Quantifying Matrix Complexity with the Calculation of Fractal Dimensions from GC×GC Chromatograms – A Potential Oil Fingerprinting Technique

By

Gustavo Felipe Herrera Vega

A Thesis Submitted to

Saint Mary's University, Halifax, Nova Scotia

in Partial Fulfillment of the Requirements for

the Degree of Geology.

March 2019. Halifax, Nova Scotia

Approved: **Todd Ventura**

Pierre Jutras

Signature: _____

Signature: _____

Date: **April 1st, 2019**

Quantifying Matrix Complexity with the Calculation of Fractal Dimensions from GC×GC Chromatograms – A Potential Oil Fingerprinting Technique

By

Gustavo Felipe Herrera Vega

A Thesis Submitted to

Saint Mary's University, Halifax, Nova Scotia

in Partial Fulfillment of the Requirements for

the Degree of Geology.

March 2019. Halifax, Nova Scotia

Approved: **Todd Ventura**

Pierre Jutras

Signature: _____

Signature: _____

Date: **April 1st, 2019**

Quantifying Matrix Complexity with the Calculation of Fractal Dimensions from GC×GC Chromatograms – A Potential Oil Fingerprinting Technique

By G.F Herrera

April 1st, 2019

Gas chromatograms record molecular patterns according to the spatial separation afforded by the analytical method used and the type of mixture being analyzed. In this respect, the molecular complexity recorded in a gas chromatogram may be definable through the calculation of its fractal dimension. Fractals are a branch of mathematics that illuminate statistical elements in the spatial patterns of various complex forms. A fractal dimension is the measure of spatial complexity. For this thesis, the question is asked whether the molecular outputs recorded using gas chromatography results in a definable fractal dimensions? To do this, we quantified the matrix complexity of maltene fractions of various oils analyzed with comprehensive two-dimensional gas chromatography (GC×GC). A fundamental question addressed in this thesis is if fractal dimensions can be used as a new oil fingerprinting method? This study aims to find fractal dimensions in multiple samples of GC×GC chromatograms. These hypotheses are tested by analyzing a suit of oils and hydropyrolysates collected from various regions. These suites of oils have been previously determined to have different molecular complexities in their gas chromatograms, subsequently it is predicted that the fractal dimension of the more complex oils will be higher. The results of this study are fractal dimensions where calculated from the three different ocean bottom hydrothermal sites at Guaymas Basin, the Middle Valley and Escanaba Trough. The fractal dimension values did not correlate with expected trends of generation and maturation of sedimentary organic matter. Due to the lack of trends we focused the project on explaining what was wrong with the method, the program and the chromatograms. The results from the three different sites were inconclusive, but we expect that such results and analysis done can be helpful for anyone trying to relate fractal dimensions with oil and gas in the future.

Acknowledgements

The culmination of this project is the start of my journey as a scientist and I would like to express my sincerest gratitude to the people whom helped me and mentored me through this process because without them, this project would not be what it is today. First, I would like to express special thanks to my thesis supervisor, Professor Todd Ventura. He was the mind behind the project and I was the executor of the ideas. Secondly, I would like to give special thanks to Connor Dalzell from the Geology Master's Program, who helped me with the data gathering for the tables as well as instruction on how to use Matlab, which was the original idea for the algorithm. Special thanks also goes to Connor Tannahill from the Department of Math and Computing Science at Saint Mary's university, who helped me realizing that Matlab was not the proper program for the image analysis algorithm, and suggested other image analysis programs including the one used for the experiment (ImageJ). Finally, I would like to thank my family, friends, Saint Mary's University and the Geology Department, staff professors and students who were with me during these five academic years, and who helped me becoming a better scientist.

Table of contents

List of figures

List of tables

1. Introduction

Comprehensive two-dimensional gas chromatography (GC×GC) and the spatial complexity metric known as the fractal dimensions (FD) are two powerful tools used to explain different natural phenomena. GC×GC enables complex mixtures of volatile, organic compound to be quantitatively measured. The FD allows for the characterization of patterns and shapes in nature that have order beyond the dimensionality we typically consider. The purpose of this project is too combine these two approaches in a way that can be useful in geosciences, chemistry, and engineering. To achieve this, this study combines the data acquired by GC×GC with a boxcounting method to calculate the FDs of different naturally occurring oils. The following hypotheses are addressed:

- To what degree is the measured FD of an organic matrix sensitive to different GC run conditions?
- Can FDs be used to determine differences in molecular make-up of different crude oils?
- If so, are there specific FDs for oils that have been weathered by different processes?
- Can the FD be used to measure changes in oil maturity?

Following these directions of study, if the method is found to be unreliable, then we will discuss what factors confounded this hybridized approach.

1.1. Gas chromatography

Gas chromatography (GC) is a commonly used analytical technique for analyzing mixtures of volatile organic compounds (analytes). The mixture is first dissolved in a solvent called

the mobile phase, which is carried though a capillary column packed with a semi-permeable material called the stationary phase. The separation of a mixture is based on the differential partitioning between the mobile and stationary phases. This is achieved by subtle differences in component partition coefficients that induce a retention lags on the effluent as the eluting matrix interacts with the stationary phase (Sparkman, Penton & Kitson, 2011). The separated matrix is then passed to a flame ionization detector or mass spectrometer where it is measured as a signal intensity of carbon atoms or cations. The resulting separated molecular data are plotted as a 2D, x-y diagram called a chromatogram, where *x* corresponds to retention time and *y* indicates the signal response created by the analytes exiting into the GC's detector. In optimal separation, this signal is proportional to the concentration of the specific analyte being separated.

- **Qualitative chromatographic analysis:** The *x-y* plot will produce an array of peaks for a sample representing the analytes present in a sample eluting from the column at different times, different retention times represent different analytes if the method conditions are held constant (Sparkman, et al, 2011).
- **Quantitative chromatographic analysis:** Each peak represents a different compound or analyte, and the area under the peak is proportional to the amount of analyte present in the chromatogram. By calculating the integrated area of the peak, the concentration of analyte in the original sample can be determined (Sparkman, et al, 2011).

Comprehensive, two-dimensional chromatography (GC×GC) is a more advanced chromatographic technique, in which a mixture is separated by passing it through two different gas capillary columns, each having a unique stationary phase. The first column

enters into a cryogenic modulator that injects parcels of the sample into the second column over a defined temporal cycle. The modulator freezes mixtures for a few seconds so that a mixture cryo-focus into a very sharp band. The modulator then fires the packet into the head of the second column. This is repeated for the duration of the sample run (Ong & Marriott, 2002). The GC×GC method offers a large increase in peak capacity, this allows the chromatogram to show more organic compounds. Peaks that would ordinarily be overprinted by coeluting bigger peaks using a traditional GC method will be sharp, clear, and separated using GC×GC chromatography. The chromatograms in this study are 3D plots with *x*-*y*-*z* axes.

1.2. Introduction to fractal dimensions

The term fractal describes geometrical objects or functions that are scale invariant and represent a part of an object or function that is itself similar to the whole i.e. having self-similarity (Feder, 1988). Fractal dimensions do not belong to the common known dimension 1, 2, and 3; but are somewhere in-between. The calculation of a FD follows a simple process. For any shape, the number of objects (N) that can define the obsject increases with magnifying factor (r) at different dimensions (Figure 1.1) following for example:

1D: a line expands linearly, according to the magnification factor 1, 2, 3…

2D: the square expands equally as the magnification factor raised to the square 1, 4, 9…

3D: the cube expands equally to the magnification factor raised to the cube 1, 8, 27…

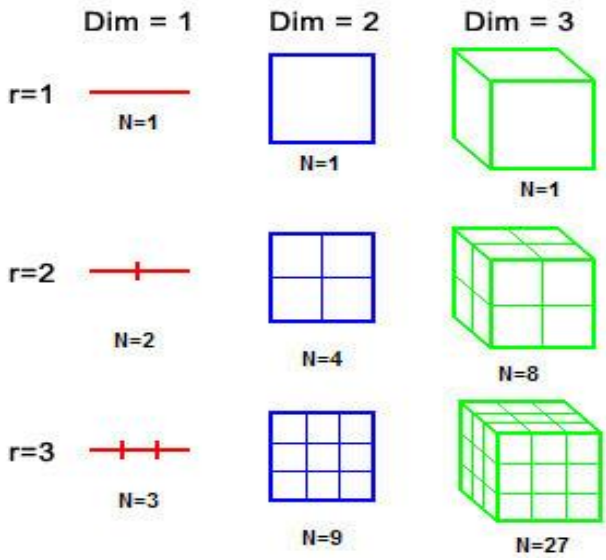


Figure 1.1 The relationship between dimension, magnifying factor (r) and the number of objects (N) (from: http://fractalfoundation.org).

A simple relationship exists between the three variables:

$$N = r^{D} \quad (1)$$

For example, if D = 3 and r = 3, N will be equal to 27. Calculation of the number of objects present in an image is achieved by:

$$\log (N) = \log (r^{D}) \text{ and } \log (r^{D}) = D * \log (r) \quad (2)$$

$$D = \log (N) / \log (r) \quad (3)$$

leaving an equation to estimate dimensions, relating the log of N and the log of r. The same equation can be used to calculate the FD in multiple self-similar synthetic objects like the Koch snowflake (Figure 1.2) or the Sierpinski triangle (Figure 1.3).

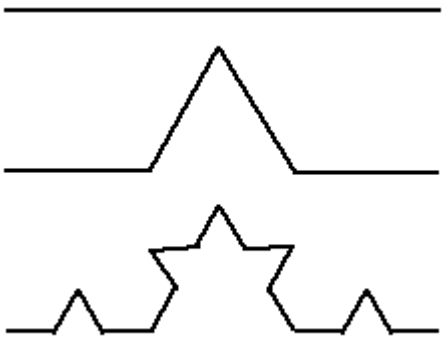


Figure 1.2 Image of Koch snowflake by multiplicative iteration of line segments showing self-similarity (from: https://www.shodor.org). The fractal image of the Koch curve is wigglier than a straight lie. It has a FD = 1.26, which is between 1D and 2D because the shape does not fill a whole 2D plane.

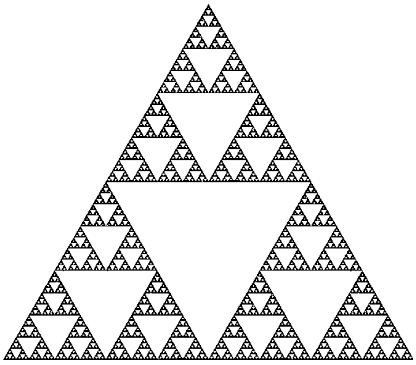


Figure 1.3 The Sierpinski triangle (from: https://www.zeuscat.com) fits in a 2D plane, but does not fill the shape completely. Here the FD = 1.58 is a little more than the Koch snowflake, meaning that the Sierpinski triangle does a better job filling up a 2D plane.

1.3. Fractal dimensions in nature

Fractal dimensions are a common property of nature. For example, the scaler size of rock fragments within sedimentary successions, faults within fault zones, and the bifurcation pattern of the roots, stems, trunks and venation of leaves for plants and trees are well known to have fractal properties (*see the fractal geometry of nature by Benoit B. Mandelbrot*). A rocky shoreline is also fractal in nature.

:

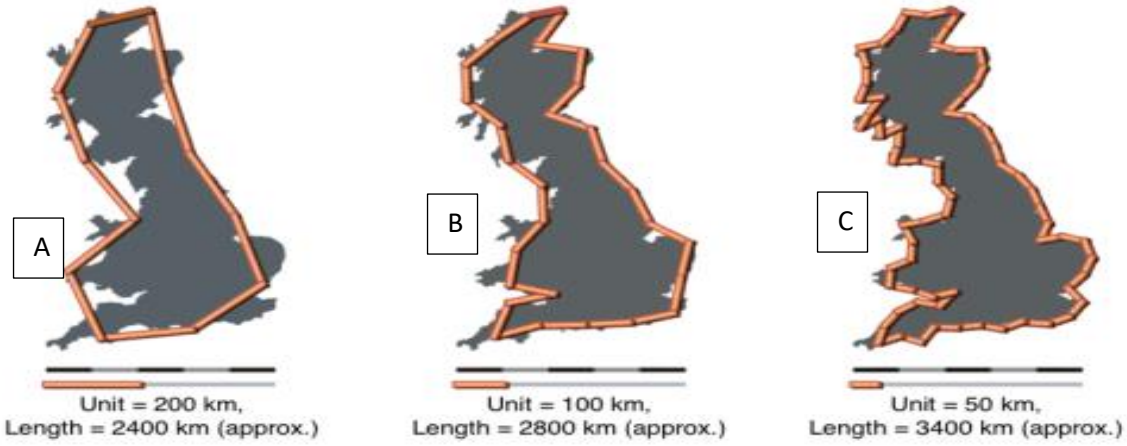


Figure 1.4 The coastline paradox proposed by Lewis Fry Richardson, shows how the length of the coast of Britain apparently increases when decreasing the measure stick (from: http://rebloggy.com).

Figure 1.4 illustrates how geographic shapes have FDs. With different unit lengths of measure, the total length of the coastline increases, meaning that the length of the coastline depends on the method in which it is measured. The paradox is that it is impossible to accurately measure the coastline of Britain (or any other coastline), because the coastline becomes progressively longer as the unit of measure is more closely tied to a more microscopic level of calculation. At the most extreme level, the length of the coastline approaches infinity. This is illustrated in Figure 1.4 with a measure stick of 200 km (r =1), N the number of sticks required to cover the whole are is equal to 12, hence the length is approximately 2400 km. At B, the measure stick length is divided by 2, 100 km, (r =2) the N number of sticks required to measure the coastline increases to 28, thus the total length approximately 2800 km. At C, a further division of the stick by two from B is 50 km (r = 4) and the N number of sticks increases to 68, with the total length approximating 3400 km.

There is a mathematical pattern in the geometric quality of fractals. As the stick shrinks r increases and N increases too. We are interested in the rate at which the perimeter changes as a function of the ruler length. In order to solve this question and understand the relationship between N, r and stick length, results are plotted in a log graph with the log (N) at the y-axis and the log (r) in the x-axis:

Table 1.1 Log (N) vs Log (r) of Britain's coastline*.

Log (r)	Log (N)
Log (1) = 0	Log (12) = 1.08
Log (2) = 0.301	Log (28) = 1.44
Log (4) = 0.6	Log (68) = 1.83

*data from: http://rebloggy.com

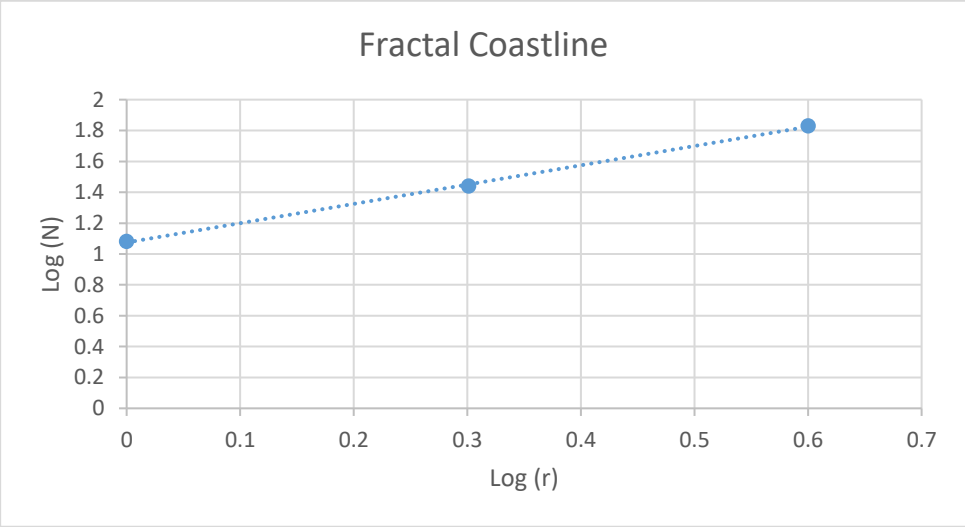


Figure 1.5 Cross-plot of Log (N) versus Log (r) for the British coastline. The regression line indicates the relationship between the number of objects and the magnifying factor.

The slope of the line that crosses the points indicates how quickly the perimeter changes versus the magnification factor. This slope will be equal to the change in y-value divided

by the change in the x-value. Hence, the slope will be equal to D. In this particular case, after calculating the slope taking r = 1 and r = 4:

$$D = \Delta \log (N)/\Delta \log (r) \quad (4)$$

$$D = 1.83 – 1.08/0.6 – 0$$

$$D = 1.25$$

The final dimension is equal to 1.25, which is the actual measured FD. The lower the dimension the straighter and smoother the coastline. The higher the dimension the more wiggled and jagged the coastline (fractal foundations). Not all natural fractals are so easy to measure and more sophisticated methods have been proposed by to count for various fractal dynamics (Zmeškal,2001).

1.4. **The Hausdorff (box-counting) method**

The box-counting method is analogous to the perimeter measuring method used in the coastline of Britain (Figure 1.6). The coastline is covered with a grid. A count is then made for how many boxes of the grid cover the coastline. This is repeated using smaller boxes in the grid so that the structure of the coastline is more accurately captured.

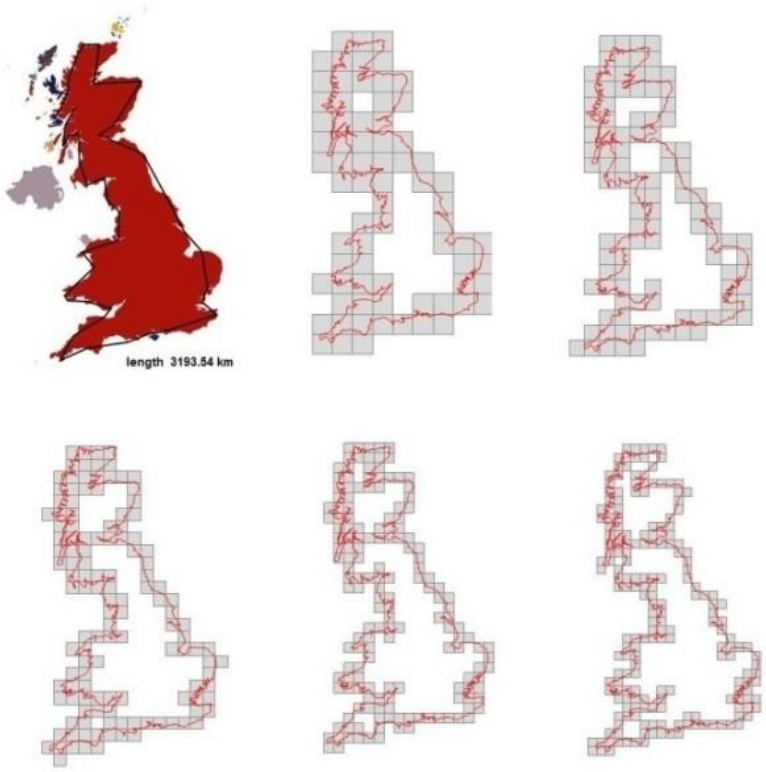


Figure 1.6 The coast of Britain is covered by boxes of a fixed size, then this size is reduced depending on the magnifying factor. This will have the same effect as Figure 1.4. The smaller the boxes, the longer the coast of Britain will be (from: https://www.researchgate.net/).

A plot with the data for the number of boxes (N) against the magnification factor (r) or the inverse of the box size can be made, and the slope of the line crossing the plotted points will be equal to the dimension (D). The same equation as before (D = log (N)/log (r)) is used to find the slope, which will again give a number between 1 and 2, but a more accurate number than the last method due to an increase in the magnification factor. A steep slope indicates that the object has a higher FD, which means that it gains complexity as the box size is reduced, a flatter slope means that the object has a lower FD and the amount of detail does not grow as quickly with decreasing box size.

Oils are known to be the most complex naturally occurring organic mixtures on Earth. Most of these molecules in oil arise by covalent bonding of just five different elements C, H, O,

N, and S and of these C and H are by far the most abundant. Due to constraints in the bond angles and the processes by which organic compounds are formed a considerable amount of symmetry should theoretically exist within organic mixture. This symmetry is also further ordered when passed through a gas chromatograph. As such, the molecular complexity of oils could be compared based on the FD that is produced from their gas chromatograms.

For this study, we hypothesize that if a FD can be accurately calculated from GC×GC chromatograms, this technique can provide a new method for calculating molecular complexity that can be used to fingerprint different types of oils or other complex mixtures. This hypothesis is tested with a suit of oils and hydropyrolysates that have been previously determined to have different complex mixtures of hydrocarbons.

2. Methods

2.1. Preparation of chromatographic images

The hypothesis was tested with three different sample sets (Table 2.1) gathered by Professor Todd Ventura. The first sample set, referred to as Cathedral Hill, is a hydrothermal vent site in the Guaymas Basin, Gulf of California. The second sample set is composed of hydropyrolysates derived from homogenized, ambient sediments collected near a hydrothermal vent site called Middle Valley, which is off axis the Juan de Fuca Ridge, off the coast of Washington State and Oregon, USA. A third site is the Escanaba Trough near the Gorda Ridge, which is also in the northeastern end of the Pacific Ocean.

These samples have different molecular complexities, they come from different sites that show different oil forming conditions and sources of organic matter.

The organic matter was solvent extracted from their sediments, and the maltene fraction was isolated by flash chromatography. The maltene fractions were then analyzed with a GC×GC-FID (a comprehensive two-dimensional gas chromatograph linked to a flame ionization detector). The samples for this project where selected to test various geochemical processes and to evaluate the effectiveness of the complexity metric to resolve changes in molecular composition of each maltene sample. After the maltene samples were analyzed with GC×GC and the molecules of the sample matrix were separated and plotted in a 3D graph as peaks represented in a 2D plane; the sample chromatogram was exported as a JPEG image (Figure 2.1).

Table 2.1 Maltene samples, collection site, and expected experimental test.

Sample set	Test	Sample number
Middle Valley, hydrothermal vent field	Hydropyrolysis, time and temperature controlled experiments	37 samples
Guaymas Basin, hydrothermal vent field	Complexity due to thermal gradient	34 samples
Escanaba Trough, hydrothermal vent field	Spot samples, difference in source material	3 samples

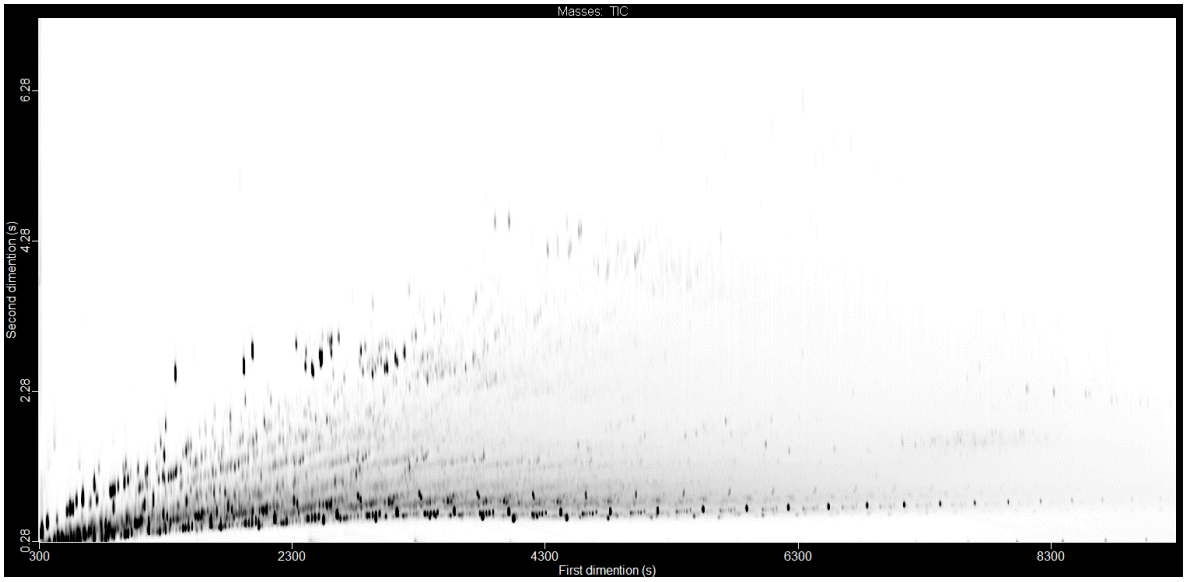


Figure 2.1 GC×GC-FID chromatogram of a maltene fraction of an oil.

The chromatogram in Figure 2.1 has well resolved molecules (black dots) that do not converge on one another. For the Guaymas Basin samples, the FD of sample extracts from four push cores were analyzed. The cores were labeled: core 3, 5, 6, and 8. Approximately 10 chromatograms were produced from each core spanning the shallowest surface sediments to the more deeply buried sediments at the bottom of the core with the depth of the sediment extracted representing the sample name. Fractal dimensions were also measured for all the samples from Middle Valley, which account for 34 of 66 different chromatograms. An additional 14 samples from the Escanaba Trough where also analyzed. Various samples from Middle Valley and Escanaba Trough sites were omitted from FD analysis if they showed notable chromatographic artifacts due to excessive column bleed or the presence of contaminants in the GC×GC-FID chromatogram (see Figure 2.2 and Figure 2.3). Figure 2.2 and 2.3 illustrates the two most common errors produced by bad sample runs.

Figure 2.2 A blank GC×GC-FID chromatogram. Black dots at the bottom of the chromatogram are septa bleed. Streaking along the bottom is the solvent peak. Arching lines across the right-hand side of the plot is column bleed. These chromatographic artifacts represent interferences that complicate the calculation of chromatographic complexity using fractal dimensions.

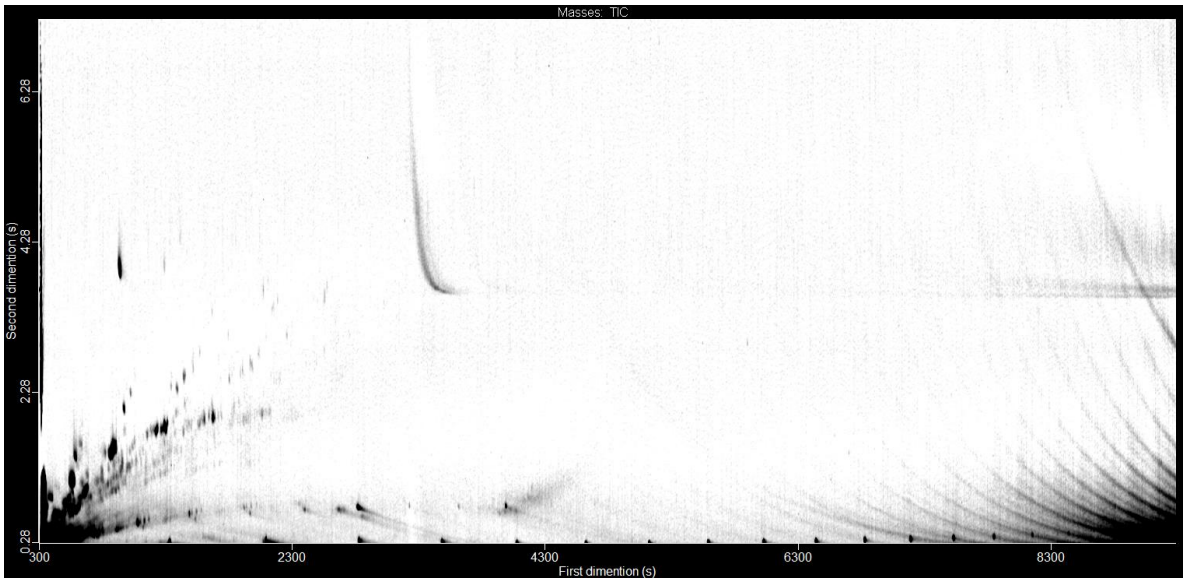


Figure 2.3 A noisy blank GC×GC-FID chromatogram. These chromatograms were omitted from box counting analysis.

FIJI and FracLac software were used to measure the FDs in our study. FIJI is a program created only for image analysis. FracLac is a plug for FIJI that is a medical tool for calculating FD and lacunarity in images. Even though FracLac was designed for image analysis in the medicine field, it appeared to be the best option for the purpose of this study because is a tool focused on finding D and human created images. Standard operating procedures of FIJI and FracLac can be found in *Appendix 1.*

2.2. Post-data processing of chromatographic images

The goal of image preparation is to make sure that every chromatogram looks the same to in order to reduce image interferences that can create errors in the calculation of an FD. The box count algorithm in FracLac works in the same way as a simple counting algorithm, dividing the image in either 1 or 0, it either exist or is absent.

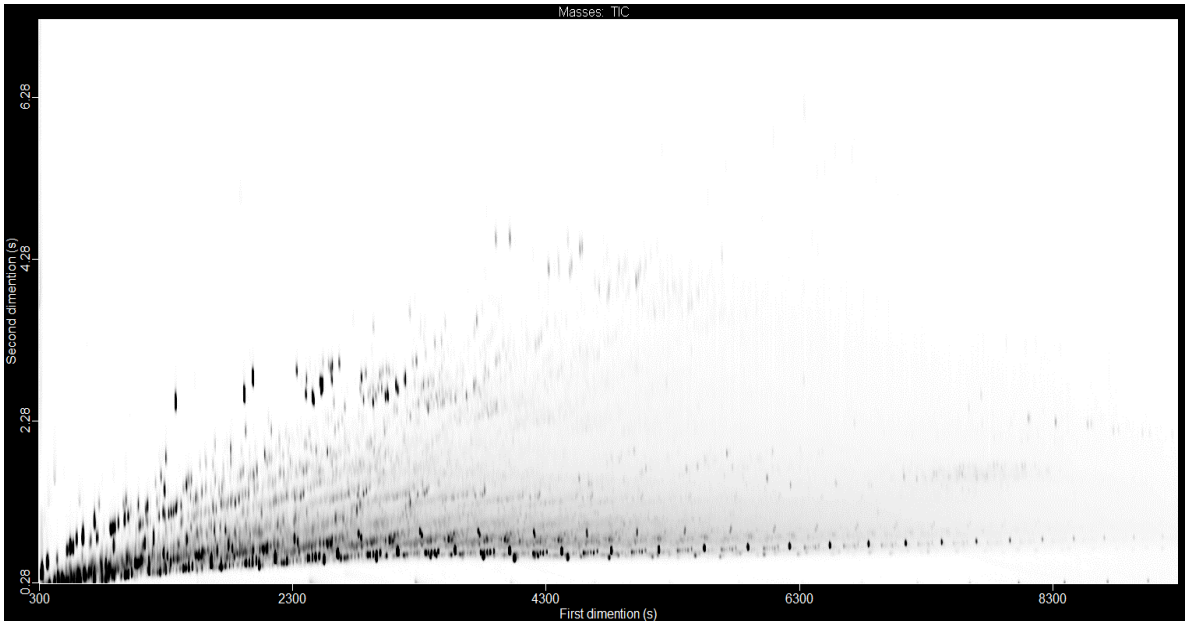


Figure 2.4 Notice that the image is composed of a background and foreground. In this case, the foreground represents the area of interest or the black peaks that are present in the chromatogram, on the other hand the background is the white regions of the chromatogram.

FracLac in this case focuses the box counting algorithm in the black peaks (or foreground) of the image; but this image will give an error if is analyzed just like it is above, due to the x-y axis black bars, where the black peaks are spread. In this case, FIJI has the best solution for cutting the image down to minimize the error that the x-y axis could cause during the run. At the same time, while editing the chromatogram FIJI also allows users to create macros in order to record operations and redo them without having to reset the edition process.

This feature of FIJI was very important during the project, allowing to edit and cut the x-y axis of all the chromatograms used for the project in the exact same shape. This feature ensures that every chromatogram look the same, reducing the error than any different edition could cause to the calculation of an FD.

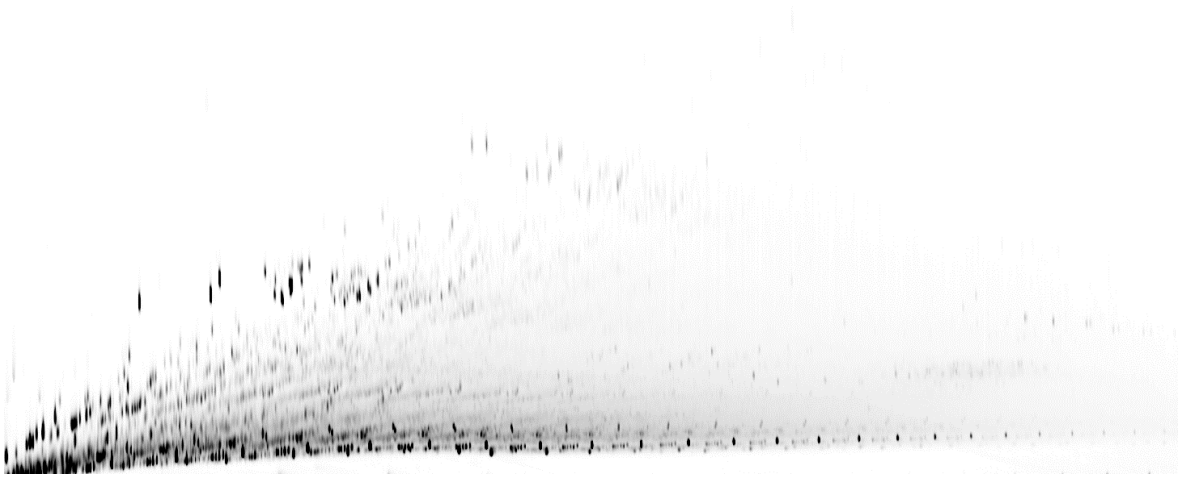


Figure 2.5 Example of a GC×GC chromatogram after edition with FIJI with black bars at the left side and base of the chromatogram removed.

After the edition is over the chromatograms look exactly like Figure 2.5; now without the x-y axis the chromatograph is ready for analysis.

Two methods where used in order to test our chromatograms to get a FD. First, a run was done using the option of grayscale box count. Second, a run was then done using a binary box count; the details about the difference between each method and how to configurator FracLac for the different runs can be found in *Appendix 1*.

3. Results

3.1. Fractal analyses using a greyscale method (3D box count)

The first method used was a greyscale box count, which should be suited for chromatogramic plots having 3 data axes. This is because the greyscale method assumes images exist in a pseudo-3D space. In the case of its application to GC×GC chromatograms it is assumed that molecular intensity can be calculated as a component of molecular complexity. The grayscale scan was performed on Guaymas Basin cores 3, 5, 6 & 8, and its results of the calculation are providing in Table 3. The FD of the core samples did not significantly vary even though the chromatograms have substantial molecular compositional differences (Table 3). This result even extended to a blank chromatogram for testing, which yielded an FD of 2.6. The method was therefor found unreliable for the purpose of the project and the Middle Valley and Escanaba Trough sample sets were not analyzed.

Table 3.1 Results differential greyscale box count analyses for the Guaymas Basin push core transect.

Core depth (cm)	Core 5 (closest to vent)		Core 6 (slightly farther from vent)		Core 8 (further from vent)		Core 3 (outside of vent)	
	FD§	**SD***	**FD**	**SD***	**FD**	**SD***	**FD**	**SD***
0-2cm	2.665	0.0043	2.6530	0.0034	2.6231	0.0060	2.6205	0.0052
2-4cm	2.645	0.0029	2.6395	0.0032	2.6141	0.0056	2.6477	0.0069
4-6cm	2.628	0.0032	2.6879	0.0043	2.6061	0.0049	2.6175	0.0043
6-8cm	2.645	0.0027	2.7035	0.0041	2.6351	0.0048	2.6614	0.0040
8-10cm	2.625	0.0031	2.6862	0.0036	2.6573	0.0080	2.6562	0.0051
10-12cm	2.673	0.0041	2.6750	0.0044	2.6566	0.0048	2.6556	0.0056
12-15cm	2.678	0.0035	2.6746	0.0052	2.6602	0.0056	2.6461	0.0055
15-18cm	2.696	0.0038	2.6693	0.0051				
18-21cm	2.688	0.0043	2.6597	0.0039				
Avg. SD*	0.026	0.0035	0.0190	0.022	0.0220	0.0057	0.018	0.0057

*The reason behind the importance of the standard deviation, lies in the experimentation done through the analysis, the standard deviation allows scientist to have an error margin, but also allows scientist to find best fits while comparing variables in a graph. Standard deviation is the error that comes along with the experiment and in this case is produced by the program that performs the box counting algorithm FracLac.

§FD = fractal dimension
*SD = standard deviation
Avg. SD = average standard deviation

Cores, 3, 5, 6 and 8 represent a very interesting set of oils for this project because there is a defined change in molecular complexity with depth due to the migration of petroleum forming compounds that are likely being cracked out of the sedimentary organic matter at a window in between 8cm to 12cm depth. Depth then was an important variable to test in this set of cores, Table 4 represents the relationship between FD and depth in grayscale.

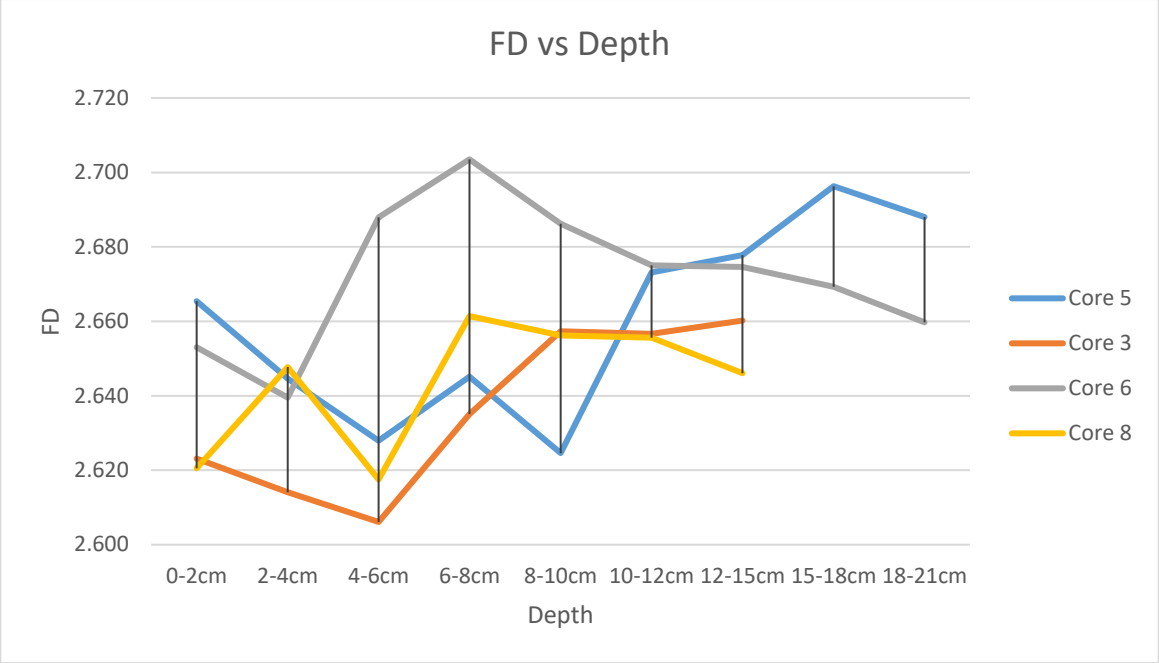


Figure 3.1 Grayscale analysis cross-plot of core depth vs FD.

The comparison of FD versus depth of core samples in Figure 3.1 does not have a significant trend and it does not represent the increase in complexity experience by these oils when cracked out from the kerogen. This lack of relationship was likely due to two errors. First, because of the way a GC×GC chromatogram is plotted, it was not possible to accurately reflect the Z dimension (or Psuedo-3D). The color change to either a mountain or bulls eye GC×GC chromatogram plot automatically creates a color contrast that scales from dark to light back to dark at the highest signal intensity. Therefore, FracLac cannot relate the intensities of a peak to a third dimension for an FD calculation. The spatial complexity is substantially biased in favor of greater complexity (or a higher FD) than actually present on the chromatogram.

3.2. Fractal analyses using a binary method (2D box count)

The binary method at first did not seem to be especially useful for the project, because it analyzes the chromatogram with a 2D box count algorithm, meaning that the magnitude of the peaks cannot be calculated. However, after the greyscale scan did not give reliable results, this alternative method was explored (see *Appendix 1* for more information about the binary option in FracLac). The binary results were more promising than the greyscale scan with variations in FDs that range between 0.87 (Figure 3.2) from a blank chromatogram to 1.62 (Figure 3.3) for a complex chromatogram.

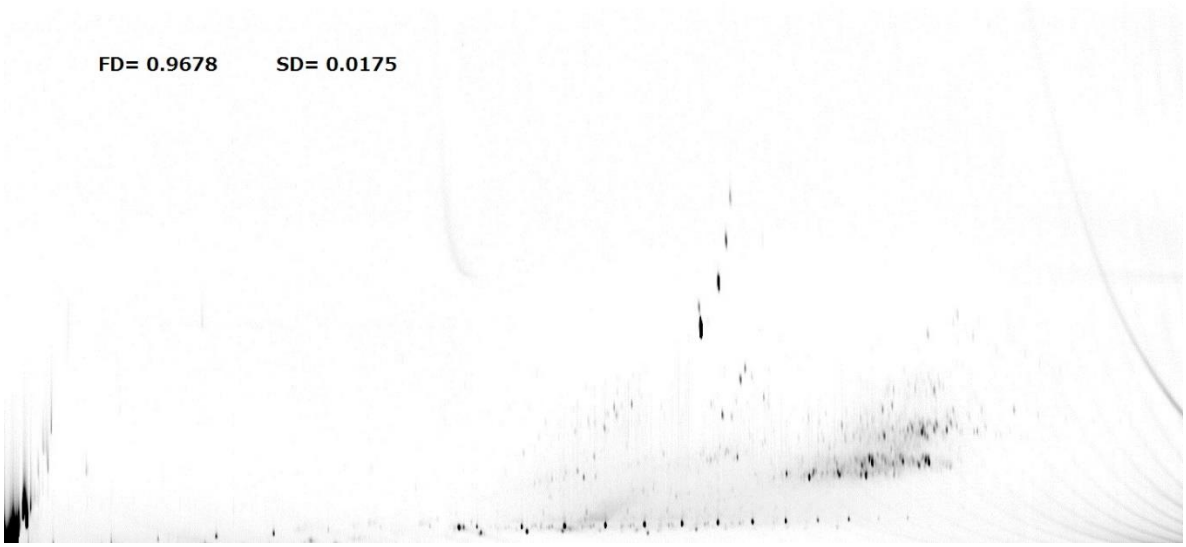


Figure 3.2 GC×GC-FID chromatogram of sample S18. This sample recorded the lowest FD by FracLac, belongs to Middle Valley sample set, and represents a hydropyrolysate heated at 250˚C for 72 hours.

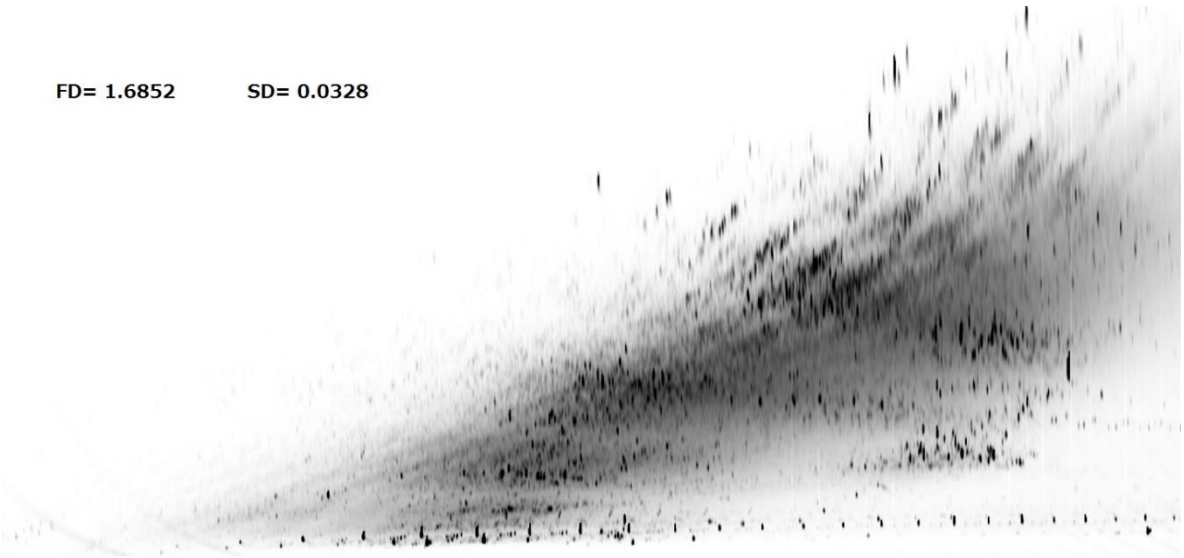


Figure 3.3 GC×GC-FID chromatogram of sample Core8, 6-12cm. This sample recorded the highest FD recorded by FracLac, belongs to Guaymas Basin, Core 8 sample set from an oil found at 6cm to 8cm.

Table 3.2 summarizes the results from the binary scans of the Guaymas Basin. Table 3.2 shows the relationship between depth and FD. Tables 3.3 and 3.4 summarize the results from the Middle Valley hydropyrolysis experiments and Escanaba Trough samples.

Table 3.2 Results of the binary analysis for a Guaymas Basin push core transects.

Core depth (cm)	Core 5 (closest to vent)				Core 6 (slightly farther from vent)				Core 8 (further from vent)				Core 3 (outside of vent)			
	FD	**SD**	**PD***	**Peak#§**	**FD**	**SD**	**PD***	**Peak§**	**FD**	**SD**	**PD***	**Peak#**	**FD**	**SD**	**PD***	**Peak§**
0-2cm	1.5783	0.0233	95.4	4533	1.5909	0.0254	96.1	4564	1.6656	0.0103	70.9	3368	1.6498	0.011	92.3306	4387
2-4cm	1.5810	0.0227	126.0	5986	1.5853	0.0111	86.2	4094	1.6306	0.0283	101.1	4804	1.6647	0.0141	50.9323	2420
4-6cm	1.5636	0.0221	124.5	5917	1.6231	0.0186	90.2	4288	1.6543	0.0107	78.6	3742	1.6018	0.0078	53.7946	2556
6-8cm	1.5795	0.0203	127.6	6063	1.5763	0.0238	101.8	4837	1.6852	0.0328	139.4	6623	1.6067	0.0219	125.879	5981
8-10cm	1.2160	0.0208	125.8	5975	1.5925	0.0238	131.8	6264	1.6718	0.0261	140.9	6694	1.5966	0.0217	101.991	4846
10-12cm	1.3395	0.0071	109.1	5184	1.3438	0.0269	103.9	4939	1.6750	0.0305	139.0	6604	1.6158	0.0307	104.369	4959
12-15cm	1.4153	0.0279	104.9	4982	1.5402	0.0307	115.2	5474	1.5986	0.0309	76.8	3648	1.5719	0.0266	80.5235	3826
15-18cm	1.4786	0.0302	89.6	4256	1.4917	0.0159	126.4	6004								
18-21cm	1.4814	0.0232	88.5	4306	1.5228	0.0206	127.48	6051								
Avg.	1.4700	0.0220	110.2	5245	1.540	0.022	108.8	5168	1.6544	0.0242	106.7	5069	1.4174	Avg.	1.47	0.0220
SD	0.1270	0.0220			0.084	0.022			0.030		SD	0.127	0.0220			0.084

* PD = Peak Density
§Peak# = Peak Number

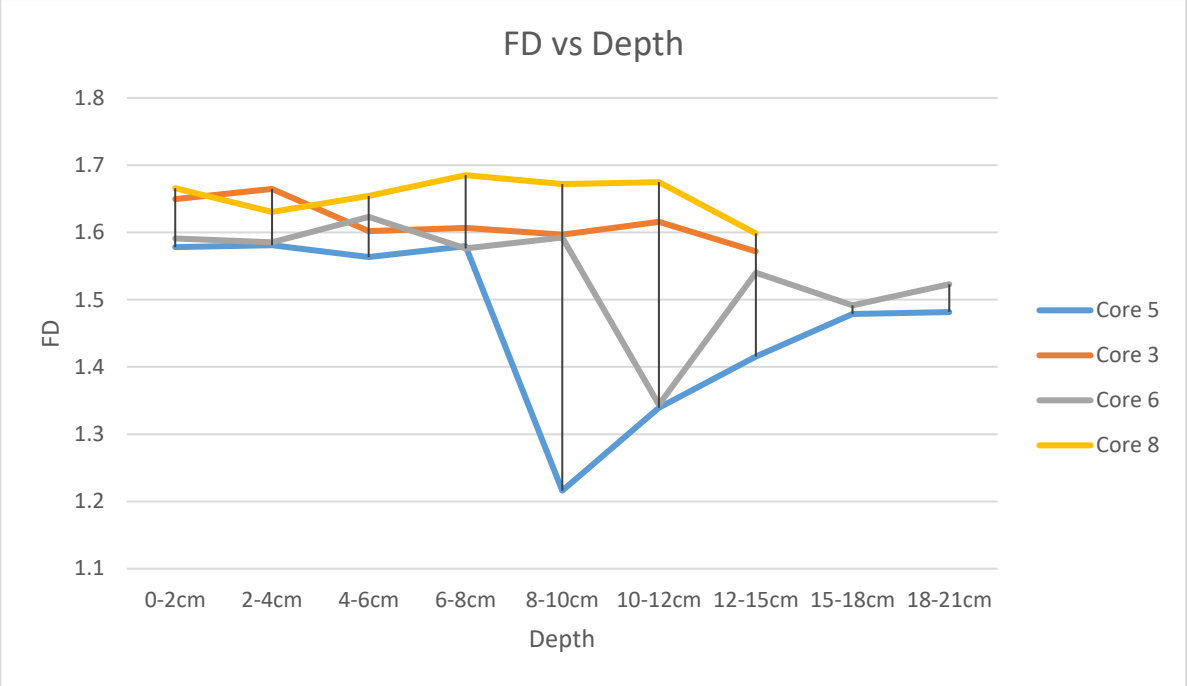


Figure 3.4 Binary cross-plot of FD vs depth for Guaymas Basin. For the 8-12cm window of cores 5 and 6 the FDs is lower and then gradually increases with deeper sediment depth. For cores 3 and 8 the FD stays nearly constant.

Cores 3 and 8 are very similar oils with respect to their molecular composition. The same results are obtained for cores 5 and 6. Compared with the greyscale analyses, these results shows a better predictive trend. However, the oil window within the Guaymas Basin samples, where vent temperatures were high enough to crack the sedimentary organic matter to form oil is found by the FD analyses to produce an opposite complexity response. The expected result was an increase in FD at 8cm-12cm across this window followed by a downward FD trend in the deeper sediment depths. The FD of the cores was plotted against peak density (Figure 3.5) and against peak number (Figure 3.6). The lack of a positive correlation between peak number/density and FD, suggests that the FD was not being calculated through the amount of peaks or its intensities.

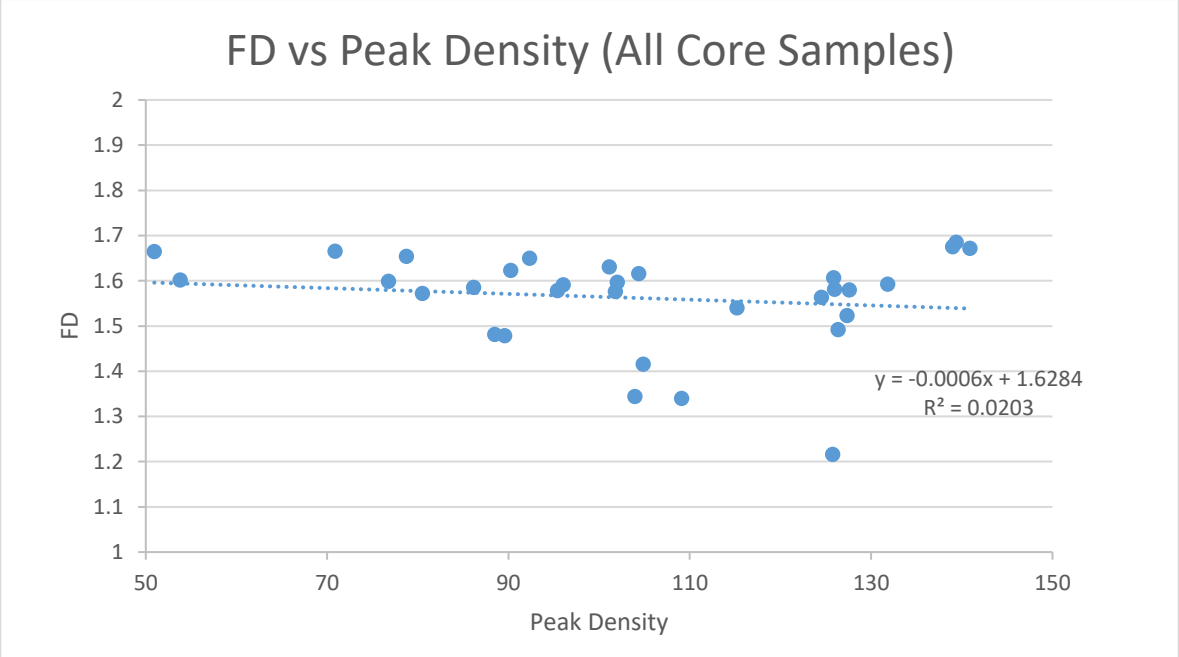


Figure 3.5 Cross-plot of FD vs peak density for all push core samples.

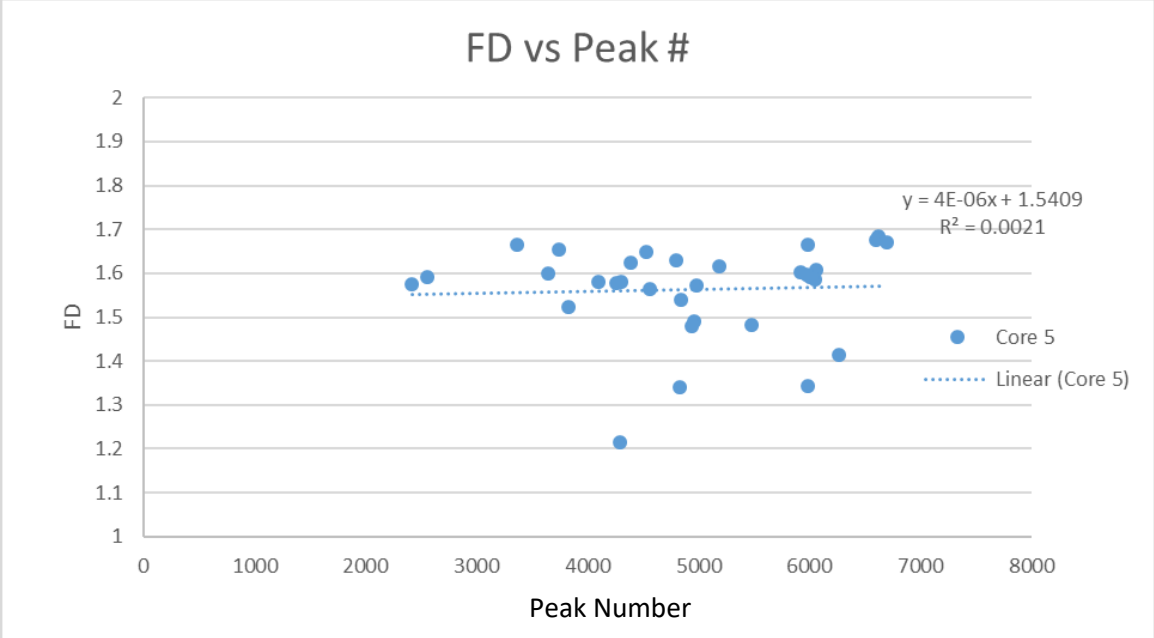
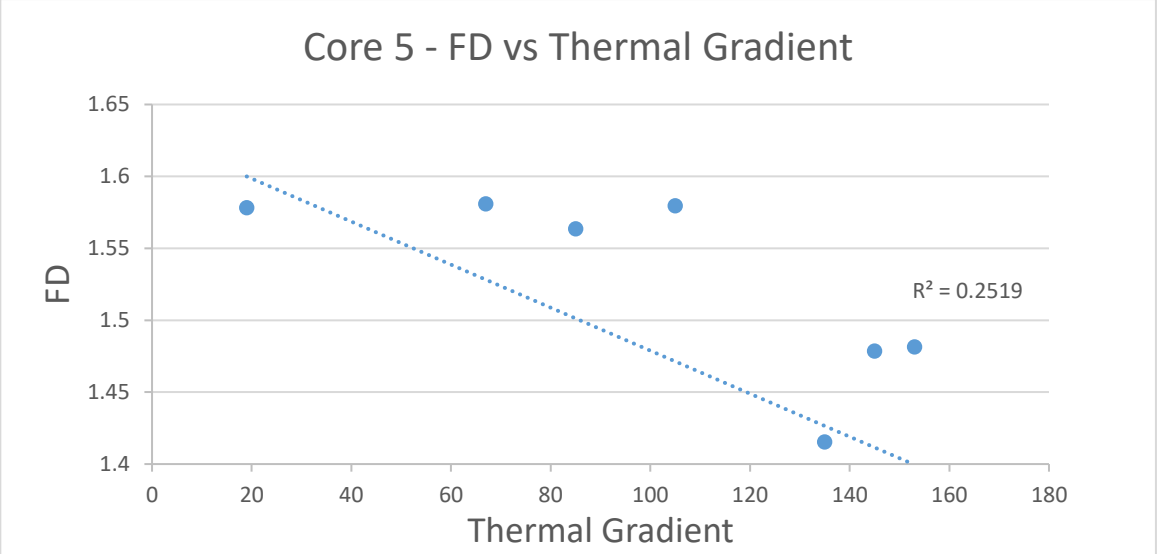
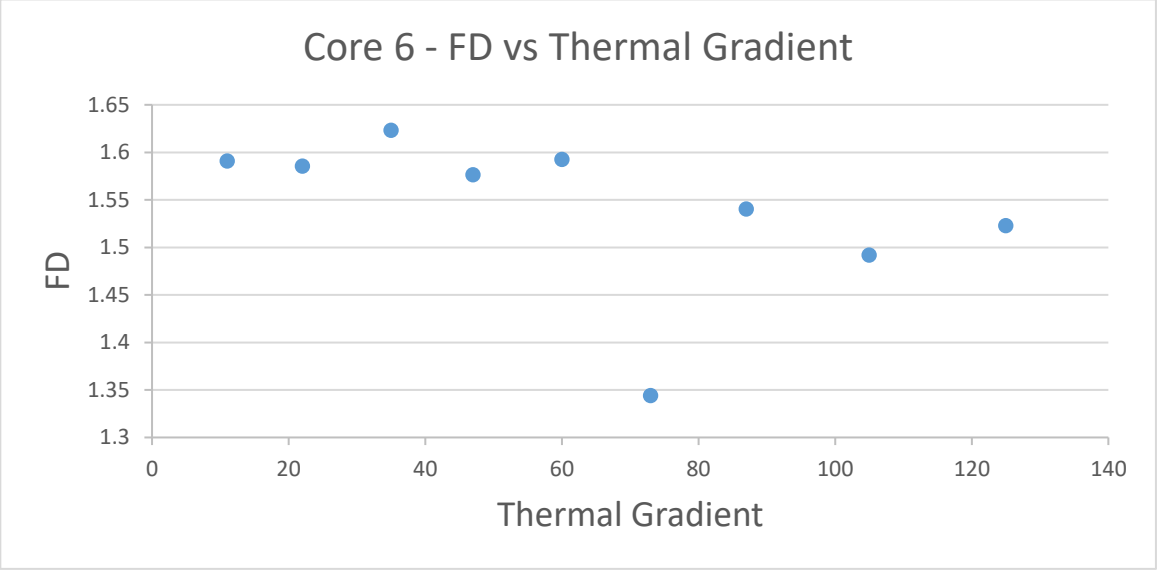


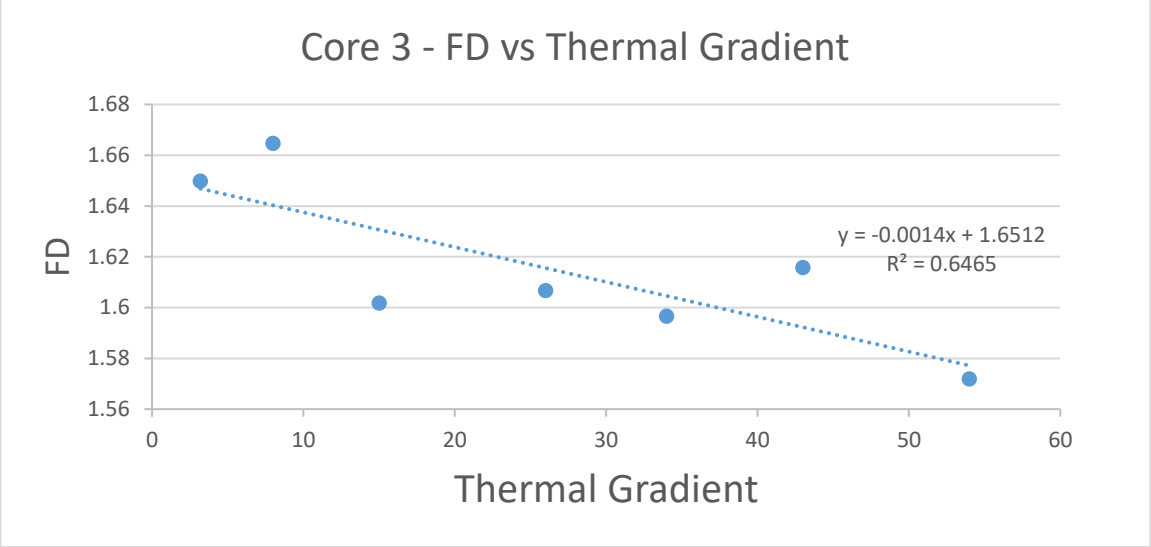
Figure 3.6 Cross-plot of FD vs peak number of all push core samples.

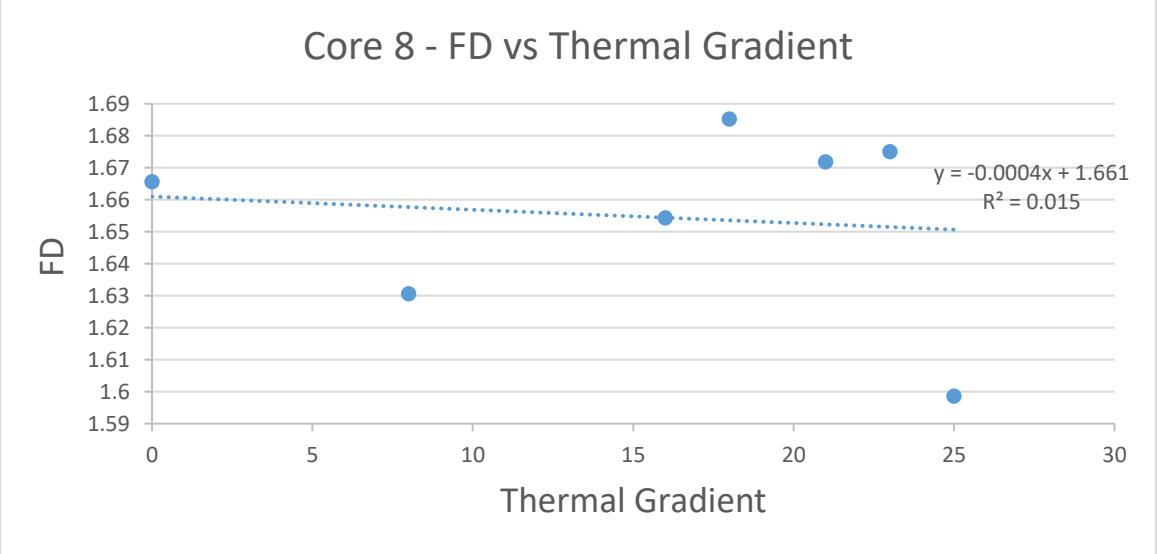
A final comparative study was performed on the cores by comparing their FDs against the recorded thermal gradients of the sediments that were extracted (Figure 3.7). Fractal

dimension versus thermal gradient was compared in order to detect whether changes in temperature affect the the FD, keeping in mind the oil window and the similarities between cores 3 and 8 and cores 5 and 6. The FD does not seem to be affected by the thermal gradient, there are low FD values in high temperature samples as well as there are high FDs with low temperatures, in fact the relationship between FDs and thermal gradient almost mimics the relationship between FD and depth.









Figures 3.7 Results of the different cores when compared with their thermal gradients, notice that even though they are similar to Figure 15, there is not a clear relationship between FD and thermal gradient.

Table 3.3 Binary box count results of the Middle Valley hydropyrolysis experiment (the three error slides (S63, S66 and S67) represent chromatograms having *granularity*).

Slide Number	Hydropyrolysis Experiment		FD	SD
	Time (hrs.)	Temperature (ºC)		
S2	Initial 2%	Initial 2%	1.5800	0.0160
S3	24	100	1.6058	0.0160
S4	48	100	1.5023	0.0195
S5	72	100	1.4952	0.0165

S6	96	100	1.2030	0.0191
S7	120	100	1.1949	0.0239
S16	24	250	1.1083	0.0238
S17	48	250	1.2569	0.0177
S18	72	250	0.9678	0.0175
S19	Blank	Blank	0.8791	0.0244
S22	48	250	1.2932	0.0143
S24	Exxon Valdez	Exxon Valdez	1.3761	0.0209
S29	24	100	1.5846	0.0134
S30	48	100	1.5673	0.0133
S31	72	100	1.5792	0.0127
S32	96	100	1.4154	0.017
S33	120	100	1.104	0.0129
S35	Initial 4%	Initial 4%	1.3354	0.0058
S36	24	150	1.4801	0.0052
S37	48	150	1.5232	0.0129
S38	72	150	1.4623	0.0249
S39	120	150	1.5530	0.1290
S40	24	200	1.4978	0.3160
S44	24	250	1.5636	0.0056
S45	48	250	1.6109	0.0079
S46	72	250	1.3058	0.0097
S47	96	250	1.1965	0.0244
S48	24	350	1.1924	0.024
S49	48	350	1.115	0.0165
S50	72	350	1.1908	0.0144
S51	96	350	1.1299	0.0182
S52	Initial 7%	Initial 7%	1.3068	0.0092
S63-error	--	--	1.6083	0.0239
S64			1.1287	0.0088
S65			1.1706	0.0237
S66-error	--	--	1.6817	0.0235
S67-error	--	--	1.5772	0.0273

.

The Middle Valley hydropyrolysis experiment used a single homogenized ambient sediment sample that was divided into a large number of sample splits. Each split was pyrolyzed at a different temperature over a specific time interval to artificially generate oil. The oil extract was measured using GC×GC-FID. The highest FDs for the Middle Valley

samples are S3 (1.6058), S29 (1.5846) and S45 (1.6109). Samples S3 & S29 belong to a hydropyrolysis experiments accomplished by heating the mixture to 100˚C for 24 hours, whereas S45 belongs to a hydropyrolysis experiment heated to 250˚C for 48 hours. The lowest recorded FD where S16 (1.1083), S18 (0.9678), S33 (1.104), S49 (1.115) and S51 (1.1299). S16, S18, S49 and S51 underwent hydropyrolysis at temperatures over 200˚C, and S33 was heated at 100˚C for 24 hours. A comparison of hydropyrolysis reaction time versus FD depending on temperature was compared for this suit, trying to find a relationship between the intensity of the FD when the oil is heated at different temperatures, but because no specific trend showed up, the relationship cannot be verified.

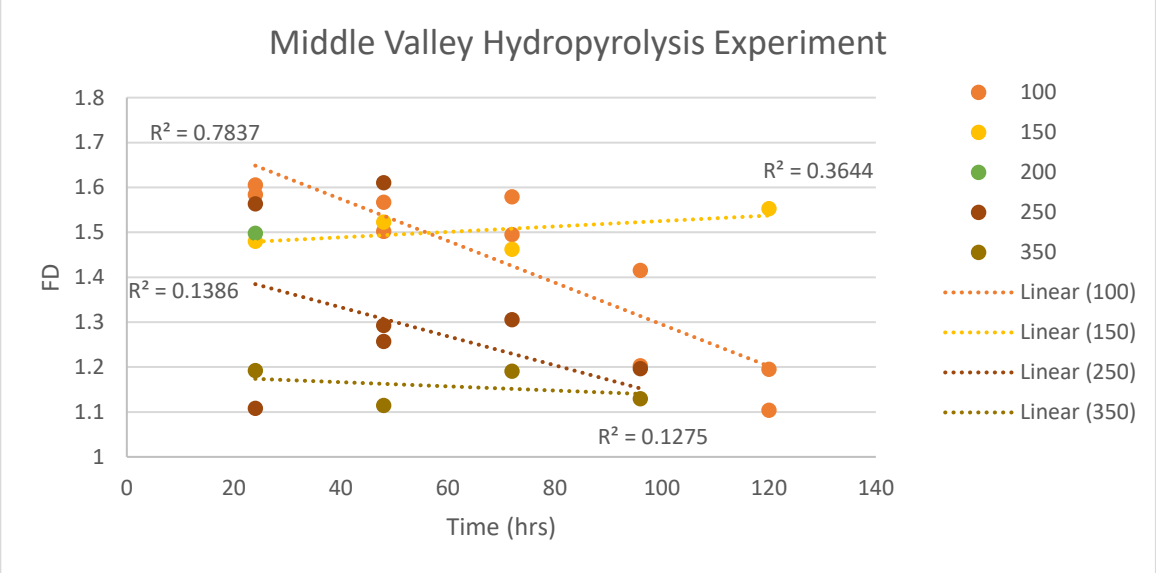


Figure 3.8 Cross-plot of FD vs time, as the Middle Valley sediments are pyrolyzed at different temperatures over different periods of time.

An important discovery in the Middle Valley experiment using binary analysis was that the way a chromatogram is created affects the FD result. Some chromatograms have high FD values, while their molecular complexities are not appreciably high. This was discovered to occur as a result of a fine-pixelated defect on the image. The granularity was identified

using CorelDraw by zooming the chromatogram to the point that one pixel occupied the entire screen (Figure 3.8). The granularity spread across all of some of the Middle Valley chromatograms. The granularity was not visible with the naked eye. It appears that the presence of granularity in the chromatographic image causes the FracLac program to assume that the whole picture was cover with peaks as though the molecular complexity of the sample was higher. This image artifact produces erroneous FD values.

Figure 3.9 GC×GC-FID chromatogram presenting granularity, the granularity in the chromatogram is caused by either a bad run in the GC×GC test or by interference of other elements during the test; granularity makes the fractal dimension increase.

Table 3.4: Result binary analysis of Escanaba Trough.

Escanaba Trough	FD	SD	Peak Number
S60	1.5471	0.0188	5,034
S61	1.6224	0.0293	6,344
S62	1.3592	0.0173	4,360

For Escanaba Trough, a set of only three samples was analyzed. The low number was due to the amount of image artifacts found on the locality's sample chromatograms. These three

samples did not display granularity and the FDs followed our hypothesis that the more complex the sample the higher it's FD.

4. Discussion

Fractal dimensions (geometry/mathematics) in chromatography has been applied successfully before; *Fractal Considerations in Chromatography* (Guillaume, Robert, Peyrin & Guinchard, 2000) uses the fractal geometry to study surface irregularities; articles such as *Fractal Chromatography: a new phase in separation science* (Edge, 2014) uses fractal geometry to analyze fractal particles identified by chromatography. Fractal dimensions in two-dimensional chromatography has never been applied before and this project is the first paper relating these two topics. However, the FDs gathered did not show specific trends as expected. This was likely due to issues caused by the preparation of the chromatograms. In figure 3.4 at the 8-12cm sediment depth for cores 5 and 6 we observed very low FDs. This was follow by an up turn in FD values until the bottom of the core. The observed pattern was the opposite scenario expected based on the number of peaks in each chromatogram. This was likely caused by an error in a lack of continuity of data transfer from one program to another. The visual display and export data values from ChromaToF are interpreted differently between the FracLac software program.

The ChromaToF software was not monochromatic. Subsequently, instead of going from light to dark with increasing peak intensity, the program generates a trend of light to dark

to light again. In this visualization scheme, the most abundant molecules in the sample have a light peak apex (peak center) producing a donut-shaped monochromatic image of a high-intensity peak. The donut-shaped peaks add false spatial complexity to the image (Figure 4.1).

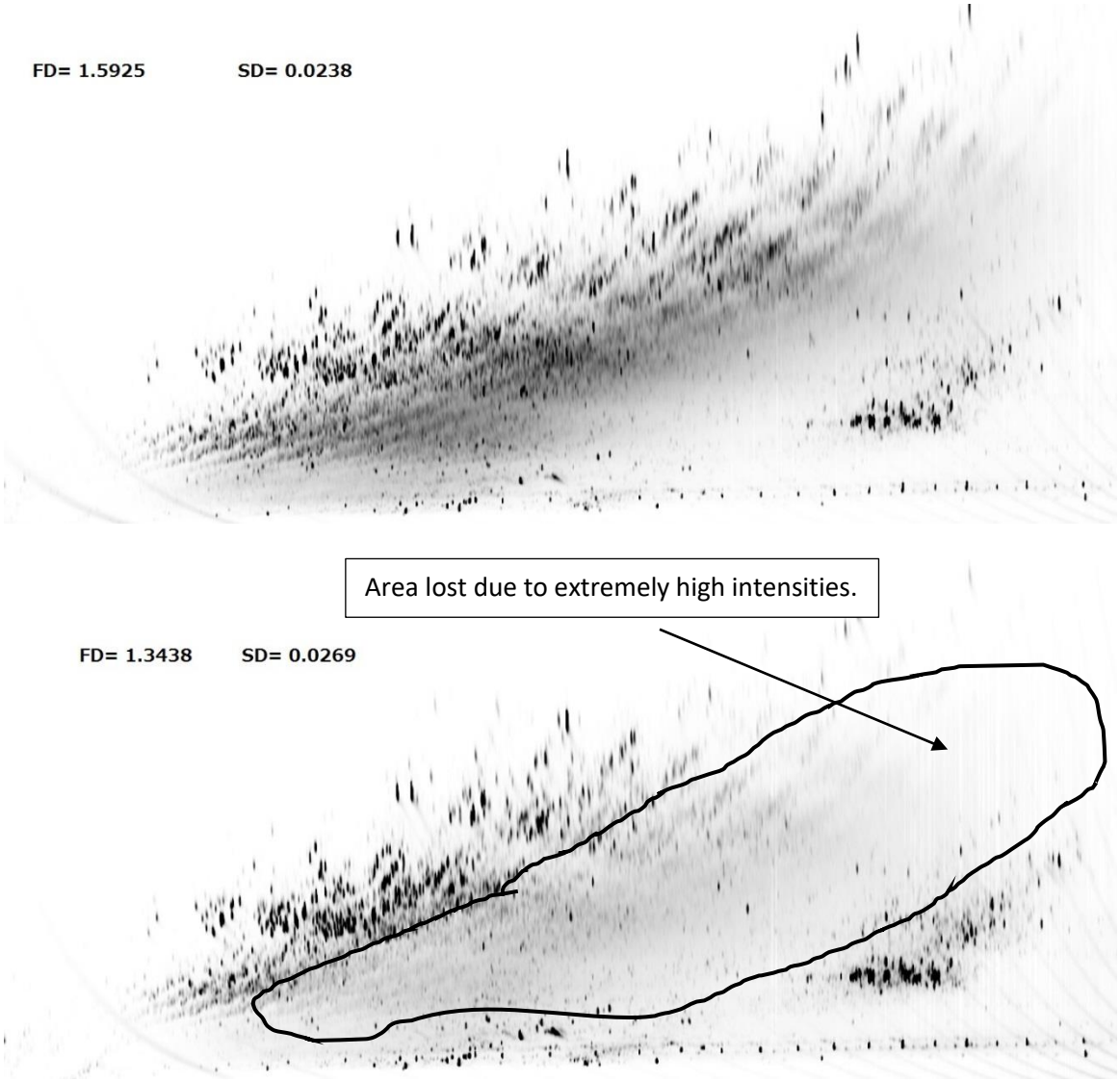


Figure 4.1 Top: GC×GC-FID chromatogram from core 6 at 8-10cm. Bottom: GC×GC-FID chromatogram from core 6 at 10-12cm, the bottom chromatogram should have more peaks than the chromatogram at the top, but due the visualization program the whole area looks white, making the FD go down to the point that it becomes the lowest in the entire core.

The lack of a trend between peak density and peak number with FD can be explained by excessive co-elution, chromatographic co-elution is a well-known phenomenon in chromatographic science and defined by the *Encyclopedia of Astrobiology* is "when two or more compounds do not chromatographically separate due to the fact that both species have retention times that differ by less than the resolution method" (Dworkin, 2011). Excessive co-elution results in concentrating high numbers of peaks into various clusters that together have the appearance of lower spatial complexity in the chromatographic image. The result is that elevated clustering will result in a lower FD as their will be more background (white space) left in the image. Alternatively, the presence of a complex mixture can also result in an overall rise in the 1^{st} and 2^{nd} dimension baseline, meaning that the amount of space becomes more uniformly dominated by one number. This signal may not be fully removed using standard baseline subtraction methods in the chromatographic software. The results of a large baseline artifact on the GC×GC chromatographic image is a patchier distribution that should result in a lower FD.

A potential resolution to this problem could be to produce a greyscale calculation, however, when this was done the image pixilation becomes complex at the baseline r value because the program's default decreases as small as 1 pixel. This means that if there are more than 2 peaks that overlap on a single pixel, FracLac will only count it as one, resulting in erroneous molecular complexity estimates. To evaluate this possibility, an experiment was conducted to minimalize the pixels per box parameter r (see equation 1). We expect the larger foreground patches can be analyzed together giving another perspective about how the algorithm functions. However, the results were only a decrease in FDs for all four experimental case (Table 4.1). The results again look more random than equally distributed, the change in pixels does not present a big difference in data gathering.

Table 4.1: Result of different minimum pixel per box.

Complexity	Sample	1 PIXEL	2 PIXELS	4 PIXELS	10 PIXELS	20 PIXELS	40 PIXELS
Low complexity	**S18**	0.9678	0.8738	0.8477	0.8520	0.8662	1.0337
	S19	0.8791	0.7868	0.7685	0.7485	0.7578	0.8075
Medium complexity	**S7**	1.1949	1.1385	1.1195	1.1118	1.1438	1.1790
	S17	1.2569	1.1869	1.1676	1.1733	1.2126	1.2319
High Complexity	**Core 8 (6-8cm)**	1.6852	1.6439	1.6280	1.6098	1.5967	1.5849
	Core 8 (10-12cm)	1.6750	1.6364	1.6203	1.6002	1.5904	1.5775

Under saturation equally causes an inadequate representation of an FD calculation because baselines and peak number decreases, this can be caused due to low injection concentration on post data processing of the images prior export.

FracLac also has a multifractal scanner. Multifractals defined by the programmer of FracLac's multifractal scanner © are fractal systems in which a single exponent D (fractal dimension) is not enough for describing its dynamics. Multifractals are common in nature. Imbedded in a multifractal system is lacunarity. Lacunarity is the space in between the patterns with its fractal distribution (Figure 4.2).

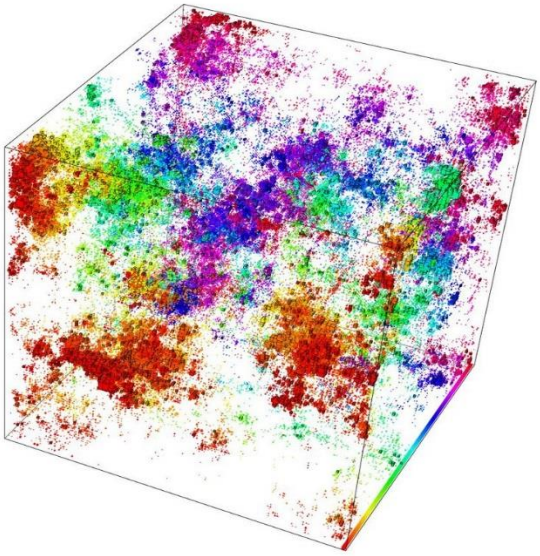


Figure 4.2 A multifractal electronic eigenstate (wave equations) at the Anderson localization transition system with 1367631 atoms (https://en.wikipedia.org/wiki/Multifractal_system).

The multifractal has been proposed in various fields like medicine, finance and geography. In geology the concept has been successfully applied to geophysics; in *scale, scaling and multifractals in geophysics* (Lovejoy & Schertzer, 2007), explains how fractals and multifractals have helped to develop our knowledge in geophysics helping scientist to organize chaos. *Multifractal measures specially for the geophysicist* (Mandelbrot, 1989) explains how multifractals and self-similarity are related to the distribution of rare minerals. It is likely that 3D GC×GC chromatograms could display mulitfactal qualities that may related to the way that molecules are distributed in oils.

5. Conclusions

The results from the three study areas produced inconclusive results. For the greyscale analyses, the results did not produce definitive predictable trends, which require other options to be explored. The binary scan proved more promising due to trends noticed in

figure 3.4 with the Guaymas Basin sample set. However, after comparing the FD against peak number and density it was realized that the method also has problems. The Middle Valley hydropyrolysis sample set showed reasonable results according to the FD and the complexity of the chromatograms, but other problems were creating errors in our chromatograms, the presence of granularity and the lack of specialization for FracLac to analyze chromatograms was the main problems here. Finally, the Escanaba Trough samples did not have granularity and the results followed our hypothesis that organic matrices with greater complexities will have higher FDs.

We have to conclude that further research must be done for the FD method to be a useful oil fingerprint technique. Specifically, a better result must be achieved relating different variables that affect complexity in oils. The ideal algorithm will create a 3D grid around not the JPEG chromatogram, but around the 3D chromatogram generated from the raw data coming from the two-dimensional chromatograph. This direct application will allow for co-eluted peaks to be counted and will minimize image artifacts common in the 2D JPEG files. Additionally, the multifractal extension of this technique should be further explored. As oil is a naturally occurring substance and is composed of molecules with multiple carbon atom linked chains, it is likely a multifractal distribution of oil forming compounds can be detected with the proper implementation of this technique. Such a direction may improve understandings of the distribution of molecules in naturally occurring substrates.

6. References

Abramoff, M.D., Magalhaes, P.J., & Ram, S.J. (2004). Image Processing with ImageJ. *BIOPHOTONICS INTERNATIONAL*, 11, 36-42.

Agterber, F.P., Cheng Q., Brown, A., & Good D., (1996). Multifractal modeling of fractures in the Lac du Bonet Batholith, Manitoba. *COMPUTERS AND GEOSCIENCE*.

Dallüge, J., Beens, J., & Brinkman, U.A.T. (2003). Comprehensive two-dimensional gas chromatography: a powerful and versatile analytical tool. *US NATIONAL LIBRARY OF MEDICINE NATIONAL INSITUTES OF HEALTH*, 1000, 69-108.

Dworkin J.P. (2011). Chromatographic Co-elution. *ENCYCLOPEDIA OF ASTROBIOLOGY*.

Yves C.G., Robert, J.F., Peyrin, E., & Guinchard, C. (2000). Fractal Considerations in Chromatography: Column Efficiency and the Multimicrocolumn System. *ASTROPHYSICAL JOURNAL*, 269, 239-249.

Karperien, A. (2012). FracLac for ImageJ. *CHARLES STURT UNIVERSITY, AUSTRALIA/CANADA*, https://imagej.nih.gov/ij/plugins/fraclac/FLHelp/Introduction.htm.

Lovejoy, S., & Schertzer, D. (2007). Scale, Scaling and Multifractal in Geophysics: Twenty Years On. *SPRINGERLINK, NONLINEAR DYNAMICS IN GEOSCIENCES,* 311-337.

Mandelbrot, B.B. (1974). Multifractal Measures Specially for Geophysicist. SPRINGERLINK, PURE AND APPLIED GEOPHYSICS, 131, 5-42.

Mandelbrot B.B. (1982) The fractal geometry of nature. *LIBRARY OF CONGRESS CATALOGING*.

Ong, R.C., & Marriott, P.J (2002). A review of basic concepts in comprehensive two-dimensional gas chromatography. *US NATIONAL LIBRARY OF MEDICINE NATIONAL INSITUTES OF HEALTH,* 40, 276-291.

Schindelin, J., Arganda, C.I., & Frise, E. et al. (2012). Fiji: an open-source platform for biological-image analysis. *NATURE METHODS*, 9, 676-682.

Schneider, C.A., Rasband, W.S., & Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *NATURE METHODS,* 9, 671-675.

De Souza, J., & Rostirolla, S.P. (2011). A fast MatLab program to estimate multifractal spectrum of multidimensional data: application to fractures. *COMPUTERS AND GEOSCIENCES,* 37, 241-249.

Sparkman, D., Penton, Z.E. & .Kitson, F.G. (2011) Gas Chromatography and Mass Spectrometry. *ACADEMIC PRESS*, Chapter 1: Introduction and History.

Springer, Berlin, Heidelberg, Rasband, W.S. (2012). ImageJ user guide, *U. S. NATIONAL INSTITUTE OF HEALTH*.

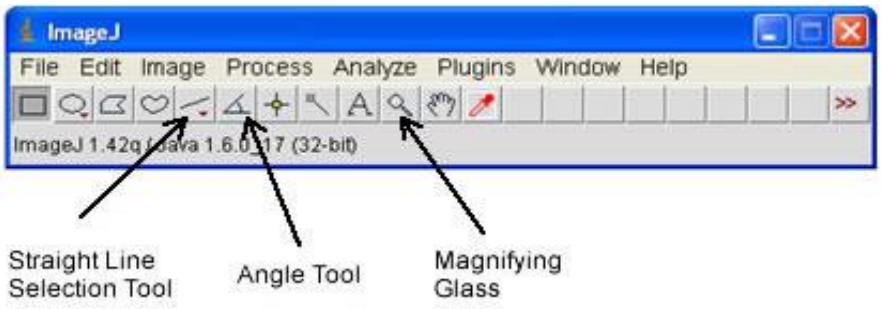
Zmeskal, O., et al. (2001). HarFA - Harmonic and Fractal Image Analysis, *FRACTAL ANALYSIS OF IMAGE STRUCTURES*, 3-5.

7. Appendix A – Porgramming FIJI and FracLac

1. Introduction to FIJI and FracLac

ImageJ is a public domain, Java-based image processing program developed at the National Institutes of Health.

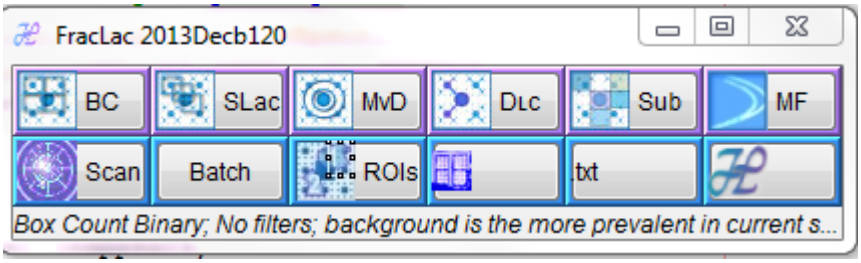
ImageJ was designed with an open architecture that provides extensibility via Java plugins and recordable macros. Custom acquisition, analysis and processing plugins can be developed using ImageJ's built-in editor and a Java compiler. User-written plugins make it possible to solve many image processing and analysis problems, from three-dimensional live-cell imaging to radiological image processing, multiple imaging system data comparisons to automated hematology systems. ImageJ's plugin architecture and built-in development environment has made it a popular platform for teaching image processing.



ImageJ is a software mostly used by biologists and doctors, but because biologist and doctors had previous thought about fractality in nature and human body, it makes this software the best choice for the purpose of this project. It comes equipped with a boxcount option, which is not very useful due to its limitations on bits and pixels per image, but which was very accurate when calculating fractal dimensions in different known geometric objects, a square, a circle, a line and a dot.

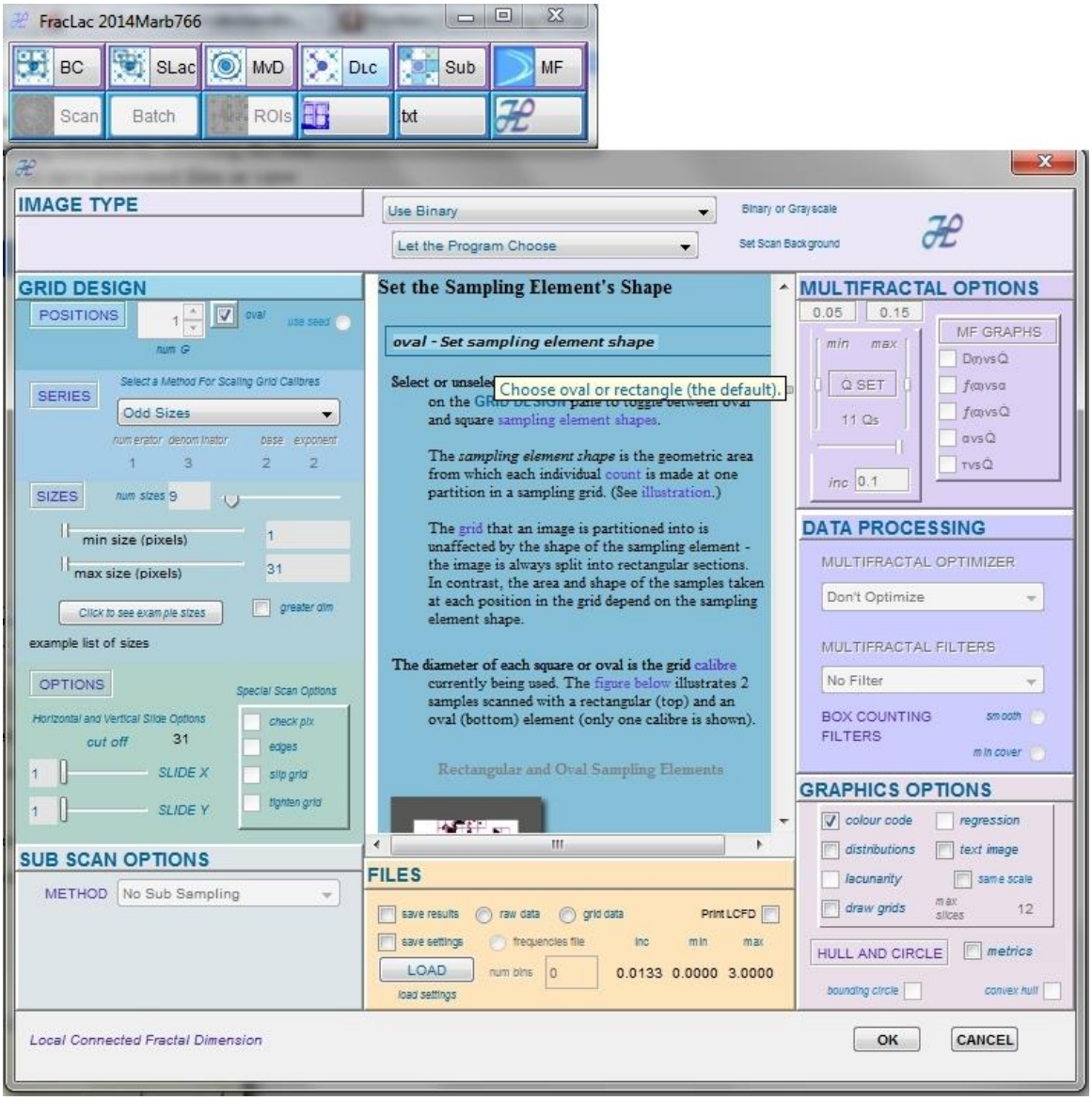
Some of the most useful features of ImageJ for this project, were the ability to create macros in order to re-do an operation multiple times, and the possibility to add plugins in order to achieve more specific goals; one of this plugins was FracLac which is a fractality-lacunarity calculator specialized for neuroscience.

FracLac software possess a much better boxcount option than the option at ImageJ, and provides a much better view of fractal dimension calculations. ImageJ was the program where this project was done and FracLac was the plugin that allowed all the calculations.



The basic box counting algorithm was originally modified from *ImageJ*'s box counting algorithm and H. Jelinek's *NIH Image* plugin, and was further elaborated based on extensive research and development. FracLac has more than one way to calculate fractal dimensions, 6 different options are present in the menu, boxcount, sliding boxcount, connected set, mass vs distance, particle analyzer, rectangular array and random scans, and finally the complex multifractal scan; it can also process more than one type of images, greyscale images, as well as binary images.

The image below shows the menu that will appear after selecting one of such methods from the FracLac plugin, is very complex due to the multiple things you can do when analyzing images here.



2. Chromatogram processing

Image processing has some specifics steps to follow before the boxcount scan can workout, *FracLac Sample 1*. Shows the whole configuration for the multiple scans the program can do; the multifractal options, data processing and sub scan options don't have any relevance for the purpose of this project.

i. ***Image type:*** Data was processed with two different methods, first by a greyscale differential scan, and then, by a binary scan, this is selected from the image type window.

1. ***Binary analysis:*** can only be done with binary images, if the image is not binary the auto convert to binary option has to be selected, this analysis will

give a fractal number between 1 and 2, because is assumed to be represented in a 2D plane.

2. ***Greyscale analysis:*** is more complex than binary, even though the data gathering method is the same the data measuring is not; here images exist in a pseudo 3D plane where a pixel is not always either on or off but somewhere else along a scale from 0 or white to 255 or dark black, this method provides a way to measure texture if we think in 3D and let the grey value be a proxy for volume, the result will give a fractal dimension between 2 and 3.

ii. ***Grid design:*** FracLac has multiple steps when configuring the grid design.

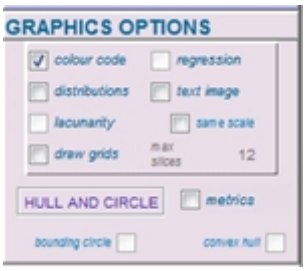
1. ***Positions:*** this option allows the boxcount scan to start from different positions, so no data is missing because the boxes didn't fully cover the whole image, then FracLac averages it and gets a fractal number for the image; for this option the max number of positions was set (12) in order to cover all different positions that could give erroneous numbers.
2. ***Series:*** this option allows the user to select different types of boxcount, including linear series, power series, scaled series, relative series, odd series and block series. For grayscale scans the block series is set as default by FracLac, this will scan a square block within an image using a series of grids calculated from the block size. In the other hand for binary power series was used; here the base is raised to the exponent added to itself to make successive sizes.
3. ***Sizes:*** The boxes will start from bigger sizes towards small ones, we can also configure how many pixels will have the starting boxes and how many pixels will have the end boxes. For the sake of this project a max number of pixels was equal to 45 and the smallest was equal to 1, the reason to do it this way was that 45 pixels was recommended by the programmers and 1 pixel was the minimum number of pixels allowed by FracLac.
4. ***Options:*** options were only available when selecting slide scans, which means that this option was not used.

iii. ***Files:*** This option allowed the user to get extra information on the data that was processed, for the project grid data was selected in order to see differences in FD when different starting positions of the boxes happened.

5. ***Graphic options:*** in graphic options two of the options were selected in order to know more about the program, the grid and the boxcounting algorithm.

5.1.Draw grids will create a GIF image showing the multiple boxes drawn through the picture and how they get smaller and smaller.

5.2.Regression will create the regression line showing the relationship between the log of box size against the log of caliber.



After all this process is, FracLac is ready to scan the previously edited chromatograms.