June 2019

Field Drilling Data Cleaning and Preparation for Data Analytics Applications

Daniel Cardoso Braga

FIELD DRILLING DATA CLEANING AND PREPARATION FOR DATA ANALYTICS APPLICATIONS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science in Petroleum Engineering

in

The Craft & Hawkins Department of Petroleum Engineering

by
Daniel Cardoso Braga
B.S., University of Campinas, 2014
August 2019

Acknowledgments

The achievements obtained in this work are the result of effort from multiple people, to whom I would like to express my gratitude:

To Dr. Mauricio Almeida for all the support in my endeavor to come to LSU and pursue my master's degree, without your help none of this would have happened. Thank you for your advice in and outside the classroom. I would also like to thank Dr Wesley Williams for accepting me as a student, and as a friend. Your sense of competition and high achievement motivated me to also always push the extra mile when things were difficult. I could not have asked for a better pair of professors to advise me in this transition of my career, with the perfect balance of experiences to help me achieve my goals.

To Tati, I would like to thank the non-stop support as my partner. The decision to move to a different country, leaving family and friends behind to pursue this dream would be much more difficult without you. Thank you for teaching me how to be a better person, and to trust my ”gut”. The award received from this work was mainly because you motivated me to go there and show my work.

To Branum, Stanley and Mark from Legacy Directional Drilling I thank the thrust deposited on me since the beginning. The plan was ambitious due to the short time frame, but with your guidance we could meet all project goals. Special thanks to Branum for his dedication and passion about extracting knowledge from data. Your help was crucial to guide me to the right way.

To all professors and employees of the Craft & Hawkins Department of Petroleum Engineering I would also like to thank for supporting us, the students, making this institution loved in the intense way it is. Thank you Dr. Wojtanowicz for giving me a chance when accepting me in this program, and for the long talks and experience shared during my time here. Also I would like to thank Dr. Gupta for being such an amazing person and professor - Your dedication is contagious, and makes us all proud of having you as a professor. To all employees from the department - Andy, Janet, Tanesha, Jeanette and Fenelon - I express my sincere gratitude for your dedication in making the department better every day.

To my family and friends from both Brazil and Baton Rouge I would like to thank for their company when the mind was not able to work anymore. I had no idea how difficult it would be to live so far from all of you, but at the same time this experience allowed us to make many new friends, so at the end we always become stronger.

Table of Contents

List of Tables

List of Figures

List of Nomenclature

3D - Three Dimensional

AFE - Authorization for Expenditure

AI - Artificial Intelligence

API - American Petroleum Institute

BHA - Bottom Hole Assembly

BOP - Blow-out Preventer

CSV - Comma Separated Values

DD - Directional Driller

DNN - Deep Neural Network

DoC - Depth of Cut

DWOB - Downhole Weight on Bit

EDR - Electronic Drilling Recorder

HHP - Hydraulic Horse Power

HL - Hook Load

HMI - Human-Machine Interface

HSI - Hydraulic Horse Power per Square Inch

IOT - Internet of Things

IQR - Interquartile Range

IT - Information Technology

KDD - Knowledge Discovery in Databases

KOP - Kick-off Point

KPI - Key Performance Indicator

LP - Landing Point

MD - Measured Depth

ML - Machine Learning

MSE - Mechanical Specific Energy

MTM - Methods-Time Measurements

MWD - Measurement While Drilling

OD - Outer Diameter

OGDQ - Operators Group for Data Quality

OPEX - Operational Expenditure

OPM - Open Porous Media

PC - Personal Computer

PDF - Probability Density Function

RMSE - Root Mean Squared Error

ROI - Return on Investment

ROP - Rate of Penetration

RPG - Rotations per Gallon

RPM - Rotations per Minute

RTOC - Real Time Operation Center

SPM - Stroke per Minute

SW - String Weight

TSE - Torsional Severity Estimate

WITS - Wellsite Information Transfer Specification

WITSML - Wellsite Information Transfer Standard Markup Language

WOB - Weight on Bit

Abstract

Throughout the history of oil well drilling, service providers have been continuously striving to improve performance and reduce total drilling costs to operating companies. Despite constant improvement in tools, products, and processes, data science has not played a large part in oil well drilling. With the implementation of data science in the energy sector, companies have come to see significant value in efficiently processing the massive amounts of data produced by the multitude of internet of thing (IOT) sensors at the rig. The scope of this project is to combine academia and industry experience to analyze data from 13 different wells drilled in an area of 2 x 4 miles. The data was collected in the same rig and contains over 12 million electronic drilling recorder data points, driller's activity logs and well profiles. The main focus is to propose a detailed workflow to clean and process real drilling data. Once cleaned, the data can be fed into data analytics platforms and machine learning models to efficiently analyze trends and plan future well more efficiently. This roadmap will serve as a basis for drilling optimization. The objective of this work is to detail the various steps needed to prepare field drilling data for business analysis, as well discuss about data analytics and machine learning application in drilling operations. The results to be presented are the detailed workflow and description of the data preparation steps, an example analysis of the drilling data and an example application of a machine learning model in drilling.

Chapter 1
Introduction

1.1. Petroleum Engineering Overview

The exploration for oil and gas is a very complex process that comprises a multitude of skilled professionals, each of them working in their area of expertise to bring petroleum fluids from the subsurface of the earth to the production facility. Once refined, petroleum products are used in multiple forms such as fuel, petrochemicals, additives to agricultural products, clothing, cosmetics, plastics and polymers, just to name a few. From the time an oil company bids for an exploration lease, to when it removes all equipment used to produce the oil and returns the area to the government, engineering work is present in all steps. An overview of the processes performed in the search, discovery and extraction of petroleum from the ground is depicted in Fig. 1.1.

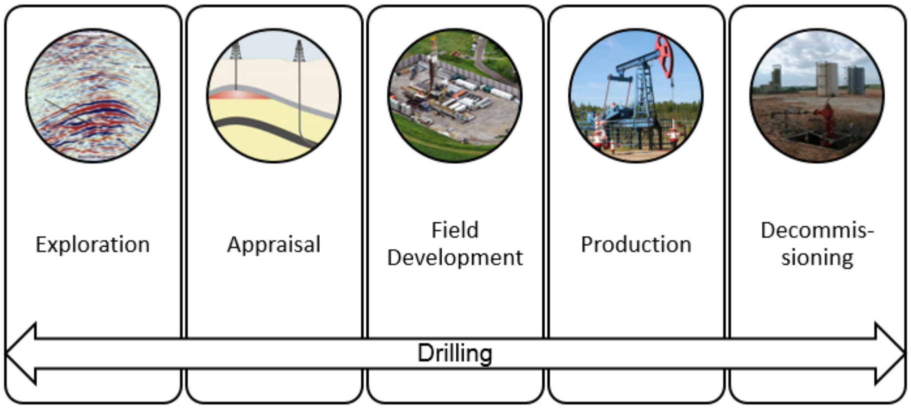


Figure 1.1. Oil and Gas Life Cycle

The typical oil and gas life cycle comprises five main phases:

1. Exploration: Oil companies, the so-called Operators, purchase land where it is believed to have oil or gas in the subsurface to be explored. Usually appraisal seismic is done by the government natural resources agency, in order to provide initial information about the lithology and geology of the area beneath the lease.

   Seismic is the technique used by the geologists in the attempt of describing the layers of rocks and sediments along the Earth's crust. With the creation of sound waves generated by explosives and then calculating the travel time taken for the waves to come back, the analyst can infer how the layers are organized. From a petroleum exploration perspective, it is very important to identify geological features like faults and sealing layers that could facilitate the accumulation of hydrocarbon fluids. A schematic of the seismic shooting process as well as an example of seismic interpretation is shown in Fig. 1.2.

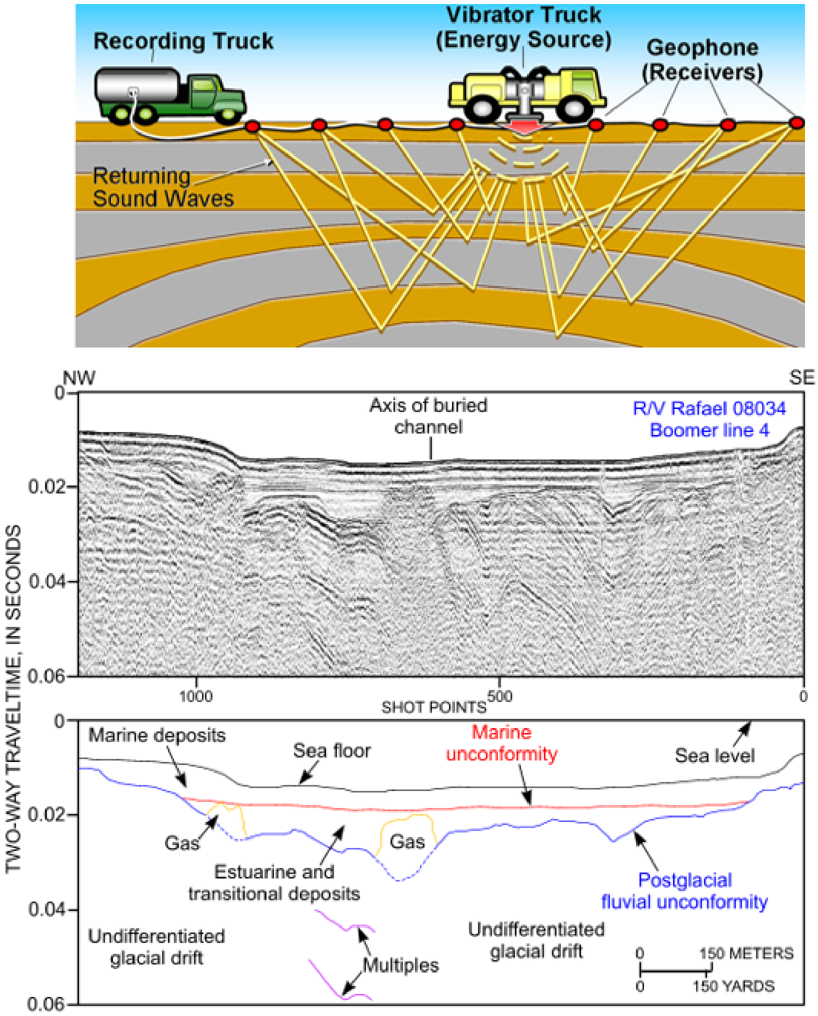


Figure 1.2. Seismic Shooting Schematic (top) – Example of Seismic Interpretation (bottom)

Petroleum geologists and petrophysicists analyze together this information and decide where to drill the first exploratory well in the area, with the objective to confirm if the target formation contains petroleum or not.

2. Appraisal: With the confirmation of oil in the first exploratory well, then the next step is to drill additional wells to evaluate the size and quality of the reservoir. The objectives of the drilling campaign now are to understand the limits of the reservoir in respect to faults or other geological features, as well as performing well tests to collect more data from the reservoir (permeability, production rates, pressures, fluid samples, etc).

3. Field Development: In this phase the field development plan is executed. This plan covers the production strategy for the whole life of the field - from 10 to 30 years depending on the size of the accumulation. Here all the wells - producers and injectors - needed are drilled, cased and completed to allow the maximum oil extraction possible (oil recovery).

4. Production: Once the wells are drilled and the production equipment that will receive, store and separate the fluids produced are installed, the production phase begins. This is denoted as the first oil of the field. As described previously, this phase is the longest of the field life cycle and can last up to three decades. Because the production can last many years, programmed stops are done to maintain the downhole and wellhead equipment.

5. Decommissioning: Finally, once it is not economical to maintain production, the operating company will decide to shut down the production facility and begins the decommissioning phase. Common reasons for the operating costs (OPEX) becoming prohibitive are the increasing injection costs to maintain reservoir pressure or the volume of water produced together with the oil reaching the surface equipment limits.

   The decommissioning phase is normally highly regulated by the national petroleum agencies, which require operators to abandon the well safely. This process usually requires the setting of cement plugs in front of producing formations and near to the surface, and the removal of the upper part of all casing strings and surface equipment.

1.2. Motivation

In the schematic shown in Fig. 1.1, the arrow below shows where drilling-related operations happen along the life of a field. From Exploration to Decommissioning, the drilling department of the operating company is involved in all steps: Drilling exploratory wells at the exploration phase; drilling delimiting wells and helping execute well tests during the appraisal; drilling the producers and injectors during the development; performing workover operations during the producing period, and finally, plugging and abandoning the well in the decommissioning phase. This description makes clear how important drilling operations are to the whole field development process. Because drilling oil wells is a very expensive activity, a unique network of service providers and contractors has been developed in the last century just to provide these services to the operating companies. These drilling service companies have multimillion-dollar budgets for research and development to employ the best technology available to the field. Some examples of such innovative technologies are the rotary steerable systems, the wired drill pipe and hybrid drill bits.

Closely linked to technology advances is automation. The drilling rig of today contains a massive amount of instrumentation that collects parameters from almost every equipment installed in the drilling rig, whether it is related to the drilling operation or not. For example, the rig control system receives multiple data streams from the Top Drive that are used to execute drilling: its position in respect to the derrick, temperature of the electric motors, drill string rotation speed, oil temperature, drill pipe elevator opening status, etc. Also other non-critical equipment have sensors measuring their state, like drilling fluid agitators, or access baskets to lift personnel, allowing its remote and safe operation. Fig. 1.3 shows an example of a modern Top Drive, with multiple sensors installed.

With the evolution of the technology applied to the drilling rig, companies started to store the data generated by these hundreds of sensors in their databases. The data usually is recorded at very high frequency (1, 5 or 10Hz), and in most cases is standardized in the Wellsite Information Transfer Standard Markup Language (WITSML), a modernization of

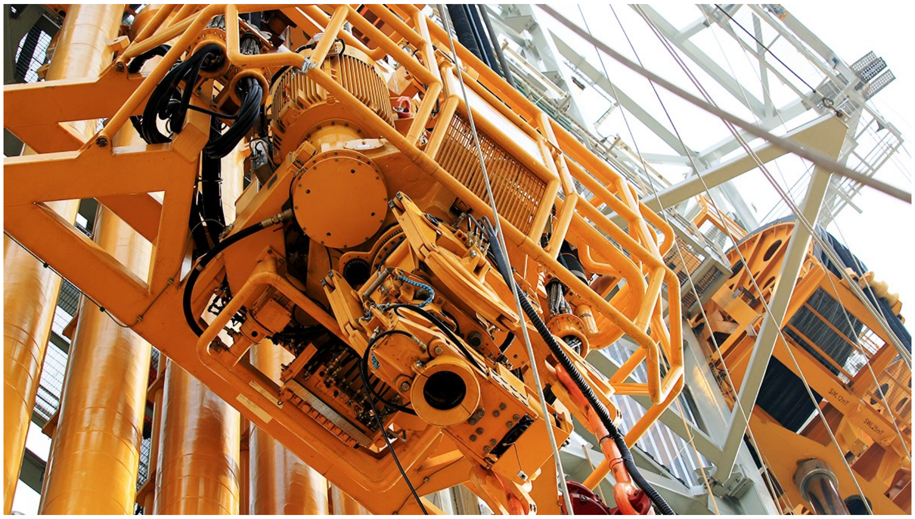


Figure 1.3. National Oilwell Varco Top Drive HPS 750

the original WITS protocol, created by an Industry consortium formed by many companies, and managed by Energistics [Energistics, 2019].

Today, many companies stream data from the sensor to the corporate office in real-time, plotted in colorful graphs on multiple screens. Fig. 1.4 shows the whole IT network that must be installed to make live data transmission possible. The data is transmitted along several means that can be located hundreds of miles apart to reach the database. Although it is very appealing to have this feature, if said company is not using it to improve how they operate, the tremendous effort behind the real-time data transmission is nothing more than a very high expense providing no return on investment (ROI) [Damski, 2014].

Putting together the advances of computational power, the exponential growth of data that is generated in every well drilled, and the strong incentive of cost reduction due do the drop of oil price in recent years, it is easy to understand why companies that own this huge amount of data wants to extract sense from it: Business managers and decision makers want to have insights from the operation they perform on a daily basis, in an attempt to improve future results and remain competitive in this very tight market. Consequently, this thesis presents a study on how to organize the raw drilling data, clean it, pre-process it and present it in a way that will enable the drilling engineer to make better decisions about the planning of a next well to be drilled.

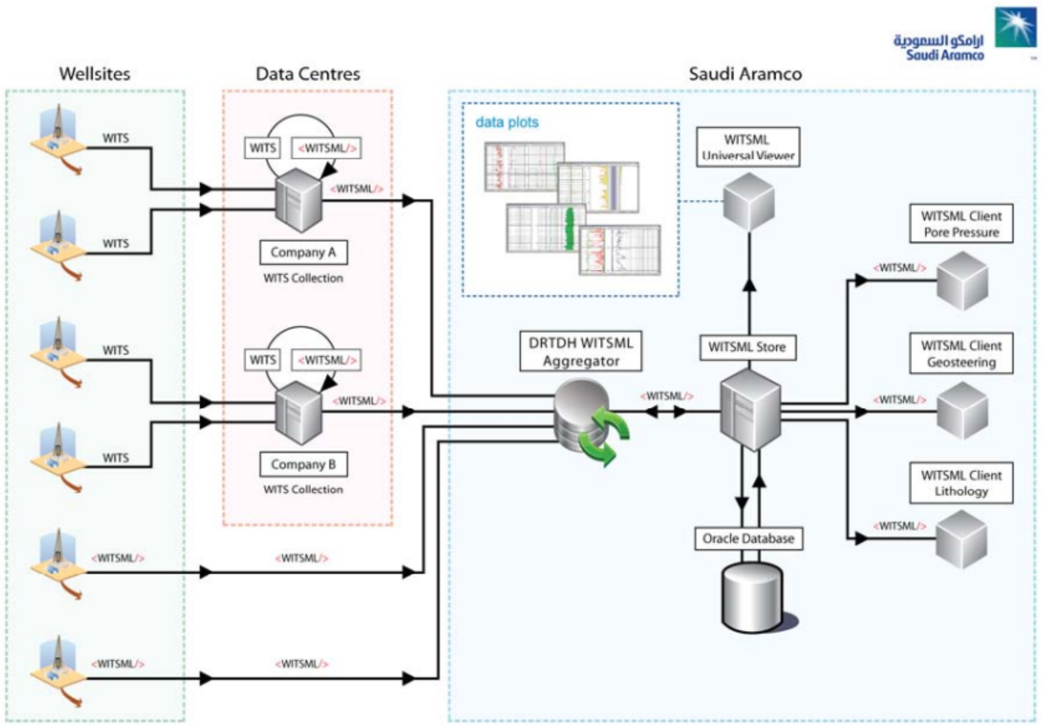


Figure 1.4. Current Real-Time Wellsite Data Flow at Saudi Aramco, From [Al-Khudiri et al., 2008]

1.3. Objectives

When dealing with massive amounts of data (a table with more than 30 columns and 1,000,000 rows for example), the analyst id responsible to extract meaning from the raw data should apply some tasks in a logical order that will help them to understand the patterns and relationship between variables. A very good approach to this process is called “The Knowledge Discovery in Databases” (KDD) and was proposed by [Fayyad et al., 1996]. Here, Fayyad proposes the KDD as a five-step process that transforms raw data into knowledge: Data selection, Preprocessing, Transformation, Data Mining and Interpretation/Evaluation. The KDD schematic is presented in Fig. 1.5.

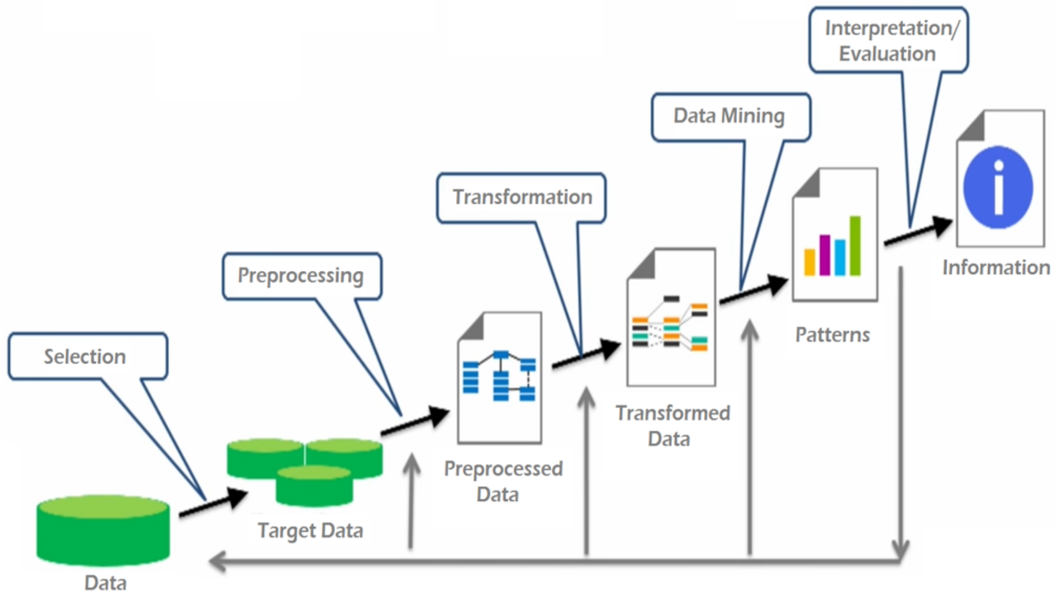


Figure 1.5. The KDD Process, Adapted From [Fayyad et al., 1996]

Having the KDD process as the base for the workflow applied to the drilling data, the objectives of the present work are:

1. Elaborate a detailed workflow of the data selection, preprocessing and transformation processes applied to the raw data collected from the Electronic Drilling Recorder (EDR).

2. Describe data analytics techniques applied to analyze and understand the drilling data.

3. Create a statistics-based (machine learning) model that will predict actual drilling performance based on previous data.

1.4. Background

1.4.1. The Well

An oil well is the physical conduit between the reservoir and surface. It is formed of several concentric steel tubulars – casing – that are drilled in a sequential order. Typically, a casing configuration goes from the largest - 30 inches of outer diameter (OD) to the smaller - 9 $\frac{3}{4}$ inches of OD. Once the target formation is reached, a production tubing – a steel pipe with 5-7 inches OD - is run until the bottom of the well and hung on top of the well head on a tubing hanger. A well schematic is presented in Fig. 1.6.

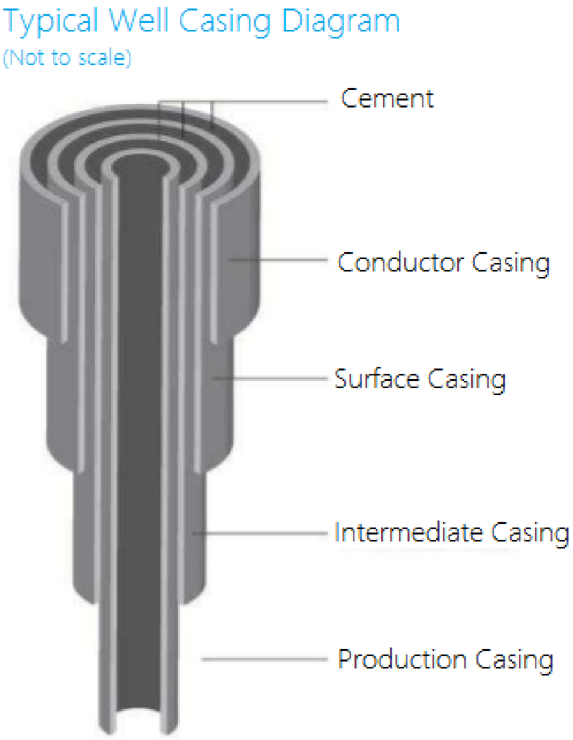


Figure 1.6. Typical Well Casing Diagram

1.4.2. Making Hole

When drilling a hole in the ground, the main objective is to break the formation, and transport the cutting material to the surface, as the well is being drilled. This is accomplished using a drill bit that rotates clockwise against the formation. The bit has teeth that are indented into the formation with applied force, known as Weight on Bit (WOB). With the indentation, usually referred to as Depth of Cut (DoC) and rotation of the bit (RPM), the drilling process occurs. The resulting performance of drilling is described as the rate of penetration (ROP), measured in feet per hour. This process is described in Figure 1.7. It can be seen from the diagram that the volume of rock cut at any revolution is simply the length of the total indentation caused by the WOB multiplied by the bit area, on every rotation.

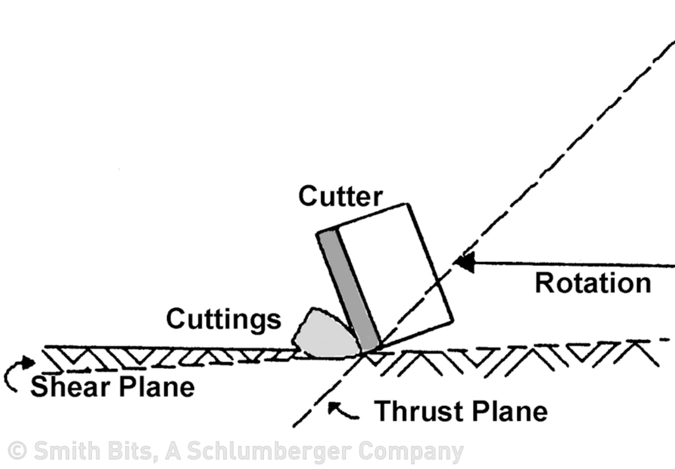


Figure 1.7. Single Cutter Representation of the Rock Cutting Mechanism

1.4.3. Drill String

Drill string is the name given to several equipment/components that are screwed together sequentially as drilling progresses. A typical drill string configuration comprises a drill bit at the tip, the Bottom Hole Assembly (BHA) that contains the main equipment that will enable the drill bit to drill in the planned direction and downhole data acquisition systems and the drill collar/drill pipe sections. Drill pipes are steel tubulars that are threaded on both ends. Its main purpose is to connect the drill bit to the top drive, and to transmit fluid from surface to the bit.

1.4.4. Measurement While Drilling Tool

The Measurement While Drilling (MWD) tool is a set of sensors that are run in to the well, just behind the drill bit (Fig 1.8). These sensors record information such as magnetic orientation, lateral and axial shock, lateral and axial vibration, temperature and acceleration. These metrics are used to tune the directional control of the bit by calculating the bit position over time. The data is transmitted to the surface via 10 Hz mud pulse telemetry, and a full log is downloaded once the tool and the whole BHA is pulled out of hole.

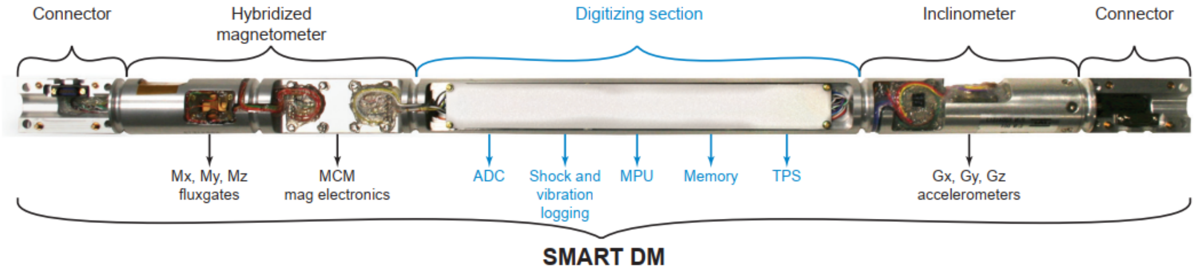


Figure 1.8. Typical MWD Assembly, From JAE Smart DM 2016 Catalog

1.4.5. Mechanical Specific Energy Concept

The concept of mechanical specific energy (MSE) was derived by [Teale, 1965]. In his work he derived a relationship between the rock strength and the energy required to destroy the rock. To prove his theory, a laboratory test was performed and Taele realized that the value of the MSE was equal to the rock compressive strength. The expression derived by Teal is:

$$MSE = \frac{Input\,Energy}{Output\,ROP} = \frac{480 \times Torque \times RPM}{ROP \times Hole\,Size^2} + \frac{4 \times WOB}{\pi \times Hole\,Size^2} \tag{1.1}$$

However, the application of MSE in drilling surveillance had only become widespread after [Dupriest et al., 2005] published the gains in drilling performance when drillers used MSE as a direct indicator of efficient drilling.

1.4.6. Electronic Drilling Recorder

The Electronic Drilling Recorder (EDR) is the computer systems that gathers information of all sensors that are related to the drilling operation (Fig. 1.9). Examples of equipment that are connected to the EDR are: Top Drive, Retractable Dolly, Drawworks, Mud Pumps, Blow-Out Preventer (BOP), Diverter, Power Slips, Pipe Handling Crane and Mud Bucket. The EDR system usually provides information about each machine that is connected to it via a Human-Machine Interface (HMI) system. With this system, the driller can easily operate and check the status of all equipment on the drill floor.

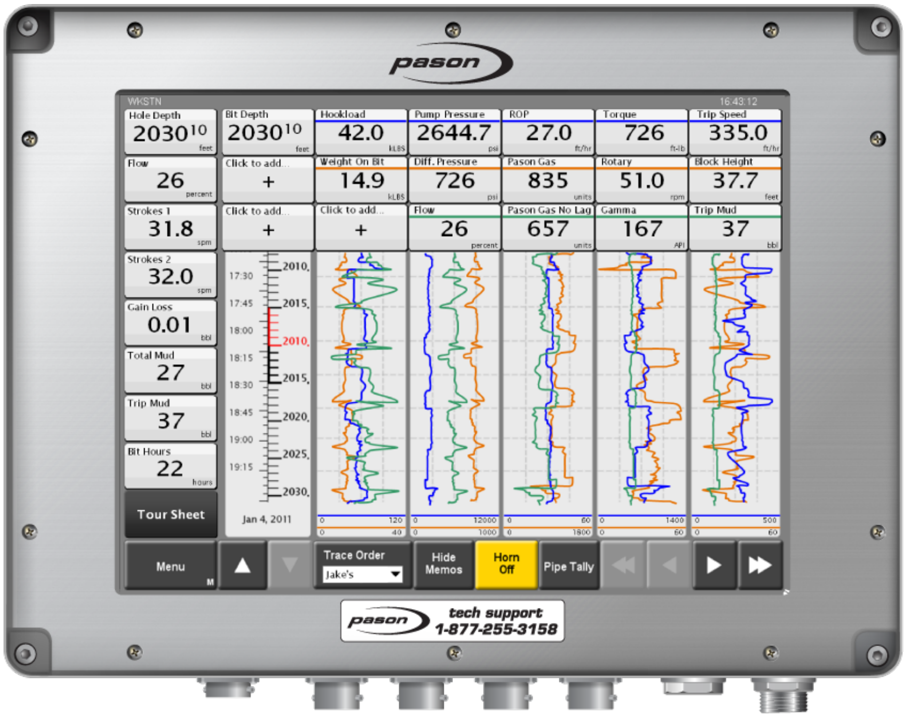


Figure 1.9. Example of an EDR Screen, From Pason Systems

1.5. Data Formats

1.5.1. EDR File

The EDR file is the download of the recorded values during a selected time interval. The data frequency can be set to display the values each 1, 5 or 10 seconds, being the last two formats the average during the selected time interval. Thus, the total number of rows depends on the time interval selected to be aggregated in the file, and the time frequency of the recording. A simple estimation can be:

$$60\, seconds \times 60\, minutes \times 24\, hours = 86,400\, rows\, (data\, points)\, per\, day$$

and, assuming that a well takes 60 days to be drilled, for a 5 seconds recording frequency we have:

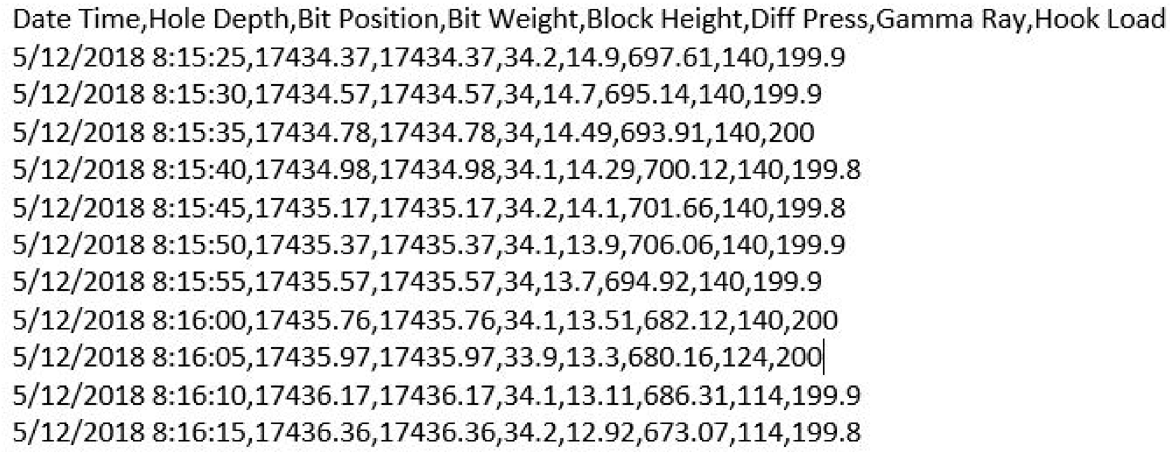
$$\frac{86,400\, rows}{5\, seconds} \times 60\, days = 1,036,800\, rows\, per\, well$$

Multiplying this number of rows by each drilling parameter that you want to analyze, the EDR file can easily be of a size that is impossible for a human to understand in a logical and meaningful way. It is therefore imperative that one use computational aid when processing such large files.

Common features (columns) that usually are analyzed in an EDR file are: Date Time (Timestamp), Hole Depth, Bit Position, Top Drive Rotation, Top Drive Torque, Pump Pressure, Weight on Bit and Block Position. However, most commercial EDR systems enable the inclusion of more than 200 readings from the various equipment installed, although this availability might depend if the sensors are in fact installed, which often must be included in the contract with the service company, among other factors.

A sample of an EDR file download is presented below. A common format for download is the comma-separated-values (CSV) format. This format describes the data in text format, being parsed into columns by a comma ",". The parsing software then understands this pattern and the data can be read as a table.

Example of a raw EDR file in CSV format:



```
Date Time,Hole Depth,Bit Position,Bit Weight,Block Height,Diff Press,Gamma Ray,Hook Load
5/12/2018 8:15:25,17434.37,17434.37,34.2,14.9,697.61,140,199.9
5/12/2018 8:15:30,17434.57,17434.57,34,14.7,695.14,140,199.9
5/12/2018 8:15:35,17434.78,17434.78,34,14.49,693.91,140,200
5/12/2018 8:15:40,17434.98,17434.98,34.1,14.29,700.12,140,199.8
5/12/2018 8:15:45,17435.17,17435.17,34.2,14.1,701.66,140,199.8
5/12/2018 8:15:50,17435.37,17435.37,34.1,13.9,706.06,140,199.9
5/12/2018 8:15:55,17435.57,17435.57,34,13.7,694.92,140,199.9
5/12/2018 8:16:00,17435.76,17435.76,34.1,13.51,682.12,140,200
5/12/2018 8:16:05,17435.97,17435.97,33.9,13.3,680.16,124,200
5/12/2018 8:16:10,17436.17,17436.17,34.1,13.11,686.31,114,199.9
5/12/2018 8:16:15,17436.36,17436.36,34.2,12.92,673.07,114,199.8
```

The same data now displayed in tabular form, Table 1.1:

Table 1.1. EDR Data in Tabular View

DATE TIME	HOLE DEPTH	BIT POSITION	BIT WEIGHT	BLOCK HEIGHT	DIFF PRESS	GAMMA RAY	HOOK LOAD
5/12/2018 8:15:25	17434.37	17434.37	34.2	14.90	697.61	140	199.90
5/12/2018 8:15:30	17434.57	17434.57	34.0	14.70	695.14	140	199.90
5/12/2018 8:15:35	17434.78	17434.78	34.0	14.49	693.91	140	200.00
5/12/2018 8:15:40	17434.98	17434.98	34.1	14.29	700.12	140	199.80
5/12/2018 8:15:45	17435.17	17435.17	34.2	14.10	701.66	140	199.80
5/12/2018 8:15:50	17435.37	17435.37	34.1	13.90	706.06	140	199.90
5/12/2018 8:15:55	17435.57	17435.57	34.0	13.70	694.92	140	199.90
5/12/2018 8:16:00	17435.76	17435.76	34.1	13.51	682.12	140	200.00
5/12/2018 8:16:05	17435.97	17435.97	33.9	13.30	680.16	124	200.00
5/12/2018 8:16:10	17436.17	17436.17	34.1	13.11	686.31	114	199.90
5/12/2018 8:16:15	17436.36	17436.36	34.2	12.92	673.07	114	199.80

1.5.2. Well Schematic

The well schematic, or well plan, is the document that describes the trajectory to be drilled. It should contain all the information needed for the correct execution of the well trajectory by the field personnel. Once the well is drilled, this document can be updated with the actual drilled path, so the planned vs actual trajectories can be compared to assess the well directional performance. Several factors must be taken into consideration for the trajectory design of a well:

- Surface location and target formation: The two basic inputs, first the designer has to understand from where the well is departing and to where it is going. This will determine the main profile of the well (J shape, S shape, Vertical, etc)

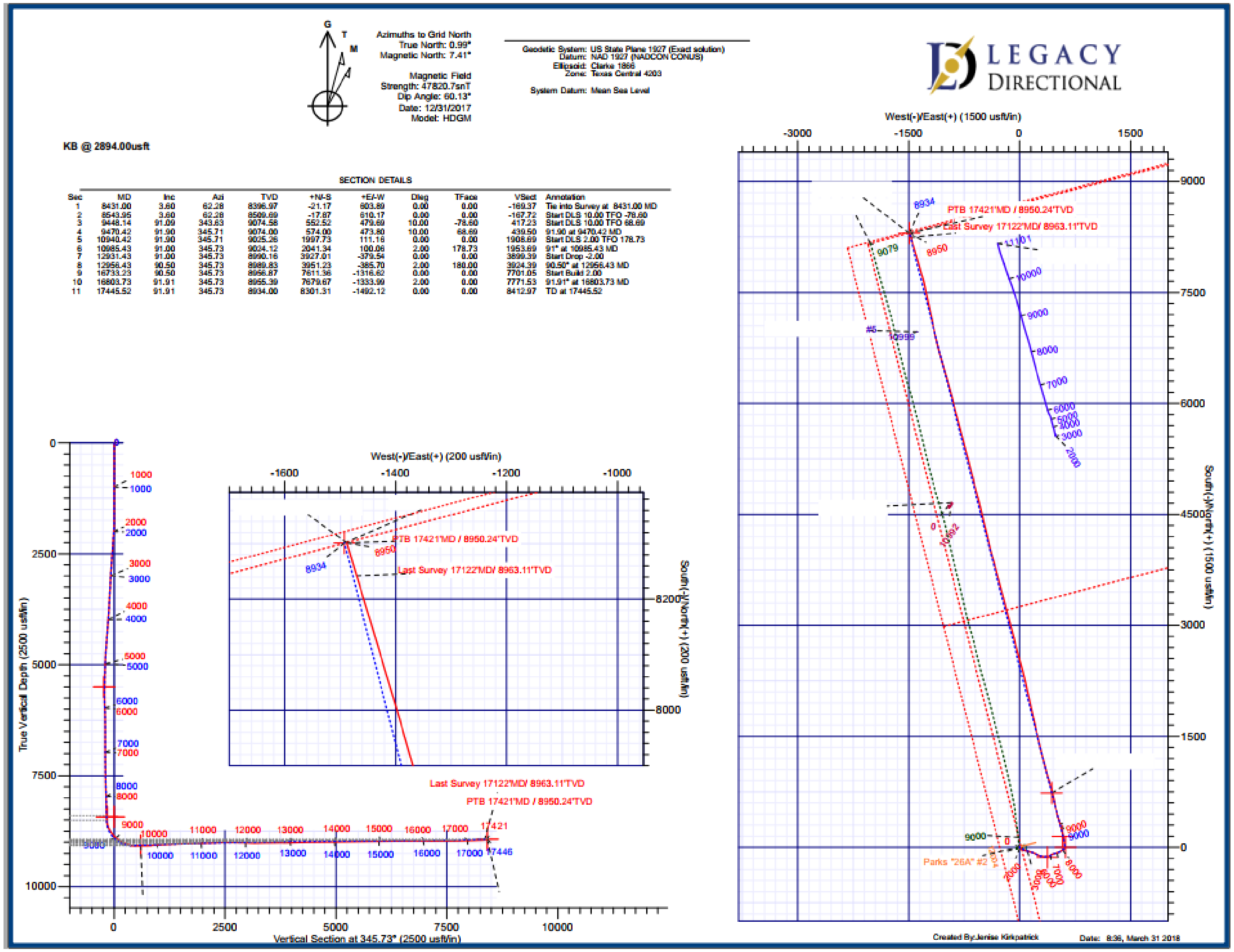
- Lease constrains: According to regulation, the limits of the leasing area apply also for the subsurface. That way, the well path must also respect these limits.

- Underground aquifer: In case of an existing aquifer along the well path, extra care has to be taken in the proper isolation of the aquifer to avoid contamination of the water body. The practice is usually to case and cement any phase that is crossing an aquifer to avoid this problem.

- Well collision: In very prolific areas, drilling activity is very intense. That way, the drilling companies have to take extra care in the directional control to avoid collision with surrounding wells. The consequences of hitting a producing well can be catastrophic. Companies operating in the same area usually share the information about their wells so the collision avoidance can be properly planned.

In the well schematic, the main geometric parameters are plotted in a lateral view (cross section of the vertical plane) and top view. A horizontal well that is typically divided into three main sections: vertical or nudge, curve and lateral or horizontal section. The vertical section is the first to be drilled, covering the shallower (thus softer) formations. At a certain point, the trajectory has to change from vertical to horizontal, so a curved section is drilled. The point at which the trajectory begins to deviate from vertical is called the kick-off-point (KOP). At the end of the curve section, the point at which the trajectory is now kept constant horizontally is called the landing point (LP). For a horizontal well as depicted in Fig. 1.10, the lateral profile can be described simply knowing the KOP and LP depths. However more information has to be given for a precise 3D description of the trajectory.

The well plan also contains georeferencing information to be used to set the directional tools such as the system datum, the ellipsoid used as reference, the direction of the true and magnetic north, the geodetic system used, etc.



LEGACY DIRECTIONAL

Azimuths to Grid North
True North: 0.99°
Magnetic North: 7.41°

Magnetic Field
Strength: 47820.7snT
Dip Angle: 60.13°
Date: 12/31/2017
Model: HDGM

Geodetic System: US State Plane 1927 (Exact solution)
Datum: NAD 1927 (NADCON CONUS)
Ellipsoid: Clarke 1866
Zone: Texas Central 4203

System Datum: Mean Sea Level

KB @ 2894.00usft

SECTION DETAILS

Sec	MD	Inc	Azi	TVD	+N/-S	+E/-W	Dleg	TFace	VSect	Annotation
1	8431.00	3.60	62.28	8396.97	-21.17	603.89	0.00	0.00	-169.37	Tie into Survey at 8431.00 MD
2	8543.95	3.60	62.28	8509.69	-17.87	610.17	0.00	0.00	-167.72	Start DLS 10.00 TFO -78.60
3	9448.14	91.09	343.63	9074.58	552.52	479.69	10.00	-78.60	417.23	Start DLS 10.00 TFO 68.69
4	9470.42	91.90	345.71	9074.00	574.00	473.80	10.00	68.69	439.50	91.90 at 9470.42 MD
5	10940.42	91.90	345.71	9025.26	1997.73	111.16	0.00	0.00	1908.69	Start DLS 2.00 TFO 178.73
6	10985.43	91.00	345.73	9024.12	2041.34	100.06	2.00	178.73	1953.69	91° at 10985.43 MD
7	12931.43	91.00	345.73	8990.16	3927.01	-379.54	0.00	0.00	3899.39	Start Drop -2.00
8	12956.43	90.50	345.73	8989.83	3951.23	-385.70	2.00	180.00	3924.39	90.50° at 12956.43 MD
9	16733.23	90.50	345.73	8956.87	7611.36	-1316.62	0.00	0.00	7701.05	Start Build 2.00
10	16803.73	91.91	345.73	8955.39	7679.67	-1333.99	2.00	0.00	7771.53	91.91° at 16803.73 MD
11	17445.52	91.91	345.73	8934.00	8301.31	-1492.12	0.00	0.00	8412.97	TD at 17445.52

G
T
M
West(-)/East(+) (1500 usft/in)
-3000
-1500
0
1500
South(-)/North(+) (1500 usft/in)
9000
7500
6000
4500
3000
1500
0
PTB 17421'MD / 8950.24'TVD
Last Survey 17122'MD/ 8963.11'TVD
8934
8950
9079
11101
10000
9000
8000
7000
6000
5000
4000
3000
2000
#5
10999
10692
0
9000
Parks "26A" #2
West(-)/East(+) (200 usft/in)
-1600
-1400
-1200
-1000
South(-)/North(+) (200 usft/in)
8200
8000
True Vertical Depth (2500 usft/in)
2500
5000
7500
10000
Vertical Section at 345.73° (2500 usft/in)
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
16000
17000
17421
17446
Created By:Jenise Kirkpatrick
Date: 8:36, March 31 2018

Figure 1.10. Well Schematic Document, Courtesy of Legacy Directional Drilling

1.5.3. Directional Driller Activity Log

At the drilling rig, several companies are hired by the operator to execute the well together. The drilling contractor provides the rig and the main personnel to operate the rig. The service companies provide expertise and equipment to execute specific tasks such as well cementing, well logging, directional drilling surveying, drilling cuttings treatment and disposal, etc.

In regards to directional drilling, the specialized professional called a Directional Driller (DD) works together with the driller in the execution of the directional path of the well. The driller is the main person responsible for the normal drilling operation and works based on the same drilling rig. He controls the top drive, mud pumps, drawworks, and all drilling related equipment. Also, he is the leader of the drilling crew (roughnecks, derrickman, assistant driller) and is responsible for assuring that safety barriers and procedures are begin put in place to avoid incidents. Finally, he is also the main person responsible for observing the drilling parameters to quickly detect kick signs, and act accordingly using the well control practices that he is trained upon.

The directional driller is a professional that is sent to the rig only when directional control is needed. Since he works for the directional drilling company and not for the drilling contractor, he/she is assigned on a job basis, usually working among several rigs and crews. This professional has the skills to execute the curve section as close as possible to the planned trajectory, and works together with the driller, providing instructions about the drilling parameters such as angle build rate, drill string rotation speed, etc.

For reporting purposes, the directional driller has to discretize every drilling operation in a chronological log. This report is used later to evaluate the performance of the drilling operation, as well to calculate any non-productive time (if any). This report contains especially useful information about operational problems that are not evident in the EDR file. For example, a very sudden stop in the drilling operation, with no apparent reason in the sensors reading can be explained due to a failure in the mud pump gearbox that was reported in the DD Activity Log. Table 1.2 shows a few rows of an activity log.

Table 1.2. Example of Directional Driller Activity Log

Job #	Rig	Job ID	Job BHA	Well BHA	Motor MFG	Motor Size	Motor Bend	Motor Stator	Motor Stages	Hole Size	Bit MFG	Bit Model	Comment	Code	Start Datetime	End Datetime	Start Depth	End Depth	Delta Hours	Delta Depth	Max Axial Shock	Max Radial Shock	Max Axial Vibe	Max Radial Vibe	Item Failed	NPT	V/C/L	MWD Run
182	Rig A	MD180037	1	1	X	8	1.5	0.88	4	12 1/4	ABC	A1B2	CUT DRILL LINE	OTHER	4/28/2019 11:30	4/28/2019 13:30	1200	1200	2	0	35.84	66.78	3.16	5.28	NONE	0	WELL SPUD	1
182	Rig A	MD180037	1	1	X	8	1.5	0.88	4	12 1/4	ABC	A1B2	DRILL CEMENT AND FLOAT EQUIPMENT	CIRCULATING	4/28/2019 13:30	4/28/2019 14:30	1200	1200	1	0	35.84	66.78	3.16	5.28	NONE	0	WELL SPUD	1
182	Rig A	MD180037	1	1	X	8	1.5	0.88	4	12 1/4	ABC	A1B2	ROTATE 1200-1285	DRILLING	4/28/2019 14:30	4/28/2019 15:00	1200	1285	0.5	85	35.84	66.78	3.16	5.28	NONE	0	VERTICAL	1
182	Rig A	MD180037	1	1	X	8	1.5	0.88	4	12 1/4	ABC	A1B2	SURVEY & CONN. @1228' INC 1.22 AZM 174.35	SURVEY & CONN.	4/28/2019 15:00	4/28/2019 15:30	1285	1285	0.5	0	35.84	66.78	3.16	5.28	NONE	0	VERTICAL	1
182	Rig A	MD180037	1	1	X	8	1.5	0.88	4	12 1/4	ABC	A1B2	DRILLING – (WOB:15.00;GPM :78.00;RPM:30)	DRILLING	4/28/2019 15:30	4/28/2019 15:50	1285	1375	0.333	90	35.84	66.78	3.16	5.28	NONE	0	VERTICAL	1
182	Rig A	MD180037	1	1	X	8	1.5	0.88	4	12 1/4	ABC	A1B2	SURVEY & CONN. @1318' INC 0.56 AZM 167.69	SURVEY & CONN.	4/28/2019 15:50	4/28/2019 16:00	1375	1375	0.167	0	35.84	66.78	3.16	5.28	NONE	0	VERTICAL	1

1.6. Data Analytics Workflow

Data analytics combine "the procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise and more accurate" [Tukey, 1962]. This definition from the famous mathematician John Tukey, known for creating the Tukey range test, inventing the box plot and coining the term "bit", among several other accomplishments, tells us two important things: First, that besides all the hype people are talking about data science, analytics, big data and artificial intelligence (AI), data analytics is nothing new. Second, this definition states that the planning of gathering the data, as well the techniques for interpreting it are equally, if not more important than the analysis itself.

For this work, the data analytics workflow chosen to be utilized is the KDD, briefly described in section 1.3 of this work. Now, we explore in more detail what each step means in the process of turning raw data into knowledge.

1.6.1. Selection

The first step in the KDD process is to define the target data set on which discovery is planned to be performed. Here the data structure that will enable the user to access the data is established. The target data set can be one or multiple tables, containing whole or a subset of variables from different sources of data.

1.6.2. Preprocessing

Data cleaning and preprocessing steps include the strategies for handling missing values, remove or account for noise, removal of unnecessary portions of the data set and fixing textual information.

1.6.3. Transformation

Transforming the data covers the actions used to create new features to better represent the data. Most times, a reduction of uncorrelated or invariant features is performed to decrease the dimensionality of the data set, thus making the processing steps more efficient. For example, the [Block_Position] column represents the instantaneous position of the top drive in respect to the rotary table, storing values such as 40 feet. If two consecutive values of the [Block_Position] column are subtracted and stored in a new column called [Block_Movement], this new column now provides much more information about the same equipment. It can be inferred the direction of the movement, given the resulting sign of the subtraction, and the speed, represented by the magnitude of the value.

1.6.4. Data Mining

From all the steps in the KDD process, without a doubt the data mining part is the most notorious. Widely discussed in the literature and in the mainstream media, data mining consists in applying basic but well understood statistical and mathematical techniques to search for patterns and try to understand them, predict future values, or both. The basic concepts from which most algorithms are derived from are: model representation (the language used to describe discoverable patterns), model evaluation (criteria or functions that evaluate how well a model meets the goals of the KDD process) and search, a concept that can be subdivided into two components (parameter search and model search, where basically both loop into searching for the best match that optimizes the model evaluation). An example of data mining is the application of the Decision Trees algorithm as a representative model, using the coefficient of determination (R^2), performing a parameter search on the optimal number of leafs and nodes of the model that result in the highest R^2.

1.6.5. Interpretation and Evaluation

This step involves visualization of the resulting data extracted from the models, or simply the patterns now visible in the data itself after all the cleaning and transformation routines. In this step also domain knowledge is applied to support the application of the knowledge created with the process to add some benefit to the user or task performed, preferably compared against a well define metric. The results should be understandable (with some postprocessing or not) and should add value to the overall process that is being represented by the data analyzed.

It is important to note that iteration can occur between all the steps, meaning that the overall process is due to find enhancements if an intermediate discovery affects a previous step. This iterative nature is common to data analysis procedures but is emphasized in the KDD process by the connecting arrows.

1.7. Literature Review – Data Analytics in Drilling Engineering

1.7.1. Drilling Data Quality and Structure

One of the main challenges in the application of data driven platforms is the setup of the data network and the quality control of the data that is being used by the system. At an enterprise level, it is normal to have files that are being used in daily workflows being copied for individual use. Once a copy is created, the person runs some analysis, modifying the data, creating metrics and visualizations from it. This might be moved forward to its direct managers and directors, being modified at each step. If another employee uses the same data to do another analysis, he/she will download it from the same source, but all the transformations will happen independently. For the management of the data in the organization, a system like this creates a complete chaos, with uncontrolled versions of business files circulating back and forth [Damski, 2014].

For the drilling domain, we are dealing with data coming from a variety of sources – from EDR data to manually generated reports. [Brannigan and Co, 1992] proposed a clear schematic on how big enterprises should organize its drilling data based on four criteria. Once the data is structured in this format, a logical relationship between the data entities (well number, basin, drilling phase, Authorization for Expenditure (AFE) number, etc) can be easily queried in a relational database network. A more recent take on the same idea of data architecture was presented by [Al-Khudiri et al., 2008]. The main advantage of this new structure was the application of the WITSML standard, created in early 2000s in a more effective manner. With focus in completion data, [Reddicharla, 2015] described an automated workflow with enhanced data quality for designing completion strings. The quality gains were achieved mainly due to standardized data entry templates applied across its organization.

Once a company has the data structure set, the next step required is to build the information technology (IT) structure that will enable its use in real time. [AlBar et al., 2018] described the path taken from data collection to performance optimization. AlBar stated that the huge bottleneck in the application of drilling data is data preparation and data control, as well the operationalization of analytical models in real-time. In the same line, [Spivey et al., 2017] applied physical concepts such as mechanical specific energy (MSE) and torsional severity estimate (TSE) in the application of an advisory system installed in the drillers chair for real time optimization. Surface data was used to estimate drilling state and indirect downhole conditions with enough accuracy to guide the driller to parameters change towards a better ROP without creating bit dysfunction.

Once the data collected is organized and the structure is put in place, the next step before application of the data is the evaluation of data quality. By that, is meant not only the pure values of the sensors, but also the combination of recordings. For values that are calculated from readings, the verification of the calculation for each recording can be performed for quality issues. One example is the calculation of WOB using the reading of the Hook Load (HL) and the input variable String Weight (SW) in the EDR system. The relationship between WOB, SW and HL is:

$$SW = HL + WOB \tag{1.2}$$

SW should be simply the weight of the drill string at particular time, measured when the drill string is being hung by the Top Drive (direct reading from HL reading). The value of SW increases as new stands are added to the drill string with drilling progress. To update the SW value in the EDR system the driller should zero (tare) the SW value before drilling with the recently added stand. From Eq. 1.2, if the SW is not updated as a new stand is being drilled, but the system is reading a HL value that is greater than the SW (due to the added weight of the new stand), the system will apply a negative value of WOB to hold the equation true, even though negative values of WOB have no physical meaning.

It is important to note that the relationship described in Eq. 1.2 is only valid for vertical wells, since the whole string weight is being supported by the hook load. For inclined and horizontal wells where part of the string is in direct contact with the wellbore wall, a correction for friction and inclination has to be made to correctly calculate the downhole weight on bit (DWOB). [Mitchell, 1976] presented a detailed description of the physics involved in weight distribution on inclined wells. More recently, [Hareland et al., 2014] modeled the accurate calculation of the DWOB using surface measurements and friction coefficient calculations, validating their results with directly measured DWOB. Although a good number of technical papers can be found dealing with this issue, the application of WOB correction for inclined wells is still not a standard practice among EDR systems.

[Neufeldt et al., 2018] did a detailed analysis of the zeroing WOB and found that drillers forget to zero WOB in around 80% of the stands, yielding to errors as large as 100% of the measured WOB. With these numbers in mind, they developed a computational routine that can correct for this error, updating the SW value based on actual readings of the HL in a correct moment. To mitigate changes in the HL load readings, the best moment to apply the artificial correction is before drilling is resumed after a new stand is connected, but after the rotation is set to the actual drilling value, the mud pumps are already on and at final value, the travelling block just starts to move down and the bit is off bottom. These conditions are important to be met to avoid errors due to friction forces acting on the drill string in drilling conditions but are not present in a static situation. Since the mud motor differential pressure (difference between the pressure when the bit is off bottom versus the pressure when drilling) also needs to be zeroed in every stand to account for changes in the pressure losses as the drill string gets longer, the algorithm proposed by Neufeldt can be applied also to the differential pressure variable.

An independent work performed by [Borjas et al., 2019] shows a very similar approach to solve the issue of data quality in EDR data for these cases where a repetitive action is needed to be executed by the driller to guarantee accurate recordings. Since both authors work for EDR manufacturers, it is clear that data quality is getting importance not only for the end user of the data, but for the designers of the systems that are collecting it. Borjas also described that the routine developed to correct WOB and differential pressure can be activated in some of the EDRs running in rigs, demonstrating a logical solution that solves the issue at the source, rather than relying on the analyst to find and fix it at a later time. The last two references dealt with correcting data that was being recorded incorrectly due to human error. But a remark has to be made in regards to the quality of the HL measure itself. The first issue arises with the fact that the total weight that is being hung by the hoisting system is usually an indirect estimation using the tension measured in the dead-line anchor load cell. The dead-line is the portion of the drill line that is set to be static during normal operations, comprising the portion after the last sheave of the crown block and the anchor that fixes the drill string. Figure 1.11 presents a schematic view of the drill line spooling scheme in a typical drilling rig.

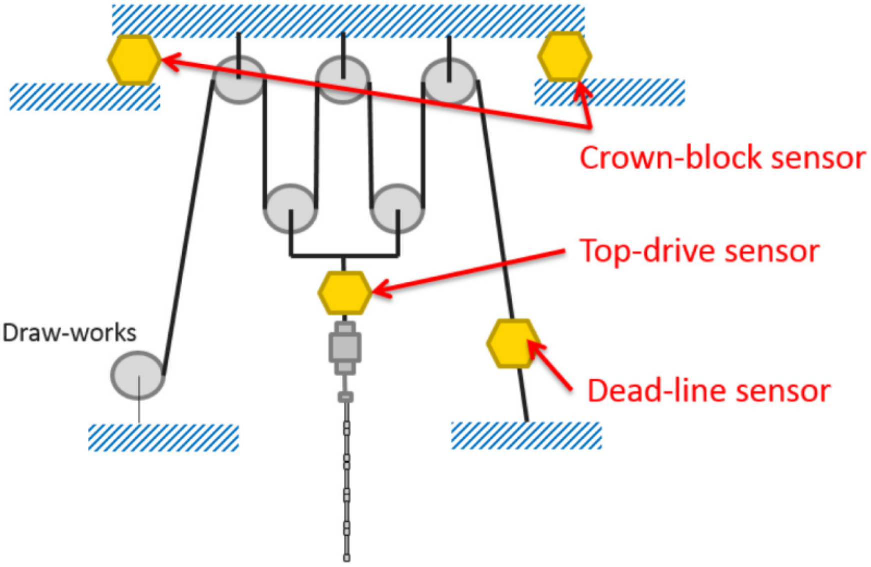


Figure 1.11. Typical Drill Line Spooling Scheme, Adapted From [Eric et al., 2015]

[Eric et al., 2015] studied the implications of this indirect reading of the hook load, including the forces generated by friction in all the pulleys in the system, the difference in height of the travelling block, the lateral forces applied by the dolly (retraction) system and the hydraulic hoses that are hung by the top drive, among other effects. In the work also a model is proposed to correct for these factors, resulting in better readings of hook load. Their work was focused primarily in the surface forces acting in the drill line spooling system that affect the hook load readings. [Kyllingstad and Thoresen, 2018] complement the analysis focusing in other parameters that also can cause errors in the measurement: well bore friction, buoyancy, well inclination, lift induced by flow and nozzle jetting, etc. These two extensive articles are good examples of all the complications that are usually neglected when analyzing a simple measurement such as the total weight hoisted by the travelling block. The analyst must have these factors in mind when using EDR data to make any conclusion about the drilling process using the WOB or hook load recordings.

Together with the data issues described so far, noise and wrong measurements can be found in practically any reading, if simple quality control measures are not put into place. [Ashok et al., 2018] detailed how other industries reduce measurement error by simply having more than one sensor taking the same reading. In case of sensor malfunction, the other readings are used to invalidate the bad reading, removing it from the recording and also raising a flag that maintenance is needed. With the problem of data recording in mind, Pradeepkumar developed a system that validates the readings of eight core readings (block position, hook load, rotary speed, rotary torque, pump strokes per minute, flow rate out, standpipe pressure and pit volume). The approach used was a statistical based model using Bayesian network to validate data in real-time.

Finally, [Maidla et al., 2018] discussed overall misconceptions about drilling data recordings, the errors associated with drilling measurements, and some guidelines that should be considered to avoid pitfalls in the analysis. Common mistakes such as evaluating gross ROP measurements instead of net ROP (when drilling is actually taking place) are highlighted in the work. Maidla also describes that in many drilling rigs the torque measurements are not direct either. In these cases, the torque is calculated using the current drawn by the top drive needed to maintain a set rotation. For these systems, it is stated that calibration is often forgotten, which implies there would be more uncertainty regarding the data.

After this literature review, it became clear that most issues on drilling data are caused by poor design of the recording systems. In my understanding, one of the causes for this issue is that the people that design these systems are not the ones that use the data. If they were and data users, it would be obvious that some readings are so wrong, or so poorly estimated, that they would take some action to do something about it. This fact is what motivated for example, [Borjas et al., 2019] to not only develop a correction for the WOB zeroing, but to automatically implement it in their systems.

Two recent initiatives that aim to solve this issue as an industry are worth mentioning. First, the work of [Halloran et al., 2018] describes the five types of data quality issues in sequence: systematic, measurement, conversion, calculation and propagation. Second, it describes the work performed by The Operators Group for Data Quality (OGDQ), a cooperation to fix the data quality issues on the key measurements used in the drilling process. Another joint effort was presented recently by [Pastusek et al., 2019], and thirteen other authors in a review of many open source models that were already published in drill string hydraulics, dynamics, directional control and bit-rock interaction models. The idea is to create a similar repository as the Open Porous Media (OPM, www.opm-project.org) but for drilling engineering.

1.7.2. Classification of Drilling Operations

The process of drilling oil wells can be described as a repetition of a few operations. The sequence of normal drilling operation can be summarized as:

- Drilling the whole in a portion equal to the drilling stand length (a drill stand usually comprises three drill pipes of around 30 ft long each.).
- Circulating: Once the top drive reaches the rotary table, a short time is spent to circulate the well prior to connection, where the mud pumps will be turned off and the cuttings may settle around the BHA.
- Connection: The drill string is put on the slips and a new stand is added to the string.
- Drilling resumes until the top drive positions reaches its lowest limit.

Less frequent operations also happen in the process of drilling wells. Tripping pipe is the term given to the operation of moving the drill bit in and out of hole between drill runs. Tripping operations are necessary for various reasons such as when the target depth is reached, when the casing setting depth is reached, when the bit is suspected to be dull (low ROP rate), when some downhole tool fails, etc. Therefore, it is also important to identify when the current operational status is:

- Tripping in: Moving the drill bit and BHA inside the whole.
- Tripping out: Removing the drill bit out of the hole.

Finally, specific operations that take a small percentage of the total drilling time of the well are also important to be distinguished for further analysis and performance evaluation. According to [Cao et al., 2018], the task of recognizing drilling activities is the basis for all other analytics routines that are performed using drilling data. In his work, Cao also included in the classification the Reaming and Back reaming operations, Slide off bottom and On Surface statuses. A good work in the automated classification of drilling activities using surface drilling parameters was performed by [Al-khudiri et al., 2015]. The scope of the work not only describes the classification process but also the application of key performance indicators (KPIs) that allow benchmarking and offset well comparison. The data structure used in the real time application, as well as the data quality control measures and examples of reports are described to provide the reader examples of analytics applications.

Other authors attempted to identify drilling operations using trend analysis or statistical and machine learning algorithms. [Serapião et al., 2007] created a model to classify drilling operations using support vector machine, which is a hyperplane classifier. Their model obtained a correctness rate (ratio between right classified cases and total cases) of 92.6%. [Arnaout et al., 2012] demonstrated a technique using discrete polynomial analysis to extract moments (patterns) for each drilling operation. Then, these patterns were applied to three test wells and obtained accuracies between 88-94%. In a similar approach of extracting traces in the data for the different types of operations, [Kristjansson et al., 2016] applied finite mixture modeling (or maximum-likelihood approach to clustering) to generate archetypes (patterns) for a group of drilling parameters. These archetypes were created for depth intervals, and then the archetypes with greater ROP were combined to determine a recommendation of the optimum drilling parameters that should yield in the best drilling rate for a future offset well, based on historic data.

1.7.3. Applications of Data Analytics in Drilling Engineering

- **Data Analytics**

Once the analyst has the drilling data organized, there are many applications of data analytics and machine learning that can be performed to extract knowledge from it, whether using past data or real-time recordings. [Eren and Kok, 2018] present their approach for comparing drilling performance between forty wells using ROP indexing. As stated by the authors, it is important to separate the flat time from the drilling time to obtain meaningful results. The indexation calculates the rate of penetration at the same depth and time basis, allowing comparison in the same order of magnitude. [Ashok et al., 2018] applied the storyboarding process to answer questions that could be asked once one has the data set at hand, such as: "what was the fastest run?" or "which well has the least tortuosity?". Again, once the data is structured, the idea is to pre-calculate some KPIs that would answer such questions, and by the time the user asks them in the platform, the answer will be presented instantaneously. In a sequence of this work, [van Oort et al., 2018] showed the back-end processes that enable the storyboarding process to happen. The main idea is to pass the data through a series of scrips – called bots – that will clean, process and index the data in a way that it can be readily accessed. If cost data is available, an analysis as the one done by [Willis and Jackson, 2018] can be performed. An important remark from their work is the highlight of the revised budget versus original budget. This will give the true baseline to compare with the actual cost, avoiding misleading over-cost scenarios.

- **Real-Time Optimization**

Several authors presented their application of data analytics to drilling operations in real-time. [Brooks et al., 2017] analyzed drilling parameters from offset wells to create a model that was later applied to a drilling campaign of three wells. The real time monitoring of MSE and ROP was used to compare the model values against the actual readings, orienting the driller to correct the drilling parameters. This approach resulted in an overall improvement of 30.9% of drilling time and invisible lost time (ILT) 47% less. Focusing on the bit-rock interaction, [Millan and Ringer, 2018] propose a workflow to estimate bit wear and in-situ-rock strength, to further estimate ROP ahead of the bit. The model can be used to estimate bit wear in real time, thus providing insights on when to pull the bit to obtain the least drilling time possible. The application of real-time data in a multi-well management level is illustrated by [Bolen et al., 2018]. Several dashboards provide a high-level picture of the drilling performance across different rigs, when drilling different wells. The data can provide insights on the ILT operations (tripping time, cementing, casing runs, etc) and resulted in an average of 40% improvement in the time spent.

- **Application of Analytics in Real Time Operation Centers (RTOC)**

When companies spend thousands of dollars in the IT structure to read drilling data from the rig in real-time, the logical approach is to combine these readings in one place, the RTOCs. Placing a few experts to analyze the real-time data in the RTOCs, the benefits can be seen in faster decision making, reduction in stuck pipe incidents, less hole cleaning issues and fluid losses events, while increasing the wells that can be monitored with the same number of personnel [Al-khudiri et al., 2015]. Other benefits of the application of RTOCs were reported by [Almeida Leon et al., 2013] as hazard prevention, better work relationships between field personnel and remote operators, reduced human intervention. Utilizing another source of data, the mud logging,[Bermúdez Martínez, 2012] combined offset wells data in a RTOC to predict in real-time the pore pressure and fracture gradient readings using correlations. With this data, the fluid engineer can calibrate the fluid density and casing set points, minimizing uncertainties in the well design as well as reducing safety risks of kicks, for example. Finally, [Mandava et al., 2018] provide best practices and applications for those who are interested in implementing RTOC in their operations.

- **Machine Learning Applications**

Further applications of drilling data to optimize drilling performance is the use of statistical based models and machine learning algorithms. Since drilling data is generated at very high volumes, statistical analysis tends to be well representative of the actual physics and phenomena happening in the drilling process. This is presented by [Liu et al., 2018], when past data was used to calibrate a statistical model to calculate wear factor to predict the optimum time to pull the bit due to decrease in cutting efficiency. Once the wear factor is calculated, a decision tree algorithm is applied to answer if the bit should be pulled or not. In twenty-five bit runs, the results showed only two false alarms and a success rate of predicting bit failure of 92%. [Evangelatos and Payne, 2016] applied spatial discretization to calibrate drilling fluid and borehole dampening coefficients, thus identifying scenarios of forward or backward whirl, or a mix of the two. Also, neural network algorithm was used to predict ROP with errors less than 20%. Bayesian Network was successfully applied to identify drilling dysfunction by [Thetford et al., 2017]. Besides identifying drilling dysfunction, this system calculates the likelihood of the cause of such dysfunction, and then recommends the change in these parameters to solve the problem in the form of operational cones.[Cao et al., 2018] applied data analytics and machine learning to create a full suite of real-time drilling optimization. The individual packages include drilling activity recognition, rotation and sliding drilling guidance, torque and drag modelling, real time hydraulics calculation, wellbore trajectory correction, among others. Another interesting application of neural networks is to generate synthetic data to calibrate automated systems. The main problem when generating drilling data is the lack of randomness and noise.[Yu et al., 2018] proved that Deep Neural Networks (DNN) can represent the noises that typically occur in EDR data, thus providing a better training data set for autonomous systems.

For a thorough review of machine learning applications in drilling operations, as well as in exploration and production, I refer the reader also to [Noshi and Schubert, 2018] and [Noshi et al., 2018] respectively.

Chapter 2
Materials and Methods

2.1. Data Set Description

The data set used in this study gathers data from thirteen wells drilled in the US Permian Basin, located in West Texas. All the wells are placed close to each other in a rectangle two by three miles long. The wells' proximity is important because the lithology found during the drilling of each well should be very similar, making any statistical analysis much more representative. Also, since the Permian Basin is one of the most active drilling sites in the country, any conclusion regarding optimum operating parameters should be easily reproducible in future wells.

Figure 2.1 depicts the basin location and what pad drilling looks like. The thin traces along the horizontal portion of the figure represent the hydraulic fracturing operation. Hydraulic fracturing a reservoir is the process of pumping fluid with proppant material to enhance the connectivity (permeability) of the reservoir, enabling the fluids to flow.

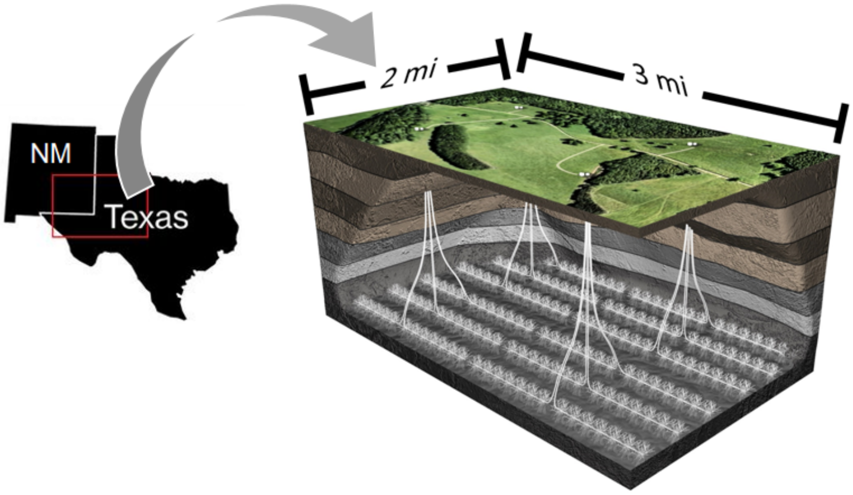


Figure 2.1. A Highlight of the Permian Basin in West Texas (left) - A Schematic of the Pad Drilling (right)

Pad drilling is the drilling layout on which multiple wells depart from a close location on surface but hit different areas in the same or multiple reservoirs. A major contributor that allowed the execution of this drilling pattern is the recent added capability of the drilling rigs to move (or “walk”) with the derrick raised and all equipment installed, including the blow out preventer (BOP). The movement is achieved with several pistons installed beneath the rig substructure that move in small increments using hydraulic power (Fig. 2.2). With this technique, the operators are now able to drill multiple wells from the same surface location, which represents significant savings in cost and time when compared to conventional drilling.

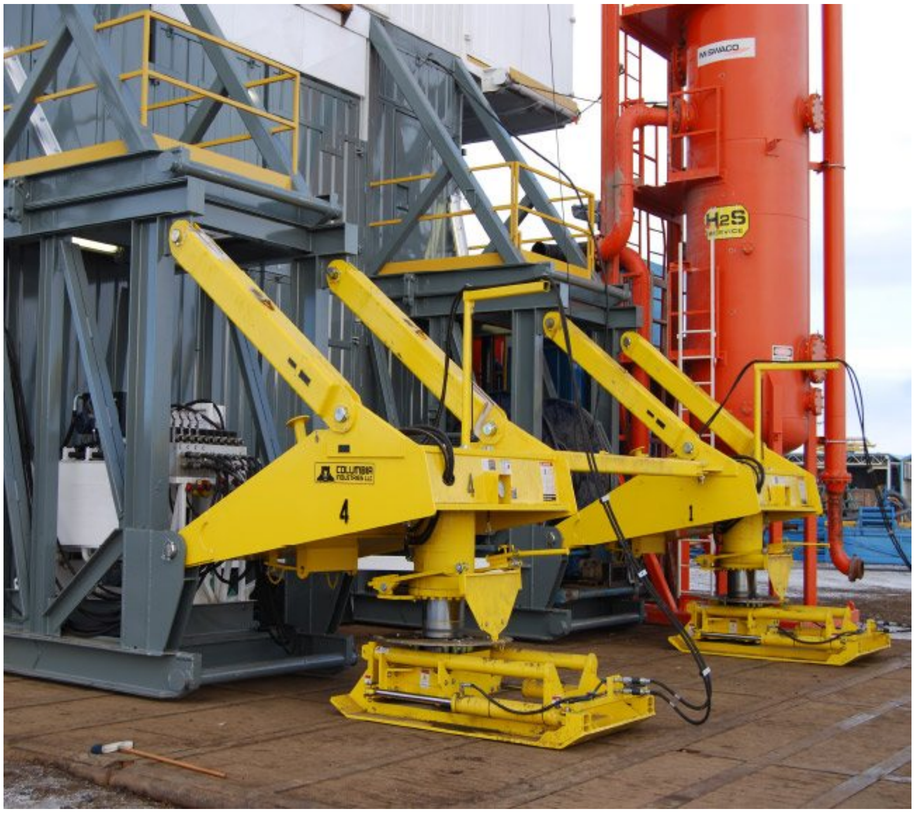


Figure 2.2. An Example of the Rig Walking System, From Columbia Industries

The thirteen wells are all of similar shape profile, horizontal wells with around 8,500 to 9,000 feet of vertical (or nudge) section, then they have between 600 to 1,000 feet of curve section, and once they achieve an inclination close to 90 degrees, the horizontal drilling continues for around 7,000 feet, reaching a total measured depth (MD) of 17,000 feet on average. Fig. 2.3 depicts the typical well profile used in this work.

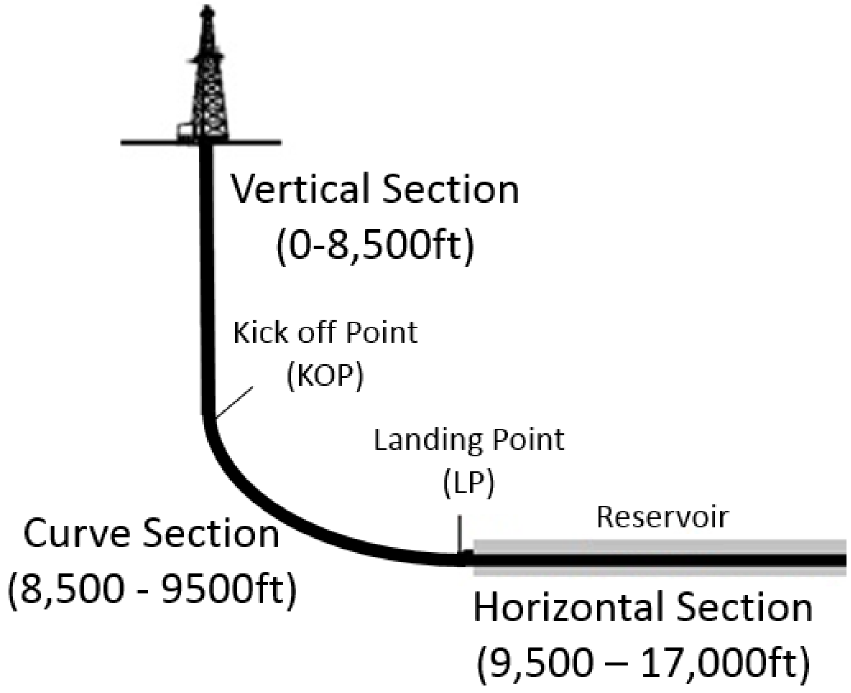


Figure 2.3. Typical Well Schematic From the Data Set

2.2. Table Schema

Once all data files from the different sources were collected, the first task was to organize the information in a structure that will allow easy access, read and update during the execution of the project. It can be noted in Table 1.2 for example, that some columns are filled with a unique value. This is not a good practice for a few reasons. First, it makes the table more crowded and difficult to read. Second, when you scale up for a table with millions of rows, adding columns with just a single value can slow down any coding or querying from that table significantly. To overcome this possible issue, the information from the different sources – the well schematic and directional driller activity log – was re-organized into two new tables, called reference tables.

The first reference table is called “Wells Overview” (Table 2.1). It contains basic information regarding the profile of the wells (KOP depth and time, LP date and time, total depth, etc). One important information is that for each well, there is a unique [Well ID] that describes this well, meaning that all information from a particular well must be in the same row. The Wells Overview table also contains generic data such as the rig that drilled the well, the well number inside the company's code, date and time for the well spud (the start of the actual drilling in a job) and ”well TD” (the moment when the drilling reaches the total depth, or TD).

Table 2.1. Wells Overview Table Example

Well ID	Pad #	Rig Name	EDR Start	EDR End	Well Spud	Well TD	Job Start Depth [ft]	Job End Depth [ft]	KOP Depth [ft]	KOP Datetime	LP Depth [ft]	LP Datetime
well1	1	Rig X	1/1/2018 00:00	3/1/2018 12:00	1/4/2018 11:33	2/28/2018 09:29	1,250	16,253	8,850	2/1/2018 02:57	10,152	2/4/2018 22:03

The second reference is Table 2.2 , called “”Wells BHA Overview”. This table gathers information describing each BHA configuration run in each well. Since all the wells are drilled with multiple bit runs, there is a not unique combination of either ”Well ID” or ”BHA Number” in this table. The solution is to create an aggregated column that combines the two columns into one (now unique) column. Having the information organized in such way is important and will allow the comparison between different bits, mud motor manufacturers, motor specifications and settings, that will provide insights regarding drilling efficiency.

Table 2.2. Wells BHA Overview

Well ID	Well BHA #	Motor Bend	Motor RPG [rpm/gal]	Motor MFG	Motor Size [in]	Motor Stator	Motor Stages	Bit Size [in]	Bit MFG	Bit Model	Datetime In	Datetime Out	Total Circ. Time [hr]	Depth In [ft]	Depth Out [ft]	Total Footage [ft]
well1	1	$1\frac{3}{4}$	0.166	MFG X	8	0.88	4	$12\frac{1}{4}$	A	ABC12	1/4/2018 11:33	1/9/2018 08:23	82	1250	5741	4491
well1	2	$2\frac{1}{4}$	0.28	MFG Z	$6\frac{3}{4}$	0.88	5	$8\frac{1}{2}$	B	CDE34	1/9/2018 17:39	1/15/2018 03:00	90	5741	12487	6746

Lastly, the EDR files are downloaded per well. The columns downloaded are:

- DateTime
- Hole Depth
- Bit Position
- Bit Weight
- Block Height
- Mud Motor Differential Pressure (Diff Pressure)
- Gamma Ray

- Hook Load
- Pump Pressure
- ROP – Average
- Top Drive RPM
- Top Drive Torque
- Flow In Rate
- Pump SPM
- ROP – Fast (instantaneous ROP value)

Fig. 2.4 shows examples of the location of the EDR readings for the main components.

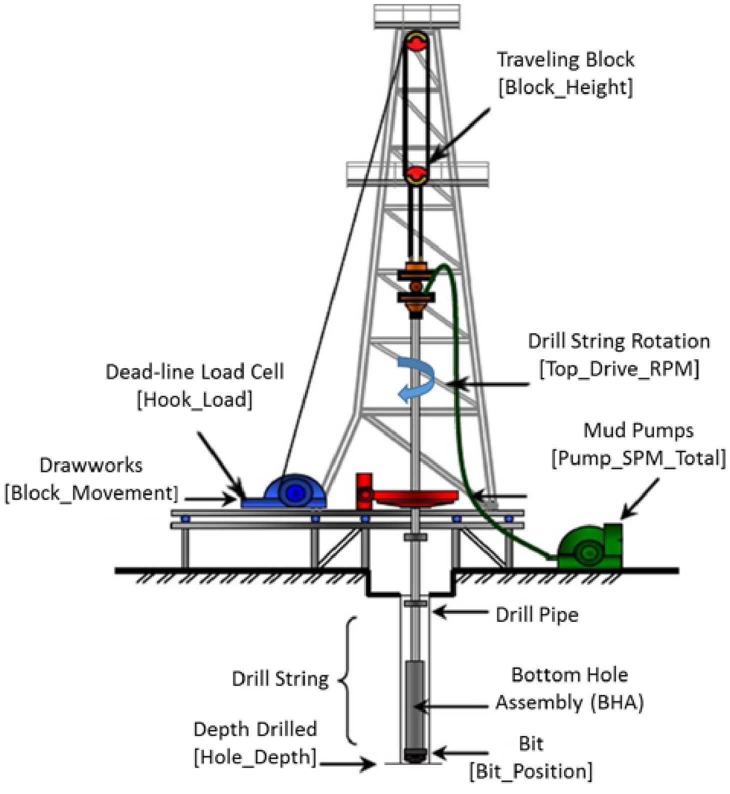


Figure 2.4. Examples of the Sensors for Main EDR Readings

Using the two reference tables, now each individual EDR file can be related to either the Wells Overview table with the [Well ID] value, or to the Wells BHA Overview table with the [Well-ID BHA] value. A schematic of this relationship is presented in Fig. 2.5.

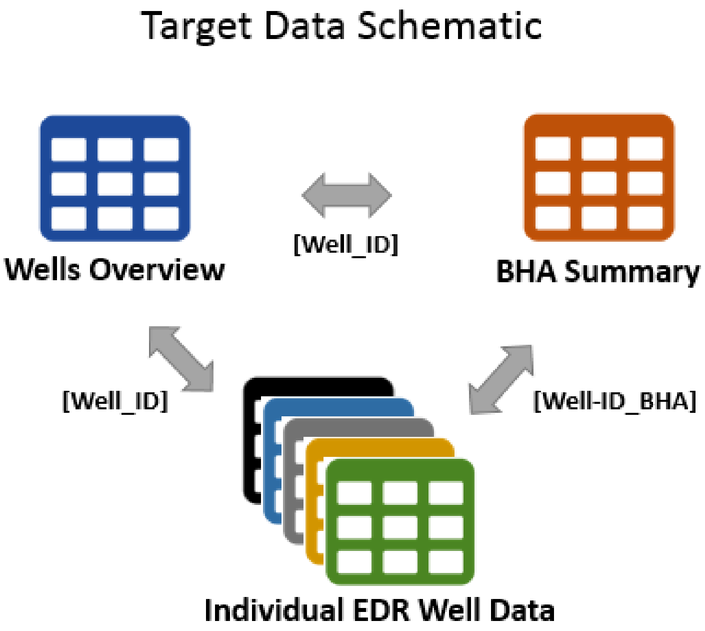


Figure 2.5. Relationship Between the Three Main Data Tables

2.3. Data Wrangling

Data wrangling defines the steps that are required to transform raw data (numeric or textual) into a format that it can be readily used for analysis, modelling, statistics, etc. Equivalent terms for these steps are Data Cleaning or Data Munging. For the KDD diagram (Fig. 1.5), the data wrangling steps comprise the data selection until data transformation. Over the next sections, we will describe the data wrangling steps needed to clean drilling data.

2.3.1. Cleaning Missing Values and Slicing the Data Set

To download an EDR file, the user has to access the EDR online repository in the manufacturer′s website, select the time interval to be downloaded and the columns wanted to be included in the file. But the fact that not all measurements are active during all the operations leads to some of the values to be null, or "-999,25" as defined as the null value for most EDR systems. This is true also if any failure happens during normal operations.

With the definition of "-999.25" as a null value, the substitution of this value for a "null" variable should be straightforward in any programming language. But first, it has to be verified if the -999.25 could be actually a valid reading, so one doesn't unintentionally delete good data. Since all the columns downloaded in the EDR should contain only positive values, this operation can be performed without concerns in this case.

Next, an obvious problem was detected by examining the EDR files: some rows have the Bit Position value greater than the Hole Depth reading. Since it is physically impossible to happen and was only observed in less than one percent of the cases, it was decided that for the rows in which [Bit Position] > [Hole Depth], the whole row will be filled with "nulls". But due to small inaccuracies in the calculations of the Bit Position by the EDR system, a more expressive number of cases is identified where Bit Position is greater than the Hole Depth no more than 0.2 feet, or 6 centimeters. These cases were kept in the analysis.

Finally, since during normal operations a total blackout of the system can happen, resulting in rows with all values as '-999.25". In some wells also, the EDR did not start to record during the first hundred feet drilled. Since the time interval selected to download the EDR came from the Wells Overview table, which reported the date time when the actual drilling started, this fact also resulted in some complete null rows until the EDR was set up properly. All these rows were deleted completely. For all the values in which a "null" variable was set in the operations described previously, then a value of 0.00 was input to it.

The second step in this initial data cleaning is to properly slice the data set removing periods of time that are either invalid or out of the interest for this analysis. For example, since the scope of this work is to only analyze data where actual drilling was being performed (hole depth was increasing), as soon as the TD was reached, the operations that took place after it (tripping out the bit, running the production casing, cement it, etc) are also removed. For the cases described in the previous paragraph, where the recordings did not begin until a certain point, we now slice these values as well based on the Well Spud date time in the Wells Overview Table. It is important to notice here that since the data came from a directional drilling company, the definition of Well Spud here is not the spud of the well (usually when the conductor pipe is put in place) but when the first feet with MWD in the BHA was drilled. This usually happens after the first two sections of casing are drilled and cemented.

After the first cleaning steps, a representation of the data set is shown in Fig. 2.6. It can be seen mainly that the missing values, as well as some of the first and last rows are now gone.

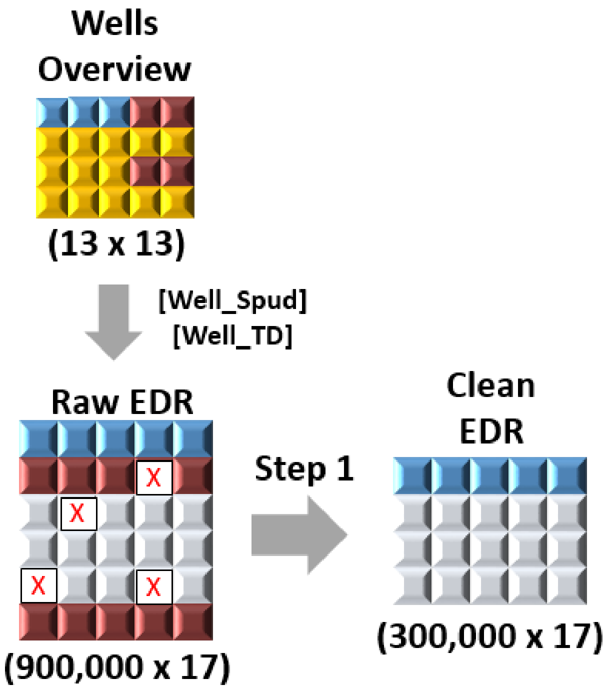


Figure 2.6. Illustration of the First Cleaning Steps. The Null Values are Cleaned, and the Wells Overview Table Provides the [Well Spud] and [Well TD] Datetime Values That Will Slice the EDR File

2.3.2. Creating Status Columns

Following the first cleaning step, next some calculated columns are created based on other values. The first calculated column is the [Making Hole] column, which is of Boolean type (TRUE or FALSE). The calculated value is simply TRUE if the actual Hole Depth is greater than the previous one, or FALSE if it is equal. This calculation will be used later to not only identify easily in which rows the drilling is progressing, but also to filter the data set for visualization and analysis.

The second status column is called [Block Movement]. It represents the difference between two consecutive block height values, being positive if the block is moving up (Actual height is greater then previous) or negative if the opposite happens. A column with this information can be very useful to infer drilling operations, when evaluated together with other values. For example, being all the other values equal, the signal of the [Block Movement] can be useful to directly distinguish between tripping in vs tripping out, or reaming vs back reaming. Also, if there is no movement, this column can support the misclassification of other operations in case of bad values (ex: the hole depth could be increasing due to an error in that reading, but if block movement is zero, it will avoid the misclassification).

A third status column is created to point out the off-set distance between the bit and the bottom of the hole. The column [Off Bottom Distance] is used as well to identify drilling operations more easily, and also to take into account small errors in the readings. As stated previously, for example, operations can still be classified as "drilling" even if the distance between the bit and the bottom of the hole is negative by a very small amount (0.2 feet), to account for problems in the bit position recording. Also, the [Off Bottom Distance] is helpful to quick access if a tripping operation is underway, when the distance is greater than one stand. The steps described in this section are represented in Fig. 2.7.

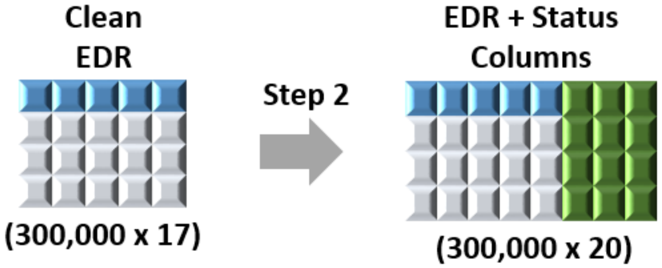


Figure 2.7. Representation of the New Status Columns Added to the EDR File

2.3.3. Referencing Well Section and BHA Number

Next, more data from the reference tables are added to the EDR file for easy slicing. First is the [Well Section], obtained with the comparison between the actual hole depth and the reference KOP and LP from the Wells Overview. If actual depth is less than KOP, its value is “Vertical”, if the value is in between KOP an LP, “Curve” and lastly if is greater than LP, [Well Section] gets a “Lateral” value.

Using the same logic, the [BHA #] from the Wells BHA Overview is read, and this value is added to the EDR file. With these both new columns in the EDR, any analysis of performance for each run, or the comparison between the drilling of the curve section in all the wells can be performed without trouble. Fig. 2.8 illustrates this step.

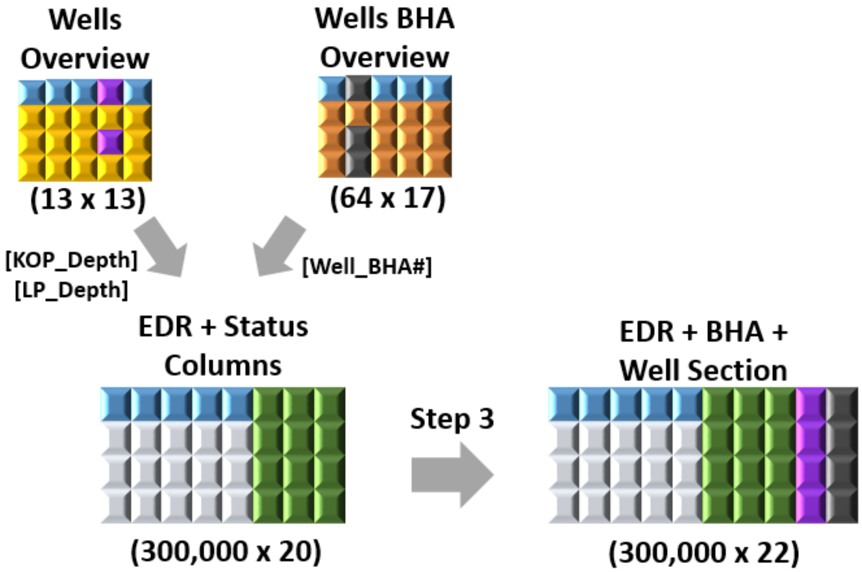


Figure 2.8. Addition of Well Section and BHA Number to the EDR File

2.3.4. Classifying Drilling Operations

As stated in Section 1.7.2, the process of drilling oil wells involves several different operations that are repeated as the drilling progresses. Therefore, it is crucial to properly classify these operations in the EDR before performing any analysis of the operation. Without proper classification, data points describing operations such as tripping, circulating, reaming will all be analyzed together with actual drilling data points. The opposite is also true: When the analyst wants to investigate tripping performance, all other data points should be excluded from the analysis. The present work classified drilling operations into 10 categories, utilizing only seven columns from the preprocessed EDR file (four preexisting columns plus the three status columns that were created). Table 2.3 summarizes the classification logic.

Table 2.3. Logic for Drilling Operations Classification

Column	Hole	Making	Off	Block	Top	Pump	Hook
Operation	Depth	Hole	Bottom Distance	Movement	Drive RPM	SPM Total	Load
Drilling		TRUE	>-0.2	<0.1	>10	>10	
Drilling/Rocking	>10000	TRUE	>-0.2	<0.1	5<x<45	>10	
Sliding		TRUE	>-0.2	<0.1	<=10	>10	
Reaming		FALSE	<= -0.2	<0	>10	>10	
Back-reaming		FALSE	<= -0.2	>0	>10	>10	
In Slip Connection		FALSE	<= -0.2				<=57
Tripping In		FALSE	<=-80	<0	<5	<5	>57
Tripping Out		FALSE	<=-80	>0	<5	<5	>57
Circulating/Survey		FALSE	<=-0.2			>10	
Other							

The criteria to classify the drilling operations was the following:

-Drilling: The Drilling label was given to the data points in which rotation drilling was happening. This is identified when the Making Hole value is TRUE, the bit is touching the bottom of the well, the travelling block is moving downwards, the drill string is rotating, and the mud pumps are on. Here we highlight that a small threshold for the block movement was given as well as for the off-bottom distance to account for errors in the measurements of block position. It was identified that in a sequence of drilling data points, sometimes the block position is recorded to move up, which is physically not correct. To account for those errors, a margin of 0.1 feet (or 3 cm) upwards is included.

-Drilling/Rocking: As the drilling advances and the drill string gets longer, more torque is developed in the drill string due to a longer contact area between the drill string and the borehole wall. In the lateral section, the driller drills with rotation unless the bit is deviating from the target formation, forcing it to slide to correct the trajectory. But the capability on applying WOB is compromised in such a long string due to longitudinal drag, the drill string elasticity causing some portions of the string to move at different velocities, and cuttings accumulation in the bottom part of the annulus due to poor hole cleaning while sliding, increasing friction significantly. These factors difficult the application of WOB, causing sudden releases of weight on the bit, and most times making the mud motor to stall. Frequent stalling of the motor can damage its components, so this problem must be corrected [Duplantis, 2016].

To control the bit orientation during a slide, the driller attempts to rotate the whole string applying successive torque clockwise and counter-clockwise, which is called "rocking" the bit. By doing this, drag is reduced and the length of the upper part of the drill string that is being rotated (up to the maximum rocking depth) increases (Fig. 2.9). The objective of the rocking operation is therefore to minimize the portion of the string that is static, i.e., does not rotate either due to the surface torque application or due to the reactive torque from the bit. With the static portion minimized, more control to apply WOB is obtained, thus the driller will spend less time pulling the bit off-bottom to release torque attempting to correct toolface orientation, resulting in better operational efficiency.

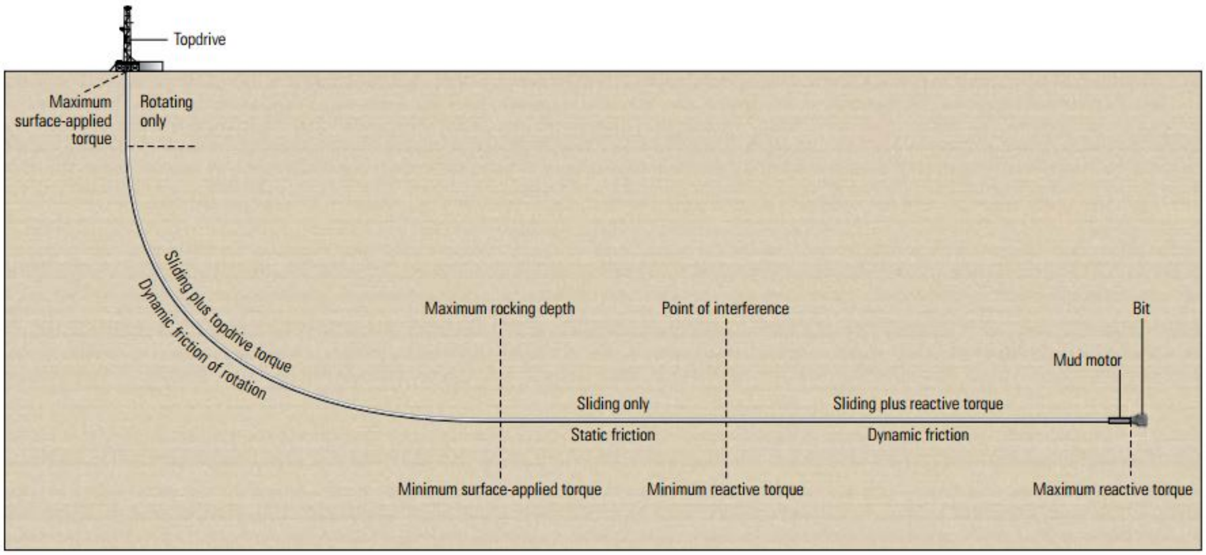


Figure 2.9. Representation of the Different Zones of the Drill String During Rocking, From [Duplantis, 2016]

This rock operation can be performed by an automated torque control system installed on the surface, or manually by the driller (as in the cases of this work). Details of this operation can be found in [Maidla et al., 2004], [Maidla et al., 2005] and [Maidla et al., 2009].

For the classification of the rocking operation in our data set, it was noticed in the directional driller activity logs that it has never occurred in depths shallower than 10,000 feet. Because of this observation, a set point on the hole depth was placed in the algorithm for the classification of this operation. Besides this depth set point, the only difference between the drilling and drilling/rocking operation is the top drive rotation. Because of the variations in rotation to control the bit orientation, it was observed that the range applied by the drillers was between 5 and 45 rpm for the majority of the cases. This range limit was applied to the top drive rpm column when classifying the drilling/rocking operation.

-Sliding: Sliding drilling is the operation when drilling is performed without rotation of the drill string. The BHA contains a bent sub with a slight inclination – 0 to 3 degrees – that will cause the well trajectory to deviate. To apply rotation to the bit, a mud motor is installed behind the bent sub. The motor will make the bit rotate as drilling mud is flown through the chamber. The chamber comprises of a rotor and stator, and the number of lobes of the assembly define the motor output of rotation and torque. The more lobes in the rotor/stator assembly, the higher the torque transmitted to the bit, but smaller is the rotary speed. Fig. 2.10 illustrates the mud motor components.

For the classification of Sliding operations, the distinction between drilling and sliding is made only with the top drive rotation. It is seen that even when sliding is being performed, sometimes the driller rotates the drill string slowly (no faster than 10 rpm) for short periods of time to overcome torque. So this value was selected as the threshold between rotating drilling and sliding. In depths greater than 10,000 ft however, the algorithm checks for the hole depth value before it checks the other columns, making the classification in data points that fall in the same interval (when top drive rotation is between 5 and 10 rpm) to be classified as drilling/rocking. This represents nature of the operation better, since at this point, sliding drilling is not expected to occur frequently; instead the driller is rocking

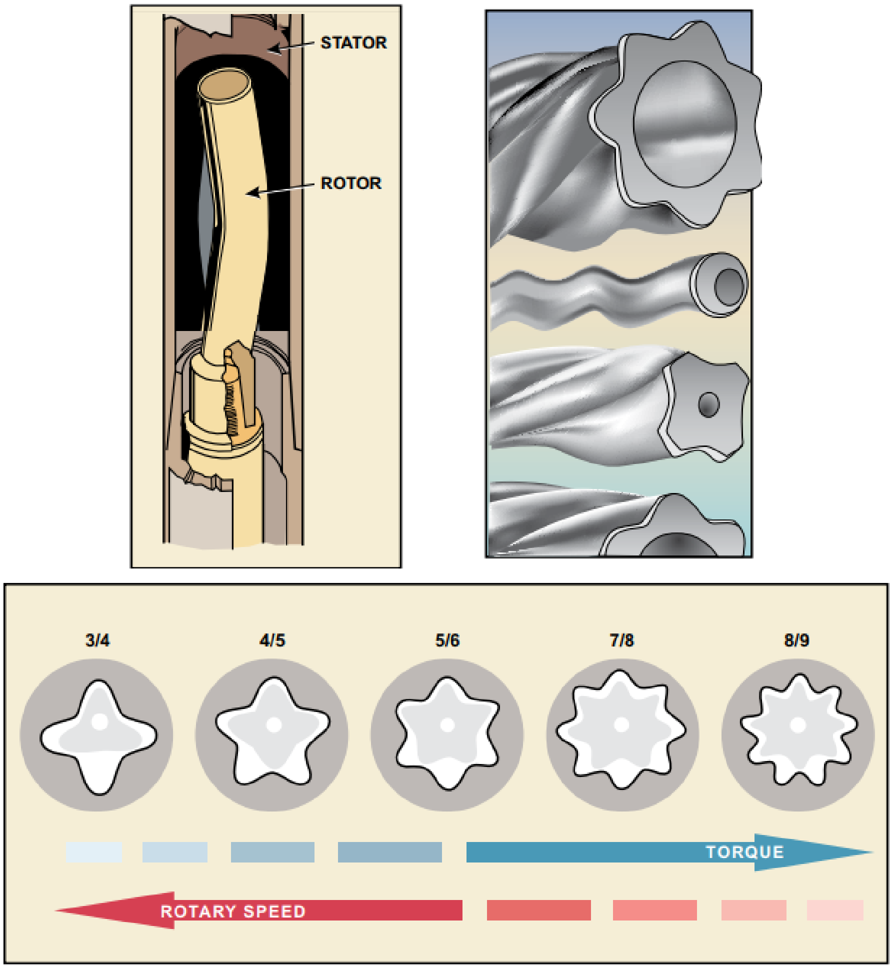


Figure 2.10. Mud Motor Assembly (top left) – Different Types of Rotors (top right) – Relationship Between Lobe Configuration and Torque and Speed Output (bottom), Adapted From [Vieira, 2009]

the drill bit in an attempt to continue drilling forward in the lateral section of the well. An example of the intermittent rotation of the drill string during the sliding operation is shown if Fig. 2.11, where it can be seen in three consecutive stands that small rotation was applied during short periods of time. The rotating drilling was performed with a top drive rotation of 55 rpm.

- Reaming: This operation is defined here simply as drilling off-bottom. Reaming is when a reamer of slightly larger diameter is installed in the BHA to enlarge the hole. But for the purpose of operation classification, the data points where the bit is pulled off bottom during drilling operations is classified as reaming. This is seen more often at the end of

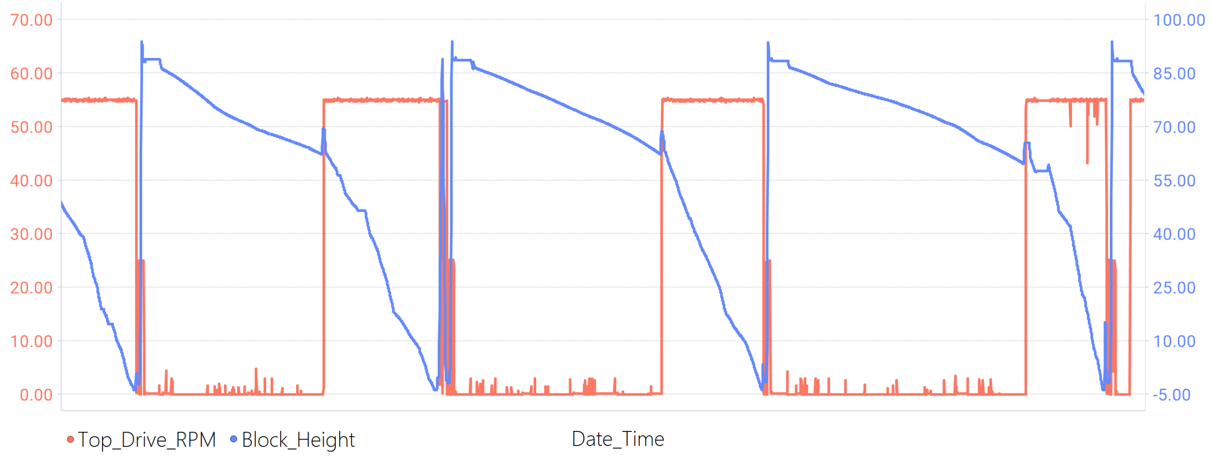


Figure 2.11. Example of Small Top Drive Rotation During Sliding Operations

the drilling of a stand, when resuming drilling after a connection in preparation for pulling out of the hole (POOH). Since this operation is not critical for classification, no threshold in the off-bottom distance and block movement columns is necessary.

-Back Reaming: Back Reaming is the reciprocal operation to Reaming, so the only difference is the direction of the block movement. Usually these two operations occur in sequence: The driller reams the well back and forth to make sure the hole is in gauge prior to trip out.

-In Slip Connection: In Slip Connection is assigned to the data points where the drill string is put on the slips. This happens more often during a drilling or tripping connection. The main indication of this operation is therefore the hook load, that can point out that the string is not being hung by the top drive with a steep drop in value. If the hook load was calibrated with the top drive and hoses installed, its value should be close to zero due to the lack of load. However, the hook load is measured indirectly by a strain gauge in the dead line, and the weight of the top drive, elevator bails, elevator, top drive hoses, etc, is not taken into account during calibration. Because of this, the values read for the hook load fluctuate around 57,000 lbf. To verify this reading, the weight of the top drive and travelling block was read from the manufacturer website in [NOV, b] and [NOV, c] and presented in Table 2.4. For hoses, elevator and other equipment, the weight is estimated.

Table 2.4. Hoisting System Element Weights

Equipment	Weight (lbs)	Weight (kg)
Top Drive TDS-11SAE	35,000	15,875
Travelling Block 650TB-500	15,725	7,133
Hoses and Elevator Assembly	6,275	2,846
Total	**57,000**	**25,855**

Table 2.4 demonstrates that the reading of around 57,000 lbf for In Slips situation is justified by the weight of all equipment of the hoisting system. Obviously, some fluctuation around this value is observed due to block movement, block movement speed, size of elevator and bails installed, mud hose elongation, if the mud hose is filled with fluid or not. [Eric et al., 2015].

The classification of the In Slip Connection is then obtained when the hook load reading is below 57,000 lbf and the bit is not touching the bottom of the well (off-bottom distance greater than 0.2 ft).

-Tripping In: Running pipe in the hole is identified when the making hole value is False, the bit is far from the hole bottom, more than one stand away, the block is moving down, and the hook load is greater than 57,000 lbf. To account for possible errors in the bit position measurements, the threshold for the off-bottom distance was reduced slightly from 90 feet (typical Range III triple stand length) to 80 feet. With the same idea, a small margin for errors in the top drive RPM and Pump Stroke-per-Minute (SPM) are accepted even though in reality, there should be no rotation or flow during tripping. A remark is made for the fact that between stands, the driller fills the well to account for the amount of drill pipe removed from the well. This operation is performed during the whole tripping operation, but since the top drive is not being moved, these data points are then captured as circulating/survey that will be described next.

-Tripping Out: Is the reciprocal of the Tripping In classification, but with the block moving upwards.

-Circulating/Survey: This operation is classified to any point where there is circulation, and the bit is not on bottom. For example, if the whole string is away from the bottom of the hole by 3 feet, with the mud pumps on, this data point is classified as circulating/survey. If the string starts to rotate, this still does not change the classification for circulating/survey. However, if later the string is moved either up or down, this action triggers the classification for Reaming or Back Reaming respectively.

-Other: The previous classification of operations covers more than 92% of the data points in a typical EDR file. These data points represent either operations not classified by the algorithm, or data points that have some sensor error that avoids its classification. Either case, these cases are simply neglected since it is a small percentage of the total number of data points in the set.

2.3.5. Correcting WOB and Differential Pressure

In an EDR system, there are usually three types of input data: some values are directly measured from a sensor or instrument; others are indirect measured or inferred from other values; and some are calibrated according to operational sequence (i.e. have to be constantly updated/zeroed). Two important values fall in the last category and deserve special attention: The WOB and mud motor differential pressure readings.

Section 1.7.1 presented how the calculation of the WOB is made, and the issues resulting from such indirect measurements. Because the EDR system has a static value for the string weight (SW) measurement, if the driller does not tare this value every time a new drill stand is added to the drill string, the system will compensate for the extra weight read in the hook load as negative weight on bit, so the equation for SW remains true. For example, in normal operation, if the EDR value for SW is set as 100 klbf, and a hook load of 95 klbf is measured using the dead line anchor sensor, then, the EDR records 100 - 95 = 5klbf as weight on bit. Now let's assume that a new drill stand weighting 2 klbf is added to the string, but the SW is not updated. Thus, when the bit is off-bottom, the hook load will hang all the weight of the string (102 klbf). But, since this value is greater than the SW, a WOB of -2 klbf is computed to be on the bit (which makes no physical sense because the bit is off-bottom). Following that, when drilling resumes and the bit touches the bottom, whatever value is being read from the WOB column will be off by the same -2 klbf. If this error propagates long enough, the off-set in WOB becomes so large that the actual weight applied to the bit is much greater than the computed value, leading to damages at the bit and inefficient drilling.

The same concept applies to the mud motor differential pressure. This value is calculated simply as the pressure difference when the bit is on-bottom minus the pressure recorded when the bit is off-bottom. Since both pressures are read in the stand pipe pressure gage (SPP), and the off-bottom pressure is called static pressure (SP), then the equation for the differential pressure DiffP becomes:

$$DiffP = P_{on-bottom} - P_{off-bottom} \tag{2.1}$$

The value of SP has to be updated every stand the same way as the SW to account for changes in the hydrostatic pressure as drilling progresses [Neufeldt et al., 2018]. In case of successive stands without zeroing, the effect in the DiffP value is the opposite to that of WOB: Since SP will not be updated (increased), the recorded value of DiffP will be greater than it really is. The effect of this error is that the motor will not run on its full capacity since the DiffP will be held to a value smaller than it could be, resulting in a smaller WOB and more likely to a slower ROP.

Combining the work of [Neufeldt et al., 2018] with [Borjas et al., 2019], an algorithm to correct for both WOB and DiifP is run on the EDR files in the data set. The first reference provided the main logic to perform the correction, whereas the second provided the actual values used to calibrate the logic. The zeroing logic starts as follows:

- A new drill stand is added to the string after a connection
- Once the connection is made up, the driller turns the mud pumps back on gradually until it reaches the stable final value. Waiting for mud flow stabilization is important because the constant rate will have a steady effect in the hook load (friction) as well

as in the DiffP (which is directly related to flow rate).

- Next, the driller resumes rotation gradually as well. Again it is important to wait for top drive rotation to stabilize for the same reasons as for the flow rate.
- The final requirement to achieve a good zeroing is to have the block moving downwards to resume drilling. [Neufeldt et al., 2018] make reference to a work where it was found that lowering the travelling block results in smaller errors when compared to raising it, thus downward movement is favored when zeroing for WOB and DiffP.

Fig. 2.12 shows the logic in the algorithm, and Fig. 2.13 depicts the same logic using a sample of EDR data set. Note in Fig. 2.12 that a verification is set for the cases where the bit is not held off-bottom long enough before all the parameters are met. If that happens, drilling will resume, and then it will avoid the zeroing in such non-ideal conditions until the next stand is drilled.

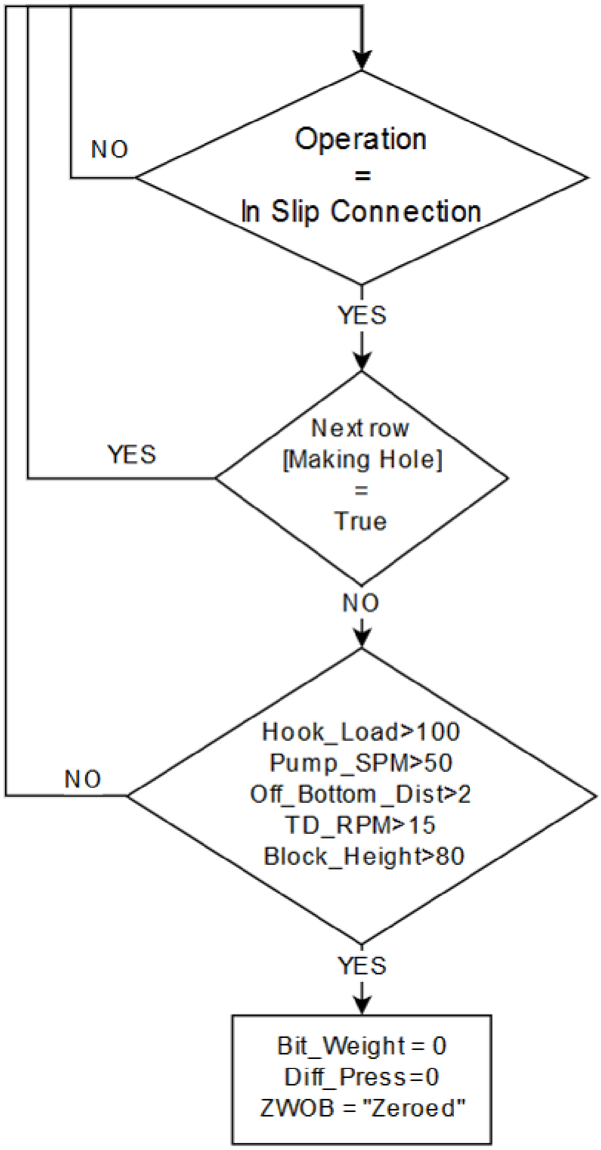


Figure 2.12. Flowchart Representing the Zero WOB and DiffP Logic

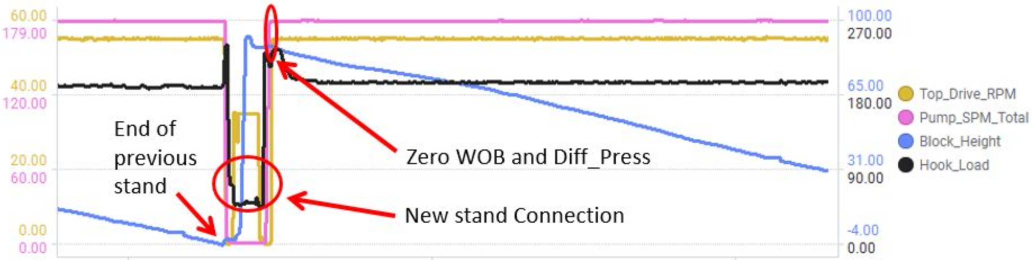


Figure 2.13. Demonstration of the Ideal Moment to Zero WOB and DiffP

2.3.6. Removing Outliers

The last step in the data wrangling process is the removal of outliers, that can be easily identified in box plots. In a box plot, the data is divided into five intervals: minimum, first quartile (Q1), median, third quartile (Q3) and maximum. The definition of this division explained by logical definition (not in order) is:

- Median: The middle value of the data set. (not the mean)
- First Quartile (Q1 or 25th percentile): the middle number between the smallest number (not the minimum) and the median of the data set.
- Third Quartile (Q3 or 75th percentile): the middle number between the highest number (not the maximum) and the median of the data set.
- Interquartile Range (IQR): The range from the 25th to the 75th percentile.
- Minimum: Q1 - 1.5 x IQR
- Maximum: Q3 + 1.5 x IQR

Following the definition above, outliers are defined as any data points that are outside the range of 1.5 times the interquartile range. Fig.2.14 depicts a box plot with its definitions, as well as a comparison with a probability density function (PDF) for a normal distribution.

It can be seen from Fig. 2.14 that considering outliers, the values outside 1.5 times the IQR, will keep 2.698 standard deviations from the mean of the data, or 99.3 % of it. This means that using this criteria will remove a very small amount of the data, still not compromising its distribution.

During the previous steps of the data cleaning process, many bad values were automatically excluded (for example, a bad data point of 100,000 ft/hr ROP at the beginning of the data set when the EDR was being set up). Because data points like this one could skew the calculation of the IQR range, this step is performed as the last one of the data

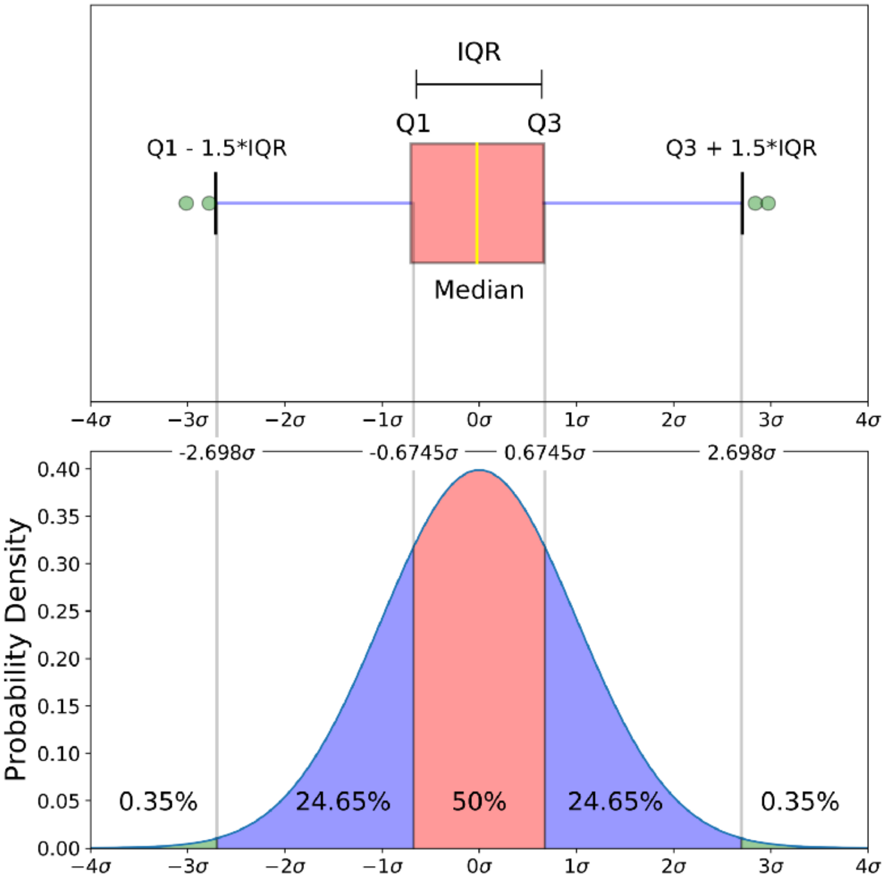


Figure 2.14. Comparison of a Box Plot of a Nearly Normal Distribution and a PDF for a Normal Distribution, From [Galarnyk, 2018]

wrangling process. To demonstrate this fact, box plots of several values are plotted below in three steps: As raw data, just after the last wrangling process (step described in 2.3.5) and the same cleaned values but filtered only for data point where drilling is happening (meaning where the drilling operation is classified either as ”Drilling”, ”Drilling/Rocking” or ”Sliding”). The reason for filtering the cleaned data point for drilling only is because it was noticed that some values can be consistently bad in other operations, for example the gamma ray values spike when the operation is circulating, or the ROP - Instant spikes during tripping operations.

2.3.7. Bit Weight

It can be seen from Fig. 2.15 that the problem with the WOB still remains after the null values are cleaned. This is mainly due to the problems already discussed of lack of

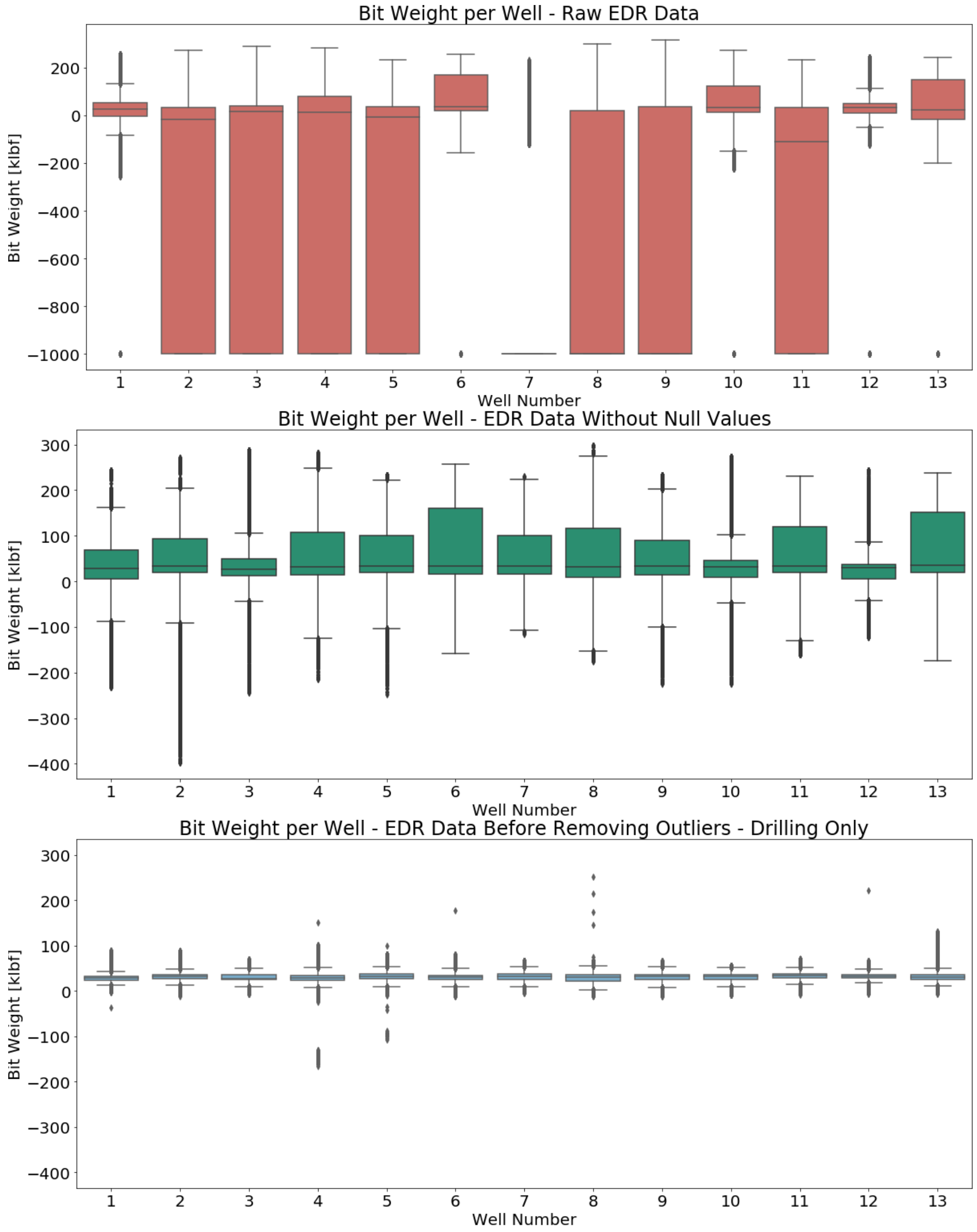


Figure 2.15. Box Plots of Bit Weight using Raw Data, Without Nulls, and Drilling Operations Only

zeroing, and affects mainly when the bit is not on bottom. When we filter the data for drilling operations only (drilling, sliding and rocking) we can see that the variance reduces significantly, with most of the data above zero and below 100,000 lbf.

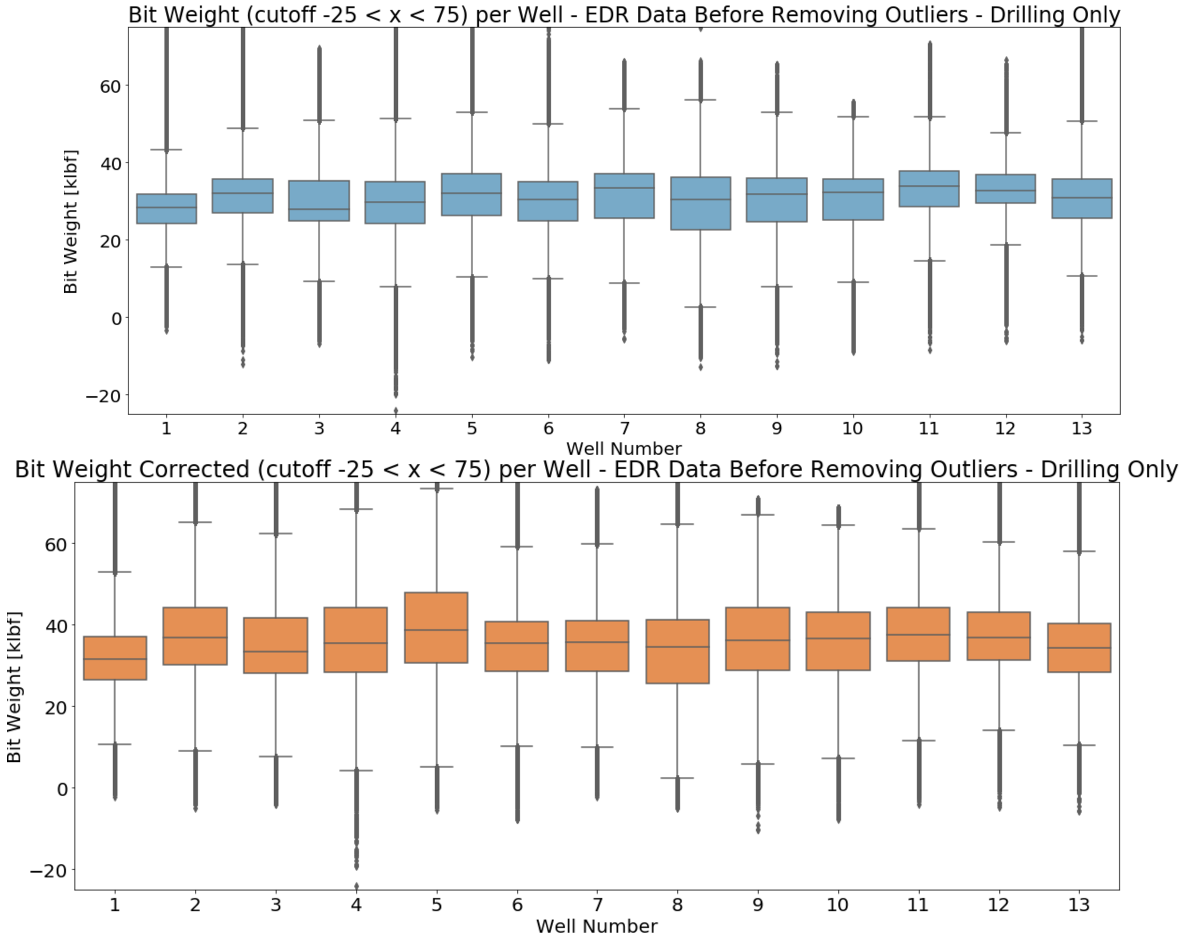


Figure 2.16. Bit Weight and Bit Weight Corrected Zoomed Between -25 and 75 klbf

Fig. 2.16 shows the expected overall increase with the correction of the WOB due to the algorithm implementation. However, it is still noted some outliers of negative values, specially in well number 4, that need further investigation.

After analyzing Figs. 2.15 and 2.16, it was decided not to remove any outliers artificially, because the very few occurrences observed in the bottom plot of Fig. 2.15, reveal that there are a very few high values, in ranges that still might be actual true values of some operational problem (such as sudden release of weight on the bit due to stuck pipe or bit stalling). Further investigation will be performed in the data analysis section of this work to understand these data points.

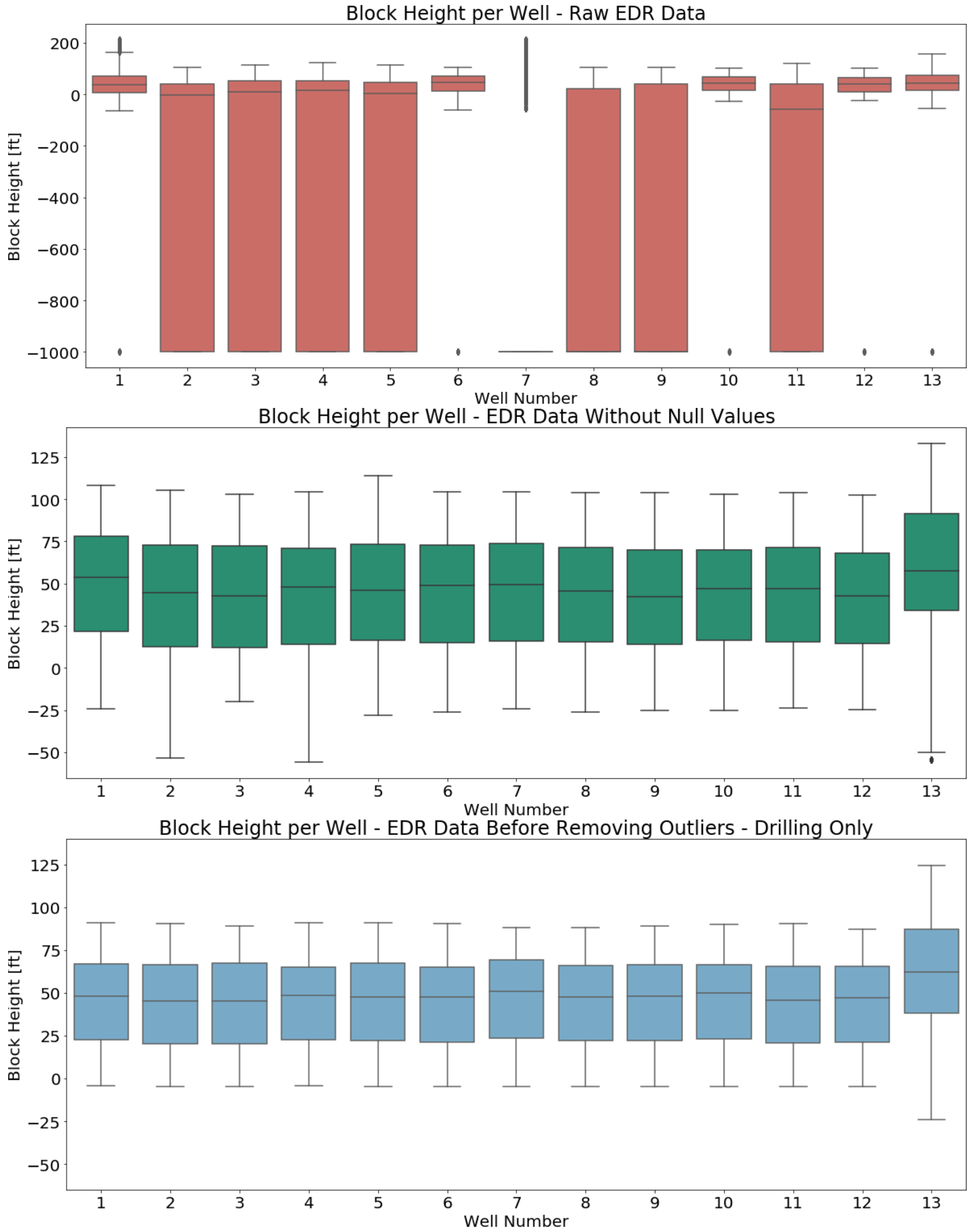


Figure 2.17. Box Plots of Block Height Using Raw Data, Without Nulls, and Drilling Operations Only

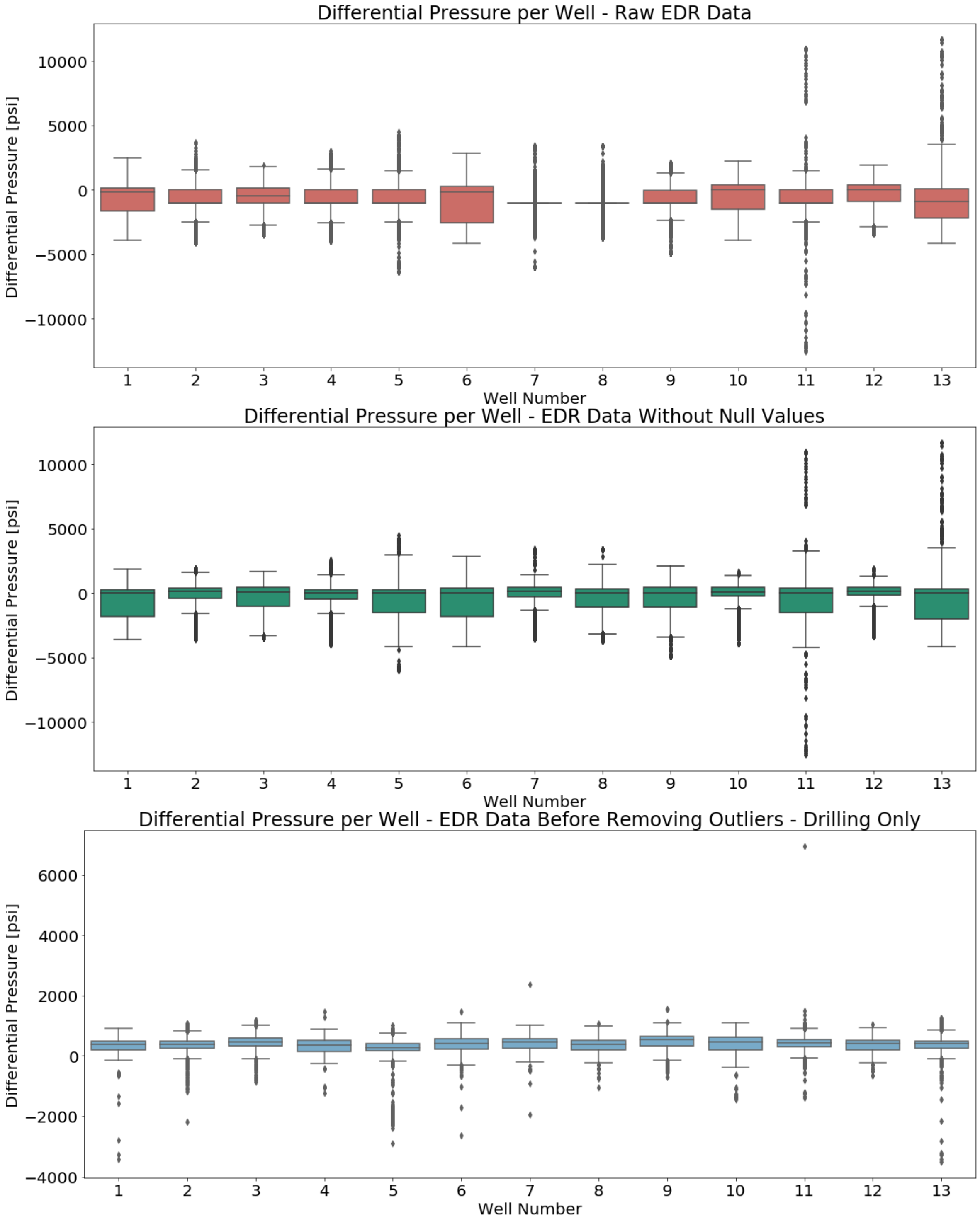
2.3.8. Block Height

An analysis of the Block Height values shows no problems with outliers. The last well presented errors in the calibration of the sensors, recording values in a wider range than really exist (the derrick has a fix value since the rig used to drill the wells is the same).

Another important observation is that most of the very negative values are removed by simply filtering the data set for drilling values only. A possible explanation for this issue is that the sensor is more likely to perform poorly during tripping operations, in which the top drive moves at much higher speeds than during drilling. Even though, for all the wells it is observed that the actual zero in the block height occurs in values below zero, being close to -4 ft. This does not pose any harm for human operated top drives, but is a problem that will have to be addressed more seriously when moving to automated operations.

2.3.9. Mud Motor Differential Pressure

Regarding the mud motor differential pressure, filtering for drilling operations only results in most of the negative values to be removed from the analysis (Fig. 2.18). This is good, and once again confirms the issues with the zeroing operation described earlier in 2.3.5. When the box plots for the differential pressure and corrected differential pressure are zoomed between -2000 and 4000 psi (Fig. 2.19), the variance of the data increased considerably for the corrected values. This is not the behavior expected with the zeroing algorithm, therefore it is an indicator that the zeroing routine for DiffP should be reviewed. A possible cause for this issue is the zeroing point selected not being at the final pressure state. Again, further investigation in the data analysis will be performed to answer this question. It is also noted however that for some of the wells (Numbers 12 and 13 for example) the zeroing routine seems to be working, i.e., not increasing data variance and also reducing the overall values of the variable, as expected by the problem with no zeroing DiffP.


Differential Pressure per Well - Raw EDR Data

Differential Pressure per Well - EDR Data Without Null Values

Differential Pressure per Well - EDR Data Before Removing Outliers - Drilling Only

Figure 2.18. Box Plots of Differential Pressure using Raw Data, Without Nulls, and Drilling Operations Only

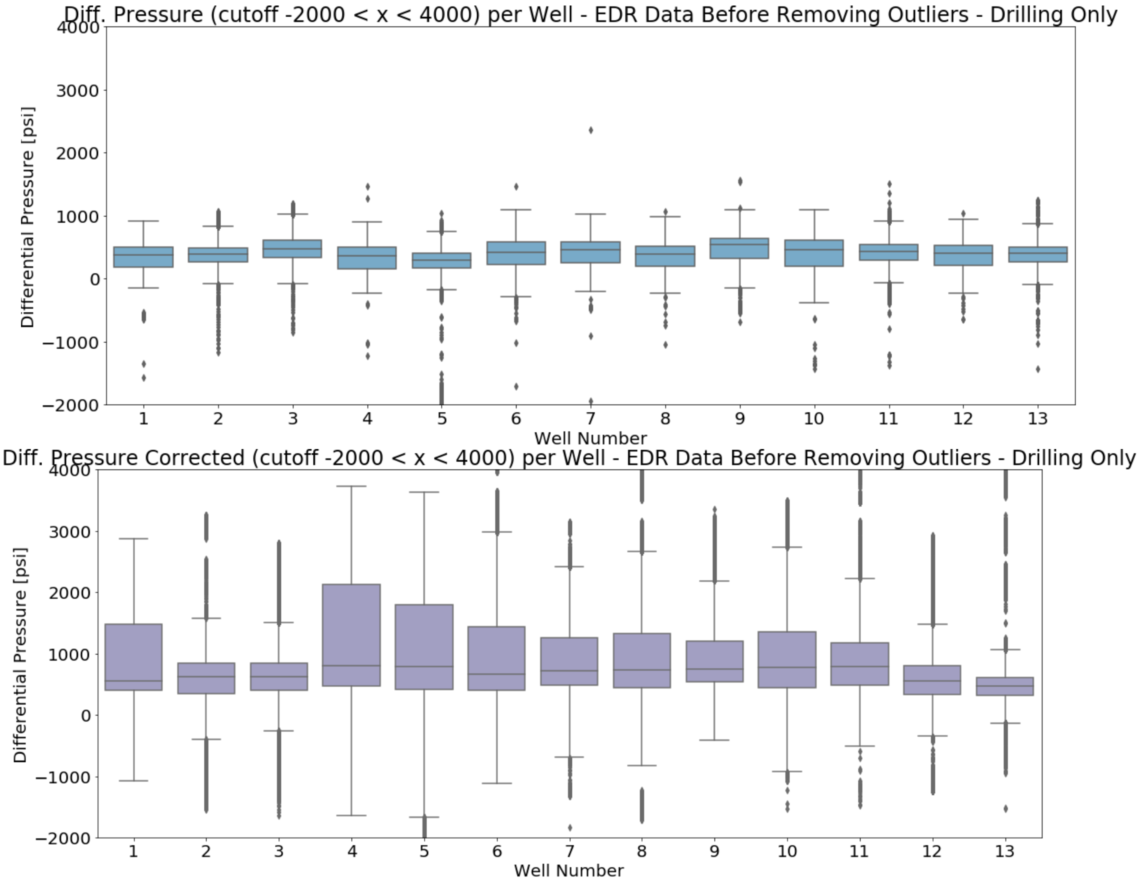


Figure 2.19. Box Plots of Differential Pressure and Differential Pressure Corrected Zoomed Between -2000 and 4000 psi

2.3.10. Gamma Ray

The gamma ray box plots show that most values are within the expected range (0 to 120°API.) after performing the null values cleaning, filtering for drilling points and zooming in. The only well that presented negative values is Well 7, with this issue being most likely an isolated recording issues. A comparison between the second and third plot shows that some outliers with very high readings disappear when filtered for drilling operations. But in the third plot it is still possible to see a consistent appearance of very high values for gamma ray even after the filtering. This phenomena is understood as a tool problem and thus all values equal to or above 200°API where replaced by 200°API to remove this issue. The value of 200°API was chosen because it still a realistic high value for gamma ray readings, thus still providing information about the high radioactivity of the formation.

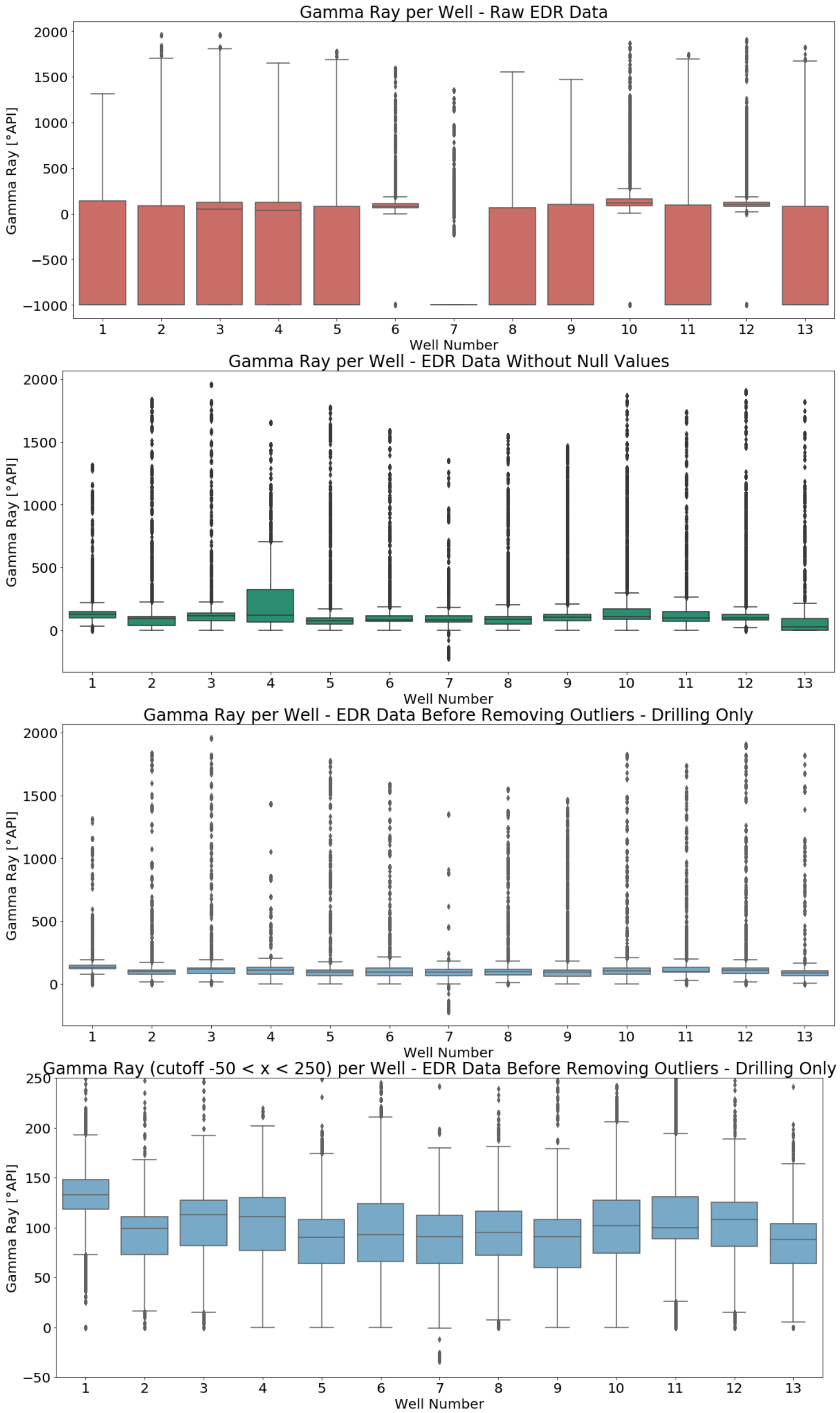


Figure 2.20. Box Plots of Gamma Ray Using Raw Data, Without Nulls, Drilling Operations Only, and Zoomed Between -50 and 250 °API

2.3.11. Hook Load

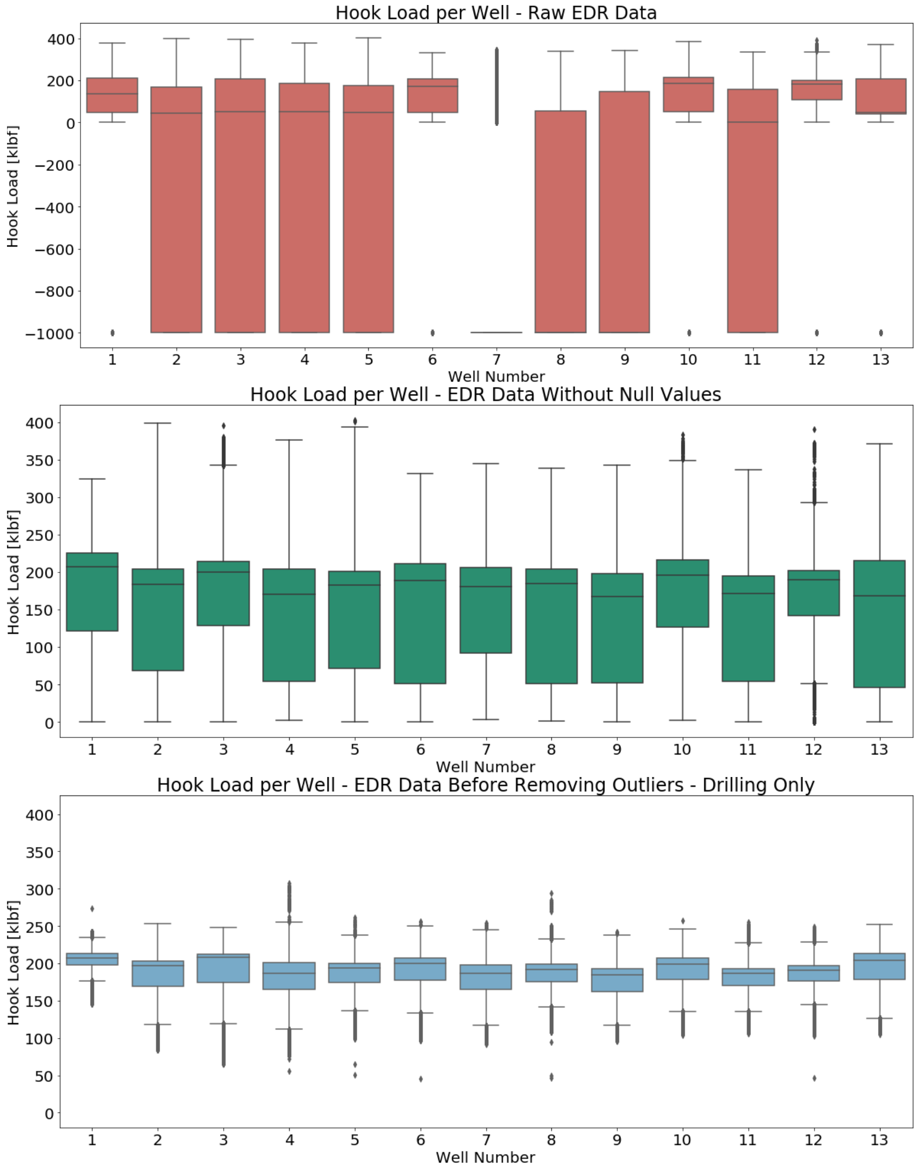


Figure 2.21. Box Plots of Hook Load Using Raw Data, Without Nulls, and Drilling Operations Only

The hook load values plotted present a consistent behavior for all wells once cleaned and filtered on drilling only data points. In the second box plot it is observed a decrease in the Q1 position, being for most of the wells, close to the In Slips value (around 57 klbf). This behavior might be explained if the overall tripping performance was slower, or more time was spent on surface (in between trips). Regarding the data quality of the values, the last plot shows no signs of problems with the measurements.

2.3.12. Pump Pressure

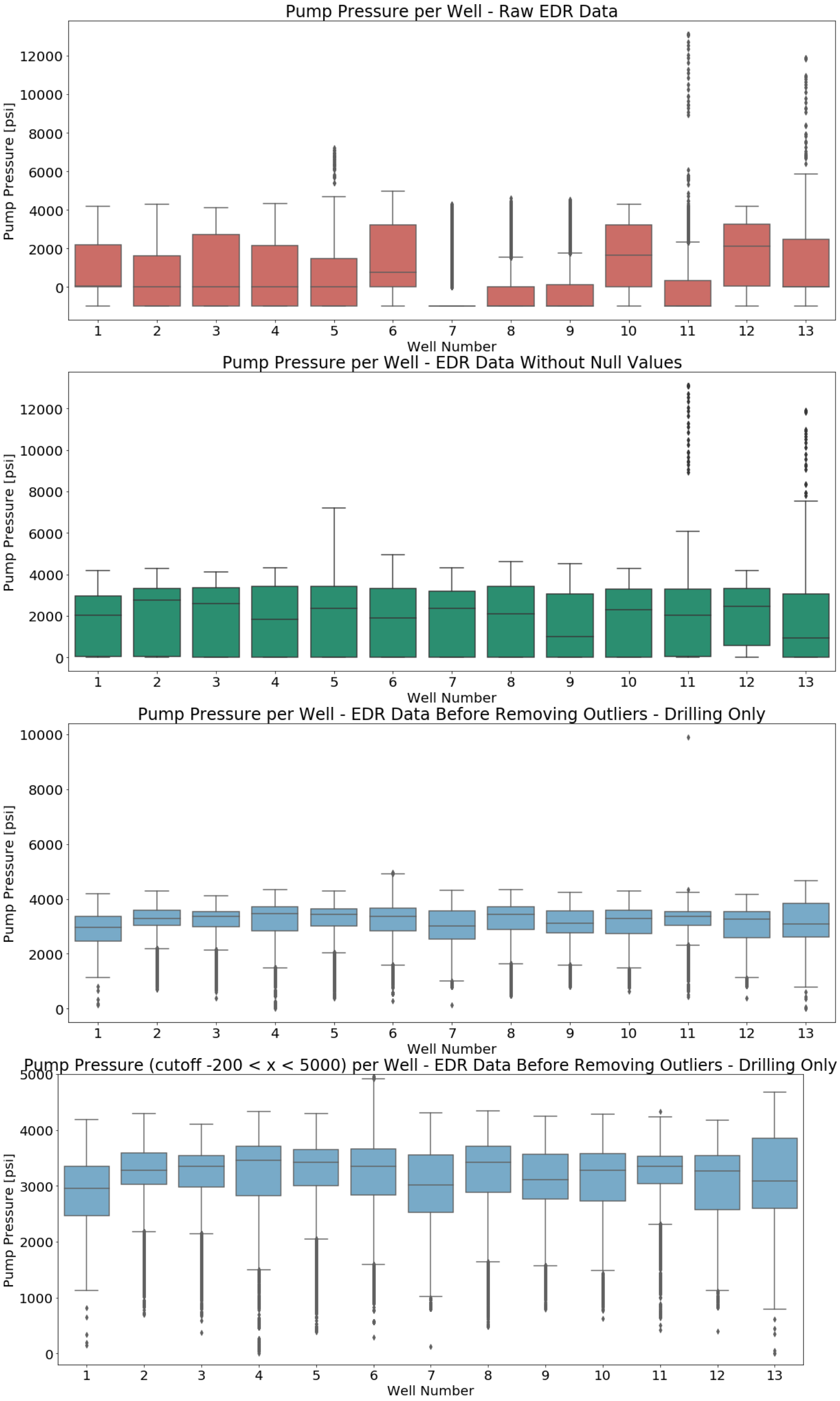


Figure 2.22. Box Plots of Pump Pressure Using Raw Data, Without Nulls, Drilling Operations Only, and Zoomed Between -220 and 5000 psi

The pump pressure, or stand pipe pressure values also do not present evident inconsistencies or problems in measurements. Since all wells are drilled up to similar measured depths, it is expected the pump pressure (hydraulics) to be similar as well. The average value when drilling is being around 3500 psi, with some wells having the data skewed towards smaller values. From Fig 2.22 it is concluded that no further treatment to this variable is necessary.

2.3.13. ROP - Average

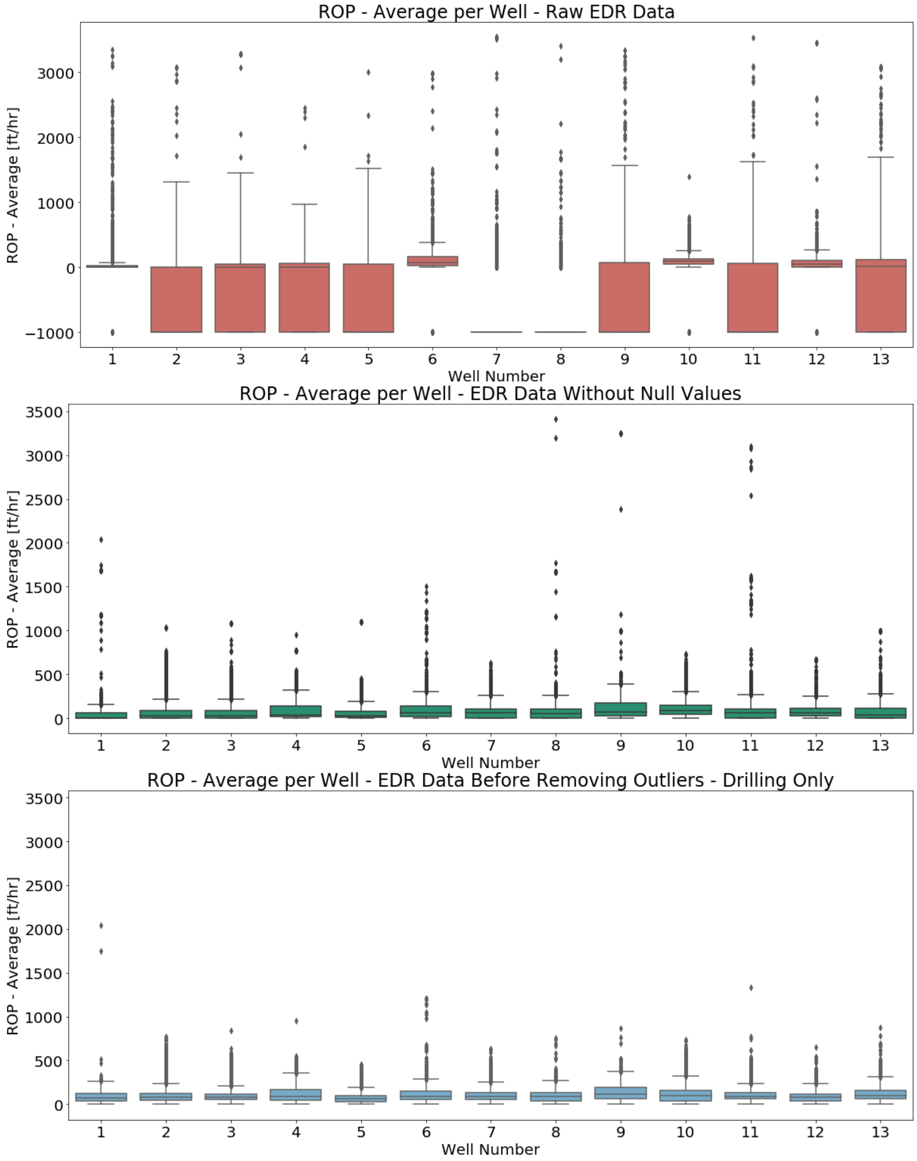


Figure 2.23. Box Plots of Average ROP Using Raw Data, Without Nulls, and Drilling Operations Only

The rate of penetration is the most important parameter in drilling analysis, since it represents the rate in which the bit advances, that can be translated to efficiency to drill. The calculation of ROP is therefore of utmost importance, since the aggregation of wrong values can result in completely wrong values and or conclusions. For this reading, the EDR calculates the average of the last 5 seconds. Fig. 2.24 shows that an increase in values for ROP by just filtering the data to drilling only operations. Since it seems that all of the wells' data show consistent outliers from 300 to 800 ft/hr, a detailed analysis may be conducted in sequence.

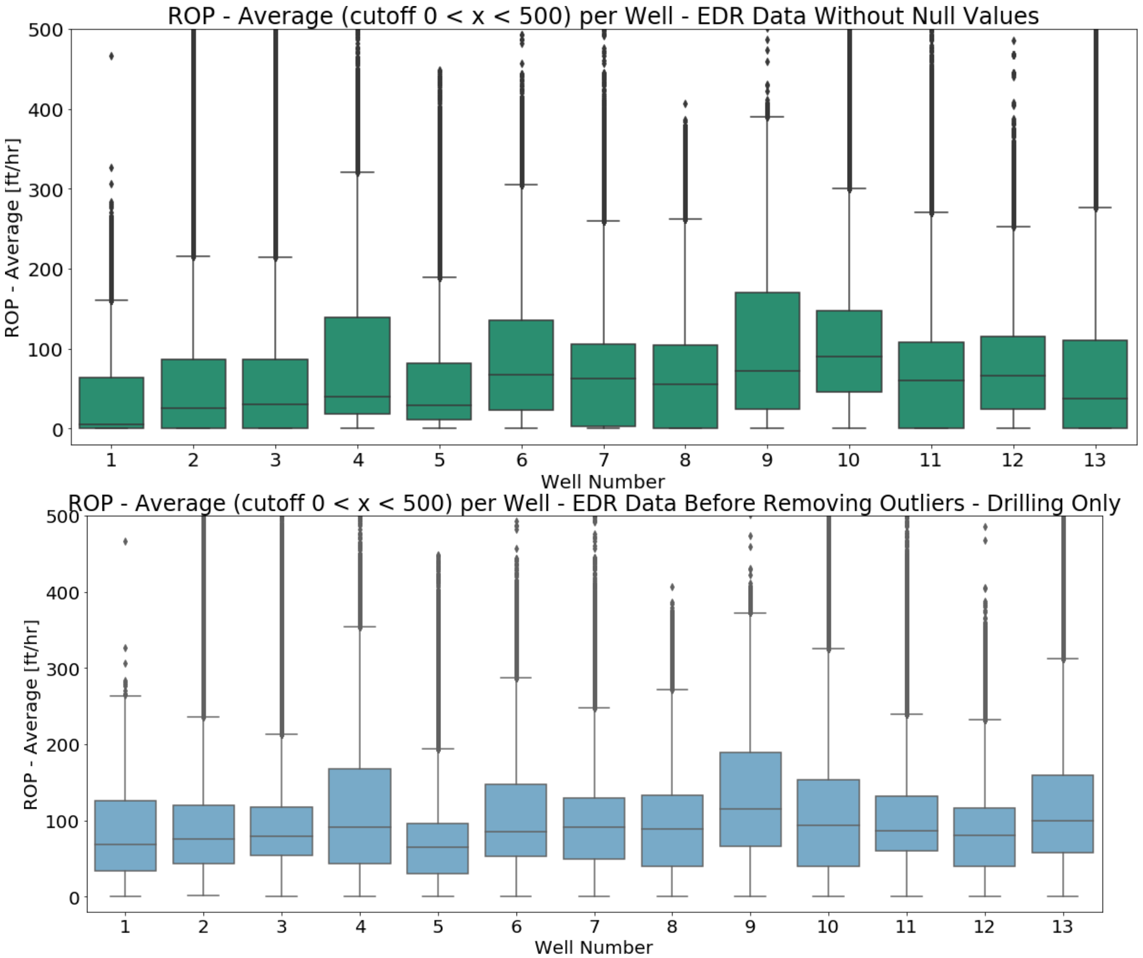


Figure 2.24. Box Plots of Average ROP for Raw Data and Cleaned Drilling-only Values Zoomed Between 0 and 500 ft/hr

2.3.14. Top Drive Rotation

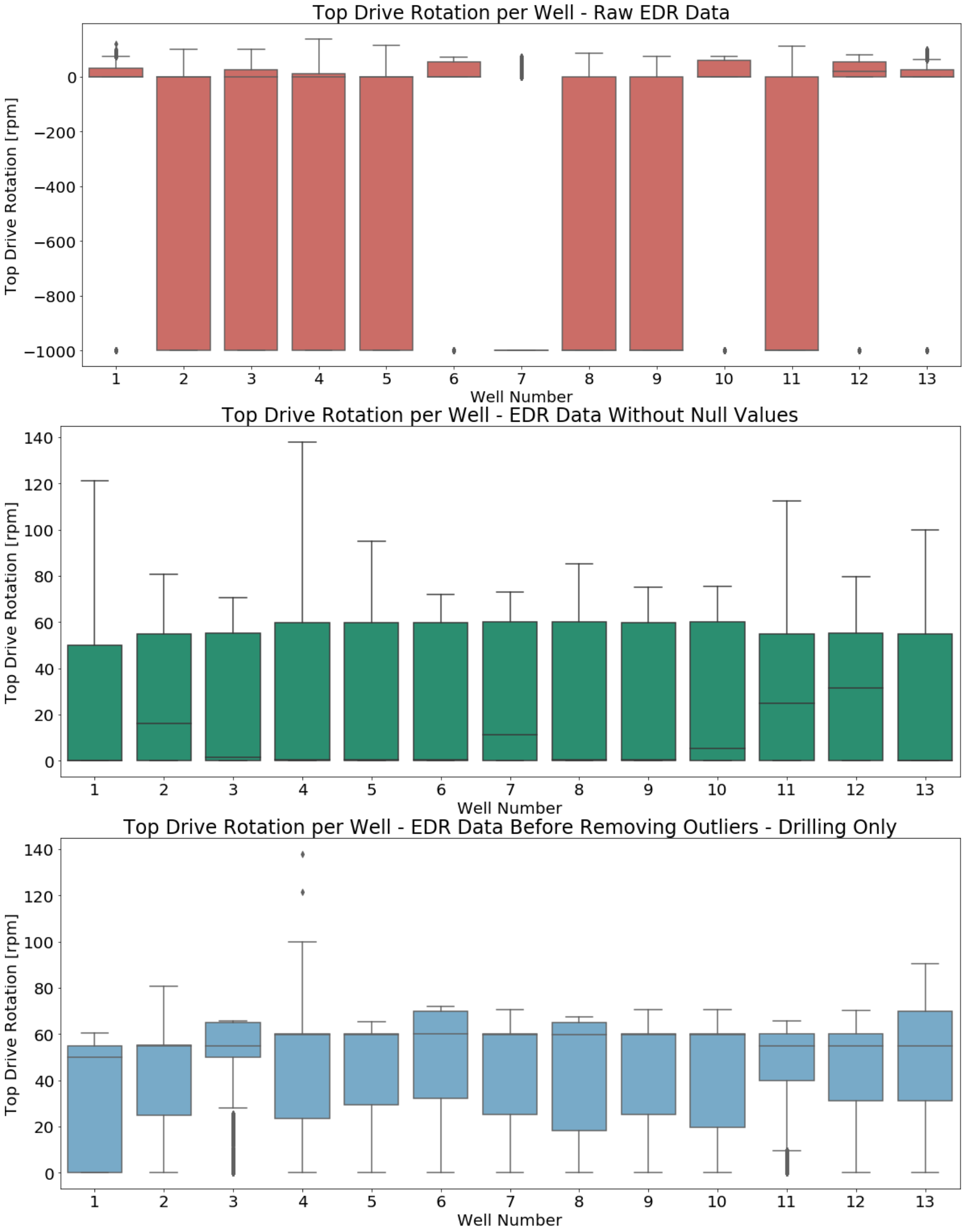


Figure 2.25. Box Plots of Top Drive Rotation Using Raw Data, Without Nulls, and Drilling Operations Only

The top drive rotation presents consistent measurements once null values are cleaned. No negative values, and most of the data for all the wells is between 0 and 60 rpm, which is the interval where most of the operations occur. For the drilling only values (third plot) the most upper limit is still at 60 rpm, however for some of the wells there are more data points pushing the distribution towards 25 rpm approximately. This could be an indication of more rocking for these wells.

2.3.15. Top Drive Torque

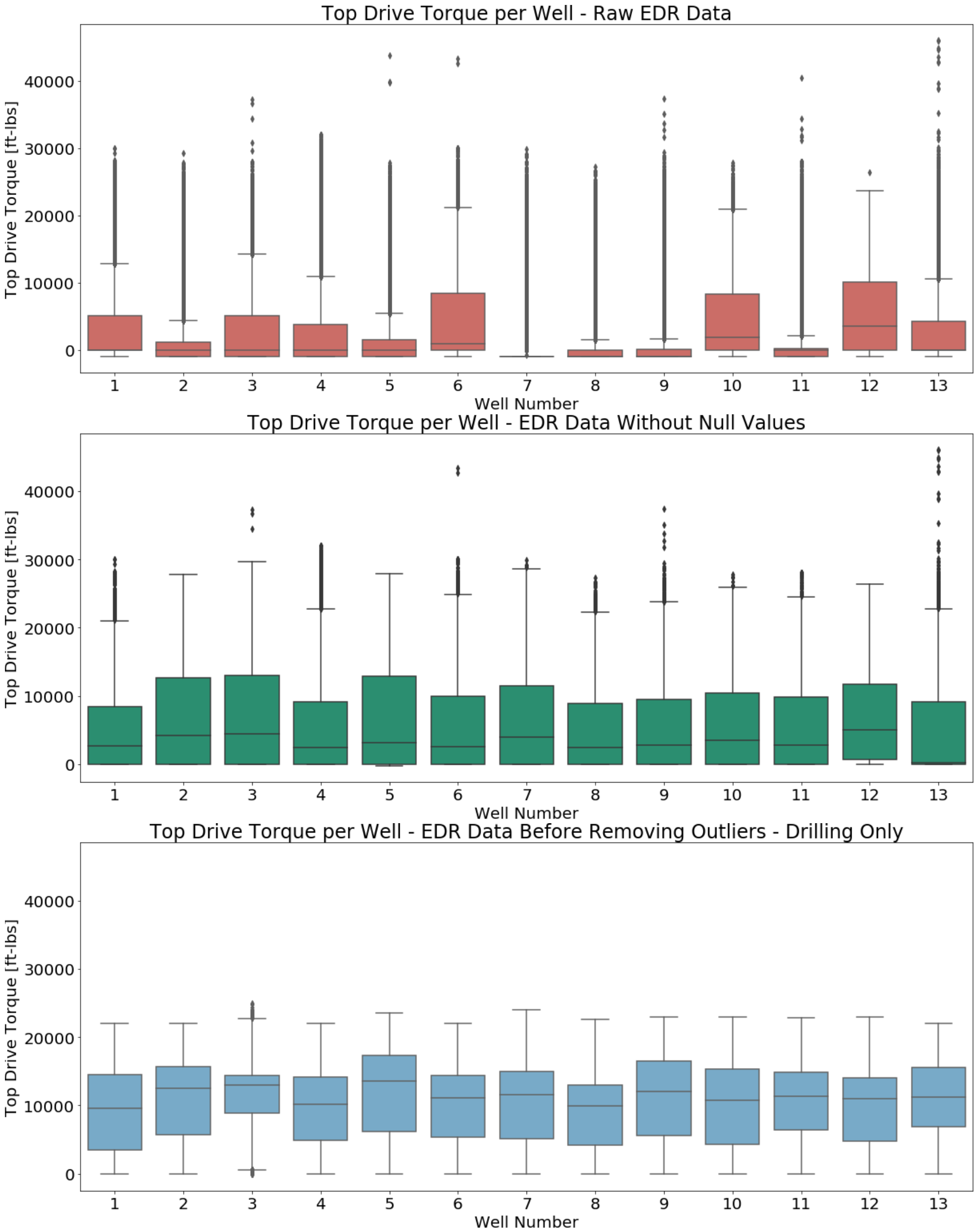


Figure 2.26. Box Plots of Top Drive Torque Using Raw Data, Without Nulls, and Drilling Operations Only

Torque measurements recorded in the thirteen wells from the data set are consistent and in accordance with the expected physical values for mud motors of 8.5 inches of diameter (but most BHAs are of this size). For the drilling only plot, only one well presented some outliers, meaning that there is no reason to believe that there are problems with this variable, so no further edits will be performed.

2.3.16. Flow Rate In

When plotting the flow rate values, a very small quantity of outliers dominated the plot, even when cleaned and filtered for drilling values only, as seen in Fig. 2.27.

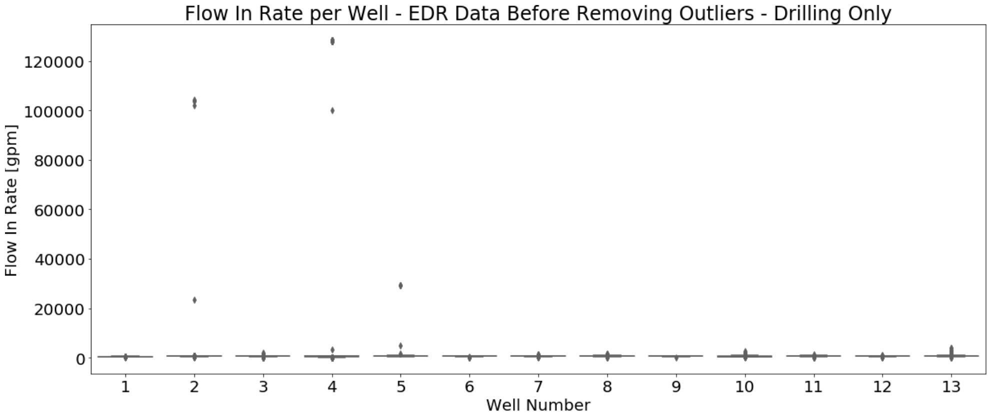


Figure 2.27. Box Plot of Flow Rate In for Data Without Nulls and Filtered for Drilling Operations Only

That way, all data values above 1600 gpm are cut from the plot, to allow interpretation. After the first cleaning routine, most of the flow rate values lie between zero and 600 gpm, which seems to be reasonable according to the pump specifications for a well of such depth. When the values are filtered for drilling only, the values are now concentrated in a higher interval - 600 to 800 gpm - indicating that the pumps are on at sufficient values to promote hole cleaning. In at least four wells however (3, 8, 10 13) some outliers show flow rates up to 1200-1400 gpm, but these values look unrealistic, especially for such a small period of time. Because the flow rate is calculated by the EDR using the pump rate (strokes per minute), liner size, stroke length and piston rod diameter, the most likely explanation for such values is that the pump rate for these points is off, resulting in an equally off calculation of the flow rate. This will be evaluated in the next column (pump rate).

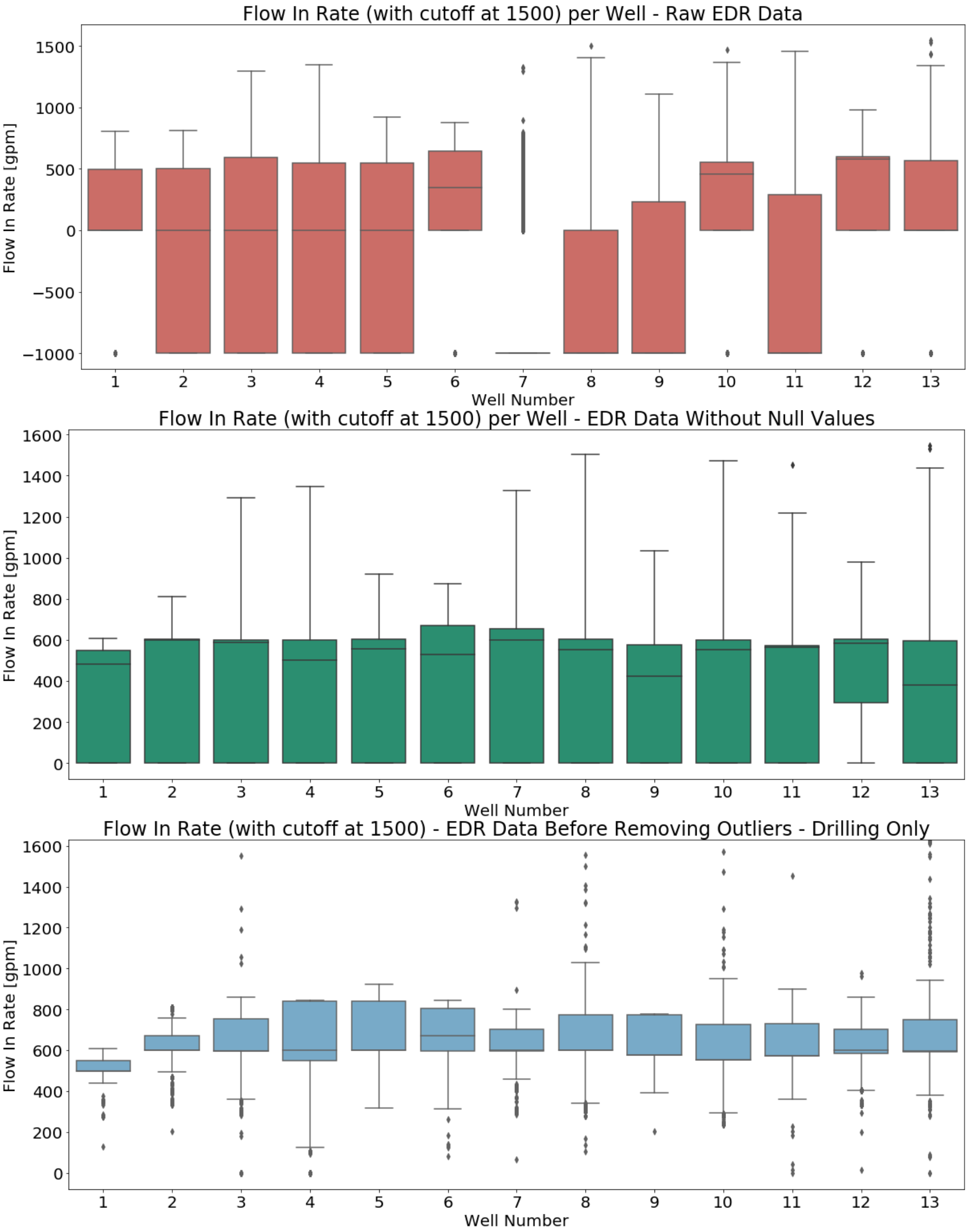


Figure 2.28. Box Plots of Flow Rate In Using Raw Data, Without Nulls, and Drilling Operations Only, With a Cutoff at 1600 gpm

2.3.17. Pump Rate

The pump rate is the count of pump speed in strokes per minute, usually measured by a micro limit switch sensor installed in one of the pistons of the pump. The sensor is actuated when the piston touches the sensor's rod, increasing the count of mud pump strokes.

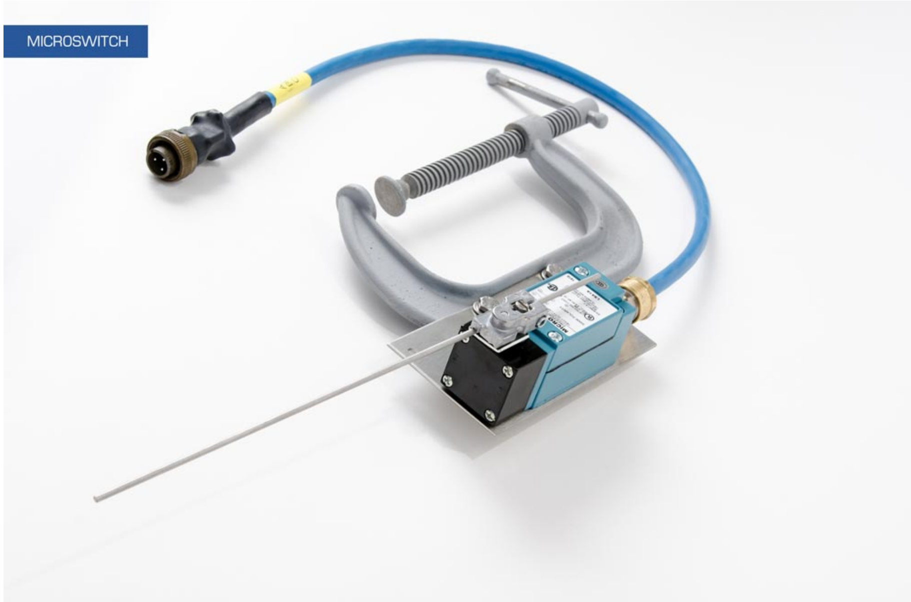


Figure 2.29. Example of Micro Limit Switch, From HDI Gauges

Since these sensors count a stroke every time it detects movement, this measure is susceptible to ”false” readings. Because the mud pump rod seals frequently present some leaking, splashing of the mud due to the piston movement could trigger the stroke counter to read before the actual piston completes a full stroke. Although the sensor is designed to withstand equipment vibration, unusual movement of the pump could also trigger false strokes. This false reading then makes the stroke per minute calculator compute an extremely high value for the pump rate.

When we compare the last plots of Fig. 2.28 and 2.30 it seems that the outliers for the pump rate indeed caused the EDR to calculate equally high flow rates in another column. According to a mud pump manufacturer [NOV, a], a pump of similar specification has a maximum pump speed of 120 spm. Since the rig used in this work has three triplex pumps and the channel feeding the EDR system calculates the sum of all pump speeds (Pump SPM - Total), it is known that any value greater than 360 spm is due to faulty measurements. Because of this reason, the values of this column will be all set to a maximum of 360 spm. Fig. 2.31 shows that this assumption is reasonable since most values lie below the threshold applied.

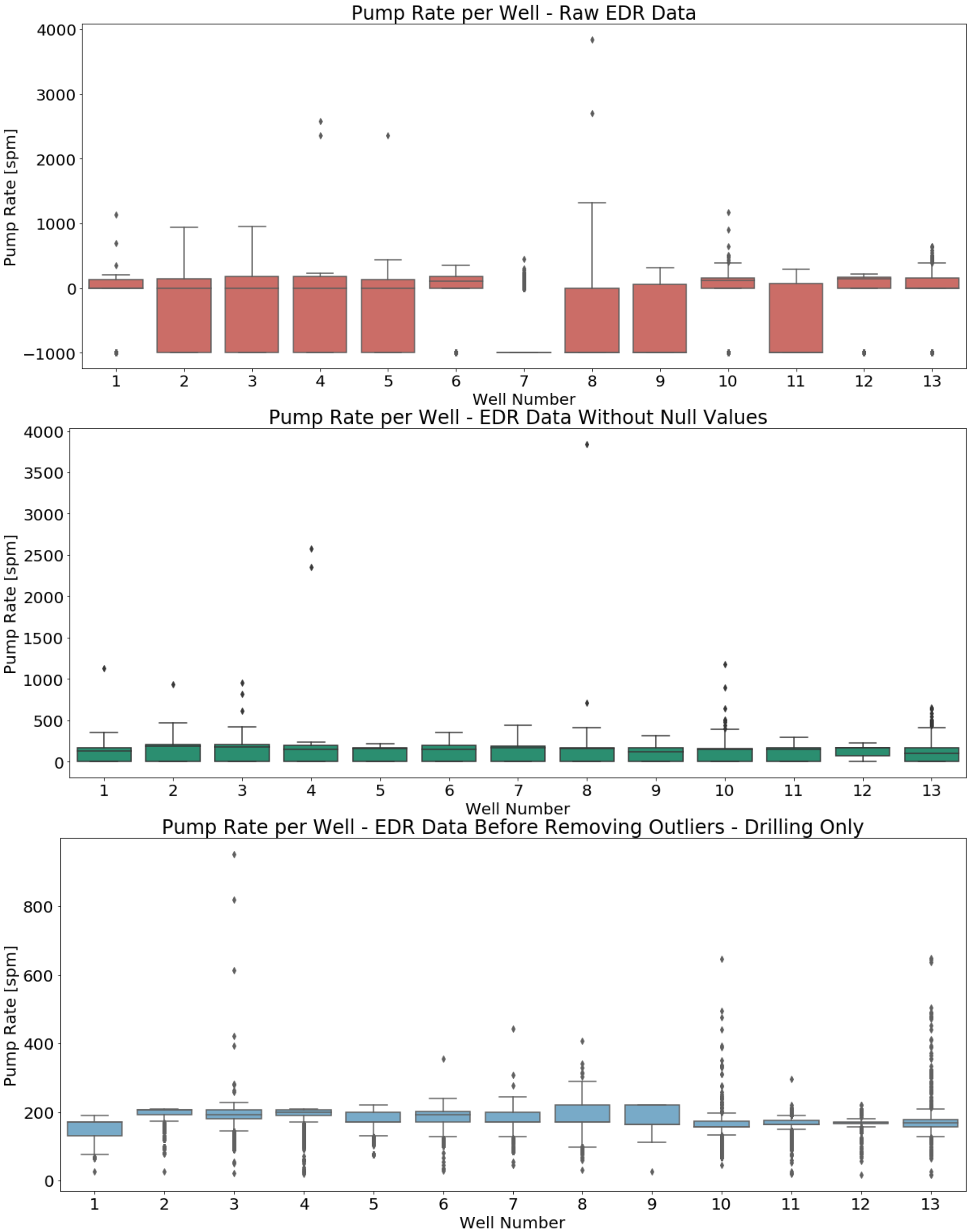


Figure 2.30. Box Plots of Pump Rate Using Raw Data, Without Nulls, and Drilling Operations Only

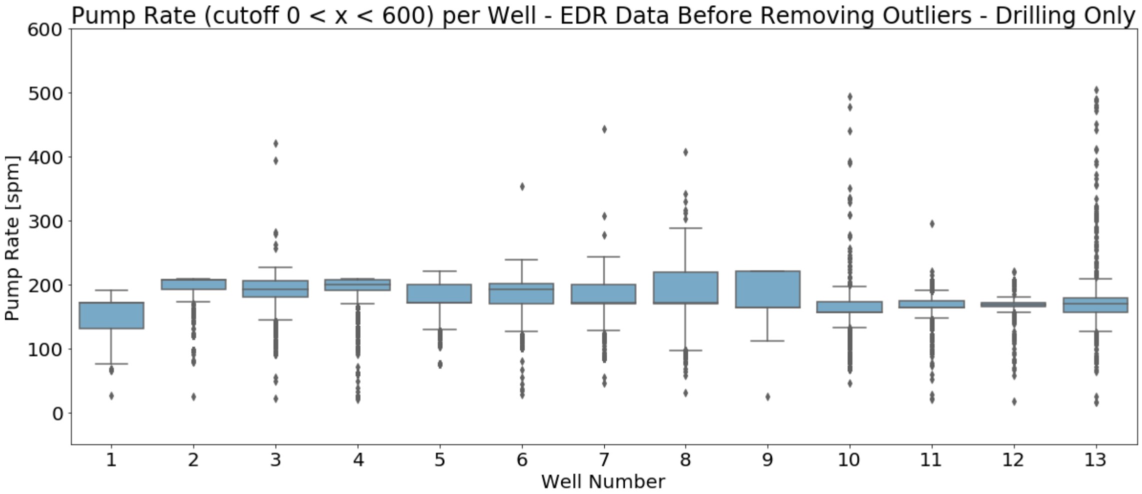


Figure 2.31. Box Plot of Pump Rate Using Data Without Nulls Filtered to Drilling Operations Only and Zoomed to Values Between 0 and 600 spm

2.3.18. Instantaneous ROP

The instantaneous ROP column (ROP - Fast) is the column name given to the top drive velocity calculation. It is measured by dividing the displacement of the top drive (using the block height measurement) over time. This calculated column is sample dependent because it is updated only once every 5 seconds. So, if the same physical movement of the top drive begins at a x:xx:00 time or xx:xx:03, the value for the ”ROP - Fast” is likely to be different. Also, for tripping operations where the block is moving at much faster speeds, this effect of sampling can result in even higher positive or negative ”ROP - Fast” recordings.

This issue is evident in Fig. 2.32 which shows instantaneous velocity of positive 400,000 ft/hr and negative of up to 1,000,000 ft/hr. The third plot shows that the negative effect is removed for drilling only operations, because in drilling the top drive is always moving downwards (ROP positive). It is also noted that this filter removed most of the very high values, since no tripping data points are included in the analysis now. However, the existence of these outliers even in drilling is attributed to bad readings in the block height, already discussed in Section 2.3.1 when dealing with bit position values greater than the hole depth. Because the reliability of this column is compromised, which in turn is useful only for qualitative analysis, these outliers are left as is. No drilling optimization

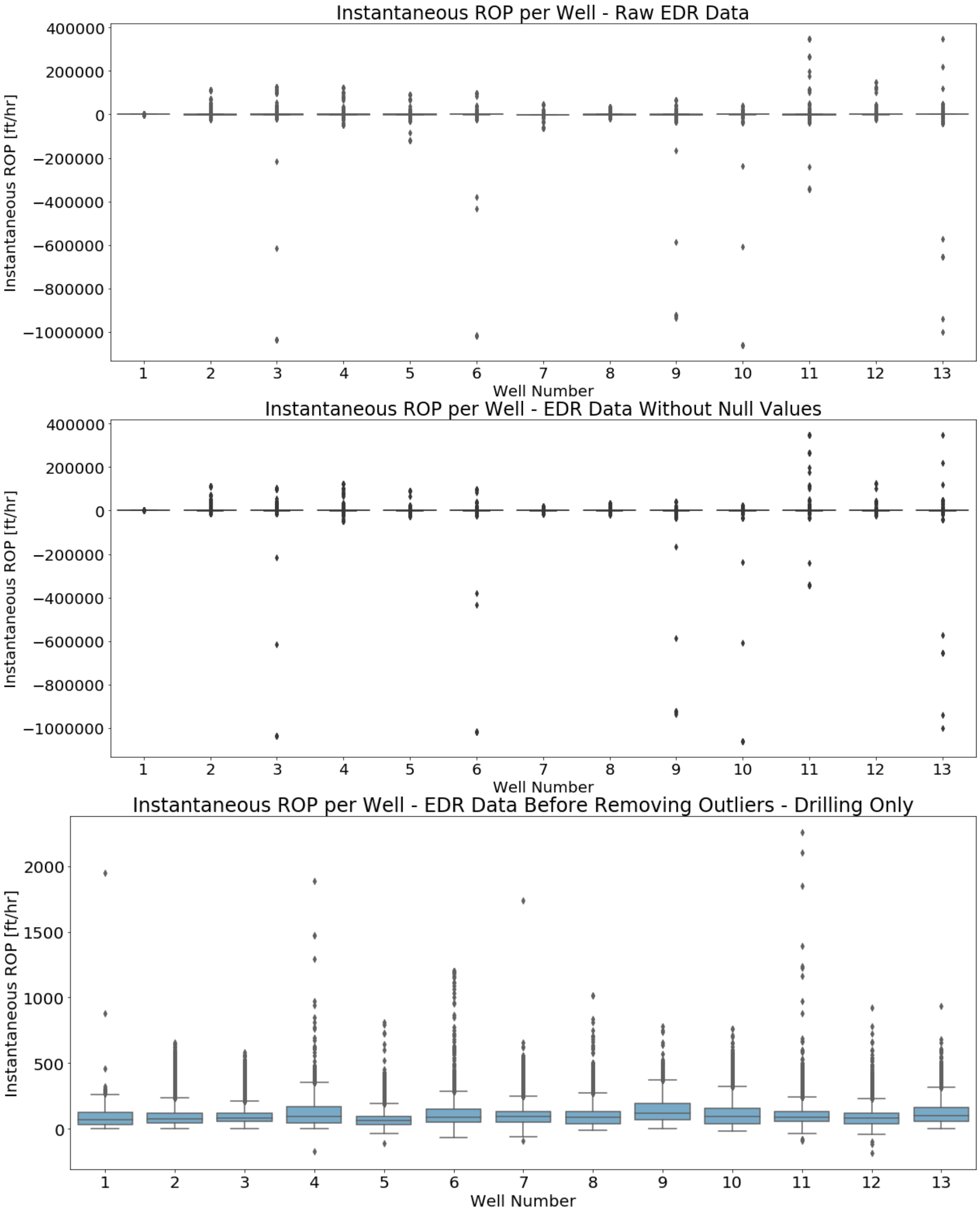


Figure 2.32. Box Plots of Instantaneous ROP Using Raw Data, Without Nulls, and Drilling Operations Only

conclusion is seen to be derived from the evaluation of this column's values, besides using it as a guideline by the driller during actual drilling. A zoom in for instantaneous ROP values below 500 ft/hr is presented in Fig. 2.33. Now it is visible that most instantaneous

ROP plots look similar to the ones in Fig. 2.24. In the zoomed plot some negative values can be seen, confirming that misreadings of block position and block movement occur in the EDR.

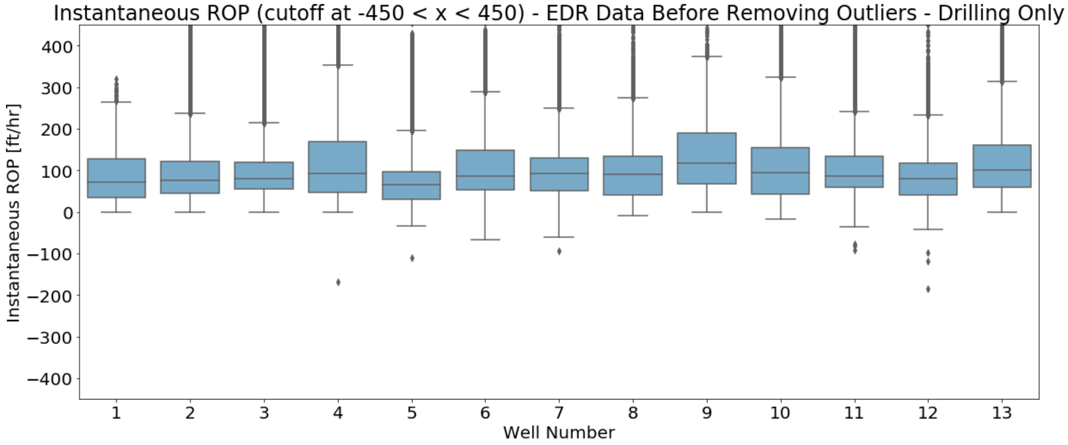


Figure 2.33. Box Plot of Instantaneous ROP Using Data Without Nulls Filtered With Drilling Operations Only, Zoomed to Values Between -450 and +450 ft/hr

The conclusion of individual analysis of all drilling parameters collected from the EDR result in only two values to be artificially corrected:

- Gamma Ray: The ”Gamma Ray” column is limited to 200 °API.

- Pump Rate: The ”Pump SPM - Total” column is limited to 320 gpm.

This fact does not mean that other columns do not need some sort of fixing. As discussed along the individual sessions, some variables affect the drilling operation more direct than others, and thus for these further investigation is done before changing its values.

Chapter 3
Data Visualization and Analysis

3.1. Data Analysis Overview

Data analysis is the process of extracting meaning from data and using it to improve a process or activity. In this work, drilling data of previously drilled wells is being used to provide insights on the performance of future wells. In the previous chapter, a complete description of the cleaning process for drilling data was proposed, based on a mixture of the experience of the analysts and published work in the literature. This resulted in a method that was not yet described in the literature with the same degree of detail.

With the data cleaned and organized, the next step is to visualize the data in plots, trying to identify patterns or relationships that would help explain a certain behavior. With the advent of computers in the last decades, an particular tool to analyze data became ubiquitous in almost every PC: Microsoft Excel®. This software contains a wide array of numerical, statistical and visualization tools that enable an easy manipulation of the data, helping tremendously in the analysis process.

But with the growth of the size of the data sets generated in the past 5-8 years, mainly because of the advance in remote sensors, wireless data transmission and computational processing power, the size of these so called ”Big Data” data sets exceeds the limit of Excel of 1,048,576 rows per file. Not to mention that the software gets slow when the number or rows is around half of the limit, depending of the number of columns and calculations performed in the spreadsheet.

To overcome this issue, data analysts needed a tool that would enable them to manipulate, transform, model, and plot the data with ease and speed. For this reason, open source languages such as Python and R gained popularity, and continue to gain space even when commercial tools were created to address the same problem (dealing with the growing size of data sets).

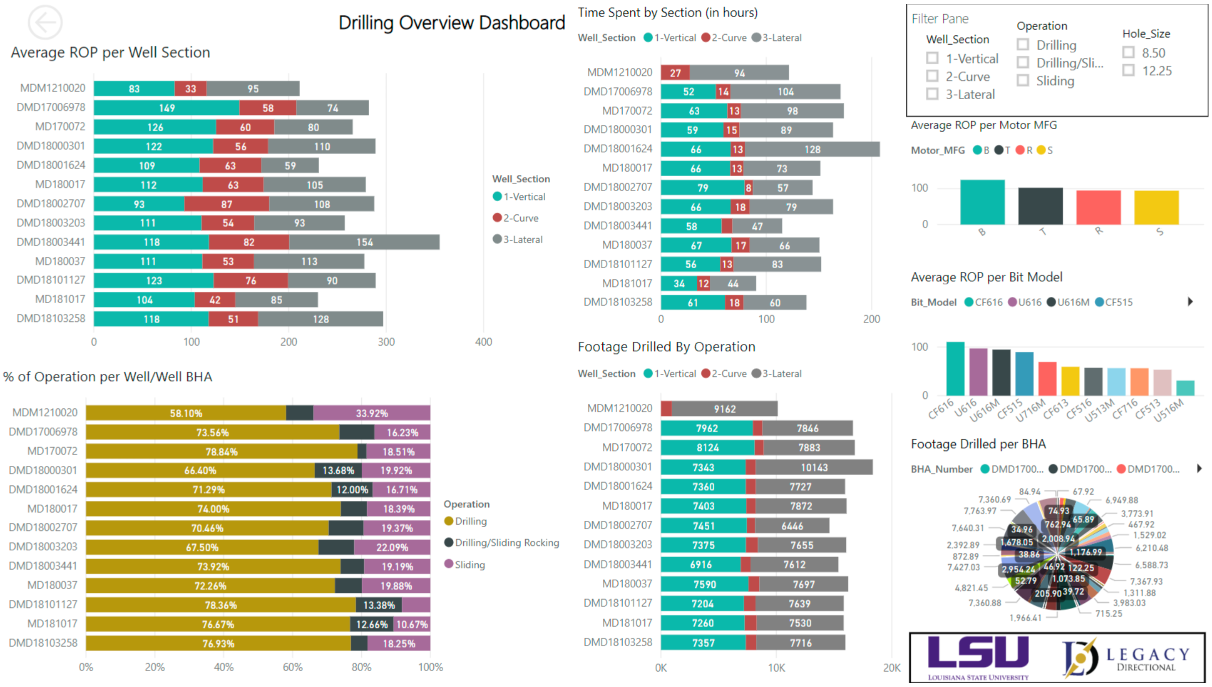
In the present work, we used Python to perform all the data manipulation and box plots described in Chapter 2. For the data visualization we utilize two widely used commercial platforms Microsoft PowerBI® and Tibco Spotfire®. The decision to use commercial programs instead of open source tools is simply to facilitate future implementation of the analysis process by the sponsor company. A great advantage of these products is the ability to interact with the visuals created, highlighting portions of data in other visuals as you select or filter a single plot. When this feature is combined with a plot of various graphs in a single screen (called a dashboard), many interesting conclusions can be derived after some analysis. Figure 3.1 illustrates a dashboard with a high level view of the data set, as well as the interactive feature of these programs.

To begin our analysis, first we summarize general information in Table 3.1 to provide an overview of the nomenclature used in the analysis.

Table 3.1. Summary of Wells Classification

Well Number	Pad Number	Well ID	Well Spud	Well TD
1	1	MDM1210020	12/13/2016	12/25/2016
2	2	DMD17006978	8/29/2017	9/30/2017
3	2	MD170072	9/3/2017	9/18/2017
4	3	DMD18000301	1/5/2018	1/31/2018
5	4	DMD18001624	2/20/2018	3/30/2018
6	4	MD180017	2/27/2018	3/15/2018
7	5	DMD18002707	4/5/2018	6/9/2018
8	5	DMD18003203	4/13/2018	6/1/2018
9	5	DMD18003441	4/20/2018	5/20/2018
10	5	MD180037	4/28/2018	5/12/2018
11	6	DMD18101127	6/17/2018	7/20/2018
12	6	MD181017	6/24/2018	7/8/2018
13	7	DMD18103258	7/26/2018	9/22/2018

A detailed look in the well spud and TD dates shows that the wells were not drilled in chronological order. In a same pad, the sequence was to first drill all the vertical sections one after another. For the last well in the pad, the drilling continued until the final depth. Then, the remaining curve and lateral section of the previous wells was drilled in reverse order. This procedure is better illustrated in Fig. 3.2.



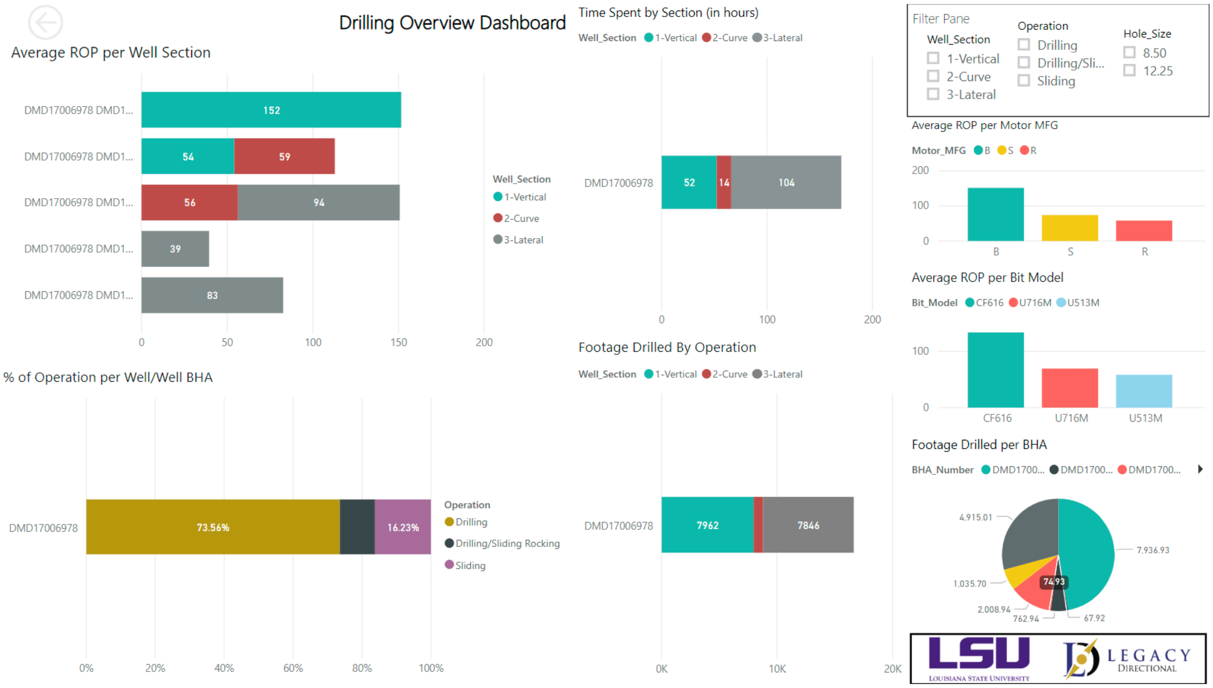

Drilling Overview Dashboard
Average ROP per Well Section
DMD17006978 DMD1... 152
DMD17006978 DMD1... 54 59
DMD17006978 DMD1... 56 94
DMD17006978 DMD1... 39
DMD17006978 DMD1... 83
0 50 100 150 200
Well_Section
● 1-Vertical
● 2-Curve
● 3-Lateral
Time Spent by Section (in hours)
Well_Section ● 1-Vertical ● 2-Curve ● 3-Lateral
DMD17006978 52 14 104
0 100 200
Filter Pane
Well_Section
□ 1-Vertical
□ 2-Curve
□ 3-Lateral
Operation
□ Drilling
□ Drilling/Sli...
□ Sliding
Hole_Size
□ 8.50
□ 12.25
Average ROP per Motor MFG
Motor_MFG ● B ● S ● R
200
100
0
B S R
Average ROP per Bit Model
Bit_Model ● CF616 ● U716M ● U513M
100
0
CF616 U716M U513M
% of Operation per Well/Well BHA
DMD17006978 73.56% 16.23%
0% 20% 40% 60% 80% 100%
Operation
● Drilling
● Drilling/Sliding Rocking
● Sliding
Footage Drilled By Operation
Well_Section ● 1-Vertical ● 2-Curve ● 3-Lateral
DMD17006978 7962 7846
0K 10K 20K
Footage Drilled per BHA
BHA_Number ● DMD1700... ● DMD1700... ● DMD1700...
4,915.01
7,936.93
74.93
1,035.70
2,008.94
762.94
67.92
LSU LOUISIANA STATE UNIVERSITY
LEGACY DIRECTIONAL

Figure 3.1. Dashboard Example Created Using a Commercial Software. All Wells View (top) and Highlight of a Single Well (bottom)

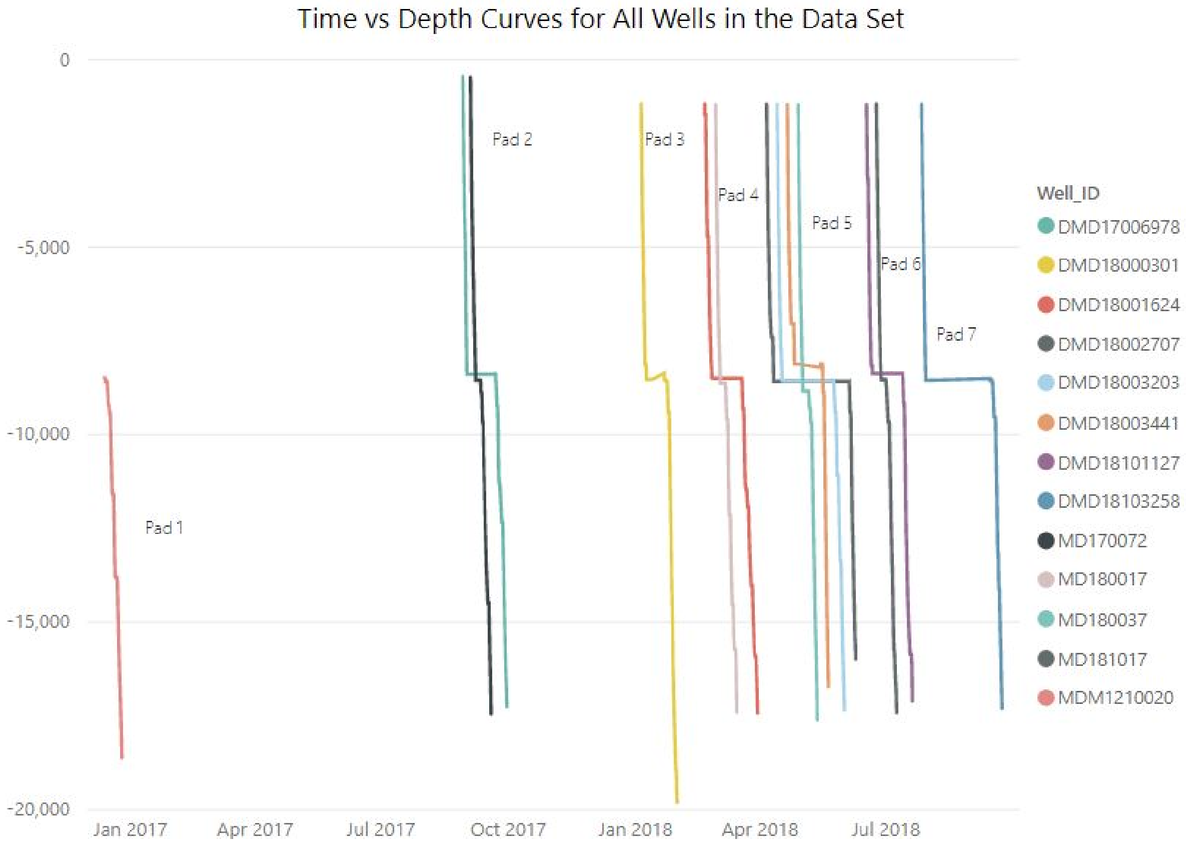


Figure 3.2. Depth vs Time For All Wells in the Data Set

3.2. Vertical Section Data Analysis

Because of the nature of the drilling process, the analysis will be divided into well sections. The first section is the vertical. It covers the shallower formations, thus the less consolidated, softer ones. Also, because the drilling cuts the formations perpendicularly, within a few feet drilled you usually find several changes in lithology.

The analysis begins comparing the average ROP per well, divided by pads. This should provide a first look into the learning curve between wells in the same pad, meaning that the crew learned with the problems from a previous well, and acted to correct in the present one. Fig. 3.3 illustrates this together with the footage drilled to allow comparison. Note that for the first well (Well 1 or MDM1210020), data from the vertical section is missing, so it was removed from this part of the analysis.

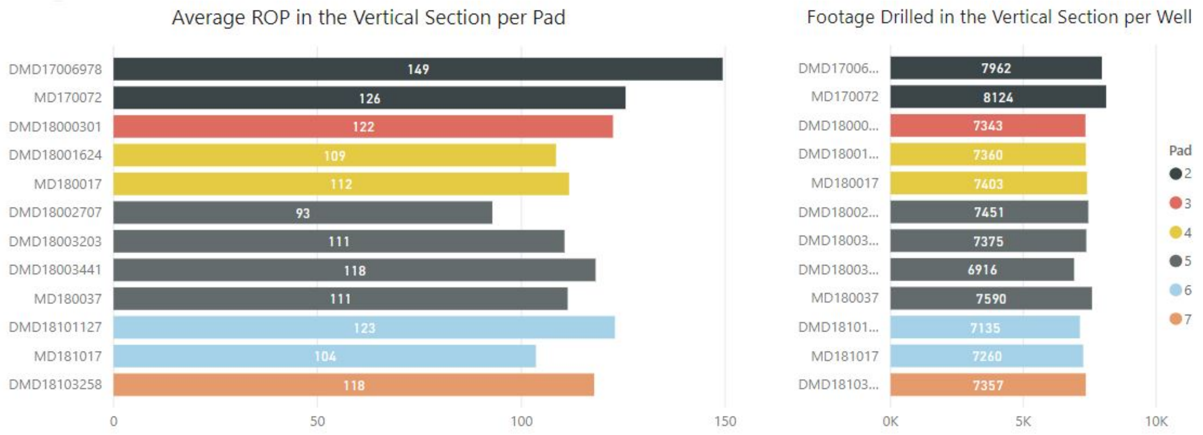


Figure 3.3. Visualization of the Average ROP in the Vertical Section

Analyzing Fig. 3.3, we cannot see any tendency of increase in ROP as new wells are being drilled. In fact, for Pads number 2 and 6 the average ROP decreased on the second well. For the pad with most wells however, Pad 5, the ROP consistently increased, showing improvements in terms of learning.

Another simple metric that can be correlated to average ROP is the percentage of drilling versus sliding for a given section. Figure 3.4 shows the percentage of the rotating and slide drilling in the vertical section for all wells but Well 1, and Fig 3.5 provides the percentage of rotating drilling against average ROP for the same wells.

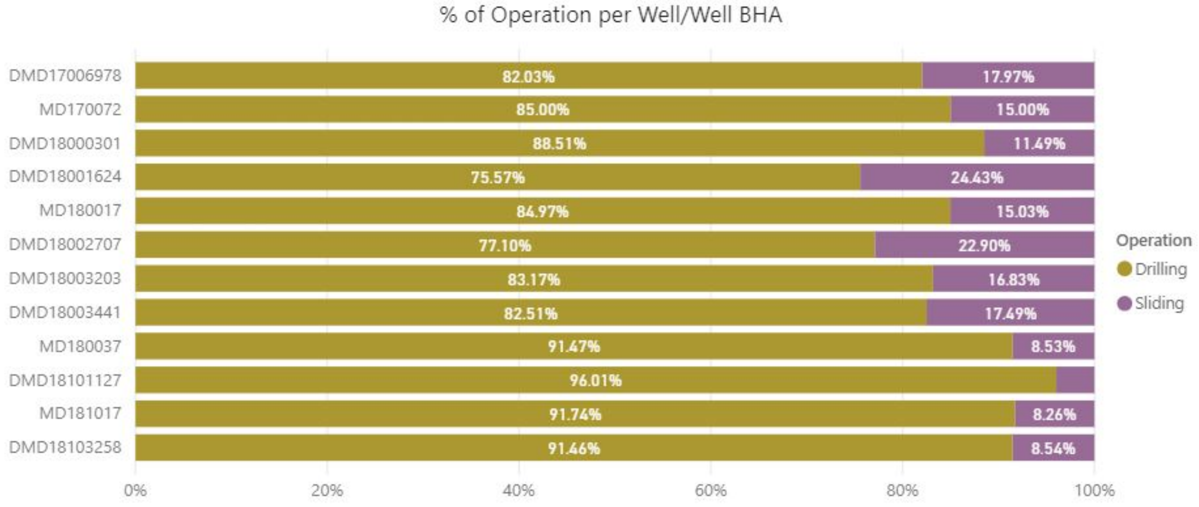


Figure 3.4. Percentage of Rotating and Slide Drilling in the Vertical Section

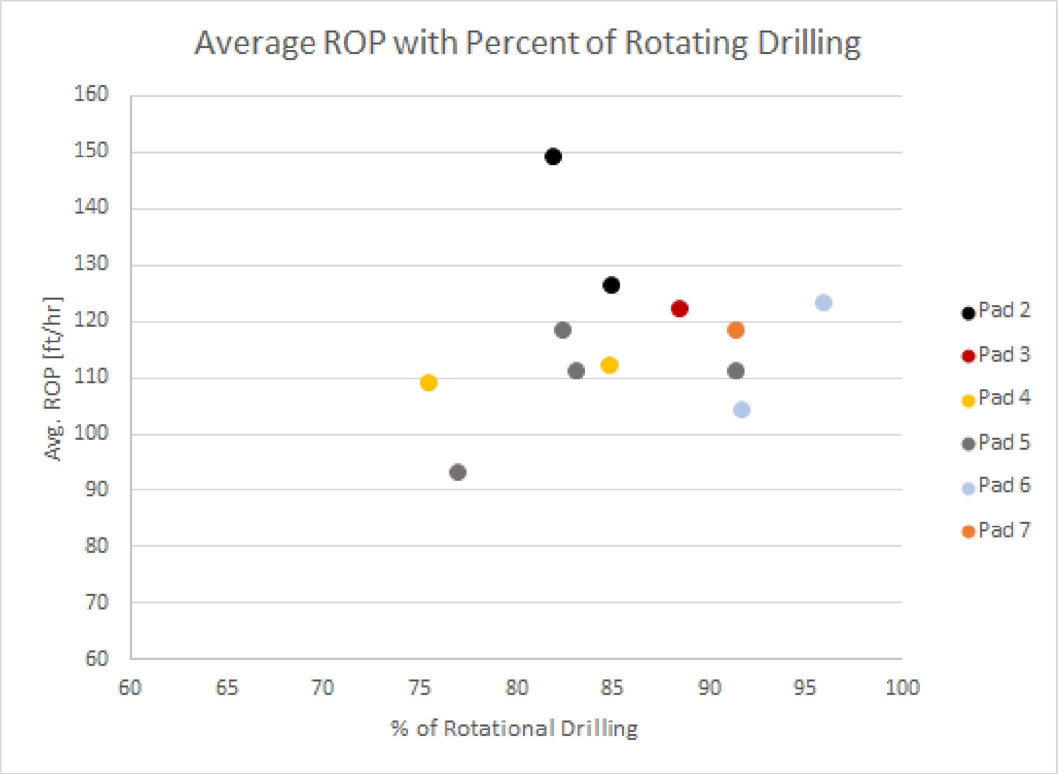


Figure 3.5. Comparison Between Percentage of Rotational Drilling Versus Average ROP in the Vertical Section

From both figures, a direct correlation is not visible when comparing all the wells in terms of average ROP, having the best performer, Well DMD17006978 not having the lowest sliding rate. The opposite is also true: the worst performers in ROP are not the ones with the highest slide percentage. But in a pad level, the tendency seems to exist in all but Pad 2, as seem in Fig. 3.5.

Next we investigate if the bit model used in the wells can provide information about the drilling performance, indicating an effect in bit design when drilling the area.

From Fig. 3.6 it is seen that all wells used bit model CF616 to drill most of the section. Outside of these cases, bit models CF513, CF516 and U513M were used to drill the curve section, but since they started a few hundred feet above the KOP, they appear on this analysis but can be neglected since their contribution to the vertical section metrics are minimal.

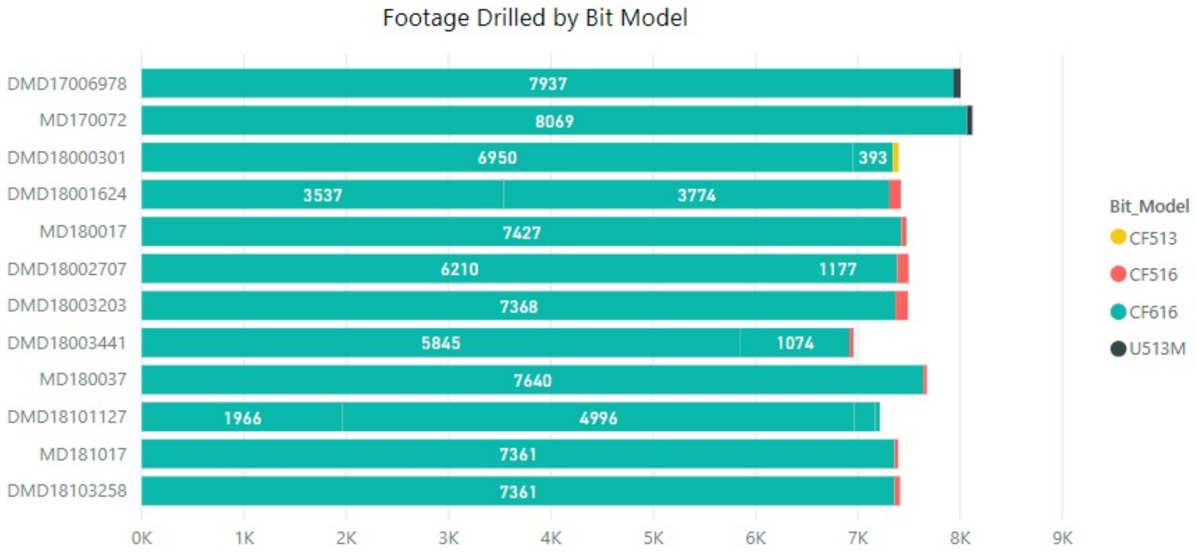


Figure 3.6. Visualization of the Footage Drilled per Bit Model in All Vertical Sections

From a manufacture's catalog, the bit nomenclature is the following [Ulterra, 2019]:

Model Example: A616S

- First letter (A): Represents the product line.
- First number (6): Number of blades of the bit.
- Second and third numbers (16): The size of the cutter in mm.
- Last letter: Denotes the bit body material (M for Matrix, S for Steel).

Although the bit used is the same in all the wells, it can be seen from Fig. 3.6 that out of thirteen wells, in seven cases (Wells 2, 3, 6, 8, 10, 12 and 13) the bit was able to drill the whole vertical section. In the remaining four wells, the bit drilled less than 4,000 feet in two of them (Wells 5 and 11). A search in the DD activity log indicates that in these wells the first bit run was interrupted due to MWD tool failure. However, for Well 11 (DMD18101127) a third run was necessary to drill only 200 feet. In Wells 4, 7 and 9 there is no evident reason for the bits to be pulled out, so we analyze how these bits were operated with the objective to find insights on why the bit had to be pulled.

To do this analysis, we first introduce the concept of hydraulic horsepower per square inch (HSI). This metric is the division of the hydraulic horsepower (HHP) provided to the bit by the hole area (bit area). The expression to calculate the bit hydraulic power in field units (flow rate in gpm and pressure in psi) is:

$$HHP = \frac{Flow\,Rate \times Pump\,Pressure}{1714} \tag{3.1}$$

With the HHP in hand, we divide it by the bit area to obtain the HSI [hp/in^2]:

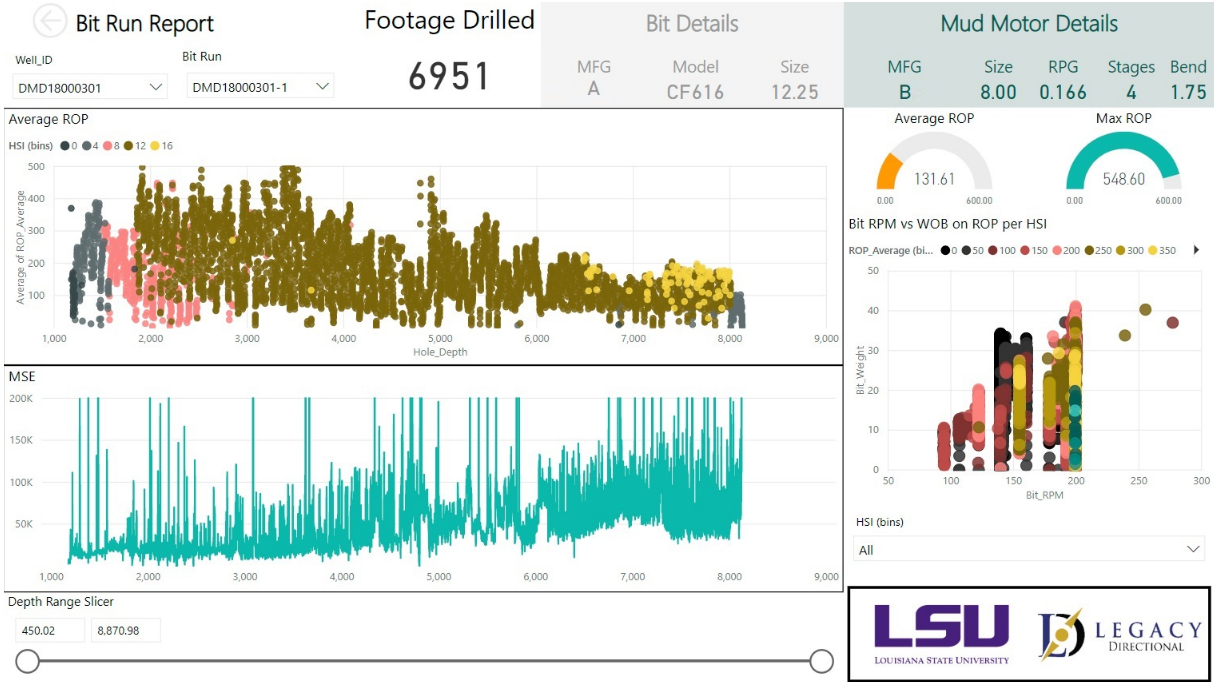
$$HSI = \frac{HHP}{\frac{Bit\,Diameter^2 \times \pi}{4}} \tag{3.2}$$

By grouping the HSI values into bins of size 4 hp/in^2, we can compare the resulting output ROP in terms of the two most important drilling parameters: the WOB and RPM - in this case, the bit RPM. Due to the use of a downhole mud motor, the total rotation of the bit when drilling in rotating mode is the sum of the top drive rotation plus the rotation due to the motor. This last term is calculated multiplying the motor factor (given in rotations per gallon, or RPG) by the total flow rate (in gpm):

$$Bit\,RPM = Motor\,RPG \times Flow\,Rate \tag{3.3}$$

First we analyze Well Number 4 with the aid of another dashboard, called Bit Report. This dashboard contains collected information about the BHA (bit and motor specifications), the plot of average ROP and MSE per foot, and the heat map of WOB vs Bit RPM for each bin of HSI. Fig. 3.7 shows in detail this view.

At a first glance, the average ROP shows a downward trend with depth. The colors in the Average ROP plot (left middle of the figure) represent the groups of HSI. The plot just below this is the surface MSE, showing an increase as drilling advances, in accordance with the ROP decrease. The right middle plot represents the heat map. Further investigation will be performed now filtering this view for the HSI Bins of 8-11, 12-15 and 16+ hp/in^2.



Bit Run Report

Well_ID: DMD18000301

Bit Run: DMD18000301-1

Footage Drilled: 6951

Bit Details		
MFG	Model	Size
A	CF616	12.25

Mud Motor Details				
MFG	Size	RPG	Stages	Bend
B	8.00	0.166	4	1.75

Average ROP: 131.61

Max ROP: 548.60

HSI (bins): All

Depth Range Slicer: 450.02 – 8,870.98

Figure 3.7. Bit Run Report Dashboard View

Fig. 3.8 shows the HSI of 8-11 hp/in^2. We can see that the well was drilled for a small portion of time with this power, and the average ROP was 143.34 ft/hr. In the heat map on the right it is seen that the best combination for the WOB and Rotation was with around 10 klbf and 200 rpm.

Because the HSI Bin of 12-15 hp/in^2 was used in almost the entire section, it was divided into two segments: from 2,000 to 5,000 feet, and from 5,000 to 8,000 feet (Figs. 3.9 and 3.10). For the first part, the decrease in ROP is due to an increase in the slide percent. The points in the left part of the Bit RPM and WOB heat map represents sliding (no top drive rotation), and the right portion denotes rotating sliding. For the second half (Fig. 3.10), we see that almost all the section was drilled with no sliding, but obtained a slower average ROP due to improper WOB. The heat map shows that for this portion of the well the WOB should be increased to around 30 klbf, showing more yellow data points in this area.

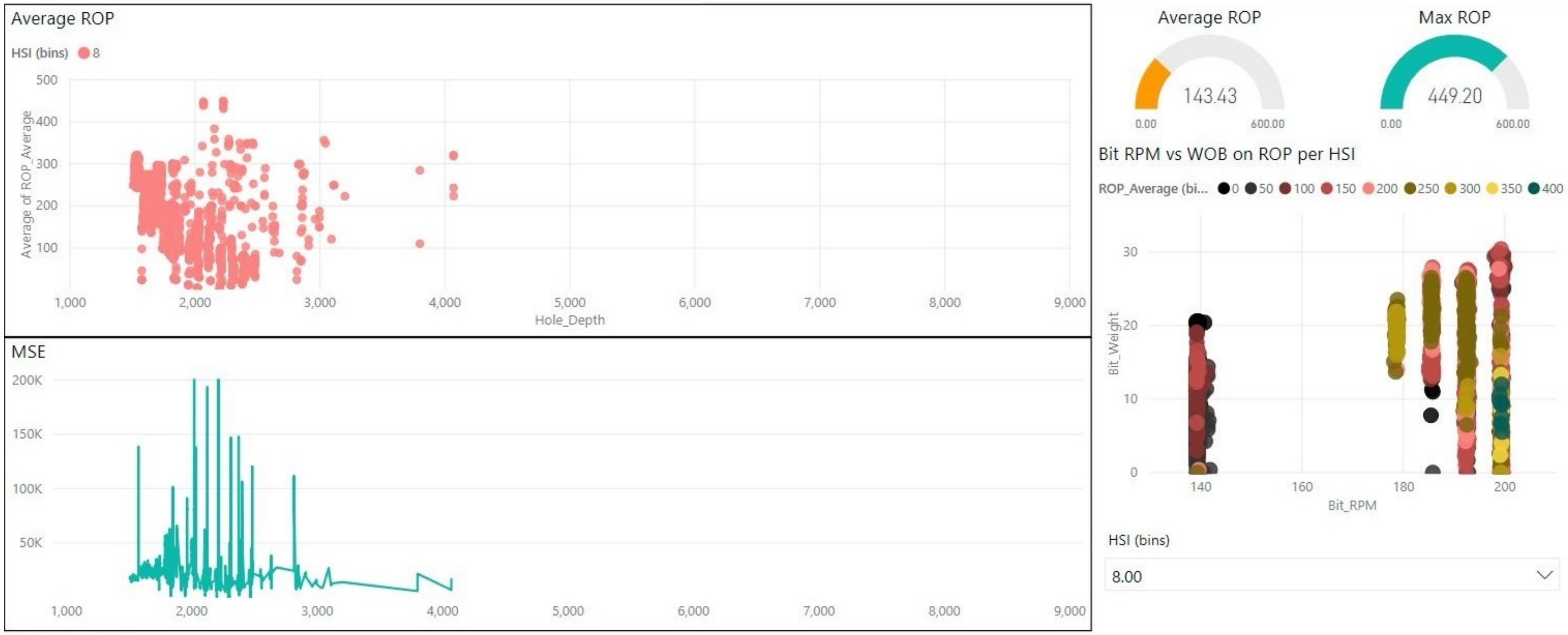


Figure 3.8. Bit Run Report for Well 4, Run 1, HSI Bin of 8 to 11 hp/in^2

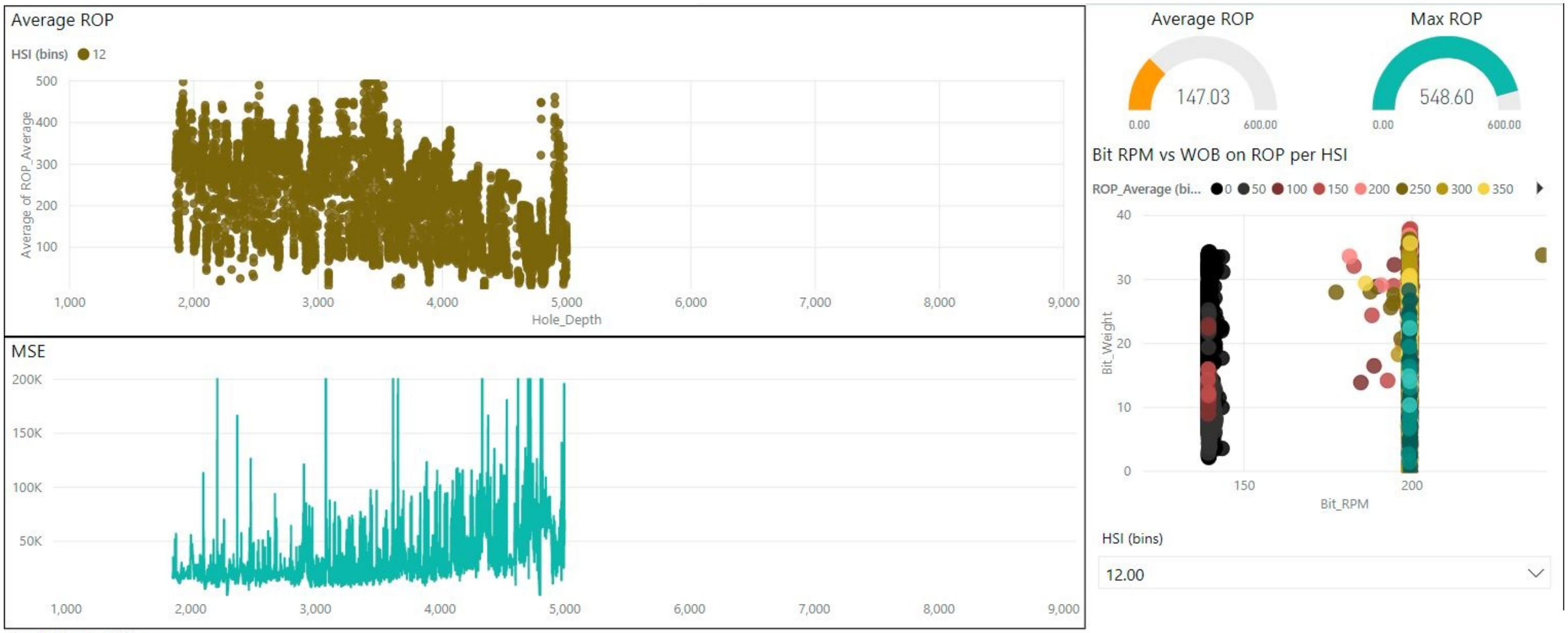


Figure 3.9. Bit Run Report for Well 4, Run 1, HSI Bin of 12 to 15 hp/in^2, From 2,000 to 5,000 feet

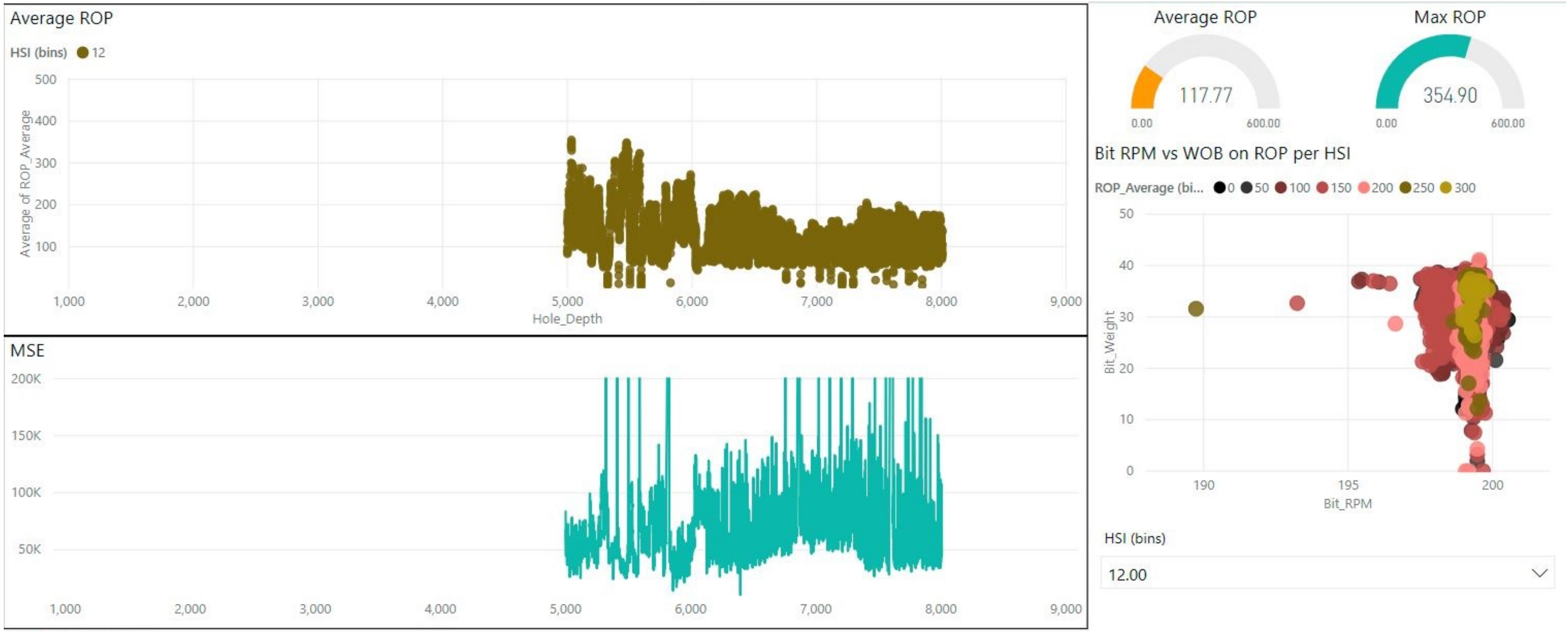


Figure 3.10. Bit Run Report for Well 4, Run 1, HSI Bin of 12 to 15 hp/in^2, From 5,000 to 8,000 feet

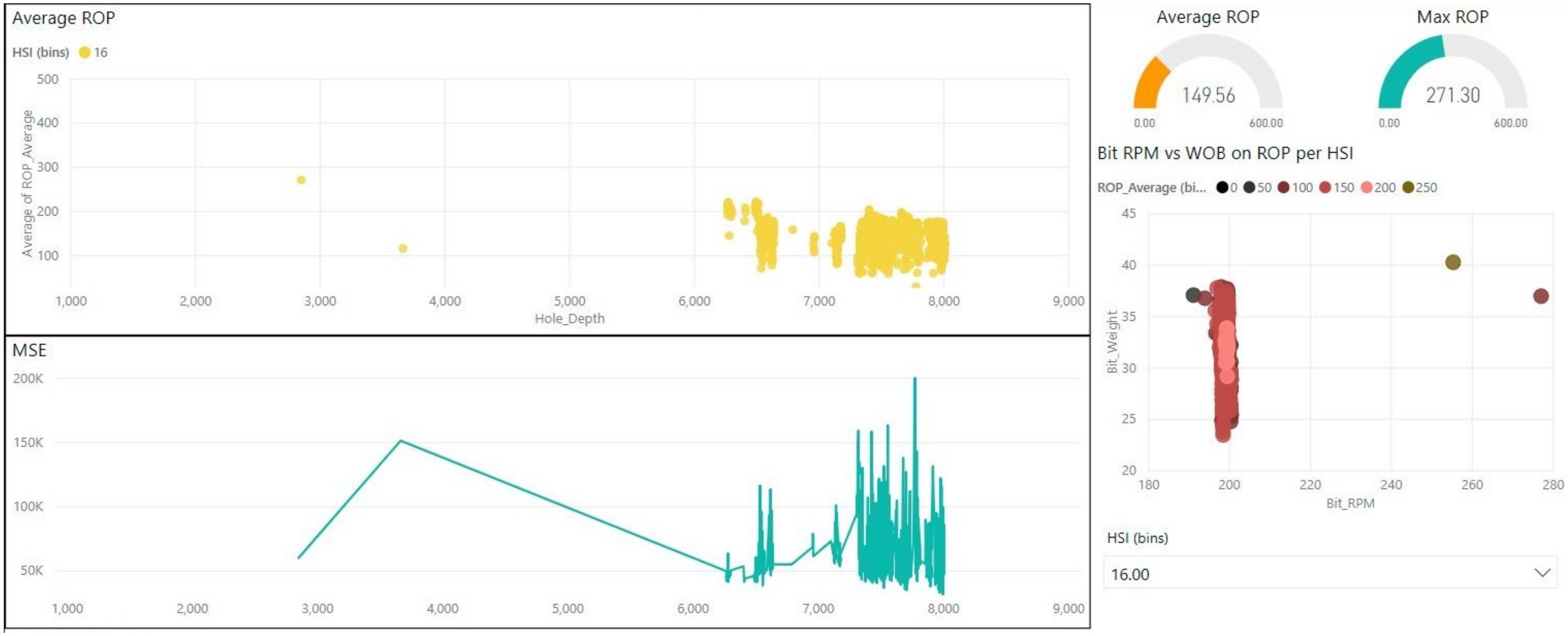
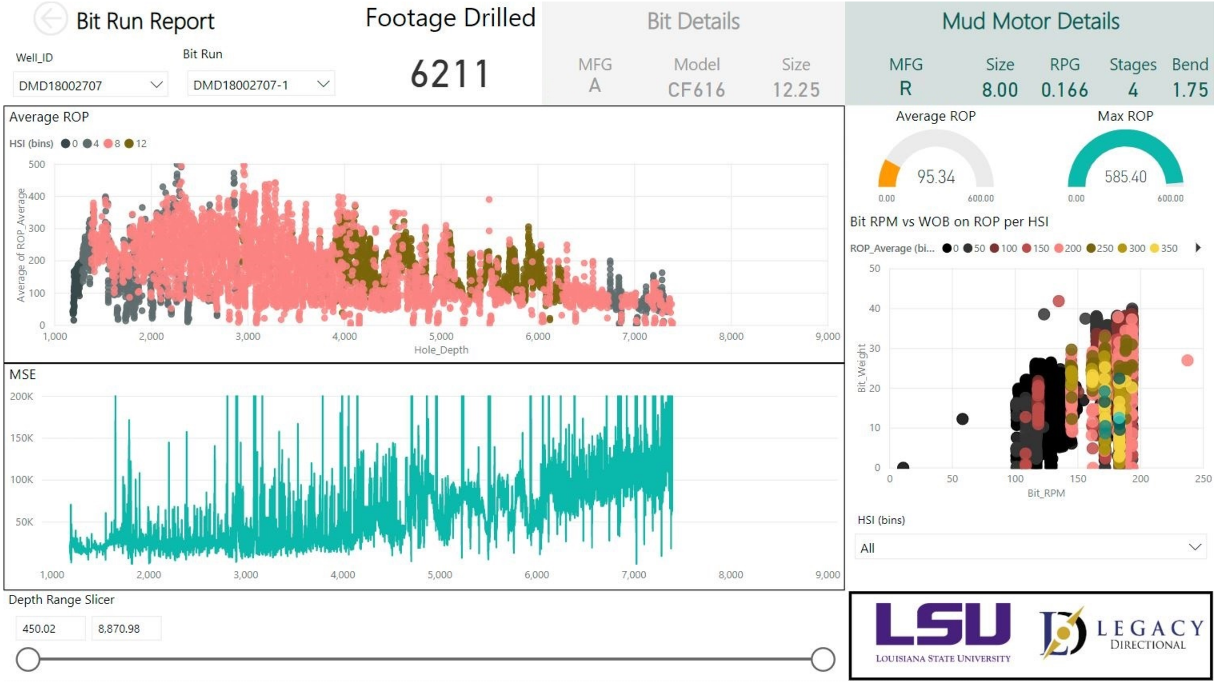


Figure 3.11. Bit Run Report for Well 4, Run 1, HSI Bin of 16 hp/in^2 and More

Lastly the final portion of the well was drilled with most of the footage at HSI of 16 hp/in^2 or more. The heat map map of Fig. 3.11 shows a maximum of 200 ft/hr (excluding outliers), meaning that the flounder point was reached for this set of parameters. The misuse of the bit during the final 1,000 feet might explain the reason the bit was not performing well, forcing the crew to pull it out of the hole.

Performing the same analysis for Wells 7, 8 and 9 (all from Pad 5) we can see that for Well 7, most of the section was drilled with a low HSI (8-11 hp/in^2 range) that resulted in inefficient drilling (Fig. 3.12).



Bit Run Report

Well_ID: DMD18002707

Bit Run: DMD18002707-1

Footage Drilled: 6211

Bit Details		
MFG	Model	Size
A	CF616	12.25

Mud Motor Details				
MFG	Size	RPG	Stages	Bend
R	8.00	0.166	4	1.75

Average ROP: 95.34

Max ROP: 585.40

HSI (bins): All

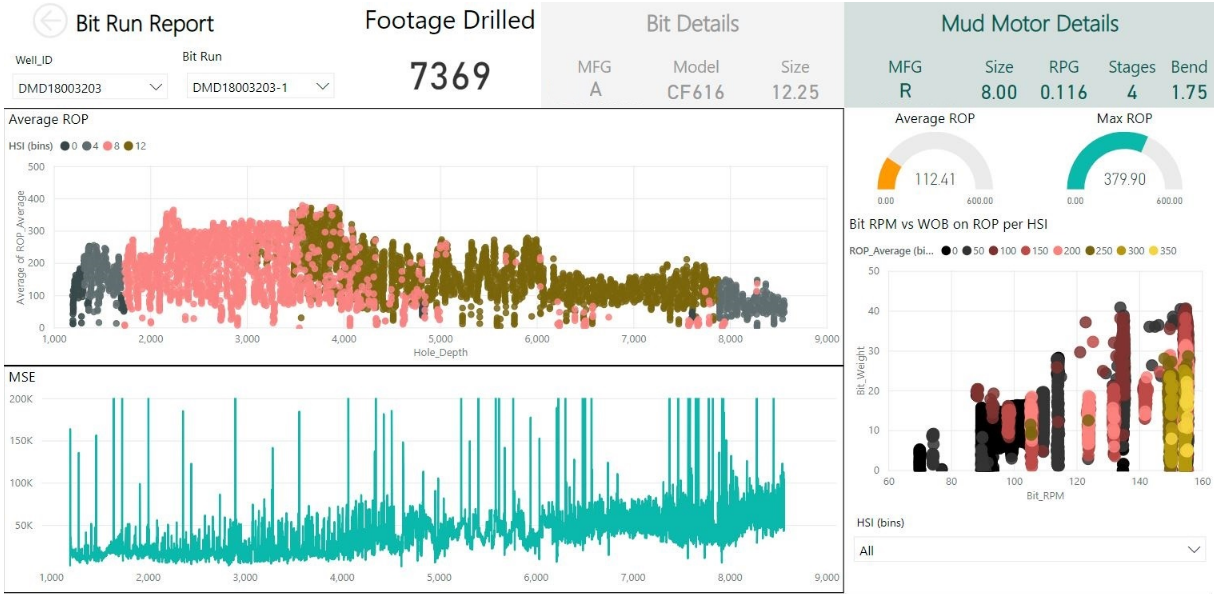
Depth Range Slicer: 450.02 – 8,870.98

Figure 3.12. Bit Run Report for Well 7, Run 1

In the same pad, the following well had the vertical section drilled up to TD, using the same bit and motor assembly. A clear difference by looking at the average ROP plot (Fig. 3.13) is that most of this well was drilled with HSI in the range of 12-15 hp/in^2.

Filtering only the HSI Bins of 8-11 and 12-15 hp/in^2, the respective heat maps show that the sweet-spot for the drilling parameters is around 150 rpm and 15-20 klbf.

Dividing the drilling of the first run of Well 8 into two segments - HSI of 8-11 and 12-15 hp/in^2, we can see the parameters that resulted in better ROP, as shown in Fig. 3.14.



Bit Run Report

Well_ID: DMD18003203

Bit Run: DMD18003203-1

Footage Drilled: 7369

Bit Details		
MFG	Model	Size
A	CF616	12.25

Mud Motor Details				
MFG	Size	RPG	Stages	Bend
R	8.00	0.116	4	1.75

Average ROP: 112.41

Max ROP: 379.90

HSI (bins): All

Figure 3.13. Bit Run Report for Well 8, Run 1

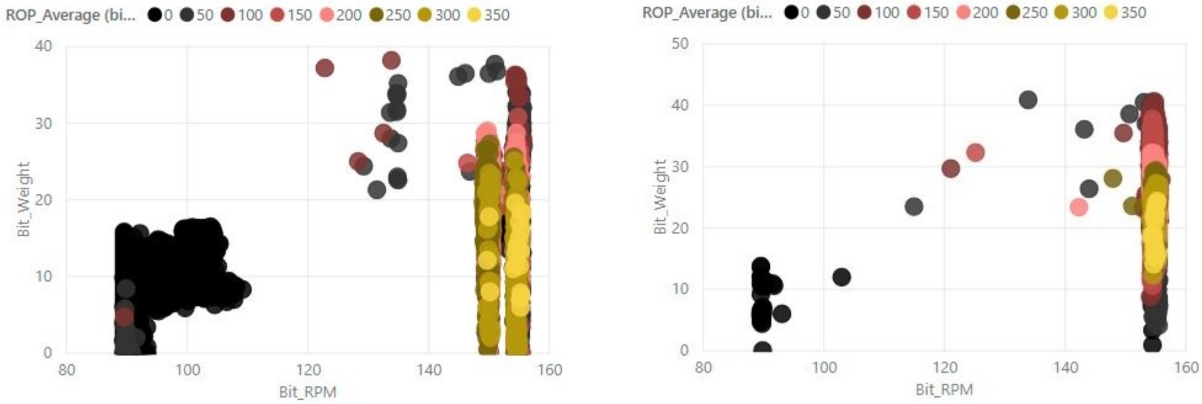


Figure 3.14. Heat Map of WOB vs Bit RPM for Well 8, Run 1. Data Points With HSI of 8-11 hp/in^2 (left), and With HSI of 12-15 hp/in^2 (right)

Fig. 3.14 shows that for the lowest HSI range, during sliding, the best ROP was achieved with lower rotation (around 90 rpm) and WOB close to 5 klbf. For rotating drilling, most points with higher ROP happened with rotation around 150 rpm and WOB of 15 klbf. When the hydraulics was increased, the new range of best performance in this well was similar, but with increased WOB to around 20 klbf.

This bit run however did not present the best performance in the whole data set. Attempting to identify the best parameters to achieve the highest performance from previously drilled wells, we first combine in the plot all wells that were able to drill the vertical section in a single run, namely Wells 2, 3, 6, 8, 10, 12 and 13. Next, we divide the total drilled section into 1,000 feet intervals of depth (besides the first and final 2,000 feet).

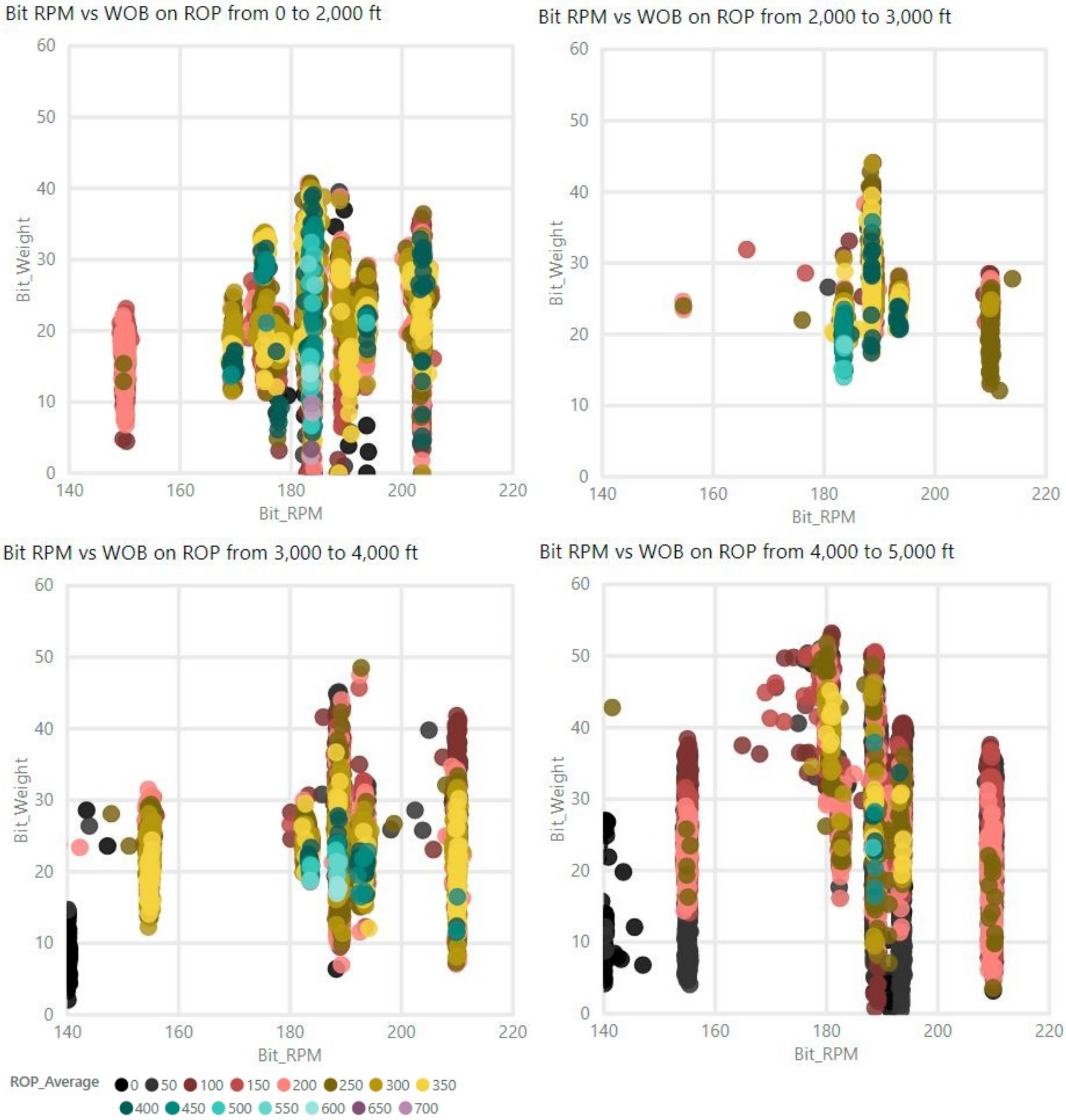


Figure 3.15. Heat Map of WOB vs Bit RPM for the First Half of Depth Intervals

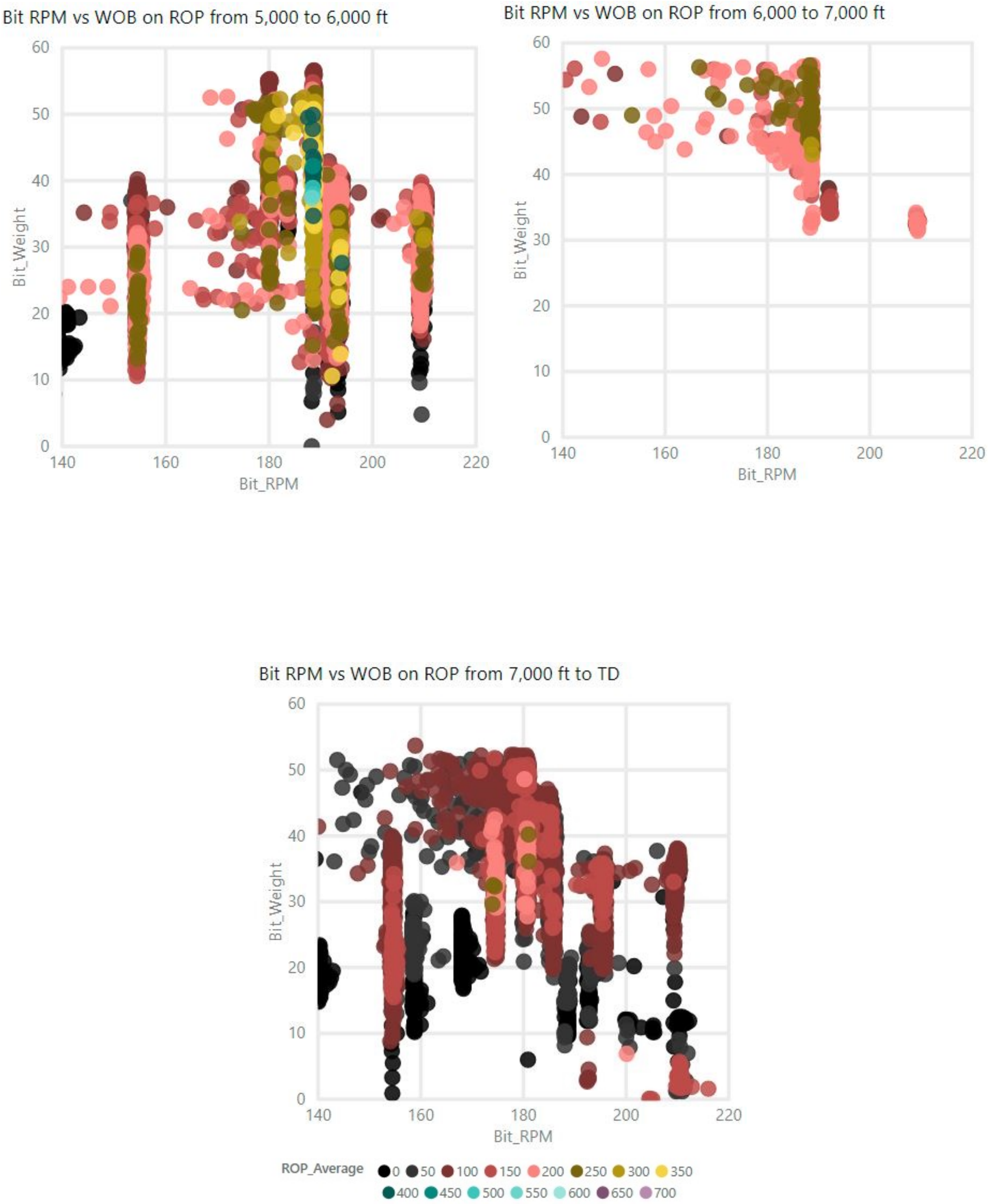


Figure 3.16. Heat Map of WOB vs Bit RPM for the Second Half of Depth Intervals

The recommendations based on Figs. 3.15 and 3.16 are:

- 0 - 2,000 feet: Drill with HSI of 8-11 hp/in^2, with bit rotation of 183 rpm and WOB of 10 klbf.
- 2,000 - 3,000 feet: Drill with HSI of 12-15 hp/in^2, with bit rotation of 183 rpm and WOB of 15 klbf.
- 3,000 - 4,000 feet: Drill with HSI of 12-15 hp/in^2, with bit rotation of 188 rpm and WOB of 20 klbf.
- 4,000 - 5,000 feet: Drill with HSI of 12-15 hp/in^2, with bit rotation of 188 rpm and WOB of 25 klbf.
- 5,000 - 6,000 feet: Drill with HSI of 12-15 hp/in^2, with bit rotation of 188 rpm and WOB of 40 klbf.
- 6,000 - 7,000 feet: Drill with HSI of 16+ hp/in^2, with bit rotation of 189 rpm and WOB of 50 klbf.
- 7,000 - Final Depth: Drill with HSI of 12-15 hp/in^2, with bit rotation of 174 or 180 rpm and WOB of 30 or 40 klbf.

3.3. Learning Curve Analysis

The behavior of ROP with time, seen in Fig. 3.3, suggests that there are issues on how the drilling contractor handles the lessons learned from previous wells. According to [Brett and Millheim, 1986], the time taken to drill a section should decrease exponentially, with the rate represented by the equation:

$$t_n = C_1 e^{C_2(1-n)} + C_3 \tag{3.4}$$

where:

- t_n: Total time to drill the n^{th} well.
- C_1: Learning potential.
- C_2: Learning rate.
- C_3: Operational/technical limit.

[Damski, 2014] explains that the learning rate C_2 reduces the gap between C_1 and C_3. The greater C_2. the faster t_n reaches C_3. A good learning curve is represented by C_2 between 0.45 and 0.80. Fig. 3.17 depicts the learning curve behavior, as well as the mathematical meaning of the constants.

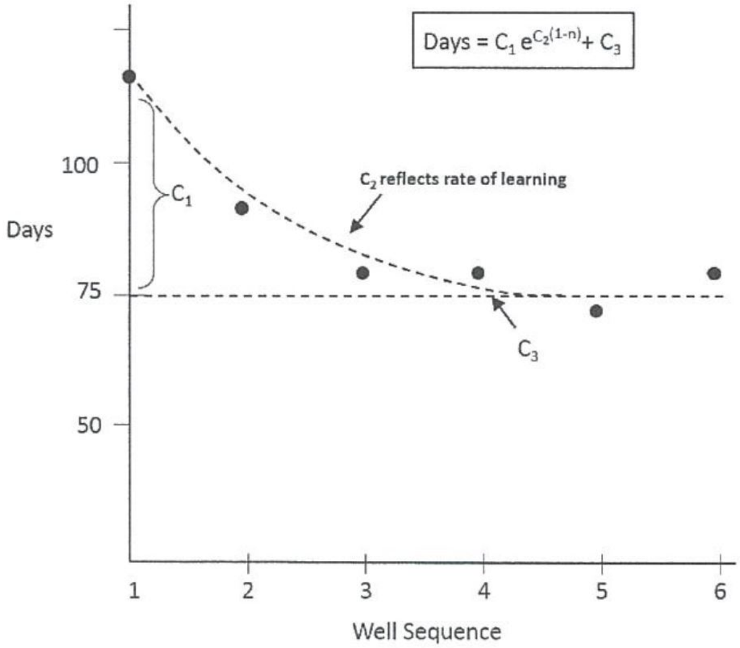


Figure 3.17. Typical Learning Curve for a Drilling Campaign, From [Damski, 2014]

Using the learning curve theory, we plot the average time taken to drill the wells versus the well sequence for the three sections: Vertical, Curve and Lateral. Fig. 3.18 shows the curve for all wells. It is important to reiterate that in all plots in this section, only data points representing drilling operations (drilling, sliding or rocking) are being used.

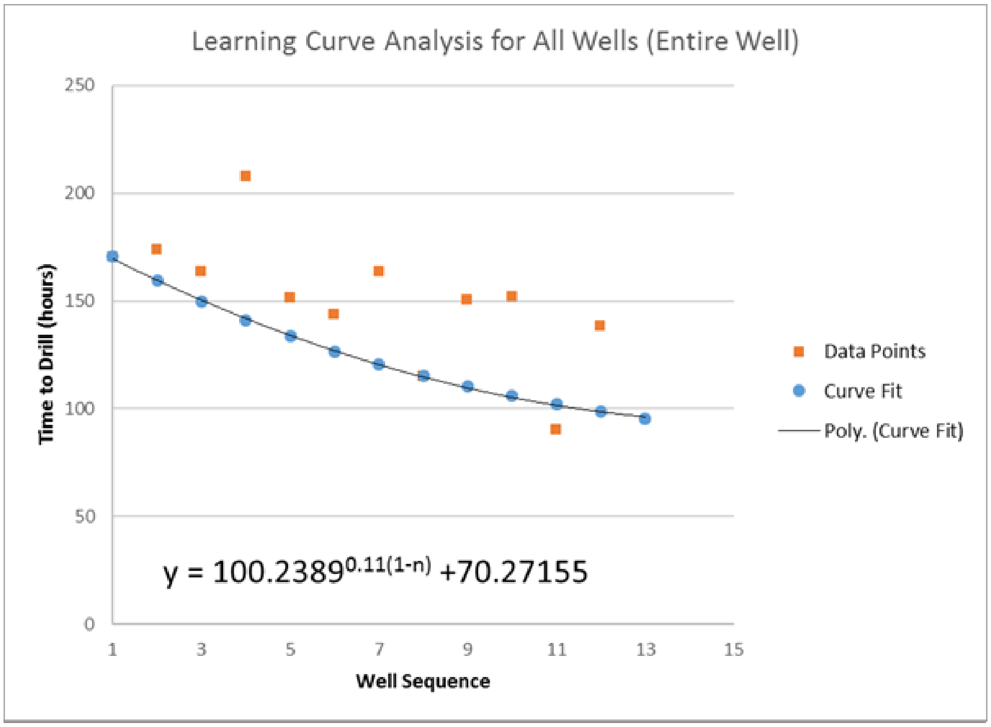


Figure 3.18. Learning Curve for All Wells in the Data Set

The overall learning coefficient when considering all wells is 0.11. This denotes that there is a room for improvement regarding previous performance to be communicated along the drilling crews. However, because the wells were drilled in the course of one year (Fig. 3.2), and we do not know if the drilling crews were the same during the whole campaign, we nail down the learning curve analysis for the wells in Pad 5, since it is the pad with the largest number of wells. The following plots (Fig. 3.19) show the learning curve behavior by well section. In the vertical part of the well, the improvement is maintained until the third well, resulting in a learning rate of 0.36. In the curve section, the behavior found is defined as ”loss of learning” [Millheim et al., 1998], where a past performance level is not obtained again. For this case, we calculate the learning curve from the second well, achieving a high learning coefficient of 0.9643, since the time to drill the curve section went from 18.1 to 10.1 hours. Lastly, the analysis for the lateral section presents inconsistent performance, with learning coefficient of 0.14.

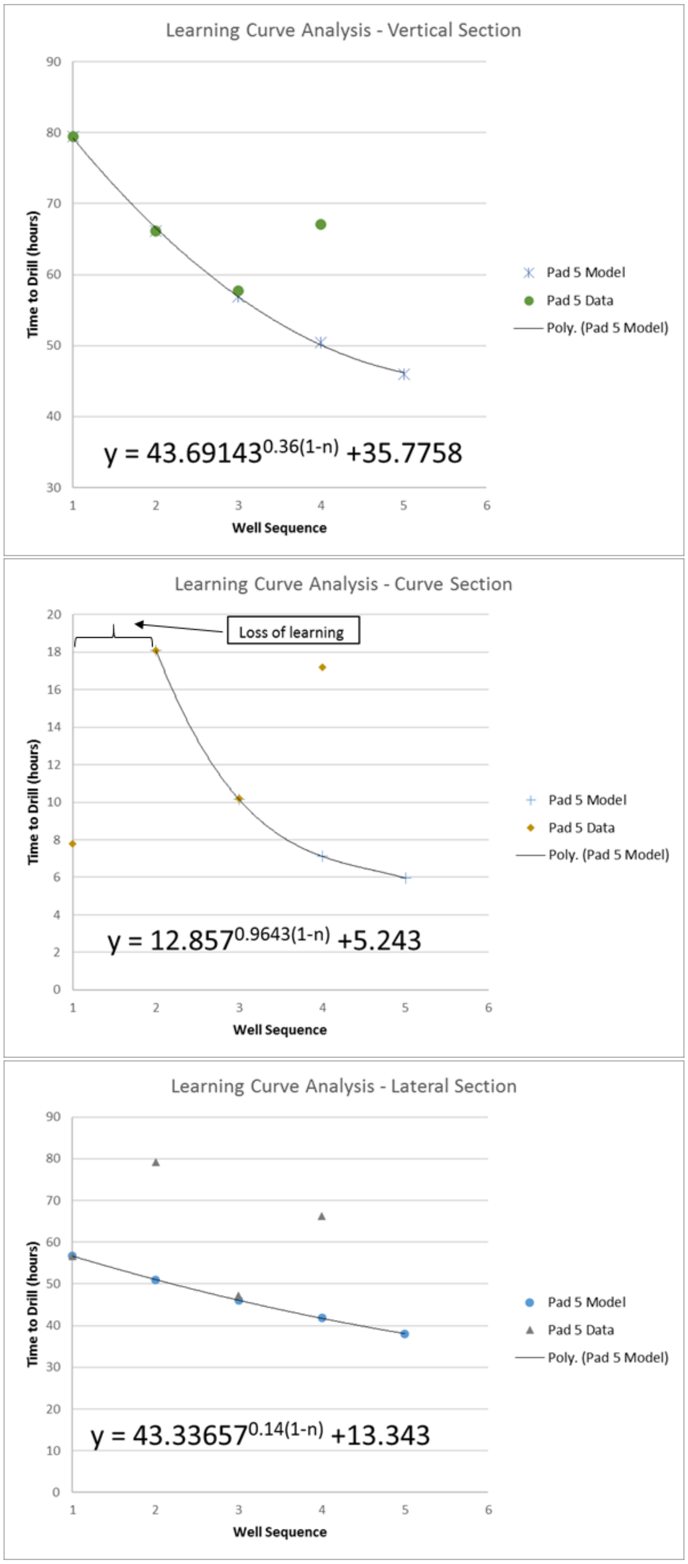


Figure 3.19. Learning Curve for Pad 5 Divided by Vertical (top), Curve (middle) and Lateral (bottom) Sections

The learning curve analysis demonstrates that there is room for improvement when evaluating the wells in the data set. Some recommendations to overcome this issue are:

- Enhance Documentation: The drillers could document and classify all drilling problems encountered in a well into a structured database for further analysis.

- Knowledge Sharing: The drilling contractor should promote knowledge sharing among the drilling crews by issuing performance alerts to be read by all relevant crew members (Toolpushers, Drillers and Assistant Drillers).

- Active Performance Monitoring: The company should keep track of the metrics as the wells are being drilled to get the benefits of the knowledge created in real time.

- Rotate Crew Members: One good approach is to rotate key crew members between crews and track the performance changes. This usually results in good practices being shared between the crews, and can raise equipment issues that are hard to detect if the rotation is done between rigs.

- Methods-Time Measurements (MTM) Analysis: This method calculates the time spent for each task as well the positioning of each member of the system (drilling equipment) to execute the task. The MTM analysis can result in significant performance improvements, for example leaving the slips ready and close to the well before a connection, rather than placing it at random on the drill floor.

Chapter 4
Machine Learning Modelling

Machine learning (ML) is a field inside computer science that has as a goal to teach the computer to provide answers (by calculating probabilities) by example. The fact that the computer learns by example differs than traditional coding because the computer was never explicitly programmed to perform a task, but it learns with the data how the data points are distributed in a data set, therefore it can infer the outcome for a new row of data never seen before.

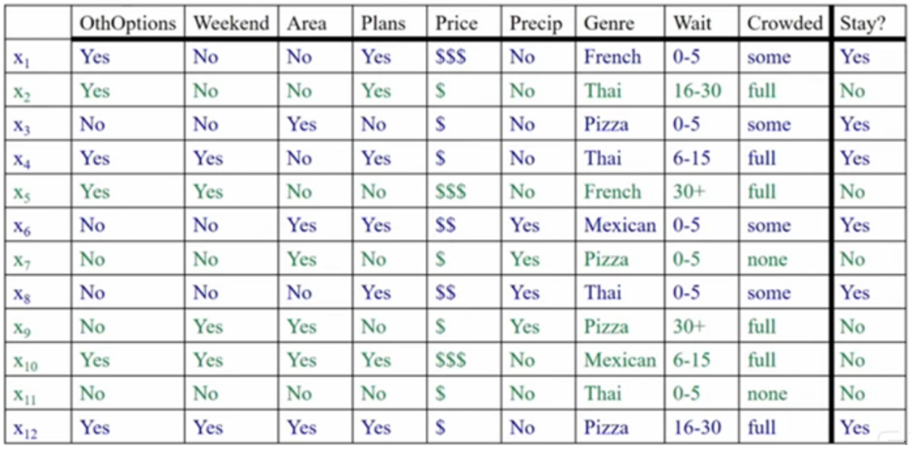
A common example of ML is picture recognition. For example, if a data set containing lots of pictures of chairs and random objects, the computer will first learn the traces that differentiate them. In this example, it is important that the pictures are labeled (have the information if they are chairs or not chairs), so the computer can learn about it. Also, it is important to highlight that no other information about the differences between a chair and "not a chair" is provided (i.e., a chair has four chairs, a back seat, etc). The computer will analyze all the pictures (in this example the pictures will be converted into one dimensional vectors containing the pixels of each picture) and identify the relationship between them. Because it is trained by example, another key aspect of ML is to have a good number of data points in the training set to result in accurate models.

Once the model is trained, meaning it calculated the important traces for all the pictures that have labels as "chair" and "not chair", it can be exposed (or tested) to a new picture, and be asked if the picture contains a chair or not. The way the computer model does that is to calculate the features (pixels) of the new picture and answer the value that the picture has the greatest probability to be. They way this procedure is normally performed, 60% to 70% of the data goes to training, and the remaining part to testing.

4.1. The Boosted Decision Trees Algorithm

Besides classification, another type of problem usually solved using ML is regression. Regression is the analog of classification, but instead of returning a class, it returns a numerical value. There are several algorithms that are used to estimate new values given a training data set. For this work, we will focus on the method that resulted in the best performance, or the smallest error, called Boosted Decision Trees.

”Decision Trees” is the name given to algorithms that provide an output by asking multiple questions to the data by comparing multiple data points and checking the outcome. The reasoning is similar to how humans perform this action, dividing possible outcomes into new branches (or leaves) after a new outcome is analyzed. For example, lets assume if we have the data set represented in Fig. 4.1. This data set contains answers of customers for the question ”if they would wait for a table in a restaurant”. In order to answer, they were given some features about the restaurant (if they have other options, type of food, location, price) as well as the occasion (special date, is it crowded, etc). After considering all the aspects of the decision, the customers provide the answer (label) if they would stay or not at the restaurant.



	OthOptions	Weekend	Area	Plans	Price	Precip	Genre	Wait	Crowded	Stay?
x_1	Yes	No	No	Yes	$$$	No	French	0-5	some	Yes
x_2	Yes	No	No	Yes	$	No	Thai	16-30	full	No
x_3	No	No	Yes	No	$	No	Pizza	0-5	some	Yes
x_4	Yes	Yes	No	Yes	$	No	Thai	6-15	full	Yes
x_5	Yes	Yes	No	No	$$$	No	French	30+	full	No
x_6	No	No	Yes	Yes	$$	Yes	Mexican	0-5	some	Yes
x_7	No	No	Yes	No	$	Yes	Pizza	0-5	none	No
x_8	No	No	No	Yes	$$	Yes	Thai	0-5	some	Yes
x_9	No	Yes	Yes	No	$	Yes	Pizza	30+	full	No
x_{10}	Yes	Yes	Yes	Yes	$$$	No	Mexican	6-15	full	No
x_{11}	No	No	No	No	$	No	Thai	0-5	none	No
x_{12}	Yes	Yes	Yes	Yes	$	No	Pizza	16-30	full	Yes

Figure 4.1. Example Data Set to Explain the Decision Tree Algorithm

Now, taking this table to a computer using decision trees, the computer will try to predict which combinations of questions will result in the answer with the best confidence (least error). To do that, it will divide each column into separate trees, and proceed to a new question until an answer is found, meaning that there is no more ambiguity left. For example, in Fig. 4.2 we see that the first questions ”if the restaurant is crowded or not” can lead to multiple answers, so another layer (node) is created. The next level checks the values in the ”do you have plans?” column, and evaluates if it still can provide a definite answer. Since it still cannot, another questions is asked, until there is a clear difference between the answers depending on the value of the column.

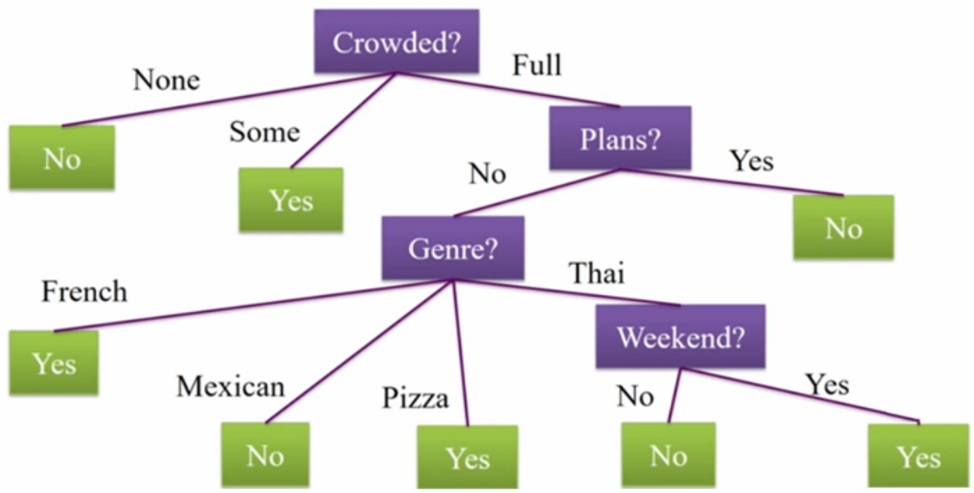


Figure 4.2. Example of a Decision Tree

Because of the way it is created, a particular tree will only perform well if the sequence of questions asked is the same. To avoid this limitation, thus making the process more generalized, boosting is performed. Boosting is the combination of several decision trees made with different orders of questions. A boosted decision tree model therefore, is the model that returns an answer to the combination of the answers from all the trees in the model.

4.2. ROP Prediction of the Curve Section Using Decision Trees

In the present work, an example of ML application will be the prediction of the ROP in the curve section of a well. The platform used to program the ML algorithm is the Microsoft® Azure Machine Learning Studio. First, we select eight wells to be used as the training wells. These wells were chosen in such way that no two wells from the same pad were used to train the model. Also, between wells of the same pad, always the first well is chosen as the training well (so we can assess the model’s ability to predict the next wells’ performance). The training workflow begins by compiling the cleaned data, after running the process described in Section 2.3.1 into a single table (Fig. 4.3).

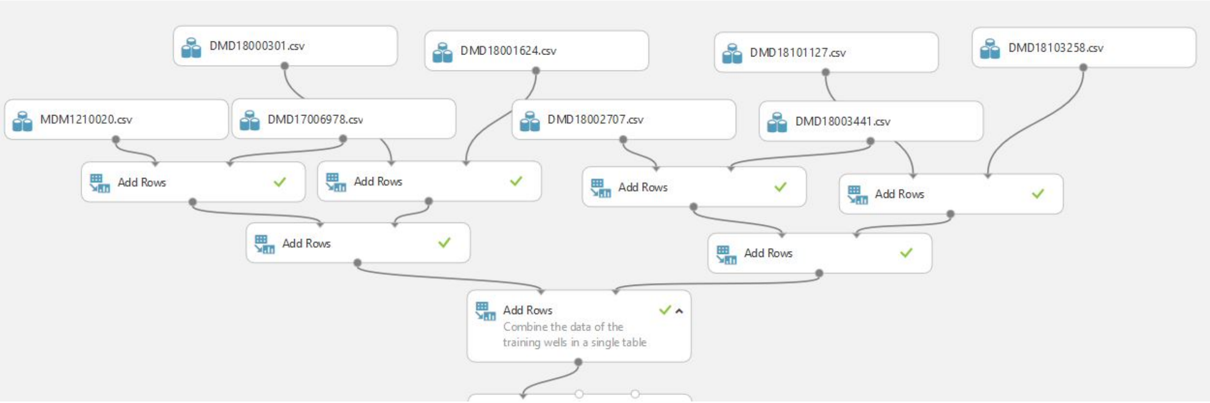
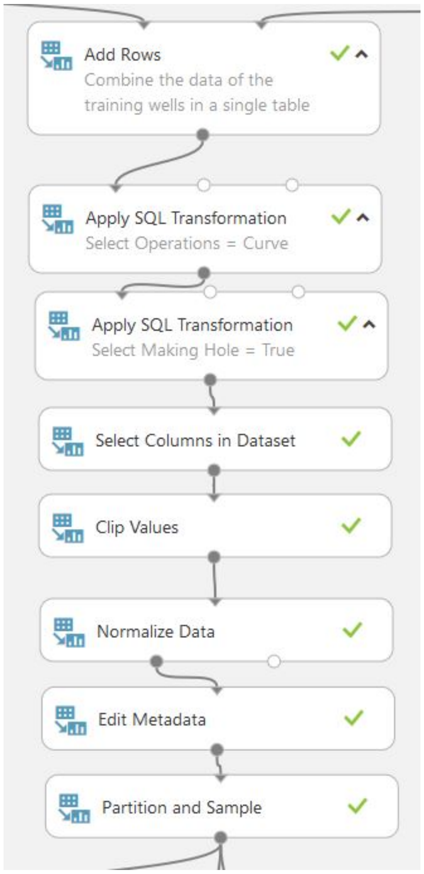


Figure 4.3. Combination the Data From the Eight Training Wells

Next, minor preparation steps are performed for the given application. In this example, we are interested only in the drilling of the curve section, so the column ”Well Section” is filtered for values of ”Curve”. Also, we look for values where ”Making Hole” is ”True”, and ”Operations” is not ”Other”, resulting in only data points where drilling occurs. Finally, to avoid bias in the training regarding the magnitude of the features, all data columns are normalized to values between 0 and 1. This way, the column of Hole Depth - average of 8,500 feet - will not have a greater weight in the training than top drive rotation - 0 to 60 rpm, even having much smaller influence in the resulting ROP. Lastly, the data set is partitioned into equally sized folds for cross validation. Cross validation is the procedure of dividing the training data set into smaller parts, and training separately in search of sampling bias. For example, lets say that in the random split between 60% of the data for the training set, and 40% for testing, no data points of ”sliding” are in the training set, so the model has never seen this operation when training. So, when asked to infer the test points containing ”sliding” drilling, the model will perform poorly because of this sampling bias. Cross validating the training set into equally divided sets can verify if the model performs with the same accuracy in all subsets. These steps as shown in Fig. 4.4.



Add Rows
Combine the data of the training wells in a single table

Apply SQL Transformation
Select Operations = Curve

Apply SQL Transformation
Select Making Hole = True

Select Columns in Dataset

Clip Values

Normalize Data

Edit Metadata

Partition and Sample

Figure 4.4. Filtering and Normalization of the Data

In the sequence of partitioning and sampling, the model is trained. Fig. 4.5 shows that the training uses also another technique called ”tuning hyperparameters”. To each boosted decision trees model, it is possible to customize its parameters such as the total number of trees in the model, the depth of each tree (number of questions), the number of leaves (nodes) in each level, etc. Since it is a heuristic model, there is no guarantee that one configuration will be better than another, being completely dependent on the data set size and values. In that way, this tuning method trains the model in several configurations until it finds the best configuration. The time spent for a computer to train all these combinations will depend on the number of the parameters sweep that is passed to the tuning method. In this work, we sampled a total of 180 combinations, varying the maximum number of leaves per tree (from 2 to 128), the minimum number of cases to form a leaf (from 1 to 100), the learning rate, which is the steps taken to reach the final answer (from 0.025 to 0.4) and the number of trees constructed (from 20 to 500). Fig. 4.6 show examples of the overall shape of the first nine trees, and Fig. 4.7 presents details of the decision logic in part of a tree.

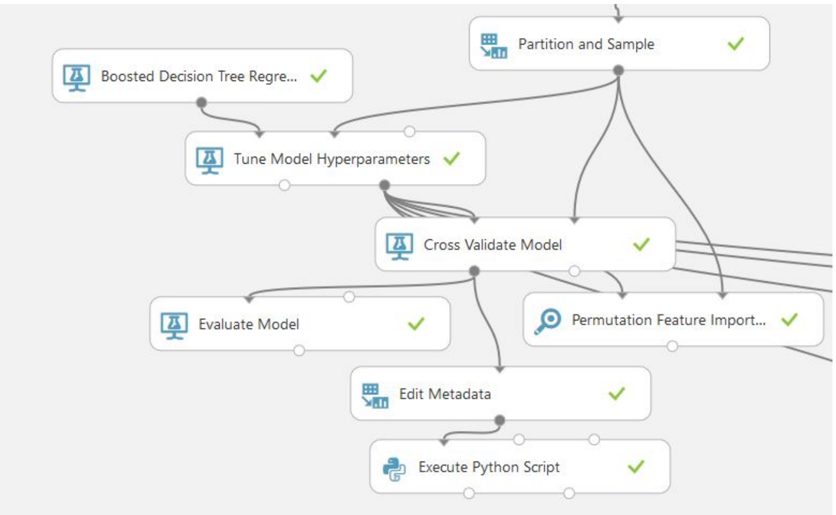


Figure 4.5. Training the Boosted Decision Trees Algorithm

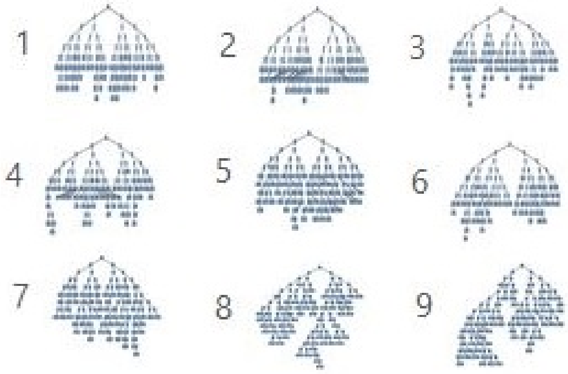


Figure 4.6. Overall Shape of the First Nine Decision Trees

The resulting best model (with the smaller error) has 350 trees, a learning rate of 0.375257, minimum of 26 cases per leaves, and 90 leaves. Besides sweeping for the best parameters to achieve good learning, and performing cross validation to avoid sampling bias, we also executed a task called ”permutation feature importance”. This procedure removes one column at a time from the training set and computes the overall performance. The individual results from each trial missing a different column is then compared with the training using all columns. The features that cause the greatest penalty in performance, are

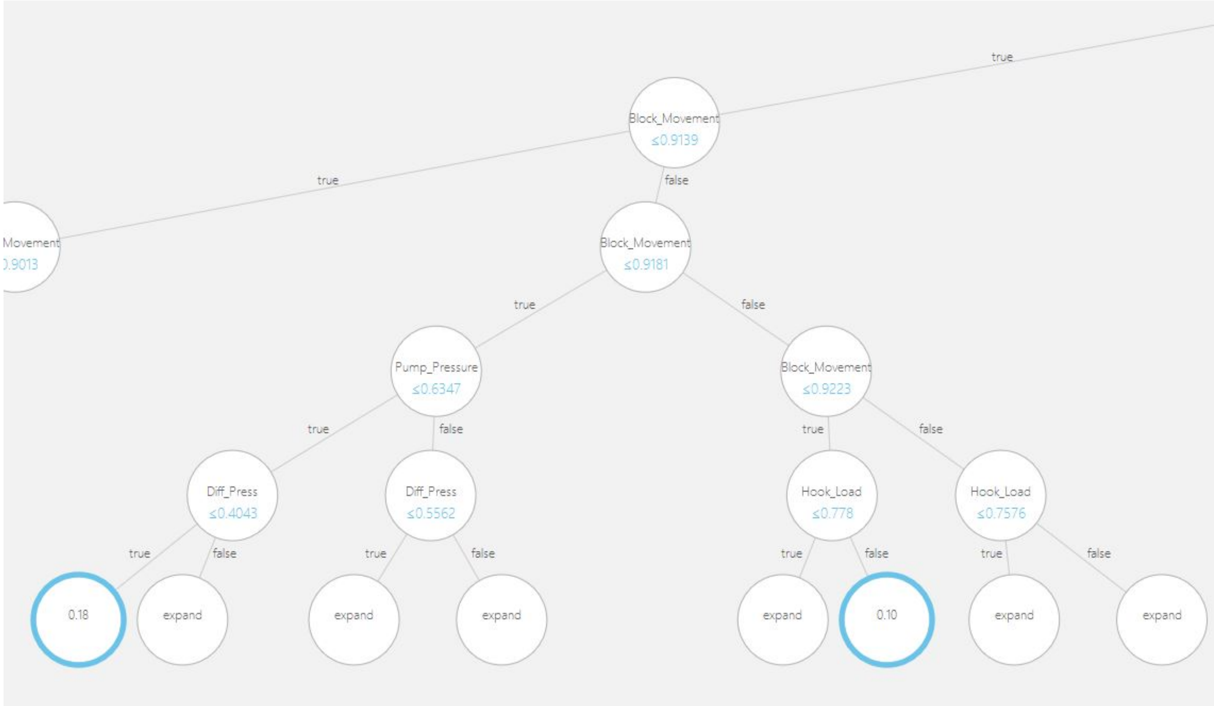


Figure 4.7. Example of the Decision Logic of Some Leaves of One Tree

attributed the highest importance; for those that barely affected the overall performance it was assigned lower importance. The resulting importance estimation in the ROP is shown in Table 4.1.

Table 4.1. Feature Importance When Predicting ROP

Feature	Score
Hook_Load	0.092441
Block_Movement	0.082389
Bit_Weight	0.080794
Pump_Pressure	0.075463
Top_Drive_Torque	0.044885
Diff_Pressure	0.044694
Block_Height	0.043641
Hole_Depth	0.038213
Flow_In_Rate	0.025674
Pump_SPM_Total	0.02061
Top_Drive_RPM	0.020033
Bit_Position	0.019743
Gamma_Ray	0.016839
Operation	0.000041

Table 4.1 shows that the hook load (and the associated WOB), the block movement speed and the pump pressure are the top features when predicting accurately the ROP. The top drive RPM resulted in a lower value mainly because almost all points represent slide drilling, so this columns has almost all values close to zero. The formation (gamma ray) also does not play a big role in the curve section since it is drilled through a fairly consistent lithology due to the angle build.

Then the model is evaluated, to check if the performance predicting the ROP is good or not. The two main metrics analyzed are the root mean squared error (RMSE) and the coefficient of determination (R^2). The RMSE represents the amount of error in the predictions, thus the less the better. The R^2 is a measure of how well the model represents the variation of the data, and the greater the better (goes between 0 and 1). The present model performed the cross validation with a RMSE of 0.033949 and a R^2 of 0.97751. These metrics suggest that the model is predicting well the ROP of the data set that it was trained upon.

The next step is to upload data from a new well (never exposed to the model) to check its ability to reproduce the results. We selected Well Number 10 (MD180037) since it is the last well of Pad 5 (Fig 4.8).

Since we used 2 wells from this pad in the training set, we expect a good result in the testing. The resulting performance for Well 10 was an RMSE of 0.137218 and R^2 of 0.560061. Although the performance is not as good as for the training set, the resulting model covers most of the variations of the ROP, as seen in Fig. 4.9.

With the satisfactory results of the model, the following steps are to change values of the test variables to see the resulting effect in the ROP, in the attempt to predict a better performance in a future drilling for a similar well. In this example, we test three scenarios multiplying the WOB by 1.3, 2 and 3. The equivalent reduction in hook load was also taken into account, since these values are dependent on each other. The resulting ROP is presented in Fig. 4.10.

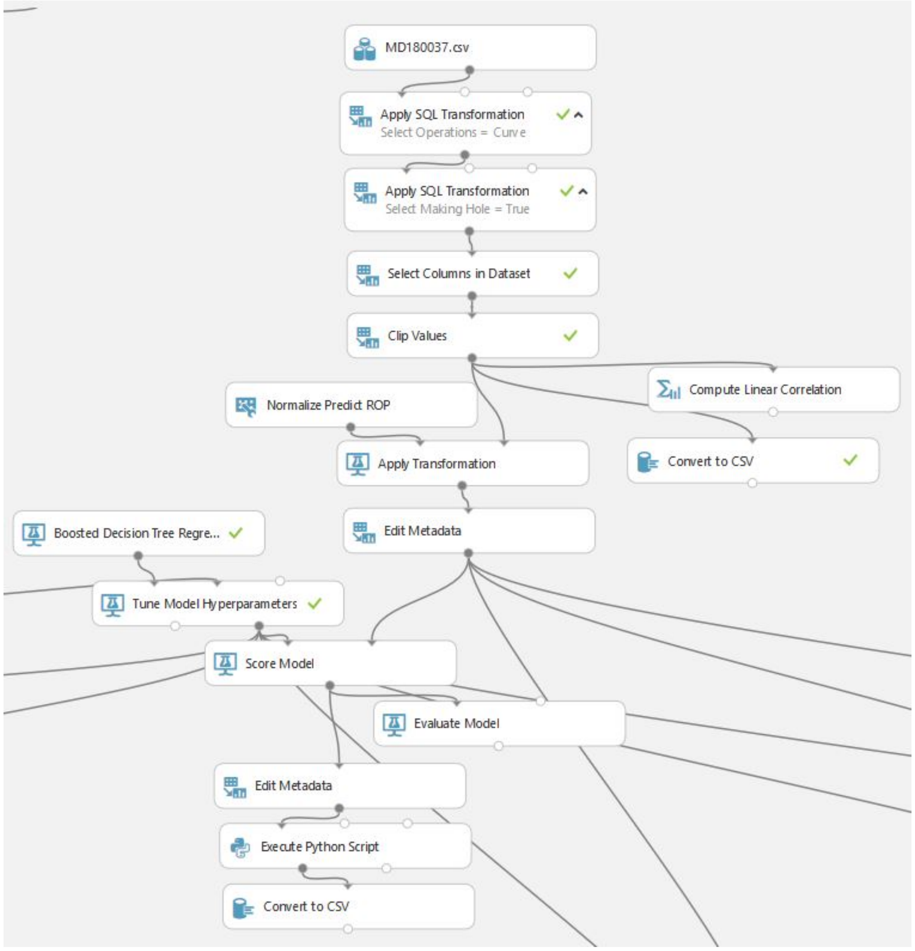


Figure 4.8. Testing of Well 10 Against the Trained Model

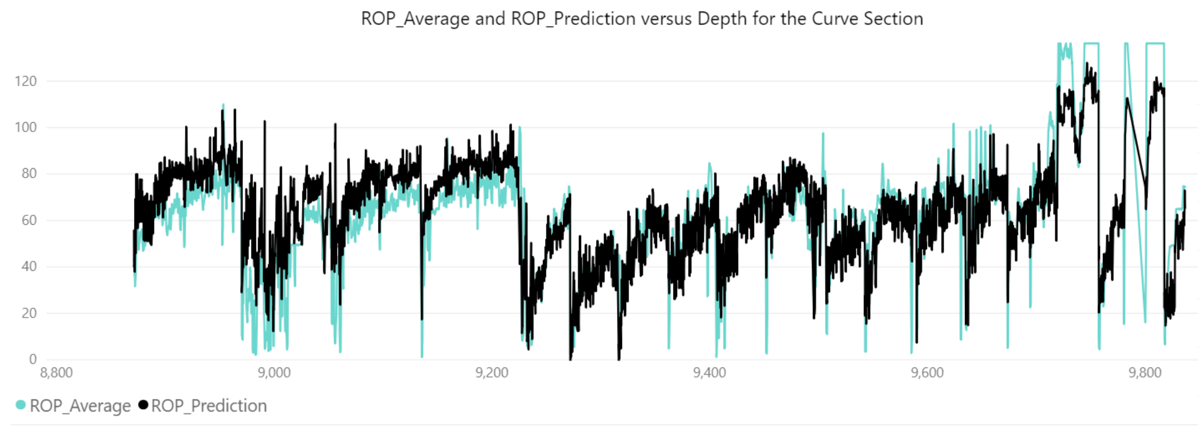


Figure 4.9. Comparison Between the Recorded and Predicted ROP for the Curve Section of Well 10

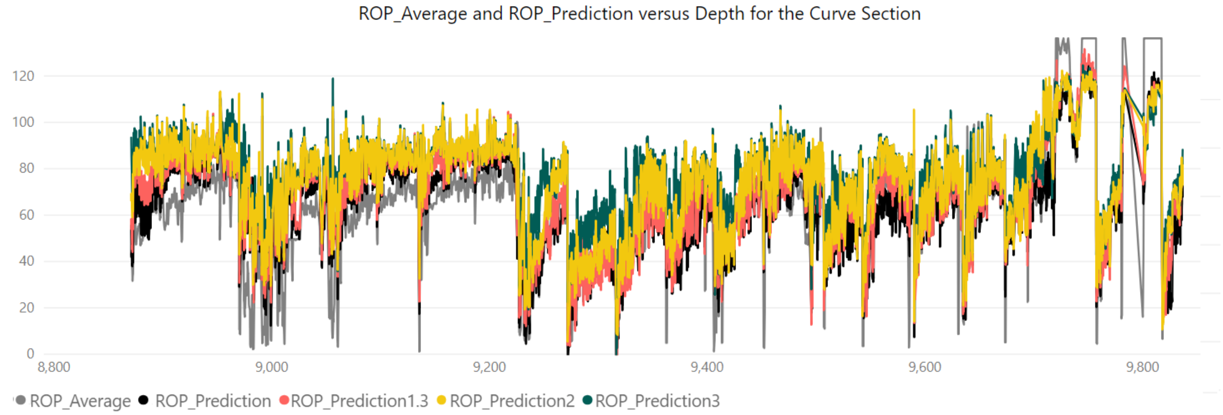


Figure 4.10. Resulting ROP Prediction for ROP of 1.3, 2 and 3 Times Higher

Fig. 4.10 presents an interesting result. The model shows that the ROP will not increase indefinitely with increase of WOB. For example, from the performance of Well 10, it can be seen that a somewhat steady performance is present from 8,900 until 9,250 feet. After this point, the ROP becomes very unsteady. Similarly, for a doubled value of the WOB (yellow curve), the ROP seems to be the best in this first interval (up to 9,250 feet). However, for the troubled part, it seems that more WOB (as represent by the green line, with a multiplier of 3) would have resulted in a better ROP. To assess this result further, plots of the two regions are presented in Figs 4.11 to 4.14.

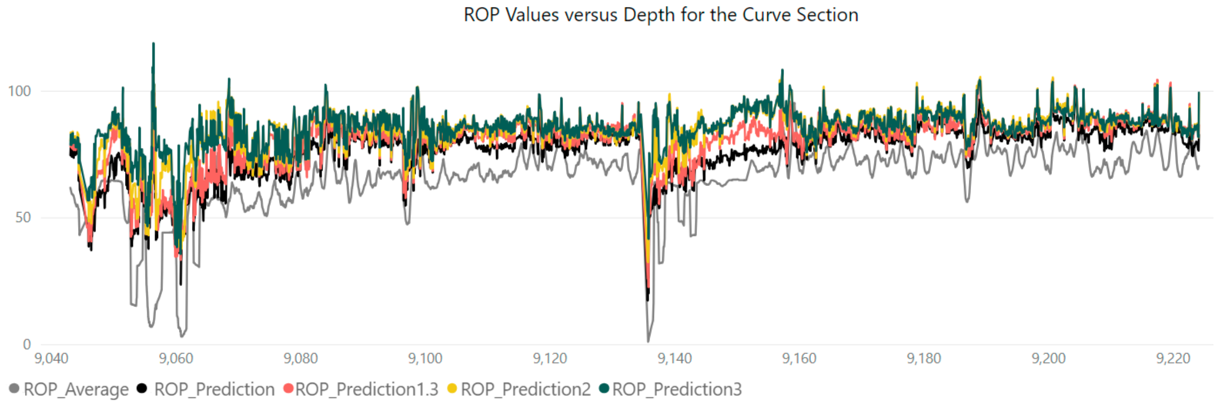


Figure 4.11. Resulting ROP Predictions Zoomed Between 9,040 and 9,250 feet

Figs. 4.11 and 4.12 show a zoom in the first part of Fig. 4.10. From the drilling parameters plot, we can see that during the two stands drilled (observed by the red line of Block Height), all important parameters were steady, namely the Differential Pressure,

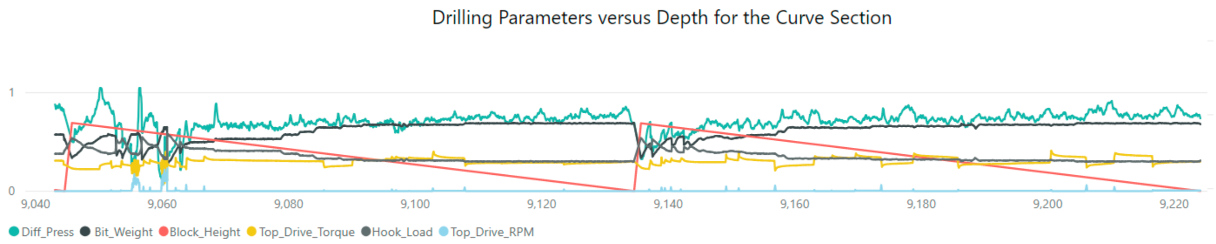


Figure 4.12. Drilling Parameters Zoomed Between 9,040 and 9,250 feet

the WOB and HL, top drive rpm and torque. This shows sections that presented efficient drilling, and an increase in the WOB, for example, would not increase rate of penetration. This is exactly what is seen in Fig. 4.11, where all predicted lines with WOB increase (the red, yellow and green lines) show almost identical values of ROP.

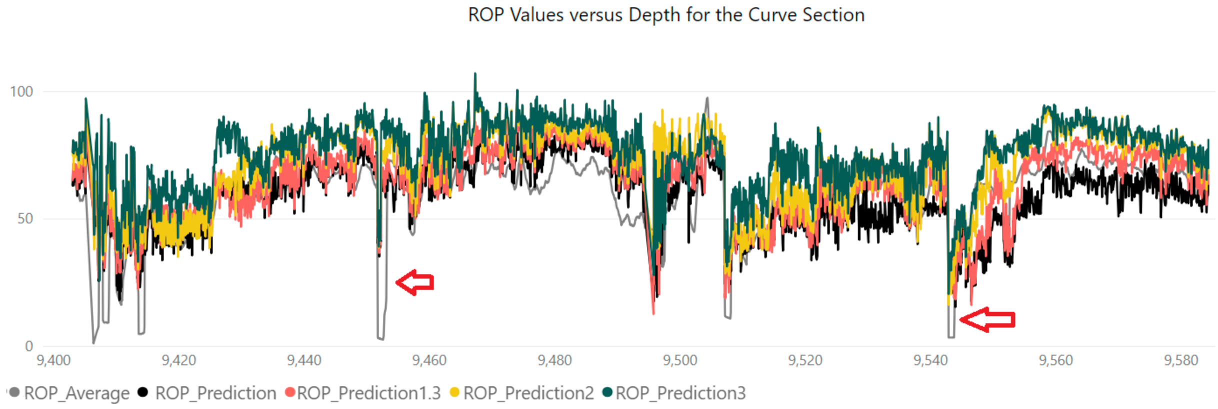


Figure 4.13. Resulting ROP Predictions Zoomed Between 9,400 and 9,600 feet

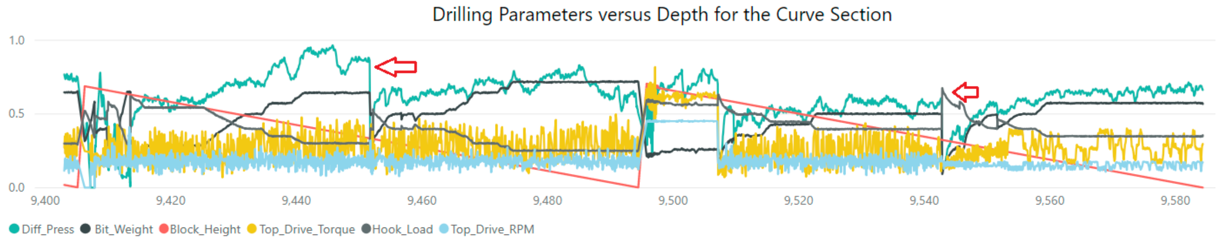


Figure 4.14. Drilling Parameters Zoomed Between 9,400 and 9,600 feet

Figs. 4.13 and 4.14 show now a zoom at the second half of Fig. 4.10, where the ROP was very inconsistent. Further examination of the drilling parameters show that around half of each drilling stand, something happened that resulted in a sudden drop of ROP. The behavior seen is a increase in WOB and Differential Pressure, high values of oscillating torque and a surface RPM greater than zero (note that the drilling parameters plot has the values normalized between 0 and 1 to be used in the model, as described previously). This behavior represents an attempt to overcome the torque and drag by rocking the drill bit, and the sudden drop in ROP is when the mud motor stalls, as explained by [Duplantis, 2016]: "At the driller's console, an impeding stall might be indicated by an increase in WOB but with no corresponding upsurge in downhole pressure to signal that an increase in downhole WOB has actually occurred. At some point, the WOB indicator will show an abrupt decrease, indicating a sudden transfer of force from the drill string to the bit". The red arrows in Figs. 4.13 and 4.14 denote the moment when the mud motor stalled. This pattern of motor stalling around the half point of the drilling stand is consistent along the second half of the curve section of Well 10. The predicted ROP from the model shows a better separation between the lines of higher WOB, suggesting that an increase in weight on bit could help prevent this issue, resulting in better performance. However, as explained in [Duplantis, 2016], [Maidla et al., 2004] and [Maidla et al., 2005], the most important parameter to overcome torque is the control of string rotation from the surface to counterbalance the reacting torque applied to the lower portion of the drill string.

The resulting ML application example in drilling could result in another ML model, one that can predict mud motor stalling. Because of the distinguishable pattern of a motor stall in Fig. 4.14, a new model could be trained in a data set where the points prior to the stall are labeled as such, as well as the actual moment when the motor stall occurs. With a model like that calibrated, a real implementation of it could help the driller predict when a motor stall is going to happen, raising a flag that can result in action, resulting in a longer motor life and wellbore quality.

Chapter 5
Conclusions and Future Directions

5.1. Conclusions

From the work of this thesis we can conclude that:

- Data Wrangling is the most difficult and time consuming task in a Data Science workflow.

- A detailed workflow to guide a drilling engineer to clean drilling data is presented in this work. This can result in significant time savings from the raw data collection to data analysis.

- Once the drilling engineer has the data cleaned and structured, a vast number of analysis can be performed with the data to extract valuable information about previous drilling performance.

- The operation of the drill bit is the most important factor in drilling. Thus the close monitoring of actual and previous operational parameters can result in significant performance gains.

- Most companies already have the data to perform data analytics in drilling. The missing part usually is an organized workflow to treat the massive data influx, and a systematic way to evaluate previous performance against well defined metrics.

5.2. Future Directions

Directions for a future work following what was done here are:

- The zero WOB and Diff Pressure algorithm needs a fine tune in its logic, since noticing that it missed the zeroing point in some drilling stands. Proper zeroing is important to correct the parameters.

- The downhole data (MWD) was available but not included in the scope of the current work. The addition of the downhole parameters (bit vibration, rotation, loads, etc) plus the geometry of the well (inclination, dog leg, tortuosity) could improve even further the analysis performed. With the 3D plot of the well geometry, the drilling parameters can be plotted in the space domain. Plotting wells drilled in the same pad in three dimensions can provide insights of the directions where the trouble formations are, helping in the well plan for future wells in the same pad.

- The machine learning modelling and analysis can be expanded in the prediction of drilling problems, if the problems are properly categorized. Thus, it is recommended to accurately identify the time portions where drilling dysfunction happened to understand the causes of drilling problems with the aid of ML techniques.

- A natural sequence of data analysis and ML modelling would be to upscale it to real-time data. Once the analysis and model are considered accurate enough for real time application, the interested company can install the structure needed for real time transmission of the data, as referenced in section 1.7.3.

References

[Al-Khudiri et al., 2008] Al-Khudiri, M., Al Shehry, M., and Curtis, J. (2008). Data Architecture of Real-Time Drilling and Completions Information at Saudi Aramco.

[Al-khudiri et al., 2015] Al-khudiri, M. M., Al-sanie, F. S., Paracha, S. A., Miyajan, R. A., Awan, M. W., Aramco, S., Kashif, M., and Ashraf, H. M. (2015). Application Suite for 24 / 7 Real Time Operation Centers 2 . Operation Centers ' Systems :.

[AlBar et al., 2018] AlBar, A. H., Alotaibi, B. M., Asfoor, H. M., and Nefai, M. S. (2018). A Journey Towards Building Real-Time Big Data Analytics Environment for Drilling Operations: Challenges and Lessons Learned.

[Almeida Leon et al., 2013] Almeida Leon, A., Hernandez, E., and Perez Bardasz, S. R. (2013). Design of an Automated Drilling Prediction System - Strengthening While-Drilling Decision Making.

[Arnaout et al., 2012] Arnaout, A., Thonhauser, G., Esmael, B., and Fruhwirth, R. K. (2012). Intelligent Real-time Drilling Operations Classification Using Trend Analysis of Drilling Rig Sensors Data.

[Ashok et al., 2018] Ashok, P., Behounek, M., Shahri, M., Chan, H.-C., van Oort, E., Thetford, T., and Saini, G. S. (2018). Spider Bots: Database Enhancing and Indexing Scripts to Efficiently Convert Raw Well Data Into Valuable Knowledge. pages 1–8.

[Bermúdez Martínez, 2012] Bermúdez Martínez, R. (2012). IMPROVING REAL-TIME DRILLING OPTIMIZATION APPLYING ENGINEERING PERFORMANCE FROM OFFSET WELLS. pages 1–15.

[Bolen et al., 2018] Bolen, M., Crkvenjakov, V., and Converset, J. (2018). The Role of Big Data in Operational Excellence and Real Time Fleet Performance Management&mdash;The Key to Deepwater Thriving in a Low-Cost Oil Environment.

[Borjas et al., 2019] Borjas, R., Creegan, A., Perdomo, A., and Shults, D. (2019). Data Quality at the Rigsite - Automated System to Zero Bit Weight and Differential Pressure. In *SPE/IADC International Drilling Conference and Exhibition*. Society of Petroleum Engineers.

[Brannigan and Co, 1992] Brannigan, J. C. and Co, M. O. (1992). The Characterization of Drilling Operations and Their Representation in Relational Databases. *at the Seventh SPE Petroleum Computer Conference held in Houston, Texas, July 19-22*, SPE 24429.

[Brett and Millheim, 1986] Brett, J. and Millheim, K. (1986). The Drilling Performance Curve: A Yardstick for Judging Drilling Performance. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

[Brooks et al., 2017] Brooks, S., Cheatham, C., Kolstad, E., Smith, G., Jarrett, C., Kumar, D., and Smith, M. (2017). Real-Time Drilling Optimization and Rig Activity-Based Models Deliver Best-In-Class Drilling Performance: Case History.

[Cao et al., 2018] Cao, D., Loesel, C., and Paranji, S. (2018). Rapid Development of Real-Time Drilling Analytics System. In *IADC/SPE Drilling Conference and Exhibition*. Society of Petroleum Engineers.

[Damski, 2014] Damski, C. (2014). *Drilling Data Vortex: Where the bits meet the bits*. Genesis Publishing and Services Pty Ltd , Australia.

[Duplantis, 2016] Duplantis, S. (2016). Slide Drilling — Farther and Faster. *Oilfield Review*, (28):50–56.

[Dupriest et al., 2005] Dupriest, F. E., Koederitz, W. L., Totco, M. D., and Company, V. (2005). Maximizing Drill Rates with Real-Time Surveillance of Mechanical Specific Energy.

[Energistics, 2019] Energistics (2019). Website. www.energistics.org. Accessed 16 Apr 2019.

[Eren and Kok, 2018] Eren, T. and Kok, M. V. (2018). A new drilling performance benchmarking: ROP indexing methodology. *Journal of Petroleum Science and Engineering*, 163(July 2016):387–398.

[Eric et al., 2015] Eric, C., Skadsem, H. J., and Kluge, R. (2015). Accuracy and Correction of Hook Load Measurements During Drilling Operations.

[Evangelatos and Payne, 2016] Evangelatos, G. I. and Payne, M. L. (2016). Advanced BHA-ROP Modeling Including Neural Network Analysis of Drilling Performance Data.

[Fayyad et al., 1996] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *Al Magazine*, 17(3):37–54.

[Galarnyk, 2018] Galarnyk, M. (2018). Understanding Boxplots. www.towardsdatascience.com/understanding-boxplots-5e2df7bcbd51. Accessed 19 Apr 2019.

[Halloran et al., 2018] Halloran, S., Hoefling, C., Vinay, N., McMullen, J., Nguyen, D., Isbell, M., Behounek, M., and Mandava, C. (2018). Operators' Group, Rig Contractors, and OEM/Service Company Work to Solve Rig Data Quality Issues.

[Hareland et al., 2014] Hareland, G., Wu, A., and Lei, L. (2014). The Field Tests for Measurement of Downhole Weight on Bit(DWOB) and the Calibration of a Real-time DWOB Model. In *International Petroleum Technology Conference*. International Petroleum Technology Conference.

[Kristjansson et al., 2016] Kristjansson, S. D., Lai, S. W., Wang, J., Tremaine, D., and Neudfeldt, A. (2016). Use of Historic Data to Improve Drilling Efficiency: A Pattern Recognition Method and Trial Results.

[Kyllingstad and Thoresen, 2018] Kyllingstad, Å. and Thoresen, K. E. (2018). Improving Surface WOB Accuracy. In *IADC/SPE Drilling Conference and Exhibition*. Society of Petroleum Engineers.

[Liu et al., 2018] Liu, Y., Kibbey, J., Bai, Y., and Wu, X. (2018). Real-Time Bit Wear Monitoring and Prediction Using Surface Mechanics Data Analytics: A Step Toward Digitization Through Agile Development.

[Maidla et al., 2004] Maidla, E., Haci, M., and Corporation, N. (2004). Understanding Torque : The Key to Slide-Drilling Directional Wells.

[Maidla et al., 2005] Maidla, E., Haci, M., Llc, S., Jones, S., Cluchey, M., Alexander, M., Corp, C., Warren, T., and Corp, T. (2005). Field Proof of the New Sliding Technology for Directional Drilling.

[Maidla et al., 2018] Maidla, E., Maidla, W., Rigg, J., Crumrine, M., Wolf-zoellner, P., and Petroleum, P. T. D. E. (2018). Drilling Analysis Using Big Data has been Misused and Abused.

[Maidla et al., 2009] Maidla, E. E., Haci, M., and Wright, D. (2009). Case History Summary: Horizontal Drilling Performance Improvement Due to Torque Rocking on 800 Horizontal Land Wells Drilled for Unconventional Gas Resources. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

[Mandava et al., 2018] Mandava, C., Guillory, R., Robert, S., Fiffick, W., Davis, D., and Myers, J. (2018). Optimizing Remote Operations Support Using an Effective Real-Time Model for Improved Drilling Performance.

[Millan and Ringer, 2018] Millan, E. and Ringer, M. (2018). A New Workflow for Estimating Bit Wear and Monitoring Drilling Efficiency in Real Time During Drilling Operations.

[Millheim et al., 1998] Millheim, K., Maidla, E., and Kravis, S. (1998). An Example of the Drilling Analysis Process for Extended Reach Wells. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

[Mitchell, 1976] Mitchell, B. (1976). Maximum Permissible Drillbit Weight from Drillcollars in Inclined (Directional) Boreholes. In *SPE Annual Fall Technical Conference and Exhibition*. Society of Petroleum Engineers.

[Neufeldt et al., 2018] Neufeldt, A. C., Lai, S. W., and Kristjansson, S. D. (2018). An Algorithm to Automatically Zero Weight on Bit and Differential Pressure and Resulting Improvements in Data Quality.

[Noshi et al., 2018] Noshi, C. I., Assem, A. I., and Schubert, J. J. (2018). The Role of Big Data Analytics in Exploration and Production: A Review of Benefits and Applications.

[Noshi and Schubert, 2018] Noshi, C. I. and Schubert, J. J. (2018). The Role of Machine Learning in Drilling Operations; A Review. (October):7–11.

[NOV, a] NOV. 12-P-160 Triplex Mud Pump. www.nov.com/Segments/Rig_Systems/Land/Drilling_Fluid_Equipment/Mud_Pumps/Triplex_Mud_Pumps/12-P-160_Triplex_Mud_Pump.aspx. Accessed on 19 Apr 2019.

[NOV, b] NOV. Top Drive Systems Web Page. www.nov.com/WorkArea/DownloadAsset.aspx?id=30945. Accessed 16 Apr 2019.

[NOV, c] NOV. Traveling Blocks Web Page. www.nov.com/Segments/Rig_Systems/Land/Hoisting/Traveling_Equipment/Traveling_Blocks.aspx. Accessed 16 Apr 2019.

[Pastusek et al., 2019] Pastusek, P., Co, E. D., Payette, G., Upstream, E., and Shor, R. (2019). SPE / IADC-194082-MS Creating Open Source Models , Test Cases , and Data for Oilfield Drilling Challenges Background - Existing Open Source and Model Examples Hydraulics & Hole Cleaning. pages 1–38.

[Reddicharla, 2015] Reddicharla, N. (2015). Automation of Down-hole Completion Data Management Workflow and Visualization.

[Serapião et al., 2007] Serapião, A. B. S., Tavares, R. M., Mendes, J. R. P., and Guilherme, I. R. (2007). Classification of petroleum well drilling operations using Support Vector Machine (SVM). *CIMCA 2006: International Conference on Computational Intelligence for Modelling, Control and Automation, Jointly with IAWTIC 2006: International Conference on Intelligent Agents Web Technologies ...*, (January).

[Spivey et al., 2017] Spivey, B. J., Payette, G. S., Wang, L., Upstream, E., and Jeffrey, R. (2017). Challenges and Lessons from Implementing a Real-Time Drilling Advisory System.

[Teale, 1965] Teale, R. (1965). The concept of specific energy in rock drilling. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, 2(1):57–73.

[Thetford et al., 2017] Thetford, T. S., Chintapalli, A., Ambrus, A., Ashok, P., Behounek, M., Nelson, B., and Ramos, D. (2017). A Novel Probabilistic Rig Based Drilling Optimization Index to Improve Drilling Performance. pages 1–14.

[Tukey, 1962] Tukey, J. W. (1962). The future of daya analysis. *The Annals of Mathematical Statistics*, 33(1):1–67.

[Ulterra, 2019] Ulterra (2019). Product Portfolio. www.ulterra.com/wp-content/uploads/2014/07/Ulterra-Drill-Bit-Portfolio-online.pdf. Acessed 19 Apr 2019 .

[van Oort et al., 2018] van Oort, E., Saini, G., Chan, H., Ashok, P., and Isbell, M. R. (2018). Automated Large Data Processing: A Storyboarding Process to Quickly Extract Knowledge from Large Drilling Datasets.

[Vieira, 2009] Vieira, J. L. (2009). *Controlled directional drilling.* Petroleum Extension Service, University of Texas at Austin.

[Willis and Jackson, 2018] Willis, J. B. and Jackson, R. (2018). Measuring Land Drilling Performance. (March):6–8.

[Yu et al., 2018] Yu, Y., Chambon, S., Liu, Q., and Belaskie, J. P. (2018). Recorded Well Data Enriches the Testing of Automation Systems by Using a Deep Neural Network Approach.

```
import numpy as np
import pandas as pd
import operator
import math



def create_status_col_slice(file,overview):
    #function to create the three status columns Making Hole, Block Movement and Off Bottom Distance
    #Also, this function slices the raw EDR file between the Well Spud and Well TD values from the MDB

    df = pd.read_csv(file)

    #Rename columns
    df.columns = df.columns.str.replace(” − ”, ”_”)
    df.columns = df.columns.str.replace(” ”, ”_”)

    #gets the name of the file to find the well file in the overview table
    wellid = file[:-4]
    #convert some columns in the overview table to datetime so the comparison below can work
    overview.iloc[:,[7,8,9,10,11,12,16,18]] = overview.iloc[:,[7,8,9,10,11,12,16,18]].apply(pd.to_datetime)

    #collect the start and end values to compare
    start = overview.Well_Spud[overview.Well_ID == wellid]
    start = start.to_string(index = False)
    end = overview.Well_TD[overview.Well_ID == wellid]
    end = end.to_string(index = False)

    #slicing the dataset
    df = df[df['Date_Time'] > start]
    df = df[df['Date_Time'] < end]

    #clean the null values
    df = df.replace(-999.25,np.NaN)
    #make null if bit position is greater than hole depth by more than 0.02 ft, then drop them
    df = df.mask(df.Hole_Depth - df.Bit_Position < -0.02)
    df = df.dropna(subset=['Hole_Depth'])
    #Drop rows with null values in all columns
    df = df.dropna(how = 'all')

    #Now replace any null value remaining with 0
    df = df.fillna(0)

    #Create the status columns
    #Making_Hole compares if Hole_Depth is increasing
    df['Making_Hole'] = df.Hole_Depth.eq(df.Hole_Depth.shift())
    df['Making_Hole'] = np.invert(df['Making_Hole'])

    #Block_Movement simply shows the different between two successive rows
    df['Block_Movement'] = df.Block_Height.diff()

    #Off_Bottom_Dist calculates the bit distance to the bottom of the hole
    df2 = df.iloc[:,[1,2]].diff(axis=1)
    df['Off_Bottom_Dist'] = df2.Bit_Position

    #Create the Well_ID Column, and then put it as the first column
    df['Well_ID'] = wellid
    outputfile = wellid + '.csv'
    df = df[['Well_ID','Date_Time', 'Hole_Depth', 'Bit_Position', 'Bit_Weight', 'Block_Height', 'Diff_Press', 'Gamma_Ray', 'Hook_Load', 'Pump_Pressure', 'ROP_Average', 'Top_Drive_RPM', 'Top_Drive_Torque', 'Flow_In_Rate', '
    Pump_SPM_Total', 'Flow_Out_Rate', 'ROP_Fast', 'Making_Hole', '
    Block_Movement', 'Off_Bottom_Dist']]



    df.to_csv(outputfile, index=False)



#this is a support function for the create bha function
def removeDuplicates(listofElements):

    # Create an empty list to store unique elements
    uniqueList = []

    # Iterate over the original list and for each element
    # add it to uniqueList, if its not already there.
    for elem in listofElements:
        if elem not in uniqueList:
            uniqueList.append(elem)

    # Return the list of unique elements
    return uniqueList

def create_bha(file, bhainfo):
     df = pd.read_csv(file)



     wellid = file[:-4]

     #find the well in bhainfo
     dfrange = bhainfo[bhainfo['Well_ID'] == wellid]

     #create the range index to be input later in the bhanumbers dataframe
     ranges = [(i, j) for i, j in zip(dfrange['Depth_In'], dfrange['Depth_Out'])]
     ranges = [num for elem in ranges for num in elem]
     ranges = removeDuplicates(ranges)

     #create the dataframe of ranges
     bhanumbers = list (enumerate(ranges))
     bhanumbers = pd.DataFrame(bhanumbers).drop(1, axis=1).drop(0,axis=0)

     #apply the ranges created as indexes for later comparison
     bhanumbers.index = pd.IntervalIndex.from_breaks(ranges)
     bhanumbers = bhanumbers.rename(columns={0:'BHA'})

     for i in df.index:
         hole_depth = df.at[i,'Hole_Depth']
         #create the bha column, comparing the depth with the ranges from bhanumbers
         df.at[i,'BHA'] = bhanumbers.loc[hole_depth].BHA

     outputfile = wellid + '.csv'

     df.to_csv(outputfile, index=False)



def create_wellsection(file,overview):
    df = pd.read_csv(file)


    wellid = file[:-4]
    #create kop and lp variables, and convert them to numbers so they can be compared with
    kop = overview.KOP_Depth[overview.Well_ID == wellid]
    kop = kop.to_numpy()
    lp = overview.LP_Depth[overview.Well_ID == wellid]
    lp = lp.to_numpy()

    for i in df.index:
        hole_depth = df.at[i,'Hole_Depth']

        #create Well Section column, populating its values based on the depth
        if hole_depth < kop:
            df.at[i,'Well_Section'] = 'Vertical'
        elif hole_depth >= kop and hole_depth <lp:
            df.at[i,'Well_Section'] = 'Curve'
        elif hole_depth >=lp:
            df.at[i,'Well_Section'] = 'Lateral'

    outputfile = wellid + '.csv'

    df.to_csv(outputfile, index=False)



def classify_well(file):
    df = pd.read_csv(file)

    for i in df.index:
        #create the variables to be used as operation definer
        hole_depth = df.at[i,"Hole_Depth"]
        making_hole = df.at[i,"Making_Hole"]
        off_bottom_dist = df.at[i,'Off_Bottom_Dist']
        block_mov = df.at[i,'Block_Movement']
        td_rpm = df.at[i,'Top_Drive_RPM']
        total_spm = df.at[i,'Pump_SPM_Total']
        hook_load = df.at[i,'Hook_Load']


        #classify the operation based on the variable values
        if hole_depth > 10000 and making_hole == True and off_bottom_dist > -0.2 and block_mov < 0.1 and td_rpm >5 and td_rpm< 45.1 and total_spm>10:
            df.at[i,'Operation'] = 'Drilling/Sliding Rocking'
        elif making_hole == True and off_bottom_dist > -0.2 and block_mov < 0.1 and td_rpm >10 and total_spm>10:
            df.at[i,'Operation'] = 'Drilling'
        elif making_hole == True and off_bottom_dist > -0.2 and block_mov < 0.1 and td_rpm <=10 and total_spm>10:
            df.at[i,'Operation'] = 'Sliding'
        elif making_hole == False and off_bottom_dist <= -0.2 and block_mov < 0 and td_rpm >10 and total_spm>10:
            df.at[i,'Operation'] = 'Reaming'
        elif making_hole == False and off_bottom_dist <= -0.2 and block_mov > 0 and td_rpm >10 and total_spm>10:
            df.at[i,'Operation'] = 'Back Reaming'
        elif making_hole == False and off_bottom_dist <= -0.2 and hook_load <=57:
            df.at[i,'Operation'] = 'In Slip Connection'
        elif making_hole == False and off_bottom_dist <= -80 and block_mov < 0 and td_rpm < 5 and total_spm<5 and hook_load>57:
            df.at[i,'Operation'] = 'Tripping In'
        elif making_hole == False and off_bottom_dist <= -80 and block_mov > 0 and td_rpm < 5 and total_spm<5 and hook_load>57:
            df.at[i,'Operation'] = 'Tripping Out'
        elif making_hole == False and off_bottom_dist <= -0.2 and total_spm > 10:
            df.at[i,'Operation'] = 'Circulating/Survey'
        else:
            df.at[i,'Operation'] = 'Other'

    df.to_csv(file, index=False)


def zero_wob_pdiff(file):
    df = pd.read_csv(file)

    #Create the variables that will be used to trigger the zeroing logic
    newstring = False
    pressure = 0.0
    hlerror = 0.0
    #Create the new column, wich zero values
    df['HL_Error'] = 0.0
    df['ZWOB'] = 'Normal'
    for i in df.index:
        #create the variables to be used as status indicator for the best moment to zero WOB and Pdiff
        operation = df.at[i,"Operation"]
        block_height = df.at[i,"Block_Height"]
        off_bottom_dist = df.at[i,'Off_Bottom_Dist']

        td_rpm = df.at[i,'Top_Drive_RPM']
        total_spm = df.at[i,'Pump_SPM_Total']
        hook_load = df.at[i,'Hook_Load']

        #This is to reset the condition variable of a new string, which will trigger the next elif as soon as we start reaming
        if operation == 'In Slip Connection':
            newstring = True
        #This checks if drilling operation started before the ZWOB could be triggered, avoiding late zeroing with wrong measures
        elif newstring == True and (operation == 'Drilling' or operation == 'Sliding' or operation == 'Drilling/Sliding Rocking'):
            newstring = False

        #This statement checks all the conditions for the zeroing algorithm to update the Error and the SP values
        elif operation == 'Reaming' and off_bottom_dist <= -2 and total_spm > 50 and hook_load > 100 and block_height > 80 and td_rpm > 15 and newstring == True:
            hlerror = abs(df.at[i,'Bit_Weight'])
            pressure = df.at[i,'Pump_Pressure']
            df.at[i,'ZWOB'] = 'Zero'
            newstring = False

        elif operation == 'Tripping Out':
            hlerror = 0
            pressure = 0

        #fill the HLERROR and SP for each row, no matter if they were updated or not
        df.at[i,'Static_Pressure'] = pressure
        df.at[i,'HL_Error'] = hlerror


    #After creating the new variables HlError and SP pressure to the whole dataset, we calculate the indicator columns SW, SWcorr, WOBcorr and Diffpresscorr
    df['String_Weight'] = df['Hook_Load'] + df['Bit_Weight']
    df['String_Weightcorr'] = df['String_Weight'] + df['HL_Error']
    df['Bit_Weightcorr'] = df['String_Weightcorr'] - df['Hook_Load']
    df['Diff_Presscorr'] = df['Pump_Pressure'] - df['Static_Pressure']

    df.to_csv(file, index=False)
```

```

def remove__outliers(file):
        df = pd.read_csv(file)

        #Clean the remaining negative values for WOB
        df.loc[df.Bit_Weight < 0, 'Bit_Weight'] = 0

        #Clean the remaining negative values for Differential Pressure
        df.loc[df.Diff_Press < 0, 'Diff_Press'] = 0

        #Fix Gamma Ray Values, setting values higher than 200 as 200
        df.loc[df.Gamma_Ray < 0, 'Gamma_Ray'] = 0
        df.loc[df.Gamma_Ray > 200, 'Gamma_Ray'] = 200

        #Fix very high flow in values
        df.loc[df.Flow_In_Rate > 1100, 'Flow_In_Rate'] = 1100

        #Fix very high Pump SPM Values
        df.loc[df.Pump_SPM_Total > 300, 'Pump_SPM_Total'] = 300

        df.to_csv(file, index=False)



def create_mse(file,bhainfo):
    df = pd.read_csv(file)


    wellid = file[:-10]
    for i in df.index:
        #get the diameter from the bhainfo, using the rowsize and bha values for each row
        diam = bhainfo.Hole_Size.loc[operator.and_(bhainfo.Well_ID == wellid,
    bhainfo.Well_BHA == df.BHA[i])].to_numpy()

        #create the variables to use in the mse calculation
        td_rpm = df.at[i,'Top_Drive_RPM']
        td_torque = df.at[i,'Top_Drive_Torque']
        rop_avg = df.at[i,'ROP_Average']
        wob = df.at[i,'Bit_Weight']


        #calculate MSE, checking if ROP is not zero, so we will not divide by
    zero
        if rop_avg != 0:
            df.at[i,'MSE'] = ((480 * td_torque * td_rpm)/(diam**2 * rop_avg))
   + ((4*wob)/(diam**2 * math.pi))
        else:
            df.at[i,'MSE'] = 0

    outputfile = wellid + '.csv'

    df.to_csv(outputfile,index=False)




def convert_to_ft_avg(file):
    #lambda function to approximate values to the nearest 0.5 decimal
    nearest_half = lambda x: round(x * 2) / 2

    df = pd.read_csv(file)

    #group the hole depth values to its nearest half, and then calculate the
   mean of all columns
    #EX: depths 401.00 and 401.02 columns will be averaged into 401.00, and
```

```
401.25, 401.40, 401.59 will be averaged into 401.5
 #Since we do not want to average all values at the same hole depth, we filter values when making hole equals true
 df = df.groupby(nearest_half(df[df['Making_Hole'] == True]['Hole_Depth'])).mean()
 df = df.drop('Hole_Depth',axis=1).reset_index()

 df.to_csv(file,index=False)
```

Vita

Daniel Cardoso Braga, born in São Paulo, Brazil, worked as a Drilling Performance Engineer for Seadrill after receiving his bachelor's degree in Mechanical Engineering from University of Campinas. His work experience is focused in performance monitoring of ultra-deepwater drilling rigs. As his interest for data science grew, he decided to pursue his master's degree at the Craft & Hawkins Department of Petroleum Engineering at Louisiana State University. Upon completion of his master's degree, he wants to apply the knowledge obtained to optimize the drilling of oil wells.