Chemometric analysis of full scan direct mass spectrometry data for the discrimination and source apportionment of atmospheric volatile organic compounds measured from a moving vehicle.

by

Larissa Christine Richards
B.Sc., Vancouver Island University, 2013

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Chemistry

I acknowledge with respect the lək̓ʷəŋən peoples on whose traditional territory the university stands, and the Songhees, Esquimalt, and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day.

Chemometric analysis of full scan direct mass spectrometry data for the discrimination and source apportionment of atmospheric volatile organic compounds measured from a moving vehicle.

by

Larissa Christine Richards
B.Sc., Vancouver Island University, 2013

Supervisory Committee

Dr. Erik Krogh, Co-Supervisor
Department of Chemistry

Dr. Tom Fyles, Co-Supervisor
Department of Chemistry

Dr. Jay Cullen, Outside Member
Department of Earth and Ocean Science

Dr. Chris Gill, Additional Member
Department of Chemistry

Abstract

Anthropogenic emissions into the troposphere can impact air quality, leading to poorer health outcomes in the affected areas. Volatile organic compounds (VOCs) are a group of chemical compounds, including some which are toxic, that are precursors in the formation of ground-level ozone and secondary organic aerosols. VOCs have a variety of sources, and the distribution of atmospheric VOCs differs significantly over time and space. Historically, the large number of chemical species present at low concentrations (parts-per-trillion to parts-per-billion by volume) have made VOCs difficult to measure in ambient air. However, with improvements in analytical instrumentation, these measurements are becoming more common place. Direct mass spectrometry (MS), such as membrane introduction mass spectrometry (MIMS) and proton-transfer reaction time-of-flight mass spectrometry (PTR-ToF-MS) facilitate real-time, continuous measurements of VOCs in air, with full scan mass spectral data capturing changes in chemical composition with high temporal resolution. Operated on-road, mobilized direct MS has been used for quantitative mapping of VOCs at the neighborhood scale, but identifying VOC sources based on the observed mixture of molecules in the full scan MS dataset has yet to be explored. This dissertation describes the use of chemometric techniques to interrogate full scan MS data, and the progression from discriminating VOC samples of known chemical composition based on full scan MIMS data through to the apportionment of VOC sources measured continuously with a PTR-ToF-MS system operating in a moving vehicle. Lab-constructed VOC samples of known chemical composition and concentration demonstrated the use of principal component analysis (PCA) to discriminate, and k-nearest neighbours to classify, samples based on normalized full scan MIMS data. Furthermore, multivariate curve resolution-alternating least squares (MCR-ALS) was used to resolve mixtures into molecular component contributions. PCA was also used to discriminate 'real-world' VOC mixtures (*e.g.*, woodsmoke VOCs, headspace above aqueous hydrocarbon samples) of unknown chemical composition measured by MIMS. Using vehicle mounted MIMS and PTR-ToF-MS systems, full scan MS data of ambient atmospheric VOCs were collected and PCA was applied to the normalized full scan MS data. A supervised analysis performed PCA on samples collected near known VOC sources, while an unsupervised analysis using PCA followed by cluster analysis was used to identify groups in a continuous, time series PTR-ToF-MS dataset measured between Nanaimo and Crofton, British Columbia (BC). In both the supervised and unsupervised analysis, samples impacted by emissions from different sources (*e.g.*, internal combustion engines, sawmills, composting facilities, pulp mills) were discriminated. With PCA, samples were discriminated based on differences in the observed full scan MS

data, however real-world samples are often impacted by multiple VOC sources. MCR-weighted ALS (MCR-WALS) was applied to the continuous, time series PTR-ToF-MS data from three field campaigns on Vancouver Island, BC for source apportionment. Variable selection based on signal-to-noise ratios was used to reduce the mass list while retaining the observed *m/z* that capture changes in the mixture of VOCs measured, improving model results, and reducing computation time. Both point (*e.g.*, anthropogenic hydrocarbon emissions, pulp mill emissions) and diffuse (*e.g.*, VOCs from forest fire smoke) VOC sources were identified in the data, and were apportioned to determine their contributions to the measured samples. The data analyzed captured fine scale changes in the ambient VOCs present in the air, and geospatial maps of each individual source, and of the source apportionment were used to visualize the distribution of VOC sources across the sampling area. This work represents the first use of MCR-WALS to identify and apportion ambient VOC sources based on continuous PTR-ToF-MS data measured from a moving vehicle. The methods described can be applied to larger scale field campaigns for the source apportionment of VOCs across multiple days to capture diurnal and seasonal variations. Identifying spatial and temporal trends in the sources of VOCs at the regional scale can help to identify pollution 'hot spots' and inform evidence-based public policy for improving air quality.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
AIC	Akaike information criterion
BIC	Bayesian information criterion
BTEX	Benzene, toluene, ethylbenzene, xylenes
CI	Chemical ionization
CIT	Cylindrical ion trap
CMB	Chemical mass balance
CV	Cross-validation
DBSCAN	Density-based spatial clustering of applications with noise
Dilbit	Diluted bitumen
DMDS	Dimethyl disulphide
DMS	Dimethyl sulphide
EI	Electron ionization
eV	Electron volt
FEP	Fluorinated ethylene propylene
FID	Flame ionization detection
FT-IR	Fourier transform infrared spectroscopy
GC	Gas chromatography
GC-FID	Gas chromatography-flame ionization detection
GC-MS	Gas chromatography-mass spectrometry
GHG	Greenhouse gas
GMM	Gaussian mixture model
GPS	Global positioning system
GWP	Global warming potential
HR	High resolution
IQR	Interquartile range
IR	Infrared
kml	Keyhole markup language
kNN	k-nearest neighbours
m/z	Mass-to-charge ratio
MCR	Multivariate curve resolution
MCR-ALS	Multivariate curve resolution-alternating least squares
MCR-WALS	Multivariate curve resolution-weighted alternating least squares

MIMS	Membrane introduction mass spectrometry
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NIR	Near infrared
ncps	Normalized counts per second
OSPW	Oil sands process-affected water
PA	Proton affinity
PBL	Planetary boundary layer
PCA	Principal component analysis
PCR	Principal component regression
PC	Principal component
PDMS	Polydimethylsiloxane
PFA	Perfluoroalkoxy alkane
PLS	Partial least squares
PM	Particulate matter
$PM_{2.5}$	Fine particulate matter (diameter < 2.5 μm)
PMF	Positive matrix factorization
ppb	parts-per-billion
ppb_v	parts-per-billion by volume
ppm	parts-per-million
ppt	parts-per-trillion
ppt_v	parts-per-trillion by volume
PTR-MS	Proton-transfer reaction mass spectrometry
PTR-ToF-MS	Proton-transfer reaction time-of-flight mass spectrometry
QIT	Quadrupole ion trap
sccm	Standard cubic centimeters per minute
SIM	Selected ion monitoring
SIMCA	Soft independent modeling of class analogy
S/N	Signal-to-noise ratio
SOA	Secondary organic aerosol
SSR	Sum of squared residuals
SVD	Singular value decomposition
SVOC	Semi-volatile organic compounds
TIC	Total ion current
ToF	Time-of-flight

TRS	Total reduced sulphur
VOC	Volatile organic compound
WSSR	Weighted sum of squared residuals

Acknowledgments

First of all, my deepest gratitude goes to my supervisor, Dr. Erik Krogh. The undergraduate courses I took with you ignited my love of Chemistry, and you were instrumental in my decision to pursue graduate studies when you pitched me a research idea that would combine my background in Chemistry and Math. I can't thank you enough for the amazing opportunities you provided me, your unwavering support, your patience, and your continual belief in me. Thank you to my co-supervisor, Dr. Tom Fyles, and to Dr. Chris Gill, for your support, guidance, and feedback throughout my degree. To Dr. Jay Cullen, thank you for being on my committee and providing an outside perspective on my work. Financial support from many sources made this research possible including NSERC, CFI, and multiple community partners. Thank you for helping turn ideas into reality.

Thank you to the members of the Applied Environmental Research Labs (AERL), past and present, who have taught me so much, supported me through challenging times, and made working in the AERL such a great experience. In particular, thank you to Dr. Nick Davey who taught my very book smart self many practical skills in the operation, repair, and maintenance of mass spectrometers. I will never forget the day we completely dismantled Griffin. To the mobile lab crew, Erik, Nick, Jon Davidson, and Trevor Michalchuk, I couldn't have asked for a better group to spend long field sampling days with. To everyone with a driver's license who drove the mobile lab on field campaigns, my dissertation would not have been possible without you. To Dr. Ryan Bell, your MATLAB tutorials made data processing so much easier. Thank you all so much, it has been such a pleasure to work with all of you.

To the VIU Chemistry Department, thank you for providing such a warm and welcoming atmosphere, and for always having your doors open for my many teaching and research questions.

Thank you to the teachers and students at the Copenhagen School of Chemometrics. The six weeks I spent learning and adventuring with you are one of the highlights of my academic career.

Christine, Steve, Eilidh, Kate, Megan, Britney, and Marian, you have been consistent cheerleaders for me throughout my degree, and your encouragement as I finished my dissertation during the pandemic was invaluable. Christine, thank you for not only supporting me as a friend, but also for providing valuable feedback on my dissertation.

Finally, thank you to my family. Mum, you are a bastion of strength and resilience that inspires me every day. Thank you for always encouraging me to pursue what I love. Bryanna and Diana, I'm so

proud to call you my sisters, and your unwavering support means so much to me. Pete, thank you for everything. You have been by my side every step of the way through my successes and failures, and it hasn't always been easy. I would not be where I am today without you. Thank you.

Dedication

For Dad, who I miss every day, Mum, and Pete.

Chapter 1 - Motivation

1.1 Air quality

Air quality has a significant impact on human and environmental health. Good air quality refers to air with low levels of air-borne pollutants, and air quality degrades as pollution levels increase. Exposure to air pollution can cause short-term health impacts, including eye irritation, coughing, and wheezing, as well as long-term health effects, such as respiratory and cardiovascular diseases, cancer, cognitive and behavioral impacts, and neurological effects.[1–8] These health impacts are not distributed evenly across the population. Children, elderly people, and those with pre-existing conditions (*e.g.,* asthma) are more susceptible to poor air quality.[2,3] Additionally, higher exposure to air pollution is observed among lower socioeconomic groups.[1,3] Ambient air pollution contributes to over 4 million premature deaths per year, globally,[9] with an estimated 15,300 deaths occurring in Canada.[10] Air-borne pollutants have a host of adverse environmental effects (*e.g.,* vegetation damage,[1,11,12] reduced biodiversity,[13,14] acid rain production[1,3,12]). Air pollution also has a large economic impact (*e.g.*, medical care, lost work hours), costing the Canadian economy billions of dollars per year.[10] Small reductions in the concentration of air pollutants can have large economic benefits.[10] On a global scale, the economic cost of air pollution is projected to reach 1 % of global gross domestic product (GDP) by 2060.[15]

Many chemical species contribute to poor air quality.[1,6,16] Some are primary pollutants, emitted directly into the atmosphere by both natural and anthropogenic processes, while others are secondary pollutants that are produced through chemical and physical transformations of primary pollutants.[1,17] In many jurisdictions regulations and air quality standards are in place for some major air pollutants, including ground-level ozone (O_3), fine particulate matter ($PM_{2.5}$), and nitrogen dioxide (NO_2), all of which are often measured on a regional scale.[1,18] These pollutants are used in Canada to generate an Air Quality Health Index,[19] informing air quality health advisories, and a record of these pollutant concentrations can be used to inform public policy. Reducing ambient concentrations of these pollutants results in measurable improvements in health outcomes,[20,21] but reducing their concentrations is rarely straightforward due to the complex chemical processes leading to their formation. In particular, O_3 and NO_2 are secondary pollutants formed through photochemical oxidation reactions involving nitrogen oxides (NO_x) and volatile organic compounds (VOCs), while $PM_{2.5}$ can be emitted directly from a source (*i.e.,* primary pollutant), such as biomass combustion and vehicle emissions, or produced from the

transformation of primary pollutants, such as when secondary organic aerosol (SOA) is formed through the oxidation and subsequent condensation of VOCs.[22] Anthropogenic VOC emissions lead to the formation of 24.6–70.0 Tg/year of SOA.[23]

Pollutant distribution varies widely over time and space[24–29], and air pollutants are typically measured at fixed-site air quality monitoring stations.[30,31] In addition to meteorological data, these stations typically report hourly concentrations of atmospheric pollutants, such as O_3, NO_2, PM, carbon monoxide (CO), sulphur dioxide (SO_2), lead (Pb), and/or total reduced sulphur (TRS).[30,31] Although VOCs are known to have both direct and indirect impacts on air quality, they are not routinely measured at most air quality monitoring stations. This is because there are many individual VOCs to consider, but also because analytical techniques for VOC measurements can be expensive, laborious, and until recently, discontinuous.[16,32,33] However, the emergence of modern direct mass spectrometry (MS) instrumentation provides a wealth of molecular level information about the composition and distribution of atmospheric VOCs.[34–37] The continuous operation of these instruments in a moving vehicle provides multivariate time series data at unprecedented temporal and spatial resolution. This dissertation examines the use of chemometric techniques applied to full scan mass spectral data to identify, apportion, and map VOC sources at local and regional scales.

1.2 Tropospheric chemistry

1.2.1 Fate and distribution of trace gases in the atmosphere

The Earth's atmosphere is thermally stratified with the portion closest to the surface known as the troposphere. It contains about 80 % of the mass of the atmosphere and is most susceptible to surface air pollutants sources. Depending on meteorological conditions and topography, a planetary boundary layer (PBL) can form in the lower portion of the troposphere. These PBLs are variable in height and duration, and act to further concentrate air pollutants emitted from the surface.[17] Trace gases are released from anthropogenic, biological, and geological processes.[17] Sources can be mobile (*e.g.*, transportation) or stationary (*e.g.*, factories, power plants), point (*e.g.*, industrial stacks) or diffuse (*e.g.*, cities, farmland), and anthropogenic or natural.[38] The wide range in the atmospheric lifetime of pollutants (seconds to centuries) has a direct impact on how far they will travel from the source and consequently, their spatial variability. Highly reactive species (*e.g.,* OH, NO_3, HO_2) are short-lived (seconds to minutes), and those with moderate lifetimes (hours to weeks) are typically dispersed on a

local or regional scale (*e.g.,* isoprene, dimethyl sulphide (DMS), NO_x). On the other hand, long-lived species (*e.g.,* methane, trichloroethane, carbon dioxide (CO_2)) are well mixed in the troposphere globally. The relationship between tropospheric lifetime and spatial scale of selected atmospheric pollutants is summarized in Figure 1.1.[17] As a result of differences in sources and removal processes, the chemical composition of the trace constituents in the troposphere is widely variable on the local, regional, and global scale.[17]

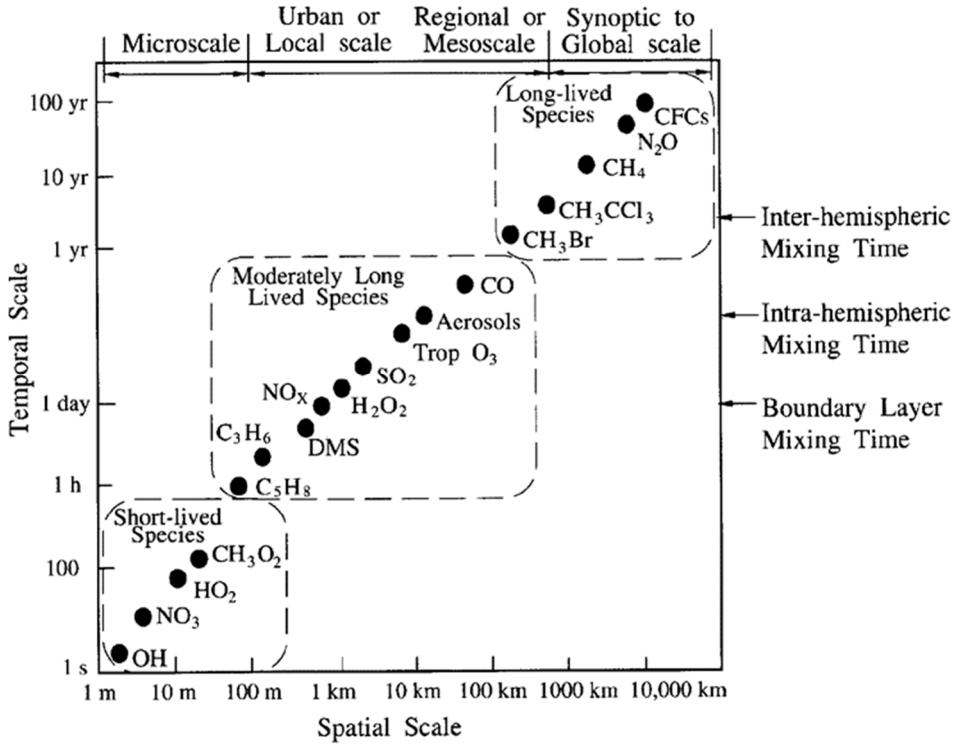


Figure 1.1: Spatial and temporal scales of variability for atmospheric constituents. (Reproduced with permission from Atmospheric Chemistry and Physics: From Air Pollution to Climate Change.[17] Copyright © 2016. John Wiley & Sons, incorporated. All rights reserved).

1.2.2 Volatile organic compounds

In general, VOCs are a collection of relatively small organic compounds (typically fewer than 15 carbons) with vapour pressures greater than 10 Pa at ambient conditions. They can be emitted into the atmosphere from a wide variety of sources including both anthropogenic (*e.g.,* combustion processes, fugitive emissions, paints and solvents, waste treatment, industrial processes) and natural (*e.g.,* vegetation, forest fires, oceans) sources.[24,39] Atmospheric concentrations can typically range from parts-per-trillion (ppt_v) to part-per-billion by volume (ppb_v).[24,40,41] VOCs include many compound classes with molecular weights typically below 250 g/mol. This includes hydrocarbons (*e.g.*, alkanes, alkenes, alkynes,

isoprene, monoterpenes, aromatics), oxygenated species (*e.g.*, alcohols, aldehydes, carboxylic acids, esters, ethers, ketones), halogenated compounds (*e.g.* chloroform, chlorofluorocarbon, hydrochlorofluorocarbons, trichloroethylene), nitrogen-containing species (*e.g.*, nitriles), and sulphur compounds (*e.g.*, DMS, methanethiol).[42] VOCs are involved in a complex series of atmospheric transformation reactions with each compound having different ozone and SOA formation potentials.[28,43] Consequently, local and regional sources of VOCs (*e.g.*, industries, traffic, agriculture, forests) in an air shed will influence the extent of secondary pollutant formation. A few representative example VOCs are shown in Figure 1.2. Table 1.1 summarizes the average ambient concentration, Canadian provincial ambient concentration guidelines (if applicable), example sources, and other information of selected VOCs.

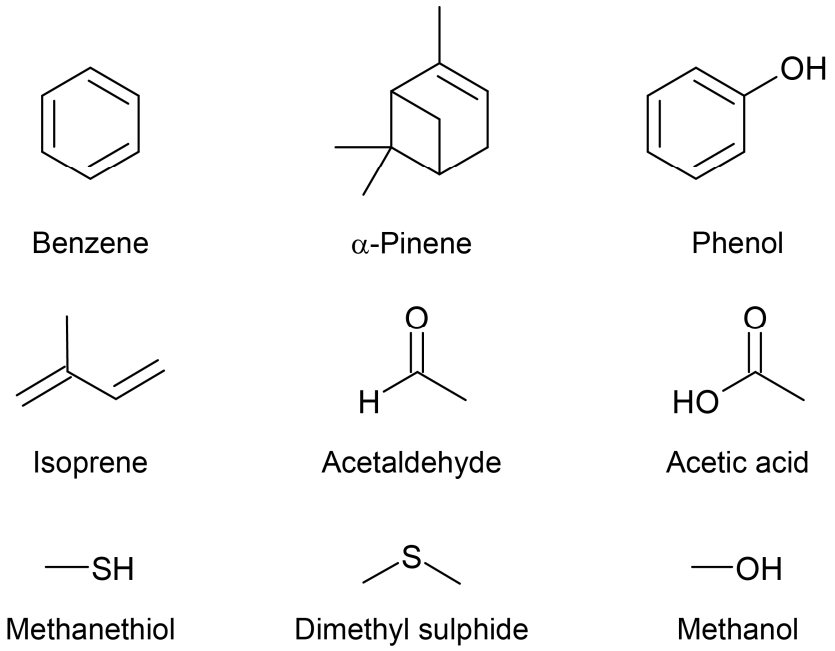


Figure 1.2: Chemical structures for some example VOCs.

Table 1.1: Some example VOCs, their average ambient concentrations, air quality guidelines, and sources.

Chemical Species	Average Conc (ppb_v)	Provincial Guidelines (ppb_v)[16]	Example sources	Other information
Acetaldehyde	0.90[a] 0.50–1.9[b] 5.5–6.8[c] 7.4[d]	1 Hour (Alberta) – 50 0.5 hours (Ontario) – 278 24 Hours (Ontario) – 278	Plants[44] Forest fires[44] Combustion[44]	Carcinogen[6] US EPA Hazardous Air Pollutant[45]
Acetic acid	N/A	N/A	Biomass burning[46] Dairy[46] Solvents[46] Forest fires[46] Composting[46]	
Acetone	0.2–3[47] 7.5[d]	1 Hour (Alberta) – 2400	Solvents[48] Biomass burning[48] Terrestrial vegetation[48] Plant decay[47] VOC oxidation[47]	
1,3-Butadiene	0.066[a] 1.64[c]	24 Hours (Ontario) – 4.5 1 year (Ontario) – 0.90 1 year (Quebec) – 0.14	Motor vehicle exhaust[49] Manufacturing and processing facilities[49] Forest fires or other combustion[49] Cigarette smoke[49]	Carcinogen[6] US EPA Hazardous Air Pollutant[45]
Benzene	0.25[a] 0.19[b] 6.8[c] 2.8[d]	1 Hour (Alberta) – 9.4 1 Year (Alberta) – 0.94 24 hours (Ontario) – 0.72 1 years (Ontario) – 0.14 24 hours (Quebec) – 3.1	Vehicle emissions[50] Industrial emissions[50] Biomass burning[50]	Carcinogen[6] US EPA Hazardous Air Pollutant[45]
Carbon Tetrachloride	0.088[a] 0.084[b] 0.18[c] 0.085[e]	24 Hours (Ontario) – 0.38 1 year (Quebec) – 0.16	Fluorocarbon production[46] Petroleum refining[46] Chemical manufacturing[46]	Carcinogen[6] US EPA Hazardous Air Pollutant[45]
Dimethyl sulphide	N/A	N/A	Landfills[46] Dairies[46] Pulp and Paper Industry[46]	

Table 1.1: Some example VOCs, their average ambient concentrations, air quality guidelines, and sources, continued.

Chemical Species	Average Conc (ppb_v)	Provincial Guidelines (ppb_v)	Example sources	Other information
Ethylbenzene	0.055[b] 8.5[c] 0.95[d]	1 Hour (Alberta) – 460 10 mins (Ontario) – 438 24 hours (Ontario) – 230 4 mins (Quebec) – 171 1 year (Quebec) – 46[16]	Styrene manufacturing[49] Consumer products[49] Fuel[49] Pesticides[49] Solvents, glues, varnishes, paints[49] Tobacco smoke[49]	US EPA Hazardous Air Pollutant[45]
Formaldehyde	2.6[a] 1.2[b] 12[c] 9.1[d]	1 Hour (BC) – 49 1 Hour (Alberta) – 53 1 hour (Manitoba) – 49 24 hours (Ontario) – 53 15 mins (Quebec) – 30	VOC oxidation[50] Biomass burning[50] Fuel combustion[51] Industrial emissions Preservative	Carcinogen[6] US EPA Hazardous Air Pollutant[45]
Isoprene	N/A	N/A	Plants[48,50] Animal Agriculture[46] Biomass burning[46] Gasoline exhaust[46]	
Methanethiol	N/A	N/A	Landfills[46] Pulp and Paper industry[46]	
Methanol	N/A	1 Hour (Alberta) – 1985 24 hours (Ontario) – 3054 4 mins (Quebec) – 4200 1 year (Quebec) – 38	Plants[48] VOC oxidation[50]	US EPA Hazardous Air Pollutant[45]
Methyl tert-butyl ether	0.22[c]	24 hours (Ontario) – 1943 4 mins (Quebec) – 50	Gasoline exhaust[46] Lawn mowers[46] Diesel exhaust[46] Gasoline evaporation[46]	US EPA Hazardous Air Pollutant[45]

Table 1.1: Some example VOCs, their average ambient concentrations, air quality guidelines, and sources, continued.

Chemical Species	Average Conc (ppb_v)	Provincial Guidelines (ppb_v)	Example sources	Other information
Naphthalene	0.11[b] 11.8[c]	10 mins (Ontario) – 9.5 24 hours (Ontario) – 4.3 4 mins (Quebec) – 38 1 year (Quebec) – 0.57	Industry[52] Combustion[52] Vehicles[52] Tobacco smoke[52] Biomass burning[52]	Carcinogen[6] US EPA Hazardous Air Pollutant[45]
α-Pinene	N/A	N/A	Vegetation[48] Green waste composting[46] Wood chips[53]	
Styrene	0.016[b] 19[c]	1 Hour (Alberta) – 51 24 hours (Manitoba) – 94 24 hours (Ontario) – 94 1 Hour (Quebec) – 35	Gasoline exhaust[46] Biomass burning[46] Wastewater treatment plants[46]	US EPA Hazardous Air Pollutant[45]
Tetrachloroethylene	0.017[a]	24 hours (Ontario) – 53 1 year (Quebec) – 0.30	Dry cleaning[46] Paint removers[46] Landfills[46] Degreasers[46] Solvents[46]	Carcinogen[6] US EPA Hazardous Air Pollutant[45]
Toluene	0.40[b] 65[c] 7.0[d]	1 Hour (Alberta) – 500 24 hours (Alberta) – 106 24 hours (Ontario) – 532 4 mins (Quebec) – 160	Solvents[50] Vehicle emissions[50]	US EPA Hazardous Air Pollutant[45]
Xylenes	0.23[b] 38[c] 4.4[d]	1 hour (Alberta) – 530 24 hours (Alberta) – 161 10 mins (Ontario) – 691 24 hours (Ontario) – 168 4 mins (Quebec) – 81 1 year (Quebec) – 4.6	Petroleum, coal and wood tar[49] Solvent[49]	US EPA Hazardous Air Pollutant[45]

[a]United States Environmental Protection Agency (US EPA) 2013 monitoring data[54]
[b]Canadian National Air Pollutants Survey (NAPS) 2009–2013 average[16]
[c]Canadian NAPS 2009–2013 maximum[16]
[d]Summer 1987 Southern California Air Quality Study (Converted to ppb from ppb of C)[17]

Some VOCs are known to be toxic[55], others are associated with nuisance odours[56,57], and some are powerful greenhouse gases (GHGs).[58] For example, benzene, formaldehyde, and 1,3-butadiene are known human carcinogens.[59] Exposure to some VOCs can lead to irritation of the eyes, nose, and throat, headaches, central nervous system damage, liver damage, kidney damage, loss of coordination, and nausea.[60] Odorous VOCs, associated with certain processes and industries (*e.g.,* pulp and paper production, waste management, combustion, petroleum refineries), can impact quality of life and may also trigger physical symptoms such as headache and nausea.[61,62] The United States (US) Environmental Protection Agency (EPA) has identified 187 pollutants as air toxics or hazardous air pollutants for regulation under the Clean Air Act.[45] Ninety percent of the US's incremental cancer risk in outdoor air is attributed to ten of the pollutants on this list, seven of which are VOCs and are included in Table 1.1.[6] VOCs that can act as GHGs (particularly halogenated compounds) have global warming potentials (GWP) orders of magnitude higher than CO_2 on a per mass basis.[58] However, as a group, the main influence of VOCs on climate is as precursors in the formation of ground-level ozone (a GHG)[63] and SOAs.[64]

Global anthropogenic emissions of non-methane VOCs were estimated to be 170 Mt in 2012, and have been increasing.[65] On a global scale, biogenic VOCs such as isoprene and monoterpenes released from vegetation constitute roughly 90 % of all emissions.[36,48] However much focus is put on anthropogenic VOCs as many of the toxic compounds have anthropogenic sources, and anthropogenic emissions can be reduced through 'greener' practices, including industrial emissions controls, use of lower VOC products, cleaner fuels, and cleaner technology. While decreases in anthropogenic VOCs have been observed in recent years in Canada, the US and the European Union, increases have been observed in other countries, such as China.[1,18,66]

The contributions of different sectors to anthropogenic VOC emissions in Canada are presented in Figure 1.3. The oil and gas industry, paints and solvents, domestic heating, transportation, and manufacturing are the major VOC source categories in Canada.[18] The mixture of individual VOCs emitted by each source will differ, and can be influenced by a number of factors. VOCs can travel in the atmosphere, and the VOCs present at a given location at a given point in time can be emitted from multiple sources. Identifying which compounds have been emitted from which sources, and in what proportions, can be done through source apportionment studies, and provide a useful means to inform development of 'greener' technologies, emission reduction strategies, and policy instruments to improve air quality.

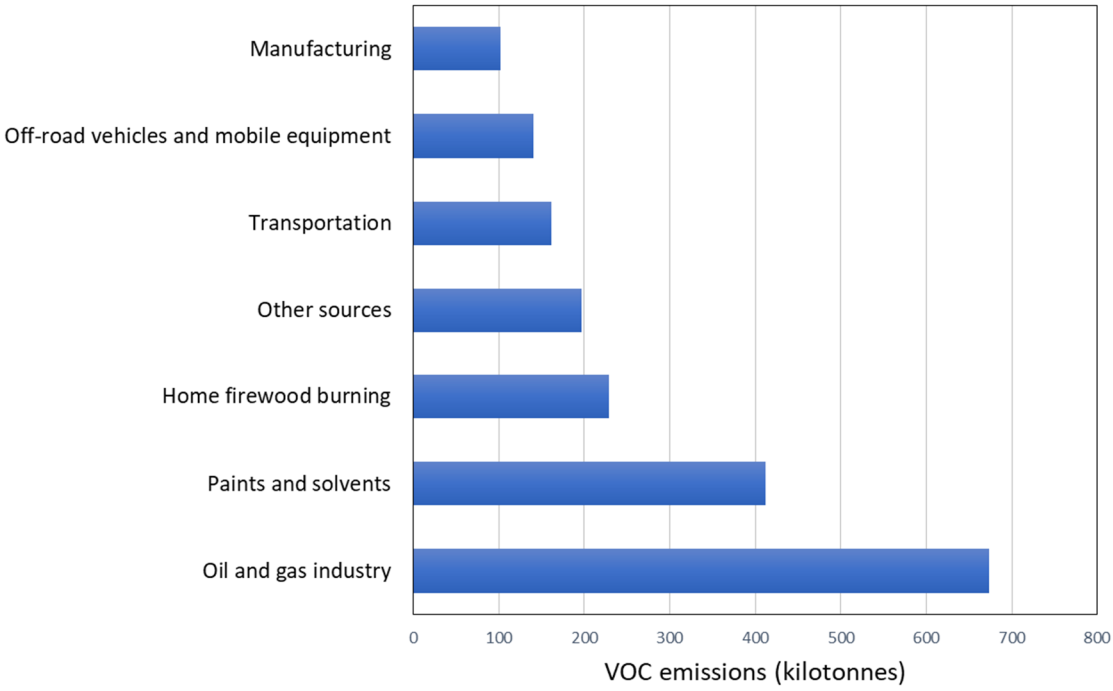


Figure 1.3: Canadian VOC emissions by sector in 2018.

1.2.3 Atmospheric transformation reactions

The troposphere is an oxidizing environment, which drives the important chemical transformations of VOCs that contribute to the formation of ground-level O_3 and SOA, both of which are known to be acutely and chronically toxic.

SOA is formed when VOCs are oxidized through a series of steps leading to increasingly polar compounds, which can then condense into microdroplets, known as aerosols. VOC oxidation reactions are complex but are generally initiated by hydroxyl radicals (OH). The reaction sequence typically leads to the formation of a carbon centred radical, which quickly reacts with O_2 to form alkylperoxyl species (ROO). Subsequent reactions result in the formation of more oxidized products, such as the ketones and aldehydes, that are more polar and consequently less volatile than their precursors.[67] The resulting SOAs droplets contribute to $PM_{2.5}$ and act as cloud condensation nuclei.[23,64,68,69]

Ground-level O_3 is produced through complex photochemical reactions involving both VOCs, NO_x, and sunlight. While described briefly in the following reaction sequence, a more comprehensive discussion of tropospheric ozone production is described elsewhere.[11,17,70] At elevated temperatures (*e.g.*, internal combustion engines) NO is produced from N_2 and O_2:

$$N_2 + O_2 \rightleftharpoons 2\,NO \qquad \text{Reaction 1.1}$$

NO oxidation produces NO_2 (where [O] = HO_2, RO_2, O_3, etc):

$$NO + [O] \rightarrow NO_2 \qquad \text{Reaction 1.2}$$

NO_2 can then photolyze with solar radiation (λ< 424 nm):

$$NO_2 + h\nu \rightarrow NO + O \qquad \text{Reaction 1.3}$$

Finally, O_3 forms in the presence of the third body (usually O_2 or N_2) via:

$$O_2 + O + M \rightarrow O_3 + M \qquad \text{Reaction 1.4}$$

The concentration of O_3 produced depends on the amounts and relative ratio of NO_x and VOCs. Contours of O_3 concentration can be plotted as a function of NO_x and VOC concentrations as shown in Figure 1.4, known as an ozone isopleth plot.[70] At high NO_x to VOC ratios, O_3 formation is limited by VOC concentrations. On the other hand, O_3 formation is limited by NO_x concentrations when the VOC to NO_x ratio is high. The diagonal line on the isopleth separates VOC and NO_x limited regions. This has important consequences because above the diagonal, a reduction in NO_x emissions will have little to no effect on reducing ground-level O_3 production and may lead to increases in O_3 concentrations. Similarly, under NO_x limited conditions, decreasing the VOC concentrations will have little impact on O_3 production.[17,71]

Ground-level O_3 concentrations typically vary from 20–60 ppb_v in urban environments, but can exceed values of 200 ppb_v in heavily polluted cities.[17] In Canada, the 8-hour ground-level O_3 air quality standard is 62 ppb_v.[72] Tropospheric O_3 concentrations are known to exhibit high spatial and temporal variation, due to the variability of precursors and conditions required in their formation.[71,72] In some urban areas, higher O_3 concentrations are observed on the weekend as compared to weekdays, despite less commuter traffic on the road and the reduced emissions of NO_x and VOC precursors. This so-called 'weekend ozone effect' results from a greater reduction in the NO_x concentrations than VOC concentrations. With less NO_x available on the weekend to titrate out the O_3, concentrations of O_3 can actually increase. In many areas, O_3 abatement strategies have focused solely on reducing NO_x concentrations, which has pushed urban ozone formation into a VOC limited regime.[71] As a result, future O_3 control strategies should target reductions in VOC concentrations.

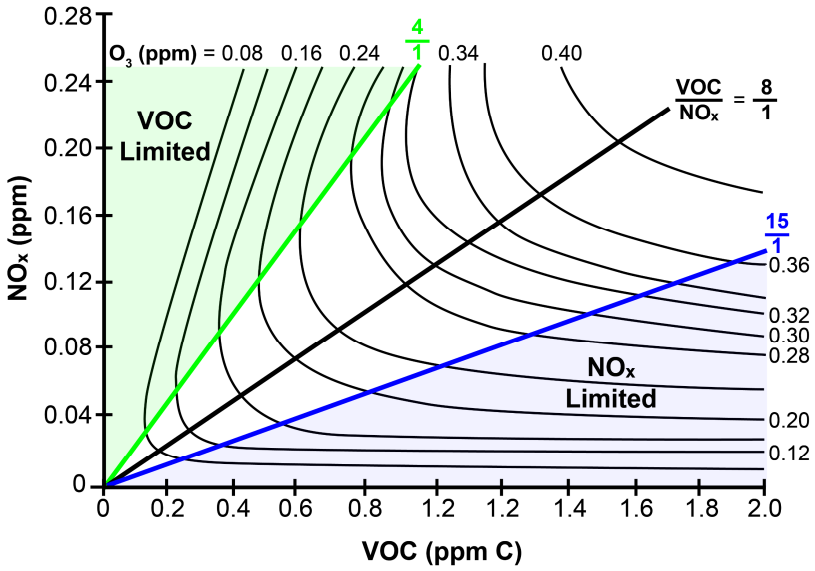


Figure 1.4: Typical ozone isopleth. At a high VOC:NO_x ratio ozone formation is NO_x limited. Adapted from Rethinking the Ozone Problem in Urban and Regional Air Pollution.[73]

Informed ozone reduction strategies require knowledge of both NO_x and VOC concentrations. While NO_x measurements from air quality monitoring stations are used for this purpose, comparable VOC data is not generally available with the same geographic coverage.[16,32,72] However, advances in analytical instrumentation have made measurements of VOCs at ambient concentrations (ppt_v-ppb_v) more feasible. In locations where VOC data is available, the data collected has been effectively used to track how the concentrations of toxic VOCs vary over time as well as identifying the sources of VOCs in an area based on the mixture of molecules that were observed.[74–76] Direct MS techniques can continuously measure VOCs at ambient concentrations using both targeted and non-targeted strategies.[77–79] Modern smaller instruments can also be deployed in a moving vehicle to provide time and location-resolved VOC data.[26,80–85] The use of this technology to identify and map VOC sources/source contributions on a regional scale is described in this dissertation.

1.3 Atmospheric VOC measurements

1.3.1 Stationary measurements

While aggregate measurements of total VOCs can be accomplished using sensor technology, such as photoionization detectors, they do not provide molecular level information about which VOCs are present. As such, they cannot identify or quantify specific air toxics (*e.g.,* benzene, acrolein), nor do they provide information that differentiates anthropogenic and biogenic sources. In general, speciated VOC measurements that are able to measure many individual VOCs are labour intensive and require more advanced instrumentation.[32] Recently, some programs such as the National Air Toxics Trends Stations (NATTS) network in

the United States[32], and Canada's National Air Pollution Surveillance (NAPS) program[16], do measure targeted VOCs in discrete, 'point-in-time' samples at strategic locations. Air samples are generally collected in evacuated canisters or onto sorbent tubes, and subsequently transported back to a laboratory where they are pre-concentrated and analyzed for a suite of VOCs. Samples are analyzed off-line using a range of instrumental techniques including gas chromatography-flame ionization detection (GC-FID) and gas chromatography-mass spectrometry (GC-MS), producing speciated concentration data in the ppt_v range.[86] This type of sample collection and analysis has been done at government air quality monitoring stations[16,32,75,76,87], as well as separate sampling sites of interest.[88–90] These techniques are sensitive and selective, but the analysis is often limited by the cost of collection, transportation, and analysis, and can take weeks or months to report. Additionally, the limited availability of this data constrains our ability to fully understand temporal and spatial trends in VOC concentrations.

More recently, instrumentation capable of continuous VOC analysis at fixed sites, such as online, automated GC-FID[91–95] or GC-MS (or combination[96]) systems, and proton-transfer reaction mass spectrometry (PTR-MS) have been providing time-resolved VOC data. These instruments produce VOC data with high temporal resolution, on the order of seconds (PTR-MS) to hours (GC systems), allowing the data to be interrogated much more quickly compared to intermittent discrete samples subsequently analyzed in a laboratory.[92,97–99] In a chromatographic analysis (*e.g.,* GC-MS) analytes are separated before being analyzed by the mass spectrometer, while in a direct MS analysis (*e.g.,* PTR-MS) the analytes are measured as a mixture, which allows changes in the composition of VOCs in the air to be captured at a much higher frequency. Stationary monitoring data may capture a range of spatial scales, ranging from meters to several kilometers.[100] Some pollutants (particularity those with short atmospheric lifetimes) have high spatial variation within an area, leading to efforts to mobilize air monitoring equipment to identify sources and understand fate.[29] These field campaigns have provided higher spatially and temporally resolved chemical measurements for targeted analytes, such as black carbon, NO_x, O_3, PM, CO, CO_2, toluene, and benzene.[29,101,102]

The complexity of the suite of VOC analytes and their low concentrations makes the use of direct MS techniques well suited for their analysis. Mass spectrometers are very sensitive, with PTR-ToF-MS systems capable of measurements down to concentrations in the ppt_v range in ambient air samples.[37] Furthermore, MS instruments can be used for both targeted and non-targeted analysis. In a mass spectrometry experiment the sample is introduced into an ion source, where the analytes are ionized. The ions are sorted in a mass analyzer by their mass-to-charge ratio (*m/z*) before passing to a detector. Mass spectrometers can be used to measure hundreds to thousands of different compounds, and depending on the mass resolution of the mass analyzer can

resolve molecules with different nominal masses (*e.g.,* quadrupoles) all the way to molecules with different exact masses (*e.g.,* Fourier-transform ion cyclotron resonance) allowing the chemical formula of each ion to be uniquely determined. Mass spectrometers are usually operated in a laboratory setting, however recent efforts to adapt the MS instrumentation for operation on mobile platforms, are intensifying. When operated in a moving vehicle, direct MS techniques such as membrane introduction mass spectrometry (MIMS) and proton-transfer reaction time-of-flight mass spectrometry (PTR-ToF-MS) continuously measure VOCs in ambient air, allowing for the collection of VOC data with high spatial (meters) and temporal (seconds) resolution.

1.3.2 Mobile VOC measurements

Environmental monitoring campaigns have used the miniaturization and ruggedization of MS systems[103] to enable their deployment in ground[26,80,108,81–83,85,104–107]-, water[109,110]-, and air[111–114]-based field campaigns. Much of this work relies on the use of direct MS of whole samples, without pre-concentration or chromatographic separation of individual molecules. While this can limit the selectivity to analyze specific molecules, it enables real-time chemical analysis and high temporal resolution. Mounting direct MS instruments on a moving platform allows for geospatial mapping of atmospheric VOCs at spatial scales ranging from meters to hundreds of kilometers. Since direct MS does not employ chromatographic separation, full scan mass spectra contain the superposition of the mass spectra of all measured VOCs, effectively 'fingerprinting' the sample. While MIMS and PTR-ToF-MS systems have been recently used for temporally and spatially resolved quantitative, targeted analysis of ambient VOCs in the ppt_v–ppb_v range[80–82,107,114] qualitative analysis that takes advantage of the wealth of information present in the full scan MS VOC data for source identification is largely unexplored. This dissertation examines chemometrics techniques to discriminate between and map VOCs sources based on the distribution of peaks in full scan mass spectra collected from a moving vehicle.

1.3.2.1 Mobile lab developments at the Applied Environmental Research Labs (AERL)

The Applied Environmental Research Labs (AERL) at Vancouver Island University has been on the forefront of mobilizing mass spectrometry for geospatial mapping of tropospheric VOCs. Initially, this work focused on the use of gas-phase MIMS systems for targeted, quantitative VOC analysis. The first-generation mobile lab (2005–2011) used a bench-top GC-MS quadruopole ion trap (QIT) instrument (Thermo GCQ) modified with an 'in-house' constructed membrane interface as the sample inlet.[108] This MIMS system was first operated in a truck to measure VOCs in Vancouver and Victoria, British Columbia (BC).[115] Subsequent work had it mounted inside the cargo space of a moving van, and powered by a towed diesel generator. Targeted quantitative analysis of toluene, benzene, and ethylbenzene/xylenes compounds (BTEX), as well as 2-methoxyphenol (guaiacol), and naphthalene was achieved using MS/MS to provide additional molecular selectivity.[108] A second-generation mobile mass spectrometer (2011–2016) used a small, ruggedized cylindrical ion trap GC-MS system (Griffin 400,

FLIR), where again the GC was replaced with a custom built, thermally controlled MIMS interface and inline permeation chamber for continuously infusing an internal standard. This system was powered by 24 VDC deep cycle battery bank and fitted with a low pressure, mini helium supply for easier operation in the back of a sports utility vehicle. The instrument was used on field campaigns in the oil sands region of Alberta, Canada for the quantitative mapping of BTEX and related compounds in air and water samples.[26,82] In both generations of mobile MIMS, response factors between a toluene-d_8 internal standard and target analytes were used for quantitation. The development of the second-generation mobile MIMS system is described in detail by Dr. Nicholas Davey in his PhD dissertation.[116]

A third-generation mobile lab built late 2016–early 2017, consists of a research purposed, customized cargo van (Mercedes-Benz Sprinter 2500, 4x4). This facility includes a PTR-ToF-MS instrument (PTR-1000, Ionicon Analytik Ges.m.b.H, Innsbruck, Austria) as the main mass spectrometer, with the second-generation MIMS system available when MS/MS capabilities are required. The mobile lab also contains an NO_x analyzer (42i, Thermo Fisher, Waltham, MA, USA), O_3 analyzer (49i, Thermo Fisher, Waltham, MA, USA), GHG analyzer for methane (CH_4), CO_2, and water vapour (FGGA 30r, Los Gatos Research, San Jose, CA, USA), optical particle sizer (OPS 3330, TSI Inc., Shoreview, MN, USA), weather station (AIO 2, Met One Instruments, Grants Pass, OR, USA), global positioning system (A45 Antenna and R330 GNSS Receiver, Hemisphere, Scottsdale, AZ, USA), and a 2-axis ultrasonic anemometer and digital compass for wind speed and direction, (RM Young 86000, Campbell Scientific, Edmonton, AB, Canada). Data from the onboard chemical sensors (excluding the mass spectrometers) is logged and mapped in real-time using custom software (Underway, Beaver Creek Analytical, Lafayette, CO, USA). On-board instrumentation is powered by a lithium-ion battery bank that provides 4–8 hours of autonomous power. Measurement frequency is generally 1 Hz, apart from the O_3 and NO_x analyzers (10 seconds), and the MIMS system (variable depending on MS scans being used). At drive speeds of 50 km/h this results in a measurement every 14 m (VOCs, GHGs, PM) and 140 m (NO_x, O_3).

Early work employing mobilized MIMS instruments focused on the quantitative analysis of targeted VOCs. However, this uses only a portion of the direct MS data available in the full scan mass spectrum. As noted by *Bell et al,* the full scan MIMS spectra measured on-road during a field campaign differed based on nearby VOC sources, as indicated by the mass spectra measured downwind from a waste-water settling pond, and while driving on fresh asphalt.[26] The relative intensity of the signals in the full scan mass spectra can, in principle, be used to identify patterns within the data related to individual source classes and apportion the relative contributions of these sources. The field of chemometrics applies multivariate statistics to chemical data to discriminate samples with different chemical composition, perform calibrations on mixtures, and mathematically

deconvolute chemical measurements of mixtures into the sources contributing to the mixture. This dissertation describes the first use of these techniques to time-resolved full scan mass spectra obtained from MIMS and PTR-ToF-MS instruments operated on-road in a moving vehicle.

1.3.2.2 Membrane introduction mass spectrometry (MIMS) and proton-reaction time-of-flight mass spectrometry (PTR-ToF-MS)

MIMS employs a semi-permeable membrane, usually hydrophobic polydimethylsiloxane (PDMS), as an interface between the bulk sample (*e.g.,* air, water, or soil) and the mass spectrometer.[117] Hydrophobic analytes are pre-concentrated in PDMS and mass transport across the membrane is driven by a concentration gradient governed by Fick's Laws of diffusion.[118] For the majority of the research described in this dissertation that utilizes MIMS, instruments equipped with electron ionization (EI) and QIT mass analyzers were employed. The resulting full scan mass spectrum obtained is determined by membrane selectivity, the extensive fragmentation of the EI source, and the unit mass resolution achieved with the QIT. The use of a QIT also allows for MS/MS, increasing specificity when used for quantitative analysis.[82]

PTR-MS systems are usually constructed with quadrupole or time-of-flight (ToF) mass analyzers, with the ToF systems providing generally higher mass resolution than a QIT.[37] In PTR-ToF-MS experiments, H_3O^+ reagent ions are generated from water vapour by a hollow cathode discharge source. Ambient samples are mixed with the reagent ions in a drift tube with uniform electric field, where ion-molecule collisions occur.[119] If the proton affinity (PA) of the analyte (M) is higher than that of water, protonated molecular ions (MH^+) are formed through the following reaction:

$$M_{(g)} + H_3O^+_{(g)} \rightarrow MH^+_{(g)} + H_2O_{(g)} \qquad \text{Reaction 1.5}$$

Proton-transfer is a soft ionization method, producing mainly MH^+ ions with little fragmentation. The major constituents of air (*e.g.,* O_2, N_2, CO_2) have lower PAs than water, while the PAs for most trace organics are higher, allowing PTR-ToF-MS to measure most atmospheric trace organics, both polar and non-polar.[119] PTR-ToF-MS produces full scan data at a mass resolution in the 1,500–15,000 m/Δm range (depending on instrument),[119] however it lacks the ability to resolve isobaric species through MS/MS experiments.

1.4 Chemometrics

1.4.1 Introduction to chemometrics

Chemometrics applies multivariate statistics to large, complex chemical data sets containing many variables to uncover underlying patterns in the data and extract information that would be missed in a univariate analysis.[120,121] Chemometric analysis makes use of a wide range of multivariate statistical techniques

for a variety of applications, mainly focused on pattern recognition, multivariate calibration, and multivariate curve resolution. Pattern recognition techniques, such as principal component analysis (PCA), allow samples to be clustered and classified based on similarities (and differences) in their overall chemical composition. Multivariate calibration methods, such as partial least squares (PLS) regression, relate a property of the sample, such as concentration, to the response across multiple variables. Multivariate curve resolution (MCR) is focused on resolving individual compounds or components from a mixture.[121]

Multivariate chemical data sets come in many forms, with differing degrees of molecular specificity. Spectroscopic methods, such as near infrared (NIR), UV-Vis, Raman, and Fourier-transform infrared (FT-IR) produce multivariate data for samples where measured signal intensities are due to the presence of specific functional groups (IR) or chromophores (UV-vis) in the analytes of interest. Instrumentation using chromatographic techniques and/or mass spectrometry provide more specific chemical information where concentrations or signal intensities are attributed to a specific compound or isomer class. In multivariate datasets each sample is described by measurements for multiple variables such as the signal intensities associated with measured *m/z*, wavenumbers, retention times, or analyte concentrations. Chemometrics has been applied to many areas of study including, but not limited to, food science[122], forensic science[123,124], metabolomics[125], chemical imaging[126], process monitoring[127], medicine[128,129], and environmental analysis.[120,130]

In this dissertation, chemometrics will be applied to direct MS data. The data is collected continuously at a high frequency (*e.g.*, 1 Hz), with signal intensities being measured across a range of *m/z (e.g.*, 50–250). The time series full scan MS data captures changes in the mixture of VOCs present in the samples, producing data that is amenable to multivariate analysis. Additionally, the molecular level specificity provided by MS allows the results of chemometric analysis to be attributed to specific molecules or isomer classes. Knowing which molecules are, for example, leading to sample discrimination or present in the mass spectrum of a resolved VOCs source, is a powerful tool when interpreting results as it allows for databases (*e.g.*, EPA Speciate) to aid in the interpretation.

1.4.2 Principal component analysis

PCA is a dimensional reduction technique used to find underlying patterns in complex, multivariate data sets.[131] The data reduction in PCA is achieved by identifying new orthogonal, latent (not directly observed) variables, called principal components (PCs). PCs are linear combinations of the original variables, as shown in Equation 1.1 where each x_i is a variable, and each a_i is the loading for that variable on the calculated PC. The first PC (PC 1) lies in the direction of highest variance in the data set, and each subsequent PC describes a direction that accounts for the next highest variability in the data set.

$$PC = a_1x_1 + a_2x_2 + \cdots + a_nx_n \qquad \text{Equation 1.1}$$

The calculation of PCs allows the data set to be represented in fewer dimensions than the original data set, with little loss of information. For each sample, a score is calculated on each PC using the (preprocessed) measurements, and PCA scores plots are produced to visualize the data. In the scores plot, samples with similar scores will be plotted closer to one another than those that differ more significantly in their scores. This allows patterns in the data related to sample variance to be uncovered, which may not have been evident using univariate methods of interrogation. A loadings plot is constructed by plotting the loading of each of the variables on the calculated PCs to determine which measurements are driving the separation observed in the scores plot. Variables with higher loadings (positive or negative) of a particular PC have a larger impact on sample scores for that PC.[131] A graphical representation of PCA being applied to a dataset with three measured variables is shown in Figure 1.5. In the left panel, the dataset is plotted in three dimensions, one for each variable. PC 1 is shown in the direction of most variance, with PC 2 and 3 being orthogonal to PC 1 in the directions of second and third most variances, respectively. The score of the samples on each PC is their coordinate along the PC axis. In the right panel, the samples scores on PC 1 and 2 have been plotted, allowing the data to be visualized in fewer dimensions. This also uncovers two groups in the data that are not obvious in the original plot.

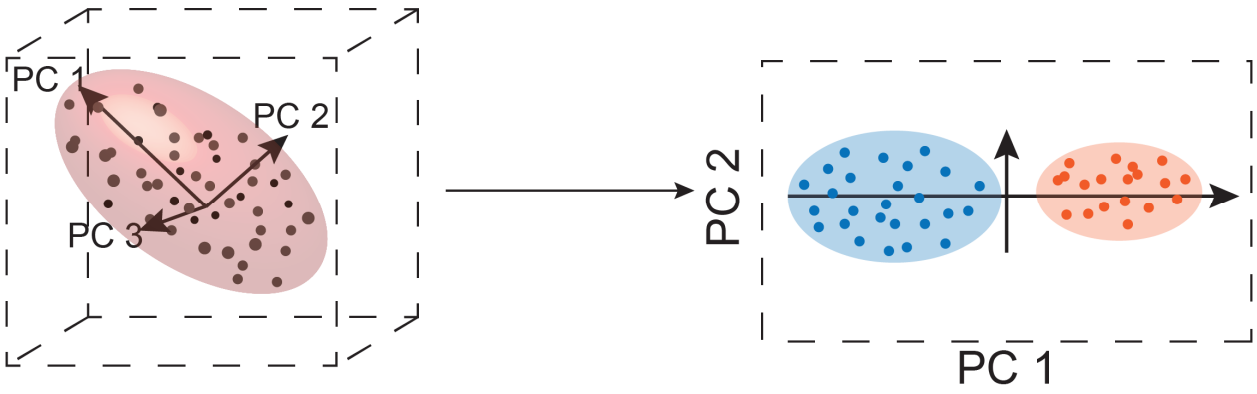


Figure 1.5: PCA schematic. A three-dimensional data set is reduced to two dimensions by plotting the data on PC 1 and PC 2, two new variables which are linear combinations of the original variables that capture the greatest variance. This allows two groups in the data to be identified.

Before performing PCA, the data is preprocessed. Most commonly, each variable is mean centered (given a mean of zero) before analysis. Additional preprocessing steps can include normalizing concentrations/signal intensities across each sample to remove concentration effects, or scaling each variable to have a standard deviation of one to normalize the variance of each variable.[131]

In environmental analysis, there is significant interest in determining the chemical profiles of real-world chemical sources contributing to a sample, as well as how much each source contributes to the sample based on the chemical concentrations detected in that sample. Although PCA is useful in finding patterns in complex datasets and identifying samples that are similar to one another, the scores and loadings calculated using PCA cannot be interpreted as real-world sources. Real-world sources are not orthogonal to one another and interpreting the PCs in this way results in sources with both negative concentrations of analytes being present in some of the source profiles and source contributions that are calculated. To determine sources impacting environmental samples, algorithms that impose non-negativity constraints on the solutions are used, preventing sources that contain negative concentrations and negative contributions.[132]

1.4.3 Source apportionment

To conduct source apportionment of environmental pollutants, mathematical receptor models are used, where the receptors are the sample sites receiving the pollutants. The mathematical method used depends on whether the possible sources of emission are known and have been well characterized. When all sources are known and characterized, linear mixing models such as the chemical mass balance (CMB) receptor model are used to determine the contribution of each source to the sample sites of interest.[133,134] However, in many environmental source apportionment problems, the sources are unknown or poorly characterized. In these cases, receptor models that do not require *a priori* knowledge of individual source profiles are used, for example multivariate curve resolution-alternating least squares (MCR-ALS) and Positive Matrix Factorization (PMF).[133]

In general, the receptor modeling problem is expressed as shown in Equation 1.2:

$$X = GF^T + E \qquad \text{Equation 1.2}$$

where $\boldsymbol{X}$ is the original data matrix (*m* by *n*), $\boldsymbol{F^T}$ is a matrix (*k* by *n*) where each row represents the chemical fingerprint of a source, $\boldsymbol{G}$ is a matrix (*m* by *k*) where each row represents the contribution of the sources to the sample, and $\boldsymbol{E}$ is a matrix containing the residuals. Depending on the algorithm used, an additional input matrix containing the uncertainties in the measurements is also required (*e.g.,* PMF or multivariate curve resolution-weighted alternating least squares (MCR-WALS)).[135]

If the sources and source compositions are known, the only unknown is matrix *G*, the source contributions. This is rarely the case however, and in many analyses the data matrix, number of samples, and number of variables are the only knowns. To determine the number of sources (*k*), the chemical composition of the sources (*F*), and the proportional contribution of each source to the samples (*G*), receptor models that do not require prior knowledge of the sources and samples are applied to the ambient chemical data set (*X*). A

graphical explanation of these self-modeling receptor models for full scan, continuous MS data are shown in Figure 1.6, where the data matrix of MS data is resolved into three sources (components). Each source is described by a mass spectrum (chemical composition) and time series (source contribution). These methods include PMF, MCR-ALS, MCR-WALS, and UNMIX.[133,136]

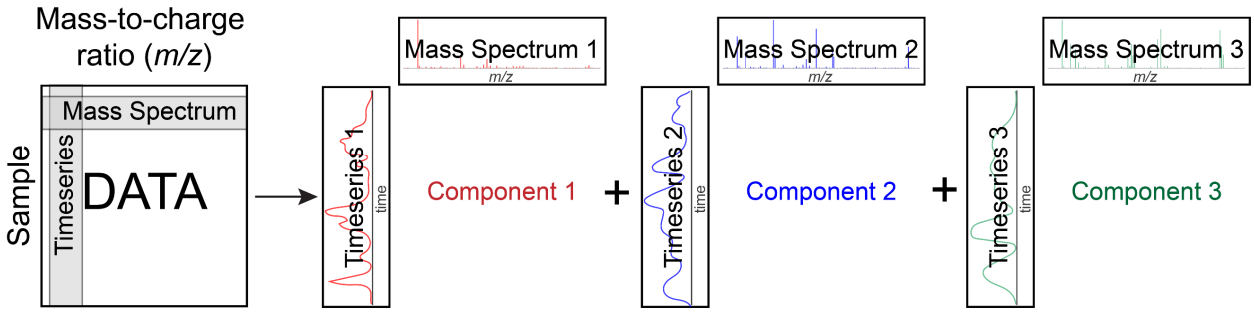


Figure 1.6: Receptor model schematic. The data matrix on the left is resolved into three sources (components), each with a mass spectrum and time series associated with it.

1.4.3.1 Source apportionment of VOCs

Receptor models have been used to apportion ambient VOCs in several cities, using ambient VOC concentrations measured at specific fixed location(s) over the course of days, months, or years.[75,78,142–144,87,92,93,137–141] GC-MS and GC-FID were commonly used for on-site VOC analysis, while some sites collected whole air samples for subsequent analysis, or used PTR-MS for on-site analysis.[75,95,137–141,143–145] UNMIX[137–141], PMF[75,95,137,138,143–147], and CMB[138,148] algorithms were used for source apportionment, and the identification of the resolved sources was informed by both the chemical composition and temporal profiles. Across these studies, vehicle exhaust (from gasoline and/or diesel), evaporative losses, and biogenic emissions were identified as sources, with industrial emissions and solvent usage also being identified in some cases.

The work described above used datasets containing concentration data for specific compounds in the receptor models, which requires a targeted list of compounds. Full scan mass spectral data contains information for compounds outside of a specific calibration suite, or for unidentified compounds that may be relevant to certain sources (*i.e.*, unknown unknowns). Receptor models have been applied to full scan data collected using aerosol mass spectrometry[149], a high resolution NO_3^--chemical ionization (CI) atmospheric pressure interface MS[79], PTR-ToF-MS[150], and an acetate HR-ToF-CI MS[151] which allows the full range of measured molecules to be considered in source identification. Most of these analysis have been conducted on datasets collected at stationary monitoring sites, with the exception of aerosol mass spectral data collected from a moving vehicle in Oakland, California[149], and acetate HR-ToF-Cl MS data of oxygenated VOCs collected from a marine vessel in the arctic[151] both of which recently employed PMF to identify sources.

1.5 Research goals

The on-road operation of direct MS systems allows complex chemical measurements to be made in real-time with temporal and spatial resolution that is not feasible using conventional methods. With the health impacts of poor air quality and inequities in the city-wide distribution of air pollutants[152–155], the merger of mobile MS with advanced chemometrics has the potential to become a key tool for the environmental monitoring of ground-level atmospheric chemistry, providing detailed information on the distribution of pollutants, and their sources, at a neighbourhood level. This dissertation aims to take concrete steps towards making this coupling a reality. The overarching goal of the work described is to enable the geospatial and temporal source identification and apportionment of ambient VOCs on a neighbourhood scale using the chemical information captured in the on-road measurement of continuous, time series direct MS data.

The application of chemometric techniques to analyze VOC measurements collected by mobile MS is presented, beginning with PCA for sample discrimination of constructed lab-based samples through to developing methodology to apportion ambient VOC sources measured from a moving vehicle. At the beginning of the work described in this dissertation, the AERL's second generation mobile MIMS system was in operation, with the mobile lab containing the PTR-ToF-MS being ready for use in 2017. Prior to the publication of Chapter 2, the application of chemometrics to full scan MIMS data of environmental air samples had not been demonstrated. Therefore, a lab-based study was designed to address if chemometrics could be used to exploit the chemical information captured in full scan MIMS data of complex mixtures of trace level VOCs. Once the methodology for sample discrimination had been established, location and time stamped mass spectral data were collected on-road in a research purposed vehicle equipped with a MIMS system (2016 field work) or PTR-ToF-MS system (2017, 2018 field work). The study areas for the measurements are discussed in Chapters 3 and 4 and include communities in the south and central area of Vancouver Island, in the southwestern areas of British Columbia, Canada, as shown in Figure 1.7. Using the full scan MS data from the field campaigns, chemometrics methods were used to address the following questions: can samples impacted by VOC from different sources be discriminated, and can the sources of the observed VOCs be identified and apportioned.

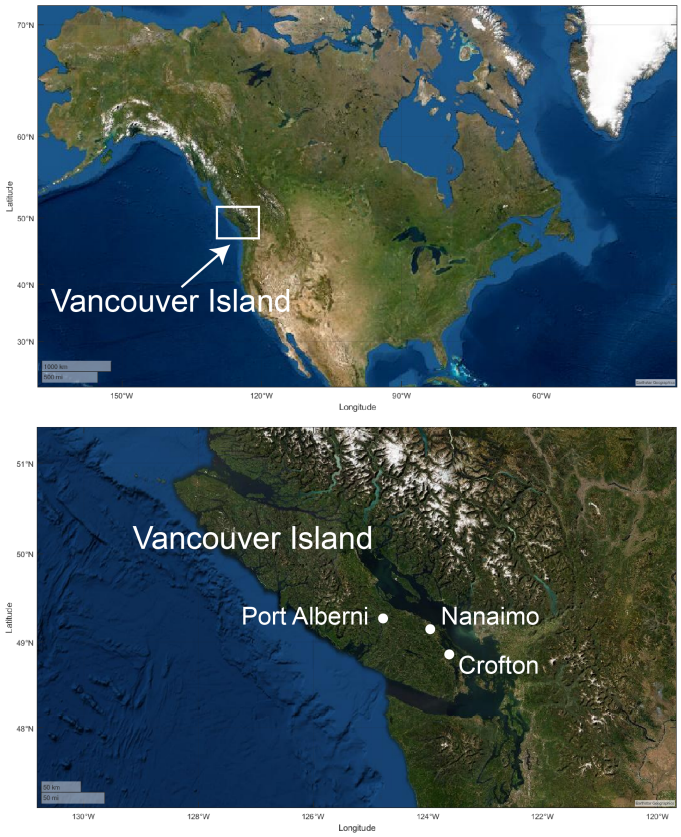


Figure 1.7: Communities visited during field-campaigns.

Chapter 2 aims to establish that the chemical information present in full scan MIMS data of complex air samples, when analyzed using chemometrics, allows sample mixtures to be discriminated, classified, and resolved into their individual components. To achieve this goal, chemometric techniques are applied to full scan MIMS data of gas phase VOCs from a variety of datasets that were constructed and measured in the lab. PCA is applied to three full scan MIMS datasets for source discrimination: a dataset of the VOCs measured in the headspace above diluted water samples that had been equilibrated with various hydrocarbon samples, such as gasoline, bitumen, and diluted bitumen; a dataset where the VOCs present in the smoke from different species of wood were measured; and finally a dataset where VOC mixtures were constructed from a set of ten different individual VOCs. For the final dataset, in addition to PCA, a hierarchical clustering algorithm was used to group samples with similar mass spectra, a k-nearest neighbour (kNN) algorithm was used to classify the samples based on the mixture of compounds they contained, and MCR-ALS was used to unmix the mixtures into their individual components.

The goal of Chapter 3 is to discriminate ambient VOC samples that are measured on-road with MIMS and PTR-ToF-MS by applying chemometrics to the full scan MS data. While sampling continuously from a moving

vehicle, the full scan data collected by MIMS and PTR-ToF-MS captures the variability in the distribution of VOCs at different locations. The work in this Chapter extends the use of the PCA to full scan MS data measured from a mobile lab to address if samples known to be impacted by different VOC sources can be discriminated, and if PCA in combination with clustering algorithms can identify similar samples within a continuous, time series, mass spectral dataset collected over a large geographic region. For both the MIMS and PTR-ToF-MS datasets, samples were identified that were collected near known VOC sources, and PCA was applied to discriminate these samples. For the MIMS dataset, non-polar VOCs such as BTEX were measured, allowing samples impacted by different hydrocarbon sources to be discriminated. For the PTR-ToF-MS dataset, both polar and non-polar compounds were measured, allowing for the discrimination of a wider range of samples that include both hydrocarbon sources, such as gas stations and auto wreckers, as well as sources emitting oxygenated and sulphur-containing compounds, such as composting facilities and pulp mills. In addition to the PCA performed on samples collected near known sources, PCA was applied to the time series PTR-ToF-MS data for the entire field campaign. Two clustering algorithms, a Gaussian mixture model and density-based spatial clustering of applications with noise, were used to cluster samples with similar mass spectra. For both the reduced and full PTR-ToF-MS datasets, maps were produced to visualize the geospatial distribution of samples.

The goal of Chapter 4 is to identify and apportion VOC sources at the neighbourhood and regional scale using data collected from a PTR-ToF-MS system operating in a moving vehicle. Three datasets collected on central Vancouver Island, BC, Canada are discussed, with MCR-WALS is used to identify and apportion VOC source profiles and contributions, with the geospatial maps being used to visualize the results. The main focus of Chapters 2 and 3 to apply PCA to full scan mass spectral data to differentiate samples containing different mixtures of VOCs. However, the VOCs measured at any given location can be due to multiple sources with the corresponding full scan mass spectrum being the superposition of the mass spectra from different sources. In this Chapter the use of a receptor model (MCR-WALS) to unmix the observed mass spectral data into interpretable VOC sources is explored. The datasets analyzed capture a range of VOC sources including biogenic (*e.g.*, forest fire smoke, emissions from forested areas) and anthropogenic emissions (*e.g.*, vehicle exhaust, industrial activities). MCR-WALS and variable selection allowed for the deconvolution of intermixed sources in time series data collected from a moving vehicle. This analysis provides additional useful information that discrimination by PCA does not provide. For all datasets, geospatial maps of the distribution of the identified sources are presented to aid in data visualization.

Chapters 2 and 3 are published manuscripts that stand alone, while Chapter 4 presents unpublished work. The required introductory and experimental details pertinent to that work have been described therein.

Chapter 5 summarizes the key results of the dissertation, including the strengths and weaknesses of the chemometrics explored with the data presented. Additionally, while the research presented makes clear progress towards establishing the pairing of mobile MS data with advanced chemometrics for environmental monitoring of ground-level atmospheric chemistry, refinement of the methodology presented will result in further improvements. Chapter 5 summarizes the research achievements and outlines next steps to address some of the challenges encountered and for future research, including incorporating measurements from the additional on-board sensors (CH_4, CO_2, NO, NO_2, NO_x, O_3, PM) operating in the mobile lab for additional resolving power. Supplementary details for Chapters 2, 3, and 4 are included in the Appendices.

Chapter 2 - Discrimination of constructed air samples using multivariate analysis of full scan membrane introduction mass spectrometry (MIMS) data

Reproduced with minor changes and permission from Richards, L. C.; Davey, N. G.; Fyles, T. M.; Gill, C. G.; Krogh, E. T. "Discrimination of Constructed Air Samples using Multivariate Analysis of Full Scan Membrane Introduction Mass Spectrometry (MIMS) Data" *Rapid Commun Mass Spectrom,* **32**, 349-360 (2018). Sample preparation and MIMS data collection of the headspace above aqueous hydrocarbon solutions and of constructed VOC mixtures by L. Richards. Sample preparation and MIMS data collected of woodsmoke emissions by N. Davey. Data analysis and final manuscript preparations performed by L. Richards. Editorial contributions to the manuscript from N. Davey, T. Fyles, C. Gill, and E. Krogh. Supplemental information for this chapter maybe be found in Appendix A.

2.1 Introduction

Volatile and semi-volatile organic compounds (S/VOCs) are important atmospheric pollutants affecting human[60] and environmental health.[41,156] Their impacts can be both direct (*e.g.*, benzene is carcinogenic[55,60]) and indirect, acting as precursors in the formation of ground-level ozone,[157] as well as SOA.[67] S/VOCs are present in outdoor and indoor air at trace levels, and have both biogenic (*e.g.*, vegetation, forest fires) and anthropogenic (*e.g.*, combustion processes, fugitive emissions) point and nonpoint sources.[38] Air samples are often influenced by S/VOCs from more than one source, and the composition of air samples can be dynamic as a result of mixing and dispersion. The chemical characterization of whole air samples using mass spectrometry provides a rich data set to which environmental forensic tools can be applied.[141,158] Recent developments in field-portable mass spectrometers can provide this data with unprecedented temporal and spatial resolution.[35] As industry and public agencies work to reduce VOC emissions,[156,159] chemometric approaches that identify and apportion sources can be used to better inform policy decisions and protect human health.

MIMS is a direct measurement technique that can be used to make continuous, online measurements of S/VOCs in air, water, and soil samples with little or no sample preparation.[117] In a MIMS experiment, a semi-permeable membrane is used as an interface between the bulk sample and the mass spectrometer. Since no chromatographic separation is employed, analytes are entrained to the mass spectrometer as a suite for analysis. Full scan MIMS data contains the overlapping mass spectra of all permeating molecules. Targeted quantitative analysis can be achieved through the use of MS/MS and/or selected ion monitoring (SIM) in

conjunction with the use of a continuously infused, isotopically labelled internal standard.[82] MIMS is well suited to mobile deployment,[81,160] and during field campaigns we have observed differences in the full scan MIMS data from distinct contamination sources.[82] We have proposed that full scan MIMS data can be used to 'fingerprint' ambient air samples and, in combination with chemometric techniques, be used to identify, and apportion, sources of contamination.[160]

Chemometrics uses multivariate statistical and mathematical techniques for the quantitative and qualitative analysis of chemical data sets containing many variables for pattern recognition, multivariate calibration, and multivariate curve resolution.[161] Pattern recognition techniques such as PCA, soft-independent modeling of class analogy (SIMCA), k-nearest neighbours (kNN), and cluster analysis are used to discriminate and classify samples based on their overall chemical composition.[162] Each of these techniques employs different strategies. In PCA, new latent variables, PCs, are calculated, which are linear combinations of the original variables. The PCs are orthogonal to one another, and describe the directions of most variance within the data. This allows the information in the dataset to be described by fewer variables, uncovering underlying patterns.[131,162] kNN is a classification technique where the class of a new sample is assigned based on a 'majority vote' between the k-nearest samples (where k is an integer) of known source.[162] Multivariate calibration methods, such as principal component regression (PCR) and PLS regression, use the response across many variables to predict sample properties, such as concentration. MCR focuses on resolving components (*e.g.*, individual molecules or specific point sources and their contributions[163]) from samples containing mixtures.[121] These chemometric methods have been used to analyze mixture data collected by mass spectrometry for a variety of applications.[164–169]

Since MIMS produces full scan spectra that are the superposition of spectra from individual compounds in a mixture, chemometric methods can, in principle, resolve both quantitative and qualitative problems. The use of chemometric methods to assist in the quantification of specific analytes (BTEX) in combined standard solutions from EI MIMS spectra has been discussed by Ohorodnik et al[170] and Gardner et al.[171] As the quantification of analytes in complex mixtures from full scan MIMS spectra is based on differences in the pattern and intensity of analyte peaks in the mass spectra, it follows that chemometric analysis should also be able to qualitatively discriminate between classes of samples. Ketola et al used PCA, mean pairwise difference, and kNN classification to discriminate and classify cola samples from different manufactures based on full scan MIMS data.[172] Ketola later used a non-linear asymmetric error function-based least mean square algorithm combined with a reference library to identify individual compounds in the mixture.[173] However, attempts to discriminate environmental samples have been limited. For example, Alberici et al applied PCA to the analysis of full scan

MIMS spectra of the water-soluble fraction of Brazilian commercial gasolines, with mixed results. While they were able to characterize variances between samples, they were unable to discriminate between types of gasoline.[174] This may have been due to a higher degree of intra- versus inter-class variability.

In this study, a series of constructed gas-phase mixtures of known composition and concentrations was constructed from 10 individual VOCs. The compounds were chosen such that some contained known isobaric interferences when ionized by electron ionization (*e.g.*, toluene, *m*-xylene, α-pinene), as well as compounds with relatively few isobars (*e.g.*, naphthalene, biphenyl). We have undertaken a systematic study of nearly 300 lab-constructed VOC mixtures to evaluate the ability of several chemometric techniques to discriminate and classify the samples, as well as apportion the individual compounds within the VOC mixtures. In addition, PCA was applied to full scan MIMS data from two real-world case studies (woodsmoke and headspace above hydrocarbon-contaminated waters) of unknown chemical composition in order to discriminate the samples based on their source. These datasets were collected using mass spectrometers that have been field mobilized, allowing for geospatial and temporal S/VOC analysis. This work is an essential step towards developing tools that identify and apportion sources based on the information contained in full scan MIMS spectra collected from a moving vehicle.

2.2 Experimental

2.2.1 Membrane introduction mass spectrometry

Three datasets are described in this work: gaseous VOC mixtures constructed from permeation tubes using various combinations of individual compounds; VOCs sampled from the woodsmoke emissions produced by the combustion of wood from different tree species; and VOCs present in the headspace above aqueous hydrocarbon samples. Each sample was flowed over a capillary hollow fiber membrane interface design that has been described previously.[175,176] A 10 cm long PDMS membrane (Silastic® brand; Dow Corning, Midland, MI, USA) with 0.47 mm outer radius and 0.26 mm inner radius was used in this study. The membrane interface was heated, and 1 mL/min of helium carrier gas (99.999% purity, Praxair, Mississauga, ON, Canada) was passed through the lumen of the membrane to transfer permeating molecules to the mass spectrometer. Three mass spectrometer systems were used for data collection. For the constructed VOC samples, a modified Griffin 400 EI (70 eV) cylindrical ion trap (CIT) mass spectrometer (FLIR, West Lafayette, IN, USA), with a base pressure of 5×10^{-5} Torr and ion trap temperature of 150 °C was used. The maximum ionization time was set to 150 ms. Data for the woodsmoke VOC samples was collected using a modified GCQ quadrupole ion trap mass spectrometer (Thermo Fisher, Austin, TX, USA), with a base pressure of $1–2 \times 10^{-4}$ Torr. Both the Griffin 400 and GCQ systems were equipped with permeation chambers made in-house, described elsewhere.[82,115] The VOCs in the headspace

above aqueous hydrocarbon samples were collected using a Griffin 300 EI CIT mass spectrometer (FLIR), using the same instrumental settings as used for the Griffin 400 system. On the Griffin systems, the membrane interface was mounted inside an aluminum block and heated to 50 °C, and full scan MIMS data was collected for *m/z* 50–250 at a frequency of approximately 10 Hz. For the GCQ, the membrane interface was housed in a temperature-controlled oven and heated to 70 °C, and full scan MIMS data (*m/z* 50–250) was collected every 12 s.

2.2.2 Data handling

The overall workflow used to construct the three datasets used for this study is shown schematically in Figure 2.1A. Each dataset contained a variety of sources, each representative of one VOC mixture (*e.g.*, benzene and toluene mixture, redcedar woodsmoke, headspace above an aqueous solution containing diluted bitumen). For each source, a number of samples were collected, with each sample measured at least once with a MIMS system. For the constructed VOC mixtures and aqueous hydrocarbon headspace VOC experiments, samples were measurements of a single source at multiple concentrations, or replicate measurements of a single source at a given concentration. For the woodsmoke VOCs, a sample represented closely related measurements where the smoke from a single fire was measured multiple times in one experiment. Each experiment produced a chronogram of the total ion current as seen in Figure 2.1B. The average full scan MIMS spectrum for a sample was calculated by averaging a subset of mass spectra across the steady-state region of the signal in the chronogram shown in Figure 2.1B. For the woodsmoke VOCs, an average full scan MIMS spectrum was calculated for each individual fire by averaging the spectra obtained for each sampling event in the experiment. The average full scan MIMS spectra were then background-subtracted to produce a final mass spectrum for analysis (example spectra for two sources are given in Figure 2.1C). A dataset for analysis was generated by constructing a data matrix where each background-subtracted average full scan MIMS spectrum (sample) was represented by a row, and the measured *m/z* values were the columns. Before chemometric analysis, each spectrum in the dataset was normalized to a unit vector in order to remove concentration effects as follows in Equation 2.1:

$$\hat{x} = \frac{x}{\sqrt{(m/z_1)^2 + (m/z_2)^2 + \cdots + (m/z_n)^2}} \qquad \text{Equation 2.1}$$

where x is the full scan mass spectrum as a row vector ($[m/z_1\ m/z_2 \ldots m/z_n]$), m/z_i is the intensity of the i^{th} m/z, and $\hat{x}$ is the full scan mass spectrum as a vector of unit length. In order for the first PC to describe the direction of most variance in PCA, each column was mean centered to give each variable (*m/z*) a mean of zero. Modifications from the workflow for specific datasets are described below, where necessary.

For the constructed VOC mixtures dataset, an apparent mass shift was sometimes observed in the full scan MIMS spectra due to space charge effects in the small volume CIT of our Griffin 400 system.[177,178] As this would have a marked effect on variance, any sample spectra with apparent mass shifting were excluded from further analysis. For this dataset, a mass range of *m/z* 50–160 was used for the analysis because the high mass range did not contain measured signals (>95 % of signal intensity was captured below *m/z* 160). For the woodsmoke VOC analysis, the emissions for each fire were sub-sampled 2–6 times. Between 5 and 20 full scan MIMS spectra were averaged in each sub-sample. The woodsmoke VOC data was originally collected for quantitative purposes, and the samples contained toluene-d_8 as a continuously infused internal standard (permeation tube at 50 °C).[82] Spectra were normalized to the base peak of toluene-d_8 (*m/z* 98). In order to remove the presence of the internal standard from the woodsmoke VOC full scan mass spectra, the averaged spectrum for clean outdoor air containing toluene-d_8, obtained at the same time as the study, was subtracted from all woodsmoke spectra.

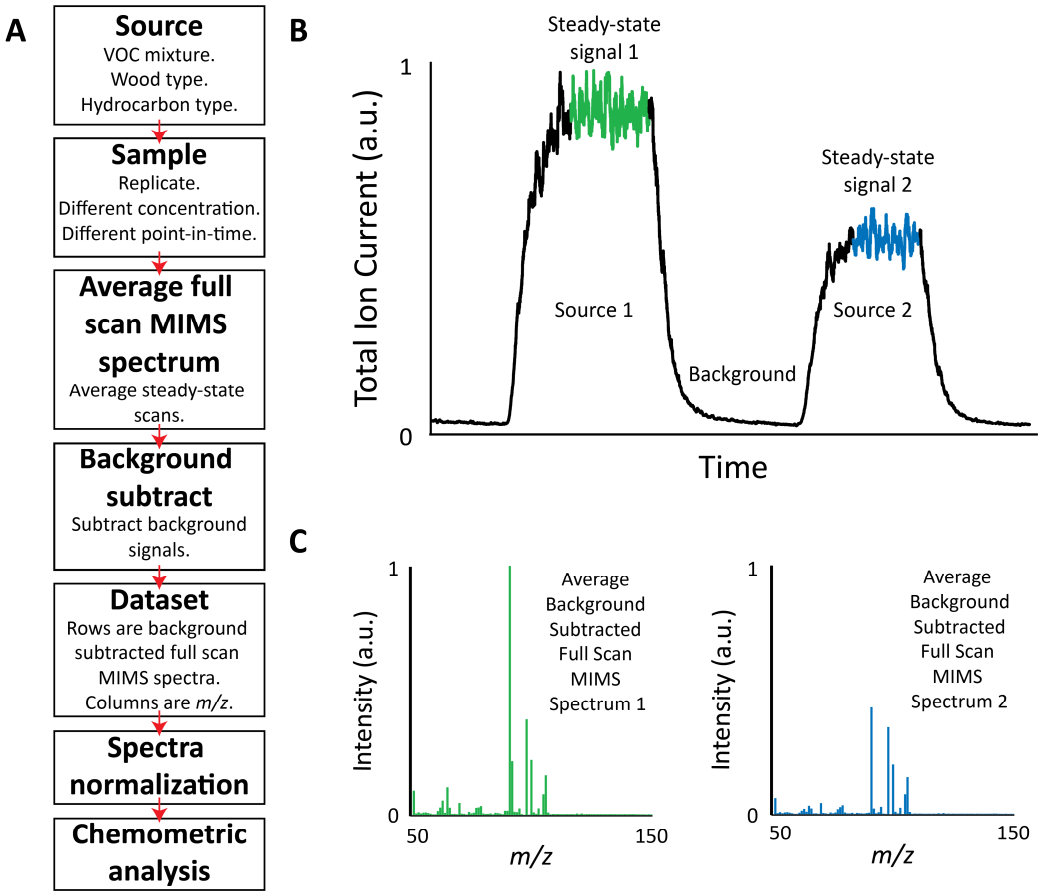


Figure 2.1: A) Schematic of the workflow to produce a dataset for chemometric analysis. B) Chronogram of two VOC samples from different sources, 1 (green, contains toluene, *m*-xylene, and toluene-d_8), and 2 (blue, contains *m*-xylene and toluene-d_8). The steady-state region of the chronogram for each sample is shown in green and blue. C) The average background-subtracted full scan MIMS spectra across both steady-state regions

2.2.3 Constructed VOC mixtures

A series of 32 gaseous VOC mixtures was constructed using a DynaCalibrator™ gas dilution system (model 340; VICI Metronics, Poulsbo, WA, USA) modified in-house to contain a second heated permeation chamber, and an in-house permeation chamber[108] present on the MIMS system based on the Griffin 400 mass spectrometer. Certified emission rate permeation tubes (Kin-Tek, La Marque, TX, USA) were used for the following analytes: benzene, toluene, *m*-xylene, chlorobenzene, naphthalene, α-pinene, trichloroethylene, carbon tetrachloride, biphenyl, and toluene-d_8. The emission rates at specific temperatures for each of the permeation tubes used are provided in Table A.1 (supporting information). Each of the 32 constructed VOC mixtures contained anywhere from one to six different compounds. Data for each mixture was collected at a variety of concentrations ranging from roughly 10 to 1000 ppb_v. The list of mixtures analyzed, the concentration ranges, relative analyte compositions, and the number of samples measured for each, is shown in Table 2.1. For each measured concentration, the MIMS signal was allowed to reach steady-state before dilution air flow was reduced to increase the concentration of the mixture. Some mixtures, such as benzene and toluene, had replicate samples measured across multiple days in order to determine if day-to-day instrumental variability would have an effect on the analysis. Hydrocarbon-scrubbed lab air generated with the DynaCalibrator™ was used as the background for these samples.

2.2.4 VOCs in woodsmoke emissions

Multiple small, wood-fueled fires were ignited using a propane torch in a new tabletop grill. Each fire contained wood from one of three different tree species: Balsam fir (Abies balsamea), Big leaf maple (Acer macrophyllum), and Western redcedar (Thuja plicata). The smoke emissions from each fire were sampled multiple times by alternating the sampling between clean outdoor air and woodsmoke emissions. Replicate fires for each wood species were sampled. Preliminary results from this dataset have been published elsewhere,[160] with the data being reprocessed for the purposes of this paper.

Table 2.1: Concentration ranges, relative abundances of each compound in the mixture, and number of samples constructed for analysis.

Mixture No.	Compounds in mixture	Concentration ranges of mixture (ppb_v)	Relative abundance	Number of samples
1	α-Pinene	12-121	100	4
2	α-Pinene, biphenyl	22-212	57:43	6
3	α-Pinene, biphenyl, chlorobenzene	30-289	42:32:26	6
4	α-Pinene, biphenyl, chlorobenzene, toluene	60-584	21:16:13:50	6
5	α-Pinene, biphenyl, chlorobenzene, toluene, toluene-d_8, *m*-xylene	99-968	13:9:8:31:23:16	6

Table 2.1: Concentration ranges, relative abundances of each compound in the mixture, and number of samples constructed for analysis, continued.

Mixture No.	Compounds in mixture	Concentration ranges of mixture (ppb_v)	Relative abundance	Number of samples
6	α-Pinene, biphenyl, chlorobenzene, toluene, *m*-xylene	76-744	16:12:10:40:22	6
7	Benzene	58-567	100	8
8	Benzene, carbon tetrachloride, toluene, trichloroethylene	105-1032	55:20:39:8	26
9	Benzene, naphthalene	79-768	74:26	4
10	Benzene, naphthalene, toluene	109-1064	53:19:28	26
11	Benzene, toluene	88-863	66:34	65
12	Benzene, toluene, trichloroethylene	95-933	61:32:7	6
13	Benzene, trichloroethylene	65-638	89:11	6
14	Biphenyl	9-91	100	6
15	Chlorobenzene	8-77	100	6
16	Chlorobenzene, carbon tetrachloride	18-175	44:56	6
17	Chlorobenzene, carbon tetrachloride, toluene, toluene-d_8, trichloroethylene, *m*-xylene	95-925	8:11:32:24:8:17	5
18	Chlorobenzene, carbon tetrachloride, toluene-d_8, trichloroethylene, *m*-xylene	64-630	12:26:16:36:11:25	6
19	Chlorobenzene, carbon tetrachloride, naphthalene	38-376	20:26:54	6
20	Chlorobenzene, carbon tetrachloride, naphthalene, trichloroethylene	46-447	17:22:45:16	12
21	Chlorobenzene, carbon tetrachloride, trichloroethylene	25-246	31:40:29	13
22	Chlorobenzene, carbon tetrachloride, trichloroethylene, *m*-xylene	41-405	19:24:18:39	6
23	Chlorobenzene, toluene, toluene-d_8, *m*-xylene	77-756	10:30:39:21	3
24	Carbon tetrachloride	10-98	100	6
25	Carbon tetrachloride, toluene, trichloroethylene	47-464	21:64:15	7
26	Carbon tetrachloride, trichloroethylene	17-169	58:42	6
27	Naphthalene	21-201	100	6
28	Naphthalene, toluene	39-496	40:60	7
29	Toluene	30-295	100	9

Table 2.1: Concentration ranges, relative abundances of each compound in the mixture, and number of samples constructed for analysis, continued.

Mixture No.	Compounds in mixture	Concentration ranges of mixture (ppb_v)	Relative abundance	Number of samples
30	Toluene-d_8	23-225	100	2
31	Trichloroethylene	7-71	100	5
32	*m*-Xylene	16-159	100	6

2.2.5 VOCs in headspace above aqueous hydrocarbon samples

In an effort to simulate real-world contaminated samples with unknown composition and concentrations, the VOCs present in the headspace above hydrocarbon-contaminated waters were measured. The hydrocarbon-contaminated samples included a diluted oil sands process water (OSPW) from the Athabasca region, Alberta, Canada, an aged mixed gasoline for two-stroke engine use (50:1), diluent (a commercial kerosene product used to reduce the viscosity of heavy oil), and diluted bitumen or 'dilbit' (sales product from heavy oil production from the Athabasca region, Alberta, Canada). The OSPW was gravimetrically diluted and used as is. The remaining solutions were diluted from saturated aqueous solutions prepared as follows. Roughly 10 mL of two-stroke engine mixed gasoline, diluent, and dilbit were added to ~1 L of deionized water (Millipore Corp., Billerica, MA, USA) in amber glass bottles, and equilibrated over 2 weeks on a shaker table (model SW23; Julabo, Seelbach, Germany) at 25 °C. The samples were transferred to 1-L separatory funnels and the aqueous fraction was collected in 250-mL amber bottles with Teflon-backed septa, minimizing the headspace. The dilbit solutions were filtered. Portions of the saturated aqueous solutions were diluted gravimetrically using deionized water into 40-mL glass vials. All samples were stored at 4 °C until use.

The experimental apparatus for the recirculation experiments is shown in Figure 2.2. Two stainless steel three-way valves (Swagelok, Solon, OH, USA) were used to alternate between hydrocarbon scrubbed room air and the recirculated headspace sample using a peristaltic pump (Masterflex L/S economy pump system with Easy-Load II pump head with Viton pump tubing, L/S 25; Cole-Parmer, Montreal, Canada). Headspace was recirculated in ¼" fluorinated ethylene propylene tubing (Cole-Parmer). Samples were prepared for headspace analysis by placing 100–200 mL of deionized water into a sealed reaction flask. Aliquots of aqueous samples were injected into the reaction flask through a Teflon-backed silicone septum. The overall dilution ratios from the saturated aqueous solutions to those sampled in the recirculation flask, and the number of samples collected for each hydrocarbon mixture, are summarized in Table A.2 (supporting information). The MIMS signals were allowed to reach steady-state, and then hydrocarbon-scrubbed room air was used to flush the sample from the membrane and return to baseline between samples.

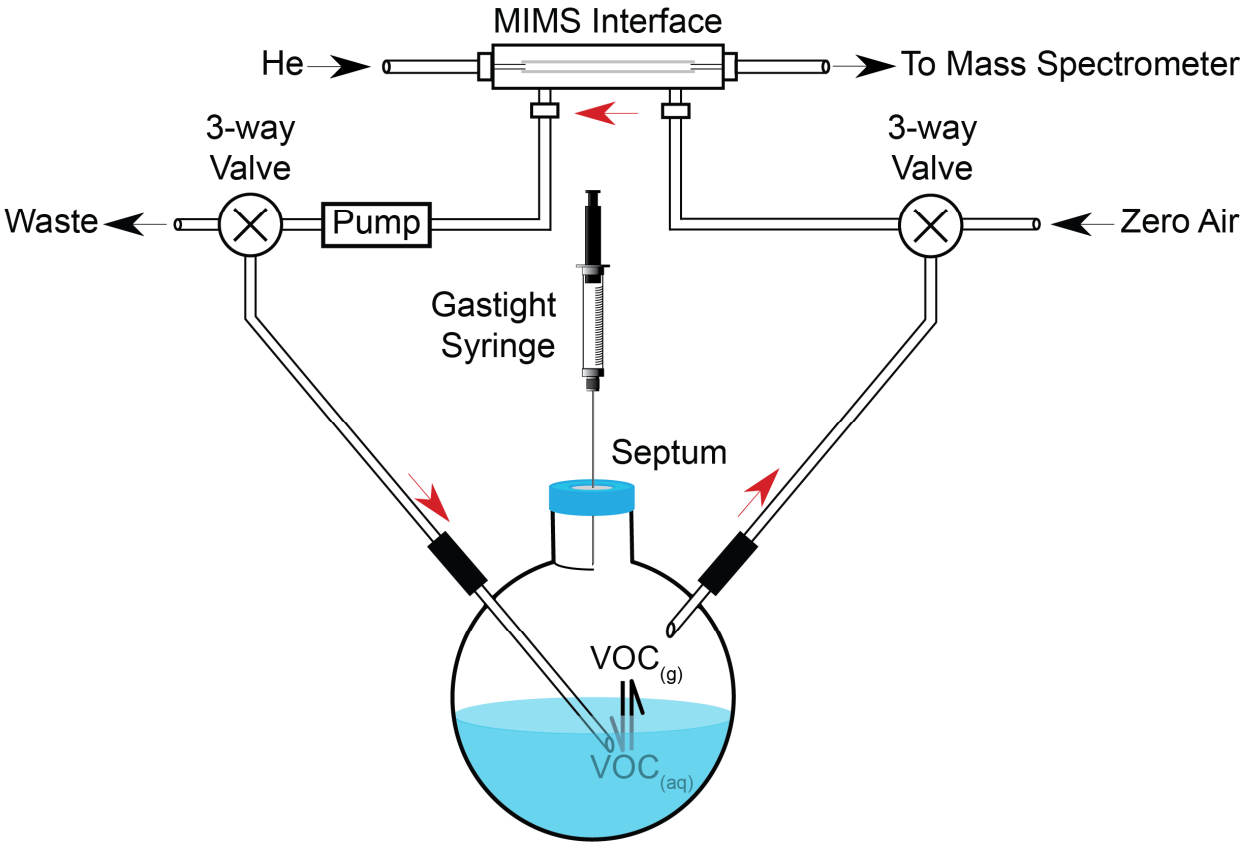


Figure 2.2: Experimental apparatus for MIMS sampling the VOCs in the headspace of a recirculation flask. The two three-way valves allow switching between sample and hydrocarbon-scrubbed air for baseline and background correction

2.2.6 Sample discrimination

For each dataset, PCA was applied to unit vector normalized and mean-centered datasets. In addition, a hierarchical cluster analysis of the unit vector normalized constructed VOC mixtures dataset was undertaken using Euclidean distances between nearest neighbours. Data analysis was carried out using MATLAB (Mathworks, Natick, MA, USA) and the PLS_Toolbox (Eigenvector Research Inc., Manson, WA, USA).

2.2.7 Sample classification

For the constructed VOC mixtures, kNN was applied to the normalized dataset for sample classification. To build a classification model, two-thirds of the samples were used as a training set, and one third used as a test set for classification. Replicate measurements of an individual mixture at the same concentration were in either the training set, or the test set, but not both. For the analysis, the Euclidean distances between samples were used to determine the nearest neighbours. To determine the value of k used for the classification, cross validation (CV) of the training set was done by randomly dividing the dataset into 10 folds. During CV, each fold is classified using the other 9 folds as the training set. This process was completed for k values of 1, 3, and 5. MATLAB and the PLS_Toolbox were used for data analysis.

2.2.8 Multivariate curve resolution

MCR was applied to the constructed VOC mixtures dataset to calculate the pure component spectra of the compounds used to construct the samples.[179–181] The data was unit vector normalized. The Scree plot and scores plots from the PCA were used to estimate the number of components for the model.[180,181] The purest component spectra in the dataset were used as initial estimates in the model, and the model was calculated using non-negativity constraints for both the spectra and component contributions. Data analysis was performed using MATLAB and the PLS_Toolbox.

2.3 Results and discussion

2.3.1 Constructed VOC Mixtures

Analysis of the constructed VOC mixtures was used to systematically test our ability to discriminate and classify samples of known composition and concentration using PCA, cluster analysis, and kNN. The sample mixtures were constructed to include compounds with relatively unique mass spectra (*e.g.*, biphenyl, naphthalene) as well as those where significant isobaric interferences are known to exist (*e.g.*, the tropylium ion ($C_7H_7^+$), *m/z* 91, is a fragment ion of toluene, α-pinene, and *m*-xylene).[104] MCR tested our ability to derive the pure component mass spectra and the contribution profiles of each component in the mixtures. A total of 298 samples in 32 mixtures containing anywhere from 1 to 6 different VOCs chosen from a set of 10 compounds (Table 2.1) were analyzed using the four techniques described above. It was possible to cluster, discriminate, and classify mixtures based on overall chemical composition, even in mixtures containing compounds with closely related mass spectra, and also to decompose the dataset into pure component spectra and their contribution profiles to each mixture.

PCA scores plots were constructed by projecting the data on the PCs explaining the most variance, with the corresponding PCA loadings plots giving a visual representation of the variables (*m/z*) contributing to each PC.[131,162] In the PCA of the constructed VOC mixtures, the first two PCs account for 57% of the variance. The scores plot for PC 1 vs PC 2 (top left), loadings plot for PC 1 vs PC 2 (top right), and three example mass spectra for the mixtures containing benzene and trichloroethylene (left); benzene and naphthalene (centre); and benzene (right) are shown in Figure 2.3. The scores plot shows that samples are well clustered. Mixtures 8, 10, and 11 (Table 2.1) had replicate measurements collected across multiple days, and while there is some spread in the data points for mixture 11, overall day-to-day instrument variability did not impede sample clustering based upon chemical composition. From the loadings plot, the main signals contributing to dataset variance are at *m/z* 78 ($C_6H_6^{+\cdot}$, benzene molecular ion), 92 ($C_7H_8^{+\cdot}$, toluene molecular ion), 91 ($C_7H_7^+$, tropylium ion), 112 ($C_6H_5Cl^{+\cdot}$, chlorobenzene molecular ion), 77 ($C_6H_5^+$, benzene fragment ion), 128 ($C_{10}H_8^{+\cdot}$, naphthalene molecular ion), and

154 ($C_{12}H_{10}^{+\cdot}$, biphenyl molecular ion). As expected, samples with negative scores on PC 1 have a high abundance of compounds whose ions display negative loadings on PC 1. For example, the molecular ions for chlorobenzene and biphenyl have negative loadings on PC 1, and samples containing these compounds have negative scores on the same PC. Examination of the Scree plot from the PCA of these mixtures indicates that 10 PCs account for 99.20% of the variance in the dataset. The scores plots for PCs 1–10 show significant structure within the dataset. In contrast, the scores plot for PC 11 is much less structured, with many mixtures having samples with both negative and positive scores on PC 11 indicating that PC 11 mainly describes noise (Figure A.1, supporting information). As a result, the dimensionality of the dataset can be reduced from 111 dimensions (corresponding to *m/z* 50–160) to 10 dimensions (PCs) without losing information.

As can be seen in Figure 2.3, many of the constructed mixtures (*e.g.*, 2, 9, 10, 21, 27, and 30) are well separated in the PC 1 versus PC 2 scores plot. Some of the more challenging mixtures (*e.g.*, 3 and 14; 25, 29, and 32; 23 and 26; 18 and 22) are somewhat resolved on the PC 1 versus PC 2 scores plot, but can be separated on other projections, such as PC 2 versus PC 3 (data not shown). There are, however, some mixtures that are not as well differentiated. For example, while mixture 9 (benzene and naphthalene) is well resolved in the scores plot, mixtures 7 (benzene) and 13 (benzene and trichloroethylene) are not. The lower panels in Figure 2.3 illustrate the normalized full scan mass spectra for these mixtures. We ascribe the small peak at *m/z* 95 to the formation of a phenyl ion water adduct, $[C_6H_6–H+H_2O]^+$, as observed by us and others in small ion traps.[182] As can be seen, the mass spectra for mixtures 13 and 7 are very similar, differing mainly by low intensity signals at *m/z* 130, 132, and 134 due to the presence of trichloroethylene in mixture 13, which is present at a low abundance relative to benzene (see Table 2.1). While mixture 9 also shares many spectral peaks with mixtures 7 and 13, it contains high intensity peaks at *m/z* 102 and 128 due to the presence of naphthalene. The ion at *m/z* 128 has a negative loading on PC 1; hence, the presence of naphthalene in mixture 9 leads to the sample discrimination.

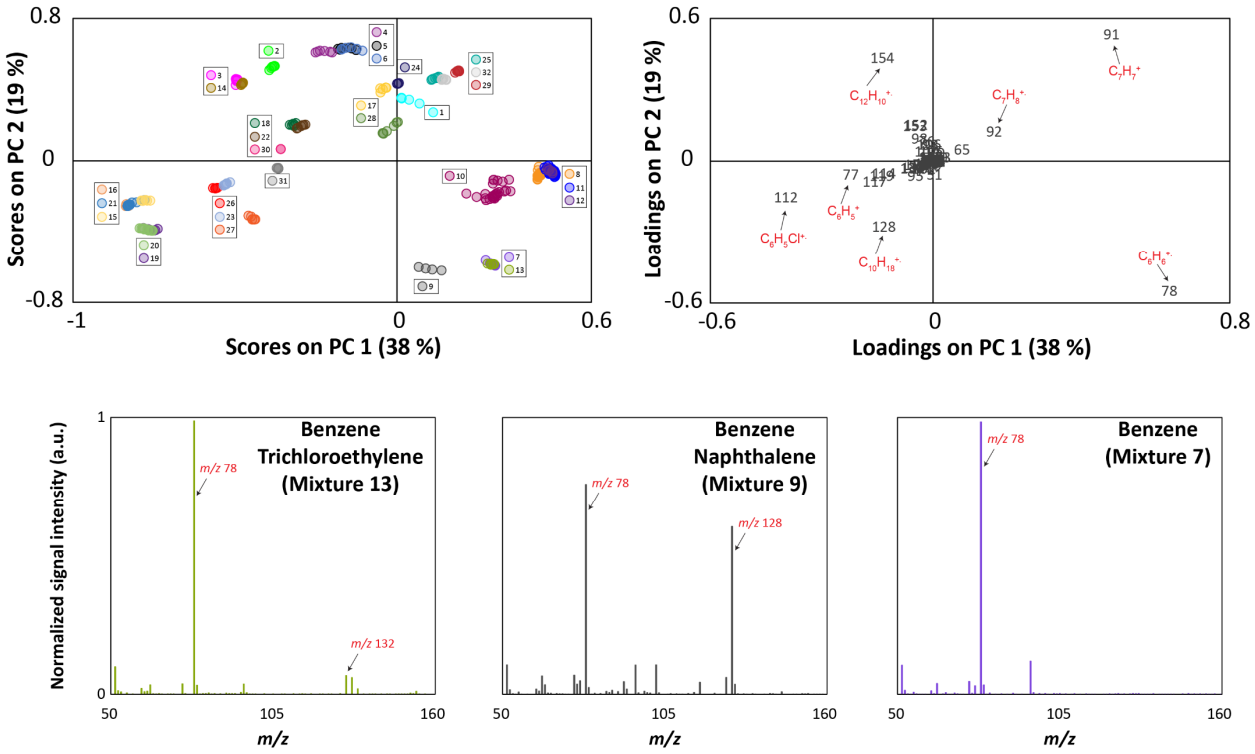


Figure 2.3: Top left: Scores plot of PC 1 vs PC 2 for the VOC mixtures constructed using permeation tubes. Legends within the scores plot identify nearby colour-coded mixtures that are numbered based on the order they are presented in Table 2.1. Top right: Loadings plot for this same dataset. Bottom left: Normalized full scan MIMS spectrum for a mixture of benzene and trichloroethylene (mixture 13). Bottom center: Normalized full scan MIMS spectrum for a mixture of benzene and naphthalene (mixture 9). Bottom right: Normalized full scan MIMS spectrum benzene (mixture 7).

Many of the overlapping mixtures in the scores plot shown in Figure 2.3 differ only by the presence of trichloroethylene and/or toluene-d_8. Discriminating these mixtures can be achieved with scores plots of PC 1 vs PC 6, and PC 2 vs PC 7, respectively (Figure A.2, supporting information).

Hierarchical cluster analysis was used to group samples based on the Euclidean distances between samples. Initially, the distance between each pair of samples was calculated, and the two samples nearest one another were joined to make a cluster. The process of joining samples to clusters based on distance to nearest neighbour was repeated until all the samples had been joined.[162] The results of the cluster analysis is shown as a dendrogram in Figure A.3 (supporting information). Each 'leaf' along the bottom of the dendrogram represents a cluster containing a single sample, colour-coded based on the VOC mixture. The horizontal bars join clusters together, and the distance of the horizontal bar on the y-axis is the distance between the nearest neighbours in the two clusters. Of the 32 mixtures, 29 have all their samples joined into a single cluster before a sample from a second mixture joins the cluster. Mixtures 7, 11, and 13 show some degree of mixing. The results of the cluster analysis are in strong agreement with those from the PCA analysis. With the exception of mixtures 23 and 25, the mixtures that overlap in the cluster analysis also overlap in PCA.

A kNN classification model was used to test sample classification based on full scan MIMS data. The classification model was built using 202 samples, with 96 samples acting as the test set. CV was done on the training set. Of the 202 training samples zero, ten, and thirty-five were misclassified using k values of 1, 3, and 5, respectively. As a result, a k of 1 was chosen for the classification. The results of the classification are shown in a confusion table in Table A.3 (supporting information). In the confusion table, the predicted mixtures of the samples in the test set are represented by the rows, the actual mixtures are represented by the columns, and the sum of each column is the number of samples tested for each mixture. If a sample is correctly classified it will be counted along the diagonal of the confusion matrix, while misclassified samples are counted off-diagonal. Table A.3 (supporting information) shows that 95 of the 96 samples were correctly classified.

Finally, MCR was carried out to calculate the spectral information for the pure components that comprise mixtures, as well as the contributions of the pure components to each sample in the original dataset.[179–181] In this analysis, non-negativity constraints were applied since neither the mass spectra, nor the relative abundances of each compound in the mixtures, contained negative values.[179,180]

Before calculating the model, the number of components in the dataset was estimated. An examination of the PCA Scree plot and scores plots indicated that the dataset contained 10 independent sources of variance (consistent with 10 individual compounds in the constructed mixtures). PCs 11–20 described small and similar amounts of variance (Figure A.1A, supporting information), indicating these components may have described noise rather than chemical information, an observation also supported by the scores plots (Figure A.1B and Figure A.1C, supporting information). As a result, a 10-component MCR model was deemed most likely. For comparison purposes 9-, 10-, and 11-component MCR models were calculated. The 9-component model resolved the pure full scan MIMS spectra for eight of the ten compounds, with the spectra for *m*-xylene and toluene-d_8 overlapping as the 9th component (data not shown). The 11-component model resolved the pure full scan MIMS spectra for all ten compounds, but included an 11th component dominated by *m/z* 65 which was not representative of any individual compound (data not shown). For the 10-component model, each of the calculated components corresponded to the full scan MIMS spectrum of the individual compounds used to make up the mixtures. Example MCR results are shown in Figure 2.4. The MCR calculated mass spectrum for component 1 is shown in the bottom left panel, with an authentic full scan MIMS spectrum for pure toluene shown in the bottom right panel. Comparing the MCR spectrum with the full scan MIMS spectrum for toluene shows little difference between the two, with the calculated and measured spectra almost perfectly matching, identifying component 1 as toluene. This was the case for all 10 of the components calculated by MCR, and

example full scan MIMS data for each compound, alongside their corresponding MCR-derived component, are shown in Figure A.4–Figure A.12 (supporting information).

The contributions of component 1 (toluene) to each of the 298 constructed VOC samples is shown in the MCR scores plot in the top panel of Figure 2.4. A high score indicates a high concentration of toluene present, and a low score indicates a low concentration of toluene. Mixture 29 in the top righthand side of the scores plot represents the samples containing only toluene. Mixture 25, which also has a high score on component 1, is the mixture containing toluene, trichloroethylene, and carbon tetrachloride. Although this mixture contains three compounds, the signal intensities of the mass spectral peaks due to trichloroethylene and carbon tetrachloride are much smaller than those of the peaks due to toluene, as shown in the example full scan MIMS spectrum (inset), resulting in a high score. All the mixtures with scores between 0.4 and 1 contain toluene as a component of the mixture. In principle, any mixture not containing toluene would be expected to have a score of zero. We do observe several mixtures with non-zero scores (typically less than 0.15) that do not contain any toluene. However, these samples did contain α-pinene and/or *m*-xylene, both of which exhibit some isobaric interference with toluene due to the presence of the tropyllium ion (and its fragments). Example full scan MIMS spectra of mixtures 11 (toluene and benzene) and 3 (α-pinene, biphenyl, and chlorobenzene) are shown for comparison as insets. The results of the MCR analysis show that it is possible to use full scan MIMS data to identify the sources contributing to each of the mixtures.

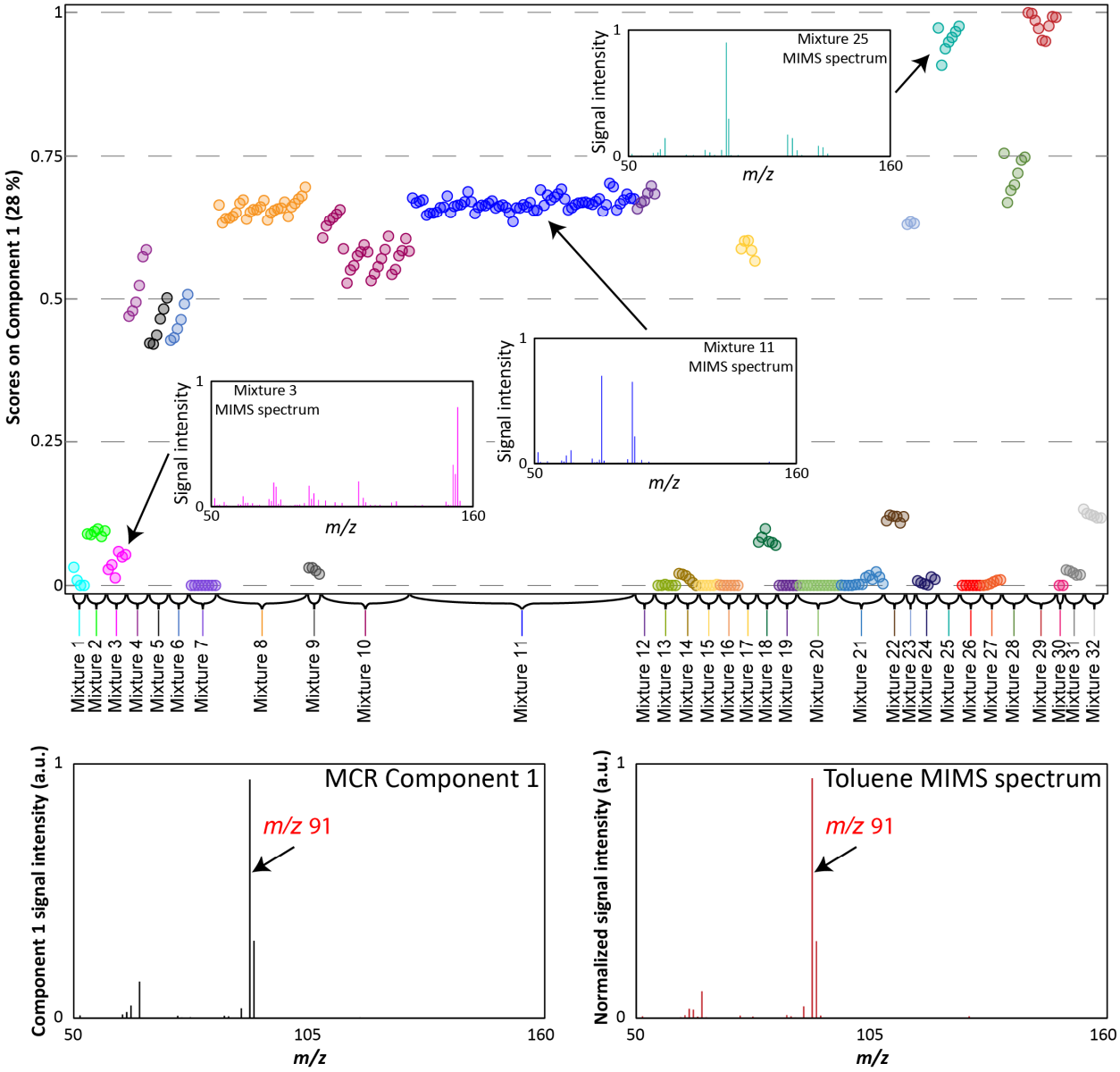


Figure 2.4: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 1, which represents the presence of toluene. Inset: full scan MIMS spectra for mixtures 3, 11, and 25. Bottom left: MCR component 1. Bottom right: Full scan MIMS spectrum of authentic toluene.

2.3.2 VOCs in woodsmoke emissions

The results of PCA on the woodsmoke dataset are shown in the top two panels of Figure 2.5, which also includes average full scan mass spectra for samples measured from the smoke of maple, balsam, and redcedar wood fires. As can be seen in the scores plot, 77% of the variance in the dataset is captured in the first two PCs, with subsequent PCs accounting for relatively little variance (Figure A.13, supporting information). The balsam samples have negative PC 1 scores and can be readily distinguished from maple and redcedar, which have positive PC 1 scores. Maple and redcedar samples can then be differentiated based on the sign of their PC 2 scores (negative and positive, respectively). The mass spectra shown in Figure 2.5 represent the average spectra

across all samples for each type of woodsmoke. The base peaks in the redcedar and maple spectra are both *m/z* 95, while the base peak in the balsam spectrum is *m/z* 77. Comparing the redcedar and maple spectra, redcedar has more intense peaks present in the high-mass range (*m/z* > 130). The loadings plot in Figure 2.5 illustrates that *m/z* 95 contributes a positive loading to both PC 1 and PC 2 and is dominant in the discrimination of redcedar. The discrimination of maple appears to be largely due to *m/z* 128 giving rise to positive PC 1 and negative PC 2 loadings. The balsam sample appears to be discriminated by *m/z* 81, 109, and 124 in the loadings plot.

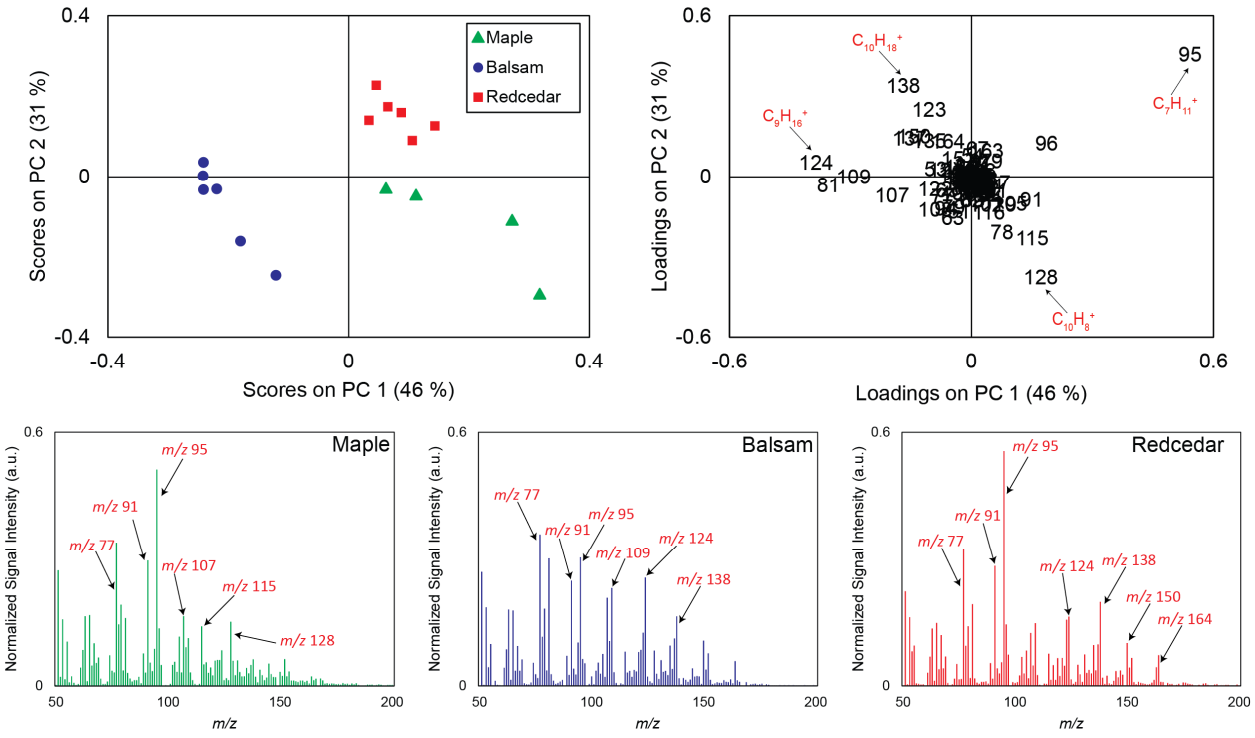


Figure 2.5: Top left: Scores plot of PC 1 vs PC 2 for the woodsmoke dataset. Top right: Loadings plot for this same dataset. Bottom left: Average normalized full scan MIMS spectrum for the maple smoke. Bottom centre: Average normalized full scan MIMS spectrum for the balsam smoke. Bottom right: Average normalized full scan MIMS spectrum for the redcedar smoke. Mass range of spectra shown have been reduced to *m/z* 50-200 for ease of viewing.

Unlike the dataset involving lab-constructed samples, the chemical composition of S/VOC mixtures in the woodsmoke samples is unknown. However, tentative formulas for the ion *m/z* values in the loadings plot can be assigned based on the known hydrophobic properties of molecules that are expected to pre-concentrate in and permeate through PDMS. For example, we attribute *m/z* 128 to $C_{10}H_8^+$ (naphthalene molecular ion), and *m/z* 138 to a $C_{10}H_{18}^+$ compound. The results clearly demonstrate the use of PCA to discriminate samples of known source, but unknown chemical composition, based on full scan MIMS data.

2.3.3 VOCs in headspace above aqueous hydrocarbon samples

Our ability to discriminate the S/VOCs above a hydrocarbon contaminant spill in water was simulated in the lab by recirculating the headspace above aqueous solutions of several hydrocarbon samples, as illustrated in Figure 2.2. The results of PCA scores and loadings for the hydrocarbon dataset are given in Figure 2.6, along with example mass spectra. The first two PCs represent 88% of the variance in the dataset, with subsequent PCs accounting for little variance (Figure A.14, supporting information). The data points for each type of hydrocarbon sample are clustered together, and the samples are well separated in the projection of PC 1 versus PC 2 scores plot allowing for the different types of contaminated waters to be differentiated. The separation of diluent and dilbit samples is of particular interest, as both samples have diluent as a constituent.

The loadings plot for this dataset is shown in the top right of Figure 2.6. On PC 1 *m/z* 78 ($C_6H_6^{+\cdot}$) has the highest loading in the positive direction, and *m/z* 91 ($C_7H_7^{+}$) and 92 ($C_7H_8^{+\cdot}$) have the highest loadings in the negative directions. These *m/z* values are indicative of the presence of BTEX (*e.g.*, *m/z* 91 being present in the EI mass spectrum of toluene, ethylbenzene, and xylenes, as well as other alkylbenzene compounds). The *m/z* variables, important in defining the direction of PC 2, are more difficult to tie to specific compounds or compound classes. In the positive direction *m/z* 51, a minor fragment ion of many aromatic species, is the most dominant, followed by *m/z* 55, 56, 59, 92, 89, and 95. In the negative direction, *m/z* 81, 91, and 106 are the most important.

Although the chemical compositions of these samples are unknown, the *m/z* values in the loadings plot indicate that it is probably the different ratios of alkyl aromatics including the BTEX compounds that are resulting in the PCA separation, with the spread along PC 1 being dominated by potential differences in benzene and toluene ratios in the samples. This data indicates that it is possible to use full scan MIMS data to differentiate between samples of different, but similar hydrocarbon sources.

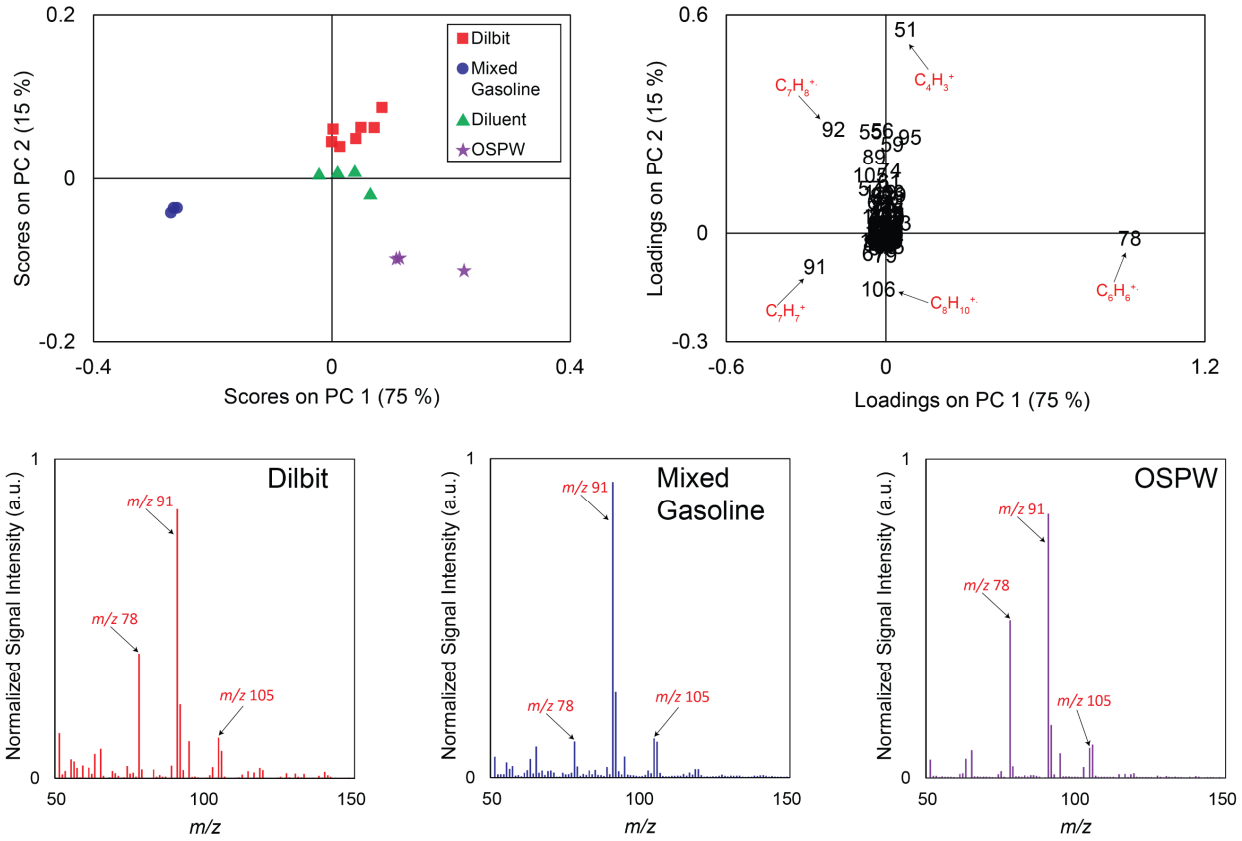


Figure 2.6: Top left: Scores plot of PC 1 vs PC 2 for the hydrocarbon dataset. Top right: Loadings plot for this same data set. Bottom left: Normalized full scan MIMS spectrum for a dilbit samples. Bottom centre: Normalizes full scan MIMS spectrum for a mixed gasoline sample. Bottom left: Normalized full scan MIMS spectrum for an oil sands process water sample. Mass range of spectra shown have been reduced to *m/z* 50-150 for ease of viewing.

2.4 Conclusions

This study indicates that the full scan mass spectral data collected by portable MIMS instruments can be effectively used for the discrimination and classification of complex VOC mixtures. A systematic chemometric study of lab-constructed VOC samples of known chemical composition and concentration demonstrates that samples can be discriminated using PCA and classified by kNN based on the chemical composition of the mixture. Furthermore, we demonstrate the use of MCR to resolve mixtures into molecular component contributions. We have successfully applied a similar PCA approach to distinguish 'real-world' sources of VOC mixtures of unknown chemical composition. Overall, these results show that qualitative analysis of whole air sample full scan MIMS data for source identification can be achieved through the use of chemometric techniques. This work is an essential first step towards using full scan MIMS data for source discrimination and apportionment based on the composition of atmospheric pollutants affecting air quality at a local and regional level. In the future this approach will be applied to real-world data collected in the field with a mobile MIMS system for source apportionment of atmospheric VOCs.

Chapter 3 - Discrimination and geospatial mapping of atmospheric VOC sources using full scan direct mass spectra data collected from a moving vehicle

Reproduced with minor changes from Richards, L. C.; Davey, N. G.; Gill, C. G.; Krogh, E. T. "Discrimination and geospatial mapping of atmospheric VOC sources using full scan direct mass spectra data collected from a moving vehicle" *Environ. Sci.: Process Impacts,* **22**, 173-186 (2020) with permissions from the Royal Society of Chemistry. Data collection by L. Richards, N. Davey, and E. Krogh. Data analysis and final manuscript preparations performed by L. Richards. Editorial contributions to the manuscript from N. Davey, C. Gill, and E. Krogh. Supplemental information for this chapter maybe be found in Appendix B.

3.1 Introduction

S/VOCs are important determinants of both environmental[41,156] and human health[60]. In the atmosphere, they are precursors in the formation of ground-level ozone[157], contribute to particulate matter through the formation of secondary organic aerosols[67], and can also have direct health effects (*e.g.*, benzene is carcinogenic).[55,60] Atmospheric S/VOCs have numerous point and diffuse sources, with both anthropogenic and biogenic origins.[17] Their distributions can vary widely over time and space depending on proximity to sources, meteorological conditions, topography, and atmospheric chemical processes.[17] Traditional methods of atmospheric VOC sampling involve the collection of samples in evacuated cannisters[183] or onto sorbent tubes[90,184] which are subsequently measured in a laboratory setting by GC-MS or related techniques.[183] While these methods allow for comprehensive VOC analysis, they are time consuming and labour intensive, and provide data for one fixed location and sampling time or period. As such, these lab-based methods do not capture fine scale variations in VOC concentrations over time and/or space. More recently, mass spectrometers have been installed at strategic stationary monitoring sites for VOC analysis, using continuous direct mass spectrometry, typically proton-transfer reaction mass spectrometry[37,99,158,185,186], and fast GC-detector systems.[187–189] The resulting datasets can be used to identify VOC sources by examining temporal trends in individual compounds, such as the differing diurnal emission patterns for biogenic VOCs and aromatic hydrocarbons[158], and by applying receptor models to the full suite of measured compounds.[186]

More recently, the miniaturization and ruggedization of analytical instrumentation has facilitated their use in moving vehicles, allowing for geospatial chemical mapping. In particular, the miniaturization of direct

mass spectrometry instrumentation allows for continuous, on-road[80–82,84,108], underwater[109,190], and airborne[37,113,191] measurements to map spatial and temporal distributions of S/VOC concentrations. Further, techniques such PTR-ToF-MS and MIMS allow for the simultaneous analysis of the mixture of S/VOCs in an air sample without chromatographic separation.[192]

MIMS employs a semipermeable membrane as an interface between the sample and the mass spectrometer.[117,160,193] Air samples are continuously passed over a PDMS membrane and S/VOCs pervaporate into a carrier gas, where they are transferred to the mass spectrometer as a mixture for analysis. EI typically employed by these systems results in the initial formation of $[M^{\cdot+}]$ ions and subsequent fragmentation. MIMS is well suited for the analysis of small, hydrophobic molecules present in an air sample. The resulting full scan mass spectra represent the superposition of all of the ions produced by the mixture of VOCs permeating the membrane.[82]

In PTR-ToF-MS analysis, VOCs in a continuously sampled air stream are reacted with H_3O^+ reagent ions in a drift tube with a uniform electric field, where ion-molecule collisions occur. If the proton affinity (PA) of the analyte (M) is higher than that of water, protonated molecular ions $[MH^+]$ are formed. PTR is a softer ionization method than EI resulting in considerably less fragmention.[194] PTR-ToF-MS is an excellent method for directly analyzing both polar and non-polar trace organic compounds in ambient air without chromatographic separation.[37,195] When coupled with time-of-flight mass spectrometry, full scan mass spectra are obtained in microseconds.

Both MIMS and PTR-ToF-MS systems have been operated on a mobile platform for temporally and spatially resolved quantitative analysis of ambient VOCs in the ppt_v to ppb_v range. PTR-ToF-MS systems have been operated on aircraft campaigns to measure and quantify VOCs in an agricultural fire plume and urban environments[113], as well as over oil and gas producing regions[114], and have been used 'on-road' to measure emissions from vehicles.[84] Several researchers have employed PTR with quadrupole mass spectrometers to assess BVOC emissions above forested areas.[111] Portable MIMS systems have been operated from a moving vehicle for quantitative VOC analysis near oil and gas activites[81,82], detect products and impurities from methamphetamine precursor production[80], and to measure VOCs associated with traffic and woodsmoke emissions.[108] These field studies focused on the identification and quantitation of individual VOCs or isomer classes by analyzing the signal intensity at specific *m/z* or by using targeted methods, such as MS/MS and/or SIM scans. Alternatively, qualitative chemometric analysis taking advantage of the wealth of non-targeted information present in a full scan mass spectrum can be used to discriminate between samples containing different VOC mixtures, but has not yet been applied to data collected on a mobile platform.[192]

PCA is a multivariate dimensional reduction technique that calculates new variables, PCs, that are linear combinations of the original variables (*m/z*), as shown in Equation 3.1:

$$PC = a_1\, m/z_1 + a_2\, m/z_2 + \cdots + a_n\, m/z_n \qquad \text{Equation 3.1}$$

with a_i being the loading for m/z_i. The PCs are orthogonal to one another, with PC 1 lying in the direction of most variance in the dataset, and each subsequent PC accounting for sequentially less variance. PCA allows the information to be described with less variables than the original data set with little information loss. For each sample, scores can be calculated by projecting the sample on the PCs. Plotting the scores allows for data visualization, potentially uncovering underlying patterns in the data set and discriminating samples based on differences in their measured mass spectra. To identify which of the measured variables influence how the samples are distributed in the scores plot, a PCA loadings plot is constructed by plotting the loadings of the variables for each PC. Loadings with larger magnitudes contribute more to sample discrimination.[131,196]

PCA has been applied to data collected from MIMS and PTR-MS systems for sample discrimination in environmental[174,192,197–199] and food[172,198,200,201] analysis. Alberici *et al* analyzed the full scan MIMS spectra of the water soluble fraction of Brazilian commercial gasolines using PCA[174], and Ketola *et al* have used a similar approach to distinguish commercial drink products based on the manufacturer.[172] PCA has also been employed to differentiate and classify terpene photo-oxidation mechanisms in atmospheric simulation chambers using data collected with a PTR-ToF-MS system.[197] We have recently employed full scan MIMS data to discriminate lab-based constructed air samples, VOCs produced in the combustion of different species of wood, and headspace samples influenced by aqueous hydrocarbon solutions.[192] None of these studies have exploited the high temporal resolution capabilities of direct mass spectrometry in a moving vehicle to provide geospatial maps of air masses influenced by different VOC sources.

Here we present MIMS and PTR-ToF-MS data collected on-road in a mobile mass spectrometry laboratory during field campaigns on central Vancouver Island, British Columbia, Canada. We use PCA to analyze the full scan mass spectra to discriminate VOC sources in ambient air samples impacted by a variety of activities including vehicle exhaust, pulp mills, sawmills, and composting facilities. We describe a supervised approach, where selected data from the field campaigns was analyzed based on known nearby point sources, as well as an unsupervised approach, where the entire data set was analyzed using PCA followed by a Gaussian Mixture Model and density-based spatial clustering of application with noise (DBSCAN) algorithm in order to identify clusters within the data set. Both approaches allow for the discrimination of real-world VOC sources and can be used in combination with spatial information to generate neighborhood-scale maps illustrating the distribution of sources.

3.2 Experimental

The data presented here was collected during field campaigns on eastern Vancouver Island, British Columbia, Canada in August 2016 (4 days) and August 2017 (2 days), using instruments mounted in a research purposed cargo van (4x4 2500 Cargo 144 Sprinter, Mercedes-Benz, Nanaimo). The study area includes residential, commercial, industrial, and rural/light agricultural uses. The surrounding area is dominated by mixed coniferous forest, and marine environments. Field notes identified suspected stationary VOC point sources (*e.g.*, gas stations, industrial sites, composting facilities) as well as mobile VOC sources (*e.g.*, vehicle exhaust, transport trucks) encountered while driving. It should be noted that it was not always possible to identify a single VOC source as some sources are co-located and the ambient air was influenced by more than one source. Details from the field campaigns are described below and in Supplemental Information.

3.2.1 Membrane introduction mass spectrometry

A modified Griffin 400 EI (70 eV) with a cylindrical quadruple ion trap (CIT) mass spectrometer (FLIR, West Lafayette, IN, USA) described previously was used for the MIMS analysis.[82] A 10 cm long PDMS capillary hollow fibre membrane (Silastic® brand; DowCorning, Midland, MI, USA) with 0.47 mm outer radius and 0.26 mm inner radius was employed, heated to 50 °C, with 1 mL/min of helium carrier gas (99.999 % purity, Air Liquide, Nanaimo, BC, Canada) continuously passing through the membrane's lumen. The mass spectrometer was operated with a base pressure of 5×10^{-5} Torr, with the ion trap at 150 °C, using a maximum ionization time of 150 ms for full scan data collection. The MIMS instrument was encased in 0.1 mm thick mu-metal ($\mu \approx$ 50,000, McMaster Carr, Elmherst, IL, USA) in order to reduce the effect of the Earth's magnetic field on the measured signal intensity.[202] The MIMS instrument described has low ppb_v sensitivity for hydrocarbons such as the BTEX suite and a compound dependent response time of 15-30 seconds due to membrane transport.

The MIMS system was operated from an independent 24 VDC power supply (4x6VDC lead acid batteries, Model S6-275AGM, Surrette Battery Company Ltd, Springhill, NS, USA) in the vehicle. Ambient air was continuously flowed over the membrane interface at 2 L/min using a 1/4 inch outer diameter and 3/16 inch inner diameter fluorinated ethylene propylene (FEP) sampling line (Cole-Parmer, Montreal, Canada) that was passed through the passenger side window and attached to the front of the vehicle 40 cm above the windshield. Two in-line stainless steel frit filters at 15 and 5 µm (Swagelok, Solon, OH, USA) were used to remove particulate matter from the sampling stream.

Field data was collected with the MIMS instrument on August 4, 5, 9, and 10, 2016 in the Nanaimo area of Vancouver Island, British Columbia (BC), Canada. Meteorological data for this field campaign is found in the Supplemental Information and summarized in Table B.1. Full scan MIMS mass spectra were collected every 7

seconds on August 4, 9, and 10, and every 1 second on August 5. Air was sampled near fresh paving at two different locations (asphalt samples), vehicle exhaust samples were collected roadside as vehicles were unloading from a ferry, as well as from vehicles encountered when sampling, gasoline samples were collected at a gas station (fresh) and from a storage container (aged), and additional road works samples were collected in the vicinity of the paving activities.

In addition to collecting full scan mass spectra for non-targeted VOC fingerprinting, several targeted quantitative scans were also included for context. Details of the quantitative analysis are found in the Supplemental Information, with Table B.2 detailing calibration information. Ambient VOC concentrations were generally observed to be in the low ppb_v range, with less than 5% of the observations above our quantitation limits (typically 3-6 ppb_v) as illustrated in the whisker plots shown in Figure B.1 (Supplemental Information). Concentration excursions above these limits were associated with sampling in close proximity to known point sources with concentrations up to 18 ppb_v for benzene, 91 ppb_v for toluene, 88 ppb_v for ethylbenzene, and 26 ppb_v for α-pinene.

3.2.2 Proton-transfer reaction time-of-flight mass spectrometry

The proton-transfer reaction time-of-flight mass spectrometer (PTR-TOF1000, Ionicon Analytik Ges.m.b.H, Innsbruck, Austria) was attached to a vibration-damping mount, with six air cushion supports (Stabl-Level, SLM-1, McMaster Carr) pressurized to 50 PSI to dampen vibration while the vehicle was in motion. The PTR-ToF-MS was powered from a custom 4800 Ah lithium ion battery bank (Canadian Electric Vehicles, Parksville, BC, Canada) and 2800 W AC/DC inverter/charger (Magnasine, MS2012 2800-12V, Sensata Tech, St. Paul, USA) that is continuously float charged from a dedicated 220 amp engine alternator. A perfluoroalkoxy alkane (PFA) sample line (4 m long, OD ¼", ID 3/16", Cole-Parmer, Montreal, Quebec, Canada) was mounted on an aluminum sampling arm extending 1 m above the roof at the front of the vehicle, with an in-line 1 µm filter (Swagelok, Solon, OH, USA). The sampling line was not heated, and ambient temperatures ranged from 24–27 °C during sampling. H_3O^+ was used as the reagent ion, and the instrument was mass calibrated every 5 s during sampling using the signals at *m/z* 21.022, 29.999, and 59.049 ($H_3{}^{18}O^+$, NO^+, $C_3H_7O^+$). The instrument resolution was 1500-2000 at *m/z* 59. An E/N of 130 Td and drift pressure of 2.20 mbar was used during sampling. The pressure inside the ToF was 1.2×10^{-6} mbar. The air sampling flow rate was 250 sccm, with a mass spectrum collected every 1 s. The transit time between the sample inlet and the mass spectrometer was about 10 seconds. The instrument is sensitive to sub parts-per-billion levels.

Field campaign data was collected on August 21 and 22, 2017 between Nanaimo and Crofton, BC. Winds were light (2-4 m/s) on all sampling days. Additional meteorological data for this field campaign is found in the

Supplemental Information and summarized in Table B.3. Samples were collected near two different pulp mills, two sawmills, wood storage site, composting facilities (commercial operator and a topsoil producer), landfill, gas station, auto wrecking facility, and ferry terminal. Some sources were sampled on both days (*e.g.*, composting facility). Mass spectra were collected at 1 Hz using the PTR-ToF-MS system, and high resolution positional data was collected using a roof mounted antenna (Hemisphere A45, Scottsdale, AZ, USA) and GPS receiver (Hemisphere R330, GNSS Receiver, Scottsdale, AZ, USA). While the focus of this work was on qualitative analysis, direct calibrations for benzene, toluene, ethylbenzene, DMS, and α-pinene were also performed using certified permeation tubes in a Dynacalibrator Gas Dilution System with concentrations ranging up to 150 ppb_v. Details of this work is found in the Supplemental Information, with Table B.4 detailing the calibration information. Observed VOC concentrations were generally below 1 ppb_v, with elevated levels up to 150 ppb_v of DMS (*m/z* 63.022) being measured in the vicinity of a pulp mill. BTEX concentrations up to 25 ppb_v benzene (*m/z* 79.049), 50 ppb_v toluene (*m/z* 93.061), and 60 ppb_v ethylbenzene (*m/z* 107.076) were measured near an auto wrecking facility. Terpene concentrations up to 56 ppb_v (as α-pinene, *m/z* 137.110) were measured near a wood chip truck and sawmills. Whisker plots for VOC concentrations over the sampling period are shown in Figure B.2 in the Supplemental Information.

3.2.3 Data analysis

Post deployment, the data collected for each day of the field campaigns was combined into a single file for interrogation. Mass lists for the PTR-ToF-MS were made using the PTRViewer 3 software (Version 3.2.8 Ionicon). Mass calibration for each of the data files was verified in the software using 5 scan averaging and signals at *m/z* 21.022, 29.997, and 59.049. After mass calibration, the spectra in each data file were summed up and then used to identify all *m/z* peaks present in the data set. For overlapping peaks, the Gauss fit tool in the PTRViewer 3 software was used to calculate overlapping mass spectral peak areas. Peak areas were corrected to account for differences in ion transmission through the mass analyzer using the instrument's transmission curve. PTR-ToF-MS data was normalized to the reagent ion signal at *m/z* 21.022 ($H_3{}^{18}O^+$) to account for ionization differences. For each field campaign a time series signal chronogram of the total ion current (TIC) data (*m/z* 30-215 excluding the signal intensities of O_2^+, $H_3OH_2O^+$, and $H_3OH_2{}^{18}O^+$ for the PTR-ToF-MS data, *m/z* 50-250 for the MIMS data) was examined to identify periods with higher VOC concentrations. We noticed some carry-over in the PTR-ToF-MS data after sampling near a composting facility (particularly at *m/z* 61.029, associated with acetic acid). On both days, the compost samples were collected at the end of the run and data impacted by carry-over was removed from the analysis.

3.2.4 Data analysis for supervised PCA analysis

Data associated with high total VOC concentrations was cross checked with the field notes to identify potential VOC sources. If a source was identified, average mass spectra were calculated by averaging the measured signal intensities across a 10-15 s time window for subsequent use in PCA. Mass spectra were background subtracted using the mass spectrum of ambient clean air. A dataset for each field campaign was generated by creating a data matrix, where each row contained an average mass spectrum for an identified sample, with the columns for the data matrix containing the measured *m/z*. Each spectrum in the dataset was normalized to a unit vector in order to remove concentration effects as follows in Equation 3.2:

$$\hat{x} = \frac{x}{\sqrt{\left(m/z_1\right)^2 + \left(m/z_2\right)^2 + \cdots + \left(m/z_n\right)^2}} \qquad \text{Equation 3.2}$$

where x is the full scan mass spectrum as a row vector ([m/z_1 m/z_2 … m/z_n]), m/z_i is the intensity of the i^{th} m/z, and $\hat{x}$ is the full scan mass spectrum as a vector of unit length. Each column was mean-centered to give each variable (m/z) a mean of zero. PCA was applied to the normalized and mean centered datasets. Data analysis was done using Microsoft Excel (Microsoft Corporation, Redmond, WA), MATLAB (Mathworks, Natick, NA), MATLAB's Statistics and Machine Learning Toolbox, and the PLS_Toolbox (Eigenvector Research Inc, Manson, WA).

3.2.5 Unsupervised data analysis

Unsupervised cluster analysis was conducted on a PCA model of the PTR-ToF-MS data set collected on August 21 and 22, 2017 using two algorithms: Gaussian Mixture Model (GMM) algorithm and a DBSCAN algorithm. The GMM algorithm calculates clusters by assuming the data set is composed of a specified number of clusters with Gaussian distributions with unknown means and covariance matrices.[162] DBSCAN identifies core points that have a certain number of neighbours within a given neighbourhood and extends the clusters outwards from these core points, resulting in clusters of high density data. The number of clusters does not need to be specified beforehand, and the clusters can have any shape.[203] For the GMM algorithm, the model was run using a random initiation for 2-20 groups, with 100 replicates per number of clusters, and a maximum of 10,000 algorithm iterations per replicate. Models were run with covariance matrices being shared among the groups, unshared among the groups, and with both full and diagonal covariance matrices using modified MATLAB code.[204] Akaike information criterion (AIC), and Bayesian information criterion (BIC) values were calculated for each of the 19 clustering options. Additional AIC and BIC values were calculated for up to 40 groups using full, unshared covariance matrices. The results from these metrics were then compared to inform

the number of groups within the dataset. The final number of groups used in the analysis was determined by producing geospatial maps with different colours used for each group. These maps were then interrogated, and compared with field notes to determine the optimum clustering model. For the DBSCAN algorithm, neighbourhood size and the number of neighbours needed to identify core points were varied until the algorithm would identify multiple clusters within the PCA model. Once again, the final number of clusters was determined using a combination of geospatial maps and field notes.

3.2.6 Geospatial maps

Geospatial maps of the August 2017 field campaign were generated to visualize the spatial distribution of VOC sources in Google Earth (Google Earth, Google LLC, Mountain View, CA, USA) by using MATLAB to produce keyhole markup language (kml) files. The Komoot Outdoor Google Earth overlay was used to aid data visualization.[205] Data was collected at a frequency of 1 Hz resulting in a spatial resolution of 15-25 m with drive speeds between 50 and 100 km/hour. As the GPS data and mass spectral data were not collected at the exact same timestamps, a location for each mass spectrum was determined by using a linear interpolation between the two latitude and longitude measurements bracketing the sampling time of the mass spectrum. For the supervised PCA, 10-15 s of data was averaged for each sample and the average sample time was used for the spatial interpolation. The data was then coloured based on a group assignment determined by the field notes and sampling location. The data points plotted on the maps were size scaled in proportion to the total ion current from the mass spectrometer which is directly related to total VOC concentration. The bin size cut offs were the 40th, 60th, and 80th percentiles for the 298 samples in the supervised PCA analysis. For the unsupervised analysis, a location for each sample (1 per second) in the dataset was calculated. The colors were determined based on group memberships calculated by the clustering algorithms. Five user defined bins were used to size the data points based on total VOC concentration.

3.3 Results and discussion

The data collected during the field campaigns described above was used to demonstrate the use of direct mass spectrometric analysis of ambient air for sample discrimination based on ambient VOC mixtures measured from a moving vehicle. The two mass spectrometers employed on the field campaigns measure different suites of compounds, capturing a wide range of sources. Tentative formula assignments for some of the measured *m/z* have been identified throughout this manuscript. While we are aware of modified PCA algorithms taking into account the error structure of the data set[206], classical PCA was chosen as the analysis technique for this preliminary study with the goal of establishing our ability to discriminate real-world samples based on their VOC composition. We have employed both a supervised approach, where only samples from

known sources were analyzed by PCA, and an unsupervised approach using multivariate clustering algorithms to group samples with similar mass spectra in the complete data set. As the data is time and location stamped, the sources can be visualized on a geospatial map. Figure 3.1 shows an example of data selection used in the supervised approach. The time series shown is the total ion current (TIC) from the PTR-ToF-MS data collected on August 21, 2017 with time of day on the x-axis. In cases where the concentration excursions are associated with identified point sources, the time series is shaded with different colours accordingly. Peaks from sources not identified in the field notes are shaded grey, and times with low VOC concentrations are unshaded. The red horizontal lines along the top of the time series indicate stationary monitoring with the engine turned off. The remaining data was collected while driving.

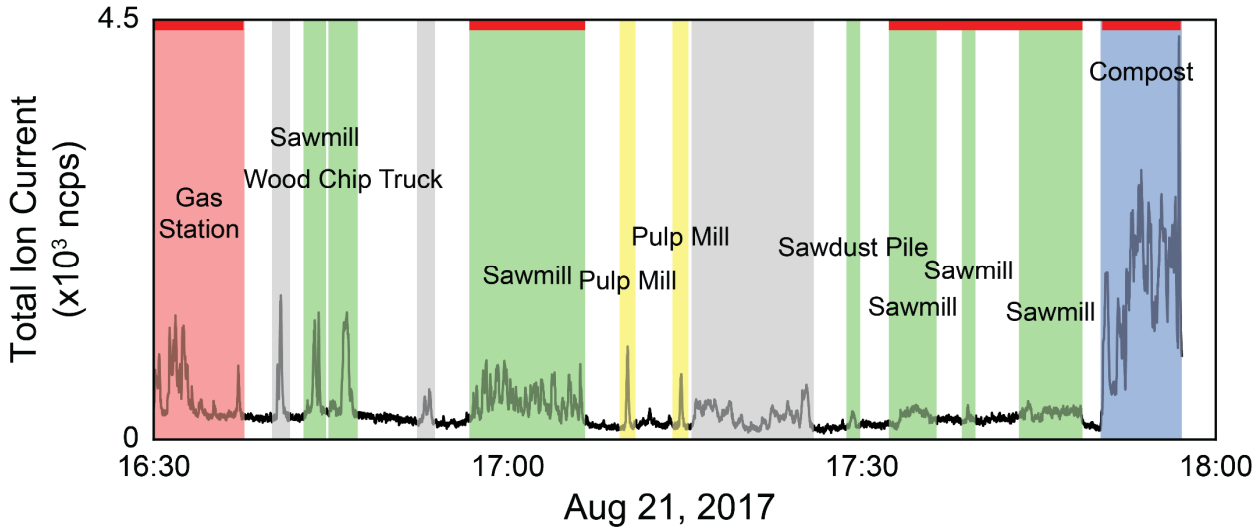


Figure 3.1: Time series of the total ion current for PTR-ToF-MS data collected on August 21, 2017. The shaded regions indicate times when higher concentrations of VOCs were being measured. Measured signals with an identified source are coloured in red, green, and yellow and labeled. Signals with unknown source are shaded in grey. The red horizontal bars along the top of the time series represent times where the vehicle was parked with the engine off for stationary sampling.

3.3.1 Membrane introduction mass spectrometry

The field campaign from August 2016 employed a MIMS instrument, which is well suited to measuring relatively small, hydrophobic VOCs such as the BTEX suite.[82] On-road measurements showed elevated signals near known hydrocarbon sources such as commercial gasoline stations, fresh asphalt paving, and vehicle exhaust. Figure 3.2 (panels A-C) show example full scan MIMS mass spectra collected during the course of these field observations. The spectrum in panel A is for fugitive hydrocarbon emissions from commercial gasoline. The base peak of this spectrum is *m/z* 91, the tropyllium ion ($C_7H_7^+$), which forms when toluene and other alkylated aromatics are fragmented during electron ionization.[207] Other major peaks in the mass spectrum of this sample are *m/z* 92 ($C_7H_8^+$, the molecular ion for toluene); 105 and 106 ($C_8H_9^+$ and $C_8H_{10}^+$ fragment and molecular ion of ethylbenzene/xylenes, respectively); and 78 ($C_6H_6^+$, the molecular ion for benzene). The spectrum in panel B was

collected near fresh asphalt, and has significantly more ions at high relative signal intensity compared to the spectrum for gasoline (panel A) and vehicle exhaust (panel C). The major ions present in this spectrum suggest the presence of a variety of hydrocarbons, and include *m/z* 67 ($C_5H_7^+$), 69 ($C_5H_9^+$), 71 ($C_5H_{11}^+$), 81 ($C_6H_9^+$), 83 ($C_6H_{11}^+$), 91 , 95 ($C_7H_{11}^+$), 97 ($C_7H_{13}^+$), and 105.[208] Similar to the mass spectrum of VOCs near commercial gasoline, vehicle exhaust (panel C) is dominated by *m/z* 91 (base peak) with additional signals at *m/z* 92, 105, and 106.

The PCA for the MIMS data is shown in Figure 3.2, with the scores plot shown in panel D, and the loadings plot in panel E. For this dataset, PC 1 describes 42 % of the variance, PC 2 describes 9 %, and PC 3 describes 6 %. The variance captured by each of the PCs can be visualized using a Scree plot, in which the relative variances are plotted in descending order. The Scree plot for this analysis is shown in Figure B.3 in the Supplemental Materials indicating that the first 10 PCs account for 71 % of the variance in the data. Samples in the scores plot are colour-coded based on nearby sources identified in the field notes, and samples from similar hydrocarbon sources are seen clustered together. PC 1 discriminates the asphalt samples from the other measured samples. The asphalt samples generally have negative scores on PC 1 and scores close to zero on PC 2. PC 2 allows for the discrimination of vehicle exhaust from fugitive emissions from gasoline and road works samples. Generally, the vehicle exhaust samples have more positive scores on PC 2. The aged gasoline samples (red diamonds) have more positive scores than the fresh gasoline samples (red triangles) on PC 1. Weathering of gasoline samples leads to the loss of the more volatile compounds which could account for the separation of the fresh and aged gasoline samples.[209] The samples associated with other road works activities plot near the gasoline samples. We hypothesize that these samples could have been influenced by fugitive emissions from hydrocarbon cleaning mixtures or improperly sealed gasoline containers. For the vehicle exhaust samples, the samples with the highest positive scores on PC 2 were collected across the street from a sawmill as vehicles were unloading from a ferry. These samples are also impacted by emissions from the sawmill, discriminating them from samples impacted solely by vehicle exhaust observed elsewhere. It was not possible for us to sample air impacted only by the sawmill without the confounding influence of vehicle exhaust on this field campaign. The scores plot for PC 1 versus PC 3 is found in Figure B.4 (panel A) in the Supplemental Materials. In this projection, the asphalt samples remain well separated, with negative scores on PC 1, but there is more overlap between the other hydrocarbon samples.

The loadings plot identifies which *m/z* lead to sample discrimination in the scores plot. Many of the measured *m/z* that are more dominant in the asphalt spectrum have negative loadings on PC 1, while *m/z* 91 demonstrates a high positive loading on PC 1, and was the base peak in most of the mass spectra from other sources. Ions due to the presence of α-pinene or other monoterpenes (*m/z* 93 ($C_7H_9^+$) and 136 ($C_{10}H_{16}^+$)) from

fresh cut forest products have positive loadings on PC 2 allowing the vehicle exhaust samples collected near a sawmill to be discriminated from those collected elsewhere in the region. The loadings plot for PC 1 versus PC 3 is found in Figure B.4 (panel B), and indicates that it is once again the asphalt samples that are differentiated from the others due to a greater number of observed compounds. Overall, this analysis demonstrates that full scan MIMS data of ambient, volatile organic hydrocarbons collected while driving can be used to discriminate samples arising from different sources.

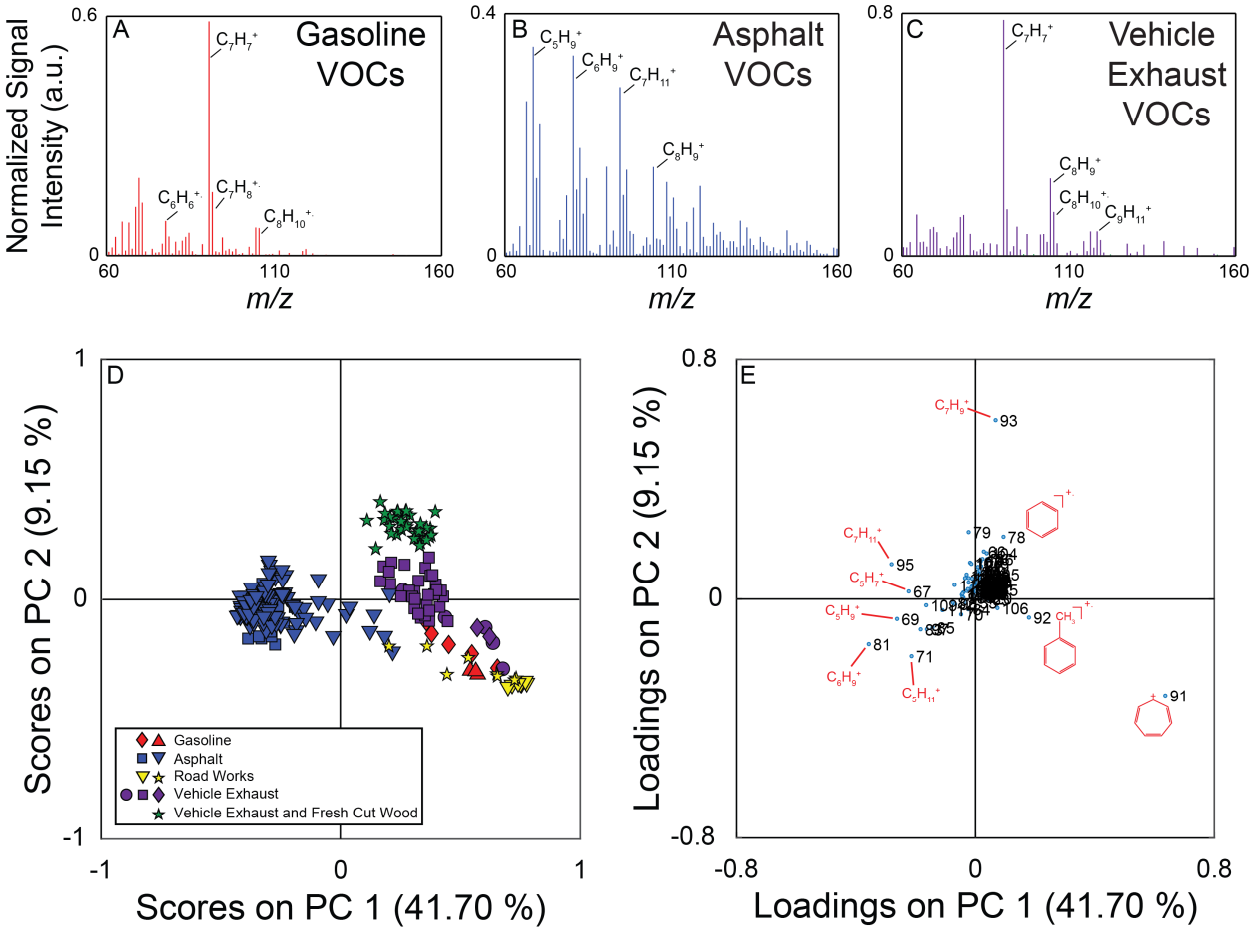


Figure 3.2: Top: Example full scan MIMS spectra collected near fresh gasoline (A), freshly laid asphalt (B), and vehicle exhaust (C). Some ions of interest have been labeled on each spectrum. Panel D: PC 1 versus PC 2 scores plot of the full scan MIMS data. Samples (based on mass spectra averaged for 15 s) are coloured based on their source type (asphalt – blue, gasoline – red, vehicle exhaust – purple, vehicle exhaust mixed with sawmill emissions – green, other road works – yellow), and the shapes represent different events when the source was encountered. Panel E: PC 1 versus PC 2 loadings plot for the MIMS analysis. The separation of asphalt samples in the scores plot is driven by the ions with negative loadings on PC 1. These ions were measured with higher relative abundance in the asphalt samples.

3.3.2 Proton-transfer reaction time-of-flight mass spectrometry

PTR-ToF-MS is more sensitive to a wider range of polar and non-polar VOCs than MIMS as well as being more selective due to the significantly greater mass resolution. Consequently, more compounds were observed

during the August 2017 field campaign. Samples were collected from both stationary (*e.g.*, industrial sites, transportation infrastructure) and moving sources (wood chip truck, farm vehicles).

A total of 290 peaks were identified between *m/z* 30-215 in the full scan data. Table B.5 in the Supplemental Materials provides a peak list of the major ions detected in the field campaign, with our measured *m/z*, possible chemical formula, calculated exact mass for the formula, potential compound identities, and observed sources. The range of compounds detected included hydrocarbons (*e.g.*, BTEX, isoprene, monoterpenes), oxygenated species (*e.g.*, methanol, acetone, acetaldehyde), and sulphur compounds (*e.g.*, DMS). As the PTR-ToF-MS does not employ chromatography or tandem mass spectrometry for additional selectivity, isomers cannot be distinguished, and the signal intensity of each *m/z* is for the sum of isomers (*e.g.*, *m/z* 59.048 would have a chemical formula of $C_3H_7O^+$ which could be both protonated acetone and protonated propanal). It should be noted that peaks at *m/z* 93.036, 107.076, 121.085, and 135.092, attributed to the protonated molecular ions of hydrocarbons associated with fugitive emissions and/or incomplete combustion of fossil fuels, have also been observed to be derived from the fragmentation of mono and sesquiterpenes, even with the softer chemical ionization of proton transfer.[210,211] This is reflected in the listed sources for these ions, as these ions had major sources near an auto wrecking facility, gas station, and ferry, but were also detected near biomass sources such as the sawmills, composting facilities, and a wood chip truck. For the purpose of the analysis described below the different samples have been broadly classified as follows: aged biomass samples – samples collected near a landfill, municipal composting facility, and topsoil producers; fresh biomass samples – samples collected near sawmills, wood chip trucks, wood storage, and sawdust piles; hydrocarbon samples – samples collected near a gas station, ferry terminal, and auto wrecking facility; pulp mill samples – samples collected near two pulp mills; and the farm vehicle samples were collected near a tractor carrying hay.

During the 2017 campaign there were several compounds measured exclusively at or near particular sources. For example, *m/z* 63.022 ($C_2H_7S^+$) was only detected near the two pulp mills on the drive route. Peaks at *m/z* 57.070 ($C_4H_9^+$), 107.076 ($C_8H_{11}^+$), and 121.085 ($C_9H_{13}^+$) were generally detected near a gas station, ferry, and an auto wrecking facility. Example mass spectra are shown in Figure 3.3 (panels A-C), with some ions of interest labeled on the spectra. The spectrum in panel A was collected near the municipal composting facility, the spectrum in panel B was collected near the auto wrecker, and the spectrum shown in panel C was collected near a sawmill. The spectrum collected near the auto wrecker is dominated by hydrocarbons, while the spectra collected near the compost and sawmill also contain ions for oxygenated species.

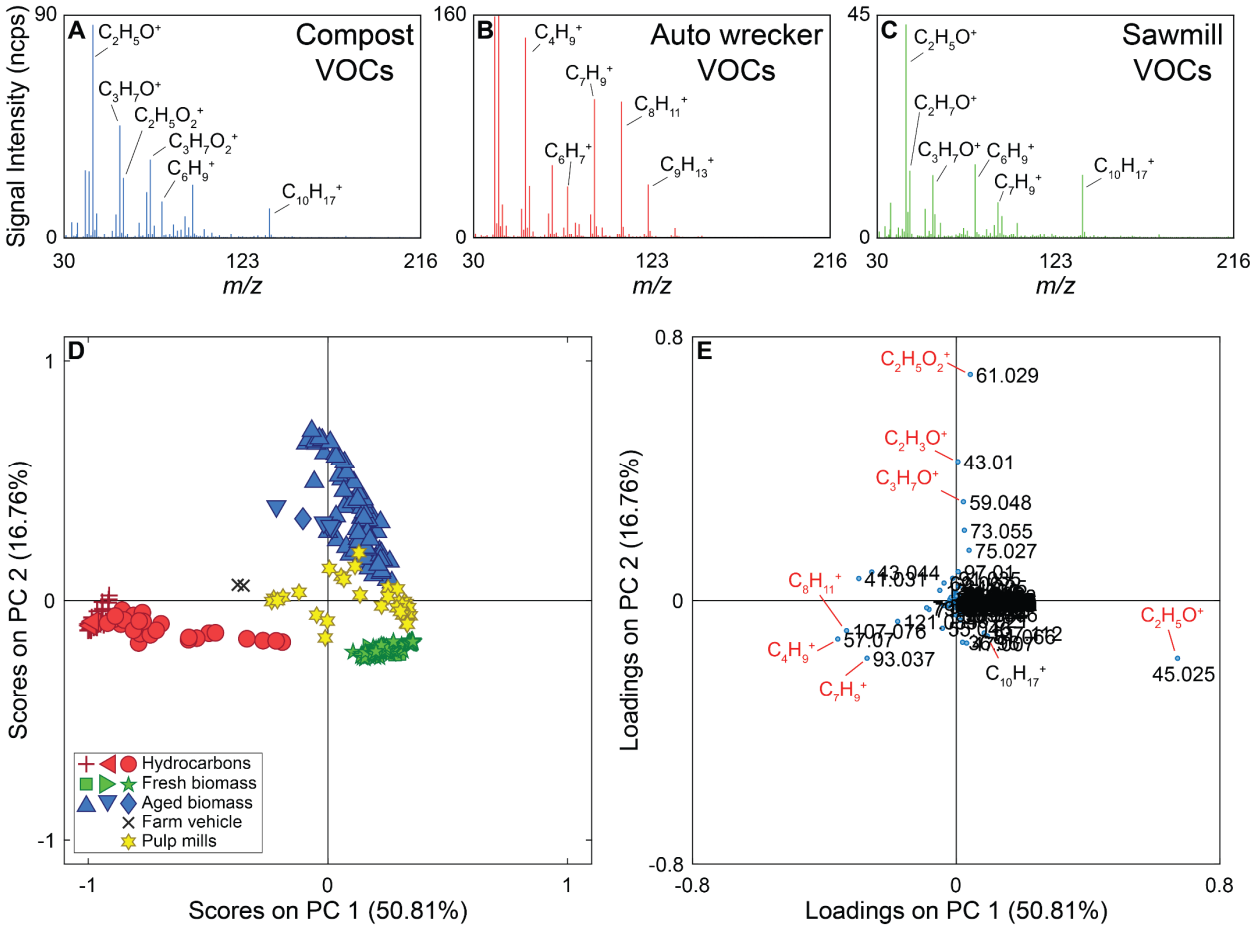


Figure 3.3: Top: Example full scan PTR-ToF-MS spectra collected near a composting facility (A), auto wrecking facility (B), and sawmill (C). Some ions of interest have been labelled on each spectrum. Panel D: PC 1 versus PC 2 scores plot of the full scan PTR-ToF-MS data. Red samples are from hydrocarbon sources, blue samples from aged biomass sources, green samples from fresh biomass sources, yellow samples from pulp mills (both fresh and processed biomass), and black samples were collected near a tractor carrying hay (hydrocarbon and biomass source). The different shapes represent different sources (*e.g.*, red stars are near a ferry, red diamonds are near an auto wrecking facility, and red circles are near a gas station). Hydrocarbon samples (red) can be discriminated from biomass samples based on the scores on PC 1. PC 2 discriminates aged biomass samples from fresh biomass. Panel E: PC 1 versus PC 2 loadings plot of the PTR-ToF-MS data. Ions associated with alkylated aromatic compounds (*m/z* 93.061, 107.076, and 121.085) allow for the discrimination of hydrocarbon sources from biomass sources due to their negative loading on PC 1. Along PC 2, acetic acid (*m/z* 61.029) has a high positive loading, while acetaldehyde (*m/z* 45.025) and monoterpenes (*m/z* 137.112) have negative loadings leading to the separation of fresh and aged biomass sources.

For the supervised PCA, 298 average mass spectra were used. The PCA scores and loadings plots for the PTR-ToF-MS data are shown in Figure 3.3 (panels D and E). PC 1 accounts for 51 % of the variance in the data set, with PC 2 accounting for 17 %. The Scree plot for the analysis is shown in Figure B.5 in the Supplemental Information with the first three PCs accounting for 78 % of the variance in the data set (>97 % of the variance is described by the first 10 PCs). Samples in the scores plot have been colour-coded based on the source type they represent (*e.g.,* hydrocarbon, fresh biomass), and the different shapes of the same colour indicate different encounters with the same source type. For example, we encountered three primarily hydrocarbon sources (red) and they are marked as triangles (ferry), plus signs (auto wrecking facility), and circles (gas station) in Figure 3.3.

In the scores plot, hydrocarbon samples have negative scores on PC 1, discriminating them from other source types. PC 2 provides discrimination between VOCs from fresh biomass, pulp mills, and aged biomass samples. The latter have positive scores on PC 2, samples with scores near zero on PC 2 are associated with pulp mill emissions, and samples with negative scores on PC 2 are due to emissions from fresh biomass near sawmills, wood chip trucks, etc. Additionally, samples collected near a farm vehicle carrying hay are located between the hydrocarbon and biomass samples on the scores plot, potentially due to the mixed nature of this source.

The loadings plot can be used to identify possible compounds leading to sample discrimination. For example, the $[M+H]^+$ ions for many hydrocarbons (*m/z* 93.061, 107.076, 121.084) have negative loadings on PC 1, which allows the hydrocarbon samples to be discriminated from the biogenic emissions. Acetaldehyde (*m/z* 45.025) has a negative loading on PC 2, while acetic acid (*m/z* 61.029) which was mainly detected in the vicinity of the compost facility has the highest positive loading on PC 2.

The PCA showing the scores and loadings for PC 1 versus PC 3 are shown in Figure B.6 in the Supplemental Materials. In this projection, the hydrocarbon samples are still separated from the biomass samples, as that discrimination falls along PC 1, but there is more overlap between the biomass and pulp mill samples. Some of the pulp mill samples have high positive scores on PC 3 due to the presence of DMS (*m/z* 63.0224) in these samples.

The mass spectra used for this analysis did not include the signal intensity for protonated methanol as this improved sample discrimination for samples impacted by biomass sources and removed the confounding influence of detecting periodic windshield washing fluid from vehicles during on-road sampling. The PCA including methanol is shown in Figure B.7 in the Supplemental Information. When methanol is included, hydrocarbon sources remain well separated, but there is some overlap between the pulp mill and fresh biomass samples and the samples associated with pulp mills and aged biomass sources overlap completely.

A regional scale map of the samples is shown in the top panel of Figure 3.4 along with the time series of the data (bottom panel). The drive route is shown by the black line, with the samples used in the supervised PCA colour-coded based on source type. The size of the dot is scaled by the total VOC concentration using the total ion current for *m/z* 30-215 (excluding the reagent ions). The labels on the map show the nearby VOC sources. Those labeled in black are stationary sources (mostly industrial), and those labeled in red were moving. The time series data shows the overlay of all measured *m/z* observed during the field campaign, with the exception of methanol as it was excluded from the chemometric analysis. Mass spectral peaks of interest are labelled on the time series, and the portions of the time series data collected near the identifiable sources used in the PCA are indicated by the coloured bars along the top of the plot. Figure B.8 (panels A-D) in the Supplemental Information

show time series of subsets of the measured compounds, with methanol in panel A, major hydrocarbon ions and their fragments in panel B, major non-methanol oxygenated species and their fragments in panel C, and sulphur compounds in panel D. It is also important to note that the on-road measurements can be influenced by mobile sources and are not always associated with stationary area sources. With the exception of acetic acid (which exhibits a slow decay time due to carry-over in the sample line), we observe only small differences in the decay times of the remaining VOCs described here (< 5 sec). These differences are dampened out for the supervised analysis (15 sec averaged mass spectra) and have only a minor influence on the unsupervised analysis using 1 second averaged mass spectra. In both cases, the decay times do not change the distribution of identified sources at the spatial resolution presented here.

Given the success of discriminating known VOC samples above, we have extended this method to identify groupings within the full data set in an unsupervised analysis. The complete dataset contains 16,737 mass spectra measured at 1 second intervals across 290 mass channels. As has been noted by others, the high sample-to-variable ratio reduces the likelihood of spurious correlations.[212] Two clustering algorithms (GMM and DBSCAN) were employed to group samples based solely on their mass spectral fingerprints. The GMM algorithm clusters all data points into groups, whereas DBSCAN only assigns dense data clusters within the PCA model. A PCA model was used as the algorithm input as this greatly reduced computational time, especially in the case of the GMM. A 15 component PCA model was used as the input. The first three PCs represent 44 % of the variance in the dataset. Most PCs past the 10^{th}, modelled noise, with only a few beyond that showing some structure related to specific geographic locations (*i.e.*, PCs 12 and 15). The 15 PC model describes 79 % of the variance in the data, with each subsequent PC accounting for less than 1.2 % of the variance. It was also found that the addition of more PCs had little impact on the results of the clustering algorithms (data not shown).

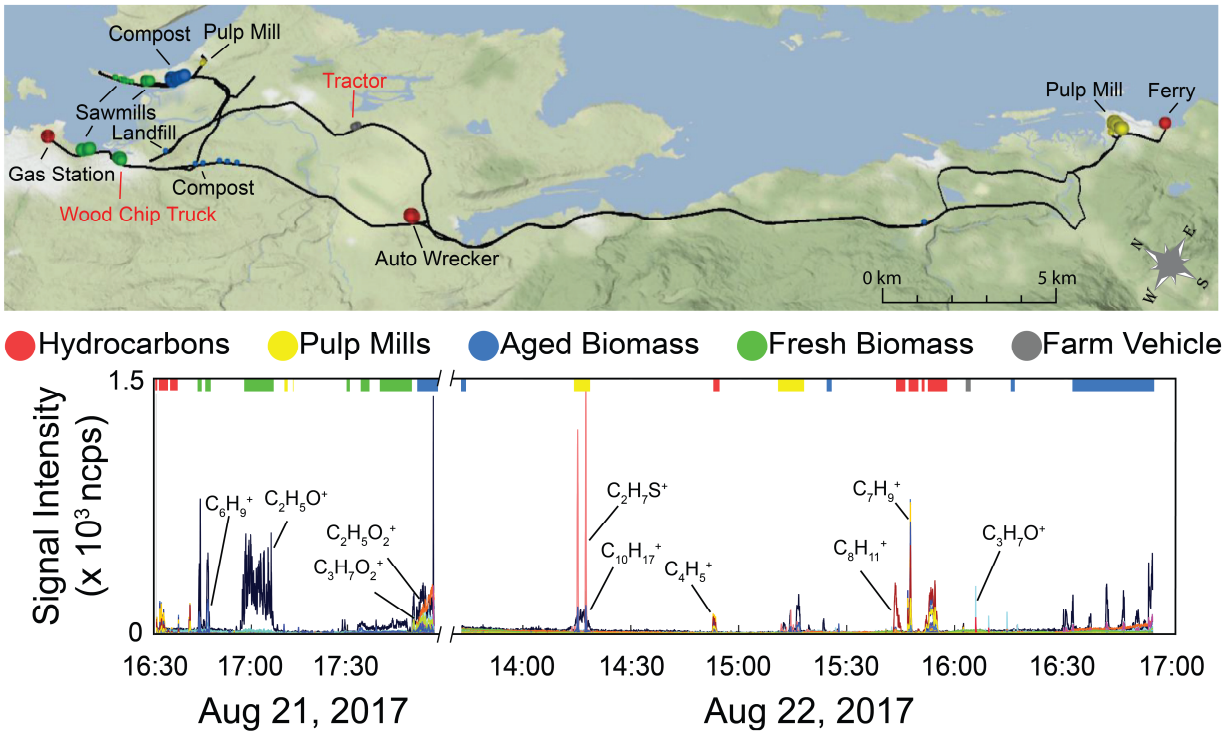


Figure 3.4: Top: Map of samples used in the PTR-ToF-MS analysis allowing the spatial distribution of the samples to be visualized. The size of the dots are proportional to total VOC concentration encapsulating the 40th, 60th, 80th, and 100th percentiles of the total ion current for the 298 samples used in the PCA. The dots are colour-coded based on sample source (red – hydrocarbons, yellow – pulp mills, blue – aged biomass, green – fresh wood, grey – farm vehicle. Bottom: Time series of all the *m/z* measured on August 21 shown on the left of the x-axis cut, and August 22, shown on the right. The coloured bars on the top of the graph indicated which data was used in the supervised PTR-ToF-MS analysis and are coloured based on source type.

For the calculated GMMs both the AIC and BIC values were minimized using unshared, full covariance matrices for 2-20 clusters. For these models, the AIC and BIC values decreased with increasing cluster number, with very similar values obtained for 12-20 clusters. Additional AIC and BIC values were calculated for up to 40 clusters using the models with full, unshared covariance matrices. The plots of the AIC and BIC values are found in Figure B.9 of the Supplemental Information. The calculated AIC values are similar for models using 12 or more clusters (they continue to decrease up to 40 clusters), whereas the BIC values reach a minimum value at 27 clusters. As BIC penalizes model complexity more than AIC[162], the BIC minimum value of 27 was used as an indicator of the most complex model needed to describe the data, with models containing 9, 10, 11, 12, 13, and 14 clusters also being calculated. Geospatial maps of the different models were compared, and the 12 group model was selected as it was able to identify sources within the dataset, without overcomplicating the model. The PCA scores plots of PC 1 versus PC 2 and PC 1 versus PC 3 for this analysis are shown in Figure 3.5 (panels A and B), with a map of the results in Figure 3.5 (panel C). Many of the point sources on the map are identified by the large dots in their vicinity (gas station, compost facility, ferry terminal, pulp mill, auto wrecker). Samples are coloured based on group membership as determined by the 12 group GMM. Figure B.10 in the Supplemental

Information show the loadings plots of PC 1 versus PC 2 and PC 1 versus PC 3 in panels A and B respectively. In the scores plots similar samples are grouped together, with many point sources grouping along the corners and edges of the data. Groups with individual sources or source types have been identified on the scores plot. For clarity, Figure B.11 (panels A-L) depicts the PC 1 versus PC 2 scores plots for the individual groups, as well as the average mass spectra for each of the groups with some ions of interest labeled. Scores along PC 2 discriminate hydrocarbon sources from pulp mill samples, fresh, and aged biomass samples. PC 1 can be used to discriminate the biomass sources, with aged biomass having negative scores on PC 1 and fresh biomass having positive scores on PC 1.

The loadings plots shown in Figure B.10 (panels A and B) indicates that similar ions lead to sample discrimination in the supervised and unsupervised data sets, with signals associated with acetic acid, acetaldehyde, formic acid/ethanol, ethylbenzene/xylenes, and toluene having high loadings. For the groups identified using GMM, some groups were associated with a single source, while others contained samples from multiple sources. For example, Group 10 (magenta triangles) was only measured near a municipal composting facility with a mass spectrum dominated by a $C_2H_3O^+$ ester or acid fragment, acetaldehyde, acetone, acetic acid, and monoterpenes signals (Figure B.11, panel J) while Group 9 (red triangles) was collected mainly in the vicinity of a pulp mill with signals associated with acetaldehyde, acetone, acetic acid, DMS, and monoterpenes (Figure B.11, panel I). Other groups were associated with multiple sources with similar emissions. For example, Group 5 contained samples collected near multiple sawmills, and from driving near a wood chip truck with a mass spectrum influenced by acetaldehyde and monoterpene fragments (Figure B.11, panel E); and Group 12 is impacted by the three hydrocarbon sources, and an unknown source, with a mass spectrum dominated by the BTEX suite and other alkylated aromatics (Figure B.11, panel L). Finally, some sources are present in multiple groups in the scores plot. For example, a group of pulp mill samples has positive scores on PC 1 and PC 2 (Group 9, red triangles), while other pulp mill samples have positive scores on PC 1 and negative scores on PC 2 (Group 3, blue diamonds). As pulp mills have multiple different sources for VOC emissions (*e.g.*, wood chip piles, pulp production, wastewater ponds), it is not surprising that not all the pulp mill samples are grouped together. Some VOC sources encountered during the field campaigns were unidentified at the time of sampling as can be seen in the grey shaded areas in Figure 3.1. We have tentatively assigned source classes to these VOC signals based on the group membership in the unsupervised data analysis. For example, the unidentified signals that appear at 16:40 in Figure 3.1 could be attributed to a hydrocarbon source, whereas those at 16:53 and 17:15-17:25 could be assigned to pulp mill emissions. The assignment of the samples collected from 17:15-17:25 is supported by field observations noting a sulphur smell. Given the limited data presented here, we have chosen not to build a

classification model. Future field campaigns will include more sources and replicates to enable us to train and verify a robust predictive model.

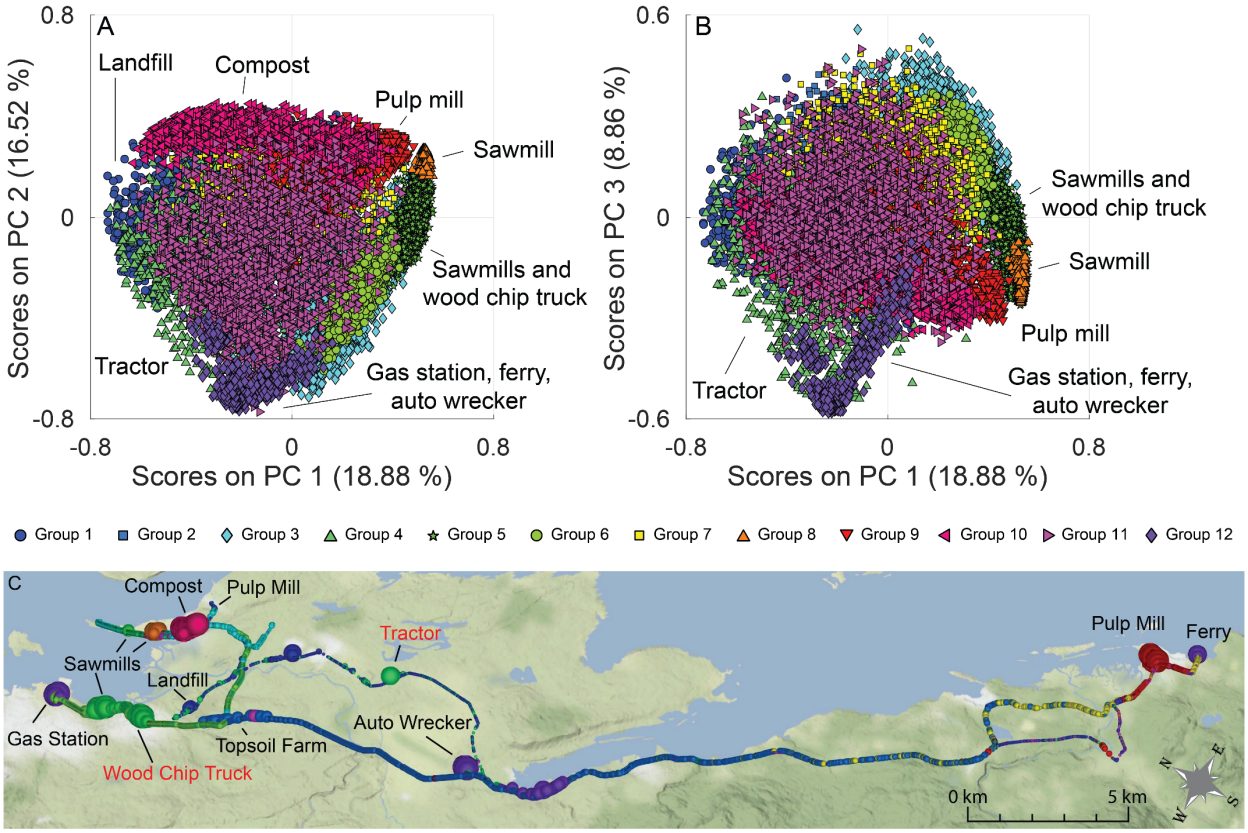


Figure 3.5: A) PCA scores plot for PC 1 versus PC 2. B) PCA scores plot for PC 1 versus PC 3. Samples are coloured based on the groups identified using the GMM algorithm. Some groups containing individual VOC sources (such as the municipal compost facility, and one of the pulp mills) or multiple VOC sources of the same type (e.g., sawmill emissions are grouped with the samples from the wood chip truck) are labeled. C) Geospatial map showing the spatial distribution of the samples, with dot size proportional to total VOC concentration. Locations of interest are labelled, with moving sources being represented by red text.

For the DBSCAN analysis a neighbourhood size of 0.022 was used with 9 neighbours being needed to identify core points. This algorithm identified 13 groups within the data (1 of which, Group 13, represents data that was not clustered). Unlike the GMM algorithm, DBSCAN has not assigned all data points to a cluster, however it has identified most of the high VOC concentration samples within the dataset. The results of the analysis are seen in Figure B.12, with the PC 1 versus PC 2 and PC 1 versus PC 3 scores plots shown in panels A-B, and maps of the analysis shown in panels C and D. The data is coloured by the DBSCAN group membership, with the black dots representing the data left ungrouped. The map in Figure B.12 (panel C) shows all data, while the map in Figure B.12 (panel D) omits the black data points for clarity. The PC 1 versus PC 2 scores plots and average mass spectra for each of the groups is shown in Figure B.13 (panels A-M) in the Supplemental Information. Samples are clustered based on source type, with most of the clusters falling on the extremities of

the scores plot. Group 8, which has scores near the origin on each of the three PCs shown contains high concentration samples measured in the vicinity of a pulp mill at approximately 14:15 on August 22, 2017. These samples were higher in DMS (*m/z* 63.022) than those measured elsewhere, but DMS does not have a significant loading until higher PCs (*e.g.*, PCs 6, 7, 12, and 15 (data not shown)). Groups 2, 4, 5, 8, and 12 have identified single sources within the data (gas station, wood chip truck, compost, pulp mill, and an unknown (possibly hydrocarbon) source respectively), while groups 3 and 10 contain samples from multiple sources of similar type (fresh biomass samples and hydrocarbon samples).

In some cases, the clusters identified by GMM have also been identified by DBSCAN, while in other cases DBSCAN has identified multiple clusters within a single group identified by GMM. For example, GMM Group 10 (Figure B.10, panel J) and DBSCAN Group 5 (Figure B.12, panel E) contain almost the same set of samples. The mass spectra associated with them have a correlation coefficient of 0.999 (Table B.6). In the case of GMM Group 12 (Figure B.10, panel L), the DBSCAN has identified 3 sub-clusters (Figure B.12, panels B, J, and L) with correlation coefficients between the identified mass spectra of 0.916, 0.990, and 0.604 for DBSCAN Groups 2, 10, and 12, respectively. In this case, DBSCAN has separately identified different hydrocarbon sources described by mass spectra containing different ratios of the measured hydrocarbons, while GMM has identified one cluster containing all the hydrocarbon samples. A full correlation matrix between the mass spectra for the clusters identified by the two methods is shown in Table B.6.

The purpose of the unsupervised analysis done using both GMM and DBSCAN algorithms was not to construct a definitive model of the data collected, but to demonstrate the ability of the techniques to identify different VOC sources measured using a PTR-ToF-MS operated in a moving vehicle. Both techniques were able to discriminate VOC sources using the normalized mass spectral fingerprint and associate these sources with specific geographic locations.

3.4 Conclusions

The work presented here represents the first application of chemometric analysis to mobilized direct mass spectrometry for VOC source discrimination and fine scale spatial distribution mapping. Supervised PCA analysis of normalized full scan mass spectra provide meaningful sample discrimination and an unsupervised analysis uncovers additional VOC source assignments. We demonstrate the use of MIMS to discriminate various hydrocarbon sources. The PTR-ToF-MS instrument is not constrained to small hydrophobic molecules and is generally more sensitive and selective, providing a richer dataset inclusive of more polar VOCs including those often associated with nuisance odours. Consequently, it can be used to discriminate samples from a wider range of sources. The data is displayed on a geospatial map for visualization and further interrogation. The use of non-

targeted full scan mass spectral data without determining the concentrations of individual VOCs reduces the amount of data processing and will facilitate real-time source identification.

Mass spectrometry yields molecular level information about the sample composition, which provides chemical insight that can be used to inform targeted analysis, or be used to identify molecular markers. Although the analysis presented here cannot identify VOC sources that have not been previously encountered, it will direct the user to 'unknowns', which can be the subject of subsequent investigation. Future work includes the application of data fusion techniques to improve source discriminating power by incorporating data from inorganic gas and particulate matter sensors. The incorporation of this data into a Geographic Information System will allow us to map plume boundaries and inform dispersion models. We are currently exploring PMF and MCR-ALS to identify sources contributing at each sampling location and apportion their relative contributions to ambient air samples.

Chapter 4 - Source apportionment of atmospheric volatile organic compounds using proton-transfer reaction mass spectrometry data collected from a moving vehicle

4.1 Introduction

Air samples are often influenced by VOCs from more than one source, and the composition of air samples can be dynamic as a result of mixing and dispersion. Given that individual VOCs differ in their O_3 and SOA forming potentials and the importance of the VOC to NO_x concentration ratio in ground-level O_3 production, understanding the chemical composition of the mixture of VOCs present at different times and locations as well as their potential sources are important factors in improving and managing air quality. Chapter 3 focused on the use of PCA and clustering algorithms to discriminate VOC samples in ambient air that had been measured using full scan direct mass spectral data collected from a moving vehicle. The goal of this chapter is to apply receptor modelling techniques to time series PTR-ToF-MS data collected on-road to identify and apportion the source(s) of VOCs that are impacting each sample.

Receptor models have been used to apportion contributions of atmospheric particulate matter[213,214], and in recent years their use has been extended to the apportionment of ambient VOCs.[75,95,148,215,216,137–142,145,146] Using ambient VOC data sources such as vehicle exhaust, biogenic emissions, petrochemical production, industrial activities, or solvent usage have been identified based on the chemical composition and temporal profiles of the sources. PMF is widely used in the air quality community for source identification of PM and VOCs.[75,95,137,138,145,146,151,215,216] MCR-WALS has been shown to resolve comparable source profiles and contributions to those identified using PMF, while those resolved by MCR-ALS tend to differ more significantly.[135,217,218] PMF, MCR-ALS, and MCR-WALS have all been applied to environmental data for source apportionment.[217,219]

While these receptor models only require the chemical data (and the uncertainties in the case of PMF and MCR-WALS) as inputs, the number of sources to be identified by the model needs to be specified. The number of sources chosen in the model impacts the calculated source profiles and contributions; three and four source models calculated for the same data may not have three sources in common. This is in contrast to PCA, where moving from an *N* component model to an *N+1* component model results in two models where the first *N* components are identical, and the $N+1^{th}$ component describes the maximum variance of the data in the

residuals of the *N* component model. Several methods can be used to estimate the initial number of sources in the data such as singular value decomposition (SVD), evolving factor analysis, and windowed factor analysis, or models with different numbers of components can be tried.[179]

The MCR-ALS algorithm takes the data matrix as an input, and for a given number of sources, calculates a solution by minimizing the sum of squared residuals (*SSR*) using an alternating least squares algorithm as shown in Equation 4.1:

$$SSR = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_{ij} - \hat{x}_{ij}\right)^2 \qquad \text{Equation 4.1}$$

where x_{ij} is the measured signal intensity for the *j*th variable in the *i*th sample, and $\hat{x}_{ij}$ is the modelled signal intensity for the same variable and sample.[218] For PTR-ToF-MS data collected from a moving vehicle, the measured *m/z* values are the variables and samples are the timestamped measurements. The MCR-ALS algorithm assumes the errors in the data are independent and identically distributed (i.i.d),[218] which is not the case for the data collected with the PTR-ToF-MS, where the uncertainty in the measurement increases with signal intensity.[220] PMF and MCR-WALS receptor models incorporate measurement errors in the analysis. While the algorithms used in MCR-WALS and PMF differ, they have been shown to provide comparable results.[135,218] MCR-WALS and PMF algorithms both require an additional input matrix, **∑** (*m* by *n*), containing the estimated measurement uncertainties.[218] PMF and MCR-WALS incorporate the uncertainties by minimizing the weighted sum of squared residuals (WSSR) as shown in Equation 4.2:

$$WSSR = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\left(x_{ij} - \hat{x}_{ij}\right)^2}{\sigma_{ij}^2} \qquad \text{Equation 4.2}$$

where σ_{ij} is the estimated uncertainty of the signal intensity for sample *i* and variable *j*.[218] In PMF analysis, WSSR is also defined as Q.

A method for calculating the error matrix for ToF-MS data is outlined in Yan *et al.*[220] The uncertainty in the measurement is calculated using Equation 4.3:

$$\sigma_{ij} = s_{ij} + s_{noise} \qquad \text{Equation 4.3}$$

where σ_{ij} is the estimated measurement uncertainty, s_{ij} is the analytical uncertainty in the measurement, and s_{noise} is the instrumental uncertainty. The instrumental uncertainty can be determined by taking the median value of the standard deviations in mass channels where recorded signals were not due to the presence of ions. The analytical uncertainty is calculated for ToF-MS data using Equation 4.4:

$$s_{ij} = a\sqrt{I_{ij}}/\sqrt{t_s} \qquad \text{Equation 4.4}$$

where I_{ij} is the signal intensity, t_s is the spectrum integration time, and a is an empirically determined factor. To calculate a from the experimental data, the signal intensity of steady-state signals, their corresponding uncertainties, and the integration time used in the data collection are fit to Equation 4.4. The values for s_{noise}, and the formula for s_{ij} can then be substituted into Equation 4.3 to calculate the uncertainty, σ_{ij}, for each measurement in the data to be analyzed. For data below the detection limit (estimated as 3 times the signal-to-noise ratio (S/N) of 'VOC free' air measurements) the uncertainty was specified as 5/6 the detection limit.[217]

The data presented in this chapter was measured using a PTR-ToF-MS system operated in a moving vehicle, with mass spectra being measured at 1 Hz. Previous studies utilizing PTR-MS data for VOC source apportionment have used longer sample averaging to identify sources impacting the sample site, however the majority of these studies have been done using data collected at stationary sites.[77,78,98,215,216] Data collected at a high frequency from a moving vehicle includes concentration excursions measured for short durations (seconds) while driving past stationary or moving VOC sources, and high frequency measurements better preserve both the location and the signal intensity associated with these sources.

PMF is widely used in the literature for source apportionment, and MCR-WALS provides similar results.[135,218,219] The EPA PMF 5.0 program is available for free from the EPA website, however, dataset size is limited and was incompatible for direct use with our high sampling frequency data. Additionally, PMF can treat rare, brief, short concentration excursion samples as outliers, further down-weighting them in the analysis[221], and these types of measurements were of interest to us as they are often made while driving. Therefore, in this chapter MCR-WALS was used as an open source alternative approach[222] and a reduced dataset was used to compare the model results to those generated by MCR-ALS and PMF.

Overall, the goal of the work presented here is to use mass spectral data collected from a PTR-ToF-MS system operated in a moving vehicle to identify and map VOC source distributions at a neighbourhood to regional level, with high-temporal (1 second) and spatial (10s of meters) resolution. The results of MCR-WALS analysis from three field campaigns on Vancouver Island, BC, Canada will be presented, with variable selection to remove the recorded mass channels with high levels of noise being performed to improve model results. The methodology used can be applied to future field campaigns which aim to determine common VOC sources in a region, identify locations with potential high emissions, and observe temporal and spatial trends.

4.2 Experimental

The analysis of three datasets collected during field campaigns on central Vancouver Island, BC, Canada are discussed in this chapter. The first was collected on August 7, 2018 in Port Alberni, BC; the second was collected between Nanaimo and Parksville, BC on August 14, 2018; and the third, also analyzed in Chapter 3 by PCA, was collected August 21–22, 2017 between Nanaimo and Crofton, BC. The study areas included residential, commercial, industrial, and rural/light agricultural uses. Port Alberni is located at the end of a marine inlet in the Alberni Valley and has a population of over 20,000.[223] The city hosts residential and commercial activities and is bisected by an important transportation corridor from and to the west coast of the island. The region includes an active forestry sector and related industry (pulp mill, sawmill) and active marine port. The sampling route includes regions of old growth forest, light agriculture, an open-aerated sewage lagoon, and a landfill site. A description of the Nanaimo area can be found in Chapter 3, with the sampling on August 14, 2018 extending North to Parksville (population 12,500), and including sampling near a pulp mill, sawmills, a pollution control center, composting facility, and waste transfer site. Field notes identified suspected stationary VOC point sources (*e.g.*, gas stations, industrial sites, composting facilities), mobile VOC sources (*e.g.*, vehicle exhaust, transport trucks), and area sources (*e.g.*, forest fire smoke) encountered while driving.

The PTR-ToF-MS (PTR-TOF 1000, Ionicon Analytik Ges.m.b.H, Innsbruck, Austria) used for this study was mounted in a research purposed cargo van (4x4 2500 Cargo 144 Sprinter, Mercedes-Benz, Nanaimo, BC, Canada) as described in Chapter 3.[224] Additional sensors in the mobile lab, including an optical particle sizer (OPS Model 3330, TSI Incorporated, Shoreview, MN, USA) for PM measurements, GHG Analyzer (FGGA 30r, Los Gatos Research, San Jose, CA, USA) to measure carbon dioxide (CO_2), methane (CH_4), and water vapour concentrations, and an NO_x analyzer (42i, Thermo Fisher, Waltham, MA, USA) measuring nitric oxide (NO), NO_2, and NO_x (NO + NO_2) concentrations, were used to help with the identification of VOC sources.

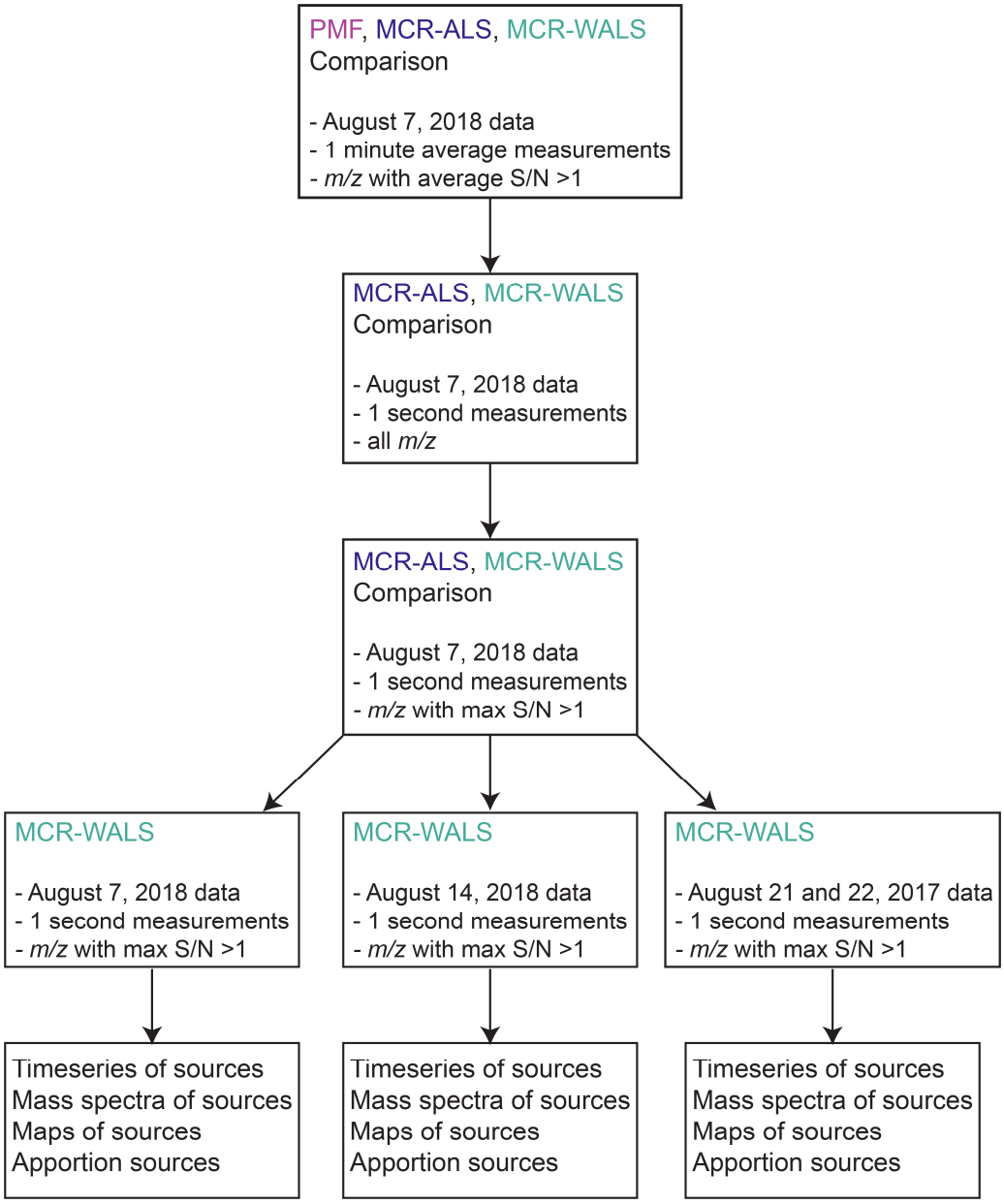


Figure 4.1: Flow chart of data analysis in this chapter.

A flow chart of the data analysis in this chapter is shown in Figure 4.1, and briefly described here. A reduced version (60 second sample averages) of the dataset collected on August 7, 2018 was used to compare PMF, MCR-ALS, and MCR-WALS. Next, MCR-ALS and MCR-WALS comparisons were made using the whole August 7, 2018 dataset (1 second samples) with and without variable selection to reduce the number of measured *m/z* included in the analysis. Finally, MCR-WALS on a subset of the measured *m/z* (chosen with variable selection) was used to identify, map, and apportion the VOC sources encountered in the three sampling campaigns. The time series and mass spectra of the resolved sources, as well as geospatial maps of the source distribution and apportionment is presented for each of the three datasets.

4.2.1 Mobile lab instrumentation

The set-up and operation of the PTR-ToF-MS in the mobile mass spectrometry lab, as well as the sample line set up for the 2017 field work, is described in Chapter 3. For the 2018 field campaigns the PFA sample line (4 m long, OD ¼", ID 3/16", Cole-Parmer, Montreal, Quebec, Canada), with an in-line 1 µm filter (Swagelok, Solon, OH, USA), was mounted on an aluminum sampling arm extending 1 m above the roof at the front of the vehicle. The sample line was not heated, and ambient temperatures during sampling ranged from 17–37 °C. H_3O^+ was used as the reagent ion. The E/N value was 130 Td, the drift pressure was 2.20 mbar, and the ToF pressure was 1.2 x 10^{-6} mbar. The instrument was mass calibrated every 5 s using the signals at *m/z* 21.022, 29.999, and 59.049 (H_3O^{18+}, NO^+, $C_3H_7O^+$), and the mass resolution was 1500–2000 at *m/z* 59. The air sampling flow rate was 250 standard cubic centimeters per minute (sccm) on August 21 and 22, 2017 and 750 sccm on the other two sampling days. A mass spectrum was collected every 1 s. The transit time between the sample inlet and the mass spectrometer was about 3 seconds (750 sccm flow rate) or 10 seconds (250 sccm flow rate). The instrument is sensitive to sub ppb_v levels for most VOCs. High resolution positional data was collected using a roof mounted antenna (Hemisphere A45, Scottsdale, AZ, USA) and GPS receiver (Hemisphere R330, GNSS Receiver, Scottsdale, AZ, USA). The OPS, mounted on the roof of the mobile lab in an enclosed case with a vertical 20 cm anodized aluminum sample inlet, collected PM data at a frequency of 1 Hz over 5 size bins (0.3–0.5 µm, 0.5–1 µm, 1–2.5 µm, 2.5–8 µm, and 8–10 µm). The GHG and NO_x analyzers were mounted inside the mobile lab, with FEP sampling lines extended on the aluminum sampling arm. Both instruments sampled using their on-board pumps (GHG, 500 sccm; NO_x 1000 sccm), with the GHG analyzer sampling at 1 Hz, and the NO_x analyzer at 0.1 Hz. Data from the onboard sensors (excluding the mass spectrometer) was logged and mapped in real-time using custom software (Underway, Beaver Creek Analytical, Lafayette, CO, USA) during the 2018 field campaigns.

4.2.2 Data analysis

The majority of the data analysis described below was performed using a combination of Excel (Microsoft Corporation, Redmond, WA, USA), MATLAB (Mathworks, Natick, MA, USA), and the PLS_Toolbox (Eigenvector Research, Manson, WA, USA), with the exception of PMF which was performed using the EPA PMF 5.0 software (US Environmental Protection Agency, Durham, NC, USA).

4.2.2.1 Exporting and preparing mass spectral data for analysis

Mass lists for the PTR-ToF-MS were made using the PTRViewer 3 software (Version 3.2.8 Ionicon). Mass calibration for each of the data files was verified in the software using 5 scan averaging. After mass calibration, the spectra in each data file were summed up and then used to identify all *m/z* peaks present in the data set. For overlapping peaks, the multipeak tool in the PTRViewer 3 software was used to calculate overlapping mass

spectral peak areas. Peak areas were corrected to account for differences in mass dependent transmission through the mass spectrometer using the instrument's transmission curve. The data collected for each field campaign was combined into a single file for interrogation. Signal intensities were normalized to the reagent ion signal at *m/z* 21.022 (H_3O^{18+}) to account for ionization differences, and background subtracted using the mass spectrum of Ultra Zero Air (99.999 %, Praxair, Nanaimo). Negative signal intensities resulting from the background subtraction were replaced with zero.

4.2.2.2 Error matrix calculation

The uncertainty in the PTR-ToF-MS measurements were estimated from lab and field data using Equation 4.3 and Equation 4.4. Instrumental uncertainty (s_{noise}) was calculated as the median value for the standard deviation of the signal intensities between *m/z* 225-279 across Ultra Zero Air measurements and a subset of field data (stationary and moving), as ions with *m/z* > 206 were rarely encountered while field sampling. Using this method s_{noise} was calculated to be 0.12 ncps as shown in Figure 4.2.

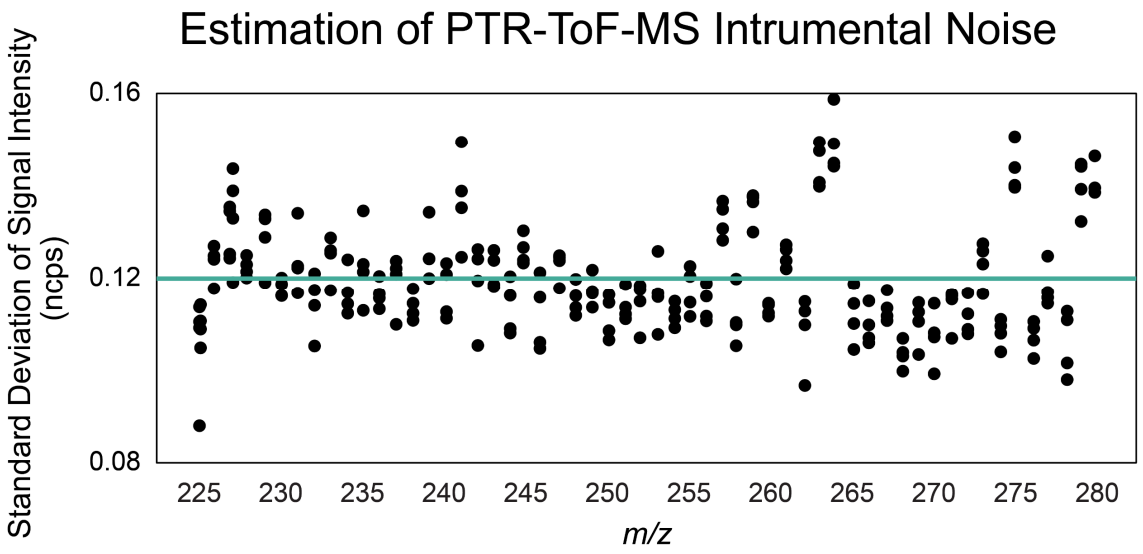


Figure 4.2: Data used to estimate instrumental noise in the PTR-ToF-MS data. Each dot is the standard deviation in the signal intensity at a particular *m/z*. The teal horizontal line at 0.12 ncps is the median value of the standard deviations. This value is the estimated instrumental noise in the data (s_{noise}).

To determine the analytical uncertainty (s_{ij}) for each measurement, *a* (Equation 4.4) needed to be determined using the average and standard deviations steady-state signal intensities generated in the lab and encountered in the field. Steady-state signals for the $[M+H]^+$ ions of benzene, toluene, ethylbenzene, DMS, methanethiol, chlorobenzene, α-pinene, *m*-xylene, and dimethyl disulphide (DMDS) were generated using a Dynacalibrator Gas Dilution System (Model 340, VICI Valco Canada, Brockville, ON, Canada), and permeation tubes (KIN-TEK Analytical, Inc, La Marque, TX, USA). For the field data, sampling times where steady-state signals

for some *m/z* were observed while driving were identified and included in the determination of *a*. Between the field data and lab data, average steady-state signal intensities and standard deviations were calculated for a selection of *m/z* between 33.03 and 137.12. To determine *a*, a plot of signal intensity (*x*) versus standard deviation (*y*) was fit to Equation 4.5.

$$y = ax^{0.5} + b \qquad \text{Equation 4.5}$$

With the exception of infrequent, very large concentration excursions for some *m/z* [*e.g.*, *m/z* 33.03 (CH_4OH^+, methanol), *m/z* 59.05 ($C_3H_6OH^+$, acetone), *m/z* 61.03 ($C_2H_4O_2H^+$, acetic acid), *m/z* 107.08 ($C_8H_{10}H^+$, ethylbenzene/xylenes)] the majority of the signal counts measured while field sampling were well below 1000 ncps, therefore calibration and field data with signal counts less than 700 ncps were used in the determination of *a*. Using this method, the resulting equation is $y = 1.3x^{0.5} - 1.4$, which gives an *a* value of 1.3 for 1 second measurements, as shown in Figure 4.3.

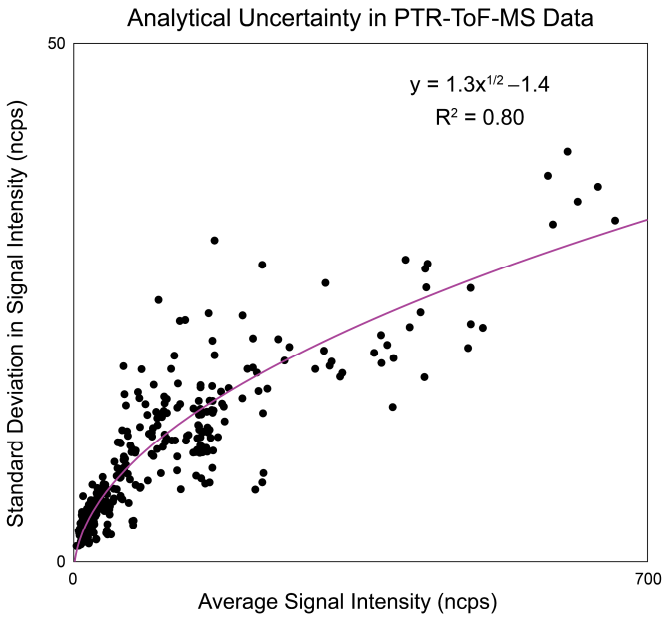


Figure 4.3: This plot of the average and standard deviation of steady-state data collected in the lab and in the field was used to determine the parameter *a*. The burgandy line is a fit of the data, resulting in an *a* value of 1.3.

Combining Equation 4.3 and Equation 4.4 with the estimated analytical and instrumental uncertainty, allowed uncertainties, σ_{ij}, above the detection limit to be estimated as shown in Equation 4.6:

$$\boldsymbol{\sigma_{ij} = 1.3\sqrt{I_{ij}} + 0.13} \qquad \text{Equation 4.6}$$

For data below the detection limit the uncertainty was estimated as 5/6 the detection limit.[217]

4.2.2.3 Variable selection

For each day of sample collection, some measured *m/z* had measured signal intensities that were slightly above, at, or below, the detection limit across the entire day of measurements. As these *m/z* only added noise to the analysis, a comparison between VOC source models with and without them were carried out. In a PMF analysis it is recommended that variables with high noise levels (calculated as average S/N), be downweighted or excluded from the analysis. S/N values for each measurement are calculated as shown in Equation 4.7, where the signal is the measurement (x_{ij}) minus the uncertainty (σ_{ij}).[217]

$$s/n_{ij} = \frac{x_{ij} - \sigma_{ij}}{\sigma_{ij}} \qquad \text{Equation 4.7}$$

Defined in this way, signals that are twice the uncertainty have a S/N of 1. In general for PMF, variables with average S/N values less than 0.5 are excluded from the analysis, variables with average S/N between 0.5 and 1 are down-weighted further, and variables with average S/N greater than 1 are used as is.[225]

Initially, the same approach was taken here, however, due to the nature of the data sets (1 second sampling, ~13 m spatial resolution at 50 km/hr drive speed), many measured *m/z* had signal intensities with an average S/N < 1 across the dataset but contained high signal intensity measurement excursions over short periods of time (seconds–minutes) that were associated with geographic locations or encounters with other mobile sources. In most cases, measured *m/z* that had little or no temporal or spatial variability in their recorded signal intensities had S/N < 1 across all measurements, while those that had measurements associated with concentration excursions due to sources encountered while sampling had a measured S/N > 1 at least once in the time series. For MCR-WALS models on 1 Hz data with variable selection, *m/z* values with all measurement having S/N < 1 were excluded from the analysis.

4.2.2.4 Determination of number of components to describe VOC sources

Initial estimates of the number of components (sources) to be resolved for each field campaign were done by examining a log plot of the eigenvalues of a PCA (using the SVD algorithm) to estimate the number of sources of variance in the datasets.[226] PCA was applied to the complete set of recorded *m/z* values up to *m/z* 205.17, (highest mass with recorded signal during the field campaigns presented here), excluding reagent ions. Before PCA, mass spectra were normalized using the 1-norm and the measurements for each recorded *m/z* was then mean centered. This resulted in initial estimates of 7 components for August 7, 2018 (Figure 4.4), 7 components for August 14, 2018, and 8 components for August 21 and 22, 2017 as starting points for receptor modeling.

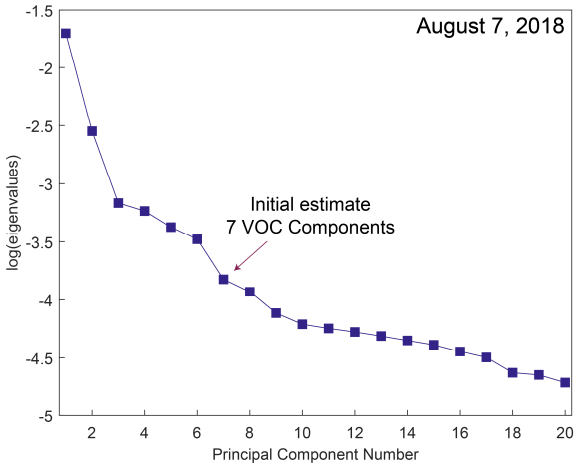


Figure 4.4: Log plot of the eigenvalues of the August 7, 2018 PCA. This plot indicates that there may be seven sources of variance in the data, and seven will be used as an initial estimate of the number of VOC sources in the data for receptor modeling.

In order to determine the likely number of VOC components present in each dataset, a combination of Meisch coefficients, scatter plots of the signal intensity measured and modelled for each recorded *m/z*, overlays of the measured and modelled time series data for each recorded *m/z*, and inspection of the resolved time series and mass spectra for each component were used. The MATLAB script to calculate Meisch coefficients and generate scatter plots was adapted from a MATLAB function provided by Glenn Johnson (personal communication).[227]

The Meisch coefficient is a variable-by-variable goodness of fit parameter. For each recorded *m/z,* the Meisch coefficient can be used to describe how well the modelled signal intensity agrees with the measured signal intensity. They can be used as an alternative to calculating an aggregate goodness of fit parameter for the whole data set. The Meisch coefficient is calculated for each variable (*m/z*) as shown in Equation 4.8.

$$r_i^2 = \frac{s(x)_i^2 - s(d_i)^2}{s(x)_i^2} \qquad \text{Equation 4.8}$$

In this equation, $s(x)_i^2$ is the variance of the variable in the original data, d_i is the residual between the measured signal and the modelled signal ($x_i - \hat{x}_i$), and $s(d_i)^2$ is the variance of the residuals. Variables with Meisch coefficients close to 1 are well modelled.[133] Some example scatter plots of measured and modelled data for a range of Meisch coefficients are shown in Figure 4.5. In the top left, and center left scatter plots, most of the data does not fall along the diagonal as the modelled data does not described the observations, and the Meisch coefficients for these two variables are close to zero. This indicates that a model with more components is required to model the observations. In the bottom center and bottom right scatter plots, the modelled data describes the observations well, with Meish coefficients of 0.999 and 1.00, respectively. The other two scatter

plots fall somewhere in between. When comparing the results of models with different numbers of components (*e.g.*, MCR-WALS with 5 or 6 resolved components), differences in the Meisch coefficients for individual variables can be compared to identify when the addition of a new component results in an improvement in the modeling of a particular variable (*e.g.*, Meisch coefficient increases from 0.2 to 0.7 when an additional component is added to the model). Used in this manner, the changes in the Meisch coefficients informed the optimum number of sources contributing to the data.

Scatter plots and overlaid time series of the observed and modelled signal intensities for each recorded *m/z* were inspected to determine which models were in good agreement with the data, and to identify samples or variables that may require a model with more components to describe the observed information.

When comparing results from the same algorithm with different numbers of components, the time series and mass spectra of the resolved components were compared to identify degenerate components which occur when one component is split into multiple components indicating that too many source components had been resolved in the model. In the case of the data discussed here, degenerate components occur when slight differences in the time series of the measured *m/z* contributing to a component lead to the component being split in two to better fit the measured data. While mathematically the model with more components represents a better fit, the degenerate components lead to a less meaningful result.

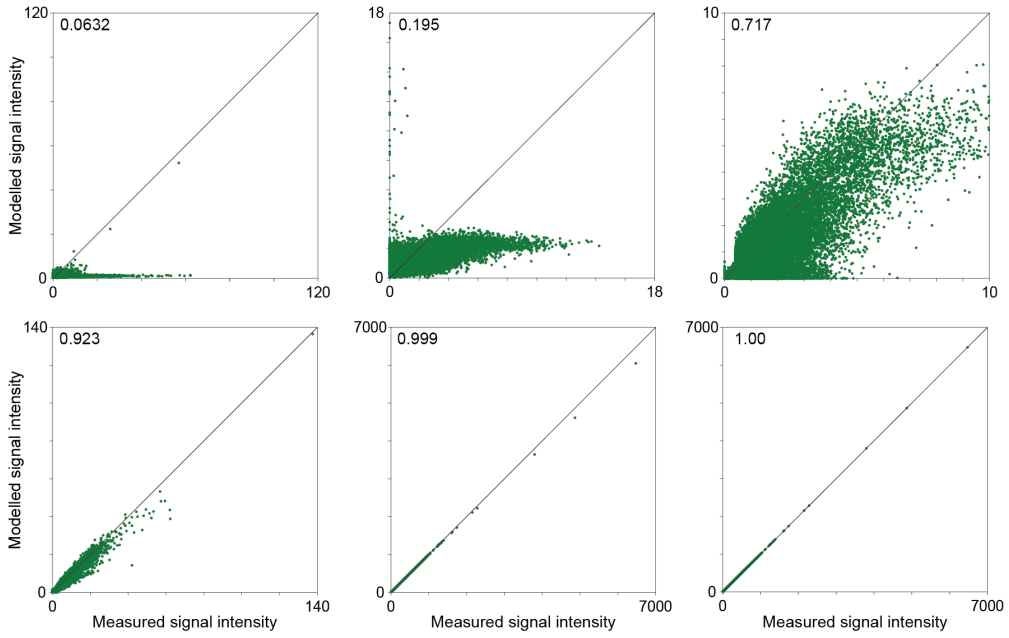


Figure 4.5: Example scatter plots of measured and modelled signal intensity for illustration purposes. For each plot the Meisch coefficient for the corresponding data is shown in the top left corner. The closer to one the Meisch coefficient is, the better the measured signal intensity for a given *m/z* is described by the calculated model. The line y = x is shown on each scatter plot. When there is significant deviation from the y = x line, as in the top left and top center scatter plots, additional components are needed to describe the data measured by the presented variable.

4.2.2.5 Identifying the resolved sources in the model

Once the number of sources needed to describe a dataset was determined, the resolved components needed to be identified as chemically meaningful sources. To identify sources the resolved time series (scores) and mass spectra (loadings) of the models were interrogated in conjunction with the field notes, geospatial maps of the resolved source concentration profiles, data from additional on-board sensors (*e.g.*, PM, CO_2, CH_4), and supplemental information (*e.g.*, the US EPA's Speciate database, previous source apportionment studies).

4.2.2.6 Apportionment of each source component in the model

The percentage of measured VOCs accounted for by each of the resolved sources was calculated using Equation 4.9 and Equation 4.10:

$$y = Gb \qquad \text{Equation 4.9}$$

$$\%\ source = \left({}^{1}\!/_{y^T}\right) \cdot G \cdot diag(b) \cdot \left(\frac{100}{\#\ samples}\right) \qquad \text{Equation 4.10}$$

where *y* is a vector (*# samples* by *1*) containing the total signal at each point in time, ***G*** is the time series source contributions (*# samples* by *# sources*), *b* is the regression vector (*# sources* by *1*) between the contributions and the total signal, *1* is a vector (*1* by *# samples*) of ones, *./* is element-wise division, and *diag(b)* is the diagonalization of the regression vector.[135]

4.2.2.7 Receptor model comparison

Algorithm comparisons (MCR-ALS, MCR-WALS, PMF) were done using the August 7, 2018 data. In order to run the PMF model in EPA PMF 5.0, the number of samples and variables in the dataset needed to be reduced. To accomplish this, 60 second sample averages were calculated, reducing the number of samples from 45537 to 748. Additionally, only recorded *m/z* with measurements having an average S/N > 1 were included. As the 60 second averaging reduced the noise level in the data, this resulted in a mass list very similar to that generated from the high frequency data including recorded *m/z* with measurements having a max S/N > 1.

PMF models with different numbers of components were run, and the PMF model used for comparison was determined by comparing the calculated Q/Qexp[221], and Meisch coefficients for the models. The PMF results were then compared to the MCR-ALS and MCR-WALS models with the same number of components. The final PMF, MCR-ALS, and MCR-WALS models were compared by calculating pairwise correlation coefficients (R^2) for both the resolved time series and mass spectra as follows:

$$R^2 = \frac{\vec{x}^T \vec{y}}{\|\vec{x}\|\|\vec{y}\|} \qquad \text{Equation 4.11}$$

where $\vec{x}$ is a time series (or mass spectrum) from the first model, and $\vec{y}$ is a time series (or mass spectrum) for a component from the second model.[135]

4.2.2.8 Mapping

Geospatial maps of the identified VOC component sources were produced using MATLAB's Mapping Toolbox. Data was collected at a frequency of 1 Hz resulting in a spatial resolution of 15–25 m with drive speeds between 50 and 100 km/hour. As the GPS data and mass spectral data were not collected at the exact same timestamps, a location for each mass spectrum was determined by using a linear interpolation between the two latitude and longitude measurements bracketing the sampling time of the mass spectrum, accounting for the lag time due to sample line transit time. For each component, a map was made with data points sized and coloured based on the signal intensity of the component, with small, white dots indicating low signal intensities and large, coloured dots indicating high signal intensities. The scale used to determine the sizes and colours of the dots differed for each source, with the size bins chosen to highlight features in the resolved time series.

Geospatial maps capturing the apportionment of the VOC sources were also produced using MATLAB's Mapping Toolbox and Adobe Illustrator. One source apportionment map was made for each field campaign, with the intensity of the colour of the dots for each source being determined by the proportion of the signal described by the source at a specific time and location. A gradient of seven colour intensities was made for each component using the following divisions: < 5 %, 5–10 %, 10–15 %, 15–20 %, 20–30 %, 30–40 %, and > 40%, with source contributions < 5 % having white dots, and those with > 40 % being the most intensely coloured. Due to the way the maps are generated in MATLAB, the calculations for the sizes of the dots for each field campaign differed slightly, but in general the dots were sized to be proportional to the log of the total ion current (TIC) (up to a maximum value of 8) multiplied by the proportion of the signal described by the measurement being plotted. As these maps had multiple dots plotted per sampling time (one per component) sizing and colouring the dots in this way allowed for more details of the contributing VOC sources to be visualized. Additionally, the data was layered on the maps so that the dots representing data points describing 30–40, or > 40 % of the VOCs in the sample were layered on top of the smaller contributions.

4.2.2.9 Particulate matter mass concentrations

The particle count time series data measured with the OPS was converted to μg/m^3 of PM for each size bin according to the OPS instruction manual. For the final mass concentration in μg/m^3 the particle density is

assumed to be 1 g/cm^3.[228] The $PM_{2.5}$ concentration for each measurement time was calculated by summing up the size bins below 2.5 µm.

4.3 Results and discussion

To evaluate several receptor modeling methodologies, PTR-ToF-MS data from August 7, 2018 was interrogated. During this field campaign samples were collected near industrial areas including a pulp mill and sawmill, a sewage treatment lagoon, urban and rural residential areas, forested areas, and commercial areas. Forest fires in the region[229] produced smoke that was observable at lower elevations for some of the day as evidenced by elevated PM levels recorded in the mobile lab. After 18:00, the vehicle was parked overnight on the pulp mill site property. VOC concentrations were generally low (< 2 ppb) with the exception of concentration excursions observed for compounds such as methanol, acetone, toluene, and ethylbenzene. The total ion chromatogram of the recorded *m/z* used in the final analysis (max S/N > 1) is shown in the top panel of Figure 4.6. Because methanol (*m/z* 33.03) excursions tended to dominate the observed MS signals, the ion chromatogram for the same data excluding methanol, is shown in the bottom panel. Gaps in the time series are data that has been redacted by request (gaps between 14:00 and 15:45) or when instrument errors occurred (gaps after 16:00).

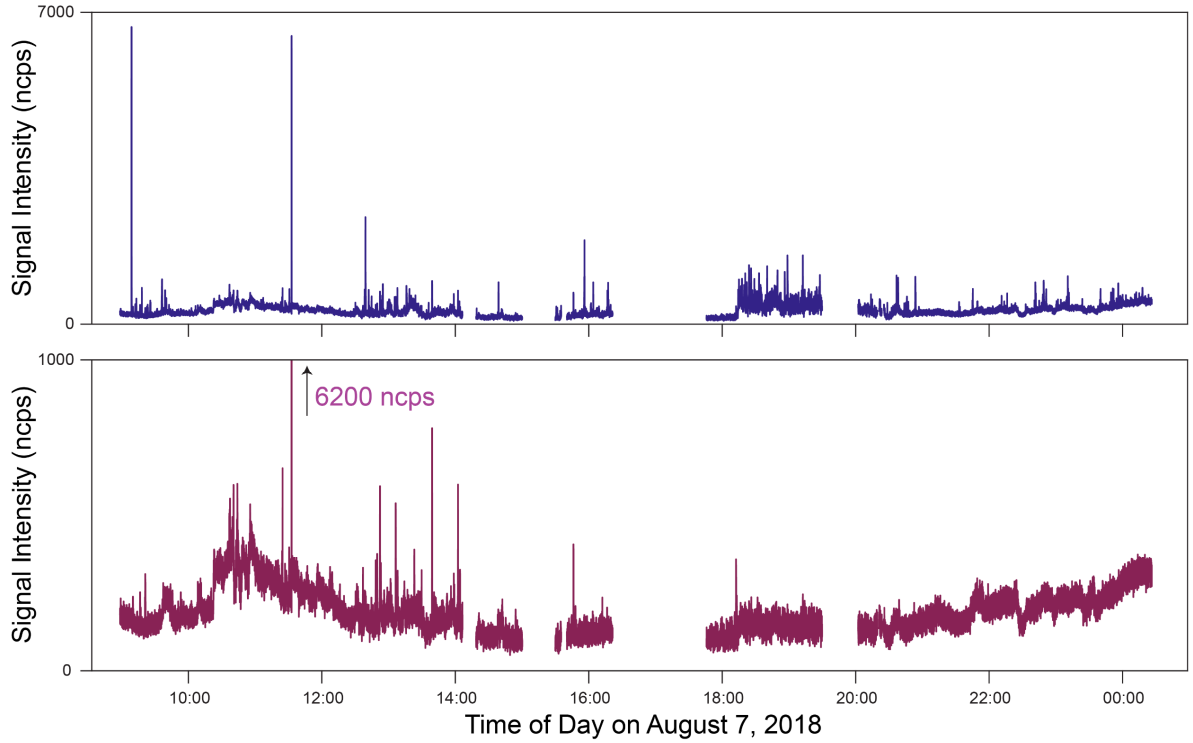


Figure 4.6: Total ion chromatogram from mobile PTR-ToF-MS for August 7, 2018 field campaign with (top) and without (bottom) the methanol signal.

The majority of VOCs encountered on August 7, 2018 were hydrocarbons and oxygenated VOC species. An overlay of the time series of ten representative *m/z* measured over the course of the day for these two compounds classes are shown in the top and bottom panels of Figure 4.7. The observed differences in the temporal profiles suggest that the sources of these compounds classes are different. As can be seen in Figure 4.7, a number of short concentration excursions of hydrocarbon VOCs were measured. Most of these were associated with vehicular traffic. The large signals recorded at ~11:30 corresponded to a vehicle with observable emissions passing the mobile lab. In contrast, the oxygenated VOCs were measured as both short concentration excursions such as acetaldehyde (*m/z* 45.03) at ~11:30 and acetone (*m/z* 59.05) observed at ~13:00 as well as broader concentrations excursions when the mobile lab was parked on the pulp mill property overnight.

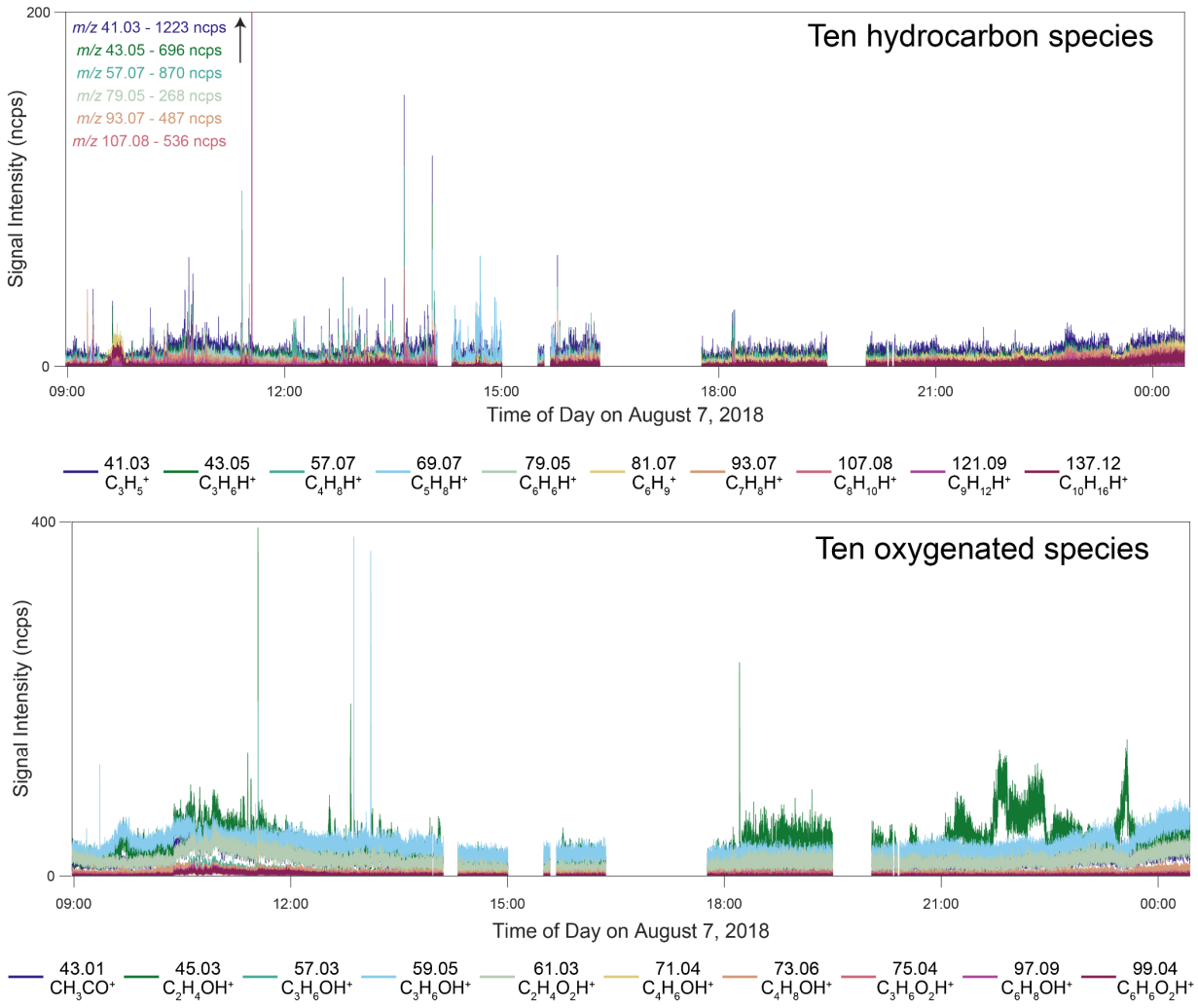


Figure 4.7: Time series for VOCs measured on August 7, 2018 in Port Alberni. Top: Time series overlay for 10 hydrocarbon VOCs measured (the y-axis in this panel has been truncated to better show the temporal profiles of the low intensity signals). Bottom: Time series overlay of 10 oxygenated VOCs species.

Example mass spectra collected on August 7, 2018 are shown in Figure 4.8. The mass spectrum on the top left was collected at an intersection on the main highway through Port Alberni. Many of the compounds in this spectrum are due to the presence of hydrocarbon species (*e.g.*, BTEX, butene), but measurements due to the presence of oxygenated compounds (*e.g.*, methanol and acetone/propanal) are also observed. The mass spectrum in the top right was collected near a commercial area in Port Alberni where multiple small businesses (including an auto body shop and boat repair business) are located. The main ions measured in this location are acetone and methanol, with other lower intensity signals due to the presence of other oxygenated species (e.g, acetic acid) also observed. The mass spectrum on the lower left was collected while parked in the old growth forest that the highway to Port Alberni passes through. This mass spectrum includes measurements for ions due to the presence of monoterpenes (*m/z* 137.12 and 81.07), as well as small oxygenated species (*e.g.* methanol, acetaldehyde, acetone/propanal, acetic acid). As the parking lot for the old growth forest is located on a highway into Port Alberni it is also possible that this mass spectrum contains signals due to traffic. The mass spectrum on the bottom right represents the VOCs associated with forest fire smoke in the area. Forest fire smoke was ubiquitous during the morning sampling in Port Alberni, and this mass spectrum was averaged from data that did not appear to contain signals from other VOC sources (*e.g.*, traffic, industry). This mass spectrum contains signals for many oxygenated compound (*e.g.*, methanol, acetaldehyde, acetic acid, acetone), and also appears to contain measurements associated with acetonitrile (*m/z* 42.03), a marker of forest fire smoke.[230] Measurements of *m/z* 42.03 can also be due to the carbon-13 isotope of $C_3H_5^+$ (*m/z* 42.04) as the mass resolution of the PTR-ToF-MS utilized here is not sufficient to distinguish the two ions, however in the case of this mass spectrum the signal intensity of the peak observed at *m/z* 42.03 is almost equal to the signal measured at *m/z* 41.03 ($C_3H_5^+$), suggesting that it is not the isotopic peak of $C_3H_5^+$.

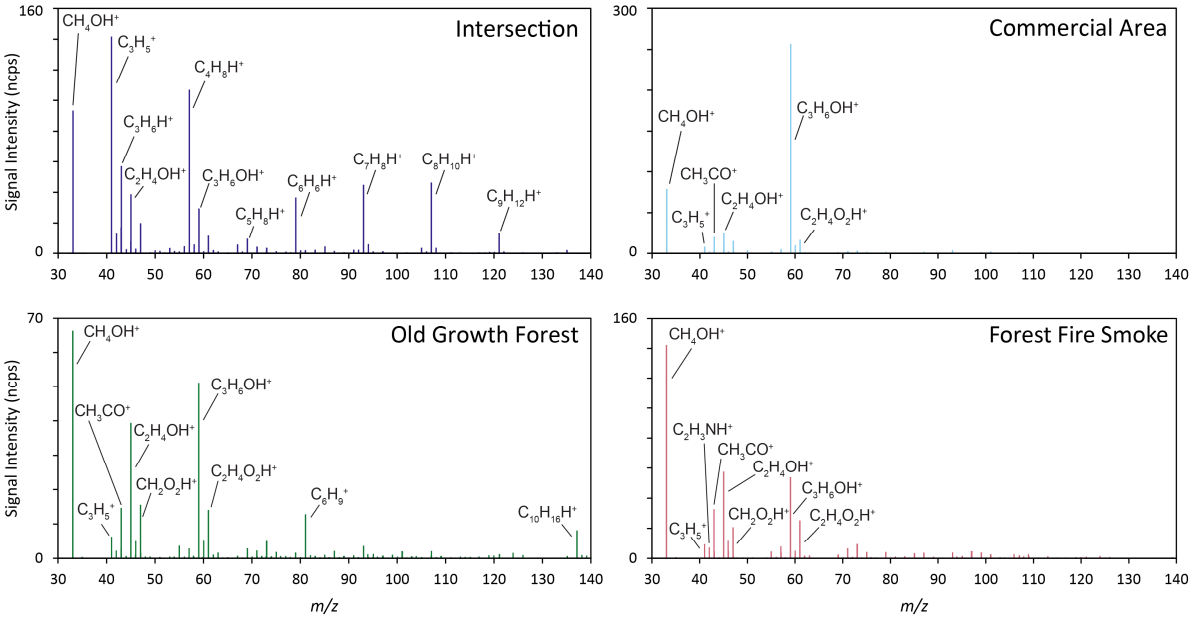


Figure 4.8: Example mass spectra collected on August 7, 2018 in the Port Alberni area. Top Left: Mass spectrum measured at an intersection. Top Right: Mass spectrum measured near a commercial area with multiple small business. Bottom Left: Mass spectrum measured while parked in an old growth forest. Bottom Right: Mass spectrum capturing the VOCs present in forest fire smoke.

Sampling on August 14, 2018 took place on eastern Vancouver Island, in the Nanaimo Regional District from Parksville (North of Nanaimo) to Cedar (South of Nanaimo). The study area includes residential, commercial, industrial (*e.g.*, compost facility, pulp mill), and rural/light agricultural uses. The surrounding area is dominated by mixed coniferous forest, and marine environments. Of note, during sampling the region was impacted by smoke from local forest fires, and a reduced sulphur plume was observed at the south end of the drive route.

The ion chromatogram of the *m/z* used in the final analysis (max S/N > 1) is shown in the top panel of Figure 4.9. Once again, methanol was measured at high signal intensity throughout the day so the ion chromatogram, excluding the signal intensity of *m/z* 33.03, is shown in the bottom panel. High VOC concentrations were recorded at multiple locations throughout the day (an auto wrecking facility at ~8:50, a regional composting facility at ~11:15, a sawmill at ~11:35, and a waste transfer site at ~15:00). Figure 4.10 show overlays of the time series for a selection of recorded *m/z* associated with hydrocarbon (top), oxygenated compound (middle), and sulphur compound (bottom) measurements. Hydrocarbons were the main species detected near the car wrecking facility at 8:50, oxygenated species are present near both the composting facility and sawmill sampled between 11:00–12:00, and sulphur compounds were mainly measured between 9:30 and 11:00, with additional measurements attributed to methanethiol measured near a pollution control center for waste water treatment at ~14:30.

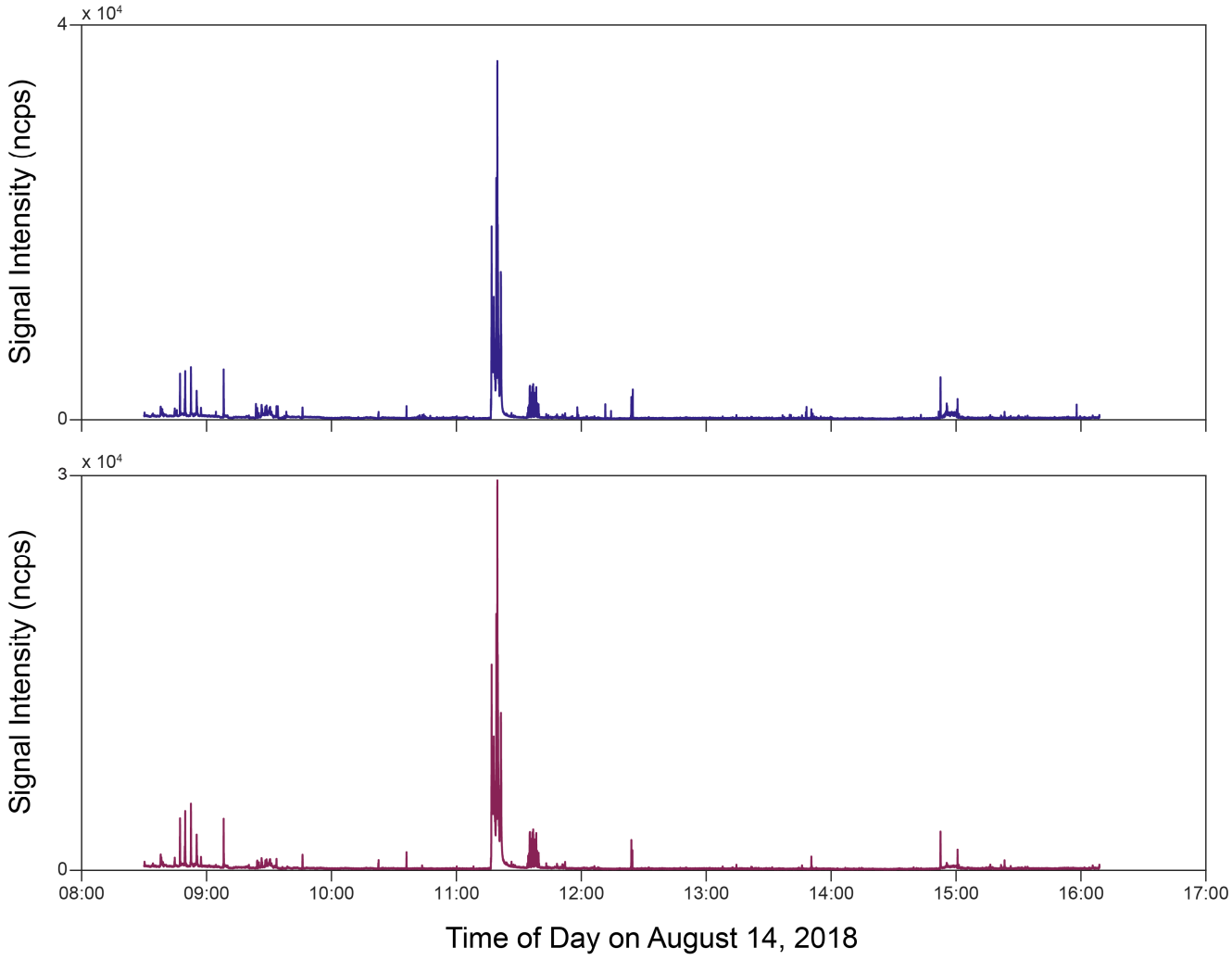


Figure 4.9: Total ion chromatogram of PTR-ToF-MS data collected on the August 14, 2018 field campaign between Nanaimo and Parksville, BC with (top) and without (bottom) the signal intensity due to methanol (*m/z* 33.03).

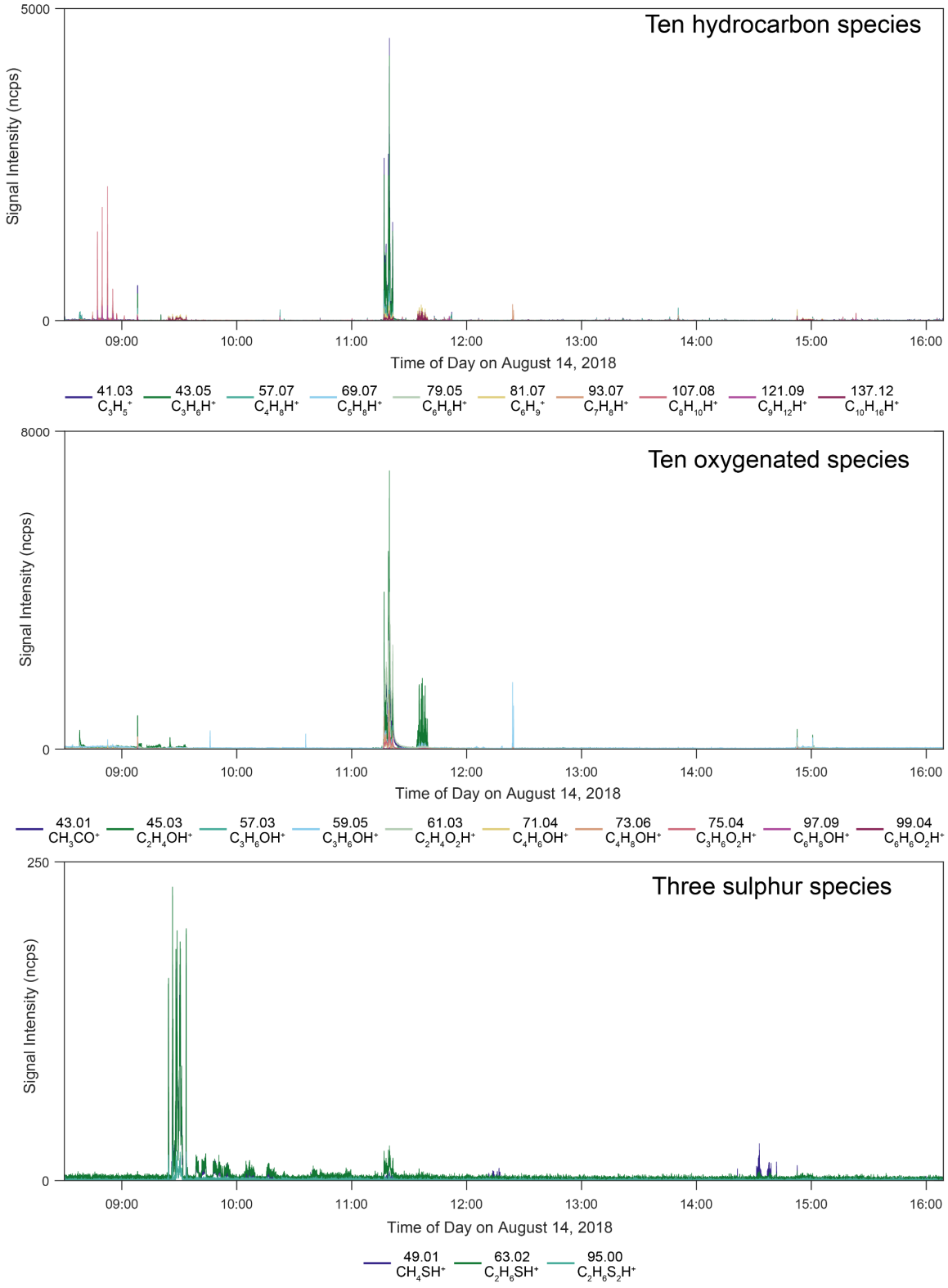


Figure 4.10: Top: Time series overlay for 10 representative hydrocarbon species measured on August 14, 2018. Middle: Time series overlay of 10 representative oxygenated species measured on August 14.2018. Bottom: Time series overlay of 3 sulphur species measured on August 14, 2018.

Example mass spectra measured on August 14, 2018 are shown in Figure 4.11. The mass spectrum shown in the top left was measured near an auto wrecking facility with recorded *m/z* measured for a variety of hydrocarbons (*e.g.*, BTEX, butene, propene) and oxygenated compounds (*e.g.*, acetaldehyde, acetone/propanal, methanol). The mass spectrum in the top right was collected near a sawmill and contains signals recorded for *m/z* attributed to monoterpenes and oxygenated compounds (*e.g.*, acetone, acetic acid, acetaldehyde, methanol). The mass spectrum on the bottom left was collected near the regional composting facility and contains signals for recorded *m/z* for a variety of oxygenated compounds (*e.g.*, acetic acid, acetaldehyde, methanol, propanoic acid, butanoic acid). The mass spectrum on the bottom right was collected near the pulp mill and contains signals associated with sulphur compounds (*e.g.*, methanethiol, ethanethiol/DMS), monoterpenes, and oxygenated compounds (*e.g.*, acetone, acetaldehyde, methanol).

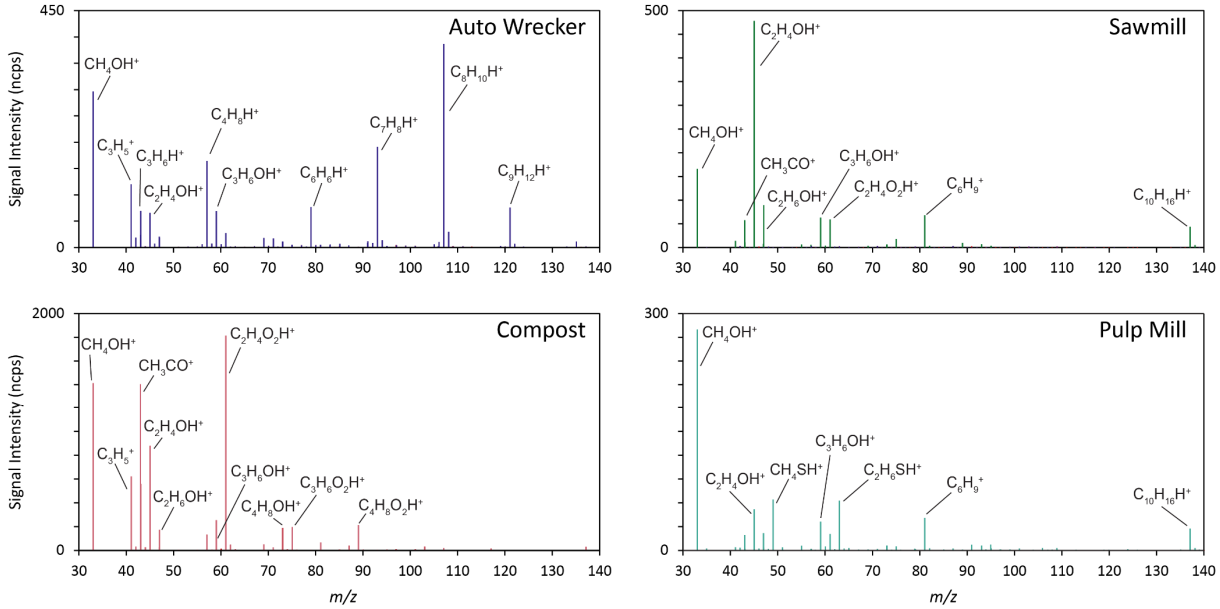


Figure 4.11: Example mass spectra collected on August 14, 2018 in the Nanaimo area. Top Left: Mass spectrum measured near an auto wrecking facility. Top Right: Mass spectrum measured near a sawmill. Bottom Left: Mass spectrum measured near the regional composting facility. Bottom Right: Mass spectrum measured near the pulp mill.

4.3.1 Observed mass-to-charge ratios

A list of the major VOC signals measured during the field campaigns are summarized in Table 4.1. The list includes 35 observed *m/z* along with possible compound identities (chemical formulas), exact masses, and the days they were observed. The PTR-ToF-MS cannot distinguish between compounds with the same chemical formula, so isomers (*e.g.,* ethylbenzene, *o*-, *m*-, and *p*-xylene, or the individual monoterpenes) are not distinguished. Some ions were measured on all sampling days summarized here, such as *m/z* 33.03 (CH_4OH^+, methanol), 61.03 ($C_2H_4O_2H^+$, acetic acid), 45.03 ($C_2H_4OH^+$, acetaldehyde), 69.07 ($C_5H_8H^+$, isoprene), and 79.05 ($C_6H_6H^+$, benzene), while others, such as *m/z* 49.01 (CH_4SH^+, methanethiol), were observed occasionally.

Chemical formulas ending in H^+ indicate the signal is associated with a protonated molecule ($[M+H]^+$), while formulas that do not end in H^+ indicate fragment ions (some signal can be due to both). In many cases there is a long list of isomers that can be measured at a particular *m/z* and the list presented here is incomplete, with more complete lists being compiled in the literature.[194,231] The chemical compounds listed here have been reported as emissions from many of the sources sampled, and compound identity could be confirmed using whole air sampling methods with subsequent GC-MS analysis. The variation in VOCs measured, and their spatial and temporal distributions, will facilitate the use of receptor models to analyze the mass spectral data to determine VOC sources impacting the sampling areas.

Table 4.1: Summary of major *m/z* measured across the field sampling days along with their potential chemical formula, identity, and exact mass.

m/z	**Possible chemical formula (possible compound)**[194]	**Exact mass**	**Day(s) observed**[a]
33.03	CH_4OH^+ (methanol)	33.03349	August 7, 2018 (Port Alberni) August 14, 2018 (Parksville–Nanaimo) August 21 and 22, 2017 (Nanaimo–Crofton)
41.03	$C_3H_4H^+$ (propyne, 1,2-propadiene) $C_3H_5^+$ (alkyl fragment from hydrocarbons, alcohols, aldehydes, esters)	41.03858	August 7, 2018 August 14, 2018 August 21 and 22, 2017
42.03[b]	$C_2H_3NH^+$ (acetonitrile)	42.03383	August 7, 2018 August 14, 2018
43.01	CH_3CO^+ (acetic acid fragment, propanol fragment, methyl vinyl ketone fragment)	43.01784	August 7, 2018 August 14, 2018 August 21 and 22, 2017
43.05	$C_3H_6H^+$ (propene, cyclopropane)	43.05423	August 7, 2018 August 14, 2018 August 21 and 22, 2017
45.03	$C_2H_4OH^+$ (acetaldehyde)	45.03349	August 7, 2018 August 14, 2018 August 21 and 22, 2017
47.01	$CH_2O_2H^+$ (formic acid)	47.01276	August 21 and 22, 2017
47.04	$C_2H_6OH^+$ (ethanol)	47.04914	August 14, 2018
49.01	CH_4SH^+ (methanethiol)	49.01065	August 14, 2018
57.03	$C_3H_4OH^+$ (acrolein) $C_3H_5O^+$ (propanoic acid fragment, R propanoate fragment)	57.03349	August 7, 2018

Table 4.1: Summary of major *m/z* measured across the field sampling days along with their potential chemical formula, identity, and exact mass, continued.

m/z	Possible chemical formula (possible compound)[194]	Exact mass	Day(s) observed[a]
57.07	$C_4H_8H^+$ (butenes)	57.06988	August 7, 2018 August 14, 2018 August 21 and 22, 2017
59.05	$C_3H_6OH^+$ (acetone, propanal)	59.04914	August 7, 2018 August 14, 2018 August 21 and 22, 2017
61.03	$C_2H_4O_2H^+$ (acetic acid, methyl formate, hydroxyacetaldehyde)	61.02841	August 7, 2018 August 14, 2018 August 21 and 22, 2017
63.02	$C_2H_6SH^+$ (dimethyl sulphide/ethanethiol)	63.02630	August 14, 2018 August 21 and 22, 2017
69.07	$C_5H_8H^+$ (isoprene, cyclopentene, pentadiene)	69.06988	August 7, 2018 August 14, 2018 August 21 and 22, 2017
71.04	$C_4H_6OH^+$ (2-butenal, methacrolein, methyl vinyl ketone)	71.04914	August 7, 2018
71.08	$C_5H_{10}H^+$ (pentene, methyl butene, cyclopentane) $C_5H_{11}^+$ (alcohol fragment, butanoic acid fragment)	71.08553	August 14, 2018 August 21 and 22, 2017
73.06	$C_4H_8OH^+$ (butanone, butanal, 2-methyl propanal)	73.06479	August 7, 2018 August 14, 2018 August 21 and 22, 2017
75.04	$C_3H_6O_2H^+$ (methyl acetate, ethyl formate, propanoic acid)	75.04406	August 7, 2018 August 14, 2018 August 21 and 22, 2017
79.05	$C_6H_6H^+$ (benzene)	79.05423	August 7, 2018 August 14, 2018 August 21 and 22, 2017
81.07	$C_6H_9^+$ (monoterpene fragment)	81.06988	August 7, 2018 August 14, 2018 August 21 and 22, 2017
85.06	$C_5H_8OH^+$ (cyclopentanone, methylbutenone)	85.06479	August 7, 2018
87.04	$C_4H_6O_2H^+$ (2,3-butanedione, methyl acrylate)	87.04406	August 7, 2018 August 14, 2018
89.06	$C_4H_8O_2H^+$ (methyl propanoate, ethyl acetate, acetoin, butanoic acid)	89.05971	August 14, 2018
91.04	$C_3H_6O_3H^+$ (lactic acid)	91.03897	August 21 and 22, 2017

Table 4.1: Summary of major *m/z* measured across the field sampling days along with their potential chemical formula, identity, and exact mass, continued.

m/z	Possible chemical formula (possible compound)[194]	Exact mass	Day(s) observed[a]
93.07	$C_7H_8H^+$ (toluene)	93.06988	August 7, 2018 August 14, 2018 August 21 and 22, 2017
95.00	$C_2H_6S_2H^+$ (DMDS)	94.99837	August 14, 2018 August 21 and 22, 2017
95.08	$C_7H_{11}^+$ (monoterpene fragment)	95.08608	August 14, 2018 August 21 and 22, 2017
99.04	$C_5H_6O_2H^+$ (furfural alcohol)	99.04406	August 7, 2018
103.06	$C_8H_6H^+$ (phenylacetylene)	103.05423	August 14, 2018
107.08	$C_8H_{10}H^+$ (ethylbenzene, *o*-, *m*-, *p*-xylene)	107.08553	August 7, 2018 August 14, 2018 August 21 and 22, 2017
117.08	$C_9H_8H^+$ (indene)	117.06988	August 14, 2018
121.09	$C_9H_{12}H^+$ (C9-aromatics)	121.10118	August 7, 2018 August 14, 2018 August 21 and 22, 2017
135.10	$C_{10}H_{14}H^+$ (C10-aromatics)	135.11683	August 14, 2018 August 21 and 22, 2017
137.12	$C_{10}H_{16}H^+$ (monoterpenes)	137.13248	August 7, 2018 August 14, 2018 August 21 and 22, 2017

[a] August 7, 2018 (09:00–24:00) in Alberni Valley; August 14, 2018 (08:30–16:15) in Nanaimo area; August 21 (16:30–18:00) and 22, 2017 (13:45–16:45) Nanaimo–Crofton.

[b] Signals measured at *m/z* 42.03 can be both the MH^+ ion for acetonitrile (exact mass 42.03383) and the carbon-13 isotope of $C_3H_5^+$ (42.04198). The mass resolution of the PTR-ToF-MS used here is insufficient to distinguish the two (requires mass resolution at *m/z* 59 of 7000 to distinguish ions with $\Delta m/z = 0.008$).

4.3.2 Quantitative analysis

The concentrations of VOCs encountered during the field campaign were generally low, with the exception of some short duration concentration excursions, and elevated concentrations near some stationary point sources (*e.g.*, composting and auto wrecker) and specific on-road vehicles in high traffic areas. Boxplots of the concentration of $C_6H_6H^+$ (*m/z* 79.05, quantified as benzene), $C_7H_8H^+$ (*m/z* 93.07, quantified as toluene), $C_8H_{10}H^+$ (*m/z* 107.08, quantified as ethylbenzene), CH_4SH^+ (*m/z* 49.01, quantified as methanethiol), $C_2H_6S_2H^+$ (*m/z* 63.02, quantified as DMS), $C_{10}H_{16}H^+$ (*m/z* 137.12, quantified as α-pinene), CH_4OH^+ (*m/z* 33.03, methanol), $C_2H_4OH^+$ (*m/z* 45.03, acetaldehyde), and $C_3H_6OH^+$ (*m/z* 59.05, acetone or propanal) for August 7 and 14, 2018 are shown in Figure 4.12. The boxplots indicate that the vast majority of observations (> 95%) were below our measured detection limits (shown with the red dotted lines) for the hydrocarbons and sulphur compounds, with

the oxygenated compounds having more observations above the detection limits. Similar boxplots for August 21–22, 2017 are provided in the supplemental information for Chapter 3. Concentrations across the sampling days reached maximums of 25, 55, 70, 55, 11, 150, 730, 750, and 180 ppb_v for benzene, toluene, ethylbenzene, α-pinene, methanethiol, DMS, methanol, acetaldehyde, and acetone, respectively. Some of the observed maximum concentrations (*e.g.,* benzene, acetaldehyde) are above the provincial ambient air guidelines for some of the averaging periods presented in Table 1.1. However, as we were sampling while moving further measurements would be required to determine if the elevated concentrations persist or are of concern. The hydrocarbons and sulphur compounds were quantified via external calibration with gas standards, while the oxygenated compounds had their concentrations estimated using the calculation in the PTR-MS Viewer Software.

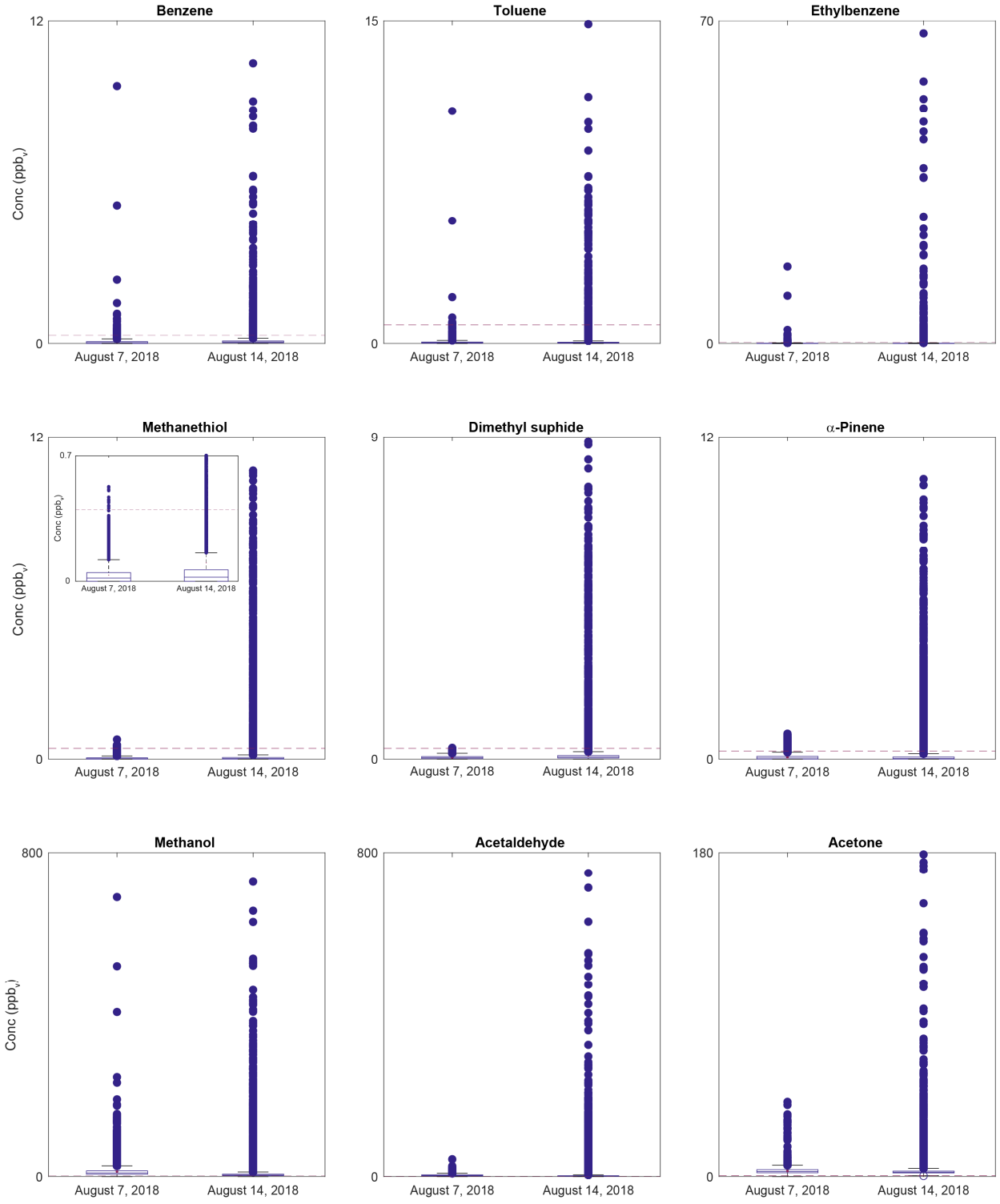


Figure 4.12: Boxplots of the concentration of some analytes measured on August 7, 2018 (9:00–24:00) in the Alberni Valley and August 14, 2018 (8:30–16:15) in the Nanaimo area. Detection limits for each compound are shown as horizontal doted lines (red). In the boxplots, the box encapsulates the interquartile range (IQR, 25th–75th percentile) in the data, the upper whisker is at the Upper Adjacent Value ($3rd\ quartile + 1.5 \times IQR$), and the lower whisker is at the Lower Adjacent Value ($1st\ quartile - 1.5 \times IQR$). External calibration were used to determine the concentrations of benzene, toluene, ethylbenzene, methanethiol, DMS, and α-pinene. Concentrations for methanol, acetaldehyde, and acetone were determine using the concentration calculation in the PTR-MS Viewer software.

4.3.3 Comparison between MCR-ALS, MCR-WALS, and PMF

The data collected on August 7, 2018 was used to compare the results of PMF, MCR-WALS, and MCR-WALS. PMF models using the EPA PMF 5.0 software were calculated for models with three to ten components using 60 second sample averages. The 60 second averaging reduced the size of the dataset, allowing analysis with EPA PMF 5.0, but at the expense of dampening the signal intensities of concentration excursions lasting only a few seconds. Additionally, the mass list was reduced to the 78 recorded *m/z* that had measured signal intensities with an average S/N > 1 over the course of the day. The estimation of measurement uncertainty shown in Equation 4.6 was modified to for account for the 60 second sample averaging.

To determine the number of components in the data, the Q/Qexp values for three to ten component models were examined, as shown in Figure 4.13. Q is the weighted sum of squares of the scaled residuals, calculated as shown in Equation 4.2, and Qexp ($Q_{expected}$) is calculated as shown in Equation 4.12:

$$Qexp = (\#\ of\ samples \times \#\ of\ strong\ variables) - (\#\ of\ elements\ in\ G + \#\ of\ elements\ in\ F) \qquad \text{Equation 4.12}$$

where ***G*** is the matrix containing the time series of the source contributions; the number of elements in it is equal to the number of samples multiplied by the number of resolved components. ***F*** is the matrix containing the source profiles of each component; the number of elements in ***F*** is equal to the number of strong variables (*m/z* with average S/N > 1) multiplied by the number of resolved components. Q/Qexp is expected to be close to one if the data follow the model described in Equation **1.2**.[221] As shown in Figure 4.13, the value of Q/Qexp decreases fairly significantly for each added component for the three–seven component PMF solutions, with diminishing returns when more than seven components are present in the model (Q/Qexp of 1.13, 0.99, 0.89, 0.76 for the seven, eight, nine, and ten component models, respectively).

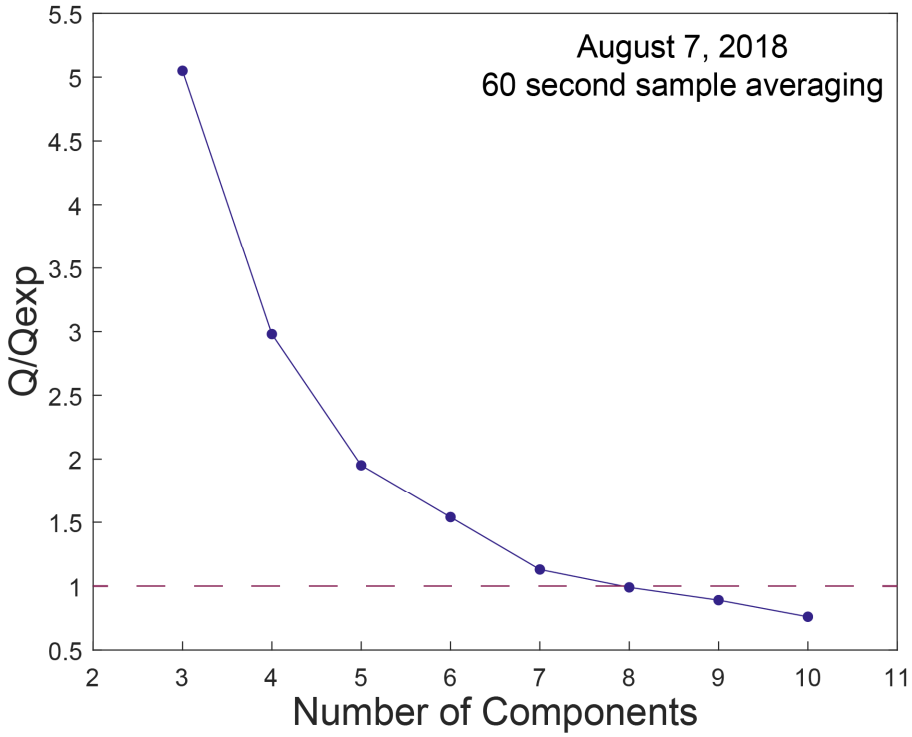


Figure 4.13: Plot of Q/Qexp for 3-10 PMF components calculated from using the reduced August 7, 2018 dataset collected in Port Alberni. The dataset contained 60 second sample averaging and the recorded *m/z* with an average S/N > 1 over the course of the sampling day.

Based on the Q/Qexp values, the six, seven, and eight component PMF models were chosen for interrogation, and six, seven, and eight component MCR-ALS and MCR-WALS models were calculated using the dataset for comparison. The Meisch coefficients for 24 of the 78 measured *m/z* used in the analysis are provided in Table 4.2 for the six, seven, and eight component PMF, MCR-ALS, and MCR-WALS models (values greater than or equal to 0.90 are bolded). For the PMF models, adding a 7th component improves the modeling of *m/z* 43.05, 57.07, 81.07, 107.08, 121.09, and 137.12, while the addition of an 8th component does not dramatically improve the modeling of any of the *m/z,* and worsens the modeling of *m/z* 81.07 and 137.12. As a result of this analysis, a seven component model was used to compare PMF, MCR-WALS, and MCR-ALS. Scatter plots of the measured and modelled signal intensity for *m/z* 45.03, 59.05, 69.07, 79.05, and 137.12 for each of the seven component models, labelled with the corresponding Meisch coefficient, are shown in Figure 4.14. This figure highlights that the MCR-WALS and PMF results are more similar to one another than to the MCR-ALS results, as the MCR-ALS scatter plots differ from the MCR-WALS and PMF plots.

Table 4.2: Meisch coefficients for the MCR-ALS, MCR-WALS, and PMF analysis of the 60 second averaged August 7, 2018 data. The column indicating the model that is discussed has been highlighted. Values greater than 0.9 have been bolded for emphasis.

	Number of MCR-ALS components			Number of MCR-WALS components			Number of PMF components		
m/z	**6**	**7**	**8**	**6**	**7**	**8**	**6**	**7**	**8**
33.03	**1.00**	**1.00**	**1.00**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
41.03	**0.975**	**0.983**	**0.990**	0.836	**0.977**	**0.978**	0.710	**0.976**	**0.976**
42.03	0.850	**0.900**	**0.907**	**0.930**	**0.928**	**0.931**	**0.927**	**0.931**	**0.938**
43.01	0.990	**0.991**	**0.991**	**0.967**	**0.969**	**0.981**	**0.963**	**0.969**	**0.986**
43.05	0.708	0.787	**0.918**	0.570	0.869	0.885	0.887	0.875	**0.909**
45.03	**0.999**	**0.999**	**0.999**	**0.994**	**0.994**	**0.997**	**0.994**	**0.995**	**0.996**
57.03	**0.932**	**0.964**	**0.964**	**0.975**	**0985**	**0.985**	**0.984**	**0.986**	**0.986**
57.07	0.660	0.826	**0.955**	0.567	**0.912**	**0.920**	**0.904**	**0.910**	**0.922**
59.05	**0.994**	**0.999**	**0.999**	**0.931**	**0.938**	**0.993**	**0.901**	**0.913**	**0.970**
61.03	**0.957**	**0.978**	**0.993**	**0.903**	0.838	**0.924**	**0.923**	**0.882**	**0.941**
69.07	0.759	0.865	**0.989**	**0.951**	**0.973**	**0.983**	-0.040	**0.972**	**0.981**
71.04	0.881	0.883	**0.888**	**0.907**	0.899	**0.913**	**0.902**	**0.902**	**0.924**
73.06	**0.920**	**0.921**	**0.921**	**0.922**	**0.913**	**0.945**	**0.900**	**0.910**	**0.936**
75.04	**0.948**	**0.950**	**0.952**	**0.931**	**0.921**	**0.945**	**0.944**	**0.932**	**0.946**
79.05	**0.908**	**0.919**	**0.955**	**0.911**	**0.966**	**0.967**	**0.968**	**0.967**	**0.971**
81.07	0.669	0.685	0.870	0.708	**0.971**	**0.953**	**0.929**	**0.953**	0.827
85.06	**0.942**	**0.961**	**0.962**	**0.969**	**0.976**	**0.977**	**0.977**	**0.977**	**0.979**
87.04	**0.935**	**0.934**	**0.939**	**0.937**	**0.941**	**0.943**	**0.938**	**0.940**	**0.946**
93.07	0.776	**0.940**	**0.940**	**0.898**	**0.955**	**0.956**	**0.950**	**0.957**	**0.963**
97.09	0.895	**0.940**	**0.939**	**0.945**	**0.961**	**0.965**	**0.959**	**0.965**	**0.969**
99.04	**0.925**	**0.957**	**0.962**	**0.960**	**0.957**	**0.952**	**0.958**	**0.956**	**0.953**
107.08	0.660	**0.915**	**0.937**	0.724	**0.936**	**0.942**	**0.917**	**0.938**	**0.953**
121.09	0.673	**0.905**	**0.934**	0.704	**0.929**	**0.947**	**0.916**	**0.933**	**0.948**
137.12	0.667	0.638	0.838	0.621	**0.953**	**0.926**	**0.900**	**0.928**	0.780

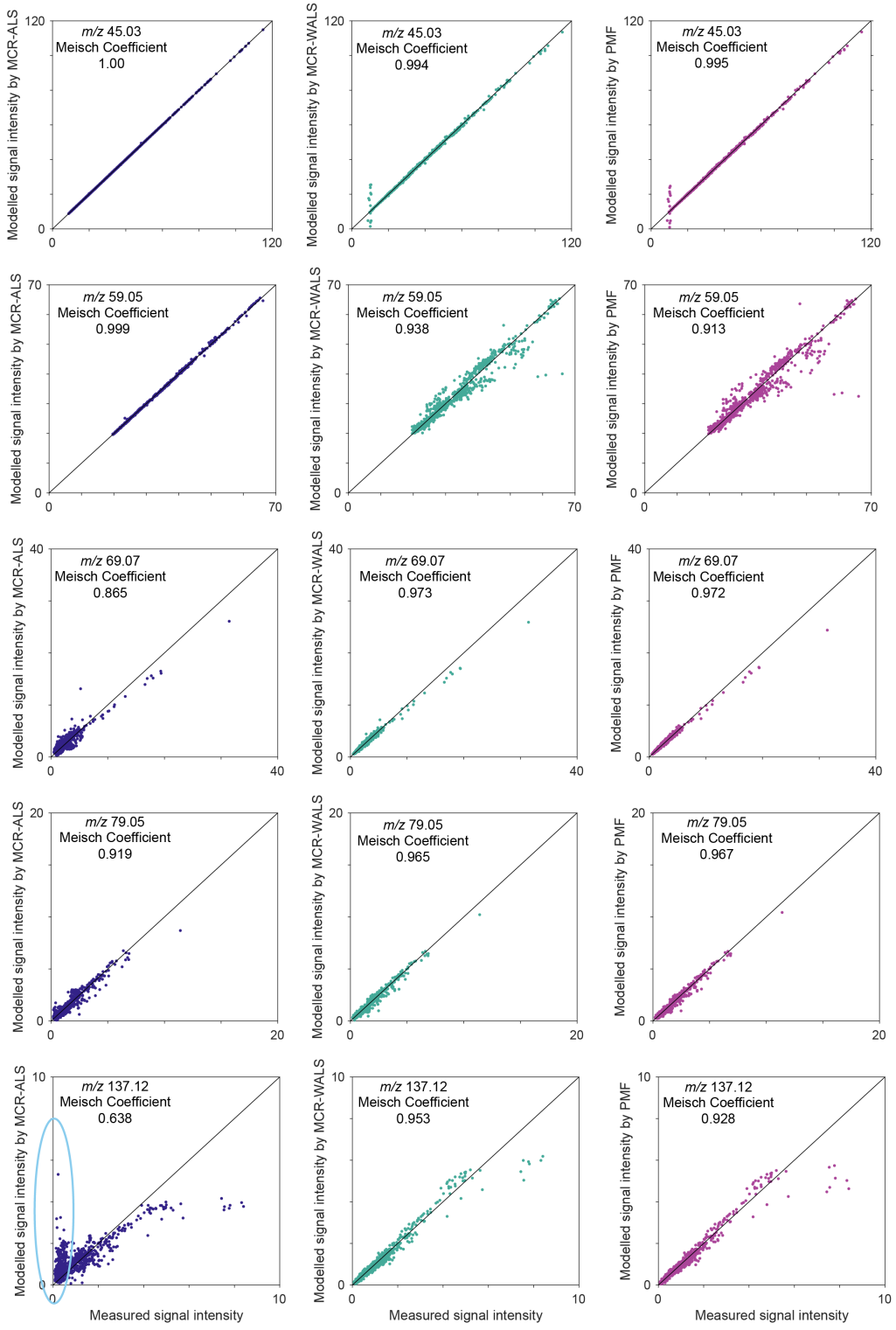


Figure 4.14: Scatter plots of the modelled and measured signal intensities for (top to bottom) *m/z* 45.03, 59.05, 69.07, 79.05, and 137.12 for the seven component models for the (left to right) MCR-ALS, MCR-WALS, and PMF models. On each scatter plot, the Meisch coefficient for comparison between the measured and modelled signal intensity for the presented *m/z* is shown. On the scatter plot for the MCR-ALS model of *m/z* 137.12, samples that have modelled signal intensities significantly higher than the observed signal intensities have been circled.

Overlays of the resolved time series and mass spectra for the seven component model for each algorithm are shown in Figure 4.15 and Figure 4.16, respectively. To correct for differences in the scale of the time series and mass spectra between PMF and MCR-(W)ALS, the data was unit vector normalized. The mass spectra are shown as overlaid bar charts with coloured shapes indicating the signal intensity for each of the three models to aid in visualizing the differences.

Pairwise correlation coefficients between the resolved time series and mass spectra for MCR-ALS and PMF, MCR-WALS and PMF, and MCR-ALS and MCR-WALS are shown in Table 4.3. As expected, the correlation coefficients indicate that the MCR-WALS results correlate more strongly with the PMF results than the MCR-ALS results do. For the seven component models, 13 of 14 correlation coefficients are greater than 0.90 when comparing the MCR-WALS and PMF models, while 8 of 14 are greater than 0.90 when comparing PMF and MCR-ALS. There are minor variations between the MCR-WALS and PMF time series and mass spectra, particularly for Component 4, while the MCR-ALS model differs more significantly from PMF, especially for the resolved time series for Components 1 and 4. When comparing the three models there are also differences in how each model reproduces the observed signal intensities of the recorded *m/z*, as shown in Figure 4.14. This is especially evident in the scatter plot for *m/z* 59.05, where all the dots fall along the diagonal for the MCR-ALS model and have more scatter with the MCR-WALS and PMF models, while in the scatter plot for *m/z* 137.12 the opposite is true. The signal intensity measured at *m/z* 137.12 is predominantly modelled in Component 1, which is where the largest differences between MCR-ALS and the other two models are observed. For Component 1, the MCR-ALS model resolved higher signal intensities for samples 50–100 and 150–400 than PMF and MCR-WALS (Figure 4.15). However, the samples where these differences occur correspond to times when no observed signal intensity was measured at *m/z* 137.12. An overlay of the time series of the measured signal intensity at *m/z* 137.12 (yellow) and the modelled signal intensity at *m/z* 137.12 by MCR-ALS (dark blue), MCR-WALS (light blue), and PMF (pink), is shown in Figure 4.17. In this plot it is evident that MCR-ALS has modelled signal at *m/z* 137.12 when none was observed. These samples have also been circled in the scatter plot in Figure 4.14.

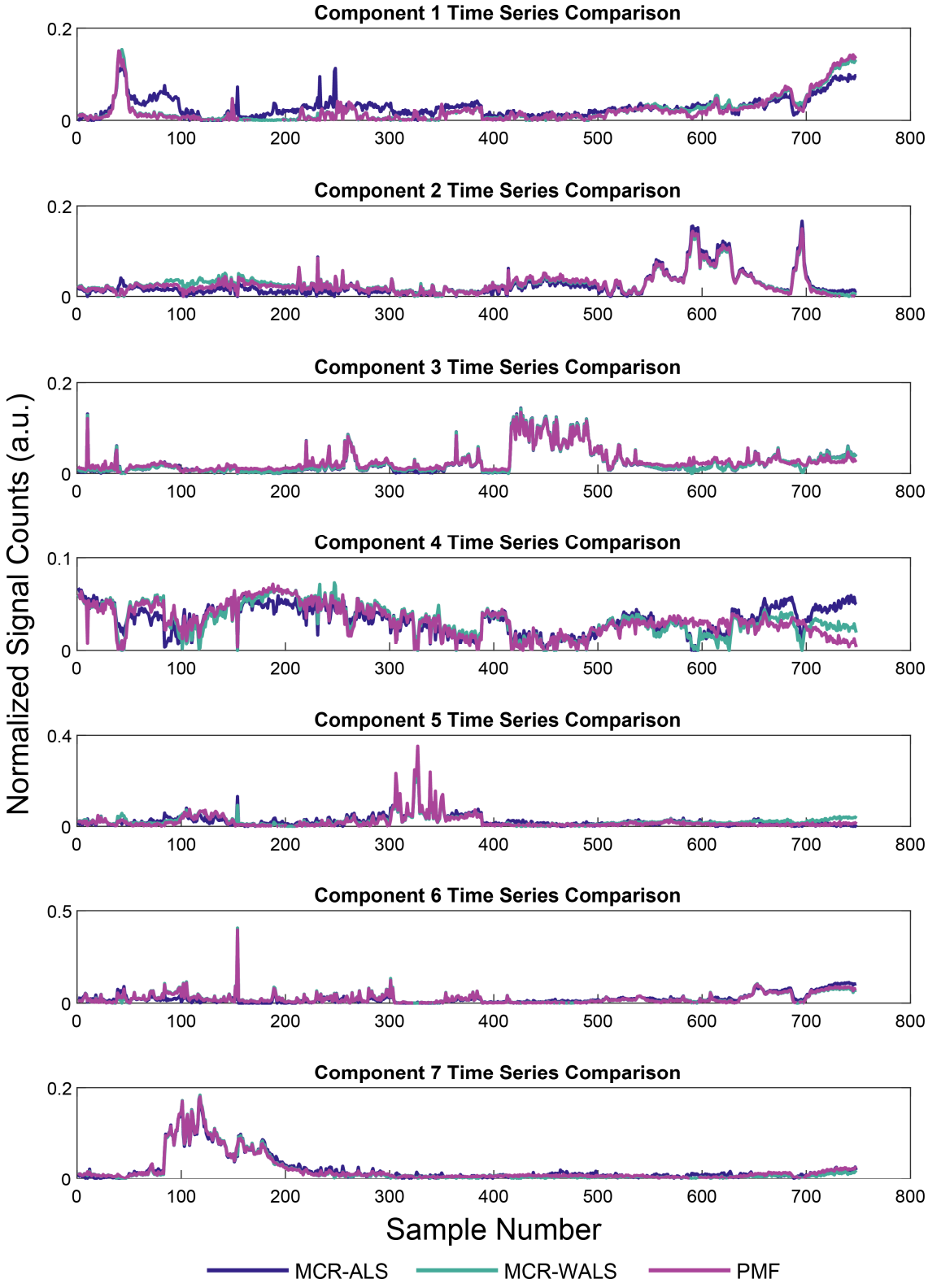


Figure 4.15: Comparison of resolved time series for seven component MCR-ALS, MCR-WALS, and PMF models. Each time series was unit vector normalized before plotting to adjust for the differences in scale between the results of the different models.

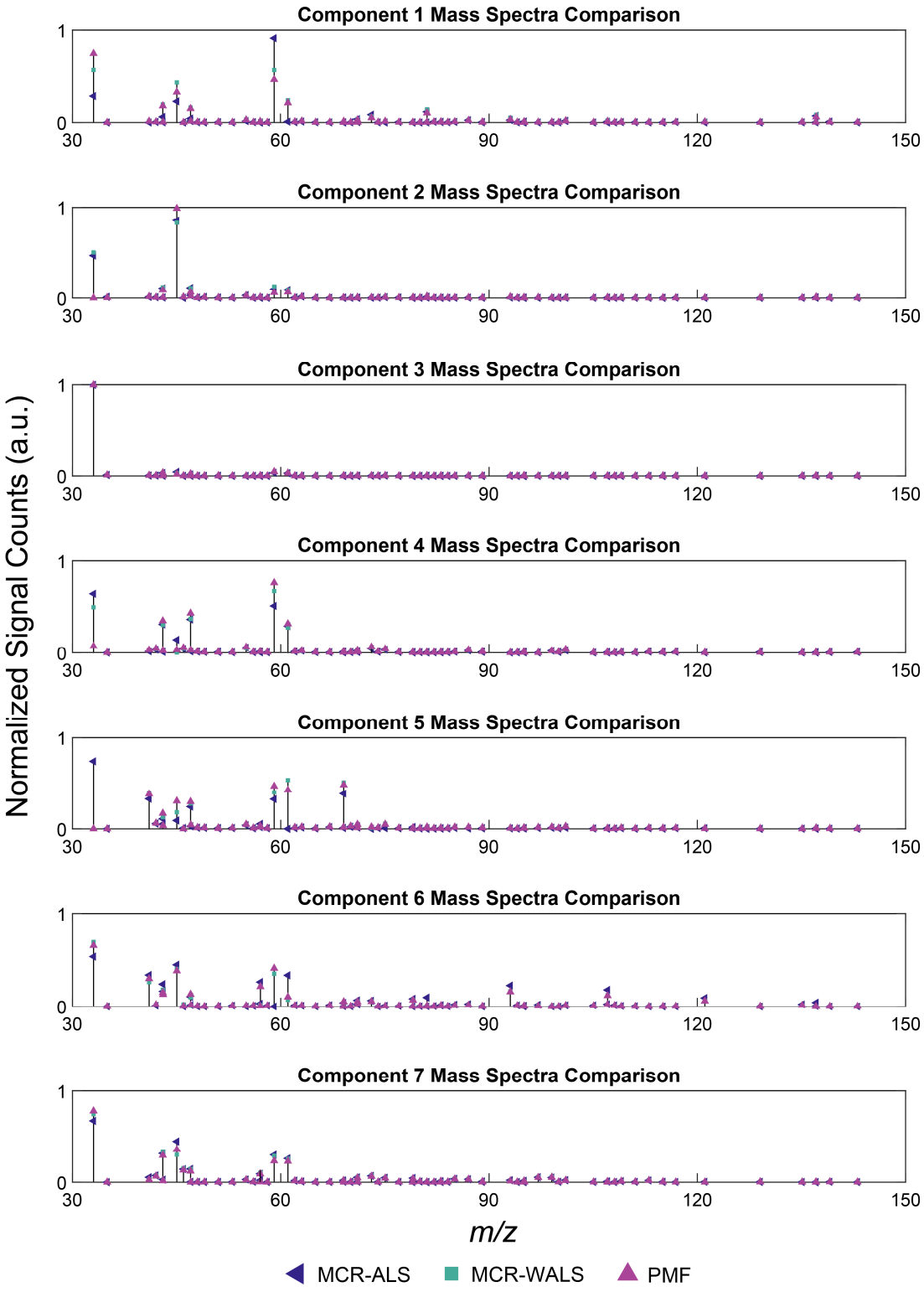


Figure 4.16: Comparison of the resolved mass spectra calculated for seven component MCR-ALS (blue triangles), MCR-WALS (aqua squares), and PMF (purple triangles) models. Each mass spectrum was unit vector normalized before plotting to adjust for the differences in scale between the results of the different models.

Table 4.3: Pairwise comparison between the resolved time series and mass spectra for each of the components in the seven component models.

Models compared	Component													
	1		2		3		4		5		6		7	
	Time series	Spectrum	Time series	Spectrum	Time series	Spectrum	Time series	Spectrum	Time series	Spectrum	Time series	Spectrum	Time series	Spectrum
PMF/ ALS	0.86	0.75	0.98	0.88	0.99	1.0	0.94	0.80	0.94	0.59	0.95	0.86	0.99	0.99
PMF/ WALS	0.99	0.97	0.99	0.86	0.99	1.0	0.98	0.90	0.92	0.98	1.0	0.99	1.0	0.99
ALS/ WALS	0.88	0.84	0.96	0.99	0.99	0.99	0.96	0.97	0.95	0.56	0.93	0.87	0.99	0.99

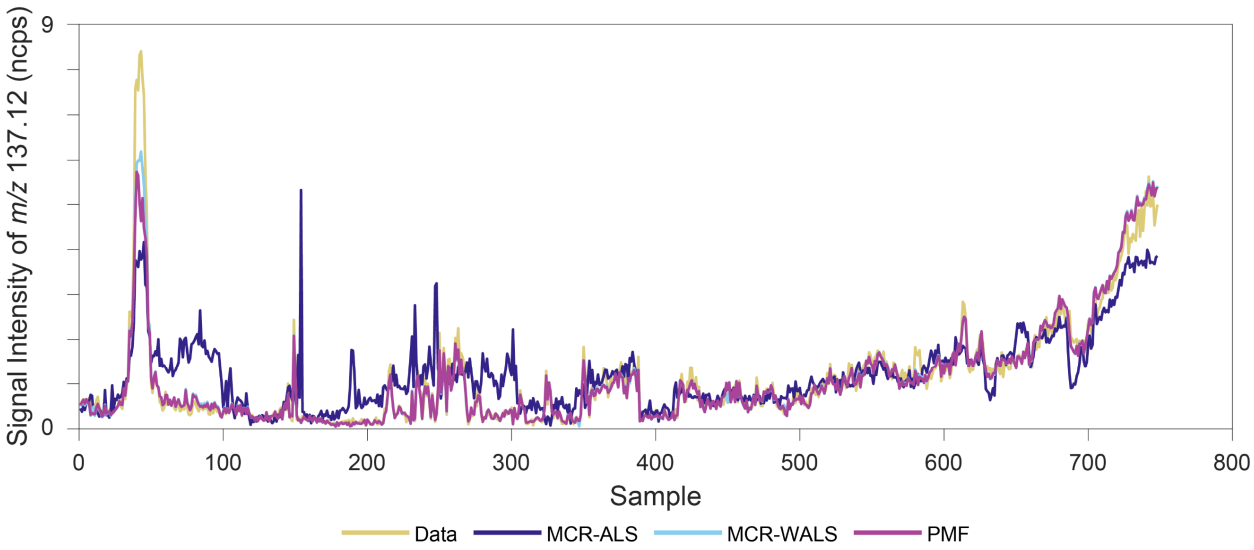


Figure 4.17: Overlay of the signal intensity at *m/z* 137.12 measured by the PTR-ToF-MS (yellow), modelled by MCR-ALS (dark blue), modelled by MCR-WALS (light blue), and modelled by PMF (pink). MCR-ALS produces a modelled signal intensity higher than the observed data, and with a different temporal profile, for samples 50–300.

This is the first comparison of PMF, MCR-WALS, and MCR-ALS for highly resolved mass spectral data collected from a moving vehicle, and these results are consistent with results reported in the literature for different data types (elemental speciation of PM_{10} aerosols, trace metals in river sediments, and simulated environmental data).[135,218,219] MCR-WALS extracts comparable source profiles and contributions to PMF and is a suitable open source alternative to PMF for datasets that are less compatible with the EPA PMF 5.0 software. MCR-WALS will be used to perform the source apportionment of VOCs using data collected by mobile mass spectrometry. In the next sections, the MCR-WALS analysis of the 1 second mass spectral data for both the

whole August 7, 2018 dataset containing all the measured *m/z* and a reduced dataset employing variable selection will be discussed.

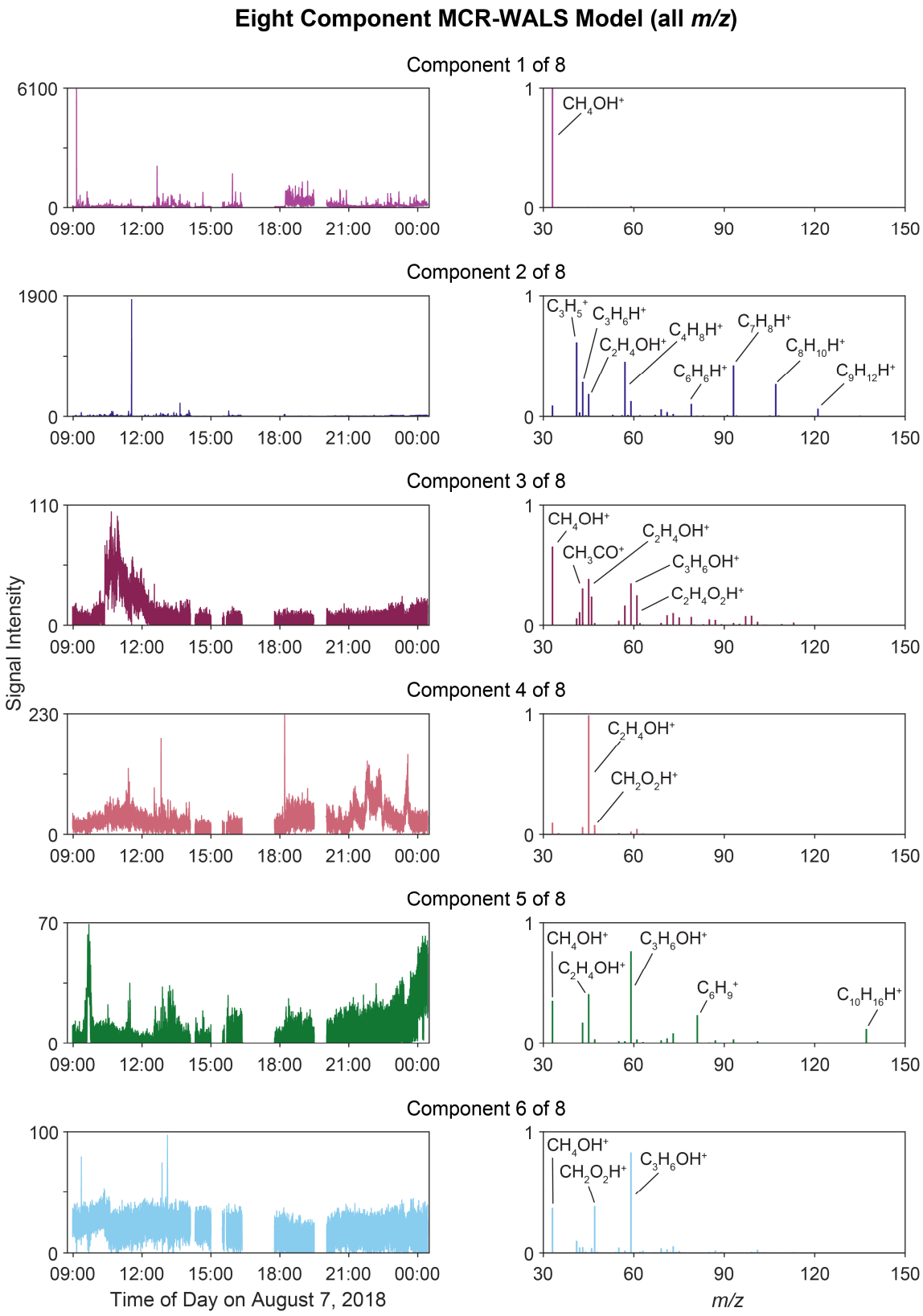
4.3.4 MCR-WALS analysis of August 7, 2018 data using all recorded m/z

On August 7, 2018 there were 307 *m/z* recorded from *m/z* 33.03–205.17, and in this section MCR-WALS will be applied to the whole dataset (1 second samples) to determine if applying the models without prior variable selection will produce results that can be interpreted for VOC source apportionment. MCR-ALS models were also calculated for interrogation, with the results discussed in Appendix C. MCR-WALS models containing four to ten components were calculated for investigation. For each model, the Meisch coefficients, x-y scatter plots and overlaid time series of the measured versus modelled signal intensity for each of the recorded *m/z* were used to evaluate the variable-by-variable goodness-of-fit. The Meisch coefficients for 26 of the recorded *m/z* observed on August 7, 2018 are shown in Table 4.4, with values greater than or equal to 0.90 bolded for emphasis. The Meisch coefficients show that for some of the recorded *m/z*, the observed data is well modelled regardless of the number of resolved components (*e.g.*, *m/z* 33.03 and *m/z* 45.03), while others, (*e.g.*, *m/z* 69.07) are modelled well in models with more components.

The time series and mass spectra for the eight and ten component MCR-WALS models will be discussed and compared. The time series and mass spectrum for each component in the eight component MCR-WALS model are shown in Figure 4.18, with the components assigned arbitrary numbers. Six of the eight components have time series with distinct structures (Components 1–6), while the time series for Components 7 and 8 are very similar and contain a high level of noise. The structure of the resolved time series are not going to be discussed in detail in this section, but for Components 1–6, many of the recorded concentration excursions coincide with sources described in field notes, or related to the sampling location. Component 1 is dominated by methanol measurements (*m/z* 33.03). Component 2 is related to hydrocarbon measurements including pentene, butene, benzene, toluene, and ethylbenzene/xylenes. The mass spectrum for Component 3 contains signals for a variety of VOCs including methanol, acetaldehyde, acetone, and acetic acid. The mass spectrum for Component 4 is mainly associated with acetaldehyde measurements. The mass spectrum for Component 5 has signals related to methanol, acetaldehyde, formic acid, acetone, acetic acid, and monoterpenes. Component 6 has a mass spectrum with measurements of methanol, formic acid, acetone, and acetic acid. Components 7 and 8 have similar resolved time series, but their resolved mass spectra differ slightly (methanol, formic acid, and acetone in Component 7; methanol, formic acid, and acetic acid in Component 8).

Table 4.4: Meisch coefficients for a selection of *m/z* measured in the August 7, 2018 field campaign for the six–ten component MCR-WALS models calculated using the full mass list. Values greater than 0.9 are bolded for emphasis.

	Number of MCR-WALS components				
m/z	**6**	**7**	**8**	**9**	**10**
33.03	**0.999**	**0.999**	**0.999**	**1.00**	**1.00**
41.03	0.819	0.816	0.816	0.816	0.868
42.03	0.440	0.460	0.461	0.463	0.462
43.01	0.732	**0.968**	**0.995**	**0.994**	**0.997**
43.05	0.726	0.720	0.720	0.720	0.725
45.03	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
47.01	0.382	0.382	0.400	0.397	0.404
55.01	0.151	0.167	0.167	**0.969**	**0.972**
57.03	0.629	0.687	0.692	0.694	0.692
57.07	0.867	0.872	0.871	0.871	0.871
59.05	0.782	0.798	0.826	0.843	0.897
61.03	0.704	0.706	**0.976**	**0.976**	**0.954**
69.07	0.061	0.065	0.068	0.071	**0.921**
71.04	0.417	0.447	0.449	0.453	0.453
73.06	0.522	0.533	0.535	0.533	0.545
75.04	0.444	0.444	0.446	0.446	0.447
79.05	0.735	0.735	0.733	0.733	0.733
81.07	0.773	0.771	0.772	0.775	0.777
85.06	0.439	0.461	0.462	0.463	0.463
87.04	0.398	0.405	0.405	0.405	0.407
93.07	0.682	0.651	0.649	0.633	0.636
97.09	0.607	0.675	0.680	0.685	0.685
99.04	0.626	0.646	0.645	0.644	0.645
107.08	0.827	0.884	0.889	0.898	**0.902**
121.09	0.512	0.529	0.530	0.535	0.535
137.12	0.711	0.710	0.711	0.714	0.703


Eight Component MCR-WALS Model (all m/z)
Component 1 of 8
6100
0
09:00
12:00
15:00
18:00
21:00
00:00
1
CH4OH+
30
60
90
120
150
Component 2 of 8
1900
C3H5+
C3H6H+
C2H4OH+
C4H8H+
C7H8H+
C8H10H+
C6H6H+
C9H12H+
Component 3 of 8
110
Signal Intensity
CH4OH+
C2H4OH+
CH3CO+
C3H6OH+
C2H4O2H+
Component 4 of 8
230
CH2O2H+
Component 5 of 8
70
C6H9+
C10H16H+
Component 6 of 8
100
Time of Day on August 7, 2018
m/z

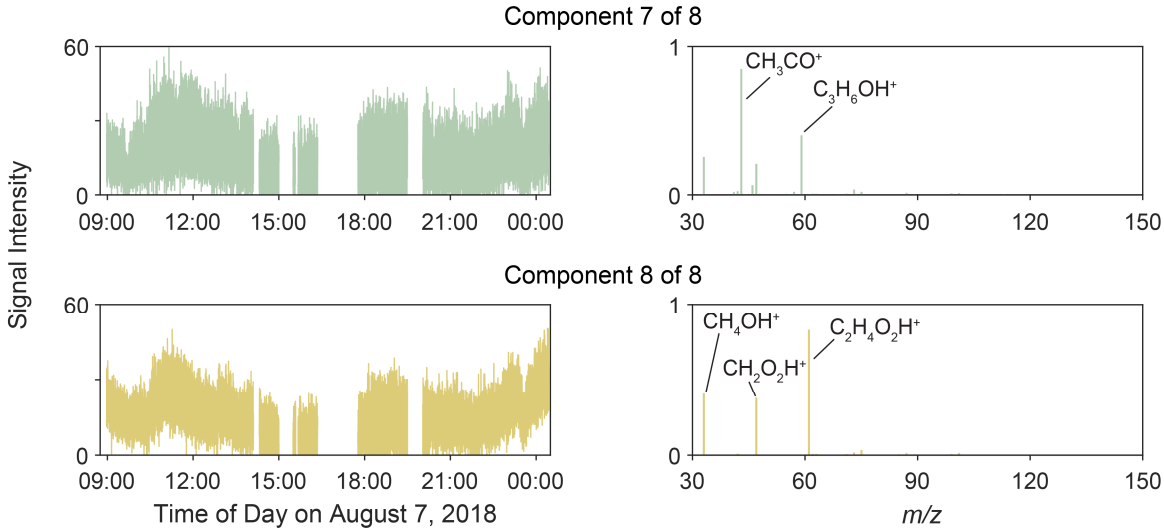


Figure 4.18: Resolved time series (left) and mass spectra (right) for the eight component MCR-WALS model calculated using all recorded *m/z* for the data collected on August 7, 2018 in Port Alberni, BC.

In this eight component MCR-WALS model, some observations that were measured at specific locations, such as signals measured at *m/z* 69.07 (isoprene), have not been well modelled, as is expected from the Meish coefficients in Table 4.4. The overlay of the observed (burgundy) and modelled (green) signal intensity at *m/z* 69.07 in the eight component MCR-WALS model is shown in Figure 4.19, with an inset showing the scatter plot of the same data. From these two plots it is evident that a component related to isoprene source(s) has not been calculated in the model. Other signals, such as *m/z* 137.12 (monoterpenes), are better modelled here as is captured by the Meisch coefficients in Table 4.4. The overlay of the measured and modelled signal intensities at *m/z* 137.12, and inset scatter plot of the same data is shown in Figure 4.20. The modelled time series in Figure 4.20 captures temporal trends in the measured data.

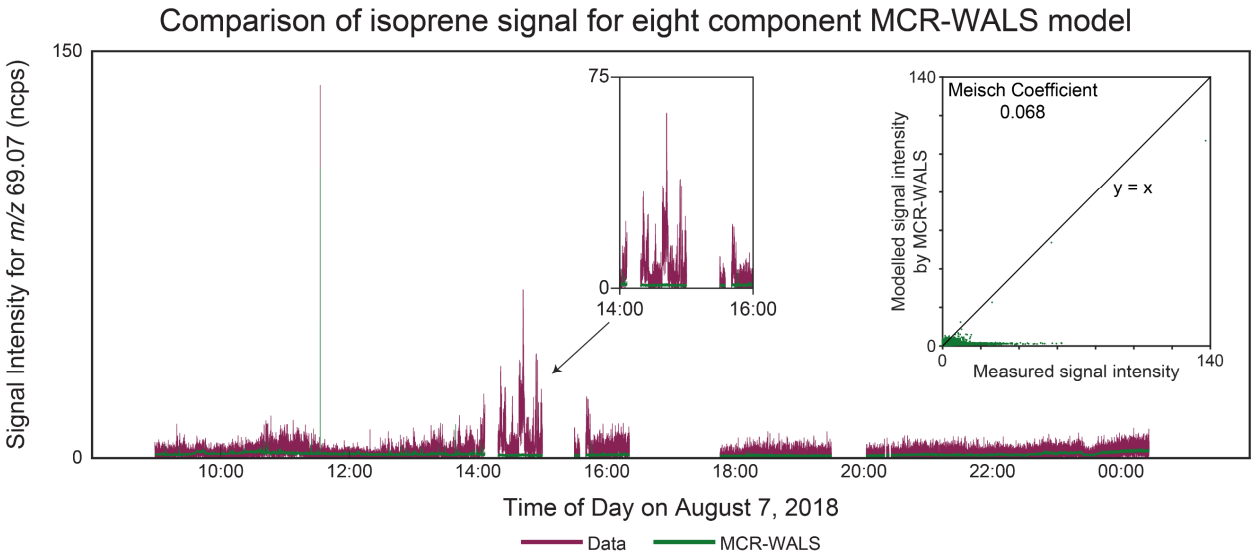


Figure 4.19: Time series and scatter plot (inset) comparison of the observed signal measured for isoprene (burgundy) and the isoprene signal modelled by the eight component MCR-WALS model (green). The observed signal intensity is not accounted for in the model, indicating that MCR-WALS models with more components are needed to account for the measurements recorded at *m/z* 69.07.

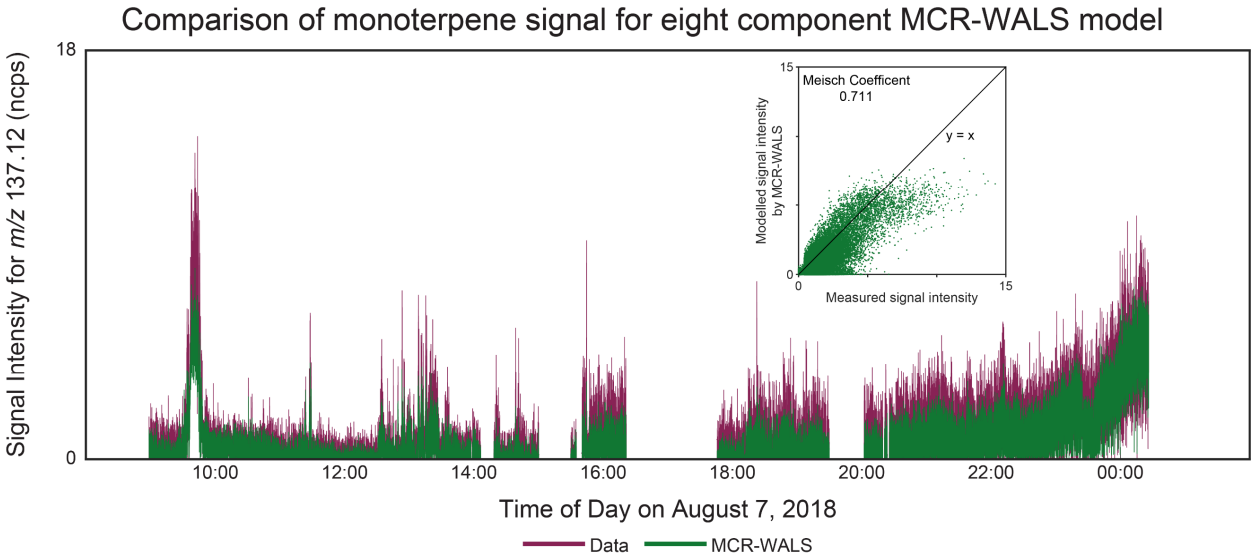
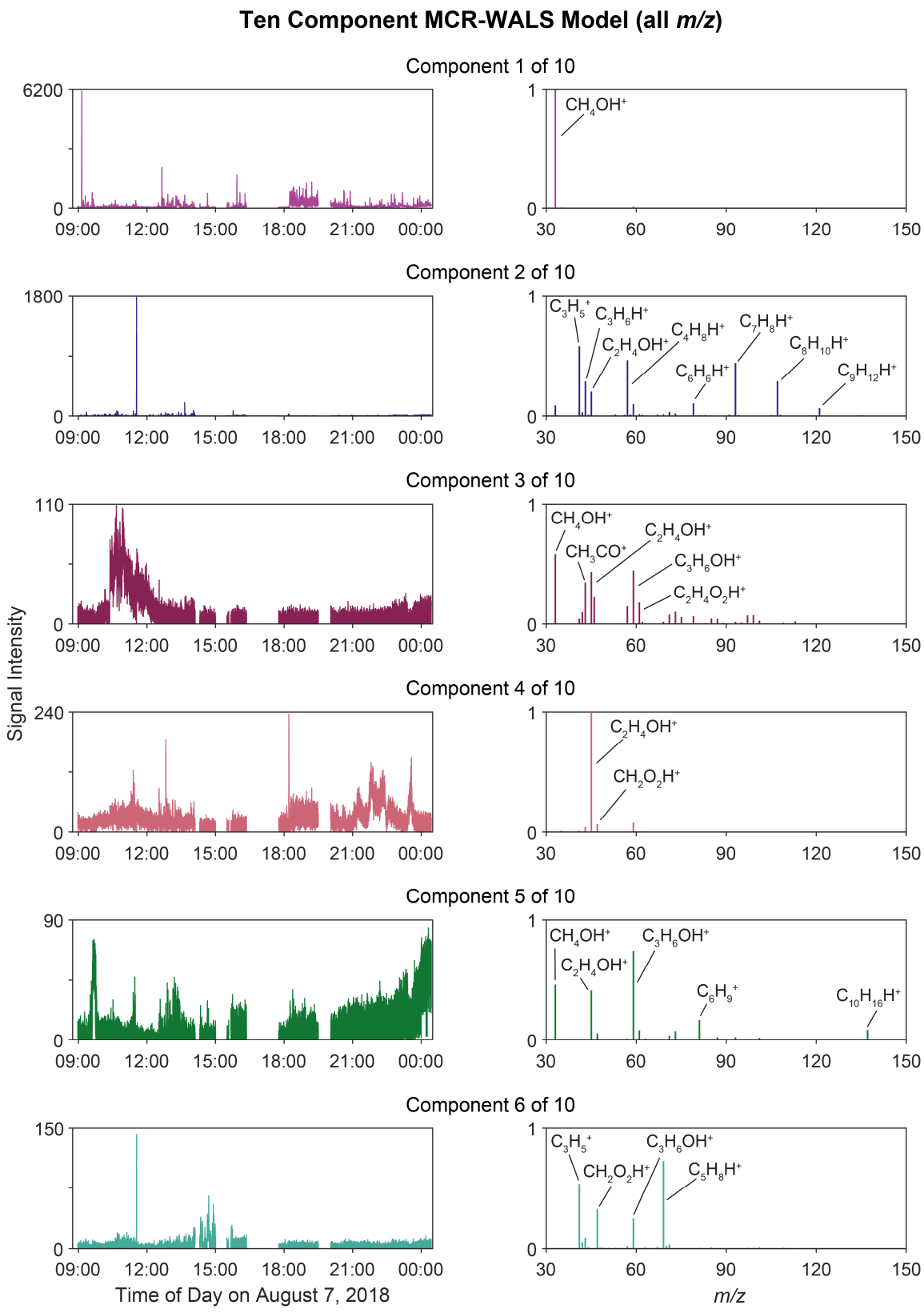


Figure 4.20: Time series and scatter plot (inset) comparison of the observed signal measured for monoterpenes (burgundy) and the monoterpene signal modelled by the eight component MCR-WALS model (green). The observed signal intensity is described well by the model.

A ten component MCR-WALS model is needed to describe the signal observed at *m/z* 69.07, with Meisch coefficient increasing from 0.071 for a nine component MCR-WALS model (not shown) to 0.921 in a ten component model. The results of the ten component model are shown in Figure 4.21. This model has seven components with unique time series, including a component that captures the isoprene measurements, and three components with similar time series with high levels of noise. The component associated with the isoprene measurements has a mass spectrum that includes measurements of isoprene, formic acid, and

acetone. The overlay time series, and scatter plot, of the measured and modelled data at *m/z* 69.07 shown in Figure 4.22 indicate that the ten component MCR-WALS model describes the observed isoprene signal.

The eight and ten component MCR-WALS models shown in Figure 4.18 and Figure 4.21 were calculated on the whole dataset (mass spectra recorded at 1 second intervals and 308 observed *m/z*). These models allowed for unique time series describing the observations in the data to be calculated, but multiple components with very similar time series, but different mass spectra, were calculated before components describing interesting temporal trends in the data, as shown with the isoprene measurements at *m/z* 69.07. This is also the case with the MCR-ALS models shown in Appendix C. Many of the recorded *m/z* with signals measured during field campaigns have a S/N < 1 for all measurements across the sampling day, and in the next section the use of a S/N-based variable selection to reduce the noise level in the dataset before MCR-WALS analysis will be explored.


Ten Component MCR-WALS Model (all m/z)
Component 1 of 10
6200
0
09:00
12:00
15:00
18:00
21:00
00:00
1
0
CH4OH+
30
60
90
120
150
Component 2 of 10
1800
C3H5+
C3H6H+
C2H4OH+
C4H8H+
C7H8H+
C8H10H+
C6H6H+
C9H12H+
Component 3 of 10
110
CH4OH+
C2H4OH+
CH3CO+
C3H6OH+
C2H4O2H+
Signal Intensity
Component 4 of 10
240
C2H4OH+
CH2O2H+
Component 5 of 10
90
CH4OH+
C3H6OH+
C2H4OH+
C6H9+
C10H16H+
Component 6 of 10
150
C3H5+
C3H6OH+
CH2O2H+
C5H8H+
Time of Day on August 7, 2018
m/z

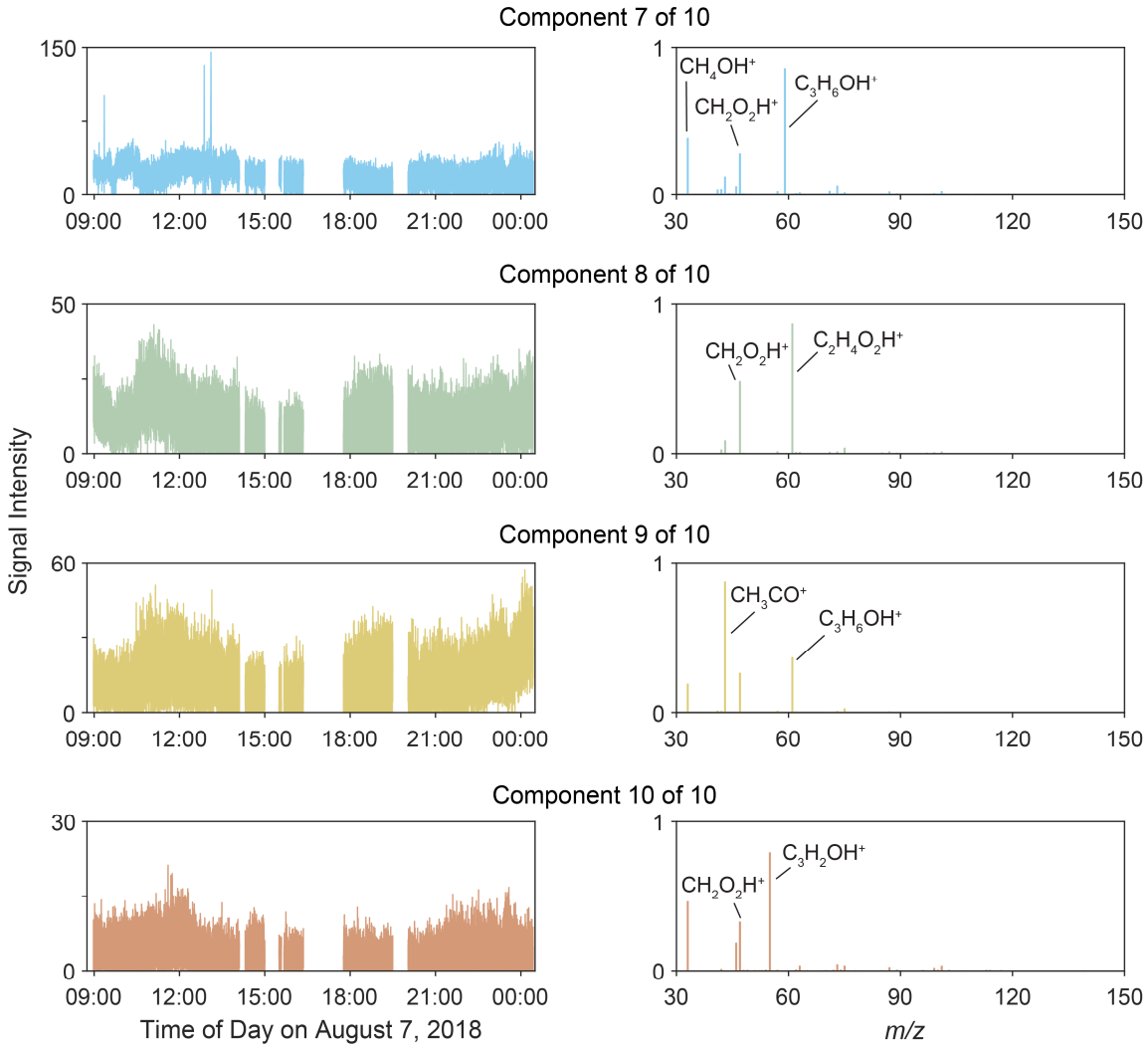


Figure 4.21: Resolved time series and mass spectra for the ten component MCR-WALS model calculated using all recorded *m/z* for the data collected on August 7, 2018 in Port Alberni, BC.

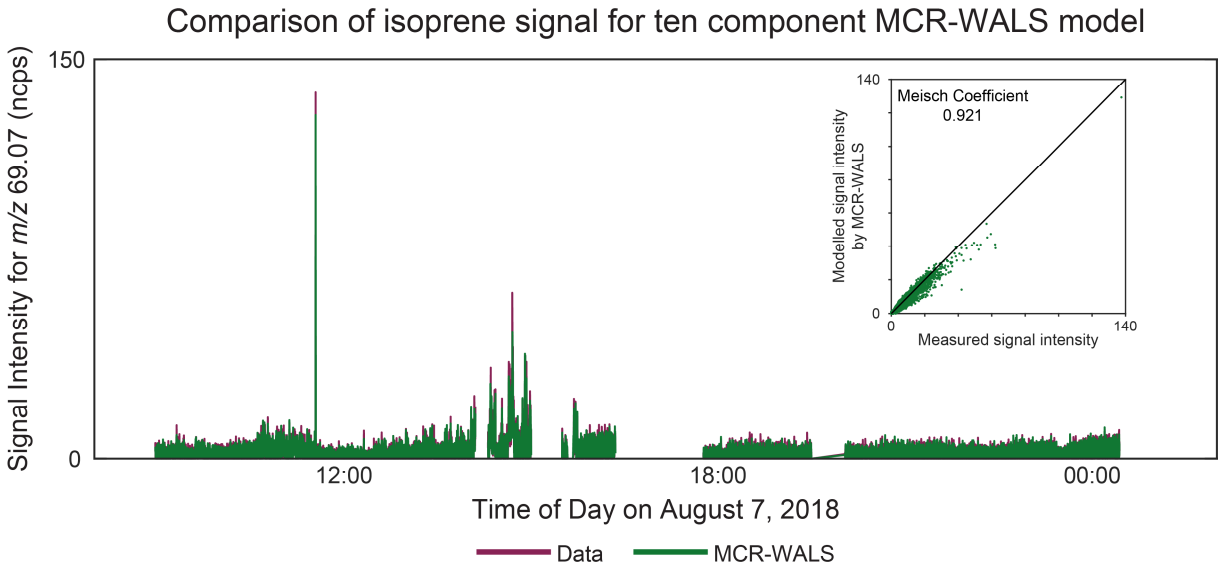


Figure 4.22: Time series and scatter plot (inset) comparison of the observed signal measured for isoprene (burgundy) and the isoprene signal modelled by the ten component MCR-WALS model (green). The observed signal intensity is accounted for in the model.

4.3.5 MCR-WALS analysis of August 7, 2018 data using variable selection

To improve the quality of results and better model the signals associated with the measured *m/z* observed to have structure in their time series related to VOC sources encountered on the field campaign, a subset of the measured *m/z* were used for subsequent MCR-WALS models. Usually in a PMF analysis variables that have observations with an average S/N < 0.5 (as defined by Equation 4.7) across the dataset are considered "poor" and excluded from the analysis, variables with average S/N between 0.5 and 1 are considered "weak" and further down weighted, and those with average S/N > 1 are considered "strong" and used as is. The data presented in this chapter was collected at a high frequency (once per second) which at city driving speeds (*e.g.*, 50 km/hr) leads to a mass spectrum being collected approximately every 14 meters. To preserve the high geospatial resolution of the data, the "poor", "weak", "strong" classification for variable selection used by PMF needed to be adjusted, as the nature of the data (noisy, some recorded *m/z* had observed concentration excursion for short periods of time compared to the length of the sampling day) resulted in very few *m/z* being preserved for the analysis using these criteria. Therefore, instead of selecting only those recorded *m/z* with high average S/N over the course of the sampling day, recorded *m/z* with at least one measurement with a S/N > 1 were included in the subsequent models. This allowed the information from short duration signals arising from driving by point sources to be preserved in the analysis. For the August 7, 2018 data collected in Port Alberni, this variable selection reduced the mass list from 307 to 71 *m/z* to be included as variables in the MCR-WALS analysis, which also significantly reduced computation times. For comparison purposes, MCR-ALS models were also calculated using the reduced mass list, and the results of this analysis are shown in Appendix C.

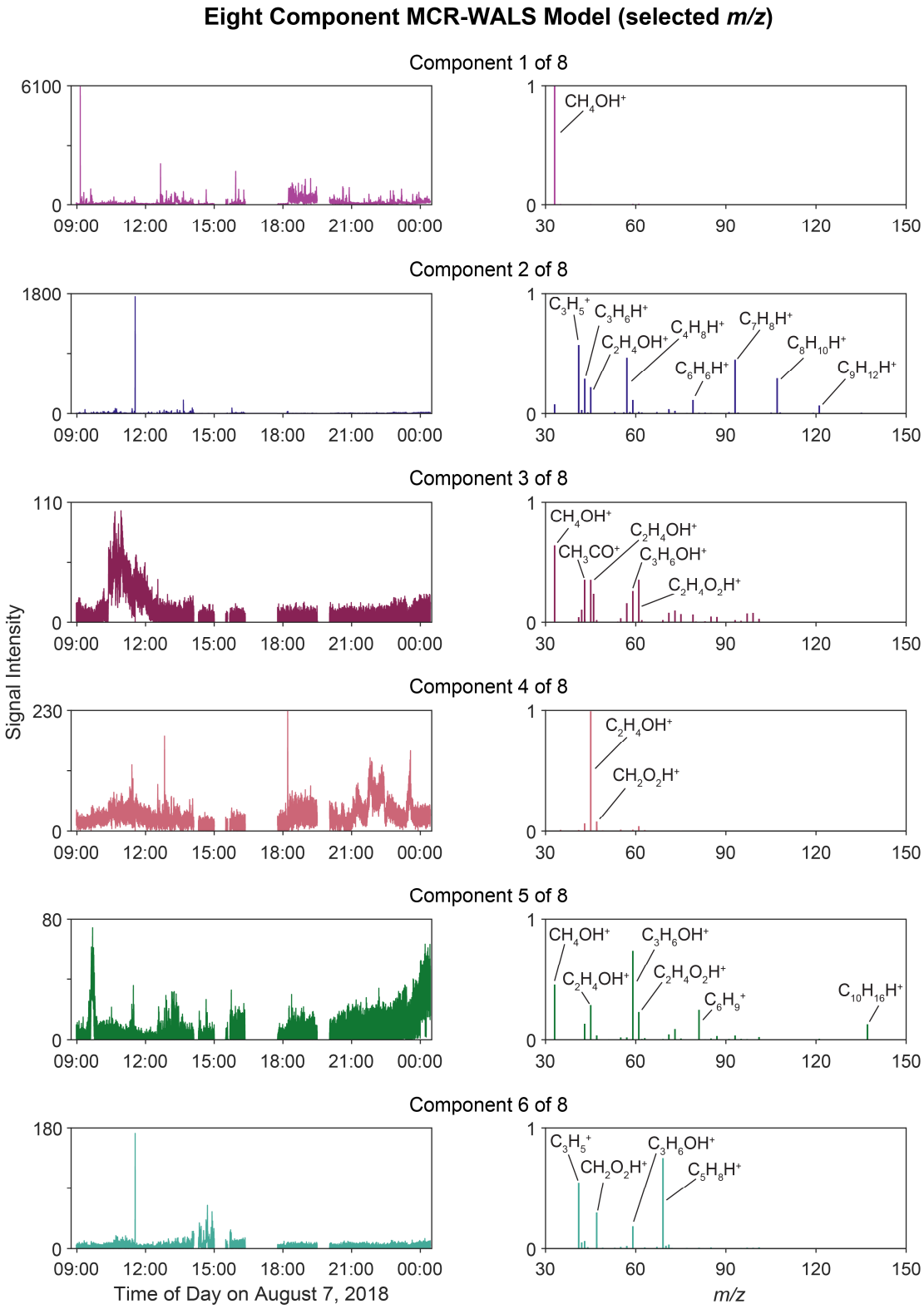
The Meisch coefficients for the main *m/z* for six to twelve component MCR-WALS models are shown in Table 4.5, with values greater than 0.9 bolded for emphasis. From these values, the time series and mass spectra for eight and ten component MCR-WALS models will be interrogated. Both these models result in the observed signals at *m/z* 69.07 (isoprene) being modelled well, and the ten component model improves the modeling of the observed signals recorded at *m/z* 59.05, and 61.03 when compared to the eight component model.

Table 4.5: Meisch coefficients for the main *m/z* measured in the August 7, 2018 field campaign for six–twelve component MCR-WALS models utilizing variable selection. Values greater than 0.9 are bolded for emphasis.

	Number of MCR-WALS components						
m/z	**6**	**7**	**8**	**9**	**10**	**11**	**12**
33.03	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
41.03	0.816	0.813	0.866	0.813	0.865	**0.992**	**0.988**
42.03	0.439	0.458	0.458	0.461	0.460	0.460	0.473
43.01	0.740	**0.970**	**0.972**	**0.995**	**0.997**	**0.998**	**0.999**
43.05	0.724	0.720	0.715	0.720	0.717	0.694	0.693
45.03	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
57.03	0.628	0.689	0.690	0.695	0.695	0.693	0.703
57.07	0.866	0.671	0.869	0.870	0.869	0.850	0.851
59.05	0.805	0.827	0.849	0.882	**0.938**	**0.941**	**0.984**
61.03	0.708	0.706	0.711	**0.970**	**0.958**	**0.958**	**0.992**
69.07	0.059	0.063	**0.923**	0.067	**0.924**	**0.993**	**0.995**
71.04	0.415	0.446	0.446	0.453	0.450	0.451	0.451
73.06	0.520	0.532	0.536	0.532	0.537	0.538	0.532
75.04	0.443	0.443	0.444	0.445	0.446	0.446	0.450
79.05	0.734	0.740	0.736	0.736	0.733	0.711	0.707
81.07	0.783	0.780	0.786	0.783	0.790	0.796	0.816
85.06	0.435	0.459	0.459	0.462	0.461	0.462	0.462
87.04	0.396	0.403	0.405	0.403	0.404	0.404	0.404
93.07	0.666	0.631	0.632	0.617	0.626	0.703	0.718
97.09	0.596	0.669	0.671	0.681	0.679	0.680	0.685
99.04	0.621	0.643	0.644	0.642	0.643	0.644	0.643
107.08	0.836	0.890	**0.901**	**0.900**	**0.905**	**0.928**	**0.926**
121.09	0.516	0.534	0.537	0.538	0.538	0.532	0.531
137.12	0.720	0.717	0.721	0.718	0.722	0.726	0.728

The time series and mass spectra for the eight component MCR-WALS model are shown in Figure 4.23. The components resolved in this model are very similar to the components shown in the ten component model without variable selection (Figure 4.21), with the exception that there is only one component (Component 8) that has little structure in the time series. By utilizing variable selection to eliminate the recorded *m/z* that contained little information, less components were needed in the model to resolve the time series with interesting temporal profiles. However, as shown in Figure 4.24, some of the observed signal intensity measured

at *m/z* 59.05 is not well modelled using eight components, and can be improved with a ten component MCR-WALS model.



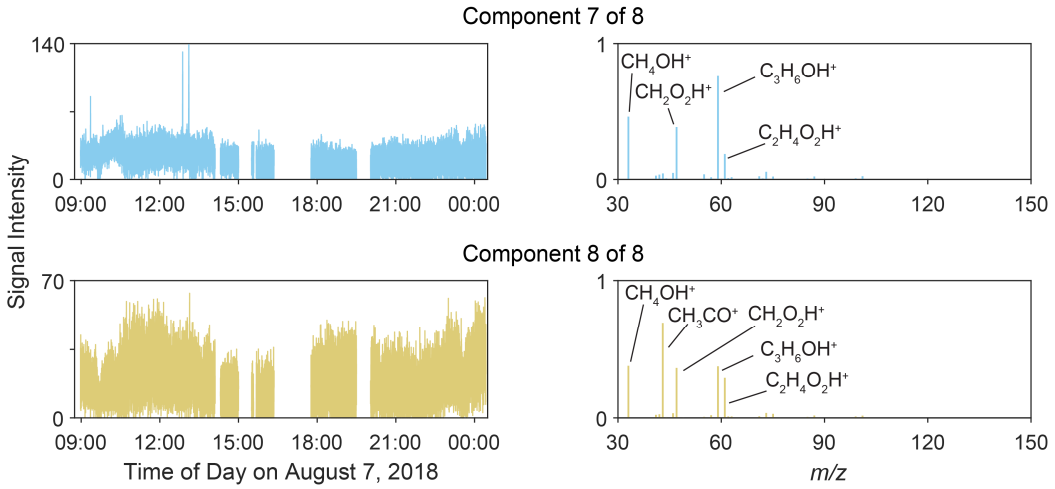


Figure 4.23: Results from the eight component MCR-WALS model for August 7, 2018 data using variable selection. Only recorded *m/z* with measurements with S/N > 1 were included in the analysis.

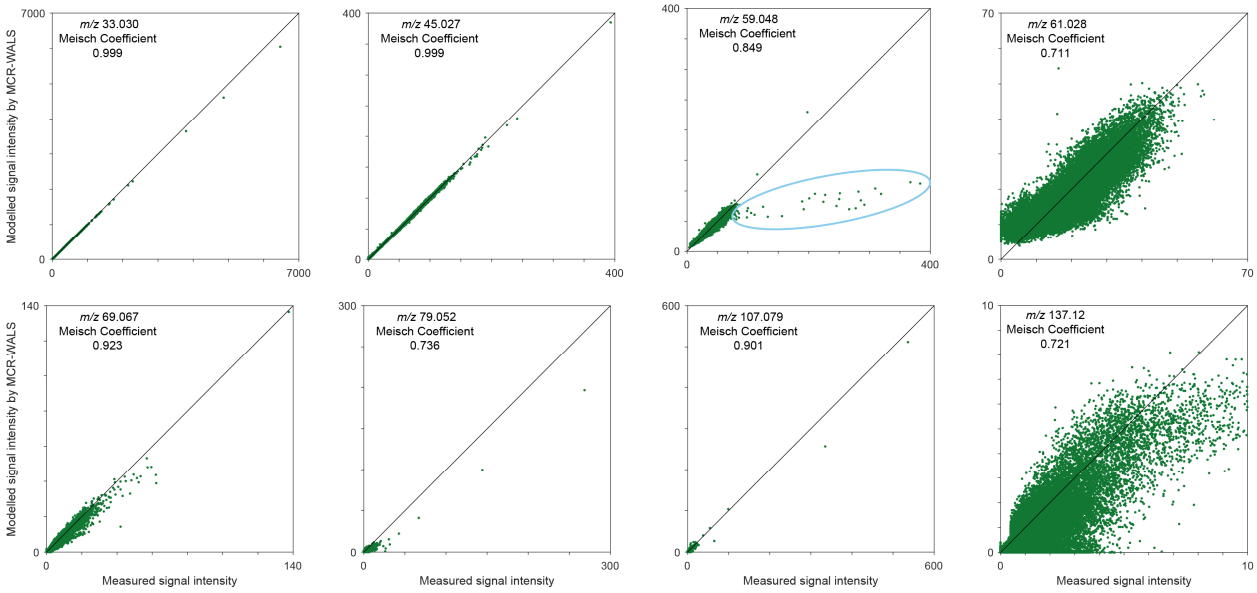
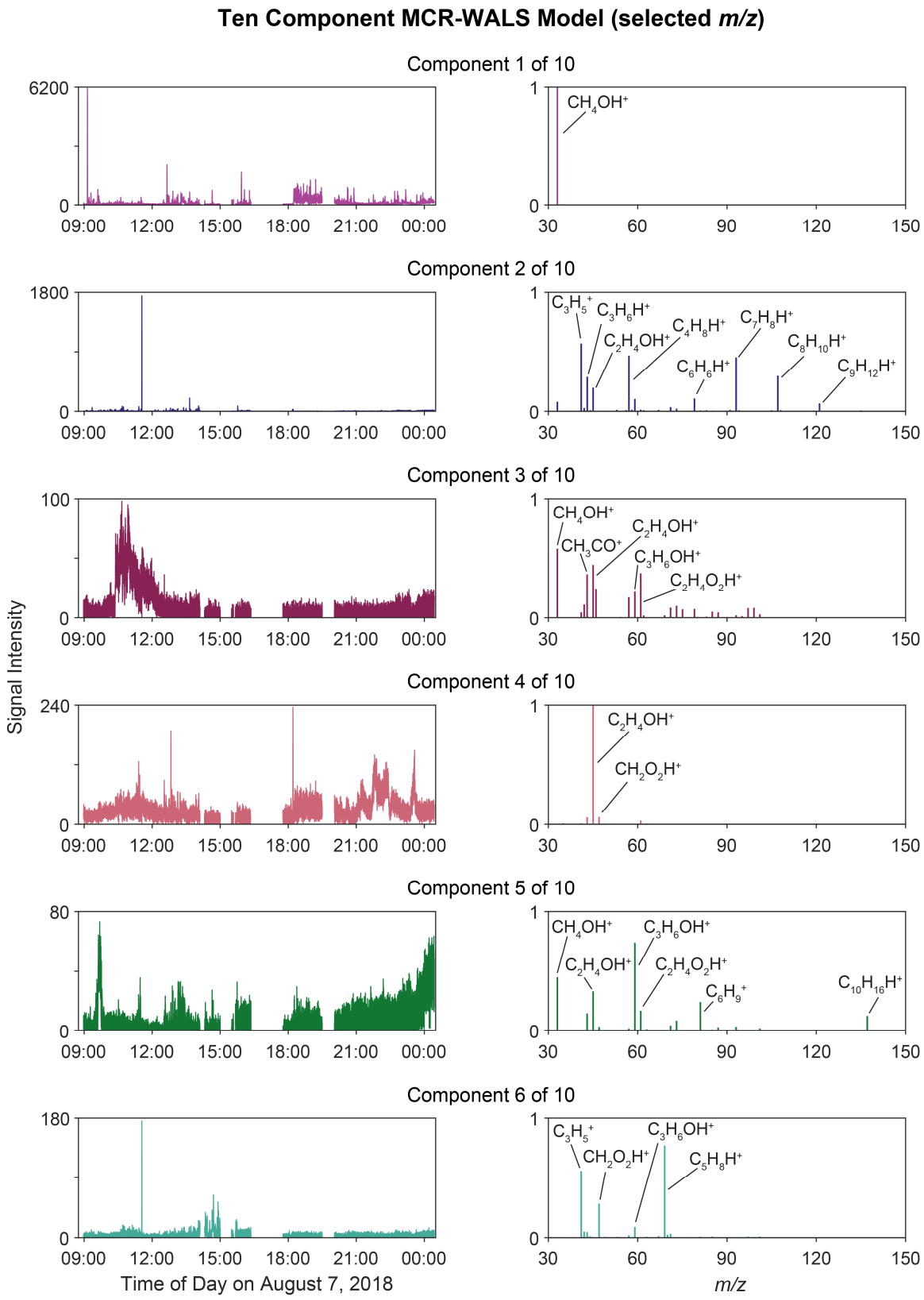


Figure 4.24: Scatter plots of the observed versus modelled signal intensity (eight component MCR-WALS) for a selection of *m/z*. The circled data on the scatter plot for *m/z* 59.05 indicate that another component(s) is required to model all the observations recorded at this *m/z*.

The ten component MCR-WALS model improves on the modeling of the observed signal at *m/z* 59.05 (Table 4.5), but as shown in Figure 4.25 the ten component MCR-WALS model has three components with almost the same time series (Components 8–10), but with different resolved mass spectra. This trend continues for models with more than ten components (data not shown), where the addition of more components to the model improves the modeling of the signals measured for some recorded *m/z*, but these models resolve degenerate components where individual components are split in two components with similar time series and different mass spectra. An example of this is shown in Figure 4.26 where in the eleven component MCR-WALS model the new component has a time series very similar to Component 2 in the eight and ten component MCR-WALS models. In Figure 4.26 the top panel shows the time series for Components 2 and 11 (reflected) in the eleven component MCR-WALS model, and the bottom panel shows the mass spectra. The time series have similar temporal profiles except for the maximum signal intensity for some of the concentration excursions, while the mass spectra are very different. Component 2 has signals associated with many hydrocarbons (*e.g.*, BTEX compounds), while Component 11 contains the signal for the ion measured at *m/z* 41.03, which is a fragment ion of many hydrocarbon compounds.


Ten Component MCR-WALS Model (selected m/z)
Component 1 of 10
6200
0
09:00
12:00
15:00
18:00
21:00
00:00
1
CH4OH+
30
60
90
120
150
Component 2 of 10
1800
C3H5+
C3H6H+
C2H4OH+
C4H8H+
C7H8H+
C6H6H+
C8H10H+
C9H12H+
Component 3 of 10
100
CH4OH+
CH3CO+
C2H4OH+
C3H6OH+
C2H4O2H+
Component 4 of 10
240
C2H4OH+
CH2O2H+
Component 5 of 10
80
CH4OH+
C3H6OH+
C2H4OH+
C2H4O2H+
C6H9+
C10H16H+
Component 6 of 10
180
C3H5+
CH2O2H+
C3H6OH+
C5H8H+
Signal Intensity
Time of Day on August 7, 2018
m/z

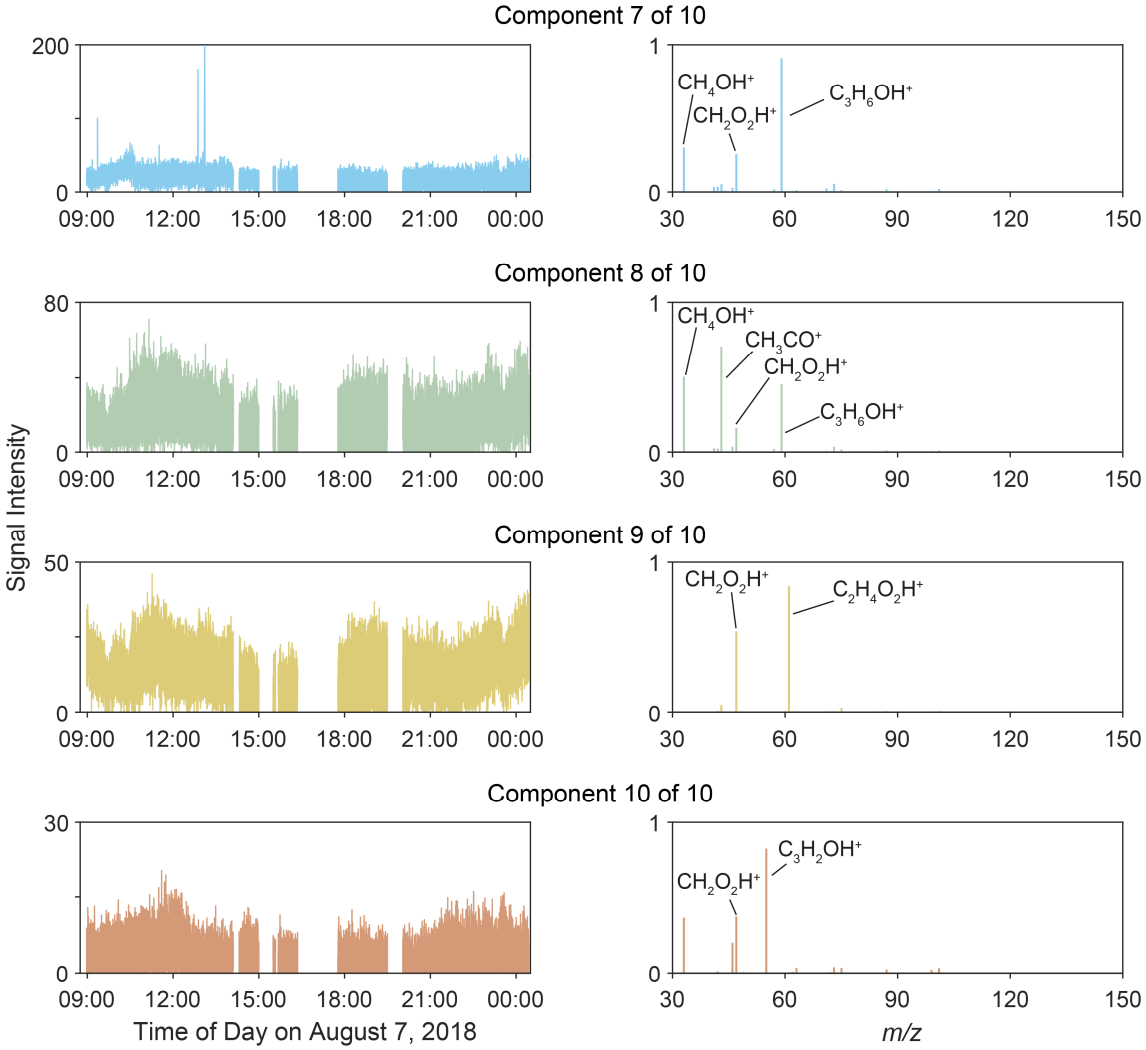


Figure 4.25: Results from the ten component MCR-WALS model for August 7, 2018 data using variable selection. Only recorded *m/z* with measurements with S/N > 1 were included in the analysis.

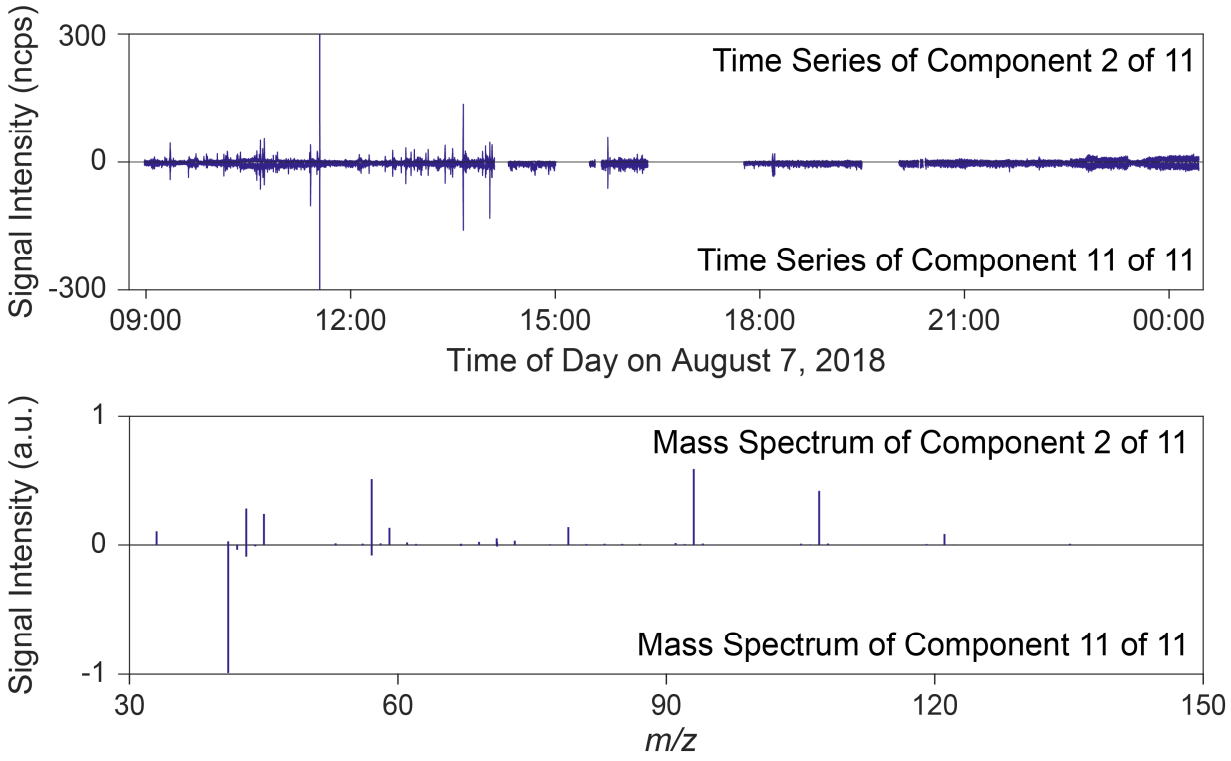


Figure 4.26: The eleven component MCR-WALS model resolves two components with similar time series, but different mass spectra. In the top figure, the time series of Component 2 of 11 is shown, with a reflection of Component 11 of 11 shown below. There are some differences in the absolute signal intensity of the concentration excursions, but the overall temporal profile of the two time series match. In the bottom panel the mass spectra for the two components are shown, with the mass spectrum for Component 11 reflected underneath the mass spectrum for Component 2. Even though the time series of the two components are very similar, the mass spectra are very different. The mass spectrum for Component 2 contains signals for many recorded *m/z* associated with hydrocarbons (*e.g.,* BTEX), while the bottom mass spectrum mainly contains the signal for the ion recorded at *m/z* 41.03, which corresponds to an alkyl fragment generated from many hydrocarbon species.

As the models containing more than eight components do not result in more unique components being resolved, the eight component MCR-WALS model that employs variable selection has been chosen for interpretation for the August 7, 2018 field campaign data, and will be discussed in the subsequent section. Employing MCR-WALS with variable selection allows for unique VOC components in the data to be calculated with simpler models compared to using MCR-ALS with or without variable selection (Appendix C), or using MCR-WALS without variable selection, and also reduces the computation time required to generate the models. MCR-WALS employing variable selection will be extended to the August 14, 2018 and August 21–22, 2017 field campaigns to identify and apportion the VOC sources encountered while sampling.

4.3.6 August 7, 2018 MCR-WALS results

Physical descriptions for the resolved components were determined by examining the field notes, time series, and geographic locations where the components were measured, the chemical species present in the resolved mass spectra, and in some instances comparing the time series of a resolved VOC component to other

air quality parameters from other chemical sensors in the mobile lab. For the August 7, 2018 dataset, seven of the eight resolved components were observed to have distinct temporal profiles (Figure 4.23) with the source contributions varying over times and space. Five components contained VOC mixtures and were associated with a type of VOC source (hydrocarbon emissions from anthropogenic activities, forest fire smoke, emissions from vegetation or wood products, emissions from vegetation, emissions from commercial activities), two components were dominated by individual VOCs (*e.g.,* methanol and acetaldehyde), and the remaining component displayed little temporal variation across the sampling period. Table 4.6 lists the major *m/z* measured for each component, along with a descriptor for the source, and the general locations or point/area sources around which the components were measured. Each component is discussed in detail below, along with expanded, annotated time series and maps showing the spatial distribution. The components are not described in numerical order, which was arbitrary.

Table 4.6: Summary of the *m/z* contributing to each component, and the locations where the components were observed, in the eight component MCR-WALS analysis of the August 7, 2018 Port Alberni data.

Component	**Main *m/z* (signal intensity) and possible chemical identity**	**Locations**
1 – Methanol	33.03 (1) – CH_4OH^+ (methanol)	Pulp mill Pulp mill odor Vehicles Landfill Greenhouses
2 – Anthropogenic hydrocarbons	33.03 (0.08) – CH_4OH^+ (methanol) 41.03 (0.55) – $C_3H_5^+$ (alkyl fragment) 43.05 (0.29) – $C_3H_6H^+$ (propene, cyclopropane) 45.03 (0.22) – $C_2H_4OH^+$ (acetaldehyde) 57.07 (0.47) – $C_4H_8H^+$ (butene) 59.05 (0.11) – $C_3H_6OH^+$ (acetone, propanal) 79.05 (0.11) – $C_6H_6H^+$ (benzene) 93.06 (0.45) – $C_7H_8H^+$ (toluene) 107.08 (0.29) – $C_8H_{10}H^+$ (ethylbenene/xylenes) 121.09 (0.07) – $C_9H_{12}H^+$ (C9-aromatics)	Intersections Vehicles
3 – Forest fire VOCs	33.03 (0.64) – CH_4OH^+ (methanol) 42.03 (0.10) – $C_2H_3NH^+$ (acetonitrile) 43.01 (0.36) – CH_3CO^+ (acetic acid fragment) 45.03 (0.36) – $C_2H_4OH^+$ (acetaldehyde) 45.99 (0.24) – NO_2^+ 57.03 (0.16) – $C_3H_4OH^+$ (acrolein) 59.05 (0.26) – $C_3H_6OH^+$ (acetone) 61.03 (0.36) – $C_2H_4O_2H^+$ (acetic acid) 71.04 (0.08) – $C_4H_6OH^+$ (methacrolein) 73.06 (0.10) – $C_4H_8OH^+$ (methyl ethyl ketone) 75.04 (0.07) – $C_3H_6O_2H^+$ (methyl acetate, ethyl formate, propanoic acid) 79.05 (0.06) – $C_6H_6H^+$ (benzene) 85.06 (0.05) – $C_5H_8OH^+$ (cyclobutanone, methylbutenone) 99.04 (0.08) – $C_6H_6O_2H^+$ (furfual alcohol)	Forest fire smoke

Table 4.6: Summary of the *m/z* contributing to each component, and the locations where the components were observed, in the eight component MCR-WALS analysis of the August 7, 2018 Port Alberni data, continued.

Component	Main *m/z* (signal intensity) and possible chemical identity	Locations
4 – Acetaldehyde	43.01 (0.06) – CH_3CO^+ (carbonyl compound fragment) 45.03 (0.99) – $C_2H_4OH^+$ (acetaldehyde) 47.01 (0.08) – $CH_2O_2H^+$ (formic acid)	Pulp mill Vehicles
5 – Emissions from vegetation or wood products	33.03 (0.46) – CH_4OH^+ (methanol) 43.01 (0.13) – CH_3CO^+ (acetic acid fragment) 45.03 (0.29) – $C_2H_4OH^+$ (acetaldehyde) 59.05 (0.74) – $C_3H_6OH^+$ (acetone) 61.03 (0.23) – $C_2H_4O_2H^+$ (acetic acid) 73.06 (0.09) – $C_4H_8OH^+$ (butanone, butanal) 81.07 (0.25) – $C_6H_9^+$ (monoterpene fragment) 137.12 (0.13) – $C_{10}H_{16}H^+$ (monoterpenes)	Old growth forest Lumber yard Woodchip piles Trucks carrying woodchips
6 – Biogenic emissions from vegetation	41.03 (0.55) – $C_3H_5^+$ (alkyl fragment) 43.01 (0.06) – CH_3CO^+ (carbonyl compound fragment) 47.01 (0.30) – $CH_2O_2H^+$ (formic acid) 59.05 (0.18) – $C_3H_6OH^+$ (acetone) 69.07 (0.75) – $C_5H_8H^+$ (isoprene)	Vegetation near river and estuary
7 – Small oxygenated species	33.03 (0.46) – CH_4OH^+ (methanol) 47.01 (0.39) – $CH_2O_2H^+$ (formic acid) 59.05 (0.76) – $C_3H_6OH^+$ (acetone) 61.03 (0.19) – $C_2H_4O_2H^+$ (acetic acid) 73.06 (0.06) – $C_4H_8OH^+$ (methyl ethyl ketone)	Shipwright Auto salvage Boat repair
8 – Other	33.03 (0.38) – CH_4OH^+ (methanol) 43.01 (0.69) – CH_3CO^+ (acetic acid fragment) 47.01 (0.37) – $CH_2O_2H^+$ (formic acid) 59.05 (0.38) – $C_3H_6OH^+$ (acetone) 61.03 (0.30) – $C_2H_4O_2H^+$ (acetic acid) 73.06 (0.03) – $C_4H_8OH^+$ (methyl ethyl ketone) 75.04 (0.03) – $C_3H_6O_2H^+$ (hydroxyacetone)	Ubiquitous

The numbers in brackets after each *m/z* indicate the signal intensity in the unit vector normalized mass spectra for each component.

4.3.6.1 August 7, 2018 Component 2 of 8 – Anthropogenic hydrocarbons

Component 2 has been attributed to anthropogenic sources of hydrocarbons, generally associated with vehicle traffic. The measured VOCs contributing to this component include propene (*m/z* 43.05, $C_3H_6H^+$, also possible fragment ion from hydrocarbons), acetaladehyde (*m/z* 45.03, $C_2H_4OH^+$), butenes (*m/z* 57.07, $C_4H_8H^+$), benzene (*m/z* 79.05, $C_6H_6H^+$), toluene (*m/z* 93.06, $C_7H_8H^+$), ethylbenzene/xylenes (*m/z* 107.08, $C_8H_{10}H^+$), as well as other hydrocarbon fragments (*m/z* 41.03, $C_3H_5^+$). As shown in the annotated time series in Figure 4.27, this component contains many sharp concentration excursions, measured near intersections and high traffic areas, and is also observed when driving in close proximity to vehicles. Port Alberni is a small city with peak traffic at a busy intersection (Redford St at 3rd Avenue) reaching 1200–1600 vehicles (all types) per hour.[232] The map of this component is shown in Figure 4.28, and many of the high signal intensities that were measured from this component are located along main roads and at busy intersections. The compounds associated with this

component are a series of hydrocarbons, plus several oxygenated species (acetaldehyde, methanol, acetone) that are present in vehicle emissions.[46]

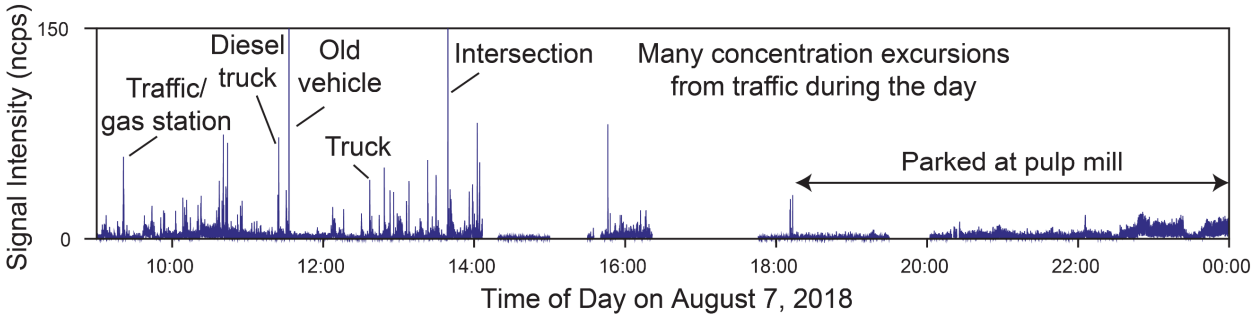


Figure 4.27: Annotated time series of the anthropogenic hydrocarbon component (Component 2 of 8) in the MCR-WALS analysis of the August 7, 2018 data.

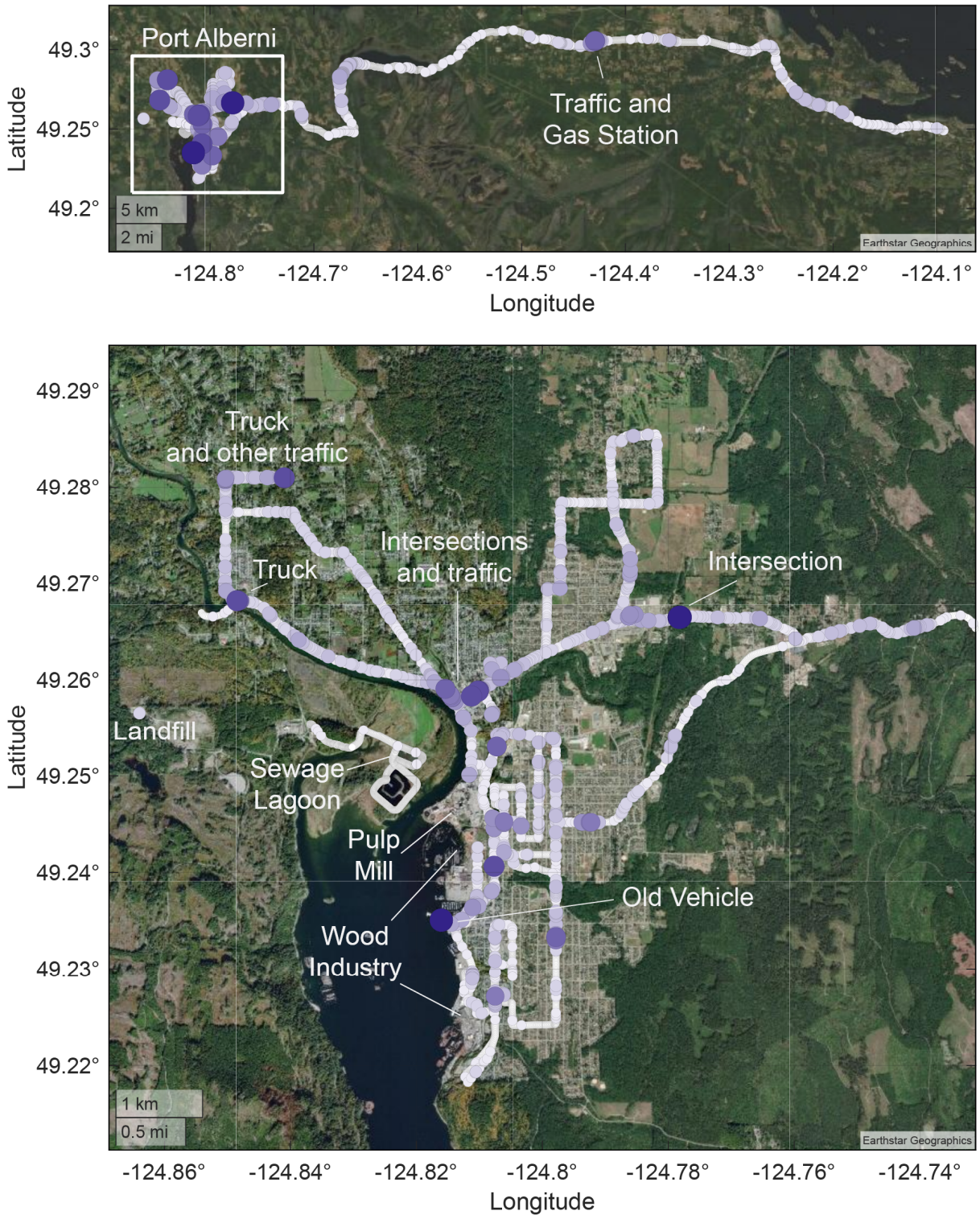


Figure 4.28: Geospatial distribution of the anthropogenic hydrocarbon component (Component 2 of 8) in the MCR-WALS model of the August 7, 2018 data. Larger and darker dots indicate higher signal intensities. The top map shows the whole drive route, and the bottom map centers on Port Alberni. The concentration excursions observed in this component are associated with intersection and higher traffic areas.

On-board sensors measuring carbon dioxide and nitrogen oxides were also in operation during the field campaign and examining the data collected by these sensors aids in validating the assignment of this component as anthropogenic hydrcarbon emissions. An overlay of the time series of the CO_2 concentration and the anthropogenic hydrocarbon component is shown in Figure 4.29, and the time series of the NO_x concentration and the anthropogenic hydrocarbon component is shown in Figure 4.30. While there are multiple sources of carbon dioxide in the area, not all of which are due internal combustion engines (*e.g.*, industrial emissions), many of the concentration excursions observed in the time series for this VOC component measured near vehicles and at intersections coincide with short duration, high (> 440 ppm) concentrations of carbon dioxide measured. The same trend is seen in the NO_x and anthropogenic hydrocarbon overlay with many of the high NO_x concentrations measured coinciding with high concentrations of hydrocarbons. In addition the time series overlay, scatter plots of the measured concentrations of CO_2 and NO_x versus the signal intensity meaured on the hydrocarbon component are presented. These scatter plots do not show a high correlation between the measurements, this may be due to many factors such as: short duration concentration excursions with measurement times that do not exactly coincide despite efforts to account for the sampling time of the different instruments, catalytic converters reduce the concentration of VOCs and NO_x emitted from vehicles but not CO_2 concentrations, the type of fuel (*e.g.*, gasoline, diesel) and type of vehicle (*e.g.*, passenger, heavy-duty vehicle) impacts ratios of pollutants being emitted, and the concentrations of VOCs measured on road are much lower than NO_x and CO_2 and disperse to below our detection limit more quickly.

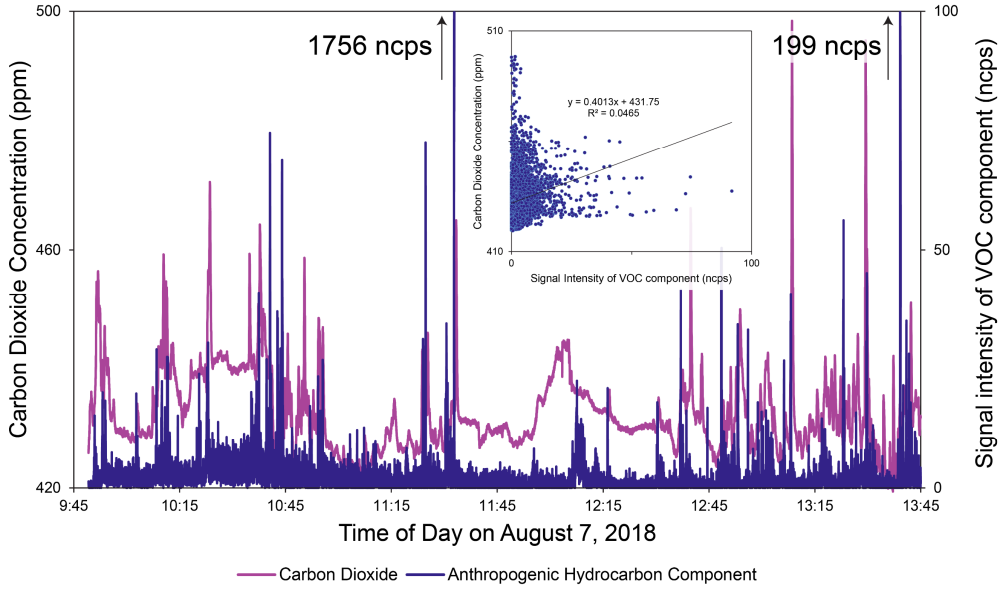


Figure 4.29: Time series overlay of the signal intensity of the anthropogenic hydrocarbon component, and concentration of CO_2 measured between 9:45–13:45 on August 7, 2018 in Port Alberni. Many of the concentration excursions observed in the anthropogenic hydrocarbon component correspond to higher concentrations of CO_2 being measured. Inset: Scatter plot of the two data sets. The very high signal intensities (off scale) observed in the anthropogenic hydrocarbon component have been removed from the scatter plot.

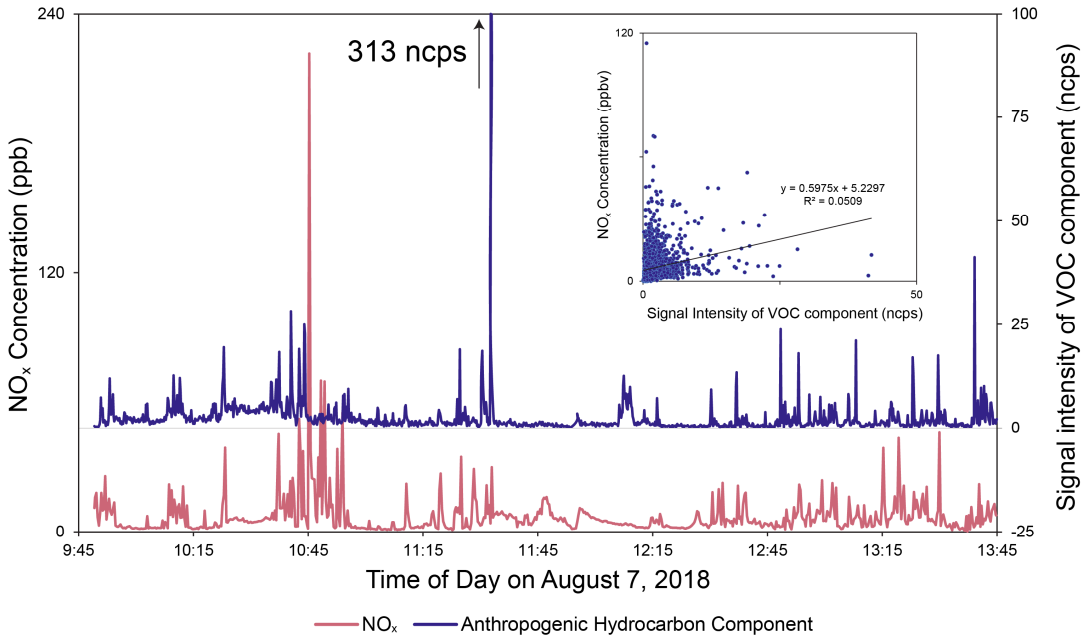


Figure 4.30: Time series overlay of the signal intensity of the anthropogenic hydrocarbon component, and concentration of NO_x measured between 9:45–13:45 on August 7, 2018 in Port Alberni. Many of the concentration excursions observed in the anthropogenic hydrocarbon component correspond to higher concentrations of NO_x being measured. NO_x measurements were made at 10 second intervals, and 10 second sample averages of the anthropogenic hydrocarbon measurements were calculated for this analysis. Inset: Scatter plot of the two data sets. The very high signal intensity (off scale) observed in the anthropogenic hydrocarbon component has been removed from the scatter plot.

The chemical profiles of exhaust from gasoline and diesel vehicles differ, both in the distribution of the volatility of the compounds emitted (organic emissions in gasoline exhaust are >90 % VOCs, while diesel emissions are about 50 % SVOCs), and the classes of compounds emitted. Diesel exhaust contains a higher proportion of alkenes and carbonyl species than gasoline exhaust, while gasoline exhaust is richer in alkanes (branched and straight chain) and aromatic species.[233] As the proton affinity of light straight chain alkanes ($<C_7$) are less than H_3O^+, they are not detected by the PTR-ToF-MS employed here. Despite these differences, the chemical compounds present in the resolved mass spectrum are present in exhaust from both gasoline vehicles (*e.g.*, butenes, toluene, benzene, ethylbenzene, acetaldehyde) (SPECIATE profile 8750a Gasoline Exhaust – Reformulated Gasoline) and diesel vehicles (*e.g.,* butenes, toluene, benzene, ethylbenzene, xylenes, C9-aromatics, acetaldehyde, acetone) (SPECIATE profiles 2514 Vehicle Exhaust Callahan Tunnel, 95335a Diesel Exhaust – Heavy Duty Truck 2011 corrected, 4758 Diesel Exhaust – Light Duty Truck at 22.6 °C Composite).[46] While the level of resolution of this component does not allow for the differentiation of different types of vehicles (*e.g.*, gasoline versus diesel, regular passenger vehicles versus trucks), it does illustrate the high temporal and spatial variability in the presence of hydrocarbon species as well as their prevalence in high traffic areas. Concentrations of VOCs near roadways have been shown to drop exponentially with distance[234], and

more targeted field campaign focused on traffic emissions at busy intersection would be well suited to differentiate the emissions from the different types of vehicle traffic and their impact on the nearby area.

4.3.6.2 August 7, 2018 Component 3 of 8 – Forest fire VOCs

The time series (Figure 4.31) for Component 3 suggests a diffuse geographic distribution rather than point sources due to the lack of fine structure related to sampling location or events. This component was primarily measured in the north-west region of Port Alberni (Figure 4.32), with contributions increasing at 10:20, reaching a peak around 10:50, and decreasing by 12:00. Field notes identified that at 10:22 the mobile lab descended into Port Alberni and entered a visible smoke plume present due to nearby forest fires. The mobile lab was also equipped with a roof-mounted OPS used to measure particulate matter, and the $PM_{2.5}$ concentration measurements from the OPS correlate with the time series of Component 3 as shown in the time series overlay and scatter plot in Figure 4.33. Both the field notes noting the presence of smoke, and the independent measurements from the OPS allowed us to identify Component 3 as the VOCs accompanying the forest fire smoke.

The mass spectrum for this component includes signals from a range of VOCs including methanol (*m/z* 33.03), acetonitrile (*m/z* 42.03), acetaldehyde (*m/z* 45.03), acrolein (*m/z* 57.03), acetone (*m/z* 59.05), acetic acid (*m/z* 61.03 and fragment ion at *m/z* 43.01), and methyl ethyl ketone/butanal (*m/z* 73.06). In addition to the field notes and independent $PM_{2.5}$ measurements, some of the compounds contributing to the component (methanol, acetone, acetic acid, acetaldehyde, acetonitrile, acrolein, methyl ethyl ketone) have been reported in the literature to be VOCs found in forest fire smoke.[89,230,235,236]

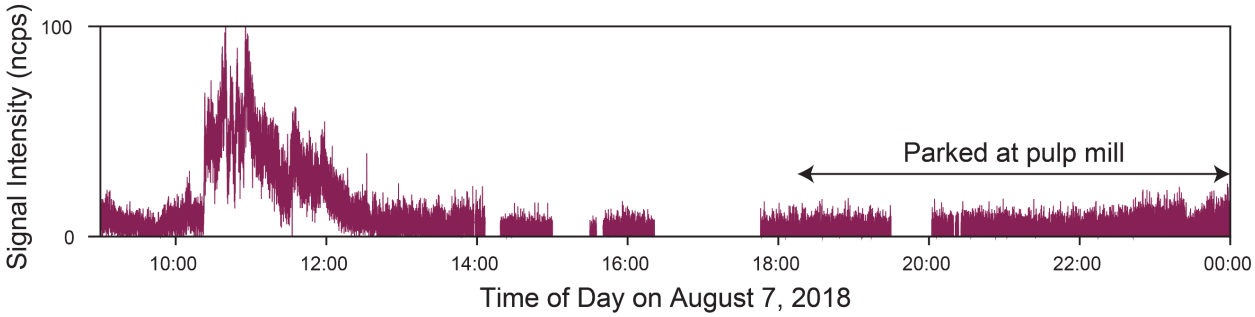


Figure 4.31: Annotated time series of the forest fire VOCs (Component 3 of 8) in the MCR-WALS analysis of the August 7, 2018 Port Alberni data.

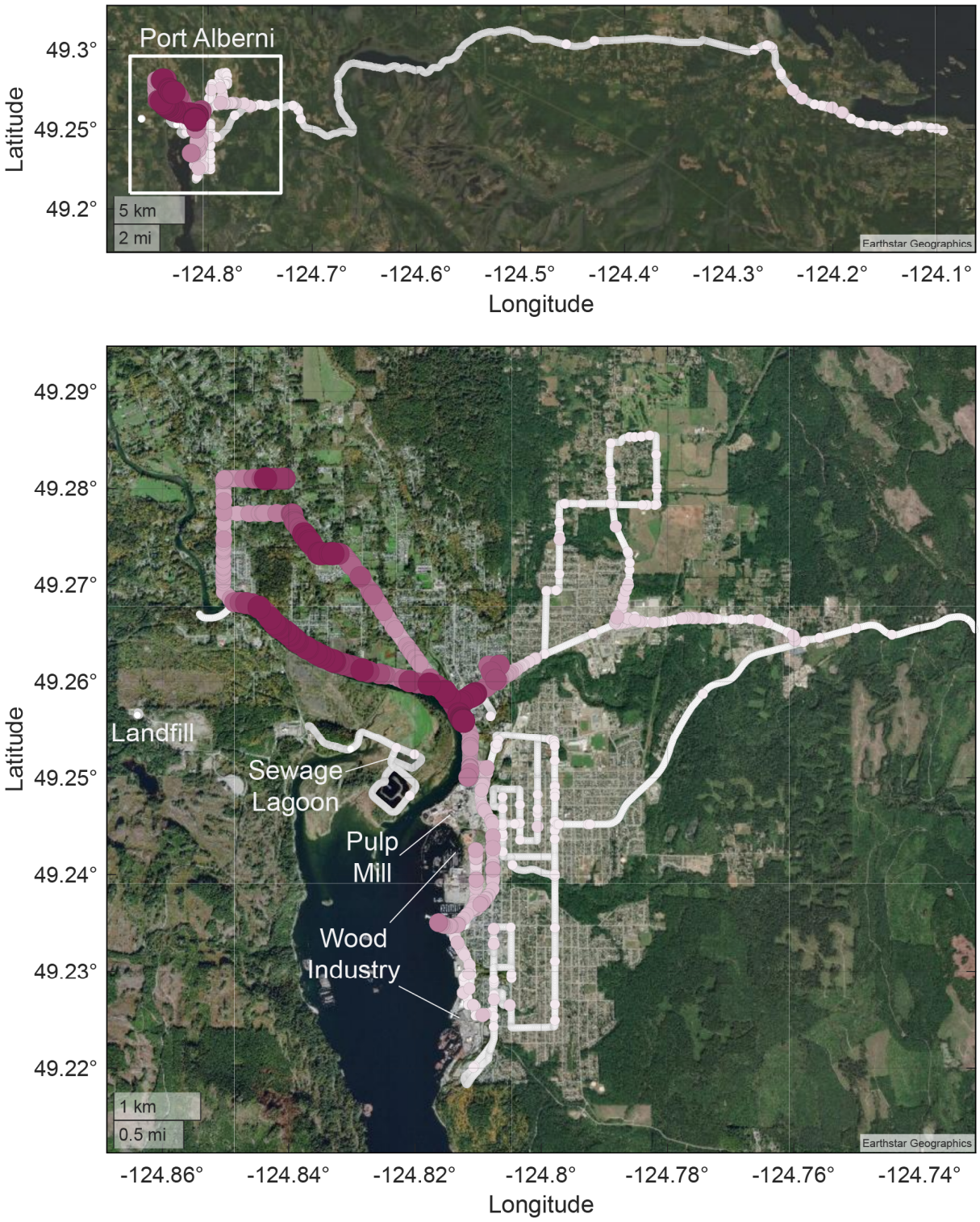


Figure 4.32: Geospatial distribution of the forest fire component (Component 3 of 8) in the MCR-WALS model of the August 7, 2018 data. Larger and darker dots indicate higher signal intensities. The top map shows the whole drive route, and the bottom map centers on Port Alberni.

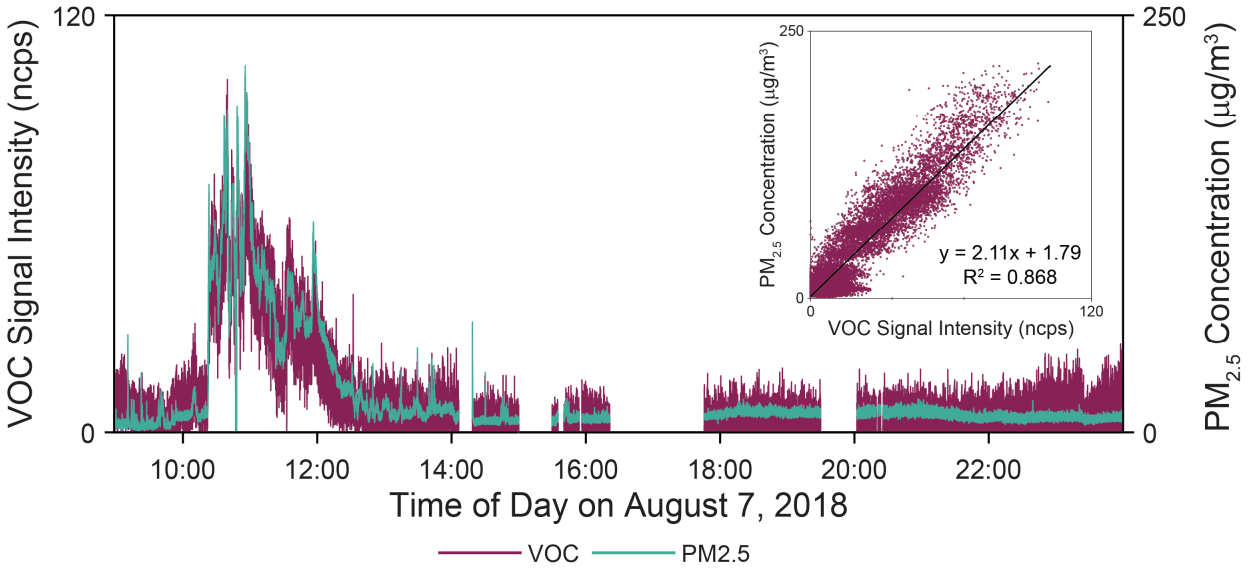


Figure 4.33: Time series overlay of the $PM_{2.5}$ measurement (teal) and Component 3 time series (burgundy) on August 7, 2018. Inset: Scatter plot of the same data.

4.3.6.3 August 7, 2018 Component 5 of 8 – Emissions from vegetation or wood products

Component 5 has been identified as the VOC emissions from vegetation or wood products, and concentration excursions were recorded throughout the sampling day as shown in the annotated time series in Figure 4.34. This component was observed in multiple locations, both rural and urban/industrial, as shown in the map in Figure 4.35. The mass spectrum for this component includes measurements associated with methanol (*m/z* 33.03), acetaldehyde (*m/z* 45.03), acetone (*m/z* 59.05), acetic acid (*m/z* 61.03 and fragment at *m/z* 43.01), and monoterpenes (*m/z* 137.12 and fragment ion at *m/z* 81.07). The drive to Port Alberni is on a highway that travels through an old growth forest of Douglas Fir and elevated signals of this component were observed while driving through, and when parked in, the forest between 9:30–9:45. Douglas Fir trees emit relatively high concentrations of monoterpenes and little isoprene, which is seen in this component,[237] and many of the other chemical species observed, such as acetaldehyde, acetone, and methanol, have been reported in elevated concentrations in Douglas Fir forests[237] and as making up a significant portion of biogenic VOC emissions globally.[48]

Once in Port Alberni, signals related to this component were observed while sampling near lumber yards, wood chip and forest residue piles (used in pulp production and/or as fuel), and while driving near vehicles carrying wood chips or yard waste. While the species of wood in the lumber yards and wood chips were not identified, many studies have characterized VOC emissions from wood chips and forest residue. Wood chips made from Western Red Cedar have been shown to emit methanol, acetone, and aldehydes (among other species) when stored due to biological and chemical oxidation processes[238], and monoterpenes have been measured as emissions from wood chips in a variety of studies.[53,238,239]

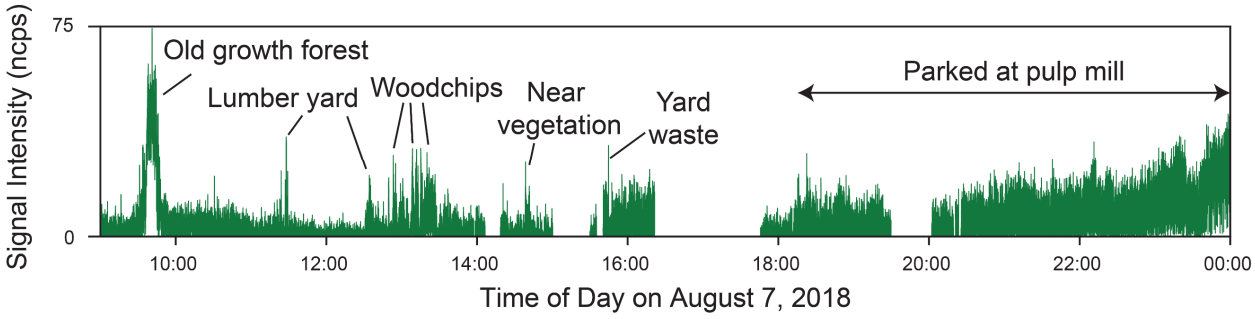


Figure 4.34: Annotated time series of the component associated with emissions from vegetation or wood products (Component 5 of 8) in the MCR-WALS analysis of the August 7, 2018 Port Alberni data.

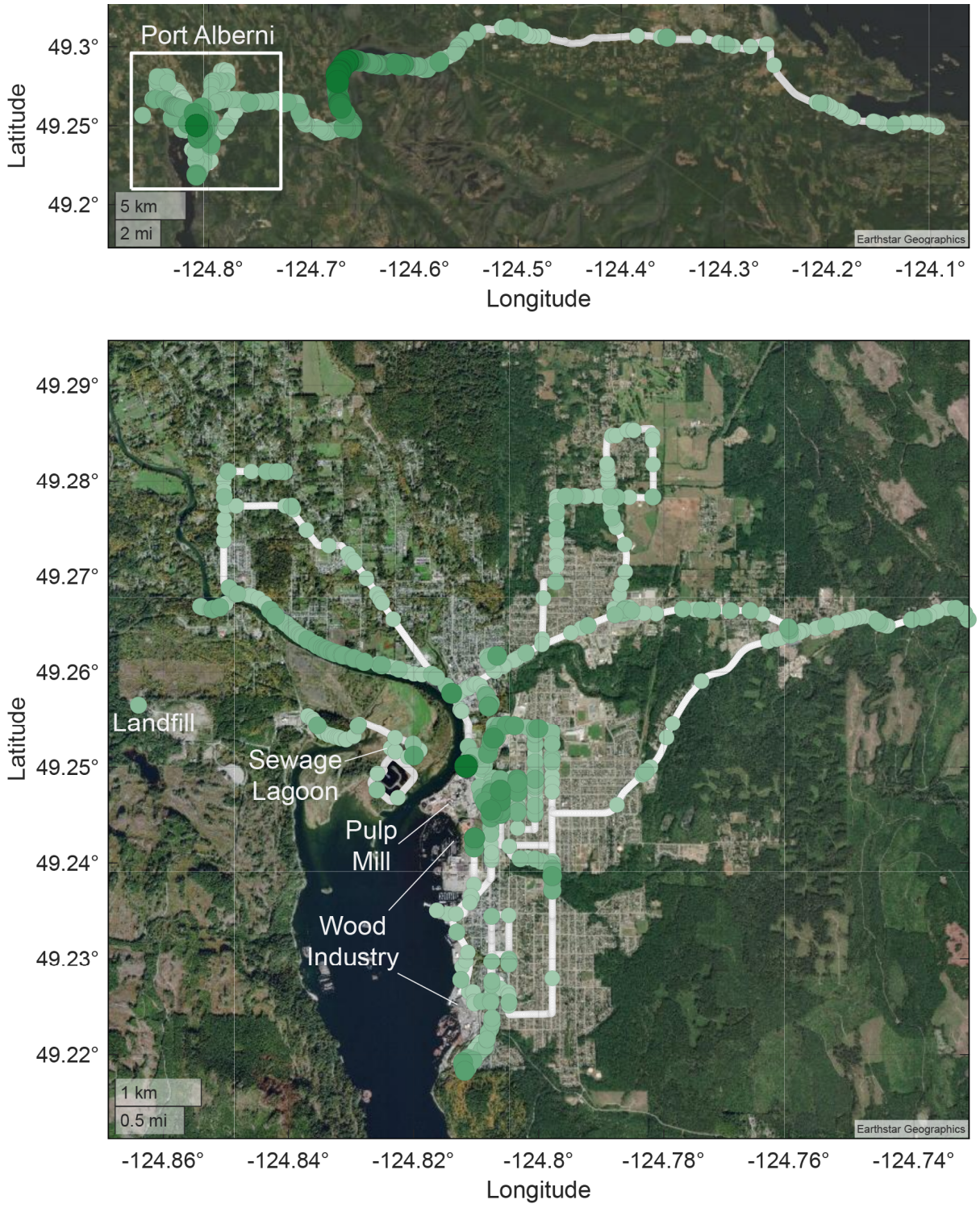


Figure 4.35: Geospatial distribution of the emissions from vegetation or wood products (Component 5 of 8) in the MCR-WALS model of the August 7, 2018 data. Larger and darker dots indicate higher signal intensities. The top map shows the whole drive route, and the bottom map centers on Port Alberni. High signal intensities for this component were measured while driving through an old growth forest, near many of the wood-based industries in Port Alberni, and near vehicles carrying wood products.

4.3.6.4 August 7, 2018 Component 6 of 8 – Biogenic emissions from vegetation

Component 6 is a second component related to VOC emissions from vegetation, but with a different temporal (Figure 4.36) and geographic (Figure 4.37) distribution than Component 5 and a mass spectrum with measured signal intensities for different biogenic VOCs. Component 5 was associated with monoterpene emissions, while Component 6 is associated with isoprene emissions. The major ions recorded in the mass spectrum for Component 6 are for formic acid (*m/z* 47.01), acetone (*m/z* 59.05), isoprene (*m/z* 69.07), and hydrocarbon fragment ions (*m/z* 41.03). This component was measured mainly while driving on the East bank of the Somass River, on the forested drive enroute to the sewage lagoon, and on the access roads around the sewage lagoon and unused wastewater lagoon, all of which are located in the central and northern region of the Somass River Estuary. The estuary is home to a wide variety of habitats (*e.g.*, tidal flats, tidal channels, levee forest, riparian forest, tidal marsh), and include a large number of non-coniferous plant species which emit higher concentrations of isoprene compared to monoterpenes.[17,240,241] When driving around the lagoons, the smaller of which was in use for city sewage treatment, measurements associated with this component were measured on sides of the lagoons that were near the marsh land on the edge of the tidal flats, and not on the more inland sections of the drive. Wind measurements recorded at two meteorological stations (Alberni Elementary (white star) and Ocean Networks Canada (blue star)) indicate the wind was coming up the inlet from the South while sampling at the lagoons. This VOC component was observed when we were sampling upwind of the sewage lagoon, while high concentrations of methane were observed while sampling downwind as shown in Figure 4.38.

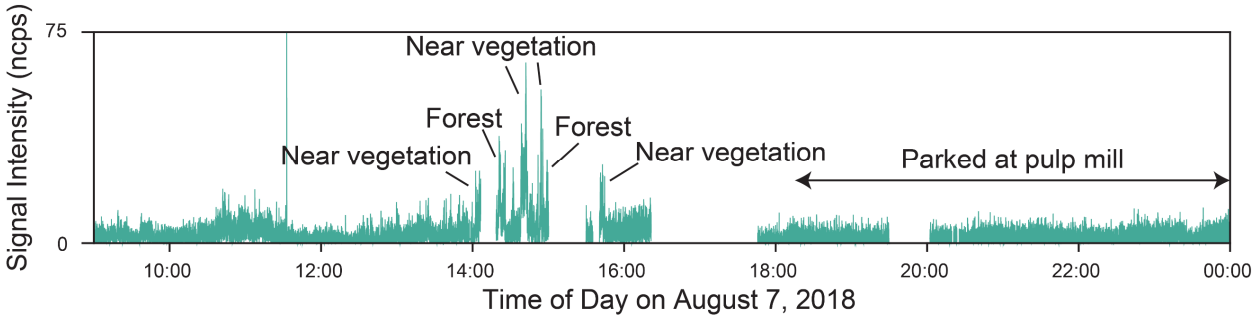


Figure 4.36: Annotated time series of the biogenic emissions from vegetation (Component 6 of 8) in the eight component MCR-WALS analysis of the August 7, 2018 Port Alberni data.

The signal at ~11:30 am is not attributed to vegetation, but instead to a vehicle in disrepair that drove near the mobile lab. The vehicle was emitting a large variety of VOCs at high signal intensity, including compounds that are isobaric with isoprene, but not biogenic in origin, and the measurements have been split across multiple components.

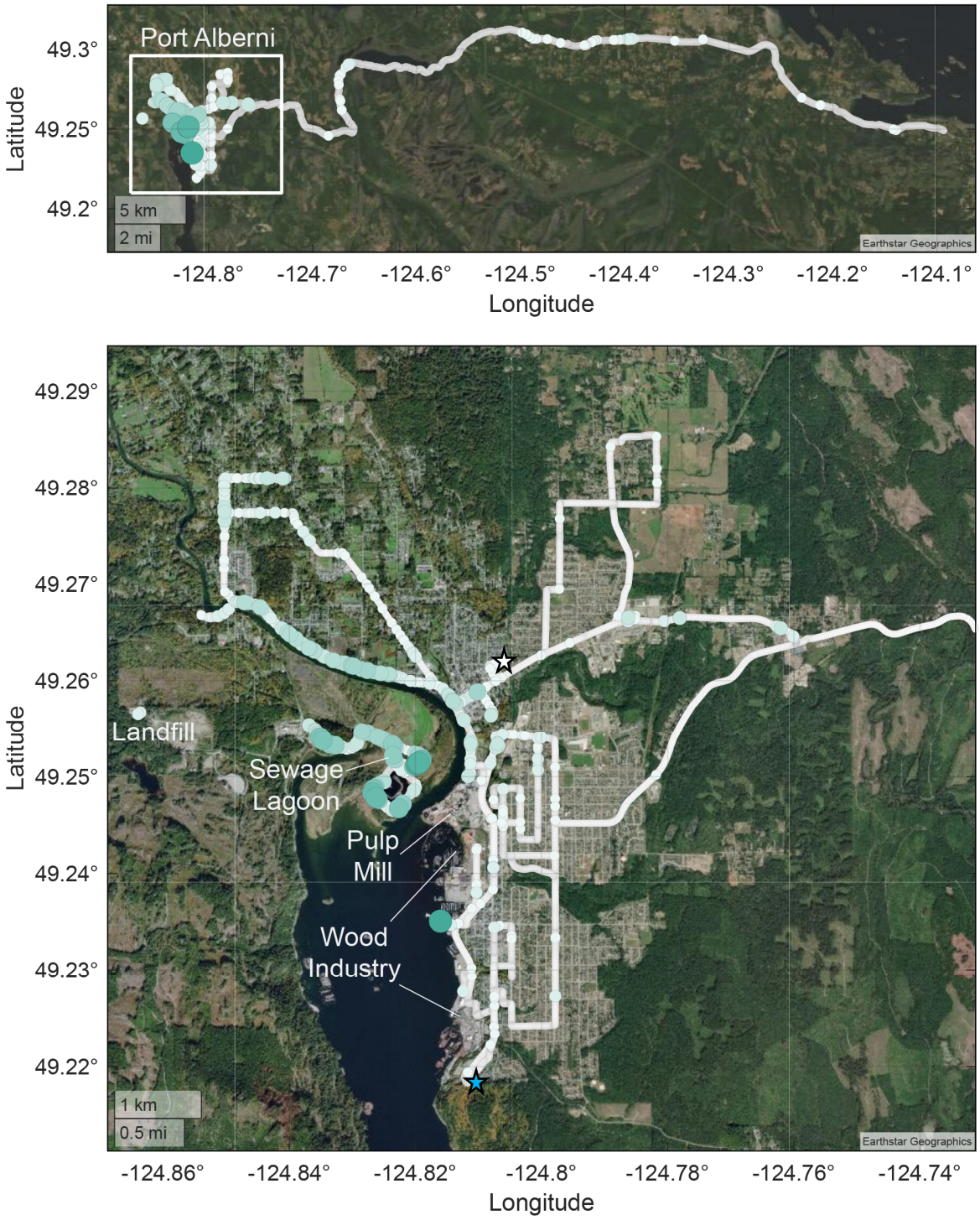


Figure 4.37: Geospatial distribution of the biogenic emissions from vegetation (Component 6 of 8) in the MCR-WALS model of the August 7, 2018 data. Larger and darker dots indicate higher signal intensities. The top map shows the whole drive route, and the bottom map centers on Port Alberni. This component was mainly measured while sampling on Somass River Estuary, and while upwind of the sewage lagoon on the estuary. The white dot indicates the location of the Port Alberni Elementary Air Quality monitoring station (hourly wind speed and direction, and $PM_{2.5}$ measurements). The blue star is the location of the Ocean Networks Canada meteorological station (wind speed and direction).

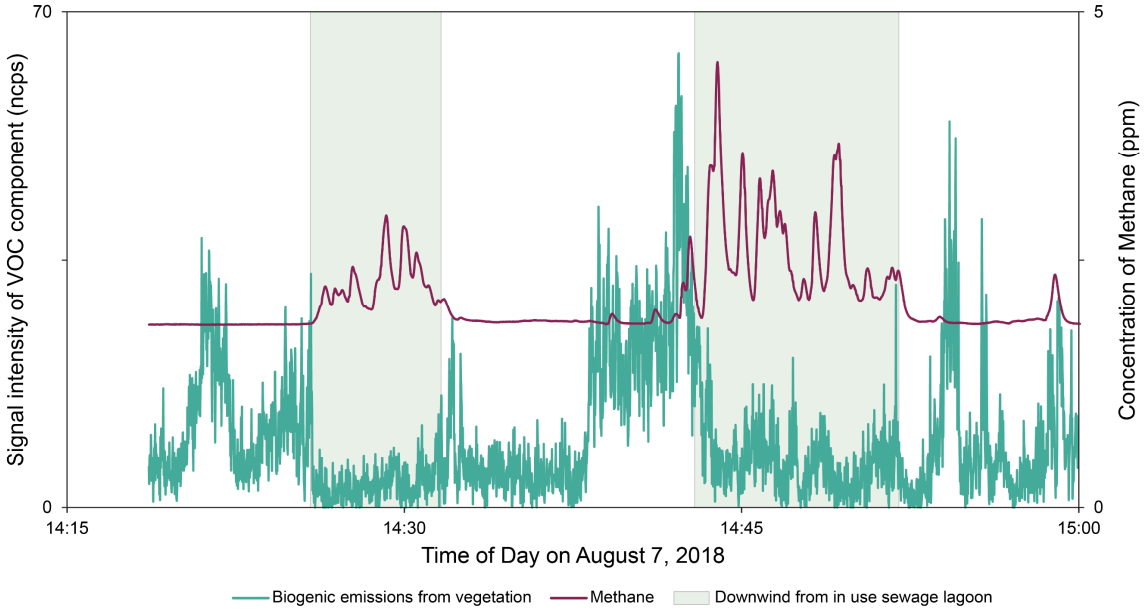


Figure 4.38: Overlay of the time series for the biogenic emissions from vegetation measurements (teal) and measured methane concentrations (burgundy) measured between 14:20–15:00 on August 7, 2018 near the sewage lagoon and unused wastewater lagoon in Port Alberni. The light green boxes indicate when sampling occurred downwind from the sewage lagoon.

4.3.6.5 August 7, 2018 Component 7 of 8 – Small oxygenated species

The mass spectrum for MCR-WALS Component 7 has major signals attributed to methanol, (*m/z* 33.03), formic acid (*m/z* 47.01), acetone (*m/z* 59.05), and acetic acid (*m/z* 61.03). This component is present at fairly constant levels throughout the sampling area, with notable concentration excursions at 9:21, 12:52, and 13:06 as shown in the time series in Figure 4.39. The concentration excursion at 9:21 occurred while sampling in traffic in an area that includes an auto salvager and shipwright, while the excursions at 12:52 and 13:06 occurred while in the vicinity of boat repair and compressed gas companies. While this component has been labelled as "small oxygenated species" it may be related to industrial manufacturing and solvent use at the nearby commercial locations. Previous studies have resolved an industrial component containing high abundances of acetone.[89,242]

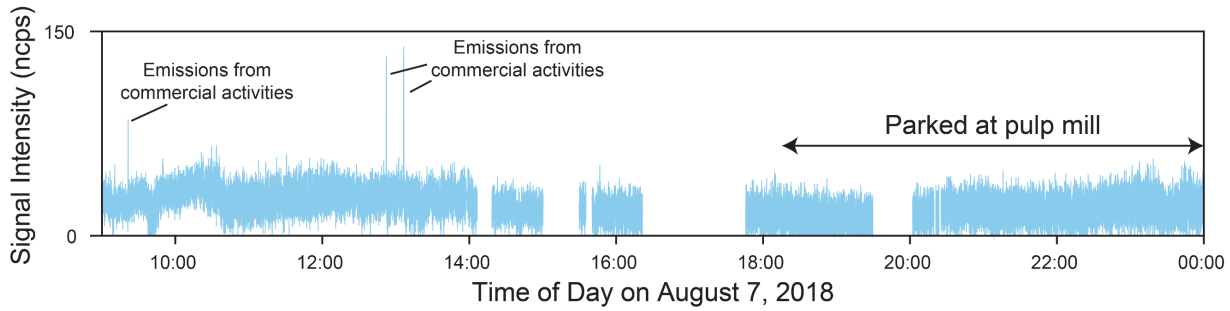


Figure 4.39: Annotated time series of the small oxygenated species (Component 7 of 8) in the eight component MCR-WALS analysis of the August 7, 2018 Port Alberni data.

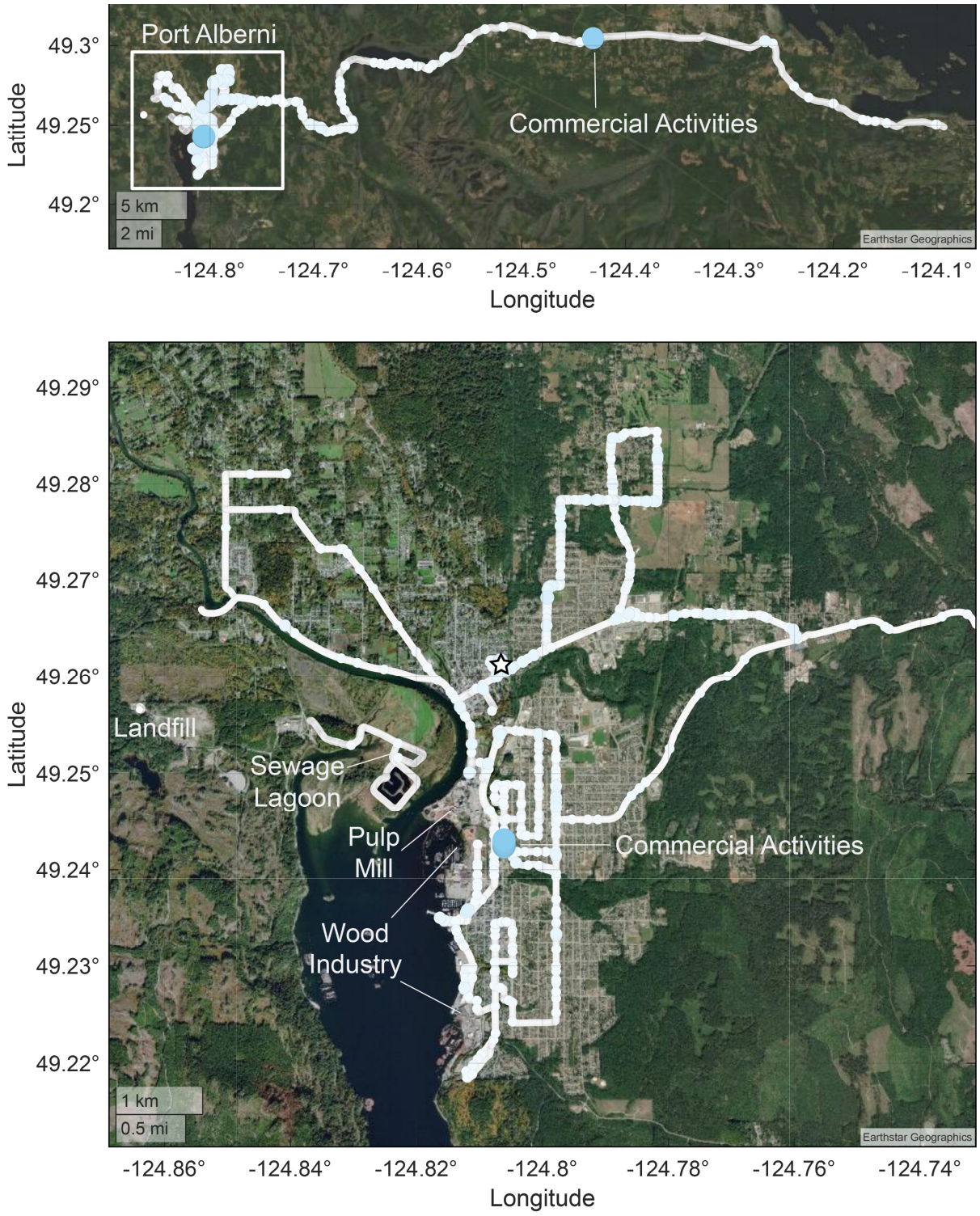


Figure 4.40: Geospatial distribution of the biogenic emissions from small oxygenated species (Component 7 of 8) in the MCR-WALS model of the August 7, 2018 data. Larger and darker dots indicate higher signal intensities. The top map shows the whole drive route, and the bottom map centers on Port Alberni. This component was measured near commercial areas on the drive route.

4.3.6.6 August 7, 2018 Component 1 of 8 – Methanol

The mass spectrum for Component 1 contains only one signal, *m/z* 33.03 (methanol). This component is measured throughout most of the day. Many short concentration excursions were measured as shown in Figure 4.41 (*e.g.,* before 10:00 and between 15:30–16:30), and are related to various point sources noted in the field notes. For example, three commercial nurseries/greenhouses on the drive to Port Alberni, driving past or near larger vehicles, intersections, and the residential garbage drop off area at the landfill. Elevated signals over longer time periods for this component coincide with field notes indicating the presence of an odor we attributed to the local pulp mill (13:00–13:30, 15:45–16:30) and while parked at the pulp mill overnight (after 18:00). Elevated methanol signals were also observed on the pulp mill property near the primary clarifier on another sampling day (data not shown). After methane, methanol is the second most abundant atmospheric organic gas, with many biogenic sources[243], but it has also been measured in stack and fugitive emissions from pulp and paper operations[46,244], as a fugitive emissions from landfills[245], and in diesel exhaust.[46] The map of the methanol component is shown in Figure 4.42, and there is a larger distribution of high signal intensity measurements for the component in the center of the map in the areas at and around the pulp mill. While methanol also contributes the mass spectrum of other components (*e.g.,* forest fire VOCs, emissions from vegetation or wood products) the time series in this component captures the changes in methanol signal that are not accounted for in the other components, and unique to the temporal profile of the methanol measurements.

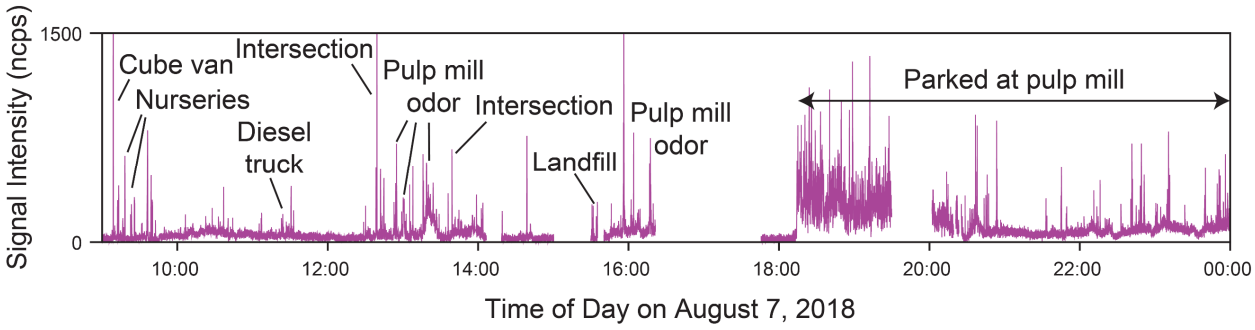


Figure 4.41: Annotated time series of the methanol component (Component 1 of 8) in the eight component MCR-WALS analysis of the August 7, 2018 Port Alberni data.

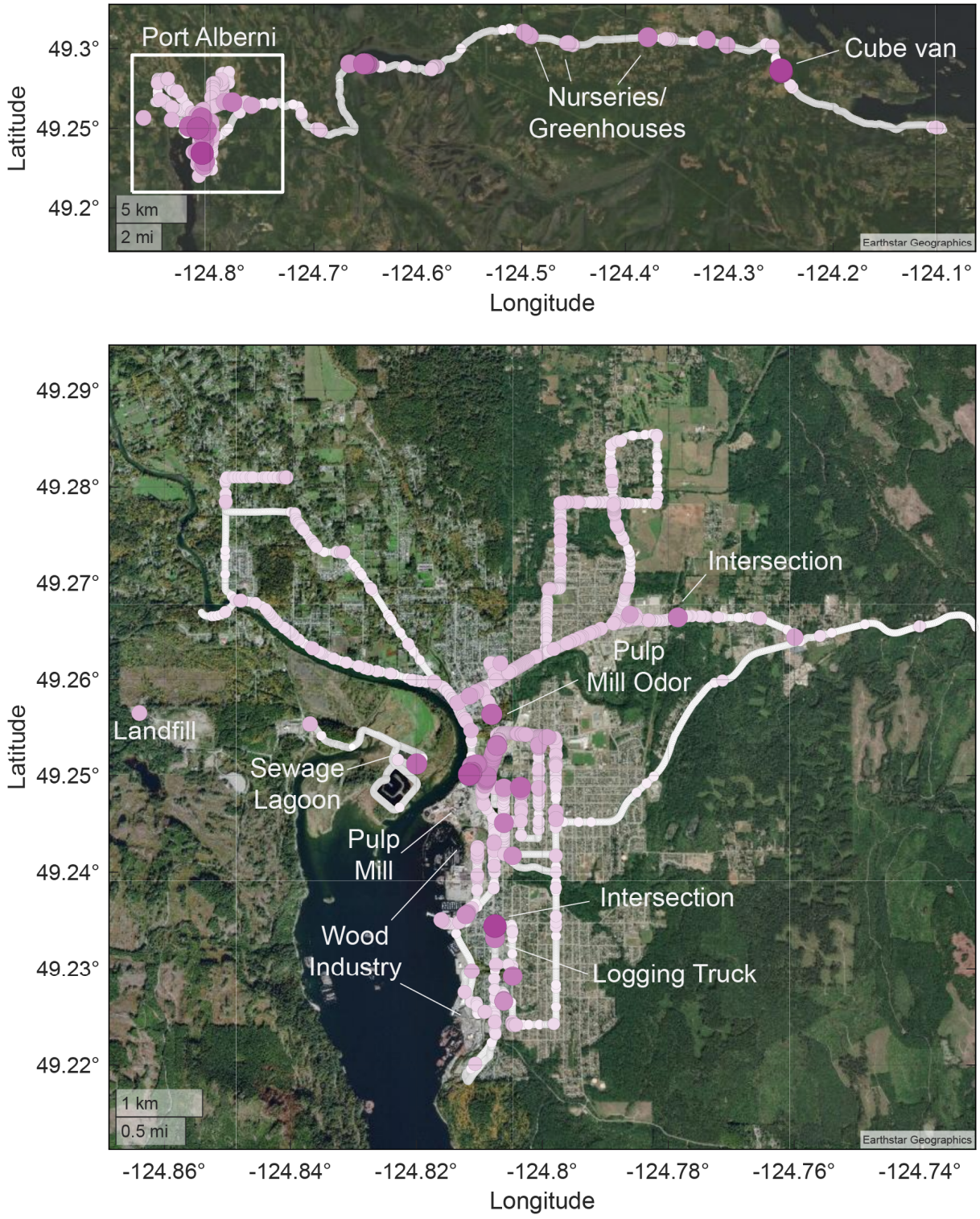


Figure 4.42: Geospatial distribution of the methanol component (Component 1 of 8) in the MCR-WALS model of the August 7, 2018 data. Larger and darker dots indicate higher signal intensities. The top map shows the whole drive route, and the bottom map centers on Port Alberni. This component was measured across the sampling area.

4.3.6.7 August 7, 2018 Component 4 of 8 – Acetaldehyde

The mass spectrum for Component 4 contains mainly the ion due to acetaldehyde (*m/z* 45.03) and was measured in many locations throughout the community as shown in Figure 4.43 and Figure 4.44. This includes near wood-based industries such as the pulp mill and sawmills, with several sharp concentration excursions coinciding with field notes indicating the presence of nearby vehicles, in particular when following large trucks. Acetaldehyde can be emitted from pulp mills[46,244], forest products[46], and is also found in some vehicle exhaust.[46] As with the methanol component, acetaldehyde measurements are present in the mass spectra for multiple components (*e.g.*, anthropogenic hydrocarbons, forest fire VOCs, emissions from vegetation or wood products), and this component captures the additional acetaldehyde measurements not associated with those components.

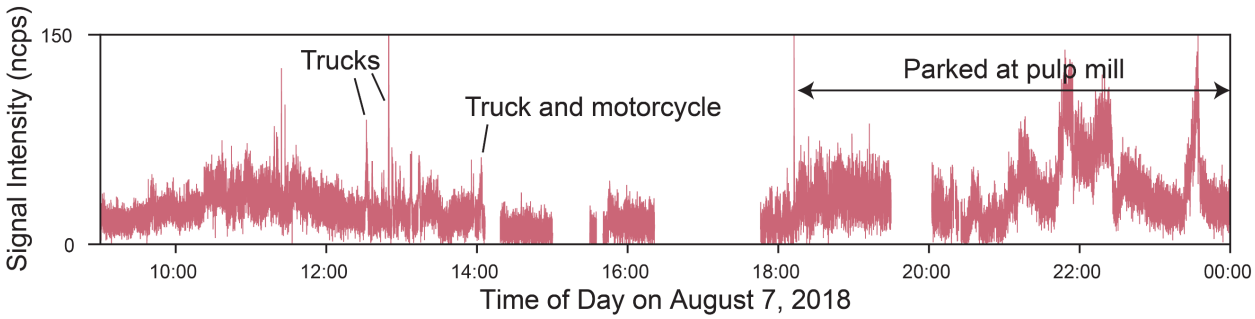


Figure 4.43: Annotated time series of the acetaldehyde component (Component 4 of 8) in the eight component MCR-WALS analysis of the August 7, 2018 Port Alberni data.

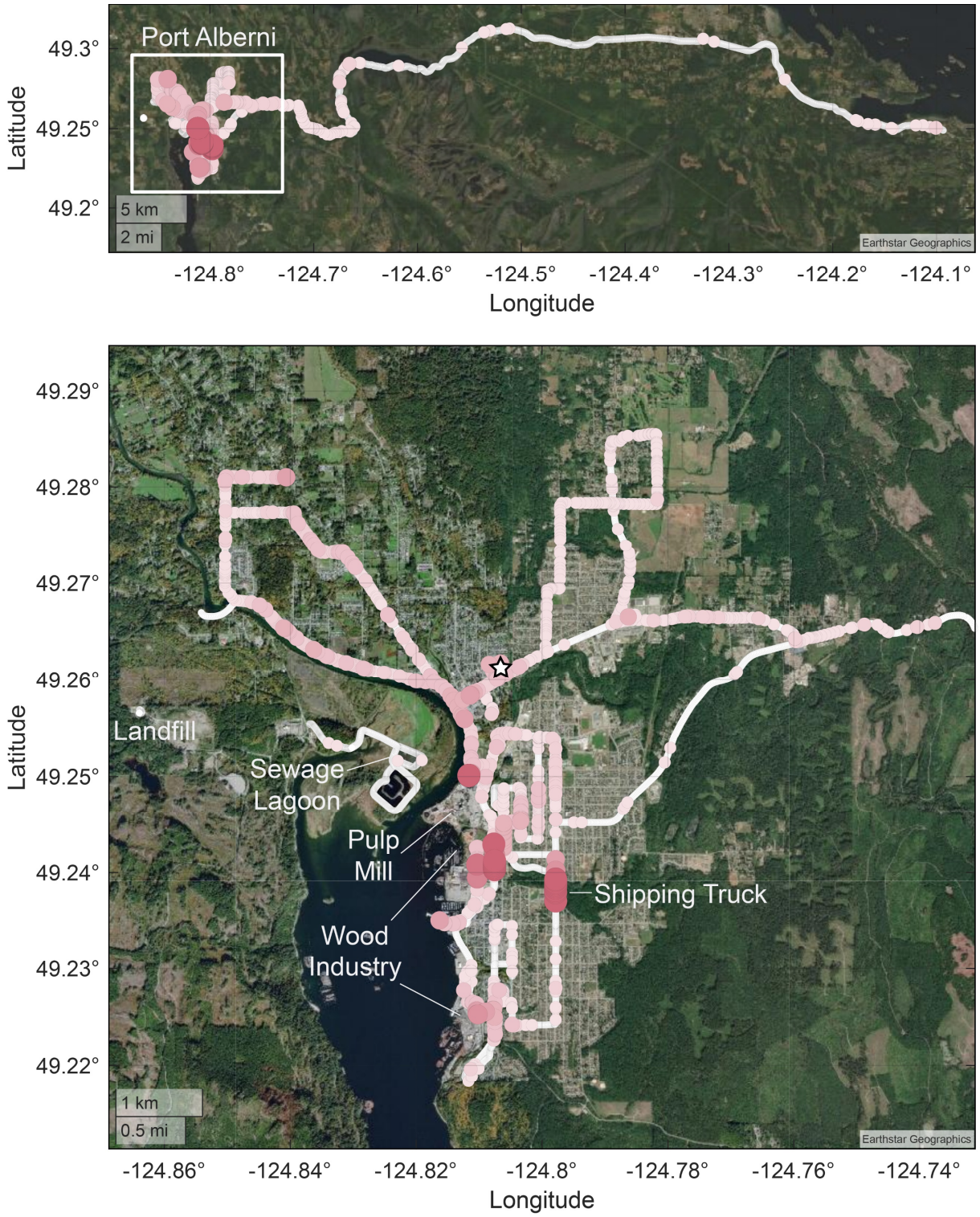


Figure 4.44: Geospatial distribution of the acetaldehyde component (Component 4 of 8) in the MCR-WALS model of the August 7, 2018 data. Larger and darker dots indicate higher signal intensities. The top map shows the whole drive route, and the bottom map centers on Port Alberni. This component was measured across the sampling area, and in particular at the pulp mill and while following some large trucks.

4.3.6.8 August 7, 2018 Component 8 of 8 – Other

Component 8 has a time series (Figure 4.45) with little geographic (Figure 4.46) or temporal features that aid in its identification, even when taken into consideration with data from the other on-board sensors (CO_2, CH_4, NO_x, O_3, PM). The mass spectra for the component contains peaks associated with methanol (*m/z* 33.03), acetone (*m/z* 59.05), acetic acid (*m/z* 61.03 and fragment at *m/z* 47.01), and alkyl fragment (*m/z* 43.01), and small amounts of methyl ethyl ketone (*m/z* 73.06) and hydroxyacetone (oxidation product of acetone, *m/z* 75.04) which is a similar set of compounds to Component 7. This component could be associated with background concentrations[47,246], VOCs that are formed through atmospheric transformations[247] with a component having a similar mass spectrum being identified as an aged plume by Stojić et al[141], be associated with inlet/sampling line effects, or be a degenerate component with Component 7 as they have very similar resolved mass spectra.

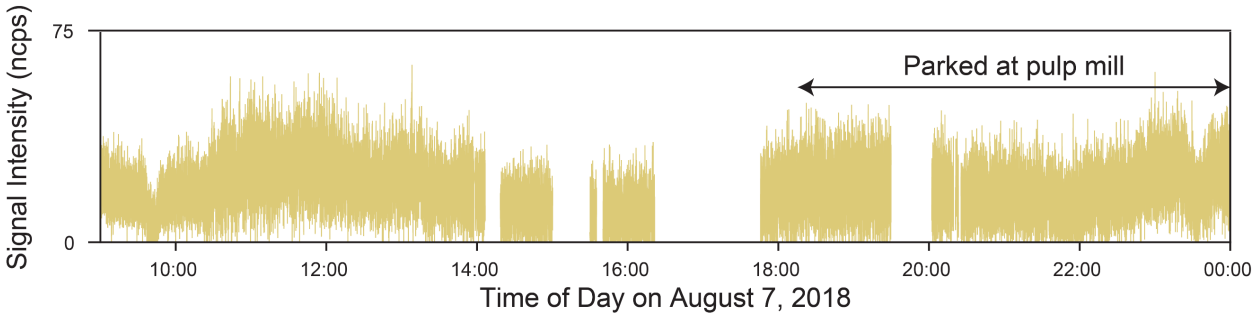


Figure 4.45: Annotated time series of the other component (Component 8 of 8) in the eight component MCR-WALS analysis of the August 7, 2018 Port Alberni data.

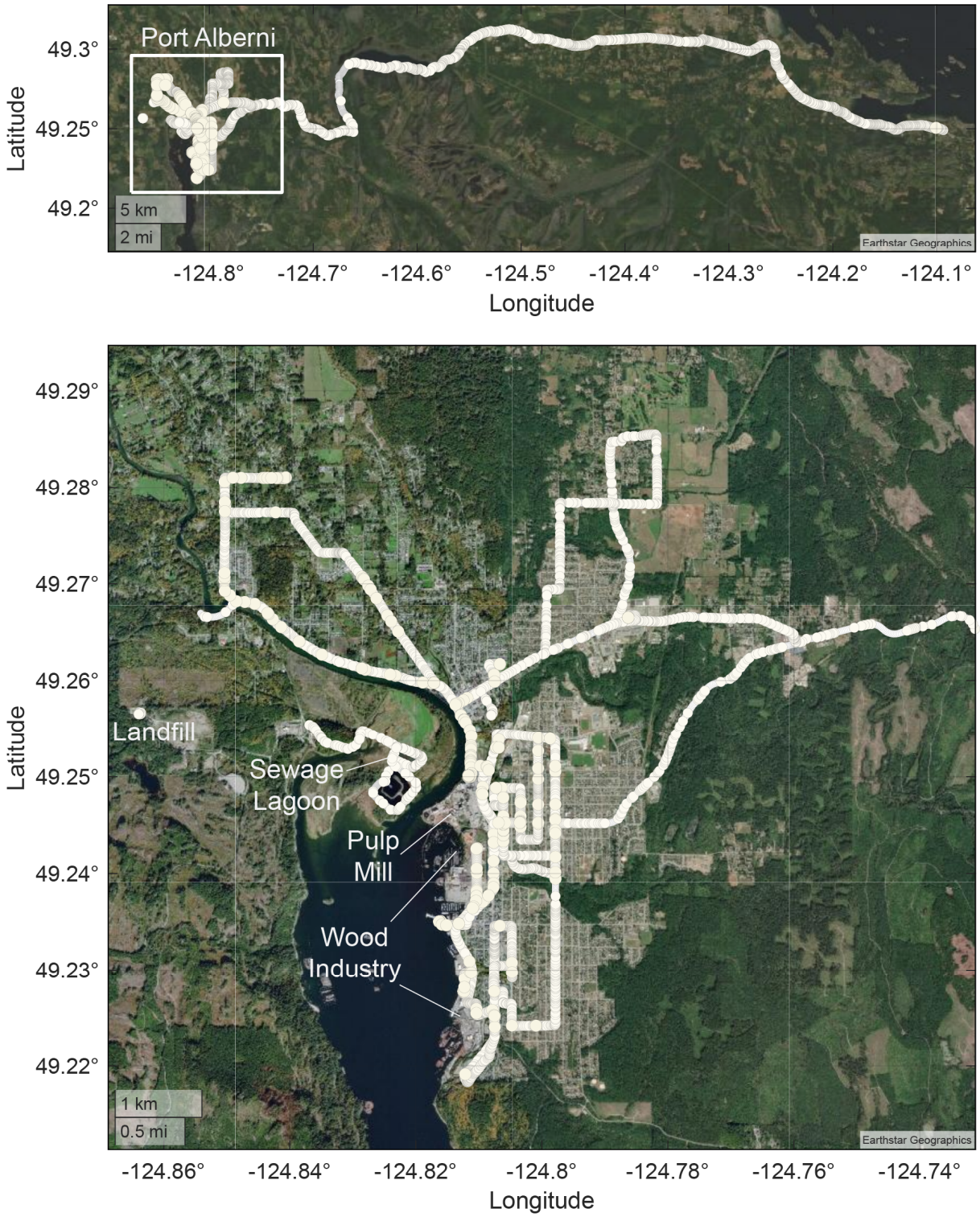


Figure 4.46: Geospatial distribution of the other component (Component 8 of 8) in the MCR-WALS model of the August 7, 2018 data. Larger and darker dots indicate higher signal intensities. The top map shows the whole drive route, and the bottom map centers on Port Alberni. This component was measured across the sampling area without much difference in the observed signal intensities.

4.3.6.9 August 7, 2018 source apportionment

Based on the eight component MCR-WALS model discussed above, the VOC sources encountered on August 7, 2018 were apportioned using Equation 4.9 and Equation 4.10. The TIC of the *m/z* included in the analysis was the total signal apportioned between the eight components. In the top panel of Figure 4.47 the TIC is shown in blue, with a reflection of the reconstructed TIC from the apportionment of the eight sources shown in burgundy. The bottom panel shows the residuals between the measured and reconstructed TIC, and with the exception of several short, high concentration excursions, the majority of the signal intensity measured in the TIC has been accounted for in the apportionment calculation.

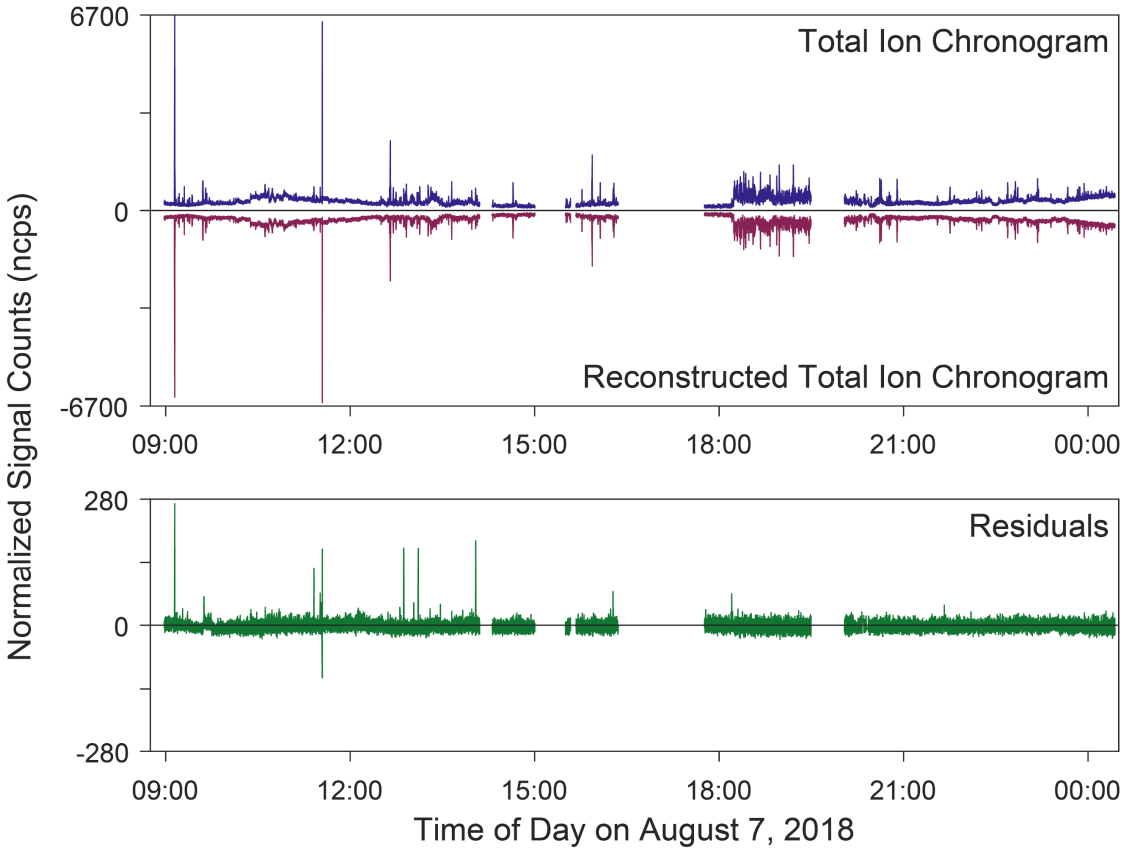


Figure 4.47: Top: The TIC of the *m/z* included in the MCR-WALS analysis is shown in the blue time series. This time series was used to apportion the VOCs observed on August 7, 2018 between the eight MCR-WALS components. The reflected time series shown in burgundy is the reconstructed TIC after the apportionment. Bottom: The residuals between the observations and apportionment of the signal measured on August 7, 2018.

A pie chart showing the percent contributions of each of the eight components over the course of the day is shown in Figure 4.48. The components that are mainly measured in short concentration excursions (*e.g.*, anthropogenic hydrocarbons, emissions from vegetation or wood products, biogenic emissions from vegetation) make up the smallest proportion of observed VOCs, 2 %, 6 %, and 3 % respectively. VOCs from forest fire smoke make up 10 % of the observed signal, with the methanol and acetaldehyde components contributing 29 % and

13 % respectively. The small oxygenated species and other component make up a large share of the signal (20 % and 17 %), and may be overrepresented as they have high background contributions relative to the other components.

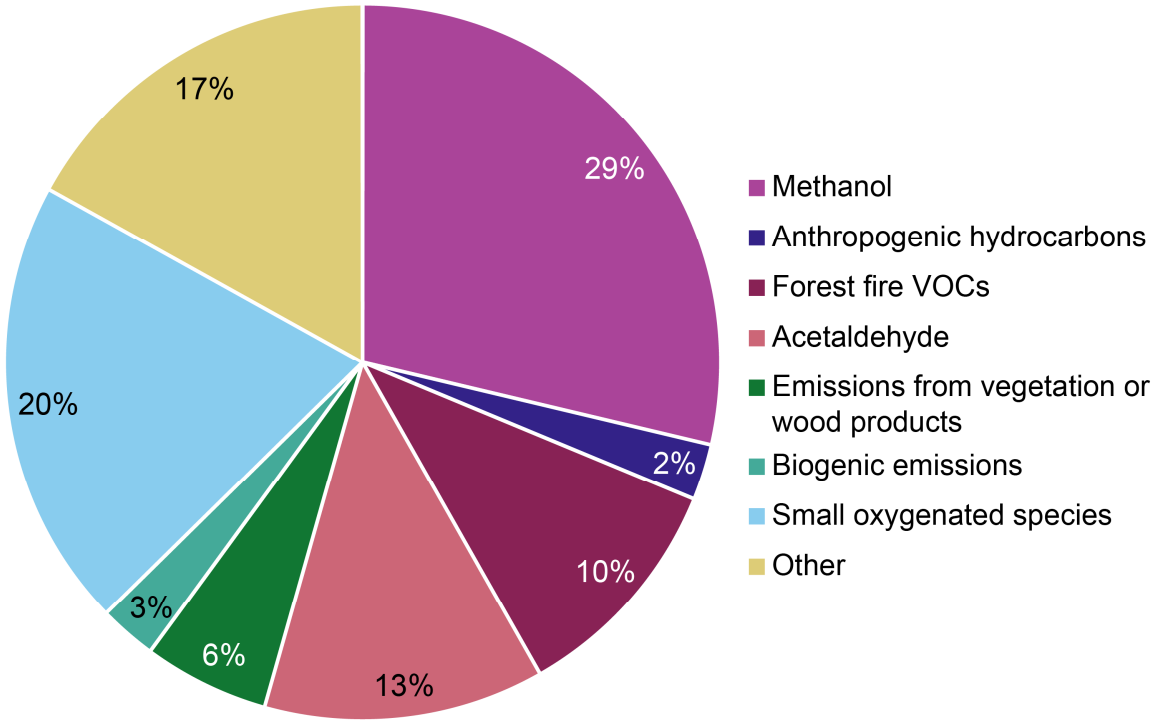


Figure 4.48: Pie chart showing the apportionment of each of the eight VOC sources observed over the course of the day on August 7, 2018.

The temporal and geographic distribution of the apportionment is shown in Figure 4.49 and Figure 4.50, respectively. In Figure 4.49, the top panel shows the measured TIC, and the bottom panel shows a stacked 100 % area plot of the eight components over the course of the day. The stacked area plot allows the change in the mixture of VOCs present throughout the day to be visualized. For example the comparatively high signals of the emissions from vegetation or wood products when driving through the old growth forest are evident just before 10:00, and the presence of forest fire smoke between 10:00–12:00 is very apparent.

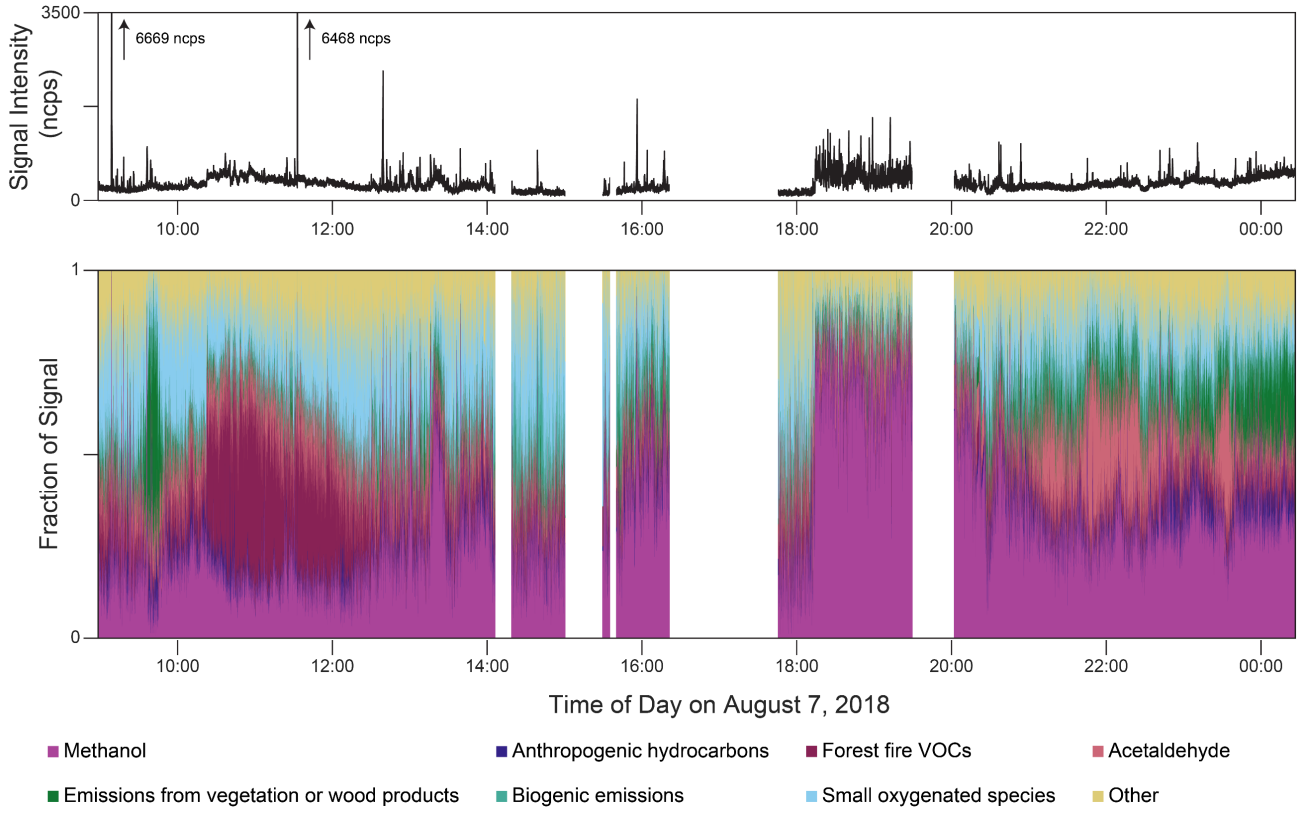


Figure 4.49: Top: TIC of the *m/z* included in the MCR-WALS analysis on August 7, 2018. Bottom: 100 % stacked area chart showing the distribution of the eight VOC components measured over the course of the day on August 7, 2018.

In the map shown in Figure 4.50 the eight components are superimposed. At each sample time, the components contributing to larger proportions (> 30 %) of the total signal have been mapped in the forefront to highlight the main VOC components observed. The size of each dot is indicative of the total amount of signal measured at that time scaled to the proportion of VOCs accounted for by the component being mapped. Mapping the components in this way produces a snapshot of the VOC sources across the region. The forest fire VOCs dominate the measurements observed in the north-west region of the map, the methanol component is present mainly in the center of Port Alberni near the pulp mill, and the anthropogenic hydrocarbon and acetaldehyde components are present in small, localized areas. Additionally, several ring charts are shown on the map, which depict the contribution of VOC sources measured at certain times and locations throughout the day. The size of the rings is scaled in proportion to the total signal intensity captured by the ring chart.

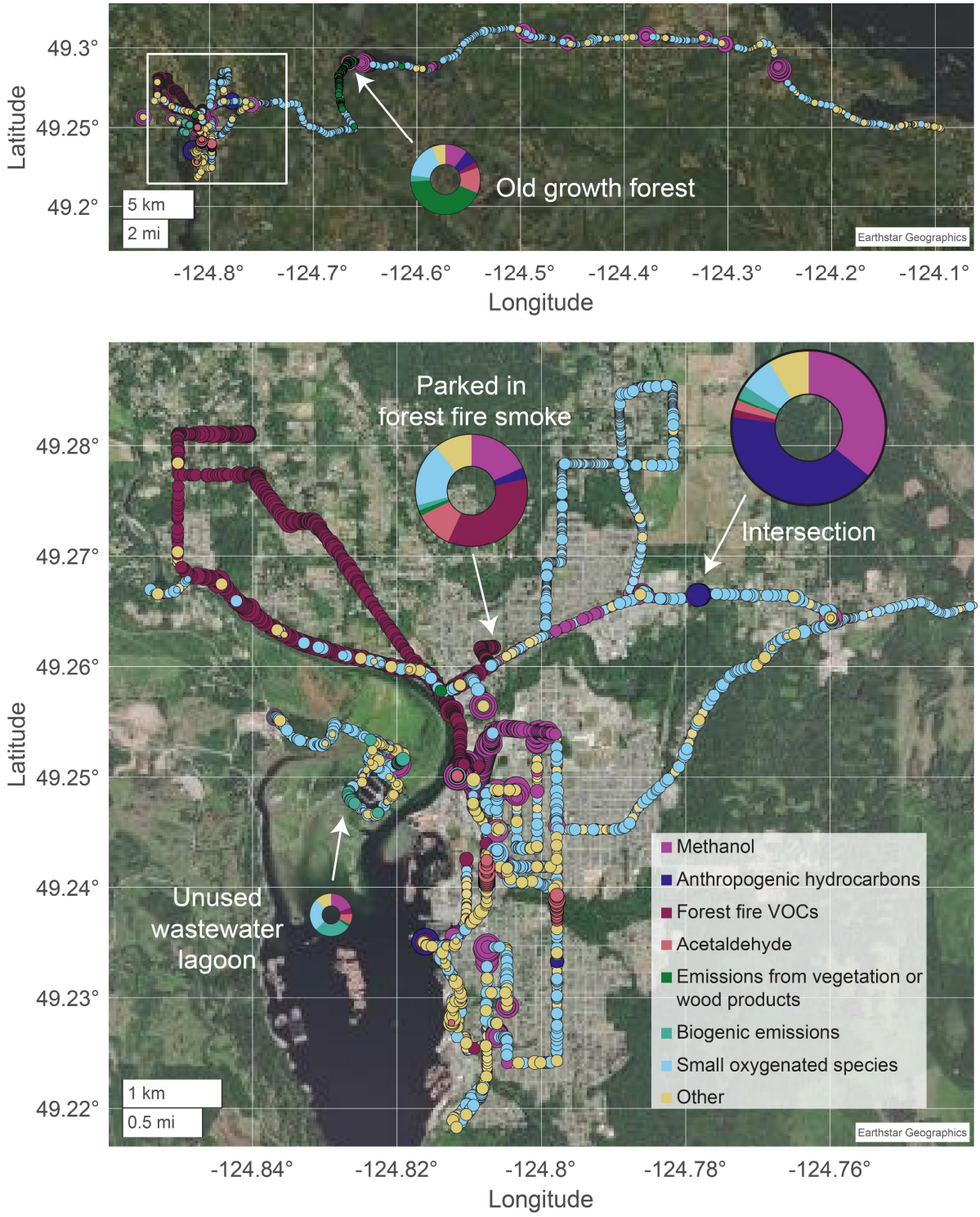


Figure 4.50: Maps of the apportionment of the VOC components measured on August 7, 2018. The dots that are visible on the maps each account for >30 % of the measured VOCs at a given location, and the larger dots describe higher signal intensities. Mapping the components together provides a visual snapshot of the distribution of VOCs present in the area. The four ring charts each describe contributions of VOC sources measured at the indicated areas, with the size of the ring proportional to the measured signal intensity at that location.

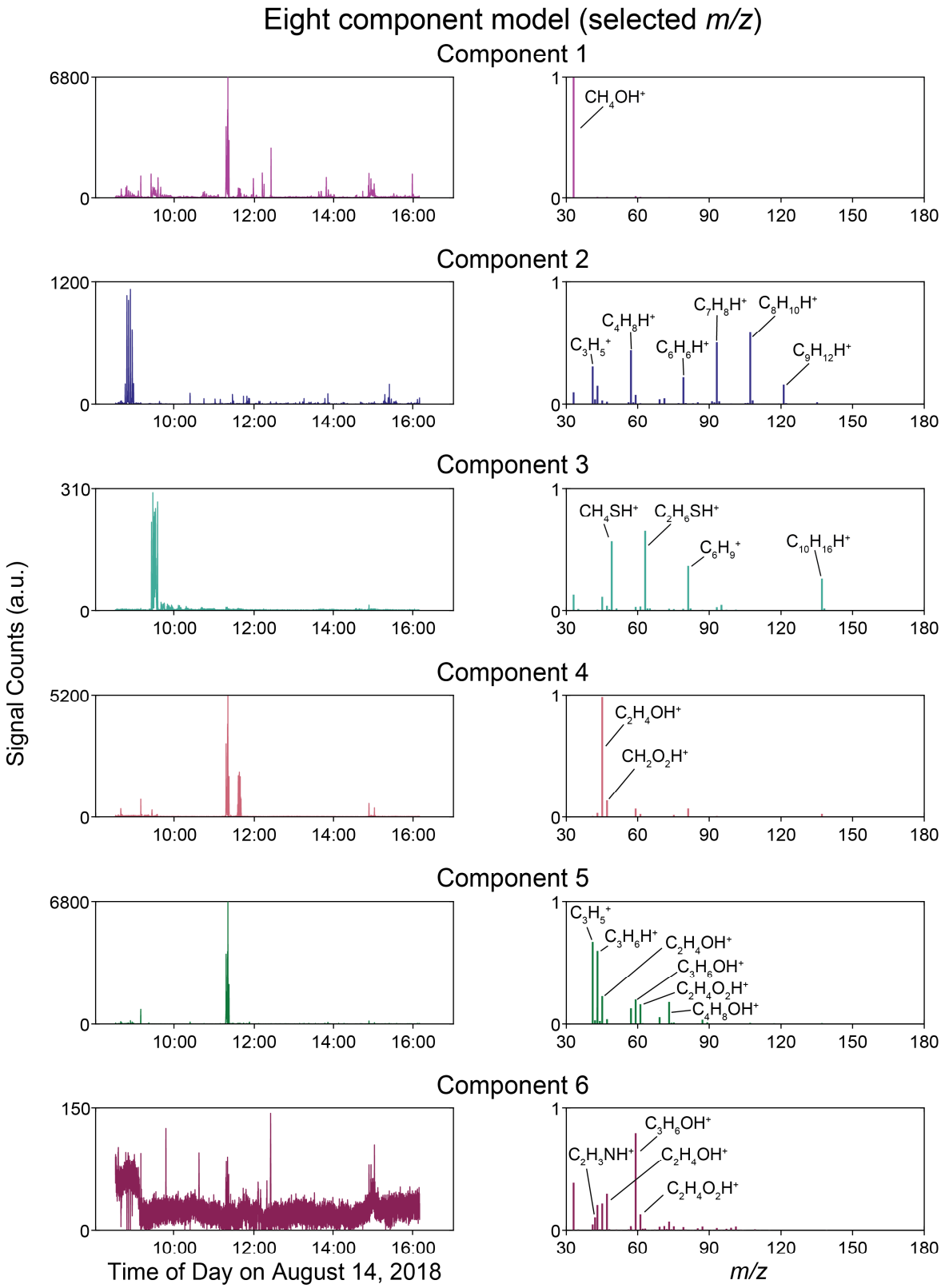
The VOC components identified on August 7, 2018 include specific sources, such as the VOCs present in the smoke from nearby forest fires and the biogenic emission from vegetation in the estuary, and more general sources such as anthropogenic hydrocarbon emissions and emissions high in monoterpenes that are due to vegetation or wood products in the sampling area. The data presented here was from a field campaign aiming to characterize the VOC emissions observed across a region, but the methodology used could also be applied to more targeted data sets to answer more specific questions at a local scale. For example, field campaigns could be designed to: explore the emissions from the different industrial areas/processes in Port Alberni (or other communities) and their impact on adjoining neighbourhoods, to identify and apportion the sources of smoke present in the colder months of the year when both the use of woodstoves for heating and backyard burning of garbage and yardwaste are commonplace, to differentiate the emissions from different vehicle types (trucks, light duty vehicles) encountered on-road, or to better characterize the biogenic emissions from the different types of vegetation present in the area. By operating a mass spectrometer in a moving vehicle these types of targeted studies become more feasible. The lab can be brought to the sample, allowing a high density of data to be collected over both time and space. Additionally, as the PTR-ToF-MS measures VOCs continuously in real-time, sampling routes can be adapted on-the-fly when interesting measurements requiring further investigation are made.

As the results described above demonstrate, applying MCR-WALS to a selection of *m*/z measured at 1 Hz using the PTR-ToF-MS operating in a moving vehicle allows for the sources of VOCs in the area to be identified, and apportioned, at unprecedented temporal and spatial resolution. This methodology has also been applied for the identification and source apportionment of VOCs observed on the August 14, 2018 and August 21 and 22, 2017 field campaigns, and the results are described in the following sections.

4.3.7 August 14, 2018 MCR-WALS results

The initial PCA analysis indicated that there were seven VOC components contributing the majority of the variance in the data collected on August 14, 2018. MCR-WALS models with six to eighteen components were calculated using the *m/z* with S/N > 1 (97 of the recorded *m/z*). The eight component model was chosen for interpretation, with details on model selection found in Appendix C.

The resolved time series and mass spectrum for each component are shown in Figure 4.51. The major ions associated with each of the components are shown in Table 4.7, along with potential chemical identities, and the locations in which the component is measured. Many of the components identified here are similar to those resolved in the August 7, 2018 Port Alberni data, including an anthropogenic hydrocarbon component, a forest fire component, and a methanol component. New components identified in this geographic region include a component related to pulp mill emissions, two components associated with emissions measured near the regional composting facility, and a biomass component. A description of each component follows, starting with the novel components identified in this dataset.


Eight component model (selected m/z)
Component 1
Component 2
Component 3
Component 4
Component 5
Component 6
Signal Counts (a.u.)
Time of Day on August 14, 2018
m/z
CH4OH+
C3H5+
C4H8H+
C6H6H+
C7H8H+
C8H10H+
C9H12H+
CH4SH+
C2H6SH+
C6H9+
C10H16H+
C2H4OH+
CH2O2H+
C3H6H+
C3H6OH+
C2H4O2H+
C4H8OH+
C2H3NH+

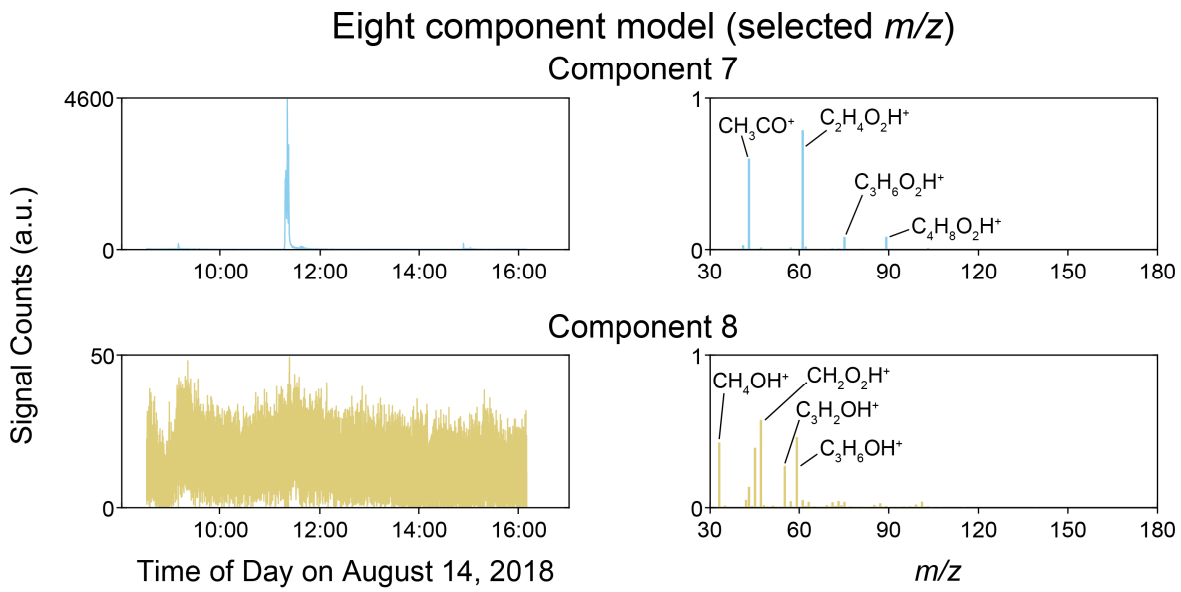


Figure 4.51: Time series (left) and mass spectra (right) for the eight component MCR-WALS model calculated for the August 14, 2018 field campaign between Nanaimo and Parksville, BC.

Table 4.7: Summary of the *m/z* contributing to each component, and the locations where the components were observed, in the eight component MCR-WALS analysis of the August 14, 2018 Nanaimo to Parksville data.

Component	**Main *m/z* and identity**	**Locations**
1 – Methanol	33.03 (1) – CH_4OH^+ (methanol)	Ubiquitous
2 – Anthropogenic hydrocarbons	33.03 (0.09) – CH_4OH^+ (methanol) 41.03 (0.31) – $C_3H_5^+$ ($C_3H_4H^+$ or fragment ion) 43.05 (0.15) – $C_3H_6H^+$(propene) 57.07 (0.07) – $C_4H_8H^+$ (butenes) 59.05 (0.11) – $C_3H_6OH^+$ (acetone) 69.07 (0.04) – $C_5H_8H^+$ (pentadiene) 71.04 (0.05) – $C_4H_6OH^+$ (2-butanal, methyl vinyl ketone) 79.05 (0.22) – $C_6H_6H^+$ (benzene) 93.06 (0.50) – $C_7H_8H^+$ (toluene) 107.08 (0.58) – $C_8H_{10}H^+$ (ethylbenzene/xylenes) 121.09 (0.16) – $C_9H_{12}H^+$ (C3-aromatics)	Auto wrecker Gas station Traffic Intersections
3 – Pulp mill plume	33.03 (0.13) – CH_4OH^+ (methanol) 45.03 (0.12) – $C_2H_4OH^+$ (acetaldehyde) 47.01 (0.04) – $CH_2O_2H^+$ (formic acid) 49.01 (0.57) – CH_4SH^+ (methanethiol) 63.02 (0.65) – $C_2H_6SH^+$ (DMS, ethanethiol) 81.07 (0.37) – $C_6H_9^+$ (monoterpenes fragment) 95.00 (0.05) – $C_2H_6S_2H^+$ (DMDS) 137.12 (0.27) – $C_{10}H_{16}H^+$ (monoterpenes)	Pulp mill plume Pollution control center

Table 4.7: Summary of the *m/z* contributing to each component, and the locations where the components were observed, in the eight component MCR-WALS analysis of the August 14, 2018 Nanaimo to Parksville data, continued.

Component	**Main *m/z* and identity**	**Locations**
4 – Biomass	45.03 (0.99) – $C_2H_4OH^+$ (acetaldehyde) 47.04 (0.13) – $C_2H_6OH^+$ (ethanol) 59.05 (0.07) – $C_3H_6OH^+$ (acetone) 61.03 (0.02) – $C_2H_4O_2H^+$ (acetic acid) 81.07 (0.07) – $C_6H_9^+$ (monoterpene fragment) 137.12 (0.02) – $C_{10}H_{16}H^+$ (monoterpenes)	Compost facility Sawmill Waste transfer station
5 – Small oxygenated species	41.03 (0.67) – $C_3H_5^+$ (alkyl fragment, oxygenated species fragment) 43.01 (0.13) – CH_3CO^+ (acetic acid fragment) 43.05 (0.60) – $C_3H_6H^+$ (propene) 45.03 (0.23) – $C_2H_4OH^+$ (acetaldehyde) 57.07 (0.13) – $C_4H_8H^+$ (butenes, oxygenated species fragment) 59.05 (0.20) – $C_3H_6OH^+$ (acetone) 61.03 (0.16) – $C_2H_4O_2H^+$ (acetic acid) 69.07 (0.05) – $C_5H_8H^+$ (isoprene) 73.06 (0.18) – $C_4H_8OH^+$ (butanone)	Compost facility Waste transfer station
6 – Forest fire VOCs	33.03 (0.08) – CH_4OH^+ (methanol) 42.03 (0.11) – $C_2H_3NH^+$(acetonitrile) 43.01 (0.20) – CH_3CO^+ (acetic acid fragment) 45.03 (0.22) – $C_2H_4OH^+$ (acetaldehyde) 47.01 (0.30) – $CH_2O_2H^+$ (formic acid) 59.05 (0.79) – $C_3H_6OH^+$ (acetone) 61.03 (0.13) – $C_2H_4O_2H^+$ (acetic acid) 73.06 (0.07) – $C_4H_8OH^+$ (butanone)	Ubiquitous Higher in the morning
7 – Compost - carboxylic acids	43.01 (0.60) – CH_3CO^+ (acetic acid fragment) 61.03 (0.78) – $C_2H_4O_2H^+$ (acetic acid) 75.04 (0.08) – $C_3H_6O_2H^+$ (propanoic acid) 89.06 (0.08) – $C_4H_8O_2H^+$ (butanoic acid)	Compost facility Waste transfer station Persists in sample lines
8 – Other	33.03 (0.43) – CH_4OH^+ (methanol) 42.03 (0.05) – $C_2H_3NH^+$(acetonitrile) 43.01 (0.14) – CH_3CO^+ (acetic acid fragment) 45.03 (0.38) – $C_2H_4OH^+$ (acetaldehyde) 47.01 (0.57) – $CH_2O_2H^+$ (formic acid) 55.01 (0.28) – $C_3H_2OH^+$ (propynal) 57.03 (0.04) – $C_3H_4OH^+$ (Acrolein, fragment ion) 59.05 (0.46) – $C_3H_6OH^+$ (acetone) 61.03 (0.05) – $C_2H_4O_2H^+$ (acetic acid) 63.02 (0.04) – $C_2H_6SH^+$ (DMS, ethanethiol) 71.04 (0.04) – $C_4H_6OH^+$ (butanal, methyl vinyl ketone) 73.06 (0.04) – $C_4H_8OH^+$ (butanone) 75.04 (0.04) – $C_3H_6O_2H^+$ (propanoic acid) 87.04 (0.03) – $C_4H_6O_2H^+$ (2,3-butanedione, methyl acrylate)	Ubiquitous

The numbers in brackets after each *m/z* indicate the signal intensity in the unit vector normalized mass spectra for each component.

4.3.7.1 August 14, 2018 Component 3 of 8 – Pulp mill plume

Component 3 is mainly associated with a plume emitted from a pulp mill that was detected between 9:30–11:00 in the south end of the drive route. Methanethiol (*m/z* 49.01), DMS (*m/z* 63.02), and monoterpenes (*m/z* 137.12, fragment at *m/z* 81.07) are the main compounds measured for this component, with methanol (*m/z* 33.03), acetaldehyde (*m/z* 45.03), formic acid (*m/z* 47.01), and DMDS (*m/z* 95.00) being minor contributors. The measurements for this component are highest near the pulp mill at approximately 9:30, and decrease with distance from the point source, as shown in the time series in Figure 4.52. The plume was carried south/south east by the wind (3.5 m/s N/NW measured at Harmac Pacific meteorological station[248], indicated on the map with the pink star) and was detected as far as 8 km from the pulp mill, as shown on the map in Figure 4.53. During this same time period (9:00–11:00) hourly total reduced sulphur concentrations of up to 2.5 ppb_v were measured at a nearby BC Ministry of Environment air quality monitoring station (Harmac Cedar Woobank[248], indicated on the map with the white star). Stack emissions from this pulp mill have been reported to contain methanol, acetaldehyde, reduced sulphur compounds (not speciated), and VOCs (also not speciated) in the Environment Canada and Climate Change emissions inventory database.[244]

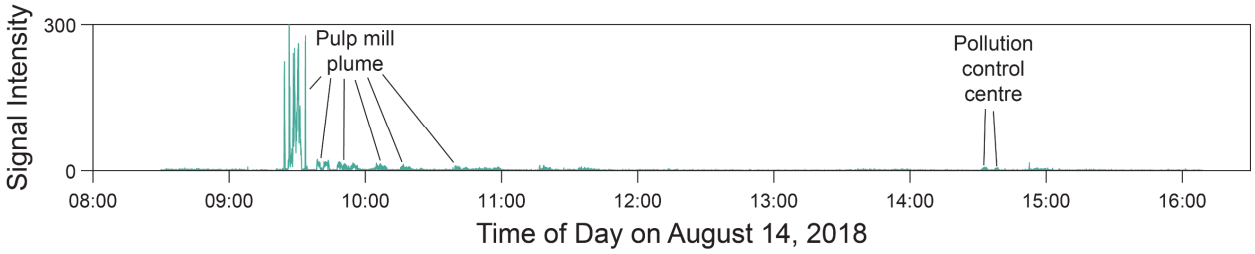


Figure 4.52: Annotated time series of the pulp mill plume (Component 3 of 8) in the eight component MCR-WALS analysis of the August 14, 2018 Nanaimo to Parksville data.

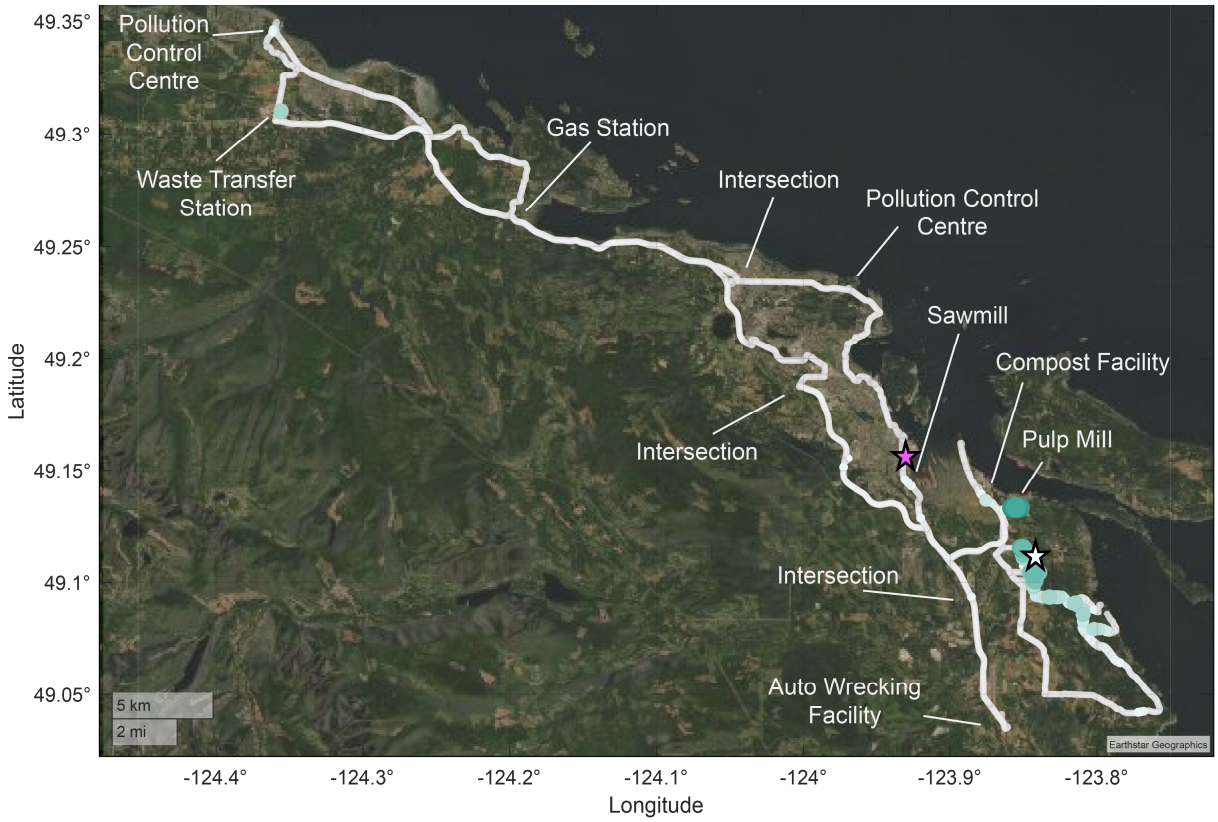


Figure 4.53: Geospatial distribution of the pulp mill plume (Component 3 of 8) in the MCR-WALS model of the August 14, 2018 data. Larger and darker dots indicate higher signal intensities This component was mainly measured at, and downwind from, the pulp mill. The white star indicates the location of the Harmac Cedar Woobank air quality monitoring station (hourly TRS and $PM_{2.5}$ measurements), and the pink star is the location of the Harmac Pacific Meteorological Station (wind speed and direction).

4.3.7.2 August 14, 2018 Component 5 of 8 – Small oxygenated species

Component 5 was measured near the regional composting facility with high signal counts (up to 6800 ncps), with short, low signal intensity (< 250 ncps) excursions measured throughout the day in various locations such as at a waste transfer station, at intersections, and near an auto wrecking facility as shown in Figure 4.54 and Figure 4.55 . Most of the *m/z* contributing to this component can be attributed to oxygenated species as either MH^+ ions (acetaldehyde, *m/z* 45.03; acetone, *m/z* 59.05; acetic acid, *m/z* 61.03; methyl ethyl ketone/butanol, *m/z* 73.06), or fragment ions (*m/z* 41.03, alkyl fragment from hydrocarbons or oxygenated compounds; *m/z* 43.01, acetic acid fragment; *m/z* 57.07, butene or fragment from oxygenated compounds). While the highest signals measured for this component are found near the composting facility, there is overlap with other types of sources (such as the auto wrecking facility) generally in locations where a wide range of compounds (*e.g.*, hydrocarbons and oxygenated compounds) were measured at high concentrations resulting in the emissions from the source being split across multiple components.

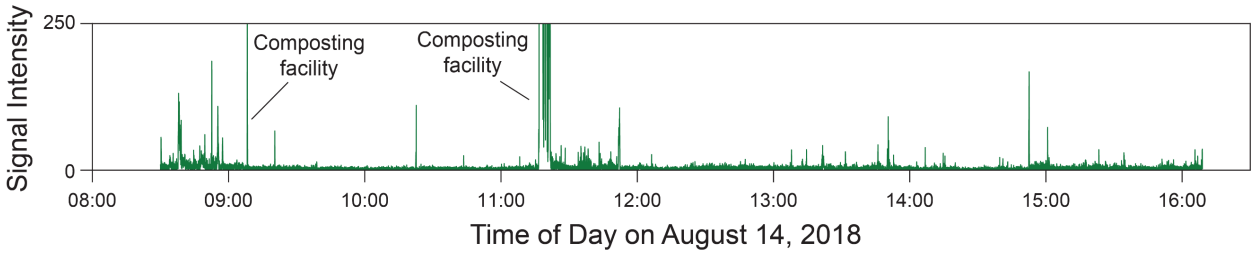


Figure 4.54: Annotated time series of the small oxygenated species (Component 5 of 8) in the eight component MCR-WALS analysis of the August 14, 2018 Nanaimo to Parksville data.

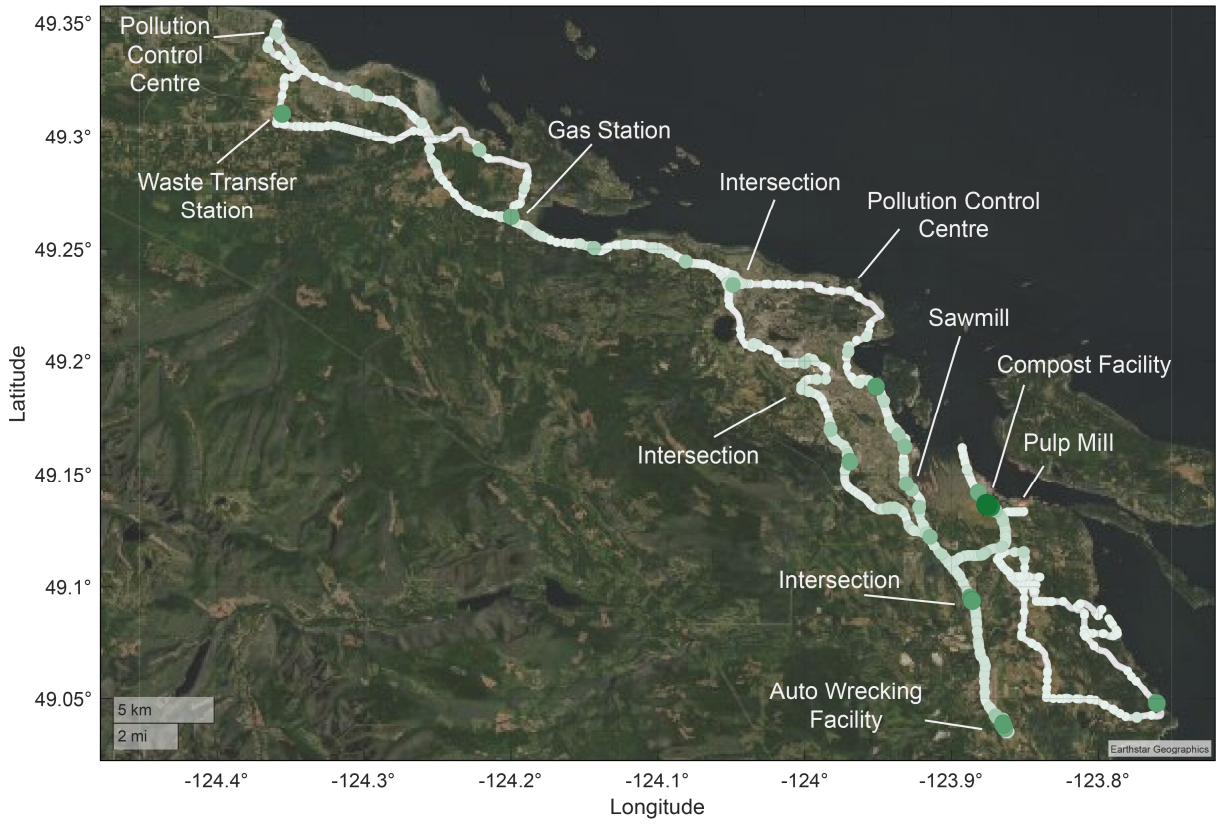


Figure 4.55: Geospatial distribution of the small oxygenated species component (Component 5 of 8) in the MCR-WALS model of the August 14, 2018 data. Larger and darker dots indicate higher signal intensities. This component was measured near many different point sources, such as the composting facility, waste transfer station, gas station, and auto wrecking facility.

4.3.7.3 August 14, 2018 Component 7 of 8 – Compost - carboxylic acids

Component 7 is associated predominately with the measurements of four ions attributed to acetic acid (*m/z* 61.03 and fragment *m/z* 43.01), propanoic acid (*m/z* 75.04), and butanoic acid (m*/z* 89.06). While these VOCs could potentially be other compounds (*e.g.*, diols, esters), we hypothesize these to be carboxylic acids due to the presence of characteristic fragment ions, such as *m/z* 43.01, a fragment of acetic acid. Fragment ions of propanoic acid and butanoic acid (*m/z* 57.03 and 71.09 respectively) were also observed in the measured data at low signal intensity near the composting facility (data not shown).

The time series associated with this component has an observably dampened decay signal (Figure 4.56) due to the more polar, less volatile species persisting in the sample.[249,250] This leads to an artifact in the mapping of this component as the molecules continue to be measured a long distance from the composting facility (Figure 4.57) as they clear from the sampling lines while driving from the composting facility back toward Nanaimo (as shown by the burgundy arrow on the map). The point sources for this component are the regional composting facility and waste transfer station, but unlike the pulp mill emissions discussed in Component 3, the persistence of this component far from the emission source cannot be attributed to a plume from the facility. Additionally, the dampened signal intensity from the more polar components may have led to the signal intensity measured near the composting facility and waste transfer center to be split across multiple components due to differences in temporal profiles of the higher and lower volatility compounds emitted from these sources.

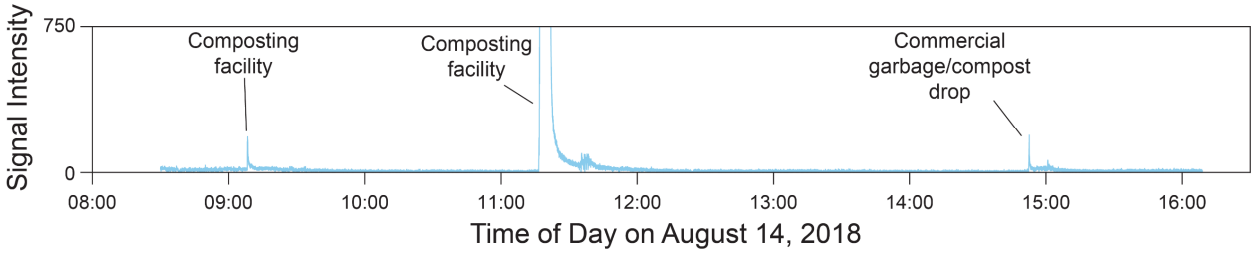


Figure 4.56: Annotated time series of the compost carboxylic acids (Component 7 of 8) in the eight component MCR-WALS analysis of the August 14, 2018 Nanaimo to Parksville data.

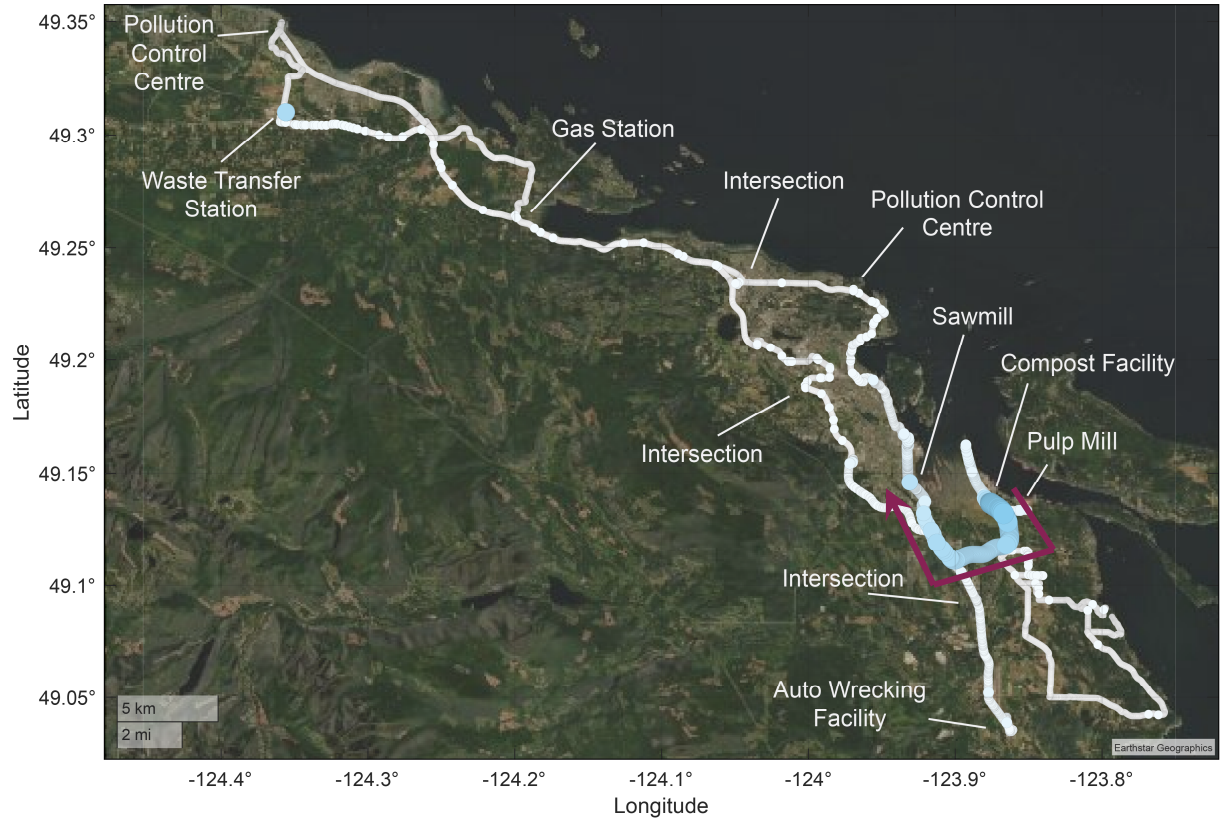


Figure 4.57: Geospatial distribution of the compost carboxylic acids (Component 7 of 8) in the MCR-WALS model of the August 14, 2018 data. Larger and darker dots indicate higher signal intensities. This component was measured at the compost facility and waste transfer station. The polar compounds present in this component had long decay times in the sampling lines resulting in the component appearing to be present on the drive route from the compost facility (indicated by burgundy arrow), however this is an artifact in the data.

4.3.7.4 August 14, 2018 Component 2 of 8 – Anthropogenic hydrocarbons

Component 2 is associated with anthropogenic hydrocarbon emissions. As with the anthropogenic hydrocarbon component in the August 7, 2018 data, the VOCs present in the mass spectrum for this component are mainly hydrocarbons (BTEX, C9-aromatics, propene, and butene) and several small oxygenated species (methanol, acetone). This component is measured while driving on a highway, on busy city streets, and at intersections. On the annotated time series (Figure 4.58) the times when the mobile lab was driving on high traffic roads are indicated by pink bars. Additionally, this component is measured at high signal intensities in proximity to an auto wrecking facility where cars are crushed before recycling. The map of this component is shown in Figure 4.59.

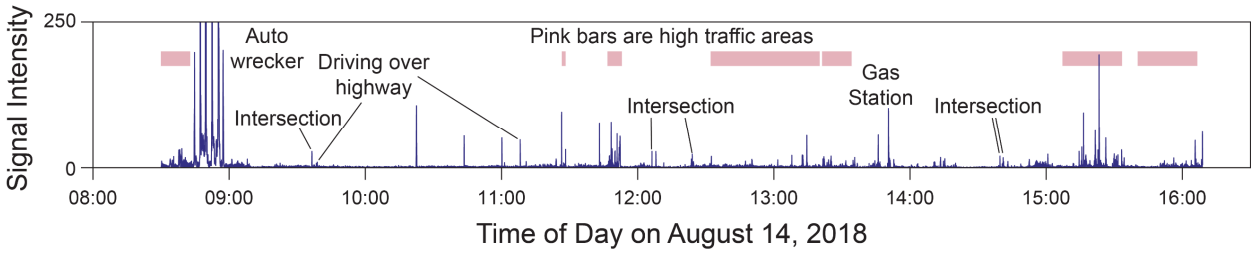


Figure 4.58: Annotated time series of the anthropogenic hydrocarbon emissions (Component 2 of 8) in the eight component MCR-WALS analysis of the August 14, 2018 Nanaimo to Parksville data.

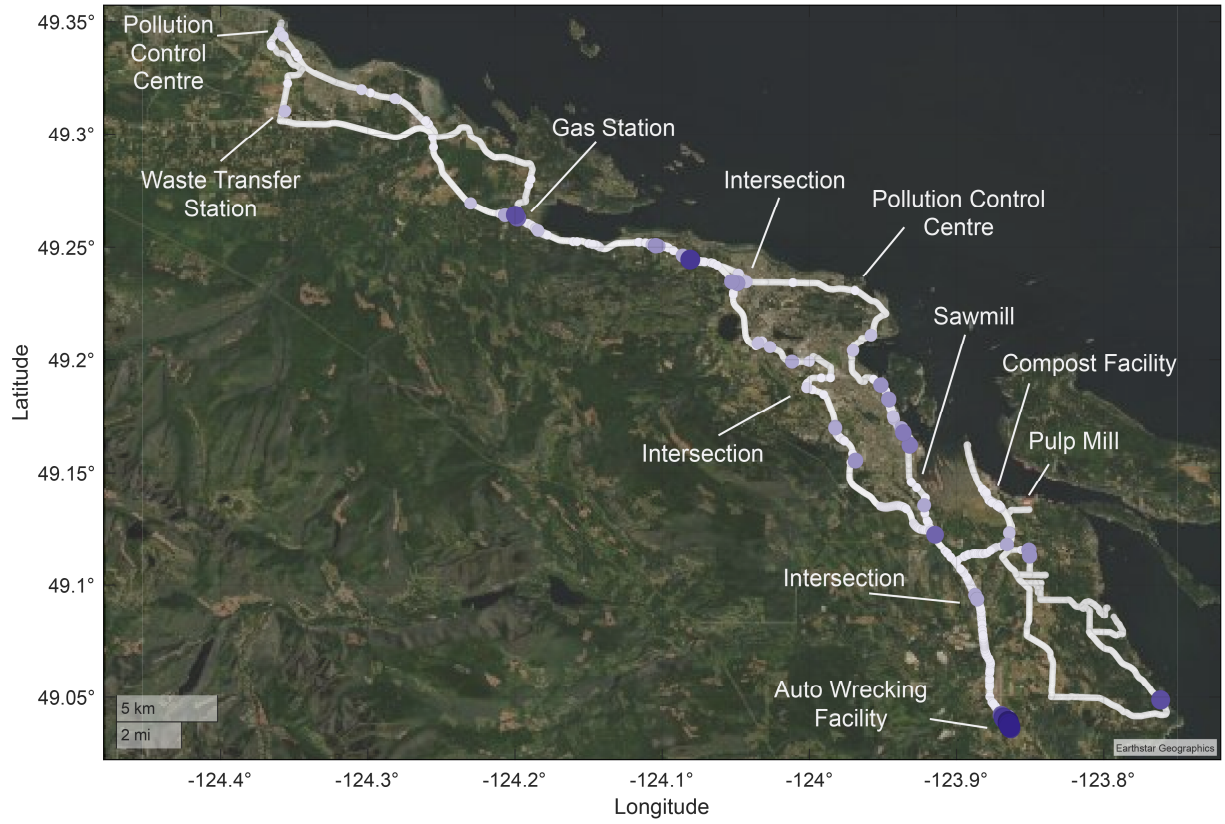


Figure 4.59: Geospatial distribution of anthropogenic hydrocarbon emissions (Component 2 of 8) in the MCR-WALS model of the August 14, 2018 data. Larger and darker dots indicate higher signal intensities. This component was measured near a gas station, auto wrecking facility, at intersections, and while driving on busy roads.

4.3.7.5 August 14, 2018 Component 6 of 8 – Forest fire VOCs

Component 6 is mostly associated with the forest fire smoke in the region. Smoke was present throughout the day, but was thicker in the morning before 9:15 when sampling on the south end of the drive route as shown in the time series and map in Figure 4.60 and Figure 4.61. The ions measured in this component are due to methanol (*m/z* 33.03), acetonitrile (*m/z* 42.03), acetaldehyde (*m/z* 45.03), formic acid (*m/z* 47.01), acetone (*m/z* 59.05), acetic acid (*m/z* 61.03 and 43.01), and methyl ethyl ketone (*m/z* 73.06). The higher signals measured at the beginning of the day correlate with on-board $PM_{2.5}$ measurements of ~100 µg/m^3, which decreased to ~30 µg/m^3 after 9:15. A time series overlay and scatter plot of the signal intensity in the forest fire VOC component and the $PM_{2.5}$ concentrations are shown in Figure 4.62. Hourly $PM_{2.5}$ concentrations recorded at the BC Ministry of Environment's Harmac Cedar Woobank air quality monitoring station (white star on the map) indicate elevated levels of $PM_{2.5}$ (> 60 µg/m^3) before 9:00, with lower levels (20–40 µg/m^3) persisting for the rest of the day, while the Nanaimo Labieux Road Station (blue star on the map) measured $PM_{2.5}$ concentrations of 15 –30 µg/m^3 throughout the day.[248] Notably, both monitoring stations recorded 24-hour average $PM_{2.5}$ concentrations over BC's air quality objective of 25 µg/m^3 on this day. Again, we observed a good correlation between this forest fire VOC component and the $PM_{2.5}$ data, although not as strong as that described earlier on August 7, 2018 in Port Alberni (Figure 4.33). This is likely due to some non-fire derived $PM_{2.5}$ arising from drives on gravel roads on August 14, 2018. Additionally, the time series for this component shows elevated signals at 11:15 (regional composting facility) and around 15:00 (waste transfer station) which do not correspond to the $PM_{2.5}$ measurements. At these locations many of the same ions were measured, but in much greater abundance than was observed in the forest fire smoke.

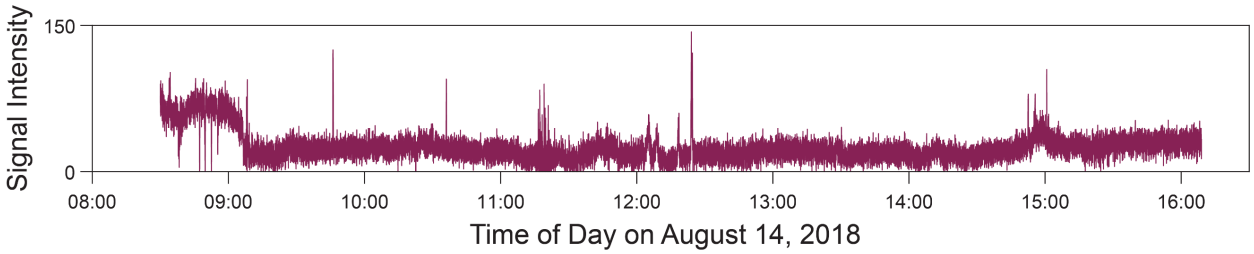


Figure 4.60: Annotated time series of the forest fire VOCs (Component 6 of 8) in the eight component MCR-WALS analysis of the August 14, 2018 Nanaimo to Parksville data.

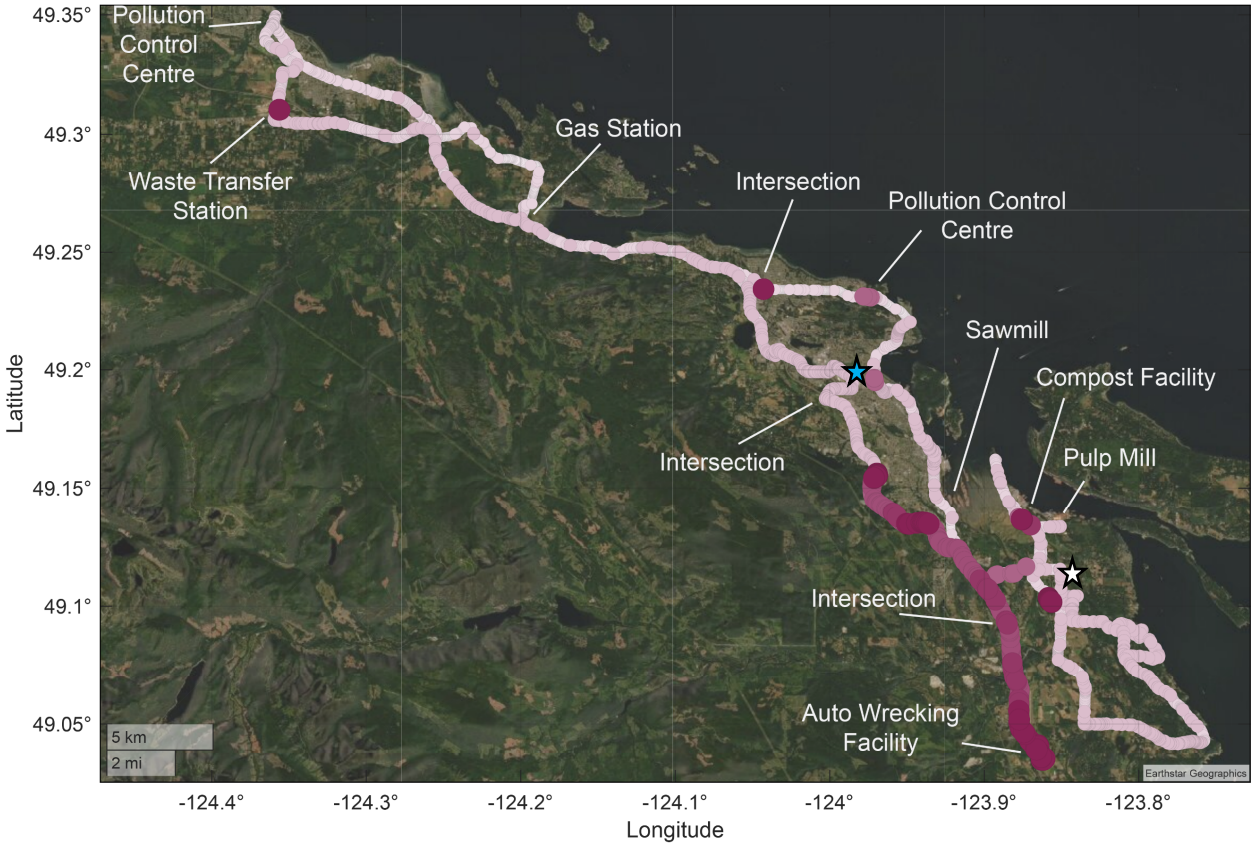


Figure 4.61: Geospatial distribution of the forest fire VOCs (Component 6 of 8) in the MCR-WALS model of the August 14, 2018 data. Larger and darker dots indicate higher signal intensities. This component was present throughout the day, with higher concentrations measured in the morning at the south end of the drive route. The white star indicates the location of the Harmac Cedar Woobank air quality monitoring station (hourly TRS and $PM_{2.5}$ measurements), and the blue star is the location of the Nanaimo Labieux air quality monitoring station (hourly $PM_{2.5}$ measurements).

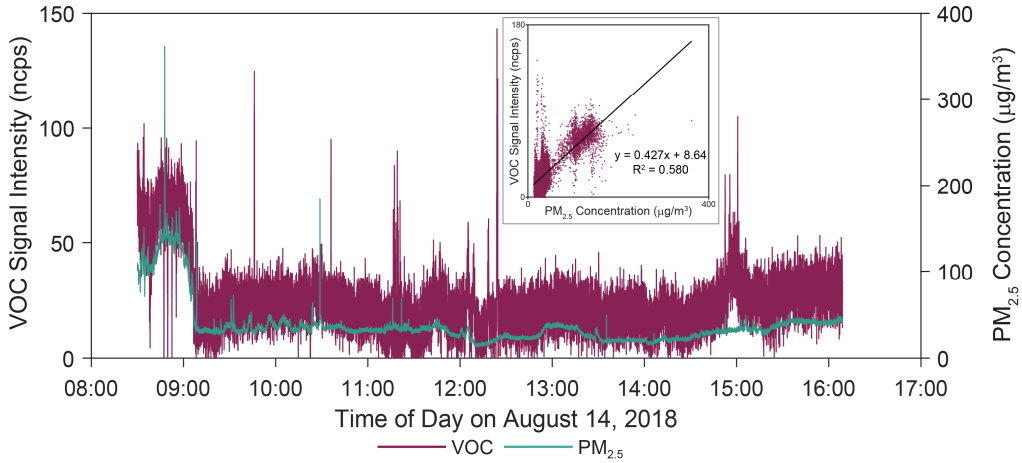


Figure 4.62: Overlay of the time series for the forest fire component and the $PM_{2.5}$ concentrations on August 14, 2018. The inset graph shows the correlation between the two datasets. Some of the high concentrations measured in the $PM_{2.5}$ data correspond with sampling occurring while driving on gravel roads. In the forest fire component time series the higher signal intensities measured at 11:15 and 15:00 are not related to forest fire smoke, but are artifacts due to the high concentration of similar compounds measured in the vicinity of the compost facility and waste transfer station.

4.3.7.6 August 14, 2018 Component 4 of 8 – Biomass

Component 4 is measured mainly in the vicinity of industrial facilities that use or process biomass such as a composting facility, sawmill, waste transfer facility, and near the wood chip piles at the pulp mill as shown in Figure 4.63 and Figure 4.64. Low intensity signals were also measured when sampling in forest fire smoke, and a signal was encountered at an intersection on the highway. The compounds measured in the mass spectrum for this component include acetaldehyde (*m/z* 45.03), ethanol (*m/z* 47.04), acetone (*m/z* 59.05), and very low monoterpene (*m/z* 137.12 and 81.07) signals.

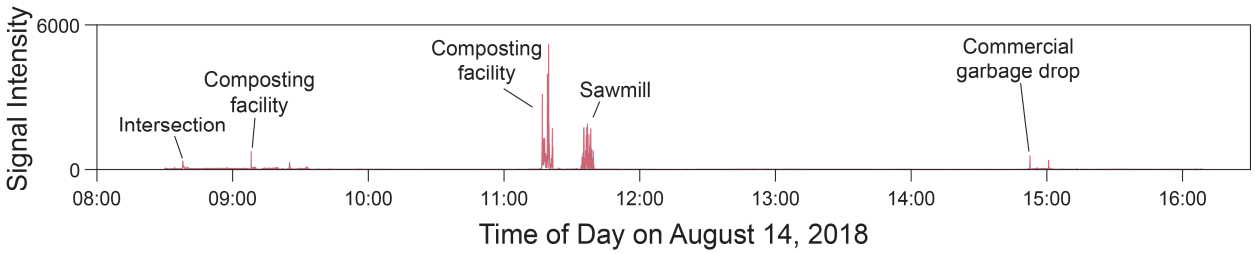


Figure 4.63: Annotated time series of the biomass emissions (Component 4 of 8) in the eight component MCR-WALS analysis of the August 14, 2018 Nanaimo to Parksville data.

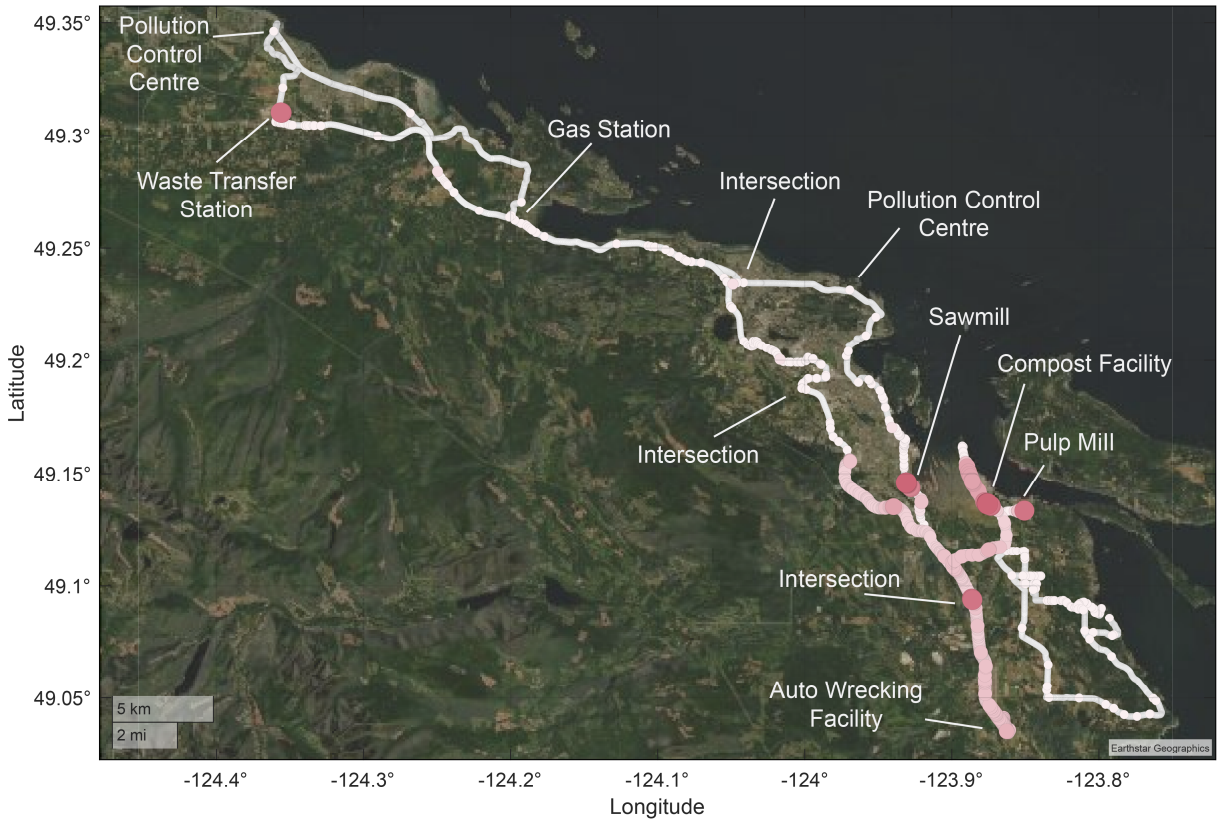


Figure 4.64: Geospatial distribution of the biomass emissions (Component 4 of 8) in the MCR-WALS model of the August 14, 2018 data. Larger and darker dots indicate higher signal intensities This component was measured mainly in the vicinity of biomass and wood-based industries.

4.3.7.7 August 14, 2018 Component 1 of 8 – Methanol

Component 1 is associated with methanol measurements and is found throughout the region. As in the August 7, 2018 analysis, methanol measurements are not associated with any one source or source type. Concentration excursions measured near composting and waste management facilities, pulp mill, sawmill, auto wrecker, and intersections were observed. The annotated time series and map of this component are shown in Figure 4.65 and Figure 4.66, respectively.

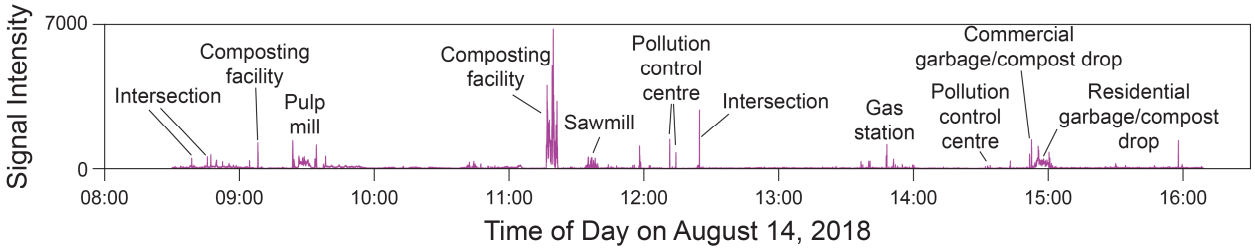


Figure 4.65: Annotated time series of the methanol component (Component 1 of 8) in the eight component MCR-WALS analysis of the August 14, 2018 Nanaimo to Parksville data.

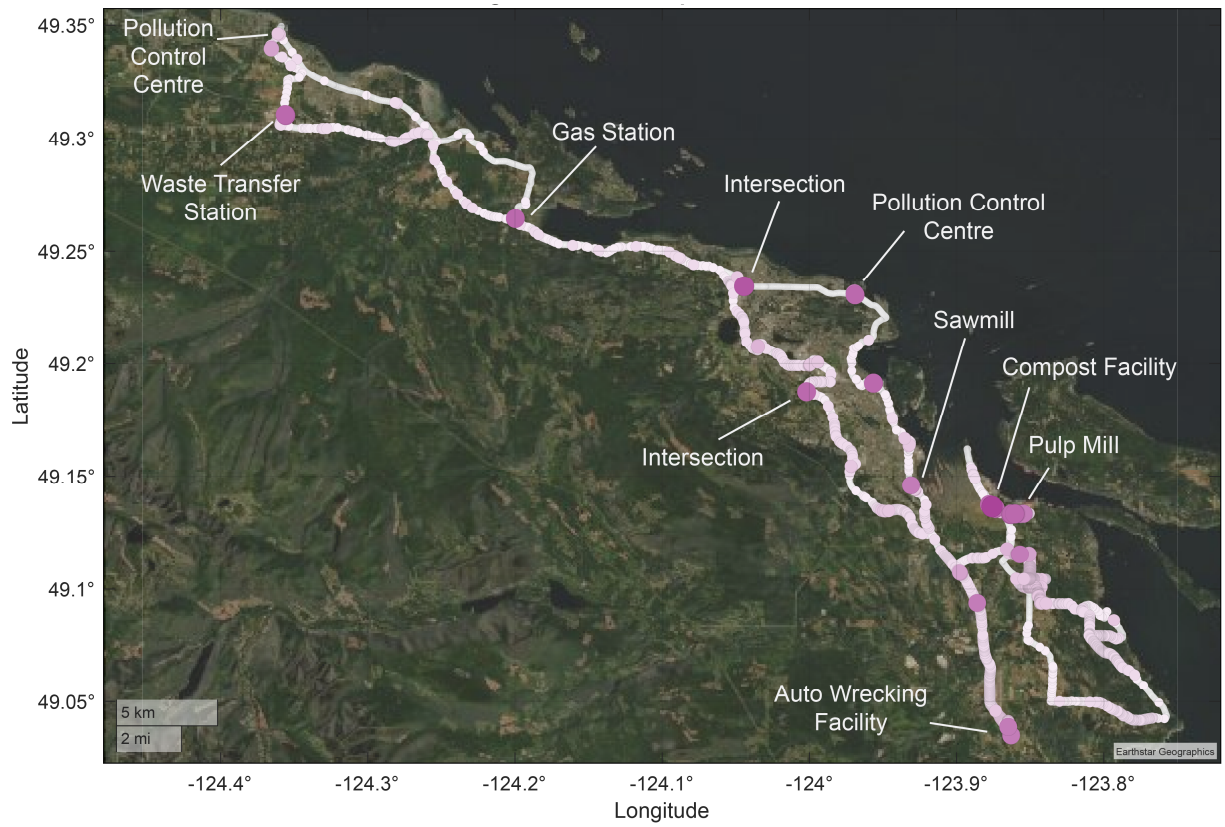


Figure 4.66: Geospatial distribution of the methanol component (Component 1 of 8) in the MCR-WALS model of the August 14, 2018 data. Larger and darker dots indicate higher signal intensities. This component was measured in many different areas such as near the pulp mill, near the waste management facilities, at intersections, and the auto wrecking facility.

4.3.7.8 August 14, 2018 Component 8 of 8 – Other

Component 8 is very similar to Component 8 in the August 7, 2018 MCR-WALS model in that it has no spatial or temporal features associated with sampling locations during the day, and therefore a map is not presented. The resolved mass spectra contains the signals for similar ions (*m/z* 33.03, 43.01, 45.03, 47.01, 55.01, and 59.05) to those in Component 8 on August 7, 2018. Like the August 7, 2018 component, there is not an apparent physical interpretation for this component related to specific regional sources.

4.3.7.9 August 14, 2018 source apportionment

The pie chart of the apportionment of the observed VOCs observed on August 14, 2018 is shown in Figure 4.67. As with the August 7, 2018 analysis components associated with point sources, or short concentration excursions (anthropogenic hydrocarbons, pulp mill plume, compost) account for less of the total measured signal intensity than those with more frequent measurements, such as the forest fire component that was observed throughout the day. The temporal distributions of the eight components are shown in Figure 4.68, along with the TIC of the measured ions included in the data analysis. In this plot, emissions from some of the major point source are evident (auto wrecking facility, pulp mill, compost facility, and waste transfer station).

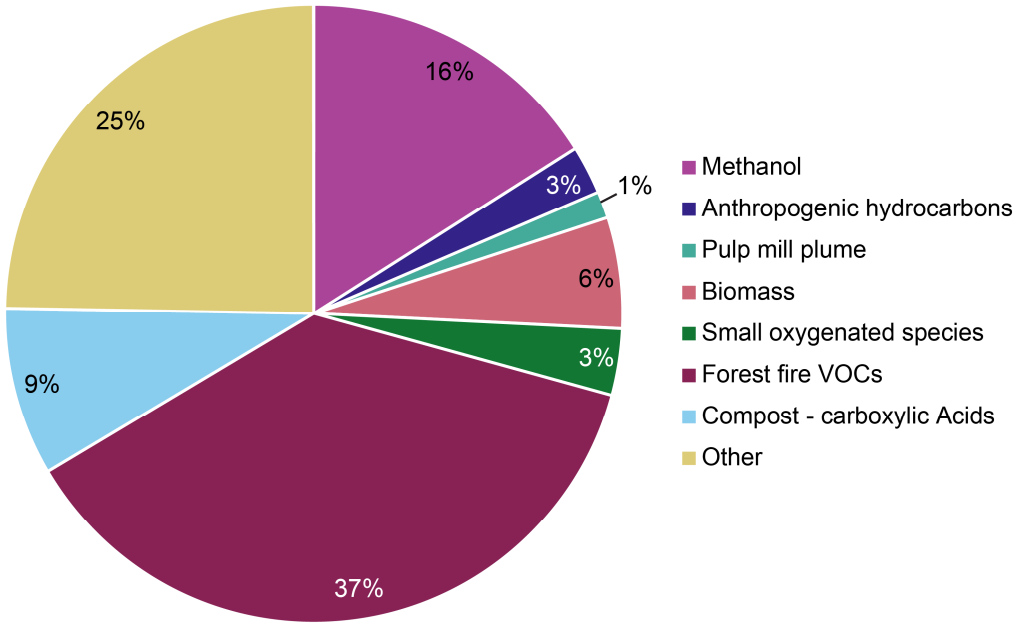


Figure 4.67: Pie chart showing the apportionment of each of the eight VOC sources observed over the course of the day on August 14, 2018 between Nanaimo and Parksville, BC.

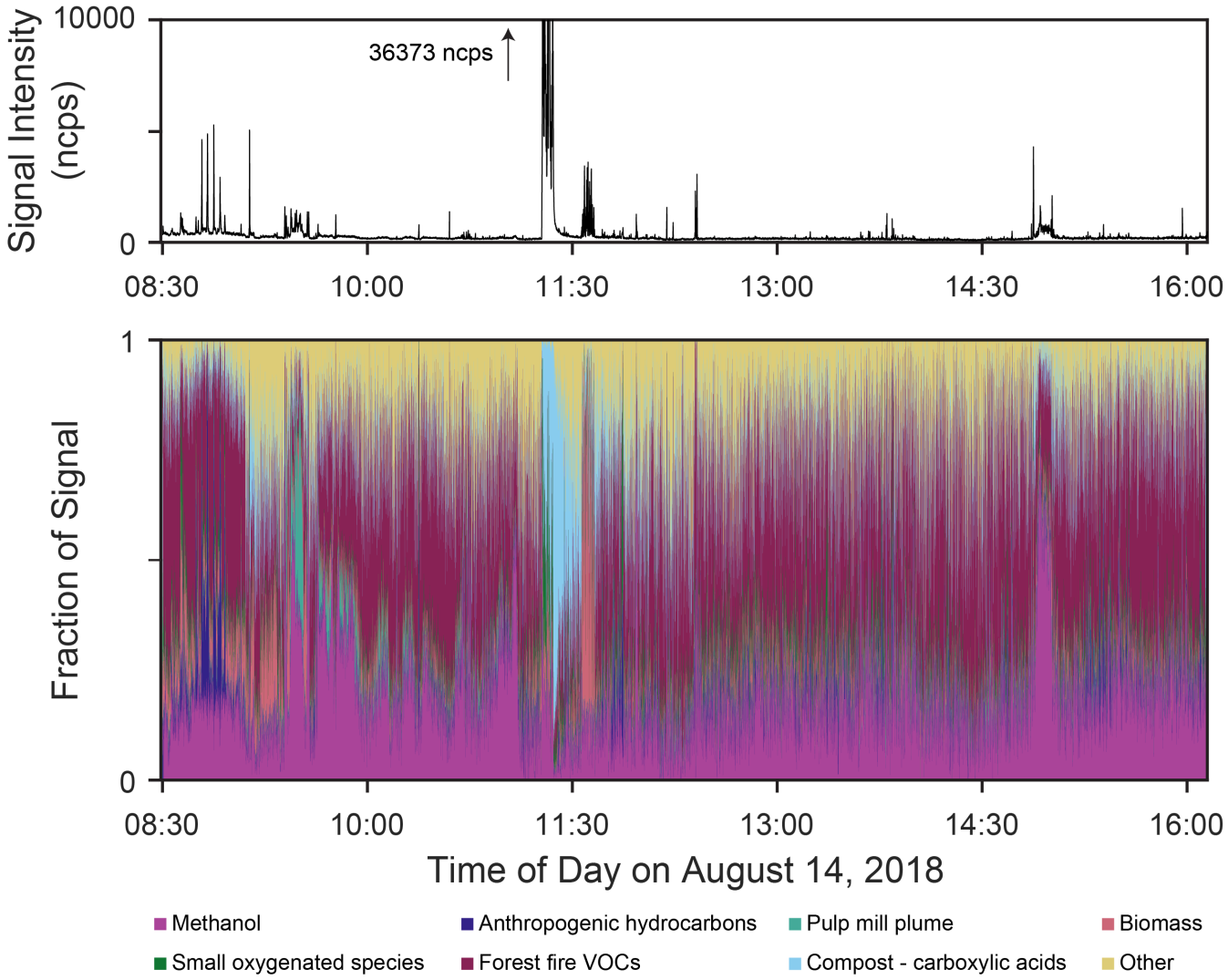


Figure 4.68: Top: TIC of the *m/z* included in the MCR-WALS analysis on August 14, 2018. Bottom: 100 % stacked area chart showing the distribution of the eight VOC components measured over the course of the day on August 14, 2018.

The spatial distribution of the sources is shown in Figure 4.69. High proportions of the biomass, anthropogenic hydrocarbon, compost, and methanol components are distributed near point source across the area. The ring charts on the map indicate the apportionment of VOCs present at particular locations, with the size of the ring proportional the total observed signal intensity. From these charts, it is evident that the distribution of VOCs measured in the air near the composting facility, pulp mill, sawmill, and auto wrecker differ significantly, with the composting facility emitting much higher concentrations of VOCs on this day compared to the other sources.

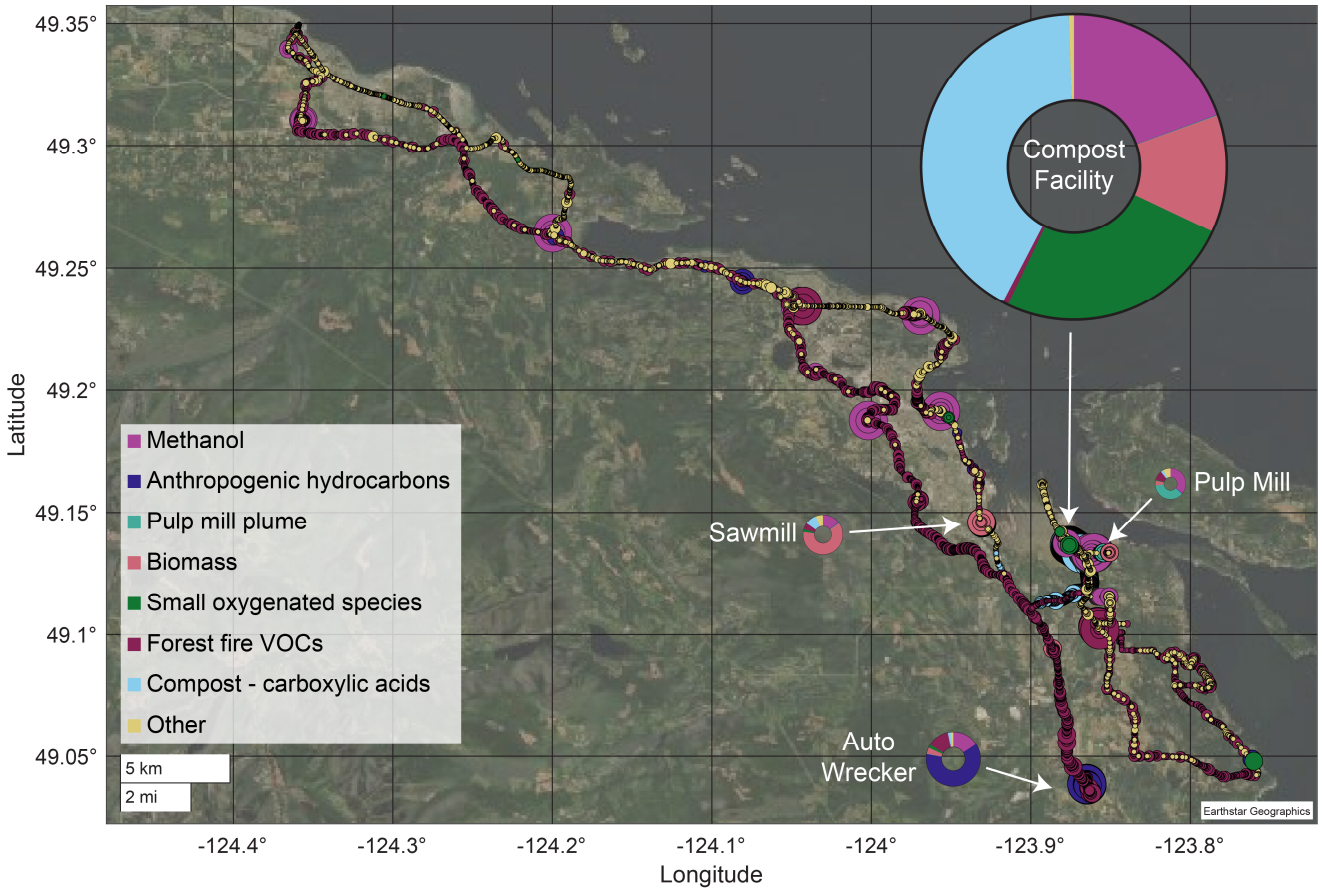


Figure 4.69: Map of the apportionment of the VOC components measured on August 14, 2018. The dots that are visible on the maps each account for >30 % of the measured VOCs at a given location, and the larger dots describe higher signal intensities. Mapping the components together provides a visual snapshot of the distribution of VOCs present in the area. The four ring charts each describe distribution of the VOCs measured at the indicated areas, with the size of the ring proportional to the measured signal intensity at that location.

4.3.7.10 August 14, 2018 challenges

The August 14, 2018 dataset was more challenging to analyze than the August 7, 2018 dataset. When choosing the number of components to use for interpretation, many different models were considered before the eight component model was chosen. Models with additional components improved the modeling of the data measured at certain *m/z* (*e.g.*, 81.07, 137.12, 107.08), but this often resulted in the resolution of components that were solely due to the presence of a single chemical compound rather than a source type, as well as splitting other components in two. For example, an eleven component model which improves the modeling of the monoterpene signals (*m/z* 81.07 and 137.1) results in a model where there are two components used to describe the carboxylic acids measured near the composting facility, a component associated with acetone measurements, a component mainly associated with monoterpene signals, and a component mainly associated with acetaldehyde signals. Additionally, the eleven component model results in a forest fire VOC component that correlates better with the $PM_{2.5}$ measurements (Figure 4.70), but the resolved components that

are due to acetone, acetaldehyde, methanol, and monoterpene signals all model a portion of the VOCs attributed to the forest fire smoke. While the eight component model does not completely model all signals measured on the field campaign, it does allow for patterns in measured data to be identified without producing a model that is resolving multiple components that are associated with individual chemical compounds or isomer classes.

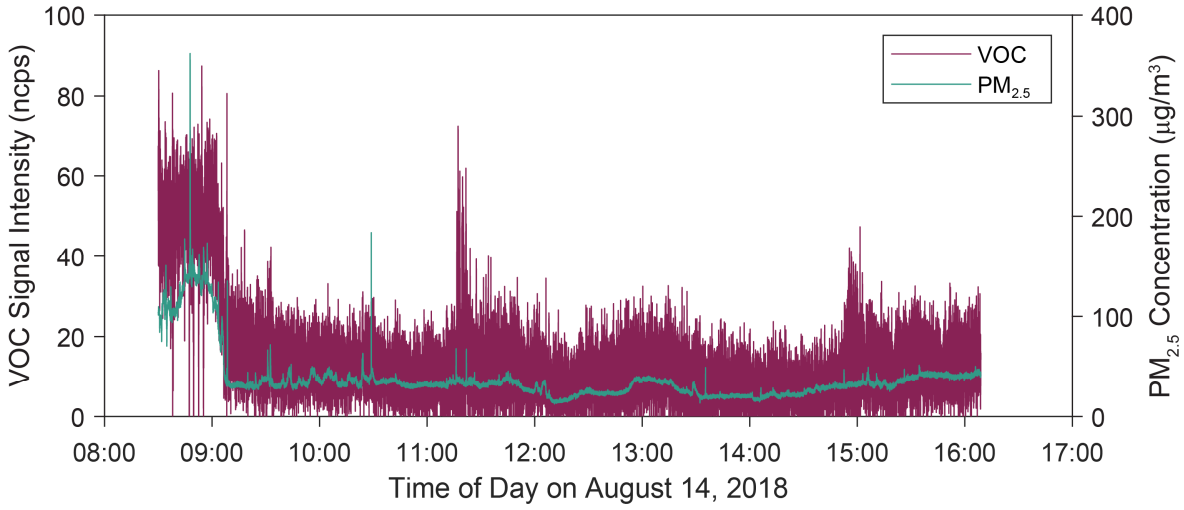
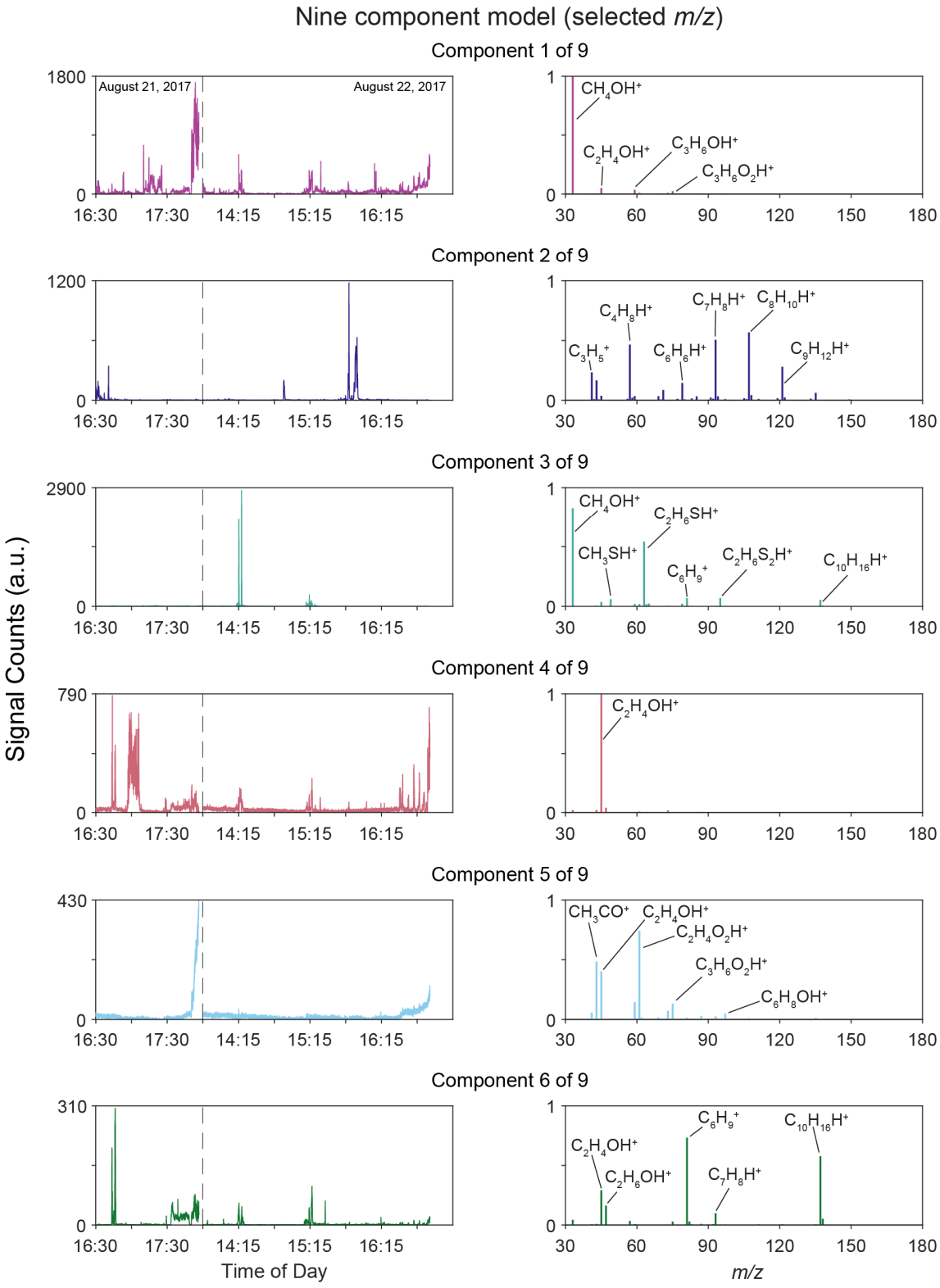


Figure 4.70: Overlay of time series of the forest fire component and $PM_{2.5}$ concentrations for the eleven component MCR-WALS model for August 14, 2018.

4.3.8 August 21 and 22, 2017 MCR-WALS results

For the data collected between Nanaimo and Crofton on August 21 and 22, 2017 MCR-WALS models with seven to twelve components were interrogated, and a nine component model was chosen for interpretation. Details on model selection are shown in Appendix C.

The resolved time series and mass spectra for the nine component MCR-WALS model are shown in Figure 4.71. Eight of the nine components have been identified with a possible source (*e.g.*, compost facility, pulp mill, biomass) or group of molecules (*e.g.*, anthropogenic hydrocarbons, small hydrocarbons and oxygenated species), with one component remaining unidentified. For each component, a source description, list of measured ions with potential chemical identities, and list of the locations where they were encountered are shown in Table 4.8. The components resolved in this model are very similar to those present in the model for the August 7, 2018 data and so will be discussed briefly in the following sections, along with the annotated time series and maps for the components.


Nine component model (selected m/z)
Component 1 of 9
August 21, 2017
August 22, 2017
CH4OH+
C2H4OH+
C3H6OH+
C3H6O2H+
Component 2 of 9
C4H8H+
C7H8H+
C8H10H+
C3H5+
C6H6H+
C9H12H+
Component 3 of 9
CH4OH+
C2H6SH+
CH3SH+
C6H9+
C2H6S2H+
C10H16H+
Component 4 of 9
C2H4OH+
Component 5 of 9
CH3CO+
C2H4OH+
C2H4O2H+
C3H6O2H+
C6H8OH+
Component 6 of 9
C6H9+
C10H16H+
C2H4OH+
C2H6OH+
C7H8H+
Signal Counts (a.u.)
Time of Day
m/z

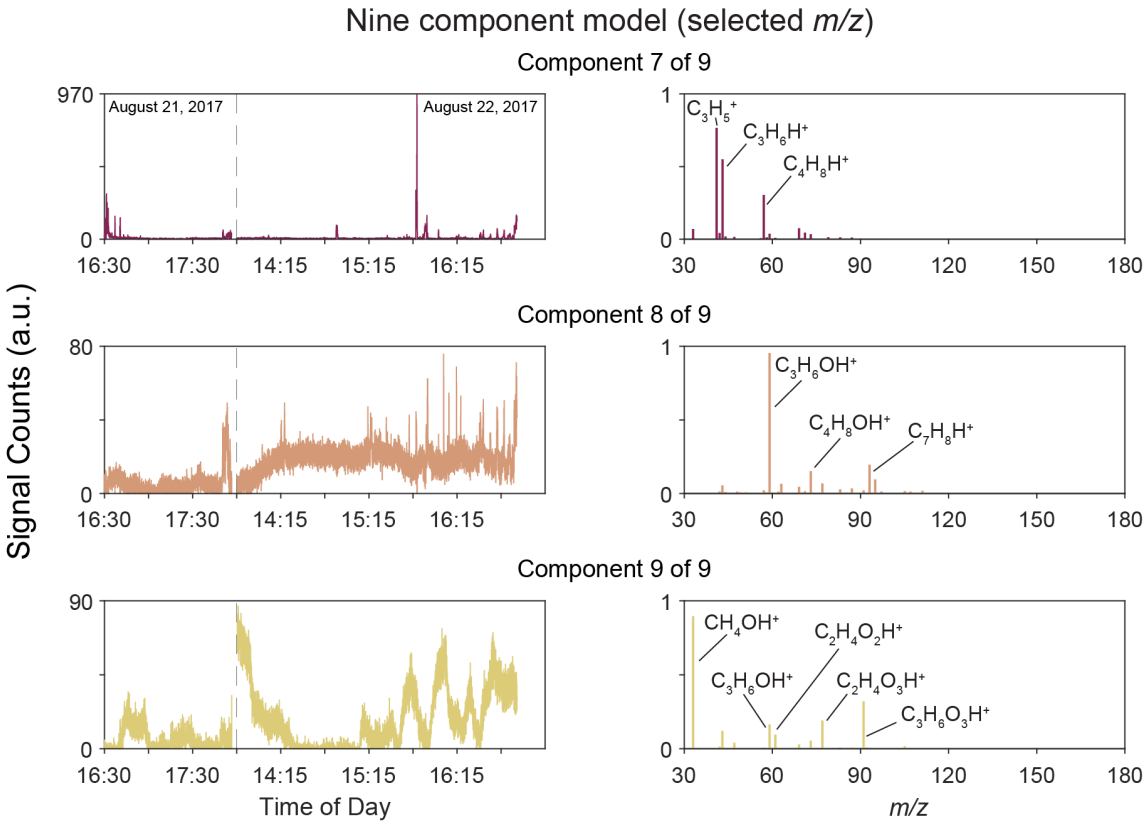


Figure 4.71: Resolved time series (left) and mass spectra (right) for the nine component MCR-WALS model calculated for the August 21 and 22, 2017 data between Nanaimo and Crofton. The vertical dotted line separates the August 21 and 22 data.

Table 4.8: Summary of the *m/z* contributing to each component, and the locations where the components were observed, in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data.

Component	Main *m/z* and identity	Locations
1 – Methanol	33.03 (0.99) – CH_4OH^+ (methanol)	Pulp mill
	45.03 (0.05) – $C_2H_4OH^+$ (acetaldehyde)	Compost facility
	59.05 (0.03) – $C_3H_6OH^+$ (acetone/propanol)	Gas station
	75.04 (0.02) – $C_3H_6O_2H^+$ (propanoic acid/hydroxyacetone)	Sawmills
		Landfill
		Auto wrecker
		Dairy farm
2 – Anthropogenic hydrocarbons	41.03 (0.23) – $C_3H_5^+$ (hydrocarbon fragment ion)	Gas station
	43.05 (0.17) – $C_3H_6H^+$ (propene)	Auto wrecker
	45.03 (0.04) – $C_2H_4OH^+$ (acetaldehyde)	Ferry terminal
	57.07 (0.47) – $C_4H_8H^+$ (butene)	Tractor
	59.05 (0.03) – $C_3H_6OH^+$ (acetone/propanol)	Wood chip truck
	69.07 (0.03) – $C_5H_8H^+$ (isoprene)	
	71.08 (0.09) – $C_5H_{10}H^+$ (pentene)	
	79.05 (0.15) – $C_6H_6H^+$ (benzene)	
	93.06 (0.51) – $C_7H_8H^+$ (toluene)	
	107.08 (0.57) – $C_8H_{10}H^+$ (ethylbenzene/xylenes)	
	121.09 (0.28) – $C_9H_{12}H^+$ (C9-aromatics)	
	135.12 (0.06) – $C_{10}H_{14}H^+$ (C10-aromatics)	

Table 4.8: Summary of the *m/z* contributing to each component, and the locations where the components were observed, in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data, continued.

Component	Main *m/z* and identity	Locations
3 – Pulp mill	33.03 (0.83) – CH_4OH^+ (methanol) 49.01 (0.06) – CH_4SH^+ (methanethiol) 63.02 (0.54) – $C_2H_6SH^+$ (dimethyl sulphide/ethanethiol) 81.07 (0.07) – $C_6H_9^+$ (monoterpene fragment) 95.00 (0.07) – $C_2H_6S_2H^+$ (DMDS) 137.12 (0.05) – $C_{10}H_{16}H^+$ (monoterpenes)	Pulp mills
4 – Acetaldehyde	33.03 (0.02) – CH_4OH^+ (methanol) 43.01 (0.02) – CH_3CO^+ (acetic acid fragment) 45.03 (0.99) – $C_2H_4OH^+$ (acetaldehyde) 47.04 (0.04) – $C_2H_6OH^+$ (ethanol)	Compost facility Sawmills Wood chip truck Pulp mill Wood storage
5 – Compost	41.03 (0.06) – $C_3H_5^+$ (hydrocarbon fragment) 43.01 (0.48) – CH_3CO^+ (acetic acid fragment) 45.03 (0.40) – $C_2H_4OH^+$ (acetaldehyde) 47.01 (0.07) – $CH_2O_2H^+$ (formic acid) 59.05 (0.15) – $C_3H_6OH^+$ (acetone) 61.03 (0.74) – $C_2H_4O_2H^+$ (acetic acid) 73.06 (0.07) – $C_4H_8OH^+$ (butanone) 75.04 (0.13) – $C_3H_6O_2H^+$ (propanoic acid) 87.04 (0.03) – $C_4H_6O_2H^+$ (2,3,-butanedione)	Compost facility
6 – Biomass	33.03 (0.04) – CH_4OH^+ (methanol) 45.03 (0.29) – $C_2H_4OH^+$(acetaldehyde) 47.04 (0.16) – $C_2H_6OH^+$ (ethanol) 81.07 (0.73) – $C_6H_9^+$ (monoterpene fragment) 93.06 (0.10) – $C_7H_8H^+$ (toluene, monoterpene fragment) 137.12 (0.58) – $C_{10}H_{16}H^+$	Sawmills Wood chip truck Lumber industry Compost facility Pulp mill Landfill
7 – Small hydrocarbons and oxygenated species	33.03 (0.07) – CH_4OH^+ (methanol) 41.03 (0.77) – $C_3H_5^+$ (hydrocarbon fragment) 43.05 (0.55) – $C_3H_6H^+$ (propene) 57.07 (0.30) – $C_4H_8H^+$ (butenes) 59.05 (0.04) – $C_3H_6OH^+$ (acetone) 69.07 (0.08) – $C_5H_8H^+$ (isoprene) 71.08 (0.04) – $C_5H_{10}H^+$ (pentene) 73.06 (0.04) – $C_4H_8OH^+$ (butanone)	Gas station Compost facility Ferry terminal Auto wrecker Tractor
8 – Acetone	43.01 (0.06) – CH_3CO^+ (acetic acid fragment) 59.05 (0.95) – $C_3H_6OH^+$ (acetone) 63.02 (0.07) – $C_2H_6SH^+$ (dimethyl sulphide/ethanethiol) 69.07 (0.05) – $C_5H_8H^+$ (isoprene) 73.06 (0.15) – $C_4H_8OH^+$ (butanone) 77.02 (0.07) – $C_2H_4O_3H^+$ (glycolic acid) 87.04 (0.04) – $C_4H_6O_2H^+$ (2,3,-butanedione) 93.06 (0.20) – $C_7H_8H^+$ (toluene) 95.00 (0.10) – $C_2H_6S_2H^+$ (dimethyl disulfide)	Compost Pulp mill Auto wrecker Landfill
9 – Other	33.03 (0.90) – CH_4OH^+ (methanol) 43.01 (0.12) – CH_3CO^+ (acetic acid fragment) 47.04 (0.04) – $C_2H_6OH^+$ (ethanol) 59.05 (0.16) – $C_3H_6OH^+$ (acetone) 61.03 (0.10) – $C_2H_4O_2H^+$ (acetic acid) 69.07 (0.03) – $C_5H_8H^+$ (isoprene) 73.06 (0.05) – $C_4H_8OH^+$ (butanone) 77.02 (0.19) – $C_2H_4O_3H^+$ (glycolic acid) 91.04 (0.32) – $C_3H_6O_3H^+$ (lactic acid)	No specific geographic region

The numbers in brackets after each *m/z* indicate the signal intensity in the unit vector normalized mass spectra for each component.

4.3.8.1 August 21 and 22, 2017 Component 1 of 9 – Methanol

The annotated time series and map of Component 1 are shown in Figure 4.72 and Figure 4.73, respectively. Like the MCR-WALS analysis of the two other datasets, the mass spectrum for Component 1 is dominated by methanol, but on this day also has low intensity signals for acetaldehyde (*m/z* 45.03), acetone/propanol (*m/z* 59.05), and hydroxyacetone/propanoic acid (*m/z* 75.04). This component is not associated with a single source or source type, but is present throughout the two sampling days. The highest intensity signals associated with this compound are measured near the regional composting facility after 17:50 on August 21, and in the vicinity of the pulp mills on the drive route. Additional signals were measured near the gas station and a sawmill. On August 21, 2017 sampling occurred in the vicinity of the pulp mill on the North end of the drive route between 17:09–17:13 and field notes indicate that odour from the pulp mill was detected until 17:22. The high signal intensities related to this odour are located to the south of the pulp mill, as shown on the map, when sampling was occurring downwind of the mill (3.5 m/s wind from the North measured at the Harmac Pacific Met 60 monitoring station).[248] Odour associated with pulp mill emissions being accounted for in the methanol component was also observed during the August 7, 2018 field campaign in Port Alberni. Sulphur compounds were also measured near the pulp mills on the drive route, and will be discussed in Component 3.

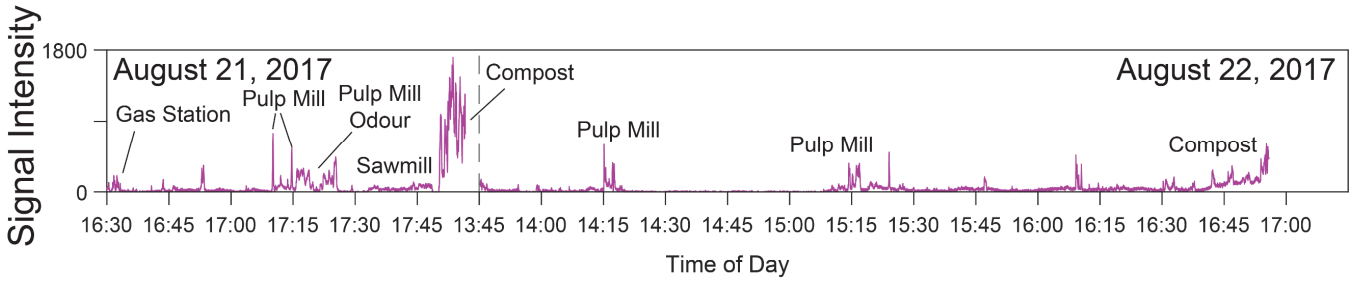


Figure 4.72: Annotated time series of the methanol component (Component 1 of 9) in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data. The vertical dotted line separates the August 21 and 22 data.

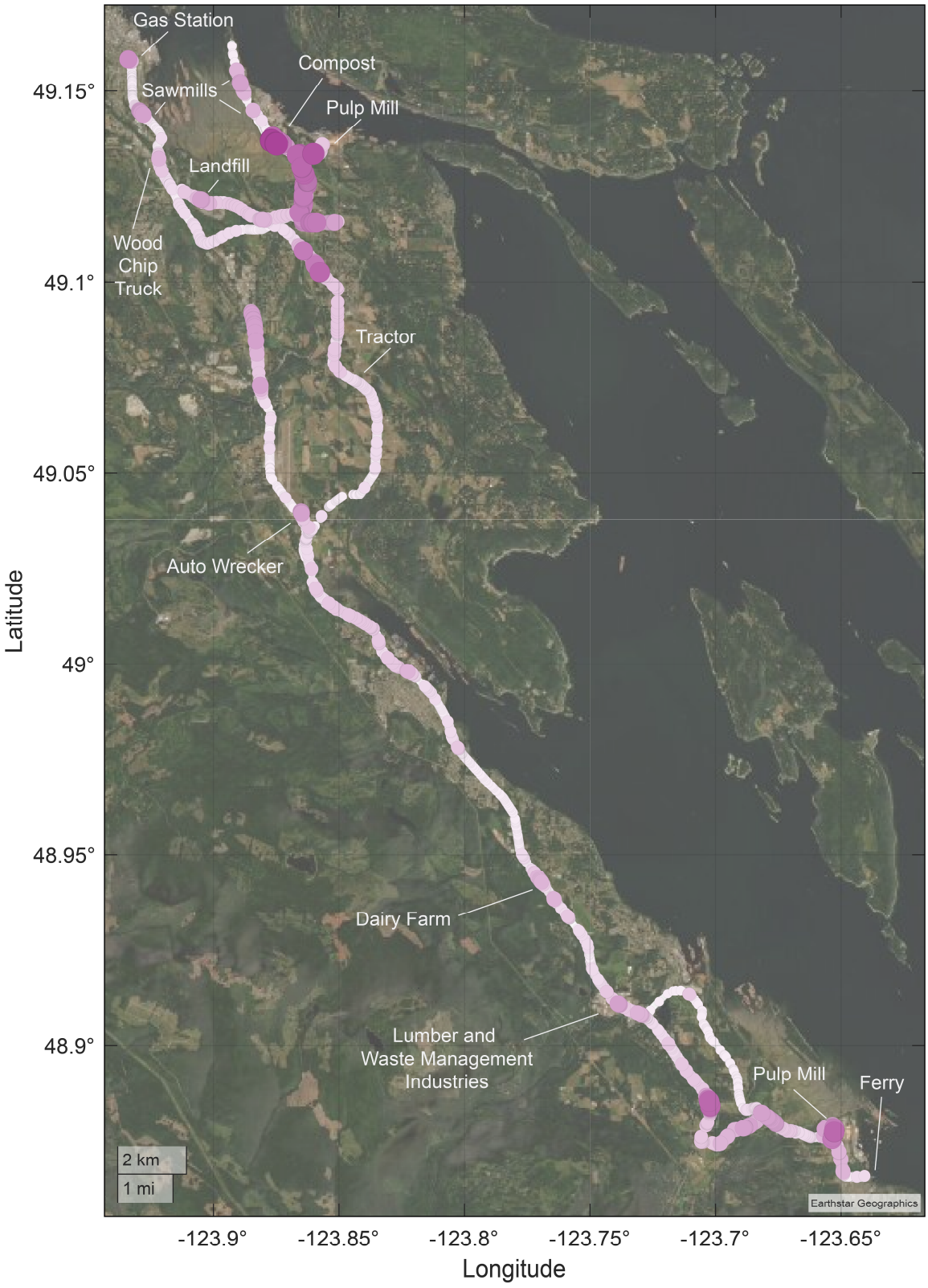


Figure 4.73: Geospatial distribution of the methanol component (Component 1 of 9) in the MCR-WALS model of the August 21 and 22, 2017 Nanaimo to Crofton data. Larger and darker dots indicate higher signal intensities. This component was measured in many locations including near the pulp mill, compost facility, dairy farm, and sawmills.

4.3.8.2 August 21 and 22, 2017 Component 2 of 9 – Anthropogenic hydrocarbon emissions

Component 2 is attributed to anthropogenic hydrocarbon emissions in the region, The mass spectrum contains measurements for many aromatic hydrocarbon species: BTEX (*m/z* 79.05, 93.06, 107.08), C9-aromatics (*m/z* 121.09), and C10-aromatics (*m/z* 135.12), as well as smaller compounds such as butene (*m/z* 57.07), pentene (*m/z* 71.08), and propene (*m/z* 43.05). As shown in the annotated time series (Figure 4.74) and map (Figure 4.75), measurements were predominantly made at a gas station while refueling the mobile lab, near an auto wrecking facility, and at a ferry terminal as shown in the map in Figure 4.75. Additional, smaller signals for this component are measured while driving and may be due to the presence of vehicle traffic. Two of the dominant sources are related to fugitive emissions (gas station and auto wrecking facility), while the signals measured at the ferry terminal may also be a combination of fugitive emissions from the ferry, and emissions from internal combustion engines.

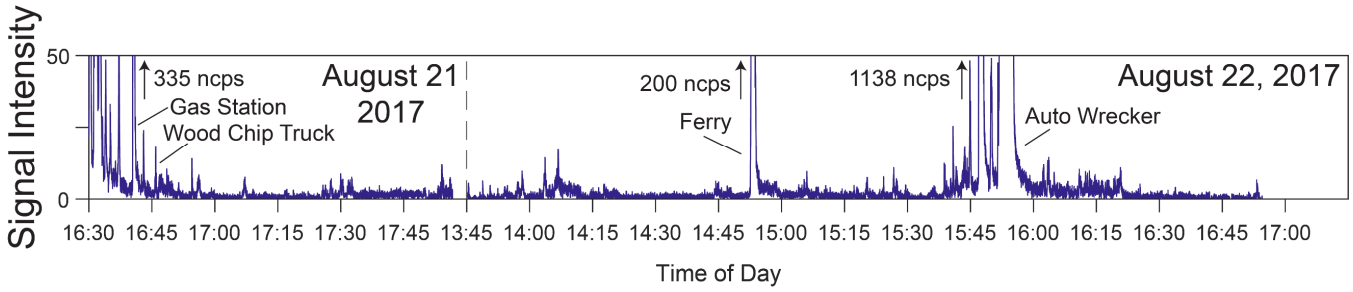


Figure 4.74: Annotated time series of the anthropogenic hydrocarbon emissions (Component 2 of 9) in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data. The vertical dotted line separates the August 21 and 22 data.

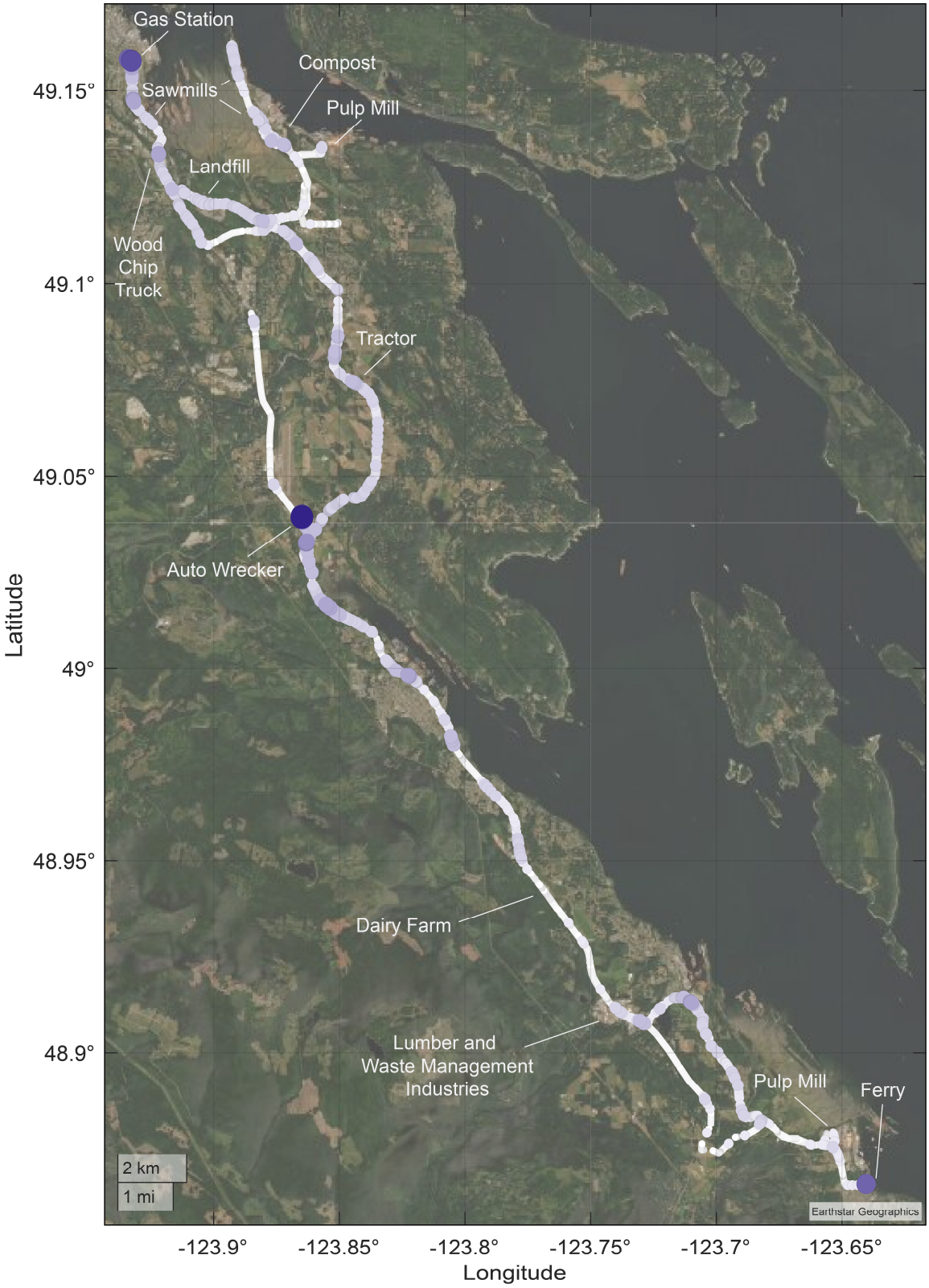


Figure 4.75: Geospatial distribution of the anthropogenic hydrocarbon emissions (Component 2 of 9) in the MCR-WALS model of the August 21 and 22, 2017 Nanaimo to Crofton data. Larger and darker dots indicate higher signal intensities. This component was measured at the ferry terminal, auto wrecker, gas station, and while driving.

4.3.8.3 August 21 and 22, 2017 Component 3 of 9 – Pulp mill emissions

Component 3 is associated with emissions from the two pulp mills on the drive route, with samples being collected in the vicinity of the pulp mill at the Northern end of the drive route on August 21, 2017, and near the pulp mill at the Southern end of the drive route on August 22, 2017. The emissions observed from the pulp mill at the South end of the drive route were higher during this field campaign as shown in the annotated time series (Figure 4.76) and on the map (Figure 4.77). The molecules contributing to this component are: methanol (*m/z* 33.03), methanethiol (*m/z* 49.01), ethanethiol/DMS (*m/z* 63.02), DMDS (*m/z* 95.00), and monoterpenes (*m/z* 137.12 and fragment ion at *m/z* 81.07). Emissions attributed to the pulp mill on the North end of the drive route were observed downwind from the source, approximately 2 km South of the pulp mill. These measurements are confirmed by the local air quality monitoring station (Harmac Cedar Woobank monitoring station, indicated by the white star on the map) which recorded hourly reduced sulphur concentrations of 1.6 ppb (17:00) and 3.8 ppb (18:00) on August 21, 2017. Both pulp mills were observed to emit similar compounds, and were identified in the same component, but the ratios of the VOCs detected differed between the two locations as shown in the mass spectra in Figure 4.78.

The molecules measured for this Component 3 are similar to the molecules measured for Component 3 during the August 14, 2018 field campaign when a reduced sulphur plume was observed downwind of the Northern pulp mill, however the ratios of the molecules in the mass spectrum differ between the models. Methanol and ethanethiol/DMS are present at higher relative abundance than the monoterpene and methanethiol signals during the August 2017 field campaign compared to the August 14, 2018 campaign.

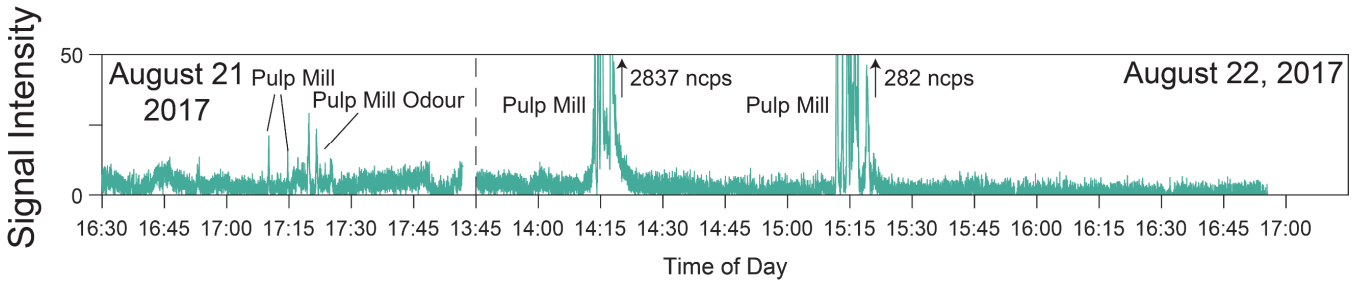


Figure 4.76: Annotated time series of the pulp mill emissions (Component 3 of 9) in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data. The vertical dotted line separates the August 21 and 22 data.

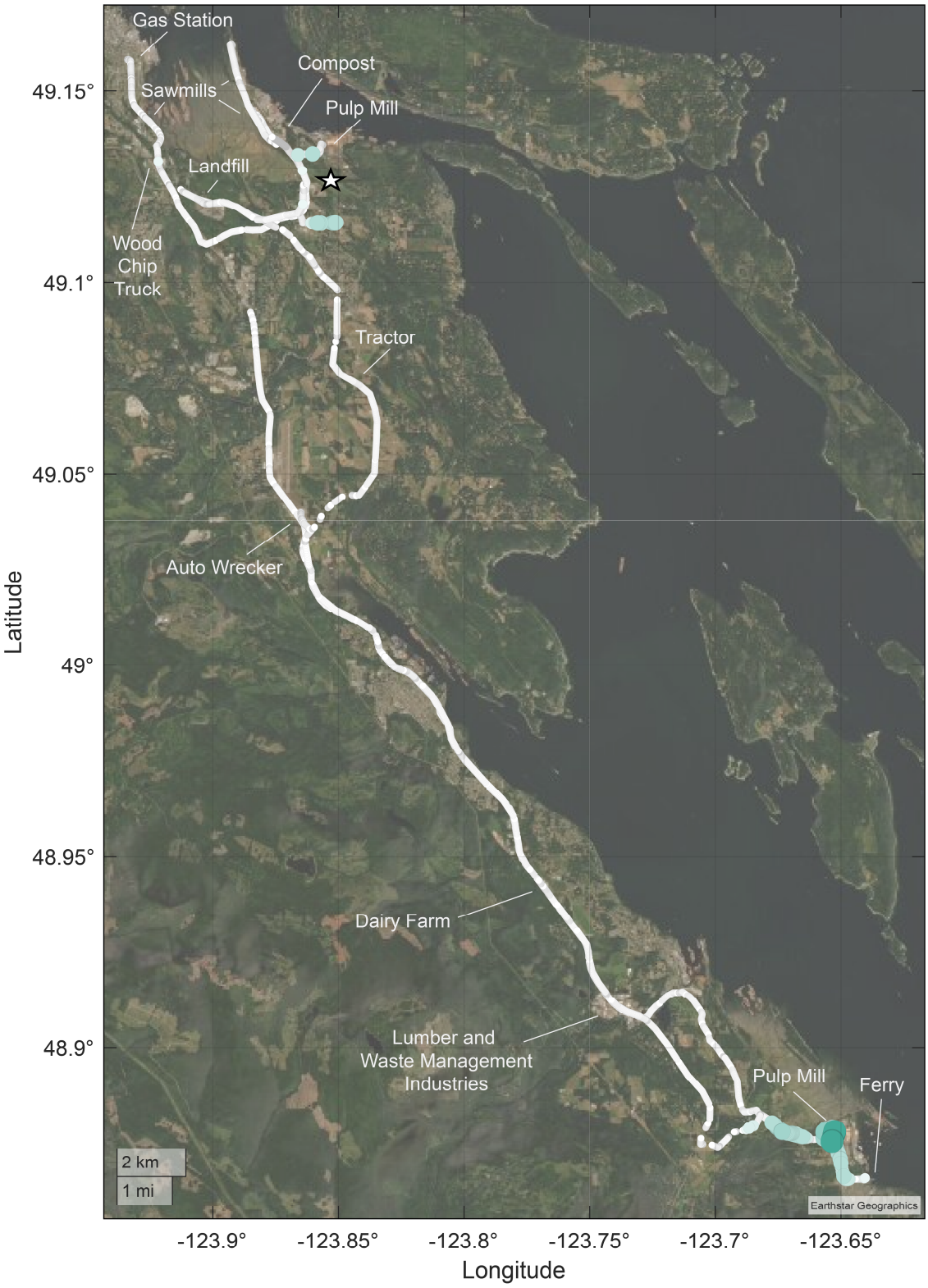


Figure 4.77: Geospatial distribution of the pulp mill emissions (Component 3 of 9) in the MCR-WALS model of the August 21 and 22, 2017 Nanaimo to Crofton data. Larger and darker dots indicate higher signal intensities. This component was measured near the pulp mills on the drive route. The white star indicates the location of the Harmac Cedar Woobank air quality monitoring station (hourly TRS measurements).

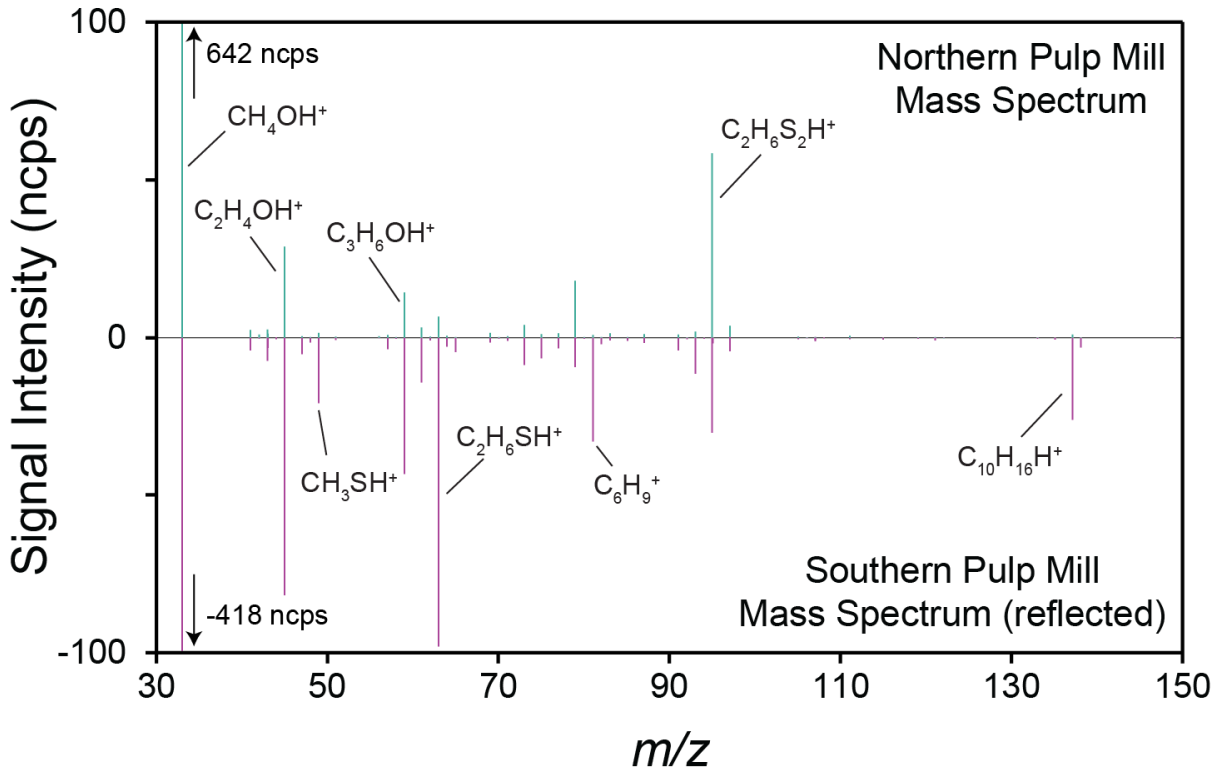


Figure 4.78: Top: Mass spectrum measured near the northern most pulp mill on the drive route. Bottom: Reflection of the mass spectrum measured near the southern most pulp mill on the drive route. Similar compounds were observed at the two pulp mills during this field campaign, but the ratios of compounds present different between the two locations.

4.3.8.4 August 21 and 22, 2017 Component 4 of 9 – Acetaldehyde

Component 4 is mainly attributed to measurements of acetaldehyde (*m/z* 45.03) not accounted for by the other components. As was found previously, the acetaldehyde component is predominately measured near biomass related industries (sawmills, composting facilities, pulp mills) with additional measurements detected near vehicles hauling these products. The annotated time series and geographic distribution of this component are shown in Figure 4.79 and Figure 4.80, respectively. The largest signal in the time series unique to this component is the signal measured near a wood storage area from 16:55–17:10 on August 21, 2017.

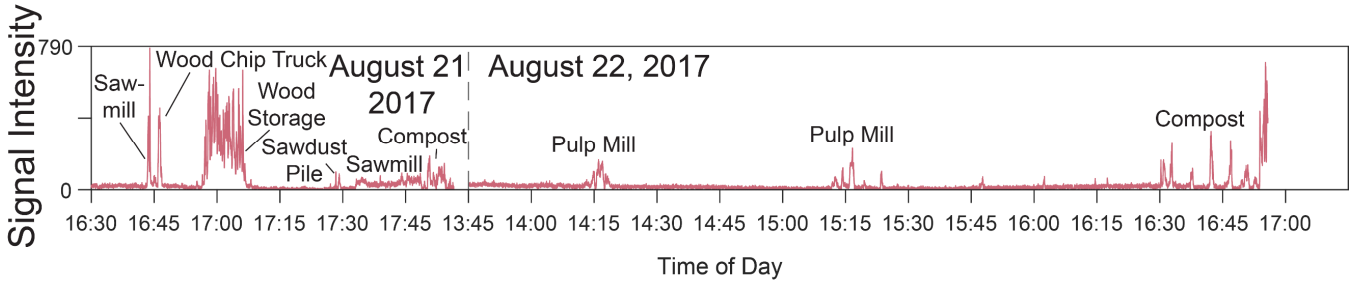


Figure 4.79: Annotated time series of the acetaldehyde component (Component 4 of 9) in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data. The vertical dotted line separates the August 21 and 22 data.

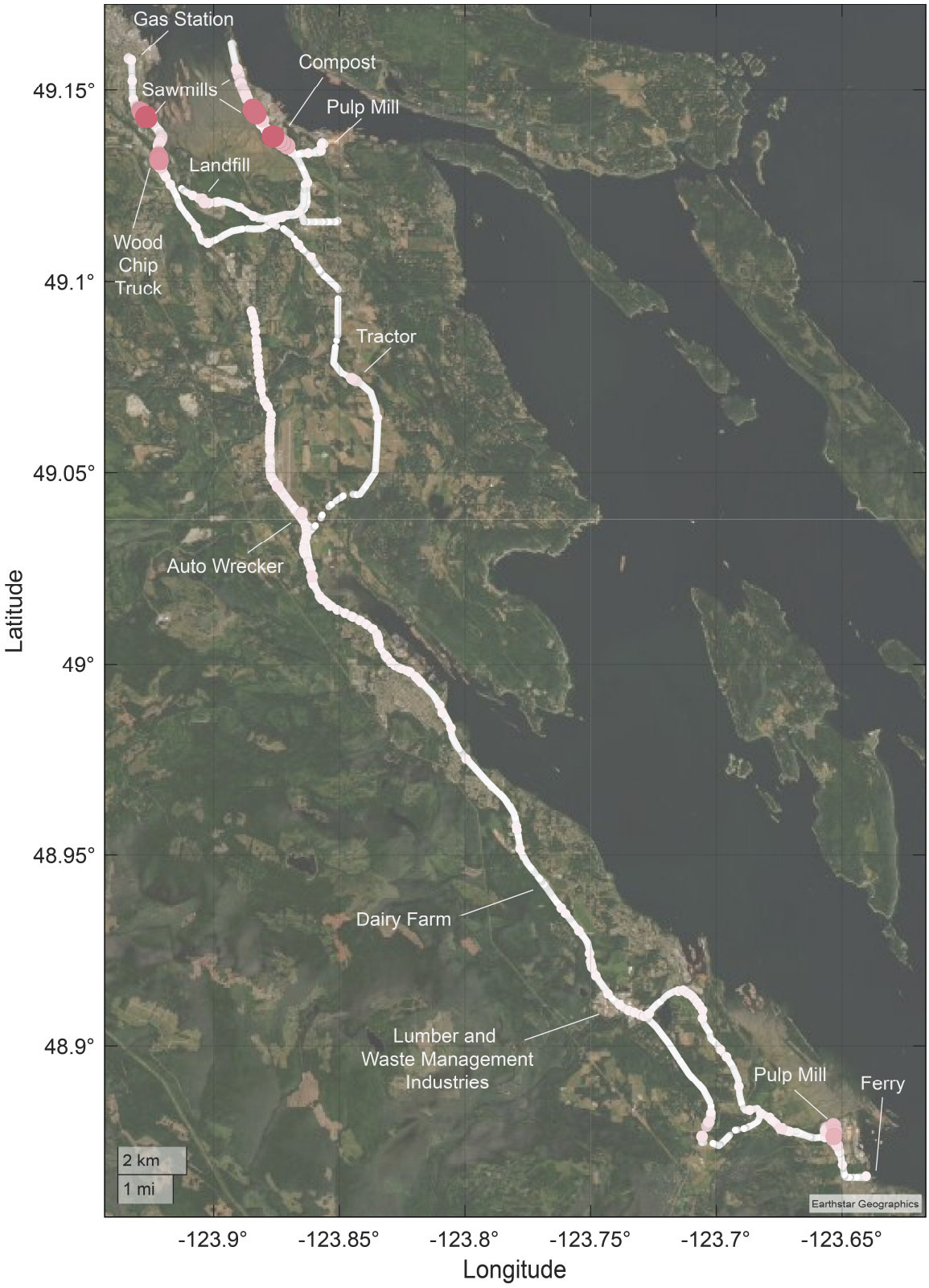


Figure 4.80: Geospatial distribution of the acetaldehyde component (Component 4 of 9) in the MCR-WALS model of the August 21 and 22, 2017 Nanaimo to Crofton data. Larger and darker dots indicate higher signal intensities. This component was measured near the sawmills, wood chip truck, and compost facility.

4.3.8.5 August 21 and 22, 2017 Component 5 of 9 – Compost

The regional composting facility is the main source of measurements associated with Component 5, as shown in the annotated time series (Figure 4.81) and map (Figure 4.82). The mass spectrum for this component contains signals attributed to many compounds including: acetaldehyde (*m/z* 45.03), formic acid (*m/z* 47.01), acetone/propanal (*m/z* 59.05), acetic acid (*m/z* 61.03 and fragment at *m/z* 43.01), and propanoic acid (*m/z* 75.04). This is the same regional composting facility that was passed on the August 14, 2018 drive route, and this component is similar to Component 7 (Compost - carboxylic acids) that was resolved for the August 14, 2018 data, in that it was only resolved near composting (2017 and 2018) and waste managements (2018) facilities, however the mass spectrum for these two components differ. For the data collected on August 21 and 22, 2017, the time series was truncated after leaving the composting facility, as the carryover in the sampling lines due to the more polar compounds measured impacted the PCA presented in Chapter 3, and that truncated dataset was used here as well. As a result, the MCR-WALS analysis of this dataset does not separate the composting signal into multiple components due to differences in decay times in the sample lines of the carboxylic acids compared to the other measured molecules.

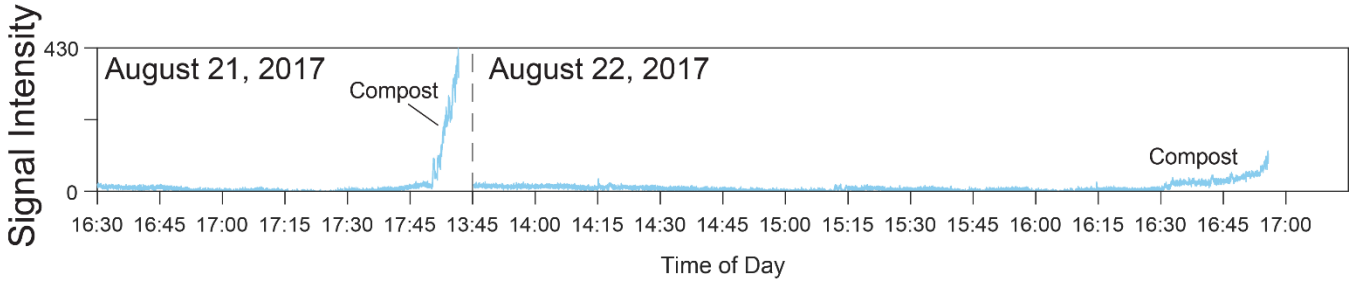


Figure 4.81: Annotated time series of the compost component (Component 5 of 9) in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data. The vertical dotted line separates the August 21 and 22 data.

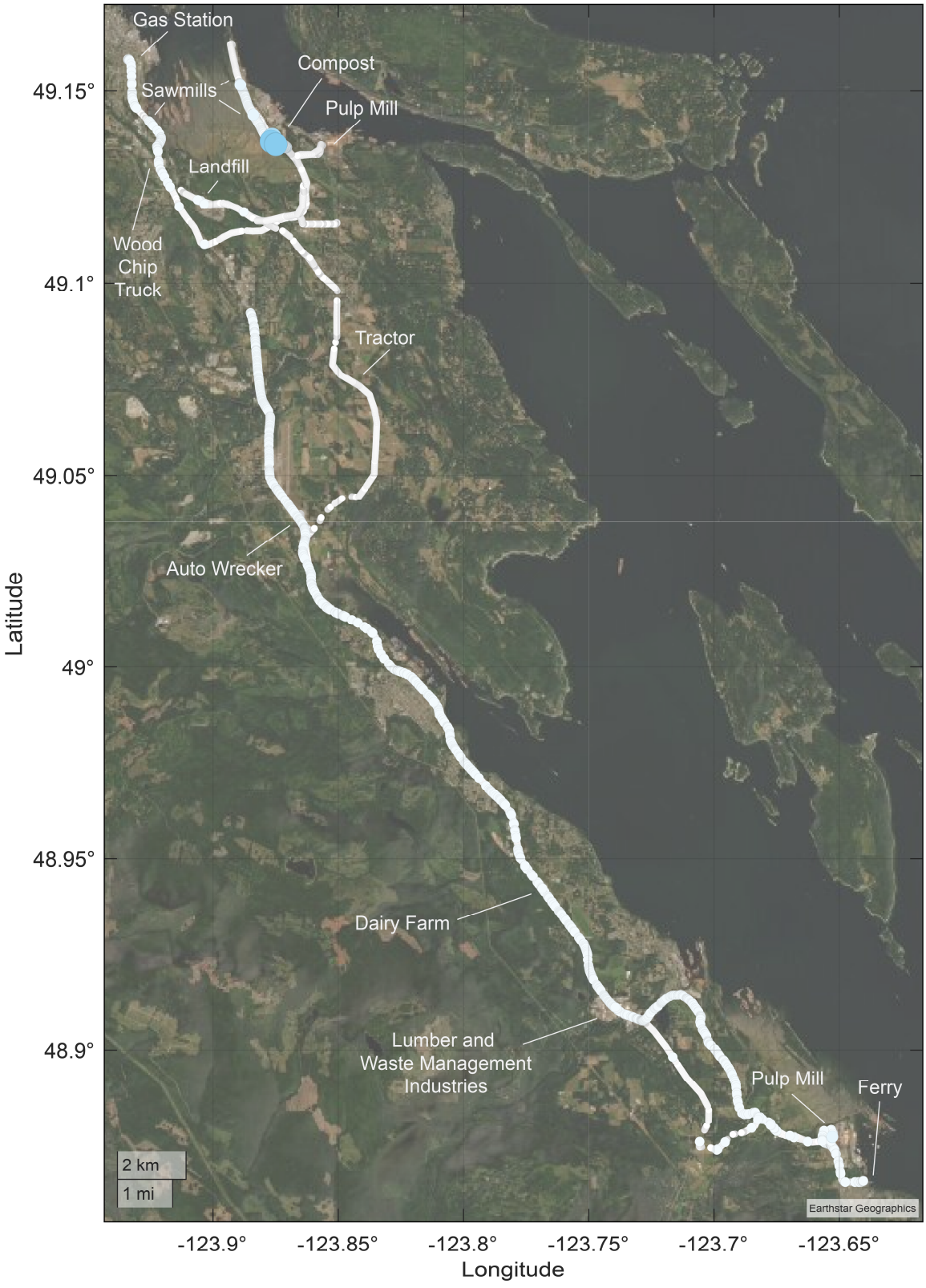


Figure 4.82: Geospatial distribution of the compost component (Component 5 of 9) in the MCR-WALS model of the August 21 and 22, 2017 Nanaimo to Crofton data. Larger and darker dots indicate higher signal intensities This component was measured near the compost facility.

4.3.8.6 August 21 and 22, 2017 Component 6 of 9 – Biomass

Similar to Component 4, Component 6 is associated with industrial activities that process biomass (sawmills, pulp mills, composting) as indicated in the annotated time series (Figure 4.83) and the geographic distribution shown in Figure 4.84. However, the compounds present in the two mass spectra differ significantly with the mass spectrum for Component 6 containing signals attributed to compounds including: acetaldehyde (*m/z* 45.03), ethanol (*m/z* 47.03), and monoterpenes (*m/z* 137.12 and 81.07), while the mass spectrum for Component 4 is mainly acetaldehyde. The major differences between the two time series are the measurements near wood storage that is present in Component 4 but not Component 6, and differences in the time series measured near the pulp mill at 14:15 on August 22, 2017. The measurements near the pulp mill were not homogenous between 14:14–14:20 with some molecules (*e.g.* acetaldehyde) being measured continuously over that time period, while others, such as the reduced sulphur compounds and monoterpenes, had higher signal intensities measured at 14:15 and 14:18. These differences are reflected in the time series of Components 1, 3, 4, and 6 which all have measurements associated with the pulp mill at the South end of the drive route. While the pulp mill emissions have been split across multiple components, it should be noted that pulp mills have multiple co-located VOCs sources, such as wood chip piles, transportation, pulp processing, and stack emissions, and the differences in the temporal profiles for the different components near the pulp mills is indicative of these different sources.

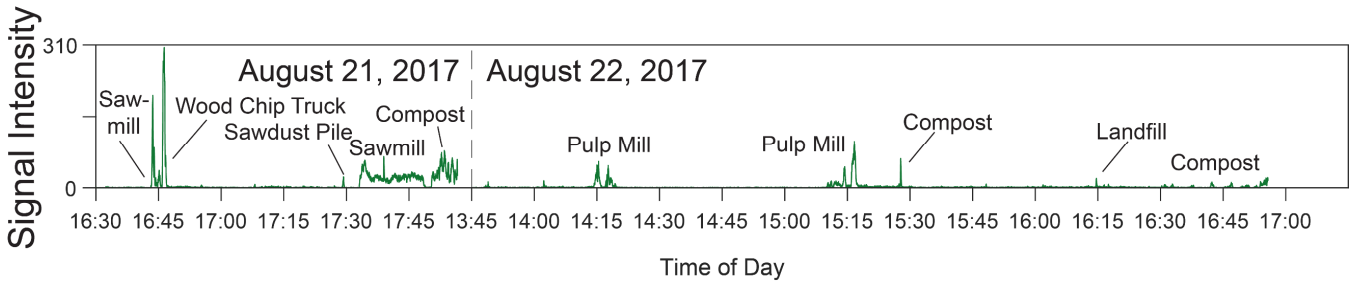


Figure 4.83: Annotated time series of the biomass component (Component 6 of 9) in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data. The vertical dotted line separates the August 21 and 22 data.

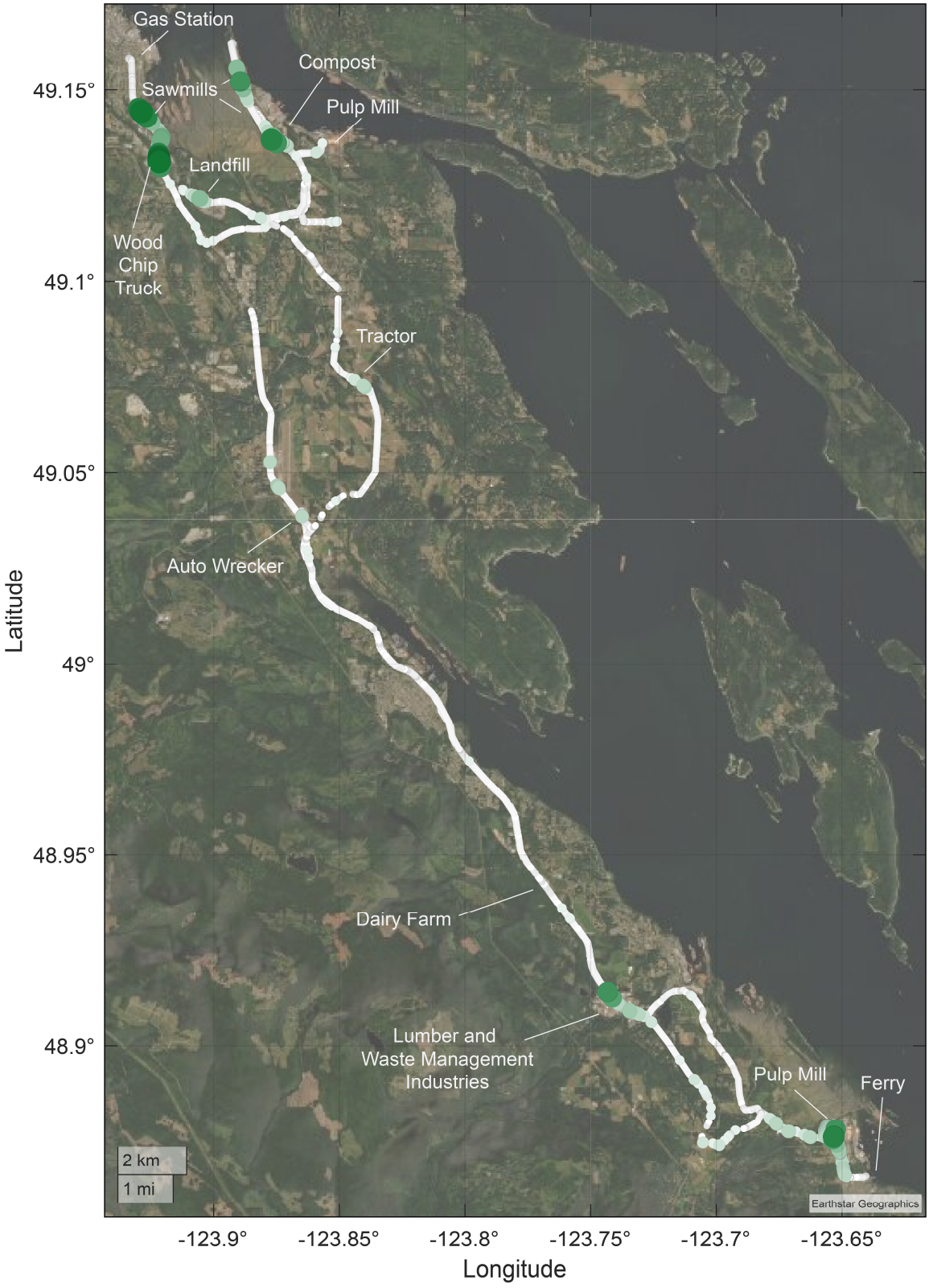


Figure 4.84: Geospatial distribution of the biomass component (Component 6 of 9) in the MCR-WALS model of the August 21 and 22, 2017 Nanaimo to Crofton data. Larger and darker dots indicate higher signal intensities. This component was measured near the industries that use wood products and the compost facility.

4.3.8.7 August 21 and 22, 2017 Component 7 of 9 – Small hydrocarbons and oxygenated species

Component 7 is measured near a variety of sources including the gas station, auto wrecking facility, composting facility, a tractor carrying hay, and the ferry terminal as shown in the annotated time series in Figure 4.85. The mass spectrum for this component contains measurements for both small hydrocarbons and small oxygenated species, as well as a large signal at *m/z* 41.03, which is a hydrocarbon fragment. The map of this component is shown in Figure 4.86.

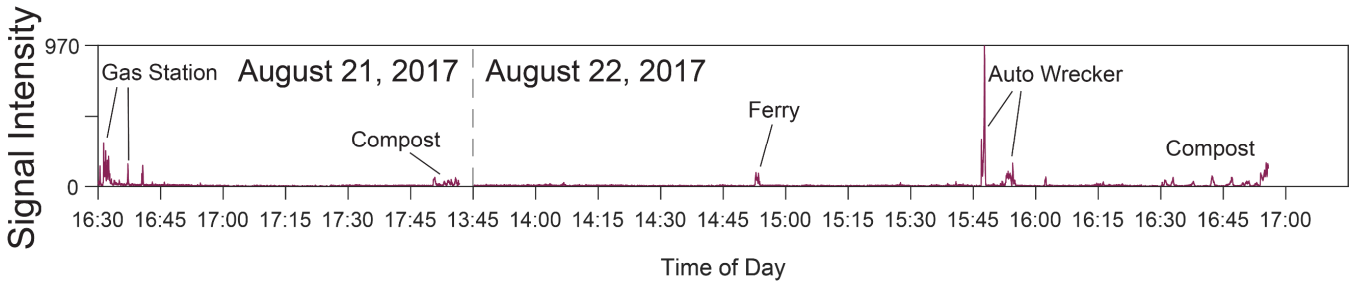


Figure 4.85: Annotated time series of the small hydrocarbon and oxygenated species component (Component 7 of 9) in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data. The vertical dotted line separates the August 21 and 22 data.

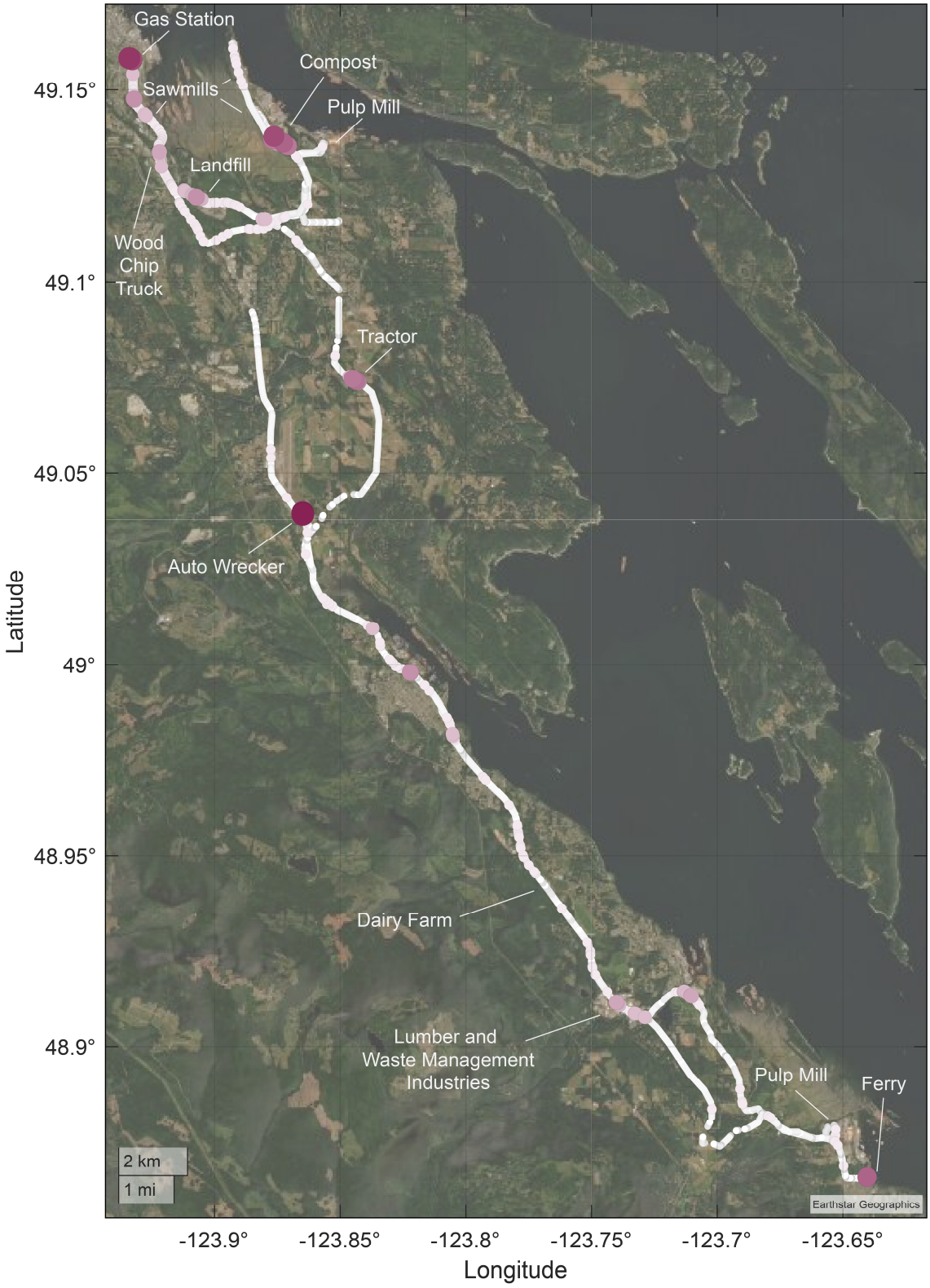


Figure 4.86: Geospatial distribution of the small hydrocarbons and oxygenated species component (Component 7 of 9) in the MCR-WALS model of the August 21 and 22, 2017 Nanaimo to Crofton data. Larger and darker dots indicate higher signal intensities. This component was measured in many locations including near the ferry terminal, auto wrecker, gas station, while passing a tractor, the compost facility, and landfill.

4.3.8.8 August 21 and 22, 2017 Component 8 of 9 – Acetone

Component 8 is measured in many short concentration excursions, predominantly on August 22, 2017. Although the mass spectrum contains signals for several VOCs, this component has been named acetone as it is predominantly associated with acetone measurements that have not been accounted for in the other components. As shown in the annotated time series in Figure 4.87 and map in Figure 4.88 it is measured near the compost facility, the pulp mill at the south end of the drive, the auto wrecker, and landfill. Additionally, it was measured from an unidentified source that was passed twice at 16:05 and 16:09 that could potentially be from a business in the area.

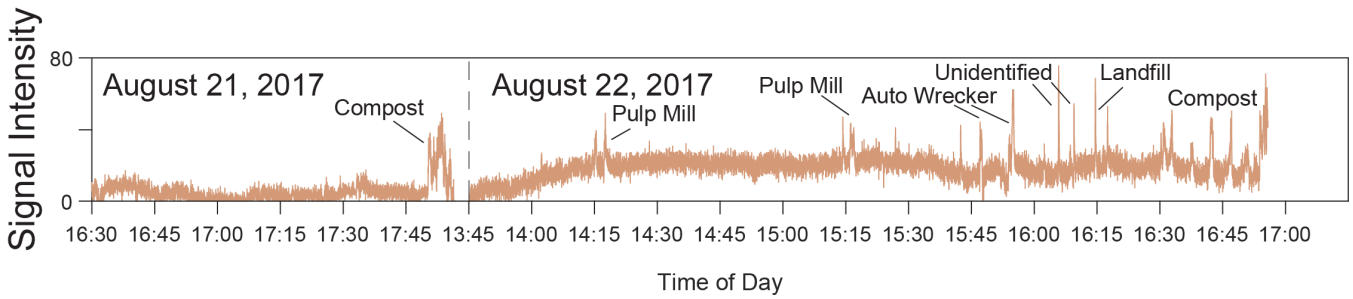


Figure 4.87: Annotated time series of the acetone component (Component 8 of 9) in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data. The vertical dotted line separates the August 21 and 22 data.

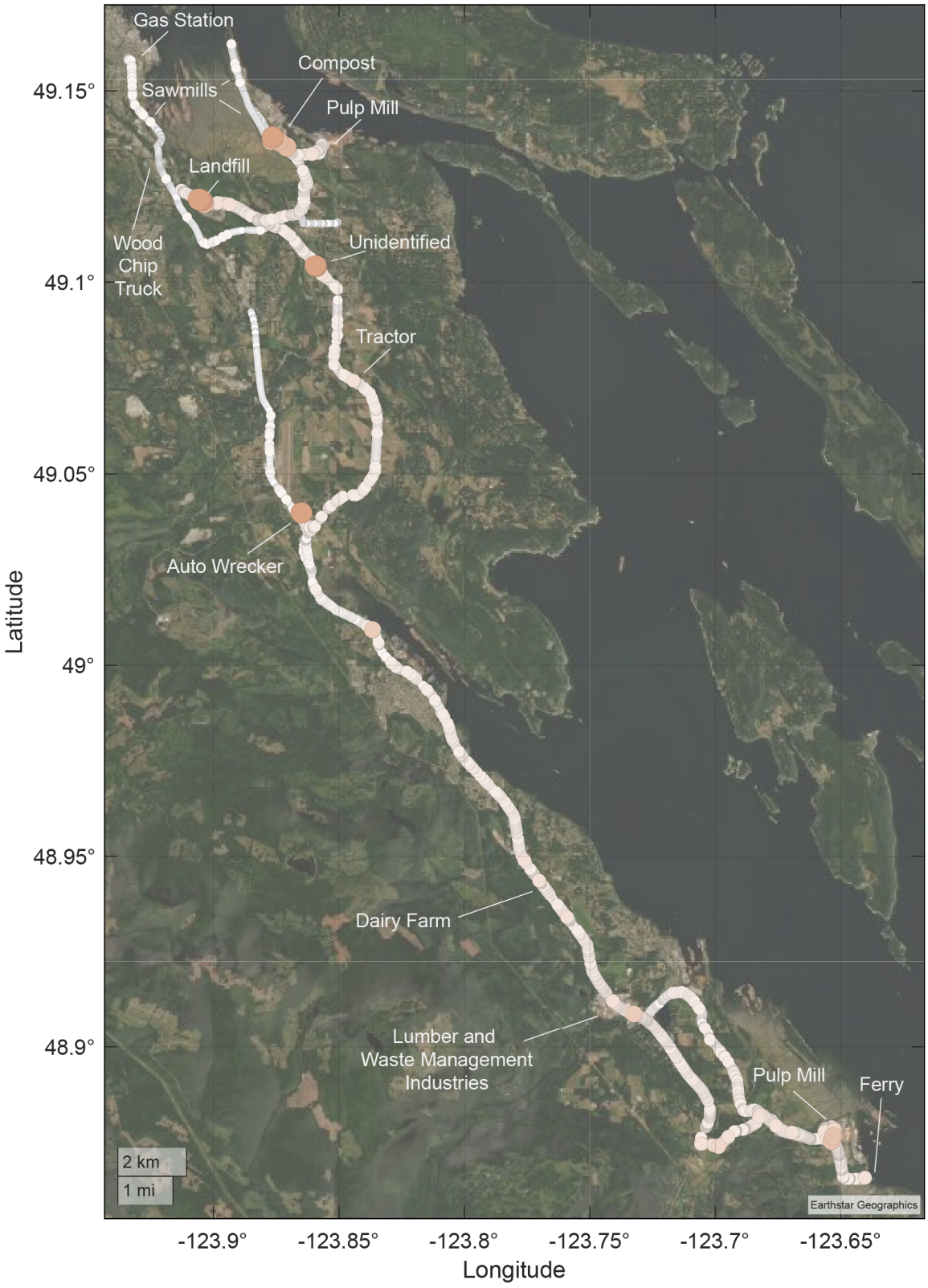


Figure 4.88: Geospatial distribution of the acetone component (Component 8 of 9) in the MCR-WALS model of the August 21 and 22, 2017 Nanaimo to Crofton data. Larger and darker dots indicate higher signal intensities. This component was measured near the auto wrecker, compost, land fill, and an unidentified point source.

4.3.8.9 August 21 and 22, 2017 Component 9 of 9 – Other

Component 9 does not have any temporal (Figure 4.89) or geographic (Figure 4.90) patterns that aid in identifying the source. However, there are polar compounds in the mass spectrum including: acetic acid (*m/z* 61.03), glycolic acid (*m/z* 77.02), and lactic acid (*m/z* 91.04), which may indicate that it is due to the photo-oxidation of less polar VOCs, or inlet effects due to the residence time of more polar compounds in the sample lines.

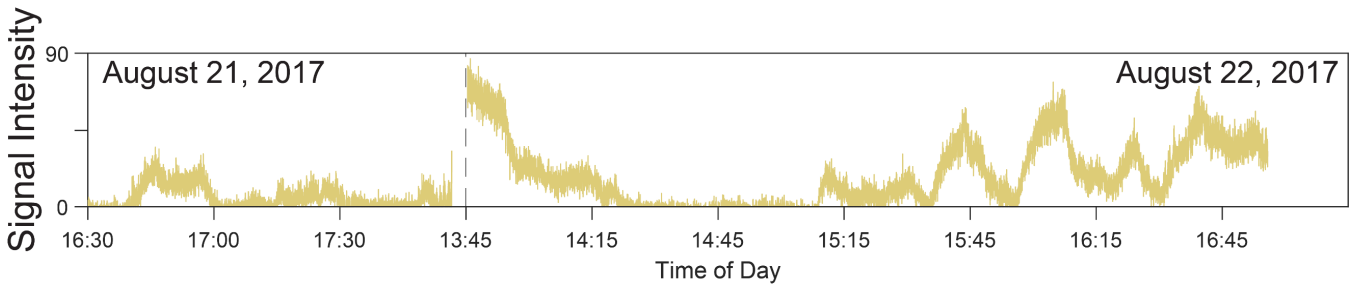


Figure 4.89: Annotated time series of the other component (Component 9 of 9) in the nine component MCR-WALS analysis of the August 21 and 22, 2017 Nanaimo to Crofton data.

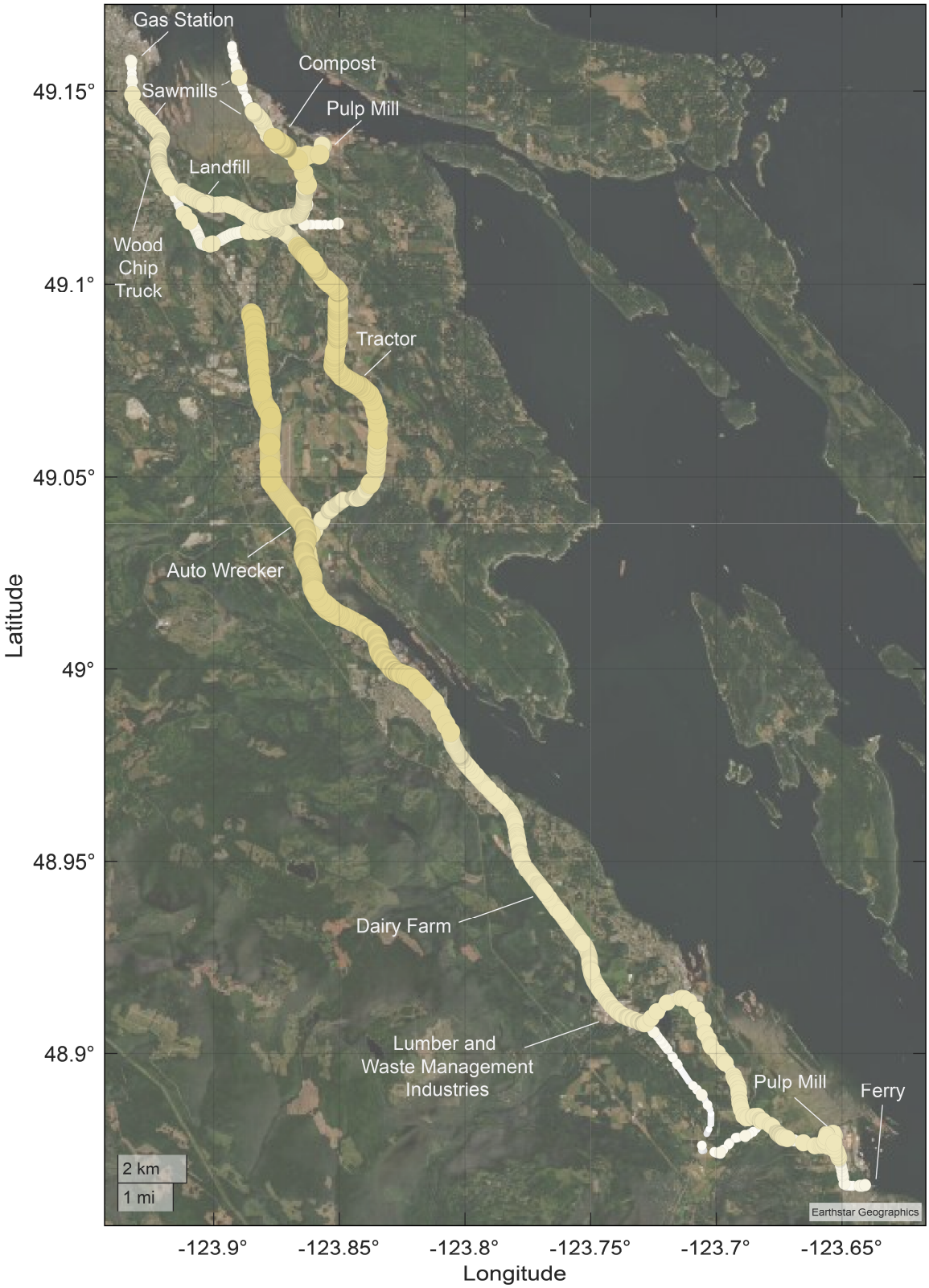


Figure 4.90: Geospatial distribution of the other component (Component 9 of 9) in the MCR-WALS model of the August 21 and 22, 2017 Nanaimo to Crofton data. Larger and darker dots indicate higher signal intensities. This component was measured throughout the region.

4.3.8.10 August 21 and 22, 2017 source apportionment

The pie chart of the source apportionment of the nine identified VOC components for the August 21 and 22, 2017 field campaign is shown in Figure 4.91. Once again, the components with the shorter concentration excursions (anthropogenic hydrocarbons, pulp mill emissions, compost facility, small hydrocarbons and oxygenated species) contribute less to the total observed VOCs, while those with more ubiquitous measurements (methanol, acetone, other) make up a large proportion.

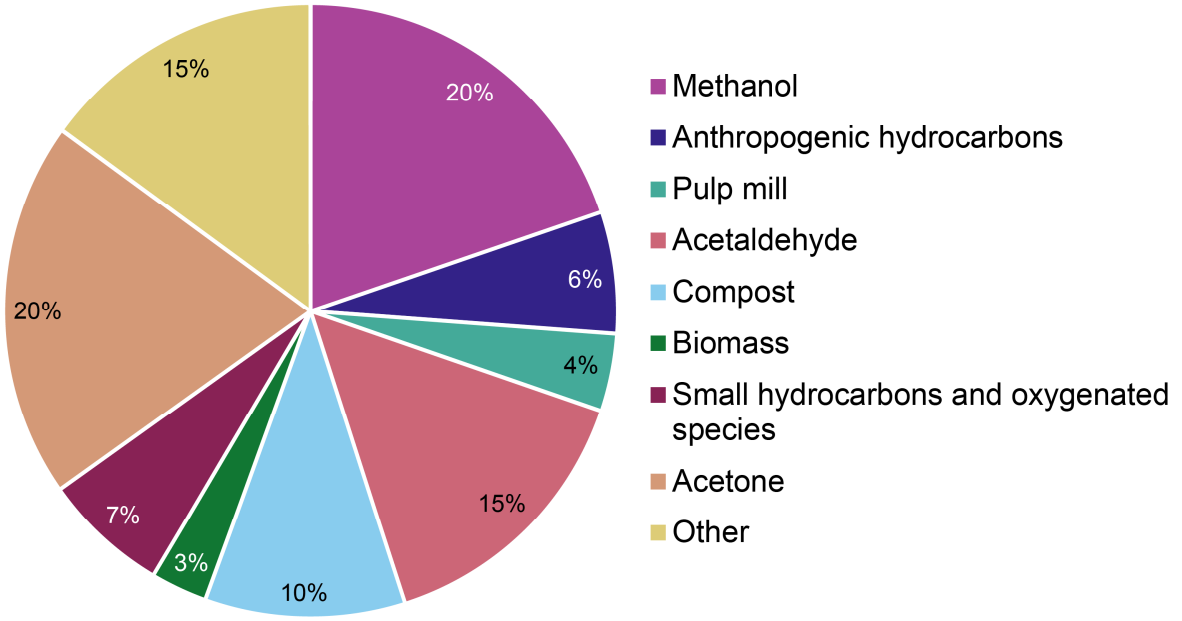


Figure 4.91: Pie chart showing the apportionment of each of the nine VOC sources observed over the course of the field campaign on August 21 and 22, 2017 between Nanaimo and Crofton, BC.

The temporal distribution of the apportionment is shown in Figure 4.92, with the TIC of the observed signal in the top panel, and the apportionment at each point in time in the bottom panel. From this plot, the change in the distribution of VOC source when sampling near point sources when higher signals were measured is evident, with anthropogenic hydrocarbon emissions making up a large percentage of the observed VOCs while at the gas station (August 21, 16:30–16:40), at the ferry terminal (August 22, 14:50), and at the auto wrecker (August 22, 15:45–15:55); acetaldehyde emissions from the wood storage area dominating between 16:55–17:05 on August 21, and biogenic emissions near the sawmill having large contributions on August 21 between 17:35 and 17:45. The emissions from the pulp mill at the south end of the drive route were split between multiple components, and this is evident in the number of components contributing to the measurements at 14:15–14:25 and 15:10–15:25 on August 22.

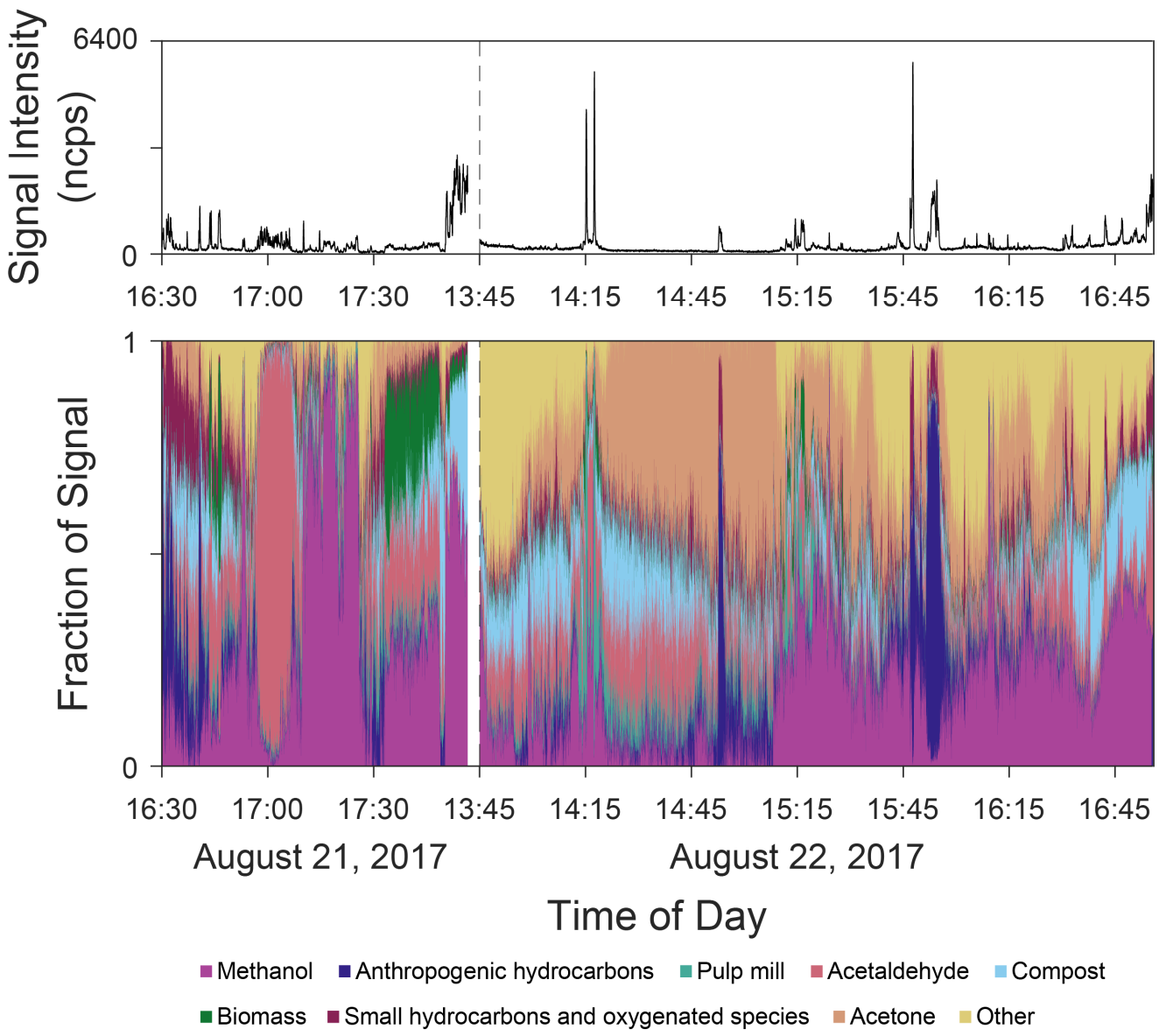


Figure 4.92: Top: TIC of the *m/z* included in the MCR-WALS analysis on August 21 and 22, 2017. Bottom: 100 % stacked area chart showing the distribution of the nine VOC components measured over the course of the field campaign on August 21 and 22, 2017. The vertical dotted line separates the August 21 and 22 data.

The geographic distribution of the apportionment is shown in Figure 4.93, with ring charts that are sized in proportion to the measured TIC at certain locations showing the distribution of components measured there. From this map it is evident that the mixture of VOCs present in the air can change dramatically over a short geographic area, and are very dependent on the local point sources.

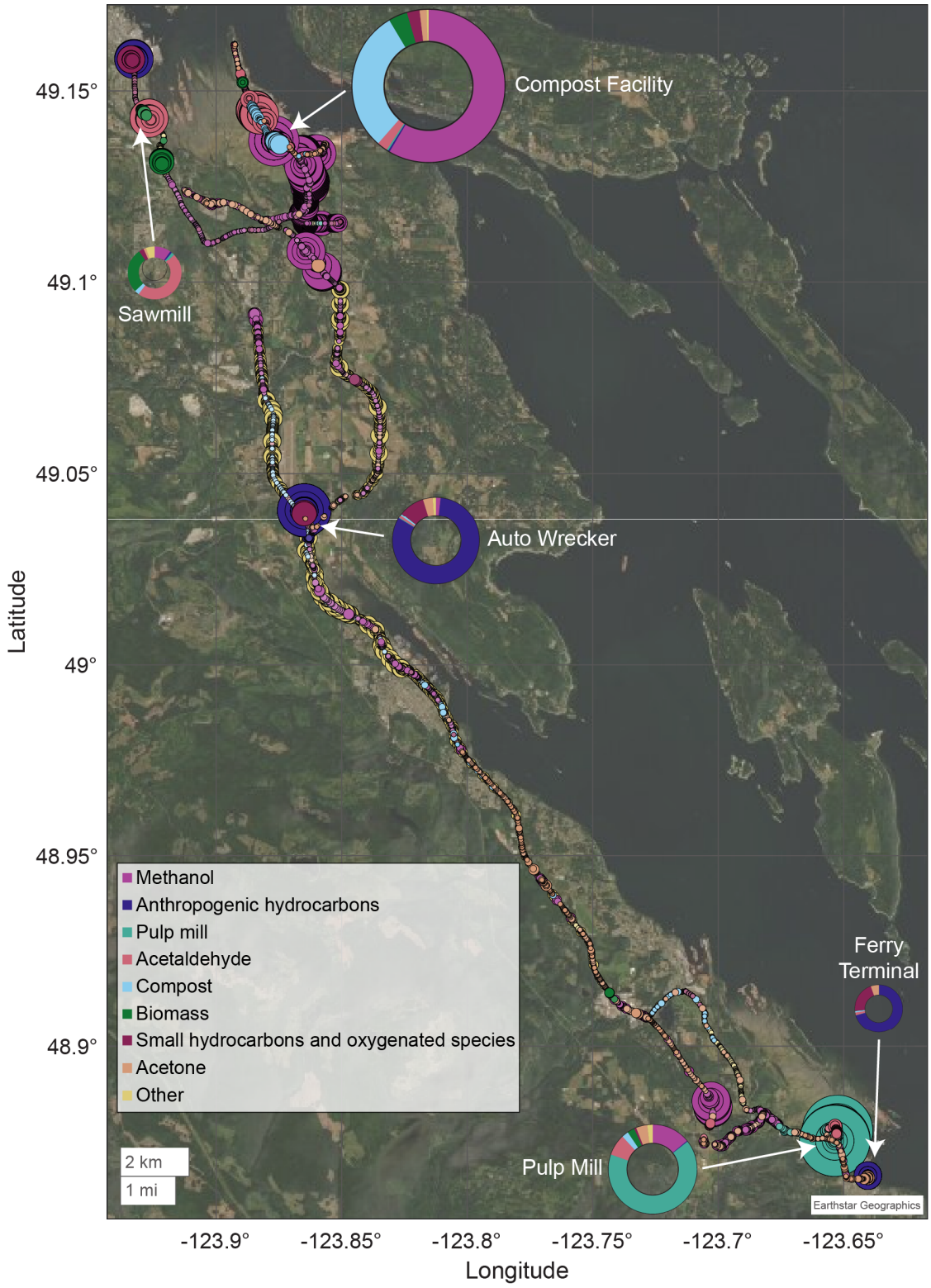


Figure 4.93: Maps of the apportionment of the VOC components measured on August 21 and 22, 2017. The dots that are visible on the maps each account for >30 % of the measured VOCs at a given location, and the larger dots describe higher signal intensities. Mapping the components together provides a visual snapshot of the distribution of VOCs present in the area. The four ring charts each describe distribution of the VOCs measured at the indicated areas, with the size of the ring proportional to the measured signal intensity at that location.

4.3.8.11 August 21 and 22, 2017 comparison between PCA (Chapter 3) and MCR-WALS

The Nanaimo to Crofton dataset analyzed here is the same data that was explored using PCA in Chapter 3. In the PCA of the complete dataset, samples were discriminated based on differences in their full scan mass spectra, with each mass spectrum (point in time) assigned to a group using two methods, a Gaussian mixing model (GMM) or density-based spatial clustering of applications with noise (DBSCAN). Using these methods samples collected near hydrocarbon sources (one group with GMM, 3 groups with DBSCAN), the southernmost pulp mill, the compost facility, and wood-based activities (two groups with GMM, one group with DBSCAN) were discriminated from one another. Before PCA, each mass spectrum was normalized to remove concentration effects, and PCA was very effective at discriminating the samples that had relatively high concentrations of VOCs before normalization. However, this resulted in some interesting VOC sources impacting samples to be missed, as the chemical species associated with them were not present at high enough relative abundance compared to the other ions measured in the same mass spectrum for them to be discriminated using PCA.

In the MCR-WALS analysis, each sample has a measured contribution from each of the nine resolved VOC components, allowing smaller contributions in the data that were missed by PCA to be identified. For example, the northern pulp mill is not distinguished in any meaningful way in the PCA analysis of the full dataset, while in the MCR-WALS analysis emissions from this pulp mill were observed both in the vicinity of, and downwind from, the source. Additionally, anthropogenic hydrocarbon emissions due to on-road traffic are resolved in the MCR-WALS analysis along with the other hydrocarbon sources, while this was not achieved through PCA. Also, there is one area with high signal intensity identified by PCA that was not resolved in the MCR-WALS analysis, which was a hydrocarbon sample measured at 15:43 on August 22, 2017. The mass spectrum for this sample has relative signal intensities for benzene/toluene/ethylbenzene of 0.12/0.02/1, which is different than the ratios in the resolved mass spectrum for the anthropogenic hydrocarbon component by MCR-WALS. The differences in the mass spectrum, along with the unmodelled signal at *m/z* 79.05 shown in the scatter plot in Figure C.10 indicate that an additional component is needed to model all the hydrocarbon VOCs observed in the data, however, none of the MCR-WALS models calculated (up to thirteen components) model the measurements at this location

Overall, the two techniques provide complementary information, with PCA identifying samples that are distinct from each other based on the overall mixture of VOCs observed in the sample, and MCR-WALS deconvoluting the measurements into the contributions from the individual sources (*e.g.*, pulp mill, compost) or source types (*e.g.*, anthropogenic hydrocarbons) observed at each location.

4.4 Successes and challenges

In this chapter the first use of MCR-WALS to identify and apportion the sources of VOCs measured by PTR-ToF-MS from a moving vehicle has been described. This was accomplished in complex ambient air samples impacted by multiple VOC sources at low ppb levels. The data was collected at one second intervals, leading to the distribution of VOC sources measured across three different field campaigns to be described at unprecedented temporal and spatial (10–30 meter) resolution. In each field campaign 8 or 9 VOC components were identified, with most being interpretable as being due to a single VOC source (*e.g.*, forest fire smoke, composting, pulp mill plume), or mixture of molecules common across multiple sources (*e.g.*, anthropogenic hydrocarbons). The use of variable selection (maximum S/N > 1) to reduce the number of *m/z* used in the analysis improved the results significantly, allowing for a higher number of interpretable VOC components to be identified using less complicated MCR-WALS models, while also significantly reducing computation time. The most interesting components that were identified are those relating to a single source that persist over a large geographic area where other point sources of VOCs were also encountered, such as the forest fire VOCs components measured on August 7 and 14, 2018, and the pulp mill emission plumes measured on August 14, 2018 and August 21, 2017. Samples impacted by these sources are more difficult to identify using the PCA method described in Chapter 3, as the relative abundance of the different sources observed in a sample have a large impact on how the samples are discriminated.

During the field campaigns some sources were encountered multiple times or for long periods of time, while others were only observed once. In general, sources with similar emission profiles (*e.g.*, anthropogenic hydrocarbon sources) that were encountered multiple times during a field campaign resulted in components with more physically interpretable sources. For example, on August 7, 2018 sources related to anthropogenic hydrocarbon emissions, forest fire VOCs, and biogenic emissions from vegetation were sampled multiple times over the course of the day, and had distinct time series with little overlap with the other resolved components. Sources emitting a mixture of compounds that were only encountered once or twice during a field campaign tended to have their contributions split across multiple components. For example, the emissions measured near the composting facility on August 14, 2018 were split across five of the eight resolved components due to the mixture of molecules encountered at that location, with only the carboxylic acid portion of the emissions appearing in a component distinct to the composting facility. In this chapter, data collected from three field campaigns across four days were presented, however additional data not discussed in this dissertation has been collected across the two sampling areas in the summer and winter. Combining multiple days of field data from one location together for analysis would produce a dataset with more replicate measurements of the VOC sources in the region, and may lead to more robust MCR-WALS components that can be identified as emissions

from an individual source. Additionally, some of the sources encountered (*e.g.*, pulp mills) are complex, with multiple VOC sources on-site and dynamic processes leading to differences in emissions profiles that lead to intra-sample variability. By including data collected across multiple field campaigns, these complex, multi-source industries could be better characterized by capturing the inter-day and intra-day variability in the emissions.

Sampling at a frequency 1 Hz allows for data with high temporal and geospatial resolution to be captured, and ensures that VOC emissions from short duration concentration excursions that are encountered while driving are recorded in close proximity to their source. However, sampling at such a high frequency lowers the sensitivity of the PTR-ToF-MS, and produces data with a higher noise level compared to if longer integration times (*e.g.*, 10 seconds for future field work) or sample averaging before analysis were used. Additionally, measuring a mass spectrum at 1 second intervals led to differences in the measured time series for compounds of differing polarities due to sample line and inlet effects, which in the case of the measurements made near the composting facility on August 14, 2018 resulted in the emissions being split over multiple components. It was also observed that very high concentrations measurements of VOC present in multiple components can lead to artifacts in the time series in the component with lower signal intensities. This can be seen in the time series for Components 2 and 6 in the August 14, 2018 results (Figure 4.51). High, short duration signal intensities are observed for Component 2 (anthropogenic hydrocarbons) in the vicinity of the auto wrecking facility, which correlate to drops in signal intensity in the time series of Component 6 (forest fire VOCs). Sample averaging, longer integration times (on future field campaigns), or further down-weighting high intensity signals may help prevent this from occurring.

In the source apportionment for each field campaign, the components with the least structure in their time series (Components 7 and 8 on August 7, 2018; Components 6 and 8 on August 14, 2018; and Component 8 on August 21 and 22, 2017) represent a large proportion of the observed VOCs on each day. The apportionment was done by calculating a regression vector between the time series for the components and the observed TIC on each day, and in many cases the regression coefficients for the components with less structure was higher than for those of the components with the more interesting temporal profiles. In examining the time series of the measured versus modelled signal intensities for the *m/z* included in the models, in some cases the modelled data agrees very well with the measured data, while in other cases the modelled data captures the temporal profile of the data, but not necessarily the noise observed in the measurements. An example of this is shown in Figure 4.94, where the isoprene signal in the top panel is modelled well including the noise in the measurement, while for the benzene signal in the bottom panel the modelled signal intensity follows the temporal trend in the data, but does not capture the entirety of the measured signal. This leads to higher regression coefficients for

the less interesting sources in order to reconstruct the observations. Once again, a reduction in noise levels achieved through sample averaging or longer integration times (in future work) could help reduce the overrepresentation of these components in the overall apportionment.

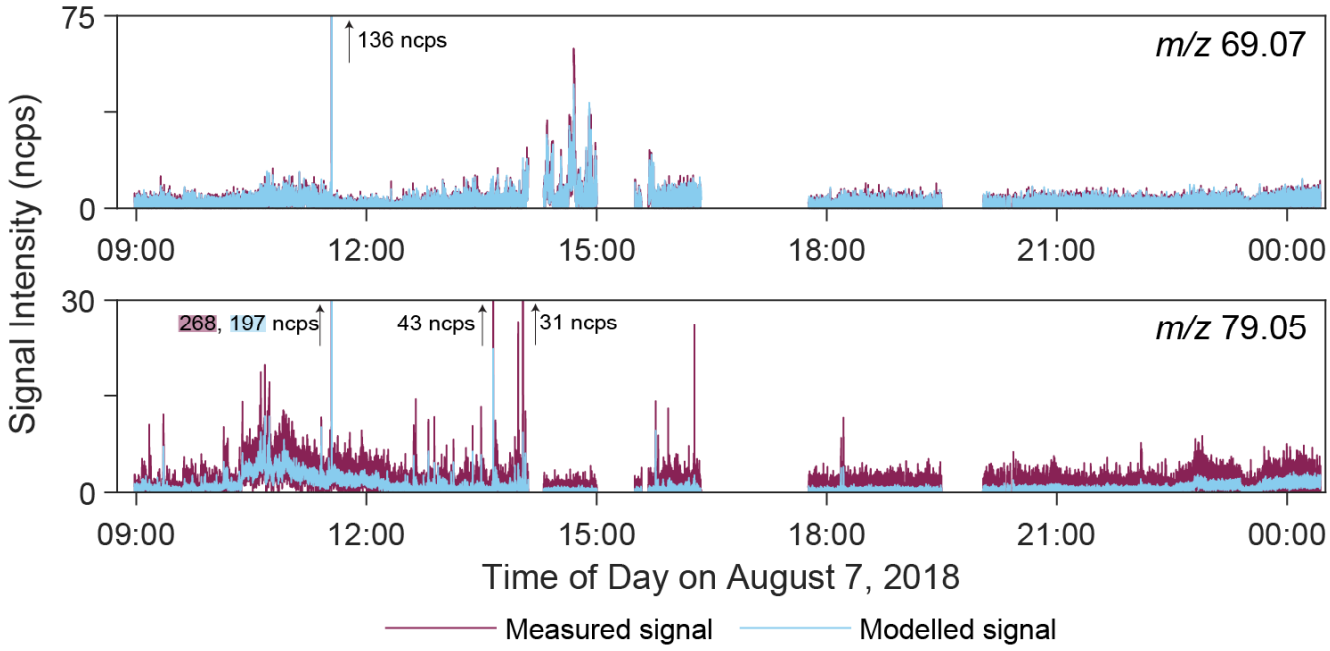


Figure 4.94: Top: Overlay of the measured and modelled signal intensity for *m/z* 69.07 on August 7, 2018. The eight component MCR-WALS model measures the signal and noise of the data well. Bottom: Overlay of the measured and modelled signal intensity for *m/z* 79.05 on August 7, 2018. The eight component MCR-WALS model measures the temporal trends in the signal, but does not capture all the noise in the measurements.

4.5 Conclusions

Using continuous, real-time mass spectral data collected from a PTR-ToF-MS system operated in a moving vehicle atmospheric VOCs can be measured over time and space at the neighborhood to regional scale. These mass spectral datasets contain a wealth of information relating to the mixture of VOCs present at each measured time and location. The use of receptor models to identify different VOC sources, both point sources, such as industrial emissions, or more diffuse area sources, such as smoke from forest fires was explored. The use of PMF for this type of analysis is common in atmospheric chemistry. In this chapter, several approaches to receptor modeling were evaluated. MCR-ALS and MCR-WALS were compared to PMF using a reduced dataset with one minute sample averages and variable selection (*m/z* with average S/N >1). From this analysis, it was found that MCR-WALS provides comparable results to PMF for the PTR-ToF-MS data. When analyzing dataset where a mass spectrum was collected every second, it was found that selecting only the *m/z* in the full scan data that had a maximum S/N > 1 over the course of the day provided a quick method for variable screening that improved the model results. While this method differed from the variable selection commonly employed in PMF

(variables with average S/N > 1 are used as it), due to the nature of the data collected during the field campaigns a lower threshold for including variable was required to retain enough *m/z* to capture the variation in the mixture of VOCs observed over time and space.

The results from the MCR-WALS analysis also allowed for the observed VOCs to be apportioned, characterizing the relative contribution of the different sources to individual samples, and across each field campaign. As each measurement was time and location stamped, the time series for each of the resolved VOC components can be mapped. In addition to aiding in the identification of the components, this is a powerful visualization of the geospatial distribution of VOC sources and their relative contributions. Many of the resolved sources were measured as short duration concentration excursions, particularly the anthropogenic hydrocarbon source for each sampling day, but others, such as the VOCs from forest fire smoke, and a sulphur plume associated with a pulp mill, were dispersed over a large area. These high-resolution maps can provide important insights into the emission sources impacting an air shed, and at a much higher temporal resolution than currently provided by government air quality monitoring stations.

Chapter 5 - Conclusions and future work

Anthropogenic emissions into the troposphere can impact air quality, leading to poorer health outcomes in the affected areas. VOCs are a group of chemical compounds that can be detrimental to air quality but have historically been difficult to measure in ambient air due to their low concentrations (ppt_v–ppb_v), and the large number of chemical compounds that fall under the umbrella of 'VOC'. However, with improvements in technology, in particular the miniaturization and ruggedization of mass spectrometers, ambient VOC measurements are becoming more commonplace. In particular, MS techniques such as MIMS and PTR-ToF-MS can be used to continuously measure VOCs in ambient air, and when operated in a moving vehicle can measure the distribution of VOCs over space and time in an area. As such, direct sampling MS is an effective tool for producing chemical fingerprints of samples, which can be interrogated for source discrimination, classification, identification, and apportionment using chemometric techniques. In this dissertation, the series of work presented described the progression from the use of PCA to discriminate lab-based samples of known and unknown chemical composition based on full scan MIMS data, to the use of PCA and clustering algorithms to discriminate ambient VOCs samples using MIMS and PTR-ToF-MS data collected from a moving vehicle, to using MCR-WALS to identify and apportion the sources of ambient VOCs over time and space in two regions on Vancouver Island, BC.

5.1 Key results from Chapter 2

In Chapter 2, chemometric techniques were applied to the full scan MIMS data of VOC mixtures for sample discrimination, classification, and mixture resolution. Three hundred samples containing 32 different VOC mixtures constructed from ten pure compounds at different concentrations were successfully discriminated using PCA. Additionally, a hierarchical clustering algorithm was used to group samples, a kNN classification model was built using 202 samples and used to successfully classify 95 of 96 samples in the test set, and MCR-ALS was used to resolve the mixtures into their pure component contributions. These lab-constructed samples contained mixtures of known chemical composition.

The full scan MIMS data of VOC mixtures measured in the woodsmoke emissions from different species of wood and from the headspace above different aqueous hydrocarbon solutions were also analyzed by PCA. These datasets were 'real-world' samples from known sources with unknown molecular compositions. In both cases, PCA successfully discriminated the different VOC sources based on the differences in the full scan MIMS data.

The results described in Chapter 2 demonstrated that the chemometric analysis of full scan MIMS data of trace VOCs in air can be used to discriminate and classify samples and were an essential first step towards using these methods for source discrimination, identification, and apportionment of atmospheric VOCs measured in the field using a mobile mass spectrometry system.

5.2 Key results from Chapter 3

In Chapter 3, the methods explored in Chapter 2 are extended to full scan mass spectral data of ambient atmospheric VOCs collected using vehicle mounted MIMS and PTR-ToF-MS systems. Samples were collected on-road and both a supervised (for the MIMS and PTR-ToF-MS data) and unsupervised (for the PTR-ToF-MS data) approach was used for sample discrimination. For the supervised analysis, PCA was applied to normalized full scan mass spectra data collected near known VOC sources. Samples containing emissions from sources such as internal combustion engines, sawmills, composting facilities, and pulp mills were discriminated based on the mixture of compounds present in the full scan MS data.

For the unsupervised analysis, PCA was applied to the continuous PTR-ToF-MS dataset measured over two days between Nanaimo and Crofton, BC. After PCA, two methods, GMM and DBSCAN, were used to identify clusters in the PCA results. For both the supervised and unsupervised PCA analysis of PTR-ToF-MS data, geospatial mapping of the results showed the geospatial distribution of the different sources encountered during the field campaign.

The results presented in Chapter 3 demonstrated that the molecular level information provided by mobile mass spectrometry allow ambient VOC sources to be discriminated. The mass spectrometry techniques employed produce continuous time series data with high temporal (seconds) and spatial (10s of meters) resolution, allowing rapid changes in the mixture of atmospheric VOCs to be identified and their distribution to be mapped at the neighbourhood and regional level.

5.3 Key results from Chapter 4

The work described in Chapter 3 allowed ambient VOC mixtures measured from a moving vehicle to be discriminated based on the observed mixture of molecules in full scan mass spectrometry data. When sampling in the real-world however, there are often multiple sources contributing to the VOCs present in the air at a given time and location, which can be unmixed using receptor models. In order to identify and apportion the different sources of VOCs, MCR-WALS was applied to the continuous time series PTR-ToF-MS data from three field campaigns on Vancouver Island, BC.

Before analysis by MCR-WALS, variable selection was applied to the data to remove the recorded *m/z* that contained no measurements with S/N greater than one. Reducing the mass list in this way retained enough of the observed *m/z* to capture changes in the mixture of VOCs measured, while improving the model results and reducing the computation time of the algorithm.

MCR-WALS identified both point (*e.g.*, anthropogenic hydrocarbon emissions, pulp mill emissions) and diffuse (*e.g.*, VOCs from forest fire smoke) VOC sources in the data. The observed VOC sources were apportioned to determine the contribution of each source to the observations made during a field campaign, near specific locations, and at each point in time.

The data analyzed was collected at 1 Hz capturing fine scale changes in the ambient VOCs present in the air. Geospatial maps of each individual source, and of the source apportionment, were used to visualize the distribution of VOC sources across the sampling area.

5.4 Research achievements and future directions

Progress on the research questions posed in Section 1.5 have been substantially addressed in the experimental work. It is now clear that chemometric techniques can be used to resolve components and apportion sources derived from temporally and spatially resolved ambient VOC datasets measured on-road. Although the available datasets are high-resolution snapshots of particular dates and locations, they nonetheless have provided robust datasets to explore and resolve essential questions that underlie this approach to ground-level atmospheric chemistry monitoring and research. PCA enabled different mixtures of trace level VOCs, both in lab-constructed samples (MIMS) and captured in field data (MIMS and PTR-ToF-MS), to be discriminated. kNN was used to classify constructed VOC mixtures (MIMS). MCR-ALS resolved constructed VOC mixtures (MIMS) into their individual chemical components. GMM and DBSCAN identified groups of samples impacted by similar VOC sources in a PCA model of continuous, time series mass spectral data (PTR-ToF-MS) collected during a field campaign. Finally, MCR-WALS was used to identify and apportion the sources of VOCs observed while field sampling (PTR-ToF-MS).

In Chapters 3 and 4 different approaches were used to prepare the data before chemometric analysis was applied to the continuous, time series PTR-ToF-MS measurements. In Chapter 3, no variable selection was applied to the data before PCA and the uncertainty in the data was not taken into account during the analysis. When using PCA, the best results are achieved when the errors in the measurements are identical, independent, and normally distributed (i.i.d.),[251] which is not the case for the PTR-ToF-MS data. For future applications on similar datasets, the use of maximum likelihood PCA (MLPCA) which, like MCR-WALS, incorporates

measurement uncertainty into the analysis may improve results.[251] Additionally, as many of the recorded *m/z* contained high levels of noise, using the variable selection method in Chapter 4 before PCA may also further improve the results by significantly reducing the level of noise in the dataset being analyzed.

In both Chapter 3 and 4 the continuous time series PTR-ToF-MS data was collected using a one second integration time, and PCA and MCR-WALS were applied to the dataset without prior sample averaging. This produced results with high temporal and spatial resolution, with sample discrimination and source apportionment being done on samples collected every 10–25 meters (depending on drive speed). While this enables very short duration signals to be identified, these high frequency measurements reduce the signal/noise ratio and generate very large datasets (28,800 mass spectra collected in 8 hours of field sampling) which can significantly increase computation times. Averaging the data across 10–30 second time bins, or using longer instrument integration times when collecting data in future field campaigns, would decrease the spatial resolution to 100–750 meters, but would decrease detection limits and improve computation times. Additionally, this sample averaging would facilitate applying MCR-WALS and PCA to multi-day datasets, which will capture the inter-day variability of the sources encountered and may provide a more representative summary of VOCs in a region. Additionally modifications to the sampling interface of the mobile lab, including a high volume, heated, sampling manifold, will aid in minimizing the carryover that was observed near certain sources.

The research described in this dissertation has many logical extensions. The data presented here provided a snapshot of the VOCs sources impacting two regions on Vancouver Island. Longer term studies over multiple days would allow for regional and seasonal trends to be identified, as well as potential local 'hot spots'. In the short-term, slight modifications to the methodology outlined in Chapter 4 will be used to analyze the regional and seasonal differences in VOCs measured in Nanaimo, BC and Port Alberni, BC. This analysis will employ the sample averaging described above to allow for multiple days of data to be modelled at the same time. These larger datasets have more replicate measurements of the sources described in Chapters 3 and 4, which will reduce the number of outlying sources present in the data. It is anticipated that this will produce MCR-WALS models of the data that contain more VOC components that can be uniquely ascribed as a specific source. For example, analyzing multiple days' worth of data from the Port Alberni area may lead to the component associated with the emissions from vegetation and wood products to be split into a vegetation component associated with the old growth forest, and a wood products component associated with the local industries. Using MCR-WALS on multiple days of data will also facilitate the identification of VOC sources present in new sampling areas, as many of the observed sources (*e.g.*, anthropogenic hydrocarbons, sawmills, forest fire

smoke) are common VOC sources across Vancouver Island and British Columbia. Most VOC source apportionment studies are done on data collected at stationary monitoring sites, and stationary deployment of the mobile lab could allow sources to be apportioned at locations that are not equipped for continuous air monitoring. In collaboration with industrial partners for site access, the methods described can be used to discriminate VOC sources within industrial sites and identify which processes may be impacting the nearby areas. VOCs are integral in the formation of secondary organic aerosols and ground-level ozone, both of which have well documented human consequences. Once VOC sources in a region have been identified, ozone formation potentials and aerosol forming potentials of the sources can be calculated to better characterize the effects of the regional VOCs on air quality. Understanding the temporal trends in the atmospheric composition, distribution, and concentrations of VOCs can help inform public policy for improvements in air quality.

These larger, multiday datasets could also be used to develop and validate a classification model (*e.g.*, kNN or SIMCA) that can be used to classify the air samples encountered while driving. In the data presented in Chapter 3, most of the VOC sources were sampled once, which did not allow for the data to be separated into a training set and test set to develop a classification model. Using multiple days of data would allow the intraday and seasonal variability of local VOC sources to be characterized, and incorporated into the classification model. A longer-term goal following the development of a classification model would allow for the measured mass spectra to be classified while sampling for on-the-fly sample discrimination.

The data from additional on-board sensors in the mobile mass spectrometry lab were discussed in Chapter 4, but have not been incorporated into the chemometric analysis presented here. Fusing the concentration data from PM, O_3, NO, NO_2, NO_x, CO_2, and CH_4 sensors is expected to further inform source discrimination and apportionment. For example, PM data is useful for identifying VOC sources associated with smoke, CO_2 to help distinguish fugitive emissions of hydrocarbons from combustion sources, and CH_4 to identify anaerobic biomass degradation. Methods for fusing together the data from the all the on-board sensors have been described in Data Fusion Methodology and Applications and will be the subject of future work.[252]

The future directions described above focused on further exploring the regional level chemical data collected using the mobile lab, but many other applications are possible. The mobile lab can be brought to industrial sites to follow processes, inform odor management strategies, or to monitor remediation efforts. Sample lines can be brought into buildings to measure indoor air quality and potentially identify occupational hazards in workplaces. The distribution of VOCs present in breath can be indicative of certain diseases such as lung cancer, and non-invasive, low risk screening methods that can identify earlier stages of lung cancer are desired.[253,254] GC-MS is often used for VOC breath analysis, but PTR-ToF-MS is becoming more common in

medical research due to the sensitivity and high time resolution of the method, and when combined with chemometrics may provide a rapid on-site screening method.

The mobilization of mass spectrometry to the field represents a paradigm shift with far-reaching potential. When combined with the chemometric techniques described here it can be used to discriminate sources, classify samples, and resolve complex mixtures. These methods contribute to our understanding of the fate and distribution of VOCs and protect environmental and human health.

Appendices

Appendix A - Supplemental information for Chapter 2

Table A.1: Permeation tube emission rate data for the ten compounds used to produce the constructed VOC mixtures.

Compound	Emissions Rate (ng/min)	Temperature (°C)
α-Pinene	447	100
Benzene	1200	50
Biphenyl	383	80
Carbon Tetrachloride	376	60
Chorobenzene	234	50
Naphthalene	697	80
Toluene	738	50
Toluene-d_8	609	50
Trichloroethylene	230	60
m-Xylene	458	50

Table A.2: Sample information for aqueous hydrocarbon samples.

Source	Gravimetric dilution ranges from saturated aqueous solution	Number of samples
OSPW	33–40	3
Diluent	910–1340	4
Dilbit	490–750	7
Mixed gasoline	680	3

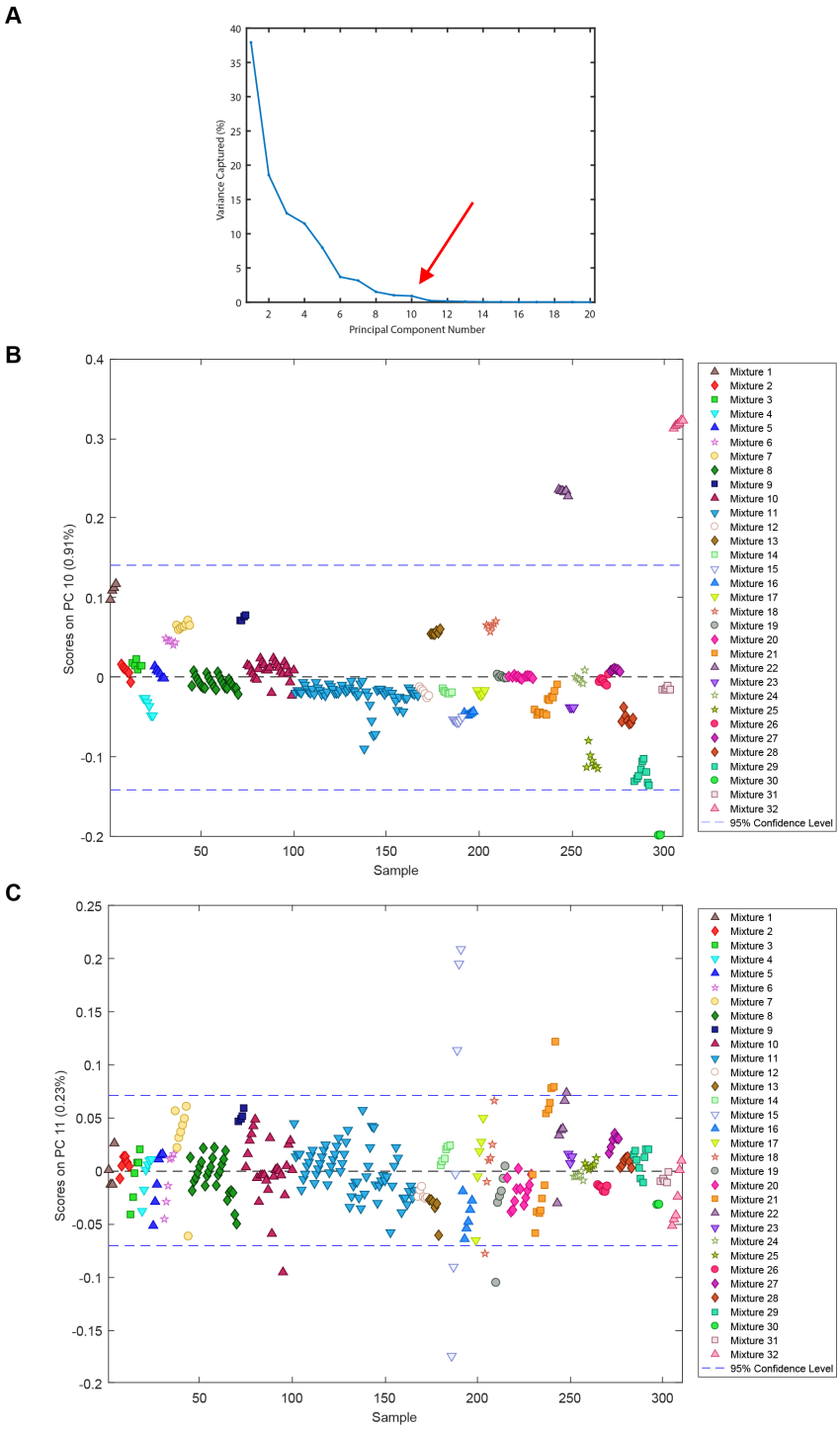


Figure A.1: A) Variance captured by each of the first 20 PCs for the permeation tube VOC mixture PCA. The first 10 PCs (indicated by the arrow) capture the majority of the variance. B) Scores plot of PC 10 versus sample number. Generally, samples containing the same mixture have similar scores on PC 10. This structure is indicative of PC 10 describing meaningful information in the dataset. C) Scores plot of PC 11 versus sample number. In this scores plot, the majority of mixtures have samples with both positive and negative scores, indicating PC 11 describes noise in the data.

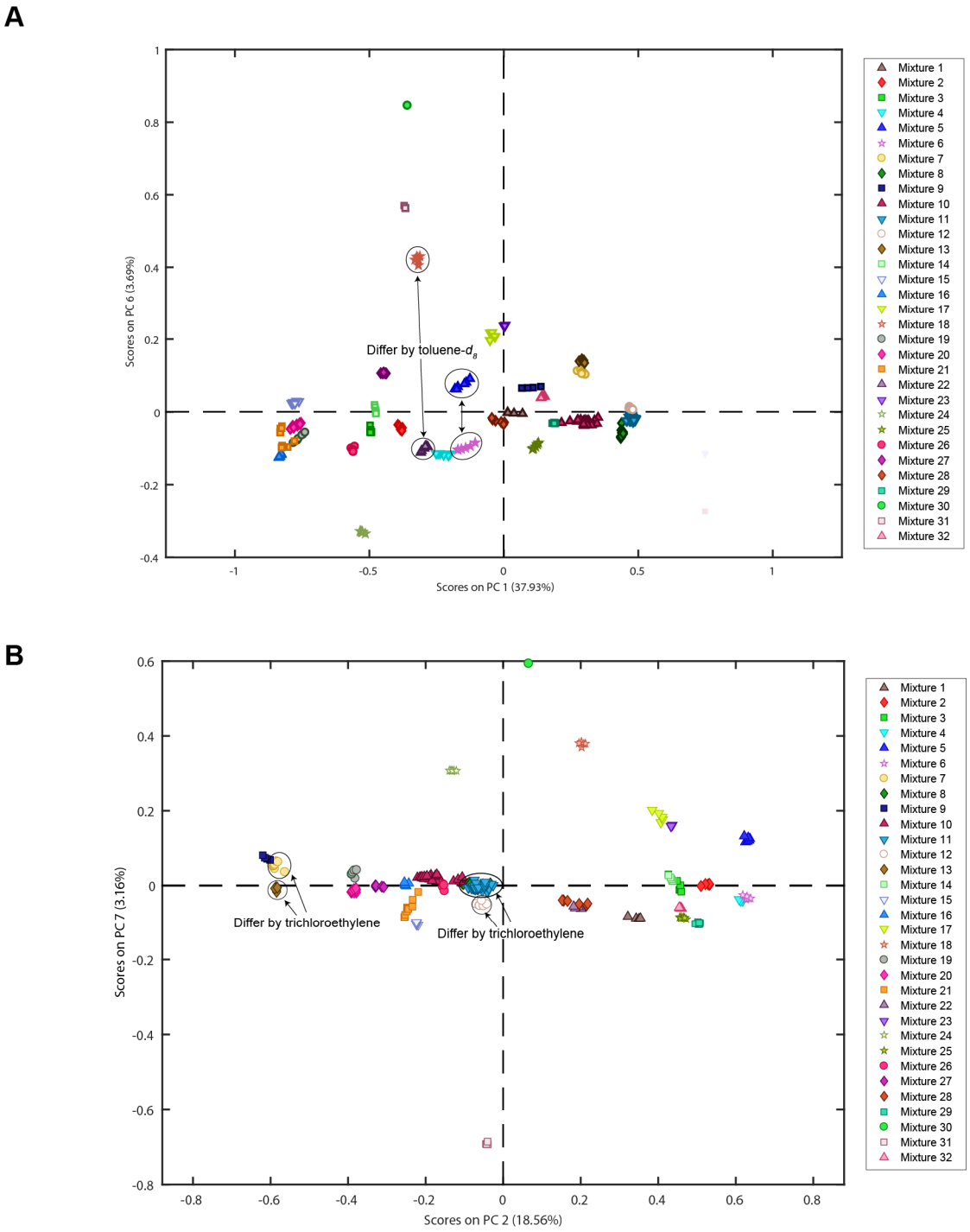


Figure A.2: A) Scores plot of PC 1 versus PC 6 for the constructed VOC mixtures. Two pairs of mixtures (mixtures 18 and 22, and mixtures 5 and 6) that differ only by the presence of toluene-d_8 are circled. B) Scores plot of PC 2 versus PC 7 for the constructed VOC mixtures. Two pairs of mixtures (mixtures 11 and 12, and mixtures 7 and 13) that differ only by the presence of trichloroethylene are circled.

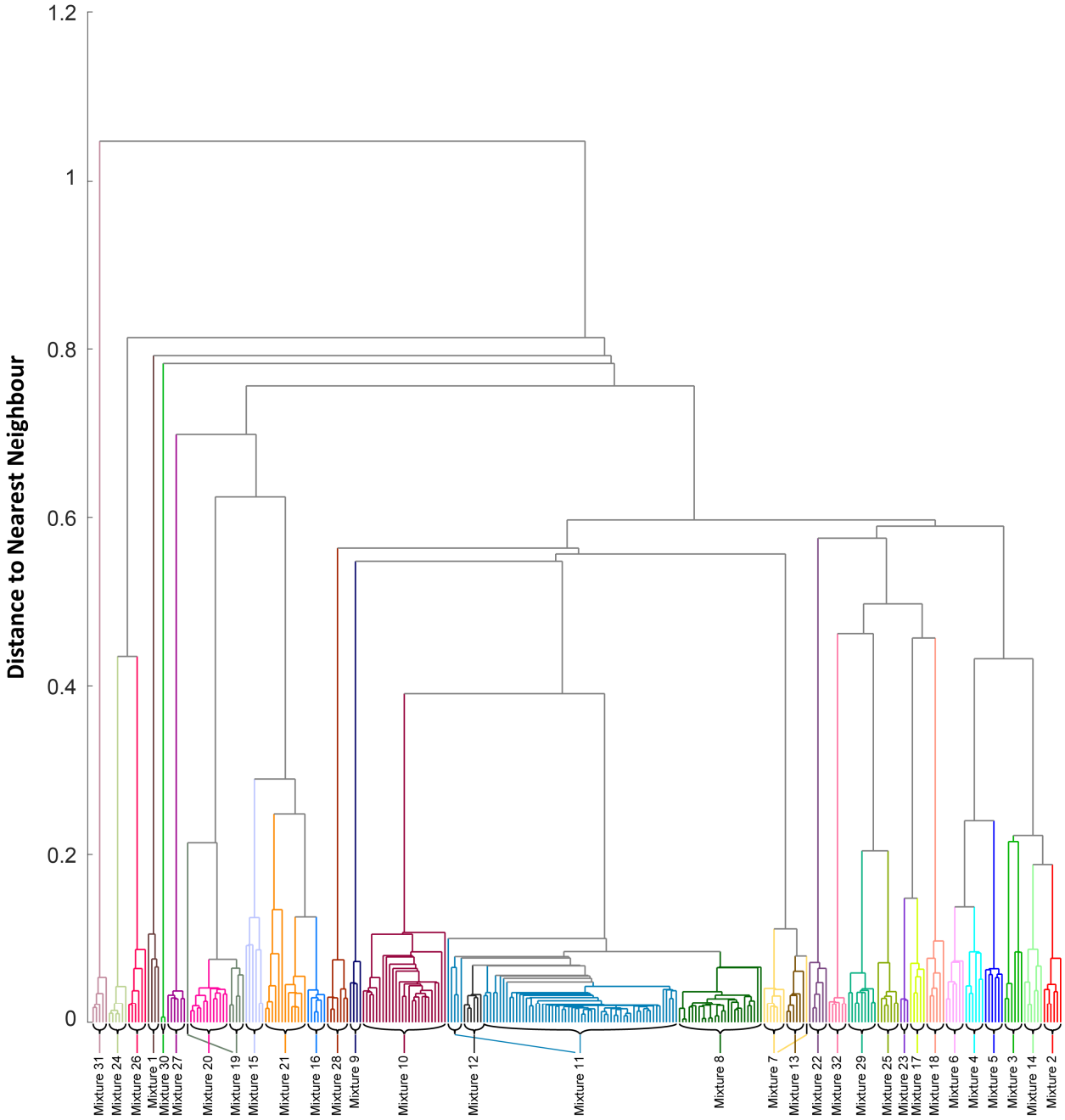


Figure A.3: Dendrogram of the cluster analysis of the constructed VOC mixtures. The average background subtracted full scan MIMS spectra were unit vector normalized before the analysis. Euclidean distances were used to measure the distance between nearest neighbors, and are shown on the y-axis.

Table A.3: Confusion table for the kNN classification (k=1) of the constructed VOC mixtures. The numbers along the diagonal represent samples that have been correctly classified. Numbers off the diagonal represent samples that have been misclassified. 95 of the 96 samples in the test set have been correctly classified.

		Actual Mixture																															
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**
Predicted Mixture	**1**	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	13	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2

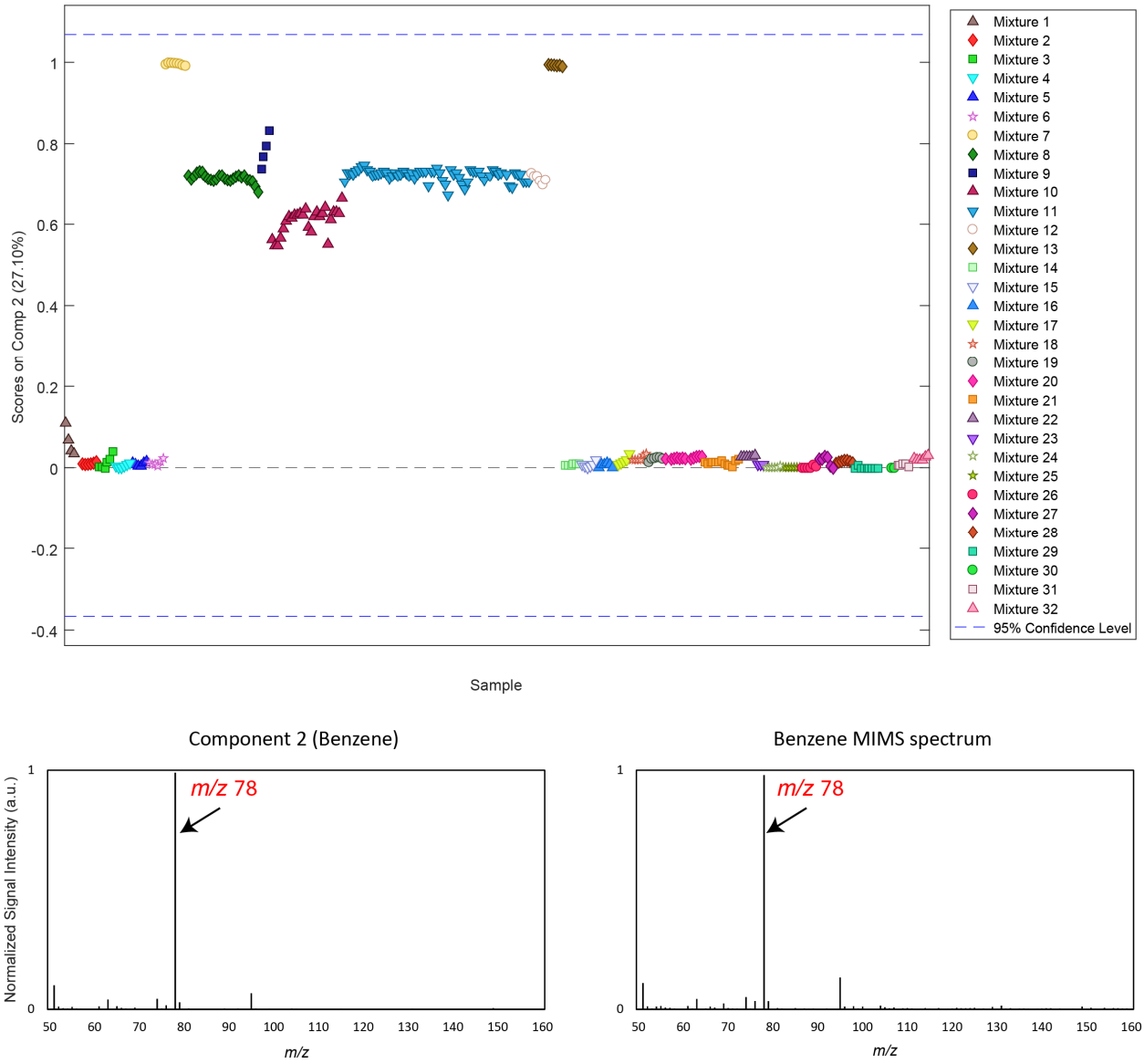


Figure A.4: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 2. Bottom left: MCR component 2. Bottom right: Full scan MIMS spectrum of benzene.

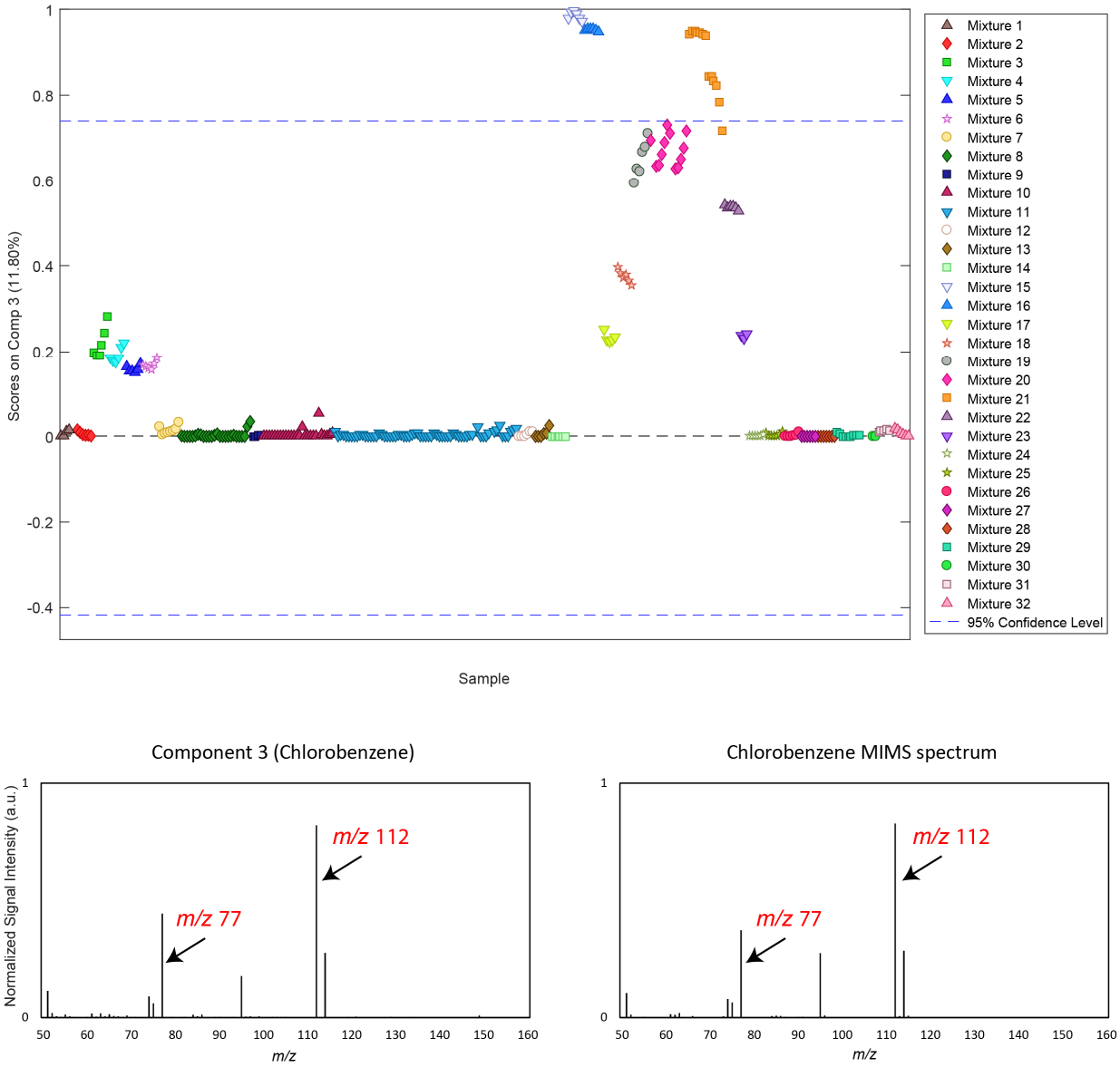


Figure A.5: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 3. Bottom left: MCR component 3. Bottom right: Full scan MIMS spectrum of chlorobenzene.

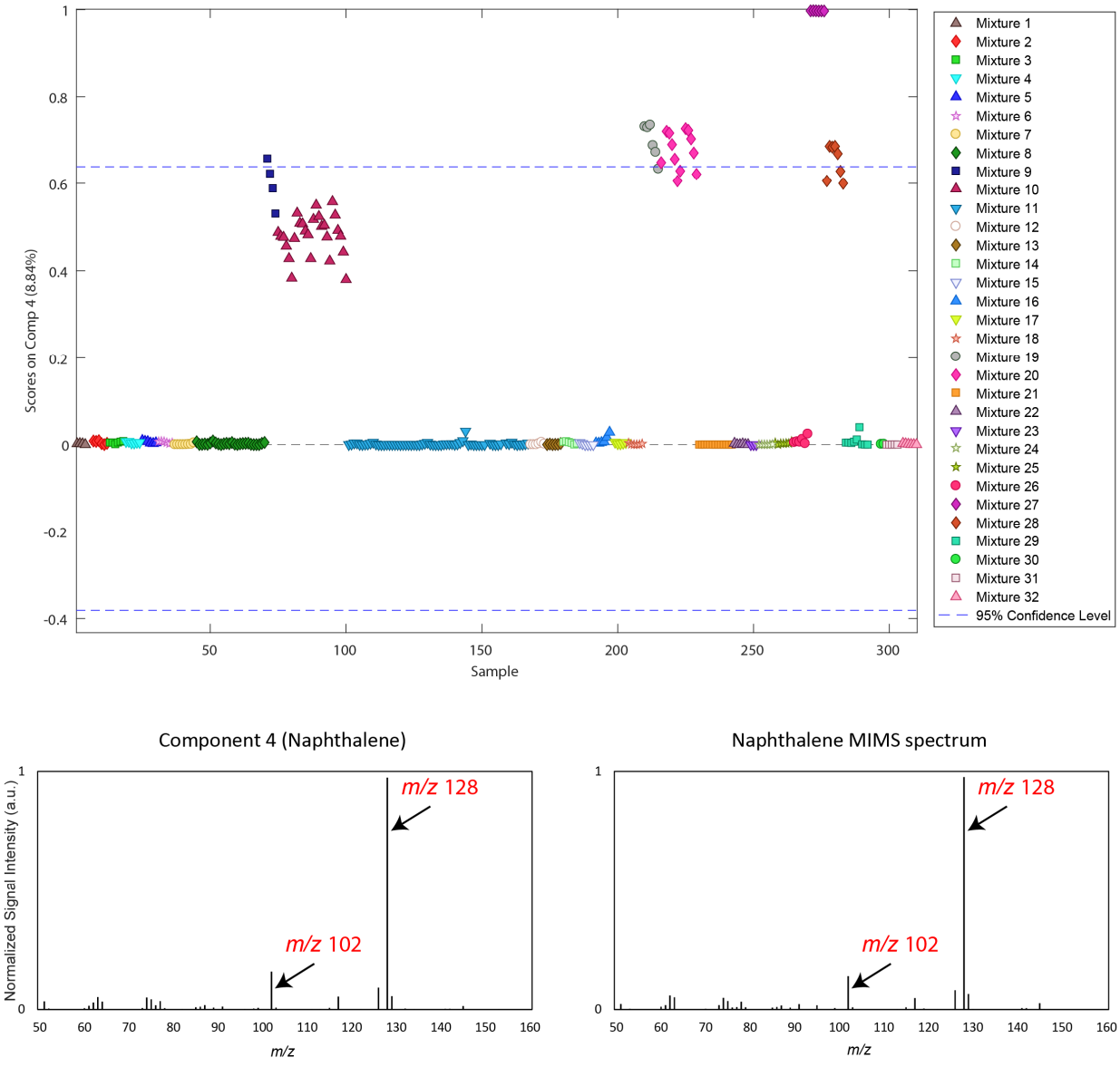


Figure A.6: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 4. Bottom left: MCR component 4. Bottom right: Full scan MIMS spectrum of naphthalene.

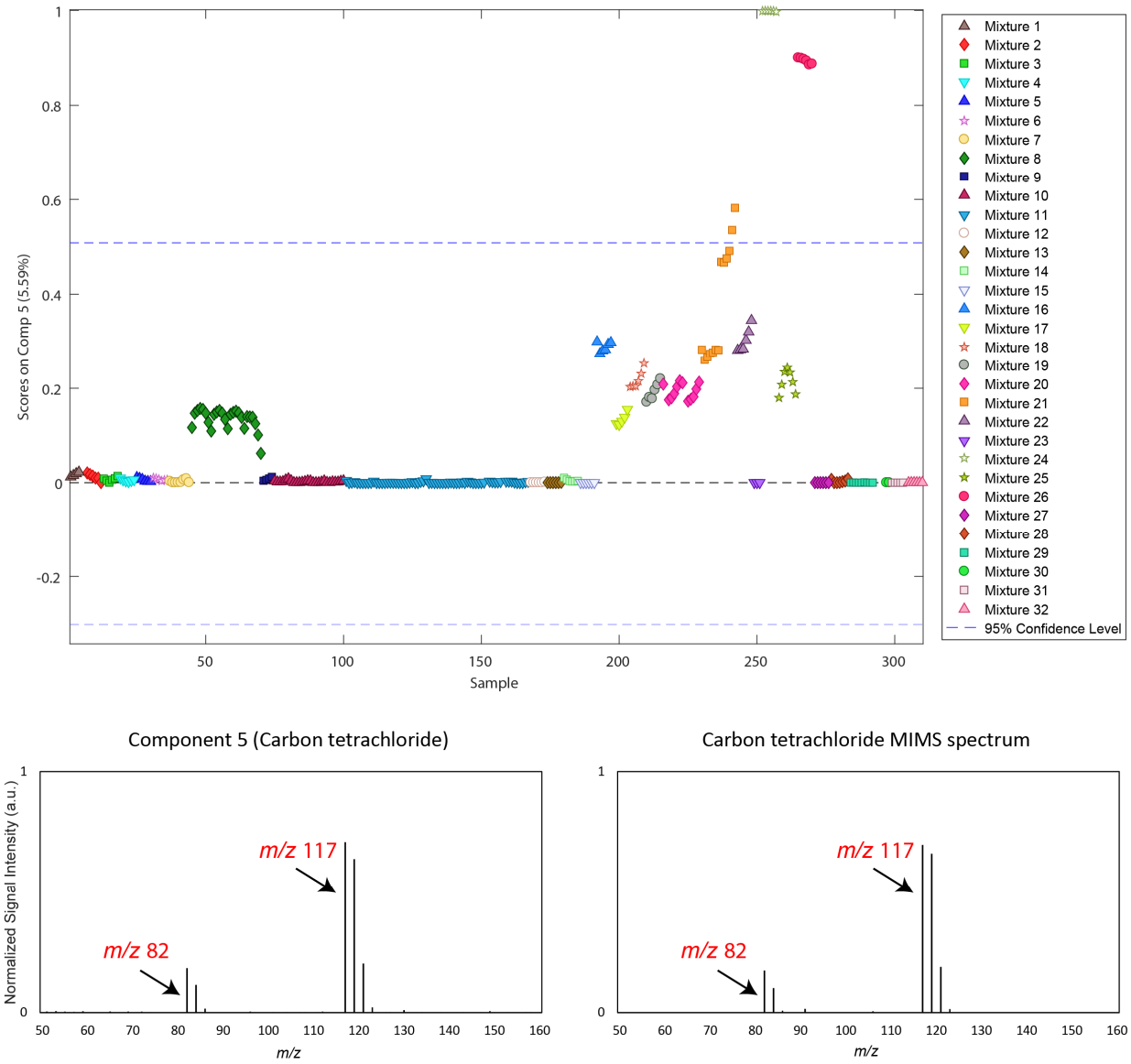


Figure A.7: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 5. Bottom left: MCR component 5. Bottom right: Full scan MIMS spectrum of carbon tetrachloride.

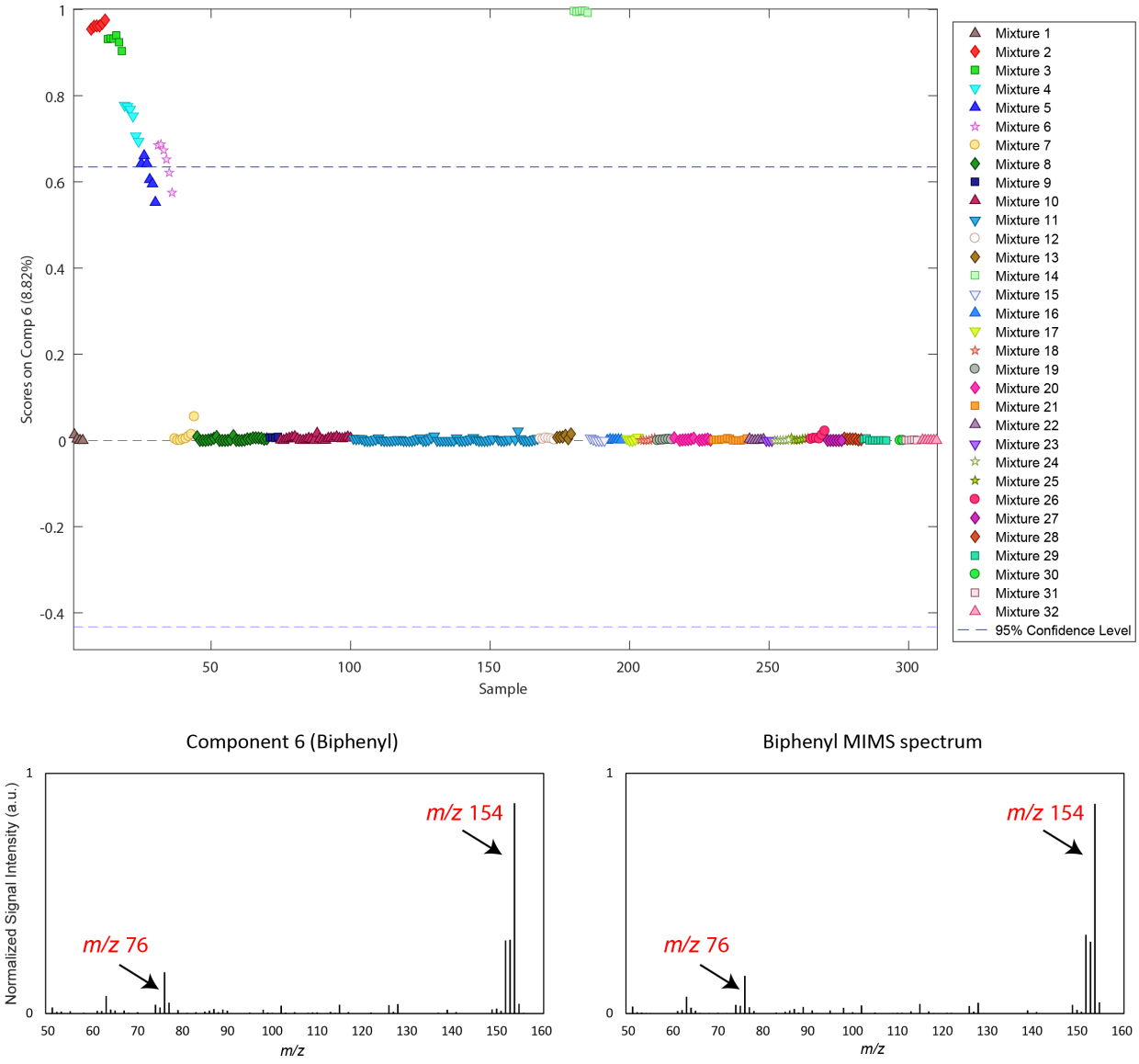


Figure A.8: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 6. Bottom left: MCR component 6. Bottom right: Full scan MIMS spectrum of biphenyl.

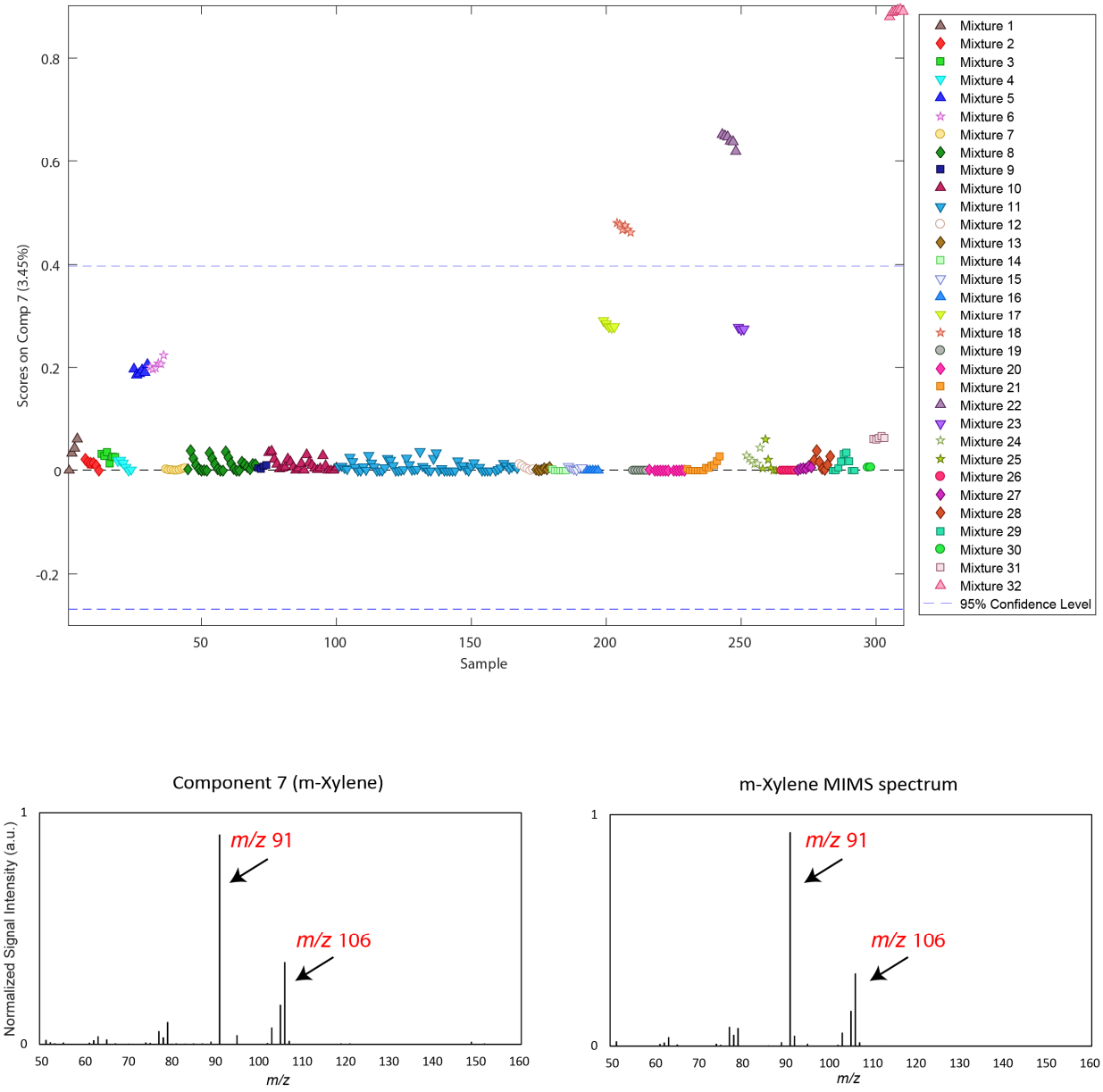


Figure A.9: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 7. Bottom left: MCR component 7. Bottom right: Full scan MIMS spectrum of *m*-xylene.

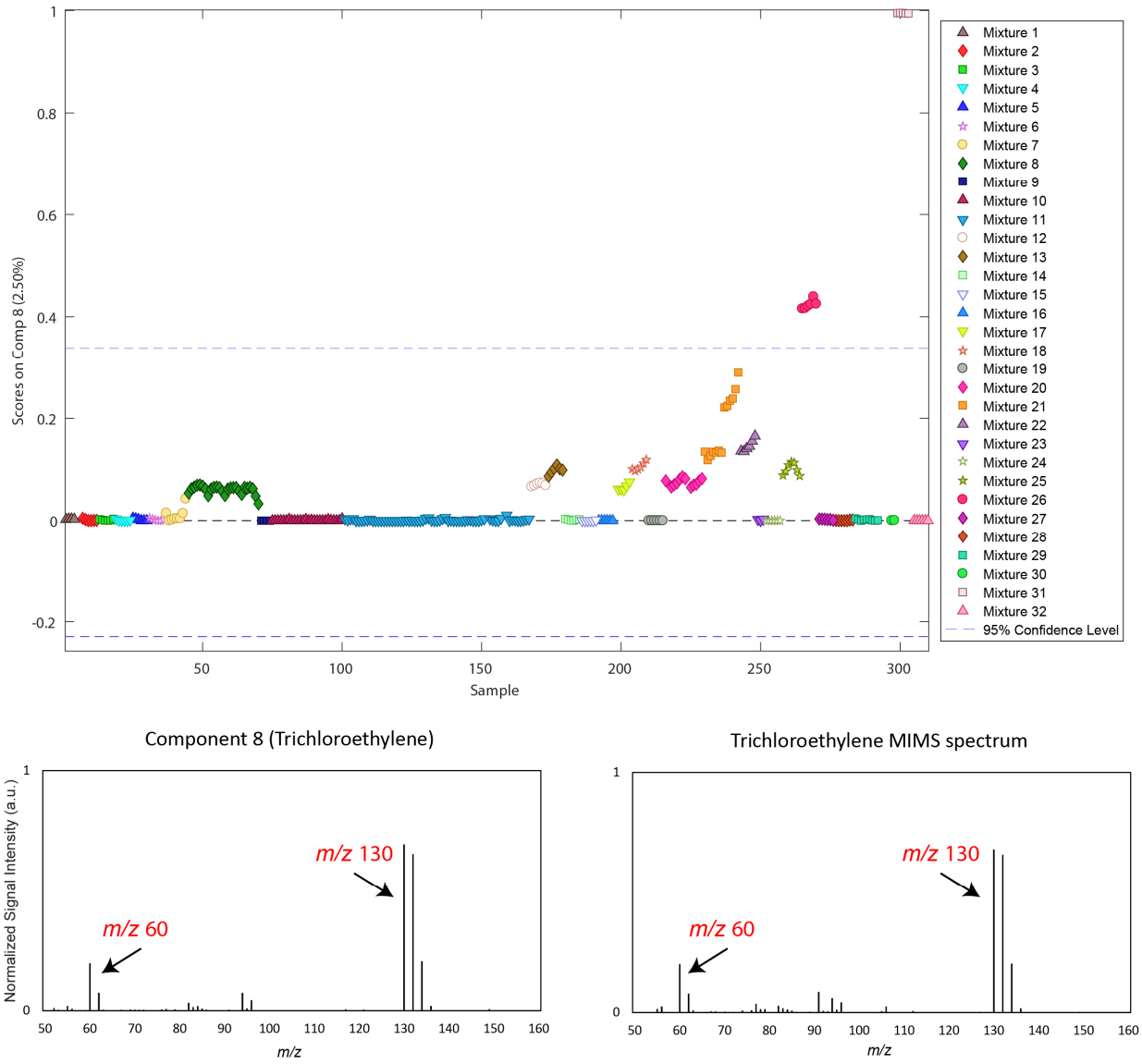


Figure A.10: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 8. Bottom left: MCR component 8. Bottom right: Full scan MIMS spectrum of trichloroethylene.

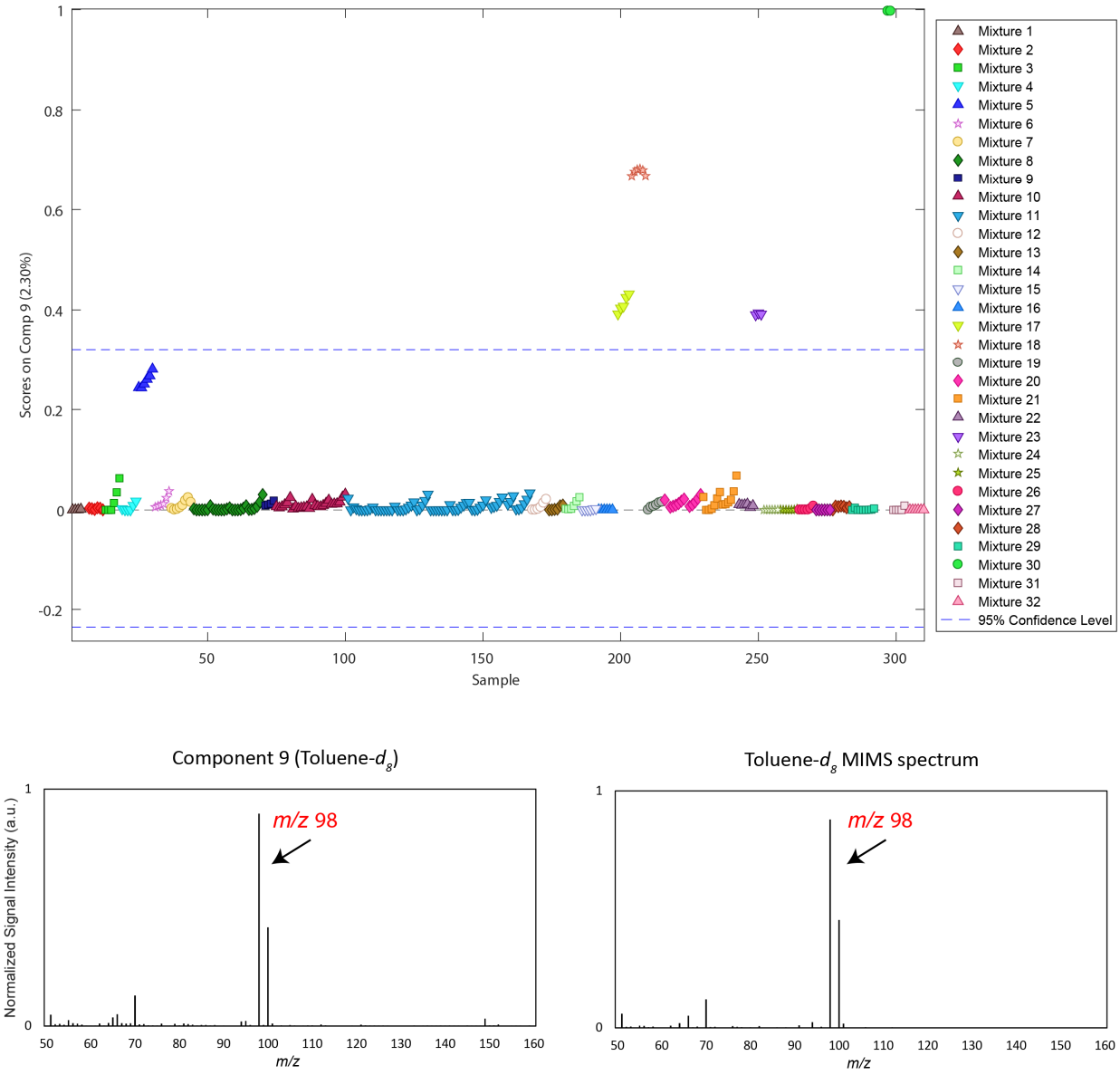


Figure A.11: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 9. Bottom left: MCR component 9. Bottom right: Full scan MIMS spectrum of toluene-d_8.

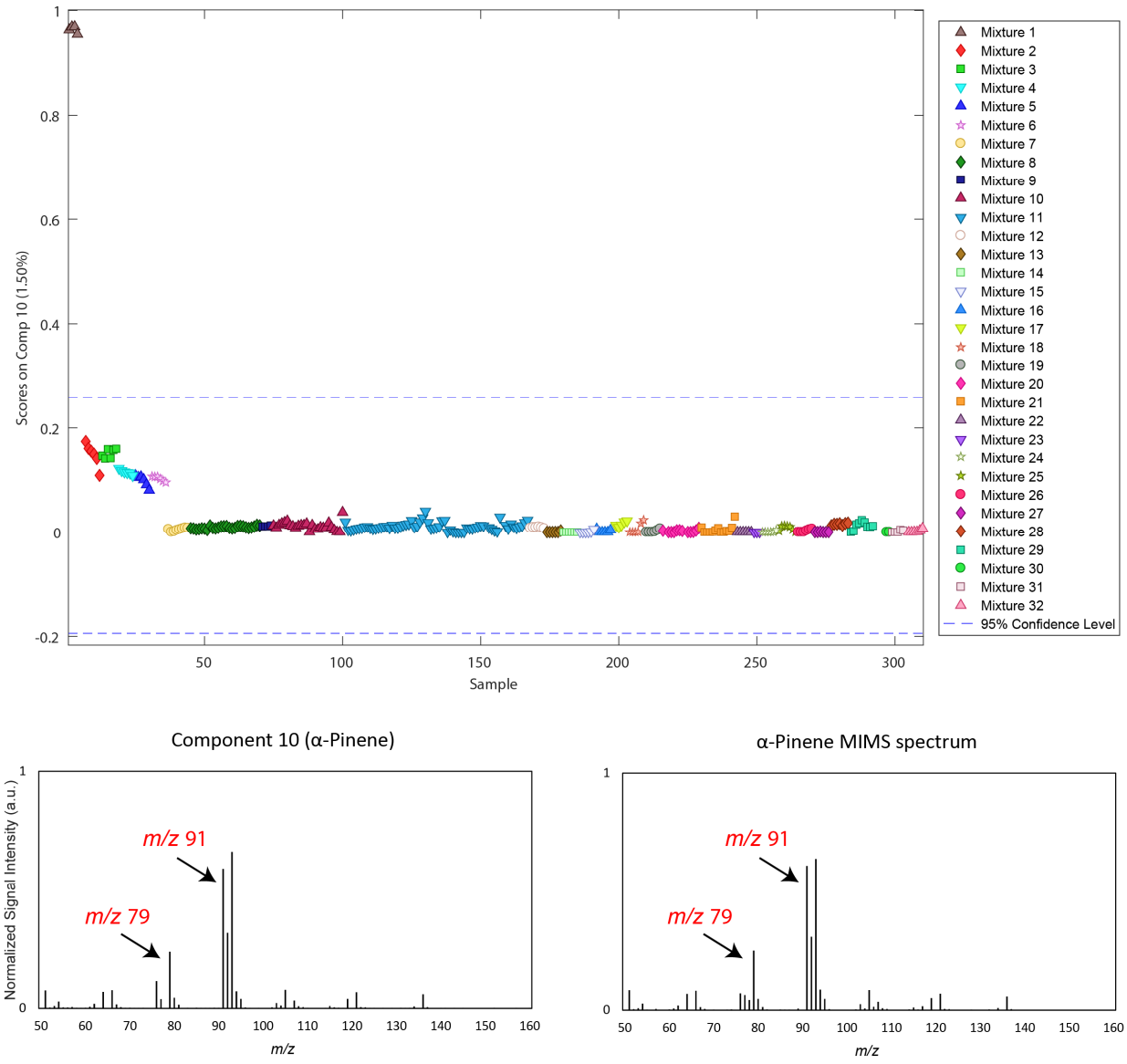


Figure A.12: MCR analysis of the constructed VOC mixtures. Top: Scores for all the samples on component 10. Bottom left: MCR component 10. Bottom right: Full scan MIMS spectrum of α-pinene.

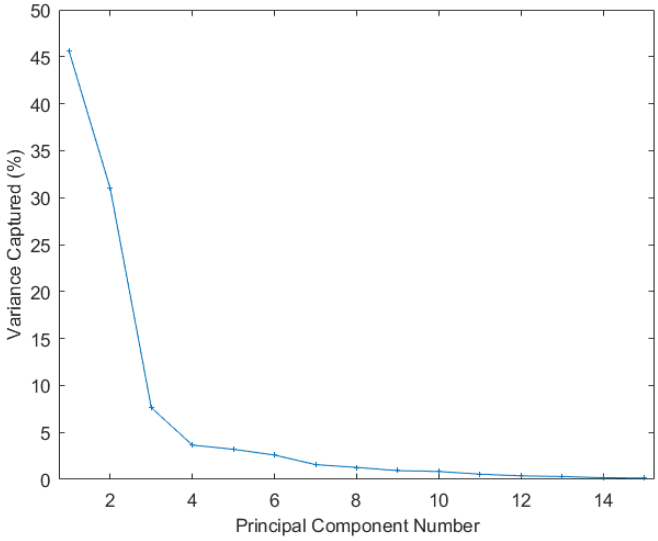


Figure A.13: Variance captured by each of the first 15 PCs for the woodsmoke PCA.

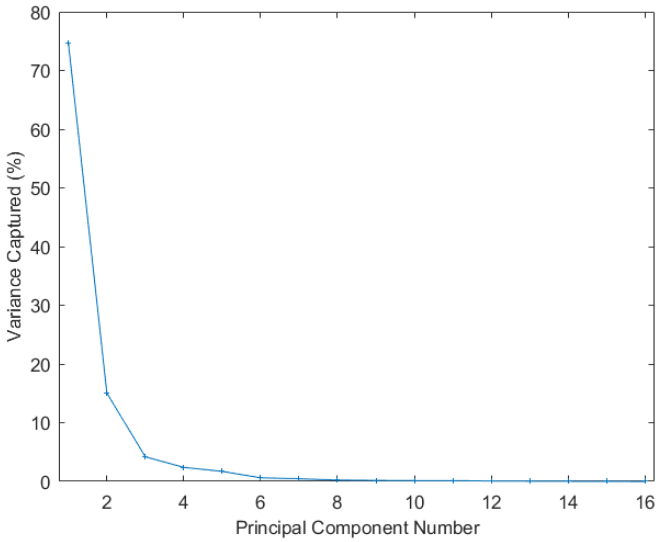


Figure A.14: Variance captured by each of the first 16 PCs for the headspace of aqueous hydrocarbon samples PCA.

Appendix B - Supplemental information for Chapter 3

August 2016 Field Campaign meteorological and ventilation index data

Meteorological[248] and ventilation index[255] data collected by the British Columbia Ministry of Environment for the Harmac Pacific weather station (meteorological data) and Central Vancouver Island Zones (ventilation index) on the four sampling days is found in Table B.1. The average wind speed ranged from 2–4 m/s, with wind directions coming from approximately 340 deg from North on August 4, 5, and 10, and from 95° from North on August 9. The afternoon ventilation indexes on August 4, 5, and 10 were Fair, and Poor on August 9, with mixing heights of 236–1399 m. Average temperatures in the region were between 18–21 °C.

Table B.1: Meteorological data for the August 2016 field campaign.

Place and Date (Ventilation Index Zone)	Wind speed (m/s)	Wind direction (deg)	Temp (°C)	Ventilation Index at 4 pm	Mixing Height (m) at 4 pm
Nanaimo Aug 4, 2016 (Central VI)	4	348	21	51 (FAIR)	947
Nanaimo Aug 5, 2016 (Central VI)	4	341	19	52 (FAIR)	1399
Nanaimo Aug 9, 2016 (Central VI)	2	95	18	20 (POOR)	236
Nanaimo Aug 10, 2016 (Central VI)	3	341	21	40 (FAIR)	745

Quantitative MIMS data

Table B.2: Quantitative information for the MIMS analysis. Detection limits were estimated using 3x the standard deviation of the baseline.

Compound	Method	Slope (counts/ppb$_v$)	R^2	Detection Limit (ppb$_v$)
Benzene	Full Scan 78	58	0.999	3
Toluene	MS/MS 91 → 64+65	56	0.999	6
Ethylbenzene	MS/MS 106 → 91	39	0.996	3
α-Pinene	SIM 121	22	0.998	11

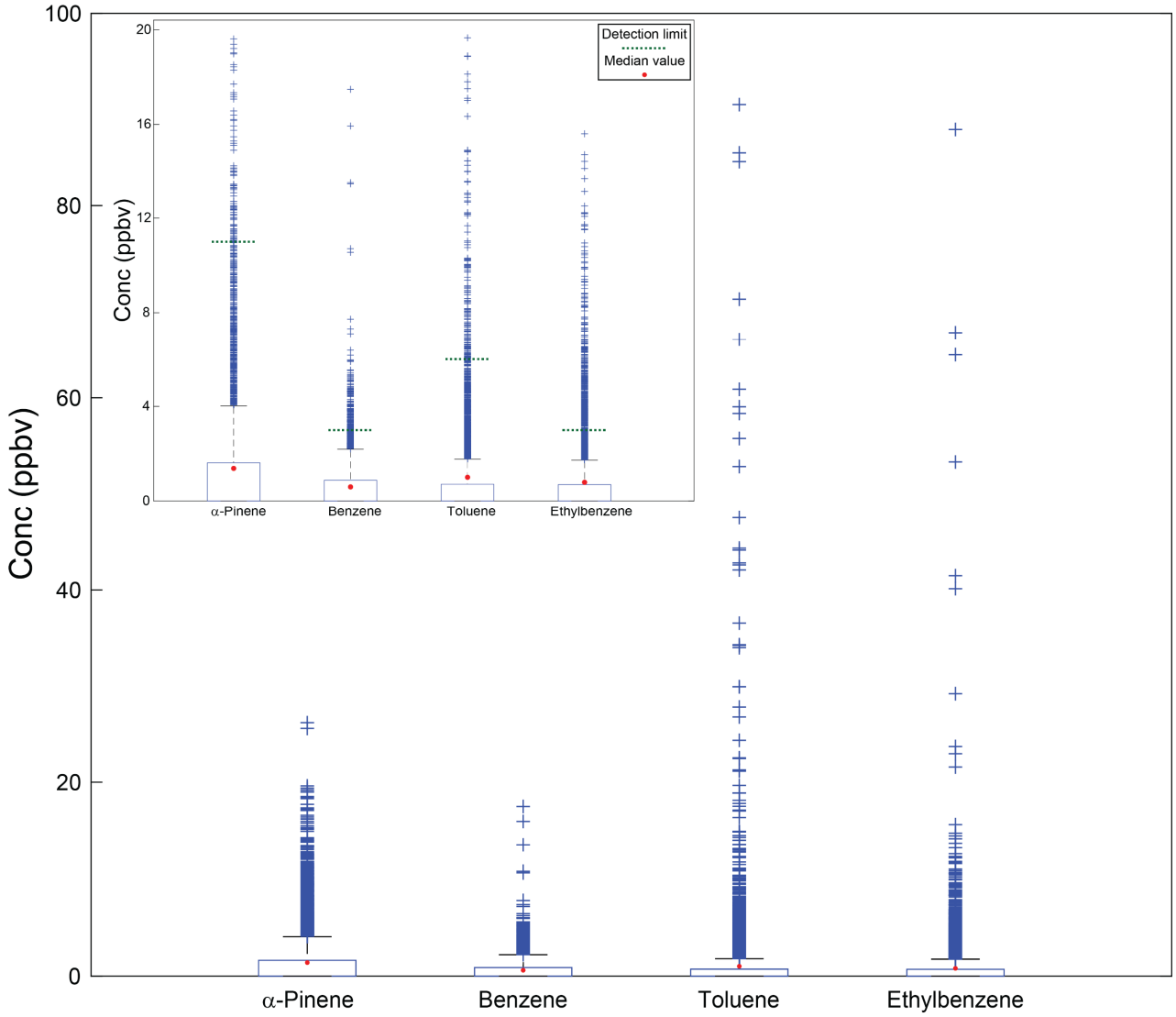


Figure B.1: Boxplot of semi-quantitative concentrations of α-pinene, benzene, toluene, and ethylbenzene. The box represents 25–75 percentiles and the whiskers display 5–95 percentiles. Data points above the 95th percentile are shown individually. The inset figure shows a close up of the concentration axis in the low concentration range. The green dotted lines in the inset represent detection limits, and the red dots represent the median concentration.

Toluene and ethylbenzene were analyzed using tandem mass spectrometry scans, whereas α-pinene was quantified using SIM, and the benzene signal intensity was extracted from the full scan data. The mass spectrometer parameters for these scans are provided in Table B.2. For quantitation, external calibration curves were generated using certified permeation tubes (Kin-Tek, La Marque, TX, USA) in a Dynacalibrator gas dilution system (Model 340; VICI Metronics, Poulsbo, WA, USA). The highest observed concentrations for BTEX compounds was measured in the vicinity of fresh asphalt. Since our focus in the current study was on sample discrimination using full scan mass spectra from a mobilized instrument, we elected not to include a continuously infused internal standard as previously described.[82] As such, corrections for instrument drift and directional dependence[82] were not employed and the measured concentrations reported here should be regarded as semi-quantitative.

August 2017 Field Campaign meteorological and ventilation index data

Meteorological[248] and ventilation index[255] data on the two drive days is found in Table B.3. The average wind speed and direction on August 21, 2017 in Nanaimo, BC during our field campaign was 3 m/s from at 253° from North. The afternoon ventilation index for the region was poor with a mixing height of 666 m. On August 22, the average wind speed was 3 m/s from 123° from North. The ventilation index was good with a mixing height of 1143 m. On August 22, our field campaign extended as far south as Crofton, BC (Data from Crofton Substation) for 1 hour of sampling. Both days were warm and sunny with average temperatures of 24 and 26 °C.

Table B.3: Meteorological data for the August 2017 field campaign.

Place and Date (Ventilation Index Zone)	Wind speed (m/s)	Wind direction (deg)	Temp (°C)	Ventilation Index at 4 pm	Mixing Height (m) at 4 pm
Nanaimo Aug 21, 2017 (Central VI)	3	253	24	32 (POOR)	666
Nanaimo Aug 22, 2017 (Central VI)	3	123	27	61 (GOOD)	1143
Crofton Aug 22, 2017 (Southern VI)	2	126	26	92 (GOOD)	1519

Quantitative PTR-ToF-MS data

Table B.4: Quantitative information for the PTR-ToF-MS analysis. Detection limits were estimated using 3x the standard deviation of the baseline.

Compound	Measured *m/z*	Slope (ncps/ppb$_v$)	R^2	Detection Limit (ppb$_v$)
Benzene	79.049	28	0.998	0.3
Toluene	93.061	45	0.996	0.9
Ethylbenzene	107.076	32	0.999	0.3
Dimethyl Sulphide	63.022	26	0.994	0.4
α-Pinene	137.112	15	0.999	0.3

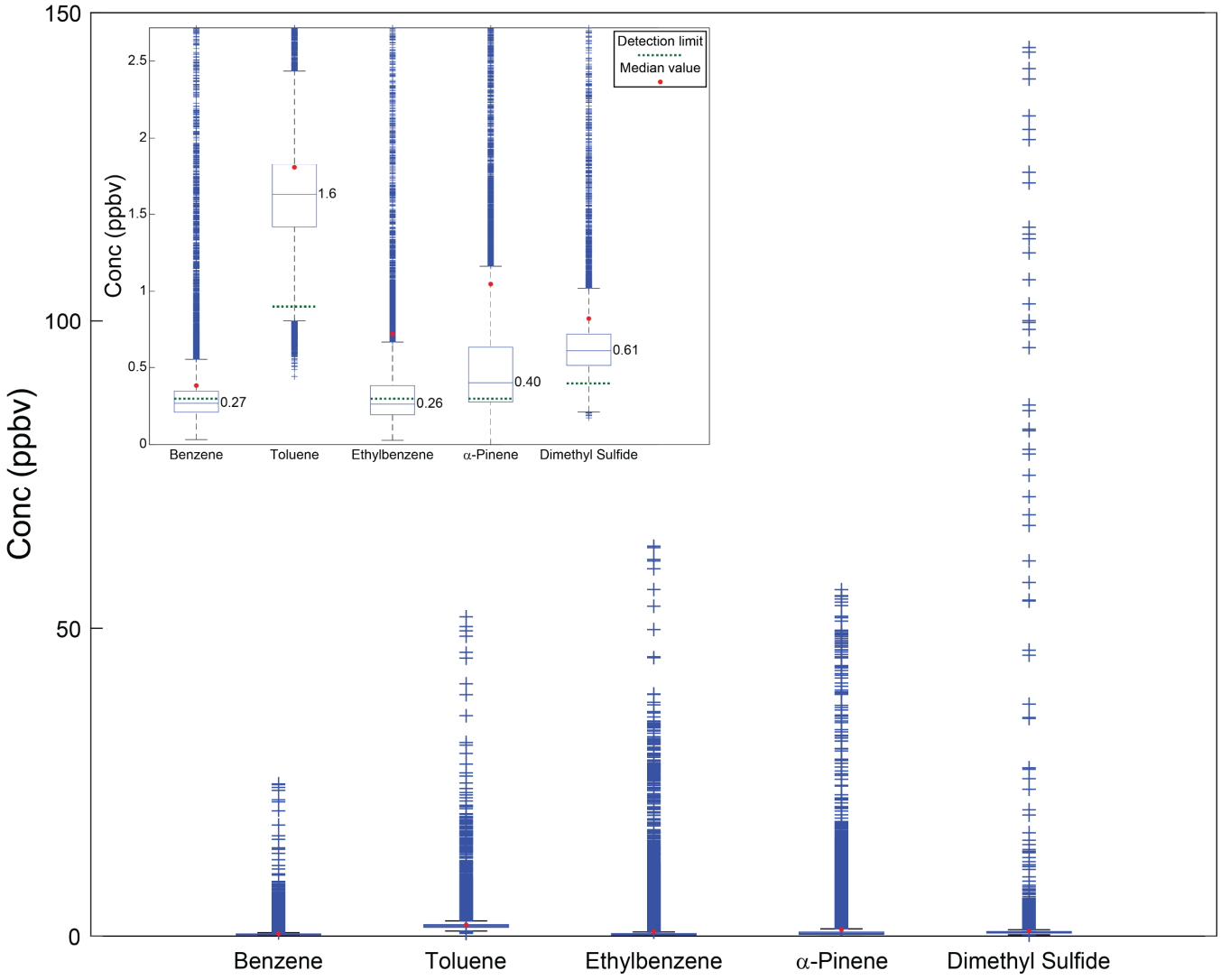


Figure B.2: Boxplot of semi-quantitative concentrations of α-pinene, benzene, toluene, and ethylbenzene. The box represents 25–75 percentiles and the whiskers display 5–95 percentiles. Data points above the 95th percentile are shown individually. The inset figure shows a close up of the concentration axis in the low concentration range. The green dotted lines in the inset represent detection limits, and the red dots represent the median concentration.

Qualitative MIMS analysis

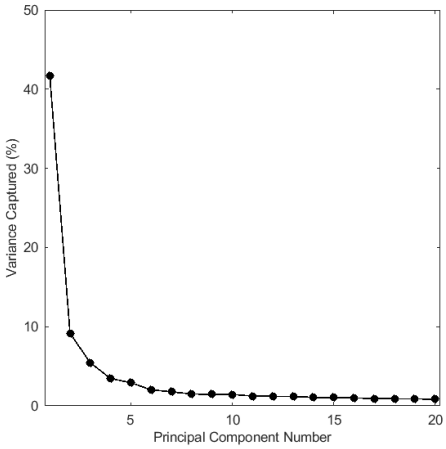


Figure B.3: Scree plot for MIMS PCA.

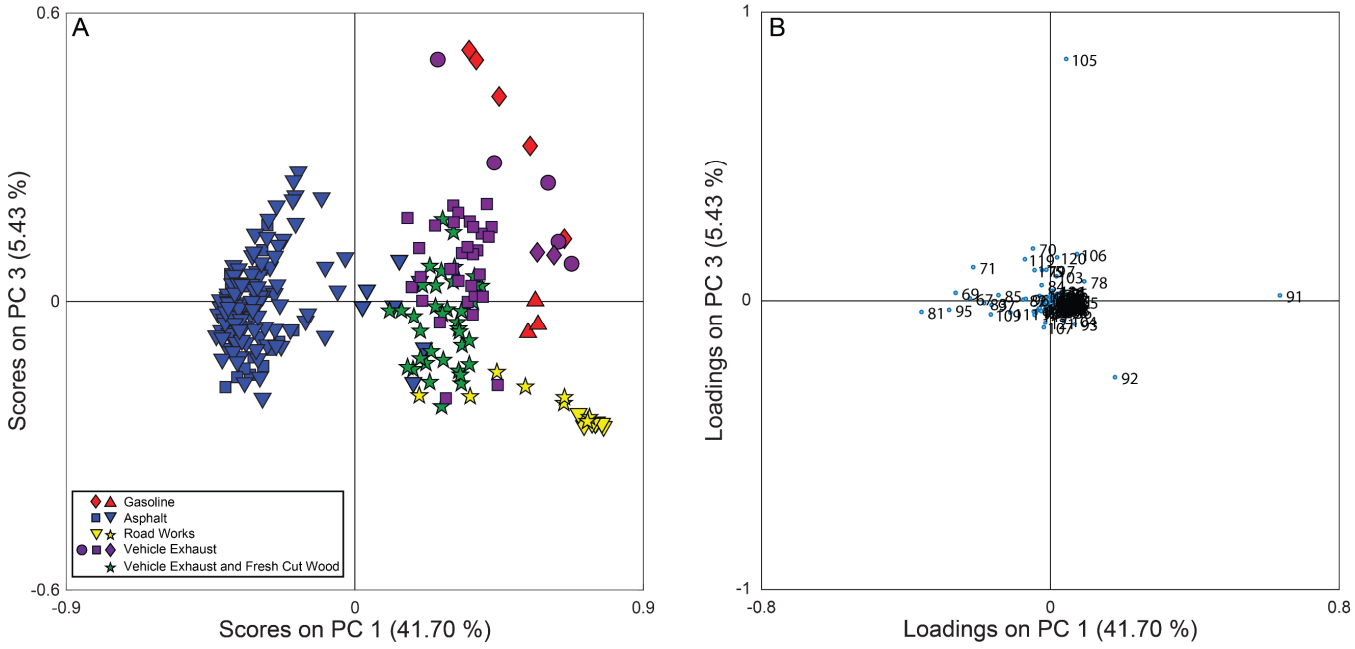


Figure B.4: PCA analysis of the MIMS data for PC 1 versus PC 3. A) PC 1 versus PC 3 scores plot. In this plot the asphalt samples are still differentiated from the other samples, but there is more overlap between the remaining hydrocarbon sources. B) PC 1 versus PC 3 loadings plot.

Qualitative PTR-TOF-MS Analysis

Table B.5: This table contains a list of major ions detected with the PTR-ToF-MS on the August 2017 field campaign. The table shows our measured *m/z*, the possible chemical formula, the exact monoisotopic mass for that formula, possible chemical compounds, and the major and minor observed sources.

Measured *m/z*	Possible Formula	Exact *m/z*	Possible Compounds	Major observed sources	Minor observed sources
33.030	CH_5O^+	33.034	Methanol[37,256,257]	Gas station Sawmill Wood chip truck Pulp mill Compost Top soil farm	Everywhere
41.031	$C_3H_5^+$	41.039	Isoprene fragment[37,257] Hydrocarbon fragment[211,258]	Gas station Compost Ferry Auto wrecking facility Farm vehicle	Wood chip truck Landfill
43.010	$C_2H_3O^+$	43.019	Ester, acid fragment[37,200,257]	Compost Auto wrecking facility	Gas station Sawmill Wood chip truck Pulp mill Ferry
43.044	$C_3H_7^+$	43.055	Propene[256] Propanal fragment[257] Hydrocarbon fragment[258,259]	Gas station Compost Ferry Auto wrecking facility	Farm vehicle Landfill
45.025	$C_2H_5O^+$	45.033	Acetaldehyde[37,256,257]	Sawmill Wood chip truck Sawmill Compost Pulp mill Auto wrecking facility Wood storage	Farm vehicle Landfill
47.010	$CH_3O_2^+$	47.012	Formic acid[37,211,256]	Sawmill Wood chip truck Compost Pulp Mill	Auto wrecking facility Landfill

Table B.5: This table contains a list of major ions detected with the PTR-ToF-MS on the August 2017 field campaign. The table shows our measured *m/z*, the possible chemical formula, the exact monoisotopic mass for that formula, possible chemical compounds, and the major and minor observed sources, continued.

Measured *m/z*	Possible Formula	Exact *m/z*	Possible Compounds	Major observed sources	Minor observed sources
57.070	$C_4H_9^+$	57.070	Butene[37,257,258] Hydrocarbon fragment[37,257–259]	Gas station Ferry Auto wrecking facility	Sawmill Wood chip truck Compost Farm vehicle Compost Top soil farm Pulp mill
59.048	$C_3H_7O^+$	59.049	Acetone[37,211,256,257] Propanal[37,211,256,257]	Compost Pulp mill Auto wrecking facility	Sawmill Top soil farm
61.029	$C_2H_5O_2^+$	61.028	Acetic acid[37,188,256,257] Glycolaldehyde[257] Ethyl acetate[257]	Compost	Pulp mill
63.022	$C_2H_7S^+$	63.026	Dimethyl sulphide[37,260]	Pulp mill	
69.067	$C_5H_9^+$	69.069	Isoprene[37,256,257] Hydrocarbon fragment[258] Methylbutenol fragment[257] Cyclohexane[257]	Gas station Compost Auto wrecking facility Ferry	Pulp mill Farm vehicle Top soil farm Landfill
71.080	$C_5H_{11}^+$	71.086	Hydrocarbon fragment[37,256,257]	Gas station Auto wrecking facility	Ferry Farm vehicle
73.055	$C_4H_9O^+$	73.065	Butanal[37,256,257] 2-Methylpropanal[256] 2-Butanone[256] Methyl ethyl ketone[37,257]	Compost Pulp mill Top soil farm Auto wrecking facility	
75.027	$C_3H_7O_2^+$	75.044	Ethyl formate[256] Methyl acetate[256] Propanoic acid[37,257] Hydroxyacetone[37,257]	Compost	Pulp mill
79.049	$C_6H_7^+$	79.054	Benzene[37,256,257] Hydrocarbon fragment[258]	Gas station Pulp mill Wood chip truck Ferry Auto wrecking facility	Top soil farm Farm vehicle Landfill

Table B.5: This table contains a list of major ions detected with the PTR-ToF-MS on the August 2017 field campaign. The table shows our measured *m/z*, the possible chemical formula, the exact monoisotopic mass for that formula, possible chemical compounds, and the major and minor observed sources, continued.

Measured *m/z*	Possible Formula	Exact *m/z*	Possible Compounds	Major observed sources	Minor observed sources
81.066	$C_6H_9^+$	81.070	Terpene fragment[37,210,257] Polyaromatic hydrocarbon fragment[37,257]	Wood chip truck Sawmill Compost Pulp mill Top soil farm	Auto wrecking facility Farm vehicle Landfill
91.035	$C_3H_7O_3^+$ or $C_7H_7^+$	91.039 or 91.055	Unidentified $C_3H_6O_3$ compound[261] or Tropyllium ion[210]	Auto wrecking facility	Gas station Ferry
93.061	$C_7H_9^+$	93.070	Toluene[37,256,257] Terpene fragment[210] Hydrocarbon fragment[258]	Gas station Sawmill Wood chip truck Auto wrecking facility Ferry	Compost Pulp mill
95.015	$C_2H_7S_2^+$ or $C_6H_7O^+$	94.998 or 95.049	DMDS or phenol[37,256,257]	Wood chip truck Pulp mill	Compost Auto wrecking facility
97.010	$C_5H_5O_2^+$	97.030	Furan-2-carbaldehyde[256] Furfural[37,257]	Compost	Pulp mill
107.076	$C_8H_{11}^+$	107.086	C_2-alkylbenzene[37,257,258] Terpene fragment[210] Hydrocarbon fragments[258]	Gas station Ferry Auto wrecking facility	Compost Pulp mill Top soil farm
121.085	$C_9H_{13}^+$	121.101	C_3-alkylbenzene[37,211,257,258] Sesquiterpene fragment[210]	Gas station Ferry Auto wrecking facility	Sawmill Wood chip truck
135.092	$C_{10}H_{15}^+$	135.117	C_4-alkylbenzene[37,257,258] Sesquiterpene fragment[210] p-Cymene[211]	Gas station Compost Ferry Auto wrecking facility	Wood chip truck Sawmill Pulp mill
137.112	$C_{10}H_{17}^+$	137.132	Monoterpenes[210]	Wood chip truck Sawmill Compost Pulp mill Top soil farm Landfill	

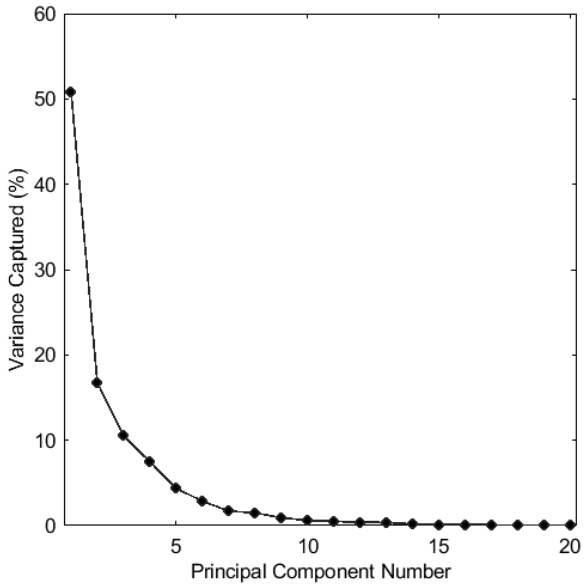


Figure B.5: Scree plot for supervised PTR-ToF-MS PCA.

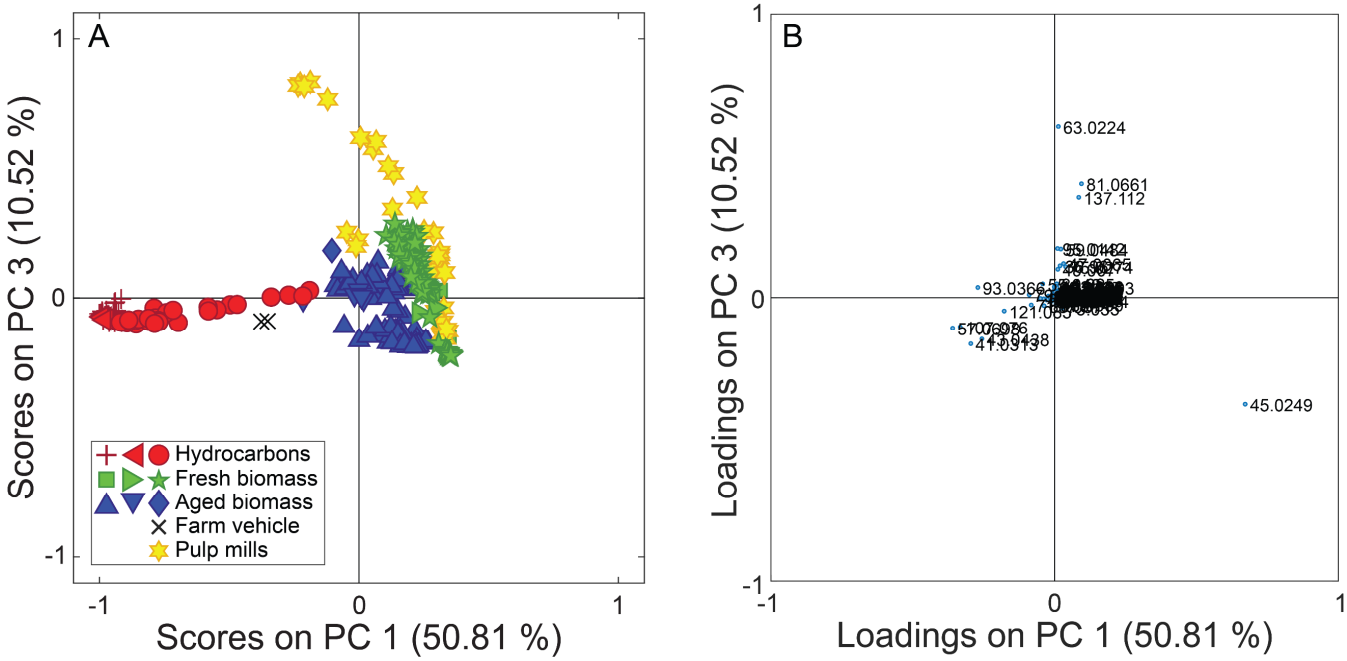


Figure B.6: Supervised PCA analysis of the PTR-ToF-MS data showing A) the PC 1 versus PC 3 scores plot, and B) the PC 1 versus PC 3 loadings plot. In this projection, the hydrocarbon samples are once again distinct from the biogenic samples, but there is less distinction between the aged biomass, fresh biomass, and pulp mill samples. Dimethyl sulphide (DMS) at *m/z* 63.022, which has a high loading on PC 3, leads to the pulp mill samples (higher abundances of DMS have more positive scores on PC 3).

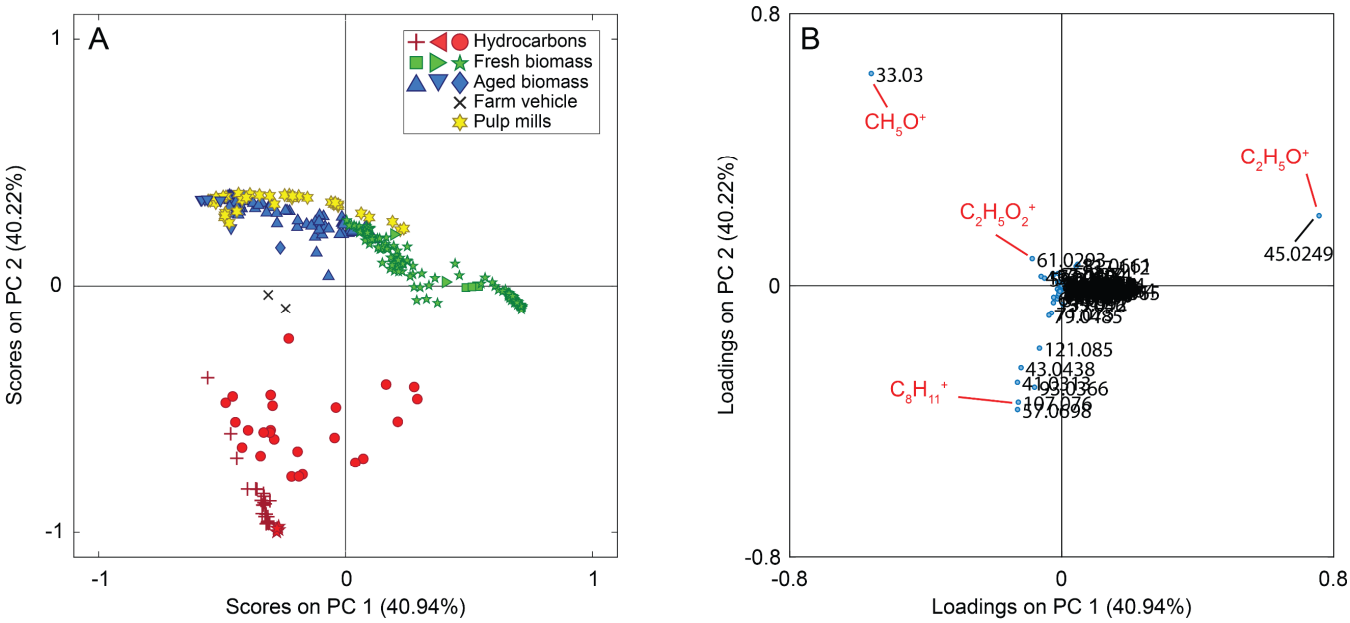


Figure B.7: PC 1 versus PC 2 scores plot (left) and loadings plot (right) for the PTR-ToF-MS analysis with the *m/z* methanol included as a variable in the data set. While the hydrocarbon (red) samples are discriminated from the biomass samples, there is significant overlap between the pulp mill samples and aged biomass samples.

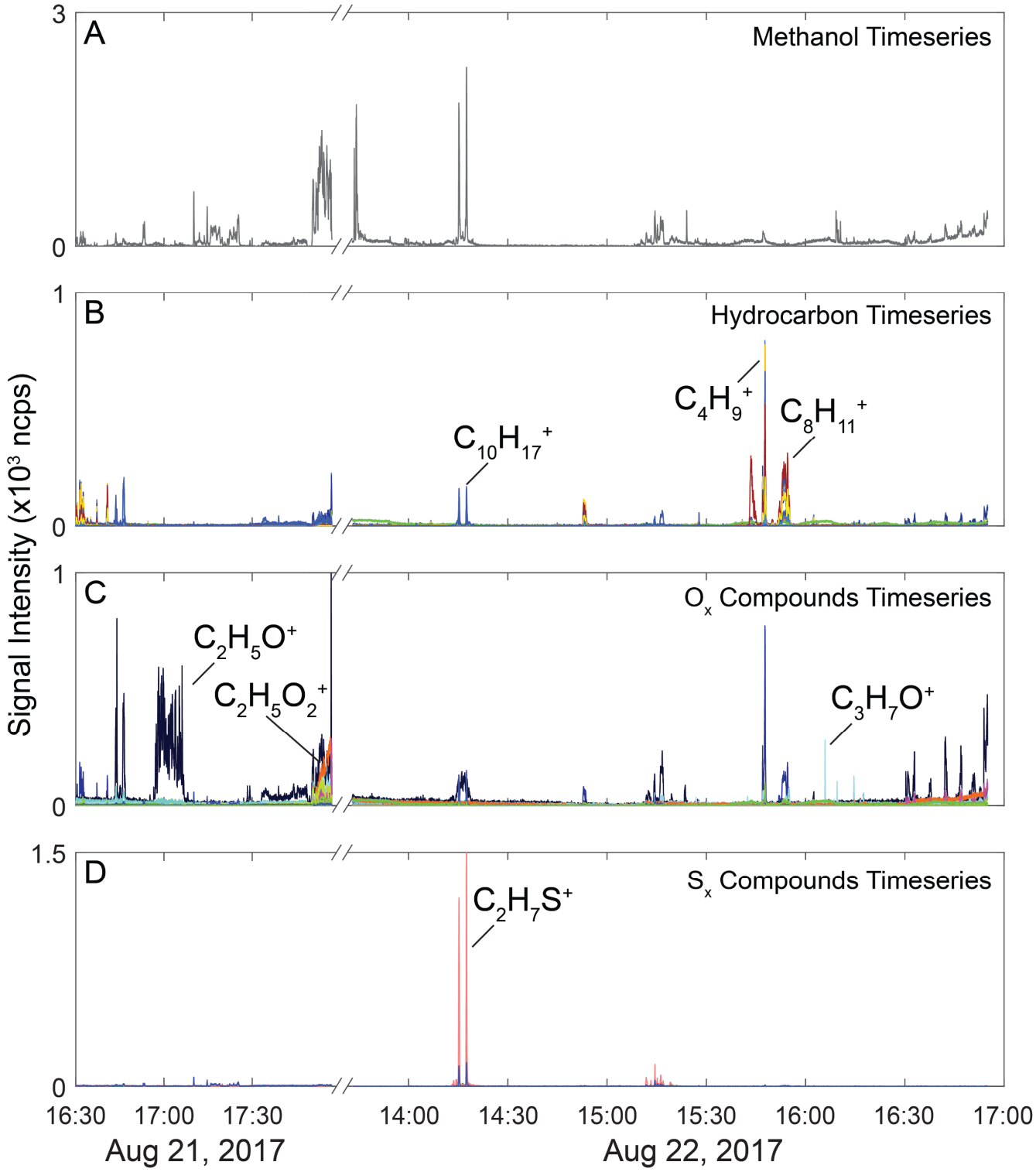


Figure B.8: A) Methanol time series. B) Time series of protonated ions and fragments from hydrocarbon species. C) Time series of protonated ions and fragments from oxygenated species. D) Sulphur compound time series.

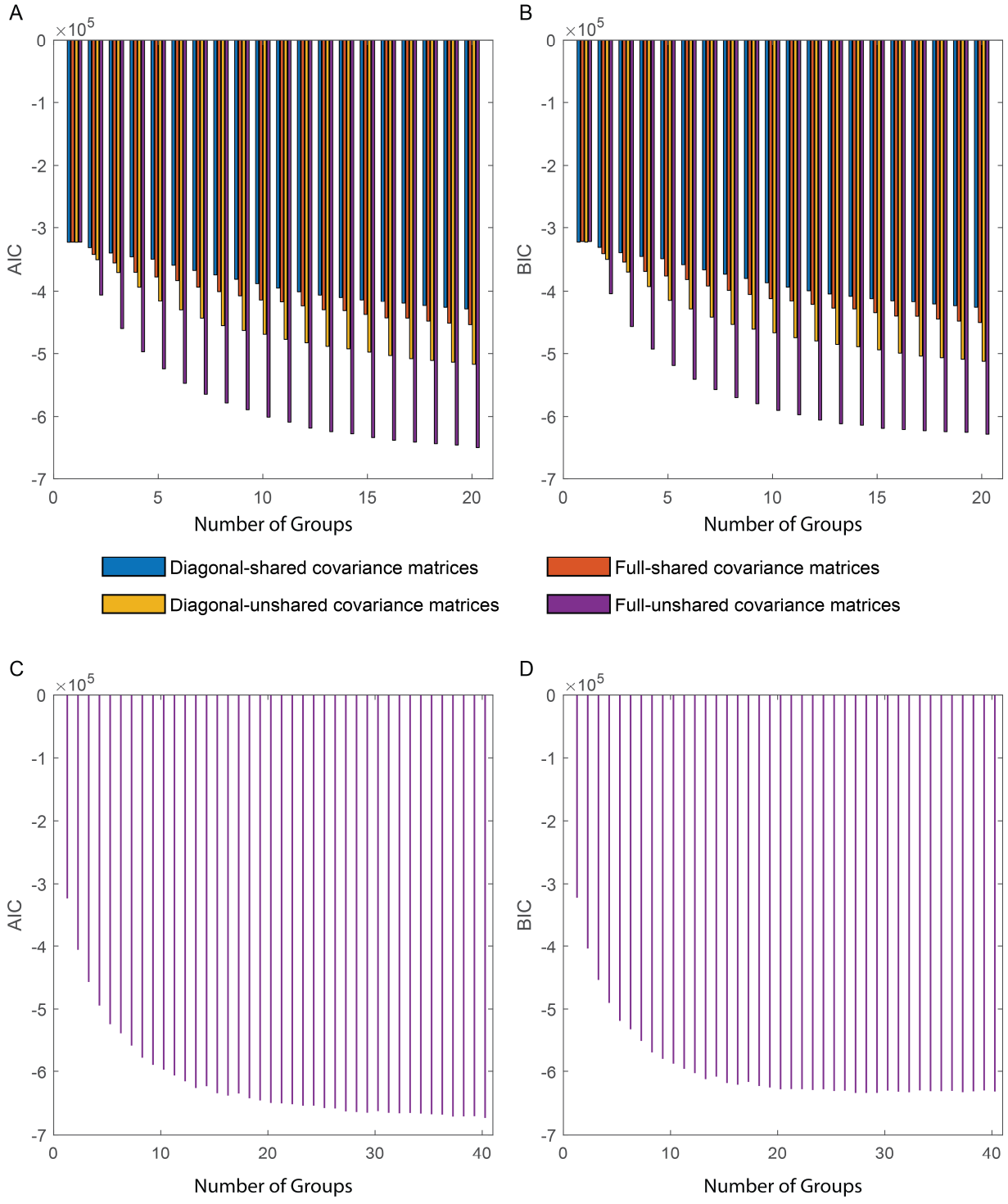


Figure B.9: A and B) AIC and BIC values (respectively) for 1–20 cluster GMM distribution models with 4 types of covariance matrices: diagonal, shared covariance matrices; full, shared covariance matrices, diagonal, unshared covariance matrices; and full, unshared covariance matrices. In all cases the full, unshared covariance matrices minimized the AIC and BIC values. Panels C and D) AIC and BIC values (respectively) for 1–40 cluster GMM models with full, unshared covariance models.

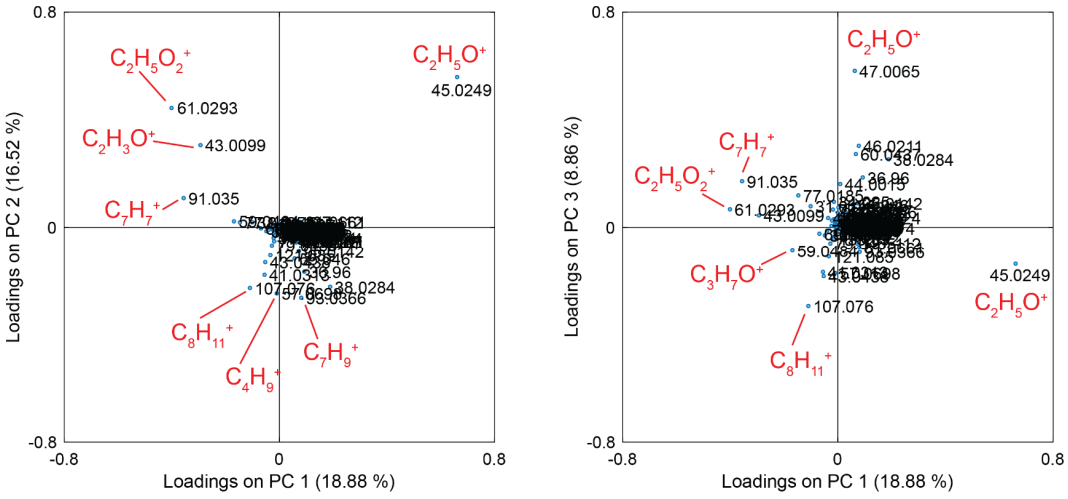


Figure B.10: Loadings plots for the unsupervised PCA of the PTR-ToF-MS data. A) Loadings for PC 1 versus PC 2. B) Loadings for PC 1 versus PC 3. Chemical formulae of some ions of interest have been labelled.

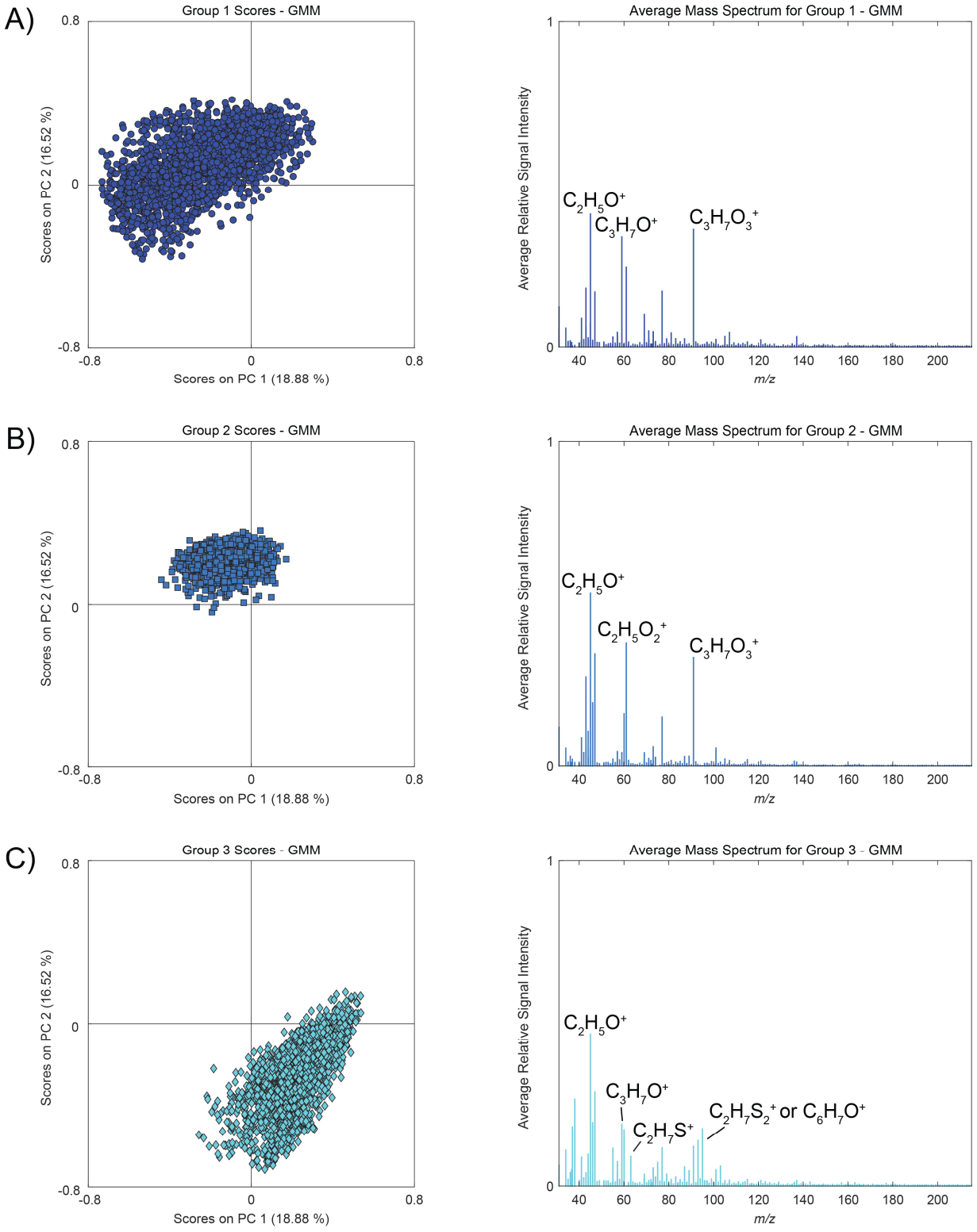


Figure B.11: The PC 1 versus PC 2 scores (left column) and average mass spectra (right column) for each of the 12 groups calculated using the GMM clustering algorithm. Chemical formulae for some important *m/z* are identified.

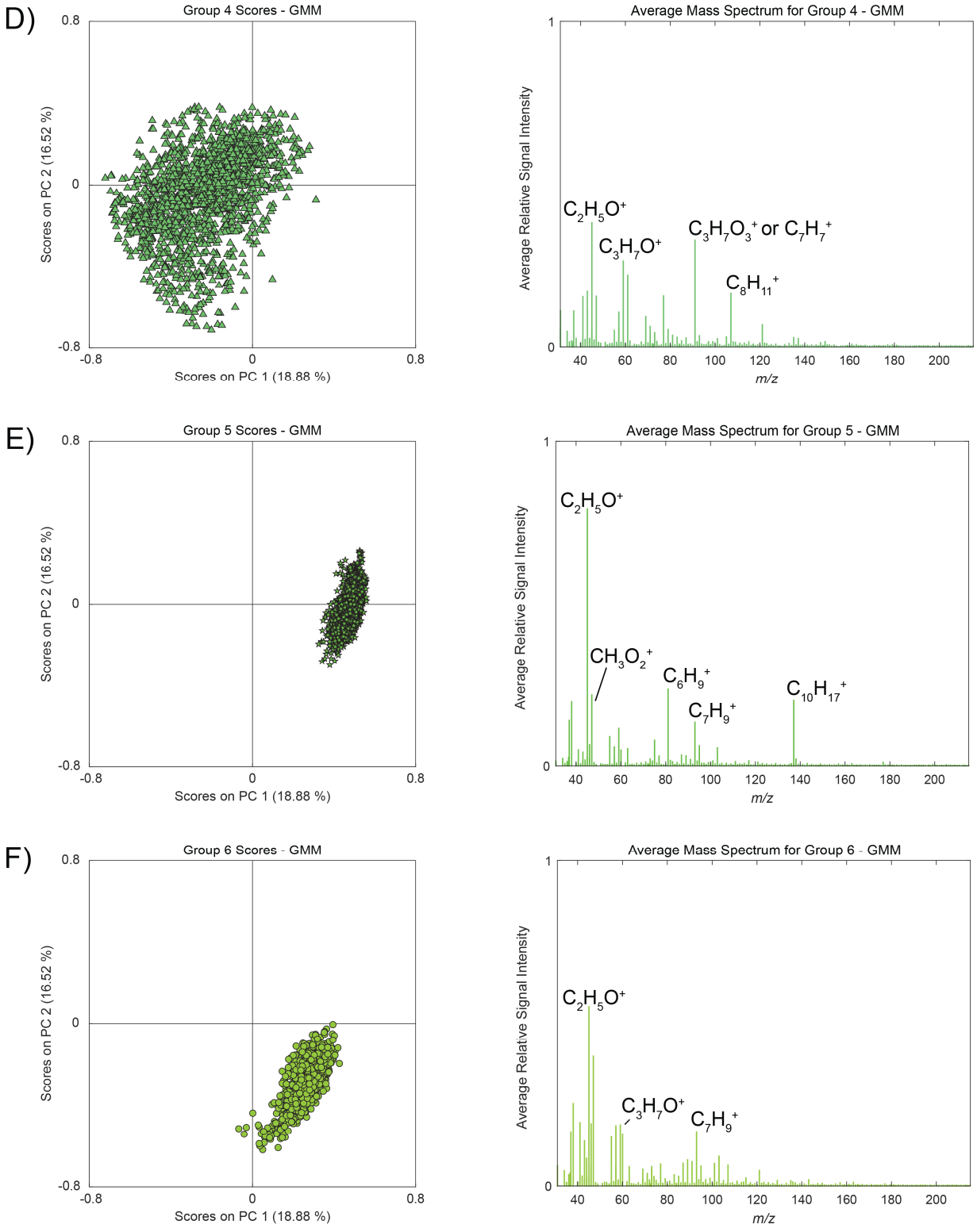


Figure B.11: (Continued) The PC 1 versus PC 2 scores (left column) and average mass spectra (right column) for each of the 12 groups calculated using the GMM clustering algorithm. Chemical formulae for some important *m/z* are identified.

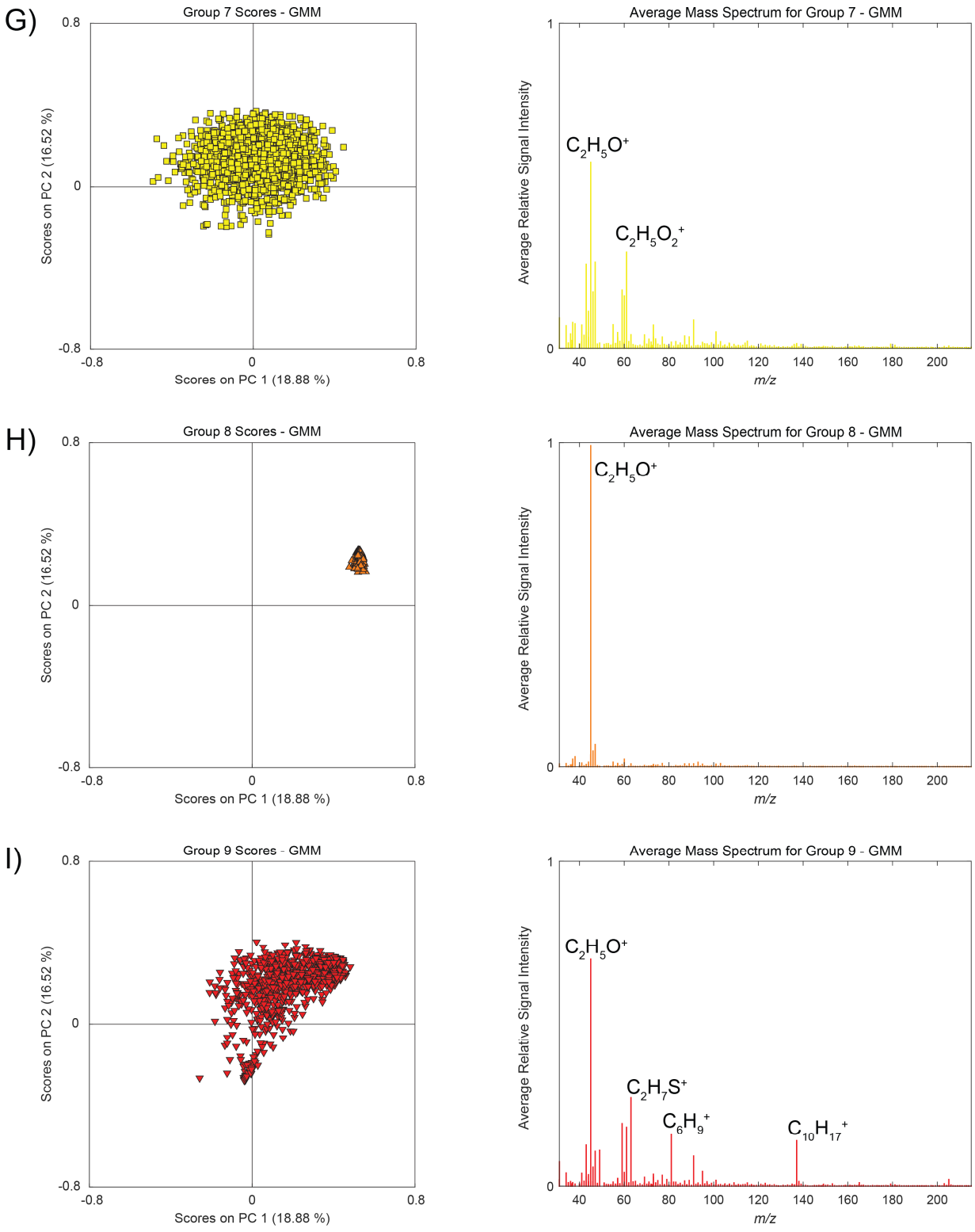


Figure B.11: (Continued) The PC 1 versus PC 2 scores (left column) and average mass spectra (right column) for each of the 12 groups calculated using the GMM clustering algorithm. Chemical formulae for some important *m/z* are identified.

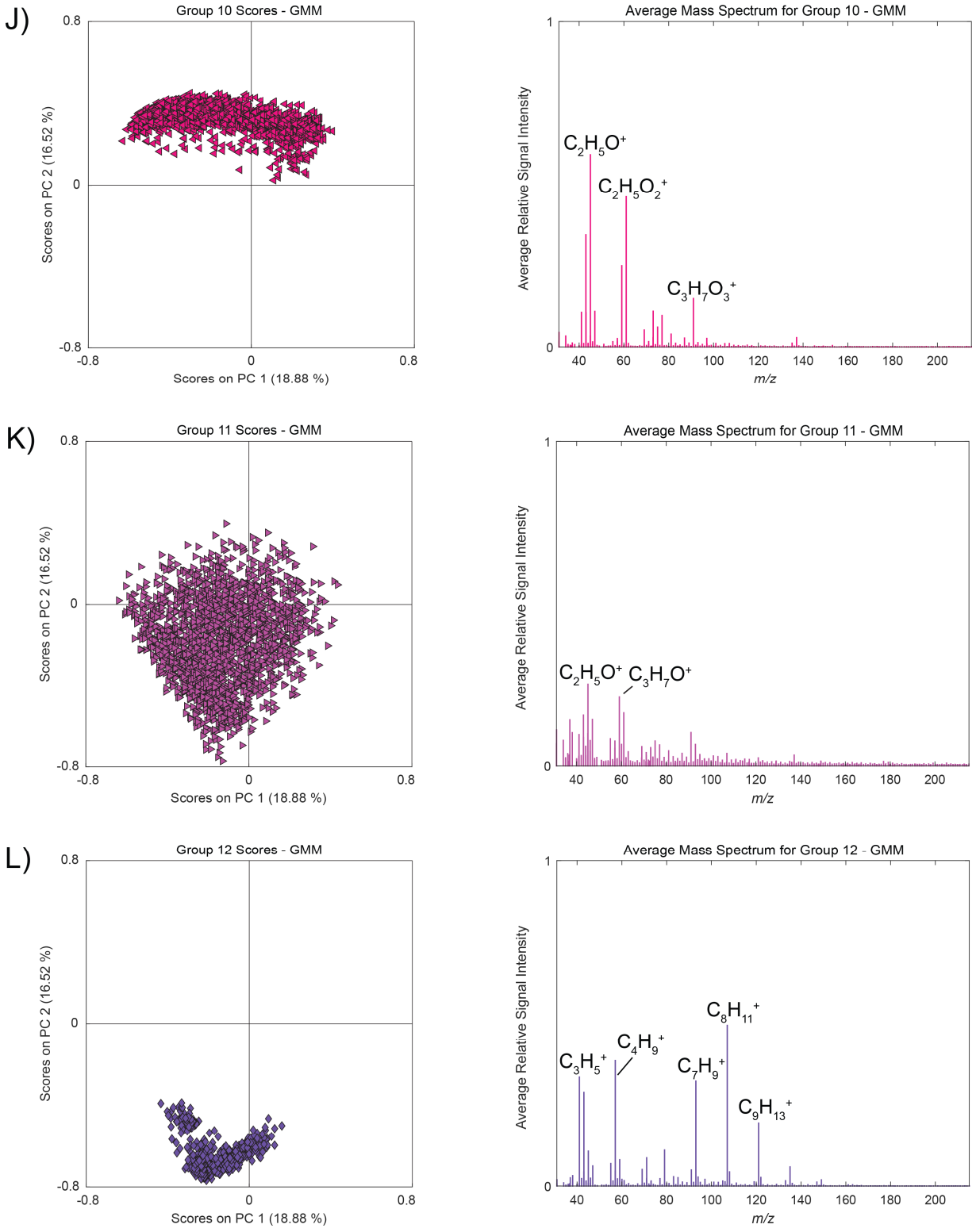


Figure B.11: (Continued) The PC 1 versus PC 2 scores (left column) and average mass spectra (right column) for each of the 12 groups calculated using the GMM clustering algorithm. Chemical formulae for some important *m/z* are identified.

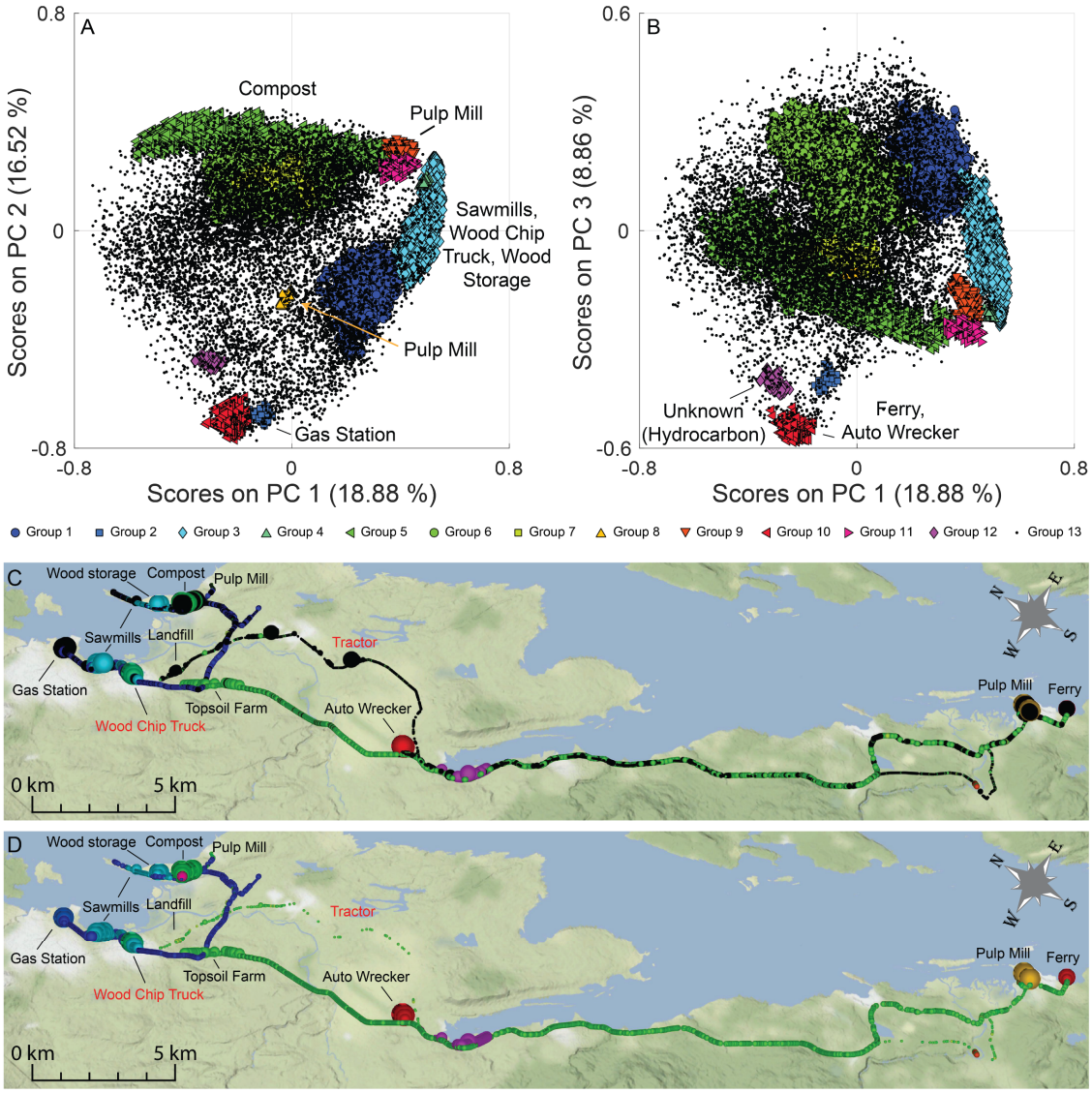


Figure B.12: Unsupervised PCA analysis of PTR-ToF-MS data, with clusters determined by DBSCAN. A) PCA scores plots for PC 1 versus PC 2. B) PCA scores plot for PC 1 versus PC 3. Samples are coloured based on the groups identified using the DBSCAN algorithm. The data points in black belong to low density regions and were not grouped with the algorithm. Some groups containing individual VOC sources (such as the municipal compost facility, and one of the pulp mills) or multiple VOC sources of the same type (*e.g.,* most sawmill emissions are grouped with the samples from the wood chip truck) are labeled. C) Geospatial map showing the spatial distribution of the samples, with dot size proportional to total VOC concentration. Locations of interest are labelled, with moving sources being represented by red text. D) Geospatial map with the black dots removed, to show the distribution of the clustered samples. The majority of high concentration VOC samples were grouped using the algorithm.

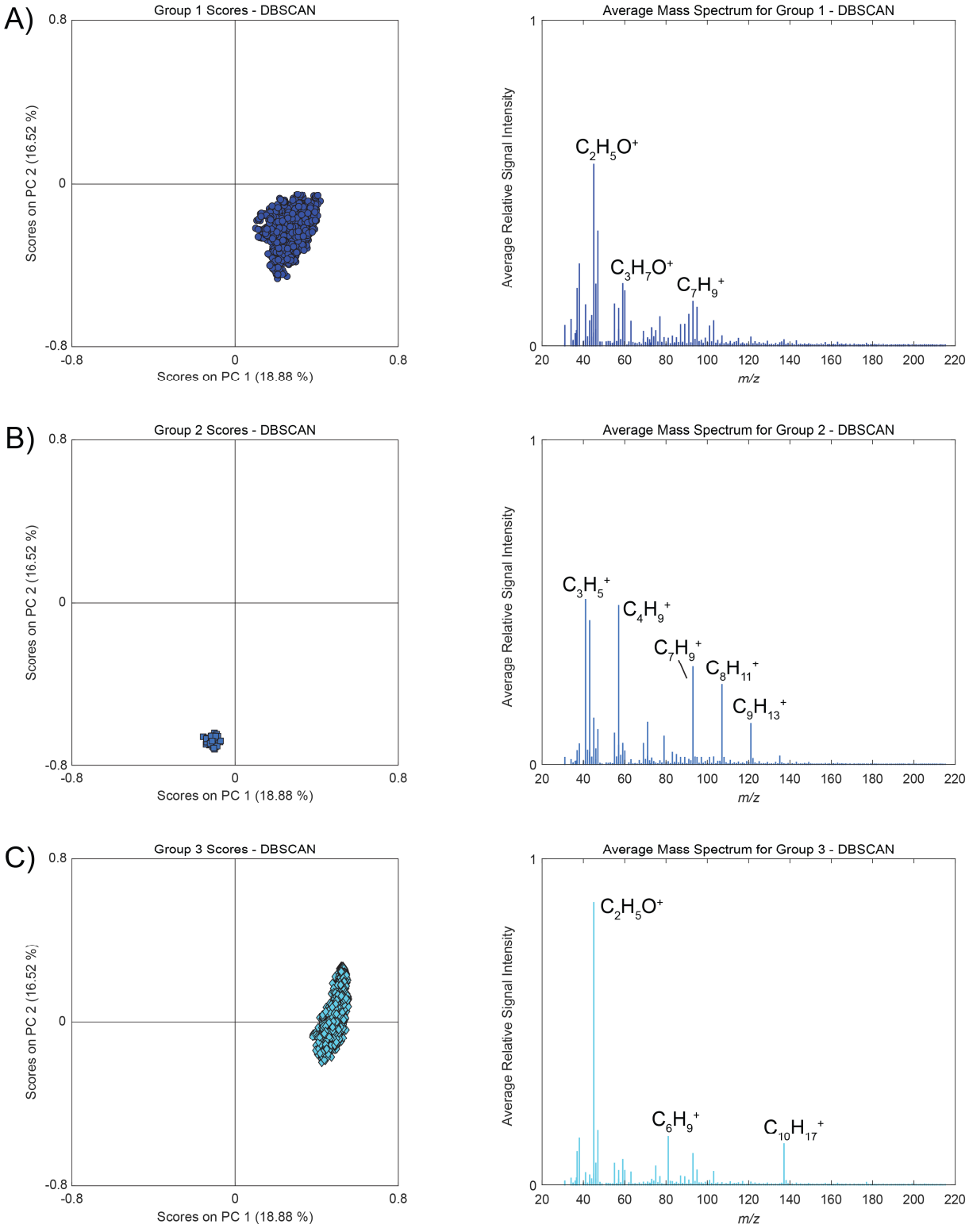


Figure B.13: The PC 1 versus PC 2 scores (left column) and average mass spectra (right column) for each of the 13 groups calculated using the DBSCAN clustering algorithm. Chemical formlae for some important *m/z* are identified.

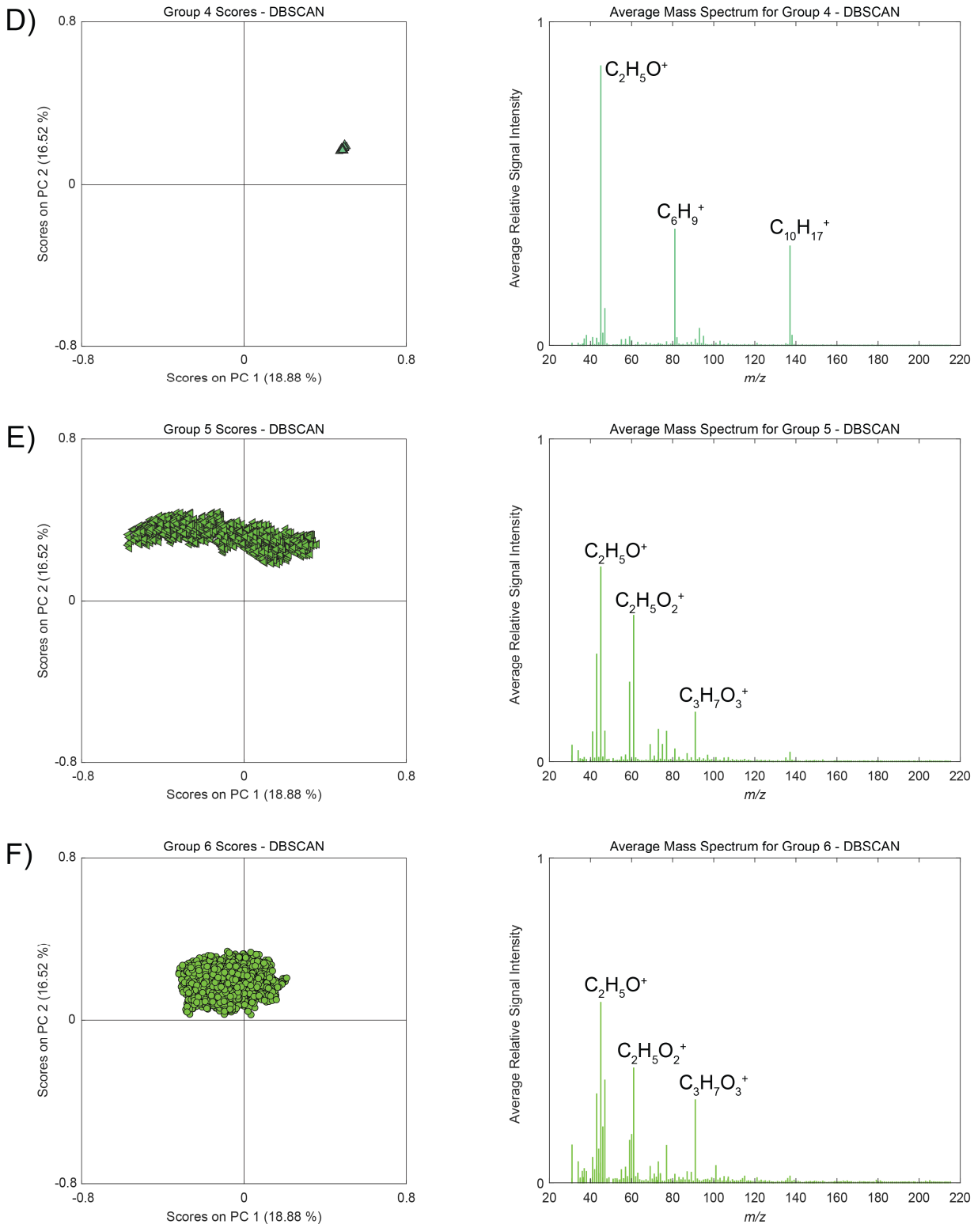


Figure B.13: (Continued) The PC 1 versus PC 2 scores (left column) and average mass spectra (right column) for each of the 13 groups calculated using the DBSCAN clustering algorithm. Chemical formlae for some important *m/z* are identified.

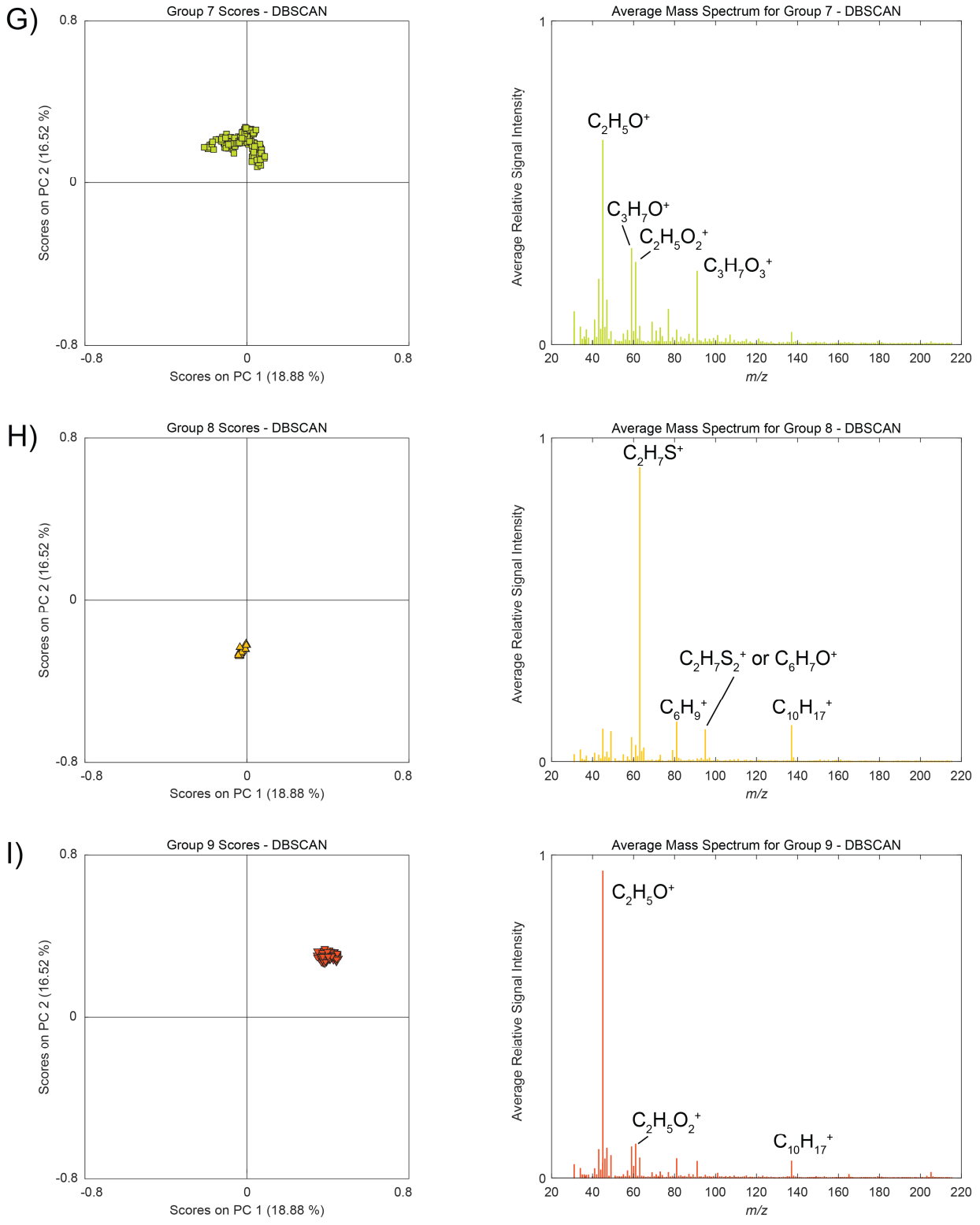


Figure B.13: (Continued) The PC 1 versus PC 2 scores (left column) and average mass spectra (right column) for each of the 13 groups calculated using the DBSCAN clustering algorithm. Chemical formlae for some important *m/z* are identified.

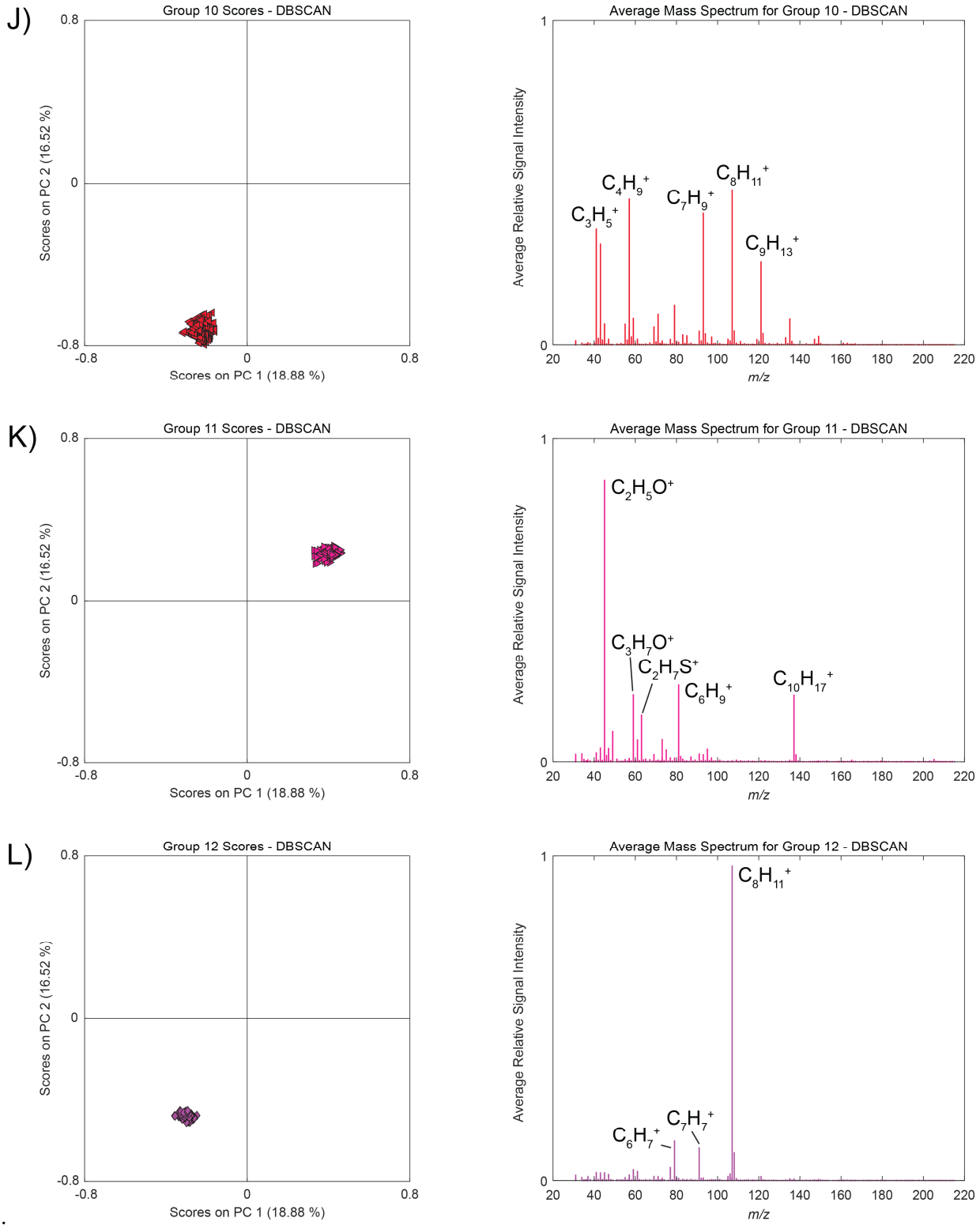


Figure B.13: (Continued) The PC 1 versus PC 2 scores (left column) and average mass spectra (right column) for each of the 13 groups calculated using the DBSCAN clustering algorithm. Chemical formlae for some important *m/z* are identified.

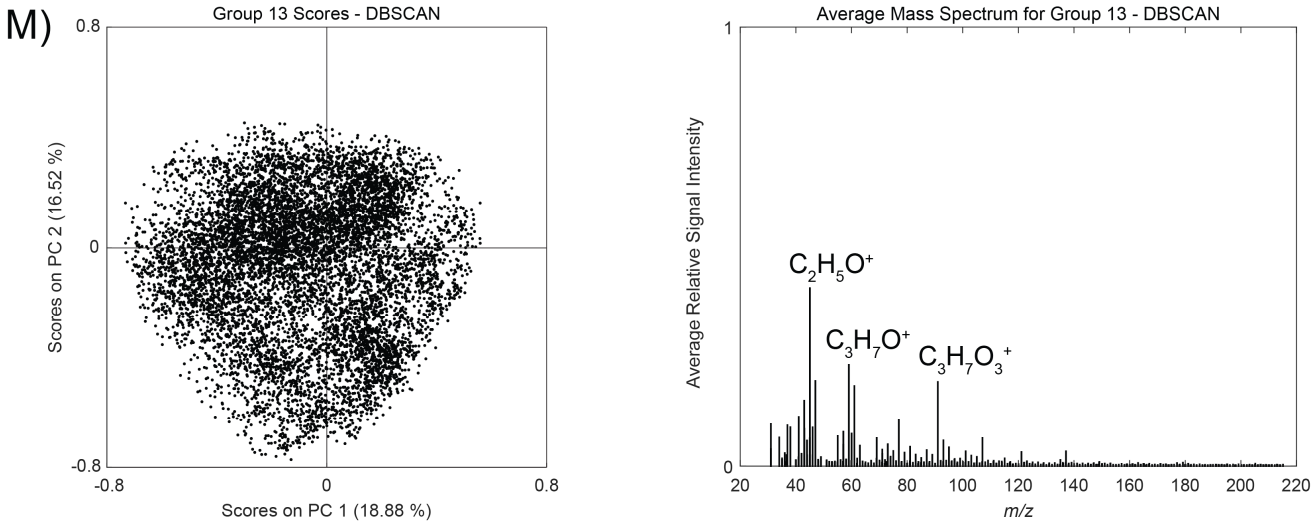


Figure B.13: (Continued) The PC 1 versus PC 2 scores (left column) and average mass spectra (right column) for each of the 13 groups calculated using the DBSCAN clustering algorithm. Chemical formlae for some important *m/z* are identified.

Table B.6: Correlation matrix between the mass spectra for the groups identified by GMM and the groups identified by DBSCAN.

		DBSCAN Group												
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
GMM Group	**1**	0.636	0.263	0.583	0.528	0.872	0.869	0.934	0.117	0.665	0.208	0.642	0.145	0.921
	2	0.708	0.246	0.636	0.566	0.848	0.987	0.866	0.100	0.710	0.138	0.596	0.085	0.882
	3	0.984	0.372	0.780	0.604	0.556	0.723	0.711	0.220	0.671	0.226	0.636	0.057	0.834
	4	0.680	0.468	0.590	0.514	0.837	0.843	0.898	0.108	0.639	0.433	0.605	0.313	0.937
	5	0.854	0.294	0.980	0.936	0.655	0.684	0.762	0.211	0.883	0.175	0.910	0.027	0.757
	6	0.983	0.563	0.793	0.624	0.576	0.735	0.711	0.156	0.689	0.396	0.634	0.102	0.842
	7	0.815	0.264	0.769	0.665	0.888	0.967	0.916	0.164	0.819	0.134	0.733	0.054	0.920
	8	0.709	0.166	0.950	0.876	0.678	0.674	0.771	0.108	0.976	0.069	0.897	0.019	0.657
	9	0.716	0.194	0.869	0.853	0.825	0.801	0.901	0.477	0.921	0.105	0.951	0.043	0.818
	10	0.596	0.259	0.672	0.619	0.999	0.892	0.933	0.127	0.773	0.181	0.731	0.070	0.870
	11	0.807	0.446	0.626	0.503	0.796	0.833	0.850	0.196	0.609	0.330	0.598	0.120	0.948
	12	0.358	0.916	0.206	0.145	0.218	0.210	0.237	0.036	0.149	0.990	0.146	0.604	0.398

Appendix C - Supplemental information for Chapter 4

MCR-ALS analysis of the August 7, 2018 data

MCR-ALS models of the 1 second measurements for the August 7, 2018 Port Alberni data were calculated with the whole set of measured *m/z* as well as with the selection of *m/z* with a maximum S/N > 1 over the course of the day. These models were calculated to compare with the MCR-WALS models of the same datasets.

MCR-ALS analysis of the August 7, 2018 data using all measured m/z

Six–ten component MCR-ALS models were calculating using the full set of *m/z,* and the Meisch coefficients for the models are shown in Table C.1 with the coefficients greater than 0.9 bolded for emphasis. The results from the six, seven, and eight component models are shown in Figure C.1, Figure C.3, and Figure C.5. In each of these models the monoterpene signals are modelled poorly, as shown in the bottom panels of Figure C.2, Figure C.4, and Figure C.6, and the isoprene signal is modelled poorly in the six and seven component MCR-ALS models (top panels of Figure C.2 and Figure C.4), and well in the eight component model (top panel of Figure C.6). The MCR-ALS models have more degenerate components compared to the MCR-WALS models calculated on the same data, with the forest fire VOC component being split across multiple components in each of these models. Additionally, the signal intensity measured at *m/z* 137.12 is not modelled adequately (Meisch coefficient of 0.611) until a ten component model is calculated, but the ten component MCR-ALS model without variable selection splits the forest fire component across four components which is not a useful result.

Table C.1: Meisch coefficients for a selection of *m/z* measured in the August 7, 2018 field campaign for the six–ten component MCR-ALS models calculated using the full mass list. Values greater than 0.9 are bolded for emphasis.

	Number of MCR-ALS components				
m/z	**6**	**7**	**8**	**9**	**10**
33.03	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
41.03	**0.927**	**0.928**	**0.959**	**0.959**	**0.960**
42.03	0.334	0.336	0.351	0.423	0.422
43.01	0.889	**0.988**	**0.993**	**1.00**	**1.00**
43.05	0.772	0.775	0.802	0.805	0.809
45.03	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
47.01	**0.997**	**0.999**	**0.999**	**1.00**	**1.00**
55.01	0.136	0.135	0.137	0.142	0.142
57.03	0.436	0.437	0.455	0.571	0.570
57.07	0.869	0.870	**0.912**	**0.913**	**0.912**
59.05	**0.998**	**0.999**	**0.998**	**0.999**	**1.00**
61.03	0.742	**0.965**	**0.953**	**0.999**	**0.999**
69.07	0.074	0.075	0.738	0.751	0.770
71.04	0.363	0.368	0.372	0.385	0.415
73.06	0.463	0.468	0.471	0.489	0.573
75.04	0.377	0.384	0.388	0.421	0.422
79.05	0.686	0.682	0.694	0.734	0.735
81.07	0.233	0.233	0.241	0.293	0.857
85.06	0.363	0.367	0.372	0.415	0.416
87.04	0.358	0.362	0.363	0.375	0.387
93.07	0.779	0.779	0.816	0.818	0.850
97.09	0.454	0.451	0.462	0.563	0.563
99.04	0.482	0.483	0.490	0.587	0.587
107.08	0.825	0.826	0.861	0.862	0.864
121.09	0.658	0.658	0.693	0.694	0.700
137.12	0.197	0.196	0.200	0.247	0.611

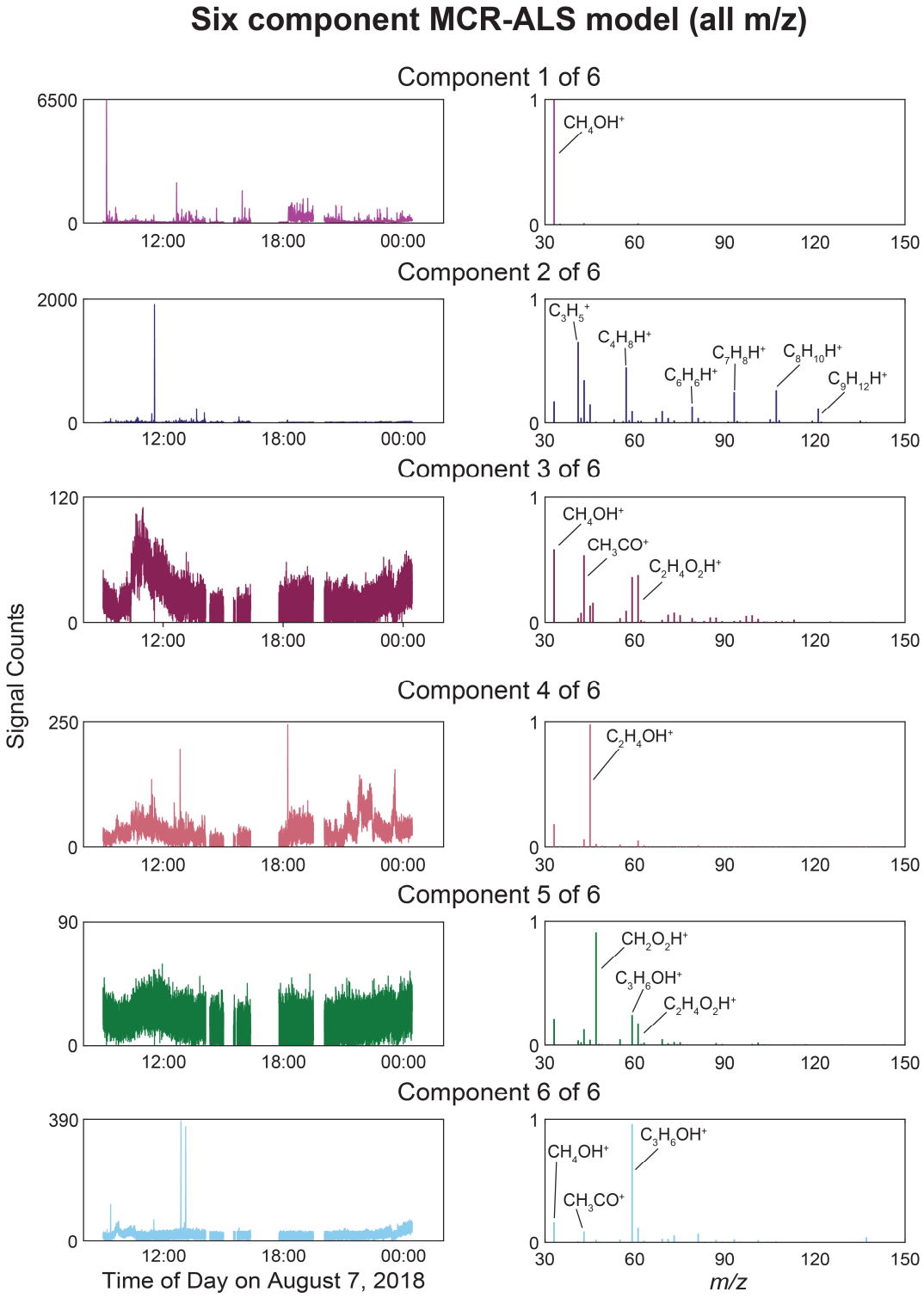


Figure C.1: Six component MCR-ALS model using all the measured *m/z*.

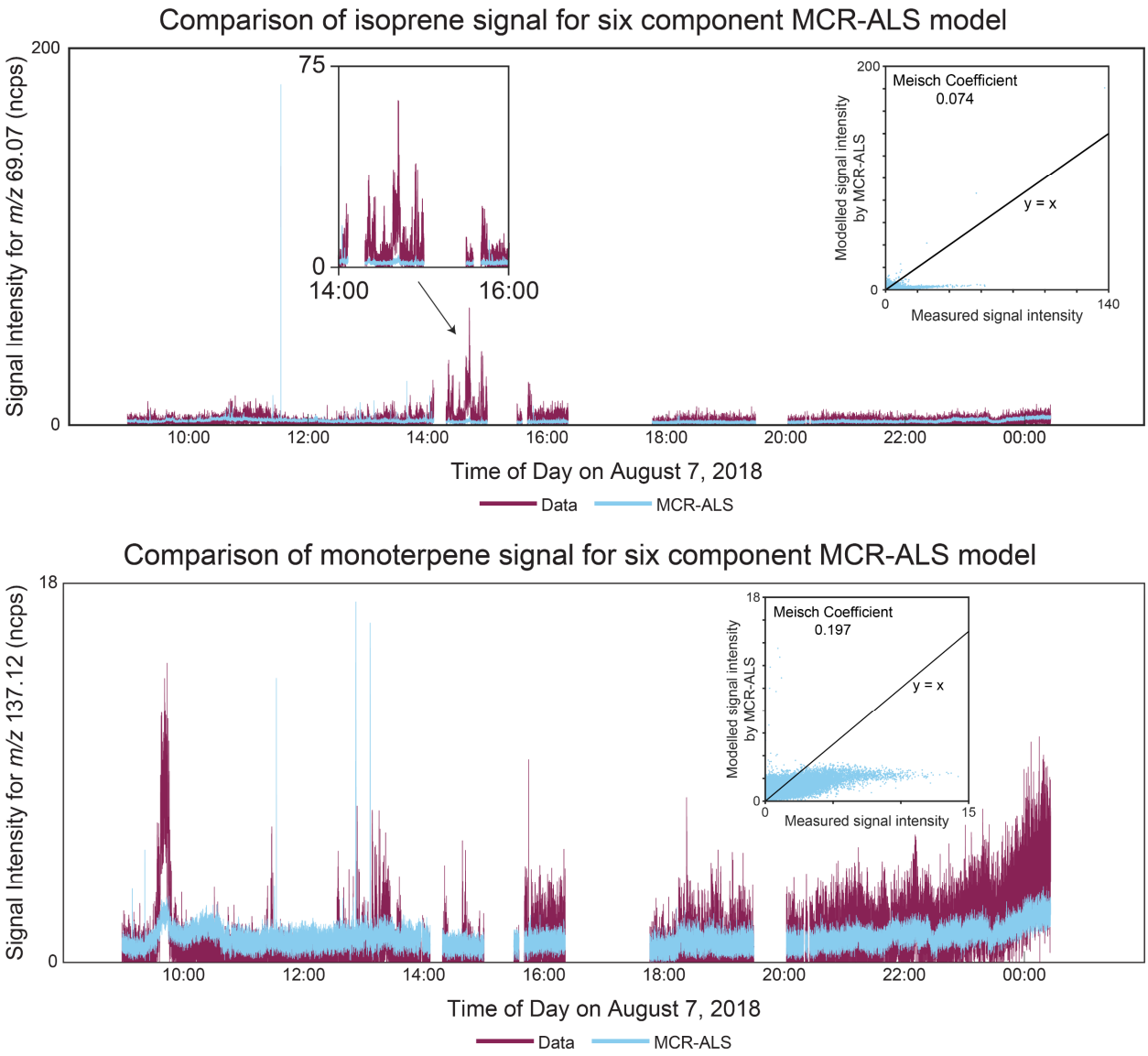
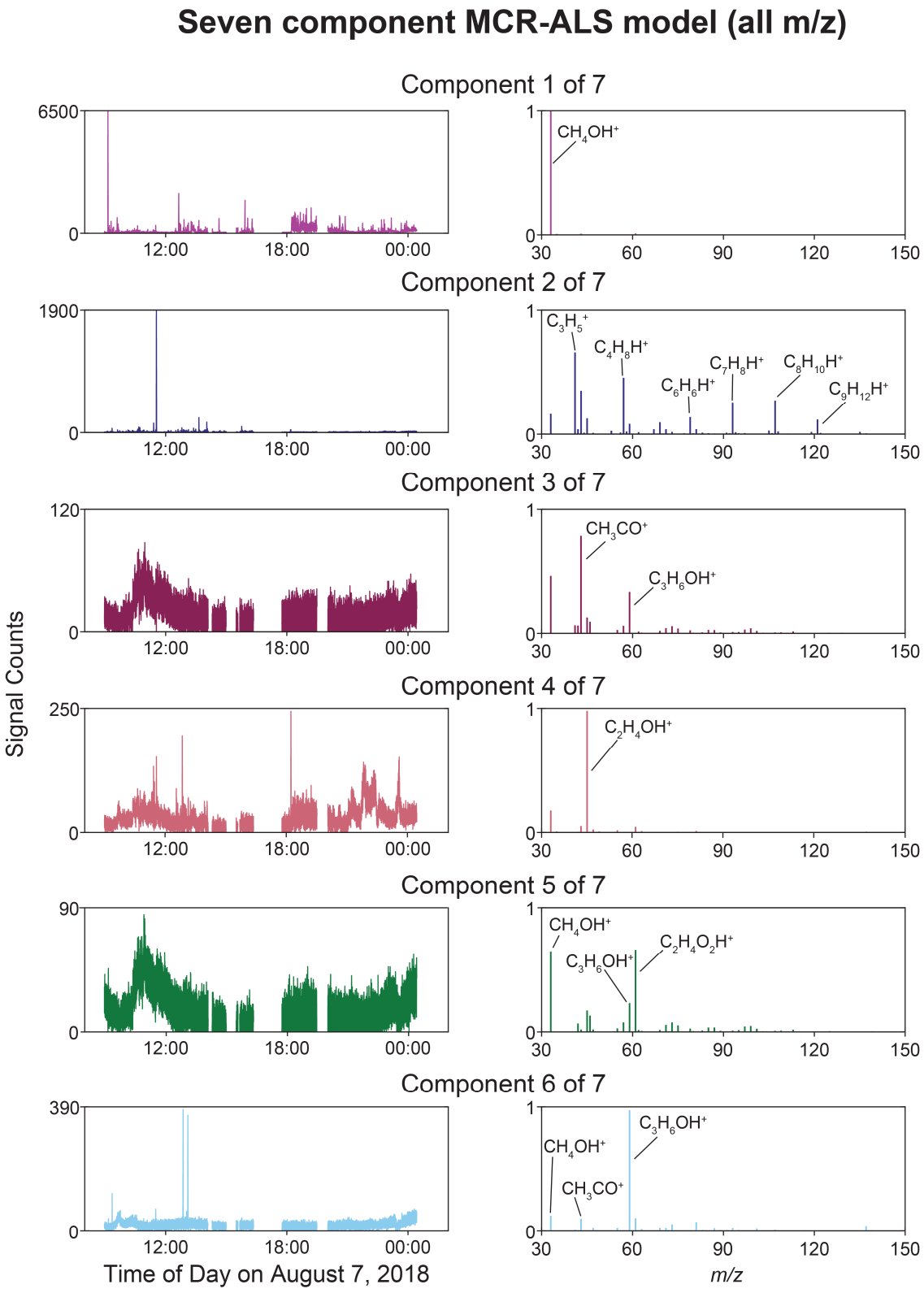


Figure C.2: Top: Overlay of measured (burgundy) and modelled (blue) signal intensity for *m/z* 69.07 in the six component MCR-ALS model with no variable selection. Bottom: Overlay of measured (burgundy) and modelled (blue) signal intensity for *m/z* 137.12 in the six component MCR-ALS model with no variable selection. Neither of these *m/z* are modelled well with six components.


Seven component MCR-ALS model (all m/z)
Component 1 of 7
Component 2 of 7
Component 3 of 7
Component 4 of 7
Component 5 of 7
Component 6 of 7
Signal Counts
Time of Day on August 7, 2018
m/z
CH4OH+
C3H5+
C4H8H+
C6H6H+
C7H8H+
C8H10H+
C9H12H+
CH3CO+
C3H6OH+
C2H4OH+
C2H4O2H+
12:00
18:00
00:00
30
60
90
120
150

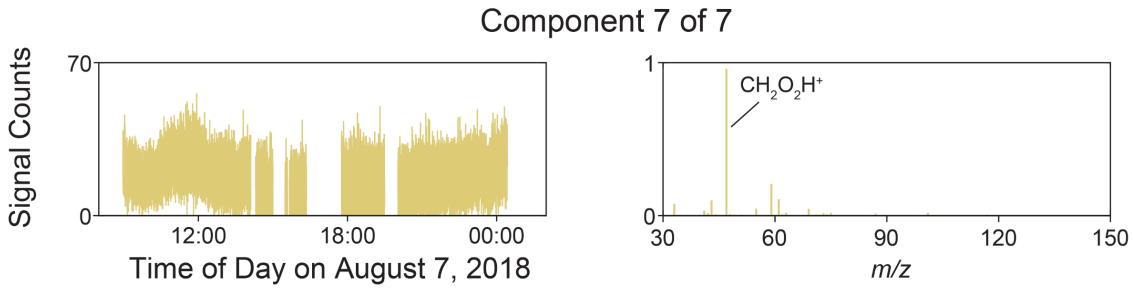
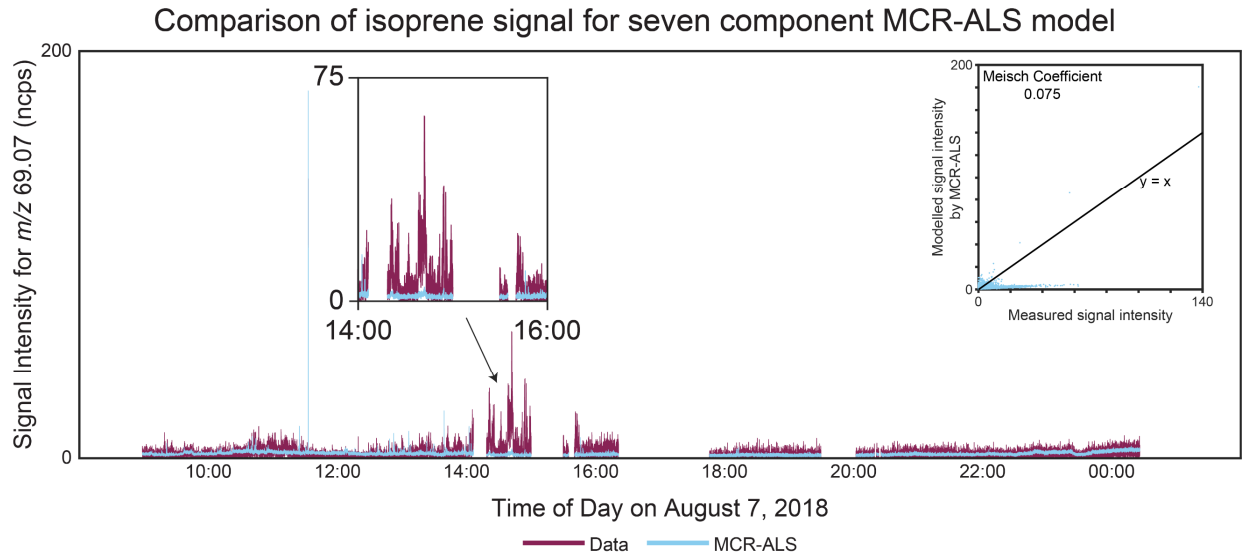


Figure C.3: Seven component MCR-ALS model with no variable selection.



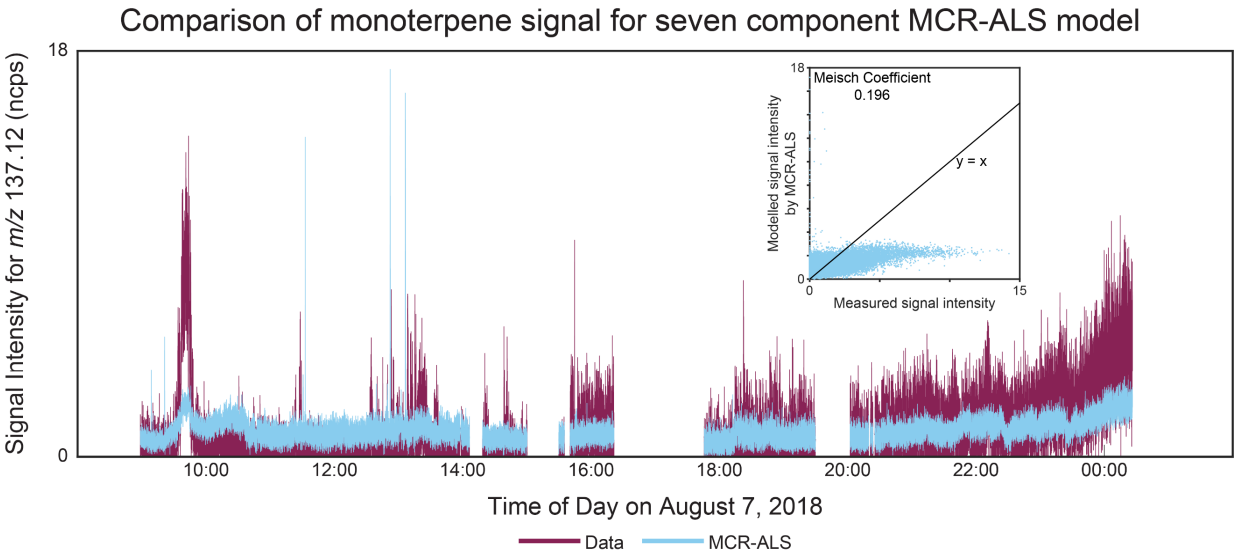
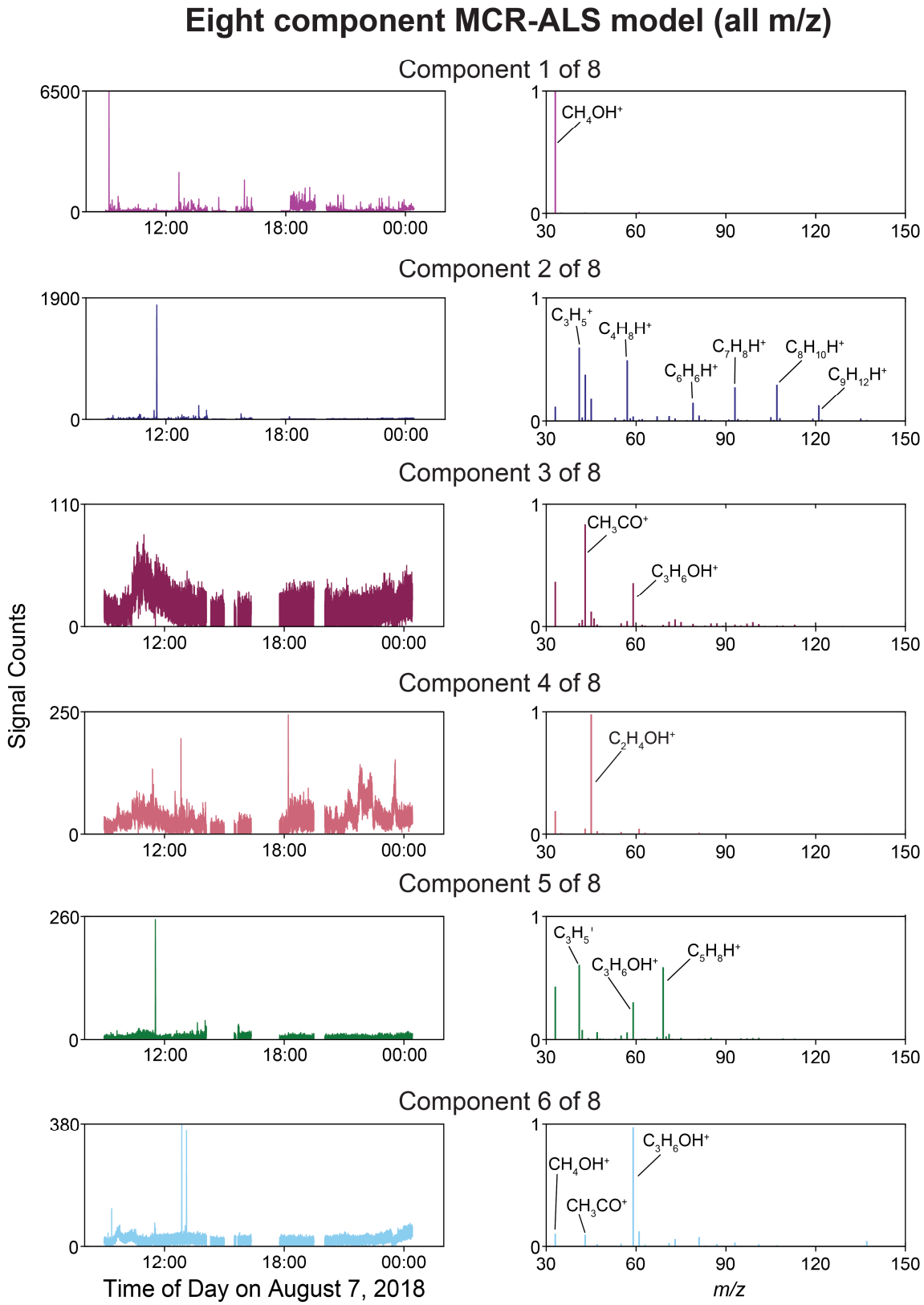


Figure C.4: Top: Overlay of measured (burgundy) and modelled (blue) signal intensity for *m/z* 69.07 in the seven component MCR-ALS model with no variable selection. Bottom: Overlay of measured (burgundy) and modelled (blue) signal intensity for *m/z* 137.12 in the seven component MCR-ALS model with no variable selection. Neither of these *m/z* are modelled well with seven components.


Eight component MCR-ALS model (all m/z)
Component 1 of 8
Component 2 of 8
Component 3 of 8
Component 4 of 8
Component 5 of 8
Component 6 of 8
Signal Counts
Time of Day on August 7, 2018
m/z
CH4OH+
C3H5+
C4H8H+
C7H8H+
C8H10H+
C6H6H+
C9H12H+
CH3CO+
C3H6OH+
C2H4OH+
C5H8H+

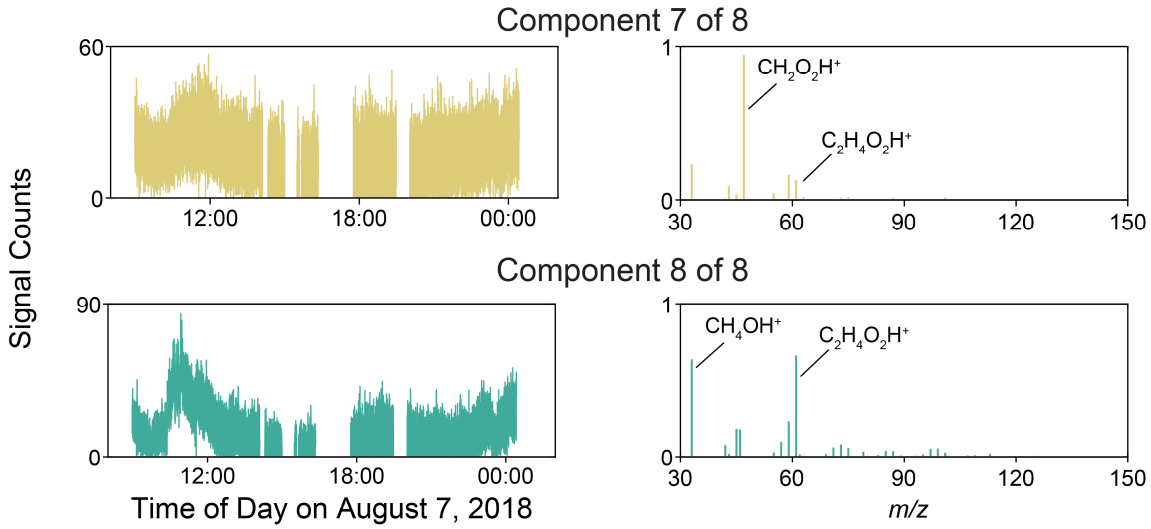


Figure C.5: Eight component MCR-ALS model with no variable selection.

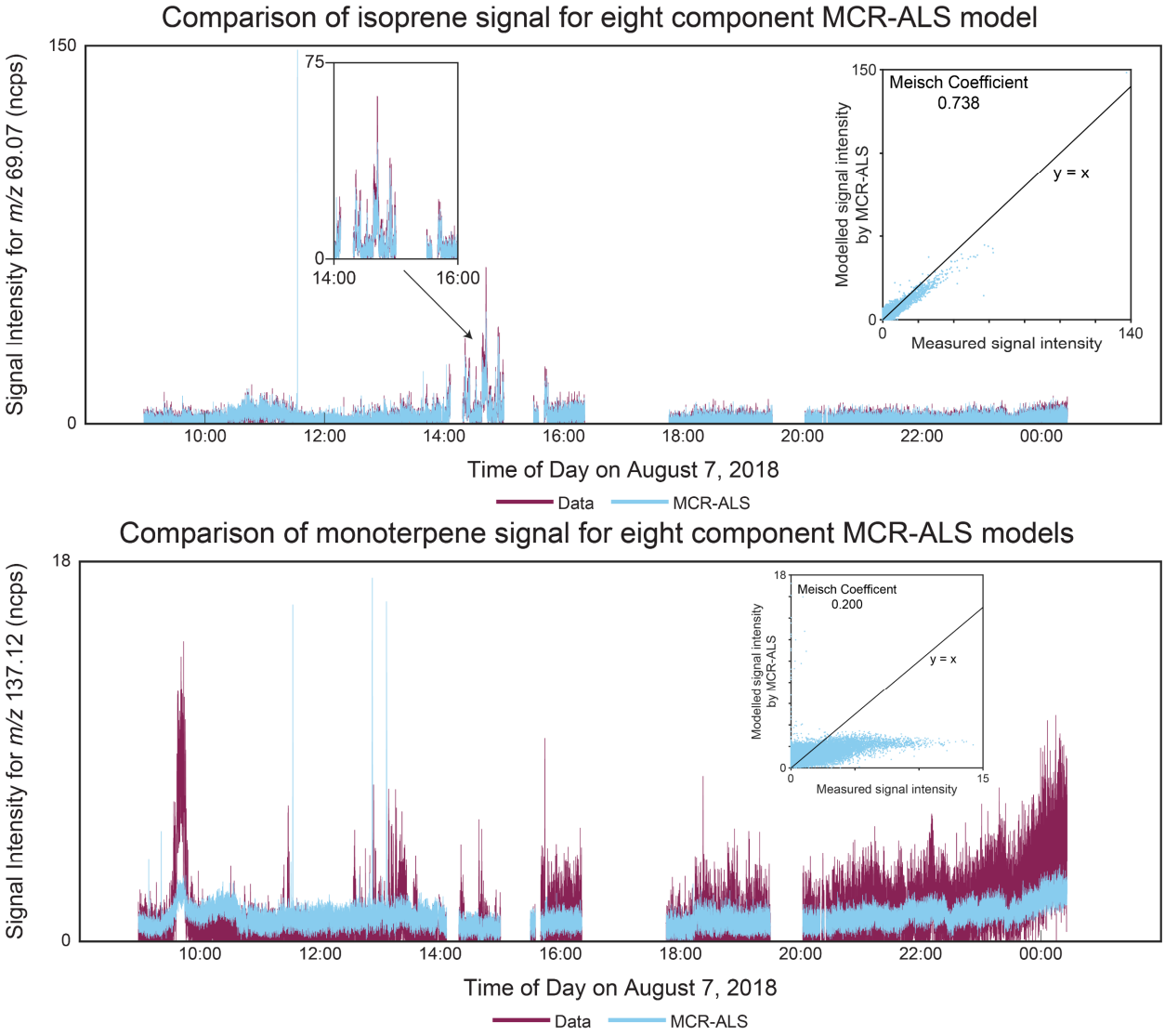


Figure C.6: Top: Overlay of measured (burgundy) and modelled (blue) signal intensity for *m/z* 69.07 in the eight component MCR-ALS model with no variable selection. Bottom: Overlay of measured (burgundy) and modelled (blue) signal intensity for *m/z* 137.12 in the eight component MCR-ALS model with no variable selection. The addition of an eight component models the *m/z* 69.07 measurements, but the observations at *m/z* 137.12 are still mostly unaccounted for.

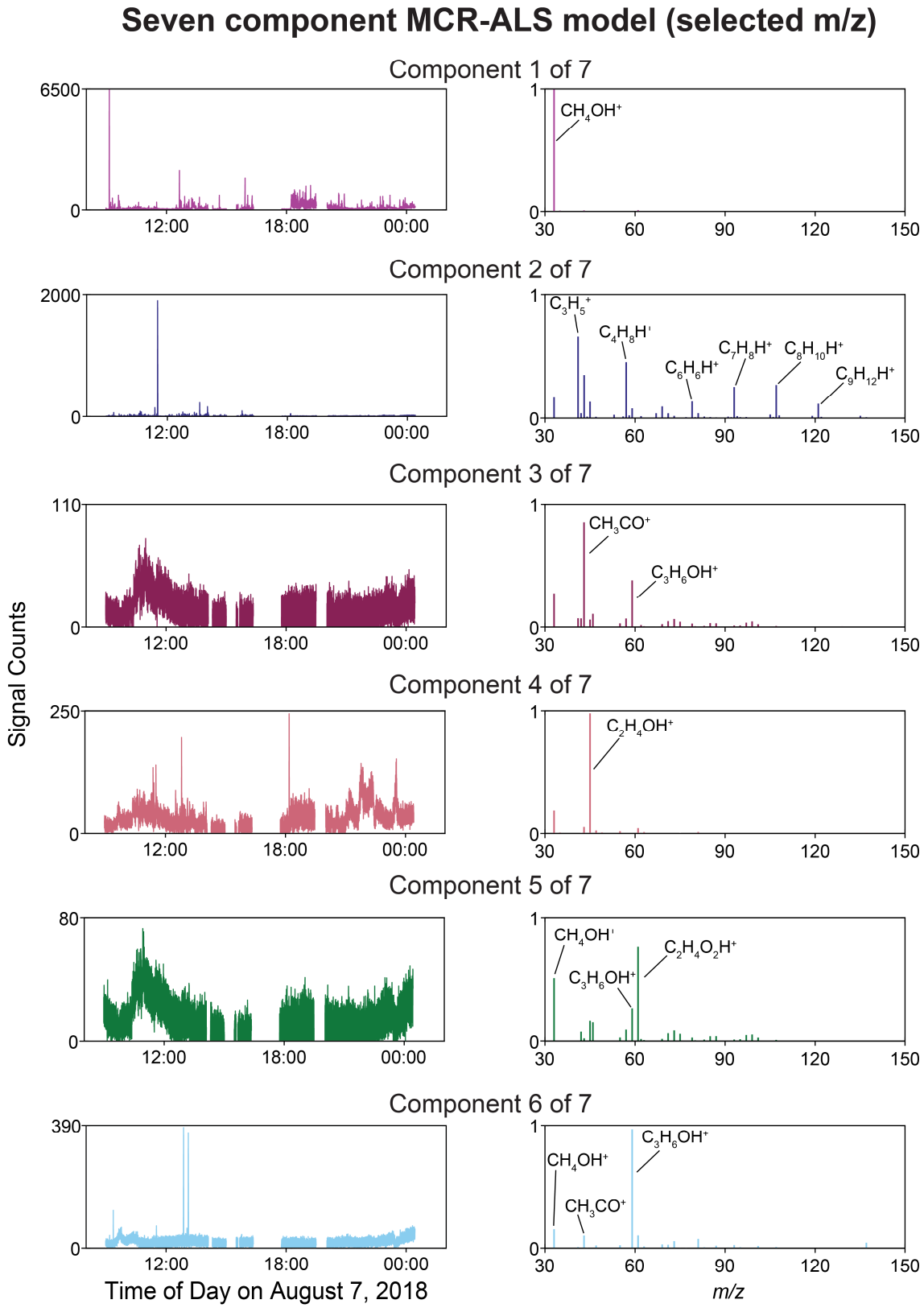
MCR-ALS analysis of the August 7, 2018 data using selected m/z

Six–twelve component MCR-ALS models were calculating using the variable selected set of *m/z,* and the Meisch coefficients for the models are shown in Table C.2 with the coefficients greater than 0.9 bolded for emphasis. The results from the seven and eight component models are shown in Figure C.7 and Figure C.8. In each of these models, the forest fire component is split across multiple components, while the signal intensity for recorded *m/z* with interesting temporal profiles related to local VOC sources are left unmodelled. As the MCR-WALS analysis resolved more meaningful components than

MCR-ALS analysis, it was chosen as the algorithm to use to identify the sources of VOCs observed during the three field campaigns.

Table C.2: Meisch coefficients for the main *m/z* measured in the August 7, 2018 field campaign for six to twelve component MCR-ALS models. Values greater than 0.9 are bolded for emphasis.

	Number of MCR-ALS components						
m/z	**6**	**7**	**8**	**9**	**10**	**11**	**12**
33.03	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
41.03	**0.927**	**0.928**	**0.959**	**0.959**	**0.959**	**0.999**	**0.999**
42.03	0.333	0.335	0.347	0.422	0.575	0.428	0.427
43.01	0.889	**0.988**	**0.991**	**0.999**	**0.999**	**0.999**	**0.999**
43.05	0.772	0.775	0.803	0.805	0.806	0.833	0.834
45.03	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
57.03	0.434	0.434	0.449	0.568	0.770	0.565	0.565
57.07	0.869	0.870	**0.912**	**0.912**	**0.916**	**0.931**	**0.931**
59.05	**0.998**	**0.998**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**
61.03	0.742	**0.968**	**0.961**	**0.999**	**0.999**	**0.999**	**0.999**
69.07	0.074	0.075	0.741	0.753	0.767	**0.989**	**0.990**
71.04	0.363	0.367	0.371	0.384	0.444	0.414	0.415
73.06	0.463	0.467	0.470	0.488	0.522	0.568	0.570
75.04	0.377	0.383	0.386	0.420	0.449	0.422	0.421
79.05	0.686	0.682	0.694	0.734	0.756	0.736	0.736
81.07	0.232	0.232	0.239	0.292	0.310	0.871	0.872
85.06	0.362	0.365	0.369	0.414	0.454	0.414	0.415
87.04	0.358	0.362	0.362	0.375	0.388	0.386	0.386
93.07	0.779	0.779	0.816	0.818	0.817	0.853	0.853
97.09	0.452	0.449	0.456	0.561	0.655	0.560	0.561
99.04	0.481	0.480	0.485	0.585	0.622	0.586	0.585
107.08	0.825	0.826	0.861	0.862	0.863	0.870	0.871
121.09	0.658	0.658	0.693	0.694	0.694	0.712	0.712
137.12	0.195	0.195	0.198	0.246	0.261	0.617	0.617


Seven component MCR-ALS model (selected m/z)
Component 1 of 7
Component 2 of 7
Component 3 of 7
Component 4 of 7
Component 5 of 7
Component 6 of 7
Signal Counts
Time of Day on August 7, 2018
m/z
CH4OH+
C3H5+
C4H8H+
C6H6H+
C7H8H+
C8H10H+
C9H12H+
CH3CO+
C3H6OH+
C2H4OH+
C2H4O2H+
12:00
18:00
00:00
30
60
90
120
150

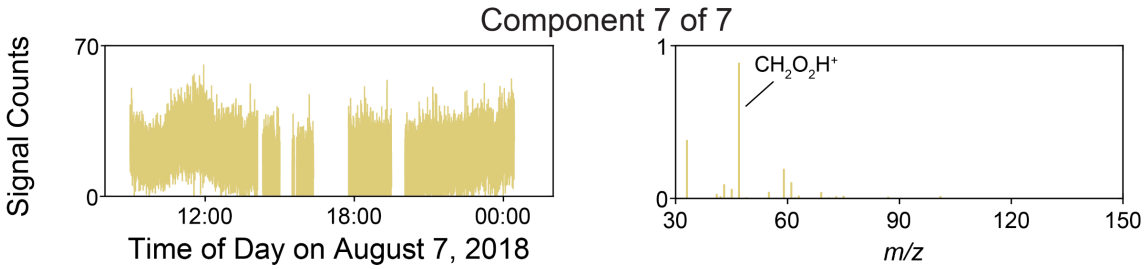
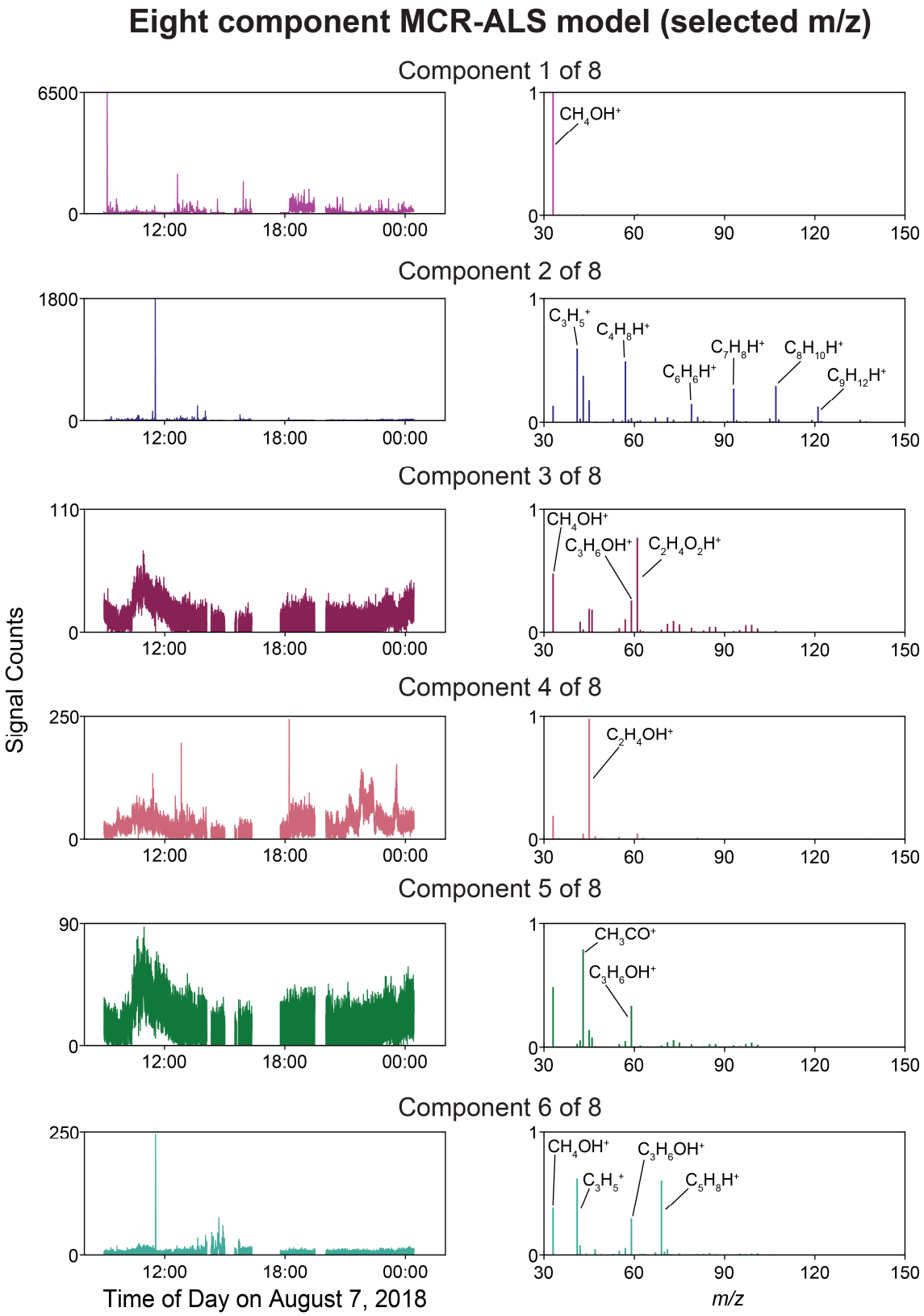


Figure C.7: Seven component MCR-ALS model using variable selection.


Eight component MCR-ALS model (selected m/z)
Component 1 of 8
Component 2 of 8
Component 3 of 8
Component 4 of 8
Component 5 of 8
Component 6 of 8
Signal Counts
Time of Day on August 7, 2018
m/z
CH4OH+
C3H5+
C4H8H+
C6H6H+
C7H8H+
C8H10H+
C9H12H+
C3H6OH+
C2H4O2H+
C2H4OH+
CH3CO+
C5H8H+
12:00
18:00
00:00

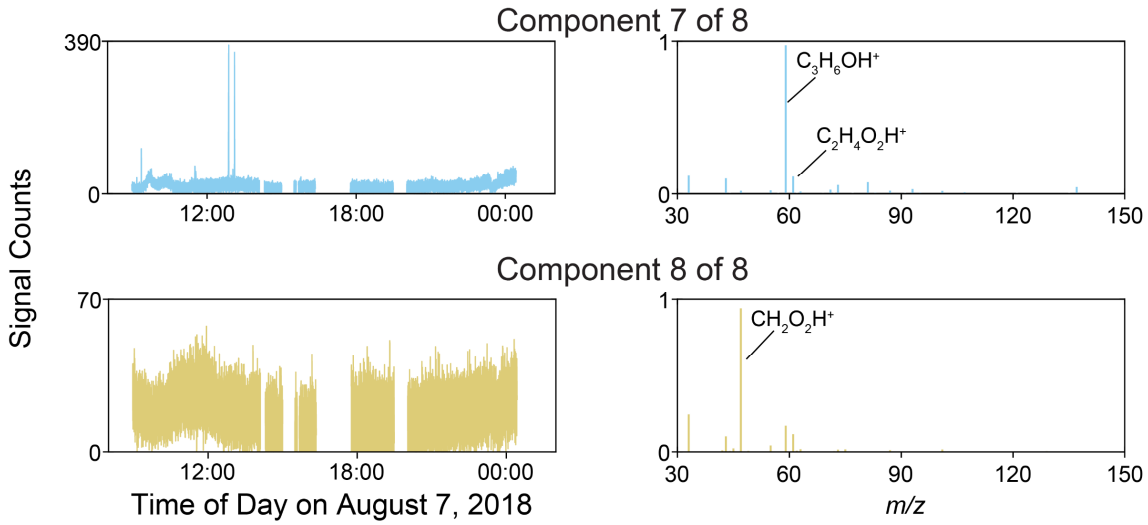


Figure C.8: Eight component MCR-ALS model using variable selection.

Determining the number of components in the August 14, 2018 MCR-WALS model

The Meisch coefficients for the major ions for the seven to thirteen component models are shown in Table C.3 with values greater than or equal to 0.9 being bolded. As the number of components is increased above seven or eight components, the modeling for most *m/z* does not improve significantly, with the exception *m/z* 107.08 and 137.12. For the eight component model, scatter plots of the measured versus modelled signal intensity for nine of the measured *m/z* are shown in Figure C.9, along with the corresponding Meisch coefficients for the presented data. For four of the *m/z* shown (*m/z* 45.027, 61.028, 63.022, and 69.067), the eight component model describes the measured data well, while the data for the other five *m/z* show some scatter. In particular, the scatter plots for *m/z* 59.049 and 137.12, (and to a lesser extent *m/z* 79.052 and 93.065), suggest that more components are needed to explain all the variation in the measured signal intensity. However, the addition of more components to the model (for example in a ten component model) does not produce more meaningful components. Rather than describing new VOC sources, additional components often resolved individual chemical species, such as acetone. The time series and mass spectra of the seven–thirteen component MCR-WALS models were evaluated, with the nine component model excluded from further evaluation as it did not model the contributions from the sulphur compounds (*m/z* 49.01, 63.02, and 95.00). An eight component model was chosen for interpretation.

Table C.3: Meisch coefficients for the main *m/z* measured in the August 14, 2018 field campaign for 7–13 component MCR-WALS models. Values greater than 0.9 are bolded for emphasis. The model chosen for interpretation is shaded.

	Number of MCR-WALS components						
m/z	**7**	**8**	**9**	**10**	**11**	**12**	**13**
33.03	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
41.03	**0.997**	**0.997**	**0.998**	**0.997**	**0.998**	**0.998**	**0.998**
42.03	0.792	0.820	0.813	0.805	0.859	0.860	0.864
43.01	**0.996**	**0.996**	**0.996**	**0.999**	**0.999**	**0.999**	**1.000**
43.05	**0.996**	**0.996**	**0.992**	**0.993**	**0.996**	**0.996**	**0.996**
45.03	**0.996**	**0.997**	**0.995**	**0.993**	**0.999**	**0.999**	**0.999**
47.04	0.835	0.834	0.848	0.857	0.844	0.842	0.842
49.01	**0.973**	**0.973**	**0.019**	**0.977**	**0.981**	**0.981**	**0.980**
57.07	**0.942**	**0.942**	**0.971**	**0.940**	**0.942**	**0.985**	**0.952**
59.05	0.784	0.809	**0.999**	**0.999**	**1.000**	**1.000**	**1.000**
61.03	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**0.999**	**1.000**
63.02	**0.971**	**0.971**	0.050	**0.978**	**0.985**	**0.985**	**0.986**
69.07	**0.941**	**0.940**	**0.951**	**0.947**	**0.939**	**0.942**	**0.940**
71.08	0.566	0.562	0.614	0.582	0.564	0.617	0.640
73.06	**0.943**	**0.942**	**0.963**	**0.952**	**0.938**	**0.953**	**0.950**
75.04	**0.978**	**0.978**	**0.978**	**0.990**	**0.989**	**0.989**	**0.989**
79.05	0.838	0.845	0.849	0.841	0.868	0.875	0.885
81.07	0.830	0.828	0.831	0.917	0.991	0.991	0.991
87.04	**0.947**	**0.947**	**0.963**	**0.955**	**0.947**	**0.957**	**0.953**
89.06	**0.929**	**0.930**	**0.930**	**0.989**	**0.991**	**0.991**	**0.993**
93.07	0.836	0.832	0.807	0.853	0.832	0.806	**0.904**
95.00	0.714	0.713	0.045	0.711	0.716	0.716	0.718
95.08	0.684	0.684	0.694	0.704	0.736	0.735	0.736
97.09	0.678	0.681	0.678	0.702	0.707	0.707	0.704
103.06	0.850	0.852	0.854	**0.930**	**0.933**	**0.933**	**0.942**
107.08	0.599	0.618	0.683	0.552	0.639	0.711	**0.999**
117.08	0.893	0.895	**0.901**	**0.899**	**0.897**	**0.904**	**0.903**
121.09	0.857	0.865	0.884	0.843	0.873	0.886	0.862
135.10	0.552	0.554	0.571	0.581	0.595	0.606	0.549
137.12	0.673	0.671	0.788	0.890	**0.984**	**0.982**	**0.983**

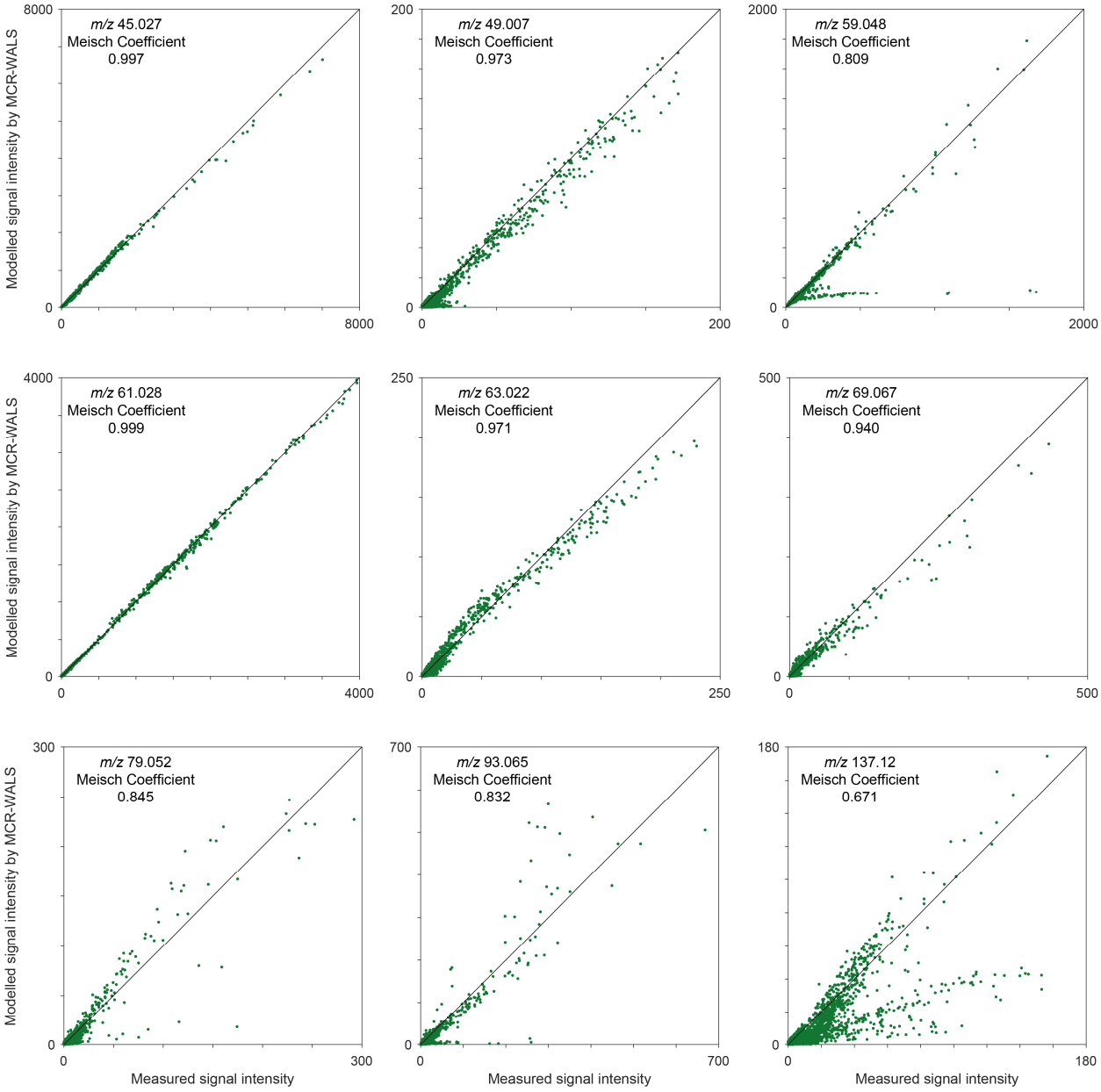


Figure C.9: Scatter plots of the measured versus modelled (with an eight component MCR-WALS model) signal intensity for nine of the recorded *m/z*, with the corresponding Meisch coefficient listed. The measured data for *m/z* 45.027, 61.028, 63.022, and 69.067 are described very well by the model, with little scatter around the y = x line. The plots for *m/z* 59.048 and *m/z* 137.12 suggest that additional components are needed to describe all the measured data at these *m/z*.

Determining the number of components in the August 21 and 22, 2017 MCR-WALS model

Fifty-eight *m/z* were included in the analysis, and the Meish coefficients for the major *m/z* observed on the day for these models are provided in Table C.4, with values over 0.9 bolded for emphasis. Scatter plots of the measured versus modelled signal intensity for nine of the observed *m/z*

are shown in Figure C.10. In this figure, the signal measured for *m/z* 45.03, 61.03, 63.02, and 137.12 are all modelled very well, with little scatter around the diagonal. There is more scatter present in the other plots, with the distribution of data present in the plots for *m/z* 49.01, 59.05, 69.07, and 79.05 suggesting that more components are required to describe all the sources of these compounds in the data. However, as with the data previously presented, additional components do not lead to more unique VOC sources being identified, but instead result in components that mainly describe one chemical compound, or that split other components in two. A nine component model was chosen for interpretation.

Table C.4: Meisch coefficients for seven to twelve component MCR-WALS models on August 21 and 22, 2017. Values greater than 0.9 are bolded for emphasis. The model chosen for interpretation is highlighted.

m/z	Number of MCR-WALS components					
	7	**8**	**9**	**10**	**11**	**12**
33.03	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
41.03	**0.992**	**0.993**	**0.992**	**0.994**	**0.994**	**0.994**
43.01	**0.931**	**0.924**	**0.944**	**0.941**	**0.972**	**0.972**
43.05	**0.988**	**0.988**	**0.988**	**0.987**	**0.988**	**0.988**
45.03	**0.999**	**0.999**	**0.999**	**0.999**	**1.00**	**1.00**
47.04	0.594	0.640	0.699	0.703	0.753	**0.998**
49.01	0.033	0.045	0.805	0.804	0.800	0.802
57.07	**0.982**	**0.981**	**0.983**	**0.989**	**0.990**	**0.990**
59.05	0.825	0.821	**0.922**	**0.957**	**0.957**	**0.956**
61.03	**0.990**	**0.988**	**0.992**	**0.992**	**0.993**	**0.993**
63.02	0.021	0.032	**0.998**	**0.998**	**0.999**	**0.999**
69.07	0.784	0.782	0.794	0.857	0.860	0.862
71.08	0.838	0.836	0.839	**0.962**	**0.958**	**0.959**
73.06	0.580	0.566	0.593	0.889	0.878	0.885
75.04	0.760	0.793	0.827	0.843	**0.927**	**0.924**
77.02	0.362	0.796	0.797	0.804	0.809	0.810
78.96	0.040	0.051	0.824	0.823	0.830	0.827
79.05	0.895	0.896	0.896	**0.912**	**0.909**	**0.909**
81.07	**0.986**	**0.990**	**0.990**	**0.990**	**0.990**	**0.996**
83.09	0.478	0.481	0.472	0.488	0.495	0.494
85.10	0.829	0.830	0.829	**0.918**	**0.914**	**0.915**
87.04	0.561	0.556	0.570	0.708	0.697	0.703
91.04	0.294	**0.950**	**0.961**	**0.961**	**0.979**	**0.980**
93.07	**0.975**	**0.976**	**0.975**	**0.991**	**0.990**	**0.991**
95.00	0.111	0.193	0.767	0.772	0.831	0.824
95.08	0.143	0.150	0.143	0.150	0.182	0.196
97.09	0.665	0.700	0.679	0.740	0.861	0.876
105.06	0.148	0.155	0.159	0.168	0.170	0.170
106.07	0.677	0.679	0.679	0.682	0.683	0.683

Table C.4: Meisch coefficients for seven to twelve component MCR-WALS models on August 21 and 22, 2017. Values greater than 0.9 are bolded for emphasis. The model chosen for interpretation is highlighted, continued.

m/z	Number of MCR-WALS components					
	7	**8**	**9**	**10**	**11**	**12**
107.08	0.766	0.767	0.767	0.765	0.767	0.766
119.08	0.850	0.851	0.852	0.893	0.892	0.891
121.09	**0.949**	**0.949**	**0.949**	**0.988**	**0.986**	**0.986**
133.07	0.714	0.713	0.719	0.806	0.800	0.800
135.10	0.683	0.698	0.687	0.754	0.827	0.841
137.12	**0.988**	**0.990**	**0.989**	**0.989**	**0.989**	**0.997**
149.11	0.362	0.360	0.366	0.468	0.462	0.462

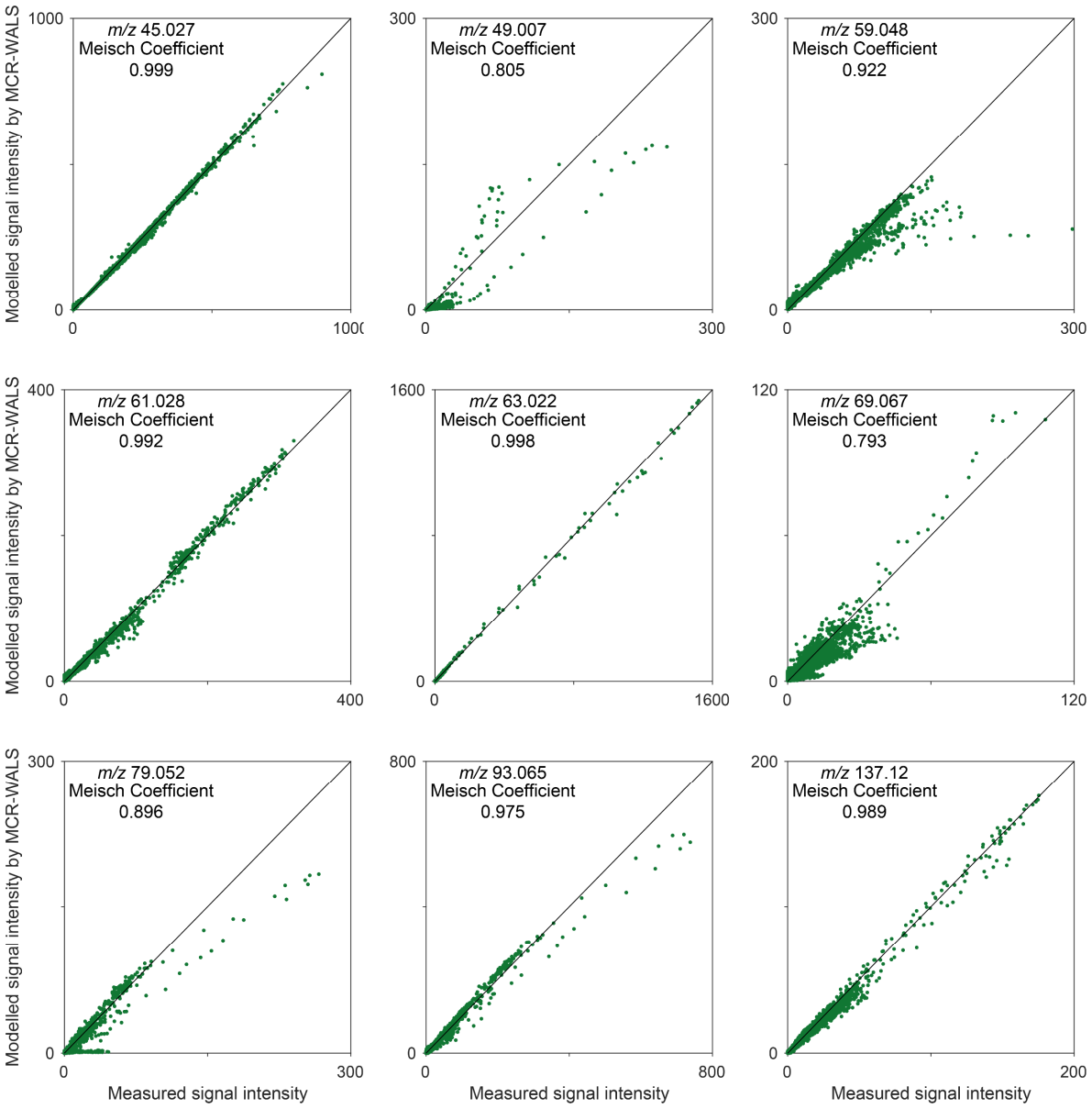


Figure C.10: Scatter plots of the measured versus modelled signal intensity for a selection of *m/z* recorded on August 21 and 22, 2017. The Meisch Coefficients for each plot are also presented. The scatter plots suggest that an additional component may be required to model the signal intensities measured at *m/z* 49.01, 59.05, 69.07, and 79.05.

Bibliography

1. European Environment Agency (EEA), *Air quality in Europe - 2020 report*, EEA Report No 09/2020, Publications Office of the European Union, Luxembourg, 2020.
2. A. Prüss-Ustün, J. Wolf, C. Corvalan, R. Bos and M. Neiraa, *Preventing disease through healthy environments: A global assessment of the environmental burden of disease*, World Health Organization, Geneva, 2016.
3. I. Manisalidis, E. Stavropoulou, A. Stavropoulos and E. Bezirtzoglou, Environmental and Health Impacts of Air Pollution: A Review, *Front. Public Heal.*, 2020, **8**, 14.
4. R. Ma, J. Ban, Q. Wang and T. Li, Statistical spatial-temporal modeling of ambient ozone exposure for environmental epidemiology studies: A review, *Sci. Total Environ.*, 2020, **701**, 134463.
5. F. J. Kelly, Oxidative stress: Its role in air pollution and adverse health effects, *Occup. Environ. Med.*, 2003, **60**, 612–616.
6. Y. Zhou, C. Li, M. A. J. Huijbregts and M. M. Mumtaz, Carcinogenic air toxics exposure and their cancer-related health impacts in the United States, *PLoS One*, 2015, **10**, 1–15.
7. P. Sicard, Y. O. Khaniabadi, S. Perez, M. Gualtieri and A. De Marco, Effect of O_3, PM_{10} and $PM_{2.5}$ on cardiovascular and respiratory diseases in cities of France, Iran and Italy, *Environ. Sci. Pollut. Res*., 2019, **26**, 32645–32665.
8. P. J. Landrigan, R. Fuller, N. J. R. Acosta, O. Adeyi, R. Arnold, N. (Nil) Basu, A. B. Baldé, R. Bertollini, S. Bose-O'Reilly, J. I. Boufford, P. N. Breysse, T. Chiles, C. Mahidol, A. M. Coll-Seck, M. L. Cropper, J. Fobil, V. Fuster, M. Greenstone, A. Haines, D. Hanrahan, D. Hunter, M. Khare, A. Krupnick, B. Lanphear, B. Lohani, K. Martin, K. V. Mathiasen, M. A. McTeer, C. J. L. Murray, J. D. Ndahimananjara, F. Perera, J. Potočnik, A. S. Preker, J. Ramesh, J. Rockström, C. Salinas, L. D. Samson, K. Sandilya, P. D. Sly, K. R. Smith, A. Steiner, R. B. Stewart, W. A. Suk, O. C. P. van Schayck, G. N. Yadama, K. Yumkella and M. Zhong, The Lancet Commission on pollution and health, *Lancet*, 2018, **391**, 462–512.
9. World Health Organization (WHO), Ambient Air Pollution, https://www.who.int/airpollution/ambient/health-impacts/en/, (accessed 16 October 2020).
10. Health Canada, *Health impacts of air pollution in Canada: estimates of morbidity and premature mortality outcomes, 2019 report*., H144-51/2019E-PDF, Health Canada, Ottawa, 2019.
11. P. S. Monks, A. T. Archibald, A. Colette, O. Cooper, M. Coyle, R. Derwent, D. Fowler, C. Granier, K. S. Law, G. E. Mills, D. S. Stevenson, O. Tarasova, V. Thouret, E. Von Schneidemesser, R. Sommariva, O. Wild and M. L. Williams, Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer, *Atmos. Chem. Phys.*, 2015, **15**, 8889–8973.
12. S. Zuhara and R. Isaifan, The Impact of Criteria Air Pollutants on Soil and Water: A Review, J. *Environ. Sci. Pollut. Res.*, 2018, **4**, 278–284.
13. R. J. Payne, N. B. Dise, C. J. Stevens, D. J. Gowing and B. Partners, Impact of nitrogen deposition at the species level, *PNAS*., 2012, **110**, 984–987.
14. M. Vellend, L. Baeten, A. Becker-scarpitta, J. L. Mccune, J. Messier, I. H. Myers-smith and D. F. Sax, Plant Biodiversity Change Across Scales During the Anthropocene, *Annu. Rev. Plant Biol.*, 2017, **68**, 563–584.
15. OECD, The Economic Consequences of Outdoor Air Pollution, OECD Publishing, Paris, 2016.

16. E. Galarneau, D. Wang, E. Dabek-Zlotorzynska, M. Siu, V. Celo, M. Tardif, D. Harnish and Y. Jiang, Air toxics in Canada measured by the National Air Pollution Surveillance (NAPS) program and their relation to ambient air quality guidelines, *J. Air Waste Manag. Assoc.*, 2016, **66**, 184–200.
17. J. H. Seinfeld and S. N. Pandis, *Atmospheric chemistry and physics: from air pollution to climate change*, John Wiley & Sons, Hoboken, NJ, 3rd edn., 2016.
18. Environment and Climate Change Canada, Canadian Environmental Sustainability Indicators: Air Pollutant Emissions, www.canada.ca/en/environment-climate-change/services/environmental-indicators/air-pollutant- emissions.html, (accessed 8 May 2021).
19. D. M. Stieb, R. T. Burnett, M. Smith-Doiron, O. Brion, H. S. Hwashin and V. Economou, A new multipollutant, no-threshold air quality health index based on short-term associations observed in daily time-series analyses, *J. Air Waste Manag. Assoc.*, 2008, **58**, 435–450.
20. E. Samoli, R. Peng, T. Ramsay, M. Pipikou, G. Touloumi, F. Dominici, R. Burnett, A. Cohen, D. Krewski, J. Samet and K. Katsouyanni, Acute effects of ambient particulate matter on mortality in Europe and North America: Results from the APHENA study, *Environ. Health Perspect.*, 2008, **116**, 1480–1486.
21. R. D. Peng, E. Samoli, L. Pham, F. Dominici, G. Touloumi, T. Ramsay, R. T. Burnett, D. Krewski, A. Le Tertre, A. Cohen, R. W. Atkinson, H. R. Anderson, K. Katsouyanni and J. M. Samet, Acute effects of ambient ozone on mortality in Europe and North America: Results from the APHENA study, *Air Qual. Atmos. Heal.*, 2013, **6**, 445–453.
22. Y. Hao, X. Meng, X. Yu, M. Lei, W. Li, W. Yang, F. Shi and S. Xie, Quantification of primary and secondary sources to PM2.5 using an improved source regional apportionment method in an industrial city, China, *Sci. Total Environ.*, 2020, **706**, 135715.
23. J. M. Kelly, R. M. Doherty, F. M. O'Connor and G. W. Mann, The impact of biogenic, anthropogenic, and biomass burning volatile organic compound emissions on regional and seasonal variations in secondary organic aerosol, *Atmos. Chem. Phys.*, 2018, **18**, 7393–7422.
24. J. Williams and R. Koppmann, in *Volatile Organic Compounds in the Atmosphere*, ed. R. Koppmann, John Wiley & Sons, Incorporated, Ames, IA, 1st edn., 2007, ch. 1, pp. 1–32.
25. S. E. Alexeeff, A. Roy, J. Shan, X. Liu, K. Messier, J. S. Apte, C. Portier, S. Sidney and S. K. Van Den Eeden, High-resolution mapping of traffic related air pollution with Google street view cars and incidence of cardiovascular events within neighborhoods in Oakland, CA, *Environ. Heal.*, 2018, **17**, 1–13.
26. N. G. Davey, R. J. Bell, C. G. Gill and E. T. Krogh, Mapping the geospatial distribution of atmospheric BTEX compounds using portable mass spectrometry and adaptive whole air sampling, *Atmos. Pollut. Res.*, 2020, **11**, 545–553.
27. M. W. Tessum, T. Larson, T. R. Gould, C. D. Simpson, M. G. Yost and S. Vedal, Mobile and Fixed-Site Measurements to Identify Spatial Distributions of Traffic-Related Pollution Sources in Los Angeles, *Environ. Sci. Technol.*, 2018, **52**, 2844–2853.
28. R. Atkinson and J. Arey, Atmospheric Degradation of Volatile Organic Compounds, *Chem. Rev.*, 2003, **103**, 4605–4638.
29. H. Z. Li, P. Gu, Q. Ye, N. Zimmerman, E. S. Robinson, R. Subramanian, J. S. Apte, A. L. Robinson and A. A. Presto, Spatially dense air pollutant sampling: Implications of spatial variability on the representativeness of stationary air pollutant monitors, *Atmos. Environ. X*, 2019, **2**, 100012.

30. Environment and Climate Change Canada, National Air Pollution Surveillance (NAPS) Program, https://www.canada.ca/en/environment-climate-change/services/air-pollution/monitoring-networks-data/national-air-pollution-program.html, (accessed 2 October 2020).
31. U.S. Environmental Protection Agency (EPA), Criteria Air Pollutants, https://www.epa.gov/criteria-air-pollutants, (accessed 2 October 2020).
32. M. Strum and R. Scheffe, National review of ambient air toxics observations, *J. Air Waste Manag. Assoc.*, 2016, **66**, 120–133.
33. Sonoma Technology Inc, *A Synthesis of Completed Community-Scale Air Toxics Ambient Monitoring Projects*, STI-910313-5469-FR, Sonoma Technology Inc, Petataluma, CA, 2013.
34. K. Sekimoto and A. R. Koss, Modern mass spectrometry in atmospheric sciences: Measurement of volatile organic compounds in the troposphere using proton-transfer-reaction mass spectrometry, *J. Mass Spectrom.*, 2020, **56**, 1–17.
35. K. Badjagbo, S. Sauvé and S. Moore, Real-time continuous monitoring methods for airborne VOCs, *TrAC - Trends Anal. Chem.*, 2007, **26**, 931–940.
36. L. K. Sahu, Volatile organic compounds and their measurements in the troposphere, *Curr. Sci.*, 2012, **102**, 1645–1649.
37. B. Yuan, A. R. Koss, C. Warneke, M. Coggon, K. Sekimoto and J. A. de Gouw, Proton-Transfer-Reaction Mass Spectrometry: Applications in Atmospheric Sciences, *Chem. Rev.*, 2017, **117**, 13187–13229.
38. Environment and Climate Change Canada, Volatile Organic Compound Emissions, https://www.ec.gc.ca/indicateurs-indicators/default.asp?lang=en&n=64B9E95D-1, (accessed 23 February 2017).
39. A. Kansal, Sources and reactivity of NMHCs and VOCs in the atmosphere: A review, J. Hazard. Mater., 2009, 166, 17–26.
40. D. Helmig and J. Bottenheim, Volatile organic compounds in the global atmosphere, *Eos*, 2009, **90**, 513–520.
41. R. G. Derwent, in *Volatile Organic Compounds in the Atmosphere*, eds. R. M. Harrison and R. E. Hester, Royal Society of Chemistry, Cambridge, UK, 1995, ch. 1, pp. 1–16.
42. G. Huang, R. Brook, M. Crippa, G. Janssens-Maenhout, C. Schieberle, C. Dore, D. Guizzardi, M. Muntean, E. Schaaf and R. Friedrich, Speciation of anthropogenic emissions of non-methane volatile organic compounds: a global gridded data set for 1970 - 2012, *Atmos. Chem. Phys.*, 2017, **17**, 7683–7701.
43. H. Zheng, S. Kong, N. Chen, Z. Niu, Y. Zhang, S. Jiang, Y. Yan and S. Qi, Source apportionment of volatile organic compounds: Implications to reactivity, ozone formation, and secondary organic aerosol potential, *Atmos. Res.*, 2021, **249**, 105344.
44. Health Canada, Residential indoor air quality guideline: acetaldehyde, https://www.canada.ca/en/health-canada/services/publications/healthy-living/residential-indoor-air-quality-guideline-acetaldehyde.html, (accessed 8 January 2020).
45. U.S. Environmental Protection Agency (EPA), Initial List of Hazardous Air Pollutants with Modifications, https://www.epa.gov/haps/initial-list-hazardous-air-pollutants-modifications, (accessed 8 January 2020).
46. SPECIATE 5.1, https://www.epa.gov/air-emissions-modeling/speciate, (accessed 10 March 2021).

47. D. J. Jacob, B. D. Field, E. M. Jin, I. Bey, Q. Li, J. A. Logan, R. M. Yantosca and H. B. Singh, Atmospheric budget of acetone, *J. Geophys. Res. Atmos.*, 2002, **107**, ACH 5-1-ACH 5-17.
48. K. Sindelarova, C. Granier, I. Bouarar, A. Guenther, S. Tilmes, T. Stavrakou, J. F. Müller, U. Kuhn, P. Stefani and W. Knorr, Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years, *Atmos. Chem. Phys.*, 2014, **14**, 9317–9341.
49. Ontario Ministry of Environment, Conservation and Parks (MECP), *Air Quality in Ontario 2017 Report*, MECP, Toronto, 2019.
50. I. Simpson and C. Volosciuk, Changing Volatile Organic Compound Emissions in Urban Environments: Many Paths to Cleaner Air, https://public.wmo.int/en/resources/bulletin/changing-volatile-organic-compound-emissions-urban-environments-many-paths, (accessed 8 January 2020).
51. D. A. Kaden, C. Mandin, G. D. Nielsen and P. Wolkoff, in *World Health Organization (WHO) Guidelines for Indoor Air Quality: Selected Pollutants*, WHO Regional Office for Europe, Copenhagen, Denmark, 2010, ch. 3, pp. 103–156.
52. C. Jia and S. Batterman, A critical review of naphthalene sources and exposures relevant to indoor and outdoor air, *Int. J. Environ. Res. Public Health*, 2010, **7**, 2903–2939.
53. A. T. Simpson, M. A. Hemingway and C. Seymour, Dangerous (toxic) atmospheres in UK wood pellet and wood chip fuel storage, *J. Occup. Environ. Hyg.*, 2016, **13**, 699–707.
54. U.S. Environmental Protection Agency (EPA), Ambient Concentrations of Selected Air Toxics, https://cfpub.epa.gov/roe/indicator.cfm?i=90, (accessed 5 October 2020).
55. World Health Organization (WHO), Exposure To Benzene: A Major Public Health Concern, WHO Public Health and Environment, Geneva, 2010.
56. P. Gostelow, S. A. Parsons and R. M. Stuetz, Odour measurements for sewage treatment works, *Water Res.*, 2001, **35**, 579–597.
57. D. González, J. Colón, A. Sánchez and D. Gabriel, A systematic study on the VOCs characterization and odour emissions in a full-scale sewage sludge composting plant, *J. Hazard. Mater.*, 2019, **373**, 733–740.
58. S. J. O'Doherty and L. J. Carpenter, in *Volatile Organic Compounds in the Atmosphere*, ed. R. Koppmann, John Wiley & Sons, Incorporated, Ames, IA, 1st edn., 2007, ch. 5, pp. 174–220.
59. G. Rao and E. P. Vejerano, Partitioning of volatile organic compounds to aerosols: A review, *Chemosphere*, 2018, **212**, 282–296.
60. U.S. Environmental Protection Agency (EPA), Volatile Organic Compounds' Impact on Indoor Air Quality, https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality, (accessed 23 February 2017).
61. C. Conti, M. Guarino and J. Bacenetti, Measurements techniques and models to assess odor annoyance: A review, *Environ. Int.*, 2020, **134**, 105261.
62. M. Aatamila, P. K. Verkasalo, M. J. Korhonen, A. L. Suominen, M. R. Hirvonen, M. K. Viluksela and A. Nevalainen, Odour annoyance and physical symptoms among residents living near waste treatment centres, *Environ. Res.*, 2011, **111**, 164–170.
63. T.F. Stocker, D. Qin, G.-K. Plattner, L.V. Alexander, S.K. Allen, N.L. Bindoff, F.-M. Bréon, J.A. Church, U. Cubasch, S. Emori, P. Forster, P. Friedlingstein, N. Gillett, J.M. Gregory, D.L. Hartmann, E. Jansen, B. Kirtman, R. Knutti, K. Krishna Kumar, P. Lemke, J. Marotzke, V. Masson-Delmotte, G.A. Meehl, I.I. Mokhov, S. Piao, V. Ramaswamy, D. Randall, M. Rhein, M. Rojas, C. Sabine, D. Shindell, L.D. Talley, D.G. Vaughan and S.-P. Xie, in *Climate Change 2013: The Physical*

Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, eds. T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013, Technical Summary, pp, 33-115.

64. M. Shrivastava, C. D. Cappa, J. Fan, A. H. Goldstein, A. B. Guenther, J. L. Jimenez, C. Kuang, A. Laskin, S. T. Martin, N. L. Ng, T. Petaja, J. R. Pierce, P. J. Rasch, P. Roldin, J. H. Seinfeld, J. Shilling, J. N. Smith, J. A. Thornton, R. Volkamer, J. Wang, D. R. Worsnop, R. A. Zaveri, A. Zelenyuk and Q. Zhang, Recent advances in understanding secondary organic aerosol: Implications for global climate forcing, *Rev. Geophys.*, 2017, **55**, 509–559.
65. M. Crippa, D. Guizzardi, M. Muntean, E. Schaaf, F. Dentener, J. A. Van Aardenne, S. Monni, U. Doering, J. G. J. Olivier, V. Pagliari and G. Janssens-Maenhout, Gridded emissions of air pollutants for the period 1970-2012 within EDGAR v4.3.2, *Earth Syst. Sci. Data*, 2018, **10**, 1987–2013.
66. Y. Zhu, L. Yang, J. Chen, X. Wang, L. Xue, X. Sui, L. Wen, C. Xu, L. Yao, J. Zhang, M. Shao, S. Lu and W. Wang, Characteristics of ambient volatile organic compounds and the influence of biomass burning at a rural site in Northern China during summer 2013, *Atmos. Environ.*, 2016, **124**, 156–165.
67. P. J. Ziemann and R. Atkinson, Kinetics, products, and mechanisms of secondary organic aerosol formation, *Chem. Soc. Rev.*, 2012, **41**, 6582–605.
68. 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, The formation, properties and impact of secondary organic aerosol: current and emerging issues, *Atmos. Chem. Phys.*, 2009, **9**, 5155–5236.
69. J. Xu, W. Hu, D. Liang and P. Gao, Photochemical impacts on the toxicity of $PM_{2.5}$, *Crit. Rev. Environ. Sci. Technol.*, 2020, DOI: 10.1080/10643389.2020.1816126.
70. B. J. Finlayson-Pitts and J. N. Pitts, Atmospheric chemistry of tropospheric ozone formation: Scientific and regulatory implications, *Air Waste*, 1993, **43**, 1091–1100.
71. P. Sicard, E. Paoletti, E. Agathokleous, V. Araminiene, C. Proietti, F. Coulibaly and A. De Marco, Ozone weekend effect in cities: Deep insights for urban air pollution control, *Environ. Res.*, 2020, **191**, 110193.
72. Environment and Climate Change Canada, Canadian Environmental Sustainability Indicators: Air Quality, www.canada.ca/en/environment-climate-change/services/environmental-indicators/air-quality.html, (accessed 8 May 2021).
73. National Research Council, *Rethinking the Ozone Problem in Urban and Regional Air Pollution*, The National Academic Press, Washington, DC, 1991.
74. X. Zhang, Y. Yin, J. Wen, S. Huang, D. Han, X. Chen and J. Cheng, Characteristics, reactivity and source apportionment of ambient volatile organic compounds (VOCs) in a typical tourist city, *Atmos. Environ.*, 2019, **215**, 116898.
75. M. C. McCarthy, Y.-A. Aklilu, S. G. Brown and D. A. Lyder, Source apportionment of volatile organic compounds measured in Edmonton, Alberta, *Atmos. Environ.*, 2013, **81**, 504–516.

76. Y. Xiong, M. A. Bari, Z. Xing and K. Du, Ambient volatile organic compounds (VOCs) in two coastal cities in western Canada: Spatiotemporal variation, source apportionment, and health risk assessment, *Sci. Total Environ.*, 2020, **706**, 135970.
77. J. Zheng, J. P. Garzón, M. E. Huertas, R. Zhang, M. Levy, Y. Ma, J. I. Huertas, R. T. Jardón, L. G. Ruíz, H. Tan and L. T. Molina, Volatile organic compounds in Tijuana during the cal-mex 2010 campaign: Measurements and source apportionment, *Atmos. Environ.*, 2013, **70**, 521–531.
78. A. Stojić, S. Stanišić Stojić, Z. Mijić, A. Šoštarić and S. Rajšić, Spatio-temporal distribution of VOC emissions in urban area based on receptor modeling, *Atmos. Environ.*, 2015, **106**, 71–79.
79. H. Li, M. R. Canagaratna, M. Riva, P. Rantala, Y. Zhang, S. Thomas, L. Heikkinen, P. M. Flaud, E. Villenave, E. Perraudin, D. Worsnop, M. Kulmala, M. Ehn and F. Bianchi, Atmospheric organic vapors in two European pine forests measured by a Vocus PTR-TOF: Insights into monoterpene and sesquiterpene oxidation processes, *Atmos. Chem. Phys.*, 2021, **21**, 4123–4147.
80. P. M. Mach, E. M. Mcbride, Z. J. Sasiene, K. R. Brigance, S. K. Kennard, K. C. Wright and G. F. Verbeck, Vehicle-Mounted Portable Mass Spectrometry System for the Covert Detection via Spatial Analysis of Clandestine Methamphetamine Laboratories, *Anal. Chem*., 2015, **87**, 11501–11508.
81. Z. L. Hildenbrand, P. M. Mach, E. M. McBride, M. N. Dorreyatim, J. T. Taylor, D. D. Carlton, J. M. Meik, B. E. Fontenot, K. C. Wright, K. A. Schug and G. F. Verbeck, Point source attribution of ambient contamination events near unconventional oil and gas development, *Sci. Total Environ.*, 2016, **573**, 382–388.
82. R. J. Bell, N. G. Davey, M. Martinsen, C. Collin-Hansen, E. T. Krogh and C. G. Gill, A field-portable membrane introduction mass spectrometer for real-time quantitation and spatial mapping of atmospheric and aqueous contaminants, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 212–223.
83. R. M. Healy, Q. F. Chen, J. Bennett and N. S. Karellas, A multi-year study of VOC emissions at a chemical waste disposal facility using mobile APCI-MS and LPCI-MS instruments, *Environ. Pollut.*, 2018, **232**, 220–228.
84. M. Zavala, S. C. Herndon, E. C. Wood, J. T. Jayne, D. D. Nelson, A. M. Trimborn, E. Dunlea, W. B. Knighton, A. Mendoza, D. T. Allen, C. E. Kolb, M. J. Molina and L. T. Molina, Comparison of emissions from on-road sources using a mobile laboratory under various driving and operational sampling modes, *Atmos. Chem. Phys.*, 2009, **9**, 1–14.
85. M. Zavala, S. C. Herndon, R. S. Slott, E. J. Dunlea, L. C. Marr, J. H. Shorter, M. Zahniser, W. B. Knighton, T. M. Rogers, C. E. Kolb, L. T. Molina and M. J. Molina, Characterization of on-road vehicle emissions in the Mexico City Metropolitan Area using a mobile laboratory in chase and fleet average measurement modes during the MCMA-2003 field campaign, *Atmos. Chem. Phys.*, 2006, **6**, 5129–5142.
86. U.S. Environmental Protection Agency (EPA), *Compendium of Methods for the Determination of Toxic Organic Compounds in Ambient Air, 2nd. Edition*, EPA/625/R-96/010b, U.S. EPA Office of Research and Development, CIncinnati, Ohio, 1999.
87. L. Shen, Z. Wang, H. Cheng, S. Liang, P. Xiang, K. Hu, T. Yin and J. Yu, A Spatial-Temporal Resolved Validation of Source Apportionment by Measurements of Ambient VOCs in Central China, *Int. J. Environ. Res. Public Health*, 2020, **17**, 791.
88. B. T. Ly, Y. Kajii, T. Y. L. Nguyen, K. Shoji, D. A. Van, T. N. N. Do, T. D. Nghiem and Y. Sakamoto, Characteristics of roadside volatile organic compounds in an urban area dominated by gasoline vehicles, a case study in Hanoi, *Chemosphere*, 2020, **254**, 126749.

89. Q. Tan, H. Liu, S. Xie, L. Zhou, T. Song, G. Shi, W. Jiang, F. Yang and F. Wei, Temporal and spatial distribution characteristics and source origins of volatile organic compounds in a megacity of Sichuan Basin, China, *Environ. Res.*, 2020, **185**, 109478.
90. Y. Liu, Q. Xie, X. Li, F. Tian, X. Qiao, J. Chen and W. Ding, Profile and source apportionment of volatile organic compounds from a complex industrial park, *Environ. Sci. Process. Impacts*, 2019, **21**, 9–18.
91. J. An, X. Su, Y. Zhang and B. Zhu, Characteristics of Volatile Organic Compounds in Nanjing and Suzhou, Two Urban Sites in the Yangtze River Delta, China, *Arch. Environ. Contam. Toxicol.*, 2020, **78**, 416–429.
92. B. T. P. Thera, P. Dominutti, F. Öztürk, T. Salameh, S. Sauvage, C. Afif, B. Çetin, C. Gaimoz, M. Keleş, S. Evan and A. Borbon, Composition and variability of gaseous organic pollution in the port megacity of Istanbul: source attribution, emission ratios and inventory evaluation, *Atmos. Chem. Phys.*, 2019, **19**, 15131–15156.
93. R. Wu, Y. Zhao, J. Zhang and L. Zhang, Variability and sources of ambient volatile organic compounds based on online measurements in a suburban region of nanjing, eastern China, *Aerosol Air Qual. Res.*, 2020, **20**, 606–619.
94. A. A. Presto, T. R. Dallmann, P. Gu and U. Rao, BTEX exposures in an area impacted by industrial and mobile sources: Source attribution and impact of averaging time, *J. Air Waste Manag. Assoc.*, 2016, **66**, 387–401.
95. J. An, J. Wang, Y. Zhang and B. Zhu, Source Apportionment of Volatile Organic Compounds in an Urban Environment at the Yangtze River Delta, China, *Arch. Environ. Contam. Toxicol.*, 2017, **72**, 335–348.
96. X. Sun, H. Wang, Z. Guo, P. Lu, F. Song, L. Liu, J. Liu, N. L. Rose and F. Wang, Positive matrix factorization on source apportionment for typical pollutants in different environmental media: A review, *Environ. Sci. Process. Impacts*, 2020, **22**, 239–255.
97. X. F. Huang, B. Zhang, S. Y. Xia, Y. Han, C. Wang, G. H. Yu and N. Feng, Sources of oxygenated volatile organic compounds (OVOCs) in urban atmospheres in North and South China, *Environ. Pollut.*, 2020, **261**, 114152.
98. X. F. Huang, C. Wang, B. Zhu, L. L. Lin and L. Y. He, Exploration of sources of OVOCs in various atmospheres in southern China, *Environ. Pollut.*, 2019, **249**, 831–842.
99. L. Kaser, T. Karl, A. Guenther, M. Graus, R. Schnitzhofer, A. Turnipseed, L. Fischer, P. Harley, M. Madronich, D. Gochis, F. N. Keutsch and A. Hansel, Undisturbed and disturbed above canopy ponderosa pine emissions: PTR-TOF-MS measurements and MEGAN 2.1 model results, *Atmos. Chem. Phys.*, 2013, **13**, 11935–11947.
100. P. A. Solomon, D. Vallano, M. Lunden, B. Lafranchi, C. L. Blanchard and S. L. Shaw, Mobile-platform measurement of air pollutant concentrations in California: Performance assessment, statistical methods for evaluating spatial variations, and spatial representativeness, *Atmos. Meas. Tech.*, 2020, **13**, 3277–3301.
101. Y. Tan, E. M. Lipsky, R. Saleh, A. L. Robinson and A. A. Presto, Characterizing the spatial variation of air pollutants and the contributions of high emitting vehicles in Pittsburgh, PA, *Environ. Sci. Technol.*, 2014, **48**, 14186–14194.
102. R. H. de Oliveira, C. de C. Carneiro, F. G. V. de Almeida, B. M. de Oliveira, E. H. M. Nunes and A. S. dos Santos, Multivariate air pollution classification in urban areas using mobile sensors and self-organizing maps, *Int. J. Environ. Sci. Technol.*, 2018, **16**, 5475–5488.

103. D. T. Snyder, C. J. Pulliam, Z. Ouyang and R. G. Cooks, Miniature and Fieldable Mass Spectrometers: Recent Advances, *Anal. Chem*., 2016, **88**, 2–29.
104. N. G. Davey, C. T. E. Fitzpatrick, J. M. Etzkorn, M. Martinsen, R. S. Crampton, G. D. Onstad, T. V. Larson, M. G. Yost, E. T. Krogh, M. Gilroy, K. H. Himes, E. T. Saganić, C. D. Simpson and C. G. Gill, Measurement of spatial and temporal variation in volatile hazardous air pollutants in Tacoma, Washington, using a mobile membrane introduction mass spectrometry (MIMS) system, J. Environ. Sci. Health A, 2014, **49**, 1199–1208.
105. N. G. Davey, R. J. Bell, E. T. Krogh and C. G. Gill, A membrane introduction mass spectrometer utilizing ion-molecule reactions for the on-line speciation and quantitation of volatile organic molecules, *Rapid Commun. Mass Spectrom.*, 2015, **29**, 2187–2194.
106. Q. Liang, X. Bao, Q. Sun, Q. Zhang, X. Zou, C. Huang, C. Shen and Y. Chu, Imaging VOC distribution in cities and tracing VOC emission sources with a novel mobile proton transfer reaction mass spectrometer, *Environ. Pollut.*, 2020, **265**, 114628.
107. P. Gu, T. R. Dallmann, H. Z. Li, Y. Tan and A. A. Presto, Quantifying urban spatial variations of anthropogenic VOC concentrations and source contributions with a mobile sampling platform, *Int. J. Environ. Res. Public Health*, 2019, **16**, 1632.
108. T. M. Rogers, E. P. Grimsrud, S. C. Herndon, J. T. Jayne, C. E. Kolb, E. Allwine, H. Westberg, B. K. Lamb, M. Zavala, L. T. Molina, M. J. Molina and W. B. Knighton, On-road measurements of volatile organic compounds in the Mexico City metropolitan area using proton transfer reaction mass spectrometry, *Int. J. Mass Spectrom.*, 2006, **252**, 26–37.
109. R. J. Bell, R. T. Short and R. H. Byrne, In situ determination of total dissolved inorganic carbon by underwater membrane introduction mass spectrometry, *Limnol. Oceanogr. Methods*, 2011, **9**, 164–175.
110. C. Janfelt, F. R. Lauritsen, S. K. Toler, R. J. Bell and R. T. Short, Method for quantification of chemicals in a pollution plume using a moving membrane-based sensor exemplified by mass spectrometry, *Anal. Chem*., 2007, **79**, 5336–5342.
111. P. K. Misztal, T. Karl, R. Weber, H. H. Jonsson, A. B. Guenther and A. H. Goldstein, Airborne flux measurements of biogenic isoprene over California, *Atmos. Chem. Phys.*, 2014, **14**, 10631–10647.
112. T. Karl, P. K. Misztal, H. H. Jonsson, S. Shertz, A. H. Goldstein and A. B. Guenther, Airborne flux measurements of bvocs above californian oak forests: Experimental investigation of surface and entrainment fluxes, OH densities, and damköhler numbers, *J. Atmos. Sci.*, 2013, **70**, 3277–3287.
113. M. Müller, T. Mikoviny, S. Feil, S. Haidacher, G. Hanel, E. Hartungen, A. Jordan, L. Märk, P. Mutschlechner, R. Schottkowsky, P. Sulzer, J. H. Crawford and A. Wisthaler, A compact PTR-ToF-MS instrument for airborne measurements of volatile organic compounds at high spatiotemporal resolution, *Atmos. Meas. Tech.*, 2014, **7**, 3763–3772.
114. A. Koss, B. Yuan, C. Warneke, J. B. Gilman, B. M. Lerner, P. R. Veres, J. Peischl, S. Eilerman, R. Wild, S. S. Brown, C. R. Thompson, T. Ryerson, T. Hanisco, G. M. Wolfe, J. M. St Clair, M. Thayer, F. N. Keutsch, S. Murphy and J. De Gouw, Observations of VOC emissions and photochemical products over US oil- and gas-producing regions using high-resolution H3O+ CIMS (PTR-ToF-MS), *Atmos. Meas. Tech.*, 2017, **10**, 2941–2968.
115. J. M. Etzkorn, N. G. Davey, a J. Thompson, a S. Creba, C. W. Leblanc, C. D. Simpson, E. T. Krogh and C. G. Gill, The Use of MIMS-MS-MS in field locations as an on-line quantitative

environmental monitoring technique for trace contaminants in air and water., *J. Chromatogr. Sci.*, 2009, **47**, 57–66.
116. N. G. Davey, PhD Dissertation, University of Victoria, 2016.
117. R. A. Ketola, T. Kotiaho, M. E. Cisper and T. M. Allen, Environmental applications of membrane introduction mass spectrometry, *J. Mass Spectrom.*, 2002, **37**, 457–476.
118. M. A. LaPack, J. C. Tou and C. G. Enke, Membrane mass spectrometry for the direct trace analysis of volatile organic compounds in air and water, *Anal. Chem*., 1990, **62**, 1265–1271.
119. A. M. Ellis and C. A. Mayhew, *Proton Transfer Reaction Mass Spectrometry: Principles and Applications*, John Wiley & Sons, Ltd, Chichester, United Kingdom, 2014.
120. P. K. Hopke, Chemometrics applied to environmental systems, *Chemom. Intell. Lab. Syst.*, 2015, **149**, 205–214.
121. B. K. Lavine and J. Workman, Chemometrics, *Anal. Chem*., 2013, **85**, 705–714.
122. D. Granato, P. Putnik, D. B. Kovačević, J. S. Santos, V. Calado, R. S. Rocha, A. G. Da Cruz, B. Jarvis, O. Y. Rodionova and A. Pomerantsev, Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing, *Compr. Rev. Food Sci. Food Saf.*, 2018, **17**, 663–677.
123. R. Kumar and V. Sharma, Chemometrics in forensic science, *TrAC - Trends Anal. Chem.*, 2018, **105**, 191–201.
124. A. Popovic, M. Morelato, C. Roux and A. Beavis, Review of the most common chemometric techniques in illicit drug profiling, *Forensic Sci. Int.*, 2019, **302**, 109911.
125. R. Madsen, T. Lundstedt and J. Trygg, Chemometrics in metabolomics-A review in human disease diagnosis, *Anal. Chim. Acta*, 2010, **659**, 23–33.
126. C. Gendrin, Y. Roggo and C. Collet, Pharmaceutical applications of vibrational chemical imaging and chemometrics: A review, *J. Pharm. Biomed. Anal.*, 2008, **48**, 533–553.
127. N. D. Lourenço, J. A. Lopes, C. F. Almeida, M. C. Sarraguça and H. M. Pinheiro, Bioreactor monitoring with spectroscopy and chemometrics: A review, *Anal. Bioanal. Chem.*, 2012, **404**, 1211–1237.
128. H. A. Gad, S. H. El-Ahmady, M. I. Abou-Shoer and M. M. Al-Azizi, Application of chemometrics in authentication of herbal medicines: A review, *Phytochem. Anal.*, 2013, **24**, 1–24.
129. B. Krakowska, D. Custers, E. Deconinck and M. Daszykowski, Chemometrics and the identification of counterfeit medicines—A review, *J. Pharm. Biomed. Anal.*, 2016, **127**, 112–122.
130. J. Chapman, V. K. Truong, A. Elbourne, S. Gangadoo, S. Cheeseman, P. Rajapaksha, K. Latham, R. J. Crawford and D. Cozzolino, Combining Chemometrics and Sensors: Toward New Applications in Monitoring and Environmental Analysis, *Chem. Rev.*, 2020, **120**, 6048–6069.
131. R. Bro and A. K. Smilde, Principal component analysis, *Anal. Methods*, 2014, **6**, 2812–2831.
132. P. K. Hopke, Recent developments in receptor modeling, *J. Chemom*., 2003, **17**, 255–265.
133. G. W. Johnson, R. Ehrlich, W. Full and S. Ramos, in *Introduction to Environmental Forensics*, eds. B. L. Murphy and R. D. Morrison, Academic Press, San Diego, 3rd ed., 2015, ch. 18, pp. 609–653.
134. J. G. Watson and J. C. Chow, in *Introduction to Environmental Forensics*, eds. B. L. Murphy and R. D. Morrison, Academic Press, San Diego, 3rd ed., 2015, ch. 20, pp. 677–706.
135. R. Tauler, M. Viana, X. Querol, A. Alastuey, R. M. Flight, P. D. Wentzell and P. K. Hopke, Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies, *Atmos. Environ.*, 2009, **43**, 3989–3997.

136. P. D. Wentzell, T. K. Karakach, S. Roy, M. J. Juanita, C. P. Allen and M. Werner-Washburne, Multivariate curve resolution of time course microarray data, *BMC Bioinformatics*, 2006, **7**, 1–19.
137. S. Demir, A. Saral, F. Ertürk, S. L. Kuzu, B. I. Goncaloğlu and G. Demir, Effect of diurnal changes in VOC source strengths on performances of receptor models, *Environ. Sci. Pollut. Res*., 2012, **19**, 1503–1514.
138. S. Yao, Q. Wang, J. Zhang, R. Zhang, Y. Gao, H. Zhang, J. Li and Z. Zhou, Ambient volatile organic compounds in a heavy industrial city: Concentration, ozone formation potential, sources, and health risk assessment, *Atmos. Pollut. Res*., 2021, **12**, 101053.
139. C. Xiong, N. Wang, L. Zhou, F. Yang, Y. Qiu, J. Chen, L. Han and J. Li, Component characteristics and source apportionment of volatile organic compounds during summer and winter in downtown Chengdu, southwest China, *Atmos. Environ.*, 2021, **258**, 118485.
140. H. Jorquera and B. Rappenglück, Receptor modeling of ambient VOC at Santiago, Chile, *Atmos. Environ.*, 2004, 38, 4243–4263.
141. Y. Song, W. Dai, M. Shao, Y. Liu, S. Lu, W. Kuster and P. Goldan, Comparison of receptor models for source apportionment of volatile organic compounds in Beijing, China, *Environ. Pollut.*, 2008, **156**, 174–183.
142. M. H. Dehghani, D. Sanaei, R. Nabizadeh, S. Nazmara and P. Kumar, Source apportionment of BTEX compounds in Tehran, Iran using UNMIX receptor model, *Air Qual. Atmos. Heal*., 2017, **10**, 225–234.
143. M. Sarkhosh, A. H. Mahvi, M. Yunesian, R. Nabizadeh, S. H. Borji and A. G. Bajgirani, Source apportionment of volatile organic compounds in Tehran, Iran, *Bull. Environ. Contam. Toxicol.*, 2013, **90**, 440–445.
144. A. Stojić, S. S. Stojić, A. Šoštarić, L. Ilić, Z. Mijić and S. Rajšić, Characterization of VOC sources in an urban area based on PTR-MS measurements and receptor modelling, *Environ. Sci. Pollut. Res*., 2015, **22**, 13137–13152.
145. H. a N. Meng, L. U. Xueqiang, Z. Chunsheng, R. a N. Liang and H. a N. Suqin, Characterization and Source Apportionment of Volatile Organic Compounds in Urban and Suburban Tianjin , China, *Adv. Atmos. Sci.*, 2015, **32**, 439–444.
146. L. K. Sahu, R. Yadav and D. Pal, Source identification of VOCs at an urban site of western India: Effect of marathon events and anthropogenic emissions, *J. Geophys. Res*., 2016, **121**, 2416–2433.
147. V. K. Holden and S. E. Hines, Update on flavoring-induced lung disease, *Curr. Opin. Pulm. Med*., 2016, **22**, 158–164.
148. E. Rodolfo Sosa, A. Humberto Bravo, A. Violeta Mugica, A. Pablo Sanchez, L. Emma Bueno and S. Krupa, Levels and source apportionment of volatile organic compounds in southwestern area of Mexico City, *Environ. Pollut.*, 2009, **157**, 1038–1044.
149. R. U. Shah, E. S. Robinson, P. Gu, A. L. Robinson, J. S. Apte and A. A. Presto, High-spatial-resolution mapping and source apportionment of aerosol composition in Oakland, California, using mobile aerosol mass spectrometry, *Atmos. Chem. Phys.*, 2018, **18**, 16325–16344.
150. Y. Zhang, O. Peräkylä, C. Yan, L. Heikkinen, M. Äijälä, K. R. Daellenbach, Q. Zha, M. Riva, O. Garmash, H. Junninen, P. Paatero, D. Worsnop and M. Ehn, A novel approach for simple statistical analysis of high-resolution mass spectra, *Atmos. Meas. Tech.*, 2019, **12**, 3761–3776.

151. E. L. Mungall, J. P. D. Abbatt, J. J. B. Wentzell, A. K. Y. Lee, J. L. Thomas, M. Blais, M. Gosselin, L. A. Miller, T. Papakyriakou, M. D. Willis and J. Liggio, Microlayer source of oxygenated volatile organic compounds in the summertime marine Arctic boundary layer, *PNAS*, 2017, **114**, 6203–6208.
152. C. W. Tessum, D. A. Paolella, S. E. Chambliss, J. S. Apte, J. D. Hill and J. D. Marshall, $PM_{2.5}$ polluters disproportionately and systemically affect people of color in the United States, *Sci. Adv*., 2021, *7*, 1–7.
153. H. Yu and A. L. Stuart, Exposure and inequality for select urban air pollutants in the Tampa Bay area, *Sci. Total Environ.*, 2016, **551–552**, 474–483.
154. M. A. G. Demetillo, A. Navarro, K. K. Knowles, K. P. Fields, J. A. Geddes, C. R. Nowlan, S. J. Janz, L. M. Judd, J. Al-Saadi, K. Sun, B. C. McDonald, G. S. Diskin and S. E. Pusede, Observing nitrogen dioxide air pollution inequality using high-spatial-resolution remote sensing measurements in Houston, Texas, *Environ. Sci. Technol.*, 2020, **54**, 9882–9895.
155. V. A. Southerland, S. C. Anenberg, M. Harris, J. Apte, P. Hystad, A. van Donkelaar, R. V. Martin, M. Beyers and A. Roy, Assessing the distribution of air pollution health risks within cities: A neighborhood-scale analysis leveraging high-resolution data sets in the Bay Area, California, *Environ. Health Perspect.*, 2021, **129**, 037006.
156. Environment and Climate Change Canada, Volatile Organic Compounds in Consumer and Commercial Products, http://www.ec.gc.ca/cov-voc/, (accessed 23 February 2017).
157. P. S. Monks, Gas-phase radical chemistry in the troposphere., *Chem. Soc. Rev.*, 2005, 34, 376–395.
158. C. Kaltsonoudis, E. Kostenidou, K. Florou, M. Psichoudaki and S. N. Pandis, Temporal variability and sources of VOCs in urban areas of the eastern Mediterranean, *Atmos. Chem. Phys.*, 2016, **16**, 14825–14842.
159. United States Environmental Protection Agency, National Emission Standards for Hazardous Air Pollutants (NESHAP), https://www.epa.gov/stationary-sources-air-pollution/national-emission-standards-hazardous-air-pollutants-neshap-9, (accessed 24 February 2017).
160. E. T. Krogh and C. G. Gill, Membrane introduction mass spectrometry (MIMS): A versatile tool for direct, real-time chemical measurements, *J. Mass Spectrom.*, 2014, **49**, 1205–1213.
161. B. K. Lavine and J. Workman, Chemometrics, *Anal. Chem*., 2002, **74**, 2763–2769.
162. K. Varmuza and P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*, CRC Press, Boca Raton, FL, 2016.
163. M. Terrado, D. Barceló and R. Tauler, Quality assessment of the multivariate curve resolution alternating least squares method for the investigation of environmental pollution patterns in surface water, *Environ. Sci. Technol.*, 2009, *43*, 5321–5326.
164. F. Peña, S. Cárdenas, M. Gallego and M. Valcárcel, Direct olive oil authentication: Detection of adulteration of olive oil with hazelnut oil by direct coupling of headspace and mass spectrometry, and multivariate regression techniques, *J. Chromatogr. A*, 2005, **1074**, 215–221.
165. R. R. Catharino, R. Haddad, L. G. Cabrini, I. B. S. Cunha, A. C. H. F. Sawaya and M. N. Eberlin, Characterization of vegetable oils by electrospray ionization mass spectrometry fingerprinting: Classification, quality, adulteration, and aging, *Anal. Chem*., 2005, **77**, 7429–7433.
166. W. Jablonski, K. R. Gaston, M. R. Nimlos, D. L. Carpenter, C. J. Feik and S. D. Phillips, Pilot-scale gasification of corn stover, switchgrass, wheat straw, and wood: 2. identification of global

chemistry using multivariate curve resolution techniques, *Ind. Eng. Chem. Res*., 2009, **48**, 10691–10701.

167. K. P. Pfeuffer, J. Caldwell, J. T. Shelley, S. J. Ray and G. M. Hieftje, Detection of counterfeit electronic components through ambient mass spectrometry and chemometrics., *Analyst*, 2014, **139**, 4505–11.
168. M. J. Pavlovich, E. E. Dunn and A. B. Hall, Chemometric brand differentiation of commercial spices using direct analysis in real time mass spectrometry, *Rapid Commun. Mass Spectrom.*, 2016, **30**, 1123–1130.
169. D. Taverna, L. Di Donna, F. Mazzotti, A. Tagarelli, A. Napoli, E. Furia and G. Sindona, Rapid discrimination of bergamot essential oil by paper spray mass spectrometry and chemometric analysis, *J. Mass Spectrom.*, 2016, **51**, 761–767.
170. S. K. Ohorodnik, R. E. Shaffner, J. H. Callahan and S. L. Rose-Pehrsson, Mixture Analysis Using Membrane Introduction Mass Spectrometry and Multivariate Calibration, *Anal. Chem*., 1997, **69**, 4721–4727.
171. W. P. Gardner, J. H. Callahan, J. E. Girard and R. E. Shaffer, Multivariate mixture analysis using reduced mass resolution-membrane introduction mass spectrometry and variable selection, *Proc. SPIE*, 1999, **3856**, 181–192.
172. R. A. Ketola, J. Heikkonen, S. Piepponen, F. R. Lauritsen and T. Kotiaho, Classification of cola beverages on the basis of mass spectra measured by membrane inlet mass spectrometry, *Rapid Commun. Mass Spectrom.*, 1998, **12**, 1011–1017.
173. R. A. Ketola, M. Ojala, V. Komppa, T. Kotiaho, J. Juujärvi and J. Heikkonen, A non-linear asymmetric error function-based least mean square approach for the analysis of multicomponent mass spectra measured by membrane inlet mass spectrometry, *Rapid Commun. Mass Spectrom.*, 1999, **13**, 654–662.
174. R. M. Alberici, C. G. Zampronio, R. J. Poppi and M. N. Eberlin, Water solubilization of ethanol and BTEX from gasoline: on-line monitoring by membrane introduction mass spectrometry, *Analyst*, 2002, **127**, 230–234.
175. C. G. Thompson, A. J., Creba, A. S., Ferguson, R. M., Krogh, E. T., Gill, A coaxially heated membrane introduction mass spectrometry interface for the rapid and sensitive on-line measurement of volatile and semi-volatile organic contaminants in air and water at parts-per-trillion levels, *Rapid Commun. Mass Spectrom.*, 2006, **20**, 2000–2008.
176. J. H. L. Nelson, E. T. Krogh, C. G. Gill and D. a. Friesen, Monitoring the TiO 2 -Photocatalyzed Destruction of Aqueous Environmental Contaminants at Parts-per-trillion Levels Using Membrane Introduction Mass Spectrometry (MIMS), *J. Environ. Sci. Heal. Part A*, 2004, **39**, 2307–2317.
177. W. R. Plass, H. Li and R. G. Cooks, Theory, simulation and measurement of chemical mass shifts in RF quadrupole ion traps, *Int. J. Mass Spectrom.*, 2003, **228**, 237–267.
178. G. Wu, R. G. Cooks and Z. Ouyang, Geometry optimization for the cylindrical ion trap: Field calculations, simulations and experiments, *Int. J. Mass Spectrom.*, 2005, **241**, 119–132.
179. A. de Juan, J. Jaumot and R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, *Anal. Methods*, 2014, **6**, 4964–4976.
180. A. de Juan and R. Tauler, in *Resolving Spectral Mixtures: With Applications from Ultrafast Time-Resolved Spectroscopy to Super-Resolution Imaging*, ed. C. Ruckebusch, Elsevier, Amsterdam, 1st edn., 2016, ch. 2, pp. 5–52.

181. D. M. Haaland, H. D. T. Jones and J. A. Timlin, in *Resolving Spectral Mixtures: With Applications from Ultrafast Time-Resolved Spectroscopy to Super-Resolution Imaging*, ed. C. Ruckebusch, Elsevier, Amsterdam, 1st edn., 2016, ch. 12, pp. 381–408.
182. C. Janfelt, R. Graesboll and F. R. Lauritsen, Characterization and optimization of membrane inlets for a miniature ion trap mass spectrometer operating at a high background pressure of humid air, *Int. J. Mass Spectrom.*, 2008, **276**, 17–23.
183. I. J. Simpson, N. J. Blake, B. Barletta, G. S. Diskin, H. E. Fuelberg, K. Gorham, L. G. Huey, S. Meinardi, F. S. Rowland, S. A. Vay, A. J. Weinheimer, M. Yang and D. R. Blake, Characterization of trace gases measured over Alberta oil sands mining operations: 76 speciated C2 – C10 volatile organic compounds (VOCs), CO_2, CH_4, CO, NO, NO_2, NO_y, O_3, and SO_2, *Atmos. Chem. Phys.*, 2010, **10**, 11931–11954.
184. United States Environmental Protection Agency, Method TO-17: Determination of Volatile Organic Compounds in Ambient Air Using Active Sampling Onto Sorbent Tubes, 1999.
185. L. Cui, Z. Zhang, Y. Huang, S. C. Lee, D. R. Blake, K. F. Ho, B. Wang, Y. Gao, X. M. Wang and P. K. K. Louie, Measuring OVOCs and VOCs by PTR-MS in an urban roadside microenvironment of Hong Kong: Relative humidity and temperature dependence, and field intercomparisons, *Atmos. Meas. Tech.*, 2016, **9**, 5763–5779.
186. C. Sarkar, V. Sinha, B. Sinha, A. K. Panday, M. Rupakheti and M. G. Lawrence, Source apportionment of NMVOCs in the Kathmandu Valley during the SusKat-ABC international field campaign using positive matrix factorization, *Atmos. Chem. Phys.*, 2017, **17**, 8129–8156.
187. C. Song, B. Liu, Q. Dai, H. Li and H. Mao, Temperature dependence and source apportionment of volatile organic compounds (VOCs) at an urban site on the north China plain, *Atmos. Environ.*, 2019, **207**, 167–181.
188. J. Li, Y. Hao, M. Simayi, Y. Shi, Z. Xi and S. Xie, Verification of anthropogenic VOC emission inventory through ambient measurements and satellite retrievals, *Atmos. Chem. Phys.*, 2019, **19**, 5905–5921.
189. T. W. Tokarek, C. A. Odame-Ankrah, J. A. Huo, R. McLaren, A. K. Y. Lee, M. G. Adam, M. D. Willis, J. P. D. Abbatt, C. Mihele, A. Darlington, R. L. Mittermeier, K. Strawbridge, K. L. Hayden, J. S. Olfert, E. G. Schnitzler, D. K. Brownsey, F. V. Assad, G. R. Wentworth, A. G. Tevlin, D. E. J. Worthy, S. M. Li, J. Liggio, J. R. Brook and H. D. Osthoff, Principal component analysis of summertime ground site measurements in the Athabasca oil sands with a focus on analytically unresolved intermediate-volatility organic compounds, *Atmos. Chem. Phys.*, 2018, **18**, 17819–17841.
190. R. T. Short, S. K. Toler, G. P. G. Kibelka, D. T. Rueda Roa, R. J. Bell and R. H. Byrne, Detection and quantification of chemical plumes using a portable underwater membrane introduction mass spectrometer, *TrAC - Trends Anal. Chem.*, 2006, **25**, 637–646.
191. B. Yuan, A. Koss, C. Warneke, J. B. Gilman, B. M. Lerner, H. Stark and J. A. De Gouw, A high-resolution time-of-flight chemical ionization mass spectrometer utilizing hydronium ions (H3O+ ToF-CIMS) for measurements of volatile organic compounds in the atmosphere, *Atmos. Meas. Tech.*, 2016, **9**, 2735–2752.
192. L. C. Richards, N. G. Davey, T. F. Fyles, C. G. Gill and E. T. Krogh, Discrimination of constructed air samples using multivariate analysis of full scan membrane introduction mass spectrometry (MIMS) data, *Rapid Commun. Mass Spectrom.*, 2018, **32**, 349–360.

193. N. G. Davey, E. T. Krogh and C. G. Gill, Membrane-introduction mass spectrometry (MIMS), *TrAC - Trends Anal. Chem.*, 2011, **30**, 1477–1485.
194. D. Pagonis, K. Sekimoto and J. de Gouw, A Library of Proton-Transfer Reactions of H3O+ Ions Used for Trace Gas Detection, *J. Am. Soc. Mass Spectrom.*, 2019, **30**, 1330–1335.
195. R. S. Blake, P. S. Monks and A. M. Ellis, Proton-Transfer Reaction Mass Spectrometry, *Chem. Rev.*, 2009, **109**, 861–896.
196. H. Abdi and L. Williams, Principal component analysis, *Wiley Interdiscip. Rev. Comput. Stat*., 2010, **2**, 433–459.
197. K. P. Wyche, P. S. Monks, K. L. Smallbone, J. F. Hamilton, M. R. Alfarra, A. R. Rickard, G. B. McFiggans, M. E. Jenkin, W. J. Bloss, A. C. Ryan, C. N. Hewitt and A. R. MacKenzie, Mapping gas-phase organic reactivity and concomitant secondary organic aerosol formation: Chemometric dimension reduction techniques for the deconvolution of complex atmospheric data sets, *Atmos. Chem. Phys.*, 2015, **15**, 8077–8100.
198. L. Cappellin, F. Biasioli, P. M. Granitto, E. Schuhfried, C. Soukoulis, F. Costa, T. D. Märk and F. Gasperi, On data analysis in PTR-TOF-MS: From raw spectra to data mining, *Sensor Actuat. B-Chem*., 2011, **155**, 183–190.
199. R. Sparrapan, M. N. Eberlin and R. M. Alberici, Natural and artificial markers of gasoline detected by membrane introduction mass spectrometry, *Anal. Methods*, 2011, **3**, 751.
200. Z. Deuscher, I. Andriot, E. Sémon, M. Repoux, S. Preys, J. M. Roger, R. Boulanger, H. Labouré and J. L. Le Quéré, Volatile compounds profiling by using proton transfer reaction-time of flight-mass spectrometry (PTR-ToF-MS). The case study of dark chocolates organoleptic differences, *J. Mass Spectrom.*, 2019, **54**, 92–119.
201. I. Khomenko, I. Stefanini, L. Cappellin, V. Cappelletti, P. Franceschi, D. Cavalieri, T. D. Märk and F. Biasioli, Non-invasive real time monitoring of yeast volatilome by PTR-ToF-MS, *Metabolomics*, 2017, **13**, 1–13.
202. R. J. Bell, N. G. Davey, M. Martinsen, R. T. Short, C. G. Gill and E. T. Krogh, The effect of the Earth's and stray magnetic fields on mobile mass spectrometer systems, *J. Am. Soc. Mass Spectrom.*, 2015, **26**, 201–211.
203. J. Han, Jiawei; Kamber, Micheline; Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, Waltham, MA, 3rd edn., 2012.
204. Mathworks, Tune Gaussian Mixture Models, https://www.mathworks.com/help/stats/tune-gaussian-mixture-models.html, (accessed 15 June 2019).
205. M. Loetzsch, Komoot Outdoor in Google Earth, http://ge-map-overlays.appspot.com/openstreetmap/komoot-outdoor, (accessed 15 July 2019).
206. P. D. Wentzell, C. C. Wicks, J. W. B. Braga, L. F. Soares, T. C. M. Pastore, V. T. R. Coradin and F. Davrieux, Implications of measurement error structure on the visualization of multivariate chemical data: Hazards and alternatives, *Can. J. Chem*., 2018, **96**, 738–748.
207. J. H. Gross, *Mass Spectrometry: A Textbook*, Springer International Publishing, Cham, Switzerland, 3rd ed., 2017.
208. F. Autelitano, F. Bianchi and F. Giuliani, Airborne emissions of asphalt/wax blends for warm mix asphalt production, *J. Clean. Prod*., 2017, **164**, 749–756.
209. T. J. Bruno and S. Allen, Weathering patterns of ignitable liquids with the advanced distillation curve method, *J. Res. Natl. Inst. Stand. Technol*., 2013, **118**, 29–51.

210. E. Kari, P. Miettinen, P. Yli-Pirilä, A. Virtanen and C. L. Faiola, PTR-ToF-MS product ion distributions and humidity-dependence of biogenic volatile organic compounds, *Int. J. Mass Spectrom.*, 2018, **430**, 87–97.
211. C. Taiti, C. Costa, W. Guidi Nissim, S. Bibbiani, E. Azzarello, E. Masi, C. Pandolfi, F. Pallottino, P. Menesatti and S. Mancuso, Assessing VOC emission by different wood cores using the PTR-ToF-MS technology, *Wood Sci. Technol.*, 2017, **51**, 273–295.
212. A. Amann, B. D. L. Costello, W. Miekisch, J. Schubert, B. Buszewski, J. Pleil, N. Ratcliffe and T. Risby, The human volatilome: Volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva, *J. Breath Res*., 2014, **8**, 034001.
213. M. Viana, T. A. J. Kuhlbusch, X. Querol, A. Alastuey, R. M. Harrison, P. K. Hopke, W. Winiwarter, M. Vallius, S. Szidat, A. S. H. Prévôt, C. Hueglin, H. Bloemen, P. Wåhlin, R. Vecchi, A. I. Miranda, A. Kasper-Giebl, W. Maenhaut and R. Hitzenberger, Source apportionment of particulate matter in Europe: A review of methods and results, *J. Aerosol Sci*., 2008, **39**, 827–849.
214. C. A. Belis, F. Karagulian, B. R. Larsen and P. K. Hopke, Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe, *Atmos. Environ.*, 2013, **69**, 94–108.
215. J. G. Slowik, A. Vlasenko, M. McGuire, G. J. Evans and J. P. D. Abbatt, Simultaneous factor analysis of organic particle and gas mass spectra: AMS and PTR-MS measurements at an urban site, *Atmos. Chem. Phys.*, 2010, **10**, 1969–1988.
216. A. Vlasenko, J. G. Slowik, J. W. Bottenheim, P. C. Brickell, R. Y. W. Chang, A. M. Maedonald, N. C. Shantz, S. J. Sjostedt, H. A. Wiebe, W. R. Leaitch and J. P. D. Abbatt, Measurements of VOCs by proton transfer reaction mass spectrometry at a rural Ontario site: Sources and correlation to aerosol composition, *J. Geophys. Res. Atmos.*, 2009, **114**, 1–16.
217. Y. Izadmanesh, J. B. Ghasemi and R. Tauler, Receptor modeling of environmental aerosol data using MLPCA-MCR-ALS, *Chemom. Intell. Lab. Syst.*, 2017, **167**, 50–62.
218. I. Stanimirova, R. Tauler and B. Walczak, A comparison of positive matrix factorization and the weighted multivariate curve resolution method. application to environmental data, *Environ. Sci. Technol.*, 2011, **45**, 10102–10110.
219. H. Chen, Y. Teng, J. Li, J. Wu and J. Wang, Source apportionment of trace metals in river sediments: A comparison of three methods, *Environ. Pollut.*, 2016, **211**, 28–37.
220. C. Yan, W. Nie, M. Aïjälä, M. P. Rissanen, M. R. Canagaratna, P. Massoli, H. Junninen, T. Jokinen, N. Sarnela, S. A. K. Häme, S. Schobesberger, F. Canonaco, L. Yao, A. S. H. Prévôt, T. Petäjä, M. Kulmala, M. Sipilä, D. R. Worsnop and M. Ehn, Source characterization of highly oxidized multifunctional compounds in a boreal forest environment using positive matrix factorization, *Atmos. Chem. Phys.*, 2016, **16**, 12715–12731.
221. A. Reff, S. I. Eberly and P. V. Bhave, Receptor modeling of ambient particulate matter data using positive matrix factorization: Review of existing methods, *J. Air Waste Manag. Assoc.*, 2007, **57**, 146–154.
222. P. D. Wentzell, wals.m (2009 update), http://groupwentzell.chemistry.dal.ca/software.html.
223. Statistics Canada, Port Alberni [Population centre], British Columbia and British Columbia [Province] (table). Census Profile. 2016 Census., https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E, (accessed 4 April 2021).

224. L. C. Richards, N. G. Davey, C. G. Gill and E. T. Krogh, Discrimination and geo-spatial mapping of atmospheric VOC sources using full scan direct mass spectral data collected from a moving vehicle, *Environ. Sci. Process. Impacts*, 2020, **22**, 173–186.
225. G. Norris, R. Duvall, S. Brown and S. Bai, EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide, U.S. Environmental Protection Agency, Washington, DC, 2014.
226. A. De Juan, J. Jaumot and R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, *Anal. Methods*, 2014, **6**, 4964–4976.
227. G. W. Johnson, ALS_scat_GWJ.m (updated 2010-2012). Personal communication
228. TSI Incorporated, Aerosol Instrument Manager Software for Optical Particle Sizer (OPS) Spectrometers. User's Manual, TSI Incorporated, Shoreview, Minnesota, 2010.
229. S. Quinn and E. Rardon, *UPDATE: Wildfire continues to burn near Great Central Lake*, Alberni Valley News, August 7, 2018.
230. T. G. Karl, T. J. Christian, R. J. Yokelson, P. Artaxo, W. M. Hao, A. Guenther, W. Min Hao and A. Guenther, The tropical forest and fire emissions experiment: method evaluation of volatile organic compound emissions measured by PTR-MS, FTIR, and GC from tropical biomass burning, *Atmos. Chem. Phys.*, 2007, **7**, 5883–5897.
231. A. M. Yáñez-Serrano, I. Filella, J. LLusià, A. Gargallo-Garriga, V. Granda, E. Bourtsoukidis, J. Williams, R. Seco, L. Cappellin, C. Werner, J. de Gouw and J. Peñuelas, GLOVOCS - Master compound assignment guide for proton transfer reaction mass spectrometry users, *Atmos. Environ.*, 2021, **244**, 117929.
232. Transtech Data Services, Vehicle Turning Movement Survey with Classification - Port Alberni, https://portalberni.ca/traffic-counts, (accessed 25 January 2021).
233. Q. Lu, Y. Zhao and A. L. Robinson, Comprehensive organic emission profiles for gasoline, diesel, and gas-turbine engines including intermediate and semi-volatile organic compound emissions, *Atmos. Chem. Phys.*, 2018, **18**, 17637–17654.
234. D. A. Olson, D. M. Hammond, R. L. Seila, J. M. Burke and G. A. Norris, Spatial gradients and source apportionment of volatile organic compounds near roadways, *Atmos. Environ.*, 2009, **43**, 5647–5653.
235. I. J. Simpson, S. K. Akagi, B. Barletta, N. J. Blake, Y. Choi, G. S. Diskin, A. Fried, H. E. Fuelberg, S. Meinardi, F. S. Rowland, S. A. Vay, A. J. Weinheimer, P. O. Wennberg, P. Wiebring, A. Wisthaler, M. Yang, R. J. Yokelson and D. R. Blake, Boreal forest fire emissions in fresh Canadian smoke plumes: C1-C10 volatile organic compounds (VOCs), CO2, CO, NO2, NO, HCN and CH3CN, *Atmos. Chem. Phys.*, 2011, **11**, 6445–6463.
236. R. J. Yokelson, Emissions of formaldehyde, acetic acid, methanol, and other trace gases from biomass fires in North Carolina measured by airborne Fourier transform infrared spectroscopy, *J. Geophys. Res. Atmos.*, 1999, **104**, 30109–30125.
237. N. Copeland, J. N. Cape, E. Nemitza and M. R. Heal, Volatile organic compound speciation above and within a Douglas fir forest, *Atmos. Environ.*, 2014, **94**, 86–95.
238. X. He, A. K. Lau and S. Sokhansanj, Effect of moisture on gas emissions from stored woody biomass, *Energies*, 2019, **13**, 128.
239. K. Rupar and M. Sanati, The release of terpenes during storage of biomass, *Biomass and Bioenergy*, 2005, **28**, 29–34.
240. J. Kesselmeier and M. Staudt, D077_An Overview on Emission, Physiology and Ecology, *J. Atmos. Chem*., 1999, **33**, 23–88.

241. Catherine Verris Associates Inc. and Gebauer & Associates Ltd., *Somass Estuary Management Plan,* 2005.
242. Y. Li, S. Yin, S. Yu, M. Yuan, Z. Dong, D. Zhang, L. Yang and R. Zhang, Characteristics, source apportionment and health risks of ambient VOCs during high ozone period at an urban site in central plain, China, *Chemosphere*, 2020, **250**, 126283.
243. K. C. Wells, D. B. Millet, L. Hu, K. E. Cady-Pereira, Y. Xiao, M. W. Shephard, C. L. Clerbaux, L. Clarisse, P. F. Coheur, E. C. Apel, J. De Gouw, C. Warneke, H. B. Singh, A. H. Goldstein and B. C. Sive, Tropospheric methanol observations from space: Retrieval evaluation and constraints on the seasonality of biogenic emissions, *Atmos. Chem. Phys.*, 2012, **12**, 5897–5912.
244. National Pollutant Release Inventory, https://www.canada.ca/en/services/environment/pollution-waste-management/national-pollutant-release-inventory.html, (accessed 10 March 2021).
245. M. Modrak, R. A. Hashmonay, R. Varma and R. Kagann, *Measurement of Fugitive Emissions at a Landfill Practicing Leachate Recirculation and Air Injection*, EPA/600/R-05/088 (NTIS PB2006-105545), U.S. Environmental Protection Agency, Washington, DC, 2005.
246. B. G. Heikes, W. Chang, M. E. Q. Pilson, E. Swift, H. B. Singh, A. Guenther, D. J. Jacob, B. D. Field, R. Fall, D. Riemer and L. Brand, Atmospheric methanol budget and ocean implication, *Global Biogeochem. Cycles*, 2002, **16**, 80-1 - 80–13.
247. X. Zhou, Z. Li, T. Zhang, F. Wang, F. Wang, Y. Tao, X. Zhang, F. Wang and J. Huang, Volatile organic compounds in a typical petrochemical industrialized valley city of northwest China based on high-resolution PTR-MS measurements: Characterization, sources and chemical effects, *Sci. Total Environ.*, 2019, **671**, 883–896.
248. BC Air Data Archive., https://envistaweb.env.gov.bc.ca/, (accessed 5 September 2019).
249. B. Deming, D. Pagonis, X. Liu, D. Day, R. Talukdar, J. Krechmer, J. A. de Gouw, J. L. Jimenez and P. J. Ziemann, Measurements of Delays of Gas-Phase Compounds in a Wide Variety of Tubing Materials due to Gas-Wall Interactions, *Atmos. Meas. Tech.*, 2019, **12**, 3453–3461.
250. G. K. Yeh and P. J. Ziemann, Gas-wall partitioning of oxygenated organic compounds: Measurements, structure-activity relationships, and correlation with gas chromatographic retention factor, *Aerosol Sci. Technol*., 2015, **49**, 727–738.
251. P. D. Wentzell, D. T. Andrews, D. C. Hamilton, K. Faber and B. R. Kowalski, Maximum likelihood principal component analysis*, J. Chemom*., 1997, **11**, 339–366.
252. *Data fusion methodology and applications*, ed. M. Cocchi, Elsevier, Cambridge, MA; Amsterdam, Netherlands; 2019.
253. M. Koureas, P. Kirgou, G. Amoutzias, C. Hadjichristodoulou, K. Gourgoulianis and A. Tsakalof, Target analysis of volatile organic compounds in exhaled breath for lung cancer discrimination from other pulmonary diseases and healthy persons, *Metabolites*, 2020, **10**, 317.
254. C. Cainap, L. A. Pop, O. Balacescu and S. S. Cainpa, Early Diagnosis and Screening for Lung Cancer, *Am. J. Cancer Res*., 2021, **10**, 1993–2009.
255. Can I Burn Outdoors Today? Check the Ventilation Index, http://www.env.gov.bc.ca/epd/epdpa/venting/, (accessed 5 September 2019).
256. E. A. Bruns, J. G. Slowik, I. El Haddad, D. Kilic, F. Klein, J. Dommen, B. Temime-Roussel, N. Marchand, U. Baltensperger and A. S. H. Prévôt, Characterization of gas-phase organics using proton transfer reaction time-of-flight mass spectrometry: Fresh and aged residential wood combustion emissions, *Atmos. Chem. Phys.*, 2017, **17**, 705–720.

257. C. Han, R. Liu, H. Luo, G. Li, S. Ma, J. Chen and T. An, Pollution profiles of volatile organic compounds from different urban functional areas in Guangzhou China based on GC/MS and PTR-TOF-MS: Atmospheric environmental implications, *Atmos. Environ.*, 2019, **214**, 116843.
258. M. Gueneron, M. H. Erickson, G. S. Vanderschelden and B. T. Jobson, PTR-MS fragmentation patterns of gasoline hydrocarbons, *Int. J. Mass Spectrom.*, 2015, **379**, 97–109.
259. M. H. Erickson, M. Gueneron and B. T. Jobson, Measuring long chain alkanes in diesel engine exhaust by thermal desorption PTR-MS, *Atmos. Meas. Tech.*, 2014, **7**, 225–239.
260. V. Perraud, S. Meinardi, D. R. Blake and B. J. Finlayson-Pitts, Challenges associated with the sampling and analysis of organosulfur compounds in air using real-time PTR-ToF-MS and offline GC-FID, *Atmos. Meas. Tech.*, 2016, **9**, 1325–1340.
261. A. R. Koss, K. Sekimoto, J. B. Gilman, V. Selimovic, M. M. Coggon, K. J. Zarzana, B. Yuan, B. M. Lerner, S. S. Brown, J. L. Jimenez, J. Krechmer, J. M. Roberts, C. Warneke, R. J. Yokelson and J. De Gouw, Non-methane organic gas emissions from biomass burning: Identification, quantification, and emission factors from PTR-ToF during the FIREX 2016 laboratory experiment, *Atmos. Chem. Phys.*, 2018, **18**, 3299–3319.