Evaluation of the Effect of Well Parameters on Oil Production

Wigwe, M. E[1]; Westfall, P. H[2]; Watson, M. C[1]; Giussani, A. [1]; Nasir, E. A.[3]

[1]Bob. L. Herd Department of Petroleum Engineering, Texas Tech University, Boston Ave, Lubbock TX 79409

[2] ISQS, Rawls College of Business Administration, Texas Tech University, Flint Ave, Lubbock TX 79409

[3]Baker Hughes a GE Company, Houston TX

Abstract

Operators in the oil and gas industry are faced with different economic decisions relating to unconventional oil wells. With the popularity of data science and big data analytics tools, a petroleum engineer applies statistical techniques to analyze oil and gas data. We use regression analysis and decision tree in R to evaluate the effect of various well parameters on oil production. Our dataset has over 5700 horizontal oil wells located in the six most productive counties in North Dakota. Two formations present are Bakken and Three Forks. Initial EDA shows that, on average, operators are applying the same drilling and completion techniques across both formations as indicated in a comparative boxplot and two-sample t-test. Linear and "loess" bivariate fit indicates that higher completion parameters lead to higher production. Recursive partitioning trees also support this finding. However, we see reduction in oil production with these parameters if we model production per the different variables. The average well costs in the Bakken increased from $6 – $6.5 million in 2008 to over $9 - $10 million in 2011. More stages or proppant does not necessarily equate to more oil, but more cost.

Key Words: regression analysis, decision trees, EDA, data analysis, big data

1.0 Introduction

In this section, a brief discussion of the terminologies that are important for a good understanding of the analytic modeling work done in the paper. This is probably not the kind of introduction a statistician would expect, but it is hoped that these descriptions would enable the reader who is not familiar with the technical terms follow along.

In the petroleum industry, oil and gas are usually produced from underground storage called reservoir. These reservoirs must have good storage capacity in the form of pore spaces, called *porosity*. In addition, the fluid stored should have the ability to flow to the wellbore via flow paths in the formation which is described as *permeability* (Ahmed, 2010). For conventional reservoirs, the permeability values range from 10 milidarcy to more than 1 Darcy, where 1 Darcy is equivalent to $9.8692 \times 10^{-13}\ m^2$. The wellbore is a conduit constructed/drilled into the target reservoir, which is necessary to bring the oil and gas to the surface. Depending on the thickness of the reservoir, it may be more economical to drill either a vertical well or a horizontal well as shown in Figure 1. In unconventional reservoir, the permeability is in the lower ranges and as a result, it is very difficult for the oil/gas to flow to the wellbore. Typical ranges of permeability for unconventional wells and tight formations are anywhere from 0.1 md to a nanodarcy. With such small conduit for oil and gas to travel through, production from such reservoir would be challenging and/or uneconomical without horizontal drilling and hydraulic fracturing. Hydraulic fractures are artificial pathways created from the wellbore into the formation to remedy or improve the natural connection in order to increase the well's ultimate recovery and enable faster production of reservoir fluid (Economides & Nolte, 2000).

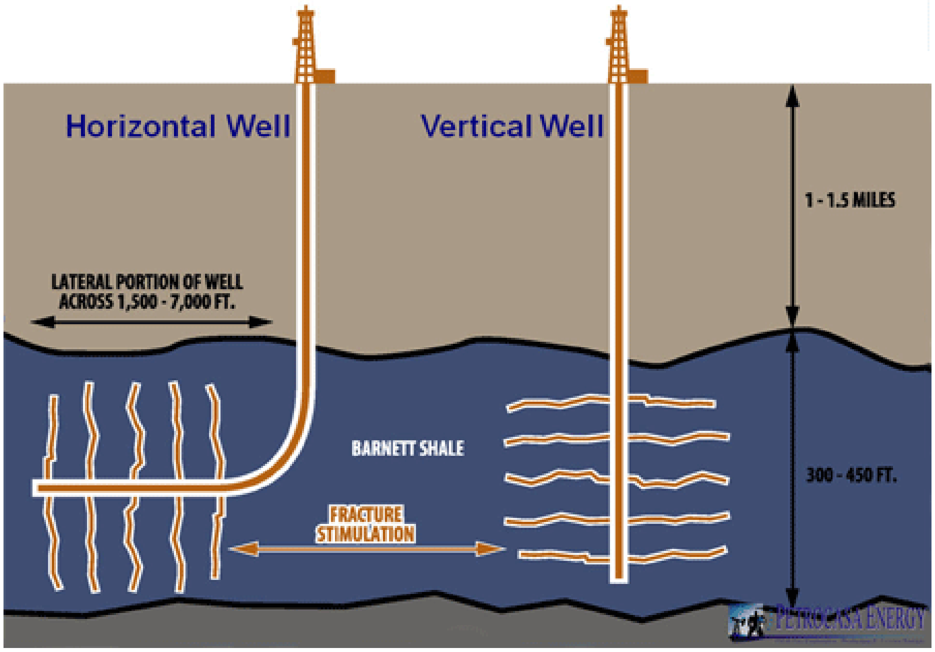


Figure 1: Horizontal vs vertical wellbore (source: Seekingalpha)

At the end of the drilling operation, steel casings are set at the total depth of the well and cemented in place to prevent the well from collapsing and provide a good seal so that formation fluid would not be able to migrate to the surface or adjacent low pressure formations. To create a hydraulic fracture, the casing needs to be perforated in stages and pump frac fluid to create the hydraulic fractures. This frac fluid is composed primarily of water, with the water-based systems accounting for more than 90% of applications and percentage of water of more than 99% have been reported (Alalli et al., 2018). We pump a mixture of water and other fit-for-purpose chemicals into the formation at a high rate for the formation/rock to break down. When this has been achieved, a fracture is created. For the fracture to grow longer into the formation, the pumping process is continued. Since these formations are at great depths, exceeding 2 miles in a lot of cases, the mass of rock from the surface to the bottom of the well will have high enough in-situ stress to close these created fractures when the pumping operation is stopped. This will defeat the entire purpose of fracturing the well in the first place. To avoid this, while the hydraulic fracture is propagating, sand is mixed with frac fluid at the surface and pumped down into the fractures to keep the fractures open. The sand pumped are called proppants.

The fracturing operation is done in stages as shown in Figure 2 for a horizontal well. The fracturing process begins from the toe of the well towards the heel until the entire lateral length or perforated interval is hydraulically fractured. Some important parameters that are measured during the fracing operation that, and used as input variables in this analysis, are briefly describe below and depicted in Figure 2:

1. **Stages:** This is the total number of frac stages on the well. Several perforations can be created in a given stage, but as long as the entire perforations are pumped through at the same time, it's one stage. Figure 2 shows six perforations that are completed/pumped at the same time. This well is an example of a 90 stage well.

2. **Perforated Interval:** This is the length of the horizontal section that was perforated. Hence the length from stage 1 to the last stage. It is a measure of the area of the well exposed to the formation and is measured in feet.

3. **Total Pounds of Proppant:** This is the sum of the amount of sand or proppant pumped in each stage, measured in pounds. Usually the amount varies per stage during implementation of a frac job, although the same amount is specified per stage in frac designs.

4. **Total Volume of Fluid:** This is the volume of water and other chemicals used for the fracing operation. Water alone makes up more than 98% of the frac fluid (King, 2012). This is measured in barrels. Detailed disclosure of chemical composition of frac fluid are reported on fracfocus.org.

5. **Injection Pressure:** The pressure at the surface is usually monitored during a frac operation. This generally provides insight into what is happening downhole. The pressure builds up until the formation breaks down and a fracture is created and begins to propagate. Several authors provide details of the interpretations given to pressure-time plots generated from frac jobs (Nolte & Smith, 1981; Pirayesh, et al, 2013; Soliman, et al, 2014; Wigwe, et al, 2018). The maximum injection pressure recorded for each well is used for analysis in this paper. The unit is pounds per square inch or psi.

6. **Injection Rate:** This is the maximum volumetric flowrate at which fluid is being injected into the well. It is measured at the surface and the unit is barrels per minute or bpm.

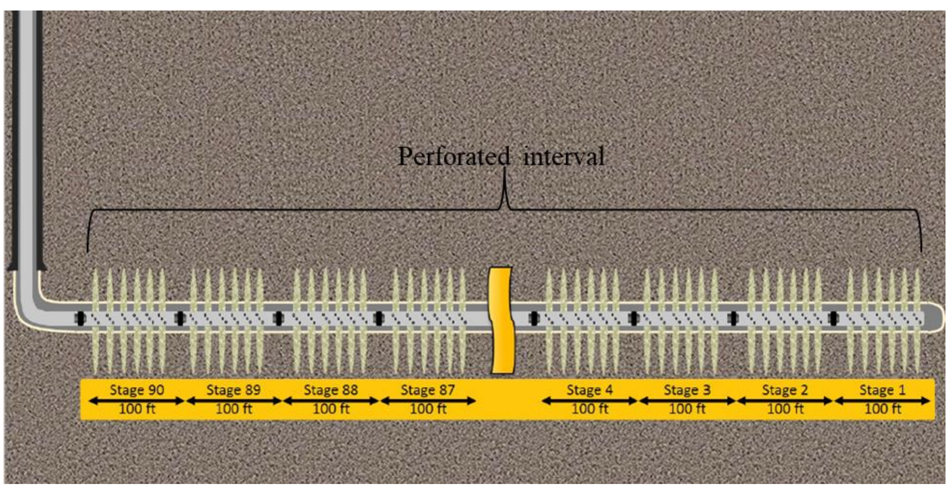


Figure 2: Wellbore schematics (modified from UOG Training, (Burton, 2013))

1.1 Motivation

Some portions of this review have been published in a recent paper with SPE (Wigwe, et al, 2019). Recent researchers in the oil and gas industry have focused on applications of the “newer” data analytics techniques such as decision trees, artificial neural network, support vector machines, deep learning and the variations in addition to traditional regression analysis. These techniques have thrived due to availability of computational resources and generally use more complex algorithms to capture the inherent (non-linear) relationship between the dependent variable and the predictor variable(s). The discussion by Cunningham, et al (2012) used multiple linear regression (MLS) model for analysis. The authors looked at four different areas in the Marcellus formation where EQT Production Company has drilled over 124 horizontal wells. They build the regression models by area and also combined all 124 observations available for analysis. Another paper by a different oil and gas company was presented by Martinez and Wray (2014). They showed that using 6 months cumulative production was a good proxy for EUR. Using non-linear regression, they were able to optimize completions design and set expectation for production performance. One particular paper (Bhattacharya et al., 2013) used the classification and regression trees (CART) for analysis (Bhattacharya et al., 2013, p. 13).

Data mining technologies have been used in the characterization of 187 wells distributed over 11 counties in northern West Virginia portion of the Marcellus (Zhou, et al, 2014). This was done to identify correlations between gas production performance of the wells and attributes of the completion and geological setting and to identify important factors useful for predicting gas recovery. The use of data analytic techniques, alongside power-law exponential decline models (Ilk, et al, 2008) have been investigated and used in forecasting water production from the Marcellus (Ettehadtavakkol & Jamali, 2019). Several other papers have been presented where data analytics techniques have been used to quantify and evaluate completion parameters in the Bakken and Three Forks (Lolon et al., 2016; Scanlan et al., 2018; Wang & Chen, 2016), while effects of geomechanical properties, interaction of natural fractures, temperature and other frac design parameters have been discussed extensively (Kolawole, et al, 2018; Kolawole & Ispas, 2019; Wigwe, et al, 2019).

Some observations made from all the analysis done in these papers are:

1. Some of the researchers focused on building models with ALL available data ONLY
2. Some others go the extra mile to split the data into training and testing sets
3. In all the papers, the models were assumed to be deterministic
4. The models are then evaluated by looking at the goodness-of-fit parameters.

In this paper:

1. We have accounted for the inherent variability or probabilistic nature in neural network models
2. We take the modeling process a step further by performing sensitivity analysis on the predictor variables.

1.2 Methodology

Exploratory data analysis is a great starting point for any data analytic workflow or project. It is useful to reveal important characteristics of the data through visualization (Hoaglin, et al, 2000). This will be used initially to get a feel of the dataset by recognizing patterns in both univariate and bivariate plots (EMC Education Services, 2015; Holdaway, 2014). The other (primary) tool that would be used is artificial neural network (ANN). Artificial neural networks were inspired by the way the human brain works. The neuron receives data sent by the external senses (vision, smell, feeling, sound, etc.), processes this information and sends a response back for an action to be taken. In ANN, the neuron acts as the CPU, where all mathematical operations are performed to generate an output from a set of inputs (Ciaburro & Venkateswaran, 2017). The inputs form the input layer, the middle layers are called the hidden layers while the output forms the output layer. All the computational work is done in the hidden layer(s) and requires a knowledge of weights, bias and activation function (SAS Institute Inc., 2018; Wigwe, et al, 2019).

Figure 3 shows the architecture of a neural network and weights and how they are used for calculation in the hidden layer. In simple terms, $y = f(x) = \Sigma x_i w_i + bias$. It has one hidden layer with 5 neurons. If there are several hidden layers (as is the case with deep learning), the output from each layer becomes input to the next layer. Some common activation functions available are linear, step, logistic, hyperbolic tangent and rectified linear unit. The configuration shown in Figure 3 is an example of a feed-forward propagation where processing from input layer to hidden layer continues to the output layer. In feedback networks, the results of the computation by each neuron is fed back to the neuron as inputs and the weights and biases are updated with the error times the derivative of the activation function. The goal of this updating sequence is to minimize the error, leading to a better fit of the model to the data.

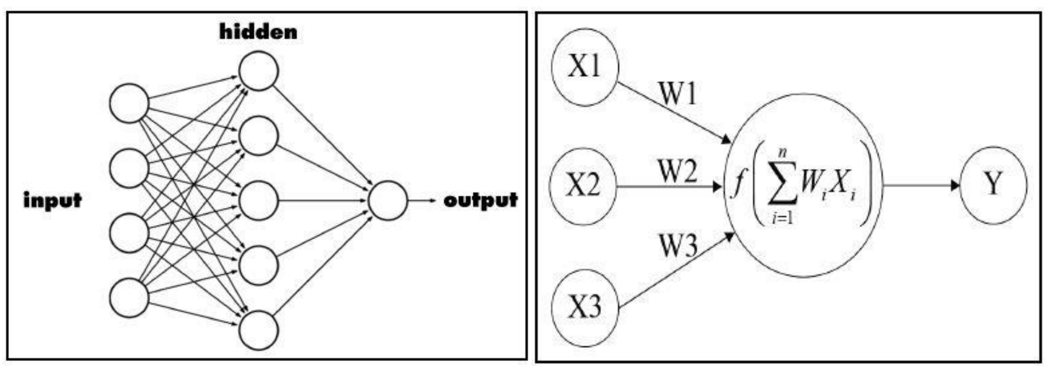


Figure 3: ANN Layers showing activation functions and weights (Ciaburro & venkateswaran, 2017)

In summary, an ANN has been described as a universal approximator of any continuous function (Hornik, et al, 1989) with primary application for building models to forecast future values of a dependent variable (Olden & Jackson, 2002). The *neuralnet* library (Guenther, 2016) was used to implement the NN methodology in R/RStudio (R Core Team, 2018; RStudio Team, 2016). The article by Schmidhuber (2015) provides a detailed historical review ANN, particularly deep learners.

In addition, “ensemble” models are used to obtain a more representative result of the analysis. Ensemble models typically leverage the power of using multiple algorithms to get a better prediction than would have been obtained if only one algorithm is used, using a voting mechanism (EMC Education Services, 2015). In ANN, predicted results changes depending on the training data used for building the model. As a result, researchers could find themselves in a dilemma of what result to present. To avoid such situation, in addition to accounting for the variability in ANN models, the dataset multiple is resampled times to get different training sets for building multiple models. We combine the results and present the average, which is a more representative solution. We present a 95% confidence interval bounding this average on a bivariate marginal plot to give an idea of the performance of the final model.

The ANN workflow would involve the following:

1. Use of a feed-forward network
2. Specification of 3 nodes in one hidden layer
3. Use of a linear activation function
4. Feature scaling the data using min-max method
5. Specify a random seed that controls which data is sampled
    a. Sample 75% of the data as training data and use the rest as test data
    b. Fit a neural network model using the training data
    c. Validate the performance of the model using test data
    d. Calculate and store model diagnostics like R^2, RMSE, RMSPE and log-likelihood
    e. Make predictions of the dependent variable, while varying one predictor and leaving the others constant. We vary the predictor from the smallest to largest value in the dataset.
6. We repeat step 5 a specified number of times (say 1000) with different random seed values. For each model, the results are stored for each variable.
7. Find the average of the 1000 predicted values of the dependent variable for each predictor and construct a 95% confidence interval about the mean.
8. Display the results as marginal plots and interpret.

2.0 Data Gathering, Cleaning and EDA

More than half of the time spent in a data science project is spent at this stage of the project, data preparation (EMC Education Services, 2015). This is because it is important to put your data in the right condition to be used for modeling. This can be a back and forth process throughout the research work, especially if the data comes from multiple sources.

2.1 Data Gathering and Cleaning

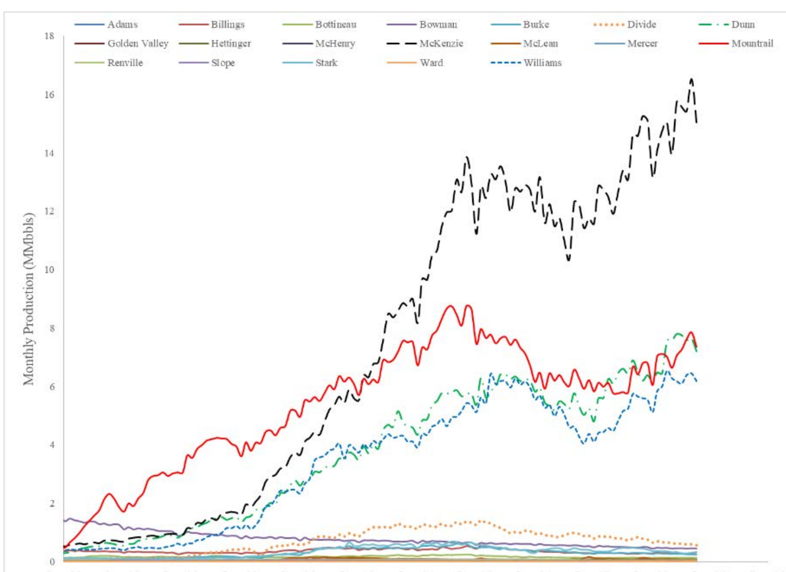
Two primary data source for this project are DrillingInfo, an oil and gas data vendor and service provider (Drillinginfo, 2018) and the North Dakota Industrial Commission's Oil & Gas Division (NDIC O&G Commission, 2018). These are public data and effort was made to correct suspicious data, rather than throw it away. In one situation, an operator had reported 28 million pounds of proppant on a 25-stage job. This did not make sense in a univariate or bivariate space, when looked at in combination with other variables (number of stages in this case). We reached out to the operator and they confirmed that the data was inputted incorrectly and provided the correct number to be 2.8 million pounds. This is more reasonable value for a 25-stage job. Another operator reported 18 million pounds of proppant, this made sense for a 47 stage frac job. A lot of the time, data scientists are working on projects in areas they have little or no knowledge about. Therefore it is important to ask the project sponsor or a knowledge expert in the field whether or not the data makes sense. Table 1 is a contingency table showing the distribution of wells in each county by formation. The most activity is clearly in the McKenzie County.

Table 1: Distribution of wells by County and Formation

County	Bakken	Three Forks	Sum
Burke	87	51	138
Divide	119	186	305
Dunn	658	365	1023
McKenzie	1185	678	1863
Mountrail	850	396	1246
Williams	843	337	1180
Sum	3742	2013	5755

3.0 Exploratory Data Analysis

The case study presented is the Bakken formation. In the Bakken and Three Fork, about 88% of the 170 barrels most likely estimate of the OOIP are located in six of the nineteen producing counties in North Dakota, according to NDGS 2010 assessment (Nordeng & Helms, 2010). These counties are Burke (10.04%), Divide (10.46%), Dunn (11.87%), McKenzie (21.51%), Mountrail (17.10%) and Williams (17.11%). The historical monthly oil production reported by the NDIC is shown in Figure 4 (top). This figure also shows these counties remain top producers during the past 10 years. The wells shown in **Table 1** were completed between 2008 and 2016, with at least one year of production recorded. Several libraries used are available in R software (R Core Team, 2018). Temporal variables were converted to date, using the *lubridate* package (Grolemund & Wickham, 2011). Figure 4 (bottom) shows the number of wells in the dataset by year completed. The peak corresponds to the period of high oil price before the price collapse of late 2014. Figure 5 shows the map of the study area.



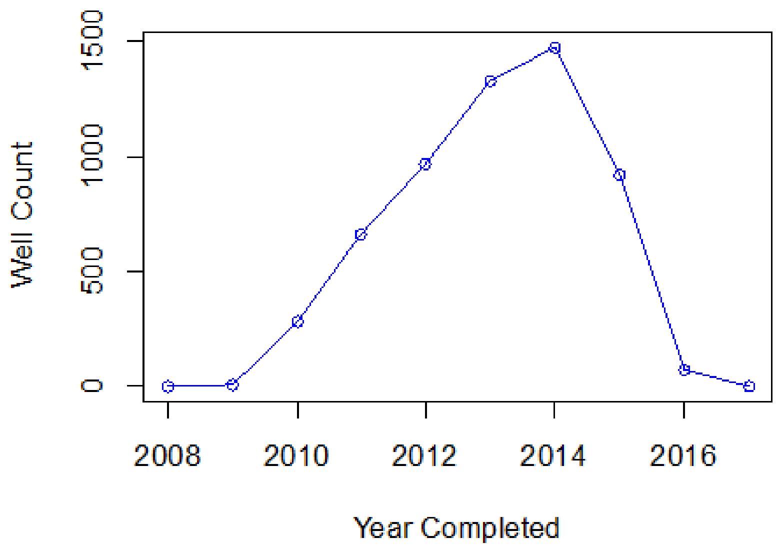


Figure 4: Historical County Production for ND (top) and Wells Completed by Year (bottom)

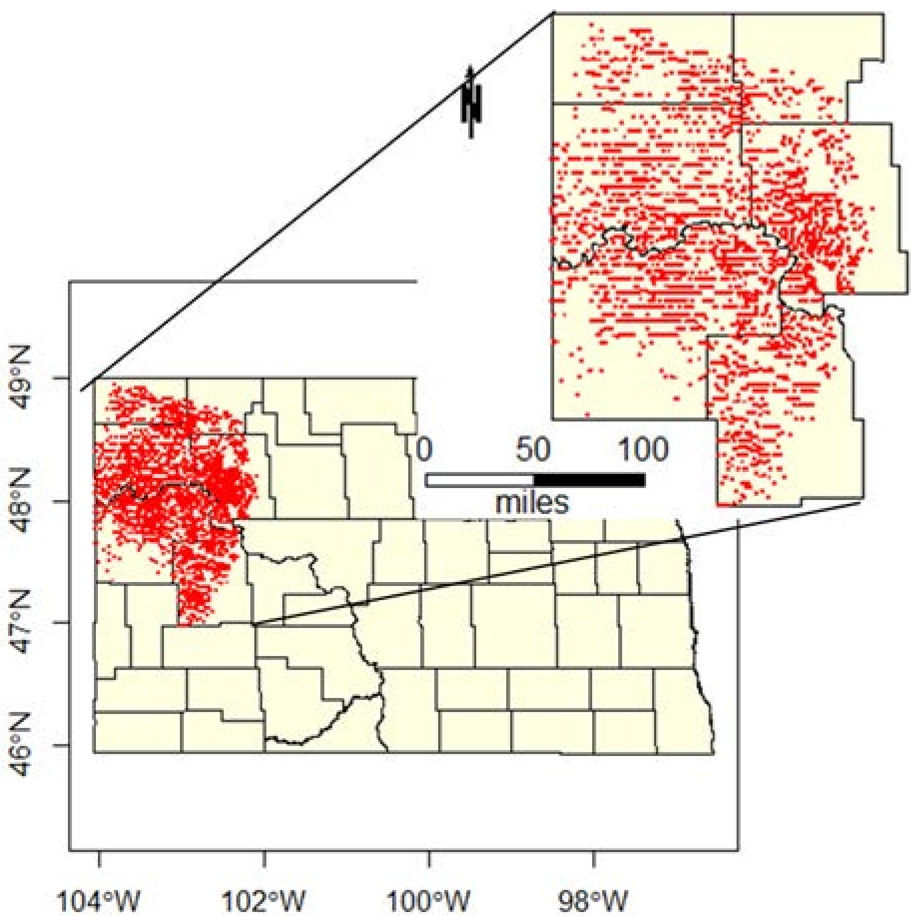


Figure 5: Map of North Dakota showing area of study with wells as red dots (Wigwe, et al, 2019)

3.1 Distribution of Completions Parameters

The distribution of the number of stages, total pounds of proppant, total volume of fluid injected and the perforated interval typically used in frac jobs in the Bakken and Three Forks formations are shown in Figure 6, Figure 7, and Figure 8. The comparative boxplot shows the number of stages, Figure 6 a. On average, operators are using the same application in both formations for the number of stages (30 stages). There is more variability in the number of stages for the Bakken compared to the Three Forks formation. Figure 6 b shows that most operators favor a perforated interval in the 8,000 ft. to 11,000 ft. range on the lateral. The distribution of total pounds of proppant used for the frac jobs is as shown in Figure 6 c. Most frac jobs used less than 5 million pounds of total proppants (the red line). As will be shown later, of the 656 occurrences of application of more than 5 million pounds of total proppants, only 83 cases occurred prior to 2014 (Figure 7 c). This indicates that the use of large pounds of proppants started becoming popular during the downturn. To summarize, on average, 75,000 barrels of fluid and 3.5 million pounds of proppants were used for the 30-stage completion of a 9,300 ft. perforated interval between 2008 and 2016.

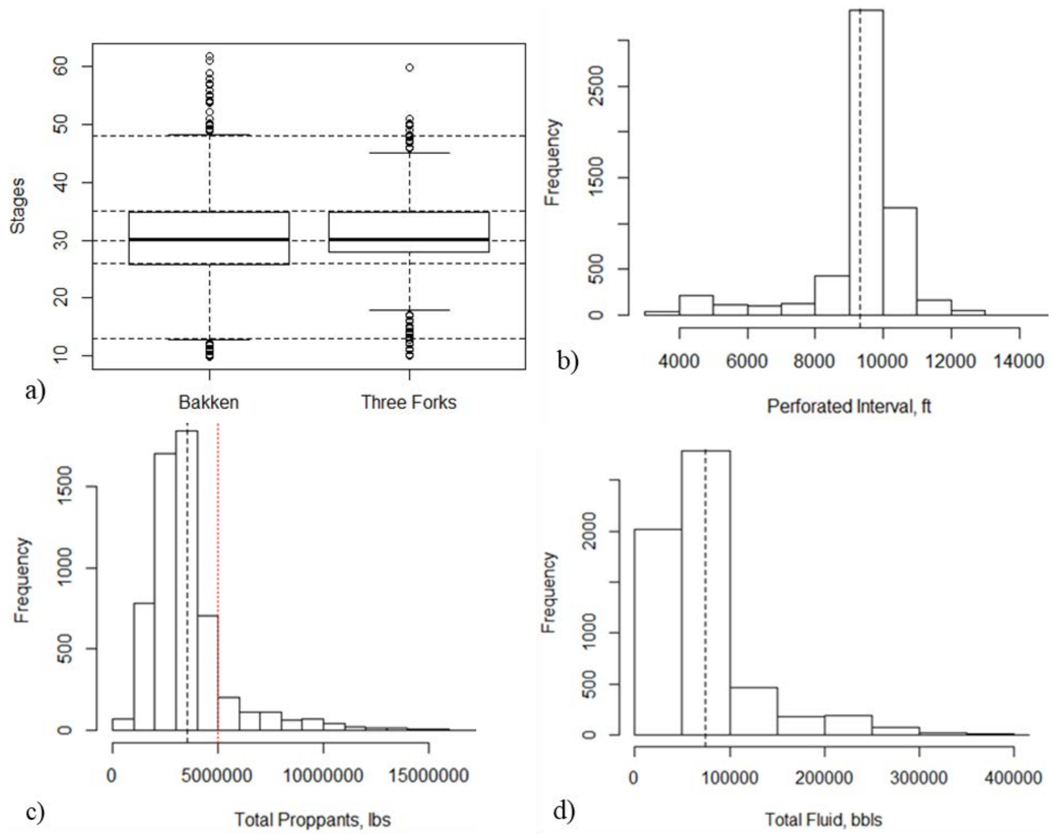


Figure 6: Distribution of Completions Parameters, vertical dashed lines are Averages

3.2 Change in Completion Parameters since 2008

Figure 7 shows the variation of completion parameters with time from 2008 to 2016. Figure 7 a shows an increasing trend in the number of frac stages used in completions. There does not appear to be a systematic change in the length of lateral and perforated interval since 2008 (Figure 7 b). However, there is an increasing tendency towards drilling longer and perforating the laterals in the 9,000 ft. - 11,000 ft. range. As mentioned previously, the use of more than 5MM lbs of proppants started becoming popular after 2014 (Figure 7 c). Most of the completions prior to 2012 utilized less than 100,000 barrels of total fluid and the use of more than 100,000 barrels of total fluid became increasingly popular from 2013 and well into the downturn (Figure 7 d). This tendency to use more proppants and higher fluid volume meant that operators could complete fewer wells with a view to "increasing" production (Figure 4) and spending less money on average.

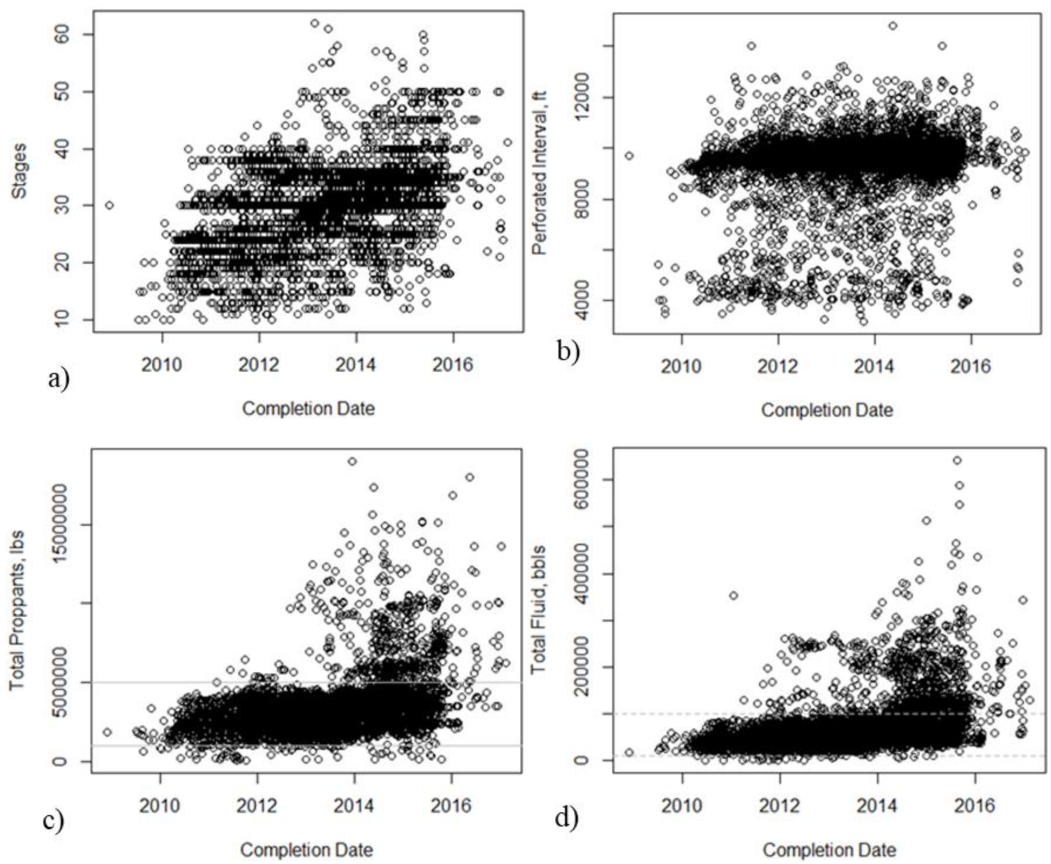


Figure 7: Distribution of Completions Parameters Cont'd

Injection rate and injection pressure are two important parameters that contribute to an efficient hydraulic fracture treatment. The limits to these two parameters are usually pre-set by the capacity of the pumping equipment used for the treatment. The injection rate seems to be similar on average for both formations with a value of around 40 bpm (Figure 8 a & b). Injection pressure on the other hand correlates with the true vertical depth of the formation and hence expect a slightly higher pressure for the Three Forks than the Bakken on average. This is because the Three Forks directly underlies the Bakken formation and hence is a slightly deeper formation. On average, the maximum injection pressure recorded is 8360 psi (Figure 8 c & d).

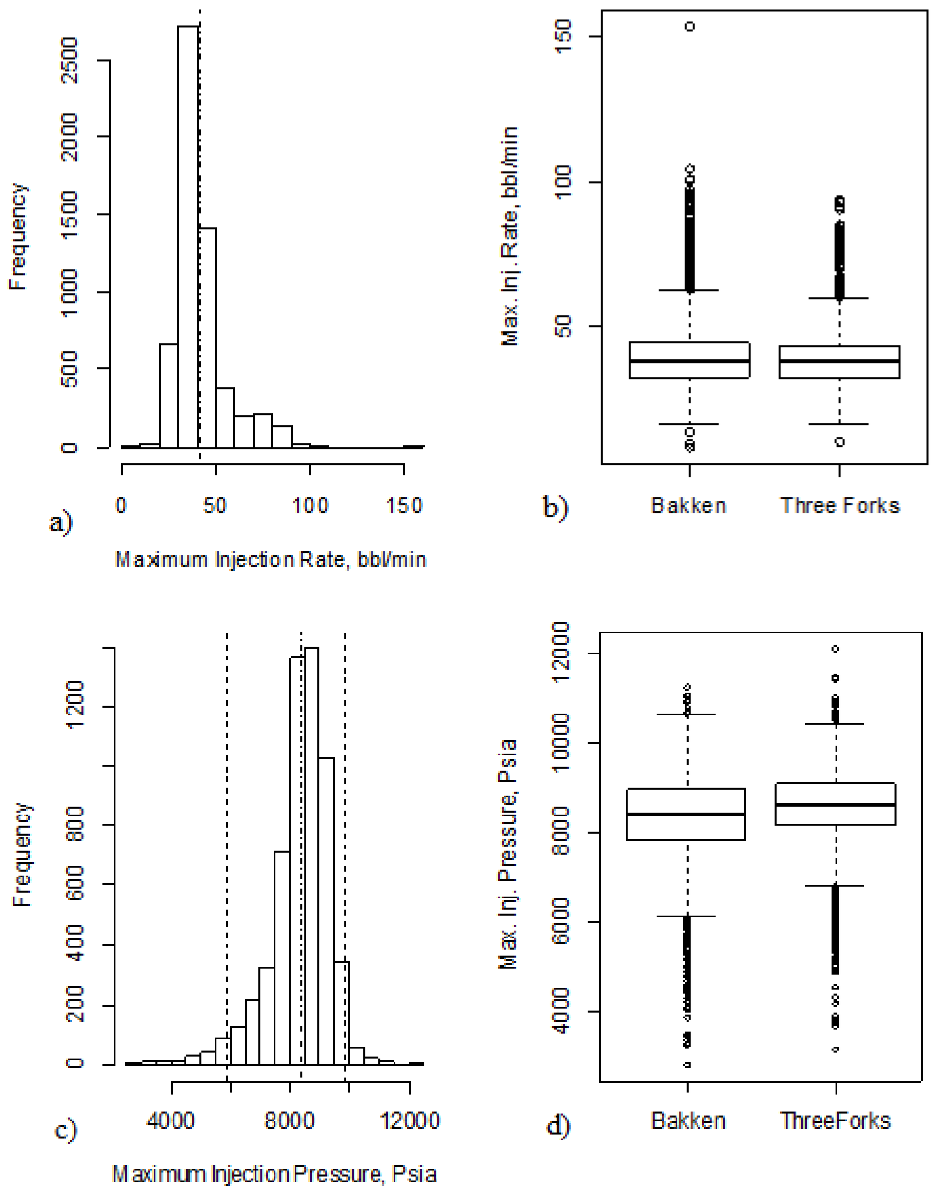


Figure 8: Distribution of Injection Rate and Pressure

Using the *corrplot* package (Wei & Simko, 2017), the plot of the correlation matrix of all the variables along with the dependent variable (six-month cumulative oil production) is shown in **Figure 9**. This plot shows that the perforated interval and total proppant have moderate correlations with the number of stages ($r = 0.53$), while as expected, the total fluid and injection rate have $r = 0.67$ and the total fluid and total proppant have $r = 0.62$. There is a very weak correlation among the other variables.

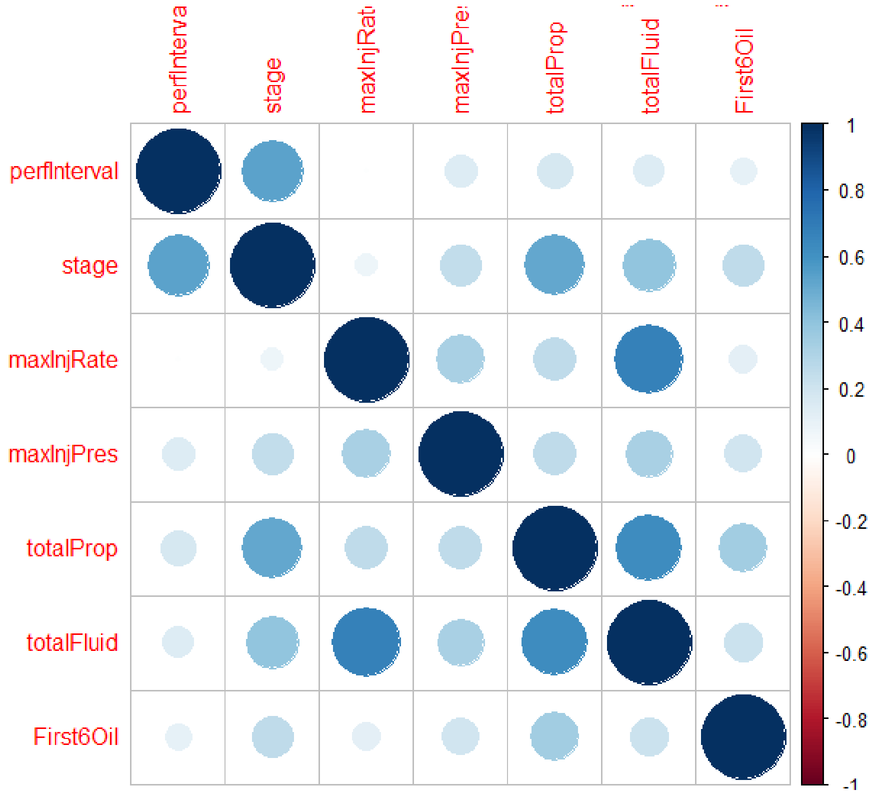


Figure 9: Correlation plot of variables

3.3 Subset Data to Complete Cases

A summary of the data with complete cases only is shown in

Table **2**. The data is normalized, split into training and testing set, and used to build ANN models as described in the workflow under the 1.2 **Methodology** section. The results will be discussed next.

Table 2: Distribution of wells by County and Formation – Complete cases only

County	Bakken	Three Forks	Sum
Burke	55	22	77
Divide	69	80	149
Dunn	340	129	469
McKenzie	436	169	605
Mountrail	512	147	659
Williams	441	96	537
Sum	1853	643	2496

4.0 Results and Discussion

The predictions were made by varying one variable at a time while keeping the others fixed at the average values as shown in Table 3. The scaled versions were used for modeling and prediction before re-scaling back to the original units. Extrapolation and interpretation of the results obtained beyond the limits of the x-axis is discouraged as the results presented holds only within the limit of the data used for analysis.

Table 3: Average values of Variables

Variable	Actual	Scaled
Perforated Interval (ft.)	9,100	0.544
Stage	28	0.337
Total Proppant (lbs.)	2,778,000	0.2197
Total Fluid (bbl)	54,400	0.1513
Injection Rate (bpm)	39	0.1997
Injection Pressure (psi)	8,100	0.5672
Sample size, n	2496	2496

4.1 Effect of Number of Stages

Figure 10 shows the variation of the number of stages on oil production. The general observation, based on the data, shows an increase in oil production up to 25 stages and oil production starts decreasing for higher stages. There seem to be a time and drainage component acting as confounding variables in this plot, which will be investigated further. This is a data-driven model, as a result the plot does tell the story of what is going on in the formation. The model does a good job based on the amount of representative data used. The 95% confidence interval is shown on the bivariate plot on the left and the conditional distribution of the predicted values at stage 30 on the right of **Figure 11**. In the dataset, about 94% of the wells were completed using 15 to 40 stages, hence there is a high level of confidence in the prediction of oil production in this range of stages. The most common stage count (30 stages) makes up 25% of the dataset. Hence, completing a 30 stage well in the Bakken, keeping the other parameters at the average values as shown in **Table 3**, will result in 52,300 – 54,600 barrels of oil production after six months, with 95% confidence.

Completions with less than 15 stages were more common when development of the Bakken began in 2008 and only few wells use completions above 45 stages by 2016. The marginal plot shows that even for a well with the same characteristics, there is a probability distribution of oil production at different stages, rather than a fixed, single value. There is more variability in the distribution (captured by the 95% confidence interval) beyond data points with high probability of occurrence in the dataset.

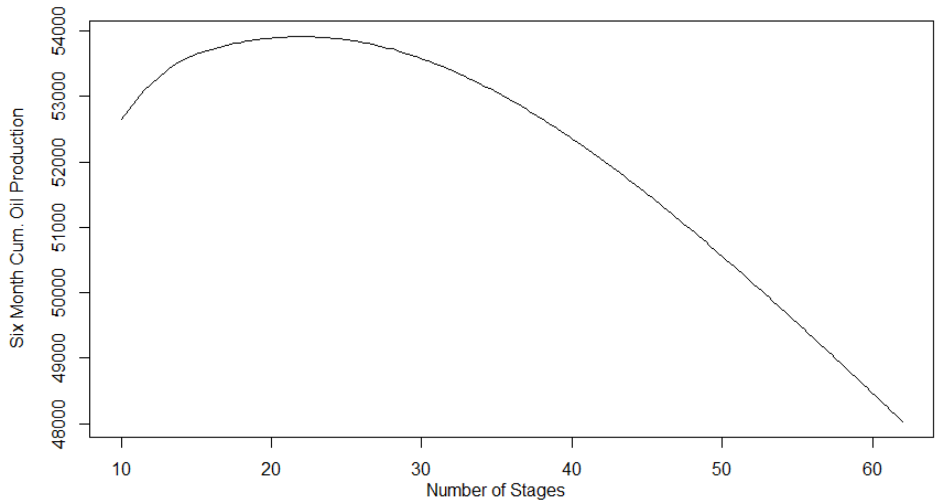


Figure 10: Variation of Oil Production with Number of Stages

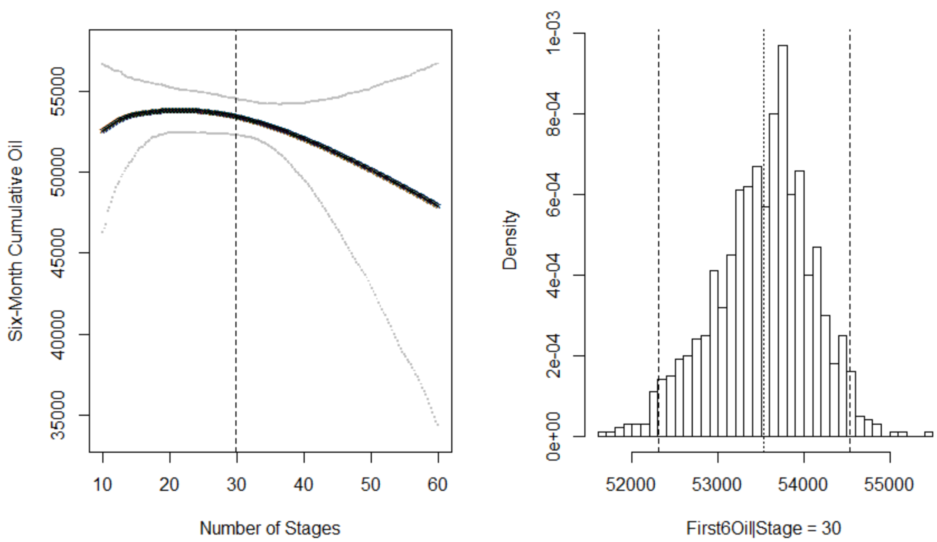


Figure 11: Conditional Distribution of Oil Production for 30 Stages showing 95% Confidence Interval

4.2 Effect of Perforated Interval

The effect of the perforated interval is shown in **Figure 12**. Similar to the number of stages, on average, oil production decreases with increasing perforated interval, showing two distinct slopes, with a transition period between. The steeper slope from 4000 – 5000 ft. corresponds to wells drilled on a one section spacing, which was typical during the early development of the Bakken. This corresponds to 6% of the wells in the data. While the smaller slope from 7000 – 10,000 ft. corresponds to later wells drilled on a two section spacing. This makes up 70% of wells in the data. The transition period is between 5000 ft. and 7000 ft. The data has wells with longer laterals, making up the rest of the data.

The result shown in **Figure 12** suggests that smaller perforated intervals correspond to higher production, and the difference in production beyond 7000 ft. of perforated interval is not very large. Similar to number of stages, earlier Bakken completions utilized less than 20 stages and one section drilling. Furthermore, there were fewer to no wells drilled in the area prior to 2008. Hence, a possible sign of depletion affecting newer wells could be observed. As a result, these newer wells relied on aggressive completions from operators to achieve high production results.

Figure 13 shows the conditional distribution of oil production, bounded by 95% confidence interval around the average. Due to the number of wells between 6000 ft. and 11,000 ft. (88% of the wells), the model's predicted oil production values are more accurate than in areas with sparse data.

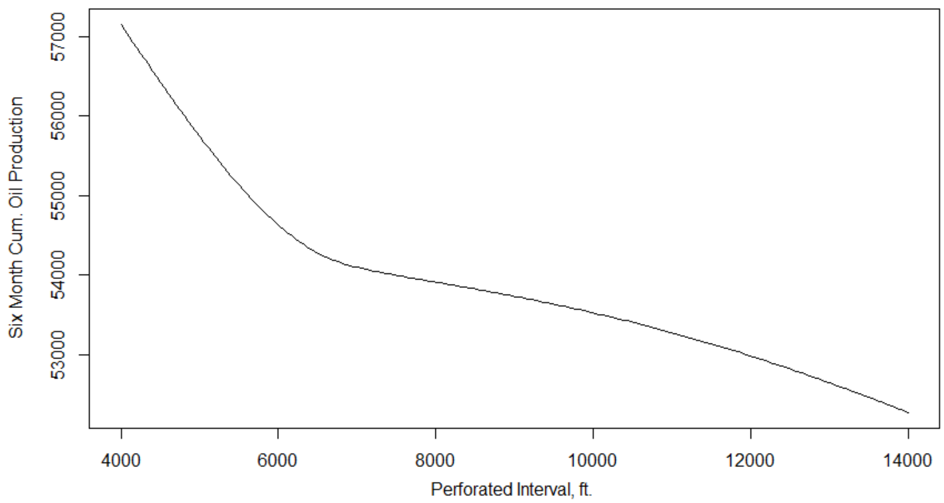


Figure 12: Variation of Oil Production with Perforated Interval

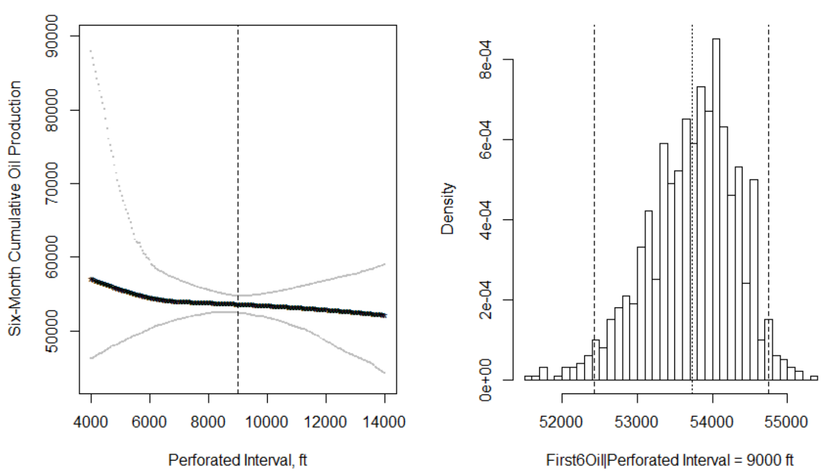


Figure 13: Conditional Distribution of Oil Production for Pi = 9000 Ft showing 95% Confidence Interval

4.3 Effect of Total Pounds of Proppants

The results for the effect of proppants injected is shown in **Figure 14**. Unlike the previous two cases, increase in the amount of proppant has a positive impact on oil production. The data suggests that 61% of the wells were completed using 1 – 3 million pounds of proppant, with 35% utilizing between 3 and 5 million pounds of proppants. Based on **Figure 15** (left), there is more variability in the oil production beyond 6 million pounds of proppants, suggesting less data in that region. Most applications utilized less than 5 million pounds of proppant. Due to the cost of proppants, length of fracture created and the limitations of the pumping equipment, it may not be feasible to pump more sand than the formation can take. For a 3 million pound frac job in the Bakken, keeping the other parameters at the average values as shown in **Table 3**, oil production of 52,800 to 55,000 barrels in six months with 95% confidence is predicted.

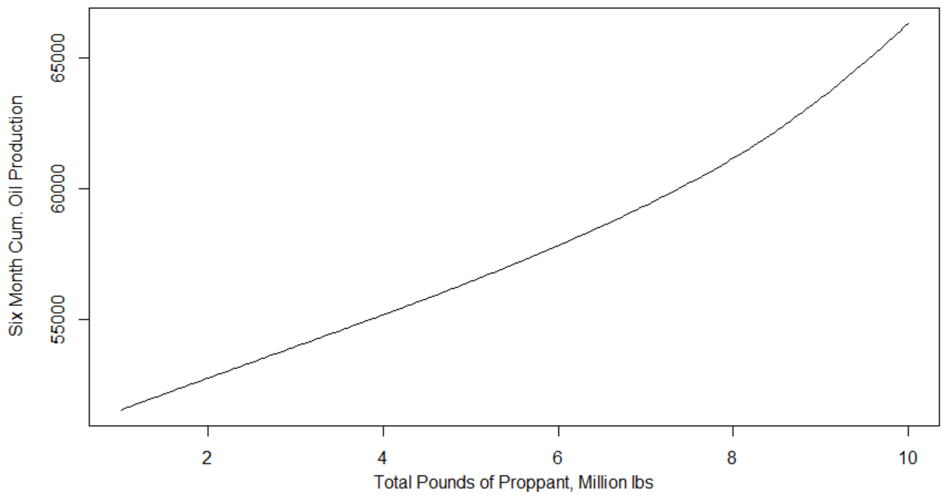


Figure 14: Variation of Oil Production with Total Proppant

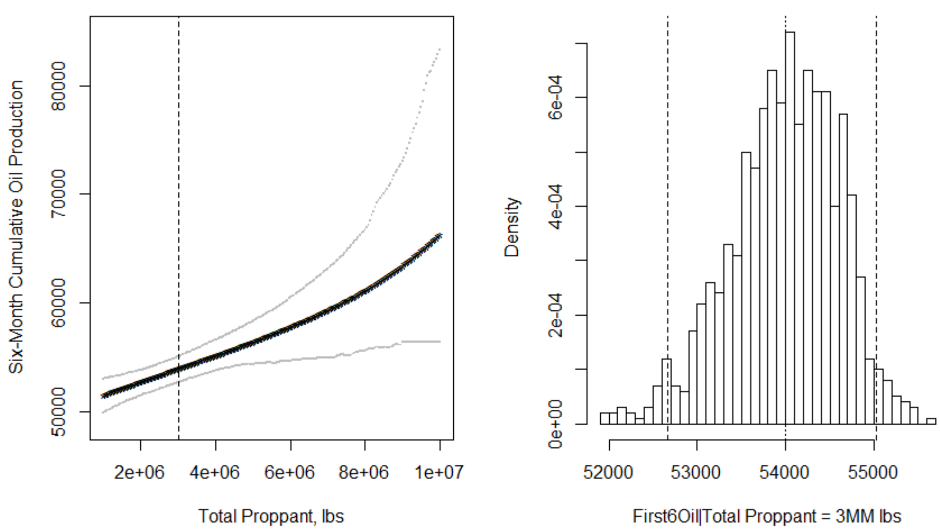


Figure 15: Conditional Distribution of oil Production for TP = 3MM lbs showing 95% Confidence Interval

4.4 Effect of Total Volume of Fluid Pumped

Similar to total proppant, the total fluid injected has a positive effect on oil production as shown in **Figure 16**. As show in **Figure 9**, the total fluid is correlated with the total proppant. Most applications (95% of cases) utilized less than 100,000 barrels of fluid, with half of that being less than 50,000 barrels. Very few jobs used more than 100,000 barrels (4% of cases). **Figure 17** (left), shows a bivariate plot of the model results with 95% confidence interval shown. The increasing variability shown beyond 100,000 barrels of fluid is indicative of less data coverage.

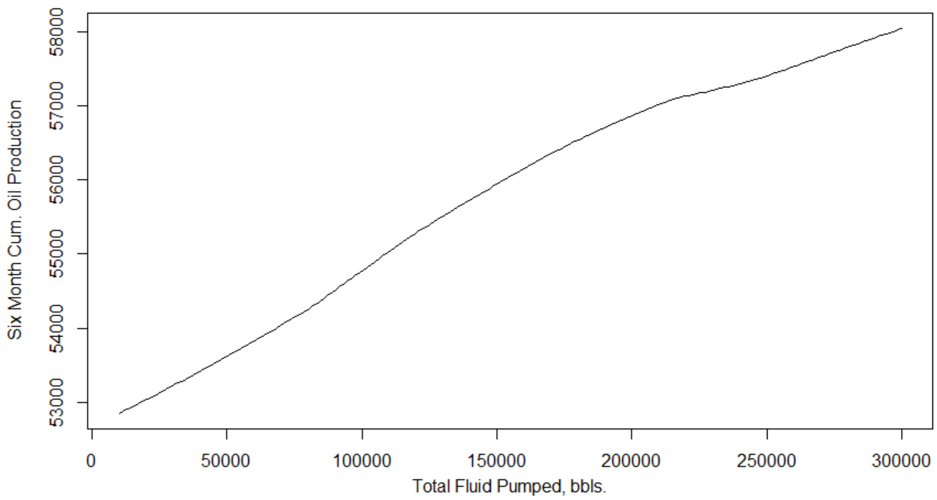


Figure 16: Variation of Oil Production with Total Fluid

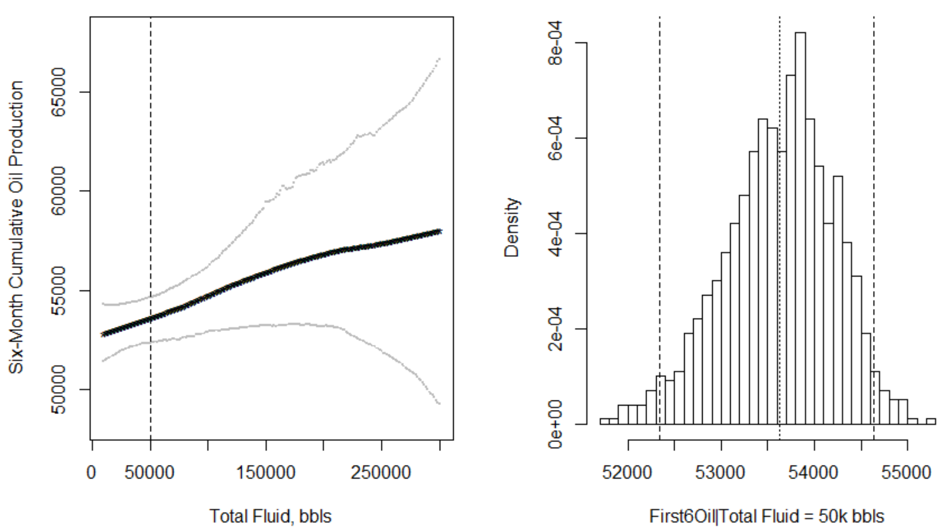


Figure 17: Conditional Distribution of oil Production for TF = 50K barrels showing 95% Confidence Interval

4.5 Effect of Maximum Injection Pressure

The results for the effect of injection pressure is shown in **Figure 18**. This plot shows a decreasing trend with increasing injection pressure. The injection pressure is a function of formation breakdown (frac) pressure, which depends on the true vertical depth of the well; this is going to be slightly different for each stage. The higher injection pressure could mean less brittleness of the rock. If this is the case, more pressure would be required to initiate a fracture. **Figure 19** (left), shows the 95% confidence interval and the conditional distribution of six-month oil given that the injection pressure is 8000 psi. About 95% of the wells have maximum injection pressure above 6,000 psi, hence the variability in the prediction is more consistent in this range.

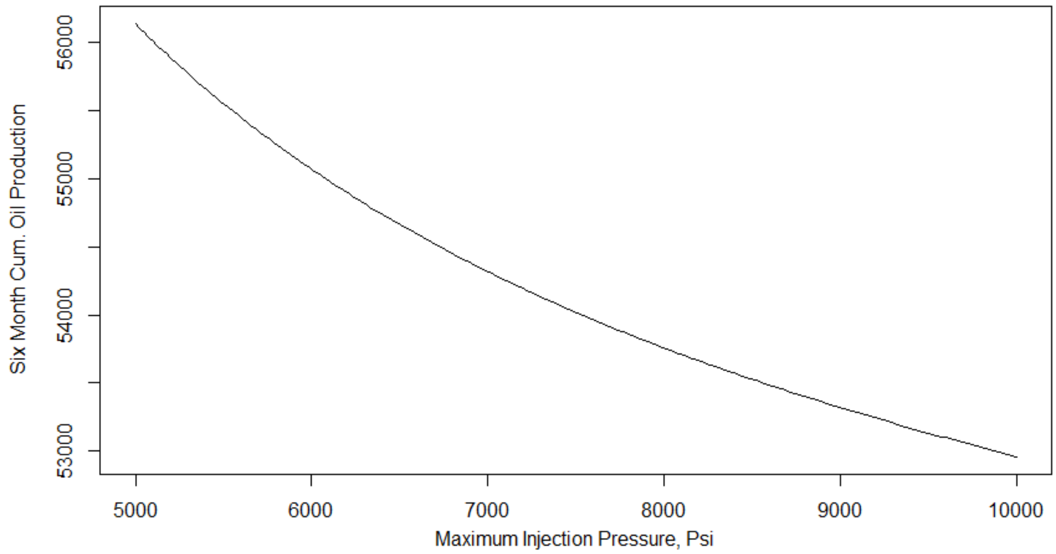


Figure 18: Variation of Oil Production with Injection Pressure

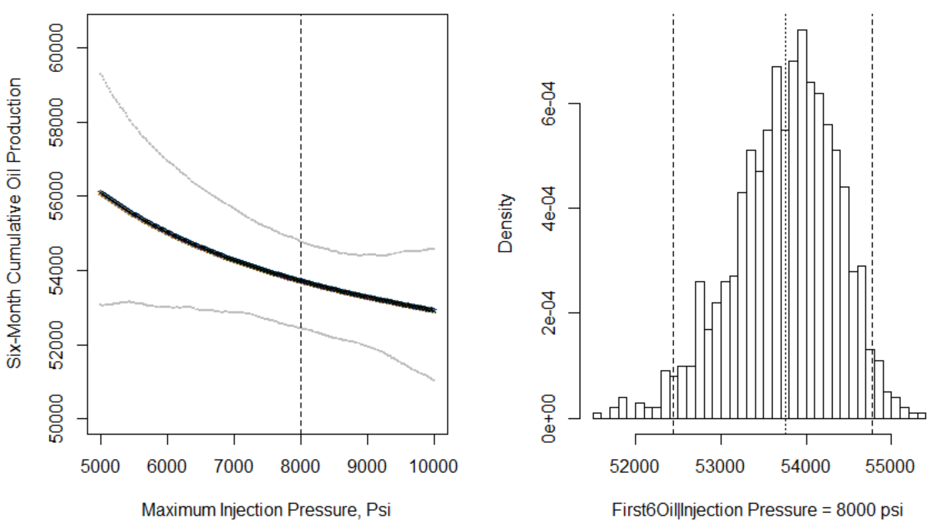


Figure 19: Conditional Distribution of oil Production for IP = 8000 psi showing 95% Confidence Interval

4.6 Effect of Maximum Injection Rate

The results for the effect of maximum injection rate is shown in **Figure 20**. This does not seem to have a reasonable impact on oil production. This is because the oil production is does not change much after six months when the injection rate is varied from 20 to 100 bpm. 90% of wells were pumped at 50 bpm or less, resulting in the level of accuracy depicted in **Figure 21** (left).

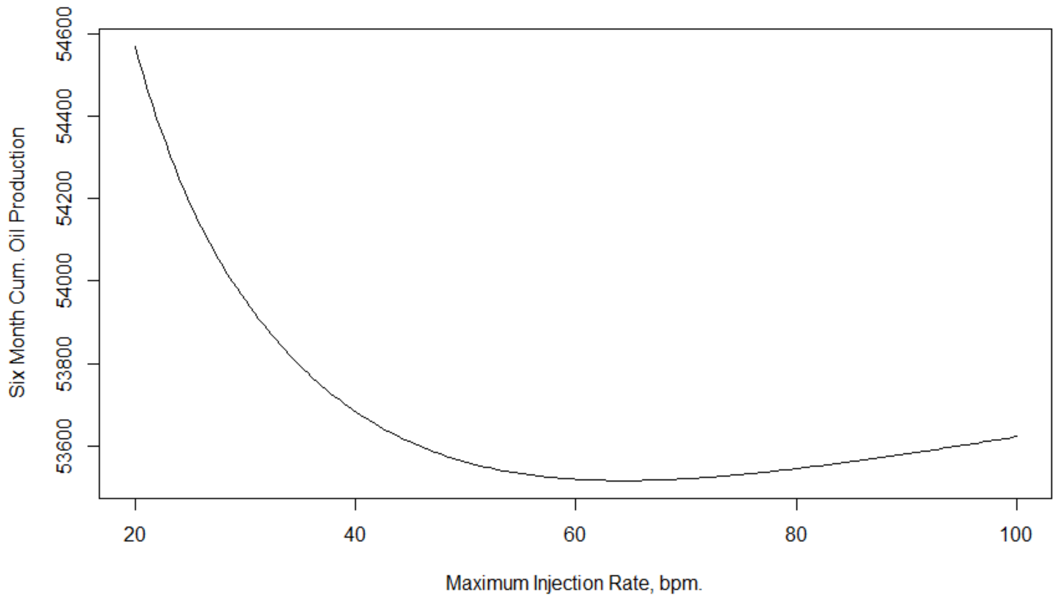


Figure 20: Variation of Oil Production with Injection Rate

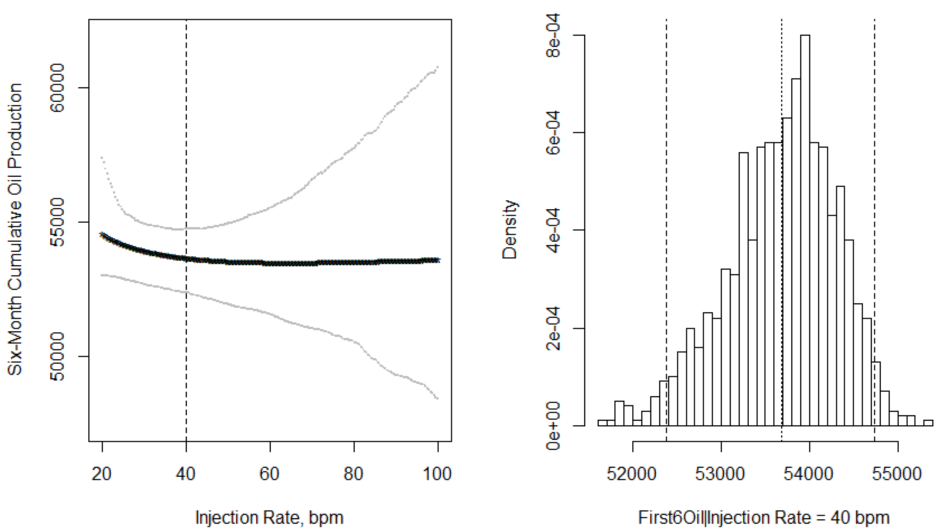


Figure 21: Conditional Distribution of oil Production for IR = 40 bpm showing 95% Confidence Interval

5.0 Conclusion

Using exploratory data analysis:

1. An increasing trend is observed for number of stages, total pounds of proppants and total fluids injected between 2008 and 2016 in the North Dakota portion of the Bakken
2. This was attributed to the fact that the downturn in 2015 forced operators to drill fewer wells. Hence, they utilized more fraturing materials in order to improve productivity. However, this did not guarantee increased oil production as the results were mixed as in previous years
3. On the other hand, the use of more stages, proppants, fluids and drilling of longer laterals would naturally result in higher cost for the operator. Hence, there is the need to evaluate the effect of these parameters on oil production.

To evaluate the effects of completion parameters on oil production:

- An "ensemble" of regression models was built using artificial neural network
- A 75/25 splitting rule was used for the training/testing set for each model
- Sensitivity studies was carried out on the completion parameters
- It was observed that oil production decreased with increasing number of stages, perforated interval, maximum injection pressure, and maximum injection rate
- On the other hand, oil production increased with total pounds of proppants and total volume of fluid pumped.

One possible confounding variable that was observed in this study is drainage or depletion, which depends on time. Oil production dropped with increasing number of stages and perforated interval. In reality, longer laterals and more stages are expected to result in higher recoveries because larger area of the reservoir is exposed to the well through stimulation. Due to possible drainage issue, earlier wells appear to perform better, forcing operators to adopt aggressive approach in completion designs in order to recover more oil. This will be the focus of future study, to ascertain how to account for drainage and incorporate it into the models.

Acknowledgement

The authors wish to acknowledge Bob L. Herd Department of Petroleum, College of Engineering, the Rawls College of Business Area of ISQS and Texas Tech University, for providing financial support for this research. We also appreciate the North Dakota Industrial Commission's Oil and Gas Division and DrillingInfo for providing datasets used for this study. Use of general public license for R programming language, RStudio IDE for R, and contributed libraries are also acknowledged.

References

Ahmed, T. (2010). *Reservoir engineering handbook*. Gulf Professional Publishing.

Alalli, A., Li, Q., Jew, A., Kohli, A., Bargar, J., Zoback, M., & Kovscek, A. (2018). Effects of Hydraulic Fracturing Fluid Chemistry on Shale Matrix Permeability. *SPE/AAPG/SEG Unconventional Resources Technology Conference*. Houston, Texas, USA: Unconventional Resources Technology Conference. https://doi.org/10.15530/URTEC-2018-2881314

Bhattacharya, S., Maucec, M., Yarus, J. M., Fulton, D. D., Orth, J. M., & Singh, A. P.

(2013). Causal Analysis and Data Mining of Well Stimulation Data Using Classification and Regression Tree with Enhancements. *SPE Annual Technical Conference and Exhibition*. New Orleans, Louisiana, USA: Society of Petroleum Engineers. https://doi.org/10.2118/166472-MS

Burton, W. A. (2013). Unconventional Completions: Which One is Right for Your Application? *SPE Annual Technical Conference and Exhibition*. New Orleans, Louisiana, USA: Society of Petroleum Engineers. https://doi.org/10.2118/166431-MS

Ciaburro, G., & Venkateswaran, B. (2017). *Neural Networks with R*. Birmingham, UK: Packt Publishing Ltd.

Cunningham, C. F., Cooley, L., Wozniak, G., & Pancake, J. (2012). Using Multiple Linear Regression To Model EURs of Horizontal Marcellus Wells. *SPE Eastern Regional Meeting*. Lexington, Kentucky, USA: Society of Petroleum Engineers. https://doi.org/10.2118/161343-MS

Drillinginfo. (2018). Oil and Gas Data Source.

Economides, M. J., & Nolte, K. G. (2000). *Reservoir Stimulation* (3rd ed.). John Wiley & Sons Inc.

EMC Education Services. (2015). *Data Science and Big Data Analytics*. Indianapolis, IN, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119183686

Ettehadtavakkol, A., & Jamali, A. (2019). A data analytic workflow to forecast produced water from Marcellus shale. *Journal of Natural Gas Science and Engineering*, *61*, 293–302. Retrieved from https://doi.org/10.1016/j.jngse.2018.11.021

Grolemund, G., & Wickham, H. (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software. Retrieved from http://www.jstatsoft.org/v40/i03/

Guenther, F. (2016). neuralnet: Training of Neural Networks. Retrieved from https://cran.r-project.org/package=neuralnet

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. T. A.-T. T.-. (2000). *Understanding Robust and Exploratory Data Analysis* (1st ed.). New York: Wiley. Retrieved from https://www.wiley.com/en-us/Understanding+Robust+and+Exploratory+Data+Analysis-p-9780471384915

Holdaway, K. (2014). *Harness Oil and Gas Big Data with Analytics: Optimize Exploration and Production with Data-Driven Models* (1st ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Ilk, D., Rushing, J. A., Perego, A. D., & Blasingame, T. A. (2008). Exponential vs. hyperbolic decline in tight gas sands: understanding the origin and implications for reserve estimates using Arps' decline curves. In *SPE annual technical conference and exhibition*. Society of Petroleum Engineers.

King, G. E. (2012). Hydraulic Fracturing 101: What Every Representative, Environmentalist, Regulator, Reporter, Investor, University Researcher, Neighbor and Engineer Should Know About Estimating Frac Risk and Improving Frac Performance in Unconventional Gas and Oil Wells. In *SPE Hydraulic Fracturing Technology Conference* (p. 80). The Woodlands, Texas, USA: Society of Petroleum Engineers. https://doi.org/10.2118/152596-MS

Kolawole, O., Federer-Kovács, G., & Szabó, I. (2018). Formation Susceptibility to Wellbore Instability and Sand Production in the Pannonian Basin, Hungary. *52nd U.S. Rock Mechanics/Geomechanics Symposium*. Seattle, Washington: American Rock Mechanics Association. Retrieved from https://doi.org/

Kolawole, O., & Ispas, I. (2019). How Hydraulic Fractures Interact with Natural Fractures: A Review and New Observations. In *ARMA Symposium*. New York City, New York: American Rock Mechanics Association.

Lolon, E., Hamidieh, K., Weijers, L., Mayerhofer, M., Melcher, H., & Oduba, O. (2016). Evaluating the Relationship Between Well Parameters and Production Using Multivariate Statistical Models: A Middle Bakken and Three Forks Case History. *SPE Hydraulic Fracturing Technology Conference*. The Woodlands, Texas, USA: Society of Petroleum Engineers. https://doi.org/10.2118/179171-MS

Martinez King, V., & Wray, L. (2014). Completion Optimization Utilizing Multivariate Analysis in the Bakken and Three Forks Formations. *SPE Western North American and Rocky Mountain Joint Meeting*. Denver, Colorado: Society of Petroleum Engineers. https://doi.org/10.2118/169534-MS

NDIC O&G Commission. (2018). Bakken Data Source. Retrieved January 8, 2018, from https://www.dmr.nd.gov/oilgas/

Nolte, K. G., & Smith, M. B. (1981). Interpretation of Fracturing Pressures. *Journal of Petroleum Technology*, *33*(09), 1767–1775. https://doi.org/10.2118/8297-PA

Nordeng, S. H., Helms, L. D., & Resources, N. D. D. of M. (2010). *Bakken Source System–Three Forks Formation Assessment*. Retrieved from https://www.dmr.nd.gov/ndgs/bakken/bakkenthree.asp

Olden, J. D., & Jackson, D. A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, *154*(1–2), 135–150. https://doi.org/10.1016/S0304-3800(02)00064-9

Pirayesh, E., Soliman, M. Y., & Rafiee, M. (2013). Make Decision on the Fly: A New Method to Interpret Pressure-Time Data During Fracturing - Application to Frac Pack. *SPE Annual Technical Conference and Exhibition*. New Orleans, Louisiana, USA: Society of Petroleum Engineers. https://doi.org/10.2118/166132-MS

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

RStudio Team. (2016). RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

SAS Institute Inc., S. (2018). JMP, Version 14 - Predictive and Specialized Modeling. Cary, NC: SAS Institute Inc.

Scanlan, W. P., Pierskalla, K. J., Sobernheim, D. W., Christian, J., Dutta, R., Sharma, A.,

… Boykin, B. (2018). Optimization of Bakken Well Completions in a Multivariate World. *SPE Hydraulic Fracturing Technology Conference and Exhibition*. The Woodlands, Texas, USA: Society of Petroleum Engineers. https://doi.org/10.2118/189868-MS

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Soliman, M. Y., Wigwe, M., Alzahabi, A., Pirayesh, E., & Stegent, N. (2014). Analysis of fracturing Pressure data in heterogeneous shale formations. *Hydraulic Fracturing Journal*, *1*(2), 8–12.

Wang, S., & Chen, S. (2016). A Comprehensive Evaluation of Well Completion and Production Performance in Bakken Shale Using Data-Driven Approaches. *SPE Asia Pacific Hydraulic Fracturing Conference*. Beijing, China: Society of Petroleum Engineers. https://doi.org/10.2118/181803-MS

Wei, T., & Simko, V. (2017). R package “corrplot”: Visualization of a Correlation Matrix (Version 0.84).

Wigwe, M.; Kolawole, O.; Watson, M.; Ispas, I.; Li, W. (2019). Influence of Fracture Treatment Parameters on Hydraulic Fracturing Optimization in Unconventional Formations. Beijing China: American Rock Mechanics Association.

Wigwe, M. E., Soliman, M. Y., & Pirayesh, E. (2018). Actual Hydraulic Fracture Length Determination Using a New Technique for Shale Fracturing Data Analysis in Real Time. In *52nd U.S. Rock Mechanics/Geomechanics Symposium* (p. 5). American Rock Mechanics Association. Retrieved from https://doi.org/

Wigwe, M. E., Watson, M. C., Giussani, A., Nasir, E., & Dambani, S. (2019). Application of Geographically Weighted Regression to Model the Effect of Completion Parameters on Oil Production – Case Study on Unconventional Wells. In *SPE Nigeria Annual International Conference and Exhibition*. Lagos, Nigeria: Society of Petroleum Engineers. https://doi.org/10.2118/198847-MS

Zhou, Q., Dilmore, R., Kleit, A., & Wang, J. Y. (2014). Evaluating gas production performances in marcellus using data mining technologies. *Journal of Natural Gas Science and Engineering*, *20*, 109–120. Retrieved from http://dx.doi.org/10.1016/j.jngse.2014.06.014