Spring 2018

# Fingerprinting the Refugio Oil Spill using Topographic Signal Processing of Two-Dimensional Gas Chromatographic Images

Rachel Bruflodt
*University of Iowa*

FINGERPRINTING THE REFUGIO OIL SPILL USING TOPOGRAPHIC SIGNAL PROCESSING OF TWO-DIMENSIONAL GAS CHROMATOGRAPHIC IMAGES

by

Rachel Bruflodt

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in the Electrical Engineering

_____

Ananya Sen Gupta
Thesis Mentor

Spring 2018

All requirements for graduation with Honors in the
Electrical Engineering have been completed.

_____

Xiaodong Wu
Electrical Engineering Honors Advisor

# Fingerprinting the Refugio Oil Spill using Topographic Signal Processing of Two-Dimensional Gas Chromatographic Images

by

Rachel Bruflodt

May 2017

Mentor: Ananya Sen Gupta

## Abstract

Oil spill fingerprinting aims to determine whether oil samples originate from the source of the spill or from a different source. This can be done by comparing the composition of hydrocarbons in each oil sample's two-dimensional gas chromatography (GC×GC) image. In this work, we compare oil samples taken from the 2015 Refugio Oil Spill with oil samples taken elsewhere. We use an approach called Peak Topography Mapping (PTM) which aims to automatically compare oil samples using both target and non-target peaks in the GC×GC image. Results suggest that the non-target peaks analyzed in PTM can provide information beyond that of solely target-based analysis.

# 1. Introduction and Background

## 1.1. Refugio Oil Spill

On May 19th, 2015, an oil pipeline owned by Plains All American Pipeline ruptured near Refugio State Beach in Santa Barbara, CA [1]. When a spill such as this occurs, it is important to be able to determine whether found oil has come from the spill or from other sources. For this work, samples were taken from the ruptured pipeline as well as other affected areas. These samples were compared to each other and to other oil samples not from the spill.

## 1.2. Two-Dimensional Gas Chromatography for Oil Spills

Each oil sample has a specific hydrocarbon profile. This profile can be seen in the form of a Two-dimensional gas chromatography image (GCxGC), where the peaks of the image correspond to biomarker compounds [3,4]. By comparing their GCxGX images, we can determine if two oil samples come from the same source or different sources.

## 1.3. Current Methods of Comparison

There are two broad approaches for fingerprinting oil samples. One method is target-based analysis [8, 9, 10] where only the peaks corresponding to target compounds are compared. This method can sometimes fail to distinguish between different oil sources because the target peaks are very similar. Target-based analysis also ignores all non-target peaks which may also contain useful information. The other approach for fingerprinting oil spills is statistical pattern recognition [11, 12, 13]. This approach can be sensitive to shifts in retention time and can also require large sets of training data. The statistical pattern recognition approach does not distinguish between target and non-target compounds.

## 2. Methods

Our approach was to use both target and non-target peaks to determine the similarity of GC×GC samples. We did this using the peak topography map concepts outlined in [2, 5, 6] with some modifications.

### 2.1. PTM

A peak topography map (PTM) is a two-dimensional matrix representation of the peaks of a GC x GC image. The first dimension of the PTM is the first dimension retention time index of the GC x GC image, and the second dimension of the PTM contains peaks in order of their elution time along the second dimension of the GC x GC image. The PTM is normalized against the maximum peak. In this case, peaks were not considered in the PTM unless their height was at least 10% of the maximum peak height in the image. This value resulted in the most consistent number of peaks across samples expected to match. As explained in [2], if the *n*th column of a GC x GC image has $k_n$ peaks, and the amplitudes and locations of the peaks are stored as $Peak_n = \{p_{1,n}, p_{2,n}, \dots, p_{k_n,n}\}$ and $Loc_n = \{m_{1,n}, m_{2,n}, \dots, m_{k_n,n}\}$, the $(l,n)^{th}$element PTM representation is described as

$$PTM[l,n] = \begin{cases} p_{l,n} + j \times m_{l,n} & if\ 1 \le l \le k_n \\ 0 & l \le k_n \end{cases}$$

### 2.2. Topography Partitioning

The PTMs of individual oil samples were compared using the topography partitioning described in [2]. This involves comparing peaks at the same locations in the PTM and determining the peak ratio. The peak ratio is defined as $\rho = \max(a, a^{-1})\ where\ a = \frac{p_{ref}}{p_{test}}$. A peak is considered "similar" if the peak ratio is below the peak threshold, and "dissimilar" is the peak ratio is above the peak threshold. For this investigation, 1.2 was used as the peak ratio threshold below

which peaks were considered to match. This number was chosen because it resulted in the largest difference between the match percentages of expected matches and expected non-matches.

### 2.3. Cross-PTM Score

The peak topography mapping technique was applied over 5 different coordinate boxes within each raw sample. Each box corresponds to a different group of compounds: Archeans, Monoaromatic Steranes, Hopanes, or Steranes. The Archean group is comprised of two separate coordinate boxes. The peak topography mapping technique produces a cross-PTM score, which is a percentage match value for each pair of samples. The percentage represents the combined similarity between the target and non-target peaks in the sample. As explained in [2], it is calculated as the weighted percentage of peaks that are "similar" between the pair of samples:

$$S_\tau(I_{test}, I_{ref}) = \frac{\left|\eta_{test} \in PTM_{test,aligned}(PTM_{ref}) : \rho(m,n) \geq 1.2\right|_w}{\eta_{test} \in PTM_{test,aligned}(PTM_{ref})}$$

where $I_{test}$ and $I_{ref}$ are the test and reference samples, $\eta_{test}$ is a node in the test PTM, and $|\cdot|_w$ is the weighted sum taken across all peaks that meet the 1.2 threshold such that larger peaks are weighted higher than smaller peaks.

## 3. Results

Expanding on the results in [7], we provide results based on 15 GC×GC samples, some of which were taken from the area of the Refugio spill.

### 3.1. Matching Samples

Table 1 provides the percentage match of each sample to RS029 in each compound region. RS029 is a sample taken from the source of the Refugio spill. The rows

highlighted green indicate the samples which were expected to match RS029 and the rows highlighted orange indicate the samples which were expected to not match RS029. Table 2 provides the number of peaks found in each sample for each compound region and uses the same color coding scheme. Table 3 provides descriptions for the samples expected to match RS029.

**Table 1** Percent matches to RS029

| Sample | Archean Index (243,440) to (260,540) match | Archean Index (212,610) to (225,645) match | Monoaromatic Steranes (80,25) to (130, 183) match | Hopanes (75,185) to (140,585) match | Steranes (140,90) to (180,346) match |
|---|---|---|---|---|---|
| Sample RS029 | 100% | 100% | 100% | 100% | 100% |
| Sample RS004 | 99.9971% | 100% | 81.3603% | 89.8363% | 87.3532% |
| Sample RS005 | 100% | 95.5480% | 66.7552% | 64.4538% | 83.7939% |
| Sample RS007 | 96.0916% | 92.0283% | 14.5531% | 26.7380% | 30.2062% |
| Sample RS009 | 94.1948% | 95.8331% | 25.2361% | 27.2325% | 33.0696% |
| Sample RS028 | 95.6403% | 96.3337% | 2.1340% | 25.9239% | 31.1829% |
| Sample RS036 | 85.1580% | 91.8798% | 25.1399% | 26.7581% | 31.9805% |
| Sample RS023 | 17.7005% | 41.0260% | 6.4315% | 3.2441% | 23.8260% |
| Sample RS070 | 85.7867% | 45.1885% | 10.4183% | 0.0146% | 24.2280% |
| Sample RS071 | 14.5794% | 71.8709% | 5.5889% | 3.1727% | 19.7485% |
| Sample Seep_Stringer | 59.3328% | 78.5981% | 99.9916% | 99.9860% | 99.7506% |
| Sample Seep_Surface | 5.2606% | 46.8358% | 68.7344% | 83.9838% | 6.7912% |
| Sample SRM1582_A | 5.9012% | 15.5048% | 9.6640% | 15.8478% | 27.2685% |
| Sample SRM1582_B | 13.7645% | 9.8375% | 7.4759% | 18.6454% | 26.1028% |
| Sample SRM1582_C | 5.8993% | 15.6132% | 7.7073% | 18.2553% | 25.3634% |

**Table 2** Number of peaks

| Sample | Archean Index (243,440) to (260,540) number of peaks | Archean Index (212,610) to (225,645) number of peaks | Monoaromatic Steranes (80,25) to (130, 183) number of peaks | Hopanes (75,185) to (140,585) number of peaks | Steranes (140,90) to (180,346) number of peaks |
|---|---|---|---|---|---|
| Sample RS029 | 28 | 34 | 68 | 25 | 256 |
| Sample RS004 | 27 | 34 | 81 | 28 | 224 |
| Sample RS005 | 27 | 36 | 84 | 29 | 214 |
| Sample RS007 | 26 | 35 | 94 | 37 | 229 |
| Sample RS009 | 26 | 36 | 93 | 35 | 229 |
| Sample RS028 | 26 | 36 | 78 | 38 | 266 |
| Sample RS036 | 26 | 37 | 83 | 37 | 215 |
| Sample RS023 | 41 | 36 | 262 | 40 | 368 |
| Sample RS070 | 81 | 32 | 175 | 47 | 281 |
| Sample RS071 | 41 | 35 | 261 | 44 | 317 |
| Sample Seep_Stringer | 37 | 36 | 72 | 27 | 221 |
| Sample Seep_Surface | 16 | 16 | 82 | 30 | 543 |
| Sample SRM1582_A | 37 | 34 | 309 | 48 | 354 |
| Sample SRM1582_B | 37 | 34 | 309 | 45 | 351 |
| Sample SRM1582_C | 38 | 34 | 310 | 46 | 347 |

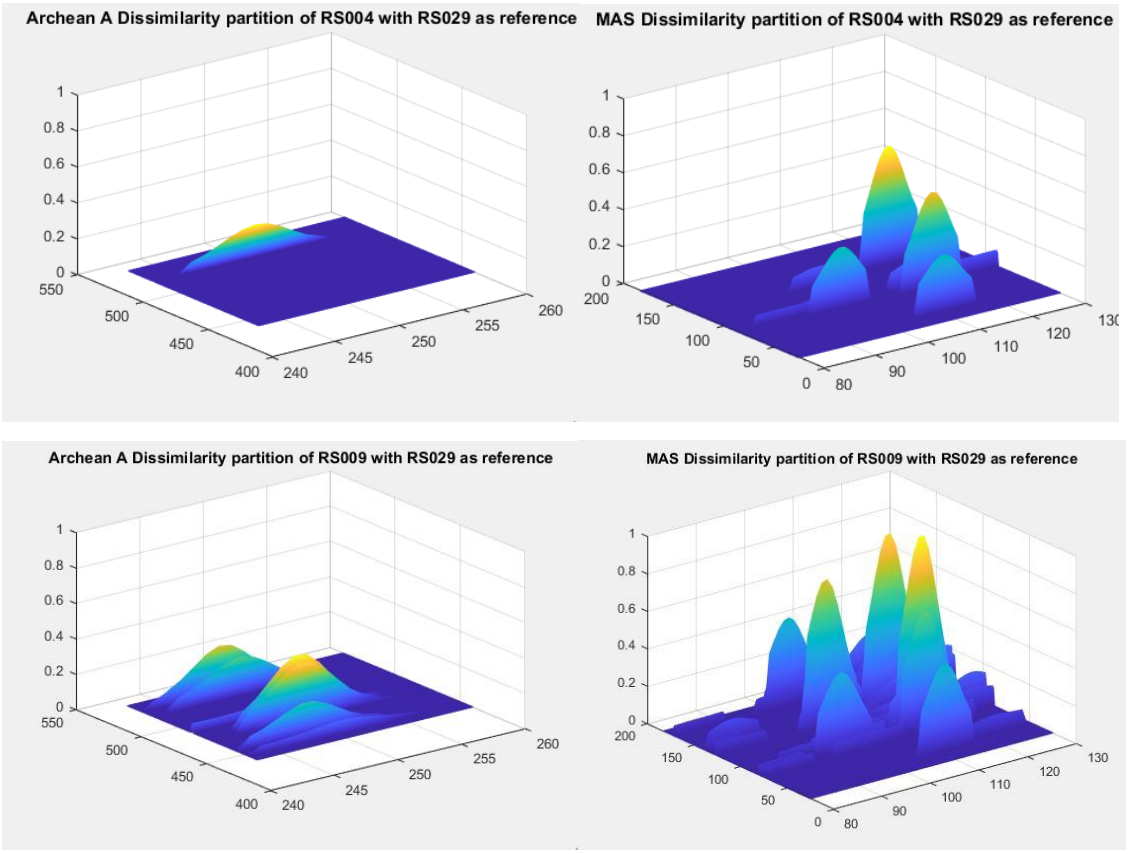
**Table 3** Sample descriptions

| Sample | Description |
|---|---|
| RS004 | Minimal weathering; collected day 0 |
| RS005 | Minimal weathering; collected day 0 |
| RS007 | Some on-water weathering; collected day 1 |
| RS009 | Some on-water weathering; collected day 1 |
| RS028 | Some weathering; collected day 1 |
| RS036 | Some on-water weathering; collected day 2 |

### 3.2. Differences in Match Percentages Across Compound Regions

Table 1 shows that RS029 matches with itself 100% in each compound region as expected. It also shows that the regions which most clearly differentiate between expected matches and non-matches are the two Archean regions. The match percentages for the samples expected to match RS029 are much lower in the Monoaromatic Sterane, Hopane, and Sterane regions. Samples collected after day 0 show the most dramatic drop in match percentages from the Archean regions to the other regions. The two samples collected on day 0 maintain relatively high match percentages in the non-Archean regions. Figure 1 shows the dissimilarity partitions for the samples RS004 and RS009 in the first Archean region and the Monoaromatic Sterane (MAS) region. That is, it shows the peaks in those regions of those two samples that did not match corresponding peaks in RS029.

**Figure 1** Dissimilarity partitions




The figure shows that the major nonmatching peaks in the Archean region are fewer and generally smaller than the major nonmatching peaks in the non-Archean Monoaromatic Sterane region. Similar results are seen across all samples in all Archean and non-Archean regions. This accounts for the strong match percentages between RS029 and expected matches to RS029 in the Archean regions. This also accounts for the less strong match percentages in non-Archean regions.

## 4. Discussion

### 4.1. SRM1582 Discrepancies

The three samples SRM1582_A, SRM1582_B, and SRM1882_C represent the same sample which was processed three different times to produce three different GC×GC images. The discrepancies in match percentages for these three samples in Table 1 are explained by small differences in peak values in the raw images. The three images have enough small differences that a few peaks in one SRM image may match corresponding peaks in the RS029 image, while the same peaks in the other SRM samples do not. This is consistent with the fact that the largest match percentage discrepancies occur in the compound regions with the fewest number of peaks, meaning that an occasional inconsistently matched peak has a larger effect on the overall match percentage. When the peak topography technique is applied to compare the SRM samples to each other, they all match each other with a match percentage of 100%.

### 4.2. Effects of Weathering

Figure 1 may show some potential effects of weathering. The sample with minimal weathering collected on day 0 (RS004) has a few, mostly small peaks in its dissimilarity partition for the MAS region. The sample collected on day 1 with some on-water weathering (RS009) on the other hand, has many small peaks in its dissimilarity partition with some larger ones. Many small peaks are characteristic of the dissimilarity partitions for weathered samples in the Monoarmatic Sterane, Hopane, and Sterane compound regions. Dissimilarity partitions for weathered samples in the Archean regions contain less peaks, and have smaller peaks overall. Dissimilarity partitions in the Archean regions therefore show less of these effects possibly influenced by weathering. More varied samples are needed to further understand the effect of weathering in each compound region.

## 5. Conclusions

The cross-PTM scores for the Archean compound region were able to differentiate oil samples that had come from the Refugio spill from oil samples that had not. The distinction was less clear when comparing cross-PTM scores for the Hopane, Sterane, and Monoaromatic Sterane regions.

## References


[1] "Refugio Oil Spill, Santa Barbara." *Oceana USA*, 25 Aug. 2016, http://usa.oceana.org/climate-and-energy/refugio-oil-spill-santa-barbara.

[2] Ghasemi Damavandi H., et al (2016), "Interpreting comprehensive two-dimensional gas chromatography using peak topography maps with application to petroleum forensics", Chem. Centr. Jnl.10 (1), 75.

[3] Ligon WV, et al (1984) Target compound analysis by two-dimensional gas chromatography-mass spectrometry, J Chromatogr A 294:77–86.

[4] Stout SA, et al (2008) Diagnostic compounds for fingerprinting petroleum in the environment. Environ. Forensics 26:54.

[5] A. S. Gupta, C. M. Reddy, and R. Nelson, "Systems and methods for topographic analysis," Sep. 16 2014, US Patent 8,838,393.

[6] H. G. Damavandi, A. S. Gupta, C. Reddy, and R. Nelson, "Oil-spill forensics using two-dimensional gas chromatography: Differentiating highly correlated petroleum sources using peak manifold clusters," in Signals, Systems and Computers, 2015 49th Asilomar Conference on. IEEE, 2015, pp. 1589–1592

[7] Bruflodt R, et al "Fingerprinting the Refugio oil spill using topographic signal processing of two-dimensional gas chromatographic images," OCEANS–Anchorage, 2017

[8] Standard practice for oil spill source identification by gas chromatography and positive ion electron impact low resolution mass spectrometry. ASTM D5739-06, developed by ASTM International. doi:10.1520/D5739-06

[9] EPA8270D semivolatile organic compounds by gas chromatography/mass spectrometry (GC/MS). developed by United States Environmental Protection Agency, http://www.epa.gov/osw/hazard/testmethods/sw846/pdfs/8270d.pdf

[10] Modified EPA8270. Developed by United States Environmental Protection Agency

[11] Ventura GT, Hall GJ, Nelson RK, Frysinger GS, Raghuraman B, Pomerantz AE, Mullins OC, Reddy CM (2011) Analysis of petroleum compositional similarity using multiway principal components analysis (MPCA) with comprehensive two-dimensional gas chromatographic data. J Chromatogr A 1218(18):2584–2592

[12] Ebrahimi D, Li J, Hibbert DB (2007) Classification of weathered petroleum oils by multi-way analysis of gas chromatography-mass spectrometry data using parafac2 parallel factor analysis. J Chromatogr A 1166(1):163–170

[13] Christensen JH, Tomasi G, Hansen AB (2005) Chemical fingerprinting of petroleum biomarkers using time warping and PCA. Environ Sci Technol 39(1):255–260