ORBITRAP AND IRMS PETROLEOMICS

A Thesis

by

DERRY XU

BS, University of Texas, 2018

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

CHEMISTRY

Texas A&M University-Corpus Christi
Corpus Christi, Texas

August 2021

August 2021

ORBITRAP AND IRMS PETROLEOMICS

A Thesis

by

DERRY XU

This thesis meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Richard Coffin, PhD
Chair

Hussain Abdulla, PhD
Committee Member

Jai Prakash, PhD
Committee Member

August 2021

ABSTRACT

Petroleum compounds from crude oil are deceptively complex; aside from the wealth of hydrocarbon compounds, crude oil may also contain other heteroatom compounds, inorganic compounds, and metals. It has been theorized that crude oil contains more compounds than genes in the human genome. Current petroleomic methods, such as the use of stable isotopes ($\delta^{13}C$, δD) as tracers for identification in an environment, would become superannuated. Advancements in analytical techniques such as Orbitrap mass spectrometry allow for the characterization of samples at resolutions at higher resolving powers and mass accuracies than conventional methods. As an initial proof of concept, 13 unique oil samples were retrieved from different processing stages and drilling environments. After these samples were processed via liquid-liquid extraction, they were analyzed using ultrahigh high-performance liquid chromatography (UPLC) coupled with Orbitrap Fusion mass spectrometer (OT-FTMS) by both electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI). I identified 19,000 compounds from the water extract fraction, most of these compounds detected in the water fraction are higher polarity heteroatom compounds than the hexane fraction. I identified a unique set of compounds (from 30 to 100s compounds) in each crude oil sample by applying volcano plot and principal component analyses. Identifying these unique compounds allows for the distinctive characterization of oil wells, spills, and processing methods. Our future goal is to apply these techniques to a wide variety of crude oil samples and form a global fuel library, using our research to protect oil companies (counterfeiting, spill responsibility) and the environment (complete spill cleanup, environmental interactions with crude oil).

ACKNOWLEDGEMENTS

I want to express my thanks to the petroleum companies (Headington Energy Partners, CITGO Petroleum of Corpus Christi, and Strategic Carbon) for providing me the crude oil samples and support for this project. I also want to acknowledge Thomas Boyd at the US Naval Research Laboratory for running the crude oil samples for the carbon isotope values.

I also want to thank my committee members (Richard Coffin, Hussain Abdulla, Jai Prakash) for their time, support, and advice in completing this project. Thank you all!

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES PAGE

LIST OF TABLES

TABLES PAGE

CHAPTER I - INTRODUCTION

Petroleum is a complex mixture of hydrocarbons composed of a variety of elements that vary between sources. High resolution and accurate analysis of petroleum composition provide unique "fingerprint" compounds that characterized each type of crude oil. Which will help track oil spill source assessment and monitoring downstream processing methods. The holistic goal of this project is to create a global petroleum library/database utilizing the capability of Orbitrap Fusion Tribrid Mass Spectrometric (OT-FT-MS) to separate and accurately identified thousands of compounds analyses. The development of this database for petroleum identification will impact several sectors in the petroleum industry, summarized by the acronym SAPOR (Survey, Account, Protect, Optimize, Research).

SURVEY--A modernized fuel database would streamline early quality assessments of crude oil, mitigating exploratory risks. One of the unique advantages of our UPLC-OT-FT-MS petroleomics method is its celerity; we can run crude oils with minimal processing (liquid-liquid extraction) compared to direct injection mass spectrometers. This allows for flexibility and risk mitigation at every stage of the crude oil development process, from upstream to downstream. Key information, such as heavy metal, sulfur, paraffin content, and heavy/light compound ratios are valuable to companies at every sector of the oil industry. Chemical analyses can be applied to wildcat and appraisal well samples, where speed matters for securing high-value reservoirs (e.g., Maalouf et al., 2017). Tests can be conducted during the midstream to verify oil composition before and after transport. The generated compound data is especially beneficial for the distribution and retailing side of the petroleum industry, where purchasers want to know the composition of their product.

ACCOUNT-- A modernized fuel database would promote accountability for hydrocarbons released in the environment. This database development provides a substantial understanding of past and present oil formation and fate. Oil spills and seepages remain a problem today, with slow leaks from existing wells and catastrophic events such as the Deepwater Horizon releasing 200 million gallons amount of petroleum (Joye 2015; Beyer et al., 2016) or the more recent 2019 Brazil oil spill (Zacharias et al., 2021; Soares et al., 2020). On land, seepages from gas stations or poorly maintained oil storage could occur, and fuels can leach into the terrestrial environment (Blake and Rowland 1995; Nganje et al., 2007). More robust methods for petroleum identification would contribute to impurities removal research with a deeper understanding of hydrocarbon sample composition. As we've discovered, a large portion of petroleum compounds is also miscible in pure water, which translates to conditions in the environment. Physical oil spill cleanup methods, such as booms, burns, and mechanical skimmers, may not be enough for environmental remediation, as the chemical impacts may still be prevalent. As a result, we should further evaluate spill, seepage, and leaching cleanup methods for their complete efficacy. It can track sample to source link by composition, and biomarkers can help absolve unwarranted blame for the release of hydrocarbons in the environment while also identifying the key players in accidents.

PROTECT-- A modernized fuel database will aid in fuel sourcing as a protection from counterfeiting and loss for energy companies. Petroleum can be profitable, but with that profit arises potential disputes. With the increase of commingling wells and well sharing agreements, it is logical that more disagreements ad lawsuits will arise over recovery ownership. Commingled production is a petroleum engineering process where a well can be designed to tap into more than one oil pool/reservoir, improving recovery efficiency and yields. This can add complexity to

certain legal agreements and ownership rights, especially with drilling operations involving well-sharing agreements or joint ventures. Fracking can also add complications to product ownership (under great disturbance, a pool's contents can migrate to a different reservoir).

Under specific mineral rights laws, the issue of horizontal wells bypassing boundaries or pipelines descending past certain depths can also arise. In addition, there are instances of theft. Individuals can tap into an owned reservoir illegally at leaks or set up illegal valves, or they can steal barrels of oil. Petroleum identification would be beneficial in all these instances, as it would provide chemical evidence of wrongdoing or absolution of guilt. Regardless of outcomes, a petroleomics database would aid in intracompany and governmental drilling/ownership disputes.

OPTIMIZE-- A modernized fuel database would optimize refining and distillation techniques while streamlining quality assessments of crude oil. When a well is drilled, and crude oil is recovered, there are at least 15 different products that the crude can be distilled and refined into, depending on the feedstock composition. A strong understanding of the feedstock chemical composition would optimize refinery and distillation techniques by detecting harmful compounds before affecting the processing systems (I.e., sulfur and sediment fouling/plugging, salts, etc.). Understanding overall chemical structures can aid in carbon rejection and hydrogen addition processes, such as various cracking, coking, and visbreaking methods, allowing for more valuable middle distillates or special petrochemicals. In addition, chemical composition data can help with performance diagnoses, troubleshooting, and maintenance.

RESEARCH-- A modernized fuel database can serve as a springboard for potentially novel research routes. Creating a fuel database with both molecular and isotopic data compiled for each sample would provide momentum for broad-spectrum paleo-geochemical assessments (Moldowan et al., 1985). Current paleontological methods involving kerogen analyses can be

further ascertained with crude oil biomarkers. In addition, Basafa and Hawboldt (2019) provide an overview of reservoir souring due to sea-water injection into offshore oil reservoirs.

CHAPTER II - BACKGROUND

Petroleum identification is not a new concept, but research groups have been running their samples with different methodologies and analytical instrumentation since its conception. The first large-scale effort towards identifying petroleum compounds was conducted from 1927-1959, dubbed "Project 6" by the American Petroleum Institute (Mair and Rossini, 1958). The project was able to identify 169 individual hydrocarbons isolated from petroleum, and the methodology to identify those hydrocarbons required manual and involved methods (without modern instrumentation or analyzers), such as distillation, selective solvent extraction, selective absorption, and crystallization. Since then, the identification of petroleum compounds has relied more on developments in chromatography, mass spectrometry, and isotope geochemistry. Today, labs still utilize instrumentation from these three analytical fields for petroleum analysis.

Isotope Ratio Mass Spectroscopy (IRMS) - Early applications of isotope ratios for petroleum analysis were first conducted in the 1960s by pioneers such as Silverman and Epstein (1958) and Eckelman et al. (1962). The primary use of petroleum stable isotope ratios was for origin formation studies or as tracers for oil/oil and oil/source-rock correlations that are resistant to the thermal evolution of hydrocarbons (Schoell 1984). Since then, applications of petroleum isotopes have extended to environmental forensics, where researchers such as Pond et al. (2002) utilize n-alkane and hydrogen isotope compositions for source identification of petroleum contamination. Carbon isotope analysis was used to fingerprint BTEX spill sources (Kelley et al., 1997) and confirm the degradation of oil (Coffin et al., 2008). Research has also extended past C/H isotopes and towards potentially toxic elements in petroleum, such as lead (Yao et al. 2015) or sulfur (Rosenberg et al., 2017). Isotope ratios are still utilized in the petroleum industry today to identify oil maturity, depositional environments, and sulfur content estimations. While relatively

sparse compared to GC or MS, there are still new developments on methodology, such as utilizing "untreated" crude oils for GC/IRMS analysis (Carrie et al., 2015). With new methods in compound-specific light element isotope analysis and thorough statistical analysis of the data (Boyd et al., 2006), there is strong potential to determine individual oil source contributions in complex environmental mixtures of natural and anthropogenic compounds.

Gas Chromatography (GC) - The use of GC was first introduced by researchers such as Desty and Whyman (1957), Eglinton et al. (1959), and Bray and Evans (1961), where petroleum compounds were able to fractionate with early, self-crafted gas-liquid chromatography systems. Researchers such as Bollet et al. (1981) and Boduszynski (1988) introduced the injection of petroleum samples into HPLC instrumentation for separation and analysis. Boduszynski (1988) bridged chromatography and MS methods for petroleum samples by utilizing field ionization mass spectroscopy (FIMS) after separation.

High-Resolution Mass Spectrometry (HRMS) - The first applications of mass spectrometry for petroleum analysis were with the MS-9 were developed by Gallegos et al. (1967) and Drushel and Sommers (1967), where they analyzed for petroleum compounds and sulfur compounds in petroleum, respectively. Since then, the developments have followed those in instrumentation: Quadrupole/triple quadrupole analyzers (Reynolds et al. 1991; Frysinger and Gaines, 1999; Eide and Zahlsen, 2005), GC-Q-TOF-MS (Qian and Dechert, 2002), and FT-ICR-MS (Qian et al., 2001). In 2008, steps were taken by Marshall and Rodgers to build a petroleum compounds database with a custom FT-ICR-MS capable of higher resolutions. Since then, the "newly born" field of petroleomics has expanded with more studies making applications towards environmental detection (Kelley et al., 1997; Wang et al. 2013).

One of the tenets of instrumental analysis is that no one instrument is the best for every scenario, for every instrumentation setup will have its strengths and weaknesses. What we have shown in our research is a synthesis of the three analytical methods mentioned: we utilized Liquid chromatography for separation and OT-FT-MS for mass/charge analysis. Petroleum identification is not a new concept, but with our instrumentation and methods, we hope to transcend the limitations of other modern setups with greater accuracy and reliability. Until now, no technique provides the ability to accurately develop a global library with absolute confidence in source determination.

Chemical Background - The hydrocarbons in petroleum can be chemically characterized as paraffin (straight chain), cycloparaffins/naphthene (rings), or aromatics. Some current identification techniques utilize the physical properties of petroleum (density, viscosity, pour point) with calculations (correlation index, UOP characterization, viscous-gravitational constant) to classify and identify petroleum based on the primary hydrocarbon type. Metadata and reservoir sourcing are other valuable method for the classification of petroleum. Classification schemes such as the Department of Energy' tertiary-oil-recovery-info-system' (TORIS) allow for the classification of reservoir/extract ID based on lithology, depositional environment, structural definition, diagenetic overprint—all of which impact crude oil composition. In other words, how petroleum was formed in the environment by Earth's processes impacts the overall composition of the reservoir, further contributing to its complexity. In addition to source rock contributions, downstream refining methods and petrochemical processing can contribute to crude oil composition. From this study, chemical engineering cavitation has shown slight changes to the original hydrocarbon sample. Heavy metals, sulfur compounds, elemental ratios, chemical

structure ratios, and metadata on petroleum additives are all potential indicators for fingerprinting petroleum samples.

Compound-specific isotope analysis (CSIA) can be applied to crude hydrocarbons as tracers relatively unaltered from the development of hydrocarbons. As a result, petroleum stable isotopes such as $\delta^{13}C$ can be used for oil-oil and oil-rock correlations. We believe that we are the first to take this isotopic data and correlate it with conventional mass spectrometry data.

In the analysis of complex samples such as petroleum hydrocarbons, it is essential to identify both the strengths and limitations of the analytical instrumentation involved. For instance, the ionization method selected can contribute to the mass feature's appearance (or disappearance) on the final spectra generated by a mass spectrometer. Therefore, the ionization method is paramount; electronic spray ionization (ESI) is best suited for ionizing polar compounds (e.g., heteroatoms), but ESI is less effective for non-polar compounds. Suppose one wanted to identify the bulk of a crude oil sample. In that case, the hydrocarbon portion (CxHy) of a sample APPI is the preferred ionization method, but compounds with higher molecular weights are more challenging to ionize with this method. APCI shares similarities with APPI and ESI and is best suited for ionizing condensed aromatics and semi-polar compounds.

Coupling chromatography instrumentation (GC/LC) is another point of consideration—gas chromatography (GC) is volatility dependent and, liquid chromatography (LC) is solubility/polarity dependent. A chromatographic method is essential for petroleomic studies because the separation over time allows for defined peaks (and thus structural identification) in a "compound-dense" sample.

When it comes to mass spectrometers, there are three main mass analyzer instrument types used for modern analytical chemistry: Q-TOF, FT-ICR, and Orbitrap. Q-TOF has limited

resolution (<60,000 resolving power), but it has rapid scan speeds and robust detection due to high kinetic energy. FT-ICR has a much higher resolution for molecular assignments, but its scan speed (1Hz) complicates the attachment of GC/LC instrumentation, which is essential for filtering compounds. The OT-FT-MS takes mass resolutions similar to the FT-ICR and combines it with faster scan speeds to allow for the use of GC/LC-- crucial for peak separation when there are so many individual compounds in an aliquot of oil.

CHAPTER III - SAMPLES AND DATA BACKGROUND

Sample Information and Processing

Figure 1-- The Cavitation sample group

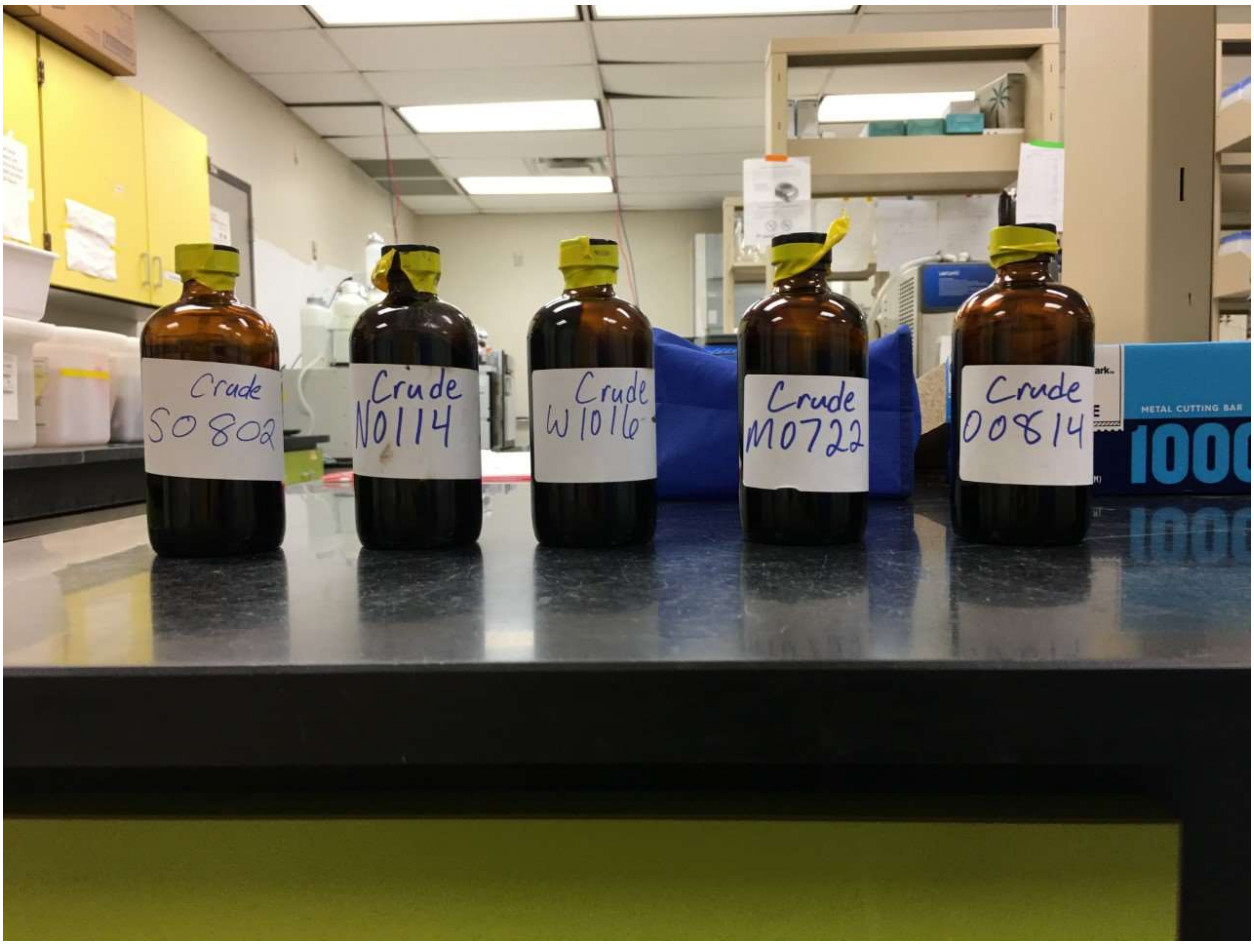

Figure 2-- The Citgo sample group

Figure 3-- The Headington sample group

This project utilizes 14 hydrocarbon samples gathered from three different companies/sources (Cavitation, Citgo, and Headington). The sample photos are indicated above, and currently, only the Headington samples have available metadata for further conclusions with data. Regarding the Headington samples, all the samples were drilled in Kennedy, Texas, in two different fields (Mifflin and Sarita). The "Diesel Mud Base" sample is a constituent of Sarita Sogu 402.

Analysis Methods and Specifications

Sample Processing prior to OT-FT-MS Analysis

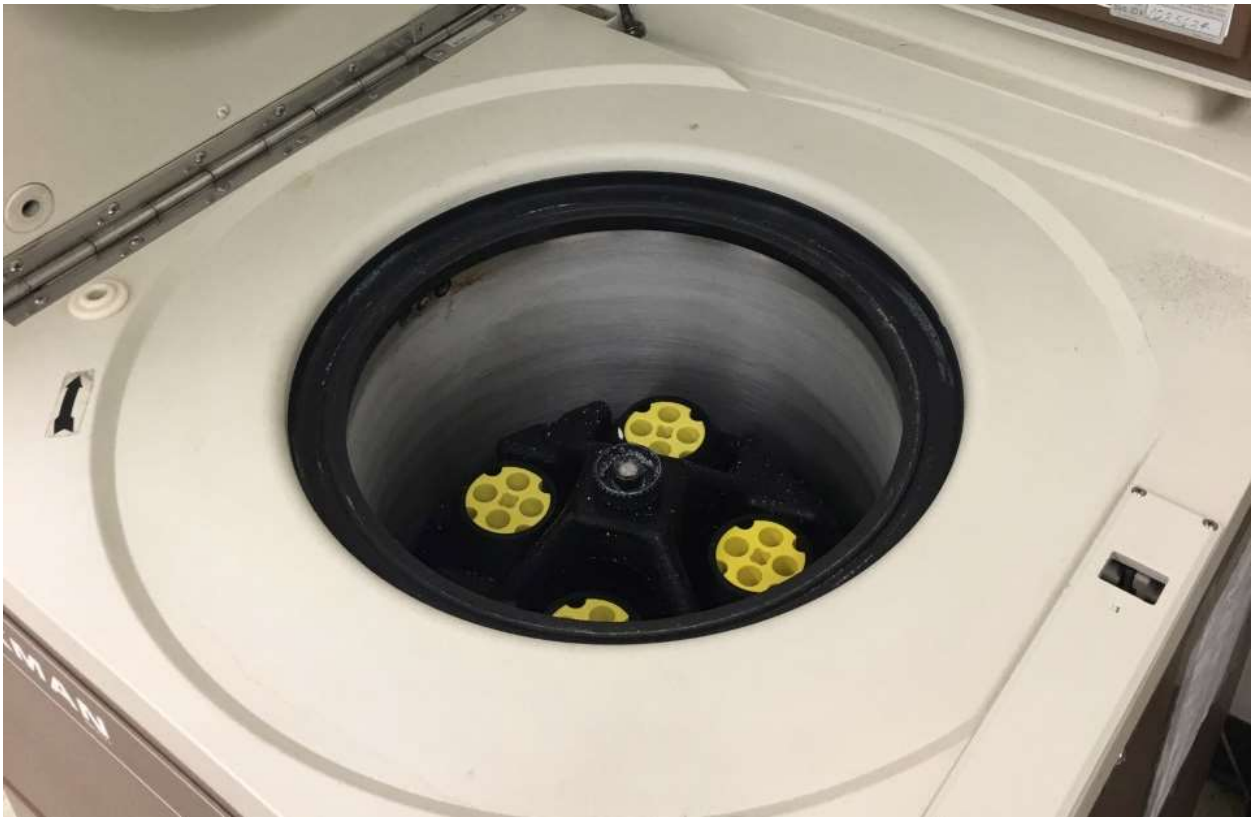

Figure 4-- The Beckman Coulter J2-21 centrifuge used for sample separation

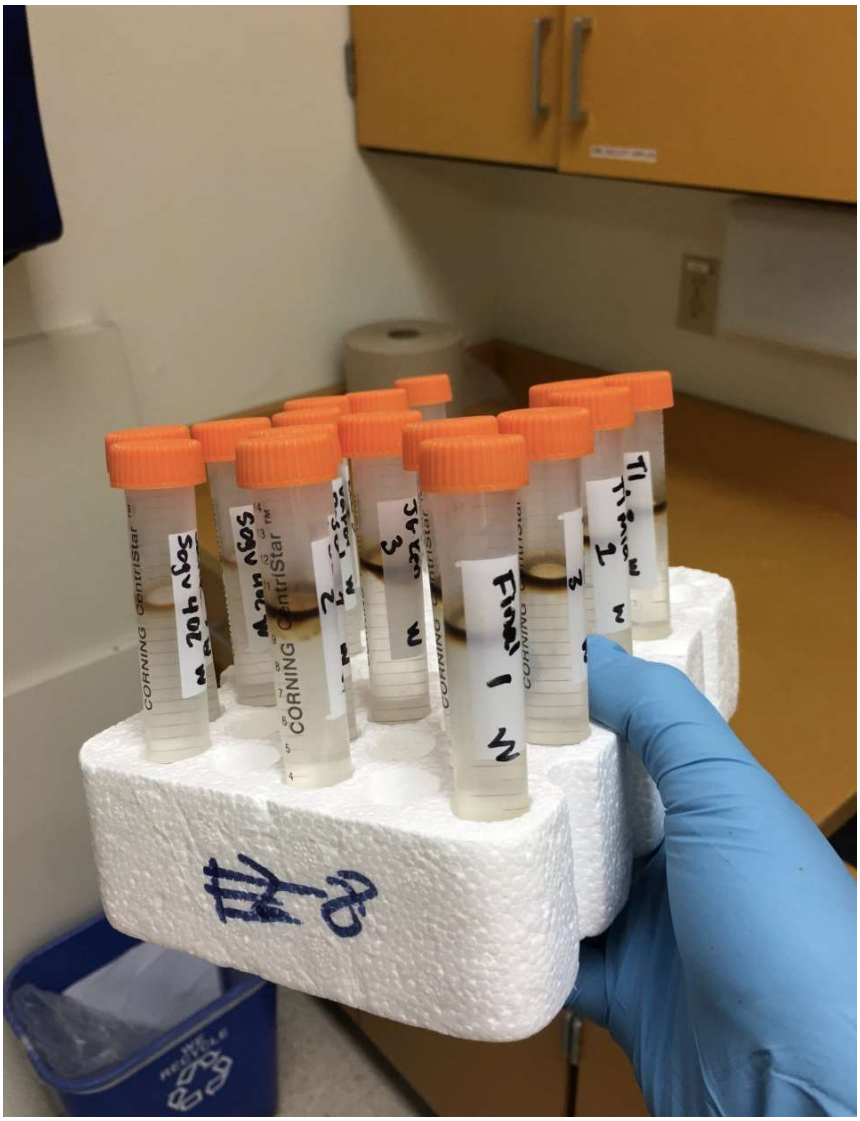

Figure 5-- Example of water-fraction samples and the separated layer

Before running petroleum samples on the OT-FT-MS, the samples must undergo processing to ensure efficient run conditions and prevent contamination. We can separate the polar and non-polar compounds of a petroleum sample through liquid-liquid extraction: I used Milli-Q water to isolate polar compounds and hexane (FisherSci Optima) to isolate non-polar compounds. To ensure the complete extraction of the respective compounds to the solvent, we centrifuged the samples at 15,000 xg RCF for 15 minutes in a Beckman Coulter J2-21 centrifuge

at 10,400xg. Extra care against contamination was taken by triple-washing all apparatus (centrifuge tubes, glassware) with the appropriate solvent before experimentation.

UPLC-APCI- Orbitrap Fusion Mass Spectrometer Analysis

The hexane extract was dried using CentriVap Vacuum Concentrators and re-dissolved in hexane. 3 μL of hexane extract was injected on Vanquish ultraperformance liquid chromatography (UPLC) coupled with Orbitrap Fusion Tribrid Mass Spectrometer (UPLC-OT-FTMS). The analysis was performed on the UPLC silica column (1.7 μm BEH HILIC column 2.1 x 150 mm) and was maintained at 30 ^{o}C via atmospheric pressure chemical ionization (APCI). Eluent A was hexane, and eluent B was hexane: isopropanol (9:1, v/v). The following gradient was used: hold at 18% B for 25 min; ramp to 35% B for 25 min; ramp to 100% B for 30 min and hold for 10 min. A 20 min column re-equilibration with the starting ratio of eluents was carried out between sample analyses. The flow rate used was 0.2 ml·min ^{1}. The APCI set at 5μA corona current, 35 Sheath gas, 10 Aux gas, 325^{o}C ion transfer tube temp, and 200^{o}C vaporizer temp. The Orbitrap full scan was run at full scan mode at 500,000 (FWHM at m/z 200) resolutions with a scan range of 50-1000 m/z and RF Lens at 40%. For MS^2, the isolation window was set at 0.7 m/z with performing both collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD) using an ion trap mass spectrometer as the detector. The AGC was set at 1.0e4 and the intensity threshold at 5.0e3.

UPLC-ESI- Orbitrap Fusion Mass Spectrometer

Water extract was analyzed on Vanquish UPLC - Orbitrap Fusion Tribrid Mass Spectrometer (UPLC-ESI-OT-FTMS). The analysis was performed on a 1.7 μm ACQUITY

UPLC BEH C18 reversed-phase column (Waters, 30Å, 1.7 µmol·L^{-1}, 2.1 mm X 100 mm) via a heated ESI (H-ESI) source to Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific) operated under positive mode. Eluent A was Milli-Q water with 0.1% (v/v) formic acid, and eluent B was acetonitrile with 0.1% (v/v) formic acid. The following gradient was used: hold at 5% B for 2 min; ramp to 65% B for 16 min; ramp to 100% B for 7 min and hold for 8 min. An 8 min column re-equilibration with the starting ratio of eluents was carried out between sample analyses. The flow rate was 0.2 ml·min^{-1} with an injection volume of 5 µL. The H-ESI setting was set at 3200 volts, 30 Sheath gas, 10 Aux gas, 325ºC ion transfer tube temp, and 200ºC vaporizer temp. The Orbitrap full scan was run at 500,000 resolutions with a scan range of 50-1000 m/z and RF Lens at 40%. For MS^2, the isolation window was set at 0.7 m/z with performing both collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD) using an ion trap mass spectrometer as the detector. The AGC was set at 1.0e4 and the intensity threshold at 5.0e3.

Carbon Isotope Ratios

Figure 6-- Florosil column to extract n-alkanes

The sample pre-processing method for isotopic analysis of n-alkanes was a urea adduction method derived from Pond et al. (2002). Urea crystals were used to filter out cyclic and highly branching chemical structures while letting straight-chain (or rarely-branching) structures such as alkanes flow through the urea crystals. The main modifications to the Pond et al. (2002) method are instrumentation and materials; a florosil column was used instead of silica gel to remove more UCMs (unresolved complex mixtures) that would otherwise cloud the C11-C23 alkane regions within the gas chromatogram data without compromising the $\delta^{13}C$ of alkanes

(Ellis and Fincannon 1998). Samples were sent over to the US Naval Research Lab to run the processed samples on a Thermo 1310 GC with an Isolink II and Delta V Advantage (IR-MS). Conformational analysis was also conducted on an Agilent GC/MS.

Data Processing

Compound Discoverer 3.2. (CD) Thermo Fisher Scientific software allows for the data generated from an OT-FT-MS to be delivered to a data processing environment to be organized. Scripts are written and customized within the program to initially identify compounds and generate broad comparisons between samples (Li 2018). A combination of Excel and Python 3.9 was used to filter and clean datasets for each sample.

Visualization

The visualization of samples allows for better analysis and comparisons between sample compounds. The main figures for this project are PCA analyses, Volcano Plots, Van Krevelen diagrams, and stable isotope graphs of alkanes (Kind and Feihn 2007; Li 2012; Koch and Dittmar 2016). The primary tools used for figure generation were Excel and Python 3.9, and the stable isotope figures were created using Sigma Plot 10.2. The following section displays all visualizations split between the two different ionization methods used, ESI and APCI.

PCA Analyses

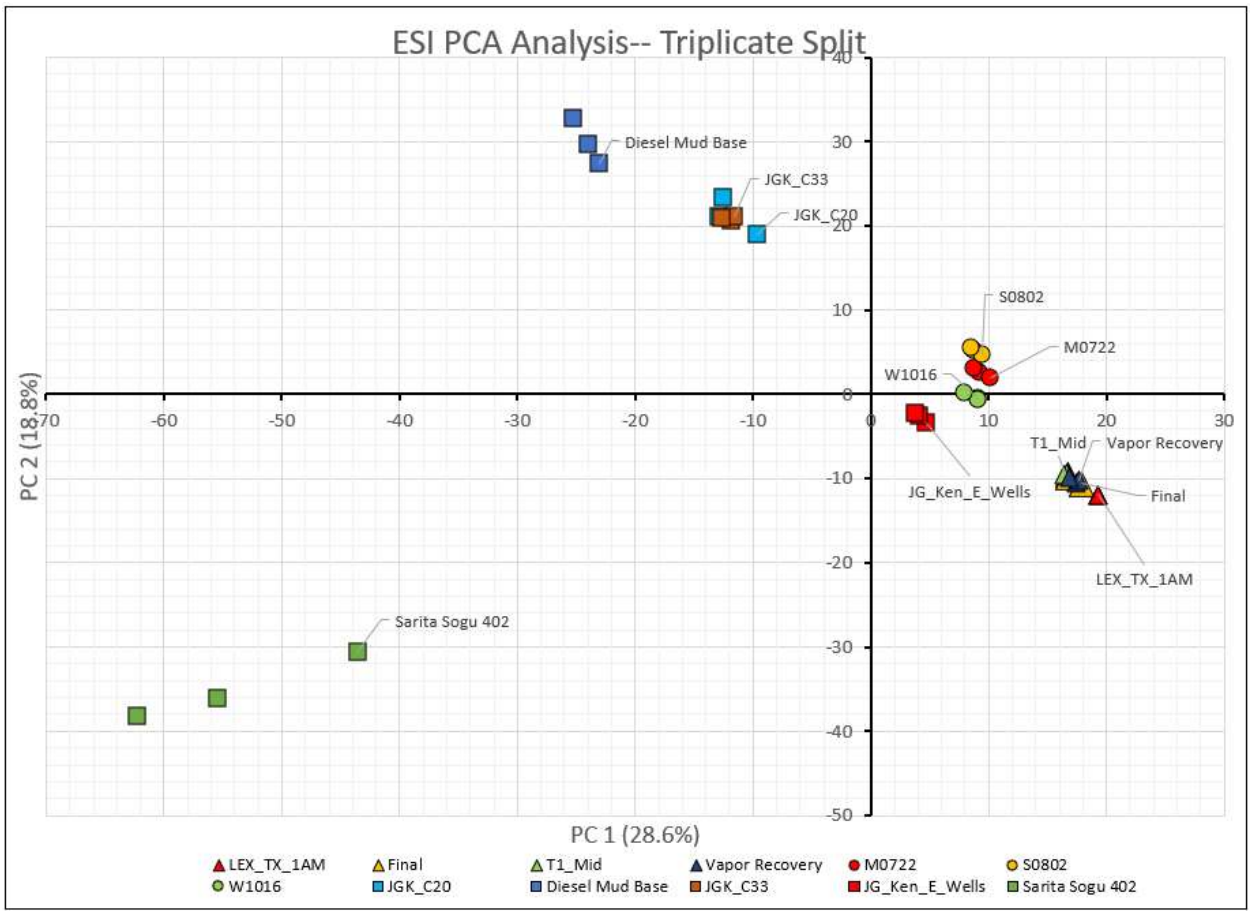

Figure 7-- PCA Analysis of ESI samples, separated by triplicate

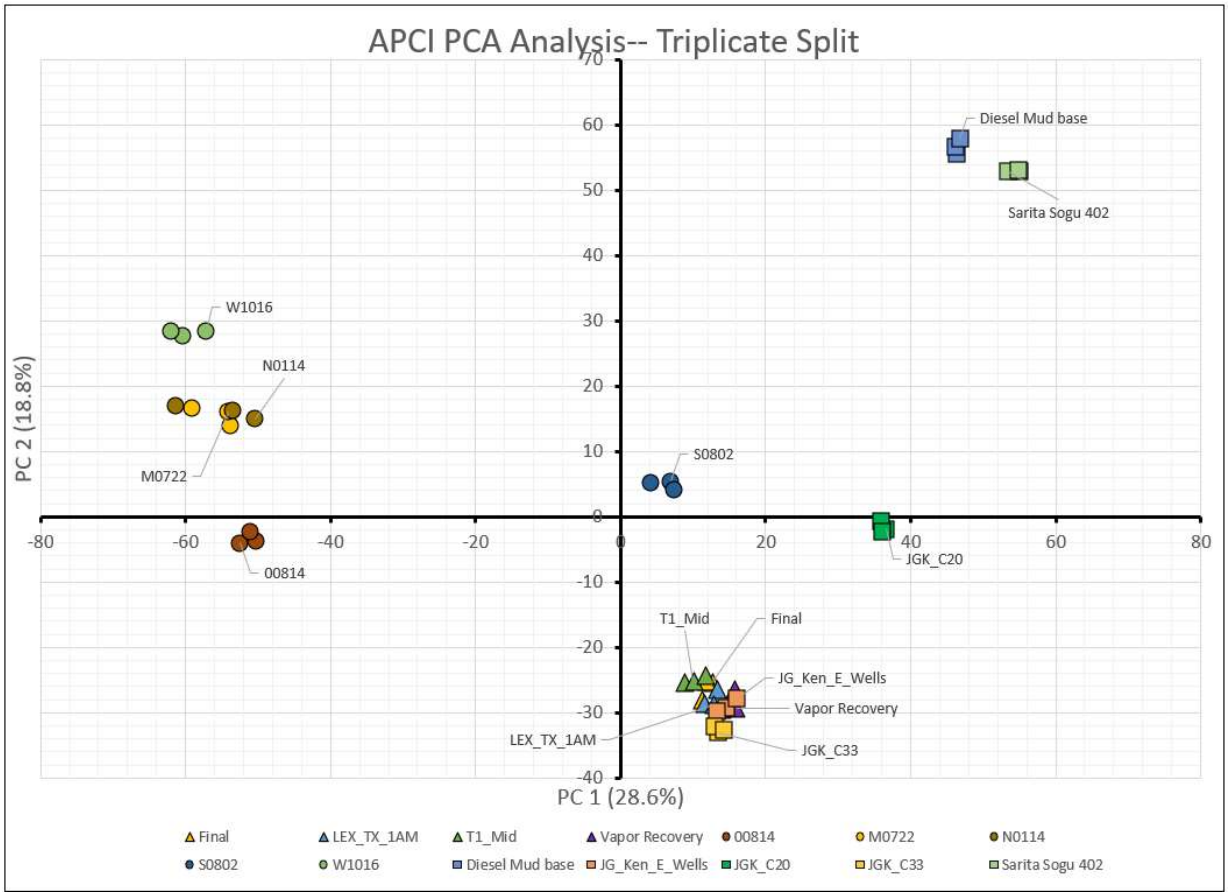

Figure 8-- PCA Analysis of APCI samples, separated by triplicate

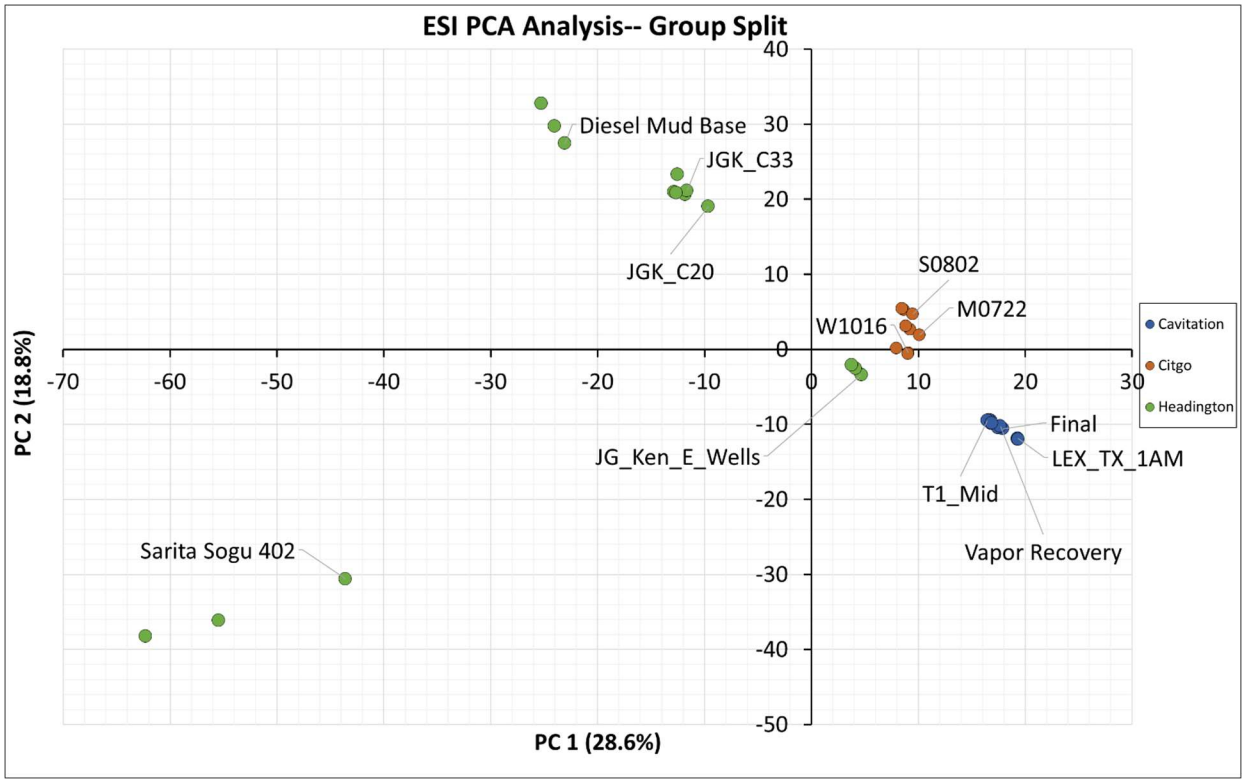

Figure 9-- PCA analysis of ESI samples, separated by sample group

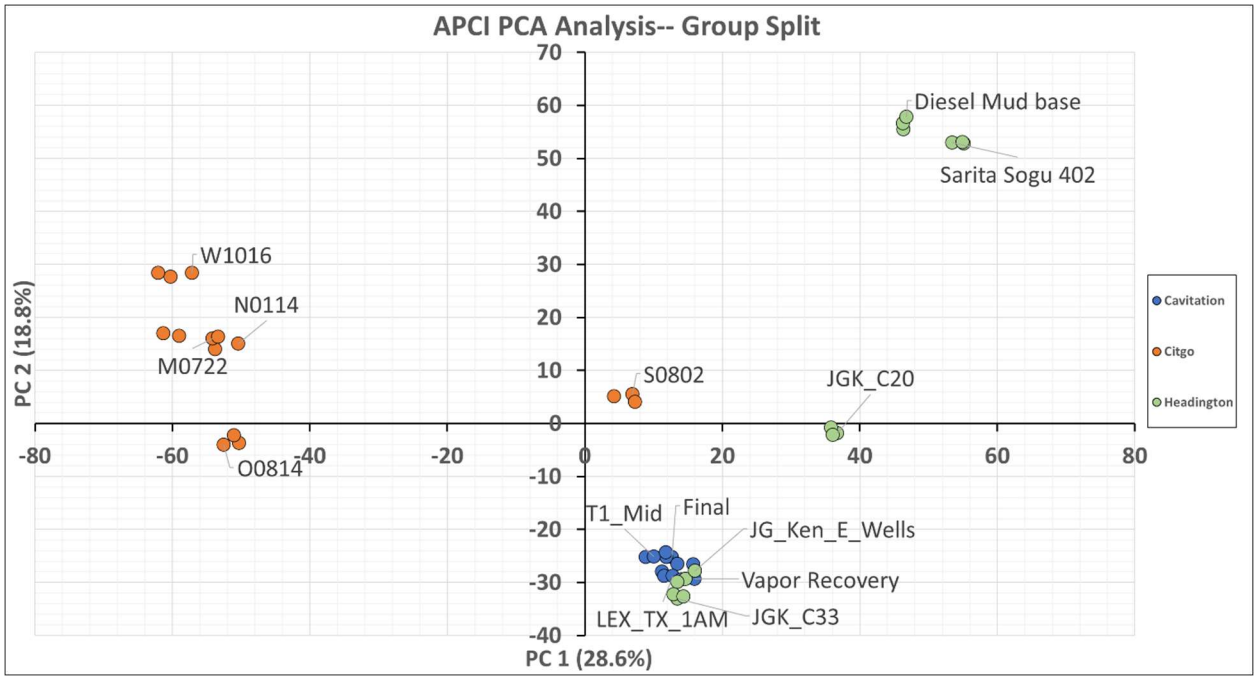

Figure 10-- PCA analysis of APCI samples, separated by sample group

Principle Component Analysis/ Loading Figure

PCA figures/ loading figures are primarily used to disseminate high dimensionality data into a group of similar samples. Samples (points) closer to one another on the graph have more similar chemical compositions than samples further away on the PCA figure. This relationship is extended in a loading figure, where every point represents a compound instead. Points closer to the center of a PCA figure also indicate more average/common properties in chemical composition. PCA figures are primarily used in life science studies with high dimensionality datasets (e.g., computational biology, metabolomics), but we believe we can extend these figures to the field of petroleomics due to the complexity of the petroleome. Attached are PCAs of our crude oil samples organized by ionization method and sample source (groupings). Triplicate analyses also indicate the reproducibility of our sample processing and instrumentation.

The PCA figures above indicate the robust method and instrumentation reproducibility, except for the Sarita Sogu 402 sample ionized by ESI. The crude oil samples can be grouped by

company, but more metadata is required to make remarks on individual samples. There seems to be a significant "clump" for the APCI PCA analysis, indicating intergroup similarities for the non-polar compounds of crude oil samples. Diesel mud base is a constituent of Sarita Sogu 402, which is evident on the ESI PCA but not the APCI PCA.

Volcano Plot

Volcano plots are specialized scatterplots that are best used for the comparison of two individual datasets. Volcano plots provide information on both the commonality and unique differences between datasets (Wentian et al., 2012). Data points (or, in this case, compounds) near the vertical axis of a volcano plot indicate similar/common compounds between the two datasets. In contrast, points further away from both axes are points/compounds unique to that dataset. Attached are volcano plots of our crude oil samples, grouped by the company/organization of origin. The groups were further color-coded by primary colors so that the "common" compounds could be logically indicated by secondary colors.

From the volcano plots created for each permutation of the three groups (Cavitation, Citgo, Headington sample groups), the most noticeable volcano plot was the Cavitation vs. Headington plot, where few unique compounds for Headington are found. A larger data pool for one side of the volcano plot (cumulative number of compounds) can be used to compare new sample groups in future assessments.

Van Krevelen

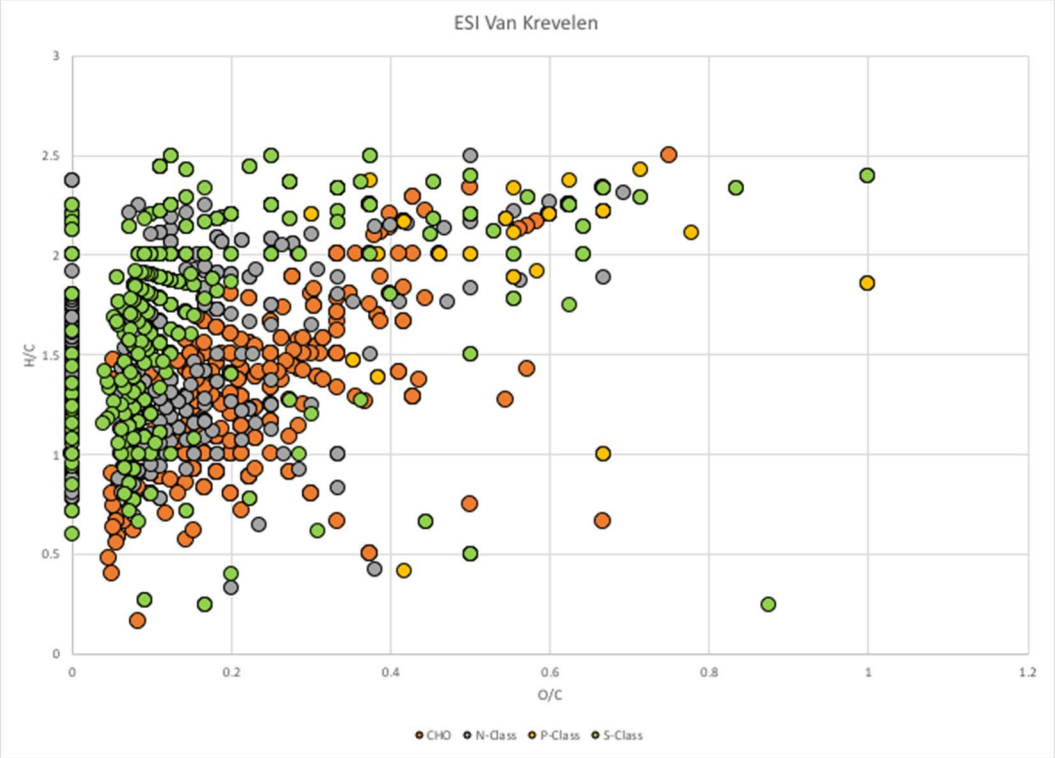

Figure 11-- Van Krevelen figure of the ESI samples

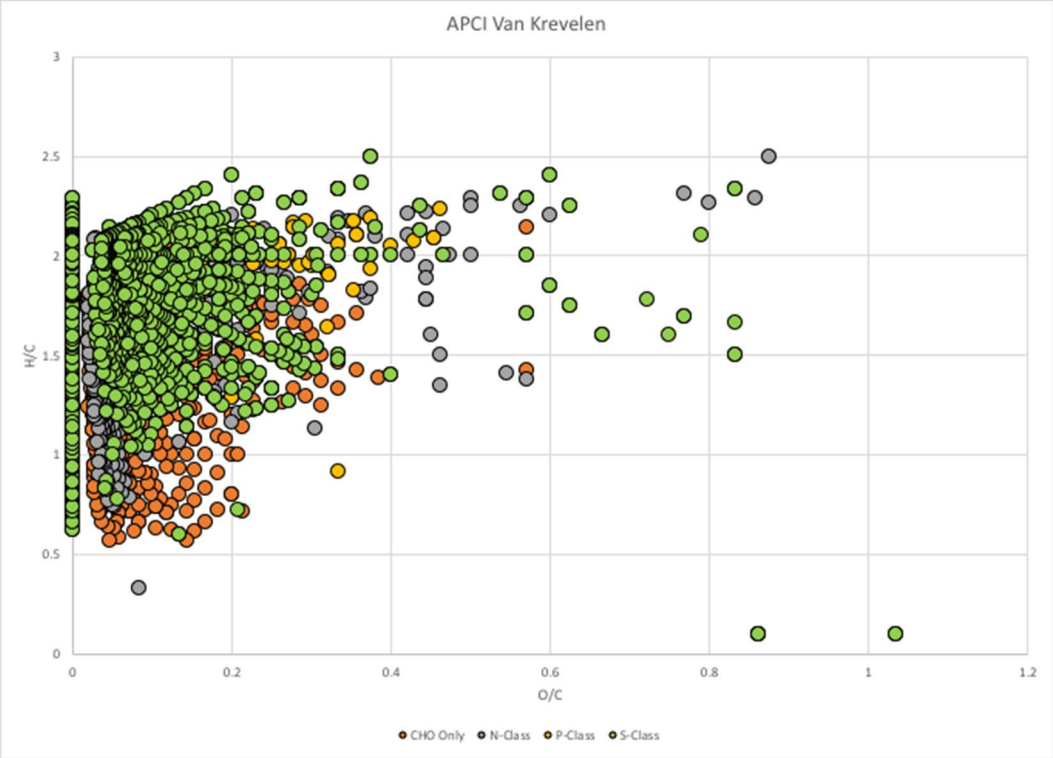

Figure 12-- Van Krevelen figure of the APCI samples

Originally developed to study the structure and reaction processes of coals and kerogens (van Krevelen, 1950), the Van Krevelen diagram has proven useful in visualizing complex spectra based on a set of core elements (Kim et al. 2003). In the case of crude oil, the core set of elements are C, H, O, N, P, S, and they are utilized in the VK diagrams shown above.

From the VK diagrams shown, APCI reveals more compounds than ESI. This is to be expected from hydrocarbons, which are primarily non-polar.

Isotopes

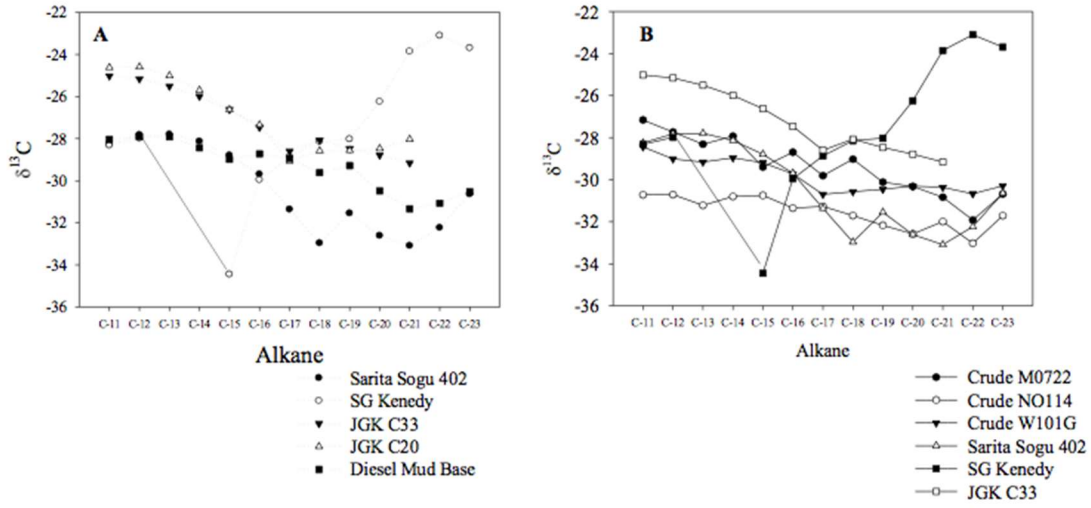

Figure 13-- Carbon isotope data of C11-C23 alkanes

While the OT-FT-MS can generate larger amount of quantitative data for a petroleum sample, isotope analyses of petroleum still have their uses in petroleomics. The study conducted by Wang et al. 2013 demonstrates a strong agreement between source and sample in isotopic readings of degraded oil spill material, and the weathered oil residue in the environment deviated no more than 0.36% from the oil tank sample (Wang et al. 2013). Studies like Wang et al. 2013 are a strong argument for isotopes as a tool for sample-source verifications, especially with heavily weathered samples.

However, based on the carbon isotope ratio data generated for the Headington and Citgo sample groups, we have decided to omit further analysis after the preliminary data generated above. The data generated by OT-FT-MS methods is more robust and can lead to stronger inferences than the isotopic information displayed in Figure 13. In the case of our research scope, we do not have any weathered samples yet, but Hydrogen and Carbon CSIA may be worth revisiting when the fuel database is expanded.

Data Table Information

In our data analysis, we tried to apply the statistical observations recognized in mass spectrometry literature (Kind and Feihn, 2007; Koch and Dittmar, 2006). Thus, the data tables includes the calculation of parameters such as aromaticity index (AI), degree of unsaturation (DU), double bond equivalent (DBE), and all their relevant modifiers. The tables provide the averages, highest, and lowest values of these parameters.

CHAPTER IV - DATA ANALYSIS

Loading Graphs/PCA VK

APCI Loading

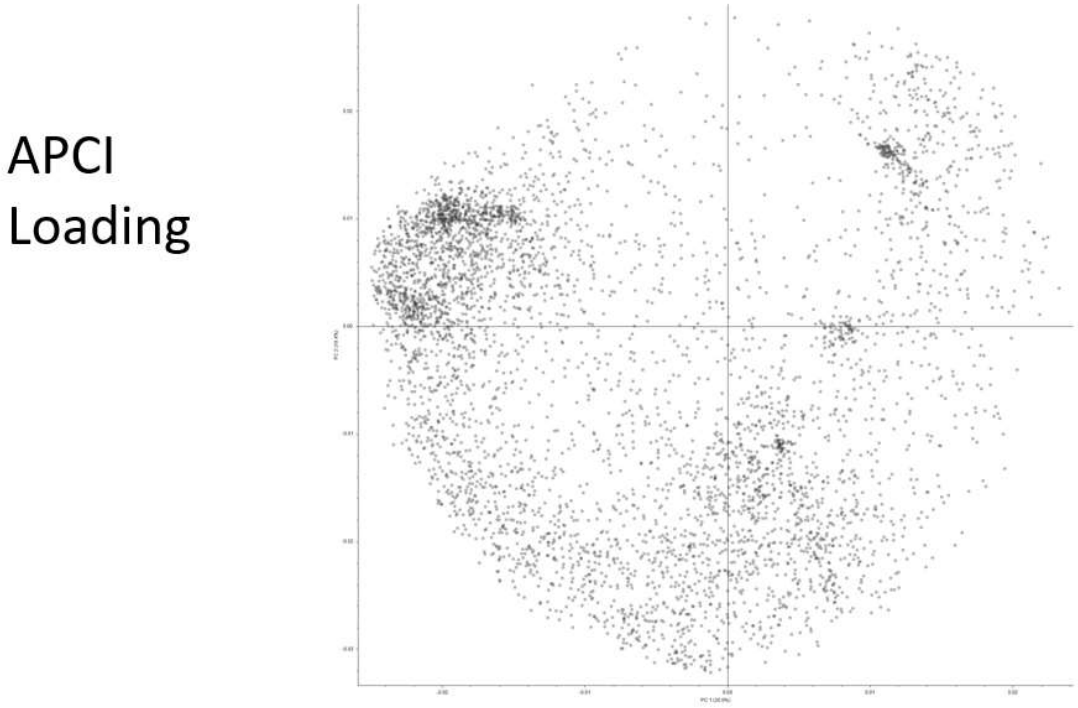

Figure 14-- APCI loading figure generated from Compound Discover

In this loading graph figure, the Citgo samples are isolated from the center and are thus collectively unique, except sample S0802. Sample S0802 is the least viscous sample in the group and the closest triplicate to the center. The Citgo APCI samples are also spread apart, indicating distinct inter-group compounds. It is also important to note that the negative side of the APCI loading graph is only occupied by the Citgo samples, further distinguishing the non-polar compounds of the Citgo sample group.

The Headington sample group has a large spread spanning the entirety of the vertical axis with groups of samples concentrating at the ends (Upper: Diesel Mud Base and Sarita Sogu 402; Lower: JGK C33 and JGK E Wells). Diesel mud base is a derivative of Sarita Sogu 402, so their

loading positioning based on their non-polar composition is logical. JGK E Wells and JGK C33 are similarly based on similar drilling environments (see metadata table).

The Cavitation sample group has the smallest spread of the three sample groups. The sample from the cavitation sample group (Mid, Final, Vapor Recovery) indicate steps of the cavitation process (with the original sample being Lex_TX_1AM).

APCI Loading Top Right

Top Right	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	4%	447	134	0	89	223	224

APCI_Loading_Top_Right	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	224	194	23	-1	18
% of total	100%	87%	10%	0%	8%
# of AI>0.5	46	44	1	0	2
# of AI>=0.67	9	8	1	0	1
# of AI-MOD>0.5	54	53	0	0	1
# of AI-MOD>=0.67	13	13	0	0	0
# of Carbons	3845	3342	398	0	215
Average O/C	0.112456554	0.101032548	0.221363552	#DIV/0!	0.323360806
Average H/C (All)	1.384811203	1.307212975	2.010226327	#DIV/0!	1.840845266
Average H/C (Has O/C)	1.376162914	1.307212975	1.995327247	#DIV/0!	2.018869464
Average Aromaticity+ Index	0.270015223	0.336932767	-0.321423135	#DIV/0!	-0.200145606
Average AI-Mod+	0.376707909	0.384031844	0.029437229	#DIV/0!	0.277135563
Average DU	5.986607143	6.536082474	1.739130435	#DIV/0!	2.833333333
Average DU/C	0.380125533	0.408795898	0.15369298	#DIV/0!	0.233570134
Average DBE	6.165178571	6.536082474	2.869565217	#DIV/0!	5.055555556
Average DBE/C	0.396540294	0.408795898	0.271091524	#DIV/0!	0.437842713
Average MW	262.7120791	257.0810896	318.032213	#DIV/0!	240.7980722
Average Carbon Count	17.16517857	17.22680412	17.30434783	#DIV/0!	11.94444444
High O/C	0.833333333	2	2.363636364	0	8
High H/C (All)	2.363636364	2	2.363636364	0	2.363636364
High H/C (Has O/C)	2.363636364	2	2.363636364	0	2.363636364
High Aromaticity Index	3.5	0.75	3.5	0	3.5
High AI-Mod	0.769230769	0.769230769	0.03030303	0	0.538461538
High DU	12	12	4	0	8
High DU/C	0.785714286	0.785714286	0.428571429	0	0.571428571
High DBE	12	12	6	0	10
High DBE/C	0.785714286	0.785714286	0.571428571	0	0.714285714
High MW	0	0	0	0	0
High C	42	42	38	0	24
Low O/C	0.027777778	0.034482759	0.027777778	0	0.041666667
Low H/C (All)	0.571428571	0.571428571	1.636363636	0	1
Low H/C (Has O/C)	0.571428571	0.571428571	1.636363636	0	1.636363636
Low Aromaticity Index	-4	-0.333333333	-4	0	-4
Low AI-Mod	0.024390244	0.024390244	0.028571429	0	0.083333333
Low DU	0	1	0	0	0
Low DU/C	0	0.047619048	0	0	0
Low DBE	0	1	0	0	2
Low DBE/C	0	0.047619048	0	0	0.181818182
Low MW	0	0	0	0	0
Low C	6	7	6	0	6

Table 1-- APCI top right loading data, separated by compound class

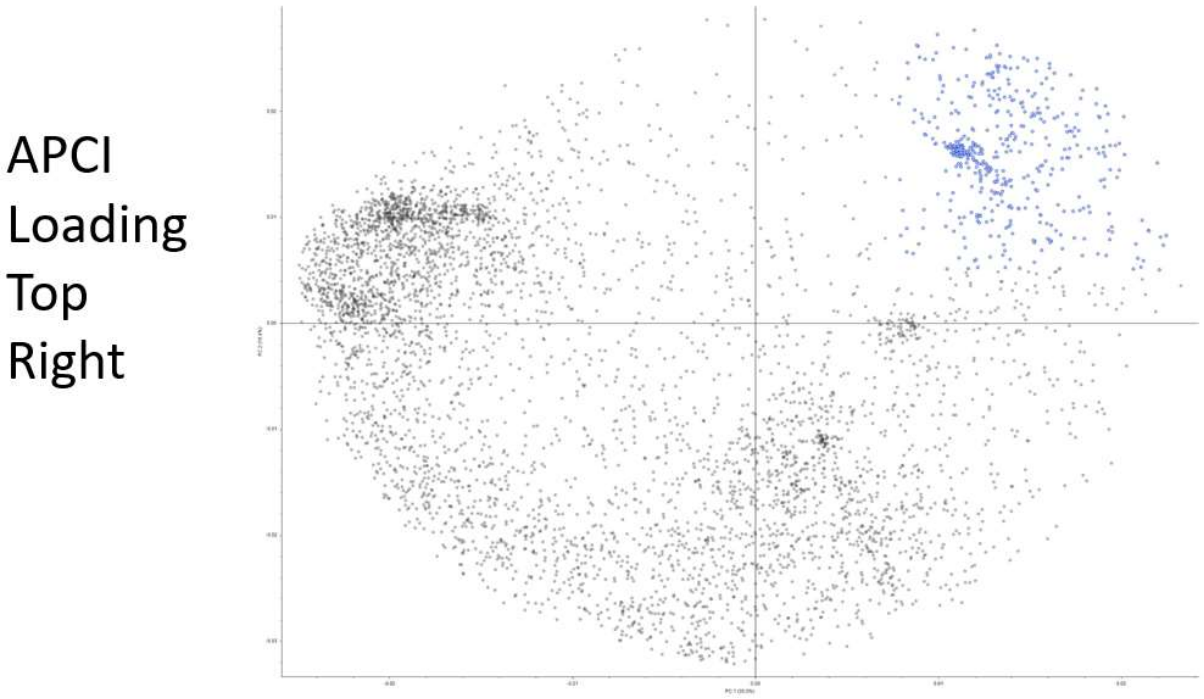

Figure 15-- APCI loading figure with the significant compounds in the top-right corner highlighted

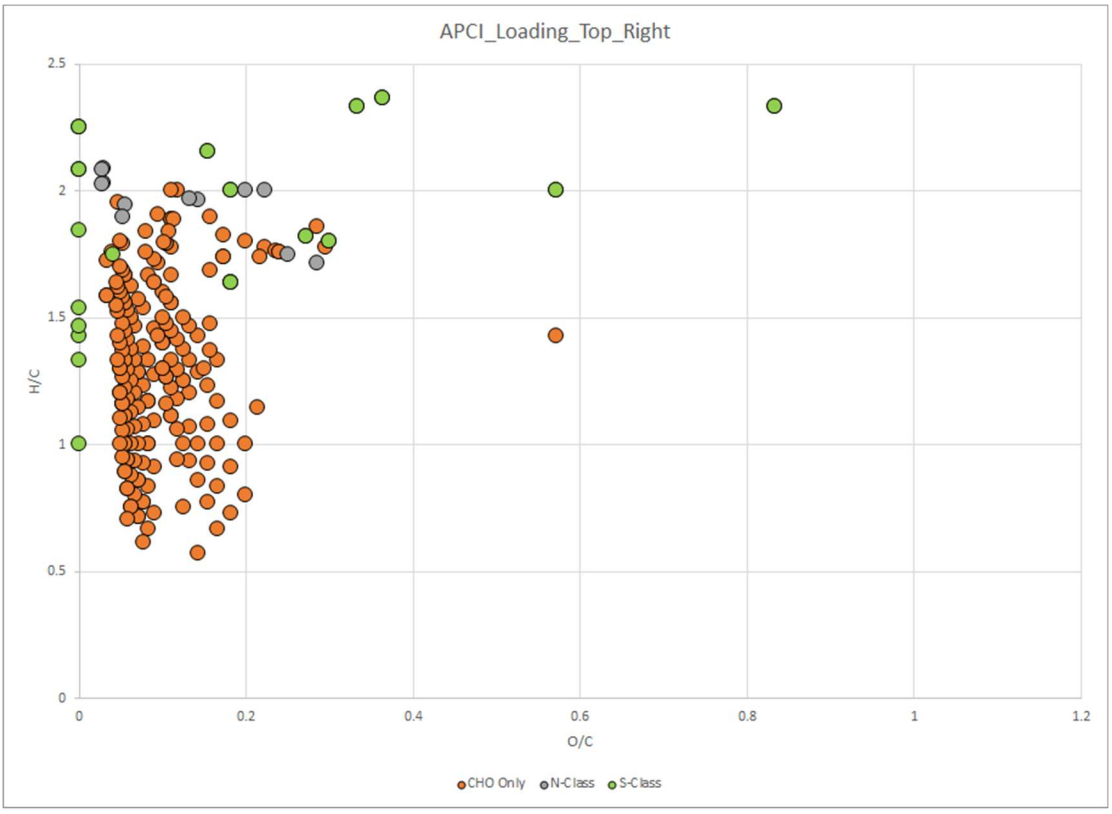

Figure 16-- Van Krevelen figure of the top-right significant compounds

In the top right selection, which primarily consists of Diesel Mud Base and Sarita Sogu 402 compounds, there are very little sulfur-containing and no phosphorus-containing compounds, as indicated by the Van Krevelen diagram.

APCI Loading Bottom Center Right

Bottom Center right	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	35%	747	97	0	140	237	510

APCI_Loading_Bottom_Center_Right	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	510	185	201	11	290
% of total	100%	36%	39%	2%	57%
# of AI>0.5	26	25	1	0	1
# of AI>=0.67	7	6	1	0	1
# of AI-MOD>0.5	30	30	0	0	0
# of AI-MOD>=0.67	8	8	0	0	0
# of Carbons	13424	5205	5598	428	7029
Average O/C	0.103783967	0.059616417	0.167893984	0.154389615	0.115768982
Average H/C (All)	1.668577014	1.37596905	1.864305461	1.90855296	1.821440964
Average H/C (Has O/C)	1.657829791	1.37596905	1.858725419	1.87855126	1.812287897
Average Aromaticity+ Index	0.053008238	0.308956725	-0.175416544	-0.157450335	-0.058121174
Average AI-Mod+	0.245334265	0.332165798	0.156710722	0.107929112	0.138825288
Average DU	5.847058824	9.27027027	4.208955224	4	3.982758621
Average DU/C	0.227335469	0.350078405	0.153701758	0.09264619	0.163860149
Average DBE	7.096078431	9.27027027	6.099502488	6.636363636	6.165517241
Average DBE/C	0.287599312	0.350078405	0.235197527	0.177526556	0.269522834
Average MW	430.6863039	403.6859068	506.8473047	682.9520873	424.3637374
Average Carbon Count	26.32156863	28.13513514	27.85074627	38.90909091	24.23793103
High O/C	0.625	1.909090909	2.4	167	17
High H/C (All)	2.4	1.909090909	2.4	2.12	2.4
High H/C (Has O/C)	2.4	1.909090909	2.4	2.12	2.4
High Aromaticity Index	5	0.727272727	5	0.157894737	5
High AI-Mod	0.75	0.75	0.288888889	0.189873418	0.333333333
High DU	21	21	17	12	17
High DU/C	0.769230769	0.769230769	0.444444444	0.255813953	0.444444444
High DBE	21	21	19	14	19
High DBE/C	0.769230769	0.769230769	0.714285714	0.325581395	0.714285714
High MW	0	0	0	0	0
High C	60	57	60	60	54
Low O/C	0.022222222	0.022222222	0.035714286	0.066666667	0.029411765
Low H/C (All)	0.615384615	0.615384615	1.333333333	1.558139535	1.333333333
Low H/C (Has O/C)	0.090909091	0.615384615	1.333333333	1.558139535	1.333333333
Low Aromaticity Index	-2.2	-0.111111111	-2.2	-0.8	-0.857142857
Low AI-Mod	0.012345679	0.048543689	0.015873016	0.012345679	0.012345679
Low DU	0	2	0	0	0
Low DU/C	0	0.090909091	0	0	0
Low DBE	0	2	0	1	2
Low DBE/C	0	0.090909091	0	0.02	0.040816327
Low MW	0	0	0	0	0
Low C	5	11	5	22	5

Table 2-- APCI bottom center right loading data, separated by compound class

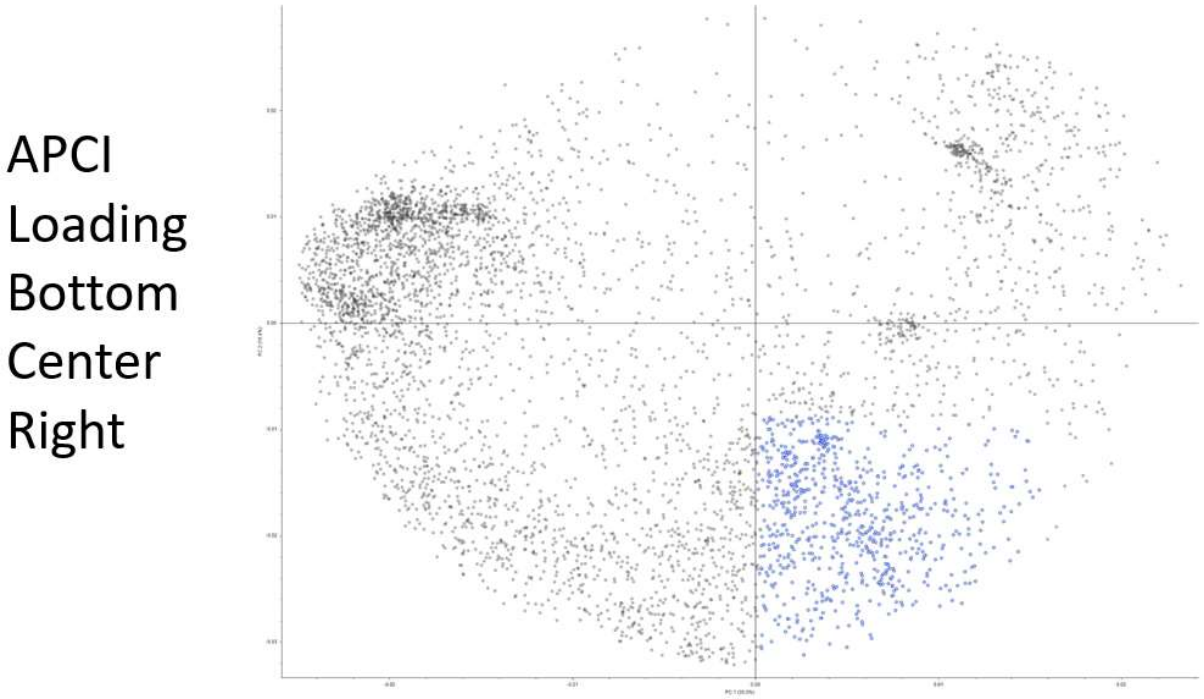

Figure 17-- APCI loading figure with the significant compounds in the bottom-right-center

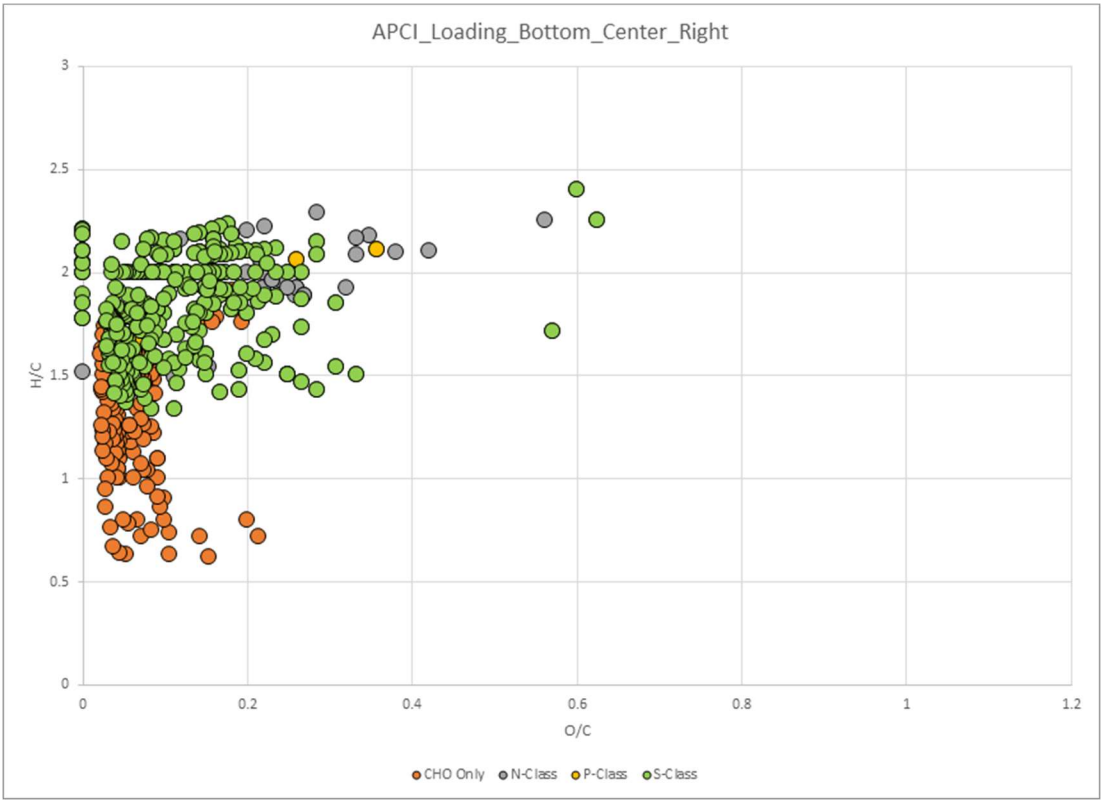

Figure 18-- Van Krevelen figure of the bottom-right-center significant compounds

In the bottom center right selection, which primarily consists of the entire Cavitation group along with JGK C33 and JGK E Wells, the molecular compositions are very clustered on the Van Krevelen figure. There is a clear separation between the clusters of CHO only compounds and the S-Class compounds. There is a significant number of un-oxygenated sulfur compounds.

APCI Loading Center Left

Center Left	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	21%	1783	53	0	277	330	1453

APCI_Loading_Center_Left	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	1453	121	912	65	667
% of total	100%	8%	63%	4%	46%
# of AI>0.5	203	9	132	0	71
# of AI>=0.67	21	0	17	0	5
# of AI-MOD>0.5	206	9	134	0	72
# of AI-MOD>=0.67	21	0	17	0	5
# of Carbons	43919	3329	28633	2090	19062
Average O/C	0.121205916	0.16745382	0.099236581	0.156333267	0.12929093
Average H/C (All)	1.475836717	1.610840449	1.459749431	1.962816907	1.54324012
Average H/C (Has O/C)	1.634432907	1.610840449	1.617100174	1.966347764	1.78160259
Average Aromaticity+ Index	0.188080913	0.065385651	0.189009773	-0.216357541	0.13369724
Average AI-Mod+	0.318953326	0.222305254	0.341881545	0.084827269	0.29631269
Average DU	8.889194769	5.94214876	9.711622807	2.584615385	7.29235382
Average DU/C	0.314532045	0.233789995	0.329345528	0.080743668	0.2830608
Average DBE	10.10392292	5.94214876	10.53289474	4.692307692	9.89505247
Average DBE/C	0.364025894	0.233789995	0.363739012	0.150565699	0.38913455
Average MW	459.7827625	442.2655156	473.4764536	571.8153331	457.616429
Average Carbon Count	30.22642808	27.51239669	31.39583333	32.15384615	28.5787106
High O/C	1.034482759	2.142857143	2.5	168	29
High H/C (All)	2.5	2.142857143	2.5	2.230769231	2.4
High H/C (Has O/C)	2.5	2.142857143	2.5	2.230769231	2.4
High Aromaticity Index	9	0.625	9	0.192307692	1
High AI-Mod	1	0.632653061	1	0.222222222	1
High DU	29	19	29	10	29
High DU/C	1	0.64	1	0.322580645	1
High DBE	31	19	31	11	31
High DBE/C	1.068965517	0.64	1.068965517	0.35483871	1.06896552
High MW	0	0	0	0	0
High C	63	42	63	55	63
Low O/C	0.021276596	0.027777778	0.021276596	0.054545455	0.02564103
Low H/C (All)	0.103448276	0.8	0.103448276	1.516129032	0.10344828
Low H/C (Has O/C)	0.103448276	0.8	0.103448276	1.516129032	0.10344828
Low Aromaticity Index	-14	-1.333333333	-14	-1.166666667	-5.5
Low AI-Mod	0.014492754	0.01754386	0.014492754	0.014492754	0.01449275
Low DU	0	0	0	0	0
Low DU/C	0	0	0	0	0
Low DBE	0	0	0	1	2
Low DBE/C	0	0	0	0.03125	0.03703704
Low MW	0	0	0	0	0
Low C	8	13	8	12	9

Table 3-- APCI center left loading data, separated by compound class

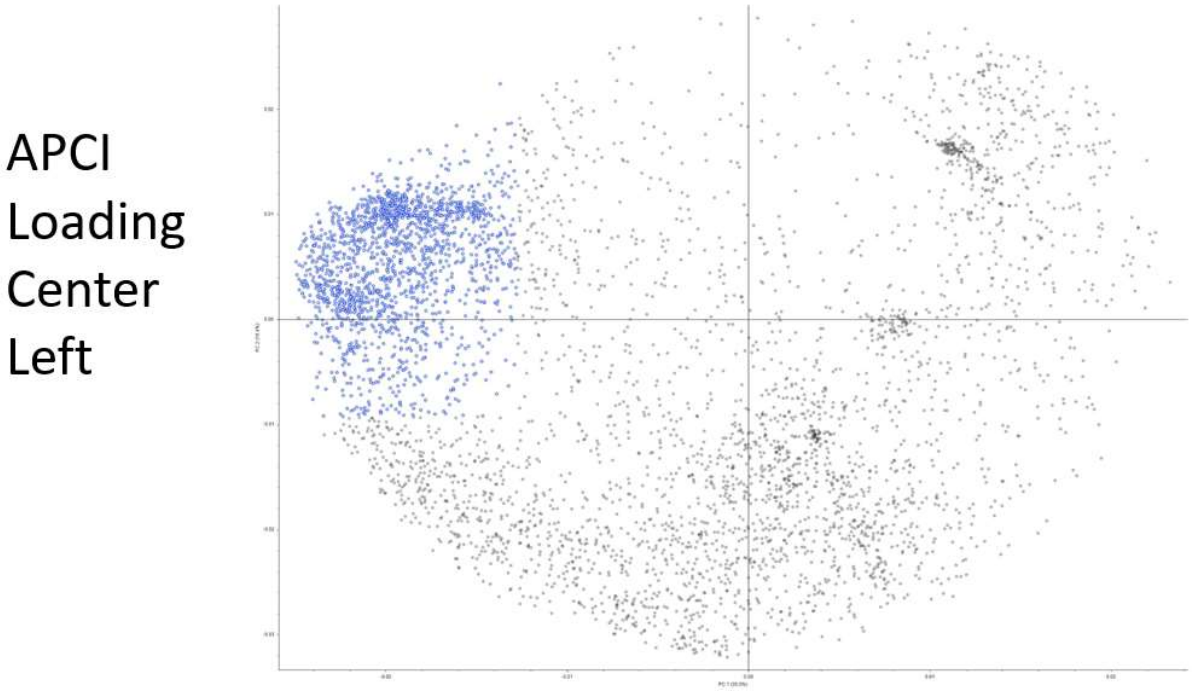

Figure 19-- APCI loading figure with the significant compounds in the left-center highlighted

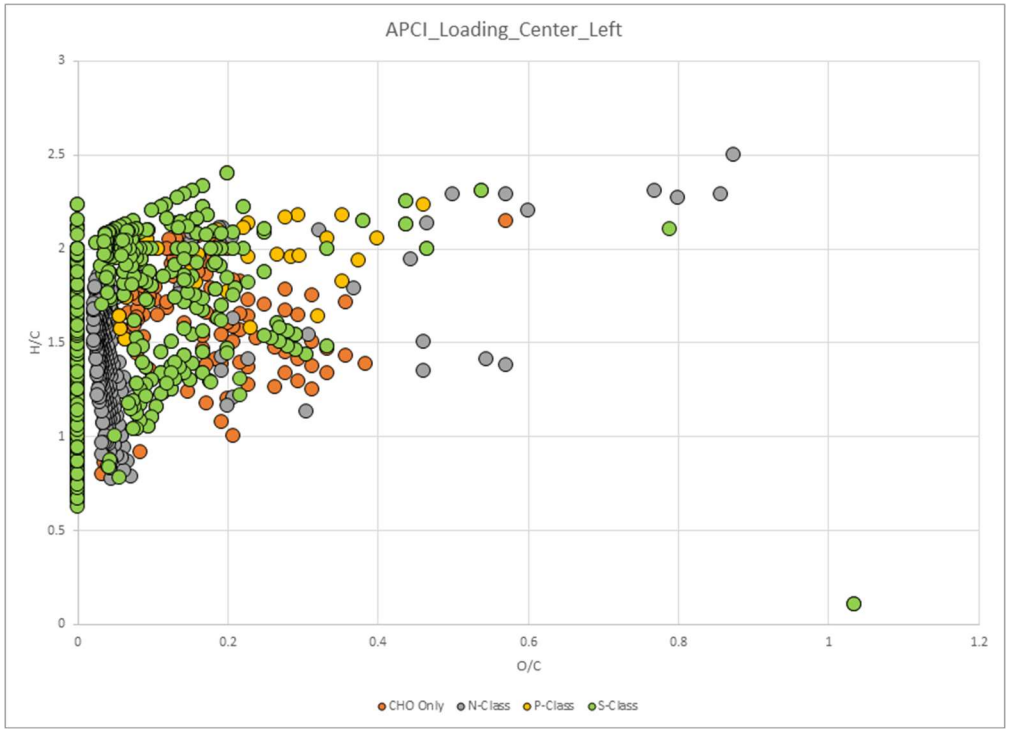

Figure 20-- Van Krevelen figure of the left-center significant compounds

In the center left selection, which primarily consists of the Citgo samples (excluding S0802), the CHO compounds are overshadowed by the large amount of S-Class and N-Class compounds. There is a large chain of N-Class compounds <0.1 O/C, and the non-oxygenated compounds are primarily sulfur-containing.

ESI Loading

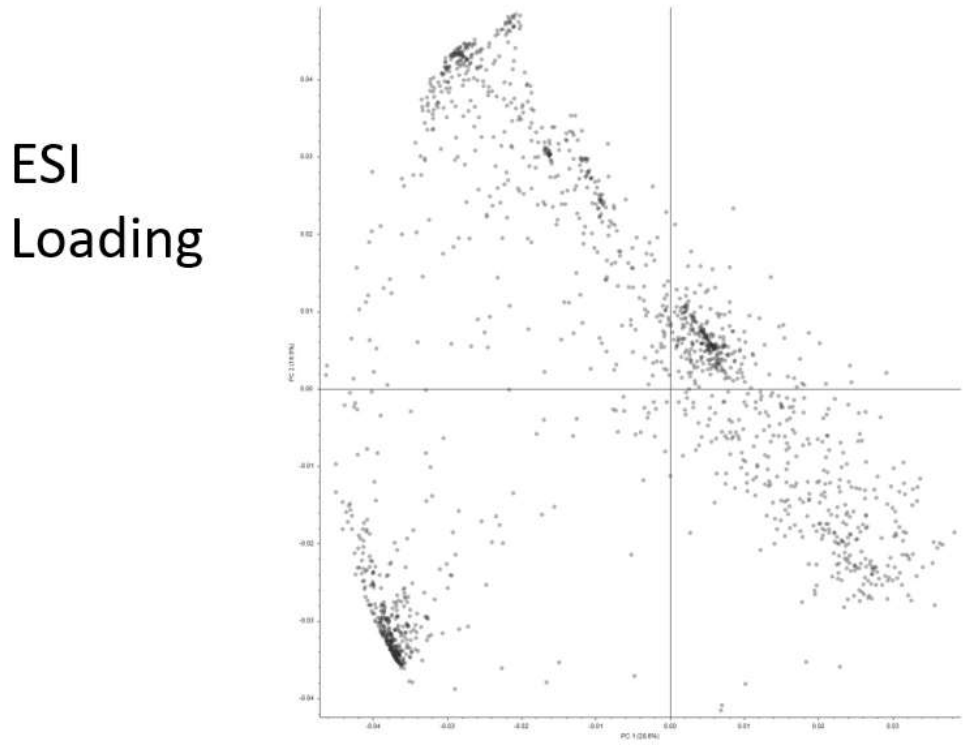

Figure 21-- ESI loading figure generated from Compound Discover

There are substantially fewer compounds in the ESI loading figure than an APCI loading figure for petroleum because petroleum is primarily non-polar. In this study, the Headington sample group is unique, occupying the entirety of the left side of the axis (except W1016). The Headington samples also have the largest inter-group sample spread of the ESI sample groups. By contrast, the Citgo samples tested have more common polar compounds (closer to the center)

with low inter-group sample spread. The Cavitation sample group features an even tighter sample spread, indicating the negligent impact of the cavitation process on the polar chemical composition of petroleum. It is worth noting that Sarita Sogu 402 has poor reproducibility compared to any other sample triplicate in this study. Further testing should be conducted to verify the statistical uniqueness of Sarita Sogu 402.

ESI Bottom Right

Bottom Right	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	0%	263	4	0	46	50	213

ESI_Loading_Bottom_Right	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	213	28	23	-1	164
% of total	100%	13%	11%	0%	77%
# of AI>0.5	27	12	9	0	6
# of AI>=0.67	6	6	0	0	0
# of AI-MOD>0.5	34	12	13	0	9
# of AI-MOD>=0.67	6	6	0	0	0
# of Carbons	3102	558	287	0	2279
Average O/C	0.118612589	0.210370309	0.168549433	#DIV/0!	0.101818821
Average H/C (All)	1.521644022	1.212873333	1.281759303	#DIV/0!	1.616885728
Average H/C (Has O/C)	1.525936501	1.212873333	1.250793651	#DIV/0!	1.620670303
Average Aromaticity+ Index	0.060890956	0.232653785	-0.416320033	#DIV/0!	0.024578222
Average AI-Mod+	0.298869332	0.389461629	0.493618435	#DIV/0!	0.244041422
Average DU	4.530516432	8.392857143	6.130434783	#DIV/0!	3.603658537
Average DU/C	0.322985774	0.445235597	0.490598773	#DIV/0!	0.275539381
Average DBE	6.539906103	8.392857143	6.391304348	#DIV/0!	6.213414634
Average DBE/C	0.47436277	0.445235597	0.518549084	#DIV/0!	0.47214487
Average MW	19.63887324	21.42171429	18.2803913	#DIV/0!	19.5457439
Average Carbon Count	14.56338028	19.92857143	12.47826087	#DIV/0!	13.89634146
High O/C	0.666666667	1.8	2.5	0	10
High H/C (All)	2.5	1.8	2.5	0	2.5
High H/C (Has O/C)	2.5	1.8	2.5	0	2.5
High Aromaticity Index	0.842105263	0.842105263	0.666666667	0	0.555555556
High AI-Mod	0.846153846	0.846153846	0.666666667	0	0.565217391
High DU	17	17	9	0	10
High DU/C	0.85	0.85	0.7	0	0.625
High DBE	17	17	9	0	12
High DBE/C	0.916666667	0.85	0.7	0	0.916666667
High MW	0	0	0	0	0
High C	28	27	18	0	28
Low O/C	0.05	0.05	0.071428571	0	0.066666667
Low H/C (All)	0.4	0.4	0.9	0	0.923076923
Low H/C (Has O/C)	0.4	0.4	0.9	0	0.923076923
Low Aromaticity Index	-7	-0.307692308	-7	0	-7
Low AI-Mod	0.025641026	0.025641026	0.214285714	0	0.041666667
Low DU	0	3	0	0	0
Low DU/C	0	0.15	0	0	0
Low DBE	1	3	1	0	3
Low DBE/C	0.111111111	0.15	0.111111111	0	0.227272727
Low MW	0	0	0	0	0
Low C	6	14	8	0	6

Table 4--ESI bottom right loading data, separated by compound class

ESI
Loading
Bottom
Right

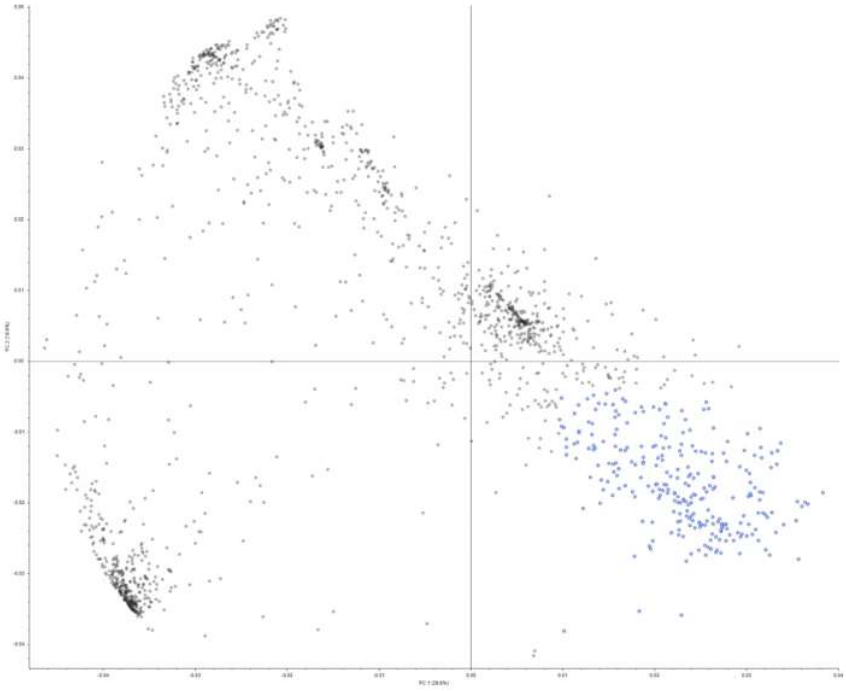

Figure 22-- ESI loading figure with the significant compounds in the bottom-right highlighted

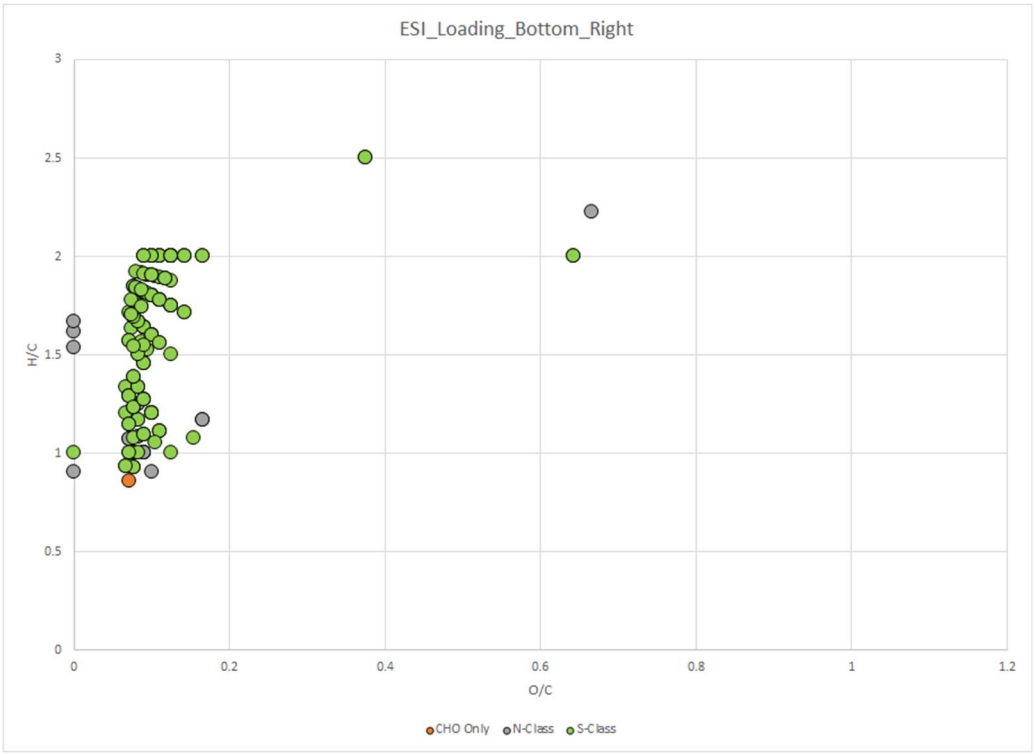

Figure 23-- Van Krevelen figure of the bottom-right significant compounds

In the bottom right section of the ESI loading figure, which consists primarily of the Cavitation sample group, the predominant sample group is S-Class samples. This contrasts with other Van Krevelen diagrams, where CHO and N-Class compounds have a more significant presence.

ESI Bottom Left

Bottom Left	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	17%	403	30	2	53	85	318

ESI_Loading_Bottom_Left	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	318	222	76	8	67
% of total	100%	70%	24%	3%	21%
# of AI>0.5	64	51	8	2	8
# of AI>=0.67	10	4	5	2	4
# of AI-MOD>0.5	84	75	4	1	5
# of AI-MOD>=0.67	14	10	2	1	2
# of Carbons	4398	3089	977	75	849
Average O/C	0.222935672	0.186646376	0.311728	0.595337	0.288290838
Average H/C (All)	1.392369279	1.216599512	1.920497	1.866766	1.878193176
Average H/C (Has O/C)	1.394259774	1.216599512	1.978957	1.866766	1.971055048
Average Aromaticity+ Index	0.049048378	0.316764909	-0.58803	-2	-0.582691488
Average AI-Mod+	0.444676939	0.426607086	0.612846	0.882353	0.675180382
Average DU	4.919137466	6.310810811	2.960526	2.625	3.447761194
Average DU/C	0.405193667	0.467931496	0.218022	0.273313	0.236457819
Average DBE	5.46361186	6.310810811	5.118421	3.875	6.358208955
Average DBE/C	0.462671561	0.467931496	0.424757	0.423016	0.497785023
Average MW	18.61879245	21.02625676	20.17879	14.35125	19.68
Average Carbon Count	13.83018868	13.85201794	12.68831	8.333333	12.48529412
High O/C	1	2.5	2.5	58	14
High H/C (All)	2.5	2.5	2.5	2.428571	2.5
High H/C (Has O/C)	2.5	2.5	2.5	2.428571	2.5
High Aromaticity Index	7	0.75	7	2.333333	7
High AI-Mod	9	0.76	9	0.882353	9
High DU	14	11	13	11	14
High DU/C	0.916666667	0.777777778	0.833333	0.916667	0.769230769
High DBE	16	11	13	12	16
High DBE/C	1	0.777777778	0.833333	1	1
High MW	0	0	0	0	0
High C	31	20	31	15	26
Low O/C	0.038461538	0.052631579	0.055556	0.142857	0.038461538
Low H/C (All)	0.416666667	0.615384615	0.833333	0.416667	0.615384615
Low H/C (Has O/C)	0.416666667	0.615384615	0.833333	0.416667	0.615384615
Low Aromaticity Index	-7	-3	-6	-7	-6
Low AI-Mod	0.03030303	0.047619048	0.030303	0.882353	0.037037037
Low DU	0	0	0	0	0
Low DU/C	0	0	0	0	0
Low DBE	0	0	1	1	2
Low DBE/C	0	0	0.052632	0.066667	0.166666667
Low MW	0	0	0	0	0
Low C	4	0	0	0	0

Table 5--ESI bottom left loading data, separated by compound class

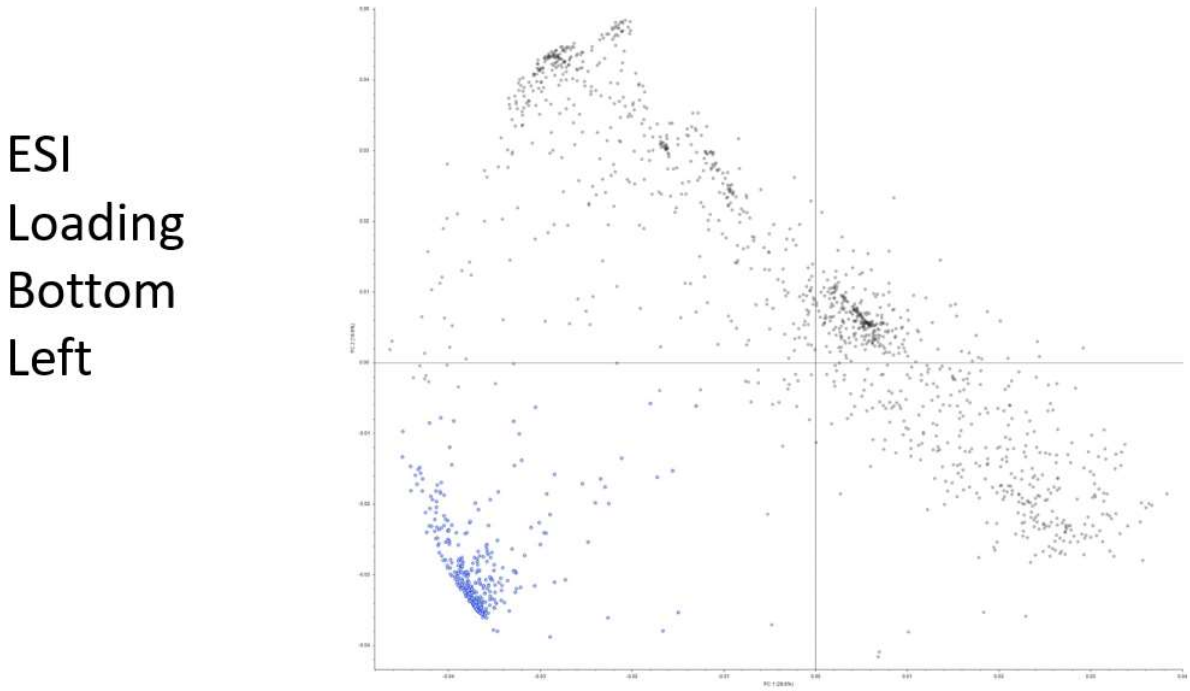

Figure 24-- ESI loading figure with the significant compounds in the bottom-left highlighted

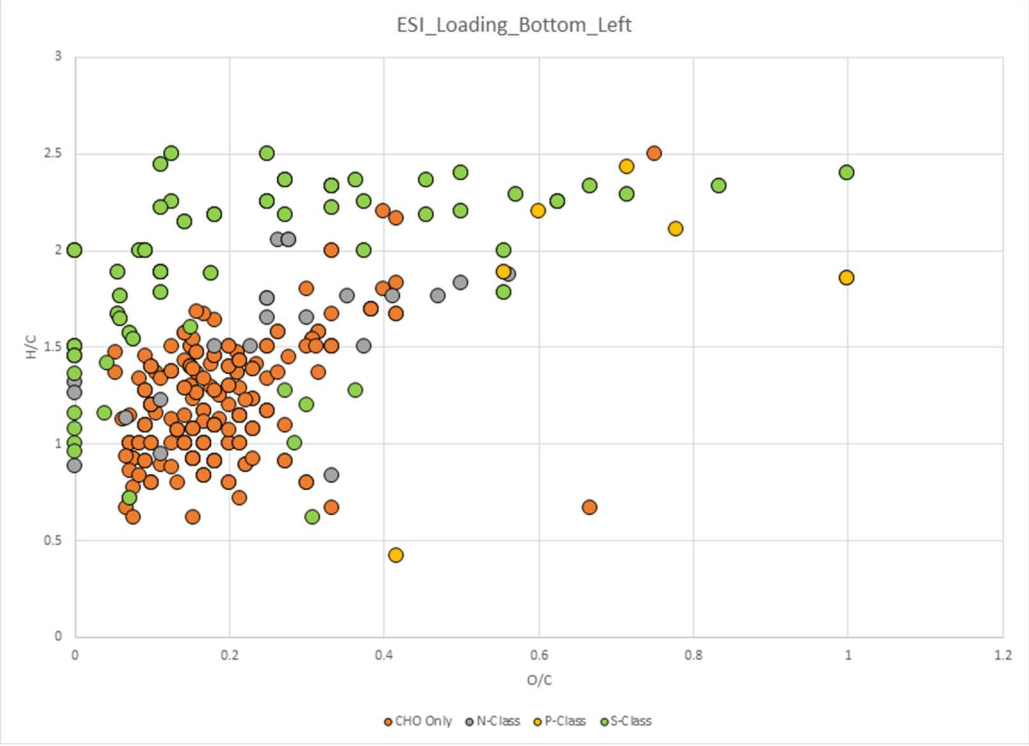

Figure 25-- Van Krevelen figure of the bottom-left significant compounds

In the bottom-left section of the ESI loading figure, which may consist solely of Sarita Sogu 402, the compounds are predominantly CHO only, with an extensive oxygenation range of S-Class compounds. There are significant S-Class and some N-Class compounds with no oxygenation.

ESI Top Left

Top Left	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	11%	403	21	14	120	155	248

ESI_Loading_Top_Left	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	248	89	110	7	69
% of total	100%	36%	44%	3%	28%
# of AI>0.5	22	4	10	0	8
# of AI>=0.67	11	838	5	0	5
# of AI-MOD>0.5	28	6	13	0	10
# of AI-MOD>=0.67	15	0	7	0	6
# of Carbons	2887	1516	1328	77	809
Average O/C	0.224203168	1.565231484	0.203720763	0.431395892	0.159365057
Average H/C (All)	1.660294693	1.565231484	1.706869162	1.720564893	1.693422814
Average H/C (Has O/C)	1.676327434	30.87640449	1.730028294	1.720564893	1.723245353
Average Aromaticity+ Index	-0.115266653	0.297444094	-0.169982391	-0.738095238	-0.21355317
Average AI-Mod+	0.36807864	0.134328358	0.400652918	0.153846154	0.492661648
Average DU	3.439516129	3.370786517	3.6	3	3.391304348
Average DU/C	0.291623583	0.318551229	0.295412697	0.333380019	0.28308513
Average DBE	4.193548387	0.318551229	4.345454545	4.285714286	6.014492754
Average DBE/C	0.365156105	17.77996723	0.36889139	0.517250233	0.539591997
Average MW	17.22392339	2971869.393	16.47838182	18.75671429	16.93985507
Average Carbon Count	11.64112903	17.03370787	12.07272727	11	11.72463768
High O/C	0.666666667	2.333333333	2.5	49	12
High H/C (All)	2.5	2.333333333	2.5	2.375	2.444444444
High H/C (Has O/C)	2.5	59	2.5	2.375	2.444444444
High Aromaticity Index	1	1	1	0	1
High AI-Mod	1	1	1	0.153846154	1
High DU	18	12	18	6	12
High DU/C	1.25	1	1.25	1.25	1.25
High DBE	18	1	18	8	16
High DBE/C	2	24.03313439	2	2	2
High MW	0	0	0	0	0
High C	24	34	24	17	22
Low O/C	0.045454545	0.166666667	0.0625	0.3	0.045454545
Low H/C (All)	0.166666667	0.166666667	0.25	0.5	0.25
Low H/C (Has O/C)	0.166666667	12	0.25	0.5	0.25
Low Aromaticity Index	-5.5	0.0625	-5.5	-1.666666667	-5.5
Low AI-Mod	0.047619048	0	0.047619048	0.153846154	0.090909091
Low DU	0	0	0	0	0
Low DU/C	0	0	0	0	0
Low DBE	0	0	0	1	2
Low DBE/C	0	0.345613653	0	0.125	0.181818182
Low MW	0	0	0	0	0
Low C	4	2	4	4	4

Table 6-- ESI top left loading data, separated by compound class

ESI
Loading
Top Left

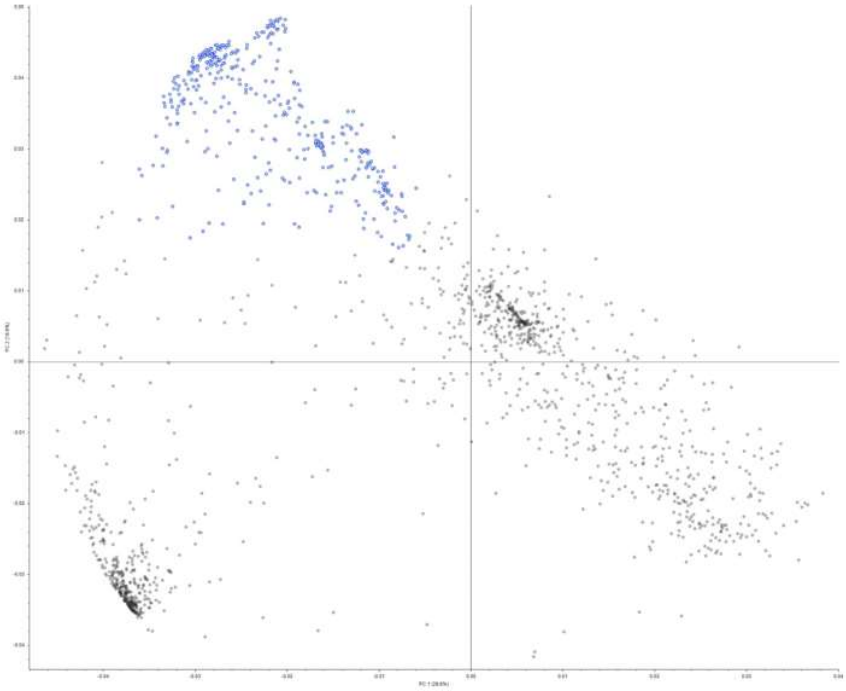

Figure 26-- ESI loading figure with the significant compounds in the top-left highlighted

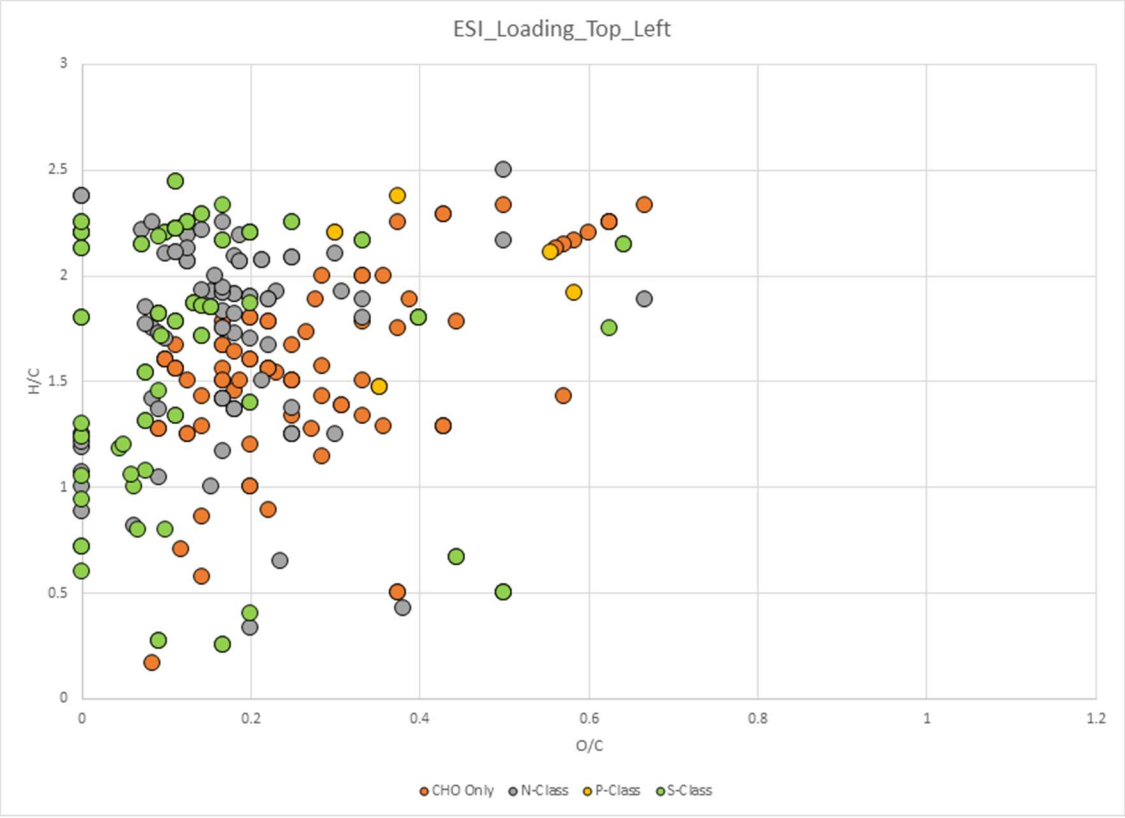

Figure 27-- Van Krevelen figure of the top-left significant compounds

In the top left section of the ESI loading figure, which consists of Headington samples (JGK C33, JCK C20, Diesel Mud Base), the Van Krevelen diagram generated is much less clustered other figures in this research. There is a significant number of unoxygenated N-Class and S-Class compounds.

APCI vs. ESI

Overall, several trends can be seen throughout the two sets of loading data.

1) Citgo is unique and isolated for APCI, while Headington is unique/isolated for ESI loading.

2) The cavitation sample group always has a tight sample spread due to LEX_TX_1AM being the base sample for all the other samples (Mid, Final, Vapor Recovery).

3) Diesel Mud base and Sarita Sogu 402 are similar on APCI because Diesel Mud base is a constituent of Sarita Sogu 402.

Volcano Plot Van Krevelens

Van Krevelen analyses were also conducted on the statistically unique compounds of every volcano plot comparison to display compounds do not present in other groups of oil visually. With a significant accumulation of different oil sample groups, this method will become a more robust method for filtering unique compounds by formation, well, or even by sample.

APCI Volcano

Citgo vs Cavitation

Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
16%	504	4	0	68	72	432

APCI_Citgo_CitgoVCavitation	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	432	63	187	9	240
% of total	100%	15%	43%	2%	56%
# of AI>0.5	114	0	82	0	36
# of AI>=0.67	13	0	12	0	1
# of AI-MOD>0.5	118	0	85	0	37
# of AI-MOD>=0.67	13	0	12	0	1
# of Carbons	11478	1665	4463	233	6712
Average O/C	0.183597187	0.205830623	0.154466	0.255287	0.16983
Average H/C (All)	1.428961159	1.712032823	1.263288	1.929768	1.544506
Average H/C (Has O/C)	1.665961028	1.712032823	1.583917	1.953884	1.803383
Average Aromaticity+ Index	0.186951205	-0.033626478	0.25615	-0.31979	0.113354
Average AI-Mod+	0.353289317	0.17051227	0.484711	0.180556	0.28033
Average DU	8.25462963	4.238095238	10.17112	3.111111	7.129167
Average DU/C	0.341316651	0.185711455	0.444527	0.107505	0.280601
Average DBE	9.81712963	4.238095238	11.32086	4.333333	9.908333
Average DBE/C	0.406948417	0.185711455	0.500406	0.157273	0.397201
Average m/z	416.75721	442.5914852	376.45	477.2196	447.1319
Average Carbon Count	26.56944444	26.42857143	23.86631	25.88889	27.96667
High O/C	0.789473684	2.142857143	2.4	120	19
High H/C (All)	2.4	2.142857143	2.4	2.230769	2.4
High H/C (Has O/C)	2.4	2.142857143	2.4	2.230769	2.4
High Aromaticity Index	0.75	0.470588235	0.75	0.192308	0.692308
High AI-Mod	0.75	0.5	0.75	0.222222	0.692308
High DU	22	13	22	10	19
High DU/C	0.769230769	0.526315789	0.769231	0.322581	0.714286
High DBE	22	13	22	11	21
High DBE/C	0.882352941	0.526315789	0.785714	0.354839	0.882353
High m/z	756.75611	668.63068	730.6453	657.5273	756.7561
High C	52	42	48	38	52
Low O/C	0.037037037	0.078947368	0.037037	0.064516	0.043478
Low H/C (All)	0.65	1.052631579	0.65	1.516129	0.714286
Low H/C (Has O/C)	0.80952381	1.052631579	0.809524	1.516129	0.869565
Low Aromaticity Index	-5.333333333	-1.333333333	-2.33333	-1.16667	-5.33333
Low AI-Mod	0.01754386	0.01754386	0.025641	0.138889	0.021277
Low DU	0	0	0	0	0
Low DU/C	0	0	0	0	0
Low DBE	0	0	0	1	2
Low DBE/C	0	0	0	0.058824	0.068966
Low m/z	172.12839	254.11521	179.0734	312.1702	172.1284
Low C	9	13	9	13	9

Table 7—APCI Citgo vs Cavitation data, Citgo side

Cavitation Side	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	17%	404	32	0	55	87	317

APCI_Cavitation_CitgoVCavitation	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	317	162	77	5	128
% of total	100%	51%	24%	2%	40%
# of AI>0.5	32	30	2	0	1
# of AI>=0.67	12	11	1	0	1
# of AI-MOD>0.5	38	35	3	0	0
# of AI-MOD>=0.67	15	15	0	0	0
# of Carbons	7024	3564	2043	206	2524
Average O/C	0.112206374	0.08058323	0.221402	0.195556	0.13313
Average H/C (All)	1.607173633	1.351801351	1.939619	1.974758	1.886283
Average H/C (Has O/C)	1.604043912	1.351801351	1.947491	1.928889	1.885766
Average Aromaticity+ Index	0.093101172	0.318264664	-0.32675	-0.17343	-0.13877
Average AI-Mod+	0.306549697	0.352549896	0.342863	0.142857	0.146904
Average DU	5.189274448	7.5	2.714286	2.8	2.5
Average DU/C	0.265221937	0.374766735	0.140501	0.057424	0.144482
Average DBE	6.056782334	7.5	4.207792	5.4	4.625
Average DBE/C	0.318069891	0.374766735	0.224111	0.155879	0.274921
Average MW	362.3022313	322.986275	503.2322	730.348	357.2597
Average Carbon Count	22.15772871	22	26.53247	41.2	19.71875
High O/C	0.625	1.909090909	2.5	167	7
High H/C (All)	2.5	1.909090909	2.5	2.06	2.5
High H/C (Has O/C)	2.5	1.909090909	2.5	2.06	2.5
High Aromaticity Index	5	0.75	5	0.111111	5
High AI-Mod	0.76	0.76	0.578947	0.142857	0.333333
High DU	17	17	12	12	7
High DU/C	0.769230769	0.769230769	0.636364	0.2	0.538462
High DBE	17	17	13	13	9
High DBE/C	0.769230769	0.769230769	0.714286	0.272727	0.714286
High MW	974.7039	960.72429	974.7039	942.7142	974.7035
High C	60	57	60	60	53
Low O/C	0.029411765	0.029411765	0.030303	0.066667	0.033333
Low H/C (All)	0.571428571	0.571428571	1	1.666667	1.230769
Low H/C (Has O/C)	0.571428571	0.571428571	1	1.666667	1.230769
Low Aromaticity Index	-4	-0.2	-4	-0.38889	-4
Low AI-Mod	0.021276596	0.025641026	0.142857	0.142857	0.021277
Low DU	0	2	0	0	0
Low DU/C	0	0.090909091	0	0	0
Low DBE	0	2	0	1	2
Low DBE/C	0	0.090909091	0	0.02	0.040816
Low MW	156.08999	170.0727	156.09	404.2708	160.0918
Low C	5	12	5	22	5

Table 8—APCI Citgo vs Cavitation data, Cavitation side

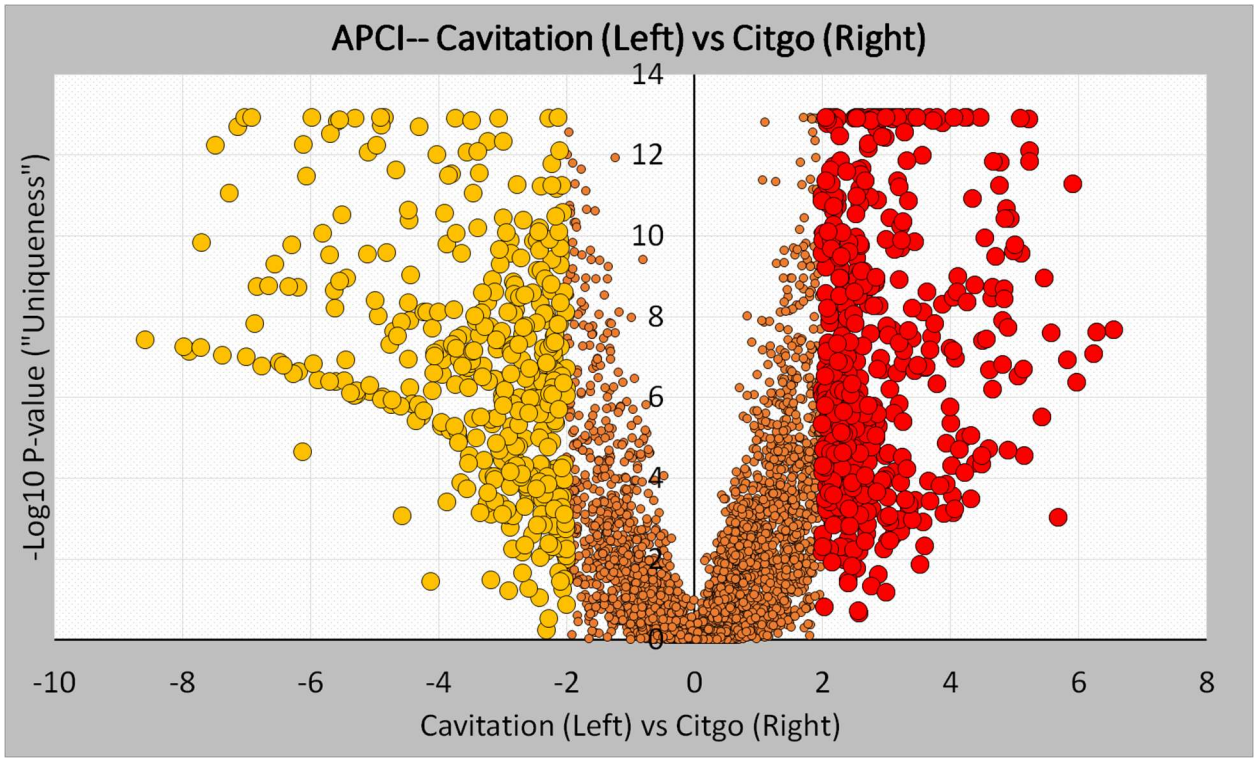

Figure 28—APCI volcano plot of the Cavitation compounds versus the Citgo compounds

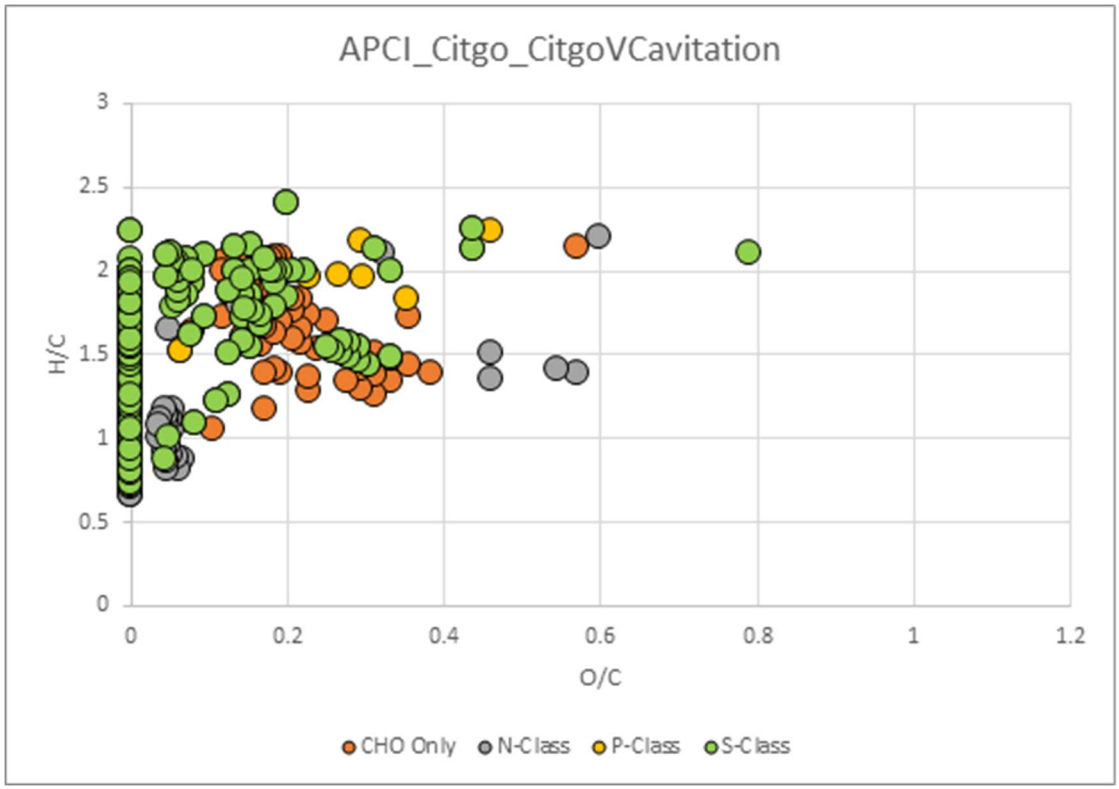

Figure 29-- Van Krevelen figure of the significant Citgo compounds (right side)

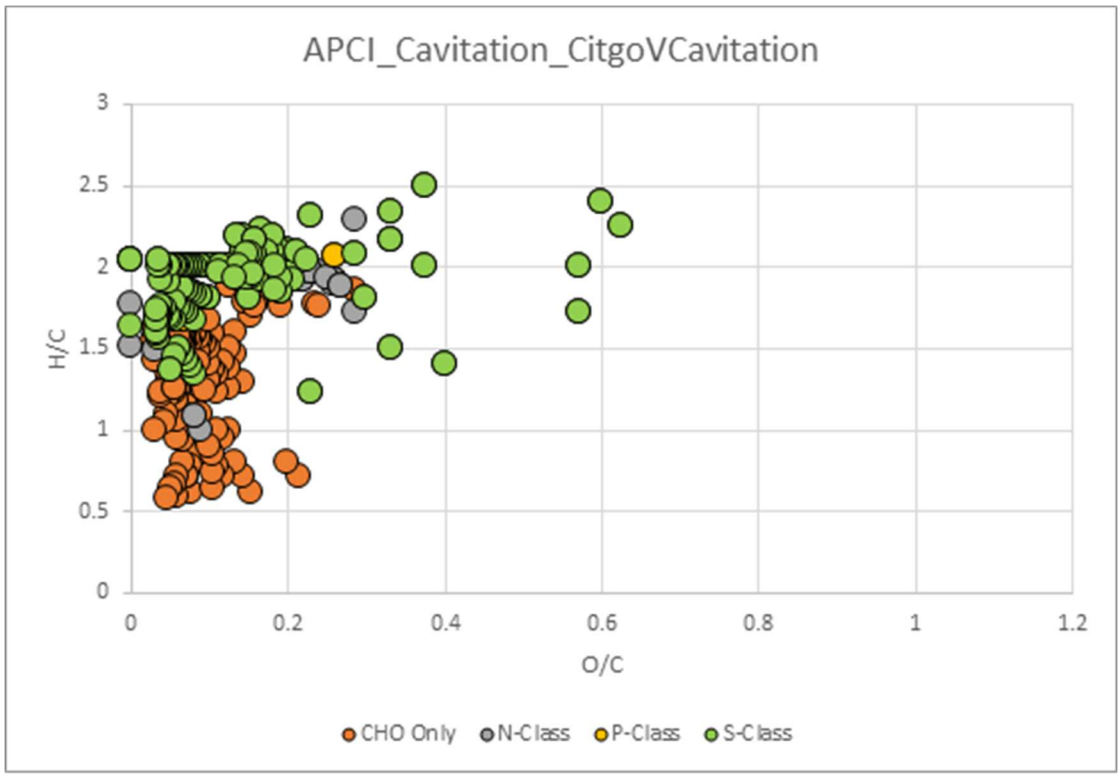

Figure 30-- Van Krevelen figure of the significant Cavitation compounds (left side)

For this volcano comparison, there are many compounds with a lot of those compounds shared between Citgo and Cavitation samples (center, high –log10 P-value). For the Citgo data, there were significant amounts of S-Class compounds, including unoxygenated S-Class compounds. A larger CHO grouping between 0.5-1.5 H/C for the Cavitation data, along with another class cluster above it at 1.5-2 H/C. There are also minimal unoxygenated compounds detected in the Cavitation sample group in comparison to Citgo samples.

Headington vs Cavitation

Headington Side	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	4%	130	13	0	20	33	97

APCI_Headington_HeadingtonVCavitation	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	97	69	11	2	19
% of total	100%	71%	11%	2%	20%
# of AI>0.5	2	1	1	0	1
# of AI>=0.67	1	0	1	0	1
# of AI-MOD>0.5	6	4	1	1	1
# of AI-MOD>=0.67	1	0	1	0	1
# of Carbons	1657	1089	298	47	328
Average O/C	0.155870036	0.123400457	0.11003	0.266667	0.297938
Average H/C (All)	1.557625588	1.510587301	1.818587	1.10119	1.610735
Average H/C (Has O/C)	1.553818804	1.510587301	1.808138	1.10119	1.593383
Average Aromaticity+ Index	-0.065168321	0.209654826	-0.02631	0.314286	-1.10008
Average AI-Mod+	0.265801767	0.271872168	0.26928	0.426901	0.239576
Average DU	3.948717949	4.565217391	4.818182	11.5	4.894737
Average DU/C	0.290980413	0.313959586	0.167967	0.547619	0.263214
Average DBE	4.341880342	4.565217391	5.636364	12.5	7.210526
Average DBE/C	0.319504311	0.313959586	0.203598	0.603571	0.402946
Average MW	283.0235247	242.9930664	449.0821	443.1679	345.7367
Average Carbon Count	17.08247423	15.7826087	27.09091	23.5	17.26316
High O/C	0.833333333	2	2.142857	90	23
High H/C (All)	2.142857143	2	2.142857	1.285714	2.142857
High H/C (Has O/C)	2.142857143	2	2.142857	1.285714	2.142857
High Aromaticity Index	0.695652174	0.625	0.695652	0.428571	0.695652
High AI-Mod	0.72	0.666666667	0.72	0.555556	0.72
High DU	23	11	23	15	23
High DU/C	0.766666667	0.7	0.766667	0.666667	0.766667
High DBE	25	11	25	16	25
High DBE/C	0.833333333	0.7	0.833333	0.75	0.833333
High MW	662.44566	576.51073	636.2967	636.2967	518.1306
High C	38	37	38	35	30
Low O/C	0.027777778	0.043478261	0.027778	0.2	0.055556
Low H/C (All)	0.6	0.8	0.6	0.916667	0.6
Low H/C (Has O/C)	0.6	0.8	0.6	0.916667	0.6
Low Aromaticity Index	-7	-0.333333333	-0.625	0.2	-7
Low AI-Mod	0.025641026	0.025641026	0.028571	0.298246	0.071429
Low DU	0	1	0	8	1
Low DU/C	0	0.052631579	0	0.428571	0.071429
Low DBE	0	1	0	9	3
Low DBE/C	0	0.052631579	0	0.457143	0.214286
Low MW	150.10415	150.10415	157.1463	250.0391	228.1179
Low C	7	7	9	12	12

Table 9—APCI Headington vs Cavitation data, Headington side

Cavitation Side	Overlap	Count (Or	Deleted CxHy	Deleted O/C>1	Blanks	Sum Dele	Count (Calculated)
	21%	482	68	0	60	128	354

APCI_Cavitation_HeadingtonVCavitation	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	354	32	192	9	197
% of total	100%	9%	54%	3%	56%
# of AI>0.5	14	4	8	0	3
# of AI>=0.67	4	3	1	0	1
# of AI-MOD>0.5	17	4	11	1	4
# of AI-MOD>=0.67	6	3	3	1	2
# of Carbons	10654	1052	6417	350	5652
Average O/C	0.100427083	0.088322141	0.13416	0.264434	0.092861
Average H/C (All)	1.679275308	1.535800463	1.614647	1.765958	1.846502
Average H/C (Has O/C)	1.781813753	1.535800463	1.793919	1.769598	1.874321
Average Aromaticity+ Index	0.090685893	0.187363959	0.10885	-0.19665	-0.02606
Average AI-Mod+	0.254341313	0.266097996	0.364552	0.453915	0.140988
Average DU	5.61299435	7.5625	6.760417	6	3.248731
Average DU/C	0.215450458	0.265365006	0.252867	0.17163	0.133585
Average DBE	6.847457627	7.5625	7.703125	7.222222	5.426396
Average DBE/C	0.270049187	0.265365006	0.283335	0.209556	0.230489
Average MW	475.5999568	496.5203841	532.4018	706.6983	473.1984
Average Carbon Count	30.0960452	32.875	33.42188	38.88889	28.69036
High O/C	1.034482759	2.052631579	2.190476	168	30
High H/C (All)	2.19047619	2.052631579	2.190476	2.166667	2.190476
High H/C (Has O/C)	2.19047619	2.052631579	2.190476	2.166667	2.190476
High Aromaticity Index	1	0.75	1	0.138889	1
High AI-Mod	1.08	0.756097561	1.08	1.08	1.08
High DU	30	16	30	30	30
High DU/C	1.034482759	0.761904762	1.034483	1.034483	1.034483
High DBE	33	16	33	33	33
High DBE/C	1.137931034	0.761904762	1.137931	1.137931	1.137931
High MW	974.7039	960.72429	974.7039	942.7142	974.7035
High C	63	57	61	60	63
Low O/C	0.018867925	0.018867925	0.028571	0.054545	0.03125
Low H/C (All)	0.103448276	0.571428571	0.103448	0.103448	0.103448
Low H/C (Has O/C)	0.103448276	0.571428571	0.103448	0.103448	0.103448
Low Aromaticity Index	-0.666666667	-0.151515152	-0.66667	-0.5	-0.66667
Low AI-Mod	0.017241379	0.048543689	0.018868	0.138889	0.017241
Low DU	0	0	0	0	0
Low DU/C	0	0	0	0	0
Low DBE	0	0	0	1	2
Low DBE/C	0	0	0	0.02	0.037037
Low MW	160.03454	232.08827	173.0839	366.2533	160.0345
Low C	8	17	10	18	8

Table 10-- APCI Headington vs Cavitation data, Cavitation side

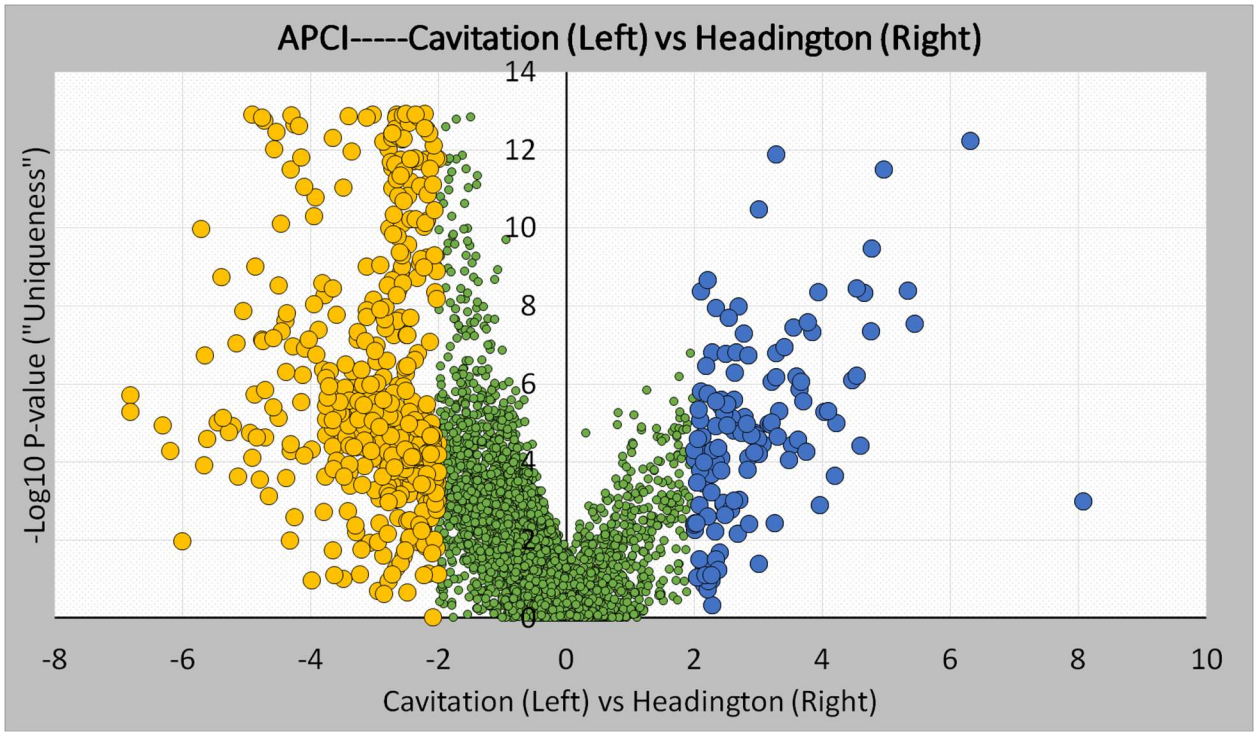

Figure 31-- APCI volcano plot of the Cavitation compounds versus the Headington compounds

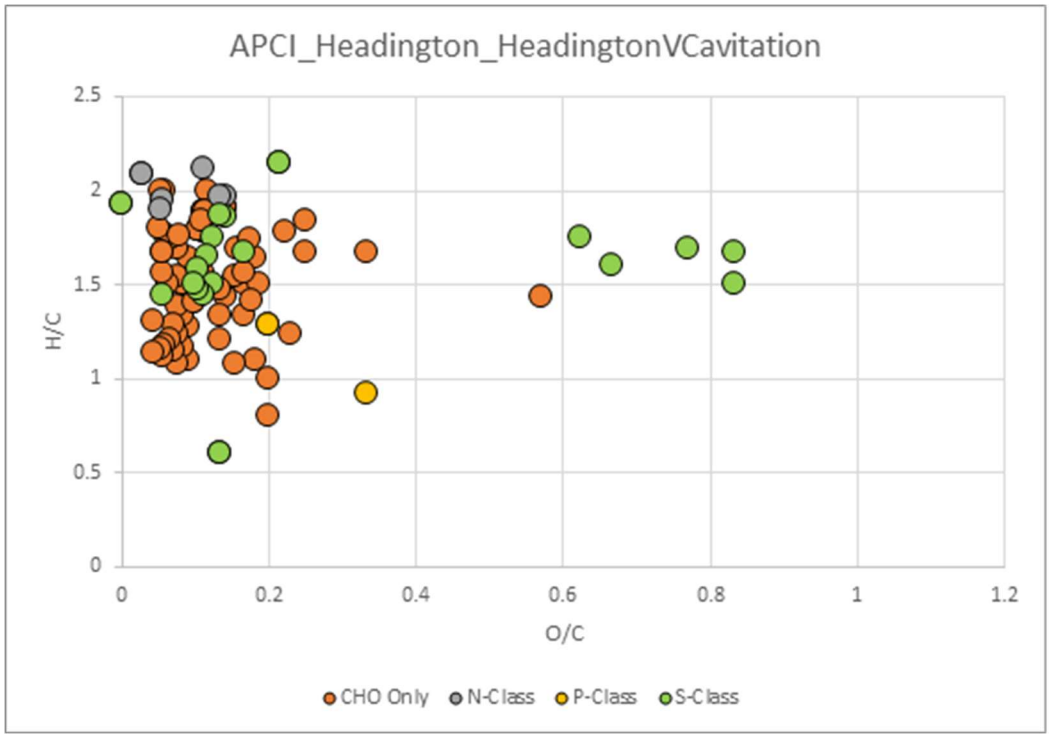

Figure 32-- Van Krevelen figure of the significant Headington compounds (right side)

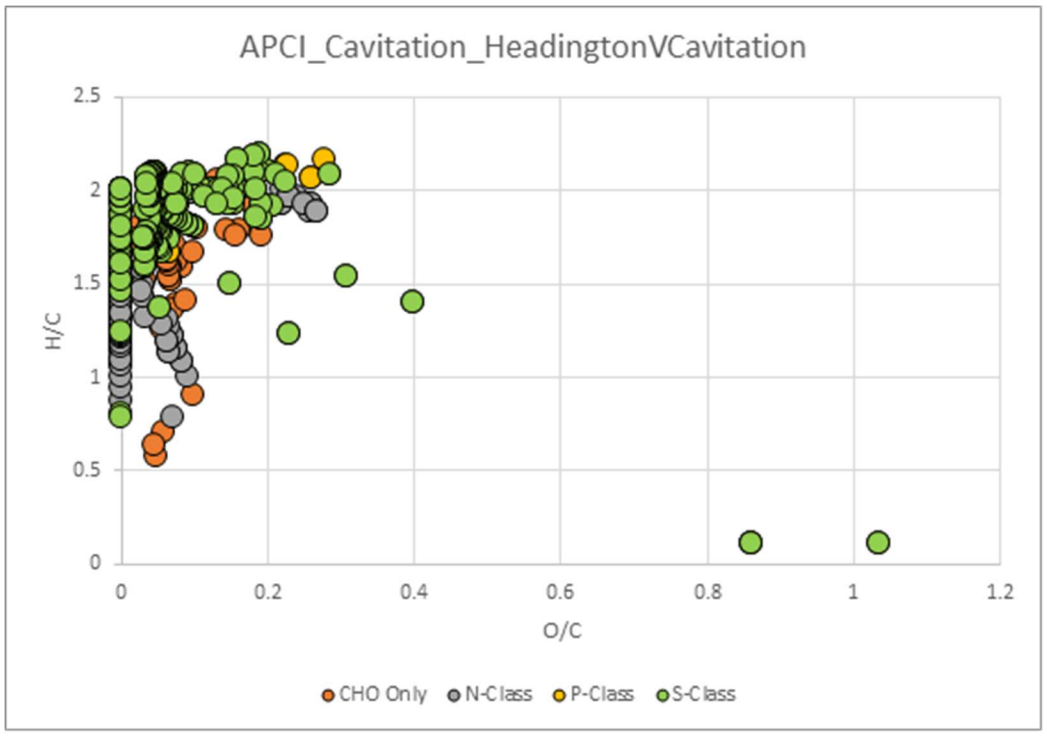

Figure 33-- Van Krevelen figure of the significant Cavitation compounds (left side)

For this volcano comparison, there are substantially fewer unique compounds detected in Headington samples than the Cavitation sample group—this also goes for the less unique compounds in the center, which is also stacked towards the Cavitation side. Regarding the Van Krevelen data, the Headington sample group primarily consists of CHO only compounds, while the Cavitation Van Krevelen consists of more N-Class and S-Class compounds.

Headington vs Citgo

Headington Side	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	6%	338	52	0	38	90	248

APCI_Headington_HeadingtonVCitgo	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	248	195	24	1	42
% of total	100%	79%	10%	0%	17%
# of AI>0.5	26	25	1	0	1
# of AI>=0.67	7	6	1	0	1
# of AI-MOD>0.5	35	34	1	0	1
# of AI-MOD>=0.67	11	10	1	0	1
# of Carbons	4396	3350	596	35	692
Average O/C	0.125770775	0.096446013	0.220181164	0.2	0.264426786
Average H/C (All)	1.47852258	1.398250344	1.95037312	1.285714286	1.741292785
Average H/C (Has O/C)	1.476684995	1.398250344	1.958096853	1.285714286	1.743852033
Average Aromaticity+ Index	0.072289856	0.291746142	-0.478789029	0.2	-0.958108093
Average AI-Mod+	0.320797511	0.333914929	0.243995443	0.298245614	0.228773889
Average DU	5.39516129	5.892307692	3.291666667	15	3.547619048
Average DU/C	0.328493462	0.363008358	0.138914053	0.428571429	0.224082636
Average DBE	5.762096774	5.892307692	4.416666667	16	5.69047619
Average DBE/C	0.354054951	0.363008358	0.224683442	0.457142857	0.37433687
Average MW	280.9483615	255.5275447	459.0410138	636.29665	331.0319921
Average Carbon Count	17.72580645	17.17948718	24.83333333	35	16.47619048
High O/C	0.833333333	2	2.333333333	90	23
High H/C (All)	2.333333333	2	2.333333333	1.285714286	2.333333333
High H/C (Has O/C)	2.333333333	2	2.333333333	1.285714286	2.333333333
High Aromaticity Index	0.75	0.75	0.695652174	0.2	0.695652174
High AI-Mod	0.76	0.76	0.72	0.298245614	0.72
High DU	23	13	23	15	23
High DU/C	0.769230769	0.769230769	0.766666667	0.428571429	0.766666667
High DBE	25	13	25	16	25
High DBE/C	0.833333333	0.769230769	0.833333333	0.457142857	0.833333333
High MW	974.7039	450.38591	974.7039	636.29665	974.7035
High C	52	32	52	35	49
Low O/C	0.027777778	0.03125	0.027777778	0.2	0.035714286
Low H/C (All)	0.588235294	0.588235294	0.6	1.285714286	0.6
Low H/C (Has O/C)	0.588235294	0.588235294	0.6	1.285714286	0.6
Low Aromaticity Index	-7	-0.333333333	-4	0.2	-7
Low AI-Mod	0.025641026	0.025641026	0.028571429	0.298245614	0.071428571
Low DU	0	1	0	15	0
Low DU/C	0	0.052631579	0	0.428571429	0
Low DBE	0	1	0	16	2
Low DBE/C	0	0.052631579	0	0.457142857	0.040816327
Low MW	152.11971	152.11971	157.14633	636.29665	178.07785
Low C	6	7	6	35	6

Table 11-- APCI Headington vs Citgo data, Headington side

Citgo Side	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	19%	1103	66	0	135	201	902

APCI_Citgo_HeadingtonVCitgo	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	902	84	542	29	416
% of total	100%	9%	60%	3%	46%
# of AI>0.5	133	7	88	0	42
# of AI>=0.67	14	0	12	0	3
# of AI-MOD>0.5	138	7	92	1	44
# of AI-MOD>=0.67	16	0	14	1	4
# of Carbons	27315	2209	17064	1000	12392
Average O/C	0.133554747	0.189319034	0.106526252	0.165647949	0.125783505
Average H/C (All)	1.491274028	1.641477272	1.455893341	1.913426192	1.613173732
Average H/C (Has O/C)	1.674005269	1.641477272	1.658273273	1.919732767	1.848169913
Average Aromaticity+ Index	0.193931932	0.022625129	0.215960265	-0.187262553	0.099061633
Average AI-Mod+	0.320475255	0.210578318	0.362701511	0.3581819	0.268534059
Average DU	8.366962306	5.202380952	9.370848708	3.310344828	6.286057692
Average DU/C	0.307063428	0.220660756	0.332723748	0.103692672	0.246436508
Average DBE	9.603104213	5.202380952	10.20295203	5.344827586	8.932692308
Average DBE/C	0.357528154	0.220660756	0.368050047	0.162902758	0.35452757
Average MW	460.1422218	432.2748111	473.5107655	611.8016503	476.1766715
Average Carbon Count	30.2827051	26.29761905	31.48339483	34.48275862	29.78846154
High O/C	1.034482759	2.142857143	2.4	168	30
High H/C (All)	2.4	2.142857143	2.4	2.176470588	2.4
High H/C (Has O/C)	2.4	2.142857143	2.4	2.176470588	2.4
High Aromaticity Index	1	0.619047619	1	0.192307692	1
High AI-Mod	1.08	0.627906977	1.08	1.08	1.08
High DU	30	17	30	30	30
High DU/C	1.034482759	0.636363636	1.034482759	1.034482759	1.034482759
High DBE	33	17	33	33	33
High DBE/C	1.137931034	0.636363636	1.137931034	1.137931034	1.137931034
High MW	914.95648	668.63068	876.84519	861.80874	914.95648
High C	63	42	63	55	63
Low O/C	0.024390244	0.035714286	0.024390244	0.054545455	0.03125
Low H/C (All)	0.103448276	0.818181818	0.103448276	0.103448276	0.103448276
Low H/C (Has O/C)	0.103448276	0.818181818	0.103448276	0.103448276	0.103448276
Low Aromaticity Index	-5.333333333	-1.333333333	-2.333333333	-0.727272727	-5.333333333
Low AI-Mod	0.016949153	0.01754386	0.016949153	0.138888889	0.016949153
Low DU	0	0	0	0	0
Low DU/C	0	0	0	0	0
Low DBE	0	0	0	1	2
Low DBE/C	0	0	0	0.045454545	0.037037037
Low MW	160.03454	254.11521	179.07337	352.23778	160.03454
Low C	9	13	9	17	9

Table 12-- APCI Headington vs Citgo data, Citgo side

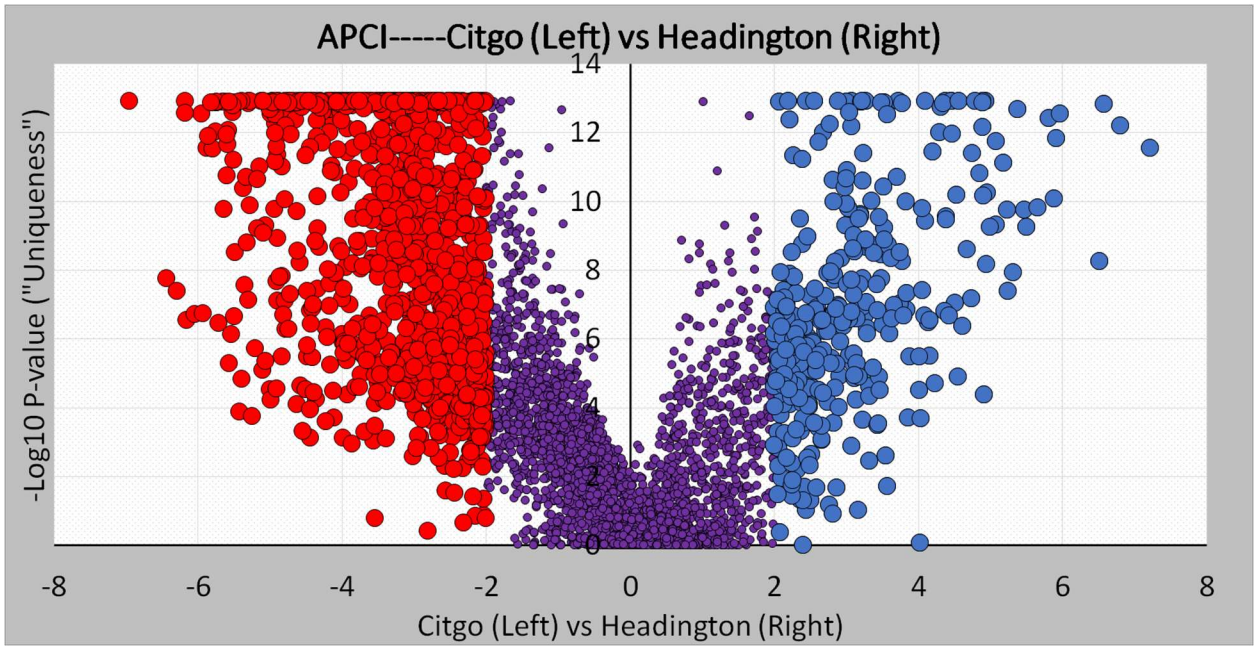

Figure 34-- APCI volcano plot of the Citgo compounds versus the Headington compounds

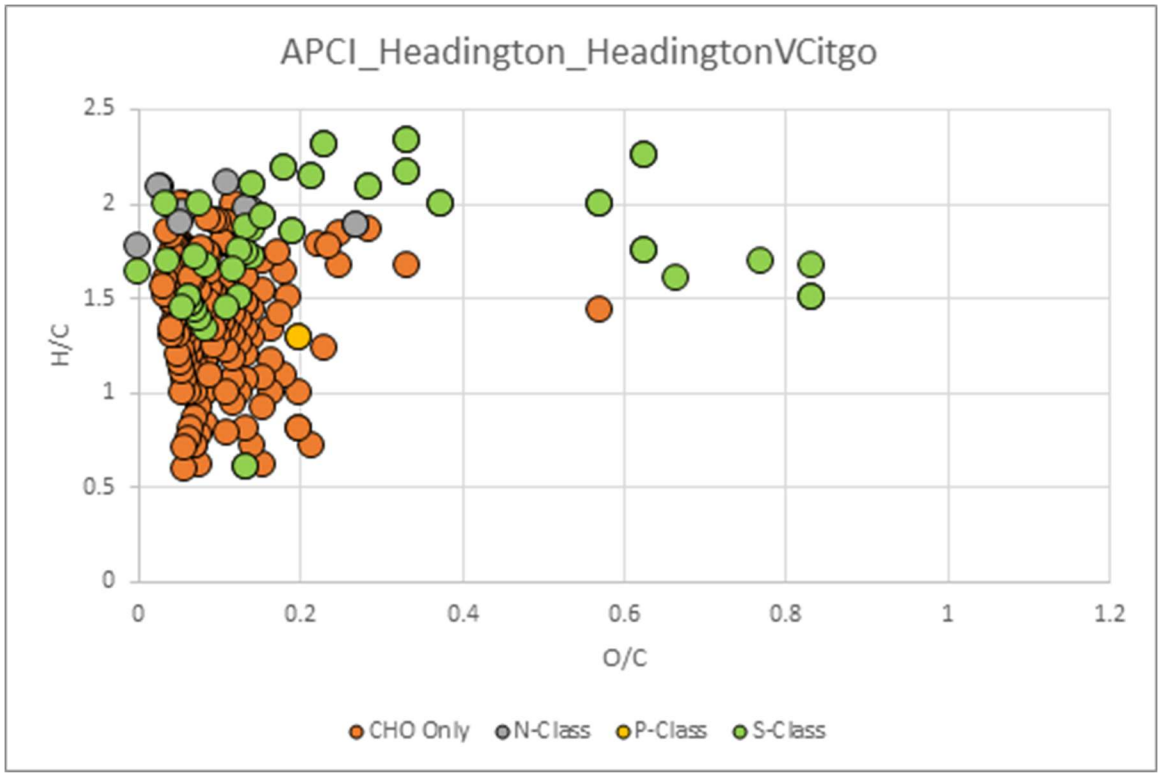

Figure 35-- Van Krevelen figure of the significant Headington compounds (right side)

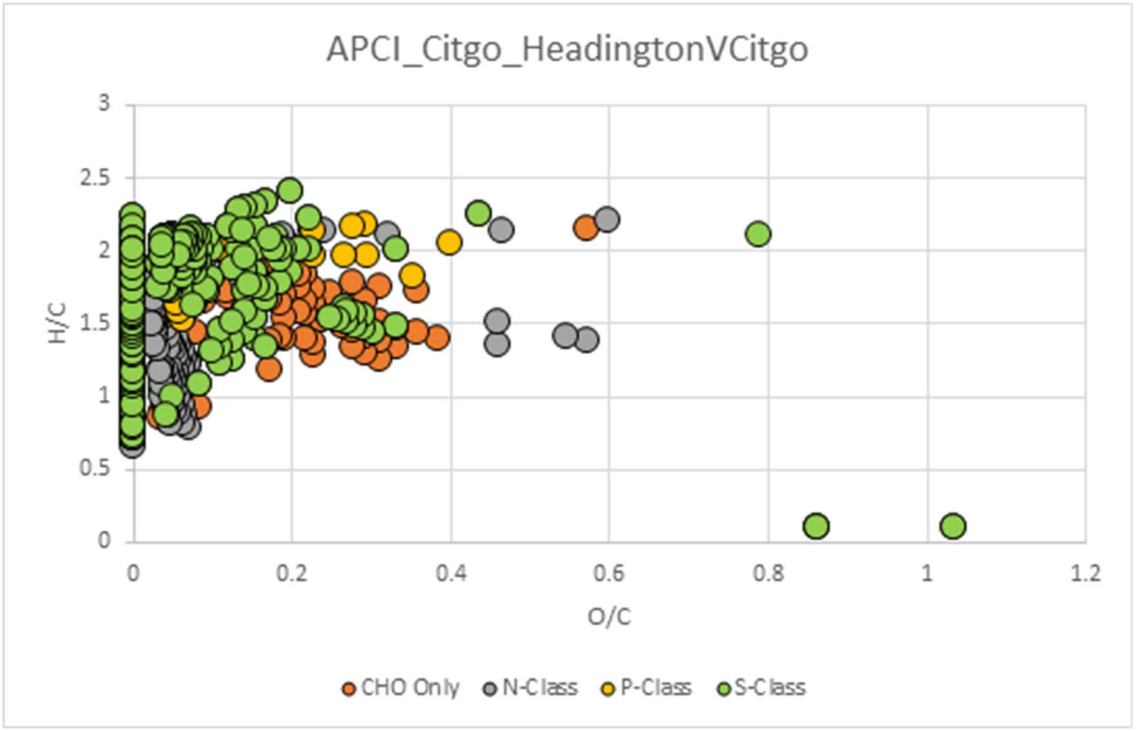

Figure 36-- Van Krevelen figure of the significant Citgo compounds (left side)

For this volcano comparison, there are many unique compounds at the –log10 P-value cap (top of the graph), meaning many unique compounds were detected for these two groups of samples. The shared compounds (center) lean towards the Citgo side, and overall, Citgo has more samples and more unique samples. Regarding the Van Krevelen data, the Headington sample group indicates many CHO only compounds relative to all the other sample types. The Citgo sample primarily consists of S-Class compounds and unoxygenated S-Class compounds.

General Notes APCI Volcano/Van Krevelen

Holistic observations are also made with the APCI Volcano Plot and Van Krevelen data.

1) The Citgo sample group has a lot of S-Class and unoxygenated compounds when ionized by APCI. This is seen with both the comparison with Cavitation and Headington.
2) The Headington sample group is primarily composed of CHO Only compounds. Compared with both Cavitation and Citgo, large clusters of CHO Only compounds are present for the Headington Van Krevelen.
3) Cavitation has similarities to both Headington and Cavitation. When compared with Headington, its unique graph is "missing a lot of CHO compounds." When compared with Citgo, it's missing a lot of "S-Class and 0-Oxy S-Class".

ESI

Headington VS Cavitation

Headington Side	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1.2	Blanks	Sum Deleted	Count (Calculated)
	11%	823	29	16	173	218	605

ESI_Headington_HeadingtonVCavitation	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	605	276	252	21	121
% of total	100%	46%	42%	3%	20%
# of AI>0.5	83	42	26	3	20
# of AI>=0.67	22	6	11	2	10
# of AI-MOD>0.5	113	64	33	2	20
# of AI-MOD>=0.67	30	13	11	1	10
# of Carbons	7415	3381	3070	209	1443
Average O/C	0.220285411	0.223972662	0.217185816	0.515292613	0.196859905
Average H/C (All)	1.485455911	1.364996988	1.602686632	1.789625765	1.644440423
Average H/C (Has O/C)	1.494150817	1.364996988	1.662206593	1.789625765	1.694244924
Average Aromaticity+ Index	0.032799838	0.166920816	-0.046700235	-1.701587302	-0.168044747
Average AI-Mod+	0.419656518	0.399581405	0.43843497	0.406636501	0.631331266
Average DU	4.58677686	4.905797101	4.242063492	2.80952381	3.983471074
Average DU/C	0.373833547	0.40535084	0.350967589	0.333439017	0.318962468
Average DBE	5.163636364	4.905797101	4.964285714	4.095238095	6.719008264
Average DBE/C	0.430567961	0.40535084	0.42321165	0.501015791	0.586046047
Average MW	216.7909479	206.4717002	222.2329662	260.134211	241.4908411
Average Carbon Count	12.25619835	12.25	12.18253968	9.952380952	11.92561983
High O/C	1	2.5	2.5	58	12
High H/C (All)	2.5	2.5	2.5	2.428571429	2.444444444
High H/C (Has O/C)	2.5	2.5	2.5	2.428571429	2.444444444
High Aromaticity Index	7	1	7	2.333333333	7
High AI-Mod	9	1	9	0.882352941	9
High DU	18	12	18	11	12
High DU/C	1.333333333	1	1.333333333	1.333333333	1.25
High DBE	18	12	18	12	16
High DBE/C	2	1	2	2	2
High MW	0	0	0	0	0
High C	32	20	32	17	26
Low O/C	0.038461538	0.052631579	0.058823529	0.142857143	0.038461538
Low H/C (All)	0.166666667	0.166666667	0.25	0.416666667	0.25
Low H/C (Has O/C)	0.166666667	0.166666667	0.25	0.416666667	0.25
Low Aromaticity Index	-7	-3	-7	-7	-5.5
Low AI-Mod	0.03030303	0.047619048	0.03030303	0.153846154	0.04
Low DU	0	0	0	0	0
Low DU/C	0	0	0	0	0
Low DBE	0	0	0	1	2
Low DBE/C	0	0	0	0.066666667	0.181818182
Low MW	0	0	0	0	0
Low C	3	4	3	3	4

Table 13-- ESI Headington vs Cavitation data, Headington side

Cavitation Side	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	0%	191	2	0	36	38	153

ESI_Cavitation_HeadingtonVCavitation	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	153	15	29		109
% of total	100%	10%	19%		71%
# of AI>0.5	22	8	12		2
# of AI>=0.67	3	2	1		0
# of AI-MOD>0.5	30	8	18		4
# of AI-MOD>=0.67	3	2	1		0
# of Carbons	2234	281	370		1583
Average O/C	0.105341263	0.185189109	0.089364604		0.097431099
Average H/C (All)	1.517077256	1.226482564	1.16396678		1.651014175
Average H/C (Has O/C)	1.517192747	1.226482564	1.030245945		1.651014175
Average Aromaticity+ Index	0.161449671	0.25277448	0.465426336		0.068007511
Average AI-Mod+	0.32560113	0.393733828	0.483144842		0.243173885
Average DU	4.424836601	7.666666667	6.689655172		3.376146789
Average DU/C	0.32660993	0.443355695	0.53800883		0.254300255
Average DBE	6.385620915	7.666666667	6.689655172		6.128440367
Average DBE/C	0.46652265	0.443355695	0.53800883		0.450691504
Average MW	255.8735702	309.9030187	195.0482107		264.6212188
Average Carbon Count	14.60130719	18.73333333	12.75862069		14.52293578
High O/C	0.35	1.8	1.25		8
High H/C (All)	2	1.8	1.5625		2
High H/C (Has O/C)	2	1.8	1.25		2
High Aromaticity Index	0.764705882	0.764705882	0.727272727		0.545454545
High AI-Mod	0.771428571	0.771428571	0.739130435		0.565217391
High DU	14	14	10		8
High DU/C	0.777777778	0.777777778	0.769230769		0.625
High DBE	14	14	10		11
High DBE/C	0.916666667	0.777777778	0.769230769		0.916666667
High MW	0	0	0		0
High C	27	24	18		27
Low O/C	0.05	0.05	0.071428571		0.071428571
Low H/C (All)	0.555555556	0.555555556	0.692307692		0.923076923
Low H/C (Has O/C)	0.555555556	0.555555556	0.692307692		0.923076923
Low Aromaticity Index	-0.307692308	-0.307692308	0.266666667		-0.25
Low AI-Mod	0.025641026	0.025641026	0.266666667		0.043478261
Low DU	1	3	5		1
Low DU/C	0.045454545	0.15	0.3125		0.045454545
Low DBE	3	3	5		3
Low DBE/C	0.15	0.15	0.3125		0.227272727
Low MW	0	0	0		0
Low C	6	9	10		6

Table 14-- ESI Headington vs Cavitation data, Cavitation side

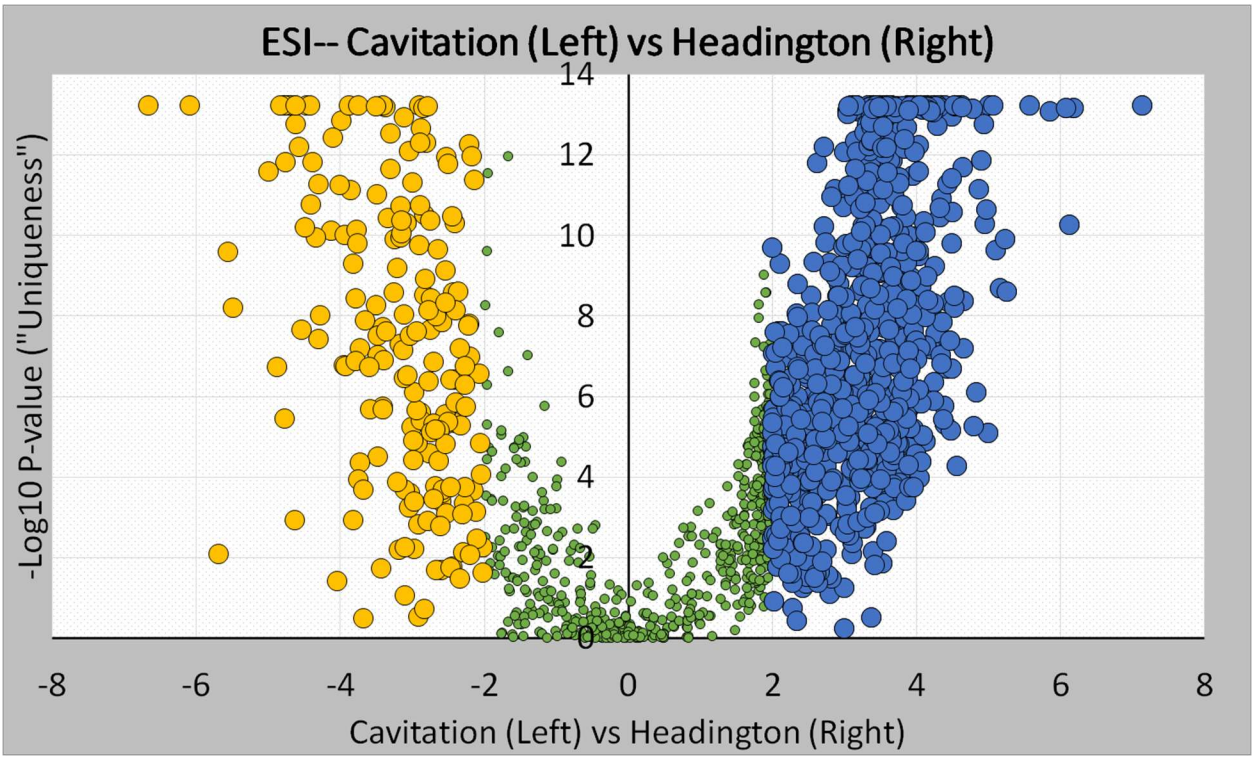

Figure 37—ESI volcano plot of the Cavitation compounds versus the Headington compounds

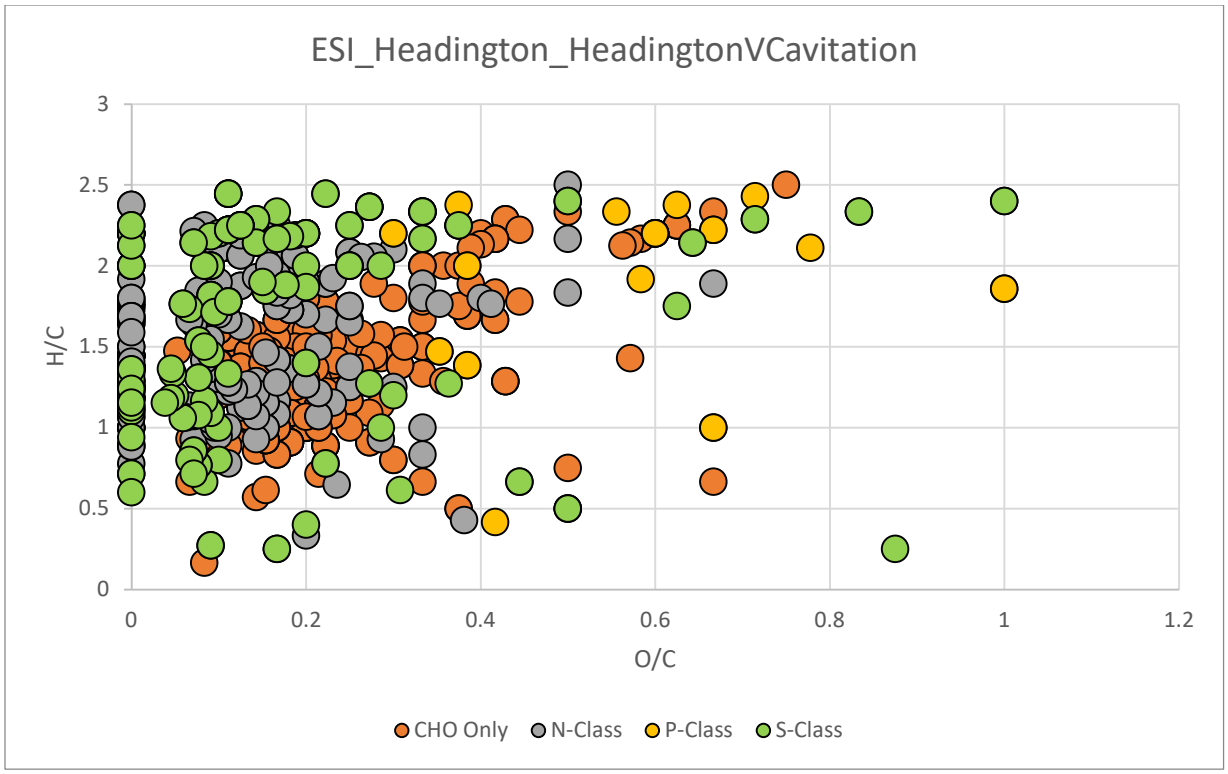

Figure 38-- Van Krevelen figure of the significant Headington compounds (right side)

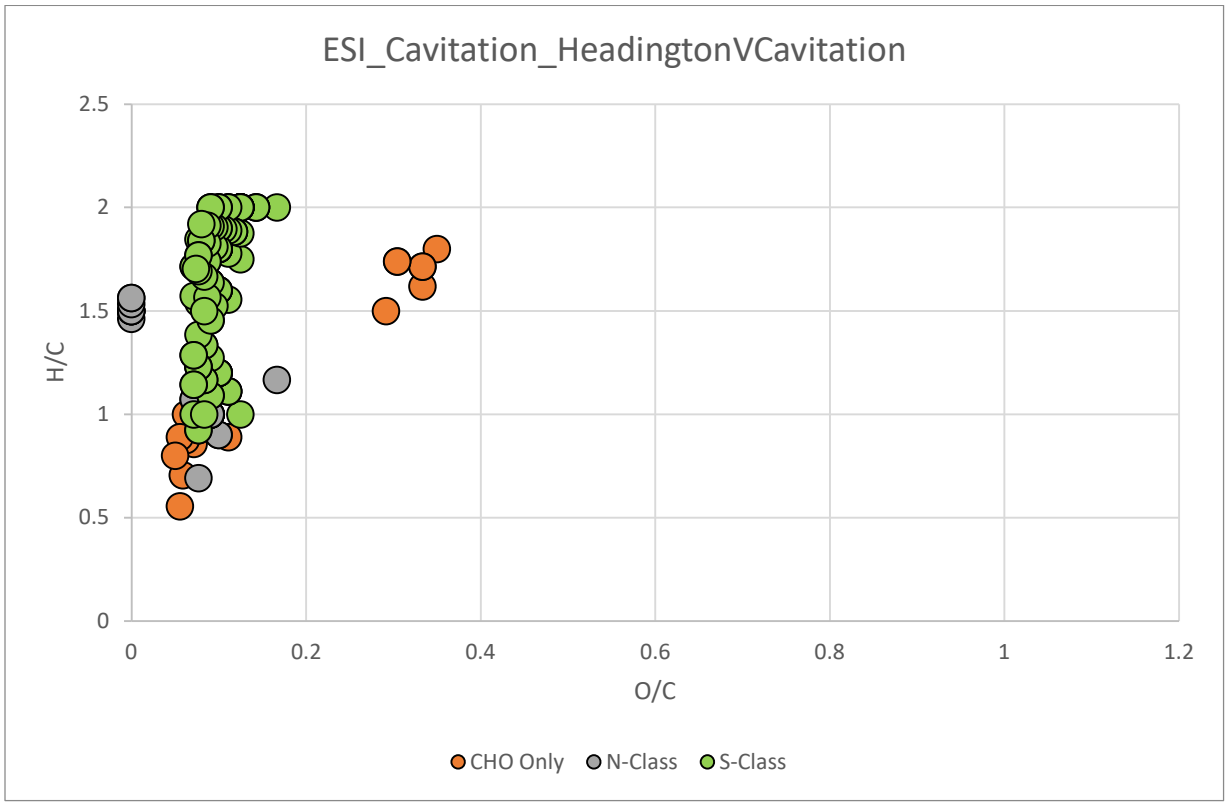

Figure 39-- Van Krevelen figure of the significant Cavitation compounds (left side)

For this volcano comparison, the compounds are stacked towards the unique Headington compound side, but both the compounds detected occupy the entirety of the X-axis and the –log10 P-values on both sides of the volcano plot, indicating that there is a wide range of unique and significant compounds. Regarding the Van Krevelen diagrams generated, the Headington sample group has many more compounds than the Cavitation sample group (as indicated on the volcano plot). There is robust clustering of N-Class, S-Class, and CHO only compounds around 0.2 O/C and between 0.5-2.5 O/C. There is also a substantial amount of unoxygenated S-Class and N-Class compounds. The Cavitation sample group is drastically different; there is a disproportionately more significant number of S-Class compounds than any other Van Krevelen, with a small CHO cluster below 1 H/C and another CHO only cluster at 0.3-0.35 O/C.

Headington VS Citgo

Headington Side	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1.2	Blanks	Sum Deleted	Count (Calculated)
	10%	299	20	7	90	117	182

ESI_Headington_HeadingtonVCitgo	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	182	118	33	9	40
% of total	100%	65%	18%	5%	22%
# of AI>0.5	28	17	3	2	9
# of AI>=0.67	7	3	2	1	3
# of AI-MOD>0.5	37	25	3	1	11
# of AI-MOD>=0.67	12	7	3	0	4
# of Carbons	2174	1388	398	91	494
Average O/C	0.20692308	0.201665678	0.258966203	0.47397811	0.152637218
Average H/C (All)	1.425827651	1.360455521	1.828585013	1.895095446	1.395231133
Average H/C (Has O/C)	1.43323296	1.360455521	1.844132006	1.895095446	1.455070525
Average Aromaticity+ Index	0.091977594	0.196690063	-0.207781836	-1.174074074	0.101124958
Average AI-Mod+	0.451881327	0.393629215	0.877381269	0.324786325	0.748042
Average DU	4.802197802	4.915254237	3.272727273	2.777777778	5.325
Average DU/C	0.393817379	0.41006792	0.263265051	0.255327316	0.427896695
Average DBE	5.423076923	4.915254237	4.818181818	4.222222222	7.975
Average DBE/C	0.449613367	0.41006792	0.412553582	0.422147007	0.663281629
Average MW	209.1430616	193.1618251	250.7408506	256.6446778	240.4387098
Average Carbon Count	11.94505495	11.76271186	12.06060606	10.11111111	12.35
High O/C	1	2.5	2.5	49	10
High H/C (All)	2.5	2.5	2.5	2.428571429	2.444444444
High H/C (Has O/C)	2.5	2.5	2.5	2.428571429	2.444444444
High Aromaticity Index	2.333333333	0.75	2.333333333	2.333333333	2.333333333
High AI-Mod	9	0.757575758	9	0.666666667	9
High DU	14	14	11	7	10
High DU/C	0.888888889	0.777777778	0.888888889	0.777777778	0.888888889
High DBE	14	14	14	10	14
High DBE/C	1.2	0.777777778	1.111111111	1.111111111	1.2
High MW	0	0	0	0	0
High C	32	20	32	17	21
Low O/C	0.052631579	0.052631579	0.066666667	0.142857143	0.066666667
Low H/C (All)	0.588235294	0.588235294	0.666666667	0.777777778	0.6
Low H/C (Has O/C)	0.588235294	0.588235294	0.666666667	0.777777778	0.666666667
Low Aromaticity Index	-6	-3	-2.2	-6	-2.2
Low AI-Mod	0.03030303	0.111111111	0.03030303	0.153846154	0.142857143
Low DU	0	0	0	0	0
Low DU/C	0	0	0	0	0
Low DBE	0	0	0	1	3
Low DBE/C	0	0	0	0.111111111	0.176470588
Low MW	0	0	0	0	0
Low C	4	4	5	7	5

Table 15-- ESI Headington vs Citgo data, Headington side

Citgo Side	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1.2	Blanks	Sum Deleted	Count (Calculated)
	1%	173	1	0	16	17	156

ESI_Citgo_HeadingtonVCitgo	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	156	8	95		55
% of total	100%	5%	61%		35%
# of AI>0.5	24	4	20		0
# of AI>=0.67	1	0	1		0
# of AI-MOD>0.5	28	4	23		1
# of AI-MOD>=0.67	1	0	1		0
# of Carbons	1809	138	1131		559
Average O/C	0.132566129	0.201585145	0.128172445		0.128255843
Average H/C (All)	1.513121678	1.347649241	1.322958539		1.871035436
Average H/C (Has O/C)	1.597511615	1.347649241	1.240030751		1.88531387
Average Aromaticity+ Index	0.235577707	0.162969659	0.329872467		-0.124280993
Average AI-Mod+	0.386906343	0.421179622	0.422607001		0.15733371
Average DU	4.406896552	6.125	5.536842105		1.745454545
Average DU/C	0.385460823	0.38860421	0.471514568		0.180906436
Average DBE	5.082758621	6.125	5.578947368		4.036363636
Average DBE/C	0.467265532	0.38860421	0.475959013		0.423764373
Average MW	191.4355796	291.4207013	179.9446093		197.291408
Average Carbon Count	11.59615385	17.25	11.90526316		10.16363636
High O/C	0.692307692	2	2.307692308		6
High H/C (All)	2.307692308	2	2.307692308		2.222222222
High H/C (Has O/C)	2.222222222	2	2.307692308		2.222222222
High Aromaticity Index	0.727272727	0.625	0.727272727		0.5
High AI-Mod	0.739130435	0.647058824	0.739130435		0.538461538
High DU	11	11	10		6
High DU/C	0.769230769	0.666666667	0.769230769		0.625
High DBE	11	11	10		8
High DBE/C	0.875	0.666666667	0.8		0.875
High MW	0	0	0		0
High C	23	23	17		23
Low O/C	0.055555556	0.055555556	0.066666667		0.076923077
Low H/C (All)	0.692307692	0.875	0.692307692		1
Low H/C (Has O/C)	0.692307692	0.875	0.692307692		1
Low Aromaticity Index	-1.666666667	-0.4	-5.5		-1.666666667
Low AI-Mod	0.043478261	0.025641026	0.185185185		0.043478261
Low DU	1	1	0		1
Low DU/C	0.066666667	0.066666667	0		0.0625
Low DBE	1	1	0		3
Low DBE/C	0.066666667	0.066666667	0		0.260869565
Low MW	0	0	0		0
Low C	4	9	7		4

Table 16--ESI Headington vs Citgo data, Citgo side

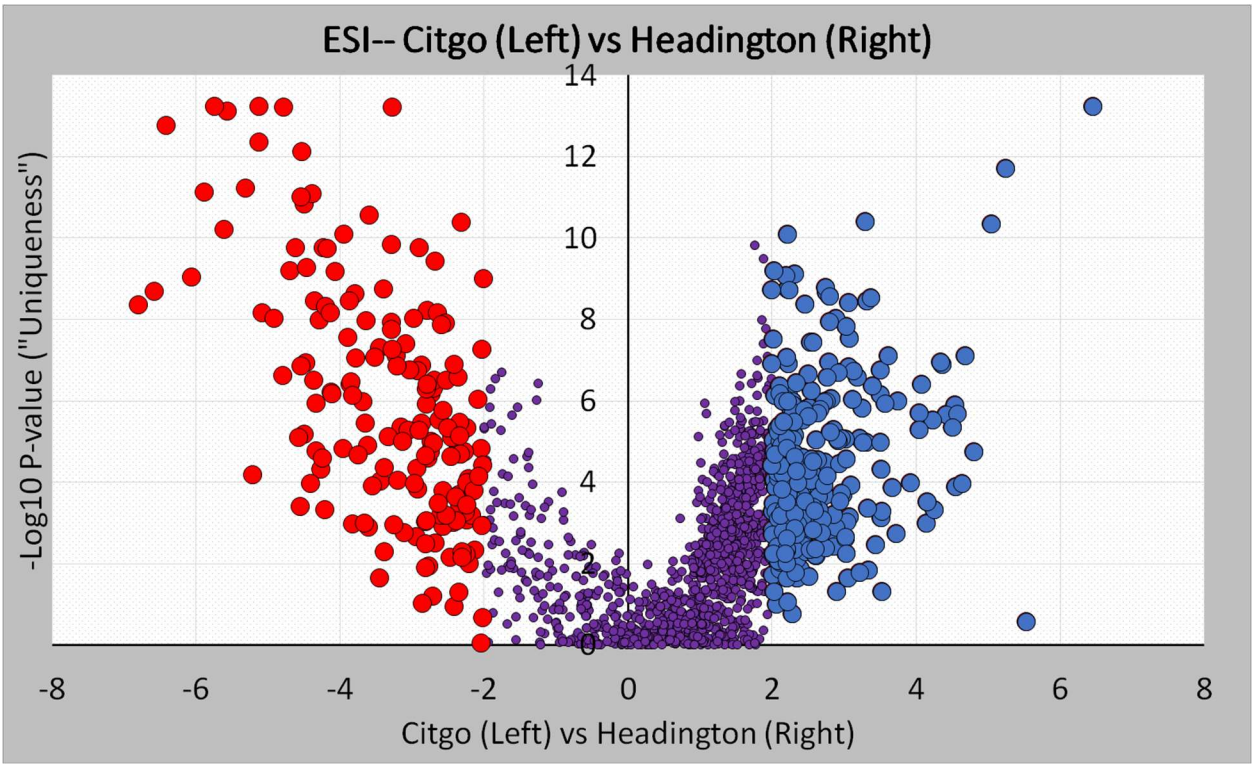

Figure 40—ESI volcano plot of the Citgo compounds versus the Headington compounds

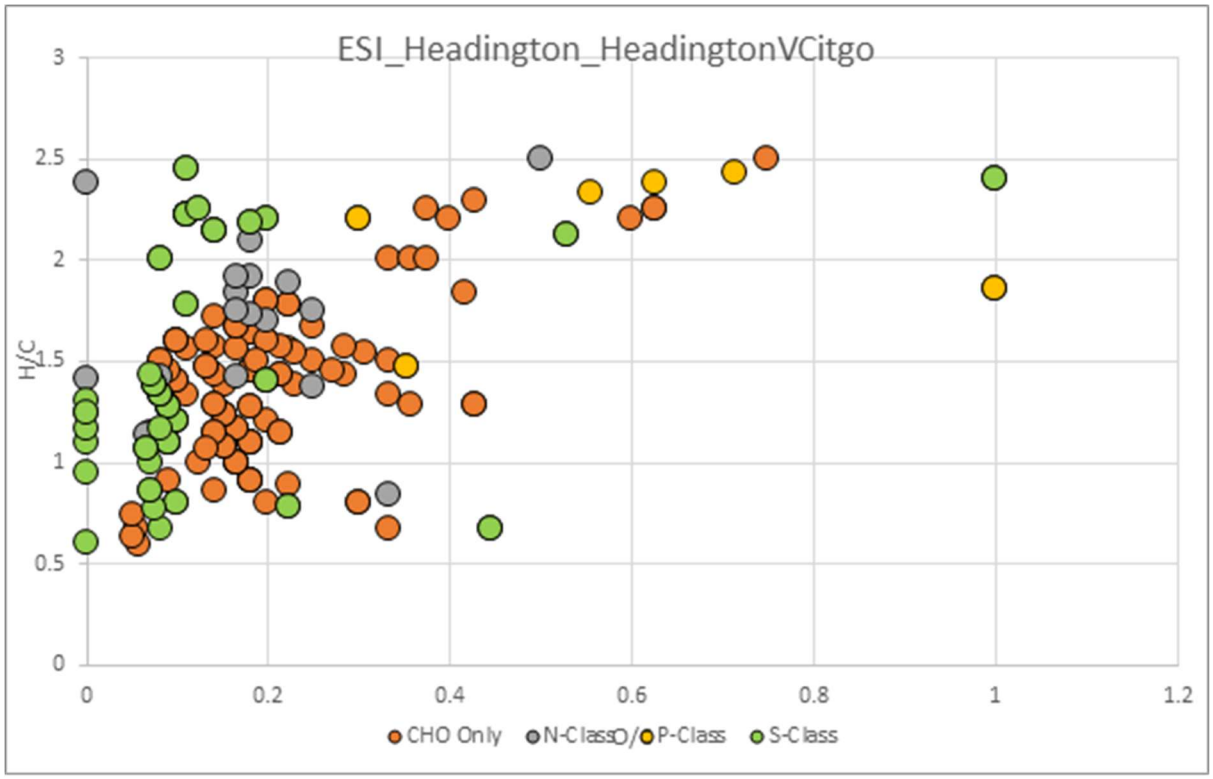

Figure 41-- Van Krevelen figure of the significant Headington compounds (right side)

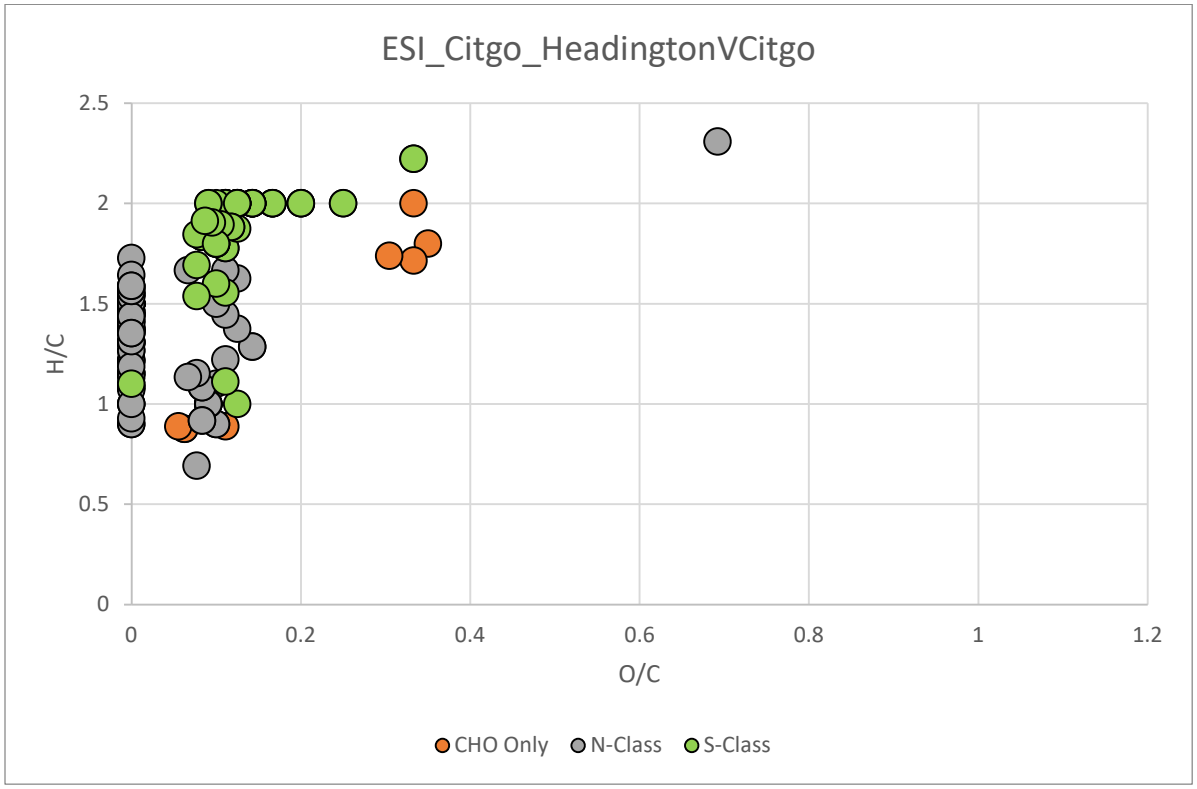

Figure 42-- Van Krevelen figure of the significant Citgo compounds (left side)

For this volcano comparison, the unique compounds are more clustered on the Headington side, while Citgo compounds are more spread out and reached higher –Log10P-Values. The standard compounds (+/- 2) are also more clustered towards the Headington side, indicating that Headington has a more significant number of compounds overall. Regarding the Van Krevelen diagrams, the Headington sample group has more compounds, with CHO "chains" and S-Class "chains" formed by the unique compounds. The Citgo Van Krevelen consists primarily of S-class compounds and N-Class compounds with small CHO clusters located at (0.3 O/C, 1.75 H/C) and (0.1 O/C, 1 H/C). There is also a disproportionate amount of unoxygenated N-Class compounds in the unique Citgo sample group.

Citgo VS Cavitation

Citgo	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	8%	301	3	3	32	38	263

ESI_Citgo_CitgoVCavitation	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	263	42	191	3	48
% of total	100%	16%	73%	1%	18%
# of AI>0.5	29	2	22	0	6
# of AI>=0.67	6	1	3	0	3
# of AI-MOD>0.5	40	7	27	0	7
# of AI-MOD>=0.67	7	1	3	0	4
# of Carbons	3053	444	2289	24	533
Average O/C	0.197598828	0.223559906	0.171116814	0.486111111	0.21725998
Average H/C (All)	1.475177569	1.458509612	1.453247228	1.597222222	1.56903336
Average H/C (Has O/C)	1.579547737	1.458509612	1.613701967	1.597222222	1.640552279
Average Aromaticity+ Index	0.154483299	0.160884354	0.192460959	-0.833333333	-0.1477859
Average AI-Mod+	0.394031408	0.361306369	0.396070756	#DIV/0!	0.465193339
Average DU	4.44486692	3.666666667	4.764397906	2.333333333	3.979166667
Average DU/C	0.393205959	0.372130581	0.412293385	0.472222222	0.355592727
Average DBE	4.920152091	3.666666667	5.083769634	4	6.541666667
Average DBE/C	0.444511606	0.372130581	0.44289382	0.791666667	0.632364641
Average m/z	192.629417	182.0156571	192.901594	221.0598833	219.6334265
Average Carbon Count	11.60836502	10.57142857	11.98429319	8	11.10416667
High O/C	0.875	2.166666667	2.444444444	43	12
High H/C (All)	2.444444444	2.166666667	2.444444444	2.375	2.444444444
High H/C (Has O/C)	2.444444444	2.166666667	2.444444444	2.375	2.444444444
High Aromaticity Index	1	0.833333333	1	0	1
High AI-Mod	1	0.846153846	1	0	1
High DU	12	6	12	5	12
High DU/C	1.25	0.857142857	1.25	1.25	1.25
High DBE	16	6	16	8	16
High DBE/C	2	0.857142857	2	2	2
High m/z	0	0	0	0	0
High C	24	18	24	12	21
Low O/C	0.047619048	0.1	0.066666667	0.375	0.047619048
Low H/C (All)	0.25	0.571428571	0.25	0.5	0.25
Low H/C (Has O/C)	0.25	0.571428571	0.25	0.5	0.25
Low Aromaticity Index	-5.5	-0.714285714	-5.5	-1.5	-5.5
Low AI-Mod	0.047619048	0.090909091	0.047619048	0	0.058823529
Low DU	0	0	0	0	0
Low DU/C	0	0	0	0	0
Low DBE	0	0	0	1	2
Low DBE/C	0	0	0	0.125	0.214285714
Low m/z	0	0	0	0	0
Low C	4	6	4	4	4

Table 17-- ESI Citgo vs Cavitation data, Citgo side

Cavitation Side	Overlap	Count (Original)	Deleted CxHy	Deleted O/C>1	Blanks	Sum Deleted	Count (Calculated)
	1%	173	8	0	0	8	165

ESI_Cavitation_CitgoVCavitation	Total Compounds	CHO Compounds	N-Class	P-Class	S-Class
# of :	138	25	12		103
% of total	100%	18%	9%		75%
# of AI>0.5	16	9	2		5
# of AI>=0.67	5	5	0		0
# of AI-MOD>0.5	21	10	4		7
# of AI-MOD>=0.67	6	6	0		0
# of Carbons	2291	491	172		1659
Average O/C	0.120726595	0.208055008	0.205659826		0.098859994
Average H/C (All)	1.429983058	1.2161357	1.38852971		1.498927767
Average H/C (Has O/C)	1.437451849	1.2161357	1.38189255		1.511319843
Average Aromaticity+ Index	0.148558866	0.245865695	-0.298247354		0.104208247
Average AI-Mod+	0.335871898	0.381200381	0.463506488		0.306825211
Average DU	5.376811594	8.16	5.916666667		4.563106796
Average DU/C	0.355924005	0.447010663	0.420277557		0.321365011
Average DBE	7.492753623	8.16	6.25		7.398058252
Average DBE/C	0.48983467	0.447010663	0.441986241		0.5007793
Average MW	292.12298	331.2129612	259.1408133		289.0031468
Average Carbon Count	16.60144928	19.64	14.33333333		16.10679612
High O/C	0.642857143	1.8	2.117647059		10
High H/C (All)	2.117647059	1.8	2.117647059		2.117647059
High H/C (Has O/C)	2.117647059	1.8	2.117647059		2.117647059
High Aromaticity Index	0.842105263	0.842105263	0.625		0.555555556
High AI-Mod	0.846153846	0.846153846	0.647058824		0.565217391
High DU	17	17	9		10
High DU/C	0.85	0.85	0.7		0.615384615
High DBE	17	17	9		12
High DBE/C	0.916666667	0.85	0.7		0.916666667
High MW	0	0	0		0
High C	28	28	18		28
Low O/C	0.05	0.05	0.0625		0.066666667
Low H/C (All)	0.4	0.4	0.9		0.923076923
Low H/C (Has O/C)	0.4	0.4	0.9		0.923076923
Low Aromaticity Index	-5	-0.307692308	-5		-5
Low AI-Mod	0.025641026	0.025641026	0.310344828		0.041666667
Low DU	1	3	1		1
Low DU/C	0.045454545	0.15	0.058823529		0.045454545
Low DBE	1	3	1		3
Low DBE/C	0.090909091	0.15	0.090909091		0.176470588
Low MW	0	0	0		0
Low C	8	10	10		8

Table 18-- ESI Citgo vs Cavitation data, Cavitation side

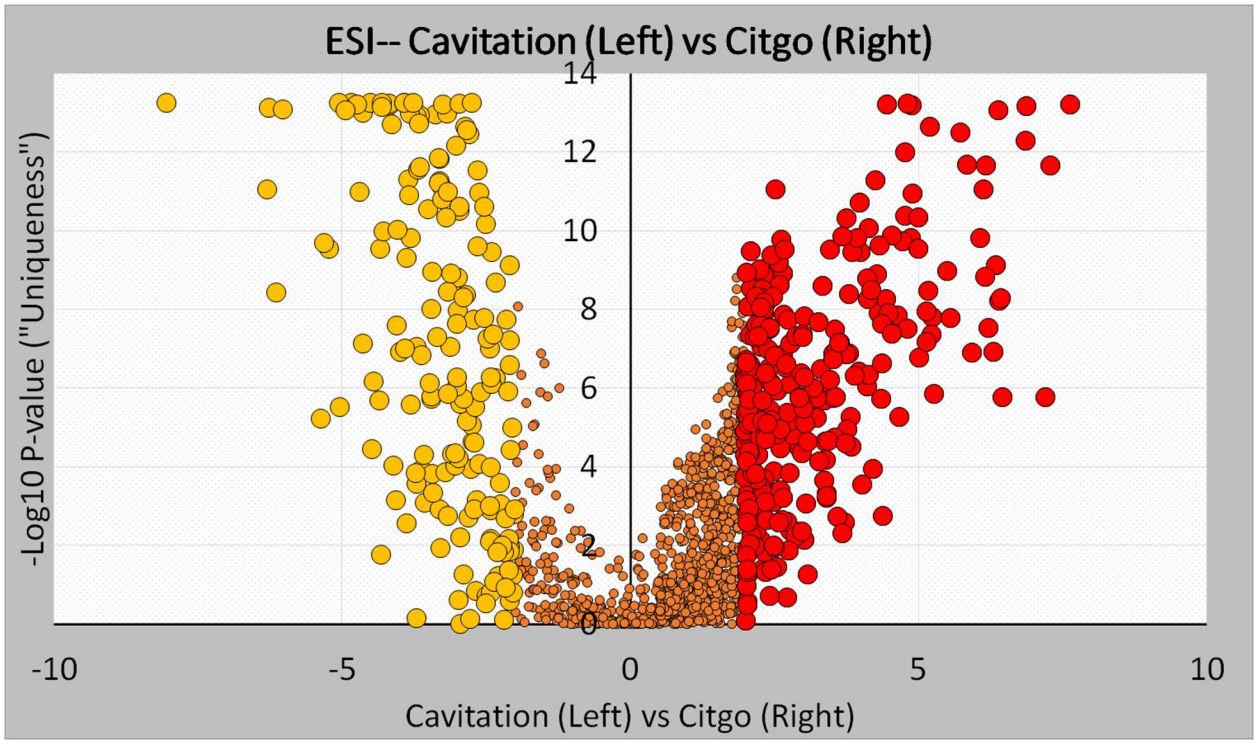

Figure 43-- ESI volcano plot of the Cavitation compounds versus the Citgo compounds

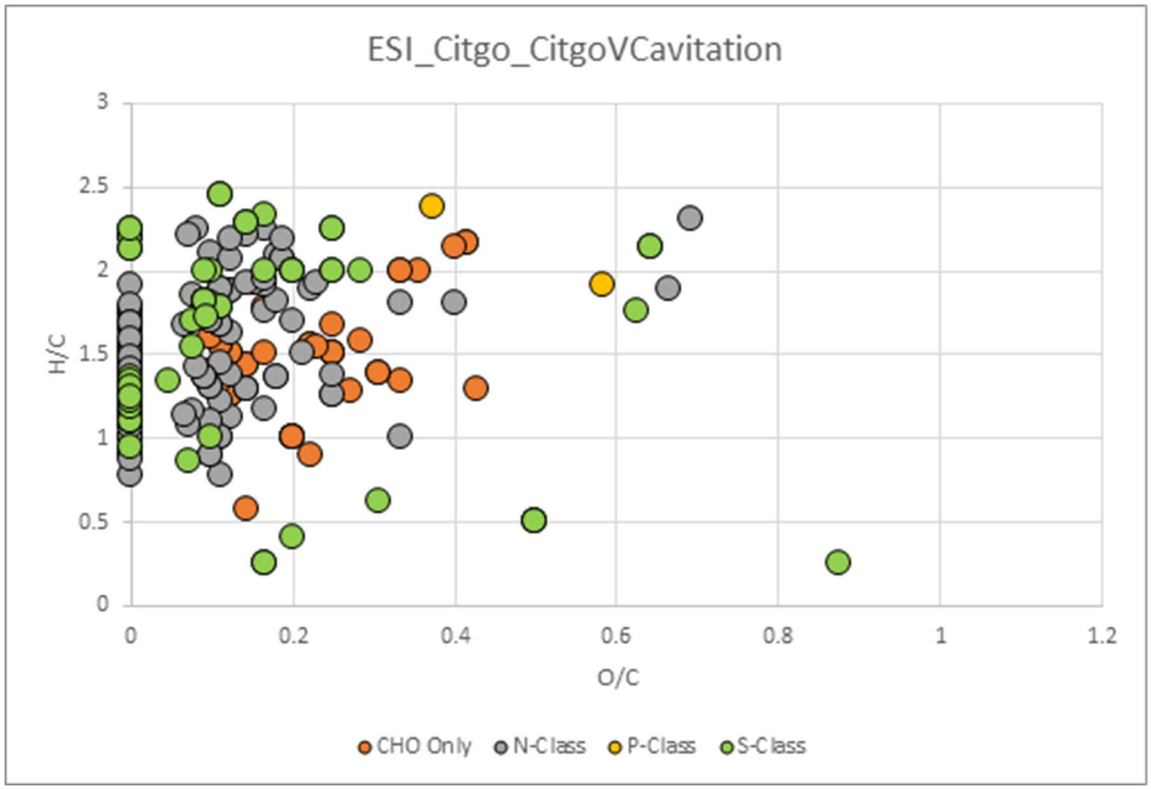

Figure 44-- Van Krevelen figure of the significant Citgo compounds (right side)

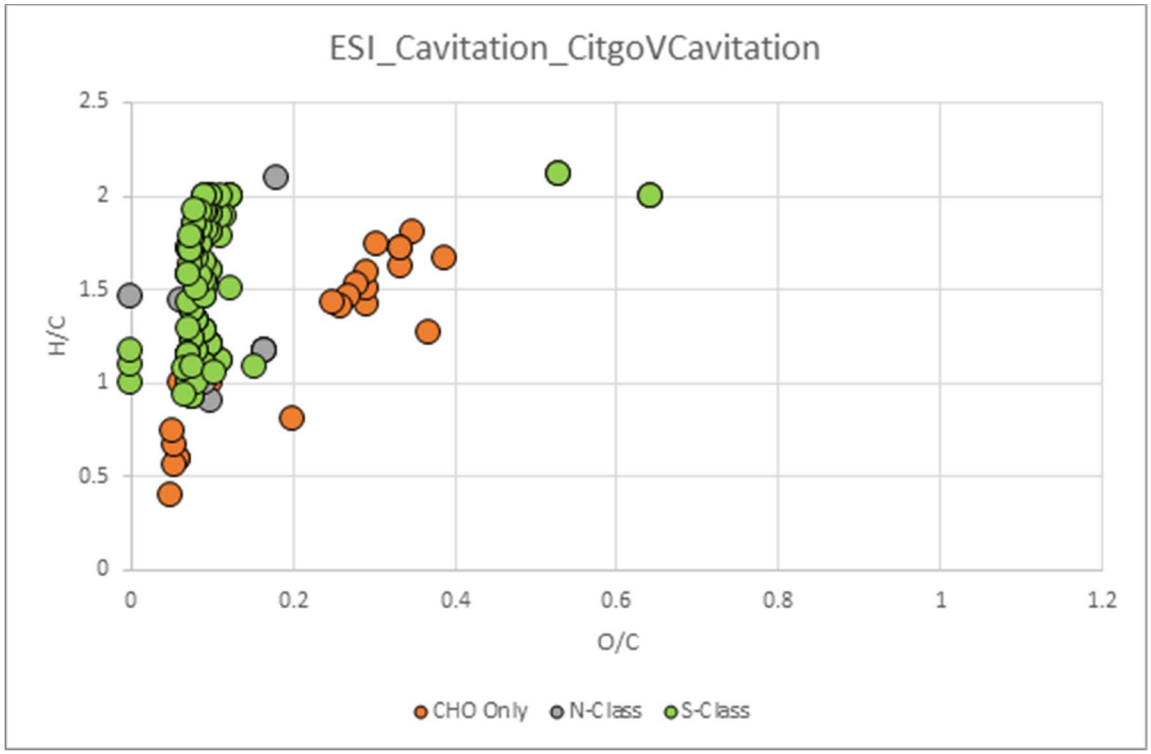

Figure 45-- Van Krevelen figure of the significant Cavitation samples (left side)

The compounds are stacked towards the Citgo side of this volcano comparison. Both sides of the volcano plot have high –Log10 P-Values. The Citgo Van Krevelen has large amounts of N-Class compounds, along with unoxygenated N-Class and S-Class compounds on the H/C axis. However, the Cavitation Van Krevelen primarily consists of a large S-Class cluster ranging from 1-2 H/C at 0.1 O/C. There is also a CHO only cluster located at (0.3 O/C, 1.5 H/C).

General Notes ESI Volcano/Van Krevelen

Holistic observations were acknowledged with the ESI volcano plot Volcano Plot and Van Krevelen data.

1) The Headington sample group had the most compounds detected with ESI ionization, sharing most of its molecular makeup with the Citgo sample group rather than the Cavitation sample group.
2) Most of the unique compounds from the Citgo sample group are unoxygenated (has an O/C value of 0).
3) The cavitation sample group ionized by ESI was revealed to have a predominant amount of unique S-Class compounds.

CHAPTER V - DISCUSSION

From preliminary data, inferences can be made with the figures created in conjunction with available metadata. These inferences pertain to the 1) Cavitation sample group and the 2) Diesel Mud Base and Sarita Sogu 402 samples.

The Cavitation samples stem from a project that utilized ultrasonic bubble cavitation to refine crude oil. While cavitation technology is a relatively new technology, cavitation technologies have improved the petroleum industry by affecting processes such as crude oil viscosity reduction, oxidative desulphurization/demetallization, and crude oil upgrading (Avvaru et al. 2018). The cavitation samples analyzed with OT-FT-MS are highly similar in composition, based on their positions in the PCA analysis (Figure 7-10). The cavitation project claims that their cavitation process fragmented heavier chain compounds (C31+) to lower molecular weight compounds, but further data analysis is required to confirm this claim.

Based on the data analyzed in this project at this time, there also seems to be many unique sulfur compounds (Figure 18; Figure 22) persisting throughout the entire cavitation process. Experimental refining methods such as ultrasonic bubble cavitation can be further studied with high-resolution mass spectrometry to confirm refining efficiency and efficacy.

The Diesel Mud Base and Sarita Sogu 402 also present a strong relationship in the data gathered. Diesel mud, or "oil-based mud," is a slurry of diesel and local sediment/dirt utilized for bore head lubrication (Osborne and Rahimtula, 1985). A significant detail observed in Figure 3 is that the Diesel Mud Base sample is red—this indicates that the fuel used was untreated, untaxed, and only allowed for non-vehicular purposes (Varughese and Khawaja, 2017). A distilled, "hydrocarbon pure" fuel such as red diesel has a unique compound signature compared to only crude oil samples, which is represented in the APCI PCA analyses (Figure 8; Figure 10). It

should be noted that Sarita Sogu 402 is highly similar to Diesel Mud Base. Obtaining the Headington sample metadata revealed the reason behind this; Sarita Sogu 402 was a sample taken immediately after mud-based drilling occurred, meaning that the field diesel contaminated the chemical composition of the initial Sarita Sogu 402 well sample. Oil-based mud contamination is a relatively modern concept (Petersen et al., 2017; Fadnes et al., 2001), but our data provides a unique, mass spectrographic perspective on the issue of oil-based mud contamination. In theory, the chemical composition of the oil-based mud becomes diluted with extensive crude oil recovery. Still, additional well samples throughout drilling would be necessary to confirm this.

Based on these two preliminary inquiries, the instrumentation, and methodology displayed in this project present research areas for further impact in addition to the areas mentioned earlier (SAPOR). Because this fuel database research is still ongoing and requires further expansion, the two strong inferences presented are still reliant on background metadata to make strong inferences. However, when more samples are accumulated in the database, progressively less background information will be needed, and background metadata can be used to confirm observations that were previously investigated solely.

CHAPTER VI - CONCLUSION

The field of petroleomics is relatively new, with the term "coined" in 2008 (Marshall and Rodgers, 2008). Since then, the number of studies merging mass spectrometry with petroleum chemistry has slowly escalated alongside improvements in instrumentation and methodology. The ultimate purpose of this study is to implement the mass spectrometry methods acquired from other fields of molecular chemistry while exploring new and original ways to disseminate and display the myriad of molecular compounds discovered in a single injection of crude oil.

Regarding the instrumentation application in this study, our petroleomics method involving OT-FT-MS and unique data analysis has shown novel and valuable techniques for the identification of petroleum-- by utilizing different ionization methods, it is possible to reveal both the polar and non-polar composition of petroleum. In the future, we hope to make more improvements to our approach through the accumulation of samples and additional ionization methods, such as Atmospheric Pressure Photoionization (APPI).

The second goal of this treatise on petroleomics is to find innovative ways to display the information generated through the spectrometric analysis of petroleum samples. This research has demonstrated visualizations currently used in the "–omics fields" (Van Krevelens, PCAs/Loading graphs, volcano plots). Still, we have applied unique approaches to our study, such as creating separate "Van Krevelen sections" for volcano plots or loading graphs. As this thesis was written, several ideas not seen in literature have been realized, such as the color-coding of sample groups *on* the volcano plots or the further separation of compound classes with Van Krevelen diagrams (CHON, CHONP, CHONS, etc.). Concepts such as these will be explored in future publications alongside new data.

REFERENCES

Avvaru, Balasubrahmanyam, et al. "Current knowledge and potential applications of cavitation technologies for the petroleum industry." *Ultrasonics sonochemistry* 42 (2018): 493-507.

Blake, Donald R., and F. Sherwood Rowland. "Urban leakage of liquefied petroleum gas and its impact on Mexico City air quality." Science 269.5226 (1995): 953-956.

Boduszynski, Mieczyslaw M. "Composition of heavy petroleums. 2. Molecular characterization." *Energy & Fuels* 2.5 (1988): 597-613.

Bollet, C., et al. "Rapid separation of heavy petroleum products by high-performance liquid chromatography." *Journal of Chromatography A* 206.2 (1981): 289-300.

Boyd, T. J., C. L. Osburn. K. J. Johnson, K. B. Birgl and R. B. Coffin. 2006 Compound-Specific Isotope Analysis Coupled with Multivariate Statistics to Source-Apportion Hydrocarbon Mixtures. Environ. Sci. Technol. 40:1916-1924.

Bray, E. E., and E. D. Evans. "Distribution of n-paraffins as a clue to recognition of source beds." *Geochimica et Cosmochimica Acta* 22.1 (1961): 2-15.

Coffin, R.B., Pohlman, J.W., Grabowski, K.S., Knies, D.L., Plummer, R.E., Magee, R.W., Boyd, T.J. 2008. Radiocarbon and stable carbon isotope analysis to confirm petroleum natural attenuation in the vadose zone. Environmental Forensics 9:75-84

Desty, D. H., and B. H. F. Whyman. "Application of gas-liquid chromatography to analysis of liquid petroleum fractions." *Analytical Chemistry* 29.3 (1957): 320-329.

Drushel, Harry V., and Adam L. Sommers. "Isolation and characterization of sulfur compounds in high-boiling petroleum fractions." *Analytical Chemistry* 39.14 (1967): 1819-1829.

Eckelmann, Walter R., et al. "Implications of carbon isotopic composition of total organic carbon of some recent sediments and ancient oils." *AAPG bulletin* 46.5 (1962): 699-704.

Eglinton, G., and A. R. Galbraith. "182. Macrocyclic acetylenic compounds. Part I. Cyclo tetradeca-1: 3-diyne and related compounds." *Journal of the Chemical Society (Resumed)* (1959): 889-896.

Eide, Ingvar, and Kolbjørn Zahlsen. "A novel method for chemical fingerprinting of oil and petroleum products based on electrospray mass spectrometry and chemometrics." *Energy & fuels* 19.3 (2005): 964-967.

Ellis, Leroy, and Ann L. Fincannon. "Analytical improvements in irm-GC/MS analyses: advanced techniques in tube furnace design and sample preparation." Organic geochemistry 29.5-7 (1998): 1101-1117.

Fadnes, Finn Hallstein, et al. "Optimization of Wireline Sample Quality by Real-Time Analysis of Oil-Based Mud Contamination-Examples from North Sea Operations." *SPE Annual Technical Conference and Exhibition*. OnePetro, 2001.

Frysinger, Glenn S., and Richard B. Gaines. "Comprehensive two-dimensional gas chromatography with mass spectrometric detection (GC× GC/MS) applied to the analysis of petroleum." *Journal of High-Resolution Chromatography* 22.5 (1999): 251-255.

Gallegos, Emilio J., et al. "Petroleum group-type analysis by high resolution mass spectrometry." *Analytical chemistry* 39.14 (1967): 1833-1838.

Kelley, Cheryl A., Beth Trust Hammer, and Richard B. Coffin. "Concentrations and stable isotope values of BTEX in gasoline-contaminated groundwater." *Environmental science & technology* 31.9 (1997): 2469-2472.

Kind, Tobias, and Oliver Fiehn. "Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry." BMC bioinformatics 8.1 (2007): 1-20.

Koch, Boris P., and T. Dittmar. "From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter." Rapid communications in mass spectrometry 20.5 (2006): 926-932.

Li, Wentian. "Volcano plots in analyzing differential expressions with mRNA microarrays." *Journal of bioinformatics and computational biology* 10.06 (2012): 1231003.

Mair, Beveridge, and Frederick Rossini. "Summary of Work of the American Petroleum Institute Research Project 6 on Hydrocarbons in the C 13 to C 38 Fraction of Petroleum." Symposium on Composition of Petroleum Oils, Determination and Evaluation. ASTM International, 1958.

Maalouf, Christophe Bassem, et al. "Unlocking the Hidden Potential of an Upper Jurassic Carbonate Reservoir After 46 Years of Field Discovery: A Case Study in Offshore Abu Dhabi." Abu Dhabi International Petroleum Exhibition & Conference. OnePetro, 2017.

Marshall, Alan G., and Ryan P. Rodgers. "Petroleomics: Chemistry of the underworld." *Proceedings of the National Academy of Sciences* 105.47 (2008): 18090-18095.

Moldowan, J. Michael, Wolfgang K. Seifert, and Emilio J. Gallegos. "Relationship between petroleum composition and depositional environment of petroleum source rocks." AAPG bulletin 69.8 (1985): 1255-1268.

Nganje, T. N., A. E. Edet, and S. J. Ekwere. "Distribution of PAHs in surface soils from petroleum handling facilities in Calabar." Environmental monitoring and assessment 130.1 (2007): 27-34.

Payne, Jerry F., et al. "Mixed-function oxygenases as biological monitors around petroleum hydrocarbon development sites: Potential for induction by diesel and other drilling mud base oils containing reduced levels of polycyclic aromatic hydrocarbons." *Marine Environmental Research* 17.2-4 (1985): 328-332.

Petersen, H. I., M. Hertle, and H. Sulsbrück. "Upper Jurassic–lowermost Cretaceous marine shale source rocks (Farsund Formation), North Sea: Kerogen composition and quality and the adverse effect of oil-based mud contamination on organic geochemical analyses." *International Journal of Coal Geology* 173 (2017): 26-39.

Pond, Kristy L., et al. "Hydrogen isotopic composition of individual n-alkanes as an intrinsic tracer for bioremediation and source identification of petroleum contamination." *Environmental science & technology* 36.4 (2002): 724-728.

Qian, Kuangnan, and Gary J. Dechert. "Recent advances in petroleum characterization by GC field ionization time-of-flight high-resolution mass spectrometry." *Analytical chemistry* 74.16 (2002): 3977-3983.

Qian, Kuangnan, et al. "Resolution and identification of elemental compositions for more than 3000 crude acids in heavy petroleum by negative-ion microelectrospray high-field Fourier transform ion cyclotron resonance mass spectrometry." Energy & Fuels 15.6 (2001): 1505-1511.

Reynolds, John G., Richard W. Crawford, and Alan K. Burnham. "Analysis of oil shale and petroleum source rock pyrolysis by triple quadrupole mass spectrometry: comparisons of gas evolution at the heating rate of 10. degree. C/min." *Energy & fuels* 5.3 (1991): 507-523.

Rosenberg, Yoav O., et al. "Study of thermal maturation processes of sulfur-rich source rock using compound specific sulfur isotope analysis." *Organic Geochemistry* 112 (2017): 59-74.

Silverman, Sol R., and Samuel Epstein. "Carbon isotopic compositions of petroleums and other sedimentary organic materials." *AAPG Bulletin* 42.5 (1958): 998-1012.

U.S. Energy Information Administration - EIA - Independent Statistics and Analysis. In: Refining crude oil - the refining process - U.S. Energy Information Administration (EIA). https://www.eia.gov/energyexplained/oil-and-petroleum-products/refining-crude-oil-the-refining-process.php. Accessed 13 Jul 2021

Varughese, Bindu Sara, and Hassan Abbas Khawaja. "Detection of red dye in diesel oil." (2017).

Wang, Chuanyuan, et al. "Fingerprint and weathering characteristics of crude oils after Dalian oil spill, China." Marine pollution bulletin 71.1-2 (2013): 64-68.

Yao, Pei-Hsuan, et al. "Lead isotope characterization of petroleum fuels in Taipei, Taiwan." *International journal of environmental research and public health* 12.5 (2015): 4602-4616.