THESIS

EXPOSING NEW COMPOSITIONAL COVERAGE OF WEATHERED PETROLEUM HYDROCARBONS THROUGH A TIERED ANALYTICAL APPROACH

Submitted by

Olivia Bojan

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2019

Master's Committee:

Advisor: Jens Blotevogel
Co-Advisor: Tom Sale

Karolien Denef

ABSTRACT

EXPOSING NEW COMPOSITIONAL COVERAGE OF WEATHERED PETROLEUM HYDROCARBONS THROUGH A TIERED ANALYTICAL APPROACH

Petroleum hydrocarbon spills are a widespread source of contamination that may threaten ecosystem services and human health, especially due to modern society's dependence on petroleum-based fuels. Remediation mainly relies on natural source zone depletion (NSZD) processes, which may generate partially oxidized transformation products of the spilled hydrocarbons through weathering or biodegradation processes. These byproducts containing one or more heteroatoms (N, S or O) – referred to as "polar hydrocarbons" – have increased water solubility and mobility in the environment. The unknown fate and toxicity of these complex mixtures of polar metabolites are causing growing concern.

The objectives of this thesis were (1) to use a tiered analytical approach to investigate polar transformation products from various sources and (2) to identify common marker compounds that can be used for a more focused characterization of weathering processes at petroleum-contaminated sites. Previous studies have shown that the majority of weathered petroleum hydrocarbon compounds could not be detected by the GC-based analyses currently required by the United States Environmental Protection Agency due to their low volatility and high molecular weight. Therefore, standard methods may yield misleading characterizations of plumes and impede effective risk management. Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS), an emerging analytical technique in the field of "petroleomics" (the characterization of petroleum at the molecular level) offers unrivaled resolving power and mass

accuracy; here it was used to determine the elemental composition of highly complex petroleum mixtures present in hydrocarbon-impacted sediment samples collected from field sites with varying redox and hydrogeological conditions.

The tiered analysis revealed that GC-based techniques could only detect select nonpolar, low-molecular weight species ($<C_{30}$) present in the sediment samples, while FT-ICR MS assigned molecular formulas to tens of thousands of individual compounds with a wide range of chemical functionalities. Principal component analyses indicated that species belonging to the O_2, N_1, and HC heteroatom classes – corresponding to carboxylic acids, pyrrolic nitrogen compounds, and PAHs, respectively – may be potential marker compounds for plume characterization.

FT-ICR MS results challenged existing site conceptual models and demonstrated the value of this technique as a forensic and source tracking tool. Toxic petroleum-derived N- and S-containing compounds were detected in background samples with "clean" GC chromatograms. Multiple core structures with characteristic double bond equivalents (DBE, a measure of aromaticity) and atomic H:C ratios were associated with unique sources of original spilled products. Asphaltenes, the most recalcitrant fraction of crude oil which is only detectable by FT-ICR MS, were unexpectedly discovered in samples from a former refinery, associating the contaminant plume with a different site owner. Finally, the distribution of polar hydrocarbons between hydrogeologically distinct zones demonstrated the impact of advective transport on the fate of water-soluble metabolites; a higher abundance of oxygenated products was found in an anoxic, low-permeability zone compared to a highly weathered oxic zone of high permeability, challenging previous expectations solely based on redox conditions.

This thesis demonstrates the unique capabilities of FT-ICR MS to enable more comprehensive site characterizations than previously possible, consequently exposing many new unknowns about the fate and transport of polar petroleum metabolites. An important limitation of this technique is the semi-quantitative nature of the results due to preferential ionization; relative abundances of the identified elemental formulas do not directly reflect concentration. More research is also needed to inform toxicological studies and risk assessment of these polar metabolite mixtures. Nevertheless, FT-ICR MS can improve understandings of natural attenuation pathways and the long-term fate of the oxidized transformation products at petroleum hydrocarbon-contaminated sites, all in support of better site management.

ACKNOWLEDGMENTS

Many people contributed to the success of this thesis, and I am extremely grateful for all of the support I received throughout this process. First, thank you to my advisors, Dr. Jens Blotevogel and Dr. Tom Sale for their ideas and guidance, and for giving me the opportunity to dive into this research and contribute to the environmental remediation community.

A portion of this work was performed at the National High Magnetic Field Laboratory, which is supported by the National Science Foundation Cooperative Agreement No. DMR-1644779 and the state of Florida. A special thanks to Amy McKenna for sharing her FT-ICR MS expertise and vast knowledge of crude oil characterization.

Thank you to Tom Sale, Mitch Olson, Maria Irianni Renno, Rick Rogers, Laura Tochko, Nate Jacobs, and Eric Roads for their work collecting sediment core samples. Maria Irianni Renno also performed microbial community analyses. Robert Young developed many of the FT-ICR MS processing scripts and introduced me to R programming. This research was funded by Center for Contaminant Hydrology sponsors. The strong foundations of past CCH students' work and the support of current group members also helped me navigate my own research. Finally, I cannot thank my family enough for their constant encouragement and for supporting all of my life goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

Hydrocarbon spills are a widespread source of contamination, especially due to the modern world's dependence on petroleum-based fuels. Inadvertent contamination of soil and water by petroleum hydrocarbons threatens water quality, ecosystem services, and human health due to the carcinogenic and toxic effects of certain hydrocarbon constituents and their byproducts. As natural source zone depletion (NSZD) processes proceed, oxygenated transformation products of the source hydrocarbons may be generated through weathering processes as a result of incomplete oxidation (Bekins et al., 2016). Previous studies observed a cascade of biotransformation products formed in the first nine months following a spill, rapidly increasing the complexity of the petroleum-derived hydrocarbon mixture (Aeppli et al., 2012). Two years after the spill, the identified peak count relative to the original crude oil decreased, implying that the products generated through weathering were removed by water-flushing, further transformed, or mineralized to CO_2.

These complex mixtures of polar metabolites are the focus of increasing concern among regulators due to their increased water solubility and mobility in the environment, and their largely unknown fate and toxicity (San Francisco Bay Regional Water Quality Control Board, 2016). Currently, the United States Environmental Protection Agency regulations only require the analysis of the nonpolar fraction – dissolved total petroleum hydrocarbons (TPH) – possibly misleading appropriate characterizations of plumes and risk management. These gas chromatography-based analyses omit most of the potentially toxic and mobile transformation products present in soil or groundwater. McKenna et al. (2013) found that most of the weathered

oil compounds could not be detected by GC-based techniques due to their low volatility and high molecular weight.

1.2 Objectives

The main objective of this work was to compare different methods of analysis for advanced site characterization of petroleum hydrocarbon-contaminated sites. These methods yielded multiple tiers of information ranging from non-specific bulk parameters to extensive datasets of compounds previously undetected by traditional analytical techniques. The unrivaled resolving power and mass accuracy of Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) enables the determination of the molecular composition of extremely complex petroleum mixtures. FT-ICR MS analyses highlight new capabilities for complete characterizations of the polar and nonpolar fractions of weathered oil.

A secondary objective was to identify indicator compounds or compound classes that can aid in the characterization and risk assessment of petroleum-contaminated sites undergoing weathering processes. These potential marker compounds would have to be selected from the enormous datasets provided by FT-ICR MS, which are informative but uneconomical for routine site monitoring.

Finally, FT-ICR MS was used to identify compositional changes at the molecular level between samples originating from distinct hydrogeologic zones at hydrocarbon-contaminated sites in order to unravel the processes leading to oxygenated hydrocarbon generation.

As the terms "oxyhydrocarbons" and "polar hydrocarbons" are often used interchangeably in the literature, but may be misleading, this document refers to the entirety of petroleum compounds containing one or more heteroatoms (N, S or O) as “polar hydrocarbons”,

and to petroleum compounds specifically with oxygen inclusion(s), such as weathered oxidation intermediates, as “oxygenated hydrocarbons” or “oxyhydrocarbons”.

2. LITERATURE REVIEW

2.1 Hydrocarbon classes

Petroleum compounds are highly complex mixtures that contain predominantly hydrocarbons (compounds containing molecules of carbon and hydrogen atoms), heteroatom compounds (hydrocarbon compounds with heteroatoms such as sulfur, nitrogen, or oxygen), and occasionally trace concentrations of metallic species. Petroleum components are classified into bulk categories based on their molecular structure. The four main groups are saturates, aromatics, resins, and asphaltenes (SARA) (TPH Criteria Working Group, 1998). The SARA classification is an experimental method used to determine the weight fractions of each of these groups within a complex petroleum mixture (Speight et al., 1991).

Saturates consist of only singly-bonded C and H atoms and do not contain any C=C double bonds. This class includes straight chain *n*-alkanes (also called *n*-paraffins), branched chain iso-alkanes (iso-paraffins), and cycloalkanes (naphthenes or cycloparaffins). Cycloalkanes are saturated hydrocarbons that contain one or more rings and may have saturated side chains. The general formula for saturates is C_nH_{2n}. Alkenes (or olefins) are unsaturated hydrocarbons that contain at least one C=C double bond. Alkynes contain at least one C≡C triple bond. Unsaturated hydrocarbons form during the cracking process of crude oil when heavier molecules are transformed into smaller ones. Unsaturated molecules are therefore found in refined products like gasoline rather than raw crude oil. Like alkanes, olefins can have straight chain, branched, or cyclic structures, and they have the general formula C_nH_{2n-2}.

Aromatic compounds are a class of cyclic unsaturated hydrocarbons that have alternating C=C bonds. Most aromatics have a benzene ring base structure; alkylbenzenes have one or more

side chain substituents attached to the ring. Polycyclic aromatic hydrocarbons (PAHs) contain two or more aromatic rings fused together. All petroleum products and crudes contain aromatics, and they are also commonly found in coal. Heterocyclic compounds are ring structures in which at least one atom in the ring is an element other than carbon. Pyrrole is an example of a heterocyclic compound commonly found in petroleum. The most common heteroatoms found in petroleum are nitrogen, oxygen, and sulfur.

Resins are compounds that contain a polar end group – composed of aromatic and naphthenic rings, often containing heteroatoms – and long alkane tails. Resins tend to be heavy liquids or sticky solids. Finally, asphaltenes are very large, polar, heteroatom-containing compounds composed of many condensed aromatic and naphthenic rings. The typical carbon number range for asphaltenes is 30-80 (Ovalles and Moir, 2018). Asphaltenes are referred to as the heavy ends, or high ends, of petroleum and are nonvolatile and dispersed as colloids in petroleum.

In addition to the four main hydrocarbon groups, smaller amounts of organometallic compounds and inorganic salts may also be present in the heavier distillation fractions and residues after the refining process.

2.2 Types of refined products

The broad category of petroleum includes both crude oil and petroleum products. Crude oil is a complex mixture of hydrocarbons found in underground geologic formations. Petroleum products are the end products of the refining process of crude oil. Refined products of crude oil produced for commercial use are either single fractions or blends of fractions from the distillation of raw crude oil. These products can be sorted into general categories based on their

composition, although their actual properties vary depending on specific refinery operations. The desired specifications for the physical properties of the product – including boiling range distribution, elemental composition, viscosity, flash point, pour point, viscosity index, API gravity, specific gravity, color, ash content, water content, demulsibility, etc. – are defined by their intended application.

Lighter fractions include naphtha and gasoline. Gasoline for automotive fuel is a blended product manufactured from the C_4 to C_{12} boiling range fractions of distilled crude oil. During the refining process, large molecules are cracked, resulting in smaller molecules with C=C double or C≡C triple bonds that fall within the desired carbon range. Components of gasoline targeted by United States Environmental Protection Agency (US EPA) methods are benzene, toluene, ethylbenzene, xylene (BTEX compounds) and naphthalene. Regulatory concern stems from their toxicity and mobility in the environment; they are sufficiently water-soluble to leach into groundwater.

The naphtha category encompasses petroleum products with a carbon range from C_6 to C_{12}. These volatile compounds have a normal boiling range between 40-220˚C, which allows for the use of gas chromatography (GC) to determine composition. Middle distillates include kerosene, jet fuels, and diesel. These compounds make up approximately 10% of crude oil and have a boiling range between 220-345˚C. Over 60% of the compounds in middle distillates have naphthenic ring structures. Diesel fuels can include up to 30-40% aromatics as a result of the cracking process during refining, but the majority is 60-90% linear, branched, and cyclic alkanes. The typical carbon range for diesel is C_8-C_{26}. Frequently, diesel fuels are mixed with additives to protect against rust, corrosions, and deposits. The increased complexity of middle distillates means that GC cannot separate all of the compounds present in these mixtures.

2.3 Current hydrocarbon analysis methods and their limitations

Historically, the evaluation of petroleum contaminated sites has focused on measuring total petroleum hydrocarbon (TPH) concentrations and establishing target cleanup levels for soil or water. Analyses required by the US EPA include dissolved TPH in the gasoline, diesel, or oil (TPHg, TPHd, TPHo) (McGuire et al., 2018). There are many published standard EPA and American Society for Testing and Materials (ASTM) analytical techniques that measure TPH concentrations in the environment (Wang et al., 1997). However, no single method measures the entire range of petroleum hydrocarbons; each method extracts and detects different fractions of these complex mixtures. Interpretation of the TPH data without understanding the capabilities and limitations of the selected method can be misleading and can lead to an inaccurate assessment of risk.

The majority of the standard TPH methods rely on gas chromatography coupled to a flame-ionization detector (GC-FID) or a mass spectrometer (GC-MS). Following sample extraction with an organic solvent, any volatile hydrocarbons that elute from the column will be detected by gas chromatography. The resulting chromatogram, the analog signal from the detector, displays a particular carbon range, with gasoline range organic (GRO) compounds (C_6-C_{10}) eluting before diesel range organic (DRO) compounds (C_{10}-C_{28}), and heavier hydrocarbons eluting later. The FID signal is proportional to the mass of hydrocarbons present, and it is insensitive to the type of hydrocarbon (e.g., aromatic, *n*-alkane and olefin). TPH measurements are calculated by integrating the area under the FID chromatogram curve and comparing the peak area to a calibration curve. GC-based methods are best suited for detecting nonpolar hydrocarbons with carbon numbers between C_6 and C_{36} and compounds that elute below 400 °C (to prevent column degradation). However, many weathered oils contain molecules with more

than 40 carbon atoms, while crude oils contain molecules with over 100 carbon atoms. These heavy hydrocarbons fall outside of the detection window of GC-based methods.

GC-FID can be useful for distinguishing unweathered gasoline and diesel and differentiating product streams in the GRO/DRO range of the chromatogram. This "fingerprint" analysis compares the chromatograms of samples to reference materials. Matching of characteristic patterns in the chromatograms can identify some fuels in a sample or mixture as long as the known source material is available as a reference. GC-FID "fingerprints" are not typically useful diagnostic tools for environmental samples because biodegradation and other weathering processes drastically alter the chromatograms.

Most of the GC-based methods for petroleum sample analysis separate the extracts into aliphatic and aromatic fractions which are then analyzed by two different detectors. The aliphatic fraction is analyzed by GC-FID and the aromatic fraction (including PAHs) is analyzed on a gas chromatograph coupled to a mass spectrometer (GC-MS) (Reddy et al., 1999). GC-MS is a highly selective technique that can identify (semi)volatile analytes by retention time and unique mass spectral pattern. The mass spectrometer selects ions of specific mass-to-charge ratios (*m/z*) and can filter analytes of interest to reduce noise the output mass spectrum. GC-MS is also highly reproducible; after injection, compounds break apart into characteristic fragment ions. The intact parent compound can be identified by the recognizable fragmentation pattern through spectral matching to references databases.

The resolution of GC-based methods is sufficient for nonpolar, volatile hydrocarbons, but many factors limit the capability to analyze complex petroleum mixtures. Isomers have identical GC-MS spectra, and many other compounds found in petroleum have very similar spectra. Incomplete volatilization of polar, oxygenated molecules and high molecular weight ($>C_{40}$)

compounds causes significant fractions of the compounds present to be undetected by these analyses. Even when polar compounds do elute in the column, they are retained on silica gel sorbents during sample preparation and therefore are not detected.

Another major limitation of GC-based techniques is the low resolution and resulting inability to separate peaks corresponding to the diverse structures found in crude oils. Nonpolar isomers that co-elute as well as most of the nonpolar and aromatic compounds in weathered oil samples appear as unresolved complex mixtures (UCMs) or "humps" in the GC chromatograms. These UCMs also challenge the characterization of petroleum compounds present because natural organic matter (NOM), plant oils and waxes may also be extracted and included in the measured TPH as long as they are dissolved in the extraction solvent. Petrogenic and biogenic hydrocarbons cannot be distinguished in GC chromatograms. Therefore, these methods can potentially overestimate TPH concentrations at a contaminated site, leading to unnecessary remediation costs and other regulatory problems (Vecchiato et al., 2017). Historically the remediation of soil contaminated with petroleum hydrocarbons has emphasized reductions in TPH load rather than reductions in risk (Brassington et al., 2007).

In summary, regulations based on target TPH levels may be insufficient because the detected TPH value (peak area) is typically unrepresentative of the concentration of petroleum-derived compounds present, and the measured concentration therefore does not accurately reflect the level of risk associated with the contamination. Oxygenated transformation products, which are mostly undetected by GC-based analyses, may represent a significant threat to ecosystems because of their higher water solubility and mobility in the environment. Clearly, a more complete molecular characterization of petrochemical species – present in both high and low abundances – may be needed for effective remediation and risk assessment, as the environmental

fate and transport of compounds released from a spill is dictated by their structure, functionality, and molecular weight (McCay et al., 2014).

2.4 Petroleum weathering

When oil is released into the environment, changes in the physical and chemical properties of the oil begin to occur, in a process known as weathering. Weathering alters the compositions of the hydrocarbon compounds over time through volatilization, sorption, dissolution, microbial transformation, photolysis, and leaching. Many of these processes form the basis of natural attenuation strategies such as natural source zone depletion (NSZD) processes (Brassington et al., 2007). Oxygenated transformation products ("oxyhydrocarbons", "polar hydrocarbons") of the source hydrocarbons are generated as a result of incomplete oxidation during these processes. Weathering also changes the toxicity and partitioning within the environment. Molecular composition (structure, functionality, molecular weight) of the compounds dictates their environmental fate. Therefore, characterizing the extent of the transformation is essential for guiding remediation strategies.

2.5 Polar transformation products

Polar petroleum compounds, defined as containing one or more heteroatoms (N, S or O), represent a minor component of crude oil (<15% by mass), but they are the focus of increasing concerns among regulators due to their increased water solubility and mobility in the environment, and their unknown fate and toxicity. Polarity increases with higher NSO:C (heteroatom to carbon) ratios because heteroatoms in functional groups are more electronegative compared to the carbon atom. In turn, water-solubility increases with polarity. After an accidental release, abiotic and biotic weathering processes form oxygenated transformation

products that can comprise 60-80% material (Chen et al., 2016). However, results based on nonvolatile dissolved organic carbon (NVDOC) analyses show that polar transformation products quantified with TPHd analysis methods only represent a fraction of the polar transformation products present at a refined fuel site.

In response to accidental petroleum releases, regulatory action is primarily driven by the Clean Water Act 40 CFR Section 311 Oil and Hazardous Substances Part 110, which prohibits "discharge to surface water resulting in: film, sheen, or discoloration" (US EPA, 1996). The first document to directly address polar petroleum-derived compounds was published by the San Francisco Regional Water Quality Control Board in 2016 (Steenson et al., 2016), which advised that the risk associated with petroleum metabolites should be assumed similar to the parent mixture. New research regarding the fate and toxicity of water-soluble polar compounds highlights the need to adapt regulations to effectively protect natural resources at contaminated sites.

2.5.1 Toxicity

Toxicity studies have shown that compounds in the extractable polar fraction of weathered oil samples can be more toxic than nonpolar compounds (Varel et al., 2011). Melbye et al. (2009) identified mutagenic nitro- and dinitro-PAHs, azaarenes, and keto-PAHs in the most polar fractions, defined as fractions containing species with at least one polar functional group. Bioactivation of less toxic parent PAHs can yield mutagenic metabolites, including products of the bay region dihydrodiol epoxide metabolic pathway (Sims et al., 1974). This mechanism refers to the deep structural pocket where a single benzene ring is attached to an angular ring fusion, as shown in Figure 1 outlines the enzyme-activated oxidation of non-carcinogenic

benzo[*a*]pyrene which forms benzo[*a*]pyrene-7,8-diol-9,10-epoxide (BPDE), a highly carcinogenic and non-GC-detectable metabolite.

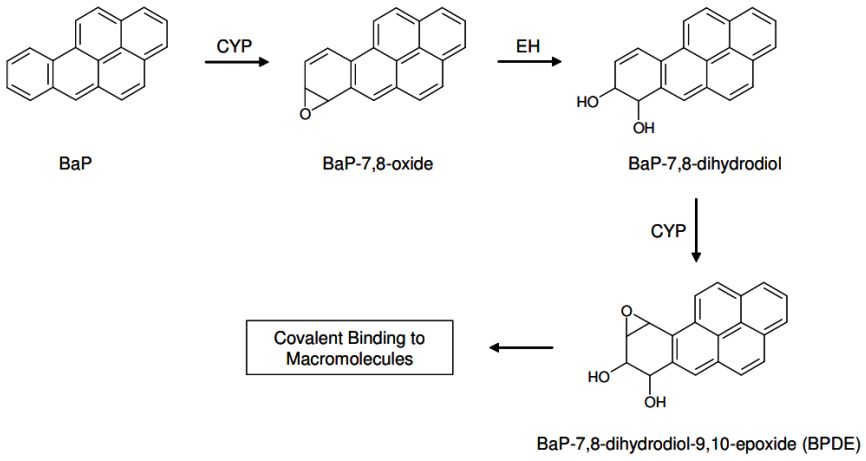

Figure 1: Bioactivation of benzo[*a*]pyrene by bay region dihydrodiol epoxide metabolic pathway forms highly carcinogenic BPDE (Naspinski, 2009).

Figure 1 details the 3-step mechanism: when the bay region epoxide opens, a benzyl cation forms. This benzyl cation is stabilized by the aromatic bonds around the bay region, and is a highly reactive electrophile which readily binds to DNA. Several other PAHs are metabolized through this pathway, including chrysene, 5-methylchrysene, phenanthrene, benzo[*c*]phenanthrene, benz[*a*]anthracene, dimethylbenz[*a*]anthracene, and dibenzo[*a,l*]pyrene (Xue and Warshawsky, 2005).

The emerging awareness of the toxicological profile of these polar products is especially concerning because the commonly monitored priority PAHs (e.g., naphthalenes, polycyclic aromatic hydrocarbons, and alkylated phenols) elute in less polar or nonpolar fractions and exhibit lower toxicity (Melbye et al, 2009).

Further highlighting the dangers of relying on GC-MS to gauge toxicity of weathered oil, most oxygenated metabolites of priority PAHs that are detectable by GC-MS are less toxic than

the parent molecules. Degradation products of fluoranthene identified by Sepic et. al. (2003) were determined to be much less toxic than fluoranthene. Zemo et. al. (2013) also used GC-MS to identify alcohols, acids, ketones, aldehydes, phenols as metabolites of weathered fuels, leading to the suggestion that polar metabolite mixtures in groundwater only showed low-to-moderate toxicological effects. These identified compounds, however, only represented a small fraction of the total ion chromatogram (Bekins et al., 2016). Based on the limited analytical coverage allowed by GC-MS, overall weathering trends appear to involve increasingly higher proportion of low-risk organic acids and esters and a lower human toxicity profile (Zemo et al., 2017), potentially misleading regulatory decisions to rely on natural attenuation.

Implications of underestimating the toxicity of these polar mixtures could be detrimental to the quality of water resources. The assumption that plumes resulting from oil and fuel spills will stabilize (based on stabilizing levels of GC-amenable indicators) then shrink due to NSZD processes may not effectively protect water resources because of the misrepresentation of polar petroleum hydrocarbon fractions in conventional analyses.

2.5.2 Solubility

Since polar compounds have elevated solubility in water, the formation of polar transformation products through weathering of oil complicates the fate and transport models used to monitor contaminated sites. Solubility influences the risk associated with these compounds. Hydrophilic compounds become more bioavailable when dissolved in water compared to being sorbed to soil; exposure to waterborne bacteria or dissolved organic matter (humic acids) may enhance biodegradation of dissolved PAHs or alter their solubility (Varel et al., 2011; Ke et al., 2009). Moderately hydrophobic compounds are soluble enough to stay in the water phase, but preferentially partition to animal lipid membranes where they can bioaccumulate. Liu et. al.

(2015) reported a lower heteroatom-to-carbon ratio (NSO:C) limit and upper mass limit for aqueous partitioning of polar crude oil-derived compounds. These mass-dependent NSO:C limits are shown in Figure 2, which presents the mass on the x-axis and NSO:C on the y-axis. The heteroatom ratios of the crude oil, the water-soluble fraction (WSF), and the water-accommodated fraction (WAF) which contains dispersed oil droplets are compared. The results demonstrated the link between a higher heteroatom-to-carbon ratio and solubility in water.

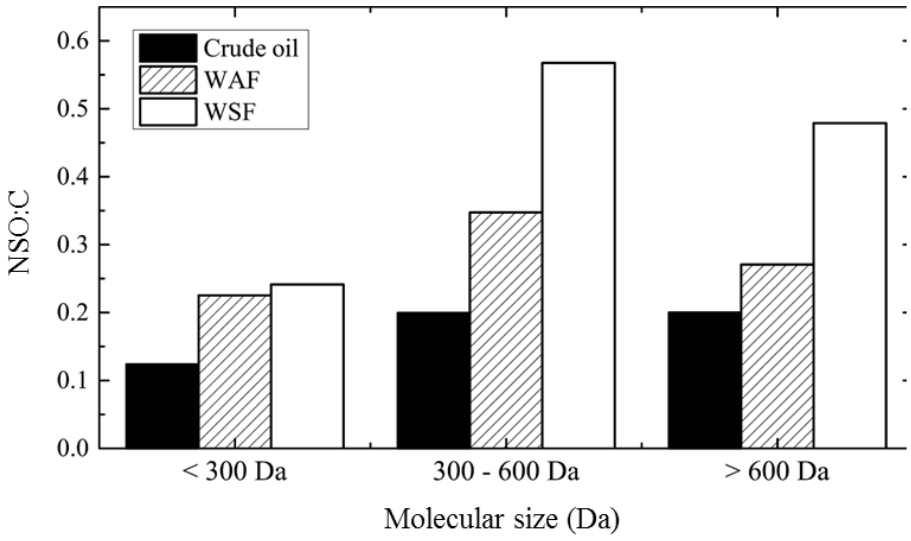

Figure 2: NSO:C vs. molecular size for various fractions of petroleum hydrocarbons. Aqueous partitioning of polar crude oil-derived compounds is associated with higher heteroatom content (Liu et al., 2015).

2.6 Biodegradation

Albeit slow, microbial degradation is the preferred remediation strategy for mature sites because it is sustainable, cost-efficient, and less environmentally invasive than most active remedies. At many sites, remediation relies on these types of natural source zone depletion (NSZD) processes for the majority of the contaminant mass removal (Karimi Askarani et al., 2018). Biodegradation processes can be very complex; the extent of biodegradation is dependent

on many factors including the type of microorganisms present, environmental conditions (temperature, electron acceptor concentrations, moisture content, oxidation-reduction potentials, soil pH or alkalinity, etc.), predominant hydrocarbon types and structural composition, and bioavailability of hydrocarbon contaminants. A wide variety of petroleum hydrocarbons can be degraded by microbes.

The degradation potential of contaminants is primarily determined by their molecular structure. *n*-Alkanes are the most readily degradable in both aerobic and anaerobic conditions. Liu et. al. established that after a crude oil release in a marine environment, the majority of low molecular weight ($<C_{14}$) *n*-alkanes evaporated within 2-3 hours of rising to the surface, while larger alkanes (C_{14}-C_{16}) were attenuated after 10-100 hours (Liu et al., 2012). Branched alkanes and single saturated ring compounds are much more stable (Stiver and Mackay, 1984; Fingas et al., 1999).

PAHs and alkylated benzene derivatives display varying degrees of degradation, with rates correlated to molecular weight, boiling point, and degree of alkylation. After a release, low molecular weight and volatile BTEX compounds are preferentially depleted. PAHs with higher numbers of aromatic rings or greater degrees of alkylation (e.g. 4- to 6-ring condensed aromatics, and alkylated thiophenes and dibenzothiophenes) degrade more slowly or are only partially metabolized (Ryerson et al., 2011; de Gouw et al., 2011).

The biodegradation of hydrocarbon contaminants occurs naturally either under aerobic or anaerobic conditions, depending on the availability of electron acceptors. The reduction-oxidation (redox) ladder illustrated in Figure 3 is the main driver of microbial activity in the subsurface.

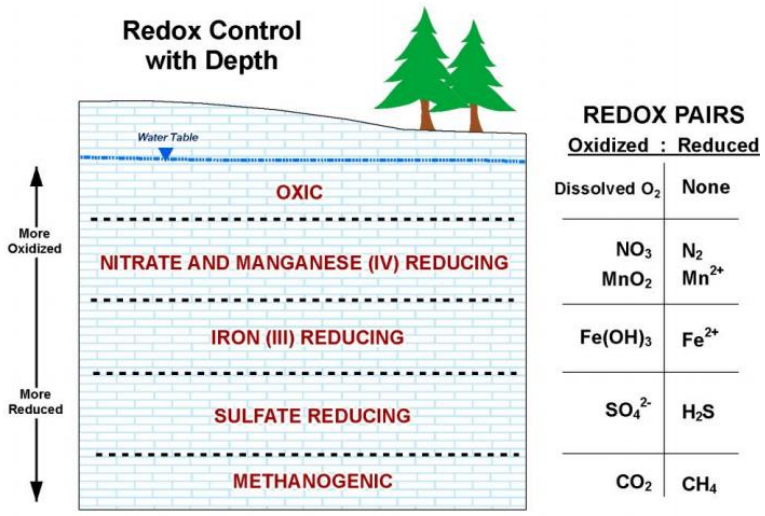

Figure 3: The redox ladder orders redox couples from highest to lowest energy potentials; processes with higher energy potentials are more thermodynamically favorable.

Oxygen is the most energetically-favorable electron acceptor, and oxygenated compounds are readily degraded compared to more recalcitrant S- or N-containing species.

Aerobic degradation mechanisms are well-studied, fast and efficient at removing organic contaminants. Aerobic conditions generally include oxic, transmissive geologic zones because they allow the flow of oxygen and water through pore space in sediments. Many studies of surface-level aquifers contaminated by the Bemidji, MN oil spill indicate that *n*-alkanes (C_6-C_{30}), toluene, and *o*-xylene are the first hydrocarbons depleted after a release. The C_{10}–C_{24} cyclohexanes, tri- and tetramethylbenzenes, acyclic isoprenoids, and naphthalenes are the most resistant to degradation (Baedecker et al., 2018). Aerobic metabolism mechanisms convert alcohol functional groups to aldehydes or ketones. Aldehydes are transformed to acids, and ketones are transformed to acetyl esters and then an acid and an alcohol. Figure 4 illustrates some of the many possible degradation pathways for toluene and carbazole, known structures in oil, under aerobic conditions.

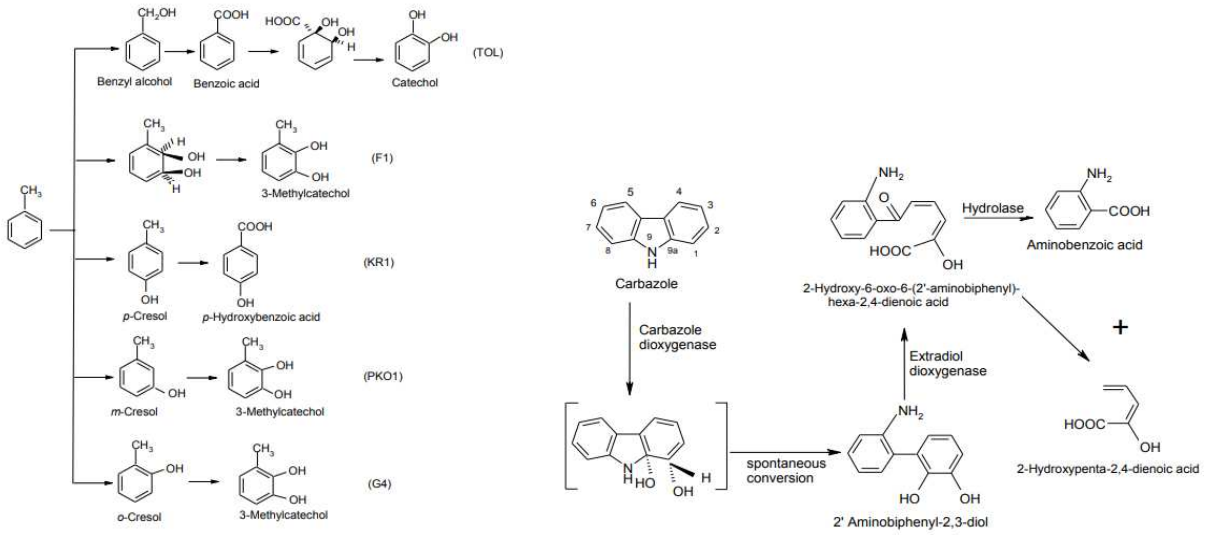

Figure 4: Aerobic degradation pathways for toluene and carbazole (Olajire and Essien, 2014).

One study of aerobic weathering of crude oil tracked molecular-level changes through time of sediment from the Deepwater Horizon spill site (Chen et al., 2016). Comparisons of weathered oil samples to the known parent material, Macondo well oil, showed a progression of oxygenation resulting from weathering – specifically a production of higher-order oxygenated transformation products via the addition of carboxylic acid functional groups onto parent hydrocarbons. Lower-order oxygenated compounds were transformed into higher order species (O_4-O_6). Aeppli et al. (2012) also noted that the formation of oxyhydrocarbons was primarily caused by biodegradation and photooxidation.

At most older, weathered spill sites, oxygen surrounding the oil has been depleted by aerobic organisms. In these anoxic environments, microbial degradation is controlled by the availability of electron acceptors. After preferred electron acceptors (O_2, Mn^{4+}, NO_3^-, and Fe^{3+}) have been depleted, sulfate reduction and methanogenesis are the dominant processes. Hydrogeology can also contribute to the anaerobic conditions. For example in low k zones, pore

throats may be too narrow to allow for advective flow of water and oxygen transport (Garg et al., 2017).

In the saturated zone, methanogenesis is the dominant redox process. Aerobic degradation processes are more likely to occur temporarily at recent releases than older weathered sites. Bekins et al. (2005) reported that at the Bemidji oil spill site, the majority of the plume was under methanogenic conditions, except for the top two meters of the oil body. Methanogenic metabolism of hydrocarbons is a complex process. Struchtemeyer et al. (2005) studied methanogenesis at a refined product (gas condensate) and determined that the primary mechanism was acetate production and oxidation, instead of hydrogenotrophic methanogenesis; these results suggested that crude oil sites may have different methanogenic pathways than refined product sites. Methanogenesis can also occur in a wide variety of geochemical environments, sometimes concurrently with sulfate reduction in the same zone (Gieg et al, 2014; Irianni Renno et al., 2016).

However, recharge and water flow facilitates microbial degradation, and can affect the oxic or anoxic nature of the soil because nutrients are transported into the contaminated zone. Recharge from the land surface has been correlated to more advanced weathering of crude oil (Bekins et al. 2005). In addition, fully or partially oxidized products may also be flushed out of the contaminated zone by water flow or fluctuating water table levels. The constant generation and removal of the degradation byproducts challenges the determination of weathering trends in these dynamic environments.

In general, the classes of petroleum hydrocarbons metabolized under anaerobic conditions are similar to the compounds degraded in aerobic conditions, but the biodegradation follows different trends. Under anaerobic conditions, the *n*-alkanes that are depleted first are

compounds with C_{18} and higher, which is the reverse of the trend observed in aerobic conditions (Bekins et al., 2005 and Baedecker et al., 2011) Another opposite trend was detected in the metabolism of alkylcyclohexanes, which involved the preferential depletion of high molecular-weight compounds of a homologous series and a resulting increase in the relative abundance of lower molecular weight compounds (Hostettler et al., 2002).

The majority of anaerobic biodegradation studies have focused on PAHs due to toxicological concerns driving the research. Some known intermediate PAH degradation products include naphthalene, phenanthrene, dibenzothiophene, fluorine, chrysene (Wang et al., 1999), and transformation products have been shown to be more toxic than their parent compounds, especially when mixture effects are taken into account (Schrlau et al., 2017). Similar to aerobic degradation pathways, fatty acids are byproducts of microbial oxidation; succinic and fumaric acids have been evaluated as indicators of anaerobic microbial activity (Young et al., 2004).

2.7 Biomarkers

Biomarker molecules have important geochemical and forensic applications; petroleum geochemists use biomarkers as "molecular fossils" to relate crude oil to its source and gauge maturity for oil exploration (Wang et al., 2006). Naturally occurring biomarkers found in crude oils, rocks, and sediments are stable and resistant to degradation, and therefore can indicate the nature, source, type, geological conditions, and thermal history of their parent product. Biomarkers can also be used for forensic source tracking to distinguish spilled oil from other sources or natural seeps, and weathering studies characterize the extent of oil transformation using biomarkers as recalcitrant tracers.

Aeppli et al. (2014) studied the degradation potential of various classes using GCxGC to separate and identify low-abundance compounds unresolvable by GC-MS. The biomarker classes included hopanes, steranes and disteranes, isoprenoids, PAHs, and triaromatic steranes (TAS). Hopanes and steranes, derived from prokaryotes and eukaryotes, have characteristic base structures illustrated in Figure 5; hopanes have four cyclohexane rings and one cyclopentane ring with a side chain emerging from C_{30}, and steranes have three cyclohexane rings and one cyclopentane ring with a side chain emerging from C_{17}. Steranes may be rearranged into disteranes during the geological process of diagenesis in sedimentary rock. Isoprenoids are composed of biosynthesized isoprene subunits. They found that over a timescale of 28 months, hopanes smaller than C_{32}, steranes, and disteranes were completely stable and resistant to biotic and abiotic degradation, while the isoprenoid, PAH, and triaromatic steranes (TAS) classes started to degrade sooner; homohopanes (C_{32}-C_{35}) and TAS compounds were shown to be susceptible to bio- and photodegradation. This timeline and the structures of the targeted petroleum biomarkers are illustrated in Figure 5.

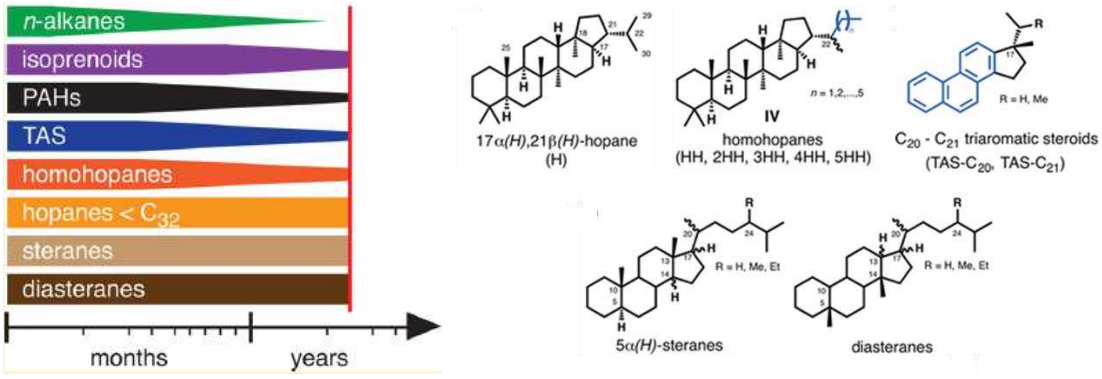

Figure 5: Relative losses of hydrocarbon classes in crude oil as a result of weathering and their corresponding structures. *n*-Alkanes are readily depleted in the first two years, while hopanes (< C_{32}), steranes, and diasteranes are recalcitrant (Aeppli et al., 2014).

Compounds used as biomarkers fall into two categories: aliphatic biomarkers and cyclic biomarkers. The two most abundant in petroleum and most commonly used biomarkers are pristine and phytane, which are isoprenoid molecules belonging to the aliphatic class. Steranes are abundant cyclic biomarkers, consisting of a tetracyclic core structure and an alkyl sidechain. Isoprenoids and steranes both originate from biogenic sources, from decaying plant material. Other examples of aromatic biomarkers include alkylated naphthalenes, phenanthrene, and chrysene, as well as sulfur-containing aromatics dibenzothiophene and benzonaphthothiophene.

2.8 Fourier transform ion cyclotron resonance mass spectrometry

Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry (MS) is an emerging analytical technique that offers the highest mass resolution and mass accuracy currently achievable. FT-ICR MS detects tens of thousands of peaks in a single mass spectrum, and selective ionization modes can target specific functional groups characteristic of compounds in the complex polar fraction of weathered oil. This technique expands the analytical window of complex petroleum hydrocarbon mixtures by detecting nonvolatile, thermally unstable, and polar basic or acidic species that are omitted from GC-based analyses (McKenna et al., 2013). As Figure 6 illustrates, FT-ICR MS provides a new view of compounds undetectable by traditional GC-based methods, which are constrained to volatile and low-molecular-weight compounds. In this figure, the compositional space of petroleum-derived species is represented by double bond equivalents (DBE = number of rings + double bonds) on the y-axis and carbon number on the x-axis. The dashed red Planar PAH Stability Limit symbolizes the theoretical maximum DBE for a planar polycyclic aromatic molecule for all molecules known to exist in petroleum (up to C_{100}) (McKenna et al., 2013).

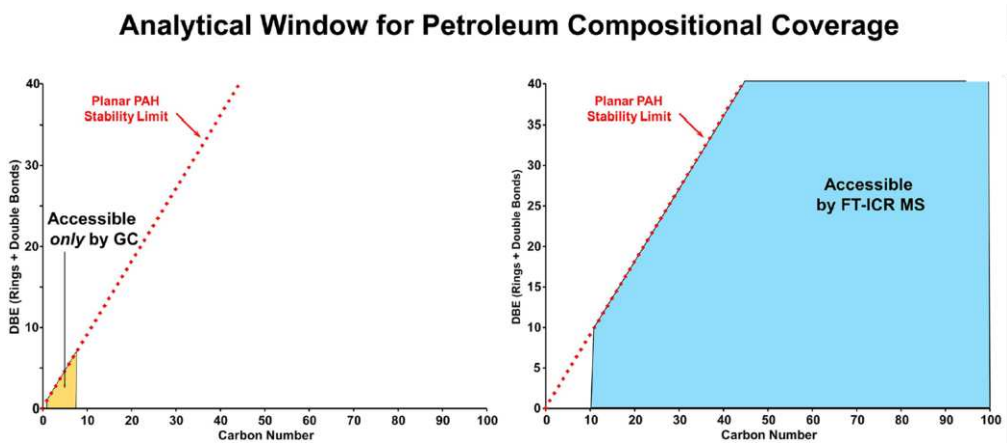

Figure 6: The compositional window accessible by only GC-based techniques is shown on the left; the right diagram shows the extensive range of compounds detectable by FT-ICR MS (McKenna et al., 2013).

The ultrahigh resolving power of FT-ICR MS has advanced the field of "petroleomics" – the characterization of petroleum at the molecular level – and enabled the direct characterization of complex mixtures without prior fractionation (Lobodin et al., 2013). Chemical speciation of petroleum and its derivatives at the molecular level reveals correlations between (and ultimately predictions of) properties and partitioning in the environment.

2.8.1 Mass resolution

The complexity of petroleum samples requires ultrahigh resolution mass spectrometry over a wide mass range (50–1500 Da), criteria only met by FT-ICR MS (Rodgers et al., 2005). Scans of petroleum samples can yield tens of thousands of mass spectral peaks, many with almost identical exact masses. FT-ICR MS regularly measures molecular mass to better than 50−100 ppb when combined with Kendrick mass sorting and isotope fine structure, enabling the assignment of a unique chemical formula, $C_cH_hN_nO_oS_s$, for each species. The sub-ppm accuracy is partly based on the calibration of the ICR frequency-to-m/z conversion procedure. Internal calibration for petroleum samples uses universally present ions of a homologous alkylation series

(differing in elemental composition by multiples of $–CH_2$) that span most of the *m*/*z* range of interest (Marshall et al., 1998).

2.8.2 Kendrick mass sorting

Elemental composition assignments can be unambiguous only up to ~400 Da. Sorting based on the Kendrick mass scale simplifies elemental composition assignment from an overwhelming number of possible combinations of elements for higher-mass ions. This scale sets the mass of the CH_2 radical to 14.0000 mass units.

$$Kendrick\ mass = IUPAC\ mass \times \frac{14.00000}{14.014565} = IUPAC\ mass \times \frac{CH_2\ value\ on\ Kendrick\ scale}{CH_2\ value\ on\ IUPAC\ scale} \quad \textbf{(1)}$$

Compounds are sorted into homologous series according to alkylation, classes (numbers of heteroatoms), and double bond equivalents (DBE). Double bond equivalents are the number of rings plus double bonds in a molecule and are a direct measure of aromaticity. Compounds with the same heteroatom composition and DBE but different numbers of CH_2 units will differ in Kendrick mass by multiples of 14.0000 Da. The advantage of converting to KM is that members of the same homologous series are easily recognized because they share the same Kendrick mass defect (KMD) which is unique to a particular series. The following equation is used to calculate the Kendrick mass defect:

$$Kendrick\ mass\ defect = nominal\ Kendrick\ mass - exact\ Kendrick\ mass \quad \textbf{(2)}$$

where nominal Kendrick mass is the Kendrick mass rounded to the nearest integer (Kendrick, 1963). The unambiguous elemental composition assignment of a single low-mass member of a series can be extended to identify all other members of the same series. Therefore, Kendrick mass sorting allows for the assignment of masses much higher than otherwise possible based solely on mass-accuracy.

2.8.3 Principles of operation

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) is based on the principles of cyclotron movement, or the circular motion of ions in a strong magnetic field (Marshall et al., 1998). Figure 7 shows a diagram of the instrument used for this work. Ions produced at atmospheric pressure are introduced into a series of octopoles. The magnetic field causes ions to resonate at their cyclotron frequency, which is directly related to their mass-to-charge ratio (*m/z*). After accumulation in the octopoles, the ions are pulsed to the ICR cell where an oscillating electric field is applied. This excitation serves to accelerate ions into a larger orbital radius easily measured by the detector plates on opposite sides of the cell, or to increase the cyclotron radius to a size larger than the radius of the ion trap so that the ions are ejected from the instrument.

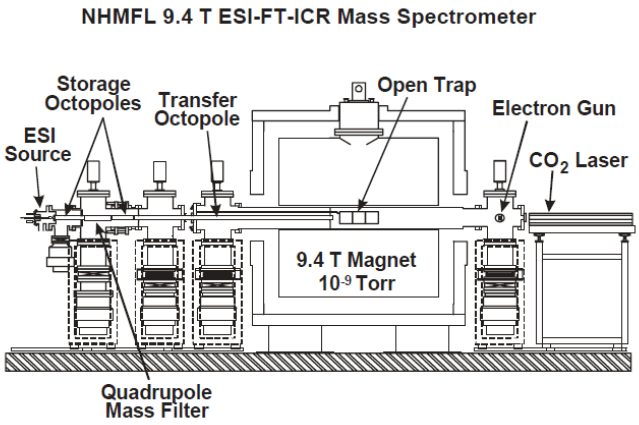

Figure 7: Diagram of the 9.4 Tesla FT-ICR mass spectrometer at the National High Magnetic Field Laboratory (McKenna et al., 2009).

The detector measures the cyclotron frequency of all of the ions in the trap and uses a Fourier transform to produce a frequency spectrum. The measured frequencies can be converted

to a mass spectrum using the relationship between frequency and *m/z*. Figure 8 illustrates this process.

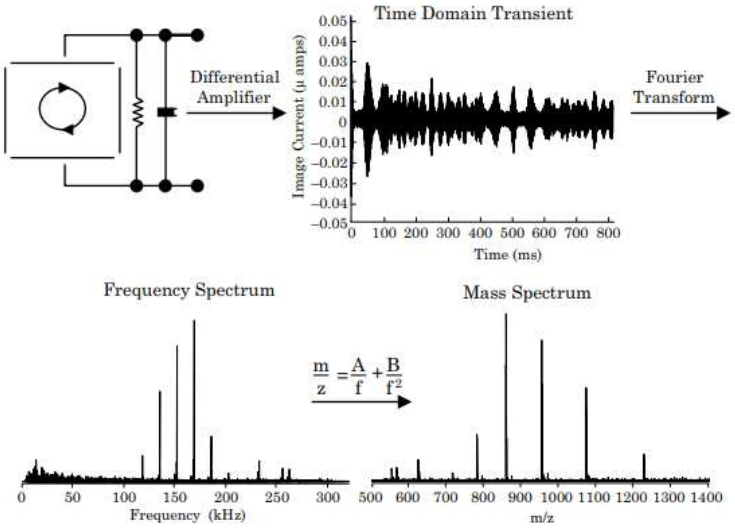

Figure 8: In FT-ICR MS analyses, ion cyclotron motion (a time-domain) is first converted to a frequency spectrum using a Fourier-transform and then to a mass spectrum using the relationship between ion frequency and ion *m/z* (Stanford et al., 2006; Purcell, 2009).

The simultaneous detection of all ion frequencies during the detection interval increases the signal-to-noise ratio. FT-ICR MS resolution can be improved by either increasing the strength of the magnet or by increasing the detection duration (Reddy et al., 2013).

2.8.4 Ionization modes

Through the generation of ions (e.g., $[M+H]^+$ and $[M–H]^–$, where M is a neutral compound) by electrospray ionization (ESI), FT-ICR MS can characterize low-volatility species previously inaccessible by GC-based techniques. Fenn and Zhan (2000) were the first to apply ESI to ionize the most polar species in petroleum samples (containing N, S, and O). Crude oil samples are complex enough without the excessive fragmentation produced by GC to further complicate the spectra. ESI generates minimal ion fragments so each neutral analyte only

corresponds to one signal in the mass spectrum. Ions are created at atmospheric pressure before being introduced into the instrument in the gas phase.

Electrospray ionization (ESI) targets polar molecules by selectively ionizing functional groups that readily protonate or deprotonate in an electric field. The selectivity for acidic or basic species can be controlled by applying a positive or negative charge to the capillary needle and counter electrode. In positive ion mode, a positive potential is applied to the needle, which repels positive ions in solution toward the capillary tip and produces a positively charge droplet that is sprayed into the mass spectrometer. Positive ESI ((+)ESI) is used for basic analytes that form cations in solution, including species with ketone- or aldehyde-containing functional groups (shown in Figure 9).

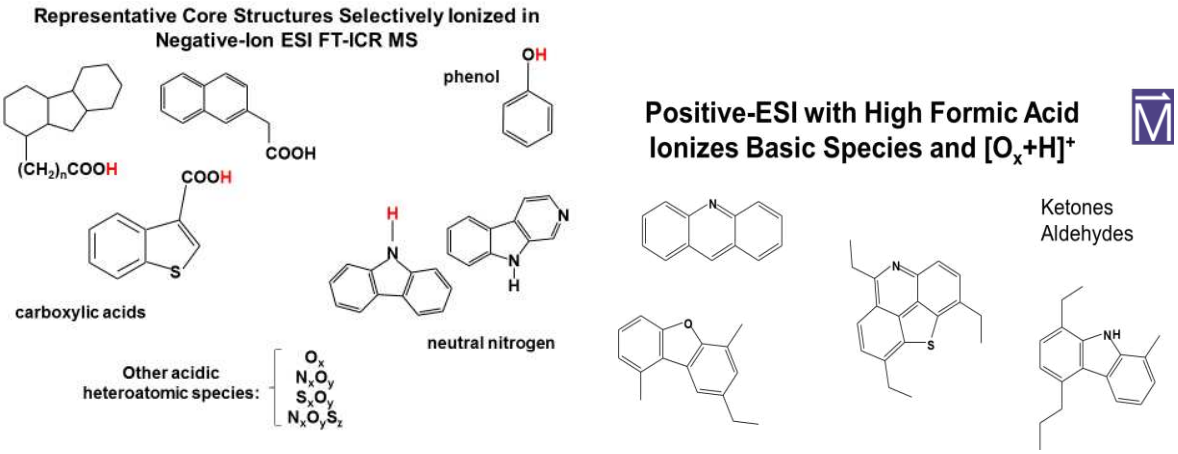

Figure 9: Examples of functional groups selectively ionized by negative-ion ESI (left) and positive-ESI (right). (Mag Lab website)

In negative ion mode, ESI preferentially ionizes molecules with acidic functional groups that readily carry a negative charge through deprotonation. Functional groups targeted by negative mode ESI include carboxylic acid, neutral nitrogen, phenol, and other heteroatomic species such as O_x, N_xO_y, S_xO_y, and $N_xO_yS_z$, as seen in Figure 9. As a result of ionization

efficiency, determined by the ability of a functional group to stabilize the charge, carboxylic acids are preferentially ionized relative to alcohols or nitrogen-containing compounds and leads to competitive effects when compounds with both types of functional groups are present in an extract. The ionization process is illustrated in Figure 10.

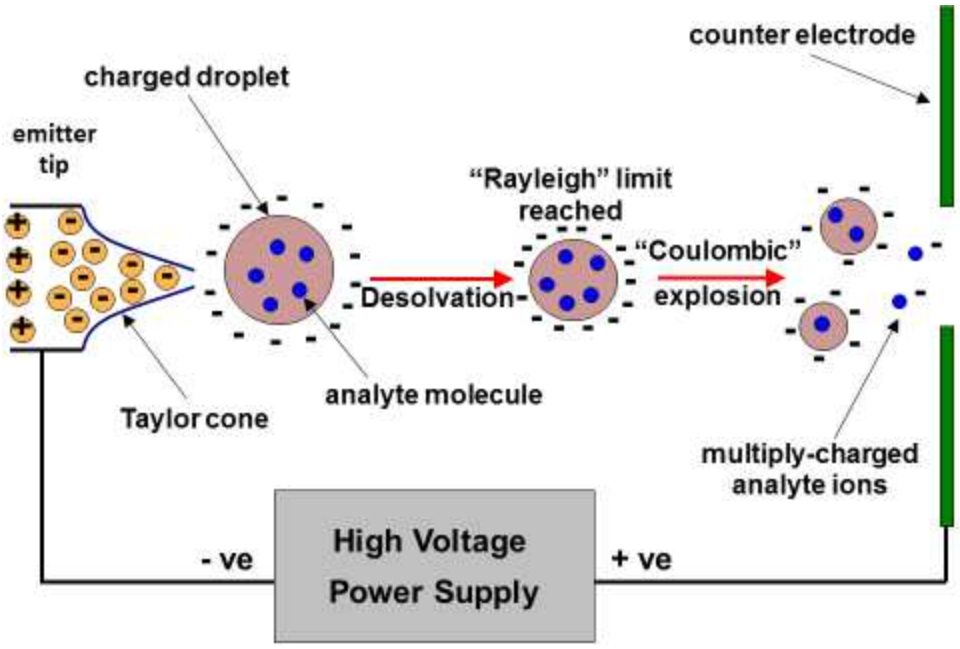

Figure 10: Schematic diagram of negative-mode electrospray ionization (Mag Lab website)

A major limitation of ESI is its inefficient ionization of nonpolar species, such as non-polar aromatic hydrocarbons, thiophenes, and furans. Atmospheric pressure photoionization (APPI) can ionize both polar and nonpolar compounds simultaneously. Figure 11 shows a schematic diagram of the APPI mechanism.

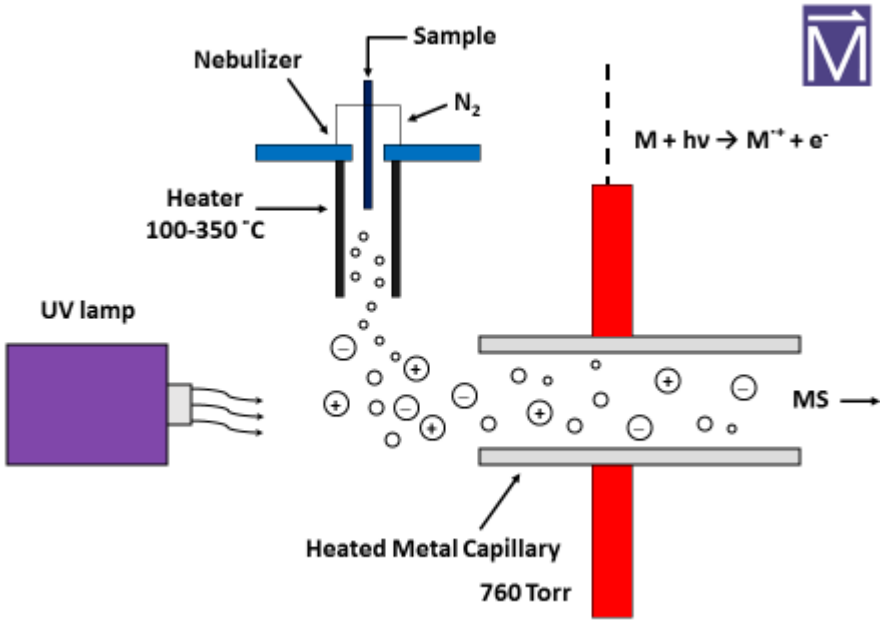

Figure 11: Schematic diagram of atmospheric pressure photoionization (Mag Lab website)

In APPI, the gas phase sample molecules enter the ionization chamber at atmospheric pressure and are exposed to ultraviolet light from a krypton lamp. The lamp emits photons with a specific energy level (10 eV) that is high enough to ionize the target molecules, but not high enough to ionize air and other background molecules, ensuring that only analyte molecules enter the mass spectrometer. Analyte molecules can be ionized through two different pathways. Direct photoionization involves a one step process where photons (hv) excite the analyte molecule (M) and cause the loss of an electron (e^-). The resulting radical cation ($M^{+\bullet}$) enters the mass spectrometer. Solvent or molecules contribute to the indirect photoionization of sample molecules. Photons excite solvent molecules, which vastly outnumber sample molecules. Collisions between solvent and sample molecules lead to a solvent-assisted chemical ionization process where the solvent molecule donates a proton (depicted by an H in the equation below) to the analyte molecule.

$$M + S + h\nu \Rightarrow [M + H]^+ + [S - H]^- \text{ (solvent dependent)} \quad \textbf{(3)}$$

This process produces two types of ion for each sample molecule – a radical cation ($M^{+\bullet}$) and a protonated molecule ($[M + H]^+$). Dopant-assisted APPI increases ionization efficiency for a diverse range of compounds. Adding a dopant, commonly toluene, also produces two types of analyte ions. In this method (explained in Figure 12), UV photons directly ionize dopant molecules, and the dopant molecules participate in the proton transfer reactions instead of the solvent.

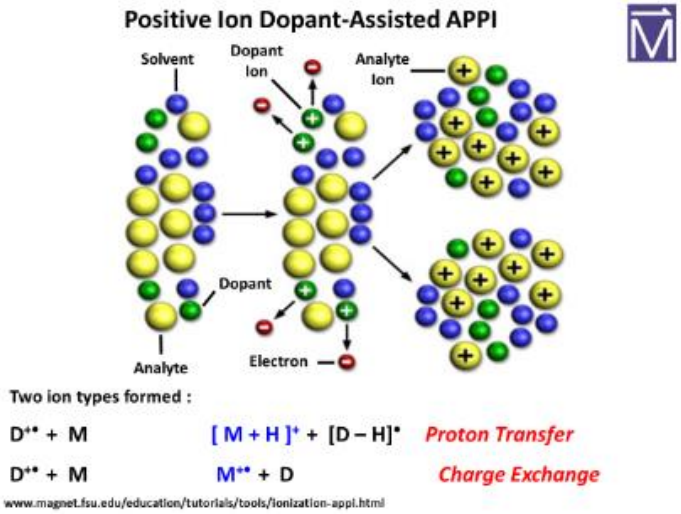

Figure 12: Diagram illustrating the mechanism of positive ion dopant-assisted APPI (Mag Lab website)

Like ESI, APPI is also a soft ionization technique that ionizes petroleum compounds with little to no fragmentation.

The ion efficiency bias makes FT-ICR an inherently non-quantitative tool. Due to selective ionization, no single ionization method can produce ions from neutral analytes with equal efficiency. The chosen ionization mode depends on the analytes of interest. Instead, FT-ICR is a valuable semi-quantitative tool for comparing changes in composition relative to total

assigned peaks within single sample, rather than between different samples. Additionally, ionization efficiency for one species can be greatly affected by the presence of other species. This matrix effect makes it difficult to relate the observed ion relative abundances to the relative abundances of their precursor neutral molecules in the original sample.

2.8.5 Data reduction

The thousands of elemental compositions assigned by FT-ICR MS enables a more comprehensive understanding of the sample composition, but it also challenges data interpretation. The next challenge becomes transforming data tables containing a plethora of elemental formulas into meaningful graphical formats. Reducing vast datasets involves using plots of the output parameters reported by PetroOrg® software developed at the National High Magnetic Field Lab (NHMFL) – including heteroatom class, abundance-weighted DBE and abundance-weighted carbon number – as well as statistical methods like principal component analysis (PCA). Comparing structural trends among different samples can help elucidate weathering trends and differentiate types or sources of the parent oil.

A plot of DBE versus carbon number plot for each heteroatom class displays all of the compounds assigned to the class in a single graph. The x-axis reflects molecular size and the y-axis reflects degree of alkylation or aromaticity. The color-coding of these "isoabundance contour plots" corresponds to the relative intensity of the compounds, after normalization to the most abundant peak in the heteroatom class. These plots can highlight structural changes when comparing the same heteroatom class for two different samples. The proximity of compounds to the planar stability line or planar aromatic limit indicates their degree of aromaticity or alkylation. The planar stability line is a straight line connecting the maximum possible DBE for each carbon number in a planar polycyclic aromatic molecule, representing the theoretical limit

between a planar and bowl-shaped structure. As Figure 13 demonstrates, DBE versus carbon number plots can differentiate between SARA (saturates, aromatics, resins, asphaltenes) fractions based on their unique slopes and y-intercepts of the planar stability lines (Cho et al., 2011).

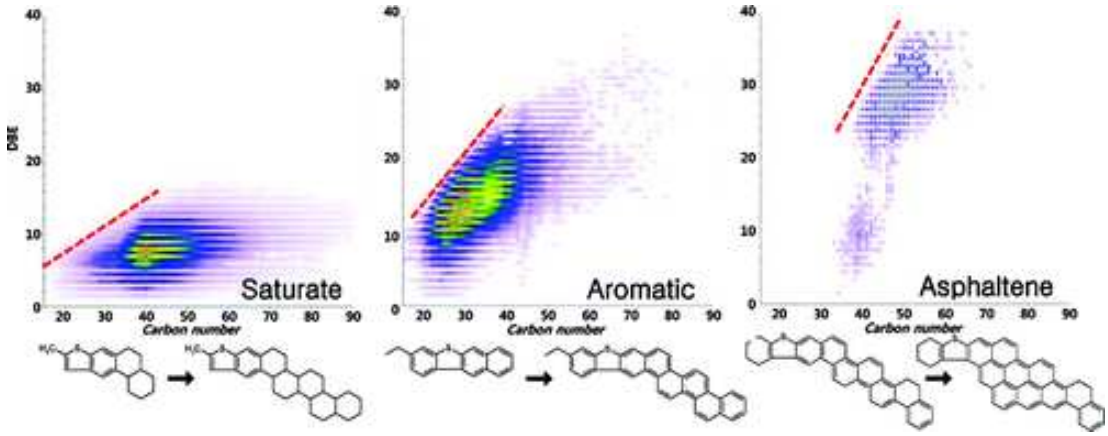

Figure 13: In DBE vs. carbon number isocontour plots, different hydrocarbon classes have characteristic planar stability line slopes and y-intercepts (Purcell et al., 2010).

Other variables commonly plotted for oil industry applications are carbon number on the x-axis and hydrogen deficiency in terms of H:C ratios on the y-axis. These plots are also color-coded for relative abundance for each heteroatom class. The H:C ratio correlates to a distinction between aliphatic (>1.5) and more condensed ring aromatic compounds (<1).

Another graphical representation of molecular compositional information is a "van Krevelen" plot, with O/C ratio on the x-axis and H/C ratio on the y-axis. Each peak in the mass spectrum is represented as a point in the van Krevelen diagram. Species of the same oxygen content and hydrogen deficiency form diagonal lines. The van Krevelen plot was originally applied to represent bulk elemental compositions in dissolved organic matter (DOM) samples and to trace potential source rocks from oil and gas products (van Krevelen, 1950). van Krevelen plots have been used to characterize extent of oil degradation and identify sources of spilled oil.

2.8.6 Applications of FT-ICR MS

The ability of negative-ESI FT-ICR MS to shed new light on the distribution of polar compounds in the O- and N-containing heteroatom classes has valuable potential for forensic applications and source tracking of different fuels or types of products. Identifying the source material is critical for effective remediation of oil spills because chemical and physical properties of the parent hydrocarbon compounds dictate weathering trends and partitioning behavior. Stanford et. al. (2006) and Teräväinen et. al. (2007) characterized the molecular composition of vacuum gas oil distillation cuts and crude oil distillation fractions and found that different fractions had unique compositional profiles.

Combining FT-ICR MS data with statistical methods – specifically principal component analysis (PCA) – can expose compositionally dependent weathering trends that are not easily recognized by visual inspection of data. Corilo et al. (2013) applied PCA to negative-mode ESI results to differentiate two different sources of spilled weathered oil. Relative abundances for heteroatom classes and DBE distributions were used as input variables for the PCA. The clustering of samples shown in the scores plot traced the detected compounds to distinct sources. The resulting loadings plot highlighted heteroatom classes primarily responsible for the distinction between potential release sources, based on corresponding large positive and negative values along the principal component axes. PCA also enabled identification of polar petroleum markers that are highly resistant to biodegradation. In this study, naphthenic acids (O_2 class) were recognized as potential polar analogues of hopane biomarkers used to understand weathering and degradation trends.

FT-ICR MS has advanced the analytical capabilities of many different types of complex mixtures beyond petroleum contamination. Like petroleum mixtures, the complexity of natural

organic matter (NOM) and dissolved organic matter (DOM) presents an analytical challenge. Molecular characterization by FT-ICR MS helped reveal structural changes in DOM due to adsorption to soil (Avneri-Katz et al., 2017). From the molecular formulas assigned, calculated degrees of unsaturation, normalized DBE/C (Koch and Dittmar, 2006), and H/C and O/C ratios were used to identify aromatic and condensed aromatic structures. Detected DOM compounds were sorted into structurally similar biogeochemical classes (lignin, lipids, carbohydrates) based on characteristic H/C and O/C ratios for each of the classes. FT-ICR MS analysis shed new light on preferential adsorption of polyphenols to mineral surfaces.

FT-ICR MS is also becoming an increasingly valuable tool for the analysis of complex biological mixtures of proteins and lipids. FT-ICR MS extends the analytical coverage of polar glycol- and phospholipids (He, 2009l Jiang, 2017). Analyzing biological samples often requires extensive sample clean up and chromatographic separation procedures, but FT-ICR MS has been used to discover protein biomarkers in serum and plasma samples, advancing clinical proteomics.

3. MATERIALS AND METHODS

3.1 Field site descriptions

The following section provides an overview of the four hydrocarbon-contaminated sites from which sediment samples for this study were taken. Samples were collected by Center for Contaminant Hydrology staff and students.

3.1.1 Site A

A former petroleum bulk storage (PBS) terminal in the northeastern United States, this site had been reported to store hydrocarbon fuels. A large freshwater tidal river borders the site to the west, and sheens attributed to historic petroleum spills appear on the river bank. An oleophilic bio-barrier (OBB) was installed as an experimental sheen remedy (Chalfant, 2015; Tochko, 2018). The OBB consisted of an oleophilic plastic geocomposite skeleton covered on the top and bottom by a geotextile, and it was designed to intercept and retain petroleum hydrocarbons in ground- and surface water, while delivering oxygen to support microbial degradation of the hydrocarbon contaminants.

Groundwater flow at this site is perpendicular to the river. Due to tidal fluctuations of approximately 1.5 m, groundwater discharges to surface water at low tide, and the OBB was submerged at high tide. The geology of the sampled sediment layer was fine-to-coarse sand with fine-to-coarse gravel, most likely from river dredge. About 6 m below ground surface is clay with minor amounts of silts; this clay layer acts as an aquitard (Arcadis, 2011).

In October 2017, the OBB was removed for analysis. Sediment samples were collected using a hand trowel from immediately below the OBB (0-10 cm) and from deeper below the OBB (10-20 cm). The upper sediment sample discussed in this thesis was given the label RU and

the two lower sediment samples were labeled RL3 and RL5. An apparent background sediment sample was also collected from farther up the bank away from the river, taken from 3-6 inches below the surface. This background sample (referred to as “Background”) was assumed to be uncontaminated based on visual inspection and a lack of petroleum odor. Samples were wrapped in aluminum foil, sealed in individual plastic re-sealable bags, placed in a cooler with dry ice and shipped overnight to Colorado State University (CSU), where they were stored in a -80 °C freezer until analysis.

3.1.2 Site B

This historic wood treating facility located on the eastern side of a sound released creosote-based wood preserving chemicals (including solvents, gasoline, mercury and other heavy metals), contaminating soil, groundwater, and sediments in the harbor. The EPA designated the area a Superfund site in 1987 (US EPA, 2015). Since then, active remedies have involved excavation of sludge and recovered oils, capping and sealing of on-site production wells, steam injection and construction of a sheet pile wall around the former process area. The groundwater level at the site varies with the tide between approximately 5 to 15 feet below ground surface (bgs). Sediments consists of both fill and native materials overlaying marine sand containing interbedded gravel, silt, and clay. Ground water flows from the upland area at the southern boundary of the site north toward the harbor and is also affected by the tidal cycle. Active ferry and marina operations in the harbor may contribute to low background sources of oil in the area.

The study area included two seep locations on the beach beyond the sheet pile wall. Soil sediment samples were collected along a horizontal line from three different distances away from the wall in the North Shoal region, labeled T1 (referring to the transect closest to the wall), T2

(transect midway between the shore and the wall), and T3 (transect farthest from the wall and closest to the shore). An OBB test site was installed farther inland from the sheet pile wall; the samples labeled 6" and 12" were collected from sediment 6 inches and 12 inches below the mat, respectively. The sample labeled "Beach" was taken upgradient of the OBB study site. Sediment samples were shipped to CSU on dry ice.

3.1.3 Site C

The site sampled was a former refinery in the United States, and samples were collected from the LNAPL smear zone by cryogenic coring. Core samples were drilled using a hollow-stem auger and were flash-frozen with liquid nitrogen before being removed from the ground. This method improved recovery of soil and pore fluids and preserved microbial activity and geologic properties (Kiaalhosseini et al., 2016). Laser induced fluorescence (LIF) profile data indicated a vertical LNAPL zone of approximately 10 feet. According to water table elevation measurements, the water table fluctuates but the smear zone is submerged for most of the year. Along the plume, groundwater flows from north to south

Two sampling events were performed. In August 2016, a core was collected from the upgradient, leading edge of plume.Soil cores were shipped on dry ice to the lab where they were cut into disc-shaped segments at known depth intervals. Two segments of the frozen core were analyzed from this site, representing two distinct hydrogeologic zones from a vertical transect: an anoxic, low k zone collected from 27 feet below ground surface (bgs) and a transmissive zone from 31.6 feet bgs. The anoxic, low k zone sample was classified as a silt layer with limestone cobbles and the oxic, transmissive zone sample was a sand.

3.1.4 Site D

Another former refinery in the Midwestern US selected as a field site is located along the southern shore of a major river. The source of the LNAPL release is an upland tank farm that closed in the 1980's. A north-flowing tributary bisects the site, and groundwater mainly flows relatively slowly (ranging from 7.2 to 21.2 feet per year) toward this creek, with some flow toward the river in the northern portion. The lithology above the bedrock is mainly clay and silty clay with discontinuous bands of sand. These intermittent zones of coarser material and macropores are believed to be preferential migration pathways for hydrocarbons toward the creek. Samples originated from a silt (~7.5 ft bgs) and a sand (~30 ft bgs) and were collected by cryogenic coring.

3.2 Sample preparation

Samples collected by cryogenic coring (from Sites C and D) were previously cut while still frozen into hockey puck-like disks. Disks were quartered and one quarter was used for hydrocarbon analysis. At the time of analysis, the frozen samples were thawed and then extracted with HPLC-grade toluene. Subsamples of roughly 10 g (including any frozen water) were extracted in 15 ml of solvent in a wide mouthed glass jar and mixed on a mechanical shaker for at least 2 hours.

3.2.1 Justification for solvent choice

Extraction experiments were conducted to optimize the solvent choice. The solvent mixtures tested included methanol, methyl *tert*-butyl ether (MTBE), hexane, hexane:acetone (70:30), and toluene. Preliminary work involved the analysis of nonpolar and polar petroleum-derived compounds sorbed onto oleophilic bio-barrier (OBB) mats by GC-FID and GC-MS.

The most common solvent choice for extracting oil-contaminated soil samples is methylene chloride. However, non-solvent peaks were detected in the GC-FID chromatogram, and polymer compounds were detected by GC-MS after extraction of an uncontaminated OBB sample with DCM, indicating that DCM dissolved the high density propylene (HDPE) interior skeleton of the OBB. Other solvents did not degrade HDPE.

Other criteria for the solvent choice were the efficient extraction of both polar and non-polar petroleum-related compounds. Extraction efficiencies of the solvents were assessed by spiking the OBB sample with a known amount of surrogate compound, and then measuring the percent recovery after GC-FID analysis and quantification. Triplicate samples were analyzed.

A diesel range organics (DRO) standard alkane mixture (EPA/Wisconsin, Restek) was used as a surrogate for nonpolar compounds, and decanoic acid was used as the surrogate for polar compounds. Calibration curves with standard concentrations ranging from 50 mg/L to 300 mg/L were used for quantification of DRO and decanoic acid spikes. Table 1 shows the percent recovery of the spiked surrogate compounds (400 μg/ml of decanoic acid and 100 μg/ml of DRO) after extraction of the OBB.

Table 1: A comparison of hydrocarbon extraction efficiencies for five different solvents for non-polar and polar surrogate compounds.

Solvent	Nonpolar (DRO)		Polar (Decanoic acid)	
	Average % Recovery	Standard Deviation	Average % Recovery	Standard Deviation (%)
MeOH	79	0.0524	34	12.38
MTBE	110	0.0968	62	1.33
Hexane	86	0.0776	80	1.70
Hexane/Acetone (70/30)	99	0.0948	50	6.27
Toluene	104	0.0574	108	7.07

Toluene was chosen as the solvent for all extractions based on the high percent recovery of the

polar and nonpolar spikes. Increasing the shaking time from 2 to 3 hours did not affect the measured recovery in DRO or decanoic acid concentrations for the spiked OBB samples. The sample-to-solvent ratio was chosen so that the sample sediment was fully covered while minimizing the solvent volume used. Natural organic matter (NOM) and dissolved organic matter (DOM) are insoluble in toluene, which eliminated interference from non-petroleum hydrocarbons (Dongbao, 2010).

3.3 Analysis

3.3.1 Gas chromatography (GC-FID and GC-MS)

After shaking the sample jars, extracts were transferred to GC vials and analyzed by GC-FID (Agilent Technologies 6890N Gas Chromatograph, Santa Clara, CA) equipped with a Flame Ionization Detector and a Restek Rtx-5 column (30 m length x 0.32 mm inner diameter x 0.25 µm film thickness, Bellfonte, PA) and GC-MS (Agilent Technologies 6890N Network Gas Chromatograph with an Agilent 5973 Network Mass Selective Detector.) Peaks were identified by spectral matching against the NIST mass spectral library.

Toluene extracts from Site C were also analyzed for (semi-)volatile organics at the Colorado State University Chemistry department's Central Instrument Facility by an Agilent 6890 gas chromatograph equipped with an Agilent 5973N Mass Selective Detector using a TG-SQC TraceGold GC column (15 m length, 0.25 mm internal diameter, 0.25 µm film thickness, ThermoFisher) and the following oven temperature program: 80 °C (held for 2 min), then increased at 15 °C min^{-1} to 315 °C (held for 5 min). Ultra-high purity helium was used as a carrier gas at a constant flow rate of 1 mL/min. Sample injections were 2 µL. Injector temperature was set at 285 °C. The GC-MS transfer line temperature was maintained at 320 °C

and the ion source temperature was held at 230 °C. The mass spectrometer was operated in electron ionization mode (70 eV). Mass spectra were recorded in full scan mode (*m/z* 45-600). Identified species were determined by comparison with spectra in the NIST library, 2011 edition (Gaithersburg, MD), with a match factor above 800.

3.3.2 Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS)

Portions of the extracts to be analyzed by FT-ICR MS were added to GC vials and dried under nitrogen gas, before being reconstituted in toluene, methanol, and tetramethylammonium hydroxide (TMAH) upon injection into the FT-ICR at the National High Magnetic Field Laboratory (NHMFL) in Tallahassee, FL.

Samples were analyzed by negative and positive mode electrospray ionization (ESI) and atmospheric photoionization (APPI) on a custom-built 9.4 T Fourier transform ion cyclotron resonance mass spectrometer at the NHMFL (Kaiser et al., 2011).

3.3.2.1 Sample extraction

All solvents were HPLC grade (Sigma-Aldrich Chemical Co., St. Louis, MO). Prior to direct injection into the mass spectrometer, ~1 mg of neat oil or weathered sample was diluted with toluene to yield a stock solution (1 mg mL^{-1}) that was further diluted to final concentration of 250 μg mL^{-1} with equal parts (v/v) methanol spiked with 0.5% v/v tetramethylammonium hydroxide (TMAH, 25% by weight in methanol) prior to FT-ICR MS analysis. A modular ICR data acquisition software package (Predator) was used for data collection (Blakney et al., 2011).

3.3.2.2 Mass spectral calibration

Internal calibration of the spectrum was possible with the use of homologous series which repeat by 14.01565 Da (or CH_2 units). Samples were manually calibrated on multiple homologous series because no single series spanned the entire *m/z* range. For samples that did

have a homologous series that spanned the entire *m/z* range so a "walking calibration" was applied to these spectra (Savory et al., 2011).

Measured IUPAC mass was converted to Kendrick mass to sort compounds that differ in mass by 14.01565 Da (mass of CH_2). Mass spectral peaks with a signal-to-noise ratio greater than six were assigned elemental compositions with MIDAS software. For elemental formula assignment, C, H, N, O, and S were the only elements considered. An rms error below 1 ppm was also an assignment criterion. If no chemical formula matched an *m/z* value within the allowed error, the peak was not included in the list of elemental formulae and was sorted into the "unassigned" category.

3.2.2.3 Data reduction

In the PetroOrg© software program developed at NHMFL, compounds with the same heteroatom content but different degree of alkylation were grouped together into heteroatom classes to enable data visualization for compositional comparison between samples.

Isoabundance contour plots were created for each heteroatom class. These color-coded plots display carbon number on the x-axis, reflecting the degree of alkylation (CH_2 units), and double bond equivalents (DBE), representing hydrogen deficiency, on the y-axis. Relative abundance on the z-axis is represented by a color scale (described by the legend) to clearly show compositional changes between samples. All plots are normalized to the most abundant peak in each heteroatom class.

Coding in RStudio simplified data visualization by creating diagrams comparing variables that highlighted clear trends in the composition of the species present. For example, Van Krevelen diagrams plot the degree of oxygenation (O:C ratios) vs. degree of aromaticity (H:C ratios); more reduced species lie farther to the left along the x-axis and more oxidized

species are farther to the right, while more saturated compounds lie higher on the y-axis and more aromatic compounds are at the bottom of the plots.

3.3.3 Elemental analysis

Five samples (listed in Table 2) were shipped to Galbraith Laboratories, Inc. (Knoxville, TN) for elemental analysis. Samples were dried at 60 °C for 24 hours before being ground for analysis.

Table 2: Description of the samples analyzed for elemental composition.

Sample Name	Site	Description
RU	A	Upper sediment layer (0-10 cm below OBB)
RL5	A	Lower sediment layer (10-20 cm below OBB)
Background	A	Sediment from 3-6 in below surface Assumed to be uncontaminated
Beach	B	Assumed to be uncontaminated
T1	B	Sediment closest to sheet pile wall

Tests were conducted to determine the content of carbon, nitrogen, oxygen, and sulfur by combustion and pyrolysis.

4. RESULTS AND DISCUSSION

The following chapter presents the findings from multiple analytical approaches used to characterize the complex mixtures of oxidized polar transformation products in the weathered hydrocarbon-contaminated samples. The analyses included elemental analysis, GC-FID and GC-MS, and FT-ICR MS. Advantages and disadvantages of each analytical technique will be discussed. This section will also describe the long-term environmental and regulatory implications of the compositional characterization of these samples.

4.1 GC-FID

Gas chromatography coupled to a flame ionization detector (GC-FID) is the most basic level of analysis for hydrocarbon-impacted samples. The majority of standard methods for quantifying petroleum hydrocarbons use GC-FID to report total petroleum hydrocarbons (TPH), which is calculated by integrating the total peak area in the chromatogram and comparing it to a calibration curve – typically within a retention time range corresponding to gasoline range organics (GRO) and diesel range organics (DRO). In this experiment, sediment samples were extracted using toluene. Toluene does not dissolve natural organic matter, so only (semi-)volatile petroleum-derived hydrocarbons are detected in these chromatograms (Fu et al., 2010).

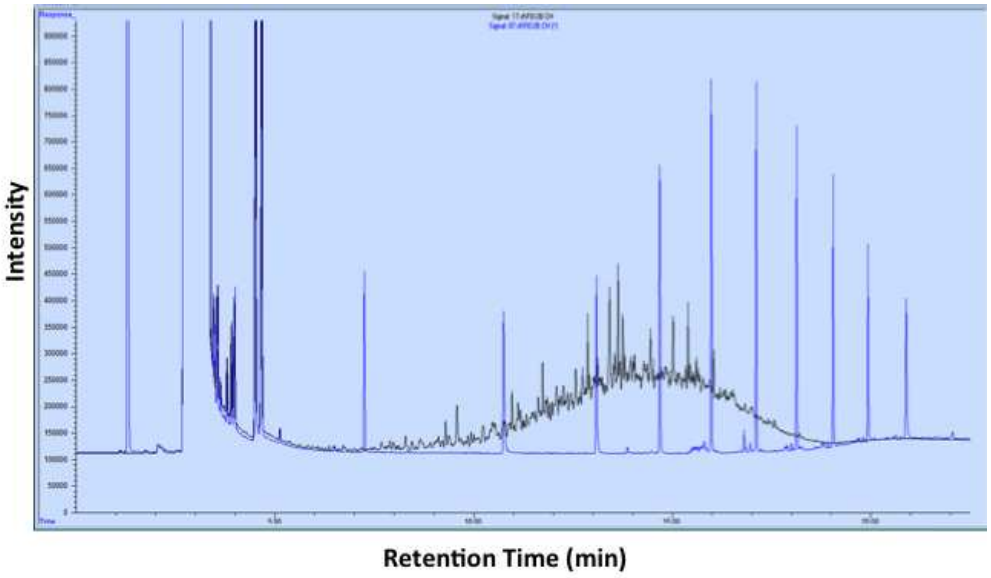

Figure 14: GC-FID chromatograms of the lower sediment zone sample from Site A (black) and 300 mg/L DRO standard (blue).

Figure 14 shows overlain chromatograms of a lower sediment zone sample from Site A and a 300 mg/L DRO standard. Based on the DRO calibration curve, the calculated concentration of hydrocarbons in the sample was 19 mg DRO/kg sample. Figure 14 demonstrates the ambiguity of individual peaks contributing to the unresolved complex mixture (UCM) typical of weathered oil; the UCM peak area was included in the calculated sample concentration. In contrast, the *n*-alkane standards appear as clean, narrow peaks. The only way to identify an unknown peak in a GC-FID chromatogram is to compare it to the retention time of a standard, but unambiguous identification of compounds in complex mixtures is impossible due to incomplete separation and lack of other compound-specific descriptors such as (accurate) mass-to-charge ratios. GC-FID analyses can be used for subjective, qualitative comparisons between samples or to generate a bulk parameter, the TPH value. TPH values for samples from all four sites were determined by GC-FID and are presented in Table 3. These results and a strong hydrocarbon odor suggested

that RL5 from Site A and the oxic transmissive sample from Site C were the most heavily contaminated samples.

Table 3: GC-FID-based TPH results

Sample Name	TPH (mg DRO/kg sample)
Site A	
RU	25
RL3	2580
RL5	2230
Background	189
Site B	
T1	54
6 inch	30
12 inch	360
Beach	15
Site C	
anoxic low k	930
oxic transmissive	3200

The sediment sample from Site A contained contaminant peaks that fall within the DRO range (C_{10} to C_{24}). The GC-FID chromatogram (Figure 15) from the anoxic low k zone sample from Site C showed a different distribution of contaminant peaks with retention times that are shorter than peaks in the Site A sample.

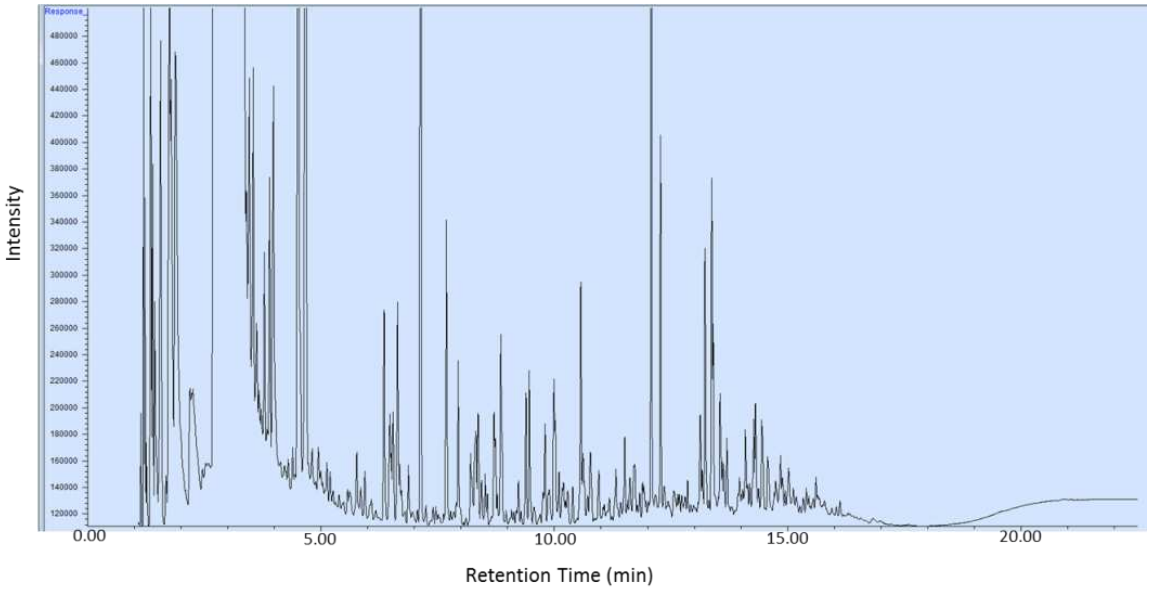

Figure 15: GC-FID chromatogram of Site C anoxic low k zone sample.

The shift in retention times indicates a difference in the boiling points of the compounds present in these samples. The Site A sediment contains higher-boiling petroleum species than the Site C sediment. The absence of *n*-alkane peaks in both the Site C and Site A samples is consistent with typical weathering patterns; the *n*-alkane fraction is the first to evaporate or degrade. However, GC-FID did not provide any information regarding oxygenated hydrocarbons.

4.2 GC-MS

Gas chromatography coupled to a mass spectrometer (GC-MS) yields some specific elemental information and provides a more detailed profile of petroleum hydrocarbons than GC-FID. This technique is still limited to volatile compounds and very few heteroatomic compounds are detected. Using spectral and retention index matching against the NIST library database for peak assignment, and excluding assignments with a quality/match factor below 800, 40-50 compounds were identified in the GC-MS chromatograms for the samples from Site C. The

quality or match factor describes how well the unknown peak matched the NIST library spectrum. The peaks with the highest match factors were labeled on the chromatogram for the anoxic, low k sample presented in Figure 15. Note that chromatographic columns and temperature gradients differed between the analyses shown in Figure 15 and Figure 16, and that analyte retention times and peaks are thus not directly comparable.

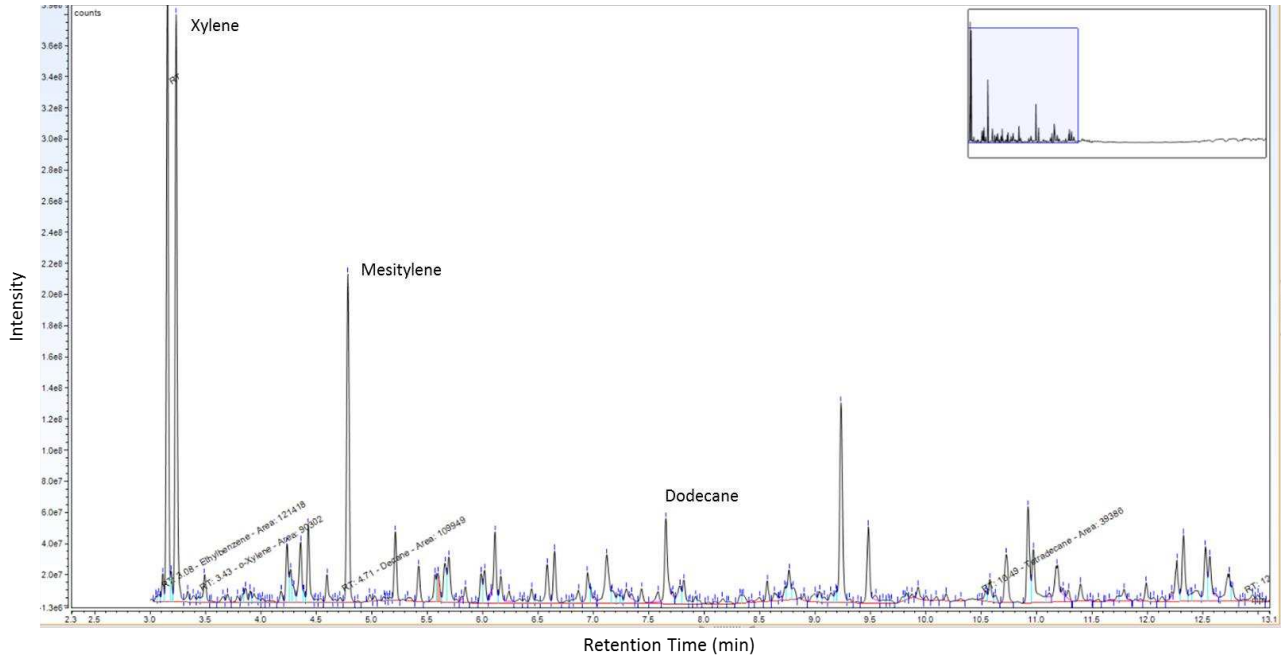

Figure 16: GC-MS chromatogram for the Site C anoxic, low k sample. The labeled peaks had the highest match factors.

The majority of the species identified were toxic PAH's – irritants and carcinogens – but only three oxygenated compounds (and no N- or S-containing molecules) were detected. In this particular sample, the most abundant compounds identified were mesitylene, 1-methylnaphthalene, and 1,3-dimethylnaphthalene. These structures are depicted in Figure 17**.**

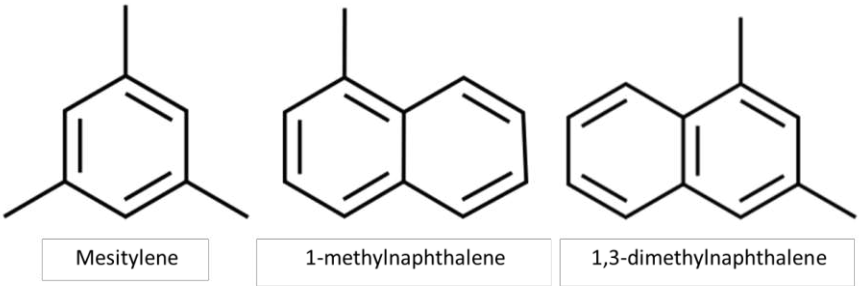

Figure 17: Most abundant species in the GC-MS chromatogram for the Site C anoxic, low k sample.

The only heteroatoms identified contained oxygen atoms: 1-methoxymethyl-4-methylnaphthalene, 1,3,5-Cycloheptatriene-1-carbaldehyde, and 1(3H)-isobenzofuranone, shown in Figure 18.

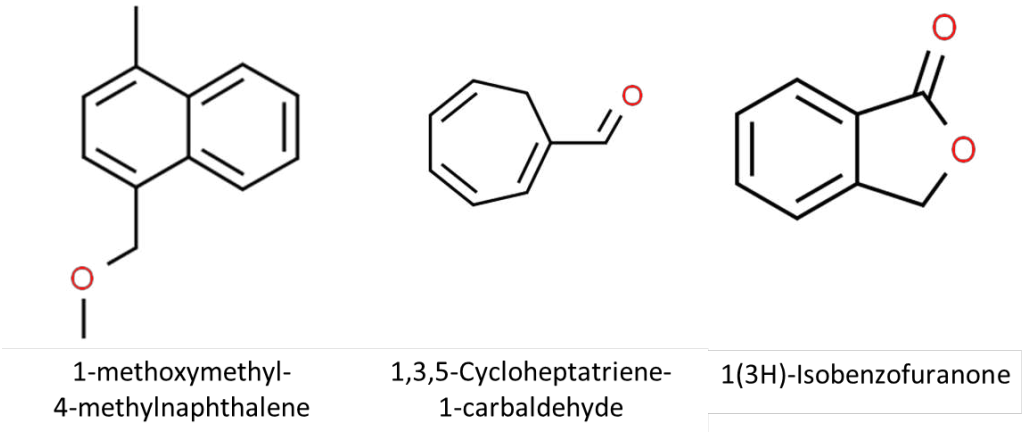

Figure 18: The three "oxyhydrocarbons" identified in the GC-MS chromatogram for the Site C anoxic, low k sample.

The peaks corresponding to the polar oxygenated species had low intensities in the chromatogram. Table 4 shows a complete list of the peaks identified in this sample and the area of the peaks, which was directly related to the intensity in the chromatogram.

Table 4: Peaks identified by GC-MS present in the Site C anoxic, low k zone sample. All of the molecules detected contained no more than 13 carbon atoms.

Retention Time (min)	Compound	Formula	Peak Area
3.05	Heptane, 2,3-dimethyl	C9 H20	40856
3.08	Heptane, 2,3-dimethyl	C9 H20	65062
3.11	Heptane, 2,3-dimethyl	C9 H20	196903
3.16	Ethylbenzene	C8 H10	7871842
3.2	octane, 3-methyl	C9 H20	162091
3.23	o-xylene	C8 H10	6976749
3.43	Cyclohexane, 1-ethyl-4-methyl-, cis	C9 H18	34652
3.9	octane, 2,6-dimethyl	C10 H22	76816
4.22	1-Formyl-1,3,5-cycloheptratriene	C8 H8 O	432214
4.27	octane, 2,5,6-trimethyl	C11 H24	339938
4.43	Mesitylene	C9 H12	609543
4.59	Benzene, 1-ethyl-2-methyl-	C9 H12	170481
4.79	Mesitylene	C9 H12	2448244
5.03	Benzene (1-methylpropyl-)	H10 H14	41491
5.21	Mesitylene	C9 H12	649930
5.43	Indane	C9 H10	417550
5.56	Benzene, 1,3-diethyl	C10 H14	272530
5.67	Benzene, 1-methyl-3-propyl-	C10 H14	232062
5.7	Benzene, 1-ethyl-3,5-dimethyl-	C10 H14	457055
5.99	Benzene, 2-ethyl-1,4-dimethyl-	C10 H14	299232
6.02	Benzene, 1-methyl-4-(1-methylethyl)-	C10 H14	248615
6.09	p-Cymene	C10 H14	672135
6.15	2,4-dimethylstyrene	C10 H12	245317
6.44	Benzene, 2-ethyl-1,4-dimethyl-	C10 H14	128334
6.6	Benzene, 1,2,4,5-tetramethyl-	C10 H14	763987
6.87	Benzene, 1-methyl-4-(1-methylpropyl)-	C11 H16	65167
6.96	2,4-dimethylstyrene	C10 H12	414002
7.12	2,4-dimethylstyrene	C10 H12	627100
7.44	Benzene, 1-methyl-4-(1-methylpropyl)-	C11 H16	91195
7.97	3-methyl-(3H)-isobenzofuran-1-one	C9 H8 O2	51269
8.58	1H-Indene, 2,3-dihydro-4,7-dimethyl-	C11 H14	193812
9.25	Naphthalene <1-Methyl->	C11 H10	873271
9.49	Naphthalene <1-Methyl->	C11 H10	458758
10.73	Naphthalene, 1,5-dimethyl-	C12 H12	297268
10.93	Naphthalene, 1,4-dimethyl-	C12 H12	496628
10.99	Naphthalene, 1,7-dimethyl-	C12 H12	478406
11.18	1-methoxymethyl-4-methylnaphthalene	C13 H14 O	347923
11.24	Naphthalene, 1,3-dimethyl-	C12 H12	151365
11.4	Naphthalene, 1,3-dimethyl-	C12 H12	289765
12.01	2,5-dimethyl-1,6-methano[10]annulene	C13 H14	92806
12.25	Naphthalene, 2,3,6-trimethyl-	C13 H14	616868
12.54	2-ethyl-3-methylnaphthalene	C13 H14	238865
12.56	Naphthalene, 2,3,6-trimethyl-	C13 H14	293854
12.74	Naphthalene, 1,4,6-trimethyl-	C13 H14	102850

GC-MS was also used to analyze for contaminant peaks from petroleum hydrocarbons in the background samples (these reference samples were only available from Sites A and B). In the chromatogram of the Site B "clean" background beach sample shown in Figure 19, no hydrocarbon-based contaminant peaks were detected by GC-MS.

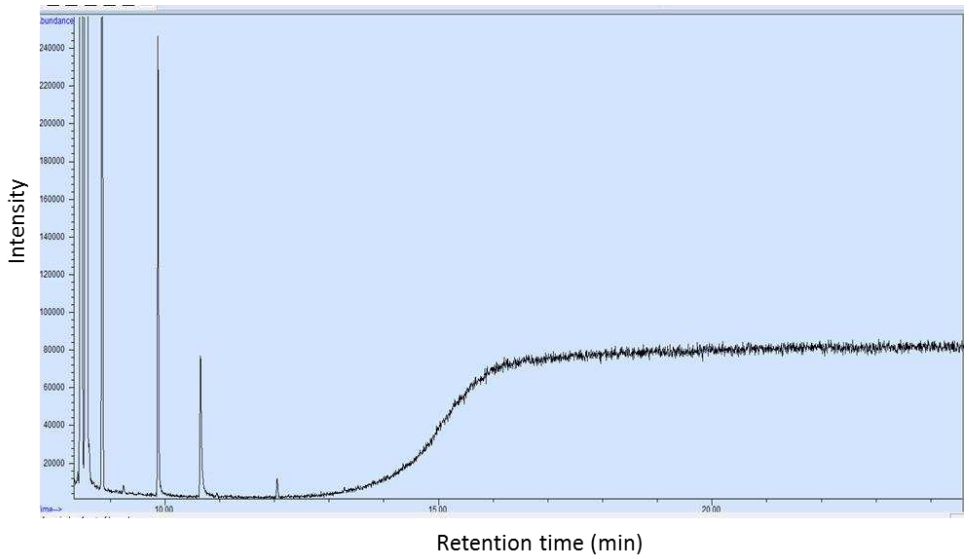

Figure 19: GC-MS chromatogram for the Site B Beach sample, which was assumed to be an uncontaminated background sample.

Table 5 lists the peaks identified in the chromatogram. These peaks were attributed to the toluene extraction solvent as *p*-xylene, benzaldehyde, and benzyl alcohol are common toluene impurities. They also appeared in the toluene blank.

Table 5: Peaks identified by GCMS present in the Site B Beach sample

Retention Time (min)	Compound
9.855	*p*-xylene
9.873	benzaldehyde
10.67	benzyl alcohol

The absence of compounds detected by GC-based techniques would support the assumption that the samples were collected from an uncontaminated location and could be used as a background for reference. However, further investigation by FT-ICR MS (see below) demonstrated the limitations of these techniques; only volatile compounds were detected, overlooking the majority of petroleum-derived species. Analysis of these contaminated sediments by GC-FID and GC-MS proved that the current standard regulatory methods do not effectively measure the transformation products of weathered oil. The toxicity, fate, and transport potential of these polar oxy-hydrocarbons remain unknown with these techniques.

4.3 Elemental composition

Elemental analysis provided another bulk parameter of the hydrocarbon-impacted samples. Table 6 shows the results of the tests, with results reported in units of % (wt/wt).

Table 6: Elemental analysis results in % by weight, conducted at Galbraith Laboratories, Inc.

	Site B		**Site A**		
	Beach	**T1**	**RU**	**RL5**	**Background**
C	< 0.50%	< 0.50%	< 0.50%	2.57%	3.14%
H	< 0.50%	< 0.50%	< 0.50%	< 0.50%	< 0.50%
N	< 0.50%	< 0.50%	< 0.50%	< 0.50%	< 0.50%
O	0.89%	1.06%	1.55%	1.05%	< 0.50%
S	0.0293%	0.0872%	0.00915%	0.0511%	0.00573%

Bulk elemental composition can provide supporting evidence for SARA classification reported by other studies (Islam et al., 2016) (Table 7).

Table 7: Characteristic oxygen content for SARA fractions of crude oil reported by Islam et al. (2016) determined by elemental analysis.

Fraction	Oxygen content (%)
Whole crude oil	0.47
Saturate	0.17
Aromatic	1.20
Resin	5.12
Asphaltene	2.58

For all five samples, elemental H and N were determined to be below the detection limit of 0.5%. Elemental C also was less than 0.5% for the Beach, T1, and RU samples. The oxygen content was slightly higher in the presumably more oxidized sample at Site A (1.55% in the upper sediment vs. 1.05% in the lower sediment). Chen and co-workers (2016) reported a 7-fold increase in oxygen in 9 month old oil samples compared to the parent oil, and ascribed this to the generation of oxidative transformation products. Thus, the elemental analysis results for the RU and RL5 samples may suggest that the petroleum hydrocarbons present at Site A had undergone some weathering along the flow path. However, overall contaminant concentrations (Table 3) were too low in the tested samples for a more detailed analysis, for instance based on elemental ratios. Furthermore, the results demonstrate that they provide limited compositional information for contaminated sediment samples.

4.4 FT-ICR MS

Fourier-transform ion cyclotron resonance mass spectrometry is a revolutionary technique because it provides an unprecedented view of compounds undetectable by traditional GC-based methods, which are constrained to (semi-)volatile and low-molecular weight compounds. As Figure 6 (from the literature review section) illustrates, FT-ICR MS can drastically expand the

analytical window for petroleum composition coverage. Using multiple ionization modes (negative- and positive-ion ESI and APPI) allowed for the detailed coverage of the array of petroleum transformation products, including polar, acidic, basic, and nonpolar species (McKenna et al., 2013).

4.4.1 Individual site comparisons

After FT-ICR MS analysis using negative-ion ESI, positive-ion ESI, and APPI, data reduction approaches were used to display the thousands of molecular formulas in intuitive graphical formats. The purpose of these graphs was to highlight compositional similarities, differences, or weathering patterns as a function of hydrogeological conditions, redox zones, or source product that were difficult to visualize from tables of formulas and DBE values alone. Sites A, B, and C discussed here were chosen because the sample coverage from these sites was the most complete; the samples analyzed were representative of distinct hydrogeological and redox conditions, which allowed for meaningful comparisons between the different zones. The three sites also included a variety of hydrocarbon products in different stages of degradation. Due to the semiquantitative nature of FT-ICR MS, the results do not reflect direct concentrations of the identified species; instead, they show compositional shifts and highlight similarities and differences between samples.

4.4.1.1 Site A

The samples were analyzed by 9.4 T FT-ICR MS by three different ionization modes: (-)ESI, (+)ESI, and (+)APPI. The broadband (-)ESI mass spectra for the Background and RL5 samples are presented in Figure 20. The spectra show *m/z* on the x-axis and relative abundance on the y-axis. In contrast to the ~40 peaks appearing in the GC-MS chromatograms, these spectra show tens of thousands of detected species with a signal-to-noise ratio of 6 (equivalent to a limit

of detection) and above. In the Background sample (assumed to represent an uncontaminated location), 27,271 peaks were assigned molecular formulas out of the 34,480 total peaks detected. As mentioned above, all samples had been extracted using toluene, which is selective towards petroleum-derived compounds and does generally not extract natural organic matter. The RL5 sample had more peaks than the Background sample, with 33,940 molecular formulas assigned out of the 40,976 total peaks detected, demonstrated in Figure 20.

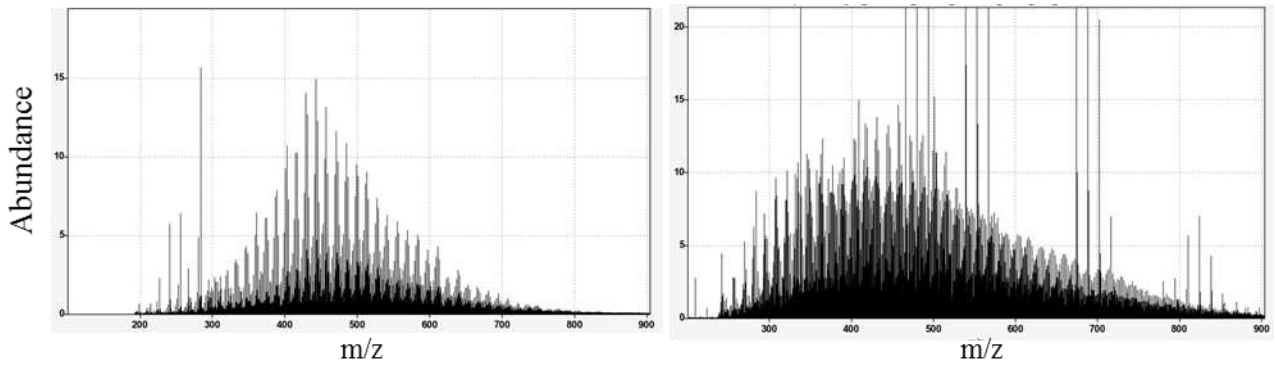

Figure 20: Broadband negative ion ESI FT-ICR MS mass spectra for the Background (left) and RL5 (right) samples.

The Background and RL5 samples both had similar peak distributions centered around 450 *m/z*, but RL5 has a slightly broader *m/z* range, with more peaks between *m/z* 700-900. The relative abundances for the RL5 peaks were also generally higher than the Background peaks. The similarities in the *m/z* range suggest that the two samples have similar chemical compositions, but there are compounds with higher molecular weight present in the RL5 sample. The other lower sediment sample (RL3) had 31,882 assigned peaks and in the upper sediment sample (RU), which was assumed to be the most oxidized, 36,021 peaks were assigned.

After PetroOrg sorted the assigned molecular formulas into heteroatom classes, the numbers of identified peaks assigned to each class were presented in a bar graph. Figure 21 compares the distribution of the dominant heteroatom classes for the species ionized by negative-

ion ESI in the upper sediment (collected from 0-10 cm below the buried oleophilic biobarrier) and lower sediment samples (from 10-20 cm below the OBB). Preliminary microbial ecology studies by Irianni Renno (2018) had found a greater abundance of aerobic microbes in the upper sediment, suggesting oxic conditions in this layer. The lower sediment zone was characterized as reducing based on the greater abundance of anaerobic microbes, which would be expected from constantly saturated sediment (Tochko, 2018).

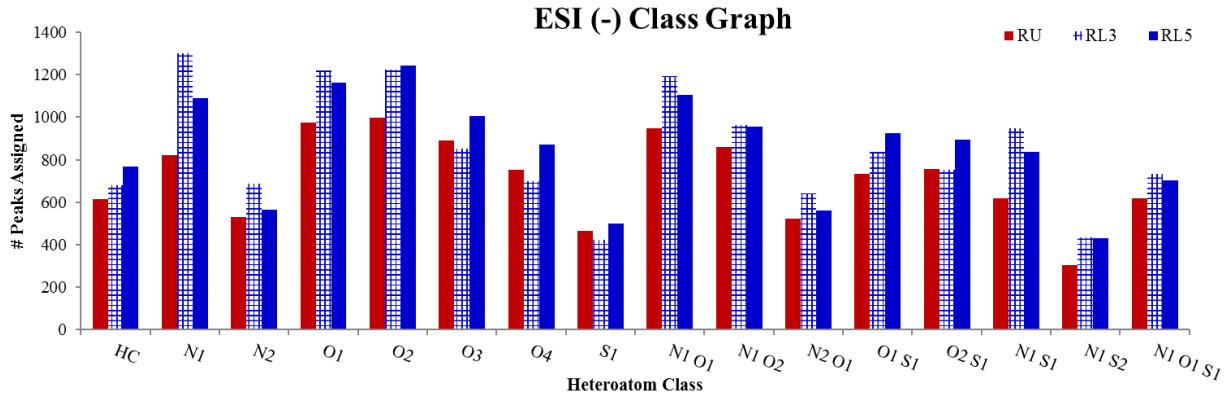

Figure 21: Heteroatom class distribution for the most prevalent classes detected by negative-ion ESI in the Site A samples.

Across all three samples, the majority of the identified peaks belonged to the O_1 and O_2 classes. The lower sediment samples also had very high relative abundances of N_1 and N_1O_1 species. The most drastic differences between the distinct sample depths were the numbers of peaks assigned to the N_1, O_1, O_2, N_1 O_1, and N_1 S_1 classes; fewer of these species were detected in the upper sediment sample (RU) compared to the lower samples (RL3 and RL5). Most of the heteroatom classes that had high relative abundances contained one or two oxygen atoms.

Even though the peak distribution in Figure 21 seems to demonstrate otherwise, the RU mass spectrum contained more assigned molecular formulas than RL3 and RL5; the class graph only presents the most prevalent heteroatom classes but in total there were 67 assigned classes in

RU and 61 in both RL3 and RL5. Since the RU assigned formulas were distributed between more heteroatom classes, fewer were assigned to the classes shown in Figure 21 compared to the RL samples. The higher number of classes correlates to a higher compositional complexity and wider range of chemical functionalities in the upper sediment, which may indicate more advanced transformation of the source products.

Oxygen containing classes were the most abundant classes detected by negative-ESI because this technique selectively targeted acidic groups. The O_1 group corresponds to alcohol functional groups, which are preferentially ionized over ketones. The O_2 class is predominantly made up of carboxylic acids, and the N_1 class corresponds to 5-membered pyrrolic nitrogen. In oil samples from the Deepwater Horizon spill, Chen and co-workers (2016) found that N_1 compounds were the most rapidly biodegradable. Thus, the lower abundance of N_1 compounds in the upper sediment may have been caused by weathering processes between the two sample depths.

RL3 and RL5 were expected to have similar composition; they were collected from the same depth but a few feet away from each other horizontally. For some of the classes (O_3, O_4, S_1, O_2, S_1), RL5 and RU appeared to be more similar to each other than RL3 and RL5. However, it is important to note that the ion efficiency bias during ionization hampers direct comparison between samples; the number of peaks assigned to each class does not directly correlate to the concentration of these compounds. High concentrations of oxygenated species can suppress the ionization of less acidic species such as N_1, NO and NS heteroatom class compounds.

The aromaticity and molecular size of the compounds detected by FT-ICR MS was visualized using plots of H-deficiency, represented by double bond equivalent (DBE) versus carbon number. In general, the lower DBE numbers correspond to more aliphatic hydrocarbons,

while higher DBE numbers represent more aromatic species. Figure 22 shows isoabundance-contoured plots of DBE on the y-axis and carbon number on the x-axis for the dominant heteroatom classes analyzed by (-)ESI.

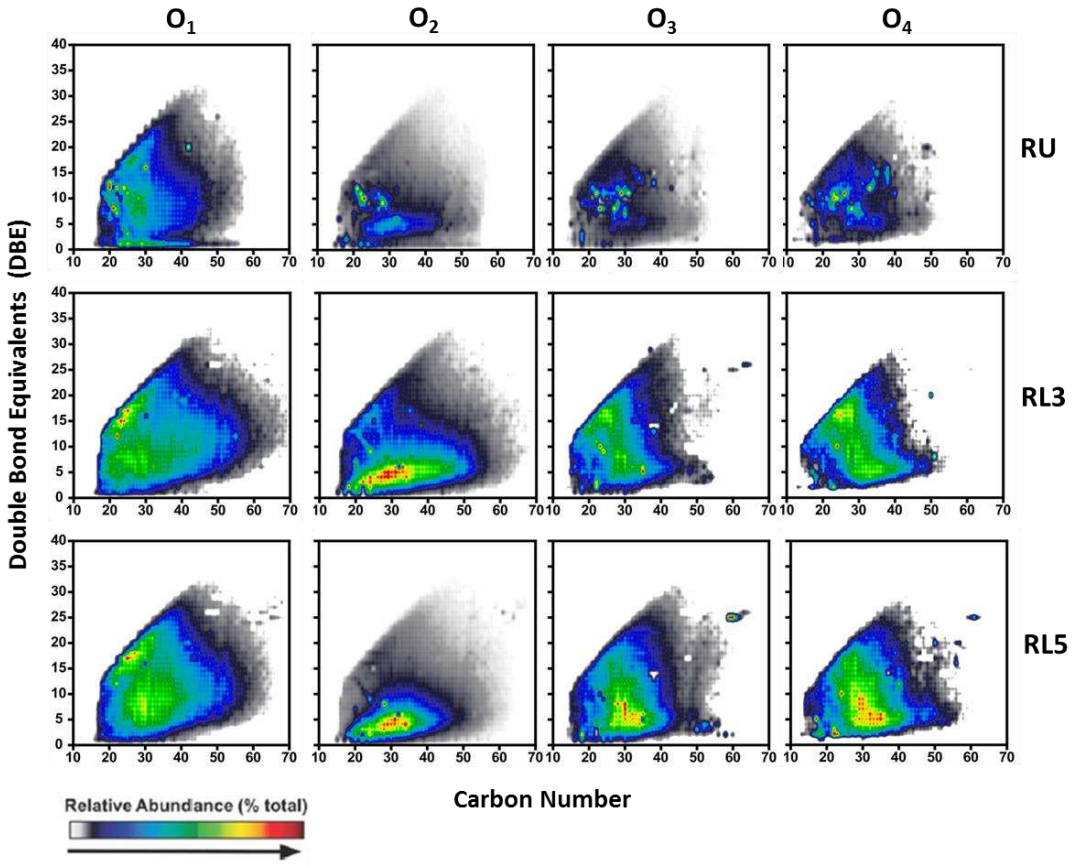

Figure 22: DBE vs. carbon number plots for the major heteroatom classes detected in the Site A samples.

The DBE and carbon number plots in Figure 22 showed bimodal distributions for the O_1-O_4 classes. These samples have one predominant distribution with low DBE and higher carbon number and another with high DBE and lower carbon number. The bimodal distribution suggests the presences of two apparent hydrocarbon distributions, which may be caused by the mixing of multiple types of fuels or different weathering trends. The O_2 class corresponds to molecules with a carboxylic acid functional group. The predominant DBE values between 2 and 5 in these

samples could suggest a high relative abundance of 1- to 6-ring naphthenic acids (Pan et al., 2017). Compounds with DBE higher than 7 are most likely multi-ring naphthenic acids or aromatic acids (Liao et al., 2012). Figure 22 shows that the 1- to 6-ring naphthenic acids are much more prevalent in the lower sediment samples compared to the upper sediment. GC-MS analyses detected methyl-, dimethyl-, and trimethylnaphthalene, which may be precursors of naphthenic acids upon oxidative transformation.

The O_1 class in (-)ESI is associated with alkylphenols, and Pan et al. (2017) identified compounds with DBE 5 (seen in the lower sediment samples) as alkylphenols containing one ring as indanols or indenols. These species are known intermediates of anaerobic biodegradation of aromatic hydrocarbons. O_4 compounds with DBE values of 4 and higher were identified in the same study as dicarboxylic acids with 2+ naphthenic rings. Definitively proving these tentative structural assignments from the elemental composition data requires more information about the original spilled product.

For all of the heteroatom classes displayed in Figure 25, there were no substantial shifts in average DBE values between the different depths, suggesting that the compounds present in both layers had a similar degree of aromaticity. However, the upper sediment sample had lower calculated abundance-weighted average carbon numbers (calculated in PetroOrg) compared to the lower sediment samples, especially for the O_1 and O_2 classes. These shifts in average carbon number reflect depth-related compositional trends. The lower average carbon number in the RU sample across all seven classes indicates the presence of lower-molecular weight compounds compared to the RL3 and RL5 samples. This is indicative of hydrocarbon transformation along the flow path from the more reducing lower to the more oxidizing upper sediment (Chen et al., 2016).

However, as stated previously, the percent relative abundance (%RA) is not necessarily directly proportional to concentration of each class in the sample. The %RA represents the amount of each class detected relative to the other detected species scaled to the most abundant peak in the spectrum. Also, ion suppression effects common in ESI cause signal magnitude of more easily ionized analytes to be higher than analytes that are not as easily ionized. Species with similar chemical properties are likely to have similar ion efficiencies. Direct quantitative comparison of %RA between samples would require an internal standard representative of all of the species of interest.

The DBE vs. carbon number plots in Figure 23 also reflected the degree of aromaticity or alkylation for the predominant species in each class through their proximity to the planar stability limit. This theoretical line, traced in the following figure, connects the maximum possible DBE for each carbon number in a planar polycyclic aromatic molecule and represents the limit between a planar and bowl-shaped structure.

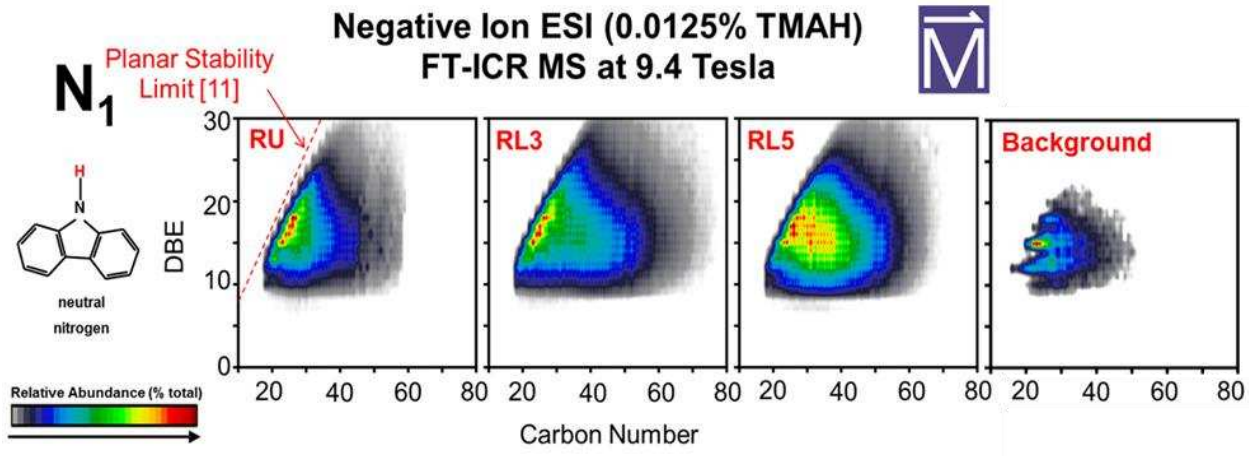

Figure 23: Negative ESI DBE vs. carbon number isocontour plots for the N_1 class for Site A samples produced by Dr. Amy McKenna, NHMFL. The high relative abundance of compounds up against the Planar Stability line indicates the presence of highly condensed aromatics.

The plots of the N_1 heteroatom class for the RU, RL3, and RL5 samples are presented in Figure 23 as an example of distributions of compounds that lie along on the planar stability line. In contrast to the O_1-O_4 classes, the N_1 plots showed a single maximum DBE and carbon number. For the N_1 class, the highest intensity DBE values ranged from 15 to 18, which suggested the presence of species with condensed aromatic benzocarbazole core structures, common compounds in petroleum mixtures (McKenna et al., 2013). Again, a decrease in higher-molecular weight species ($>C_{60}$) along the flow path from the lower to the upper sediment was indicative of ongoing weathering processes.

Compounds with DBE values of 9, 12, and 15 were identified as carbazoles, dicarbazoles, and benzocarbazoles, respectively by Pan et al. (2017). Pyrrolic nitrogen (ionized by negative ESI) has been reported as one of the most recalcitrant nitrogen compounds in many petroleum refinery processes. The Background sample did not show species that are close to the planar stability line, indicating that the compounds present are less aromatic. The presence of any N_1 compounds in the Background sample instantly signals that the sample is contaminated by petroleum hydrocarbons, contrary to the conclusions drawn previously based on GC analysis.

Positive ESI results confirmed the presence of highly condensed aromatic compounds in the RU, RL3, and RL5 samples. Figure 24 shows extremely narrow carbon number distributions. Nearly all of the detected compounds in these samples are in very close proximity to the planar stability limit.

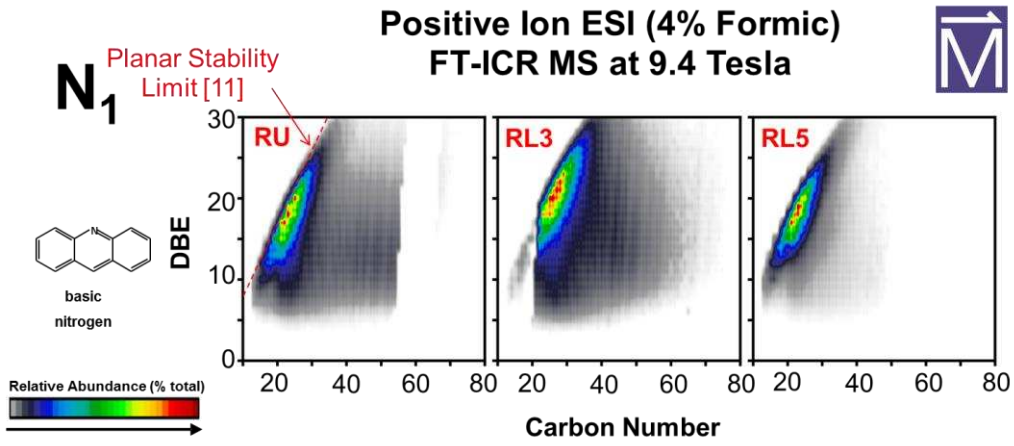

Figure 24: Positive ESI DBE vs. carbon number isocontour plots for the N_1 class for Site A samples produced by Dr. Amy McKenna, NHMFL.

In positive ESI, the N_1 is associated with pyridinic (basic) nitrogen in a 6-membered ring, a highly aromatic structure. Based on the isocontour plots, the detected highly aromatic species likely correspond to acridine or benzoquinoline core structures. As Figure 25 demonstrates, unique structures can be highlighted by each ionization mode.

DBE vs. carbon number plots of compounds ionized by APPI provided another example of the extended petroleum compound coverage enabled by FT-ICR MS. RL3 and RL5 were the only samples from Site A that had enough material for APPI analysis.

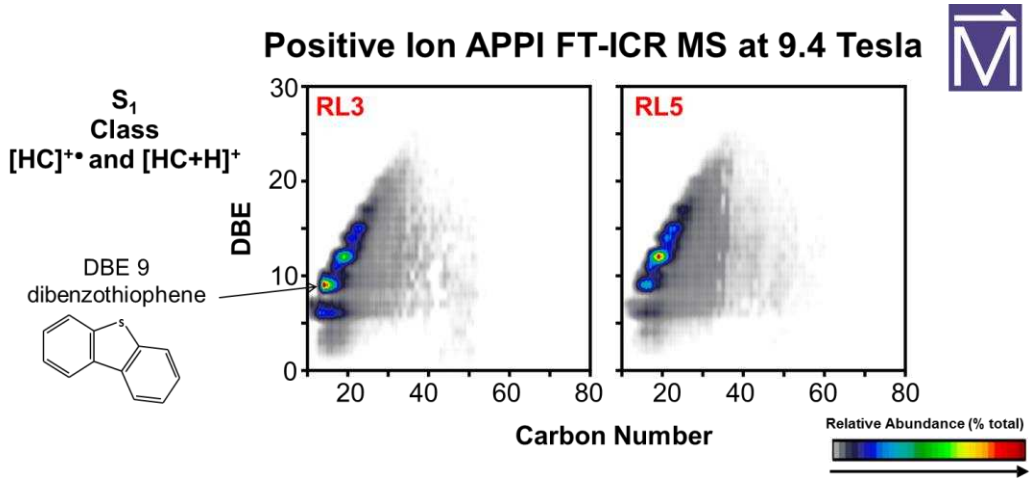

Figure 25: Positive APPI DBE vs. carbon number isocontour plots for the S_1 class for Site A samples produced by Dr. Amy McKenna, NHMFL.

Figure 25 presents a similar profile of S_1 compounds that are tightly pressed against the planar stability line. Sulfur was detected in the RL5 sample by elemental analysis, and Figure 25 shows that FT-ICR MS can help suggest the structure of these compounds. Dibenzothiophene is widely used as a model compound to represent the organic sulfur in coal and oil, and organosulfur compounds are environmentally hazardous. Compounds with a DBE of 6 or higher with at least 8 carbons correspond to benzothiophene derivatives while dibenzothiophene derivatives would appear at a DBE of 9 or higher with at least 12 carbons (Barrow et al., 2015). Figure 25 identified the highly abundant dibenzothiophenes in the RL3 and RL5 samples. Both alkylbenzothiophenes and alkyldibenzothiophenes are soluble in water, enhancing their potential for migration in the environment.

Plots of the H/C ratio on the y-axis vs. carbon number on the x-axis were created for the oxygenated classes detected by (-)ESI; these plots are commonly used by the oil industry. The H/C reflects the degree of aromaticity and can help distinguish between asphaltene and maltene products, or aromatic vs. aliphatic compounds. The plots follow the same isocontour scaling for relative percent abundance as the DBE vs. carbon number images. Figure 26 presents the H/C vs. carbon number plots for the upper and lower sediment samples.

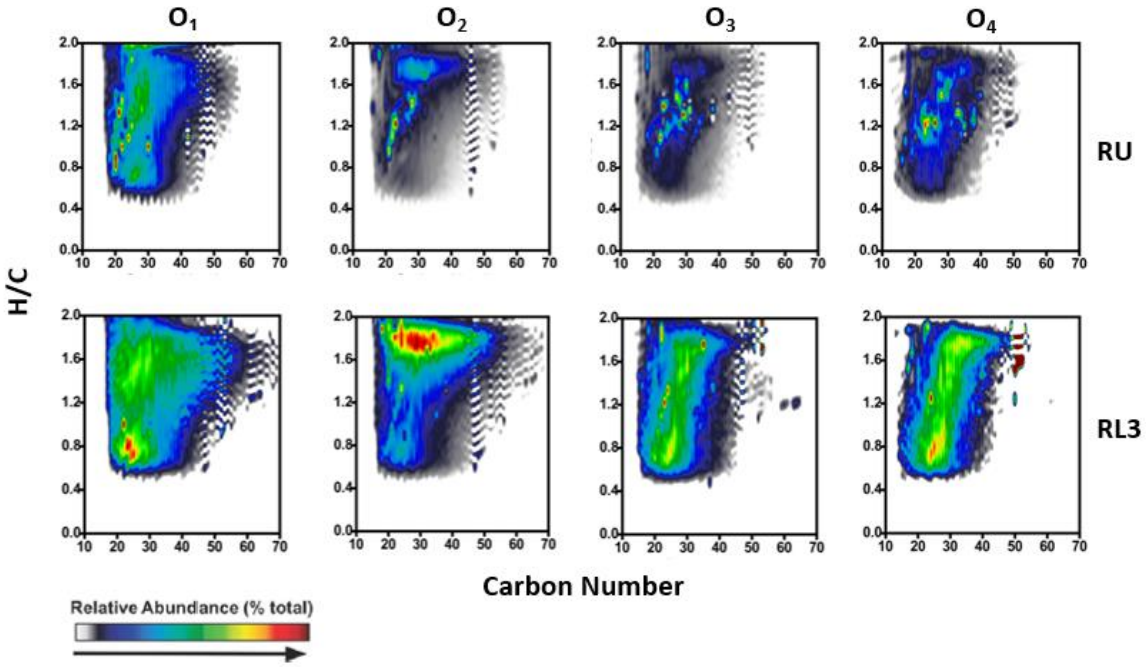

Figure 26: Negative ESI H/C vs. carbon number isocontour plots for the O_x classes detected in Site A samples.

These plots clearly show bimodal distributions in the O_x classes for the lower sediment sample but less clearly-defined distributions in the upper sediment. The predominant H/C ratios for the lower sediment layer converge around 0.8 and the 1.4-1.8 range. In the upper sample, there are distributions of similar relative abundances spanning the H/C and carbon number ranges; the distributions in both samples span roughly the same carbon number range, from around C_{25}-C_{30}. In the O_1 class, the carbon numbers of species with high H/C ratios extends over much higher carbon numbers than the species with lower H/C ratios. In the O_2 class, the higher H/C cluster has a much higher relative abundance than the lower H/C cluster.

The dominant H/C ranges in these samples match the characteristic ratios for asphaltenes and maltenes, which are ~0.9 and 1.6, respectively (McKenna et al., 2013). The maltene fraction is less aromatic and has a lower heteroatom content than asphaltenes, resulting in a lower H/C ratio. Asphaltenes have higher DBE values for the same carbon number compared to maltenes

and represent an extremely complex, polar, and recalcitrant fraction of petroleum. The O_1, O_3 and O_4 plots clearly show that the lower sediment samples contain both asphaltenes and maltenes. The bimodal distribution provides additional evidence of a blended fuel at this site.

Overall, these results suggest that the supposedly uncontaminated Background sample did indeed contain petroleum hydrocarbons, a surprising finding with important risk assessment implications. The sample appeared to be clean according to the standard method using GC-based analyses. Dissolved organic matter does not dissolve in toluene (Fu et al., 2010), so the compounds detected by FT-ICR MS were all petroleum-derived. Analyses by FT-ICR MS provided some unexpected insights into the existing site conceptual model. In contrast to initial assumptions, some of the compounds detected in the sediment were heavy asphaltenes rather than only the lighter refined products that were thought to be present.

4.4.1.2 Site B

The negative-ion ESI broadband mass spectra for the Beach sample and 12" sample from Site B are compared in Figure 27. The Beach sample was collected from an area apart from the study area and was assumed to be uncontaminated by hydrocarbons while the 12" sample originated from the sediment 12 inches beneath an OBB test site where the subsurface was determined to be contaminated based on visual appearance, odor, and TPH values.

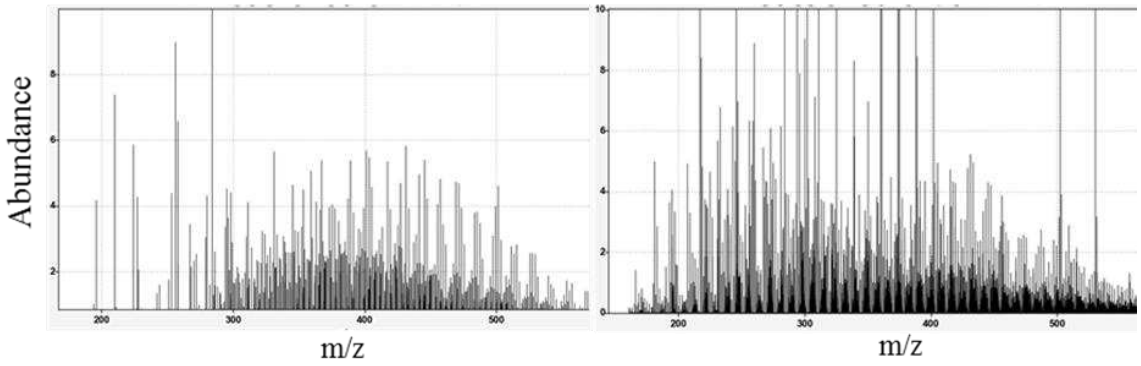

Figure 27: Broadband negative ion ESI FT-ICR MS mass spectra for the Beach (left) and 12” (right) samples from Site B.

The Beach sample spectrum contained 17,204 assigned peaks while the 12” spectrum contained 18,393 assigned peaks, suggesting that the 12” sample was more structurally complex than the Beach sample. The 12” sample had a peak distribution that spanned an overlapping but slightly wider *m/z* range and had higher peak intensities. Both distributions were centered around 400 *m/z*. The assigned formulas were sorted into heteroatom classes and the dominant classes in the 6” and 12” samples were presented in Figure 28**.** The 6” sample was collected from the sediment directly on top of the 12” sample; there were 21,412 assigned molecular formulas in the 6” sample, indicating higher compositional complexity compared to the 12” sample.

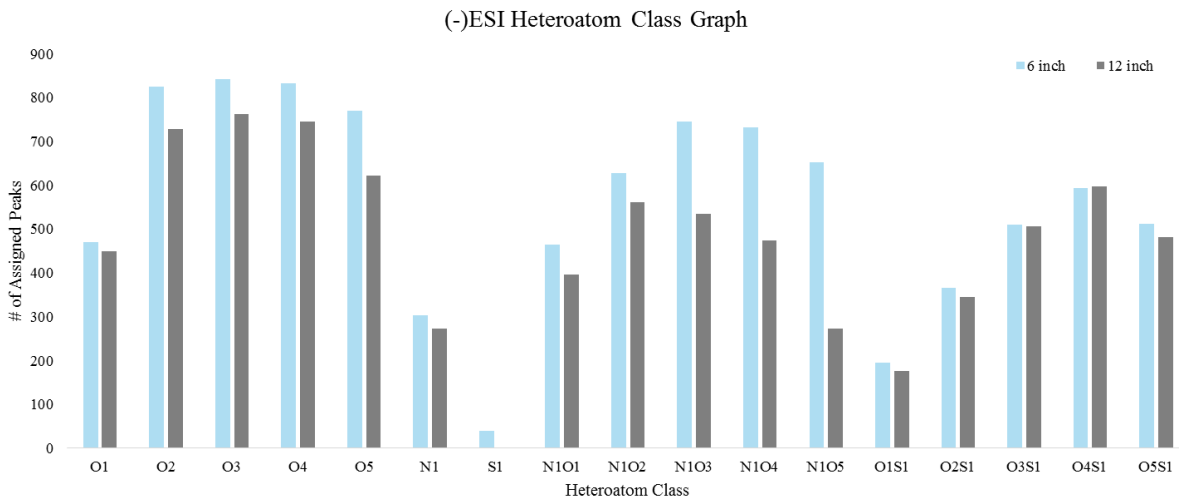

Figure 28: Heteroatom class distribution for the most prevalent classes detected by negative-ion ESI in the 6" and 12" samples from Site B.

Figure 28 describes the relative distributions of heteroatom classes within a single sample and the ion efficiency bias challenges direct comparison between different samples. For all of the samples, lower-oxygen-containing species (O_2, O_3, and O_4) were the major compound classes detected by negative ESI. The O_1 class contained the fewest number of assigned peaks out of all of the O_x classes. The major O_xS_1 classes for all of the samples were O_3S_1 and O_4S_1. O_xS_1 compounds most likely correspond to soluble partial degradation products of sulfur-containing constituents in crude oil, which are present primarily in aromatic, resin, and asphaltene fractions.

Relative distributions for most of the heteroatom classes were similar between the two depths with the exception of the N_1O_x species. In the 6" sample, there was a higher relative abundance of N_1O_3-O_5 compared to lower-order N_1O_1-O_2, and the opposite was seen for the 12" sample. N_1 species are likely converted to N_1O_1, N_1O_2 and higher-order oxygenated species through biodegradation pathways (Pan et al., 2017). Thus, the shallow 6" sample appears to be more weathered than the lower 12" sample. However, fully understanding these biodegradation

trends requires more information about the redox conditions and microbial communities present. Similar to the Site A Background sample, the presence of N_1 compounds (associated with benzocarbazoles) provided evidence of oil contamination challenging the use of the Beach sample as a reference "uncontaminated" location.

DBE vs. carbon number plots were created and the two sample depths were compared. Figure 29 includes the O_x classes which were the predominant classes detected.

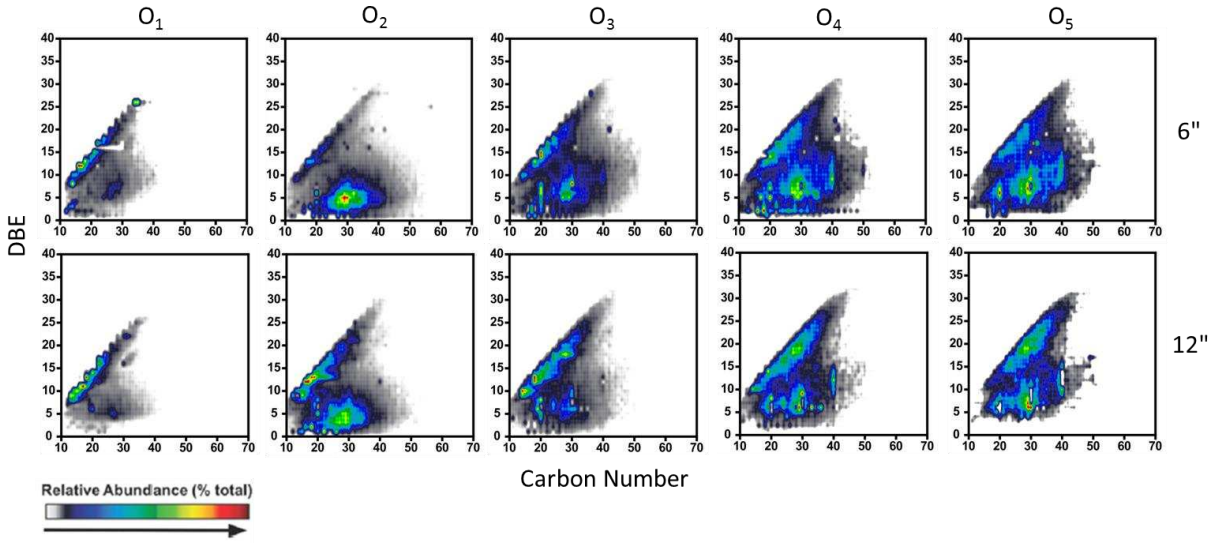

Figure 29: DBE vs. carbon number plots for negative ESI analysis of Site B samples.

First, bimodal distributions were evident in all of the classes for the 6" and 12" samples. The 6" and 12" sample also had compounds pressed tightly up against the planar stability limit; these compounds belonged to the higher-DBE distribution. The second distribution in each of the diagrams corresponded to lower DBE values. Based on previous structural identification in the literature, higher-DBE compounds are likely correlated to additional aromatic rings in the core structures with varying degrees of bioresistance (Pan et al., 2017). Neither the abundance-weighted average DBE values nor the abundance-weight average carbon numbers appeared to change with sample depth. The presence of the highly condensed aromatics up against the PSL

was consistent with the account in the historical records that wood-treating oil was stored at the site.

Figure 30 presents the H/C vs. carbon number plots for the O_2-O_5 classes. Again, the bimodal distributions shown in the plots were immediately evident.

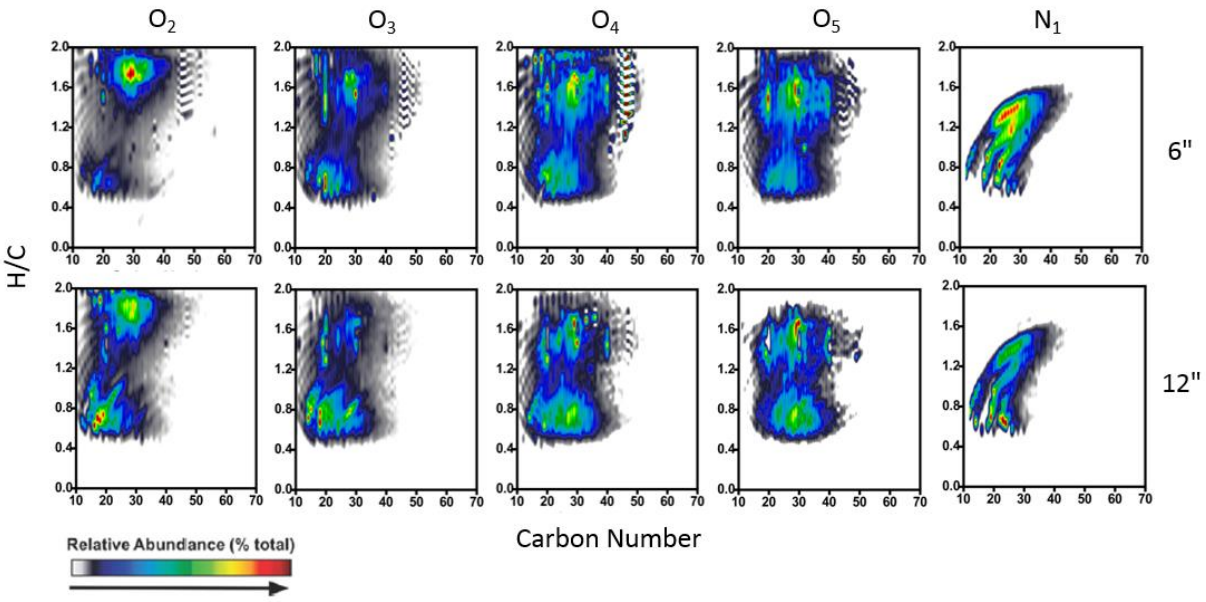

Figure 30: H/C vs. carbon number plots for the O_2-O_5 classes ionized by (-)ESI in the Site B samples.

For both samples, the compounds with the highest relative abundance were centered around ~C_{25}-C_{30} and H/C ratio of ~1.6. The 6" and 12" samples had an additional group of highly abundant compounds with an H/C of ~0.8. Similar to Site A, these H/C ratios correspond to the maltene and asphaltene fractions of crude oil. These plots suggest that the 6" and 12" samples were contaminated by highly condensed aromatic asphaltenes in addition to lower molecular-weight maltenes. The bimodal distribution in plot provided evidence of blended fuels or two stable core structures.

FT-ICR MS data also enables tracking the average oxidation state of carbon (OSC) as a function of carbon number or molecular size. Figure 31 displays the average OSC value on the y-axis calculated using Equation 4 and the carbon number on the x-axis

$$Avg.OSC = 4 - \left[\frac{4C + H - 3N - 2O - 2S}{C} \right] \qquad \textbf{(4)}$$

where C, H, N, O, and S refer to the stoichiometric numbers of each carbon, hydrogen, nitrogen, oxygen, and sulfur atom per formula.

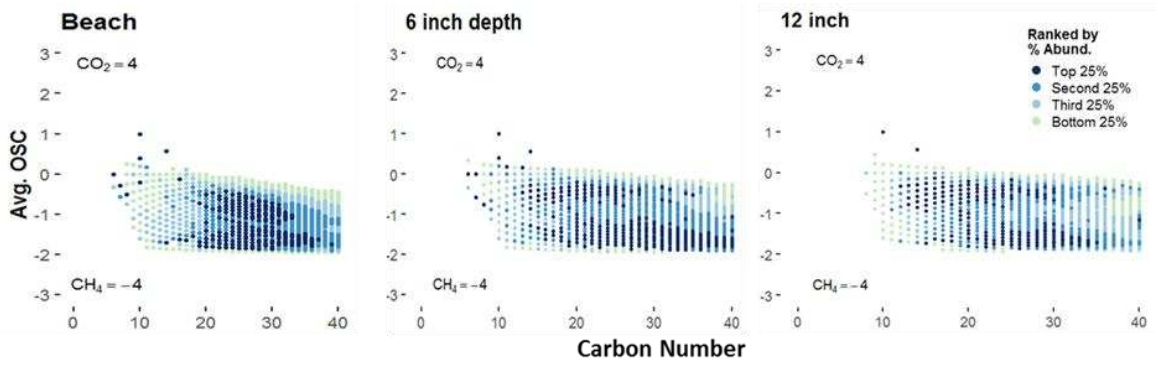

Figure 31: Average oxidation state of carbon (OSC) vs. carbon number plots for the different depths at Site B. Each point on the plot represents an assigned formula and the color corresponds to its relative percent abundance.

The maximum and minimum oxidation limits are represented by CO_2 and CH_4, respectively. The negative values for all three samples indicate that carbon is in a reduced state, far from being fully mineralized. All samples share a similar OSC and carbon number range, and these plots give additional evidence of the bimodal distribution of compounds. Based on these results, tidal flushing appeared to affect all samples evenly. The T1-T3 samples collected along a transect between the sheet pile wall and the shore may have furthered the understanding of weathering processes along the major flow path. Unfortunately, the T2 and T3 samples did not contain enough extractable material so these samples were not discussed in this work.

4.4.1.3 Site C

The samples from Site C represented two distinct hydrogeological zones – an anoxic, low k layer and an oxic, transmissive layer. Figure 32 shows the heteroatom class bar graph for the negative ESI results, with the number of peaks assigned on the y-axis and the O_x heteroatom classes on the x-axis.

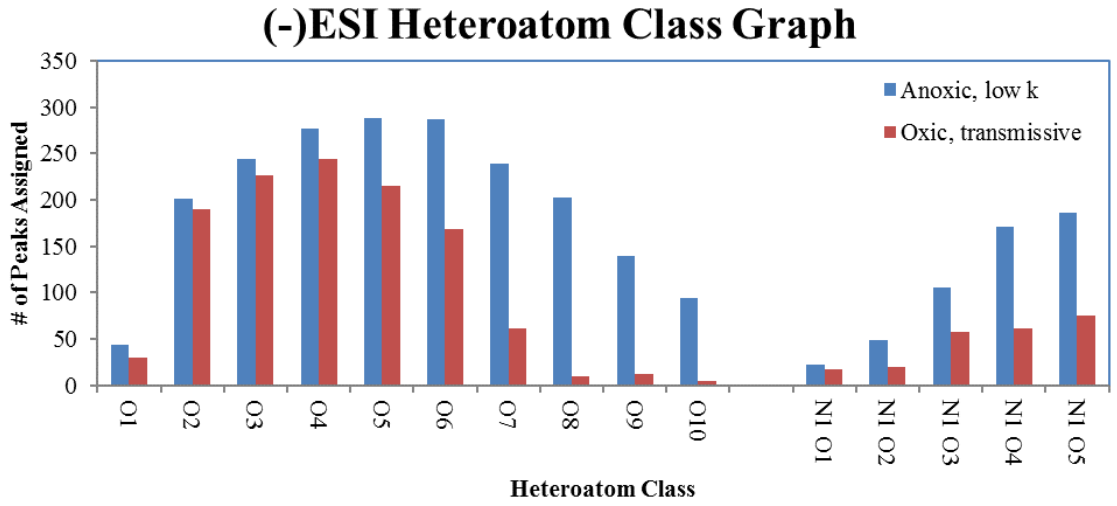

Figure 32: Heteroatom class distribution for the O_x heteroatom classes detected by negative-ion ESI for the Site C samples.

In the oxic transmissive zone sample, the most abundant class detected was O_4, and the O_7-O_{10} classes were relative low in abundance. The most abundant classes in the anoxic low k sample were O_5 and O_6, and the sample appeared to be enriched in highly-oxygenated O_7-O_{10} peaks. Overall, more higher-oxygen-containing species in the O_x class were detected at Site C compared to Sites A and B.

The molecular composition of species detected was also displayed in van Krevelen diagrams, with O/C ratio on the x-axis and H/C ratio on the y-axis shown in Figure 33. Although

the two samples had similar H/C ranges, the oxic transmissive sample (left) contained almost no compounds with O/C ratios greater than 0.6.

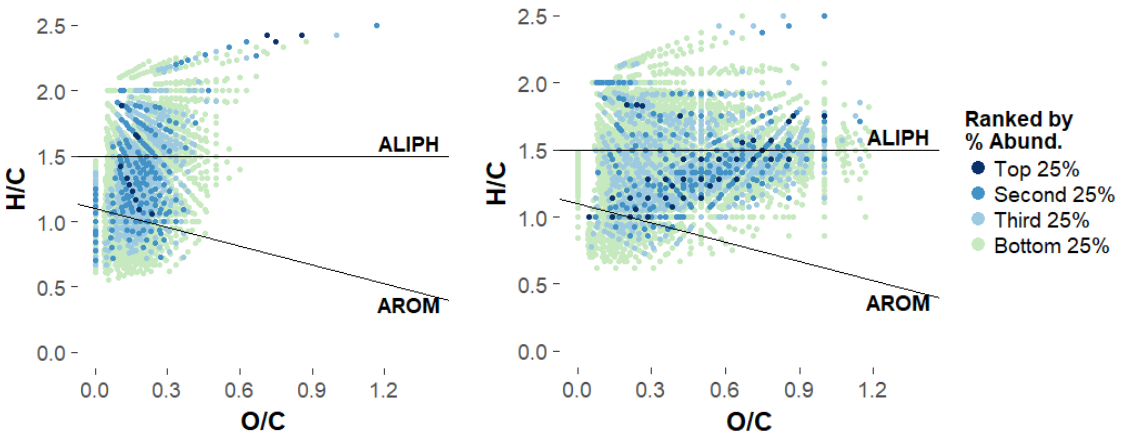

Figure 33: Van Krevelen diagrams for the oxic transmissive (left) and anoxic low k sample (right) display the degree of oxygenation in terms of the O/C ratio on the x-axis and the degree of aromaticity in terms of the H/C ratio on the y-axis.

The Van Krevelen diagrams clearly highlighted the lack of highly oxygenated species in the oxic transmissive sample in contrast to the anoxic sample. To address the impact of redox conditions on the relative abundance of polar intermediates in the two samples, microbial community analyses quantifying the relative abundance of bacterial and archaeal 16S rRNA transcripts were compared (data not shown). Consistent with the hydrogeology, results revealed that the anoxic, low k zone was dominated by methanogenic and fermentative communities, while aerobic communities dominated the oxic transmissive zone (Irianni Renno et al., 2016).

The molecular size distributions for both samples were compared by plotting the carbon number vs. average carbon oxidation state for each molecule, shown in Figure 34; CO_2 represents the upper limit of hydrocarbon oxidation while CH_4 is the most reduced species.

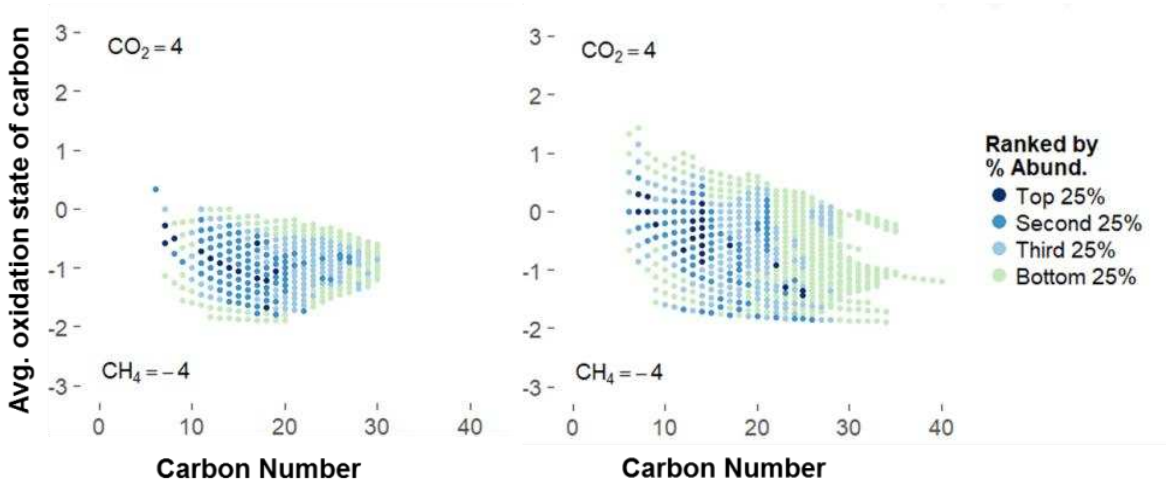

Figure 34: Carbon number vs. average oxidation state of carbon plots show that species larger than C_{30} are depleted in the oxic transmissive zone sample (left) while species in the anoxic low k zone sample (right) are more highly oxidized.

In both samples, the compounds with the highest relative abundance were lower molecular weight compounds, in the range of C_{10}-C_{20}. However, compounds with over 30 carbons were surprisingly detected in this refined gasoline LNAPL. The larger, high-molecular weight compounds (> C_{30}) appear to be depleted in the oxic sample. The anoxic sample contains low-molecular weight compounds that are more oxidized compared to the oxic sample.

The heteroatom class graph and van Krevelen diagrams show that, perhaps counterintuitively, the anoxic zone sample had a higher relative abundance of highly oxygenated compounds, associated with higher polarity, than the oxic zone sample. The oxic zone would be expected to be more weathered than the anoxic zone due to the presence of thermodynamically more favorable O_2 available for microbial transformation. Indeed, the lack of large molecular weight compounds in the oxic zone does indicate a higher degree of weathering (Chen et al., 2016). Thus, it seems plausible that more oxygenated, hydrophilic transformation products formed in the high transmissivity of the oxic zone were mobilized and flushed away (Chen et al., 2016).

In addition to negative ESI, samples were also analyzed by APPI. APPI produces radical cations and protonated ions from aromatic species and can ionize more nonpolar compounds than negative ESI. The DBE vs. carbon number plots for the anoxic low k sample and oxic transmissive sample are presented in Figure 35 for the dominant heteroatom classes.

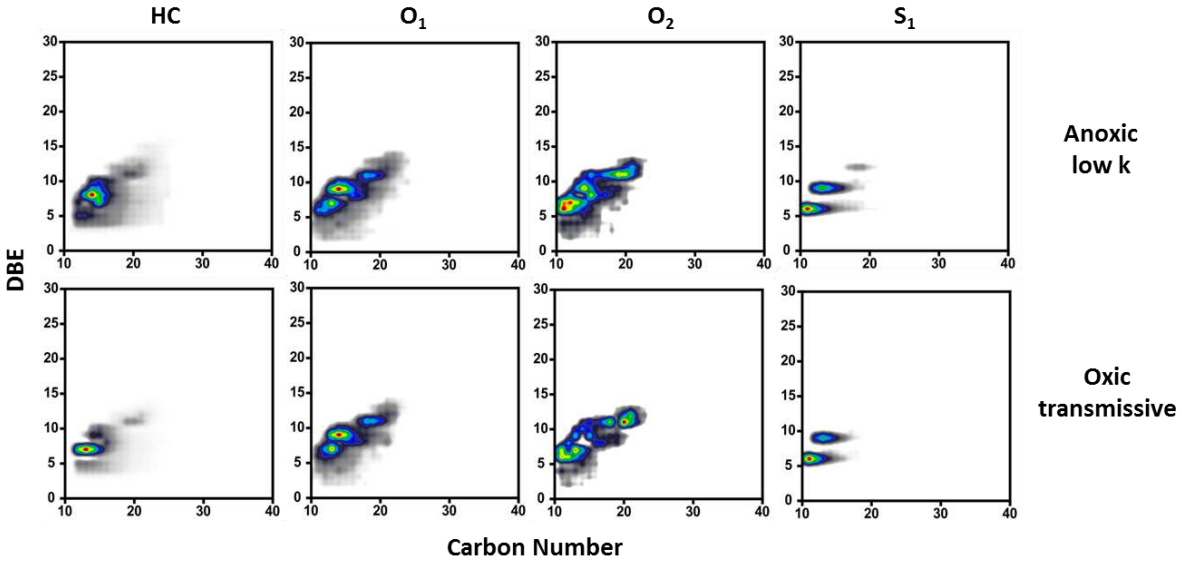

Figure 35: APPI DBE vs. carbon number plots for the anoxic low k and oxic transmissive samples from Site C.

The plots in Figure 35 appear almost identical and did not highlight any structural differences between the two hydrogeological zones. APPI analyses were not pursued further because APPI targets nonpolar aromatic compounds in crude oil. Figure 35 highlighted the strong influence of ionization method on the coverage of components detected by FT-ICR MS. APPI analyses likely identified the extremely recalcitrant compounds common to crude oils; these compounds would resist biodegradation, which may explain the similarities in the DBE vs. carbon number plots. For the investigation of differences between the water-soluble species in the two hydrogeologically distinct samples, the discussion instead focused on the water-soluble polar species ionized by negative ESI.

4.4.2 Additional data reduction strategies

Connecting the extensive characterization to the original goals of this research, data reduction strategies were investigated to develop a practical screening method that scientists, practitioners, and regulators could use to monitor these contaminated sites – ideally using widely available analytical techniques instead of time-consuming FT-ICR MS analyses only currently available in a few research laboratories. The compounds with the potential to serve as biomarkers for these new methods would be limited to transformation products rather than naturally-occurring components in crude oil. These data reduction strategies performed in this project involved identifying compounds that appeared in the (-)ESI spectra of all 15 analyzed samples from Sites A-D, including the reference "background" samples.

The capability of principal component analysis (PCA) to identify weathering trends or significant differences between samples was also explored. By reducing complex datasets to the minimum number of dimensions while keeping the maximum amount of information, PCA is meant to facilitate easier pattern recognition.

4.4.2.1 Methods to find common biomarker formulas

In order to identify elemental formulas common across all of the 15 samples analyzed, an entire list of assigned formulas from each sample reported by PetroOrg was compiled in Excel. Each sample had a separate column in the spreadsheet. Using the "sort" function, the duplicates were grouped and any formulas unique to a single sample were deleted. The duplicates appeared in at least two samples and this step simply reduced the number of formulas in the list. The "COUNTIF" function was applied to the remaining entries; if a given formula was found in all 15 columns, the output was reported as TRUE. After deleting the FALSE entries, a table of 360

common formulas (Table A1 in the Appendix) was compiled and the formulas were sorted by heteroatom class.

Table 8: Number of formulas present in every sample assigned to each heteroatom class.

Heteroatom Class	# of Formulas
N_1	19
N_1O_1	3
N_1 S_1	1
O_1	6
O_2	74
O_3	114
O_4	129
O_5	9
O_3S_1	5

These results provided more evidence once again of the dominance of O_2-O_4 classes in the (-)ESI spectra. Since oxidation converts lower-order oxygen compounds (O_1-O_3) to higher-order oxygen compounds (O_4-O_6) (Chen et al., 2016), it makes sense that the O_4 class would have the highest number of common formulas for these weathered oil samples. However, there is a great deal of ambiguity associated with this set of compounds; based on elemental composition alone, the structure cannot be determined without additional information. O_3 and O_4 species could possibly be composed of a combination of hydroxyl, ketone, ester, and carboxylic functionalities.

Compounds in this list were consistent with formulas reported in literature that were commonly detected in crude oil in general (Chen et al., 2016). In particular, S_1O_3 species have been reported in across all oil-related processes; they can be naturally-occurring in crude oil, products of degradation by thermo-sulfate reducing bacteria, surfactants injected downhole to

improve recovery or as clean-up efforts, injected downhole, or introduced during sample preparation.

Deriving the desired meaning from this set of formulas requires more structural information -- about the original products stored at these sites and the formation of the source material (including the formation geology, temperature and pressure conditions, and active microbial communities present in the formation).

4.2.2.2 Principal component analysis (PCA)

Out of all four field sites, the most complete characterization came from Site A, with samples that were analyzed by all three ionization modes. This site was therefore selected for principal component analysis (PCA) to be performed on the elemental compositions obtained from the (-)ESI, (+)ESI, and (+)APPI mass spectra individually. Inputs to the PCA function of PetroOrg were the assigned formulas from all heteroatom classes (except those containing ^{13}C) from the four Site A samples: Background, RL3, RL5, RU. Relative abundance (%RA) was selected as the variable and the PCA was based on nonlinear iterative partial least squares (NIPALS) (Lohninger, 1999).

In the PCA for (-)ESI data (presented in Figure 36), PC1 and PC2 account for 98.7% of the total explained variance. The score plot shows a clear separation in PC1 between the Background and the three samples that were collected from underneath the OBB.

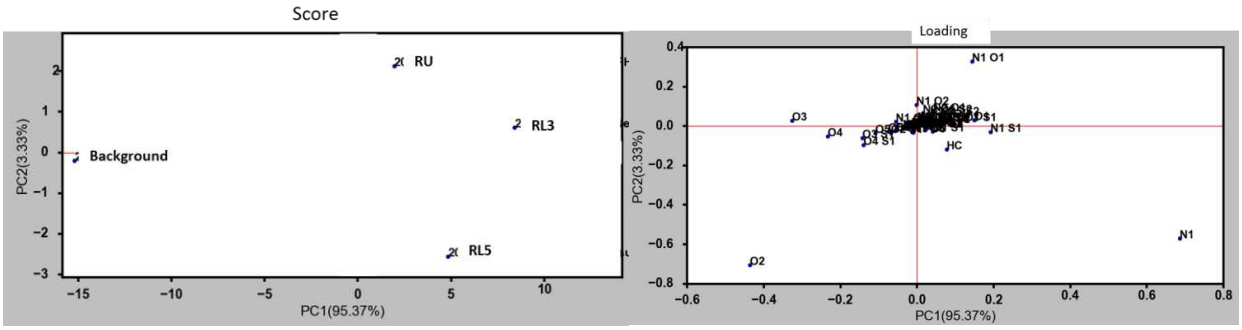

Figure 36: PCA score (left) and loading (right) plots of %RA for (-)ESI to determine compositional differences between the assigned heteroatom classes in Site A samples.

In the loading plot, almost all of the heteroatom classes clustered around the origin except for N_1 and O_2, which have opposite signs in PC1 but negative PC2 values. The relatively large distance from the origin indicates that the N_1 and O_2 classes drive the compositional differences between samples. In negative-mode ESI, N_1 corresponds to problematic and recalcitrant five-membered ring (pyrrolic) nitrogen. In the loading plot, the angles between the vectors drawn from the origin to a given heteroatom class represents how the two heteroatom classes correlate with one another (Ngo, 2018). The vectors for the N_1 and O_2 classes meet at roughly 90°, meaning they are most likely not correlated.

The PCA score plot for (+)ESI data according to %RA shows a different separation of the four Site A samples. As Figure 37 demonstrates, RL3 and RU are similar in PC1 and PC2, while RL5 and Background are spread out along both components. Since clustering in the PCA score plot gives an indication of similarity between samples, the arrangement in Figure 40 suggests that RL3 and RU were more similar to each other than RL3 and RL5. As the separation in the loading plot suggests, the N_1 group is mainly responsible for the variation between samples; nitrogen in (+)ESI corresponds to highly aromatic 6-ring nitrogen.

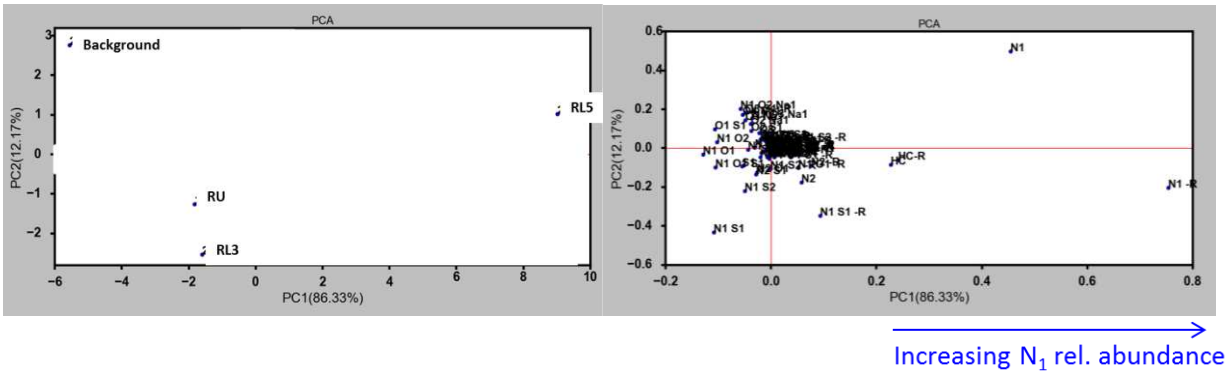

Figure 37: PCA score (left) and loading (right) plot of %RA for (+)ESI analysis of Site A samples. PC1 and PC2 account for 98.5% of the total variance.

The small angle formed by the N_1 and N_1-R vectors in the loading plot suggests that these two classes are positively correlated. The N_1-R class refers to radical anion N_1 compounds that result from this ionization technique. The PCA for the APPI data was inconclusive because only two samples (RL3 and RL5) had enough material for the APPI analysis.

Principal component analysis showed distinctions between the Site A samples and hinted at significant heteroatom classes that may cause these distinctions. For all three ionization modes, the heteroatom classes that seem to drive the variation between samples have the highest relative abundance. Based on the observation that RU and RL3 seem more similar than RL3 and RL5, the geology of the sediment layers may have been fairly heterogeneous. Alternatively, the spacing of the sample collection may not have allowed for enough biodegradation processes to differentiate the samples.

Since there was not enough information to deduce weathering trends using PCA results from a single site, PCA was also performed using samples from all four sites. All 15 of the samples were analyzed by (-)ESI, and their PetroOrg datasets were loaded into the PCA function in the program with %RA as the selected variable. The loading plot is shown in Figure 38**,** and the score plots is included in the Appendix (Figure A1).

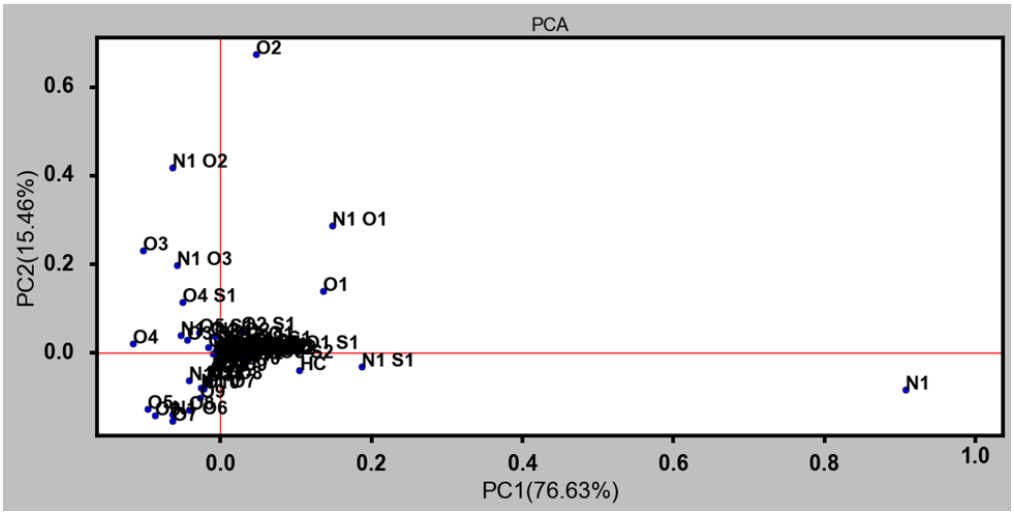

Figure 38: Plot of PCA loadings of %RA for formulas detected by (-)ESI. Across all four sites, the N_1 and O_2 classes had the most impact on differentiating samples.

Based on their distance away from the origin in the loading plot, the N_1 and O_2 classes were primarily responsible for the differences between samples. Figure 38 shows that the two classes are separated by a 90° angle, suggesting that they vary independently of each other. A high percentage of the variance, 92.09%, was explained by PC1 and PC2 indicating the strength of the model.

Samples from Site A and Site D (6 samples total) were analyzed by (+)ESI. The loadings plot is shown in Figure 39; the score plot is shown in Figure A2 of the Appendix. The first two principal components accounted for 98.51% of the variance, confirming the high model strength.

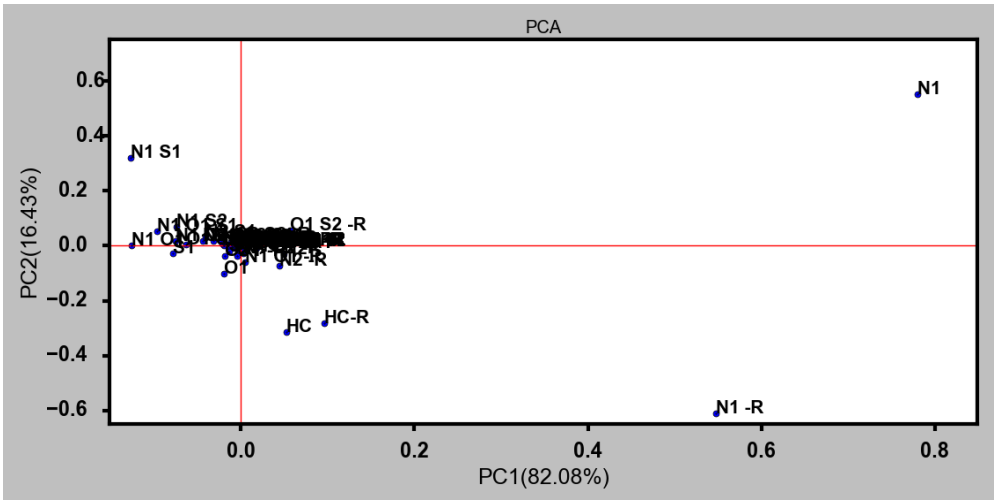

Figure 39: PCA loading plot of %RA for formulas detected by (+)ESI. The most significant compounds belonged to the N_1 and N_1-R classes.

Similar to the results displayed in Figure 38 for the single-site PCA, Figure 39 points to the significance of the N_1 and N_1-R classes. (+)ESI selectively ionized basic functional groups, so these groups correspond to compounds containing highly aromatic 6-ring nitrogen. Finally, Figure 40 presents the loading plot for the six samples analyzed by (+)APPI from Sites A, C, and D.

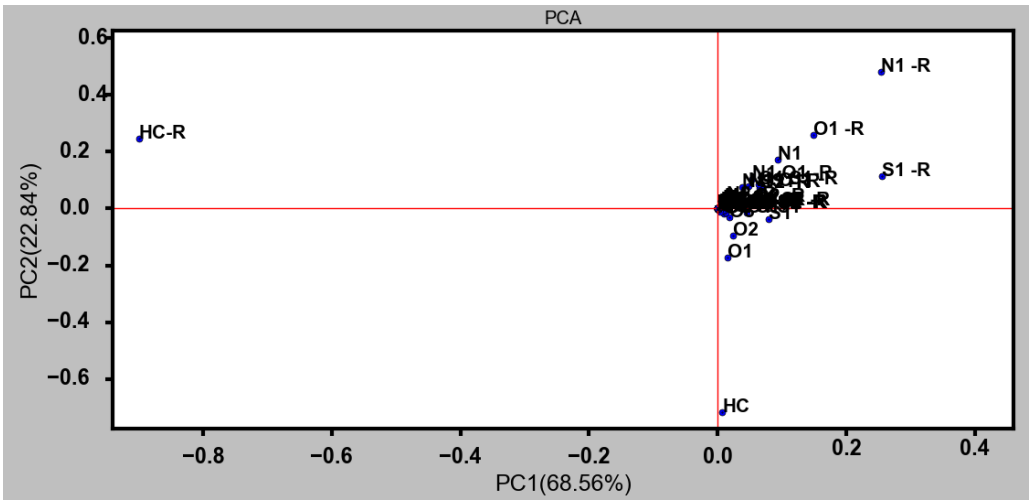

Figure 40: PCA loading plot of %RA for (+)APPI data. The most significant compounds belonged to the HC and HC-R classes.

PC1 and PC2 can account for 91.40% of the total variance. This loading plot shows the HC and HC-R classes widely separated from the others, suggesting that these classes have the greatest impact on differentiating the samples. The HC and HC-R species in (+)APPI correspond to aromatic nonpolar PAHs.

The significant classes identified by PCA for (-)ESI and (+)ESI using samples from all sites were the same as those identified by the PCA using only Site A, reinforcing the conclusions that the N_1 and O_2 classes (species containing pyrrolic nitrogen and carboxylic acid functionalities) drove the distinctions between samples. The (+)APPI PCA loading plot demonstrated that HC compounds had a significant impact. A larger number of samples used for PCA yields more reliable results, but the analyses using a single site compared to all sites showed similar significant heteroatom classes. Future analyses could focus on the N_1 and O_2 classes in (-)ESI analysis and the HC class in APPI as potential candidates for marker selection.

5. SUMMARY AND CONCLUSIONS

The objectives of this project were (1) to utilize a tiered analytical approach to investigate the polar "oxyhydrocarbons" formed by the natural weathering and biodegradation of petroleum products and (2) to identify potential marker compounds that can be used for a more focused characterization of weathering processes at petroleum-contaminated sites. Different analytical techniques gave increasingly detailed levels of characterization, ranging from the most basic GC-based methods to the ultrahigh resolution achievable only by FT-ICR MS.

Consistent with other published studies of crude oil, standard GC-FID and GC-MS could not detect the oxygenated, partly heteroatomic, high-molecular weight species present in these sediment samples. The characterization was limited to compounds with fewer than ~30 carbons and only a few polar transformation products. For elemental analysis, contaminant concentrations in the sediment cores were too low, limiting this approach to source zones where LNAPL is still present. The incomplete information about oxyhydrocarbons and heteroatomic species obtained from the suite of standard analyses pointed to the need for new methods to more accurately estimate the risk posed by oil-contaminated sites.

In contrast, FT-ICR MS analysis revealed tens of thousands of individual compounds in the petroleum-contaminated core samples as well as a variety of chemical species that were not detected by conventional gas chromatographic methods. The results qualitatively expanded the compositional coverage of polar functionalities beyond the classic naphthenic acid compounds identified in the GC-MS spectrum to highly polar, multifunctional oxidized transformation products including ketone, hydroxyl, and carboxylic acid classes of molecules. These species

have been recently identified in produced water, emulsions, and petroleum production deposits, highlighting the broad potential for this advanced analytical technique.

The molecular-level insights from FT-ICR MS analysis altered the existing site conceptual models. In these samples, a wide range of heteroatoms was identified, including environmentally harmful nitrogen- and sulfur-containing species that were not previously thought to be present. Another surprising finding was the detection of a plethora of petroleum-based contaminants in samples collected from background locations that had been deemed uncontaminated based on standard GC-based analysis. Plots of DBE vs. carbon number identified bimodal distributions of oxyhydrocarbons in the Site A and B samples. The presence of multiple highly-abundant structures indicates different sources of original spilled oil. FT-ICR also identified asphalt products, even though Site A was believed to contain only refined product. Asphaltenes are a long-term legacy of the oil industry, and FT-ICR MS is the only available instrument that can detect these extremely recalcitrant compounds. These conclusions highlight the valuable application of FT-ICR MS as a forensic fingerprinting tool that can trace weathered metabolites back to the original product. From a regulatory standpoint, source tracking can help resolve environmental liability issues.

At Site A, depletions in high-molecular weight compounds and compounds with a low degree of oxygenation (i.e., O_1 and O_2 classes) along the flow path from anoxic to more oxic conditions offered indication for ongoing weathering processes. However, at Site C this new technique also pointed out the impact of site-specific hydrogeological conditions on the fate of oxygenated transformation products. Here, the oxic zone of a higher degree of weathering displayed less oxygenated products than the anoxic zone, challenging previous expectations that extended weathering would be correlated with a higher degree of oxygenation in the petroleum

hydrocarbon composition. Due to the high transmissivity of the oxic zone, polar oxygenated metabolites were likely flushed out, while the anoxic low k zone retained these compounds. The major impact of advective transport on the fate of water-soluble transformation products complicates the objectives of identifying and relying on indicator compounds.

Principal component analyses provide the potential to unravel degradation trends if a more targeted investigation can include a range of representative samples with known redox, geological, and microbiological properties. PCA was performed with data from all three ionization modes using samples from Site A only, and the analysis was also run using samples from Sites A-D. The resulting score plots reflected the compositional similarity of samples based on where they clustered relative to each other. Loading plots indicated that the N_1 and O_2 classes (species containing pyrrolic nitrogen and carboxylic acid functionalities) drove the distinctions between samples, while the (+)APPI PCA showed that HC compounds (aromatic PAHs) had a strong influence on distinguishing samples. Species in these classes may be potential marker compounds for plume characterization. PCA score plots readily detect outliers when the sample size is large enough, but the small number of samples in this experiment limited the conclusions that could be made. Different sediment layers were expected to have compositions differences, but as the sampling locations were too similar or close together, ongoing biodegradation processes have not had enough time or distance to reveal clear weathering trends.

The results from this project also highlighted the ongoing challenges of crude oil analysis. No representative model compounds for oil have been found that reflect the bulk properties of the whole oil, and there are no commercial standards for the naphthenic acids produced by oil-degrading microorganisms or those present in the reservoirs where the original crude oil was formed. Weathering and biodegradation results in higher-order oxygenated and

other polyfunctional compounds, which cannot be separated by LC methods. Only one successful separation approach has been developed, but this aromatic ring separation can only be applied to unweathered oil (Podgorski, 2013).

Nevertheless, this project has created a strong foundation for future studies characterizing petroleum weathering processes, and has demonstrated the unique capabilities of FT-ICR MS to uncover significant environmental forensic information. Recognizing the transport and ongoing transformation of weathered mixtures is the first step toward understanding the NSZD processes and evaluating the risk posed by weathered petroleum plumes. More research is required on this topic, which is outlined in the following chapter.

6. FUTURE WORK

Results of the analyses described in this project demonstrated that FT-ICR MS can reveal environmentally relevant heteroatom classes that are undetectable by standard bulk analytical techniques. The methods used to monitor contaminated sites need to move beyond the traditional GC-based analyses, but the limitations revealed by the list of common formulas and PCA show that deriving structural trends depends on site-specific, detailed knowledge about the original petroleum products historically present, which is often not available.

Building on the findings described in this thesis and dealing with the on-going challenges of crude oil characterization, future work will require more in-depth and specific FT-ICR MS analysis. Using an internal standard could optimize the analysis of the acidic oxygenated compound distribution and enable calculating relative concentrations. Pan et al. (2017) used octadecane- D35 acid as an internal standard for the O_1, O_2, O_3, and O_4 class species. For more complex polyfunctional species, selection of the internal standard requires careful consideration of preferential ionization in a given ionization mode.

The next steps toward deducing more structural information about polar oxyhydrocarbons would involve isolation of the emulsion-stabilizing compounds, referred to as interfacial material (IM). These compounds influence the transport potential of the petroleum metabolites. Also, biodegradation requires solubilization, so hydrophobic petroleum hydrocarbon degradation depends on interfacially active compounds. Biosurfactants produced by oil-degrading microorganisms promote the solubility and enhance the biodegradation of complex hydrocarbons like asphaltenes and resins. Consequently, although not directly derived from petroleum, biosurfactants may be potential marker compounds to track hydrocarbon weathering.

Next, chromatographic fractionation by a modified aminopropyl silica (MAPS) method can optimize the analysis of complex crude oil mixtures by simultaneously isolating the highly abundant acidic compounds and separating them into distinct molecular weight ranges. The acids in each fraction have similar ionization efficiencies, so ion suppression effects are reduced and the instrumental dynamic range is increased (Rowland 2014). MAPS fractionation of IM acids has been shown to significantly extend the detection of high-*m/z* (>600-Da) IM compounds (Clingenpeel 2017). The species that skew the relative abundance distribution in electrospray ionization can also be removed by anion exchange before mass spectral analysis (Rodgers 2013).

Only after the low-molecular weight acids were removed were important biosurfactants ($>C_{30}$ acids) detected, because low- molecular weight acids dominated the spectra and caused signal suppression of the biosurfactants (Rowland et al., 2014). FT-ICR MS analysis of the biosurfactants produced during weathering coupled with microbial community analyses could give insight into the biodegradation mechanisms occurring in hydrocarbon-contaminated environments. The specific microorganisms actively biodegrading the petroleum compounds are correlated to the DBE values of the acidic metabolites found at the oil-water interface.

Further fractionation of the oil-contaminated samples could also shed new light on the transport potential of the remaining recalcitrant asphaltenes. As the most recalcitrant fraction of crude oil, these compounds are assumed not to migrate far in the environment through groundwater flow or tidal flushing. However, structural analysis of the fractionated asphaltenes can explore the possibility that these complex mixtures were deposited as aggregates and slowly separated into monomers over time (Duran et al., 2019).

Overall, these procedures would help overcome the ion efficiency bias and yield different compositional coverage than the results from the unfractionated samples described in this work.

After more accurate characterizations of the polar fraction of petroleum transformation products present, targeted regulatory methods can be developed. Additional analyses would require more material from the contaminated sites and a larger sample quantity that is representative of the hydrogeology and redox variations. Unfortunately, obtaining original material from these contaminated sites is often challenging due to the ambiguity of the historical use and ownership. A possible recommendation is to analyze and archive fresh spill samples for future assessments of weathering. In conjunction with redox measurements and microbial analyses, the detailed molecular characterization from FT-ICR MS data could provide valuable insight about weathering trends, potential toxicity effects, and the long-term fate of these oxidized transformation products over time.

REFERENCES

Aeppli, C.; Carmichael, C.A.; Nelson, R.K.; Lemkau, K.L.; Graham, W.M.; Redmond, M.C.; Valentine, D.L.; Reddy, C.M. (2012) Oil Weathering after the Deepwater Horizon Disaster Led to the Formation of Oxygenated Residues. *Environmental Science & Technology*, 46(16), 8799-8807. dx.doi.org/10.1021/es3015138

Aeppli, C.; Nelson, R.K.; Radovic, J.R.; Carmichael, C.A.; Valentine, D.L.; Reddy, C.M. (2014) Recalcitrance and Degradation of Petroleum Biomarkers upon Abiotic and Biotic Natural Weathering of Deepwater Horizon Oil, *Environmental Science & Technology*, 48(12), 6726−6734. dx.doi.org/10.1021/es500825q

Avneri-Katz, S.; Young, R.B.; McKenna, A.M.; Chen, H.; Corilo, Y.E.; Polubesova, T.; Borch, T.; Chefetz, B. (2017) Adsorptive fractionation of dissolved organic matter (DOM) by mineral soil: Macroscale approach and molecular insight. *Organic Geochemistry*, 103, 113–124. https://doi.org/10.1016/j.orggeochem.2016.11.004

Baedecker, M.J.; Eganhouse, R.P.; Qi, H.; Cozzarelli, I.M.; Trost, J.J.; Bekins, B.A. (2018). Weathering of Oil in a Surficial Aquifer, *Groundwater*, 56, 5, 797-809. doi: 10.1111/gwat.12619

Barrow, M.P.; Peru, K.M.; Fahlman, B.; Hewitt, L.M.; Frank, R.A.; Headley, J.V. (2015) Beyond Naphthenic Acids: Environmental Screening of Water from Natural Sources and the Athabasca Oil Sands Industry Using Atmospheric Pressure Photoionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, *Journal of The American Society for Mass Spectrometry*, 26(9), 1508-1521. doi: 10.1007/s13361-015-1188-9

Bekins, B.A.; Cozzarelli, I.M.; Erickson, M.L.; Steenson, R.A.; Thorn, K.A. (2016) Crude Oil Metabolites in Groundwater at Two Spill Sites, *Groundwater*, 54(5), 681-691. doi: 10.1111/gwat.12419

Blakney, G.T.; Hendrickson, C.L.; Marshall, A.G. (2011). Predator data station: A fast data acquisition system for advanced FT-ICR MS experiments. *International Journal of Mass Spectrometry*, 306(2-3), 246-252. https://doi.org/10.1016/j.ijms.2011.03.009

Brassington, K.J.; Hough, R.L.; Paton, G.I.; Semple, K.T.; Risdon, G.C.; Crossley, J.; Hay, I.; Askari, K.; Pollard, S.J.T. (2007). *Critical Reviews in Environmental Science and Technology*, 37(3), 199-232. https://doi.org/10.1080/10643380600819625

Chalfant, M. W. (2015*). Oleophilic biobarriers for control of hydrocarbon sheens at groundwater surface water interfaces* (Master's thesis). Colorado State University Libraries.

Chen, H.; Hou, A.; Corilo, Y.E.; Lin, Q.; Lu, J.; Mendelssohn, I.A.; Zhang, R.; Rodgers, R.P.; McKenna, A.M. (2016). 4 Years after the Deepwater Horizon Spill: Molecular Transformation of Macondo Well Oil in Louisiana Salt Marsh Sediments Revealed by FT-ICR Mass Spectrometry. *Environmental Science & Technology*, 50, 9061-9069. doi: http://dx.doi.org/10.1021/acs.est.6b01156

Cho, Y.; Kim, Y.H.; Kim, S. (2011) Planar Limit-Assisted Structural Interpretation of Saturates/Aromatics/Resins/Asphaltenes Fractionated Crude Oil Compounds Observed by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry*, 83(15), 6068-6073. https://doi.org/10.1021/ac2011685

Clingenpeel, A.C.; Rowland, S.M.; Corilo, Y.E.; Zito, P.; Rodgers, R.P. (2017) Fractionation of Interfacial Material Reveals a Continuum of Acidic Species That Contribute to Stable Emulsion Formation, *Energy & Fuels*, 31(6), 5933-5939. doi: 10.1021/acs.energyfuels.7b00490

Corilo, Y.E.; Podgorski, D.C.; McKenna, A.M.; Lemkau, K.L.; Reddy, C.M.; Marshall, A.G.; Rodgers, R.P. (2013) Oil Spill Source Identification by Principal Component Analysis of Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectra. *Analytical Chemistry*, 85(19), 9064–9069. doi: 10.1021/ac401604u

de Gouw, J.A.; Middlebrook, A.M.; Warneke, C.; Ahmadov, R.; Atlas, E.L.; Bahreini, R.; Blake, D.R.; Brock, C.A.; Brioude, J.; Fahey, D.W.; Fehsenfeld, F.C.; Holloway, J.S.; Le Henaff, M.; Lueb, R.A.; McKeen, S.A.; Meagher, J.F.; Murphy, D.M.; Paris, C.; Parrish, D.D.; Perring, A.E.; Pollack, I.B.; Ravishankara, A.R.; Robinson, A.L.; Ryerson, T.B.; Schwarz, J.P.; Spackman, J.R.; Srinivasan, A.; Watts, L.A. (2011). Organic aerosol formation downwind from the Deepwater Horizon oil spill. *Science*, 331(6022), 1295-1299. doi: 10.1126/science.1200320

Dongbao, F.; Woods, J.R.; Kung, J.; Kingston, D.M.; Kotlyar, L.S.; Sparks, B.D.; Mercier, P.H.J.; McCracken, T.; Ng, S. (2010) Residual Organic Matter Associated with Toluene-Extracted Oil Sands Solids and Its Potential Role in Bitumen Recovery via Adsorption onto Clay Minerals. Energy & Fuels, 24(4), 2249-2256. https://pubs.acs.org/doi/abs/10.1021/ef900885p

Duran, J.A.; Casas, Y.A.; Xiang, L.; Zhang, L.; Zeng, H.; Yarranton, H.W. (2019). Nature of Asphaltene Aggregates. Energy & Fuels, 33, 3694−3710. doi: http://dx.doi.org/10.1021/acs.energyfuels.8b03057

Fingas, M.F. (1999). The Evaporation of Oil Spills: Development and Implementation of New Prediction Methodology. *International Oil Spill Conference Proceedings*, 1999(1), 281-287. https://doi.org/10.7901/2169-3358-1999-1-281

Fu, D.; Woods, J.R.; Kung, J.; Kingston, D.M.; Kotlyar, L.S.; Sparks, B.D.; Mercier, P.H.J.; McCracken, T.; Ng, S. (2010) Residual Organic Matter Associated with Toluene-Extracted Oil Sands Solids and Its Potential Role in Bitumen Recovery via Adsorption onto Clay Minerals, Energy & Fuels, 24(4), 2249-2256. https://doi.org/10.1021/ef900885p

Garg, S.; Newell, C.J.; Kulkarni, P.R.; King, D.C.; Adamson, D.T.; Irianni Renno, M.; Sale, T.C. (2017). Overview of Natural Source Zone Depletion: Processes, Controlling Factors, and Composition Change. *Groundwater Monitoring & Remediation*, 37(3), 62-81. doi: 10.1111/gwmr.12219

He, H. (2009). *Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry in Biological Applications: Lipids and Proteins* (Master's thesis). Florida State University Libraries.

Hostettler, F.D.; Kvenvolden, K.A. (2002). Alkylcyclohexanes in Environmental Geochemistry. Environmental Forensics, 3(3-4). 293-301. https://doi.org/10.1006/enfo.2002.0100

Irianni Renno, M.; Akhbari, D.; Olson, M.R.; Byrne, A.P.; Lefèvre, E.; Zimbron, J.; Lyverse, M.; Sale, T.C.; S.K. De Long. (2016). Comparison of Bacterial and Archaeal Communities in Depth-Resolved Zones in an LNAPL Body, *Applied Microbiology and Biotechnology*, 100(7), 3347-3360. doi: 10.1007/s00253-015-7106-z

Islam, A.; Ahmed, A.; Thorn, K.; Kim. S. (2016). Molecular-level evidence provided by ultrahigh resolution mass spectrometry for oil-derived doc in groundwater at Bemidji, Minnesota. *Journal of Hazardous Materials*, 320(15), 123-132. http://dx.doi.org/10.1016/j.jhazmat.2016.08.018

Jiang, T. (2017). Top-down and Middle-down Proteomics by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (Master's thesis). Florida State University Libraries.

Kaiser, N.K.; Quinn, J.P.; Blakney, G.T.; Hendrickson, C.L.; Marshall, A.G. (2011). A Novel 9.4 Tesla FTICR Mass Spectrometer with Improved Sensitivity, Mass Resolution, and Mass

Range. Journal of The American Society for Mass Spectrometry, 22(8), 1343-1351. https://doi.org/10.1007/s13361-011-0141-9

Karimi Askarani, K.; Stockwell, E.B.; Piontek, K.R.; Sale, T.C. (2018). Thermal Monitoring of Natural Source Zone Depletion. Groundwater Monitoring & Remediation, 38(3), 43-52. doi:10.1111/gwmr.12286

Karimi Askarani, K. (2019). Thermal Monitoring of Natural Source Zone Depletion (Doctoral dissertation, Colorado State University. Libraries).

Ke, L.; Bao, W.; Chen, L.; Wong, Y.S.; Tam, N.F.Y. (2009). Effects of humic acid on solubility and biodegradation of polycyclic aromatic hydrocarbons in liquid media and mangrove sediment slurries. *Chemosphere*, 76(8), 1102–1108. doi: 10.1016/j.chemosphere.2009.04.022

Kendrick, E. (1963) A Mass Scale Based on CH_2 = 14.0000 for High Resolution Mass Spectrometry of Organic Compounds. *Analytical Chemistry*, 35, 2146–2154. https://doi.org/10.1021/ac60206a048

Kiaalhosseini, S.; Johnson, R.L.; Rogers, R.C.; Irianni Renno, M.; Lyverse, M.; Sale, T. (2016). Cryogenic Core Collection (C3) from Unconsolidated Subsurface Media. Groundwater Monitoring and Remediation, 36(4), 41-49. doi: 10.1111/gwmr.12186

Koch, B.P.; Dittmar, T. (2006). From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Communications in Mass Spectrometry*, 20(5), 926-932. https://doi.org/10.1002/rcm.2386

Liao, Y.; Shi, Q.; Hsu, C.S.; Pan, Y.; Zhang, Y. (2012). Distribution of acids and nitrogen-containing compounds in biodegraded oils of the Liaohe Basin by negative ion ESI FT-ICR MS. *Organic Geochemistry*, 47, 51-65. https://doi.org/10.1016/j.orggeochem.2012.03.006

Linh Ngo. (2018) Principal component analysis explained simply, BioTuring Blog. https://blog.bioturing.com/2018/06/14/principal-component-analysis-explained-simply/

Liu, Z.; Liu, J.; Zhu, Q.; Wu, W. (2012). The weathering of oil after the Deepwater Horizon oil spill: insights from the chemical composition of the oil from the sea surface, salt marshes and sediments. *Environmental Research Letters*, 7(3). doi:10.1088/1748-9326/7/3/035302

Liu, Y.; Kujawinski, E.B. (2015). Chemical Composition and Potential Environmental Impacts of Water-Soluble Polar Crude Oil Components Inferred from ESI FT-ICR MS. *PLoS ONE*, 10, https://doi.org/10.1371/journal.pone.0136376

Lobodin, V.V.; Juyal, P.; McKenna, A.M.; Rodgers, R.P.; Marshall, A.G. (2013). Tetramethylammonium hydroxide as a reagent for complex mixture analysis by negative ion electrospray ionization mass spectrometry. *Analytical Chemistry*, 85(16), 7803-7808. doi: 10.1021/ac401222b

Lohninger, H. (1999). Teach/Me Data Analysis; Springer: Berlin.

Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews*, 17, 1-35. https://doi.org/10.1002/(SICI)1098-2787(1998)17:1<1::AID-MAS1>3.0.CO;2-K

McCay, D. F.; Rowe, J. J.; Whittier, N.; Sankaranarayanan, S.; Etkin, D. S. (2004). Estimation of potential impacts and natural resource damages of oil. *Journal of Hazardous Materials*, 107(1-2), 11−25. doi: 10.1016/j.jhazmat.2003.11.013

McGuire, J.T.; Cozzarelli, I.M.; Bekins, B.A.; Link, H.; Martinović-Weigelt, D. (2018) Toxicity Assessment of Groundwater Contaminated by Petroleum Hydrocarbons at a Well-Characterized, Aged, Crude Oil Release Site, *Environmental Science & Technology,* 52(21), 12172-12178. doi: 10.1021/acs.est.8b03657

McKenna, A.M. (2009). *Detailed Characterization of Heavy Crude Oils and Asphaltenes by Ultrahigh Resolution Fourier Transform Cyclotron Resonance Mass Spectrometry*. (Doctoral Thesis) Florida State University Libraries.

McKenna, A.M.; Nelson, R.K.; Reddy, C.M.; Savory, J.J.; Kaiser, N.K.; Fitzsimmons, J.E.; Marshall, A.G.; Rodgers, R.P. (2013). Expansion of the Analytical Window for Oil Spill Characterization by Ultrahigh Resolution Mass Spectrometry: Beyond Gas Chromatography. *Environmental Science & Technology,* 47(13), 7530−7539. dx.doi.org/10.1021/es305284t

Melbye, A.G.; Brakstad, O,G,; Hokstad, J.N.; Gregersen, I.K.; Hansen, B.H.; Booth, A.M.; Rowland, S.J.; Tollefsen, K.E. (2009) Chemical and toxicological characterization of an unresolved complex mixture-rich biodegraded crude oil. *Environmental Toxicology and Chemistry*, 28(9), 1815-1824. doi: 10.1897/08-545.1

Napinski, C.S. (2009). Toxicity Analysis of Polycyclic Aromatic Hydrocarbon Mixtures (Doctoral Thesis). Texas A&M University Libraries.

Office of Solid Waste and EPA 542-R-05-013 Emergency Response (2005) Remediation System Evaluation Final Report. https://nepis.epa.gov/Exe/ZyNET.exe/P1000TJI.txt?ZyActionD=ZyDocument&Client=EPA&Index=2000%20Thru%202005&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&UseQField=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5CZYFILES%5CINDEX%20DATA%5C00THRU05%5CTXT%5C00000015%5CP1000TJI.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=x&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=1&ZyEntry=4

Olajire, A.A.; Essien, J.P. (2014). Aerobic Degradation of Petroleum Components by Microbial Consortia, Journal of Petroleum & Environmental Biotechnology, 5. doi: 10.4172/2157-7463.1000195

Ovalles, C.; Moir, M.E. (2018) *The Boduszynski Continuum: Contributions to the Understanding of the Molecular Composition of Petroleum*. ACS Symposium Series; American Chemical Society: Washington, DC. https://pdfs.semanticscholar.org/46ee/691068a1e7a964f13781e9df5ead462a9f77.pdf

Pan, Y.; Liao, Y.; Shi, Q.; Hsu, C.S. (2013) Acidic and Neutral Polar NSO Compounds in Heavily Biodegraded Oils Characterized by Negative-Ion ESI FT-ICR MS. *Energy & Fuels*, 27, 2960−2973. dx.doi.org/10.1021/ef400191h

Pan, Y.; Liao, Y.; Shi, Q. (2017). Variations of Acidic Compounds in Crude Oil during Simulated Aerobic Biodegradation: Monitored by Semiquantitative Negative-Ion ESI FT-ICR MS. *Energy & Fuels*, 31(2), 1126-1135. doi: 10.1021/acs.energyfuels.6b02167

Podgorski, D.C.; Corilo, Y.E.; Nyadong, L.; Lobodin, V.V.; Bythell, B.J.; Robbins, W.K.; McKenna, A.M.; Marshall, A.G.; Rodgers, R.P. (2013). Heavy Petroleum Composition. 5. Compositional and Structural Continuum of Petroleum Revealed. *Energy & Fuels,* 27(3), 1268-1276. doi: dx.doi.org/10.1021/ef301737f

Purcell, J.M. (2007). *Petroleum Analysis by Atmospheric Pressure Photoionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry* (Doctoral thesis). Florida State University Libraries

Purcell, J.M.; Merdrignac, I.; Rodgers, R.P.; Marshall, A.G.; Gauthier, T.; Guibard, I. (2010). Stepwise Structural Characterization of Asphaltenes during Deep Hydroconversion Processes Determined by Atmospheric Pressure Photoionization (APPI) Fourier Transform Ion Cyclotron Resonance (FT-ICR) Mass Spectrometry, *Energy & Fuels*, 24(4), 2257-2265. https://doi.org/10.1021/ef900897a

Reddy, B.M. (2013) *An Investigation of Deepwater Horizon Heavy End Environmental Transformation by High Resolution Detection and Isolation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry* (Doctoral thesis) Florida State University Libraries

Reddy, C.M.; Quinn, J.G. (1999). GC-MS Analysis of Total Petroleum Hydrocarbons in Seawater Samples After the North Cape Oil Spill. *Marine Pollution Bulletin*, 38(2), 126-135. https://doi.org/10.1016/S0025-326X(98)00106-4

Rodgers, R. P.; Schaub, T. M.; Marshall, A. G. (2005). Petroleomics: MS Returns to its Roots. *Analytical Chemistry*, 77(1), 20A−27A. https://doi.org/10.1021/ac053302y

Rodgers, R.P.; Mapolelo, M.M.; Robbins, W.K.; Chacón-Patiño, M.L.; Putman, J.C.; Niles, S.F.; Rowland, S.M.; Marshall, A.G. (2019). Combating selective ionization in the high resolution mass spectral characterization of complex mixtures. Faraday Discussions, 218, 29-51. doi: 10.1039/c9fd00005d

Rowland, S.M.; Robbins, W.K.; Corilo, Y.E.; Marshall, A.G.; Rodgers, R.P. (2014). Solid-Phase Extraction Fractionation To Extend the Characterization of Naphthenic Acids in Crude Oil by Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Energy & Fuels*, 28(8), 5043-5048. doi: https://doi.org/10.1021/ef5015023

Ruddy, B.M.; Huettel, M.; Kostka, J.E.; Lobodin, V.V.; Bythell, B.J.; McKenna, A.M.; Aeppli, C.; Reddy, C.M.; Nelson, R.K.; Marshall, A.G.; Rodger, R.P. (2014). Targeted Petroleomics: Analytical Investigation of Macondo Well Oil Oxidation Products from Pensacola Beach. Energy & Fuels, 28(6), 4043-4050. https://doi.org/10.1021/ef500427n

Ryerson, T.B.; Aiken, K.C.; Angevine, W.M.; Atlas, E.L.; Blake, D.R.; Brock, C.A.; Fehsenfeld, F.C.; Gao, R.-S.; de Gouw, J.A.; Fahey, D.W.; Holloway, J.S.; Lack D. A.; Lueb, R. A.; Meinardi, S.; Middlebrook, A.M.; Murphy, D.M.; Neuman, J.A.; Nowak, J.B.; Parrish, D.D.; Peischl, J.; Perring, A.E.; Pollack, I.B.; Ravishankara, A.R.; Roberts, J.M.; Schwarz, J.P.; Spackman, J.R.; Stark, H.; Warneke, C.; Watts, L.A. (2011). Atmospheric emissions from the Deepwater Horizon spill constrain air- water partitioning,

hydrocarbon fate, and leak rate. *Geophysical Research Letters*, 38(7). doi:10.1029/2011GL046726.

Savory, J.J.; Kaiser, N.K.; McKenna, A.M.; Xian, F.; Blakney, G.T.; Rodgers, R.P.; Hendrickson, C.L.; Marshall, A.G. (2011). Parts-per-billion Fourier transform ion cyclotron resonance mass measurement accuracy with a "walking" calibration equation. *Analytical Chemistry*, 2011, 83(5), 1732–1736. doi: 10.1021/ac102943z

Sepic, E.; Bricelj, M.; Leskovsek, H. (2003). Toxicity of fluoranthene and its biodegradation metabolites to aquatic organisms. Chemosphere, 52(7), 1125–1133. doi: 10.1016/S0045-6535(03)00321-7

Schrlau, J.E.; Kramer, A.L.; Chlebowski, A.; Truong, L. (2017). Formation of Developmentally Toxic Phenanthrene Metabolite Mixtures by Mycobacterium sp. ELW1. *Environmental Science & Technology*, 51(15), 8569-8578. doi: 10.1021/acs.est.7b01377

Sims, P.; Grover, P.L.; Swaisland, A.; Pal, K.; Hewer, A. (1974). Metabolic activation of benzo(a)pyrene proceeds by a diol-epoxide. *Nature*, 252, 326-328. doi: 10.1038/252326a0

Speight, J.G. 1991. The Chemistry and Technology of Petroleum. New York: Marcel Dekker. 2nd edition.

Stanford, L.A.; Kim, S.; Rodgers, R.P.; Marshall, A.G. (2006). Characterization of compositional changes in vacuum gas oil distillation cuts by electrospray ionization FT-ICR mass spectrometry. *Energy & Fuels*, 20(4), 1664–1673. https://doi.org/10.1021/ef060104g

Steenson, R.; Hellmann-Blumberg, U.; Elias, D.; Brown, K.; Fry, N.; Naugle, A.; Meillier, L.; Prowell, C. (2016). Petroleum Metabolites: Literature Review and Assessment Framework. San Francisco Bay Regional Water Quality Control Board. https://www.waterboards.ca.gov/rwqcb2/publications_forms/documents/SF_WB_Petroleum_Metabolites.pdf

Stiver, W.; Mackay, D. (1984) Evaporation rate of spills of hydrocarbons and petroleum mixtures. *Environmental Science & Technology*, 18(11), 834-840. doi: 10.1021/es00129a006

Teräväinen, M.J.; Pakarinen, J.M.H.; Wickström, K.; Vainiotalo, P. (2007). Comparison of the composition of Russian and North Sea crude oils and their eight distillation fractions

studied by negative-ion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry: The effect of suppression. *Energy & Fuels*, 21(1), 266–273. https://doi.org/10.1021/ef060294v

Tochko, L. (2018) *Processes Governing the Performance of Oleophilic Bio-barriers (OBBs) – Laboratory and Field Studies* (Master's thesis). Colorado State University Libraries.

US EPA, 1996. Clean Water Act 40 CFR Section 311 Oil and Hazardous Substances Part 110. https://www.epa.gov/sites/production/files/2014-04/documents/b_40cfr110.pdf

van Krevelen, D.W. (1950). Graphical-Statistical Method for the Study of Structure and Reaction Processes of Coal. *Fuel*, 29, 269-284.

Varel, U.L.; Machala, M.; Ciganek, M.; Neca, J.; Pencikova, K.; Palkova, L.; Vondracek, J.; Löffler, I.; Streck, G.; Reifferscheid, G.; Flückiger-Isler, S. Weiss, J.M.; Lamoree, M.; Brack, W. (2011). Polar compounds dominate in vitro effects of sediment extracts. *Environmental Science & Technology*, 45(6), 2384–2390. https://pubs.acs.org/doi/10.1021/es103381y

Vecchiato, M.; Bonato, T.; Bertin, A.; Argiriadis, E.; Barbante, C.; Piazza, R. (2017). Plant Residues as Direct and Indirect Sources of Hydrocarbons in Soils: Current Issues and Legal Implications, *Environmental Science & Technology Letters*, 4(12), 512−517. DOI: 10.1021/acs.estlett.7b00464

Wang, Z.; Fingas, M. (1997). Review: Developments in the analysis of petroleum hydrocarbons in oils, petroleum products and oil-spill-related environmental samples by gas chromatography. *Journal of Chromatography A*, 774(1-2), 51-78. doi: 10.1016/s0021-9673(97)00270-7

Wang, Z.; Stout, S.A.; Fingas, M. (2006). Forensic Fingerprinting of Biomarkers for Oil Spill Characterization and Source Identification. *Environmental Forensics*, 7(2), 105-146. doi: 10.1080/15275920600667104

Weisman, W. (1998). *Analysis of Petroleum Hydrocarbons in Environmental Media*. Total Petroleum Hydrocarbon Criteria Working Group Series. http://www.hawaiidoh.org/references/TPHCWG%201998.pdf

Xue, W.; Warshawsky, D. (2005). Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review. *Toxicology and Applied Pharmacology*, 206(1), 73-93. doi: 10.1016/j.taap.2004.11.006

Young, L.Y.; Phelps, C.D. (2004). Metabolic Biomarkers for Monitoring in Situ Anaerobic Hydrocarbon Degradation, *Environmental Health Perspectives*, 113(1), 62-67. doi: 10.1289/ehp.6940

Zemo, D.A.; O'Reilly, K.T.; Mohler, R.E.; Tiwary, A.K.; Magaw, R.I.; Synowiec, K.A. (2013). Nature and Estimated Human Toxicity of Polar Metabolite Mixtures in Groundwater Quantified as TPHd/DRO at Biodegrading Fuel Release Sites. *Groundwater Monitoring & Remediation*, 33(4), 44–56. https://doi.org/10.1111/gwmr.12030

Zemo, D.A.; O'Reilly, K.T.; Mohler, R.E.; Magaw, R.I.; Devine, C.E.; Ahn, S.; Tiwary, A.K. (2017). Life Cycle of Petroleum Biodegradation Metabolite Plumes, and Implications for Risk Management at Fuel Release Sites, *Integrated Environmental Assessment and Management*, 13(4), 714-727. https://doi.org/10.1002/ieam.1848

Zhan, D. L.; Fenn, J. B. (2000). Electrospray mass spectrometry of fossil fuels. *International Journal of Mass Spectrometry*, 194(2-3), 197-208. doi: 10.1016/S1387-3806(99)00186-4

APPENDIX

Table A1: Elemental formulas assigned by FT-ICR MS that were detected in all 15 samples by negative ESI.

N_1	O_1	O_2			
C20 H23 N1	C19 H18 O1	C15 H28 O2	C19 H24 O2	C23 H28 O2	C26 H42 O2
C21 H21 N1	C20 H18 O1	C16 H20 O2	C19 H26 O2	C23 H30 O2	C27 H24 O2
C21 H25 N1	C22 H16 O1	C16 H24 O2	C19 H28 O2	C23 H32 O2	C29 H36 O2
C21 H27 N1	C23 H18 O1	C16 H26 O2	C19 H30 O2	C23 H34 O2	C30 H30 O2
C22 H27 N1	C23 H20 O1	C17 H16 O2	C19 H32 O2	C23 H38 O2	C61 H72 O2
C22 H29 N1	C24 H20 O1	C17 H18 O2	C19 H34 O2	C23 H42 O2	
C23 H29 N1		C17 H20 O2	C19 H36 O2	C24 H20 O2	
C23 H31 N1		C17 H22 O2	C17 H26 O2	C24 H24 O2	
C24 H27 N1		C17 H30 O2	C19 H38 O2	C24 H26 O2	
C24 H31 N1		C17 H28 O2	C20 H16 O2	C24 H30 O2	
C24 H33 N1		C18 H16 O2	C20 H18 O2	C24 H34 O2	
C25 H33 N1		C18 H18 O2	C20 H20 O2	C24 H36 O2	
C25 H35 N1		C18 H20 O2	C20 H24 O2	C24 H38 O2	
C26 H35 N1		C18 H22 O2	C20 H26 O2	C24 H40 O2	
C26 H37 N1		C18 H24 O2	C20 H36 O2	C25 H20 O2	
C27 H37 N1		C18 H26 O2	C20 H38 O2	C25 H26 O2	
C27 H39 N1		C18 H28 O2	C22 H36 O2	C25 H34 O2	
C28 H39 N1		C18 H30 O2	C22 H38 O2	C25 H36 O2	
		C19 H14 O2	C22 H40 O2	C25 H38 O2	
N1 O1		C19 H16 O2	C23 H20 O2	C25 H40 O2	
C23 H21 N1 O1		C19 H18 O2	C23 H22 O2	C25 H42 O2	
C24 H23 N1 O1		C19 H20 O2	C23 H24 O2	C26 H22 O2	
C25 H25 N1 O1		C19 H22 O2	C23 H26 O2	C26 H26 O2	

O3					**O3 S1**
C11 H8 O3	C18 H26 O3	C21 H28 O3	C23 H26 O3	C25 H40 O3	C13 H12 O3 S1
C15 H18 O3	C18 H28 O3	C21 H30 O3	C23 H32 O3	C26 H22 O3	C17 H26 O3 S1
C15 H22 O3	C19 H18 O3	C21 H32 O3	C23 H34 O3	C26 H24 O3	C18 H28 O3 S1
C15 H24 O3	C19 H20 O3	C21 H34 O3	C23 H36 O3	C26 H26 O3	C19 H30 O3 S1
C15 H28 O3	C19 H22 O3	C21 H36 O3	C23 H38 O3	C26 H28 O3	C24 H36 O3 S1
C16 H14 O3	C19 H24 O3	C21 H38 O3	C23 H40 O3	C26 H30 O3	
C16 H16 O3	C19 H28 O3	C21 H40 O3	C24 H20 O3	C26 H32 O3	**O5**
C16 H20 O3	C19 H30 O3	C30 H30 O3	C24 H22 O3	C26 H34 O3	C19 H14 O5
C16 H22 O3	C19 H32 O3	C22 H16 O3	C24 H30 O3	C26 H36 O3	C22 H20 O5
C16 H24 O3	C19 H34 O3	C22 H18 O3	C24 H34 O3	C26 H38 O3	C24 H20 O5

C16 H26 O3	C19 H36 O3	C22 H20 O3	C24 H36 O3	C26 H42 O3	C24 H24 O5
C17 H14 O3	C19 H38 O3	C22 H24 O3	C24 H38 O3	C27 H26 O3	C25 H24 O5
C17 H18 O3	C20 H14 O3	C22 H26 O3	C24 H40 O3	C27 H28 O3	C26 H24 O5
C17 H22 O3	C20 H18 O3	C22 H28 O3	C24 H42 O3	C27 H30 O3	C26 H26 O5
C17 H28 O3	C20 H20 O3	C22 H30 O3	C25 H20 O3	C27 H40 O3	C27 H28 O5
C18 H16 O3	C20 H22 O3	C22 H32 O3	C25 H22 O3	C27 H42 O3	C28 H26 O5
C18 H14 O3	C20 H38 O3	C22 H34 O3	C25 H24 O3	C28 H28 O3	
C17 H24 O3	C20 H40 O3	C22 H36 O3	C25 H26 O3	C28 H32 O3	
C17 H26 O3	C21 H16 O3	C22 H38 O3	C25 H28 O3	C28 H36 O3	
C18 H18 O3	C21 H18 O3	C23 H18 O3	C25 H30 O3	C28 H42 O3	
C18 H20 O3	C21 H20 O3	C23 H20 O3	C25 H34 O3	C29 H36 O3	
C18 H22 O3	C21 H22 O3	C23 H22 O3	C25 H36 O3	C29 H42 O3	
C18 H24 O3	C21 H26 O3	C23 H24 O3	C25 H38 O3		

O4					
C15 H12 O4	C19 H30 O4	C21 H34 O4	C24 H20 O4	C26 H22 O4	C27 H34 O4
C15 H16 O4	C19 H32 O4	C21 H36 O4	C24 H22 O4	C28 H42 O4	C27 H36 O4
C15 H28 O4	C19 H34 O4	C21 H40 O4	C24 H24 O4	C29 H28 O4	C27 H38 O4
C16 H16 O4	C19 H36 O4	C22 H16 O4	C24 H26 O4	C29 H32 O4	C27 H40 O4
C16 H18 O4	C20 H16 O4	C22 H18 O4	C24 H36 O4	C29 H34 O4	C27 H42 O4
C17 H16 O4	C20 H18 O4	C22 H22 O4	C24 H38 O4	C28 H32 O4	C28 H26 O4
C17 H22 O4	C20 H20 O4	C22 H26 O4	C25 H30 O4	C28 H34 O4	C26 H24 O4
C17 H24 O4	C20 H22 O4	C22 H28 O4	C25 H32 O4	C28 H36 O4	C28 H28 O4
C17 H28 O4	C20 H24 O4	C22 H30 O4	C24 H28 O4	C28 H38 O4	C28 H30 O4
C17 H32 O4	C20 H26 O4	C22 H32 O4	C24 H30 O4	C28 H40 O4	
C18 H14 O4	C20 H34 O4	C22 H34 O4	C24 H32 O4	C29 H36 O4	
C18 H16 O4	C20 H36 O4	C22 H36 O4	C24 H34 O4	C29 H40 O4	
C18 H20 O4	C20 H38 O4	C22 H38 O4	C24 H40 O4	C29 H42 O4	
C18 H22 O4	C20 H40 O4	C22 H40 O4	C24 H42 O4	C30 H36 O4	
C18 H24 O4	C21 H14 O4	C23 H18 O4	C25 H20 O4	C26 H30 O4	
C18 H26 O4	C21 H16 O4	C23 H20 O4	C25 H22 O4	C26 H32 O4	
C18 H30 O4	C21 H18 O4	C23 H22 O4	C25 H24 O4	C26 H34 O4	
C19 H16 O4	C21 H20 O4	C23 H24 O4	C25 H26 O4	C26 H36 O4	
C19 H18 O4	C21 H22 O4	C23 H26 O4	C25 H28 O4	C26 H38 O4	
C19 H20 O4	C21 H24 O4	C23 H30 O4	C25 H34 O4	C26 H40 O4	
C19 H22 O4	C21 H26 O4	C23 H32 O4	C25 H36 O4	C26 H42 O4	
C19 H24 O4	C21 H28 O4	C23 H34 O4	C25 H38 O4	C27 H24 O4	
C19 H26 O4	C21 H30 O4	C23 H38 O4	C25 H40 O4	C27 H30 O4	
C19 H28 O4	C21 H32 O4	C23 H42 O4	C25 H42 O4	C27 H32 O4	

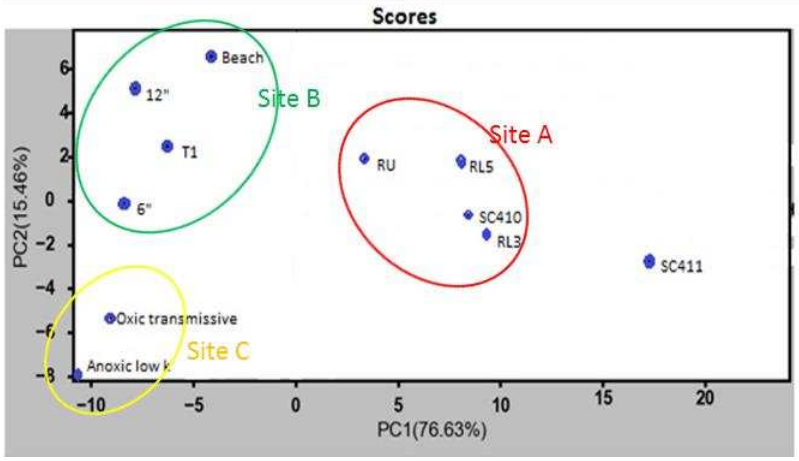

Figure A1: PCA score plot based on %RA for all samples analyzed by (-)ESI. Samples from within the same site generally clustered together for Sites A, B and C. (Site A: RL3, RL5, RU; Site B: T1, 6", 12", Beach; Site C: Anoxic low k, oxic transmissive)

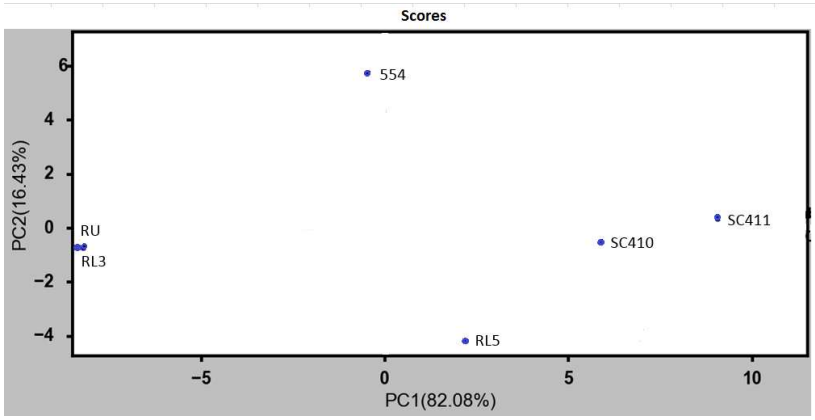

Figure A2: PCA score plot based on %RA for all samples analyzed by (+)ESI.

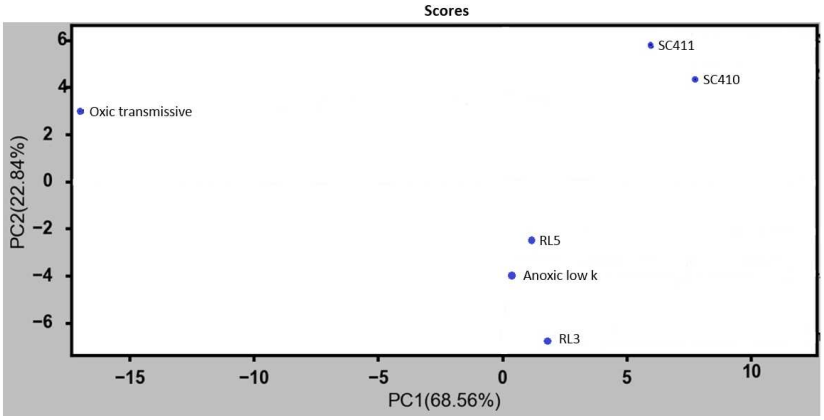

Figure A3: PCA score plot based on %RA for all samples analyzed by (+)APPI.