, 1
NH4-SO4-NO3 system case
Na  SO4  NH3  NO3  Cl  Ca  K  Mg  RH  TEMP")

# function that appends header to file
write.table_with_header <- function(x, file, header,
...){cat(header, "\n", file = file)
```

```
write.table(x, file, append = T, ...)}

#ignore warning message for line below
write.table_with_header(coniso[, (2:length(coniso))], "/Users/Khoi/Dropbox/Isorropia/v22_errck/WORLDMAP.INP", header, sep='   ', row.names=F, col.names=F)

library(maps)
library(mapproj)
library(mapplots)

#data <-
data <- data[order(data$Lon),]
data$textcolor[data$textcolor=="pink"] <- "magenta"

map('world', fill = TRUE, col = "grey90", xlim=c(-150,150), ylim=c(10,70), mar=c(rep(6,4)),oma=c(0,0,0,0))
points(x=data$Lon,y=data$Lat,col='blue',pch=data$pch,cex=1.1)
par(new=T)
plot(NA,NA, xlim=c(-150,150), ylim=c(10,70),las=1,
ylab="",xlab="",col="darkblue",xaxt="n",yaxt="n")
axis(1,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
axis(2,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
axis(3,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
axis(4,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)

par(new=T,mar=c(0,0,0,0))
plot(NA,NA, xlim=c(-1,1), ylim=c(-1,1) ,las=1,
ylab="",xlab="",xaxt="n",yaxt="n",bty="n")
a=1
for (i in seq(-0.8,0.8,0.2)){

add.pie(z=c(data[a,2],data[a,3],data[a,4],data[a,5],data[a,6]), x=i, y=0.77, radius=0.12,
labels="",col=c("red","blue","orange","green","lightblue"))
text(data$Location[a], x=i, y=0.96,cex=1,col=data$textcolor[a])
a=a+2
a = 2
for (i in seq(-0.8, 0.8, 0.2)) {
  add.pie(z = c(data[a, 2], data[a, 3], data[a, 4], data[a, 5], data[a, 6]), x = i, y = -0.77, labels = "", col = c("red", "blue", "orange", "green", "lightblue"))
  text(data$Location[a], x = i, y = -0.96, cex = 1, col = data$textcolor[a])
  a = a + 2
}
text(c("Urban", "Urban Downwind", "Rural"), x = c(0.6, 0.8, 1), y = c(-1.05), cex = 1, col = c("blue", "black", "magenta"))

dev.off()

### NEW DATA WITH REANALYSIS DATA

con <- read.table("/Users/Khoi/Google Drive/Rworldmap/Dataset_v.2.csv", header = T, sep = ",", colClasses = c("character", rep("numeric", 12), "character", "character"), na.strings = "NA")
coniso = data.frame(Location = con$Location, Na = c(0), SO4 = con$Sulfate, NH3 = con$Ammonium, NO3 = con$Nitrate, Cl = c(0), Ca = c(0), K = c(0), Mg = c(0), RH = 0.01 * as.numeric(con$RH), TEMP = as.numeric(con$Temp))
write.table_with_header(coniso[, (2:length(coniso))], "/Users/Khoi/Dropbox/Isorropia/v22_errck/WORLDMAP_v.2.INP", header, sep = " " , row.names = F, col.names = F)

data.iso <- read.csv("~/Google Drive/Rworldmap/worldmap_v.2/WORLDMAP_v.dat", sep = " ")
con$water = data.iso$WATER
data = con
data <- data[order(data$Lon),]
data$textcolor[data$textcolor == "pink"] <- "magenta"
write.csv(data, "~/Google Drive/Rworldmap/datawithwater_v.3.csv", row.names = F)

########## START here

data = read.csv("~/Google Drive/Rworldmap/datawithwater_v.3.csv")
write.csv(data, "~/Google Drive/Rworldmap/datawithwater_v.3.5.csv", row.names = F)

########## NEW PLOT

pdf(file = "~/Users/khoi/Google Drive/Rworldmap/map_v.2.pdf", width = 13, height = 7)
map('world', fill = TRUE, col = "grey90", xlim=c(-150,150), ylim=c(-60,70), mar=c(rep(10,4)), oma=c(0,0,0,0))
points(x=data$Lon,y=data$Lat,col='blue',pch=0,cex=1.5)
par(new=T)
plot(NA,NA, xlim=c(-150,150), ylim=c(-60,70) ,las=1,
ylab="",xlab="",col="darkblue",xaxt="n",yaxt="n")
axis(1,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
axis(2,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
axis(3,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
axis(4,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
lines(c(-170,170),c(0,0),lty=2,col="grey65")

par(new=T,mar=c(0,0,0,0))
plot(NA,NA, xlim=c(-1,1), ylim=c(-1,1) ,las=1,
ylab="",xlab="",xaxt="n",yaxt="n",bty="n")
a=1
for ( i in seq(-0.8,0.8,0.175)){
  add.pie(z=c(data[a,2],data[a,3],data[a,4],data[a,5],data[a,6]),
x=i, y=0.77, radius=0.12,
labels="",col=c("red","blue","orange","green","lightblue"))
  text(data$Location[a], x=i, y=0.96,cex=1,col=as.vector(data$textcolor[a]))
  a=a+2 }
a=2
for ( i in seq(-0.8,0.8,0.175)){
  add.pie(z=c(data[a,2],data[a,3],data[a,4],data[a,5],data[a,6]),
x=i, y=-0.77, radius=0.12,
labels="",col=c("red","blue","orange","green","lightblue"))
  text(data$Location[a], x=i, y=-0.96,cex=1,col=as.vector(data$textcolor[a]))
  a=a+2 }
text(c("Urban", "Urban Downwind","Rural"), x=c(0.6, 0.8, 1), y=c(-1.05),cex=1,col=c("blue","black","magenta"))

dev.off()

mean((data$Water[data$Sitetype=="urban"]))
mean((data$Water[data$Sitetype=="rural"]))
mean((data$Water[data$Sitetype=="urbandownwind"]))
data$sum=data$Sulfate+data$Nitrate+data$Ammonium+data$Organics+data$Water
data$waterfrac=data$Water/data$sum
data$dry = data$Sulfate+data$Nitrate+data$Ammonium+data$Organics
data$ratio = data$Water/data$dry
mean((data$ratio[data$Sitetype=="urban"]))
mean((data$ratio[data$Sitetype=="rural"]))
mean((data$ratio[data$Sitetype=="urbandownwind"]))
### FIGURING OUT ORGANIC WATER CONTRIBUTION

data = read.csv("~/Google Drive/Rworldmap/datawithwater_v.3.csv")
data$korg = NA
data$korg[data$Sitetype=="urban"] = 0.08
data$korg[data$Sitetype=="urbandownwind"] = 0.11
data$korg[data$Sitetype=="rural"] = 0.13
data$Worg = (data$Organics/1.4)*data$korg*(0.01*data$RH/(1 - 0.01*data$RH))
data$Wtot = data$water + data$Worg

data$pctdiff = (data$Wtot - data$water)/data$water*100

library(maps)
library(mapproj)
library(mapplots)

### new map

pdf(file="/Users/khoi/Google Drive/Rworldmap/map_v.3_withorgW.pdf", width = 13, height = 7)

map('world', fill = TRUE, col = "grey90", xlim=c(-150,150), ylim=c(-60,70), mar=c(rep(10,4),oma=c(0,0,0,0))
points(x=data$Lon,y=data$Lat,col='blue',pch=0,cex=1.5)
par(new=T)
plot(NA,NA, xlim=c(-150,150), ylim=c(-60,70),las=1, ylab="",xlab="",xaxt="n",yaxt="n")
axis(1,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
axis(2,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
axis(3,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
axis(4,las=1,col="darkblue",col.axis="darkblue",cex.axis=1.3)
lines(c(-170,170),c(0,0),lty=2,col="grey65")

par(new=T,mar=c(0,0,0,0))
plot(NA,NA, xlim=c(-1,1), ylim=c(-1,1),las=1, ylab="",xlab="",xaxt="n",yaxt="n",bty="n")
a=1
for (i in seq(-0.8,0.8,0.175)) {
  add.pie(z=c(data[a,2],data[a,3],data[a,4],data[a,5],data[a,20]),
x=i, y=0.77, radius=0.12,
  labels="",col=c("red","blue","orange","green","lightblue"))
  text(data$Location[a], x=i, y=0.96,cex=1,col=as.vector(data$textcolor[a]))
a=a+2
}
a=2
for (i in seq(-0.8,0.8,0.175)) {
  add.pie(z=c(data[a,2],data[a,3],data[a,4],data[a,5],data[a,20]),
x=i, y=-0.77, radius=0.12,
  labels="",col=c("red","blue","orange","green","lightblue"))
  text(data$Location[a], x=i, y=-0.96,cex=1,col=as.vector(data$textcolor[a]))
a=a+2
}
mean((data$Wtot[data$Sitetype=="urban"]))
mean((data$Wtot[data$Sitetype=="urbandownwind"]))
mean((data$Wtot[data$Sitetype=="rural"]))

data$dry = data$Sulfate+data$Nitrate+data$Ammonium+data$Organics
data$ratio2 = data$Wtot/data$dry

mean((data$ratio2[data$Sitetype=="urban"]))
mean((data$ratio2[data$Sitetype=="urbandownwind"]))
mean((data$ratio2[data$Sitetype=="rural"]))

mean((data$pctdiff[data$Sitetype=="urban"]))
mean((data$pctdiff[data$Sitetype=="urbandownwind"]))
mean((data$pctdiff[data$Sitetype=="rural"]))

mean((data$Organics[data$Sitetype=="urban"]))
mean((data$Organics[data$Sitetype=="urbandownwind"]))
mean((data$Organics[data$Sitetype=="rural"]))

data$ALWfraction = data$Wtot/(data$dry+data$Wtot)
data[order(data$ALWfraction),]

##### stacked bar plot
#####
data=data[order(data$Location),]

## urban
data.urban = data[data$Sitetype=="urban",]
urbanavg=data.frame(water=data.urban$water,
Worg=data.urban$Worg,org=data.urban$Organics,
ammonium=data.urban$Ammonium,
nitrate=data.urban$Nitrate,SO4=data.urban$Nitrate)
urbanavg2=t(urbanavg)
colnames(urbanavg2)=data.urban$Location
urban=as.matrix(urbanavg2)

## urban downwind
data.ud = data[data$Sitetype=="urbandownwind",]
udavg=data.frame(water=data.ud$water,
Worg=data.ud$Worg,org=data.ud$Organics, ammonium=data.ud$Ammonium,
nitrate=data.ud$Nitrate,SO4=data.ud$Nitrate)
# udavg[3:8,]=c(NA)
udavg2=t(udavg)
colnames(udavg2)=data.ud$Location
# colnames(udavg2)=c(data.ud$Location, rep(NA,6))
# colnames(udavg2[1:2])=c("Weybourne","Hohenpeissenberg")
ud = as.matrix(udavg2)

## rural
data.rural = data[data$Sitetype == "rural",]
ruralavg = data.frame(water = data.rural$water,
                      Worg = data.rural$Worg,
                      org = data.rural$Organics,
                      ammonium = data.rural$Ammonium,
                      nitrate = data.rural$Nitrate,
                      SO4 = data.rural$Nitrate)
ruralavg2 = t(ruralavg)
colnames(ruralavg2) = data.rural$Location
rural = as.matrix(ruralavg2)

## vertical
pdf(file = "~/Google Drive/Rworldmap/stackedbar.pdf", width = 6.6, height = 6.6)
#par(mfrow=c(3,1), mar=c(2,5,1,1), tcl = -0.5)
par(mfrow=c(1,1), mar=c(2,3,1,0), oma=c(1,2,0,2), tcl = -0.5)
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE), widths=c(3,1))
barplot(urban, col=c("lightblue", "cornflowerblue", "green", "orange", "blue", "red"), las=1, cex.axis = 1.24, ylim = c(0,160))
text(x=6.5, y=137, "Urban", cex = 1.25, adj = c(0,0))
legend("topleft", inset=0,
barplot(rural, col=c("lightblue", "cornflowerblue", "green", "orange", "blue", "red"), las=1, cex.axis = 1.24, ylim = c(0,14))
text(x=0.25, y=13, "Rural", cex = 1.25, adj = c(0,0))
barplot(ud,
col=c("lightblue", "cornflowerblue", "green", "orange", "blue", "red"), las = 1, cex.axis = 1.24, ylim = c(0,40))
text(x=0.25, y=37, "Urban", cex = 1.25, adj = c(0,0))
text(x=0.25, y=34, "Downwind", cex = 1.25, adj = c(0,0))
dev.off()

## horizontal
all.data = cbind(urbanavg2, udavg2, ruralavg2)
pdf(file = "~/Google Drive/Rworldmap/stackedbar_v.2.pdf", width = 6.6, height = 6)
par(mar=c(2,3,1,0), oma = c(3,2,0,0))
par(fig = c(0,1,0,0.8))
barplot(all.data, col = c("lightblue", "cornflowerblue", "green", "orange", "blue", "red"), cex.axis = 1.24, ylim = c(0,45), xpd = F, las=2)
legend("topright", inset=0,
c("Sulfate","Nitrate","Ammonium","Organics", "Organic
Water","Inorganic Water"),
pch=NA,text.col=c("red","blue","orange","green","cornflowerblue","lightblue"),
  cex=1,box.lwd = 0,box.col = "transparent",bg = "transparent",horiz=F)

mtext(expression(paste("Mass Concentration (\textmu g m^{-3},")", sep=" ")), side=2, line=3,cex=1.25)

library(Hmisc)

minor.tick(nx=0, ny=4, tick.ratio=0.5)

par(fig=c(0,1,0.7,1), new=TRUE)

barplot(all.data,col=c("lightblue","cornflowerblue","green","orange",
  "blue","red"),las=1,cex.axis=1.24,ylim=c(70,160),
xaxt="n",bty="n",xpd=F)

dev.off()
APPENDIX F: R CODES FOR CHAPTER 2

svdmadat <- read.csv("/Users/Khoi/Dropbox/SOAS/SOAS
DATA/fixed/svdmadat20130906.csv")
library (lubridate)
svdmadat$timelocal <- ymd_hms(as.character(svdmadat$timelocal))

library(Hmisc)
library(plyr)
## minor tick function
my.minor.tick <-
function (nx = 2, ny = 2, tick.ratio = 0.5, side=NA)
{ # added the side parameter
   ax <- function(w, n, tick.ratio) {
      range <- par("usr")[if (w == "x")
                             1:2
                             else 3:4]
      tick.pos <- if (w == "x")
                   par("xaxp")
                   else par("yaxp")
      distance.between.minor <- (tick.pos[2] -
                                tick.pos[1])/tick.pos[3]/n
      possible.minors <- tick.pos[1] - (0:100) *
                         distance.between.minor
      low.minor <- min(possible.minors[possible.minors >=
                                  range[1]])
      if (is.na(low.minor))
         low.minor <- tick.pos[1]
      possible.minors <- tick.pos[2] + (0:100) *
                         distance.between.minor
      hi.minor <- max(possible.minors[possible.minors <=
                                   range[2]])
      if (is.na(hi.minor))
         hi.minor <- tick.pos[2]
      if (.R.) # Next three lines have only the
               modifications
         axis(if (w == "x" & is.na(side) )
              1
              else if (!is.na(side)) side
              else 2, seq(low.minor, hi.minor, by =
                            distance.between.minor),
               labels = FALSE, tcl = par("tcl") * tick.ratio)
      else axis(if (w == "x")
                 1
                 else 2, seq(low.minor, hi.minor, by =
                               distance.between.minor),
                 labels = FALSE, tck = par("tck") * tick.ratio)
   } if (nx > 1)
      ax("x", nx, tick.ratio = tick.ratio)
   if (ny > 1)
      ax("y", ny, tick.ratio = tick.ratio)
invisible()}
mintick <- function(ax,n,t.ratio=0.5,mn,mx,...){

  lims <- par("usr")
  if(ax %in%c(1,3)) lims <- lims[1:2] else lims[3:4]

  major.ticks <- pretty(lims,n=5)
  if(missing(mn)) mn <- min(major.ticks)
  if(missing(mx)) mx <- max(major.ticks)

  major.ticks <- major.ticks[major.ticks >= mn & major.ticks <= mx]

  labels <- sapply(major.ticks,function(i)
    as.expression(bquote(10^ .(i)))
  )
  axis(ax,at=major.ticks,label=s=labels,...)

  n <- n+2
  minors <- log10(pretty(10^major.ticks[1:2],n))-major.ticks[1]
  minors <- minors[-c(1,n)]

  minor.ticks = c(outer(minors,major.ticks,`+`))
  minor.ticks <- minor.ticks[minor.ticks > mn & minor.ticks < mx]

  axis(ax,at=minor.ticks,tcl=par("tcl")*t.ratio,labels=FALSE)
}

library(ggplot2)
summarySE <- function(data=NULL, measurevar, groupvars=NULL,
na.rm=FALSE, conf.interval=.95, .drop=TRUE) {
  require(plyr)
  # New version of length which can handle NA's: if na.rm==T,
  # don't count them
  length2 <- function (x, na.rm=FALSE) {
    if (na.rm) sum(!is.na(x))
    else       length(x)
  }
  # This is does the summary
  datac <- ddply(data, groupvars, .drop=.drop,
  .fun= function(xx, col, na.rm) {
    c( N    = length2(xx[[col]]),
      mean = mean   (xx[[col]]),
      sd   = sd     (xx[[col]]),
      median = median(xx[[col]]),
      quartile25 =
    }
  }
}
quartile75 =
as.vector(summary(xx[[col]], na.rm=na.rm))[5]
)

  ),
  measurevar,
  na.rm
}

# Rename the "mean" column
datac <- rename(datac, c("mean"=measurevar))
datac$se <- datac$sd / sqrt(datac$N)  # Calculate standard error of the mean
# Confidence interval multiplier for standard error
# Calculate t-statistic for confidence interval:
# e.g., if conf.interval is .95, use .975 (above/below), and use df=N-1
ciMult <- qt(conf.interval/2 + .5, datac$N-1)
datac$ci <- datac$se * ciMult
return(datac)
}

## Figure 2: QA plot

#20130715
mass <- read.csv("/Users/Khoi/Dropbox/SOAS/soas DATA/mass/20130714.csv")
mass2 <- read.csv("/Users/Khoi/Dropbox/SOAS/soas DATA/mass/20130715.csv")
smith <- read.csv("/Users/Khoi/Dropbox/SOAS/soas DATA/smithfixed/smith20130714.csv")
smith2 <- read.csv("/Users/Khoi/Dropbox/SOAS/soas DATA/smithfixed/smith20130715.csv")

mass$hourlocal <- (mass$timehr-5)
mass2$hourlocal <- (mass2$timehr-5)+24
mass2<-mass2[mass2$timehr<5,]
mass <- rbind(mass,mass2)
smith$hourlocal <- (smith$timehr-5)
smith2$hourlocal <- (smith2$timehr-5)+24
smith2 <- smith2[smith2$timehr<5,]
smith <- rbind(smith,smith2)
mass <- mass[mass$hourlocal>0&m&mass$hourlocal<24,]
smith <- smith[smith$hourlocal>0&m smith$hourlocal<24,]

mass3 <- read.csv("/Users/Khoi/Dropbox/SOAS/soas DATA/mass/20130608.csv")
mass4 <- read.csv("/Users/Khoi/Dropbox/SOAS/soas DATA/mass/20130609.csv")
data2 <- read.csv("/Users/Khoi/Dropbox/SOAS/soas DATA/data2/20130608.csv")
data2$timehr <- data2$Time_String/3600
data3 <- read.csv("/Users/Khoi/Dropbox/SOAS/soas DATA/data2/20130609.csv")
data3$timehr <- data3$Time_String/3600
mass3$hourlocal <- (mass3$timehr-5) + 24
mass4 <- mass4[mass4$timehr<5]
data2$hourlocal <- (data2$timehr-5)
data3$hourlocal <- (data3$mass4$hourlocal<5)
mass3 <- rbind(mass3, mass4)
data2 <- rbind(data2, data3)
mass3 <- mass3[mass3$hourlocal>0 & mass3$hourlocal<24]
data2 <- data2[data2$hourlocal>0 & data2$hourlocal<24]

CPC <- NULL
CPC$CPC <- (data2$CPC_3772[data2$CPC_3772 >= 1200])/1000
data2 <- data2[data2$hourlocal>0 & data2$hourlocal<24]
CPC$CPC <- as.data.frame(CPC)

cairo_pdf(file="/Users/Khoi/Dropbox/Papers/Software\figures/figure2u.pdf", family="Swis721 BT", width = 6.6, height = 6)
par(mfrow=c(2,1), mar=c(0,5,1,1), oma=c(5,0,1,0), tcl=0.5)
plot(mass3$hourlocal[mass3$state==1],
mass3$CPCb[mass3$state==1]/1000, col="white", pch="", cex=0.8, ylim=c(0,6), type="l", yaxt="n", xaxt="n", ylab="", xlab="",
xlim=c(0,24))
points(mass3$hourlocal[mass3$state==1],
mass3$CPCb[mass3$state==1]/1000, col="red", pch=19, cex=0.7)
points(mass3$hourlocal[mass3$state==2],
mass3$CPCb[mass3$state==2]/1000, col="blue", pch=1, cex=0.7)
points(mass3$hourlocal[mass3$state==3],
mass3$CPCb[mass3$state==3]/1000, col="black", pch=15, cex=0.7)
points(CPC$timehr, CPC$CPC, col="grey60", pch="*", cex=0.5)
mtext(expression(paste("", " (", cm^-3, " x ", 10^3, sep="")),
side=2, line=2.25, cex=1.2)
#my.minor.tick(nx=6, ny=0, side=1)
# my.minor.tick(nx=5, ny=0, side=1)
# my.minor.tick(nx=0, ny=5, side=2)
# my.minor.tick(nx=0, ny=5, side=4)
axis(1, at=c(0,6,12,18,24), labels=F, cex.axis=1, las=1)
axis(2, pretty(c(0,6)), cex.axis=1, las=1)
axis(3, at=c(0,6,12,18,24), labels=F, cex.axis=1, las=1)
axis(4, pretty(c(0,10)), cex.axis=1, labels=F)
legend(1, 6, c("Ambient", "Dry-humidified", "Dry", "QA Check"),
pch=c(19,1,15,19), col=c("red","blue","black","grey60"),
cex=1, box.lwd = 0, box.col = "white", bg = "white")

plot(smith$hourlocal, smith$hourlocal, col="white", lwd=2,
ylim=c(0,6), xlim=c(0,24),
type="l", yaxt="n", xaxt="n", ylab="", xlab=""
)
points(smith$hourlocal, smith$concentration/1000, pch=".", 
cex=0.4,col="grey60", lwd=3)
mtext(expression(paste("", " (",cm^-3,") x ",10^3, sep="")), 
cex=1.2,side=2,line=2.25)
mtext(expression("Hours past midnight (local time)"), side=1, 
line=2.3, cex=1.2)
# my.minor.tick(nx=5, ny=0, side=3)
# my.minor.tick(nx=5, ny=0, side=1)
# my.minor.tick(nx=0, ny=5, side=2)
# my.minor.tick(nx=0, ny=5, side=4)
axis(1, at=c(0,6,12,18,24), labels=c(0,6,12,18,24), cex.axis=1, 
las=1)
axis(2, pretty(c(0,6)), cex.axis=1, las=1)
axis(3, at=c(0,6,12,18,24), labels=F, cex.axis=1, las=1)
axis(4, pretty(c(0,6)), cex.axis=1, labels=F)
dev.off()

## Figure 3
## # spectradat <- read.csv("/Users/Khoi/Dropbox/SOAS/DOAS 
## DATA/averagedist/averagedist20130604.csv")
## spectradat <- read.csv("/Users/Khoi/Dropbox/DOAS/DOAS 
## DATA/spectradatflagged/20130604withb.csv")
## spectradat$dn.dp <- spectradat$Inverted.CPC/(spectradat$Dup- 
## spectradat$Dlow)
## spectradat$dn.dp2 <- spectradat$dn/(spectradat$Dup-
## spectradat$Dlow)

spectradat <- read.csv("/Users/Khoi/Dropbox/DOAS/DOAS 
Data/dlogdpfits/dVdlogdpfit.csv")
spec <- read.csv("/Users/Khoi/Dropbox/DOAS/DOAS 
specfig3.csv")
spectradat2 <- read.csv("/Users/Khoi/Dropbox/DOAS/DOAS 
Data/dlogdpfits/dNdlogdpfit.csv")

spec2 <- spectradat2[i:(1+59),]
spec2 <-read.csv("/Users/Khoi/Dropbox/DOAS/DOAS 
Data/dlogdpfits/165721.csv")
spectradat2[i:165900,] <- spec2
## Number
# par(mfrow=c(2,3), mar=c(0,0,1,0), oma=c(5,5,1,1))
# i <- 165721
# plot(log10(0.001*spectradat2$Dmid[i:(i+59)]), 
# 0.001*spectradat2$dN.dlogDp[i:(i+59)],type="s", xaxt="n", xlab="", 
# ylab="", yaxt="n", ylim=c(0,15), xlim=c(-3,1))
# lines(log10(0.001*spectradat2$Dmid[i:(i+59)]), 
# 0.001*spectradat2$dN.dlogdpfit[i:(i+59)],type="l", col="red")
# axis(2, las=1, cex.axis=1.2)
# mtext(expression(paste("dN/", dlog[10], D[p], "", "("cm^-3, " 
# x", 10^3, sep="")), side=2, line=2.5, cex=0.8)
# axis(1,at=c(0.01, 0.10,1.00),labels=F, cex.axis=1.2)
# mintick(1,9,mm=-3,mx=1)
# minor.tick(nx=0,ny=5)
# pdf(file="/Users/Khoi/Dropbox/Papers/DOAS/DOAS/figures/figure3.pdf", 
# width = 6.6, height = 3.3)
par(mfrow=c(2,3), mar=c(0,0,1,0), oma=c(5,5,1,1),tcl=0.5)
i <- 165721
plot((0.001*spectradat2$Dmid[i:(i+59)]),
0.001*spectradat2$dN.dlogDp[i:(i+59)],type="s", xaxt="n", xlab="", ylab="", yaxt="n", ylim=c(0,15),log="x"
lines((0.001*spectradat2$Dmid[i:(i+59)]),
0.001*spectradat2$dN.dlogdpfit[i:(i+59)],type="l", col="red")
axis(2, las=1, cex.axis=1.2)
mtext(expression(paste("dN/", dlog[10], D[p], " (", cm^-3, " cm^-3)), side=2, line=2.5, cex=0.8)
axis(1,at=c(0.01, 0.10, 1.00), labels=F, cex.axis=1.2)
i <- i+60
plot((0.001*spectradat2$Dmid[i:(i+59)]),
0.001*spectradat2$dN.dlogDp[i:(i+59)],type="s", log="x", xaxt="n", xlab="", ylab="", yaxt="n", ylim=c(0,15))
lines((0.001*spectradat2$Dmid[i:(i+59)]),
0.001*spectradat2$dN.dlogdpfit[i:(i+59)],type="l", col="red")
axis(1,at=c(0.01, 0.10, 1.00), labels=F, cex.axis=1.2)
i <- i+60
plot((0.001*spectradat2$Dmid[i:(i+59)]),
spectradat2$dV.dlogDp[i:(i+59)],type="s", log="x", xaxt="n", xlab="", ylab="", yaxt="n", ylim=c(0,3))
lines((0.001*spectradat2$Dmid[i:(i+59)]),
spectradat2$dV.dlogdpfit[i:(i+59)],type="l", col="red")
axis(2, at=seq(0,3,1), las=1, cex.axis=1.2)
mtext(expression(paste("dV/", dlog[10], D[p], " (", \(\mu\)m^-3," cm^-3," ), sep="")), side=2, line=2.5, cex=0.8)
axis(1, at=c(0.01, 0.10, 1.00), labels=as.character(c(0.01, 0.1, 1)), cex.axis=1.2)
mtext(c("Mobility Diameter (\(\mu\)m")", side=1, line=2.6)
i <- i+60
plot((0.001*spectradat2$Dmid[i:(i+59)]),
spectradat2$dV.dlogDp[i:(i+59)],type="s", log="x", xaxt="n", xlab="", ylab="", yaxt="n", ylim=c(0,3))
lines((0.001*spectradat2$Dmid[i:(i+59)]),
spectradat2$dV.dlogdpfit[i:(i+59)],type="l", col="red")
axis(2, at=seq(0,3,1), las=1, cex.axis=1.2, labels=F)
axis(1, at=c(0.01, 0.10, 1.00), labels=as.character(c(0.01, 0.1, 1)),
cex.axis=1.2)
mtext(c("Mobility Diameter (\text{\(\mu\)m})"), side=1, line=2.6)

plot(0.001*spec$Dmid, spec$dV.dlogDp, type="s", log="x", xlab="", ylab="", yaxt="n", ylim=c(0,3))
lines(0.001*spec$Dmid, spec$dV.dlogdpfit, type="l", col="red")
axis(2, at=seq(0,3,1), las=1, cex.axis=1.2, labels=F)
axis(1, at=c(0.01, 0.10, 1.00), labels=as.character(c(0.01, 0.1, 1)),
cex.axis=1.2)
mtext(c("Mobility Diameter (\text{\(\mu\)m})"), side=1, line=2.6)
dev.off()

## Figure 4

specn <- read.csv("/Users/Khoi/Dropbox/SOAS/SOAS Data/dlogdpfits/dNdlogdpfit.csv")
specn1 <- specn[specn$state==1,]
specn2 <- specn[specn$state==2,]
specn3 <- specn[specn$state==3,]

specv <- read.csv("/Users/Khoi/Dropbox/SOAS/SOAS Data/dlogdpfits/dVdlogdpfit.csv")
specv1 <- specv[specv$state==1,]
specv2 <- specv[specv$state==2,]
specv3 <- specv[specv$state==3,]

specs <- read.csv("/Users/Khoi/Dropbox/SOASdSdlogdpfit.csv")
specs1 <- specs[specs$state==1,]
specs2 <- specs[specs$state==2,]
specs3 <- specs[specs$state==3,]

dn1 <- summarySE(specn1, measurevar="dN.dlogdpfit", groupvars=c("Dmid"))
dn2 <- summarySE(specn2, measurevar="dN.dlogdpfit", groupvars=c("Dmid"))
dn3 <- summarySE(specn3, measurevar="dN.dlogdpfit", groupvars=c("Dmid"))

dv1 <- summarySE(specv1, measurevar="dV.dlogdpfit", groupvars=c("Dmid"))
dv2 <- summarySE(specv2, measurevar="dV.dlogdpfit", groupvars=c("Dmid"))
dv3 <- summarySE(specv3, measurevar="dV.dlogdpfit", groupvars=c("Dmid"))

ds1 <- summarySE(specs1, measurevar="dS.dlogdpfit", groupvars=c("Dmid"))
ds2 <- summarySE(specs2, measurevar="dS.dlogdpfit", groupvars=c("Dmid"))
ds3 <- summarySE(specs3, measurevar="dS.dlogdpfit", groupvars=c("Dmid"))
plot(0.001*dn$Dmid, dn$median, col="white", yaxt="n", xaxt="n", ylab="", xlab="", log="x", ylim=c(0,400))
polygon(c(0.001*dn$Dmid, rev(0.001*dn$Dmid)), c((dn$quartile25), rev(dn$quartile75)), col="grey75", border="grey75")
lines(0.001*dn$Dmid,dn$median, col="black", lty=1)
axis(1,at=c(0.01, 0.10,1.00),labels=F, cex.axis=0.8, lwd=0, lwd.ticks=0.6)
axis(2, pretty(c(0,400)), labels=F,las=1, lwd=0, lwd.ticks=0.6)

mtext(expression(paste("dV/", dlog[10], D[p], " (",µm^3," cm^-3, "), sep="")), side=2, line=2.5, cex=0.6)
mtext(c("Mobility Diameter (µm)"), side=1, line=2.6,cex=0.7)

plot(0.001*dn$Dmid, dn$median, col="white", yaxt="n", xaxt="n", ylab="", xlab="", log="x", ylim=c(0,400))
polygon(c(0.001*dn$Dmid, rev(0.001*dn$Dmid)), c((dn$quartile25), rev(dn$quartile75)), col="grey75", border="grey75")
lines(0.001*dn$Dmid,dn$median, col="black", lty=1)
axis(1,at=c(0.01, 0.10,1.00),labels=F, cex.axis=0.8, lwd=0, lwd.ticks=0.6)
axis(2, pretty(c(0,400)), labels=F,las=1, lwd=0, lwd.ticks=0.6)

mtext(c("Mobility Diameter (µm)"), side=1, line=2.6,cex=0.7)
dev.off()

## Figure 5: Water and RH diurnal trend
water <- summarySE(svdmadat, measurevar="waterRHadj",
groupvars=c("hourlocal"))
RH <- summarySE(svdmadat, measurevar="RHamb",
groupvars=c("hourlocal"))
dry <- summarySE(svdmadat, measurevar="dry",
groupvars=c("hourlocal"))

cairo_pdf(file="/Users/Khoi/Dropbox/Papers/SOAS/figures/figure5.pdf",
  family="Swis721 BT", width = 6.6, height = 6)
par(oma=c(0,0,0,1),tcl=0.5)
plot(water$hourlocal, water$waterRHadj, col="white", ylim=c(0,20),
yaxt="n", xaxt="n", ylab="", xlab="", bty="n")
polygon(c(0:23, rev(0:23)), c((water$quartile25), rev(water$quartile75)),
col="cornflowerblue", border="cornflowerblue")
lines(water$hourlocal,water$median, col="black", lty=1)
my.minor.tick(nx=5, ny=0, side=3)
my.minor.tick(nx=5, ny=0, side=1)
my.minor.tick(nx=0, ny=5, side=2)
axis(1, at=c(0,6,12,18,24), labels=c(0,6,12,18,24),cex.axis=1,las=1)
axis(2, pretty(c(0,20)), cex.axis=1,las=1)
axis(3, at=c(0,6,12,18,24), cex.axis=1, labels=F)
par(new=T)
plot(RH$hourlocal, RH$RHamb, col="white", ylim=c(0,100), yaxt="n",
xaxt="n", ylab="", xlab="")
polygon(c(0:23, rev(0:23)), c((RH$quartile25), rev(RH$quartile75)),
col="grey85", border="grey85")
lines(RH$hourlocal,RH$median, col="black", lty=1)
axis(4, pretty(c(0,100)), cex.axis=1, las=1)
my.minor.tick(nx=0, ny=5, side=4)
mtext(expression(paste("Particle-phase liquid water", " (µg",m^-
-3,")", sep="")), side=2, line=2, cex=1)
mtext(expression("Hour past midnight (local time)"), side=1,
line=2.2, cex=1)
mtext(expression("Relative Humidity (%)"), side=4, line=2.2, cex=1)
#legend(16, 100, c(expression(paste(H[2], O[ptcl], " (\(\mu g\),m^-
-3,")", sep="")), "RH (%)", pch=c(15,15),col=c("blue","grey"),
cex=0.7,box.lwd = 0,box.col = "white",bg = "white")
dev.off()

cairo_pdf(file="/Users/Khoi/Dropbox/Papers/SOAS/figures/figure5b.pdf",
  family="Swis721 BT", width = 6.6, height = 6)
par(oma=c(0,0,0,1),tcl=0.5)
plot(water$hourlocal, water$waterRHadj, col="white", ylim=c(0,20),
yaxt="n", xaxt="n", ylab="", xlab="", bty="n")
polygon(c(0:23, rev(0:23)), c((water$waterRHadj-water$ci),
rev(water$waterRHadj+water$ci)), col="cornflowerblue", border="cornflowerblue")
lines(water$hourlocal,water$waterRHadj, col="black", lty=1)
my.minor.tick(nx=5, ny=0, side=3)
my.minor.tick(nx=5, ny=0, side=1)
my.minor.tick(nx=0, ny=5, side=2)
axis(1, at=c(0,6,12,18,24), labels=c(0,6,12,18,24), cex.axis=1, las=1)
axis(2, at=c(0,20), cex.axis=1, las=1)
axis(3, at=c(0,6,12,18,24), cex.axis=1, labels=F)
par(new=T)
plot(RH$hourlocal, RH$RHamb, col="white", ylim=c(0,100), yaxt="n",
xaxt="n", xlab="", ylab="")
polygon(c(0:23, rev(0:23)), c((RH$quartile25), rev(RH$quartile75)),
col="grey85", border="grey85")
lines(RH$hourlocal, RH$median, col="black", lty=1)
axis(4, at=c(0,100), cex.axis=1.5, las=1)
my.minor.tick(nx=5, ny=0, side=4)
my.minor.tick(nx=5, ny=0, side=1)
axis(4, at=c(10,100), cex.axis=1.5)
boxplot(svdmadat$RH ~ svdmadat$hourlocal, varwidth=T, col="grey",
ylim=c(10,100), yaxt="n", xaxt="n", cex.axis=1.5, las=1, col="grey")
my.minor.tick(nx=5, ny=0, side=3)
my.minor.tick(nx=5, ny=0, side=1)
axis(4, at=c(0,100), cex.axis=1.5)
axis(1, at=c(0,24), cex.axis=1.5)
library(Hmisc)
my.minor.tick(nx=0, ny=5, side=4)
# boxplot(svdmadat$waterRHadj ~ svdmadat$hourlocal, varwidth=T, col="grey",
ylim=c(10,100), xaxt="n", cex.axis=1.5, col="cornflowerblue")
my.minor.tick(nx=0, ny=5, side=3)
my.minor.tick(nx=0, ny=5, side=1)
# boxplot(svdmadat$HH ~ svdmadat$hourlocal, varwidth=T, col="grey",
ylim=c(10,100), xaxt="n", cex.axis=1.5, col="cornflowerblue")
my.minor.tick(nx=0, ny=5, side=3)
my.minor.tick(nx=0, ny=5, side=1)
# boxplot(svdmadat$HH ~ svdmadat$hourlocal, varwidth=T, col="grey",
ylim=c(10,100), xaxt="n", cex.axis=1.5, col="cornflowerblue")
# legend(13, 74, c(expression(paste(H[2], 0[ptcl], " (\(\mu g/\),m^3,")),
sep="")), "RH (%)"), lty=1, lwd=c(1,1), col=c("blue","grey"),
cex=1.5, box.lwd = 0, box.col = "white", bg = "white")
devoff()
# cairo_pdf(file="/Users/Khoi/Dropbox/Papers/figure5dry.pdf",
# width = 6.6, height = 6)
par(oma=c(0,0,0,1), tcl=0.5)
my.minor.tick(nx=5, ny=0, side=3)
my.minor.tick(nx=5, ny=0, side=1)
my.minor.tick(nx=0, ny=5, side=2)
plot(water$hourlocal, water$waterRHadj, col="white", ylim=c(0,20), yaxt="n", xaxt="n", ylab="", xlab="", bty="n")
polygon(c(0:23, rev(0:23)), c((dry$quartile25), rev(dry$quartile75)), col="brown", border="brown")
lines(dry$hourlocal,dry$median, col="black", lty=1)
my.minor.tick(nx=0, ny=5, side=2)
axis(1, at=c(0,6,12,18,24), labels=c(0,6,12,18,24),cex.axis=1,las=1)
axis(2, pretty(c(0,20)), cex.axis=1,las=1)
axis(3, at=c(0,6,12,18,24), cex.axis=1, labels=F)
dev.off()

## Figure 6: Kappa diurnal trend (both modes and volume-based)

svdmadat <- read.csv("/Users/Khoi/Dropbox/SOAS/soas DATA/fixed/svdmadat20130906.csv")
svdmadat <- svdmadat[svdmadat$kappadiam1<5 & svdmadat$kappadiam1 >-0.3,]
summary(svdmadat$kappadiam1 [svdmadat$kappadiam1<5 & svdmadat$kappadiam1 >-0.3])

kappanuc <- summarySE(svdmadat, measurevar="kappadiam1", groupvars=c("hourlocal"))
kappaacc <- summarySE(svdmadat, measurevar="kappadiam2", groupvars=c("hourlocal"))
kappavol <- summarySE(svdmadat, measurevar="kappa", groupvars=c("hourlocal"))

cairo_pdf(file="/Users/Khoi/Dropbox/Papers/SOAS/figures/figure6.pdf", family="Swis721 BT",width = 6.6, height = 6)

par(mfrow=c(3,1), mar=c(0,5,1,1), oma=c(5,0,1,0),tcl=0.5)
plot(kappanuc$median, col="white", ylim=c(0,1), xaxt="n", ylab="", xlab="", cex.axis=1.3)
polygon(c(1:24, rev(1:24)), c((kappanuc$quartile25), rev(kappanuc$quartile75)), col="grey87", border="grey87")
lines(kappanuc$median, col="black", lty=1)
mtext(expression(paste(K[D1]))), side=2, line=2.3, cex=1)
axis(1, at=c(0,6,12,18,24), labels=F,cex.axis=1,las=1)
axis(2, pretty(c(0,1)), cex.axis=1.5, labels=F)
axis(3, at=c(0,6,12,18,24), labels=F,cex.axis=1.3,las=1)
axis(4, pretty(c(0,1)), cex.axis=1, labels=F)

plot(kappaacc$median, col="white", ylim=c(0,1), xaxt="n", ylab="", xlab="", cex.axis=1.3)
polygon(c(1:24, rev(1:24)), c((kappaacc$quartile25), rev(kappaacc$quartile75)), col="grey87", border="grey87")
lines(kappaacc$median, col="black", lty=1)
mtext(expression(paste(K[D2]))), side=2, line=2.3, cex=1)
axis(1, at=c(0,6,12,18,24), labels=F,cex.axis=1,las=1)
axis(2, pretty(c(0,1)), cex.axis=1.5, labels=F)
axis(3, at=c(0,6,12,18,24), labels=F,cex.axis=1,las=1)
axis(4, pretty(c(0,1)), cex.axis=1, labels=F)
plot(kappavol$median, col="white", ylim=c(0,1), xaxt="n", ylab="", xlab="", cex.axis=1.3)
polygon(c(1:24, rev(1:24)), c((kappavol$quartile25), rev(kappavol$quartile75)), col="grey87", border="grey87")
lines(kappavol$median, col="black", lty=1)
mtext(expression(paste(K[vol])), side=2, line=2.3, cex=1)
axis(1, at=c(0,6,12,18,24),
labels=c(0,6,12,18,24),cex.axis=1.3,las=1)
axis(2, pretty(c(0,1)), cex.axis=1.5, labels=F)
axis(3, at=c(0,6,12,18,24), labels=F,cex.axis=1,las=1)
axis(4, pretty(c(0,1)), cex.axis=1, labels=F)
mtext(expression("Hour past midnight (local time)"), side=1,
line=2.5, cex=1)
dev.off()

## Figure 7: Growth factor diurnal trend (both modes and volume-based)
gf <- summarySE(svdmadat, measurevar="GFRHadj", groupvars=c("hourlocal"))

### kappa isolines: Vw = Vdry * K *(RH^-1 -1)^-1
RH <- summarySE(svdmadat, measurevar="RHamb", groupvars=c("hourlocal"))
dry <- summarySE(svdmadat, measurevar="dry", groupvars=c("hourlocal"))
svdmadat$k0.1 <-((svdmadat$dry* 0.1*((0.01* svdmadat$RHamb)^( -1)-1)^(-1)+ svdmadat$dry)/svdmadat$dry
svdmadat$k0.2 <-((svdmadat$dry* 0.2*((0.01* svdmadat$RHamb)^( -1)-1)^(-1)+ svdmadat$dry)/svdmadat$dry
svdmadat$k0.3 <-((svdmadat$dry* 0.3*((0.01* svdmadat$RHamb)^( -1)-1)^(-1)+ svdmadat$dry)/svdmadat$dry
svdmadat$k0.4 <-((svdmadat$dry* 0.4*((0.01* svdmadat$RHamb)^( -1)-1)^(-1)+ svdmadat$dry)/svdmadat$dry
svdmadat$k0.5 <-((svdmadat$dry* 0.5*((0.01* svdmadat$RHamb)^( -1)-1)^(-1)+ svdmadat$dry)/svdmadat$dry
svdmadat$k0.6 <-((svdmadat$dry* 0.6*((0.01* svdmadat$RHamb)^( -1)-1)^(-1)+ svdmadat$dry)/svdmadat$dry
svdmadat$k0 <-((svdmadat$dry* 0.6*((0* svdmadat$RHamb)^( -1)-1)^(-1)+ svdmadat$dry)/svdmadat$dry
k0 <- summarySE(svdmadat, measurevar="k0", groupvars=c("hourlocal"))
k0.1 <- summarySE(svdmadat, measurevar="k0.1", groupvars=c("hourlocal"))
k0.2 <- summarySE(svdmadat, measurevar="k0.2", groupvars=c("hourlocal"))
k0.3 <- summarySE(svdmadat, measurevar="k0.3", groupvars=c("hourlocal"))
k0.4 <- summarySE(svdmadat, measurevar="k0.4", groupvars=c("hourlocal"))
k0.4 <- summarySE(svdmadat, measurevar="k0.4", groupvars=c("hourlocal"))
k0.5 <- summarySE(svdmadat, measurevar="k0.5", groupvars=c("hourlocal"))
k0.6 <- summarySE(svdmadat, measurevar="k0.6", groupvars=c("hourlocal"))

# k0.1 <- (dry$dry* 0.1* ((0.01*RH$RHamb)^(-1)-1)^(-1)+dry$dry)/dry$dry
# k0.2 <- (dry$dry*0.2* ((0.01*RH$RHamb)^(-1)-1)^(-1)+dry$dry)/dry$dry
# k0.3 <- (dry$dry*0.3* ((0.01*RH$RHamb)^(-1)-1)^(-1)+dry$dry)/dry$dry
# k0.4 <- (dry$dry*0.4* ((0.01*RH$RHamb)^(-1)-1)^(-1)+dry$dry)/dry$dry
# k0.5 <- (dry$dry*0.5* ((0.01*RH$RHamb)^(-1)-1)^(-1)+dry$dry)/dry$dry
cairo_pdf(file="/Users/Khoi/Dropbox/Papers/SOAS/figures/figure7b.pdf", width = 6.6, height = 6,family="Swis721 BT")

par(oma=c(0,0,0,1),tcl=0.5)
plot(gf$median, col="white", ylim=c(0,4), yaxt="n", xaxt="n",
ylab="", xlab="")
polygon(c(1:24, rev(1:24)), c((gf$qquartile25), rev(gf$qquartile75)),
col="lightblue1", border="lightblue1")
lines(gf$median, col="black", lty=1)
lines(k0$median, type="l", lty=3, col="grey65")
lines(k0.1$median, type="1", lty=3, col="grey65")
lines(k0.2$median, type="1", lty=3, col="grey65")
lines(k0.3$median, type="1", lty=3, col="grey65")
lines(k0.4$median, type="1", lty=3, col="grey65")
lines(k0.5$median, type="1", lty=3, col="grey65")
lines(k0.6$median, type="1", lty=3, col="grey65")

#my.minor.tick(nx=0, ny=5, side=2)
#my.minor.tick(nx=0, ny=5, side=4)
axis(1, at=c(0,6,12,18,24),
labels=c(0,6,12,18,24),cex.axis=1.3,las=1)
#axis(1, pretty(c(0,24)), cex.axis=1.3)
axis(2, pretty(c(0,4)), cex.axis=1, las=1)
axis(3, at=c(0,6,12,18,24), labels=F,cex.axis=1.3,las=1)
axis(4, pretty(c(0,4)), cex.axis=1,labels=F)

mtext(expression(paste(gf[vol])), side=2, line=2.3, cex=1)
mtext(expression("Hours past midnight (local time")), side=1,
line=2.5, cex=1.3)

dev.off()
# boxplot(kappadiam1~hourlocal, data=svdmadat, ylim=c(0,1), cex=0.5, pch=19, xaxt="n", cex.axis=1.5, varwidth=T, border="black", col="papayawhip")
# mtext(expression(bold(paste(K[D1]))), side=2, line=3, cex=1.5)
# my.minor.tick(nx=0, ny=5, side=2)
# my.minor.tick(nx=0, ny=5, side=4)
# boxplot(kappadiam2~hourlocal, data=svdmadat, ylim=c(0,1), pch=19, cex.axis=1.5, xaxt="n", varwidth=T, border="black", col="limegreen")
# mtext(expression(bold(paste(K[D2]))), side=2, line=3, cex=1.5)
# my.minor.tick(nx=0, ny=5, side=2)
# my.minor.tick(nx=0, ny=5, side=4)
# boxplot(kappa~hourlocal, data=svdmadat, ylim=c(0,1), pch=19, xaxt="n", cex.axis=1.5, varwidth=T, border="black", col="salmon")
# mtext(expression(bold(paste(K[vol]))), side=2, line=3, cex=1.5)
# my.minor.tick(nx=0, ny=5, side=2)
# my.minor.tick(nx=0, ny=5, side=4)
# mtext(expression(bold("Hour past midnight (local time)")), side=1, line=3, cex=1.45)
# axis(1, pretty(c(0,24)), cex.axis=1.5)
# my.minor.tick(nx=0, ny=5, side=2)
# my.minor.tick(nx=0, ny=5, side=4)
# my.minor.tick(nx=5, ny=0, side=1)
# my.minor.tick(nx=5, ny=0, side=3)
# mtext(expression(bold("Hour past midnight (local time)")), side=1, line=2.2, cex=1)
# dev.off()

## Figure 8: SV diurnal trend (with mean and confidence interval)

dfc <- summarySE(svdmadat, measurevar="sv", groupvars=c("hourlocal"))

cairo_pdf(file="/Users/Khoi/Dropbox/Papers/SOAS/figures/figure8b.pdf ", family="Swis721 BT",width = 6.6, height = 6)

par(oma=c(0,0,0,1),tcl=0.5)

plot(dfc$sv, col="white", ylim=c(-2,2), xaxt="n",yaxt="n", ylab="",xlab=""

polygon(c(1:24, rev(1:24)), c((dfc$sv-dfc$ci), rev(dfc$sv+dfc$ci)),
col="navyblue", border="navyblue")

lines(dfc$sv, col="grey80")

axis(1, at=c(0,6,12,18,24), labels=c(0,6,12,18,24),cex.axis=1.3,las=1)

axis(2, pretty(c(-2,2)), cex.axis=1)

axis(3, pretty(c(0,24)), cex.axis=1, labels=F)

axis(4, pretty(c(-2,2)), cex.axis=1, labels=F)

#my.minor.tick(nx=0, ny=5, side=2)
#my.minor.tick(nx=0, ny=5, side=4)
#my.minor.tick(nx=5, ny=0, side=1)
#my.minor.tick(nx=5, ny=0, side=3)

mtext(expression(paste(V[sv], " (\mu m^3 cm^{-3}, \)")), side=2, line=2.2, cex=1)

mtext(expression("Hour past midnight (local time)"), side=1, line=2.2, cex=1)

dev.off()

#lines(c(0:24), rep(0,25))
library(scales)
library(grid)
pdf(file='/Users/Khoi/Dropbox/Papers/SOAS/figures/figure7.pdf',
width = 11, height = 11)
ggplot(svdmadat,aes(x=hourlocal,y= sv)) + stat_summary(fun.data = 
"mean_cl_normal", geom="smooth",
fill="darkseagreen2",size=1.5,col="darkgreen") +
ylab(expression(bold(paste("SVOCs", " (",\n\mu m^3,"/",cm^3, "\)",
sep="")))) +xlab(expression(bold(paste("Hour past midnight (local
\time)", sep="")))) +theme_bw()+theme(axis.text = element_text(colour 
= "black", size=19),axis.ticks.length = unit(-0.25 , "cm"),
axis.ticks.margin = unit(0.5, "cm"), axis.title =
element_text(face="bold", size=24),
panel.grid.major=element_blank(),panel.grid.minor=element_blank(),
panel.border = element_rect(colour = "black",size=1.25))
dev.off()

# Table 1
k <- which ((0.001*spectradat$Dmid)<0.1695)
data <- spectradat[k,]
for (i in seq(1,nrow(data), 40)){data$averageC[i:(i+39)] <-
sum(data$Inverted.CPC[i:(i+39)])}
summary(data$averageC[data$state==1])
summary(data$averageC[data$state==2])
summary(data$averageC[data$state==3])
data$ns <- pi*(0.001*data$Dmid)^2*data$Inverted.CPC
for (i in seq(1,nrow(data), 40)){data$averageS[i:(i+39)] <-
sum(data$ns[i:(i+39)])}
summary(data$averageS[data$state==1])
summary(data$averageS[data$state==2])
summary(data$averageS[data$state==3])

k1 <- which ((0.001*spectradat$Dmid)>0.1695)
data2 <- spectradat[k1,]
for (i in seq(1,nrow(data2), 20)){data2$averageC[i:(i+19)] <-
sum(data2$Inverted.CPC[i:(i+19)])}
data2$ns <- pi*(0.001*data2$Dmid)^2*data2$Inverted.CPC
for (i in seq(1,nrow(data2), 20)){data2$averageS[i:(i+19)] <-
sum(data2$ns[i:(i+19)])}
summary(data2$averageC[data2$state==1])
summary(data2$averageC[data2$state==2])
summary(data2$averageC[data2$state==3])
summary(data2$averageS[data2$state==1])
summary(data2$averageS[data2$state==2])
summary(data2$averageS[data2$state==3])

k3 <- which ((0.001*spectradat$Dmid)>0.175)
data3 <- spectradat[k3,]
for (i in seq(1,nrow(data3), 20)){data3$averageC[i:(i+19)] <-
sum(data3$Inverted.CPC[i:(i+19)])}
data3$ns <- pi*(0.001*data3$Dmid)^2*data3$Inverted.CPC
for (i in seq(1,nrow(data3), 20)){data3$averageS[i:(i+19)] <-
sum(data3$ns[i:(i+19)])
}

summary(data3$averageC[data3$state==1])
summary(data3$averageC[data3$state==2])
summary(data3$averageC[data3$state==3])

summary(data3$averageS[data3$state==1])
summary(data3$averageS[data3$state==2])
summary(data3$averageS[data3$state==3])

k4 <- which ((0.001*spectradat$Dmid)<0.175)
data4 <- spectradat[k4,]
data4$ns <- pi*(0.001*data4$Dmid)^2*data4$Inverted.CPC
for (i in seq(1,nrow(data4), 40)){data4$averageS[i:(i+39)] <-
sum(data4$ns[i:(i+39)])
}

summary(data4$averageS[data4$state==1])
summary(data4$averageS[data4$state==2])
summary(data4$averageS[data4$state==3])

#Table 3
(log(max(svdmadat$waterRHadj), base=exp(1))-log(min(svdmadat$waterRHadj), base=exp(1)))/log((max(svdmadat$dry))-log(min(svdmadat$dry)), base=exp(1))

(max(svdmadat$waterRHadj)-
min(svdmadat$waterRHadj))/(max(svdmadat$RHamb)-min(svdmadat$RHamb))
(max(svdmadat$waterRHadj)-
min(svdmadat$waterRHadj))/(max(svdmadat$kappa)-min(svdmadat$kappa))

### OTHER
x<- svdmadat$RHamb;y<- svdmadat$waterRHadj;
plot(x,y, cex=0.2, pch=19, col="cornflowerblue", ylim=c(0,20),
xlab="Ambient RH (%)", ylab="Aerosol Water (µg/m3)")
f <- function(x,a,b,d) {(a*x^2) + (b*x) + d}
fit <- nls(y ~ f(x,a,b,d), start = c(a=1, b=1, d=1))
co <- coef(fit)
curve(f(x, a=co[1], b=co[2], d=co[3]), add = TRUE, col="black",
lwd=2)

### alexis
# The two days I am going to use for the intercomparison are 6/22
# and 7/2
# dVdlogDp matrix
# dNdlogDp matrix
# Total volume concentration
# Total number concentration
spectradat <- read.csv("/Users/Khoi/Dropbox/SOAS/SOAS
DATA/spectradat/20130622.csv")
library(lubridate)
spectradat$Dlow <- c(spectradat$Diameter.Array[2:60], 0)
spectradat$Dup <- spectradat$Diameter.Array[1:60]
spectradat$Dmid <- (spectradat$Dlow+spectradat$Dup)/2
spectradat$timeUTC <- ymd_hms(as.character(spectradat$timeUTC))
spectradat$timelocal <- with_tz(spectradat$timeUTC)
spectradat2 <- spectradat[day(spectradat$timelocal)==22,]
spectradat22 <- rbind(spectradat2, spectradat3)

spectradat22$dN.dlogDp <-
spectradat22$Inverted.CPC/(log(spectradat22$Dup, base=10)-
log(spectradat22$Dlow, base=10))
spectradat22$dV.dlogDp <-
spectradat22$dN.dlogDp*pi/6*(0.001*spectradat22$Dmid)^3

spectradat22$dN.dlogDp[spectradat22$dN.dlogDp<0] <- 0
spectradat22$dN.dlogDp[spectradat22$dN.dlogDp>700] <- 0
spectradat22 <- spectradat22[spectradat22$state==3,]
spectradat <- data.frame(timelocal=spectradat22$timelocal, dN.dlogDp = spectradat22$dN.dlogDp, dV.dlogDp=spectradat22$dV.dlogDp, CPC = spectradat22$CPC, volume=spectradat22$mass)
spectradat$Dmid <- spectradat22$Dmid
write.table(spectradat, 
"/Users/Khoi/Dropbox/SoAS/alexis_20130622.txt",row.names=F)

###

spectradat <- read.csv("/Users/Khoi/Dropbox/SoAS/SOAS
DATA/spectradat/20130702.csv")
library(lubridate)
spectradat$Dlow <- c(spectradat$Diameter.Array[2:60], 0)
spectradat$Dup <- spectradat$Diameter.Array[1:60]
spectradat$Dmid <- (spectradat$Dlow+spectradat$Dup)/2
spectradat$timeUTC <- ymd_hms(as.character(spectradat$timeUTC))
spectradat$timelocal <- with_tz(spectradat$timeUTC)
spectradat2 <- spectradat[day(spectradat$timelocal)==2,]
spectradat72 <- rbind(spectradat2, spectradat3)

spectradat72$dN.dlogDp <-
spectradat72$Inverted.CPC/(log(spectradat72$Dup, base=10)-
log(spectradat72$Dlow, base=10))
spectradat72$dV.dlogDp <-
spectradat72$dN.dlogDp*pi/6*(0.001*spectradat72$Dmid)^3
spectradat72$dN.dlogDp[spectradat72$dN.dlogDp<0] = 0
spectradat72$dN.dlogDp[spectradat72$Dmid>700] = 0
spectradat72 <- spectradat72[spectradat72$state==3,]

spectradat <- data.frame(timelocal=spectradat72$timelocal, dN.dlogDp = spectradat72$dN.dlogDp, dV.dlogDp=spectradat72$dV.dlogDp, CPC = spectradat72$CPC, volume=spectradat72$mass)

write.table(spectradat, 
"/Users/Khoi/Dropbox/SOAS/alexis_20130702.txt",row.names=F)

######

spectradat <- read.csv("/Users/Khoi/Dropbox/SOAS/20130604.csv")
library(lubridate)
spectradat$Dlow <- c(spectradat$Diameter.Array[2:60], 0)
spectradat$Dup <- spectradat$Diameter.Array[1:60]
spectradat$Dmid <- (spectradat$Dlow+spectradat$Dup)/2

spectradat$Time_string <- spectradat$timehr*3600
spectradat$timeUTC <- as.POSIXct(paste("2013-06-04", format(as.POSIXct('0001-01-01 00:00:00') + spectradat$Time_string, "%H:%M:%S"), format="%Y-%m-%d %H:%M:%S", tz="UTC")
#spectradat$timeUTC <- ymd_hms(as.character(spectradat$timeUTC))
spectradat$timelocal <- with_tz(spectradat$timeUTC)

spectradat2 <- spectradat[day(spectradat$timelocal)==4,]

spectradat <- read.csv("/Users/Khoi/Dropbox/SOAS/20130605.csv")
library(lubridate)
spectradat$Dlow <- c(spectradat$Diameter.Array[2:60], 0)
spectradat$Dup <- spectradat$Diameter.Array[1:60]
spectradat$Dmid <- (spectradat$Dlow+spectradat$Dup)/2

spectradat$Time_string <- spectradat$timehr*3600
spectradat$timeUTC <- as.POSIXct(paste("2013-06-05", format(as.POSIXct('0001-01-01 00:00:00') + spectradat$Time_string, "%H:%M:%S"), format="%Y-%m-%d %H:%M:%S", tz="UTC")
#spectradat$timeUTC <- ymd_hms(as.character(spectradat$timeUTC))
spectradat$timelocal <- with_tz(spectradat$timeUTC)

spectradat3 <- spectradat[day(spectradat$timelocal)==4,]

spectradat72<- rbind(spectradat2,spectradat3)

spectradat72$dN.dlogDp <- spectradat72$Inverted.CPC/(log(spectradat72$Dup, base=10)-log(spectradat72$Dlow, base=10))
spectradat72$dV.dlogDp <-
spectradat72$dN.dlogDp*pi/6*(0.001*spectradat72$Dmid)^3

spectradat72$dN.dlogDp[spectradat72$dN.dlogDp<0] = 0
spectradat72$dV.dlogDp[spectradat72$dV.dlogDp<0] = 0
spectradat72$dN.dlogDp[spectradat72$Dmid>800] = 0
spectradat72$dV.dlogDp[spectradat72$Dmid>800] = 0
spectradat72 <- spectradat72[spectradat72$state==3,]
spectradat <- data.frame(timelocal=spectradat72$timelocal, Diameter = spectradat72$Dmid,dN.dlogDp = spectradat72$dN.dlogDp, dV.dlogDp=spectradat72$dV.dlogDp, CPC = spectradat72$CPC, volume=spectradat72$mass)
write.table(spectradat, 
"/Users/Khoi/Dropbox/SOAS/alexis_20130604.txt",row.names=F)

### SI Figure 1
n=21
m=6
library(lubridate)
svdamat$timelocal <- ymd_hms(as.character(svdmadat$timelocal ))
svdamat$k.0.1 <-svdamat$dry* 0.1*((0.01* svdamat$RHamb)^(-1)-1)^(-1)+ svdamat$dry
svdamat$k.0.2 <-svdamat$dry* 0.2*((0.01* svdamat$RHamb)^(-1)-1)^(-1)+ svdamat$dry
svdamat$k.0.3 <-svdamat$dry* 0.3*((0.01* svdamat$RHamb)^(-1)-1)^(-1)+ svdamat$dry
svdamat$k.0.4 <-svdamat$dry* 0.4*((0.01* svdamat$RHamb)^(-1)-1)^(-1)+ svdamat$dry
svdamat$k.0.5 <-svdamat$dry* 0.5*((0.01* svdamat$RHamb)^(-1)-1)^(-1)+ svdamat$dry
svdamat$k.0.6 <-svdamat$dry* 0.6*((0.01* svdamat$RHamb)^(-1)-1)^(-1)+ svdamat$dry
svdamat$k.0 <-svdamat$dry* 0.6*((0* svdamat$RHamb)^(-1)-1)^(-1)+ svdamat$dry

day21 <- svdamat[day(svdmadat$timelocal)==n,]
day21 <- day21[month(day21$timelocal)==m,]
cairo_pdf(file="/Users/Khoi/Dropbox/Papers/SOAS/figures/SI-1.pdf", 
family="Swis721 BT",width = 6.6, height = 6)
par(oma=c(0,0,0,1),tcl=0.5)
plot(day21$timelocal, day21$GFRHadj, type="l", las=1, xlab="Time of day (local time)", ylab=c("Volumetric growth factor"),
col="blue",ylim=c(1,4.5))
lines(day21$timelocal,day21$k.0.1, lty=3, col="grey65")
lines(day21$timelocal,day21$k.0.2, lty=3, col="grey65")
lines(day21$timelocal,day21$k.0.3, lty=3, col="grey65")
lines(day21$timelocal,day21$k.0.4, lty=3, col="grey65")
lines(day21$timelocal, day21$k0.5, lty=3, col="grey65")
lines(day21$timelocal, day21$k0.6, lty=3, col="grey65")
par(new=T)
plot(day21$RHamb~ day21$timelocal, type="l", lty=2, xaxt="n",
yaxt="n", xlab="", ylab="", ylim=c(0,100))

axis(4, las=1)
title(c("June 21, 2013"))
mtext(c("RH (%)"), side=4, line=2)
dev.off()

### water data for other people
waterdata <- data.frame(timeUTC = svdmadat$timeUTC,
timelocal=svdmadat$timelocal, RH=svdmadat$RHamb, dry=svdmadat$dry,
sv=svdmadat$sv, water=svdmadat$waterRHadj, kappa=svdmadat$kappa)
write.table(waterdata, "/Users/Khoi/Dropbox/SOAS/waterdata.txt",
row.names=F)

### dry data for rebecca
specn <- read.csv("/Users/Khoi/Dropbox/SOAS/soasData/dlogdpfits/dNdllogDpfit.csv")
specn3 <- specn[specn$state==3,]
spectradat72 <- specn3
spectradat72$dN.dlogDp[spectradat72$dN.dlogDp<0] = 0
spectradat72$dV.dlogDp[spectradat72$dV.dlogDp<0] = 0
spectradat72$dN.dlogDp[spectradat72$Dmid>900] = 0
spectradat72$dV.dlogDp[spectradat72$Dmid>900] = 0
spectradat72 <- spectradat72[spectradat72$state==3,]

library(lubridate)
spectradat72$date <- ymd(spectradat72$date)
spectradat72$Time_string <- spectradat72$timehr*3600
spectradat72 $timeUTC <-
as.POSIXct(paste(as.character(spectradat72$date),
format(as.POSIXct(’0001-01-01 00:00:00’)) + spectradat72
$Time_string, "%H:%M:%S"), format="%Y-%m-%d %H:%M:%S", tz="UTC")
#spectradat72$timeUTC <- ymd_hms(as.character(spectradat72$timeUTC))
spectradat72$timelocal <- with_tz(spectradat72$timeUTC,
tz="America/Chicago")

y<- which(specn3$date>20130703)
#spectradat72$dV.dlogDp <- specn3$dV.dlogDp

spectradat72$intCPC2 <- spectradat72$intCPC
spectradat72$intvolume2 <- spectradat72$intvolume

spectradat72$intCPC[y] <- 2*(spectradat72$intCPC[y])
spectradat72$intvolume[y] <- 2*spectradat72$intvolume[y]

for (i in seq(1,nrow(spectradat72),60))
{spectradat72$intCPC[i:(i+59)] <-
{sum(spectradat72$dN.dlogDp[i:(i+58)]}*
(log10(spectradat72$Dup[i:(i+58)]) -
log10(spectradat72$Dlow[i:(i+58)])))
for (i in seq(1,nrow(spectradat72),60))
{spectradat72$intvolume[i:(i+59)] <-
(sum(spectradat$dV.dlogDp[i:(i+58)]* 
(log10(spectradat72$Dup[i:(i+58)]) -
log10(spectradat72$Dlow[i:(i+58)]))))
}
spectradat <- data.frame(timelocal=spectradat72$timelocal, Diameter = spectradat72$Dmid,dN.dlogDp = spectradat72$dN.dlogDp, 
dV.dlogDp=spectradat72$dV.dlogDp, CPC = spectradat72$intCPC, 
volume=spectradat72$intvolume)

k<-which(spectradat$volume > 30)
k<- which(round(spectradat$volume, digits=4) == 30.4361)
#spectradat<-spectradat[-k,]
write.table(spectradat, 
"/Users/Khoi/Dropbox/SOAS/dry_20130715.txt",row.names=F)

### MET SECTION
met <- read.csv ("/Users/Khoi/Dropbox/Isorropia/SOAS/soasmet.csv", 
skip=6, na.strings=c("Null"))
met$Start_LT <- mdy_hm(as.character(met$Start_LT))
met$date <- as.Date (as.character(met$Start_LT))
met$date2 <- ymd(as.character(met$date))
met2<- aggregate (PCP_mm~date, data=met, sum)
met2$date <- ymd(as.character(met2$date))
met2 <- met2[met2$PCP_mm>0,]
met2 <- met2[day(met2$date)!=2,]
met2 <- met2[day(met2$date)!=1,]
met <- met[day(met$date)!=2,]
met <- met[day(met$date)!=1,]
svdmadat$date <- as.Date(svdmadat$timelocal)

par oma=c(0,0,0,1))
plot(svdmadat$timelocal,svdmadat$waterRHadj, col="cornflowerblue", 
pch=19, cex=0.3, ylim=c(0,80),type="n", 
yaxt="n",xaxt="n",xlab="",ylab="")
lines(met2$PCP_mm-met2$date, type="h",col="grey65", lwd=15, 
lend="square")
axis(2, at=seq(0,25,5),las=1,col="grey60")
par(new=T)
plot(svdmadat$timelocal,svdmadat$waterRHadj, col="cornflowerblue", 
pch=19, cex=0.3, ylim=c(-120,30),type="n", 
yaxt="n",xaxt="n",xlab="",ylab="",bty="n")
points(met$Start_LT,met$T_degC, pch=19, cex=0.1, col="tomato4", 
ylim=c(-120,30))
axis(4, at=seq(10,30,10),las=1,col="tomato4")
par(new=T)
plot(svdmadat$timelocal,svdmadat$RHandb, col="darkgreen", pch=19, 
cex=0.4, ylim=c(-200,150),yaxt="n",xaxt="n",xlab="",ylab="",bty="n")
axis(2, at=seq(40,100,20), las=1, col="darkgreen")
par(new=T)
plot(svdmdat$timelocal, svdmadat$dry, col="salmon", pch=19, cex=0.4, ylim=c(-35,50), yaxt="n", xaxt="n", xlab="", ylab="", bty="n")
axis(2, at=seq(0,15,5), las=1, col="salmon")
par(new=T)
plot(svdmdat$timelocal, svdmadat$waterRHadj, col="darkblue", pch=19, cex=0.4, ylim=c(0,80), yaxt="n", xaxt="n", xlab="", ylab="", bty="n")
axis(4, at=seq(0,60,10), las=1, col="darkblue")
axis(4, at=seq(0,60,5), las=1, col="darkblue", labels=F, tcl=-0.3)
par(new=T)
plot(svdmdat$date, svdmadat$waterRHadj, col="cornflowerblue", pch=19, cex=0.3, ylim=c(0,80), type="n", yaxt="n", xaxt="n", xlab="", ylab="", bty="n")
axis.Date(1, at=seq(head(svdmdat$date)[1], tail(svdmdat$date)[1], 5))
axis.Date(1, at=seq(head(svdmdat$date)[1], tail(svdmdat$date)[1], 1), las=2, labels=F, tcl=-0.3)
dev.off()
APPENDIX G: R CODES FOR CHAPTER 3

```r
#### READ IN DATA

data = read.csv("/Users/Khoi/Dropbox/NEWdata-with2012.csv")
org = read.csv("/Users/Khoi/Dropbox/orgcombined-till2012")
co.o3 = read.csv("/Users/khoi/Dropbox/O3.CO_ALLsites.csv")

library(lubridate)
data$date = ymd_hms(as.character(data$date))
org$date = ymd(as.character(org$date))
co.o3$date = ymd_hms(as.character(co.o3$date))

data$year = year(data$date)
data$month = month(data$date)

org$year = year(org$date)
or$month = month(org$date)
or$= or[or$year>2000,]

co.o3$year = year(co.o3$date)
co.o3$month = month(co.o3$date)
co.o3= co.o3[co.o3$year>2000,]

### urban and rural
urban = data[ data$type=="Urban",]
rural = data[ data$type=="Rural" | data$type=="Suburban",]

urban.org = or[ or$type=="Urban",]
rural.org = or[ or$type=="Rural" | or$type=="Suburban",]

urban.co.o3 = co.o3[co.o3$type=="Urban",]
rural.co.o3 = co.o3[co.o3$type=="Rural" | co.o3$type=="Suburban",]

##MAKE PLOTS
library(ggplot2)
library(lattice)
library(gridExtra)

urban.loess = predict(loess(urban$water~urban$year))
rural.loess = predict(loess(rural$water~rural$year))
data.loess = predict(loess(data$water~data$year))

urban.oc.loess = predict(loess(urban.org$OC~urban.org$year))
rural.oc.loess = predict(loess(rural.org$OC~rural.org$year))
data.oc.loess = predict(loess(org$OC~org$year))

write.csv(urban.loess,"/Users/Khoi/Google Drive/Papers/Trends/ Figures/Raw figures/urban.loess.csv", row.names=F)
```
write.csv(rural.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/rural.loess.csv", row.names=F)
write.csv(data.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/data.loess.csv", row.names=F)
write.csv(urban.oc.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/urban.oc.loess.csv", row.names=F)
write.csv(rural.oc.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/rural.oc.loess.csv", row.names=F)
write.csv(data.oc.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/data.oc.loess.csv", row.names=F)

urban.loess=read.csv("/Users/Khoi/Dropbox/Papers/Trends/Figures/Raw figures/urban.loess.csv")
rural.loess=read.csv("/Users/Khoi/Dropbox/Papers/Trends/Figures/Raw figures/rural.loess.csv")
data.loess=read.csv("/Users/Khoi/Dropbox/Papers/Trends/Figures/Raw figures/data.loess.csv")

urban.oc.loess=read.csv("/Users/Khoi/Dropbox/Papers/Trends/Figures/Raw figures/urban.oc.loess.csv")
rural.oc.loess=read.csv("/Users/Khoi/Dropbox/Papers/Trends/Figures/Raw figures/rural.oc.loess.csv")
data.oc.loess=read.csv("/Users/Khoi/Dropbox/Papers/Trends/Figures/Raw figures/data.oc.loess.csv")

par(mar=c(1,0.5,0.5,1), mfrow = c(1,1), oma=c(1,1,1,1))

plot(unique(urban$year),
     unique(urban.loess$x),col="grey65",type="l",lty=2,cex=1.5,xlab="Year ",ylab="",ylim=c(0,80),las=1,lwd=4)
lines(unique(rural$year), unique(rural.loess$x),col="orange",lty=2, cex=1.5,lwd=4)
lines(unique(data$year),
     unique(data.loess$x),col="black",lty=1,cex=3.5,lwd=4)
par(new=T)
plot(unique(org$year),
     unique(data.oc.loess$x),type="l",lty=1,cex=1.5,ylim=c(0,4),las=1,col ="darkgreen",lwd=4,xlab="n",ylab="")
#points(unique(rural.org$year),
     unique(rural.oc.loess$x),col="orange", lty=2,cex=1.25,lwd=4,pch=1,type="b")
#lines(unique(urban.org$year), unique(urban.oc.loess$x), col="grey", lty=2, cex=1.25, lwd=4, type="b", pc h=1)
axis(4, las=1)
mtext(expression(paste("Water \( (\mu g \ m^{-3}) \)\) ), side=2, line=2, cex=1.2)

#plot(org$year, )
par(new=T)
plot(NA, NA, xlim=c(0, 10), ylim=c(0, 10), xlab="", ylab="", axes=F)
leg = c("Urban", "Rural", "Overall Water", "Organics")
col=c("grey65", "orange", "black", "darkgreen")
space=c(2, 4, 6, 8)
for (i in c(1:4))
{text(leg[i], x=space[i]+0.8, y=10, col=col[i])}
dev.off()

######## Six-figure panel

# TEMP
urbanTEMP.loess = predict(loess(urban$TEMP~urban$year))
ruralTEMP.loess = predict(loess(rural$TEMP~rural$year))
dataTEMP.loess = predict(loess(data$TEMP~data$year))
write.csv(urbanTEMP.loess, "/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/urbanTEMP.loess.csv", row.names=F)
write.csv(ruralTEMP.loess, "/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/ruralTEMP.loess.csv", row.names=F)
write.csv(dataTEMP.loess, "/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/dataTEMP.loess.csv", row.names=F)

# RH
urbanRH.loess = predict(loess(urban$RH~urban$year))
ruralRH.loess = predict(loess(rural$RH~rural$year))
dataRH.loess = predict(loess(data$RH~data$year))
write.csv(urbanRH.loess, "/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/urbanRH.loess.csv", row.names=F)
write.csv(ruralRH.loess, "/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/ruralRH.loess.csv", row.names=F)
write.csv(dataRH.loess, "/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/dataRH.loess.csv", row.names=F)

# SO4
urbanSO4.loess = predict(loess(urban$SO4~urban$year))
ruralSO4.loess = predict(loess(rural$SO4~rural$year))
dataSO4.loess = predict(loess(data$SO4~data$year))
write.csv(urbanSO4.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/urbanSO4.lo ess.csv",row.names=F)
write.csv(ruralSO4.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/ruralSO4.loess.csv",row.names=F)
write.csv(dataSO4.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/dataSO4.loess.csv",row.names=F)

#NO3
urbanNO3.loess = predict(loess(urban$NO3~urban$year))
ruralNO3.loess = predict(loess(rural$NO3~rural$year))
dataNO3.loess = predict(loess(data$NO3~data$year))
write.csv(urbanNO3.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/urbanNO3.loess.csv",row.names=F)
write.csv(ruralNO3.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/ruralNO3.loess.csv",row.names=F)
write.csv(dataNO3.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/dataNO3.loess.csv",row.names=F)

#NH3 or NH4
urbanNH3.loess = predict(loess(urban$NH3~urban$year))
ruralNH3.loess = predict(loess(rural$NH3~rural$year))
dataNH3.loess = predict(loess(data$NH3~data$year))
write.csv(urbanNH3.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/urbanNH3.loess.csv",row.names=F)
write.csv(ruralNH3.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/ruralNH3.loess.csv",row.names=F)
write.csv(dataNH3.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/dataNH3.loess.csv",row.names=F)

pdf(file="/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/6fig.pdf", width=6.6,height=8)
par(mfrow=c(3,2),mar=c(5,4,0.5,1),oma=c(0,0.5,0,0))
plot(unique(urban$year), (unique(urban$TEMP.loess)-273.15),type="l",lty=1,lwd=1,cex=1.5,xlab="",ylab="Temperature (deg. C)",las=1,col="grey65",ylim=c(14,22))
lines(unique(rural$year), (unique(rural$TEMP.loess)-273.15),type="l",lty=1, lwd=1,cex=1.5,col="orange")
#lines(data$year, predict(loess(((data$TEMP) -273.15)~rural$year)),type="l",lty=1, lwd=1,cex=1.5,col="black")
plot(unique(urban$year),
unique(urban$RH.loess),type="l",lty=1,lwd=1,cex=1.5,xlab="",ylab="RH"
,las=1,col="grey65",ylim=c(0.6,0.8))
lines(unique(rural$year), unique(ruralRH.loess), type="l", lty=1, lwd=1, cex=1.5, col="orange")

plot(unique(urban$year),
unique(urbanSO4.loess), type="l", lty=1, lwd=1, cex=1.5, xlab="Year", ylab = expression(paste("Sulfate", " (µg "m^-3", " )", sep="")), las=1, col="grey65", ylim=c(0,6))
lines(unique(rural$year), unique(ruralSO4.loess), type="l", lty=1, lwd=1, cex=1.5, col="orange")

plot(unique(urban$year),
unique(urbanNO3.loess), type="l", lty=1, lwd=1, cex=1.5, xlab="Year", ylab = expression(paste("Nitrate", " (µg "m^-3", " )", sep="")), las=1, col="grey65", ylim=c(0,2))
lines(unique(rural$year), unique(ruralNO3.loess), type="l", lty=1, lwd=1, cex=1.5, col="orange")

plot(unique(urban$year),
unique(urbanNH3.loess), type="l", lty=1, lwd=1, cex=1.5, xlab="Year", ylab = expression(paste("Ammonium", " (µg "m^-3", " )", sep="")), las=1, col="grey65", ylim=c(0,3))
lines(unique(rural$year), unique(ruralNO3.loess), type="l", lty=1, lwd=1, cex=1.5, col="orange")

plot(unique(urban.org$year),
unique(urban.oc.loess), type="l", lty=1, lwd=1, cex=1.5, xlab="Year", ylab = expression(paste("OC", " (µg "m^-3", " )", sep="")), las=1, col="grey65", ylim=c(0,5))
lines(unique(rural.org$year), unique(rural.oc.loess), type="l", lty=1, lwd=1, cex=1.5, col="orange")

dev.off()

# orig data - no screening
RH80 <- data[data$RH>.85,]
RH80 <- RH80[RH80 $RH>.80,]

RH85 <- data[data$RH>.90,]
RH85 <- RH85[RH85 $RH>.90,]

# data screened for fires
co.o3_700 = co.o3[co.o3$CO < 700,]
co.o3_700 = co.o3_700[!is.na(co.o3_700$date),]

fire = data[paste(data$date, data$site) %in% paste(co.o3_700$date, co.o3_700$site),]
fire=fire[!fire$site="JST",]
RH801 <- fire[fire$RH>.85,]
RH801 <- RH801[RH801 $RH>=.80,]
RH851 <- fire[fire$RH<.90,]
RH851 <- RH851[RH851 $RH>=.85,]
RH901 <- fire[fire$RH<.95,]
RH901 <- RH901[RH901 $RH>=.90,]

months <-
c("January","February","March","April","May","June","July","August",
"September","October","November","December")

pdf(file="/Users/Khoi/Dropbox/Papers/Trends/ Figures/ Raw figures/trends-noJST.pdf", width=6.6, height=7.7)

for (n in c(1:12)){
  y= RH80$water[RH80 $month==n]
x= RH80$year[RH80 $month==n]

  plot (x,y, ylim=c(0,lim),
xlab="",ylab="",cex.main=1,pch="",cex=0,las=1,labels=F)
  mtext(months[n])
  loess_fit <- loess(y ~ x)
  lines(x, predict(loess_fit), col = "red", lwd=2)
  lines(RH801$year[RH801 $month==n], predict(loess(RH801$water[RH801 $month==n])~RH801$year[RH801 $month==n])),col="red",lty=2)

  par(new=T)
  y= RH85 $water[RH85 $month==n]
x= RH85 $year[RH85 $month==n]

  plot (x,y, ylim=c(0,lim),
xlab="",ylab="",bty="n",cex.main=1,pch=19,cex=0,lab
els=F)
  loess_fit <- loess(y ~ x)
  lines(x, predict(loess_fit), col = "green", lwd=2)
  lines(RH851$year[RH851 $month==n], predict(loess(RH851$water[RH851 $month==n])~RH851$year[RH851 $month==n])),col="green",lty=2)

  par(new=T)
  y= RH90 $water[RH90 $month==n]
x= RH90 $year[RH90 $month==n]

  plot (x,y, ylim=c(0,lim),
xlab="",ylab="",bty="n",cex.main=1,pch=19,cex=0,lab
els=F)
  loess_fit <- loess(y ~ x)
  lines(x, predict(loess_fit), col = "blue", lwd=2)
  lines(RH901$year[RH901 $month==n], predict(loess(RH901$water[RH901 $month==n])~RH901$year[RH901 $month==n])),col="blue",lty=2)
}

par(new=T, mfrow=c(1,1), oma=c(0,0,0,0))
plot(NA,NA, xlim=c(0,10), ylim=c(0,10), xlab="",ylab="",axes=F)
mtext(expression(paste("Water ("mu g " , m^-3, "), sep="\n"))), side=2, line=-1,cex=1.2)
mtext("Year", side=1, line=1.7,cex=1.2)
leg = c("80 ,â§ RH < 85", "85 ,â§ RH < 90", "90 ,â§ RH < 95")
col=c("red","green","blue")
space=c(2,4,6)
for (i in c(1:3))
{text(leg[i], x=space[i]+1.5, y=10,col=col[i]) }
dev.off()

######## monthly trends sulfate and ammonium
months <-
c("January","February","March","April","May","June","July","August", "September","October","November","December")
lim=8
par(mar=c(2.5,0.5,0.5,1), mfrow = c(4,3), oma=c(2,5,5,1))
for (n in c(1:12))
{y= RH80$SO4[RH80 $month==n]
x= RH80$year[RH80 $month==n]
y1= RH80$NH3[RH80 $month==n]
x1= RH80$year[RH80 $month==n]
plot (x,y, ylim=c(0,lim),
xlab="",ylab="",cex.main=1,pch="",cex=0,las=1,labels=F)
mtext(months[n])
loess_fit <- loess(y ~ x)
lines(x, predict(loess_fit), col = "red",lwd=2)
# axis(2)
par(new=T)
loess_fit1=loess(y1~x1)
plot(x1, predict(loess_fit1), col = "red",lwd=2,xaxt="n",yaxt="n",ylim=c(0,lim),lty=2,type="l")
# axis(4)
par(new=T)
y= RH85 $SO4[RH85 $month==n]
x= RH85 $year[RH85 $month==n]
y1= RH85$NH3[RH85 $month==n]
x1= RH85$year[RH85 $month==n]
plot (x,y, ylim=c(0,lim),
yaxt="n",xaxt="n",xlab="",ylab="",bty="n",cex.main=1,pch=19,cex=0,labels=F)
loess_fit <- loess(y ~ x)
lines(x, predict(loess_fit), col = "green",lwd=2)
par(new=T)
loess_fit1=loess(y1~x1)
plot(x1, predict(loess_fit1), col = "green",lwd=2,xaxt="n",yaxt="n",ylim=c(0,lim),lty=2,type="l")
# axis(4)
par(new=T)
y = RH90 $SO_4[^{RH90}]$ month==n
x = RH90 $year[^{RH90}]$ month==n
y1 = RH90$NH_3[^{RH90}]$ month==n
x1 = RH90$year[^{RH90}]$ month==n
plot (x,y, ylim=c(0,lim),
yaxt="n",xaxt="n",xlab="",ylab="",bty="n",cex.main=1,pch=19,cex=0,labels=F)
loess_fit <- loess(y ~ x)
lines(x, predict(loess_fit), col = "blue",lwd=2)
par(new=T)
loess_fit1=loess(y1~x1)
plot(x1, predict(loess_fit1), col = "blue",lwd=2,xaxt="n",yaxt="n",ylim=c(0,lim),lty=2,type="l")
# axis(4)
par(new=T, mfrow=c(1,1),oma=c(0,0,0,0))
plot(NA,NA, xlim=c(0,10),ylim=c(0,10), xlab="",ylab="",axes=F)
mtext(expression(paste(NO[3], " or ", NH[4]," (µg ",m^-3, ")"),
sep=" ")) , side=2, line=-1,cex=1.2)
mtext("Year", side=1, line=1.7,cex=1.2)
leg = c("80 '§ RH < 85","85 '§ RH < 90", "90 '§ RH < 95")
col=c("red","green","blue")
space=c(2,4,6)
for (i in c(1:3))
{text(leg[i], x=space[i]+1.5, y=10,col=col[i])}
dev.off()

######### pie chart for overall, august, january
######### pie chart for overall, august, january
org$OM = 1.4*org$OC
orgavg <- aggregate (cbind(OC, OM) ~ year, data=org, mean)
dataavg <- aggregate (cbind(SO4,NO3, NH3,water,TEMP,RH) ~ year,
data=data, mean)
urbanavg <- aggregate (cbind(SO4,NO3, NH3,water,TEMP,RH) ~ year,
data=urban, mean)
ruralavg <- aggregate (cbind(SO4,NO3, NH3,water,TEMP,RH) ~ year,
data=rural, mean)
urban.org.avg = aggregate(OC~year, data=urban.org,mean)
rural.org.avg = aggregate(OC~year, data=rural.org,mean)
write.csv(urban.org.avg,
"/Users/khoi/desktop/urban.org.avg.csv",row.names=F)
write.csv(rural.org.avg,
"/Users/khoi/desktop/rural.org.avg.csv",row.names=F)
dataavg$frac =
dataavg$water/(dataavg$SO4+dataavg$NO3+dataavg$NH3+dataavg$water+org
avg$OM)
(dataavg$frac[1]-dataavg$frac[12])/dataavg$frac[1]
(dataavg$water[1]-dataavg$water[12])/dataavg$water[1]
(urbanavg$water[1]-urbanavg$water[12])/urbanavg$water[1]
(ruralavg$water[1]-ruralavg$water[12])/ruralavg$water[1]

library(mapplots)
data.aug = data[data$month==8,]
org.aug = org[month(org$date)==8,]
org.aug.orgavg <- aggregate (cbind(OC, OM) ~ year, data=org.aug, mean)
data.aug.avg <- aggregate (cbind(SO4,NO3, NH3,water,TEMP,RH) ~ year, data=data.aug, mean)

data.jan = data[data$month==1,]
org.jan = org[month(org$date)==1,]
org.jan.orgavg <- aggregate (cbind(OC, OM) ~ year, data=org.jan, mean)
data.jan.avg <- aggregate (cbind(SO4,NO3, NH3,water,TEMP,RH) ~ year, data=data.jan, mean)
org.jan.orgavg = org.jan.orgavg[org.jan.orgavg$year!=2001,]
at = c(2001:2012)

pdf(file="~/Users/Khoi/Google Drive/Papers/Trends/ Figures/Raw figures/piechart.pdf", width=6.6,height=6.6)

par(mfrow=c(3,1), mar=c(0,5,1,1), oma=c(5,0,2,0), tcl=-0.5)

plot(dataavg$water~dataavg$year,las=1,xaxt="n",ylab="",xlab="",pch=" ",ylim=c(-10,70),type="l",lwd=2,lty=1,cex.axis=2)
axis(1, at=at,labels=F)
a=1
for (a in c(1:12)){
  add.pie(z=c(dataavg[a,2],dataavg[a,3],dataavg[a,4],orgavg[a,3],dataavg[a,5]), x=dataavg$year[a], y=dataavg$water[a], radius=8.5, labels="",col=c("red","blue","orange","green","lightblue"),border=T)
  text(x=2009,y=65, "All Months",cex=1.8,adj = c(0,0))
}

plot(data.aug.avg$water~data.aug.avg$year,las=1,xaxt="n",ylab="",xlab="",pch=" ",ylim=c(-10,150),type="l",lwd=2,lty=1,cex.axis=2)
axis(1, at=at,labels=F)

mtext(expression(paste("Water (\(\mu g\) \(m^\{-3\}\), \"", sep=" "")), side=2, line=3,cex=1.5)

text(x=2009,y=134, "August",cex=1.8,adj = c(0,0))
a=1
for (a in c(1:12)){
  add.pie(z=c(data.aug.avg[a,2],data.aug.avg[a,3],data.aug.avg[a,4],org.aug.orgavg[a,3],data.aug.avg[a,5]), x=data.aug.avg$year[a], y=data.aug.avg$water[a], radius=17, labels="",col=c("red","blue","orange","green","lightblue"),border=T)
  text(x=2009,y=65, "All Months",cex=1.8,adj = c(0,0))
}
### Jan

```r
plot(c(NA, data.jan.avg$water) ~ c(2001, data.jan.avg$year), las=1, xaxt="n", ylab="", xlab="", pch="", ylim=c(-10, 70), type="l", lwd=2, lty=1, cex.axis=2)
axis(1, at=at, cex.axis=2)
mtext("Year", side=1, line=2.7, cex=1.5)
text(x=2009, y=65, "January", cex=1.8, adj = c(0, 0))
```

```r
a=1
for (a in c(1:11)){
  add.pie(z=c(data.jan.avg[a,2], data.jan.avg[a,3], data.jan.avg[a,4], org.jan.orgavg[a,3], data.jan.avg[a,5]), x=data.jan.avg$year[a],
  y=data.jan.avg$water[a], radius=8.5, labels="", col=c("red", "blue", "orange", "green", "lightblue"), border=T)
}
```

```r
par(new=T, mfrow=c(1,1), mar=c(0,0,0,0), oma=c(0,0,2,0), tcl=-0.5)
plot(c(NA, data.jan.avg$water) ~ c(2001, data.jan.avg$year), las=1, xaxt="n", bty="n", ylab="", xlab="", ylim=c(0, 70), pch="", lwd=2, yaxt="n")
legend("top", inset=0,
  c("Sulfate", "Nitrate", "Ammonium", "OM", "Water"),
  pch=c(16), col=c("red", "blue", "orange", "green", "lightblue"),
  cex=1.2, box.lwd = 0, box.col = "transparent", bg = "transparent", horiz=T)
```

```r
dev.off()
```

##### stacked bar plot

```r
# all months
orgavg <- aggregate (cbind(OC, OM) ~ year, data=org, mean)
dataavg <- aggregate (cbind(SO4, NO3, NH3, water, TEMP, RH) ~ year, data=data, mean)
dataavg$OM=orgavg$OM
dataavg2=data.frame(water=dataavg$water, OM=dataavg$OM, NH3=dataavg$NH3, NO3=dataavg$NO3, SO4=dataavg$SO4)
dataavg3=t(dataavg2)
years=as.character(c(2001:2012))
colnames(dataavg3)=years
all=as.matrix(dataavg3)
```

```r
# August
orgavg=org.aug.orgavg
dataavg=data.aug.avg
dataavg$OM=orgavg$OM
dataavg2=data.frame(water=dataavg$water, OM=dataavg$OM, NH3=dataavg$NH3, NO3=dataavg$NO3, SO4=dataavg$SO4)
dataavg3=t(dataavg2)
```
colnames(dataavg3)=years
aug=as.matrix(dataavg3)
## january
orgavg=org.jan.orgavg
dataavg=data.jan.avg
dataavg$OM=orgavg$OM
dataavg2=data.frame(water=c(NA, dataavg$water),OM=c(NA, dataavg$OM),
NH3=c(NA,dataavg$NH3), NO3=c(NA,dataavg$NO3),SO4=c(NA,dataavg$SO4))
dataavg3=t(dataavg2)
colnames(dataavg3)=years
jan=as.matrix(dataavg3)
## plot
pdf(file="/Users/Khoi/Dropbox/Papers/Trends/Figures/Raw
figures/stackedbar.pdf", width=6.6,height=7.7)
par(mfrow=c(3,1), mar=c(2,5,0,0), oma=c(2.5,1,2,0),tcl=-0.5)
#barplot(all,legend=rownames(all),col=c("lightblue","green","orange",
"blue","red"),las=1,cex.axis=1.24,ylim=c(0,70),xaxt="n")
barplot(all,col=c("lightblue","green","orange","blue","red"),las=1,cex.axis=1.24,ylim=c(0,70),xaxt="n")
legend("top", inset=0,
c("Sulfate","Nitrate","Ammonium","OM","Water"),
pch=c(15),col=c("red","blue","orange","green","lightblue"),
cex=1.25,box.lwd = 0,box.col = "transparent",bg =
"transparent",horiz=T)
text(x=12.5,y=56, "All Months",cex=1.5,adj = c(0,0))
barplot(jan,
col=c("lightblue","green","orange","blue","red"),las=1,cex.axis=1.24
,ylim=c(0,70),xaxt='n')
mtext(expression(paste("Average Mass Concentration (\(\mu g \times m^{-3}\))", sep=" ")
), side=2, line=3,cex=1.25)
text(x=12.5,y=56, "January",cex=1.5,adj = c(0,0))
barplot(aug,col=c("lightblue","green","orange","blue","red"),las=1,cex.axis=1.24,
ylim=c(0,150),xlab="Year")
text(x=12.5,y=137, "August",cex=1.5,adj = c(0,0))
dev.off()

########### Water from OC estimates
sites <- read.csv("/Users/khoi/Google Drive/TTN/IGAC
poster/searchsites.csv")
org$OM = 1.4*org$OC
data$date2 <- floor_date(as.POSIXct(data$date), "day")
data24avg <- aggregate (cbind(SO4,NO3, NH3,water,TEMP,RH)~date2+site,data=data, mean)
org24avg <- aggregate (cbind(OC, OM)~date+ site,data=org, mean)
org24 <- org24avg[paste(as.character(org24avg$date), org24avg$site),
%in% paste(as.character(data24avg$date), data24avg$site),]
avg24 <- data24avg[paste(as.character(data24avg$date2),
data24avg$site) %in% paste(as.character(org24$date),org24$site),]

avg24 <- avg24[order(avg24$date2,avg24$site),]
org24 <- org24[order(org24$date,org24$site),]

avg24$OC = org24$OC
avg24$OM = org24$OM

seq(0.01,0.21,0.04)
#0.01 0.05 0.09 0.13 0.17 0.21
avg24$excess0.01 <- 0.01*avg24$OM/1.4*avg24$RH/(1-avg24$RH)
avg24$excess0.05 <- 0.05*avg24$OM/1.4*avg24$RH/(1-avg24$RH)
avg24$excess0.09 <- 0.09*avg24$OM/1.4*avg24$RH/(1-avg24$RH)
avg24$excess0.13 <- 0.13*avg24$OM/1.4*avg24$RH/(1-avg24$RH)
avg24$excess0.17 <- 0.17*avg24$OM/1.4*avg24$RH/(1-avg24$RH)
avg24$excess0.21 <- 0.21*avg24$OM/1.4*avg24$RH/(1-avg24$RH)

avg24$year <- year(avg24$date2)
avg24 = avg24[avg24$excess0.21!=max(avg24$excess0.21),]

for (i in c(1:8)){
  avg24$type[avg24$site%in%sites$name[i]] <- as.vector(sites$type[i])
}

rural24 <- avg24[avg24$type=="Rural" | avg24$type=="Suburban",]
urban24 <- avg24[avg24$type=="Urban",]

newdata <- aggregate
  (cbind(water,excess0.01,excess0.05,excess0.09,excess0.13,excess0.17,
excess0.21)~year, data=avg24,mean)
rural24avg <- aggregate
  (cbind(water,excess0.01,excess0.05,excess0.09,excess0.13,excess0.17,
excess0.21)~year, data=rural24,mean)
urban24avg <- aggregate
  (cbind(water,excess0.01,excess0.05,excess0.09,excess0.13,excess0.17,
excess0.21)~year, data=urban24,mean)

pdf(file="/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/LWC-OC.pdf", width=6.6,height=6)
par(mfrow=c(1,1),mar=c(4,4,1,1),oma=c(0,0.5,0,0))
plot(newdata$year,newdata$water,type="l",lwd=5,ylim=c(0,80),las=1,ylab=
"")
mtext(expression(paste("Water (\(\mu g \frac{m^3}{}\)), \text{sep}=" "))),
side=2,line=2.5,cex=1.5)
mtext("Year", side=1,line=2.5,cex=1.5)
for (i in 3:length(newdata)){
  lines(newdata$year,newdata$water+newdata[,i],col="black",lty=2,lwd=5
0/(i^2))
}
lines(rural24avg$year,rural24avg$water,type="l",lwd=5,col="orange")
for (i in 3:length(rural24avg)){
  lines(rural24avg$year, rural24avg$water+rural24avg[,i], col="orange", lty=2, lwd=50/(i^2))
}
lines(urban24avg$year, urban24avg$water, type="l", lwd=5, col="grey65")
for (i in 3:length(urban24avg)){
  lines(urban24avg$year, urban24avg$water+urban24avg[,i], col="grey65", lty=2, lwd=50/(i^2))
}
legend("topright", inset=0, c("Overall", "Urban", "Rural"), pch=c("-"), col=c("black", "grey65", "orange"), cex=1, box.lwd = 0, box.col = "transparent", bg = "transparent", lwd=3, horiz=F)
dev.off()

# percent difference
newdata[,9:14]=NA
for (i in 3:8){
  newdata[, (6+i)]=newdata[,i]/newdata[,2]
}
rural24avg[,9:14]=NA
for (i in 3:8){
  rural24avg[, (6+i)]=rural24avg[,i]/rural24avg[,2]
}
urban24avg[,9:14]=NA
for (i in 3:8){
  urban24avg[, (6+i)]=urban24avg[,i]/urban24avg[,2]
}

mean(c(newdata[,9],newdata[,10],newdata[,11],newdata[,12],newdata[,13],newdata[,14]))
mean(c(rural24avg[,9],rural24avg[,10],rural24avg[,11],rural24avg[,12],rural24avg[,13],rural24avg[,14]))
mean(c(urban24avg[,9],urban24avg[,10],urban24avg[,11],urban24avg[,12],urban24avg[,13],urban24avg[,14]))

######## Isoprene = see isoprene.R

######## Ozone

o3 = co.o3[!is.na(co.o3$O3),]
o3rural = o3[o3$type=="Rural" | o3$type=="Suburan",]
o3urban = o3[o3$type=="Urban",]
urban.o3.loess = predict(loess(o3urban$O3~o3urban$year))
rural.o3.loess = predict(loess(o3rural$O3~o3rural$year))
data.o3.loess = predict(loess(o3$O3~o3$year))

write.csv(urban.o3.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/urban.o3.loess.csv", row.names=F)
write.csv(rural.o3.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/rural.o3.loess.csv", row.names=F)
write.csv(data.o3.loess,"/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/data.o3.loess.csv", row.names=F)

plot(unique(o3urban$year),
one(unique(urban.o3.loess),type="l",lty=2,cex=1.5,xlab="Year",ylab="Ozone (ppb)",las=1,col="grey65",ylim=c(0,70),lwd=2)
lines(unique(o3rural$year), unique(rural.o3.loess),type="l",lty=2, cex=1.5,col="orange",lwd=2)
lines(unique(o3rural$year), unique(data.o3.loess),type="l",lty=1, cex=1.5,col="black",lwd=2)
dev.off()

######## ######## ######## ######## ########
######## pH Index
######## ######## ######## ######## ########
data$Hions = 1000* (2*data$SO4/96 + data$NO3/62 - data$NH3/18)
data$pH = -1*log((data$Hions/data$water), base=10)
data$pHisorropia = -1*log((data$H/data$water), base=10)
data$pHisorropia2 = -1*log((data$H), base=10)
dataph <- data[data$pH>(-100)&data$pH<(100)&data$pHisorropia<(100),]
dataph <- dataph[!is.na(dataph$pH),]

ph = aggregate (pH~year, data=dataph, mean)
summary(lm(pH~year,data=ph))

pdf(file="/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/pHindex.pdf", width=6.6,height=6)
ggplot(dataph, aes(year, pH))+ stat_smooth()+theme_bw()+coord_cartesian(xlim=c(2001, 2012),ylim=c(-0.7,0))+theme(text = element_text(size = 24))+xlab("Year")+ylab("pH Index - Ions")
#ph2= ggplot(dataph, aes(year, pHisorropia))+
stat_smooth()+theme_bw()+coord_cartesian(xlim=c(2001, 2012))+theme(text = element_text(size = 24))+xlab("Year")+ylab("pH Index - Isorropia")
dev.off()}
OC to SO4 ratios

### fires screened

```r
fire$day = floor_date(as.POSIXct(fire$date), "day")
fire.agg = aggregate(cbind(SO4, NO3)~site+day, data=fire, mean)
org.agg = aggregate(OC~site+date, data=org, mean)

fire.new = fire.agg[paste(fire.agg$day,
fire.agg$site)%in%paste(org$date, org$site),]
oc.new = org.agg[paste(org.agg$date,
org.agg$site)%in%paste(fire.new$day, fire.new$site),]

fire.new <- fire.new[order(fire.new$day,fire.new$site),]
oc.new <- oc.new[order(oc.new$date,oc.new$site),]

fire.new$OM = 1.4*oc.new$OC
fire.new$OC = oc.new$OC
#fire.new$date = as.Date(fire.new$day, "%y-%m-%d")
fire.new$OCSO4ratio=fire.new$OC/fire.new$SO4
fire.new$OCSO4NO3ratio=fire.new$OC/(fire.new$SO4+fire.new$NO3)
fire.new$year = year(fire.new$day)
```

### fires not screened

```r
data$day = floor_date(as.POSIXct(data$date), "day")
data.agg = aggregate(cbind(SO4, NO3)~site+day, data=data, mean)
or agg = aggregate(cbind(OC,EC)~site+date, data=org, mean)

data.new = data.agg[paste(data.agg$day,
data.agg$site)%in%paste(or g$date, org$site),]
oc.new2 = org.agg[paste(or g.agg$date,
or g.agg$site)%in%paste(data.new$day, data.new$site),]

data.new <- data.new[order(data.new$day, data.new$site),]
oc.new2 <- oc.new2[order(oc.new2$date,oc.new2$site),]

data.new$OM = 1.4*oc.new2$OC
data.new$OC = oc.new2$OC
data.new$OCSO4ratio=data.new$OC/data.new$SO4
data.new$OCSO4NO3ratio=data.new$OC/(data.new$SO4+data.new$NO3)
data.new$year = year(data.new$day)
```

```r
p1 = ggplot(fire.new, aes(year, OCSO4ratio)) +
  stat_smooth() + theme_bw() + coord_cartesian(xlim=c(2001, 2012), ylim=c(0,2)) + theme(text = element_text(size = 24)) + xlab("Year") + ylab("OC/SO4 - screened")
p2 = ggplot(fire.new, aes(year, OCSO4NO3ratio)) +
  stat_smooth() + theme_bw() + coord_cartesian(xlim=c(2001, 2012), ylim=c(0,2)) + theme(text = element_text(size = 24)) + xlab("Year") + ylab("OC/(SO4+NO3) - screened")
p3 = ggplot(data.new, aes(year, OCSO4ratio)) +
  stat_smooth() + theme_bw() + coord_cartesian(xlim=c(2001,
```

library(gridExtra)
pdf(file="/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/OCratios.pdf", width=6.6,height=6)
grid.arrange(p1,p2,p3,p4,ncol = 2)
dev.off()

## secondary SOC calculations

#AE5 = read.csv("/Users/khoi/Google Drive/SEARCH/search_2003_AE5_screened.csv")
searchAE5file <- "/Users/khoi/Documents/SOC/search_2003_AE5_new.csv"

AE5$ratio = AE5$AORGPA/AE5$modEC
AE5$date = mdy_hm(as.character(AE5$TimeOn))

# data.AE5 = data.agg[paste(month(data.agg$day), day(data.agg$day), data.agg$site)]%in%paste(month(AE5$date), day(AE5$date), AE5$site),]
# AE5.new = AE5[paste(month(AE5$date), day(AE5$date), AE5$site)]%in%paste(month(data.agg$day), day(data.agg$day), data.agg$site),]

AE5$month = month(AE5$date)
AE5.bymonth = aggregate(ratio-month+siteID,data=AE5,mean)
#AE5.bhm =AE5.bymonth[AE5.bymonth$siteID=="BHM",]

data.new <- data.new[order(data.new$day, data.new$site),]
oc.new2 <- oc.new2[order(oc.new2$day, oc.new2$site),]

data.new$EC = oc.new2$EC
#data.new2 = data.new[data.new$site=="BHM",]
#data.new2$month = month(data.new2$day)
data.new$month = month(data.new$day)

## BHM ONLY

# for (i in c(1:12)){

}
# data.new2$ratio[data.new2$month%in%AE5.bhm$month[i]] =
# as.vector(AE5.bhm$ratio[i])
#
for (i in c(1:96)){
  data.new$ratio[(data.new$month%in%AE5.bymonth$month[i])&data.new$sit
  e%in%AE5.bymonth$siteID[i]] = as.vector(AE5.bymonth$ratio[i])
  }

#data.new2$OCsec = data.new2$OC-min(c(data.new2$OC,
data.new2$ratio*data.new2$EC))
data.new$multiplier = data.new$ratio*data.new$EC

for (i in c(1:nrow(data.new))){
  data.new$min[i] = min(c(data.new$OC[i], data.new$multiplier[i]))
  }

data.new$OCsec = data.new$OC-data.new$min

data.new$POC = data.new$OC-data.new$OCsec

#boxplot(data.new2$OCsec~data.new2$year,varwidth=T)
boxplot(data.new$OCsec~data.new$year,varwidth=T)

### LOESS - S7

data.oc.loess=read.csv("/Users/Khoi/Google
Drive/Papers/Trends/ Figures/Raw figures/data.oc.loess.csv")
data.soc.loess = predict(loess(data.new$OCsec~data.new$year))
write.csv(data.soc.loess,"/Users/Khoi/Google
Drive/Papers/Trends/ Figures/Raw figures/data.soc.loess.csv",
row.names=F)

pdf(file="/Users/Khoi/Google Drive/Papers/Trends/ Figures/Raw
figures/SOC.OC.loess.pdf", width=6.6,height=6)

plot(unique(org$year),
as.vector(unique(data.oc.loess$x)),type="l",lty=1,lwd=2,cex=1.5,ylab ="",
las=1,col="darkgreen",ylim=c(0,4),xlab="Year")
lines(unique(data.new$year), unique(data.soc.loess),type="l",lty=1,
lwd=2,cex=1.5,col="red")
mtext(expression(paste("OC or SOC (\mu g \ m^{-3}, "", sep="" ")),
side=2, line=2,cex=1.35)
text(2011,3, "OC",col="darkgreen")
text(2011,1.75, "SOC",col="red")

#axis(at=c(2001:2012),side=1)

dev.off()

pdf(file="/Users/Khoi/Google Drive/Papers/Trends/ Figures/Raw
figures/OCsec.pdf-BHM", width=6.6,height=6)
ggplot(data.new2, aes(year, OCsec))+
stat_smooth()+theme_bw()+coord_cartesian(xlim=c(2001,
2012))+theme(text = element_text(size =
24))+xlab("Year")+ylab("Secondary OC at BHM")
dev.off()
pdf(file="/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/OCsec.pdf", width=6.6, height=6)
ggplot(data.new, aes(year, OCsec)) +
  stat_smooth() + theme_bw() + coord_cartesian(xlim=c(2001, 2012)) +
  theme(text = element_text(size = 24)) + xlab("Year") + ylab("SOC")
dev.off()

####
d = data.new[, c("year", "OC", "OCsec")]
d2 = reshape(d, varying=c("OC", "OCsec"),
  v.names="concentration", timevar="compound", times=c("OC", "OCsec"),
  direction="long")
d2$color = "red"
d2$color[which(d2$compound=="OC")]="seagreen"
 #d2$color[which(d2$compound=="POC")]="black"

pdf(file="/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/OCsec.and.OC.pdf", width=6.6, height=6)
ggplot()+
  stat_smooth(data=d2, aes(x=year, y=concentration, group=compound, color=color)) +
  theme_bw() + coord_cartesian(xlim=c(2001, 2012), ylim=c(0, 5)) +
  theme(text = element_text(size = 24)) + xlab("Year") + ylab("SOC") +
  theme(legend.position = "none")
dev.off()

## org by month
pdf(file="/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/OCbymonth.pdf", width=6.6, height=7)
lm=10
avg24$month=month(avg24$date2)
par(mar=c(2.5, 2, 1, 2), mfrow = c(4, 3), oma=c(2, 5, 5, 1))
for (n in c(1:12)){
  y = avg24$OC[avg24$month==n]
  x = avg24$year[avg24$month==n]
  y1 = avg24$water[avg24$month==n]
  x1 = avg24$year[avg24$month==n]

  plot(x,y, ylim=c(0, lm),
       xlab="", ylab="", cex.main=1, pch=" " , cex=0, las=1)
  mtext(months[n])
  loess_fit <- loess(y ~ x)
  lines(x, predict(loess_fit), col = "red")

  par(new=T)
  plot(x1, y1, ylim=c(0, 70),
       xlab="", ylab="", cex.main=1, pch=" " , cex=0, las=1, xaxt="n", yaxt="n")
  axis(4, las=1, col="blue", col.axis="blue")
  loess_fit2 <- loess(y1 ~ x1)
  lines(x1, predict(loess_fit2), col = "blue", lty=2)
}
dev.off()

##SOC by month for BHM
```r
pdf(file="/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/SOCbymonth.pdf", width=6.6,height=7)
lim=10
avg24$month=month(avg24$date2)
par(mar=c(2.5,2,1,2), mfrow = c(4,3), oma=c(2,5,5,1))
for (n in c(1:12)){
  y= data.new2$OCsec[data.new2 $month==n]
x= data.new2$year[data.new2 $month==n]

  y1 = avg24$water[avg24$month==n]
x1 = avg24$year[avg24$month==n]

  plot (x,y, ylim=c(0,lim),
xlab="",ylab="",cex.main=1,pch="",cex=0,las=1)
mtext(months[n])
loess_fit <- loess(y ~ x)
lines(x, predict(loess_fit), col = "red")

  par(new=T)
  plot (x1,y1, ylim=c(0,lim),
xlab="",ylab="",cex.main=1,pch="",cex=0,las=1,xaxt="n",yaxt="n")
  axis(4,las=1,col="blue",col.axis="blue")
  loess_fit2 <- loess(y1 ~ x1)
  lines(x1, predict(loess_fit2), col = "blue",lty=2)
}
dev.off()

##SOC and OC by month for all - S8
pdf(file="/Users/Khoi/Google Drive/Papers/Trends/Figures/Raw figures/SOCbymonth-ALL.pdf", width=6.6,height=7)
lim=5
avg24$month=month(avg24$date2)
par(mar=c(2,1,1,1), mfrow = c(4,3), oma=c(2,4,5,1))
for (n in c(1:12)){
  y= data.new$OCsec[data.new $month==n]
x= data.new$year[data.new $month==n]

  y1 = avg24$water[avg24$month==n]
x1 = avg24$year[avg24$month==n]

  plot (x,y, ylim=c(0,lim),
xlab="",ylab="",cex.main=1,pch="",cex=0,las=1,labels=F)
mtext(months[n])
loess_fit <- loess(y ~ x)
lines(x, predict(loess_fit), col = "red")
lines(x, predict(loess(data.new$OC[data.new $month==n]~x)), col = "seagreen")

  par(new=T)
  plot (x1,y1, ylim=c(0,70),
xlab="",ylab="",cex.main=1,pch="",cex=0,las=1,xaxt="n",yaxt="n",labels=F)
  axis(4,las=1,col="blue",col.axis="blue",labels=F)
  loess_fit2 <- loess(y1 ~ x1)
  lines(x1, predict(loess_fit2), col = "blue",lty=2)
}
dev.off()

## Table of correlation

##### parameter vs. time

#overall water, rural water, urban water
summary(lm(dataavg$water~dataavg$year))
summary(lm(ruralavg$water~ruralavg$year))
summary(lm(urbanavg$water~urbanavg$year))

#water fraction
dataavg$waterfraction = dataavg$water/(sum(dataavg[,2:5])+orgavg$OM)
summary(lm(dataavg$waterfraction~dataavg$year))

#OC
summary(lm(orgavg$OM~orgavg$year))

#SO4, NO3, NH4
summary(lm(dataavg$SO4~dataavg$year))
summary(lm(dataavg$NO3~dataavg$year))
summary(lm(dataavg$NH3~dataavg$year))

##### parameter vs. water

#OC
ruralorgavg <- aggregate (cbind(OC) ~ year, data=rural.org, mean)
urbanorgavg <- aggregate (cbind(OC) ~ year, data=urban.org, mean)

summary(lm(ruralavg$water~ruralorgavg$OC))
summary(lm(urbanavg$water~urbanorgavg$OC))
summary(lm(dataavg$water~orgavg$OM))
summary(lm(unique(data.loess)~unique(data.oc.loess)))

#SO4, NO3, NH4
summary(lm(dataavg$water~dataavg$SO4))
summary(lm(dataavg$water~dataavg$NO3))
summary(lm(dataavg$water~dataavg$NH3))

# met
summary(lm(dataavg$water~dataavg$TEMP))
summary(lm(dataavg$water~dataavg$RH))

#####
### 8 figure panel

#### site.avg=aggregate(water-year+site,data=data,FUN=mean)
orgsite.avg=aggregate(OC-year+site,data=org,FUN=mean)
sites=c("CTR","BHM","YRK","JST","OAK","GFP","OLF","PNS")

pdf(file="~/Users/Khoi/Dropbox/Papers/Trends/Figures/Raw figures/8-figALLsites.pdf", width=6.6, height=6)
par(mar=c(2.5,2,1,2), mfrow = c(4,2), oma=c(2,5,5,1))

for(i in 1:8){
  loess =
predict(loess(data$water[data$site==sites[i]]~data$year[data$site==sites[i]]))
  loessorg =
predict(loess(org$OC[org$site==sites[i]]~org$year[org$site==sites[i]]))
  plot(unique(data$year[data$site==sites[i]]),unique(loess),type="l",las=1,col="black",ylim=c(0,80),xlim=c(2001,2012))
    text(2011.5,76, sites[i])
  par(new=T)
  plot(unique(org$year[org$site==sites[i]]),unique(loessorg),type="l",las=1,col="darkgreen",ylim=c(0,6),xlim=c(2001,2012),yaxt='n',xaxt='n')
    axis(4,las=1)
}
dev.off()

### seasonal means table

gsom = 1.4*org$OC
orgavg <- aggregate (cbind(OC, OM) ~ year, data=org, mean)
data.avg=aggregate(water~year+site,data=data,FUN=mean)
olddata= data
data=olddata[!(olddata$site%in%c("JST")),]
data=olddata[(olddata$site%in%c("CTR")),]
data=olddata
winter = data[data$month%in%c(12,1,2),]
winteravg = aggregate(water~year+site,data=winter,FUN=mean)
spring = data[data$month%in%c(3,4,5),]
springavg = aggregate(water~year+site,data=spring,FUN=mean)
summer = data[data$month%in%c(6,7,8),]
summeravg = aggregate(water~year+site,data=summer,FUN=mean)
fall = data[data$month%in%c(9,10,11),]
fallavg = aggregate(water~year+site,data=fall,FUN=mean)

# winterorg = org[org$month%in%c(12,1,2),]
# winterorgavg = aggregate(OC~year+site,data=winterorg,FUN=mean)
# springorg = org[org$month%in%c(3,4,5),]
# springorgavg = aggregate(OC~year+site,data=springorg,FUN=mean)
# summerorg = org[org$month%in%c(6,7,8),]
# summerorgavg = aggregate(OC~year+site,data=summerorg,FUN=mean)
# fallorg = org[org$month%in%c(9,10,11),]
# fallorgavg = aggregate(OC~year+site,data=fallorg,FUN=mean)
# aug = data[data$month%in%c(9),]
# augavg = aggregate(water~year+site,data=aug,FUN=mean)

summary(lm(data.avg$water~data.avg$year))
summary(lm(winteravg$water~winteravg$year))
summary(lm(springavg$water~springavg$year))
summary(lm(summeravg$water~summeravg$year))
summary(lm(fallavg$water~fallavg$year))

ruralavg = aggregate(cbind(water,RH)~year,data=rural,FUN=mean)

# seasonal plot of water

data$dry=data$SO4+data$NH3+data$NO3
monthlywater = aggregate (cbind(water,dry)~month+year, data=data,mean)
monthlywater$date = ymd(paste(monthlywater$year," - ",monthlywater$month," -",15,sep=" "))
monthlywater$time=monthlywater$year+(monthlywater$month)/12
#plot(monthlywater$date,monthlywater$water,type="l")
data$season = NA
data$season[data$month%in%c(12,1,2)]="winter"
data$season[data$month%in%c(3,4,5)]="spring"
data$season[data$month%in%c(6,7,8)]="summer"
data$season[data$month%in%c(9,10,11)]="fall"
seasonalwater = aggregate (cbind(water,dry,month)~season+year, data=data,mean)
seasonalwater$value = NA
seasonalwater$value[seasonalwater$season%in%c("winter")]=0
seasonalwater$value[seasonalwater$season%in%c("spring")]=0.25
seasonalwater$value[seasonalwater$season%in%c("summer")]=0.5
seasonalwater$value[seasonalwater$season%in%c("fall")]=0.75
seasonalwater$time = seasonalwater$year+seasonalwater$value
seasonalwater=seasonalwater[order(seasonalwater$time),]

## at surface

pdf(file="/Users/khoi/Dropbox/Proposal/seasonalavgs1.pdf", width=6, height=6)
plot(monthlywater$time,monthlywater$water,pch=19,col="grey60",cex=0.7,las=1,xlab="Year",ylab="",ylim=c(0,100))
#lines(monthlywater$time,monthlywater$water,type='l',lwd=1,col="grey30")
mtext(expression(paste("Water ("m^\-3", ")", sep=" ")), side=2, line=2,cex=1.2)
lines(seasonalwater$time,seasonalwater$water,type='l',lwd=3,col="darkblue")
legend("topright", inset=0, c("Monthly averages","Seasonal averages"), pch=c(19,19),col=c("grey60","darkblue"), cex=1,box.lwd = 0,box.col = "transparent",bg = "transparent",horiz=F)
dev.off()

## aloft
rhum_long = read.csv("/Users/Khoi/Google Drive/SEARCH/NARR/aloftwaterdata_273.csv")
rhum_long$dateUTC = mdy_hms(as.character(as.chron(ISOdatetime(1800,1,1,0,0,0,"UTC") + 3600*rhum_long$localtime)))
rhum_long$date = with_tz(rhum_long$dateUTC, tz="America/Chicago")
rhum_long$year = year(rhum_long$date)
aloft800 = rhum_long[rhum_long$press==800,]
aloft800$month=month(aloft800$date)

waterdata=data
data=aloft800

#### go run monthlywater and seasonalwater again above
pdf(file="/Users/khoi/Dropbox/Proposal/seasonal.aloft.avgs.pdf",
width=6, height=6)
plot(monthlywater$time,monthlywater$water,pch=19,col="grey60",cex=0.7,las=1,xlab="Year",ylab="",ylim=c(0,20))
#lines(monthlywater$time,monthlywater$water,type='l',lwd=1,col="grey 30")
mtext(expression(paste("Water (µg m^{-3}," sep="" ))), side=2, line=2,cex=1.2)
lines(seasonalwater$time,seasonalwater$water,type='l',lwd=3,col="darkblue")
#lines(seasonalwater$time,seasonalwater$dry,type='l',lwd=3,col="brown3")
legend("topright", inset=0, c("Monthly averages","Seasonal averages"), pch=c(19,19),col=c("grey60","darkblue"), cex=1,box.lwd = 0,box.col = "transparent",bg = "transparent",horiz=F)
dev.off()
library(chron)
library(lubridate)
improvesites = read.csv("/Users/khoi/Google
Drive/SEARCH/NARR/improve-sites-latlon.csv")
improvesites = improvesites[order(improvesites$SiteCode),]

## surface RH
sfcrhfiles <- list.files("/Users/Khoi/Documents/NARR/sfcrh",full=T)
sfcrh_read <- lapply (sfcrhfiles, read.table,
header=F,col.names=c("timestring","sfcrh" ))
for (i in 1:nrow(improvesites)){
  sfcrh_read[[i]]$site = improvesites$SiteCode[i]
  sfcrh_read[[i]]$lat = improvesites$lat[i]
  sfcrh_read[[i]]$lon = improvesites$lon[i]
}
sfcrh <- do.call ("rbind", sfcrh_read)
sfcrh$date=(as.chron(ISODatetime(1800,1,1,0,0,0,"UTC") + 3600* sfcrh
$timestring,format=c("month day year","h:m:s")))
sfcrh $date = as.character(sfcrh $date)
sfcrh $date = mdy_hms(sfcrh $date)
sfcrh $year=year(sfcrh $date)
sfcrh $month=month(sfcrh $date)
sfcrh $hour=hour(sfcrh $date)
write.csv(sfcrh, "/Users/Khoi/Documents/NARR/sfcrh.csv",
row.names=F)

head(sfcrh, 10)

## surface temp
sfctempfiles <-
list.files("/Users/Khoi/Documents/NARR/sfctemp",full=T)
sfctemp_read <- lapply (sfctempfiles, read.table,
header=F,col.names=c("timestring","sfctemp" ))
sfctemp <- do.call ("rbind", sfctemp_read)
sfcrh$sfctemp = sfctemp$sfctemp

## surface pressure
pressfiles <-
list.files("/Users/Khoi/Documents/NARR/sfcpress",full=T)
press_read <- lapply (pressfiles, read.table,
header=F,col.names=c("timestring","sfcpress" ))
sfcpress <- do.call ("rbind", press_read)
sfcrh$sfcpress = sfcpress$sfcpress

### pbl - old years
pblfiles <- list.files("/Users/Khoi/Documents/NARR/pbl",full=T)
pbl_read <- lapply(pblfiles, read.table, header=F, col.names=c("timestring","pbl"))

for (i in 1:nrow(improvesites)){
  pbl_read[[i]]$site = improvesites$SiteCode[i]
  pbl_read[[i]]$lat = improvesites$lat[i]
  pbl_read[[i]]$lon = improvesites$lon[i]
}

pbl <- do.call("rbind", pbl_read)
### pbl - new years

pblfiles2 <- list.files("/Users/Khoi/Documents/NARR/pbl -2", full=T)
pbl2_read <- lapply(pblfiles2, read.table, header=F, col.names=c("timestring","pbl"))

for (i in 1:nrow(improvesites)){
  pbl2_read[[i]]$site = improvesites$SiteCode[i]
  pbl2_read[[i]]$lat = improvesites$lat[i]
  pbl2_read[[i]]$lon = improvesites$lon[i]
}

pbl2 <- do.call("rbind", pbl2_read)

pbl = rbind(pbl1, pbl2)
pbl = pbl[order(pbl$site),]

sfcrh$pbl = pbl$pbl

write.csv(pbl, "/Users/Khoi/Dropbox/Satellites/pbl.csv", row.names=F)
write.csv(sfcrh, "/Users/Khoi/Dropbox/Satellites/sfcrh.csv", row.names=F)

### ### ### ### ### ###
### START HERE
### ### ### ### ### ###
sfcrh = read.csv("/Users/Khoi/Dropbox/Satellites/sfcrh.csv")

### connect with input data
imp <- read.csv("/Users/khoi/Dropbox/Satellites/improve.txt", na.strings=c(-999, "-999"))
library(lubridate)
imp$date <- ymd_hms(imp$Date)
imp <- imp[order(imp$date),]

### ### ### ### ### ### ### ### ###
### MAKING ISORROPIISA INPUT FILE
### ### ### ### ### ### ### ### ###
data=imp
coniso <- data.frame (date = data$date, site = data$SiteCode, state=data$State, Na=c(0), SO4 = data$SO4f.Value,
coniso <- coniso[!is.na(coniso$SO4),]
coniso <- coniso[!is.na(coniso$NO3),]
coniso <- coniso[!is.na(coniso$date),]
## connect dates
coniso$day = floor_date(as.POSIXct(coniso$date), "day")
conisoavg = aggregate(cbind(Na, SO4, NH3, NO3, Cl, Ca, K, Mg, RH, TEMP)~day+site+state, data=coniso, mean)

sfcrh$day = floor_date(as.POSIXct(sfcrh$date), "day")
sfcrh.avg = aggregate(cbind(sfcrh,sfctemp,sfcpress)~day+site+lat+lon, data=sfcrh, mean)

sfcrh.avg=sfcrh.avg[order(sfcrh.avg$day, sfcrh.avg$site),]
head(sfcrh.avg)

sfcrh.avg2 = sfcrh.avg[paste(sfcrh.avg$site,as.character(sfcrh.avg$day))%in%paste (conisoavg$site,as.character(conisoavg$day)),]
conisoavg2 = conisoavg[paste(conisoavg$site,as.character(conisoavg$day))%in%paste (sfcrh.avg2 $site,as.character(sfcrh.avg2 $day)),]

tail(conisoavg2)
tail(sfcrh.avg2)

iso = conisoavg2
iso$RH = 0.01*as.numeric(as.character(sfcrh.avg2$sfcrh))
iso$TEMP = as.numeric(as.character(sfcrh.avg2$sfctemp))
iso$pressure_mb = 0.01*as.numeric(as.character(sfcrh.avg2$sfcpress))
iso$lat = sfcrh.avg2$lat
iso$lon = sfcrh.avg2$lon

write.csv(iso, "/Users/Khoi/Documents/NARR/iso_surface.csv", row.names=F)

###write iso file

header <- c(  
"Input units (0=umol/m3, 1=ug/m3)  
1  

Problem type (0=forward, 1=reverse); Phase state (0=solid+liquid, 1=metastable)  
1, 1  

SO4-N03 system case  
Na  SO4  NH3  NO3  Cl  Ca  K  Mg  RH  TEMP"")
# function that appends header to file
write.table_with_header <- function(x, file, header, ...){cat(header, '\n', file = file)
    write.table(x, file, append = T, ...)}

#ignore warning message for line below
write.table_with_header(iso[,4:13],
    "~/Users/Khoi/Dropbox/Isorropia/v22_errck/improve-surface.INP",
header, sep='     ', row.names=F, col.names=F)

## after running ISORROPIA
surface_water = read.table("~/Documents/NARR/surface run/improve-surface.dat",
header=TRUE, quote=" ", na.strings="NaN")
iso$water = surface_water$WATER

iso = iso[order(iso $day),]
iso$year = year(iso$day)

Northeast <- c("ME","NH","VT","MA","RI","CT","NY","NJ","DE","PA")
Southeast<- c("WV","VA","KY","TN","NC","SC","GA","FL","AL","MS","LA","AR", "MD", "DC")
Central <- c("TX","OK","MN","IA","MO","KS","NE")
Midwest <- c("OH","IN","IL","MI","WI","ND","SD")
MountainWest <- c("AZ","NM","MT","ID","NV","UT","WY","CO")
WestPacific <- c("WA","OR","CA")

iso$region=NA
iso$region[iso$state%in% Northeast]= "Northeast"
iso$region[iso$state%in% Southeast]= "Southeast"
iso$region[iso$state%in% Central]= "Central"
iso$region[iso$state%in% Midwest]= "Midwest"
iso$region[iso$state%in% WestPacific]= "West Pacific"
iso$region[iso$state%in% MountainWest]= "Mountain West"

regions = (unique(iso$region))
regions= regions[!is.na(regions)]
iso$season[month(iso$day)%in%c(12,1,2)] = "winter"
iso$season[month(iso$day)%in%c(3,4,5)] = "spring"
iso$season[month(iso$day)%in%c(6,7,8)] = "summer"
iso$season[month(iso$day)%in%c(9,10,11)] = "fall"

write.csv(iso, "~/Users/Khoi/Dropbox/Satellites/iso-surface-withwater.csv", row.names=F)

### START HERE IN FUTURE ###
iso = read.csv("~/Users/Khoi/Dropbox/Satellites/iso-surface-withwater.csv")
library(lubridate)
iso_seasons = iso[iso$season%in%c("winter","summer"),]
iso_seasons = iso_seasons[order(iso_seasons$site),]
iso_seasons_avg = aggregate
cbind(water,RH,TEMP,pressure_mb)~site+state+season+lat+lon+region+year, data=iso_seasons,mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site,
iso_seasons_avg$year),]
iso_seasons_avg$seasdiff=NA
for (i in seq(2,nrow(iso_seasons_avg),2)){
  if
    (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$state[i])){
      iso_seasons_avg$seasdiff[i] =
      abs(as.numeric(iso_seasons_avg$water[i+1]) -
      as.numeric(iso_seasons_avg$water[i]))
    }
}

improvesites = read.csv("/Users/khoi/Google
Drive/SEARCH/NARR/improve-sites-latlon.csv")
regions = c("West Pacific","Northeast","Mountain
West","Southeast","Central","Midwest")

library(maps)
library(mapproj)
library(mapplots)
## Goldstein et al paper: March 2000 to February 2007
# iso_0007 = iso_seasons_avg[iso_seasons_avg$year%in%(2000:2007),]
## 2000-2006
iso_0007 = iso_seasons_avg ### ALL YEARS
iso_0007=as.data.frame(iso_0007)
iso_0007_avg = aggregate
cbind(seasdiff,water)~site+state+region+lat+lon,
data=iso_0007,mean)
iso_0007_avg$seasdiff_real = iso_0007_avg$seasdiff
#iso_0007_avg$seasdiff[iso_0007_avg$seasdiff_real>15]=15
rbPal <-
colorRampPalette(c('blue','turquoise','green','yellow','orange','red '))
iso_0007_avg$color <-
rPal(70)[as.numeric(cut(iso_0007_avg$seasdiff,breaks = 70))]

iso_0007_avg=iso_0007_avg[order(iso_0007_avg$site,decreasing = T),]
iso_0007_avg=data.frame(iso_0007_avg)

color.bar <- function(lut, min, max=-min, nticks=11, ticks=seq(min,
max, len=nticks), title='') {
  scale = (length(lut)-1)/(max-min)
  plot(c(0,10), c(min,max), type='n', bty='n', xaxt='n', xlab='',
yaxt='n', ylab='', main=title)
  axis(4, ticks, las=1)
  for (i in 1:(length(lut)-1)) {
    y = (i-1)/scale + min
  }
rect(0,y,10,y+1/scale, col=lut[i], border=NA)
}
}

#### map goldstein

df(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/Surface_Map.pdf", width=6.6,height=6)
par(mfrow=c(1,1),mar=c(0,0,0,0),oma=c(0,0,0,3))
map('state', fill = TRUE, col = "grey90")
map.scale(x=-120.5,y=27.5,ratio=F, relwidth=0.2,cex=0.8)
points(x=iso_0007_avg$lon,y=iso_0007_avg$lat,col=iso_0007_avg$color, cex=1.1,pch=19)

par(new=T,mar=c(0,17,0,1),oma=c(10,12,10,2))
iso_0007_avg$scale=round(iso_0007_avg$seasdiff,digits=0)
color.bar(rbPal(50), min(iso_0007_avg$scale),max=max(iso_0007_avg$scale))

dev.off()

## START OF ISO_SUB
#iso_sub = iso[iso$year%in%c(2000:2008),]
#iso_sub = iso_sub[iso_sub$season%in%c("winter", "summer"),]

########## CHANGE THIS FOR OTHER PRESSURES
#iso_sub = iso_sub[iso_sub$pressure == 1000,]
########## CHANGE THIS FOR OTHER PRESSURES
iso_sub = iso
iso_sub$year2 = iso_sub$year
iso_sub$month = month(iso_sub$day)
iso_sub$year2[iso_sub$season=="summer"] = iso_sub$year[iso_sub$season=="summer"]+0.5
iso_sub=iso_sub[order(iso_sub$year2),]

summarySE <- function(data=NULL, measurevar , groupvars=NULL, na.rm=FALSE, conf.interval=.95, .drop=TRUE) {
  require(plyr)
  # New version of length which can handle NA's: if na.rm==T, don't
  # count them
  length2 <- function (x, na.rm=FALSE) {
    if (na.rm) sum(!is.na(x))
    else length(x)
  }
  # This is does the summary
  datac <- ddply(data, groupvars, .drop=.drop,
     .fun= function(xx, col, na.rm) {
       c( N   = length2(xx[[col]], na.rm=na.rm),
         mean = mean (xx[[col]], na.rm=na.rm),
         sd   = sd   (xx[[col]], na.rm=na.rm),
         median = median(xx[[col]], na.rm=na.rm),
         quartile25 = as.vector(summary(xx[[col]],
          na.rm=na.rm))[2],
    na.rm=na.rm))}
quartile75 = as.vector(summary(xx[[col]],
na.rm=na.rm))[5]
}

# Rename the "mean" column
datac <- rename(datac, c("mean"=measurevar))
datac$se <- datac$sd / sqrt(datac$N)  # Calculate standard error of the mean
# Confidence interval multiplier for standard error
# Calculate t-statistic for confidence interval:
# e.g., if conf.interval is .95, use .975 (above/below), and use
df=N-1
  ciMult <- qt(conf.interval/2 + .5, datac$N-1)
datac$ci <- datac$se * ciMult
return(datac)

#### Figure 2 of paper
#iso_sub = iso[iso$year%in%c(2000:2008),]
#iso_sub = iso_sub[iso_sub$pressure == 1000,]
pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/F2.pdf",
width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,4,2,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
measurevar="water", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  plot(NA,NA, col="white", ylim=c(0,20), ylab="", xlab="Year",
xlim=c(1988,2013),las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
c((summary_data$quartile25), rev(summary_data$quartile75)),
col="cornflowerblue", border="cornflowerblue")
  lines(summary_data$time,summary_data$median, col="black",
  lty=1,lwd=3,cex=1)
  title(regions[i])
}

par(new=T,mfrow=c(1,1),mar=c(0,3,0,0),oma=c(0,0,0,0))
plot (NA,NA, xlim = c(0,1), ylim=c(0,1), yaxt="n", xaxt="n",
frame.plot=F)
mtext(expression(paste("Water", " (µg \, m^{-3}, )", sep="" ))),
side=2, line=1, cex= 1.3)

dev.off()
for (i in 1:length(regions)) {
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
    measurevar="RH", groupvars=c("year", "season"))
  summary_data = summary_data[, is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,1), ylab="", xlab="Year",
    xlim=c(1988,2013),las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
    c(summary_data$quartile25), rev(summary_data$quartile75)),
    col="grey75", border="grey75")
  lines(summary_data$time,summary_data$median, col="black",
    lty=1,lwd=3,cex=1)
  # lines(summary_data$time,summary_data$RH, col="orange",
  # lty=1,lwd=1.5,cex=1)
  # mtext("RH", side=2, line=2)
  title(regions[i])
}

par(new=T,mfrow=c(1,1),mar=c(0,3,0,0),oma=c(0,0,0,0))
plot (NA,NA, xlim = c(0,1), ylim=c(0,1), yaxt="n", xaxt="n",
  frame.plot=F)
mtext(expression(paste("RH", sep="")), side=2, line=1, cex=1.3)
dev.off()

### surface to boundary layer
pbl = read.csv("/Users/Khoi/Dropbox/Satellites/pbl.csv")
library(lubridate)
library(chron)
pbl$date=(as.chron(ISOdatetime(1800,1,1,0,0,0,"UTC") + 3600* pbl
  $timestring,format=c("month day year","h:m:s")))
pbl $date = as.character(pbl $date)
pbl $date = mdy_hms(pbl $date)
pbl $year=year(pbl $date)
pbl $month=month(pbl $date)
pbl $hour=hour(pbl $date)
pbl$day = floor_date(as.POSIXct(pbl$date), "day")
pbl.avg = aggregate(cbind(pbl)-day+site+lat+lon, data=pbl, mean)
pbl.avg=pbl.avg[order(pbl.avg$day,pbl.avg$site),]
head(pbl.avg)

## merge pbl and surface data
# pbl.avg2 =
pbl.avg[paste(pbl.avg$site,as.character(pbl.avg$day))%in%paste(sfcrh.
  avg2$site,as.character(sfcrh.avg2$day)),]
```r
# sfcrh.avg3 = sfcrh.avg2[paste(sfcrh.avg2$site,as.character(sfcrh.avg2$day))%in%paste(pbl.avg2$site,as.character(pbl.avg2$day))],

iso2 = iso
pbl.avg2 = pbl.avg[paste(pbl.avg$site,as.character(pbl.avg$day))%in%paste(iso2$site,as.character(iso2$day))],
iso3 = iso2[paste(iso2$site,as.character(iso2$day))%in%paste(pbl.avg2$site,as.character(pbl.avg2$day))],

head(pbl.avg2)
head(iso3)
iso3$pbl = pbl.avg2$pbl
write.csv(iso3, "/Users/Khoi/Dropbox/Satellites/iso-surface-withpbl.csv", row.names=F)

##
iso.pbl = read.csv("/Users/Khoi/Dropbox/Satellites/iso-surface-withpbl.csv")
iso.all = read.csv("/Users/Khoi/Dropbox/Satellites/iso-ALLYEARS-withwater.csv")

iso.all = iso.all[order(iso.all$site,iso.all$day),]
summary(iso.pbl$pbl)

#####
### Seasonal differences map
#####

color.bar <- function(lut, min, max=-min, nticks=11, ticks=seq(min, max, len=nticks), title='') {
  scale = (length(lut)-1)/(max-min)
  plot(0,10, c(min,max), type='n', bty='n', xaxt='n', yaxt='n', ylab='', main=title)
  axis(4, ticks, las=1)
  for (i in 1:(length(lut)-1)) {
    y = (i-1)/scale + min
    rect(0,y,10,y+1/scale, col=lut[i], border=NA)
  }
}
rbPal <- colorRampPalette(c('blue','turquoise','green','yellow','orange','red '))

## (a) surface from 2000-2007
iso_seasons = iso[iso$season%in%c("winter","summer"),]
iso_seasons.a = iso_seasons[iso_seasons$year%in%c(2000:2007),]
```
#mean for summer and winter
iso_seasons_avg = aggregate
  (cbind(water,RH,TEMP,pressure_mb)~site+state+season+lat+lon+region+year, data=iso_seasons.a,mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site,
    iso_seasons_avg$year),]
iso_seasons_avg$seasdiff=NA

for (i in seq(2,nrow(iso_seasons_avg),2)){
  if
    (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$state[i])){
      iso_seasons_avg$seasdiff[i] =
    abs(as.numeric(iso_seasons_avg$water[i+1])-
        as.numeric(iso_seasons_avg$water[i]))
    }
}

iso_0007 = iso_seasons_avg
iso_0007=as.data.frame(iso_0007)
iso_0007_avg = aggregate
  (cbind(seasdiff,water)~site+state+region+lat+lon,
    data=iso_0007,mean)
iso_0007_avg$color <-
  rbPal(70)[as.numeric(cut(iso_0007_avg$seasdiff,breaks = 70))]
iso_0007_avg=iso_0007_avg[order(iso_0007_avg$site,decreasing = T),]
iso_0007_avg=as.data.frame(iso_0007_avg)

## (b) 1000 mb from 1988-2013
iso_seasons = iso[iso$season%in%c("winter","summer"),]

#mean for summer for winter
iso_seasons_avg = aggregate
  (cbind(water,RH,TEMP,pressure_mb)~site+state+season+lat+lon+region+year, data=iso_seasons,mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site,
    iso_seasons_avg$year),]
iso_seasons_avg$seasdiff=NA

for (i in seq(2,nrow(iso_seasons_avg),2)){
  if
    (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$state[i])){
      iso_seasons_avg$seasdiff[i] =
    abs(as.numeric(iso_seasons_avg$water[i+1])-
        as.numeric(iso_seasons_avg$water[i]))
    }
}

iso_0007 = iso_seasons_avg
iso_0007=as.data.frame(iso_0007)
iso_0007_avg = aggregate
  (cbind(seasdiff,water)~site+state+region+lat+lon,
  data=iso_0007,mean)
iso_0007_avg$color <-
  rbPal(70)[as.numeric(cut(iso_0007_avg$seasdiff,breaks = 70))]
iso_0007_avg=iso_0007_avg[order(iso_0007_avg$site,decreasing = T),]
iso_0007_avg=data.frame(iso_0007_avg)
iso_0007_avg.b = iso_0007_avg

## (c) aloft to pbl from 2000-2007
iso_seasons = iso.all[iso.all$season%in%c("winter","summer"),]
iso.pbl_seasons = iso.pbl[iso.pbl$season%in%c("winter","summer"),]
lubridate
iso.pbl Seasons$day2 = ymd(iso.pbl Seasons$day)

iso Seasons2 =
  iso seasons[paste(iso seasons$site,as.character(iso seasons$day2))%i
  n%paste(iso.pbl seasons $site,as.character(iso.pbl seasons $day2)),]

## iso Seasons2$pbl = NA
## iso Seasons2$surface mb = NA
## for (i in seq(3511756,nrow(iso_seasons2),29)){
  ##  iso Seasons2$pbl[i:(i+28)] =
  ##  iso.pbl seasons[paste(iso.pbl seasons$day2,iso.pbl seasons$site)
  ##  %in%paste(iso seasons2$day2[i],iso seasons2$site[i])]
  ##  iso Seasons2$surface mb[i:(i+28)] =
  ##  iso.pbl seasons$pressure mb[paste(iso.pbl seasons$day2,iso.pbl seaso
  ##  ns$site)%in%paste(iso seasons2$day2[i],iso seasons2$site[i])]
## }

write.csv(iso Seasons2, " /Users/Khoi/Documents/NARR/iso. all-
  withpbl.csv")

iso Seasons2 = read.csv("/Users/Khoi/Documents/NARR/iso. all-
  withpbl.csv")

iso Seasons3 = iso Seasons2[(iso Seasons2$pressure <
  iso Seasons2$pbl),]
iso Seasons3 = iso Seasons3[(iso Seasons3$pressure >
  iso Seasons3$surface mb),]

iso Seasons4 =
data.frame(day2=iso Seasons3$day2,site=iso Seasons3$site,
  state=iso Seasons3$state, season=iso Seasons3$season,
  region=iso Seasons3$region,
  year=iso Seasons3$year,water=iso Seasons3$water)
iso2=data.frame(day2=iso$day,site=iso$site, state=iso$state,
  season=iso$season, region=iso$region, year=iso$year,
  water=iso$water)
iso Seasons4$day2=ymd(iso Seasons4$day2)
iso2$day2 = ymd(iso2$day)
iso_seasons5 = rbind(iso_seasons4, iso2)
iso_seasons5 = iso_seasons5[order(iso_seasons5$site, iso_seasons5$day2),]
iso_seasons.c = iso_seasons5[iso_seasons5$year%in%c(2000:2007),]

# mean for summer for winter
iso_seasons_avg = aggregate(cbind(water)~site+state+season+region+year, 
data=iso_seasons.c, mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site, iso_seasons_avg$year),]
iso_seasons_avg$seasdiff=NA
for (i in seq(2,nrow(iso_seasons_avg),2)){
  if (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$state[i])){
    iso_seasons_avg$seasdiff[i] = abs(as.numeric(iso_seasons_avg$water[i+1]) -
    as.numeric(iso_seasons_avg$water[i]))
  }
}
for (i in 1:nrow(iso_seasons_avg)){
  iso_seasons_avg$lat[i]=(improvesites$lat[improvesites$SiteCode%in%iso_seasons_avg$site[i]])
  iso_seasons_avg$lon[i]=(improvesites$lon[improvesites$SiteCode%in%iso_seasons_avg$site[i]])
}
iso_0007 = iso_seasons_avg
iso_0007=as.data.frame(iso_0007)
iso_0007_avg = aggregate(cbind(seasdiff, water)~site+state+region+lat+lon, 
data=iso_0007, mean)
iso_0007_avg=color <- rbPal(70)[as.numeric(cut(iso_0007_avg$seasdiff, breaks = 70))]
iso_0007_avg=iso_0007_avg[order(iso_0007_avg$site, decreasing = T),]
iso_0007_avg=as.data.frame(iso_0007_avg)

## (d) 800-1000 mb from 1998-2013

# mean for summer for winter
iso_seasons_avg = aggregate(cbind(water)~site+state+season+region+year, 
data=iso_seasons5, mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site, iso_seasons_avg$year),]
iso_seasons_avg$seasdiff=NA
for (i in seq(2,nrow(iso_seasons_avg),2)){
if (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$state[i])){
    iso_seasons_avg$seasdiff[i] = abs(as.numeric(iso_seasons_avg$water[i+1]) - as.numeric(iso_seasons_avg$water[i]))
}
}
for (i in 1:nrow(iso_seasons_avg)){
iso_seasons_avg$lat[i] = (improvesites$lat[improvesites$SiteCode%in%iso_seasons_avg$site[i]])
iso_seasons_avg$lon[i] = (improvesites$lon[improvesites$SiteCode%in%iso_seasons_avg$site[i]])
}
iso_0007 = iso_seasons_avg
iso_0007 = as.data.frame(iso_0007)
iso_0007_avg = aggregate(cbind(seasdiff,water)~site+state+region+lat+lon, data=iso_0007,mean)
iso_0007_avg$color <- rbPal(70)[as.numeric(cut(iso_0007_avg$seasdiff,breaks = 70))]
iso_0007_avg=iso_0007_avg[order(iso_0007_avg$site,decreasing = T),]
iso_0007_avg = data.frame(iso_0007_avg)
iso_0007_avg = iso_0007_avg[order(iso_0007_avg$site,decreasing = T),]
iso_0007_avg = data.frame(iso_0007_avg)
iso_0007_avg.d = iso_0007_avg
## getting color scales the same
iso_0007_avg.a$letter = c("a")
iso_0007_avg.b$letter = c("b")
iso_0007_avg.c$letter = c("c")
iso_0007_avg.d$letter = c("d")
seasdiff_data = rbind(iso_0007_avg.a,iso_0007_avg.b,iso_0007_avg.c,iso_0007_avg.d)
seasdiff_data$seasdiff2 = seasdiff_data$seasdiff
seasdiff_data$seasdiff2[seasdiff_data$seasdiff>13] = 13
seasdiff_data$color = rbPal(70)[as.numeric(cut(seasdiff_data$seasdiff2,breaks = 70))]
iso_0007_avg.a = seasdiff_data[seasdiff_data$letter=="a",]
iso_0007_avg.b = seasdiff_data[seasdiff_data$letter=="b",]
iso_0007_avg.c = seasdiff_data[seasdiff_data$letter=="c",]
iso_0007_avg.d = seasdiff_data[seasdiff_data$letter=="d",]
library(maps)
library(mapproj)
library(mapplots)
### the overall plot
pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/F1_mean2.pdf", width=8,height=5,useDingbats=FALSE)
cex_n = 0.7
par(mfrow=c(2,2),mar=c(0,0,0,0),oma=c(0,0,0,5))

map('state', fill = TRUE, col = "grey90", lwd=0.6)
points(x=iso_0007_avg.a$lon,y=iso_0007_avg.a$lat,col=iso_0007_avg.a$color,cex=cex_n,pch=19)

map('state', fill = TRUE, col = "grey90", lwd=0.6)
points(x=iso_0007_avg.b$lon,y=iso_0007_avg.b$lat,col=iso_0007_avg.b$color,cex=cex_n,pch=19)

map('state', fill = TRUE, col = "grey90", lwd=0.6)
points(x=iso_0007_avg.c$lon,y=iso_0007_avg.c$lat,col=iso_0007_avg.c$color,cex=cex_n,pch=19)

map('state', fill = TRUE, col = "grey90", lwd=0.6)
points(x=iso_0007_avg.d$lon,y=iso_0007_avg.d$lat,col=iso_0007_avg.d$color,cex=cex_n,pch=19)

#map.scale(x=-120.5,y=27.5,ratio=F, relwidth=0.2,cex=0.8)

par(new=T,mfrow=c(1,1),mar=c(0,17,0,1),oma=c(7,19,7,2))
seasdiff_data$scale=round(seasdiff_data$seasdiff 2,digits=0)
color.bar(rbPal(50),
min(seasdiff_data$scale),max=max(seasdiff_data$scale))
dev.off()

### all pressures
#### NARR data
summarySE <- function(data=NULL, measurevar, groupvars=NULL, na.rm=FALSE, conf.interval=.95, .drop=TRUE) {
  require(plyr)
  # New version of length which can handle NA's: if na.rm==T, don't count them
  length2 <- function (x, na.rm=FALSE) {
    if (na.rm) sum(!is.na(x))
    else length(x)
  }
  # This is does the summary
  datac <- ddply(data, groupvars, .drop=.drop,
    .fun= function(xx, col, na.rm) {
      c( N    = length2(xx[[col]], na.rm=na.rm),
         mean = mean   (xx[[col]], na.rm=na.rm),
         sd   = sd     (xx[[col]], na.rm=na.rm),
         median = median(xx[[col]], na.rm=na.rm),
         quartile25 = as.vector(summary(xx[[col]],
          na.rm=na.rm))[2],
         quartile75 = as.vector(summary(xx[[col]],
          na.rm=na.rm))[5]
improvesites = read.csv("/Users/khoi/Google Drive/SEARCH/NARR/improve-sites-latlon.csv")
sites = unique(iso[, c("site", "state", "region")])
for (i in 1:nrow(sites)){
  sites$lat[i]=(improvesites$lat[improvesites$SiteCode%in%sites$site[i]])
  sites$lon[i]=(improvesites$lon[improvesites$SiteCode%in%sites$site[i]])
}

##map of improvesites
pdf("/Users/Khoi/Dropbox/Satellites/Figures/map_sites.pdf")
map('state',fill = TRUE, col="white")
points(x=sites$lon[sites$region=="Northeast"],y=sites$lat[sites$region=="Northeast"],col="blue",pch=19,cex=1.1)
points(x=sites$lon[sites$region=="West Pacific"],y=sites$lat[sites$region=="West Pacific"],col="lightgoldenrod1",pch=19,cex=1.1)
points(x=sites$lon[sites$region=="Mountain West"],y=sites$lat[sites$region=="Mountain West"],col="purple",pch=19,cex=1.1)
points(x=sites$lon[sites$region=="Central"],y=sites$lat[sites$region=="Central"],col="indianred1",pch=19,cex=1.1)
points(x=sites$lon[sites$region=="Midwest"],y=sites$lat[sites$region=="Midwest"],col="cornflowerblue",pch=19,cex=1.1)
points(x=sites$lon[sites$region=="Southeast"],y=sites$lat[sites$region=="Southeast"],col="darkgreen",pch=19,cex=1.1)
map.scale(x=-121.5, y=28.5, ratio=FALSE, relwidth=0.2)
#par(new=T)
#plot(NA,NA, xlim=c(-1,1), ylim=c(-1,1) ,las=1, ylab="",xlab="",xaxt="n",yaxt="n",bty="n")
legend("bottomright", inset=0, c("Northeast","Southeast","Midwest","Central","Mountain West","West..."
## TEMPERATURE

tempfiles <- list.files("/Users/Khoi/Documents/NARR/Satellite/temp", full=T, pattern="*temp.txt")

for (i in 1:nrow(improvesites)) {
  temp_read[[i]]$site = improvesites$SiteCode[i]
}

temp <- do.call ("rbind", temp_read)
library(chron)
temp$date = mdy_hms(as.character(as.chron(ISOdatetime(1800,1,1,0,0,0,"UTC") + 3600*temp$timestring)))
temp$year = year(temp$date)
temp$month = month(temp$date)
temp$hour = hour(temp$date)

## SPECIFIC HUMIDITY

rhfiles <- list.files("/Users/Khoi/Documents/NARR/Satellite/shum", full=T, pattern="*shum.txt")

for (i in 1:nrow(improvesites)) {
  rh_read[[i]]$site = improvesites$SiteCode[i]
}

shum <- do.call ("rbind", rh_read)
shum$date = mdy_hms(as.character(as.chron(ISOdatetime(1800,1,1,0,0,0,"UTC") + 3600*shum$timestring)))
shum$year = year(shum$date)
shum$month = month(shum$date)
shum$hour = hour(shum$date)

#### Specific humidity to RH - NARR data
qair2rh <- function(qair, temp, press) {
  es <- 6.112 * exp((17.67 * temp)/(temp + 243.5))
  e <- qair * press / (0.378 * qair + 0.622)
  rh <- e / es
  rh[rh > 1] <- 1
  rh[rh < 0] <- 0
  return(rh)
}

press = as.numeric(c("1000", "975", "950", "925", "900", "875",
                     "850", "825", "800", "775", "750", "725", "700", "650", "600",
                     "200", "175", "150", "125", "100"))
rhum = shum
for (i in 2:30) {
  rhum[,i]=as.numeric(qair2rh(qair = (shum[,i]/1000),
                             temp=(temp[,i]-273.15), press=press[(i-1)]))
}

# write.csv(rhum, "/Users/Khoi/Documents/NARR/Satellite/rhum.csv", row.names=F)
# write.csv(temp, "/Users/Khoi/Documents/NARR/Satellite/temp.csv", row.names=F)

rhum = read.csv("/Users/Khoi/Documents/NARR/Satellite/rhum.csv")
temp = read.csv("/Users/Khoi/Documents/NARR/Satellite/temp.csv")

#### connect with input data

imp <- read.csv("/Users/khoi/documents/organosulfates/improve.txt", na.strings=c(-999,"-999"))
library(lubridate)
imp$date <- ymd_hms(imp$Date)
imp <- imp[order(imp$date),]

#### #### #### #### #### #### #### ####
#### MAKING ISORROPISA INPUT FILE
#### #### #### #### #### #### #### ####
data=imp
conso <- data.frame (date = data$date, site = data$SiteCode,
                     state=data$State,Na=c(0),SO4 = data$SO4f.Value,
                     NH3=c(0),NO3=data$NO3f.Value,Cl=c(0),Ca=c(0), K=c(0),Mg=c(0),
                     RH=c(0), TEMP=c(0))
conso = coniso[order(coniso$date,coniso$site),]
coniso <- coniso[!is.na(coniso$SO4),]
coniso <- coniso[!is.na(coniso$NO3),]
coniso <- coniso[!is.na(coniso$date),]

## connect dates
library(lubridate)
coniso$day = floor_date(as.POSIXct(coniso$date), "day")
conisoavg = aggregate(cbind(Na, SO4, NH3, NO3, Cl, Ca, K, Mg, RH, TEMP)~day+site+state, data=coniso, mean)
conisoavg=conisoavg[order(conisoavg$day,conisoavg$site),]

rhum$day = floor_date(as.POSIXct(rhum$date), "day")
rhumavg = aggregate(cbind(X1000, X975, X950, X925, X900, X875, X850, X825, X800, X775, X750, X725, X700, X650, X600, X550, X500, X450, X400, X350, X300, X275, X250, X225, X200, X175, X150, X125, X100)~day+site, data=rhum, mean)

temp$day = rhum$day
temp$site = rhum$site
tempavg = aggregate(cbind(X1000, X975, X950, X925, X900, X875, X850, X825, X800, X775, X750, X725, X700, X650, X600, X550, X500, X450, X400, X350, X300, X275, X250, X225, X200, X175, X150, X125, X100)~day+site, data=temp, mean)
rhumavg=rhumavg[order(rhumavg$day),]
tempavg=tempavg[order(tempavg$day),]

write.csv(rhumavg,
"/Users/Khoi/Documents/NARR/Satellite/rhumavg.csv", row.names=F)
write.csv(tempavg,
"/Users/Khoi/Documents/NARR/Satellite/tempavg.csv", row.names=F)

rhumavg2 = rhumavg[paste(rhumavg$site,as.character(rhumavg$day))%in%paste(conisoavg$site,as.character(conisoavg$day))]
tempavg2 = tempavg[paste(tempavg$site,as.character(tempavg$day))%in%paste(conisoavg$site,as.character(conisoavg$day))]
conisoavg2 = conisoavg[paste(conisoavg$site,as.character(conisoavg$day))%in%paste(rhumavg2$site,as.character(rhumavg2$day))]

iso = conisoavg2
iso$pressure=NA

iso_backup = iso

row= 258475
for (i in 1:29){
  iso$RH[((i-1)*row+1):(i*row)]= rhumavg2[,,(i+2)]
  iso$TEMP[((i-1)*row+1):(i*row)] = tempavg2 [,,(i+2)]
  iso$pressure[((i-1)*row+1):(i*row)]=c(press[i])
  if (i != 29){
    i=i+1
  }
iso[((i-1)*row+1):(i*row),]= iso_backup
}

write.csv(iso,
"/Users/Khoi/Documents/NARR/Satellite/iso_withstates.csv",
row.names=F)

###write iso file

header <- c(
"Input units (0=umol/m3, 1=ug/m3)
1

Problem type (0=forward, 1=reverse); Phase state (0=solid+liquid, 1=metastable)
1, 1

SO4-NO3 system case
Na  SO4  NH3  NO3  Cl  Ca  K  Mg  RH  TEMP")

# function that appends header to file
write.table_with_header <- function(x, file, header, ...
){cat(header, '
', file = file)
write.table(x, file, append = T, ...)}

#ignore warning message for line below
write.table_with_header(iso[,4:13],
"/Users/Khoi/Dropbox/Isorropia/v22_errck/improve2000onwards-
withstates.INP", header, sep='     ', row.names=F, col.names=F)

#old one was improve2000onwards
water.improve = read.table("~/Documents/NARR/Satellite/Isorropia Run
2/improve2000onwards-withstates.dat", header=TRUE, quote="\"
, na.strings="NaN")
#iso = read.csv("/Users/Khoi/Documents/NARR/Satellite/iso.csv")
isowater = water.improve$WATER

# states = unique (imp[,c("SiteCode", "State")])
# states$State=as.character(states$State)
# a= 2044693
# #iso$state=NA
# for (i in a:nrow(iso)){
#   # iso$state[i]=(states$State[states$SiteCode%in%iso$site[i]])
# }

iso = iso[order(iso $day),]
isoyear = year(iso$day)
iso$region=NA
iso$region[iso$state%in% Northeast]= "Northeast"
iso$region[iso$state%in% Southeast]= "Southeast"
iso$region[iso$state%in% Central]= "Central"
iso$region[iso$state%in% Midwest]= "Midwest"
iso$region[iso$state%in% WestPacific]= "West Pacific"
iso$region[iso$state%in% MountainWest]= "Mountain West"

regions = (unique(iso$region))
regions= regions[!is.na(regions)]

write.csv(iso, "/Users/Khoi/Dropbox/Satellites/iso-withwater.csv", row.names=F)

### BOXPLOTs

pdf(file="/Users/Khoi/Dropbox/Satellites/Figures/regions_timeseriesboxplots.pdf", width=6.6,height=7)

par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
  boxplot(iso$water[iso$region==regions[i]]~year(iso$day[iso$region==regions[i]]), varwidth=T, las=1,ylim=c(0,20),main=regions[i])
}

dev.off()

### FINDING SEASONAL DIFFERENCES

iso$season[month(iso$day)%in%c(12,1,2)] = "winter"
iso$season[month(iso$day)%in%c(3,4,5)] = "spring"
iso$season[month(iso$day)%in%c(6,7,8)] = "summer"
iso$season[month(iso$day)%in%c(9,10,11)] = "fall"

### START HERE IN FUTURE

#iso = read.csv("/Users/Khoi/Documents/NARR/Satellite/iso-withwater.csv")
#iso = read.csv("/Users/Khoi/Dropbox/Satellites/iso-withwater.csv")
iso = read.csv("/Users/Khoi/Dropbox/Satellites/iso-ALLYEARS-withwater.csv")
iso$region=NA

Northeast <- c("ME","NH","VT","MA","RI","CT","NY","NJ","DE","PA")
Southeast<- c("WV","VA","KY","TN","NC","SC","GA","FL","AL","MS","LA","AR","MD","DC")
Central <- c("TX","OK","MN","IA","MO","KS","NE")
Midwest <- c("OH","IN","IL","MI","WI")
MountainWest <- c("AZ","NM","MT","ID","NV","UT","WY","CO","ND","SD")
WestPacific <- c("WA","OR","CA")
iso$region[iso$state%in% Northeast]= "Northeast"
iso$region[iso$state%in% Southeast]= "Southeast"
iso$region[iso$state%in% Central]= "Central"
iso$region[iso$state%in% Midwest]= "Midwest"
iso$region[iso$state%in% WestPacific]= "West Pacific"
iso$region[iso$state%in% MountainWest]= "Mountain West"

regions = (unique(iso$region))
regions= regions[!is.na(regions)]

library(lubridate)

improvesites = read.csv("/Users/khoi/Google Drive/SEARCH/NARR/improve-sites-latlon.csv")
regions = c("West Pacific","Northeast","Mountain West","Southeast","Central","Midwest")

library(maps)
library(mapproj)
library(mapplots)

iso_seasons = iso[iso$season%in%c("winter","summer"),]

#mean for summer and winter
iso_seasons_avg = aggregate
(cbind(water)~site+state+season+region+year, data=iso_seasons,mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site,
iso_seasons_avg$year),]
iso_seasons_avg$seasdiff=NA
for (i in seq(2,nrow(iso_seasons_avg),2)){
  if
  (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$state[i])){
    iso_seasons_avg$seasdiff[i] =
      abs(as.numeric(iso_seasons_avg$water[i+1])-as.numeric(iso_seasons_avg$water[i]))
  }
}

iso_seasons_avg$ratio = NA
for (i in seq(2,nrow(iso_seasons_avg),2)){
  if
  (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$state[i])){
    iso_seasons_avg$ratio[i] = (as.numeric(iso_seasons_avg$water[i-1])/as.numeric(iso_seasons_avg$water[i]))
  }
}

for (i in 1:nrow(iso_seasons_avg)){

iso_seasons_avg$lat[i]=(improvesites$lat[improvesites$SiteCode%in%iso_seasons_avg$site[i]])
iso_seasons_avg$lon[i]=(improvesites$lon[improvesites$SiteCode%in%iso_seasons_avg$site[i]])

### plotting mass ratios
par(oma=c(0,0,0,0),mar=c(2,2,0,0))
boxplot(ratio~year, data=iso_seasons_avg,ylim=c(0,3))
summary(iso_seasons_avg$ratio)
iso_seasons_avg_SE = iso_seasons_avg[iso_seasons_avg$region%in%c("Southeast"),]
ggplot(iso_seasons_avg_SE, aes(year,ratio)) + geom_point() + geom_smooth()
iso_seasons_avg_NE = iso_seasons_avg[iso_seasons_avg$region%in%c("Northeast"),]
ggplot(iso_seasons_avg_NE, aes(year,ratio)) + geom_point() + geom_smooth()
iso_seasons_avg_MW = iso_seasons_avg[iso_seasons_avg$region%in%c("Mountain West"),]
ggplot(iso_seasons_avg_MW, aes(year,ratio)) + geom_point() + geom_smooth()

### ratio trendline
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
  plot(iso_seasons_avg$year[regions_avg$region==regions[i]&regions_avg$season=="summer"],regions_avg$ratio[regions_avg$region==regions[i]&regions_avg$season=="summer"],las=1,pch=19,ylim=c(0,10),xlim=c(2000,2008),col="red")
  title(regions[i])
}

##Goldstein et al paper: March 2000 to February 2007
iso_0007 = iso_seasons_avg[iso_seasons_avg$year%in%(2000:2007),]
iso_0007 = iso_seasons_avg
iso_0007=as.data.frame(iso_0007)
iso_0007_avg = aggregate
{cbind(seasdiff,water)~site+state+region+lat+lon, data=iso_0007,mean)
rbPal <- colorRampPalette(c('blue','turquoise','green','yellow','orange','red '))
iso_0007_avg$color <- rbPal(70)[as.numeric(cut(iso_0007_avg$seasdiff,breaks = 70))]
iso_0007_avg=iso_0007_avg[order(iso_0007_avg$site,decreasing = T),]
```
iso_0007_avg=data.frame(iso_0007_avg)

color.bar <- function(lut, min, max=-min, nticks=11, ticks=seq(min, max, len=nticks), title='') {
  scale = (length(lut)-1)/(max-min)
  plot(c(0,10), c(min,max), type='n', bty='n', xaxt='n', xlab=''
  yaxt='n', ylab='', main=title)
  axis(4, ticks, las=1)
  for (i in 1:(length(lut)-1)) {
    y = (i-1)/scale + min
    rect(0,y,10,y+1/scale, col=lut[i], border=NA)
  }
}

#### map goldstein

pdf(file="/Users/Khoi/Dropbox/Satellites/Figures/map_goldstein-ALLyearsALLpressure.pdf", width=6.6,height=6)
par(mfrow=c(1,1),mar=c(0,0,0,0),oma=c(0,0,0,3))
map('state', fill = TRUE, col = "grey90")
map.scale(x=-120.5,y=27.5,ratio=F, relwidth=0.2,cex=0.8)
points(x=iso_0007_avg$lon,y=iso_0007_avg$lat,col=iso_0007_avg$color, cex=1.1,pch=19)

par(new=T,mar=c(0,17,0,1),oma=c(10,12,10,2))
is0_0007_avg$scale=round(iso_0007_avg$seasdiff,digits=0)
color.bar(rbPal(50),
min(iso_0007_avg$scale),max=max(iso_0007_avg$scale))
dev.off()

#### line plot

#iso_lineplot = summarySE(iso_0007, measurevar = "water",groupvars = "year")
regions_avg = aggregate(water~region+year+season,data=iso_0007,mean)
regions_avg=regions_avg[order(regions_avg$year),]
regions_avg$year2 = regions_avg$year
regions_avg$year2[regions_avg$season=="summer"] =
regions_avg$year[regions_avg$season=="summer"]+0.5
iso_0007$year2 = iso_0007$year
iso_0007$year2[iso_0007$season=="summer"] =
iso_0007$year[iso_0007$season=="summer"]+0.5
iso_0007=iso_0007[order(iso_0007$year2),]

pdf(file="/Users/Khoi/Dropbox/Satellites/Figures/summer_winter_points_means.pdf", width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
```
```r
plot(regions_avg$year2[regions_avg$region==regions[i]&regions_avg$season=="summer"],regions_avg$water[regions_avg$region==regions[i]&regions_avg$season=="summer"],las=1,pch=19,ylim=c(0,10),xlim=c(2000,2008),col="red")
points(regions_avg$year2[regions_avg$region==regions[i]&regions_avg$season=="winter"],regions_avg$water[regions_avg$region==regions[i]&regions_avg$season=="winter"],las=1,pch=15,col="blue")
title(regions[i])
}
dev.off()

df(file="/Users/Khoi/Dropbox/Satellites/Figures/summer_winter_points_ALL-withtrendline.pdf", width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
plot(iso_0007$year2[iso_0007$region==regions[i]&iso_0007$season=="summer"],iso_0007$water[iso_0007$region==regions[i]&iso_0007$season=="summer"],las=1,pch=19,ylim=c(0,10),xlim=c(2000,2008),col="red")
points(iso_0007$year2[iso_0007$region==regions[i]&iso_0007$season=="winter"],iso_0007$water[iso_0007$region==regions[i]&iso_0007$season=="winter"],las=1,pch=15,col="blue")
loess_fit = predict(loess(iso_0007$water[iso_0007$region==regions[i]]~iso_0007$year[iso_0007$region==regions[i]]))
lines(iso_0007$year2[iso_0007$region==regions[i]],loess_fit,lwd=3)
title(regions[i])
}
dev.off()

df(file="/Users/Khoi/Dropbox/Satellites/Figures/summer_winter_points-withseasonalrendline.pdf", width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
plot(iso_0007$year2[iso_0007$region==regions[i]&iso_0007$season=="summer"],iso_0007$water[iso_0007$region==regions[i]&iso_0007$season=="summer"],las=1,pch=19,ylim=c(0,10),xlim=c(2000,2008),col="red")
points(iso_0007$year2[iso_0007$region==regions[i]&iso_0007$season=="winter"],iso_0007$water[iso_0007$region==regions[i]&iso_0007$season=="winter"],las=1,pch=15,col="blue")
loess_fit = predict(loess(iso_0007$water[iso_0007$region==regions[i]]~iso_0007$year[iso_0007$region==regions[i]]))
loess_fit = predict(loess(iso_0007$water[iso_0007$region==regions[i]]~iso_0007$year[iso_0007$region==regions[i]]))
lines(iso_0007$year2[iso_0007$region==regions[i]&iso_0007$season=="summer"],loess_fit,lwd=3, col="pink")
loess_fit = predict(loess(iso_0007$water[iso_0007$region==regions[i]]~iso_0007$year[iso_0007$region==regions[i]]))
```
ason=="winter"] iso_0007$year[iso_0007$region==regions[i] & iso_0007$season=="winter"]

lines(iso_0007$year2[iso_0007$region==regions[i] & iso_0007$season=="winter"], loess_fit, lwd=3, col="cornflowerblue")

title(regions[i])
}
dev.off()

## START OF ISO_SUB
iso_sub = iso[iso$year<2008,]
#iso_sub = iso_sub[iso_sub$season%in%c("winter", "summer"),]

######### CHANGE THIS FOR OTHER PRESSURES
iso_sub = iso_sub[iso_sub$pressure == 1000,]

######### CHANGE THIS FOR OTHER PRESSURES

iso_sub$year2 = iso_sub$year
iso_sub$month = month(iso_sub$day)
iso_sub$year2[iso_sub$season=="summer"] = iso_sub$year[iso_sub$season=="summer"]+0.5
iso_sub=iso_sub[order(iso_sub$year2),]

par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
  measurevar="water", groupvars=c("year","month"))
  summary_data$time=summary_data$year+summary_data$month/12
  plot(NA,NA, col="white", ylim=c(0,10), yaxt="n",
  ylab="", xaxt="n", xlim=c(2000,2008))
  axis(1, at=c(2000:2008), labels=c(2000:2008), cex.axis=1, las=1)
  axis(2, pretty(c(0,10)), cex.axis=1, las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
  c((summary_data$quartile25), rev(summary_data$quartile75)),
  col="cornflowerblue", border="cornflowerblue")
  lines(summary_data$time,summary_data$median, col="grey35",
  lty=1, lwd=3, cex=1)
  lines(summary_data$time,summary_data$water, col="orange",
  lty=1, lwd=1.5, cex=1)
  mtext(expression(paste("Water", " (µg \left(m^{-3}\right)")), side=2, line=2)
  title(regions[i])
}
dev.off()

par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){

summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
measurevar="water", groupvars=c("year", "season"))
summary_data = summary_data[!is.na(summary_data$year),]
summary_data$time = summary_data$year + c(0.75, 0.25, 0.5, 0)
summary_data= summary_data[order(summary_data$time),]
plot(NA,NA, col="white", ylim=c(0,15), yaxt="n",
ylab="", xlab="Year", xlim=c(2000,2008))
axis(1, at=c(2000:2008), labels=c(2000:2008), cex.axis=1, las=1)
axis(2, pretty(c(0,15)), cex.axis=1, las=1)
polygon(c(summary_data$time, rev(summary_data$time)),
c((summary_data$quartile25), rev(summary_data$quartile75)),
col="cornflowerblue", border="cornflowerblue")
lines(summary_data$time,summary_data$median, col="grey35",
lty=1, lwd=3, cex=1)
lines(summary_data$time,summary_data$water, col="orange",
lty=1, lwd=1.5, cex=1)
mtext("Water", side=2, line=2)
title(regions[i])
}
dev.off()

# mass weighted kappa value and for RH

# RH

par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions))
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
measurevar="RH", groupvars=c("year", "season"))
summary_data = summary_data[!is.na(summary_data$year),]
summary_data$time = summary_data$year + c(0.75, 0.25, 0.5, 0)
summary_data= summary_data[order(summary_data$time),]
plot(NA,NA, col="white", ylim=c(0,1), yaxt="n",
ylab="", xlab="Year", xlim=c(2000,2008))
axis(1, at=c(2000:2008), labels=c(2000:2008), cex.axis=1, las=1)
axis(2, pretty(c(0,1)), cex.axis=1, las=1)
polygon(c(summary_data$time, rev(summary_data$time)),
c((summary_data$quartile25), rev(summary_data$quartile75)),
col="cornflowerblue", border="cornflowerblue")
lines(summary_data$time,summary_data$median, col="grey35",
lty=1, lwd=3, cex=1)
lines(summary_data$time,summary_data$water, col="orange",
lty=1, lwd=1.5, cex=1)
mtext("RH", side=2, line=2)
title(regions[i])
}
dev.off()

# TEMP

par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
                           measurevar="TEMP", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(270,320), yaxt="n",
       xlab="Year", xlim=c(2000,2008))
  axis(1, at=c(2000:2008), labels=c(2000:2008), cex.axis=1, las=1)
  axis(2, cex.axis=1, las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
          c((summary_data$quartile25), rev(summary_data$quartile75)),
          col="cornflowerblue", border="cornflowerblue")
  lines(summary_data$time,summary_data$median, col="grey35",
        lty=1,lwd=3, cex=1)
  lines(summary_data$time,summary_data$TEMP, col="orange",
        lty=1,lwd=1.5, cex=1)
  mtext("TEMP", side=2, line=2)
  title(regions[i])
}
dev.off()

## Sulfate
pdf(file="/Users/Khoi/Dropbox/Satellites/Figures/summarySE-
      SO4byseasonandyear-1000MB.pdf", width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
                           measurevar="SO4", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,8), yaxt="n",
       xlab="Year", xlim=c(2000,2008))
  axis(1, at=c(2000:2008), labels=c(2000:2008), cex.axis=1, las=1)
  axis(2, cex.axis=1, las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
          c((summary_data$quartile25), rev(summary_data$quartile75)),
          col="cornflowerblue", border="cornflowerblue")
  lines(summary_data$time,summary_data$median, col="grey35",
        lty=1,lwd=3, cex=1)
  lines(summary_data$time,summary_data$SO4, col="orange",
        lty=1,lwd=1.5, cex=1)
  mtext("Sulfate", side=2, line=2)
  title(regions[i])
}
dev.off()

## Nitrate
pdf(file="/Users/Khoi/Dropbox/Satellites/Figures/summarySE-
      NO3byseasonandyear-1000MB.pdf", width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){

summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
measurevar="NO3", groupvars=c("year","season"))
summary_data = summary_data[!is.na(summary_data$year),]
summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
summary_data= summary_data[order(summary_data$time),]
plot(NA,NA, col="white", ylim=c(0,3), yaxt="n",
ylab="", xlab="n", xlab="Year", xlab=c(2000,2008))
axis(1, at=c(2000:2008), labels=c(2000:2008),cex.axis=1,las=1)
axis(2, cex.axis=1, las=1)
polygon(c(summary_data$time, rev(summary_data$time)),
c((summary_data$quartile25), rev(summary_data$quartile75)),
col="cornflowerblue", border="cornflowerblue")
lines(summary_data$time,summary_data$median, col="grey35",
lty=1, lwd=3, cex=1)
lines(summary_data$time,summary_data$NO3, col="orange",
lty=1, lwd=1.5, cex=1)
mtext("Nitrate", side=2, line=2)
title(regions[i])
}

dev.off()

## Total INORGANICS
iso_sub$inorg = iso_sub$SO4+iso_sub$NO3
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
measurevar="inorg", groupvars=c("year","season"))
summary_data = summary_data[!is.na(summary_data$year),]
summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
summary_data= summary_data[order(summary_data$time),]
plot(NA,NA, col="white", ylim=c(0,8), yaxt="n",
ylab="", xlab="n", xlab="Year", xlab=c(2000,2008))
axis(1, at=c(2000:2008), labels=c(2000:2008),cex.axis=1,las=1)
axis(2, cex.axis=1, las=1)
polygon(c(summary_data$time, rev(summary_data$time)),
c((summary_data$quartile25), rev(summary_data$quartile75)),
col="cornflowerblue", border="cornflowerblue")
lines(summary_data$time,summary_data$median, col="grey35",
lty=1, lwd=3, cex=1)
lines(summary_data$time,summary_data$IN3, col="orange",
lty=1, lwd=1.5, cex=1)
mtext("Sulfate + Nitrate", side=2, line=2)
title(regions[i])
}

dev.off()

# mass weighted kappa
#Sum eiki eqn
iso_sub$k_index = 0.5*(iso_sub$SO4/(iso_sub$SO4+iso_sub$NO3))+0.7*(iso_sub$NO3/(iso_sub$SO4+iso_sub$NO3))
iso_sub$k_index[which(is.na(iso_sub$k_index))] = 0

iso_sub$k_index = (0.5*1.77*iso_sub$SO4+0.7*1.74*iso_sub$NO3)/(1.77*iso_sub$SO4+1.74*iso_sub$NO3)

iso_sub$k_index = iso_sub$water/(1.8*(iso_sub$SO4+iso_sub$NO3))*((1-is_sub$RH)/iso_sub$RH)
iso_sub$k_index[which(is.na(iso_sub$k_index))] = 0

par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],], measurevar="k_index", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,1), yaxt="n",
   ylab="",xaxt="n", xlab="Year", xlim=c(2000:2008))
  axis(1, at=c(2000:2008), labels=c(2000:2008),cex.axis=1,las=1)
  axis(2, cex.axis=1, las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
  c(summary_data$quartile25, rev(summary_data$quartile75)),
  col="cornflowerblue", border="cornflowerblue")
  lines(summary_data$time,summary_data$median, col="grey35",
     lty=1,lwd=3,cex=1)
  lines(summary_data$time,summary_data$k_index, col="orange",
      lty=1,lwd=1.5,cex=1)
  mtext("Kappa Index", side=2, line=2)
  title(regions[i])
}
dev.off()

library(lubridate)
imp$date <- ymd_hms(imp$Date)
imp <- imp[order(imp$date),]
imp$year = year(imp$date)

library(maps)
library(mapproj)
library(mapplots)

Northeast <- c("ME", "NH", "VT", "MA", "RI", "CT", "NY", "NJ", "DE", "PA")
Southeast <-
Central <- c("TX", "OK", "MN", "IA", "MO", "KS", "NE")
Midwest <- c("OH", "IN", "IL", "MI", "WI", "ND", "SD")
MountainWest <- c("AZ", "NM", "MT", "ID", "NV", "UT", "WY", "CO")
WestPacific <- c("WA", "OR", "CA")

imp$region=NA
imp$region[imp$State%in% Northeast]= "Northeast"
imp$region[imp$State%in% Southeast]= "Southeast"
imp$region[imp$State%in% Central]= "Central"
imp$region[imp$State%in% Midwest]= "Midwest"
imp$region[imp$State%in% WestPacific]= "West Pacific"
imp$region[imp$State%in% MountainWest]= "Mountain West"

imp$season[month(imp$date)%in%c(12,1,2)] = "winter"
imp$season[month(imp$date)%in%c(3,4,5)] = "spring"
imp$season[month(imp$date)%in%c(6,7,8)] = "summer"
imp$season[month(imp$date)%in%c(9,10,11)] = "fall"
imp_seasons = imp[imp$season%in%c("winter","summer"),]

#mean for summer and median for winter for OM
imp_seasons_avg = aggregate(cbind(OMCf.Value)~SiteCode+State+season+region+year+Latitude+Longitude, data=imp_seasons,mean)
imp_seasons_avg = imp_seasons_avg[order(imp_seasons_avg$SiteCode, imp_seasons_avg$year),]
imp_seasons_avg$seasdiff=NA
for (i in seq(2,nrow(imp_seasons_avg),2)){
  if (as.vector(imp_seasons_avg$State[i+1])%in%as.vector(imp_seasons_avg$State[i])){
    imp_seasons_avg$seasdiff[i] =
    abs(as.numeric(imp_seasons_avg$OMCf.Value[i+1])-as.numeric(imp_seasons_avg$OMCf.Value[i]))
  }
}

color.bar <- function(lut, min, max=-min, nticks=11, ticks=seq(min, max, len=nticks), title='') {
  scale = (length(lut)-1)/(max-min)
  plot(c(0,10), c(min,max), type='n', bty='n', xaxt='n', xlab='', yaxt='n', ylab='', main=title)
  axis(4, ticks, las=1)
  for (i in 1:(length(lut)-1)) {
    y = (i-1)/scale + min
    rect(0,y,10,y+1/scale, col=lut[i], border=NA)
  }
}
imp_0007 = imp_seasons_avg[imp_seasons_avg$year%in%(2000:2007),]
imp_0007=as.data.frame(imp_0007)
imp_0007_avg = aggregate
  (cbind(seasdiff,OMCf.Value)~SiteCode+State+region+Latitude+Longitude,
  data=imp_0007,mean)

imp_0007_avg.2 = aggregate
  (cbind(seasdiff,OMCf.Value)~SiteCode+State+region+Latitude+Longitude,
  data=imp_seasons_avg,mean)

## getting color scales the same
imp_0007_avg$letter = c("a")
imp_0007_avg.2$letter = c("b")

org_data = rbind(imp_0007_avg,imp_0007_avg.2)
rbPal <-
colorRampPalette(c('blue','turquoise','green','yellow','orange','red'))
org_data$color = rbPal(70)[as.numeric(cut(org_data$seasdiff,breaks =
70))]

iso_0007_avg.a = org_data[org_data$letter=="a",]
iso_0007_avg.b = org_data[org_data$letter=="b",]

pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/F3.pdf",
width=6.6,height=6)
## a 2000-2007
par(mfrow=c(2,1),mar=c(0,0,0,0),oma=c(0,0,0,3))
map('state', fill = TRUE, col = "grey90")
#map.scale(x=-120.5,y=27.5,ratio=F, relwidth=0.2,cex=0.8)
points(x=iso_0007_avg.a$Longitude,y=iso_0007_avg.a$Latitude,col=iso_0007_avg.a$color,cex=1.1,pch=19)
## b ALL years
map('state', fill = TRUE, col = "grey90")
points(x=iso_0007_avg.b$Longitude,y=iso_0007_avg.b$Latitude,col=iso_0007_avg.b$color,cex=1.1,pch=19)
par(new=T,mar=c(0,17,0,1),oma=c(10,12,2,2))
org_data$scale=round(org_data$seasdiff,digits=0)
color.bar(rbPal(70), min(org_data$scale),max=max(org_data$scale))
dev.off()

## time series of OM
#pdf(file="/Users/Khoi/Dropbox/Satellites/Figures/OM.byseasonandyear.pdf",
width=6.6,height=6)
pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/F4_v.2.pdf",
width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,5,5,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(imp_sub[imp_sub$region==regions[i],],
    measurevar="OMCf.Value", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,5), yaxt="n", ylab="",xaxt="n",
   xlab="Year", xlim=c(2000,2008))
  axis(1, at=c(2000:2008), labels=c(2000:2008),cex.axis=1,las=1)
  axis(2, cex.axis=1, las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
    c((summary_data$quartile25), rev(summary_data$quartile75)),
    col="grey75", border="grey75")
  lines(summary_data$time,summary_data$median, col="black",
    lty=1,lwd=3,cex=1)
  # lines(summary_data$time,summary_data$OMCf.Value, col="orange",
  #  lty=1,lwd=1.5,cex=1)
  mtext("OMCf.Value", side=2, line=2)
  title(regions[i])
}
dev.off()

### RH and nitrate

par(mar=c(2.5,2,1,5), mfrow = c(3,2), oma=c(2,5,5,3))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
    measurevar="NO3", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,3), yaxt="n", ylab="",xaxt="n",
   xlab="Year", xlim=c(2000,2008))
  axis(1, at=c(2000:2008), labels=c(2000:2008),cex.axis=1,las=1)
  axis(2, cex.axis=1, las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
    c((summary_data$quartile25), rev(summary_data$quartile75)),
    col="darkblue", border="darkblue")
  lines(summary_data$time,summary_data$median, col="grey65",
    lty=1,lwd=3,cex=1)
  # lines(summary_data$time,summary_data$NO3, col="orange",
  #  lty=1,lwd=1.5,cex=1)
  mtext("Nitrate", side=2, line=2.5,col="darkblue")
  title(regions[i])
  par(new=T)
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
    measurevar="RH", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
summary_data= summary_data[order(summary_data$time),]
plot(NA,NA, col="white", ylim=c(0,1), yaxt="n", ylab="", xaxt="n",
xlab="Year", xlim=c(2000,2008))
axis(4,pretty(c(0,1)), cex.axis=1, las=1)
polygon(c(summary_data$time, rev(summary_data$time)),
c("grey65", border="cornflowerblue"),
lines(summary_data$time,summary_data$median, col="grey35",
lty=1,lwd=3,cex=1)
lines(summary_data$time,summary_data$RH, col="orange",
lty=1,lwd=1.5,cex=1)
mtext("RH", side=4, line=2.5,col="grey65")
}
dev.off()

## RH and kappa

pdf(file="/Users/Khoi/Dropbox/Satellites/Figures/summarySE-
 kappaandRH-1000MB.pdf", width=6.6,height=6)
par(mar=c(2.5,2,1,5), mfrow = c(3,2), oma=c(2,5,5,3))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
measurevar="k_index", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0.45,1), yaxt="n",
ylab="", xaxt="n", xlab="Year", xlim=c(2000,2008))
  axis(1, at=c(2000:2008), labels=c(2000:2008), cex.axis=1, las=1)
  axis(2, cex.axis=1, las=1)
polygon(c(summary_data$time, rev(summary_data$time)),
c("cornflowerblue", border="cornflowerblue"),
lines(summary_data$time,summary_data$median, col="grey35",
lty=1,lwd=3,cex=1)
lines(summary_data$time,summary_data$k_index, col="orange",
lty=1,lwd=1.5,cex=1)
mtext("Kappa Index", side=2, line=2.5, col="cornflowerblue")
title(regions[i])
}
par(new=T)
summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
measurevar="RH", groupvars=c("year","season"))
summary_data = summary_data[!is.na(summary_data$year),]
summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
summary_data= summary_data[order(summary_data$time),]
plot(NA,NA, col="white", ylim=c(0,1), yaxt="n", ylab="", xaxt="n",
xlab="Year", xlim=c(2000,2008))
axis(4,pretty(c(0,1)), cex.axis=1, las=1)
polygon(c(summary_data$time, rev(summary_data$time)),
c("grey65", border="cornflowerblue")
}
lines(summary_data$time, summary_data$median, col="grey35",
lty=1, lwd=3, cex=1)
lines(summary_data$time, summary_data$RH, col="orange",
lty=1, lwd=1.5, cex=1)
mtext("RH", side=4, line=2.5, col="grey65")
}
dev.off()

### ### ### ###
### WALKER BRANCH STUFF
### ### ### ###
walker <- read.csv("~/Documents/NARR/Satellite/970101_091231_Walker_Branch.lev20", header=T, na.strings="N/A", skip=4)
library(lubridate)
walker$date = dmy_hms(paste(walker$Date.dd.mm.yy., walker$Time.hh.mm.ss.))
walker$year = year(walker$date)
iso$year = year(iso$day)
se = iso[iso$region%in%c("Southeast"),]

##### boxplots of AOT(440 nm), SE Water, Solar Zenith Angle
pdf(file="/Users/Khoi/Dropbox/Satellites/Figures/boxplot_AOT_WaterSE.pdf", width=6.6, height=6)
par(mar=c(3,5,1,2), mfrow=c(3,1), oma=c(1,1,1,1))
boxplot(AOT_440~year, data=walker, las=1, ylim=c(0,0.5), ylab="AOT (440 nm)", varwidth=T)
boxplot(water~year, data=se, las=1, ylim=c(0,10), ylab="SE Water", xlim=c(12,22), varwidth=T)
boxplot(Solar_Zenith_Angle~year, data=walker, las=1, ylab="Solar Zenith Angle", varwidth=T, ylim=c(40,80))
dev.off()

## Correlation between AOT and water
walker$month = month(walker$date)
walker$season[month(walker$date)%in%c(12,1,2)] = "winter"
walker$season[month(walker$date)%in%c(3,4,5)] = "spring"
walker$season[month(walker$date)%in%c(6,7,8)] = "summer"
walker$season[month(walker$date)%in%c(9,10,11)] = "fall"
walker.avg = aggregate(AOT_440~year+season, data=walker, FUN=mean)
walker.avg$year2 = NA
walker.avg$year2[walker.avg$season == "winter"] = walker.avg$year[walker .avg$season == "winter"]
walker.avg$year2[walker.avg$season == "spring"] = walker.avg$year[walker .avg$season == "spring"] + 0.25
walker.avg$year2[walker.avg$season == "summer"] = walker.avg$year[walker .avg$season == "summer"] + 0.5
walker.avg$year2[walker.avg$season=="fall"] = walker.avg$year[walker.avg$season=="fall"] + 0.75

### numbers

iso_sub = iso[iso$year%in% c(2000:2007),]
iso_sub = iso_sub[iso_sub$season%in% c("winter","summer"),]

summary(iso_sub$water[iso_sub$region=="Northeast"])  
summary(iso_sub$water[iso_sub$region=="Southeast"])  
summary(iso_sub$water[iso_sub$region=="Midwest"])  
summary(iso_sub$water[iso_sub$region=="Central"])  
summary(iso_sub$water[iso_sub$region=="Mountain West"])  
summary(iso_sub$water[iso_sub$region=="West Pacific"])  

##### Figures for paper

#### Seasonal differences map

```r
color.bar <- function(lut, min, max=-min, nticks=11, ticks=seq(min, max, len=nticks), title='') {
  scale = (length(lut)-1)/(max-min)
  plot(c(0,10), c(min,max), type='n', bty='n', xaxt='n', xlab='', yaxt='n', ylab='', main=title)
  axis(4, ticks, las=1)
  for (i in 1:(length(lut)-1)) {
    y = (i-1)/scale + min
    rect(0,y,10,y+1/scale, col=lut[i], border=NA)
  }
}
rbPal <- colorRampPalette(c('blue','turquoise','green','yellow','orange','red'))
```

### (a) 1000 mb from 2000-2007

iso_seasons = iso[iso$season%in%c("winter","summer"),]
iso_seasons.a = iso_seasons[iso_seasons$pressure==1000,]
iso_seasons.a = iso_seasons.a[iso_seasons.a$year%in%c(2000:2007),]

#mean for summer and median for winter
iso_seasons_avg = aggregate
  (cbind(water)~site+state+season+region+year, 
   data=iso_seasons.a,mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site, iso_seasons_avg$year),]

iso_seasons_avg$seasdiff = NA
for (i in seq(2, nrow(iso_seasons_avg), 2)) {
  if (as.vector(iso_seasons_avg$state[i+1]) %in% as.vector(iso_seasons_avg$state[i])){
    iso_seasons_avg$seasdiff[i] = abs(as.numeric(iso_seasons_avg$water[i+1]) - as.numeric(iso_seasons_avg$water[i]))
  }
}

iso_seasons_avg$ratio = NA
for (i in seq(2, nrow(iso_seasons_avg), 2)) {
  if (as.vector(iso_seasons_avg$state[i+1]) %in% as.vector(iso_seasons_avg$state[i])){
    iso_seasons_avg$ratio[i] = (as.numeric(iso_seasons_avg$water[i-1])/as.numeric(iso_seasons_avg$water[i]))
  }
}

for (i in 1:nrow(iso_seasons_avg)) {
  iso_seasons_avg$lat[i] = (improvesites$lat[improvesites$SiteCode %in% iso_seasons_avg$site[i]])
  iso_seasons_avg$lon[i] = (improvesites$lon[improvesites$SiteCode %in% iso_seasons_avg$site[i]])
}

iso_0007 = iso_seasons_avg
iso_0007 = as.data.frame(iso_0007)
iso_0007_avg = aggregate(cbind(seasdiff, water) ~ site + state + region + lat + lon, data = iso_0007, mean)
iso_0007_avg$color <- rbPal(70)[as.numeric(cut(iso_0007_avg$seasdiff, breaks = 70))]
iso_0007_avg = iso_0007_avg[order(iso_0007_avg$site, decreasing = T),]
iso_0007_avg = data.frame(iso_0007_avg)

iso_0007_avg.a = iso_0007_avg

## (b) 1000 mb from 1998-2013
iso_seasons = iso[iso$season %in% c("winter", "summer"),]
iso_seasons.b = iso_seasons[iso_seasons$pressure == 1000,]

# mean for summer and median for winter
iso_seasons_avg = aggregate(cbind(water) ~ site + state + season + region + year, data = iso_seasons.b, mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site, iso_seasons_avg$year),]
iso_seasons_avg$seasdiff = NA
for (i in seq(2, nrow(iso_seasons_avg), 2)) {
if (as.vector(iso_seasons_avg$state[i+1]) %in% as.vector(iso_seasons_avg$state[i])){
    iso_seasons_avg$seasdiff[i] = abs(as.numeric(iso_seasons_avg$water[i+1]) - as.numeric(iso_seasons_avg$water[i]))
}
}

iso_seasons_avg$ratio = NA
for (i in seq(2,nrow(iso_seasons_avg),2)){
    if (as.vector(iso_seasons_avg$state[i+1]) %in% as.vector(iso_seasons_avg$state[i])){
        iso_seasons_avg$ratio[i] = (as.numeric(iso_seasons_avg$water[i-1]) / as.numeric(iso_seasons_avg$water[i]))
    }
}

for (i in 1:nrow(iso_seasons_avg)){
    iso_seasons_avg$lat[i] = (improvesites$lat[improvesites$SiteCode %in% iso_seasons_avg$site[i]])
    iso_seasons_avg$lon[i] = (improvesites$lon[improvesites$SiteCode %in% iso_seasons_avg$site[i]])
}

iso_0007 = iso_seasons_avg
iso_0007 = as.data.frame(iso_0007)
iso_0007_avg = aggregate(cbind(seasdiff,water) ~ site + state + region + lat + lon,
data = iso_0007, mean)
iso_0007_avg$color <- rbPal(70)[as.numeric(cut(iso_0007_avg$seasdiff, breaks = 70))]
iso_0007_avg = iso_0007_avg[order(iso_0007_avg$site, decreasing = T),]
iso_0007_avg = data.frame(iso_0007_avg)
iso_0007_avg.b = iso_0007_avg

## (c) 800-1000 mb from 2000-2007
iso_seasons = iso[iso$season %in% c("winter","summer"),]
iso_seasons.c = iso_seasons[iso_seasons$pressure %in% c(800:1000),]
iso_seasons.c = iso_seasons.c[iso_seasons.c$year %in% c(2000:2007),]

# mean for summer and median for winter
iso_seasons_avg = aggregate(cbind(water) ~ site + state + season + region + year,
data = iso_seasons.c, mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site, iso_seasons_avg$year),]
iso_seasons_avg$seasdiff = NA
for (i in seq(2, nrow(iso_seasons_avg), 2)){
    if (as.vector(iso_seasons_avg$state[i+1]) %in% as.vector(iso_seasons_avg$state[i])){
        iso_seasons_avg$seasdiff[i] = abs(as.numeric(iso_seasons_avg$water[i+1]) - as.numeric(iso_seasons_avg$water[i]))
    }
}
iso_seasons_avg$seasdiff[i] =
abs(as.numeric(iso_seasons_avg$water[i+1])-
as.numeric(iso_seasons_avg$water[i]))
    }
}

iso_seasons_avg$ratio = NA
for (i in seq(2,nrow(iso_seasons_avg),2)){
    if
        (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$
state[i])){
            iso_seasons_avg$ratio[i] = (as.numeric(iso_seasons_avg$water[i-1]
)/as.numeric(iso_seasons_avg$water[i]))
    }
}
for (i in 1:nrow(iso_seasons_avg)){
    iso_seasons_avg$lat[i]=(improvesites$lat[improvesites$SiteCode%in%is
o_seasons_avg$site[i]])
    iso_seasons_avg$lon[i]=(improvesites$lon[improvesites$SiteCode%in%is
o_seasons_avg$site[i]])
}
iso_0007 = iso_seasons_avg
iso_0007=as.data.frame(iso_0007)
iso_0007_avg = aggregate
    (cbind(seasdiff,water)~site+state+region+lat+lon,
data=iso_0007,mean)
iso_0007_avg=color <-
rPalm(70)[as.numeric(cut(iso_0007_avg$seasdiff,breaks = 70))]
iso_0007_avg=iso_0007_avg[order(iso_0007_avg$site,decreasing = T),]
iso_0007_avg=data.frame(iso_0007_avg)
iso_0007_avg.c = iso_0007_avg

## (d) 800-1000 mb from 1998-2013
iso_seasons = iso[iso$season%in%c("winter","summer"),]
iso_seasons.d = iso_seasons[iso_seasons$pressure%in%c(800:1000),]

#mean for summer and median for winter
iso_seasons_avg = aggregate
    (cbind(water)~site+state+season+region+year,
data=iso_seasons.d,mean)
iso_seasons_avg = iso_seasons_avg[order(iso_seasons_avg$site,
    iso_seasons_avg$year),]
iso_seasons_avg$seasdiff=NA
for (i in seq(2,nrow(iso_seasons_avg),2)){
    if
        (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$state[i])){
            iso_seasons_avg$seasdiff[i] =
abs(as.numeric(iso_seasons_avg$water[i+1])-as.numeric(iso_seasons_avg$water[i]))
        }
    }
}
iso_seasons_avg$ratio = NA
for (i in seq(2,nrow(iso_seasons_avg),2)){
  if (as.vector(iso_seasons_avg$state[i+1])%in%as.vector(iso_seasons_avg$state[i])){
    iso_seasons_avg$ratio[i] = (as.numeric(iso_seasons_avg$water[i-1])/as.numeric(iso_seasons_avg$water[i]))
  }
}

for (i in 1:nrow(iso_seasons_avg)){
iso_seasons_avg$lat[i]=(improvesites$lat[improvesites$SiteCode%in%iso_seasons_avg$site[i]])
iso_seasons_avg$lon[i]=(improvesites$lon[improvesites$SiteCode%in%iso_seasons_avg$site[i]])
}

iso_0007 = iso_seasons_avg
iso_0007=as.data.frame(iso_0007)
iso_0007_avg = aggregate(cbind(seasdiff,water)~site+state+region+lat+lon, 
data=iso_0007,mean)
iso_0007_avg$color <- rbPal(70)[as.numeric(cut(iso_0007_avg$seasdiff,breaks = 70))]
iso_0007_avg=iso_0007_avg[order(iso_0007_avg$site,decreasing = T),]
iso_0007_avg = data.frame(iso_0007_avg)
iso_0007_avg.d = iso_0007_avg

## getting color scales the same
iso_0007_avg.a$letter = c("a")
iso_0007_avg.b$letter = c("b")
iso_0007_avg.c$letter = c("c")
iso_0007_avg.d$letter = c("d")

seasdiff_data = rbind(iso_0007_avg.a,iso_0007_avg.b,iso_0007_avg.c,iso_0007_avg.d)

seasdiff_data$color = rbPal(70)[as.numeric(cut(seasdiff_data$seasdiff,breaks = 70))]

iso_0007_avg.a = seasdiff_data[seasdiff_data$letter=="a",]
iso_0007_avg.b = seasdiff_data[seasdiff_data$letter=="b",]
iso_0007_avg.c = seasdiff_data[seasdiff_data$letter=="c",]
iso_0007_avg.d = seasdiff_data[seasdiff_data$letter=="d",]

### the overall plot
pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/F1.pdf", 
width=8,height=5)

cex_n = 0.7
par(mfrow=c(2,2),mar=c(0,0,0,0),oma=c(0,0,0,5))

map('state', fill = TRUE, col = "grey90", lwd=0.6)
points(x=iso_0007_avg.a$lon,y=iso_0007_avg.a$lat,col=iso_0007_avg.a$color,cex=cex_n,pch=19)
map('state', fill = TRUE, col = "grey90", lwd=0.6)
points(x=iso_0007_avg.b$lon,y=iso_0007_avg.b$lat,col=iso_0007_avg.b$color,cex=cex_n,pch=19)
map('state', fill = TRUE, col = "grey90", lwd=0.6)
points(x=iso_0007_avg.c$lon,y=iso_0007_avg.c$lat,col=iso_0007_avg.c$color,cex=cex_n,pch=19)
map('state', fill = TRUE, col = "grey90", lwd=0.6)
points(x=iso_0007_avg.d$lon,y=iso_0007_avg.d$lat,col=iso_0007_avg.d$color,cex=cex_n,pch=19)
#map.scale(x=-120.5,y=27.5,ratio=F, relwidth=0.2,cex=0.8)
par(new=T,mfrow=c(1,1),mar=c(0,17,0,1),oma=c(7,19,7,2))
seasdiff_data$scale=round(seasdiff_data$seasdiff,digits=0)
color.bar(rbPal(50),
min(seasdiff_data$scale),max=max(seasdiff_data$scale))
dev.off()

########
### Quartiles in aerosol water
########
iso_sub=iso
#iso_sub = iso[iso$year%in%c(2000:2008),]
#iso_sub = iso_sub[iso_sub$pressure == 1000,]
pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/F2.pdf",
width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,4,2,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
measurevar="water", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,15), yaxt="n", ylab="",xaxt="n",
  xlab="Year", xlim=c(2000,2008))
  axis(1, at=c(2000:2008), labels=c(2000:2008),cex.axis=1,las=1)
  axis(2, pretty(c(0,15)), cex.axis=1, las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
  c((summary_data$quartile25), rev(summary_data$quartile75)),
  col="cornflowerblue", border="cornflowerblue")
  lines(summary_data$time,summary_data$median, col="black",
  lty=1,lwd=3,cex=1)
  # lines(summary_data$time,summary_data$water, col="orange",
  lty=1,lwd=1.5,cex=1)
  title(regions[i])
}
par(new=T,mfrow=c(1,1),mar=c(0,3,0,0),oma=c(0,0,0,0))
plot (NA,NA, xlim = c(0,1), ylim=c(0,1), yaxt="n", xaxt="n",
frame.plot=F)
mtext(expression(paste("Water", " (\mu g \, m^{-3}, ")", sep="" ))),
side=2, line=1, cex= 1.3)

dev.off()

########
## OM map & quartiles
########

imp_sub = imp[!is.na(imp$OMCf.Value),]

pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/F4.pdf",
width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,4,2,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(imp_sub[imp_sub$region==regions[i],],
measurevar="OMCf.Value", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,6), ylab="",
lab="Year",las=1,xlim=c(1988,2013))
  polygon(c(summary_data$time, rev(summary_data$time)),
c((summary_data$quartile25), rev(summary_data$quartile75)),
col="lightgreen", border="lightgreen")
  lines(summary_data$time,summary_data$median, col="black",
lty=1,lwd=3,cex=1)
  #   lines(summary_data$time,summary_data$OMCf.Value, col="orange",
lty=1,lwd=1.5,cex=1)
  #   mtext("OMCf.Value", side=2, line=2)
  title(regions[i])
}
par(new=T,mfrow=c(1,1),mar=c(0,3,0,0),oma=c(0,0,0,0))
plot (NA,NA, xlim = c(0,1), ylim=c(0,1), yaxt="n", xaxt="n",
frame.plot=F)
mtext(expression(paste("Organic Matter", " (\mu g \, m^{-3}, ")", sep="" ))),
side=2, line=1, cex= 1.3)

dev.off()

########
## Quartiles for sulfate + nitrate
########

iso_sub$inorg = iso_sub$SO4+iso_sub$NO3

pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/F5.pdf",
width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,4,2,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
               measurevar="inorg", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,10), ylab="",
       xlim=c(1988,2013), las=1)
  polygon(c(summary_data$time, rev(summary_data$time)),
          c((summary_data$quartile25), rev(summary_data$quartile75)),
          col="indianred1", border="indianred1")
  lines(summary_data$time,summary_data$median, col="blue3",
        lty=1,lwd=3,cex=1)
  # lines(summary_data$time,summary_data$inorg, col="orange",
  #       lty=1,lwd=1.5,cex=1)
  # mtext("Sulfate + Nitrate", side=2, line=2)
  title(regions[i])
}
par(new=T,mfrow=c(1,1),mar=c(0,3,0,0),oma=c(0,0,0,0))
plot (NA,NA, xlim = c(0,1), ylim=c(0,1), yaxt="n", ylab="",
      xaxt="n",xlab="Year", xlim=c(2000,2008))
axis(1, at=c(2000:2008), labels=c(2000:2008),cex.axis=1,las=1)
axis(2, pretty(c(0,1)), cex.axis=1, las=1)
polygon(c(summary_data$time, rev(summary_data$time)),
       c((summary_data$quartile25), rev(summary_data$quartile75)),
       col="grey75", border="grey75")
lines(summary_data$time,summary_data$median, col="black",
      lty=1,lwd=3,cex=1)
# lines(summary_data$time,summary_data$RH, col="orange",
#       lty=1,lwd=1.5,cex=1)
# mtext("RH", side=2, line=2)

dev.off()

########
## quartiles for RH
########

pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/F5.pdf",
    width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,4,2,1))
for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[iso_sub$region==regions[i],],
               measurevar="RH", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,1), yaxt="n", xaxt="n",
       frame.plot=F,ylab="",xlab="")
  mtext(expression(paste("Total Inorganics", " (µg m^{-3}, "
              sep="" ))), side=2, line=1, cex= 1.3)
}

par(new=T,mfrow=c(1,1),mar=c(0,3,0,0),oma=c(0,0,0,0))
plot (NA,NA, xlim = c(0,1), ylim=c(0,1), yaxt="n", ylab="",
      xaxt="n")
mtext(expression(paste("Total Inorganics", " (µg m^{-3}, "
              sep="" ))), side=2, line=1, cex= 1.3)

dev.off()
title(regions[i])
}

par(new=T,mfrow=c(1,1),mar=c(0,3,0,0),oma=c(0,0,0,0))
plot (NA,NA, xlim = c(0,1), ylim=c(0,1), yaxt="n", xaxt="n",
frame.plot=F)
mtext(expression(paste("RH", sep="" ))), side=2, line=1, cex= 1.3)

dev.off()

# f(RH) F= b0+b1(100/(100-RH))+b2(100/(100-RH))^2
# bext = 3*f(RH)*[sulfate]+3*f(RH)*[nitrate]+4*[OM]+10*[EC]+1*[Fine Soil]+0.6*[Coarse Mass]+10

extinction_data <- read.csv("~/Documents/NARR/extinction_data.txt", na.strings = c(NA))
iso = read.csv("/Users/Khoi/Dropbox/Satellites/iso-surface-withwater.csv")

library(lubridate)
extinction_data$day = ymd(extinction_data$Date)
extinction_data = extinction_data[order(extinction_data$day,extinction_data$SiteCode),]

ext = extinction_data_avg[paste(extinction_data_avg$SiteCode,as.character(extinction_data_avg$day))%in%paste(iso$site,as.character(iso$day)),]
iso_2 = iso[paste(iso$site,as.character(iso$day))%in%paste(ext$SiteCode,as.character(ext$day))]

nrow(ext)
nrow(iso_2)

ext = ext[order(ext$day,ext$SiteCode),]
iso_2 = iso_2[order(iso_2$day,iso_2$site),]

ext$RH = iso_2$RH
ext$region = iso_2$region
ext$season = iso_2$season
ext = ext[ext$RH<=0.98,]
ext$f_RH = 0.33713+0.58601*(100/(100-100*ext$RH))+0.09164*(100/(100-100*ext$RH))^2
#units = Mm^-1
ext$bext = 3*ext$f_RH*ext$SO4f.Value+3*ext$f_RH*ext$NO3f.Value+4*1.8*ext$OCf.Value+10*ext$ECf.Value+1*ext$SOILf.Value+0.6*ext$CM_calculated.Value+10
# quartiles for bext

regions = c("West Pacific", "Northeast", "Mountain West", "Southeast", "Central", "Midwest")
iso_sub = ext
iso_sub$year = year(iso_sub$day)
iso_sub$year2 = iso_sub$year
iso_sub$month = month(iso_sub$day)
iso_sub$year2[iso_sub$season=="summer"] =
iso_sub$year[iso_sub$season=="summer"]+0.5
iso_sub=iso_sub[order(iso_sub$year2),]

summarySE <- function(data=NULL, measurevar, groupvars=NULL,
na.rm=FALSE, conf.interval=.95, .drop=TRUE) {
require(plyr)
  # New version of length which can handle NA's: if na.rm==T, don't count them
  length2 <- function (x, na.rm=FALSE) {
    if (na.rm) sum(!is.na(x))
    else       length(x)
  }
  # This is does the summary
  datac <- ddply(data, groupvars, .drop=.drop,
    .fun= function(xx, col, na.rm) {
      c( N    = length2(xx[[col]], na.rm=na.rm),
        mean = mean   (xx[[col]], na.rm=na.rm),
        sd   = sd     (xx[[col]], na.rm=na.rm),
        median = median(xx[[col]], na.rm=na.rm),
        quartile25 = as.vector(summary(xx[[col]],
                na.rm=na.rm))[[2]],
        quartile75 = as.vector(summary(xx[[col]],
                na.rm=na.rm))[[5]]
    },
    measurevar,
    na.rm
  )
  # Rename the "mean" column
  datac <- rename(datac, c("mean"=measurevar))
  datac$se <- datac$sd / sqrt(datac$N)  # Calculate standard error of the mean
  # Confidence interval multiplier for standard error
  # Calculate t-statistic for confidence interval:
  # e.g., if conf.interval is .95, use .975 (above/below), and use
  df=N-1
  ciMult <- qt(conf.interval/2 + .5, datac$N-1)
  datac$ci <- datac$se * ciMult
  return(datac)
}
pdf(file="/Users/Khoi/Dropbox/Papers/Satellites/Figures/bext_timeseries.pdf", width=6.6,height=6)
par(mar=c(2.5,2,1,2), mfrow = c(3,2), oma=c(2,4,2,1))

for (i in 1:length(regions)){
  summary_data <- summarySE(iso_sub[ext$region==regions[i],],
  measurevar="bext", groupvars=c("year","season"))
  summary_data = summary_data[!is.na(summary_data$year),]
  summary_data$time=summary_data$year+c(0.75, 0.25, 0.5, 0)
  summary_data= summary_data[order(summary_data$time),]
  plot(NA,NA, col="white", ylim=c(0,150), ylab="",
  xlab="Year",las=1,xlim=c(1988,2013))
  polygon(c(summary_data$time, rev(summary_data$time)),
  c((summary_data$quartile25), rev(summary_data$quartile75)),
  col="grey75", border="grey75")
  lines(summary_data$time,summary_data$median, col="black",
  lty=1,lwd=3,cex=1)
  # lines(summary_data$time,summary_data$RH, col="orange",
  # lty=1,lwd=1.5,cex=1)
  # mtext("RH", side=2, line=2)
  title(regions[i])
}

par(new=T,mfrow=c(1,1),mar=c(0,3,0,0),oma=c(0,0,0,0))
plot(NA,NA, xlim = c(0,1), ylim=c(0,1), yaxt="n", xaxt="n",
frame.plot=F)
mtext(expression(paste("Extinction Coefficient (1/Mm)", sep=" " )))
side=2, line=1, cex= 1.3)

dev.off()