UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IDENTIFYING KEY STIMULATION PARAMETERS IN EAGLE FORD SHALE: DATA MINING AND STATISTICAL APPROACH

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

DANA K SAEED
Norman, Oklahoma
2017

IDENTIFYING KEY STIMULATION PARAMETERS IN EAGLE FORD SHALE:
DATA MINING AND STATISTICAL APPROACH

A THESIS APPROVED FOR THE
SCHOOL OF AEROSPACE AND MECHANICAL ENGINEERING

BY

____
Dr. Zahed Siddique, Chair

____
Dr. Farrokh Mistree

____
Dr. Maysam Pournik

Dedication

To my mother and father

Acknowledgements

I would like to thank my committee chair, Dr. Zahed Siddique, and my committee members, Dr. Farrokh Mistree, and Dr. Maysam Pournik for their guidance and support throughout my journey at the Baker Hughes 21st Century Coop Program and the course of this research.

I would also like to extend my gratitude to the people at Baker Hughes, especially Rustom Mody for establishing the Baker Hughes 21st Century Coop program and Larry Watkins for managing the program and making it a great learning experience that I was lucky enough to be part of, and Sergio Centurion, Randall Cade, and Partha Ganguly for offering their expertise and guidance on this study and providing me the completion data.

Finally, thanks to my mother and father for their encouragement and to my friends for their patience and love.

Table of Contents

List of Tables

List of Figures

Abstract

The principal question for this thesis is as follows:

What are the key hydraulic fracturing parameters that affect the amount of hydrocarbons that can be recovered in Eagle Ford Shale?

The problem considered in this thesis is establishing a relationship between hydraulic fracturing parameters and the amount of hydrocarbons produced.

The importance of this problem is identifying the stimulation parameters that can be used to improve the efficiency and effectiveness of fractures and ultimately increase the production of hydrocarbons. This will assist engineers to make better decisions related to hydraulic fracturing by focusing on the set of key stimulation parameters that are identified in this thesis.

The proposed method to the problem is to consider all the stimulation parameters together and use data mining and statistical techniques. In this thesis, the use of four different data mining and statistical approaches, Logistic Regression, Decision Trees, Support Vector Machines, and Neural Networks, are proposed. The foundation is based on the fact that the stimulation parameters are highly interrelated and need be considered together as a whole. These approaches allow the analysis of such a system and have the capability of capturing nonlinear relationships between the input and output parameters.

The major findings are identifying eight hydraulic fracturing parameters, Perforated Length Interval (ft), Injection Rate per Stage (bpm), Number of Clusters per Stage, Volume of Proppant per Stage (lbs), Volume of Water per Stage (gals), Number of Stages, Average Treating Pressure per Stage (psi), Maximum Treating Pressure per

Stage (psi) as the key stimulation factors that have a direct impact on the amount of hydrocarbons produced, determining an efficient method of analyzing and comparing multiple variables for multiple wells, establishing a production metric that reflects long term production performance, and identifying the best performing regression method.

CHAPTER 1

BACKGROUND AND PROBLEM FORMULATION

In this chapter, a brief description of the Baker Hughes challenge problem and the sustainability triangle is first presented to provide the background and motivation for selection of the research focus. The context and importance of the selected research focus is provided by presenting an overview of unconventional reservoirs and hydraulic fracturing. Then the problem along with the research questions are defined and the thesis objectives are presented. The proposed approach to the problem and the limitations of the proposed approach are also addressed in this chapter.

1.1 BACKGROUND AND MOTIVATION FOR SELECTION OF RESEARCH FOCUS

1.1.1 Baker Hughes Challenge Problem Defined

The Baker Hughes 21st Century Co-Op at the University of Oklahoma School of Aerospace and Mechanical Engineering is a five year BS/MS degree program in mechanical engineering aimed at developing technical competencies and meta-competencies needed by engineers to hit the road running and succeed in the oil and gas industry. In addition to core courses in mechanical engineering, the curriculum includes customized courses jointly offered by company engineers and faculty during summer internships, a senior capstone experience and graduate theses that are of relevance to the sponsoring company, and graduate cross-disciplinary courses from the School of

Industrial and Systems Engineering and the Mewbourne School of Petroleum Engineering.

Larry Watkins of Baker Hughes Inc. presented the BHI Scholars with the "challenge problem" in the beginning of the 2014 Fall Semester. Below was the problem presented to the team:

The BHI-Class of 2013 team focused on an overview of unconventional hydrocarbon resources, primarily shale plays. The challenge for BHI-Class of 2014 team is to extend the efforts from where BHI-Class of 2013 ended. The challenge for BHI-Class of 2014 is to review and identify the go forward challenges facing development of shale. For this challenge, consider the following dimensions (question areas) for developing shale:

- Technical Issues
- Political Issues
- Economics of Shale Development
- Recovery Factors in Shale

Political Issues: Identify and discuss factors in the political realm that currently influence development of shale resources. Provide thoughts on ways to mitigate these factors including but not limited to education or improved operating methods. Economics of Shale Development: Identify key factors that currently limit the economics of shale development. These factors include but are not limited to knowledge required for planning well paths and completion methods, approaches of different types of E&P companies to well placement and planning, costs/supply of components for well completions and fracture operations. Discuss how these factors influence the overall economics for E&P companies in shale operations.

Technical Issues: Identify and discuss the limitations of current technologies used in shale development. Discuss if the limitations are specific hardware, methods, materials, fundamental knowledge or combinations of these. Describe which technology areas influence the other elements of this challenge and then rank the technology issues in order of greatest positive impact on shale development going forward.

Recovery Factors in Shale: Identify the factors that currently determine initial recovery factors in shale development areas. Describe how uncertainty in the input parameters influences the recovery from shale reservoirs. Provide a prioritized list of which information would provide the greatest reduction in uncertainty when initially estimating recovery. Discuss what actions might be possible to improve recovery.

As a group, the BHI Scholars framed the shale development problem in the industry today looking at the four different perspectives: technical, political, economics, and recovery factors. Identifying the drivers, focuses, issues, and major dilemmas within the perspective further expanded each perspective using the sustainability triangle introduced by Dr. Farrokh Mistree to the Baker Hughes Scholars. The research approach to the challenge problem and the sustainability triangle is explained next.

1.1.2 Baker Hughes Challenge Problem: Research Approach

The BHI scholars broke into two interdisciplinary teams in order to tackle the challenge problem. Mechanical and petroleum engineering backgrounds were represented in both groups. The perspectives were split on terms of apparent connectivity. Each perspective was framed using the sustainability triangle.

A sustainability triangle was used by the team to organize complexity. The team assessed the perspectives from three different drivers including: social, environment, and economic. We further analyzed the drivers by determining the focus of the driver and the issues that are present in the industry from the corresponding perspective. The next step is to connect each issue from each driver to another issue of another driver. This connection needs to reveal tension present between the issues. These tensional connections are used to indicate dilemmas. The three types of dilemmas we analyzed are social/economic, social/environment, and economic/environment. This step is repeated for each of the issues in each of the drivers connected to each of the other issues in each of the other drivers. Once the team identified multiple dilemmas around each perspective, we were able to narrow down the choices to focus on the most relevant challenges that the industry is facing today. The end goal for the Baker Hughes Challenge Problem report was to identify these industry dilemmas, thereby allowing for research questions and, ultimately, Master's Thesis topics to be identified by the BHI scholars. A completed sustainability triangle for the recovery factor perspective is developed from the Baker Hughes Challenge problem and is shown in Figure 1.1. In the next section, a brief description of Figure 1.1 and how one of the tensions identified led to this thesis topic are presented.

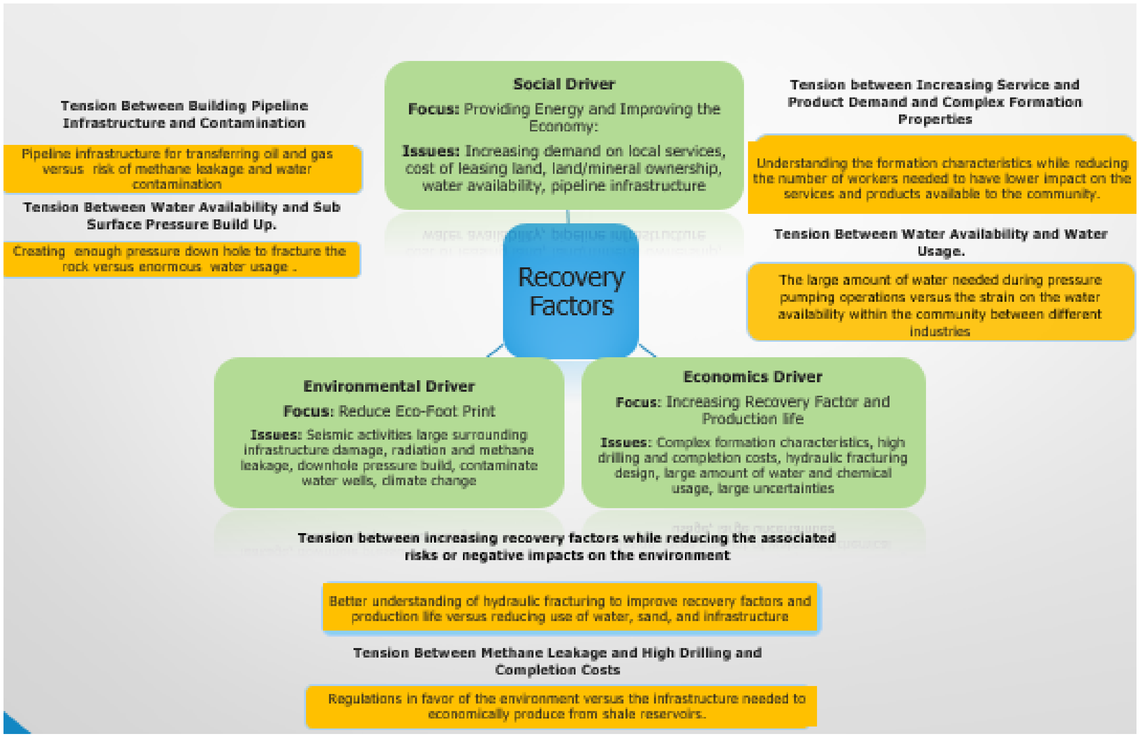


Figure 1.1: Completed Sustainability Triangle for the Recovery Perspective (Baker Hughes Challenge Problem Report)

1.1.3 Sustainability Triangle for Recovery Factors

A completed sustainability triangle for the recovery perspective is developed from the Baker Hughes Challenge problem and is shown in Figure 1.1. This figure is taken from the Baker Hughes Challenge Problem report and is used to show the tensions that have been identified for the Recovery Factors perspective. We have identified six tensions for the Recovery Factors perspective and each of these tensions can be developed into a research question for a Master's thesis. For example, we have identified a tension between the environmental driver and economic driver. The tension is between increasing the recovery factors and production life of shale wells and reducing the negative impacts on the environment. This tension arises from the fact that production of oil from shale wells requires hydraulically fracturing the formation and this process could potentially have negative impacts on the surrounding environment such increasing uses of water, sand, and infrastructure to transport equipment and material to the wellsite. Resolving this tension requires a better understanding of hydraulic fracturing and the key factors in the design of hydraulic fracturing that can be changed and used to reduce the negative impacts on the environment. We chose to further investigate this tension in this thesis. However, since the scope of this topic is too broad and big for a master's thesis, we decided to break down the topic into smaller subtopics that are better suited for a master's thesis. This tension is broken down into the following subtopics:

1. Increasing our understanding of hydraulic fracturing by identifying the key stimulation parameters

2. Establishing a relationship between the key stimulation parameters and the production performance

3. Identifying the optimal value of each key stimulation parameter that can be used to increase the recovery factors and production life of shale wells

4. Demonstrating how the key stimulation parameters can be manipulated to reduce the associated risks or negative impacts on the environment

We select subtopic number 1 to further investigate in this thesis. Thus the focus in this thesis is "Increasing our understanding of hydraulic fracturing by identifying the key stimulation parameters". The following are the reasons why we select subtopic 1:

- Personal interest in increasing our understanding of hydraulic fracturing
- Knowledge and experienced in hydraulic fracturing operation due to multiple internships that I have had at Baker Hughes with the completions group
- The high importance of hydraulic fracturing in shale development
- Actual field data is available to analyze and extract information from

The rest of the other subtopics that have been identified are potential master's thesis topics that can be further investigated in the future work.

In Section 1.2, an overview of hydraulic fracturing and unconventional reservoirs is provided (1) to provide a context for the thesis topic that has been identified and (2) to provide the reader an understanding of shale reservoirs and hydraulic fracturing process. In Section 1.3, the problem and the research questions are formulated in regards to hydraulic fracturing, and the proposed solution to the problem is presented.

1.2 BACKGROUND AND IMPORTANCE OF UNCONVENTIONAL RESERVOIRS AND HYDRAULIC FRACTURING: "SHALE BOOM"

1.2.1 Background and Importance of Hydraulic Fracturing: "Shale Boom"

The "Shale Boom" or "Shale Revolution" (Meisenhelder, 2013) that started mainly since 2008 has been a game changer and has reshaped both the US energy industry and the global energy landscape. “The US oil boom has had profound implications for the rest of the world, boosting economic growth and enhancing America’s global influence” (Crooks, 2015). Due to the massive unlocking of unconventional oil and gas reserves, the US has become the number one producer of the unconventional oil and gas in the world. As shown in Figure 1.4 and 1.5, since 2008, oil and gas production from unconventional reservoirs has emerged significantly and has continuously increased until the downturn in 2014. Gas is projected to increase by 49% of the total US gas production by 2035. “Production from tight oil plays surpassed 50% of the total U.S. oil production in 2015 when tight oil production reached 4.9 million (b/d)”, and “…tight oil development continues to be the main driver of total U.S. crude oil production, accounting for about 60% of the total cumulative domestic production in the Reference case domestic between 2016 and 2040” (EIA Annual Energy Outlook 2017). Furthermore, as shown in Figure 1.6, since 2008, the US net import of natural gas has drastically decreased and the US has gained more energy security and has gotten much closer to energy independence.

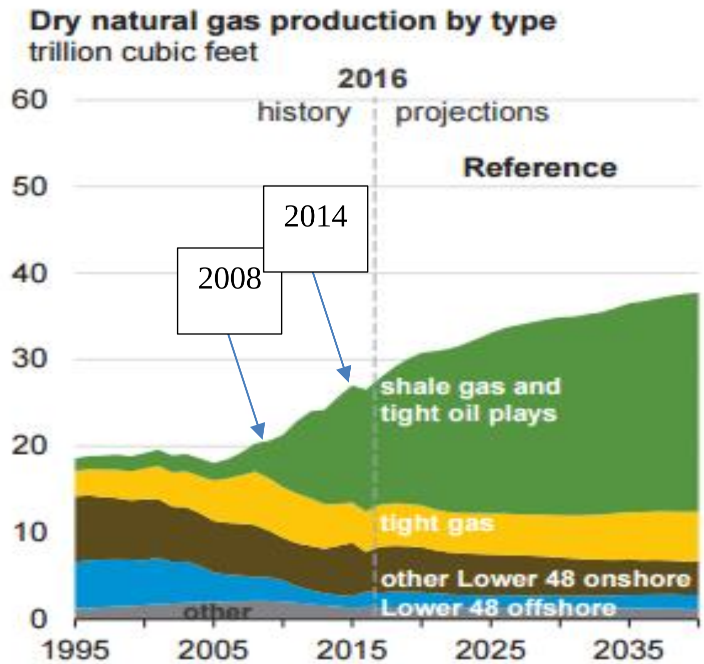


Figure 1.2: US Oil and Gas Production Historic Data and Projection (1995-2040) (EIA Annual Energy Outlook 2017)

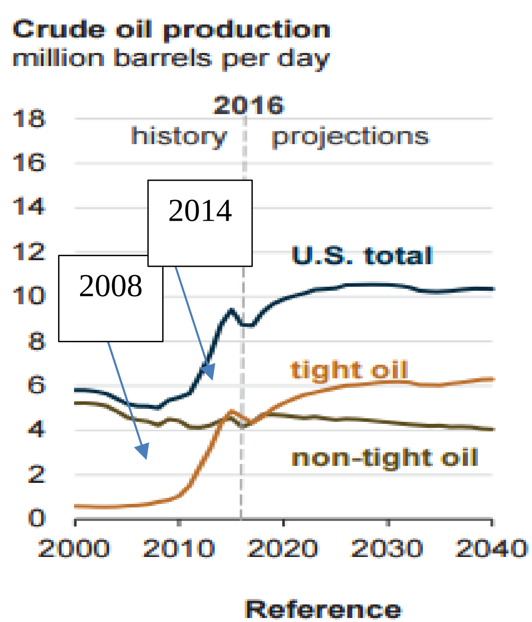


Figure 1.3: US Tight Oil Production Historic Data and Projection (2000-2040) (EIA Annual Energy Outlook 2017)

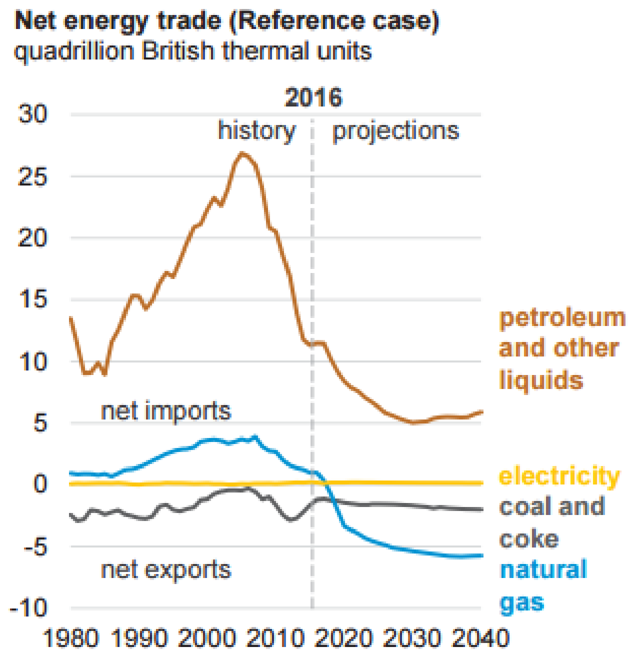


Figure 1.4: Impact of Shale Boom on US Energy Independence (EIA Annual Energy Outlook 2017)

In addition, the "Shale Boom" has greatly affected the job market in the oil and gas industry and has added a large number of jobs. According to a report by Reuters (2015), "A U.S. oil and gas drilling boom fueled by hydraulic fracturing technology added about 725,000 jobs nationwide between 2005 and 2012". In addition to such huge economic impacts, the "Shale Boom" has had political, social, and environmental effects as well, but they are beyond the scope of this work and will not be discussed here.

The main cause of "Shale Boom" is the recent technological advancement in Horizontal Drilling and Multistage Hydraulic Fracturing that enables drilling of horizontal wells that can reach to a large section of the reservoir laterally and hydraulically fracturing the formation to create fractures for the petroleum to flow through into the wellbore. Horizontal Drilling and Hydraulic Fracturing are defined and further discussed in Section 1.3.3. In the next section, the difference between conventional and unconventional reservoirs is explained to establish a better understanding of unconventional reservoirs and to define some of the terminology used in this section.

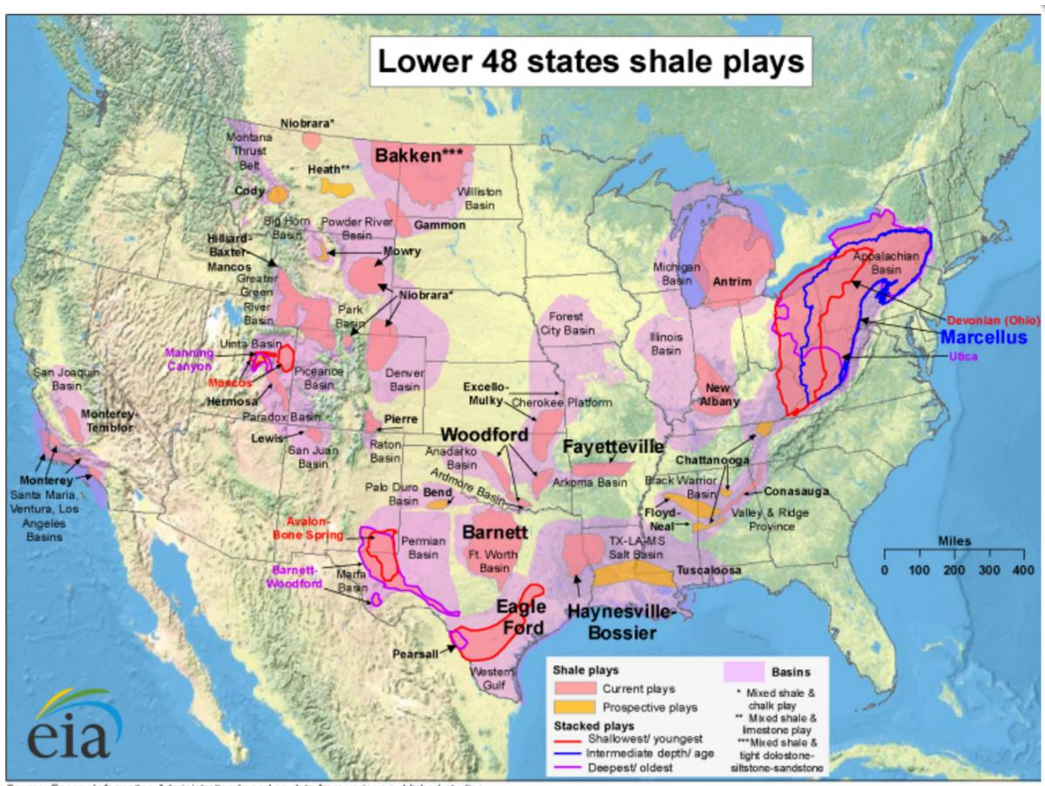
1.2.2 Overview of Unconventional Reservoirs in the Context of Petroleum Reservoir Systems and "Shale Boom"

In the field of oil and gas, reservoir refers to "that portion of the trapped formation that contains oil and/or gas as a single hydraulically connected system"**.** In a simpler term, a reservoir rock is a rock that petroleum migrates to and is trapped in. Petroleum reservoirs are classified into conventional and unconventional reservoirs. In the literature, to have a conventional petroleum reservoir, certain requirements must be fulfilled. The first requirement is to have a source rock which is a material from which petroleum is formed. Second, the petroleum must have been under enough pressure and temperature or have been "cooked" for it to flow. The third requirement is that the petroleum must have migrated from the source rock into a porous and permeable rock which is the fourth requirement of a conventional reservoir system.

Porosity of a rock refers to the void space in the rock and is used as an important parameter in the calculation of oil or gas volume in a rock. Permeability refers to how connected the pore spaces of a rock are and is an important parameter in the calculation of fluid flow rate. Last but not least, there needs to be a trap which is defined as a subsurface condition that is restricting or preventing the petroleum from further migration or movement so that it accumulates in a reservoir. Once these requirements are met, there is going to be an accumulation of hydrocarbons in a rock and those hydrocarbons can be brought to the surface by drilling a vertical well into the reservoir.

However, unconventional reservoirs do not meet these requirements. An unconventional reservoir refers to a rock that has petroleum in it and due to its low permeability, the petroleum has been trapped and not migrated to anywhere. In an unconventional rock there is not sealing or trap. The reservoir rock acts as a trap, a source rock, and a reservoir rock, and that is the difference between conventional and unconventional reservoirs. However, the major difference between the two types of reservoirs that affects how petroleum is being extracted from them is the permeability.

Permeability of a reservoir determines how easily or fast the fluids can be moved into the wellbore and brought to the surface. Conventional reservoirs are characterized as relatively high permeability reservoirs while unconventional reservoirs have extremely low permeability. Due to relatively high permeability of conventional reservoirs, when a vertical well is drilled into the reservoir rock, the fluids within the rock can relatively easily travel through the rock and reach to the wellbore. However, in unconventional reservoirs, since the permeability is so small, the fluids cannot travel and reach to the wellbore. This is the main reason why the industry has been producing from conventional reservoirs for hundreds of years with vertical wells but has not been able to produce from unconventional reservoirs until the recent years. The sources of unconventional reservoirs are shale oil or “tight oil”, shale gas, coalbed methane, tight sandstones, and methane hydrates. The most important of those are shale oil or “tight oil” and shale gas. The US is rich in shale. The US shale plays are shown in Figure 1.7.



Source: Energy Information Administration based on data from various published studies.
Updated: May 9, 2011

Figure 1.5: Shale Gas and Oil Plays in the US(EIA)

1.2.3 Overview of Horizontal Drilling and Hydraulic Fracturing

As mentioned earlier, the ability to produce from unconventional reservoirs is attributed to Horizontal Drilling and Multistage Hydraulic Fracturing. Horizontal Drilling refers to drilling of a well horizontally to reach out to a larger section of the reservoir. Hydraulic Fracturing is defined as the injection of fluids which mainly consist of water and sand into the reservoir at pressures higher than the formation pressure to fracture the formation open and create flow paths for the fluid to travel through and reach to the wellbore. These two technologies or methods have led to a successful production of oil and gas from unconventional reservoirs.

Due to such importance of horizontal drilling and hydraulic fracturing, there has been a lot of interest in this topic from both the academia and the industry. There has recently been a lot of studies and research to improve the efficiency and effectiveness of hydraulic fracturing while coming up with ways of reducing the environmental footprints. However, even though a lot of technological advancement in the area of horizontal and directional drilling has been made, there still remains a lot to be unknown about hydraulic fracturing (Centurion, et al., 2013). Hydraulic fractures are not very efficient or effective (Shelley, 2016). Improving the performance of a hydraulic fracturing technique in a specific formation is largely based on trial and error and past experience. There is a large demand for improving the efficiency and effectiveness of hydraulic fracturing to increase the oil recovery (World Oil, 2016).

In this section, an overview of the horizontal drilling and hydraulic fracturing is provided to show the importance of hydraulic fracturing in shale reservoirs, and a brief literature review is provided to support that the hydraulic fractures are not efficient or effective and there still remains a lot to be known about hydraulic fracturing. In Section 1.3, the problem with hydraulic fracturing is more specifically defined, and a more specific description of the hydraulic fracturing process is provided to indicate the challenges in making hydraulic fractures more efficient and effective.

1.3 PROBLEM FORMULATION AND THESIS OBJECTIVES

1.3.1 Problem Definition

Over the years, there have been several different hydraulic fracturing techniques developed such as Plug and Perf (PNP), Ball Activated Completion System (BACS), and Coil Tubing Activated Completion System (CTACS). However, Plug and Perf (PNP) is the most commonly applied and accepted technique of hydraulic fracturing across the industry (Casero, et al., 2013), and the operational step by step of the technique is described as follows: The production casing or liner is first run into the well. The process of creating fractures starts from the end of the wellbore called the toe of the wellbore to the beginning of the lateral length which is called the heel of the wellbore as shown in Figure 1.5. First, the perforation guns are sent downhole to perforate the casing, in other words, to punch holes into the casing so that there is communication between the wellbore and the formation. In each stage, a set of perforations are created and those perforations are called clusters. Once the clusters have been created, the hydraulic fracturing fluid which mainly consists of water and sand is pumped down at higher pressures than the formation pressure and at a specific injection rate to fracture the formation. Then the first plug is sent downhole and is set in place to provide through tubing isolation between the stages. Once the fractured stage has been isolated by the plug, the perforating guns are sent downhole again to create a set of clusters in the second stage. Once the clusters are created, the fracturing fluid is pumped down to fracture the formation in the second stage. This process of plug, perforate, and pump down fluid is repeated until all stages are fractured as shown in Figure 1.5. Once the process is complete, the plugs are milled out and the well is ready for production (Burton, 2013).

However, due to the fact that plugging, perforating, and pumping each single stage separately is very time consuming and costly, the industry has adopted the perforation cluster approach which basically means perforating and treating multiple stages or groups of cluster perforations simultaneously together (Casero, et al., 2013). However, when treating multiple stages or groups of clusters together, the fractures form and grow in the weak points of the formation since fracture forms at the path of least resistance. This creates unevenly distributed fractures and leaves some parts of the formation untreated (Burton, 2013). Thus, whether each stage or perforation clusters are treated separately or multiple stages are treated together is one of the important factors on how efficient and effective the fractures turn out to be.

The other variables that can be changed to affect the performance of the hydraulic fracturing technique are the lateral length, the number of stages, the space between the stages, the length of each stage, the number of clusters per stage, the space between the clusters, the rate of fluid injection, the fluid pressure, the amount of water injected, the amount of proppant injected, the type of fluid mixture, the type of proppant pumped, and a few more that is beyond the scope of this thesis such as perforating shot gun phases and density (Ferguson, et al., 2012).

When it comes to improving fracture efficiency and effectiveness, the above-mentioned parameters are utilized and changed. However, one of the main challenges in doing so is that there are too many parameters and there still remains a lot of uncertainty about which one of these parameters are the key to creating better fractures. For example, we cannot tell how the effectiveness and efficiency of a hydraulic fracturing technique or the production performance of a well is going to be affected if we increase the number of stages, the fluid rate, or the amount of proppant. It is uncertain which parameters need to be changed to increase fracture efficiency and effectiveness and ultimately increase the production performance of the well.

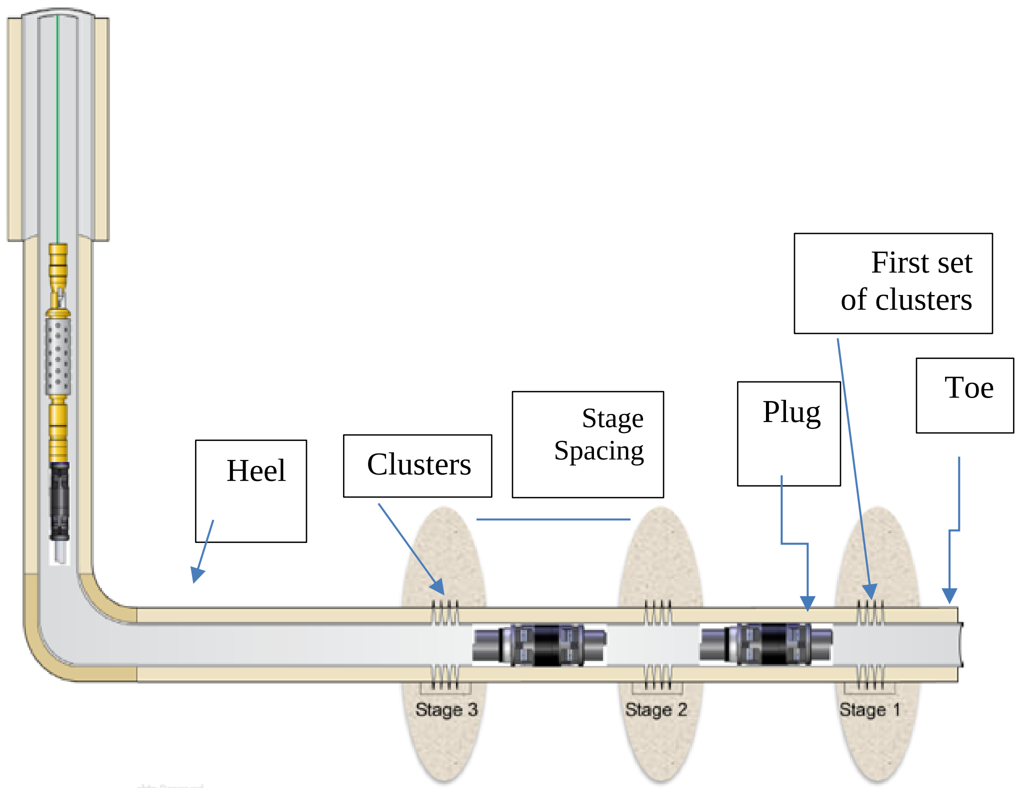


Figure 1.6: Configuration of a Hydraulic Fracturing Technique in Eagle Ford Shale (Anderson, 2014)

For example, Shelley and co-authors (Shelley, et al., 2012) conducted a data driven modeling study in 2012 to determine the best practices for completion of Eagle Ford Shale Wells. They analyzed and modeled the completion, frac, production, mud log, and lateral data of 55 wells, and they had 30 days cumulative production data to evaluate productivity. A preliminary analysis of the data indicated no apparent relationship between the completion parameters and 30 days cumulative production data but it showed a strong relationship between the depth of the wells and the 30 days cumulative production. Then they used artificial neural network (ANN) to model the data. The data was divided into 39 wells for training, 9 wells for testing, and 6 wells for validation. The model had 13 input parameters and used cumulative production volume per average estimated pressure drawdown for 30 days as a productivity metric.

In their study, they concluded that data driven modeling approach provides a pragmatic perspective and identifies the key completion parameters and ways to improve the effectiveness of completion techniques. They found geological and reservoir variations among the wells to have a dominant effect on Eagle Ford production, and in addition they identified the completion related parameters that affect well production and economics which are treatment fluid type/volume, number of frac treatments, proppant type/conductivity, perforating strategy, treatment rate and lateral length. The data driven model was also used to estimate the effect of alternative completion and frac scenarios on well productivity and it was found that economic evaluation of data driven model predictions can be used to determine appropriate completion and frac procedure to maximize return.

However, since Initial production (IP) is a poor indicator of long term well performance and can show reverse correlation (Taylor, et al., 2011), the results of the above study do not necessarily reflect the effect of the considered parameters on the long term well production performance.

Centurion and co-authors (Centurion, et al., 2013, 2014) performed a series of multivariate analysis of reservoir, well geometry, and completion related parameters for a large number of wells in Eagle Ford Shale using Linear Regression analysis. They concluded that multivariate analysis is a good technique for identifying the most important variables among the many factors that affect well productivity. They found well lateral length, stage spacing, proppant per ft, pressure, cluster spacing, thickness, average porosity, and perforation length to be the most influential factors that drive production. However, since they used only Linear Regression analysis, they were not able to identify any nonlinear relationship between the input variables and the output variable.

Gao and co-authors (Gao, et al., 2013) conducted a multivariate analysis of 9 variables (TVD, total proppant volume, lateral length, total volume of fluid, oil API gravity, GOR, flowing tubing pressure, midpoint of a well lateral in x and y direction) in 273 wells in Eagle Ford Shale using Multivariate Adaptive Regression Splines (MARS) which is a form of regression analysis that can model non linearity and interactions among variables. They used peak equivalent barrels of oil (PeakBOED) as a measure of early time production performance. They found that the maximum total vertical depth and the flowing tubing pressure have the most impact on early time production performance, and GOR and API gravity were found to have the next most impact on the early production performance. Contrary to the previously mentioned studies, they found proppant volume, lateral length, and frac fluid to have less contributions to the early time production performance. However, they concluded that reducing the ratio of proppant to fluid volume could increase the early time production.

Although they have considered the non-linearity and interactions between the variables, this study does not take into account the effect of most of the other hydraulic fracturing parameters such as number of stages, stage spacing, and number of clusters since it only includes 3 hydraulic fracturing parameters.

LaFollette and co-author (LaFollette, et al.,2014) used boosted tree regression model to analyze the production impacts of well location, well architecture, completion, and stimulation on the production metrics which are defined as best producing month in the first 12 producing months (BO) and barrels of oil per completed lateral length (BO/ft). An initial analysis of the input variables and output variables using scatterplot matrix showed no linear patterns between the input and output variables. Using relative influence to measure impact on the target variable, they found well location and GOR to have a strong impact on the production performance which agrees with most of the previously mentioned studies. Even though longer laterals led to higher total production, wells with longer laterals were found to be less efficient, and larger treatment with more proppant was found to be associated with better productivity.

Even though the results of this modeling study are similar to the results of the previously mentioned studies, the authors do not test or validate the method on a set of non-training data to determine the accuracy of the predicted values.

Viswanathan and co-authors (Viswanathan, et al., 2011) utilized neural network trained self-organizing maps and numerical simulations to evaluate different completion techniques and compare them to channel fracturing in Eagle Ford Shale. First month equivalent gas rate was used as key performance indicator (KPI). Scatter plots of the completion parameters and KPI did not show any strong correlation, but it did show a strong correlation between GOR and produced condensate rate per ft. Based on the neural network analysis, fluid volume per stage, proppant volume per stage, and 100 mesh sand were found to have a small impact on KPI while high proppant concentration and cluster spacing had a significant impact on the production performance. Even though the results of the model are compared and confirmed with the results of a simulation analysis, the study does not consider the effect of the parameters on the long term production performance.

There have been many other statistical analyses done in attempt to determine the key production variables and establish a relationship between input variables and output variable in other shale formations. Voneiff and co-authors (Voneiff, et al., 2013) used multivariate regression analysis in Montney Formation to determine how completion parameters affect production rates, and they found the number of stages and the number of perforation clusters per stage to be the most important completion parameters. Alqatrani and co-authors (Alqatrani, et al., 2016) used multiple regression modeling to study the influence of completion parameters in hybrid treated water and linear gel stimulations and treated water fracture stimulations in Glauconite Formation in southern Chile, and they found the total perforations to have the most impact in both cases. Ling and co-authors (Ling, et al., 2016) performed a statistical analysis using simple scatter plots to identify the optimal values of the completion parameters and were able to come up with optimal values for these parameters.

Thus, the problem addressed in this thesis is that **it is uncertain what the key stimulation parameters are that can be used to improve the efficiency and effectiveness of the hydraulic fractures.**

In this thesis, we analyze actual field hydraulic fracturing data for 65 wells in Eagle Ford Shale to determine the key stimulation parameters that have a direct impact on the production performance of the wells.

1.3.2 Research Questions

The principal question for this thesis is as follows:

What are the key hydraulic fracturing parameters that affect the amount of hydrocarbons that can be recovered in Eagle Ford Shale?

Even though the primary thesis question seems to be a simple and straightforward question that can be easily answered, there are multiple sub questions that need to be addressed and answered first to reach to an answer to the primary question. The sub questions are as follows:

1. ***How can the effect of a stimulation variable be measured?*** The first sub question regards determining a metric or parameter to measure the influence of a hydraulic fracturing variable. It is important to determine an accurate and realistic metric that can reflect the changes made by a stimulation variable. Some examples of the metrics used in the industry as it is shown in the studies mentioned in Section 1.3.1 are six months cumulative oil production, peak oil production, estimated ultimate recovery. However, each one of these has its own advantages and disadvantages, and selecting the right metric depends on the goals of the study and the type of data available to analyze. In Chapter 3, in the data understanding and preparation, the right metric for this thesis is determined.

2. ***How can the effects of geological and petrohpysical variables and reservoir fluid properties be eliminated or minimized?***

   Shale reservoirs are characterized as highly heterogeneous and anisotropic.

That means from one county to another and even from one well to another, the geological and petrohysical properties and the reservoir fluid properties could be different. Therefore, two wells next to each other that are drilled and completed exactly the same way could have different oil and gas production. Therefore, it is important that we eliminate or account for the effects of those variables early on in the study so that they don't mask the effects of hydraulic fracturing parameters. A method to mitigate the effects of this variability between the wells is discussed in Chapter 3.

3. What parameters of hydraulic fracturing should be considered for analysis? As mentioned in the earlier sections, there are a lot of parameters that go into a hydraulic fracturing technique, and attempting to consider and analyze all these parameters together is not feasible and could potentially influence the results of the study. Some of these parameters need to be eliminated early on in the study and the number of the parameters should be reduced to a manageable number without leaving out any important variables. A dimension reduction analysis is performed in Chapter 3 to answer this question and determine only the important parameters to further investigate.

4. What kind of data mining and statistical approach can be used to capture both linear and nonlinear relationships between the input and the output variables?

The relationship between the hydraulic fracturing parameters and the amount of hydrocarbons is not a simple direct relationship that can be easily captured with a two dimensional plot. It is important that we determine and use techniques that have the ability to capture both linear and nonlinear relationship between the hydraulic fracturing parameters and the amount of hydrocarbons that can be recovered. In Chapter 3, this question is answered.

5. ***How do we assess and compare the performance of the techniques?***

Once we have identified and used the regression techniques, the next step is to compare and assess the performance of the techniques to select the best regression technique for analysis of hydraulic fracturing data. This question is also answered in Chapter 3.

All the research questions are answered first in Chapter 3, and as these questions are answered, we get a better understanding of what the solution is going to look like and what methods can be used to get to the solution to the problem that has been addressed. In Chapter 4, the primary research question is answered.

The objective in this thesis is to determine an efficient method of analyzing and comparing multiple variables for multiple wells and a method that can be used to identify the key stimulation parameters.

1.4 PROPSED METHOD TO THE PROBLEM

1.4.1 Data Mining and Statistical Method

In this thesis, hydraulic fracturing is treated as a system in which multiple variables interact and affect the performance of the system. We analyze the impact of those variables and their interactions together on the performance of the system. To be able to do that, a technique is required that allows us to analyze multiple variables for multiple wells together in order to discover and extract unknown relationships and patterns in the data.

Data mining and statistical analysis which refers to the use of statistical tools and analytical techniques to “extract previously unknown, comprehensible, and actionable information from large databases” (Zekulin, 2015) is the proposed method for this study due to its ability for multivariate analysis. Data driven approach is a more pragmatic method to understanding the complex relationship between hydraulic fracturing and a non-homogenous formation, and it provides a higher level of understanding of the main drivers of hydrocarbon production. In addition, since data driven modeling is holistic and cross disciplined by nature, it is useful in providing direction for discrete evaluation of completion techniques (Shelley, et al., 2012). With 3-4 years of production and completion data, data mining can be used to possibly discover unrevealed trends and the underlying data structure (Gao, et al., 2013). LaFollette and co-authors have shown that multivariate statistical analysis allows the modeling of the impact of particular well architecture, completion, and stimulation parameters on the production outcome” (LaFollette, et al., 2014).

When it comes to data mining and statistical analysis, a popular approach is a method called CRISP_DM which stands for Cross Industry Standard Process for Data Mining. This is a systematic and structured way of conducting data mining studies and it increases the likelihood of getting accurate and reliable results (Delen, 2010).

The CRISP-DM is a six step process and is outlined according to the research questions in this thesis as shown in Figure 1.6. The first step is understanding the project and developing the goals of the project which we stated in Section 1.3. The second step is understanding the data quality, quantity, and relevancy to the problem which is discussed in Section 3.1. The third step is processing and preparing the data by answering the research questions which is presented in Section 3.2. The fourth step focuses on applying and evaluating the regression techniques which is discussed in Chapter 4. The fifth step is analyzing the results of the regression technique and drawing conclusions which is discussed in Chapter 5. The last step is critical evaluation of the concluding points and the method which is done in Chapter 5.

Thus, this thesis is a statistical analysis of hydraulic fracturing data. We are given field data and we use data mining and statistical tools and techniques to observe and extract trends and relationships from the data. Since this is not an analytical or numerical analysis, it is possible that the results may not exactly follow the laws of physics or the common understanding of shale formations and hydraulic fracturing, and that is one of the limitations of this approach as it is discussed further in the next section.

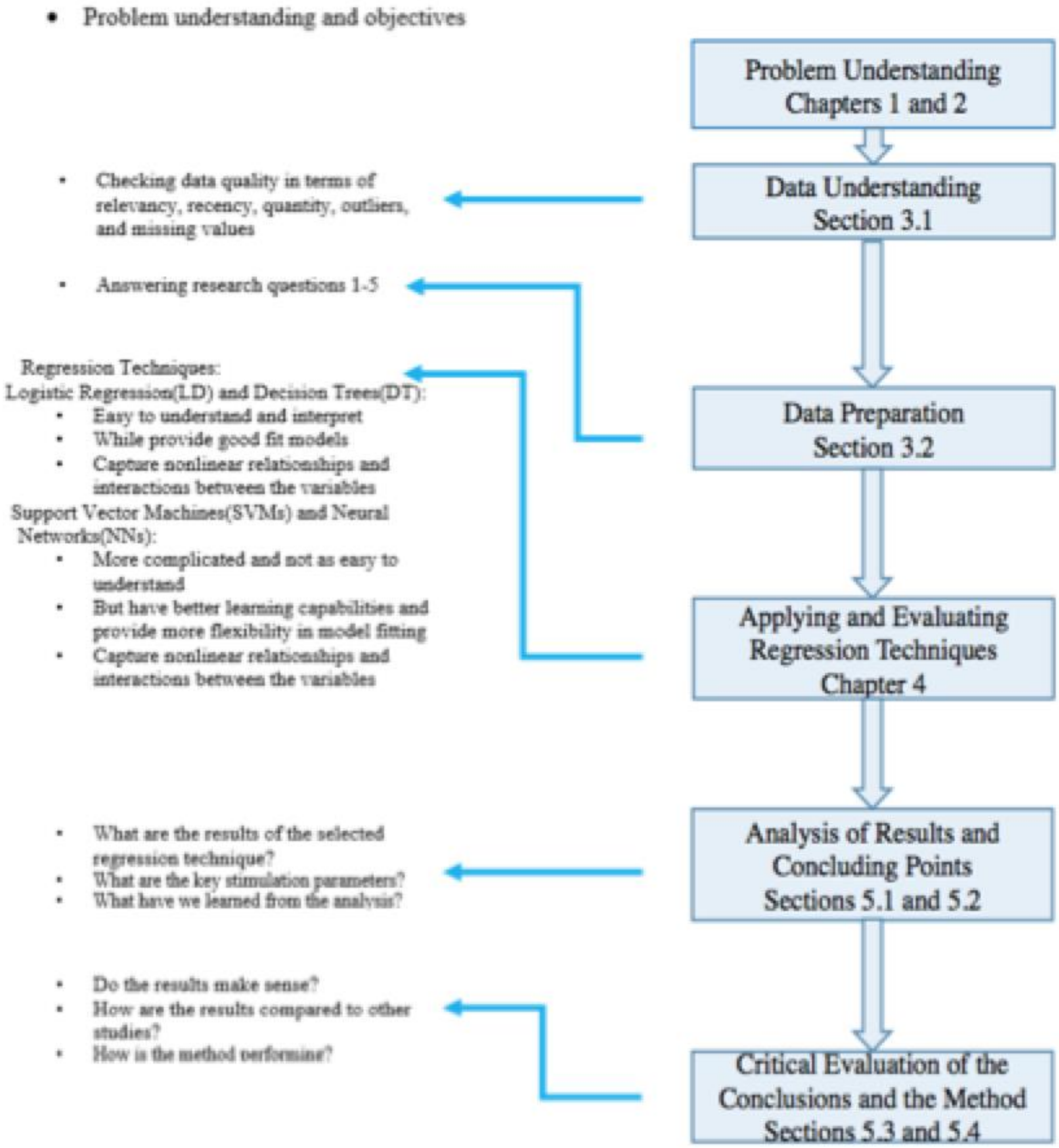


Figure 1.7: CRISP-DM for Analysis of Hydraulic Fracturing Data

1.4.2 Limitations of the Proposed Approach

When it comes to statistical analysis, having enough data plays a very important role. It is usually better to have more data in statistical analysis than small data for the sample to be more representative. However, due to the fact that most of hydraulic fracturing data is not publically available, there is limitation to how much data we can have access to and use in this thesis. The hydraulic fracturing data used in this thesis is provided by Baker Hughes, and it includes 65 wells in Eagle Ford Shale as shown in the Appendix Tables A.1 and A.2. The wells are distributed throughout 10 counties, and concentrated mainly in 3 counties. Due to confidentially agreement, the name and exact location of the wells cannot and will not be disclosed in this thesis.

Moreover, since the wells are spread out throughout the counties, we understand that it is very likely that geological, petrohysical, and reservoirs characteristics of the wells vary and that could affect the production performance of the wells regardless of how they are hydraulically fractured. However, since we don't have access to most of the geological and reservoir properties of those wells, we cannot include those parameters in our study. To mitigate the effects of the geological and reservoir properties, we have classified and normalized the wells based on their location and a few of the reservoir parameters that we have data for such as thickness, pressure, depth, and porosity. This is discussed further in Chapter 3 when answering research question 2.

The wells are also owned and operated by different operators, and that means it is very likely that the production method is different among the wells. This could also affect the production performance of the wells. However, since we don't have any knowledge of the production method of the wells, we cannot include this in our analysis.

Last but not least, the techniques that we use in this thesis are purely data driven and are not based on the physics of the formation and fluid properties. It is possible that we may have results that do not necessarily agree with the physics of rock formation and fluid flow. The results are closely studied and verified in Chapter 5.

1.5 ORGANIZATION OF THE THESIS

The organization of this thesis is shown in Figure 1.8. The first chapter of this thesis is designed as an introduction to the problem. In Section 1.1, the background and motivation for the problem along with an overview of the Baker Hughes challenge problem is discussed. In Section 1.2, an overview of the conventional and unconventional reservoirs, horizontal drilling and hydraulic fracturing is presented to provide context for the problem and define some of the oil and gas industry terminologies. In Section 1.3, the problem formulation and the objectives are presented. In Section 1.4, the proposed approach to the problem and the limitations of the approach are presented. In Section 1.5, the organization of the thesis is demonstrated.

In Chapter 2, a critical literature review and justification of the problem are discussed. In Section 2.1, the formation and fluid characteristics of the Eagle Ford Shale play are presented. In Section 2.2, the justification of the research questions is provided.

In Chapter 3, the data understanding and the answer to the research questions are presented. In Section 3.1, the data understanding process which includes the methods and tools for Section 3.1.1 analysis of data quality and Section 3.1.2 identifying outliers are explained. In Section 3.2, the data preparation process which includes answering the five research questions sequentially is presented.

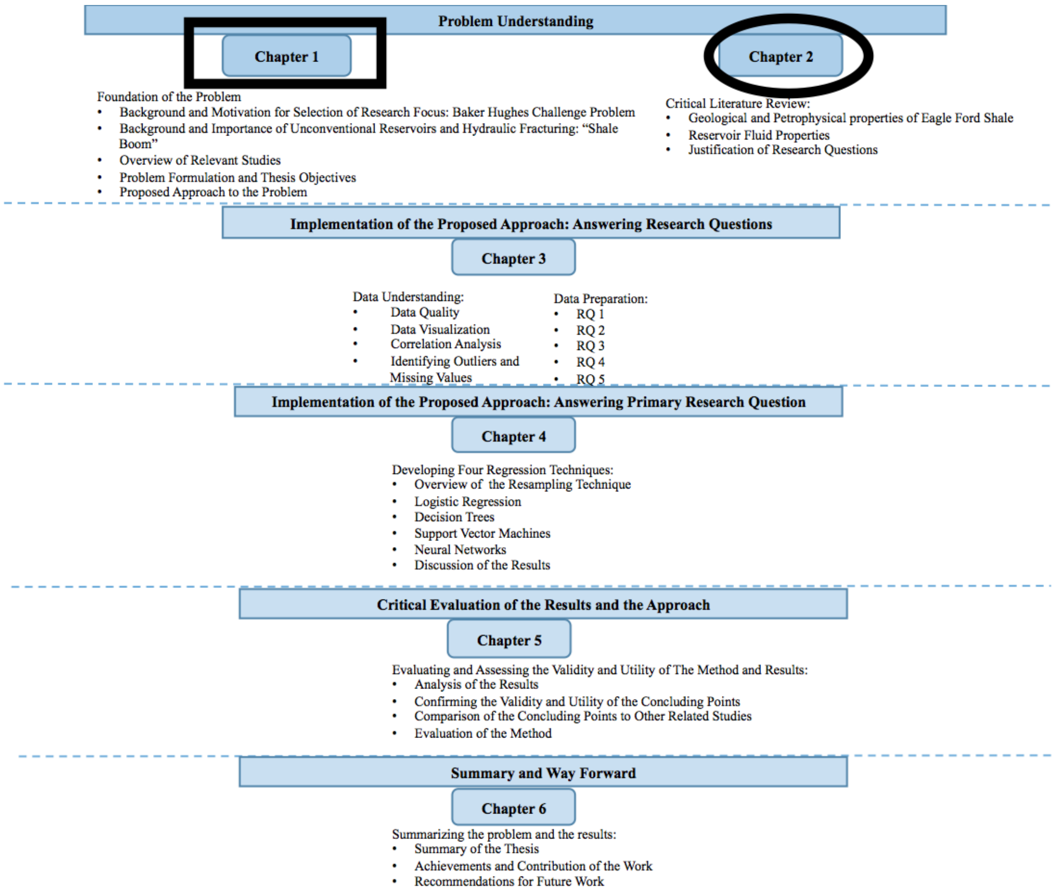


Figure 1.8: Organization of the Thesis -Present (Boxed) and Next (Circled)

In Chapter 4, the four proposed regression techniques and the resampling method are presented as follows: In Section 4.1 overview of the resampling method, in Section 4.2 Logistics Regression, Section 4.3 Decision Trees, Section 4.4 Support Vector Machines, and Section 4.5 Neural Networks. In Section 4.6, a discussion of the results of is presented.

In Chapter 5, the results and the method are evaluated. In Section 5.1, the analysis of the results is presented, and in Section 5.2, the validity and utility of the concluding points are confirmed. The concluding points are compared to relevant studies in the literature in Section 5.3, and the evaluation of the method is presented in

Section 5.4.

In Chapter 6, the thesis is summarized and reviewed to determine if the objectives have been met. Then the contributions are discussed and the recommendations for future work are presented.

CHAPTER 2

CRITICAL LITERATURE REVIEW AND JUSTIFICATION OF RESEARCH QUESTIONS

In Chapter 1, the problem along with the research questions are identified and the proposed method to the problem is presented. In this chapter, a critical literature review in regards to the problem is provided. In Section 2.1, an overview of the geological and reservoir properties of the Eagle Ford Shale which is the formation that is being studied in this thesis is presented. Understanding the formation characteristics and how they vary throughout Eagle Ford Shale is crucial to a better understanding of well performance. North America is rich in shale and the many shale plays that exist in the US have different formation and fluid properties that are crucial to the extraction of hydrocarbons. Most of the decisions regarding the drilling, completion, stimulation, and production methods are based on the knowledge of the formation and fluid properties available. "Success is dependent primarily on understanding the geology and reservoir properties" (Jaripatke, et al., 2013). In Section 2.2, the research gaps in the current studies and methods for analysis of hydraulic fracturing are presented and the research questions presented in Chapter 1 are justified. The justification of the research questions is provided based on the studies presented in Section 1.3.1. Once we have provided an overview of the Eagle Ford Shale formation characteristics and justified the research questions, we start looking at our data and answering the research questions which are done in the Chapter 3.

2.1 CLASSIFICATION AND ANALYSIS OF EAGLE FORD SHALLE FORMATION PROPERTIES

Eagle Ford Shale Play, approximately 11 million acres, stretches from the Texas border with Mexico east ward as shown in Figure 2.1. The production of the Eagle Ford Shale began in 2008 with the drilling of the first well in La Salle County, Texas (Fan et al., 2011). Since then, the number of drilled wells has tremendously increased and has led to a large economic development. From 2009 to 2013, the number of producing oil leases and gas leases has increased by about 6300 percent and 3600 percent, respectively (Munir, et al., 2017). Eagle Ford Shale is considered the largest economic development in the history of Texas and the largest oil and gas development in the world based on capital invested (Okeahialam, 2013). It has had an impact of more than $60 billion dollar on the local economy of Texas, and has added over 116 000 jobs to nearby areas of the play (Eagle Ford Shale, 2017).

The formation rock properties and fluid properties vary substantially (Cook, et al., 2014). The formation depth and thickness vary from northeast to southwest (Fan, et al., 2011). Further geological and petrophyiscal properties are discussed in the next section. The formation fluid is divided into three windows: oil, condensate, and gas as shown in Figure 2.1. Further discussion of the fluid properties of the formation and area of interest is provided in Section 2.1.2.

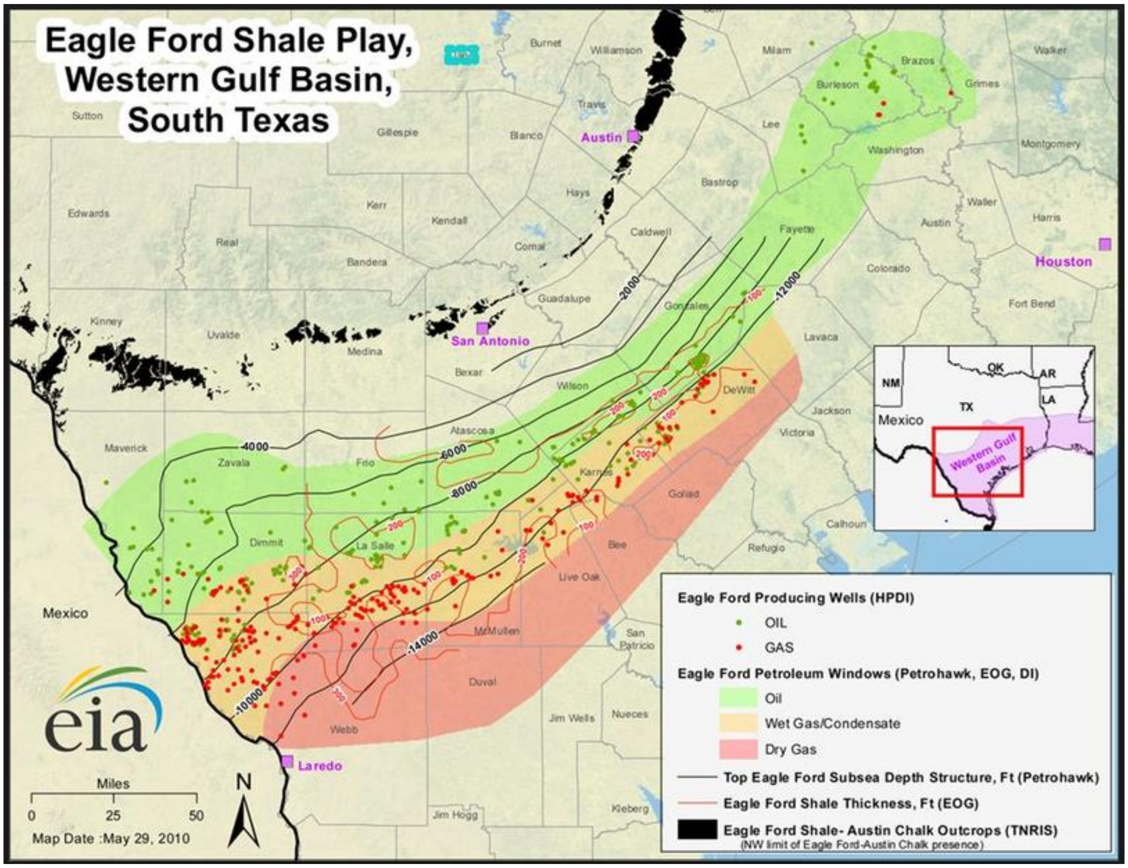


Figure 2.1: Location of Eagle Ford Shale Play (EIA, 2011)

2.1.1 Geological and Petrophysical Properties

Eagle Ford Shale was deposited in the late cretaceous period during which shelf slope, basin turbidities, and deltas formed along the shelf margin; and it affected the deposition of the Eagle Ford Shale and led to a laterally and vertically varying reservoir quality and thickness. The formation is bounded Austin Chalk from the top and Buda Limestone from the bottom. The play is divided into Upper Eagle Ford and Lower Eagle Ford units as shown in Figure 2.2 (Hentz, et al., 2011). The Upper Eagle Ford is characterized by relatively low gamma ray indicating low shale content while the lower Eagle Ford contains mainly high GR zones indicating high shale content and high TOC (Driskill, et al., 2012).

Due to regional dip, the formation depth increases from North to South, and as shown in Figure 2.1, the formation depth ranges from 4000 ft in the North to 14000 ft in the South (EIA,2011). Therefore, as it is discussed in the next section, we expect to have higher reservoir pressures in the South than in the North part of the formation due to higher depth of the formation in the South, and this information is useful to answer research question 2.

Moreover, as shown in Figures 2.3 and 2 .4, the thickness of both lower and upper Eagle Ford Shale decreases down dip from North to South and from West to East except that the Lower Eagle Ford thickens in the very east side and in the south of la Salle County. The total Eagle Ford thickness interval is the largest in the Maverick Basin and decreases to the minimum over San Marcos Arch (Hentz, et al., 2011).

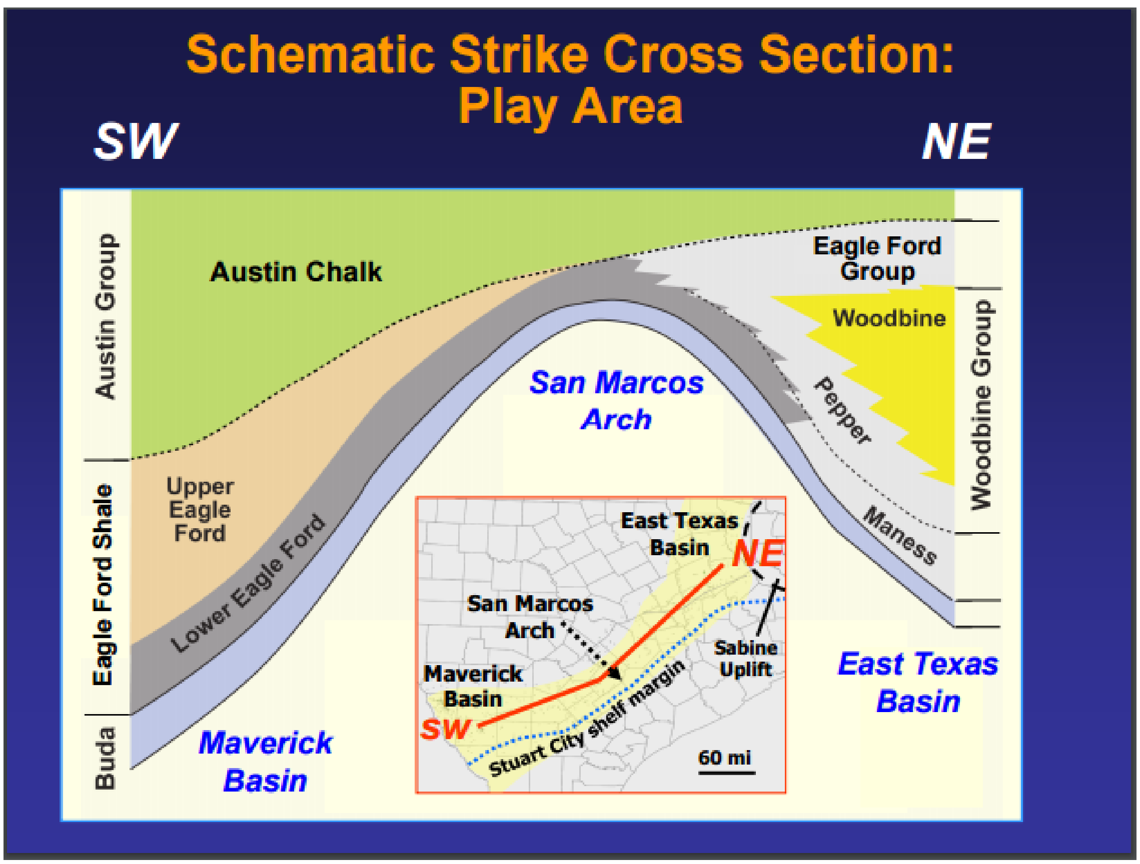


Figure 2.2: Schematic Interpretation of Eagle Ford Shale Play (Hentz, et al., 2011)

In overall Eagle Ford Shale, calcite ranges from 40% to 68%, clay content is 15% on average, quartz and feldspar content is roughly 15%, and TOC is 4% on average (Mullen, 2010). The carbonate content of the formation can be as high as 70% (McMillan, et al., 2016) and such high carbon content and subsequently lower clay content makes Eagle Ford Shale more brittle and well suited for hydraulic fracturing (Eagle Ford Shale, 2017)

Effective porosity is between 3% to 10% while permeability ranges between 3 nd and 405 nd (Martin, et al., 2011). However, low gamma ray intervals which is a characteristic of the Upper Eagle Ford Shale are brittle rocks rich in calcite, while high gamma ray intervals which is a characteristic of the Lower Eagle Ford Shale have carbonate and high clay and TOC content and are ductile rocks (Tian, et al., 2013, 2014). TOC, maturity, and kerogen type are the critical parameters that control good shale source rocks (Passey, et al., 2010). Thus, according to this, the lower Eagle Ford Shale is a good source rock. However, since the Upper Eagle Ford is rich in calcite and is brittle, it is a better target for hydraulic fracturing because rock brittleness is a key factor for effective completions and fracture treatment (Rickman, et al., 2008). Since the upper Eagle Ford is the thickest in the south west, we expect to have a better fracture performance among the wells located in the southwest than the north east. This information comes handy when we compare group 1wells located in the south to group 3 wells located in the north east.

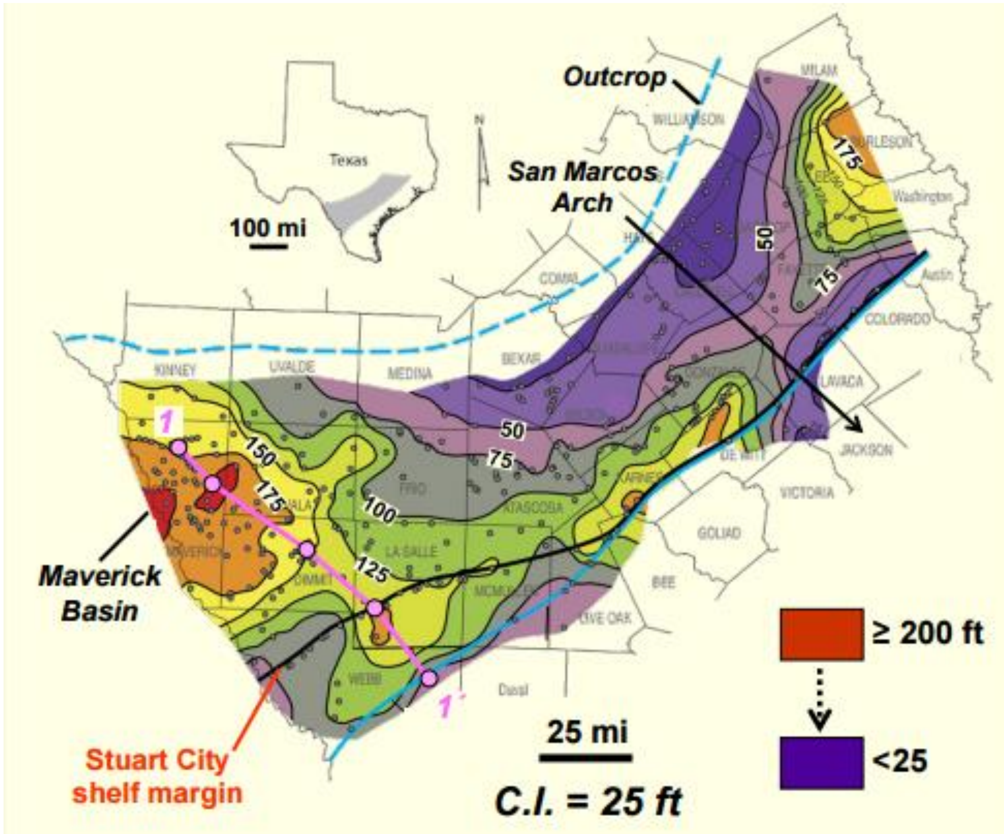


Figure 2.3: Isopach or Thickness of Lower Eagle Ford Shale (Hentz, et al., 2011)

Since our wells are spread out from south west to north east, we expect to have a large variability among our wells in terms of reservoir depth, thickness, clay and TOC content, and calcite and carbonate. These properties could impact well performance. Since the wells located in the south west are deeper and thicker and are overall richer in TOC, they are expected to have higher production performance assuming everything else is the same. This variability in geological and pterophysical properties could mask the impact of the hydraulic fracturing parameters. The methods to mitigate this effect are discussed in Chapter 3.

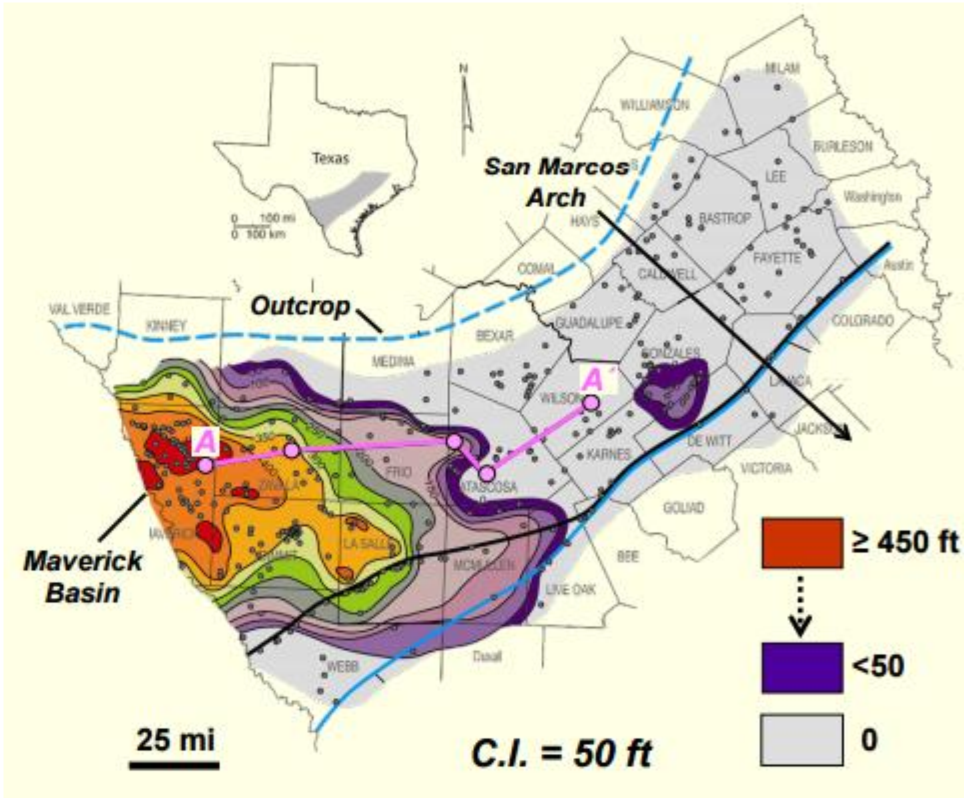


Figure 2.4: Isopach or Thickness of Upper Eagle Ford Shale (Hentz, et al., 2011)

2.1.2 Reservoir Fluid Properties

The Eagle Ford fluid types change from black oil, to volatile oil, gas condensate, and to dry gas with increasing depth and thermal maturity (Tian, et al, 2012). Reservoir pressure usually increases with increasing depth and pressure gradient. In Eagle Ford Shale, the reservoir pressure increases from less than 6000 psi at depth of approximately 7500 ft in the Southern Dimmit County to more than 10000 psi at a depth of approximately 12000 ft in Karnes and DeWitt Counties in the northwest (Tia, et al., 2014). The pressure gradient compares well with the reservoir pressure as it was found that the reservoir pressure gradient increases from less than 0.68 psi/ft in the southeast of Eagle Ford Shale to approximately 0.85 psi/ft in the northeastern (Tia, et al., 2014).

The oil API gravity in the Eagle Ford Shale increases from 40 in northwest to 65 in southeast as shown in Figure 2.5 (McMillan, et al., 2016), and the gas specific gravity increases from 0.6 in southeast to more than 0.85 in northwest as shown in Figure 2.6 (Tia, et al., 2015). These maps show that the thermal maturity of the reservoir fluid increases from northwest to southeast which corresponds to the increasing depth trend mentioned earlier.

Based on these trends, we expect the wells located in the southwest to have higher reservoir pressures and pressure gradient and high API gravity with greater thermal maturity which could potentially contribute to a better production performance of these wells.

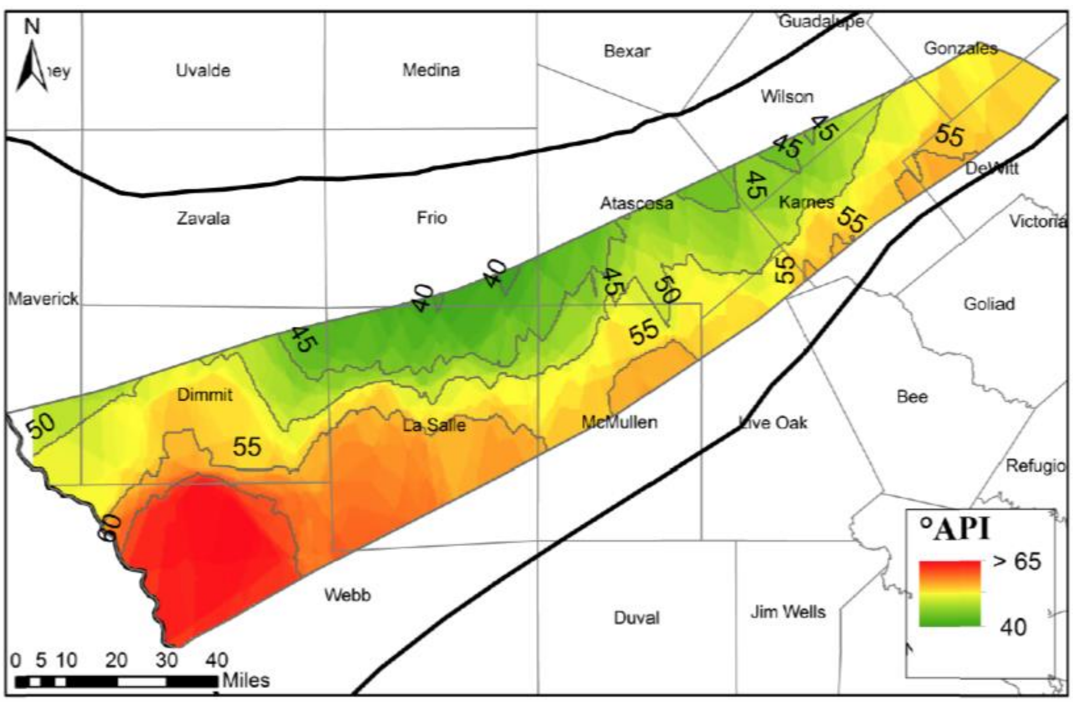


Figure 2.5: Oil API gravity in Eagle Ford Shale (McMillan, et al., 2016)

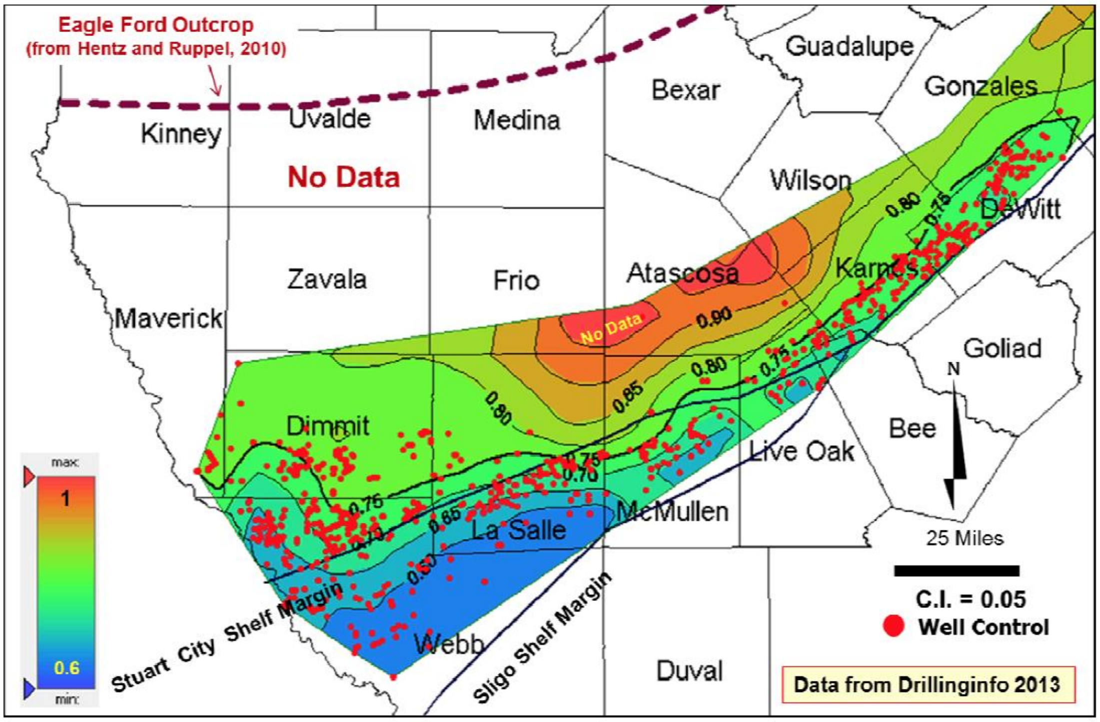


Figure 2.6: Gas Specific Gravity in Eagle Ford Shale (Tia, et al., 2014)

2.2 IDENTIFYING RESEARCH OPPORTUNITIES AND JUSTIFICATION

In this chapter, a literature review of the properties of the Eagle Ford Shale Play is presented. In Section 2.1.1, the geological and petrophysical properties of the Eagle Ford Shale Formation are discussed and analyzed. The variability in Eagle Ford Shale in terms of depth, thickness, gamma ray, TOC content, brittleness, and porosity, and permeability are discussed, and how each of these variables could affect the performance of our wells is presented. In Section 2.1.2, the fluid properties of Eagle Ford Shale Formation are discussed, the distribution and potential impact of reservoir pressure, pressure gradient, GOR, oil API gravity and gas specific gravity, and thermal maturity are demonstrated. In this section, the research opportunities identified through critical literature review are presented and the connection between the identified research opportunities and research questions proposed in Chapter 1 is established.

RQ 1. How can the effect of a stimulation variable be measured? In the studies presented in Section 1.3.1, in almost all the studies, a different metric as a measure of production performance is used. Some of them are using and considering peak production or first month production to be a good indicator of performance while some others use 6 months cumulative production or best months in 12 months production. It is clear that there no consensus or established criteria to determine a good performance indicator even though selecting this value is critical to the entire study. It can be concluded that there is a need to establish a good performance indicator which is considered in this thesis.

RQ 2. How can the effects of geological and petrohpysical variables and reservoir fluid properties be eliminated or normalized? In Section 2.1, the geological, petrophyiscal, and reservoir fluid properties are explained and the potential impact of each of those properties are discussed. In addition, some of the studies presented in Section 1.3.1 demonstrated the influence of geological and reservoir properties such as depth and GOR of the wells. It is clear that there is a strong relationship between the geological and reservoir properties and production performance in Eagle Ford Shale. To understand the impact of hydraulic fracturing parameters, the influence of the geological and reservoir parameters need to be eliminated if possible, or minimized. The methods to mitigate the impact of the reservoir related parameters are discussed in Chapter

3.

RQ 3. What parameters of hydraulic fracturing should be considered for analysis? The studies presented in Section 1.3.1 also show that even though the objective of all of them is to determine the key production parameters, the studies selected somehow different stimulation parameters to analyze and consider. It is clear that there is a lot of hydraulic fracturing related parameters, and acquiring data for all these parameters is not possible. The hydraulic fracturing parameters analyzed in this thesis are presented in Chapter 3.

RQ 4. What kind of data mining and statistical approach can be used to capture both linear and nonlinear relationships between the input and the output variables? Some of the data mining and statistical approaches used to identify the key stimulation parameters are discussed in the studies presented in Section 1.3.1. It was shown that in some of the studies, logistic regression or simple scatter plot methods are used, both of which do not have the ability to capture non linearity between the input and output variables. While in some of the other studies a modified version of logistic regression or neural networks that have the capability to capture non linearity are used, they did not consider dividing the data into training and test data to be able to test and validate the methods. Furthermore, none of the studies considered using more than one regression technique on the same data to see the difference between the different techniques and determine the best data driven method for hydraulic fracturing analysis. It is important that a comparison of some of the powerful statistical techniques is done to determine the best performing technique. Several techniques are used and discussed in Chapter 4.

RQ 5. How do we assess and compare the performance of the techniques?

None of the studies reviewed in the literature and presented in Section 1.3.1 used more than one technique to analyze hydraulic fracturing and thus they didn't present any performance assessment techniques. Several performance assessment techniques are discussed and used in Chapter 3 to determine the best performing regression technique.

2.3 SYNOPSIS OF CHAPTER 2

In this chapter, a literature review of the Eagle Ford Shale Formation properties and the statistical analysis of hydraulic fracturing are presented. In Section 2.1.1, the geological and petrophysical properties of the Eagle Ford Shale and the variation of those properties throughout the formation are discussed. In Section 2.1.2, the reservoir fluid properties of the Eagle Ford Shale and how the variation in those properties could affect the production of performance of our wells are discussed. In Section 2.2, the justification of the research questions is presented. Now that the problem and research questions are defined, context and justification of the research questions have been provided, and an overview of the geological and reservoir properties of the formation are presented, in the next chapter, we start analyzing the data and use it to answer the research questions. In Section 3.1, the data quality is first examined to make sure the data is valid and well suited to answer the research questions. In Section 3.2, the data is used to answer the five research questions, and then in Chapter 4, the primary research question is answered.

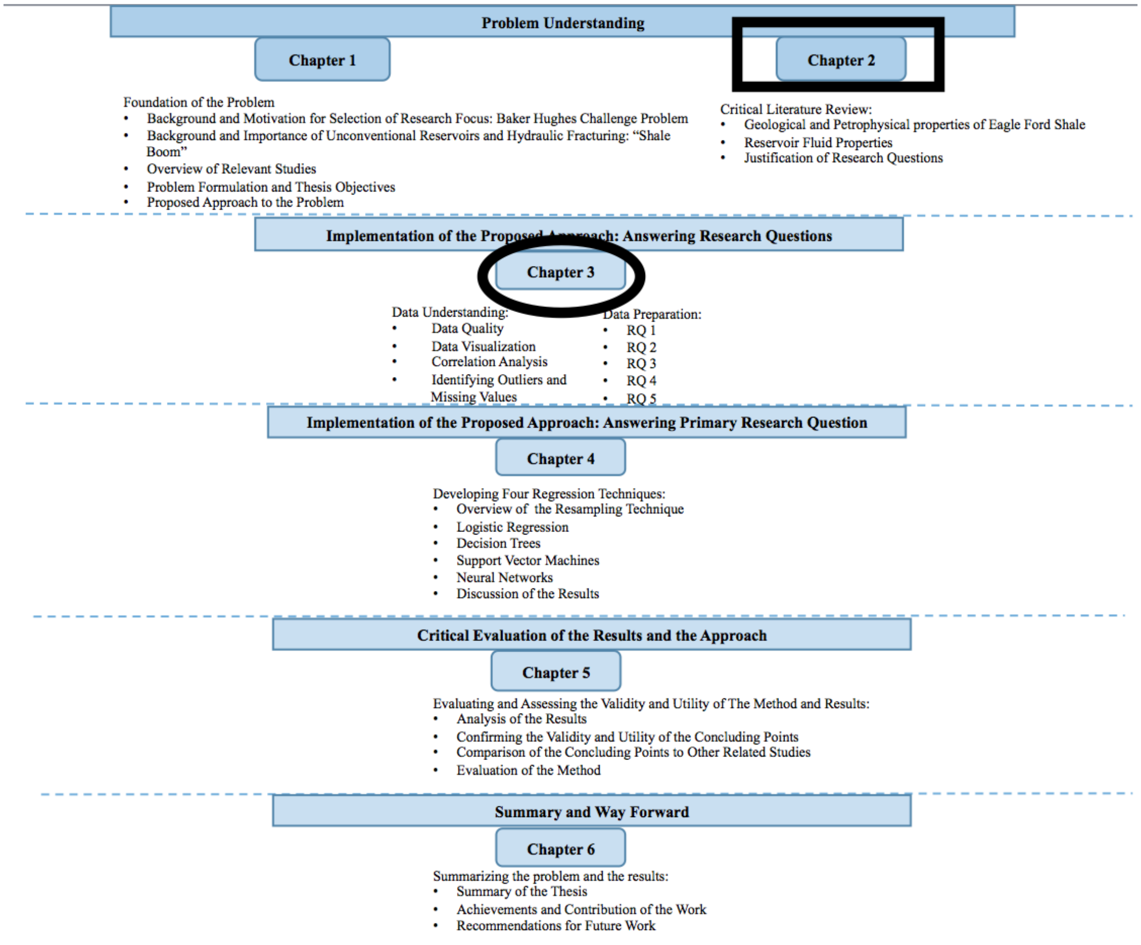


Figure 2.7: Organization of the Thesis -Present (Boxed) and Next (Circled)

CHAPTER 3

DATA UNDERSTANDING AND PREPARATION

In Chapters 1 and 2, the problem and the research questions are defined. Based on the literature, an overview of the Eagle Ford Shale formation characteristics which are crucial for understanding well performance is presented. Based on the studies reviewed in the literature, the research questions are justified and the research gaps are identified. In the previous two chapters, we cover the problem understanding, and in this chapter, we cover the data understanding. In addition, the research questions are answered in this chapter and the data is prepared for analysis which is done in Chapter 4 to answer the primary research question. Understanding and preparing the data is an important step in data mining and statistical analysis. Data understanding and preparing can affect the performance of a regression technique in terms of its classification and prediction capabilities. Data understanding and preparation refer to the improvement of data quality by visualizing, adding, cleansing, deleting, and transforming the data. This chapter is divided into two major sections. In Section 3.1, data understanding which refers to the process of exploring the data and identifying any outliers, missing values, and discrepancies to gain insight into the data in general and with respect to the objectives of this thesis is presented. In Section 3.2, the data preparation which in this thesis refers to the process of answering the research questions and preparing the data for regression analysis. In this section, the five research questions are answered, and in the next chapter, the primary research question is answered.

3.1 DATA UNDERSTANDING

Data understanding helps us understand the data better and gain insight into the data quality, quantity, recency, relevancy, and validity. This process helps us answer the following questions regarding the data:

- What kind of data do we have available?
- Is the data quality, quantity, and recency sufficient?
- Is the data relevant to the problem?
- Are there any outliers or missing values?
- How is the data distribution?
- Does the data suit the problem?

To answer these questions about the data, we first explore the data quality in the next section.

3.1.1 Data Quality

In this section, we examine the quality of the data by checking the accuracy, completeness, and timeliness of the data. High quality of data is crucial to a good data analysis. If the data has low quality, it is hard to trust the results of the analysis as the saying goes “garbage in, garbage out”.

We first look at the accuracy of the data. We compare our data to the data available in Drillinginfo to see if there are any errors or discrepancy in the data given to us. In DrillingInfo, we are able to find data about the production history of the wells, the depth of the first and last perforation, the perforation interval, and the completion date of the wells. We compared our data against the data available in DrillingInfo for these parameters, and we found out that there is a large discrepancy between the production data that we have and the production data available in Drillinginfo for most of the 65 wells. The reason for such a big discrepancy is that we found out the production data given to us is production data per lease not per well even though it has been recorded as a production data per well. A lease usually has more than one well, and thus the production data for a lease is usually much higher than the production data for a well. That also confirms why we see the production data that we have is much higher than the one we get from DrillingInfo for each well. It is worth mentioning that DrillingInfo reported only production data per lease until recently they released a new software called Production Workspace that reports production data per well and per lease.

Another discrepancy that we found between the data that we have and the data reported in DrillingInfo is in the depth of the first and last perforation of some wells. This discrepancy occurs only for a few wells, and we corrected this information according DrillingInfo as its data made more sense when we compared the depth of those wells to the surrounding wells.

After we confirmed the accuracy of as much of the data as we had access to, we looked at the completeness of the data. We checked our data to see if there are any missing records or missing attribute values. Using R programming, each column of the data is checked and reported as shown in Table 3.1. The number in front of each variable represents the number of missing values for that variable. As it can be seen, all the data is complete except for the production related variables. There are missing values for all the production related variables except the PEAK_BOERC and CUMGOR_RC which stand for peak equivalent barrels of oil and cumulative gas and oil ratio, respectively. The reason for the missing values for the production related variables could simply be that those wells haven't been on production for more than a month at the time this data was recorded. These wells have a complete data for peak oil which means they have been on production when this data was recorded, but they haven't been on production for long enough to have 3 months or more of production data.

To resolve the missing values and the discrepancy issues in production data, we accessed the production data of the 65 wells from Production Workspace in DrillingInfo and used it to replace our production data.

The timeliness of the data is another quality aspect of the data that is important in regards to the problem. The data is checked to see how recent is the data and if it reflects the current nature of the problem domain. The completion date or the last frac date which tells us about the timeliness of the data is available for all the wells and is checked to see how recent is the hydraulic fracturing of these wells. All the wells are completed between 2009 and 2011. This tells us two things: first, the data is recent enough in regards to the problem domain and can be used to approach the problems addressed in this thesis. The fundamental hydraulic fracturing technique is still the same and the problems associated with it and addressed in this thesis remain to be unsolved.

Second, it means that we have at least three years of production data for each one of these wells assuming that they all came on production soon after their completion date.

Now that we have this data that is complete, accurate, and recent, we can use it to derive a metric for long term production performance. 3 years of production data is long enough to reflect the long term production performance of the wells. In Section 3.2, we define the long term production and discuss this further to answer the first research question.

Table 3.1: Number of Missing Values in Each Variable of the Data Given

Variable	Count
Watergals_perWell	0
Watergals_perStg	0
MaxTreatPresspsi_perWell	0
AvgTreatPresspsi_perWell	0
AvgInjRatebpm_perWell	0
Proppanttlbsperwell_no100num	0
TotalProppanttlbsperstage_no100num	0

Missings per variable:

Variable	Count
WELLAZIMUTH_CAL	1
Max_TVD	0
Horiz.length_ft	1
CorrectedTVDs	0
PETROPHS_RCGROSS	0
PETROPHS_RCNET	0
PETROPHS_RCNGR	0
PETROPHS_RCPHIH	0
PETROPHS_RCSOPHIH	0
PETROPHS_RCPHIA	0
PETROPHS_RCTOCH	0

Missings per variable:

Variable	Count
Perforatedlength	0
FirstFracDate	0
Stages	0
Clustersperstg	0
TopPerfftperWell	0
BottomPerfftperWell	0
StgLengthft	0
ClusterSpacingft	0

Missings per variable:

Variable	Count
PEAK_BOERC	0
CUM_OIL_3MO	4
CUM_OIL_6MO	15
CUM_OIL_12MO	47
CUM_WATER_3M	4
CUM_BOE_3MO	4
CUM_BOE_6MO	15
CUM_BOE_12MO	47
CUMGOR_RC	0

3.1.2 Outliers

An outlier is defined as “an observation that appears to deviate markedly from the other observations in the sample” (Engineering Statistics, 2013). It is important to detect outliers because an outlier could indicate bad data or could indicate something exceptional or interesting that may need further investigation. Two different methods are selected to detect outliers: boxplot and the generalized extreme studentized deviates(ESD).

Boxplot (box and whisker diagram) is a visualization technique to identify outliers in the data. It is a graph of the data that displays a line drawn from the minimum value to the maximum value. Within that a box with lines drawn at the first quartile, the median, and the third quartile is depicted as shown in Figure 3.1. The diagram on the left is a simple boxplot with no outliers while the one on the right is a more complicated boxplot with outliers and potential outliers. An outlier is either 3 x IQR or more below the first quartile or 3 x IQR or more above the third quartile, and suspected outliers are either 1.5 x IQR or more below the first quartile or 1.5 x IQR or more above the third quartile (Hoffmann, 1981).

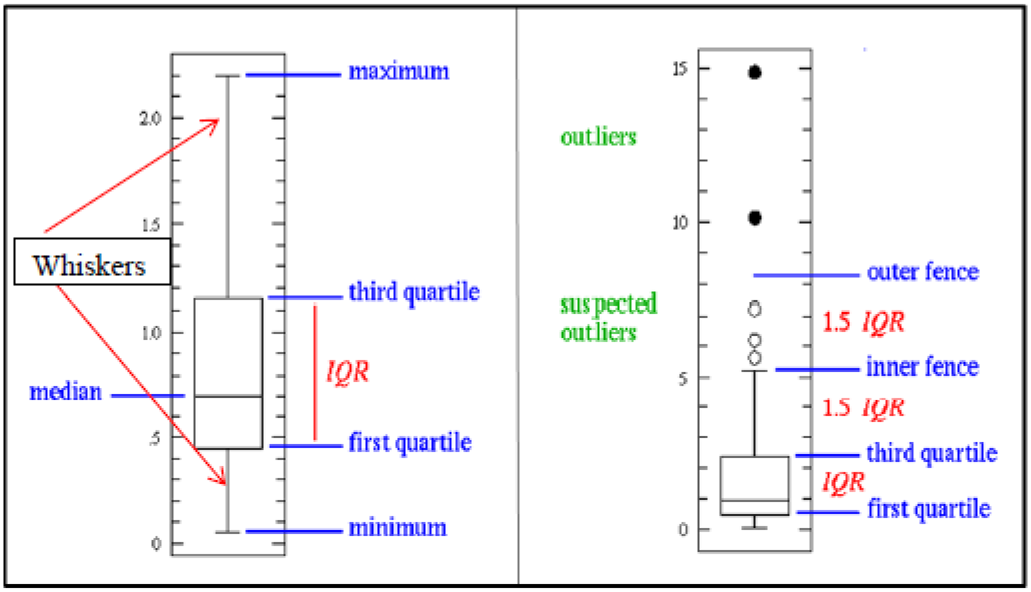


Figure 3.1: Boxplot Diagram (Hoffmann, 1981)

The ESD is used to detect one or more outliers. Given an upper bound r, ESD performs r separate tests to identify r outliers. The generalized ESD test is defined for the hypothesis as follows:

H_0: There are no outliers in the data set

H_a: There are up to r outliers in the data set

Test Statistic: Compute:

$$R_i = \frac{\max_i |x_i - \bar{x}|}{s}$$

where x and s are the sample mean and standard deviation, respectively. The critical values are calculated as follows:

$$\lambda_i = \frac{(n-i)\, t_{p,\, n-i-1}}{\sqrt{(n-i-1+t^2_{p,\, n-i-1})(n-i+1)}} \qquad i = 1, 2, \ldots, r$$

Where $t_{p.v}$ is the 100p percentage point from the t distribution and v is the degrees of freedom.

The number of outliers is determined by the largest i such that $R_i > \lambda_i$ (Engineering Statistics, 2013)

We first use the boxplot technique to visually check the outliers as shown in Figure 3.2. We found five outliers and confirmed it with the ESD test. Shown in Figure 3.2 is a boxplot on top of a scatterplot to show the five outliers compared to the rest of the data points. The x-axis is the number of wells and y-axis is the cumulative oil of each well. Thus, we have five wells that have production performance way larger than the rest of the wells.

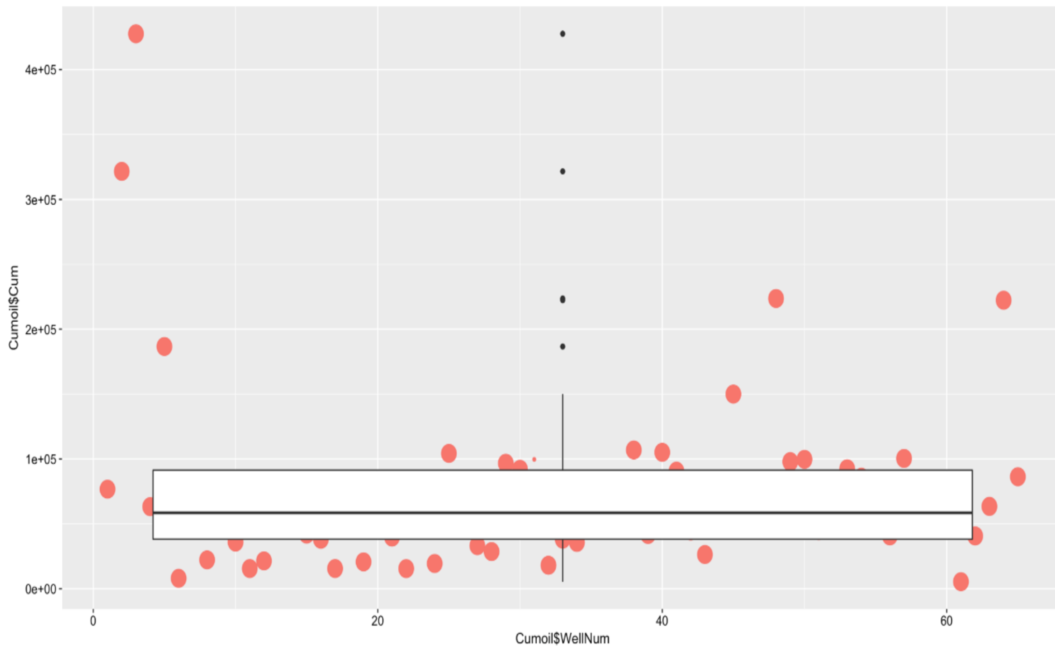


Figure 3.2: Boxplot on Top of a Scatterplot Showing Five Outliers

We decided to investigate why we have such high production for these five wells that make them outliers. We went through our data and checked all the parameters for these three wells, and compared them to the rest of the other wells. There wasn't anything drastically different between those wells. We then checked those five wells in Drilling Info, and we found out that three of those wells have been side tracked as indicated in DrillingInfo by the change in the last two digits of the API number. The production data reported for these wells includes the side track production which adds more to what the wells were originally producing. That explains why these three wells have such a high well performance. We decided not to include these wells in our analysis as it will not be a fair comparison with the other wells.

The other two wells have slightly less production performance compared to the three outlier wells. However, we haven't found anything drastically different between the two wells and the rest of the other wells on DrillingInfo to explain why they have such high production performance. We cannot include those two wells in our analysis as they skew the data largely and affect the results of the model, and we have no explanation as to why they have such high production compared to the rest of the other wells. Therefore, including them in the analysis will not add any value as we will still have no explanation for their high performance.

3.2 DATA PREPARATION

In this section, the five research questions are answered by using the data and the statistical tools and techniques such as scatterplot matrix. By answering the research questions, we are also preparing the data for regression analysis which is done in the next chapter. We need to first establish a performance metric to be able to measure the impact of the hydraulic fracturing parameters. We then need to eliminate or minimize the geological and reservoir related differences between the wells so that we can compare wells with similar reservoir and geological properties. This then takes us to the third research question which is selecting a manageable number of hydraulic fracturing parameters for analysis. Once we have established a performance metric, minimized the geological and reservoir differences between the wells, and selected the hydraulic fracturing parameters, we select the regression techniques that are suitable for multivariate analysis of hydraulic fracturing data. Last but not least, the performance assessment techniques to compare the regression techniques are discussed in this section.

3.2.1 Research Question 1: *How can the effect of a stimulation variable be measured?*

The first research question regards determining a metric that can be used to indicate long term production performance of a well. The metric is used to compare the performance of multiple wells and determine the impact of changing the hydraulic fracturing parameters. As mentioned earlier, we are interested in determining the impact of the hydraulic fracturing parameters on the long term production performance of a well. To be more specific, the long term production performance in this thesis refers to the first 3 years of production. Even though shale wells usually continue producing for more than 3 years, a typical shale well produces the most in the first 6 months to 3 years time period. As shown in Figure 3.3, the production of a shale well dramatically decreases after that time period and the well continues producing at a very low rate. The average decline rate in Eagle Ford Shale is as high as 74%, and gradually decreases to 47 and 19 % during the second and third years, and after 3 years, only 11% of the initial production remains (Wachtmeister, et al., 2017). Therefore, most of the oil is produced during the first 3 years and maximizing the production during the first 3 years is of great interest. We select 3 years cumulative production as a metric for production performance, and as indicated earlier in the data understanding section, we have 3 years production data available for all the wells.

During that time period, the well also produces gas and we account for the gas produced by converting the gas into oil and add it to the oil production as the following:

Barrels of oil equivalent (BOE)= oil production (bbls) + Mcf/6

The conversion factor used to convert gas to oil is approximately 6 and it is based on the heating value.

In addition, the 3 years cumulative production metric is more reflective of the reservoir behavior than the initial peak oil production as shown by the spearman's correlation in Table 3.2. In the table, the shorthand BOE stands for 36 months cumulative Barrels of Oil Equivalent, resep stands for Reservoir Pressure, satu stands for Oil Saturation, vis stands for Oil Viscosity, Thick stands for Reservoir Thickness, Poros stands for Porosity, and Peak stands for Initial Peak Oil Production which refers to the highest oil production in the first six months of oil production. We notice that there is a correlation of 0.82 between the initial Peak oil production and the 3 years cumulative BOE which means that the initial Peak oil does not completely reflect the long term or 3 years production performance. We also notice that there is a higher correlation between the cumulative BOE and all the reservoir properties except oil viscosity than there is between the reservoir properties and the initial Peak oil production. This means that the 3 years cumulative BOE is more reflective of the reservoir behavior than the initial Peak oil production.

Thus, in this thesis, the 36 months cumulative barrels of equivalent oil for which we use the shorthand BOE is selected as a production performance metric because it is a better representative of long term production performance and a better representative of reservoir behavior than the initial Peak oil production.

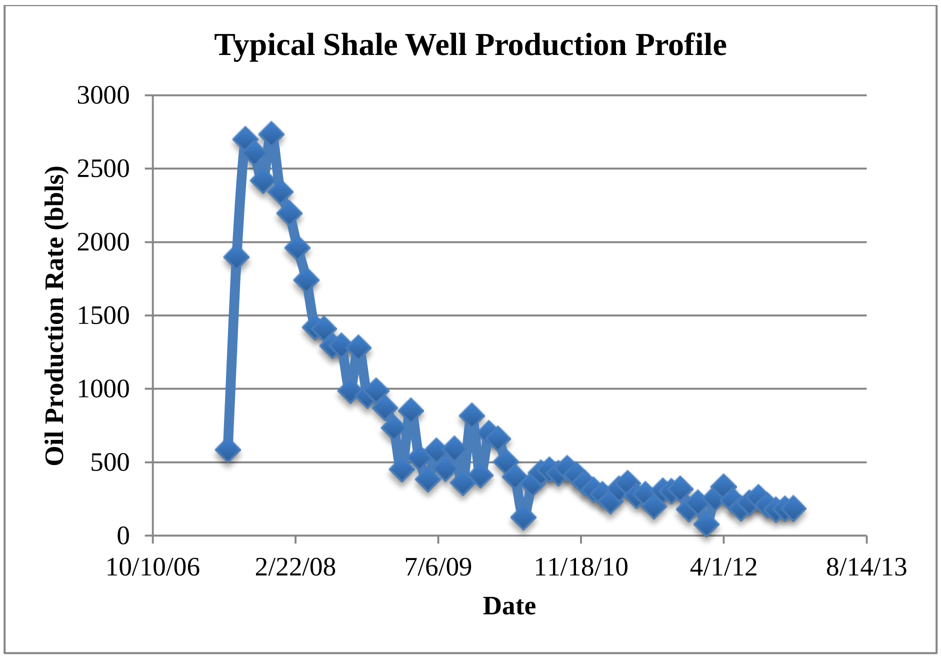


Figure 3.3: Typical Shale Well Production Profile

Table 3.2: Spearman's Correlation of the Reservoir Properties and Production Performance Metric

	BOE	resep	satu	vis	Thick	Poros	Peak
BOE	1.00	0.45	0.19	-0.38	-0.39	0.23	0.82
resep	0.45	1.00	0.66	-0.73	-0.78	0.61	0.39
satu	0.19	0.66	1.00	-0.46	-0.64	0.72	0.16
vis	-0.38	-0.73	-0.46	1.00	0.36	-0.51	-0.44
Thick	-0.39	-0.78	-0.64	0.36	1.00	-0.48	-0.19
Poros	0.23	0.61	0.72	-0.51	-0.48	1.00	0.22
Peak	0.82	0.39	0.16	-0.44	-0.19	0.22	1.00

3.2.2. Research Question 2: *How can the effects of geological and petrophysial variables and reservoir fluid properties be eliminated or minimized?*

Shale reservoirs are known to be highly heterogeneous and anisotropic meaning that the reservoir properties vary throughout the formation. Reservoir and geological properties have a great impact on the production performance of a well. The relationship between reservoir and geological properties and the production performance of a well are shown in the two equations below.

Original Oil In Place

$$= 7{,}758 A h \phi S_{oi}$$

where: A = surface area of the reservoir, acres.
h = net-pay thickness of the formation, ft.
ϕ = porosity, fraction.
S_{oi} = initial oil saturation, fraction.

Oil Production Rate

$$q = \frac{kh\Delta p}{\mu_o}$$

q_o = surface production rate, STB/D
k = formation permeability, mD.
h = formation thickness, ft.
μ_o = oil viscosity, cP.
Δp= reservoir pressure-wellbore pressure(psi)

The first equation is used to measure the oil capacity of a reservoir and the second equation is used to indicate the production capacity of the reservoir. If two wells have the same hydraulic fracturing design, the one that has better reservoir and geological properties will have a higher production performance than the other well. Since we are interested in determining the impact of the hydraulic fracturing parameters, we need to make sure we normalize the effects of the reservoir and geological differences between the wells so that any difference in production performance can be attributed to the differences in hydraulic fracturing design between the wells. To minimize the effects of the differences in geological and reservoir parameters, we divide the wells into groups of similar reservoir and geological properties. According to the above equations, the important reservoir and geological related parameters are the following:

- Area of the reservoir
- Thickness of the reservoir
- Porosity of the reservoir
- Oil saturation of the reservoir
- Permeability of the reservoir
- Pressure of the reservoir
- Oil viscosity of the reservoir

Thus, we divide the wells based on these parameters into 3 groups as shown in Table 3.3 below. A scatterplot matrix of the complete geological and reservoir related data of the wells of each group is shown in the Appendix in Figures B.1, B.2, and B.3.

Table 3.3: The Reservoir and Geological Related Parameters of the Three Groups

Reservoir and geological Properties	Group 3	Group 2	Group 1
Pressure (psi)	2759	4511	6277
Thickness (ft)	381	197	159
Porosity	0.085	0.100	0.094
Oil Saturation	0.56	0.65	0.64
Viscosity (cp)	0.68	0.56	0.42
Permeability	Nano darcy	Nano darcy	Nano darcy
Area of the reservoir	Unavailable data	Unavailable data	Unavailable data

Since the area of the wells are unknown, we can't account for this parameter, and since the permeability of the wells is not available, we can't use this parameter to group the wells either. However, based on the literature review presented in Chapter 2, the permeability of the Eagle Ford Shale is extremely low and is in the range of nano darcies throughout the formation. The viscosity data isn't available for the wells either. However, since we have oil API gravity of each well, we calculated the viscosity of the wells using the Beggs and Robinson correlation equation (Beggs, et al., 1978). The correlation requires the oil API gravity and reservoir temperature as inputs. We don't have the reservoir temperature of the wells, but since we have the depth of each well, we were able to calculate the reservoir temperature using a temperature gradient of 0.02 F/ft determined by Gong and co-authors (Gong, et al. 2013) and adding a surface temperature of 60 F. The calculated viscosity numbers presented in Table 3.3 were checked and confirmed with the viscosity heatmap of the Eagle Ford Shale developed by Cander (Cander, 2013).

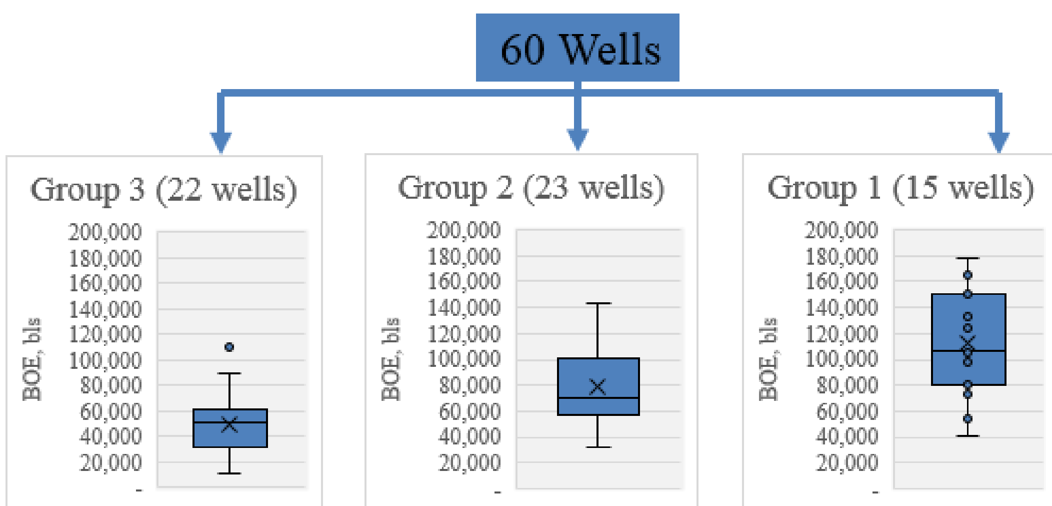


Figure 3.4: Production Performance of Each Group

Based on the reservoir properties listed in Table 3.3, Group 1 has a more favorable reservoir quality than Group 3 since it has a higher reservoir pressure, porosity, oil saturation, and lower viscosity than Group 3 even though Group 3 has a larger reservoir thickness. The production performance of Group 1 is on average higher than that of Group 3 as shown in Figure 3.4. This difference in production performance between the wells could be due to the differences in the reservoir and geological properties between the groups or it could be due to a combination of the reservoir and geological and hydraulic fracturing parameters.

The distribution of the production performance of the wells in each group are shown in the boxplots, and we can see that even though the wells in each group have fairly similar reservoir properties, their production performances are different which could be due to the differences in hydraulic fracturing design between the wells in each group.

The differences in production performance between the groups and between the wells within a group are explained in Chapter 5 after determining the importance of the hydraulic fracturing parameters.

The grouping of the wells based on the reservoir and geological properties matches the trends in Eagle Ford Shale reservoir and geological properties reviewed in the literature as presented by the blue arrows in Figure 3.5.

Thus far, the research question 2 has been answered and the differences in reservoir and geological properties between the wells have been minimized by dividing the wells into groups of similar reservoir and geological properties.

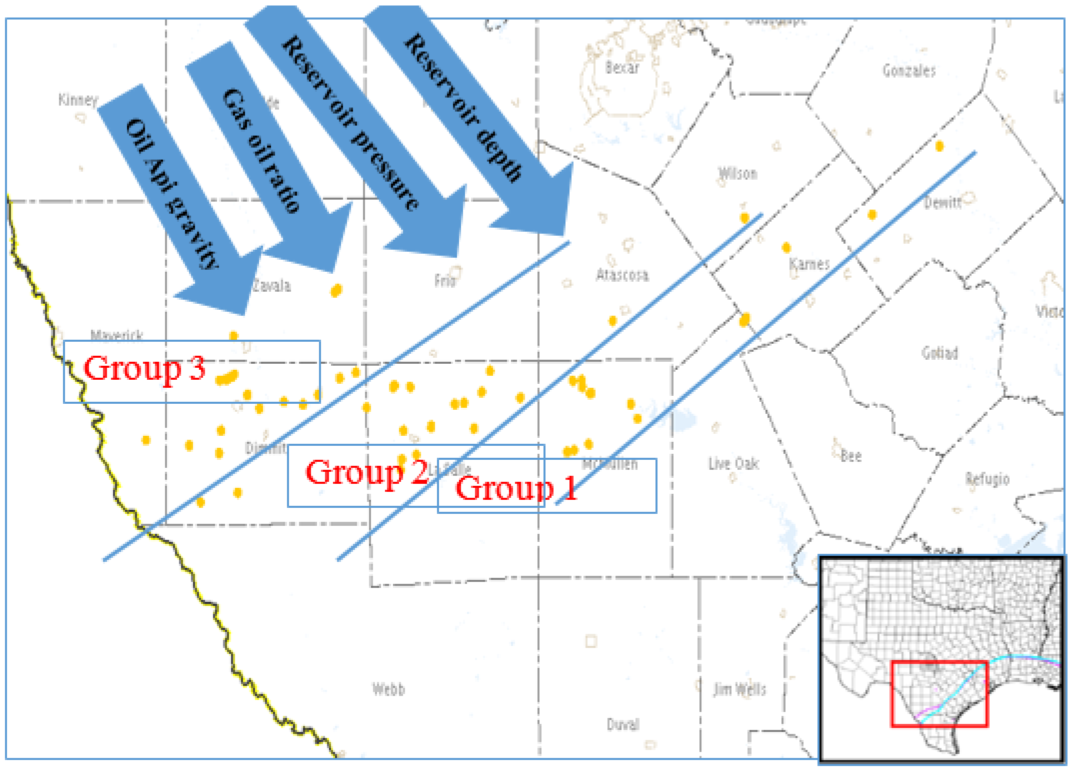


Figure 3.5: Location of the 3 Groups of Wells and Changing Reservoir Properties

3.2.3. Research Question 3: *What parameters of hydraulic fracturing should be considered for analysis?*

The hydraulic fracturing parameters selected for analysis are the following:

- Perforated Length Interval (pli)(ft)
- Injection Rate per Stage (irps) (bpm)
- Number of Clusters per Stage (nocps)
- Volume of Proppant per Well (voppw)(lbs)
- Volume of Water per Well (vowpw)(gals)
- Number of Stages per Well (nos)
- Average Treating Pressure per Well (atppw)(psi)
- Maximum Treating Pressure per Well (mtppw)(psi)

These parameters are selected to represent both the geometry and the fluid design of the hydraulic fracturing process, and these are all measured parameters. We also have a complete set of data points for each one of these parameters. However, one other parameter that would have been useful to include is the type of fluid mixture added to the hydraulic fracturing fluid so that we could analyze the impact of fluid type on hydraulic fracturing. Since we don't have any data available regarding the type of the mixture, we can't include this in our analysis.

An initial analysis of those hydraulic fracturing parameters using a scatterplot matrix correlation indicates that there is a strong correlation between average treating pressure and maximum treating pressure as shown in Figure 3.6. The symbols in the square boxes represent the hydraulic fracturing parameters and are defined above at the beginning of this section. We notice that there is a strong correlation between average treating pressure and maximum treating pressure, and we use the Spearman's Rank correlation coefficient to quantify the correlation between the variables. Spearman's rank correlation coefficient is used to consider the ordering of values and rank them. The mathematical form of Spearman's rank correlation coefficient is as follows:

$$\rho = 1 - 6\frac{\sum_{i=1}^{n}(r(x_i) - r(y_i))^2}{n(n^2 - 1)}$$

Where $r(x_i)$ is the rank of value (x_i) when the list $(x_{1,}\ldots . x_n)$ is sorted in increasing order. $R(y_i)$ is define analogously. $-1 \leq \rho \leq 1$ and larg values of $/\rho/$ mean the rankings of the x and y are in a similar or exact order (Engineering Statistics, 2013).

Based on the Spearman's Rank correlation coefficient, there is a 0.83 correlation between average treating pressure and maximum treating pressure as shown in Table 3.4. It is safe to say that higher average treating pressure most of the time means higher maximum treating pressure. We can use average treating pressure to represent the maximum treating pressure and reduce the dimension to 7 variables from 8 variables. However, we include both variables in the analysis as maximum treating pressure tells us about the breakdown pressure of the formation while the average treating pressure tells us about the treatment pressure of the fractures and both are useful information to have in the analysis.

It is also shown that there is a slightly strong correlation between number of clusters per stage (nocps) and the perforated length interval (pli) and the number of stages which means that wells with higher number of clusters and number of stages usually have longer lateral length. However, checking with the spearman correlation matrix, the correlation between those parameters is found to be below 0.80 which means that the correlation is not strong enough to represent those parameters with one and eliminate the others. We proceed with the 8 parameters for regression analysis.

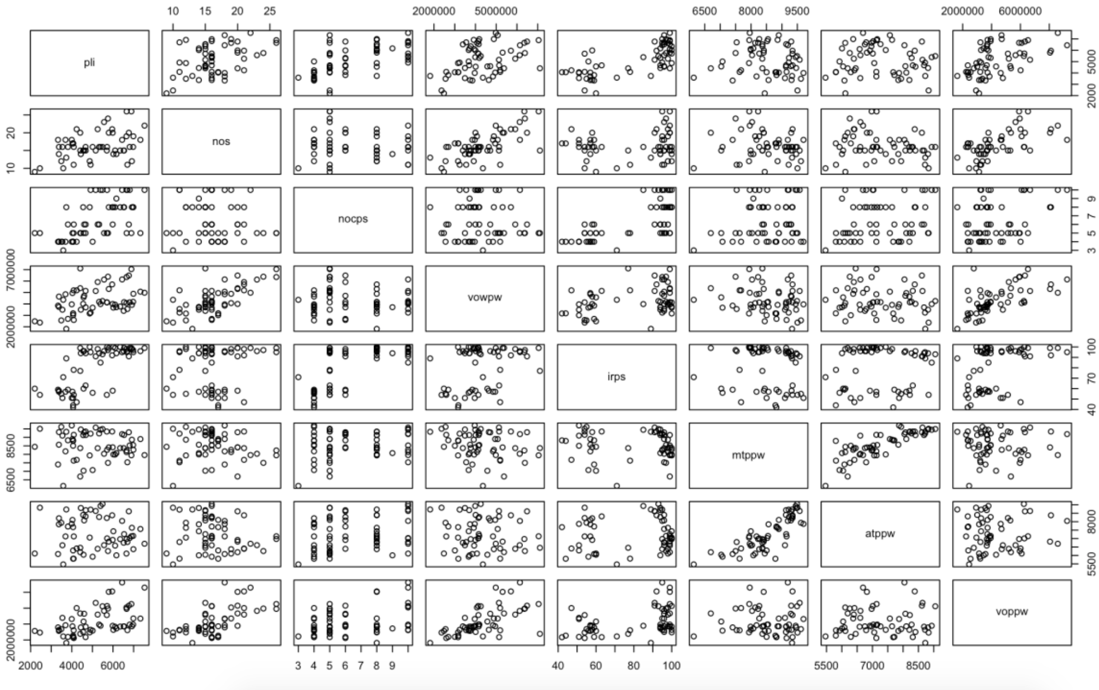


Figure 3.6: Scatterplot Matrix of the Hydraulic Fracturing Parameters

Table 3.4: Spearman's Correlation of the Hydraulic Fracturing Parameters

	BOE	pli	nos	nocps	vowpw	irps	mtppw	atppw	voppw
BOE	1.000	-0.039	0.140	-0.210	-0.051	-0.130	0.180	0.220	0.029
pli	-0.039	1.000	0.300	0.650	0.390	0.690	-0.170	-0.030	0.540
nos	0.140	0.300	1.000	-0.110	0.590	0.046	-0.170	-0.170	0.610
nocps	-0.210	0.650	-0.110	1.000	0.032	0.670	0.110	0.260	0.310
vowpw	-0.051	0.390	0.590	0.032	1.000	0.290	-0.200	-0.088	0.670
irps	-0.130	0.690	0.046	0.670	0.290	1.000	-0.310	-0.052	0.380
mtppw	0.180	-0.170	-0.170	0.110	-0.200	-0.310	1.000	0.830	-0.035
atppw	0.220	-0.030	-0.170	0.260	-0.088	-0.052	0.830	1.000	0.030
voppw	0.029	0.540	0.610	0.310	0.670	0.380	-0.035	0.030	1.000

Thus, up to this point, we have answered the first 3 research questions and we have prepared the following data for regression analysis:

- Production Performance Metric, BOE
- 3 Groups of Wells
    - Group 1: 15 wells

    - Group 2: 23 wells
    - Group 3: 22 wells
- 8 Hydraulic Fracturing Parameters
    - Perforated Length Interval (pli)(ft)
    - Injection Rate per Stage (irps) (bpm)
    - Number of Clusters per Stage (nocps)
    - Volume of Proppant per Well (voppw)(lbs)
    - Volume of Water per Well (vowpw)(gals)
    - Number of Stages per Well (nos)
    - Average Treating Pressure per Well (atppw)(psi)
    - Maximum Treating Pressure per Well (mtppw)(psi)

In the next section, we answer the fourth research question by identifying the regression techniques that are suitable for analysis of the above data that we have prepared, and then in the last section, we answer the fifth research question by presenting the methods of assessing the regression techniques and selecting the best performing regression technique for analysis of hydraulic fracturing data.

3.2.4. Research Question 4*: What kind of data mining and statistical approach can be used to capture both linear and nonlinear relationships between the input and output variables?*

The problem that has been addressed in this thesis falls into Supervised Statistical Learning which means the input and output variables are known, and the regression technique is trained to identify a relationship between the known input and output data. The goal is to fit a regression technique that relates response to predictors or inputs variables so we understand the relationship between the response and the predictors. There are multiple supervised learning techniques such as multiple linear regression, decision trees, elastic net, neural networks, and so on. Depending on the nature of the problem, each of these techniques have different performances, and choosing the right one depends on factors like the goal of the study (inference, prediction, or classification), whether there is linear or nonlinear relationship between input and output variables, whether the data is continuous or binary, and so on.

In this thesis, we have identified four regression techniques that are well suited for the purposes of this study. The techniques are Logistic Regression (LG), Decision Trees (DT), Support Vector Machines (SVMs), and Neural Networks (NNs). All of these techniques have the ability to capture nonlinear relationships which is an import factor in this study as it is likely that there is nonlinear relationship between the input and output variables. The second reason is that our data is continuous and these techniques can be used and output continuous values.

The first two techniques LG and DT are easy to understand and interpret and provide good fits to the data. SVMs and NNs are more complicated and not as easy to interpret and understand, but they have the advantage of being flexible in fitting the data and have better learning capabilities when compared with LG and DT (Moro, et al., 2014).

There are some other methods that provide these advantages of capturing non linearity and interactions between the variables. One of these models is Multivariate Adaptive Regression Splines (MARS). However, this technique is not considered in this thesis due to its unpopularity and the limited scope of this thesis. As mentioned by

Gao and co-authors, their paper is the first documented application of the MARS algorithm to analyze and interpret petroleum industry data (Gao, et al., 2013). The methods that we have considered in this thesis are much more widespread and popular in the analysis of petroleum industry data. The second reason is that there is clearly not enough time and space to consider all the regression techniques in a master's thesis. With the four regression techniques that we have selected, we have covered most of the advantages offered by other regression techniques such as ability to capture non linearity and interactions between the input and output variables, flexibility in fitting the data, and ease of interpretability.

In the next chapter, a brief description of each of these techniques is provided, and these techniques are used on the data. Then they are compared to select the best of them to proceed with.

In the next section, the methods of assessing the performance of these techniques and comparing them to each other are presented.

3.2.5. Research Question 5: *How do we assess and compare the performance of the techniques?*

We use R^2 and RMSE to compare the regression techniques and select the technique with the best fit. R^2 is used to measure the correlation between the observed and predicted values while RSME is used to measure the predictive ability of the fit. They are both simple methods to assess the quality of a regression fit.

R^2 statistics uses a proportion of variance explained to assess the measure of fit and is always between 0 and 1. R^2 is calculated as the following:

$$1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}$$

Where $\mathrm{TSS} = \sum (y_i - \bar{y})^2$ and $\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

TSS is the total sum of squares and measures the total variance in the response which means it is the amount of variability inherent in the response before regression is performed. However, RSS measures the amount of variability in the response that is not explained after performing the regression. Thus, TSS – RSS is the measure of the variability that is explained by the regression model. R^2 is the measure of the variability in Y that can be explained by using X. A technique with an R^2 value close to 1 indicates that a large proportion of the variability in the response has been explained while an R^2 of close to 0 indicates the opposite. However, since R^2 is a measure of correlation not accuracy and is highly affected by the variance in the output variable, it can have systematic bias. We use RMSE to confirm the model assessment.

RMSE which stands for root mean square error is used to characterize the technique's predictive capabilities and compare the four methods. RMSE uses the regression residuals which are calculated as the observed values minus the regression predictions. Then the residuals are squared and summed and then divided by the number of samples to calculate the mean squared error(MSE).

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2,$$

The square root of MSE is taken to get RMSE so that it is in the same units as the original data. RMSE is usually interpreted as the average distance between the observed values and the regression predictions.

3.3. SYNOPSIS OF CHAPTER 3

In this chapter, we gain an understanding of our data in regards to the problem that has been addressed and prepare the data for regression analysis which is done in the next chapter. In Section 3.1, data understanding is presented. The quality of the data in regards to how relevant, recent, and complete is the data is examined and improved. Some discrepancy, missing values, and outliers are identified and resolved.

In Section 3.2, data preparation is presented. The five research questions are answered using the statistical tools and analysis. Based on an understanding of shale well behavior, a production performance metric is determined. The wells are grouped based on the reservoir and geological related parameters to minimize the differences between the wells regarding the reservoir properties. The hydraulic fracturing parameters are selected for analysis, and the regression techniques are identified. Last but not least, the methods of assessing and comparing the regression techniques are identified.

Moving forward, in the next chapter, a brief description of the four regression techniques is first provided. Then the techniques are used on the data, and they are assessed and compared to determine the best performing regression technique. Then the best performing technique is used to determine the key stimulation parameters which is the answer to the primary research question and is presented in the next chapter.

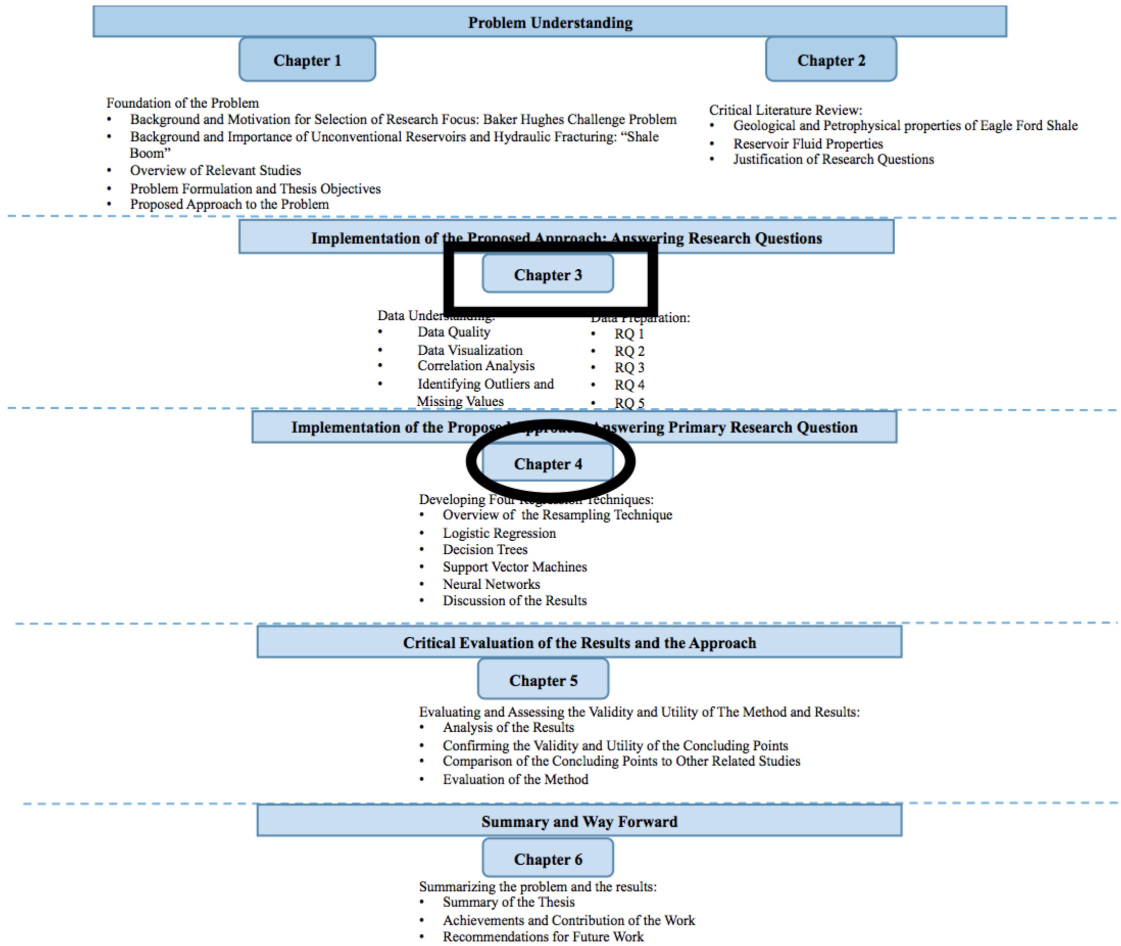


Figure 3.7: Organization of the Thesis -Present (Boxed) and Next (Circled)

CHAPTER 4

OVERVIEW OF THE REGRESION TECHNIQUES AND RESULTS

In Chapter 3, data understanding in regards to the data quality, missing values, and outliers is provided. The quality of the data is improved and any issues relating to missing values and outliers are resolved. Then the data is used to answer the five research questions. A production metric to measure well performance is established, the geological and reservoir variations among the wells are minimized by grouping wells of similar properties together. Eight hydraulic fracturing parameters are selected to analyze, and four regression techniques are identified to use for analysis of the data. Last but not least, two statistical techniques are determined to measure the performance of the regression techniques and compare them. Now that we have prepared the data and identified the tools and techniques of data analysis, in this chapter, we use the data and these techniques to answer the primary research question. We first present a resampling method called 10 Fold Cross Validation used by regression techniques to learn about the data and make a better fit. Then the four modeling techniques which are Logistic Regression (LR), Decision Trees (DT), Support Vector Machines (SVMs), and Neural Networks (NNs) are explained. We then run the four techniques on the group 2 data since this group has the largest number of data points. Once we have run the techniques on the group 2 data, we then evaluate and compare the results of the techniques to select the best performing method. We run the best performing method on the rest of the other groups. We analyze the results for each group and draw conclusions to answer the primary research question. In the next chapter, the concluding points and the method are evaluated and compared to other studies that have been done in this area.

4.1 OVERVIEW OF THE RESAMPLING TECHNIQUE

One of the most common resampling methods is 10-fold cross validation which we use in this thesis due to its superiority over the other methods. Resampling methods involve repeatedly drawing samples from the data and refitting the regression technique on each sample to gain additional information about the data. This method allows us to calculate the test error associated with each technique and select the method with the smallest test error.

10-fold cross validation (CV) method randomly divides the data into 10 groups or folds of approximately equal size. The first set of data points are used as a validation set. The regression technique is then fit on the remaining 10-1 folds, and mean RMSE is calculated between the predicted and actual values in the held out group or fold. This procedure is repeated 10 times, and each time a different group of data points are used as a validation set. This procedure results in 10 estimates of the test error RMSE.

Using 10-fold cross validation is very common as one of the advantages of using 10 folds instead of more than 10 is the computational time which would be much longer if we had used a larger number of folds. Another advantage of 10-fold cross validation is that it has been empirically shown that using k=10 gives more accurate estimates of the test error that suffer neither from high bias nor high variance (Kuhn, et al., 2016).

The reason we use cross validation in this thesis is that we are interested in the minimum point in the estimated test RMSE curve and use this minimum RMSE with R^2 to identify a method that results in the lowest test error.

4.2 LOGISTIC REGRESSION

Logistic linear regression is one of the simplest and most straight forward technique that can be written in the following form:

$$y_i \approx f(X_{.i}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi}$$

Where $\beta_1, \ldots, \beta_p$ are the regression coefficients for the ith predictors, and β_0 is the estimated intercept. X_{ij} represents the value of the jth predictor for the ith sample which in our case corresponds to 8 predictors for 60 data points, and Y_i is the numeric response for the ith sample. Since we don't know the value of the coefficients β, we estimate the β value based on the observed data. The technique uses the ordinary least squares method which minimizes the sum of the squared differences between the fitted and the observed values to estimate the value of β and is written in the following form:

$$L(\hat{\beta}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = ||Y - \mathbf{X}\hat{\beta}||_2^2$$

$L(\hat{\beta})$ is called the method of ordinary least squares (OLS).

The main advantage of the above mentioned method and equation is that it is simple and easy to interpret. The estimated coefficients of the predictors can be used to interpret the relationship between the predictors and the response variable. Another advantage is that this model can easily be extended to a polynomial regression that is capable of capturing nonlinearity between the input and output variables and account for the interaction within the input variables. Each of the original predictors or parameters can be easily raised to a power such as x^2,x^3…x^n (Kuhn et al., 2016).

In this thesis, to fit a linear regression that has the capability of capturing the nonlinear relationships and accounting for the interaction between the variables, a specific formula is created as shown in the Appendix D. Each predictor has been extended to include all the linear parameters, square of each parameter, and the two factor interactions between the parameters.

We run the technique on group 2 with 10 cross validation and we got an RSME of 288544.4 and R^2 value of 0.85. We will use these two numbers to compare this technique to the rest of the other techniques that are discussed next.

4.3 DECISION TREES

Decision tree uses a set of rules to divide a large heterogeneous data into smaller and more homogenous groups and creates a hierarchical structure with respect a target variable. Building a tree from root to leaves involves the selection of the splits first which is based on an evaluation measure for the predictors or input parameters. A predictor with the best evaluation measure is chosen which is the predictor that results in the smallest tree or that produces the purest nodes. Then the data is divided according to the splits. This procedure is then repeated and applied to the subsets as well. A decision on whether a node should be split or is a terminal is made at each node, and finally each terminal node is assigned a class (James, et al., 2013). Shown in Figure 4.1 is an example of a tree that we built for the hydraulic fracturing parameters. The Code is shown in the Appendix D. Since the number of data points are small and have values close to each other in group 2, we have only two branches of the tree. For this method, the smallest RSME value is 30283 and the value of R^2 is 0.8379

As we can see the previous regression technique has a smaller RSME value and a higher R^2 value and thus it is considered a better technique than the decision tree method. Therefore, we choose the regression method over the decision tree method.

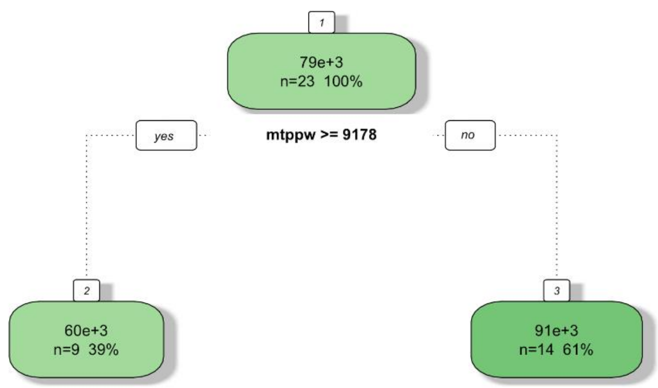


Figure 4.1: Decision Tree Model of Group 2 Data

4.4 SUPPORT VECTOR MACHINES

Support vector machines are a class of very powerful and flexible techniques that are not limited to linear models, are robust to outliers, and have different types of penalized regression. The ϵ-insensitive regression minimizes the following:

$$c \sum_{i=1}^{n} L_{\epsilon} (y_i - \hat{y}_i) + \sum_{j=1}^{p} \hat{\beta}_j^2$$

Where c is a cost parameter that is set by the user, and $L\epsilon(.)$ is a loss function. The cost parameter is associated with the residuals not the regression coefficients like in the linear regression. Thus with the following loss function:

$$L_\epsilon (y_i - \hat{y}_i) = \begin{cases} 0 & \text{if } |y_i - \hat{y}_i| \leq \epsilon \\ |y_i - \hat{y}_i| - \epsilon & \text{otherwise} \end{cases}$$

Points with small residuals do not contribute to the regression fit, and larger residuals contribute a linear amount. SVM is similar to linear regression in the sense that the parameter estimates can be written as a function of a set of unknown parameters α_i and the training sample as shown below:

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{p} x_{ji} u_j \right)$$

There are as many unknown parameters α_i as there are data points in the above equation, and the training data represented by x is required to make a prediction. Training data points with $\alpha_i \neq 0$ are called support vectors, and only points with large residuals are used for prediction. The above equation is usually rewritten in the follow form:

$$f(\mathbf{u}) = \hat{\beta}_0 + \sum_{i=1}^{n} \alpha_i K(X_{.i}, \mathbf{u})$$

The K(.) is called a kernel function, and a linear kernel function is written as:

$$K(X_{.i}, \mathbf{u}) = \sum_{j=1}^{p} x_{ji} u_j = X'_{.i}\mathbf{u}$$

There are other nonlinear kernel functions such as polynomial, radial basis function, and hyperbolic tangent. The radial basis function has been shown to be effective and is used in this thesis as shown in the written code in the Appendix D (James, et al., 2013). After we run the SVMs technique, the kernel parameter was analytically estimated to be 0.05245, and the smallest RSME value and R^2 value are found to be 26146.97 and 0.91. We can see there is a big improvement on both the RSME and R^2 values compared to the previous two techniques. Thus, we choose the SVMs over the previous methods.

4.5 NEURAL NETWORKS

Neural networks are nonlinear regression techniques that use an intermediary set of unobserved variables called hidden variables or hidden units to model the outcome. These hidden units are linear combinations of some or all the predictor variables, and this linear combination is transformed by a nonlinear function as the following:

$$h_k(\mathbf{x}) = g\left(\beta_{0k} + \sum_{i=1}^{P} x_j \beta_{jk}\right)$$

Where

$$g(u) = \frac{1}{1+e^{-u}}.$$

g(.) is the nonlinear function, and the coefficient β_{ik} is the effect of the ith predictor on the kth hidden unit. The hidden units can be specified by the user, and a NN usually uses multiple hidden units to model the outcome. Once the number of the hidden units is defined, a linear combination connects the hidden units to the outcome as the following:

$$f(\mathbf{x}) = \gamma_0 + \sum_{k=1}^{H} \gamma_k h_k$$

Where h is the number of hidden units.

The parameters are usually initialized to random values and then optimized to minimize the sum of the squared residuals. We created a sequence of numbers starting from 1 to 27 by a step of 2 to change the number of hidden units as the method is running and find the optimal number of layers in the method.

Since there is a large number of regression coefficients, neural networks tend to over-fit the relationship between the predictors and the response. To mitigate over-fitting, we use weight decay which is a penalization method to regularize the regression technique. We have assigned a value of 0.1, 0.01, and 0.001 to the decay weight and these are commonly used values as shown in the code in Appendix D. The objective function thus has been changed to error+λf(Θ) where the function f(Θ) grows larger as the components of Θ become larger and λ is the weight decay that represents how much we want to protect against overfitting. Having a value of 0 for the decay weight means we don't want any protection against overfitting while having a large value means we want the technique to make Θ as small as possible (Kuhn, et al., 2016).

To make sure the regression coefficients are on the same scale. We centered and scaled the parameters prior. Shown in Figure 4.2 is the regression that we have fit to the data.

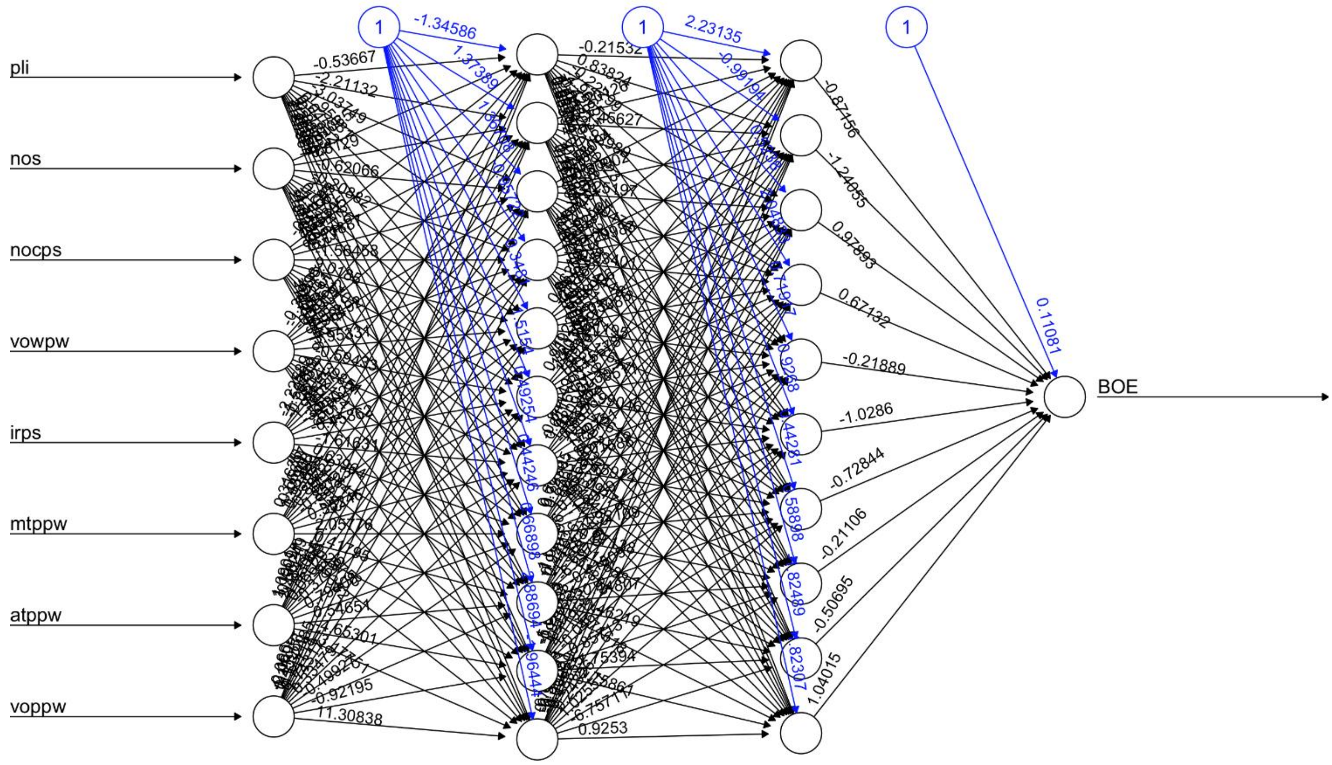


Figure 4.2: Visualization of the Neural Networks Technique for Group 2 Data

The variables on the left side are the input parameters going into the method, and the circles represent the hidden layers. The black numbers represent the weighted vectors between the neurons and the blue lines represent the bias added. On the right hand side is the BOE which is the output parameter. We are interested in the assigned weights to each parameter, but unfortunately there is not much we can tell based on this visualization. However, with the use of Garson method and function, we can easily determine the relative importance of each variable as shown in the next section.

We found the smallest RMSE to be 28032 and R^2 to be 0.92. Even though the SVMs method has a smaller value of RSME, the neural networks have a smaller R^2 value. Choosing any of the last two methods would not make any significant differences in our analysis since there is not a significant difference between the two methods. However, since we are not performing any predictions on the data due to the fact that there are not enough data points to divide into training and testing data, we are more interested in the correlation and the effect of the parameters than the predictive ability of the method. Thus, we choose the method with a larger R^2 value which is the neural networks.

4.6 DISCUSSION OF THE RESULTS

Now that we have identified and selected the best performing method, we run this method on the data for all the groups and present the results for each group in this section. We run the Neural Networks for the 3 groups, and we got the relative importance of each parameter in each of the three groups as show in Figures 4.3,4.4, and 4.5. The hydraulic fracturing parameters are listed on the x axis and the relative importance number is on the y axis.

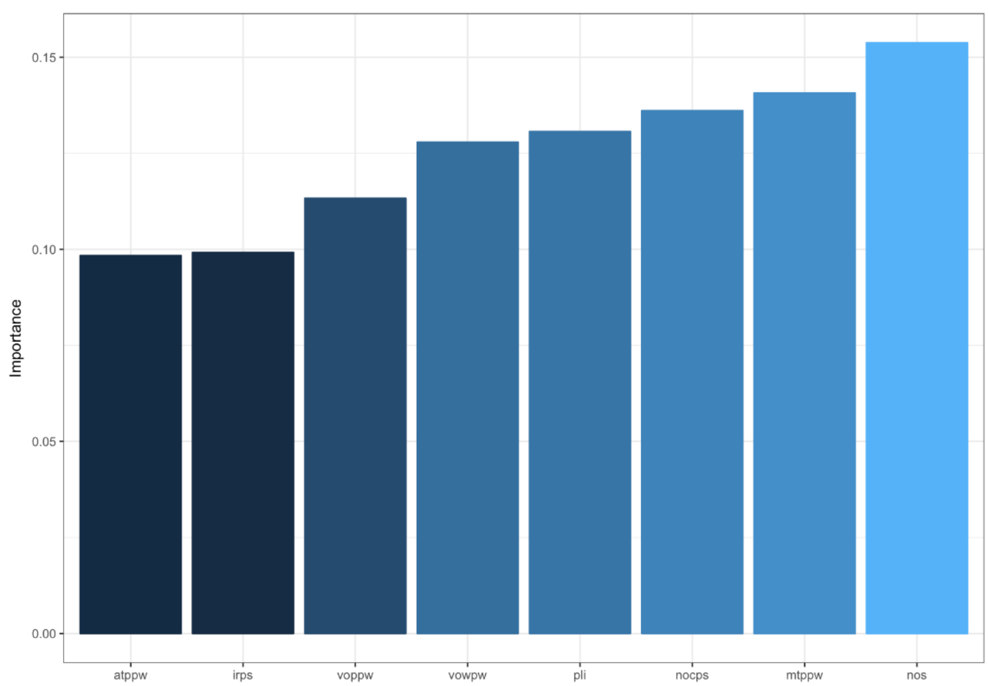


Figure 4.3: Relative Importance of Each Parameters on the Output for Group 1

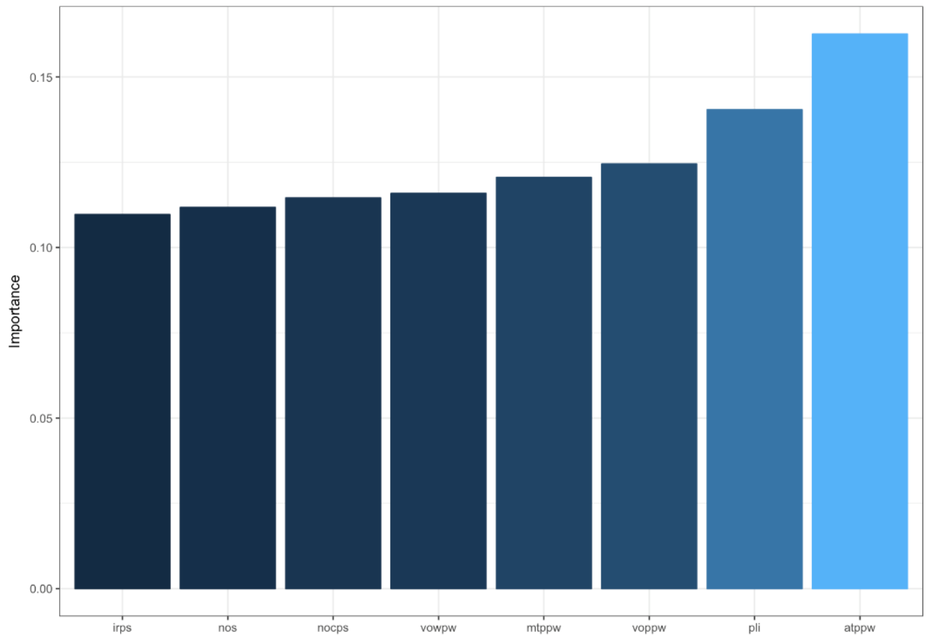


Figure 4.4: Relative Importance of Each Parameters on the Output for Group 2

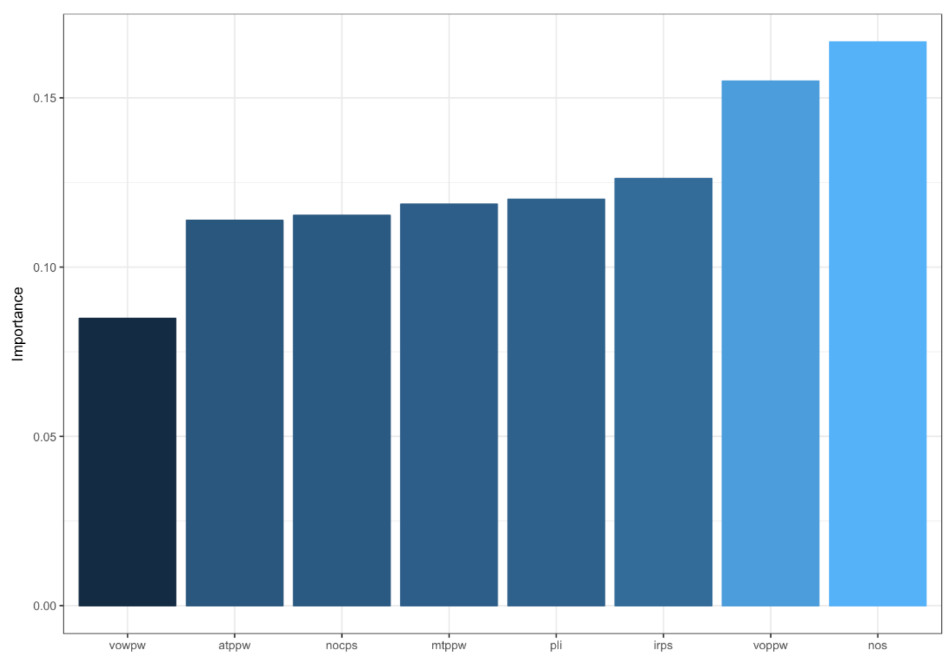


Figure 4.5: Relative Importance of Each Parameters on the Output for Group 3

The first thing we notice from these figures is that all the parameters have relatively high importance in each group meaning that these 8 hydraulic fracturing parameters have a large impact on the output parameter which is production performance. Thus, these results indicate that the key stimulation parameters are the following:

- Perforated Length Interval (pli)(ft)
- Injection Rate per Stage (irps) (bpm)
- Number of Clusters per Stage (nocps)
- Volume of Proppant (voppw)(lbs)
- Volume of Water (vowpw)(gals)
- Number of Stages (nos)
- Average Treating Pressure (atppw)(psi)
- Maximum Treating Pressure (mtppw)(psi)

We also notice that there is a slight difference between the parameters in terms of relative importance, and that changes for each group. For example, in Group 1, the 3 parameters with the highest relative importance are Number of stages, Maximum treating pressure, and number of clusters. However, the in Group 2, the 3 parameters with the highest relative importance are Average treating pressure, Perforated lateral length, and Volume of proppant. The top 3 parameters for in Group 3 are Number of stages, Volume of proppant, and Injection rate per stages.

Thus, the results of the regression method indicate that (1) each of the 8 hydraulic fracturing parameters that we have investigated have a large impact on the production performance, and (2) the top 3 parameters have the highest relative importance and are different from one group to another.

In the next chapter, we evaluate these results and the method. We check to see if these statistical results can explain the differences in production performance between the 3 groups and the wells within each group.

4.7 SYNOPSIS OF CHAPTER 4

In this chapter, the primary research question is answered. An overview of the resampling method and regression techniques is first presented. In Section 4.1, the 10 cross validation method that is used in the regression techniques is discussed. In Section 4.2, an overview of the logistic regression and the results of the technique are presented. In Section 4.3, a brief description of the decision tree method is provided and then the results of the method are presented. In Section 4.4, an overview of the Support Vector Machines and the results of the method are presented. In Section 4.5, the Neural Networks are presented and the results of the technique are discussed. Once all the techniques are run on the data, we selected the NNs as the best performing technique and run it on the rest of the data. The results of the NNs method are presented and analyzed in Section 4.6. The results of the method indicate two main points: first, all eight hydraulic fracturing parameters have relatively high importance in regards to the production performance of the wells, and second, different weights or relative importance are assigned to the hydraulic fracturing parameters meaning that some of the hydraulic fracturing parameters have relatively higher importance than the rest.

In the next chapter, we first analyze the first point by looking at the hydraulic fracturing parameters and the production performance between the 3 groups of wells and see if the variations in group performance can be explained by the differences in the hydraulic fracturing parameters. We then analyze this point further by comparing wells of similar geological and reservoir properties and determine if the differences in their production performance can be explained by the differences in the hydraulic fracturing parameters. Then the second point is evaluated by assigning the weights to each parameters and see if increase in production performance from one well to another is due to the change in the hydraulic fracturing parameters between the two wells. Based on the evaluation, we then draw concluding points about the results and the method. Then the concluding points are compared with the relevant studies from the literature.

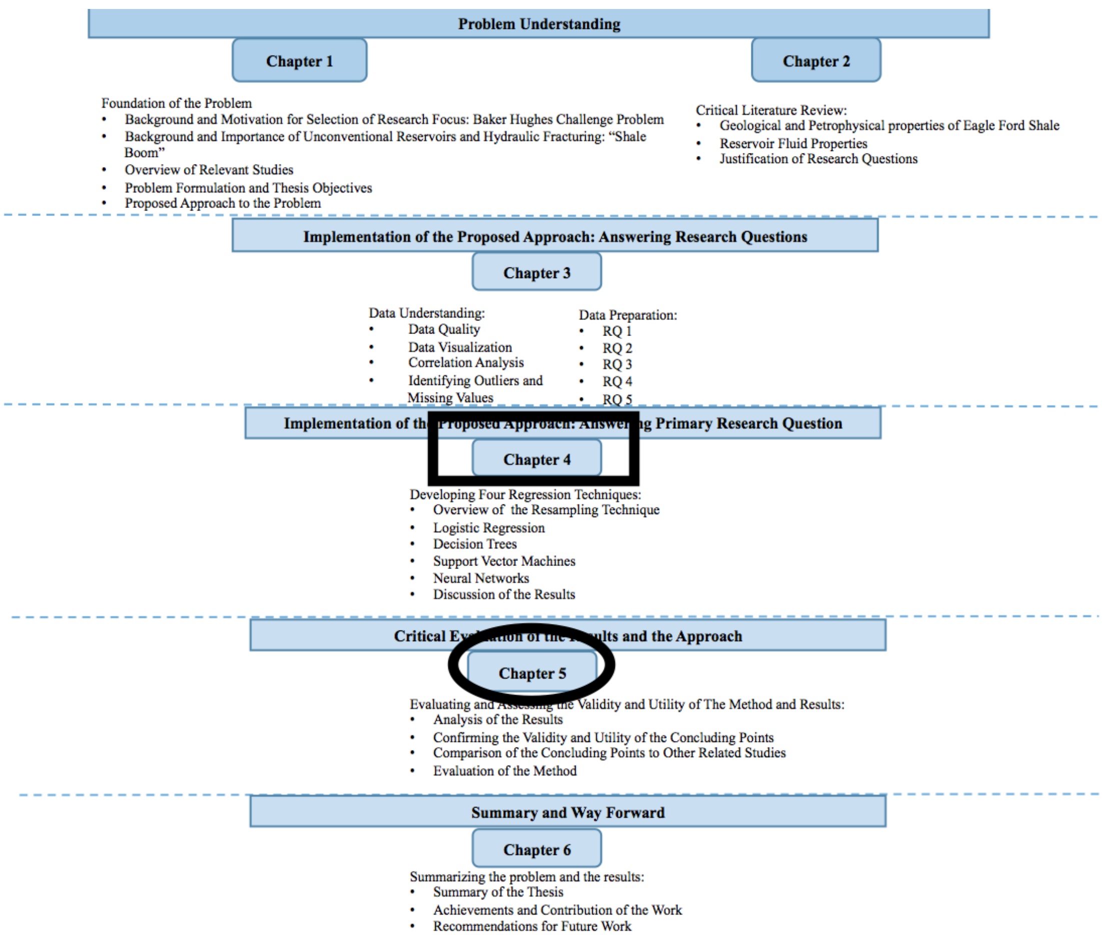


Figure 4.6: Organization of the Thesis -Present (Boxed) and Next (Circled)

CHAPTER 5

CRITICAL EVALUATION OF THE RESULTS AND THE APPROACH

In the previous chapter, the primary research question is answered by running the regression techniques on the data. The results of the four techniques are shown and it is found that the NNs and SVMs are the best performing techniques. NNs is selected to process with for the rest of the analysis and it is run on the rest of the data. Based on the results of the NNs method, the eight hydraulic fracturing parameters all have relatively high importance on the production performance of the wells, and some of the parameters have relatively higher importance than the rest. In this chapter, we evaluate the results of the regression technique and the approach. We first evaluate the first point by comparing the hydraulic fracturing design and the production performance of the 3 groups while keeping in mind that these groups have different reservoir and geological properties. We then evaluate this point by comparing wells of similar geological and reservoir properties. The differences in production performance and the hydraulic fracturing parameters are examined between a set of wells in each group. Once we have confirmed the importance of the hydraulic fracturing parameters, we evaluate the second point or result of the NNs method by assigning weights to each hydraulic fracturing parameters and calculating the increase in production performance of two wells. We then draw concluding points from the evaluation and compare the concluding points to the results of relevant studies reviewed in the literature. This chapter is ended with the concluding points and the summary of the thesis is provided in the next chapter in which the achievements and recommendations for future work are presented as well.

5.1 ANALYSIS OF THE RESULTS

We divide the parameters into two groups: the first group we call geometry related parameters which include the parameters related to the geometry of the completion system as presented in Section 1.3.1 and are the following: perforated interval length (pli)(ft), number of stages (nos), and number of clusters per stage (nocps). The second group is called fluid related parameters which includes: volume of proppant per well (voppw)(lbs), volume of water per well (vowpw)(gallons), injection rate per stage (irps)(bpm), average treating pressure per well (atppw)(psi), and maximum treating pressure per well (mtppw)(psi) as presented in Section 1.3.1.

To understand how each one of these parameters plays an important role in the hydraulic fracturing of shale reservoirs, we compare the 3 groups in terms of how their production performances and the hydraulic fracturing parameters are different.

As mentioned in Chapter 3, the first group has a higher production performance than Group 2 and Group 3, and Group 2 has a higher production performance than Group 3. This difference in production performance could be due to the differences in reservoir and geological properties shown in Table 5.1. On average, Group 1 has a better reservoir quality than Group 2 and Group 3, and that could be the reason why it has a better production performance.

However, even though this provides strong evidence or explanation as to why Group 1 is performing better than Group 3, it doesn’t completely explain the big difference in production performance observed between the two groups. If we assume the hydraulic fracturing design between the two groups are the same, and the pressure drawdown is the same, the ratio of the production performance of Group 1 to Group 3 is calculated to be 2.027. However, the ratio of the actual production performance of Group 1 to Group 3 is 2.27. There is 23 percent difference in production performance between the two groups that is left unexplained. Therefore, in addition to the reservoir and geological differences, there must be another factor that is affecting the production performance and contributes to such a big difference in the production performance of the two groups.

Table 5.1: Reservoir Related Parameters of the Three Groups

Reservoir and geological Properties	Group 3	Group 2	Group 1
Pressure (psi)	2759	4511	6277
Thickness (ft)	381	197	159
Porosity	0.085	0.100	0.094
Oil Saturation	0.56	0.65	0.64
Viscosity (cp)	0.68	0.56	0.42
Permeability	Nano darcy	Nano darcy	Nano darcy
Area of the reservoir	Unavailabl e data	Unavailable data	Unavailable data

Since it is shown in results of the NNs regression analysis that is done in Section 4.5 that all the hydraulic fracturing parameters have an impact on the production performance of the groups, we investigate the effect of those parameters further and provide a complete explanation as to why there is such a big difference in production performance between Group 1 and Group 3.

In each group we calculate the average value of each parameter so that we can represent the entire group with one number for each parameter and compare it with the other two groups. The equation to calculate the average value of each parameter in each group is shown in appendix equations C.1 and C.2, and the equation to calculate percent increase is shown in Equation C.3. This way it is easier and more efficient to observe trends across the groups and make reasonable comparisons. Shown in Table 5.2 is the average value of each of the parameters for the three groups. The last row represents the numbers used to scale the parameters so that they can all be represented as a two digit number as shown in Figure 5.1.

Table 5.2: The Average Value of Each Parameter in Each Group

Groups	pli(ft)	vowpw (gallons)	nos	nocps	voppw (lbs)	irps (bpm)	atppw (psi)	mtppw (psi)	BOE
Group 1	5209	4327146	15.47	6.467	4091630	79.4	8151	9074	112.629
Group 2	4444	3812685	15.91	5.826	3980458	71.48	7395	8737	79.04
Group 3	5792	4834696	17.55	7.09	4795956	90.55	6578	7996	49.71
Scaling Parameters	/100	100,000			/100,000		/100	/100	/1000

We notice that Group 3 has the largest average value for all the parameters except average treating pressure and maximum treating pressure as shown in Figure 5.1. The values of the largest parameters such as volume of water and proppant in Figure 5.1 have been scaled so that they can all be represented in the same plot. The numbers on the x axis represent group number, and the color boxes and the numbers inside the color boxes represent the value of the hydraulic fracturing parameters mentioned in Table 5.1

As shown in Figure 5.1, the overall fracture design is different between the groups, and the largest difference is between Group 1 and Group 3. Since these wells or groups of wells are hydraulically fractured on a stage basis as mentioned in the earlier chapters, it is important to know how these groups compare on a stage basis. We divide all the parameters by the number of stages except number of clusters, injection rate, and treating pressure as these are already on a stage basis. The normalization also creates a new parameter called stage length which is the division of perforated lateral length by the number of stages. We scale the large variables according to Table 5.2 and plot the values as shown in Figure 5.2.

We notice there is a smaller difference in the hydraulic fracturing design between Group 1 and Group 3 on a stage basis. For example, on a well basis, it is shown that the wells in Group 3 have a larger volume of water than the wells in Group 1. However, on a stage basis, it is shown that the volume of water pumped down for each stage of the wells in Group 1 is more than the volume of the water pumped down for each stages of the wells in Group 3. The ratio of water volume of Group 1 to Group 3 is 0.896 on a well basis while the ratio is 1.018 on a stage basis. We observe a similar trend in the other parameters. For proppant volume, the ratio of Group 1 to Group 3 on a well basis is 0.852 while on a stage basis the ratio is 0.967. Thus, even though comparing the hydraulic fracturing data on a well basis provides us a good idea about the overall hydraulic fracturing design and performance, it doesn’t necessarily reflect how each stages of the wells are being treated. The hydraulic fracturing data on both a well basis and stage basis are investigated to know the overall hydraulic fracturing design and performance of the wells and the design and performance of each stages, respectively.

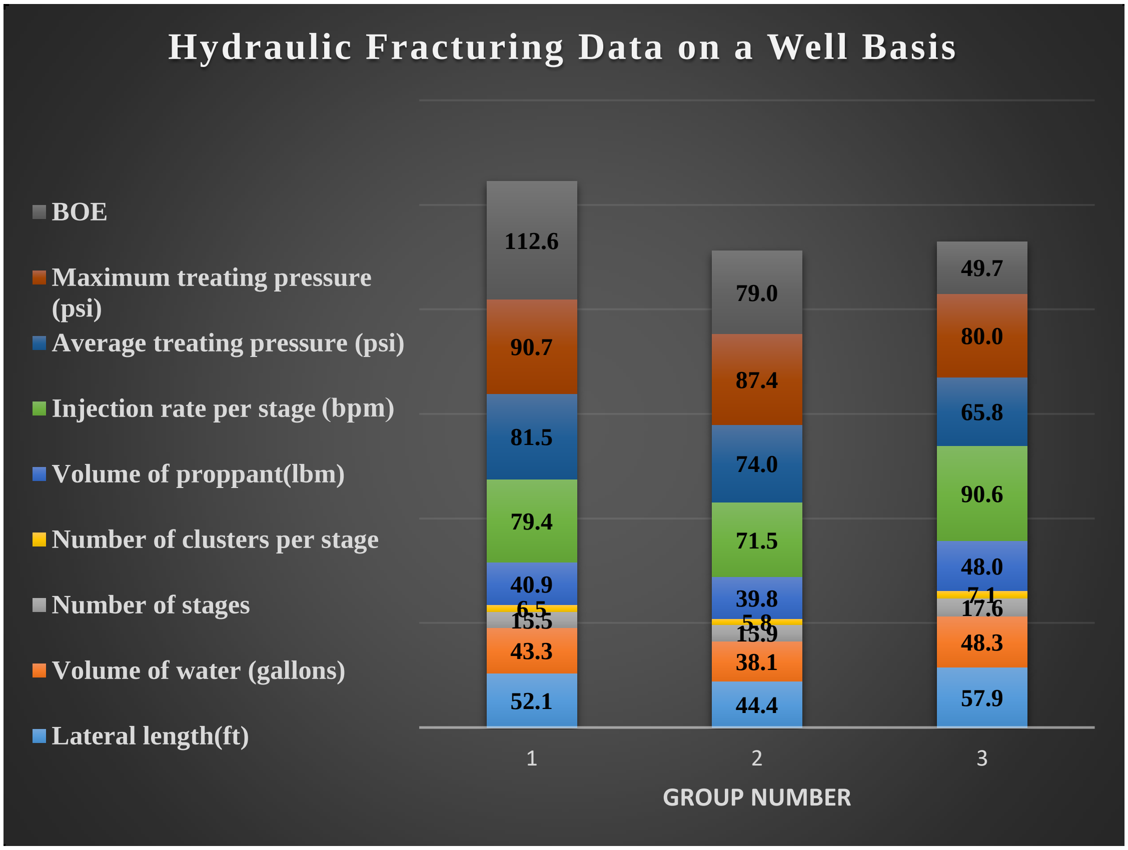


Figure 5.1: Average Value of Hydraulic Fracturing Parameters and BOE for Each Group

Table 5.3: Scaling of the Hydraulic Fracturing Parameters

Groups	pli(ft)	vowpw (gallons)	nos	nocps	voppw (lbs)	irps (bpm)	atppw (psi)	mtppw (psi)	BOE
Scaling Parameters	/100	10,000			/10,000		/10	/10	/100

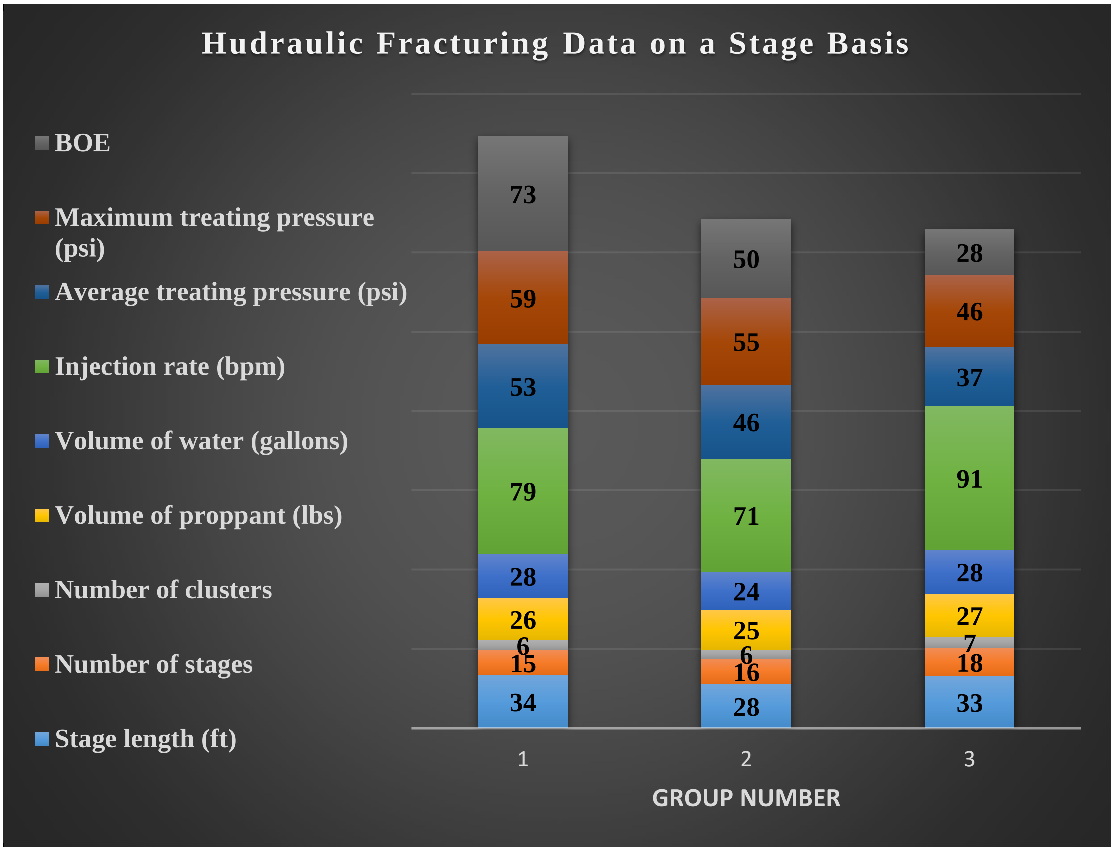


Figure 5.2: Average Value of Hydraulic Fracturing Parameters on a Stage Basis

As shown in Figure 5.2, the stage length in Group 1 is as short as the stage length in Group 3, the volume of water pumped per stage, the proppant volume per stage, and the number of clusters per stage are fairly similar for Groups 1 and 3. The biggest difference between Groups 1 and 3 is in the injection rate and treating pressure. Even though Group 3 has a higher injection rate than Group 1, Group 1 is treated with higher pressure on a stage basis. The percent increase in injection rate from Group 1 to Group 3 is only 14.1 percent while the percent increase in average treating pressure and maximum treating pressure from Group 3 to Group 1 is 28.7 and 40.5 percent, respectively. Thus, even though Group 3 has a higher injection rate than Group 1, its treating pressure is much less than Group 1. In hydraulic fracturing, high treatment pressure that is above rock break down pressure is required to initiate and propagate fractures (Ling, et al., 2016). It is possible that a high treatment pressure is applied in Group 1 because Group 1 has a higher reservoir pressure and thus a higher treatment pressure is required to fracture the formation. It is also possible that the high treatment pressure in Group 1 has resulted in creating better fractures and fracture propagation than the fractures in Group 3 even though Group 1 has slightly smaller injection rate. This coupled with the effect of the geological and reservoir properties could explain why Group 1 has such a higher production performance on a stage and well basis. However, due to data limitation, we have no way of confirming whether the high treatment pressure led to better fractures or not in Group 1, but in the next section we show the importance of having high treatment pressure by comparing wells of similar reservoir properties.

Table 5.4: Reservoir Related Parameters of a Set of Wells in Group 1

Well Number	1	2	3	4	5
BOE	165133	124307	106965	97086	53616
Reservoir Pressure (psi)	4262	4412	5781	7888	7208
Oil Saturation	0.71	0.67	0.65	0.67	0.70
Oil Viscosity (cp)	0.48	0.47	0.50	0.27	0.36
Reservoir Thickness (ft)	147	164	150	165	181
Porosity	0.10	0.10	0.09	0.10	0.10

5.2 CONFIRMING THE VALIDITY AND UTILITY OF THE CONCLUDING POINTS

In this section, we compare and analyze a set of wells from each group to confirm the importance of the hydraulic fracturing parameters that we have identified as the key stimulation parameters.

Group 1:

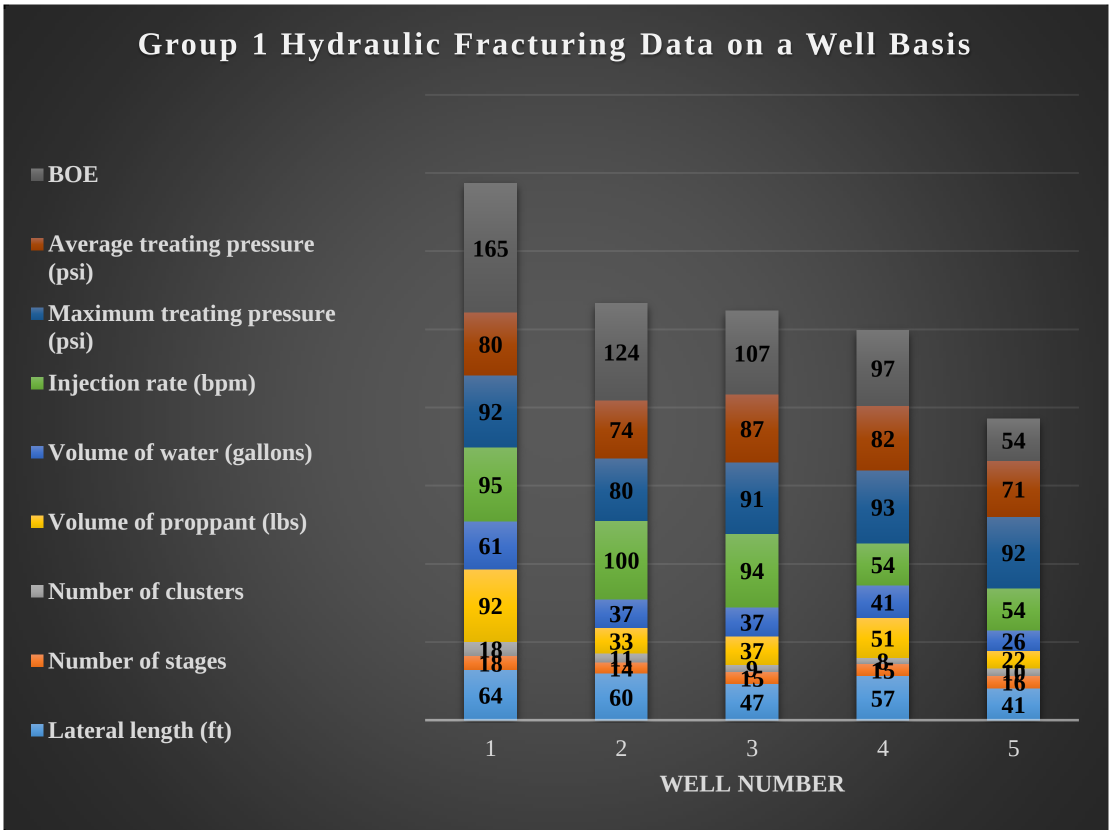


Figure 5.3: The Hydraulic Fracturing Parameters and Production Performance of a Set of Wells in Group 1

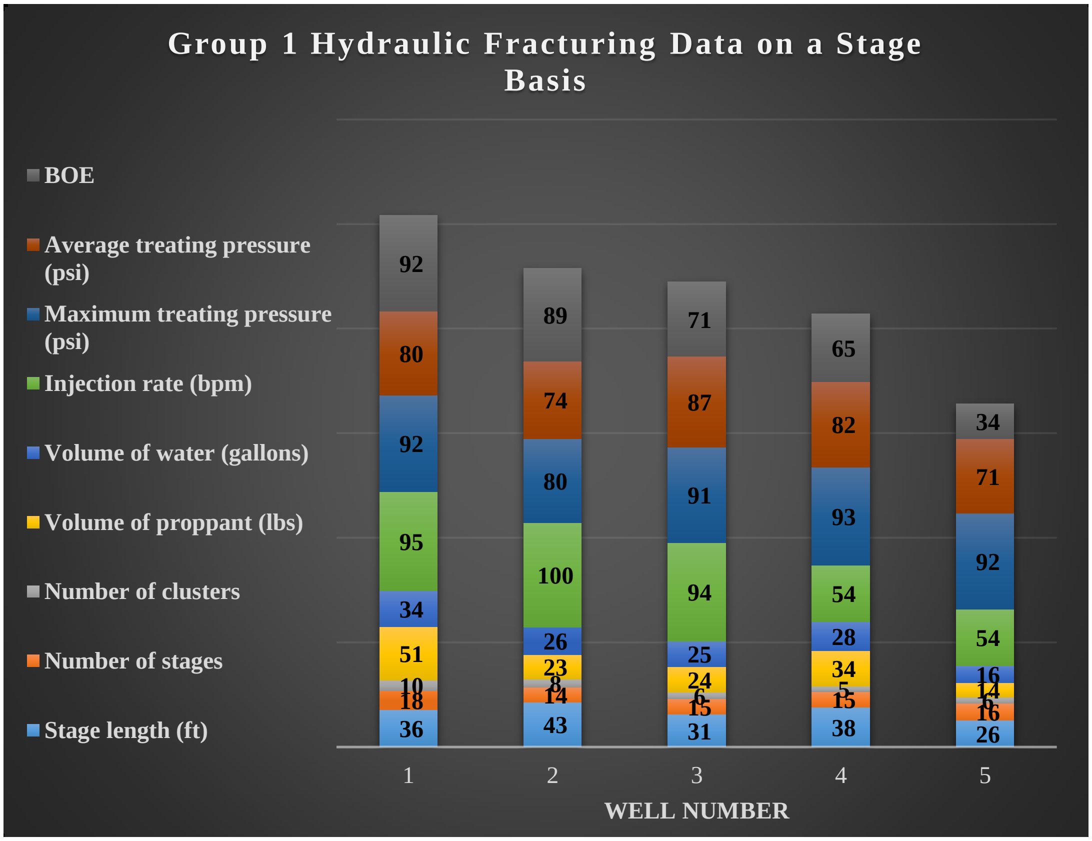


Figure 5.4: The Hydraulic Fracturing Parameters and Production Performance of a Set of Wells in Group 1

Table 5.5: Calculated Production Increase Based on the Weights Assigned to Each Parameter

Paramters	Assigned Wieghts	Wells 1-2	Wells 2-3	Wells 3-4	Wells 4-5
Lateral length (ft)	0.130	55.3	174.5	(133.0)	208.5
Number of stages	0.155	0.620	(0.155)	-	(0.155)
Number of clusters	0.135	0.270	0.270	0.135	(0.135)
Volume of proppant (lbs)	0.115	685,809	(45,154)	(168,684)	333,045
Volume of water (gallons)	0.128	316,297	(3,054)	(56,642)	198,315
Injection rate (bpm)	0.100	(0.50)	0.60	4.00	-
Maximum treating pressure (psi)	0.140	167.6	(159.7)	(29.0)	26.7
Average treating pressure (psi)	0.098	59.9	(122.5)	45.5	107.0
Calculated BOE increase		1,002,389	(48,314)	(225,438)	531,702
Actual BOE increase		40,826	17,342	9,879	43,470

The following are the main observations for a Group 1 wells:

- The Reservoir properties of Wells 1 and 2 are fairly similar as shown in Table 5.4. However, the production performance of the two wells are different which could be due to the difference in the hydraulic fracturing design. On a well basis, Well 1 has a slightly longer lateral length, higher number of stages, and larger volume of water and proppant than Well 2 as shown in Figure 5.3. This could be the reason why Well 1 has a higher production performance than Well 2. On a stage basis, Well 1 has a higher production performance than Well 2 which could be due to the fact that more volume of water and proppant are pumped down for each stages of Well 1 than the stages of Well 2 as shown in Figure 5.4. The rest of the other parameters are fairly similar between Well 1 and Well 2 on a stage basis. Thus, Well 1 and 2 fractures are treated with fairly similar pressures and injection rate, but since Well 1 has a longer lateral length with higher number of stages and clusters and has more volume of water and proppant pumped down for each stage, it has a higher production performance.

  According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between the two wells should contribute to about 1 million BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 40,826 barrels as shown in Table 5.5. The equations for the calculation of the numbers in Table 5.5 are provided in the appendix Equations C.4 and C. 5. Thus, the weights or relative importance value assigned to each parameters over estimates the production differences between the two wells.

- Wells 2 and 3 have very similar reservoir related properties as shown in Table 5.4. Even though the reservoir pressure of Well 3 is slightly larger than the reservoir pressure of Well 2, Well 2 has a higher production performance than Well 3 which could be due to the differences in the hydraulic fracturing design of the two wells. On a well basis, Wells 2 and 3 have almost the same value for all the parameters except the lateral length as shown in Figure 5.3. Well 2 has a much longer lateral length than Well 3 which could be the reason why it has a higher production performance. Well 2 is 1300 ft longer than Well 3, and the production increase in Well 2 is 17000 bbls. Assuming that this production increase is due to the increase in the lateral length, the production increase is 13 bbls/ ft, and the percent increase in production is 16 percent.

If we were to assume that the hydraulic fracturing design of the two wells are the same, and the drawdown pressure is the same, the ratio of the production performance of Well 2 to Well 3 based on the geological and reservoir properties is 0.82. This means that Well 3 is in fact supposed to have a 21 percent higher production performance than Well 2 based on the geological and reservoir properties. However, in the actual data it is indicated that Well 2 has a higher production performance than Well 3 and the ratio of the production performance of Well 2 to Well 3 is 1.16. This means that the increase in the lateral length accounts for about 37 percent improvement in production performance.

On a stage basis, the production performance of each stage in Well 2 is higher than the production performance of each stage in Well 3 even though the two wells have almost the same treatment on a stage basis. The only difference between the stages of the two wells is that the stages in Well 2 are longer than the stages in Well 3 which is also the reason why Well 2 has a longer lateral length. Thus, the increase in the lateral length has contributed to an increase in the stage length and production performance on both a stage basis and a well basis.

According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between the two wells should contribute to about -48132 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 17342 barrels as shown in Table 5.5.. Thus, the weights or relative importance value assigned to each parameters over estimates the production differences between the two wells.

- According to the reservoir related properties shown in Table 5.4, Well 4 has a much better reservoir quality than any of the other wells. However, it has a smaller production performance than Wells 1,2, and 3. This could be due to the differences in the hydraulic fracturing design. On a well basis, Well 4 has slightly higher volume of water and proppant and longer lateral length than well 3, but Well 4 has a much lower injection rate than Well 3 which could be why it has a lower production performance than the other wells as shown in Figure 5.3. On a stage basis, the length of the stages, the volume of proppant and water, and the treating pressure are fairly similar for the two wells as shown in Figure 5.4. However, the injection rate of each stage in Well 3 is much higher than the injection rate of each stages of Well 4 which could be the reason why the production performance of each stages of Well 4 is smaller than the production performance of each stages of Well 3.

According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between the two wells should contribute to about -225,444 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 9,879 barrels as shown in Table 5.5.. Thus, the weights or relative importance value assigned to each parameters over estimates the production differences between the two wells.

- Well 5 has the smallest value for almost all the parameters compared to the other wells on both a well basis and a stage basis as shown in Figures 5.3 and 5.4, and that could be the reason why it has such a low production performance compared to the other wells even though it has a better reservoir quality than the other wells as shown in Table 5.4.

  According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between Wells 4 and 5 should contribute to about 531,740 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 43470 barrels as shown in Table 5.5.. Thus, the weights or relative importance value assigned to each parameters over estimates the production differences between the two wells.

The histogram plots of the hydraulic fracturing parameters and production performance on both a well basis and stage basis for the rest of the other wells in Group 1 are shown in the appendix in Figures C.1, C.2, C.3, C.4.

Group 2:

Table 5.6: Reservoir Related Parameters of a Set of Wells in Group 2

Well Number	1	2	3	4	5
BOE	132700	113376	109542	67626	66210
Reservoir Pressure (psi)	3595	4305	7298	4705	3582
Oil Saturation	0.603	0.690	0.652	0.722	0.555
Oil Viscosity (cp)	0.375	0.628	0.473	0.540	0.556
Reservoir Thickness (ft)	381	152	177	154	192
Porosity	0.100	0.100	0.120	0.090	0.080

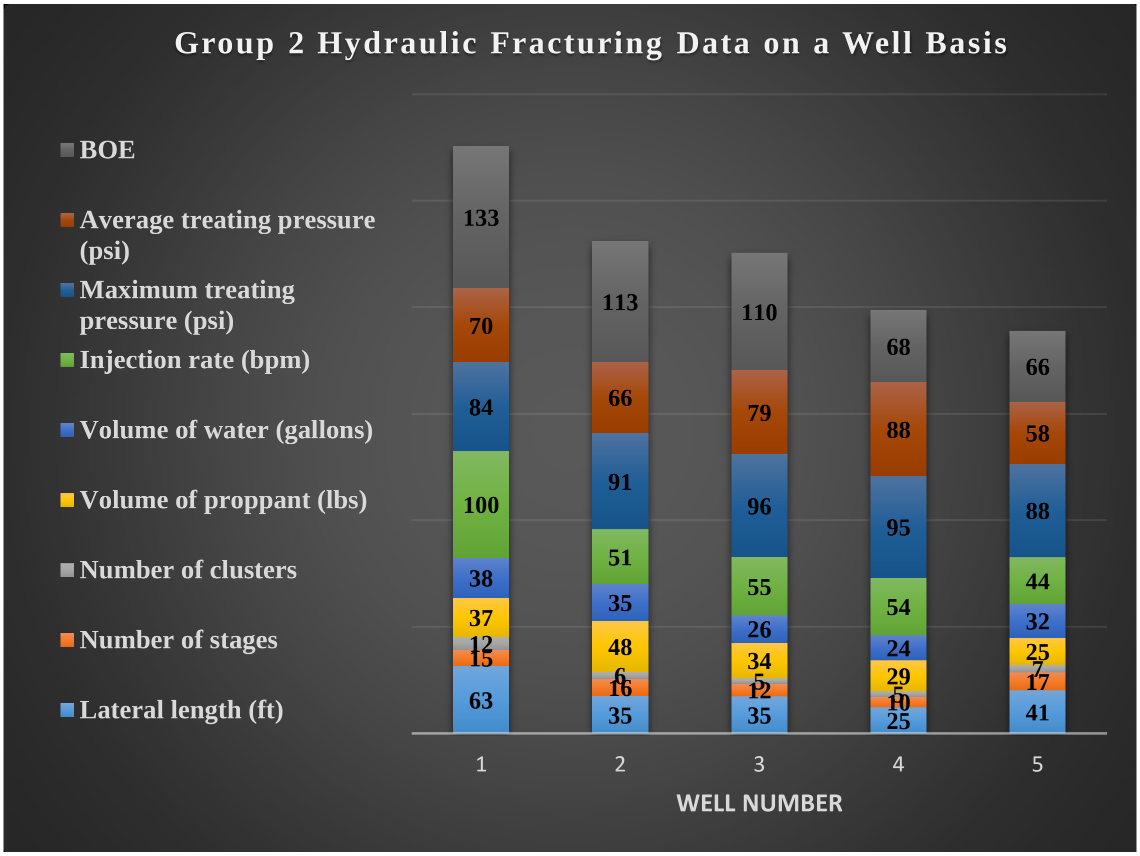


Figure 5.5: The Hydraulic Fracturing Parameters and Production Performance of a Set of Wells in Group 2

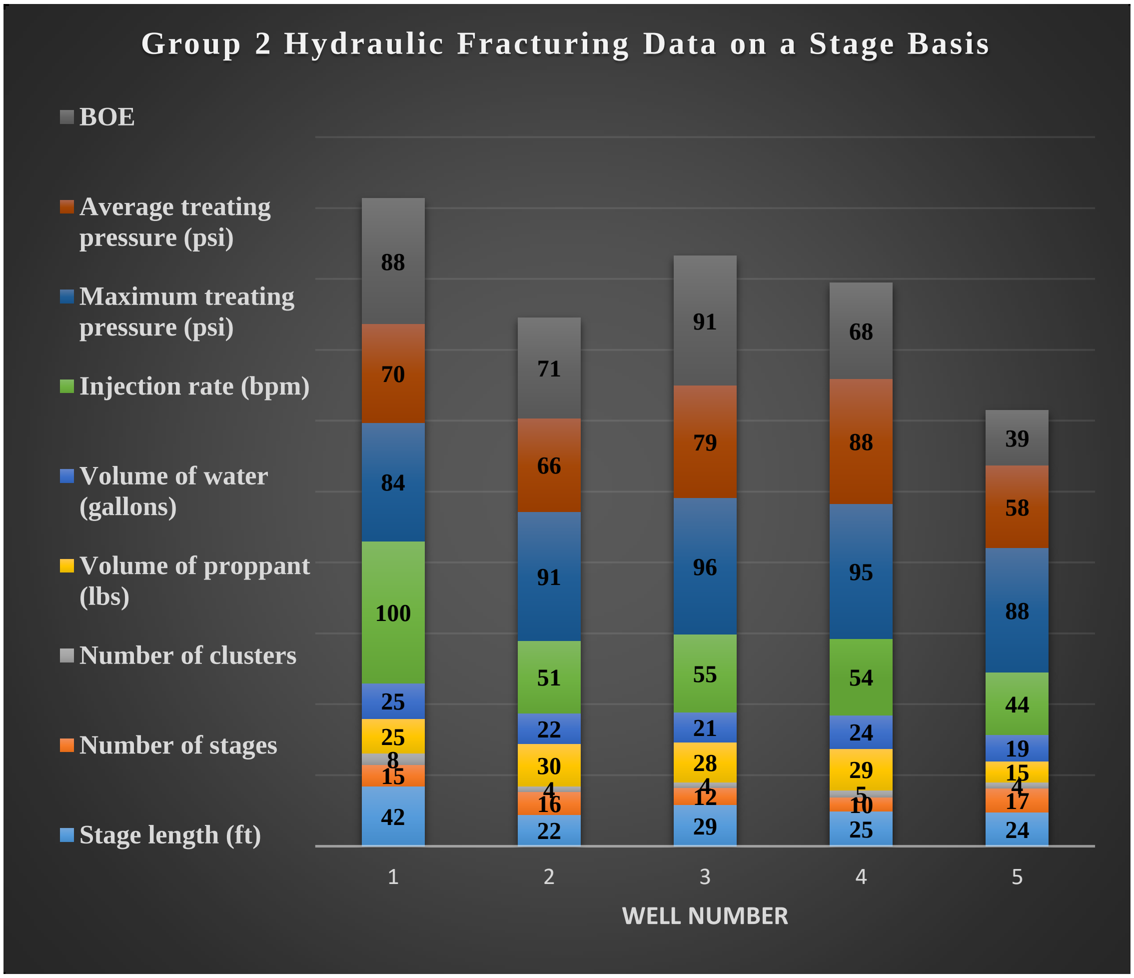


Figure 5.6: The Hydraulic Fracturing Parameters and Production Performance of a Set of Wells in Group 2

Table 5.7: Calculated Production Increase Based on the Weights Assigned to Each Parameter

Paramters	Assigned Wieghts	Wells 1-2	Wells 2-3	Wells 3-4	Wells 4-5
Lateral length (ft)	0.14	392.70	7.84	145.18	(225.68)
Number of stages	0.12	(0.12)	0.46	0.23	(0.81)
Number of clusters	0.12	0.47	-	(0.12)	0.12
Volume of proppant (lbs)	0.13	(137,971)	177,061	54,912	47,713
Volume of water (gallons)	0.12	33,022	110,192	22,748	(95,666)
Injection rate (bpm)	0.11	5.59	(0.46)	0.11	1.14
Maximum treating pressure (psi)	0.12	(87.48)	(62.16)	13.56	85.80
Average treating pressure (psi)	0.17	59.73	(219.78)	(148.01)	499.62
Calculated BOE increase		(104,579)	286,979	77,672	(47,593)
Actual BOE increase		19,324	3,834	41,916	1,416

The following are the main observations for Group 2 wells:

- The Reservoir related parameters of Wells 1 and 2 are fairly similar except for the oil viscosity and reservoir thickness parameters as shown in Table 5.6. Well 2 has slightly larger oil viscosity and smaller reservoir thickness than Well 1. The production performance of Well 2 is smaller than that of Well 1 which could be due to the differences in the oil viscosity and reservoir thickness or a combination of reservoir related parameters and the hydraulic fracturing parameters.

  On a well basis, Wells 1 and 2 have fairly similar values for all the hydraulic fracturing parameters except the lateral length and injection rate as shown in Figure 5.5. Well 1 has a much longer lateral length and much larger injection rate than Well 2 which could contribute to the production performance increase in Well 1. We previously showed that longer lateral length and higher injection rate leads to higher production performance in the Group 1 wells.

On a stage basis, we observe a similar trend as shown in Figure 5.6. The fracture treatment of the stages of Wells 1 and 2 are fairly similar except for the injection rate and stage length. The stages of Well 1 are longer and have higher injection rate than the stages of Well 2 and that could be the reason why the production performance of each of the stages of Well 1 are higher than the production performance of the stages of Well 2. Thus, Wells 1 and 2 have fairly similar fracture treatment pressures with similar volumes of water and proppant, but since the fractures in Well 1 are treated with much higher injection rate than the fractures in Well 2, and the lateral length of well 1 is much larger, Well 1 has a higher production performance

According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between the two wells should contribute to about -104,579 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 19,324 barrels as shown in Table 5.7. Thus, the weights or relative importance value assigned to each parameters are not accurate.

- According to the reservoir related parameters shown in Table 5.6, Well 3 has a better reservoir quality than Well 2. However, the production performance of Well 2 is similar, if not higher, than the production performance of Well 3. This could be due to the difference in the hydraulic fracturing parameters. On a well basis, Wells 2 and 3 have fairly similar values for all the parameters except for the volume of proppant and water and number of stages. Well 2 has a much larger number of stages and slightly larger volume of water and proppant than Well 3. This could be the reason why Well 2 has a similar production performance as that of Well 3 even though it has a lower reservoir quality than Well 3.

  However, on a stage basis, the production performance of each stages of Well 3 is higher than the production performance of the stages of Well 2. The hydraulic fracturing parameters on a stage basis are fairly similar between the two wells as shown in Figure 5.6. The stages of Well 3 are producing more oil than the stages of Well 2 even though the stages of both wells are treated almost the same way and have very similar stage length. This could be due to the fact that the reservoir quality of Well 3 is better than the reservoir quality of Well 2 as mentioned earlier.

  In short, the overall production performance of Well 2 is higher or similar to the production performance of Well 3 because Well 2 has a larger number of stages than Well 3. The production performance of each stages of Well 3 is better than the production performance of each stages of Well 2 because Well 3 has a better reservoir quality.

According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between the two wells should contribute to about 286,979 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 3,834 barrels as shown in Table 5.7. Thus, the weights or relative importance value assigned to each parameters are not accurate.

- The reservoir properties of Wells 3 and 4 are fairly similar except for the reservoir pressure as shown in Table 5.6. Well 3 has a higher reservoir pressure than Well 4. The production performance of Well 3 is much higher than the production performance of Well 4 which could be due to the higher reservoir pressure or a combination of the reservoir pressure and the hydraulic fracturing parameters. On a well basis, the hydraulic fracturing parameters are all very similar except the lateral length as shown in Figure 5.5. The lateral length of Well 3 is much longer than the lateral length of Well 4, and we have already shown that wells with longer lateral length perform better.

On a stage basis, the production performance of each stages of Well 3 is higher than the production performance of each stages of Well 4 even though the stages of both wells have been treated very similarly and have similar lengths as shown in Figure 5.6. This could be due to the higher reservoir pressure in Well 3. Thus, we could say that Well 3 has a higher production performance than Well 4 because each stages of Well 3 performs better than each stages of Well 4 due to higher reservoir quality of Well 3 and Well 3 has a longer lateral length than Well 4.

According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between the two wells should contribute to about 77,672 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 41,916 barrels as shown in Table 5.7. Thus, the weights or relative importance value assigned to each parameters are not accurate.

- The reservoir quality of Well 4 is slightly better than the reservoir quality of Well 5 due to higher reservoir pressure, oil saturation, slightly lower oil viscosity, and slightly higher porosity than the

Well 5 even though Well 5 has a slightly larger reservoir thickness as shown in Table 5.6. However, the production performance of the two wells are fairly similar. This could be due to the differences in the hydraulic fracturing design of the two wells. On a well basis, Well 4 has a higher treating pressure, injection rate than Well 5 as shown in Figure 5.5. However, Well 5 has a longer lateral length and larger number of stages than Well 4. Thus, the overall production performance of Well 5 is similar to that of Well 4 because it has a much larger number of stages and longer lateral length than Well 4.

On a stage basis, the production performance of each stages of Well 4 is much larger than the production performance of each stages of Well 5 as shown in Figure 5.6. This could be due to the fact that the volume of proppant and water, the injection rate, and the treating pressure of the stages of Well 4 are higher than those of the stages of Well 5. Thus, each of the stages of Well 4 is producing better than the stages of Well 5 not only because the Well 4 has a better reservoir quality than Well 5 but also because the stages of Well 4 are treated much better than the stages of Well 5.

In short, the overall production performance Well 4 and Well 5 are similar even though Well 4 has a slightly better reservoir quality and a better production performance on a stage basis because Well 5 has a much longer lateral length and a larger number of stages. In Well 5, the relatively poor reservoir quality and production performance on a stage basis is compensated for by increasing the lateral length and the number of stages.

According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between the two wells should contribute to about 47,593 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 1,416 barrels as shown in Table 5.7. Thus, the weights or relative importance value assigned to each parameters are not accurate.

The histogram plots of the hydraulic fracturing parameters and production performance on both a well basis and stage basis for the rest of the other wells in Group 2 are shown in the appendix in Figures C.5, C.6, C.7, C.8.C.9, C.10, C.11, C.12.

Table 5.8: Reservoir Related Parameters of a Set of Wells in Group 3

Well Number	1	2	3	4	5
BOE	109354	72270	57973	56596	11269
Reservoir Pressure (psi)	2490	2490	3075	2501	2879
Oil Saturation	0.518	0.517	0.574	0.517	0.613
Oil Viscosiy (cp)	0.609	0.760	0.567	0.676	0.866
Reservoir Thickness (ft)	400	400	331	403	469
Porosity	0.080	0.080	0.090	0.080	0.090

Group 3:

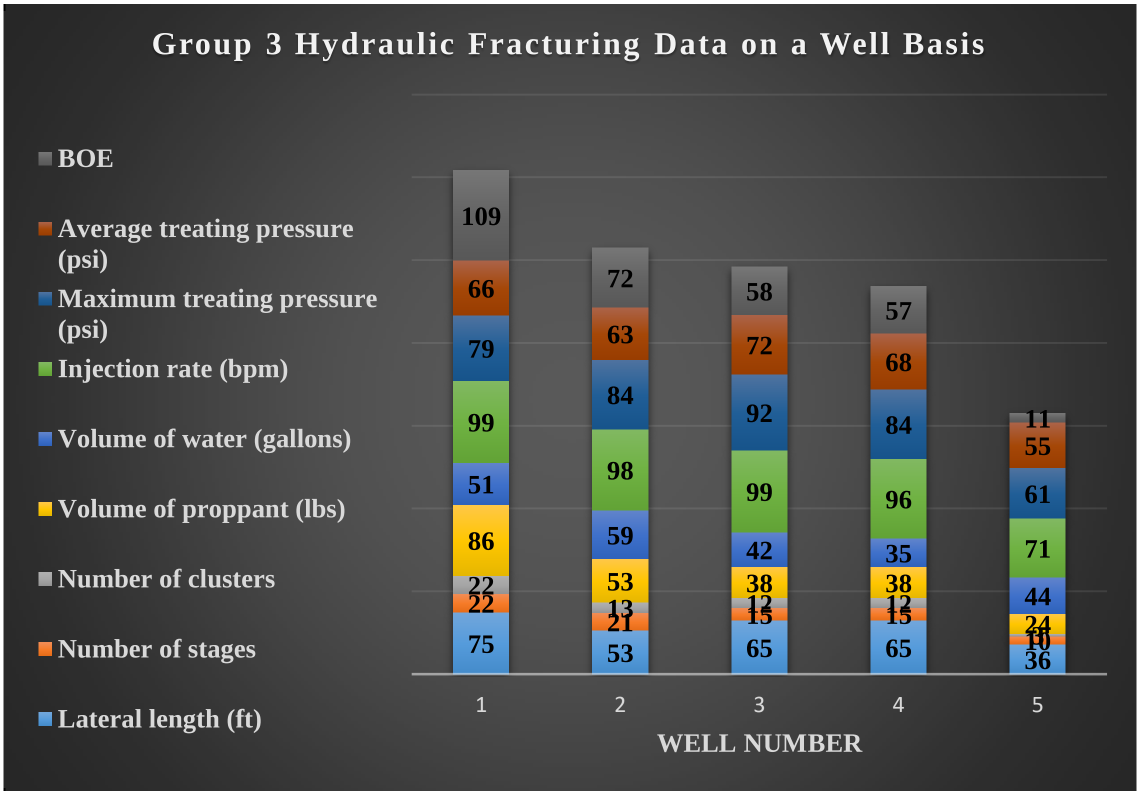


Figure 5.7: The Hydraulic Fracturing Parameters and Production Performance of a Set of Wells in Group 3

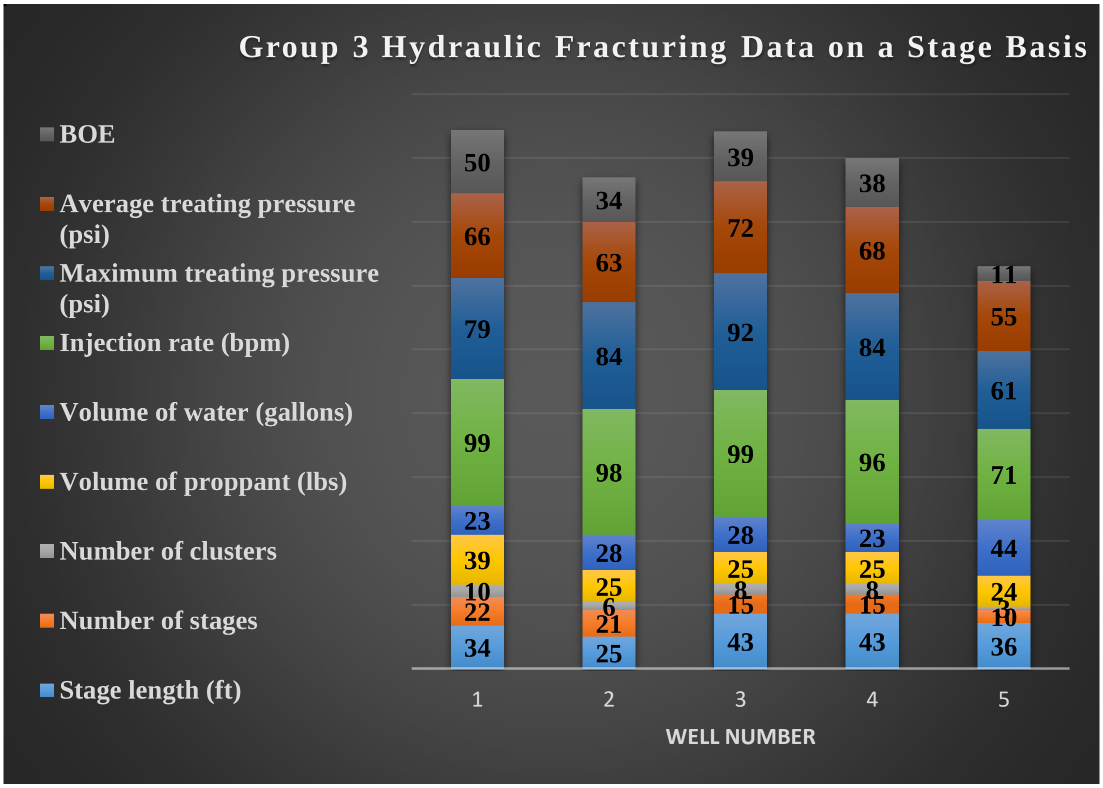


Figure 5.8: The Hydraulic Fracturing Parameters and Production Performance of a Set of Wells in Group 3

Table 5.9: Calculated Production Increase Based on the Weights Assigned to Each Parameter

Paramters	Assigned Wieghts	Wells 1-2	Wells 2-3	Wells 3-4	Wells 4-5
Lateral length (ft)	0.120	264.00	(144.00)	-	348.00
Number of stages	0.165	0.17	0.99	-	0.83
Number of clusters	0.117	0.47	(0.23)	-	0.59
Volume of proppant (lbs)	0.155	516,150	232,500	-	209,250
Volume of water (gallons)	0.085	(69,700)	142,800	63,750	(80,750)
Injection rate (bpm)	0.125	0.13	(0.13)	0.38	3.13
Maximum treating pressure (psi)	0.119	(57.12)	(89.25)	89.25	273.70
Average treating pressure (psi)	0.115	34.50	(103.50)	51.75	143.75
Calculated BOE increase		446,692	374,964	63,891	129,270
Actual BOE increase		37,000	14,000	1,000	46,000

The following are the main observations for Group 3 wells:

- As shown in Table 5.8, the reservoir related properties of Wells 1 and 2 are very similar, but their production performances are very different. This could be due to the differences in the hydraulic fracturing design. On a well basis, Wells 1 and 2 have fairly similar treatment pressure, injection, water volume, and number of stages and clusters, but Well 1 has a longer lateral length and larger volume of proppant than well 2 as shown in Figure 5.6. This could be the reason why Well 1 has a higher production performance than Well 2 as we have shown longer lateral length and larger proppant volume lead to higher production performance.

  On a stage basis, we observe a similar trend as shown in Figure 5.7. The stages of the wells have fairly similar values for all the parameters except for the proppant volume and stage length. The volume of proppant pumped and the stage length of Well 1 are higher than those of Well 2 and that could be the reason why the stages of Well 1 perform better than the stages of Well 2.

  According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between the two wells should contribute to about 446,667 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 37,000 barrels as shown in Table 5.9. Thus, the weights or relative importance value assigned to each parameters are not accurate.

- The reservoir properties of Wells 2 and 3 are fairly similar, but their production performance are very different as shown in Table 5.8. This could be due to the differences in the hydraulic fracturing design. On a well basis, Well 3 has a longer lateral length and higher treatment pressures while Well 2 has a much larger number of stages and volume of water and proppant than Well 3as shown in Figure 5.6. Since we have demonstrated that having a longer lateral length or a large number of stages contributes to the overall production performance, and in this case one has a longer later and the other has a higher number of stages, we need to look the stage performance of the wells first to be able to explain the difference in the overall production performance of the two wells.

  On a stage basis, the production performance of each stages of Well 3 is higher than the production performance of each stages of Well 2 as shown in Figure 5.7. The stages of each of the two wells have almost exactly the same values for all the parameters except the treatment pressures and stage length. The stages of Well 3 have higher treatment pressures and are much longer than the stages of Well 2, and that could be the reason why the stages of Well 3 perform better. Even though Well 3 has better stage performance, the overall production performance of Well 2 is better because it has much larger number of stages.

In short, the stages of Well 3 perform better than the stages of Well 2 because they are treated better and are longer than the stages of Well 2. However, the overall production performance of Well 2 is higher than the overall production performance of Well 3 because Well 2 has a much larger number of stages and this way compensates for relatively poor performance of the stages.

According to the weights assigned to each parameter by the regression technique, the sum of the differences in the hydraulic fracturing parameters between the two wells should contribute to about 374,700 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 14,000 barrels as shown in Table 5.9. Thus, the weights or relative importance value assigned to each parameters are not accurate.

- The reservoir properties of Wells 3 and 4 are fairly similar, and their production performances are fairly similar as shown in Table 5.8. This could also be due to the fact that the hydraulic fracturing design of the two wells are fairly similar too. On a well basis and a stage basis, the hydraulic fracturing parameters of Wells 3 and 4 are almost exactly the same as show in Figures 5.5 and 5.6.

  According to the weights assigned to each parameter by the regression technique, the sum of the slight differences in the hydraulic fracturing parameters between the two wells should contribute to about 63,844 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 1,000 barrels as shown in Table 5.9. Thus, the weights or relative importance value assigned to each parameters are not accurate.

- The reservoir properties of Wells 4 and 5 are fairly similar, but their production performance are very different as shown in Table 5.8. This could be due to the differences in the hydraulic fracturing design. On a well basis, Well 4 has a longer lateral length, larger number of stages, larger volume of proppant, and larger injection rate than Well 5 as shown in Figure 5.5.

This could be the reason why Well 4 is performing better than Well 5.

On a stage basis, we observe a similar trend as shown in Figure 5.6. The stages of Well 4 are longer and have higher injection rate and produce better than the stages of Well 5 even though the stages of Well 5 have larger volume of proppant.

Thus, the production performance of Well 4 is higher than the production performance of Well 5 because the stages of Well 4 perform better than the stages of Well 5 and Well 4 has larger number of stages.

According to the weights assigned to each parameter by the regression technique, the sum of the slight differences in the hydraulic fracturing parameters between the two wells should contribute to about 128,678 BOE difference in production performance between the two wells. However, the actual BOE difference between the two wells is 46,000 barrels as shown in Table 5.9. Thus, the weights or relative importance value assigned to each parameters are not accurate

The histogram plots of the hydraulic fracturing parameters and production performance on both a well basis and stage basis for the rest of the other wells in Group 1 are shown in the

Appendix C in Figures C.13, C.14, C.15, C.16, C.17, C.18, C.19, C.20.

The following are the main takeaways from the above observations:

The difference in production performance of multiple wells of similar reservoir and geological properties is due to and can be explained by the differences in the hydraulic fracturing parameters that have been identified as the key stimulation parameters as the following:

- Two wells with similar values of the geometry and fluid related parameters have similar production performance (Wells 3 and 4 in Group 3, Wells 2 and 3 in Group 2)
- If two wells have similar fluid related parameters, the one with longer later, higher number of stage, or higher number of cluster has a higher production performance (Wells 2 and 3 in Group 1, Wells 3 and 4 in Group 2, Wells 5 and the rest of the wells in Group 3)
- If two wells have similar geometry related parameters, the well with the highest fluid related parameters has a higher production performance (Wells 1 and 2 in Group 1)
- Increasing the geometry related parameters at the expense of the fluid related parameters or the other way around is not going to improve the production performance of the well (Wells 3 and 4 in Group 1, Wells 4 and 5 in Group 2, Wells 5 and the rest of the other wells in Group 3)

- The well with the smallest value for both geometry and fluid related parameters has the lowest production performance (Well 5 in Group 1, Well 5 in Group 3)
- Increasing both geometry and fluid related parameters together results in the highest production performance (Well 1 in Group 1, Well 1 in Group 2, Well 1 in Group 3)
- Increasing the number of stage or the lateral length contributes to the overall production performance indicated by the histogram plots of the data on a well basis, while increasing the fluid related parameters and number of clusters contributes to stage performance indicated by the histogram plots of the data on a stage basis (Wells and 3 in Group 2, Well 4 and 5 in Group 2, Wells 2 and 3 in Group 3)
- Having a better reservoir quality contributes to stage performance (Wells 3 and 4 in Group 2)

Table 5.10: Some of the Wells That Do Not Follow the Conclusions

BOE	sl	nos	nocps	vowpw	irps	mtppw	atppw	voppw
116,900	239	17	4	176,921	51	500	370	140,338
80,187	239	17	4	176,921	42	519	452	140,338
113,376	222	16	4	218,130	51	569	413	298,871
51,840	222	18	4	220,162	51	539	437	257,415
132,700	423	15	8	251,328	100	558	464	245,211
76,159	407	15	8	262,388	99	532	439	244,353
60,464	333	16	6	251,538	94	577	517	323,019
55,250	307	15	6	271,317	96	619	538	460,709

Even though the majority of the wells follow these trends mentioned above, there are several wells that do not follow these trends. As far as the data that we have regarding these wells, there is nothing too drastically different in terms of their production performance, the hydraulic fracturing parameters, and the geological and reservoir related parameters as shown in Tables 5.10 and 5.11. Therefore, we have no way of finding out or showing exactly why these wells do not follow these trends. However, several pairs of these wells show that they have almost exactly the same value for the hydraulic fracturing parameters as shown in Table 5.10.

Our theory is that since these pairs of wells have almost exactly the same value of the hydraulic fracturing parameters, there is something else that is being tested and is causing the difference in the production performance. For example, it could be that they kept all the parameters constant for a pair of wells, and they wanted to see the effect of treating multiple stages together instead of single stage treatment. This difference in treatment method is probably causing the difference in the production performance. However, since we don't have any data regarding the treatment method, we can't confirm our theory.

Based on the trends that we have observed in the data, we conclude that the eight hydraulic fracturing parameters investigated in this thesis have an impact on the production performance of the wells and contribute to the differences in production performance between the wells. The knowledge of the reservoir and geological related parameters coupled with the knowledge of the hydraulic fracturing parameters can be used to understand and explain the differences in production performance between the wells.

Table 5.11: Geological and Reservoir Related Parameters of the Wells Mentioned in Table 5.10

BOE	resep	satu	vis	Thick	Poros
116,900	4,260	0.66	0.59	162	0.11
80,187	3,599	0.56	0.61	192	0.08
113,376	4,305	0.69	0.63	152	0.10
51,840	4,859	0.76	0.58	146	0.11
132,700	3,595	0.60	0.37	381	0.10
76,159	3,591	0.60	0.39	381	0.10
60,464	4,851	0.67	0.38	165	0.11
55,250	4,854	0.67	0.41	165	0.11

The improvement of a hydraulic fracturing design from one well to another should be done on a stage basis, and the design that maximizes both the geometry and fluid related parameters together results in the highest production performance. However, more study and data is required in the future to accurately quantify the impact of each one of these parameters and to determine the optimal value of each one of these factors given the geological and reservoir properties.

5.3 COMPARISON OF THE CONCLUDING POINTS TO RELEVANT STUDIES IN THE LITERATURE

In this section we compare the concluding points to the other studies that have been done on hydraulic fracturing parameters to evaluate our analysis in relation to other studies in the literature. The following are some those studies:

Shelley and co-authors identified treatment fluid type/volume, number of frac treatments, proppant type/conductivity, perforating strategy, treatment rate and lateral length as the important parameters that have an impact on production (Shelley, et al., 2012).

Our concluding points agree with their results in regards to the treatment volume and rate, number of frac treatments, and lateral length being the important parameters. However, we do not have data for the other parameters that they included in their analysis such as proppant type and conductivity, and therefore we cannot comment on those variables.

Contrary to our study, Gao and co-authors found proppant volume, lateral length, and frac fluid to have less contributions to the early time production performance (Gao, et al., 2013). This could be due to the fact that contrary to our method, they included the reservoir related parameters such as oil API gravity, GOR, and flowing tubing pressure in their analysis instead of minimizing the effect of those parameters as we did in our analysis. In their analysis, they found that the maximum total vertical depth and the flowing tubing pressure have the most impact on the early time production performance, and GOR and API gravity were found to have the next most impact on the early production performance. However, since we interested in the effect of the hydraulic fracturing parameters, it is important to eliminate or minimize the effects of the reservoir and geological related parameters first.

LaFollette and co-authors showed that even though longer laterals led to higher total production, wells with longer laterals were found to be less efficient, and larger treatment with more proppant was found to be associated with better productivity (LaFollette, et al., 2014).

This agrees with our results as we have indicated that longer laterals could lead to higher production performance given larger treatment with larger volume of proppant and water. Whether it is efficient to have a longer later or not is an economic concern that is beyond the scope of this work.

Viswanathan and co-authors found fluid volume per stage, proppant volume per stage, and 100 mesh sand to have a small impact on KPI while high proppant concentration and cluster spacing had a significant impact on the production performance (Viswanathan, et al., 2011).

Our conclusions do not agree with this study regarding the importance of fluid volume per stage and proppant volume per stage. The rest of the other parameters are not investigated in this study. Although they have identified the volume of fluid and proppant as less important, they have identified the high proppant concentration as a high important parameter even though proappant concentration is the ratio of proppant volume to water volume.

Thus, in this section we conclude that our results agree with most of the studies that have been done on this topic, and there are some studies that do not agree which could be due to several reasons mentioned above.

5.4 EVALUATION OF THE APPROACH

When we run the NNs on the data, based on the results, we showed two main points about the relationship between the inputs which are the hydraulic fracturing parameters and the output which is the production performance parameter: First, it was shown that all the hydraulic fracturing parameters that we have included in this study have high relative importance on the production performance of the wells. In the earlier sections, we investigated this point and confirmed that these hydraulic fracturing parameters that we have analyzed have a large impact on the production performance of the wells. Thus, in regards to the importance of the hydraulic fracturing parameters, the method, NNs, has been successful at identifying the important parameters in hydraulic fracturing.

Second, the second point based on the results of the NNs method was the assigned different magnitude of relative importance for the different hydraulic fracturing parameters. For example, as shown in Figure 4.4, the relative importance of the average treating pressure (atppw) is 0.16 while the relative importance of the perforated lateral length (pli) is 0.14. This shows that the different parameters have different magnitude of impact on the production performance of the wells.

To validate these numbers and see if they reflect the importance of each parameter or not, we considered the relative importance as the weight and assigned it to the parameters as shown in Tables 5.4, 5.7, and 5.7. We first calculated the difference in the hydraulic fracturing parameters between two wells and multiplied it by the assigned weight. We then summed the weighted value of all the parameters according to the equation below to calculate the BOE increase between the two wells. We compared that calculated BOE increase to the actual BOE increase between the two wells and found out that the method over estimated and inaccurately predicted the BOE increase between the wells.

$$\mathbf{Y = 0.14pli + 0.11nos + 0.115nocps + 0.125voppw + 0.118vowpw + 0.108irps + 0.12mtppw + 0.16atppw}$$

Thus, the weight assigned to each parameter does not reflect the true weight of that parameter on the production performance, and cannot be used to quantify the impact of the hydraulic fracturing parameters on the production performance accurately.

We conclude that the NNs method was able to successfully show the important parameters but failed to quantify the impact of each parameter on the production performance of the wells accurately.

5.5 SYNOPSIS OF CHAPTER 5

In this chapter, we evaluate the results and the NNs method. In Section 5.1, we analyze the results between the three groups and drew conclusions from the analysis. It is found that even though there is a big difference in the geological and reservoir properties of the groups and that could potentially cause the difference in the production performance between the wells, the geological and reservoir proeprties alone couldn't account for the big difference in the production performance observed between the groups. There is a significant difference between the treating pressure and injection rates of the three groups and that could potentially affect the production performance of the wells and together with the differences in the geological and reservoir properties could explain why group 1 performs better than the other groups. In Section 5.2, we analyze the results between the sets of wells from each group and confirmed the importance of eight hydraulic fracturing parameters. In Section 5.3, the results of other relevant studies are presented and compared to our concluding points. It is found that most of the other studies' results agree with our concluding points. In Section 5.4, the NNs method is evaluated, and it is concluded that the method is able to accurately identified the important parameters but fails to accurately quantify the importance of the parameters. This concludes this thesis. In the next chapter, the summary of the work and the contributions of the work are presented.

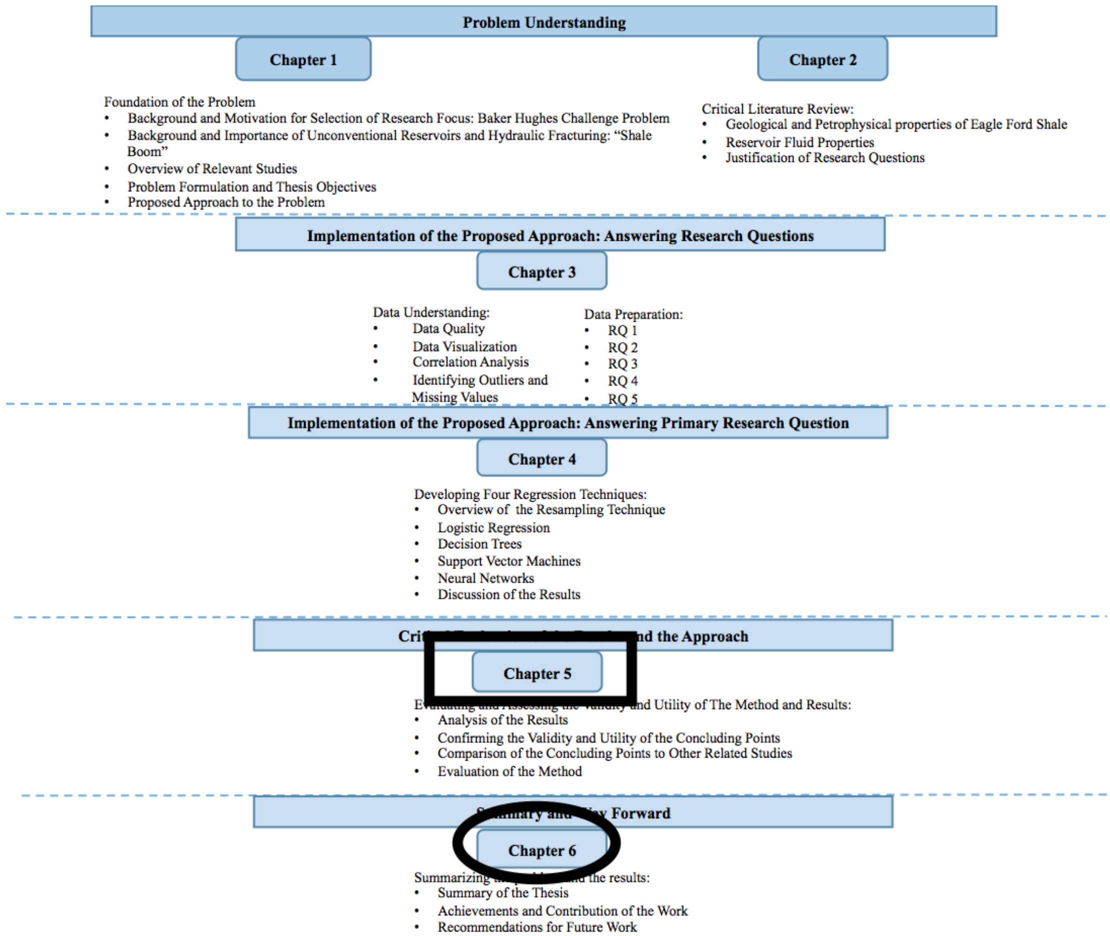


Figure 5.9: Organization of the Thesis -Present (Boxed) and Next (Circled)

CHAPTER 6

CLOSURE

In the previous chapter, the results of the NNs regression method were evaluated and compared to other studies. It was found that the eight hydraulic fracturing parameters have relatively high importance, and the method was able to successfully show that but failed to quantify the relative importance of the parameters accurately. In this chapter, a summary of the work is provided and the main points of the chapters are highlighted in Section 6.1. The research questions and the answers are briefly mentioned in Section 6.2. Then the relevant achievements and contributions of this thesis are presented in Section 6.3. The remaining questions and research gaps for future work are discussed in Section 6.4 and this brings the work to a close.

6.1 SUMMARY OF THE THESIS

The main goal of this thesis is to identify the key stimulation parameters using data mining and statistical analysis. The chapters are structured in the way that all the work leads up to identifying the key parameters which is done in Chapter 4.

In Chapter 1, the background and motivation for selection of the research focus is presented, and an overview of the Baker Hughes challenge problem that led to the selection of the research focus is also presented. The background and importance of the research focus which is hydraulic fracturing is discussed to establish context and define some of the oil and gas industry related terminologies. The problem is then defined in regards to the research focus as the following: **it is uncertain what the key stimulation parameters are in hydraulic fracturing**. The objectives are then defined, and the main objective is to identify the key stimulation parameters that have an impact on production performance. Once the problem and the objectives have been defined, the proposed approach to the problem is presented. An approach that treats hydraulic fracturing as a system composed of multiple variables is used in this thesis. Data mining and statistical analysis is selected and the CRISP-DM which is a common process for data mining and statistical analysis is used to carry out the analysis

In Chapter 2, a critical literature review is presented. This chapter is divided into two main parts. In Sections 2.1, a literature review of the properties of the Eagle Ford Shale formation is provided. The geological, petrophyiscal, and reservoir properties of the formation that affect the production performance of the wells are discussed to show the variation of those properties throughout the formation and how that might affect the performance of the wells. This information about the formation properties is also helpful and used to answer research question 2 which is done in Chapter 3. In Section 2.2, the achievements and shortcomings of the relevant studies are presented, and the justification of the research questions are then provided.

In Chapter 3, the proposed approach is implemented to answer the research questions. In Section 3.1, the data quality is first examined to make sure the data is complete, recent, accurate, and sufficient for analysis. In Section 3.2, the research questions are answered using the data and the statistical tools and methods. By answering the research questions, the data is also prepared for regression analysis which is done in the next chapter to answer the primary research question.

In Chapter 4, the proposed method is implemented to answer the primary research question. The four regression techniques are presented. An overview of each technique is discussed to show the pros and cons of each method. Then the four regression techniques are run on the Group 2 data, and the results of each technique are compared using RSME and R^2. The NNs was found to be the best performing technique, and this method was used on the rest of the data. In the results of the NNs method, it was shown that all the hydraulic fracturing parameters have relatively high importance on the production performance. To confirm the importance of each parameter and validate the results of the regression analysis, a critical evaluation of the results and the method are presented in the next chapter.

In Chapter 5, a critical evaluation of the results and the approach is discussed. The results of the NNs method are evaluated first on the three groups of wells that have different reservoir and geological properties. It was shown that the biggest difference is between Group 1 and Group 3 in terms of production performance and the hydraulic fracturing parameters. It is possible that Group 1 performs better than Group 3 because it has a better reservoir quality and higher treating pressure than Group 3. Then the results are further validated and evaluated on a set of wells of similar reservoir and geological properties from each group. The importance of the hydraulic fracturing parameters was confirmed by comparing wells of similar reservoir and geological properties. Several concluding points are drawn from the analysis, and they are compared to other relevant studies presented in Chapter 2. Our concluding points were found to agree with the results of most of the studies. The NNs method used in this thesis is also evaluated. It was concluded that NNs is a good method to show trends and patters about the importance of the hydraulic fracturing parameters but it cannot accurately quantify the impact of each parameter on production performance accurately.

6.2 ANSWERING THE RESEARCH QUESTIONS

The primary research question for this thesis is:

What are the key hydraulic fracturing parameters that affect the amount of hydrocarbons that can be recovered in Eagle Ford Shale?

To answer this question, there are several other questions that need to be answered first. The research questions are presented in Section 1.3.2, and the support and answers to these research questions are presented in Section 3.2. We have stated those questions and answered each one of them as the following:

1. ***How can the effect of a stimulation variable be measured?***

   36 months of cumulative equivalent oil is determined as a metric to measure the impact of a stimulation variable, and the details of how this question is answered are presented in Section 3.2.

2. ***How can the effects of geological and petrohpysical variables and reservoir fluid properties be eliminated or minimized?***

   The effects of geological, petrophysical, and reservoir fluid properties are minimized by dividing the wells into 3 groups based on their geological, petrophysical, and reservoir fluid properties. Wells of similar properties are grouped together, and the details of how this question is answered are presented in Section 3.2.

3. **What parameters of hydraulic fracturing should be considered for analysis?**

   The hydraulic fracturing parameters selected for analysis are the following:

- Perforated Length Interval (pli)(ft)
- Injection Rate per Stage (irps) (bpm)
- Number of Clusters per Stage (nocps)
- Volume of Proppant (voppw)(lbs)
- Volume of Water (vowpw)(gals)
- Number of Stages (nos)
- Average Treating Pressure (atppw)(psi)
- Maximum Treating Pressure (mtppw)(psi)

The details of how this question is answered are presented in Section 3.2.

4. **What kind of data mining and statistical method can be used to capture both linear and nonlinear relationships between the input and the output variables?**

The four regression techniques used and compared are the following:

- Linear Regression
- Decision Tree
- Support Vector Machines
- Neural Networks

These techniques have the ability of capturing both linear and nonlinear relationships between the input and output variables and the ability to analyze continuous data such as our data. The details of how this question is answered are presented in Section 3.2.

5. **How do we assess and compare the performance of the techniques?**

RSME and R^2 are used to assess the performance of each regression technique and compare their performance to select the best performing technique. The details of how this question is answered are presented in Section 3.2.

As we answered these questions, we got a better understanding of our data, and we were able to answer the primary research questions as the following:

What are the key hydraulic fracturing parameters that affect the amount of hydrocarbons that can be recovered in Eagle Ford Shale?

The key hydraulic fracturing parameters that affect the amount of hydrocarbons that can be recovered in Eagle Ford Shale are the following:

- Perforated Length Interval (pli)(ft)
- Injection Rate per Stage (irps) (bpm)
- Number of Clusters per Stage (nocps)
- Volume of Proppant (voppw)(lbs)
- Volume of Water (vowpw)(gals)
- Number of Stages (nos)
- Average Treating Pressure (atppw)(psi)
- Maximum Treating Pressure (mtppw)(psi)

6.3 ACHIEVMENTS AND CONTRIBUTIONS

The primary goal in this thesis is to identify the hydraulic fracturing parameters that have an impact on the production performance of Eagle Ford Shale wells. This goal has been achieved in this thesis by analyzing multiple variables of multiple shale wells and identifying the key stimulation parameters that impact the production performance of the wells. In addition to identifying the key stimulation parameters mentioned in the previous section, the following are the main findings:

- When comparing the hydraulic fracturing design of multiple wells, the comparison should be done on a stage basis to indicate the effect of fracture treatment on the production performance of each stage
- Comparing the hydraulic fracturing parameters of multiple wells on a well basis indicates the importance of increasing the number of stages and the lateral length
- The improvement of a hydraulic fracturing design from one well to another can be done by increasing the lateral length or number of stages or by improving the stage performance which requires increasing the fluid related parameters and number of clusters per stage
- When trying to improve the design of the hydraulic fracturing parameters, both the geometry and fluid related parameters need to be improved to maximize the production
- SVMs and NNs have a higher performance than LR and DT in the analysis of hydraulic fracturing data
- NNs is a good method to show trends and patterns regarding the important variables that affect production performance, but it fails to accurately quantify the impact of each parameter on production performance

- Understanding the geological and reservoir properties is the key to a better analysis of the hydraulic fracturing parameters

6.4 FUTURE WORK

In this thesis, we have identified the key stimulation parameters in hydraulic fracturing of Eagle Ford Shale wells. In Section 1.4.2, one of the limitations of this study is presented as having hydraulic fracturing data that is limited to only Eagle Ford Shale. Thus, the data represents only the Eagle Ford Shale and the results apply to only Eagle Ford Shale. This study can be further expanded by repeating the same analysis on a different set of hydraulic fracturing data that is from a different shale formation to see if the same result is obtained or the important hydraulic fracturing parameters differ from one shale formation to another. Another limitation that is mentioned in Section 1.42 is that we can not include the effect of production method since we don't have any data in regards to the production method of the wells. This study can be expanded to include or eliminate the effect of the production method and more accurately quantify the impact of each of the key stimulation parameters on the production performance.

Last but not least, as mentioned in Section 1.4.2, this analysis is purely data driven, and it can be expended by understanding how it compares to the physics of rock formation. It would be helpful to determine which one of these parameters contributes the most to the production performance of shale wells. We have shown that increasing the value of these hydraulic fracturing parameters together results in increase in production performance, and it would be helpful to know how much increasing those parameters will increase the production performance. This can even be further expended to determine some sort of equation that ties production performance to the hydraulic fracturing parameters directly, so that one can predict the production performance results from different scenarios of hydraulic fracturing design. Those parameters can be changed until an optimal value is found.

The End

References

Alqatrani, G., Britt, L. K., Otzen, G., Guzman, M., Kusanovic, G., & Dunn-Norman, S. (2016, August 24). A Statistical Approach to Fracture Optimization of the Glauconite Formation in Southern Chile. Society of Petroleum Engineers. doi:10.2118/181864-MS

Anderson (Baker Hughes Presentation). May 2015.

Beggs, H. D., And Robinson, J. R. (1975, September 1). Estimating the Viscosity of Crude Oil Systems. Society of Petroleum Engineers. doi:10.2118/5434-PA

Burton, W. A. (2013, November 11). Unconventional Completions - an Operational Comparison. Society of Petroleum Engineers. doi:10.2118/167030-MS

Cander, H., 2013, Finding Sweet Spots in Shale Liquids and Gas Plays: (with Lessons from the Eagle Ford Shale): AAPG Search and Discovery Article #40936 (2012):http://www.searchanddiscovery.com/documents/2012/40936cander/ndx_cander.pdf(Accessed June 5, 2017)

Casero, A., Adefashe, H. A., And Phelan, K. M. (2013, January 28). Open Hole Multi-Stage Completion System in Unconventional Plays: Efficiency, Effectiveness and Economic. Society of Petroleum Engineers. doi:10.2118/164009-MS

Centurion, S. M. (2011, January 1). Eagle Ford Shale: A Multistage Hydraulic Fracturing, Completion Trends and Production Outcome Study Using Practical Data Mining Techniques. Society of Petroleum Engineers. doi:10.2118/149258-MS

Centurion, S. M., Cade, R., And Luo, X. L. (2012, January 1). Eagle Ford Shale: Hydraulic Fracturing, Completion, and Production Trends: Part II. Society of Petroleum Engineers. doi:10.2118/158501-MS

Centurion, S., Cade, R., Luo, X. L., And Junca-Laplace, J.-P. (2013, September 30). Eagle Ford Shale: Hydraulic Fracturing, Completion and Production Trends, Part III. Society of Petroleum Engineers. doi:10.2118/166494-MS

Centurion, S., Junca-Laplace, J.-P., Cade, R., & Presley, G. (2014, October 27). Lessons Learned From an Eagle Ford Shale Completion Evaluation. Society of Petroleum Engineers. doi:10.2118/170827-MS

Cook, D., Downing, K., Bayer, S., Watkins, H., Sun Chee Fore, V., Stansberry, M., … Peck, D. (2014, April 1). Unconventional Asset Development Work Flow in the Eagle Ford Shale. Society of Petroleum Engineers. doi:10.2118/168973-MS

Crooks, E. (2015, April 24). The US shale revolution. Financial Times. https://www.ft.com/content/2ded7416-e930-11e4-a71a-00144feab7de (Accessed on May 10, 2017)

D. Delen. (2010, June 17). A comparative analysis of machine learning techniques for student retention management.Decision Support Systems.

Driskill, B., Suurmeyer, N., Rilling-Hall, S., Govert, A. M., And Garbowicz, A. (2012, January 1). Reservoir Description Of The Subsurface Eagle Ford Formation, Maverick Basin Area, South Texas, USA. Society of Petroleum Engineers. doi:10.2118/154528-MS

Eagle Ford Shale. *Eagle Ford Shale Geology*. Available at: http://eaglefordshale.com/geology/ Accessed on May 15, 2017)

EIA, Annual Energy Outlook, (2017, January 5). Available at : www.eia.gov/aeo (accessed on May 5, 2017)

EngineringStatistics.Available at:http://www.itl.nist.gov/div898/handbook/index.htm(Accessed on May 15, 2017)

Fan, L., Martin, R. B., Thompson, J. W., Atwood, K., Robinson, J. R., And Lindsay, G. J. (2011, January 1). An Integrated Approach for Understanding Oil and Gas Reserves Potential in Eagle Ford Shale Formation. Society of Petroleum Engineers. doi:10.2118/148751-MS

Ferguson, H. C., Thomas, C. G., Wellhoefer, B. J., And Ingram, S. R. (2012, January 1). Cementing Sleeve Fracture Completion in Eagle Ford Shale will Forever Change the Delivery of Hydraulic Fracturing. Society of Petroleum Engineers. doi:10.2118/158490-MS

Gao, C., And Gao, H. (2013, September 30). Evaluating Early-Time Eagle Ford Well Performance Using Multivariate Adaptive Regression Splines (MARS). Society of Petroleum Engineers. doi:10.2118/166462-MS

Gong, X., Tian, Y., McVay, D. A., Ayers, W. B., And Lee, J. (2013, November 5). Assessment of Eagle Ford Shale Oil and Gas Resources. Society of Petroleum Engineers. doi:10.2118/167241-MS

Hentz, T.; Ruppel, S. 2011. Regional Stratigraphic and Rock Characteristics of Eagle Ford Shale in Its Play Area: Maverick Basin to East Texas Basin. AAPG (American Association of Petroleum Geologists) 2011 Annual Convention, 10-13 April , Houston, Texas.

Hoffmann,R. Boxplot. Available at: http://www.physics.csbsju.edu/stats/box2.html (Accessed on May 25, 2017)

INTEL.(Nov 6, 2015). U.S fracking boom added 725,000 jobs-study. Reuters. http://www.reuters.com/article/usa-fracking-employment-study-idUSL8N13159X20151106 (Accessed on May 10, 2017)

Jaripatke, O., And Pandya, N. (2013, August 12). Eagle Ford Completions Optimization - An Operator's Approach. Unconventional Resources Technology Conference.

James, G., Witten, D., Hastie, T., Tibshirani, R., An Introduction to Statistical Learning, Springer, New York, (2013).

Kuhn.M., Johnson, K., Applied Predictive Modeling, Springer, New York, (2013).

LaFollette, R. F., Izadi, G., And Zhong, M. (2014, February 4). Application of Multivariate Statistical Modeling and Geographic Information Systems Pattern-Recognition Analysis to Production Results in the Eagle Ford Formation of South Texas. Society of Petroleum Engineers. doi:10.2118/168628-MS

Ling, K., Wu, X., Han, G., And Wang, S. (2016, August 24). Optimising the Multistage Fracturing Interval for Horizontal Wells in Bakken and Three Forks Formations. Society of Petroleum Engineers. doi:10.2118/181788-MS

Martin, R., Baihly, J. D., Malpani, R., Lindsay, G. J., And Atwood, W. K. (2011, January 1). Understanding Production from Eagle Ford-Austin Chalk System. Society of Petroleum Engineers. doi:10.2118/145117-MS

McMillan, M. D., Nwachukwu, C. R., And Rahnema, H. (2016, September 26). Geospatial Analysis of Compositional Variations across the Eagle Ford and Barnett Shale Plays. Society of Petroleum Engineers. doi:10.2118/181702-MS

Meisenhelder, J. (2013, July 1). Shale 2.0: From Efficient to Effective Shale Development. Society of Petroleum Engineers. doi:10.2118/0713-0018-JPT

Moro, S., Cortez, P., Rita, P. (2014, March 13). A data-driven approach to predict the success of bank telemarketing. Decision Support Systems.

Mullen, J. (2010, January 1). Petrophysical Characterization of the Eagle Ford Shale in South Texas. Society of Petroleum Engineers. doi:10.2118/138145-MS

Mullen, J., Lowry, J.C., and Nwabuoku, K.C. 2010. Lesson Learned Developing the Eagle Ford Shale. Paper 138446 Presented Tight Gas Completions Conference in San Antonio, 18-20 March.

Munir, H., Baddour, D.What is the Eagle Ford Shale?. State Impact. https://stateimpact.npr.org/texas/tag/eagle-ford-shale/page/2/.html (Accessed on May 25, 2017)

Okeahialam, I., Yang, M., Shinde, D. B., Sahai, V., Araque-Martinez, A., And Rai, R. (2017, May 1). Completion Optimization Under Constraints: An Eagle Ford Shale Case Study. Society of Petroleum Engineers. doi:10.2118/174057-PA

Passey, Q. R., Bohacs, K., Esch, W. L., Klimentidis, R., And Sinha, S. (2010, January 1). From Oil-Prone Source Rock to Gas-Producing Shale Reservoir - Geologic and Petrophysical Characterization of Unconventional Shale Gas Reservoirs. Society of Petroleum Engineers. doi:10.2118/131350-MS

Priyadharshini. Data Mining vs. Statistics- How Are They Different?. Simplilearn. https://www.simplilearn.com/data-mining-vs-statistics-article (Accessed on June 5, 2017)

Rickman, R., Mullen, M. J., Petre, J. E., Grieser, W. V., And Kundert, D. (2008, January 1). A Practical Use of Shale Petrophysics for Stimulation Design Optimization: All Shale Plays Are Not Clones of the Barnett Shale. Society of Petroleum Engineers. doi:10.2118/115258-MS

Shelley, F.R. (2016, March 1). WorlOil. http://www.worldoil.com/magazine/2016/march-2016 (Accessed on April 10, 2016)

Shelley, R. F., Saugier, L. D., Al-Tailji, W., Guliyev, N., And Shah, K. (2012, January 1). Understanding Hydraulic Fracture Stimulated Horizontal Eagle Ford Completions. Society of Petroleum Engineers. doi:10.2118/152533-MS

Taylor, R. S., Barree, R. D., Aguilera, R., Hotch, O., And Storozhenko, K. K. (2011, January 1). Why Not to Base Economic Evaluations on Initial Production Alone. Society of Petroleum Engineers. doi:10.2118/148680-MS

Tian, Y., Ayers, W. B., And McCain, W. D. (2013, March 26). The Eagle Ford Shale Play, South Texas: Regional Variations in Fluid Types, Hydrocarbon Production and Reservoir Properties. International Petroleum Technology Conference. doi:10.2523/IPTC-16808-MS

Tian, Y., Ayers, W. B., And McCain, W. D. (2014, April 1). Regional Impacts of Lithologic Cyclicity and Reservoir and Fluid Properties on Eagle Ford Shale Well Performance. Society of Petroleum Engineers. doi:10.2118/169007-MS

Tian, Y., Ayers, W. B., Sang, H., McCain, W. D., And Ehlig-Economides, C. (2017, February 15). Quantitative Evaluation of Key Geological Controls on Regional Eagle Ford Shale Production Using Spatial Statistics. Society of Petroleum Engineers. doi:10.2118/185025-MS

Tian, Yao, Ayers, W. B., and McCain, W. D., Jr. 2012. Regional Analysis of Stratigraphy, Reservoir Characteristics, and Fluid Phases in the Eagle Ford Shale. Paper presented at Gulf Coast Association of Geological Societies and the Gulf Coast Section of SEPM, Austin, Texas, 21-24 October.

Viswanathan, A., Altman, R. M., Oussoltsev, D., Kanneganti, K. T., Xu, J., Grant, D., Rhein, T. K. (2011, January 1). Completion Evaluation of the Eagle Ford Formation With Heterogeneous Proppant Placement. Society of Petroleum Engineers. doi:10.2118/149390-MS

Voneiff, G., Sadeghi, S., Bastian, P., Wolters, B., Jochen, J., Chow, B., … Gatens, M. (2013, November 5). A Well Performance Model Based on Multivariate Analysis of Completion and Production Data from Horizontal Wells in the Montney

Formation in British Columbia. Society of Petroleum Engineers. doi:10.2118/167154-MS

Wachtmeister, H., Lund, L., Aleklett, K. et al. Nat Resour Res (2017, January 19). Production Decline Curves of Tight Oil Wells in Eagle Ford Shale.Natural Resource Research. doi:10.1007/s11053-016-9323-2

Appendix A: Complete Data of the 65 Wells

The hydraulic fracturing data used in this thesis is provided by Baker Hughes, and it includes 65 wells in Eagle Ford Shale as shown in Tables A.1 and A.2 Due to confidentially agreement, the name and exact location of the wells cannot and will not be disclosed in this thesis. The data shown in Tables A.1 and A.2 are used throughout this thesis to do the datamining and statistical analysis on.

Acronym definitions for Tables A.1:

- Perforated Length Interval (pli)(ft)
- Injection Rate per Stage (irps) (bpm)
- Number of Clusters per Stage (nocps)
- Length of a stage (los) (ft)
- Volume of Proppant per Well (voppw)(lbs)
- Volume of Water per Well (vowpw)(gals)
- Number of Stages per Well (nos)
- Average Treating Pressure per Well (atppw)(psi)
- Maximum Treating Pressure per Well (mtppw)(psi)
- Barrels of Equivalent Oil (BOE) (barrels)

Acronym definitions for the Table A.2:

- Barrels of Equivalent Oil (BOE) (barrels)
- Reservoir Pressure (resep) (psi)
- Oil Saturation (satu)
- Oil viscosity (vis) (cp)
- Reservoir thickness (Thick) (ft)
- Porosity (Poros)

Table A.1: Hydraulic Fracturing Data of the 65 Wells with BOE

Well Num	BOE	pli	nos	los	nocps	vowpw	irps	mtppw	atppw	voppw
1	97577.8	4570	15	305	5	4574346	60	7033	6043	4313659
2	528414	5357	14	383	8	3135846	57	10849	9681	4916900
3	711984	5691	11	407	5	2791656	47	9813	9053	3562360
4	100861	3694	18	205	4	4844868	57	8177	7128	3854420
5	267997	6161	25	252	5	6697446	97	8680	7762	6665250
6	11269.3	3584	10	358	3	4358424	71	6143	5469	2444080
7	77886.8	6660	26	256	5	6354810	95	8228	6973	5909040
8	31764.3	6166	14	440	9	3714270	94	8075	6007	3444280
9	72269.9	5255	21	250	6	5917674	98	8364	6393	5194620
10	50210.9	6988	19	368	8	4885608	95	8266	6372	4705672
11	41083.2	5713	16	271	10	4130154	100	8267	7011	6274780
12	34162.3	5756	24	240	5	6382740	97	7499	6261	5956020
13	52875.9	5566	23	242	5	6131538	96	8033	6608	5658730
14	60630.5	6891	26	265	5	7045878	99	7943	7136	6525400
15	60927.2	5985	20	299	6	3995166	58	8415	6919	3780080
16	47209.9	4234	21	202	4	5188134	47	7677	6412	6012860
17	30518.3	6698	11	609	10	3553074	96	7533	6794	3229480
18	142600	5011	16	313	4	3670926	56	7074	5925	3171920
19	42141.8	6809	16	351	10	4140486	99	8523	7031	3931980
20	54420.9	4869	12	489	8	3159702	100	7934	7059	2949820
21	56596	6518	15	351	8	3390912	96	8367	6753	3706040
22	25839.7	6977	16	351	8	4181730	97	8549	7156	3938940
23	76158.5	6102	15	351	8	3935820	99	7986	6582	3665300
24	23459.2	4410	15	246	5	7107030	77	8367	6453	2767940
25	132700	6349	15	351	8	3769920	100	8370	6965	3678160
26	109354	7531	22	271	10	4978050	99	7958	6693	8590400
27	57973.1	6539	15	337	8	4157202	99	9138	7263	3705440
28	53038.2	5795	20	211	8	5307792	99	8409	6819	8052020
29	109542	3488	12	260	4	2556246	55	9617	7935	3365440
30	153234	4032	17	257	5	4949742	57	8900	8105	3718582
31	129667	7331	18	403	5	5070719	96	8898	7568	4691720
32	31320.8	3733	13	297	8	1813938	89	9328	8723	1626260
33	55250.3	4598	15	307	6	4712148	96	9290	8075	5298800
34	60463.7	5327	16	333	6	4180974	94	9228	8277	3933170
35	116900	4063	17	253	4	3007662	51	8503	6294	2385746
36	80187.1	4063	17	253	4	3163272	42	8821	7677	2341040
37	69793.3	3331	18	185	4	3848040	59	7403	6085	3576760
38	132570	6917	12	459	8	3916668	98	8861	7639	3720580
39	66210.4	4063	17	252	4	3174192	44	8789	5804	2544440
40	124307	6018	14	430	8	3673740	100	7993	7433	3271160
41	113376	3544	16	136	4	3490074	51	9099	6603	4781930
42	51840.1	3996	18	136	4	3962910	51	9699	7869	4633470
43	52987.3	5489	16	345	10	4254558	93	9490	9050	6240360
44	56553.4	5960	21	211	8	5707590	92	9342	8379	8115455
45	165133	6443	18	271	10	6144810	95	9190	8044	9234720
46	67626	2451	10	201	5	2363466	54	9504	8832	2926140
47	57736.3	5159	15	271	10	5131518	91	9576	8633	6669000
48	249009	5851	27	121	5	5075322	92	9230	8559	6800870

49	150076	3371	14	241	4	3738840	57	8566	7845	3137260
50	177911	3375	16	169	4	4642092	58	9311	8205	3273941
51	53616.1	4095	16	256	6	2590770	54	9150	7127	2234580
52	70663.8	3811	16	238	6	2689386	59	8825	7713	2383160
53	99260.2	6768	20	338	6	6486144	91	9301	8613	6314120
54	89397.8	4550	15	303	8	4007136	96	9129	8401	3003480
55	73230.6	4437	14	241	5	3491544	94	9461	8814	3570000
56	41064.4	4912	11	447	10	3237738	95	9218	8715	3278160
57	106965	4676	15	312	6	3697596	94	9134	8683	3663800
58	97086	5699	15	309	5	4140108	54	9341	8219	5130620
59	104491	4669	16	292	5	5140548	63	9343	8315	2352560
60	80829.7	5393	16	121	10	3986136	85	9445	8933	3782640
61	10474.4	4089	11	247	5	5189898	78	7628	5810	3214660
62	40290	2201	9	201	5	2477958	60	8443	6115	3133380
63	78608.3	4408	20	220	5	5375958	99	6696	6193	5340420
64	227417	7670	17	451	7	6301512	98	8628	6008	6012040
65	88782.3	6652	19	271	10	4071900	98	8134	6111	6087380

Table A.2: Reservoir Related Data of the 65 Wells with BOE

Well Num	BOE	resep	satu	vis	Thick	Poros
1	97577.8	4434	0.61738	0.72774	124.84	0.09
2	528414	9397	0.61446	0.21789	166.97	0.1
3	711984	9746	0.65918	0.26884	213.52	0.11
4	100861	3328	0.61838	0.89052	229.6	0.08
5	267997	3663	0.63138	0.73135	204.85	0.09
6	11269.3	2879	0.61296	0.86577	468.74	0.09
7	77886.8	3290	0.54622	0.58097	305.52	0.08
8	31764.3	2484	0.51611	0.69999	402.91	0.08
9	72269.9	2490	0.51742	0.76036	399.96	0.08
10	50210.9	2489	0.51902	0.64069	402.34	0.08
11	41083.2	2477	0.51695	0.72869	406.44	0.08
12	34162.3	2472	0.51779	0.71712	407.81	0.08
13	52875.9	3256	0.57502	0.62234	310.63	0.09
14	60630.5	3082	0.57342	0.56686	329.32	0.09
15	60927.2	2738	0.53225	0.64154	384.68	0.09
16	47209.9	3289	0.59675	0.63952	273.63	0.08
17	30518.3	2739	0.53225	0.5956	384.68	0.09
18	142600	3963	0.61889	0.31043	369.4	0.11
19	42141.8	2461	0.52562	0.68841	415.57	0.08
20	54420.9	3253	0.5751	0.59714	310.76	0.09
21	56596	2501	0.51663	0.67631	402.5	0.08
22	25839.7	2991	0.60569	0.472	359.24	0.08
23	76158.5	3591	0.60289	0.38948	380.91	0.1
24	23459.2	2988	0.59973	0.58224	414.14	0.09

25	132700	3595	0.60276	0.37461	380.89	0.1
26	109354	2490	0.51773	0.60877	400.08	0.08
27	57973.1	3075	0.57356	0.56707	330.78	0.09
28	53038.2	3220	0.61104	0.50801	414.18	0.09
29	109542	7298	0.65239	0.47284	177.27	0.12
30	153234	8453	0.57242	0.32378	149.74	0.09
31	129667	3665	0.58628	0.54978	181.69	0.08
32	31320.8	4855	0.67916	0.56402	163.32	0.11
33	55250.3	4854	0.6687	0.40689	165.07	0.11
34	60463.7	4851	0.6687	0.38474	165.1	0.11
35	116900	4260	0.65577	0.59251	161.91	0.11
36	80187.1	3599	0.56154	0.61463	191.9	0.08
37	69793.3	4034	0.6787	0.91219	149.15	0.1
38	132570	4451	0.67129	0.43987	163.92	0.1
39	66210.4	3582	0.55518	0.5557	191.8	0.08
40	124307	4412	0.67155	0.47224	163.82	0.1
41	113376	4305	0.69018	0.62794	151.9	0.1
42	51840.1	4859	0.75534	0.58233	145.52	0.11
43	52987.3	5081	0.67276	0.32453	164.76	0.11
44	56553.4	4262	0.71457	0.49342	147.39	0.1
45	165133	4262	0.71401	0.47505	147.4	0.1
46	67626	4705	0.72213	0.53972	154.21	0.09
47	57736.3	4237	0.67898	0.42938	156.95	0.1
48	249009	4236	0.68122	0.41671	156.71	0.1
49	150076	8584	0.56006	0.34036	149.74	0.09
50	177911	8476	0.56573	0.3277	150.15	0.09
51	53616.1	7208	0.7	0.36101	181.18	0.1
52	70663.8	7085	0.70428	0.50033	180.7	0.1
53	99260.2	5143	0.67019	0.50543	148.95	0.09
54	89397.8	5301	0.65414	0.53218	146.71	0.09
55	73230.6	5468	0.65221	0.50425	149.86	0.09
56	41064.4	5766	0.64954	0.50193	150.88	0.09
57	106965	5781	0.6489	0.50485	150.46	0.09
58	97086	7888	0.6664	0.27408	165.41	0.1
59	104491	7228	0.63621	0.36033	150.52	0.1
60	80829.7	7375	0.68824	0.36432	181.6	0.1
61	10474.4	2462	0.58368	0.77855	513.48	0.09
62	40290	4397	0.57462	1.07169	134.72	0.11
63	78608.3	2641	0.56442	1.18613	252.77	0.09
64	227417	2636	0.55954	0.78313	263.3	0.09
65	88782.3	2226	0.54827	0.83971	407.09	0.09

Appendix B: Scatterplot Matrix of the Geological and Reservoir Properties of the Groups of Wells

A scatterplot matrix of the complete geological and reservoir related data of the wells of each group is shown in Figures B.1, B.2, and B.3. The data shown in Figures B.1, B.2, and B.3 are used in Section 3.2.2 to answer the research question 2.

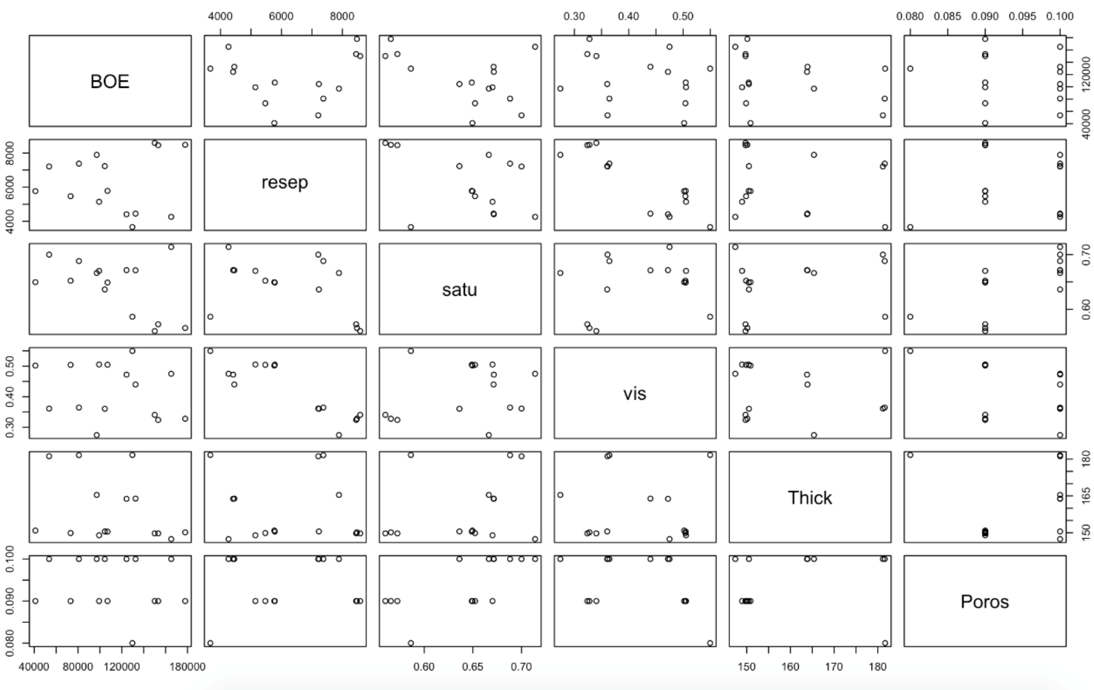


Figure B.1: Geological and Reservoir Properties of Group 1

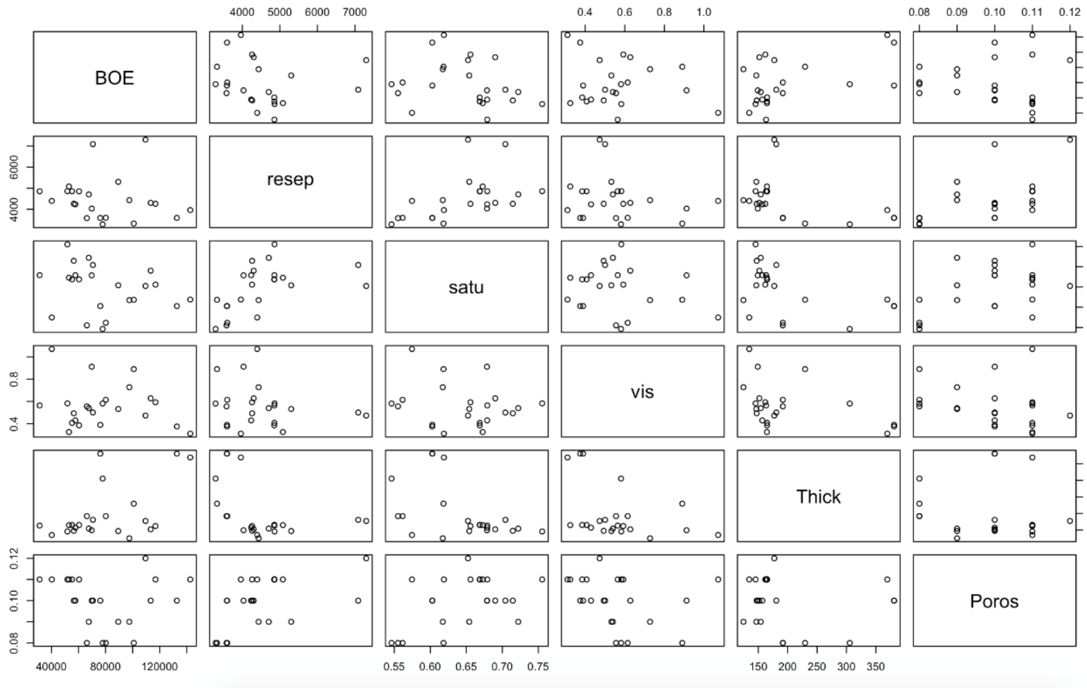


Figure B.2: Geological and Reservoir Properties of Group 2

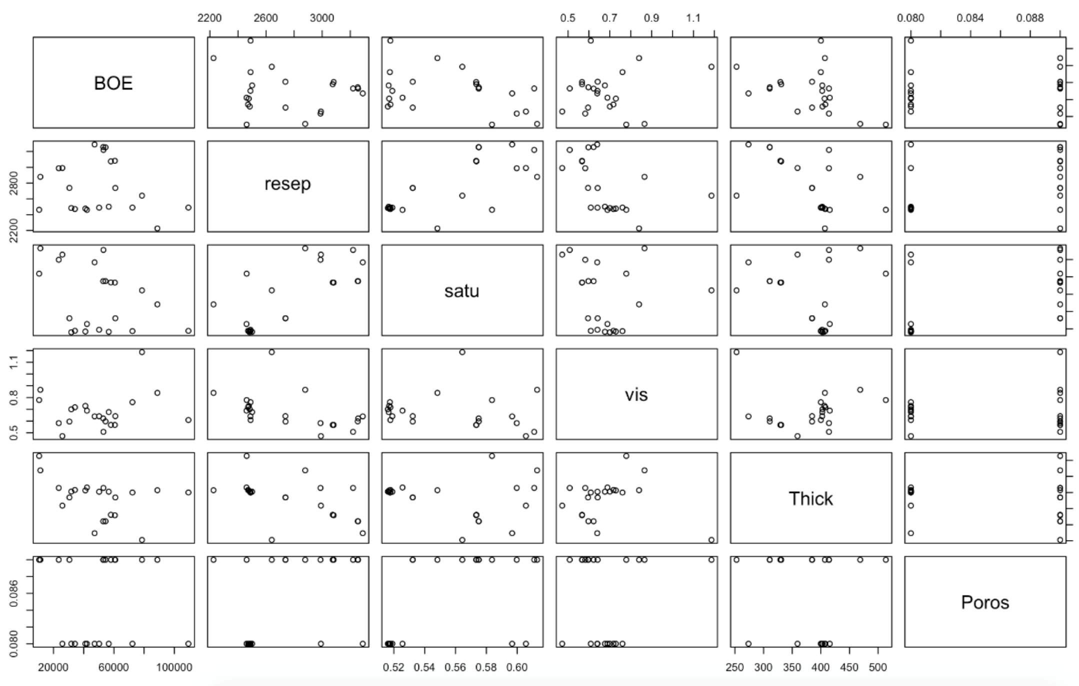


Figure B.3: Geological and Reservoir Properties of Group 3

Appendix C: Equations and Hydraulic Fracturing Data on a Well Basis and Stage Basis for the Rest of the Other Wells

The equation to calculate the average value of each parameter in each group is shown in Equations C.1 and C.2, and the equation to calculate percent increase is shown in Equation C.3. These equations are used to calculate the values in Table 5.2 in Section 5.1. Equation C.4 is used I Section 5.1 to calculate the difference in value of a parameter between two wells, and Equation C.5 is used to assign weight to each parameter and calculate the increase in production performance of a well in Section 5.2.

Equation C.1:

Average value of each parameter= sum of the value of the parameter for all the wells in the group/ number of wells in the group

Equation C.2:

Scaling parameter= the value of the parameter/ the scaling number

Equation C. 3:

Percent increase= (new number-original number / original number)*100

Equation C.4:

difference in value= new number or parameter – original number or parameter

Equation C.5: Assigning weight to each parameter

Y= 0.14pli + 0.11nos + 0.115nocps + 0.125voppw + 0.118vowpw +0.108irps + 0.12mtppw + 0.16atppw

Y= is the increase in production performance caused by the difference in the hydraulic fracturing parameters

Group 1:

The histogram plots of the hydraulic fracturing parameters and production performance on both a well basis and stage basis for the rest of the other wells in Group 1 are shown in Figures C.1, C.2, C.3, C.4. The observations mentioned in Section 5.2 for Group 1 wells apply to the wells shown in Figures C.1, C.2, C.3, and C.4.

Second set of wells

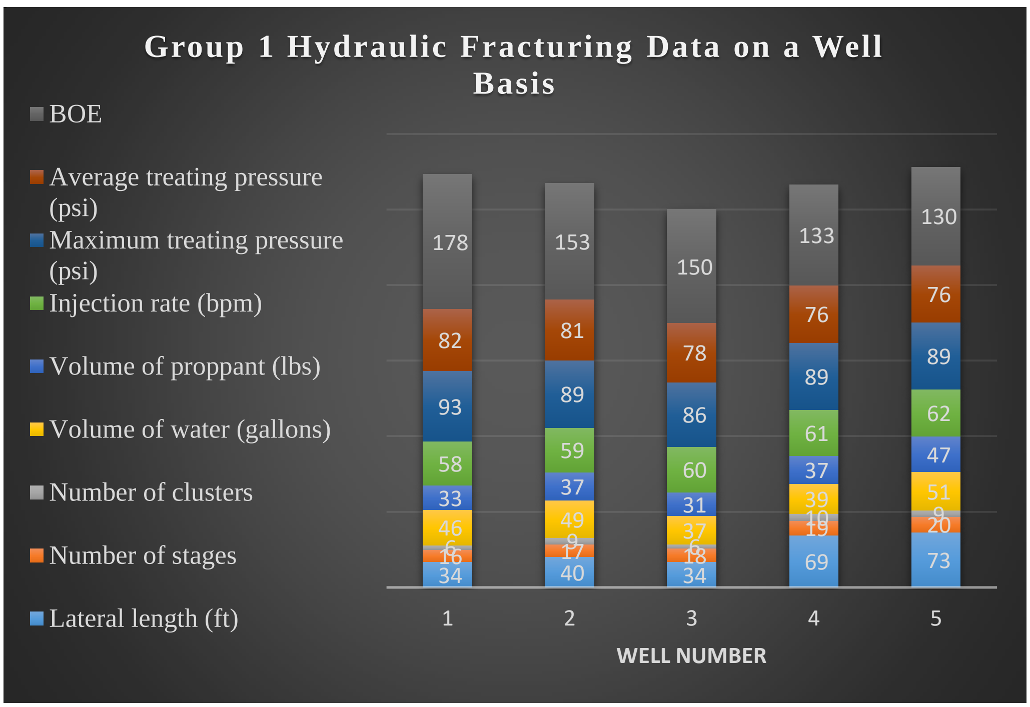


Figure C.1: The Hydraulic Fracturing Parameters and Production Performance of a Second Set of Wells in Group 1

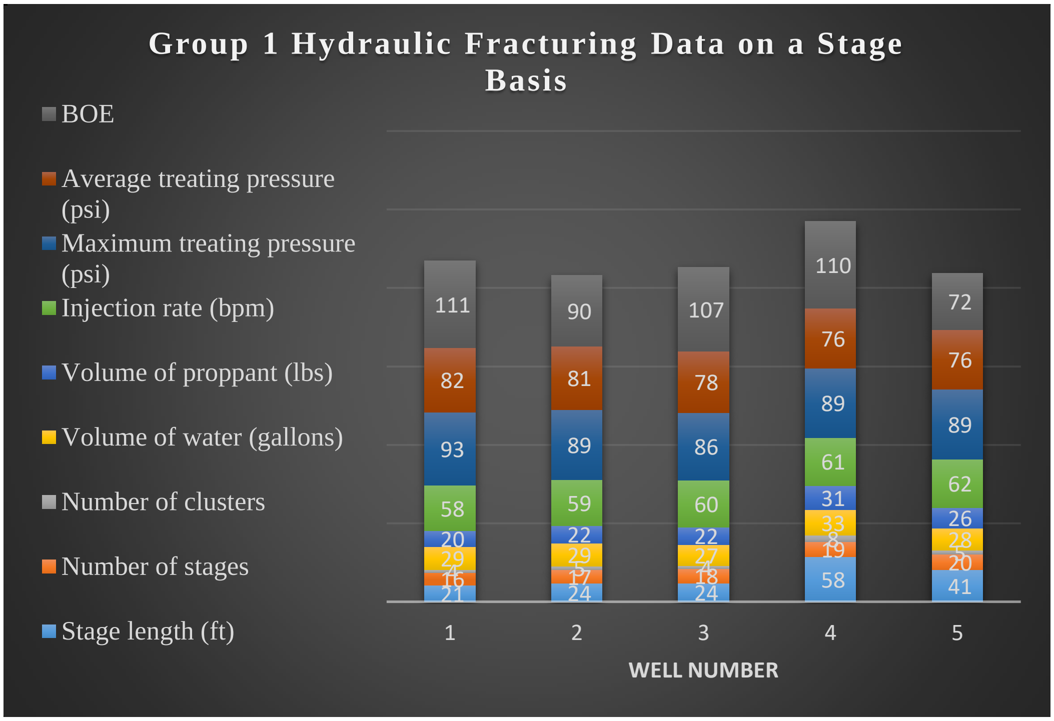


Figure C.2: The Hydraulic Fracturing Parameters and Production Performance of a Second Set of Wells in Group 1

Third Set of Wells:

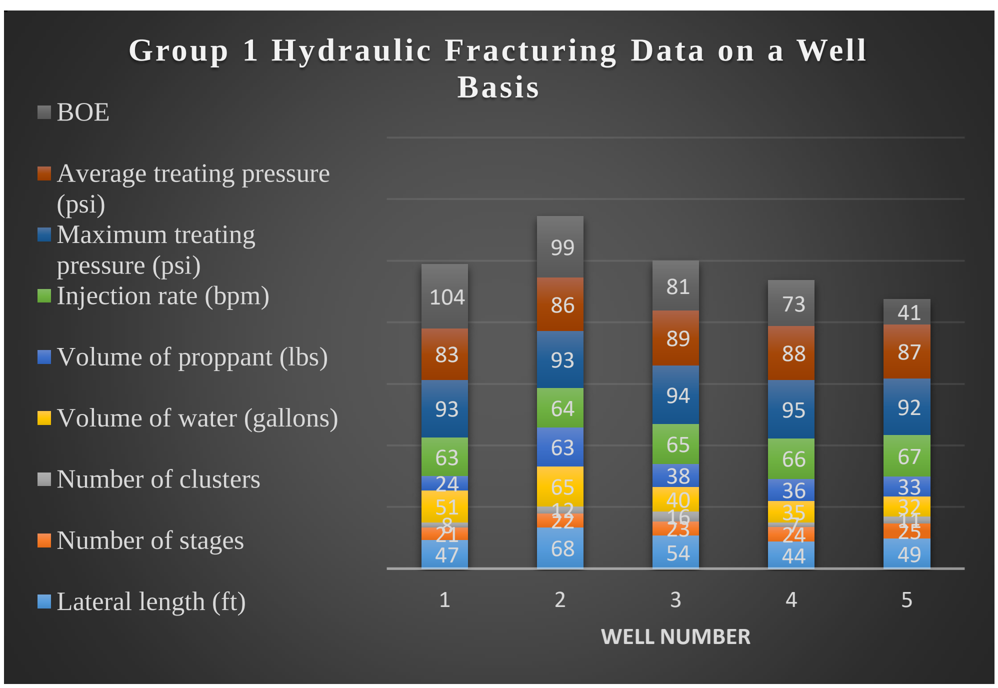


Figure C.3: The Hydraulic Fracturing Parameters and Production Performance of a Third Set of Wells in Group 1

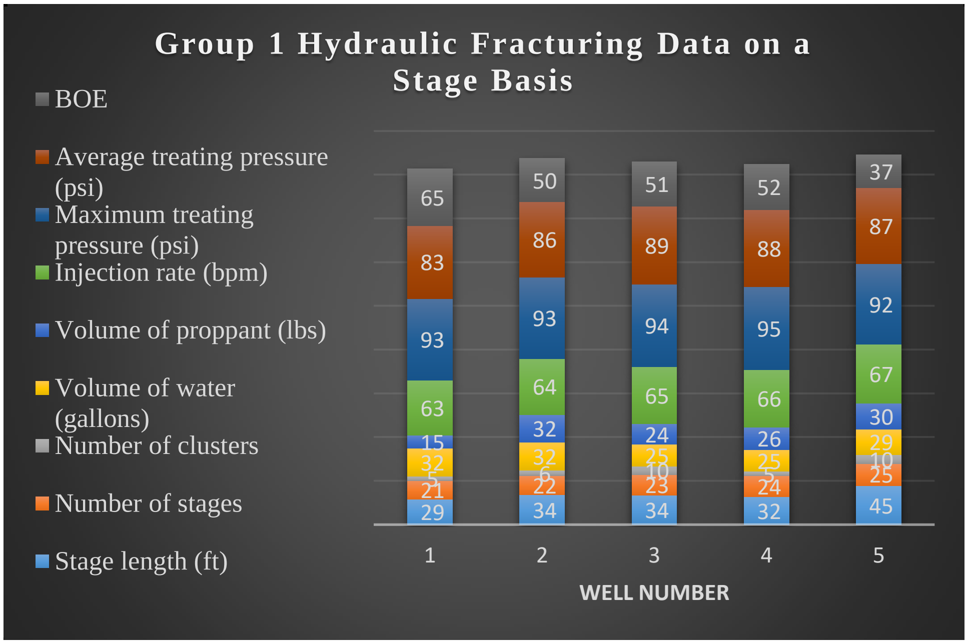


Figure C.4: The Hydraulic Fracturing Parameters and Production Performance of a Third Set of Wells in Group 1

Group 2:

The histogram plots of the hydraulic fracturing parameters and production performance on both a well basis and stage basis for the rest of the other wells in Group 2 are shown in Figures C.5, C.6, C.7, C.8.C.9, C.10, C.11, C.12. The observations mentioned in Section 5.2 for Group 2 wells apply to the wells shown in Figures C.5, C.6, C.7, C.8.C.9, C.10, C.11, C.12.

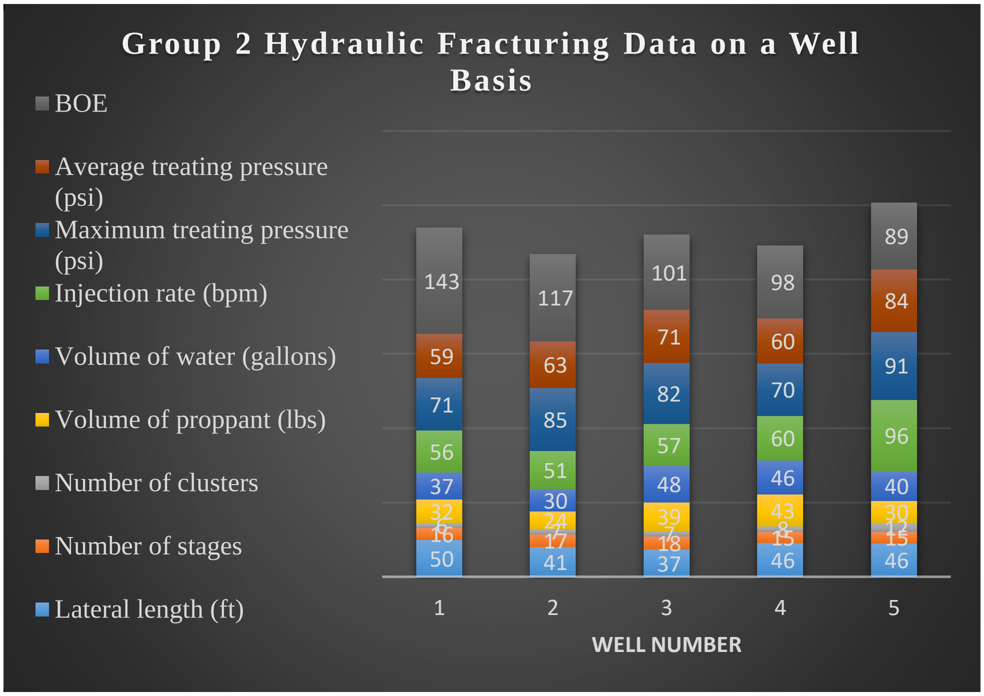


Figure C.5: The Hydraulic Fracturing Parameters and Production Performance of a Second Set of Wells in Group 2

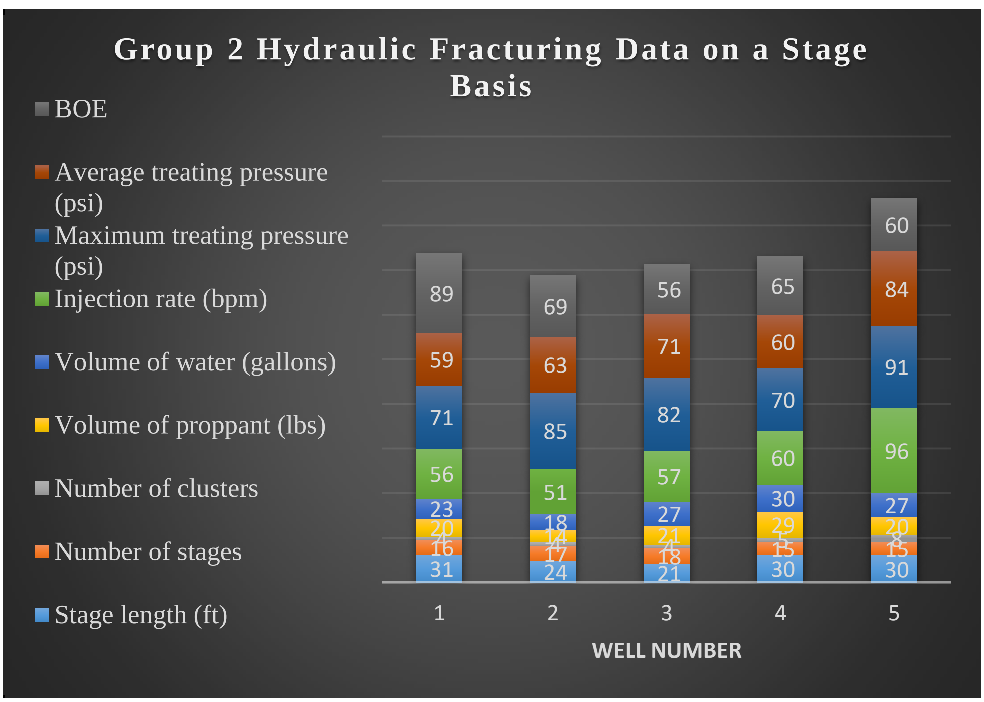


Figure C.6: The Hydraulic Fracturing Parameters and Production Performance of a Second Set of Wells in Group 2

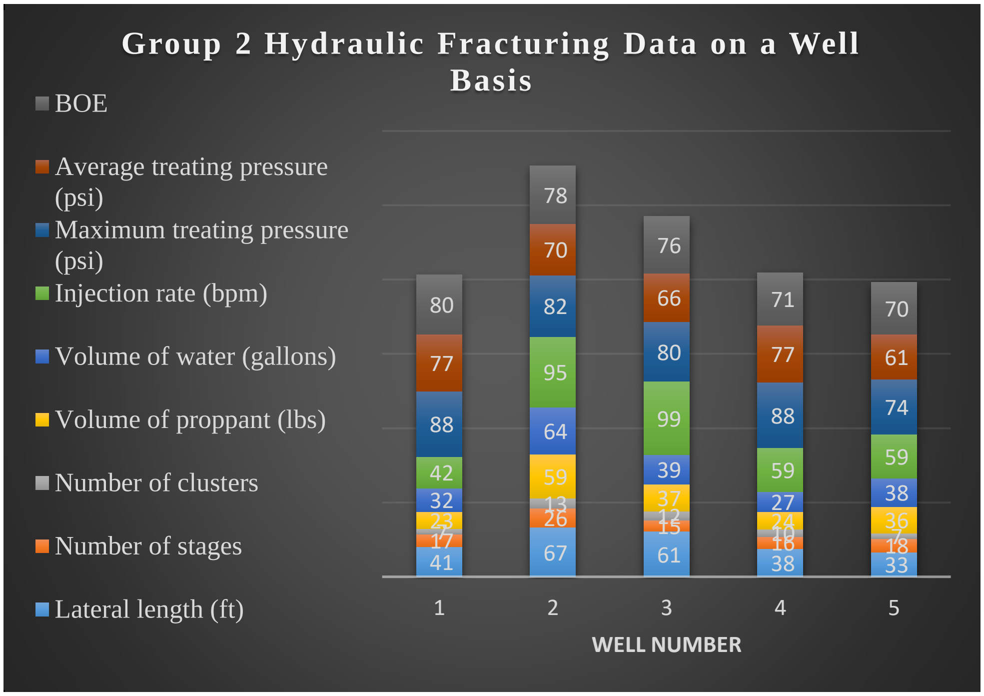


Figure C.7: The Hydraulic Fracturing Parameters and Production Performance of a Third Set of Wells in Group 2

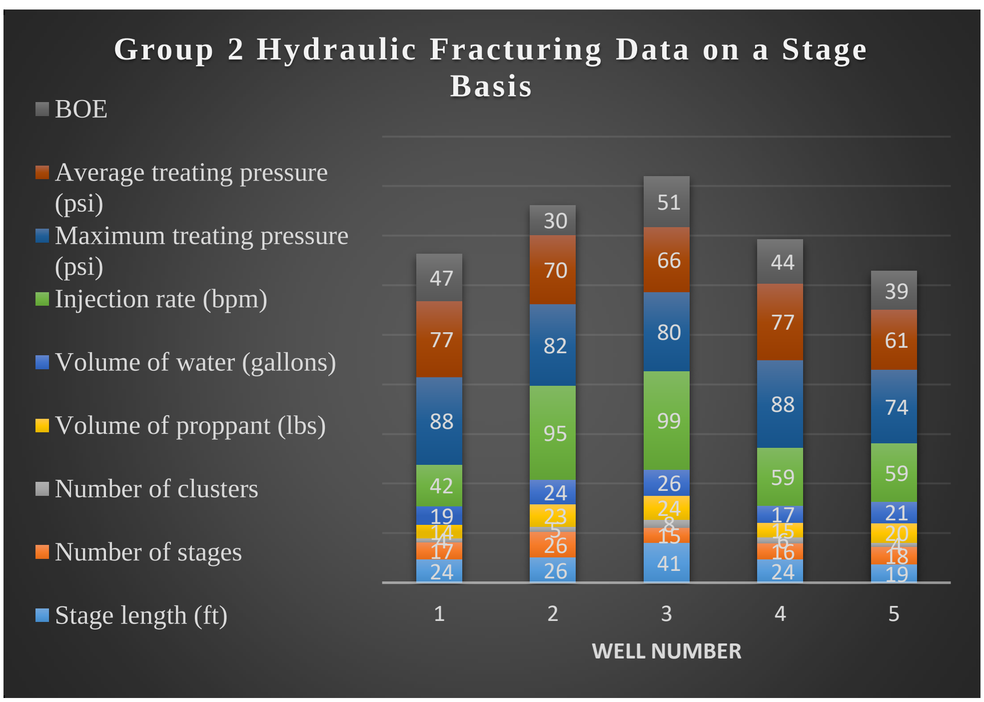


Figure C.8: The Hydraulic Fracturing Parameters and Production Performance of a Third Set of Wells in Group 2

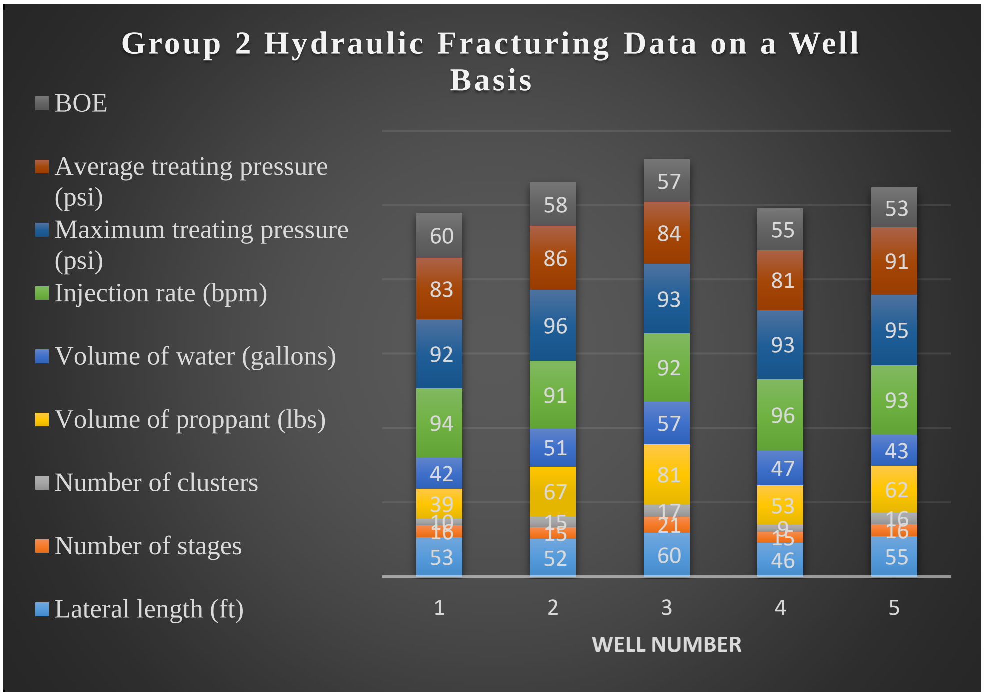


Figure C.9: The Hydraulic Fracturing Parameters and Production Performance of a Fourth Set of Wells in Group 2

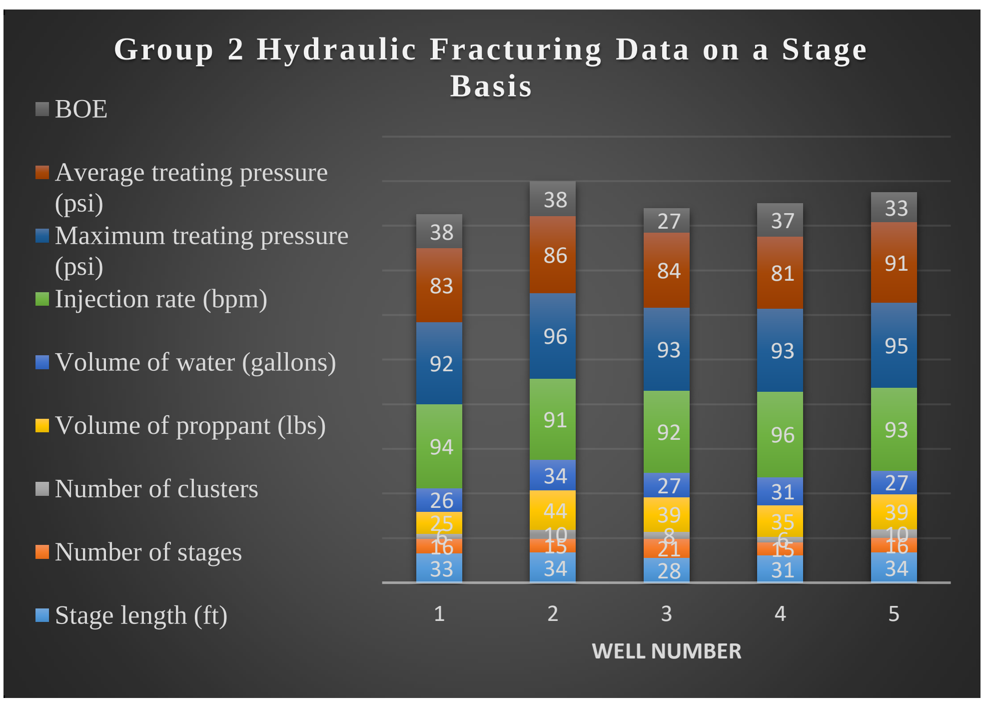


Figure C.10: The Hydraulic Fracturing Parameters and Production Performance of a Fourth Set of Wells in Group 2

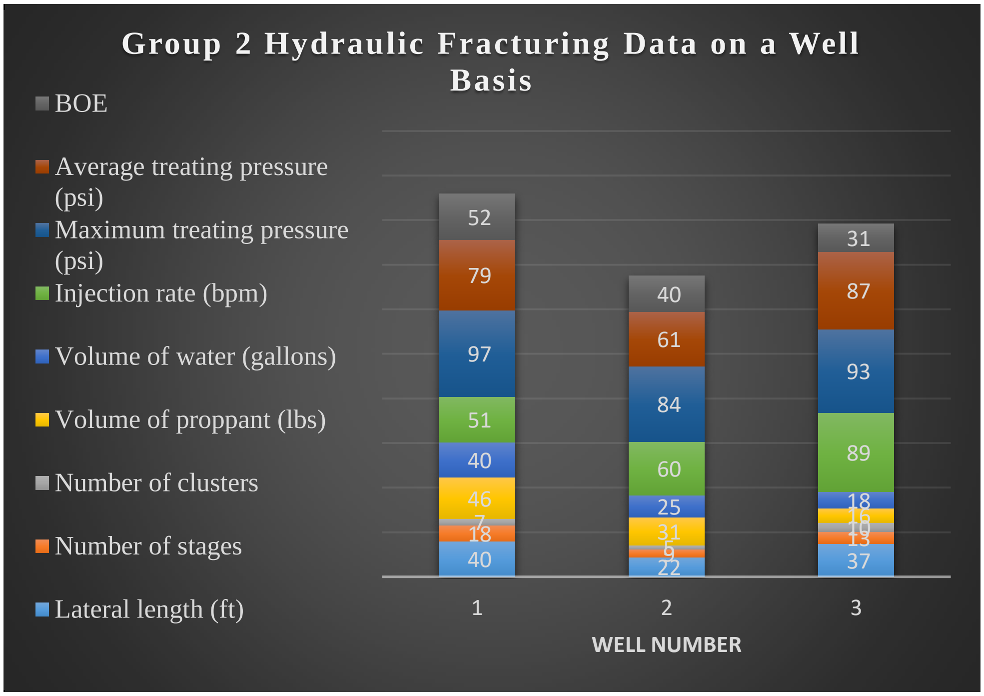


Figure C.11: The Hydraulic Fracturing Parameters and Production Performance of a Fifth Set of Wells in Group 2

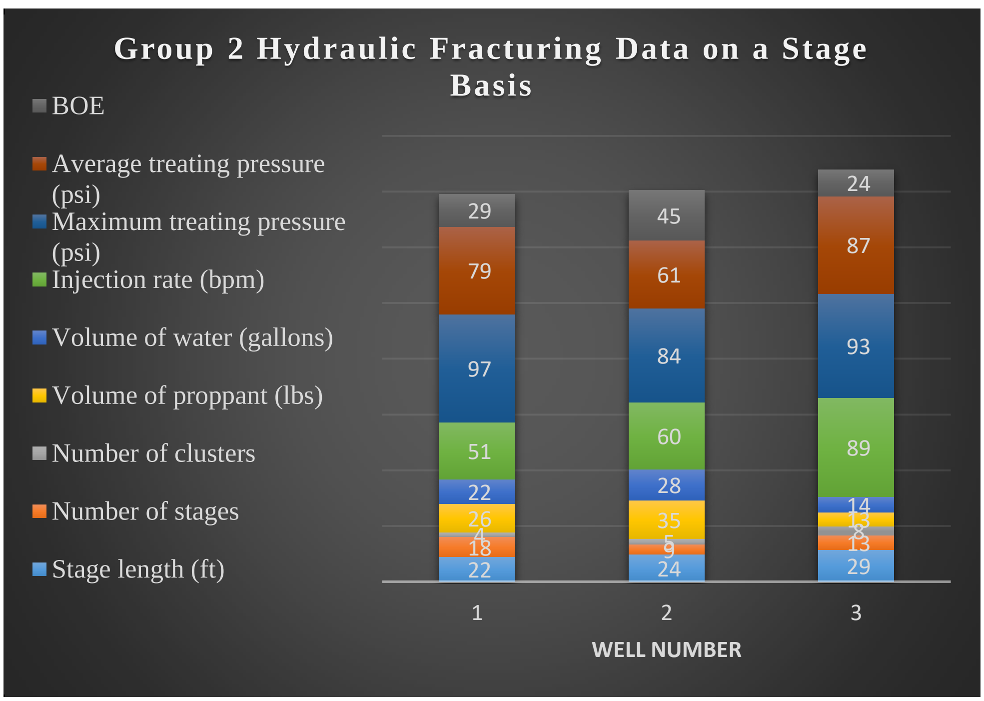


Figure C.12: The Hydraulic Fracturing Parameters and Production Performance of a Fifth Set of Wells in Group 2

Group 3:

The histogram plots of the hydraulic fracturing parameters and production performance on both a well basis and stage basis for the rest of the other wells in Group 3 are shown in Figures C.13, C.14, C.15, C.16, C.17, C.18, C.19, and C.20. The observations mentioned in Section 5.2 for Group 3 wells apply to the wells shown in Figures C.13, C.14, C.15, C.16, C.17, C.18, C.19, and C.20.

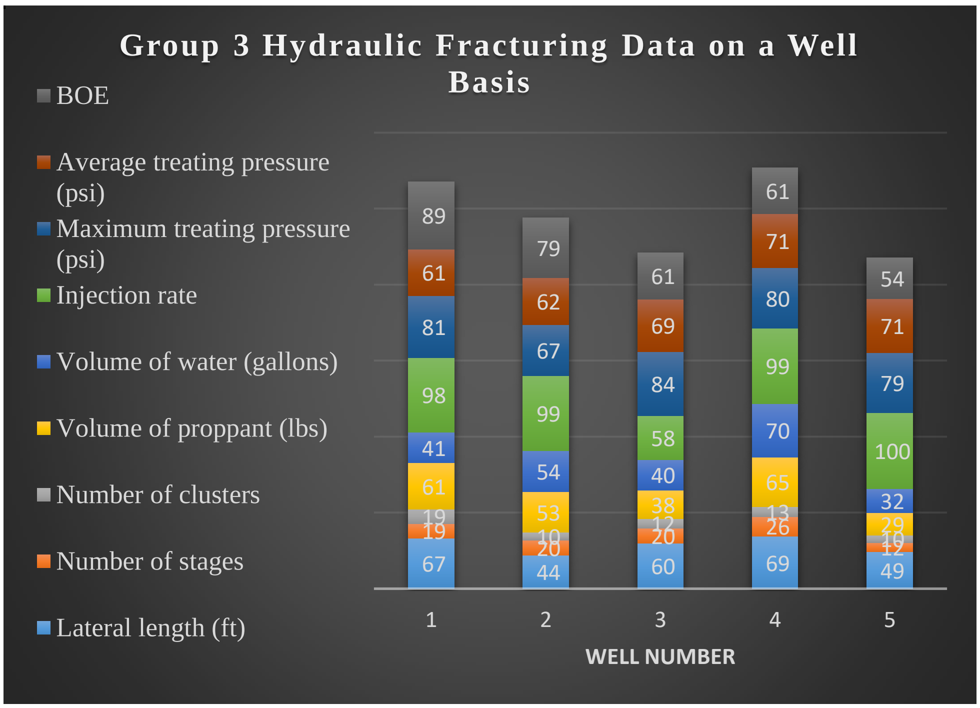


Figure C.13: The Hydraulic Fracturing Parameters and Production Performance of a Second Set of Wells in Group 3

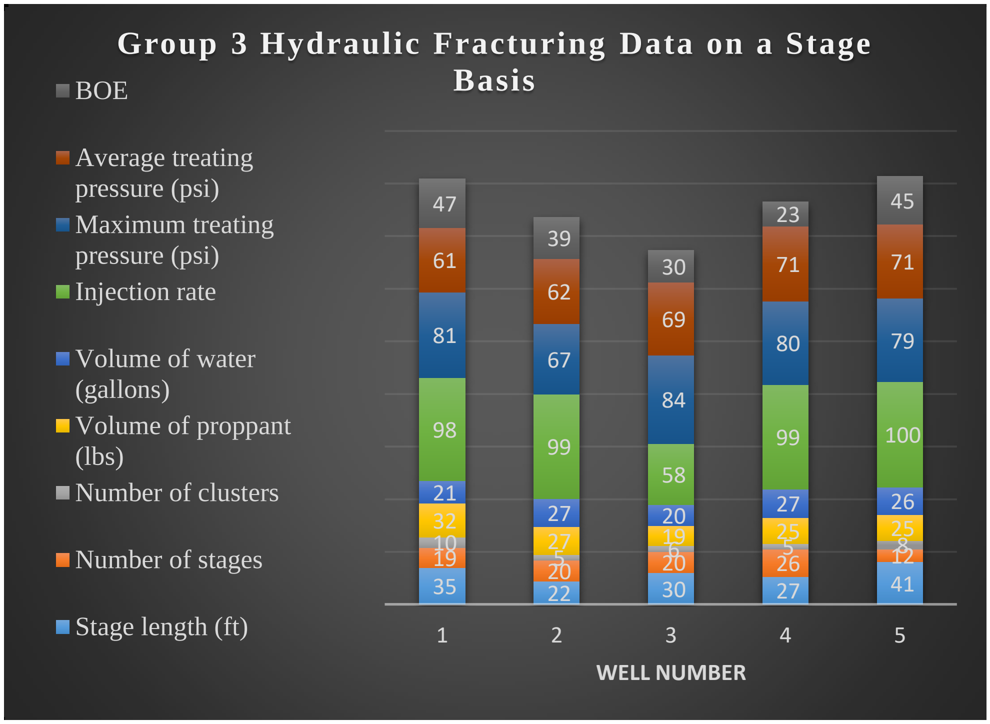


Figure C.14: The Hydraulic Fracturing Parameters and Production Performance of a Second Set of Wells in Group 3

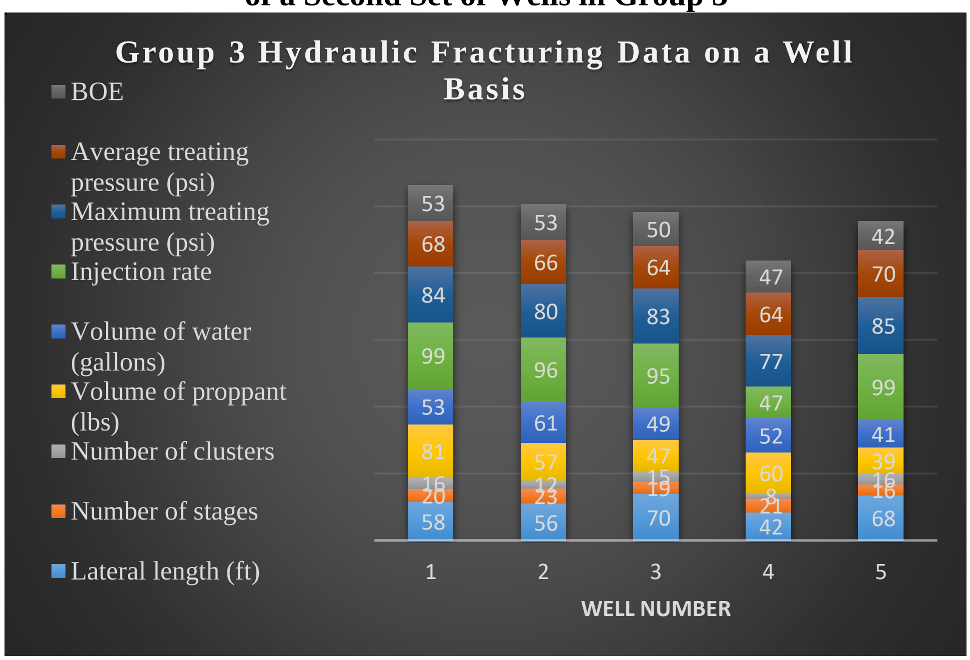


Figure C.15: The Hydraulic Fracturing Parameters and Production Performance of a Third Set of Wells in Group 3

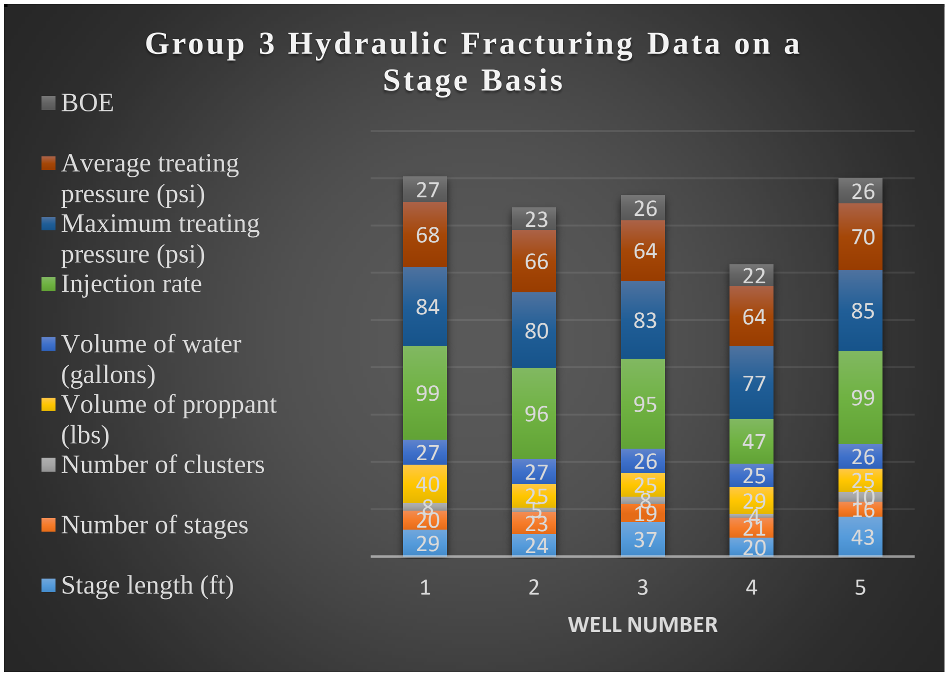


Figure C.16: The Hydraulic Fracturing Parameters and Production Performance of a Third Set of Wells in Group 3

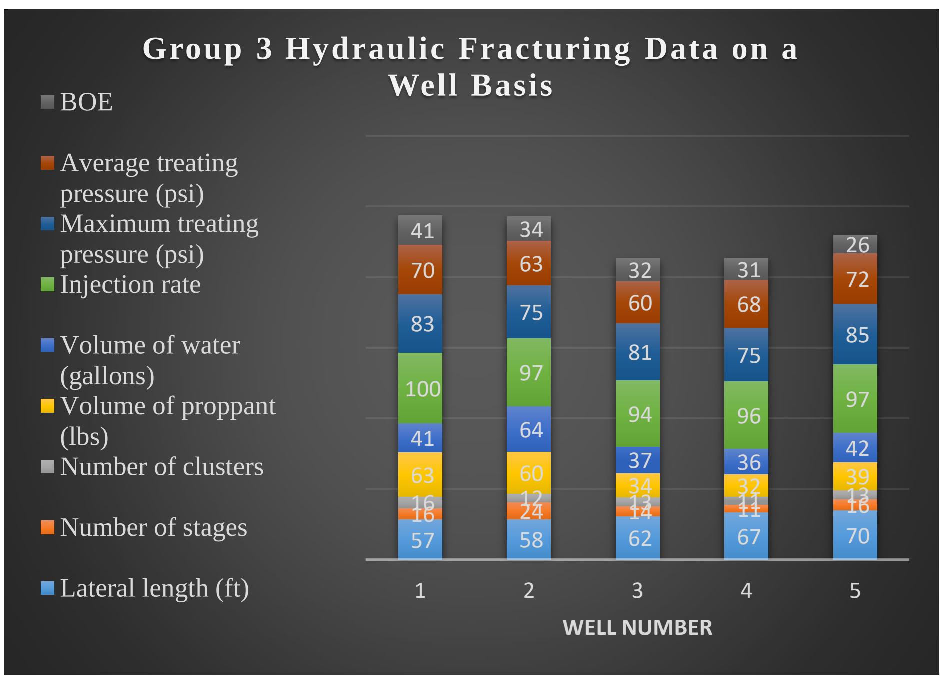


Figure C.17: The Hydraulic Fracturing Parameters and Production Performance of a Fourth Set of Wells in Group 3

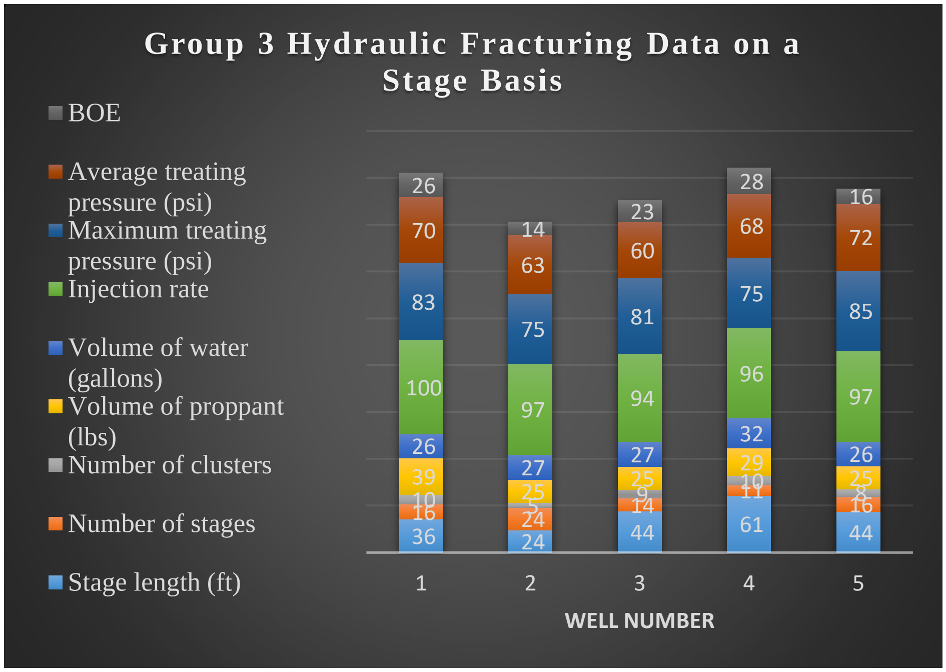


Figure C.18: The Hydraulic Fracturing Parameters and Production Performance of a Fourth Set of Wells in Group 3

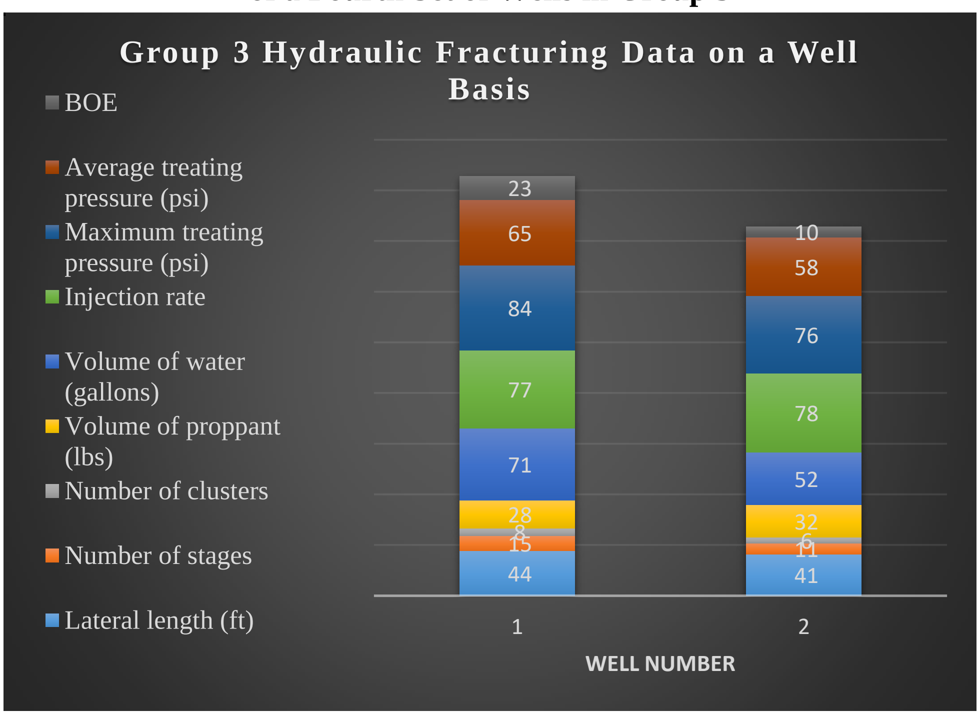


Figure C.19: The Hydraulic Fracturing Parameters and Production Performance of a Fifth Set of Wells in Group 3

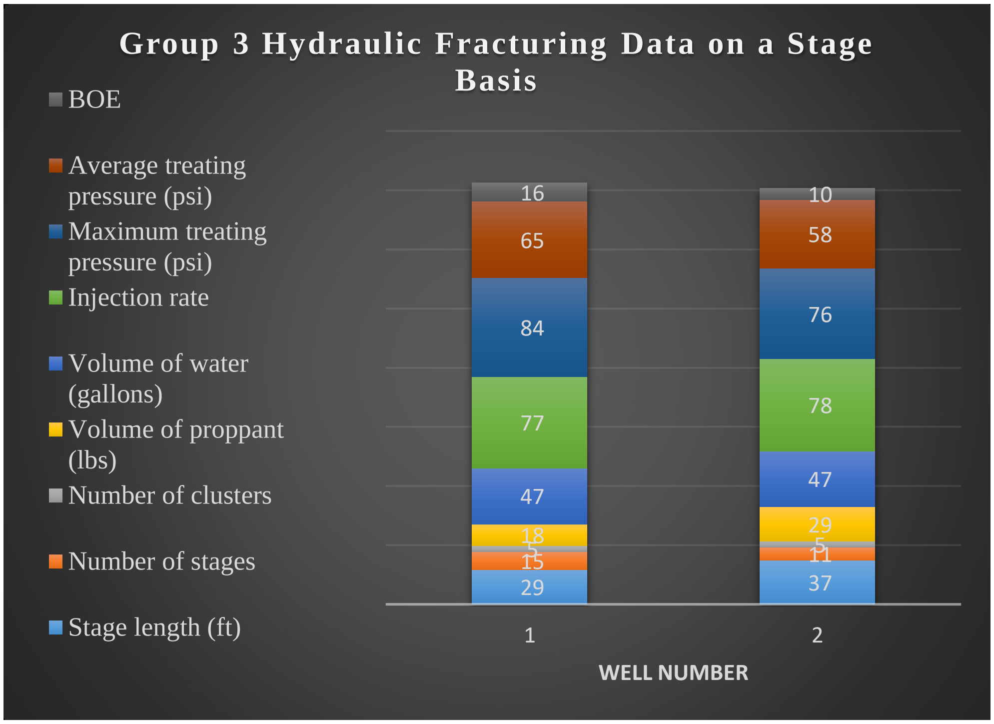


Figure C.20: The Hydraulic Fracturing Parameters and Production Performance of a Fifth Set of Wells in Group 3

Appendix D: R Code of the Four Regression Techniques

The r code used to program the four different techniques of regression analysis is shown in Appendix D below. These techniques are used and run on the data in Chapter 4 to answer the primary research question. After these techniques are run, the best performing technique is selected which is the NNs method and is run on the rest of the data as shown in Chapter 4.

```
#Linear Regression
gr2<-ddply(gr2,
            .(pli,nos,nocps,vowpw,irps,mtppw,atppw,voppw),
                function(x)c(BOE=mean(x$BOE)))
Object<- traincontrol (method= "repeatedcv ", repeats= 1, number =10)
set.seed(955)
linear <- train(BOE ~ (.)^2 + I(pli^2)+
            I(nos^2) + I(nocps^2)+
           I(irps^2) + I(vowpw^2)+
           I(mtppw^2)+
           I(atppw^2) + I(voppw^2),data = gr2,
           method = "lm",
           trControl = Object)

linear


#Decision Tree
set.seed(955)
rpart <- train(BOE ~ .,
```

```
                    data = gr2,
                    method = "rpart",
                    tuneLength = 30,
                    trControl = Object)

rpart
#SVM Technique
set.seed(955)
svm<- train(BOE ~ ., data = gr2,
                    method = "svmRadial",
                    tuneLength = 15,
                    preProc = c("center", "scale"),
                    trControl = Object)



svm

#NNs technique
tGrid <- expand.grid(.decay = c(0.001, .01, .1),
                     .size = seq(1, 27, by = 2))
NNs<- train(BOE ~ .,
                    data = gr2,
                    method = "nnet",
                    tuneGrid = tGrid,
                    preProc = c("center", "scale"),
                    linout = TRUE,
                    trace = FALSE,
```

```
        maxit = 1000,
        trControl = Object)
NNs
garson(NNs)
```