UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DATA-DRIVEN ANALYSIS OF HORIZONTAL WELL PERFORMANCE IN THE ANADARKO BASIN USING DIGITAL WELL LOG CLUSTERING TECHNIQUES AND SEQUENCE STRATIGRAPHY

A THESIS

SUBMITTED TO THE GRADUATE
FACULTY

in partial fulfillment of the requirements for

the Degree of

MASTER OF SCIENCE

By

JORDAN RENNER
Norman,
Oklahoma
2021

DATA-DRIVEN ANALYSIS OF HORIZONTAL WELL PERFORMANCE IN THE ANADARKO BASIN USING DIGITAL WELL LOG CLUSTERING TECHNIQUES AND SEQUENCE STRATIGRAPHY

A THESIS APPROVED FOR THE SCHOOL OF GEOSCIENCES

BY THE COMMITTEE
CONSISTING OF

Dr. John D. Pigott, Chair

Dr. Matthew J. Pranter

Dr. Heather Bedle

ACKNOWLEDGEMENTS

This research was funded through the generous support of ConocoPhillips, AAPG, SIPES Foundation, the Oklahoma Geological Foundation, and Cosmo Energy. I would like to thank Dr. John Pigott for his mentorship, guidance, and support. I would also like to thank Dr. Matthew Pranter and Dr. Heather Bedle for their input and support. I would like to thank my colleagues at Cosmo Energy for investing in me and this research. I would like to thank Jerry for assisting me in Petrel. Most of all I would like to thank God, my family and my friends for their never-ending patience, encouragement, and support throughout this process.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Mississippian reservoirs are among the most prolific oil and gas producing reservoirs in the mid-continent. Understanding regionally and locally where Mississippian sweet spots for petroleum exploitation are located is a key to fully exploiting their resources. Assuming a petroleum system exists, rock properties as determined from petrophysical logs are one of the most important factors in determining how a well will perform. As decline curves can be used to predict future performance of a well, based on direct production measurements, petrophysical log signatures should be related to predicting future well performance. Based on this premise, a novel approach using petrophysics and data science is tested to find and rank similar packages of rock to each other based on percent sameness using cosine similarity and K-means clustering. This study explores the methodology, practicality, and limitations of evaluating the relationship of basic petrophysical log signature packages to production volumes within the Anadarko Basin, as well as the greater ancient "Oklahoma Basin" using an integrated petrophysical, data science, and sequence stratigraphic approach. The workflow discussed herein is designed to search through a data set comprised of 25,673 petrophysical well logs in search of Gamma Ray signatures that are correlative with that of a designated Mississippian sweet spot from one given well. Not only is this method important in the search for petroleum, but it can also be applied to other commercial means of global resource exploitation.

INTRODUCTION

The Anadarko Basin of Oklahoma has had successful hydrocarbon exploration and production since the early 20th century (Brown, 2020), and still today there are many untapped resources for both conventional and unconventional drilling (Brown, 2020). The Anadarko Basin has also been the location of many technological advancements in oil and gas (Brown, 2020). One problem encountered while working in this basin that most certainly exists in all other petroleum basins, is the use of close proximity exploration in order to purchase acreage and drill new wells. Horizontal drilling may appear to be statistical, but that would only be true to the extent of the reservoir. That is, after a few good wells are completed in an area, the price to purchase acreage in adjoining sections suddenly increases, which can mean that companies can pay more for the riskiest assets. With the exception of Ball et al. (1991) Petroleum Geology of the Anadarko Basin and the USGS system assessment (2010), few studies have looked at this amount of data and tried to make a regional comparison of lithological units with the advent of contemporary unconventional exploration with respect to close proximity exploration. Presented herein is a novel approach to comparing reservoir electrofacies across the basin, in which log attribute analysis was performed, and then clustering based on sameness, and finally a comparison of well production profiles.

Previous studies have investigated the usefulness of machine learning in different aspects of geosciences. For example, with respect to geophysical data, Bougher (2016) argued that most geophysical data analysis was based on visual correlations and pattern recognition, which can be used to train a supervised learning framework. He argued that with the complexity of the Earth's sedimentary layers and with issues that often come up during seismic acquisition and processing, using unsupervised learning could prove to be more useful in extracting information directly from the data than using conventional physical models. With respect to geological exploration and exploitation, others have used machine learning to try and determine what the most important factors are for successful well performance. For example, when comparing the effects of different geologic and completion parameters, Luo et al. (2018) found that normalized volume of proppant, thickness, and structural depth of Middle Bakken Shale reservoirs proved to have the greatest effects on reservoir performance. Others have successfully used Artificial Neural Networks (ANN) to predict lithology, such as has been used in the area of investigation by Wethington (2017), Drummond (2018), Miller (2018), Hickman (2018), Miller (2019), and others. Of these, Wethington (2017), Drummond (2018) and Hickman (2018) also employed the use of Derivative Trend Analysis which can aid in the interpretation of high order relative sea level cycles (Appendix B-2). It is important to note when using Derivative Trend Analysis, that it is up to the interpreter to determine if the cycles are a result of lateral or vertical accommodation fill. That is, that if the cycles are local, they are likely horizontal, but if they are regional, they are likely vertical (Pigott and Bradley, 2014). For this study the focus will be aimed at a few of the most prolific oil and gas reservoirs in Oklahoma, primarily the Mississippian-aged Meramec, Osage, Sycamore Lime and Caney Shale. A big data approach is taken to try and answer some questions that have been posed from previous studies and determine the correlation between Anadarko Basin Meramec reservoirs and its laterally equivalent constituents in other locations of the Anadarko as well as Ardmore and possibly Arkoma Basins. A regional picture of the tectonic provinces of the study area is shown in Figure 1.

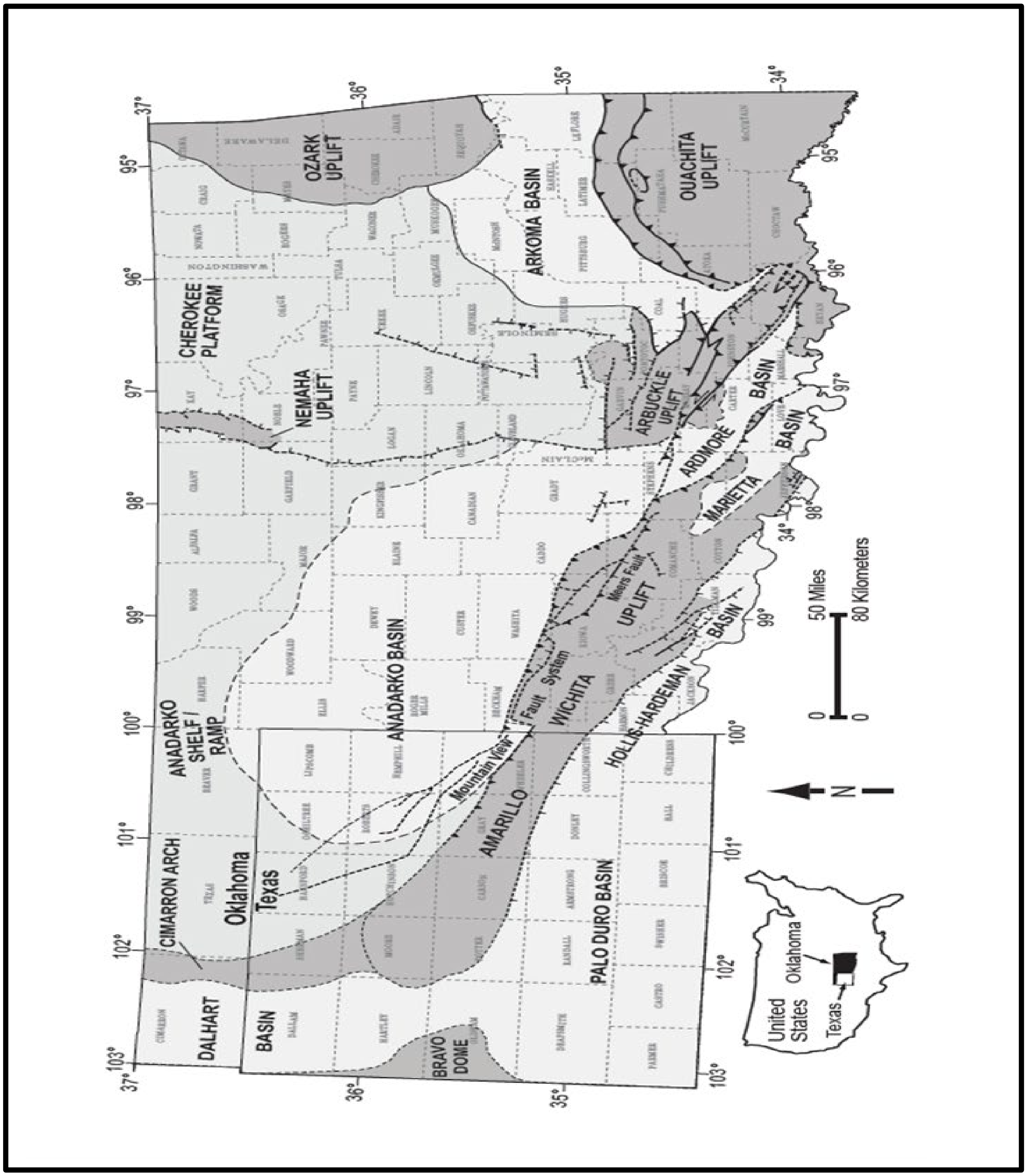

Figure 1: Regional map of tectonic provinces of Oklahoma and Texas panhandle. (from Miller (2019) and Miller et al. (2021); modified from Dutton, 1984; Campbell et al., 1988; McConnell 1989; Northcutt and Campbell, 1995; Johnson and Luza, 2008; LoCriccho, 2012)

Mississippian time in Oklahoma consisted of many cycles of high stands and low stands, with carbonates dominating during transgressions and siliciclastics dominating during regressions. The Meramec of the STACK (Sooner Trend of the Anadarko (Basin) in Canadian and Kingfisher (Counties)) consists primarily of carbonate and siliciclastic sediments that were deposited on a large shallow carbonate ramp and move laterally into a basinal setting (Price et al., 2017; Miller, 2019; Miller et al., 2021). Miller (2019) and Miller et al. (2021) identified five primary lithologies in his study using cored wells and applied a form of supervised machine learning for non-cored wells. He designated five lithologies: mudstone, argillaceous siltstone, argillaceous-calcareous siltstone, calcareous siltstone, and silty limestone. Other previous studies have found comparisons to the Meramec of the STACK in different locations in Oklahoma. Shelley (2016) and Shelley et al. (2019) found promising relation between the rock outcrops of the Pryor Quarry in northeast Oklahoma to the Shaffer 1-23 cored well in Blaine County, further showing how siliciclastics were widely deposited during low stands in Oklahoma in the Mississippian. Shelley (2016) also determined that the best Meramec reservoir quality in Oklahoma would be located at the base of the sequences, and that it could even be predicted through human and machine learning (ML) recognition of the Facies 1 signature from his study using a sequence stratigraphic approach. Miller and Cullen (2018) found that when looking at thin sections from Meramec siltstones in Blaine County and comparing them to Sycamore siltstones of the Arbuckle Mountains, there are no discernable differences.

This study takes the contributions of others and attempts to use a new approach to show how lithology is the dominant factor in horizontal well production results. Machine learning techniques are applied to analyze quantitative curve shape across the entire data set through comparison and cluster analysis. The geologic background work for these ideas was provided through previous studies (Wethington, 2017; Miller, 2018). The methods here are purely computational based well log and production analysis. This investigation asks three essential questions:

1. How can one optimize vast amounts of petrophysical data to delineate unconventional sweetspots?
2. Can machine learning be used to effectively compare lithology and classify log sequence stratigraphy across a basin?
3. Similarly, can machine learning incorporate vast amounts of data to predict well production?

GEOLOGIC SETTING

The North American paleogeography during the Mississippian and the eventual formation of the Anadarko Basin in Oklahoma were related to three principal tectonic events: the Acadian, Antler, and Proto-Oachita orogenies (Gutschick and Sandberg, 1983). The energy from these events gave way to the formation of the Transcontinental Arch which striked from northeast to southwest and divided the Madison Carbonate Ramp to its northwest and the Burlington ramp to its southeast (Appendix A-1).

The prominent oil and gas basins of Oklahoma were not defined to their current state until Pennsylvanian time (Johnson and Others, 1988). Throughout the Cambrian through late Mississippian, this entire region was a part of the ancient “Oklahoma Basin”, a shallow epicontinental sea in a tropical to subtropical environment (Figure 2; Appendix A-4) (Johnson and Others, 1988). The primary well known source rock for much of Oklahoma and specifically the Mississippian reservoirs of the study area is the Woodford Shale (Cardott and Lambert, 1985). The Woodford was deposited in euxinic seas during late Devonian to early Mississippian time (Cardott and Lambert, 1985). Mississippian time in ancient Oklahoma represented shallow well oxygenated seas and the deposition of fossiliferous limestones and downdip of those, siltstones and mudstones (Price et al., 2017). During the late Mississippian, subsidence of the Oklahoma aulacogen failed rift was followed by a collision of Gondwana and the North American plate (Johnson and Luza, 2008). More Orogenic activity during the Pennsylvanian compartmentalized the Oklahoma Basin into the distinct tectonic provinces of today. The Anadarko Basin where the test well is located is considered to be an asymmetrical foreland basin striking from the northwest to the south east (Johnson and Others, 1988). Spatial confinement of all Oklahoma basins are provided by the Wichita uplift to the west, the Arbuckle uplift to the south, the Nemaha ridge to the east, and the Ouachita uplift to the south east (Figure 3).

The primary target of this study is Mississippian reservoirs. During the Mississippian, there were four main depositional episodes: Kinderhookian, Oseagean, Meramecian, and Chesterian (Figure 4). The main focus will be on the Osagean and the Meramecian units at the top of the Kaskaskia sequence. Both the Osagean and Meramecian were deposited in mostly greenhouse conditions, but towards the upper Meramecian the transition to icehouse can be seen with higher frequency sea level changes (Appendix A-2). The Meramec is often characterized as being deposited on a gently sloping ramp with carbonate rich deposits proximally and mixed carbonate and siliciclastics basinward (Johnson and Others, 1988). One question that will be answered in the results of this study is what sort of relationship exists if any between Meramec, Osage, Sycamore Lime, and Caney Shale, chronostratigraphically and as reservoirs (Appendix A-3).

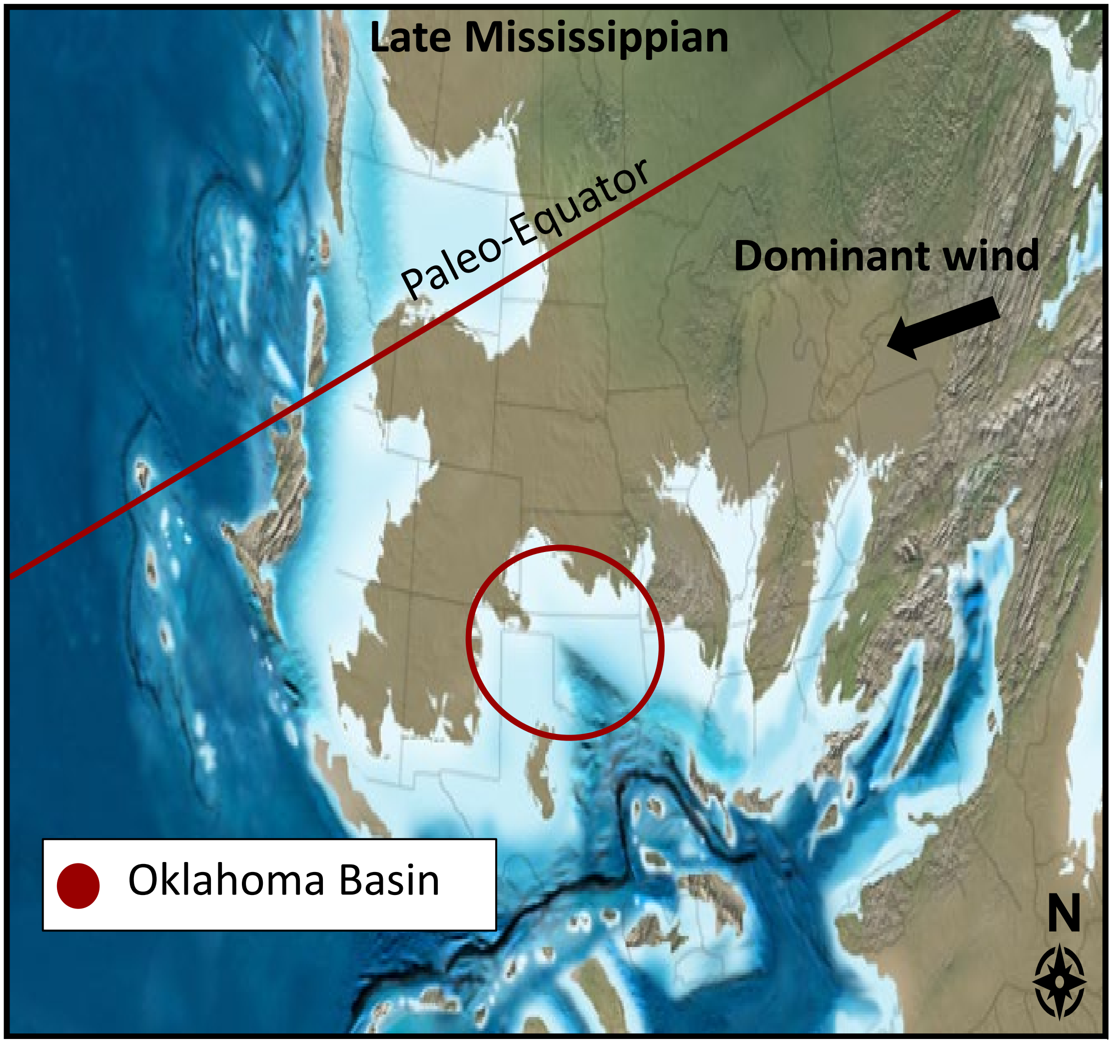

Figure 2: Paleogeographic representation of Late Mississippian (325 Ma) Oklahoma. Exposed land is shown in brown and green and the color blue of the water grades darker with increasing depth. Ancient Oklahoma Basin is outlined. Modified from Blakey (2011).

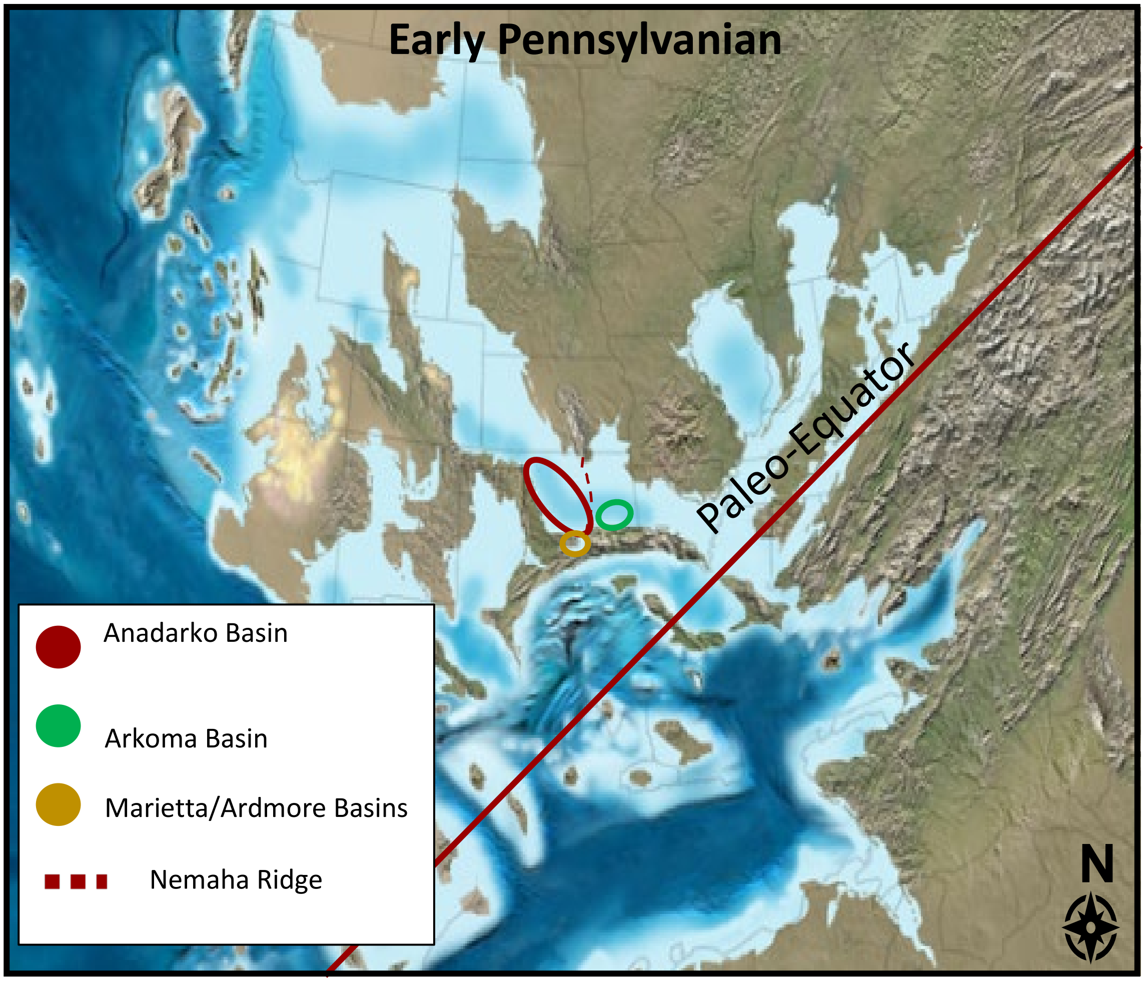

Figure 3: Paleogeographic representation of Early Pennsylvanian (315 Ma) Oklahoma. Exposed land is shown in brown and green and the color blue of the water grades darker with increasing depth. Orogenic activity during this time separated the Oklahoma Basin into the Anadarko, Arkoma, Marietta and Ardmore Basins of today. Modified from Blakey (2011).

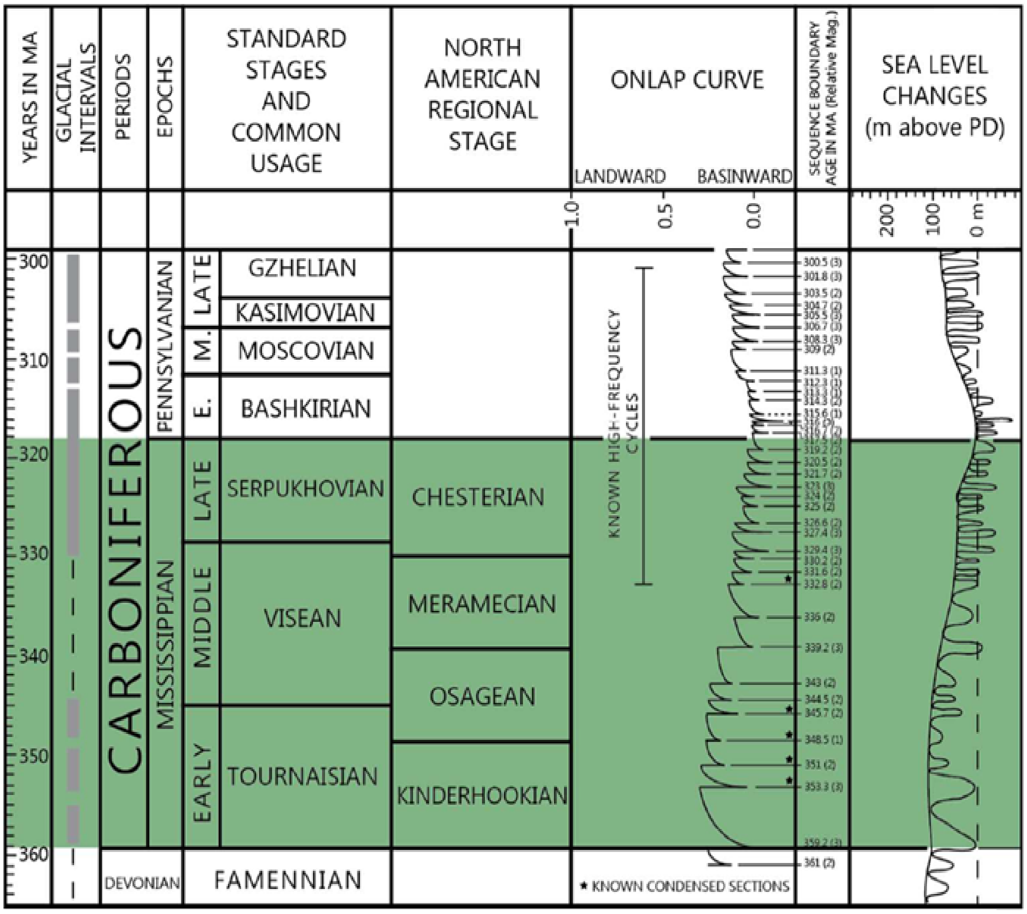

Figure 4: Stratigraphic, onlap and sea level curves for the Carboniferous Period. The Mississippian has been highlighted in green as it is the primary Epoch of study. Modified from Haq and Schutter, (2008).

PETROLEUM BACKGROUND

For this study, the primary type log is a Mississippian Meramec log signature package from the STACK play in Oklahoma. Miller (2018) and Miller et al. (2019) identified the prominent reservoir lithologies in a coarsening upward pattern to be primarily mudstones, argillaceous-calcareous siltstones, silty limestones, and fossiliferous sandstones. The maximum flooding surface caps the Lower Meramec, a retrogradational parasequence set (Figure 5)(Appendix B-1). The most favorable reservoir quality is found within the higher porosity and permeability argillaceous lithology units, primarily the parasequences that are directly above and below the maximum flooding surface. The Gulf Oil 1-25 Rohling used in the comparison is clastic dominated and would be located somewhere on the paleo ramp to basin transition. Previous studies have found lithostratigraphic correlation between the lower Meramec and upper Sycamore, as well as the upper Meramec with the lower Caney unit, though it is unclear where exactly this ramp to basin transition is regionally and stratigraphically (Miller (2018) and Miller et al. (2019)). According to Droege and Vick (2018), the Meramec is the highest producer and most commonly developed reservoir in the STACK, while the cherty Lower Osage formation in the western part of the STACK is estimated to have the highest amounts of oil-in-place volumes. In 2017, Devon Energy completed a Meramec well that produced at a peak rate of 6,000 BOE per day in Kingfisher County (Droege and Vick, 2018). This would be a new BOE peak rate record for the STACK play (Droege and Vick, 2018).

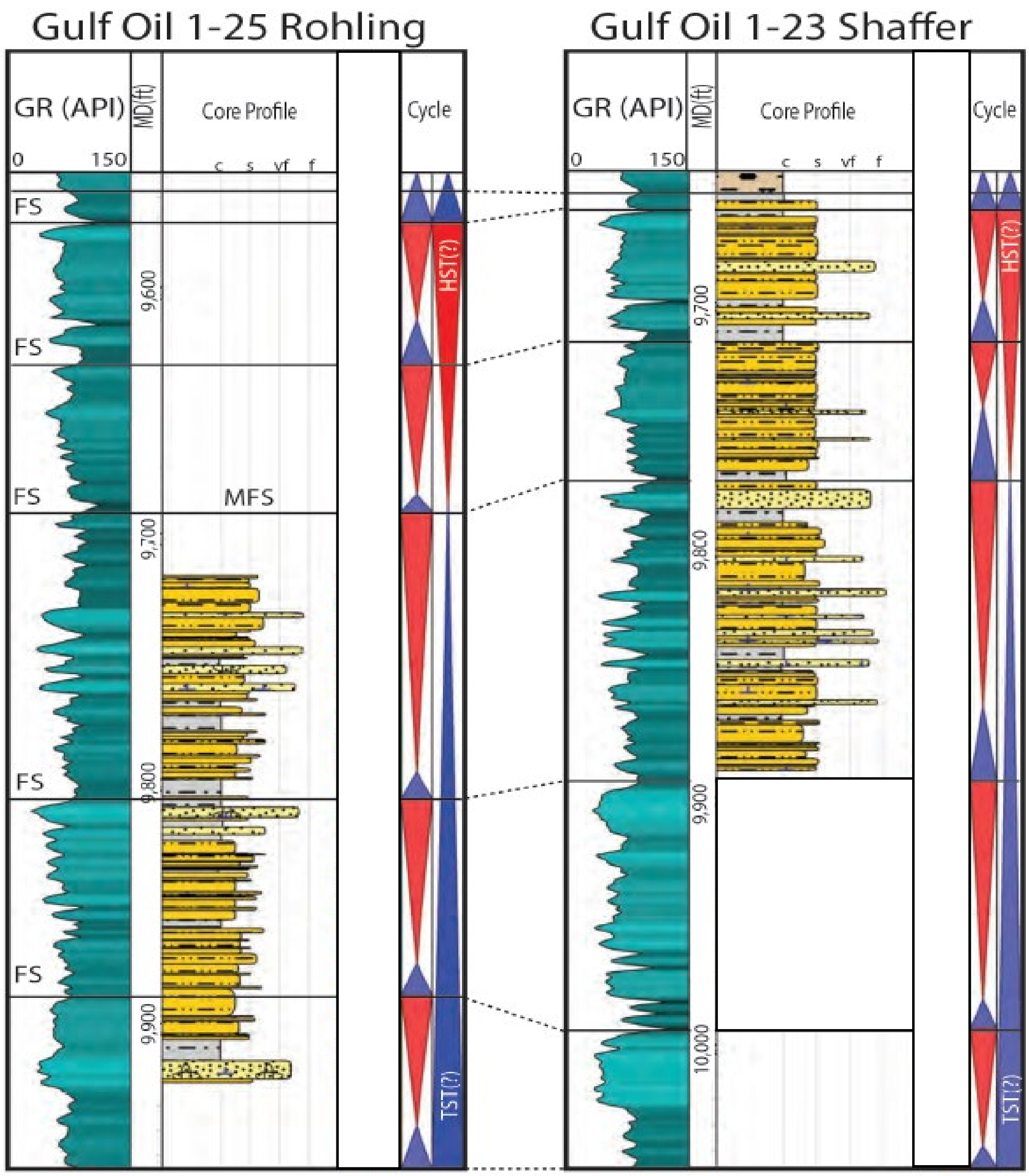

Figure 5: Well log responses alongside core descriptions for the Gulf Oil 1-25 Rohling and the Gulf Oil 1-23 Shaffer. Lower to middle Meramec can be seen in the Gulf Oil 1-25 Rohling and middle to upper Meramec can be seen in the Gulf Oil 1-23 Shaffer. The Meramec lithofacies from these wells was used as the prevailing lithology for the formation, and used to compare across the entire data set. The Meramec here has seven sequence stratigraphic zones separated by shaley flooding surfaces. The Maximum Flooding surface is seen between the retrogradational and progradational parasequence sets. Directly above and below the maximum flooding surface is where the optimal reservoir quality is found. Modified from Miller (2018) and Miller et al. (2019).

METHODS

Data Availability and Quality

With Oklahoma's long history of oil and gas exploration and production, a tremendous amount of data has been generated. The dataset includes 43,111 digital well logs from across the state of Oklahoma, spanning the Anadarko, Arkoma, Marietta and Ardmore Basins (Figure 6). These logs were provided by Cosmo Energy in Oklahoma City. With the goal of the project being to take a big data approach to some traditional and novel well log pattern recognition techniques, the LAS data set had to be normalized and transformed from the raw form into a more useful dataset.

The first step in order to be able to compare all of the logs together was the need for a tabular format database in which all well data could be stored. For this a free python package that is very well maintained and published by the Canadian Well Logging Society that can read and write LAS files called "Lasio" was used. Lasio is compliant with many types of LAS files even those with errors and non-traditional formatting (Inverarity, 2020). Post Lasio extraction it was imperative to breakout the tables into a "well_header" table, a "master_table" and a "well_digits" table to get a better grasp on the data and for simplicity of normalization and calculations. The data is stored in a MySQL database (Figure 7).

Normalization

Next the digital well log mnemonics and the digital well log digits were normalized. The curves that were chosen to retain are the primary lithology indicator logs which are Gamma Ray, Bulk Density, Neutron Density, Spontaneous Potential, Sonic, and Photoelectric Factor (Appendix C-1). According to Hancock (1992), these are the most useful of the logs that are normally recorded. Hancock (1992) also said that when determining lithology from logs, the best logs would be those that take into account rock properties over fluid properties, thus giving a more accurate measurement of the actual rock down the well bore. The mnemonic normalization was completed by determining the most common mnemonic unit that goes with each of the curve descriptions of interested (i.e. Gamma Ray and GR.GAPI). Next it was important to rename all of the descriptions that go with GR.GAPI to “Gamma Ray” and this step will be repeated for each of the key curves of interest. After completing the curve naming standardization, the database could now be filtered to only include the curves and values of interest. The well digits of the usable 25,673 well logs could then be normalized to be on a scale from 0-1 to ensure a level playing field before any further calculations or comparisons. For this a code from the sklearn preprocessing normalization module called “MinMaxScaler” was used. This coding allowed the data set to have the range be between 0 and 1 and scaled the data set appropriately (Pedregosa et al., 2011) (Appendix C-2).

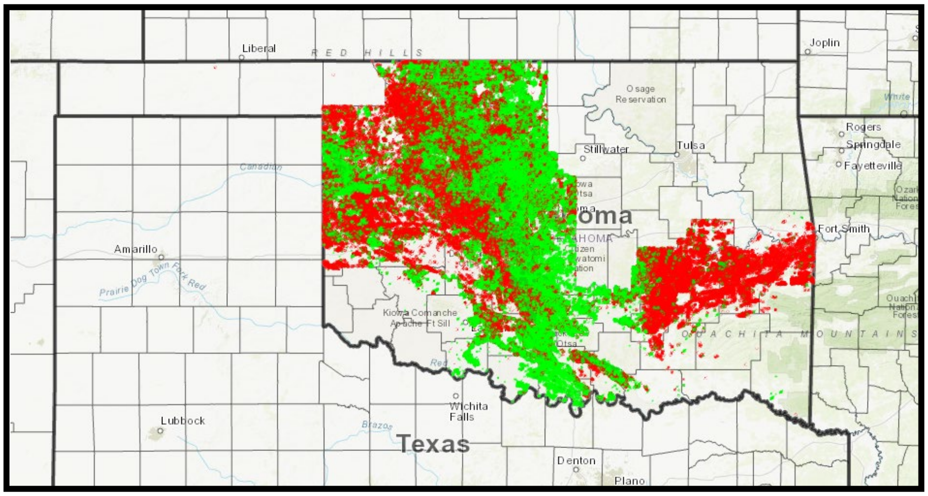

Figure 6: Well log data set comprised of 43,111 digital well logs spanning all major petroleum Basins of the state. Curves utilized are GR, Density, Sonic, SP, and PEF. Map from Drilling Info

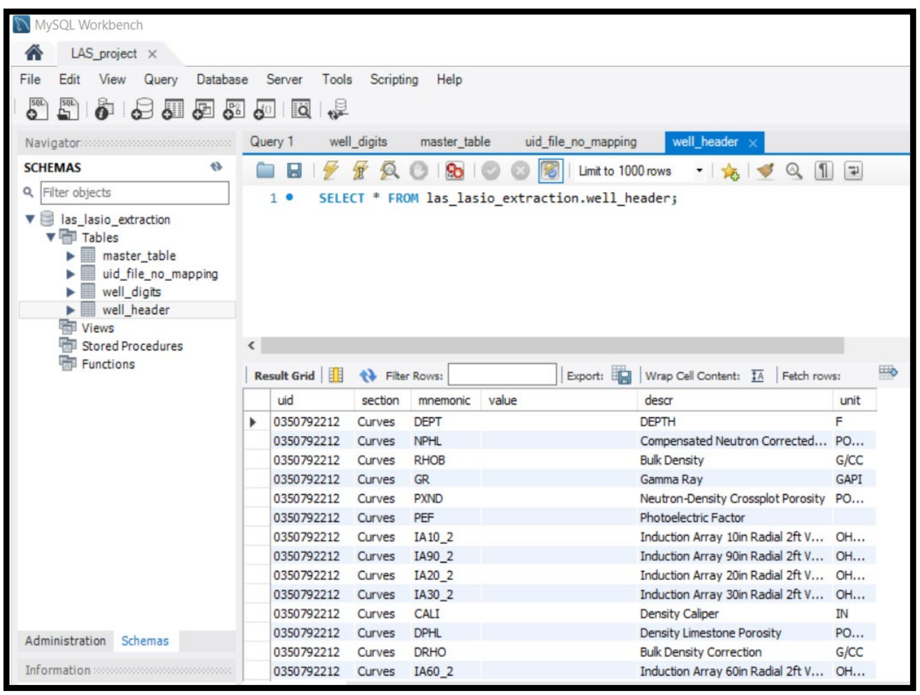

Figure 7: Example image taken from the MySQL data base where all log data and calculations are stored and viewed.

Quantitative Curve Shape

In order to compare sections of rock between different well bores, a relative curve shape based off of the change from point to point on a foot-by-foot basis needed to be created. This was completed using the previously calculated normalized well digit values that were scaled from 0-1 and finding the difference between each point and the point above it.

The main assumption when using this method is that the logging tools used in oil and gas exploration and production measure consistently down one hole, but should not be trusted to measure consistently across multiple holes of spatial variability in log properties. Therefore, curve shape must be generated based on relative values rather than absolute values so that an attempt can be made to compare signatures between wells. Figure 8 shows an example of a calculated curve shape.

GR
75.5278
73.9489
72.3023
71.4688
69.5753
66.0759
65.1409
68.8326
72.1582

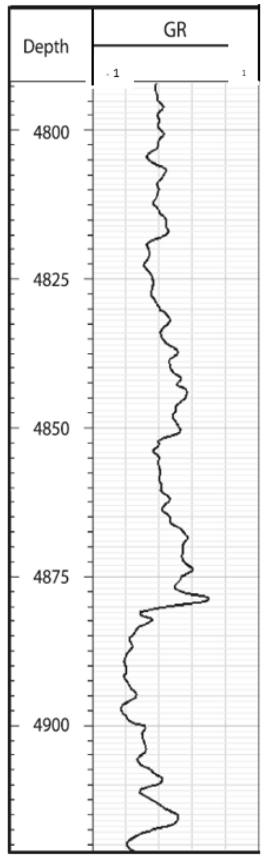

Figure 8: Relative quantitative curve shape generated by scaling original well digit data to the range of 0-1, and then taking the difference from point to point on a foot by foot basis and re-plotting the curves.

DTA Modeling in Python

Derivative trend analysis (DTA) is a novel form of log attribute analysis developed by Techlog of Schlumberger (Figure 9). It is designed similarly to other forms of attribute analysis such as those applied in seismic interpretation, with the goal being to enhance subtle information in hopes of being able to draw better geologic conclusions from the data. DTA can be incredibly useful in quick glance petrophysical analysis. When Geologists examine a cross section with the intent of interpretation, they are not simply looking at absolute measured values across wells. What they are looking for is trends, patterns, and shapes throughout the depth section. This is a similar approach that adds another dimension to the method above when it was discussed how the quantitative curve shape was created to compare relative values between different wells. DTA will place emphasis on the geologically significant curve shapes while drowning out the insignificant noise. This approach does not take into account quite the detail as the quantitative curve shape model, but what it does do greatly is aid in the interpretation of sequence stratigraphic cycles. These cycles can then be compared and correlated between multiple wells in an area, or wells with a high percent sameness score in different areas.

The first step in creating a Gamma Ray derivative curve is to create a smoothed curve based on a specified window of the original curve. Techlog's version of DTA uses a technique by Shapiro and Stockman (2000) in which a Gaussian smoothing function is applied to the data. Using their methodology, all values that fall within the specified window are weighted according to their distance from the original point. These points are recalculated as a weighted average of the surrounding points. What was found to work better when using python, is to apply this methodology using exponential weighted moving average (Pandas Dev team 2020)(Appendix C-3). Gaussian smoothing mentioned in the Shapiro and Stockman article was primarily designed for smoothing images, while exponential weighted moving average had much more literature in regards to usage in python and it achieved the same results. The goal of the weighted averages throughout the defined window is to smooth curves by calming high-frequency noise and preserving low frequency trends. The central difference equation is then applied to differentiate the smoothed curves (Wethington, 2017).

$$Derivative(i) = (Value(i+1) - Value(i-1)) / (Depth(i+1) - Depth(i-1))$$

The central difference method calculates the slope on a curve of neighboring points in order to determine the approximate derivative of the point of interest. The end results will show positive and negative derivative values. Positive values are indicating the original curve is decreasing upwards or shallowing upwards and becoming sandier. Negative values are indicating the original curve is increasing upwards or becoming shalier. Large positive or negative values are likely indicating a distinct change in lithology (Wethington, 2017).

Wethington (2017) also mentioned that improving this methodology to work on multiple wells in an area would greatly help with geological interpretation, and this is what has been done here in python.

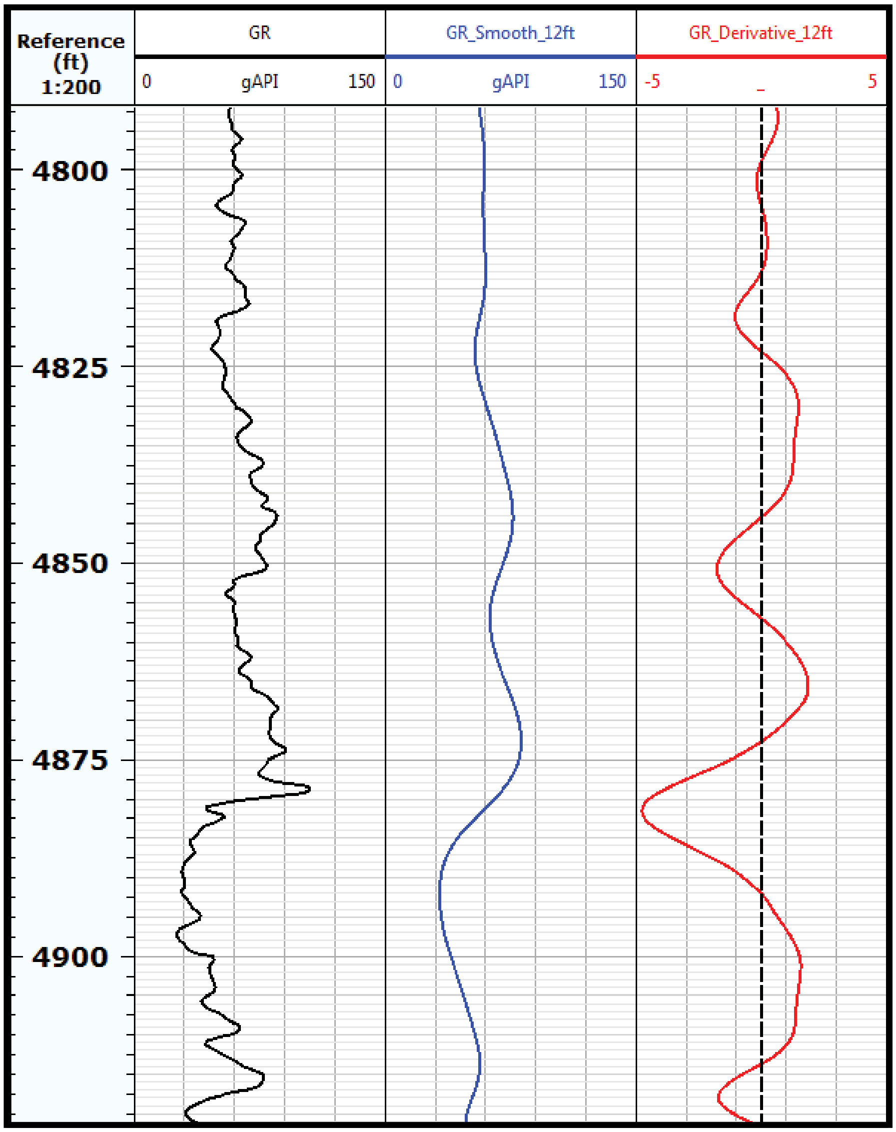

Figure 9: Here is an example of the 3 curves of interest in the Derivative Trend Analysis (DTA) study. The curves from left to right are as follows: Original Gamma Ray curve, smoothed GR curve, and the derivative of the smoothed GR curve. Note when the GR derivative curve is positive, the original GR curve is decreasing upwards, while the opposite is true for a negative GR derivative curve (Modified from Wethington, 2017).

Bipartite Petrophysical Process Energy Motifs

Sequence stratigraphy is one of the most common and most useful approaches used to infer how and when sediment accumulated in a basin. It can be an especially insightful tool to use when exploring for oil and gas and trying to determine the spatial extent of reservoirs. A sequence stratigraphic (allostratigraphic) approach is much more desirable than a basic lithostratigraphic approach especially in oil and gas exploration. An allostratigraphic method will provide a more holistic picture to the interpreter by taking into account different timing of sediment accumulation and accounting for erosional, non-depositional, and flooding surfaces (Robinson, 2014) (Appendix B-3). The two common approaches are the Galloway method and the Vail method. The Vail et al (1987) approach places sequence boundaries at the top of high stand systems tracts (HSTs) at unconformities where the shoreline is stationary at a high point, while the Galloway (1989) approach places sequence boundaries at maximum flooding surfaces which are at the top of transgressive systems tracts (TSTs)(Appendix B-1). Two other abbreviations are also used in the model. These are regressive systems tracts (RSTs) which categorize the falling stage of relative sea level, and lowstand systems tract (LSTs) where the shoreline is stationary at a low point (Pigott and Bradley, 2014). The best and most ideal way to use sequence stratigraphy in exploration geology is to use a combination of seismic, logs, core, and outcrop data to get a full understanding of the geology. The focus of this study is to determine how existing well log data can be used to draw more complete conclusions of the geology and why a well performed the way that it did. So in this case the Bipartite Petrophysical Process Energy Motif (2021) (Figure 10) adapting from the Galloway method will be used which has a higher vertical resolution but not much horizontal resolution. But with a large log dataset, lateral conclusions similarly to that of seismic can be drawn. Derivative Trend Analysis coupled with the Bipartite Petrophysical Process Energy Motif (2021) will help give better local and regional sequence stratigraphic interpretations for areas and formations of interest.

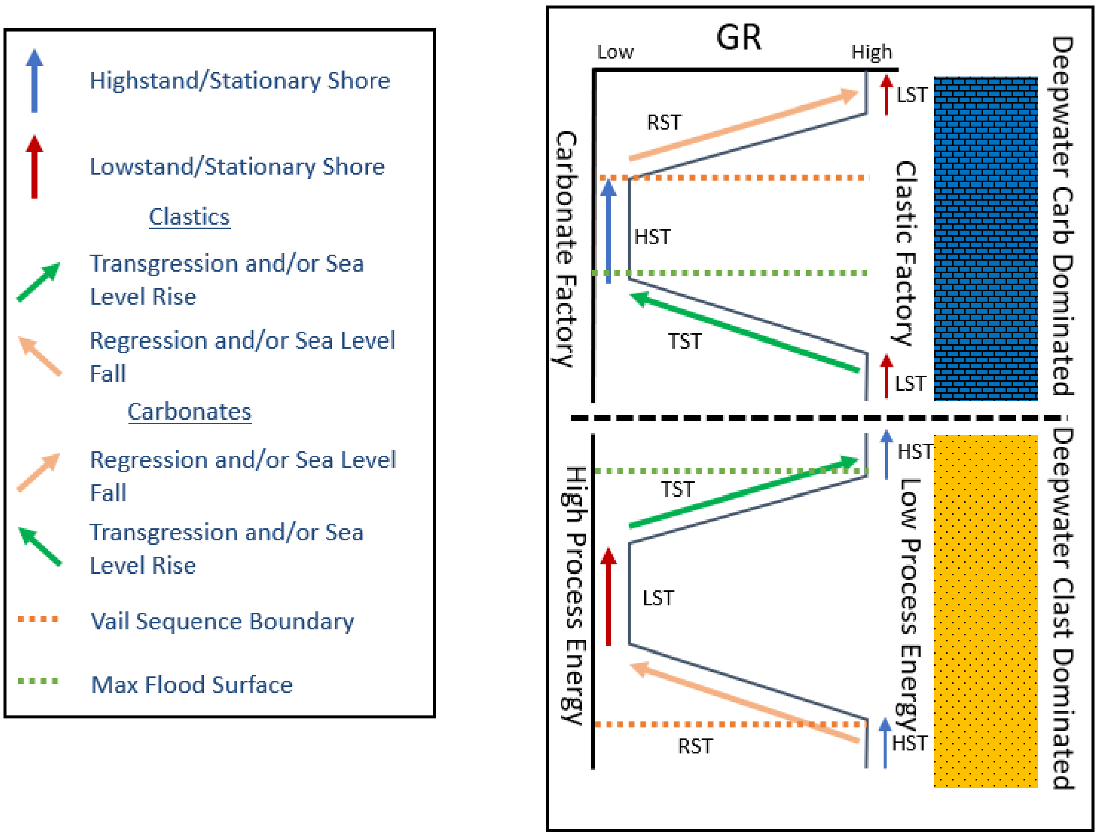

Figure 10: Modified from Pigott et al., 2021. Showing petrophysical method for categorizing GR log signatures with a sequence stratigraphic approach of determining basin accommodation changes through the use of Gamma Ray motifs.

Clustering Techniques

After the quantitative curve shapes were created, a technique was needed that could determine if the geologic section of interest in the Meramec was seen anywhere else in the data set regardless of location and to what degree of similarity. For this cosine similarity (Appendix C-3) was chosen. Cosine similarity appeared to be appropriate, as it was designed to be used with text formatted documents such as Word where it would count the number of words that any documents had in common. However, a common flaw with this technique would be that as the document's size increased the chances of it having many words in common regardless of the similarity of the topic with another document would be high. This problem would operate the same in geology, where most well logs are very likely to have multiple of the same points in common. However, the order of these points in which they occur determines if the words are in a sense telling the same story. When using this operation on a text document, even if 500 words are in common with another paper, if those words are not in a similar order, the documents are more than likely not talking about the same topic (Prabhakaran, 2020). The same logic applies to the application of such an analysis to petrophysical logs.

To use Cosine Similarity with the well logs, a list of vectors over 500ft intervals needed to be created, with stem of 200ft (0-500, 200-700, 400-900 etc). This vector list was created across all 25,673 files. For this study, the test was to use a specific 500ft section of file number 4384 in the data base that is known to be upper and lower Meramec from previous log and core studies, and compare the similarity of all other vectors across the 25,673 well data set to it. The results from the comparison analysis were then clustered by percent sameness using the most optimal method of K-means clustering and plotted on a chart in order to separate outliers. K-means is used to determine the smallest amount of squared distances from every point of interest to the nearest centroid that exists inside of each cluster (Hartigan, 1975). First a decision on an appropriate number of centroids to be dispersed throughout the data had to be made. After different statistical tests and comparisons, nine was chosen to be the optimal K. Once K was chosen, the rest of the data points are grouped in relation to the nearest centroid. Data points and centroids are then reassigned and recalculated until each data point has been placed closely to an appropriate centroid (Hartigan, 1975) (Appendix C-4).

RESULTS

Cluster Analysis

The main purpose of employing machine learning techniques to the data set was to determine if comparing well bore lithology signatures from one well to many others was a viable way to draw conclusions about what production curves may look like. Limitations will be discussed in a later section, but while preliminary results show some promising results, unfortunately many low correlation values also exist. A five-hundred-foot interval of the Gulf Oil 1-25 Rohling from 9500’ to 10000’ was used in the comparison as the Meramec model to compare against all other wells. The figures below show the three curves that were available for the 1-25 Rohling, and how those curve values compare with all 25,673 wells across five-hundred-foot intervals of the other wells, regardless of depth and location.

The three lithology indicating curves that were available to test from the Gulf 1-25 Rohling were Gamma Ray, Bulk Density, and Spontaneous Potential. When looking at all of the data, cluster group “max p” which was the maximum positive correlation group had an average %sameness across the board of 49%. This is including the Rohling which matched at 100% for all three curves. Separating out the Rohling, and Gamma Ray and Spontaneous Potential had the worst correlation of a maximum proportion for each of around 30% (Figures 11 and 12). Where as Bulk Density had a maximum proportion of 46% (Figure 13).

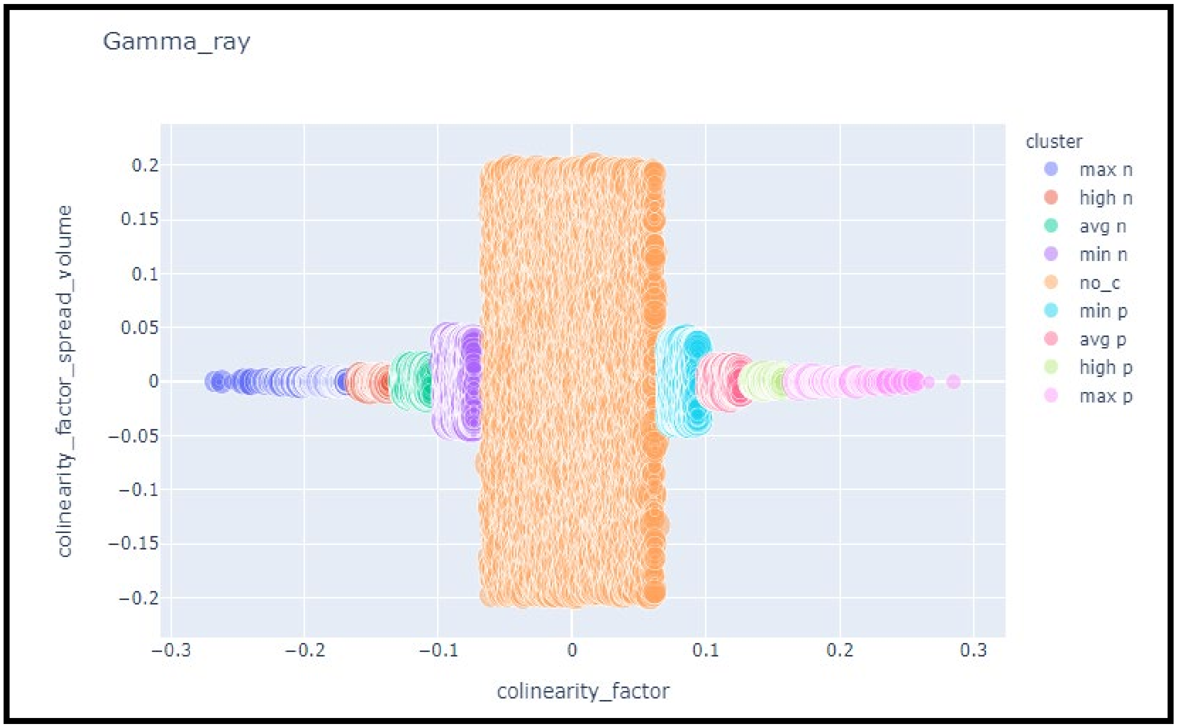

Figure 11: Chart showing %sameness of Cosine similarity of Gamma Ray on the 500' section of the Gulf Rohling well, compared against all 25,673 wells in the data set. Maximum correlation found for Gamma Ray with this method of K-means clustering was 29%.

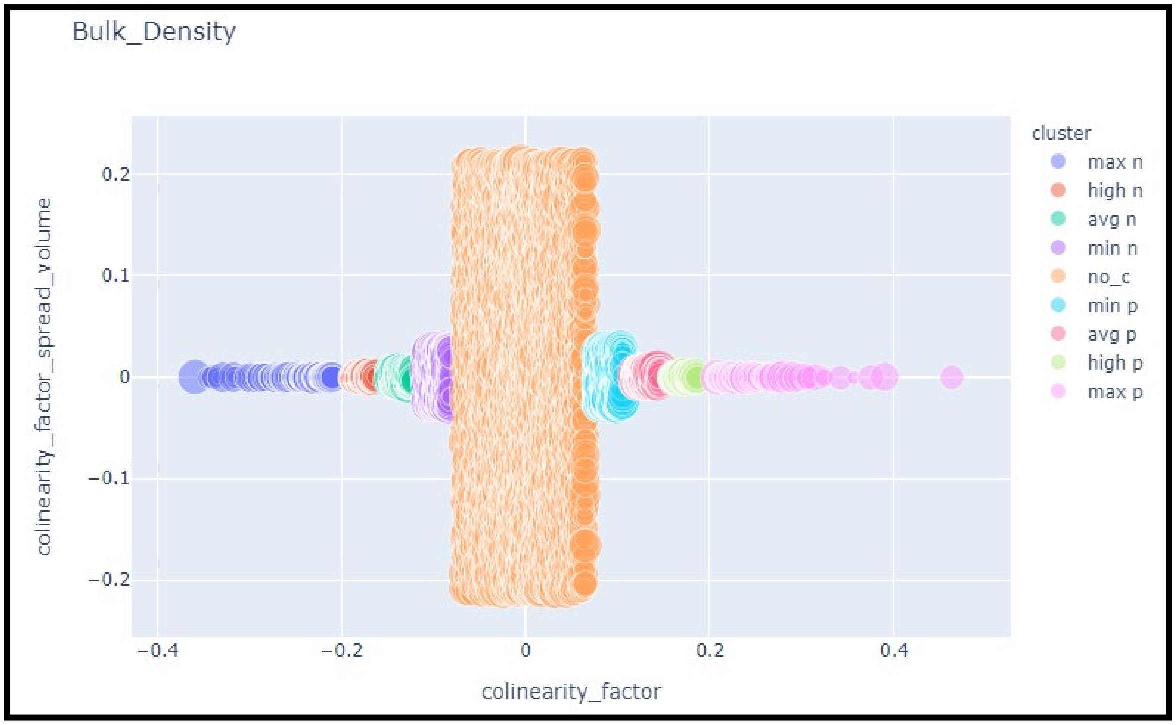

Figure 12: Chart showing %sameness of Cosine similarity of Bulk Density on the 500' section of the Gulf Rohling well, compared against all 25,673 wells in the data set. Maximum correlation found for Bulk Density with this method of K-means clustering was 46%.

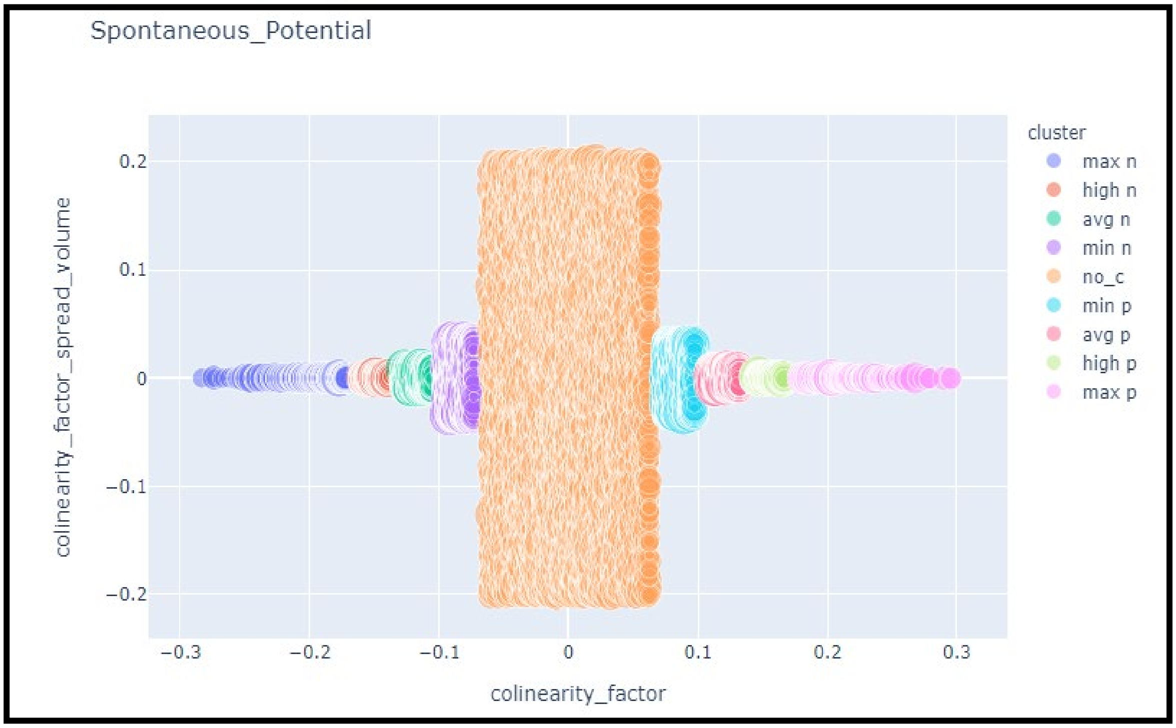

Figure 13: Chart showing %sameness of Cosine similarity of Spontaneous Potential on the 500' section of the Gulf Rohling well, compared against all 25,673 wells in the data set. Maximum correlation found for Spontaneous Potential with this method of K-means clustering was 30%.

Curve Shape Comparison

In order to compare well log values and curve shapes to each other, the well digits had to be normalized for the selected curves on a 0 to 1 scale. The next step was to find the change from point to point on a foot-by-foot basis. These are termed the "up-down-delta" curves. The up-down-delta curves allow everything to be put on a level playing field in order to compare relative curve shape between wells even if the well logs were recorded by different service companies. Relative curve shape was graphed out in python using plotly. This is a quick-look tool to compare what different wells look like that matched similarly.

The cosine similarity and K-means cluster analysis that was completed was based on the relative curve shape from the up-down-delta curves. With Gamma Ray having a maximum positive correlation of 29%, Bulk Density a maximum positive correlation of 46%, and Spontaneous Potential a maximum positive correlation of 30%. Figure 14 below shows how the up-down-delta relative curve shape that was generated captures the shape of the curve without dealing with absolute values. Figures 15 and 16 below show a comparison between the wells with maximum positive correlation of GR and RHOB, respectively, and that of the Rohling 1-25.

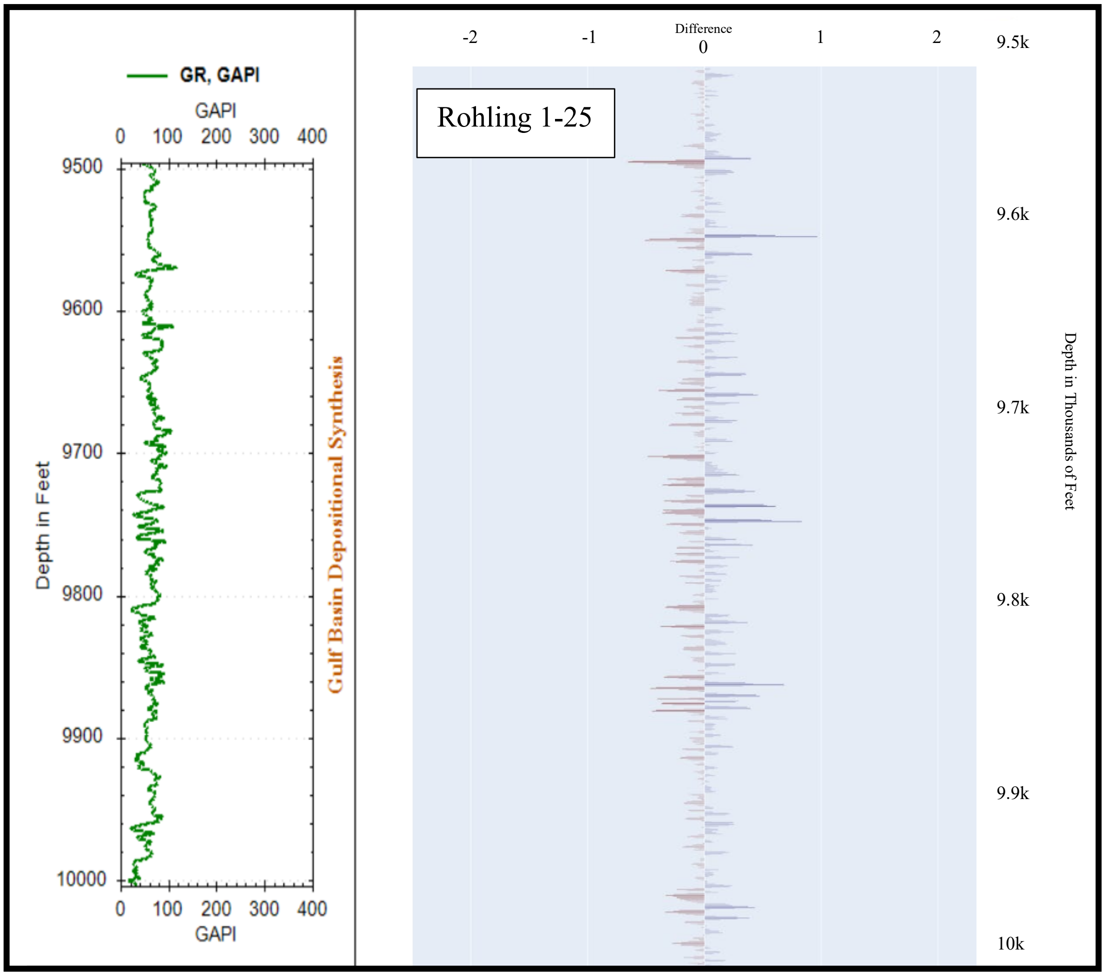

Figure 14: Plot showing original Gamma Ray curve for the Gulf Rohling 1-25 well on the left. On the right is the up-down-delta curve which is showing the relative shape of the curve based on the normalized data. The depth interval for both the original GR curve and the up-down-delta curve are both zoomed to show the 500' interval from 9500' to 10000' which is the Meramec model for the study. When comparing the two side by side, the major changes in shape are all captured by the up-down-delta curve with more emphasis given to the higher magnitude changes.

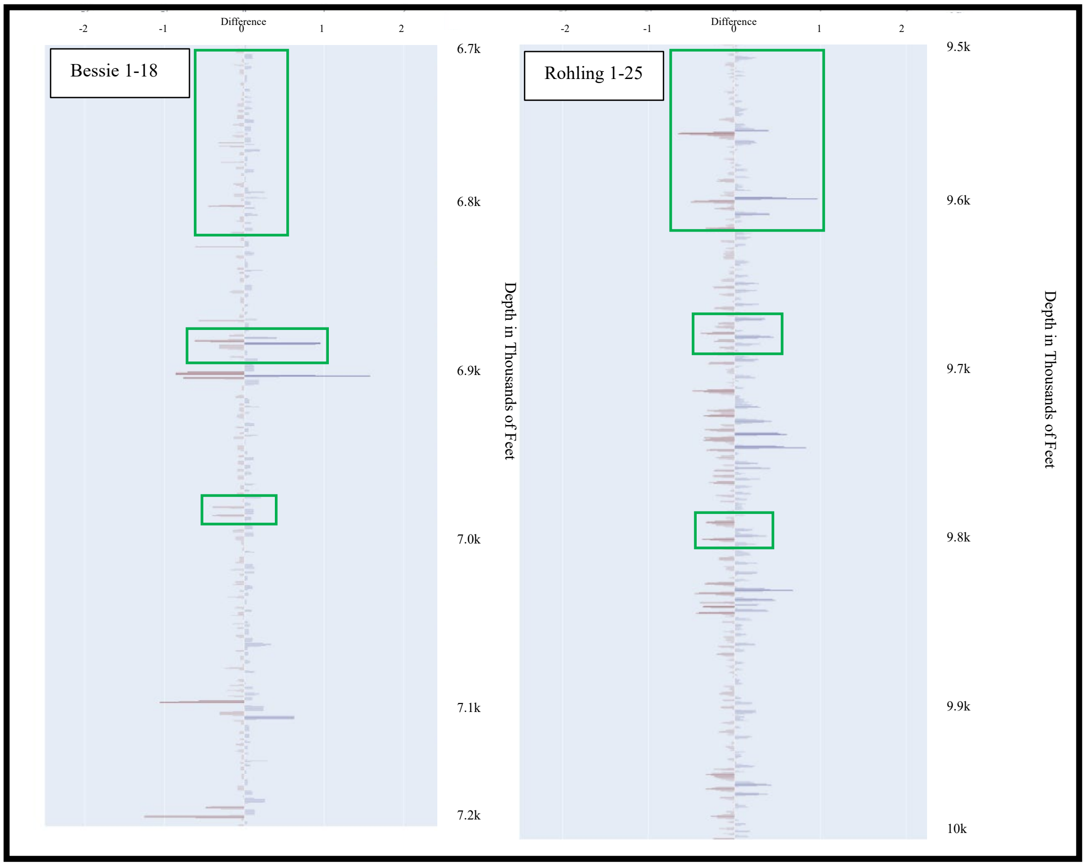

Figure 15: The plot on the right is the up-down-delta curve for Gamma Ray for the Rohling 1-25 which is showing the relative shape of the curve based on the normalized data. The plot on the left is for the Bessie 1-18 which matched at 29% sameness score. The depth interval for the Rohling 1-25 on the right is zoomed to show the 500' interval from 9500' to 10000' which is the Meramec model for the study. The Bessie 1-18 is zoomed in to 6700' to 7200' which is the depth window at which the highest match was found. When comparing the two side by side, the initial 140' of these wells are similar in shape (shown with the green boxes). Then moving down the log there are a couple other packages of sediment that also have similar characteristics.

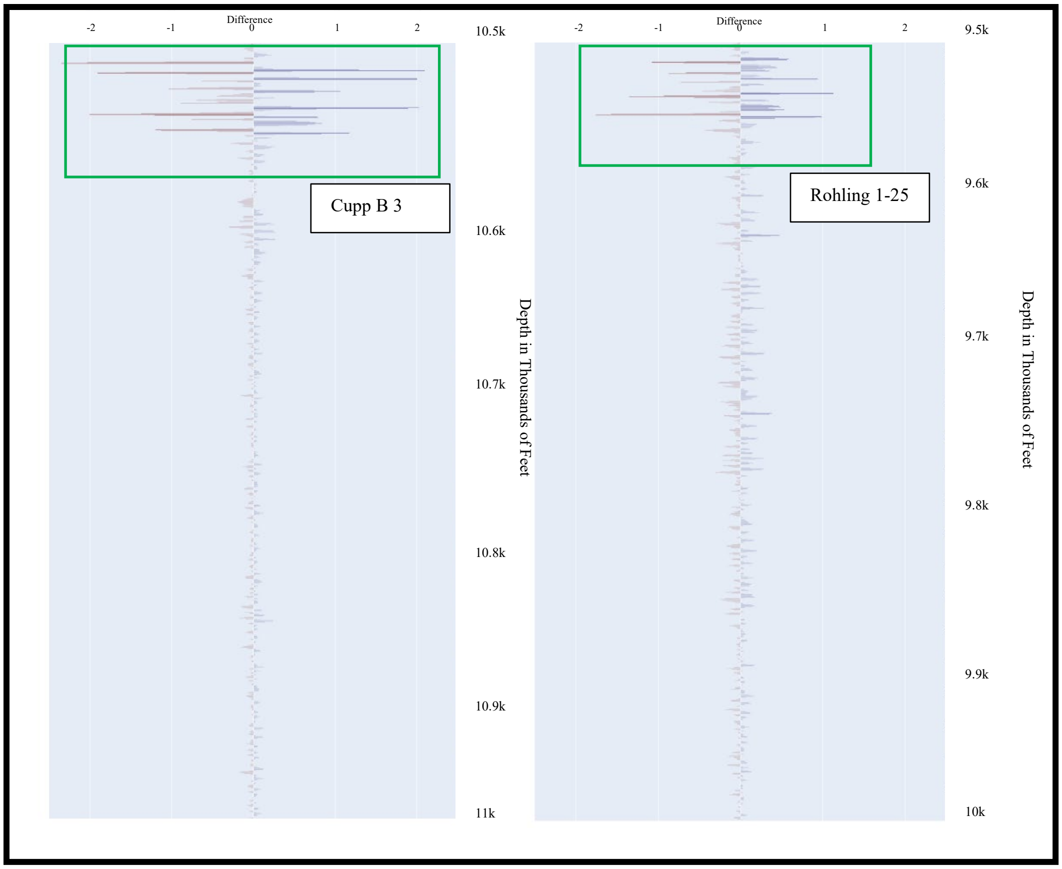

Figure 16: The plot on the right is the up-down-delta curve for Bulk Density for the Rohling 1-25 which is showing the relative shape of the curve based on the normalized data. The plot on the left is for the Cupp B 3 which matched at 46% sameness score. The depth interval for the Rohling 1-25 on the right is zoomed to show the 500' interval from 9500' to 10000' which is the Meramec model for the study. The Cupp B 3 is zoomed in to 10500' to 11000' which is the depth window at which the highest match was found. When comparing the two side by side, the initial 60' to 80' of these wells are similar in shape and both have a lot of movement (shown with the green boxes). Then moving down the log they have similar straight line characteristics, a likely indication that these wells are seeing similar lithology at these depths.

Log Attribute Analysis

Similarly to attribute analysis in seismic data, log attributes are used to attempt to enhance subtle information in hopes of being able to draw better geologic conclusions from the data. The attribute that was applied to the data is a modified novel approach to derivative trend analysis (DTA). To geologists, the absolute values of measured GR signatures may not mean anything, but the curve shape created from those values is what is useful for interpretation. DTA is used here to place emphasis on the geologically significant curve shapes while drowning out the insignificant noise. Exponential weighted moving average was applied over a 12ft window to create a smoothed GR derivative curve.

The results shown in Figure 17 indicate that positive values on the derivative curve translates to the original curve decreasing upwards or shallowing upwards and becoming sandier, while negative values are indicating the original curve is increasing upwards or becoming shalier. This has proven to be an effective tool to aid in interpreting sequence stratigraphic cycles, and even more granularity can be obtained when overlaying a Bipartite Petrophysical Process Energy Motif (2021) on to the derivative curve (Figure 18). When looking into the production volume analysis, the DTA Bipartite Petrophysical Process Energy Motifs (2021) of multiple wells can be compared alongside each other to determine if it is possible that similar basin evolution and depositional episodes were at work in different areas.

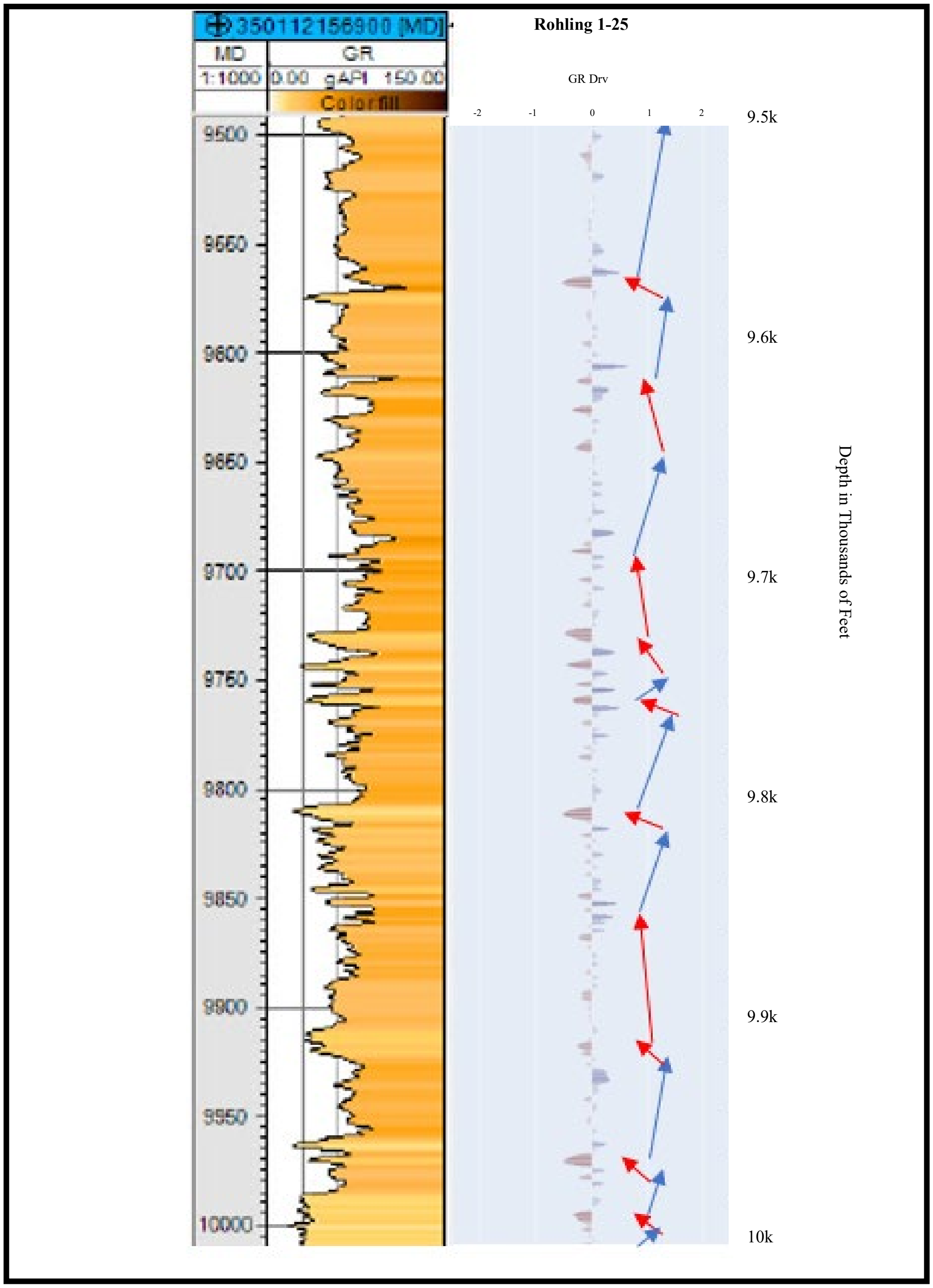

Figure 17: The first track on the left is the original Gamma Ray curve of the Rohling 1-25 exported from Petrel using a yellow to brown color scale with the lighter color being sandier, and darker shalier. The track on the right is the GR Derivative curve used to enhance geologically significant curve shapes while drowning out the insignificant noise. The increasing positive values from base to top are shown with blue arrows, and they indicate generally that the original curve is decreasing upwards and becoming sandier. The decreasing negative values are shown with red arrows, and they indicate the original curve is increasing upwards and becoming shalier.

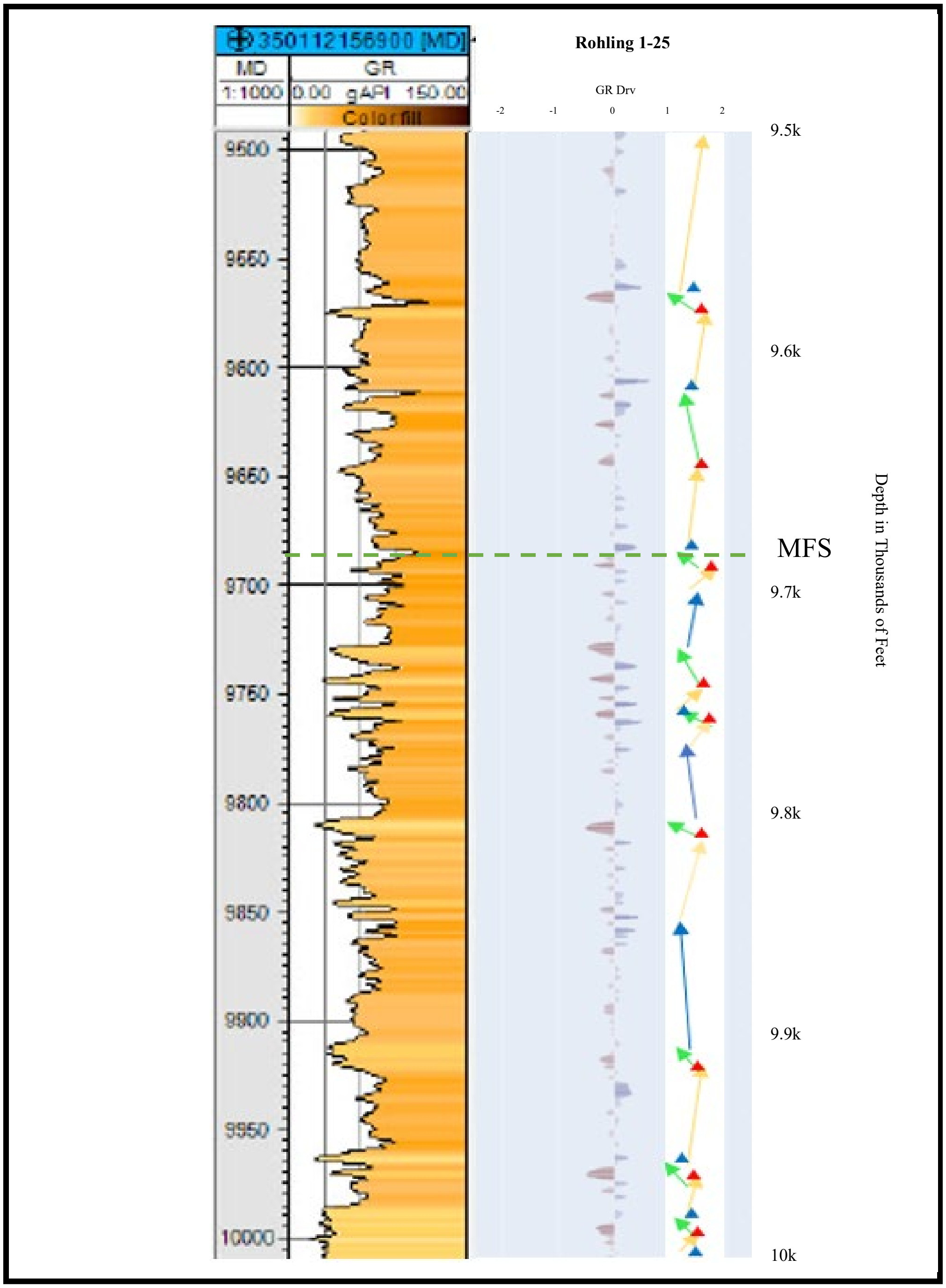

Figure 18: The first track on the left is the original Gamma Ray curve of the Rohling 1-25 exported from Petrel using a yellow to brown color scale with the lighter color being sandier, and darker shalier. The track on the right is the GR Derivative curve used to enhance geologically significant curve shapes while drowning out the insignificant noise. A Process Energy Motif (2021) has been overlain onto the GR Derivative curve in order to aid in sequence stratigraphic interpretation. Blue arrows represent HSTs, yellow RSTs, red LSTs, and green TSTs. The Meramec here appears to have seven sequence stratigraphic units capped by the maximum flooding surface illustrated with the dashed green line.

Production Analysis

The intent of this study was to create a new and practical tool to analyze and compare horizontal well performance across the state, regardless of location. While this tool is still in rudimentary form, it could still prove valuable to take a look at the production curves and target formations of the top matched wells. The wells that matched best to the benchmark, the Rohling 1-25, are shown on the map in figure 19. A few of them are regionally close by, but others are spread much farther away from the Rohling, and would be located in entirely different fields. The question to answer was could there be any relationship between the rock at these different well sites. Was it possible that when these reservoir units were deposited in the ancient Oklahoma Basin, that similar depositional events were happening simultaneously in different parts of the basin. Other previous studies have found lithostratigraphic and thin section correlation between the lower Meramec and upper Sycamore, as well as lithostratigraphic correlation of the upper Meramec with the lower Caney unit (Miller (2018) and Miller et al. (2019)).

The horizontal wells most closely associated with the Rohling 1-25 are the Oppel wells. The parent well was drilled by Continental in 2016 and had an IP of around 900 BOE. The increased densities were drilled by Ovintiv in 2019 and they came on at around 600 BOE. All of these wells are located in sections 24 and 25 of 16N-10W Blaine County Oklahoma and were completing the Mississippian Meramec with roughly 7000ft laterals and perforating in the depth range of the lower Meramec from 9700ft to 10000ft vertical depth (Figure 20).

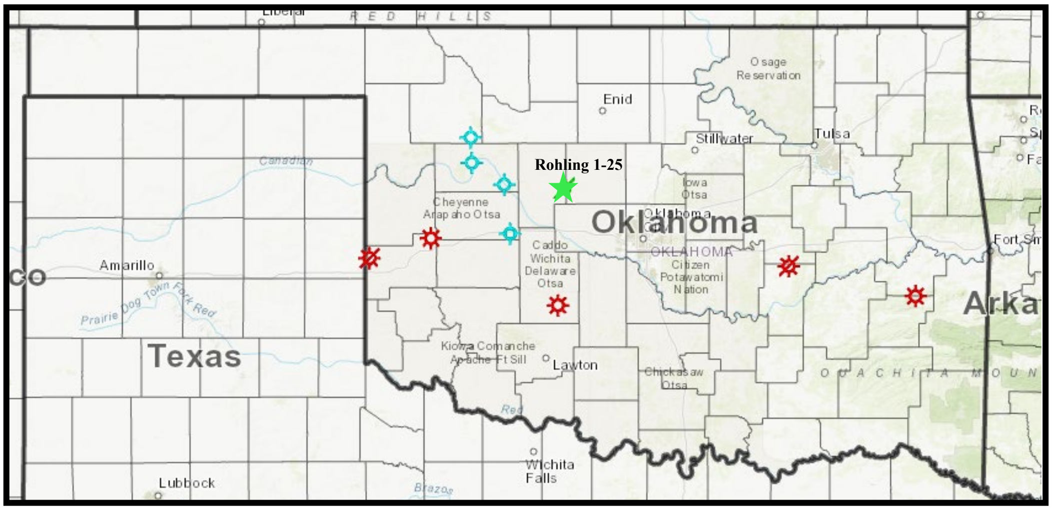

Figure 19: A map of Oklahoma obtained from Drilling Info showing the locations of the Meramec marker well, the Rohling 1-25, and the wells which had lithology indicator logs with the highest match percentage in the cluster analysis when searching for Mississippian Meramec signatures.

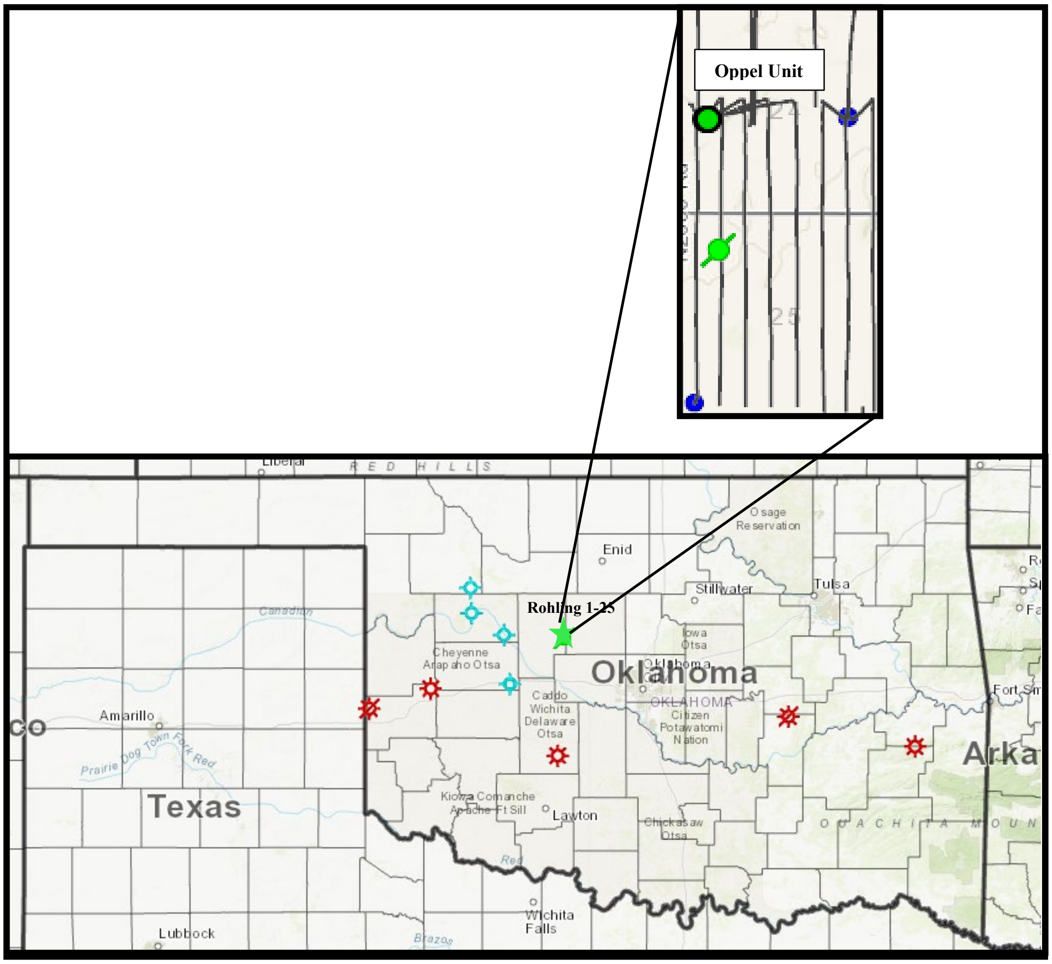

Figure 20: A map of Oklahoma obtained from Drilling Info showing the locations of the Meramec marker well, the Rohling 1-25, the wells which had lithology indicator logs with the highest match percentage in the cluster analysis, and the inset map shows the zoomed in location of the Ovintiv Oppel wells.

The first matched well to investigate is the Bessie 1-18 located in Woodward County Oklahoma. This well was only a 29% match with the benchmark of the Rohling 1-25, but when comparing the GR derivative curve with the Bipartite Petrophysical Process Energy Motif (2021) there are a lot of comparisons that can be seen (Figure 21). The horizontal that is most closely associated with the Bessie is the Johnson 24 1H, which was drilled by Tapstone Energy in 2018 and had an IP of around 800 BOE (Figure 22). Tapstone's target formation for this well was Mississippian, which from the data available it is not known which reservoir exactly in the Miss, but based on the Bipartite Petrophysical Process Energy Motif (2021) and the production curves, it looks like it very well could be Meramec or one of its lateral constituents. Looking at the production curves for these two wells, a conclusion can be drawn that reservoir properties must be reasonably similar, being that both have a similar IP and similar decline rate (Figure 23) (Pigott and Bradley, 2014). The Mississippian reservoir formation here was seen at a shallower depth than that of the Oppel's in Blaine County, but based on what is known about the Anadarko, the Oppel's are located down dip of the Johnson, thus the sediments that the Johnson is tapping into were deposited farther up on the low relief ramp setting.

The next well to look into is the Thompson 2-11 located in Hughes County Oklahoma. The relative curve shape of this well matched with the benchmark at a meager 27%, but again, there are many similarities that can be seen when comparing the production curves as well as the GR derivative curve with the Bipartite Petrophysical Process Energy Motif (2021) overlain

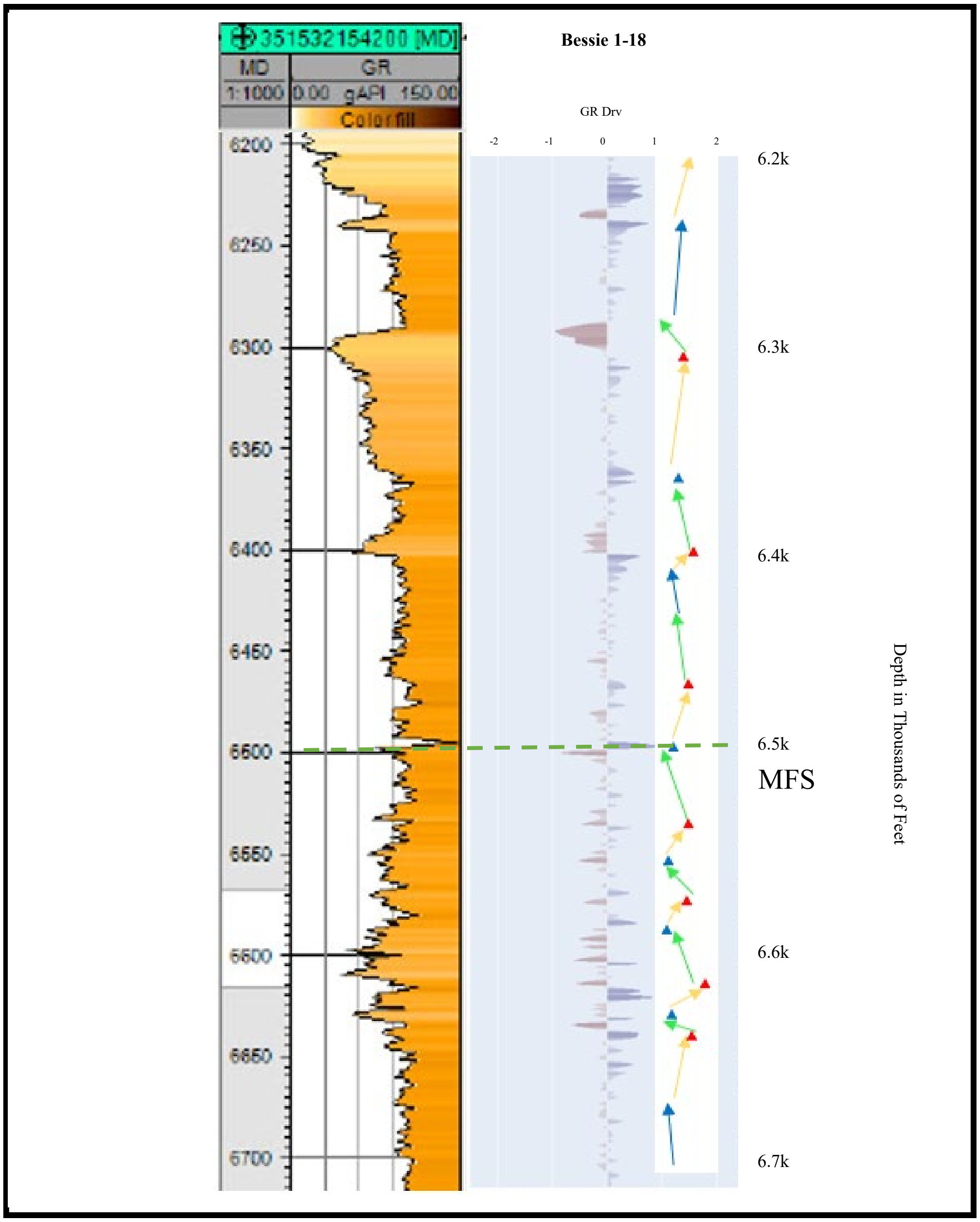

Figure 21: The first track on the left is the original Gamma Ray curve of the Bessie 1-18 exported from Petrel using a yellow to brown color scale with the lighter color being sandier, and darker shalier. The track on the right is the GR Derivative curve used to enhance geologically significant curve shapes while drowning out the insignificant noise. A Process Energy Motif (2021) has been overlain onto the GR Derivative curve in order to aid in sequence stratigraphic interpretation. Blue arrows represent HSTs, yellow RSTs, red LSTs, and green TSTs. Comparing with figure 18, there are very similar depositional energies at work and this very well could be the Meramec or one of its lateral constituents.

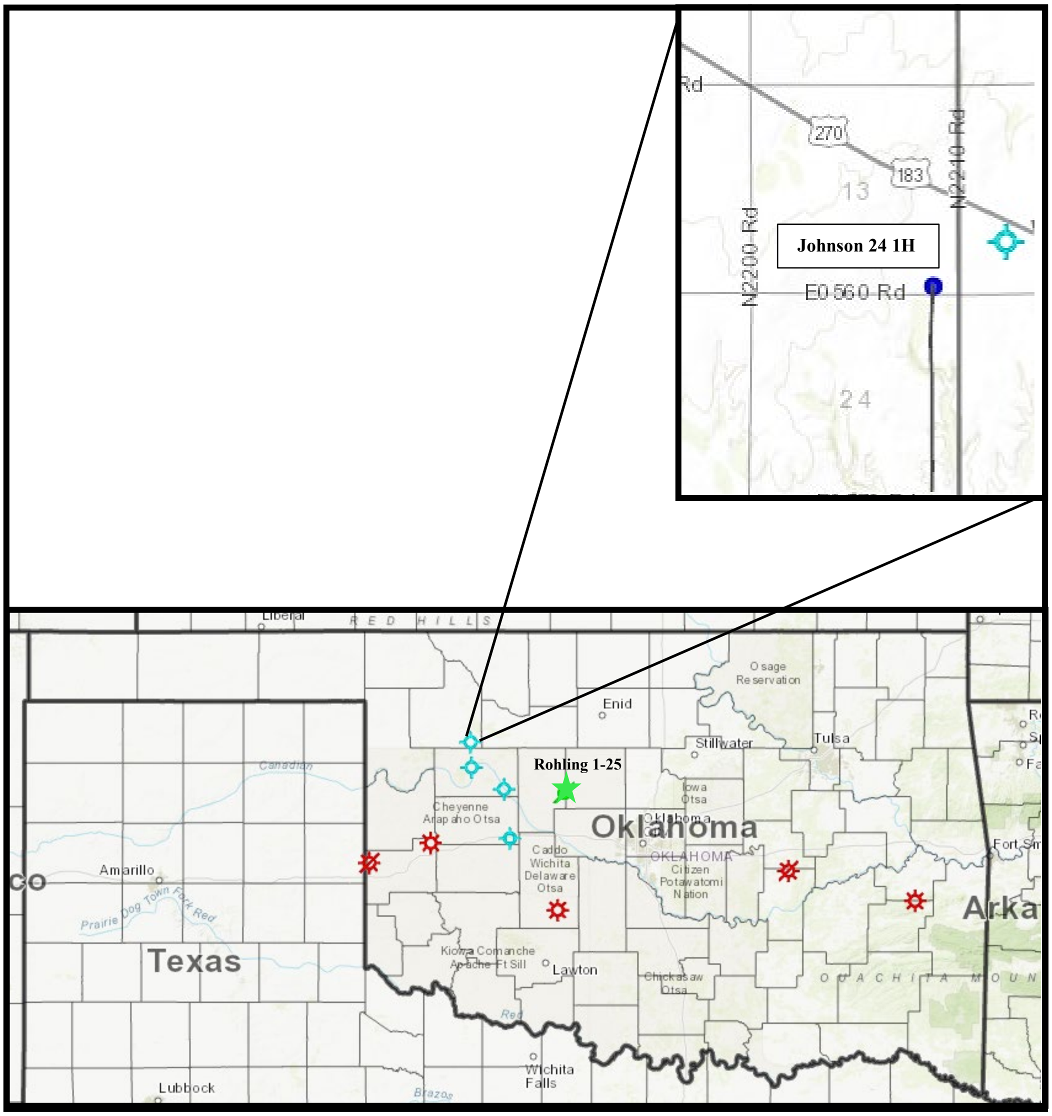

Figure 22: A map of Oklahoma obtained from Drilling Info showing the locations of the Meramec marker well, the Rohling 1-25, the wells which had lithology indicator logs with the highest match percentage in the cluster analysis, and the inset map shows the zoomed in location of the Tapstone Johnson 24 1H well which was completed in sec. 24 of 20N-18W and is the horizontal well most closely related to the Bessie 1-18.

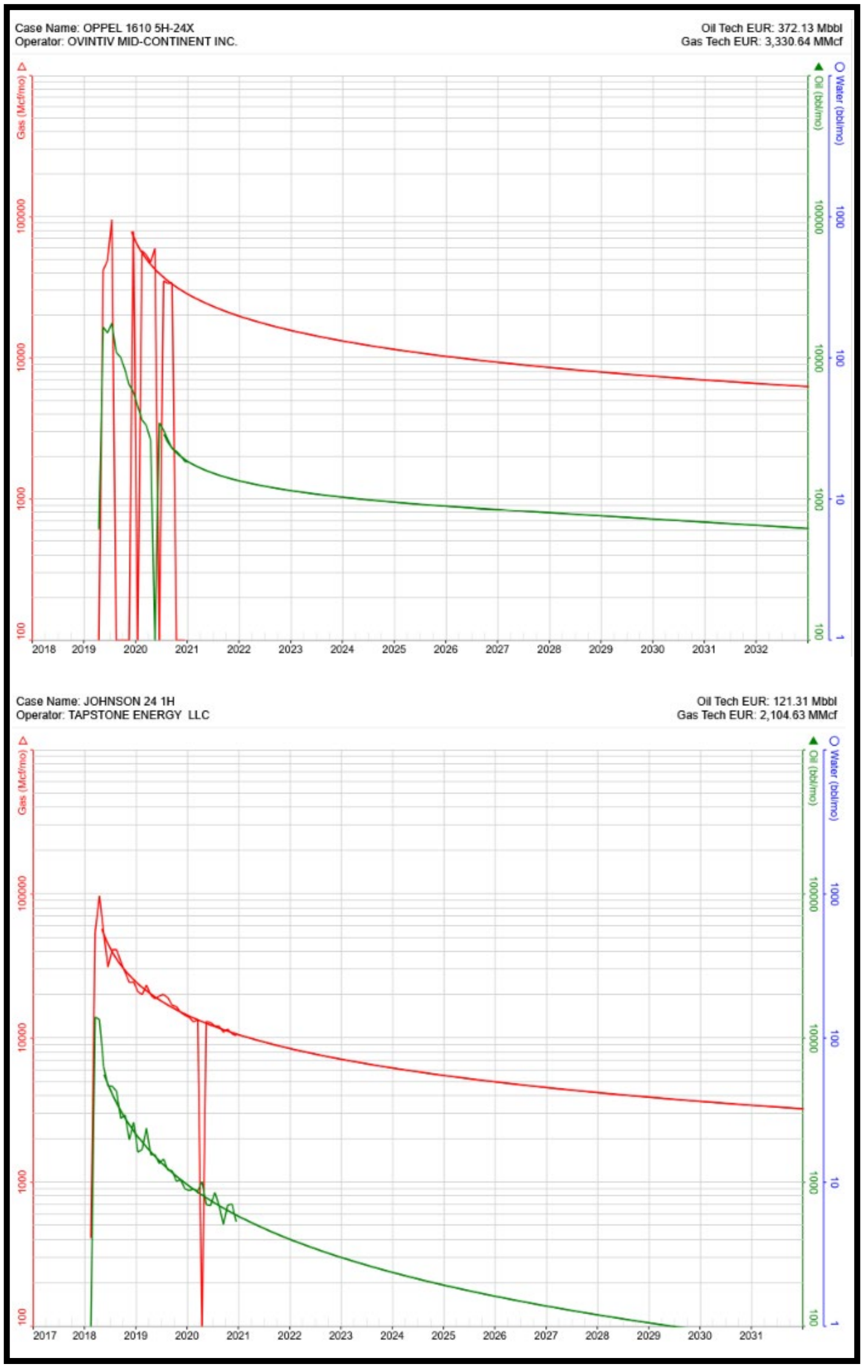

Figure 23: Comparison of the production curves with public production data obtained from Drilling Info and decline curves created in PHDwin. The Oppel unit wells on top which are located in Blaine County OK, and the Johnson 24 1H on the bottom which is located in Woodward County OK are over 50 miles apart, yet with their similar Process Energy Motifs (2021) and their similar production volumes and decline rates, it looks as though they could be neighbors.

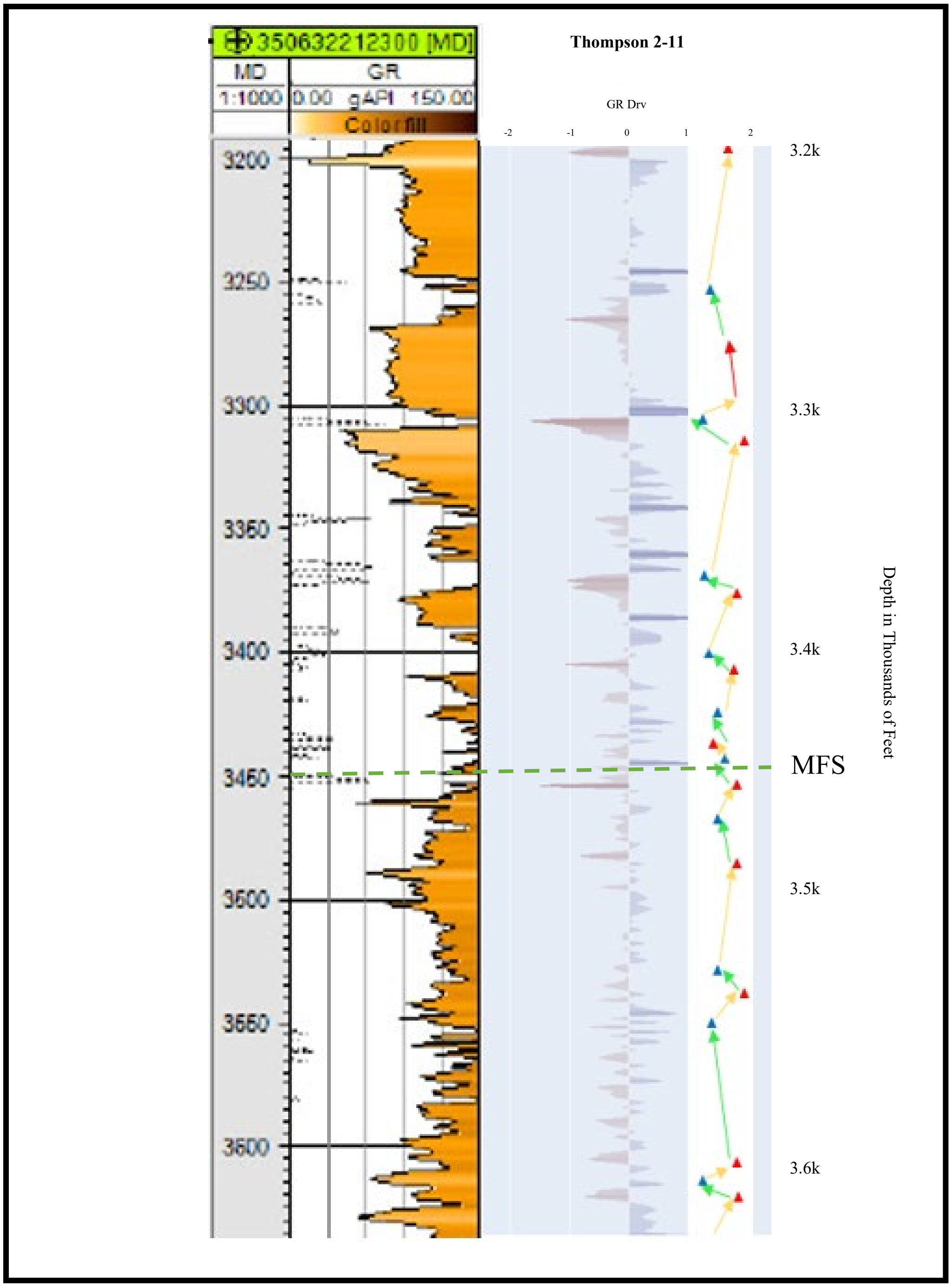

Figure 24: The first track on the left is the original Gamma Ray curve of the Thompson 2-11. The track on the right is the GR Derivative curve used to enhance geologically significant curve shapes while drowning out the insignificant noise. A Process Energy Motif (2021) has been overlain onto the GR Derivative curve in order to aid in sequence stratigraphic interpretation. Blue arrows represent HSTs, yellow RSTs, red LSTs, and green TSTs. Comparing with figure 18, there are very similar depositional energies at work and this very well could be a lateral constituent of the Meramec or a genetically related package.

(Figure 24). The horizontal well that is most closely related to the Thompson 2-11 is the Calyx Energy Fields 19-1HX well (Figure 25). The Fields was drilled and completed in 2016 and came on at around 300 BOE, which is a much lower IP, but this is a Gas well with very little oil production. Calyx Energy's target formation with the Fields was the Mississippian, which again with the available data we cannot be sure which reservoir in the Miss Calyx was targeting. Based on the results that were generated from the Bipartite Petrophysical Process Energy Motif (2021) and the production curves, it does appear that it could be a lateral constituent of the Meramec, or a related package that had similar factors acting upon it during the time of deposition. When analyzing the production curves of the Oppels and the Fields, it is distinct that the Gas curves are relatively similar, likely due to similar reservoir properties (Figure 26). With the Mississippian occurring at a much shallower depth on the East side of the state, it would definitely be expected to see this as a primarily gas producer. Another interesting fact about the Fields well is that it is in line with the Oppels and the Johnson on the north west trending ramp of the ancient Oklahoma Basin.

The last well to look into was the highest match from the clustering results. The Cupp B 3 is located in Beckham County Oklahoma and the RHOB curve matched with that of the Rohling 1-25 at 46% (Figure 16). The horizontal well that is most closely associated with the Cupp is the Mustang Fuel Mills 8-19H (Figure 27). The Mills drilled and completed in the Granite Wash play in western Oklahoma, so comparing the production curve to that of the Oppel's is rather irrelevant. However, being that the RHOB curves matched below the granite wash, it would be beneficial to point out that Mississippian strata does in fact exist below the granite wash in this area (Figure 28).

Based on the results from the rest of this study, it is possible there may be similar reservoir quality in the Mississippian strata below that Granite Wash to that of the Oppels, the Johnson, and the Fields, but more research would need to be done to test this theory.

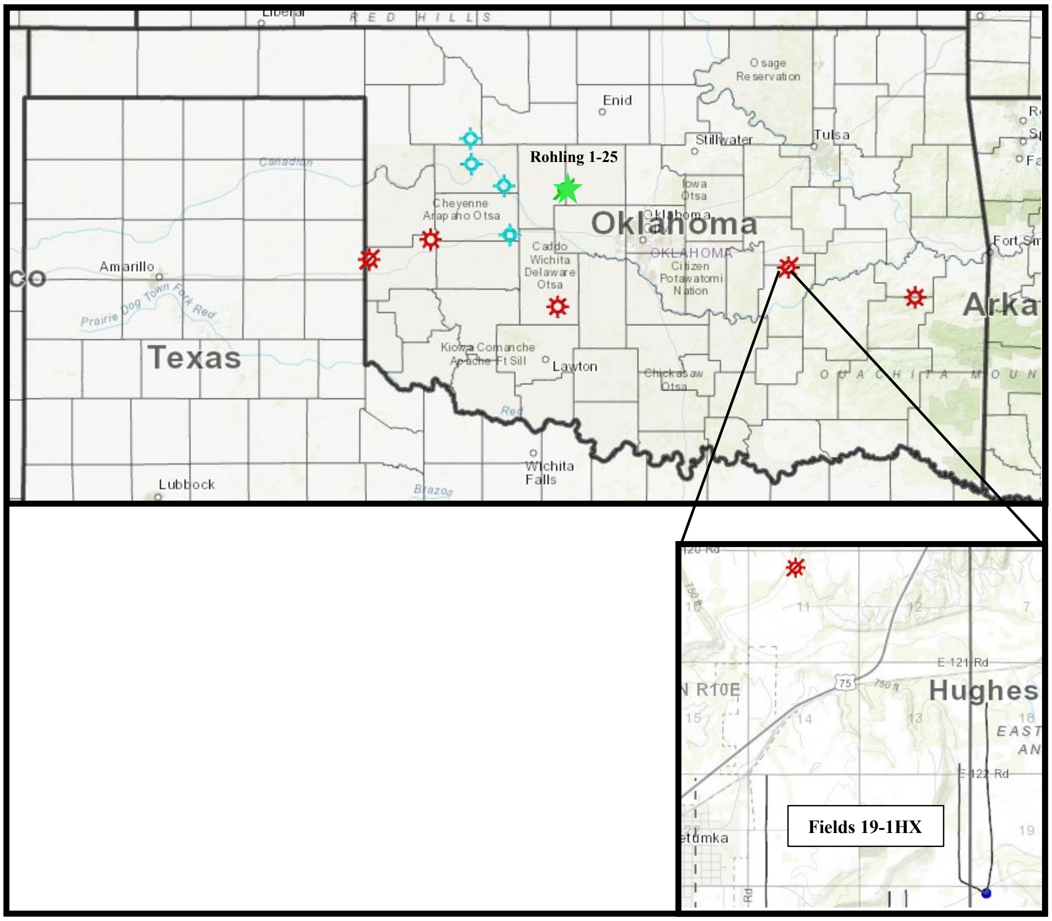

Figure 25: A map of Oklahoma obtained from Drilling Info showing the locations of the Meramec marker well, the Rohling 1-25, the wells which had lithology indicator logs with the highest match percentage in the cluster analysis, and the inset map shows the zoomed in location of the Calyx Fields 19-1HX well which was completed in sec. 24 of 09N-10E and sec. 18 and 19 of 09N-11E and is the horizontal well most closely related to the Thompson 2-11.

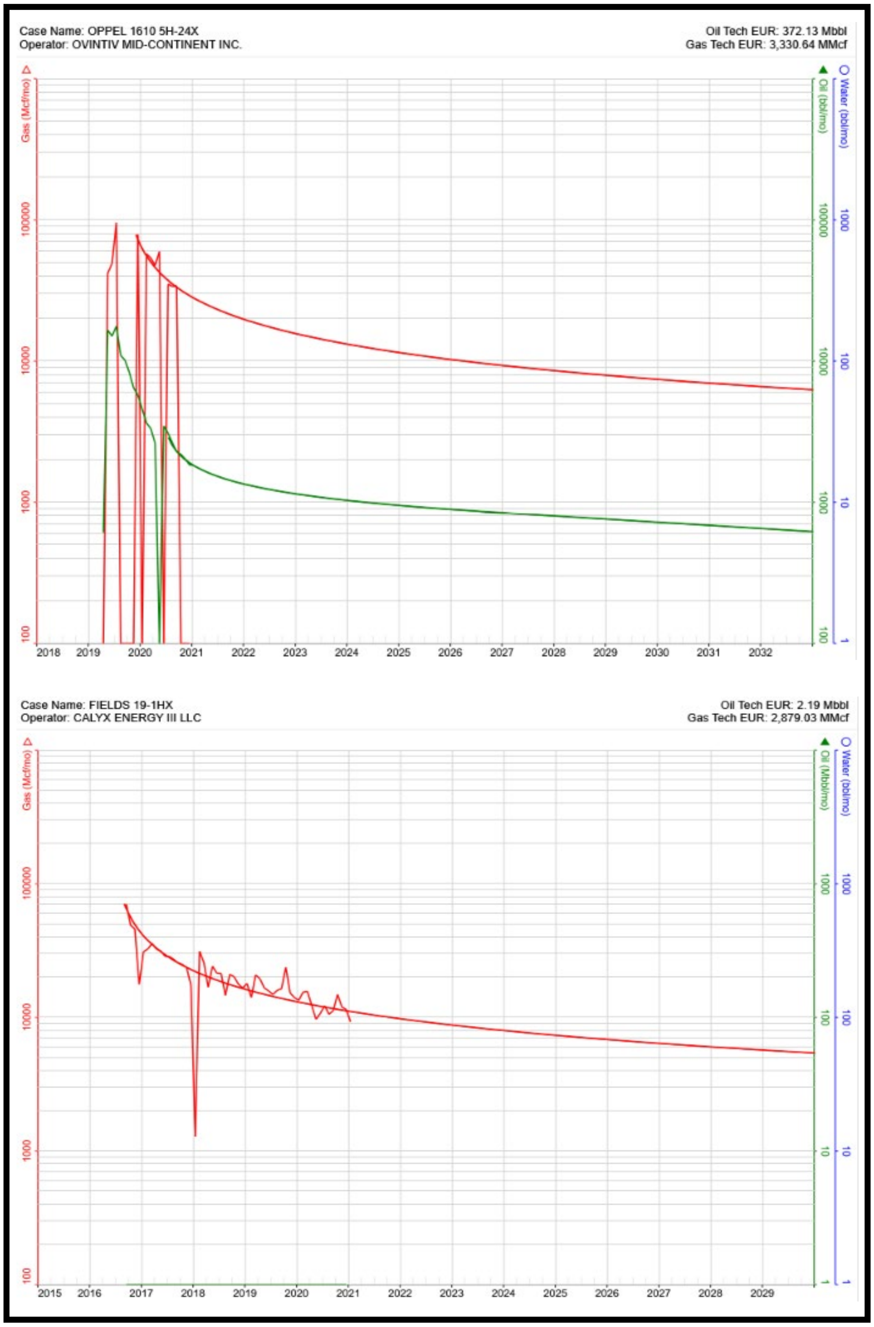

Figure 26: Comparison of the production curves with public production data obtained from Drilling Info and decline curves created in PHDwin. The Oppel unit wells on top which are located in Blaine County OK, and the Fields 19-1HX on the bottom which is located in Hughes County OK are over 180 miles apart, yet with their similar Process Energy Motifs (2021) and their similar gas production volumes and decline rates, it is plausible that the reservoirs are quite similar.

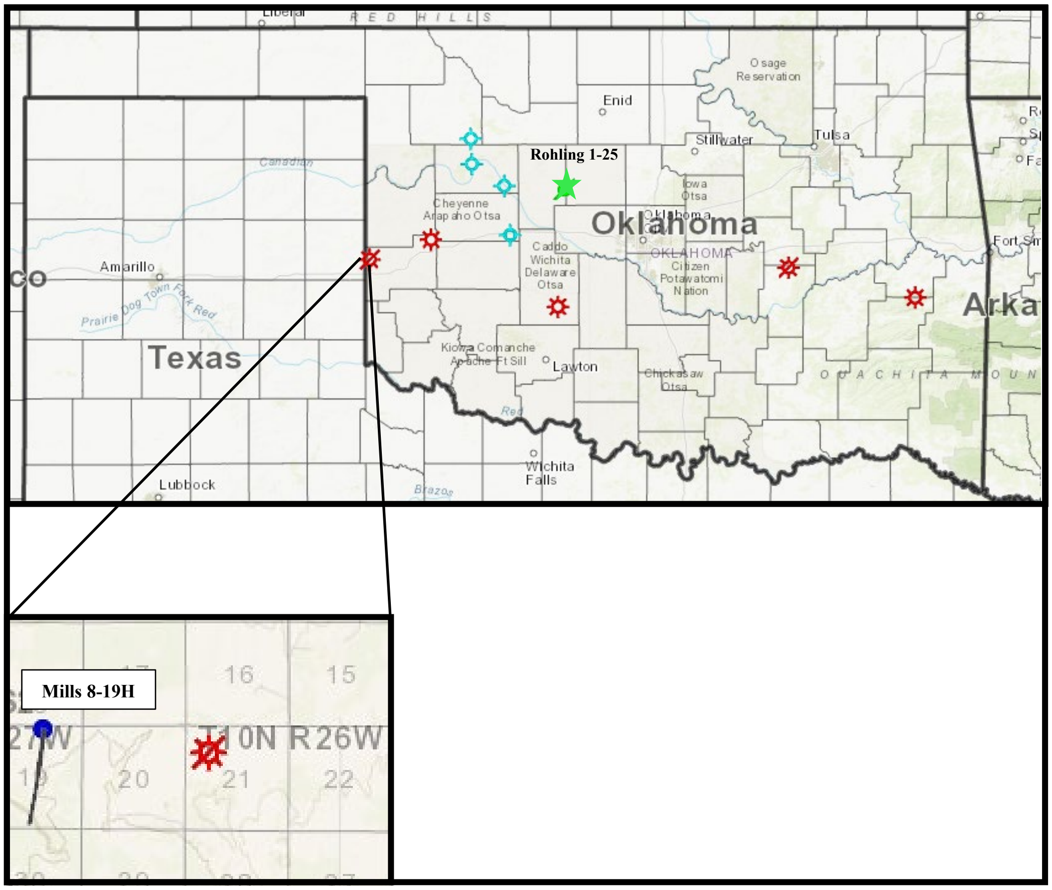

Figure 27: A map of Oklahoma obtained from Drilling Info showing the locations of the Meramec marker well, the Rohling 1-25, the wells which had lithology indicator logs with the highest match percentage in the cluster analysis, and the inset map shows the zoomed in location of the Mustang fuels Mills 8-19H well which was completed in sec. 19 of 10N-26W and is the horizontal well most closely related to the Cupp B 3. This well matched RHOB with the highest percentage of 46%.

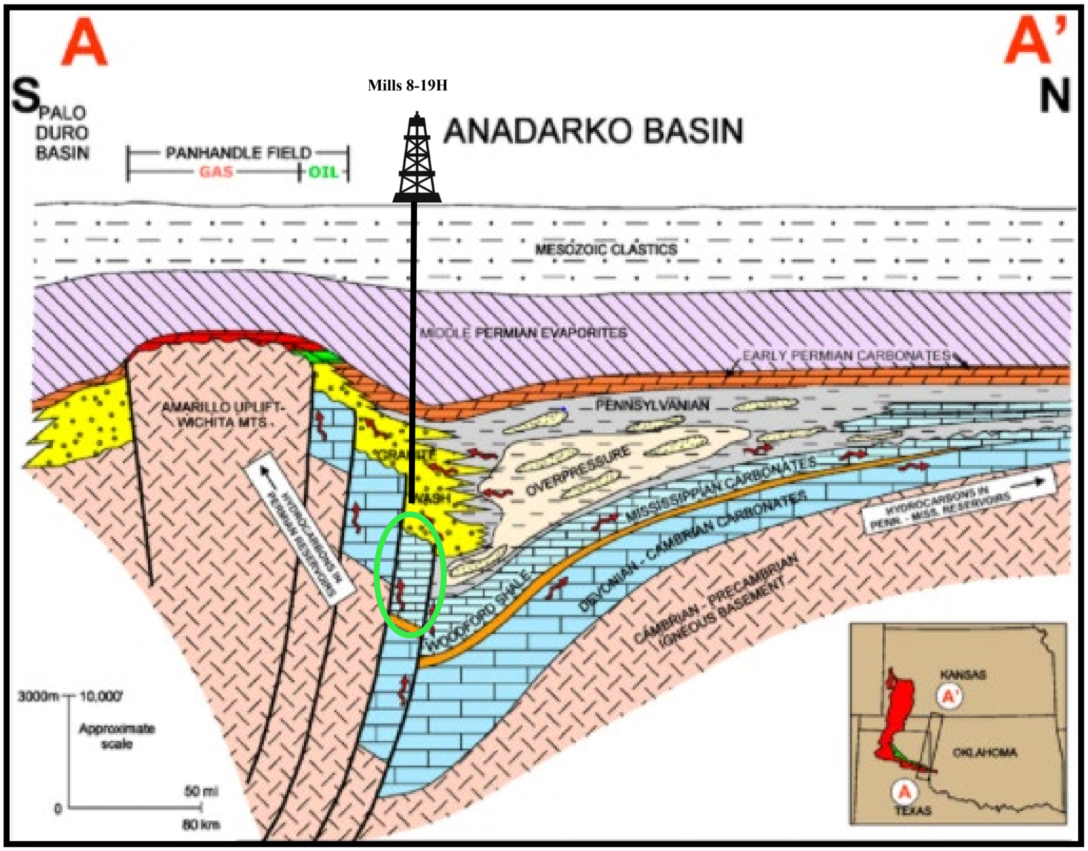

Figure 28: A south to north cross-section in far western Oklahoma capturing the area where the Mills 8-19H was drilled. The Mills drilled and completed in the Granite Wash play as many wells do in Beckham County. The green circle on the map shows where Mississippian sediments are located below the Granite Wash, and this is due to the nature of the Basin during Mississippian time when the entire Oklahoma Basin was a shallow epicontinental sea. The RHOB curve match of 46% was located below the Granite Wash, could this mean that there is a Mississippian reservoir with similar quality to that which the Oppels, Johnson, and Fields tapped into? More research would need to be done to test this theory (Fierstien, 2014).

Limitations

A discussion of some of the limitations that were discovered while working through this project and how they could be improved is now warranted. First, the log data set that was available was extensive, but many logs were very low quality, and did not have the lithology indicating curves that the study required which reduced the amount of wells that could be used in the study. One of the main issues that was encountered was with the strength of the computer that this volume of data required. A very powerful AWS machine was used to perform the operations, but it was only enough to process cosine similarity with stems of 200ft. Financially speaking it would be incredibly expensive to get a machine with the power to run stems of 10ft or even 1ft, but that is the granularity that is truly needed to make this work as it was envisioned. In addition to cosine similarity, the K-means clustering could have also benefited from a more powerful machine, as clustering all of the well log signatures is a complex combination of inputs and drags down accuracy when the operations that can be performed are limited due to power and time. The ability to experiment with different clustering techniques and analysis would have also been useful, but due to time constraints with how long it takes for the operations to run, it was not possible for this study. This thesis research is a preliminary study and the first time that machine learning has been used to attempt to compare well log signatures for similarity across an entire basin, so it should be expected that results will improve with time and additional work.

CONCLUSIONS

This study presents a novel approach to using well logs and clustering techniques to find and rank similar packages of rock to each other based on percent sameness, and then applying log attributes to compare log responses to production results. The basis behind this study was a desire to push the envelope and try and draw conclusions from the conglomerate of well log data that has been amassed in the Anadarko Basin using modern technology such as advanced clustering techniques. Machine learning as viewed in this investigation is learning from the past (previous wells) in order to predict the future (exploration wells). Decline curves are used to predict future performance of a well, based on direct production measurements; log signatures should also be used to predict future performance of wells, which are based on direct rock measurements. The only way to do this with the shear amount of data that is out there is with machine learning, and an unsupervised learning study like that which has been performed will let the data speak for itself. With the current method of acquiring new acreage being either long expensive studies, or using close proximity exploration, it seems time for a new, practical method that can serve as a quick look analysis.

While the novel methods that have been performed here are quite rudimentary, the potential future application of a workflow designed to search through thousands of well logs to find specific Gamma Ray signature packages of interest should not be taken lightly. The same methods can be applied to all other digital log curves of interest. While this study chose to analyze Gamma

Ray signatures of interest in oil and gas, it can also be applied to other natural resource exploitation such as the search for Helium. In the future it will be necessary for this tool to improve its accuracy, however even the current results were able to scan through 25,673 digital well logs and match Mississippian rock from one well with other sections of wells that show Mississippian rock, proving that even in the rudimentary form it can be used to drastically reduce time spent on searching through logs and increase time spent on interpretation and resource development. This investigation yields promising results showing that this technology has the possibility to be refined and used as an everyday tool.

REFERENCES

Ball, M., Mitchell, H.E., Frezon, S.E., Petroleum Geology of the Anadarko Basin Region Province (115), Kansas, Oklahoma, and Texas, U.S. Geological Survey, 1991.

Banerjee, S., 2020, K-means clustering from scratch, Machine Learning, Towards AI—Multidisciplinary Science Journal

Blakey, R., 2011, Paleogeography and geologic evolution of North America, http://cpgeostystems.com/paleomaps.html, (accessed August 2020).

Bougher, B.B., 2016, Machine learning applications to geophysical data analysis, M.S. Thesis, The University of British Columbia, Vancouver, B.C., 86p.

Boyd, D.T., 2008, Stratigraphic guide to Oklahoma oil and gas reservoirs: Oklahoma Geological Society, Special Publication 2008-1, 2 p

Brown, D., 2020, Anadarko Basin: A Super Super Basin, AAPG Explorer, American Association of Petroleum Geologists. https://explorer.aapg.org/story/articleid/57181

Campbell, J. A., C. J. Mankin, A. B. Schwarzkopf, and J. J. Raymer, 1988, Habitat of petroleum in Permian rocks of the Midcontinent region; in, Permian Rocks of the Midcontinent, W. A. Morgan and J. A. Babcock, eds.: Midcontinent Society of Economic Paleontologists and Mineralogists, Special Publication No. 1, p. 13–35.

Cardott, B. J. and M. W. Lambert, 1985, Thermal maturation by vitrinite reflectance of Woodford Shale, Anadarko Basin, Oklahoma: AAPG Bulletin, v. 69, no. 11, p. 1982-1998

Childress, M., 2015, High resolution sequence stratigraphic architecture of a mid-continent Mississippian outcrop in southwest Missouri, M.S. Thesis, Oklahoma State University, Stillwater, Oklahoma, 272 p.

Doveton, J., 2003, Reading the rocks from wireline logs, oil and gas information, Kansas geological survey. www.kgs.ku.edu/PRS/ReadRocks/portal.html

Droege, L., Vick, H., 2018. Redefining the STACK Play from Subsurface to Commercialization: Identifying Stacked Pay Sweet Spots in the Northern Anadarko Basin. Search and Discovery 11104.

Drummond, K., 2018, Regional stratigraphy and proximal to distal variation of lithology and porosity within a mixed carbonate-siliciclastic system, Meramec and Osage series (Mississippian), central Oklahoma, M.S. Thesis, University of Oklahoma, Norman, Oklahoma, 165 p.

Dutton, S.P., 1984, Fan-Delta Granite Wash of the Texas panhandle, Oklahoma City Geological Society, vol. Short Course Notes.

Fierstien, J., 2014, The Anadarko Basin makes Oklahoma oil & gas more the OK, Enverus Innovator Blog, https://www.enverus.com/blog/anadarko-Basin-oklahoma-oil-and-gas-industry-ok/#

Galloway, W., E., 1989, Genetic stratigraphic sequences in Basin analysis; 1, Architecture and genesis of flooding-surface bounded depositional units, AAPG bulletin, 73,125-142.

Gutschick, R. C., and C. A. Sandberg, 1983, Mississippian continental margins of the conterminous United States: in D. J. Stanley, and G. T. Moore, eds., The Shelfbreak: Critical Interface on Continental Margins: SEPM Special Publication 33, p. 79–96.

Hancock, N., 1992, Quick-Look Lithology from Logs: Part 4. Wireline Methods. ME 10: Development Geology Reference Manual, pp. 174-179.

Haq, B.U., and S.R. Schutter, 2008, A chronology of Paleozoic sea-level changes: Science, v. 322, p. 64–68.

Hartigan, J.A., 1975, Clustering algorithms: New York, John Wiley & Sons, p. 84–107.

Hickman, G., 2018, Parasequence-scale stratigraphic variability of lithology and porosity of Mississippian Meramec reservoirs and the relationships to production characteristics, STACK trend, Oklahoma, M.S. Thesis, University of Oklahoma, Norman, Oklahoma, 105 p.

Inverarity, K., 2020. Kinverarity1/Lasio. [online] GitHub. Available at: https://github.com/kinverarity1/lasio

Johnson, K.S., 1988, Anadarko Basin Symposium, Oklahoma Geological Survey Circular 90

Johnson, K.S. and K.V. Luza, 2008, Earth sciences and mineral resources of Oklahoma, Educational Publication 9, Oklahoma Geological Survey, 22 p.

Kerans, C., and S. Tinker, 1997, Sequence stratigraphy and characterization of carbonate reservoirs: SEPM Short Course Notes 40, 130 p.

LoCricchio, E., 2012, Wash Play Overview, Anadarko Basin: Stratigraphic framework and controls on Pennsylvanian granite wash production, Anadarko Basin, Texas and Oklahoma, AAPG Search and Discovery Article, no. 110163.

Luo, G., Tian, Y., Sharma, A., Ehlig-Economides, C., 2019, Eagle ford well insights using data-driven approaches, International Petroleum Technology Conference, Beijing, China, 26-28 March 2019.

Manning, C.D., Raghavan, P., Schütze, H. (2008), Introduction to Information Retrieval. Cambridge University Press. https://nlp.stanford.edu/IR-book/html/htmledition/the-vector-space-model-for-scoring-1.html

McConnell, D.A., M.J. Goyda, G.N. Smith, and J.P. Chitwood, 1989, Morphology of the frontal fault zone, southwest Oklahoma: Implications for deformation and deposition on the Wichita Uplift and Anadarko Basin: Geology, vol. 18, no. 7, p.34–637

Miller, J., Cullen, A., 2018, Sycamore Limestone Arbuckle Mountains, Shale Shaker, 69 2. 10.13140/RG.2.2.25005.59361

Miller, J. C., M. J. Pranter, and A. B. Cullen, 2019, Regional stratigraphy and organic richness of the Mississippian Meramec and associated strata, Anadarko Basin, central Oklahoma: Shale Shaker, 70, 50–79.

Miller, J., 2018, Regional stratigraphy and organic richness of the Mississippian Meramec and associated strata, Anadarko Basin, central Oklahoma, M.S. Thesis, University of Oklahoma, Norman, Oklahoma, 154 p.

Miller, M., 2019, Mississippian Meramec Lithologies and Petrophysical Property Variability, Stack Trend, Anadarko Basin, Oklahoma, M.S. Thesis, University of Oklahoma, Norman, Oklahoma, 106 p.

Miller, M., M. J. Pranter, I. Gupta, D. Devegowda, K. J. Marfurt, C. Sondergeld, C. Rai, C. T. McLain, J. Packwood, R. E. Larese, 2021, Mississippian Meramec lithologies and petrophysical property variability, stack trend, Anadarko Basin, Oklahoma, Interpretation, Vol. 9, No. 2 (May 2021); p. SE1–SE21.

Northcutt, R. A. and J. A. Campbell., 1995, Geologic provinces of Oklahoma: Oklahoma Geological Survey Open-File Report 5–95, 1 sheet, scale 1: 750000, 6-page explanation and bibliography.

Pandas Dev. Team, 2020, pandas.DataFrame.ewm, Available at: https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.ewm.html#pandas-dataframe-ewm

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, Vol. 12, pp. 2825-2830.

Pigott, John D. and Bradley, Bryant W., 2014, Application of Production Decline Curve Analysis to Clastic Reservoir Facies Characterization within a Sequence Stratigraphic Framework: Example- Frio Formation, South Texas, GCAGS Journal, Vol 3, p. 112-133

Pigott, John D., Rui Zhai, Matthew P. Lynch, Cyril S. Frazier, Tyler W. Bickley, Kulwadee L. Pigott, and Esra Yalcin, 2021, 3D seismic geomorphology and sequence stratigraphy of the Northern Delaware Basin Bone Spring Formation: Constraint from deepwater bipartite petrophysical motifs, American Association of Petroleum Geologists (AAPG) Global Super Basins Leadership Conference, 25-26 January, 2021.

Prabhakaran, S., 2020, Cosine Similarity – Understanding the Math and How it Works (with Python), [online] ML+, Available at: https://www.machinelearningplus.com/nlp/cosine-similarity/

Price, B., K. Haustveit, and A. Lamb, 2017, Influence of stratigraphy on barriers to fracture growth and completion optimization in the Meramec Stack Play, Anadarko Basin, Oklahoma: Unconventional Resources Technology Conference (URTEC), Article #2697585, 8 p.

Robinson, M.C., 2014, Using interpreted digital well-logs and sequence stratigraphy to develop high resolution subsurface models, Enverus technical library, 12 p.

Shapiro, L. G., and G. C. Stockman, 2000, Computer vision: Upper Saddle River, NJ, Prentice Hall, p. 170.

Shelley, S. A., 2016, Outcrop-based sequence stratigraphy and reservoir characterization of an upper Mississippian mixed carbonate-siliciclastic ramp, Mayes County, Oklahoma, M.S. Thesis, Oklahoma State University, Stillwater, Oklahoma, 92 p.

Shelley, S., G. M. Grammer, and M. J. Pranter, 2019, Outcrop-based reservoir characterization and modeling of an Upper Mississippian mixed carbonate–siliciclastic ramp, northeastern Oklahoma, in G. M. Grammer, J. M. Gregg, J. O. Puckette, P. Jaiswal, S. J. Mazzullo, M. J. Pranter, and R. H. Goldstein, eds., Mississippian reservoirs of the midcontinent: AAPG Memoir 122, p. 207–225.

Vail, P. R., 1987, Seismic Stratigraphy Interpretation Using Sequence Stratigraphy: AAPG Studies in Geology 27, volume 1:Atlas of Seismic Stratigraphy.

Wethington, N., 2017, Stratigraphic architecture of the Mississippian Limestone through integrated electrofacies classification, hardtner field area, Kansas and Oklahoma, M.S. Thesis, University of Oklahoma, Norman, Oklahoma, 112 p.

Appendix A. Geologic Setting

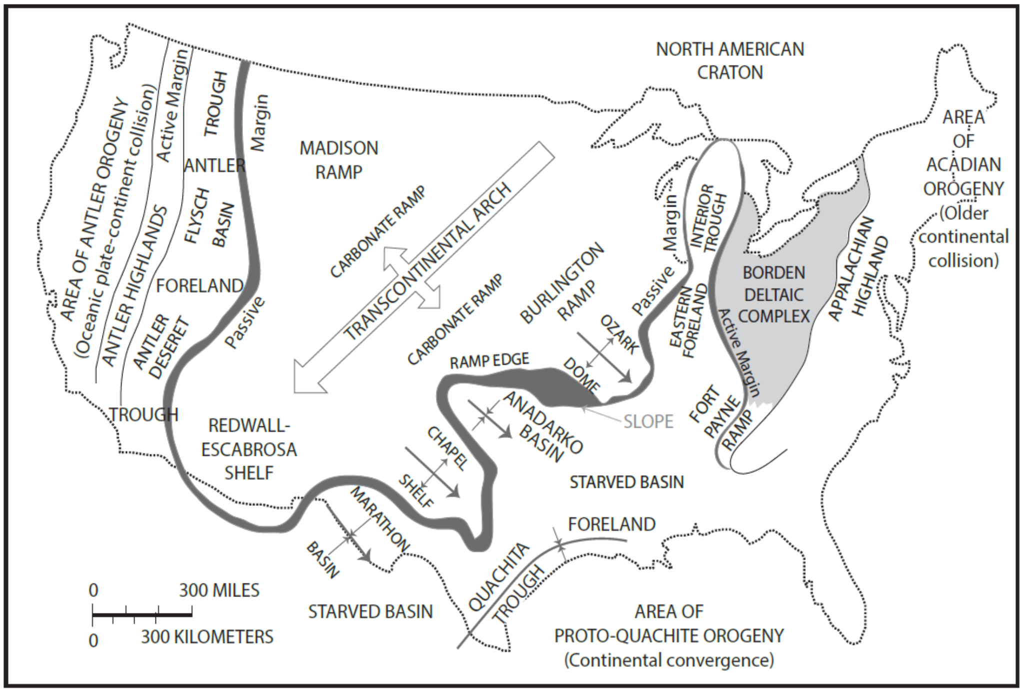

Figure A-1. Paleogeographic map of North America 325 million years ago during the Mississippian and after the formation of the Transcontinental Arch which separated the Madison Ramp and the Burlington Ramp. The Burlington carbonate ramp fed the proto-Anadarko Basin with sediments (modified from Gutschick, 1983 and Wethington, 2017).

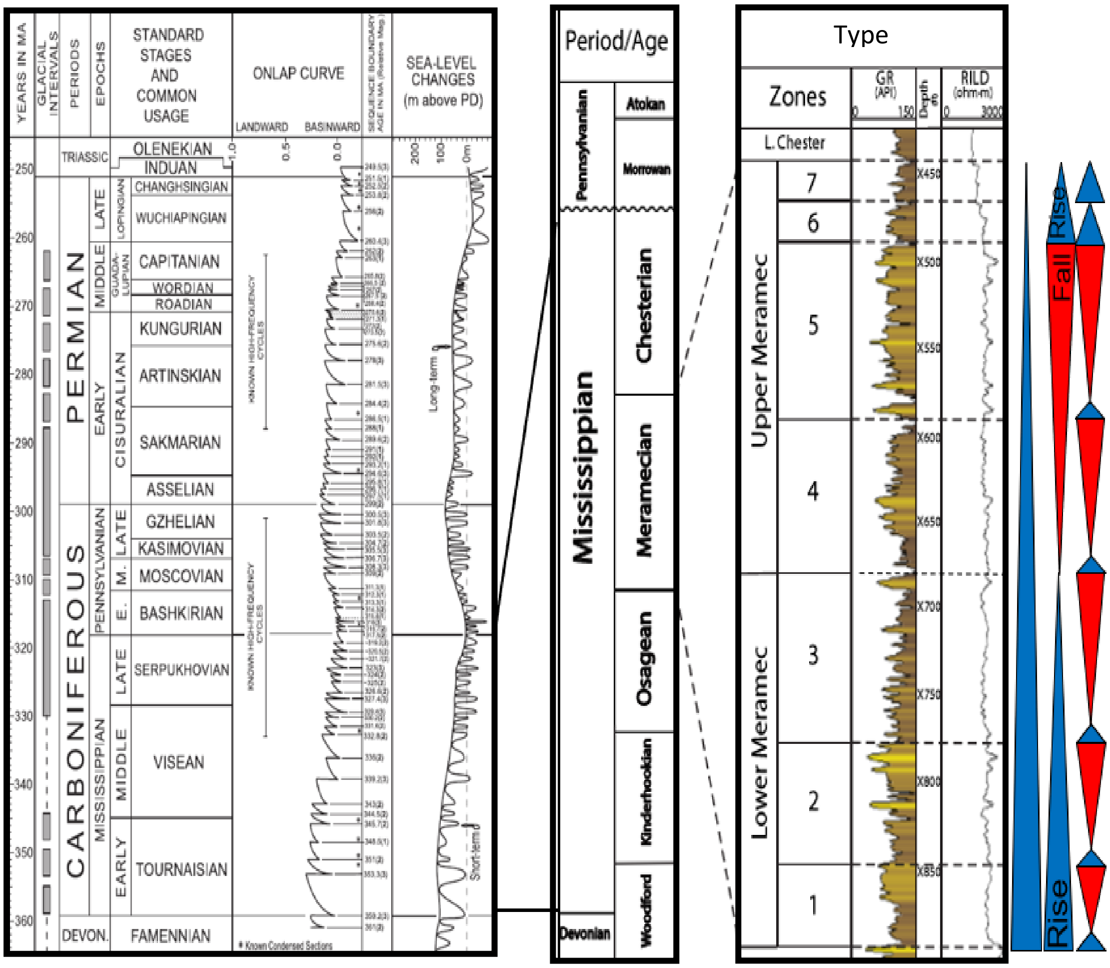

Figure A-2: During the Mississippian, there were four main depositional episodes: Kinderhookian, Oseagean, Meramecian, and Chesterian. This study focused on the Osagean and the Meramecian units at the top of the Kaskaskia sequence. Both the Osagean and Meramecian were deposited in mostly greenhouse conditions, but towards the upper Meramecian the transition to icehouse can be seen with higher frequency sea level changes. The Meramec is often characterized as being deposited on a gently sloping ramp with carbonate rich deposits proximally and mixed carbonate and siliciclastics Basinward (Modified from Johnson and Others, 1988; Boyd, 2008; Haq and Schutter, 2008; Williams, 2020).

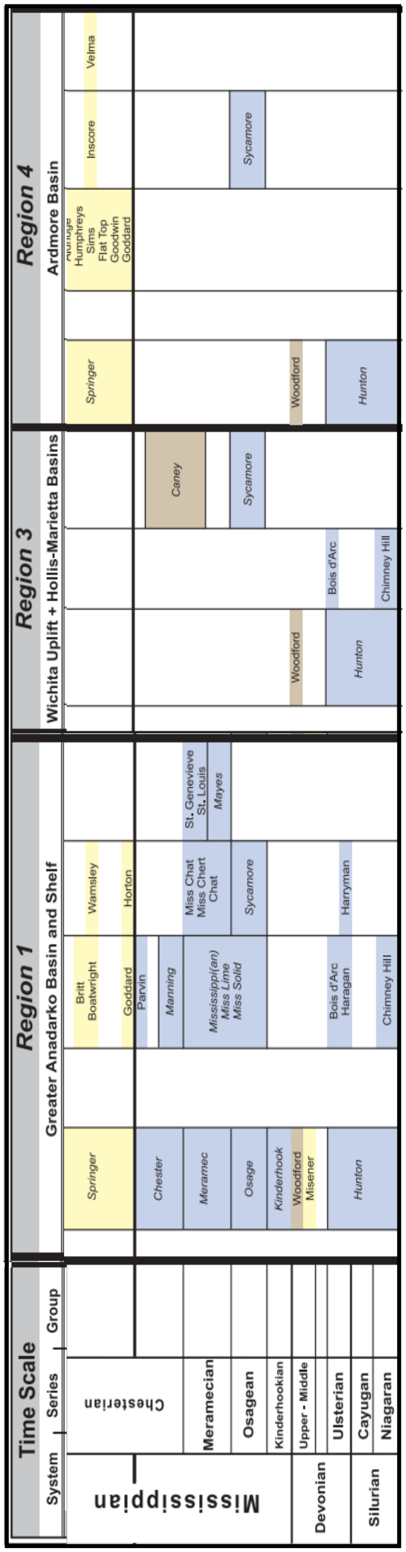

Figure A-3: One of the questions that hoped to be answered in the results of this study were what sort of relationship can be seen if any between Meramec, Osage, Sycamore Lime, and Caney Shale, chronostratigraphically and as reservoirs, factoring in the fact that portions of these formations were all deposited at the same time in the ancient Oklahoma Basin (Figure modified from Boyd, 2008).

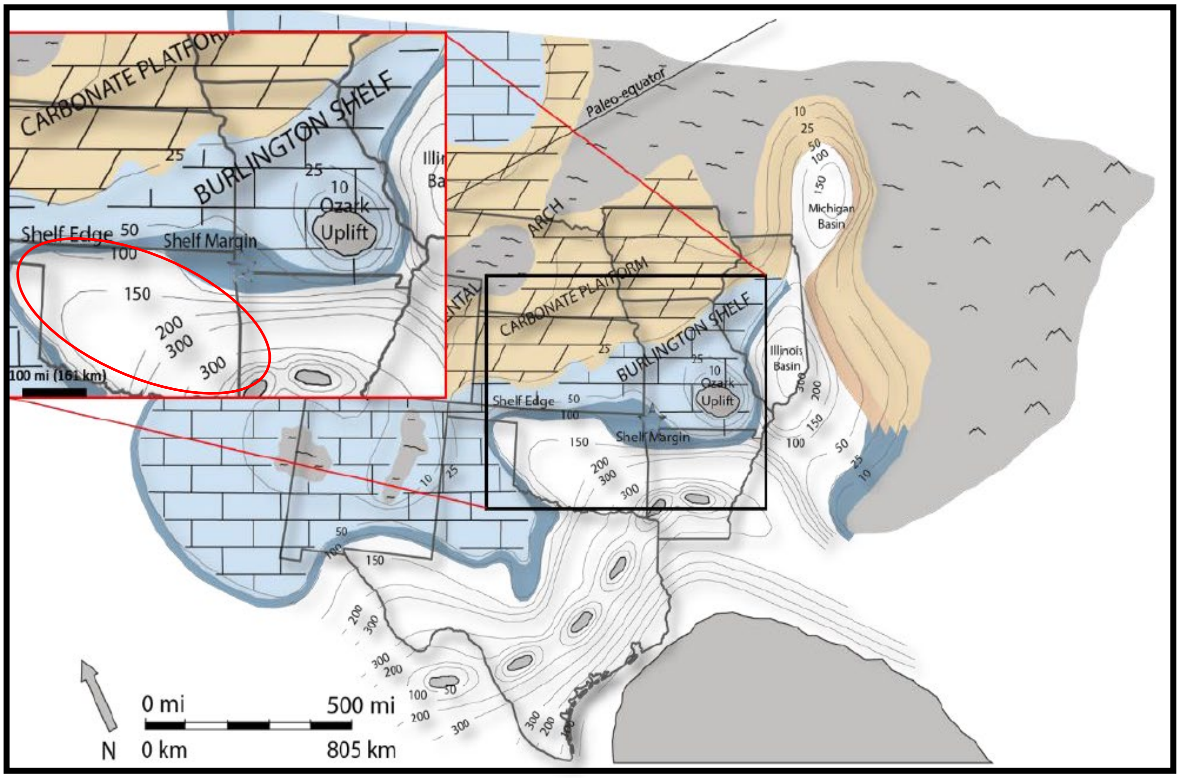

Figure A-4: Paleo-depositional map of North America 325 million years ago during the Mississippian and after the formation of the Transcontinental Arch which separated the Madison Ramp and the Burlington Ramp. The Burlington carbonate ramp fed the proto-Anadarko Basin with sediments. Blue colors showing the distribution of primarily carbonate sediments while tan shows the distribution of primarily clastic sediments. The red outlined box is zoomed in to show the Ancient Oklahoma Basin area in location to the paleo shelf. (modified from Gutschick, 1983 and Childress, 2015).

Appendix B: Sequence Stratigraphy

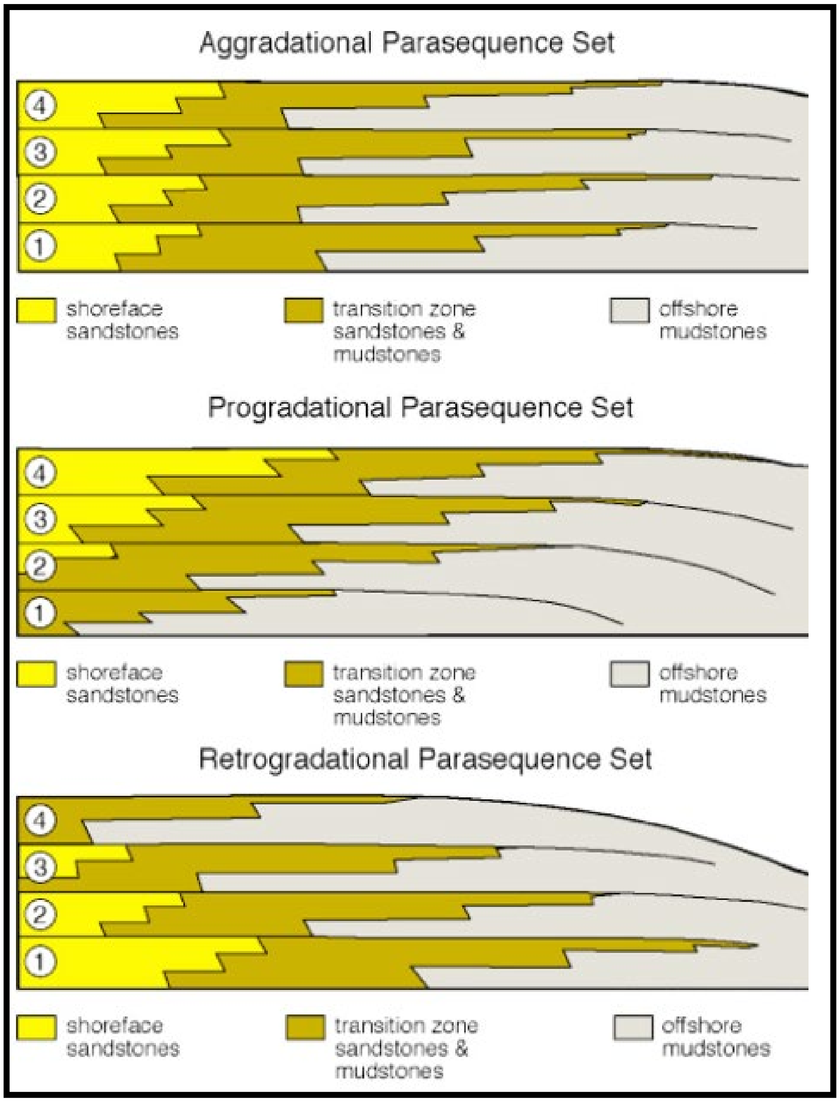

Figure B-1: From Van Wagoner et al. (1990)

- **Parasequence** genetically related package of strata bounded by flooding surfaces
- **Transgression**: ocean moves towards shore; result of sea level rise
- **Regression**: ocean moves away from shore; result of sea level fall
- **Progradational**: shore and nearshore deposits move outward into the ocean and overlie deeper water deposits
- **Retrogradational**: deeper water deposits move towards land and overlie shallow water deposits
- **Aggradational**: facies remain in the same general location and stack atop others of the same facies

Tectono-Eustatic/ Eustatic Cycle Order	*Sequence Straigraphic Unit*	Duraction (my)	*Relative Sea Level Amplitude (m)*	*Relative Sea Level Rise/Fall Rate (cm/1,000 yr)*
First		>100		<1
Second	Supersequence	10-100	50-100	1-3
Third	Depositional Sequence Composite Sequence	1-10	50-100	1-10
Fourth	High Energy Sequence, Parasequence and Cycle Set	0.1-1	1-150	40-500
Fifth	Parasequence, High-Frequency Cycle	0.01-0.1	1-150	60-700

(From SEPM#40)

Figure B-2: Chart displaying the stratigraphic cycle hierarchy showing the ranges of sea level variations through cycle order naming, sequence stratigraphy unit naming, time in millions of years, sea level amplitude, and sea level rise and fall rate (Kerans and Tinker 1997).

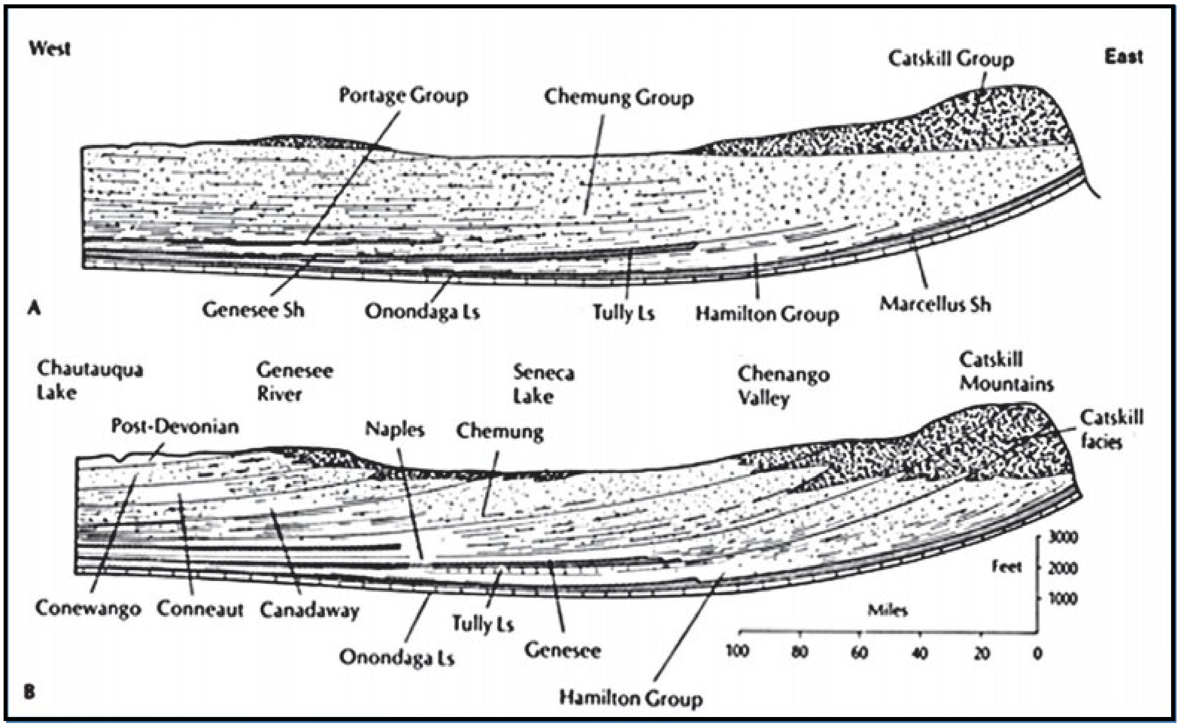

Figure B-3: Example of the difference between a common lithostratigraphic interpretation (A) versus an allostratigraphic approach (B). The allostratigraphic approach will provide a more wholistic picture to the interpreter by taking into account different timing of sediment accumulation and accounting for erosional, non-depositional, and flooding surfaces (Robinson, 2014; Prothero and Schwab, 1996; Dunbar and Roger, 1957).

Appendix C: Clustering and Data Techniques

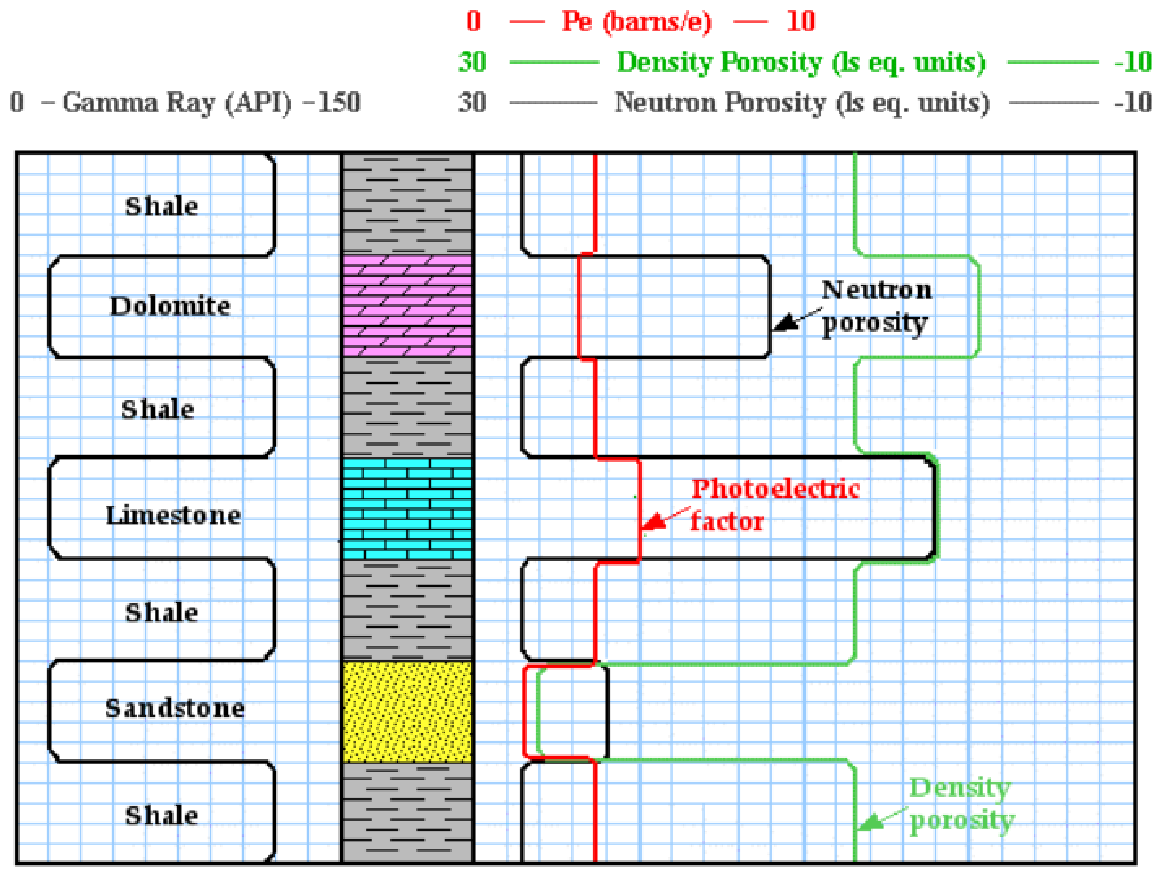

Figure C-1: Example figure of four of the most common lithology indicator well logs, and what the curve shape would be generally expected to look like for shale, dolomite, limestone and sandstone (Doveton, 2003).

```
X_std = (X - X.min(axis=0)) / (X.max(axis=0) - X.min(axis=0))
X_scaled = X_std * (max - min) + min
```

Figure C-2: Equation for Sklearn preprocessing normalization module called "MinMaxScaler". Scales features individually into a given range; in the case of this project the range is between 0 and 1. Min and max would be 0 and 1 respectively (Pedregosa et. al., 2011).

`cosine_similarity` computes the L2-normalized dot product of vectors. That is, if x and y are row vectors, their cosine similarity k is defined as:

$$k(x, y) = \frac{xy^\top}{\|x\| \|y\|}$$

This is called cosine similarity, because Euclidean (L2) normalization projects the vectors onto the unit sphere, and their dot product is then the cosine of the angle between the points denoted by the vectors.

Figure C-3: Table explaining the usage of cosine similarity function, which was used to compare well log signatures, before comparisons were clustered (Manning et al., 2008).

A *K-means algorithm*

Input: k (number of clusters), D (data points)

1. *Choose random k data points as initial clusters mean*
2. *Associate each data point in D to the nearest centroid. This will divide the data into k clusters.*
3. *Recompute centroids*
4. *Repeat step 2 and step 3 until there are no more changes of cluster membership of the data points.*

B

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

Figure C-4: Part A is explaining the basic steps to run K-means clustering, and part B is the equation for Euclidean Distance between two points. Each data point is assigned to a cluster determined by the smallest squared Euclidian distance between them (Banerjee, 2020).

Appendix D: DTA Modeling

Parameters: **com** : *float, optional*

Specify decay in terms of center of mass, $\alpha = 1/(1 + com)$, for $com \geq 0$.

span : *float, optional*

Specify decay in terms of span, $\alpha = 2/(span + 1)$, for $span \geq 1$.

halflife : *float, str, timedelta, optional*

Specify decay in terms of half-life, $\alpha = 1 - \exp(-\ln(2)/halflife)$, for $halflife > 0$.

If `times` is specified, the time unit (str or timedelta) over which an observation decays to half its value. Only applicable to `mean()` and halflife value will not apply to the other functions.

New in version 1.1.0.

alpha : *float, optional*

Specify smoothing factor α directly, $0 < \alpha \leq 1$.

min_periods : *int, default 0*

Minimum number of observations in window required to have a value (otherwise result is NA).

adjust : *bool, default True*

Divide by decaying adjustment factor in beginning periods to account for imbalance in relative weightings (viewing EWMA as a moving average).

- When `adjust=True` (default), the EW function is calculated using weights $w_i = (1 - \alpha)^i$. For example, the EW moving average of the series $[x_0, x_1, ..., x_t]$ would be:

$$y_t = \frac{x_t + (1 - \alpha)x_{t-1} + (1 - \alpha)^2 x_{t-2} + ... + (1 - \alpha)^t x_0}{1 + (1 - \alpha) + (1 - \alpha)^2 + ... + (1 - \alpha)^t}$$

- When `adjust=False`, the exponentially weighted function is calculated recursively:

$$\begin{aligned} y_0 &= x_0 \\ y_t &= (1 - \alpha)y_{t-1} + \alpha x_t, \end{aligned}$$

ignore_na : *bool, default False*

Ignore missing values when calculating weights; specify `True` to reproduce pre-0.15.0 behavior.

- When `ignore_na=False` (default), weights are based on absolute positions. For example, the weights of x_0 and x_2 used in calculating the final weighted average of $[x_0$, None, $x_2]$ are $(1 - \alpha)^2$ and 1 if `adjust=True`, and $(1 - \alpha)^2$ and α if `adjust=False`.
- When `ignore_na=True` (reproducing pre-0.15.0 behavior), weights are based on relative positions. For example, the weights of x_0 and x_2 used in calculating the final weighted average of $[x_0$, None, $x_2]$ are $1 - \alpha$ and 1 if `adjust=True`, and $1 - \alpha$ and α if `adjust=False`.

axis : *{0, 1}, default 0*

The axis to use. The value 0 identifies the rows, and 1 identifies the columns.

times : *str, np.ndarray, Series, default None*

New in version 1.1.0.

Times corresponding to the observations. Must be monotonically increasing and `datetime64[ns]` dtype.

If str, the name of the column in the DataFrame representing the times.

If 1-D array like, a sequence with the same shape as the observations.

Only applicable to `mean()`.

Figure D-1: Chart explaining how exponential weighted moving average runs. This was used while performing python DTA analysis in order to generate a smoothed curve based on a specified window of the original curve where all values that fall within the specified window are weighted according to their distance from the original point. These points are recalculated as a weighted average of the surrounding points. The goal of the weighted averages throughout the defined window is to smooth curves by calming high-frequency noise and preserving low frequency trends for easier more accurate quick-glance interpretation (Pandas Dev Team, 2020).