Use of Biomarker Ratio Database and Search Tool to Quickly Identify Similar Oil Samples

Laird Henkel and Marida Martin

California Department of Fish and Game
Office of Spill Prevention and Response
Outline

Why is oil fingerprinting important?
Oil fingerprinting/biomarker ratios
Goal of this project (database search tool)
MATLAB search tool
Results
Next steps
Why is Oil Fingerprinting Important?

- Comparison of suspect samples to known spill source
- Mystery spills
  - Natural petroleum seeps
  - Anthropogenic
    - Acute (bilge cleaning, sudden leaks/spills)
    - Chronic (shipwrecks, slow leaks)
Why is Oil Fingerprinting Important?

Natural Petroleum Seeps

- Worldwide, ~180 million gallons (4.3 million bbls) into marine environment per year
- In Santa Barbara Channel, ~6 million gallons (143,000 bbls) released per year

Source: Kvenolden and Cooper 2003
Why is Oil Fingerprinting Important?

Natural Petroleum Seeps

- Oiled Wildlife Care Network (OWCN) intakes an average of about 275 miscellaneous oiled birds per year.
Why is Oil Fingerprinting Important?

Shipwrecks

- S.S. *Jacob Luckenbach* estimated to have killed >50,000 birds
- S.S. *Montebello* poses potential risk
Petroleum Fingerprinting

Basics

- Gas Chromatography/Mass Spectrometry (GCMS)
- Comparison of chromatograms
Petroleum Fingerprinting

Biomarker Ratios

- Biomarker = organic compounds from dead things
- Ratios provide way of quantifying comparisons
Petroleum Fingerprinting

Biomarker Ratios

- Biomarker = organic compounds from dead things
- Ratios provide way of quantifying comparisons
Petroleum Fingerprinting

Biomarker Ratios

- Biomarkers can be affected by:
  - Biodegradation
  - Weathering

- USGS identified 19 ratios that are relatively stable

<table>
<thead>
<tr>
<th>Index</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>del 13C</td>
</tr>
<tr>
<td>2</td>
<td>Ts/Tm</td>
</tr>
<tr>
<td>3</td>
<td>C_{26}/Tet (triplet)</td>
</tr>
<tr>
<td>4</td>
<td>C_{28}/C_{29}</td>
</tr>
<tr>
<td>5</td>
<td>PAH-RI</td>
</tr>
<tr>
<td>6</td>
<td>C2D/C2P</td>
</tr>
<tr>
<td>7</td>
<td>C3D/C3P</td>
</tr>
<tr>
<td>8</td>
<td>C_{28}/C_{29} TT</td>
</tr>
<tr>
<td>9</td>
<td>C_{20}/C_{23} TT</td>
</tr>
<tr>
<td>10</td>
<td>C_{22}/C_{21} TT</td>
</tr>
<tr>
<td>11</td>
<td>C_{24}/C_{23} TT</td>
</tr>
<tr>
<td>12</td>
<td>C_{26}/C_{25} TT</td>
</tr>
<tr>
<td>13</td>
<td>C_{31} S/H</td>
</tr>
<tr>
<td>14</td>
<td>C_{29} H/H</td>
</tr>
<tr>
<td>15</td>
<td>C_{35}/C_{34} S</td>
</tr>
<tr>
<td>16</td>
<td>BNH/H</td>
</tr>
<tr>
<td>17</td>
<td>OI/H</td>
</tr>
<tr>
<td>18</td>
<td>G/H</td>
</tr>
<tr>
<td>19</td>
<td>C_{29} Ts/C_{29} H</td>
</tr>
</tbody>
</table>

USGS identified 19 ratios that are relatively stable

Lorenson et al. 2009
Petroleum Fingerprinting

Biomarker Ratios

- USGS and others have used ratios to model similarity of samples, using PCA, etc.
Project Goals

Ratio Comparison Tool:

- Able to quickly compare a mystery sample to a large number of other known samples
- Be simple to use and to interpret
- Is not dependent on the number or variability of other samples in the database
Methods

- Developed two simple MATLAB routines with different algorithms
- Tested validity of results with standard visual comparisons of chromatograms
## Methods

### Method 1: Mean Percentage Difference (MPD)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-23.4</td>
<td>0.59</td>
<td>6.2</td>
<td>1.10</td>
<td>10</td>
<td>0.02</td>
<td>0.03</td>
<td>0.86</td>
<td>0.17</td>
<td>0.23</td>
<td>0.63</td>
<td>0.88</td>
<td>0.32</td>
<td>0.59</td>
<td>0.83</td>
<td>0.30</td>
<td>0.20</td>
<td>0.11</td>
<td>0.35</td>
</tr>
<tr>
<td>B</td>
<td>-22.8</td>
<td>0.28</td>
<td>4.4</td>
<td>1.10</td>
<td>44</td>
<td>0.95</td>
<td>0.06</td>
<td>0.70</td>
<td>0.14</td>
<td>0.56</td>
<td>0.40</td>
<td>0.76</td>
<td>0.56</td>
<td>0.77</td>
<td>1.60</td>
<td>0.64</td>
<td>0.05</td>
<td>0.13</td>
<td>0.19</td>
</tr>
</tbody>
</table>
# Methods

## Method 1: Mean Percentage Difference (MPD)

\[
\text{MPD} = \frac{\text{abs}(0.03 - 0.06)}{\text{avg}(0.03, 0.06)} = 0.67
\]

<table>
<thead>
<tr>
<th>A</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>-23.4</td>
<td>0.59</td>
<td>6.2</td>
<td>1.10</td>
<td>10</td>
<td>0.02</td>
<td>0.03</td>
<td>0.86</td>
<td>0.17</td>
<td>0.23</td>
<td>0.63</td>
<td>0.88</td>
<td>0.32</td>
<td>0.59</td>
<td>0.83</td>
<td>0.30</td>
<td>0.20</td>
<td>0.11</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>-22.8</td>
<td>0.28</td>
<td>4.4</td>
<td>1.10</td>
<td>44</td>
<td>0.95</td>
<td>0.06</td>
<td>0.70</td>
<td>0.14</td>
<td>0.56</td>
<td>0.40</td>
<td>0.76</td>
<td>0.56</td>
<td>0.77</td>
<td>1.60</td>
<td>0.64</td>
<td>0.05</td>
<td>0.13</td>
<td>0.19</td>
</tr>
</tbody>
</table>
**Methods**

**Method 1: Mean Percentage Difference (MPD)**

\[
MPD = \frac{\text{abs}(0.03 - 0.06)}{\text{avg}(0.03, 0.06)}
\]

\[
= 0.67
\]

\[
MPD = \frac{\text{abs}(0.86 - 0.70)}{\text{avg}(0.86, 0.70)}
\]

\[
= 0.21
\]
## Methods

### Method 1: Mean Percentage Difference (MPD)

MPD = \frac{\text{abs}(0.03-0.06)}{\text{avg}(0.03, 0.06)}

= 0.67

MPD = \frac{\text{abs}(0.86-0.70)}{\text{avg}(0.86, 0.70)}

= 0.21

Total MPD (mean of 19 MPD individual values) = 0.57
Methods

Method 2: Standardized Slope

|     | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   |
|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| A   | -23.4| 0.59 | 6.2  | 1.10 | 10   | 0.02 | 0.03 | 0.86 | 0.17 | 0.23 | 0.63 | 0.88 | 0.32 | 0.59 | 0.83 | 0.30 | 0.20 | 0.11 | 0.35 |
| B   | -22.8| 0.28 | 4.4  | 1.10 | 44   | 0.95 | 0.06 | 0.70 | 0.14 | 0.56 | 0.40 | 0.76 | 0.56 | 0.77 | 1.60 | 0.64 | 0.05 | 0.13 | 0.19 |

- Convert ratio to slope
- Standardize distance between peaks to height of second peak
Methods

Method 2: Standardized Slope

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-23.4</td>
<td>0.59</td>
<td>6.2</td>
<td>1.10</td>
<td>10</td>
<td>0.02</td>
<td>0.03</td>
<td>0.86</td>
<td>0.17</td>
<td>0.23</td>
<td>0.63</td>
<td>0.88</td>
<td>0.32</td>
<td>0.59</td>
<td>0.83</td>
<td>0.30</td>
<td>0.20</td>
<td>0.11</td>
<td>0.35</td>
</tr>
<tr>
<td>B</td>
<td>-22.8</td>
<td>0.28</td>
<td>4.4</td>
<td>1.10</td>
<td>44</td>
<td>0.95</td>
<td>0.06</td>
<td>0.70</td>
<td>0.14</td>
<td>0.56</td>
<td>0.40</td>
<td>0.76</td>
<td>0.56</td>
<td>0.77</td>
<td>1.60</td>
<td>0.64</td>
<td>0.05</td>
<td>0.13</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Ratio 0.03 = slope 44.1°
## Methods

### Method 2: Standardized Slope

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-23.4</td>
<td>0.59</td>
<td>6.2</td>
<td>1.10</td>
<td>10</td>
<td>0.02</td>
<td>0.03</td>
<td>0.86</td>
<td>0.17</td>
<td>0.23</td>
<td>0.63</td>
<td>0.88</td>
<td>0.32</td>
<td>0.59</td>
<td>0.83</td>
<td>0.30</td>
<td>0.20</td>
<td>0.11</td>
<td>0.35</td>
</tr>
<tr>
<td>B</td>
<td>-22.8</td>
<td>0.28</td>
<td>4.4</td>
<td>1.10</td>
<td>44</td>
<td>0.95</td>
<td>0.06</td>
<td>0.70</td>
<td>0.14</td>
<td>0.56</td>
<td>0.40</td>
<td>0.76</td>
<td>0.56</td>
<td>0.77</td>
<td>1.60</td>
<td>0.64</td>
<td>0.05</td>
<td>0.13</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Ratio 0.03 = slope 44.1°
Ratio 0.06 = slope 43.2°
Difference = slope 0.09°
## Methods

### Method 2: Standardized Slope

|    | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A  | -23.4 | 0.59 | 6.2 | 1.10 | 10  | 0.02 | 0.03 | 0.86 | 0.17 | 0.23 | 0.63 | 0.88 | 0.32 | 0.59 | 0.83 | 0.30 | 0.20 | 0.11 |
| B  | -22.8 | 0.28 | 4.4 | 1.10 | 44  | 0.95 | 0.06 | 0.70 | 0.14 | 0.56 | 0.40 | 0.76 | 0.56 | 0.77 | 1.60 | 0.64 | 0.05 | 0.13 |

**Difference = slope 8.7°**

**Average Diff = 10.2°**
## Methods

### Summary of 2 Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Percentage Difference (MPD)</td>
<td>Simple, intuitive</td>
<td>May overestimate difference when values very small or large</td>
</tr>
<tr>
<td>Standardized Slope</td>
<td>Does not overestimate difference when values very small or large</td>
<td>Less intuitive (values range from 0 to 135°)</td>
</tr>
</tbody>
</table>
Results

Comparison of 2 methods (53 paired samples + 6 paired replicates)
Results

Comparison of 2 methods (53 paired samples + 6 paired replicates)

Threshold of 0.10 = 91% consistent
Results

Comparison of 2 methods (53 paired samples + 6 paired replicates)

Threshold of 2.3 = 74% consistent
Results

Comparison of 2 methods (53 paired samples + 6 paired replicates)

Mean Percentage Difference (MPD) method performed better, with threshold of 0.10 (average of 10% difference in ratios) predicting 100% of matching samples with 9% error.
Summary

Benefits of MATLAB search tool:

- Can quickly compare samples to large database (676 USGS samples; >120 OSPR-PCL samples)
- Can help identify groups of matching samples that could have a chronic anthropogenic source
- Simple program can be easily shared (1 KB)

MATLAB search tool is NOT a statistical test to determine similarity (not intended to replace visual comparison)
Summary

Helped identify/refine groups of matching/similar samples
Summary

MATLAB dendrogram

Platform
A

Cosco
Busan
Summary

Next Steps:
- Continue to build OSPR-PCL ratio database
- Additional validation with larger sample size
- Investigate refinement of algorithm (some ratios better than others?)
Acknowledgments

- OSPR-PCL: Susan Sugarman, Shane Stahl, Dave Crane
- USGS biomarker ratio pioneers: Tom Lorenson, Bob Rosenbauer, Fran Hostettler, Ken Peters
- OWCN: Mike Ziccardi and many member organizations (especially IBRRC)
- OSPR-MWVCRC: Erin Dodd and Hannah Nevins

THANKS!