The Thesis Committee for Scott Pine Wallace
Certifies that this is the approved version of the following thesis:

ROP in Horizontal Shale Wells:
Field Measurements, Model Comparisons,
and Statistical Learning Predictions

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: _________________________________
Kenneth E. Gray

_______________________________
Hugh C. Daigle
ROP in Horizontal Shale Wells:
Field Measurements, Model Comparisons,
and Statistical Learning Predictions

By:
Scott Pine Wallace, B.S.

Thesis
Presented to the Faculty of the Graduate School of
The University of Texas at Austin
In Partial Fulfillment
Of the Requirements
For the Degree of

Master of Science in Engineering

The University of Texas at Austin
August 2015
Dedication

To my family, friends, and everyone who has helped and encouraged me along my long and non-traditional academic career.
Acknowledgements

The experience of conducting my graduate studies in the Petroleum Engineering department at the University of Texas at Austin has been a fantastic opportunity for personal and academic growth. I have been soundly impressed by the entire program, from the outstanding professors to the innovative research, as well as the caliber of my fellow students within the graduate program.

I am particularly grateful to my research advisor, Dr. Kenneth E. Gray, who has been a source of solid advice and guidance since my first meeting with him before starting my time at UT. Dr. Gray’s knowledge about this subject matter stretches over the duration of the field’s development, and his perspective has been essential to guide the shape of my research and work within his program. I would also like to thank Dr. Hugh C. Daigle for being the second reader of my thesis.

I would also like to thank the sponsors of the Wider Windows Industrial Affiliate Program, which are: British Petroleum, Chevron, ConocoPhillips, Marathon Oil, National Oilwell Varco, Occidental Oil and Gas, and Shell. These sponsors have been key to the IAP’s success by providing guidance to the projects, and attending our program review meetings where they have offered constructive comments on our research. I would like to thank National Oilwell Varco and Marathon Oil in particular for providing field data to the research group.

Finally, I would like to thank the other students of the Wider Windows IAP: Tyler Adams, Lucas Barros, Yongcun Feng, Arjang Gandomkar, Bishwas Ghimire, Chiranth
Hegde, Anthony Ho, Xiaorong Li, Ebrahim Rasromani, Cesar Soares, and Peidong Zhao.

Our collaboration and discussions have been very useful, and I wish you all the best in your academic and professional careers.
Abstract

ROP in Horizontal Shale Wells:
Field Measurements, Model Comparisons,
and Statistical Learning Predictions

Scott Pine Wallace, M.S.E.
The University of Texas at Austin, 2015

Supervisor: Kenneth E. Gray

Rate of Penetration (ROP) is one of the most important indicators of drilling efficiency available to drillers and engineers. Optimizing the ROP on a well allows the operator to decrease the amount of time spent drilling, which reduces cost. Further reductions in cost can come from utilizing and accurate performance model to understand whether a trip to the surface for a new bit is necessary, or if a bit trip would just increase Non Productive Time (NPT) without significantly benefitting performance. Clearly, understanding the factors that affect ROP is an essential part of drilling a successful well.

Models for ROP have been developed over the academic history of Petroleum Engineering. One of the first models was the model developed by Bingham (1964), which offered a simple formula relating the RPM, Weight on Bit (WOB), and the diameter of the bit to a calculated value of ROP. Further work has continued in ROP modeling by Bourgoyne and Young (1974), who created a much more detailed ROP model including eight input parameters, Hareland and Rampersad (1994), who developed a drag-bit
specific model, and Motahhari et al. (2010), who developed a model specific to wells drilled with a positive displacement motor (PDM) and a polycrystalline diamond compact (PDC) bit. These models have a varying number of input parameters, and each rely on the tuning of between three and eight empirical coefficients in order to optimize them to the well which is being studied.

This study applies these traditional ROP models to data collected while drilling modern horizontal shale wells. These wells were drilled with a rotary steerable system, as well as a downhole PDM, and PDC bit. The traditional models were first fit to the drilling data by using the full range of the horizontal section of the well to optimize the empirical coefficients. This method resulted in the traditional models acting largely like a moving average of the drilling performance over the horizontal region. Then, the empirical coefficients were optimized based on 50 ft sections of the horizontal region, which produced a much tighter fit between the calculated and actual ROPs. However this fitting methodology was found to be erroneous, since it was generating a forced overfit of the model to the actual data.

Finally, the Wider Windows Statistical Learning Model was applied to the drilling data. This produced the best fits of any of the models which were considered, and was the only one of the models which followed the high-frequency changes in the actual ROP data. As a result, this was the only one of the models which could be considered accurate for not only the estimation, but also the prediction and optimization of ROP in horizontal shale wells.
Table of Contents

Chapter 1 - Introduction ........................................................................................................... 1
  1.1 - Motivation .................................................................................................................... 1
  1.2 - Thesis Organization .................................................................................................... 2

Chapter 2 – Literature Review and Background ....................................................................... 3
  2.1 - Traditional ROP Models ............................................................................................ 3
    2.1.1 - Bingham’s ROP Model (1964) ........................................................................... 4
    2.1.2 - Bourgoyne and Young ROP Model (1974) ....................................................... 12
    2.1.3 - Hareland and Rampersad’s Drag Bit ROP Model (1994) ...................... 18
    2.1.4 - Motahhari et al. Positive Displacement Motor and PDC Bit ROP Equation (2010) ........................................................................................................... 21
    2.1.5 - Review of Traditional Models ........................................................................... 25
    2.2 - Statistical Modeling and Drilling .............................................................................. 26
    2.3 - Related Work within Wider Windows IAP .............................................................. 27

Chapter 3 - Data Sources and Management .......................................................................... 29
  3.1 - Data Compression ....................................................................................................... 30
  3.2 - Properties of the Wells ............................................................................................... 35
    3.2.1 - Well #1 Information ......................................................................................... 36
    3.2.2 - Well #2 Properties ......................................................................................... 39
    3.2.3 - Well #3 Properties ......................................................................................... 41
Chapter 4 - Traditional Model Fits ................................................................. 45

4.1 - Fitting Models with Full Range of Lateral Data ................................. 45

4.1.1 - Bingham Model Fit with Full Range of Lateral Data ...................... 45

4.1.1.1 - Well #1 Bingham Model Fit with Full Range of Lateral Data...... 48

4.1.1.2 - Well #2 Bingham Model Fit with Full Range of Lateral Data..... 50

4.1.1.3 - Well #3 Bingham Model Fit with Full Range of Lateral Data...... 52

4.1.2 - Bourgoyne and Young Model Fit with Full Range of Lateral Data.... 54

4.1.2.1 - Well #1 Bourgoyne and Young Model Fit with Full Range of Lateral Data ........................................................................................................ 57

4.1.2.2 - Well #2 Bourgoyne and Young Model Fit with Full Range of Lateral Data ........................................................................................................ 58

4.1.2.3 - Well #3 Bourgoyne and Young Model Fit with Full Range of Lateral Data ........................................................................................................ 60

4.1.3 - Hareland and Rampersad Model Fit with Full Range of Lateral Data 62

4.1.3.1 - Well #1 Hareland and Rampersad Model Fit with Full Range of Lateral Data ........................................................................................................ 64

4.1.3.2 - Well #2 Hareland and Rampersad Model Fit with Full Range of Lateral Data ........................................................................................................ 65

4.1.3.3 - Well #3 Hareland and Rampersad Model Fit with Full Range of Lateral Data ........................................................................................................ 67
List of Tables

Table 1 - Bingham Input Parameters ........................................................................................................ 8
Table 2 - Bourgoyne and Young Input Parameters .................................................................................. 14
Table 3 - Hareland and Rampersad Input Parameters ......................................................................... 20
Table 4 - Motahhari et al. Input Parameters ......................................................................................... 22
Table 5 - Summary of Traditional Model Input Parameters .............................................................. 26
Table 6 - BHA Ranges for Well #3 ......................................................................................................... 42
Table 7 - Bingham Model Input Parameters and Data Types .............................................................. 46
Table 8 - Bourgoyne & Young Model Input Parameters and Data Types .......................................... 56
Table 9 - Hareland and Rampersad Model Input Parameters and Data Types .......................... 63
Table 10 - Motahhari et al. Model Input Parameters and Data Types .............................................. 70
Table 11 - RMSE Summary for Traditional ROP Models with Full Range of Data .................. 77
Table 12 - RMSE Summary for Traditional ROP Models with Increased Coefficient Resolution ........................................................................................................................................................................ 99
Table 13 - Empirical Coefficients for Hareland Model from First 10 Sections of Increased Coefficient Resolution Data Table ........................................................................................................ 100
Table 14 - RMSE Values for All Models and Wells, Including Averages ......................................... 104
Table 15 - Input Parameters for Wider Windows Statistical Learning Model .......................... 108
Table 16 - RMSE Values of Traditional Models Compared to WWSLM ....................................... 113
List of Figures

Figure 1 - Bingham Well Comparison ................................................................. 7
Figure 2 - Drilling Performance Differences Between Formations .................... 8
Figure 3 - Bingham Correlating W/D and ROP ............................................... 9
Figure 4 - Bingham Unit Bit Loading per Revolution ...................................... 10
Figure 5 - Bingham Regions of Drilling ......................................................... 11
Figure 6 - Hareland and Rampersad Cutter and Rock Interaction .................. 19
Figure 7 - Motahhari et al. ROP Predictions .................................................. 24
Figure 8 - Comparison of the Drilling Performance of Three Horizontal Wells in the Eagle Ford .......................................................................................... 30
Figure 9 - Driving Parameters for Compression Program ............................... 33
Figure 10 - Intermediary Checks for Compression Program ............................. 34
Figure 11 - Drilling Program Profile View for Well #1 .................................... 36
Figure 12 - Input Data for Well #1 ................................................................. 38
Figure 13 - Drilling Program Profile View for Well #2 .................................... 39
Figure 14 - Input Data for Well #2 ................................................................. 41
Figure 15 - Drilling Program Profile View for Well #3 .................................... 42
Figure 16 - Input Data for Well #3 ................................................................. 44
Figure 17 - Bingham Calculated ROP over Actual ROP for Well #1 ................. 50
Figure 18 - Bingham Calculated ROP over Actual ROP for Well #2 ............... 52
Figure 19 - Bingham Calculated ROP over Actual ROP for Well #3 ............... 54
Figure 20 - Bourgoyne & Young Calculated ROP over Actual ROP for Well #1 ............... 58
Figure 21 - Bourgoyne & Young Calculated ROP over Actual ROP for Well #2 ............ 60
Figure 22 - Bourgoyne & Young Calculated ROP over Actual ROP for Well #3 ............ 62
Figure 23 - Hareland & Rampersad Calculated ROP over Actual ROP for Well #1 ........ 65
Figure 24 - Hareland & Rampersad Calculated ROP over Actual ROP for Well #2 ........ 67
Figure 25 - Hareland & Rampersad Calculated ROP over Actual ROP for Well #3 ........ 69
Figure 26 - Motahhari et al. Calculated ROP over Actual ROP for Well #1 ................. 72
Figure 27 - Motahhari et al. Calculated ROP over Actual ROP for Well #2 ................. 74
Figure 28 - Motahhari et al. Calculated ROP over Actual ROP for Well #3 ................. 76
Figure 29 - Straight Line Fit for Well #1 Data .................................................................... 79
Figure 30 - Comparison of Bingham Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #1 ............................................................ 83
Figure 31 - Comparison of Bingham Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #2 ............................................................ 84
Figure 32 - Comparison of Bingham Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #3 ............................................................ 85
Figure 33 - Comparison of Bourgoyne and Young Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #1 ........................................... 87
Figure 34 - Comparison of Bourgoyne and Young Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #2 ........................................... 88
Figure 35 - Comparison of Bourgoyne and Young Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #3............................................................. 89

Figure 36 - Comparison of Hareland and Rampersad Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #1.............................................. 91

Figure 37 - Comparison of Hareland and Rampersad Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #2................................................ 92

Figure 38 - Comparison of Hareland and Rampersad Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #3......................................... 93

Figure 39 - Comparison of Motahhari et al. Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #1..................................................... 95

Figure 40 - Comparison of Motahhari et al. Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #2..................................................... 96

Figure 41 - Comparison of Motahhari et al. Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #3..................................................... 97

Figure 42 - WWSLM Model Fit over Actual ROP for Well #1............................................. 110

Figure 43 - Comparison of Full-Range, Increased Coefficient Resolution, and WWSLM Model Fits .................................................................................................................. 112

Figure 44 - Motahhari et al. Model Compared to WWSLM with Only 500 ft of Training Data ......................................................................................................................... 114
Chapter 1 - Introduction

1.1 - Motivation

This project came about based on the other similar research within the Wider Windows IAP on drilling performance and ROP modeling. There were already studies in place considering the application of traditional models to vertical wells and how the models dealt with changes in parameters that occurred across formation boundaries, as well as efforts to model the torque and drag in these wells. Considering that so many of the wells drilled in domestic USA plays are horizontal shale wells, it became a good task to investigate whether these traditional models were able to capture the peculiarities of this drilling style and be useful in the modern drilling worlds.

As the project progressed, a project on developing a real-time drilling optimization system also sprang up alongside this one, so a special attention was focused on whether these traditional models could not only model the drilling environment in the horizontal shale wells, but also be predictive in nature. In order to be predictive, they would have to be able to respond to the high-frequency changes in the actual ROP values which are seen in drilling data.

This thesis seeks to demonstrate the model fitting of these traditional ROP models to the horizontal shale well drilling data which has been provided to the Wider Windows IAP. Along the way, the input parameters and their effects on the ROP will be inspected and evaluated. The traditional models will be fit with two different methods in an attempt
to find the best way to use them with horizontal shale well data. These results will be compared and evaluated for which are the best options. Finally, the Wider Windows Statistical Learning Model will be tested on this same dataset, and demonstrated as another tool for evaluating drilling data.

1.2 - Thesis Organization

There are seven chapters to this thesis. The second chapter will provide a literature review on the models which have been considered in this study, as well as a brief look at the history of statistical modeling in relation to drilling data. The third chapter contains all of the traditional model fits, first with the full range of the horizontal well, and then with an increased coefficient resolution. Each of the four models is applied to all three of the wells in the project for both of these phases, resulting in 24 different model fits. The fifth chapter shows the model fits for the Wider Windows Statistical Learning Model, demonstrating it as a strong tool for drilling evaluation. The sixth chapter discusses some of the work which is continuing in Wider Windows that is related to this project, while the seventh chapter presents the conclusions of this thesis.
Chapter 2 – Literature Review and Background

2.1 - Traditional ROP Models

Since the speed that a well is drilled can generally be correlated to the cost of drilling the well, optimizing rate of penetration (ROP) has long been a goal of drillers and drilling engineers. As a result, researchers have been trying to build models for ROP and drilling performance since the academic study of drilling began. These models have advanced over the decades, adding more input variables and taking advantage of the increase of drilling data which has become available.

One thing that is consistent with all of the traditional ROP models is their reliance on empirical coefficients. These empirical coefficients are used to modify the input parameters (or sub-functions which are made from the input parameters) based on a weighting. The models vary in their reliance on these empirical coefficients, with between three and eight of them utilized in the models considered by this project. With some of the models there is information on the upper and lower bounds of these empirical coefficients, but for other models this range must be established by engineering judgement. The problems caused by this variability and the ways that it was minimized will be addressed later in this document.

The traditional models can be divided into three categories based on the types of bits which they are designed to model. Some models are designed specifically for tri-cone or drag / polycrystalline diamond compact (PDC) bits, while others are generic in nature.
and could be applied to bits of either type. Since this study is particularly interested in modeling the rate of penetration in horizontal shale wells, only a subset of the available ROP models was considered. In all of the field data from program sponsors which was available for this study, the horizontal sections of shale wells were drilled only with PDC bits. As such, it would not be valid to attempt to fit these sections to the models which are based on tri-cone bits, so only the models which were designed for either a PDC bit or a generic bit type were considered.

The models which were ultimately chosen to be considered in this project were: Bingham’s ROP Model (1964), Bourgoyne and Young’s ROP Model (1974), Hareland and Rampersad’s Drag Bit ROP Model (1994), and Motahhari et al’s Drag Bit ROP Model (2010). These models are defined in greater detail in the following pages.

2.1.1 - Bingham’s ROP Model (1964)

M. Grant Bingham’s model was the first ROP model to gain traction within the academic and professional community. Bingham had been working as an engineer with Shell in South Texas, and his interest with the modeling of drilling performance began when he was tasked with evaluating the then new jet bits which were being introduced to the industry. These jet bits had shown increased performance, but without an understanding of what was occurring down-hole, the performance increases were an anomaly. Bingham’s interest and continued work on this project caused him to be assigned to the Shell Development Company in Houston in 1957, where he continued his studies on improving drilling performance. In 1964, Bingham resigned from Shell and
returned to the academic world at University of Houston, where he began publishing a series of articles in The Oil and Gas Journal which related drilling performance to rock strength and rock drillability.

In the introduction of his journal series, the editors of the journal stated that “Rotary drilling . . . stands as the last remaining art in an industry where virtually all other operations have been reduced to science. A single link remains to be forged before drilling, too, can elevate itself to scientific status. That vital link is a thorough understanding of what happens on the bottom of the hole as the bit rolls over it.” The editors define the challenge of understanding downhole bit dynamics so succinctly, although this is a problem that has continued to try engineers and researchers ever since these first models proposed by Bingham. Even in the modern drilling world, drilling a well still has a form of art to it, as so many of the decisions in real time are based on the experience, feel, and knack of the driller. Bingham recognized that his models would probably not be the end-all of drilling performance evaluation, stating in his article “We are unlikely to find soon an exact and all-encompassing definition of the problem. Drilling is too complex.” Bingham goes on to state that “more than 26 variables are said to influence drilling” although far fewer variables are captured by the equations that he developed.

Bingham’s first task was to establish that drilling performance was even a consistent and predictable phenomenon. He initially looked at high frequency drilling data (for the time) which recorded the ROP at 2-ft intervals, and was faced with the highly
erratic behavior which drilling rate exhibits. Bingham found that over short intervals, there is usually at least a two-fold difference in the ROP, which at the time he attributed to the lack of understanding of hydraulics and cuttings removal. Due to this variability, Bingham dismissed short interval testing since it would require “an unusually uniform, homogeneous section of rock” among other reasons.

Looking at longer intervals proved to be the solution to this variability. When Bingham considered the entire length of a well, he was surprised by how little the on-bottom rotating time varied from well to well. Bingham studied a West Texas field where a “Contractor A” had drilled three wells within one year inside a two mile radius which had each taken 312.0, 309.5, and 312.5 on-bottom rotating hours to drill to a depth of 8,000 ft. “Contractor B” drilled a well in the same field, which took over 400 hours to reach the same depth. This showed Bingham that the consistency achieved by the Contractor A is due to their drilling technique and not just to the variability of the rock. This consistency is plotted below in Figure 1, showing the drilling performance of each contractor in the same field, and leading Bingham to continue to look for the predictability of drilling performance.
Bingham recognized that for his long-interval study to be effective, he needed to break the well into suitable intervals. By plotting the depth of a well over on-bottom time, the slope of the given line represents the ROP of the well. This data was plotted by Bingham as shown below in Figure 2. It can be seen that the slope changes at each of the formation boundaries, and continues to be representative of a lower and lower ROP as the depth increases. Based on this analysis, Bingham decided to evaluate drilling performance independently in each formation, a very reasonable decision which has been continued in nearly all drilling performance studies since, including this one.
Figure 2 - Drilling Performance Differences between Formations (Bingham 1964)

The parameters which are included in the Bingham equation for ROP, as well as his variable names for them and their units are shown below in Table 1. Bingham also initially included the rock strength variable “S (psi)” in his analysis, however he dropped it as in input variable as at the time there was no way to measure the in-situ rock strength while drilling a well.

Table 1 - Bingham Input Parameters

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Abbreviation</th>
<th>Bingham Variable</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of Penetration</td>
<td>ROP</td>
<td>R</td>
<td>ft/sec</td>
</tr>
<tr>
<td>Rotary Speed</td>
<td>RPM</td>
<td>N</td>
<td>rev/sec</td>
</tr>
<tr>
<td>Bit Diameter</td>
<td>---</td>
<td>D</td>
<td>ft</td>
</tr>
<tr>
<td>Bit Load</td>
<td>WOB</td>
<td>W</td>
<td>lb</td>
</tr>
</tbody>
</table>
Bingham was seeking a correlation between these variables that could produce consistent and repeatable results. He first combined the WOB and Bit Diameter terms into the “Unit Bit Loading” parameter (W/D), which represents the pounds of bit loading per foot of bit diameter. He then attempted to correlate R (ROP) with W/D (WOB/Bit Diameter), which produced the plot shown in Figure 3. This plot shows no common limit, with the ROP varying at each of the RPMs which were tested.

Figure 3 - Bingham Correlating W/D and ROP (Bingham 1964)

Bingham then plotted the unit bit loading against the penetration per revolution of the bit, represented by R/N and shown below in Figure 4. This provided a better correlation, including some apparent lines of limiting characteristics. Bingham defined the regions of this plot with a number of terms. The initial line that the penetration parameter starts on from the origin is called the “Performance Region,” and as the data fall off of
this line to the right, they enter the “transport-limit region.” If the ROP remains constant despite increased bit loading, the performance is in the poorest region, the “volume removal region,” which signifies that the bit is balled up, or there is no circulating fluid. These regions are shown on the plot below in Figure 5.

Figure 4 - Bingham Unit Bit Loading per Revolution (Bingham 1964)
Now that a reasonable correlation between the performance and load parameters has been established, Bingham goes on to develop it into an equation. It is here when the first of the dimensionless empirical constants are introduced into this model. Bingham defines these as “a” and “b” and describes that they are different for each formation encountered along the wellbore. He places them in the equation as shown below:

$$\frac{R}{N} = a \left(\frac{W}{D}\right)^b$$

(1)

Bingham identifies that these coefficients have vast variability, with his estimate for the range of valid values for “a” spanning nine orders of magnitude. He states about the variation of a and b between formations that “in my opinion, it is this great variation which has frustrated the prediction of drilling to date.” Bingham goes on to declare that the variability will be addressed by using a simple linear approximations for the intervals that are considered.
The above equation can be rearranged to isolate the R term representing ROP. Additionally, a third empirical coefficient “c” can be added to the RPM term, known as the RPM factor. This presents the final form of the Bingham ROP model, as shown below in Bingham variables in Equation 2 as well as common variable names in Equation 3:

\[ R = a N^c \left( \frac{W}{D} \right)^b \]  

(2)

\[ ROP = a (RPM)^c \left( \frac{WOB}{D_{bit}} \right)^b \]  

(3)

2.1.2 - Bourgoyne and Young ROP Model (1974)

The Bourgoyne and Young model was published ten years after Bingham’s model was released. This model was initially presented at the Society of Petroleum Engineers & American Institute of Mining Engineers Sixth Conference on Drilling and Rock Mechanics, held in Austin, Texas in January of 1973. It was then published in the Transactions journal in 1974 under the title “A Multiple Regression Approach to Optimal Drilling and Abnormal Pressure Detection”. At the time of the publication, A. T. Bourgoyne Jr. was a researcher at Louisiana State University, and F. S. Young Jr. was an engineer in the Baroid division of N L Industries in Houston. In the abstract of their paper, Bourgoyne and Young acknowledge that a number of previous models for drilling performance had been created, but that the limited data which was used to develop these previous models made them inaccurate. Bourgoyne and Young go on to state that “Recent developments in onsite well monitoring systems have made possible the routine determination of the best
mathematical model for drilling optimization...” Bourgoyne and Young emphasize the importance of optimizing drilling performance, as they note that the cost of a well per foot has been increasing at around 7.5 percent per year.

In introducing their model, Bourgoyne and Young relate that many operators of the time were using different models for optimizing bit weight and rotary speed, optimizing jet bit hydraulics, and detecting abnormal pore pressure. It was Bourgoyne and Young’s goal to combine all of these facets of the drilling procedure into a unifying model.

The Bourgoyne and Young model is very comprehensive for its time, and incorporates eight different input variables. These variables are combined in an overarching equation, which is shown in Equation 4 below. In this equation, ROP is represented by the differential equation expression $dD/dt$ for change in depth over time. The exponential function “$e^x$” is represented by “exp.” The summation function combines the products of $a_2x_2, a_3x_3, \ldots, a_8x_8$. Each of terms $a_j$ represent an empirical coefficient which is applied to the formula, while each of the terms $x_j$ represent the input variable for that particular value.

$$\frac{dD}{dt} = \exp \left( a_1 + \sum_{j=2}^{8} a_jx_j \right)$$

(4)

The names, abbreviations, and variable names as referenced in Bourgoyne and Young’s paper are shown in Table 2 below for all of the input variables used in this model.
As mentioned above, the “a” and “x” terms represent empirical coefficients and input variables respectively. Only $a_1$ stands alone as an “a” term without a corresponding “x” term. This $a_1$ term is a catch-all term that represents not only the effect of formation strength on penetration rate, but also any other factors which affect penetration rate and are not captured by the following terms in the model.
The second and third terms $x_2$ and $x_3$ represent the effects of compaction on the penetration rate. These terms assume an exponential decrease in penetration rate with depth in a normally compacted formation. This was an assumption made by Bourgoyne and Young at the time based on research by Murray and Combs, but they acknowledge that this was not verified experimentally yet. Also, the second and third terms were normalized to equal 1.0 for a normally compacted formation at 10,000 ft depth. The equations for $x_2$ and $x_3$ are shown below:

$$x_2 = 10,000 - D$$  \hspace{1cm} (5)

$$x_3 = D^{0.69}(g_p - 9.0)$$  \hspace{1cm} (6)

The $x_4$ term represents the effects of pressure differential on the penetration rate, and is based on an exponential relation between penetration rate and excess bottom-hole pressure up to around 1,000 psi. This term’s equation is as follows:

$$x_4 = D(g_p - \rho_c)$$  \hspace{1cm} (7)

The fifth term models the effect of weight on bit and bit diameter on penetration rate, and does so by assuming that the penetration rate is directly proportional to $(W/d)^{4.5}$. Even though the equation below is in a natural log form, the entire summation is already inside an expression which raises everything as a power of “$e$”, so these operators cancel out. This also assumes a normalization where this entire term is equal to 1.0 when 4,000 lb per inch of bit diameter. Additionally, this term invokes the “threshold bit weight” which is the value of $W/d$ at which the drillstring begins drilling,
and must be determined by drill-off tests in the field or estimated. The equation for $x_5$ is shown below:

$$x_5 = \ln \left[ \frac{\left(\frac{W}{d}\right) - \left(\frac{W}{d}\right)_t}{4 - \left(\frac{W}{d}\right)_t} \right]$$

(8)

The $x_6$ term models the effect of rotary speed on the penetration rate, and assumes that penetration rate is directly proportional to RPM (again with the natural log function canceling out the exp in the overarching equation). This term is normalized so that it equals 1.0 with an RPM of 100.

$$x_6 = \ln \left( \frac{N}{100} \right)$$

(9)

The seventh term includes the effects of bit wear on the penetration rate, and does so by introducing the variable “h” which represents the percentage of bit tooth height which has been worn away. The equation for $x_7$ is very simple:

$$x_7 = -h$$

(10)

The final and eighth term represents the effects of bit hydraulics on the penetration rate. This equation is based on microbit experiments which were performed by Eckel, who discovered that the rate of penetration was proportional to a Reynolds number group raised to the 0.5 power. This method has been included in the Bourgoyne and Young model with the following equation:

$$x_8 = \frac{\rho q}{350 \mu d_n}$$

(11)
Bourgoyne and Young’s model can then be combined into a giant expression which combines each of the variables, which is shown below in Equation 12. This is a rather unmanageable equation however, so it is recommended to look at each of the terms independently and then combine them at the end of the calculation.

\[
\frac{dD}{dt} = \exp\left( a_1 + a_2(D - 10,000) + a_3(D)^{0.69}(g_p - 9.0) + a_4(D)(g_p - \rho_c) \\
+ \ln\left( \frac{W}{d} - \left(\frac{W}{d}\right)_t \right) + \ln\left( \frac{N}{100} \right) + -a_7h + a_8\left( \frac{\rho q}{350 \mu d_n} \right) \right) \tag{12}
\]

Bourgoyne and Young’s equation is very verbose, but depends on the accuracy of the factors \(a_1 - a_8\) to be able to model drilling performance. To find appropriate values for these coefficients, Bourgoyne and Young performed a multiple regression analysis of the equations on drilling data which had been collected on short intervals. In their research, Bourgoyne and Young also conducted a sensitivity analysis and found that at least 30 depth points are needed to perform this regression with accuracy for all of the above parameters. It should be noted though that the data in Bourgoyne and Young’s paper which they fit all of their data to (from an onshore Louisiana Shale field) involves a very generalized set of results. The depth intervals for each entry of data are between 100-3,000 ft, and in each of these intervals the input data appears to have been averaged. Also the data points often start with a new bit run (of which there were 40 in this well).
While Bourgoyne and Young wrote that their study was conducted on small intervals of drilling data, from analyzing their paper, it appears that they followed similar methodology to Bingham by considering the rate of penetration on entire bit runs or formation sections, rather than by considering the high-frequency data of true foot-by-foot resolution.

2.1.3 - Hareland and Rampersad’s Drag Bit ROP Model (1994)

The Hareland and Rampersad model was first presented by the authors at the Latin American / Caribbean Petroleum Engineering Conference in Buenos Aires, Argentina in April of 1994, in a paper titled “Drag-Bit Model Including Wear.” G. Hareland and P.R. Rampersad were both researchers at the New Mexico Institute of Mining and Technology at the time of this research. This model concentrated on bit cutter geometry as the dominant factor which influenced rate of penetration. It also addressed specifically drag bits (including Natural Diamond Bits (NDB), Polycrystalline Diamond Compact Bits (PDC), as well as Geoset Bits), as they had gained significant popularity due to the increased usage of downhole motors as well as their ability to resist wear. Hareland and Rampersad state that while drag bits are easy to describe geometrically, the wide variety of drag bit designs available as well as the sensitivity of drag bits to formation properties and operating conditions make them have a wide range of drilling performance.

As this model is so focused on the cutter geometry, it also deals with approximating the mechanism by which the cutters are actually breaking the rock in their path. In previous research there were differing theories, with Weavind and Dyer
presenting cases for both tensile and brittle failure as well as shearing plastic failure. Additionally, Appl and Rowley showed failure criteria to be plastic when under high confining pressure. For the purposes of this model, Hareland and Rampersad used a conservation of mass approach to modeling the cutter-rock interaction, where the rate of penetration was dependent on the rate of rock removal from the front of the cutters. This analysis considers the group of cutters interacting with a flat surface, and assumes equal loading on each cutter. It also assumes that as the bit is rotating, the rock has been removed on the aft side of each cutter, so it is only in contact with the rock on the surface in front of the advancing cutter. This interaction between the cutter and the rock can be seen below in Figure 6 from Hareland and Rampersad’s paper.

![Figure 6 - Hareland and Rampersad Cutter and Rock Interaction (Hareland and Rampersad 1974)]

From here, Hareland and Rampersad go on to derive the equation for ROP by starting with the assumption that the compressive strength of the rock can be related to the mechanical weight on bit over the projected contact area. From here they develop
the equation to find the penetration of the diamond bit for the specified bit weight and cutter geometry. Then they determine the area being removed at the front of the cutter due to this penetration depth, and finally the rock volume which is removed per revolution of the bit. From this, Hareland and Rampersad are able to solve for the rate of penetration by dividing the removed rock volume by the bit face area. Table 3 below shows all the variables which are used in the Hareland and Rampersad equation, and the final equation that was developed by Hareland and Rampersad is shown below.

Table 3 - Hareland and Rampersad Input Parameters

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Abbreviation</th>
<th>Hareland and RampersadVariable</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Cutters</td>
<td>---</td>
<td>$N_c$</td>
<td>---</td>
</tr>
<tr>
<td>Bit Rotational Speed</td>
<td>RPM</td>
<td>RPM</td>
<td>rev/min</td>
</tr>
<tr>
<td>Bit Diameter</td>
<td>---</td>
<td>$D_b$</td>
<td>in</td>
</tr>
<tr>
<td>Cutter Diameter</td>
<td>---</td>
<td>$d_c$</td>
<td>in</td>
</tr>
<tr>
<td>Weight on Bit per Cutter</td>
<td>---</td>
<td>$W_{mech}$</td>
<td>lbs</td>
</tr>
<tr>
<td>Uniaxial Compressive Rock Strength</td>
<td>---</td>
<td>$\sigma_c$</td>
<td>psi</td>
</tr>
<tr>
<td>Lithology Correction Factor</td>
<td>---</td>
<td>$a$</td>
<td>---</td>
</tr>
<tr>
<td>RPM Correction Factor</td>
<td>---</td>
<td>$b$</td>
<td>---</td>
</tr>
<tr>
<td>WOB Correction Factor</td>
<td>---</td>
<td>$c$</td>
<td>---</td>
</tr>
</tbody>
</table>

$$
ROP = \left( \frac{14.14 N_c \, RPM}{D_b} \right) \left[ \left( \frac{d_c}{2} \right)^2 \cos^{-1} \left( 1 - \frac{4W_{mech}}{N_c d_c^2 \pi \sigma_c} \right) \right] - \left( \frac{2W_{mech}}{N_c \pi \sigma_c} - \frac{4W_{mech}^2}{(N_c d_c \pi \sigma_c)^2} \right) \left( \frac{d_c}{2} - \frac{2W_{mech}}{N_c d_c \pi \sigma_c} \right)^{1/2} \right]
$$

(13)
As with the other traditional models, Hareland and Rampersad’s model also relies on a few empirical coefficients. There are three coefficients in this case, the weight on bit factor, the lithology factor, and the RPM factor. In the case of this model however, they are all combined into one term “COR”, which is shown below:

\[ COR = \frac{a}{(RPM^b)(WOB^c)} \]  

(14)

Including the correction factor, the final Hareland and Rampersad ROP equation is shown below:

\[
ROP = \left( \frac{a}{(RPM^b)(WOB^c)} \right) \left( \frac{14.14 N_c RPM}{D_b} \right) \left[ \left( \frac{d_c}{2} \right)^2 \cos^{-1} \left( 1 - \frac{4W_{mech}}{N_c d_c^2 \pi \sigma_c} \right) \right] \\
- \left( \frac{2W_{mech}}{N_c \pi \sigma_c} - \frac{4W_{mech}^2}{(N_c d_c \pi \sigma_c)^2} \right)^{\frac{1}{2}} \left( \frac{d_c}{2} - \frac{2W_{mech}}{N_c d_c \pi \sigma_c} \right) \]  

(15)

2.1.4 - Motahhari et al. Positive Displacement Motor and PDC Bit ROP Equation (2010)

This model was developed by H.R. Motahhari, G. Hareland, J.A. James, and M. Bartlomowicz, and was first presented at the Canadian International Petroleum conference in June of 2008, before being published by the SPE in 2010 with the title of “Improved Drilling Efficiency Technique Using Integrated PDM and PDC Bit Parameters”. At the time of publication, Motahhari and Hareland were researchers at the University of Calgary, while James and Bartlomowicz were employed by Husky Energy. This model was developed as part of an effort to optimize drilling using positive displacement motors and PDC bits.
Motahhari et al. recognized that the previously developed traditional models for drilling optimization were not designed for PDM motors, which had become very prevalent due to the increase of directional and horizontal drilling operations. PDM motors convert hydraulic energy from the drilling mud to rotational energy by using an assembly of a helical rotor inside a helical stator. The rotor has one less lobe than the stator, creating sealed chambers which are pressurized by the incoming mud, and cause the rotor to be turned eccentrically within the stator. Positive displacement motors with more lobes are able to create higher torque, but cannot provide as much RPM for a given flow rate, so it is important to be able to optimize the PDM selection for the well.

The rate of penetration model which was used in this study was an evolution of Hareland and Rampersad’s 1994 work which made it more suitable for working with positive displacement motors. The variables used in this equation are shown below in Table 4 and the equation itself is shown in the following equation. A lot of the complexity of Hareland’s 1994 model is captured in the $W_f$ term of the Motahhari et al. model.

Table 4 - Motahhari et al. Input Parameters

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Abbreviation</th>
<th>Motahhari et al. Variable</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit Wear Function</td>
<td>---</td>
<td>$W_f$</td>
<td>percent</td>
</tr>
<tr>
<td>Total Bit Rotational Speed</td>
<td>RPM</td>
<td>$RPM_t$</td>
<td>rpm</td>
</tr>
<tr>
<td>Weight on Bit</td>
<td>WOB</td>
<td>WOB</td>
<td>lbs</td>
</tr>
<tr>
<td>Bit Diameter</td>
<td>---</td>
<td>$D_B$</td>
<td>in</td>
</tr>
<tr>
<td>Confined Compressive Strength</td>
<td>CCS</td>
<td>$S$</td>
<td>psi</td>
</tr>
<tr>
<td>Bit Design Coefficient</td>
<td>---</td>
<td>$G$</td>
<td>---</td>
</tr>
<tr>
<td>RPM Correction Factor</td>
<td>---</td>
<td>$\gamma$</td>
<td>---</td>
</tr>
<tr>
<td>WOB Correction Factor</td>
<td>---</td>
<td>$\alpha$</td>
<td>---</td>
</tr>
</tbody>
</table>
In demonstrating the application of their positive displacement motor and PDC bit model, Motahhari et al. ran some simulations of model predictions. In the following chart shown in Figure 7, the model was used to plot the predicted rate of penetration for three different PDMs throughout a well. This model was run with a log of confined compressive strength as one of the inputs, and RPM and WOB data were input from pre-recorded drilling data.
Figure 7 - Motahhari et al. ROP Predictions

The plot shows that the model can discern between the three different PDMs, but that also there is a fair bit of noise to the prediction due to the variations in the input.
data. Unfortunately, the true ROP was not plotted along with this data, so a comparison to the actual drilling performance is not possible. It is also important to note that Motahhari et al. emphasize that the high frequency ROP is not what should be optimized, but instead engineers should focus on optimizing the average ROP for a chosen interval of a well.

2.1.5 - Review of Traditional Models

These four models described above are the traditional models which will be examined in this study. There is quite a range of theories for how to best model rate of penetration, and also quite a long time spanned within the development of these models. Only the most recent model by Motahhari et al. is specifically designed to work with positive displacement motors and PDC bits, which all of the horizontal shale wells in the sample dataset for this study were drilled with. It would then be expected that the Motahhari et al. model would be the most accurate performer. However, despite the other models being predominantly developed for modelling applications of vertical wells, they will be tested in the horizontal shale regime as well.

As a summary, Table 5 below shows the input variables that are used by each of the models, as well as the number of empirical coefficients or model factors which are utilized. It can be seen that some input variables are considered essential by each model, while others may only be factored in once. It is interesting to observe that the Bourgoyne and Young model has by far the most comprehensive set of input variables, despite being one of the oldest models in the set. Among the things that are only considered by the
Bourgoyne and Young model are over/underbalance, depth, and jet impact force. The comparisons later in this project will make interesting note of whether all of these additional variables add accuracy to the Bourgoyne and Young model or whether they just contribute to the noise of its predictions.

Table 5 - Summary of Traditional Model Input Parameters

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Num. of Empirical Coefficients</th>
<th>RPM</th>
<th>WOB</th>
<th>Bit Diam.</th>
<th>TVD</th>
<th>Pore Pressure</th>
<th>ECD</th>
<th>Fractional Bit Wear</th>
<th>Jet Impact Force</th>
<th>Rock Compressive Strength</th>
<th>Number of Cutters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bingham</td>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bourgoyne and Young</td>
<td>8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hareland and Rampersad</td>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Motahhari et al.</td>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2 - Statistical Modeling and Drilling

Applying statistical modeling to ROP prediction has been previously worked on by Bilgesu et al (1997) who developed an artificial neural network to predict ROP, which was also taken on by Jahanbakhshi et al (2012). Dunlop et al (2011) show the potential gains that a closed loop automation system can provide, and have demonstrated experience deploying their model from simulation to field tests. The work by Hegde, Wallace, and Gray (2015) is an expansion of the statistical study of drilling performance, concentrating
on applying supervised learning techniques to the challenge of ROP prediction. The development of the Wider Windows Statistical Learning Model (WWSLM) as detailed in Hegde, Wallace, and Gray (2015) is not the focus of this project, however since the model is available and has been created within the Wider Windows IAP, it will be tested alongside these traditional models for its ability to match and predict the ROP of a horizontal shale well.

2.3 - Related Work within Wider Windows IAP

This project has been conducted within the Drilling Parametrics group, a subset of the Wider Windows Industrial Affiliate Program (IAP), and is one of a few projects that are focusing on the modeling and optimization of downhole drilling performance. The aforementioned work by Chiranth Hegde on the development of the WWSLM is very related, as the WWSLM model has been tested with horizontal shale well data (among other types), and is included in this project. The inclusion of the WWSLM in this project allows for some preliminary validation of the model which will be useful in establishing accuracy and credibility of the WWSLM as its development progresses.

Work by Cesar Soares on the “ROPPlotter” software has employed similar modeling techniques of the traditional models, and collaboration between these projects has resulted in expanded understanding of how to implement these models and how to deal with the peculiarities of each one. The project by Lucas Barros operates in a similar fashion, by evaluating the vertical sections of wells with the traditional ROP models, and observing what parameters change at the formation boundaries. Additionally, Anthony
Ho’s work toward modeling the Torque and Drag at the bottom of the hole is another part of the drilling parametrics environment, and will add more knowledge within this group.
Chapter 3 - Data Sources and Management

This project would not have been possible without the data which has been donated to the Wider Windows IAP by our sponsors. The data which has been used was provided by Marathon Oil, Chevron, and National Oilwell Varco, and they are due many thanks for making it available to the research group.

Since this project is focusing on horizontal shale wells, the data from NOV became the most important. There were four wells provided, all of which were drilled in the Eagle Ford shale with a horizontal section. Three of these wells were on one pad, and the other one was at a second location. This made them an ideal set of data to be used for this study. Unfortunately, the well that was on a different pad did not include a complete dataset in the lateral. Once the horizontal tools were being run in, the data record only includes one point for every 90 ft stand of pipe. This is not nearly enough data to base a model simulation on, so this well was removed from the study. The remaining three wells are shown in Figure 8’s plot of drilling performance below, which includes annotation of where the curve segment and lateral segment began for each of the wells. There are some anomalies shown in this plot, such as the abrupt drop in depth from 4,000 ft to 8,000 ft shown in Well #2, however this can be explained by the source data being incomplete for that range. Luckily this will not be a problem for this project as the incomplete portion is in the vertical section of the well. Additionally, trips to the surface and back can be identified by long flat sections of Non-Productive Time (NPT). This will be explored in greater detail further on in the project.
Figure 8 - Comparison of the Drilling Performance of Three Horizontal Wells in the Eagle Ford

3.1 - Data Compression

The data was provided in the form of CSV files which contain a data row for each second of the well’s progress. There are over 25 different columns of data, some differing
between each well dependent on which particular MWD/LWD or data recording equipment was being run at the time. Each file contains the standard sources of data however, including parameters such as Measured Depth, Bit Position, Bit Weight, Flow Weight, and recorded ROP, among others. Extra columns of data which were available in some of the files included data from NOV’s SoftSpeed system for reducing stick-slip vibration, or StringSense, another NOV product which uses strain gauges located in the top drive to extrapolate for downhole conditions, or CoPilot data from a Baker Hughes LWD tool. The data from these extra sensors was interesting to look at, however it was not used in this study as one of the main objectives of this work is to come up with a method for predicting ROP based entirely on parameters that are always available at surface independent of the drilling or LWD contractor used.

As mentioned before, the source data was with a resolution of every 1 second of drilling time. This was quite excessive for this project. The datasets with 1 second resolution included over 600,000 rows for a full well, and created CSV files of over 700 MB. These file sizes were unworkable for computational purposes, requiring excessive time to open and manipulate on a personal computer. As a result, the first task was to evaluate this dataset and reduce it to a more manageable file size. Previous efforts to reduce file size had just selected a row of data to retain every “x” feet, however this methodology results in a compressed dataset which is not as representative of the full resolution data, as it is not considering any of the data points which are dropped. As a
result, these methods were abandoned for this project and a better solution was developed.

The first step of this data compression effort was to discard the rows where the drill bit was not on bottom. For a study that is only concerned with drilling ROP, it is unimportant to show the progress of the bit up and down the hole during a trip, or when it is picked up to circulate off bottom. These pieces of the drilling process can still be identified in the compressed data by gaps in the time records, or by incrementation of the BHA number, however they are not important points for the calculation of ROP, so they were dropped from the datasets. Luckily these off-bottom points were easy to identify, as the data included a “on bottom” bit, which was either “0” for off bottom, or “1” for on bottom.

To further compress the data, an Excel VBA program was written which would handle the task. The program allows the user to import the full uncompressed data record, and specify parameters which will reduce the file size to a manageable value. To make this program versatile and usable by other members of the Wider Windows team, it was written in an Excel sheet which can be shared. In order to execute the program, the user must first copy the full uncompressed data file into the second worksheet of the file (leaving the control sheet in the first position). The uncompressed data must have column headings of only one row, and must have columns for Measured Depth (MD) as well as relative time, the cumulative amount of time spent drilling in seconds. These requirements were chosen with the available datasets in mind, and make using this
program on all sponsor provided data manageable. The user fills in the form shown below in Figure 9 with the required parameters, including the MD that they want their compressed log to start and end from, as well as the desired resolution of the compressed log. For the resolution, any value can be chosen depending on how much compression is desired, although for this project a value of 0.5 ft was chosen as the standard.

<table>
<thead>
<tr>
<th>Start Depth</th>
<th>13500 ft</th>
<th>(leave blank for 0 ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>End Depth</td>
<td>18500 ft</td>
<td>(leave blank for TD)</td>
</tr>
<tr>
<td>Resolution</td>
<td>0.5 ft</td>
<td>(must be filled in)</td>
</tr>
<tr>
<td>Rel Time Column</td>
<td>4 -</td>
<td>number (from 1)</td>
</tr>
<tr>
<td>MD Column</td>
<td>1 -</td>
<td>number (from 1)</td>
</tr>
</tbody>
</table>

Figure 9 - Driving Parameters for Compression Program

When the “Get Parameters” button above is pressed, the program evaluates the uncompressed data in the second sheet of the file, and finds a number of intermediary values. It identifies the total number of columns and rows in the source data, and then identifies what row number correlates to the desired start and end of the compressed log. It also determines how many total rows will exist in the compressed log after the process is completed. These values for the above example are shown below in Figure 10. If the user is satisfied with the intermediary values, they can click the “Perform Compression” button, and the process will begin.
Once the compression is started, VBA builds an array with the dimensions of the original dataset (for the range of depths to be included in the compressed dataset) and populates it with the original values. It also builds an array with the dimensions of the compressed dataset, which will be filled in as it goes. It then moves through this array with a custom function similar to an “AVERAGEIF” statement. The program establishes the range that it is currently evaluating based on the start depth and the resolution, so the first range would be from 13,500.0 ft to 13,500.5 ft. It then looks at the source data array, and starting from the beginning of the MD column, identifies the range of rows which fall within the desired compression range. With this range determined, the program then evaluates each column and averages the values for the uncompressed range into the final compressed value. It should be noted that the program can also be told to ignore certain values when calculating this average, for example some of the datasets report “-999.25” when they do not have a current reading, and the program can be told to ignore these values for the sake of the average. After the program has
calculated the compressed value for this first range, it increments the depth range to the next step of resolution, and again identifies the rows in the original dataset which correspond to this range. The program carries on like this until the entire desired compressed range has been evaluated and calculated, and then exports this compressed array back into a new worksheet in the original Excel file. This sheet can be saved as a new, compressed version, with a much more manageable file size. For example, on one of the wells this process reduced the dataset dimensions from 260,000+ rows to around 11,000 rows, a reduction of around 95% of the file size. This new file was much more manageable to work with on a personal computer, while still capturing a sufficient amount of high-speed data to show the rapid changes in input parameters which can affect ROP.

Previous efforts to reduce file size had just selected a row of data every “x” feet, however this methodology results in a compressed dataset which is not as representative of the full resolution data, as it is not considering any of the data points which are dropped.

3.2 - Properties of the Wells

Each well was examined to determine where the lateral started, and what the quality of the data was for the extent of the well.
3.2.1 - Well #1 Information

Well #1 was inspected both in the drilling data and in the associated documentation that was provided to the research group. From looking at this information, the well profile was found, which is shown below in Figure 11 (credit NOV). This well was drilled with a positive displacement motor (PDM), a rotary steerable system (RSS), and a polycrystalline diamond compact (PDC) bit. This is good because this is the most modern system of drilling horizontal sections, and will represent the currently available technology well. The well was drilled with a total of 3 bottom-hole assemblies (BHAs), with only one BHA necessary to drill the lateral. This is the indicator of a good well, as the operator was able to drill the lateral without accumulating any non-productive time (NPT) from trips out of the hole in the lateral section.

![Figure 11 - Drilling Program Profile View for Well #1 (NOV)](image)

It can be seen from this well plan that the curve section ends at a TVD of 12,975 ft. By looking at the corresponding data, this can be found to correlate to 13,421 ft measured depth. This will be rounded to 13,500 ft for convenience. The well continues to a Total Depth (TD) of 19,099 ft (measured), which will be rounded to 19,000 ft MD. This
leaves a 5,500 ft lateral to use for the model fitting. The data was recorded every second for all of the columns of information along this lateral, so this data will be run through the compression program detailed above. The data compression program intelligently reduces the number of rows from one row per second to one row per half foot of measured depth. The result of the data compression program was a reduction from 260,000+ rows of data to a much more manageable 11,000 rows.

After the Well #1 data was processed in the data compression program, the data still shows a large amount of variation between points, which is typical for drilling data of all kinds. A plot of some of the input variables for Well #1 is shown in Figure 12 below in its post-compressed form. The ROP data can be seen to be averaging around 100 ft/hr, although the data is quite noisy, bounding between 50-120 ft/hr very frequently. There is also a noticeable dip in the ROP around 14.5 k-ft, which seems to correspond to a decrease in WOB at the same depth. The values for Bit RPM are relatively consistent at 400 RPM for most of the well, with a drop to 200 RPM between 16.5 k-ft and 17.7 k-ft measured depth. The RPM values also have a decent amount of noise in the range of 15.7 k-ft to 16.5 k-ft, with the value dropping to around 50 RPM on numerous occasions. It would appear that the reduction to 200 RPM at the bit was in response to this section where it was difficult to maintain RPM, and then by the time RPM was increased back to 400, the issue was not nearly as prevalent. The WOB is a generally noisy signal between 20-40 k-lbf, including one large trend lower as mentioned before as corresponding to the ROP dip around 14.5 k-ft. The flow rate data starts at 500 gal/min, and reduces with a
linear trend as the hole gets deeper and the pumps have more difficulty overcoming the increased hydraulic friction of the drillpipe and annulus. There are a few spots of noise between 15 k-ft and 18 k-ft, although of these input parameters this is by far the least noisy of the bunch. With the exception of the ROP dip at 14.5 k-ft, none of these input sources seem to correlate to the trend of the ROP values, suggesting that complex interactions between the input parameters and the actual ROP are taking place.

Figure 12 - Input Data for Well #1
3.2.2 - Well #2 Properties

Well #2 was drilled in a very similar fashion to Well #1, utilizing the same planned well profile, which is shown below in Figure 13. There is a curve segment with build rates of 5°/100ft and then 8°/100ft, landing at 13,000 ft TVD, with an associated MD of just under 13,500 ft. It was also drilled with a downhole PDM, rotary steerable system, and PDC bit. This was a very good well, and only required two BHA to drill from surface until the end of the lateral, with one BHA being used for the entirety of the horizontal section included in this data range. TD was reached at just over 19,000 ft, and drilling the lateral took 101 hrs. This well has a data row every second, so it was also run through the data compression program, which reduced it to around 11,000 rows of data from over 250,000 rows.

![Figure 13 - Drilling Program Profile View for Well #2 (NOV)](image)

By inspecting the input parameters for Well #2 alongside the actual ROP, we can view the performance in greater detail. A plot of some input parameters with the actual ROP is shown below in figure 14. The ROP trace is reasonably steady, with values predominantly between 50-150 ft/hr. There are some swings up and down, but they are
mostly slow events. The amount of noise is decently constant, with the exception of the range between 14.5 k-ft and 15.0 k-ft, where the noise is less. Looking at the RPM signal, it shows periods where it is noisy (as would be expected), and other periods where it is consistently at 350 RPM. The regions where the RPM is stuck at 350 RPM appear to be missing data in the source file, as they report a far too high value of RPM (around 7,500 RPM), which was trimmed by the data compression program. This is a potential source of error, although in this study, the empirical coefficients for modifying the effect of the RPM parameter should be able to minimize the error that is introduced. Looking at the WOB trace, it is pretty consistent, with setpoints ranging from 20-40 k-lbf. There are no significant anomalies in the WOB plot, and the amount of noise is reasonable for this data.
3.2.3 - Well #3 Properties

The final well considered in this project was well #3. It was drilled in a similar fashion to wells #1 and #2, but with a slightly different profile, which is shown below in Figure 15. This well kicks off at 12,200 ft, and builds with BUR of 5°/100ft and then 8°/100ft until it reaches an inclination of 89°. The curve lands at a measured depth of 13,158 ft (which will be rounded up to 13,500 ft), and it reaches its TD at 17,937 ft. This is

Figure 14 - Input Data for Well #2
the same build and lateral inclination as the previous two wells, which makes for a good basis for comparison. This well had more BHAs required to reach TD than the previous wells, with four in total. There were three BHAs that played a part in the lateral, with their starting and ending depths shown in Table 6 below. Having multiple BHAs in the lateral will provide an opportunity to examine how the ROP models deal with new bits being introduced in the course of the lateral.

![Drilling Program Profile View for Well #3 (NOV)](image)

**Figure 15 - Drilling Program Profile View for Well #3 (NOV)**

**Table 6 - BHA Ranges for Well #3**

<table>
<thead>
<tr>
<th>BHA</th>
<th>MD In (ft)</th>
<th>MD Out (ft)</th>
<th>Range (ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>11,989.7</td>
<td>13,651.3</td>
<td>1,661.6</td>
</tr>
<tr>
<td>3</td>
<td>13,651.3</td>
<td>14,585.4</td>
<td>934.1</td>
</tr>
<tr>
<td>4</td>
<td>14,585.4</td>
<td>17,937.0</td>
<td>3,351.6</td>
</tr>
</tbody>
</table>

The data for Well #3 was not recorded in the same one-second resolution as the previous two wells, it was instead recorded with one point at every foot of measured depth. From examining the data it can be reasoned that the service company who was recording the data on location already applied their own averaging algorithm in order to
reduce the number of saved points in the well log. As a result, the Well #3 data does not need to run through the compressor program, and can be analyzed straight away. Since this data is in 1.0 ft resolution instead of the 0.5 ft resolution which has been generated in the compressed files, it will be interesting to see how the quality of the model fits compares.

The traces for some of the input parameters for Well #3 are shown below in Figure 16. This plot also includes lines to indicate where each new BHA starts in the data. The ROP signal is the smoothest of the three wells, with most of the values clustered around 50 ft/hr. There is still some noise in the ROP signal, but it does not have as many wild swings as the previous wells. This is most likely due to the 1.0 ft resolution of this data, and even more so due to the compression algorithm which the data had already been put through. There are less significant data anomalies in this well compared to the previous two. It is also interesting to observe that the ROP was not significantly degraded right before each trip out of the hole. Typically, the BHA is tripped out because the current configuration is not producing high enough ROP any more, usually due to a worn bit. However in this case, it can be surmised that the BHA was tripped out due to other causes, such as problems with the MWD tools. Unfortunately the driller’s notes are not included along with this drilling data, so a concrete reason for why the drillstring was pulled is not available. The RPM trace is also very smooth, and exhibits the behavior of being set at one power level, and then decreasing as the well got deeper due to the drag on the drillstring. The WOB is the noisiest of the traces, but still has the characteristics of a value
that was not varied significantly along the lateral. Finally the flow rate, which is the cleanest of the signals, only shows a couple of changes in value, and nothing drastic. So based on the general smoothness of these traces, the models should have a very good chance of fitting Well #3 accurately.

Figure 16 - Input Data for Well #3
Chapter 4 - Traditional Model Fits

This section will demonstrate the methodology that was conducted in order to fit the traditional models to the field data. These model fits were conducted in a couple different ways to evaluate the relative accuracy of these methods.

4.1 - Fitting Models with Full Range of Lateral Data

The first method used was to expose the traditional model to the entire range of the horizontal section of the well. This was chosen as the focal point for this project because the homogeneity of the horizontal section made for as consistent of a drilling environment as can be expected in the field. This also kept the true vertical depth (TVD) almost constant (these wells were drilled at an incline of 89 degrees, causing a slight but nearly negligible increase in TVD). This should also give the traditional models the most data to work with in order to create an accurate fit over the length of the lateral. The procedure for the full-length fit is shown below.

4.1.1 - Bingham Model Fit with Full Range of Lateral Data

The Bingham model was the first to be evaluated, and model fits were conducted on each of the three wells, which will be detailed individually in the following sections. The fitting of the traditional models occurred in an Excel environment, for convenience in using the VBA programming environment, and the ease of being able to share these documents with collaborators if necessary. A new workbook was created for each well, and a worksheet within that workbook was created for each model. Each model has four
types of input parameters: constants, empirical coefficients, depth-based data, and special parameters (which will be addressed later). The constants are terms such as the bit diameter or number of cutters, the empirical coefficients are the multipliers and exponents which modify the strength of each of the model’s terms, and the depth-based data is factors that change while drilling such as WOB or RPM. Table 7 below shows the input parameters for the Bingham model.

Table 7 - Bingham Model Input Parameters and Data Types

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Abbreviation</th>
<th>Bingham Variable</th>
<th>Units</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of Penetration</td>
<td>ROP</td>
<td>R</td>
<td>ft/sec</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Rotary Speed</td>
<td>RPM</td>
<td>N</td>
<td>rev/sec</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Bit Diameter</td>
<td>---</td>
<td>D</td>
<td>ft</td>
<td>Constant</td>
</tr>
<tr>
<td>Bit Load</td>
<td>WOB</td>
<td>W</td>
<td>lb</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Formation Factor</td>
<td>---</td>
<td>a</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>Drillability Factor</td>
<td>---</td>
<td>b</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>RPM Factor</td>
<td>---</td>
<td>c</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
</tbody>
</table>

Tables were set up in the Excel worksheet for each well’s Bingham model, one for the constants (only bit diameter in this case), and another for the empirical coefficients. The empirical coefficient table also includes limits for the lower and upper bounds of the coefficients. Where available, these limits come from literature and previous work, however for some of the models they had to be established as an engineering judgement, and their ranges were set to span similar distances to the established coefficients from other models. Initially, these empirical coefficients were set as a guessed value from the
middle of the upper/lower bounds range, since the program requires an initial value to begin iterating from.

The ROP equation for the model itself was then entered into a column of the worksheet next to the depth-based data. Its references were locked to the tables of constants and empirical coefficients for those values, and allowed to vary with depth for the input parameters which were depth based. This column of “Calculated ROP” was filled down through the range of the entire lateral. These calculated ROP values are not expected to be accurate at this point, since they are including the guessed values for the empirical coefficients as stated above. However it is necessary to generate a column of calculated values in order to check the error between the true and calculated ROPs.

Mean-Square Error (MSqE) was chosen as the error equation to use for this purpose. Mean-Square Error is a good choice for this error property because as the second moment of the error term about the origin, it includes both the bias and the variance of the estimator in its calculation. The formula for calculating the MSqE is shown in the equation below. The values of the squared error were calculated on each row of the depth-based data, comparing the actual measured ROP with the calculated ROP from the model. These squared error values were then summed and divided by the number of rows in order to create the MSqE term, a single value which represented the average error throughout the entire range of ROP values. The units of the MSqE value are (ft/hr)^2. For display purposes later, the square root of this MSqE term was taken, which provides the Root-Mean-Square-Error (RMSE) term. This is a more convenient error term, as the units
of the error are the same as the units of ROP as ft/hr. This RMSE value will be the term for each model fit that will be used to compare the accuracy of each of the different models.

\[
\text{Mean Square Error (MSqE)} = \frac{1}{n} \sum_{i=1}^{n} (\text{ROP}_{\text{calc}} - \text{ROP}_{\text{actual}})^2
\] (17)

Now that there is a single term which represents the average error in the ROP predictions throughout the lateral, attempts can begin to improve this accuracy. To accomplish this, Excel Solver was utilized to iterate the input parameters. The only input parameters which are allowed to be modified are the empirical coefficients, so that is what will be focused on. Solver was run with the goal of reducing the MSqE value to its minimum value, while being allowed to modify the empirical coefficients. The upper and lower bounds for the empirical coefficients were also included in the rules for solver, forcing it to keep them within the prescribed ranges. The program continues iterating until there are no more changes to the empirical coefficients which can reduce the MSqE value. At this point, the model is as tuned for the full range of the lateral data as possible, and the accuracy can be evaluated.

4.1.1.1 - Well #1 Bingham Model Fit with Full Range of Lateral Data

After the model fitting iteration was completed, the full-range Bingham model produced the fit shown in Figure 17 for the calculated ROP values. This fit is rather decent considering that the only input parameters for this model are RPM, WOB, and bit diameter. The Bingham model responds to the dip in Well 1’s ROP at 14.5 k-ft very quickly,
indicating that it was able to capture the source of this excursion in its equation. However the general shape of this plot is more of a moving average than a high-frequency trace of ROP. It tends to stay around the average ROP of 100 ft/hr, and has the tendency to allow noise on the lower side of this average more than the upper side. However, when the actual ROP spikes to a higher value around 16.5 k-ft, the Bingham model does not follow it up, seemingly stuck at the moving average. This model ended up with a RMSE of 23.3 ft/hr. This will be compared to the accuracy of the other models and other wells after their data is calculated later in the project.
Figure 17 - Bingham Calculated ROP over Actual ROP for Well #1

4.1.1.2 - Well #2 Bingham Model Fit with Full Range of Lateral Data

Well #2 was also evaluated with the Bingham model over the full range. This well was drilled with an almost identical profile and plan as Well #1, so the same range of 13,500 ft to 19,000 ft was used for the lateral data. After the model’s three empirical coefficients were iterated, the fit of calculated ROP to actual ROP which is shown in Figure
18 was generated. The calculated ROP trace has more variance than the plot from Well #1, but it still fails to respond to some of the significant events that occur in the actual ROP. There are noticeable increases in ROP at around 14.2 k-ft, 15.3 k-ft, and 17.1 k-ft. The Bingham model fails to respond to any of these trends, and instead stays near the average ROP in those sections. The model responds more readily to excursions to the lower side of the average, following dips in the actual ROP that occur at around 14.5 k-ft, 16.2 k-ft, and 18.2 k-ft. It is interesting to note that while the Bingham model is usually a moving average trend, this plot has segments of the calculated ROP that seem to follow different trends, and there is not a predominant moving average that the model clusters around. These segments can be seen between 17.0 k-ft and 18.2 k-ft, as well as between 18.2 k-ft and 19.0 k-ft. Although the moving average tendency is less prevalent in this model fit, that does not improve the accuracy. The model deviated from the average ROP in poor places, and as a result, the RMSE for this model came out to be 30.2 ft/hr.
4.1.1.3 - Well #3 Bingham Model Fit with Full Range of Lateral Data

Well #3 was run through the same model fitting methodology as was laid out for wells #1 and #2, and the plot of the ROP comparison is shown in Figure 19. It should be noted that this well data was provided to the WW IAP in 1.0 ft resolution, and had already been averaged and smoothed by the service company that was recording data on the
wellsite. As a result, the input data traces were smoother than the first two wells, as was shown in Section 3.2.3. The horizontal section for Well #3 was shorter than the other wells, going only 4,500 ft. The actual ROP trace has a couple of slow swings to faster and slower trends, with a faster-than-average trend occurring between 14.2 k-ft and 14.6 k-ft, and a slower-than-average trend taking place between 14.7 k-ft and 15.4 k-ft. The full-range Bingham model does not follow either of these deviations, and maintains an almost straight line at the average ROP throughout those ranges. The only instance where the calculated ROP trace deviates from the average for a meaningful amount of time occurs at 15.7 k-ft, where the calculated ROP goes higher than the average for around 100 ft. The actual ROP was higher than average here, but not as significantly as other points in the well. This Bingham full-range model is the most “moving average” type of result that is seen throughout this study. The most likely cause for this is the pre-averaged data that was supplied to the research group. However, since the model’s ROP values are so clean and consistent with the average ROP of the well, the RMSE for this model fit comes out to 14.8 ft/hr, the best accuracy for a full-range Bingham model.
4.1.2 - Bourgoyne and Young Model Fit with Full Range of Lateral Data

The Bourgoyne and Young model was the next one to be evaluated. The data sources for the Bourgoyne and Young model are shown below in Table 8. This model has the most input parameters by far, and some of them had to be sought out by researching the well or the area of the drilling in either the provided documentation or in other
resources. These more involved input parameters are marked with the “Special” data type in the table below. The value for the pore pressure was found from the included well planning documentation for a particular depth within this formation, and then was modified with a hydrostatic gradient to account for the difference in depth within the wellbore. The ECD was estimated in a similar fashion, based on the properties of the mud as detailed in the documentation, and the true vertical depth. The value for Threshold Bit Weight was taken as a constant of 0.1 k-lb/in, which was found in literature as a reasonable value. The fractional bit wear was estimated based on the length of the bit run, assuming that the bit had entered the hole with 0% wear, and wear occurred linearly until the end of that bit run. The Jet Impact Force was calculated from the mud and hydraulic properties, as well as the known dimensions of the bit. All of these estimates of the special factors have some error to them, however they are the best values that can be determined based on the available information. In order to improve the accuracy of these special parameters to the best case, they would have to be recorded in the depth-based logs along with all of the standard parameters. For the sake of this project, these estimates are reasonable enough, especially since they are still modified heavily by the empirical coefficients.
Table 8 - Bourgoyne & Young Model Input Parameters and Data Types

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Abbreviation</th>
<th>Bourgoyne and Young Variable</th>
<th>Units</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of Penetration</td>
<td>ROP</td>
<td>$\frac{dD}{dt}$</td>
<td>ft/hr</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Depth</td>
<td>TVD</td>
<td>D</td>
<td>ft</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Pore Pressure</td>
<td></td>
<td>$p_p$</td>
<td>lb/gal</td>
<td>Special</td>
</tr>
<tr>
<td>Equivalent Circulating Density</td>
<td>ECD</td>
<td>$\rho_c$</td>
<td>lb/gal</td>
<td>Special</td>
</tr>
<tr>
<td>Weight on Bit</td>
<td>WOB</td>
<td>W</td>
<td>k-lb</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Bit Diameter</td>
<td></td>
<td>$d_b$</td>
<td>in</td>
<td>Constant</td>
</tr>
<tr>
<td>Threshold Bit Weight</td>
<td></td>
<td>$\left(\frac{W}{d_b}\right)$</td>
<td>k-lb/in</td>
<td>Special</td>
</tr>
<tr>
<td>Rotary Speed</td>
<td>RPM</td>
<td>N</td>
<td>rpm</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Fractional Bit Wear</td>
<td></td>
<td>h</td>
<td>percent</td>
<td>Special</td>
</tr>
<tr>
<td>Jet Impact Force</td>
<td></td>
<td>$F_j$</td>
<td>Lbf</td>
<td>Special</td>
</tr>
<tr>
<td>Mud Density</td>
<td></td>
<td>$\rho$</td>
<td>lb/gal</td>
<td>Constant</td>
</tr>
<tr>
<td>Flow Rate</td>
<td></td>
<td>q</td>
<td>gal/min</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Apparent Viscosity</td>
<td></td>
<td>$\mu$</td>
<td>cp</td>
<td>Constant</td>
</tr>
<tr>
<td>Nozzle Diameter</td>
<td></td>
<td>$d_n$</td>
<td>in</td>
<td>Constant</td>
</tr>
<tr>
<td>Formation Strength Factor</td>
<td></td>
<td>$a_1$</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>Compaction Factor</td>
<td></td>
<td>$a_2$</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>Undercompaction Factor</td>
<td></td>
<td>$a_3$</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>Pressure Differential Factor</td>
<td></td>
<td>$a_4$</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>Bit Weight Factor</td>
<td></td>
<td>$a_5$</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>Rotary Speed Factor</td>
<td></td>
<td>$a_6$</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>Tooth Wear Factor</td>
<td></td>
<td>$a_7$</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>Hydraulic Factor</td>
<td></td>
<td>$a_8$</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
</tbody>
</table>

There are 8 empirical coefficients for the Bourgoyne and Young model, with $a_1$-$a_8$ each modifying one of the input terms. These were set up in a table of empirical coefficients, and had their lower and upper bounds entered. With the input parameters
now determined, the same procedure as was detailed on the Bingham model was
followed to create a depth-based “calculated ROP” value from the constants, special
variables, empirical coefficients, and depth-based parameters. The same procedure for
calculating the MSqE and using Excel Solver to iterate the values of the empirical
coefficients was followed, producing a best-fit calculated ROP for each of the wells below.

4.1.2.1 - Well #1 Bourgoyne and Young Model Fit with Full Range of Lateral Data

Well #1 data was run through the iteration program, which produced the plot of
calculated over actual ROP shown below in Figure 20. Compared to the Bingham model
fit for Well #1, the Bourgoyne and Young data is much more erratic. It still responds to
the dip in ROP at 14.5k, although after that it seems to act like a moving average with an
upper limit. There are also three sections that seem to be prevalent, one from 15 k-ft to
16.4 k-ft where it has a maximum of 120 ft/hr, another between 16.4 k-ft and 17.6 k-ft
with a maximum of 75 ft/hr, and a third section from 17.6 k-ft until TD with a maximum
of 100 ft/hr. The second section from 16.4 k-ft and 17.6 k-ft is the least accurate, with
many of the spikes of the calculated ROP going directly opposite of the actual ROP. As a
result of the accumulation of this gross error despite using the most optimized empirical
coefficients, the overall RMSE for the Bourgoyne and Young model on Well #1 is 31.9
ft/hr.
4.1.2.2 - Well #2 Bourgoine and Young Model Fit with Full Range of Lateral Data

The same process was applied to the data from Well #2, and the resulting plot is shown below in Figure 21. This model fit is noisy, but the Bourgoine and Young model follows the trends of the calculated ROP here more than in most wells. It can be seen to be following the actual ROP’s trend of increasing from 13.5 k-ft to 14.5 k-ft, and then
follows the actual ROP’s drop down to a lower value. It continues tracing the actual ROP’s trend very well from 14.5 k-ft to 14.8 k-ft, where the actual ROP data was the least noisy. For the rest of the well, the model’s calculated ROP remains a solid follower of the actual ROP’s trend, although the noise of the calculated ROP increases. However, when the RMSE is calculated, a value of 34.7 ft/hr error is determined, which is rather lackluster. This model fit captured the trends of the actual ROP well, although the noise in the calculated ROP caused the RMSE to remain higher than desired.
4.1.2.3 - Well #3 Bourgoyne and Young Model Fit with Full Range of Lateral Data

The drilling data from Well #3 was also run through the same process in order to generate the most optimal full-range fit of the Bourgoyne and Young model. The plot of the comparison of ROP calculations is shown in Figure 22. Similar to the Well #3 model fit for the Bingham model, this Bourgoyne and Young Well #3 fit is the smoothest of all the
Bourgoyne and Young fits, due to the 1.0 ft resolution of the source data, as well as the pre-averaging of the data. This model fit is minimally noisy, and does the best job of all the Bourgoyne and Young fits of following the actual ROP values. Particularly around 14.8 k-ft, the calculated ROP follows the actual ROP below the average. The one significant deviation of the calculated ROP is at 15.7 k-ft, where it tries to follow the calculated ROP which is trending above the average, but overshoots substantially, and takes around 80 ft to recover. Overall, this is the most accurate of the Bourgoyne and Young model fits, with an RMSE value of 16.9 ft/hr.
4.1.3 - Hareland and Rampersad Model Fit with Full Range of Lateral Data

The third model in the project was the Hareland and Rampersad model for ROP, which was also evaluated for all of the wells, first by exposing it to the full range of lateral data. The input parameters and data types for the Hareland and Rampersad model are shown below in Table 9. There are a two input parameters which required special
attention for this model, with the Weight on Bit per Cutter being calculated as a depth-based parameter solved simply by dividing each WOB by the number of cutters on the bit, and the Uniaxial Compressive Rock Strength being assumed as a constant for this formation, since only limited data was available in the associated documentation. As before, there is expected to be slight error introduced at this juncture, but this error can be mitigated by the fitting of the empirical coefficients and should be minimal in the long run.

Table 9 - Hareland and Rampersad Model Input Parameters and Data Types

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Abbreviation</th>
<th>Hareland and Rampersad Variable</th>
<th>Units</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Cutters</td>
<td>---</td>
<td>N&lt;sub&gt;c&lt;/sub&gt;</td>
<td>---</td>
<td>Constant</td>
</tr>
<tr>
<td>Bit Rotational Speed</td>
<td>RPM</td>
<td>RPM</td>
<td>rev/min</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Bit Diameter</td>
<td>---</td>
<td>D&lt;sub&gt;b&lt;/sub&gt;</td>
<td>in</td>
<td>Constant</td>
</tr>
<tr>
<td>Cutter Diameter</td>
<td>---</td>
<td>d&lt;sub&gt;c&lt;/sub&gt;</td>
<td>in</td>
<td>Constant</td>
</tr>
<tr>
<td>Weight on Bit per Cutter</td>
<td>---</td>
<td>W&lt;sub&gt;mech&lt;/sub&gt;</td>
<td>lbs</td>
<td>Special</td>
</tr>
<tr>
<td>Uniaxial Compressive Rock Strength</td>
<td>---</td>
<td>σ&lt;sub&gt;c&lt;/sub&gt;</td>
<td>psi</td>
<td>Special</td>
</tr>
<tr>
<td>Lithology Correction Factor</td>
<td>---</td>
<td>a</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>RPM Correction Factor</td>
<td>---</td>
<td>b</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>WOB Correction Factor</td>
<td>---</td>
<td>c</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
</tbody>
</table>

The same fitting procedure was conducted as has been discussed for the other models, with a depth-based column of “Calculated ROP” values being solved, including the effects of the empirical coefficients, which are then iterated with Excel Solver to
determine which values of these coefficients minimize the mean-square-error over the length of the lateral.

4.1.3.1 - Well #1 Hareland and Rampersad Model Fit with Full Range of Lateral Data

The drilling data from Well #1 was the first dataset that the Hareland and Rampersad model was tested on. The resulting plot of calculated over actual ROP is shown below in Figure 23. For this well, the Hareland and Rampersad model produces ROP values very similar to the Bingham model, despite the increased complexity of the Hareland and Rampersad model. It reacts to the dip in ROP at 14.5 k-ft quickly, but then after that acts as a moving average with an upper limit, similar to what has been seen before. As far as moving averages go, it is quite accurate, fitting right down the middle of the actual ROP data. Despite this, the model fails to respond to many of the ROP excursions above from this central tendency, such as the increase in actual ROP at 16.6 k-ft, or the increase at 17.5 k-ft. It does follow some of the decreases in ROP more closely, such as the minor dip at 16.1 k-ft, and the slightly longer lasting dip at 17.2 k-ft. It is interesting to observe that the Hareland and Rampersad model is much more responsive to deviations below the average rather than above. Overall this model has a RMSE of 23.2 ft/hr, which will be the most accurate full-range model for Well #1.
Figure 23 - Hareland & Rampersad Calculated ROP over Actual ROP for Well #1

4.1.3.2 - Well #2 Harel& and Rampersad Model Fit with Full Range of Lateral Data

The same procedure was applied to the data from Well #2, which generated the plot of the calculated vs actual ROP shown in Figure 24 below. This model fit tends to stay around the overall moving average of the actual ROP, and has more deviations to the lower side of the average than the upper. There is almost a limit that regulates the upper
extent of the calculated ROP, where despite how high the actual ROP goes, the calculated ROP cannot rise to follow it. There is the brief section at 17.1 k-ft where the calculated ROP follows the actual ROP up, but it does not go as far above average as the actual. The biggest glaring places where this model produces a lot of error are the section between the beginning and 14.5 k-ft, as well as the section from 15.0 to 15.6 k-ft. Overall, this model fit does a poor job of following the trends put forth by the actual ROP data. The RMSE value for well #2’s Hareland and Rampersad model fit came out to 29.4 ft/hr.
Figure 24 - Hareland & Rampersad Calculated ROP over Actual ROP for Well #2

4.1.3.3 - Well #3 Hareland and Rampersad Model Fit with Full Range of Lateral Data

The final well was fit with the Hareland and Rampersad model was Well #3. This is the smoothest of the model fits for the Hareland and Rampersad model, as it was for the Bingham and Bourgoyne and Young models, as can be seen in the comparison plot in Figure 25. Again, this is due to the data from this well having been pre-compressed to 1.0
ft resolution, and not some innate parameter about this particular well. Similar to the Bingham fit for Well #3, this model takes on the appearance of a very straight moving average of the overall well performance. The model ignores deviations of the actual ROP above and below the average, with the exception of an 80ft section around 15.7 k-ft. However, even though this model just stays near a straight average, it does so without introducing extra noise and therefore error. As a result, it gives a RMSE value of 14.8 ft/hr, which is the best that will be calculated for Well #3.
4.1.4 - Motahhari et al. Model Fit with Full Range of Lateral Data

The final model considered in this study was the Motahhari et al. model, which was developed specifically for the modeling of wells drilled with positive displacement motors and PDC bits. As with all models in this section, the model was first tested by applying it to the full range of data available for the wells in this project. The Motahhari
et al. model has the input parameters shown in Table 10, including two special parameters. The bit wear function was estimated in the same fashion as in the Bourgoyne and Young model, by stating that the bit was 0% worn when each new bit run began, with wear increasing linearly with the amount of depth that the bit drills. The confined rock strength was determined from the associated documentation about this field, and then assumed to be a constant for the length of the lateral. Again, this will introduce some error, but since each of these laterals exists only in one formation and at a nearly constant TVD, these are not unreasonable approximations to make.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Abbreviation</th>
<th>Motahhari et al. Variable</th>
<th>Units</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit Wear Function</td>
<td>---</td>
<td>Wf</td>
<td>percent</td>
<td>Special</td>
</tr>
<tr>
<td>Total Bit Rotational Speed</td>
<td>RPM</td>
<td>RPM_t</td>
<td>rpm</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Weight on Bit</td>
<td>WOB</td>
<td>WOB</td>
<td>lbs</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Bit Diameter</td>
<td>---</td>
<td>D_B</td>
<td>in</td>
<td>Constant</td>
</tr>
<tr>
<td>Confined Compressive Strength</td>
<td>CCS</td>
<td>S</td>
<td>psi</td>
<td>Special</td>
</tr>
<tr>
<td>Bit Design Coefficient</td>
<td>---</td>
<td>G</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>RPM Correction Factor</td>
<td>---</td>
<td>γ</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
<tr>
<td>WOB Correction Factor</td>
<td>---</td>
<td>α</td>
<td>---</td>
<td>Empirical Coef.</td>
</tr>
</tbody>
</table>

The model was once again programmed into the Excel sheet and the empirical coefficient were iterated to find the values which caused the mean-square-error of the entire lateral to be the least.
4.1.4.1 - *Motahhari et al. Model Fit with Full Range of Lateral Data*

The plot of the Motahhari et al. ROP model over the actual ROP for Well #1 is shown below in Figure 26. This plot is similar to those of the Bingham and Hareland and Rampersad models. The Motahhari et al. model tends to follow more of the deviations to the lower side of the average ROP, including the significant spike at 14.5 k-ft, as well as smaller spikes at 15.3 k-ft, 16.1 k-ft, and 17.2 k-ft. However, it acts as if there is an upper limit on the moving average, which is decreasing almost linearly with depth (likely due to the effect of the bit wear function). Also, as with the Hareland and Rampersad model, this model does not do a good job of following deviations above the moving average, in fact it only spikes above the moving average trend once, and that is an erroneous data point. Overall this model produces an RMSE of 24.0 ft/hr.
4.1.4.2 - Well #2 Motahhari et al. Model Fit with Full Range of Lateral Data

A similar plot was made for the Well #2 data of the model calculated ROP over the actual ROP, which is shown below in Figure 27. The outcome of this model fit is similar to that of the Hareland and Rampersad model for Well #2, although the Motahhari et al. model. Both models act like a moving average with an upper limit, and create sections of
data that follow similar trends for around 1,000 ft before changing. However there are some differences in this Motahhari et al. model as well, one of which is that it does not attempt to follow the actual ROP up above the average around 17.1 k-ft. There is noise on this model, but it is not that dramatic, with most of the points being clustered around a good section-based moving average for the actual ROP. As a result, this well shows a RMSE value of 27.5 ft/hr, which is the best accuracy for Well #2.
Figure 27 - Motahhari et al. Calculated ROP over Actual ROP for Well #2

4.1.4.3 - Well #3 Motahhari et al. Model Fit with Full Range of Lateral Data

The final model fit for the Motahhari et al. model was for Well #3, which is shown below in Figure 28. Well #3 is again the smoothest of the wells, since it was a 1.0 ft resolution, and had pre-smoothed data. As with the other Motahhari et al. models, a trend of decreasing ROP over the length of the lateral is apparent, due to the bit wear
parameter in this model. Since there were multiple BHAs used to drill Well #3, we can also examine what happens at the BHA boundaries. The most apparent boundary is between BHAs 3 and 4, at a depth of about 14.6 k-ft. The model-fit ROP jumps up here, as the bit wear parameter resets with the new bit, however the model still does not fit the actual ROP well around that trip out and back into the hole. As with other Well #3 fits, the model fit data is the smoothest, and it approximates a moving average with the aforementioned decreasing trend. It does not follow ROP excursions above or below the average, and remains very clean of noise. As a result of this clean data signal, this model produced a RMSE value of 15.5 ft/hr.
4.1.5 - Conclusions based on Full Range Model Fits

After evaluating each of the 4 models on these wells, it has been shown that when provided with the full range of the lateral for fitting, they tend to perform as moving averages. These traditional models react to the most extreme of ROP deviations, but seem biased to respond to decreases in ROP much more than increases. Additionally, the
traditional models do not respond with the high frequency that would be hoped with a model that has access to a datapoint every 0.5 ft of the well (or 1.0 ft for Well #3). The RMSE values of each of the wells for the full-range tests are shown below in Table 11:

<table>
<thead>
<tr>
<th>Model</th>
<th>Well #1</th>
<th>Well #2</th>
<th>Well #3</th>
<th>Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bingham Full Range</td>
<td>23.3</td>
<td>30.2</td>
<td>14.8</td>
<td>22.8</td>
</tr>
<tr>
<td>Bourgoyne and Young Full Range</td>
<td>31.9</td>
<td>34.7</td>
<td>16.9</td>
<td>27.8</td>
</tr>
<tr>
<td>Hareland and Rampersad Full Range</td>
<td><strong>23.2</strong></td>
<td>29.4</td>
<td><strong>14.8</strong></td>
<td>22.5</td>
</tr>
<tr>
<td>Motahhari et al. Full Range</td>
<td>24</td>
<td><strong>27.5</strong></td>
<td>15.5</td>
<td><strong>22.3</strong></td>
</tr>
<tr>
<td><strong>Averages</strong></td>
<td><strong>25.6</strong></td>
<td><strong>30.45</strong></td>
<td><strong>15.5</strong></td>
<td><strong>23.85</strong></td>
</tr>
</tbody>
</table>

It can be seen that for Well #1, the average RMSE value was 25.6 ft/hr, with the best performer being the Hareland and Rampersad model, with an RMSE of 23.2 ft/hr. For Well #2, the average accuracy was 30.45 ft/hr, and the most accurate model was the Motahhari et al. model, at 27.5 ft/hr. Finally, Well #3 had an average accuracy of 15.5 ft/hr, with the best model being the Hareland and Rampersad model which produced an RMSE of 14.8 ft/hr. Wells #1 and #2 have similar input parameters, as well as similar results for their model fits. On each of the models, Well #3 allowed a markedly better fit, again due to the decreased resolution of 1.0 ft/row, and even more due to the fact that the source data had already been run through a compression and smoothing algorithm.
When taking all of the models and all of the wells into consideration, the full-range model fits produced an overall RMSE value of 23.85 ft/hr.

These traditional models performed so consistently as moving average functions that they were also compared to a straight-line fit of the data from Well #1 over the same range, which is shown below in Figure 29. The slope of this line was iterated with the same error-reducing method which had been used to dial in the empirical coefficients for each of the traditional models. After optimizing the slope, this straight line had an RMSE of 27.5 ft/hr. This is within the range of the RMSE values based on the traditional models, demonstrating that an actual straight line would be around the same level of accuracy. However this does not imply that the traditional models are garbage, just that they should be used for macro-scale predictions of ROP. They will be valuable to predict the ROP for an entire formation or bit run, but should not be relied on to predict the ROP on a foot-by-foot basis.
4.2 - Fitting Models with Increased Coefficient Resolution

After evaluating the results of the full-range model fits, an effort was made to increase the accuracy of these model fits. Despite the hope that providing the models with the full range of the lateral would allow them to find the most accurate empirical
coefficients that would cause the model to accurately match the ROP for the entire wellbore, this did not seem to be the case. The models were not as responsive as desired when given the entire range of the lateral to fit against and iterate for the most accurate empirical coefficients, so the lateral was broken up into a number of smaller sections, and the empirical coefficients were then optimized within these sections. For the following section, the sections were set at 50 ft. For the first 50 ft of the well, one set of empirical coefficients was solved for to minimize the error in this section. Then the next 50 ft of the well was considered, and a new set of empirical coefficients was determined which would minimize the error in this section.

To accomplish this, further programming in Excel and Excel VBA was conducted. This program was written on top of the previous worksheets where all of the depth based, constant, and special data had been present. A new depth-based table was built, although the column of depth values did not increment by 0.5 ft, but rather by 50 ft, such that the values were 13,500 ft, 13,550 ft, 13,600 ft, etc. The next column was the “coefficient start row”, a value which represented the row number (in the full 0.5 ft resolution data set) at which the current row started. So for the 13,500 ft coefficient, it would start at row #2, while the 13,550 ft coefficient started at row 102. These rows which corresponded to the start of each 50 ft section were found with the “MATCH” function, so that this same program could be used on each well and each model. This 50-ft resolution table then included columns for each of the empirical coefficients required for the model, so between 3-8 columns. Finally, it included a column for the “Segment Sum Error-Squared”
which is the total squared error for each of the sections of the well. This is calculated based on another set of depth-based data.

In addition to this per-section summary table including the empirical coefficients, there was also data added to the full resolution depth-based data table. For each data row (so each 0.5 ft) there is an entry in the “reference row” column, where the program looks up which row of the sectioned table contains the current row of the full resolution data. Then the calculated ROP is found, by using the same equation of the model as before, although for the empirical coefficients, it instead uses the reference row value to look up the appropriate empirical coefficients from the per-section table. Additionally, the square-error for each row of the full resolution data is calculated, and these are the values which are then summed up by the “Segment Sum Error-Squared” column in the per-section table. At the end of the entire process, the Segment Sum Error-Squared values are totaled up for the entire lateral, and then divided by the number of full-resolution rows to get the Mean Square Error, and the square root of the MSqE is taken to provide the Root-Mean-Square-Error.

Finally, the program uses Solver within VBA to iterate each of the rows of empirical coefficients in the per-section table so that the segment sum error-squared for that particular section is minimized. The program steps down the entire per-section table, iterating a new set of empirical coefficients for every 50 ft of lateral. When this process is complete, the RMSE value has been minimized for this process.
4.2.1 - Bingham Model Fit with Increased Coefficient Resolution

This method was first run on the Bingham model with Well #1 data. The same input parameters were used as in the first phase of the model fits, the only difference was that the empirical coefficients were allowed to change every 50 ft. The resulting fit of this increased coefficient resolution Bingham model is shown below in Figure 30, in addition to the previously generated model fit using the full length of the lateral. The effect of the increased resolution is obvious – there is much greater tracing of the actual ROP. The model no longer is behaving like a moving average with a forced upper limit. It is now following deviations of the ROP above the average as well as below, and is capturing more of the high-frequency jumping than before. As a result of this increase in coefficient resolution, the RMSE of the Bingham model fit for Well #1 has improved from 23.3 ft/hr to 16.6 ft/hr. This is a significant step, although there is still room for further accuracy gains. By examining the plot, the underlying actual ROP trace is still much noisier than the model fit ROP, and the model fit ROP is still acting as a moving average, except only over a 50 ft range.
This method was also run for Wells #2 and #3, and the plots comparing them to their full-range Bingham model fits are shown below in Figure 31 and 32. For Well #2, the increased coefficient resolution produced a RMSE value of 19.8 ft/hr, while for Well #3 the RMSE came in at 10.3 ft/hr.
Figure 31 - Comparison of Bingham Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #2
4.2.2 - Bourgoyne and Young Model Fit with Increased Coefficient Resolution

The same methodology was carried out on the Bourgoyne and Young model with an increased empirical coefficient resolution of 50 ft. The same input parameters and upper/lower bounds of the empirical coefficients were used. Figure 33 shows the comparison between the full-range model fit and the 50 ft resolution model fit. The noise of the model is dramatically increased, and it is a much more erratic fit than the Bingham
model in the previous example. While the Bingham model still looked largely like a moving average which was reset every 50 ft, this Bourgoyne and Young model is attempting to fit the inherent noise of the actual ROP data much more closely. It follows the ROP through swings both above and below the normal trend, and even matches some of the erratic, errant looking points on the extremes of the ROP ranges. This fit is certainly an improvement over the full-range fit, although it may have gone too far in its attempt to recreate the messy signal of the actual ROP. The RMSE for Well #1 with the Bourgoyne and Young model improved from 31.9 ft/hr to 23.6 ft/hr. This is the most drastic change in ROP accuracy seen yet, although even with the increased resolution, this RMSE value is only just in the same range as the full-range RMSE values for the other three model types.
Figure 33 - Comparison of Bourgoyne and Young Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #1

This method was also run on the Bourgoyne and Young model for Wells #2 and #3, and the plots comparing them to their full-range Bourgoyne and Young model fits are shown below in Figure 34 and 35. For Well #2, the increased coefficient resolution produced a RMSE value of 28.8 ft/hr, while for Well #3 the RMSE came in at 15.0 ft/hr.
Figure 34 - Comparison of Bourgoyne and Young Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #2
Figure 35 - Comparison of Bourgoyne and Young Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #3

4.2.3 - Hareland and Rampersad Model Fit with Increased Coefficient Resolution

This increased resolution fitting procedure was also applied with Well #1 data and the Hareland and Rampersad model. As before, there were no changes made to the constant or special input parameters, or the upper/lower bounds of the empirical coefficients. The plot in Figure 36 shows the comparison of the full-range model fit and the model fit with increased coefficient resolution. The transformation of the model fit
here is similar to that of the Bingham model above – the increased resolution fit no longer looks like a moving average of the actual ROP. It follows the ROP excursions above and below the average quite readily. However in the regions where the Bingham increased coefficient resolution fit line was almost straight and in the middle of the actual ROP range, this Hareland and Rampersad fit shows more variation in those regions. An example of this trend can be seen between 15.5 k-ft and 15.7 k-ft in both the Bingham and Hareland and Rampersad increased coefficient resolution plots. The RMSE for the Hareland and Rampersad model improved from 23.2 ft/hr to 16.7 ft/hr. As with the Bingham model above, this improvement is a step in the right direction, but still does not come close to actually following the high-frequency of variation in the drilling rate.
The same procedure was also applied to Wells #2 and #3, and the plots comparing them to their full-range Hareland and Rampersad model fits are shown below in Figure 37 and 38. For Well #2, the increased coefficient resolution produced a RMSE value of 23.3 ft/hr, while for Well #3 the RMSE came in at 10.8 ft/hr.
Figure 37 - Comparison of Harelond and Rampersad Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #2
4.2.4 - Motahhari et al. Model Fit with Increased Coefficient Resolution

Finally, this method was also applied to the Motahhari et al. model with 50 ft resolution for the empirical coefficients. The original input parameters and special variables were retained, as well as all of the depth-based data. The comparison between the full-range fit and the increased coefficient resolution fit are plotted below in Figure 39. This plot of the Motahari 50 ft fit is strikingly similar to the Bingham and Hareland and
Rampersad fits, with the greatest similarity to the Bingham model. As with the Bingham model, it follows the ROP excursions when there are extreme or longer lasting deviations, but when the actual ROP signal is the cleanest, the model is also very stable, almost acting like a stepped moving average with a 50 ft period. Increasing the empirical coefficient resolution improved the RMSE value from 24.0 ft/hr to 16.7 ft/hr. As with the other models, this is a good improvement but not a significant enough improvement to allow this model to be used in this fashion to predict high-speed changes in the ROP.
Figure 39 - Comparison of Motahhari et al. Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #1

The Motahhari et al. model was also applied to Wells #2 and #3 with increased coefficient resolution, and the plots comparing them to their full-range Motahhari et al. model fits are shown below in Figure 40 and 41. For Well #2, the increased coefficient resolution produced a RMSE value of 18.3 ft/hr, while for Well #3 the RMSE came in at 10.4 ft/hr.
Figure 40 - Comparison of Motahhari et al. Full Range Model Fit (Left) and Increased Coefficient Resolution Model Fit (Right) for Well #2
4.2.5 - Conclusions based on Increased Coefficient Resolution Model Fits

Based on the model fits with the increased coefficient resolution, the RMSE error has been reduced by at least 30% for each of the model types. After allowing the coefficients to vary every 50 ft, the issue of the models behaving like a moving average over the length of the lateral has been resolved, and they are now much more responsive to local changes. Additionally, they all are able to follow the ROP excursions both below
and now above the median ROP readily. Table 12 below shows the RMSE values for each of the increased coefficient resolution model fits, as well as the average values for the full-range model fits for comparison purposes. The best fit for each well has been highlighted with bold, and it can be seen that the Bingham increased coefficient resolution model fit was the most accurate for Well #1 and Well #3, while the Motahhari et al. increased coefficient resolution model had the lowest error for Well #2. Compared to the average RMSE values for the full-range model fits, each of the increased coefficient resolution fits is a strong increase. The only exception to this is the increased coefficient resolution Bourgoyne and Young model fit for Well #3, which only achieves a RMSE of 15.0 ft/hr, hardly an improvement over the average full-range RMSE of 15.5 ft/hr. This is most likely due to the high degree of noise that is prevalent in the Bourgoyne and Young model fits, and despite being allowed to vary the empirical coefficients every 50 ft, the noise in the ROP calculation still causes it to have a similar RMSE value to the full-range data. However, when initially comparing across the board, this method looks like quite an improvement on the quality of these model fits.
Table 12 - RMSE Summary for Traditional ROP Models with Increased Coefficient Resolution

<table>
<thead>
<tr>
<th>Model Fit Averages</th>
<th>Well #1</th>
<th>Well #2</th>
<th>Well #3</th>
<th>Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Range Model</td>
<td>25.6</td>
<td>30.45</td>
<td>15.5</td>
<td>23.85</td>
</tr>
<tr>
<td>Bingham Increased Coef. Resolution</td>
<td>16.6</td>
<td>19.8</td>
<td>10.3</td>
<td>15.6</td>
</tr>
<tr>
<td>Bourgoyne and Young Increased Coef. Resolution</td>
<td>23.6</td>
<td>28.8</td>
<td>15.0</td>
<td>22.5</td>
</tr>
<tr>
<td>Hareland and Rampersad Increased Coef. Resolution</td>
<td>16.7</td>
<td>23.3</td>
<td>10.8</td>
<td>16.9</td>
</tr>
<tr>
<td>Motahhari et al. Increased Coef. Resolution</td>
<td>16.7</td>
<td>18.3</td>
<td>10.4</td>
<td>15.1</td>
</tr>
<tr>
<td>Increased Coef. Resolution Model Fit Averages</td>
<td>18.4</td>
<td>22.55</td>
<td>11.625</td>
<td>17.5</td>
</tr>
</tbody>
</table>

However, these model fits still present a large issue, which is overfitting. By allowing the empirical coefficients to be reset every 50 ft, the error is certainly being decreased, but the empirical coefficients are being so wildly varied that they are only applicable to that particular 50 ft section of the well. Since the empirical coefficients have so much strength to modify the calculated ROP, adjusting them can essentially force the calculated ROP curve to approximate any trend in the actual ROP over a short interval. Shown below in Table 13 are the empirical coefficients for the Hareland model for the first ten 50 ft sections of Well #1. By observing the variation line-to-line in the empirical coefficients, especially coefficient “a” which is the Lithology Correction Factor, this becomes apparent. Between 13,800 ft and 13,950 ft this value goes from 222 to 18,847 and then back down to 227, exhibiting a wild swing between the minimum and maximum.
values seen in this range. Also, since this empirical coefficient is the Lithology Correction Factor, it is implying that the effect of the lithology on the ROP is varying this widely within just a 500 foot range, while it is known that the entire lateral was drilled within one formation with relatively consistent rock properties. The other two coefficients do not swing as wildly between each individual point, but do show a very wide range of values within this range.

Table 13 - Empirical Coefficients for Hareland Model from First 10 Sections of Increased Coefficient Resolution Data Table

<table>
<thead>
<tr>
<th>Coef. Start MD (ft)</th>
<th>Coef. End MD (ft)</th>
<th>Empirical Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>13500</td>
<td>13550</td>
<td>2772.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00010</td>
</tr>
<tr>
<td>13550</td>
<td>13600</td>
<td>6977.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00010</td>
</tr>
<tr>
<td>13600</td>
<td>13650</td>
<td>14620.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00012</td>
</tr>
<tr>
<td>13650</td>
<td>13700</td>
<td>973.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00016</td>
</tr>
<tr>
<td>13700</td>
<td>13750</td>
<td>1437.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00010</td>
</tr>
<tr>
<td>13750</td>
<td>13800</td>
<td>11196.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00010</td>
</tr>
<tr>
<td>13800</td>
<td>13850</td>
<td>222.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00010</td>
</tr>
<tr>
<td>13850</td>
<td>13900</td>
<td>18847.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00010</td>
</tr>
<tr>
<td>13900</td>
<td>13950</td>
<td>227.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.12765</td>
</tr>
<tr>
<td>13950</td>
<td>14000</td>
<td>3082.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.00010</td>
</tr>
</tbody>
</table>

Considering the effects of increasing the coefficient resolution to its extreme shows where this methodology would end up if continued. Assume that the coefficient resolution was reduced to 0.5 ft, now there is a new set of empirical coefficients for every single row of the depth-based data. These coefficients would be iterated between their upper and lower bounds until they drove the calculated ROP as close as possible to the actual ROP. When viewed as a model fit, this would suggest that there was very low error,
however it would be meaningless data, since the model would be so drastically overfit to the actual ROP curves. The smaller the resolution that the empirical coefficients are fit to, the more they are being forced to produce an overfit to the actual model, and the less meaningful their results become.

4.3 - Sources of Error in Traditional Model Fits

There are definitely some sources of error that could be preventing these traditional models from achieving their best possible accuracy. One source is the special input variables which have been discussed on each of the above models. These special variables were not recorded in the depth-based data, and needed to be estimated based on the other documentation that accompanied this field data, or approximated based on other data which was found in research. These estimations were definitely not perfectly accurate, however they should have been reasonable enough to serve as a good approximation. Of course the error of the models could be reduced if there were actual depth-based data for each of these variables, such as pore pressure or rock compressive strength, however this was not available in the available field data.

Another potential source of error is the upper and lower bounds which were established for the empirical coefficients. This could actually be considered both a source of error or a source of unrealistic accuracy. Since these coefficients have such an impact on the value of the calculated ROP, they are largely determinant of whether the fit will be close at a point or not. The upper and lower bounds for the empirical coefficients were set based on guidelines from the model authors as well as other published studies which
have utilized these models, but the ranges were still typically quite wide. When the coefficients that have been solved for are examined, some of them are forced to their lowest limit such as the instances of coefficient “b” equaling 0.0001 in Table 13. In that model, “b” represents the RPM factor, and is applied to the RPM value as an exponent. So by driving the empirical coefficient as close to zero as possible, the model is trying to eliminate the effect of the RPM by raising it to the power of zero. This would imply that the RPM has zero effect on the ROP at that point, which is a largely unreasonable assumption.

A final source of error particular to Well #3 is the data quality, as was discussed in Section 3.2.3. This data was provided to the Wider Windows group in the form of one row of drilling data for each 1.0 ft of well, as opposed to the other wells which were provided with one row for every second of drilling. This means that the data was averaged and managed by the drilling contractor or service company, and is not as noisy as the data which was compressed by the compression program discussed earlier in this project. The data for Well #3 is still a good test of the model fitting accuracy, but the RMSE values for Well #3 should only be compared in magnitude against other Well #3 values, and the trends in these accuracies can then be compared to the RMSE values from Wells #1 and #2.

4.4 - Conclusions about Traditional Models and Fit Accuracy

The total results for the study of the traditional ROP models are shown below in Table 14, which includes the RMSE values for the full-range model fits, the increased
coefficient resolution model fits, and the averages of the RMSE values for each well and each model. After evaluating the traditional models on each of the wells, and with both full range data and increased coefficient resolution methods, a number of conclusions can be reached. First, it can be said that the traditional models when supplied with the full range of lateral data can be reasonably accurate at matching the overall trend of the ROP throughout a section. Each of the models tends to look like a moving average of the actual ROP, and the bulk of the error comes from the traditional models not matching the noise of the actual ROP signal. This establishes that the traditional models have value when used to evaluate the ROP over a long range of drilling. Using these models to estimate the drilling performance for an entire lateral section would be reasonable.
Table 14 - RMSE Values for All Models and Wells, Including Averages

<table>
<thead>
<tr>
<th>RMSE Values (ft/hr)</th>
<th>Well #1</th>
<th>Well #2</th>
<th>Well #3</th>
<th>Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bingham Full Range</td>
<td>23.3</td>
<td>30.2</td>
<td>14.8</td>
<td>22.8</td>
</tr>
<tr>
<td>Bourgoyne and Young Full Range</td>
<td>31.9</td>
<td>34.7</td>
<td>16.9</td>
<td>27.8</td>
</tr>
<tr>
<td>Hareland and Rampersad Full Range</td>
<td><strong>23.2</strong></td>
<td>29.4</td>
<td><strong>14.8</strong></td>
<td>22.5</td>
</tr>
<tr>
<td>Motahhari et al. Full Range</td>
<td>24</td>
<td><strong>27.5</strong></td>
<td>15.5</td>
<td>22.3</td>
</tr>
<tr>
<td>Full Range Model Fit Averages</td>
<td>25.6</td>
<td>30.45</td>
<td>15.5</td>
<td>23.85</td>
</tr>
<tr>
<td>Bingham Increased Coef. Resolution</td>
<td>16.6</td>
<td>19.8</td>
<td><strong>10.3</strong></td>
<td>15.6</td>
</tr>
<tr>
<td>Bourgoyne and Young Increased Coef. Resolution</td>
<td>23.6</td>
<td>28.8</td>
<td>15.0</td>
<td>22.5</td>
</tr>
<tr>
<td>Hareland and Rampersad Increased Coef. Resolution</td>
<td>16.7</td>
<td>23.3</td>
<td>10.8</td>
<td>16.9</td>
</tr>
<tr>
<td>Motahhari et al. Increased Coef. Resolution</td>
<td>16.7</td>
<td>18.3</td>
<td>10.4</td>
<td>15.1</td>
</tr>
<tr>
<td>Increased Coefficient Resolution Model Fit Averages</td>
<td>18.4</td>
<td>22.55</td>
<td>11.625</td>
<td>17.5</td>
</tr>
</tbody>
</table>

Attempting to increase the accuracy of these traditional models by increasing the empirical coefficient resolution has been shown to be a fallacy. Yes, the accuracy of the models can be increased, although it comes as a result of overfitting the model to a smaller range of points. The empirical coefficients which were solved for in this method are not at all representative of the entire lateral, and as the coefficient resolution is further increased become meaningless. Other methods of varying the empirical
coefficients over the length of the section may still be valid, but they would have to be carefully structured as to not allow this same problem to occur.

Overall, the traditional models can be seen as a valid tool for estimating the ROP over a section of the well, but cannot be considered useful for foot-by-foot prediction of the ROP. The models all fail to follow the high-frequency variation in the actual ROP, since this variation is not typically caused by implicit changes in measured input factors, but rather by the erratic nature of the drilling process. Without being able to follow the foot-by-foot ROP of the well, these models are poor tools for predicting and optimizing the drilling process in real time, since they are unable to account for which factors would be contributing to the performance at each individual point. As such, these models would not be useful as the engines behind a real-time performance optimization and control system.
Chapter 5 - Wider Windows Statistical Learning Model

In addition to the traditional models, the horizontal well data was also modeled with the Wider Windows Statistical Learning Model (WWSLM). This model was developed within the Wider Windows research group by Chiranth Hegde, and first presented at the May 2015 Wider Windows sponsor meeting.

5.1 - Introduction to the Wider Windows Statistical Learning Model

The WWSLM has been developed specifically to address the challenge of predicting and optimizing drilling performance. The model includes on a number of supervised techniques, including least squares regression, random forests and ensemble techniques. These statistical learning methods as well as their application to the prediction of ROP with field data are further detailed in the paper by Hegde, Wallace, and Gray (2015). Further methods are being investigated, and will be integrated into the SLM as they are validated.

It is very important to note that this SLM relies only on data which are available at the surface while drilling, and does not depend on any downhole measurements, BHA characteristics, or formation properties for its analysis, presenting a number of advantages when using the model to predict ROP. The WWSLM can be quickly adapted to a new well, BHA, or formation since there are no constants to update, and does not require an operator to run expensive tools to produce downhole measurements of every variable that could influence the ROP. The number of factors that the driller has control
over at the surface is limited, and while there are traditional wisdoms about how to increase ROP by changing these parameters, they are only effective to a certain point. For example, increasing WOB is typically a way to increase ROP, but this only works up to a threshold, and then is no longer effective. The same could be said about changes to RPM, mud flow rate, or the under/overbalance of the drilling mud. There are complex interactions taking place between all of the input parameters, and these interactions are not readily apparent to even the most skillful and experienced drillers. That is where the WWSLM will excel – the supervised learning methods which make up the model are able to identify complex multi-parameter interactions within a system, and can describe the performance of the well using a much more complicated formula than any of the traditional models.

This project does not attempt to detail the inner workings of the WWSLM, as that is being addressed in a parallel Wider Windows project. However since this model has been developed to the point where it can be applied to drilling data, it will be applied to the horizontal well data and compared to the traditional models.

5.2 - Wider Windows Statistical Learning Model Fit Methodology

The data which was provided to the WWSLM was first run through the same compression program which was utilized for the traditional models. For Well #1 (which will be used for this comparison) this resulted in data between 13,500 ft and 19,000 ft with a row every 0.5 ft. The input parameters which were included in this dataset are shown below in Table 15. Each of these parameters was depth based, and are all available
to be read from the rig’s Electronic Drilling Recorder (EDR). Some of these parameters are not seen as typical factors which influence drilling performance, such as block height or hook load, however they have been included anyways since they are available. The beauty of using a statistical model to determine the interactions between these input parameters is that it will decide whether a parameter is actually important to the output variable or not, so including additional parameters which are not immediately evident as being tied to drilling performance is not a concern.

Table 15 - Input Parameters for Wider Windows Statistical Learning Model

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Abbreviation</th>
<th>Units</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measured Depth</td>
<td>MD</td>
<td>ft</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Weight on Bit</td>
<td>WOB</td>
<td>k-lbf</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Block Height</td>
<td>---</td>
<td>ft</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Differential Pressure</td>
<td>---</td>
<td>psi</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Bit RPM</td>
<td>RPM</td>
<td>RPM</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Flow Rate</td>
<td>---</td>
<td>gal/min</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Hook Load</td>
<td>---</td>
<td>k-lbf</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Pump Pressure</td>
<td>---</td>
<td>psi</td>
<td>Depth-Based</td>
</tr>
<tr>
<td>Actual ROP</td>
<td>ROP</td>
<td>ft/hr</td>
<td>Depth-Based</td>
</tr>
</tbody>
</table>

One thing that is important to note about the WWSLM (and all similar statistical models) is that it must be given some data to learn from before it can produce predictions. This data is called the “training data” and it is the first exposure to a data set that the statistical model will have. Within this training data, it has access to not only the values for the input parameters, but also the values of the output parameter of “actual ROP”. The WWSLM analyzes the input parameters and builds a complex network of interactions
that can be used to explain the output variable. After a statistical model is trained, it is then applied to the “validation data”, which is a separate dataset that the model has never seen before, and which does not contain the actual ROP data. The WWSLM then analyzes the validation data, and applies the effects of the complex interactions that it learned with the training data to produce a “predicted ROP” value for each point of the validation dataset.

This process was applied to the horizontal section of Well #1, with the first 500 ft of the 5,500 ft lateral as training data, and the following 5,000 ft as the validation data. After analyzing the data, the WWSLM output its predicted values for the range of the validation data, which has been plotted below in Figure 42. The green line for the WWSLM predicted values starts 500 ft after the actual ROP data, since there were no predictions made during the training phase of the model. Looking at this plot, it is clearly the best fit of any of the models or methods which have been considered so far in this study. The WWSLM has an almost perfect fit with the actual data, and only in the most extreme and brief outliers does it fail to follow the actual ROP data. This WWSLM prediction responds significantly faster than any of the traditional models, and does not exhibit any of the lag which has been seen before. The WWSLM exhibited a RMSE of only 5.0 ft/hr, a very significant improvement over the traditional models.
This model is the first of the models considered in this study that could actually be used for real-time prediction and optimization, since it is able to not only respond to the high-speed drilling data, but make predictions of the more extreme portions of the actual ROP curves. It is essential to have a model which considers the points of highest variance as well as the more "normal" points, because if the goal is to optimize drilling
performance in real-time, that can only be done by knowing what to improve in order to address the swings which are most adversely affecting the ROP.

Comparing the WWSLM model to the previously tested models is shown below in Figure 43. This comparison considers Well #1 against the Hareland and Rampersad model, in both the full range configuration and the increased coefficient resolution configuration. The improvements of the WWSLMs accuracy are very evident when viewed alongside the previous model fits from the traditional models. Additionally, Table 16 below shows the numerical comparison of these three model types, again reinforcing the strength of the WWSLM for predicting drilling performance.
Figure 43 - Comparison of Full-Range, Increased Coefficient Resolution, and WWSLM Model Fits
Table 16 - RMSE Values of Traditional Models Compared to WWSLM

<table>
<thead>
<tr>
<th>Model</th>
<th>Well #1</th>
<th>Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bingham Full Range</td>
<td>23.3 ft/hr</td>
<td>25.6 ft/hr</td>
</tr>
<tr>
<td>Bourgoyne and Young Full Range</td>
<td>31.9 ft/hr</td>
<td></td>
</tr>
<tr>
<td>Hareland and Rampersad Full Range</td>
<td>23.2 ft/hr</td>
<td></td>
</tr>
<tr>
<td>Motahhari et al. Full Range</td>
<td>24.0 ft/hr</td>
<td></td>
</tr>
<tr>
<td>Bingham Increased Coef. Resolution</td>
<td>16.6 ft/hr</td>
<td>18.4 ft/hr</td>
</tr>
<tr>
<td>Bourgoyne and Young Increased Coef. Resolution</td>
<td>23.6 ft/hr</td>
<td></td>
</tr>
<tr>
<td>Hareland and Rampersad Increased Coef. Resolution</td>
<td>16.7 ft/hr</td>
<td></td>
</tr>
<tr>
<td>Motahhari et al. Increased Coef. Resolution</td>
<td>16.7 ft/hr</td>
<td></td>
</tr>
<tr>
<td><strong>Wider Windows Statistical Learning Model</strong></td>
<td><strong>5.0 ft/hr</strong></td>
<td></td>
</tr>
</tbody>
</table>

Again, it should be noted that the WWSLM has only been exposed to 500 ft of the data as training data, and is inferring the rest of its ROP predictions based on its internal model. To compare this to the performance of a traditional model, the Motahhari et al. model was run again, and given the first 500 ft of the well to optimize the empirical coefficients in. Then, the final optimized empirical coefficients were applied to all of the remaining well data for the rest of the lateral. This comparison is shown below in Figure 44. It shows that for the first 500 ft, the Motahhari et al. model tries to follow the actual ROP, but after the empirical coefficients are no longer allowed to be iterated at a depth of 14,500 ft, the model reverts back to being an almost straight-line moving average of the drilling performance. This highlights the unique ability of statistical models to base their learning on only a fraction of the total data and still build an accurate picture of what will cause changes further down the wellbore.
Figure 44 - Motahhari et al. Model Compared to WWSLM with Only 500 ft of Training Data

5.3 - Wider Windows Statistical Learning Model Conclusions

This model has been shown to be the most accurate model for producing a calculated ROP to match the actual ROP in this horizontal drilling data. Furthermore, this model is the only one of the models evaluated which is capable of following and modeling
the foot-by-foot variations in the ROP data. This is due to the WWSLM’s ability to detect complex interactions between the input parameters which are unable to be described by a straightforward equation, and are not detectable to a driller or engineer in the field. Additionally, the WWSLM achieved this result by only evaluating parameters which are available as real-time data on the surface of the drilling rig. This makes the WWSLM the only one of the considered models which is truly predictive in nature, instead of just being formed to the current well. As a result, the WWSLM is the only one of the tested models which could be useful in optimizing and predicting the ROP in real-time.
Chapter 6 - Future Research and Continuing Work

There are continuing efforts underway within the Drilling Parametrics sub-group of Wider Windows that are related to this project. These projects will continue to expand the Wider Windows group’s understanding of the phenomena that affect drilling performance in the downhole environment.

Work is continuing on ROPPlotter, which so far has been designed to create plots based on the traditional ROP models and serves as a good evaluator of depth-based records. ROPPlotter’s capabilities will be expanded to include Torque & Drag as well as Mechanical Specific Energy (MSE).

Based on the success of the Wider Windows Statistical Learning Model in predicting drilling performance based only on the surface-readable input parameters, additional work will expand the WWSLM to include Torque, Drag, and MSE. By including all of these parameters, a much more comprehensive model of drilling performance can be developed. There will also be an effort to use more efficient statistical methods, and incorporate feature engineering so that the model has a stronger basis for predictions.

The paper by Wallace, Hegde, and Gray (2015) titled “A System for Real-Time Drilling Performance Optimization and Automation Based on Statistical Learning Methods,” which is set to be presented at the SPE Intelligent Oil & Gas conference in Abu Dhabi in September of 2015 gives a roadmap for some of the follow-on work in Wider Windows.
Chapter 7 - Conclusions

This project set out to investigate whether the traditional ROP models developed by Bingham, Bourgoyne and Young, Hareland and Rampersad, and Motahhari et al. were effective at modeling the ROP in horizontal shale wells. In order to accomplish this, a dataset of three wells drilled in the Eagle Ford was used. All of these wells were drilled with a rotary steerable system, a downhole positive displacement motor, and a PDC bit, making them very representative of the current state of shale drilling technology. These wells had laterals from 4,500-5,500 ft, with the horizontal section starting at a measured depth of around 13,500 ft and a true vertical depth of 13,000 ft.

The first task was to compress the provided data from having a row every one second down to having a row for every 0.5 ft of measured depth. This was accomplished by created an Excel VBA program that intelligently compressed this data, resulting in a reduction of the dataset from around 260,000 rows to 11,000, and making the model fitting much more manageable.

These three wells were then evaluated with the four traditional models when they were optimized over the entire range of the lateral. This involved iterating the empirical coefficients so that the total root mean square error (RMSE) between the model’s calculated ROP and the actual ROP was minimized. In order to accomplish this, another Excel program was written to handle the optimization task. This produced 12 model fits. These model fits showed that when the full range of the lateral was used for error minimization, the traditional models tend to behave like a moving average of the actual
ROP. Overall, the traditional models were more able to follow deviations of the actual ROP that went lower than the average ROP rather than those that exceeded the average. Each model had its own peculiarities, but when viewed as a group, they were largely representative of a moving average. The overall average RMSE value for this phase of the project was 23.85 ft/hr. This showed that the traditional models could be used as an accurate model of the total performance of a drillstring over a horizontal section of a well, however they are not responsive enough to react to foot-by-foot changes in the input parameters.

The next phase of model fit testing was to increase the resolution of the traditional models’ coefficients so that they were allowed to vary every 50 ft of measured depth. This resulted in the development of another Excel program that performed optimization on every 50 ft interval as it progressed down the length of the well. The resulting RMSE values for these increased coefficient resolution fits were much improved over the full-range fits, with an average RMSE of 15.1 ft/hr. This is a significant increase, and was visible not only in the RMSE value, but also by viewing the plots of the model fits, which clearly were much more responsive and better fitting in this phase of the project. However, this method was found to be a fallacy – by increasing the coefficient resolution the model is just being forced to overfit to the actual ROP data, and these results, while encouraging, are rather meaningless. This was still a good exercise in exploring a different methodology of model fitting, and the program which was developed in this phase can be useful going
forward if applied to sections that encompass different formations, instead of 50 ft sections within the same lateral.

The final model which was tested with the data was the Wider Windows Statistical Learning Model. This model was developed to predict and optimize ROP based only on the parameters which can be measured at the surface of a well, and with no information about the BHA or formation necessary. The WWSLM works by being provided a range of test data (in this case 500 ft of Well #1), and then it uses that information to predict the values of the ROP in the range of the validation data (5,000 ft in this case). This model provided the best accuracy by far, with a RMSE value of just 5.0 ft/hr. From inspecting the plot of this model fit, it shows a near-exact fit of the model’s predictions to the actual ROP data. This is very encouraging since this WWSLM has been developed within the IAP, and with these early signs of its validity, a number of further goals for this model will continue.

This project has been an excellent opportunity to learn how to manipulate ROP models and has provided great insight into what factors affect the drilling performance of a horizontal shale well. The case studies presented here have established that traditional models can model the ROP performance in a horizontal shale well with reasonable accuracy over large ranges of measured depth, but are not accurate enough to be used in an optimization project where predicting and understanding foot-by-foot changes to the ROP is required.
List of Acronyms

BHA: Bottom Hole Assembly
CCS: Confined Compressive Strength
ECD: Equivalent Circulating Density
IAP: Industrial Affiliate Program
MD: Measured Depth (ft)
MSE: Mechanical Specific Energy
MSqE: Mean-Square Error
MWD: Measurement While Drilling
NDB: Natural Diamond Bit
NPT: Non-Productive Time
PDC: Polycrystalline Diamond Compact
PDM: Positive Displacement Motor
RMSE: Root-Mean-Square Error
ROP: Rate of Penetration
RPM: Rotations per Minute
RSS: Rotary Steerable System
SLM: Statistical Learning Model
SPE: Society of Petroleum Engineers
TVD: Total Vertical Depth
VBA: Visual Basic for Applications
WITS: Wellsite Information Transfer System
WITSML: Wellsite Information Transfer System XML
WOB: Weight on Bit
WWSLM: Wider Windows Statistical Learning Model
References


