QUANTITATIVE ANALYSIS OF SECONDARY ORGANIC AEROSOL USING A HIGH-RESOLUTION MASS SPECTRAL DATABASE

William James Dixon

PHD

UNIVERSITY OF YORK

CHEMISTRY

NOVEMBER 2018

Abstract

Organic aerosol is an important factor in the atmospheric pollution, contributing up to 90% of the total aerosol. The complex system of sources and pathways of functionalisation these compounds can undergo, lead to a large number of compounds which exist in the organic fraction. The state of this organic fraction is in flux, due to the varying contributions of sources and the oxidative aging of the air mass.

A purpose-built, UPLC-MS^2 method, combined with a semi-targeted automated processing software was developed to compliment large datasets required for high temporal resolution. This method was validated using previous analysis of organic aerosol from London and Birmingham.

The organic fraction in Beijing, China was observed from high resolution off-line filter samples during two field campaigns during winter 2016 and summer 2017. In total 646 compounds were identified during these campaigns, of which, the nitrophenolics and organosulfates were observed to be the most important groups. Hierarchical cluster analyses were applied to these two groups to determine the most important clusters within these heteroatom groups.

43 nitrophenolics were observed in the organic aerosol and 191 organosulfates. Nitrophenolics observed by this method show very high variation in ionisation efficiency based on structural isomerism, with ionisation efficiency varying of factors up to three orders of magnitude. The most significant organosulfates are more locally sourced to the Beijing area.

Table of contents

List of tables

List of figures

Figure 2.3: Gradient of elution for UPLC analysis with longer time for gradient change in eluent composition. Length of cleaning time is half high eluent B, half a gradient return to high eluent A conditions. Eluent A: UPLC grade H_2O, with 0.1% by volume of formic acid. Eluent B: 100% UPLC grade MeOH.

Figure 2.4: Base peak chromatogram for the shorter length gradient. BPC shows most abundant peaks from ambient sample from EROS campaign.

Figure 2.5: Base peak chromatogram for the longer length gradient. BPC shows most abundant peaks from ambient sample from EROS campaign, including peaks which were not observed in the shorter gradient method BPC (particularly at the later retention time).

Figure 2.6: Base peak chromatogram for the longer length gradient. BPC shows most abundant peaks from ambient sample from ClearfLo campaign.

Figure 2.7: Ion signal for three (pictured) of the five standard chemicals from direct injection. Each vertical, black line represents a change in ESI settings to induce a change in the ion signal. Each vertical red line separation shows a different setting change, the settings varied are as follows: Auxiliary flow rate, Sweep flow rate, S Lens Angle, Auxiliary gas temperature, Capillary temperature.

Figure 2.8: A simplified flow diagram of the base process that the SAPA program performed when processing the data read from Tracerfinder. Green elements show plots which are produced, orange shows external data read by the software, purple elements show processes and the blue element shows user-defined manual inputs.

Figure 3.1: Calibration curves for two standards used to quantify the filter analysis. 95% confidence interval is shaded around the average trend line (in blue). Low variation across the analysis batches is seen in the narrow shaded area. Average relative standard deviation of gradients from this calibration is 4.03%.

Figure 3.2: (Upper) Chromatogram of an ambient winter aerosol sample, dominated by a large peak around the dead volume at 0.69 mins. (Lower) Chromatogram of an ambient summwe aerosol sample, dominated by a large peak around the dead volume at 0.69 mins and a large broad peak between 0.75 - 5 mins.

Figure 3.3: Mass spectra of m/z 251.00 – 251.10, multiple ions can be observed in this mass range, mostly consisting of organosulfates.

Figure 3.4: Van Krevelen plots for the averaged winter and summer filter sample respectively. Each point is colour coded by its heteroatom group and is sized proportionally by its abundance in standardised peak area (peak area m^{-3}) in the sample. In both plots, a standard point of 10,000,000 peak area m^{-3} is inserted in pink at O:C ratio = 3, H:C ratio = 0 in order to give a frame of reference when assessing point sizes, all the point sizes in each plot are normalised to their respective standard point.

Fig. 3.5: Averaged composition of filter samples by heteroatom groups in winter and summer campaigns (left and right respectively) from the detected signal. The winter composition was: CHO – 5.1%, CHON – 82.3%, CHOS – 11.5% and CHONS – 1.0%. The summer composition was: CHO – 23.2%, CHON – 43.6%, CHOS – 28.9% and CHONS – 4.3%.

Figure 3.6: Correlation of AMS and database analysis of total organic aerosol for winter (upper) and summer (lower) campaigns. Database analysis of TOC measured in volume corrected peak area.

Figure 3.7: Double bond equivalencies of compounds found in the averaged filter sample for both winter (left) and summer (right) campaigns. In both plots, a standard point of 10,000,000 peak area m^{-3} is inserted in pink at carbon number = 20, DBE = 0 in order to give a frame of reference when assessing point sizes, all the point sizes in each plot are normalised to their respective standard point.

Figure 3.8: Average oxidation state of carbon for compounds found in the averaged filter sample for both winter (left) and summer (right) campaigns. In both plots, a standard point of 10,000,000 peak area m^{-3} is inserted in pink at carbon number = 20, OSc = -2 in order to give a frame of reference when assessing point sizes, all the point sizes in each plot are normalised to their respective standard point.

Figure 3.9: Temporal trends of each heteroatom group and the total organic carbon (total OC) for the winter campaign. Temporal trends with fixed axis for better comparison can be found in figure 3.10)

Figure 3.9: Temporal trends of each heteroatom group and the total organic carbon (total OC) for the summer campaign. Temporal trends with fixed axis for better comparison can be found in figure 3.10)

Figure 3.10: Time series for each of the campaigns (winter – left, summer – right). Each heteroatom group has been separated into different time series,

with total OC for each. The scales have been fixed for both winter and summer keep both plots for each group comparable.

Figure 3.11: Low temporal resolution (24 hourly) time series plots for each of the heteroatom groups in the winter campaign.

Figure 3.12a: Time series for non-refractory organic matter (NR-OM) measured by AMS and database analysis for the winter campaign. Database data is quantified as time corrected peak area.

Figure 3.12b: Time series for non-refractory organic matter (NR-OM) measured by AMS and database analysis for the summer campaign. Database data is quantified as time corrected peak area.

Figure 3.13: The range of levels for each of the heteroatom groups along with upper and lower quartiles for both summer (red) and winter (blue).

Figure 4.1: Common oxidation pathways for formation of nitro-aromatic compounds. Taken from Vione 2005.

Figure 4.2: Common oxidation pathway for formation of ortho-nitro-aromatic compounds. Taken from Harrison 2005.

Figure 4.3: Extracted ion chromatographs of methylated nitrophenolic compounds, each peak is annotated with the structure of the corresponding compound

Figure 4.4: The peak area responses of nitro-containing aromatic compounds in a mixed solution of 1 ppm. Points are colour coded for position of nitro group: green – meta, red – ortho, blue – para and orange – not on the same ring as hydroxy group.

Figure 4.5: Separated EICs of standards in mixed solution, including isomers of nitrophenol, methyl nitrophenol, a dimethylnitrophenol and a methoxynitrophenol

Figure 4.6: Overlaid EICs of standards in mixed solution, including isomers of nitrophenol, methyl nitrophenol, a dimethylnitrophenol and a methoxynitrophenol

Figure 4.7: Dendrogram produced by HCA, each of the six groups selected are colour coded for ease of identification, the axis shows the height of the dendrogram, which is a representation of dissimilarity. The higher the height, the more dissimilar the branches being linked are.

Figure 4.8: Co-variance plot of 43 CHON compounds observed, with enough data for correlation analysis, in the samples colour-coded by correlation. More red points show higher correlation, bluer points show higher anti-correlation and greener points show no significant correlation between species. Horizontal lines have been added to easily identify the group clusters.

Figure 4.9: Source wind rose plots and summed time series for the 6 cluster groups, the averaged time points corresponding with the filter sample are represented by a point and colour coded by the summed abundance of the components of each group.

Figure 4.10: Temporal trends of the summed peak areas for the 6 cluster groups defined by HCA of nitrophenolics

Figure 4.11: Total summed contributions of each component across the time series to group 1

compounds which co-vary highly together to show patterns of high co-variance groups, these groupings show the initial formation of subsets within the heteroatom group and highlight the need for big data analysis.

Figure 5.3: Dendrogram displaying the HCA clusters for 84 CHOS compounds, the 11 cluster groups are shown by different colours.

Figure 5.4: Polar plots for each of the cluster group 1 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Figure 5.5: Polar plots for each of the cluster group 2 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Figure 5.6: Polar plots for each of the cluster group 3 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Figure 5.7: Polar plots for each of the cluster group 4 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Figure 5.8: Polar plots for each of the cluster group 5 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Figure 5.9: Polar plots for each of the cluster group 6 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Figure 5.29: Temporal trends of the four organosulfate isomers at *m/z* 215.0230 observed by the HILIC method.

Figure 5.30: Normalised temporal trends of the four organosulfate isomers at *m/z* 215.0230 observed by the HILIC method.

Acknowledgements

When I was a child, my Grandad told me I could do anything I set my mind to. Like with everything he said, I believed him and from that point have had blind faith in myself to accomplish anything I wanted. Nothing worth having is ever easy and this PhD is no exception, there were many low periods and many high ones too. But the memory of my Grandad was what kept me going through the darkest of moments. This thesis is dedicated to you, Grandad. I hope it would make you proud.

There are many people I would like to thank for their support. Firstly I would like to thank my main supervisor, Jacqui Hamilton. Your guidance and advice have been invaluable during this period and without you I fear I would not have accomplished half of what I have. I would also like to thank my second supervisor and IPM, Andrew Rickard and Ally Lewis, who gave me good advice for additions and challenged my knowledge on the subject and encouraged me to learn more. I would also like to give a particular thank you to Kelly Pereira, who took over as my supervisor for a very crucial period in my PhD and gave me a great deal of help with the orbi and I cannot be grateful enough for all the help. I have been incredibly fortunate to have the supervision team that I do.

I would also like to thank my friends and family. In particular, Danielle Dale, who has dealt with a great deal of my rants and rambling over the years and continues to listen and support whenever I need help. Also, to Kathryn Chamberlain who, undergoing your own PhD, relates to my day to day dramas and stresses more than anyone. You are an inspiring person in both professional and personal life and I hope you remember that.

Finally, I would like to thank Jeroen Peters. You were a most excellent supervisor and an insight into the world of science. You were the first person to suggest I do a PhD and had confidence I was capable. While some days I wondered why I listened, I appreciate everything you've done.

Declaration

I declare that the work presented in this thesis is original work and I am the sole author, except where otherwise stated and not been previously submitted to any institution for any award.

The work presented was partially submitted for peer-review in the following article:

William J. Dixon; Kelly L. Pereira, Freya A. Squires, Rachel E. Dunmore, James. R. Hopkins, James. D. Lee, Daniel J. Bryant, Andrew R. Rickard, Yele Sun, Weiqi Xu and Jacqueline F. Hamilton; Automated high-throughput method for the detection of organic aerosols using a novel high-resolution mass spectrometry database; ACS Earth Space Chem.; submitted.

Chapter 1 – Introduction and background

1.1 History and effects

In 1952 the city of London, UK suffered from a mass smog which descended on the city[1]. This smog affected the ability of the visibility within the local area, leading to a shutdown of the city as traffic ground to a halt. In the following days, an unprecedented level of hospitalisations were reported and around 12000 deaths have been linked to this smog event[1]. In the midst of the aftermath, the smog was linked to the poor air quality of the city and the pollution from the area and air quality became a hotly studied topic.

It is now commonly known that the quality of the air in the environment has major impacts on human health and climate. Exposure to poor air quality is a significant environmental contributor to the increased risk factors of premature mortality and numerous diseases; affecting an estimated 4 million premature deaths globally, in 2015[2–6]. Heart disease and strokes are the most common causes of premature deaths resulting from exposure to excessive air pollution[7]. Respiratory and cardiovascular disease and cancer[3] are common, chronic conditions which can develop as a result of prolonged exposure to high pollution levels. Aerosol particulate matter (PM), particularly that with a diameter ≤ 2.5 μm ($PM_{2.5}$) severely impacts health, being classified air pollution as a known carcinogen (group 1) by The International Agency for Cancer Research (IARC) due to the associated risk of increased rates for several forms of cancer[6]. Particulate matter also has detrimental effects in areas other than human health; compounds within organic aerosol also have effects on wildlife and

plant life. Water soluble organic compounds (WSOC) for example are able to enter the water cycle[8] leading to larger scale impacts from WSOCs compared to compounds which only remain in the gaseous and particulate phases alone. WSOCs such as nitrophenols are known phytotoxins, capable of damaging plant life[8–10] if taken up and are an example of the larger scale effects of air pollution which are not often linked to the atmosphere. Additionally, WSOC also have detrimental effects on aquatic life due to their water solubility[8].

In addition to the negative health effects of $PM_{2.5}$ on humans, animals and plant-life, aerosol has negative impacts on climate. The earth's energy budget can be driven by a number of factors, one of the most impactful and difficult to predict is the forcing from aerosol. The Inter-governmental Panel on Climate Change (IPCC) published the most significant contributions on the planetary radiative forcing[11]. The radiative forcing from aerosol (both OC and EC) ranges between -0.77 and 0.23 W m^{-2} and as such cannot be assumed to all behave the same. No other group is capable of both positive and negative radiative forcing to the degree of particulate matter and hints at the extent of the intricacies of the complexity of compounds with aerosol. Aerosol particulates can act as cloud condensation nuclei (CCN) and contribute to the increased cloud number and size, increasing reflection of solar radiation and driving a negative radiative forcing[12,13]. Brown carbon particulates are strong light absorbers in the visible spectrum (giving rise to the name) and absorbers across the spectrum, and hence contribute to warming within the atmosphere and having an effect on the radiative forcing[14–18]. As a result of a combination of direct and indirect effects from aerosol this leads to the large variation on radiative forcing and subsequently can cause difficulties in predicting the effects of aerosol (for example using predictive tools such as climate models)

as the specific forcing of aerosol in a given air mass is dependent on the composition, and composition varies vastly between regions.

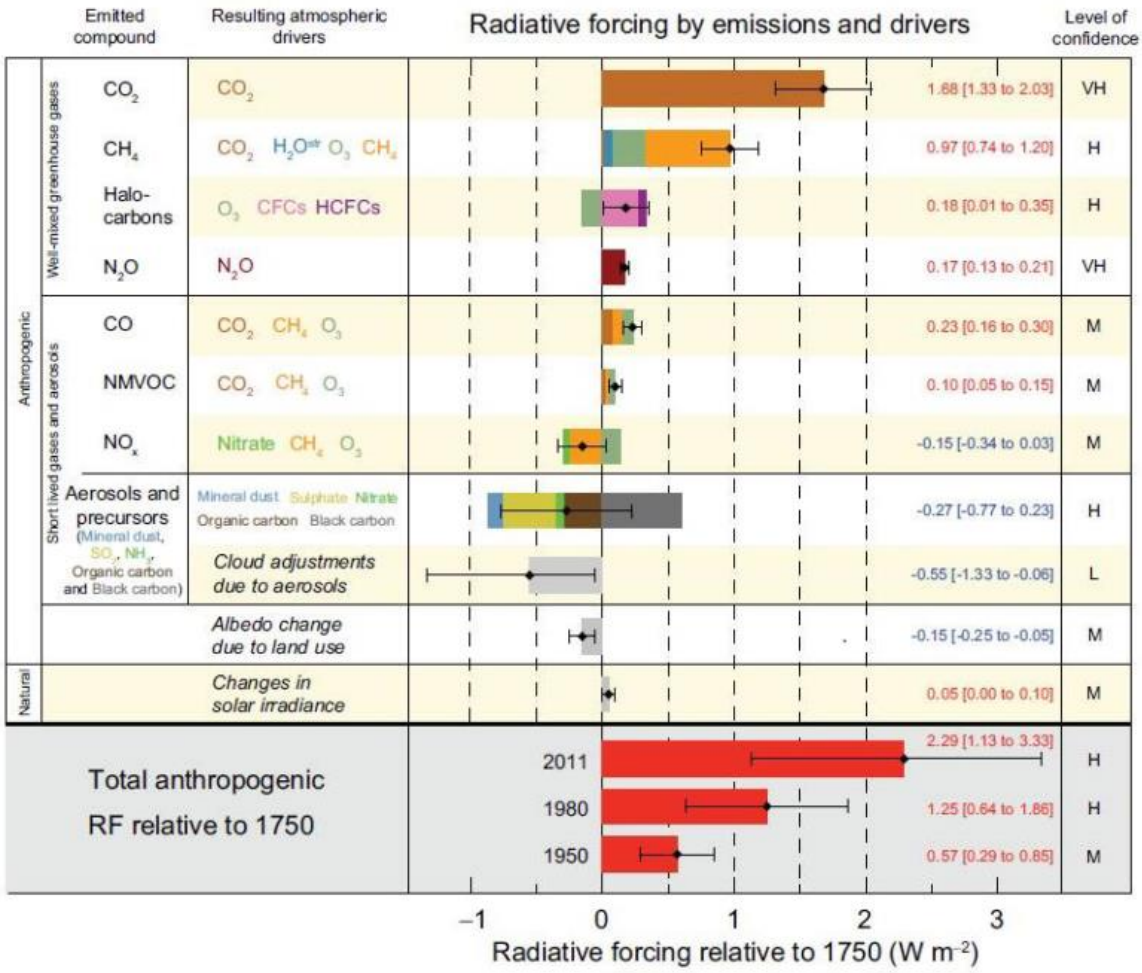

Figure 1.1: IPCC report on the radiative forcing of atmospheric factors in 2011[11]. On the right hand side of the figure shows forcing relative to 1750 Wm^{-2}, with uncertainties. The confidence for each of the factors is at the far right of the figure (L - Low, M - Medium, H – High and VH – Very high).

1.2 Composition and source

Carbon-based material dominates urban samples of aerosol and is made up of both elemental carbon (EC) and organic carbon (OC), contributing up to 90% of total aerosol[12]. Previous work has shown that particulates in urban areas are composed of more than 10,000 individual organic molecules[8,19] making organic aerosol the most complex and challenging component of urban air pollution; and therefore is the least understood aspect of air pollution, due to this complexity of composition. In addition, the flux nature of the composition leads to changes in the composition over time, this is due to reactions which the compounds undergo as air masses age and become more oxidised. The sources of organic aerosol (OA) are varied. OA can be directly emitted to the atmosphere (primary organic aerosol, POA) from a range of anthropogenic sources, including diesel exhaust[20], industry[21], manufacturing, cooking[22,23] and biomass fuel combustion[24] and natural sources such as biogenic plant emissions[25–27], forest fires[28] and from the sea surface microlayer[29].

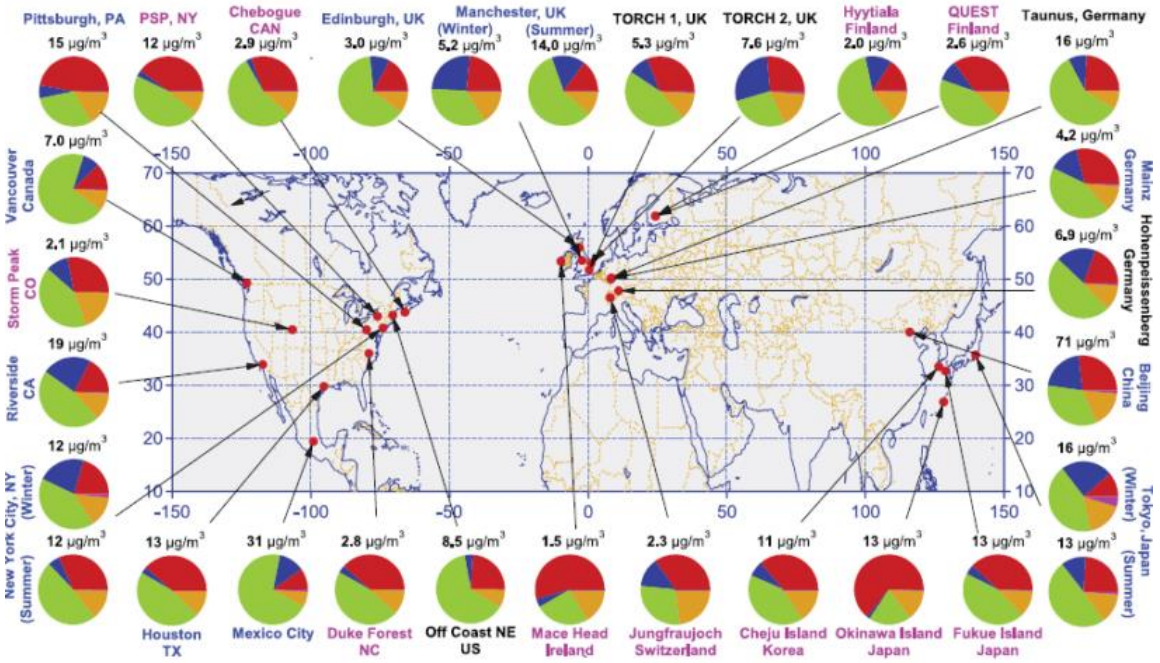

Figure 1.2: Composition of PM taken from different sampling areas, the names of urban area are in blue, rural areas in pink and downwind of major cities in black. The composition shown in pie charts, organic in green, sulfate in red, nitrate in blue and ammonium in orange. Taken from Zhang 2007[30].

Aerosol which forms from the oxidation of volatile organic compounds (VOCs) in the atmosphere and the subsequent loss in volatility of these components is known as secondary organic aerosol (SOA)[30]. Both biogenic (from living organisms) and anthropogenic (man-made) sources are capable of VOC production and as such SOA can be biogenic and anthropogenic in nature. Biogenic VOCs (BVOCs) are a large source of gas phase pollution, accounting for ~1100 Tg C y^{-1} [31]; one of the most common biogenic VOC groups being isoprene and terpenoids, (compounds derived from isoprene; $C_{10}H_{16}$ for monoterpenes or $C_{15}H_{24}$ for sesquiterpenes). BVOC has been previously

estimated to contribute 88 Tg C annually to SOA[32], several times higher than the estimated contribution from anthropogenic contribution[28,32] of around 10 Tg C annually and as such is a significant factor in the global SOA composition as longer-lived species, produced in such large quantities from highly forested areas can impact regions down-wind.

Despite this, the contribution from anthropogenic sources is still significant and cannot be understated. Arguably, anthropogenic SOA would be easier to comprehend due to its lower abundance and hence much research has been done to find the sources of anthropogenic SOA, with a view to reducing the impact from it. SOA tracer compounds such as benzene, toluene and phenol are good indicators of anthropogenic sources and can be sourced around industrial sites and transport lines (shipping lanes, railways and roads) and indeed has been used to identify said transport lines. In addition to the anthropogenic VOCs produced, inorganic anthropogenic aerosol (such as sulfate particulate) and inorganic gas phase pollutants (such as O_3 and NO_x) are produced[33–36], these act upon the VOC emissions and facilitate the aging processes and generation of SOA.

More recently, global and regional models are employed to best estimate the levels and contributions of aerosol, these models rely on inventories of emissions to maintain accuracy. Longer run models can run-away with their uncertainties leading to discrepancies between the prediction from a model and the reality that follows. Despite this, models are a fast, reliable and simple visual technique for predicting how aerosol evolves and how the movement of air parcels within the atmosphere leads to dilution and movement of pollution globally. Air back trajectory plots can be used to track the movement of air

parcels, and the implementation of these can demonstrate where air has come from and over what sort of terrain. This information can be used to identify potential sources of compounds and any aging which may have occurred if the compound was long-lived enough to have had time to oxidise, for example during periods of stagnancy within the air parcel.

1.3 Oxidative species

The ageing of VOC leads to a large number of tracer compounds and increase the complexity of components in organic aerosol. As compounds get more oxidised and become aged, their volatility lowers; as a result these compounds are more likely to partition into the particulate phase of a dynamic particulate/gaseous system. Volatility is used as an indicative measurement of the age of an aerosol[37]. Figure 1.3 shows where commonly observed groups of species occur within the relationship between volatility and oxidation (given in the form of average oxidation state, OSc); LV-OOA – lower volatility oxygenated organic aerosol, SV-OOA – semi volatile oxygenated organic aerosol, BBOA – biomass burning organic aerosol, HOA – Hydrocarbon-like organic aerosol, ELVOC – extremely low volatile organic compounds, LVOC – low volatile organic compounds, SVOC – semi volatile organic compounds, IVOC – intermediate volatile organic compounds and VOC – Volatile organic compounds. A general trend is seen in these groups, as the oxidation of a compound increases, its volatility becomes lower.

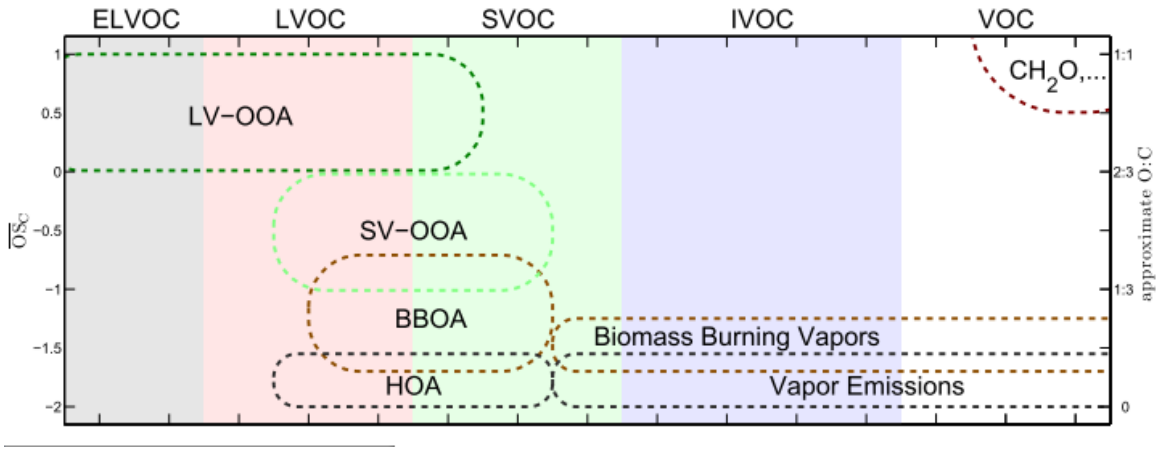

Figure 1.3: The relationship between the average oxidation state of the carbon and the volatility of the compound, commonly observed groups of aerosol are highlighted on the figure (LV-OOA – lower volatility oxygenated organic aerosol, SV-OOA – semi-volatile oxygenated organic aerosol, BBOA – biomass burning organic aerosol, HOA – Hydrocarbon-like organic aerosol, ELVOC – extremely low volatile organic compounds, LVOC – low volatile organic compounds, SVOC – semi-volatile organic compounds, IVOC – intermediate volatile organic compounds and VOC – Volatile organic compounds). taken from Donahue 2012[37]

The large number of gas-phase pollutants and common oxidants for VOCs, leads to a large number of possible products, depending on the ambient conditions of the air mass at the time. Figure 1.4 shows several of the most common oxidation pathways which can be undertaken by VOC species. Common gas-phase oxidants including O_3 NO_x and OH typically form radicals when aliphatic VOCs are exposed to them. These are particularly relevant and potent during ozone oxidation of double bonds in aliphatic VOCs. The oxidation of this double bond forms a five-membered double radical species, commonly

known as a Creigee intermediate or ozonide[38,39]. This species quickly collapses into carbonyls and esters. Creigee intermediates, in themselves, can impact composition of tropospheric chemistry. Near the surface, these Creigee intermediates can become stabilized and oxidise SO_2 into SO_3 and subsequently into H_2SO_4, thus directly emitted sulfates can be turned into acid and particulates at the surface level, directly causing harm to humans and the local environment.

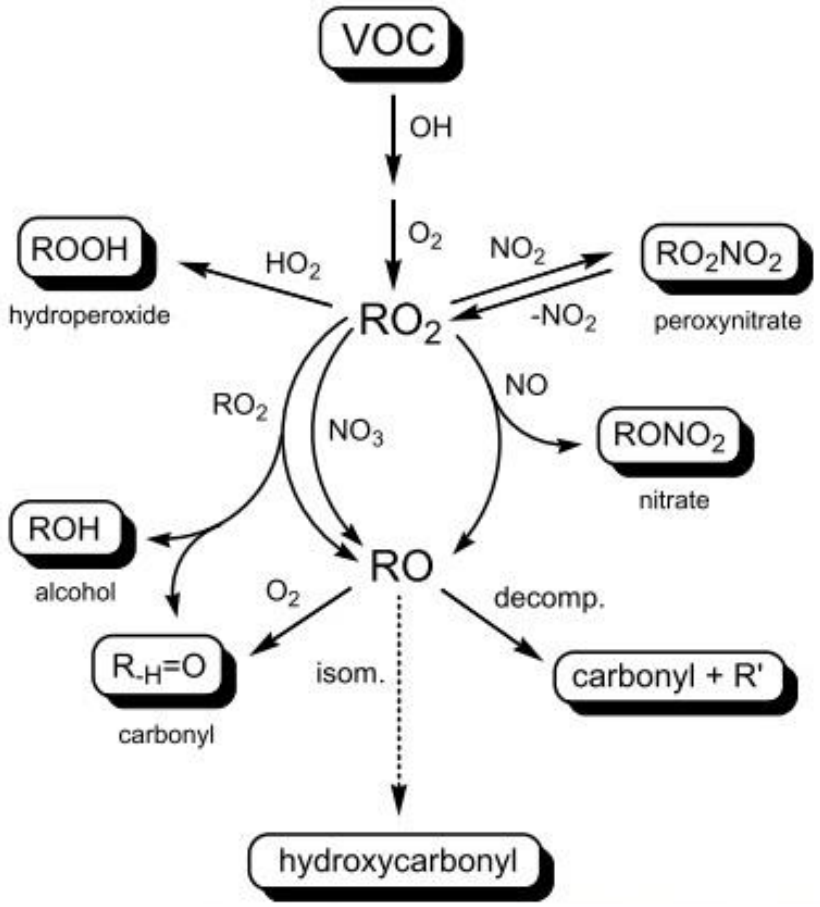

Figure 1.4: Schematic of the OH mechanisms commonly undergone by VOCs, taken from Hallquist 2009[28]

1.4 Nitroaromatics

Nitro-aromatic compounds are a large scale atmospheric pollutant and are some of the most abundant compounds within the organic fraction of both the gaseous and particulate phases. The generation of nitro-aromatics are from direct traffic emissions, biomass burning and oxidation of other directly emitted aromatics[10,16,40–43]. Nitro-aromatic compounds have detrimental effects on human and plant health, but also have additional notably detrimental effects after their deposition into water[44–46]; and taken up by plants and aquatic life (such as algae and fish). The correlation between levels of nitrated-aromatics (specifically nitrated phenols) and damage to plant-life in both *in situ* and *in vitro* experiments have been observed; this phytotoxicity is attributed with forest decline in America and Europe. 2,4-dinitrophenol in particular is a known coupling agent in oxidative phosphorylation and can damage plant metabolism even at levels below 10 μM[47]. Despite their clear detrimental interactions with the biosphere however, little work into the behaviour of isomers of individual nitro-aromatics has been performed and how these individual isomers behave. This is because different nitro-aromatics behave differently and have differing formation pathways[16], so isolating and characterising their behaviours presents difficulties without specifically focusing on this family of compounds.

It is known that structural isomers have differing abundances within ambient air and show variation in partitioning coefficient between gas phase and particulate phase[45,48–51]. In ambient particulate samples, the abundance of para-nitro aromatics is much higher than that of ortho-nitro aromatics[40,48,49,52]; this is unexpected based on previously observed nitration of aromatics[53]. Vione *et al* propose the reason for the higher levels of 4-nitrophenol compared to 2-

nitrophenol found in ambient air is the susceptibility for 2-nitrophenol to undergo degradation in the gas phase[45]; in the gas phase, ortho-nitrophenols are reported to be a source of nitrous acid (also known as HONO)[9] and would account for some of this gaseous phase degradation. Interestingly, Al-Naiema and Stone observed more 4-nitrophenol in the gaseous phase compared to particulate phase during sampling in Iowa, USA[41]; this is opposed to Cecinato *et al* who observed higher abundance in the particulate phase from samples in Rome, Italy[48]. As such, a constant ratio of gaseous to particulate phase cannot be assumed.

Seasonal variation is also observed in nitro-aromatic compounds, the reason for which is not completely decided upon but is determined to be a combination of factors ultimately deciding the abundances. Zhang *et al* observed a large increase (over 2x) in 4-nitrophenol during the winter, compared to summer, compared to a slight decrease in 2-nitrophenol during the winter period when compared to summer[49]. Using aerosol mass spectrometry (AMS) Xu *et al* detected an increase in biomass burning organic aerosol (BBOA) and less oxidised oxygenated organic aerosol (LO-OOA)[54]; whilst these are not purely nitro-aromatic measurements, nitro-aromatics are sourced from biomass burning and their semi-volatile nature would place them in LV-OOA, the increase in abundance for each of these two aerosol groups would firmly suggest increase in nitrophenol levels.

It is clear that the behaviours of nitro-aromatics vary based on their structural isomerism within the atmosphere, and yet during quantification using mass spectrometry (MS) based analysis, these isomers are often grouped together when observing the sensitivities of MS analysis to nitro-aromatics[8,16], leading

to less detail in the analysis of the abundance of nitro-aromatics within the atmosphere. Oxidation of aromatics into nitrophenolics are typically from OH radicals and nitrogen oxides (NOx)[55]. Equations 1.1 and 1.2 show common formation pathways of nitrification of aromatics which can occur within the VOC.

$$Ar - H + \cdot NO2\ (N2O4) \leftrightharpoons Ar(H) - NO2 \cdot (+\ \cdot NO2) \qquad \text{(Eq 1.1)}$$

$$Ar(H) - NO2 \cdot + O2\ \rightarrow Ar(H) - NO2 + \cdot HO2 \qquad \text{(Eq 1.2)}$$

Directly emitted nitrophenolic compounds come from a diverse range of sources, such as from traffic (particularly from the catalytic converters in diesel engines) and the manufacture of products such as dyes, pesticides and from the treatment of wood and paper during production[47]. During winter months, this is noticeable fraction of the total organic carbon. In some cases, the levels of particular nitrophenolic compounds are hundreds of times higher in the winter months compared to the summer months[56].

For European countries, such as Slovenia, approximately 1% of the total OC comes from nitrophenolics found in brown carbon[56]. For countries with much higher populations and more densely populated areas, such as China or India, this percentage is expected to be much higher due to the prolific use of biomass burning as a fuel source and with peak PM levels reaching upwards of 400 μg m^{-3}, this means 1% of this would be >4 μg m^{-3} from nitrophenolics alone. Again, leading to significant risks of damage to plant-life and pollution of the water-cycle of the local area. Recent events of significant visibility loss during

extreme pollution events (particularly from light absorbing, brown carbon) are a very apparent consequence of the impact of carbonaceous material and have led to its increased visibility in the public zeitgeist. Studies of light absorbing organic carbon from India and China peak during 30 - 35 mg m^{-3} [56,57]. It is thus expected from these levels of brown carbon, that a large amount of light absorbing nitrophenolics will be present.

1.5 Organosulfates

The first evidence for organosulfate compounds in the atmosphere was discovered in 2007[58], this uncovered new potential impacts from sulfate; in addition to the direct effects of sulfate and the links to acid rain, new evidence of SOA formation by the reaction of biogenic VOCs with ambient sulfate emerged[25]. The formation of organosulfates from biogenic VOC will partially account for the underestimation of biogenic SOA observed in models[59,60], this underestimation is largely due to the current limited knowledge of a complete set of reactions which VOCs undergo. The most common pathway for formation of the organosulfate is the opening of epoxy rings, usually as a result of double bond oxidation and functionalization.

Organosulfates (OS) are also directly emitted, most commonly long-chain aliphatic compounds, the result of oxidation of fuels or a product of surfactant manufacturing. As such, directly emitted OS tend to be anthropogenic.

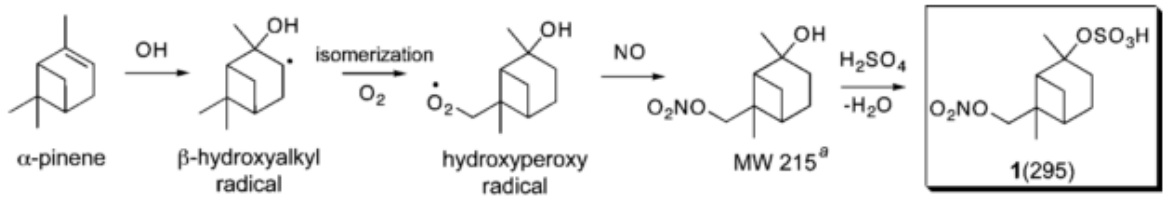

Figure 1.5: An example of the formation of biogenic nitroxy-organosulfate from a terpene tracer compound. Taken from Surrat 2008[25]

This variation of the formation and emission of organosulfates allows for a variety of structures and a range of functionality in addition to the sulfate group. This means that only relatively small proportions of the overall organosulfate group have similar properties to other organosulfate species. As such, the identification and characterization of organosulfates in complex matrices (such as atmospheric aerosol) becomes a challenge as there is no set formula to the group. The rates of formation of organosulfates are significantly increased in environments with low relative humidity (RH), more acidic environments and during the presence of high levels of seed aerosol[61,62]. Lower pH levels increase the abundance of acidic and ionized species, allowing for easier formation of the sulfate ester. In a similar way, higher RH dilutes the relative concentration of these acidic, ionized species and thus lower the rate of the sulfate esterification. Higher RH also increases the formation of functional species from H_2O and diversifies the pathways which these species can undergo. This would subsequently reduce the amount of precursor material which would undergo sulfonation[58,61,63,64]. Acidic seed particles also increase

the formation of organosulfate as these offer a surface on which reactions can be undergone and facilitate the decrease in volatility of species which can react.

In some cases, species can be functionalized with both nitroxy and sulfate groups. These compounds tend to form from biogenic species with higher levels of unsaturation, leading to higher availability for functionality, terpene species therefore have higher levels of nitroxy-sulfate derivatives[65,66].

1.6 Detection

As previously mentioned, the sources of SOA are quite complex and as such can be difficult to completely map. In different regions, sources can have varying levels of dominance. The degree to which these sources contribute to the total emissions can be estimated by employing source apportionment techniques, such as a Chemical Mass Balance (CMB) modelling. CMB uses pre-defined chemical speciation profiles based on measurements carried out at the emission source, usually in laboratory test conditions[67,68]. Using aerosol mass spectrometry (AMS) it is observed that oxidised material dominates particle mass in both rural and urban locations, and that this material is formed in the atmosphere from the oxidation of VOCs. Unfortunately, the high temperatures and high energy ionisation conditions used in aerosol mass spectrometry destroys the majority of the compositional detail during analysis and so provides little information on the sources of this oxidised material[69].

The picture of organic aerosol that emerges is one of a composition that is in flux, changing as emissions undergo complex chemical reactions as they are transported away from their source. In recent years there have been many

projects to study SOA formation, leading to rapid advances in mechanistic and process level understanding. Unfortunately, we can still only characterize a very small fraction of the organic mass of particles, inhibiting our ability to draw conclusions about their source. There have been attempts to use CMB to quantify the sources of SOA, using a very limited subset of tracer compounds based on chamber experiments[20,70]; however, the dynamic nature of the composition of SOA means this approach is difficult. Thus even fundamental questions, such as whether natural or anthropogenic emissions contribute more SOA within cities, and what factors influence the formation of SOA, remain mostly unanswered.

One way to determine the sources of SOA is to use high-resolution offline techniques coupled to filter sampling in the field. High mass resolution techniques, such as Fourier-transform ion-cyclotron resonance mass spectrometry (FT-ICR MS) with electrospray ionisation or Filter Inlet for Gases and AEROsols for Chemical Ionisation Mass Spectrometry (FIGAERO-CIMS), enable individual compounds to be identified by their chemical formula[8,71,72]. However, structural isomers cannot be separated by this method without the employment of chromatography. This leads to the potential for smaller peaks becoming eclipsed by abundant species or compounds with a higher ionisation efficiency, and so the exact identity of all of the unknown compounds cannot be fully confirmed. To overcome such limitations, hyphenated techniques such as the combination of liquid chromatography (LC) or gas chromatography (GC) in tandem with MS are often employed[73,74]. Such techniques allow for the separation of compounds with the same *m/z* values based on their affinity to a stationary phase, potentially allowing for identification of isomers.

The majority of studies investigating SOA use a targeted analytical approach, where specific SOA tracers are identified and quantified, based on the composition observed in previous simulation chamber experiments. The quantification of compounds by this method requires authentic standard compounds to match retention time (for LC-MS analysis) and calibrate the instrument response for factors such as ionisation efficiency. In many cases, standards are not commercially available, making quantification using standards difficult without specialised organic synthesis and due to variations in ionisation efficiency, proxy compounds (compounds which possess similar properties to the target analyte) can not necessarily be relied upon to give accurate quantification. These challenges make the full quantification of thousands of compounds a costly and time-consuming using a traditional LC-MS or AMS approach. Due to the complexity of the organic fraction of aerosol, the identification and quantification of these compounds becomes problematic without some foreknowledge of the composition. Targeted screening is a potential solution and is an approach often used in other fields of research[75–77], metabolomics (and other –omic fields) in particular uses targeted screening for analysis[78,79], but has not been previously employed within atmospheric research, despite the apparent potential for the implementation of such a technique.

Targeted screening typically employs the use of a database of observed compounds and all samples are checked for the presence of entries in the database; examples of commonly used analysis include, drug detection[80], multiplex bioassays for food safety[81,82] and structural matches in crystallography[83]. Screening requires the development of libraries of commonly found targets to screen against and many libraries have been generated for

both liquid and gas chromatography, ranging from proprietary to open source; prevalent libraries within mass spectrometric analysis include m/z cloud[84] and FiehnLib[85] and these are commonly used as a reference spectrum during MS analysis. An advantage of targeted screening is that it tends to be a high-throughput and rapid analysis for the targets with very good sensitivity during quantification. However, a limitation of targeted screening is that no analysis is performed on anything outside of the targets within the database; as such, any application of targeted screening to atmospheric composition must have a thorough and extensive library to be able to account for the thousands of compounds previously described in organic aerosol[8,86,87]. This type of methodology also requires a high number of data points to provide any statistically meaningful analysis in temporal variation.

The advantages of targeted screening are that it can be high-throughput, allowing rapid analysis of the target compounds, with very good sensitivity and can allow for accurate quantification of numerous species in a complex matrix. However, it is limited in its scope as no analysis is performed on species not in the database and with open-source databases being limited and proprietary databases being limited in what their licences will permit in terms of their use, the development of a master database for use in atmospheric chemistry is still in its early stages.

Due to the complexity of the organic fraction of aerosol (>10000 chemical species within an air mass), the identification and quantification of the compounds within a large number of samples becomes increasingly problematic and time consuming without some foreknowledge of the composition. Targeted screening offers a potential solution, targeted screening

typically compares mass spectra of unknowns to a database of mass spectra of previously observed compounds, for example the NIST database contains thousands of mass spectra. However, mass spectral databases for tandem mass spectrometry (such as CID or HCD) are less common,[76,77,88] but are often used in metabolomics (and other –omic fields)[79,89]. Examples of commonly used analysis include, drug detection[80], multiplex bioassays for food safety[81,82] and structural matches in crystallography[83]. Targeted screening requires the development of mass spectral libraries of commonly found tracer targets, and the use of both proprietary and open source tandem MS libraries to produce a master database or smaller more niche databases, for example just biogenic tracers within air masses. Examples of commonly used open source tandem mass spectral libraries include m/zcloud[90] (specifically for Orbitrap MS) and FiehnLib[85] and these may be the first point-of-call for developing a master database for the use in targeted screening.

A different approach is a non-targeted, whole sample approach to look for covariance and to identify trends without any *a priori* knowledge of the sources. The use of high complexity analysis coupled to emerging data science methodologies has revolutionised biomedicine, (including metabolomics, proteomics and phenomics). An untargeted approach has been used within the field of atmospheric chemistry on samples to get a more complete characterization of organic aerosol. The work of Rogge *et al* is an example of this, the group comprehensively identified and quantified the organic components directly emitted from a number of sources such as wood combustion and natural gases[91,92]. Comprehensive studies like these are invaluable when building databases of emission tracer compounds and determining source fingerprints for use in source apportionment. However, in

the cases identified by Rogge *et al*, only a few samples from each source were taken and as such these studies may not be fully applicable as a reference for a long term source apportionment study of ambient samples; particularly if looking at more aged tracer compounds. The best practice for the development of tracer compound and emission databases are therefore a combination of the comprehensive chamber experiments and the fingerprinting of direct sources. This type of analysis tends not to be fully quantitative using 'relative quantification' (peak area compared to the peak area of a different species) as a rough indication of abundance, in the same way proxy compounds are sometimes used in drug development when a isotopically labelled version of the target analyte is not available.

The translation and application of these techniques to atmospheric chemistry, could potentially be used to highlight the contributions of sources and reaction mechanisms to the overall composition, in a similar way as CMB uses tracer compounds to determine source apportionment. Already, these techniques are being applied in environmental chemistry focusing on pollutants and their decomposition products and metabolites albeit with the practice in its infancy.

A semi-targeted screening technique offers a compromise between targeted screening and untargeted screening. Semi-targeted screening allows for quantification of the targeted analytes along with the identification of unknowns within a sample and has been employed in the field of metabolomics for the identification of both the known and unknown compounds within the metabolome[93,94]. Unfortunately, there are currently no tandem mass spectral databases suitable for most environmental analysis, including atmospheric aerosols. However, it is clear of the benefits of a previously described master

database to the rapid analysis of organic aerosol. The time and costs associated with analysis of organic aerosol would be greatly reduced by an automated screening process, and a completely automated system would even lower the skills barrier required for analysis of organic aerosol composition and increasing the scope for who can produce data and in turn increase the amount of data available to understand the complexities of atmospheric aerosol. A high-throughput LC-MS method using automated processing would be able to analyse a high number of data points in a very small time-scale and used in tandem with an automated sampling technique and datamining; the potential for this system is unprecedented in its ability to understand the atmospheric composition. The employment of statistically meaningful analysis of the variation of organic tracers in ambient aerosol using regression techniques such as hierarchical cluster analysis (HCA)[95,96], multi-variate analysis (MVA)[97]or positive matrix factorisation (PMF) would also be possible if the dataset was increased using this automated system.

1.6.1 Liquid Chromatography

The fundamentals of chromatography, stem from gravimetric column chromatography where a mix of compounds are run through a vertical column of stationary phase. The affinity of compounds to the stationary phase slows the progression of the compound through the column, in the case of gravitational chromatography, the movement through the column is facilitated by gravity[98]. This method is time consuming and the resolution of separation is limited to the height of the column used (and number of chromatographic plates), which tended to fill a room.

In more recent times, chromatography employs the use of pressurised mobile phase to push compounds through a column, in liquid chromatography (LC) the mobile phase is liquid, this mobile phase tends to possess an opposing property to the stationary phase in the column, and hence the difference in affinities for the two phases dictates the time taken for a compound to elute from the column (retention time, RT)[99]. This allows for higher resolution of separation (an increased number of chromatographic plates) within the shorter column length as the packing also encourages increased path length as analytes transfer along the column do not follow a linear path, instead must navigate around the packing and this increase in path length leads to more time for interaction with the stationary phase (increasing the effective number of chromatographic plates). This pressurised chromatography is called high performance liquid chromatography (HPLC) or ultra-high performance liquid chromatography (UHPLC)[100,101]. The most common LC is known as reversed phase, as it is the reverse of the first phase of chromatography used (known as normal phase); reverse phase consists of a polar mobile phase and a non-polar stationary phase. Using reverse phase chromatography, polar compounds would elute before less polar compounds, as such RT can be used as an indicator for polarity in reverse phase. Naturally, normal-phase chromatography has the opposite phases as polar and non-polar.

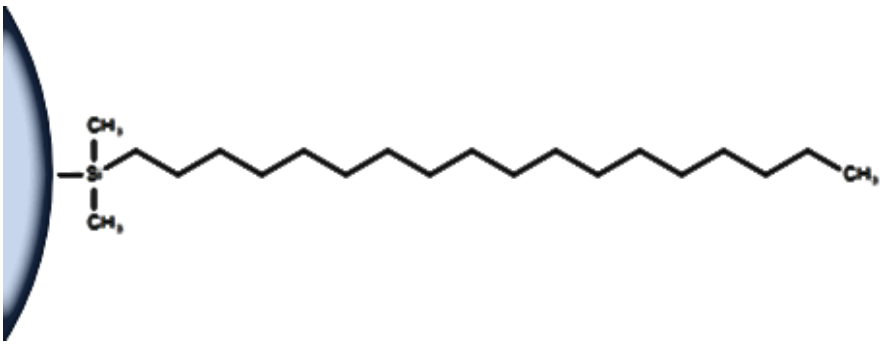

Figure 1.6: A simplified diagram of a C18-based stationary phase packing showing the chain used for primary interactions with analytes.

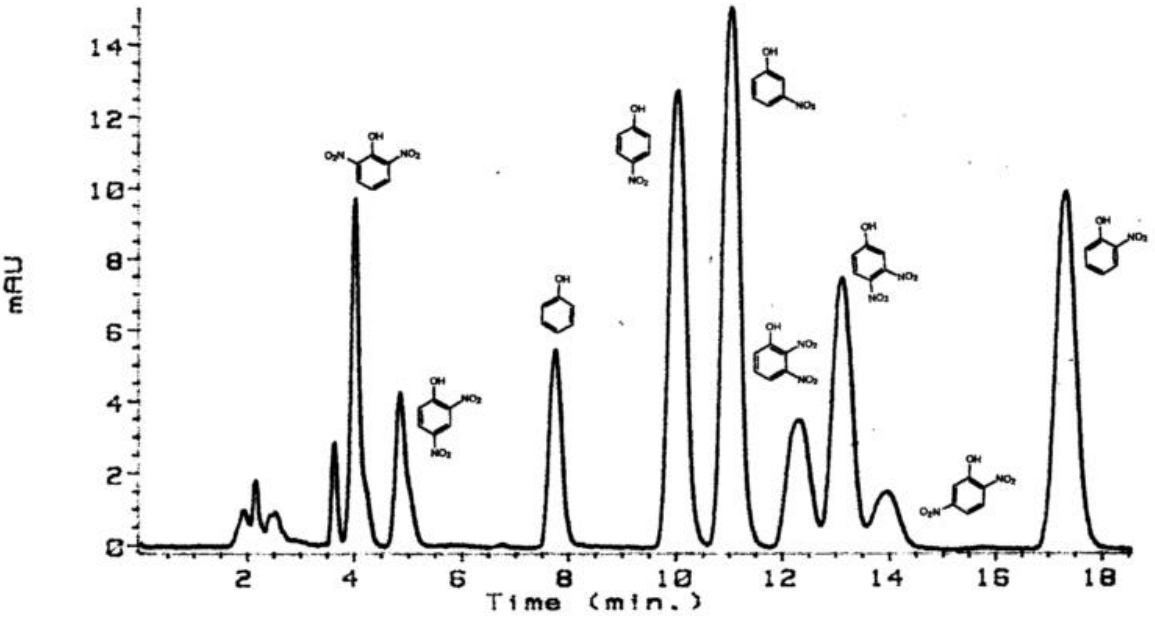

Figure 1.7: A chromatograph of phenolic compounds found in ambient air. The chromatography shows good separation of the compounds and aids easy identification. Taken from Belloli 1999[102]

An orthogonal method of separation to traditional C18 back-boned LC, is that of Hydrophilic interaction chromatography (HILIC). C18 back-boned chains

work on hydrophobic interactions such as Van der Waal's forces to cause the affinity between an analyte and the solid phase. HILIC, on the other hand, can use these mechanisms, however the shorter chain length of the packing combined functionality on the chains cause the hydrophobic interactions to be less impactful. Instead, the solid phase attracts a layer of hydrophilic region around the packing, this acts as a secondary interaction, as the analyte has an affinity for the hydrophilic region, rather than the general mobile phase, creating a bi-phasic system and thus is still inhibited from being eluted based on it's hydrophilicity. This bi-phasic system is less robust that a clear split solid vs mobile phase of RPLC, it can be disturbed and broken down by higher flow rates or by rapidly changing mobile phase conditions (this tends to mean gradients are shallower and longer to allow the column enough time to adapt the bi-phase to the changing mobile phase conditions. However, the combination of the secondary interactions in the bi-phasic layer and primary interactions with the chain of the packing leads to a lot greater selectivity and sensitivity when it comes to the separation of analytes.

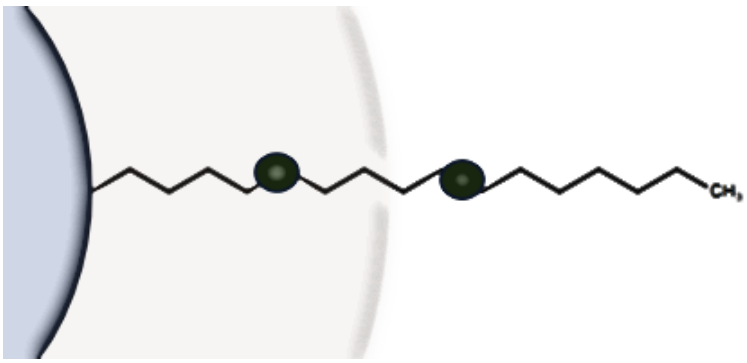

Figure 1.8: A simplified diagram of a HILIC stationary phase packing showing the chain used for primary interactions with analytes and a grey-scale region of highly aqueous nature which allows for secondary interactions with analytes.

1.6.2 Mass spectrometry

Mass spectrometry is a powerful technique for the rapid detection and identification of numerous compounds, simultaneously without the need for *a priori* knowledge of the compounds and has been applied to many different analysed in various fields[79,82,103]. Mass spectrometry generally consists of three basic principles; ionisation of compounds, separation of compounds by mass and detection of ions. Historically, compounds were ionised using high energy ionisation techniques, such as electron impact ionisation (EI) to create ions consistently from compounds[104]. These ions are accelerated by an electrical field and separated by a magnetic field perpendicular to the electrical field. The degree to which the ion path is distorted is related to the momentum of the ion (i.e. smaller ions are distorted more than larger ions), the ramping of this magnetic field allows different ion masses to separately detected creating a

mass spectra of mass to charge ratio (*m/z*). These earlier techniques would detect the ions on photographic paper[105].

As the field progressed, new techniques were developed. EI, a very high energy ionisation leads to high fragmentation of compounds, this destructive technique is good for learning about the structure of the compound based on the fragments, however the ion which corresponds to the compound (precursor ion) is often in low abundance. In a complex mixture of compounds, the loss of these precursor compounds can make determining the contents of the mixture more difficult as the abundance of the fragments become an eclipsing factor in the spectrum.

Softer techniques, such as electrospray ionisation (ESI)[106,107], leads to much less fragmentation, making the precursor much more significant in the spectrum and such a mixture of different compounds can be better identified by their precursor ions. The trade-off between soft and hard ionisations lead to the development of tandem mass spectrometry[108]. An initial soft ionisation leads to a spectra of precursor ions which are then fed into a collision cell, where the ions are fragmented, producing a spectrum of product ions.

An alternative soft ionisation technique, Chemical ionisation mass spectrometry (CIMS) is a commonly used technique, typically hyphenated to gas chromatography (GC) where a reagent species (for example ammonia) is ionised prior to analyte injection and this ionisation is passed via reaction from the reagent species to the analyte. CI was developed prior to ESI and is therefore more of a staple technique in older fields such as environmental science and food safety detection[109–111].

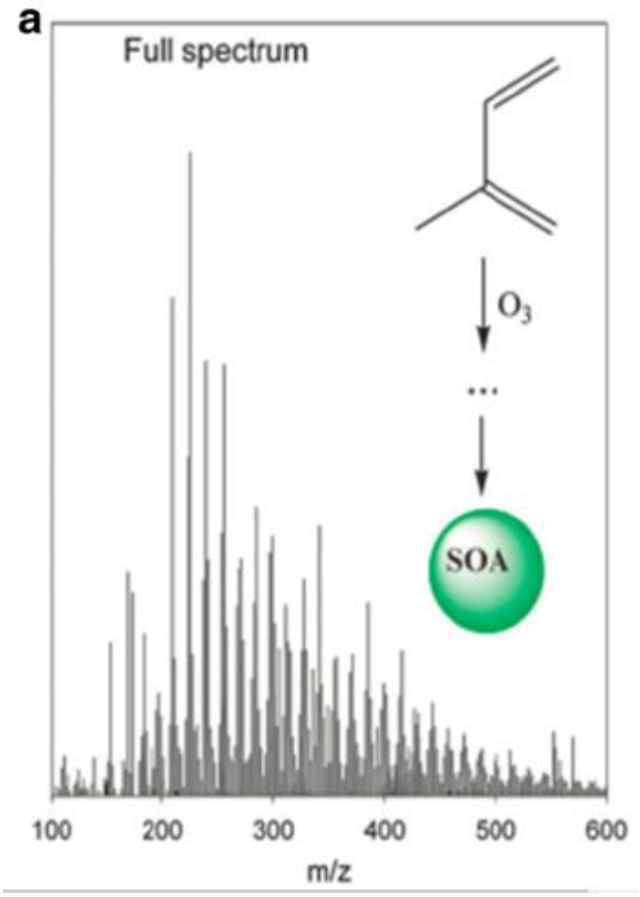

Figure 1.9: Mass spectra of SOA precursor ions from a sample of biogenic sourcing. Taken from Parshintsev 2015[112]

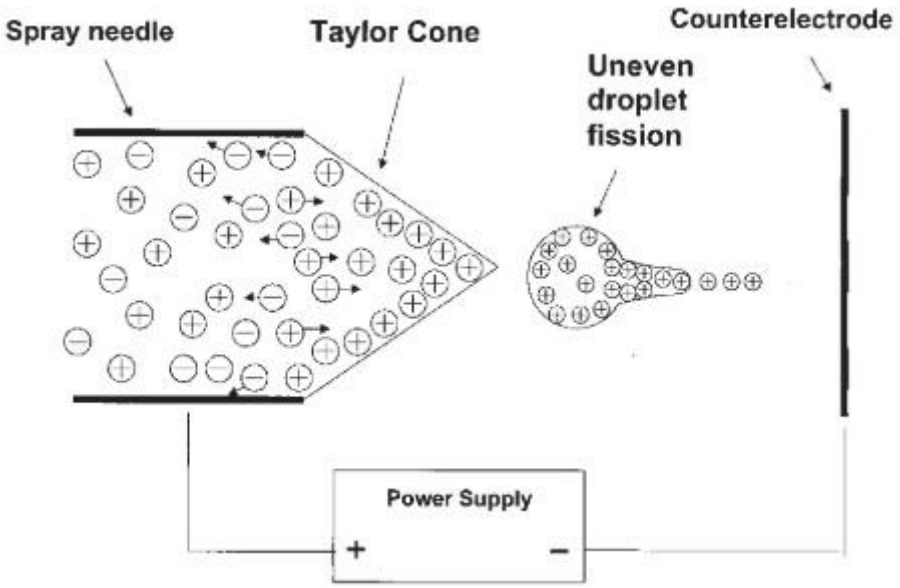

Figure 1.10: Diagram of electrospray ionisation (ESI) source needle and the build up of charge within a droplet until the charge overcomes the surface tension of the solvent droplet. Taken from Pramanik[113].

Detection and separation of ions based on their mass to charge ratio (*m/z*) is the central principle concept for mass spectrometry. Once all compounds within the analyte are ionised by the aforementioned ionization techniques, these compounds are then subjected to an electrical field and the differences in mass between these compounds, leads to differences in behaviour between these compounds when subjected to this electrical field. The most common example of this is time-of-flight mass spectrometry. Where ionised species are excited in an electrical field to a constant kinetic energy and then forced through a flight tube by electrodes. The time taken to complete the distance of the flight tube is related to the mass of each compounds (equation relating kinetic energy to the mass of the body is shown in equation 1.3), as all of those compounds possess the same kinetic energy upon firing[114,115]; the higher the mass, the

longer time taken to flow down the tube. At the other end of the tube, the detection of the ionised species (by use of photomultiplier tubes or other analogous detectors) gives the level of charge of the species, and thus *m/z* can be calculated from the two components. Using longer flight tubes, higher mass resolution can be achieved; however larger fight tubes make the system very cumbersome and does not offer a favourable trade-off between optimal resolution while maintaining bench-top suitability.

$$Kinetic\ energy\ (E_k) = ½\ m\ V^2 \qquad \text{(Eq 1.3)}$$

Where *m* is the mass of the body in motion and *V* is the velocity of the body

Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometry[116] offers a solution to the limited sizes of common mass spectrometry units. By subjecting the ionised species to a perpendicular field while in the flight tube, the ions undergo a helical motion. This significantly increases the flight-path lengths without the need to increase the size of the flight tube and hence MS unit. The advantage of this is two-fold. Higher resolution from the increased path length and higher selectivity of the species. This is due to the helical nature of the path length not having a constant radius, depending on the *m/z* values of the species. Species with lower m/z values would be more affected by the perpendicular field and thus would accelerate more in the field, spiralling with an increasing radius and eventually 'spinning out' by colliding with the walls of the fight tube before reaching the detector. As such, there is less of an ion count and the techniques becomes more sensitive. These perpendicular

fields can be varied across the scan or across numerous scans to be more selective of the detection of the desired *m/z*; as such, FT-ICR-MS offers very high resolution.

Orbitrap mass spectrometry is based on a similar principle to that of Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometry. Ions are trapped between two electrodes and rotated laterally around a spindle shaped electrode. The ions form rings around the spindle and the rings separate based on the *m/z* values of compounds. The frequency of the rings moving across the spindle is related to the *m/z* of the compound[117]. Orbitrap mass spectrometry has comparable resolution to the FT-ICR MS whilst offering a bench-top size; comparably, the hardware of an FT-ICR MS can be up to the size of a small car and thus, an orbitrap is an appealing alternative for smaller laboratories or laboratories which are tight on bench-top real-estate. However, orbitrap is not able to provide comparable resolution to TOF instruments and therefore a laboratory must decide if this trade-off is suitable for their in-house needs.

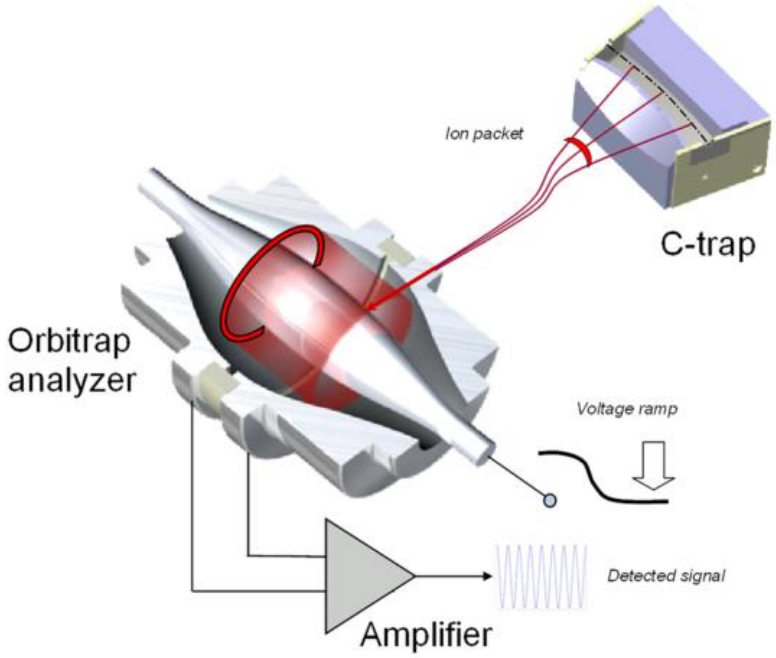

Figure 1.11: Spindle electrode in an orbitrap system. In red, a ring of ions can be seen moving along the spindle electrode and the signal being interpreted. Taken from Zubarev 2013[117]

1.7 The current state of air quality in China

China is a rapidly developing country with an estimated population of over 1.3 billion and its population is rapidly increasing due to recently overturned policies which formerly limited family sizes and social need to counter the problems arising from an aging population. This massively increasing population, particularly in densely populated areas, means more pollution and more people at risk from the effects of that pollution. Inhabitants of larger cities experience daily average $PM_{2.5}$ levels in excess of 100 μg m^{-3}, putting these people at risk of serious negative health effects[118]. Examples of this include Lang *et al* who reported that the average $PM_{2.5}$ levels for Beijing were

106 μg m^{-3}, ten times higher than the WHO's recommended maximum safe limits[7]; leading to increased risks from cancers and other chronic diseases. Similarly, Wang *et al* reported average $PM_{2.5}$ levels in Shanghai of 103.07 μg m^{-3} demonstrating that this is not just an issue limited to the capital, instead all densely populated zones are at potential risk. The Ministry of Environmental Protection of People's Republic of China (MEPPRC) reported that for 2013, the average $PM_{2.5}$ levels across China were 72 μg m^{-3}, suggesting this problem still affects populations outside of the larger cities to a lesser, but still significant degree. This demonstrates a wider-spread risk of diseases across China as a whole.

During high pollution events, $PM_{2.5}$ has been seen to reach particularly high levels, to the point where the pollution is visible[119,120] with maximum $PM_{2.5}$ levels within the developed city of Lazhou reported as 356.8 μg m^{-3} in 2006 by Zhou *et al*[34] (>35 times the safe exposure levels set by the WHO). This was compared to the peak levels of $PM_{2.5}$ for a less developed, forested area, Waliguan which were measured at 8.8 μg m^{-3} [34] in 2010. Rohde and Muller present a similar report for Beijing, where $PM_{2.5}$ levels of 250 μg m^{-3} during a spring campaign were observed[121]. The maximum $PM_{2.5}$ concentration observed for the Baoshan district of Shanghai was 168 μg m^{-3} for a winter campaign[118]. This consistent exposure to $PM_{2.5}$ levels in excess of WHO limits is detrimental to the health of the people of China. It is estimated that the death toll attributed to air pollution in China is between 1-2 million[121,4].

Understanding the chemical nature of this particulate matter is a key method to understanding its source and is the first step towards improving the air quality of China; as previously discussed more information on the sources leads

to better informed policies and more cost-effectively implementation of solutions.

The Chinese environmental yearbook in 2010 estimated that 507 Mt of soot, 1147 Mt of SO_2 and 936 Mt of NO_x were emitted from China. Reports state that sulfate accounts for 18 ±10 μg m^{-3} (19.5%) and nitrate accounts for 7.7 ± 5.7 μg m^{-3} (7.3%) of the $PM_{2.5}$ in China[34]. This leaves a vast amount of $PM_{2.5}$ unaccounted for. Zhou *et al* report that approximately one quarter of $PM_{2.5}$ (25.7%) is carbonaceous species and of this fraction 76% is organic carbon (OC) in nature[34].

The composition of $PM_{2.5}$ at a specific location varies depending on source and season. In Beijing, the key sources of $PM_{2.5}$ were determined (by emission inventory and numerical modelling of the emissions) to be vehicular (attributed to 31.5% of $PM_{2.5}$), industry (22.9%), residential coal combustion - RCC (14.5%) and industrial coal combustion – ICC (14.5%). Similarly, in Guangdong the main sources of ambient $PM_{2.5}$ were attributed to industrial (22.5%), residential (13.1%), traffic (19.3%), agriculture (17.5%) and biogenic (15.3%)[24].

Differences in the contribution proportions were observed in this data set; the biogenic contribution varied between 4.8 and 26.7% of ambient $PM_{2.5}$ for winter and summer respectively. This reflects the difference in solar radiation intensity during these periods and hence plant emissions. A maximum contribution from agricultural sources was observed in autumn months (19.5%), as a result of the seasonal harvesting of crops around this period[24].

The peak emissions from traffic (24.0% of total $PM_{2.5}$) occur during the spring, however the minimum contribution period is only 17.7% (during the autumn)[24], suggesting a constant emission from traffic which is not significantly affected by season. This contribution from traffic is expected to rise as vehicle ownership has consistently increased by 17% year on year[34].

Thesis outline

The purpose of this thesis is to investigate the composition and sources of organic aerosol. To accomplish this, a custom built high-throughput ultra-high performance liquid chromatography with tandem mass spectrometry (UHPLC-MS^2) method and data processing software were developed and implemented on the analysis of off-line filter samples of $PM_{2.5}$ from Beijing, China. A brief overview of the work contained within this thesis is as follows:

Chapter 2 – This work describes the method development and optimisation of the UPLC-ESI-orbitrap-MS^2 analysis methodology. This chapter also describes the development of the database used in the software and validated using samples from previous campaigns in London, UK and Birmingham, UK.

Chapter 3 – In this study, the organic aerosol from two campaigns in Beijing, China are investigated. The overall composition of the observed aerosol is described and variations within campaign and between campaigns are explored.

Chapter 4 – This study investigates the nitrophenolic compounds in greater depth. A higher resolution method for specific nitrophenolic analysis is demonstrated and abundant species are quantified, and preliminary source apportionment performed using gas-phase and AMS data to corroborate the analysis from the UPLC-MS^2 analysis.

Chapter 5 – This work then explores, the organosulfate fraction of the aerosol from Beijing is explored. Greater source apportionment is performed and observed species are compared with HILIC method performed in collaboration with the Surratt group in North Carolina, USA.

Chapter 6 – A summary of the work performed and future work is discussed with links to applications in informing policy and links to previous analysis.

Chapter 2 – Development of method for detection of organic aerosol and Software for the Automated Processing of Aerosol (SAPA)

2.1 Introduction

Poor air quality and the associated risks are the top environmental risk factor, with over 4 million linked deaths from long-term exposure in 2015[2–4]. Acute conditions which occur from extreme exposure to high air pollution are strokes and heart disease[7]; less severe chronic effects, which are of high concern, include cardiovascular and respiratory diseases including cancer. Due to these effects, the International Agency for Cancer Research (IARC) has classified air pollution as a group one, known carcinogen due to particulate matter pollution being a strong risk factor for cancer[122].

In urban areas, $PM_{2.5}$ is mostly dominated by species containing carbon, usually as a mix of both elemental carbon and organic carbon-based molecules. Prior studies have shown that $PM_{2.5}$ from urban sites contain 10,000s of individual organic molecules[8,19]. The most complex fraction of particulate matter is organic aerosol (OA) and therefore it is the most difficult part of air pollution to understand fully. Fortunately, the complexities of organic aerosol offer insight and can be used to determine the sources of pollution, with unique combinations of tracer compounds to form 'fingerprints' for a source. Organic aerosol can be both directly emitted (primary organic aerosol, POA) and formed from reactions from species within the atmosphere (Secondary organic aerosol (SOA). Sources of POA come from a range of sources, including anthropogenic

sources, such as diesel exhaust[20], industry[21], manufacturing, cooking[22,23] and biomass fuel combustion[24]. POA can also have natural sources such a plant emissions[25–27], forest fires[28] and from the sea surface microlayer[29]. SOA formation is predominantly from the oxidation of volatile organic compounds (VOCs) from sources such as plant emissions like monoterpenes and anthropogenic VOCs[20,39,49,65,123–125]. The oxidation of these volatile compounds results in lower volatility and hence the partitioning of these compounds into the particulate phase.

Source apportionment techniques, such as a Chemical Mass Balance (CMB) modelling, can be used to assess the contributions of these sources to the composition of organic aerosol. CMB uses pre-defined chemical speciation fingerprint profiles to assess the contributions from each source type. These fingerprint profiles are based on laboratory tests, typically chamber experiments, sampling various sources or replicating source conditions[67,126].

Aerosol Mass Spectrometry analysis (AMS) is a measurement of the bulk aerosol and has shown that more oxidised material dominates particle mass in both rural and urban locations. These more highly oxidised compounds are mostly products from the oxidation of VOCs and as such the dominant species in aerosol tend to be SOA[30]. However, due to the increased complexity that comes with the number of possible reaction pathways VOCs can undergo, and the composition therefore exponentially increasing in complexity and with continually changing formation conditions the composition is constantly in flux. The composition will also change as emissions undergo complex chemical reactions and transport of material to and from sampling site, controlling the supply of VOCs and changing the possible formation pathways. These

numerous points of variability lead to the nature of the aerosol being in constant flux. It thus becomes increasingly clear that the need for bulk analysis is significant to gaining a more complete understanding of organic aerosol *in situ*. Unfortunately, the high temperatures and energetic ionisation conditions used in AMS destroys the majority of the compositional detail during analysis and as such is not always appropriate for the whole analysis of this dominant oxidised material, particularly in urban environments[69].

Advances in mechanistic and process level understanding have been achieved in recent years[28]; however as our understanding of these source fingerprints are limited mostly to chamber and laboratory replications, our ability to characterize all of the potential compounds is limited. There have been attempts to use CMB to quantify the sources of SOA, using a very limited subset of tracer compounds based on these chamber experiments[20,70]. However, the dynamic nature of the composition of SOA means this approach is difficult, particularly on very aged air masses which have undergone larger levels of oxidation and thus the composition may not resemble the freshly emitted pollution. Thus even fundamental questions, such as whether natural or anthropogenic emissions form more SOA in cities, and what factors influence the formation of SOA, remain mostly unanswered.

Direct infusion of aerosol extracts into high resolution mass spectrometry, with a soft ionisation (such as electron-spray ionisation, ESI or atmospheric-pressure chemical ionisation, APIC) enable ions to be identified by their chemical formula[8,77,127] without losing the precursor ion to fragmentation; and implementation of short scan times of these high resolution techniques allow for thousands of compounds to be rapidly analysed by this technique. Examples

of this method include orthogonal acceleration time of flight mass spectrometry (oa-TOFMS)[128,129], Orbitrap mass spectrometry[64,130] and Fourier-transform ion-cyclotron resonance mass spectrometry (FT-ICR MS)[73,130]. These techniques tend to be fast, allowing for near-real time sampling if the cycle time of an injection is less than the sampling rate. In very recent times, systems such as the Filter Inlet for Gases and AEROsols (FIGAERO-CIMS) allow for the collection of both aerosols and gases pseudo-simultaneously; aerosols are adsorbed onto a filter unit and thermally desorbs the organics into a chemical ionisation mass spectrometer[127] and gas-phases analysis is performed in between aerosol samplings, usually in two or three hourly sampling resolution, allowing for better insight in to the variability of aerosol across a day as opposed to an averaged sample across the whole day. This technique adds a new dimension in analysis and for near real-time comparison of gas phase and particulate phase of organic aerosol.

Such a direct comparison allows for investigation of aspects such as air mass aging, the oxidation of VOCs and transference into particulate SOA. In all of these methods, structural isomers cannot be separated, and low concentration species may be eclipsed by the signal from more abundant isomers or those with a higher ionisation efficiency. To overcome such limitations, hyphenated techniques such as the combination of liquid chromatography (LC) or gas chromatography (GC) combined with MS are often employed[73,74]. These techniques separate out compounds with the same *m/z* values based on their affinity to a stationary phase, potentially allowing for identification of isomers assuming sufficient resolution or enough of a different in affinity to stationary phases between isomers (although, variable phases, such as polarity and chirality can be employed as alternatives to ensure separation[131,132]).

The use of a relatively short LC method offers a trade-off between resolution and utility for the method, and the needs for resolution and applicability in different analysis is explored more thoroughly in chapter 1. Most tailored analysis used in fields such as analytical and bio-analytical science use chromatography which is less than 8 minutes for the gradient, however, the target analytes tend to be fewer than ten per analytical method. A method which therefore aims to identify thousands of compounds in one method could feasibly have its chromatographic gradient extended up to the hour mark to achieve adequate resolution of all compounds; albeit the tailoring of a gradient which provides adequate separation of a vast number of species, each with significantly different chemical properties from one another, would be incredibly complex. It follows that such a method would be cumbersome to utilise and develop, and still be limited to known species.

Another limitation which presents itself from such a method would be the selection of the chromatographic column itself. Due to the lowered volatility of the compounds found in aerosol, the obvious starting point would be HPLC columns. The most common starting point would be C18 packing, and indeed this offers a very robust separation of compounds for most functionalities; however, for compounds with very similar properties, this packing material can insufficiently separate compounds with similar polarities. With *a priori* knowledge of the composition of the samples, there are C18-based packings with additional functionality which may offer improved separation for otherwise poorly-resolved species. Without any *a priori* knowledge, it is arguable if packing with C18-based variations would offer a sizable benefit over a straight C18 packing.

The employment of a different based packing such as HILIC, again would change the retention order of compounds and offer improved separation of species with certain characteristics. In the case of HILIC, improved separation of polar compounds is seen, with a significant reduction in separation of non-polar species. Other packings, such as phenyl-hexyl or pentafluoro-phenyl (PFP) show good separation for species containing aromatic groups and species of different sizes and steric compositions, (due to the larger steric activity of phenyl-containing packing, when compared to the more slender and flexible C18 packing) but show poorer separation for aliphatic species. This again presents the *a priori* knowledge issue. When theses column packings are employed, it is because the analyst has an idea of the types of compounds they want to resolve and potentially what they are resolving from. Thus, with no *a priori* knowledge of the sample, it is not guaranteed what success would be gained in terms of separation of complete unknowns in a complex system. In fact, employing these columns may offer no separation to the most prevalent species within the aerosol sample.

A C18 analysis can be performed to 'scout-out' the types of compounds within an aerosol sample, and then tailor an analysis using a more targeted packing in a column, either by two separate analyses or by using a tandem, 2-dimentional chromatographic method. However, the longer run times required by the C18 separation alone to achieve sufficient separation to identify the species to begin with, could be so long, that adding an entire second analysis may not be worth the time investment; particularly if the number of samples is very large or if the analyst is wanting to get a pseudo-real time analysis (i.e. sampling every hour).

In the field of agricultural analysis, monolithic columns are used for the separation of Humic-like samples, with many components (in an analogous way to aerosol). Monolithic columns approach chromatography differently, as they do not possess the traditional packed 'beads' with coating to induce the solid-phase, but instead are one continuous, porous rod (giving rise to the name) which itself possess the solid-phase coating. This in a way acts like a better packed column, leading to more consistent, robust and improved separation compared to a regularly packed column. These can also be put in tandem to increase column length to 25 cm, compared to a regular 5 or 10 cm column. This increased length and improved porosity leads to a significantly improved separation; however, being a relatively new column technology, they are not produced by many companies, have not been adopted often by analytical organisations (due to their sometimes prohibitive expense) and do not come with a range of all the solid phase coatings which might commonly be employed during analysis. As a result of this, monolithic columns are not often used in routine analysis, however research with their use is increasing in utilisation in recent years.

The solution to the challenge of separation and identification comes from the high-resolution mass spectrometry (HRMS), which can be employed to aid the chromatography. Even if peaks for different species may not be fully separated chromatographically, the species can be separately identified using the mass spectra and the abundancies of their ions. Therefore, the HRMS can be used as pseudo-separation in the identification of compounds. The removal of the necessity to have long gradients (and there for cycle times) means a method can be tailored to be shorter than the sampling rate for aerosol and potentially be used as pseudo real-time measurements.

The limitations of this technique should not be overlooked; the shorter run time means poorer chromatographic resolution and relying on HRMS to accommodate for the limitations, does not solve all of the problems. Compounds which cannot be distinguished by m/z and have similar chemical properties would be indistinguishable in this method and therefore not be identifiable in isolation. Examples of such compounds which may be lost to this shorter method include enantiomers and structural isomers.

When identifying distinct groupings and clusters of similar compounds, maximum distance between the mid-points of groups of similar compounds dictate the separation of clusters. The advantage of clustering such data using HCA, is that it presents a very clear and simple dendrogram for easy interpretation, however, since this is based on the logical distances between clusters, the data may not, in actuality, possess the structure which the HCA implies. A further limitation is that the number of clusters is often determined by the researcher who, when determining clustering criterion (or cut-off point for number of clusters), is inherently introducing a source of bias, this can be particularly problematic when interpreting very large data sets. Treiger *et al* reviewed some clustering techniques and offered some solutions to overcoming some of the limitations which arise from their application to large datasets[96] and some of their caveats are included in more recent work. The easiest method recommended and often employed in order to by-pass this limitation is the sub-setting of the data prior to analysis.

This work describes the development and validation of a high-throughput, selective and sensitive method to identify organic compounds present in aerosol extracts. A mass spectral database of organic aerosol tracer molecules

was built using literature data and samples from previous field studies and simulation chamber experiments. A reverse-phase ultra-performance liquid chromatography coupled to high-resolution orbitrap tandem mass spectrometry (UPLC – Orbitrap MS/MS) method was developed and the data produced analysed using a bespoke, automated program based in the language R. This method was used to investigate the trends in organic aerosol during two field campaigns in Beijing, China in the winter and summer of 2016-2017.

2.2 Experimental Methods

2.2.1 Filter sample collection and extraction of $PM_{2.5}$

Aerosol samples were collected during two campaigns (9th November to 9th December 2016 and 18th May to 24th June 2017) at the Institute of Atmospheric Physics (IAP) in Beijing, China. This work was done as part of the Sources and Emissions of Air Pollutants in Beijing (AIRPOLL-Beijing) project funded by the Natural Environment Research Council (NERC) in the UK and the National Natural Science Foundation of China (NSFC) in China. $PM_{2.5}$ filter samples were collected using an ECOTECH HiVol 3000 (Ecotech, Australia) high volume air sampler with a $PM_{2.5}$ selective inlet at a flow rate of 1.33 m^3 min^{-1}. Compatible filters (Pallflex Filters, 200 x 250 mm; Ecotech, Australia) were baked at 500 ^oC for 5 hours prior to use, this was to purge the filter of any VOCs or PM which may have become adhered during the manufacturing process or during shipment.

After collection, samples were wrapped in foil and stored and shipped to the laboratory at -20 ^oC to ensure that the only matter on the filter was from

collection periods and avoid contamination. Samples were collected at a height of approximately 8 m from ground-level, on top of a building with the IAP complex, which is situated in Beijing, China (reference co-ordinates: 39.9780° N, 116.3871° E). Samples were collected every 3 hours during the daytime (approximately from 08:30 to 17:30) and one sample was collected over night between 17:30 and 08:30. On 10 highly polluted days (where $PM_{2.5}$ levels exceeded 100 μg m^{-3}), daytime samples were collected hourly to get a more precise temporal resolution which maintaining enough loading per filter to ensure a good quality of data.

The extraction of organic aerosol from filter samples was performed in a method similar to that of Hamilton *et al*[61]. To summerise, the method is as follows: approximately 1/8th of the filter was cut into small pieces and solvated using 4 mL of LC-MS grade water for 2h to extract the WSOC. The samples were sonicated for 30 minutes to ensure full dissociation of PM and filter paper, then the mulch was filtered through 0.45 μm pore PTFE syringe filters (Millipore, Germany). The extracts were evaporated to dryness using a vacuum solvent evaporator (Biotage, Sweden), at 25 ◦C and reconstituted in 1 mL of 50:50 methanol/water to allow sufficient solubility for water soluble and less water soluble species.

2.2.2 Standard mixtures

Reference calibration curves were used to determined semi-quantification of species as by grouping into common functionalities. The generation of the co-spiked calibration solutions of which the calibration curves were comprised,

were as follows: primary stock standard solutions at 2 ppm were generated for 13 nitro-aromatic compounds; including four organosulfates, four organic acids and one terpene-derived nitroxysulfate (full list of standards in table 2.1.) using commercially available standards (all compounds from Sigma Aldrich, USA). All solutions were dissolved in 50:50 HPLC grade methanol and LC/MS grade water (both from Fisher Chemical, USA) to match the loading conditions of the reconstituted extracts. These were further diluted to 100, 50, 25, 12.5, 6.125, 3.0625 and 1.53125 ppb using the same 50:50 methanol/water diluent. Standard solutions were stored at -20 ◦C until use. All of these calibrations showed good linearity and good precision (R^2 >0.9 for all but two compounds, 2-nitrophenol and 2-nitro-1-naphthol, which have R^2 >0.85). This suggested that using these compounds as proxies for calibration in quantification would be suitable for the purposes of this analysis.

Table 2.1: Compounds used as standards for quantitation during the analysis. The percentage relative standard division (%RSD), R^2 and limit of detection (LOD).

Compound	Formula	Group	%RSD	R^2
***cis*-Pinonic Acid**	$C_{10}H_{16}O_3$	CHO	1.81	0.9835
Azelaic Acid	$C_9H_{16}O_4$	CHO	5.12	0.9912
2-Methyl-3-Nitrophenol	$C_7H_7NO_3$	CHON	1.22	0.9887
2-Methyl-5-Nitrophenol	$C_7H_7NO_3$	CHON	2.82	0.9694
4-Methyl-2-Nitrophenol	$C_7H_7NO_3$	CHON	7.48	0.9036
2,4-Dinitrophenol	$C_6H_4N_2O_5$	CHON	2.21	0.9646
4-Nitro-*m*-Cresol	$C_7H_7NO_3$	CHON	2.99	0.9934
2,6-Dimethyl-4-Nitrophenol	$C_8H_9NO_3$	CHON	7.14	0.9980
2-Methyl-4-Nitrophenol	$C_7H_7NO_3$	CHON	2.99	0.9919
3-Methyl-2-Nitrophenol	$C_7H_7NO_3$	CHON	1.71	0.9786
4-Nitrophenol	$C_6H_5NO_3$	CHON	1.74	0.9778
3-Nitrophenol	$C_6H_5NO_3$	CHON	1.89	0.9736
2-Nitrophenol	$C_6H_5NO_3$	CHON	7.48	0.8901
2-Nitro-1-Naphthol	$C_{10}H_7NO_3$	CHON	5.85	0.8637
4-Methoxy-2-Nitrophenol	$C_7H_7NO_4$	CHON	9.98	0.9954
2-Naphthyl Sulfate	$C_{10}H_8SO_4$	CHOS	2.07	0.9703
Octyl Sulfate	$C_8H_{18}SO_4$	CHOS	4.85	0.9724
6-Mannose-D- Sulfate	$C_6H_{12}O_8S$	CHOS	3.31	0.9503
Dodecyl Sulfate	$C_{12}H_{26}SO_4$	CHOS	2.52	0.9745
4-Nitrophenol Sulfate	$C_6H_5NO_6S$	CHONS	13.3	0.9923

2.2.3 UPLC-MS^2 method

2.2.3.1 UPLC method development

The diverse range of characteristics within the organic aerosol in Beijing which is documented in literature necessitated the choice of column was chosen so as to be as generic as possible with little functionality outside of standard non-polarity seen in reverse phase chromatography. This broad scope approach to the choice of packing allowed the method to give an overview understanding of the major groups of compounds, without focussing one set of properties over another; but also allow for scope for more specific, tailored analysis going forward (such as chromatography with HILIC methodology). As previously discussed, most broad-scope analysis was performed using C18 based packing. Due to availability, a 10 cm HPLC C18 column was selected; this was to allow the greater level of separation in the 10 cm column compared to analogous 5 cm columns with no favouring of a particular property or functionality; which can occur from C18 end-capped columns.

As no *a priori* knowledge was obtainable regarding the full complexity of organic aerosol, selecting packing other than a standard C18-based packing would lead to biases on which compounds would be analysed, for example aromatics would be retained much better on bi-phenyl packing but aliphatic surfactants would be lost nearer the dead-volume. Therefore, it was decided that a C18 based packing be used as this is one of the more universal column packing chemistry and allows for retention of a greater range and number of compounds potentially relevant to this study. C18 based packing tends to be more robust than the more complex chemistry used for chromatography. Given the large number of injections and runs required for this size of an analysis

(and any potential re-analysis), a consistent, robust column would be less likely to fail during a run and result in the loss of data; it also allows the retention times to be consistent between runs (and indeed between batches of columns) allowing for a steady RT which can reliably by analysed by an automated processing without the need for human oversight and tweaking of detection parameters. This allows the whole process to be much more hands-off.

The development of the UPLC method was performed using a mixed solution of standards as described in Pereira *et al* 2004[133]. Five compounds (2-hydroxyhexanoic acid, cis-pinonic acid, adipic acid, 4-methoxybenzoic acid and 2,6-dimethyl-4-nitrophenol) at concentrations of 1, 0.5, 0.25, 0.125 and 0.0625 ppm were run on C18, reversed phase 5 µm, 4.6 x 100 mm, Accucore column (Thermo scientific, UK) to test the separation of the different calibration compounds. It should be noted that Pereira *et al* used a HPLC method with a flow rate of 1.3 mL/min, with the employment of UPLC, a much higher resolution can be achieved at a much lower flow rate due to the higher pressures that can be achieved. Unfortunately, we did not have available columns with smaller packing (i.e. 1.7 µm or 2.5 µm, rather than the 5 µm packing available) at the start of the work and as such we were unable to maximise the resolution possible with UPLC. Since the work going forward was based on fixed RTs for compounds, this could not be updated midway through the work; however, offers an interesting opportunity in any future work. The reduction in flow rate compared to the method by Pereira *et al* allows for better separation as fast flow rate encourages the faster elution of peaks, leading to both thinner peaks and with better resolution, assuming the peaks are separated; however, it also reduces the time between elutions and leads to poorer separation.

The compounds were chosen to best represent a range of functionality and sources which commonly occur in aerosol. For example, cis-pinonic acid is a common biogenic tracer, methoxybenzoic acid is an anthropogenic tracer and nitrophenol is a tracer of transportation emissions; as such with a small number of compounds, a method which can adequately characterise different tracer groups using their representative compounds, could be developed. The compounds chosen are well documented as reference standards for atmospheric analysis and are robust as standards, making them ideal reference materials for the development of an analytical technique.

With concentrations of individual species ranging from ppt (parts per trillion) to ppm (parts per million) depending on the compound and its sourcing, getting a representative concentration range for a broad range of compound is difficult. The issue of selecting the correct range of concentrations when developing a calibration range, is compounded by the non-linearity which can occur in calibration curves when concentration ranges exceed 1000-2000 fold. This is largely due to detector sensitivity when using mass spectrometry; typically, mass spectrometry uses electron-multiplier tubes as detectors. The automatic gain (the system which amplifies signals from ions to allow the computer to read it as a 'true' peak) used to detect low-level signals can become saturated at the high levels, leading to a plateauing of the calibration curve at the higher end and thus making them inefficient at detecting high concentrations, if the sensitivities are tailored for low concentrations.

In routine analysis, this can be solved by separating the samples into a high-range analysis and low-range analysis, adjusting the automatic gain settings to suit each method. However, this is not suitable for very large variations in

concentration, and ionisation efficiency; finding a single set of parameters to suit every compound is near-impossible with current technology. Additionally, creating a high and a low range assay for time resolved aerosol would double the run time for any set of samples being analysed in both assays and this could result in numerous extra hours of analysis time for little gain.

As a starting point, the parameters were optimised to best suit the most prevalent of the compounds observed in the aerosol, since less prevalent compounds (or those with significantly poorer ionisation efficiency) will likely be eclipsed by more prevalent species in the analysis and therefore be indistinguishable from the baseline (i.e. not have a signal to noise ratio adequate to confirm a signal). It should be noted that there is scope to tailor parameters such as the gain controls to suit the smaller signals in future work.

Regarding the choice of mobile phase, this was also done to provide the most broad-scope possible. In routine analysis, typically, a relatively more polar mobile phase A (such as water) is paired with a less polar mobile phase B (such as methanol, MeOH or acetonitrile, ACN). Acetonitrile is less polar than MeOH, and since we expect the composition of aerosol to be more functionalised (and hence more polar) as it loses volatility, as seen in Donahue 2012[37], it is expected that better separation can be achieved from two more similar phases. As such water and methanol were chosen as mobile phases A and B respectively. In literature, acidifying mobile phase A has shown better resolution; therefore, to achieve the broader scope of analysis, the generic acidification of the water was performed. The final mobile phase A was UPLC-grade water: Formic acid (100:0.1 v/v).

To begin with, a commonly used 10-90% B gradient will be employed to get a sense of the distribution of characteristics in the reference material and tailored then based on outcome (Figure 2.1). Compounds which elute very quickly from the column (less than 1 minute and with a retention factor of <1.5) were not retained by the column and hence not properly resolved. Due to the expected concentrations of species in the ambient samples, an increased time for compounds to be washed from the column was proposed. This was done by increasing the length of time at the 95% MeOH plateau (Figure 2.2).

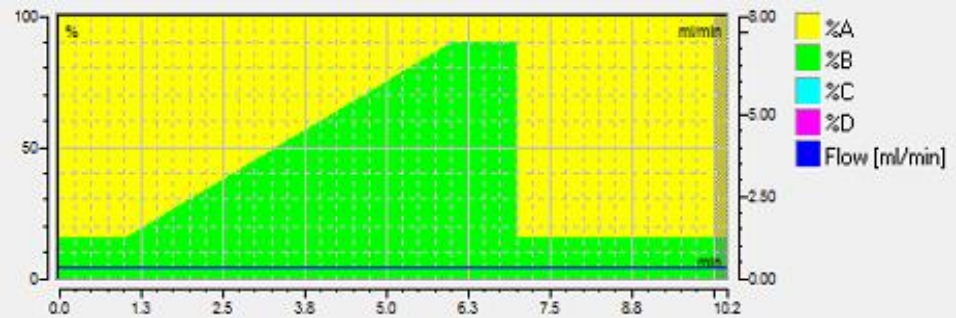

Figure 2.1: Gradient of elution of default short method for UPLC analysis. Eluent A: UPLC grade H_2O, with 0.1% by volume of formic acid. Eluent B: 100% UPLC grade MeOH.

Samples from the EROS and ClearfLo campaigns were analysed on the method with longer cleaning steps, no carry-over was observed between samples. As previously discussed, the abundances and numbers of compounds were significant and an increased gradient length was implemented to have better separation of species and isomers within the sample (figure 2.3). To avoid

severely long run-time, the wash-step time was reduced to the minimal time to ensure no carry-over.

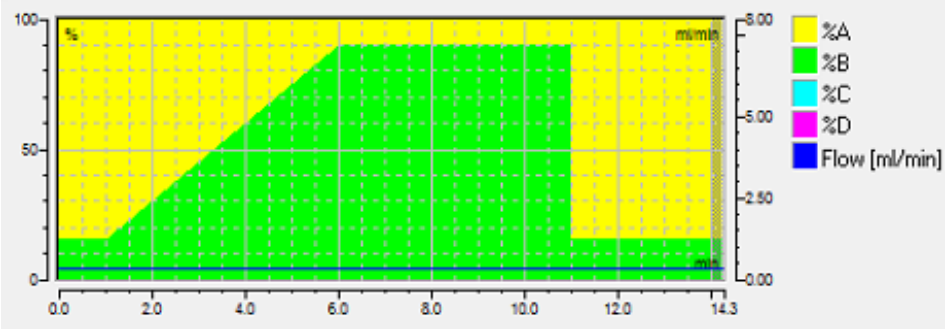

Figure 2.2: Gradient of elution of UPLC analysis with extra long high eluent B time (5 mins). Eluent A: UPLC grade H_2O, with 0.1% by volume of formic acid. Eluent B: 100% UPLC grade MeOH.

The length of the gradient from 10% MeOH to 80% MeOH was increased by from 5 to 8 mins and the cleaning step time was reduced to maintain a short run time of less than 16 minutes. This 16 minute run time was chosen to ensure all of the samples could be analysed in a timely manner. With pre-injection washes, post-injection washes and autosampler motions, the cycle time of this analysis was 20 minutes. A sample from the EROS campaign was analysed by both the shorter (figure 2.4) and longer (figure 2.5) methods compounds seen in the shorter method around 8.52 mins can be seen to elute later in the longer method. In the space formed, smaller peaks can now be seen in the longer method; these compounds would have been eclipsed in the shorter method by the larger peaks. As such the longer method was chosen to effectively separate as many components as possible.

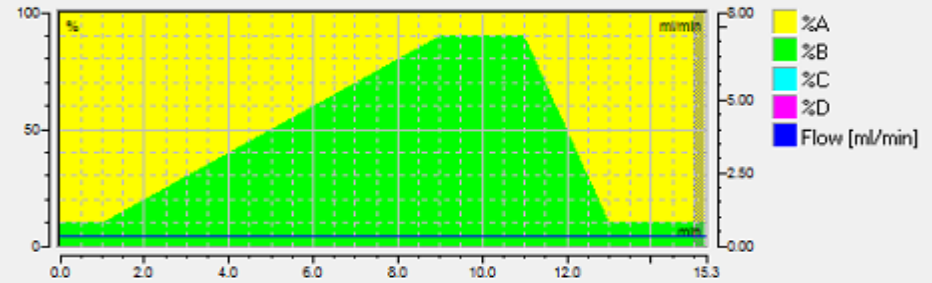

Figure 2.3: Gradient of elution for UPLC analysis with longer time for gradient change in eluent composition. Length of cleaning time is half high eluent B, half a gradient return to high eluent A conditions. Eluent A: UPLC grade H_2O, with 0.1% by volume of formic acid. Eluent B: 100% UPLC grade MeOH.

This longer gradient gave very good separation of the significant peaks within the base peak chromatogram (BPC). A representative BPC for a ClearfLo sample seen in figure 2.6. The most prevalent, obvious features with the ClearfLo sample are the sharp peaks between 0.80 mins and 2.00 mins. This collection of features are the very poorly retained and completely unretained, polar species. These species include the majority of organosulfates and demonstrate one of the limitations of C18 use in analysis; whilst most compounds are retained and show good resolution from each other as seen in the EROS sample in figures 2.4 and 2.5, it is not a flawless analysis. For a better understanding and better separation of this group, other chromatography, such as HILIC packed LC, is recommended; albeit this may lead to a poor retention of other groups. Towards the very end of the run, dimers and very hydrophobic compounds can be seen in the BPC (after 9 mins). These compounds are significantly less abundant, compared to the poorly

resolved, polar species at the start of the run and thus gives some overview insight into the aerosol; for example, the nature of the composition is that of a more aged air-mass, with more functionalisation and less volatility, hence the abundance of the more polar species compared with the non-polar region. Additional features can be seen across the run, which correspond to a range of other key groups of differing properties.

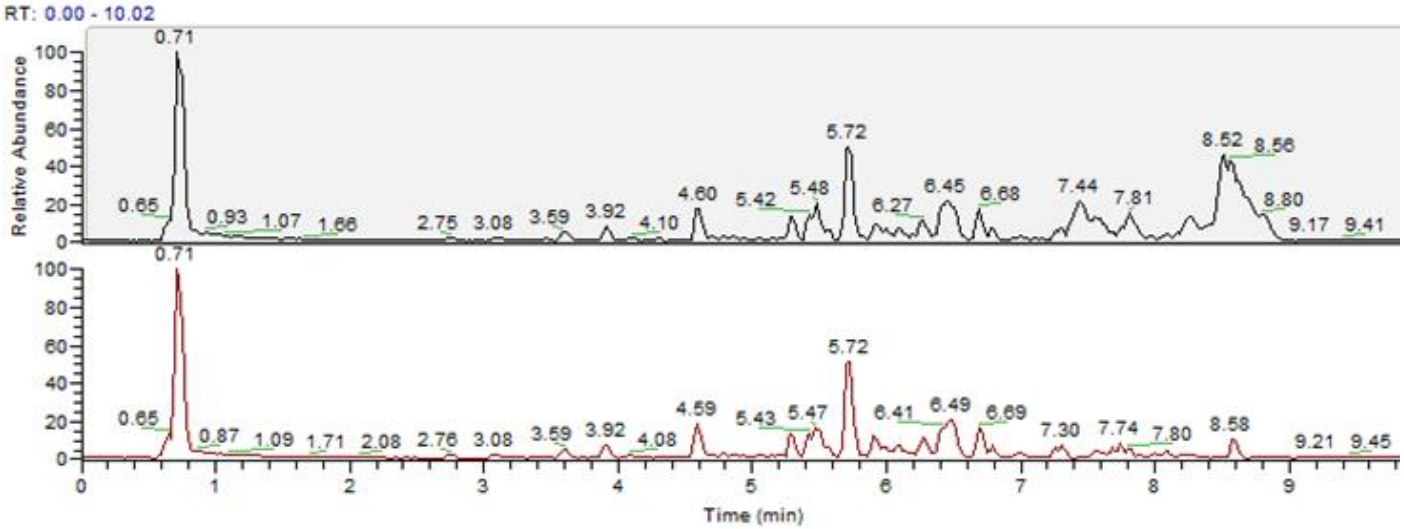

Figure 2.4: Base peak chromatogram for the shorter length gradient. BPC shows most abundant peaks from ambient sample from EROS campaign.

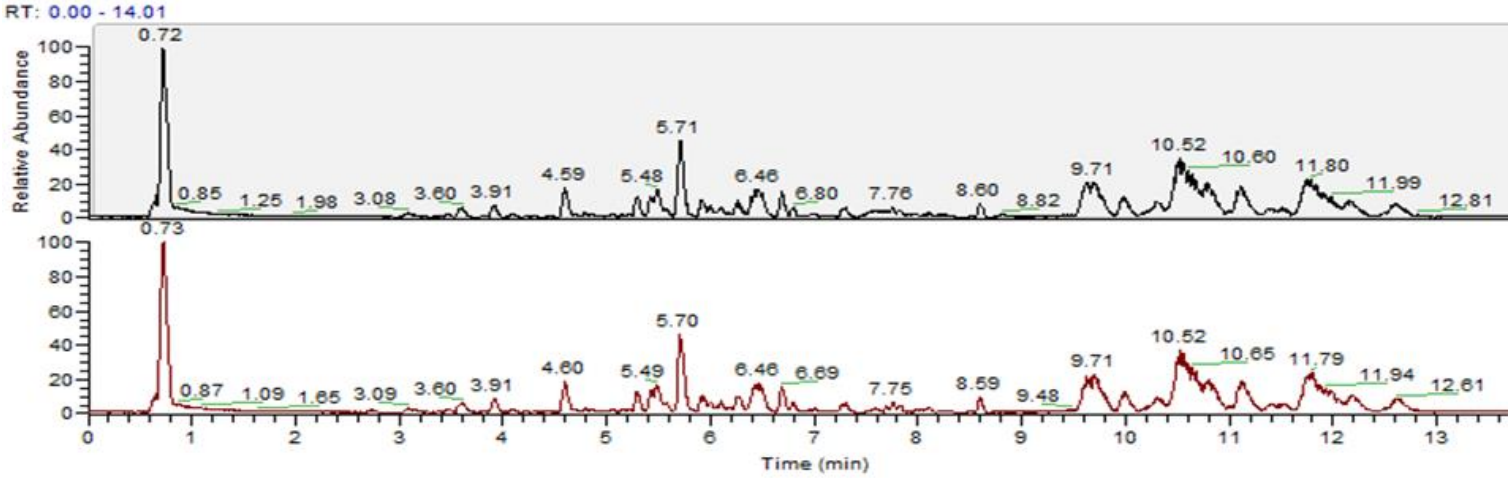

Figure 2.5: Base peak chromatogram for the longer length gradient. BPC shows most abundant peaks from ambient sample from EROS campaign, including peaks which were not observed in the shorter gradient method BPC (particularly at the later retention time).

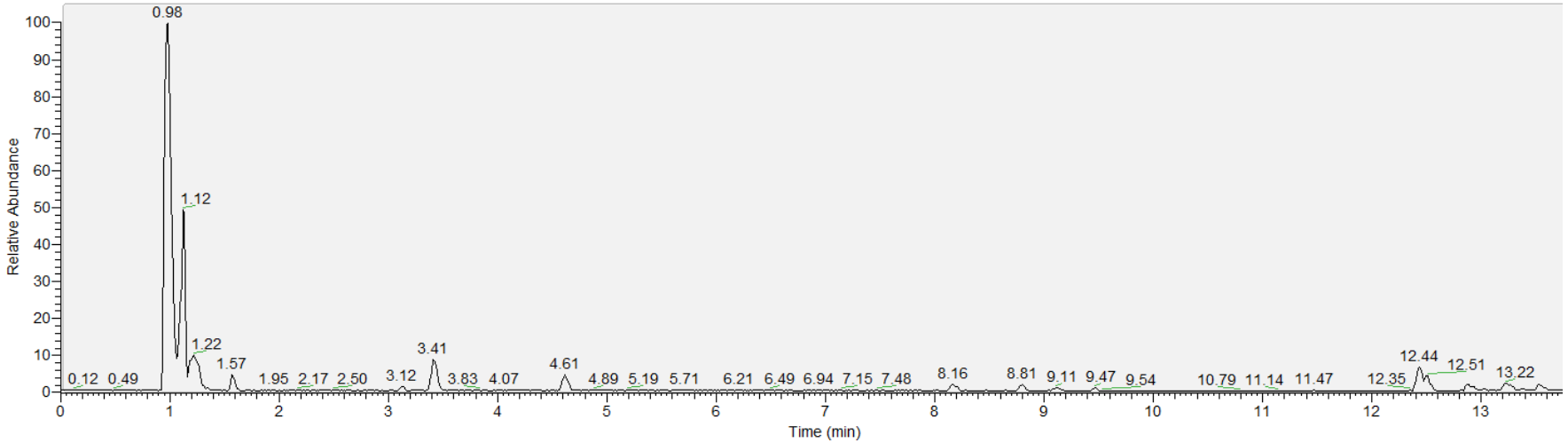

Figure 2.6: Base peak chromatogram for the longer length gradient. BPC shows most abundant peaks from ambient sample from ClearfLo campaign.

In the short method, a defined peak can be seen at 5.72 mins, this was used as a reference point for the longer method alongside two broader peaks at 7.44 and 8.52 mins respectively. This was the desired region to expand. In the longer method, the peak at 5.72 mins stayed very similar, however the two broader peaks eluted much later, at 9.71 and 10.52 mins respectively showing they are retained much better retained; increasing their retention factor (k, the equation is given in equation 2.1) from 9.48 and 11.00 to 12.68 and 13.82 respectively.

$$k = \frac{Tr - T0}{T0} \qquad \text{(Eq 2.1)}$$

Where k is the retention factor, Tr is the retention time of analyte and T0 is the dead time (also known as solvent front, or retention time for unretained analyte).

A very broad peak eluting at 8.52 mins in the short method, shows shouldering. This occurs when the amount of a sample injected is too high or when two peaks co-elute and are not resolved; in this case, the levels injected is not enough to cause the shouldering effect, this suggests a peak is being eclipsed. In the longer method, an additional peak can be seen, being removed from the broad peak. To work out how well the compounds are separated and to show the effectiveness of the chromatography the separation factor (α, the equation is given in equation 2.2) can be calculated. In the short method, the separation factor for the peak and its shoulder was 1.03 and in the longer method the broad peak and the newly separated peak have a separation factor of 1.06.

Another peak appears to be coming from the broad peak, suggesting more analytes are being eclipsed or merged into this broad peak, highlighting the need and subsequent utility for chromatography in this sort of analysis.

$$\propto = \frac{Tr2 - T0}{Tr1 - T0} \qquad \text{(Eq 2.2)}$$

Where α is the separation factor, Tr1 and Tr2 are the retention times of two analytes analyte and T0 is the dead time (also known as solvent front, or retention time for unretained analyte).

To assess the resolution of the longer method, compared to the shorter method, two clear, adjacent peaks with no shouldering or co-eluting peaks to affect the width of the peak should be analysed for their resolution (i.e. how well the peaks are separated based on their distance apart and widths. In this case, two peaks at RT 3.59 and 3.92 mins respectively in the short method and the same two peaks at 3.60 and 3.91 in the longer method. The equation for resolution, R, is given in equation 2.3. Calculating the resolutions for these two peaks in the shorter method is shown in equation 2.4 and for the longer method in equation 2.5. The resolution for the shorter method is 1.78 and the longer method is 2.30, showing an increase in separation for these two peaks despite their retention times appearing closer together.

$$R = 2\frac{Tr2 - Tr1}{Wb1 + Wb2} \quad \text{(Eq 2.3)}$$

Where R is the resolution, Tr1 and Tr2 are the retention times of two analytes analyte and Wb1 and Wb2 are the peak widths at the base of the peak for two analytes.

$$R = 2\frac{3.92 - 3.59}{(4.00 - 3.81) + (3.63 - 3.45)} = 1.78 \quad \text{(Eq 2.4)}$$

$$R = 2\frac{3.91 - 3.60}{(4.00 - 3.80) + (3.65 - 3.58)} = 2.30 \quad \text{(Eq 2.5)}$$

The gradient could be extended further to improve resolution, however it was decided that the need for a high-throughput sample was significant enough to limit the run time for each sample and as the cycle time was approximately 20 minutes from injection to injection, this was deemed a reasonable time to achieve a compromise between resolution and high throughput analysis. Fortunately, with the employment of high-resolution mass-spectrometry, this can be used as a pseudo-chromatographic technique. Using lower resolution (for example unit resolution in a multiple reaction monitoring, MRM, method) such as those used in drug detection would mean the chromatogram would show numerous peaks having similar masses and potentially lead to co-eluting compounds. Much higher mass resolution means that only compounds with that exact mass (and hence formula) are shown in the chromatogram and can

be individually identified even if possessing the same retention time as the EIC mass window can be made much narrower.

Due to the closeness of the very polar groups and organosulfates, a two-step gradient was considered with a much longer gradient in the initial minutes of the run to encourage the retention of the very polar compounds. However, this was decided against; in this case, as the number of compounds within the sample is so large, that finding the appropriate time to change the gradient mid run would be difficult, as any compounds which elutes at the time of the change would experience peak broadening across the two gradients and as such would result in very poor resolution leading to poor quantification from integration of this peak. Additionally, the issue of extending the cycle time would also present itself, if creating a shallower gradient initially, this would increase the cycle time, unless the 2nd gradient were much steeper and thus compromise the resolution of the species within that part of the run. An interesting observation between the two previous campaign samples is, as previously mentioned, the increased levels of very polar species in the ClearfLo samples and the increased non-polar species in the EROS sample. The composition of aerosol at the EROS site suggests much more non-polar, less functionalised species; indicative of a fresher, more primary emission with little aging of the air-mass.

This clearly demonstrates the different chemical composition between the two cities (Birmingham and London, respectively) and highlights the need for city specific, or at least region specific, understanding of chemical composition and by extension the risks associated from this aerosol.

2.2.3.2 Tandem MS method development

To get the best response from our analysis and see more of the less abundant species within the run, we optimised the mass spectrometer parameters, post UPLC. As discussed above, this is a fine balance, optimisation parameters are not universal; for example, collision energies favoured by some compounds may not be favoured by others and indeed, using tandem mass spectrometry, different transitions favour different collision energies. As a result of this, the parameters used by the engineer who initially set up and optimised the orbitrap for lab use were employed. These default parameters may not be optimal, however finding the optimum settings for each species within the library of compounds would take months and not be a productive use of time. Consequently, the default parameters were considered sufficient for the purposes of this analysis. One of the settings which could be optimised without too much time investment and provide some optimisation was the Heated Electro-Spray Ionisation (HESI, also known as ESI).

These settings are usually constant for all scans and transitions analysed by the MS and thus would only need to optimised once for the whole analysis. The HESI parameters were optimised to improve sensitivity of the method to the standards used in calibration mix described above. The five compounds (2-hydroxyhexanoic acid, cis-pinonic acid, adipic acid, 4-methoxybenzoic acid and 2,6-dimethyl-4-nitrophenol), each at a concentration of 1 ppm, were continuously directly injected at a rate of 300 μL m^{-3}. This rate of injection was chosen to best replicate the conditions of a run of ambient samples. The following parameters of the HESI and inlet were varied to determine how each affected the ion signal for each of the compounds: auxiliary flow rate, sweep

flow rate, S-lens angle and auxiliary gas temperature. The response from the infusion across each iteration of the settings can be seen in figure 2.7 as a plot of ion signal from each of the compounds against scan number.

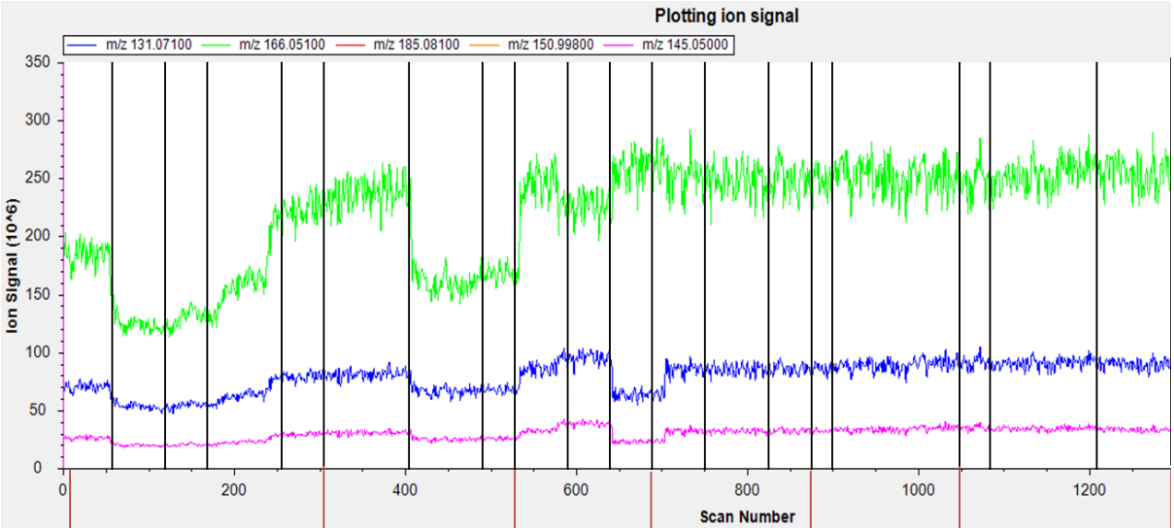

Figure 2.7: Ion signal for three (pictured) of the five standard chemicals from direct injection. Each vertical, black line represents a change in ESI settings to induce a change in the ion signal. Each vertical red line separation shows a different setting change, the settings varied are as follows: Auxiliary flow rate, Sweep flow rate, S Lens Angle, Auxiliary gas temperature, Capillary temperature.

The optimal conditions were determined as follows: the electrospray voltage was 4.00 kV, with capillary and auxiliary gas temperatures of 320 °C. The scan range was set between *m/z* 50-750 and the resolution was 70,000 at *m/z* 200. The sheath gas flow rate and auxiliary gas flow rates were 70 and 3 arbitrary units respectively. However, it can be seen from figure 2.7 that this

optimisation did not make significant improvements to the ion signal from the standards, the signal was relatively stable. Regardless, the optimised settings were used from this point forward; as minor improvements were seen in the optimal settings and gave an increased likelihood that less abundant species would be observed in the run, or that signal of all species would be higher and therefore easier to quantify against the background levels (increasing the signal to noise ratio).

2.2.3.3 Final UPLC-MS^2 method

After investigation and optimisation, a final satisfactory method was achieved and in the interest of time, this was not revisited or revised afterwards as the mass spectral library would be tied to set retention times for compounds using that specific method. Any change to the UPLC-MS method would result in rending the library moot as the reference retention times would no longer fit with any new analytical methods. The extracted samples and standard mixtures would be analysed using the following ultra-performance liquid chromatography tandem mass spectrometry (UPLC-MS^2), with an Ultimate 3000 (Thermo Scientific, USA) system attached to a Q-Exactive Orbitrap MS (Thermo Fisher scientific, USA) with a heated electro-spray ionisation source (HESI). The UPLC method employed a reversed-phase 5 µm 4.6 x 100 mm, Accucore column (Thermo scientific, UK) held at 40 ^{o}C.

The mobile phase A was HPLC grade water acidified with formic acid (100:0.1 v/v) and mobile phase B was 100% MeOH (Fisher Chemical, USA). The flow rate for the blended mobile phases was 300 µL min^{-1}. A summary of the

gradient was an isobaric hold at 90% A and 10% B between 0 and 1 minute. Between 1 and 10 minutes, the gradient changed to a linear increase to 10% A and 90% B. This was isobaric between 10 and 12 minutes. Then, between 12 and 12.4 minutes the gradient was lowered linearly to 90% A and 10% B. Finally, this gradient was at an isobaric hold until the end of the separation (16 minutes). Including washes and autosampler manoeuvres, the cycle time from one injection to the next was approximately 20 minutes.

A linear gradient was chosen over a multi-gradient as compounds are eluted across the whole gradient and changing the gradient as a compound is eluting would risk changing the shape of the peak and reducing the resolution. The injection volume was 2 µL. The mass spectrometer was operated in negative mode using full scan MS^2. The electrospray voltage was 4.00 kV, with capillary and auxiliary gas temperatures of 320 ^oC. The scan range was set between *m/z* 50-750 and the resolution was 70,000 at *m/z* 200. The sheath gas flow rate and auxiliary gas flow rates were 70 and 3 arbitrary units respectively. MS^2 was performed in data dependant top 10 method (DD-MS^2 TOPN) mode with a MS^2 resolution of 17,500. The normalised collision energy (NCE) of HCD collision cell was 35 eV with an isolation window of 4.0 *m/z*.

2.2.4 Method validation

The validation of the method was performed during the method development. Each calibration sample contained a mix of standard compounds to produce multiple curve of standard compounds, with concentrations ranging between

50 and 1 ppb, these calibration curves were run in triplicate over the course of a batch (described in more detail in section 4). In supplementary figure 1, example calibration curves are shown for 2 compounds (for 2,4-dinitrophenol and 2-nitro-1-naphthol) within the mixed calibration sample over the course of four days. The calibrations showed good linearity over the calibration range (1 – 100 ppb) with almost all $R^2 > 0.9$ (compounds with lower ionisation efficiency show lower linearity). The reproducibility of the experiment was investigated by analysing a mixture of six of the 20 standard compounds, 6 times at a concentration of 100 ppb. The average standard deviation of the six compounds at 10.6% (see supplementary table s1).

2.3 Analysis method development

2.3.1 Construction of mass spectral database.

When developing a database to use with targeted screening, the source of these compounds must be considered. As previously discussed, a bespoke database is a better option than pre-existing databases, as currently available databases are limited in their scope and often not applicable to the field of atmospheric chemistry. The starting point for the development of the database used in this work was observed chemical species in literature typically using mass spectrometric analysis to best identify m/z and elemental formulae. When populating the database manually, issues can arise from inherent bias from the researcher. An example of this bias is the availability of information; not all research is open-access and as such, research held behind a pay-wall

was not included in the database. Additionally, specific species analysis tend to be on a smaller scale, only analysing dozens of species per paper, typically focusing on well-studied tracer compounds for better comparison of sources and compositions of air masses in different location (for example, much work has been done on the biogenic tracer group, monoterpenes and as such monoterpene-derived species were more readily available in comparison to other groups). In chapter 1, it was mentioned that the comprehensive analysis of the composition of species specific to a particular source is limited in nature but offers a very solid avenue for populating a database with tracers.

In the case of this work, literature was the first calling point for database population. Literature was sourced based on the paper's inclusion of *m/z* or elemental formulae for tracers and the abundance thereof, this decision was made to ease the inclusion into the compound list function in TraceFinder™ which favours *m/z*, elemental formulae and fragment data to identify species.

The literature was sourced using the google scholar and web of science search functions, using keywords and topic filters; specifically focusing on organosulfates, tracers (both for biogenic and anthropogenic) and commonly occurring species which occur in urban environments as our reference samples and expected sampling sites were to be urban in nature. Again, this can lead to some inherent bias as the literature search is limited to those which appear on those two search engines using key search terms inputted by the user. In addition, due to the time constraints of building the database and progressing to the next steps in the research, the papers which were included were the ones found during the time to populate the database (around 2 to 3 months of populating, identifying species and optimising the search function). As such,

more useful literature or literature with broader scope of tracers could have been released after the data gathering phase and database building phase of the work, as such this is a potential place for improvement going forward.

With a list of elemental formula, this was entered into a brief screening method. These high-resolution masses were screened for in the reference samples from EROS and ClearfLo campaigns, but also screened for known standards of individual compounds. This was done to assess the peak spread of compounds and determine the retention time (or retention times, if multiple isomers exist of the same elemental formula) which occur for the elemental formulae of the species. This was done to pair elemental formula with RT and create a species in the database, if multiple peaks occurred, this would be considered as isomers occurring and each isomer included as a different entry into the database. Due to the variation of peak size which occurs reasonable caveats were taken into account when deciding if a peak was 'real' or not. This would include a significant peak of at least 3:1 signal to noise ratio for that transition. This could potentially lead to an additional level of bias to the selection; however the screening method was also in Tracefinder™ and as such, parameters to selectively identify peaks and avoid false-positives were implemented using that software. This removes some of the user-defined control and therefore removes some of the potential bias which can occur from manual peak selection from an analyst looking at a TIC or BPC chromatogram.

Species were unincluded from the database if the peak was not observed in the extracted EROS and ClearfLo samples or if they appeared to be duplicates of another species (for example if the RT was similar to another compound, within a margin of error for the RT or if the functional groups in once group

could be misread as another combination of functional groups). Within this, there is a risk that a real compound would be dropped from the database if it is too similar to another species; however, the removal of these compounds would reduce the likelihood of false positives and identifying a compound twice, mistaking it for two different compounds. The avoidance of false positives was deemed more favourable compared to having a larger database with potential duplicates.

The employment of a semi-targeted screening method to apply to the aerosol samples requires a database to compare the spectra against, due to the lack of an already existing database for compounds found within atmospheric aerosol, once was purpose built for this analysis. The mass spectral database was built using the compound database function in TraceFinder™ v.3.2 software (Thermo Fisher Scientific, USA). Extracted $PM_{2.5}$ samples, from the Elms Road Observatory Site (EROS), Birmingham, UK[74] and the Clean air for London (ClearfLo) project[134] (London, UK), were initially used to build the database by comparison to product ion mass spectra found in previous literature studies[8,25,139–144,61,87,133–138]. Mass spectra scans were scanned for the *m/z* for observed species in literature, the high resolution mass spectrometry (HRMS) allows for positive identification of the exact formulae for the species and avoided possible mis-identification of species if lower resolution *m/z* values are used (multiple possible elemental formulae can produce the same *m/z* values if only measured to 1 decimal place). Extracted ion chromatograms for each *m/z* values were used to obtain the retention time (RT) of each of the resolved isomers in chromatograms of the filter extracts, due to the multiple possible structures which can occur from the same elemental formulae (including isomers of the same compound). Each of the compounds from the

database were identified in the ambient $PM_{2.5}$ samples from the ClearfLo and EROS campaigns and SOA from chamber experiments. This was done for the verification of the compounds within the database prior to use in this analysis going forward, if the species is observed in literature and observed in samples from EROS and/or ClearfLo campaigns it can be assumed a real species and used in the database for use in analysis. A more detailed exploration of the database is discussed later. Compounds were classified by their molecular elemental formulae using the generic form: $C_cH_hO_oN_nS_s$, where c is the number of carbon atoms, h is the number of hydrogen atoms, o is the number of oxygen atoms, n is the number of nitrogen atoms and s is the number of sulfur atoms.

A major advantage of the high resolution of the MS was that it allowed identification of elemental formula with low errors and the use of a very narrow *m/z* range (0.05 Da) to plot extracted ion chromatograms. If multiple peaks were present within the specified *m/z* range, each of the distinct peaks would be noted and multiple entries for the same elemental formula would be included in the database, each with a unique tag of elemental formula and RT. Mass spectra of abundant peaks in samples from chamber generated SOA of biogenic VOCs[61] and gasoline vehicle exhaust[38] were also included in the database. In cases where more information was present, additional levels of tagging were added to the database entries, *i.e.* if the compound was identified from simulation chamber experiments of a BVOC, then it would be labelled as a biogenic secondary organic aerosol (BSOA) tracer along with the specific precursor (i.e. α-pinene BSOA). Commercially available standard compounds (including 13 CHON, 4 CHOS, 2 CHO and 1 CHONS compound) were also run using this method and the database tags updated to include compound identity. In total 646 compound mass spectra were added to the database.

2.3.2 Automated method development

The UPLC-MS data obtained was analysed using a batch method built in TraceFinder™. The mass tolerance of the method was 5 ppm, the retention time window was set to 10 seconds, to allow for a small tolerance of variability in RT, and the peak tailing factor was limited to 1.75, to avoid co-detection of closely eluting peaks. The minimum signal to noise (S/N) ratio needed for a positive qualitative result was set to 3.0 and quantification was performed if the S/N ratio was greater than 10. Peak smoothing was set to the lowest value of 1. The TraceFinder™ results were outputted as a .CSV file.

A purpose built code was written in R to extract the time code from the file outputted from TraceFinder™. Peak areas for each of the compounds identified by comparison with the database were matched to their corresponding filter sample. These were matched with their corresponding filter sampling times and the peak areas were normalised by volume of air sampled to allow comparison between filter samples, with units of peak area per volume of air (peak area m^{-3}). Where possible, the known compounds were quantified using appropriate calibration curves ran during the same sequence. For each compound in the database, common aerosol metrics were calculated automatically from the elemental formula including oxygen : hydrogen (O:C) ratios, hydrogen : carbon (H:C) ratios, double bond equivalencies (DBE) and average oxidation state of carbon (OSc, equation in supplementary information) to interpret the chemical nature of the aerosol[37,145–149].

2.4 Overview of the SAPA software functionality

A key part of this project, which underlies the work presented within this thesis, is the automated processing of the data. The purpose of the software is to make this type of analysis significantly more user friendly, lowering the bar-to-entry of data analysis and interpretation. Currently, interpretation of LC-MS/MS data is done in two ways, either by-hand peak picking or using in-built software analysis. The choice of which to use is based upon the nature of the data which is being analysed. In drug analysis, in-built quantification can be used using software such as Analyst (Sciex, USA); this is a targeted approach and therefore used only for known compounds in a specific analysis. This software allows for the accurate quantification of these known species. By developing a calibration curve from the peak area response of designated calibration standards and back calculating samples of unknown concentration based on their peak area response. This form of analysis can be completely automated and leads to very easy to use analysis of known compounds; however like any targeted screening, it is limited to only known compounds in the analysis.

The second type of data interpretation is by-hand analysis, where the chromatography is read by a trained analyst to see trends of peak areas in response to peak area of concentration to determine the location and attribute them to individual species. This form of analysis has a much higher barrier-to-entry with regard to the interpretation as subtle variation in retention times and the presence of artefacts in the chromatogram can be mis-read by untrained data analysts. This un-targeted approach allows to identification of

both known and unknown compounds. In some cases, a known compound shows clear evidence of the compound from large, well defined peaks, sometimes these are much more subtle; hence the necessity of a level of training and expertise in the technique of data interpretation.

By-hand offers much more flexible analysis, which a piece of software is unable to distinguish, such as the distinguishing of interference from actual signal from the target analyte or loss in peak shape due to issues with the chromatography. This means the by-hand approach offers a much more accurate approach to data interpretation and avoids false-positives and false-negatives which might occur from an automated approach; by-hand can also prevent mis-quantification by a more distinguishing look at potential interferences which may impact a peak.

However, this is a much slower process, with every peak needing to be carefully examine one at a time by a trained analysis and care taken to be as uniform as possible in the integration of peaks (particularly if the integration is manual also). Additionally, if the analyst is inexperienced or unfamiliar with the matrix they are working with, peaks from noise or matrix effect can be mis-interpreted as authentic signal from compound, or one peak being mis-interpreted as another (for example where two species have similar RTs and the windows of RTs for each species overlap, this could lead to an analyst assuming the peak corresponds to one species at a slightly different RT, when in reality it is caused be a different species at an unusual RT). The semi-targeted integration of the data offers a good middle ground to best suit the mostly unknown nature of aerosol composition and is described in more detail in chapter 3.

The semi-targeted analysis of compounds in the aerosol is performed by Tracefinder™ as described previously and the resulted data is exported in the universal file format of a .CSV file (comma separated value file). The software will then read this file and begin its processes, a simplified summary of the processes performed by the SAPA software is shown in figure 2.8. Upon reading the raw output of the Tracefinder™ application, the .CSV file is formatted by the SAPA into a way more easily ready by the software going forward and strips the data of any non-relevant factors which are included by Tracefinder™ (due to its additional functionalities, Tracerfinder™ will add dummy or blank data to the .CSV file as it has one output file format which is uniform for all implementations of its functions).

The software then processes all 'empty' or 'non-applicable' entries into one uniform format so the data can process the whole file without erroring or missing pieces due to a different designation terminology for essentially the same state of that piece of data. The software will then identify the heteroatom group from the elemental formula and separate out the species accordingly, calculating common metrics for aerosol characterisation such as DBE and OSc. Once these characteristics are calculated, the software will then attribute the sample number to a time period of sampling, organise them chronologically and build temporal trends for heteroatom groups and notable individual species.

For compounds which are fully identified and have a corresponding calibration curve, the software will separate these out, calculate and plot a calibration curve and use the trendline to calculate the concentrations of the known analytes. For species which do not have an exact match to a calibration

compound, the software can be told to use one or multiple calibrations to estimate a range of concentrations based on the calibration standards which best fit the unknown analyte. This is more indicative of concentration than fully quantitative and is not compulsory, instead the raw peak area can be used as a metric instead. Calibrations will also be run periodically and plotted for easy detection of drift; if the drift is insignificant, these different calibrations can be run across the whole batch to calculate a representative gradient for the whole run of 300 samples. Next the software will bring in auxiliary data available on the AIRPOLL-Beijing sharing point. This auxiliary data includes gas phase data for inorganic pollutants and AMS data (including data from sub-setted groups, such as BBOA). The software will then generate timelines for each of these factors and align the time-lines with the off-line filter samples. Where appropriate (such as when sampling rates are different) averages are taken of data points to better align the different timelines and ensure compatibility of date-stamps and comparability of data.

The software will then be able to plot metrics of co-variation between groups of species, individual species or the TOC in comparison with the auxiliary data to see if there is any significant co-variation to help identify possible links between the formation of a group of species or individual compound (or indeed the loss of a group or compound) during the presence or absence of inorganic aerosol or gas-phase pollutants. This can also be used to see co-variation of individual species with AMS factors, and lend insight into the properties of these compounds, potentially allowing them to be included as a part of that factor given significant co-variation. A caveat with this approach has been discussed previously, with the filter sampling analysing the individual species and AMS measuring bulk aerosol, any links or co-variation is not a guaranteed link.

Conversely, compounds may co-vary poorer than would be expected with AMS factors due to the sampling techniques and subsequent averaging of data. It should therefore be noted that any links drawn are drawn tentatively until further data can support such links or if that data already exists in literature.

The software will then go on to map out abundances of species with the averaged wind-speed and direction during that sampling period. At this point the software can plot wind-rose graph as seen in chapter 4. This can thus demonstrate links between direction of sourcing of air-masses with the presence or lack of species or groups of species. To compare and assess abundances of compounds, the software will calculate averages and interquartile ranges for user-defined species, these are then plotted as side by side box and whisker plots for easy comparison of mean abundances and consistency of abundances. It is recommended that species compared in this way should by fully identified and have reference standards for full quantification. This is because of the possible bias that can be introduced when comparing two species by peak area abundance alone as the two given species may exhibit differing ionisation efficiencies using a given ionisation method.

A major functionality of this software is that of the species to species co-variance analysis. This allows a user to see if two species co-vary and allow for investigation of similar sourcing or similar formation pathways, as if both species are either generated or eliminated during the same time period and at similar rates, this strongly suggests links between the two. Further investigation into their elemental formulae may uncover similar structural backbones suggesting similar sourcing or alternatively, similar functionalities,

suggesting similar formation pathways (such as a similar oxidant producing their functionalisation).

By hand, this type of analysis is very difficult as abundance plots must be generated comparing levels at the same time point and calculating a measure of how well a gradient can represent the co-variance. An example of this is the coefficient of determination (or R^2). Calculating R^2 values for a hand-full of compounds is well within a typical analyst's capabilities. However, when looking for hundreds of compounds, each compared to the other soon raises the time associated with this task (for example, 600 individual compounds compared to each other would require 359400 separate R^2 calculations.

This demonstrates the need and utility for such types of big-data analysis such as those performed by the SAPA software as manual calculation of these (even with the use of tools such as excel) would take excessively long periods of time by trained analysts and thus be very expensive in terms of man-hours. Comparatively, this can be automatically performed by the use of adaptable software such as the SAPA produced for this thesis. As a major part of this thesis, the SAPA software was produced to be adaptable for a variety of time-resolved data, not just atmospheric composition. As a result, this software could be changed to accept any compound library built in Tracefinder™, be in for water pollution analysis, bio-marker analysis or other forms of high-throughput analysis that would need a big-data solution. The raw code for the SAPA software is available through GitHub (via https://github.com/Kaitam506/WJD506-SAPA) and is open-source as part of the terms of the research. This software can be modified and improved by any user wishing to speed up their analysis.

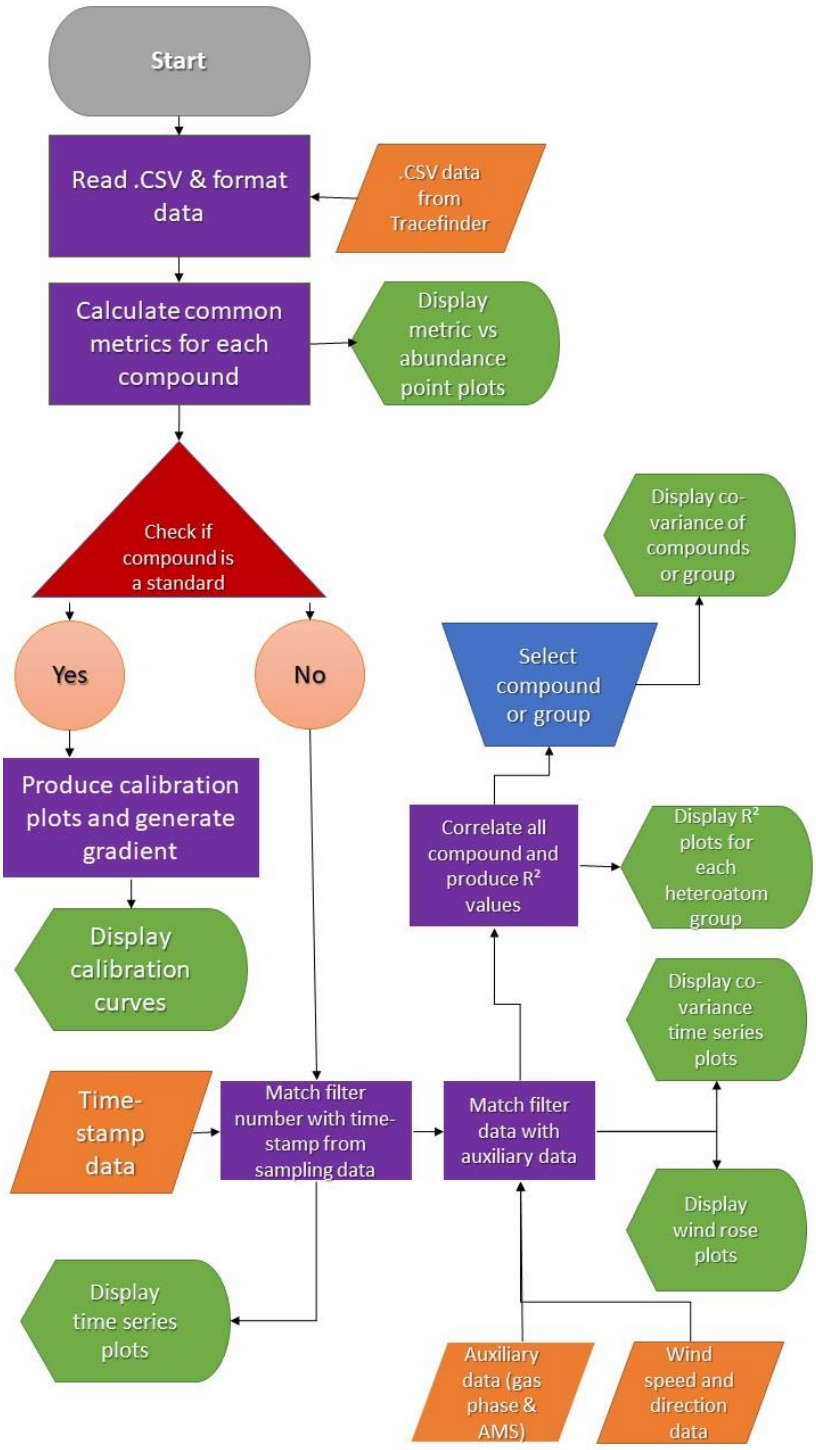

Figure 2.8: A simplified flow diagram of the base process that the SAPA program performed when processing the data read from Tracerfinder. Green elements show plots which are produced, orange shows external data read by the software, purple elements show processes and the blue element shows user-defined manual inputs.

2.5 Conclusions and impact on research going forward

A short UPLC-MS method (approximately 16 minutes per sample) was developed to allow for the rapid analysis of offline filter samples. This analytical technique was custom-built to be suitable for high-throughput analysis while maintaining sufficient resolution and separation of structural isomers for full identification of commonly observed species within organic aerosol. This method was applied to, in total, 260 filter samples from two sampling campaigns in Beijing, China to characterise the observed aerosol extract. Purpose-built, automated data processing software was employed to the mass spectra data to overcome the common limiting factors present which such large-scale off-line analysis of aerosol. The normal limiting factor for high-throughput MS analysis is the identification and integration of peaks; normally this would take many weeks if 1000s of peaks were manually integrated by a user, whereas this is reduced to a matter of days with automation.

A combination of extracts from previous analysis from urban aerosol along with observed components in literature from a variety of sampling sites and a spectrum of compositions, were used to produce a library of 646 compounds. These compounds were identified fully where applicable and given identifying tags where appropriate. The analysis of the aerosol from Beijing was performed using method, using this library as a reference in a proof of principle study to demonstrate how the use of how this automated analysis can be employed to shed light on the composition of organic aerosol and glean data from the compounds present.

A bespoke software was built to automatically process the data, manipulate the output .CSV from Tracefinder™ software into a readable file. From this the software performs calibrations, plots variations in the calibration line to detect drift in the signal across the run and use these calibrations to quantify known compounds. Finally, the software formats the data into a timeseries against averaged auxiliary data and plots comparisons between individual compounds, groups and/or auxiliary data (such as AMS or meteorological data).

After performing this analysis, it became clear that the limiting factor for this type of analysis is the extraction of the organic matter from the filters. Without the use of automation in the extraction process, the extraction of approximately 300 filter samples took multiple weeks. As such, this would need to be a consideration for such large-scale sampling going forward. The complete automation of filter sampling, extraction and extract analysis would allow for much larger scale analysis with increasing temporal resolution. Ultimately, the effectiveness of such a high-throughput automated method in the analysis of organic aerosol is shown, even when analysis large amounts of data.

Chapter 3 – Overview of ambient organic aerosol in Beijing, China

3.1 Introduction

The people's republic of China is one of the largest countries in the world, with a population of over 1.3 billion people. The rapid development in China over the past years have resulted in large growth in infrastructure and industry, leading to an increase in pollution[119]. Due to the local topography near Beijing, this results of periods of significant levels of $PM_{2.5}$[120,121,150]; Lang *et al* reported that the average $PM_{2.5}$ levels for Beijing were 106 μg m^{-3}, ten times higher than the WHO's recommended maximum safe limits[7], with maximum $PM_{2.5}$ levels within the developed city of Lazhou reported as 356.8 μg m^{-3} in 2006[34]. During these periods, visibility is vastly diminished and the risks of breathing difficulties and respiratory diseases increase[137], with haze days occurring on around 50% of days in Beijing and Shanghai[120]. It is estimated that the death toll attributed to air pollution in China is between 1-2 million[121,4].

The composition of $PM_{2.5}$ at a specific location varies depending on source and season. In Beijing, the key sources of organic $PM_{2.5}$ were determined to be mostly vehicular and industrial (attributed to 31.5% and 22.9% of $PM_{2.5}$, respectively)[21]. Similarly, in Guangdong the main sources of ambient $PM_{2.5}$ were also attributed to industrial and traffic (22.5% and 19.3% respectively)[24]; showing that the largest fraction of particulate is a consequence of large-scale industrialisation and mass adoption of consumer vehicles, markers of the rapid development. The contributions from traffic are consistent throughout the year

(24.0% of total $PM_{2.5}$ during maxima and 17.7% during minima)[24] and is expected to rise as vehicle ownership has consistently increased by 17% annually[34]. For inorganic aerosol, sulfate accounts for 18 ±10 μg m^{-3} (19.5%) and nitrate accounts for 7.7 ± 5.7 μg m^{-3} (7.3%) of the $PM_{2.5}$ in China[34].

This rapid development also leads to significant contribution from coal-combustion tracers within the PM as coal accounts for up to 70% of China's energy source[120] and China's energy consumption has increased by 120% between 2000 and 2010 and continues to rise and development continues[120]. As a response to this, the Chinese government have taken steps in order to reduce the levels of pollution in the main cities within China and minimise the population to exposure to pollutants such as SO_2, O_3 and PM up to 25%[119,121].

The work covered within this chapter will explore the overall composition of the organic aerosol observed across two sampling campaigns, winter 2016 and summer 2017. The individual species will be investigated to determine important species and groups which contribute the most to the overall organic aerosol and finally differences between the two campaigns and within the sampling periods will be explored.

3.2 Analysis of aerosol

3.2.1 Analysis overview

From the two campaigns, 260 $PM_{2.5}$ extracts from Beijing were analysed by UPLC-MS/MS. These were run in batches of 30, in the following repeating

sequence; blank, 5-point calibration of 3 mixed standard solutions (total of 20 compounds), 10 samples. The short analysis time allowed 150 samples to be run, with blanks and calibrations in less than 1 week. The variation in peak area for the calibration compounds for batches within the same day and between days was very small. The average standard deviation for the calibration curves was 4.03%, examples of the calibration curves for two compounds across all of the samples can be seen in figure 3.1. We observed no carry over in either calibration points nor from samples was found in blanks run between samples despite the high concentrations of organic material observed in the ambient samples. The high concentrations of the samples, particularly from sulfate was observed to cause a small matrix effects on the levels of organosulfates. However, upon investigation, the effect on peak area was < 5% and so was deemed insignificant compared the average variance in calibration levels.

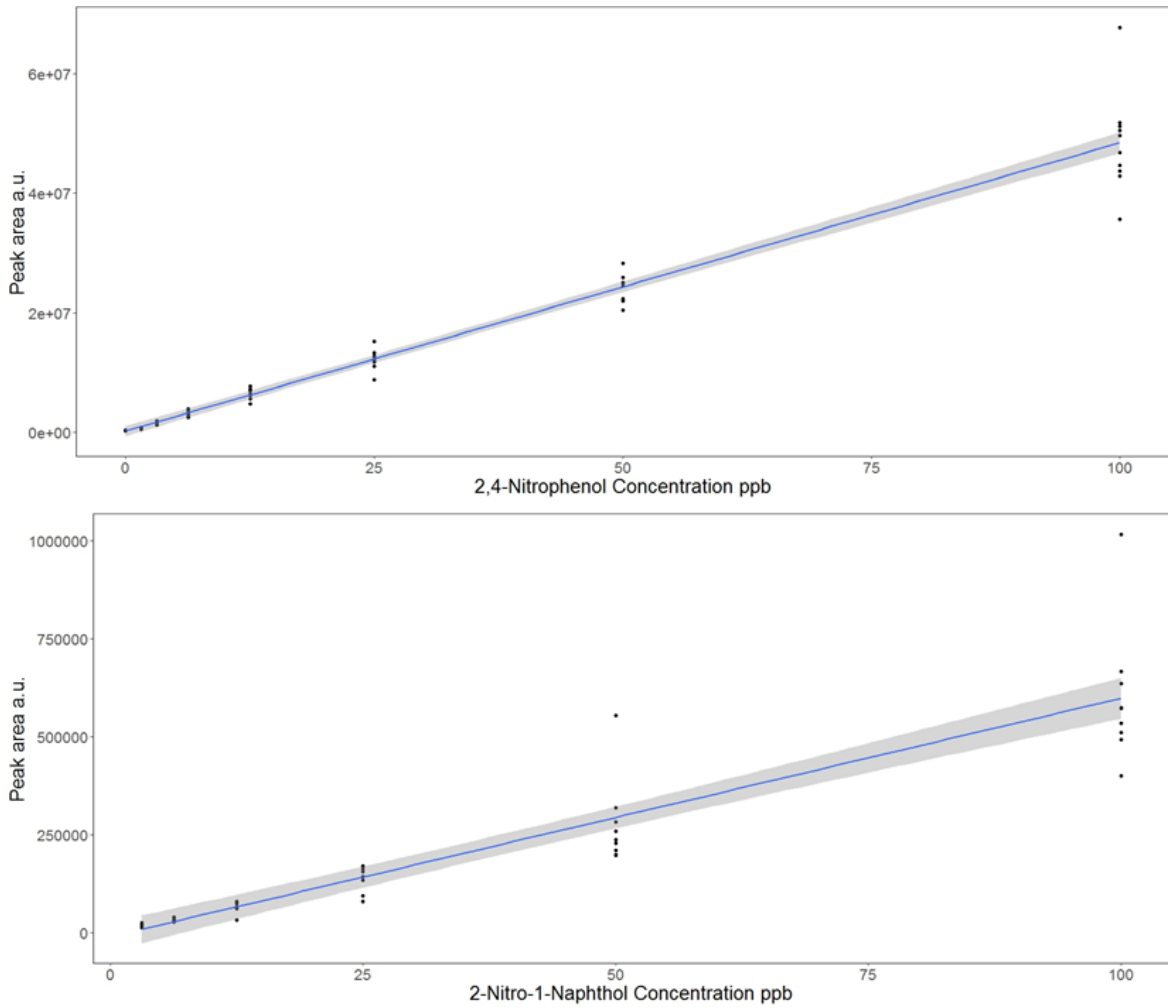

Figure 3.1: Calibration curves for two standards used to quantify the filter analysis. 95% confidence interval is shaded around the average trend line (in blue). Low variation across the analysis batches is seen in the narrow shaded area. Average relative standard deviation of gradients from this calibration is 4.03%.

3.2.2 Analysis of Beijing $PM_{2.5}$ extracts by UPLC -MS/MS

The UPLC-MS data was compared to the database and the data automatically processed using the SAPA software developed, discussed in detail in chapter 2. The $PM_{2.5}$ extracts from Beijing were extremely complex, containing many

thousands of resolved peaks. A typical base peak chromatogram (BPC) of a filter sample (20th November 2016, 12:27) from the winter Beijing campaign is shown in figure 3.2 and is dominated by a large peak at 0.69 mins associated with both inorganic sulfate (*m/z* 97) and organosulfates. Additional small peaks at 4.18 mins, 5.06 mins, 6.76 mins, were identified as nitrophenols and will be discussed below. Additionally, in the summer BPC (7th June 2017, 12:57), has a large broad peak at the start of the run (between 0.75 mins and 5 mins), viewing the mass spectra in this area shows this to largely consist of a mix of organosulfates. A mass spectrum of the region between *m/z* 251.0 – 251.1 at RT = 3.97 mins is shown in figure 3.3. Six different organosulfate ions (C_8 to C_{13}) can be observed in a small mass range window, highlighting the power of this analytical technique when applied to complex matrixes such as that found in organic aerosol.

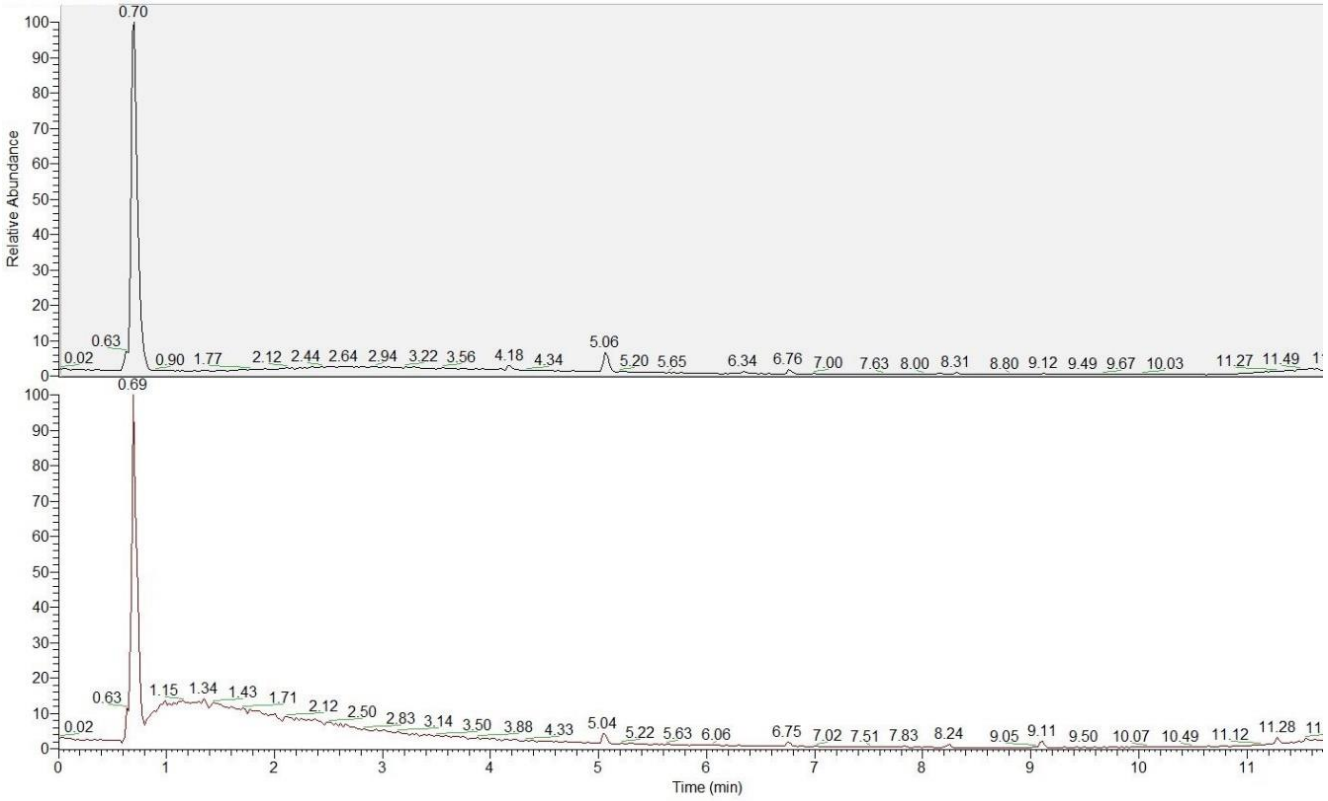

Figure 3.2: (Upper) Chromatogram of an ambient winter aerosol sample, dominated by a large peak around the dead volume at 0.69 mins. (Lower) Chromatogram of an ambient summer aerosol sample, dominated by a large peak around the dead volume at 0.69 mins and a large broad peak between 0.75 - 5 mins.

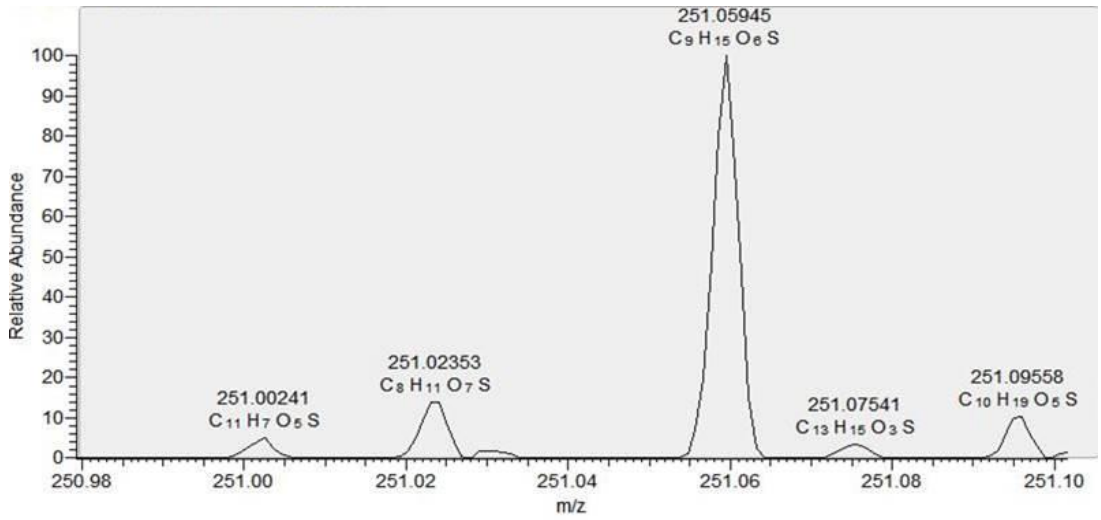

Figure 3.3: Mass spectra of m/z 251.00 – 251.10, multiple ions can be observed in this mass range, mostly consisting of organosulfates.

3.2.3 Organic aerosol overview

The Beijing winter observation period was characterised by periods of low wind speed resulting in high levels of $PM_{2.5}$ lasting around 4-5 days and visibly impairing vision, suggesting high levels of brown carbon. These low wind periods were usually followed by a sudden change in wind where a higher speed, North-Westerly wind would prevail; this resulted in a purging of the air mass build-up and much lower levels of $PM_{2.5}$ for a period around 1-4 days. The mean hourly $PM_{2.5}$ concentration was 96 μg m^{-3}, with a maximum of 438 μg m^{-3} (observed on the 4th December 2016). NO_x, CO and VOC levels were enhanced during high pollution events and ozone depletion due to titration by NO. In summer, $PM_{2.5}$ levels were lower; with an average of 62 μg m^{-3} and a maximum of 216 μg m^{-3}. Ozone levels followed a photo-chemically driven diurnal profile,

with a mean daytime average of 54 ppb, reaching a maximum of 180 ppb. This higher level of wind speeds, air mass mixing and favourable conditions for reactivity (from higher solar activity and higher ozone levels) lead to a predominantly diurnal profile from a majority of compounds.

A total of 646 compounds from the database were observed in the filter samples. 268 CHO compounds were observed ranging from C_2 to C_{22} and between 1 and 16 oxygens. 191 CHOS compounds were observed ranging from C_3 to C_{20}, 3 to 12 oxygens and 1 to 2 sulphurs. The smallest group in terms of numbers was that of the CHON group, containing 46 compounds ranging between C_6 and C_{18}. These compounds contained 1 to 9 oxygen atoms and 1 to 2 nitrogens, with the majority containing a nitro- group. Finally, 141 CHONS compounds were observed ranging between C_4 and C_{16}, containing 1 sulfur, 1 to 2 nitrogens and 6 to 12 oxygens. These compounds most commonly included nitrate and sulfate groups. A full list of molecular formulae and retention times are given in the supplementary data Table S2.

The number of compounds identified was consistent between winter and summer filter samples with 90% overlap, suggesting that the same sources of compounds are seen during both campaigns, but the source strength changes with season. Van Krevelen plots, where the O:C ratio is plotted against the H:C ratio, can be useful for visual interpretation of organic aerosol composition. The compounds observed are plotted in Van Krevelen space in figure 3.4, separated by season. The point size is scaled to the campaign average peak area per m^3 (relative to the standard point of 10,000,000 peak area m^{-3}) and coloured by heteroatom content. The standard point is an arbitrary point used as a

visualisation aid to normalise the peak areas of the heteroatom groups to one standard point.

In winter, high levels of CHON and CHOS species were observed and a pie chart showing the campaign average contribution of each heteroatom group to the total peak area per m^{-3} is shown in figure 3.5. It should be noted that this only relates to compounds that match the database and is not an accurate reflection of the organic aerosol mass as a whole. CHON compounds accounted for, on average, 80% of the observed signal for species found in the database for the winter campaign and this will be discussed in chapter 4.3.

The most abundant species were identified as octyl sulphate (based on comparison to an authentic standard) and 4 CHON species. The 4 CHON species had high double bond equivalency (DBE) values >4 and were identified as a series of nitro-aromatic compounds based on subsequent comparison with authentic standards (4-nitrophenol, 4-methyl-2-nitrophenol and 2-nitro-1-naphthol). In summer, the abundance was spread more widely among the heteroatom groups. A significant relative reduction in the abundance of CHON compounds was observed and a larger contribution to the total peak area per m^{-3} was from the CHO compounds. The increase in the amount of CHO compounds is the result of enhanced biogenic emissions of terpenoids in summer and increased photochemistry[151].

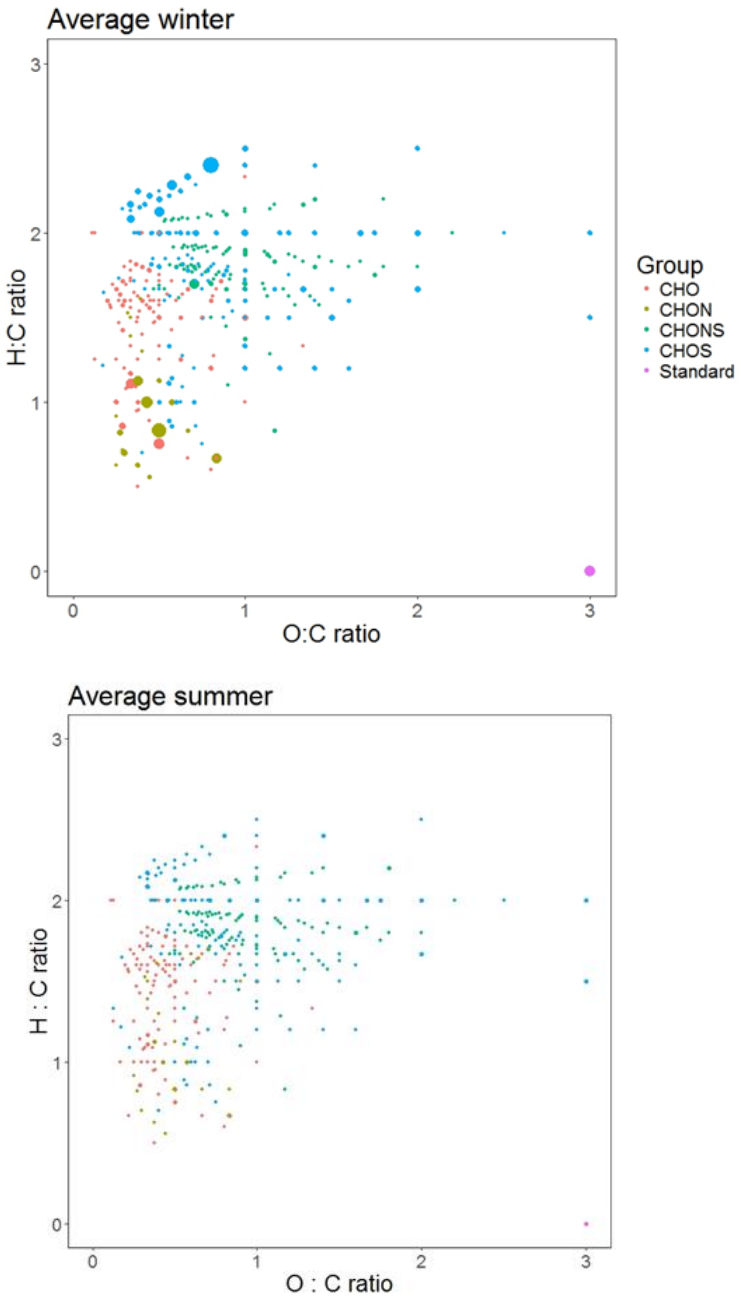

Figure 3.4: Van Krevelen plots for the averaged winter and summer filter sample respectively. Each point is colour coded by its heteroatom group and is sized proportionally by its abundance in standardised peak area (peak area m^{-3}) in the sample. In both plots, a standard point of 10,000,000 peak area m^{-3} is inserted in pink at O:C ratio = 3, H:C ratio = 0 in order to give a frame of reference when assessing point sizes, all the point sizes in each plot are normalised to their respective standard point.

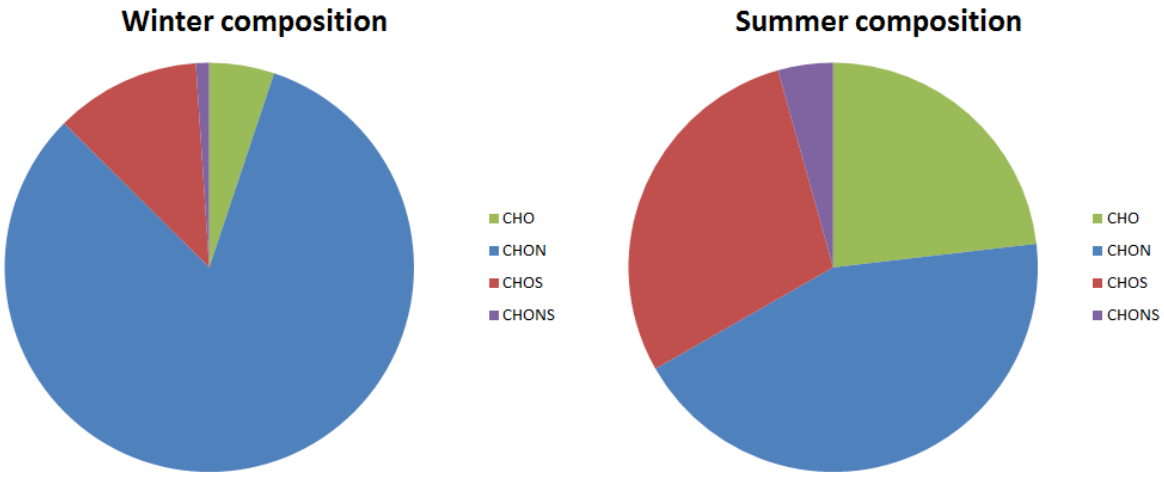

Fig. 3.5: Averaged composition of filter samples by heteroatom groups in winter and summer campaigns (left and right respectively) from the detected signal. The winter composition was: CHO – 5.1%, CHON – 82.3%, CHOS – 11.5% and CHONS – 1.0%. The summer composition was: CHO – 23.2%, CHON – 43.6%, CHOS – 28.9% and CHONS – 4.3%.

In order to determine if the number of organic tracers observed in this study is sufficient enough to accurately capture the variability of organic aerosol in Beijing, the results were compared with a co-located Aerosol Mass Spectrometer at the Beijing sampling site. The AMS organic mass concentrations were averaged to match the filter sampling times and plotted against the total peak area m^{-3} of tracers observed in each filter, shown in Figure 3.6.

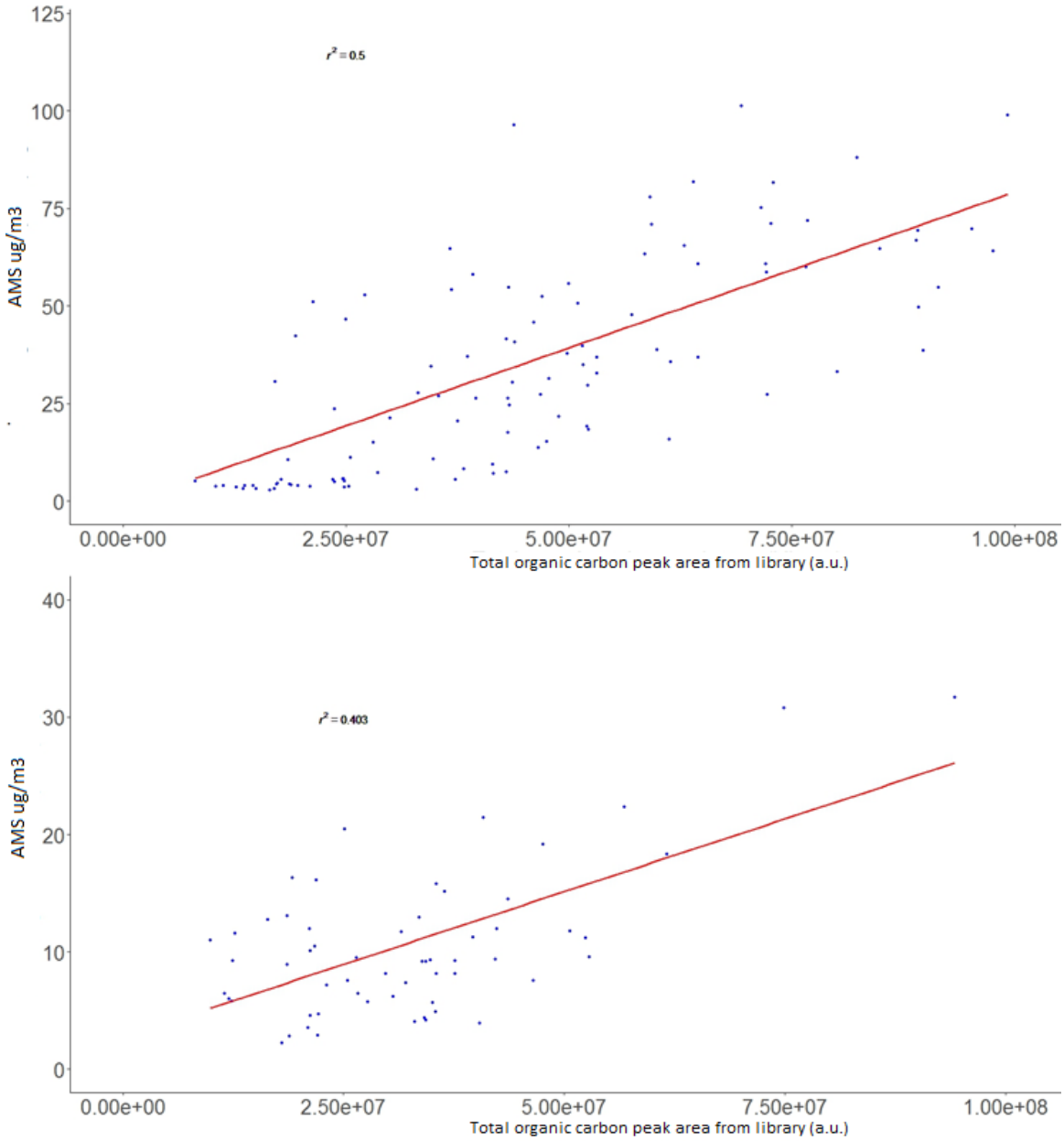

Figure 3.6: Correlation of AMS and database analysis of total organic aerosol for winter (upper) and summer (lower) campaigns. Database analysis of TOC measured in volume corrected peak area.

A moderate correlation is observed between the two sets of data; R^2 = 0.5 in winter and R^2 = 0.4 in summer. In winter, there is a cluster of data points at low concentrations, where the UPLC-MS method observed a higher relative

signal than the AMS. This is most likely an artefact caused by the high sensitivity of electrospray ionisation to CHON compounds, which are dominated by nitro-aromatics. A strong diurnal pattern in the CHON can be observed between 22nd and 27th of November. This occurred during a period of low $PM_{2.5}$ levels and due to the high sensitivity of CHON compounds, this heavily influences the total OC trend line during this period. To account for this, a method to normalise the signal for the high CHON intensity was attempted. Using the calibration curves of each of the standard compounds, an average calibration for each heteroatom group was calculated. Unfortunately, this resulted in a poorer correlation with the AMS data, primarily due to the large difference in sensitivity within each heteroatom group. For example, the sensitivity of ESI towards the nitrophenol standard compounds (N=13) varied by up to a factor of 100. Therefore no further attempt was made to calibrate compounds without standards using a group type or proxy calibration approach. However, the >600 organic tracers observed using the database method appear to capture the large-scale variability of the organic aerosol in Beijing to a suitable degree.

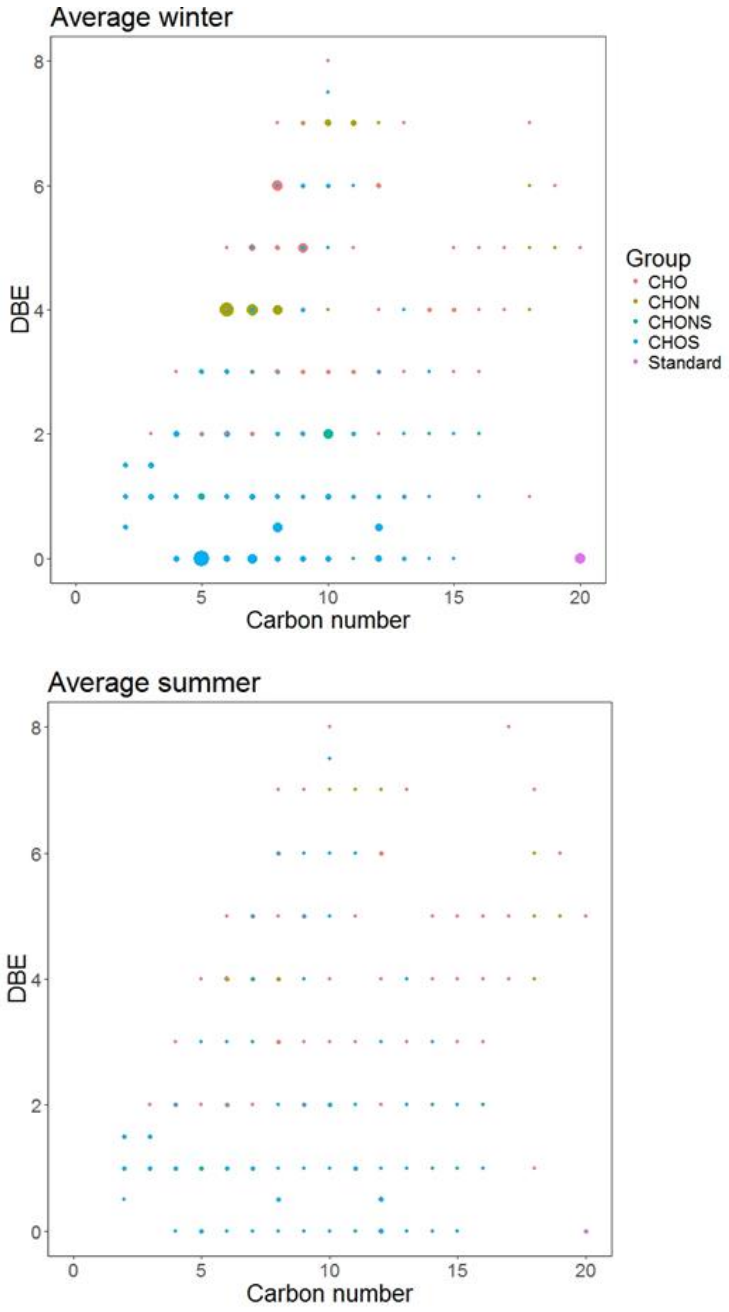

Figure 3.7: Double bond equivalencies of compounds found in the averaged filter sample for both winter (left) and summer (right) campaigns. In both plots, a standard point of 10,000,000 peak area m^{-3} is inserted in pink at carbon number = 20, DBE = 0 in order to give a frame of reference when assessing point sizes, all the point sizes in each plot are normalised to their respective standard point.

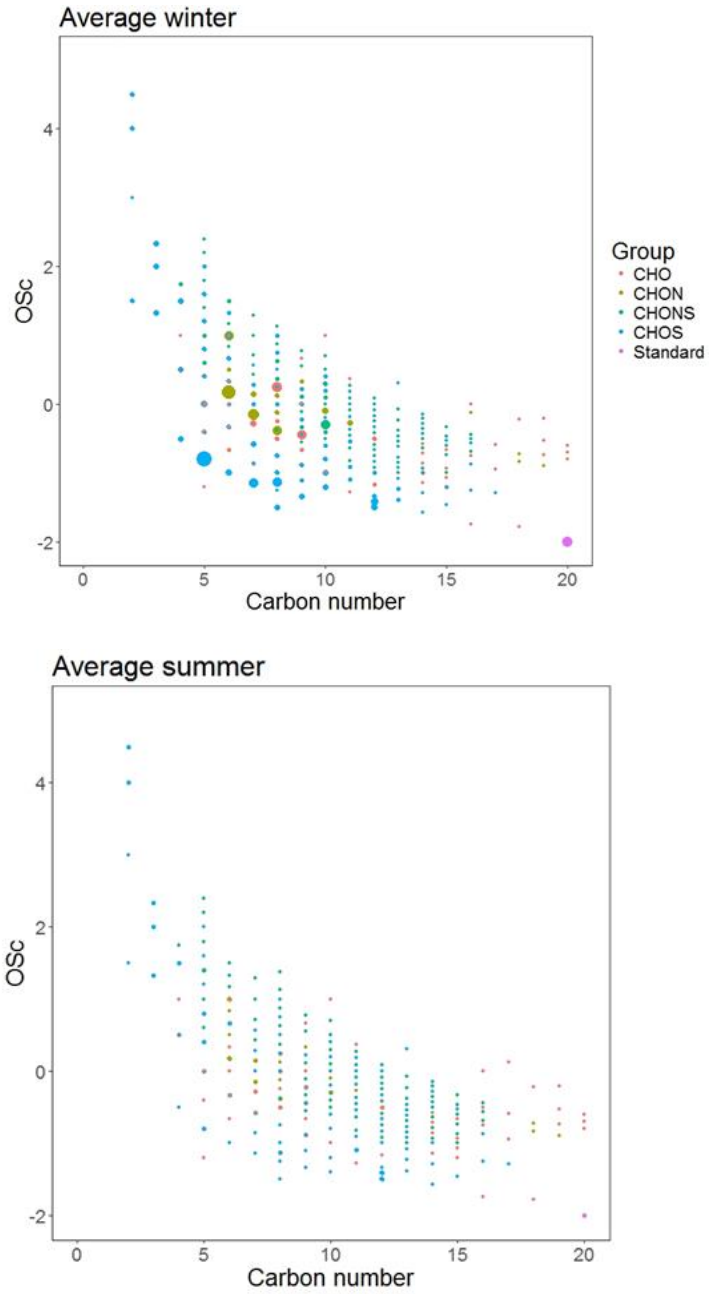

Figure 3.8: Average oxidation state of carbon for compounds found in the averaged filter sample for both winter (left) and summer (right) campaigns. In both plots, a standard point of 10,000,000 peak area m^{-3} is inserted in pink at carbon number = 20, OSc = -2 in order to give a frame of reference when assessing point sizes, all the point sizes in each plot are normalised to their respective standard point.

3.2.4 Temporal trends within the heteroatom groups

The time-series of the summed peak area per m^{-3} for each of the heteroatom groups, and the total observed peak area m^{-3} (total OC) in each filter is shown in figure 3.9. In the winter campaign, three periods of high pollution can be observed in the total OC time series: 12th to 14th November, 19th to 21st November and 3rd to 5th of December. Unfortunately, due to an electrical fault, no data was obtained between the 14th and 19th November. The high time resolution sampling has allowed the temporal evolution of the complex organic aerosol composition to be studied in detail, with strong diurnal profiles observed for many species.

Along-side the high temporal resolution samples, 24 hour samples were collected using a co-located high volume air sampler. The results obtained using this method for the 24 hour samples during the winter campaign can be found in figure 3.11. The large-scale pollution events can be observed in this data, however the loss of finer detail leads to poorly defined trends and makes any attempt to identify the sources of organic aerosol more challenging. Additionally, the poorer resolution in the filter samples may make comparisons to higher time resolution data, such as the organic aerosol factors determined by AMS using PMF, also much more difficult.

The high sensitivity and rapid analysis afforded by the UPLC-MS/MS method coupled to semi-targeted screening using the mass spectral database, provides a step change in our ability to understand the evolution of organic aerosol.

To begin to demonstrate this, the time series of the heteroatom groups will be presented to investigate the trends of different oxidised OA functionalities. This will be followed by a more detailed investigation of the trends in the evolution of the organosulfates, to show the potential of this approach.

A comparison of the total signal from the total organic carbon (TOC) observed in the off-line measurements to the non-refractory organic matter (NR-OM) observed by AMS measurements taken at the same site, are seen in figure 3.12. Very good agreement is seen between these two techniques in both summer and winter campaigns, showing that the applicability of this high-resolution offline sampling in determining TOC is compatible with AMS measurements. However, AMS measurements detect bulk aerosol, whereas this method detects individual species. This means the way TOC is calculated by the library is different to how NR-OM is calculated by AMS. Using the summed levels of all species in a library is inherently limited to the limits of the species in that library. This fact, in combination with different sampling rates (AMS sampling every 5 minutes and this method is on average hourly) will lead to variations between the two temporal trends. Despite these expected variations, the agreement between the methods is very good.

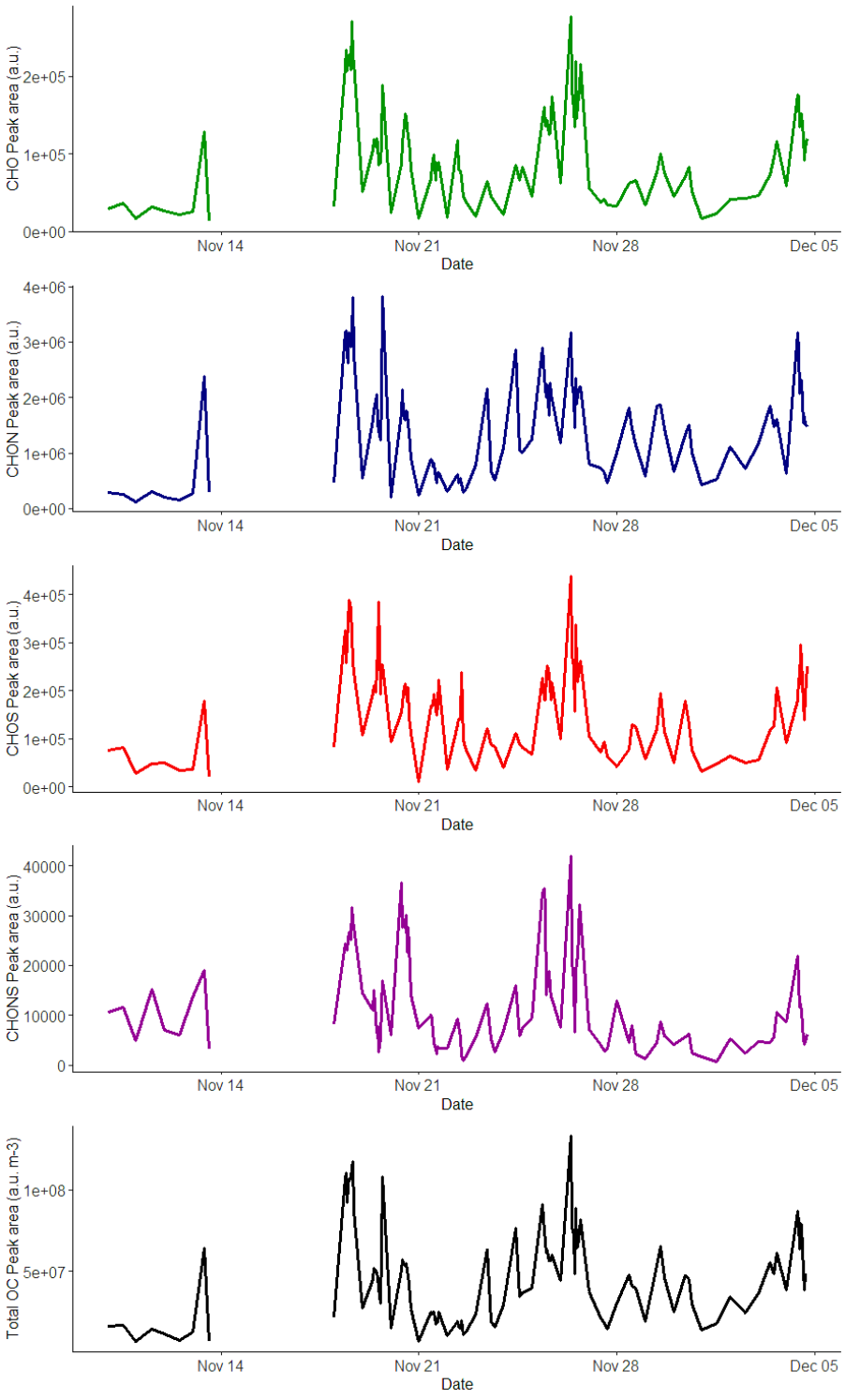

Figure 3.9a: Temporal trends of each heteroatom group and the total organic carbon (total OC) for the winter campaign. Temporal trends with fixed axis for better comparison can be found in figure 3.10)

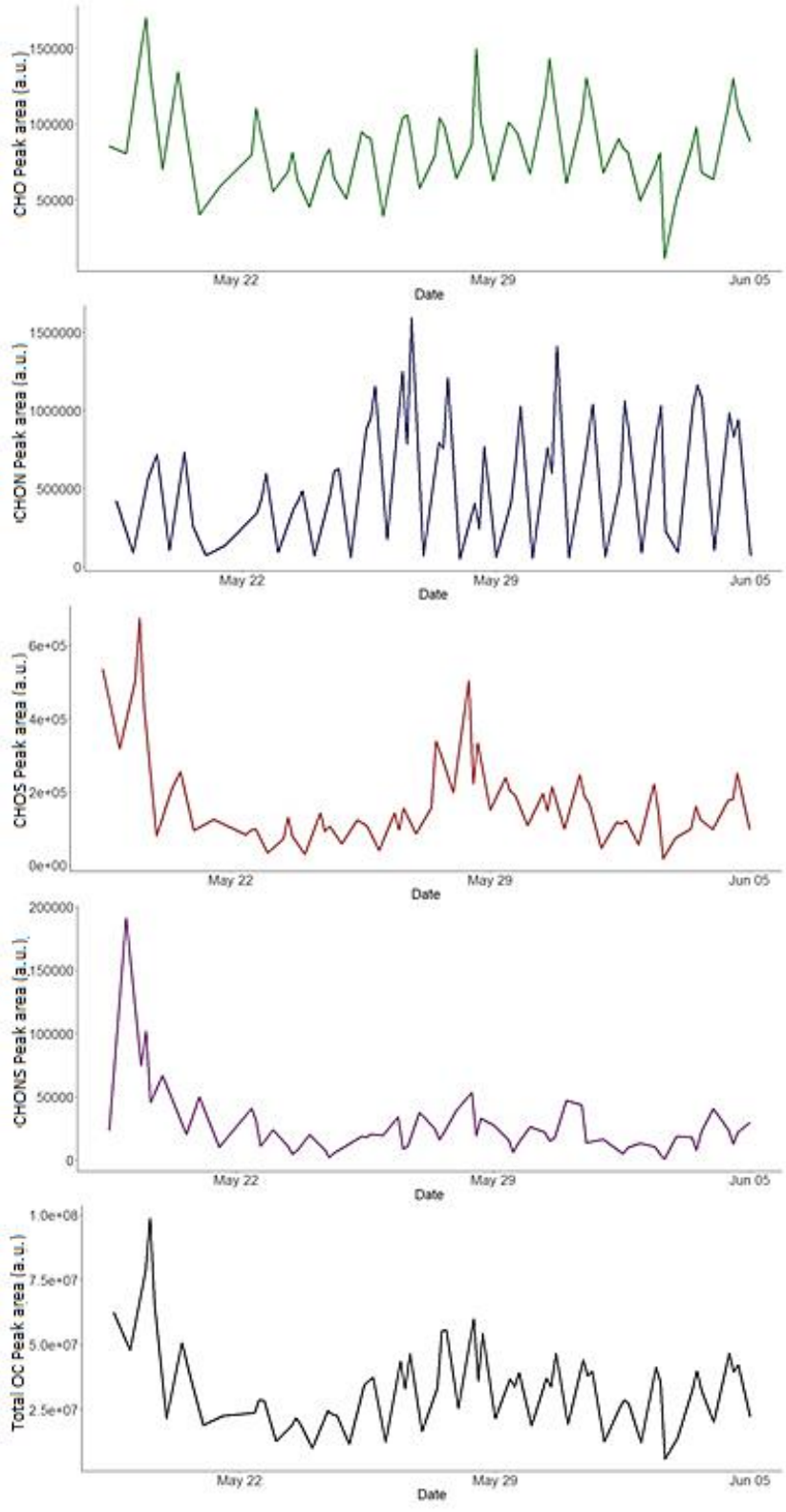

Figure 3.9b: Temporal trends of each heteroatom group and the total organic carbon (total OC) for the summer campaign. Temporal trends with fixed axis for better comparison can be found in figure 3.10)

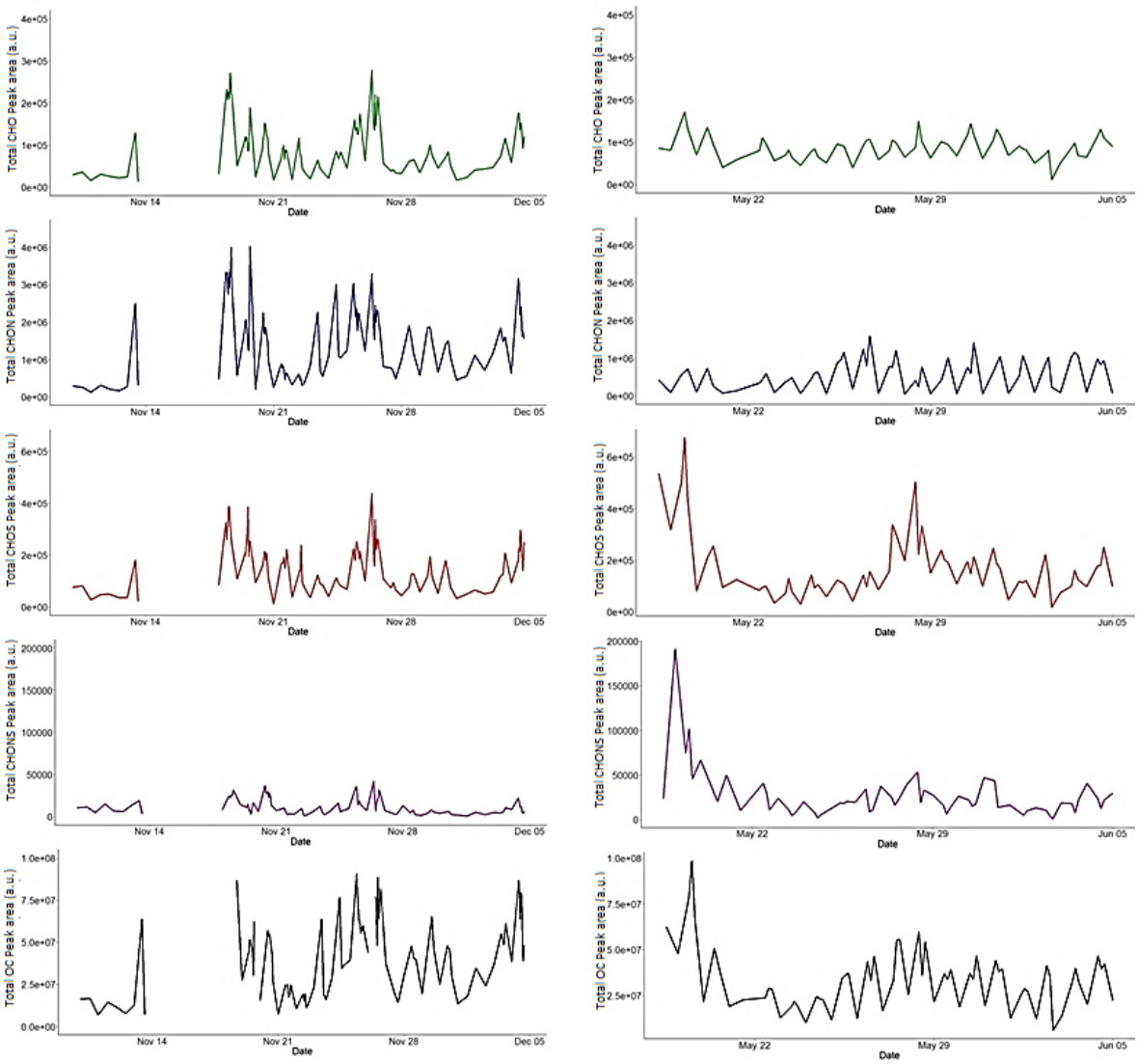

Figure 3.10: Time series for each of the campaigns (winter – left, summer – right). Each heteroatom group has been separated into different time series, with total OC for each. The scales have been fixed for both winter and summer to keep both plots for each group comparable at a glance.

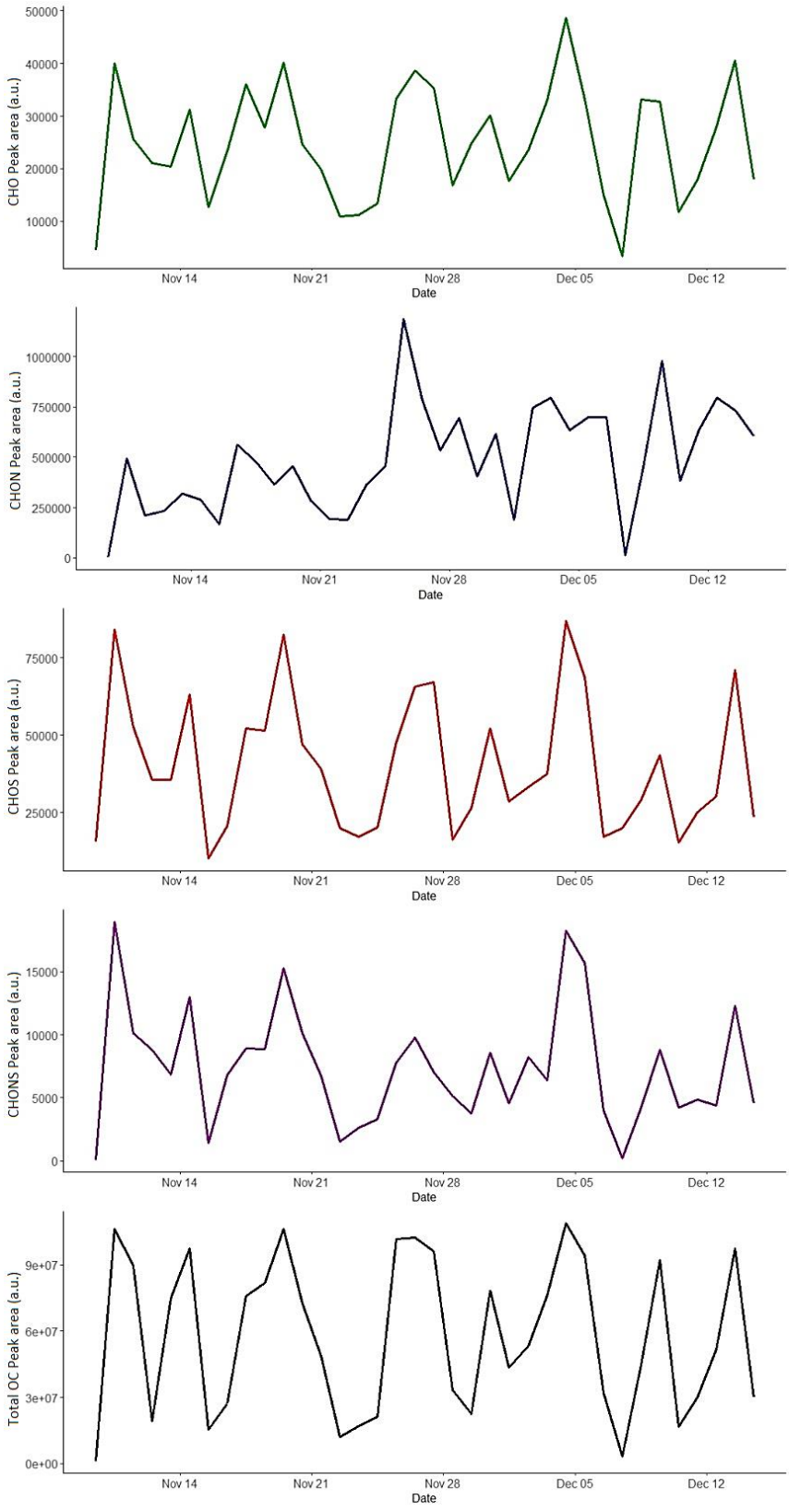

Figure 3.11: Low temporal resolution (24 hourly) time series plots for each of the heteroatom groups in the winter campaign.

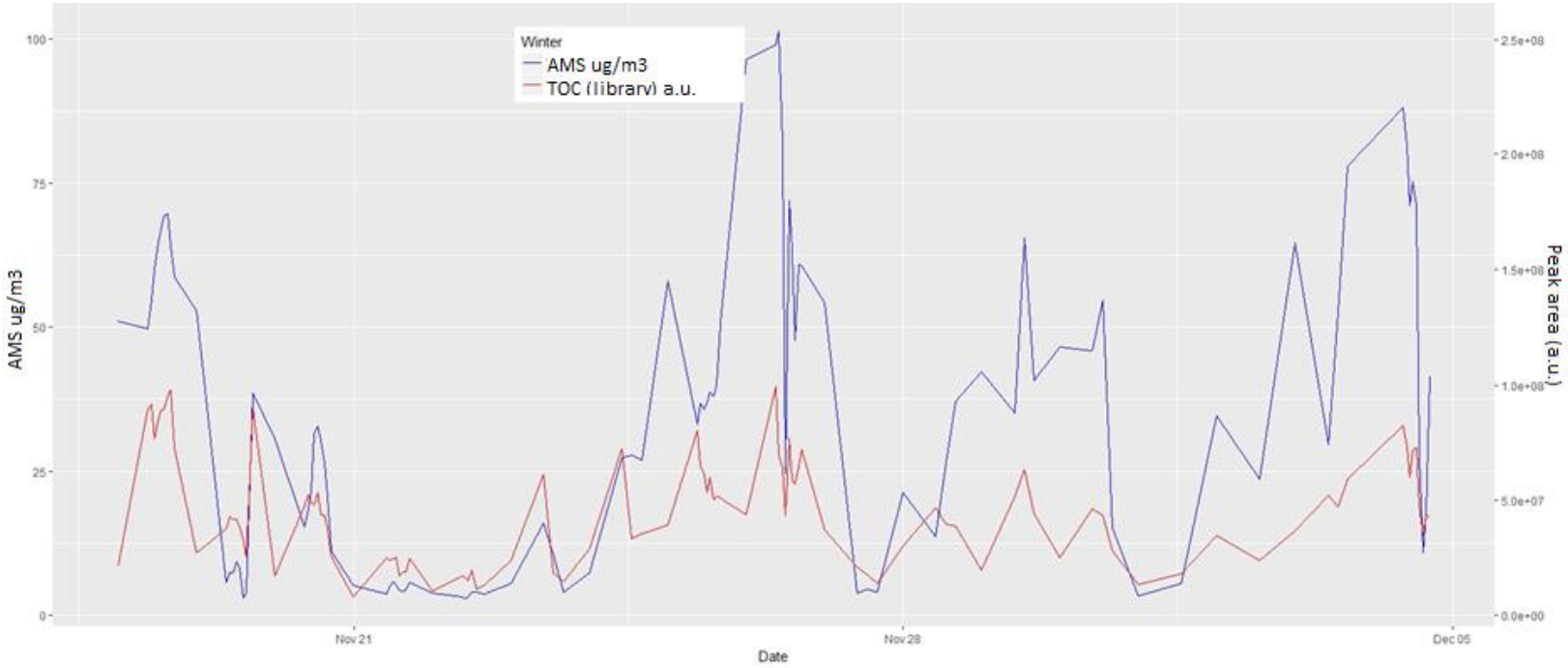

Figure 3.12a: Time series for non-refractory organic matter (NR-OM) measured by AMS and database analysis for the winter campaign. Database data is quantified as time corrected peak area.

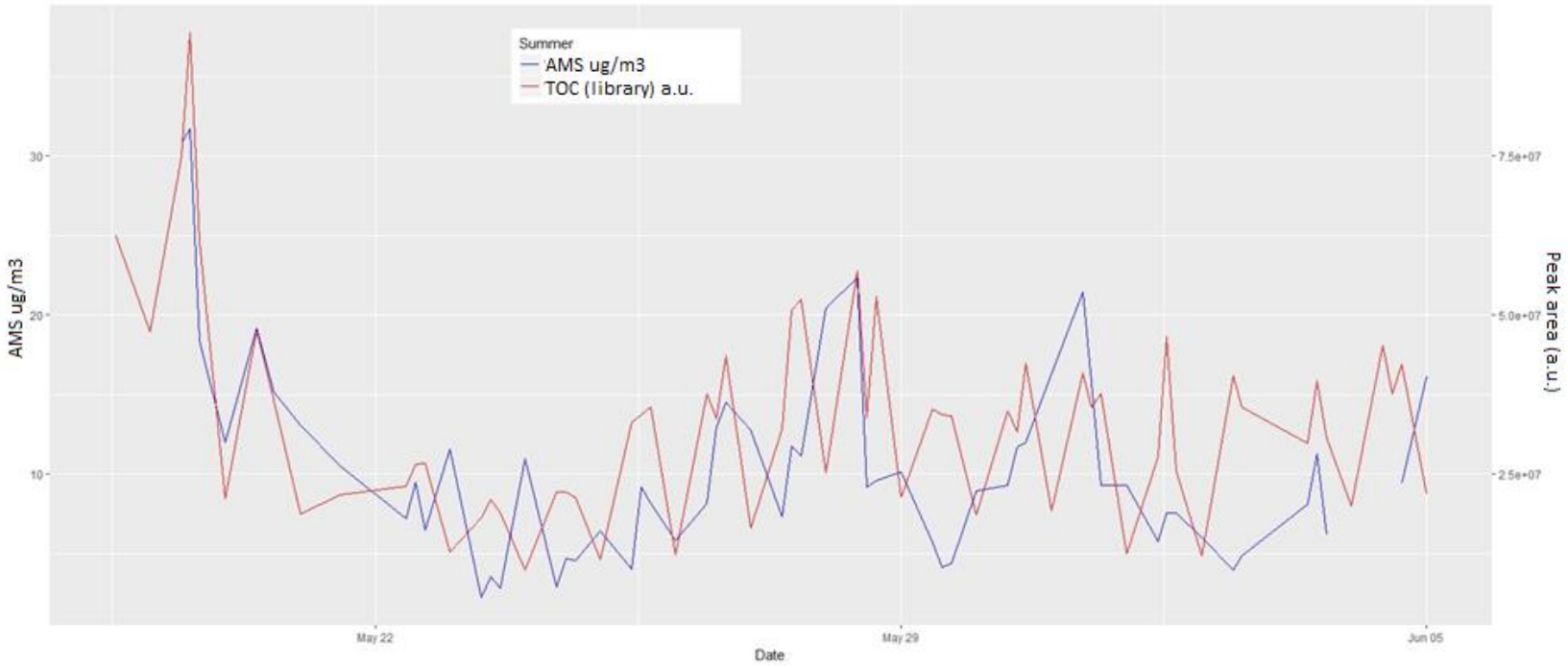

Figure 3.12b: Time series for non-refractory organic matter (NR-OM) measured by AMS and database analysis for the summer campaigns. Database data is quantified as time corrected peak area.

3.2.4.1 CHON species

A strong diurnal pattern in the CHON compounds (which are mainly nitrophenols) can be observed. In winter, the peaks occur in the early morning sample (between 08:30-11:30) with minima during the overnight samples. These peaks could be associated with a rush-hour increase in traffic emissions. An evening rush hour signal is not observed as the daytime high frequency sampling was stopped at 17:30 and so any traffic linked pollution spikes from an evening 'rush hour' would be integrated into the overnight sample. In addition, photolysis of nitrophenols would act as a sink during the day[9] contributing to the decay of nitrophenol levels; leading to a shark-fin shape in the trend.

An alternate possibility, is that the nitrophenols form as a result of reaction between oxygenated aromatics from biomass burning and NO_3[137,147,152]; however, the levels of NO_3 radicals at the surface are likely to be low during winter due to the very high levels of NO_x (up to 300 ppb). During the night, NO_3 chemistry of regional pollutants in a residual layer above the nocturnal boundary layer (NBL) could produce nitrophenolic compounds. The filters collected at the surface may be in a shallow NBL during this period and dominated by local emissions with low CHON formation. As the sun rises in the morning and the air warms, the two layers (upper residual layer and NBL) would mix and increase CHON levels at the surface during the early morning, leading to the peak during the first filter of the day as the increase in CHON levels prior to this would be averaged overnight and appear lower.

In summer, the CHON levels are much lower, with a winter: summer ratio of the mean peak area m^{-3} of 3.69. Again, a strong diurnal cycle is observed; with a peak during the early morning filters sample. However, on many days a second peak is observed at 13:30, possibly as a result of traffic emissions in the more favourable weather, or removal of CHON during the highest solar intensity in the gas phase due to photolysis or OH chemistry, leading to a dip during peak solar activity which is less significant in the early morning and early afternoon. This would lead to the appearance of two peaks. However, at the time of writing more information is required to draw fully accurate conclusions about the source of the CHON species in PM in Beijing.

3.2.4.2 CHO and CHOS species

In winter, the abundances of the CHO and CHOS groups track organic AMS closely (R^2 = 0.338). Low levels are observed during clean periods, associated with high wind speeds from the north-west, during the aforementioned purging of the air-mass. A minimum between 21st and 22nd November, when organic AMS levels were < 3 μg m^{-3}. In winter the most abundant CHOS compounds have a double bond equivalency (DBE) of 0, as shown in figure 3.8, indicating saturated aliphatic organosulfates, following the generic formula of $C_nH_{2n+2}SO_4$[153]. Such compounds could represent primary emissions, such as octyl sulfate and dodecyl sulfate, these are common surfactants used in cleaning and hygiene products which would logically be present in high quantities in highly-urbanised areas. There is also some evidence for their generation from alkane oxidation[153] and as such the source of these compounds

cannot be fully attributed, but are most probably attributed to house-hold emissions.

The CHOS compounds have minorly higher average levels in winter, with a mean average winter to summer peak area m^{-3} ratio of 1.06; in contrast, the CHO ratio was higher in summer. The averages do not, in this, case explain the full story and the higher averages for CHOS in winter may go against expectations. In figure 3.13, it can be seen that interquartile range for both CHO and CHOS compounds between the 1st and 3rd quartiles are much higher in winter, this spread is the cause for higher mean CHOS levels in winter. The median values for both CHOS and CHO compounds are higher in summer, this aligns more with the expected outcome as an increased biogenic component would contribute the summer CHO and CHOS levels.

In summer, strong diurnal profiles are seen in both the CHO and CHOS groups, peaking in the sample collected around midday. This suggests a large contribution to SOA from oxidation by hydroxyl radicals and an increased contribution from biogenic VOC oxidation. Isoprene also had a strong daytime diurnal profile, with an average mid-day mixing ratio of ~ 1 ppb and a maximum of 2.8 ppb. The CHOS time series also shows additional features, including an increased signal between 19th and 22nd May and the 27th and 29th May; which are concurrent with periods of increased SO_2 and sulphate aerosol concentrations (R^2 = 0.218), suggesting a link between the oxidation of VOCs in the presence of inorganic sulphates and the production of CHOS. In summer there is a shift in the chemical composition of the organosulfates, with an increase in abundance of compounds with higher O:C ratios (winter intensity weighted average O:C = 0.555, summer intensity weighted average O:C =

0.624) reflecting increased reaction rates under the highly oxidising polluted atmosphere in summer in Beijing. This may also reflect a change in the strength of the sources of CHOS species, with increased levels of biogenic emissions with high O:C ratios occurring during periods of high solar activity, so it is difficult to confirm which of the two is the leading cause, or if a combination is the most likely to be the underlying issue.

AMS data from the same period reflects this trend of increased oxidation in the summer sampling, the average AMS O:C ratio during the winter campaign is 0.451 and 0.597 for the summer campaign. There are discrepancies between the values for the O:C ratios of the AMS compared to this method, this is again due to the AMS measuring bulk aerosol whereas this method detects individual compounds. Nonetheless, the increase in more oxidised material during the summer period is observed in both detection methods.

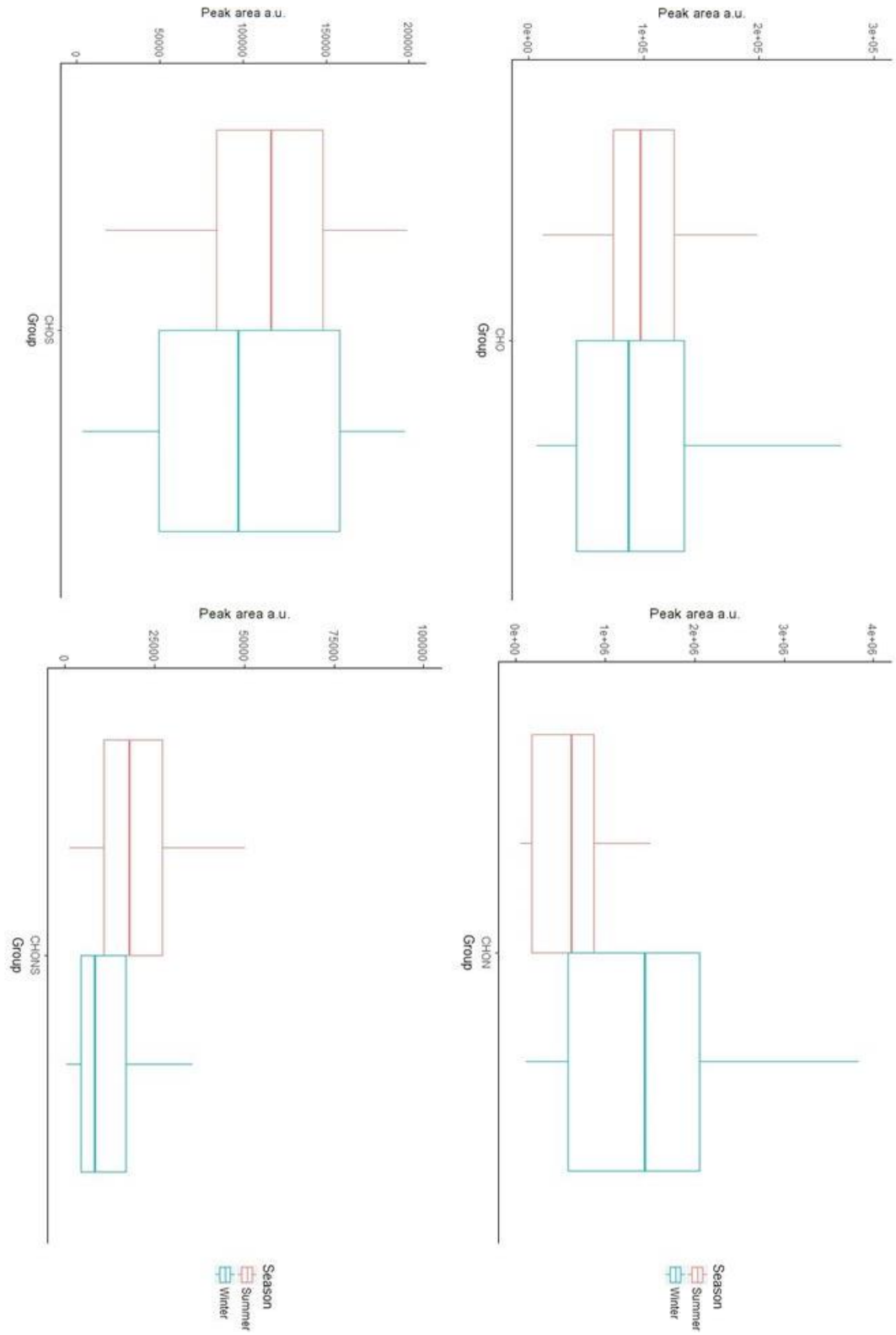

Figure 3.13: The range of levels for each of the heteroatom groups along with upper and lower quartiles for both summer (red) and winter (blue).

3.2.4.3 CHONS species

In the winter campaign CHONS compounds have 2 periods of significant abundance, 19th of November to 21st November and 25th to 26th November; with peaks occurring in the day, particularly at 09:00 and 13:00. In periods of lower abundance (between 22nd and 24th November) strong diurnal patterns are still seen, again with peak occurring around 09:00. In the winter, there is a higher number and abundance of compounds with a lower O:C ratio, suggesting the generation of these compounds is due to secondary oxidation of VOCs, as such the generation produces less oxidised species (compounds with lower O:C ratios). In summer, the CHONS compounds follow a diurnal profile that peaks during the overnight sample, with a minimum in the afternoon sample; this suggests that the limiting factor is the inorganic oxidants (such as NOx) which are more abundant overnight rather than photo-oxidation being the limiting factor. This is surprising due to the season; however, the higher average temperatures may contribute to this unusual oxidation pathways. On some days, there is also a high abundance of CHONS in the early morning sample (8:30-11:30), possibly as a result of the mixing down of aged regional aerosol.

The measured CHONS compounds were more than twice as abundant in the summer, with an average winter to summer peak area m^{-3} ratio of 0.41. There was a large spike in the CHONS on the 19th May, similar to the CHOS, although it was shorter in duration and the peak concentration occurred at 20th May rather than 19th. As with previous heteroatom groups, there was also a shift towards more oxidised species in summer, with abundances from compounds

with higher O:C ratios. The interquartile range for both summer and winter were very similar.

3.3 Challenges

In automated processing of a large number of compounds, tailoring the analysis for optimum peak detection and integration for each of the analyte targets can become very time consuming and difficult, particularly in semi-targeted approaches where unknowns are involved. During the development, the importance of resolution in the chromatography became apparent, in trying to decrease run-time and improve the usefulness as a high-throughput technique, the resolution in the chromatography was limited. As a result of this, we observed difficulty with separating compounds with very similar polarity and hence differentiating RTs between isomers in this method; this particularly was an issue with nitro-aromatics which we could fully identify and ensuring the different RTs for the isomers was imperatives.

To rectify this, two methods of peak detection were employed. Firstly, a method for general detection. This method had a larger peak width limit to allow for smaller, broader peaks to be detected and peaks which were closer to background were able to be identified. The second method was tailored for the detection of nitro-phenolic compounds. This method had a smaller peak width limit, this was done to distinguish between two peaks which otherwise could not be fully resolved. The disadvantage to this was the peak area was not fully observed, as the overlap of the peaks were not included in either integration. This leads to a small underestimation of the compounds.

During this method development recovery tests were not performed, as such potential underestimation from losses are not accounted for. Potential sources of losses include the evaporation and reconstitution of extracts along with the solubility of each compound within the solvents. In a water extraction method the majority of compounds which will be extracted will be WSOC in nature. These compounds may not be very soluble in methanol and as such, remnants may occur as a consequence of using this extraction method as very water soluble compounds may not become fully solubilised in the final water: methanol solvent and may precipitate out of solution. Since this is common extraction method, documented in literature, these potential losses were deemed acceptable as they are comparable with literature data.

3.4. Conclusions

In total, 260 filter samples were analysed by the automated method. The short UPLC-MS method (approximately 16 minutes per sample) allowed for the rapid analysis of offline filter samples, making this a suitable technique for high-throughput analysis. Normally, the limiting factor for high-throughput MS analysis is the identification and integration of peaks; with 100s of compounds in 100s of samples, such an analysis done manually by a user could take many weeks. With the automated peak detection and data processing, this entire task can be performed within a day, drastically reducing the time and associated costs with data processing. The whole analytical process from extracted samples to data output for 260 samples and calibration curves was 8 days. The limiting factor for this type of analysis is the extraction of the

organic matter from the filters. Without the use of automation in the extraction process, the extraction of 300 filter samples took multiple weeks.

This work demonstrated both a time and cost effective use of a high-throughput analytical technique for off-line aerosol measures and automated data processing software on high temporal resolution filters samples from a field campaign. The ability of the method to automatically identify the common characteristics of aerosol composition and identify individual compounds within the bulk aerosol was presented. From this, the compounds which contributed largely to the composition of the aerosol were determined from two field campaigns in Beijing, China.

Of the four major heteroatom groups, the CHOS and CHON were the most abundant, with 80% of the total response signal coming from the CHON group alone and nearly 1/3 of all compounds observed were CHOS.

The winter campaign was dominated by low DBE organosulfates and nitro-aromatics and the summer campaign less influenced by the nitro-aromatics, instead seeing an increased abundance of higher O:C compounds, particularly CHO in nature. Another interesting find was the high degree to which solar activity is a controlling factor on the composition of atmospheric aerosol in Beijing.

From this work, the need for higher temporal resolution within off-line aerosol measurement to capture the finer details in compound trends is clear and thus the limitations which come with standard 24 hour or 12 hourly sampling of ambient aerosol become apparent. The limitations which come with the quantification of unknowns by their peak area in MS analysis were highlighted;

in addition to how vast differences in sensitivity between compounds can lead to mis-quantification and the 'loss' of compounds with lower sensitivity in the method by being eclipsed by larger peaks or being indistinguishable from the background noise.

Going forward, smaller groupings of compounds will be analysed to gain more a informative look at the nature of the composition; the groupings of the compounds were determined by comparisons of trends in terms of abundance. As such, compounds which had significant co-variation were formed into groups.

Grouping compounds into sets with similar sensitivities and functionality should give more accurate results during quantification when applying the normalised calibration curves as it can be reasoned that compounds with similar functionalities will have more similar sensitivities than groups of compounds with significantly different functionality. As a result, using analogous calibration curves could be applied to groups with similar functionalities to gain semi-quantification if not full quantification.

Finally, hierarchical clustering was employed to separate out sources of CHOS and CHON compounds to demonstrate how the chemistry of organic compounds depends on the interactions of primary emissions from their various sources, the relative contributions of said sources and the formation pathways as those primary emissions interact with inorganic oxidants during their life-times. The exact composition of organic aerosol varies largely depending on the time of day and the meteorological conditions in the local area, and how the relative contribution of both local and regional sources varies across the sampling periods. This work also highlights the need for higher

resolution off-line sampling as without this, the contributions of some of the sources can be overlooked during 24-hour samples as the impacts of short-term events (such as a high transport impact from 'rush-hours') could be averaged out with diurnal variations of other sources such as solar activity.

Chapter 4 – Investigation of the nitrophenolic component of aerosol

4.1 Introduction

Nitro-aromatic compounds are a large scale atmospheric pollutant and are some of the most abundant compounds within the organic fraction of both the gaseous and particulate phases with typical ambient concentrations in the ng m^{-3} range in air[55]; these compounds have detrimental effects on human and plant health in air masses, but also have a notable detrimental effects after deposition into water[44,45] particularly affecting plant-life as they are taken up as part of the water cycle. During events with high relative humidity (RH) and high water saturation levels (such as periods of fog), nitrophenolic concentrations are particularly high due to these compounds being more favourable to partitioning into water-based systems. During these periods, people exposed to such high levels of nitrophenolics experience significant breathing difficulties and indeed, exposure has been liked to issues with the reduction in the efficiency of blood cells in carrying oxygen and also are linked to increased risks of several forms of cancer[55,154].

Nitrophenolic compounds have a high affinities with rain water and river water, with ambient concentrations in the order of μg m^{-3}, significantly higher than in the atmosphere[55] as atmospheric levels do contribute to nitrophenolic levels in the water cycle through wet deposition. Despite the lower concentrations in the atmosphere when comparted to water sources, the risks to the general public from atmospheric nitrophenols are still significant in their own right.

This affinity for aquatic systems allows nitrophenolic compounds to be easily uptaken by planet life and by microbial life, and thus their known phytotoxic effects[55] occur. The generation of nitrophenolic compounds is generally from direct traffic emissions (where levels can be as high at mg m^{-3} directly at the exhaust[10]), degradation of pesticides[155], biomass burning and notable, oxidation of directly emitted aromatics[10,16,41–43,45]. Oxidation of aromatics into nitrophenolics are typically from interactions of aromatics with OH radicals and nitrogen oxides (NOx)[55] following the generalised pathway shown in figures 4.1 and 4.2.

$$\mathrm{Ar{-}H + {}^{\bullet}NO_2\ (N_2O_4) \rightleftarrows Ar(H){-}NO_2^{\bullet}(+\ {}^{\bullet}NO_2)}$$

$$\mathrm{Ar(H){-}NO_2^{\bullet} + O_2 \rightarrow Ar{-}NO_2 + HO_2^{\bullet}}$$

Figure 4.1: Common oxidation pathways for formation of nitro-aromatic compounds. Taken from Vione 2005[156]

Different nitro-aromatics possess variable properties and have differing formation pathways[16], despite this, little work into the behaviour of structural isomers of individual nitro-aromatics has been performed. It is known that structural isomers have differing abundances within ambient air and show variation in the partitioning coefficient between gas phase and particulate phase[45,48,49]; this would posit that structural isomers would possess inherent individual behaviours in other factors.

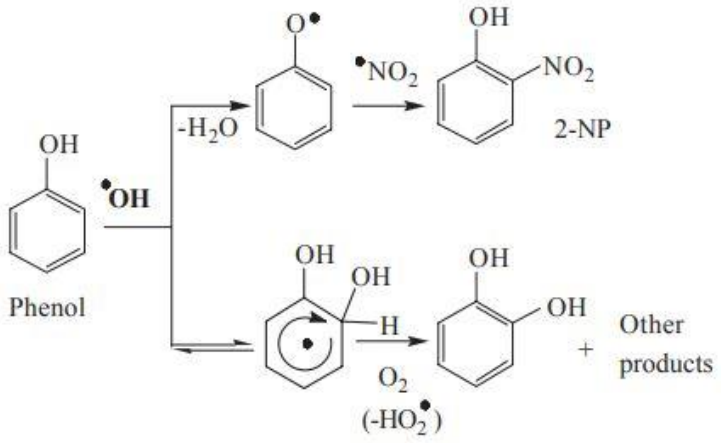

Figure 4.2: Common oxidation pathway for formation of ortho-nitro-aromatic compounds. Taken from Harrison 2005[55]

In ambient particulate samples, the abundance of para-nitro aromatics are much higher than that of ortho-nitro aromatics[48,49]; this is unexpected based on previously observed nitration of aromatics[53], where nitrification of aromatics favours the ortho positions due to the directing effects of a hydroxy group on an aromatic system with regard to addition to the aromatic ring.

Vione *et al* propose the reason for the higher levels of 4-nitrophenol compared to 2-nitrophenol found in ambient air is the susceptibility for 2-nitrophenol to undergo degradation in the gas phase[45]. In the gas phase, ortho-nitrophenols are reported to be a source of nitrous acid (also known as HONO)[9] and this degradation would account for some of this gaseous phase degradation; however the degradation of ortho-nitroaromatics and the production of HONO would not produce a complete account for the degradation of the othro-nitrophenolic species, and so other losses of ortho-nitrophenolics therefore must occur.

Interestingly, Al-Naiema and Stone observed more 4-nitrophenol in the gaseous phase compared to particulate phase during sampling in Iowa, USA[41]; this is opposed to Cecinato *et al* who observed higher abundance in the particulate phase from samples in Rome, Italy[48]. As such, a constant ratio of gaseous to particulate phase cannot be assumed when comparing observations between different sources and different sampling sites.

Seasonal variation is also observed in nitro-aromatic compounds, the reason for which is not completely understood but is best determined to be a combination of factors ultimately deciding the abundances of isomers, as discussed in chapter 1, abundances between ortho- and para- vary vastly, with para-nitrophenolics having a higher abundance in winter. Therefore, it cannot be assumed that a ratio of isomers in one location cannot be assumed to be consistent across a time-scale.

It is clear that the behaviours of nitro-aromatics vary based on their structural isomer within the atmosphere, and yet during quantification using mass spectrometry (MS) based analysis, these isomers are often grouped together; assuming uniform behaviour from isomers, despite observations that demonstrate variations in the sensitivities of nitro-aromatics in MS analysis[8,16]. This subsequently, leads to less detail in the analysis of the abundance of nitro-aromatics within the atmosphere and potentially leads to under-estimation of very harmful species in both atmospheric and water systems.

Exposure of Nitrophenolics to UV light[9] is shown to produce SOA in the presence of no other oxidants other than air (which is shown to generate OH radicals when exposed to UV light), in chamber experiments at EUPHORE atmospheric chamber (Valencia, Spain). Observing that 150 μg m^{-3} of 2-nitrophenol

produces up to 30 μg m^{-3} of aerosol and noting ortho- nitrophenols producing similar proportions of SOA relative to their concentration.

This work describes the identification of commonly occurring nitrophenolic compounds observed across samples from two field campaigns in Beijing, China. Where known standards are available, species are quantified across the campaigns and preliminary source apportionment is performed and with comparisons to common AMS factors, the contributions of the sources to each of the species are determined. The effects of stereo-isomerism within nitrophenolic compounds on the ionisation efficiency by electro-spray ionisation (ESI) techniques are investigated and the implications these variations have upon the quantification of nitro-phenolic compounds are explored. Finally, recommendations are made on how to most effectively reduce nitrophenolic levels based on the sources which contribute the most to the total nitrophenolic levels.

4.2 Methods

4.2.1 UPLC-MS method

All samples were analysed by ultra-performance liquid chromatography mass spectrometry (UPLC-MS), specifically an Ultimate 3000 (Thermo Scientific, USA) system attached to a Q-Exactive orbitrap (Thermo Fisher scientific, USA) using a heated electro-spray ionisation source (HESI) and a resolution of 70,000 at m/z 200. Using reverse phase UPLC, 5 μm 4.6 x 100 mm, C18 column at 40 ^{o}C. The mobile phase composed of HPLC grade water and 100% MeOH (both from Fisher Chemical, USA) the water was acidified with 0.1% of

formic acid. The flow rate was 300 µL min^{-1}. The mass spectrometer was operated in negative mode using full scan MS^2. The electrospray voltage was 4.00 kV, with capillary and auxiliary gas temperatures set to 320 ^oC.

4.2.2 Adapted automated peak detection for improved isomer selectivity

In the 16 minute UPLC method, the detection of certain nitrophenolic isomers was not optimally reliable when using the very sensitive peak integration parameters used on Tracerfinder software. For example the settings used when using the universal method for detecting all organic compounds within the aerosol described in chapter 2. To get a more targeted approach, the peak detection threshold was increased. This meant that less of the peak was integrated, rather than the peak area being calculated from the peak to the baseline, a small threshold must be overcome before a peak is considered 'present'. This prevented the automated software from mis-interpreting two close peaks as being the same peak (the system's automated processing can assume a level of splitting which the peak may have undergone due to high concentration of analyte). The higher threshold meant the automated processing in the Tracefinder™ step was more reliable, and whilst potentially leads to a very minor under-estimation of the peak area, the quantification was deemed more accurate without the co-detection of peaks.

4.2.3 Adapted UPLC-MS method for nitro-phenolic detection

The confirmation of the identity of the nitronaphthol compound was performed using an ACQUITY UPLC (by Waters, USA) coupled to an Obritrap Fusion MS^2 system (Thermo Fisher, USA). Using ionisation settings as stated previously. The chromatography run was lengthened to increase separation of peaks around the retention time of the compounds in the range of the nitronaphthol species and allow for better identification whilst keeping all other variable as consistent as possible. The new gradient is as follows: the gradient was held at 90% of H_2O (A) and 10% MeOH (B) between 0 and 1.5 minutes. Between 1.5 and 27 minutes, the gradient changed linearly to 10% A and 90% B; this was held between 27 and 29 minutes. Between 29 and 31 minutes the gradient was changed linearly to 90% A and 10% B; this gradient was held until the end of the separation (33 minutes). At a flow rate of 300 µL min^{-1}.

4.3 Results and Discussion

The resolution of the chromatography gave good separation of compounds allowing for identification of different nitro-phenolics within the sample. In total, 43 nitro-aromatic compounds were observed in the samples as described in chapter 3. Variation of the response peak area from identical nitrophenolics at the same concentration was observed in both individual standards of each nitrophenolic compound and a mixed standard of nitrophenolics; suggesting this was not ion suppression, caused by a mixed standard. Further enquiry into the effect of relative positions of functional groups on the response ionisation efficiency within nitro-phenolics showed that when the nitro group is ortho to

the hydroxyl group, the ionisation efficiency is diminished by orders of magnitude (100 - 1000 times lower peak area) compared to meta- or para-positioning of the nitro group relative to the hydroxyl group. This effect is constantly observed across many types of nitrophenolic compounds. The implications of this effect are significant; if some form of chromatography, or otherwise, identifying and separating isomers (a practice not always done in atmospheric analysis) is not present in analysis, this can lead to mis-quantification of nitrophenolics.

Either the ionisation efficiency is assumed to be the same as the meta-nitrophenolics and as such any (commonly occurring) ortho-nitrophenolics are significantly under-estimated. Alternatively, any mix of nitrophenolics would lose contributions from ortho-nitrophenolics as their relative peak area would be dwarfed by any presence of para- or meta- nitrophenolics, insufficient or non-existent chromatography would leave the peaks from ortho-nitrophenolics eclipsed by other peaks.

In figure 4.3, the extracted ion chromatograms (EIC) from a mixed calibration standard solution of methyl nitrophenol isomers are shown. The lower two EIC panels in figure 4.3 containing the nitrophenolics with a single nitro group in the ortho position relative to the hydroxy group, are orders of magnitude smaller when compared to those of the meta and para counter parts. These two species, are also present in the EIC from the upper left panel of figure 4.3. However, the peak sizes are so small in comparison to the isomer with the nitro group in the para position, further highlighting the sheer scale of variation of ionisation efficiency between the different isomers. This pattern for a lower ionisation efficiency was also observed in nitrophenolic compounds with no

other functionalisation, although the peak areas were not the same as their methylated counter parts.

In figure 4.4, the relationship between relative position of functional groups and the ionisation efficiencies of various nitro-containing compounds is seen. Consistently, species with nitro functionalisation with have between 10 and 1000 times lower peak area. Due to this vast variability of ionisation efficiencies, the risk of eclipsing smaller peaks without proper chromatographic separation becomes apparent.

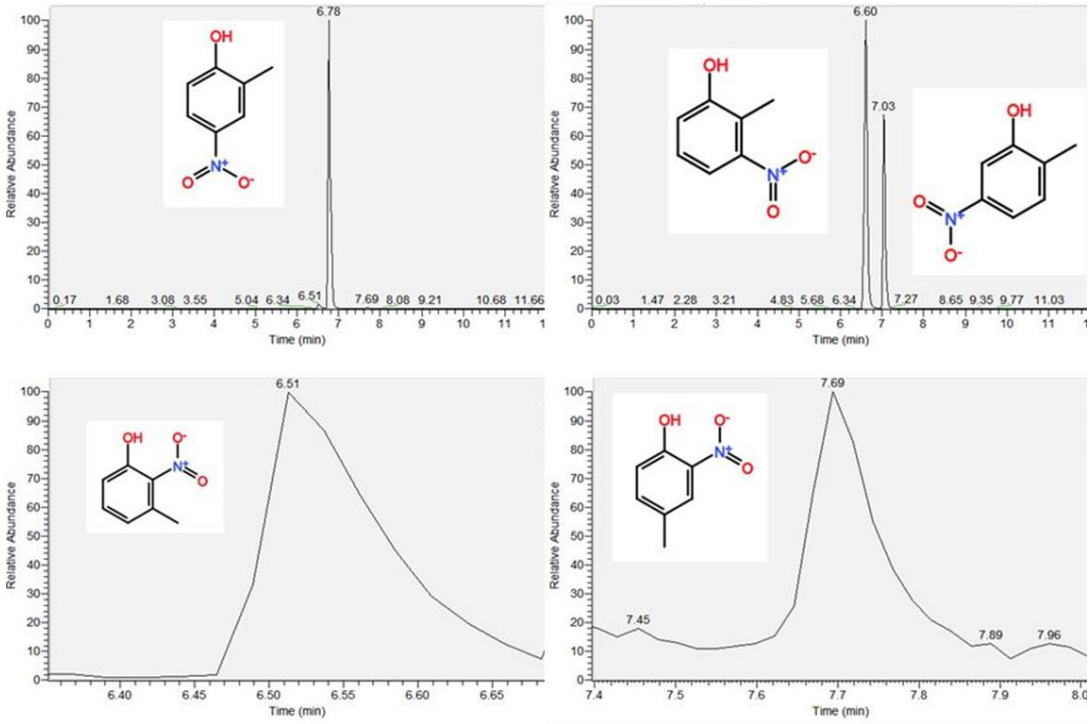

Figure 4.3: Extracted ion chromatographs of methylated nitrophenolic compounds, each peak is annotated with the structure of the corresponding compound. It should be noted the ortho-nitrophenolics in the lower two panels can be observed as very small peaks in the upper two panels demonstrating the significant differences in peak size at the same concentration.

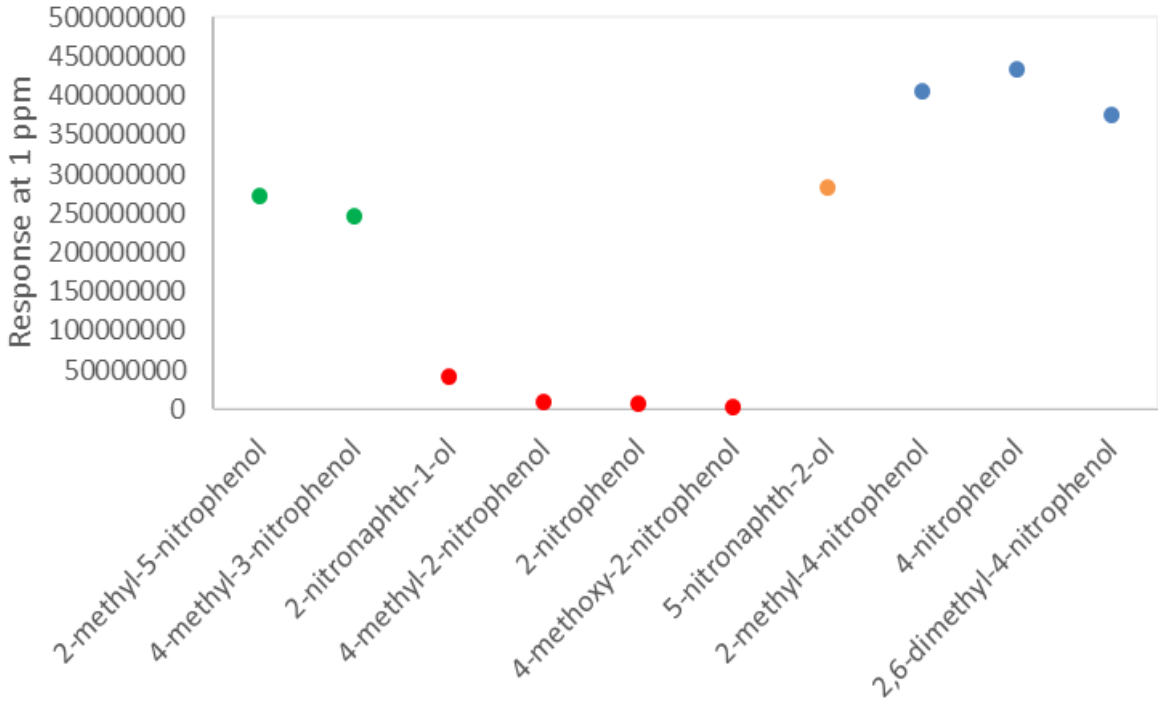

Figure 4.4: The peak area responses of nitro-containing aromatic compounds in a mixed solution of 1 ppm. Points are colour coded for position of nitro group: green – meta, red – ortho, blue – para and orange – not on the same ring as hydroxy group. It should be noted that the different positions of the nitro- group lead to clusterings in intensity of the response.

Figures 4.5 and 4.6 shows the EIC chromatographs of nitro-phenolic standards, isomers of nitrophenol, methyl nitrophenol, a dimethylnitrophenol and a methoxynitrophenol to ensure sufficient separation for identification. From the mixed calibration, the individual nitrophenolic species were observed to have a minor overlap in their respective retention time windows and thus were deemed sufficiently separated in the method for identification, using the adjusted Tracefinder™ method specific to nitrophenolic compounds.

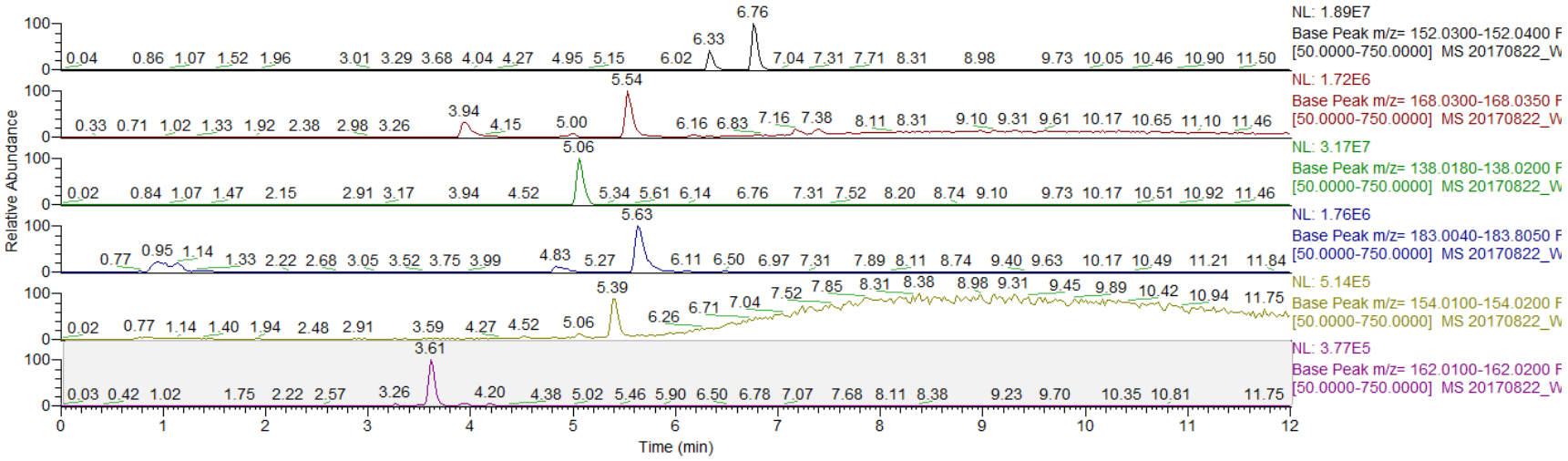

Figure 4.5: Separated EICs of standards in mixed solution, including isomers of nitrophenol, methyl nitrophenol, a dimethylnitrophenol and a methoxynitrophenol

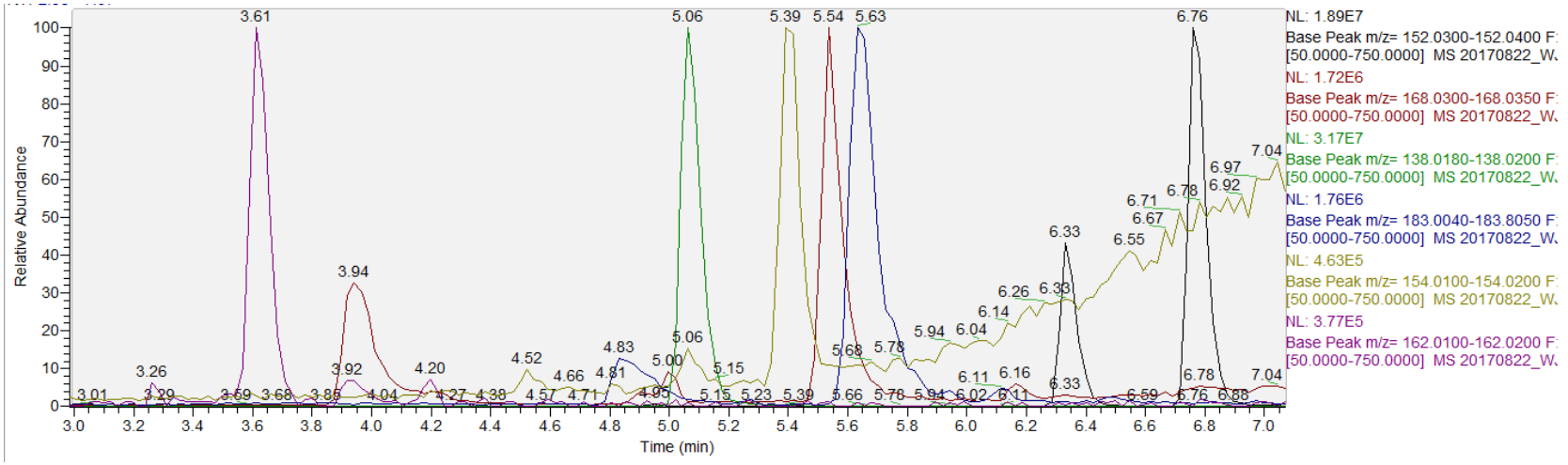

Figure 4.6: Overlaid EICs of standards in mixed solution, including isomers of nitrophenol, methyl nitrophenol, a dimethylnitrophenol and a methoxynitrophenol

4.3.2 Time series of nitrogen containing components

The temporal trends for the nitrogen containing fraction of the organic aerosol in the winter and summer sampling periods show very different profiles. Strong, consistent diurnal patterns are observed in summer. The maxima are seen in the first filter sample of the day (around 9 a.m.) and minima are observed during the over-night samples. This is in contrast with the winter time series, which shows much lower levels of consistency within the temporal trends. In winter, three periods of high abundance are observed: 19th to 21st November, 23th to 27th November and 3th to 5th December. The period between 23rd and 27th November occurred during a period of low $PM_{2.5}$ and demonstrate very defined diurnal profiles, more closely matching the summer diurnal pattern; again, with maxima around the first filter sample of the day and minima during the overnight sample. This is consistent with the total organic fraction observation during these periods; this shows that during this very low $PM_{2.5}$ period, the nitrogen containing fraction is the most abundant driving factor for the total observed signal in this method.

4.3.3 Source apportionment and cluster analysis

As previously discussed chapter 3, 43 nitro-containing species were observed by the automated peak detection. Hierarchical cluster analysis (HCA) was performed on these species to determine the relationships between the compounds and best estimate cluster groupings from the timeseries of the nitro-aromatics. HCA is a form of divisive clustering, this is a 'top-down'

approach to cluster analysis, i.e. from the common branches out to smaller sub-branches, the hclust function in R is a common and easy to use way to produce such an analysis, this calculates the distances between two clusters as seeks to maximise distances between two clusters as discussed in greater depth in chapter 2. If measured in the form of a dendrogram, a 'slice' can be taken through the branches to clearly define where the separations are and match individual species into the clusters, this slice is usually user defined and for the purposes of this, usually a manageable number of clusters are chosen (which still represent the data).

The HCA analysis highlighted six cluster groups as shown in figure 4.7, using the as.dendrogram function in R, the groups selected were compared against the groupings selected by the CorPlot function in R[157] as seen in figure 4.8. From both cluster analysis techniques, 6 cluster groups can be determined. The groups which contribute the most significantly to the total nitro containing signal are groups 5 and 6, which contribute more than 80% of the total signal.

On the surface, figure 4.7 shows that most compounds have positive correlation with each other except for group 1, containing the 2-nitrophenol. It would be logical that as most of these aromatics have similar sources and precursor compounds, if the activity of these sources or the abundance of the precursor compounds increases, there would be a general increase in the general levels of nitrophenolics, as such it is expected that the nitrophenolics generally have positive correlation. Based on favourability of reaction mechanisms to for nitrophenolics selecting ortho- positioning over para- and much more over meta- positioning; it would follow that formation of ortho-nitrophenolics would be in competition with para-nitrophenolics for source

components. Hence, their formation would most likely have anti-correlation as one isomer would form in preference over the other with source aromatic material being the limiting factor.

Another possibility for the anti-correlation is the loss of ortho-nitrophenolics by degradation into HONO, which does not occur in the other isomers as such, their degradation would cause losses in the 2-nitrophenol time series which would not occur in the others and would lead to an apparent anti-correlation. In a few cases, the is an absolute correlation or anti-correlation (i.e. where R^2 = 1 or -1), this occurs when compounds only are observed at certain time points and as such, time points where two compounds occur are also limited, and so if two compounds occur only at a couple of points, this leads to an absolute correlation or anti-correlation.

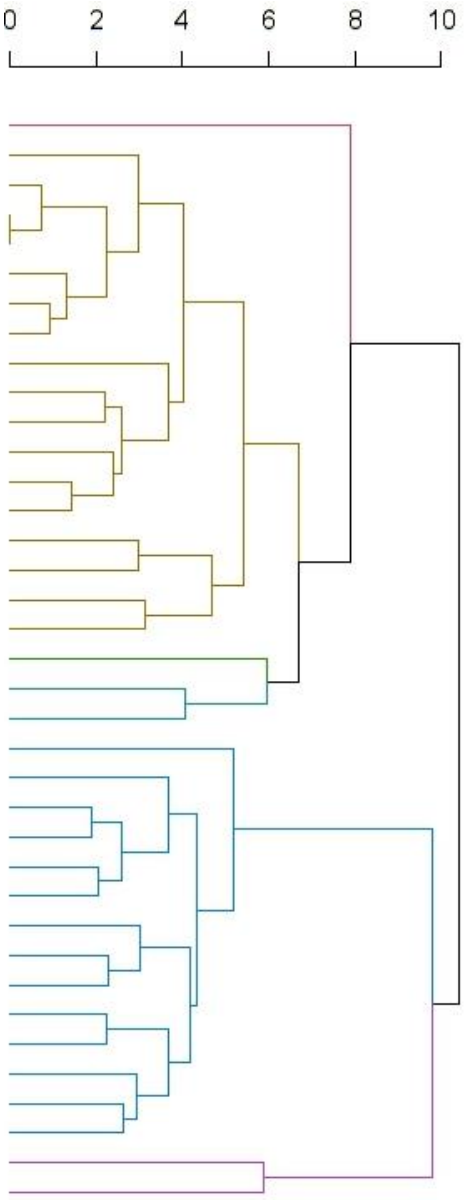

Figure 4.7: Dendrogram produced by HCA, each of the six groups selected are colour coded for ease of identification, the axis shows the height of the dendrogram, which is a representation of dissimilarity. The higher the height, the more dissimilar the branches being linked are.

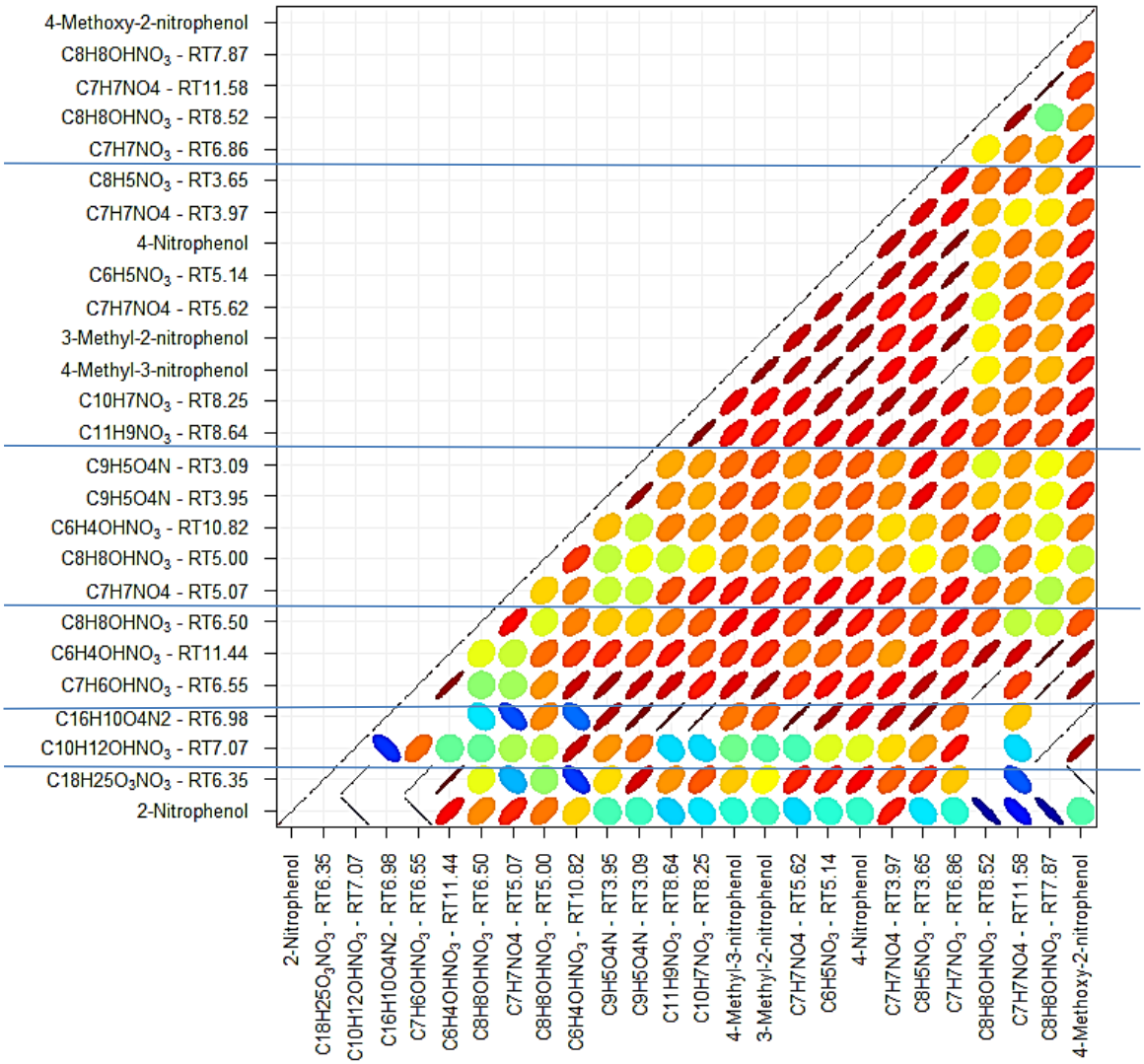

Figure 4.8: Co-variance plot of 43 CHON compounds observed, with enough data for correlation analysis, in the samples colour-coded by correlation. More red points show higher correlation, bluer points show higher anti-correlation and greener points show no significant correlation between species. Horizontal lines have been added to easily identify the group clusters.

In figure 4.10, the time series for each of the groups can be seen. In general, each of the groups are defined by very sharp and spiky peaks, suggesting very quick formation, or at least rapid introduction to the sampling site, and quick

removal, most likely by further reaction rather than dissipation of the air mass (given general low wind speeds and movement of air-masses would take more time than is observed for the removal of nitrophenolics to occur) and given that the losses of the CHON compounds are during high photo-activity periods.

As discussed previously, two air masses are observed during the winter sampling period, Group 1 has the highest concentrations during the earlier sampling period. This group of compounds, consisting of two compounds which evenly contribute to the group, appear to be longer-range transported compounds, becoming more prevalent during higher speed winds and typically during south-easterly winds as observed in figure 4.9. A possible source of this could be longer-range transported airmasses from more populated, coastal cities (such as Tianjin) based on the direction of the source winds. Interestingly, 2-nitrophenol should be the most common species based on previous analysis and the preference for ortho positioning in the formation of nitrophenolics[40,48,155]; however, we see that non-ortho species dominate in the later days of the sampling period. Potentially, this apparent lack of 2-nitrophenol could be due to the previously described inhibition in ionisation efficiency of nitro-phenolics with ortho-functionalisation. Additionally, due to the air-mass being subjected to transport and having a longer life-time, this would also expose the airmass to a greater change of oxidation and thus be more susceptible to chemical change or degradation. It would be therefore fair to assume that some of the 2-nitrophenol has experienced self-reaction and formed HONO as previously discussed in section 1.4. Ultimately, the most likely cause for this apparently lack of a common compound is a combination of the two proposed events. Unfortunately, the other compound in this group does

not match any of the standards and as such, not knowing much about its structure, cannot lend much insight into the sourcing of 2-nitrophenol.

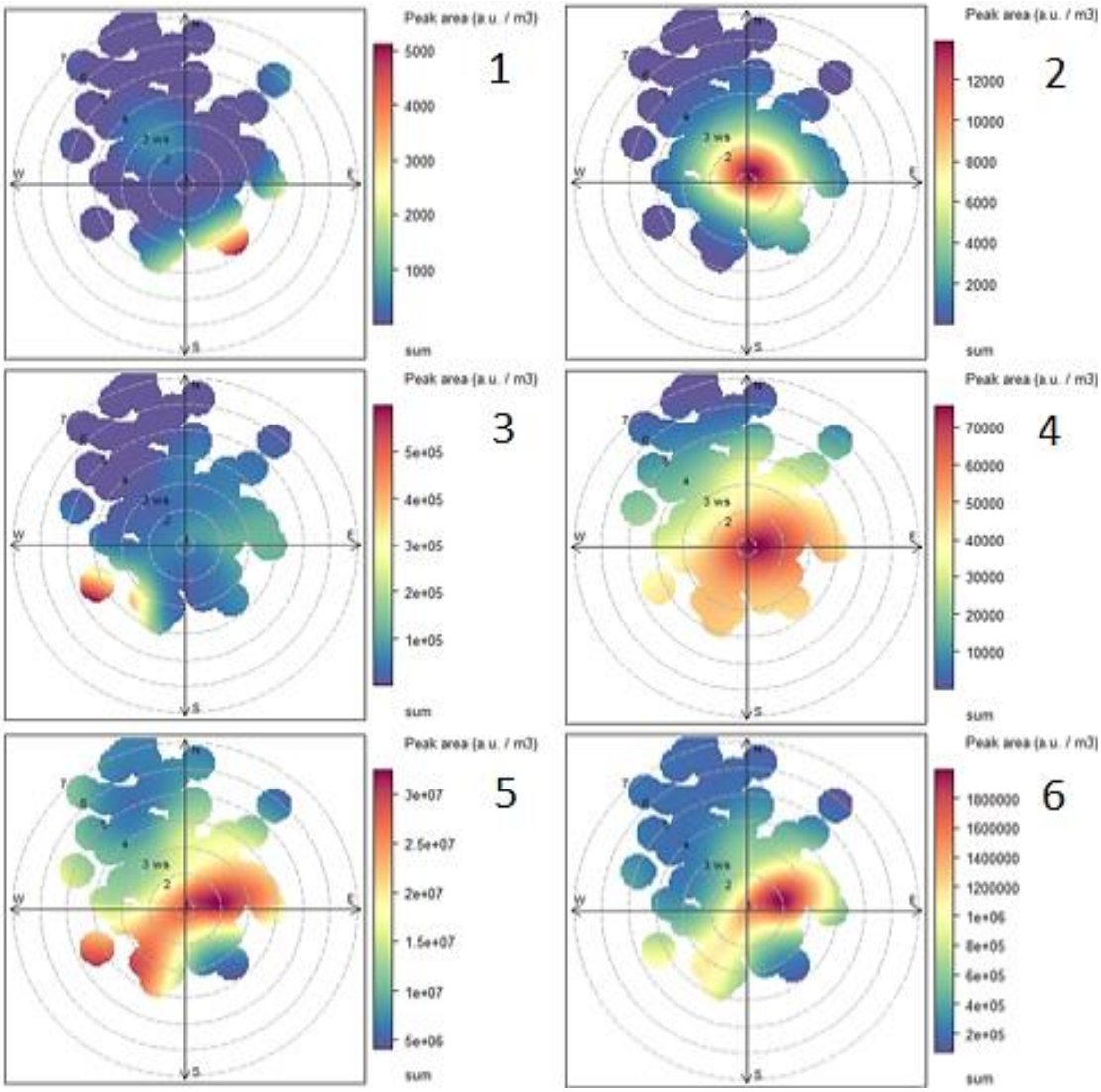

Figure 4.9: Source wind rose plots and summed time series for the 6 cluster groups, the averaged time points corresponding with the filter sample are represented by a point and colour coded by the summed abundance of the components of each group.

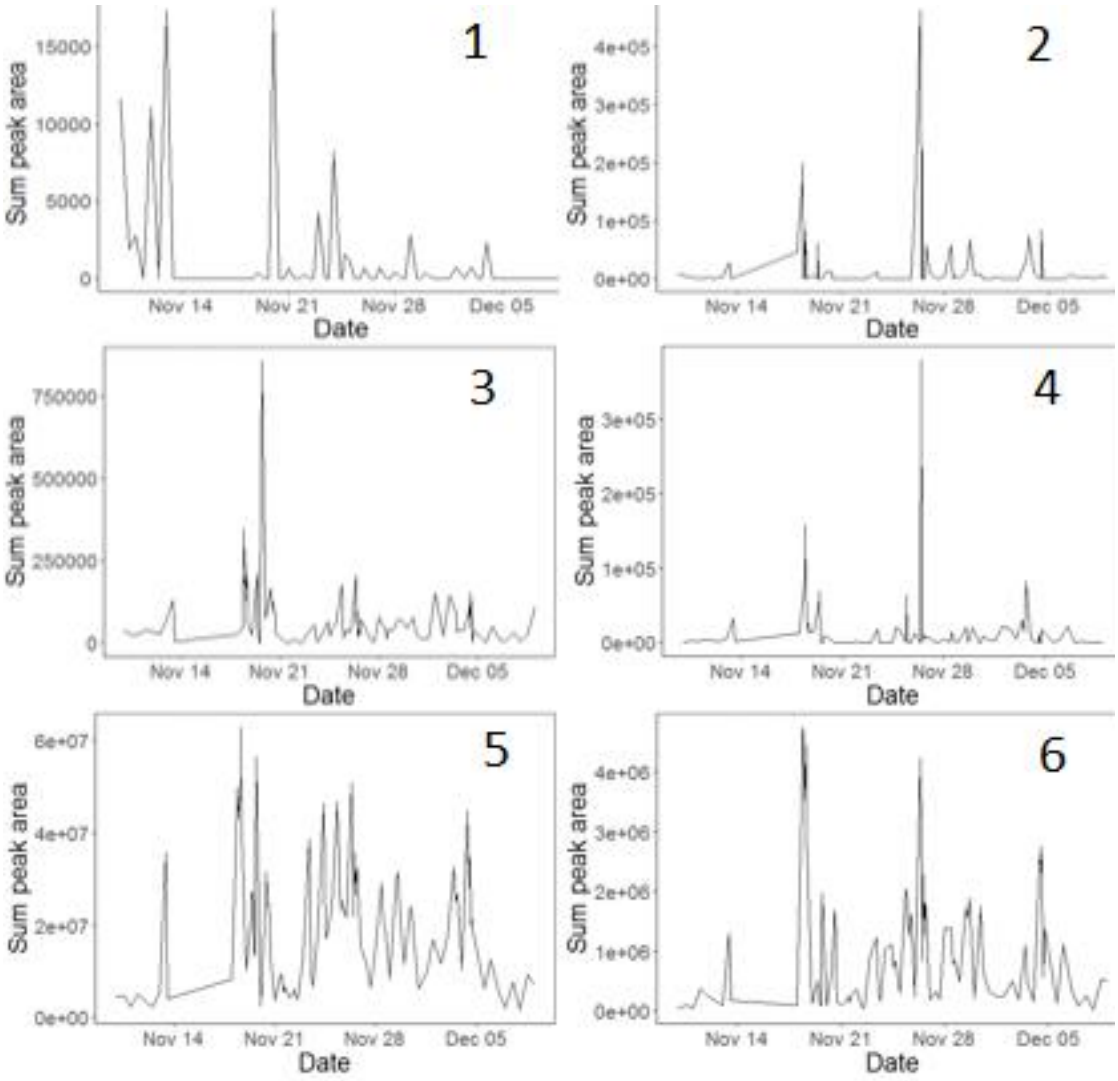

Figure 4.10: Temporal trends of the summed peak areas for the 6 cluster groups defined by HCA of nitrophenolics

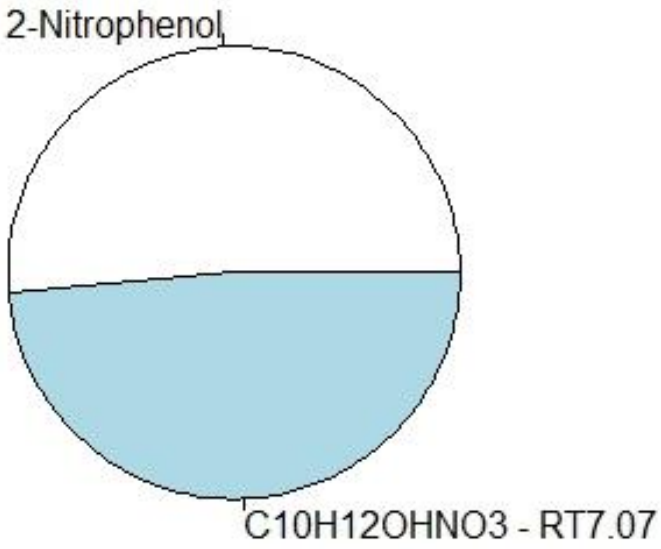

Figure 4.11: Total summed contributions of each component across the time series to group 1

Group 2 has relatively small peaks, with the highest being on the 27th November and a second significant peak on 19th November, occurring mostly during low wind speeds, thus is most likely to have a very localised sourcing. The compounds which comprise group 2 are mostly dominated by one nitrophenolic compound, however none of the species matched any of the standards used in this analysis and so could not be fully identified. As a result of this, it is hard to determine structure or functionality which may shed more light onto the source or sources of the compounds in group 2. The compounds which are less abundant ($C_{18}H_{25}O_3NO_3$ and $C_{16}H_{10}O_4N_2$) are far more functionalised and have higher DBE values suggesting these are much more aromatic in nature and, given the functionalisation of these compounds, are most likely anthropogenic in nature. Since the group is most prevalent during

low wind speeds, it could be inferred that these are from traffic emissions which have been oxidised during the build-up of pollution during the low-wind speeds. However, as they are a minor contributor to the group, it is more likely the most significant contributor to that group being a direct emission from traffic, influencing the wind rose plot.

On the other hand, group 3 shows similarities to group 1, in that its component compounds demonstrate a longer ranged sourcing; albeit with a directionality that is more towards the south west rather than south east. The compounds which contribute to this group are also most abundant around the 19th of November, during the first observed air mass during the winter sampling campaign. The observed signal from group 3 is evenly composed of four compounds between C_6 and C_8, most likely further functionalised nitrophenolics in nature. The components are seen more consistently than groups 1 and 2. There is a repeating (albeit, less abundant) diurnal pattern occurring between 23-27th November with maxima occurring at 9am and minima occurring during the overnight sample. Given the higher abundance during higher wind speeds, this suggests that the compounds are formed south of Beijing, transported north and mix in with the local air in the morning as air-masses mix.

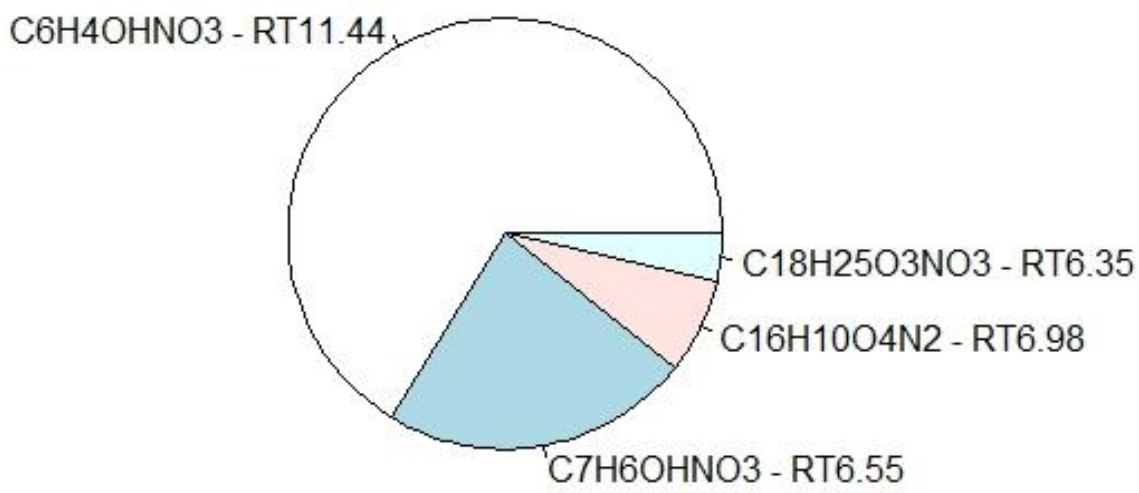

Figure 4.12: Total summed contributions of each component across the time series to group 2

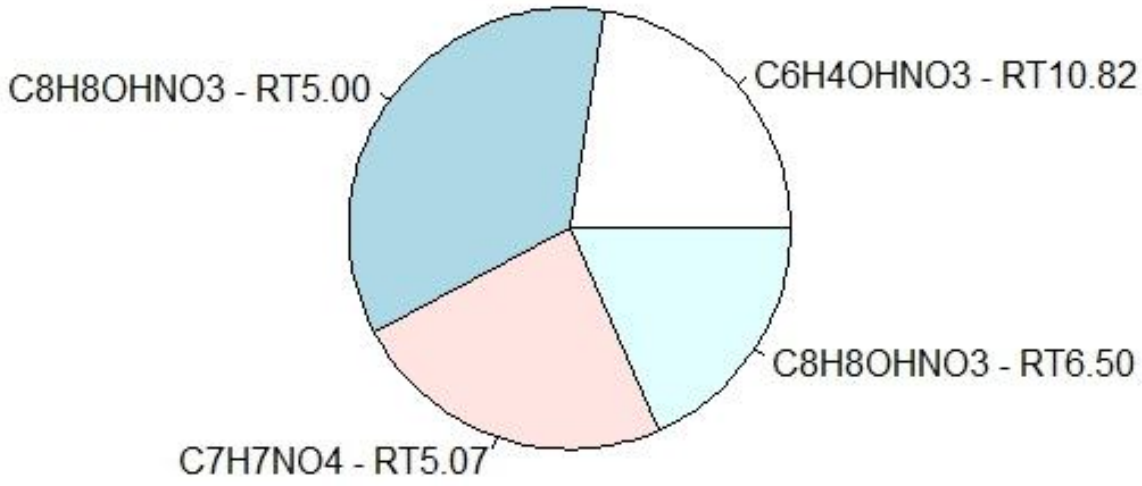

Figure 4.13: Total summed contributions of each component across the time series to group 3

Group 4 is the most unique out of all of the groups. The directional source seems to be almost a point source, but not as localised. This suggests that the

true source is quite wide spread and local with a particularly high abundance close to the sampling site; as opposed to being transported from longer-range air masses travelling which the other three groups have demonstrated up until now. The particularly high level point source can be explained by increased abundances during low wind speeds, for example the higher abundance on the 27th November where wind-speeds were notably low. Based on the compounds which comprise group 4, including 4-methoxy-2-nitrophenol, are most likely sourced from biomass burning. This is further corroborated by the time series, which is very much different to the other groups.

In many parts of China, and near Beijing in particular, biomass burning is still a common heating source, due to the abundance of fuel and its relatively low cost. The Chinese government is making major pushes to put residents on the grid of district electrical heating and power[158,159], however this is a slow and costly process and as such, biomass burning is still very prevalent. Group 4 also has a large contribution to the overall signal produced by nitro-containing components in the organic aerosol. As such, this group may become less of a significant contributor to the overall CHON levels within Beijing as the Chinese government's policies to crack down on biomass based fuels and replace them with a more nationalised, electric heating system will reduce the numbers of people turning to biomass burning as a source of heating, the lowering the source of the compounds in group 4.

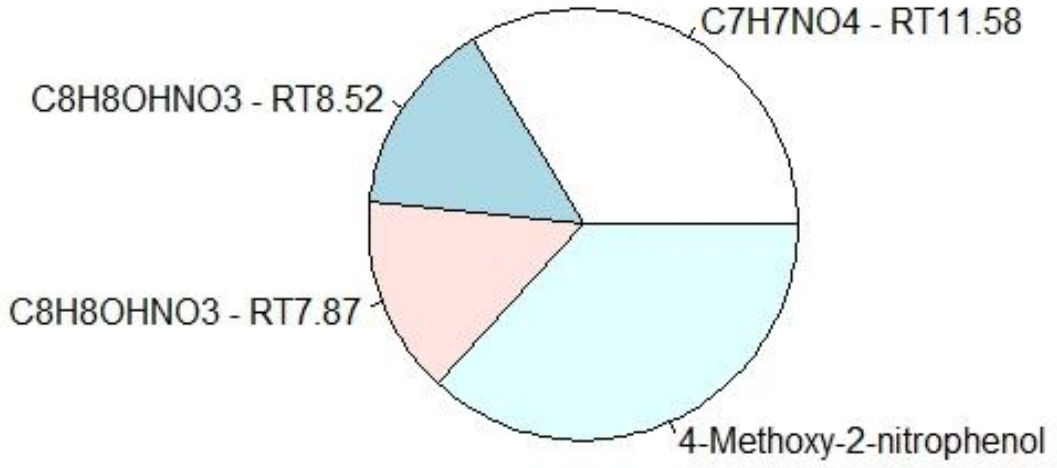

Figure 4.14: Total summed contributions of each component across the time series to group 4

The time series for groups 5 and 6 strongly follow the major features of the total CHON containing time series (as seen in S.I. figure 1, the total timeseries for all nitro-containing species). Three periods of high concentrations are seen between 19 and 21st November, 25 and 27th November and 2 and 4th December. In both groups 5 and 6, a period of very distinct diurnal patterns is observed between 23-27th November with maxima occurring at 9am and minima occurring during the overnight sample. However, while group 5 shows very similar shaped peaks with a gradual build up from day to day, group 6 has two days of lower maxima and a sharp increase in signal for the second two peaks during that diurnal pattern. This suggests that the main source of compounds within group 6 was much stronger during the second couple of days compared to the first. This is in contrast to the sourcing of compounds within group 5, which was more consistent and the differences within maxima was more due to meteorological conditions.

Low wind speeds and the lack of venting of that air mass allowed compounds to build up during this period, leading to the distinct features of this event. The sources of these two groups are also very localised, showing that the peaks seen in this period are local and building up during low wind speeds, rather than longer-range transport bringing these compounds in. In the case of higher wind speeds, this would flush that air mass and remove compounds from the area local to the sampling site.

Interestingly, the directional sourcing is not a point source, the sourcing is almost linear between very low easterly winds and low south-westerly winds, suggesting that the sources of these compounds are more wide-spread. This would suggest a potential traffic emission source. The potential traffic sourcing is further corroborated by the compounds of which groups 5 and 6 are comprised, tending to be mono- or bi-aromatic in nature and are most likely to phenolic, based on their chemical formula. The compounds which comprise group 5 evenly contribute to the total signals from the group; whereas group 6 is largely dominated by the nitronaphtholic compound.

These very aromatic unfunctionalized compounds are most likely anthropogenic in nature based on their chemical formulae and the location of the sourcing suggested these compounds. Group 6 have more sudden losses after their peaks than group 5, this could be due to the nitrophenolics having the nitro group in the ortho- position. As previously discussed, these compounds are more susceptible to degradation and production of HONO compared to meta- and para- nitrophenolics' since the compounds comprising group 5 are meta- and para- in nature, this would explain why group 6 appears to have much sharper losses of signal compared to group 5. Additionally, group

6 appears to have less intensity than group 5 while following similar temporal patterns. This again lends credence to the compounds in group 6 being ortho- in nature as we have previously seen nitrophenolics with an ortho- positioning have generally lower ionisation efficiency, leading to a lower abundance when measuring peak area; it is clear that 2-nitro-1-naphthol is ortho- in nature and therefore it may be the case that the ortho-nitrophenolics and ortho-nitronaphtholics co-vary well and cluster together.

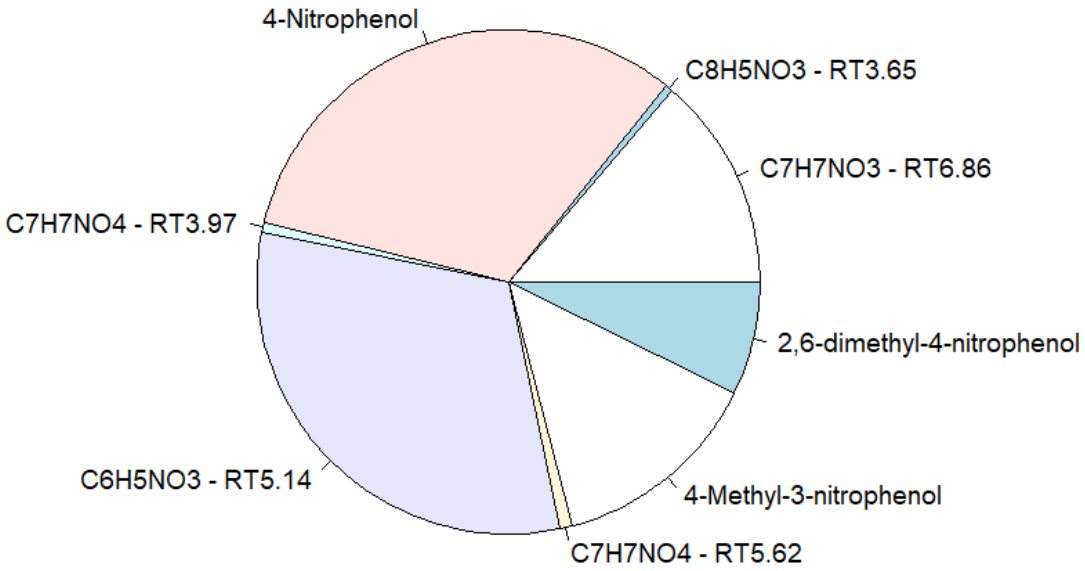

Figure 4.15: Total summed contributions of each component across the time series to group 5

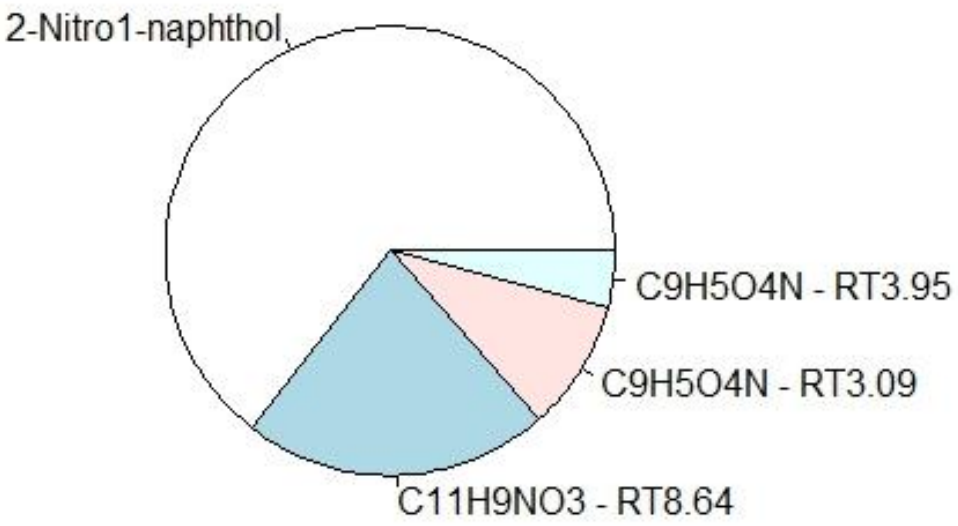

Figure 4.16: Total summed contributions of each component across the time series to group 6

4.3.3 Quantified nitrophenolic temporal trends

Due to the significant variation in ionisation efficiency demonstrated by the nitrophenolic group of compounds, compounds were only fully quantified if the isomer is confirmed and a commercially available standard could be found. In total, 10 isomers were full quantified, and their levels were taken as a percentage of the total organic carbon (%NR-OM) present during these samples, determined by AMS analysis. The most abundant compounds were found to be the 4-nitrophenol, 2-nitrophenol and 2-methyl-3-nitrophenol components, there we found to have maxima %NR-OM of 1.246%, 0.544% and 0.405% respectively as seen in table 4.1. The nitrophenol isomers have previously been found to be very abundant in populated cities[48] with the ortho and para isomers being the most prevalent. Interestingly, it is the meta isomer of methyl nitrophenol which was the most abundant, potentially showing the

directional preferences during nitrification with additional functionalisation. It is notable that with just these top three compounds having a maxima of over 2% of NR-OM combined, this could be as much as 4-5 ng m^{-3} given the upper averages for TOC observed in Beijing, these levels pose serious risk to populations exposed to this atmospheric hazard; additionally as these are all water-soluble there is the aforementioned additional risk of pollution to water supplies from wet deposition. Given the observations observed in chamber experiments by Bejan *et al*, this suggests that 20-25 ng m^{-3} of gas phase nitrophenolics were present in Beijing, assuming similar conversions of nitrophenolics into SOA for all nitrophenolics, not just ortho-nitrophenolics.

The compounds with the lowest minima were 4-methoxy-2-nitrophenol and 2,6-dimethyl-4-nitrophenol (both approximately 0.0005% of NR-OM). 4-Methoxy-2-nitrophenol is a common biomass burning tracer, the large variation in %NR-OM contribution from this compound demonstrates the impact of biomass burning on the organic aerosol and its variation. The other nitrogen containing compounds shower smaller variation between their maxima and minima contributions, whereas 4-methoxy-2-nitrophenol is more tied to periods of heating during particularly cold nights and evenings, as demonstrated by sharp peaks towards the colder nights as opposed to a consistent overnight peak. As previously mentioned, the contribution of this is likely to fall as Chinese government reforms will encourage the transference of heating from biomass burning to a more regional or national electrical source.

When quantifying the nitrophenolics compounds, the mixed calibration standards were used to confirm identification and form calibration curves, as discussed in chapter 2 of this thesis. This calibration range was approximately

100-fold in range to maintain linearity and allow for more accurate extrapolation should the compounds exist outside of the expected range. Herein exposes the limitation with such a calibration, due the nature of nitrophenolics occurring in many different levels in aerosol and some rapidly degrading during the filter's sampling period, this meant that some compounds were observed below the quantification limit of the calibration range. Since the calibration curve was designed with linearity in mind, it was assumed that this would provide a linear gradient from which to extrapolate from. However, since some species such as 3-nitrophenol existed far below this limit of detection, the software's calculation stated that the compound had a concentration of <0 ng m^{-3}. Naturally, this is impossible and demonstrates that extrapolation from such a limited calibration range may not be applicable to real-world samples. As discussed in chapter two, calibration ranges which exceed 1000 fold may exhibit significant non-linearity, making them also unsuited for analysis as broad-scale as this project.

Instead, the author posits that such an analysis may be best suited to a split method, with higher concentration ranges be analysed by one method with a 500 to 1000 fold range and another lower concentration method be performed to capture the compounds which are below the limit of detection of the higher concentration method. This sort of analysis may also benefit from a minor over-lap of calibration curves for compounds on the threshold. Unfortunately, this was not possible in the scope of this work but presents an interesting direction to approach in future work following this project.

In table 4.2, maxima for methyl nitrophenolics and also a di-nitrophenolic compounds reach significant levels. Whilst in the data, these seem to be

momentary spikes, un-characteristic of the averages in Beijing, reaching 15-31 ppt, these concentrations are seen to begin to have toxic effects on organisms[160] and if allowed to deposit into the soil via wet deposition, can build up and cause much more severe harm. As previously noted however, these concentrations are the exception, rather than the rule and average levels are much lower.

Table 4.1 : Percentage contributions to total organic carbon during the winter sampling campaign (%NR-OM, calculated by peak area comparisons to summed total observed peak area for that time point) by the quantified nitrophenolic compounds.

	Minimum %NR-OM	Maximum %NR-OM	Mean %NR-OM
Total Nitrophenolics	0.22	3.45	1.835
2,4-Dinitrophenol	0.109	0.24	0.175
2,6-Dimethyl-4-Nitrophenol	0.000583	0.208	0.104
4-Nitrophenol	0.00756	1.246	0.627
3-Methyl-2-Nitrophenol	0.00142	0.148	0.075
4-Methyl-3-Nitrophenol	0.00203	0.374	0.188
4-Methoxy-2-Nitrophenol	0.000551	0.252	0.126
2-Methyl-5-Nitrophenol	0.00541	0.0311	0.018
2-Methyl-3-Nitrophenol	0.00206	0.405	0.204
2-Nitrophenol	0.0928	0.544	0.318

Table 4.2 : Quantified nitrophenolics maxima and minima across the campaigns where observed peak area were within calibration ranges for quantification

	Maximum ($ng\ m^{-3}$)	Minimum ($ng\ m^{-3}$)
2,4-Dinitrophenol	31.60	0.06
2,6-Dimethyl-4-Nitrophenol	5.16	0.18
3-Methyl-2-Nitrophenol	15.18	0.50
4-Methyl-3-Nitrophenol	15.18	0.64
4-Methoxy-2-Nitrophenol	15.30	0.07
2-Methyl-3-Nitrophenol	15.18	0.64
2-Nitrophenol	3.07	0.09

4.3.4 Potential for mis-quantification using inappropriate standards

The nitro-naphthol compound did not correspond to the standards which were commercially available at the time of this research project. Using an analogous standard is a common alternative choice when exact standards are not available or if there is no separation of isomers (via chromatography or using other methods) and all isomers assumed to have homogeneous behaviour and treated as such. Herein, the potential scale for the mis-quantification of

compounds is quantitatively demonstrated when using inappropriate standards.

The difference in ionisation efficiency between the two nitronaphthol standards, the former with the nitro on the ortho position and the latter with the nitro group on a different ring to the hydroxy group is approximately seven-fold. Using this as a factor in calibrations shows the potential differences in concentrations observed by using inappropriate standards.

Figure 4.17 shows the time series with concentrations determined from the 2-nitro-1-naphthol standard. As the observed peak concentrations are over 1.6 μg m^{-3}; levels which are unprecedentedly high for a single species, this suggests that this may be mis-quantification. Whilst this does not inherently prove that this unknown observed nitronaphthol compound is not observed at this concentration, the levels being so unexpectedly high would suggest that the true concentrations are potentially lower. Figure 4.18 shows the time series of the nitronaphol compound calibrated using 5-nitro-2-napthol, with peak concentrations of around 250 ng m^{-3} which are in keeping with other levels observed for a single species. However, since neither species have the same retention time under the extended method, they cannot be said to be the same species.

As such, both commercially available standards are demonstrated to be equally invalid as calibration standards, giving significantly different concentrations to each other. The actual concentrations of the species when quantified using appropriate standards could be significantly different and most probably much lower than observed (if the abundance of the nitronaphthol species is similarly abundant to the rest of the nitrophenolic group). As such, the apparent

variation in observed quantification using inappropriate standards could leave values approximately factors of 10 away from the true value.

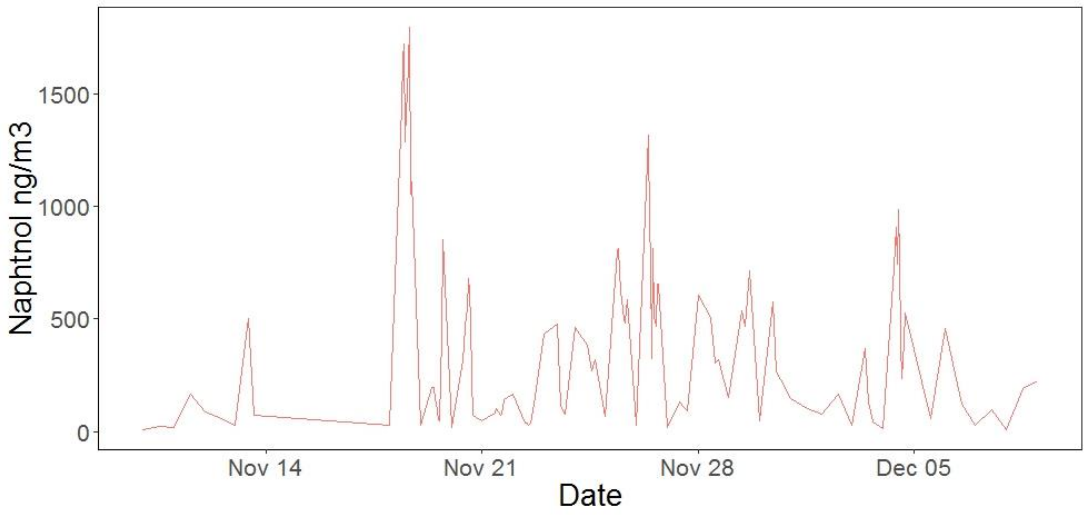

Figure 4.17: Time series for observed nitronaphthol, calibrated using 2-nitro-1-naphthol

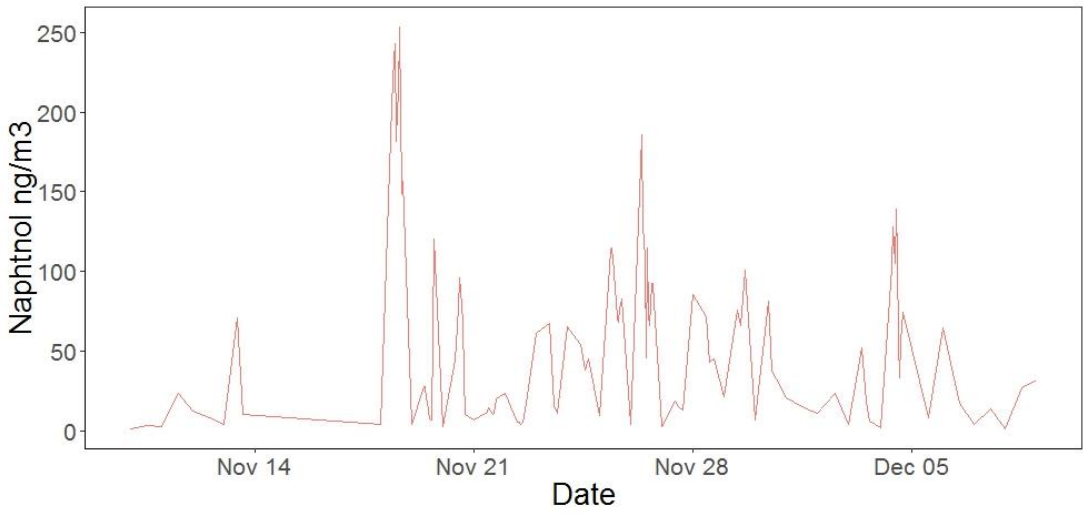

Figure 4.18: Time series for observed nitronaphthol, calibrated using 5-nitro-2-naphthol

4.3.5 Comparisons of observations with analysis of other cities

Total nitrophenolics were observed in this work to contribute up to 1.8% of NR-OM; with a total of 13 compounds, identified fully using this method. A similar number of nitrophenolics was observed in Shandong province (northern China)[56] by ESI-ToF-MS from cloud water, also identifying many of the individual species, such as 4-nitrophenol and 2-methyl-4-nitrophenol. In cloud water, it was observed that at most 3% of the total brown carbon (absorbing at 300 - 400 nm) was caused by nitrophenolics. Highlighting the similarities in components and transference of nitrophenolics between atmospheric and water systems. This is also consistent with levels seen in European countries such as Slovenia, who see 1% of their BrC being comprised of nitrophenolics[56]. In the south of China, near Guangzhou, para-nitrophenols were observed in the soil[161] with capacity for soils in the area to contain 160 mg/kg of para-nitrophenol, highlighting a storage for nitrophenolics post wet deposition, with soils capable of storing up to 160 ppm of nitrophenols and then later transferring them into water cycles. It is also suggested by the US environmental protection agency that the safe levels of nitrophenolics in water should not exceed 1 mg/L, as such large amounts of wet deposition from atmospheric nitrophenol should be considered when determining if water potentially exceeds this limit. Schurrmann *et all*[160] observed 50% inhibition of growth of organisms at nitrophenolic levels between 0.1 and 300 mg/L, as such there is a clear and present issue presented by nitrophenolics at this level in the atmosphere.

In Chongqing, a mega-city known for production of electronics and cars, Tian *et al* reported deposition of oxygenated Polyaromatics (OPAHs) averaged as high as 536 ng/m^2/deposition[162]. The nitro-naphtholic compounds observed within this work would fit within this group and could be used as a tracer to represent the group. If most nitro-naptholics have ionisation efficiencies similar to 5-nitro-2-naphthol, then based on our observances, the concentrations observed in this work the average observance for a nitro-naptholic compound would be 130 ng m^{-3}, which would fit with the work of Tian *et al;* assuming not all of the OPAHs would be deposited during wet/dry deposition, it would be logical to infer that the atmospheric levels of OPAHs are much higher than that observed during a deposition event. Some variation in concentrations observed can be attributed to differences in meteorological conditions and geolocation. Chongqing is further south than Beijing, closer to the more humid region of China, this is discussed in the paper as aiding the generation of OPAHs and their partitioning into the particulate phase. Interestingly, Tian *et al* state that nitrated PAHs are on the lower end of those observed in deposition events, compared to other PAHs and as such it is most likely these remain in the atmosphere. Finally, their PCA analysis attributes the majority of the PAH to be from secondary formation and combustion emission. This is in keeping with the findings from this work in Beijing, with the majority of nitro-aromatics appearing to be caused from biomass burning and from overnight nitration of aromatics.

4.4 Conclusions

It is clear that the picture of nitro-containing compounds is complex. With the majority of these species being nitro-phenolic in nature. A large amount of these compounds appear to be locally sourced, as opposed to being brought to the sampling site via long-range transport of air masses. The compounds were observed to be best fitted into six major clusters when HCA was applied to their temporal trends. The most important of these cluster groups, show defined diurnal patterns during low PM periods suggesting that during this low period, the total signal is most influenced by nitrophenolic components. This may suggest that at low $PM_{2.5}$ periods, where wind speeds tend to be higher, the nitro-containing fraction of the organic aerosol may be cleared from the Beijing area; but may end up contributing to another regions poor air quality or be moved to a region where it can be readily transfer into the water cycle via wet deposition and present its hazardous behaviour in other ways. The low PM periods also show that even when the air mass is being cleared, there are significant sources of nitro-containing species to contribute and influence the local atmosphere in Beijing, making them a consistent threat which at times may be overlooked due to more abundant groups.

During periods influenced most significantly be CHON, these components showed peak levels around 9 am, however are most likely to be generated overnight and reside in the upper nocturnal layer, away from sampling, leading to an apparent low concentration at a sampling height of 8m. This would suggest that in the morning as the air heats up and beginning to mix these layers, these compounds would then mix down and subsequently be sampled during the morning sampling period (8:30-9:30 am). This could also point to a

traffic emission during a morning 'rush hour', however without the presence of an evening 'rush hour' event; the most likely explanation, which agrees with literature is the nocturnal boundary layer separation theory.

The concentrations of nitrophenolics tend to decay during the day. The degradation of nitrophenolics due to photoxidiation has been previous documented[163] and this degradation leads to the defined 'shark-fin' shape of the most abundant compounds and hence the shape of the time series of the total signal for nitrogen-containing components. The confirmation of this explanation can be performed in future studies by sampling above and below the boundary between the separated nocturnal layers; indeed, at the sampling site, there is a 300 m tower which would be ideal for this sort of experiment. However, for this work, there was only one sampler available and with only one sampler at a fixed height of 8 m from ground level, this was not possible in this study. Additionally, during the high frequency sampling, this would mean going up and down the tower to replace and store the filters at -20 °C every 30 mins, which would be impractical and without due care, may expose the filters to the risk of being damaged during transit from the sampler to the freezer.

Gas phase nitro-phenolics produce aerosol nitrophenolics when exposed to UV light, without the need for addition oxidants (other than the OH radicals generally produced by exposure of air to UV light). This demonstrates that both directly emitted and indirectly emitted nitrophenolics can both easily produce large quantities of aerosol (roughly 20% of its concentration can become SOA, based on chamber experiments of ortho-nitrophenolics).

The signal from biomass burning is unlike the more common nitrophenolics and does not follow the distinct diurnal cycles, instead demonstrates sharp peaks during particularly cold periods corresponding with increased biomass burning; and comparatively little biomass burning tracer is seen in the warmer, summer sampling period. As previously discussed, going forwards, this is the group of compounds most likely to reduce in the coming years as China is reviewing it's policies on heating in cities and replacing biomass burners with electrical grid heaters.

The most abundant isomers of nitrophenolics have the nitro- group at the ortho and para positions. These are also observed to be the most toxic of the compounds with Wang *et al*[164] observing that the nitrophenolics with strong electron-withdrawing properties (i.e. ortho or para functionalised) are the more likely to inhibit growth in organisms reinforcing the need for accurate quantification of ortho-nitrophenolics and under-estimation of these particular compounds lead to the biggest unknown risks relative to the other isomers. It is observed that individual compounds can peak as high as 30 ppt; however this is the exception, rather than the rule, but still highlights the severe effects of pollutant build up during low wind speeds. When compared to observation in other Chinese mega-cities, the concentrations of nitro-aromatics are similar to very industrial cities; however, observations performed in literature posit that the majority of nitro-aromatics remain in the atmosphere and are not as susceptible to deposition as other PAHs might appear to be.

Some nitrophenolics are observed to have levels outside of the quantifiable range, as such extrapolation of the calibration range was required. However, in some cases, the extrapolation proved to be invalid for compounds such as

3-nitrophenol. Future analysis would benefit from a split-level approach to quantification with one method targeted at higher concentration compounds and a second method targeting lower concentration compounds. This would minimise the need for extrapolation and make quantification more valid for difficult compounds with low concentrations and also for compounds which demonstrate concentration variation too vast for a calibration curve to be able to quantify without excessive ranges.

Chapter 5 – Organosulfate contribution to ambient aerosol in Beijing

5.1 Introduction

Oxidation of volatile organic compounds (VOCs) within the atmosphere forming secondary organic aerosol is a large contributor to fine particulate matter. One commonly prevalent group of components which are susceptible to photooxidative aging and oxidation within the atmosphere are biogenic volatile organic compounds (BVOCs). These compounds and show particular sensitivity to oxidation from hydroxyl radicals and inorganic nitrate[25]. Common BVOCs, such as isoprene and monoterpenes (species with the generic formula of $C_{10}H_{16}$, such as α-pinene and β-caryophyllene) are found abundantly across the planet. Current investigation of emissions from forested regions have put the figure for isoprene and terpene emissions at approximately 500 Tg of carbon annually[31]. As a result of the sheer abundance for this group of compounds and the resulting oxidised tracers, any negative effects to health or atmosphere are inherently going to be significant and thus significant work is required to investigate the ultimate fate of the resulting SOA.

One particular group of SOA from this form of oxidation are organosulfates. Organosulfates were first identified within atmospheric aerosol by Surratt *et al* 2007[58]. These sulfate esters were discovered in chamber experiments during the oxidation of biogenic VOC in the presence of acidic seed aerosol. Due to the lack of thermodynamic favourability of such reactions, the formation of organosulfates are very sensitive to the conditions of the oxidation[58]; as a result of the sensitivity to these factors, the abundance of organosulfates tends

to be variable across different regions. Factors such as pH, temperature and humidity are the most important in controlling formation of organosulfates due the formation pathway[25,165]. Properties of seed aerosol also play a factor in the formation of aerosol and much work is currently being undertaken to understand the intricacies of their effects.

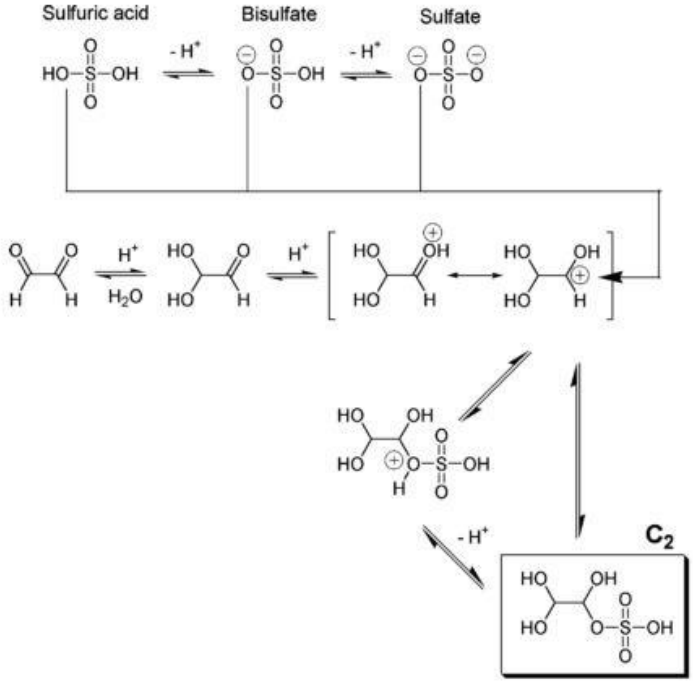

Figure 5.1 A Proposed mechanism for the oxidation of glyoxal to its sulfate ester. Taken and formatted from Liggo *et al* 2015[166].

Due to the variation in structures of organosulfates, being both directly emitted (for example, surfactants found in common hygiene products) and products of oxidation of VOCs (such as the aforementioned biogenic organosulfates), it has become a challenge to identify and characterise the complete composition of the organosulfate contribution. Individual organosulfates display a significant

variation in structure functionality, in addition to the sulfate ester and isomerism within individual species.

It becomes apparent that the only commonality between the compounds would be the sulfate ester group. This leads to difficulty when determining the exact chemical structure and nature of each of the individuals within the group as a whole, since the large number of possibilities means there are no set rules or tendencies that the compounds must abide by, in order to be considered part of this group. Thus assumptions in chemical properties or behaviour, cannot be formed as would be possible with other groups, such as nitrophenolics.

At present, the simplest way to identify organosulfates is to identify and break the group, as a whole, down into common sub-sets present in the group[153,167,168]. One common characteristic organosulfates share is the high polarity, due to the hydrophilicity of the sulfate group. As a result of this, organosulfates elute in the very hydrophilic region of LC analysis (for reversed-phase LC (RPLC) this would be very early in the run time) potentially leading to poor separation of isomers as most organosulfates possess very similar hydrophilicity and thus similar elution times. Instead, hydrophilic interaction chromatography (HILIC) is used as an alternative to RPLC[66]. The different packing allows for a good separation of hydrophilic components, however its use on less hydrophilic compounds is less successful, leading to poorer shaped peaks. HILIC also requires longer gradients to allow the bi-phase to adapt to the changing mobile phase and not break-down.

This work investigates in greater detail the complexities of the organosulfate components within the organic aerosol. This is performed by using a combination of the previously described automated analysis of UPLC-MS^2

analysis and from a HILIC-based LC-MS system and a comparison of these two techniques with regard to their differing insight into the organosulfate group.

5.2 Methods and materials

5.2.1 UHPLC-MS

All liquid standard samples were analysed by ultra-performance liquid chromatography mass spectrometry (UPLC-MS), specifically an Ultimate 3000 (Thermo Scientific, USA) system attached to a Q-Exactive orbitrap (Thermo Fisher scientific, USA) using a heated electro-spray ionisation source (HESI) and a resolution of 70,000 at m/z 200. A reverse phase method was employed for the UPLC analysis, 5 µm 4.6 x 100 mm, C18 column at 40 ^{o}C. The working mobile phase was composed of HPLC grade water and 100% MeOH (both from Fisher Chemical, USA) the water was acidified using formic acid to improve peak resolution; the final concentration was 0.1% of formic acid. At a flow rate of 300 µL min^{-1}, the gradient was as described in chapter 2. The full separation took around 16 minutes. The mass spectrometer was operated in negative mode using full scan MS^2 using data dependant MS^2 scans for the top ten observed signals.

5.2.2 HILIC LC-MS

15 extracted biogenic samples from the summer campaign were run on the purpose built HILIC method and were performed at University of North

Carolina using the method described in Cui et al[66]. The total chromatography run time was 30 mins.

5.3 Results

5.3.1 Overview of organosulfate fraction

Analysing the heteroatom groups by peak area signal for an overall look into the composition, the sheer signal from the organosulfate group is not as large as the nitrogen-containing group in terms of pure peak area response using MS detection. This is due to the differences in ionisation efficiency between the two functional groups (NO_2 and SO_4); however, the sheer numbers of observed species are much larger.

In total, 191 individual species were observed across the two sampling campaigns, these species varied vastly in terms of their properties, such as in the number of carbons in the molecule (for both chain length and branching), DBE and polarity (determined by RT on the LC method).

Figure 5.1 shows the spread of carbon number and oxidation states of observed organosulfates. The majority of the compounds are in one of two areas of the plot. The first group encompasses the smaller, more oxidised species (most likely to be a result of VOC oxidation) and the second is larger, less oxidised more aliphatic compounds (more likely primary emissions such as surfactants). The more abundant compounds (or at least the compounds with the largest observed signal) mostly fit into these two groupings as did the majority of the

individual species observed across the two campaigns, meaning these two groupings are the most significant.

It should also be noted how low the average oxidation state of carbon (OSc) values are for these molecules when compared to the average OSc of other species in literature[37]. This is due to how we have calculated OSc in this work. The justification for this choice is that only one oxygen from the sulfate group contributes to the oxidation state of the carbon backbone (and hence the average oxidation state of carbons within the species) and therefore only one of the 4 oxygens from the sulfate is relevant for OSc calculations. As a result, the number of oxygens in our OSc calculation was been reduced by three, to remove the non-contributing oxygens from the oxidation state of the carbon backbone.

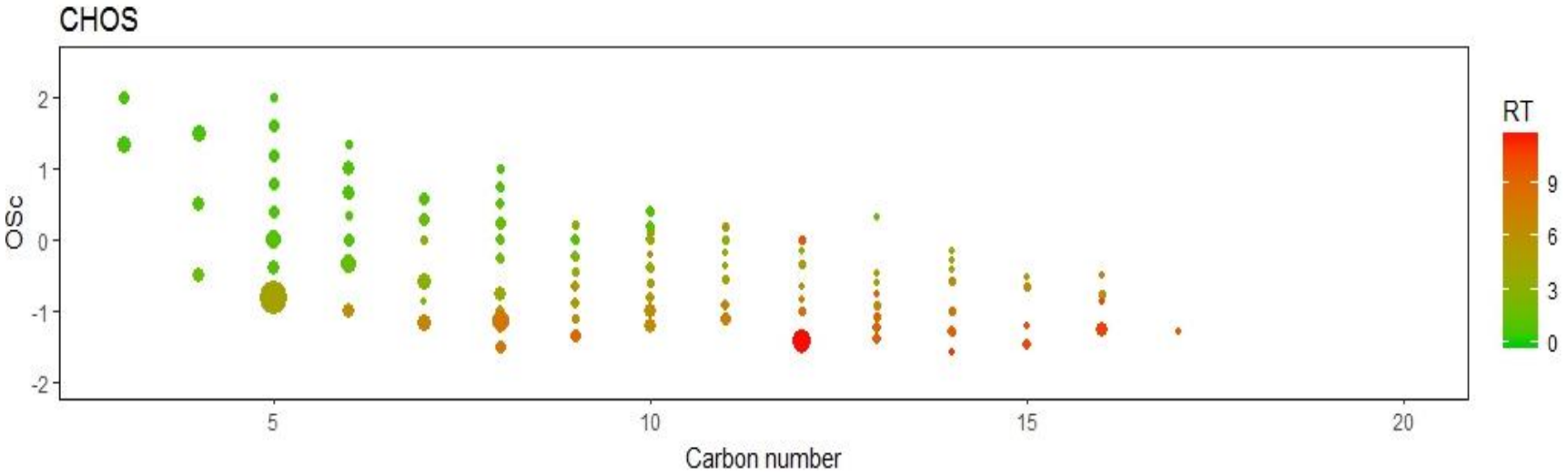

Figure 5.1: The average oxidation states of carbon (OSc) of organosulfate components in the aerosol samples. Point sizes are indicative of signal level and point colour is indicative of retention time (a proxy for polarity in the chromatography) with greener colours representing more polar compounds and more red colours showing less polar compounds.

5.3.2 Quantification of organosulfates using automated UPLC-MS^2

Due to the variation of structure in organosulfates and the previously observed variation in behaviour between compounds possessing differing structures (and indeed, in certain cases isomers of the same species), finding suitable and representative compounds to use as an accurate representation for all organosulfates which occur within the organic aerosol is challenging. The ideal for this type of analysis would be to employ the use of 100s of standards in complex calibration curves. However, this was not possible in this study as we were limited to the commercially available standards as we did not have the synthetic abilities at hand to produce the numerous standards required. In this study, the standards used were those commercially available and as such were limited in number but were more varied in terms of their structure and functionality.

The observed organosulfate species were dominated by a high number of low DBE (usually DBE = 0 or 1) compounds, suggesting a higher number of aliphatic or surfactant like species. Given the urban nature of Beijing this is to be expected due to the large amounts of cleaning products used in domestic environments and produced by the industrial district towards the south of Beijing.

5.3.3 Preliminary source apportionment

To get a better understanding of the organosulfate fraction, the compounds must be sub-set into groups with similarities; for example compounds which

possess similar sourcing or similar mechanisms of formation. This can be performed by the analyst reviewing the compounds by eye and using their personal judgement to group compounds, however this introduces inherent bias as it requires a human input as to which groups 'should' behave more similar than others. As such compounds in manually chosen groups may not demonstrate similar behaviours as would be expected.

A more hands-off approach, which avoids the inherent bias of human input, comes from clustering techniques, applied to the time series of each individual species. Commonalities within these time series would indicate common sourcing or activation of similar mechanisms of formation or aging. Using co-variation as a proxy for similar behaviours are more likely to closely resemble accurate groupings of similarities compared to manual groupings based on appearance or reported properties.

Hierarchical cluster analysis (HCA) was applied to the 191 CHOS compounds during the winter samples in order to determine the co-variance of compounds within the CHOS heteroatom group. Figure 5.2 shows the co-variance of all 191 observed components with more red points showing high positive correlation between the two species. Points which are more blue show stronger anti-correlation between the two species. Finally, greener points show no significant correlation neither positive nor negative. From this, it is clear that manually selecting groups in such complex matrices would present a very-time consuming and problematic challenge.

As a result of the HCA analysis, groups of compounds which tend to co-vary highly together are put closer together in the figure. This leads to pockets of red in the figure showing how cluster groups are starting to form and show

similarities between these compounds. And in some cases, groups of compounds which strongly anti-correlate. However, at this scale, it is initially difficult to glean too much information from these groups outside of number of groups. It is notable though that the largest group of red, highly correlating species have the tendency to be smaller, polar compounds which group together.

It is seen in figure 5.2, that gaps in the co-variances are seen between certain compounds. This is due to missing data, whether the signal for an individual compound being below the limit of detection for example or no compound being present due to the source of the compound not being as prevalent (such as a change in wind direction leading to the introduction of a different air mass). These gaps in data mean that a good correlation cannot be established for these compounds. As a measure of data control, a threshold of 40% of data points was used to ensure good quality of data in the clustering analysis and ensure that sufficient data was entered before full clustering.

After this filtering, 84 compounds were analysed by HCA. As seen in figure 5.3, the cluster analysis separated the compounds into eleven distinct clusters, termed clusters 1 - 11. The time series and polar plots of each of the groups are seen in figures 5.4 to 5.14.

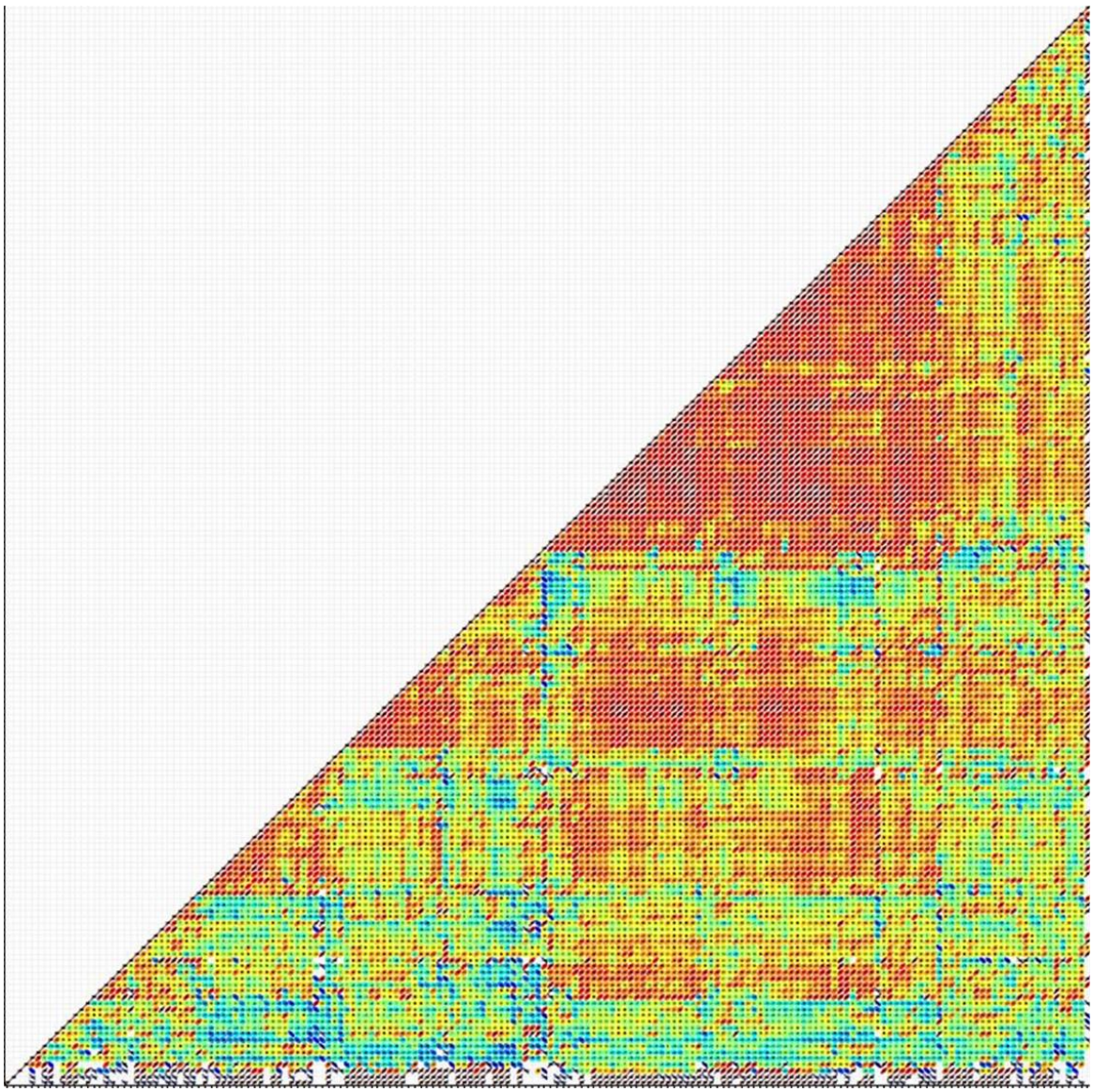

Figure 5.2: Correlations of 191 observed organosulfates, colour-coded by correlation. More red points show higher correlation, bluer points show higher anti-correlation and greener points show no significant correlation between factors. Hierarchical cluster analysis was applied to this cluster plot to group compounds which co-vary highly together to show patterns of high co-variance groups, these groupings show the initial formation of subsets within the heteroatom group and highlight the need for big data analysis.

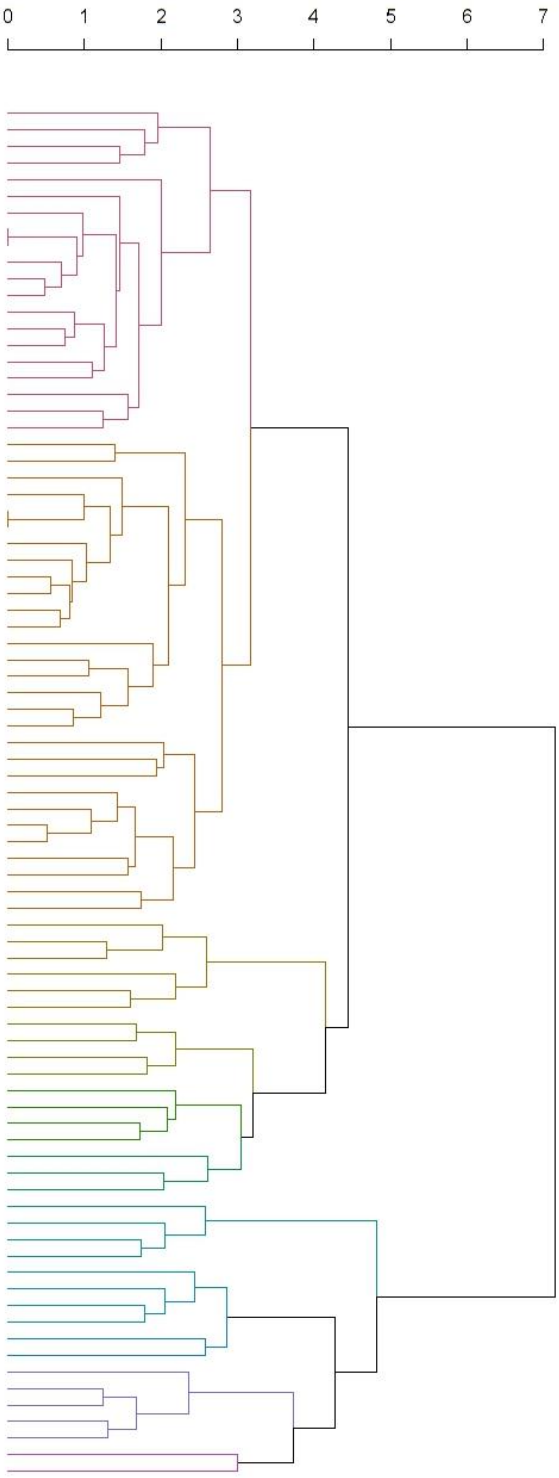

Figure 5.3: Dendrogram displaying the HCA clusters for 84 CHOS compounds, the 11 cluster groups are shown by different colours.

Group 1 consists of 20 compounds and has a higher concentration from southerly winds, with the highest levels coming from a south-by-southwesterly wind. The compounds within group 1 have a high number of C5 and C10 species which have are potentially biogenic in nature due to the higher DBE, this would fit as towards the south are large green patches. The compounds are more likely to also be longer lived and are more able to have long-range travel and become aged as SOA. As such compounds in the air mass present complications to other areas once this air mass leaves the Beijing area.

Group 2 consists of 29 compounds, largely dominated by one compound ($C_5H_{12}SO_4$, constituting over 50% of the total peak area for the group) and most compounds are smaller than C10. Group 2 largely has a point sourcing from the east, with a smaller contribution to the total peak area being more localised to the detection site. The nature and directional sourcing of this group suggests that the group consists of primary emissions from the residential. Group 2 peaks around early afternoon (12:00 - 14:00) during high photo-activity, so this could account for the formation of the other half of the peak area in this group.

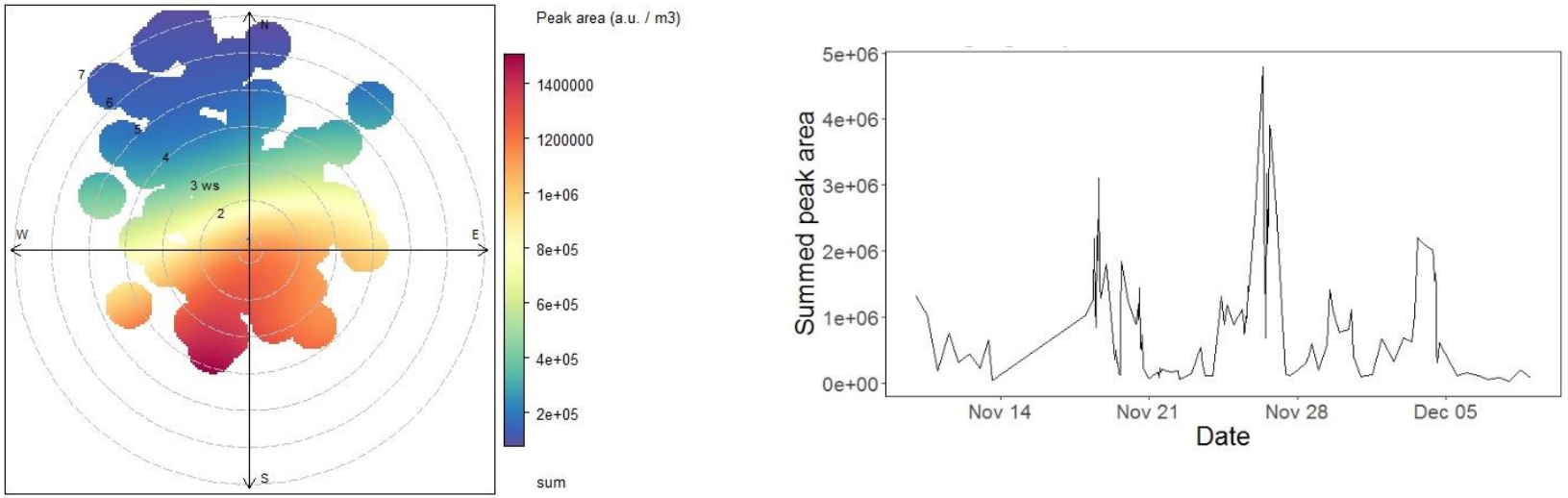

Figure 5.4: Polar plots for each of the cluster group 1 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

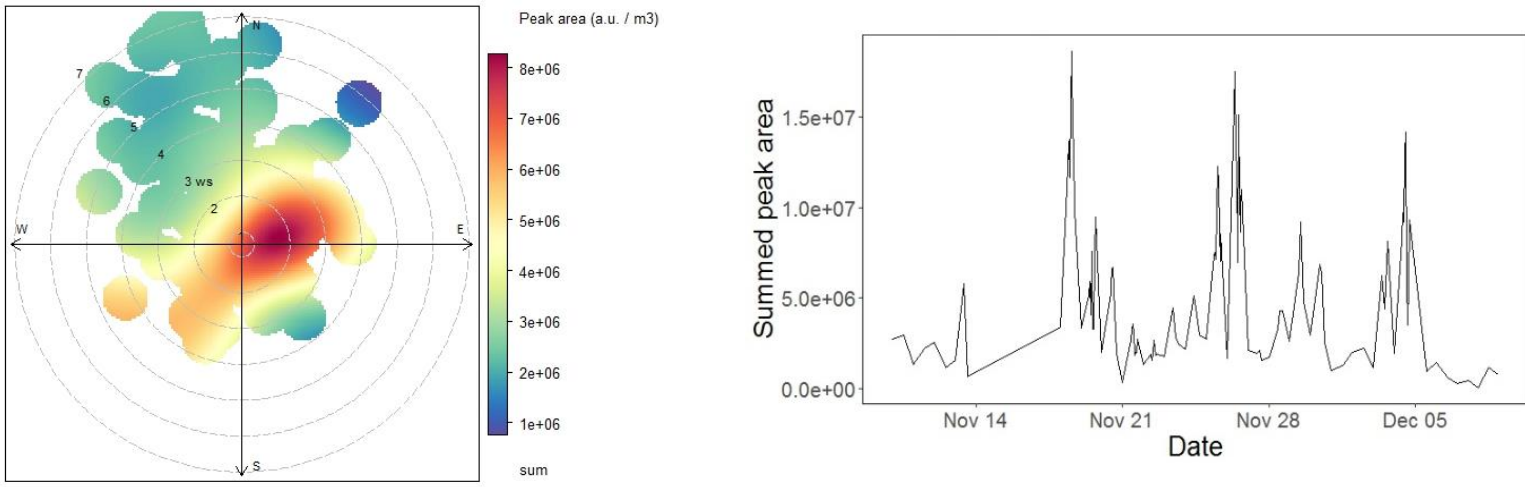

Figure 5.5: Polar plots for each of the cluster group 2 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Group 3 has a very local point sourcing. It is dominated by one compound ($C_{12}H_{26}SO_4$, contributing over 50% of the total peak area) with two further compounds ($C_{10}H_{22}OSO_4$ and $C_8H_{14}O_2SO_4$) contributing almost all of the remaining peak area. Due to the aliphatic nature of these dominant compounds, it can be suggested that they are from local traffic emissions as the sampling site is very close to the 3rd and 4th ring roads of Beijing. Group 4 is composed of four compounds who contribute near equally to this group. Group 4 appears to have a strong point source from the south west and a less source to the east. These compounds were also seen in near the Yangtze river by Wang et al[137], who suggested that as well as the sulfate group, these compounds may also contain hydroxy or carbonyl groups. Due to the proximity of the sourcing and the compounds having also been observed near the Yangtze river, this could be a possible source, suggesting this air-mass is much longer lived and would explain the additional functionality of the compounds in group 4.

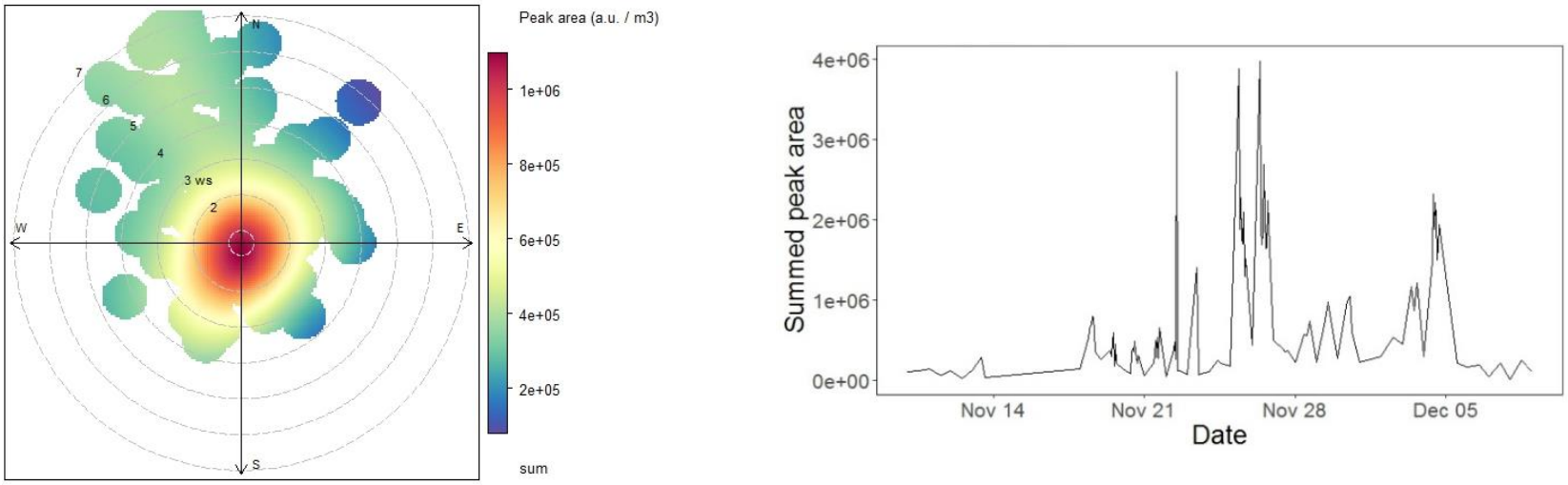

Figure 5.6: Polar plots for each of the cluster group 3 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

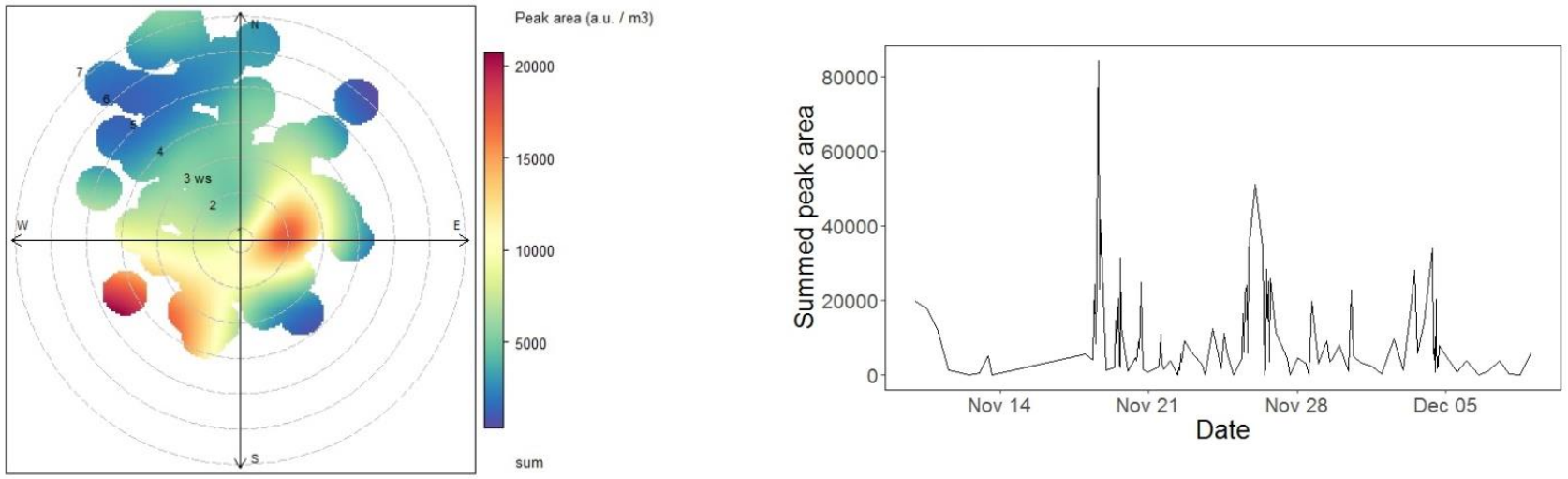

Figure 5.7: Polar plots for each of the cluster group 4 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Group 5 has a similar directional sourcing to group 4, however the point sources are much more defined, have two similarly proportioned sources from the east and south west. Group 5 is dominated by one compound ($C_7H_{16}SO_4$, contributing 75% of the total peak area), due to the low DBE of the compounds, this group appears to consist of aliphatic compounds, potentially with hydroxy functionalisation. Indeed, groups 4 and 5 are grouped very closely in the HCA analysis suggesting they have very similar trending. Based on the time series of both groups, the similarities occur between 19th and 21st of November. This suggests that both groups were present in the first air mass and thus would have similar sourcing, however discrepancies between their formation path ways could be the reason behind their slightly difference sourcing profiles as the oxidants may have different levels at different times.

Group 6 is dominated by two compounds ($C_7H_{14}OSO_4$ and $C_{10}H_{20}OSO_4$, contributing over 90% to the total peak area), the trends of these compounds peak over-night and in the later afternoon (after 16:00) and tend to show very little functionalisation in addition to their sulfate groups. As such, these compounds are likely to be primary emissions from residential emissions, coming from the very densely populated centre of Beijing. As the directional source is rather broad, as opposed to the more point sources of the other groups, this would suggest that the source is not from one region but from all of the residential districts.

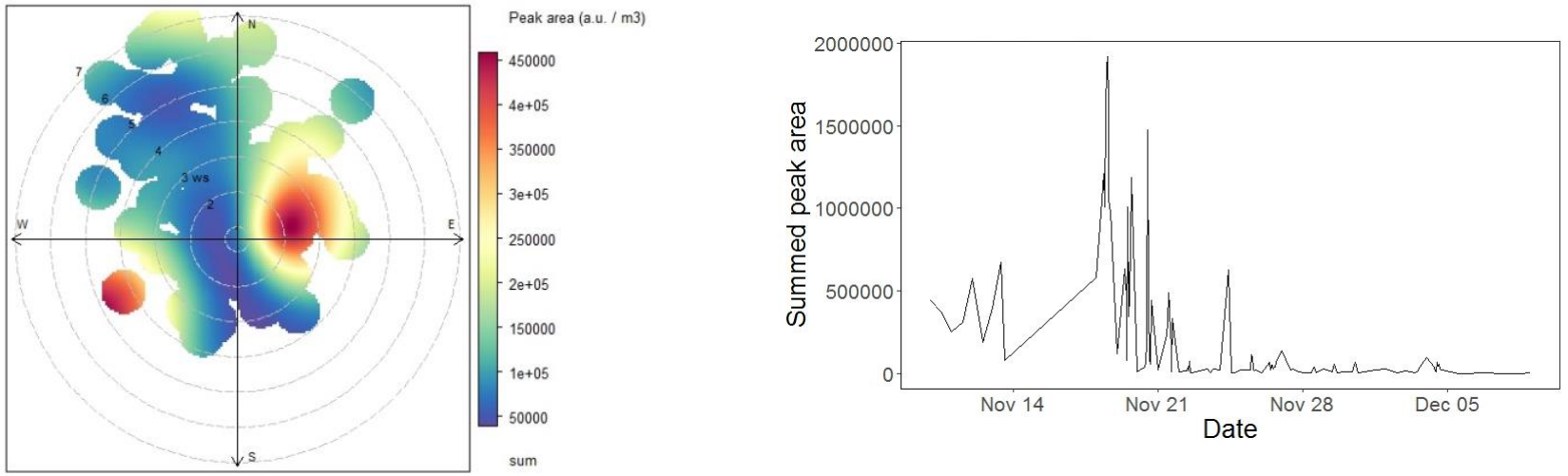

Figure 5.8: Polar plots for each of the cluster group 5 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

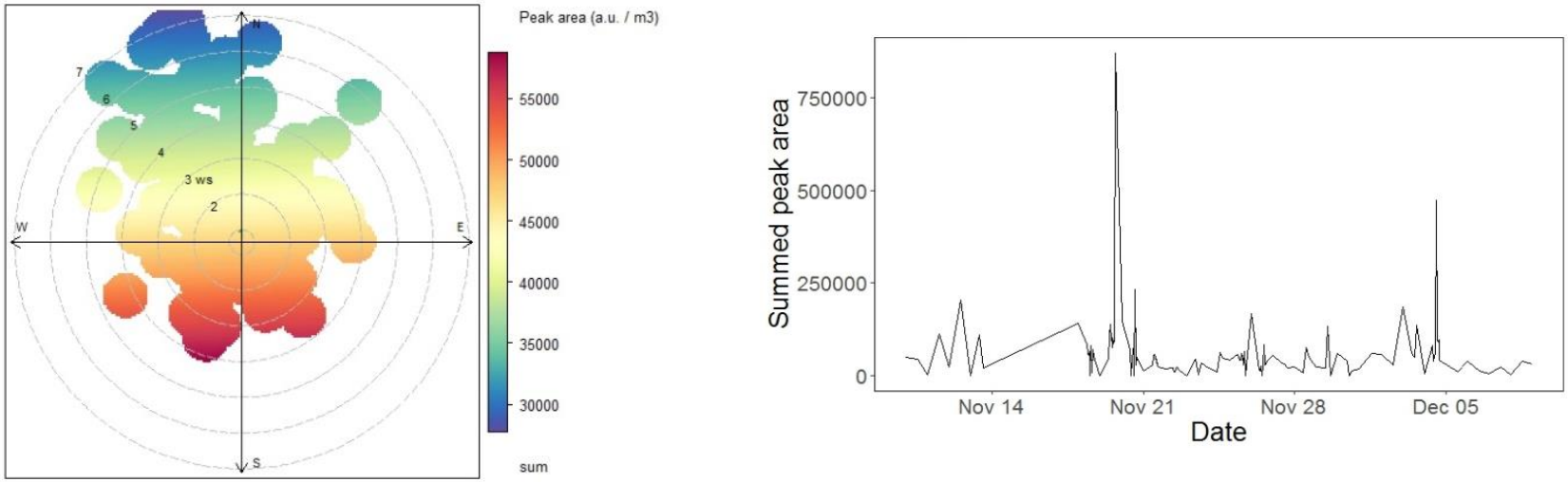

Figure 5.9: Polar plots for each of the cluster group 6 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Group 7 is unlike the previous groups as its source seems to not come from any local source, the directional source seems to come from the north, potentially from the mountainous range towards the north-east Beijing. Group 7 is dominated by one compound ($C_4H_{10}SO_4$, contributing 80% of the total peak area to this cluster group) and is the most abundant between the period of 21st of November to 23rd of November.

Group 8 has the least localised source of all of the cluster groups, with the exception of the very sharp peak on the 19th of November, which accounts for the very high point in the north west of the directional source. This group is almost completely made up from octyl sulfate, a common primary emission from industry and traffic emissions (with octane being an anti-knocking agent in fuels). Due to the high numbers of large and small road across Beijing and the prevalence of smaller personal motorbikes and petrol scooters, the source of octyl sulfate is going to be wide-spread. With the removal of the one anomalously high peak, the source map is more homogenous across directions and wind speeds. Group 9 is another unusual grouping in the cluster analysis and is composed only of $C_5H_{10}OSO_4$, which seems to be sourced regionally from the areas outside of Beijing and has very defined significant peaks and throughs without consistent diurnal cycles. This erratic behaviour is unlike any of the other groups.

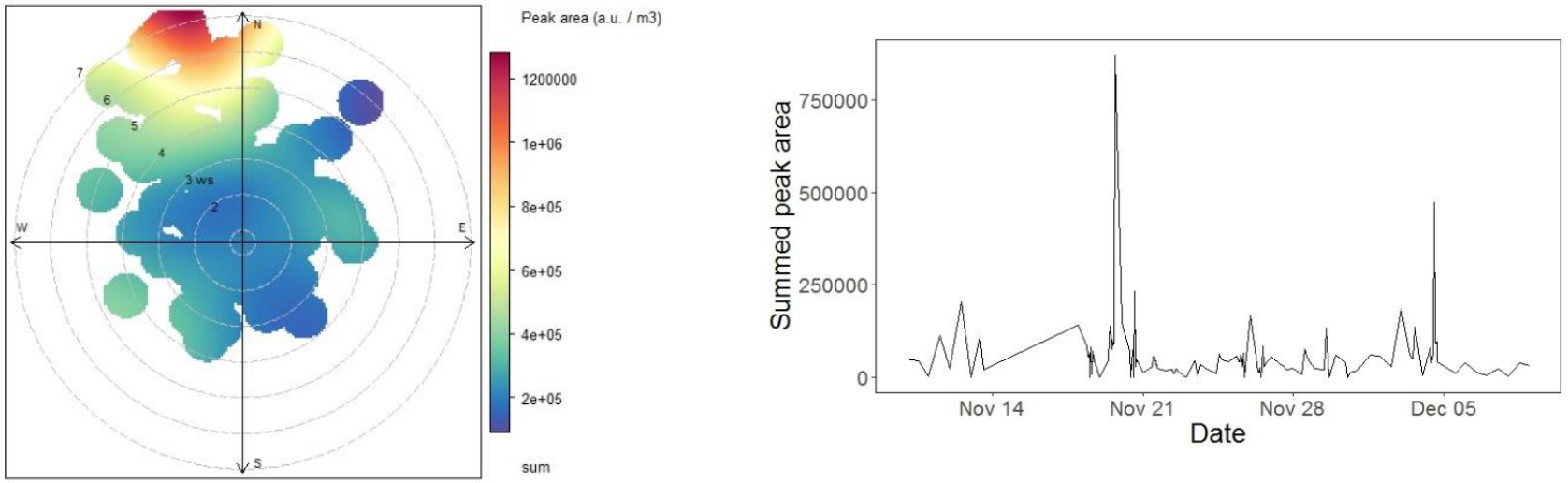

Figure 5.10: Polar plots for each of the cluster group 7 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

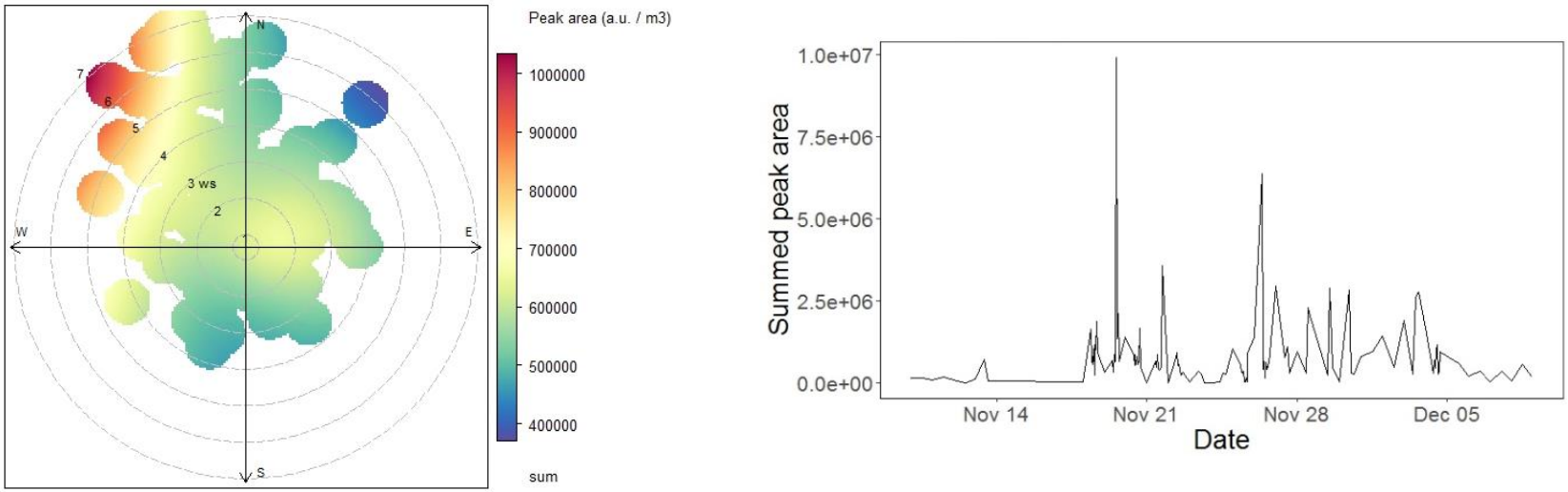

Figure 5.11: Polar plots for each of the cluster group 8 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

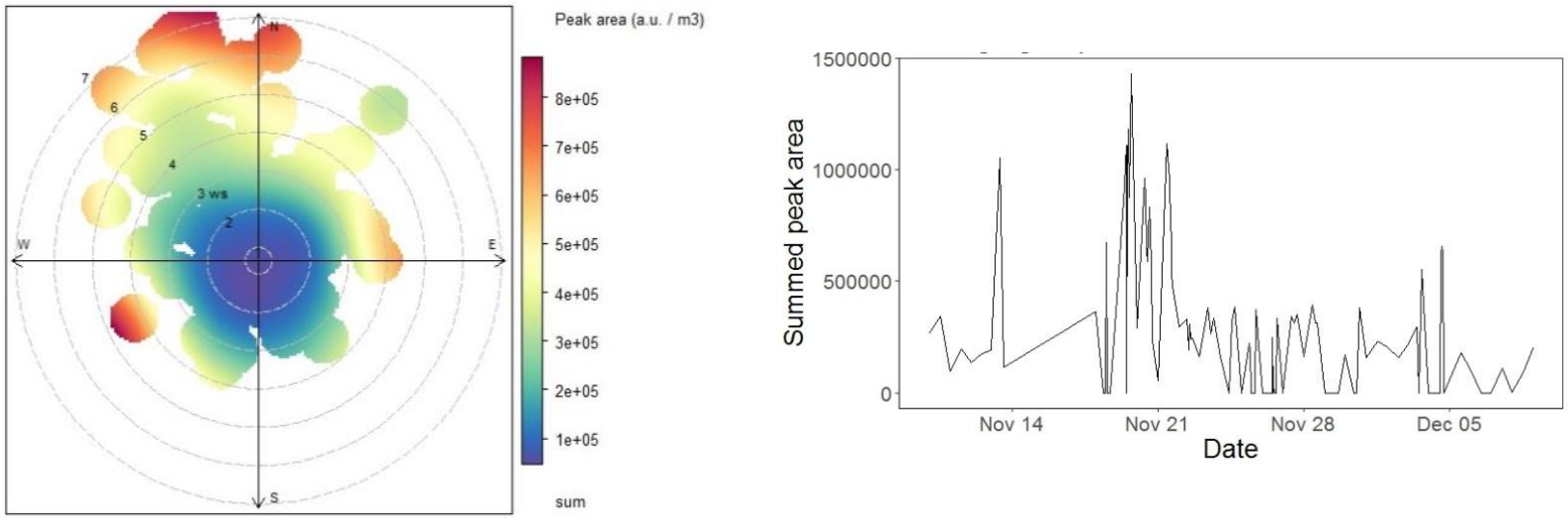

Figure 5.12: Polar plots for each of the cluster group 9 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

Group 10 has two directional sources, during high wind speeds and peaking during the day, giving rise to a defined diurnal cycle. This suggests that precursors are brought in during high wind speeds, are reacted with and further functionalised, during the day to produce the compounds in group 10 and thus, the generation of these compounds is limited by photoactivity. Group 10 is dominated by two compounds, dodecyl sulfate and ethyl sulfate, contributing over 90% of the total peak area of group 10. This would suggest that the precursors are aliphatic compounds which become sulphated as they approach the sampling site. Group 11 consists of two compounds, $C_{11}H_{22}OSO_4$ and $C_{12}H_{26}OSO_4$, with the former contributing slightly more to the group. The source seems to be largely local and slightly to the west.

The separation of sources demonstrates how each source contributes to the overall organic composition and how the contribution from each of these sources varies over time. It can be seen that differing air masses produce a different set of compounds within the $PM_{2.5}$ which lead to differing effects on the populous. The usefulness of the high temporal resolution off-line sampling can also be seen. Without this hourly resolution, sources for these groups would be harder to assign, for example group 6 compounds peaking in the afternoon lend credence to the assignment of residential sourcing as the peaks coincide with people returning from work.

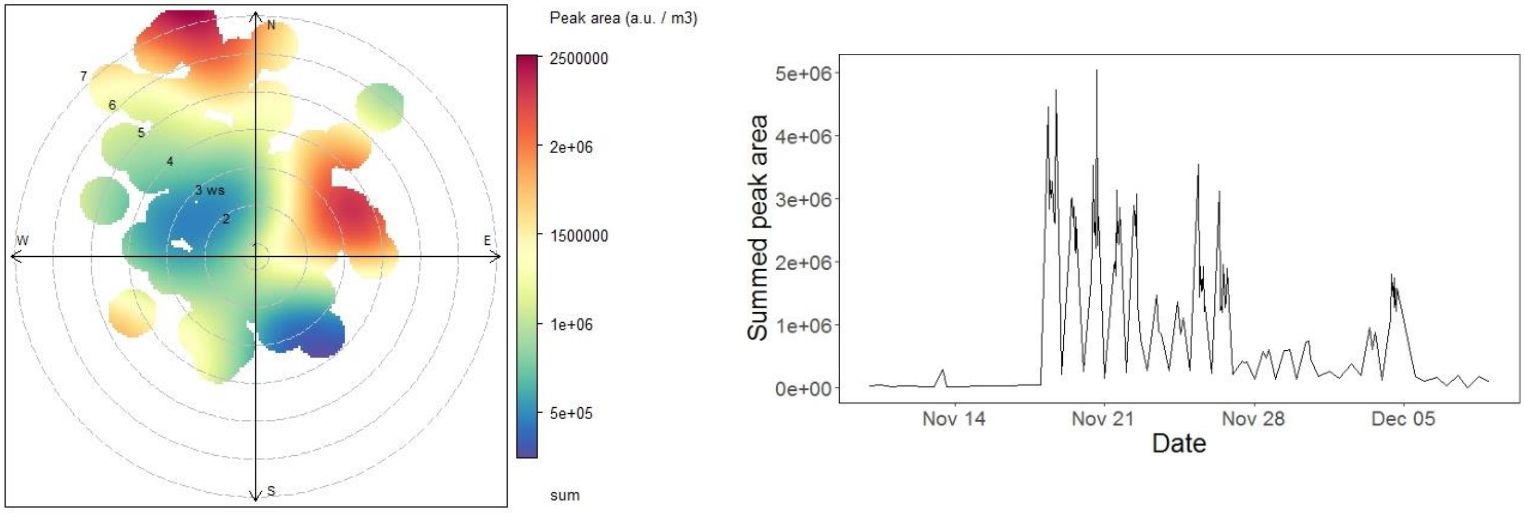

Figure 5.13: Polar plots for each of the cluster group 10 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

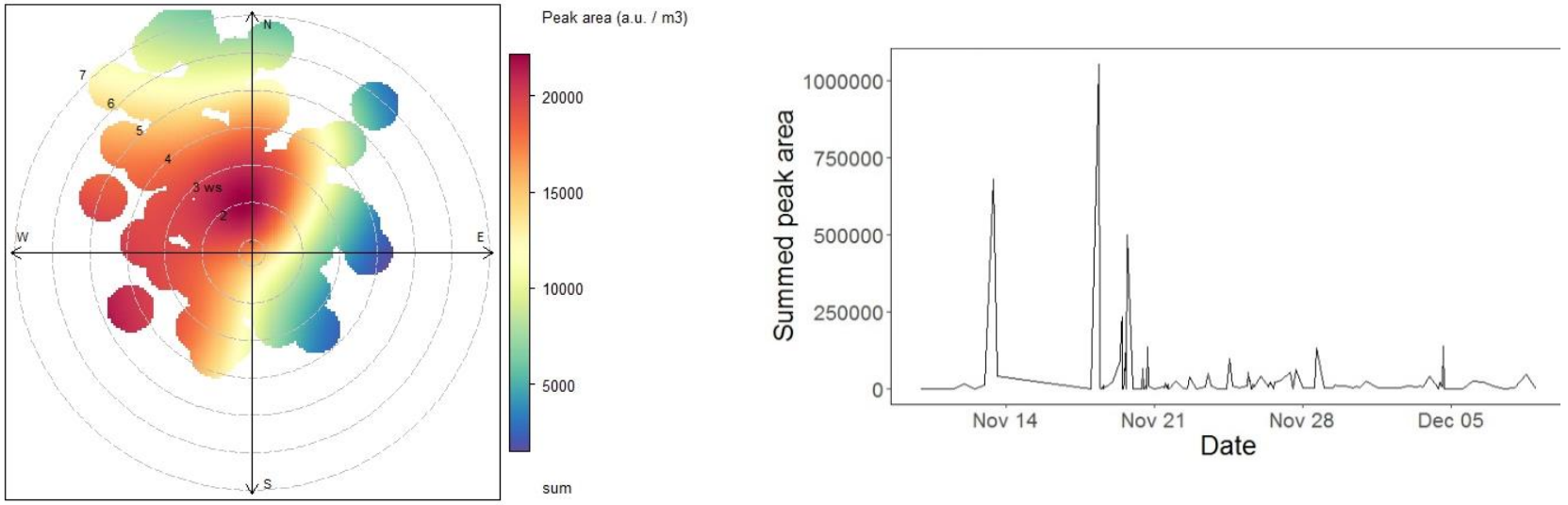

Figure 5.14: Polar plots for each of the cluster group 11 showing the summed peak area at different meteorological conditions (left) and time series of summed peak area for each group (right).

The contribution to the total CHOS from the 11 clusters across the winter campaign is shown in figures 5.15 and 5.16. Two distinct periods can be observed during the winter campaign; the first, prior to 21st November and the second after 21st November. This first period can be defined by the higher contribution to the total CHOS from groups 3 and 5 and a lower contribution from group 8. The second period is characterised as a higher contribution from groups 8 and 10 and very low contribution from groups 5 and 6. In both periods, the contribution from group 2 is largely consistent, apart from small scale variation across the days, suggesting the overall levels of these locally sourced compounds are the largest contributors of OS and unaffected by meteorological conditions.

Figure 5.15 demonstrates that during the peaks of the overall CHOS summed peak area, groups 2 and 8 co-vary very well. Due to the relationship with the total CHOS, octyl sulfate (the dominant compound in group 8) and the long chain aliphatic compounds in group 2; the overall trends of total CHOS are strongly controlled by longer chain, more aliphatic components, hence have strong sourcing from residential, traffic and industrial; largely sources which use fossil fuels. During the period of very low total CHOS, group 10 dominates, however, this is more of a case where the levels of the other groups are significantly reduced, rather than the levels of group 10 being significantly more abundant.

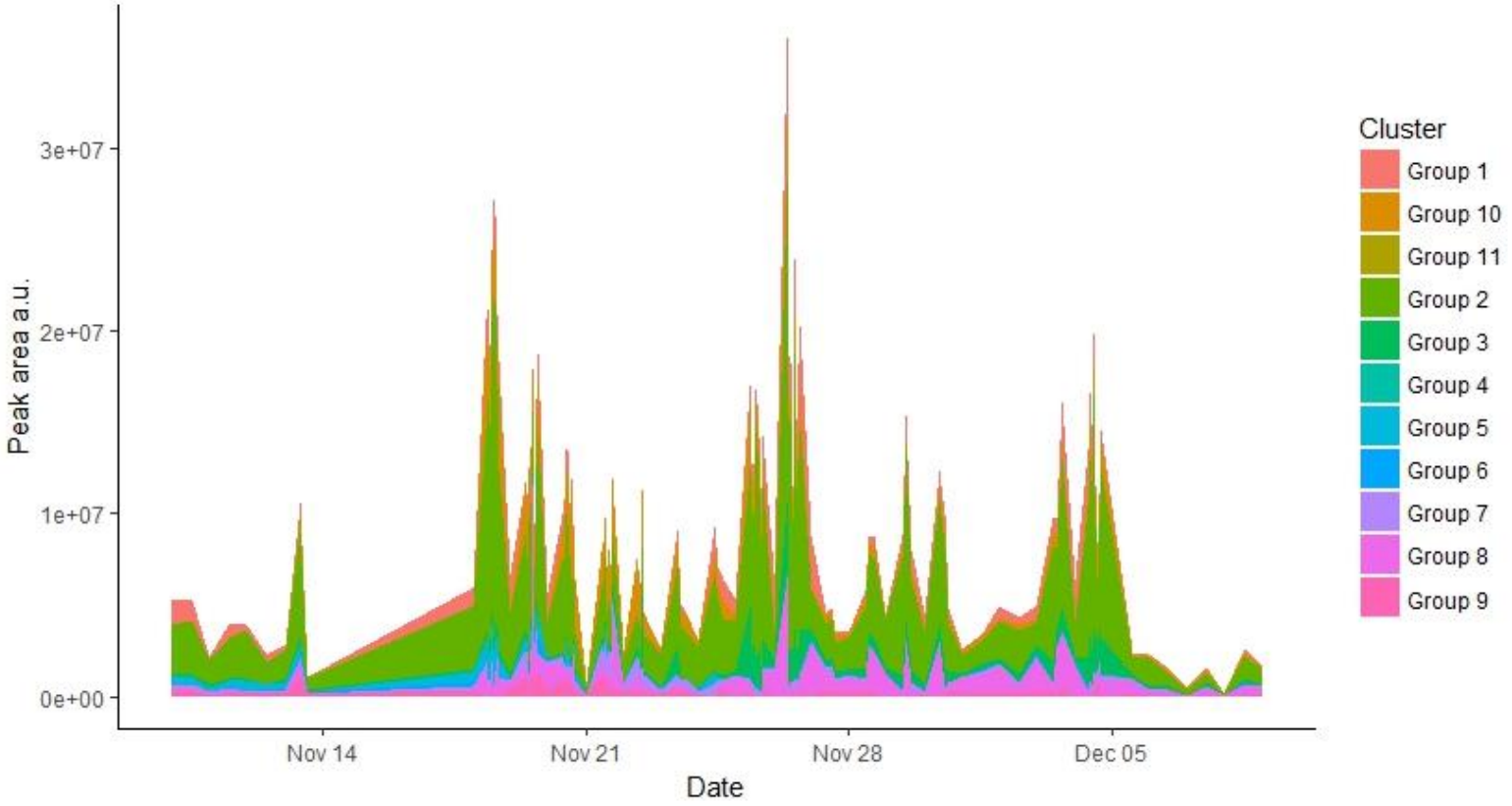

Figure 5.15: Variability of contribution from cluster groups to total CHOS composition over winter campaign as a proportion of total peak area.

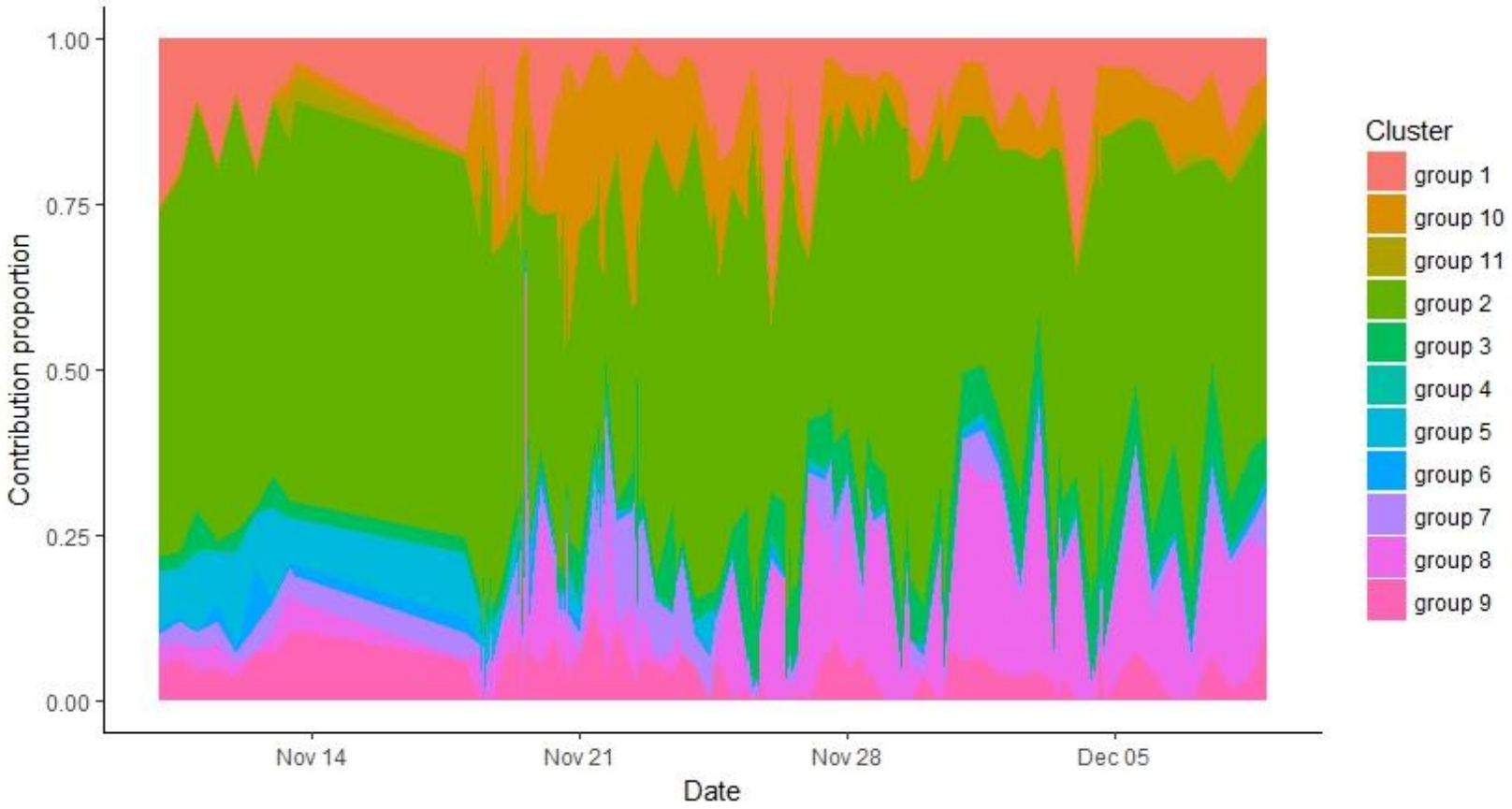

Figure 5.16: Normalised contribution from cluster groups to total CHOS composition over winter campaign as a proportion of total peak area.

5.3.4 Comparisons with auxiliary data

Each of the groups was compared against AMS factors and gas-phase inorganics; whilst not all of the groups showed significant correlation with all of the auxiliary factors, some interesting correlations are observed for some of the groups.

Group 1 has some interesting correlations with the auxiliary data; figure 5.17 shows, as expected, the abundance of group 1 components correlate well with SO_2 levels and anti-correlate with O_3 levels. However, the abundance of group 1 correlates minorly with humidity levels. This would suggest that these compounds are not directly sulphated, rather that they are further oxidised existing sulfates This is because, as discussed previously in this thesis, higher humidity inhibits the production of organosulfates as the presence of ions water interferes with the sulfation process. This, combined with the directionality observed earlier would suggest that the compounds are longer-lived sulfates, being further oxidised as they are transported to Beijing rather than being directly emitted sulfates or compounds being sulfonated in the Beijing area.

The idea that these are longer-lived species being oxidised is lent further credence by the good correlation of group 1 compounds with both less oxidised oxygenated organic aerosol and more oxidised oxygenated organic aerosol (LO-OOA and MO-OOA) AMS factors suggesting these are very oxidised components, rather than freshy or directly emitted species.

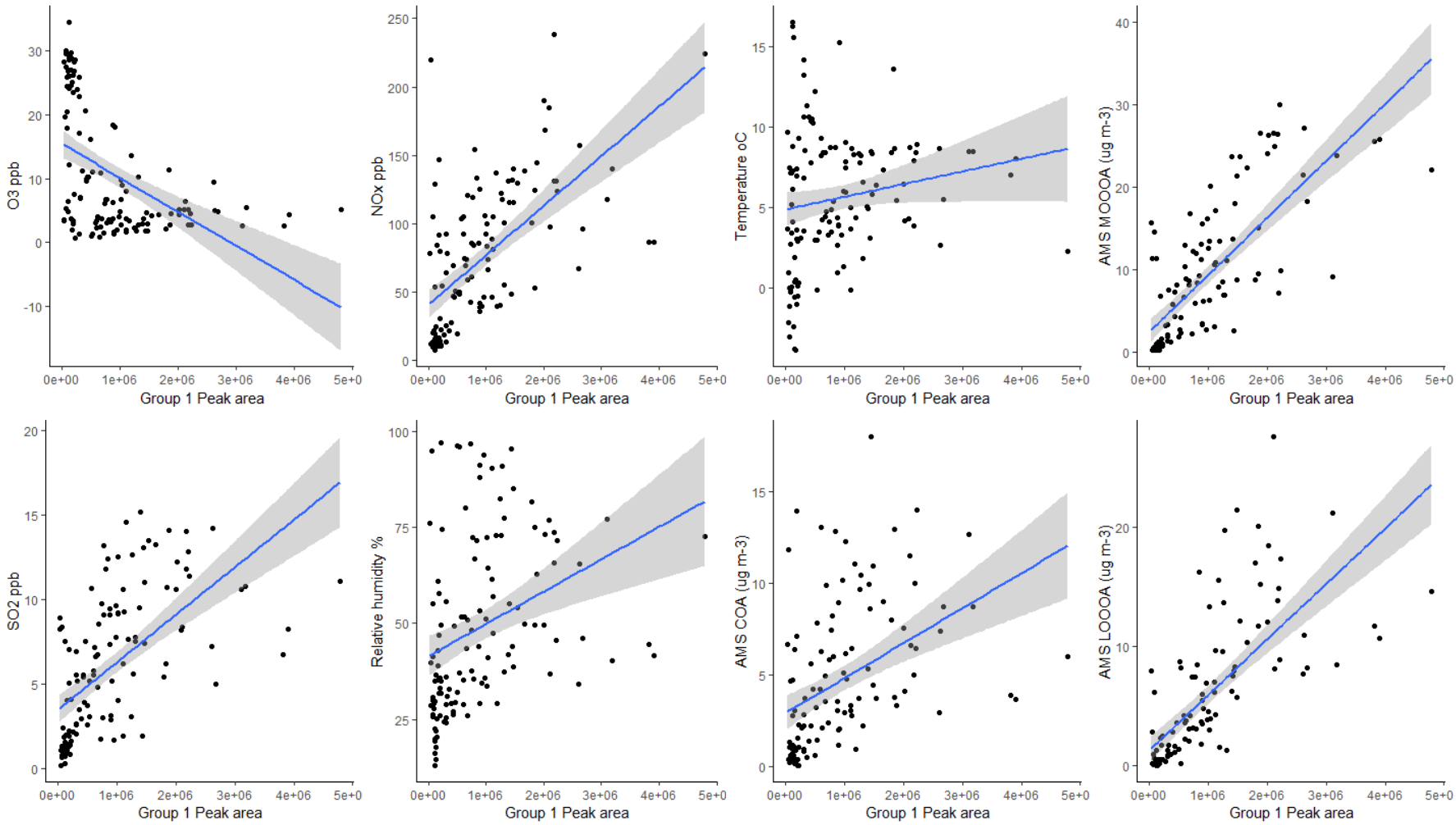

Figure 5.17: Notable correlations between AMS and gas phase oxidants with group 1 averaged peak area at each time point.

Like group 1, group 2 has positive correlation with SO_2 and anti-correlation with O_3 as expected. Figure 5.18 shows that the group seems to have a narrower, better correlation with LO-OOA than MO-OOA, this is probably due to the dominant compound being octyl sulfate and the other contributors are primary emissions and being oxidised at source. This is based on the peaks of abundance occurring around the highest solar activity and the source apportionment being more localised. These factors would all point to the group being locally emitted and oxidised locally, leading them to correlated well with the LO-OOA AMS factor.

Group 2 components have a small, broad, correlation with humidity; however since this is such a poor correlation, it may just be a few anomalous points leading the trendline astray rather an indicative of unusual behaviour. Group 2 also shows correlation with cooking sourced aerosol (COA), based on its local sourcing and correlation with COA, this could indicate this group be dominated by domestic emissions, such as cleaning surfactants and cooking emissions.

Figure 5.19 shows that the correlation of auxiliary data with group 3 abundances seem mostly dominated by a few points, whilst the majority of the other points appear to show little correlation. As with the other two groups discussed, group 3 also shows correlation with SO_2 and anti-correlation with O_3 along with a minor preference to lower humidity (as would be expected by more directly sulphated species). As this group does not significantly correlate with any of the more oxidised AMS factors and the localised sourcing of the emissions, this would also lend credence to the previous suggestion that this factor is perhaps dominated by traffic emissions as these would be locally sourced and have not had much time to become further oxidised.

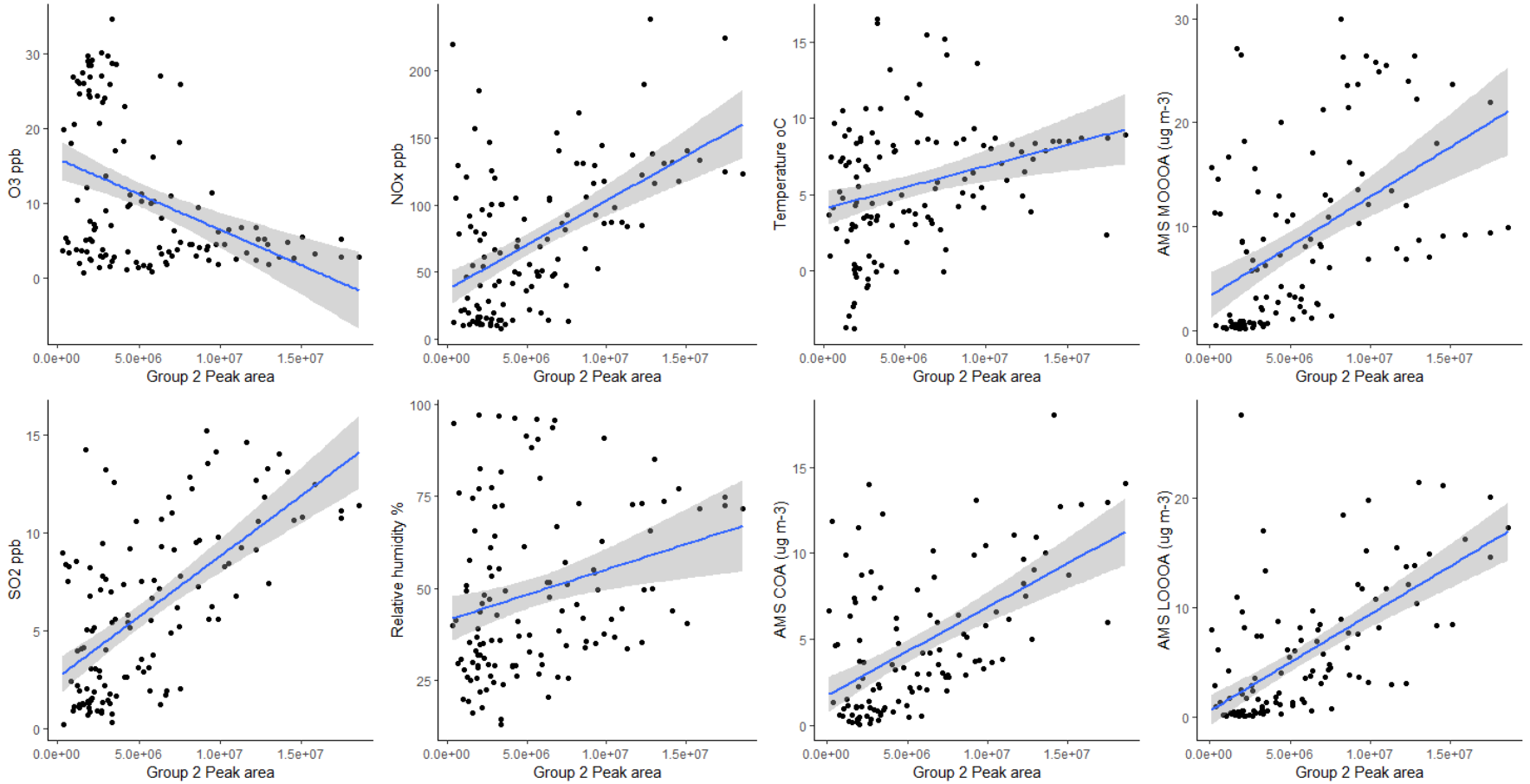

Figure 5.18: Notable correlations between AMS and gas phase oxidants with group 2 averaged peak area at each time point.

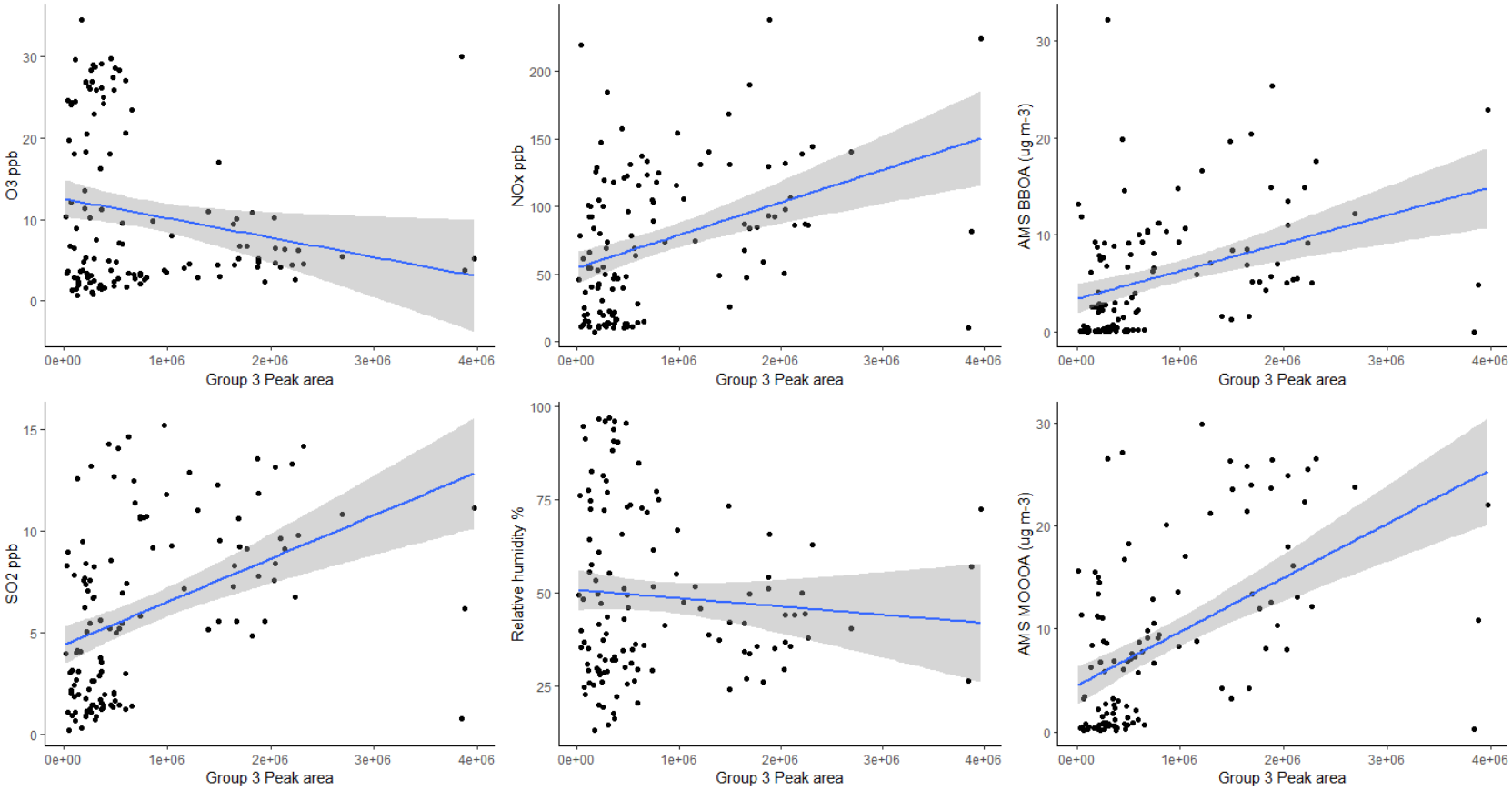

Figure 5.19: Notable correlations between AMS and gas phase oxidants with group 3 averaged peak area at each time point.

Group 4 does not have strong correlations with the auxilliary data, it is clear that a few points dominate the whole trend, as these points are when the peak area is the highest, they correlate to the very sharp peaks seen in the time-series. Interestingly, the compounds do not correlate to either of the LO-OOA or MO-OOA factors, which brings doubt upon the suggestion the compounds in group 4 are longer lived, highly oxidised species; however, the elemental formulae of the compounds would offer very little other options for what they are and due to the compounds being observed in the region the aim-mass is seen to originate from, it is still a strong possibility.

Group 5 correlates minorly with SO_2 and anti-correlates with O_3, as is the trend with organosulfate groups, and seems to correlate with both temperature and LO-OOA. Based on the low DBE values for the components of this group, this suggests that these are more volatile aliphatic compounds which are emitted when temperatures increase and are oxidised locally. The caveat to this is that group 5 is dominated by one compound ($C_7H_{16}SO_4$, contributing nearly 75% of the average peak area), this one compound would exhibit the increased abundance from volatiles in higher temperatures. The remaining compounds in the group may not correlate highly with any of the auxiliary factors and the high abundance peaks (which would account for a large amount of the average contribution) would drive the apparent correlations.

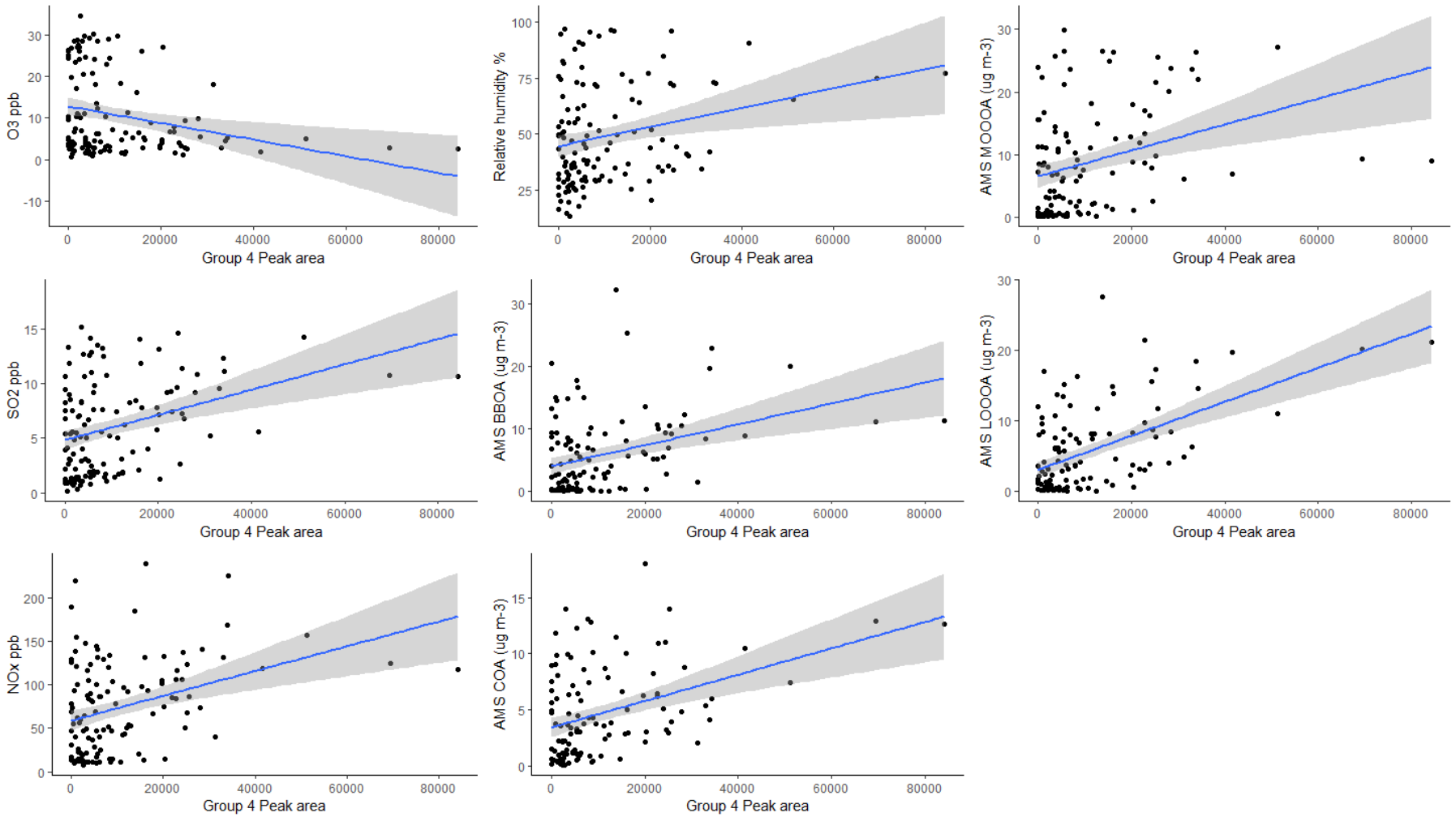

Figure 5.20: Notable correlations between AMS and gas phase oxidants with group 4 averaged peak area at each time point.

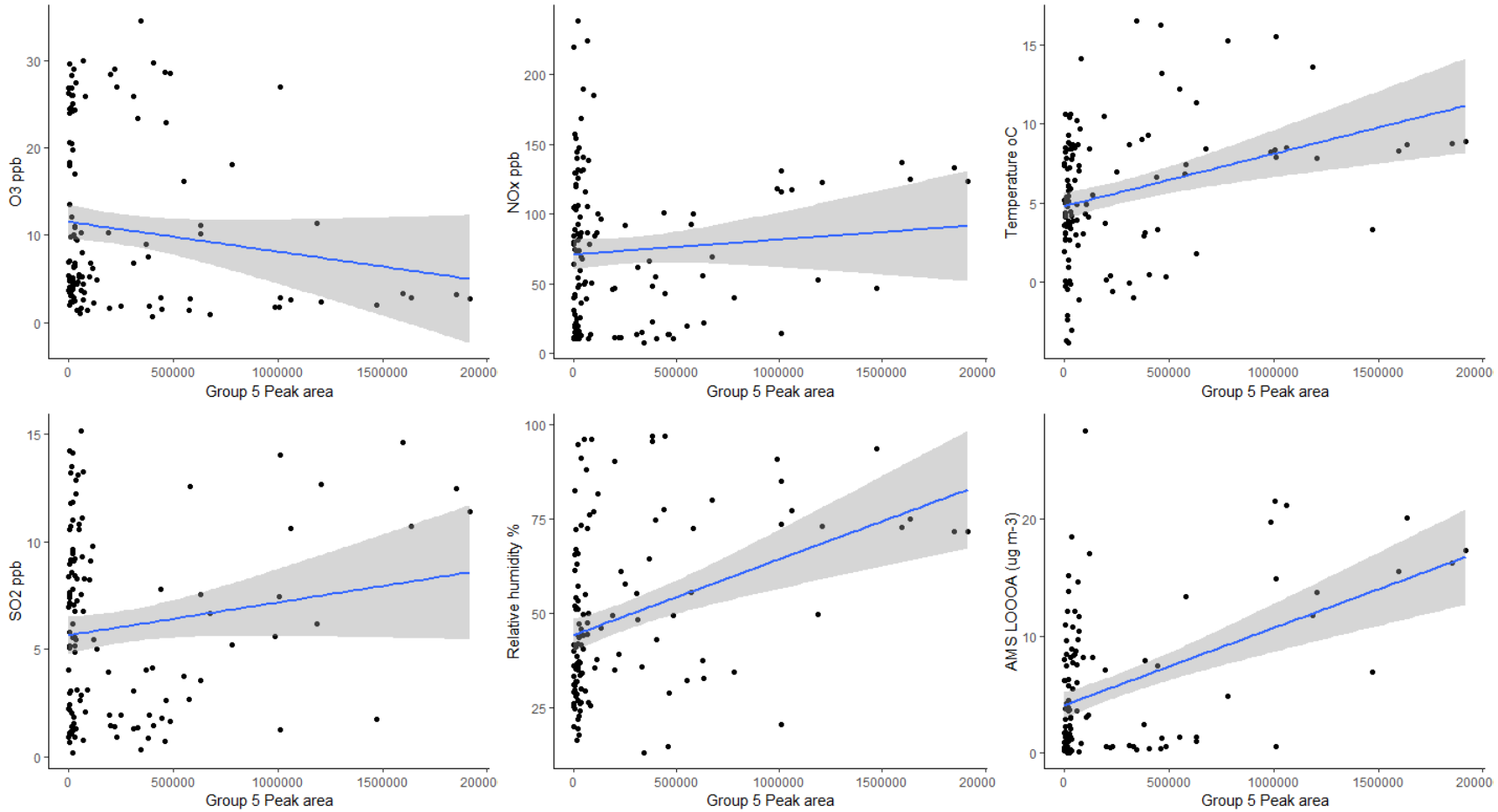

Figure 5.21: Notable correlations between AMS and gas phase oxidants with group 5 averaged peak area at each time point.

Group 7 behaves the most differently to the other groups, having correlation with O_3 and anti-correlation with SO_2; however has strong anti-correlation with humidity and temperature, as seen in figure 5.22. Based on the source direction and the prevalence of the group during very high wind speeds, it would seem that the components in this group are very long lived and longer range transported to the sampling site. This would perhaps explain the higher abundances with lower temperatures; as the temperture drops, the atmospheric layers contract, bringing the components of group 7 closer to the sampling site at ground level.

It can be reasonably implied that the source of this group is not local combustion, as there is a strong anti-correlation with coal-combution organic aerosol (CCOA) and you would expect if it was localised heating, there to be a strong correlation with CCOA as temperatures dropped. In fact, the opposite is true.

As seen in figure 5.23, group 9 anti-correlates with most auxillary factors, except ozone, which correlates quite well with the compounds of group 9. Similarly to the compounds which comprise group 8, these compounds appear to be longer ranged transported species, and do not correlate with coal-combustion organic aerosol or bio-mass burning organic aerosol (commonly used heating mechanisms in China). Instead are transported in from further afar; however, due to the anti-correlation which occurs between the AMS factor LO-OOA and compounds comprising group 9.

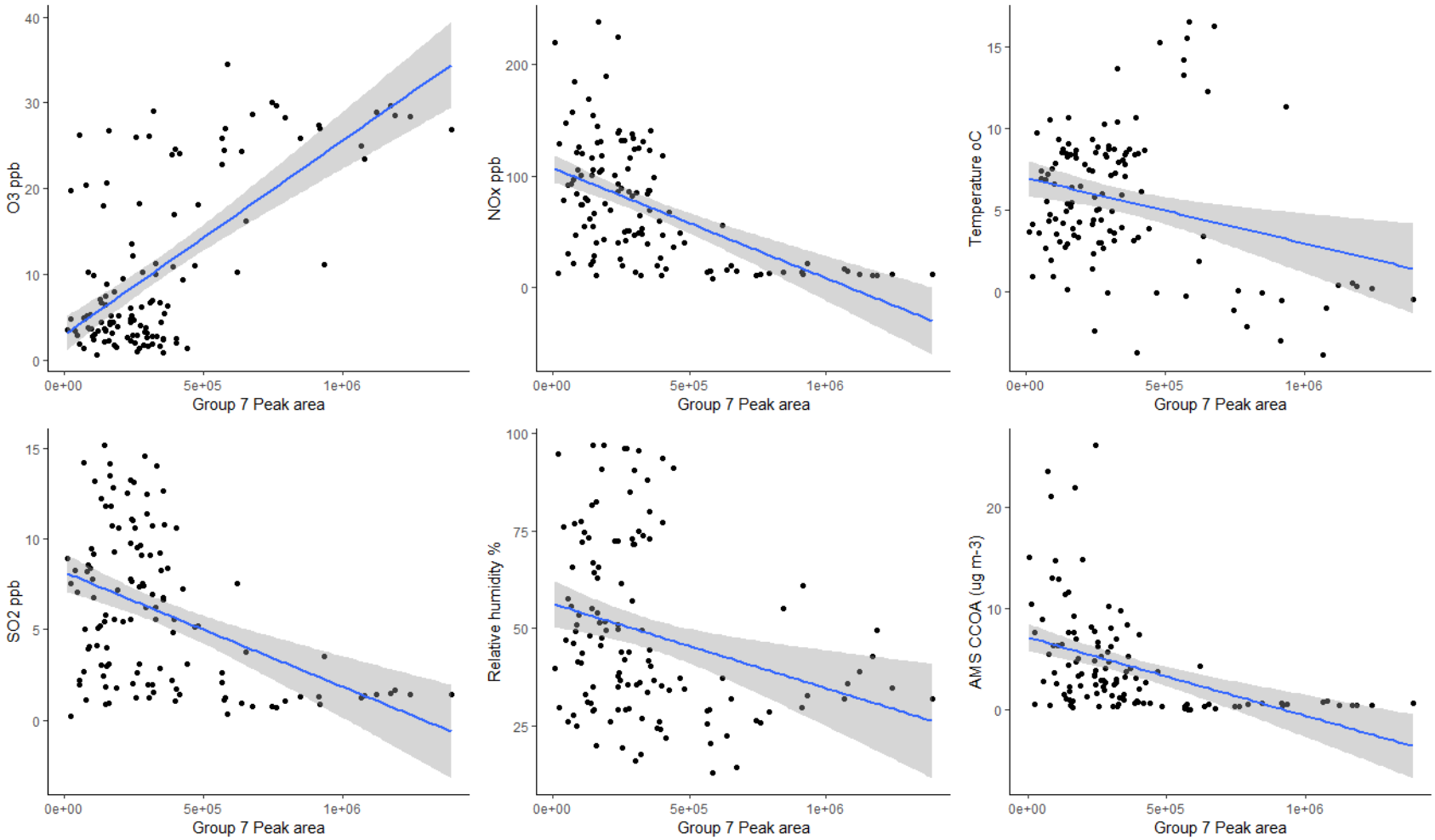

Figure 5.22: Notable correlations between AMS and gas phase oxidants with group 7 averaged peak area at each time point.

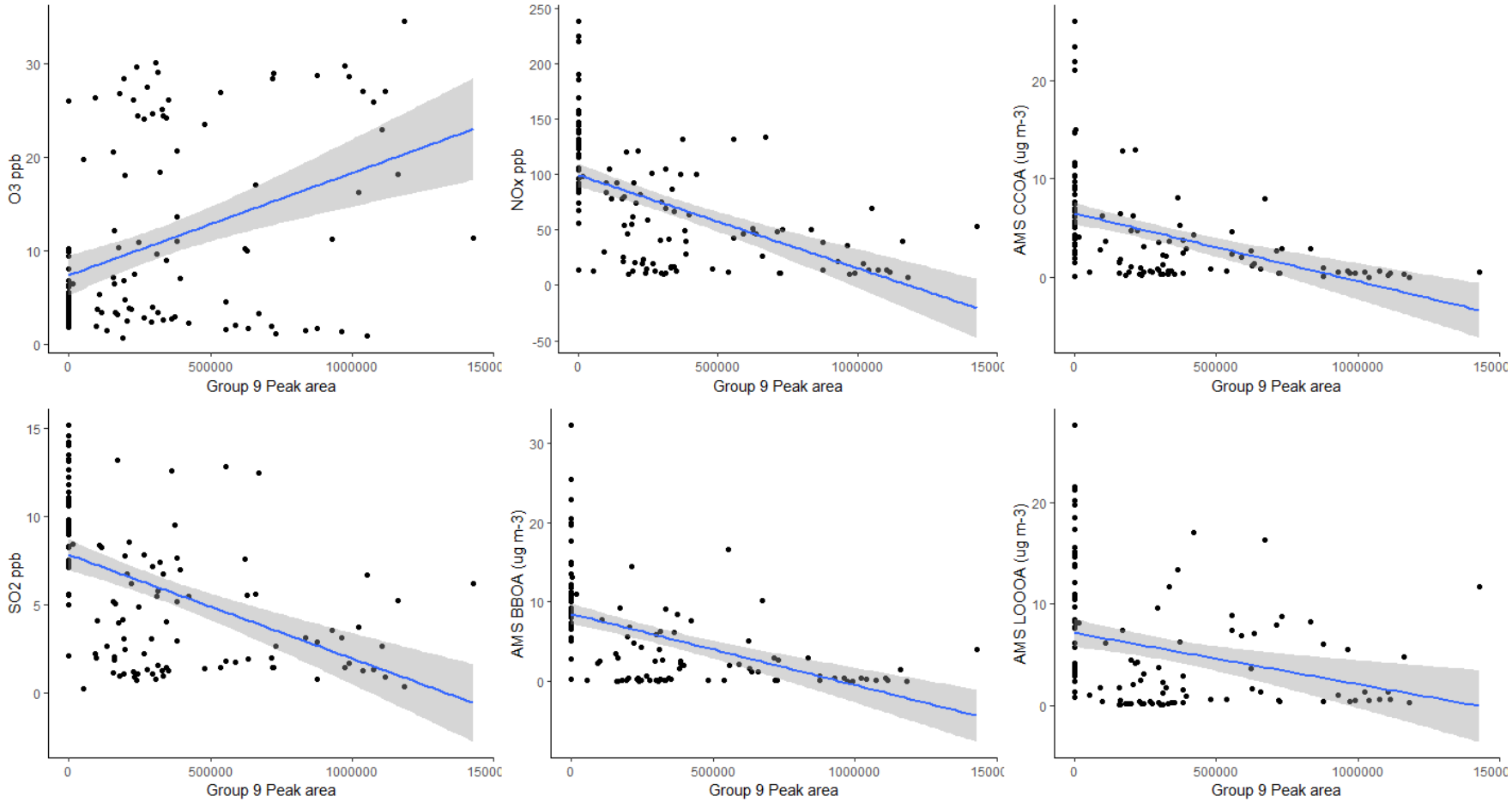

Figure 5.23: Notable correlations between AMS and gas phase oxidants with group 9 averaged peak area at each time point.

The final group which shows interesting correlations is group 10. Correlating more with LO-OOA and anti-correlating with MO-OOA, this matches with what the time-series and directional sourcing show; longer ranged compounds coming in during high wind speeds and producing very high peaks, then are oxidised during the peak solar activity, leading to a sharp reduction in abundance. This matches what is observed in the comparisons with AMS factors; as longer ranged, oxidised species come in and then when they are oxidised into a different compound, the abundance is less. As opposed to a source species being locally oxidised to produce the species being investigated (in this case, would correlate well with MO-OOA). It should be noted however that a large portion of the average peak area in group 10 is from dodecyl and ethyl sulfate and these compounds would behave differently to longer range transported species and lessen the correlations seen from the oxidised species.

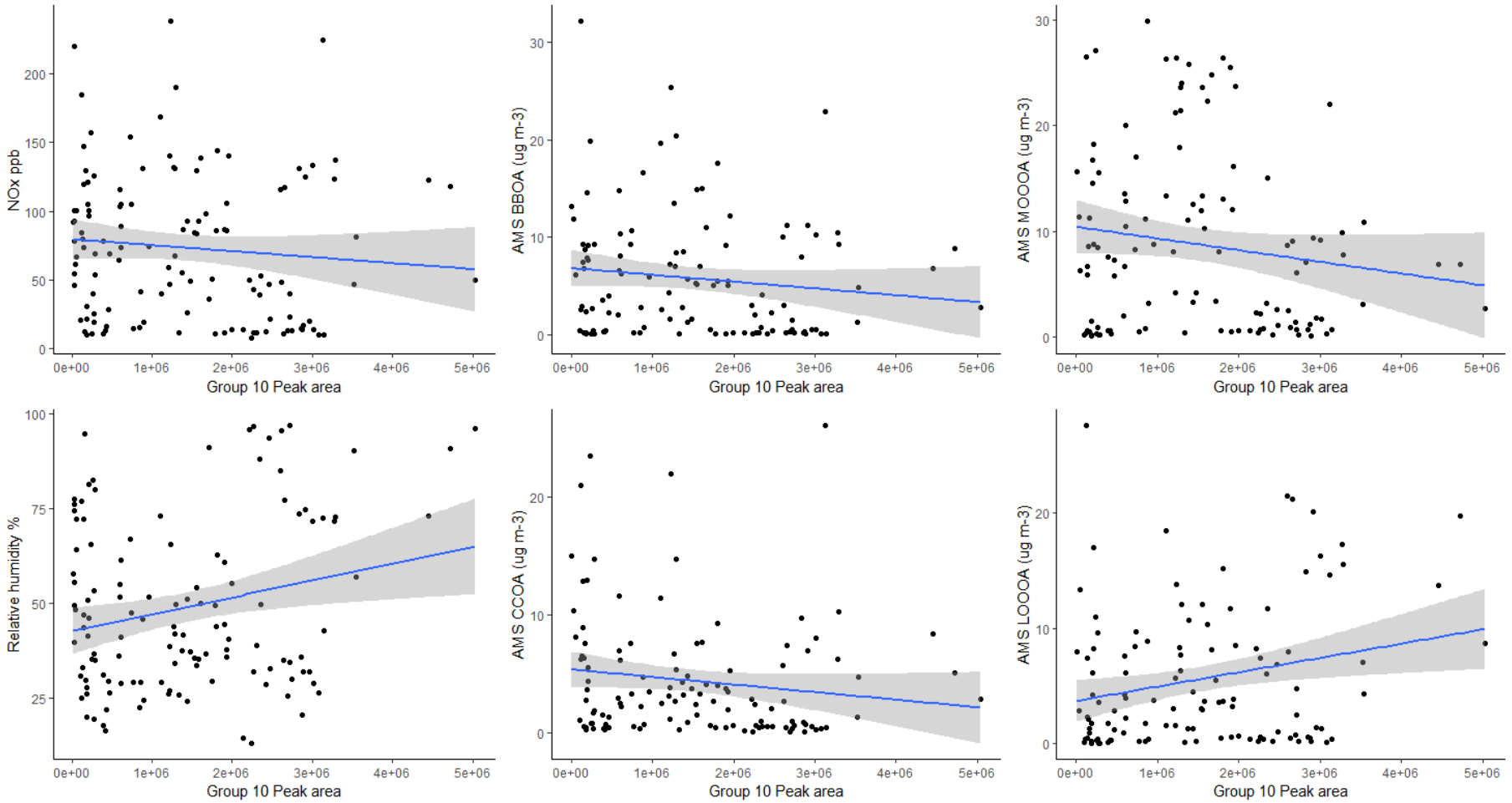

Figure 5.24: Notable correlations between AMS and gas phase oxidants with group 10 averaged peak area at each time point.

5.3.5 Important CHOS species

It is clear from the cluster analysis that one of the most important organosulfates is octyl sulfate, a common surfactant used for many fields such as cleaning agents. The broad sourcing of this compound is indicative of its wide spread usage. The large peak around the 20th November is the cause of the sharp northerly peak on the sourcing. A lose diurnal cycle is observed in the temporal trends in figure 5.25 suggesting a large amount of this sourcing is more anthropogenic in nature as it follows human activity (either from industrial or domestic sourcing). Similarly, figure 5.26 shows intense peaks during the day and particularly towards the end of sampling during warmer temperatures periods with lower relative humidity (RH). Drier periods are seen to significantly increase the production of organosulfates, particularly in air with high acidity. This highly acidic air is commonly seen near industrial sites, such as the high level of industry to the south of Beijing. Higher southerly winds would carry this acidic, highly sulfonated air into the centre and allow for the larger scale production of organosulfates.

Two notable unknown organosulfates are observed. These two species contain both nitrogen and sulfur and are most likely biogenic tracer compounds. The first of these two species occurs at *m/z* 295, $C_{10}H_{17}NO_3SO_4$ is most likely an α-pinene tracer (shown in figure 5.27). This species has been previously observed in chamber studies[58,169]. This compound follows a much more significant diurnal cycle, peaking around the overnight sample, during low solar activity and photo-oxidation, but increased NOx levels. This leads to the longer-lived tracer compounds from terpenes being oxidised further, however at the time of

writing, it is unclear where this process is the sulfonation of a nitroxy species or the nitrification of sulfates.

The second species observed is at *m/z* 215 and occurs as a peak very early in the chromatography and as result, not very resolved. This is a common occurrence for a number of organosulfates in a reversed-phase chromatography, due to their very high polarity, occurring very close to the dead-volume. As such, while a reversed-phase method is effective for a universal method of detection, its usefulness in a particular sub-set, such as very polar compounds, is limited. Therefore, other methods of separation were employed for better resolution of this sub-set of compounds.

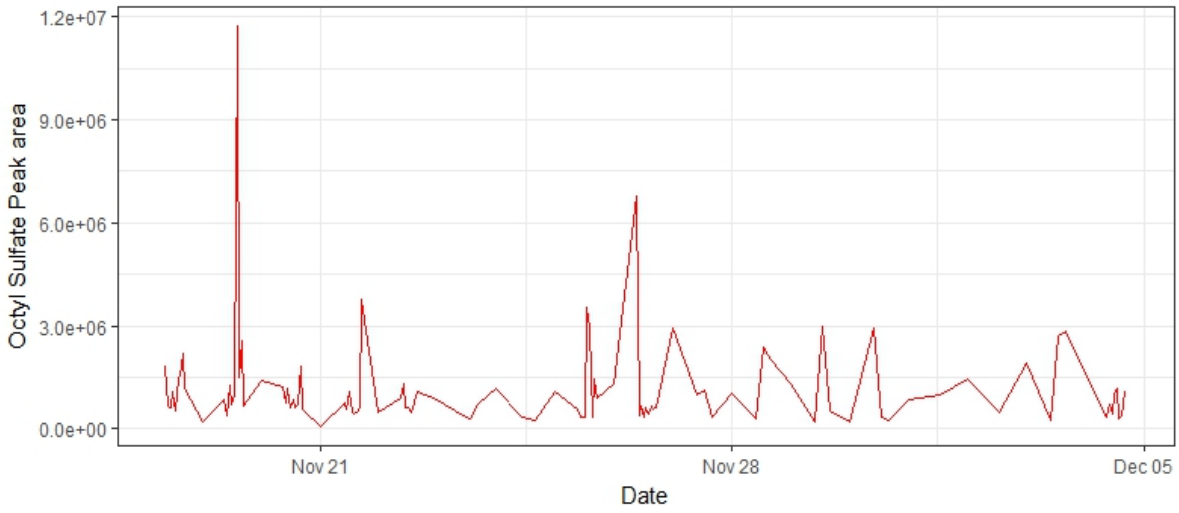

Figure 5.25: Temporal trends of octyl sulfate in peak area response in filter samples during the winter sampling campaign

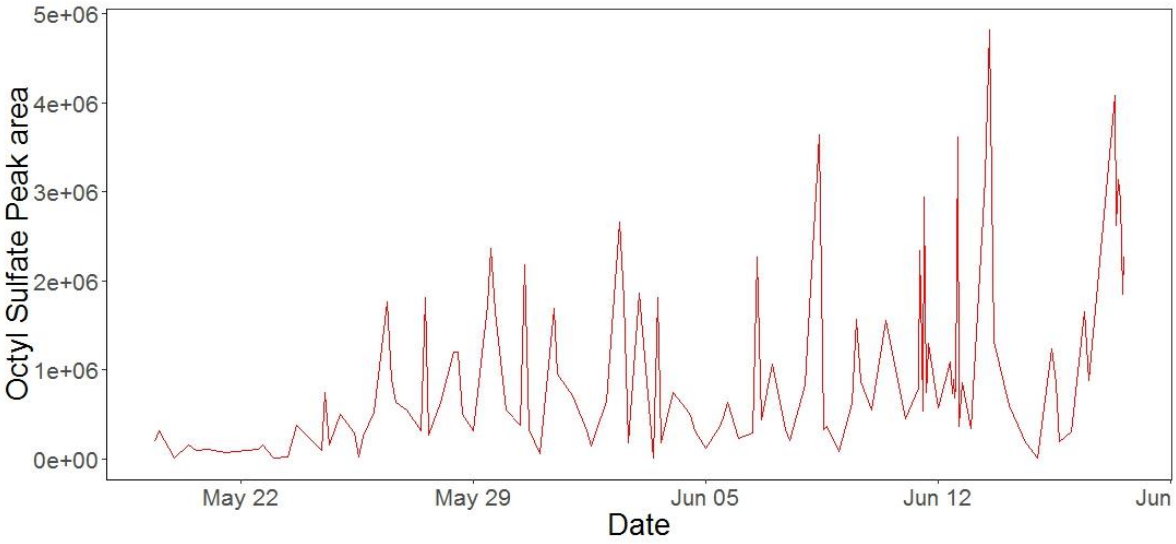

Figure 5.26: Temporal trends of octyl sulfate in peak area response in filter samples during the summer sampling campaign.

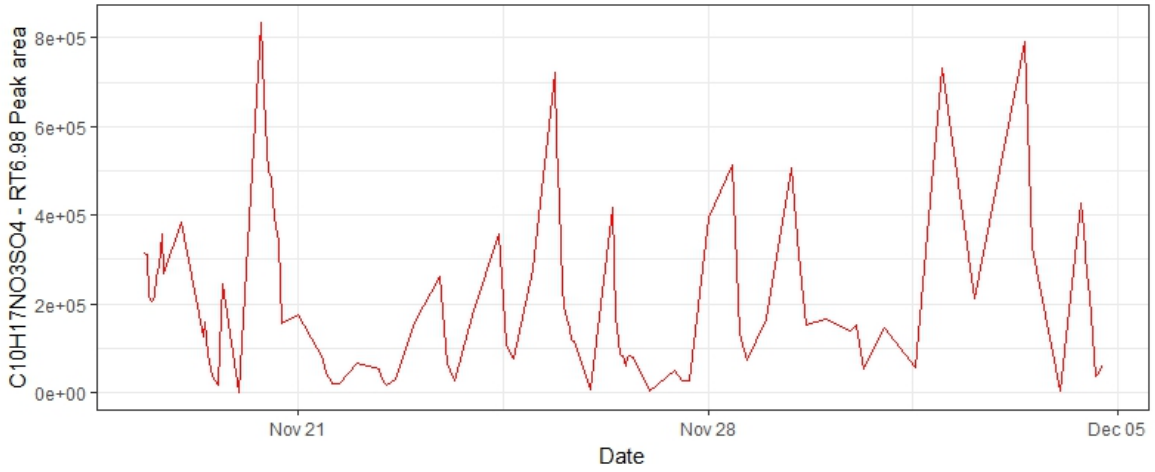

Figure 5.27: Temporal trends of the m/z 295 species in peak area response in filter samples. This species is most likely an α-pinene oxidation tracer.

5.3.6 Comparison of to HILIC method detection

15 samples from the summer campaign were analysed at UNC using the HILIC method. The observed peak at *m/z* 215 was separated into four isomers. Figure 5.28 shows the EIC of a representative sample, these samples usually follow the pattern of a couple of peaks which are less intense eluting early (2.44 mins and 3.85 mins) and two more intense peaks which elute later (4.77 mins and 5.92 mins). In the work of Cui *et al*, these compounds are characterised as organosulfates derived from methyltetrols, including E & Z isomers from the functionalisation of the double bond[66]. However, specifically the confirmation of which species corresponding to which peak has not been performed at the time of writing.

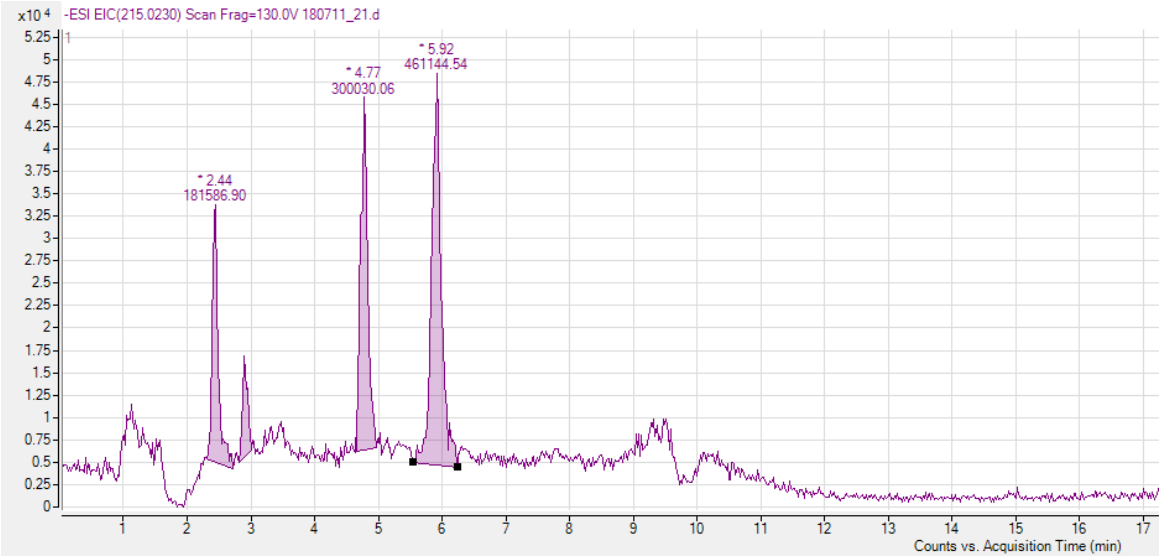

Figure 5.28: EIC chromatogram of the suspected organosulfate at *m/z* 215.0230. The four peaks observed as methyltetrol derived organosulfates are highlighted in purple.

These compounds were quantified and the trends of the four organosulfates plotted to see their variation, these trends can be seen in figures 5.29 and 5.30. It can be seen that in the last few samples, all four compounds co-varied well and had similar trending. However, the first ten samples show differences in the variation of the organosulfate species, these compounds do not co-vary as well. These show that despite the source compound (the methyltetrol species) being the source of all of the organosulfates, they are not necessarily even formed. It follows that the formation is mechanistically controlled, as opposed to source controlled; additionally, the favoured mechanism of organosulfate formation varies across the sub-set of samples, suggesting minor variations on the conditions which allowed for sulfonation of the methyltetrol precursor.

It can be seen across all the samples; the two later eluting compounds are consistently higher than the two peaks which elute earlier. At the time of writing, it is not known whether these compounds are more abundant or have a higher ionisation efficiency than the other two species, as the only available standards are a mix of isomers.

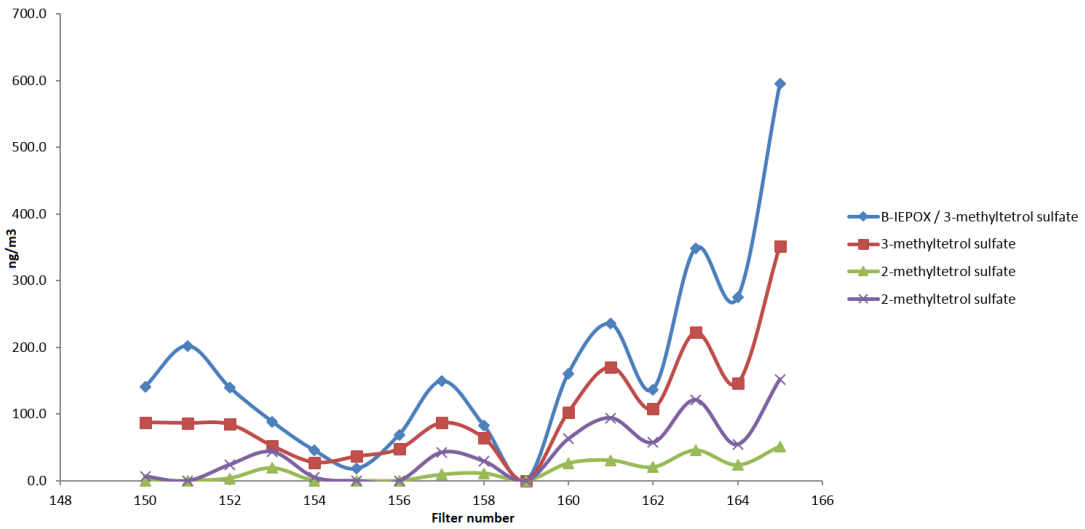

Figure 5.29: Temporal trends of the concentrations of four organosulfate isomers at *m/z* 215.0230 observed by the HILIC method across a small subset of filters, with the corresponding filter number shown on the x-axis.

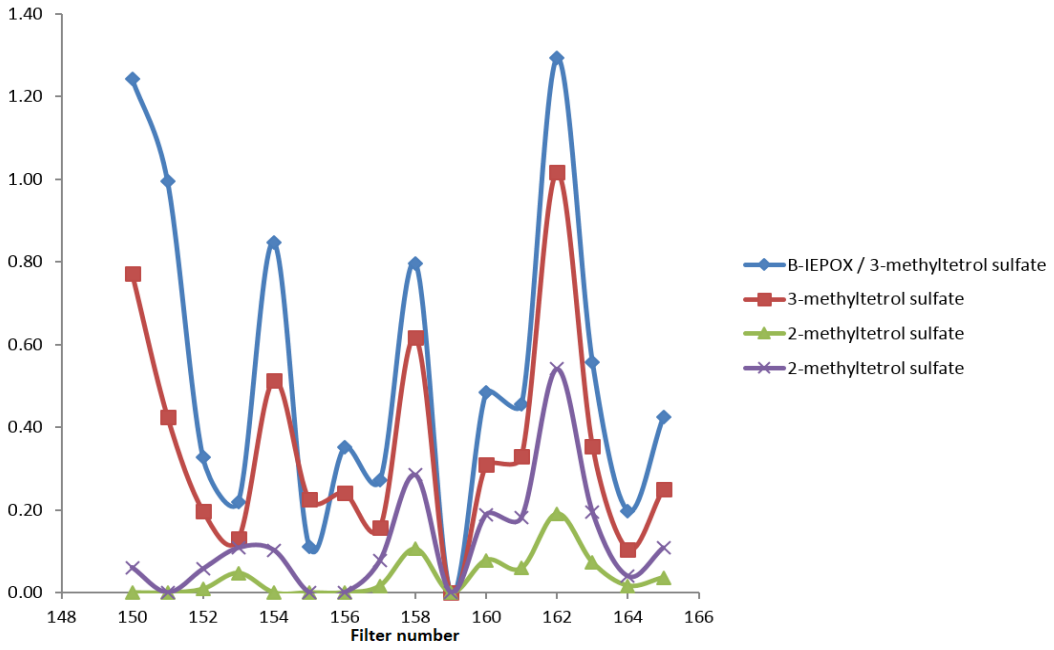

Figure 5.30: Normalised temporal trends of the four organosulfate isomers at *m/z* 215.0230 observed by the HILIC method across a small subset of filters, with the corresponding filter number shown on the x-axis.

5.4 Conclusions

Organosulfates are the most varied heteroatom group observed across the two campaigns in Beijing. They possess the largest variation in number of carbons, structure and chemical properties making predicting behaviours and trends very difficult.

In total, 191 organosulfates were observed by the UPLC-MS method. And after data quality assurance was employed, 84 compounds were clustered into 12 total cluster groups, with six of these groups being the most important in terms of contribution to the total signal and most significant variation and dominance. The six major groups averaged over 80% of the total signal from organosulfate group with group 2 averaging consistently around 40-50% of the total signal, making this group the most consistently important. In terms of importance and contribution to the total levels of CHOS associate peak area, group 2 is followed by groups 1 and 8.

Group 2 has localised source apportionment, typically from near the residential areas, largely dominated by one compound ($C_5H_{12}SO_4$, constituting over 50% of the total peak area) and most compounds contain less than 10 carbons, so are smaller in nature. The directional sourcing of this group suggests that the group consists of primary emissions from the local residential areas surrounding Beijing. Group 2 temporal trends show peaks around early afternoon (12:00 - 14:00) suggesting a strong links with photo-activity which is most likely the source of the other half of the compounds in this group. In comparison with auxiliary data, group 2 components have correlation with

humidity and with cooking sourced aerosol (COA). Based on its local sourcing and correlation with COA, this could indicate this group be dominated by domestic emissions, such as cleaning surfactants and cooking emissions or during lunch-time cooking from local restaurants and street-food vendors.

Group 1 largely comes from a south-by-southwesterly wind, the compounds are potentially biogenic based on their DBE values and the large green regions where the air-masses come from. The compounds are more likely to also be longer lived SOA being oxidised during transport towards Beijing. Group 1 components occur more during higher humidity levels; lending credence that these compounds are not directly sulphated, rather that they are further oxidised existing sulfates; as humidity inhibits the production of organosulfates. Additionally, group 1 compounds with both less oxidised oxygenated organic aerosol and more oxidised oxygenated organic aerosol (LO-OOA and MO-OOA) AMS factors suggesting these are very oxidised components, rather than freshy or directly emitted species.

Compounds which comprise group 8 are, in general, sourced over a wider spread, with the exception of an anomalously high abundance on the 19th of November which skewed the wind rose plot. This group is almost completely comprised of octyl sulfate, a common primary emission from industry and traffic emissions. Group 8 shows no significant trending with auxiliary data, this is most likely as the group is comprised of a commonly occurring compound present in many things from, house-hold cleaning products to anti-knocking agents in fueld; as such if some factors do correlate with a particular auxiliary factor, there may be others which anti-correlate and thus there would be no net correlation. In general octyl sulfate is more likely to be a primary emission

rather than a SOA product. It would therefore be the most probable candidate to address with regard to policies for cleaner air in China. Finding alternatives to the octyl sulfate surfactant in products like household cleaners or use a different anti-knocking agent in fuel as it would be difficult to reduce the amount of fuels used by the personal vehicles typically employed by the Beijing populous.

It is clear that whilst the majority of organosulfate compounds are sourced locally, a significant portion are sourced from outside of Beijing. As a result, even if the sourcing were addressed within Beijing, there would still be notable levels of organosulfates within the Beijing region.

Two air masses are observed in the winter sampling period, this is seen by the differences in contributions from the main groups during two periods within the winter campaign. The first air mass is characterised by higher contributions from group 5 compounds and occuring before 19th November. With the second air mass that is observed occurring after the 20th November and is characterised by increased contributions to the signal from groups 8 and 10.

It is seen that despite the reversed-phase method being effective as a universal detection method, it is not as effective when analysing a specific sub-set of compounds which have very little variation in polarity. Using HILIC chromatography as an alternate method of separation is a more effective separation due to the higher selectivity in very polar compounds and the addition of secondary interactions between analytes and the stationary phase packing.

Four important organosulfate isomers were observed as methyltetrol derived species, the formation of these species are shown to be mechanistically controlled and the levels of formation for each of these species do not co-vary with consistency, demonstrating the conditions of formation are variable during the sampling period, leading to the favoured formation of some methyltetrol derives organosulfates more than others.

Chapter 6 – Summary and future work

6.1 Summary

Organic aerosol makes up a significant proportion of particulate matter. The large number of different sources of VOCs and the numerous reaction pathways VOCs undergo lead to a vast complexity within the SOA. This complexity of SOA leads to challenges during identification and quantification. Techniques such as mass spectrometry allow for simultaneous identification of thousands of compounds at low concentrations; however, one of the caveats to such techniques is variation of ionisation efficiency using electrospray ionisation, can lead to mis-quantification if not properly catered for. Tracer compounds for different sources can be used as a fingerprint to identify contributions to SOA from different sources and show different source apportionment; these tracer compounds can be used as a proxy to allow for smaller scale analysis to be used to identify behaviours of a much larger scale complexity.

A short UPLC-MS method (approximately 16 minutes per sample) was developed to allow for the rapid analysis of offline filter samples. This technique has been shown to be suitable for high-throughput analysis of organic aerosol, while maintaining resolution and sufficient separation of most commonly found structural isomers of the most abundant species for complete identification. Purpose-built, automated data processing software was employed to the mass spectra data. This method was applied to 260 filter samples in total from two field campaigns in Beijing, China during both the winter and summer.

The normal limiting factor for high-throughput MS analysis, such as the technique developed for this work, is the identification and integration of peaks.

Manual peak integration for 1000s of compounds in 100s of samples, would normally take many weeks. This has been shown to be reduced to a matter of days by implementing automation for the detection and integration of peaks within mass spectra. This sort of automated data interpretation is the first of its kind and thus this time-efficient approach further benefits the analyst when it comes to big data analysis; a notable example is the analysis of co-variation of compounds to determine groupings. To analyse the co-variance between over 600 compounds would mean plotting and calculating the coefficient of variance over 300,000 times.

When automation of peak integrations in employed, the limiting factor for this type of analysis then becomes the extraction of the organic matter from the filters. Without the use of automation in the extraction process, the extraction of approximately 300 filter samples took multiple weeks. As such, this would need to be a consideration for such large-scale sampling going forward. Whilst automatic samplers can replace filters in short time increments to give good temporal resolution; at the time of writing, there exists no way to automate the extraction of the filters. A complete automation of filter sampling, extraction and extract analysis would allow for much larger scale analysis with increasing temporal resolution.

Additionally, the method can be shown to be adaptable to investigate specific groupings of compounds more closely but is generally very capable of giving overviews of whole organic composition, in some cases additional methods of chromatography has been employed to enhance the separation of some groups which C18-based reverse phased chromatography would be unable to do. In this case of this work, HILIC-based separation was used instead of the C18-

based separation due to HILIC's better effectiveness on compounds with similar and high polarity. This is due to the secondary interactions occurring in HILIC chromatography allowing for more specific interactions for organosulfates. The conclusion from this approach is that, whilst a broad-spectrum analysis is very good for rapid identification and understanding of the overview characterisation of organic aerosol, to properly gain more refined understanding, both the chromatography and library must be tailored to suit the desired group an analyst is investigating; as no one method would be both specific and detailed and truly broad enough to encompass the whole spectrum of organic aerosol.

Due to the novel nature of this work, full confirmation of the conclusions discussed in this would require conformational studies (discussed in future work, later in this thesis); at present, the conclusions seem to be in good standing with current literature and understanding of the nature of organic composition of atmospheric chemistry. The results from the off-line filter samples show that the winter campaign was mostly dominated by low DBE organosulfates and nitro-aromatics. This was in contrast to the summer campaign, which was less influenced by the nitro-aromatics. Instead, an increased abundance of higher O:C compounds was observed, most notably CHO and CHOS in nature (particularly abundant species were biogenic sourced, such as organosulfates). The high degree to which solar activity is a controlling factor on the composition of atmospheric aerosol in Beijing is significant; but is not the only defined limiting factor. For a notable amount of organosulfates and nitro-containing compounds, the abundance of inorganic oxidants is just as important, if not more so, for some species which form overnight.

In some cases this is as much of an influence on the composition as local traffic sources and as such, highlights the complexities observed in the sourcing of organic aerosol. Two air masses are observed in the winter sampling period, this is seen by the differences in contributions from the main groups during two periods within the winter campaign occurring either side of the 21st of November. This first period shows higher contribution from groups 3 and 5 and a lower contribution from group 8. The second period is the opposite, demonstrating a higher contribution from groups 8 and 10 and very low contribution from groups 5 and 6. Interestingly, the contribution from group 2 is consistent between both air masses. This highlights the need for specific understanding of individual species or at least more refined groupings of aerosols (for example, highly co-varying species rather than the current approach in literature which is of heteroatom composition).

The composition of air-masses can vary radically based on region and thus it is important to identify the specific case for a region when approaching policies which aim to reduce air pollution as a broad-brush approach to policy may not be applicable or effective in all regions it is applied to; and emphasises the need for further work to be performed on identifying individual species both as tracers of pollution sources, but understanding the results of oxidation and aging as air-masses move from the site of their sourcing.

The data showed good correlation with auxiliary AMS data and good correlation with total organic aerosol, capturing about 50% of the total variability. This highlights the validity of this method, but also demonstrates the limitations of this method of sampling. Offline filter sampling is limited, by nature, to the frequency of its filters. For the purposes of this work, this was an hourly

sampling with an overnight sample. Necessitating averaging of data to match other sampling techniques. This difference in sampling would account for a large part of the differences in the measurements and so the results observed in this work can be given credence.

One of the most impactful conclusions from this work is the observations that compounds with different structures show different ionisation efficiencies, this includes structural isomers of the same species. The extent of this variance in ionisation efficiency in structural isomers is up to three orders of magnitude. From this, limitations arise with regards to the quantification of unknowns by MS analysis and this can result in mis-quantification. In addition, there is a potential for 'losses' of compounds which possess lower sensitivity by ionisation in the method either being eclipsed by more abundant peaks without sufficient separation of isomers by chromatography or having a poor limit of detection for isomers with a lower ionisation efficiency. This also highlights the need for higher temporal resolution within off-line filter sampling, as without this, the contributions of some of the sources can be overlooked during 24 hour samples as the averages of these species tend to be averaged and sharper peaks are lost.

This is most negatively impactful to human and animal health, both in the atmospheric sense and in a broader sense. Not only are nitroaromatics hazardous in particulate matter and are known phytotoxins, their transfer into the water cycle through wet deposition leads to additional routes of exposure and thus increase the total risk factors posed by nitroaromatics. If water safety analysis is performed in a similar way to organic aerosol analysis, this could also have the limitations and lack of understanding of the links of isomerism

and ionisation efficiency and may indeed be another form of mis-quantification of nitroaromatics.

Nitrogen-containing species are observed to be a dominant contributor to the aerosol during the winter-time. These compounds are complex in nature; with the majority of these species being nitrophenolic in nature. Sourcing of these nitrophenolic compounds appear to mostly be more localised, as opposed to being brought to the sampling site via long-range transport of air masses. After HCA, six clusters were shown to best describe the nitrogen-containing species, the most important of these cluster groups, show defined diurnal patterns during periods of low $PM_{2.5}$.

The peak concentrations for the nitrogen-containing species typically occurred around 9 am. However, it is most likely that they are generated overnight and reside in the upper nocturnal layer; and mix down during the heating and subsequent mixing of these layers during the morning sampling period (8:30-9:30 am), as opposed to being directly emitted from traffic as could reasonably be suggested. This conclusion is based on period studies of the nocturnal generation of nitro-aromatics[170]. The concentrations of nitrophenolics tend to decay during the day, which is also consistent with previous observations. However whether this decay is due to deposition, further oxidation and functionalisation into species not in this work's database or whether they decay via HONO losses, is not known at this point.

The signal from biomass burning is unlike the more common nitrophenolics and does not follow the distinct diurnal cycles, instead demonstrates sharp peaks during particularly cold periods corresponding with increased biomass burning;

and comparatively little biomass burning tracer is seen in the warmer, summer sampling period.

Given that not all the groups follow the same apparent sourcing (and indeed, the variations of wind rose plots in the CHOS sourcing groups), it is unlikely that the CHON groups share similar localised sourcing, not as an artefact of the slackened wind speeds, but instead as a result of actual sourcing during low windspeeds.

In the other most dominant group, 191 organosulfates were observed by the UPLC-MS method across the two campaigns. After strict data quality assurance was employed, 84 compounds remained for further investigation. The trendlined data from these remaining species were clustered into six major groups and six minor groups. The major groups averaged over 80% of the total signal from organosulfates with CHOS group 2 averaging consistently around 40-50% of the total signal, making this group the most consistently important in the sulfur-containing aerosol.

Hierarchical clustering was employed to investigate potential commonalities and sources within the sulfur-containing compounds. This demonstrated how the chemistry of organic compounds depends on the interactions of the various sources. The exact composition of organic aerosol is in flux and its variation is dependent on the time of day and the meteorological conditions in the local area. The relative contribution of both local and regional sources varies across the sampling periods is also a contributing factor; again, highlighting the importance of further investigation into the nuances of individual species and their highly-covarying cluster groups. An interesting curiosity which was unfortunately not fully understood from this work as the highly anti-covarying

groups. These were less common in the HCA analysis but offer an interesting insight to the chemistry of organosulfates. The current thinking is that these anti-covarying species may compete for oxidants which cause their formation, as such when one is produced more favourably the other may not. Alternatively, it may be that one source is more active when the other source is not, but given our current understanding of organosulfate formation it is more likely to be a variation in the limiting factors controlling formation pathways.

An example of these variations in organosulfate formation can be seen in the methyl-tetrolsulfate-derived species. These compounds possess very similar structures and the same *m/z* values; however, have different sites of functionalisation (similarly to the nitrophenolics) and different stereochemistry (from their functionalisation of the alkene bond). It is notable that despite having the same source compounds and similar oxidants, these compounds are observed to co-vary poorly in the second time-period of our analysis, suggesting as their abundance increases (due to more favourable conditions of formation) the limiting factor becomes the source methyl-tetrol and thus as one organosulfate is formed more favourably, the others do not form as much. It should also be noted that these organosulfate isomers display variation in their ionisation efficiency, albeit not as significantly as the nitrophenolic group. More work is therefore required to better understand what the factors are which influence which organosulfate isomer gets formed over which. This would lead to a better understanding of formation pathways and may aid policy as these formation pathways can potentially be directly addressed reducing the amount of organosulfates (and potentially all SOA) occurring.

6.2 The place of this work in a greater scope

Going forward, this work demonstrates a technique which is not exclusive to the field of atmospheric chemistry. The implementation of rapid analysis of samples and quantification of hundreds of compounds can be translated and employed to a few fields, such as water safety or food safety. In the example of water safety testing, it would be better to tailor such a library to specifically water-soluble species; it is still possible to directly use the library described in this work, since the extraction method used in atmospheric aerosol used water-based extraction and therefore favoured water-soluble species. Additionally, there is a large over-lap of species which can occur in aerosol and within the water cycle, due to wet-deposition of aerosol or leaching of aerosol particles into water systems.

Notably, from the compound library described in this work, the group of compounds with the most importance to water systems are the nitrophenolics. With life-times of 1-2 months for nitrophenols, there is scope for nitrophenolics to complete the water cycle and enter drinking domestic water supplies or enter into livestock drinking water. Due to the toxic and phytotoxic nature of the nitrophenolic family of compounds, it is important to be able to assess safe levels within drinking water. And as demonstrated, the ionisation efficiency majorly varies depending on its structural isomerism, the importance of properly identifying all of the isomers to quantify correctly is discussed in more detail above. This is equally important in hydrological and atmospheric fields of study and regulation.

A greater understanding of the composition of aerosol also offers utility with regard to informing models and improving their databases to improve predictive capabilities. Using detailed time-resolved concentrations of individual species as a basis for machine learning will allow for models to run and make predictions of how an air mass will continue to evolve and age. Additionally, this could be used to test 'dummy' scenarios and test impacts of policies without affecting daily life. With individual species linked to sources, it would be possible to use the model to test what would happen if a source is increased or decreased and the effect on the composition. Testing in such a way to see which recommendations and policies bring about the most impactful change on a region would greatly improve the economical and practical impacts from policy makers. For example, in Europe, swapping to low CO_2 producing vehicles (diesel) did not take into account the impact of the catalytic converters using high levels of NOx and the subsequent impact of high NO_x levels on air quality; and indeed, one problem was created by the solution to another. Being able to more accurately simulate this using tracers and better understanding of the complexities of aerosol composition would lead to better informed policies going forward and avoid costly solutions which may have a further detrimental effect down the line.

In a broader scope of environmental science, an understanding of the composition of pollution gives insight into the sources. This is useful when establishing or informing policy. With a greater understanding of sources and how the pollution is formed, this allows for a more targeted and effective approach to reducing the pollution, and potentially making this more cost-effective too if resources and time are employed more efficiently too. Additionally, rather than a broad-brush approach to reducing the pollution in a

country, it may be much more effective to look at each region's specific composition and source to improve the knowledge base and understanding of the specific needs and employ a more tailored approach to reducing emissions in each region.

6.3 Recommendations to informing policy on improving air quality

Based on the outcomes of this work, some preliminary recommendations can be made. Is it clear that a significant fraction of the aerosol is produced from nitrophenolics. This would be the most logical, first point of call when trying to address the air quality in Beijing. To address this, the approach should be two-fold, the most obvious approach would be to minimise the traffic contributions of direct emissions of nitrophenolics. The author understands that personal transportation and commercial transportation is integral to the lifestyle of the Beijing population and as such, reducing the number of vehicles would lead to more severe impacts further down the line to people's livelihoods and lifestyles. Therefore, reducing the number of vehicles is not an effective method of reducing the pollution. Instead, the types of vehicles should be addressed.

During the campaign, it was observed that a large number of personal motorbikes and scooters are employed by the people and a lot of the larger vehicles are older and thus have higher levels of pollution. It follows that replacing these old, cheaper, personal motorbikes and scooters with newer vehicles with lower emissions or encouraging people to use more public transport (albeit with newer engines and less emissions) would more effectively target the direct emissions.

This would also help the indirect production of nitrophenolics. It was discussed earlier in the thesis, that the production of nitrophenolics from the nitrification of aromatics occurs during the nightime in the upper nocturnal layers of the atmosphere and then mix down in the warmer morning period. Reducing the emissions from transport would also, in part, address the emission of aromatics from transport and hence reduce the supply of aromatics which can be nitrified.

Aromatics are also generated by industrial sources; indeed it was shown in chapter 1 that approximately 20% of the pollution in Beijing comes from industry. In the lead up to the 2008 olympics, industrial zones were relocated, or their output reduced to improve the air quality for the athletes competing at the olympics. Repeating this on a larger scale would also lead to a reduction of emitted aromatics and VOCs from industrial sources, again limiting the levels of aromatics which can be nitrified.

It should be noted however, that relocating industry to the southern parts of Beijing would not be beneficial to the air quality of Beijing. It was shown that strong southerly winds occur around Beijing and a mountainous region to the north can lead to a build up of pollution as the air-masses become trapped in the region and have nowhere to dissipate; this is particularly the case during cold periods and slower wind-speeds (as seen in the winter campaign). Therefore, the selection of a region where industry can be relocated to must consider the wind directions which enter Beijing and might allow longer-lived transported VOCs and SOA back into Beijing.

The most prevalent compound occurring from the organosulfates is octyl sulfate and similar low DBE organosulfates are also in high abundance. These compounds tend to be locally sourced and based on their structure, they are

most likely to be surfactants and household emissions. As such, a reduction of these compounds being emitted cannot be achieved by reducing the number of products containing surfactants as they tend to be cleaning products and thus would reducing the level of cleaning would cause issues down the line. Instead, alternatives to surfactants should be put into the cleaning products used in Beijing. This would lead to a reduction in the amount of VOC aliphatics entering the atmosphere. Even if the compounds sourced within Beijing were addressed, there would still significant organosulfates brought in from further afield. This means that additional policies targeted to the surrounding areas would also be effective in improving the air quality in Beijing.

A final consideration is the level of SOA from biomass burning which comes into Beijing from long-range transport of air-masses. This is something already being addressed by the Beijing government, as a new initiative is being rolled-out replacing wood-burning heaters with electric heating linked up to a centralised grid. This is a positive step; however, given that the biomass burning SOA is able to be transported long distances; work should be done to ensure that this form of heating is rolled out effectively nationwide as air-masses from further west of Beijing may still impact the air-quality in Beijing as the VOCs produced there could become oxidised during transit and become SOA affecting Beijing.

6.4 Future work

Going forward, trends of individual compounds within the aerosol will be analysed in order to generate smaller, more distinct groups within the

heteroatom groups, such as the very complex sulfur-containing fraction. The need for higher precision and more detailed clustering analysis is increasingly abundant; grouping compounds into smaller clusters, with more similar sensitivities and functionality should give more accurate results during quantification, when applying the normalised calibration curves using calibration standards as analogous proxy compounds.

The confirmation of the sourcing of the nitrophenolic compounds can be achieved in future studies by sampling above and below the boundary between the separated nocturnal layers, however with only one sampler at a fixed height of 8m from ground level, this was not possible in this study. An automated filter sampler would be best practice for any off-line sampling over 300m from ground level, as manually swapping these filters would be both dangerous and time consuming having to travel up and down a tower or equivalent site where the higher level sampling would take place.

Ideally, given sufficient time and funding to synthesise the thousands of compounds observed in the organic aerosol would lead to the best quantification of components. However, as previously discussed, this would be impractical under current conditions and at present, representative samples for each cluster group is the best practice at the time of writing.

Alternatively, proxies (compounds possessing very similar structures and functionality) may offer the closest simulation to the actual chemical being quantified, but again with such high variation in ionisation efficiencies between structural isomers, there is no guarantee that a proxy would imitate the behaviours of the desired target analyte perfectly.

Another area of improvement within the field that is highlighted from this work, which requires further insight, is the need for higher temporal resolution within off-line aerosol measurement, in order to capture the finer details in compound trends is clear and from this, the limitations which come with standard 24 hour or 12 hourly sampling of ambient aerosol become apparent. High temporal resolution off-line filter sampling using high-volume samplers, could be best achieved using an automated filter sampler, to speed up the exchanging of filters, however this was not possible in this study. Whilst the swapping of filters could be automated and sped-up, there currently exists no way to automate the extraction. As such, this remains a significant limiting factor to increasing the through-put of such a method.

Alternatively, using the FIGAERO-CIMS as either an accompaniment to high temporal off-line measurements or instead of, would allow for an easy way to get higher resolution sampling.

List of abbreviations

AMS – Aerosol mass spectrometry

AVOC – Anthropogenic volatile organic compounds

BBOA – Biomass burning organic aerosol

BPC – Base peak Chromatogram

BSOA – Biogenic secondary organic aerosol

BVOC – Biogenic volatile organic compounds

CCN – Cloud condensation nuclei

CCOA – Coal-combustion organic aerosol

CHO – Compounds containing carbon, hydrogen and oxygen

CHON – Compounds containing carbon, hydrogen, nitrogen and oxygen

CHONS – Compounds containing carbon, hydrogen, nitrogen, sulfur and oxygen

CHOS – Compounds containing carbon, hydrogen, sulfur and oxygen

CMB – Chemical mass balance

COA – Cooking-based organic aerosol

DBE – Double bond equivalences

DD-MS^2 TOPN – Data dependant tandem mass spectrometry, top ion analysis

EC – Elemental carbon

EI – Electron impact ionisation

ELISA – Enzyme-Linked ImmunoSorbant Assay

ESI – Electro-spray ionisation

FIGAERO-CIMS – Filter inlet for gas and aerosol with chemical ionisation mass spectrometry

FT-ICR MS – Fourier transform ion cyclotron resonance mass spectrometry

GC – Gas chromatography

H:C – Hydrogen:Carbon

HCA – Hierarchical cluster analysis

HCD – Heated collision dissociation

HILIC - Hydrophilic interaction chromatography

HPLC – High-performance liquid chromatography

ICC – International committee on climate

LC – Liquid chromatography

LO-OOA – Less-oxidised oxygenated organic aerosol

LV-OOA – Low-volatility oxygenated organic aerosol

m/z – Mass to charge ratio

MO-OOA – More-oxidised oxygenated organic aerosol

MRM – Multiple reaction monitoring

MS – Mass Spectrometry

MS/MS – Tandem mass spectrometry

MVA – Multi-variance analysis

NP – Nitrophenol

NR-OM – Non-refractory organic matter

O:C – Oxygen:Carbon ratio

oa-TOFMS – Orthogonal acceleration time-of-flight mass spectrometry

OC – Organic carbon

OS - Organosulfates

OSc – Average oxidation state of carbon

PM – Particulate matter

PM2.5 – Particulate matter of 2.5 microns of less in diameter

POA – Primary organic aerosol

RCC – Residential coal combustion

RH – Relative humidity

RP-LC – Reverse phase liquid chromatography

RT – Retention time

S/N – Signal:Noise ratio

SAPA – Software for the automated processing of aerosol

SOA – Secondary Organic Aerosol

TIC – Total ion chromatogram

UNC – University of North Carolina, USA

UPLC – Ultra-performance liquid chromatography

VOC – Volatile Organic Compounds

WHO – World Health Organisation

WSOC – Water soluble organic compounds

xMAP – Multi-Analyte Profiling

References

1 D. L. Davis, M. L. Bell, D. L. Davis, M. L. Bell and D. Bates, *Environ. Health Perspect.*, 2002, **110**, 734–735.

2 B. Brunekreef and S. T. Holgate, *Lancet*, 2002, **360**, 1233–1242.

3 World Health Organization, *Geneva World Heal. Organ.*, 2013, 1–309.

4 J. Lelieveld, J. S. Evans, M. Fnais, D. Giannadaki and A. Pozzer, *Nature*, 2015, **525**, 367–71.

5 C. Arden Pope III and D. W. Dockery, *J. Air Waste Manage. Assoc.*, 2006, **56**, 1368–1380.

6 IARC, IARC : Outdoor air pollution a leading environmental cause of cancer deaths IARC : Outdoor air pollution a leading environmental cause of cancer deaths, https://www.iarc.fr/en/media-centre/iarcnews/pdf/pr221_E.pdf.

7 World Health Organization, *Geneva World Heal. Organ.*, 2006, 1–22.

8 L. R. Mazzoleni, P. Saranjampour, M. M. Dalbec, V. Samburova, A. G. Hallar, B. Zielinska, D. H. Lowenthal and S. Kohl, *Environ. Chem.*, 2012, **9**, 285–297.

9 I. Bejan, A. El, I. Barnes, T. Benter, B. Bohn and P. Wiesen, *Phys. Chem. Chem. Phys.*, 2006, **8**, 2028–2035.

10 J. Tremp, P. Mattrel, S. Fingler and W. Giger, *Water Air Soil Pollut.*, 1993, **68**, 113–123.

11 Intergovernmental Panel on Climate Change, *Climate Change 2013*, 2013.

12 Q. Zhang, J. L. Jimenez, M. R. Canagaratna, I. M. Ulbrich, N. L. Ng, D. R. Worsnop and Y. Sun, *Anal. Bioanal. Chem.*, 2011, **401**, 3045–3067.

13 P. H. McMurry, *Atmos. Environ.*, 2000, **34**, 1959–1999.

14 B. J. Finlayson-Pitts, *Faraday Discuss.*, , DOI:10.1039/C7FD00161D.

15 E. A. Bruns, I. El Haddad, G. S. Jay, D. Kilic, F. Klein, U. Baltensperger and A. S. H. Prévôt, 2016, 1–9.

16 C. Mohr, F. D. Lopez-hil, P. Zotter, A. S. H. Pre, L. Xu, N. L. Ng, S. C. Herndon, L. R. Williams, J. P. Franklin, M. S. Zahniser, D. R. Worsnop, W. B. Knighton, A. C. Aiken, K. J. Gorkowski, M. K. Dubey, J. D. Allan and J. A. Thornton, *Environ. Sci. Technol*, , DOI:10.1021/es400683v.

17 Y. Desyaterik, Y. Sun, X. Shen, T. Lee, X. Wang, T. Wang and J. L. Collett, *J. Geophys. Res. Atmos.*, 2013, **118**, 7389–7399.

18 H. Lukács, A. Gelencsér, S. Hammer, H. Puxbaum, C. A. Pio, M. Legrand, A. Kasper-Giebl, M. Handler, A. Limbeck, D. Simpson and S. Preunkert, *J. Geophys. Res. Atmos.*, 2007, **112**, 1–9.

19 J. Hamilton, P. Webb, A. Lewis, J. Hopkins, S. Smith and P. Davy, *Atmos. Chem. Phys. Discuss.*, 2004, **4**, 1393–1423.

20 S. Guo, M. Hu, Q. Guo, X. Zhang, M. Zheng, J. Zheng, C. C. Chang, J. J. Schauer and R. Zhang, *Environ. Sci. Technol*, 2012, **46**, 9846–9853.

21 M. Dan, G. Zhuang, X. Li, H. Tao and Y. Zhuang, *Atmos. Environ.*, 2004,

38, 3443–3452.

22 N. J. Farren, N. Ramírez, J. D. Lee, E. Finessi, A. C. Lewis and J. F. Hamilton, *Environ. Sci. Technol.*, 2015, **49**, 9648–9656.

23 F. Klein, S. M. Platt, N. J. Farren, A. Detournay, E. A. Bruns, C. Bozzetti, K. R. Daellenbach, D. Kilic, N. K. Kumar, S. M. Pieber, J. G. Slowik, B. Temime-roussel, N. Marchand, J. F. Hamilton, U. Baltensperger, A. S. H. Pre and I. El Haddad, *Environ. Sci. Technol*, 2016, **50**, 1243–1250.

24 X. Yin, Z. Huang, J. Zheng, Z. Yuan, W. Zhu, X. Huang and D. Chen, *Atmos. Res.*, 2017, **186**, 63–71.

25 J. D. Surratt, Y. Gómez-González, A. W. H. Chan, R. Vermeylen, M. Shahgholi, T. E. Kleindienst, E. O. Edney, J. H. Offenberg, M. Lewandowski, M. Jaoui, W. Maenhaut, M. Claeys, R. C. Flagan and J. H. Seinfeld, *J. Phys. Chem. A*, 2008, **112**, 8345–8378.

26 A. H. Goldstein, C. D. Koven, C. L. Heald and I. Y. Fung, *Nature*, 2009, **106**, 8835–40.

27 Q. Chen, D. K. Farmer, J. Schneider, S. R. Zorn, C. L. Heald, T. G. Karl, A. Guenther, J. D. Allan, N. Robinson, H. Coe, J. R. Kimmel, T. Pauliquevis, S. Borrmann, U. Pöschl, M. O. Andreae, P. Artaxo, J. L. Jimenez and S. T. Martin, *Geophys. Res. Lett.*, 2009, **36**, 1–5.

28 M. Hallquist, J. C. Wenger, U. Baltensperger, Y. Rudich, D. Simpson, M. Claeys, J. Dommen, N. M. Donahue, C. George, a. H. Goldstein, J. F. Hamilton, H. Herrmann, T. Hoffmann, Y. Iinuma, M. Jang, M. E. Jenkin, J. L. Jimenez, a. Kiendler-Scharr, W. Maenhaut, G. McFiggans, T. F.

Mentel, a. Monod, a. S. H. Prévôt, J. H. Seinfeld, J. D. Surratt, R. Szmigielski and J. Wildt, *Atmos. Chem. Phys.*, 2009, **9**, 5155–5236.

29 A. Cincinelli, A. M. Stortini, M. Perugini, L. Checchini and L. Lepri, *Mar. Chem.*, 2001, **76**, 77–98.

30 Q. Zhang, J. L. Jimenez, M. R. Canagaratna, J. D. Allan, H. Coe, I. Ulbrich, M. R. Alfarra, A. Takami, A. M. Middlebrook, Y. L. Sun, K. Dzepina, E. Dunlea, K. Docherty, P. F. Decarlo, D. Salcedo, T. Onasch, J. T. Jayne, T. Miyoshi, A. Shimono, S. Hatakeyama, N. Takegawa, Y. Kondo, J. Schneider, F. Drewnick, L. Cottrell, R. J. Griffin, J. Rautiainen, J. Y. Sun and Y. M. Zhang, *Geophys. Res. Lett.*, 2007, **34**, 1–6.

31 J. F. Hamilton, M. R. Alfarra, N. Robinson, M. W. Ward, A. C. Lewis, G. B. McFiggans, H. Coe and J. D. Allan, *Atmos. Chem. Phys.*, 2013, **13**, 11295–11305.

32 Q. T. Nguyen, M. K. Christensen, F. Cozzi, A. Zare, A. M. K. Hansen, K. Kristensen, T. E. Tulinius, H. H. Madsen, J. H. Christensen, J. Brandt, A. Massling, J. K. Nøjgaard and M. Glasius, *Atmos. Chem. Phys.*, 2014, **14**, 8961–8981.

33 R. J. van der A, B. Mijling, J. Ding, M. E. Koukouli, F. Liu, Q. Li, H. Mao and N. Theys, *Atmos. Chem. Phys. Discuss.*, 2016, 1–18.

34 X. Zhou, Z. Cao, Y. Ma, L. Wang, R. Wu and W. Wang, *Chemosphere*, 2016, **144**, 518–526.

35 R. Atkinson, *J. Phys. Chem. Ref. Data*, , DOI:10.1063/1.556012.

36 J. G. Calvert, R. Atkinson, J. a. Kerr, S. Madronich, G. K. Moortgat, T. J.

Wallington and G. Yarwood, *Oxford Univ. Press*, , DOI:10.1016/j.jphotochem.2010.03.004.

37 N. M. Donahue, J. H. Kroll, S. N. Pandis and A. L. Robinson, *Atmos. Chem. Phys.*, 2012, **12**, 615–634.

38 F. A. Mackenzie-rae, H. J. Wallis, A. R. Rickard, K. Pereira, S. M. Saunders, X. Wang and J. F. Hamilton, *Atmos. Chem. Phys.*, 2017, 1–34.

39 K. L. Pereira, University of York, 2014.

40 D. Vione, V. Maurino, C. Minero, M. Duncianu, R. Olariu, C. Arsene, M. Sarakha and G. Mailhot, *Atmos. Environ.*, 2009, **43**, 2321–2327.

41 I. M. Al-naiema and E. A. Stone, 2017, 2053–2065.

42 M. A. J. Harrison, M. R. Heal and J. N. Cape, *Atmos. Chem. Phys.*, 2005, 1679–1695.

43 M. D. Hurley, T. J. Wallington and K. H. Becker, 2001, **35**, 1358–1366.

44 R. Martin-skilton, M. W. H. Coughtrie and C. Porte, 2006, **79**, 24–30.

45 D. Vione, V. Maurino, C. Minero, M. Duncianu, R. Olariu, C. Arsene, M. Sarakha and G. Mailhot, *Atmos. Environ.*, 2009, **43**, 2321–2327.

46 L. Ganranoo, S. K. Mishra, A. K. Azad, A. Shigihara, P. K. Dasgupta, Z. S. Breitbach, D. W. Armstrong, K. Grudpan and B. Rappenglueck, *Anal. Chem.*, , DOI:10.1021/ac101015y.

47 C. Schummer, C. Groff, J. Al Chami, F. Jaber and M. Millet, *Sci. Total Environ.*, , DOI:10.1016/j.scitotenv.2009.06.051.

48 A. Cecinato, V. Di Palo, D. Pomata and M. Possanzini, *Chemosphere*, 2005, **59**, 679–683.

49 Y. Y. Zhang, L. Müller, R. Winterhalter, G. K. Moortgat, T. Hoffmann and U. Pöschl, *Atmos. Chem. Phys.*, 2010, **10**, 7859–7873.

50 M. R. Heal, M. A. J. Harrison and J. Neil Cape, *Atmos. Environ.*, , DOI:10.1016/j.atmosenv.2007.02.003.

51 J. Lüttke and K. Levsen, *Atmos. Environ.*, , DOI:10.1016/S1352-2310(96)00228-2.

52 R. I. Olariu, B. Klotz, I. Barnes, K. H. Becker and R. Mocanu, *Atmos. Environ.*, , DOI:10.1016/S1352-2310(02)00202-9.

53 H. Sukuzi and T. Mori, *J. Chem. Soc.*, 1995, 41–44.

54 W. Xu, Y. Sun, Q. Wang, W. Du, J. Zhao, X. Ge, T. Han, Y. Zhang, W. Zhou, J. Li, P. Fu, Z. Wang and D. R. Worsnop, *ACS Earth Sp. Chem.*, , DOI:10.1021/acsearthspacechem.7b00106.

55 M. A. J. Harrison, S. Barra, D. Borghesi, D. Vione, C. Arsene and R. Iulian Olariu, *Atmos. Environ.*, 2005.

56 Y. Desyaterik, Y. Sun, X. Shen, T. Lee, X. Wang, T. Wang and J. L. Collett, *J. Geophys. Res. Atmos.*, , DOI:10.1002/jgrd.50561.

57 H. Lukács, A. Gelencsér, S. Hammer, H. Puxbaum, C. A. Pio, M. Legrand, A. Kasper-Giebl, M. Handler, A. Limbeck, D. Simpson and S. Preunkert, *J. Geophys. Res. Atmos.*, , DOI:10.1029/2006JD008151.

58 J. D. Surratt, J. H. Kroll, T. E. Kleindienst, E. O. Edney, M. Claeys, A.

Sorooshian, N. L. Ng, J. H. Offenberg, M. Lewandowski, M. Jaoui, R. C. Flagan and J. H. Seinfeld, *Environ. Sci. Technol.*, 2007, **41**, 517–527.

59 K. A. Pratt, M. N. Fiddler, P. B. Shepson, A. G. Carlton and J. D. Surratt, *Atmos. Environ.*, 2013, **77**, 231–238.

60 R. Ots, D. E. Young, M. Vieno, L. Xu, R. E. Dunmore, J. D. Allan, H. Coe, L. R. Williams, S. C. Herndon, N. L. Ng, J. F. Hamilton, R. Bergström, C. Di Marco, E. Nemitz, I. A. Mackenzie, J. J. P. Kuenen, D. C. Green, S. Reis and M. R. Heal, *Submitt. to ACP*, 2015, **3**, 6453–6473.

61 J. F. Hamilton, M. Rami Alfarra, K. P. Wyche, M. W. Ward, A. C. Lewis, G. B. McFiggans, N. Good, P. S. Monks, T. Carr, I. R. White and R. M. Purvis, *Atmos. Chem. Phys.*, 2011, **11**, 5917–5929.

62 M. I. Schurman, T. Lee, Y. Sun, B. A. Schichtel, S. M. Kreidenweis and J. L. C. Jr, 2015, **15**, 737–752.

63 a. P. S. Hettiyadura, E. a. Stone, S. Kundu, Z. Baker, E. Geddes, K. Richards and T. Humphry, *Atmos. Meas. Tech.*, 2015, **8**, 2347–2358.

64 P. Lin, J. Z. Yu, G. Engling and M. Kalberer, *Environ. Sci. Technol.*, 2012, **46**, 13118–13127.

65 N. L. Ng, P. S. Chhabra, A. W. H. Chan, J. D. Surratt, J. H. Kroll, A. J. Kwan, D. C. McCabe, P. O. Wennberg, A. Sorooshian, S. M. Murphy, N. F. Dalleska, R. C. Flagan and J. H. Seinfeld, *Atmos. Chem. Phys.*, 2007, **7**, 5159–5174.

66 T. Cui, Z. Zeng, E. O. dos Santos, Z. Zhang, Y. Chen, Y. Zhang, C. A. Rose, S. H. Budisulistiorini, L. B. Collins, W. M. Bodnar, R. A. F. de Souza,

S. T. Martin, C. M. D. Machado, B. J. Turpin, A. Gold, A. P. Ault and J. D. Surratt, *Environ. Sci. Process. Impacts*, , DOI:10.1039/C8EM00308D.

67 J. Yin, S. A. Cumberland, R. M. Harrison, J. Allan, D. E. Young, P. I. Williams and H. Coe, *Atmos. Chem. Phys.*, 2015, **15**, 2139–2158.

68 U. S. E. P. A. (EPA), Chemical Mass Balance (CMB) Model - US EPA, https://www3.epa.gov/scram001/receptor_cmb.htm.

69 J. T. Jayne, D. C. Leard, X. Zhang, P. Davidovits, A. Kenneth, C. E. Kolb, D. R. Worsnop, J. T. Jayne, D. C. Leard, X. Zhang, P. Davidovits, A. Smith, C. E. Kolb, D. R. Worsnop, J. T. Jayne, D. C. Leard, X. Zhang, K. A. Smith and C. E. Kolb, *Aerosol Sci. Technol.*, 2000, **6826**, 49–70.

70 T. E. Kleindienst, M. Jaoui, M. Lewandowski, J. H. Offenberg and K. S. Docherty, *Atmos. Chem. Phys.*, 2012, **12**, 8711–8726.

71 J. Laskin, A. Laskin and S. A. Nizkorodov, *Mass Spectrometry Analysis in Atmospheric Chemistry Mass Spectrometry Analysis in Atmospheric Chemistry*, 2017.

72 F. D. Lopez-Hilfiker, C. Mohr, M. Ehn, F. Rubach, E. Kleist, J. Wildt, T. F. Mentel, A. Lutz and M. Hallquist, 2014, 983–1001.

73 E. Finessi, R. T. Lidster, F. Whiting, T. Elliott, M. R. Alfarra, G. B. McFiggans and J. F. Hamilton, *Anal. Chem.*, 2014, **86**, 11238–11245.

74 M. Z. Özel, M. W. Ward, J. F. Hamilton, A. C. Lewis, T. Raventós-duran and R. M. Harrison, *Aerosol Sci. Technol.*, 2011, **6826**, 109–116.

75 A. A. Bletsou, J. Jeon, J. Hollender, E. Archontaki and N. S. Thomaidis, *Trends Anal. Chem.*, 2015, **66**, 32–44.

76 H. Lee and B. J. Lee, *Food Addit. Contam.*, 2011, **0049**, 396–407.

77 J. Laskin, A. Laskin and S. A. Nizkorodov, *Anal. Chem.*, 2017, **90**, 166–189.

78 G. J. Patti, O. Yanes and G. Siuzdak, *Nat. Publ. Gr.*, 2012, **13**, 263–269.

79 N. R. Kitteringham, R. E. Jenkins, C. S. Lane, V. L. Elliott and B. K. Park, *J. Chromatogr. B*, 2009, **877**, 1229–1239.

80 T. G. Rosano, M. Wood and T. A. Swift, *J. Anal. Toxicol.*, 2011, **35**, 411–423.

81 J. Peters, R. Van Dam, R. Van Doorn, D. Katerere, F. Berthiller, W. Haasnoot and M. W. F. Nielen, *PLoS One*, 2017, **10**, 1–27.

82 E. De Dominicis and M. Suman, *J. Mass Spectrom.*, 2012, 1232–1241.

83 F. H. Allen, *Acta Crystallogr. Sect. B*, 2002, 380–388.

84 HighChem LLc, mzCloud, https://www.mzcloud.org/.

85 T. Kind, G. Wohlgemuth, D. Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz and O. Fiehn, *Anal. Chem.*, 2009, **81**, 10038–10048.

86 X. Wang, N. Hayeck, M. Brüggemann, L. Yao, H. Chen and C. Zhang, , DOI:10.1002/2017JD026930.

87 A. Masalaite, R. Holzinger, V. Remeikis, T. Röckmann and U. Dusek, 2017, **148**, 62–76.

88 A. A. Bletsou, J. Jeon, J. Hollender, E. Archontaki and N. S. Thomaidis, *Trends Anal. Chem.*, 2015, **66**, 32–44.

89 G. J. Patti, O. Yanes and G. Siuzdak, *Nat. Publ. Gr.*, 2012, **13**, 263–269.

90 HighChem LLc, mzCloud.

91 W. F. Rogge, L. M. Hildemann, M. A. Mazurek, G. R. Cass and B. R. T. Simoneit, *Environ. Sci. Technol.*, 1993, **27**, 2736–2744.

92 W. F. Rogge, L. M. Hildemann, M. A. Mazurek, G. R. Cass and B. R. T. Simoneit, *Environ. Sci. Technol.*, 1998, **32**, 13–22.

93 I. Bobeldijk, M. Hekman, J. D. V. Der Weij, L. Coulier, R. Ramaker, R. Kleemann, T. Kooistra, C. Rubingh, A. Freidig and E. Verheij, *J. Chromatogr. B*, 2008, **871**, 306–313.

94 K. Burgess, D. Creek, P. Dewsbury, K. Cook and M. P. Barrett, *Rapid Commun. Mass Spectrom.*, 2011, **25**, 3447–3452.

95 R. M. De Miranda, M. De Fátima Andrade, A. Worobiec and R. Van Grieken, *Atmos. Environ.*, 2002, **36**, 345–352.

96 B. Treiger, I. Bondarenko, H. Van Malderen and R. Van Grieken, *Anal. Chim. Acta*, 1995, **317**, 33–51.

97 P. K. Hopke, E. S. Gladney, G. E. Gordon, W. H. Zoller and A. G. Jones, *Atmos. Environ.*, 1976, **10**, 1015–1025.

98 A. J. Shusterman, P. G. McDougal and A. Glasfeld, *J. Chem. Educ.*, 1997, **74**, 1222.

99 B. L. Karger, *J. Chem. Educ.*, 1997, **74**, 45.

100 L. Nováková, L. Matysová and P. Solich, *Talanta*, 2006, **68**, 908–918.

101 J. C. Van De Steene and W. E. Lambert, *J. Am. Soc. Mass Spectrom.*, 2008, **19**, 713–718.

102 R. Belloli, B. Barletta, E. Bolzacchini, S. Meinardi, M. Orlandi and B. Rindone, *J. Chromatogr. A*, 1999, **846**, 277–281.

103 L. Dolejs, P. Beran and J. Hradec, *Org. Mass Spectrom.*, 1968, **1**, 563–566.

104 R. S. Gohlke and F. W. McLafferty, *J. Am. Soc. Mass Spectrom.*, 1993, **4**, 367–371.

105 K. M. Downard, *Eur. J. Mass Spectrom. (Chichester, Eng).*, 2007, **13**, 177–190.

106 J. B. Fenn, *J. Biomol. Tech.*, 2002, 13, 101–118.

107 M. Pulfer and R. C. Murphy, *Mass Spectrom. Rev.*, 2003, **22**, 332–364.

108 D. T. Snyder, C. J. Pulliam, Z. Ouyang and R. G. Cooks, *Anal. Chem.*, 2016, **88**, 2–29.

109 S. Yang, J. Ding, J. Zheng, B. Hu, J. Li, H. Chen, Z. Zhou and X. Qiao, *Anal. Chem.*, 2009, **81**, 2426–2436.

110 H. R. Mottram, S. E. Woodbury and R. P. Evershed, *Rapid Commun. Mass Spectrom.*, 1997, **11**, 1240–1252.

111 J. W. Martin, D. C. G. Muir, C. A. Moody, D. A. Ellis, W. C. Kwan, K. R. Solomon and S. A. Mabury, 2001, **74**, 323–329.

112 J. Parshintsev and T. Hyötyläinen, *Anal. Bioanal. Chem.*, 2015, 1–21.

113 B. N. Primanik, A. K. Ganguly and M. L. Gross, *Applied electrospray mass spectrometry*, 2002.

114 I. V. Chernushevich, A. V. Loboda and B. A. Thomson, *J. Mass Spectrom.*, , DOI:10.1002/jms.207.

115 W. C. Wiley and I. H. McLaren, *Rev. Sci. Instrum.*, , DOI:10.1063/1.1715212.

116 A. G. Marshall, C. L. Hendrickson and G. S. Jackson, , DOI:10.1002/9780470027318.a6006m.

117 R. A. Zubarev and A. Makarov, *Anal. Chem.*, 2013, **85**, 5288–5296.

118 J. Wang, Z. Hu, Y. Chen, Z. Chen and S. Xu, *AEA*, 2013, **68**, 221–229.

119 R. J. Huang, Y. Zhang, C. Bozzetti, K. F. Ho, J. J. Cao, Y. Han, K. R. Daellenbach, J. G. Slowik, S. M. Platt, F. Canonaco, P. Zotter, R. Wolf, S. M. Pieber, E. A. Bruns, M. Crippa, G. Ciarelli, A. Piazzalunga, M. Schwikowski, G. Abbaszade, J. Schnelle-Kreis, R. Zimmermann, Z. An, S. Szidat, U. Baltensperger, I. El Haddad and A. S. Prevot, *Nature*, 2014, **514**, 218–222.

120 Y. S. Wang, L. Yao, L. L. Wang, Z. R. Liu, D. S. Ji, G. Q. Tang, J. K. Zhang, Y. Sun, B. Hu and J. Y. Xin, *Sci. China Earth Sci.*, 2014, **57**, 14–25.

121 R. A. Rohde and R. A. Muller, *PLoS One*, 2015, **10**, 1–14.

122 IARC, IARC : Outdoor air pollution a leading environmental cause of cancer deaths IARC : Outdoor air pollution a leading environmental cause of cancer deaths.

123 L. E. Mael, M. I. Jacobs and M. J. Elrod, *J. Phys. Chem. A*, 2015, **119**, 4464–4472.

124 M. S. Shalamzari, R. Vermeylen, F. Blockhuys, T. E. Kleindienst, M. Lewandowski, R. Szmigielski, K. J. Rudzinski, G. Spólnik, W. Danikiewicz, W. Maenhaut and M. Claeys, *Atmos. Chem. Phys. Discuss.*, 2015, **15**, 29555–29590.

125 C. M. Boyd, T. Nah, L. Xu, T. Berkemeier and N. L. Ng, *Environ. Sci. Technol.*, 2017, acs.est.7b01460.

126 U. S. E. P. A. (EPA), Chemical Mass Balance (CMB) Model - US EPA.

127 F. D. Lopez-Hilfiker, C. Mohr, M. Ehn, F. Rubach, E. Kleist, J. Wildt, T. F. Mentel, A. Lutz and M. Hallquist, *Atmos. Meas. Tech.*, 2014, **7**, 983–1001.

128 F. Drewnick, S. S. Hings, P. DeCarlo, J. T. Jayne, M. Gonin, K. Fuhrer, S. Weimer, J. L. Jimenez, K. L. Demerjian, S. Borrmann and D. R. Worsnop, *Aerosol Sci. Technol.*, 2005, 39, 637–658.

129 P. F. DeCarlo, E. J. Dunlea, J. R. Kimmel, A. C. Aiken, D. Sueper, J. Crounse, P. O. Wennberg, L. Emmons, Y. Shinozuka, A. Clarke, J. Zhou, J. Tomlinson, D. R. Collins, D. Knapp, A. J. Weinheimer, D. D. Montzka, T. Campos and J. L. Jimenez, *Atmos. Chem. Phys.*, 2008, **8**, 4027–4048.

130 B. Nozière, M. Kalberer, M. Claeys, J. Allan, B. D'Anna, S. Decesari, E. Finessi, M. Glasius, I. Grgić, J. F. Hamilton, T. Hoffmann, Y. Iinuma, M. Jaoui, A. Kahnt, C. J. Kampf, I. Kourtchev, W. Maenhaut, N. Marsden, S. Saarikoski, J. Schnelle-Kreis, J. D. Surratt, S. Szidat, R. Szmigielski and

A. Wisthaler, *Chem. Rev.*, 2015, **115**, 3919–3983.

131 B. Schuur, A. B. de Haan, M. Kaspereit and M. Leeman, in *Comprehensive Biotechnology, Second Edition*, 2011.

132 L. R. Snyder, J. J. Kirkland and J. W. Dolan, *Introduction to Modern Liquid Chromatography*, 2010.

133 K. L. Pereira, J. F. Hamilton, A. R. Rickard, W. J. Bloss, M. S. Alam, M. Camredon, A. Muñoz, M. Vázquez, E. Borrás and M. Ródenas, *Atmos. Chem. Phys.*, 2014, **14**, 5349–5368.

134 S. I. Bohnenstengel, S. E. Belcher, A. Aiken, J. D. Allan, G. Allen, A. Bacak, T. J. Bannan, J. F. Barlow, D. C. S. Beddows, W. J. Bloss, A. M. Booth, C. Chemel, O. Coceal, C. F. Di Marco, M. K. Dubey, K. H. Falon, Z. L. Fleming, M. Furger, J. K. Gietl, R. R. Graves, D. C. Green, C. S. B. Grimmond, C. H. Halios, J. F. Hamilton, R. M. Harrison, M. R. Heal, D. E. Heard, C. Helfter, S. C. Herndon, R. E. Holmes, J. R. Hopkins, A. M. Jones, F. J. Kelly, S. Kotthaus, B. Langford, J. D. Lee, R. J. Leigh, A. C. Lewis, R. T. Lidster, F. D. Lopez-Hilfiker, J. B. McQuaid, C. Mohr, P. S. Monks, E. Nemitz, N. L. Ng, C. J. Percival, A. S. H. Prévôt, H. M. A. Ricketts, R. Sokhi, D. Stone, J. A. Thornton, A. H. Tremper, A. C. Valach, S. Visser, L. K. Whalley, L. R. Williams, L. Xu, D. E. Young and P. Zotter, *Bull. Am. Meteorol. Soc.*, 2015, **96**, 779–804.

135 J. Laskin, A. Laskin, P. J. Roach, G. W. Slysz, G. A. Anderson, S. A. Nizkorodov, D. L. Bones and L. Q. Nguyen, 2010, **82**, 2048–2058.

136 E. A. Bruns, I. El Haddad, G. S. Jay, D. Kilic, F. Klein, U. Baltensperger and A. S. H. Prévôt, *Sci. Rep.*, 2016, 1–9.

137 X. K. Wang, S. Rossignol, Y. Ma, L. Yao, M. Y. Wang, J. M. Chen, C. George and L. Wang, *Atmos. Chem. Phys. Discuss.*, 2015, **15**, 21415–21448.

138 Y. Wang, J. Ren, X. H. H. Huang, R. Tong and J. Z. Yu, *Environ. Sci. Technol*, 2017, **51**, 6791–6801.

139 N. Sareen, A. G. Carlton, J. D. Surratt, A. Gold, B. Lee and F. D. Lopez-hilfiker, *Atmos. Chem. Phys.*, 2016, **16**, 14409–14420.

140 J. J. Schauer, M. J. Kleeman, G. R. Cass and B. R. T. Simoneit, *Environ. Sci. Technol*, 2001, **35**, 1716–1728.

141 S. Staudt, S. Kundu, H. J. Lehmler, X. He, T. Cui, Y. H. Lin, K. Kristensen, M. Glasius, X. Zhang, R. J. Weber, J. D. Surratt and E. A. Stone, *Atmos. Environ.*, 2014, **94**, 366–373.

142 J. Xue, Y. Li, X. Xie, C. Xiong, H. Liu, S. Chen, Z. Nie, C. Chen and J. Zhao, *Atmos. Environ.*, 2017, **159**, 55–65.

143 P. J. Ziemann and R. Atkinson, *Chem. Soc. Rev.*, 2012, **41**, 6582.

144 M. Camredon, J. F. Hamilton, M. S. Alam, K. P. Wyche, T. Carr, I. R. White, P. S. Monks, a. R. Rickard and W. J. Bloss, *Atmos. Chem. Phys.*, 2010, **10**, 2893–2917.

145 A. S. Willoughby, A. S. Wozniak and P. G. Hatcher, *Atmos. Chem. Phys.*, 2014, **14**, 10299–10314.

146 S. L. Blair, A. C. Macmillan, G. T. Drozd, A. H. Goldstein, R. K. Chu, L. Paša-, J. B. Shaw, N. Toli, P. Lin, J. Laskin, A. Laskin and S. A. Nizkorodov, *Environ. Sci. Technol*, 2016, **51**, 119–127.

147 R. E. O'Brien, A. Laskin, J. Laskin, C. L. Rubitschun, J. D. Surratt and A. H. Goldstein, *J. Geophys. Res.*, 2014, 6636–6637.

148 J. H. Kroll, N. M. Donahue, J. L. Jimenez, S. H. Kessler, M. R. Canagaratna, K. R. Wilson, K. E. Altieri, L. R. Mazzoleni, A. S. Wozniak, H. Bluhm, E. R. Mysak, J. D. Smith, C. E. Kolb and D. R. Worsnop, *Nat. Chem.*, 2011, **3**, 133–139.

149 J. H. Kroll, C. Y. Lim, S. H. Kessler and K. R. Wilson, *J. Phys. Chem. A*, 2015, **119**, 10767–10783.

150 X.-F. Huang, D.-L. Chen, Z.-J. Lan, N. Feng, L.-Y. He, G.-H. Yu and S.-J. Luan, *Atmos. Res.*, 2012, **114**–**115**, 28–37.

151 A. Calogirou, B. R. Larsen and D. Kotzias, *Atmos. Environ.*, 1999, **33**, 1423–1439.

152 P. M. Edwards, K. C. Aikin, W. P. Dube, J. L. Fry, J. B. Gilman, J. Z. de Gouw, M. G. Graus, T. F. Hanisco, J. Holloway, G. Hübler, J. Kaiser, F. N. Keutsch and B. M. Lerner, *Nat. Geosci.*, 2017, **10**, 490–495.

153 M. Riva, T. Da Silva Barbosa, Y.-H. Lin, E. A. Stone, A. Gold and J. D. Surratt, *Atmos. Chem. Phys. Discuss.*, 2016, **16**, 1–39.

154 F. Pilar, M. B. Josep, A. Joan, G. Magda and M. S. Anna, *Environ. Sci. Technol.*, , DOI:10.1021/es00028a024.

155 W. E. Shafer and J. Schönherr, *Ecotoxicol. Environ. Saf.*, , DOI:10.1016/0147-6513(85)90071-5.

156 D. Vione, V. Maurino, C. Minero and C. Analitica, 2005, **39**, 7921–7931.

157 D. C. Carslaw and K. Ropkins, *Environ. Model. Softw.*, 2012, **27**, 52–61.

158 L. Chen, N. Heerink and M. Van Den Berg, 2006, **58**, 407–420.

159 W. Chen and R. Xu, *Energy Policy*, 2010, **38**, 2123–2130.

160 G. Schüürmann, R. K. Somashekar and U. Kristen, *Environ. Toxicol. Chem.*, , DOI:10.1897/1551-5028(1996)015<1702:SARFCA>2.3.CO;2.

161 J. Zhang, C. Wu, A. Jia and B. Hu, *Appl. Surf. Sci.*, , DOI:10.1016/j.apsusc.2014.01.130.

162 M. Tian, F. M. Yang, S. J. Chen, H. B. Wang, Y. Chen, L. Y. Zhang, L. M. Zhang, L. Xiang and B. Q. Qiao, *Chemosphere*, , DOI:10.1016/j.chemosphere.2017.08.077.

163 N. Daneshvar, M. A. Behnajady and Y. Zorriyeh Asghar, *J. Hazard. Mater.*, , DOI:10.1016/j.jhazmat.2006.06.045.

164 X. Wang, J. Yu, Y. Wang and L. Wang, *Chemosphere*, , DOI:10.1016/S0045-6535(01)00082-0.

165 J. H. Kroll and J. H. Seinfeld, *Atmos. Environ.*, 2008, **42**, 3593–3624.

166 J. Liggio, S. M. Li and R. McLaren, *Environ. Sci. Technol.*, 2005, **39**, 1532–1541.

167 E. A. Stone, L. Yang, L. E. Yu and M. Rupakheti, *Atmos. Environ.*, 2012, **47**, 323–329.

168 K. E. Altieri, B. J. Turpin and S. P. Seitzinger, *Atmos. Chem. Phys. Discuss.*, 2009, **8**, 17439–17466.

169 M. Riva, S. Tomaz, T. Cui, Y. H. Lin, E. Perraudin, A. Gold, E. A. Stone, E. Villenave and J. D. Surratt, *Environ. Sci. Technol.*, 2015, **49**, 6654–6664.

170 P. M. Edwards, K. C. Aikin, W. P. Dube, J. L. Fry, J. B. Gilman, J. Z. de Gouw, M. G. Graus, T. F. Hanisco, J. Holloway, G. Hübler, J. Kaiser, F. N. Keutsch and B. M. Lerner, *Nat. Geosci.*, , DOI:10.1038/NGEO2976.

Supplementary Data

Average oxidation state of carbon (OSc), corrected for sulfate groups, is calculated as follows:

$$\overline{OSc} \approx 2\,\frac{O-(3*S)}{C}-\frac{H}{C}$$

(eq. S1)

Where: O is number of oxygens, S is number of sulfurs, C is number of carbons and H is number of hydrogens present in the compound

Double bond equivalency (DBE) is calculated as follows:

$$DBE = C-\frac{H}{2}-\frac{N}{2}+1$$

(eq. S2)

Where: C is number of carbons, H is number of hydrogens and N is number of nitrogens present in the compound

. . .

Table S1: The variation in observed peak area for 6 standard compounds. All peak areas measured in arbitrary units.

	Compound peak area / a.u.					
Sample	**Pinonic Acid**	**2,6-Dimethyl-4-Nitrophenol**	**2-Nitrophenol**	**2-Nitro-1-Naphthol**	**4-Nitro-m-Cresol**	**Phthalic Acid**
1	366000	1400000	112000	99300	13200000	483000
2	384000	1670000	96500	92100	11600000	394000
3	388000	1660000	92400	84800	12800000	415000
4	378000	1600000	83400	97300	12800000	308000
5	378000	1600000	86600	75000	13200000	228000
6	350000	1470000	86900	73500	13500000	234000
Average	374000	1566667	92967	87000	12850000	343667
Standard Deviation %	3.39	6.32	10.24	11.63	4.75	27.53
Average Standard Deviation %	10.60					

Table S2: List of compounds in the database and the associated retention time for each compound.

Compound	Retention time (mins)	Compound	Retention time (mins)
2-(2-carboxyethyl)-3,3-dimethylcyclobutanecarboxy acid (B-Caryophyllene SOA)	4.75	C3H4O4	0.81
2-(2-hydroxy-ethyl)-3,3-dimethylcyclobutanecarboxylic acid (B-Caryophyllene SOA)	6.44	C3H6OSO4	11.72
2,4-Dinitrophenol	5.67	C3H6OSO4	0.71
2,6-dimethyl-4-nitrophenol	7.47	C3H6SO6	0.78
2-hydroxyhexanoic acid	4.63	C3H8SO4	1.02
2-Methyl-3-nitrophenol	6.57	C4H10SO4	1.70
2-Methyl-4-nitrophenol	6.86	C4H4O4	8.62
2-Methyl-5-nitrophenol	7.03	C4H4O4	0.84
2-Napthyl Sulfate	5.75	C4H4O4	9.88

2-Nitro-1-Naphthol	9.54	C4H6O4 (B-caryophyllene SOA)	1.00
2-Nitrophenol	6.16	C4H6SO6	0.76
3-(4-(5-hydroperoxy-5-hydroxypent-1-en-2-yl)-2,2-dimethylcyclobutyl)-3-hydroxypropanoic acid (B-Caryophyllene SOA)	8.20	C4H7NO3SO4	0.71
3,3-dimethyl-2-(3-oxobutyl)cyclobutanecarboxylic acid (B-Caryophyllene SOA)	7.3	C4H8OSO4	0.94
3,5-Dinitrocatechol	0.97	C4H8SO7	0.75
3-Methyl-2-nitrophenol	6.55	C5H10O(NO3)2SO4	2.81
3-Nitrophenol	5.40	C5H10O(NO3)2SO4	1.93
4-Methoxy-2-nitrophenol	6.97	C5H10O(NO3)2SO4	1.20
4-Methyl-2-nitrophenol	7.71	C5H10O1(NO3)2SO4	2.77
4-Methyl-3-nitrophenol	6.62	C5H10O11N2S	2.81

4-Nitrocatechol	4.11	C5H10O11N2S	1.91
4-Nitro-m-cresol	6.37	C5H10O11N2S	1.20
4-Nitrophenol	5.08	C5H10SO5	1.43
4-Nitrophenyl Sulfate	3.14	C5H10SO5	0.94
9-Nitroanthracene	9.12	C5H10SO6	0.80
Adipic Acid	2.58	C5H10SO7	0.78
Azelaic acid	6.53	C5H11NO3SO4	0.79
B-Caryophyllonic acid (B-Caryophyllene SOA)	9.10	C5H11NO3SO4	5.74
Benzoic Acid	5.87	C5H11NO3SO4	3.83
Benzyl Sulfate	1.09	C5H11NSO9	1.04
B-nocaryophyllonic acid (B-Caryophyllene SOA)	7.32	C5H11O2NO3SO4	0.96
C10H10O6S	6.96	C5H11O2NO3SO4	1.04
C10H10O6S	8.09	C5H11O3SO4	6.00
C10H10O6S	6.53	C5H11O3SO4	5.19
C10H10O6S	1.32	C5H11O7S	0.75
C10H10O6S	0.77	C5H11ONO3SO4	1.19
C10H10O7S	4.24	C5H11ONO3SO4	1.21
C10H10O7S	6.53	C5H12O3SO4	0.70
C10H11NO9S	4.25	C5H12OSO4	0.98
C10H11NO9S	3.11	C5H12SO4	5.50
C10H11O9NS	0.84	C5H12SO4	4.67

C10H12O2	1.15	C5H12SO4	4.32
C10H12O3	3.18	C5H12SO7	0.75
C10H12O4	1.37	C5H6O2SO4	0.73
C10H12O7S	4.22	C5H6O3SO4	0.71
C10H12O7S	1.73	C5H6O4	1.24
C10H12O7S	2.92	C5H6O4	0.78
C10H12O7S	2.06	C5H6O4SO4	0.71
C10H14O4 (Pinene SOA)	0.98	C5H6O5SO4	0.71
C10H14O5 (Pinene SOA)	1.52	C5H6OSO4	0.83
C10H14O6 (Pinene SOA)	4.68	C5H8O4 (Pinene SOA)	1.80
C10H15NO2S	6.38	C5H8O4 (Pinene SOA)	0.83
C10H15ONO3SO4	0.99	C5H8O4SO4	0.77
C10H16O2 (a-pinene SOA)	7.07	C5H8O5	5.88
C10H16O2SO4	3.37	C5H8O5	1.34
C10H16O3 (B-Caryophyllene SOA)	4.77	C5H8O5	0.88
C10H16O3 (Limonene SOA)	6.98	C5H8SO5	0.94
C10H16O3 (Limonene SOA)	6.48	C5H8SO8	0.77

C10H16O3 (B-Caryophyllene SOA)	6.83	C5H9COOH	2.38
C10H16O3(Pinene SOA)	5.37	C5H9COOH	4.51
C10H16O3SO4	7.53	C5H9COOH	1.79
C10H16O4 (Limonene SOA)	5.68	C5H9NO3SO4	0.71
C10H16O4 (a-pinene SOA)	3.92	C5H9O2NO3SO4	1.40
C10H16O4 (Limonene SOA)	6.53	C5H9O3NO3SO4	0.94
C10H16O4 (Limonene SOA)	1.28	C5H9ONO3SO4	1.77
C10H16O4SO4	6.16	C6H10O2SO4	1.16
C10H16O5 (Limonene SOA)	6.26	C6H10O3SO4	0.87
C10H16O5 (Pinene SOA)	1.32	C6H10O4 (Pinene SOA)	2.56
C10H16O5SO4	3.67	C6H10O4 (Pinene SOA)	1.18
C10H16O6	4.93	C6H10O4 (Pinene SOA)	3.77
C10H16O6SO4	5.39	C6H10O4 (Pinene SOA)	1.51

C10H16OSO4	6.76	C6H10O4SO4	0.77
C10H16SO4	1.28	C6H10O5	11.60
C10H16SO7	3.25	C6H10O5	10.35
C10H17NO3SO4	2.98	C6H10O5	1.08
C10H17NSO8	1.02	C6H10O5	8.68
C10H17NSO9	4.99	C6H10O5	7.04
C10H17O3NO3SO4	4.04	C6H10O5	5.25
C10H17O4NO3	3.83	C6H10O5SO4	0.71
C10H17O4NO3SO4	4.31	C6H11NO3SO4	0.71
C10H17O4SO4	6.95	C6H11O2NO3SO4	2.68
C10H17O4SO4	3.72	C6H11O3NO3SO4	1.10
C10H17O4SO4	0.90	C6H11ONO3SO4	2.44
C10H17O5NO3	4.58	C6H12(COOH)2	5.76
C10H17O8NS	6.53	C6H12O4	2.39
C10H17O9NS	4.22	C6H12O4	1.01
C10H17SO6	3.85	C6H12O5SO4	6.03
C10H18O(SO4)2	9.08	C6H12OHSO4	1.59
C10H18O3SO4	3.81	C6H12OSO4	1.36
C10H18O3SO4	1.21	C6H12SO5	1.59
C10H18O3SO4	1.61	C6H12SO6	0.91
C10H18O4 (Pinene SOA)	7.42	C6H12SO9	0.70
C10H18OSO4(NO3)2	6.63	C6H13NO3SO4	5.21

C10H18OSO4(NO3)2	8.50	C6H13ONO3SO4	1.88
C10H18SO5	5.69	C6H14OSO4	1.64
C10H18SO6	4.22	C6H14SO4	6.12
C10H19NO3SO4	8.72	C6H14SO4	5.62
C10H19O2NO3SO4	5.14	C6H14SO4	6.10
C10H19O3NO3SO4	1.42	C6H3OH(COOH)2	11.55
C10H19OH(NO3)2	10.05	C6H3OH(COOH)2	6.82
C10H19OH(NO3)2	8.22	C6H3OH(COOH)2	8.42
C10H19OH(NO3)2	1.26	C6H3OH(COOH)2	7.94
C10H20O3SO4	5.07	C6H3OH(COOH)2	4.57
C10H20O3SO4	3.97	C6H3OH(COOH)2	1.60
C10H20O3SO4	1.55	C6H4NO6S	11.28
C10H20OSO4	6.71	C6H4NO6S	3.27
C10H20SO6	6.98	C6H4NO6S	2.08
C10H21ONO3SO4	6.31	C6H4OHNO3	11.44
C10H22OSO4	6.64	C6H4OHNO3	10.82
C10H22SO10	0.81	C6H5COOHNO3	6.43
C10H22SO4	10.34	C6H5COOHNO3	5.87
C10H7NO3	8.25	C6H5COOHNO3	5.51
C11H12O4	7.50	C6H5NO3	5.14
C11H12O7S	5.38	C6H5NO3SO4	1.29
C11H12O7S	1.41	C6H5NO3SO4	2.96
C11H12O7S	0.82	C6H5O4S	5.5

C11H12O7S	6.55	C6H5O4S	6.25
C11H12O7S	3.09	C6H5O4S	4.31
C11H14O7S	4.93	C6H5O4S	1.45
C11H14O7S	2.51	C6H5OHCOOH	6.01
C11H14O7S	0.84	C6H5OHCOOH	3.87
C11H14O9	0.86	C6H5OHCOOH	2.96
C11H18O3	8.14	C6H5OHNO3	11.44
C11H18O6	6.94	C6H5OHNO3	4.11
C11H18O7	4.89	C6H5OHNO3	10.82
C11H19NO9	5.34	C6H6O6	0.78
C11H19O2NO3SO4	5.39	C6H8O3	2.73
C11H19O3NO3SO4	4.46	C6H8O3	1.36
C11H19O4NO3SO4	4.94	C6H8O6	0.76
C11H20O3SO4	4.75	C6H8SO6	0.79
C11H20O4	8.15	C6H9(COOH)3	3.26
C11H20O4SO4	3.42	C6H9(COOH)3	2.23
C11H20O5SO4	3.24	C6H9(COOH)3	1.81
C11H20OSO4	5.21	C7H10O2 (Pinene SOA)	3.20
C11H21NO3SO4	9.23	C7H10O2 (Pinene SOA)	1.76
C11H21O2NO3SO4	7.64	C7H10O4	2.36
C11H21O3NO3SO4	5.55	C7H11O2NO3SO4	0.71

C11H21ONO3SO4	8.05	C7H11O3NO3SO4	1.32
C11H22(COOH)2	9.29	C7H11O3SO4	1.04
C11H22OSO4	7.37	C7H12O2 (Pinene SOA)	5.74
C11H22SO6	7.93	C7H12O2 (Pinene SOA)	5.61
C11H23NO3SO4	9.85	C7H12O2 (Pinene SOA)	3.07
C11H23ONO3SO4	8.23	C7H12O2SO4	8.98
C11H24OSO4	8.45	C7H12O2SO4	1.25
C11H9NO3	8.64	C7H12O3SO4	1.02
C12H11NO3	4.80	C7H12O4	5.06
C12H14O4	8.09	C7H12O4	4.16
C12H14O4	6.39	C7H12SO6	8.98
C12H18O4 (B-Caryophyllene SOA)	6.51	C7H12SO6	1.26
C12H20O2SO4	5.91	C7H13O2NO3SO4	3.81
C12H20O3	8.21	C7H13O3NO3SO4	1.32
C12H20O3	8.77	C7H13ONO3SO4	0.73
C12H20O3SO4	10.31	C7H14COOHSO4	9.14
C12H20O4SO4	4.31	C7H14COOHSO4	8.09
C12H20O5SO4	3.74	C7H14OSO4	2.77
C12H20O6SO4	9.82	C7H14SO5	3.29

C12H21O2NO3SO4	7.73	C7H14SO5	2.81
C12H21O3NO3SO4	5.98	C7H15NO3SO4	6.34
C12H21O4NO3SO4	6.04	C7H15ONO3SO4	3.38
C12H22O4	8.77	C7H16OSO4	2.66
C12H23NO3SO4	6.07	C7H16SO4	7.48
C12H23O2NO3SO4	6.46	C7H16SO4	6.89
C12H23O3NO3SO4	6.46	C7H16SO4	8.04
C12H23ONO3SO4	8.92	C7H16SO4	7.47
C12H24SO6	8.57	C7H16SO4	6.66
C12H25NO3SO4	10.47	C7H6O2	4.06
C12H25ONO3SO4	8.41	C7H6O2	3.64
C12H26OSO4	8.81	C7H6O2	1.95
C12H26SO4	11.5	C7H6O4	2.44
C13H20O5 (B-Caryophyllene SOA)	6.14	C7H6O4	5.08
C13H20O5 (B-Caryophyllene SOA)	5.92	C7H6O4	3.50
C13H20O8SO4	8.77	C7H6O4	1.69
C13H20O8SO4	6.53	C7H6O4S	1.85
C13H20O8SO4	1.38	C7H6O5S	1.12
C13H20O8SO4	0.99	C7H6OHNO3	11.73
C13H21ONO3SO4	8.25	C7H6OHNO3	6.55
C13H22O3	8.97	C7H6OHNO3	5.07

C13H22O4 (B-Caroyphyllene SOA)	6.62	C7H7COOH	5.25
C13H22O4 (B-Caryophyllene SOA)	6.95	C7H7COOH	4.40
C13H23O2NO3SO4	7.01	C7H7NO3	11.74
C13H23O3NO3SO4	6.62	C7H7NO3	6.86
C13H23O4NO3SO4	5.35	C7H7NO4	11.58
C13H24O2SO4	7.07	C7H7NO4	5.07
C13H24O3SO4	9.22	C7H7NO4	3.97
C13H24O4SO4	4.60	C7H7NO4	5.62
C13H24O5SO4	5.44	C7H8O4S	1.07
C13H25O2NO3SO4	8.57	C7H8O4S	3.22
C13H25O3NO3SO4	6.94	C7H9ONO3SO4	0.73
C13H25OHSO4	9.4	C8H10O	8.68
C13H25ONO3SO4	9.72	C8H10O	5.74
C13H26OSO4	9.34	C8H10O2	1.60
C13H26SO6	8.83	C8H10O2	6.35
C13H26SO6	9.55	C8H10O3	1.42
C13H27NO3SO4	11.24	C8H10O4	0.98
C13H27ONO3SO4	9.32	C8H10O5	3.47
C13H28OSO4	9.59	C8H11ONO3SO4	0.79
C13H28SO4	9.59	C8H12O2SO4	1.13

C14H11NO3	4.59	C8H12O3 (Limonene SOA)	4.12
C14H20O5	7.74	C8H12O3SO4	5.26
C14H22O3 (B-Caryophyllene SOA)	8.87	C8H12O3SO4	3.85
C14H22O4 (B-Caryophyllene SOA)	7.63	C8H12O3SO4	1.28
C14H22O4 (B-Caryophyllene SOA)	7.34	C8H12O4 (Limonene SOA)	4.23
C14H22O6	5.82	C8H12O4 (Limonene SOA)	3.07
C14H22O7	5.88	C8H12O4SO4	1.02
C14H23O6NO3	9.58	C8H12O5 (Limonene SOA)	2.90
C14H23O6NO3	6.55	C8H12O5 (Limonene SOA)	1.29
C14H24O4	8.85	C8H12O5SO4	0.77
C14H24O4	9.32	C8H12O6 (Pinene SOA)	3.17
C14H24O4	7.86	C8H12O6 (Pinene SOA)	7.76
C14H24O4	7.18	C8H12O6SO4	0.77
C14H24O4SO4	9.12	C8H12OSO4	1.98

C14H24O4SO4	7.45	C8H13NSO12	3.66
C14H24O4SO4	8.50	C8H13NSO12	1.48
C14H24O4SO4	8.01	C8H13NSO12	0.95
C14H24O4SO4	5.14	C8H13O2NO3SO4	0.81
C14H24O5	8.07	C8H13O3NO3SO4	1.28
C14H24O5	7.57	C8H13O4NO3SO4	2.36
C14H24O5 (B-Caryophyllene SOA)	7.28	C8H14(COOH)2	5.41
C14H24O5SO4	4.31	C8H14(COOH)2	6.98
C14H24O6SO4	3.81	C8H14(COOH)2	4.98
C14H24O7	7.17	C8H14(COOH)2	3.92
C14H24O7SO4	3.95	C8H14O2SO4	2.12
C14H24O8	5.06	C8H14O2SO4	1.24
C14H24O8	9.1	C8H14O3SO4	3.04
C14H24O8	5.88	C8H14O3SO4	1.49
C14H24O8	4.13	C8H14O3SO4	0.89
C14H24OSO4	8.41	C8H14O4	5.76
C14H25O2NO3SO4	7.98	C8H14SO10	3.07
C14H25O3NO3SO4	7.37	C8H14SO10	0.79
C14H25O4NO3SO4	6.34	C8H15NO3SO4	0.85
C14H27O2NO3SO4	8.41	C8H15O2NO3SO4	2.86
C14H27O3NO3SO4	8.05	C8H15O3NO3SO4	1.98
C14H27OHSO4	9.47	C8H15OHSO4	4.38

C14H27OHSO4	10.09	C8H15OHSO4	3.85
C14H27ONO3SO4	10.51	C8H15ONO3SO4	4.42
C14H29ONO3SO4	9.87	C8H16(COOH)2	7.42
C14H30SO4	11.08	C8H16OSO4	4.19
C15H22O4	7.01	C8H16SO5	4.38
C15H22O7	5.02	C8H16SO5	3.86
C15H24O3SO4	6.48	C8H17COOHSO4	5.88
C15H24O4 (B-Caryophyllene SOA)	7.90	C8H17NO3SO4	7.42
C15H24O4SO4	4.69	C8H17ONO3SO4	4.89
C15H24O5	7.86	C8H18OSO4	5.05
C15H24O7	7.01	C8H18SO3	8.11
C15H25NO3SO4	10.91	C8H18SO4	9.44
C15H25O5NO3	7.86	C8H18SO4	8.54
C15H25O6NO3	8.34	C8H18SO4	8.55
C15H25O6NO3 - RT7.26	7.26	C8H18SO4	8.06
C15H25ONO3SO4	8.66	C8H18SO4	7.43
C15H25SO4	11.54	C8H5NO3	3.65
C15H25SO4	10.1	C8H6O4	4.06
C15H26O5	8.63	C8H6O4	9.60
C15H26O5	6.92	C8H6O4	1.96
C15H27O2NO3SO4	8.41	C8H6O4	1.41
C15H27O3NO3SO4	8.05	C8H6O6S	2.74

C15H27O4NO3SO4	6.66	C8H6O6S	1.02
C15H28O6SO4	9.29	C8H8O2	5.25
C15H28O6SO4	6.03	C8H8O2	4.39
C15H29ONO3SO4	10.91	C8H8O3	7.40
C15H31ONO3SO4	10.36	C8H8O3	5.29
C15H32OSO4	10.36	C8H8O3	4.77
C16H10O4N2	6.98	C8H8O4S	1.37
C16H24O8	6.07	C8H8O5S	8.40
C16H25O5SO4	7.42	C8H8O5S	3.12
C16H25O5SO4	8.03	C8H8O5S	2.71
C16H25SO3	11.78	C8H8O5S	4.16
C16H26O13	1.70	C8H8O5S	1.73
C16H26O13	9.49	C8H8OHNO3	8.52
C16H26O6	8.83	C8H8OHNO3	7.87
C16H26O7	7.34	C8H8OHNO3	6.5
C16H27ONO3SO4	9.76	C8H8OHNO3	5.00
C16H28O4SO4	6.44	C8H9NO3	7.58
C16H28O4SO4	6.26	C9H10O3	8.72
C16H28O8	6.44	C9H10O3	6.08
C16H29O2NO3SO4	8.75	C9H10O3	5.46
C16H29O3NO3SO4	8.57	C9H10O3	4.68
C16H29O4NO3SO4	8.05	C9H10O3	1.15
C16H32O2SO4	10.67	C9H10O5S	3.24

C16H32O5SO4	10.51	C9H10O5S	2.12
C16H32O5SO4	5.18	C9H12O3 (Limonene SOA)	3.86
C17H20O11	3.92	C9H12O3 (Limonene SOA)	2.83
C17H20O11	3.05	C9H12O5S	6.29
C17H20O11	1.19	C9H12O5S	3.86
C17H24O3(NO3)2	9.39	C9H12O5S	2.01
C17H26O8	6.53	C9H13ONO3SO4	0.94
C17H26O8	5.47	C9H14O2 (Limonene SOA)	5.70
C17H27SO3	9.47	C9H14O2 (Limonene SOA)	4.87
C17H28O3S	9.52	C9H14O2 (Limonene SOA)	3.25
C17H28O6	10.67	C9H14O3 (Limonene SOA)	4.87
C17H28O6	9.49	C9H14O3 (Limonene SOA)	5.80
C18H24O10	4.77	C9H14O3 (Limonene SOA)	4.45

C18H24O10	5.14	C9H14O4 (B-Caryophyllene SOA)	4.43
C18H25O3NO3	10.1	C9H14O4 (B-Caryophyllene SOA)	2.53
C18H25O3NO3	6.35	C9H14O4 (Limonene SOA)	4.90
C18H25O3NO3	8.57	C9H14O4 (Limonene SOA)	2.85
C18H25O3NO3	7.84	C9H15O2NO3SO4	4.19
C18H25O3NO3	7.17	C9H15O3NO3SO4	3.4
C18H27O3NO3	9.92	C9H15O4NO3SO4	1.55
C18H27O3NO3	7.02	C9H15O8NS	4.01
C18H27O3NO3	6.14	C9H15O8NS	5.36
C18H29O4NO3	10.09	C9H16O2SO4	3.52
C18H29O4NO3	9.25	C9H16O3SO4	1.14
C18H29SO3	10.96	C9H16O3SO4	2.75
C19H18O7	6.98	C9H16O3SO4	2.74
C19H28O7	8.25	C9H16O4	6.53
C19H28O7	7.77	C9H16O4SO4	3.29
C19H28O9	6.87	C9H16O4SO4	1.32
C19H28O9	5.90	C9H16O4SO4	0.83

C19H29O3NO3	10.19	C9H16O5SO4	0.73
C19H29O3NO3	7.49	C9H16SO6	3.61
C19H29O3NO3	6.67	C9H16SO6	1.82
C19H31SO3	11.82	C9H17NO3SO4	7.69
C19H31SO3	9.61	C9H17O2NO3SO4	3.22
C20H32O10	6.96	C9H17O3NO3SO4	3.81
C20H32O10	11.22	C9H17ONO3SO4	5.48
C20H32O8	7.77	C9H18O2SO4	5.87
C20H32O9	6.44	C9H18OSO4	5.73
C20H33O7SO4	10.57	C9H18SO6	5.88
C20H33O7SO4	9.54	C9H19NO3SO4	8.25
C20H33O7SO4	6.55	C9H19ONO3SO4	5.89
C20H33O7SO4	5.73	C9H20OSO4	6.39
C21H20O8	6.99	C9H20SO4	9.46
C21H20O8	9.32	C9H20SO4	9.05
C21H20O8	5.92	C9H20SO4	9.48
C21H36O7	9.71	C9H5O4N	3.95
C21H36O7	8.94	C9H5O4N	3.09
C2H3O2SO4	1.40	C9H6O6	4.95
C2H3O6S	0.75	C9H6O6	1.97
C2H3OSO4	0.77	C9H8O4	5.25
C2H4SO6	0.75	C9H8O5	8.11

cis-pinonic acid (pinene SOA)	5.74	C9H8O5	7.86
D-Mannose 6-Sulfate	0.7	C9H8O5	6.41
Dodecyl Sulfate	11.66	C9H8O5	4.76
M-Toluic acid	7.13	C9H8O5S	3.06
Nitro-catechol	4.11	C9H8O5S	0.91
nor-caryophyllenic acid (B-Caryophyllene SOA)	3.68	C9H8O5S	2.27
Octyl Sulfate	8.06	Pinic Acid (a-pinene SOA)	4.93