

Peak-cognizant Signal Processing of Raw Instrument Signals to Quantify Environmental Weathering of Contaminants from the Deepwater Horizon Spill

Bernice Kubicek

*Electrical and Computer Engineering
University of Iowa*

Iowa City, IA, United States

bernice-kubicek@uiowa.edu

Ananya Sen Gupta

*Electrical and Computer Engineering
University of Iowa*

Iowa City, IA, United States

ananya-sengupta@uiowa.edu

Fabian MullerDahlberg

*Electrical and Computer Engineering
University of Iowa*

Iowa City, IA, United States

fabian-mullerdahlberg@uiowa.edu

Alexandra Zelenski

*Electrical and Computer Engineering
University of Iowa*

Iowa City, IA, United States

alexandra-zelenski@uiowa.edu

Roberto Wong

*Environmental Sciences
Louisiana State University*

Baton Rouge, LA, United States

evwong@lsu.edu

Edward Overton

*Environmental Sciences
Louisiana State University*

Baton Rouge, LA, United States

ebovert@lsu.edu

Abstract—In this work, we present peak-cognizant quantification of environmental weathering of crude oil from the from the *Deepwater Horizon* oil spill. The key idea is to autonomously extract peak information from raw gas chromatography-mass spectrometry (GC-MS) signals from crude oil samples, and represent the relative weathering of different peaks in a graph-based quantitative computational framework. We also present results from pre-processing the raw signals with baseline correction and signal normalization. Retention time alignment is performed by first aligning the source oil by determining the retention time drift between prominent peaks within the signals and applying the calculated drift to the weathered oil samples. Peak finding, validation, and grouping of the five weathered oil samples to a source oil sample allows compound associations to be discovered. We present preliminary results as graphical visualizations allowing for rapid and precise interpretation of weathering compounds within polycyclic aromatic hydrocarbons (PAH). Results presented were generated with oil samples showing different degrees of weathering collected from the Deepwater Horizon spill.

Index Terms—gas chromatography-mass spectrometry, peak processing, oil weathering, retention time alignment

I. INTRODUCTION

On April 20, 2010 an explosion occurred on the Deepwater Horizon oil rig, resulting in the release of an estimated 4.9 million barrels of oil into the Gulf of Mexico [1]. Over the next few months the shorelines of Alabama, Florida, Louisiana, and Mississippi were contaminated with crude oil [1]. Ten years later, the important question of quantifying whether and how

This research was funded by the National Science Foundation under grant number 1808463. This research was also made possible in part by a grant from The Gulf of Mexico Research Initiative. Data are publicly available through the Gulf of Mexico Research Initiative Information Data Cooperative (GRIIDC) at <https://data.gulfresearchinitiative.org>

much the different toxic contaminants that constituted the spilled crude oil weathered over time, still lingers.

Complex mixtures, such as crude oil released by the Deepwater Horizon disaster, are composed of hundreds or thousands of compounds. These compounds manifest as peaks in raw instrument signals produced from common chromatographic techniques, such as gas chromatography mass spectrometry (GC-MS). Many of these compounds are known to be toxic to humans and can persist in marine and coastal environments in varying degrees. Due to the thousands of compounds it is impractical and challenging to manually examine individual compounds present in each sample [2]. Furthermore, traditional compound-cognizant analysis utilizes a limited quantity of the raw GC-MS signal, where only the known (target) peaks are considered. Chemometric statistical studies analyzing compounds in both crude oils and petroleum fuels while the remaining GC-MS signal remains unused [3]. The remaining unused information in the raw GC-MS signals will allow quantification of compound weathering and degradation in the marine environment, aiding in the study of long-term impact of off-shore drilling and oil spills to public health, safety, and environment sustainability.

The remaining of the paper is organized as follows: Section II describes recent work in the field and highlights the key contributions of this work. Section III introduces the field data and describes the algorithm. Resulting visualizations are presented in Section IV for two analyzed ions, and concluding remarks are listed in Section V.

II. RELATED WORK

The impact of the Deepwater Horizon spill on the environment and wildlife has been substantially studied over the past decade [3]–[12]. Of specific interest to the presented work is the weathering of harmful PAHs as described in [3]–[5], [8]–[10].

Numerous researchers have created algorithms for retention time correction when examining raw GC-MS signals. Two of the most common retention time correction algorithms are the correlation optimized warping (COW) algorithm and the dynamic time warping (DTW) algorithm [13], [14]. Typically, after retention time alignment associations are determined via principle component analysis (PCA) or linear discriminant analysis (LDA), both have been successful when applied to raw oil GC-MS signals in [15], [16].

We differentiate ourselves from the previously described research through retention time alignment via a novel algorithm - allowing alignment of heavily weathered raw signals - and generating associations via a peak-cognizant iterative approach prior to generating meaningful, compact representations allowing for rapid and precise interpretation of PAH compounds.

III. TECHNICAL APPROACH

Of the four states impacted by the Deepwater Horizon spill, we demonstrate our algorithm on samples taken from five locations throughout the Louisiana marshland between 2010 and 2015. After raw signal extraction from the GC-MS, there are four main stages of the algorithm: (i) calibration of the raw source and weathered oil signals, (ii) peak (compound) searching and validation in the source and weathered sample data, (iii) grouping of the weathered peaks to a corresponding source peak, and (iv) generation of meaningful visualizations for quick interpretation on compound weathering. A brief description of the stages follow.

A. Chemical Analyses

An Agilent 7890A Gas Chromatograph with an inert XL mass selective detector Agilent 5975C fitted with a HP-5MS high resolution capillary column was used for chemical analysis. The capillary column was 30 m long, 0.25 mm in diameter, with 0.25 μm thick film. Ultrahigh purity helium from Air Liquide, Houston, TX at a flow rate of 1ml/min was used as the carrier gas. A Hewlett-Packard single tapered deactivated borosilicate liner was attached to the injection port, ran in splitless mode and set to 250°C. The temperature was initially set and held at 60°C for 3min, then increased by 5°C each minute until 280°C was reached, this was held for 3 min. Lastly, the temperature was increased by 1.5°C each minute until 300°C was reached, this temperature was held constant for 2 min. The interface between the mass selective detector and MS was 280°C. Specific PAHs were quantified by running the mass selective detector in selective ion monitoring mode.

Response factors calculated by internal and calibration standards were used to find specific target PAH concentrations. A 5-point calibration curve generated for calibration standards

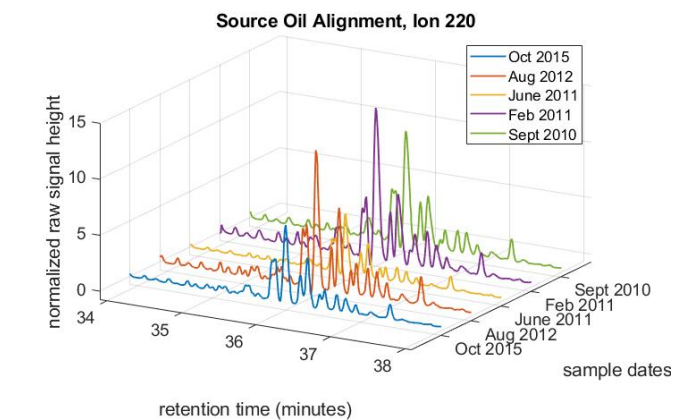


Fig. 1. Baseline corrected, C30 hopane normalized, and retention time corrected signals for C-3 Phenanthrene, Ion 220.

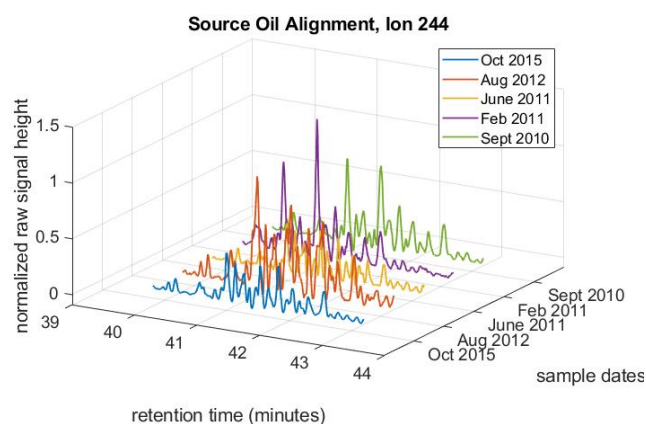


Fig. 2. Baseline corrected, C30 hopane normalized, and retention time corrected signals for C-3 Pyrene, Ion 244.

included parent PAHs. From the calibration curves the instrument response factors were used to calculate concentrations of individual analytes in the extracted samples. The quantified results of alkylated homologues are semi-quantitative, as they were calculated by response factors generated by un-alkylated parent PAH compounds.

An Oil Analysis Standard, Lot #121004, from Absolute Standards was used to prepare the 5-point calibration standards. Amber vials with PTFE-lined caps held the solution of calibration standards. These were frequently checked for degradation and replaced if necessary. In each extracted sample batch or 12-hour period during analyses one point of the original 5-point calibration standard was processed. If this calibration standard was within $\pm 20\%$ of the mean relative response factor of the initial 5-point curve it was accepted. The procedural blanks were below 5% of the target analytes. In each sample batch the source oil was analyzed as a QA/QC sample. The average surrogate recovery was approximately 74%, ranging between 55% and 106% for the Phenanthrene- d_{10} surrogate standard. The resulting target analyte concentrations were recovery corrected.

B. Calibration of Raw Data

The first step in calibration of the raw data is baseline correction of all signals to remove the baseline offsets from the data. Next, normalization of both the source and weathered data is performed through division with the corresponding C30 hopane biomarker compound found in the samples respective Hopane ion. Normalization with the hopane biomarker is done to mitigate machine variance present during GC-MS calibration, as this hydrocarbon biomarker is resistant to oil weathering.

Retention time alignment is done to correct the difference in retention time between runs, or retention time drift. The retention time drift is calculated for all source oil signals by finding the differences between prominent peaks, the found retention time drift is then used to shift the weathered oil signals. Correctly aligned source oil signals for Ions 220 and 244 are shown in Fig. 1 and Fig. 2.

C. Peak Identification and Validation

After calibration, peak identification is first performed through empirically selecting a minimum prominence value to filter out relatively small, noisy peaks. Peak validation is performed through a coelution check, where if two or more peaks overlap the data with the largest peak is used in the remaining calculations. Valid peaks are considered the output of the coelution check.

The signal-to-noise ratio (SNR) in decibel for each peak is calculated by finding the retention times at 1/3 of a peak's height, say t_1 and t_2 . The peak area, A_p is then calculated by:

$$A_{pk} = \sum_{t=t_1}^{t_2} |S(t)|^2 \quad (1)$$

where S denotes the signal's height. Peak power, P_p , is calculated using the peak area and 1/3 width as shown in:

$$P_{pk} = A_s / (t_2 - t_1) \quad (2)$$

Similar analysis is performed to calculate the noise power, P_n . Lastly, the SNR is calculated by:

$$SNR_{pk} = 10 \log_{10}(P_{pk}/P_n) \quad (3)$$

D. Peak Grouping

Peak grouping is performed through looking at individual weathered peaks for each sample date, if the retention time of the weathered peak is similar to that of a single source oil peak these weathered and source oil peaks are considered grouped. If the retention time of the weathered peak is similar to that of two source oil peaks, the weathered peak is grouped with the source oil peak closest in retention time. Lastly, if the weathered peak is similar to that of three or more source oil peaks, the weathered peak is grouped with the source oil peak with the highest SNR. If a weathered peak does not find a source oil peak it is labeled as an 'unpaired' peak, which is iteratively cycled through after the remaining weathered peaks have been grouped.

E. Weathering Visualizations

The weathered sample date order is determined based off degradation profile rather than chronological sample harvest date as data was selected based off varying degrees of weathering to show the robust nature of the peak-cognizant weathering quantification algorithm. As such, the chronological ordering has been removed since the weathering is dependant on the circumstances at the point of collection of the sample (e.g. quantity of oil, covered or uncovered, aerobic or anaerobic, length of time as the collection location).

After the weathered sample date order is determined, two visualizations are created:

1. The grouped peaks height ordered from source oil to most weathered signals. The grouped peaks are enumerated as their average retention time with a unique color for each group. These will be referred to as 'degradation profile plots'.
2. As a graph representing the weathering profile, originally discussed in [17]. This graph-based visualization shows the individual peak heights, peak retention time, peak sample, and ratio of the decline in the peak height. These will be referred to as 'persistent weathering profile graphs'. A sample outlined in red denotes the peak height was less than 5% of the overall maximum peak height - indicating the peak might be machine noise. If a sample is missing in the persistent plot this means the peak was either not validated, there was no peak grouped to that retention time, or the peak degraded.

An example and key for interpreting the two compounds in the persistent plots is given in Fig. 3. The peak height and retention time is denoted by p and τ , respectively across n samples. Sample dates are encoded by d , and the relative ratio of weathering is denoted by r . Peaks are grouped by their

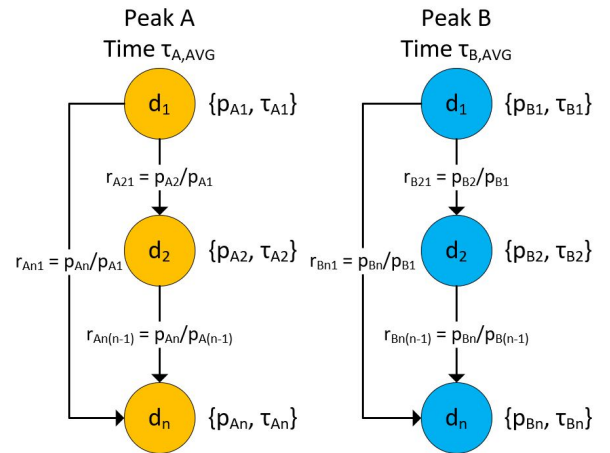


Fig. 3. Graph framework for persistent plots, recreated with permission from [17]. Peaks are grouped by average retention time τ_{AVG} , assigned a unique color. Peak height, retention time, and sample date can be easily read via p , τ , d . The relative rate of weathering between samples is denoted by r , or the ratio of the latter peak height to the former.

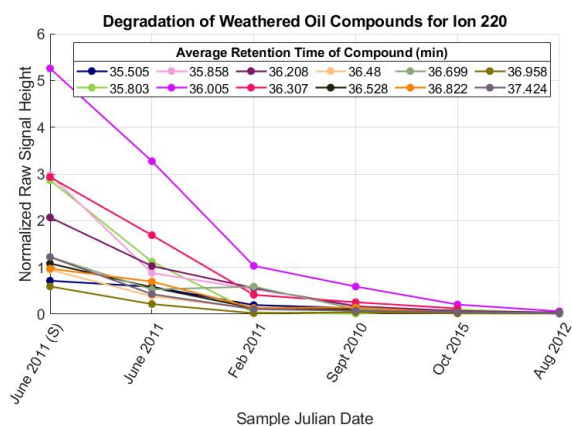


Fig. 4. Weathering degradation profile for C-3 Phenanthrene, Ion 220. Peak groups are specified via unique colors corresponding to the average retention time of the group. The plot shows the C30 normalized signal height against the sample harvest date.

average retention time, τ_{AVG} , which also correspond to the average retention time used in the degradation plots.

IV. RESULTS

Results are separated in two sections, one for each of type of visualization described in subsection III-E.

A. Degradation Profile Plots

Fig. 4 and Fig. 5 show the grouped C30 normalized peak heights ordered from the source oil to the most weathered sample. Each peak group is specified by a unique color, corresponding to the average retention time of the respective group.

An expected degradation profile is seen in Ion 220 (Fig. 4), where generally all signal heights decrease from the source to the final peak. Fig. 5 illustrates the power of these degradation profiles, an increase in normalized signal height occurs in the June 2011 sample. This occurs in the compound at retention time 40.823 minutes (grouped as yellow). This increase would inform a chemist to further examine the compound at this retention time - as it may be machine noise, a contaminant accidentally added in the GC-MS, or - in the extreme case - an unexpected and unexplained increase in a harmful toxin requiring further research and characterization.

B. Persistent Weathering Profile Graphs

The persistent weathering profile graphs for C-3 Phenanthrene, Ion 220 are shown in Fig. 6 and Fig. 7. These graphs span multiple insets as the resolution would be too small to fit onto a single panel. There are three persistent graphs for Ion 244, for the same reason.

Fig. 6 shows the persistent peak profile of C-3 Phenanthrene for average retention times of 35.51 min to 36.31 min. All but one relative ratio decrease across all sample dates and average retention times. A relative ratio of 12.139 is seen between sample dates 2 and 6 at an average retention time of 35.8 min. However, both the sample dates are enclosed in a red circle

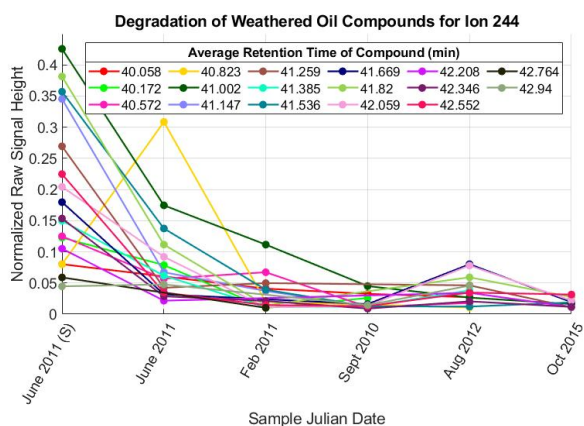


Fig. 5. Weathering degradation profile for C-3 Pyrene, Ion 244. Peak groups are specified via unique colors corresponding to the average retention time of the group. The plot shows the C30 normalized signal height against the sample harvest date. There is an abnormal increase in normalized signal height for retention time 40.823 at sample date June 2011 when compared to the remaining compounds.

meaning this increase could be due to random fluctuations of the signal. This is similar to the relative ratios of greater than one in Fig. 7, where increases in peak height at average retention times of 36.82 min and 36.96 min occur between samples enclosed in red. Further examination should be given to average retention time 36.7 min, between sample dates 4 and 3 - as this relative ratio shows an increase in normalized signal abundance.

Of the 17 average retention times shown in Fig. 8-10 the most prominent relative ratio greater than one is seen in Fig. 8 at average retention time 40.82 min between the source sample and September 2010 sample. Neither normalized abundance height was less than the 5% maximum height, suggesting there was an increase in this compound. The remaining relative ratios greater than one are between either one or both sample dates enclosed in red and may be due to fluctuations within the raw signals.

V. CONCLUDING REMARKS

We introduced preliminary work on autonomous generation of degradation profiles for weathered polycyclic aromatic hydrocarbons from raw gas chromatography-mass spectrometry (GC-MS) signals. The algorithm includes preprocessing methods such as baseline correction, C30 hopane biomarker normalization, and retention time alignment between persistent peaks in the source oil. Peak validation is performed to identify coeluting peaks. An iterative approach is taken to create associations between weathered and source peaks, where peaks in a weathered sample are matched to the closest source oil peak. Preliminary results on degradation profiles and related graph-based visualization of peak persistence were presented for C-3 Phenanthrene and C-3 Pyrene.

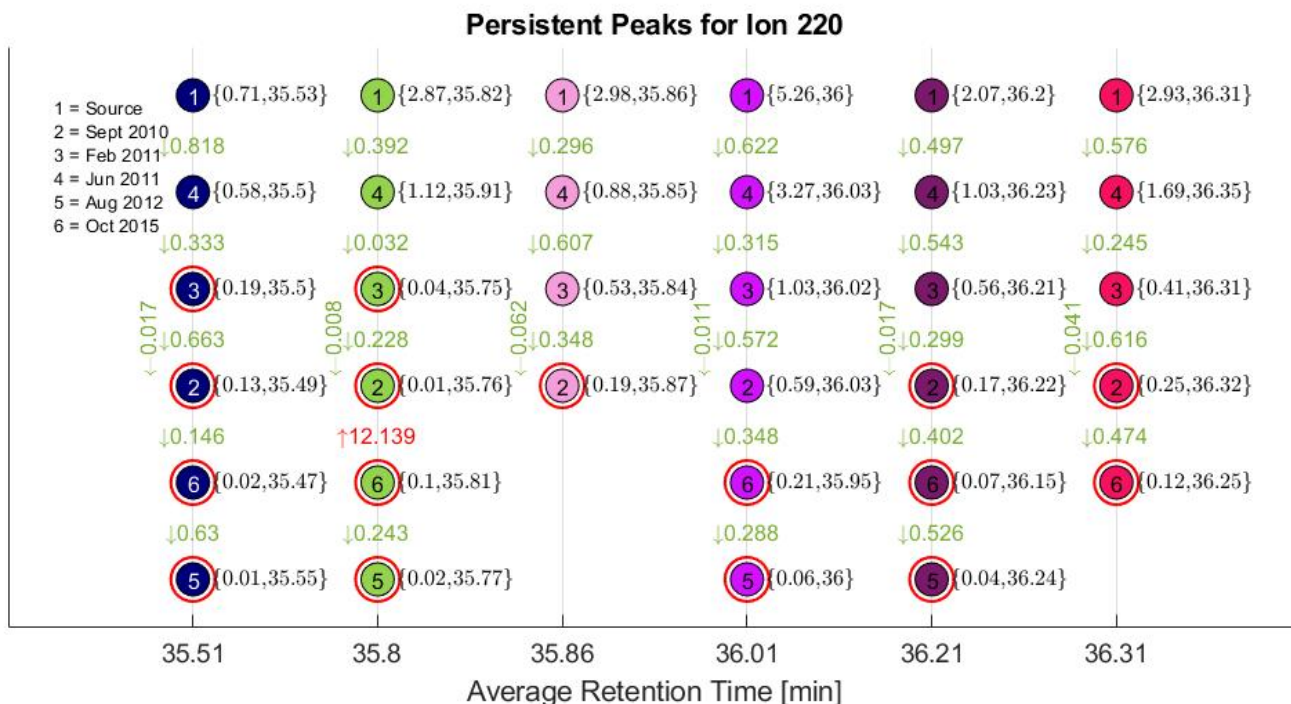


Fig. 6. Persistent peak profile of C-3 Phenanthrene, Ion 220, for average retention times of 35.51 minutes to 36.31 minutes. Refer to Fig. 3 for graph interpretation. Peaks are grouped by retention time with the same coloring as in the degradation plot for quick interpretation. In all cases except retention time 35.8 the compounds are shown to decrease. At retention time 35.8, between samples 2 and 6 a relative ratio of 12.139 is seen. However, both samples contain a red circle - indicating the increase could be noise as the peaks are less than 5% of the maximum peak height.

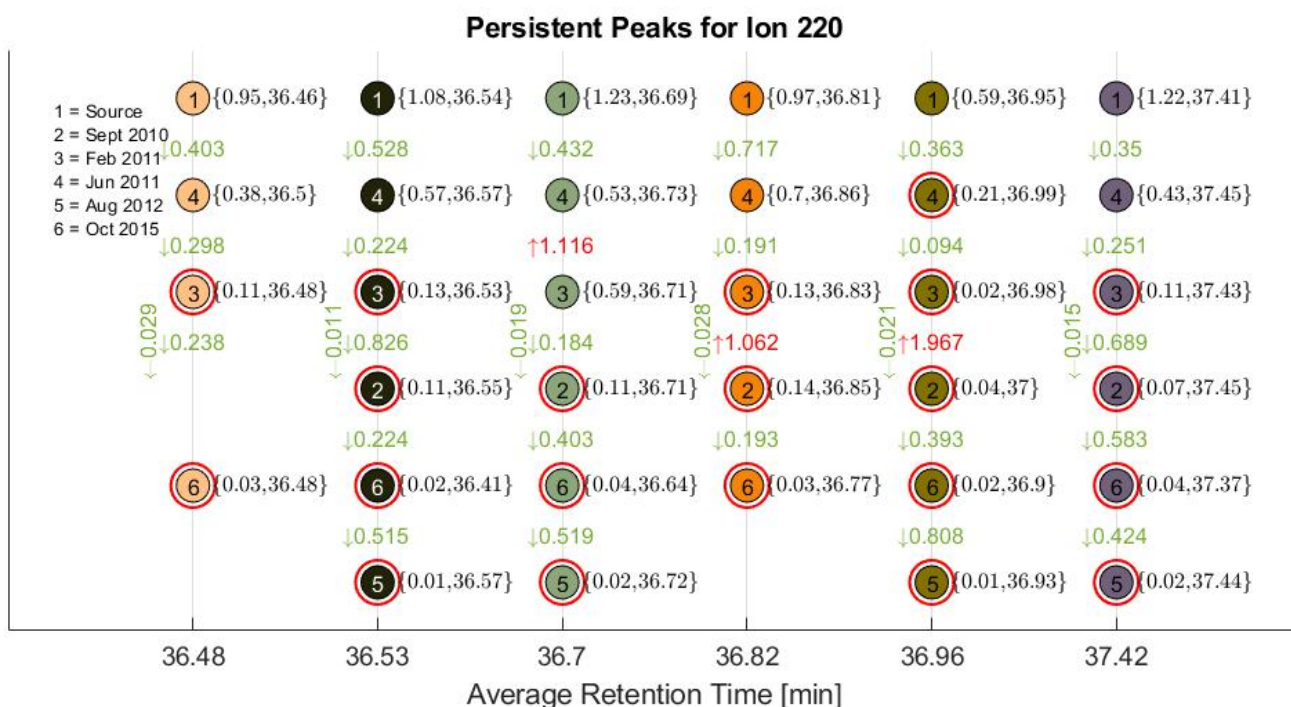


Fig. 7. Persistent peak profile for the remaining peaks in the C-3 Phenanthrene, Ion 220, sample. Refer to Fig. 3 for graph interpretation. Peaks are grouped by retention time with the same coloring as in the degradation plot (Fig. 4) for quick interpretation. An increase of relative ratio 1.116 is seen at average retention time 36.7 between sample 4 and 3. Neither sample is enclosed with a red circle - meaning both peaks are not considered as noise. From this analysis, the development of this compound should further be examined to determine where this increase originates. A relative peak ratio of greater than 1 is seen between samples 3 and 2 of average retention time 36.82 min and 36.96 min - however these samples are enclosed with a red circle meaning they may be considered random fluctuations.

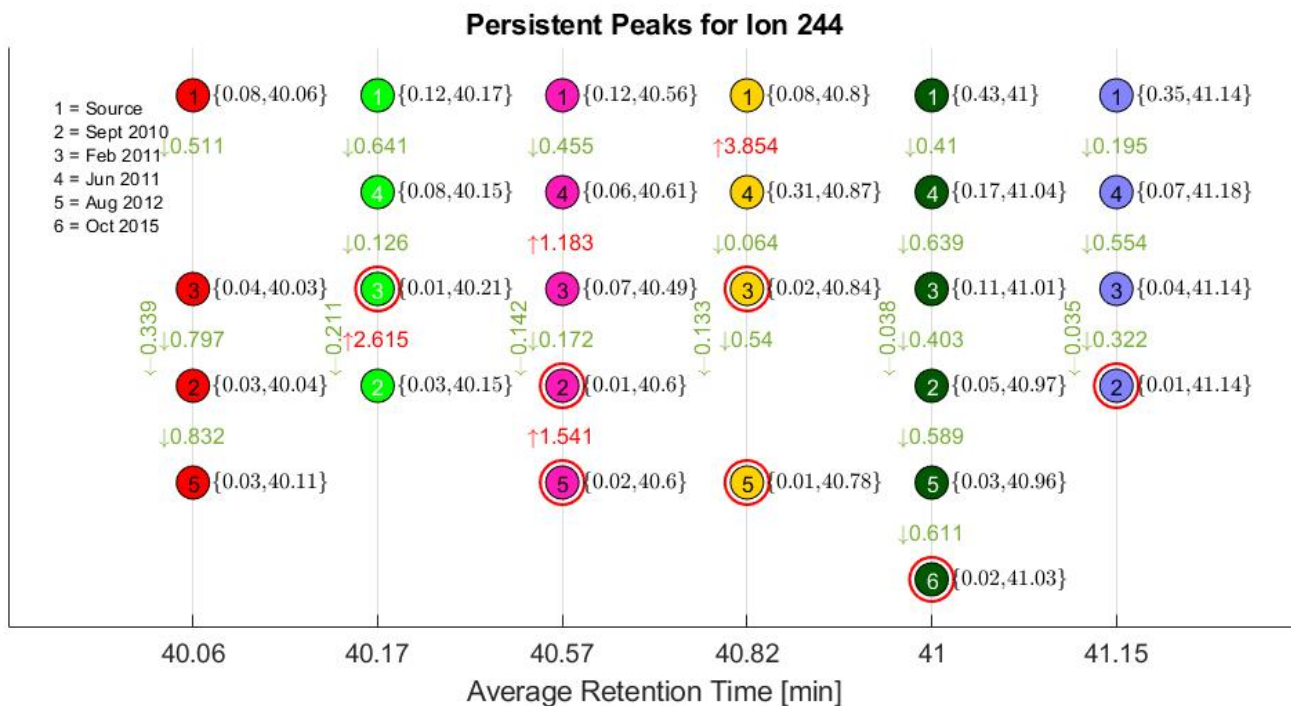


Fig. 8. Persistent peak profile of C-3 Pyrene, Ion 244, for average retention times of 40.06 minutes to 41.15 minutes. Refer to Fig. 3 for graph interpretation. Peaks are grouped by retention time with the same coloring as in the degradation plot for quick interpretation. A relative increase of 3.854 is seen at retention time 40.82 min between sample dates 1 and 4. Neither sample is enclosed in red indicating this is not random fluctuations in the signal and should be further examined. A small increase in peak height is seen at average retention time 40.57 min between sample dates 4 and 3. Again, neither sample is enclosed in red and this phenomenon should be further investigated. A relative ratio of 2.615 is seen at average retention time 40.17 min between sample dates 3 and 2. Sample date 3 (February 2011) is enclosed in red while sample date 2 (September 2010) is not - indicating sample date 3 was less than 5% of the maximum peak height. The remaining sample dates (5 and 6) are not present - indicating the peak degraded out. Therefore, the increase in sample height between sample dates 3 and 2 should not be given as much priority as the previously listed relative ratios greater than one as it is most likely random fluctuations within the signal. The remaining relative ratio greater than one occurs at retention time 40.57 min between sample dates 2 and 5, both sample dates enclosed in red circles - indicating this increase may be random fluctuations within the signal.

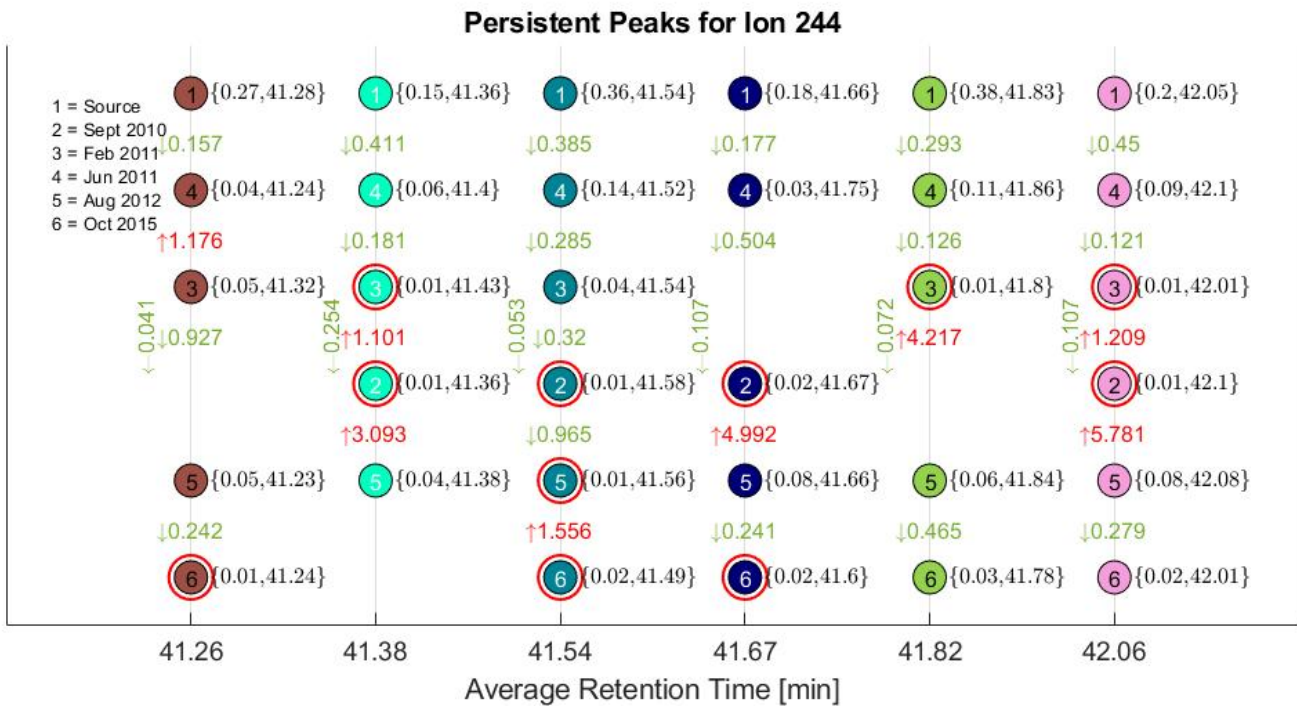


Fig. 9. Persistent peak profile of C-3 Pyrene, Ion 244, for average retention times of 41.26 to 42.06 min. Refer to Fig. 3 for graph interpretation. Of the relative ratios greater than one, the most prominent occurs at average retention time 41.26 min between sample dates 4 and 3 as neither are enclosed in red circles. The relative ratios seen at the remaining average retention times occur between sample dates with either one or both sample dates enclosed in red circles.

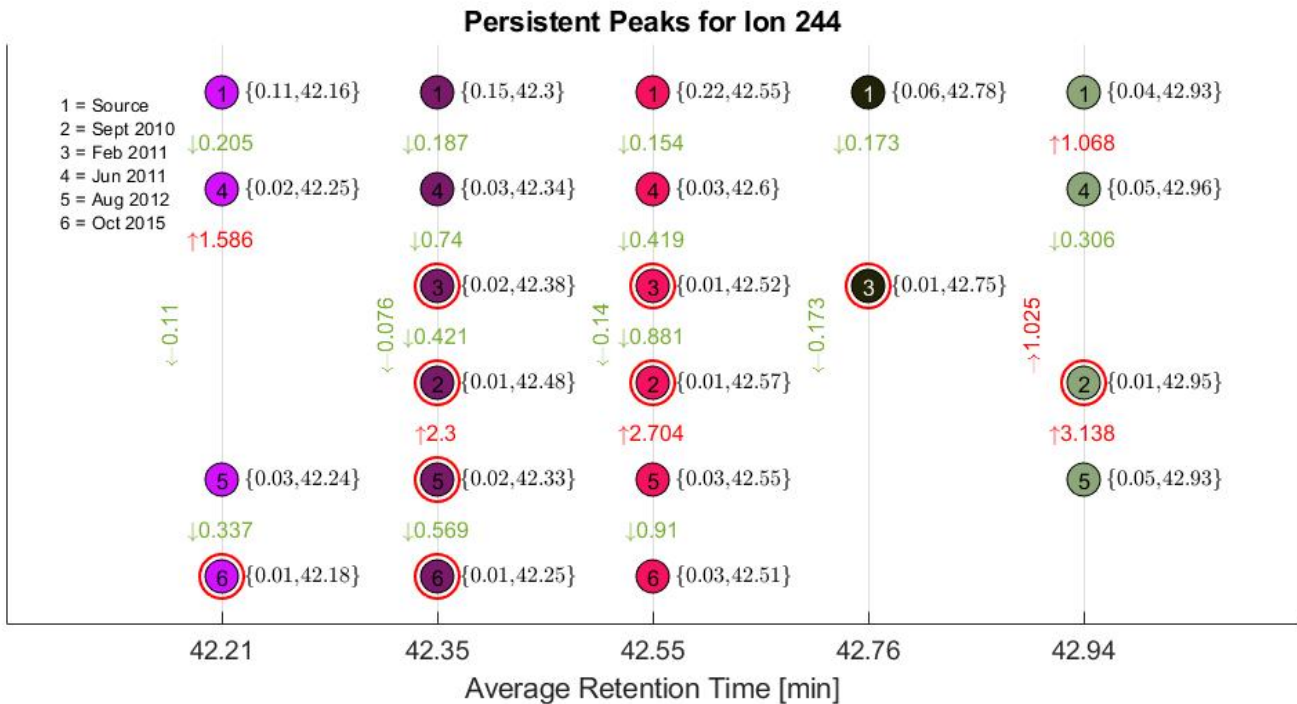


Fig. 10. Persistent peak profile of C-3 Pyrene, Ion 244, for average retention times of 42.21 to 42.94 min. Refer to Fig. 3 for graph interpretation. The relative ratio of 1.586 at average retention time 42.21 min may be considered fluctuations of the signal as sample dates 3 and 2 are not shown, indicated they degraded. The samples in average retention time 42.94 min should be furthered examined as there is an increase of 1.025 from the first sample to the last sample.

REFERENCES

- [1] U. S. C. Guard and N. R. T. (US), *On Scene Coordinator Report: Deepwater Horizon Oil Spill*. US Department of Homeland Security, US Coast Guard, 2011.
- [2] J. R. Radović, C. Aeppli, R. K. Nelson, N. Jimenez, C. M. Reddy, J. M. Bayona, and J. Albaigés, “Assessment of photochemical processes in marine oil spill fingerprinting,” *Marine Pollution Bulletin*, vol. 79, no. 1-2, pp. 268–277, 2014.
- [3] M. A. Tarr, P. Zito, E. B. Overton, G. M. Olson, P. L. Adhikari, and C. M. Reddy, “Weathering of oil spilled in the marine environment,” *Oceanography*, vol. 29, no. 3, pp. 126–135, 2016.
- [4] I. A. Mendelssohn, G. L. Andersen, D. M. Baltz, R. H. Caffey, K. R. Carman, J. W. Fleeger, S. B. Joye, Q. Lin, E. Maltby, E. B. Overton *et al.*, “Oil impacts on coastal wetlands: implications for the mississippi river delta ecosystem after the deepwater horizon oil spill,” *BioScience*, vol. 62, no. 6, pp. 562–574, 2012.
- [5] R. E. Turner, E. B. Overton, B. M. Meyer, M. S. Miles, G. McClenachan, L. Hooper-Bui, A. S. Engel, E. M. Swenson, J. M. Lee, C. S. Milan *et al.*, “Distribution and recovery trajectory of macondo (mississippi canyon 252) oil in louisiana coastal wetlands,” *Marine Pollution Bulletin*, vol. 87, no. 1-2, pp. 57–67, 2014.
- [6] A. J. Esbaugh, E. M. Mager, J. D. Stieglitz, R. Hoenig, T. L. Brown, B. L. French, T. L. Linbo, C. Lay, H. Forth, N. L. Scholz *et al.*, “The effects of weathering and chemical dispersion on deepwater horizon crude oil toxicity to mahi-mahi (*coryphaena hippurus*) early life stages,” *Science of the Total Environment*, vol. 543, pp. 644–651, 2016.
- [7] A. A. Perez-Umphrey, C. M. B. Burns, P. C. Stouffer, S. Woltmann, and S. S. Taylor, “Polycyclic aromatic hydrocarbon exposure in seaside sparrows (*ammodramus maritimus*) following the 2010 deepwater horizon oil spill,” *Science of the total environment*, vol. 630, pp. 1086–1094, 2018.
- [8] F. Yin, G. F. John, J. S. Hayworth, and T. P. Clement, “Long-term monitoring data to describe the fate of polycyclic aromatic hydrocarbons in deepwater horizon oil submerged off alabama’s beaches,” *Science of the Total Environment*, vol. 508, pp. 46–56, 2015.
- [9] G. F. John, Y. Han, and T. P. Clement, “Weathering patterns of polycyclic aromatic hydrocarbons contained in submerged deepwater horizon oil spill residues when re-exposed to sunlight,” *Science of the Total Environment*, vol. 573, pp. 189–202, 2016.
- [10] J. Duan, W. Liu, X. Zhao, Y. Han, S. O’Reilly, and D. Zhao, “Study of residual oil in bay jimmy sediment 5 years after the deepwater horizon oil spill: persistence of sediment retained oil hydrocarbons and effect of dispersants on desorption,” *Science of the Total Environment*, vol. 618, pp. 1244–1253, 2018.
- [11] Q. Lin, I. A. Mendelssohn, S. A. Graham, A. Hou, J. W. Fleeger, and D. R. Deis, “Response of salt marshes to oiling from the deepwater horizon spill: Implications for plant growth, soil surface-erosion, and shoreline stability,” *Science of the Total Environment*, vol. 557, pp. 369–377, 2016.
- [12] A. K. Huba and P. R. Gardinali, “Characterization of a crude oil weathering series by ultrahigh-resolution mass spectrometry using multiple ionization modes,” *Science of the Total Environment*, vol. 563, pp. 600–610, 2016.
- [13] N.-P. V. Nielsen, J. M. Carstensen, and J. Smedsgaard, “Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping,” *Journal of chromatography A*, vol. 805, no. 1-2, pp. 17–35, 1998.
- [14] D. J. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series.” in *KDD workshop*, vol. 10, no. 16. Seattle, WA, USA., 1994, pp. 359–370.
- [15] J. H. Christensen, G. Tomasi, and A. B. Hansen, “Chemical fingerprinting of petroleum biomarkers using time warping and pca,” *Environmental science & technology*, vol. 39, no. 1, pp. 255–260, 2005.
- [16] P. Sun, K. Bao, H. Li, F. Li, X. Wang, L. Cao, G. Li, Q. Zhou, H. Tang, and M. Bao, “An efficient classification method for fuel and crude oil types based on m/z 256 mass chromatography by cow-pca-lda,” *Fuel*, vol. 222, pp. 416–423, 2018.
- [17] A. S. Gupta, B. Meyer, and E. Overton, “Quantifying weathering profiles of environmental contaminants from marine and coastal oil spills using signal processing techniques,” in *OCEANS 2018 MTS/IEEE Charleston*. IEEE, 2018, pp. 1–4.