Fall 2019

Insights into well performance and completion optimization using a coupled FracFocus and DrillingInfo database

Mustafa A. Al-Alwani

INSIGHTS INTO WELL PERFORMANCE AND COMPLETION OPTIMIZATION

USING A COUPLED FRACFOCUS AND DRILLINGINFO DATABASE

By

MUSTAFA ADIL AL-ALWANI

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

In

PETROLEUM ENGINEERING

2019

Approved by:

Dr. Shari Dunn-Norman, Advisor
Dr. Ralph Flori
Dr. Mingzhen Wei
Dr. Abdulmohsin H. Imqam
Dr. Fateme Razaei

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following eight articles, formatted in the style used by the Missouri University of Science and Technology:

Paper I: Pages 9-47 have been published at SPE Annual Technical Conference and Exhibition (ATCE).

Paper II: Pages 48-82 have been published American Rock Mechanics Association (ARMA).

Paper III: Pages 83-124 have been published American Rock Mechanics Association (ARMA).

Paper IV: Pages 125-164 have been been published at SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition

Paper V: Pages 165-196 have been published at SPE Liquids-Rich Basins Conference–North America

Paper VI: Pages 197-226 have been published at Asia Pacific Unconventional Resources Technology Conference

Paper VII: Pages 227-251 have been published at Asia Pacific Unconventional Resources Technology Conference

Paper VIII: Pages 252-279 have been published at SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition

ABSTRACT

The hydraulic fracturing designs, drilling, and completion trends are undergoing continuous change as the oil and gas industry pushing the boundaries of the traditional designs to increase wells productivity and meet the energy demand. The lateral length in addition to the amount of proppant and water used in well stimulation have witnessed a paradigm shift over the past few years. The motivation for this work began with a desire to better understand well performance as a function of the type of fracturing fluid treatment applied. An early objective was to compare whether East Texas Cotton Valley wells stimulated with water fracs had better long-term performance than similar wells stimulated with gelled fluids.

This initial work revealed challenges in acquiring sufficient data for the research, which led to the objective of building a new database from publically available information and then using that information to address the research questions. Developing this database required novel methods of determining at least two key parameters not readily available. One parameter was fluid and proppant volumes, as the FracFocus database reports mass percentages. The second parameter was perforated lateral length as a proxy for fracturing stage data, which was not available for this research.

This work resulted in a methodology for combining FracFocus fluid data with DrillingInfo. A dashboard was also developed to facilitate easy analysis with the newly created database. Once the database and dashboard were available, secondary objectives were to identify completion trends in specific shale play areas, such as the Marcellus and Permian Basin, and to apply data analytics to specific shale play areas.

ACKNOWLEDGMENTS

First, I am forever indebted to my academic advisor and mentor, Dr. Shari Dunn-Norman. Since my early days at Missouri S&T as a Master's degree Fulbright scholar, Dr. Dunn-Norman believed in me and gave me an endless opportunities to succeed. I am very thankful for the generous funding and all the support I got under her mentorship throughout the course of the PhD degree. I will always look up to her as a natural leader, very intelligent professional, a brave spirit, a heart of gold and everything I hope to be one day.

I am also very grateful to Professor Larry Britt for supervising this work. He has been a tremendous help for me. His guidance and knowledge were an essential drive in this research.

I wish to thank the committee members, Dr. Flori, Dr. Wei, Dr. Imqam, and Dr. Razaei. Their consistent support and constructive criticism were highly appreciated.

I would also like to acknowledge my friends and research group members Dr. Abo Taleb Al-Hameedi, Husam Al-Kinani, and Dr. Miguel Cedeno. They have always provided me with great support and continuous encouragement.

Last but not least, I would like to express my deepest appreciation and gratitude to my family and friends in Iraq and in the United States. Special thanks goes to Carl Haberl and my family in Pennsylvania, their love and infinite support are beyond words to be described. The Diaz and the Branson families have been my true family in Rolla and I am very proud of their warm love and endless support. I am very thankful to Gerry Funkhouser and all the people that believed in me and encouraged me during my difficult times and I dedicate all my success to them.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

PAPER IV

SECTION

1. INTRODUCTION

1.1. DEVELOPMENT OF OIL AND NATURAL GAS FROM SHALE

Geologists have known for years that substantial deposits of oil and natural gas are trapped in deep shale formations. These shale reservoirs were created tens of millions of years ago. In the U.S. and around the world today modern horizontal drilling techniques and hydraulic fracturing have helped to access and produce oil and natural gas in these shale reservoirs Figure 1.1 shows different shale plays distributed on the lower 48 United States map.

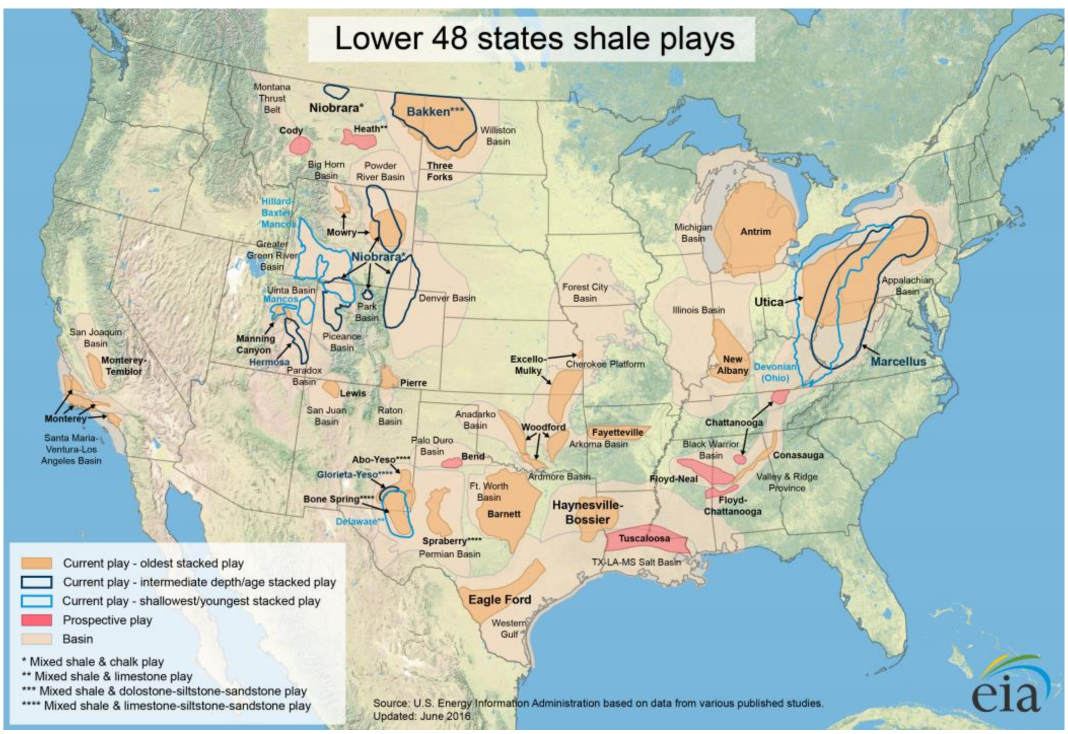


Figure 1.1 Shale Plays in the United States (E.I.A., 2019)

1.2. UNCONVENTIONAL RESOURCES OIL AND GAS

Conventional and unconventional resources are different in their reservoir characteristics. Traditional method used to evaluate conventional reservoirs are not necessarily applicable in analyzing unconventional reservoirs. Therefore, optimizing well completions and production fromunconventional shale resources have presented many challenges to the oil and gas industry.

Unconventional resources, such as tight/shale gas, tight oil, and oil sand are important to the future of the oil and gas industry as they play significant role in the world's energy supply. The world's growing economies require sustained use of oil and gas for the foreseeable future as the need for supply of energy increases. As renewables supply grows, unconventional resources role in the future energy will remain significant, particularly as the unconventional resources are an order of magnitude larger than the conventional deposits that were the primary targets of exploration and production (Leimkuhler, 2012).

Unconventional deposits can be grouped into three general categories:

1. Unconventional reservoirs, which comprise source rocks and ultra-tight sandstone and carbonates.
2. Unconventional fluids, including heavy oil, bitumen, and sour/acid gases.
3. Hydrocarbons that are locked in rocks, such as methane hydrates and shale oil that is trapped in an immature source rock.

Based on the latested reported data by the U.S. Energy Information Administration (eia), As of December 2018, the U.S. tight plays and shale have produced about 7 million BBL/Day and 65 billion cubic feet/Day, respectively.

The U.S. record of natural gas production is credited to the shale gas production that became readily available after the introduction of hydraulic fracturing techniques to the unconventional resources.

Figure 1.2 shows the average daily crude oil production (Million BBL/D) and the average daily production of natural gas (Bcf/d) between 2004 and 2018. It illustrates that the United States has increased its daily production over the years and the tight oil has contributed 60% of the total production as well as shale gas has provided 70% of the total U.S. production by the end of 2018.

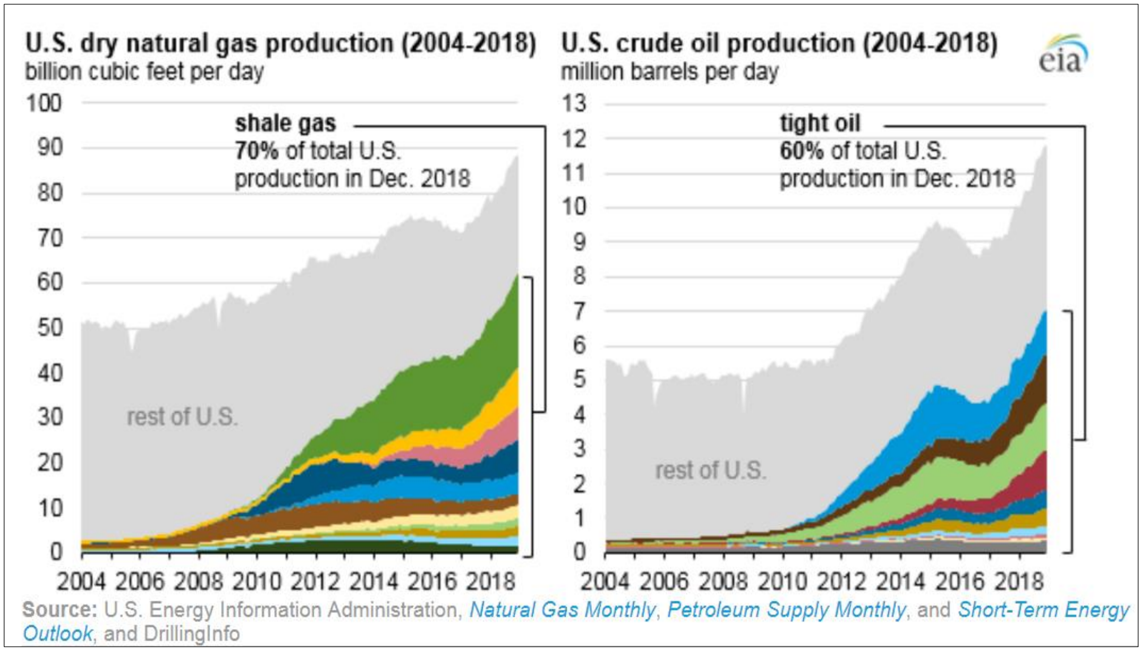


Figure 1.2. U.S. Oil and Natural Gas Production 2004-2018 (EIA, 2019)

1.3. HYDRAULIC FRACTURING

Hydraulic Fracturing (H.F.) is a process which consists of pumping a mixture of mostly water and sand, plus a few chemicals, under controlled conditions into deep underground reservoir formations. The chemicals are generally used for lubrication, to keep bacteria from forming, and to provide enough viscosity to help carry the sand. These chemicals typically range in concentrations from 0.1 to 0.5 percent by volume and help to improve the performance of the stimulation process. Figure 1.3 represents an example of hydraulic fracturing fluid components, 99.5% of the fluid mixture is mainly water and sand with 0.5% other chemicals.

A majority of chemicals used in hydraulic fracturing are now disclosed by the fracturing company in a registry referred to as FracFocus, maintained by the Groundwater Protection Council. The website can be accessed on www.fracfocus.org. This site lists the chemicals and the amounts used on a particular hydraulic fracturing job. The information provided lists the components, ingredient, purpose, and the amount as a percent composition by mass from the total hydraulic fracturing fluid components. In water based hydraulic fracturing fluids, water is the main ingredient, sand is the second largest ingredient followed by small quantities of acid, biocides, scale inhibitors. Friction reducers, and gelling agents. Each one of the chemical additives has a specific purpose, e.g. acid removes near wellbore damage, cleans up perforations prior to injecting fracturing fluids; biocides control bacterial growth in hydraulic fracturing fluids to prevent clogging and reduced performance of the stimulation process; corrosion inhibitor prevents corrosion of tanks and pipes; friction reducers decrease the pumping friction in pipes and inside the created fractures; and gelling agent improves proppant placement and provide higher viscosity to carry proppant particles further into the fractures.

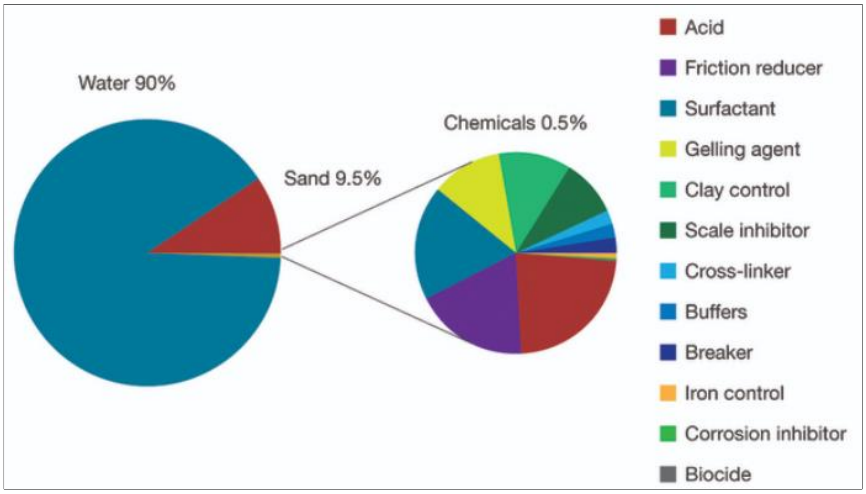


Figure 1.3. Example of a Typical Fracturing Fluid Composition (wa.gov.au, 2018)

1.4. DATA MINING AND USING BIG DATA APPROACHES

Big data analytics are emerging as a useful tool to optimize well completion parameters such as the type of fracturing fluid or proppant used. The use of data analytics is being facilitiated through the accumulation and online availability of public data like FracFocus. The ability to integrate chemical data with the completion and production data in a single data baseenables each well to serve as a real-scale geophysical experiment to come up with a scientific and data driven insights and models that help to address many of the questions and concerns expressed by operators, the public, environmentalists, regulatory representatives, and non-profit organizations related to the hydraulic fracturing stimulation.

The application of data science in the oil and gas industry is an emerging technology. Some major companies have established new divisions within the company to deal with their accumulated data. Data science and analytics help the company's management and engineers to gain better insights about the performance of their processes with an ultimate goal to optimize the process and increase productivity. In the upstream part of the oil and gas industry, data scientists are faced with enormous amount of data including the production, completion, hydraulic fracturing treatment components, and other treatment parameters. In most cases, the database would include tens of millions of raw data inputs. The role of the data scientist or as it referred in some companies as to a hybrid engineer, is to de-silo the dataset and clean the data from outliers, inaccuracy, and solve formatting issues. Once the data is clean it will be ready for data analysis and visualizations.

Outcomes from big data analysis offer the operational engineers and managements a fuller picture of their managed assets and enables them to drill down through the data and gain specialized analyses and visualizations for different levels of interests. The use of this evolving technology enabled the engineers to shift from the old approaches of being stuck to offline spreadsheets to fully dynamic dashboards that enable them to display different outputs in the same time while being able to interact with the data, run simulations, and test several ideas with simple mouse clicks.

2. RESEARCH MOTIVATIONS AND OBJECTIVES

2.1. RESEARCH MOTIVATION

The motivation for this work began with a desire to better understand well performance as a function of the type of fracturing fluid treatment applied. An early objective was to compare whether East Texas Cotton Valley wells stimulated with water fracs had better long-term performance than similar wells stimulated with gelled fluids. This initial work revealed challenges in acquiring sufficient data for the research, which led to the objective of building a new database from publically available information and then using that information to address the research questions. Developing this database required novel methods of determining at least two key parameters not readily available. One parameter was fluid volume and proppant mass, as the FracFocus database reports mass percentages. The second parameter was perforated lateral length as a proxy for fracturing stage data, which was not available for this research.

This work resulted in a methodology for combining FracFocus fluid data with DrillingInfo. A dashboard was also developed to facilitate easy analysis with the newly created database. Once the database and dashboard were available, a primary motive was to use this database to identify completion trends in specific shale play areas, such as the Marcellus and Permian Basin, and to apply data analytics to specific shale play areas.

Another movtive of the work was to use the database to investigate operational trends. For example, well pad density as of January 1, 2015 and again as of January 1, 2019 were plotted for U.S. shale plays and provided to the Health Effects Institute (HEI)*. This work will be published separately from the publications in this dissertation.

2.2. RESEARCH OBJECTIVE

Objectives of this research include

- Constructing a comprehensive database that combines FracFocus chemical chemical data and DrillingInfo production and completion data for several major U.S. Shale Plays.
- Develop data gathering and processing methodologies to serve as a guideline to update the created database.
- Develop classification system to identify the hydraulic fracturing treatment types based on the type and percentages of the chemicals used.
- Develop interactive descriptive models in the form of modern dashboards that are user-friendly powered by TIBCO SpotFire.
- Build statistical models to compare and predict the performance of wells based on type of the treatments, completion, and stimulation parameters.

---

*HEI, 2019 "Potential Human Exposures Associated with Unconvetional Oil and Gas Development: Literature Survey and Research Planning", Health Effects Institute – Energy, Boston, MA. Report expected September 2019.

PAPER

I. LONG-TERM COMPARISON OF GEL AND WATER FRACTURE STIMULATION IN MARCELLUS SHALE PLAY

Mustafa A. Al-Alwani, Larry K Britt, Shari Dunn-Norman, Husam H. Alkinani, Abo Taleb T. Al-Hameedi, Atheer M. Al-Attar

Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA

ABSTRACT

The goal of any hydraulic fracturing stimulation is to design and execute the appropriate treatment that is best suited for the stimulated reservoir. Selecting the best treatment must achieve the desired fracture geometry to maximize long-term well productivity and reserve recovery. The main objective of this study is to conduct detailed short and long-term production and well-to-well comparisons of the different types of fracture stimulation fluids in the Marcellus Shale play.

A database of more than 4,000 horizontal, stimulated Marcellus wells was constructed for this study. The wells were divided into four groups according to the type of treating fluid: water, gel, cross-linked, and hybrid fracs. Chemical data from FracFocus were gathered and processed then combined with completion and production data to investigate the gas short and long-term production. Detailed monthly production data for the studied wells were captured from DrillingInfo database and utilized in this study.

This paper reports and compares the Marcellus gas initial production, the gas cumulative production at the end of the first month, first 6 months, first year, 2 years, and 5 years, according to the type of hydraulic fracturing fluid used in primary stimulation. The work provides insights into Marcellus well performance as a function of stimulation parameters such as the volume of stimulation fluid and the amount of pumped proppants. The impact of perforated lateral length is taken into consideration and used to normalize production and stimulation parameters. The study shows that water fracturing fluids outperformed the other types of hydraulic fracturing fluids.

This paper introduces several data processing workflows that serve as a reference for individuals who are interested in extracting and processing data from the FracFocus database. It also documents the occurrence in hydraulic fracturing fluid types and measures the effects of the fracturing fluid volume and total proppant pumped on the initial and cumulative production.

1. INTRODUCTION

Advancements in horizontal drilling technology coupled with the ability to economically place repeated hydraulic fracturing stimulations along the horizontal well have driven the growth in developing unconventional resources as a leading source of oil and gas to the energy supply in the United States. The resulting boom in drilling and stimulation have led to an accumulation of a vast amount of drilling, completion, stimulation, and production data. Gathering and analyzing all of the accumulated data aid in optimizing the productivity of the wells and identify the greatest economic exploitation methods for future well drilling and completion.

A large number of industries have been applying data descriptive and predictive approaches in their operations' optimization in addition to decision making and future planning. Data analytics have been defined as one of the highest growing fields within the oil and gas industry (Khvostichenko and Makarychev-Mikhailov, 2018). Many of the operators and service contractors have realized the imperative of data-driven decisions and started to appropriate many of their resources to expand their data utilization in everyday decisions. The literature shows an abundance of articles that depicts the growing interest in data analysis studies. The output of this kind of studies are very insightful and shows a wide range of applicability (Yang et al., 2013; Arthur et al., 2014; Romero and Poston, 2016; Myers et al., 2017; Luo and Zhang, 2018; Olaoye and Zakhour, 2018; Srinivasan et al., 2018; Carman and Wheeler, 2018; Weijers et al., 2019, Al-Alwani et al., 2019).

In hydraulic fracturing, fluids ladened with proppant, are injected under high pressure to create a fracture which provides a pathway for oil and gas to readily flow into a well. There are four stages found in most stimulation treatments. The spearhead pre-fracturing stage or it can also be called as the acid stage involves pumping water with diluted acid to clean and clear the perforations and the near wellbore area from any potential debris. The next stage is the pad stage which involves pumping the mixed water and chemicals fluid under a certain pressure to initiate and propagate the fracture. The following stage is the slurry or the proppant stage which involves the pumping of a mix of hydraulic fracturing fluid and propping agents. The proppant will be transported into the fracture and remain in the fractures to keep them propped open. The final and last stage is the flush stage which involves pumping clean base fluid to displace any proppant from the wellbore.

There are several types of hydraulic fracturing fluids. Water fracs consist of water, friction reducer, and may also include some clay control agents and other chemical ingredients. Gel fracs (linear gel) includes water, gelling agents and other chemical components. Example of gelling agents can be guar, hydroxypropyl guar (HPG) or hydroxyethyl cellulose (HEC). Gelling agents can be prone to bacterial growth; therefore, biocides are usually added to the gel fluid to eliminate bacterial problems. Cross-linked gel consists of water, gelling agents and cross-linking agents. The use of cross-linked gel fracturing fluid requires adding breakers to the mixed fluid to help break the cross-linked gel fluid in the fractures and flow back to the surface once the job is completed. In some of the stimulated wells, a combination of the three different types of hydraulic fracturing fluids can be used. For example, treated water can be pumped in some stages then followed by linear gel in the other stages. Another example can be a linear gel to cross-linked gel. The combination of different fluid types used in the same fracturing job classified as hybrid fluid fracturing.

The Marcellus shale play is located in the Appalachian basin. It is a source rock that belongs to the Middle Devonian. The Marcellus Shale extends across Pennsylvania, West Virginia, New York, Kentucky, and Ohio. The shale characteristics of the Marcellus vary from the Northeastern side to the Southwestern side. Modern applications of hydraulic fracturing in the Marcellus began in 2003 and 2004, when operators in the Southwestern part of Marcellus utilized large slickwater treatments to test the applicability of horizontal well multi-stage fracturing in the Marcellus shale after it was found to be successful in the Barnett shale in Texas (Myers, 2018).

The ultimate objective of this study is to compare both short- and long-term Marcellus stimulated well performance as a function of different hydraulic fracturing fluid treatments. The publicly available FracFocus database was used to identify the type of hydraulic fracturing fluids used in wells. Data obtained from FracFocus was combined with commercially available well production data (DrillingInfo, 2019). The following discussion presents details regarding the data handling and workflows created to uniquely identify hydraulic fracturing treatments, and then associate those treatments with each Marcellus well production history. It also compares Marcellus well performance as a function of fluid volumes, proppant loads, normalized on lateral well length.

2. APPROACH

A core requirement of this study was to establish a methodology to utilize the publicly accessible FracFocus database and establish procedures to extract the data and transform it to usable inputs. For example, FracFocus provides each single chemical treatment applied in a well as a separate row of data. Identifying all stimulation treatments associated with the primary completion in a particular well, requires some aggregation method to be applied to the rows of data. This is but one example of numerous such adjustments required to clean and format FracFocus data. Importantly, an understanding of unconventional well hydraulic fracturing fluids and procedures is really required to process these data and ensure a robust and verified database.

FracFocus database contains chemical data for wells completed after 2011, when the FracFocus 1.0 registry began. However, it is important to note that reporting methods varied over time and many wells completed in the earliest time period may not have complete data. These wells were excluded from study.

The FracFocus data were then merged with the completion and production database of DrillingInfo. Merging the two databases provides the ability to associate hydraulic fracturing fluid treatments in a well with its drilling and completion parameters and production history. The merged database created in this study includes all stimulated wells throughout the United States reported in FracFocus. Only Marcellus wells are included in this work.

The following subsections provide additional detail regarding the technical approach used to process the data. This discussion can serve as a guide for future FracFocus data mining and processing.

2.1. HANDLING RAW FRACFOCUS DATA

More than 4 million rows of chemical ingredients data were downloaded from the FracFocus website. The data are available online to be downloaded in Excel files. Due to the limitations of the number of rows that one Excel sheet can accept (maximum of 1,048,576 rows) the Fracfocus data were downloaded and condensed into 5 Excel files, which were then combined and processed together. Once all the FracFocus data were stacked together, the following step was to identify parameters to be included in the database (e.g. well location, ingredient name, water volume, etc) and ignore the unnecessary inputs (e.g. file upload key). The format of some data columns were adjusted for further use. For example, the date data was adjusted from FracFocus format of mm/dd/yyyy hh:mm to mm/dd/yyyy and well location data (surface longitude and latitude) were used to create centroid points to map the wells' surface location on the map. After all the data merging and formatting, the selected columns were cleaned from any leading or trailing white spaces and then saved in the format of tab-delimited text file to be used in the next phase of the data processing. Figure 1 illustrates the FracFocus initial data processing workflow.

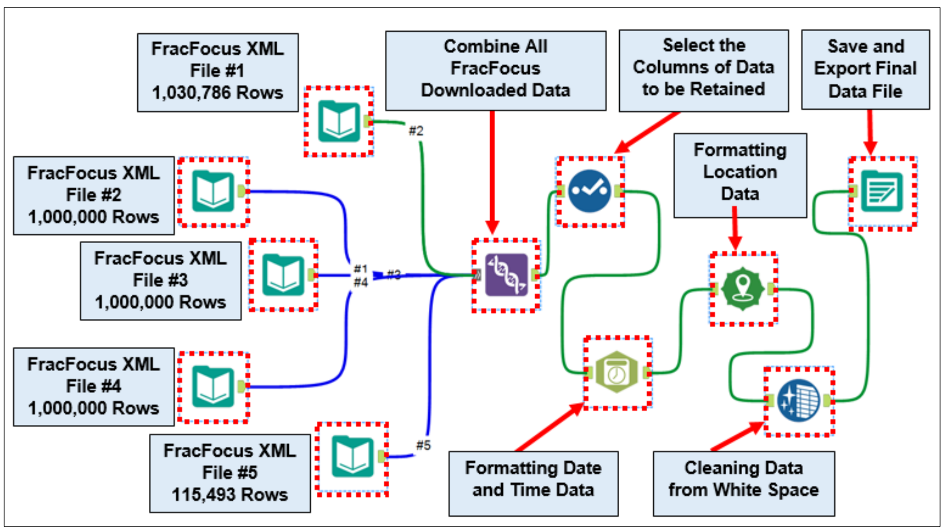


Figure 1. Data Processing Workflow Illustrating the Initial Steps to Process FracFocus Data Files

2.2. GROUPING FRACFOCUS CHEMICAL INGREDIENTS

One of the biggest challenges in using the FracFocus data is that the data are presented in a stacked format. For a single well report, there are several rows of the same chemical groups that represent different mass percentage of a chemical ingredient. This work classifies all the chemicals into 19 groups and then process each group of chemicals separately to aggregate the total mass percentage of each group's chemical used in a particular well. The chemical groups designated are as follows: water, proppant, gelling agents, cross linkers, breakers, friction reducers, surfactants, biocides, clay control, acid, pH adjuster, corrosion inhibitor, iron control, scale inhibitor, non-emulsifiers, solvents, liquid N2, others, and other non-CAS chemicals.

Classifying FracFocus data into the aforementioned groups is not an easy task and requires knowledge of hydraulic fracturing treatments applied in unconventional oil and gas well. It also requires data analysis and grouping methods. In FracFocus, there are no standards in reporting the chemicals' name or function which makes it impractical to try to classify the ingredients based only on the name or the purpose. To properly classify the chemical groups, a combination of the chemical ingredient name along with the chemical abstract service number (CAS-Number) was used. CAS-Number is a unique number assigned for any chemical substance and each CAS-Number linked to a list of information that describes the chemical substance. A treemap chart technique was applied in a hierarchy style to group chemicals. The chemicals that are heavily present in the highest number of fractured wells are listed on top of the hierarchy. The ingredient name will be assigned based on the most frequently mentioned name among all the wells which was associated with a specific CAS-Number. Applying the "most common" function in assigning the ingredient name eliminates the FracFocus pitfalls of naming the chemicals in different styles and spelling errors. The size of the box in the treemap is based on the count of the number of the wells that the CAS-Number existed.

Figure 2. Illustrates a treemap chart with the basic setup steps to create the chart based on the FracFocus database. In this illustration, the CAS-Number 14808-60-7 is the unique number for sand that is used as a proppant in hydraulic fracturing. Sand is on top of the hierarchy followed by water (7732-18-5) because they comprise the greatest mass

percent ingredients in all the stimulated wells. Based on the CAS-Number all the chemicals were grouped into one of the 19 groups. Two sources of data check were used to validate the accurate function of each CAS-Number in order to assign it to the appropriate group. Environmental protection agency (EPA) safer chemical ingredients list and a study conducted by Helmholtz Zentrum München, Institute of groundwater ecology which lists complete list of all H.F. fracturing chemicals extracted from the Waxman list and the FracFocus database were utilized in identifying the prospective chemical groups (Elsner & Hoelzer, 2016; EPA, 2019).

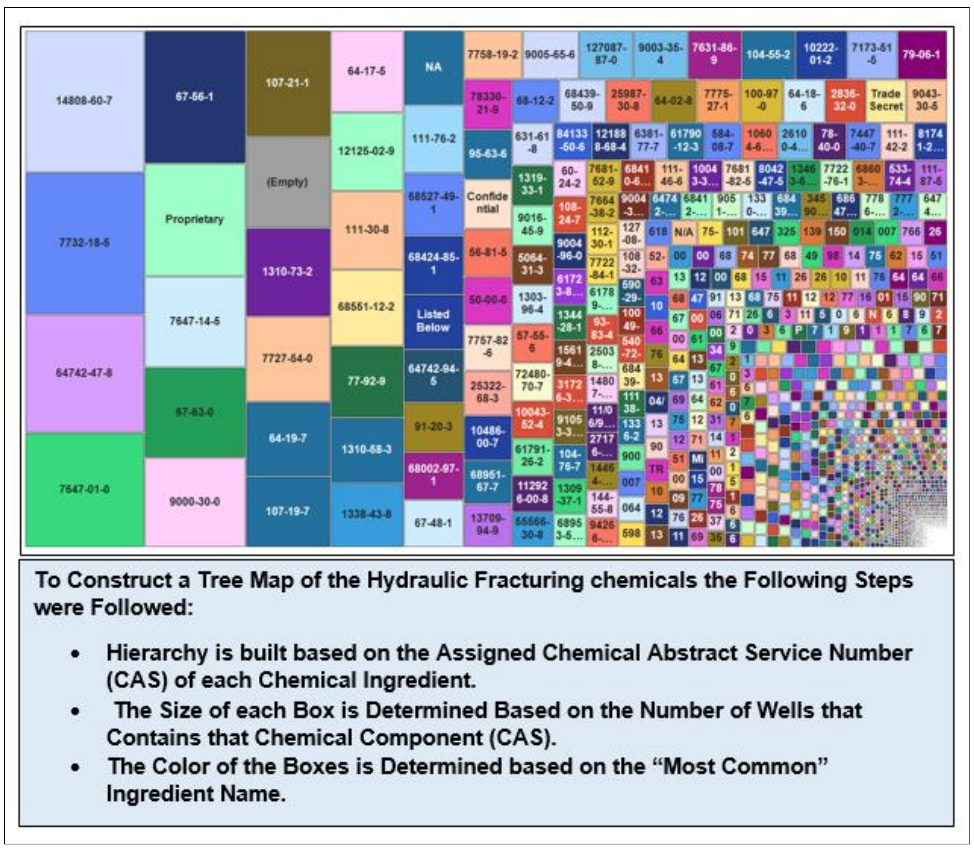


Figure 2. TreeMap Chart Utilized to Group FracFocus Chemical Ingredients

2.3. PROCESSING AND AGGREGATING EACH CHEMICAL GROUP

Once the 19 chemical groups were created, each group was processed separately to inspect for duplications in data entry within FracFocus reports. Once duplicate rows were eliminated, an aggregation function was utilized to add up the chemicals' mass percentage for every well based on the unique well API number. It is worth mentioning that for the same well, the same chemical can be mentioned in several rows with different mass percentages or the same group of chemicals can contain different ingredients' names and CAS-Numbers, yet serve the same purpose (e.g. surfactants). The workflow also inspects for refracturing operations and assigns a fracturing job number for each well. For example, if the same well's API number appears to have two separate fracturing dates and different chemical components associated with each date, this means that the well has been refractured and based on how many different report dates available, the well stimulation treatments will be numbered in ascending order with the primary completion designated as '1' and associated with the first fracturing job start date.

Figure 3 illustrates the data processing workflow to calculate the total mass percentage of chemicals for a stimulation treatment. There are two output files generated in this workflow. The first one will report the summed-up mass percentages of all chemical groups per well and the second one will keep all the original components without adding the mass percent together. The first file output is used in this study to calculate all the actual mass and concentrations of all the classified groups while the second file output can be used to identify the different chemical components and their mass percentages, for example all the different proppant types and sizes.

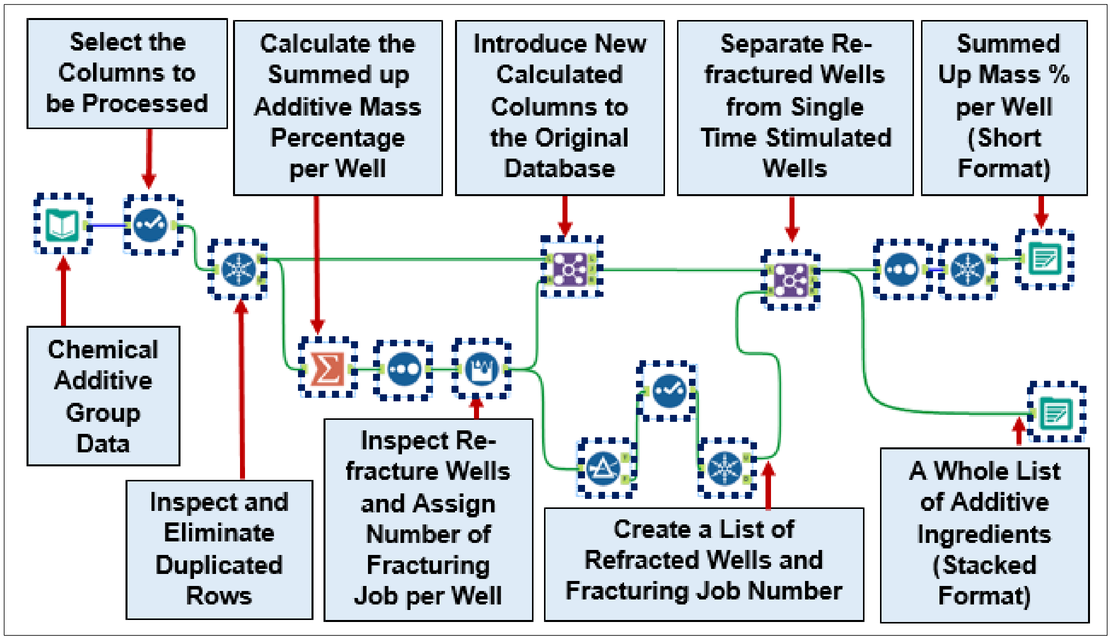


Figure 3. Data Processing Workflow to Aggregate the Total Mass % of Chemical Ingredients per Well

2.4. CREATING PROCESSED FRACFOCUS DATABASE

Figure 4 shows the steps of building the processed FracFocus database by merging and combining the 19 chemical groups altogether. The workflow starts by joining the water and proppant data based on the API number. In this step, the workflow investigates the well's API number and joins the data of the water group with the data of the proppant group. If one well has only proppant and no water data, the data will be held aside into a separate output, the same is true for the wells with only water and no proppant data. After combining the water and the proppant data into a new output, the proppant mass in pounds will be calculated based on the total mass percentage, the water mass percentage, and the total water volume. As a check, the actual reported proppant mass was compared with reported proppant mass. A 98% match indicated the mass percentage aggregating workflows were setup properly. The other remaining chemical groups were introduced into the workflow using union tools which means that only the wells that have a mass percentage of that particular chemical group will obtain a mass percentage value and the remaining wells will be assigned a zero value for the lack of that chemical group into the stimulation fluids recipe. The line break in the workflow diagram represents the entry of all the remaining chemical groups. Once all the chemical groups were introduced and matched together, several new columns were introduced to the database by using the formula options to perform calculations to obtain the mass in pounds for all the chemical groups as well as the total concentration of the chemicals by dividing the mass of the material by the liquid phase base fluid volume. Finally, the total hydraulic fracturing percentage recipe was calculated to compare the accuracy of the process.

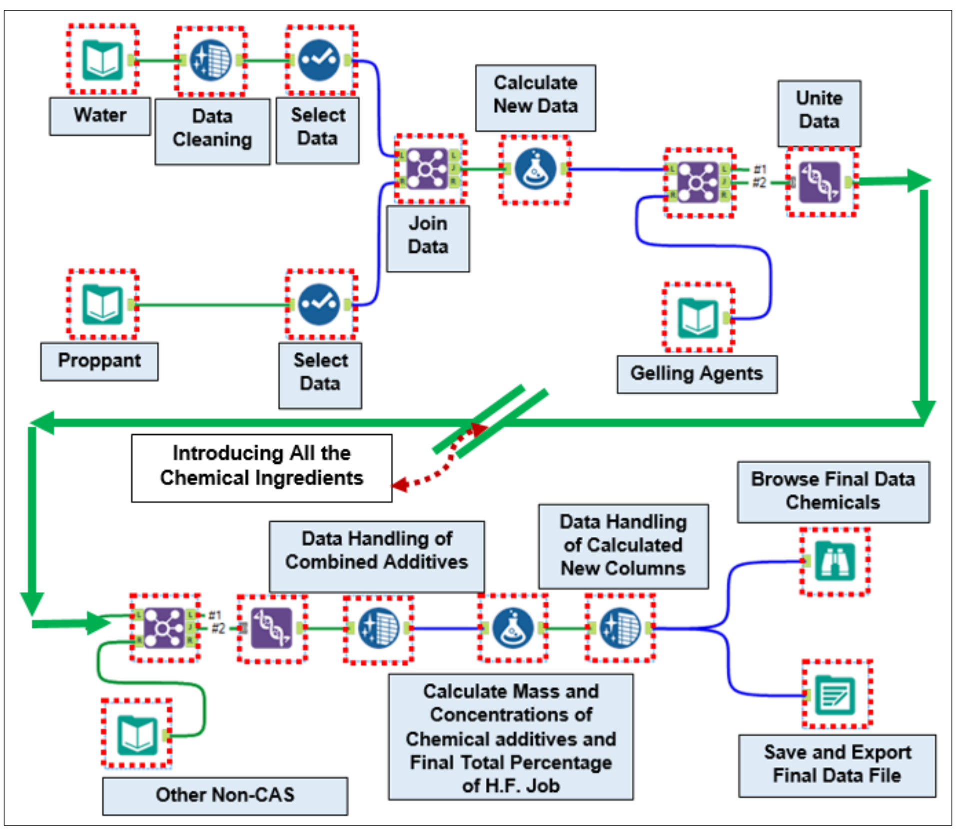


Figure 4. Data Processing Workflow Illustrating the Steps of Building the FracFocus Processed Database

2.5. ACCURACY SCREENING

After calculating the hydraulic fracturing recipe total mass percent, a screening criterion was used to select only wells whose total chemical components mass percentage adds up to the range between 99 % and 101 %. FracFocus on their website disclaimer stated that the total mass percent of the disclosed chemical may be less than 100% because not all the non-MSDS materials are required to be disclosed as per some of the states' policies. The total mass percent may exceed 100 as some of the chemicals' mass percentage might be expressed as its maximum concentration. Figure 5 shows the final selected 82,351 wells with the total mass percent that adds up to 99-101 % (FracFocus, 2019). Of these total US well population, 4000 Marcellus wells were identified for study.

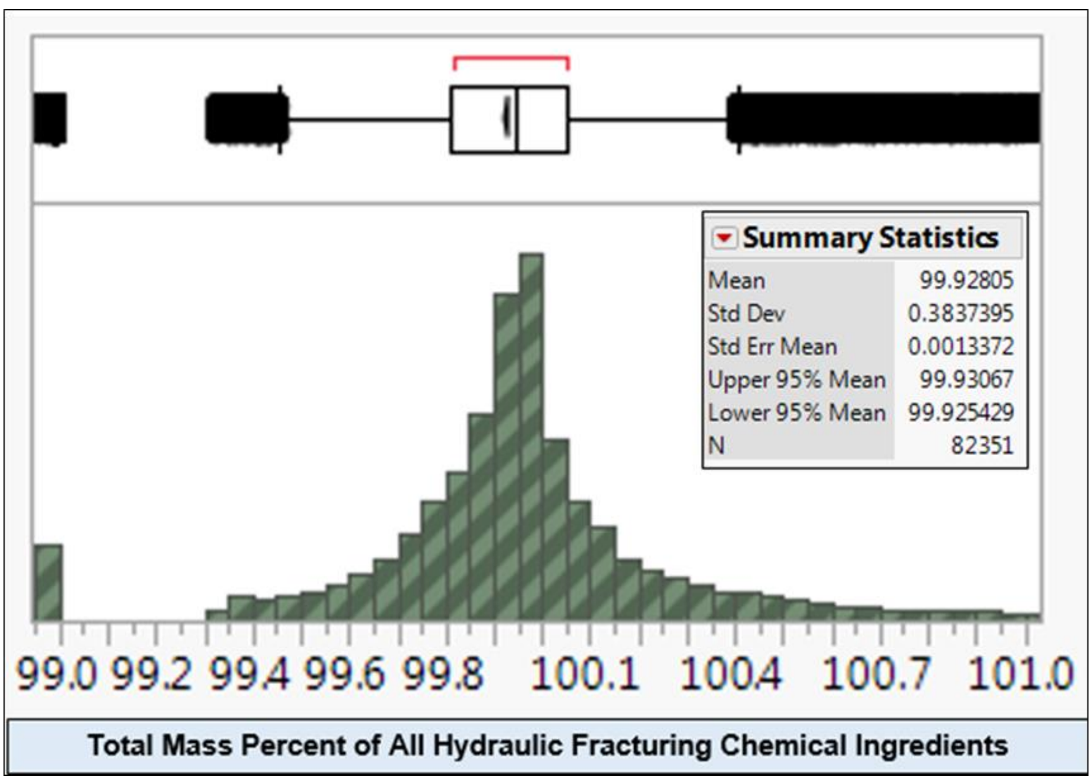


Figure 5. Box Plot and Distribution Bars of the Total Mass Percentage of the Hydraulic Fracturing Total Recipe

2.6. INTEGRATING FRACFOCUS DATA WITH DRILLINGINFO DATA

The developed FracFocus database was merged with the DrillingInfo database to generate a comprehensive database that enables comparisons between the stimulation and completion parameters and their impact on production and well performance. Figure 6 illustrates the data merging workflow where all the screened FracFocus data were matched by the API number with the DrillingInfo database. The wells in this database are populated between the years 2012 and 2018. DrillingInfo data has production and completion data extend prior to the year 2012 but for the sake of this study and to utilize the chemical registry database, only the wells that are completed after the year 2012 were included.

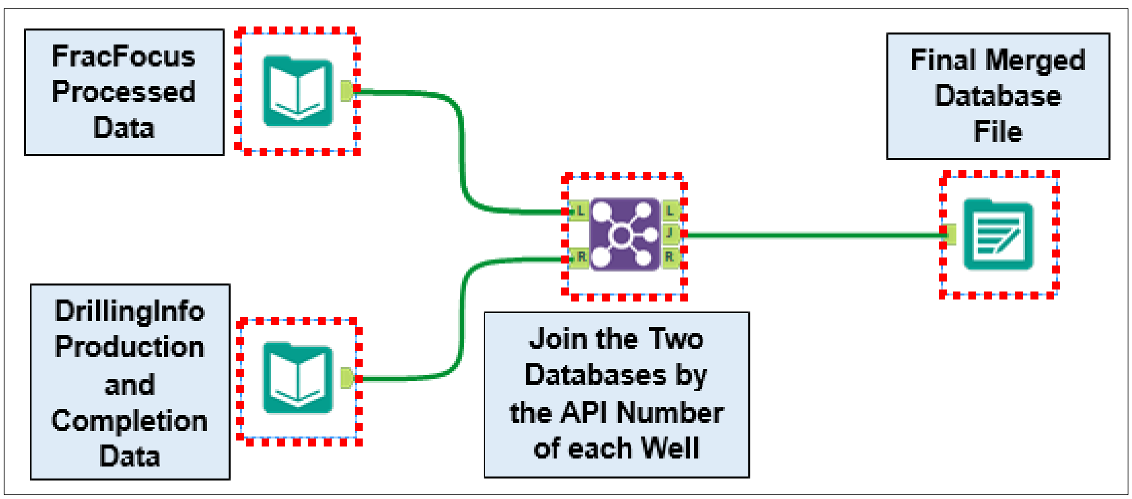


Figure 6. Data Processing Workflow Merging Final FracFocus Processed Chemical Database with DrillingInfo Production and Completion Data

2.7. FINAL DATA CLEANING AND OUTLIERS DETECTION

All the data columns in the final merged FracFocus and DrillingInfo were subjected to outliers' detection and elimination procedure by utilizing the box plots techniques and removing all the suspected outliers to prepare the data for rigorous classifications and comparisons studies (Alkinani et al., 2018). Figure 7 shows an example of the cleaned water and proppant data, respectively.

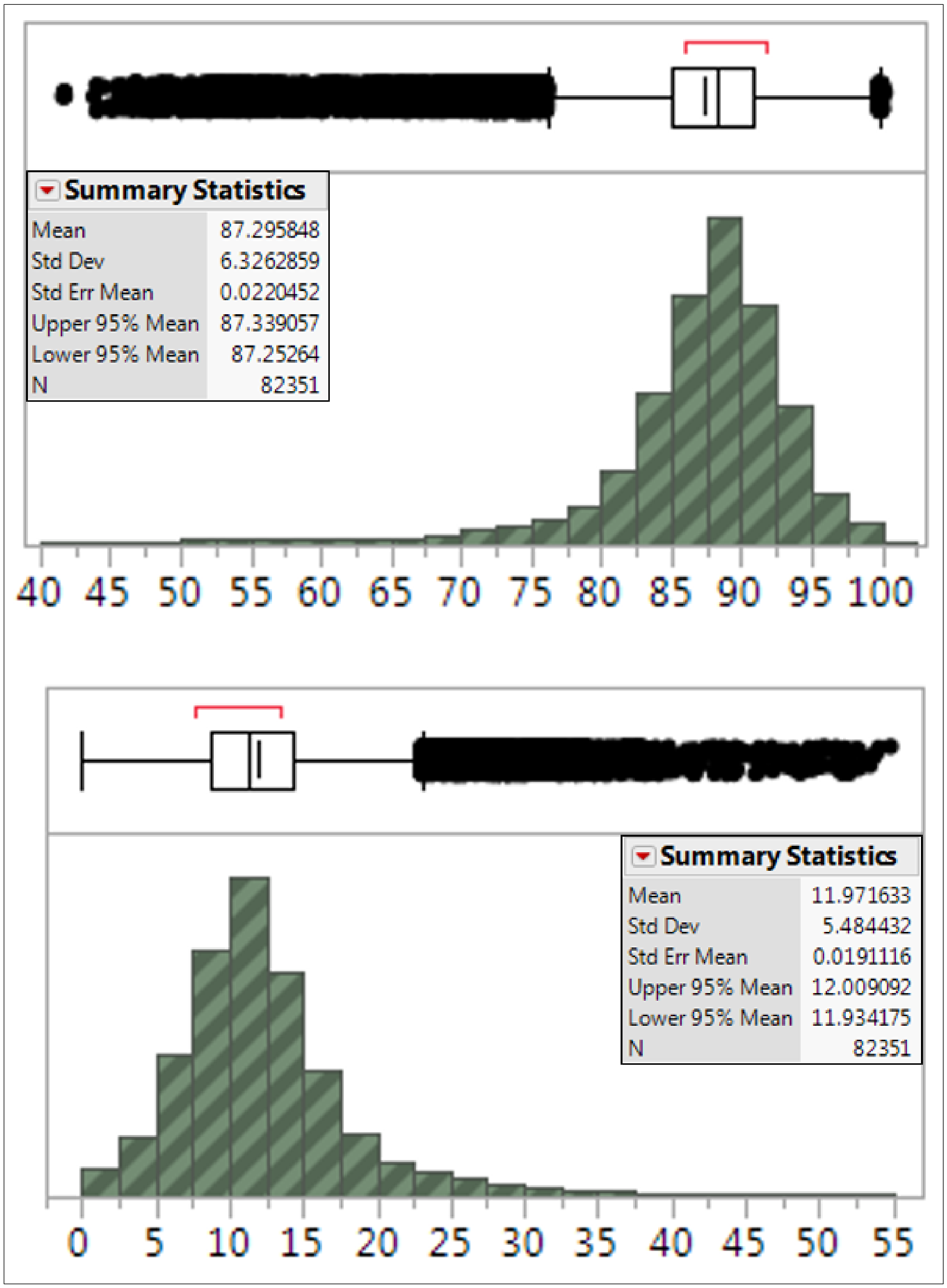


Figure 7. Illustration of JMP Software Utilization to Perform Outliers Detection by Applying the Concepts of Box Plot and Data Distribution Bars

2.8. CLASSIFYING STIMULATION TREATMENT FLUIDS

The hydraulic fracturing fluid types were classified as water treatment, gel, cross-linked or hybrid based on the presence of key chemical ingredients: gelling agents, cross-linking agents, and breakers. A formula was integrated into the workflow to perform the classification based on knowledge of the four fluids types and their application. For example, the presence of a chemical breaker or a cross-linking chemical would indicate the fluid was cross-linked. Figure 8 shows the classification workflow with the built-in conditional statement classification tool.

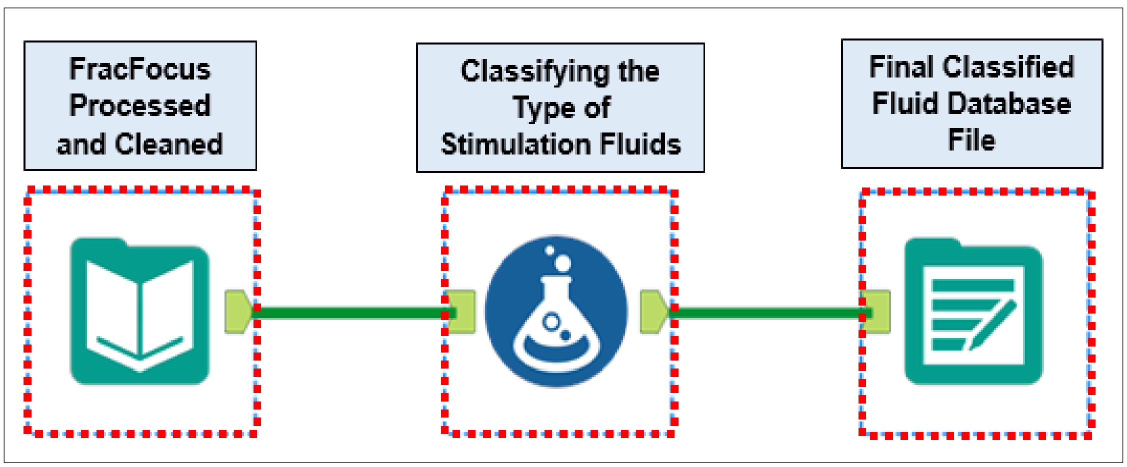


Figure 8. Data Processing Workflow to Assign Stimulation Fluid Treatment Type for Each Well Based on the Chemical Ingredients' Percentage

3. RESULTS AND DISCUSSION

This section discusses the Marcellus wells identified for study, including their treatment fluid types, completion trends, and the initial production and cumulative recovery with time for Marcellus wells stimulated with each type of hydraulic fracturing fluid. Box plots and statistical summaries of production data are also provided.

3.1. MARCELLUS SHALE DATA COVERAGE

More than 4,000 Marcellus wells completed between 2012 and 2018 have been included in this study. The data cover most of the Marcellus Shale play across most of the active counties in Pennsylvania and West Virginia. Figure 9 shows a pie chart with all the major counties and the percentage of the wells that span across them. Figure 10 shows a pie chart with the percentage of the four treatment fluid types. As shown, more than 55% of the Marcellus wells were treated with water fracturing fluids. More than 26% of the wells were treated with hybrid fracturing fluid, which means that the wells have been fractured in some stages by water then gel fluid or gel and crossed-linked or any combination of the treatment fluids. Linear gel was used to treat about 13% of the wells, and crossed linked fluid were utilized to treat less than 5% of the wells. Figure 11 pictures a map of the Marcellus area focusing on the wells in Pennsylvania and West Virginia. The wells are color-coded based on the treatment fluid type, i.e. water wells are colored with blue, gel with red, crossed-linked gel with dark red, and hybrid wells with purple. Each state on the map was colored differently to make the border between the states more visible.

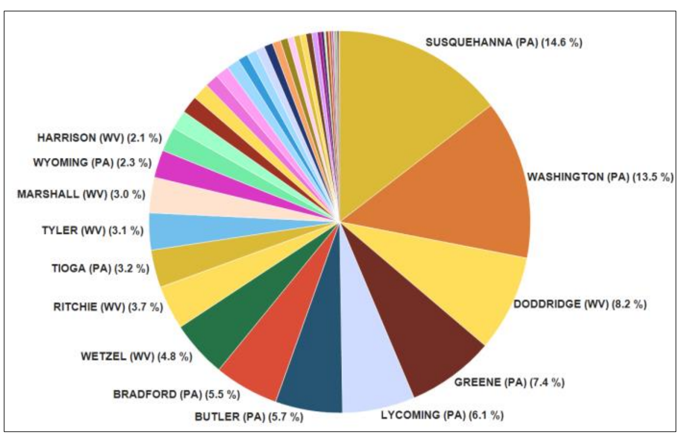


Figure 9. Pie Chart of the Counties and the Wells' Distribution Percentages of the Studied Area

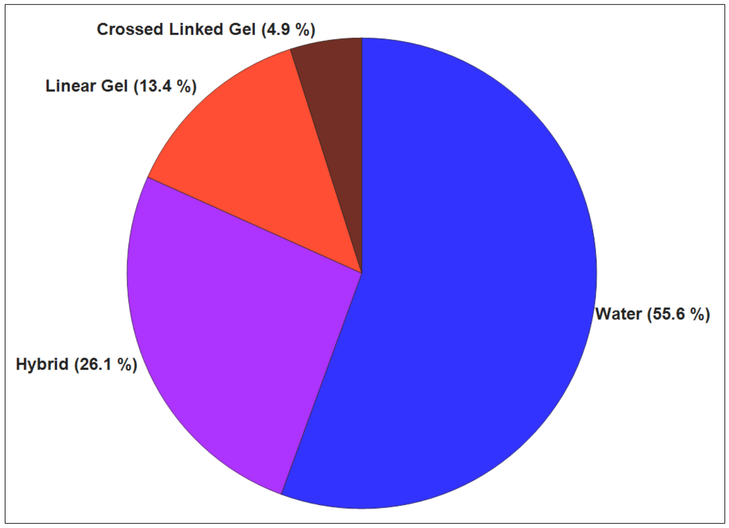


Figure 10. Pie Chart of the Stimulation Fluid Type Percentages of the Studied Area

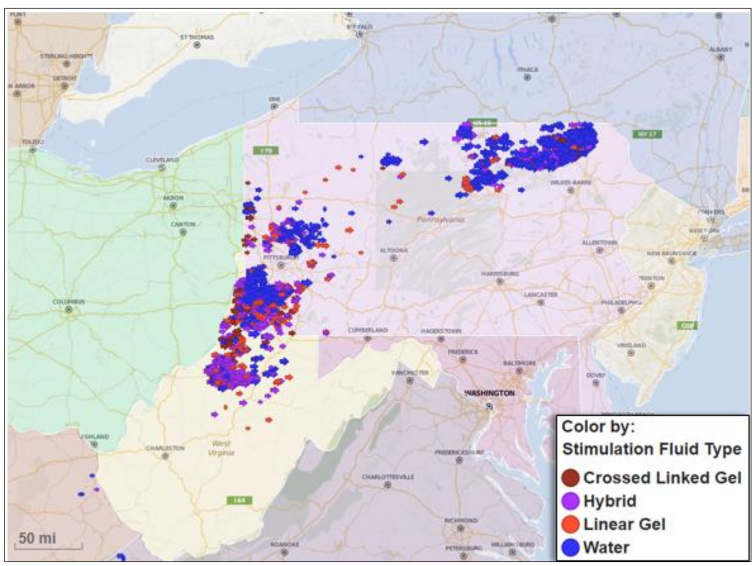


Figure 11. Appalachian Basin Map Showing Stimulated Wells which are Colored Based on Treatment Type

3.2. TREATMENT FLUID TYPES AND WELL PERFORMANCE IN THE MARCELLUS SHALE

The primary focus of this work is to investigate Marcellus shale well production performance as a function of time for the four hydraulic fracturing fluid types. A user-friendly dashboard was developed and linked to the database previously described. The dashboard facilitates creating a wide variety of figures to illustrate well performance as a function of completion parameters.

Figure 12 shows the average total stimulation fluid volume per well for each hydraulic fracturing fluid type. To take into consideration the difference in lateral length between the wells, the stimulation fluid volume was normalized to the perforated lateral length by dividing the total stimulation volume per well by the lateral length as shown in Figure 13 This analysis shows that wells treated with hybrid fluid consumed more stimulation fluid than the other treatment fluid types and the normalized treatment fluid volume per stimulated length showed that cross-linked gel wells consumed more fluid per foot in comparison to other types of fluids.

Figure 14 shows the amount of pumped proppants per well for each fluid type. Wells treated with hybrid fluid type used the highest average proppant pounds per well. Figure 15 shows proppant load based on the normalized per each foot of perforated lateral length. In this analysis, water fractured wells consumed more proppant per foot compared to the other types, whereas cross-linked fluids exhibit the highest proppant load in lb/gallon as shown in Figure 16 Wells treated with water may exhibit higher normalized proppant loads because these wells have relatively shorter average lateral length compared to the wells treated with other fluid types. Figure 17 shows the average perforated lateral length for all the fluid types in this study. The longest laterals were stimulated with hybrid fluid. The shortest average perforated lateral length was stimulated with water fracturing.

Figure 18 shows the progression of lateral length increase per time for all the types of treatment fluids. It also shows that in recent years, water fracs are competing with hybrid fluids to stimulated longer laterals. Figure 19 shows the number of completed wells over time and the corresponding percentage of each fluid type. These results indicate a growing trend in utilizing more water fracs in the Marcellus shale compared to the other stimulation fluid types.

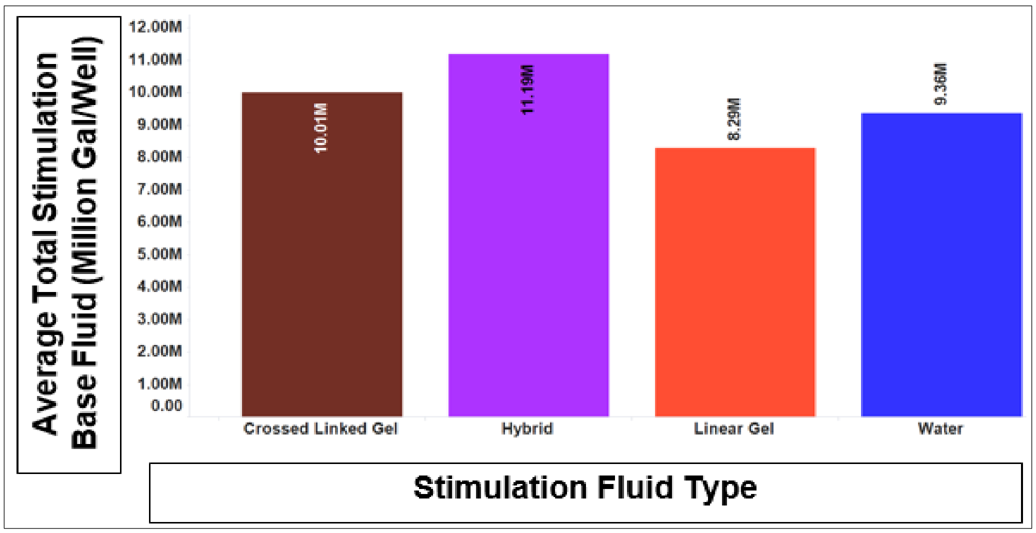


Figure 12. Bar Chart Comparing the Marcellus Average Stimulation Fluid Volume Usage per Well for Different Stimulation Fluid Types

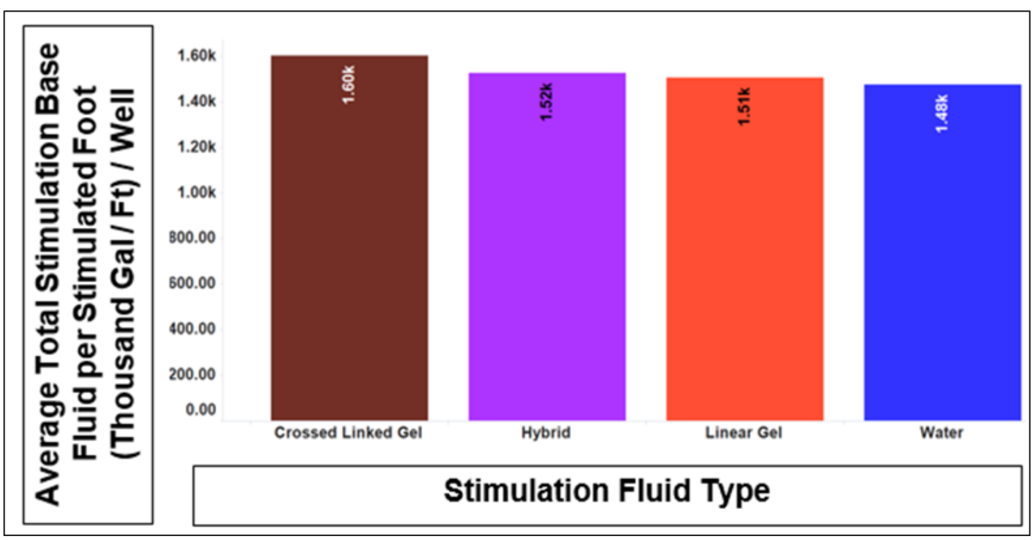


Figure 13. Bar Chart Comparing the Marcellus Average Stimulation Fluid Volume Usage per Stimulated Foot per Well for Different Stimulation Fluid Types

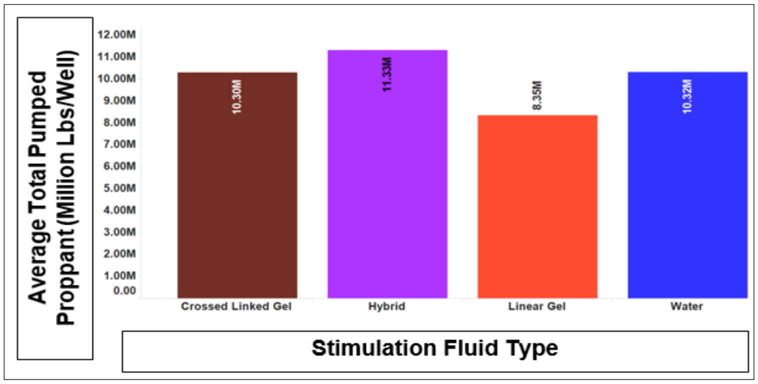


Figure 14. Bar Chart Comparing the Marcellus Average Total Proppant Usage per Well for Different Stimulation Fluid

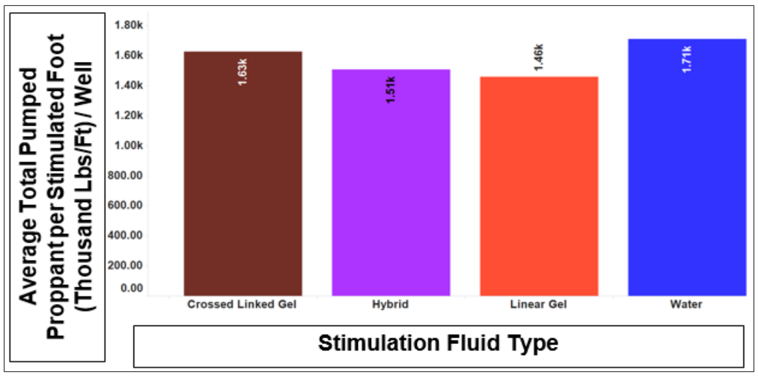


Figure 15. Bar Chart Comparing the Marcellus Average Total Proppant Usage per Stimulated Foot per Well for Different Stimulation Fluid Types

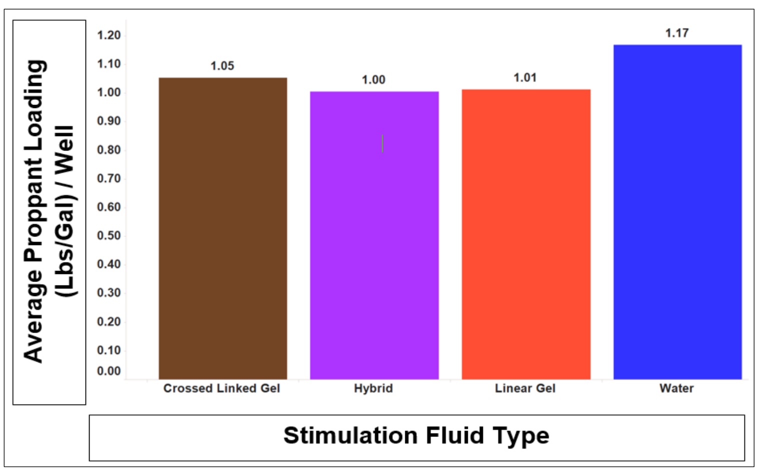


Figure 16. Bar Chart Comparing the Marcellus Average Proppant Loading per Well for Different Stimulation Fluid Types

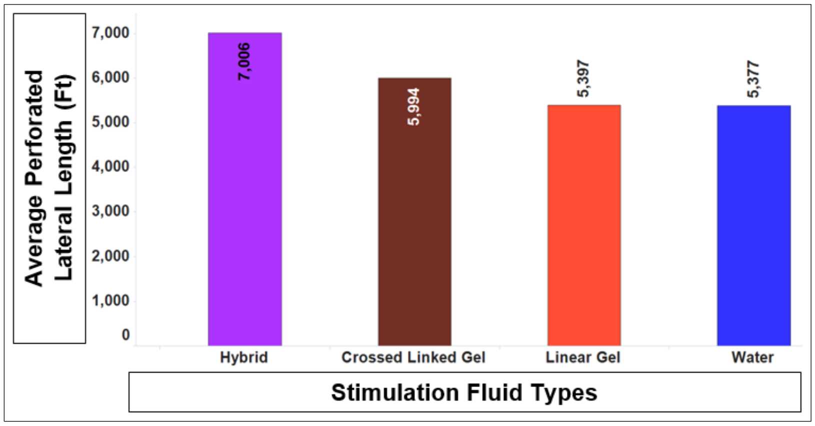


Figure 17. Bar Chart Representing the Marcellus Average Perforated Lateral Length for the Different Types of Stimulation Fluids Between 2012-2017

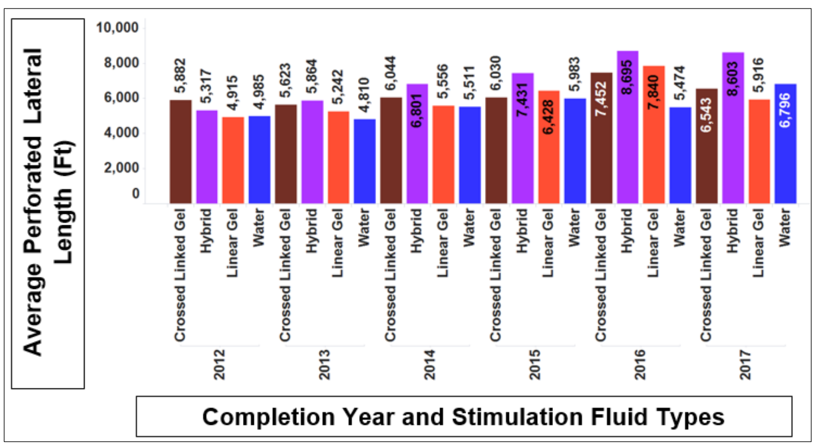


Figure 18. Bar Chart Representing the Marcellus Average Perforated Lateral Length between 2012 and 2017 for the Different Types of Stimulation Fluids

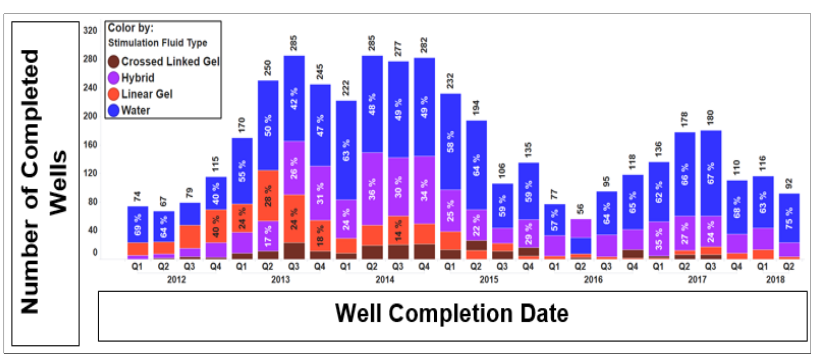
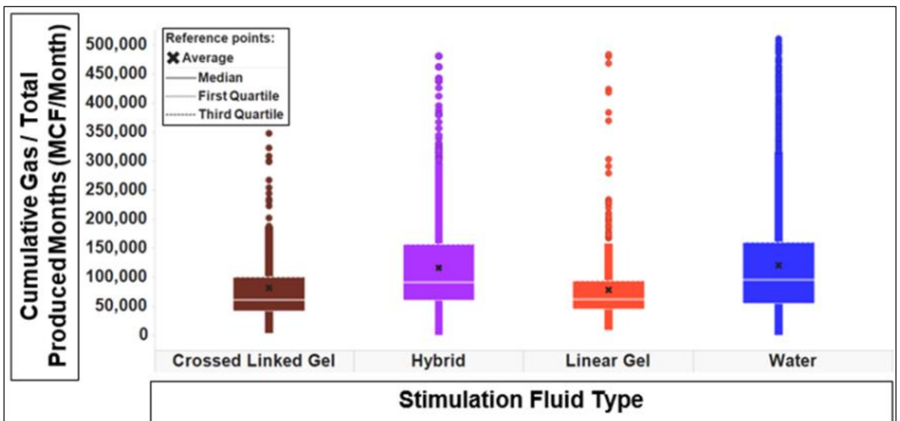


Figure 19. Stacked Bar Chart of the Number of Completed Wells over the Recent Years Colored Based on the Stimulation Fluid Types

3.3. BOX PLOTS AND STATISTICAL TABLES OF THE MARCELLUS PRODUCTION DATA

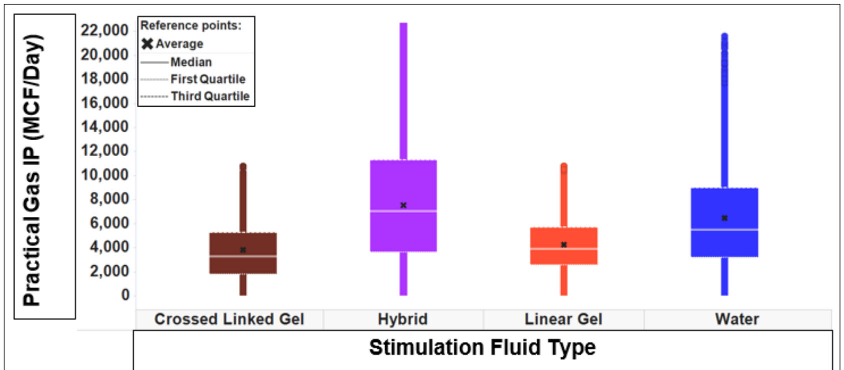
Figure 20 and Figure 21 provide production data distributions using box plots along with the associated summary statistical table. Each fluid type's box plot is color coded based on the type of stimulation fluid. The average value is marked as an x mark on the box plot. The median value is marked as a white solid line and the first and third quartiles are marked as dotted lines. The statistical tables shows the count of number of wells per stimulation type, the average value, median value, the number of outliers, the value of the first quartile, the value of the third quartile, the value of the first 10% of the data and the value of the 90% of the data. These boxplots present the final production data from this study. Figure 20 shows a box plot and statistical table of the cumulative production divided by the number of produced months in (MCF/month). Figure 21 shows a box plot and statistical table of the practical IP (MCF/Day).



	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	203	1,082	546	2,240
Avg	82,072	116,500	77,956	121,200
Median	61,251	91,175	61,973	95,558
Outliers	13	37	34	118
Q1	41,802	60,052	44,156	54,747
Q3	100,181	156,622	93,335	159,628
P10	28,716	39,507	31,285	32,366
P90	150,043	226,685	134,543	250,148

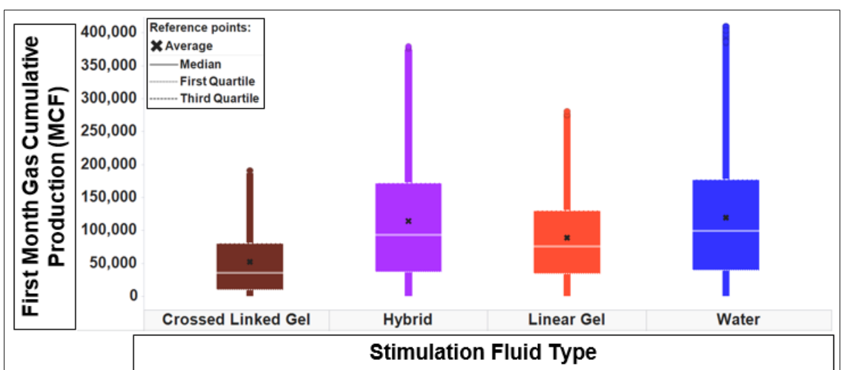
Figure 20. Box Plot and Statistical Table of Marcellus Cumulative Gas Production per Total Produced Months

The box plots and statistical tables of the cumulative gas production for the first month, first 6 months, first year, two years, and 5 years are presented in Figure 22, Figure 23, Figure 24, Figure 25, and Figure 26, respectively.



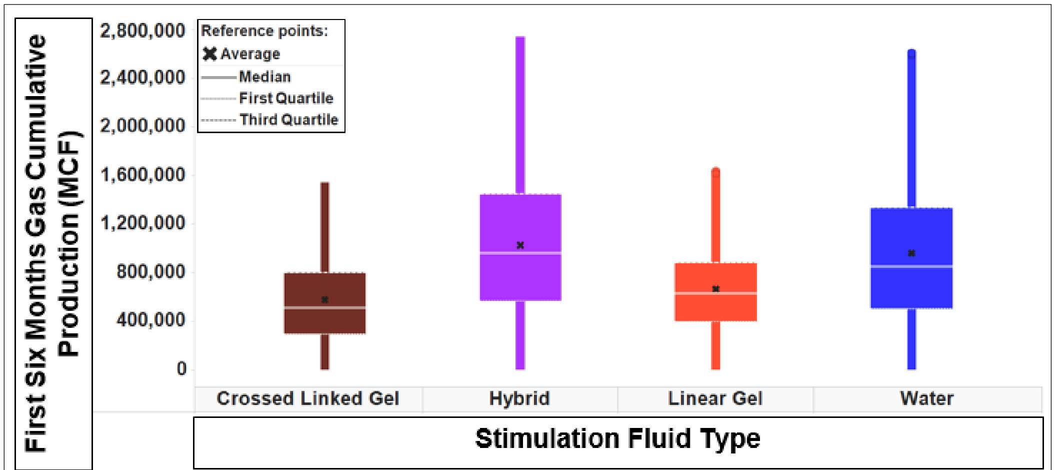
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	184	1,065	518	2,195
Avg	3,836	7,538	4,229	6,470
Median	3,304	7,035	3,923	5,497
Outliers	3	0	8	50
Q1	1,765	3,607	2,577	3,170
Q3	5,243	11,313	5,680	8,976
P10	696	1,874	1,238	1,618
P90	8,022	13,975	7,679	12,857

Figure 21. Box Plot and Statistical Table of Marcellus Practical Gas Initial Production



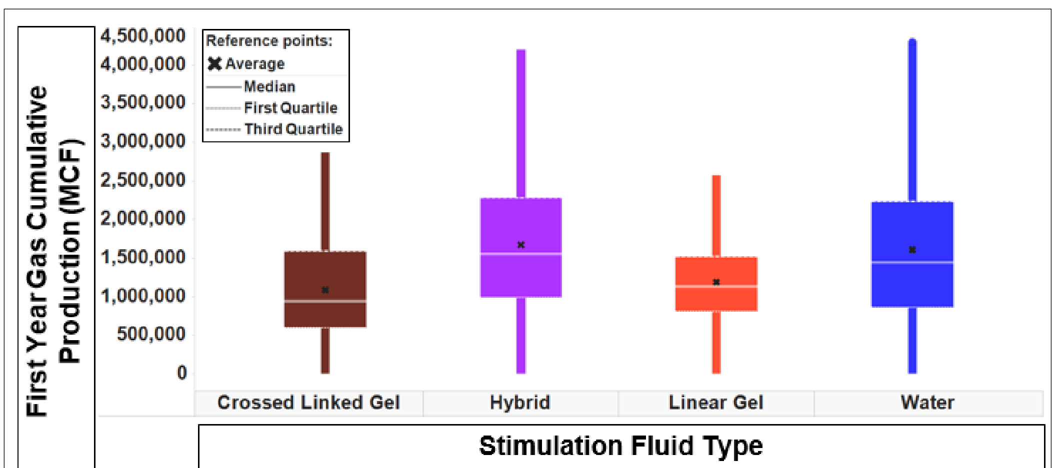
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	182	1,044	514	2,176
Avg	52,627	114,466	88,604	119,289
Median	36,076	93,426	76,283	99,292
Outliers	1	4	3	25
Q1	9,717	36,272	33,899	39,358
Q3	80,577	171,870	130,146	177,380
P10	263	6,286	6,863	8,720
P90	130,339	260,715	185,522	261,293

Figure 22. Box Plot and Statistical Table of Marcellus First Month Gas Cumulative Production



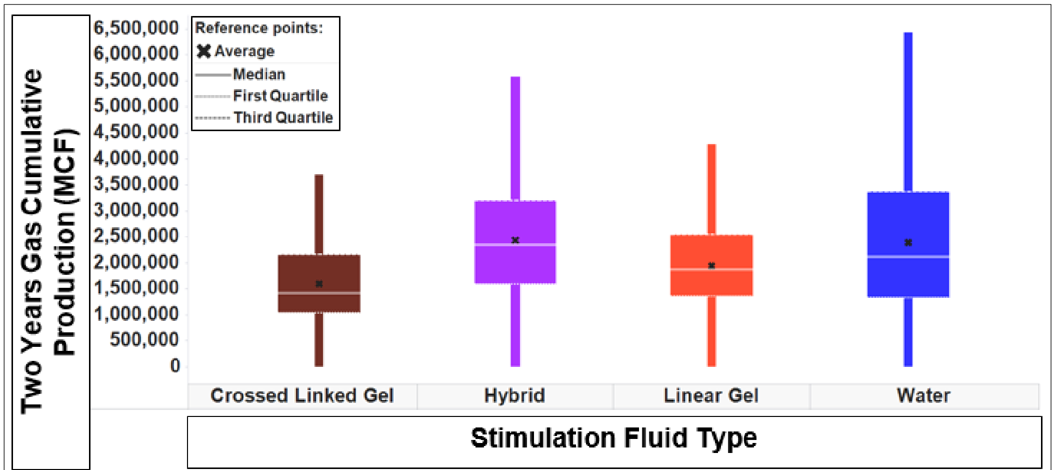
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	183	1,037	504	2,100
Avg	576,319	1,028,977	665,561	961,112
Median	513,952	961,580	628,300	853,831
Outliers	0	0	3	7
Q1	292,511	564,542	396,153	500,149
Q3	798,454	1,446,377	882,232	1,334,520
P10	162,376	315,755	241,435	261,250
P90	1,118,705	1,846,016	1,168,850	1,856,167

Figure 23. Box Plot and Statistical Table of Marcellus First Six Months Gas Cumulative Production



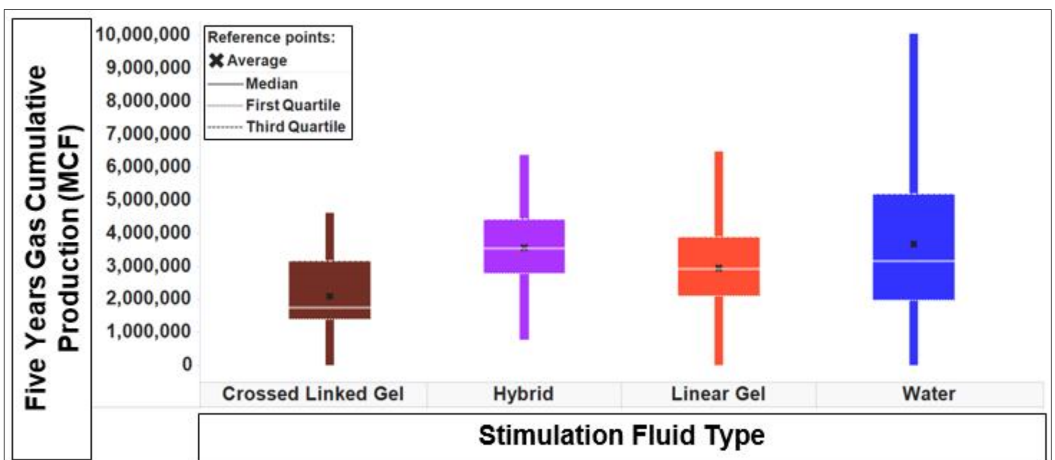
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	180	963	486	1,910
Avg	1,084,118	1,675,263	1,191,574	1,608,913
Median	942,689	1,554,910	1,135,483	1,444,737
Outliers	0	0	0	1
Q1	598,928	992,418	802,354	856,442
Q3	1,588,559	2,274,851	1,515,674	2,229,467
P10	381,683	559,376	516,067	490,875
P90	1,959,442	2,884,981	1,967,100	3,062,153

Figure 24. Box Plot and Statistical Table of Marcellus First Year Gas Cumulative Production



	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	158	789	461	1,580
Avg	1,605,814	2,440,254	1,948,247	2,399,947
Median	1,421,165	2,352,285	1,877,817	2,122,385
Outliers	0	0	0	0
Q1	1,035,665	1,588,407	1,355,369	1,321,001
Q3	2,166,476	3,199,050	2,545,956	3,371,041
P10	634,235	969,643	896,415	774,436
P90	2,916,482	4,114,239	3,164,396	4,443,171

Figure 25. Box Plot and Statistical Table of Marcellus First Two Years Gas Cumulative Production



	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	33	134	195	359
Avg	2,101,067	3,571,196	2,959,239	3,671,438
Median	1,762,129	3,558,076	2,926,392	3,163,584
Outliers	0	0	0	0
Q1	1,387,610	2,790,660	2,086,810	1,962,562
Q3	3,159,399	4,427,494	3,888,574	5,201,279
P10	1,062,192	1,603,725	1,303,425	1,105,469
P90	4,040,278	5,209,940	4,841,501	7,245,098

Figure 26. Box Plot and Statistical Table of Marcellus Five Years Gas Cumulative Production

3.4. COMPARING TREATMENT FLUID TYPES WITH GAS IP, CUMULATIVE FIRST MONTH, 6 MONTHS, FIRST YEAR, 2 YEARS, AND 5 YEARS GAS PRODUCTION

Well performance was found to vary between initial producing and longer-term recovery, for the four hydraulic fracturing fluid types studied. Figure 27 and Figure 28 show comparisons between the four types of treatment fluids for the average initial production (IP) per well and the average IP per foot per well, respectively. Practical initial production is measured by dividing the second-month cumulative production by the number of produced days (MCF/Day). Comparing the treatment fluids showed that hybrid fluids reflected slightly higher average practical IP than water treatment. Linear and cross-linked gel performed lower than water and hybrid fluids. To eliminate the bias related to the difference between the lateral length of the stimulated wells, the practical IP data were divided by the perforated lateral length to be compared with the average production per perforated and stimulated length. Water treatment showed the highest productivity per stimulated foot followed by hybrid, then linear gel and finally crossed-linked gel. Figure 29 shows the average cumulative first-month production in Marcellus wells and Figure 30 shows the average first-month cumulative gas production per perforated foot in Marcellus. The cumulative 6 months gas and 6 months gas/ft are represented in Figure 31 and Figure 32, respectively. The average first year, two years and five years cumulative gas production of the Marcellus wells are shown in Figure 33, Figure 34 and Figure 35, respectively. The average cumulative gas per perforated length (MCF/Ft) for the Marcellus first year, two years and five years are presented in Figure 36, Figure 37, and Figure 38, respectively. Figure 39 shows the cumulative gas production divided by the number of produced wells for the wells completed after the year 2017.

In all presented production cases, water fractured wells in the Marcellus outperformed the other types of stimulation fluid. This is especially evident when the average production is normalized to the perforated lateral length.

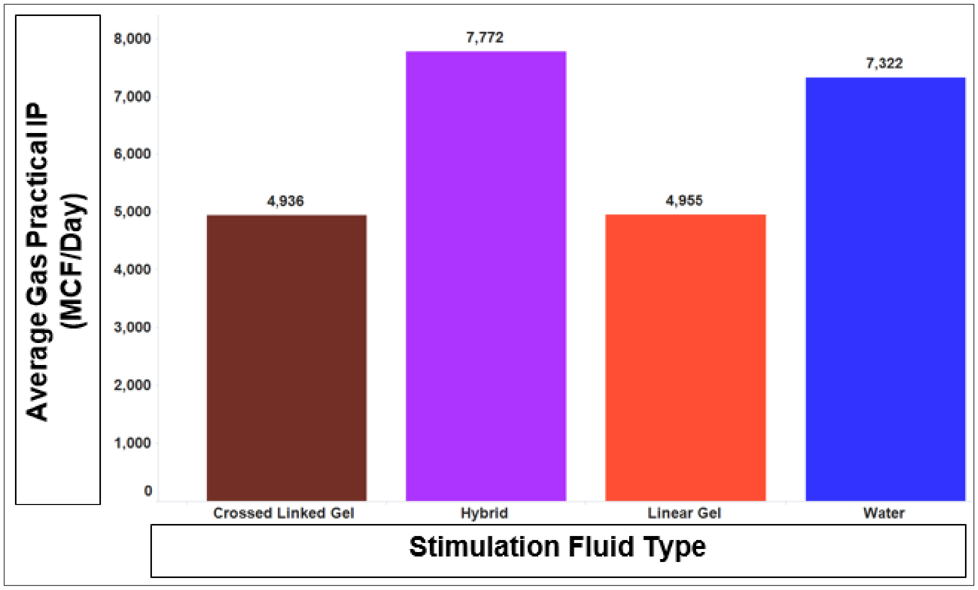


Figure 27. Bar Chart Comparing the Marcellus Average Practical IP for Different Stimulation Fluid Types

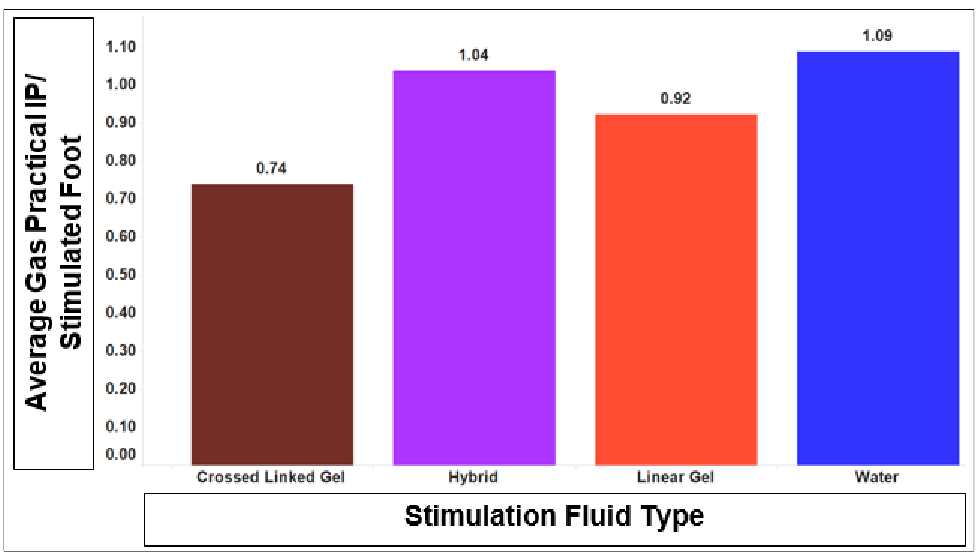


Figure 28. Bar Chart Comparing the Marcellus Average Practical IP per Stimulated Foot for Different Stimulation Fluid Types

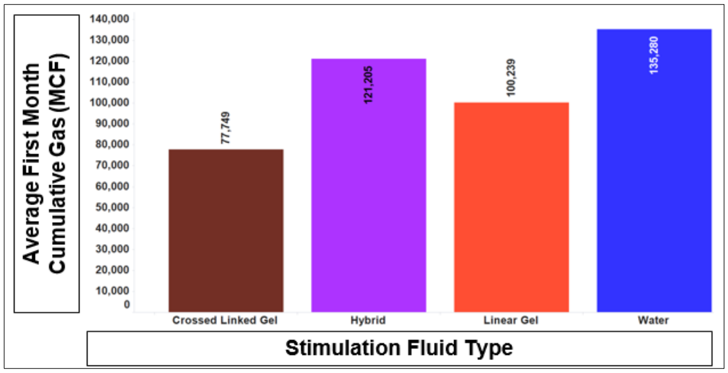


Figure 29. Bar Chart Comparing the Marcellus Average First Month Cumulative Gas for Different Stimulation Fluid Types

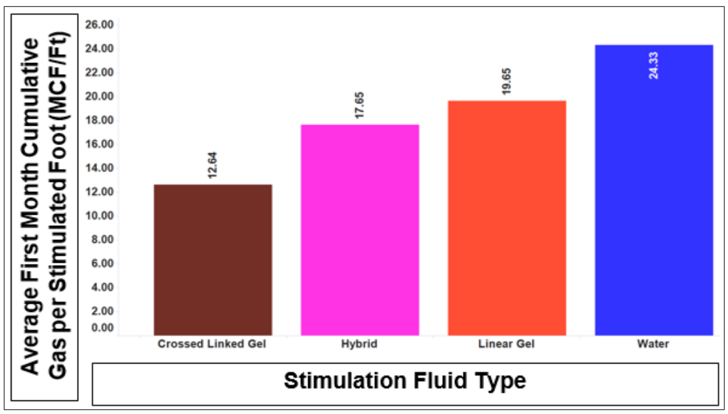


Figure 30. Bar Chart Comparing the Marcellus Average First Month Cumulative Gas per Stimulated Foot for Different Stimulation Fluid Types

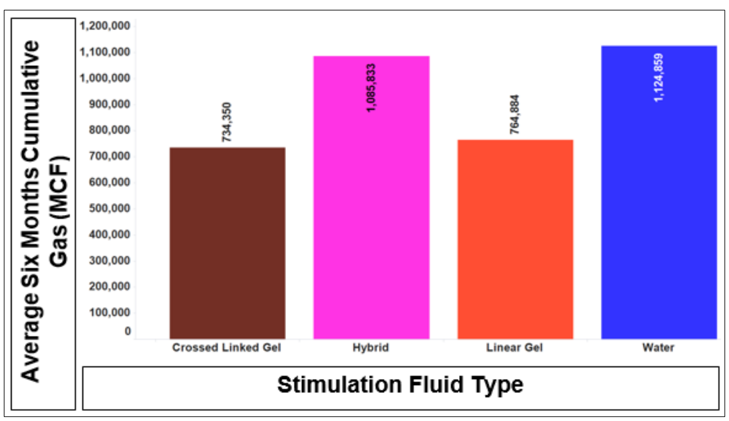


Figure 31. Bar Chart Comparing the Marcellus Average Six Months Cumulative Gas for Different Stimulation Fluid Types

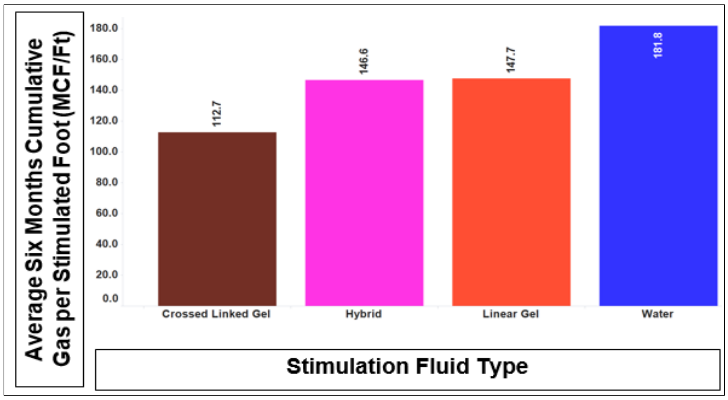


Figure 32. Bar Chart Comparing the Marcellus Average Six Months Cumulative Gas per Stimulated Foot for Different Stimulation Fluid Types

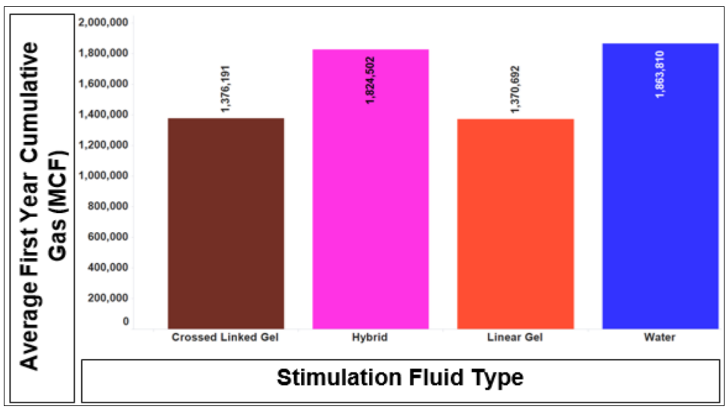


Figure 33. Bar Chart Comparing the Marcellus Average First Year Cumulative Gas for Different Stimulation Fluid Types

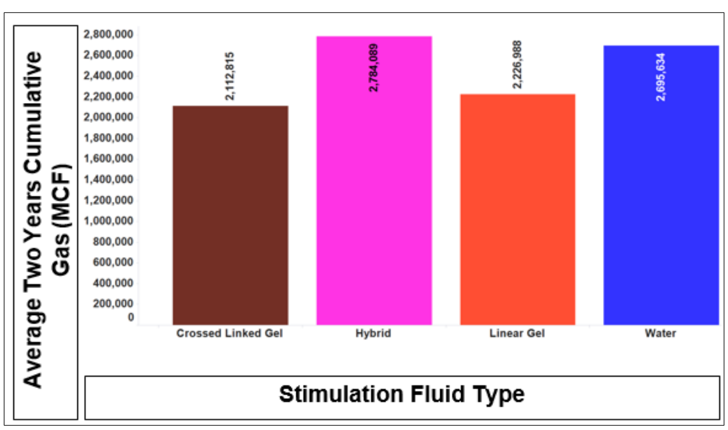


Figure 34. Bar Chart Comparing the Marcellus Average Two Years Cumulative Gas for Different Stimulation Fluid Types

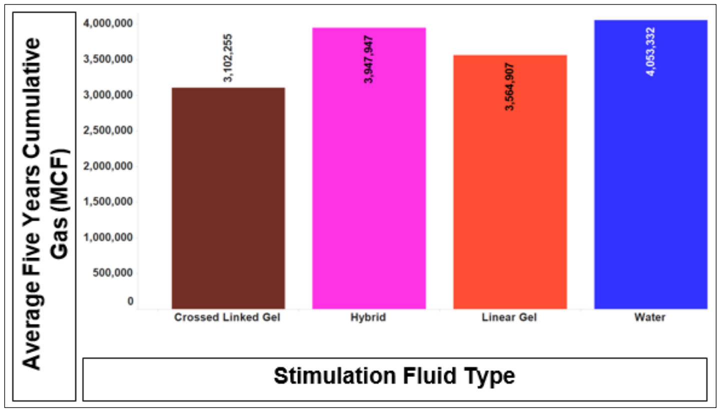


Figure 35. Bar Chart Comparing the Marcellus Average Five Years Cumulative Gas for Different Stimulation Fluid Types

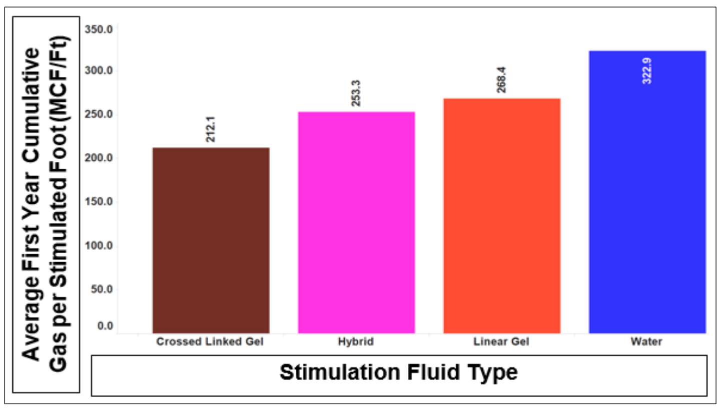


Figure 36. Bar Chart Comparing the Marcellus Average First Year Cumulative Gas per Stimulated Foot for Different Stimulation Fluid Types

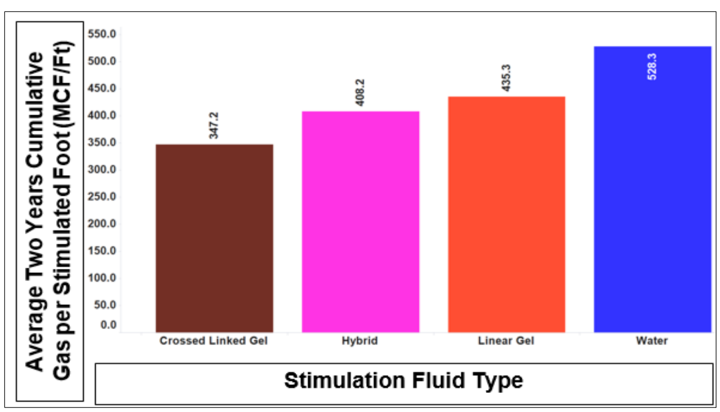


Figure 37. Bar Chart Comparing the Marcellus Average Two Years Cumulative Gas per Stimulated Foot for Different Stimulation Fluid Types

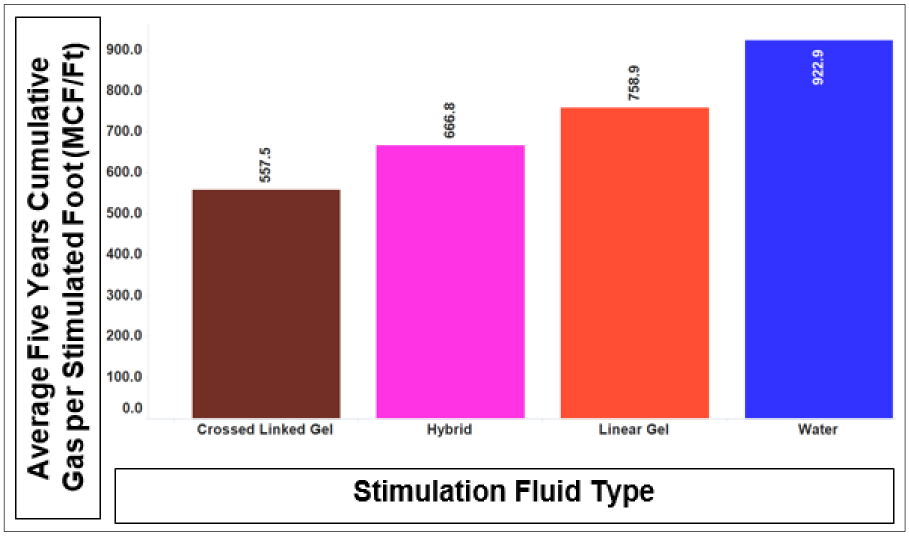


Figure 38. Bar Chart Comparing the Marcellus Average Five Years Cumulative Gas per Stimulated Foot for Different Stimulation Fluid Types

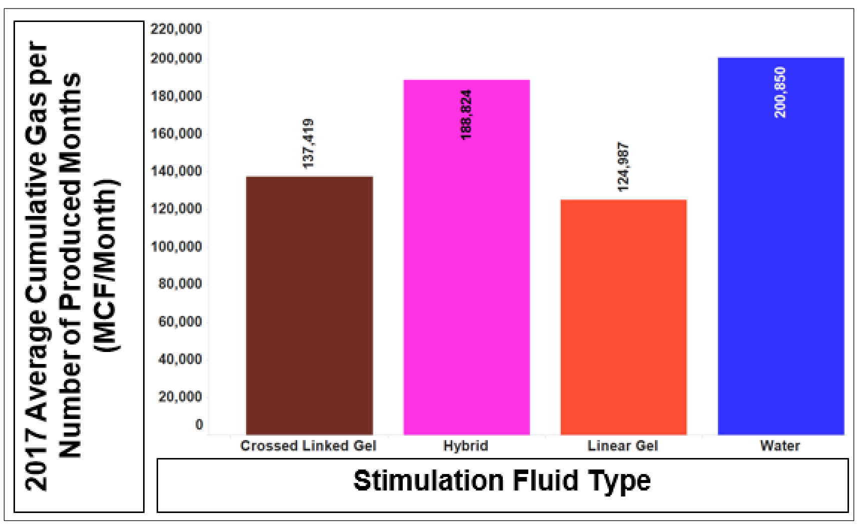


Figure 39. Bar Chart Comparing the Marcellus Average 2017 Cumulative Gas per Number of Produced Months for Different Stimulation Fluid Types

3.5. COMPARING THE EFFECT OF STIMULATION PARAMETERS WITH AVERAGE GAS PRODUCTION

Figure 40 through Figure 45 show average proppant mass per perforated lateral length and average water volume per perforated lateral length versus the cumulative gas production per perforated lateral length for the first year, 2 years and 5 years, respectively. Within each of these figures, the blue bar charts represent the average proppant mass per foot (Lbs/Ft) and the green chart represents the average water volume per foot (Gal/Ft). The solid black line chart represents the average cumulative production per foot (MCF/ft).

This analysis indicates, Marcellus water fracs result in higher cumulative recovery compared to other hydraulic fracturing fluid types. Water fracs used relatively more proppant per foot compared to cross-linked gel but produced 75% incremental cumulative gas production within 2 years and more than double the production of cross-linked fluids at 5 years cumulative recovery. Water fracs used the least amount of water per stimulated foot with respect to all of the four treatment fluid types while performed the best in terms of short and long-term cumulative production.

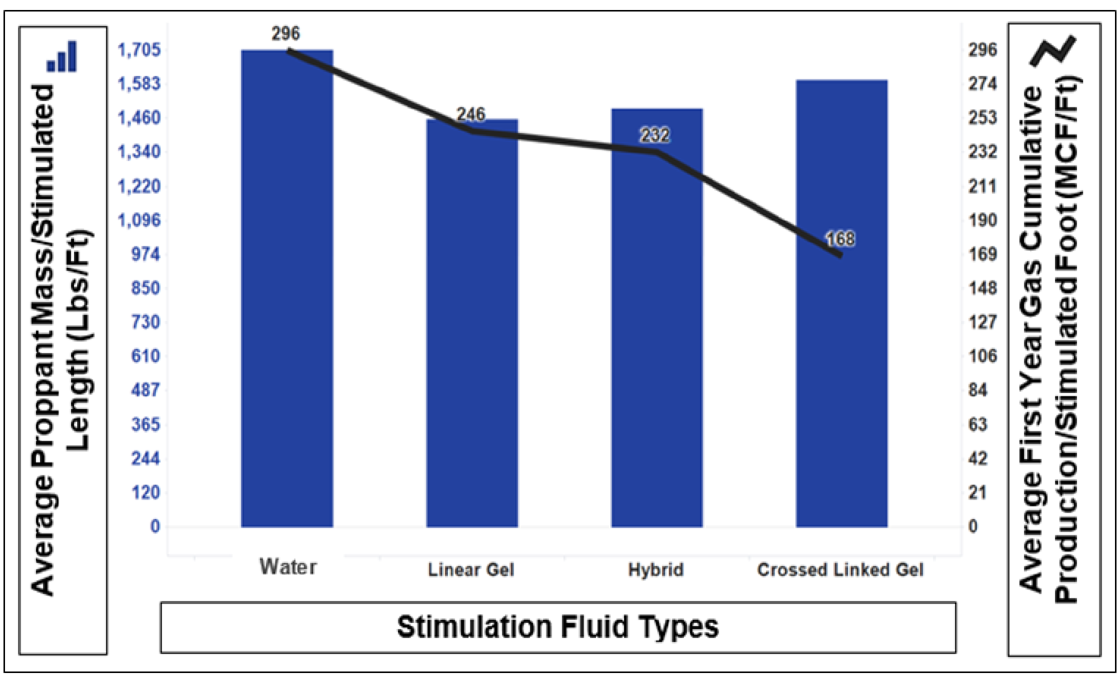


Figure 40. Combined Bar and Line Charts Comparing the Marcellus Average Pumped Proppant Mass per Foot and the First Year Cumulative Gas Production per Stimulated Length for each Stimulation Fluid Type

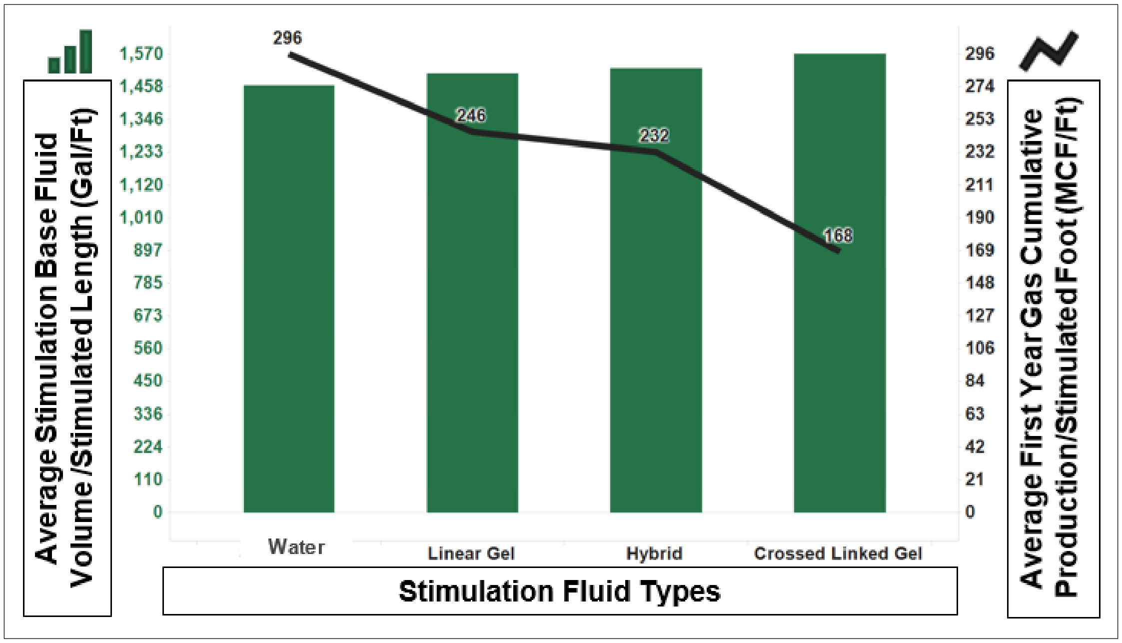


Figure 41. Combined Bar and Line Charts Comparing the Marcellus Average Pumped Stimulation Base Fluid per Foot and the First Year Cumulative Gas Production per Stimulated Length for each Stimulation Fluid Type

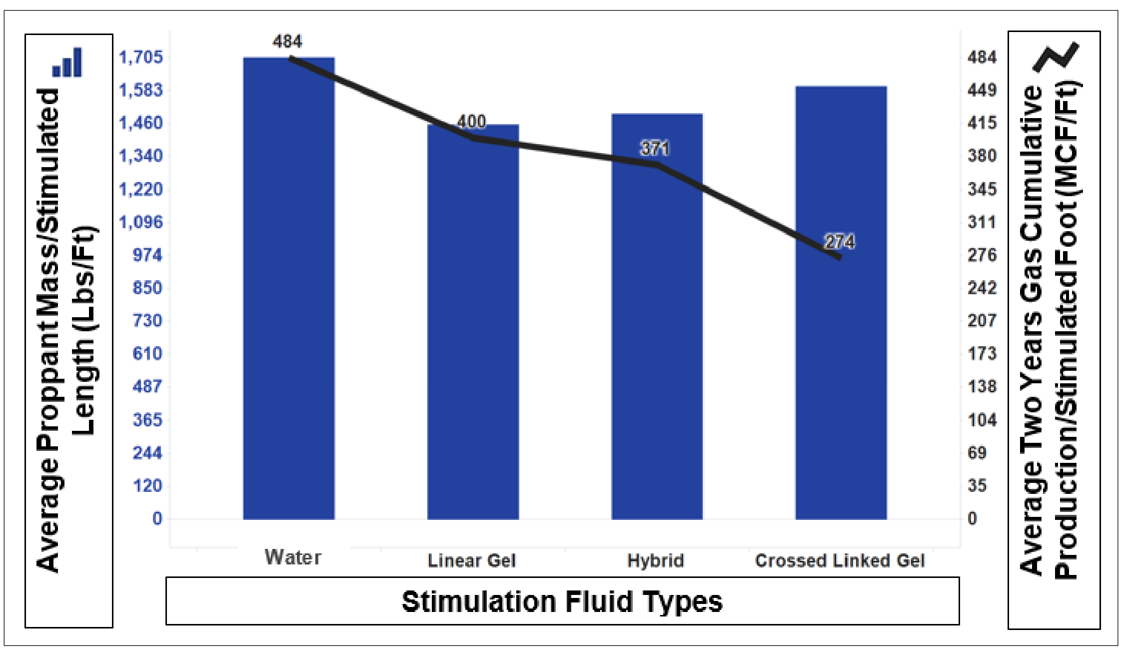


Figure 42. Combined Bar and Line Charts Comparing the Marcellus Average Pumped Proppant Mass per Foot and the Two Years Cumulative Gas Production per Stimulated Length for each Stimulation Fluid Type

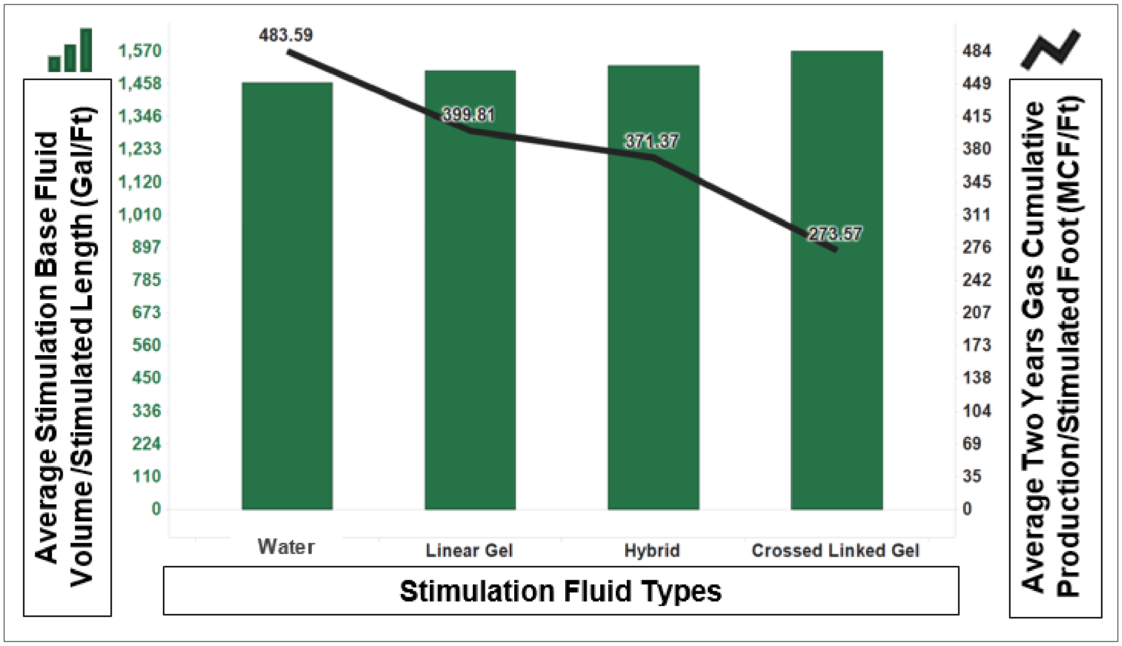


Figure 43. Combined Bar and Line Charts Comparing the Marcellus Average Pumped Stimulation Base Fluid per Foot and the Two Years Cumulative Gas Production per Stimulated Length for each Stimulation Fluid Type

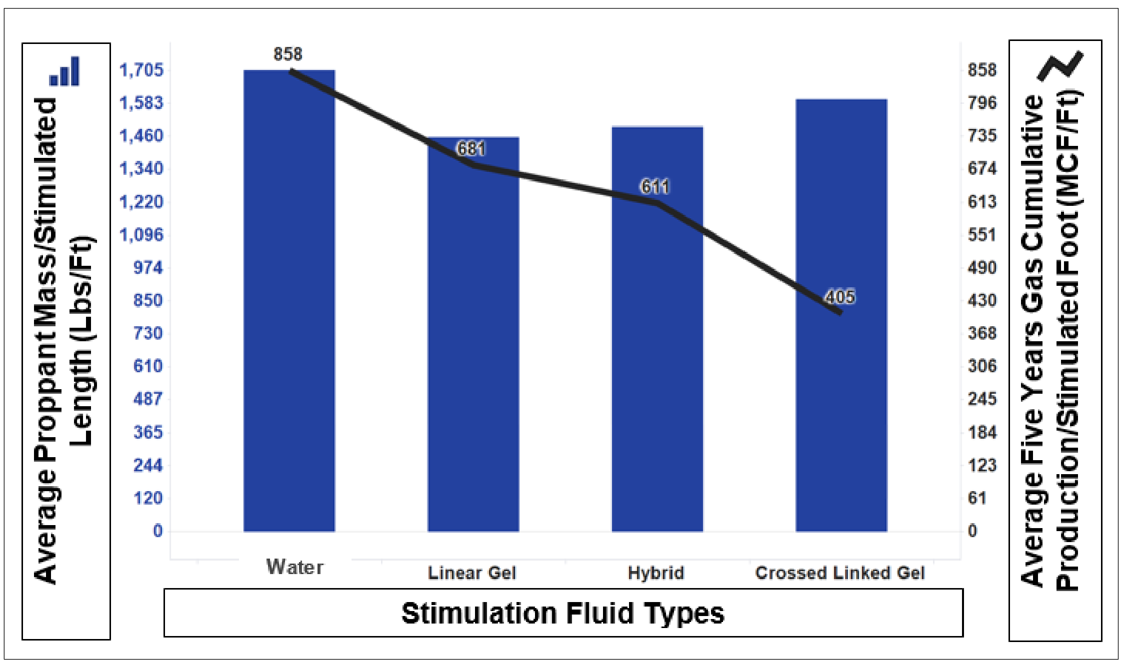


Figure 44. Combined Bar and Line Charts Comparing the Marcellus Average Pumped Proppant Mass per Foot and the Five Years Cumulative Gas Production per Stimulated Length for each Stimulation Fluid Type

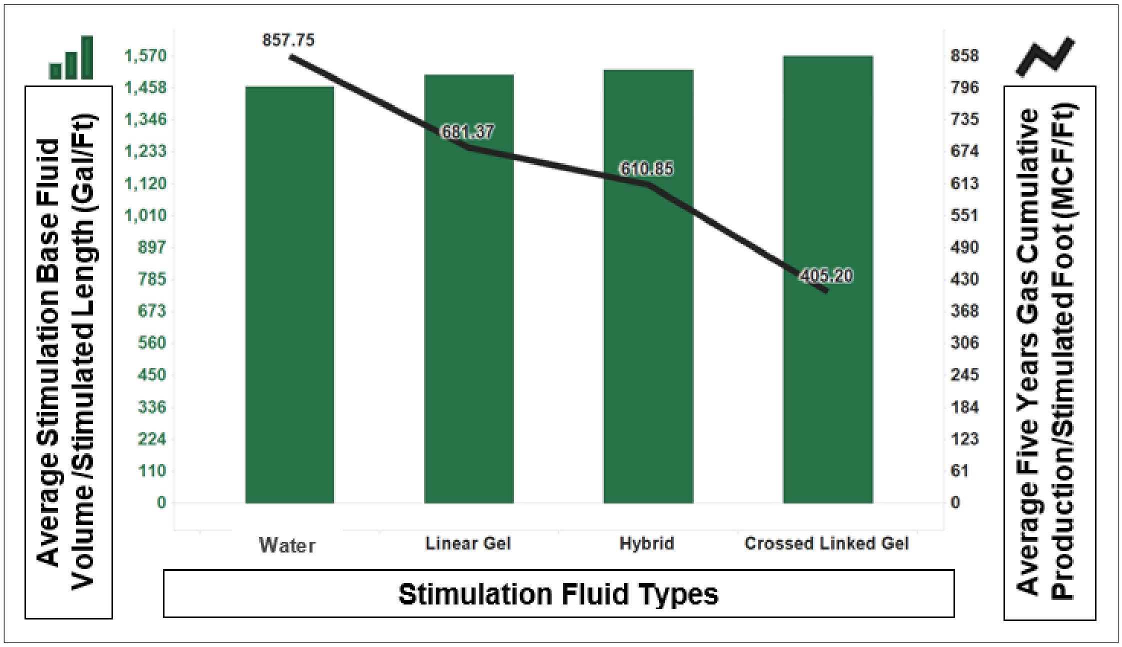


Figure 45. Combined Bar and Line Charts Comparing the Marcellus Average Pumped Stimulation Base Fluid per Foot and the Five Years Cumulative Gas Production per Stimulated Length for each Stimulation Fluid Type

4. CONCLUSIONS

In this study, chemical data from FracFocus were combined with well completion and production data from DrillingInfo, to create a robust database for investigating Marcellus well production as a function of hydraulic fracturing fluid type.

The main conclusions of this study were summarized in the following points:

- FracFocus chemical registry is a valuable source of hydraulic fracturing fluid, chemical and proppant data for wells stimulated after 2012. The data are available for download to Excel. The raw data must be reformatted for practical use, which may require some fundamental understanding of hydraulic fracturing treatments.

- This study provides detailed methodologies to utilize and process FracFocus public chemical registry database of stimulated wells in the United States. The workflows include raw data manipulation, grouping chemicals, calculating and verifying wells mass percent treatments, and merging final FracFocus data with DrillingInfo.
- Marcellus completions vary as a function of hydraulic fracturing fluid used. The greatest average treatment fluid volume and proppant are associated with hybrid fluid treatments.
- When normalized for lateral length, water frac treatments exhibited the highest average proppant load per ft/well. Cross-linked fracturing fluids used the highest average fluid volume/ft/well.
- The wells treated with water fracs had slightly shorter lateral length than other types of treatment fluids while hybrid fracs were associated with the longest lateral length.
- Hybrid fluid fracs showed slightly higher average practical IP than water treatments. Average 2-year cumulative recovery for hybrid fluids are slightly higher than water frac treatments.
- Water fractured Marcellus wells 5-year cumulative gas production significantly exceeds other fracturing fluid treatments. Data suggests an increasing number of Marcellus water fracs with time over the period of study.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to DrillingInfo team and the co-founder and director of university outreach Mr. Mark Nibbelink for their generosity in providing access to the U.S. production and completion data.

REFERENCES

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar, A. M. Review of Stimulation and Completion Activities in the United States Shale Plays. Paper ARMA 19-200, presented at the 53rd US Rock Mechanics / Geomechanics Symposium held in New York City, New York, USA, 23-26 June 2019

Alkinani, H. H., Al-Hameedi, A. T. T., Dunn-Norman, S., Flori, R. E., Hilgedick, S. A., Al-maliki, M. A., Amer, A. S. (2018). Journal of King Saud University – Science Examination of the relationship between rate of penetration and mud weight based on unconfined compressive strength of the rock. Journal of King Saud University - Science. https://doi.org/10.1016/j.jksus.2018.07.020

American Chemical society, CAS, Retrieved on February/ 05/ 2019, from, https://www.cas.org/support/documentation/chemical-substances/faqs

Arthur, J. D., Layne, M. A., Hochheiser, H. W., & Arthur, R. (2014, February 4). Spatial and Statistical Analysis of Hydraulic Fracturing Activities in US Shale Plays and the Effectiveness of the FracFocus Chemical Disclosure System. Society of Petroleum Engineers. doi:10.2118/168640-MS

Carman, P., Li, C., & Wheeler, R. (2018, September 24). Unorthodox Fluids for Unconventional Completions: Maximizing Production and Avoiding Pitfalls. Society of Petroleum Engineers. doi:10.2118/191751-MS

DrillingInfo website. Retrieved on March/01/2019, from https://info.drillinginfo.com/

Elsner, Martin & Hölzer, Kathrin. (2016). es5b02818 si 001

EPA website. Retrieved on February/ 05/ 2019, from https://www.epa.gov/saferchoice/safer-ingredients

FracFocus website. Retrieved on May/01/2019, from http://fracfocus.org/

Khvostichenko, D., & Makarychev-Mikhailov, S. (2018, August 9). Oilfield Data Analytics: Linking Fracturing Chemistry and Well Productivity. Unconventional Resources Technology Conference. doi:10.15530/URTEC-2018-2903086

Luo, H., & Zhang, Y. (2018, October 5). Mining FracFocus and Production Data for Efficacy of Fracturing Fluid Formulations. Society of Petroleum Engineers. doi:10.2118/191787-18ERM-MS

Myers, R. (2018, October 5). LPG Fracturing in the Marcellus Shale Using Public Domain Information. Society of Petroleum Engineers. doi:10.2118/191815-18ERM-MS

Myers, R. R., Knobloch, T. S., Jacot, R. H., & Ayers, K. L. (2017, October 4). Production Analysis of Marcellus Shale Well Populations Using Public Data: What Can We Learn from The Past? Society of Petroleum Engineers. doi:10.2118/187505-MS

Olaoye, O., & Zakhour, N. (2018, September 24). Utilizing a Data Analytics Workflow to Optimize Completion Practices While Leveraging Public Data: A Permian Basin Case Study. Society of Petroleum Engineers. doi:10.2118/191416-MS

Romero, D. H., & Poston, S. W. (2016, May 10). Data Mining and Analysis - Eagle Ford Formation, South Texas. Society of Petroleum Engineers. doi:10.2118/179994-MS

Srinivasan, K., Ajisafe, F., Alimahomed, F., Panjaitan, M., Makarychev-Mikhailov, S., & Mackay, B. (2018, August 28). Is There Anything Called Too Much Proppant? Society of Petroleum Engineers. doi:10.2118/191800-MS

Weijers, L., Wright, C., Mayerhofer, M., Pearson, M., Griffin, L., & Weddle, P. (2019, January 29). Trends in the North American Frac Industry: Invention through the Shale Revolution. Society of Petroleum Engineers. doi:10.2118/194345-MS

Yang, Y., Robart, C. J., & Ruegamer, M. (2013, February 4). Analysis of U.S. Hydraulic Fracturing Design Trends. Society of Petroleum Engineers. doi:10.2118/163875-MS

II. PRODUCTION PERFORMANCE ESTIMATION FROM STIMULATION AND COMPLETION PARAMETERS USING MACHINE LEARNING APPROACH IN THE MARCELLUS SHALE

Mustafa A. Al-Alwani, Larry K Britt, Shari Dunn-Norman, Husam H. Alkinani, Abo Taleb T. Al-Hameedi, Atheer M. Al-Attar

Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA

ABSTRACT

Harnessing the power of predictive statistical modeling and advanced machine learning techniques to evaluate and assess the impact of various well completion and stimulation parameters on the wells' production performance have become one of the recent growing interest in the oil and gas industry and especially in the development of the unconventional resources. The main objective of this study is to utilize the partial least square (PLS), a machine learning technique, to create predictive models to evaluate the impact of several well completion parameters such as lateral length and gross perforated interval coupled with different stimulation parameters such as the total pumped water and sand volumes in addition to the concentrations of different stimulation additives. The completion and stimulation parameters will be used as a predictor variable to the short- and long-term gas production in the Marcellus shale play. Another outcome from this study is to gain insights about the relationship between those parameters and the wells production in order to optimize the future designs of those wells. Data of more than 2700 wells were utilized from two sources, the stimulation parameters were gathered from FracFocus website and the well completion and production data were gathered from DrillingInfo database. Three production predictive models were presented in the form of mathematical

functions that were created based on the wells' data which can be considered as the training dataset. Unlike many other prediction modeling approaches, this study applies cross validation procedures to assure the reliability of the constructed prediction models.

1. INTRODUCTION

The recent breakthrough in the development of the unconventional resources has served as the main attributor behind the surge in the oil and gas production in the United States. Unlike conventional resources, the source rock which is usually organic-rich formations has become the target for production. Unfortunately, most of those resources suffer from extremely low permeability and require hydraulic fracturing stimulation to create flow paths and allow the hydrocarbons to be produced. Nowadays, most of the hydraulically fractured wells are also drilled horizontally to allow more exposure to the targeted formation. One of the most challenging and difficult tasks is to simulate the unconventional reservoirs and create accurate reservoir models. The reason behind the stimulation difficulty is that any proposed model must encounter many of the complex flow features in addition to the physical and chemical interactions. A good reservoir model must accommodate the following functions: adsorption and desorption, turbulent flow in nano-scale pores, flow behavior in induced and natural fractures, overall stress effects on the proppant and the created fractures, the effects of stimulation chemical residues and their damaging skin effects, etc.

During the past few years, there has been a shift in the research area in the oil and gas industry towards a more attractive alternative of using data-driven statistical approaches to understand the reservoir performance and utilize it in optimizing the production in the unconventional reservoirs. Mining the data and applying several statistical and visualization techniques to investigate insights into any system performance and to discover important patterns and trends have been used earlier by other industries such as health care and environmental applications (Hastie et al., 2001). There are two types of predictive modeling techniques: supervised and unsupervised. Unlike the unsupervised learning techniques, the outcome variable is predicted using the number of variables called the predictors or the input variables and the dataset is usually divided into two segments: training and testing. The training dataset is used to build the model by utilizing regression, neural networks, support vector machine, decision trees, etc. Unsupervised learning is referred to the other forms of statistical and data-driven analysis that are applied to simply investigate patterns in different input variables or to determine the structure of the data and their clusters by utilizing cluster analysis, mapping, and principal component analysis (Schuetter et al, 2015).

In this paper, a supervised PLS model is used to predict short and long-term production based on completion and stimulation variables.

2. LARGE DATA UTILIZATION AND STATISTICAL MODELING LITERATURE

In the fast development of the unconventional oil and gas industry, many companies employ the power of big data and data-driven statistical modeling to determine the latent correlations between the geological and engineering parameters, and which of those parameters have the most effect on the initial and long-term production performance.

Predictive modeling is the approach of building mathematical or algorithmic models to produce accurate predictions. The model usually consists of one variable known as the response or the output and one or more independent variables known as the predictors. Predictive modeling is not only used to detect a pattern within the data and to quantify the response based on the predictors, but also used to create a model that is capable of quantifying outcomes based on future assumed values (Khun and Johnson, 2013). It must be mentioned that statistical approaches should not be considered as a replacement for intuition and field experience. It can be used as a supplemental to the experience and helps in decision making.

Khun and Johnson (2013) highlighted that no matter how widely the predictive models are used, the models at some point will fail to provide accurate predictions. To avoid the pitfalls of models inaccuracy, 1) the data should go through rigorous quality control and pre-processing techniques, 2) the models should be validated, 3) use the model within its scope of application 4) use the model within the data range that the model was originally built from to avoid excessive extrapolation, 5) select the appropriate number of the influencing factors to prevent the model from being overfitted to the data, 6) pay attention to data dependence on time and location (Harris, 2014).

The utilization of data analysis in enhanced oil recovery (EOR) and drilling application has been discussed by the following studies: (Al-Dhafeeri et al., 2005) (Alvardo et al., 2008), (Baker et al., 2012), (Aldhaheri et al., 2016), (Leite Cristofaro et al., 2017), (Hegde et al., 2015) and (Wallace et al., 2015)

Many studies in the literature discussed some simple linear regression models and tried to model some of the completion variables and their relationship to production. Other studies focused on using the approach of random forests and gradient boosting machine which are decision tree methods (Lolon et al., 2016). The random forest models are built by creating a collection of several uncorrelated regression trees (Hastie & Tibahirani, 2008) (Wilson, 2015).

Several studies have utilized the applications of data mining in the assessment of unconventional resources. LaFollette and Holcomb (2011) documented a study from the Barnett shale in north Texas. They have used wells' architecture data to find and explain hidden trends that are not visible using simple scatter plots by combining geographical information system (GIS) pattern recognition with traditional statistical techniques.

Gong et al. (2011) used the Markov-chain Monte Carlo simulation (MCMC) to quantify the uncertainty in reserve in a probabilistic decline curve analysis method. The presented approach resulted in limited use of discrete P10, P50, and P90 production forecast which did not match or came close to the actual production data.

Bhattacharya et al. (2013) utilized data-driven deterministic model called classification and regression tree (CART) techniques to investigate the possibility of finding patterns in significant input variables that affect the values of hydraulic fracturing job pause time (JPT) in certain regions in an effort to identify the variables that cause proppant screenout. The model prediction capability in this study was enhanced by using normal score transform (NST) and by dividing the dataset into smaller groups by using clustering.

Mohaghegh (2013) presented a method for asset management by utilizing advanced data-driven and predictive analytics. In this webinar presentation, the author talked about the applications of advanced data-driven analytics in providing many insights in hydraulic fracturing applications in unconventional formations. Mohaghegh (2013) also highlighted the facts that this type of techniques utilizes “Hard Data”. Hard data referred to variables that are obtained from field measurements such as type and amount of hydraulic fracturing fluids, proppant, chemicals, break down pressure, instantaneous shut-in pressure (ISIP), closure pressure, etc. While soft data referred to the variables that inferred or estimated without directly measuring them such as fracture half-length, fracture height and width, conductivity, etc.

Gupta et al. (2014) discussed an alternative method to reservoir simulation and history matching in the unconventional reservoirs by utilizing data mining and time series analysis to forecast production. This paper used the neural network techniques (NN) to predict the future production performance of gas wells in shale formations based on the historical production data from previous years. This study also used time series analysis to detect trends in the rate of decline which is similar to the techniques used in stock markets to predict the stock ticks.

Esmaili and Mohaghegh (2014) presented a novel approach to model and history match gas production in the Marcellus shale using advanced data mining, machine learning, and pattern recognition techniques. They incorporated the use of hard data such as production data, well logs, completion, and stimulation data to guide the model and determine its behavior.

Fulford et al. (2015) expanded the work of Gong et al. (2011) and presented a machine learning study to predict production in liquid-rich unconventional shale wells. Fulford et al. (2015) utilized transient hyperbolic model (THM) to reflect the different flow regime existing in the unconventional shale wells. They also increased the accuracy of discrete forecasts by constructing type curves from the analyzed wells and modified the likelihood forecasts.

Schuetter et al. (2015) used data analytics for production optimization in unconventional reservoirs. The paper discussed issues on how to build a good predictive model and showed a systematic approach on how to decide on the well's good from a bad performance. The study focused on data from the Wolfcamp shale in the Permian Basin. The statistical approach in this study was mainly focused on building predictive models for production using simple regression and some advanced methods such as random forests, support vector regression, gradient boosting machine and multidimensional kriging. The decision tree approach was used to provide useful insights into the most influential variables on production performance.

Al-Hameedi et al. (2018) utilized PLS techniques to predict lost circulation volumes, ECD, and penetration rate (ROP). They used large dataset of drilling parameters to construct their loss circulation prediction model and compared the new model to previously developed models.

3. MODELING STIMULATION AND COMPLETION PARAMETERS IN THE MARCELLUS SHALE PLAY

Gathering and analyzing large data by utilizing the public domains and the available commercial databases have become a trendy and fruitful research topic in the area of hydraulic fracturing. The importance of the topic has gain popularity due to the vast amount of data that are being generated and recorded every day in every aspect and phase of the oil and gas wells' lifecycle. Harnessing the power of big data analysis techniques helps in grasping knowledge and insights into the relationships between parameters and general trends of different aspects of any process to optimize future performance.

In this paper, a newly developed model is introduced to investigate the Marcellus shale gas wells production performance based on the stimulation and completion parameters. The model investigates parameters such as completed lateral length, volume of pumped sand (proppant), the total pumped stimulation fluid (water), the concentration of the following additives: friction reducer, biocides, acid, clay control, scale inhibitor, corrosion inhibitor, and iron control. Those parameters were used to predict the initial production, the first 12, and 24 months of gas cumulative production within the Marcellus shale play.

Data from over 2700 horizontal wells form the Marcellus shale were utilized in this study, two sources of data were gathered, filtered, and combined to be used in building the prediction models. The first source of data was from data mining FracFocus chemicals disclosure registry website. The FracFocus data were gathered and then processed through a series of data workflows to extract accurate data of the amount and the concentration of all the chemicals that have been used in stimulating every well. The second source of data was from DrillingInfo database where the completion and production data were gathered. Then, the two data sources were combined and some parameters were normalized to help in presenting more accurate prediction models. Most data were utilized in training the proposed models, the models were also tested with an independent dataset that has not been used in creating the models.

The main goal of this new procedure was to use an advanced machine learning algorithm by utilizing the partial least square (PLS) regression techniques to estimate the potential initial production and longer-term production of gas wells in the Marcellus play before completing and stimulating the future wells. These performance prediction models will enable operators in the area to selectively choose the best completion and stimulation parameters that will yield the best cumulative gas production based on the history of the previously treated wells' performance.

4. PARTIAL LEAST SQUARE (PLS) REGRESSION ALGORITHM

The Principal component analysis which is usually referred as (PCA) is one of the unsupervised learning methods in the artificial intelligence applications. PCA is one of the common approaches that are utilized to infer low dimensional features from a large dataset with many investigated variables. To simply put it, PCA is mainly considered as a mathematical approach that changes a group of correlated variables to a smaller group of uncorrelated variables that is referred as principal components. In PCP, the X-variables represent the predictor variables and the y-variables represent the response variables (output). The data variability is honored within the PCA techniques, most of the variability of the predictors are represented in the first principal component. This approach is classified as unsupervised method because the response variables are not used in identifying the principal components. The unsupervised learning techniques cannot be tested within the model to determine the reliability in predicting the output variable within acceptable margin of errors. However, the supervised methods such as partial least square (PLS) allows the analyst to test the model with testing data from the predictors variables that have never been used during the model creation. PLS is considered as the supervised alternative method to the unsupervised PCA method. PLS method generate a number of inherent factors which express the combination linearity in the dataset. After the number of inherent factors are computed, the least squares method will be used to present a fitted model. One of the major differences between PCA and PLS is that the latter determines the newly generated latent factors in a manner that is considered to be supervised that distinguish both the predictor (X-variables) and the response (Y-variable). As it is explained by (James et al., 2013), PLS uses x and y variables to determine optimum direction which best demonstrate variability of x-variables while distinguishing y-variable. Due to the PLS high efficiency in dealing with large datasets with many variables which may possess collinearity and due to the recommendations in the literature to use this technique in the oil and gas industry applications (Tuffery, 2011), this method was chosen from other regression techniques to be implemented in this study.

To start with the PLS, the analyst should make sure that the regression algorithm data are centered and scaled. This requires that the predictors and the response must have a standard deviation equal one and a mean value of zero. The mean value of the variable and its variation are engaged in creating the generated latent factors, hence centering the data is very important to the successes of this method. When the variables are centered, this will mean that the change of standard deviation of one x-variable will be equivalent to the change of standard deviation of another x-variable. To demonstrate the interaction term and its calculation, assume X1 and X2 are two predictors variables and the interaction term (IT) can be calculated using the following Eq. (1) (SAS, 2008).

$$IT = \left(\frac{X_1 - mean\ (X_1)}{STD(X_1)}\right) * \left(\frac{X_2 - mean\ (X_2)}{STD(X_2)}\right) \qquad (1)$$

Non-linear iterative partial least square (NIPALS) is one of the popular algorithms in PLS. It functions by taking one factor at a time. Assume X=Xo as the x-variable matrix which is scaled and centered. Also assume Y=Yo as the y-variable matrix which is scaled and centered. A linear combination of the predictors will be created and represented as: t= Xo.w, t and w are defined as score vector and its associated weight vector respectively. Eq. (2) and Eq. (3) show how the PLS is predicting Xo and Yo using t regression.

$$\widehat{X_o} = t\acute{p}, \text{ Where } \acute{p} = (\acute{t}t)^{-1}\acute{t}\, X_o \quad (2)$$

$$\widehat{\mathrm{Y_o}} = \mathrm{t}\acute{c}, \text{ where } \acute{c} = (\acute{t}t)^{-1}\acute{t}\, Y_o \quad (3)$$

The c vector is called the Y-loading and p vector is called X-loading. The term $t=X_o.w$ will be utilized to maximize $\acute{t}\,u$ **(**covariance) with a combination of output (response) variable $u=Y_o.q$. Also, first eigenvectors of $\acute{X}_o Y_o \acute{Y}_o X_o$ and $\acute{Y}_o X_o \acute{X}_o Y_o$ are proportional to w, q, x-weight, y-weight. The previous steps show how to get the first latent factor. Similarly, the second latent factor can be obtained by replacing X_o with X-residual and Y_o with Y-residual of first latent factor as indicated in Eq. (4) and (5). Score vectors are obtained in the same process and repeated for the desired number of latent factors.

$$X_1 = X_o - \widehat{X_o} \quad (4)$$

$$Y_1 = Y_o - \widehat{Y_o} \quad (5)$$

4.1. CROSS VALIDATION

In order to confirm the number of the latent factors to be used in the model, a cross-validation process must be applied on all the computed latent factors. The selected number of latent factors must satisfy the following two requirements: 1) to aim on capturing the variation in the response variables while honoring the predictive parameter; 2) to avoid model overfitting. Scores plots and root mean of the predicted residual sum of squares (PRESS) are used to perform cross-validation. Figure 1 shows PRESS finding process for specific number of factors.

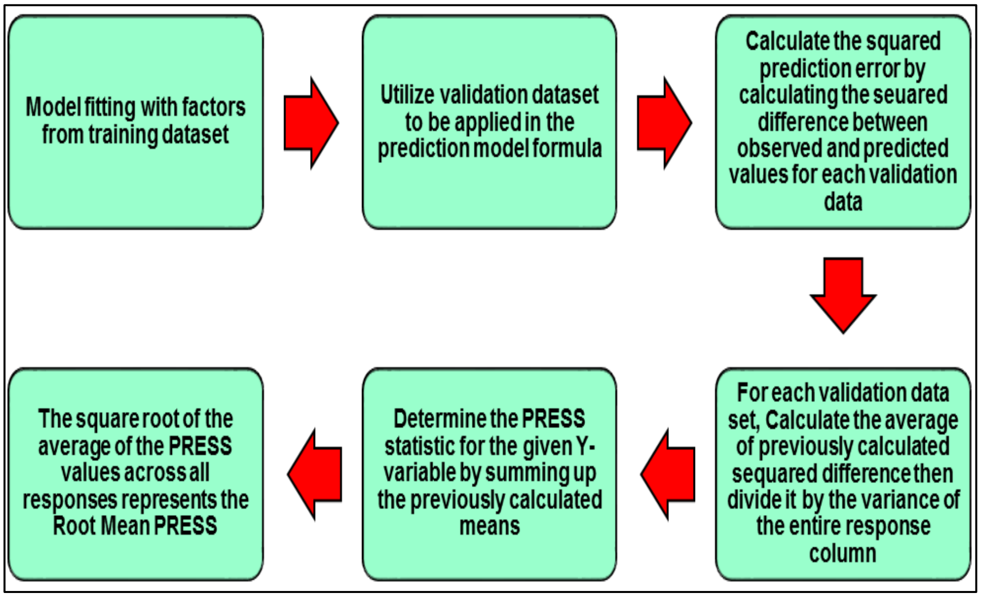


Figure 1. The Process of Determining the Root Mean of PRESS

4.2. VARIABLES IMPORTANCE IN PROJECTION

In any prediction analysis, one of the most important steps is to decide and select the response variables that greatly affects the prediction model from the variables of minimal importance and to be eliminated from the model. In PLS algorithm, this can be achieved by utilizing the variable importance in projection (VIP). VIP is generated for each x-variable and a cutoff limit must be agreed on to be used as a threshold value to eliminate the variables that are deemed to be below the cutoff value. As referenced by Eriksson et al, a cutoff of 0.8 is generally regarded as a low value, therefore, variables with less than 0.8 VIP would be dropped as referenced by (Eriksson et al, 2006).

For any y-variable (j1, j2… jn) The VIP calculation for each predictor is demonstrated in Eq. (6).

$$VIP_j = \sqrt{d\sum_{k=1}^{A} v_k\left(w_{kj}\right)^2 / \sum_{k=1}^{A} v_k} \quad (6)$$

Where d and A are the number of variables, and the number of latent factors respectively, vk is the variance of X and can be calculated in Eq. (7):

$$v_k = {c_k}^2 \acute{t}_k t_k \quad (7)$$

Where Ck is determined for every vector of t score and the predicted y as shown in Eq. (8):

$$c_k = \frac{\acute{t}_k y(k)}{\acute{t}_k t_k} \quad (7)$$

Figure 2 is a summary flowchart of the PLS algorithm (Al-Hameedi et al., 2018).

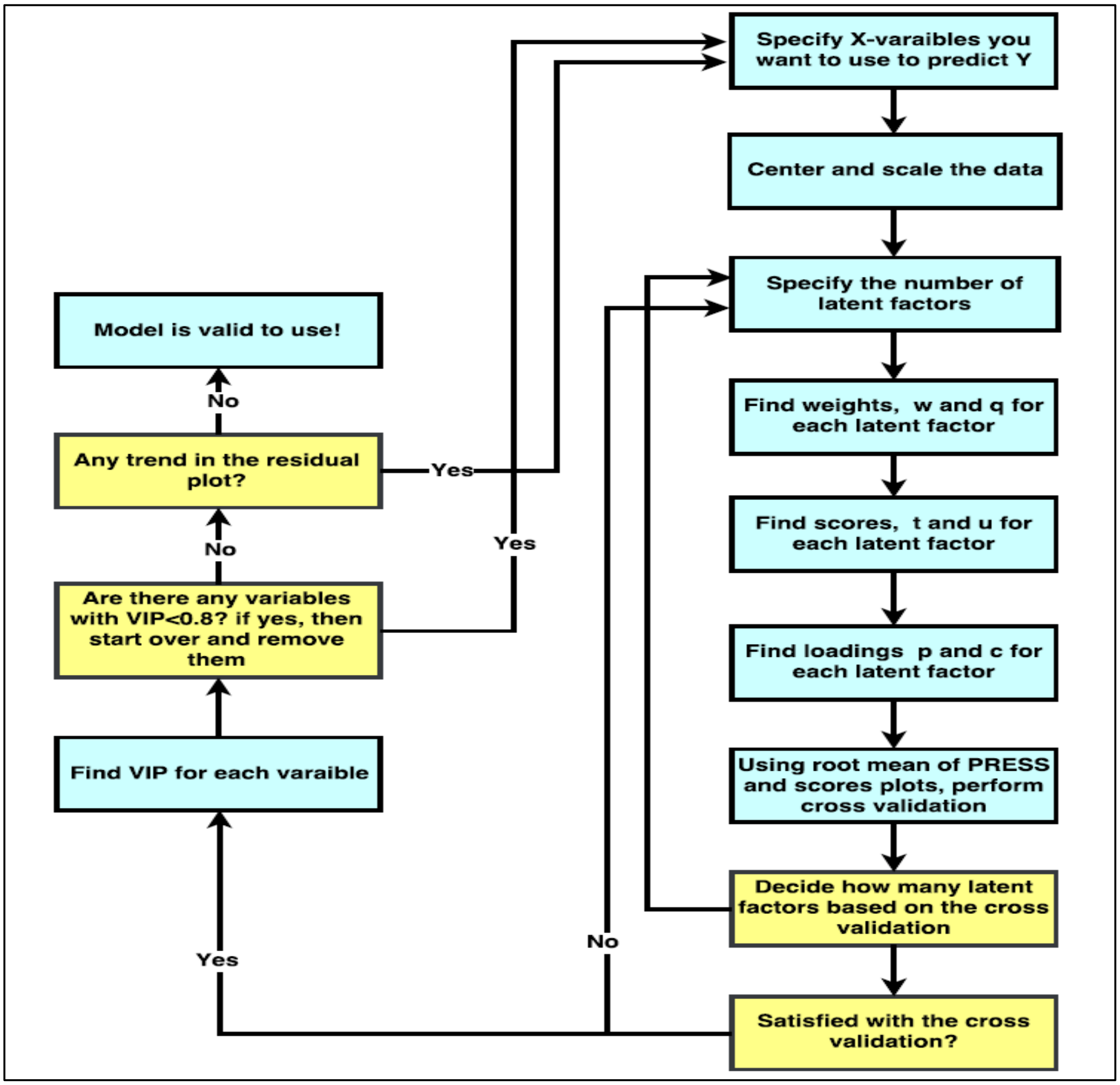


Figure 2. PLS Algorithm Summary (Al-Hameedi et al., 2018)

5. APPROACH

There is a large number of different completion and stimulation parameters that affect the production performance of gas wells in the Marcellus. Those parameters often have complex interrelationship among each other. The operators and engineers are often challenged to decide the optimum value for each parameter in the effort to optimize the overall well performance and get more return for the invested capital. The purpose of this study is to develop advanced regression models to estimate the initial production of gas wells as well as the first 12, and 24 months of cumulative gas production using machine learning techniques. The developed models were also tested with a new dataset of parameters to test the reliability of the models in estimating production based on the completion and stimulation parameters.

The first step in this research was to build a comprehensive database that combines the chemical data, the volume of the pumped fracturing fluids, and proppant mass. Fracfocus website and DrillingInfo for production and completion databases were selected to be the major two sources to be downloaded and merged. Figure 3 illustrates the merging of the databases workflow.

FracFocus chemical disclosure registry provides public disclosure of hydraulic fracturing chemical additives used in hydraulically fractured wells in the United States. As of April 2019, more than 127700 wells are available on the FracFocus website. The website is designed to provide the public with a pdf downloadable document that lists the main well identification headers such as the well name, state, county, API number, operator name, the geographic coordinates, production type, the true vertical depth, and the total volume of water used in the well. The second part of the well document on FracFocus is a list of all the chemical components identified by trade name, supplier name, the purpose of use, chemical ingredients of the component, the concentration of the added component and the maximum ingredient concentration in the hydraulic fracturing fluid (% by mass). To search for any well on FracFocus website, there are two options: search by map or standard search. In standard search you can choose to filter and narrow down the search by selecting state name, county name, operator name, job submission exact date or a date range, API number, and well name. If the well API number or the well name is known, the search will render one pdf file related to the exact well name, if not, then the search will render a list of pdf files for all the wells within the search-specified parameters. Using the document provided by FracFocus is mainly intended for transparency and the public's education. The direct use of these individual documents for the purpose of reflecting trends and performing analysis is not possible without sourcing the database and aggregate the chemical components by its summed-up value for each well.

FracFocus website is hosted by the Ground Water Protection Council (GWPC) and the Interstate Oil and Gas Compact Commission (IOGCC). It is sponsored by the oil and gas industry. It was launched during the second quarter of 2011 and the disclosure by the industry was voluntary. Once the FracFocus was implemented, more companies started to submit the disclosure reports. Several states adopted laws and regulations requiring operators to submit the chemical disclosure of each newly fractured wells to the registry. As of today, 26 states require all operators to report to FracFocus registry. Some states require the operators to report the hydraulic fracturing data to state agencies and the submission to FracFocus is optional.

A proper analysis of FracFocus information is not an easy objective and requires considerably more efforts. As in every big data set and especially in FracFocus, many sources of errors are encountered. The data extracted from FracFocus is subjected to human errors and typos or can happen by simply putting information in the wrong box as a data entry error. There are some limitations to the database, examples of that would be the % mass of the additives are expressed in the maximum concentration and the total % mass of the whole ingredients percentage may exceed 100%, also the total sometimes will be less than 100% due to the absence of non MSDS ingredients which may or may not be listed depending upon state reporting requirements (FracFocus, 2019). One of the challenges in gathering the data of the chemicals for the purpose of identifying the hydraulic fracturing fluids' components, is that for the same chemical you might find several different names that each service company uses differently as well as different trade names, e.g., Ethylene Glycol a chemical used as a non-emulsifier has different names in the registry such as Ethylene alcohol, Tescol, HOCH2CH2OH, Dowtherm SR, Monoethylene glycol, Macrogol 400 BPC, etc. Using multi-name makes the utilization of search by name for the chemical components difficult task. The best way to search and classify the chemicals is to search by the Chemical Registry Abstract Number (CAS-Number) which standardizes the chemicals name by a special number. Using the CAS website can help in bringing more detailed information about the chemical components and serve as a material safety data sheet (MSDS). Employing CAS-Number to extract data should be done carefully by cross-checking the data with other related columns before properly labeling. CAS-Number in the database is not always entered properly and any typos or data entry error would change the assigned group during the data extraction process.

DrillingInfo is a commercial database that provides drilling and completion data in addition to many other services. For this study, a list of wells that were drilled and completed between 2010 and 2018 was download and then matched with the wells list generated from FracFocus. For each well in DrillingInfo, a list of several variables was generated. An example of the data includes the following: well name, API number, lease number, operator name, field name, state and county name, longitude, latitude, spud date, completion date, initial production, second month production, first 6, 12, 24, and 60 months of productions, lateral length, perforated intervals.

To perform this study, more than 2700 horizontal wells from the Marcellus shale play that have been drilled and completed between 2011 and 2018 were gathered and prepared to provide the necessary production, completion, and stimulation parameters to generate production prediction models.

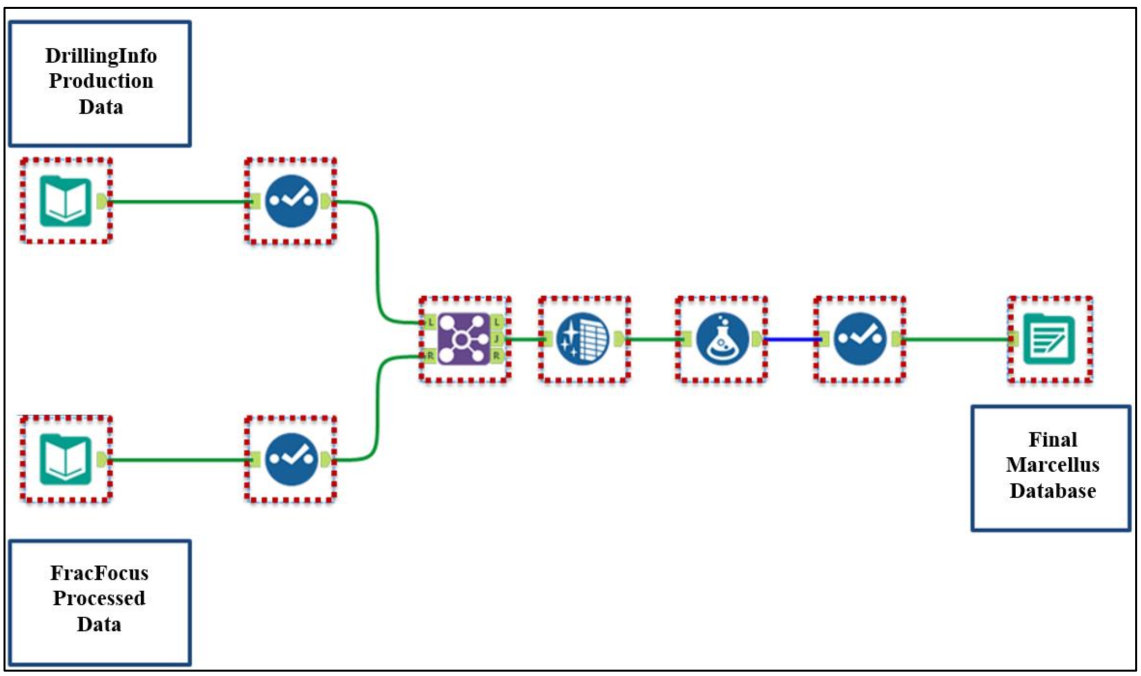


Figure 3. Workflow Combining FracFocus and DrillinInfo Data into One File

6. RESULTS AND DISCUSSION

In this section, three models were developed to predict the initial production of gas (IP), 12 months cumulative gas production (MCF), and 24 months cumulative gas production (MCF) for the Marcellus shale play. The initial production (IP) data were calculated in terms of practical IP (MCF/day) by taking the second month's cumulative production and divides it by the number of produced days in that month in order to get the first full month of production. The predictors were the following parameters: Total Base Water Volume in Gal (W.V), Proppant Mass in lbs (P.M), Lateral Perforated Length in ft (L.L), Biocide Concentration in ppg (Bio), Clay Control Concentration in ppg (C.C), Corrosion Inhibitor Concentration in ppg (C.I), Friction Reducer Concentration in ppg (F.R), Iron Control Concentration in ppg (I.C), Scale Inhibitor Concentration in ppg (S.I), and Surfactants Concentration in ppg (Surf).

6.1. PRACTICAL IP MODEL

Figure 4 shows the root mean of PRESS with the number of latent factors. The number of latent factors that minimize the root mean of PRESS is 5. Figure 5 shows the score plots for the five latent factors. By inspecting Figure 5, all score plots have trends. Thus, all latent factors are necessary and should be included in the model

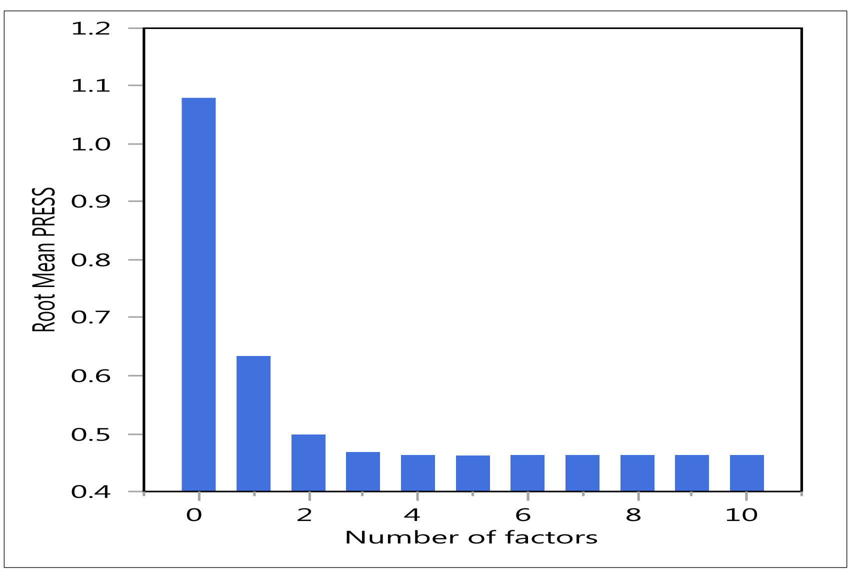


Figure 4. Root Mean of PRESS for All Factors

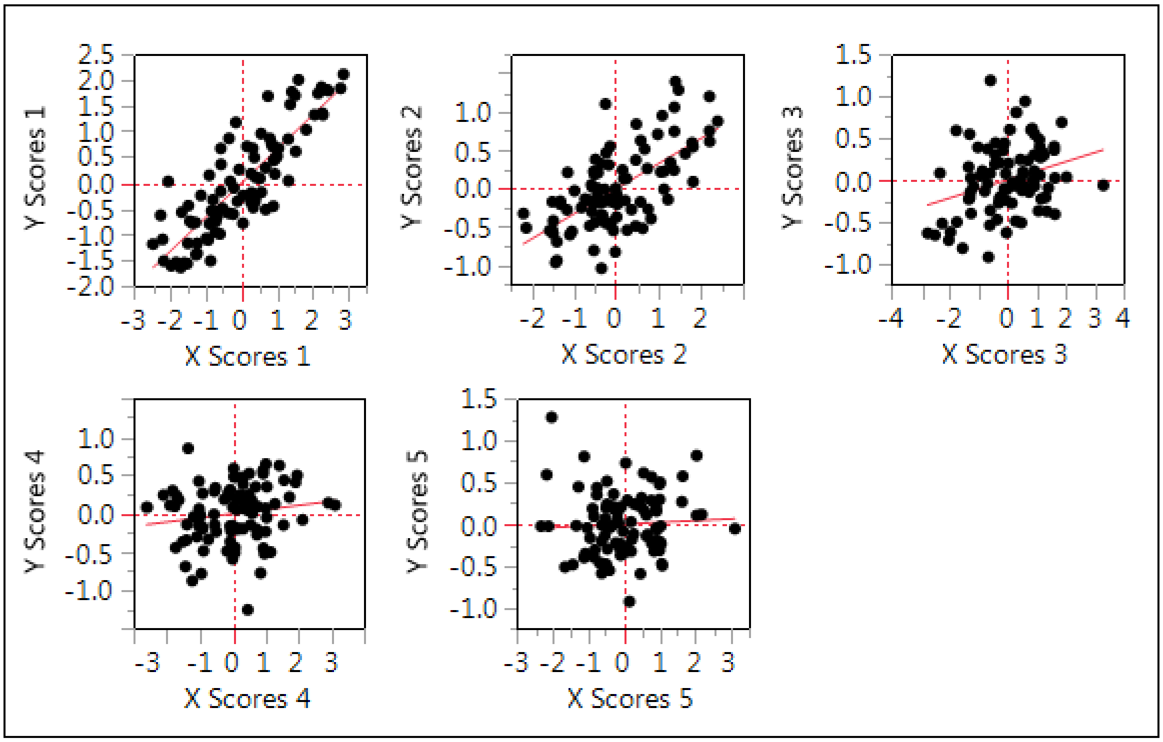


Figure 5. Score Plots of All Latent Factors

Figure 6 and Figure 7 show the percentage of variation explained by each factor for y and X-variables, respectively. As can be seen from Figure 7, all factors are contributing to the explanation of the variability in the X-variables. Also, Figure 7 shows that factor 5 is tremendously contributing to the variability of the total base water volume. Based on Figure 6, factors 1 and 2 contribute about 80 % to the variability of y variable. Thus, factors 3-5 are only contributing to the variability of X-variables.

Figure 8 shows the predicted and actual practical IP. The R-squared of this model is 0.86. Figure 9 shows the residual plot of the practical IP model. The residual plot is a diagnostic plot utilized to validate the created model. The points in the residual plot have to be randomly distributed with no trends, any trend in the residual plot indicates the model is not valid. Since there is no trend in the residual plot, the model is said to be valid. Figure

9 showed no trend is observed in the residual plot. Thus, the model is valid. Eq. (9) can be used to estimate practical IP for the Marcellus shale gas wells.

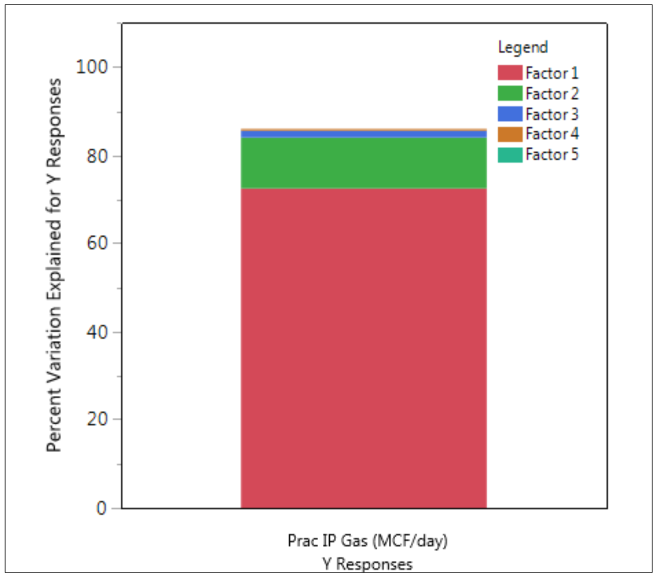


Figure 6. Variation Explained by each Factor for Y Variable

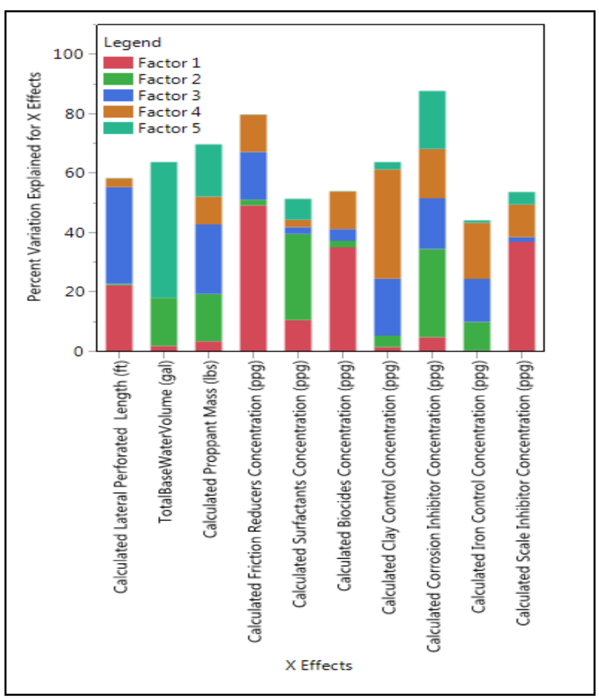


Figure 7. Variation Explained by each Factor for X-Variables

Gas IP (MCF/day) = - 290.9364 - 3539.3907 * Bio - 2413.2584 * C.C + 3429.3187 C.I + 7813.3781* F.R – 2233.7530 * I.C + 0.1583 * L.L + 0.0000882843 * P.M + 2924.7865 * S.I + 1283.9070 * S.C + 0.0000743278 * W.V (9)

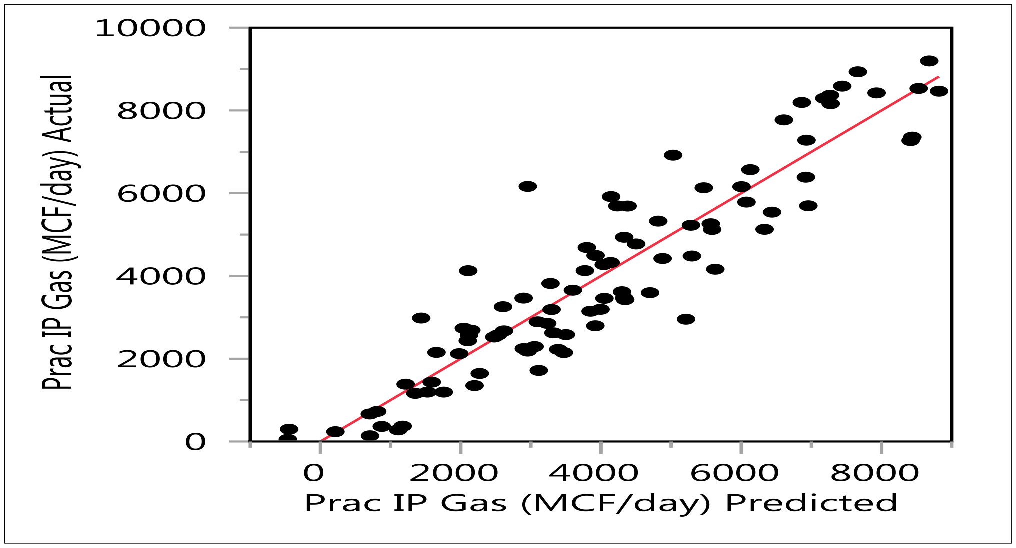


Figure 8. Predicted and Actual Practical IP Model

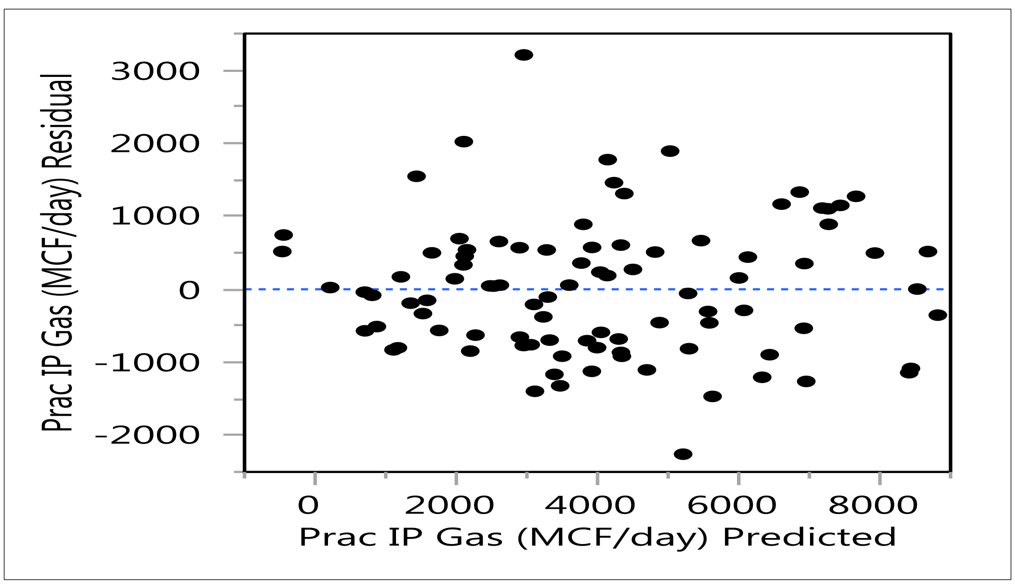


Figure 9. Residual Plot of Practical IP Model

6.2. 12 MONTHS CUMULATIVE PRODUCTION MODEL

Figure 10 shows the root mean of PRESS with the number of latent factors. The number of latent factors that minimize the root mean of PRESS is 2. Figure 11 shows the score plots for the two latent factors. By inspecting Figure 11, all score plots have trends. Thus, all latent factors are necessary and should be included in the model.

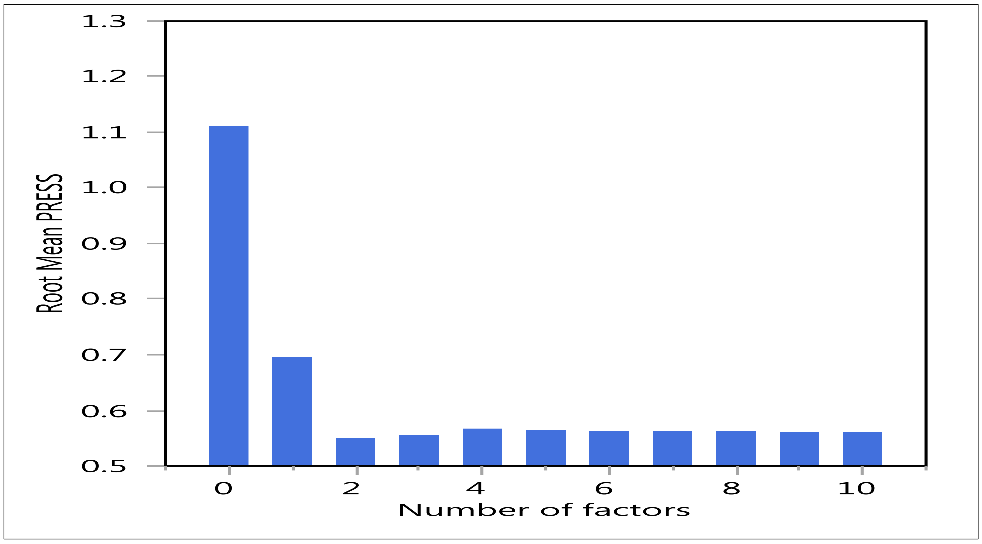


Figure 10. Root Mean of PRESS for All Factors

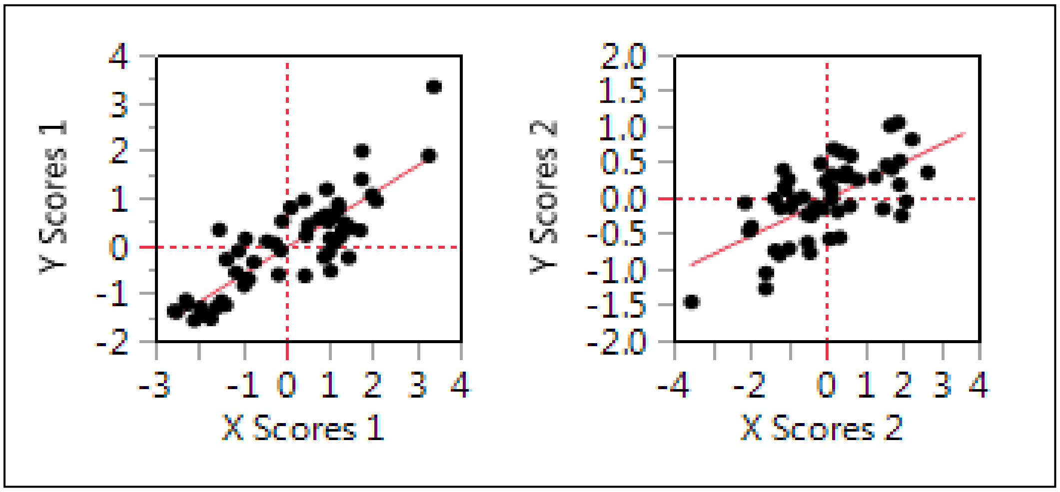


Figure 11. Score Plots of All Latent Factors

Figure 12 and Figure 13 show the percentage of variation explained by each factor for y and X-variables, respectively. As can be seen from those figures, all factors are contributing to the explanation of the variability in the X-variables and y variable.

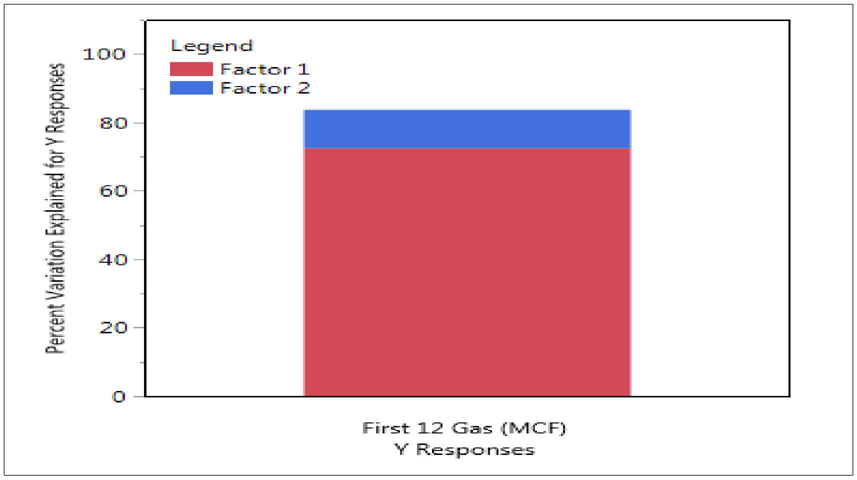


Figure 12. Variation Explained by each Factor for Y Variable

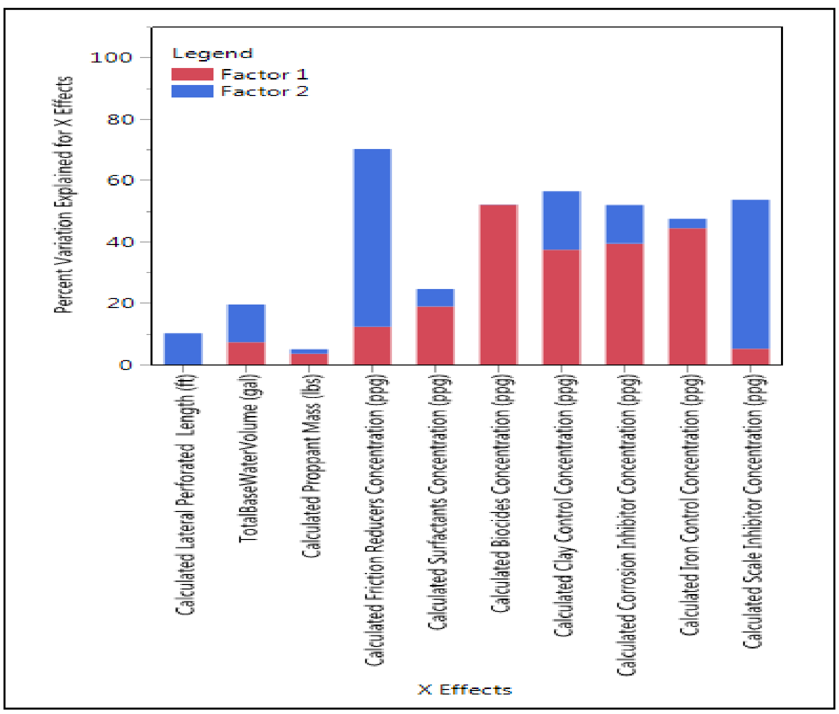


Figure 13. Variation Explained by each Factor for X-Variables

Figure 14 shows the predicted and actual first 12 months of production. The R-squared of this model is 0.84. Figure 15 shows the residual plot of the first 12 months of production, no trend is observed in the residual plot. Thus, the model is valid. Eq. (10) can be used to estimate the first 12 months of production.

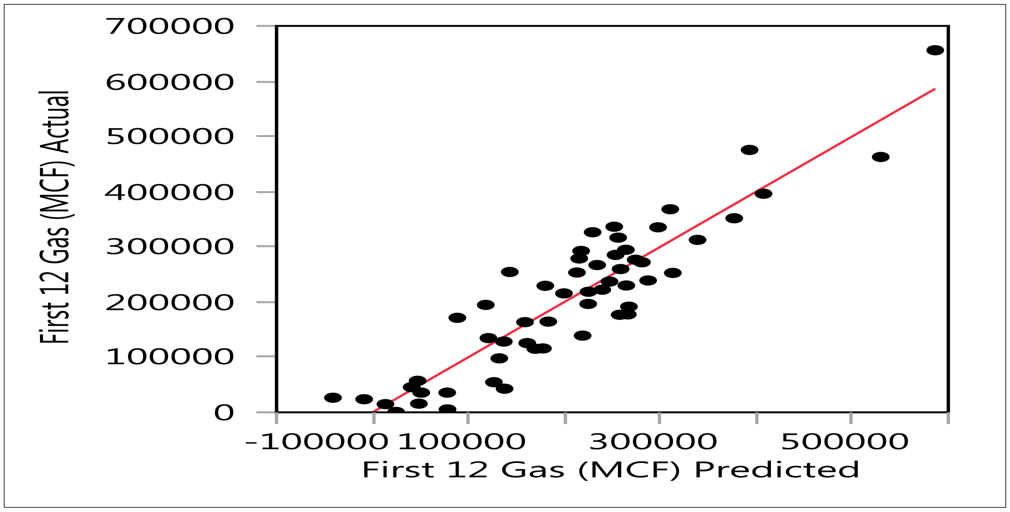


Figure 14. Predicted and Actual First 12 Months Production Model

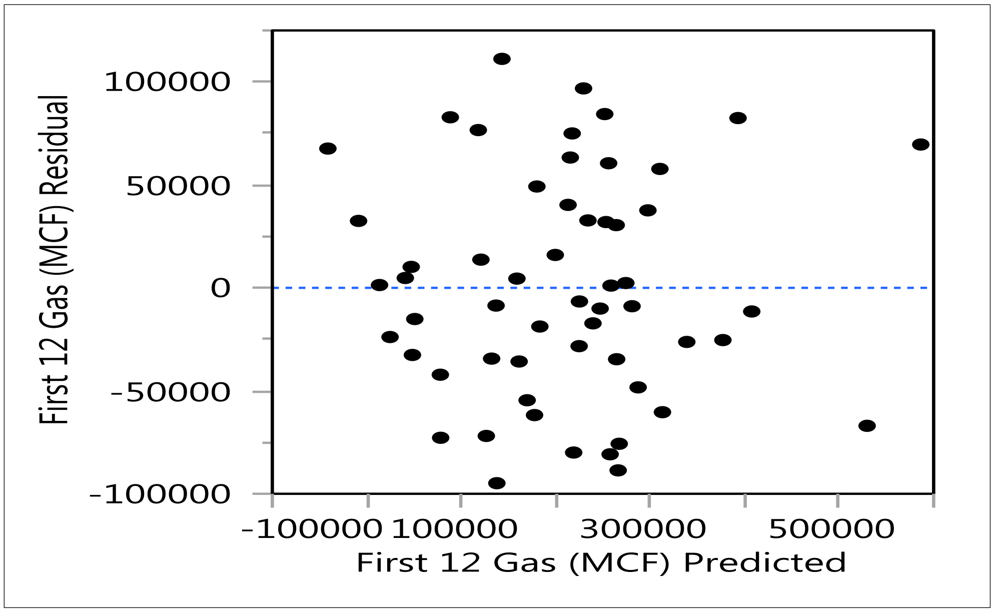


Figure 15. Residual Plot of First 12 Months Production Model

12 Months Cumulative (MCF) = 189227.9753 - 211896.6315 * Bio + 86665.8332 * C.C + 245532.2476 * C.I – 93087.5674 * F.R + 117695.0319 * I.C + 3.9562 * L.L + 0.00232897 * P.M – 69179.3296 * S.I - 107287.0095 * Surf + 0.00966696 W.V (10)

6.3. 24 MONTHS CUMULATIVE PRODUCTION MODEL

Figure 16 shows the root mean of PRESS with the number of latent factors. The number of latent factors that minimize the root mean of PRESS is 4. Figure 17 shows the score plots for all latent factors. By inspecting Figure 17, all score plots have trends. Thus, all latent factors are necessary and should be included in the model.

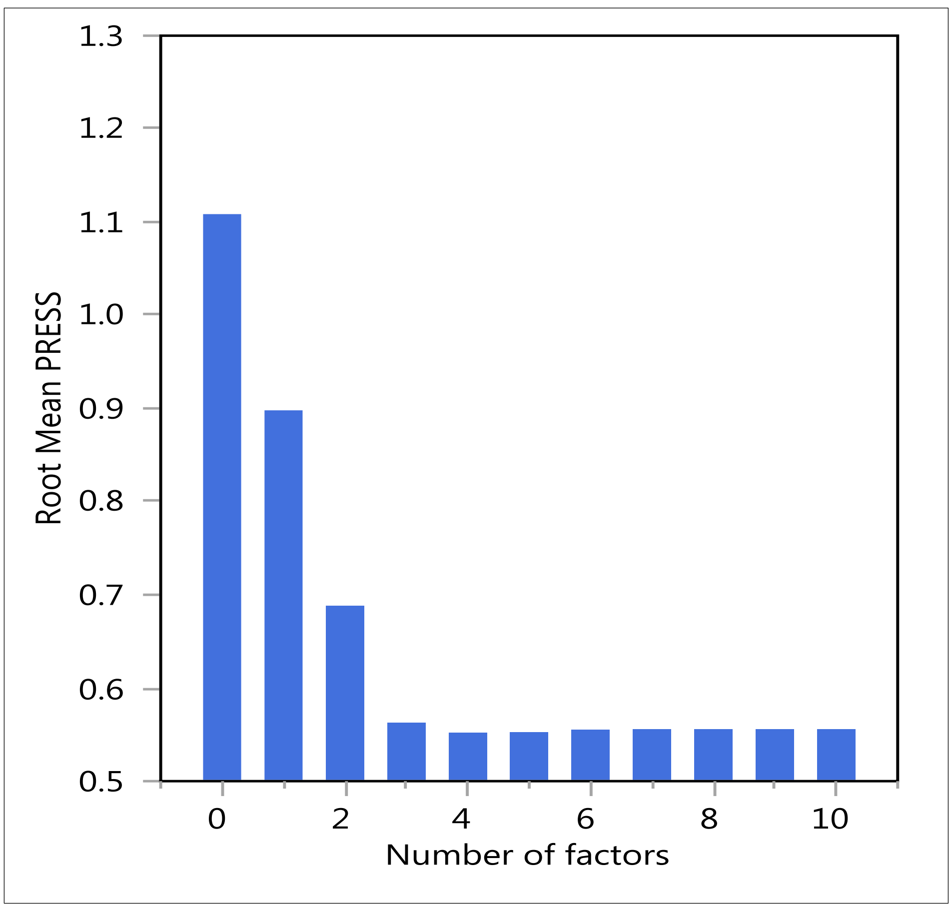


Figure 16. Root Mean of PRESS for All Factors

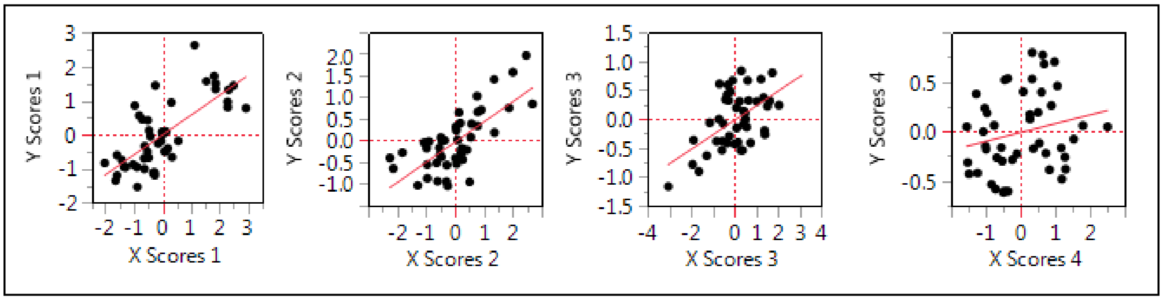


Figure 17. Score Plots of All Latent Factors

Figure 18 and Figure 19 show the percentage of variation explained by each factor for x-variables and y, respectively. As can be seen from Figure 18, all factors are contributing to the explanation of the variability in the X-variables. Also, Figure 18 shows that the factor 4 is tremendously contributing to the variability of the biocide concentration. Based on Figure 19, factors 1, 2, and 3 contribute about 80 % to the variability of y variable. Thus, factors 4 is mainly contributing to the variability of X-variables.

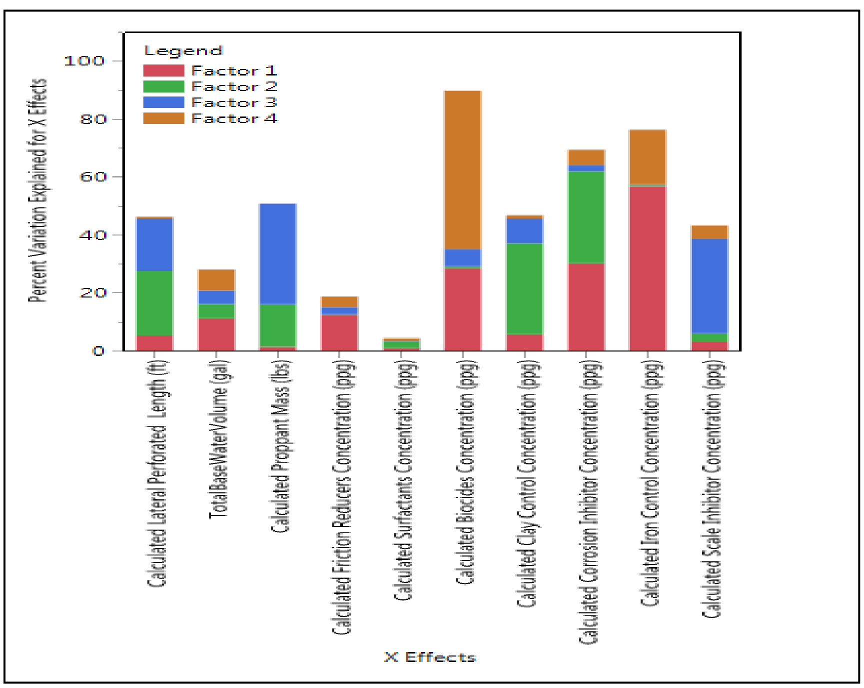


Figure 18. Variation Explained by each Factor for X-Variables

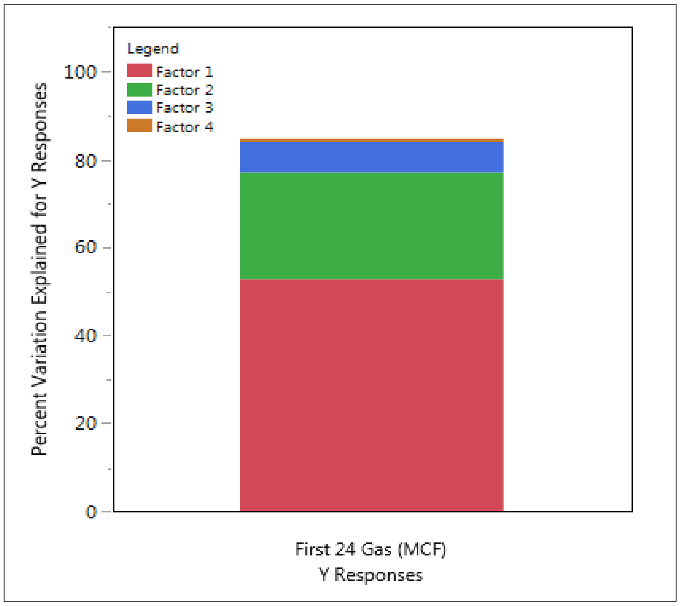


Figure 19. Variation Explained by each Factor for – Variable

Figure 20 shows the predicted and actual first 24 months of production. The R-squared of this model is 0.85. Figure 21 shows the residual plot of the first 24 months of production, no trend is observed in the residual plot. Hence, the model is valid. Eq. (11) can be used to estimate the first 24 months of production.

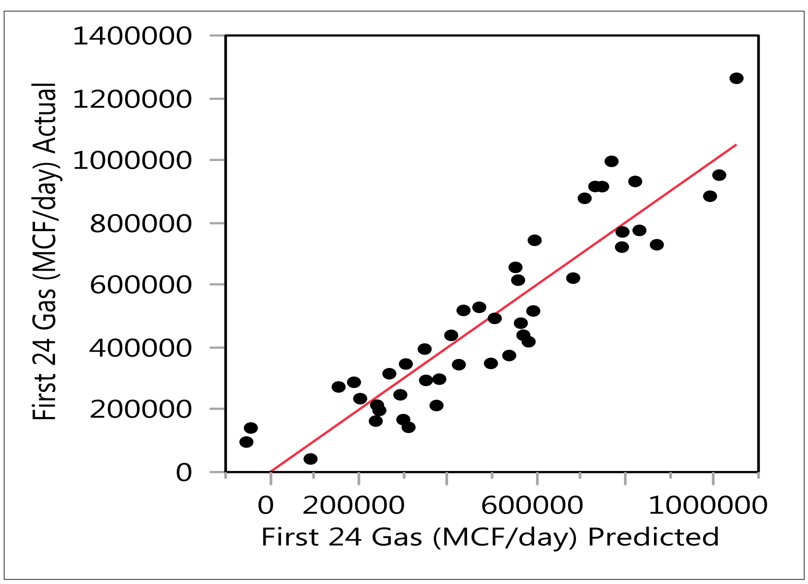


Figure 20. Predicted and Actual 24 Months Production Model

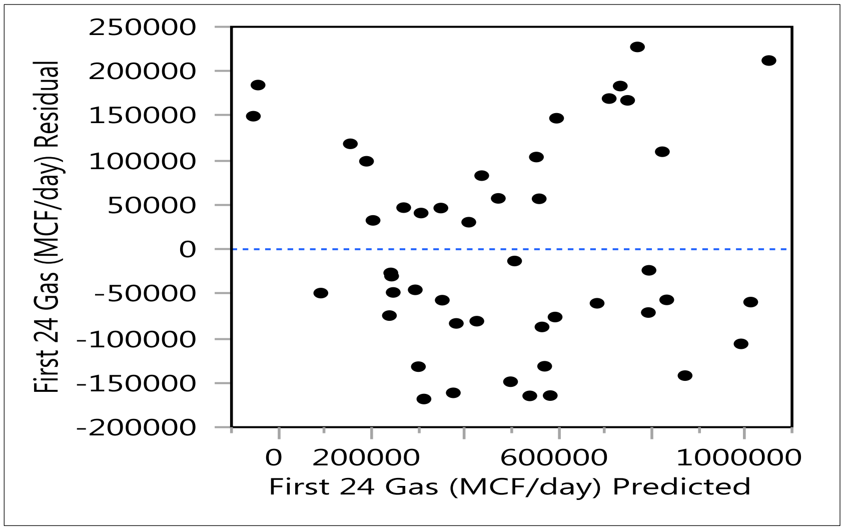


Figure 21. Residual Plot of 24 Months Production Model

$$\textbf{24 Months Cumulative (MCF)} = 843267.9223 - 707348.6493 * Bio - 282003.2427 * C.C - 633185.7206 * C.I - 1033832.082 * F.R. + 870451.6125 * I.C + 9.31861 * L.L + 0.0053218177 * P.M + 134326.32181 * S.I + 76864.074062 * S.I + 0.0094119918 * W.V \quad (11)$$

7. MODELS VERIFICATION AND TESTING

The three models developed in this study were tested on new data of 15 wells (data not used in creating the models). Figures 22, 23, and 24 show the predicted and the actual production of practical IP, 12 months production, 24 months production, respectively. A strong correlation between the actual and the predicted data for all models can be observed (the data closely overlap with the black 45 degrees line). Therefore, the models are valid and can be used to predict practical IP, 12 months production, and 24 months production.

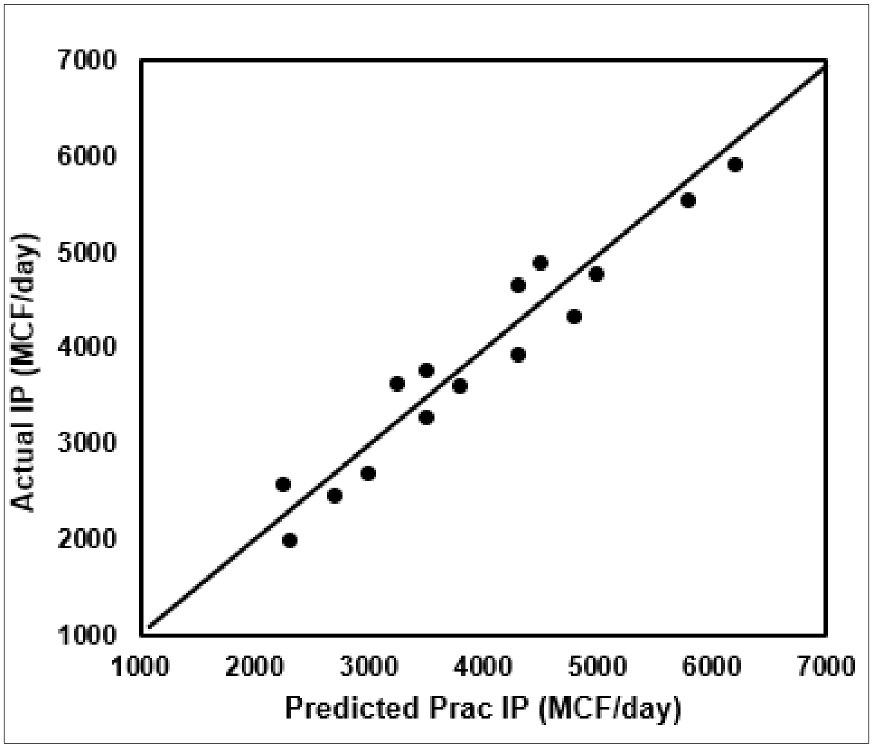


Figure 22. Actual and Predicted IP Model Testing

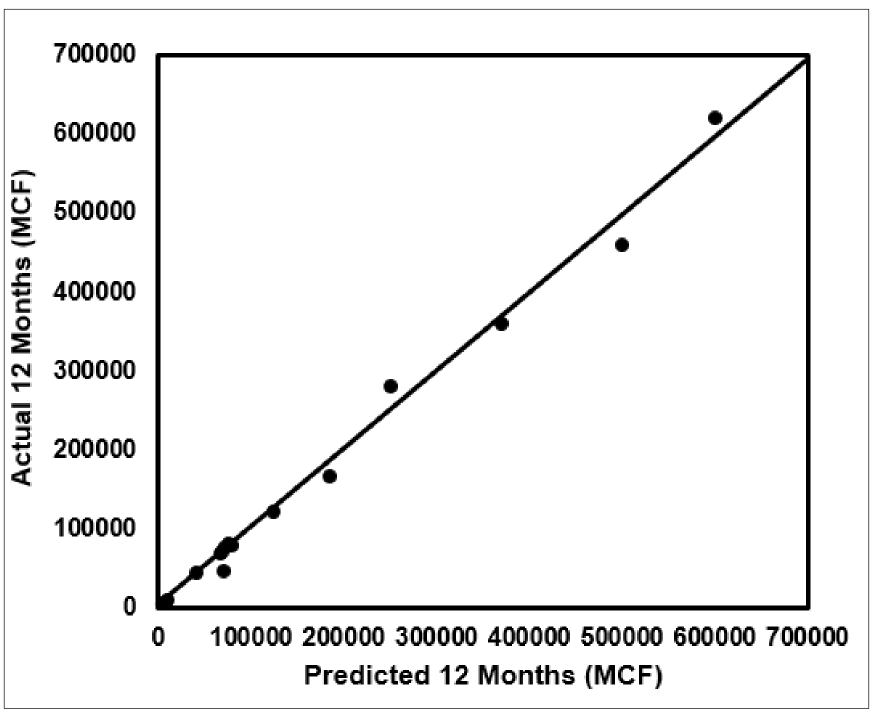


Figure 23. Actual and Predicted 12 Months Model Testing

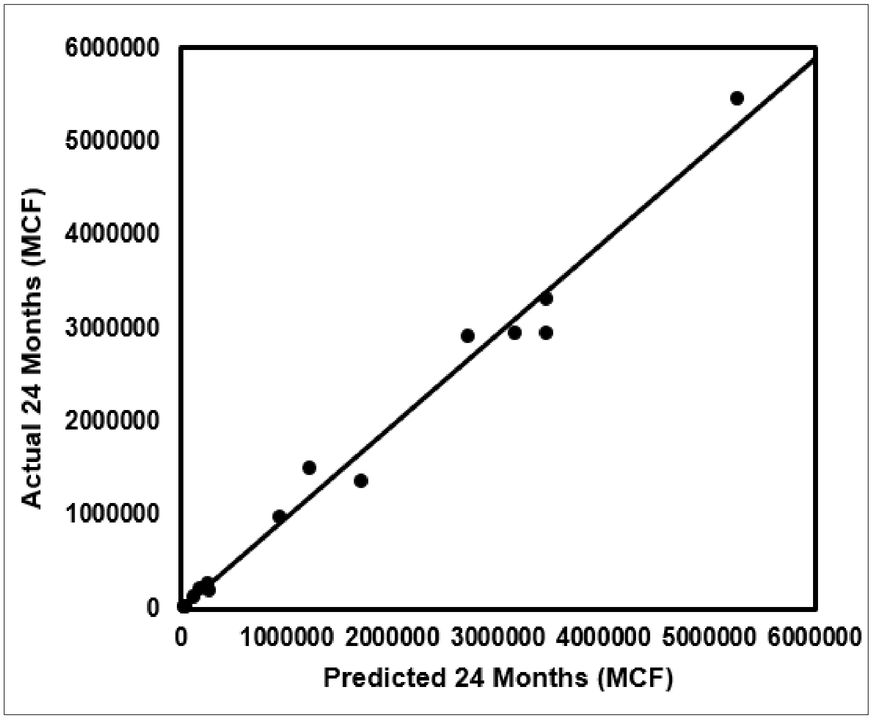


Figure 24. Actual and Predicted 24 Months Model Testing

8. CONCLUSIONS

This paper presents production predictive models by using data-driven statistical analysis of more than 2700 horizontal gas wells in the Marcellus shale. This work demonstrates the application of advanced techniques to develop mathematical models to estimate the IP (MCF/day) of wells, 12 and 24 months of cumulative gas production for wells in the Marcellus.

Those three developed models can be used by operators and completion engineers to estimate the well productivity based on the predictor parameters prior to drill and complete the proposed wells. Based on the desired production optimization and economics, the completion and stimulation parameters can be tweaked to match the optimized production. The main conclusions of this study are the following:

- Three advanced statistical models were developed using PLS approach to predict the initial production rate and the first and second years of cumulative gas production in the Marcellus shale area.
- The PLS approach has never been applied to predict production in the unconventional plays. All of the three developed models showed a very good correlation with the actual data with R2 value ranging from 0.84 to 0.86.
- Using the PLS technique will enhance the prediction of gas production.
- Selecting the optimum number of latent factors is very important part in the PLS method and requires careful inspection of the root mean of PRESS plot, the scores plots, and the percent variations of y and x plots.
- The three models developed in this study were tested on new data of 15 wells and proved their ability to predict IP, 12 months of production, and 24 months of production within an acceptable range of error.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to DrillingInfo team and the co-founder and director of university outreach Mr. Mark Nibbelink for their generosity in providing access to the U.S. production and completion data.

REFERENCES

Al-Dhafeeri, A. M., Nasr-El-Din, H. A., Seright, R. S., & Sydansk, R. D. (2005, January 1). High-Permeability Carbonate Zones (Super-K) in Ghawar Field (Saudi Arabia): Identified, Characterized, and Evaluated for Gel Treatments. Society of Petroleum Engineers. doi:10.2118/97542-MS

Aldhaheri, M. N., Wei, M., & Bai, B. (2016, April 11). Comprehensive Guidelines for the Application of In-Situ Polymer Gels for Injection Well Conformance Improvement Based on Field Projects. Society of Petroleum Engineers. doi:10.2118/179575-MS

Al-Hameedi, Abo Taleb & Alkinani, Husam & Dunn-Norman, Shari & E. Flori, Ralph & A. Hilgedick, Steven & S. Amer, Ahmed & Alsaba, Mortadha. (2018). Mud loss estimation using machine learning approach. Journal of Petroleum Exploration and Production Technology. 10.1007/s13202-018-0581-x

Alvarado, V., Thyne, G., & Murrell, G. (2008, January 1). Screening Strategy for Chemical Enhanced Oil Recovery in Wyoming Basins. Society of Petroleum Engineers. doi:10.2118/115940-MS

Baker, R. O., Stephenson, T., Lok, C., Radovic, P., Jobling, R., & McBurney, C. (2012, January 1). Analysis of Flow and the Presence of Fractures and Hot Streaks in Waterflood Field Cases. Society of Petroleum Engineers. doi:10.2118/161177-MS

Bhattacharya, S., Maucec, M., Yarus, J. M., Fulton, D. D., Orth, J. M., & Singh, A. P. (2013, September 30). Causal Analysis and Data Mining of Well Stimulation Data Using Classification and Regression Tree with Enhancements. Society of Petroleum Engineers. doi:10.2118/166472-MS

Eriksson, L., Johansson, E., Kettaneh-Wold, S., Trygg, J., Wikstr¨om, C., & Wold, S. (2006). Multi- and megavariate data analysis. Part I. Basic principles and applications. Umetrics Academy, 2006

FracFocus, https://fracfocus.org/, retrieved Jan, 24th 2019

Fulford, D. S., Bowie, B., Berry, M. E., Bowen, B., & Turk, D. W. (2015, September 28). Machine Learning as a Reliable Technology for Evaluating Time-Rate Performance of Unconventional Wells. Society of Petroleum Engineers. doi:10.2118/174784-MS

Gong, X., Gonzalez, R. A., McVay, D., & Hart, J. D. (2011, January 1). Bayesian Probabilistic Decline Curve Analysis Quantifies Shale Gas Reserves Uncertainty. Society of Petroleum Engineers. doi:10.2118/147588-MS

Gupta, S., Fuehrer, F., & Jeyachandra, B. C. (2014, September 30). Production Forecasting in Unconventional Resources using Data Mining and Time Series Analysis. Society of Petroleum Engineers. doi:10.2118/171588-MS

Harris, C. 2014. Potential Pitfalls in Exploration and Production Application of Machine Learning, Presented at the SPE Western Northern American and Rocky Mountain Joint Regional Meeting, Denver, Colorado 16-18 April, 2014 SPE169523-MS

Hastie, T., Tibshirani, R., and Friedman, J.. 2008. The Elements of Statistical Learning, Springer, 2nd edition

Kuhn, M. and Johnson, K. 2013. Applied Predictive Modeling. Springer

LaFollette, R. F., & Holcomb, W. D. (2011, January 1). Practical Data Mining: Lessons-Learned From the Barnett Shale of North Texas. Society of Petroleum Engineers. doi:10.2118/140524-MS

Lolon, E., Hamidieh, K., Weijers, L., Mayerhofer, M., Melcher, H., & Oduba, O. (2016, February 1). Evaluating the Relationship Between Well Parameters and Production Using Multivariate Statistical Models: A Middle Bakken and Three Forks Case History. Society of Petroleum Engineers. doi:10.2118/179171-MS

Schuetter, J., Mishra, S., Zhong, M., & LaFollette, R. (2015, July 20). Data Analytics for Production Optimization in Unconventional Reservoirs. Unconventional Resources Technology Conference. doi:10.15530/URTEC-2015-2167005

Hegde, C. M., Wallace, S. P., & Gray, K. E. (2015a, September 15). Use of Regression and Bootstrapping in Drilling Inference and Prediction. Society of Petroleum Engineers. doi:10.2118/176791-MS

Hegde, C., Wallace, S., & Gray, K. (2015b, September 15). Using Trees, Bagging, and Random Forests to Predict Rate of Penetration During Drilling. Society of Petroleum Engineers. doi:10.2118/176792

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. "An Introduction to Statistical Learning." With Applications in R | Gareth James | Springer. Springer-Verlag New York

S. Mohaghegh, "Shale asset management: advanced data driven and predictive analytics", SPE Webinar, November 2013

Leite Cristofaro, R. A., Longhin, G. A., Waldmann, A. A., de Sá, C. H. M., Vadinal, R. B., Gonzaga, K. A., & Martins, A. L. (2017, October 24). Artificial Intelligence Strategy Minimizes Lost Circulation Non-Productive Time in Brazilian Deep-Water Pre-Salt. Offshore Technology Conference. doi:10.4043/28034-MS

SAS Institute Inc. 2008. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc

Soodabeh Esmaili, Shahab D. Mohaghegh, Full field reservoir modeling of shale assets using advanced data-driven analytics, Geoscience Frontiers, Volume 7, Issue 1, 2016, Pages 11-20, ISSN 1674-9871

Stéphane Tufféry, S. 2011. Data Mining and Statistics for Decision Making. Chichester, West Sussex: Wiley

T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning: Data Moning, Inference, and Prediction, New York: Springer, 2001

Thanh N. Tran, Nelson Lee Afanador, Lutgarde M.C. Buydens, Lionel Blanchet, Interpretation of variable importance in Partial Least Squares with Significance Multivariate Correlation (sMC), Chemometrics and Intelligent Laboratory Systems, Volume 138, 2014, Pages 153-160, ISSN 0169-7439, https://doi.org/10.1016/j.chemolab.2014.08.005

Wallace, S. P., Hegde, C. M., & Gray, K. E. (2015, September 15). A System for Real-Time Drilling Performance Optimization and Automation Based on Statistical Learning Methods. Society of Petroleum Engineers. doi:10.2118/176804-MS

Wilson, A. (2015, October 1). Do Data-Mining Methods Matter? A Wolfcamp Shale Case Study. Society of Petroleum Engineers. doi:10.2118/1015-0087-JPT

Wold, S., Johansson, E., Cocchi, M. PLS-partial least squares projections to latent structures in 3D-QSAR in drug design; theory methods and applications (1993) PLS-Partial Least Squares Projections to Latent Structures, pp. 523-550

III. REVIEW OF STIMULATION AND COMPLETION ACTIVITIES AND TRENDS IN THE UNITED STATES SHALE PLAYS: PERMIAN BASIN CASE STUDY

Mustafa A. Al-Alwani, Larry K Britt, Shari Dunn-Norman, Husam H. Alkinani, Abo Taleb T. Al-Hameedi, Atheer M. Al-Attar

Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA

ABSTRACT

The hydraulic fracturing designs, drilling, and completion trends are undergoing continuous change as the oil and gas industry pushing the boundaries of the traditional designs to increase wells productivity and meet the energy demand. The lateral length in addition to the amount of proppant and water used in well stimulation have witnessed a paradigm shift over the past few years. This paper represents a large-scale descriptive analysis of the stimulation and completion trends in the Permian Basin between the first quarter of 2013 and the second quarter of 2018. The aim of this analysis is to shed a light on the trends in completion and stimulation and to show the statistical values of each parameter over the studied period. The data sources have been compiled from two sources. FracFocus is the first source for stimulation parameters. It is publically accessible but needs rigorous data processing to render useable parameters to be integrated in this study. The second data source is DrillingInfo which is a commercial database. The completion and production data in this study were obtained from there and integrated with the FracFocus data. The compiled data base for the Permian Basin has more than 15,000 wells covers the Midland, Delaware, and the Central Basin Platform. Analyzing the trends in the presented visualizations elucidate an obvious growing trend over time in the amount of stimulation

proppant and water. The overall trends of stimulation and completion in the Permian Basin are presented here in addition to the trends in pad drilling and how the density of the wells in a single pad site has been increasing over the past few years. Large data set utilization will demonstrate a sensible and comprehensive representation of the trends in the studied area. This work will give the readers an established baseline and will serve as a reference point to many of the completion and stimulation parameters by comparing the statistical values such as the median, mean, and standard deviation of each parameter of interest during the analyzed period.

1. INTRODUCTION

The last decade advancement in horizontal drilling applications coupled with the hydraulic fracturing have led to tremendous growth in utilizing the unconventional resources as a vital source of oil and gas in the United States. As a result, to the boom in drilling and stimulation in the unconventional resources, large amount of drilling, completion, and stimulation data have accumulated over the last years. Gathering and analyzing data such as reservoir properties, drilling, completion, and stimulation parameters will help in optimizing the wells' productivity to achieve more economical benefits from those plays.

Many industries have been utilizing descriptive and predictive data analytics as part of their processes optimization, future planning, and making decisions. In the oil and gas sector, data analytics have been described as one of the most dynamically growing field in this sector (Khvostichenko and Makarychev-Mikhailov, 2018). After realizing the influence and the importance of the data-driven decisions, many of the big operators and leading service companies are allocating many of their resources and dedicating specialized teams to the trending area of data science. Despite the fact that large scale data analytics are relatively new, the literature shows the growing interest in this type of studies and the outputs were very insightful (Yang et al., 2013; Arthur et al., 2014; Romero and Poston, 2016; Myers et al., 2017; Luo and Zhang, 2018; Olaoye and Zakhour, 2018; Srinivasan et al., 2018; Carman and Wheeler, 2018; Weijers et al., 2019).

In the oil and gas industry, there are many pitfalls and challenges to overcome in order to execute a data science project. The first and eternal challenge is the availability of comprehensive data. For many years in the past, the availability of drilling, completion, and stimulation data were limited to the operating and service companies and data sharing culture between the companies was not common. The availability of wide range data is an essential part for the success of any data-driven analysis. Missing parts in the dataset will hinder the studies and might result in biased decisions. In order to see the big picture in any data analytics, the availability of a comprehensive database that combines collaborative data from all the operators and service providers in the studied area is deemed necessary. When the data are withheld by each operator and were not shared across the board, each operator will end up possessing only one piece of the puzzle and less efficient analyses might be obtained due to the limitation of the data availability. The second challenge in any data analytics is the quality and ingestion form of the data, for example, the accuracy of the reported data and the type and format of the saved data whether the information is saved on scanned paper reports, uneditable electronic file, or saved in an easy to copy and transform data format. The third obstacle to successful data analytics is large scale data handling and processing tools and software availability. The good old way of spreadsheet calculations might reach its limit before covering the large data constituents. Microsoft Excel, in particular, cannot accept and process in a single spreadsheet more than 1,048,576 rows of data. Data handling software limitations may limit the capabilities of utilizing large datasets such as FracFocus where it contains several million rows of stimulation chemicals' data. Therefore, the need for more advanced processing and visualization software becomes a necessity to perform such studies.

In this study, all the available FracFocus raw data between 2011 and 2018 were downloaded, processed, and then combined with data from another commercial database to integrate the United States completion and production data to the stimulation data. The created database will serve as a source for several analytical studies to identify the completion and stimulation activities and to highlight the trends in different parts of the United States.

Over the last few years, the unconventional resources in the United States have gained the industry focus and many operators and service providers began to implement different drilling and completion strategies. They have been pushing the design and operating boundaries to find the best practices to optimize and maximize wells' productivity.

Collecting data about drilling, completion, and stimulation parameters over the last years will provide a good base to analyze the industry's general trends and approaches. It will show how the strategies have been evolving along the journey of experimentation and refinement to reach the target of optimizing the productivity of the wells while reducing the footprint as reflected by the adoption of multiple wells pads instead of single well pads.

The drilling technologies advancement has led to longer horizontal laterals which demand a higher number of stimulated stages and a need for more proppant and water quantities to achieve more effective stimulated reservoir volume. Historically, the operators in the unconventional resources during the early development stages of their drilling and completion programs have started their designs with basic engineered procedures while trying to establish a baseline for their treatment. The early designs were limited by logistics, equipment rating and availability, material cost, and the targeted or desired initial production. Other operators simply copied and adopted the successful designs from their neighboring operators. Starting from the baseline designs, there have been many development and steps that the pioneer operators had to go through to reach to the best and optimized designs that match their productivity expectations. The developments were not limited to only the completion and stimulation operational designs, but it also included the products types as well, such as proppant, additives and fluids.

Many publications in the literature tried to document the history and the trends of the shale best practices. The early studies did not contain representative formation descriptive data that can be utilized to generalize and extrapolate trends to the field level. Early studies were usually conducted on data from a limited number of localized wells and reflected a very area specific trend (Carman and Wheeler, 2018). King (2010) enriched the literature with a published study that included a comprehensive review of the trends of the shale plays best practices. At that time, that study was the best summary study and it is still considered as a very valuable reference. As more data are shared, and more comprehensive databases are built, presenting historical case studies can be very helpful in tracking the industry progression towards the best and optimum completion strategies. Having high-quality data of the monumental number of parameters that describes the formation, drilling and completion will serve the purpose of finding meaningful trends and correlations by using data visualization and descriptive analytics techniques.

This study is investigating the general trends in the completion and stimulation in the Permian Basin over the time frame of Q1 of 2013 to Q2 of 2018. The trends of the proppant and water consumption over time are presented along with the parameter's distribution and statistical values. The normalized values of stimulation parameters by the perforated lateral length is also discussed. Pad drilling approach in the Permian is also addressed in this study. Finally, the effects of the amount of the proppant and water on the initial production are discussed in the results and discussion section to show the effects of the completion parameters on well productivity.

2. METHODOLOGY

The first step in this research was to build a comprehensive database that contains all the chemical data including the volume of pumped fracturing fluids and proppant mass. The Fracfocus website and DrillingInfo for production and completion commercial database were selected to be the two major databases to be downloaded, processed and then merged together.

The FracFocus website (FrackFocus, 2019) is a public source to report the types and percentages of all fluids and chemicals that are pumped as part of a hydraulic fracturing stimulation operation. The data of each well is reported separately on a pdf file. The website allows searching for the wells by state, county, etc. down to the unique API number. The website is managed by the Ground Water Protection Council and Interstate Oil and Gas Compact Commission. Since 2011, more than 127,000 well sites have been registered into the website as of April 2019. Due to the limitation of the maximum number of rows that each excel sheet can handle, the FracFocus database was downloaded in several separate excel spreadsheets. The database contains the well identifiers information such as the longitude and latitude, the state and county, the water volume used, and the mass percentage of most of the chemical additives including the sand which is used as a propping agent. The data quality obtained from the raw parameters of FracFocus is challenging and requires a lot of processing and cleansing procedures to eliminate the typos and errors. Data processing algorithms were developed in this study and many data workflows were integrated to eliminate errors and repeated fields to come up with an accurate calculated mass value for each chemical ingredient and to group the additives based on their intended purpose of use. The data workflows were also developed to detect re-fractured wells and many built-in validation processes were integrated to make sure that the outcomes of the data were accurate. The concentration in pound/gallon of all the chemical additives was also determined in the newly created database. More than eighty thousand wells with validated data quality were extracted from FracFocus. Those wells spanning the whole U.S. Figure 1 shows a data workflow example.

Production and completion data were extracted from a commercial database source (DrillingInfo, 2019). It compiles and verifies data retrieved from the States' public records and third-party sources. Since the data quality is already verified by DrillingInfo (D.I.), routine data quality check and outlier detection procedures were applied to D.I. parameters. Practical initial production was calculated as the second-month production divided by the number of produced days (BBL/Day, MCF/Day), the perforated lateral length was used as a proxy to stage count as well as a normalizing parameter. It was calculated by subtracting the upper perforation depth from the lower perforation depth. Proppant density (Proppant Loading) was calculated by dividing the total amount of the pumped proppant mass (lbs) by the total volume of used water (gal).

3. DATA DISTRIBUTION OF THE WELLS IN THE PERMIAN BASIN

Having data accessibility of many wells as presented in this analysis enhances the quality of the trends investigation and gives many insights for the overall use of proppant and water volumes across the Permian Basin. It can also provide insights into the progression of trends within the basin and can be filtered and drilled down to show different levels of details.

Figure 2 shows the map of Texas and New Mexico with all the wells integrated into this study. The wells' markers were color-coded based on the type of drilling (Green: Vertical, Blue: Directional, and Red: Horizontal). The shapes of the markers were also identified based on the well's drilling type (Star: Directional, Horizontal Arrow: Horizontal, and Vertical Arrow: Vertical).

Figure 3 shows the number of wells for the three major geological components in the Permian and the overall percentage of data coverage for each basin. The pie chart sectors are sized based on the number of wells in each basin and all the sectors were color coded based on the geological basins.

4. RESULTS AND DISCUSSION

Completion trends across the United States have been showing a continuous upswing in the amount of proppant and fluid used in well stimulation. In previous publications, Data from all the major U.S. unconventional basins were utilized to visualize the growing trends over time. In the aggregated U.S. plays trends paper the completion trends in all the formations showed an upswing correlation in term of the amount of proppant and water as well as the stimulated lateral length. In this study, the focus will be directed to the Permian Basin activities over the studied period. It covers all the wells drilled and completed in the Midland Basin (8866 wells), Delaware Basin (4370 wells), and the Central Basin Platform (2280 wells). The following subsections will discuss the trends in stimulation and completion parameters.

The following sub-sections will discuss the Permian Basin stimulation and completion parameters:

4.1. PROPPANT QUANTITIES AND TRENDS

In hydraulic fracturing stimulation, a propping agent (such as sand or ceramics) which is usually called proppant are used to keep the fractures open. Proppant is usually mixed with water and other different chemicals to hydraulically fracture and stimulate the low and ultra-low permeability formations. It has been used in hydraulic fracturing for about 70 years. There has been a continuous growing demand on proppant which can be attributed to the relatively recent advancement in horizontal drilling and the ability to drill longer laterals which requires multiple fracturing stages. The continuous growing demand for energy sources and the escalation of oil and gas prices have pushed the operational limits of drilling, completion, and stimulation to expand and hence proppant demand.

Figure 4 shows combined box plots and distribution histograms of the proppant data for the Permian Basin hydraulic fracturing stimulation proppant utilization over time. The plots are also associated with a statistical table that shows the count of the participating wells; average value of the proppant mass pumped for each particular year; median value of the proppant which are used to describe the variables with wide range of values to eliminate any bias results on the average value due to the high value of the suspected outliers. In many cases the median and the average are close in values which indicates a good data quality and there are no extreme outliers and skewness that influence the average value; summation which represents the total mass of proppant consumed for each particular year; maximum and minimum values to show the range of the data, standard deviation; the first and third quartile values of the data distribution; interquartile range; and the number of suspected outliers. Each figure was color-coded based on the stimulation year. Comparing the median and the average values of the proppant mass over the years shows an undeniable trend that the operators have been utilizing more proppant as time progressed.

The years 2011 and 2012 had the least number of wells because of the fact that FracFocus was released during the second quarter of 2011 and reporting the stimulation data to the website was voluntarily, after 2013 more operators started to participate in data reporting and many states started to require all operators to report the stimulation data as part of the regulations. In this analysis, data from the years 2011 and 2012 were eliminated because of the small sample representation. The year 2018 only includes the wells that were stimulated during the first two quarters. This figure illustrates that the average proppant mass started around 1.8 million pounds in 2013 and as the time progressed and the lateral length became longer, more proppant was pumped surging over 13 million pounds during 2017 and 2018. The median values of proppant mass in the Permian Basin corroborated with average values and confirmed that the data is clean from extreme outliers or skewness. To account for the increase in lateral length over time, box plots and distribution histograms were created for the normalized value of proppant mass over the perforated lateral length (Lbs/ft) as shown in Figure 5 It is still evident that the median and average values of the amount of proppant pumped per perforated foot of the lateral were increasing with time and the data distribution range started to get wider during 2015 onward. As time progressed, the operators started to push the proppant usage range further by trying to pump more proppant per foot as evident in the data distribution histograms. More wells started to appear on the density distribution bars (histograms) at the high proppant/foot ranges. The suspected outliers were kept in these figures to show the high ends trials that some operators conducted to test the operational limits. Only the extreme outliers which reflect no applicable values were removed in this analysis.

Figure 6 shows the proppant average and median values in the Permian wells for each quarter of the studied period. The red bars represent the average value of proppant for each quarter, the black line series represents the median value of proppant for each quarter, and the green line series represents the number of wells that were completed and stimulated during each quarter. The overall trend shows a continuous increase in the amount of proppant being pumped as part of the hydraulic fracturing stimulation. It also shows that the average and median of the amount of sand started to escalate as the number of the drilled wells started to go down associated with the crash of the oil and gas prices during the last quarter of 2014. The average sand per well has increased from 1.8 million pounds per well in 2013 to almost 14 million pounds per well during early 2018.

To show the normalized average and median proppant per perforated foot in the Permian Figure 7 was generated. The trend still confirms the continuous increase in the amount of proppant being pumped regardless of the lateral length. The amount of the proppant per stimulated lateral foot more than tripled from almost 500 lb/ft during early 2013 to more than 1800 lb/ft during early 2018.

Figure 8 shows in the red bar chart the average value of proppant per well sub-classified based on the type of the drilled wells (D: Deviated or Directional, H: Horizontal, and V: Vertical). The black line series represents the number of stimulated wells for each type of wells. The figure illustrates the continuous increase in the average amount of proppant being pumped. It also demonstrates that horizontal wells consume the highest amount of proppant when compared to the other types of wells as a result of multi-stage fracturing operations within the same horizontal well. It is also evident that in the Permian Basin vertical wells drilling used to dominate other types of drilling during 2013 and 2014. In 2015 the operators started to switch their focus to drill most of the wells horizontally. The figure also captures the fact that over the past five years the average proppant per horizontal well has tripled from 5 million pounds per well in 2013 to 15 million pounds per well in 2018.

Figure 9 shows the average proppant consumption per well for each of the three major basins in the Permian. It also shows that the Central Basin Platform has the minimum average proppant consumption per well due to the fact that more than 70 percent of the vertical wells were drilled within this basin. Delaware and Midland basins average proppant consumption per well is relatively close in value with the latter being slightly higher since the second quarter of 2015.

4.2. WATER QUANTITIES AND TRENDS

Water is used in hydraulic fracturing to serve as the primary carrier base fluid. Water and proppant (sand) can make up 99.5 % of the hydraulic fracturing fluid. Water is sourced from lakes, rivers, municipal supplies, and groundwater when sufficient quantities are available.

Figure 10 shows combined box plots and distribution histograms for the volume of water utilized in hydraulic fracturing operations in the Permian Basin. Similar to proppant trends, the water quantities have been increasing with time since 2013. The water volume has increased from an average of 2 million gallons per treated well to almost an average of 13 million gallons per stimulated well. It is also evident from the distribution histograms that over the past three years, the water data distribution became wider as more wells began to utilize larger stimulation volumes due to the increase in drilled lateral length and more hydraulic fracturing stages have been implemented within a single well. The median and the average values are close in their values especially for the last three years which indicate a good data quality and the data is clean from unrealistic outliers which could introduce some biases in the average values.

To investigate the water volume in the Permian Basin and to isolate the effect of the lateral length difference from well to well, a normalized water volume per lateral length parameter was generated. Figure 11 represents distribution histograms combined with box plots over time to show the values of water gallons usage per each stimulated lateral foot.

The increasing trend in water usage was confirmed to be evident regardless of the lateral length. The volume of water per foot has increased from 600 Gal/ft in 2013 to approximately 1800 gal/ft in 2018.

Figure 12 shows the water average and median values in the Permian wells for each quarter of the studied period. The blue bars represent the average value of water (million Gal) for each quarter, the black line series represents the median value of proppant for each quarter, and the green line series represents the number of wells that were completed and stimulated during each quarter. The overall trend shows a continuous increase in the volume of water being pumped as part of the hydraulic fracturing stimulation. Similar to proppant, the average and median of stimulation water volume started to escalate as the count of the drilled wells started to go down as a result to the crash of oil and gas prices during the last quarter of 2014. The average water per well has increased from two million gallons per well in 2013 to almost fifteen million gallons per well during Q2 of 2018. To show the average and median of normalized water per perforated foot in the Permian Figure 4.10 was generated. The trend still confirms the continuous increase in pumped stimulation water volume regardless of the lateral length. The amount of water gallons per stimulated lateral foot have more than tripled from almost 600 Gal/ft during early 2013 to more than 1800 lb/ft during 2018.

Figure 13 shows in the blue bar chart the average value of water volume per well sub-classified based on the type of the drilled wells (D: Deviated or Directional, H: Horizontal, and V: Vertical). The black line series represents the number of stimulated wells for each type of wells. The figure illustrates the continuous increase in the average volume of water being pumped. It also demonstrates that horizontal wells consume the highest amount of water when compared to the other types of wells as a result of multi-stage fracturing operations within the same horizontal well. The figure also captures the fact that over the past five years the average pumped water per horizontal well has tripled from 5 million gallons per well in 2013 to 15 million gallons per well in 2018.

Figure 15 shows the average water consumption per well for each of the three major basins in the Permian. It also shows that the Central Basin Platform has the minimum average water consumption per well due to the fact that more than 70 percent of the vertical wells were drilled within this basin. Delaware and Midland basins average water consumption per well is relatively close in value with the latter being slightly higher since Q1 of 2015.

4.3. PROPPANT LOADING QUANTITIES AND TRENDS

Proppant loading is defined as how much proppant were pumped relative to one gallon of water. The proppant loading parameter in this study was calculated by dividing the total amount of proppant (in pounds) by the total volume of water (in gallons).

Figure 16 shows a combined distribution histograms and box plots of the proppant loading over each year of the studied period. The median, average, and the 90th percentile are marked on each box plot. The distribution histograms show a practical value range from 0.5 to 1.5 Lbs/Gal. The average and median are close in values which indicates a low skewness in the data distribution with the average being slightly higher than the median which indicates a slight right skewness.

Figure 17 shows a combined bar and line charts of the proppant loading average and median values respectively. The red bar shows the average proppant loading for each quarter of the studied period associated with the black line series that shows the median value of the proppant loading for each quarter. The average value fluctuated between 0.9 to 1.1 Lbs/Gal.

4.4. HORIZONTAL LATERAL LENGTH AND PERFORATED INTERVALS DATA AND TRENDS

Horizontal wells are usually drilled vertically or directionally up to a certain depth (kick-off point) and then steered horizontally into the target zone. The horizontal section can be extended to thousands of feet to create a larger surface contact area with the reservoir. The drilled lateral then will be perforated and stimulated in several different stages. To understand the lateral and perforated interval lengths and trends over time in the Permian Basin the following figures were generated.

Figure 18 shows combined distribution histograms and box plots of lateral lengths in the Permian for each year of the studied period. The average and median values are marked on each box plot. The distribution histograms show that the lateral length can vary from 5 to 11 thousand feet. The box plots show that the lateral length average and median values have increased over time. The average lateral length increased from six thousand in 2013 to more than eight thousand in 2018.

Figure 19 shows a combined distribution histograms and box plots of the perforated intervals of the lateral length for each year of the studied period. The average and median values are marked on each box plot. The distribution histograms show that the perforated lateral length can vary from low values of few feet to 10 thousand feet. The low values of perforated intervals are associated with vertical wells. The box plots show that the perforated lateral intervals average and median values have increased over time. The average lateral length increased from 2900 ft in 2013 to more than seven thousand in 2018.

In the previous two figures, the data distribution and average values for 2017 and 2018 showed similar trends for the horizontal and perforated lengths. If the same trend continues for 2019, it will indicate that the operators in the Permian have optimized their drilling programs and reached to their desired lateral length that balances the drilling and completion costs with the production enhancement from a longer lateral.

Figure 20 shows a combined line and bar charts for each quarter of the studied period. The bar chart represents the average value for the horizontal lateral length, the purple line series represents the median value of the horizontal lateral lengths, and the green line series represents the median value of the perforated lateral intervals. The figure confirms that as time progressed the lateral length and the perforated interval have increased. It also showed that the gap between perforated interval and lateral length became narrower with time which indicates that about 90% of the horizontal well lateral lengths have been perforated and stimulated. In 2013 the perforated intervals were approximately 60% of the horizontal length while starting in Q2 2016 the percentage of the perforated lateral to the horizontal length became closer to 90%. The perforated lateral interval can be used as a proxy to a number of stimulation stages and this trend shows an increase in the number of stimulation stages as time progressed in the Permian Basin.

4.5. DRILLING TYPES AND ACTIVITIES OVER TIME

Drilling activities are measured by how many wells were drilled and completed for a certain period of time. There are three types of wells drilled in the Permian: vertical, horizontal and directional. Figure 21 shows a bar chart of wells count in each quarter of the study period. The drop of wells count confirms the recent industry downturn as a result of oil and gas prices crash during the last quarter of 2014. Each bar subdivided and colored based on the type of the wells drilled during that quarter. Green represents vertical wells percentage, red represents a horizontal well percentage, and blue represents directional wells percentage. Before the oil and gas prices drop, more than half of the stimulated wells in the Permian were drilled horizontally. Starting in 2015 the industry shifted their focus to drill and stimulate more horizontal wells and by 2017 more than 90% of the wells were drilled horizontally.

4.6. SPACIAL AND TEMPORAL PAD DRILLING ACTIVITIES

Pad drilling is the approach of drilling multiple wells from one single drilling site. It reduces the drilling and completion surface footprint. Pad density is defined as how many wells are drilled within close surface proximity on a single drilling site. To study the evolution of pad drilling practices in the Permian Basin the following figures were prepared to understand the trends and activities of pad densities.

Figure 22 represents the map of Texas and part of New Mexico. More than 20,500 horizontal and directions wells that drilled between Q1 2013 and Q2 2018 were included in this study. Since more than 99% of the wells were drilled in a pad of six wells or smaller, the data in this study were limited to pad density range of one to six. The wells on the map are color-coded based on the pad density. i.e. a well which is drilled on a single well pad will be labeled as pad density of 1 and colored in light blue on the map. Similarly, a well which is drilled on a pad that contains 6 wells, will be labeled as pad density of 6 and will be colored on the map with black color.

Figure 23 shows combined box plots and distribution histograms for the pad densities from 2013 to 2018. Earlier in the Permian (2013 and 2014), more than 50% of the wells were drilled on a single well pad site as indicated by the median value which is equal to one. After 2015 the median became 2 and the average value started to shift upward which indicates an increase in the number of wells that drilled on a multiple wells pad.

Figure 24 shows a bar chart that illustrates the wells count of each pad density over time. The bars indicate that at earlier time in the study, the single well pads were dominating the distribution. As time progressed, multiple wells pads increased in well counts and in 2018, the double wells pad dominated the distribution.

Figure 25 shows multiple line charts and each line colored based on the pad density value. This figure associates the lateral length to the pad density to understand the trends of drilling strategies in the Permian Basin. It shows that the wells drilled on a single pad were shorter on average lateral length than the wells drilled on multiple wells pad.

Figure 26 shows 6 pie charts illustrate the percentage of wells in each pad density over the past 6 years. It is clearly evident that single well pads are diminishing with time. Single well pads in 2013 represented 69% of the wells' population, in 2014 it became 56% and continued to be less every following year to become 22% in 2018 while double well pad became 30% and three wells pad became 23%.

From the previous five figures, it is evident that the operators in the Permian Basin are planning to expand the utilization of pad drilling by drilling multiple wells from one surface location to reduce the drilling rig move costs and stimulation logistics. It is also evident that the operators are recently trying to maximize their well designs by drilling the longer laterals on multiple wells pads.

4.7. STIMULATION AND COMPLETION VARIATION EFECTS ON AVERAGE AND MEDIAN PRODUCTION OF OIL AND GAS IN THE PERMIAN BASIN

In the unconventional resources, the initial production is considered as an effective parameter to measure the effectiveness of the drilling and stimulation operations for each reservoir. In this study, the practical initial production parameter was used as the initial production. Practical IP is calculated by dividing the second-month total production by the number of produced days of that month. The following 8 figures will discuss the production response of oil and gas practical IP and practical IP per stimulated foot to the water and proppant per stimulated foot parameters.

Figure 28 shows a bar chart of four binned ranges of water volume per perforated lateral length on the x-axis, and the right-hand side y-axis represents the average value of practical gas IP in (MCF/Day), while the left-hand side of y-axis represents the median value of practical gas IP in (MCF/Day). The figure shows that at water per foot value of more than 3000 Gal/ft, the average and median Prac. IP of gas decreased.

Figure 4.25 shows a bar chart of four binned ranges of water volume per perforated lateral length on the x-axis, and the right-hand side y-axis represents the average value of practical oil IP in (BBL/Day), while the left-hand side of y-axis represents the median value of practical oil IP in (BBL/Day). The figure shows that at water per foot value of more than 3000 Gal/ft, the average and median Prac. IP of oil decreased.

The effect of excess stimulation water volume on initial production could be attributed to the excess water flowback during the initial production which competes with the hydrocarbon production.

Figure 29 shows a bar chart of eight binned ranges of proppant mass per perforated lateral length on the x-axis, and the right-hand side y-axis represents the average value of practical gas IP in (MCF/Day), while the left-hand side of y-axis represents the median value of practical gas IP in (MCF/Day). The figure shows that at proppant per foot value of more than 3000 Lbs/ft, the average and median Prac. IP of gas decreased.

Figure 30 shows a bar chart of eight binned ranges of proppant mass per perforated lateral length on the x-axis, and the right-hand side y-axis represents the average value of practical oil IP in (BBL/Day), while the left-hand side of y-axis represents the median value of practical oil IP in (BBL/Day). The figure shows that at water per foot value of more than 3000 Lbs/ft, the average and median Prac. IP of oil decreased.

To investigate the lateral length effect on production, normalized parameters of practical IP of gas and oil to the lateral length were generated by dividing the average and median production value for each well by its lateral length. The following four figures show the production per stimulation foot response to proppant per foot and water per foot.

Figure 31 shows a bar chart of six binned ranges of proppant mass per perforated lateral length on the x-axis, and the right-hand side y-axis represents the average value of practical gas IP per stimulated foot in (MCF/Day)/ft, while the left-hand side of y-axis represents the median value of practical gas IP per stimulated foot in (MCF/Day)/ft.

Figure 32 shows a bar chart of six binned ranges of proppant mass per perforated lateral length on the x-axis, and the right-hand side y-axis represents the average value of practical oil IP per stimulated foot in (BBL/Day)/ft, while the left-hand side of y-axis represents the median value of practical oil IP per stimulated foot in (BBL/Day)/ft.

Figure 33 shows a bar chart of six binned ranges of water volume per perforated lateral length on the x-axis, and the right-hand side y-axis represents the average value of practical gas IP per perforated lateral foot in (MCF/Day)/ft, while the left-hand side of y-axis represents the median value of practical gas IP per perforated lateral foot in (MCF/Day)/ft.

Figure 34 shows a bar chart of six binned ranges of water volume per perforated lateral length on the x-axis, and the right-hand side y-axis represents the average value of practical oil IP per perforated lateral foot in (BBL/Day)/ft, while the left-hand side of y-axis represents the median value of practical oil IP per perforated lateral foot in (BBL/Day)/ft.

All the previous four figures confirm the incremental increase of initial production per foot as a result of increasing the stimulation parameters represented by the amount of proppant per foot and the volume of water per foot.

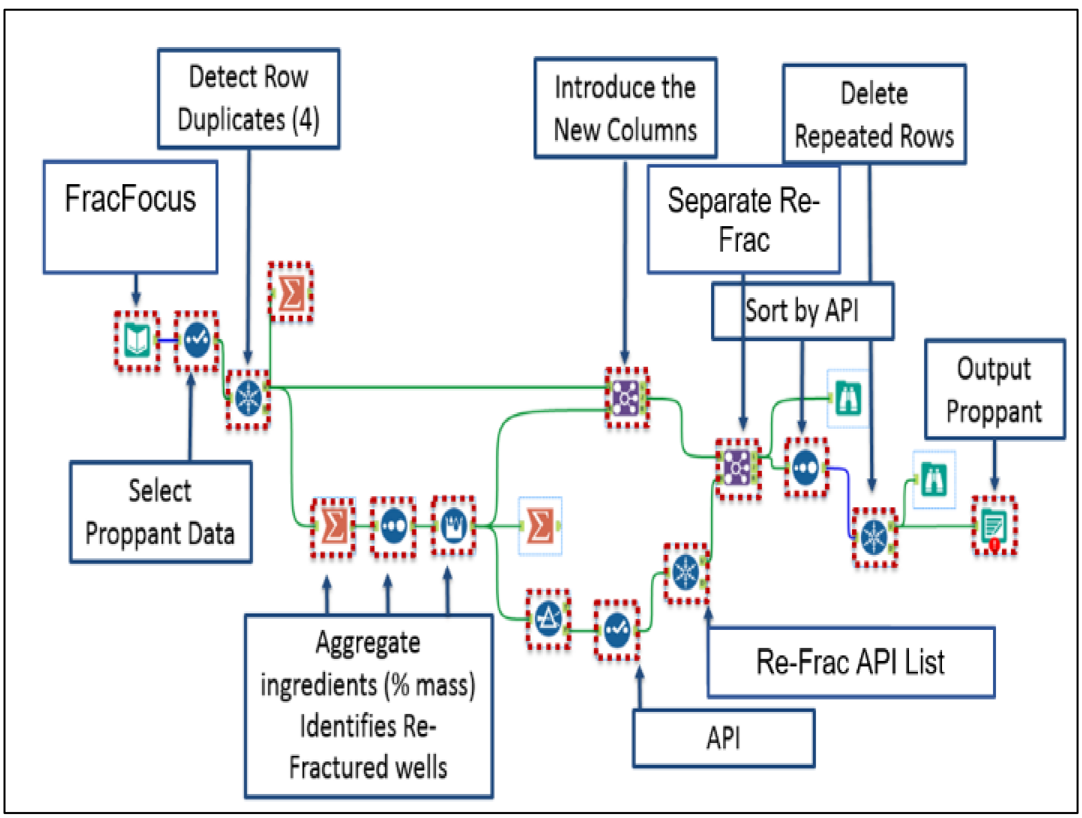


Figure 1. Data Processing Workflow to Handle Data from FracFocus

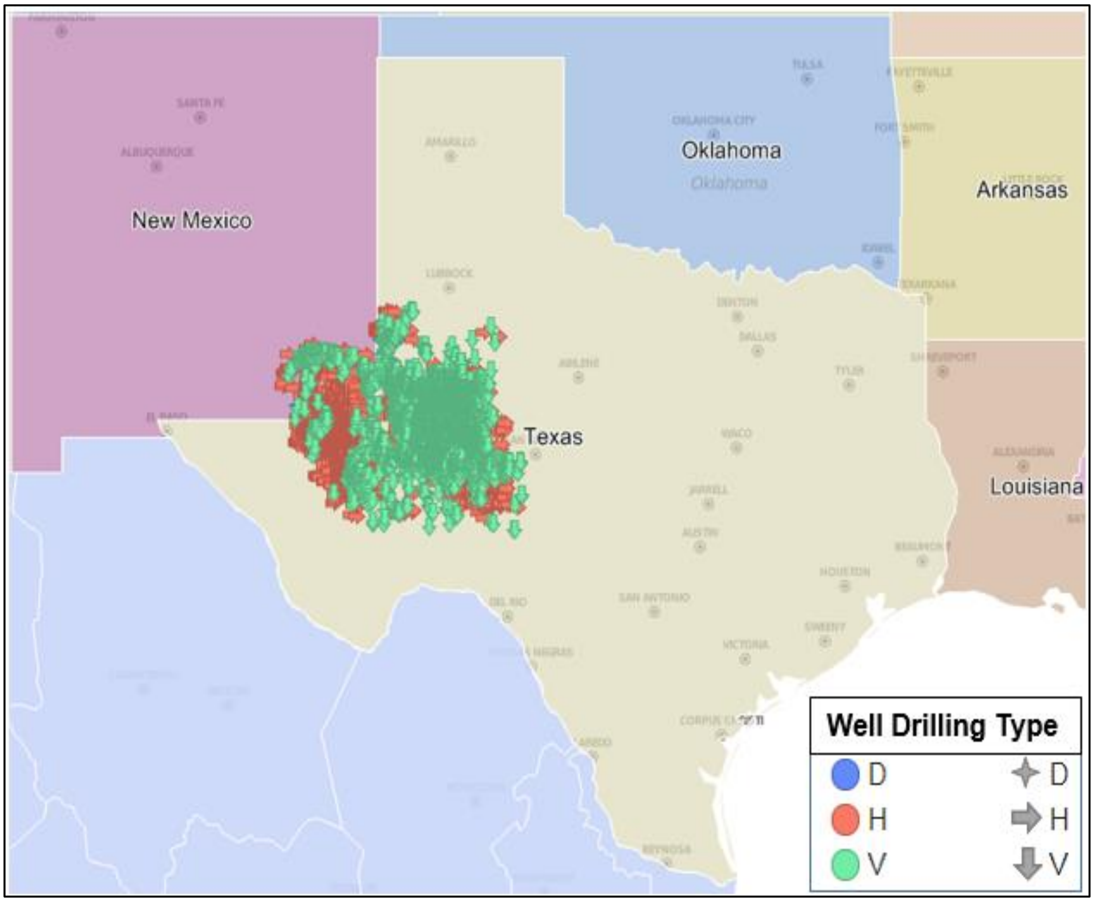


Figure 2. Wells' Locations on the U.S. Map

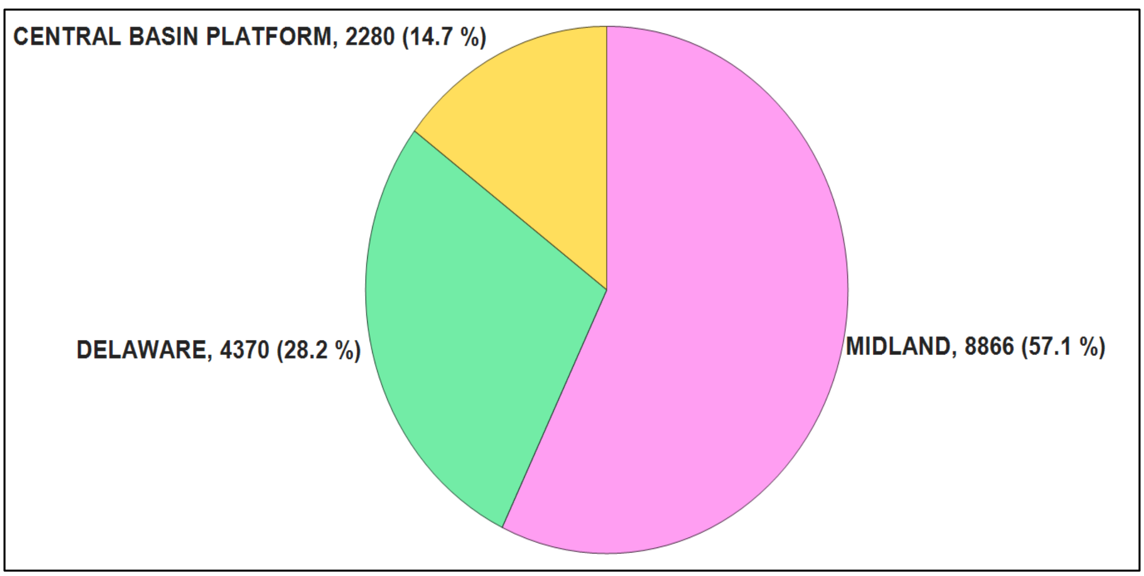
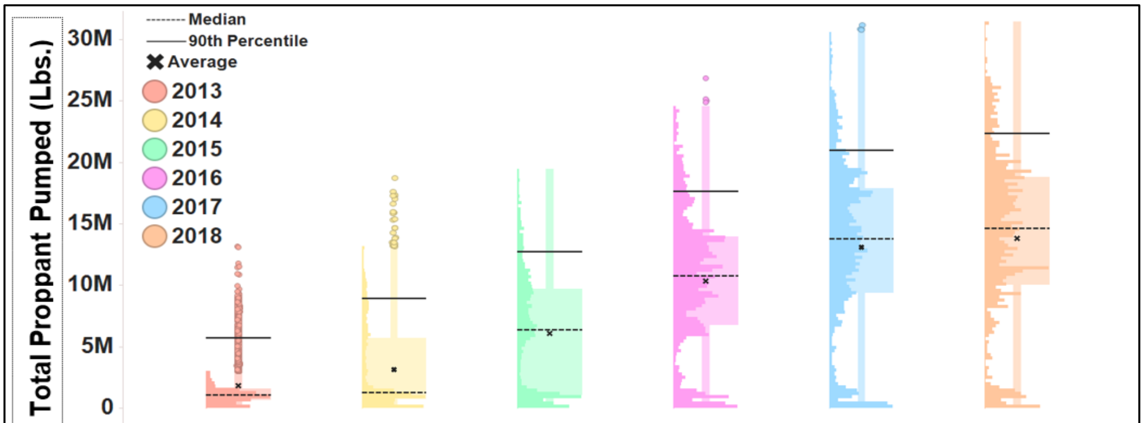


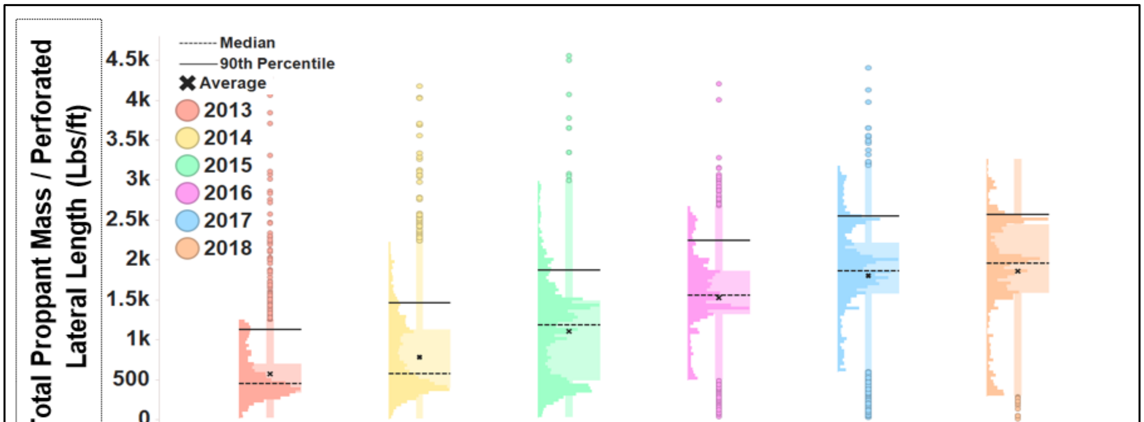
Figure 3. Data Coverage of the Major Three Components in the Permian Basin



	2013	2014	2015	2016	2017	2018
Count	3045	4430	2651	1793	2906	1041
Avg	1.84301e+06	3.14311e+06	6.11823e+06	1.03731e+07	1.31368e+07	1.38311e+07
Median	1.06443e+06	1.26251e+06	6.35434e+06	1.07489e+07	1.37646e+07	1.45859e+07
Sum	5.61195e+09	1.3924e+10	1.62194e+10	1.8599e+10	3.81756e+10	1.43981e+10
Max	1.31644e+07	1.87613e+07	1.94518e+07	2.6881e+07	3.11797e+07	3.14451e+07
Min	10327.3	11842.1	24883.3	16531	25206.3	28233.4
StdDev	2.21737e+06	3.52278e+06	4.82289e+06	5.71489e+06	6.84405e+06	6.89717e+06
Q1	667118	713671	1.06634e+06	6.74295e+06	9.37346e+06	1.0025e+07
Q3	1.60221e+06	5.70655e+06	9.71591e+06	1.39366e+07	1.79086e+07	1.87813e+07
IQR	935096	4.99287e+06	8.64958e+06	7.19366e+06	8.5351e+06	8.75635e+06
Outliers	495	27	0	4	4	0

Completion Year

Figure 4. Proppant Mass Box Plot and Data Distribution Histogram



	2013	2014	2015	2016	2017	2018
Count	3043	4419	2648	1786	2850	959
Avg	575.042	777.352	1107.62	1526.86	1804.35	1859.61
Median	443.398	576.405	1181.98	1552.72	1856.7	1952.06
Sum	1.74985e+06	3.43512e+06	2.93299e+06	2.72697e+06	5.14241e+06	1.78336e+06
Max	4587.31	4181.25	4556.6	4205.72	4410.17	3262.93
Min	14.5156	9.44803	28.6748	36.6295	28.0496	9.42057
StdDev	401.117	516.307	615.245	600.862	677.381	668.878
Q1	331.43	389.28	482.539	1314.43	1569.07	1581.93
Q3	699.842	1126.07	1484.87	1859.15	2212.65	2444.83
IQR	368.412	736.792	1002.33	544.719	643.588	862.9
Outliers	169	51	12	218	282	15

Completion Year

Figure 5. Proppant/Perforated Lateral Box Plot and Data Distribution Histogram

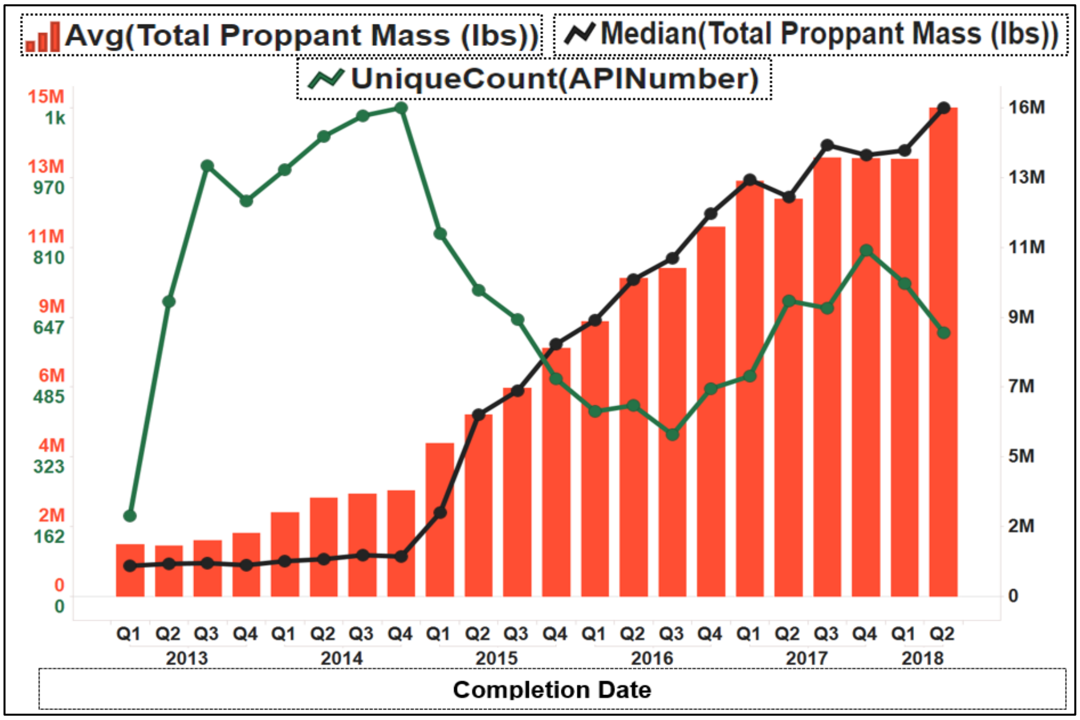


Figure 6. Average, Median Proppant Mass and Wells Count/Time

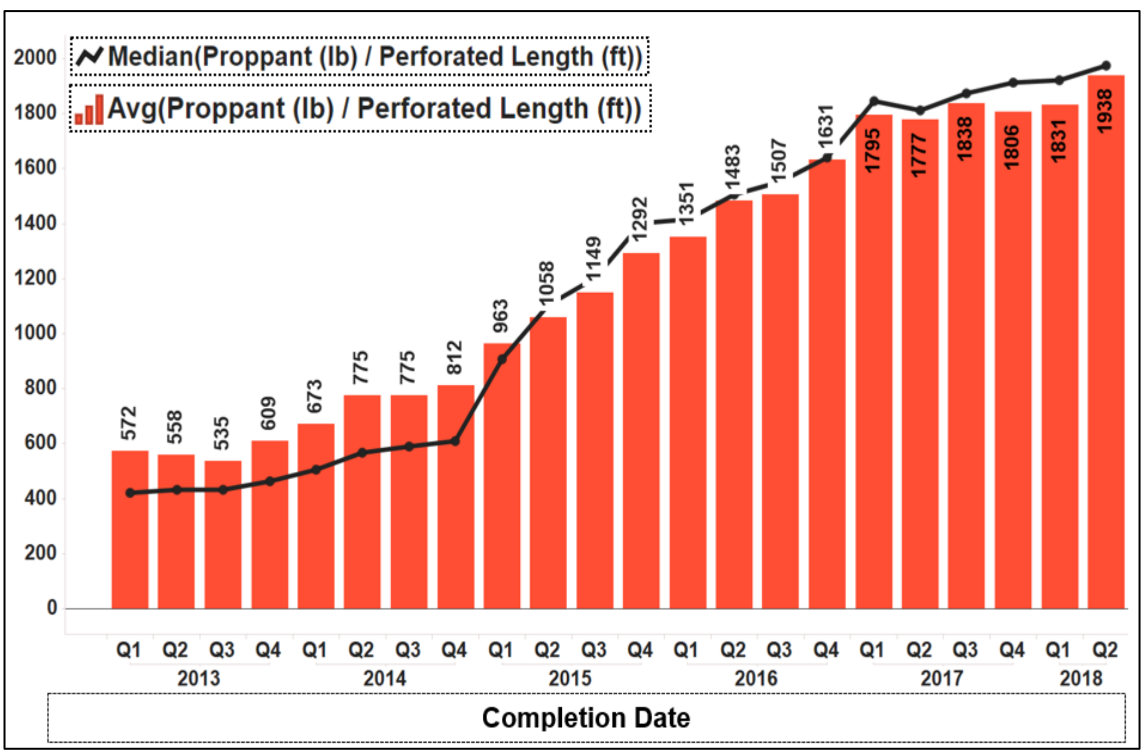


Figure 7. Average and Median Proppant Mass / Perforated Lateral over Time

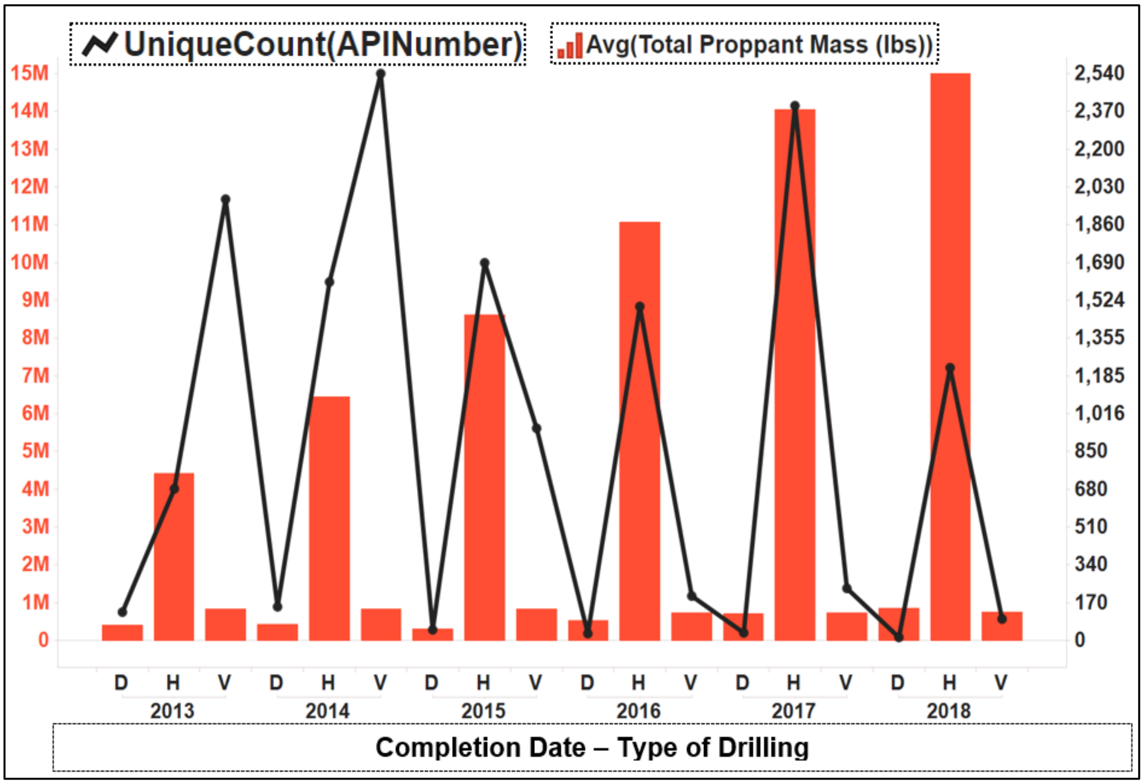


Figure 8. Average Proppant and Number of Wells over Time for each Well Type

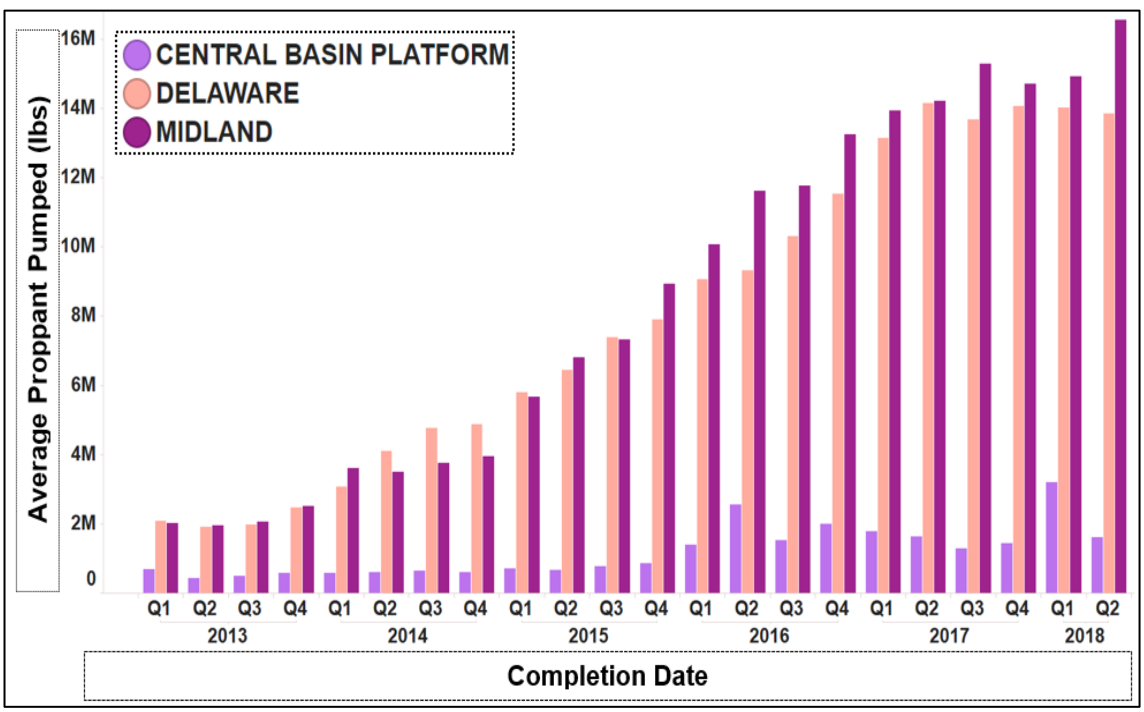
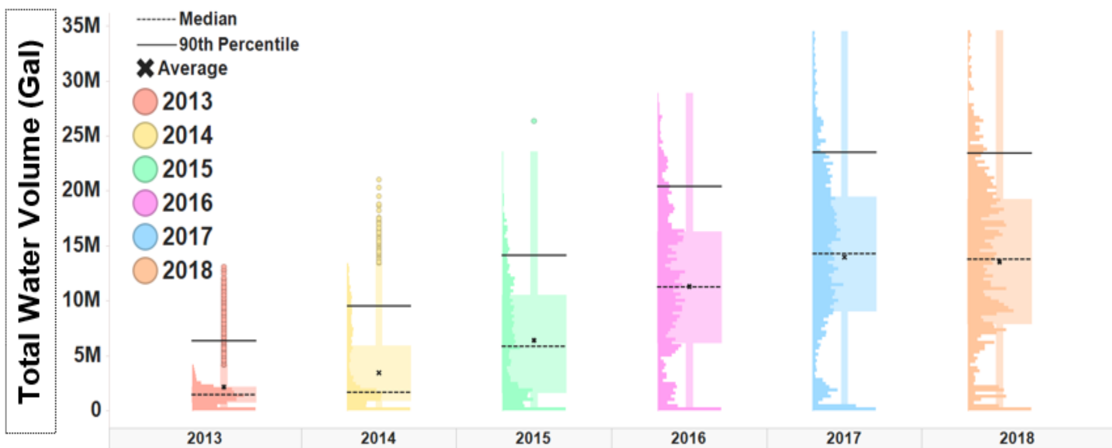


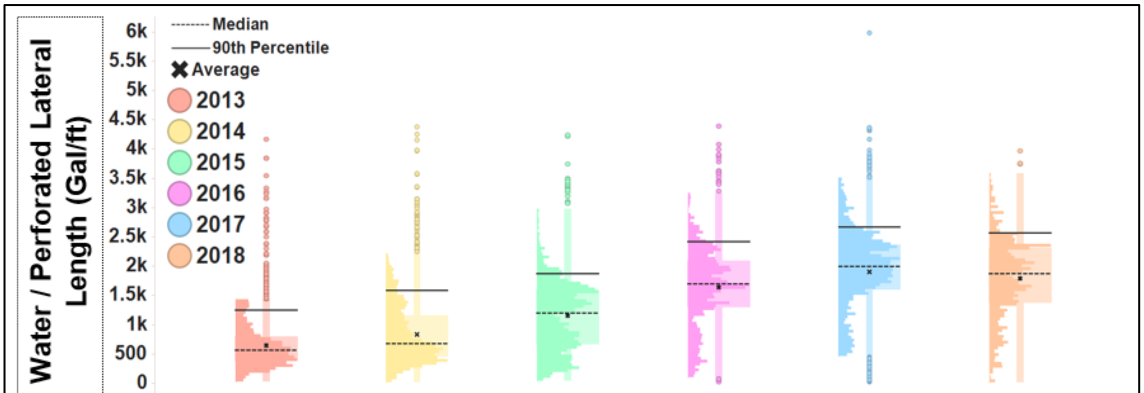
Figure 9. Average Proppant Consumption by Major Basins in the Permian



	2013	2014	2015	2016	2017	2018
Count	3045	4430	2651	1793	2906	1041
Avg	2.14566e+06	3.43769e+06	6.4472e+06	1.13525e+07	1.39607e+07	1.35807e+07
Median	1.41847e+06	1.67911e+06	5.86302e+06	1.12554e+07	1.43052e+07	1.37513e+07
Sum	6.53354e+09	1.5229e+10	1.70915e+10	2.03551e+10	4.05699e+10	1.41375e+10
Max	13147260	21049747	26349792	28912884	34544328	34587126
Min	7938	8274	8721	23036	25675	23868
StdDev	2.458e+06	3.80399e+06	5.281e+06	6.7942e+06	7.78042e+06	7.64446e+06
Q1	760496	849251	1.56202e+06	6.10403e+06	9.02422e+06	7.86715e+06
Q3	2.13637e+06	5.88575e+06	1.04941e+07	1.62781e+07	1.94324e+07	1.92465e+07
IQR	1.37588e+06	5.0365e+06	8.93204e+06	1.0174e+07	1.04082e+07	1.13794e+07
Outliers	435	78	1	0	0	0

Completion Year

Figure 10. Water Volume Box Plot and Data Distribution Histogram



	2013	2014	2015	2016	2017	2018
Count	3043	4419	2648	1786	2850	959
Avg	652.399	834.843	1153	1643.97	1897.86	1786.95
Median	560	676.953	1197.94	1686.36	1984.23	1865.16
Sum	1.98525e+06	3.68917e+06	3.05315e+06	2.93612e+06	5.4089e+06	1.71369e+06
Max	4166.17	4375.34	4246.2	4389.13	5983.69	3972.33
Min	20.9738	12.9234	40.3815	19.8457	25.1269	11.8559
StdDev	426.705	530.347	615.045	681.359	734.785	699.291
Q1	371.346	444.358	657.018	1289.35	1594.13	1368.13
Q3	800.051	1153.6	1587.97	2086.87	2360.1	2313.51
IQR	428.705	709.24	930.952	797.519	765.968	945.38
Outliers	130	62	22	24	185	3

Completion Year

Figure 11. Water Volume / Lateral Box Plot and Data Distribution Histogram

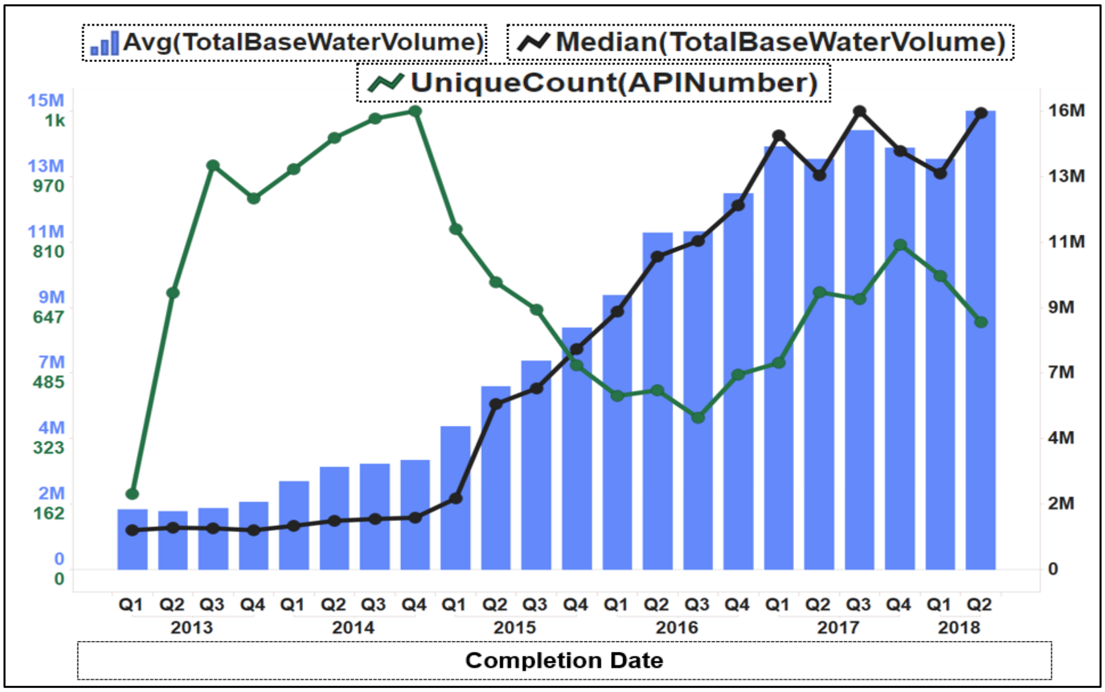


Figure 12. Average, Median Water Volume and Well Count/Time

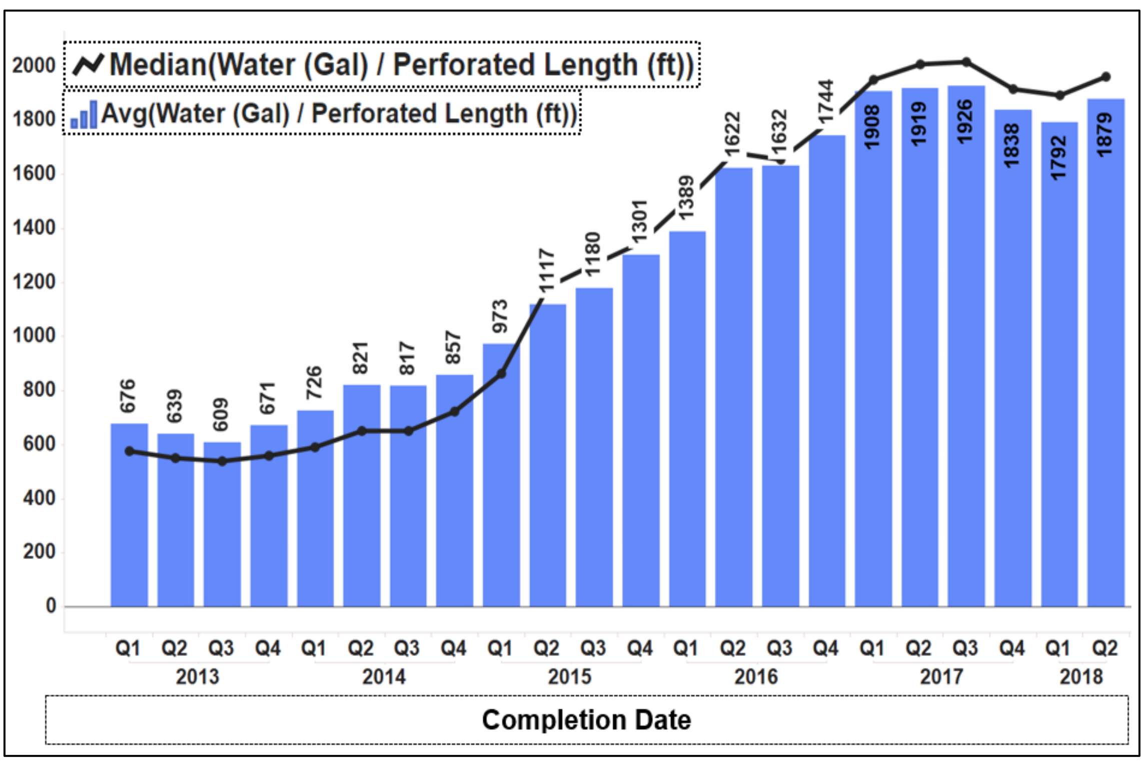


Figure 13. Average and Median Water Volume / Perforated Lateral over Time

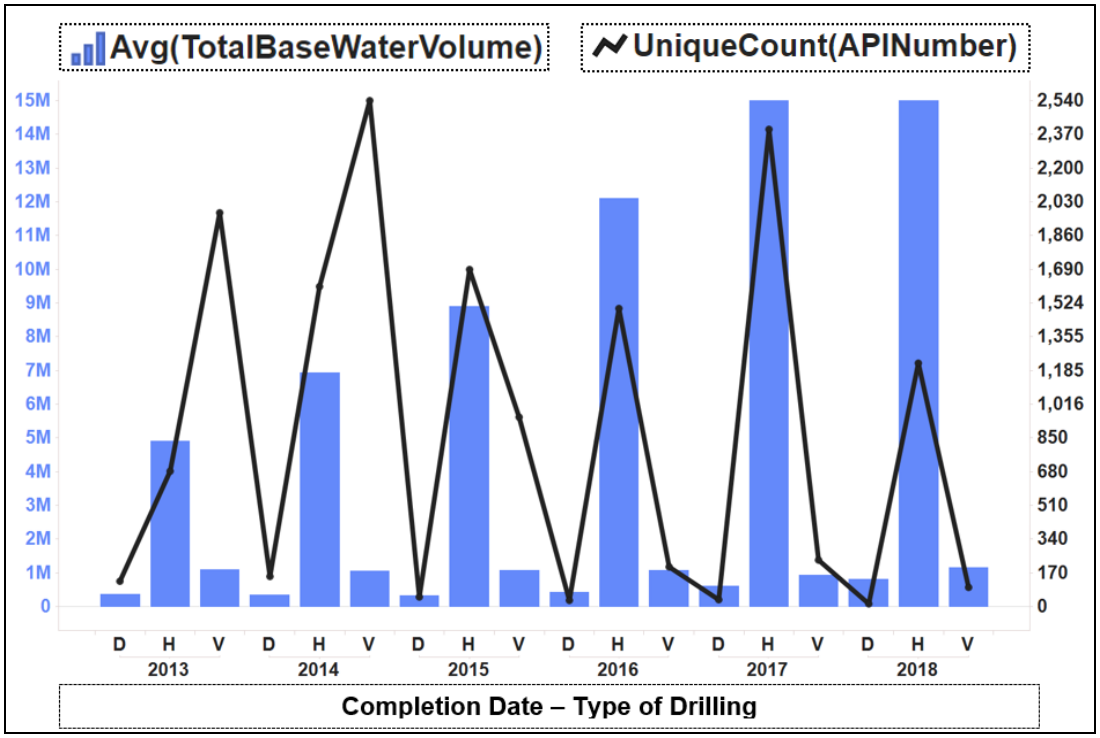


Figure 14. Average Water use and Number of Wells over Time for each Well Type

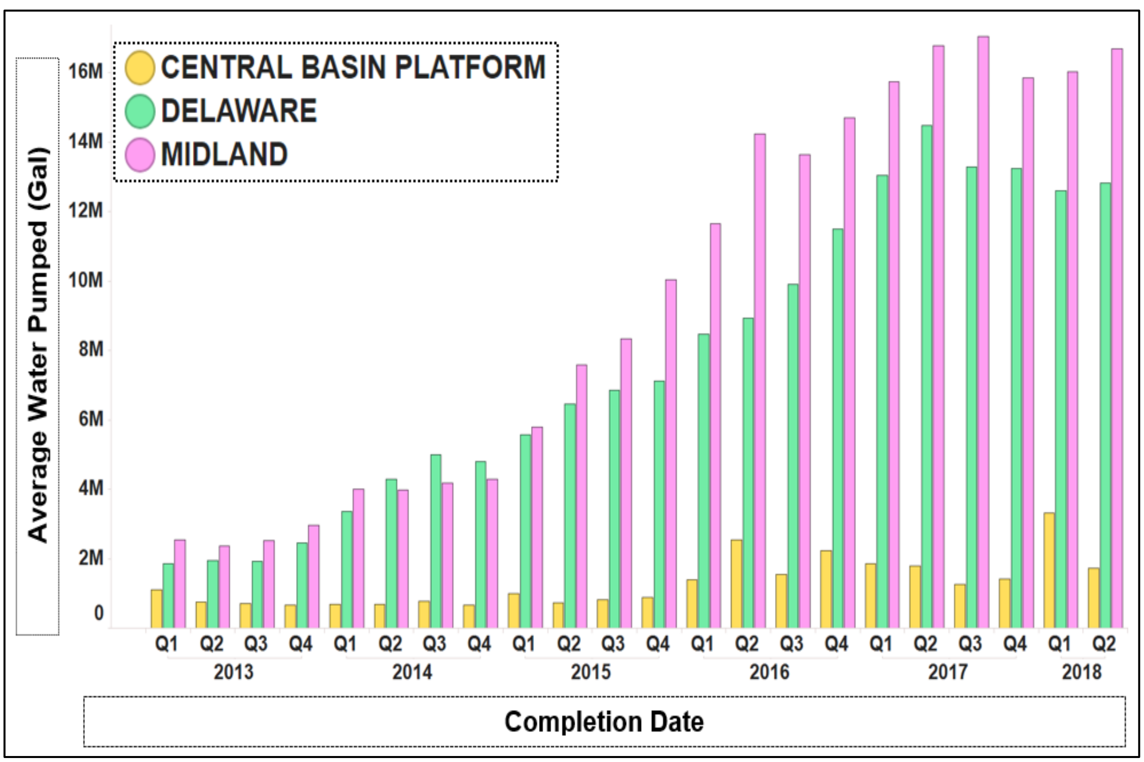
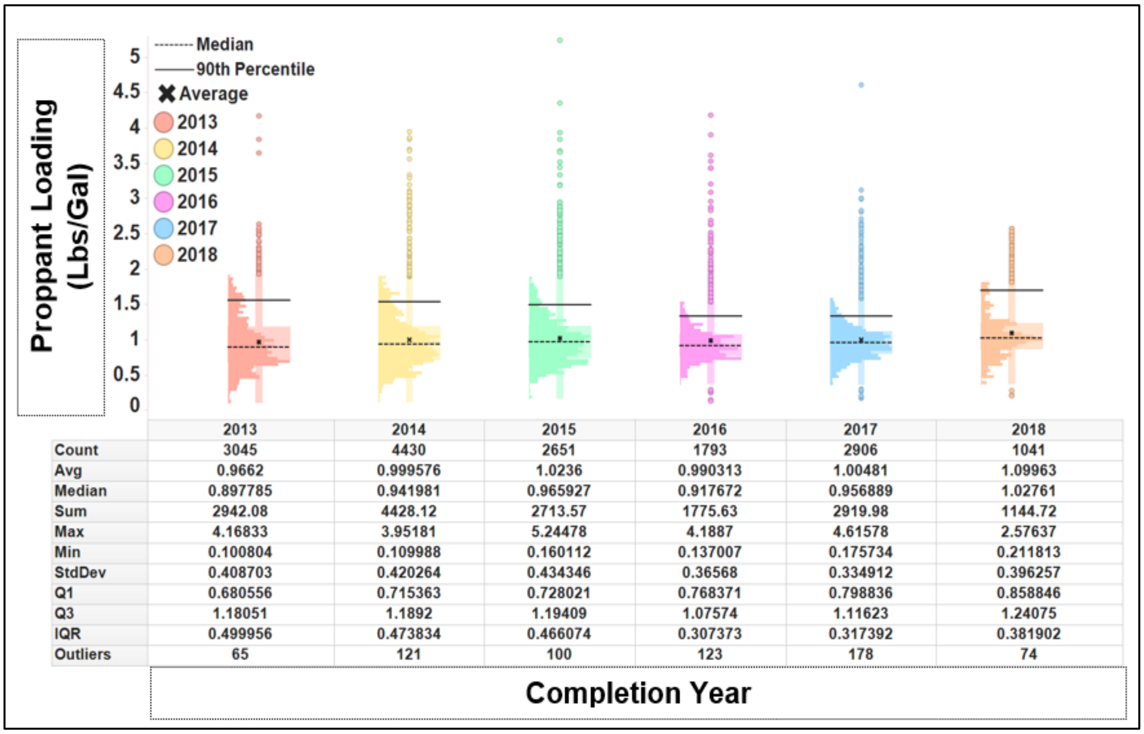


Figure 15. Average Water Consumption by Major Basins in the Permian



	2013	2014	2015	2016	2017	2018
Count	3045	4430	2651	1793	2906	1041
Avg	0.9662	0.999576	1.0236	0.990313	1.00481	1.09963
Median	0.897785	0.941981	0.965927	0.917672	0.956889	1.02761
Sum	2942.08	4428.12	2713.57	1775.63	2919.98	1144.72
Max	4.16833	3.95181	5.24478	4.1887	4.61578	2.57637
Min	0.100804	0.109988	0.160112	0.137007	0.175734	0.211813
StdDev	0.408703	0.420264	0.434346	0.36568	0.334912	0.396257
Q1	0.680556	0.715363	0.728021	0.768371	0.798836	0.858846
Q3	1.18051	1.1892	1.19409	1.07574	1.11623	1.24075
IQR	0.499956	0.473834	0.466074	0.307373	0.317392	0.381902
Outliers	65	121	100	123	178	74

Figure 16. Proppant Mass/ Water Volume Box Plot and Distribution Histogram

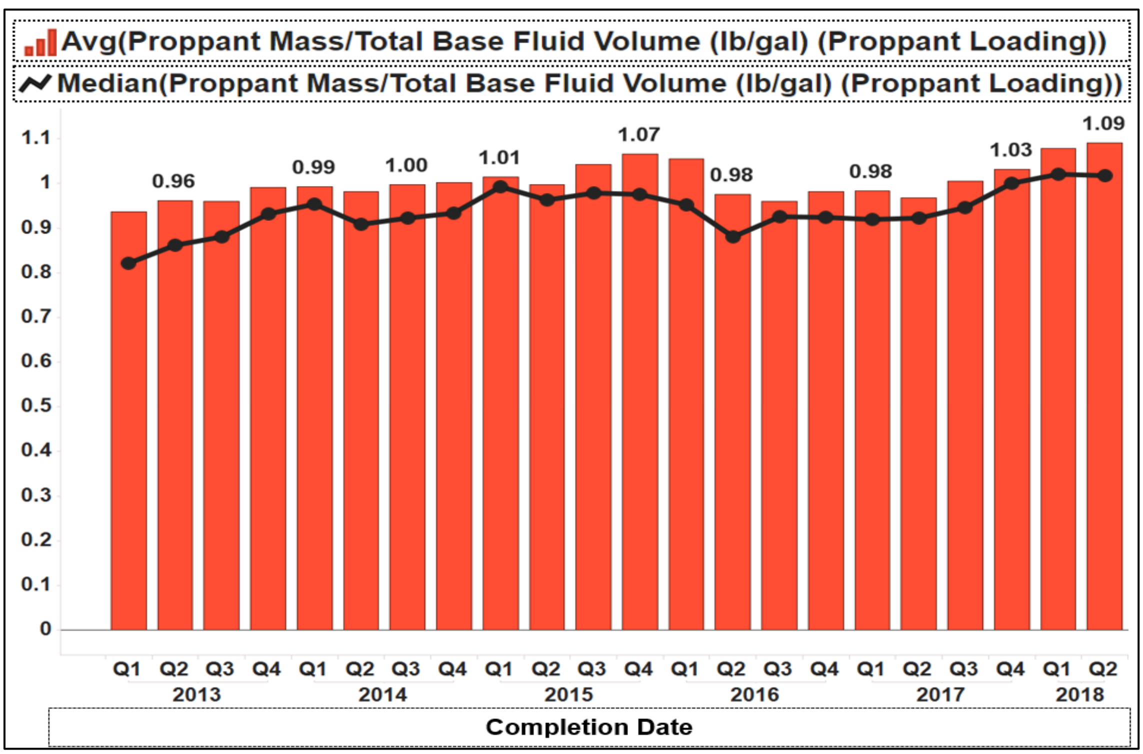
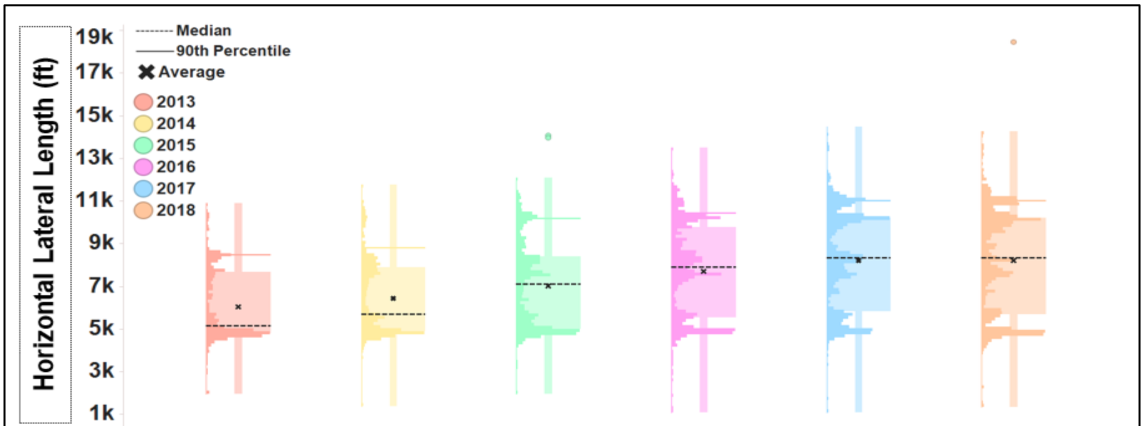


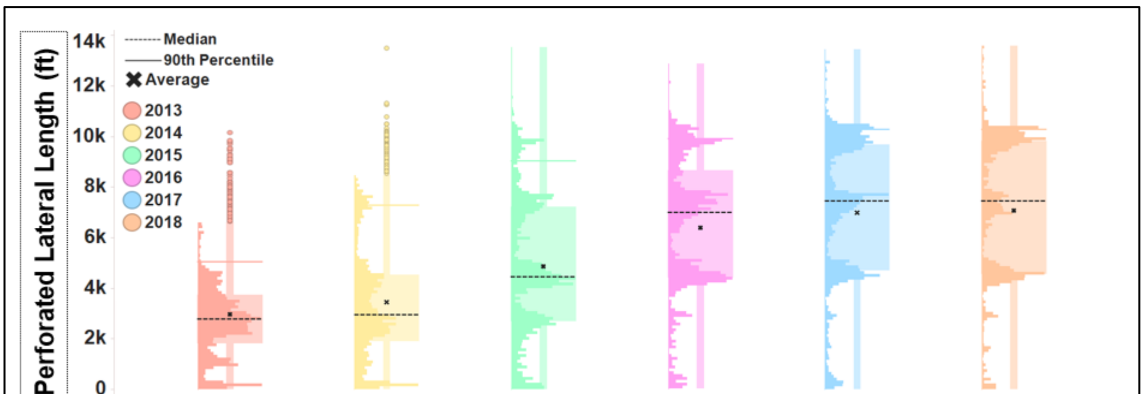
Figure 17. Average and Median Values of Proppant Loading



	2013	2014	2015	2016	2017	2018
Count	775	1691	1707	1538	2571	930
Avg	6048.11	6438.82	7030.36	7720.73	8232.03	8200.59
Median	5133.45	5680.62	7097.39	7884.14	8338.59	8325.68
Sum	4.68728e+06	1.0888e+07	1.20008e+07	1.18745e+07	2.11645e+07	7.62655e+06
Max	10900.6	11767.3	14090.6	13492.5	14488.9	18462.8
Min	1965.03	1388.52	1970.25	1088.19	1087.29	1330.62
StdDev	1678.45	1807.73	2035.82	2181.1	2294.58	2475.83
Q1	4763.24	4884.42	5013.7	5523.4	5839.97	5682
Q3	7692.48	7894.69	8388.84	9790.93	10213.6	10208.6
IQR	2929.24	3010.27	3375.15	4267.53	4373.65	4526.64
Outliers	0	0	2	0	0	1

Completion Year

Figure 18. Horizontal Length Box Plot and Data Distribution



	2013	2014	2015	2016	2017	2018
Count	3044	4420	2648	1786	2850	961
Avg	2970.21	3439.51	4853.46	6398.51	6990.54	7078.43
Median	2781	2950	4447	6983	7453	7438
Sum	9.04133e+06	1.52026e+07	1.2852e+07	1.14277e+07	1.9923e+07	6.80237e+06
Max	10152	13508	13527	12865	13450	13581
Min	0	0	14	21	9	0
StdDev	1885.89	2359.48	2779.67	2764	2916.81	2988.11
Q1	1820	1891.5	2690.25	4413	4695	4569
Q3	3748.5	4525	7223	8650.25	9672	9829
IQR	1928.5	2633.5	4532.75	4237.25	4977	5260
Outliers	229	135	0	0	0	0

Completion Year

Figure 19. Perforated Lateral Length Box Plot and Data Distribution Histogram

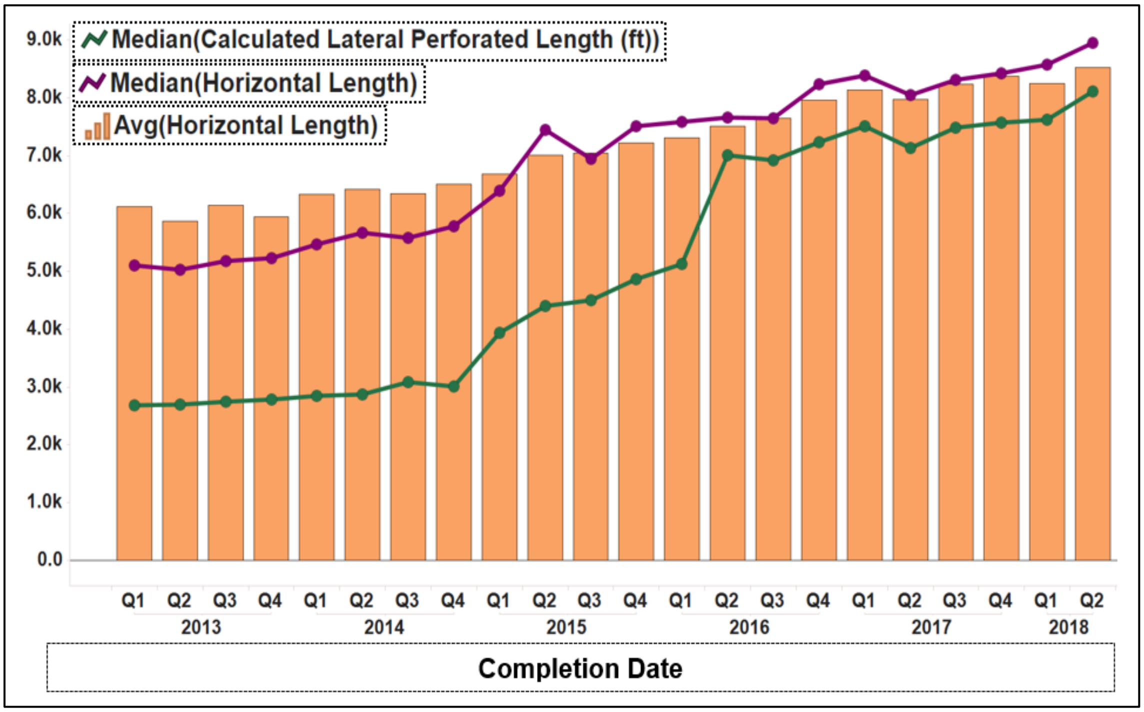


Figure 20. Average and Median Horizontal Length, Perforated Lateral Length

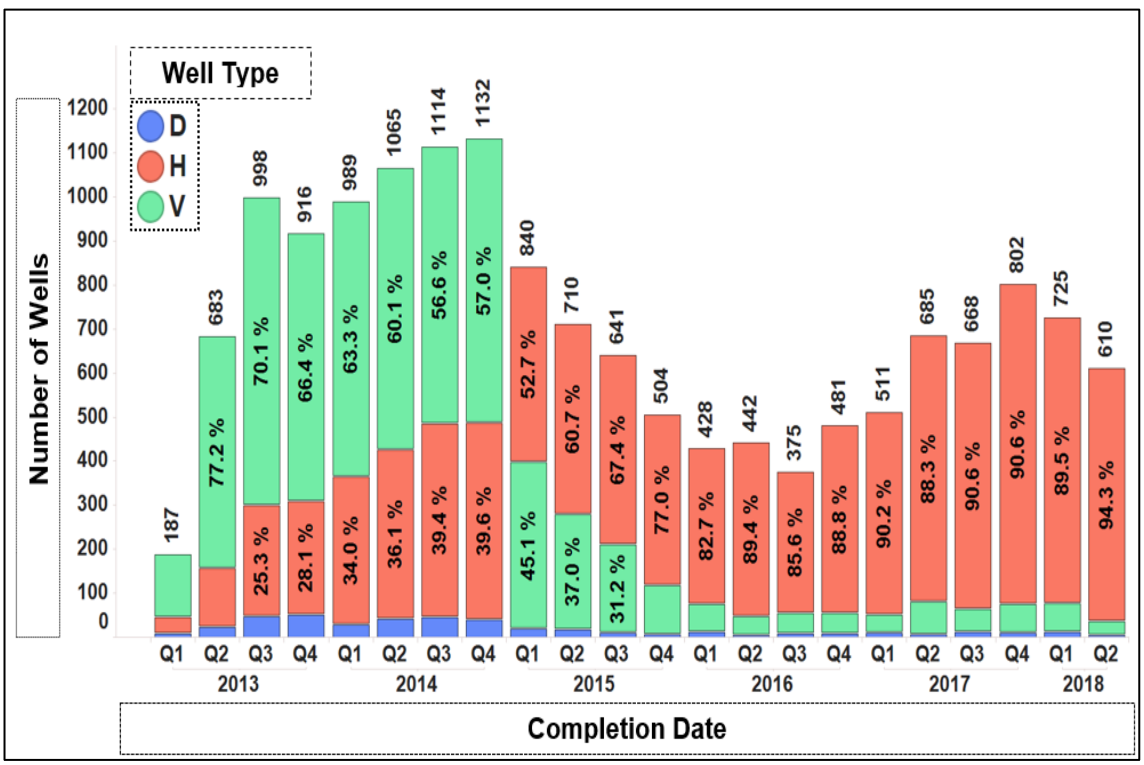


Figure 21. The Data Sample of the Number of wells and Completion over Time

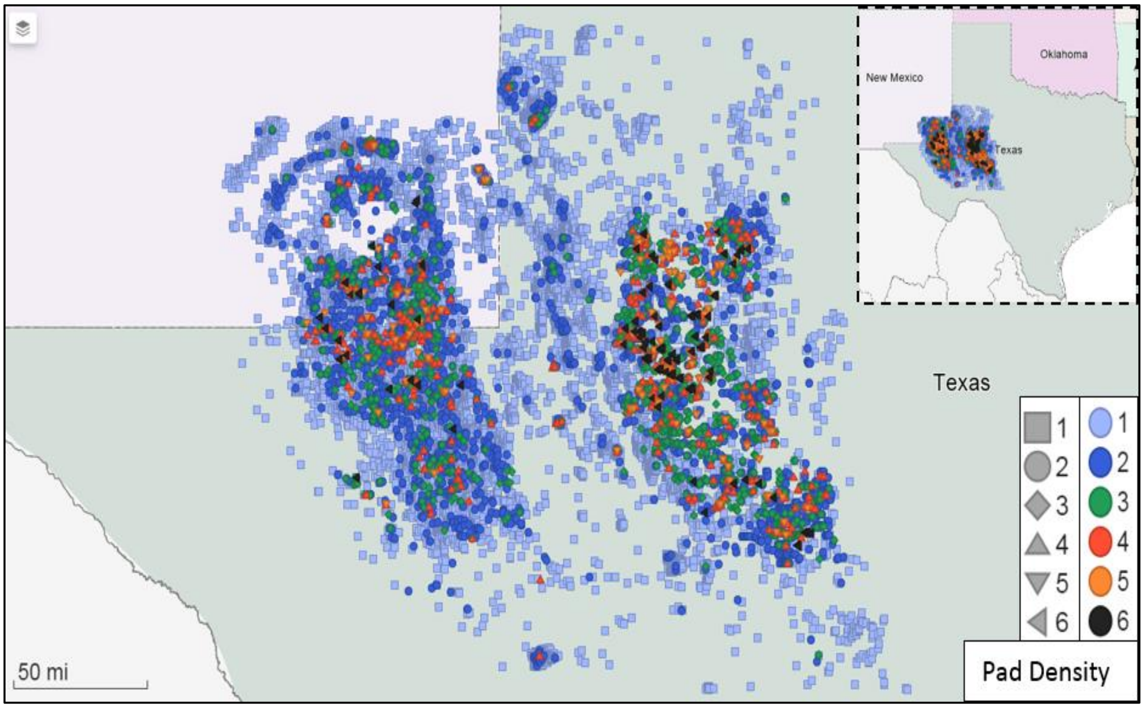
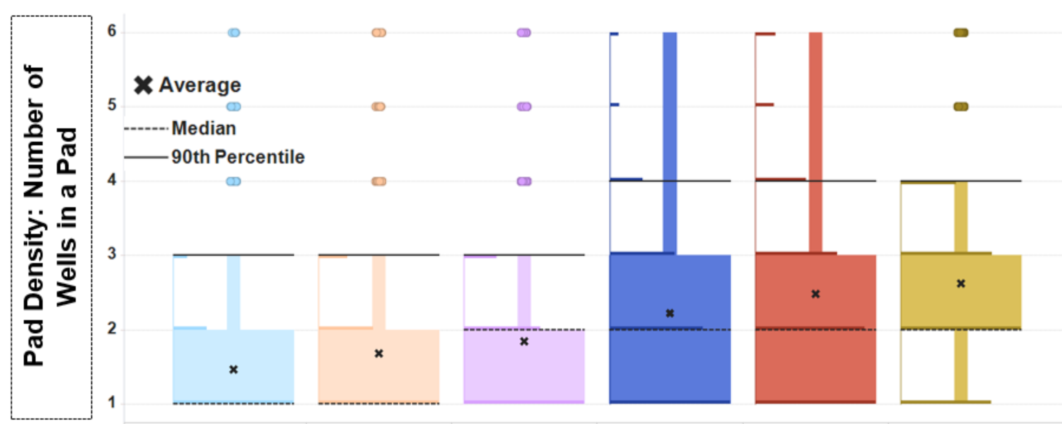


Figure 22. Map of Texas and New Mexico Shows Well Distribution and Pad Density



	2013	2014	2015	2016	2017	2018
Count	2483	3605	3014	2353	4011	5126
Median	1	1	2	2	2	2
Avg	1.5	1.7	1.8	2.2	2.5	2.6
Outliers	97	167	248	0	0	512
StdDev	0.8	0.9	1.0	1.3	1.4	1.4
Q1	1	1	1	1	1	2
Q3	2	2	2	3	3	3
P10	1	1	1	1	1	1
P90	3	3	3	4	4	4
MostCommon	1	1	1	1	1	2
IQR	1	1	1	2	2	1

Drilling Year

Figure 23. Box Plot and Distribution Histogram of Wells Count for each Pad Density

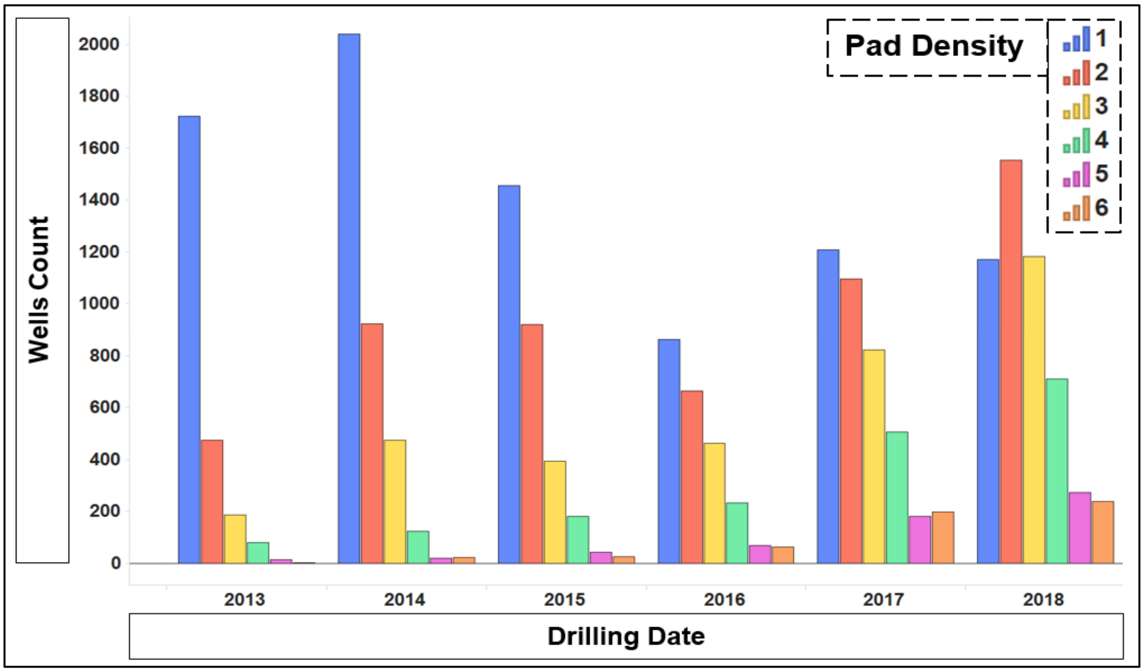


Figure 24. Bar Chart of Pad Density over Time

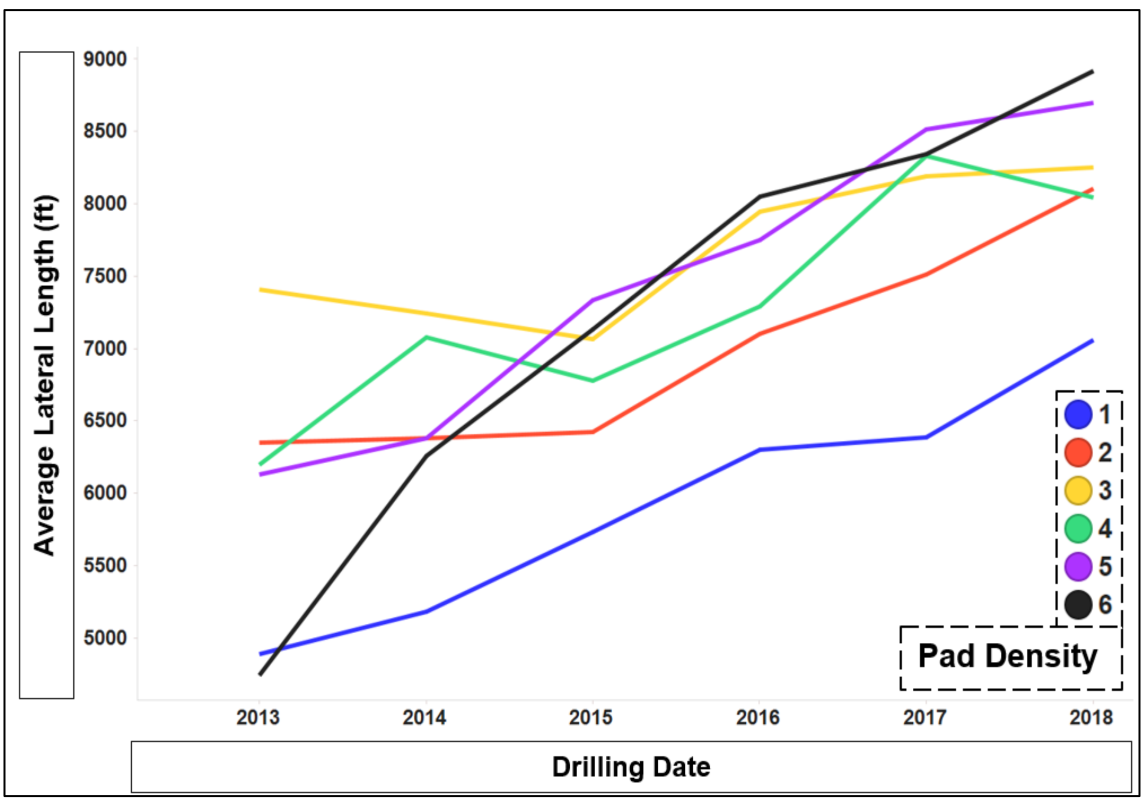


Figure 25. Line Chart of the Average Lateral Length/Time for each Pad Group

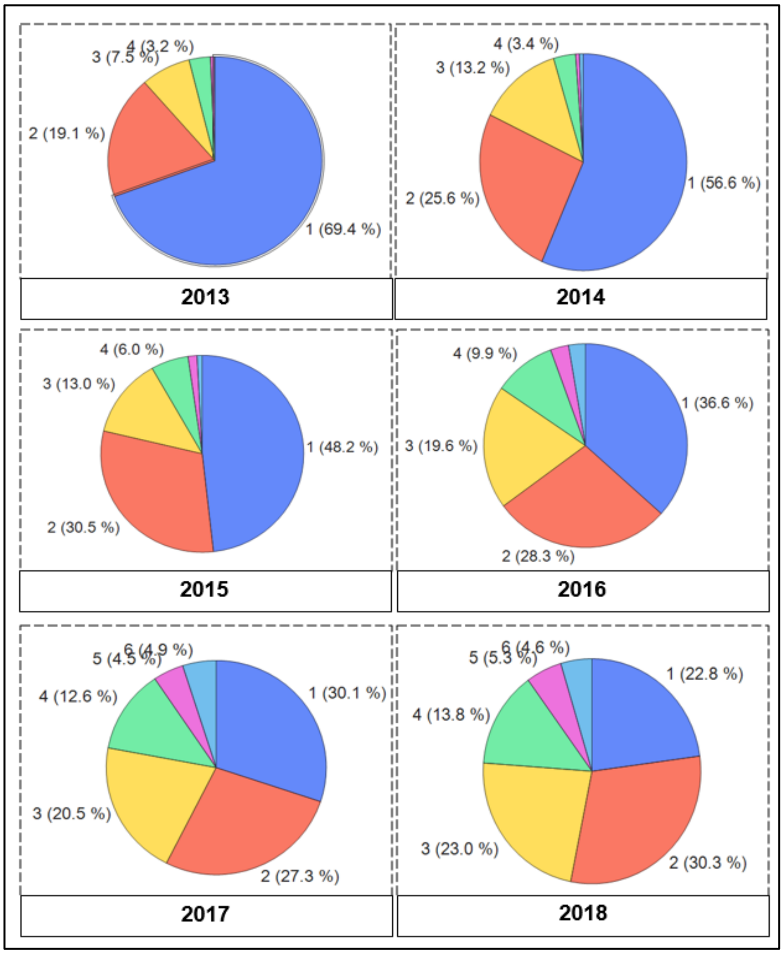


Figure 26. Pie Charts of Pad Densities-Wells' Percentages over Time

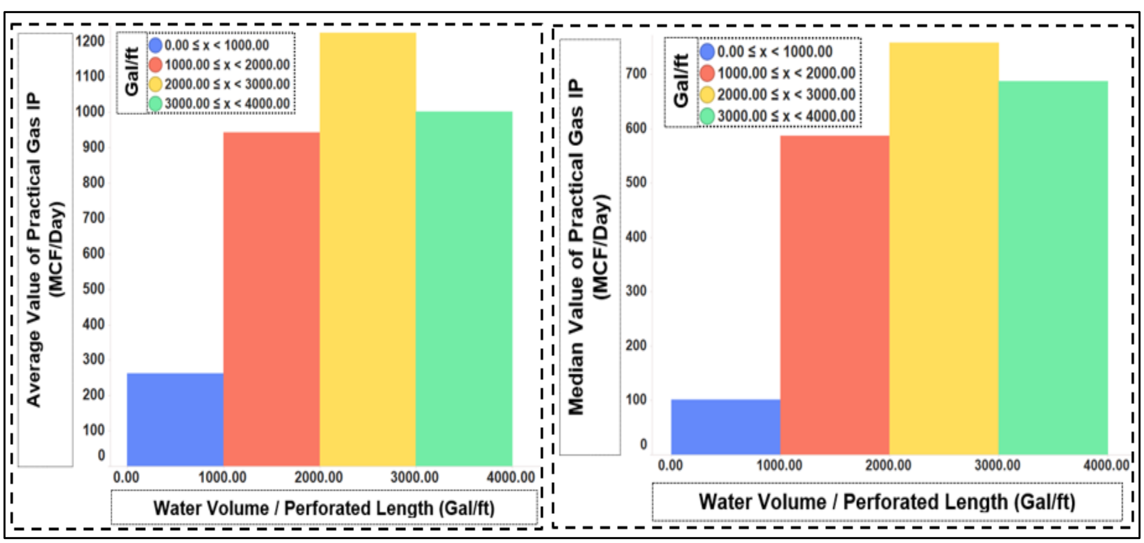


Figure 27. Water/Ft Effect on Average and Median Gas IP

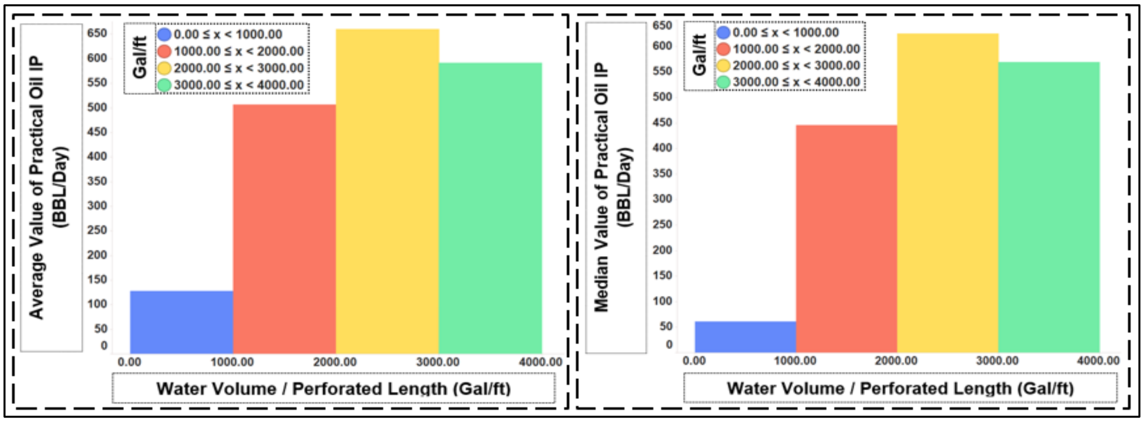


Figure 28. Water/Ft Effect on Average and Median Oil IP

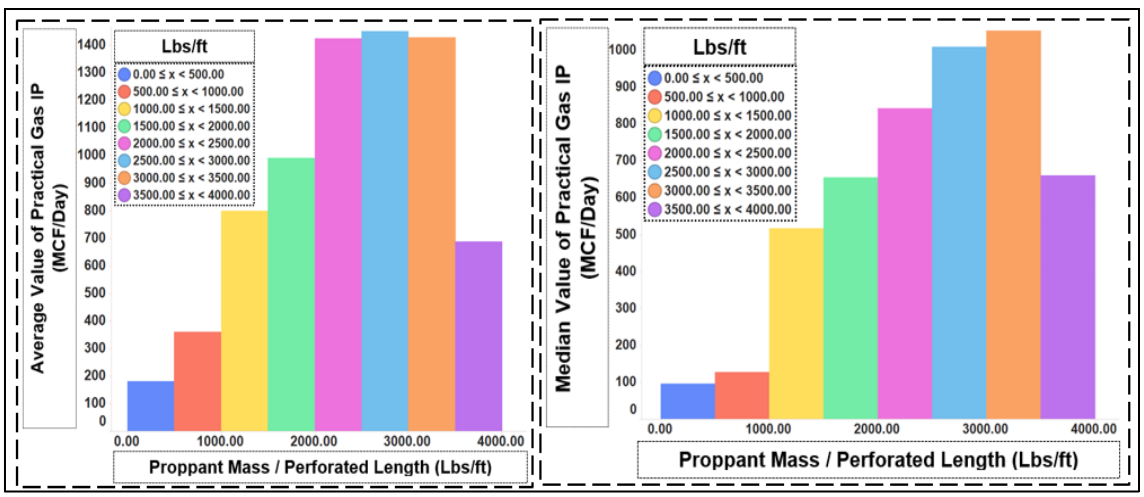


Figure 29. Proppant/Ft Effect on Average and Median Gas IP

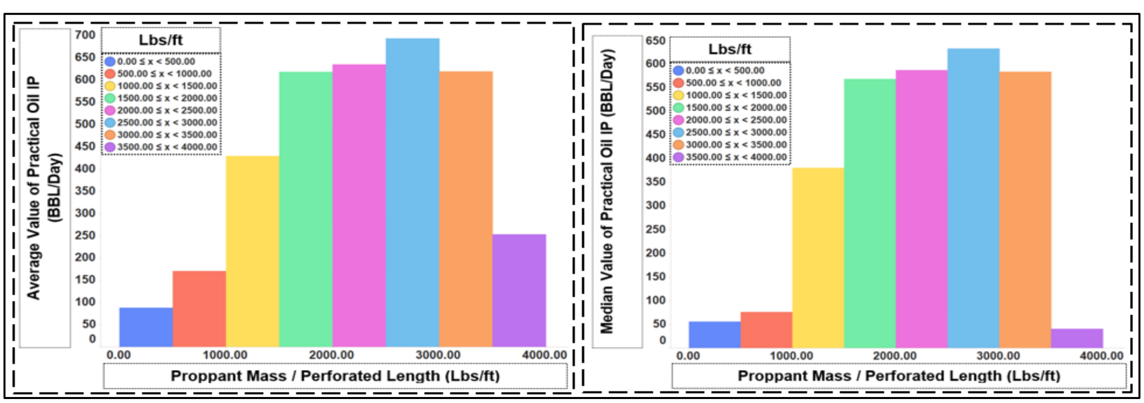


Figure 30. Proppant/Ft Effect on Average and Median Oil IP

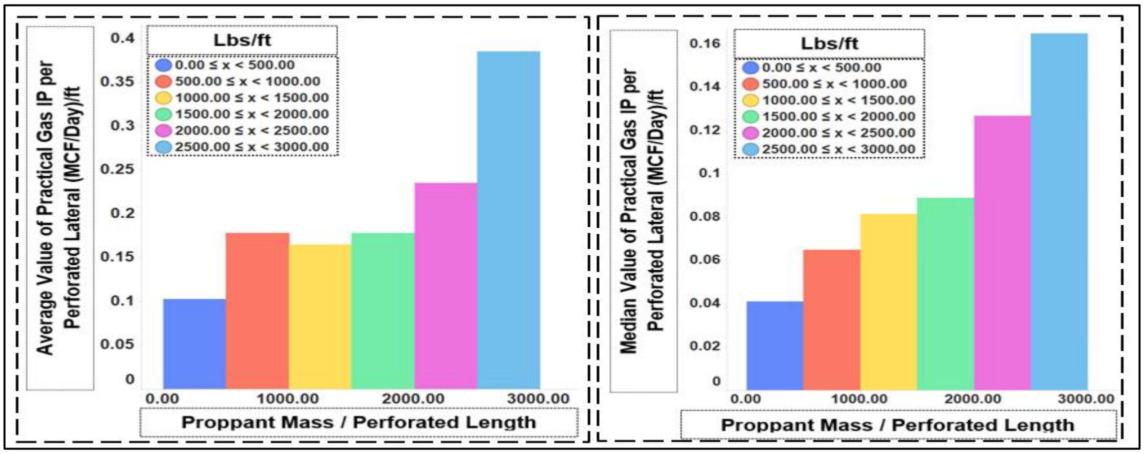


Figure 31. Proppant/Ft Effect on Average and Median Gas IP / Perforated Foot

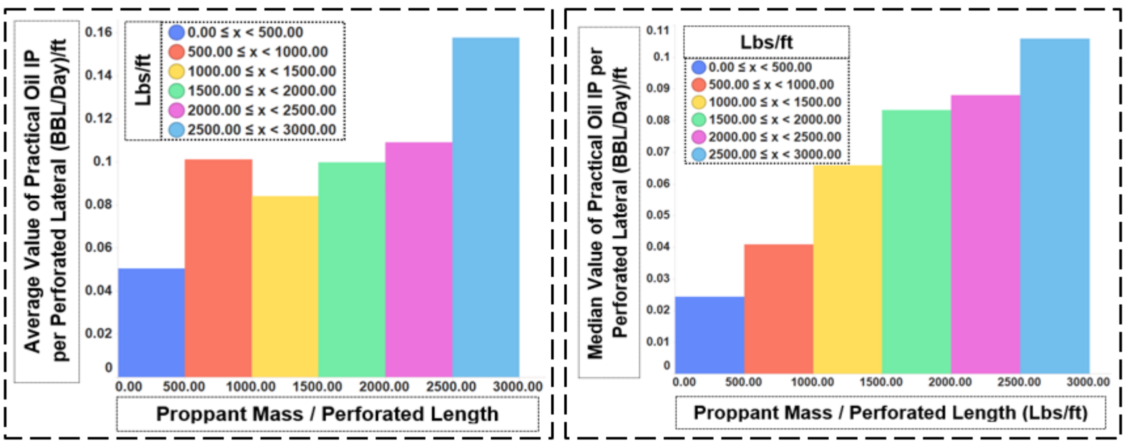


Figure 32. Proppant/Ft Effect on Average and Median Gas IP per Perforated Foot

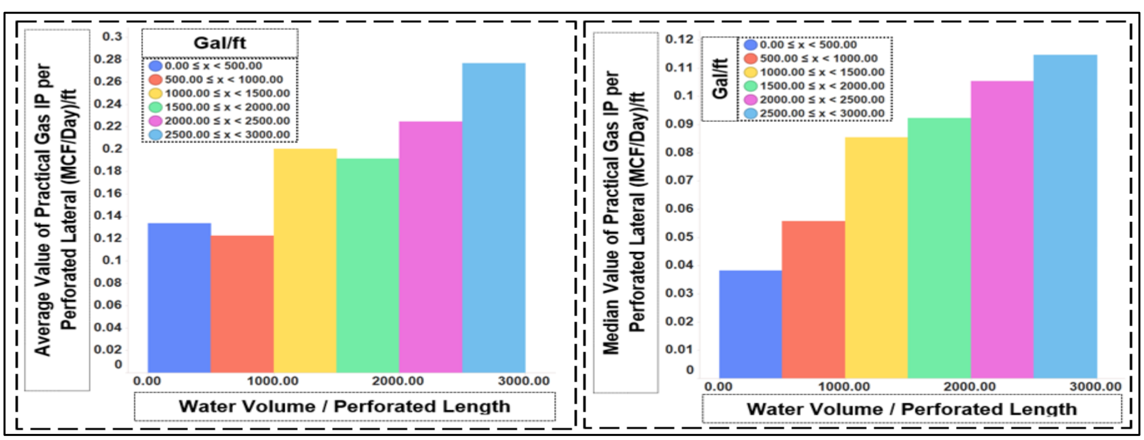


Figure 33. Water/Ft Effect on Average and Median Gas IP per Perforated Lateral

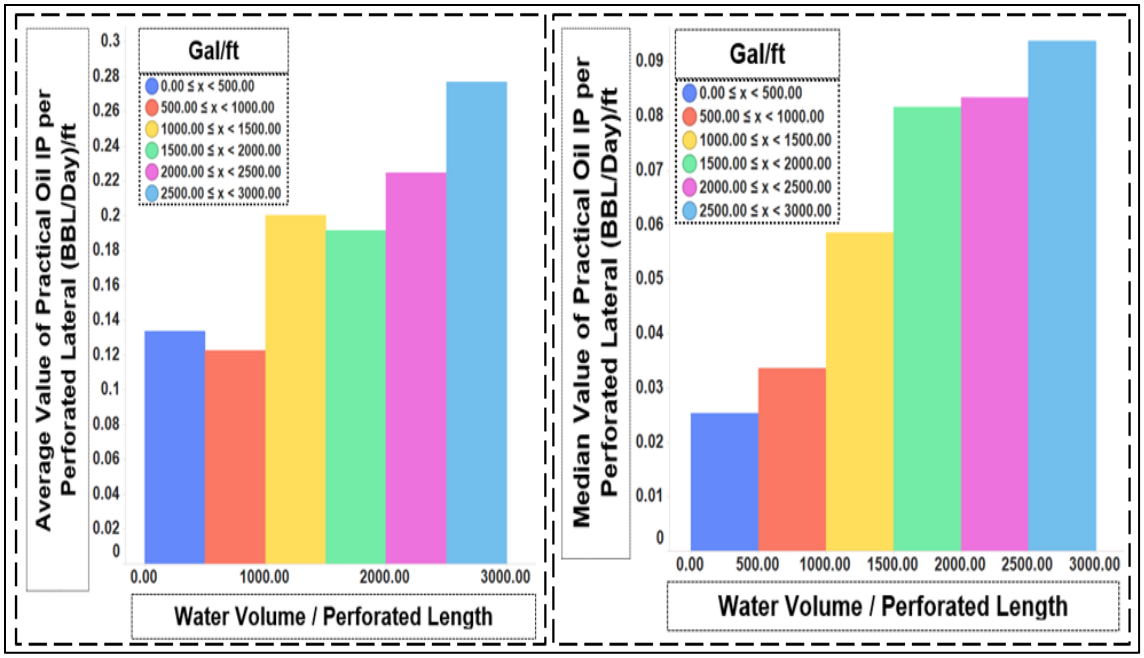


Figure 34. Water/Ft Effect on Average and Median Oil IP per Perforated Lateral

5. CONCLUSIONS

The following points will summarize the main conclusions that are observed about the Permian Basin completion and stimulation trends:

- Stimulation chemicals database were gathered from FracFocus. The proppant data were processed and checked for accuracy then combined with completion and production data acquired from DrillingInfo. The combined stimulation, completion, and production database were used to study the overall activates and trends in the Permian Basin.
- The study focused on the wells drilled and stimulated between the first quarter of 2013 and the second quarter of 2018.

- Both of the average and median values were presented and discussed in this study to eliminate the bias effect of the high and small range values on the average.
- Over time, the operators in the Permian have increased the total proppant mass and water volume pumped per well as well as the proppant mass per stimulated foot and water volume per stimulated foot.
- When the oil and gas prices crashed in Q4 2014, the wells count dropped and the industry focus was shifted towards drilling more horizontal wells compared to vertical drilling and the lateral length and stimulation volumes got bigger to increase the productivity while reducing the wells count.
- More than 70% of the drilled vertical wells were concentrated in the Central Basin Platform.
- Proppant loading in the Permian ranged at 0.5 to 1.5 Lbs/Gal. The lower proppant loading represents the well fractured with slick water and the higher proppant loading represents the wells fractured with gel and crossed linked fluids.
- The horizontal lateral length and the perforated length have increased with time. Many laterals in the past two years were drilled in excess to 10 thousand feet. After Q2 of 2016, more than 90% of the lateral length was perforated and stimulated.
- The trend of pad drilling has been adopted in the Permian Basin with more than 77 % of the wells drilled on multiple wells pads.
- The initial production in this paper was represented by the practical IP which is obtained by dividing the second month of production by the number of production days.

- The effect of water volume per foot on the initial production was investigated and the practical IP of gas and oil showed that the production got improved up to a value of 3000 Gal/ft.
- The effect of proppant mass per foot on the initial production was investigated and the practical IP of gas and oil showed that the production got improved up to a value of 3000 Lbs/ft.
- The practical IP per stimulated lateral were also investigated to test the improvement in production as a result of increasing normalized water and proppant volumes. The Prac IP/ft confirmed the benefits of increasing the stimulation volume up to the optimized values.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to DrillingInfo team and the co-founder and director of university outreach Mr. Mark Nibbelink for their generosity in providing access to the U.S. production and completion data.

REFERENCES

Arthur, J. D., Layne, M. A., Hochheiser, H. W., & Arthur, R. (2014, February 4). Spatial and Statistical Analysis of Hydraulic Fracturing Activities in US Shale Plays and the Effectiveness of the FracFocus Chemical Disclosure System. Society of Petroleum Engineers. doi:10.2118/168640-MS

Carman, P., Li, C., & Wheeler, R. (2018, September 24). Unorthodox Fluids for Unconventional Completions: Maximizing Production and Avoiding Pitfalls. Society of Petroleum Engineers. doi:10.2118/191751-MS

Carman, P., Li, C., & Wheeler, R. (2018, September 24). Unorthodox Fluids for Unconventional Completions: Maximizing Production and Avoiding Pitfalls. Society of Petroleum Engineers. doi:10.2118/191751-MS

DrillingInfo website. Retrieved on February/07/2019, from https://info.drillinginfo.com/

FracFocus website. Retrieved on April/02/2019, from http://fracfocus.org/

Khvostichenko, D., & Makarychev-Mikhailov, S. (2018, August 9). Oilfield Data Analytics: Linking Fracturing Chemistry and Well Productivity. Unconventional Resources Technology Conference. doi:10.15530/URTEC-2018-2903086

King, G. E. (2010, January 1). Thirty Years of Gas Shale Fracturing: What Have We Learned? Society of Petroleum Engineers. doi:10.2118/133456-MS

Luo, H., & Zhang, Y. (2018, October 5). Mining FracFocus and Production Data for Efficacy of Fracturing Fluid Formulations. Society of Petroleum Engineers. doi:10.2118/191787-18ERM-MS

Myers, R. R., Knobloch, T. S., Jacot, R. H., & Ayers, K. L. (2017, October 4). Production Analysis of Marcellus Shale Well Populations Using Public Data: What Can We Learn from The Past? Society of Petroleum Engineers. doi:10.2118/187505-MS

Olaoye, O., & Zakhour, N. (2018, September 24). Utilizing a Data Analytics Workflow to Optimize Completion Practices While Leveraging Public Data: A Permian Basin Case Study. Society of Petroleum Engineers. doi:10.2118/191416-MS

Romero, D. H., & Poston, S. W. (2016, May 10). Data Mining and Analysis - Eagle Ford Formation, South Texas. Society of Petroleum Engineers. doi:10.2118/179994-MS

Srinivasan, K., Ajisafe, F., Alimahomed, F., Panjaitan, M., Makarychev-Mikhailov, S., & Mackay, B. (2018, August 28). Is There Anything Called Too Much Proppant? Society of Petroleum Engineers. doi:10.2118/191800-MS

Weijers, L., Wright, C., Mayerhofer, M., Pearson, M., Griffin, L., & Weddle, P. (2019, January 29). Trends in the North American Frac Industry: Invention through the Shale Revolution. Society of Petroleum Engineers. doi:10.2118/194345-MS

Yang, Y., Robart, C. J., & Ruegamer, M. (2013, February 4). Analysis of U.S. Hydraulic Fracturing Design Trends. Society of Petroleum Engineers. doi:10.2118/163875-MS

IV. SHORT AND LONG-TERM PRODUCTIVITY COMPARISON OF HYDRAULIC FRACTURING FLUID SYSTEMS IN EAST TEXAS COTTON VALLEY FORMATION

Mustafa A. Al-Alwani, Larry K Britt, Shari Dunn-Norman, Husam H. Alkinani, Abo Taleb T. Al-Hameedi, Miguel Cedeno, Atheer M. Al-Attar, Waleed Al-Bazzaz

Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA

ABSTRACT

The aim of most hydraulic fracturing treatment is to design and implement the best stimulation strategies that help to increase the short and long-term productivity of the stimulated reservoir section. This can be achieved by selecting the best treatment fluid types that are compatible with the formation to achieve the best combination of fracture geometry and well's productivity. The main objective of this study is to investigate the short and long-term wells' productivity for the different types of hydraulic fracturing treatment fluids in the East Texas Basin Cotton Valley formation.

Two hundred and sixty one Cotton Valley wells were identified and selected for this study. All the wells were horizontal and hydraulically fractured that were completed after 2012. The wells were classified based on the type of hydraulic fracturing fluid into four major groups: gel, water, cross-linked gel, and hybrid fracs. The hydraulic fracturing chemicals were obtained from FracFocus public record database. FracFocus data were processed then combined with production and completion data obtained from DrillingInfo database. Several data workflows were developed to classify the treatment fluid types based on the chemical ingredients of each fracturing treatment.

The productivity in this paper is reported in terms of equivalent barrel of oil which converts every 6 MCF of gas into 1 equivalent barrel of oil. This paper investigates the Cotton Valley initial production (BOE/Day), cumulative 6 months, 1 year, 2 years, and 5 years and compares the productivity for each type of hydraulic fracturing fluids. This work also provides several insights about the wells performance as a function to stimulation parameters such as the amount of pumped proppant, the volume of pumped water, and the length of the perforated horizontal lateral. The normalized production and stimulation parameters to the perforated lateral are also discussed in this paper.

This study discusses several data processing workflows that will help to illustrate the procedures to extract data from FracFocus and DrillingInfo. It provides an understanding of the hydraulic fracturing fluid types' occurrence and measures the effect of stimulation and completion parameters on the short and long-term productivity.

1. INTRODUCTION

Advancements in horizontal drilling technology coupled with the ability to economically place repeated hydraulic fracturing stimulations along the horizontal well have driven the growth in developing unconventional resources as a leading source of oil and gas to the energy supply in the United States. The resulting boom in drilling and stimulation have led to an accumulation of a vast amount of drilling, completion, stimulation, and production data. Gathering and analyzing all of the accumulated data aid in optimizing the productivity of the wells and identify the greatest economic exploitation methods for future well drilling and completion.

A large number of industries have been applying data descriptive and predictive approaches in their operations' optimization in addition to decision making and future planning. Data analytics have been defined as one of the highest growing fields within the oil and gas industry (Khvostichenko and Makarychev-Mikhailov, 2018). Many of the operators and service contractors have realized the imperative of data-driven decisions and started to appropriate many of their resources to expand their data utilization in everyday decisions. The literature shows an abundance of articles that depicts the growing interest in data analysis studies. The output of this kind of studies are very insightful and shows a wide range of applicability (Yang et al., 2013; Arthur et al., 2014; Romero and Poston, 2016; Myers et al., 2017; Luo and Zhang, 2018; Olaoye and Zakhour, 2018; Srinivasan et al., 2018; Carman and Wheeler, 2018; Weijers et al., 2019, Al-Alwani et al., 2019a; Al-Alwani et al., 2019b).

In hydraulic fracturing, fluids ladened with proppant, are injected under high pressure to create a fracture which provides a pathway for oil and gas to readily flow into a well. There are four stages found in most stimulation treatments. The spearhead pre-fracturing stage or it can also be called as the acid stage involves pumping water with diluted acid to clean and clear the perforations and the near-wellbore area from any potential debris. The next stage is the pad stage which involves pumping the mixed water and chemicals fluid under a certain pressure to initiate and propagate the fracture. The following stage is the slurry or the proppant stage which involves the pumping of a mix of hydraulic fracturing fluid and propping agents. The proppant will be transported into the fracture and remain in the fractures to keep them propped open. The final and last stage is the flush stage which involves pumping clean base fluid to displace any proppant from the wellbore.

There are several types of hydraulic fracturing fluids. Water fracs consist of water, friction reducer, and may also include some clay control agents and other chemical ingredients. Gel fracs (linear gel) includes water, gelling agents and other chemical components. Example of gelling agents can be guar, hydroxypropyl guar (HPG) or hydroxyethyl cellulose (HEC). Gelling agents can be prone to bacterial growth; therefore, biocides are usually added to the gel fluid to eliminate bacterial problems. Cross-linked gel consists of water, gelling agents and cross-linking agents. The use of cross-linked gel fracturing fluid requires adding breakers to the mixed fluid to help break the cross-linked gel fluid in the fractures and flow back to the surface once the job is completed. In some of the stimulated wells, a combination of the three different types of hydraulic fracturing fluids can be used. For example, treated water can be pumped in some stages then followed by linear gel in the other stages. Another example can be a linear gel to cross-linked gel. The combination of different fluid types used in the same fracturing job classified as hybrid fluid fracturing.

The Cotton Valley formation is contained in the East Texas Basin. It is a large hydrocarbon-bearing formation spanning across the eastern part of Texas and the western part of Louisiana. It is classified as a tight gas sandstone. The formation is comprised of very fine to fine grains of sandstone that are tightly cemented. It is also interbedded with carbonate, siltstone, and mudstone (Westcott, 1984). The overall porosity and permeability are very low and the wells are stimulated with different types of hydraulic fracturing fluid treatments. The majority of the reservoir intervals contain gas and the production is mainly gas associated with liquid hydrocarbon and water (Liu et al., 2011).

The ultimate objective of this study is to compare both short- and long-term Cotton Valley stimulated well performance as a function of different hydraulic fracturing fluid treatments. The publicly available FracFocus database was used to identify the type of hydraulic fracturing fluids used in wells. Data obtained from FracFocus were combined with commercially available well production data (DrillingInfo, 2019). The following discussion presents details regarding the data handling and workflows created to uniquely identify hydraulic fracturing treatments, and then associate those treatments with each selected Cotton Valley well's production history. It also compares the Cotton Valley well performance as a function of fluid volumes, proppant loads, normalized on lateral well length.

2. APPROACH

A core requirement of this study was to establish a methodology to utilize the publicly accessible FracFocus database and establish procedures to extract the data and transform it into usable inputs. For example, FracFocus provides each single chemical treatment applied in a well as a separate row of data. Identifying all stimulation treatments associated with the primary completion in a particular well requires some aggregation method to be applied to the rows of data. This is but one example of numerous such adjustments required to clean and format FracFocus data. Importantly, an understanding of unconventional well hydraulic fracturing fluids and procedures is really required to process these data and ensure a robust and verified database.

FracFocus database contains chemical data for wells completed after 2011, when the FracFocus 1.0 registry began. However, it is important to note that reporting methods varied over time and many wells completed in the earliest time period may not have complete data. These wells were excluded from this study.

The FracFocus data were then merged with the completion and production database of DrillingInfo. Merging the two databases provides the ability to associate hydraulic fracturing fluid treatments in a well with its drilling and completion parameters and production history.

The following subsections provide additional detail regarding the technical approach used to process the data. This discussion can serve as a guide for future FracFocus data mining and processing.

2.1. HANDLING RAW FRACFOCUS DATA FOR THE U.S. WELLS

More than 4 million rows of chemical ingredients data were downloaded from the FracFocus website. The data are available online to be downloaded in Excel files. Due to the limitations of the number of rows that one Excel sheet can accept (maximum of 1,048,576 rows) the Fracfocus data were downloaded and condensed into 5 Excel files, which were then combined and processed together. Once all the FracFocus data were stacked together, the following step was to identify parameters to be included in the database (e.g. well location, ingredient name, water volume, etc) and ignore the unnecessary inputs (e.g. file upload key). The formats of some data columns were adjusted for further use. For example, the date data were adjusted from FracFocus format of mm/dd/yyyy hh:mm to mm/dd/yyyy and well location data (surface longitude and latitude) were used to create centroid points to map the wells' surface location on the map. After all the data merging and formatting, the selected columns were cleaned from any leading or trailing white spaces and then saved in the format of a tab-delimited text file to be used in the next phase of the data processing. Figure 1 illustrates the FracFocus initial data processing workflow for the entire U.S. database.

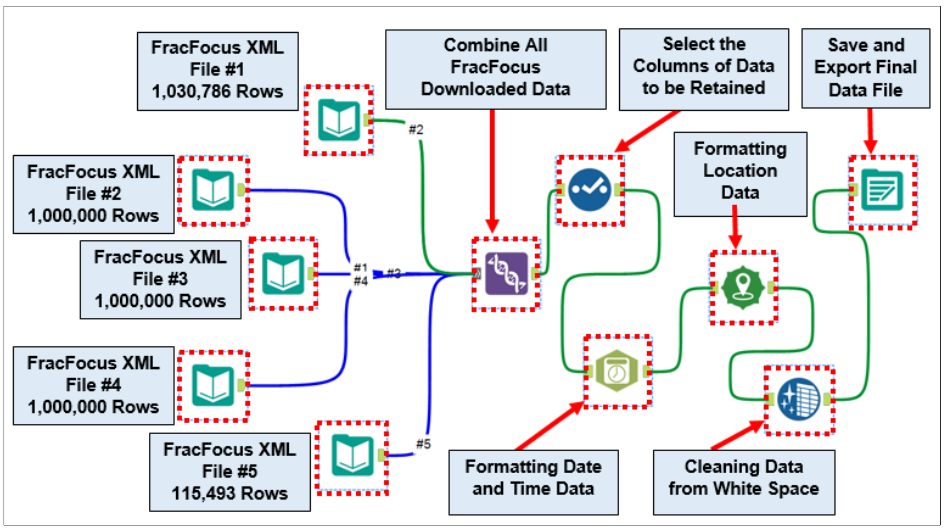


Figure 1. Data Processing Workflow Illustrating the Initial Steps to Process FracFocus Data Files

2.2. GROUPING FRACFOCUS CHEMICAL INGREDIENTS

One of the biggest challenges in using the FracFocus data is that the data are presented in a stacked format. For a single well report, there are several rows of the same chemical groups that represent different mass percentage of a chemical ingredient. This work classifies all the chemicals into 19 groups and then process each group of chemicals separately to aggregate the total mass percentage of each group’s chemical used in a particular well. The chemical groups designated are as follows: water, proppant, gelling agents, cross linkers, breakers, friction reducers, surfactants, biocides, clay control, acid, pH adjuster, corrosion inhibitor, iron control, scale inhibitor, non-emulsifiers, solvents, liquid N2, others, and other non-CAS chemicals.

Classifying FracFocus data into the aforementioned groups is not an easy task and requires knowledge of hydraulic fracturing treatments applied in unconventional oil and gas well. It also requires data analysis and grouping methods. In FracFocus, there are no standards in reporting the chemicals' name or function which makes it impractical to try to classify the ingredients based only on the name or the purpose. To properly classify the chemical groups, a combination of the chemical ingredient name along with the chemical abstract service number (CAS-Number) was used. CAS-Number is a unique number assigned for any chemical substance and each CAS-Number linked to a list of information that describes the chemical substance. A treemap chart technique was applied in a hierarchy style to group chemicals. The chemicals that are heavily present in the highest number of fractured wells are listed on top of the hierarchy. The ingredient name will be assigned based on the most frequently mentioned name among all the wells which was associated with a specific CAS-Number. Applying the "most common" function in assigning the ingredient name eliminates the FracFocus pitfalls of naming the chemicals in different styles and spelling errors. The size of the box in the treemap is based on the count of the number of the wells that the CAS-Number existed.

Figure 2 illustrates a treemap chart with the basic setup steps to create the chart based on the FracFocus database. In this illustration, the CAS-Number 14808-60-7 is the unique number for sand that is used as a proppant in hydraulic fracturing. Sand is on top of the hierarchy followed by water (7732-18-5) because they comprise the greatest mass percent ingredients in all the stimulated wells. Based on the CAS-Number all the chemicals were grouped into one of the 19 groups. Two sources of data check were used to validate the accurate function of each CAS-Number in order to assign it to the appropriate group. Environmental protection agency (EPA) safer chemical ingredients list and a study conducted by Helmholtz Zentrum München, Institute of groundwater ecology which lists complete list of all hydraulic fracturing chemicals extracted from the Waxman list and the FracFocus database (Elsner & Hoelzer, 2016; EPA, 2019).

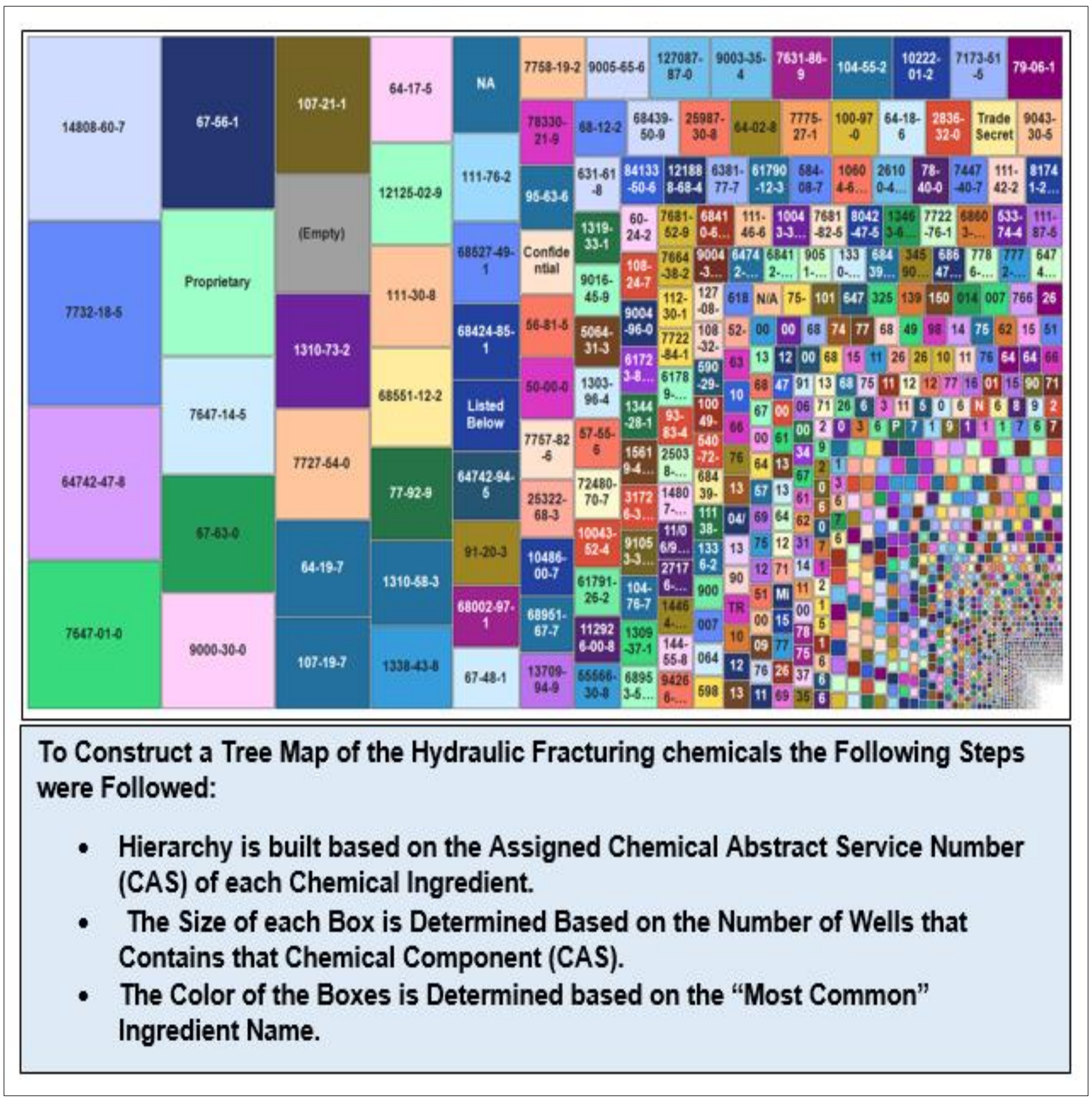


Figure 2. Treemap Chart Utilized to Group FracFocus Chemical Ingredients

2.3. PROCESSING AND AGGREGATING EACH CHEMICAL GROUP

Once the 19 chemical groups were created, each group was processed separately to inspect for duplications in data entry within FracFocus reports. Once duplicate rows were eliminated, an aggregation function was utilized to add up the chemicals' mass percentage for every well based on the unique well API number. It is worth mentioning that for the same well, the same chemical can be mentioned in several rows with different mass percentages or the same group of chemicals can contain different ingredients' names and CAS-Numbers, yet serve the same purpose (e.g. surfactants). The workflow also inspects for refracturing operations and assigns a fracturing job number for each well. For example, if the same well's API number appears to have two separate fracturing dates and different chemical components associated with each date, this means that the well has been refractured and based on how many different report dates available, the well stimulation treatments will be numbered in ascending order with the primary completion designated as '1' and associated with the first fracturing job start date.

Figure 3 illustrates the data processing workflow to calculate the total mass percentage of chemicals for a stimulation treatment. There are two output files generated in this workflow. The first one will report the summed-up mass percentages of all chemical groups per well and the second one will keep all the original components without adding the mass percent together. The first file output is used in this study to calculate all the actual mass and concentrations of all the classified groups while the second file output can be used to identify the different chemical components and their mass percentages for example, all the different proppant types and sizes can be identified for every treated well.

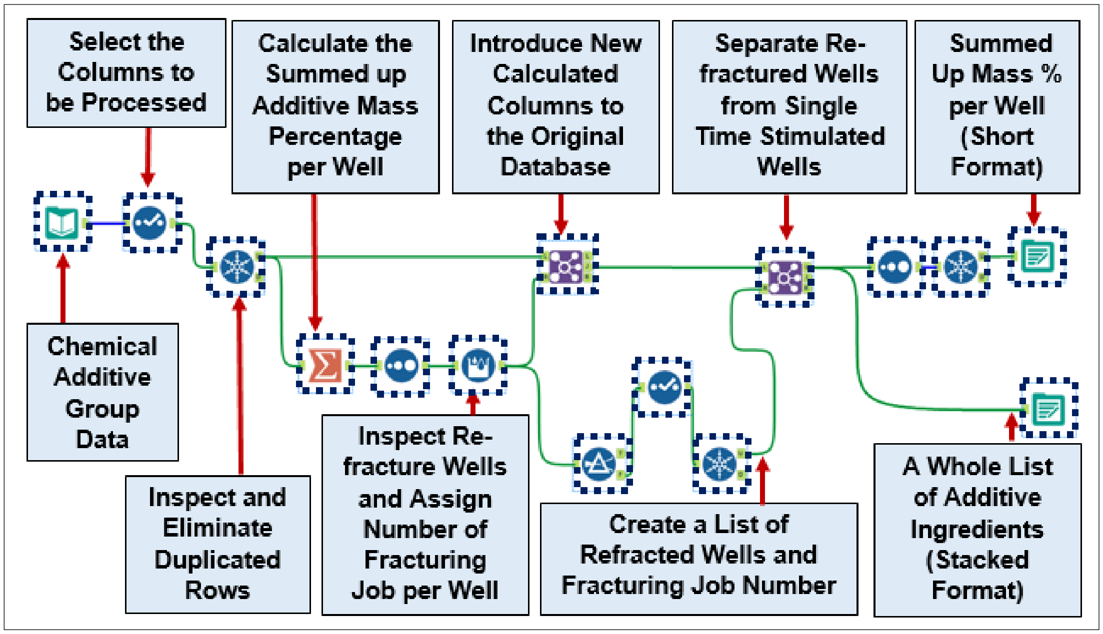


Figure 3. Data Processing Workflow to Aggregate the Total Mass % of Chemical Ingredients per Well

2.4. CREATING PROCESSED FRACFOCUS DATABASE

Figure 4 shows the steps of building the processed FracFocus database by merging and combining the 19 chemical groups altogether. The workflow starts by joining the water and proppant data based on the API number. In this step, the workflow investigates the well's API number and joins the data of the water group with the data of the proppant group. If one well has only proppant and no water data, the data will be held aside into a separate output, the same is true for the wells with only water and no proppant data. After combining the water and the proppant data into a new output, the proppant mass in pounds will be calculated based on the total mass percentage, the water mass percentage, and the total water volume. As a check, the actual reported proppant mass was compared with reported proppant mass. A 98% match indicated the mass percentage aggregating workflows were setup properly. The other remaining chemical groups were introduced into the workflow using union tools which means that only the wells that have a mass percentage of that particular chemical group will obtain a mass percentage value and the remaining wells will be assigned a zero value for the lack of that chemical group into the stimulation fluids recipe. The line break in the workflow diagram represents the entry of all the remaining chemical groups. Once all the chemical groups were introduced and matched together, several new columns were introduced to the database by using the formula options to perform calculations to obtain the mass in pounds for all the chemical groups as well as the total concentration of the chemicals by dividing the mass of the material by the liquid phase base fluid volume. Finally, the total hydraulic fracturing percentage recipe was calculated to compare the accuracy of the process.

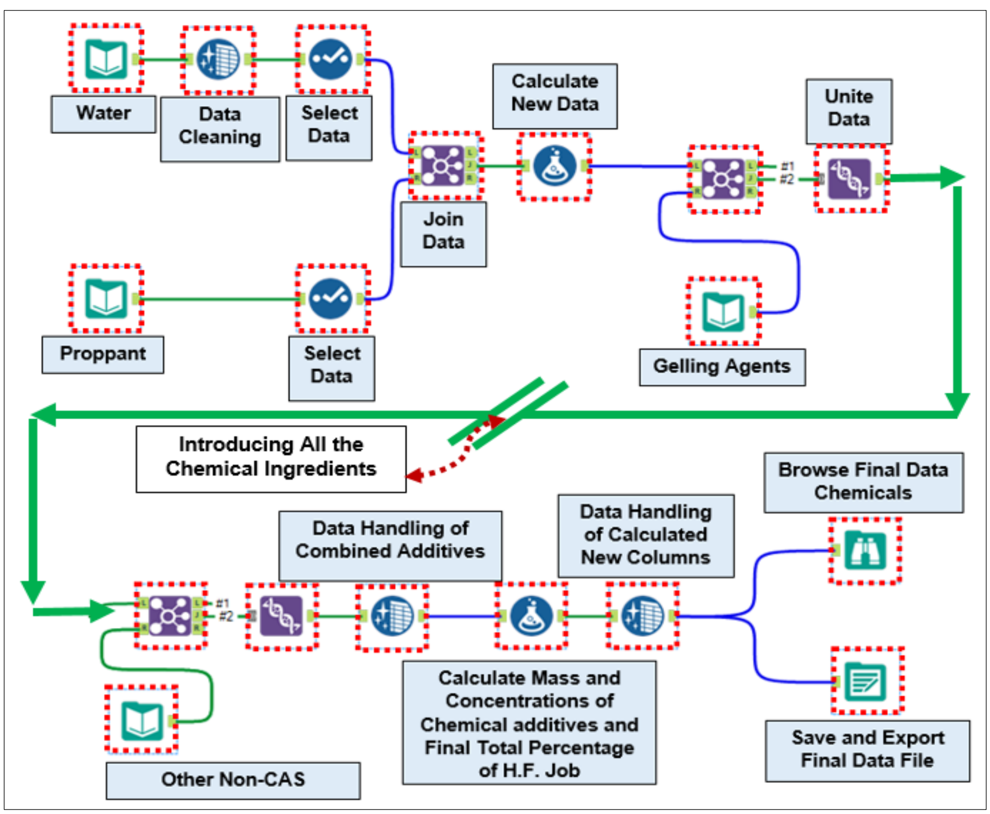


Figure 4. Data Processing Workflow Illustrating the Steps of Building the FracFocus Processed Database

2.5. INTEGRATING FRACFOCUS DATA WITH DRILLINGINFO DATA

The developed FracFocus database was merged with the DrillingInfo database to generate a comprehensive database that enables comparisons between the stimulation and completion parameters and their impact on production and well performance. Figure 5 illustrates the data merging workflow where all the screened FracFocus data were matched by the API number with the DrillingInfo database. The wells in this database are populated between the years 2012 and 2018.

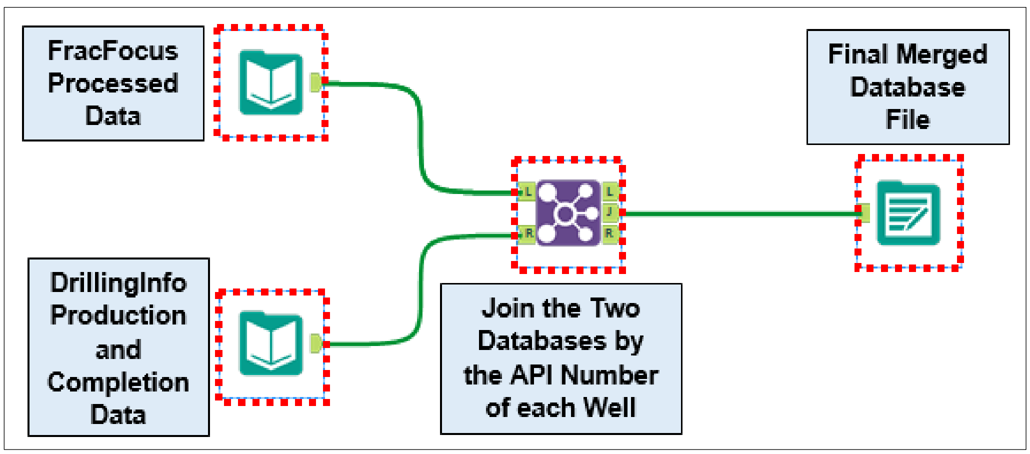


Figure 5. Data Processing Workflow Merging Final FracFocus Processed Chemical Database with DrillingInfo Production and Completion Data

2.6. CLASSIFYING STIMULATION TREATMENT FLUIDS

The hydraulic fracturing fluid types were classified as water treatment, gel, cross-linked or hybrid based on the presence of key chemical ingredients: gelling agents, cross-linking agents, and breakers. A formula was integrated into the workflow to perform the classification based on knowledge of the four fluids types and their application. For example, the presence of a chemical breaker or a cross-linking chemical would indicate the fluid was cross-linked. Figure 6 shows the classification workflow with the built-in conditional statement classification tool.

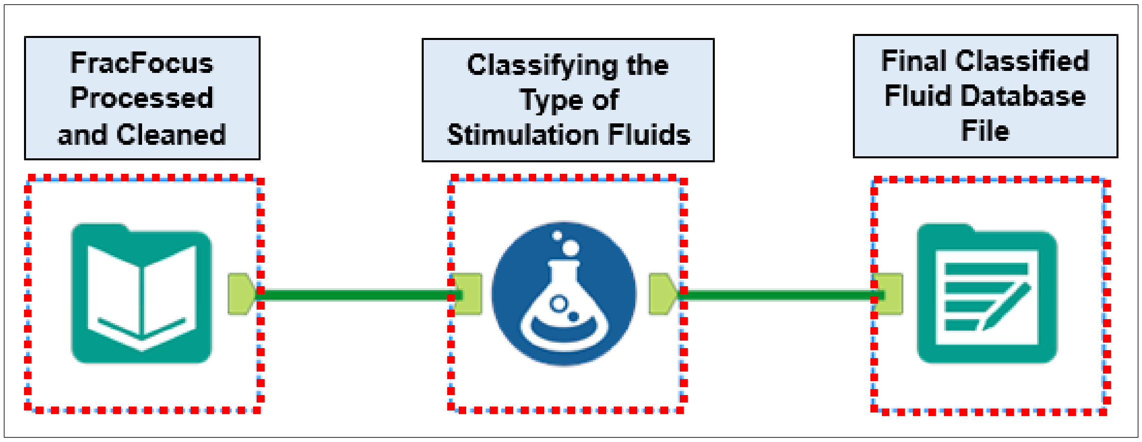


Figure 6. Data Processing Workflow to Assign Stimulation Fluid Treatment Type for Each Well Based on the Chemical Ingredients' Percentage

3. RESULTS AND DISCUSSION

This section discusses the identified Cotton Valley wells including their treatment fluid types, completion trends, and the initial production and cumulative recovery with time for Cotton Valley wells stimulated with each type of hydraulic fracturing fluid. Box plots and statistical summaries of production and some completion data are also provided.

3.1. COTTON VALLEY TIGHT GAS PLAY DATA COVERAGE

In this study, 261 horizontal wells completed between 2012 and 2018 have been selected. The data cover most of the Cotton Valley tight gas play across most of the active counties in Texas and Louisiana. Figure 7 shows a pie chart with all the major counties and the percentage of the wells that span across them. Figure 8 shows a pie chart with the percentage of the four treatment fluid types. As shown, more than 41% of the Cotton Valley wells were treated with cross-linked fracturing fluids. 28% of the wells were treated with hybrid fracturing fluid, which means that the wells have been fractured in some stages by water then gel fluid or gel and crossed-linked or any combination of the treatment fluids. Linear gel was used to treat about 15.7% of the wells, and water frac fluids were utilized to treat less than 15% of the wells. Figure 9 pictures a map of the study area focusing on the wells in Texas and Louisiana. The wells are color-coded based on the treatment fluid type, i.e. water wells are colored with blue, gel with green, crossed-linked gel with red, and hybrid wells with yellow. Each state on the map was colored differently to make the border between the states more visible.

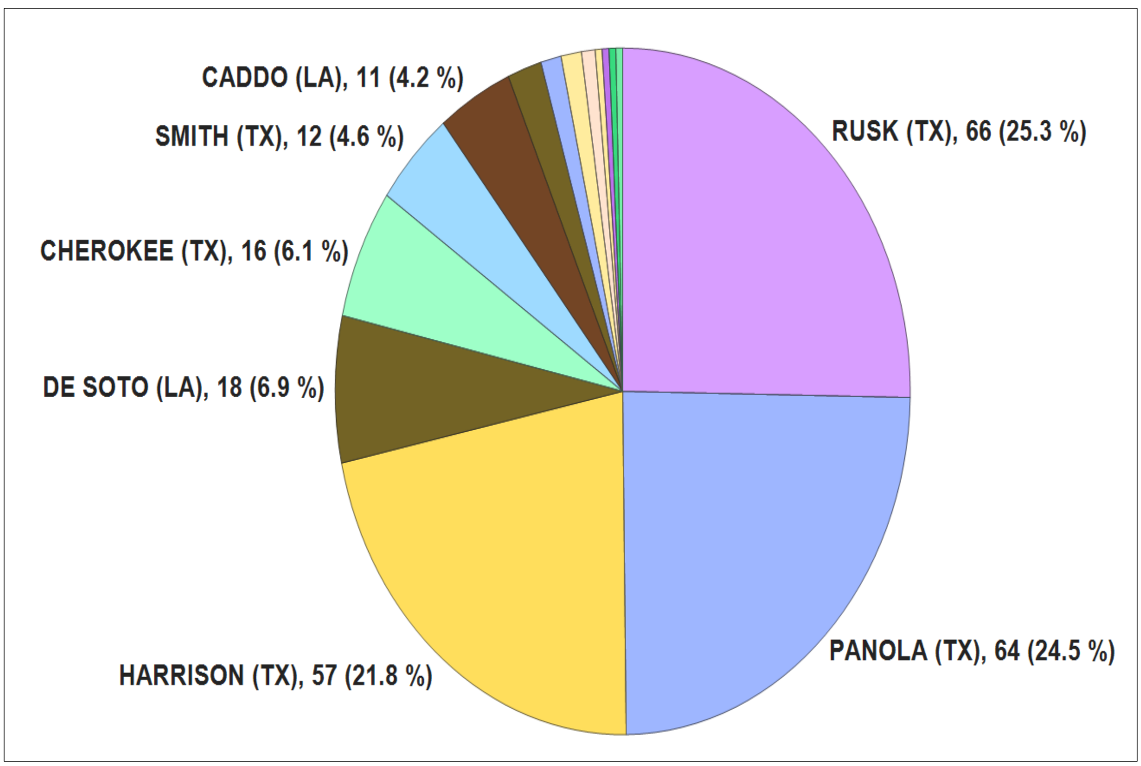


Figure 7. Pie Chart of the Counties and the Wells' Distribution Percentages of the Studied Area

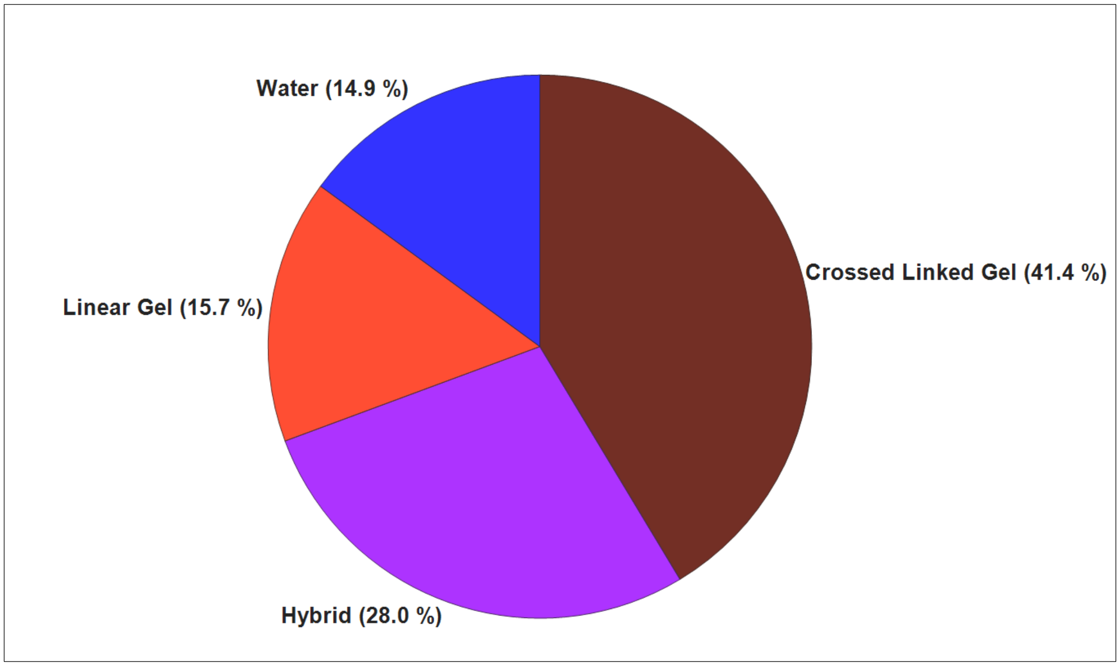


Figure 8. Pie Chart of the Stimulation Fluid Type Percentages of the Studied Area

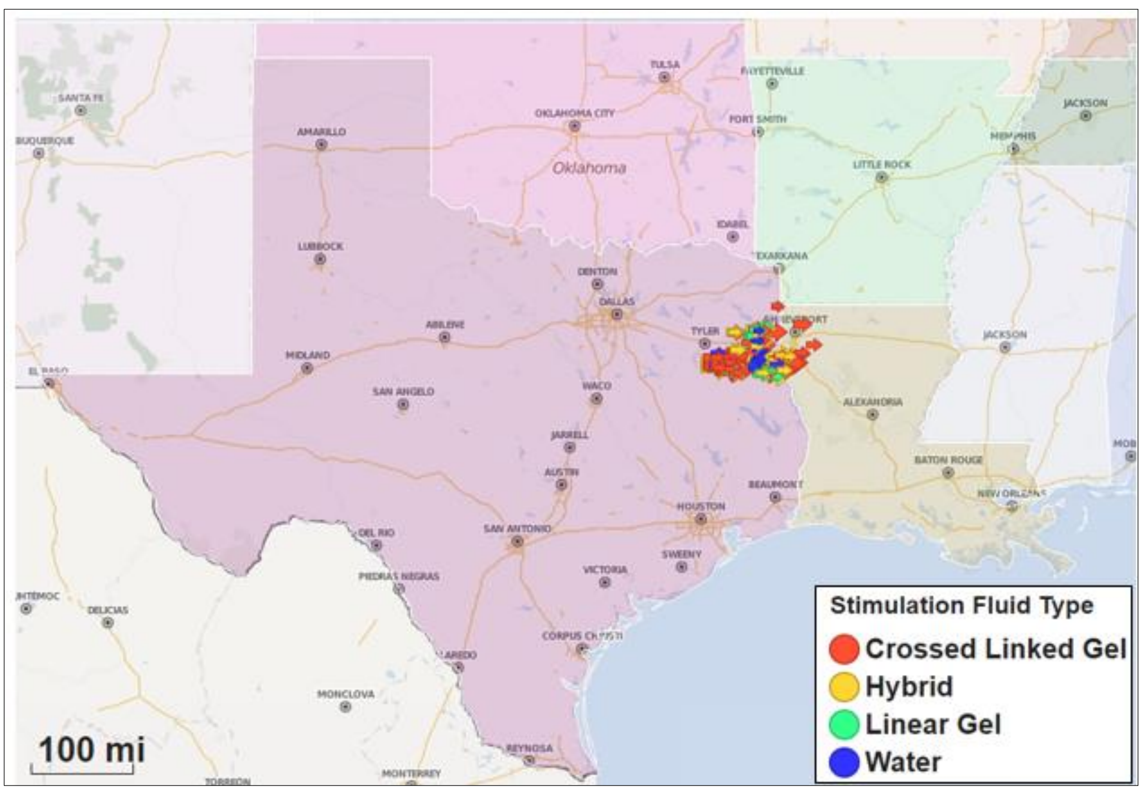


Figure 9. East Texas Cotton Valley Map Showing Stimulated Wells which are Colored Based on Treatment Types

3.2. TREATMENT FLUID TYPES AND WELL PERFORMANCE IN THE COTTON VALLEY PLAY

The primary focus of this work is to investigate the Cotton Valley formation well production performance as a function of time for the four hydraulic fracturing fluid types. User-Friendly dashboards were developed and linked to the database previously described. The dashboard facilitates creating a wide variety of figures to illustrate well performance as a function of completion parameters.

Figure 10 shows the average total stimulation base fluid volume per well for each hydraulic fracturing fluid type. To take into consideration the difference in lateral length between the wells, the stimulation fluid volume was normalized to the perforated lateral length by dividing the total stimulation volume per well by the lateral length as shown in Figure 11 This analysis shows that wells treated with hybrid fluid consumed more stimulation fluid than the other treatment fluid types and the normalized treatment fluid volume per stimulated length showed that cross-linked gel wells consumed more fluid per foot in comparison to other types of fluids.

Figure 12 shows the amount of pumped proppants per well for each fluid type. Wells treated with cross-linked gel fluid type used the highest average proppant pounds per well. Figure 13 shows proppant mass normalized to the perforated lateral length. In this analysis, wells fractured with cross-linked gel consumed more proppant per foot compared to the other types, whereas linear gel fluids exhibited the highest proppant load in lb/gallon as shown in Figure 14 Wells treated with water were associated with the lowest proppant loading which means the wells treated with water frac used less proppant compared to other treatment types. Figure 15 shows the average perforated lateral length for all the fluid types in this study. The longest laterals were stimulated with linear gel fluids. The shortest average perforated lateral length was stimulated with water fracturing.

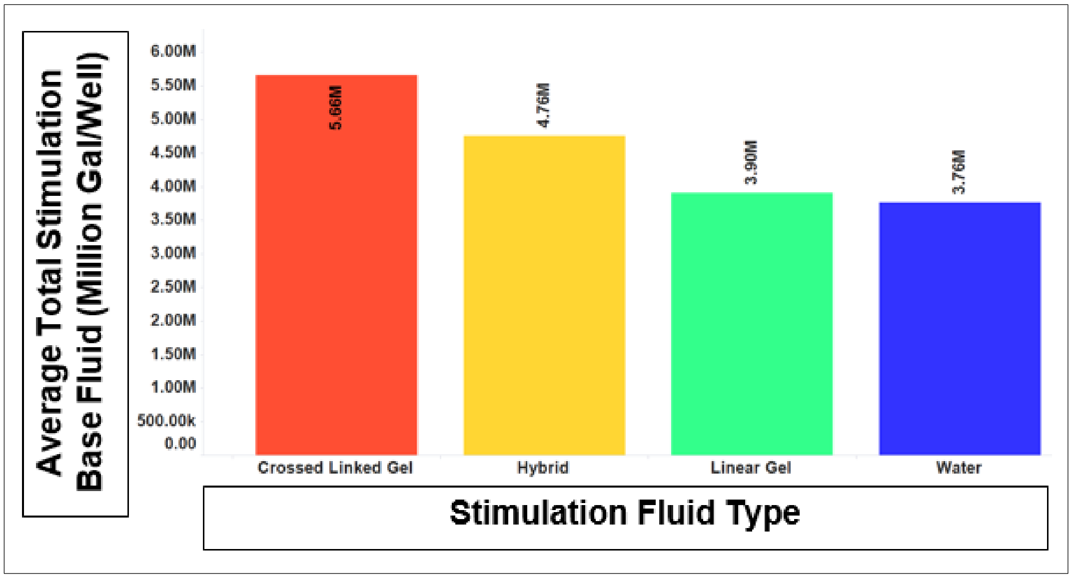


Figure 10. Bar Chart Comparing the Cotton Valley Average Stimulation Fluid Volume Usage per Well for Different Stimulation Fluid Types

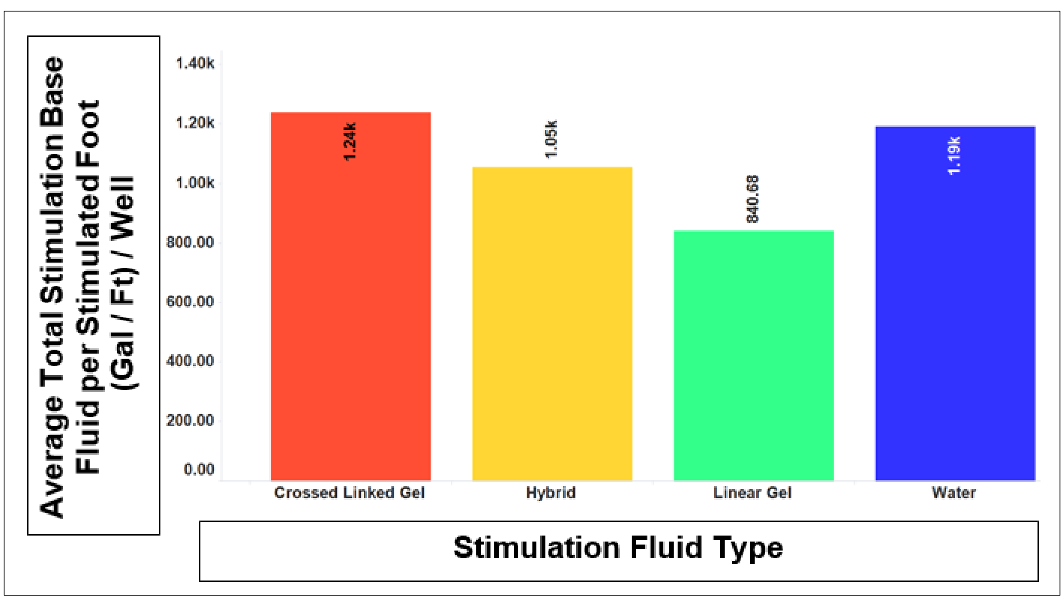


Figure 11. Bar Chart Comparing the Cotton Valley Average Stimulation Fluid Usage per Stimulated Foot per Well for Different Stimulation Fluid Types

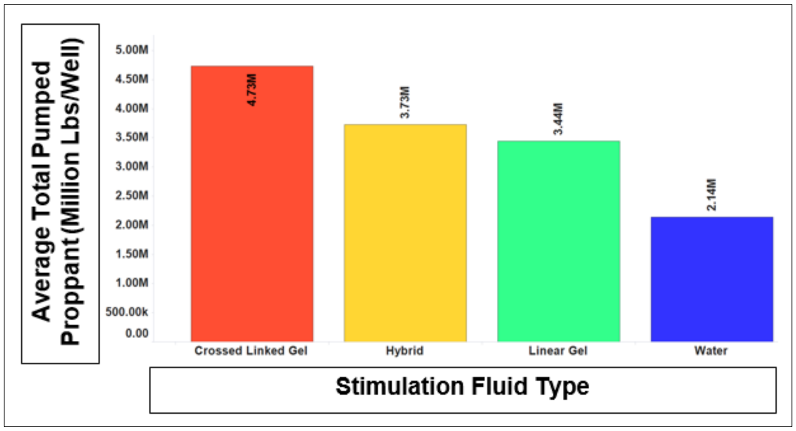


Figure 12. Bar Chart Comparing the Cotton Valley Average Total Proppant Usage per Well for Different Stimulation Fluid Types

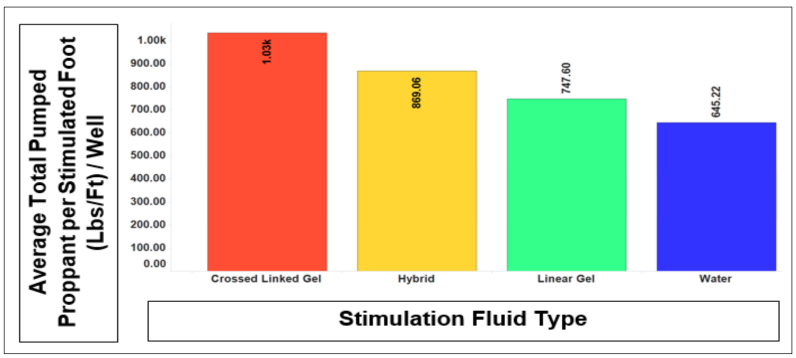


Figure 13. Bar Chart Comparing the Cotton Valley Average Total Proppant Usage per Stimulated Foot per Well for Different Stimulation Fluid Types

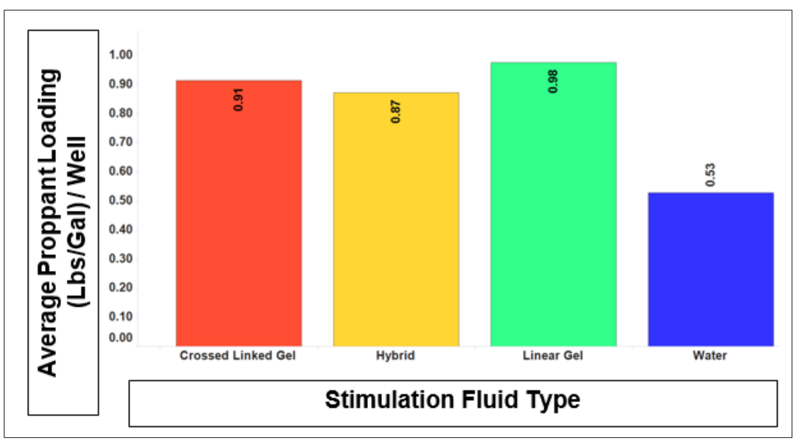


Figure 14. Bar Chart Comparing the Cotton Valley Average Proppant Loading per Well for Different Stimulation Fluid Types

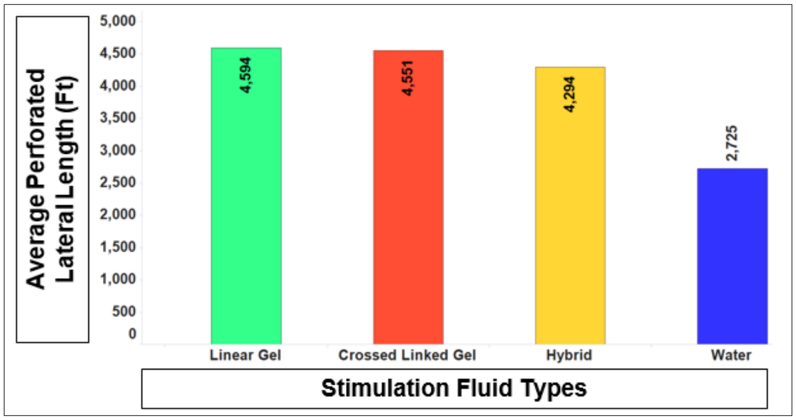


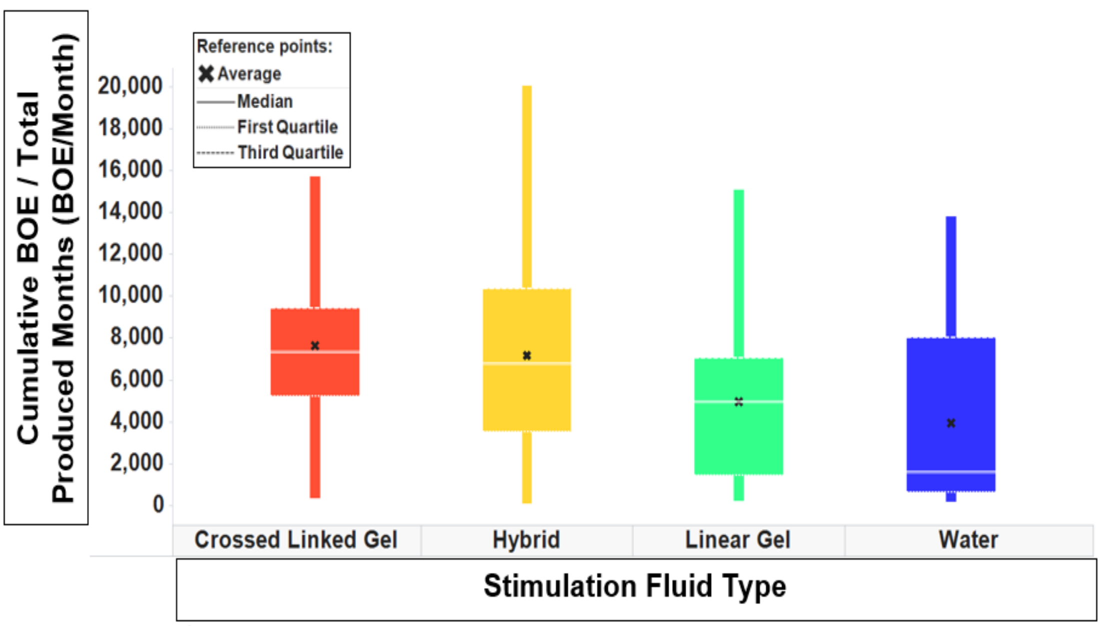
Figure 15. Bar Chart representing the Cotton Valley Average Perforated Lateral Length for The Different Types of Stimulation Fluids Between (2012-2018)

3.3. BOX PLOTS AND STATISTICAL TABLES OF THE COTTON VALEY PRODUCTION DATA

Figure 16 to Figure 24 provide the production and completion data distributions using box plots along with the associated summary statistical table. Each fluid type's box plot is color-coded based on the type of stimulation fluid. The average value is marked as an x mark on the box plot. The median value is marked as a white solid line and the first and third quartiles are marked as dotted lines. The statistical tables shows the count of number of wells per stimulation type, the average value, median value, the number of outliers, the value of the first quartile, the value of the third quartile, the value of the first 10% of the data and the value of the 90% of the data. Since this study compares the production and completion data as well as some normalized production per foot parameters for different types of hydraulic fracturing fluids, these boxplots were chosen to show all the production and completion data distribution.

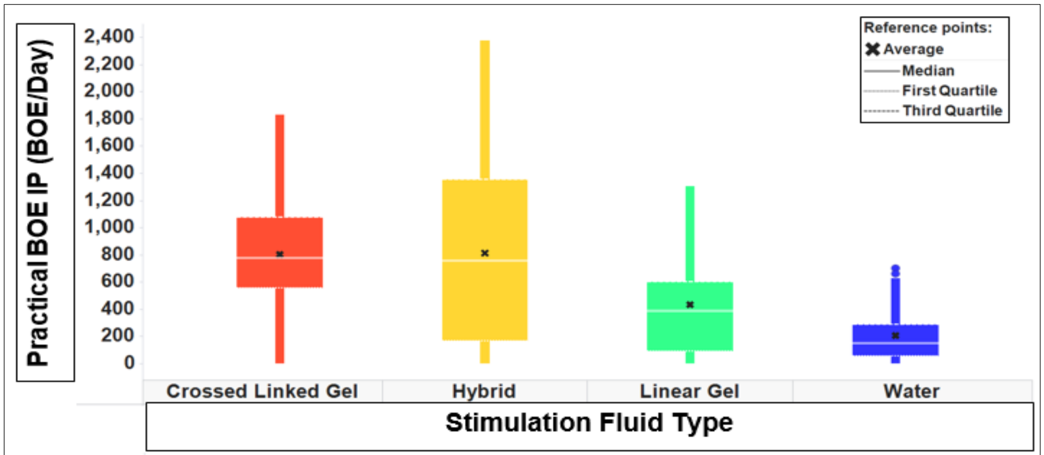
Finally, all the suspected outliers were cleaned out and removed from this dataset to make the data comparison more reliable.

Figure 16 shows a box plot and statistical table of the cumulative production in barrel of equivalent oil (6 MCF=1BOE) divided by the number of produced months in (BOE/month). Figure 17 shows a box plot and statistical table of the practical IP (BOE/Day). The box plots and statistical tables of the cumulative BOE production for the first 6 months, first year, two years, and 5 years are presented in Figure 18, Figure 19, Figure 20, and Figure 21, respectively. Lastly, the box plots and statistical tables of total proppant, total water, and perforated lateral length are shown in Figure 22, Figure 23, and Figure 24, respectively.



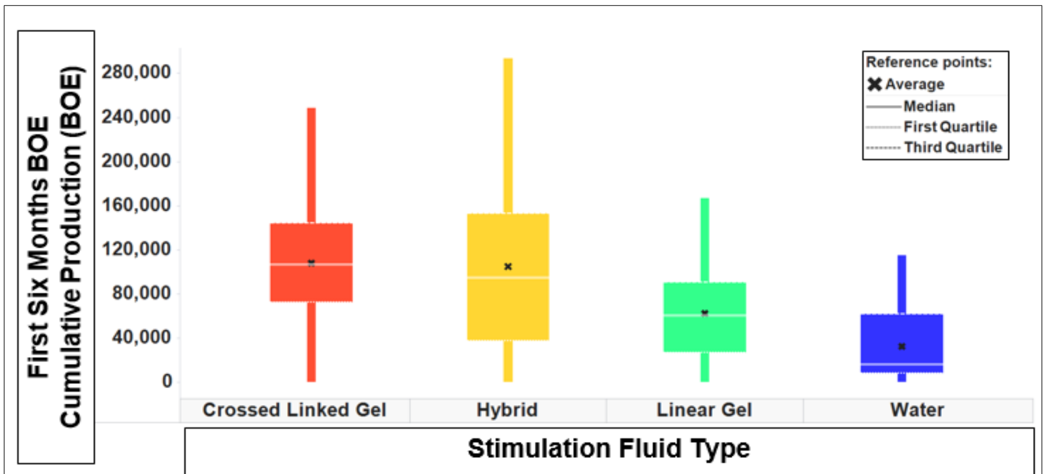
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	106	71	38	33
Avg	7,627	7,174	4,967	3,960
Median	7,335	6,800	4,970	1,592
Outliers	0	0	0	0
Q1	5,213	3,510	1,440	638
Q3	9,417	10,334	7,057	8,003
P10	3,796	547	414	414
P90	13,147	13,350	8,880	9,867

Figure 16. Box Plot and Statistical Table of Cotton Valley Cumulative BOE Production per Total Produced Months



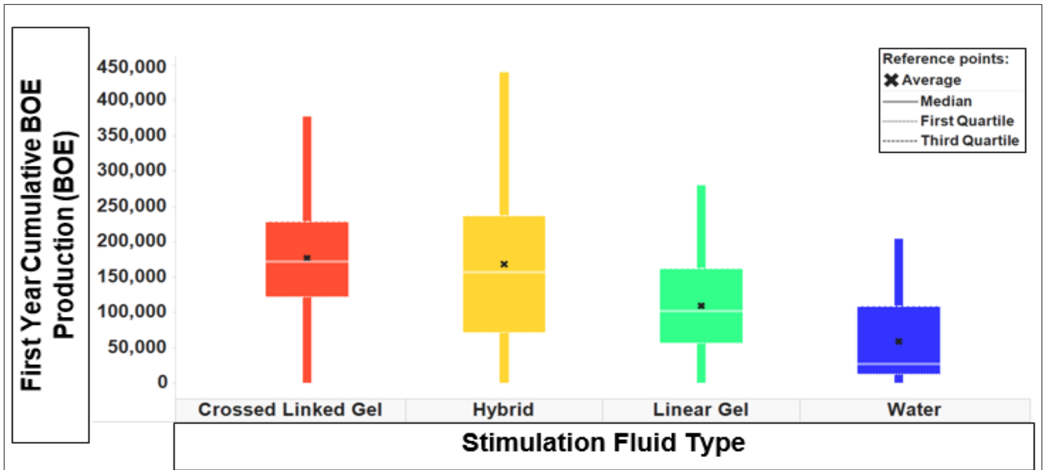
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	108	72	40	38
Avg	802	816	433	207
Median ▶	778	759	389	153
Outliers	0	0	0	2
Q1	558	167	90	54
Q3	1,079	1,353	604	289
P10	254	88	51	0
P90	1,377	1,728	987	520

Figure 17. Box Plot and Statistical Table of Cotton Valley Practical BOE Initial Production



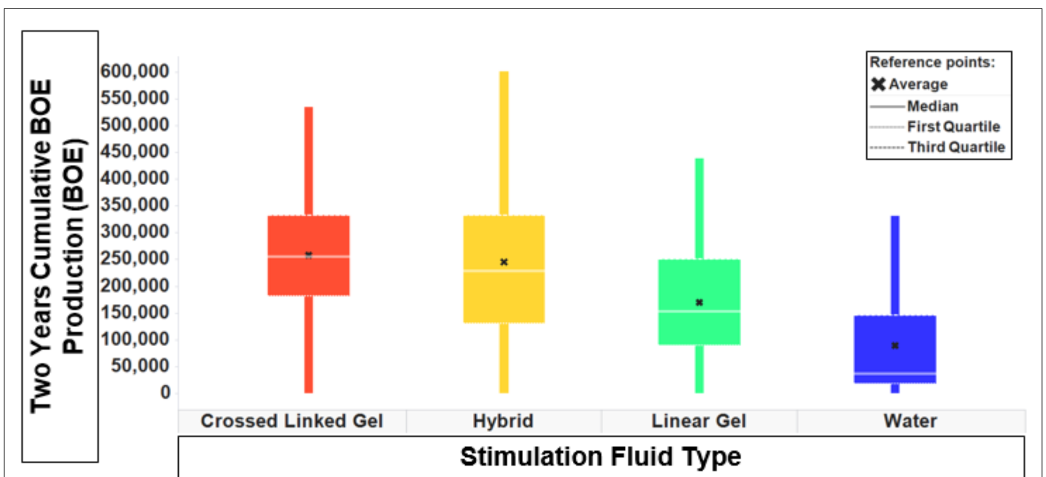
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	107	72	40	37
Avg	107,935	105,100	62,495	32,552
Median ▶	106,953	94,588	60,520	16,025
Outliers	0	0	0	0
Q1	72,526	37,691	26,871	8,320
Q3	143,909	153,217	90,729	61,807
P10	52,106	10,779	7,666	0
P90	162,786	219,829	116,410	89,983

Figure 18. Box Plot and Statistical Table of Cotton Valley First Six Months Cumulative BOE Production



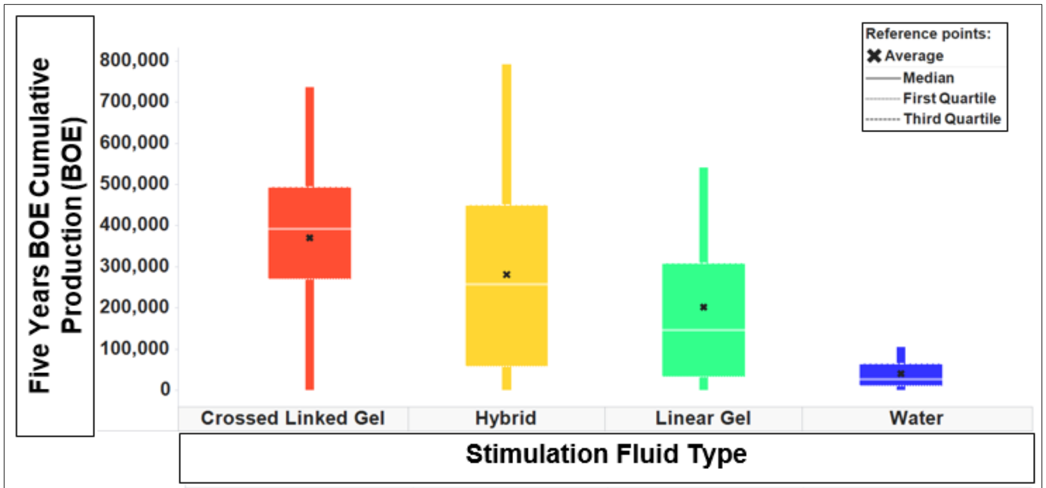
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	103	70	40	37
Avg	176,903	168,578	109,635	59,309
Median	172,128	157,360	102,205	26,861
Outliers	0	0	0	0
Q1	121,787	70,731	56,368	12,443
Q3	227,768	236,317	161,765	108,671
P10	91,421	16,107	10,872	0
P90	262,750	342,334	206,401	175,057

Figure 19. Box Plot and Statistical Table of Cotton Valley First Year Cumulative BOE Production



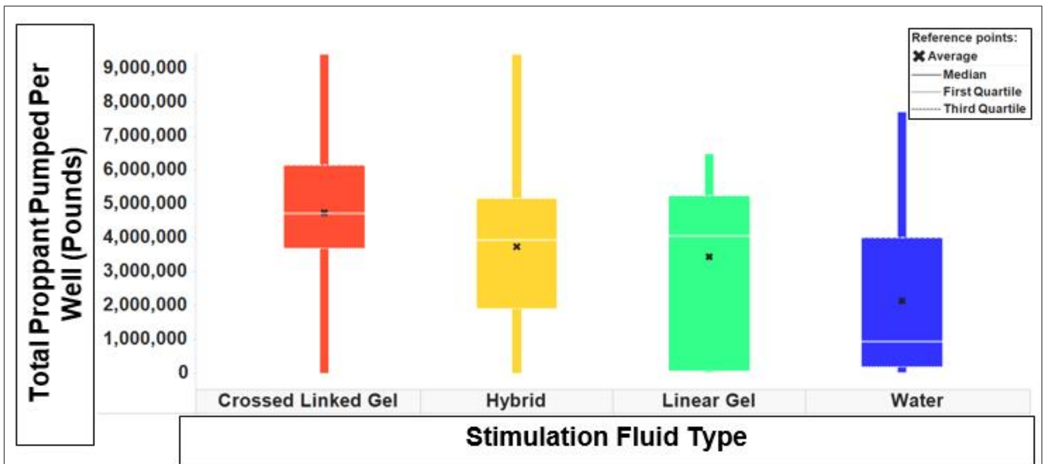
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	90	67	40	34
Avg	258,606	245,977	170,499	89,826
Median	255,403	228,589	153,695	37,316
Outliers	0	0	0	0
Q1	181,223	129,859	89,827	17,305
Q3	331,838	332,344	250,008	146,210
P10	129,380	26,602	15,044	0
P90	410,714	477,814	323,444	290,929

Figure 20. Figure 14 Box Plot and Statistical Table of Cotton Valley First Two Years Cumulative BOE Production



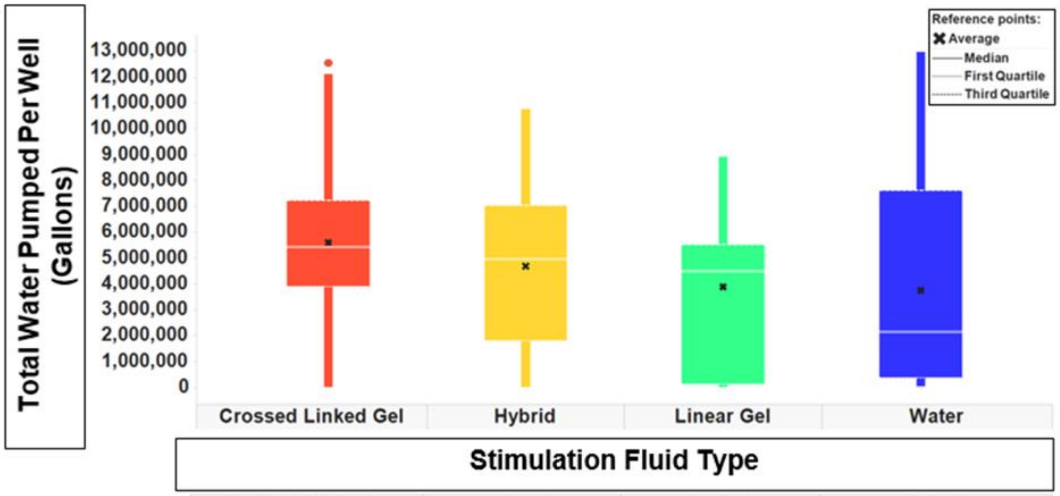
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	36	29	22	19
Avg	370,357	280,782	201,252	39,875
Median	393,027	258,022	147,136	26,879
Outliers	0	0	0	0
Q1	269,753	56,459	32,330	9,264
Q3	493,141	448,806	307,954	63,345
P10	110,277	14,918	14,033	0
P90	548,050	590,457	502,741	95,366

Figure 21. Box Plot and Statistical Table of Cotton Valley Five Years Cumulative BOE Production



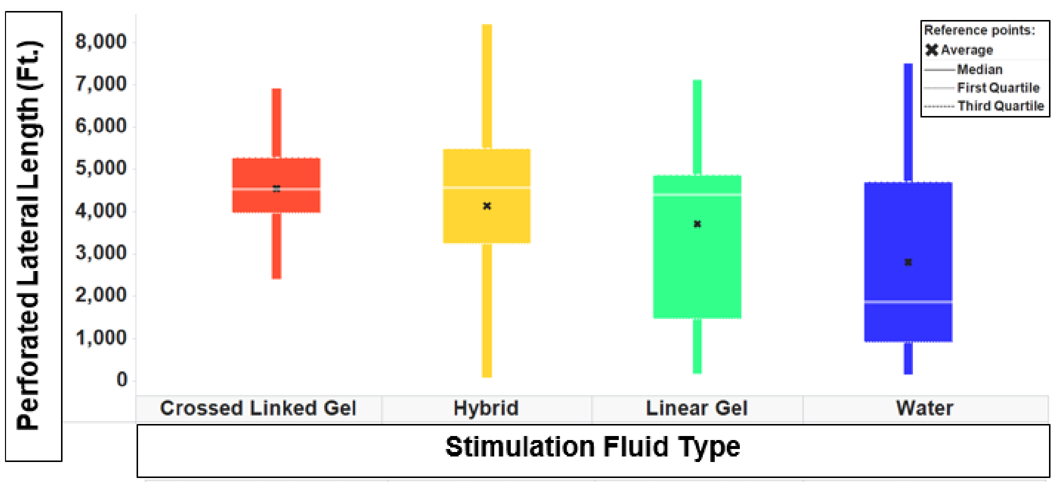
	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	108	73	41	40
Avg	4,731,075	3,729,819	3,442,212	2,137,621
Median	4,708,056	3,926,081	4,053,821	938,853
Outliers	0	0	0	0
Q1	3,673,634	1,907,832	68,523	179,925
Q3	6,128,738	5,162,005	5,238,348	4,000,109
P10	1,948,760	150,117	30,286	49,895
P90	7,512,081	6,892,832	5,929,707	5,061,931

Figure 22. Box Plot and Statistical Table of Cotton Valley Total Proppant Pumped Per Well



	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	108	73	41	40
Avg	5,600,021	4,685,777	3,882,487	3,751,053
Median	5,428,604	4,949,922	4,502,956	2,144,629
Outliers	1	0	0	0
Q1	3,894,912	1,786,652	114,052	364,112
Q3	7,229,373	7,029,204	5,507,031	7,603,890
P10	2,467,766	170,583	36,960	150,016
P90	8,758,825	8,203,463	6,489,423	8,992,913

Figure 23. Box Plot and Statistical Table of Cotton Valley Total Water Per Well



	Crossed Linked Gel	Hybrid	Linear Gel	Water
Count	107	73	41	40
Avg	4,549	4,134	3,707	2,803
Median	4,523	4,556	4,402	1,859
Outliers	0	0	0	0
Q1	3,960	3,232	1,447	911
Q3	5,284	5,482	4,863	4,706
P10	3,379	925	868	215
P90	5,696	6,600	5,634	5,848

Figure 24. Box Plot and Statistical Table of Cotton Valley Lateral Length

3.4. COMPARING TREATMENT FLUID TYPES AND BOE IP, CUMULATIVE FIRST 6 MONTHS, FIRST YEAR, 2 YEARS, AND 5 YEARS BOE PRODUCTION

Although the Cotton Valley is a tight gas formation and most of the production is gas, there is also some reported oil production. In this study, all the production comparisons were based on the equivalent barrel of oil method were 6 MCF of gas equal 1 barrel of equivalent oil. Figure 25 shows the five years cumulative production of gas, oil, and the equivalent barrel of oil for the four fracturing fluid types. Wells' performance was found to vary between initial producing and longer-term recovery for the four hydraulic fracturing fluid types studied. Figure 26 and Figure 27 show comparisons between the four types of treatment fluids for the average initial production (IP) per well and the average IP per foot per well, respectively. Practical initial production is used in this study and it is measured by dividing the second-month cumulative production by the number of produced days (BOE/Day). Calculating the practical IP based on the second month of production is more representative as in the first month production the number of produced days may vary based on when the well was brought to production after completion during that month. Comparing the IP of the different treatment fluid types showed that hybrid fluids reflected slightly higher average practical IP than cross-linked gel treatment. Linear and water fracs performed lower than hybrid and cross-linked fluids. To eliminate the bias related to the difference between the lateral length of the stimulated wells, the practical IP data were divided by the perforated lateral length to be compared based on the normalized (BOE/Day)/foot. Hybrid treatment showed the highest productivity per stimulated foot followed by cross-linked gel, then linear gel and finally water fracs. Figure 28 shows the average cumulative first six months production in Cotton Valley wells and Figure 29 shows the average first six months cumulative BOE production per perforated foot. The average first year, two years and five years cumulative BOE production of the Cotton Valley wells are shown in Figure 30, Figure 31, and Figure 32, respectively. The average cumulative BOE per perforated length (BOE/Ft) for the Cotton Valley first year, two years and five years are presented in Figure 33, Figure 34, and Figure 35, respectively.

In summary, hybrid and cross-linked gels performed the best in terms of initial production rate. The cross-linked gels and hybrid frac treatments performed the highest and showed slightly similar behavior in the average cumulative production up to 2 years. The five years cumulative BOE showed that the cross-linked gel fluids produced more 30% higher than hybrid fracs and 84% better than gel fracs. Cross-linked fracs performed superior with 9 folds higher average cumulative 5 years BOE compared to water fracs. The average cumulative production per stimulated foot analysis also confirmed the same conclusion with water fracs performing the lowest while cross-linked, hybrid, and gel fracs exhibited slightly similar production behavior as indicated in the figures below.

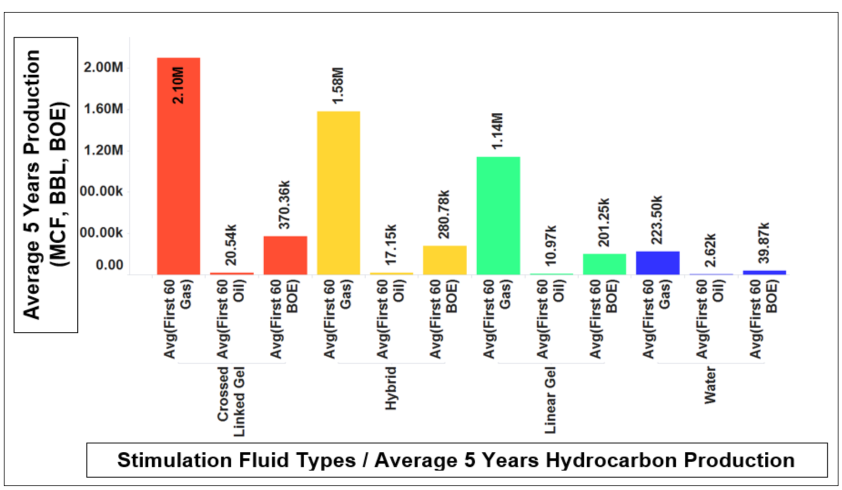


Figure 25. Bar Chart of Average Five Years Cumulative Production of Gas, Oil and Equivalent Oil for each Hydraulic Fracturing Fluid Type

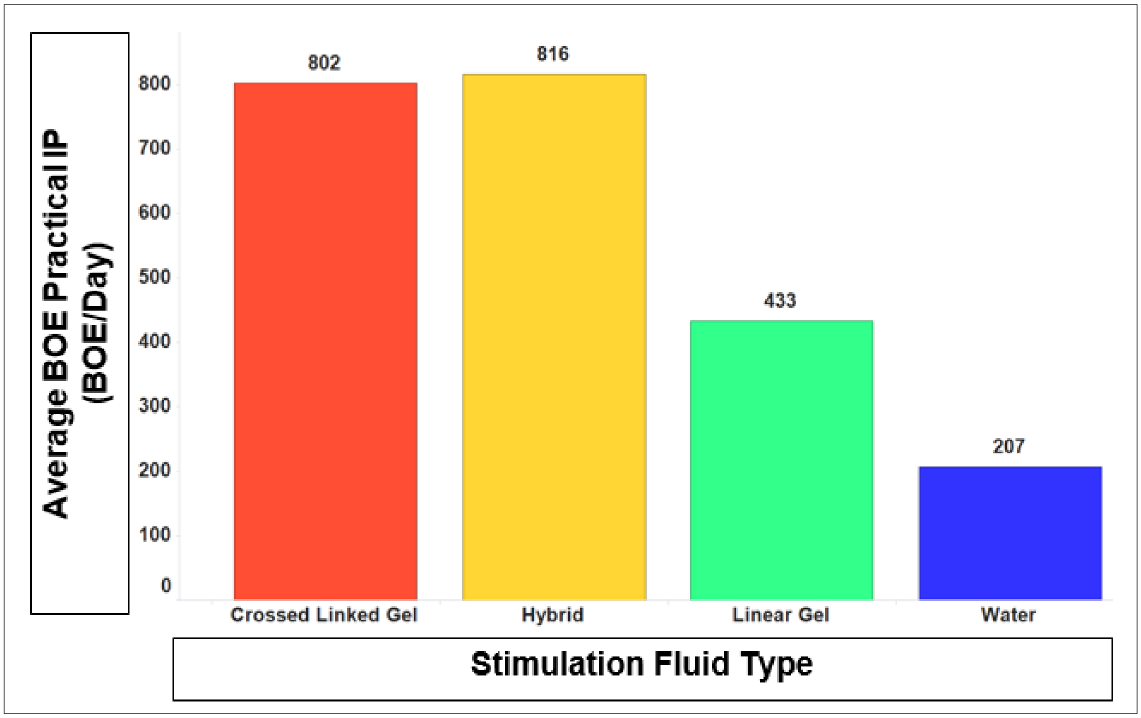


Figure 26. Bar Chart Comparing the Cotton Valley Average Practical IP for Different Stimulation Fluid Types

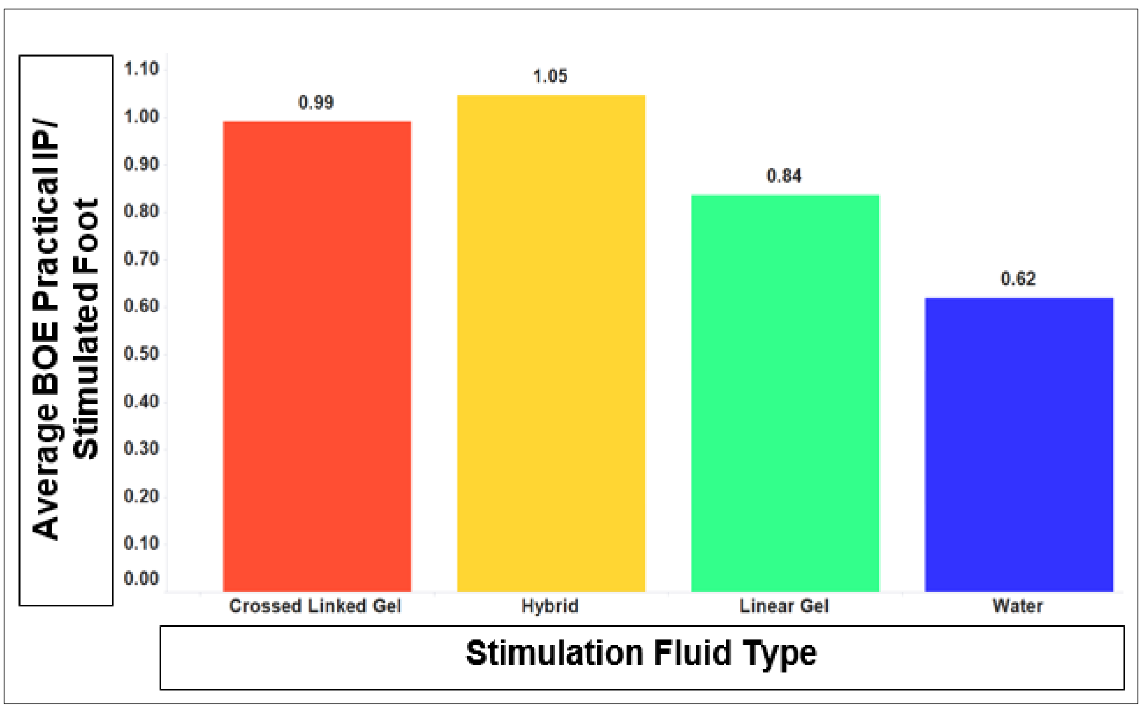


Figure 27. Bar Chart Comparing the Cotton Valley Average BOE Practical IP per Stimulated Foot for Different Stimulation Fluid Types

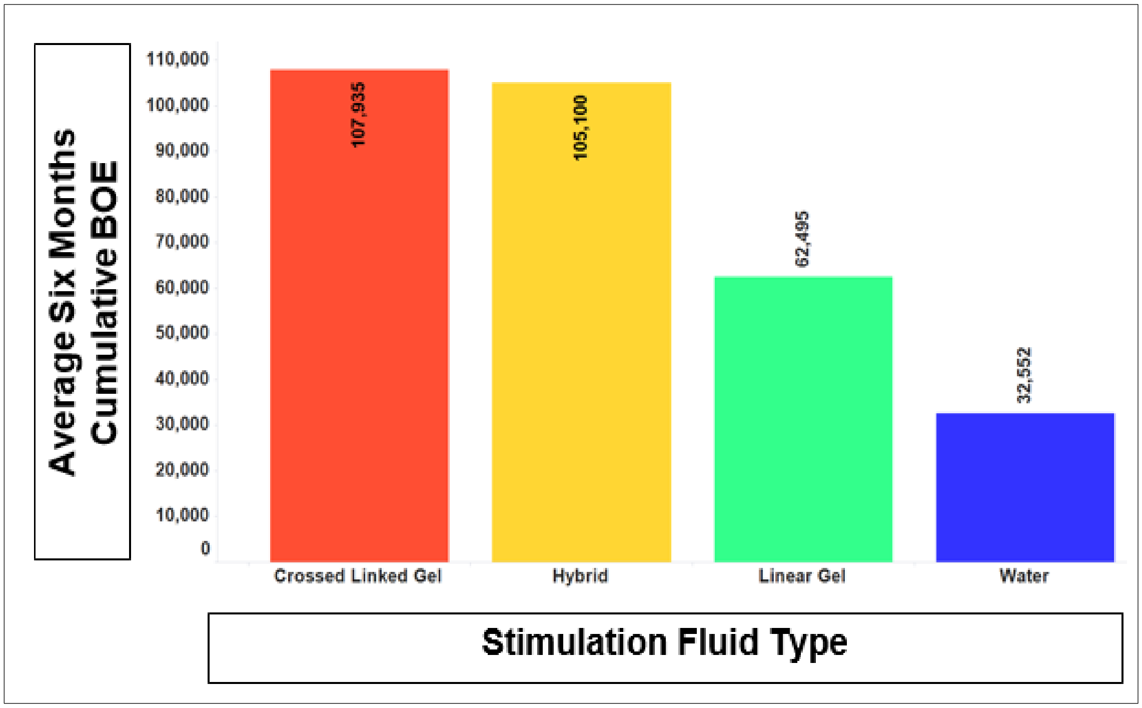


Figure 28. Bar Chart Comparing the Cotton Valley Average Six Months Cumulative BOE for Different Stimulation Fluid Types

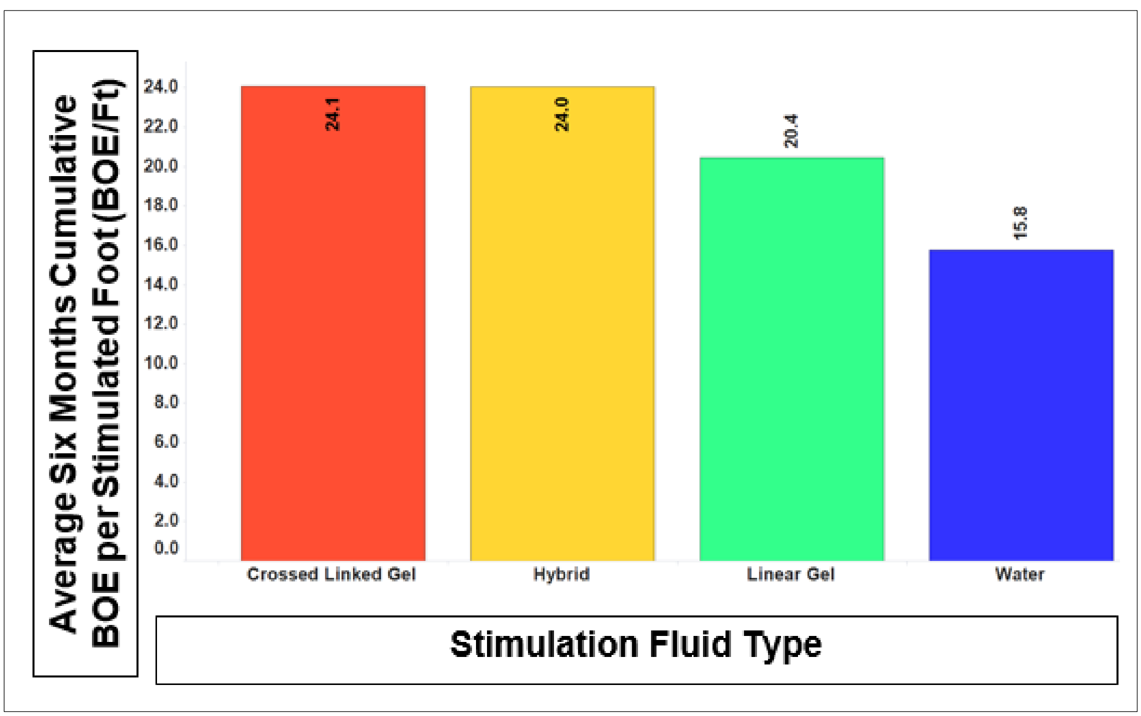


Figure 29. Bar Chart Comparing the Cotton Valley Average Six Months Cumulative BOE per Stimulated Foot for Different Stimulation Fluid Types

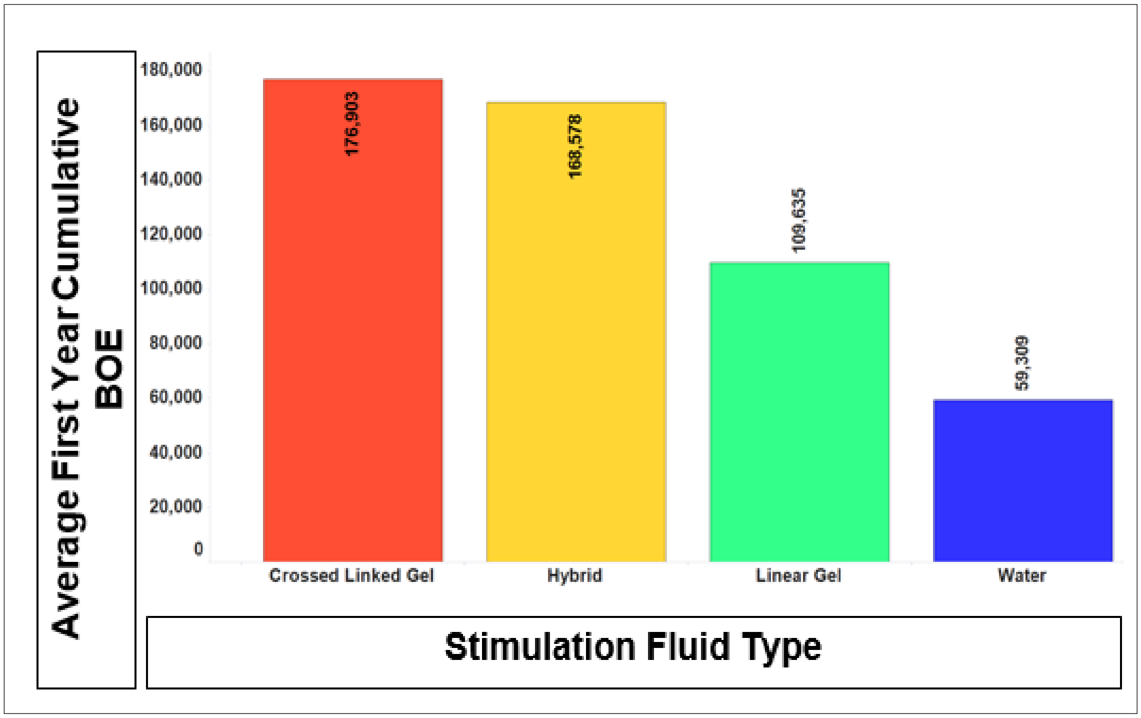


Figure 30. Bar Chart Comparing the Cotton Valley Average First Year Cumulative BOE for Different Stimulation Fluid Types

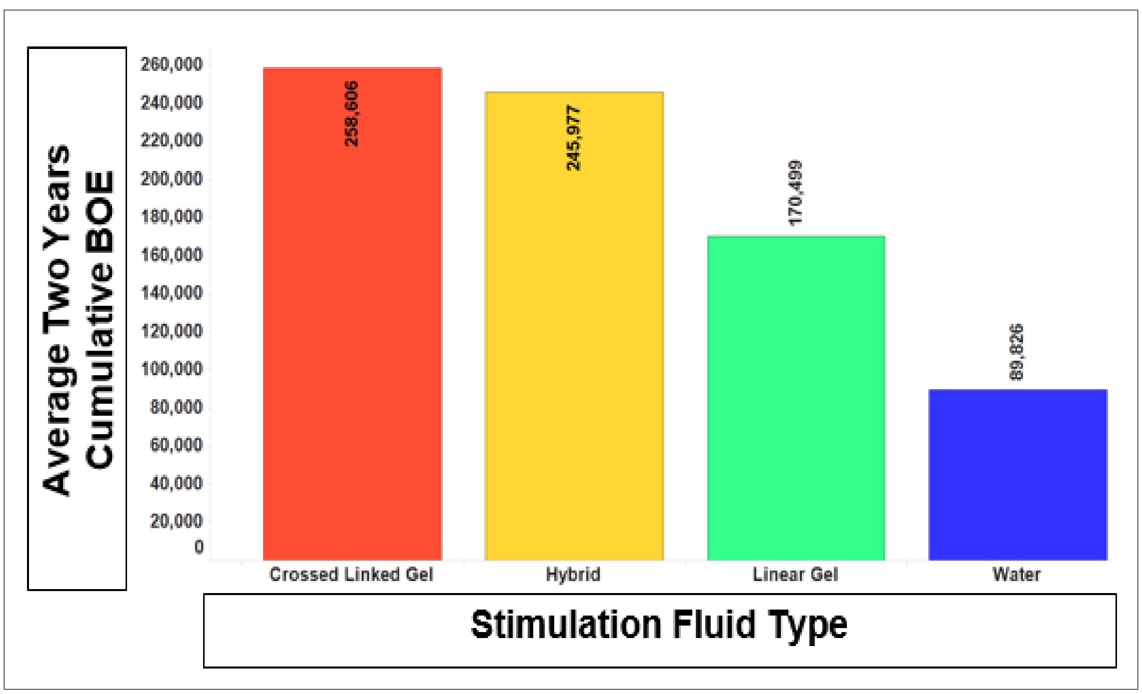


Figure 31. Bar Chart Comparing the Cotton Valley Average Two Years Cumulative BOE for Different Stimulation Fluid Types

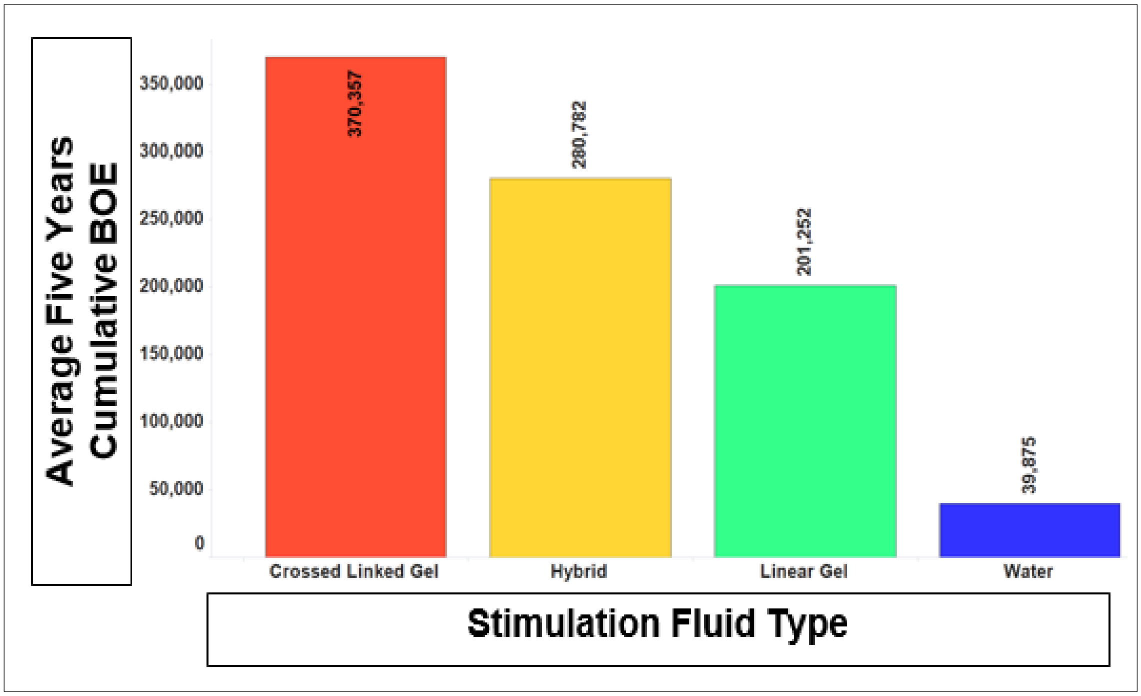


Figure 32. Bar Chart Comparing the Cotton Valley Average Five Years Cumulative BOE for Different Stimulation Fluid Type

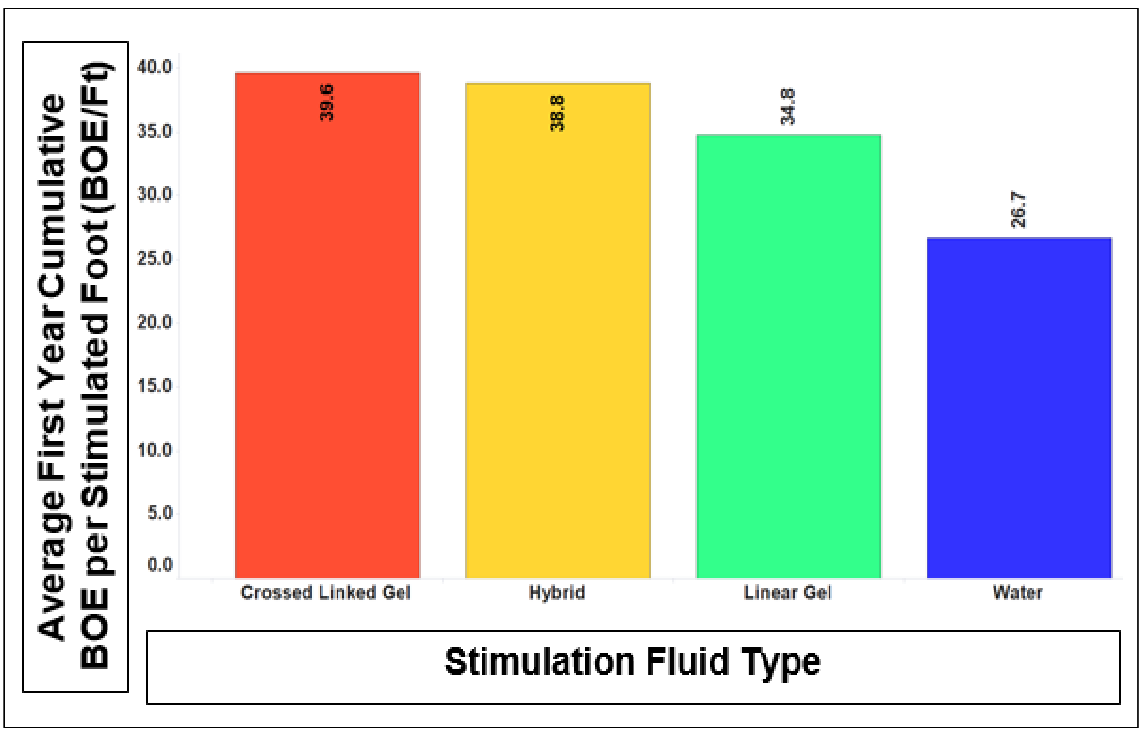


Figure 33. Bar Chart Comparing the Cotton Valley Average First Year Cumulative BOE per Stimulated Foot (BOE/ft) for Different Stimulation Fluid Types

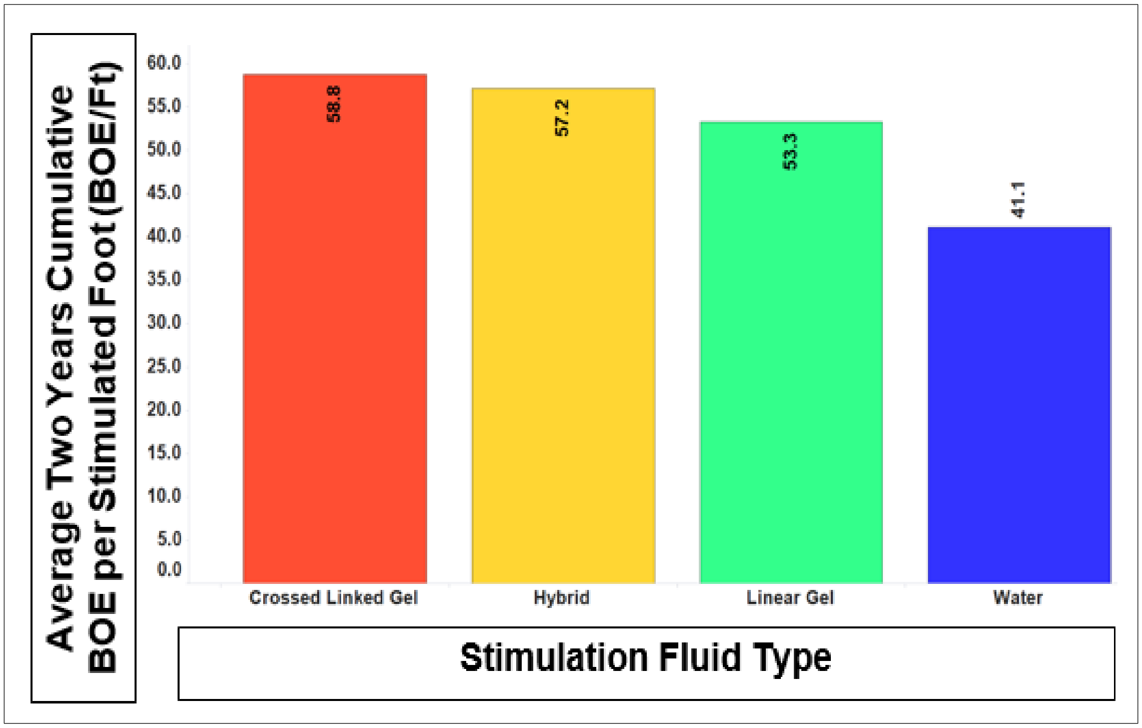


Figure 34. Bar Chart Comparing the Cotton Valley Average Two Years Cumulative BOE per Stimulated Foot for Different Stimulation Fluid Types

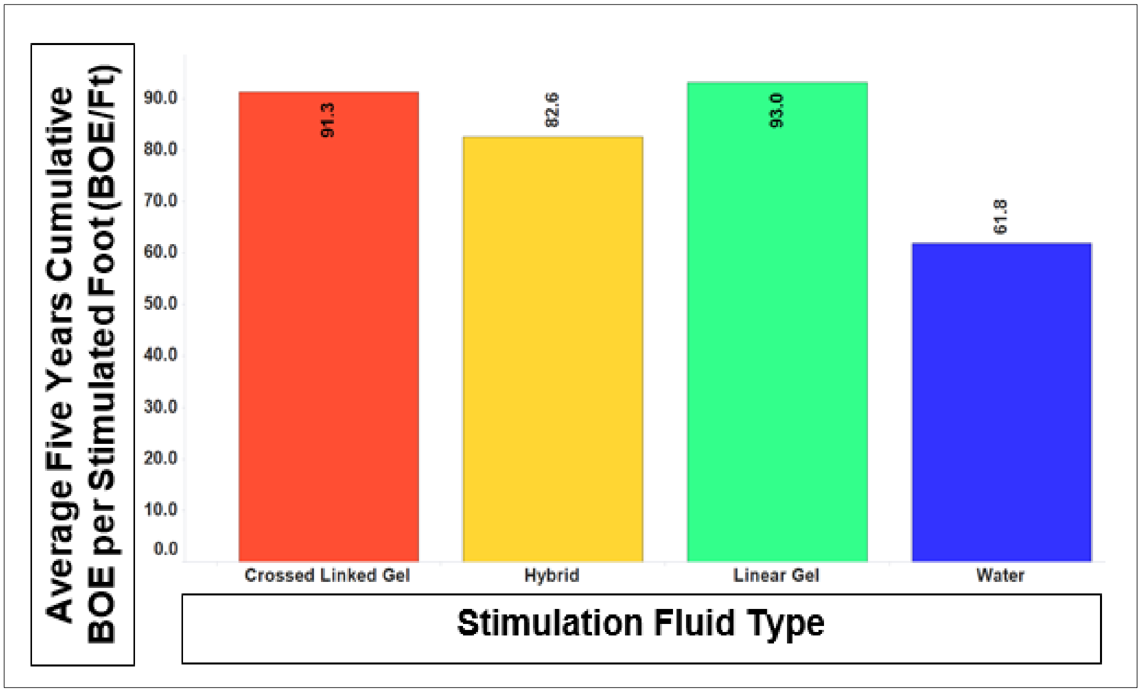


Figure 35. Bar Chart Comparing the Cotton Valley Average Five Years Cumulative BOE per Stimulated Foot for Different Stimulation Fluid Types

3.5. COMPARING THE EFFECT OF STIMULATION PARAMETERS ON AVERAGE BOE PRODUCTION

Figure 36, Figure 37, Figure 38, Figure 39, Figure 40, and Figure 41 present the average proppant mass per perforated lateral length and average water volume per perforated lateral length versus the cumulative BOE production per perforated lateral length for the first year, 2 years and 5 years, respectively. Within each of these figures, the blue bar charts represent the average proppant mass per foot (Lbs/Ft) and the green chart represents the average water volume per foot (Gal/Ft). The solid black line chart represents the average cumulative production per foot (BOE/ft).

This analysis indicates that cross-linked gels delivered the highest proppant per foot and consumed the highest water per foot. Cross-linked gels performed the highest BOE/ft for the first and second years and slightly lower than gel fracs five years BOE/ft. On the other hand, water fracs delivered the least amount of proppant per stimulated foot and consumed the second largest amount of water. Water fracs produced the least BOE/ft compared to the other four hydraulic fracturing fluid systems in the Cotton Valley that are addressed in this study.

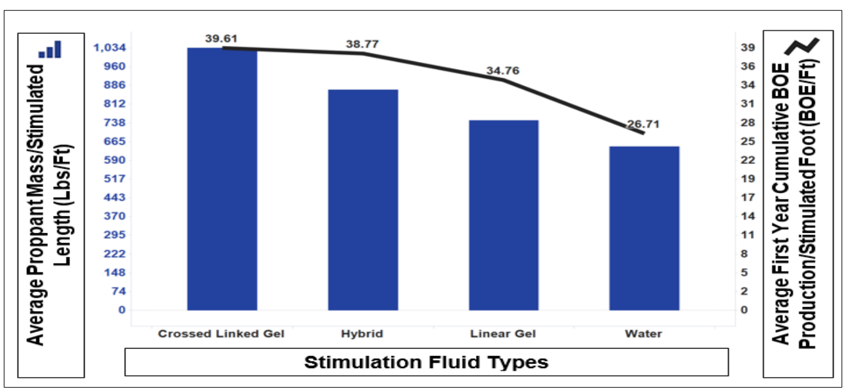


Figure 36. Combined Bar and Line Charts Comparing the Cotton Valley Average Pumped Proppant Mass per Foot and the First Year Cumulative BOE Production per Stimulated Length for each Stimulation Fluid Type

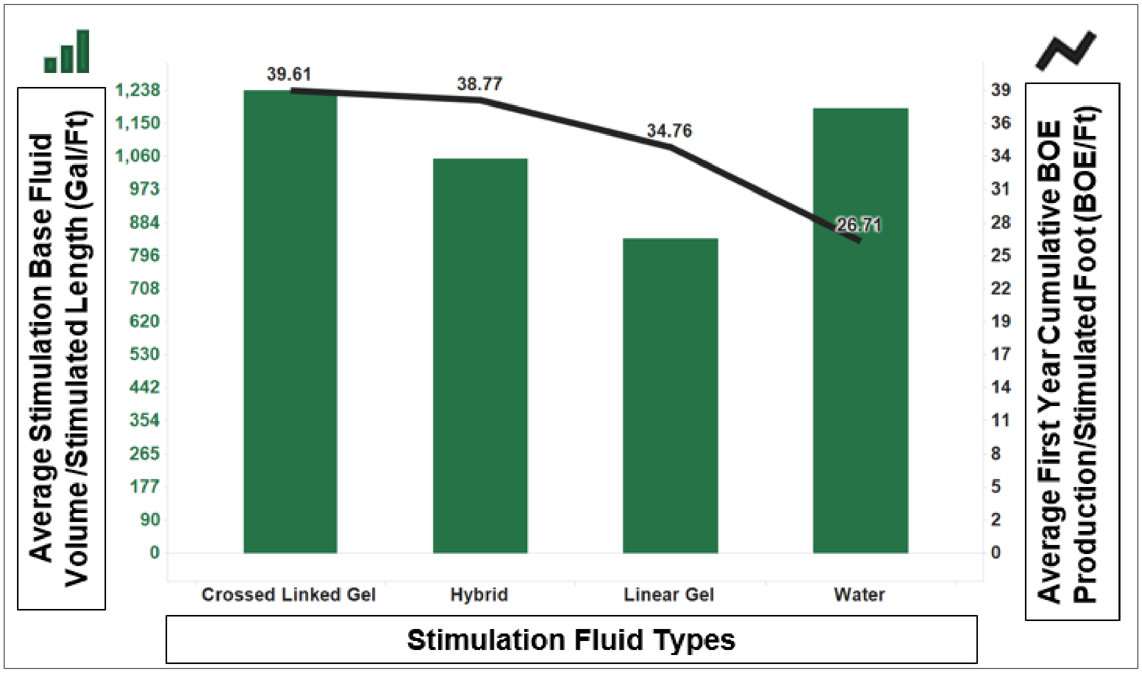


Figure 37. Combined Bar and Line Charts Comparing the Cotton Valley Average Pumped Stimulation Base Fluid per Foot and the First Year Cumulative BOE Production per Stimulated Length for each Stimulation Fluid Type

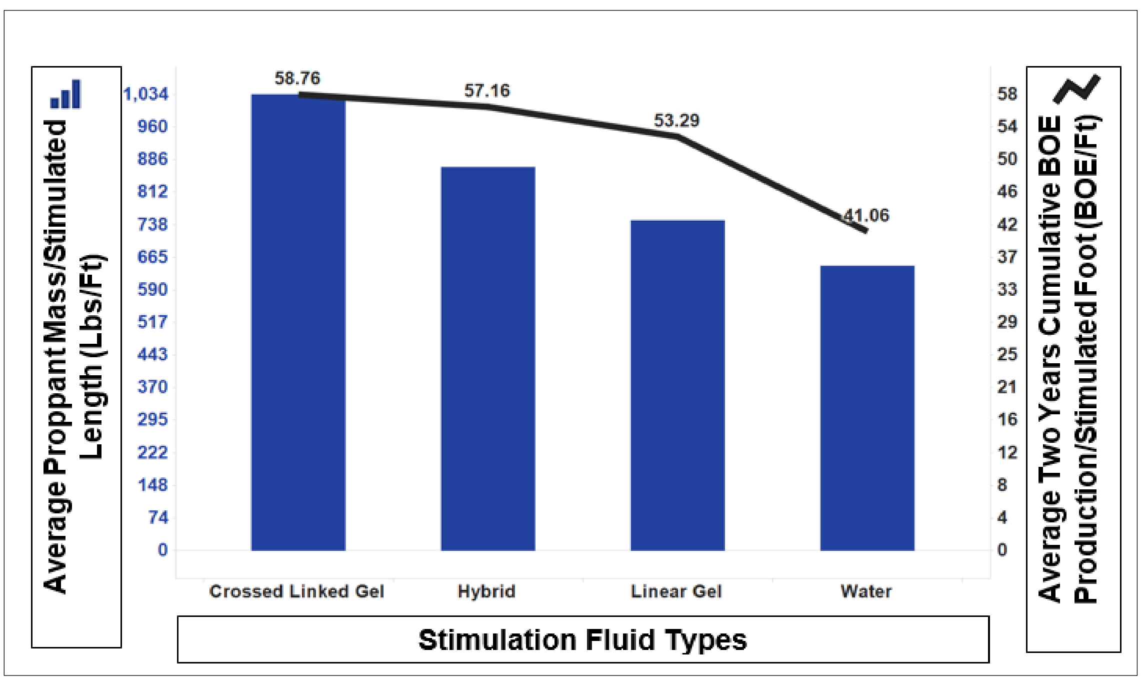


Figure 38. Combined Bar and Line Charts Comparing the Cotton Valley Average Pumped Proppant Mass per Foot and the Two Years Cumulative BOE Production per Stimulated Length for each Stimulation Fluid Type

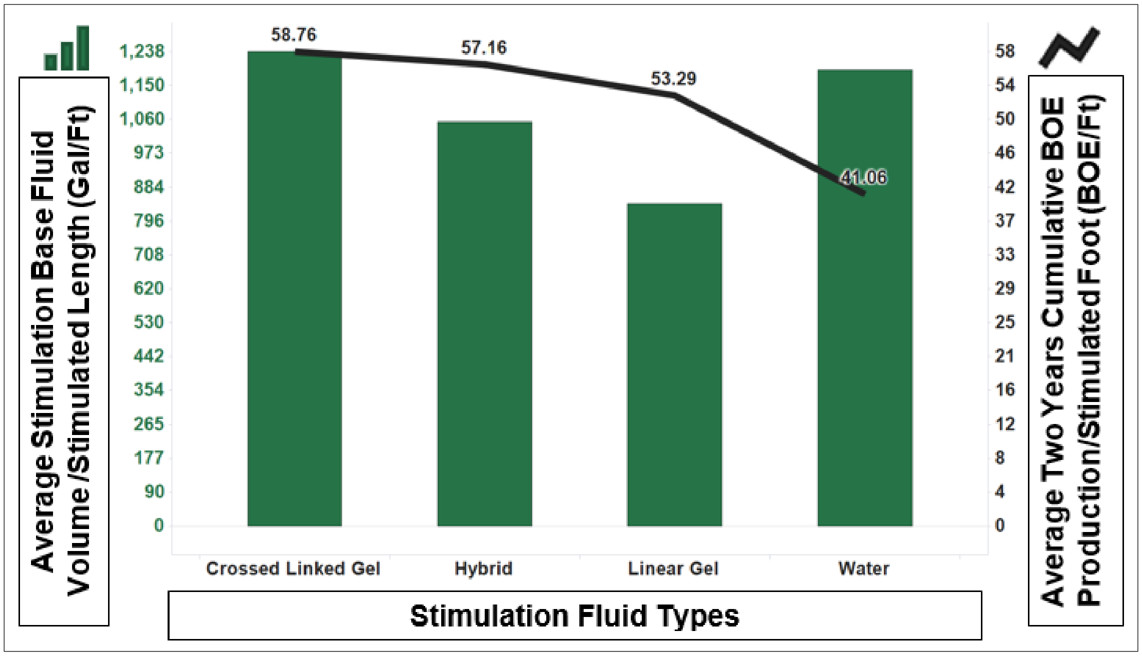


Figure 39. Combined Bar and Line Charts Comparing the Cotton Valley Average Pumped Stimulation Base Fluid per Foot and the Two Years Cumulative BOE Production per Stimulated Length for each Stimulation Fluid Type

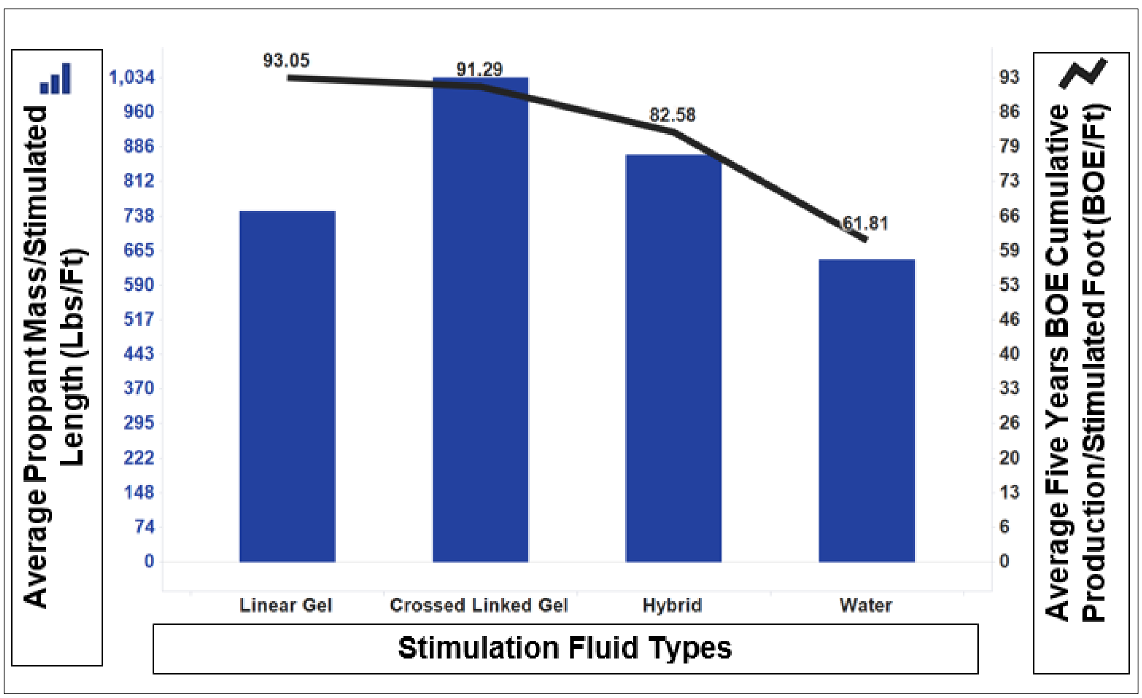


Figure 40. Combined Bar and Line Charts Comparing the Cotton Valley Average Pumped Proppant Mass per Foot and the Five Years Cumulative BOE Production per Stimulated Length for each Stimulation Fluid Type

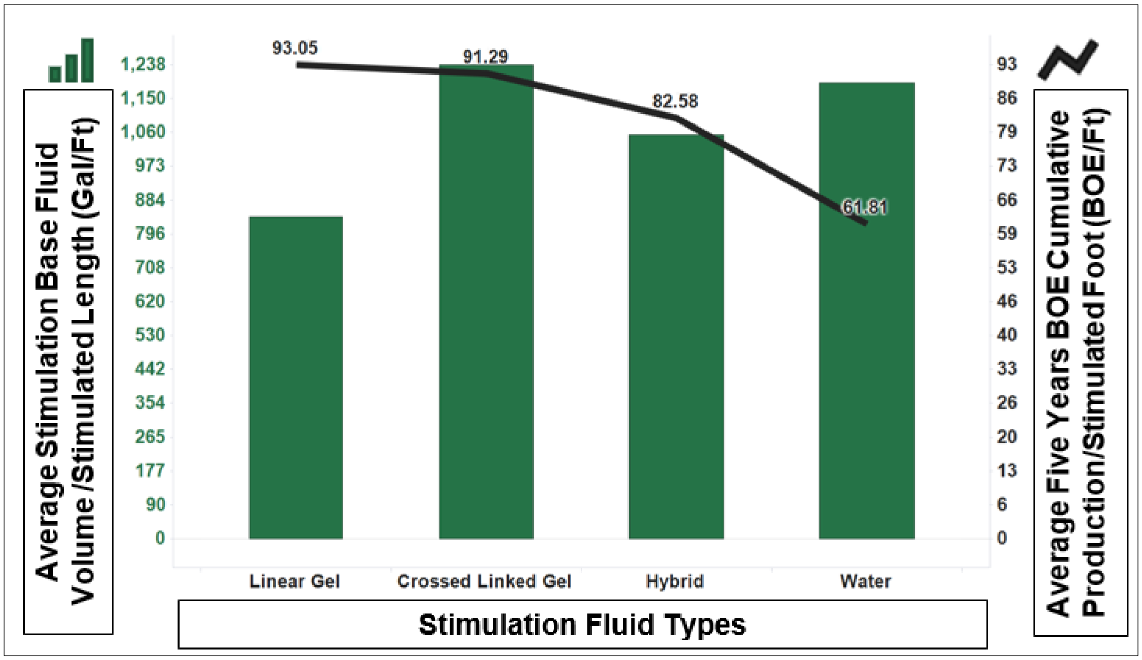


Figure 41. Combined Bar and Line Charts Comparing the Cotton Valley Average Pumped Stimulation Base Fluid per Foot and the Five Years Cumulative BOE Production per Stimulated Length for each Stimulation Fluid Type

4. CONCLUSIONS

In this study, chemical data from FracFocus were combined with well completion and production data from DrillingInfo, to create a robust database for investigating Cotton Valley well production as a function of hydraulic fracturing fluid type.

The main conclusions of this study were summarized in the following points:

- FracFocus chemical registry is a valuable source of hydraulic fracturing fluid, chemical and proppant data for wells stimulated after 2012. The data are available for download to Excel. The raw data must be reformatted for practical use, which may require some fundamental understanding of hydraulic fracturing treatments.

- This study provides detailed methodologies to utilize and process FracFocus public chemical registry database of stimulated wells in the United States. The workflows include raw data manipulation, grouping chemicals, calculating and verifying wells mass percent treatments, and merging final FracFocus data with DrillingInfo.
- A workflow was developed to classify the types of the hydraulic fracturing fluid based on the chemical ingredients of the treatment fluid.
- Cotton Valley completions vary as a function of hydraulic fracturing fluid and proppant used. The greatest average treatment fluid volume and proppant are associated with cross-linked gel fluid treatments.
- When normalized for lateral length, cross-linked gel fracs also exhibited the highest average proppant load per ft/well. It also consumed the highest average fluid volume/ft/well.
- The wells treated with water fracs were associated with shorter perforated lateral length compared to other types of treatment fluids while cross-linked gel fracs were associated with the longest perforated lateral length.
- Hybrid fluid fracs showed slightly higher average practical IP than cross-linked gel treatments.
- Cross-linked gel fracs outperformed all the other fluid types in terms of cumulative BOE for the first 6 months, first year, 2, and 5 years, respectively.
- The average 5 years cumulative BOE production for the wells fractured with cross-linked gel fluids was 9 folds higher than the average 5 years cumulative BOE for the wells fractured with water frac fluid.

- In cross-linked gel fracs, the first year cumulative BOE production represented 42% of the final 5 years cumulative BOE production while first year cumulative production of BOE for hybrid, gel, and water represented 46%, 42%, and 43%, of the five years cumulative BOE production, respectively.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to DrillingInfo team and the co-founder and director of university outreach Mr. Mark Nibbelink for their generosity in providing access to the U.S. production and completion data.

REFERENCES

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar, A. M. Review of Stimulation and Completion Activities in the United States Shale Plays. Paper ARMA 19-200, presented at the 53rd US Rock Mechanics / Geomechanics Symposium held in New York City, New York, USA, 23-26 June 2019a

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar, A. M. Long-Term Comparison of Gel and Water Fracture Stimulation in Marcellus Shale Play. Paper SPE-195990-MS accepted, and it will be presented at the SPE Annual Technical Conference and Exhibition held in Calgary, Canada, 30 Sep–2 Oct 2019b

American Chemical Society, CAS, Retrieved on February/ 05/ 2019, from, https://www.cas.org/support/documentation/chemical-substances/faqs

Alkinani, H. H., Al-Hameedi, A. T. T., Dunn-Norman, S., Flori, R. E., Hilgedick, S. A., Al-maliki, M. A., Amer, A. S. (2018). Journal of King Saud University – Science Examination of the relationship between rate of penetration and mud weight based on unconfined compressive strength of the rock. Journal of King Saud University - Science. https://doi.org/10.1016/j.jksus.2018.07.020

Arthur, J. D., Layne, M. A., Hochheiser, H. W., & Arthur, R. (2014, February 4). Spatial and Statistical Analysis of Hydraulic Fracturing Activities in US Shale Plays and the Effectiveness of the FracFocus Chemical Disclosure System. Society of Petroleum Engineers. doi:10.2118/168640-MS

Carman, P., Li, C., & Wheeler, R. (2018, September 24). Unorthodox Fluids for Unconventional Completions: Maximizing Production and Avoiding Pitfalls. Society of Petroleum Engineers. doi:10.2118/191751-MS

DrillingInfo website. Retrieved on March/01/2019, from https://info.drillinginfo.com/

Elsner, M., & Hoelzer, K. (2016). Quantitative Survey and Structural Classification of Hydraulic Fracturing Chemicals Reported in Unconventional Gas Production. Environmental Science & Technology, 50(7), 3290-3314. doi:10.1021/acs.est.5b02818

EPA website. Retrieved on February/ 05/ 2019, from https://www.epa.gov/saferchoice/safer-ingredients

FracFocus website. Retrieved on May/01/2019, from http://fracfocus.org/

Gas Formation Pay Recognition: American Association of Petroleum Geologists Bulletin, V. 67. p. 1002-1013

Khvostichenko, D., & Makarychev-Mikhailov, S. (2018, August 9). Oilfield Data Analytics: Linking Fracturing Chemistry and Well Productivity. Unconventional Resources Technology Conference. doi:10.15530/URTEC-2018-2903086

Liu, S., Spain, D. R., Devier, C., Buller, D., & Murphy, E. (2011, May 14). Integrated Petrophysical Study of a North America Tight Gas Sand: Cotton Valley Formation, East Texas. Society of Petrophysicists and Well-Log Analysts

Luo, H., & Zhang, Y. (2018, October 5). Mining FracFocus and Production Data for Efficacy of Fracturing Fluid Formulations. Society of Petroleum Engineers. doi:10.2118/191787-18ERM-MS

Myers, R. R., Knobloch, T. S., Jacot, R. H., & Ayers, K. L. (2017, October 4). Production Analysis of Marcellus Shale Well Populations Using Public Data: What Can We Learn from The Past? Society of Petroleum Engineers. doi:10.2118/187505-MS

Olaoye, O., & Zakhour, N. (2018, September 24). Utilizing a Data Analytics Workflow to Optimize Completion Practices While Leveraging Public Data: A Permian Basin Case Study. Society of Petroleum Engineers. doi:10.2118/191416-MS

Romero, D. H., & Poston, S. W. (2016, May 10). Data Mining and Analysis - Eagle Ford Formation, South Texas. Society of Petroleum Engineers. doi:10.2118/179994-MS

Srinivasan, K., Ajisafe, F., Alimahomed, F., Panjaitan, M., Makarychev-Mikhailov, S., & Mackay, B. (2018, August 28). Is There Anything Called Too Much Proppant? Society of Petroleum Engineers. doi:10.2118/191800-MS

Weijers, L., Wright, C., Mayerhofer, M., Pearson, M., Griffin, L., & Weddle, P. (2019, January 29). Trends in the North American Frac Industry: Invention through the Shale Revolution. Society of Petroleum Engineers. doi:10.2118/194345-MS

Wescott, W. A. 1984. Diagenesis of Cotton Valley Sandstone (Upper Jurassic), East Texas: Implications for Tight

Yang, Y., Robart, C. J., & Ruegamer, M. (2013, February 4). Analysis of U.S. Hydraulic Fracturing Design Trends. Society of Petroleum Engineers. doi:10.2118/163875-MS

V. DESCRIPTIVE DATA ANALYTICS TO INVESTIGATE STIMULATION AND COMPLETION TRENDS IN THE UNITED STATES: HOW PROPPANT AND WATER UTILIZATION HAVE CHANGES OVER TIME?

Mustafa A. Al-Alwani, Larry K Britt, Shari Dunn-Norman, Husam H. Alkinani, Abo Taleb T. Al-Hameedi, Atheer M. Al-Attar, Waleed Al-Bazzaz

Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA

ABSTRACT

The United States hydraulic fracturing and completion activities have seen a paradigm shift in the pumped water and proppant volumes along the increasing completed lateral over time. Hydraulic fracturing has become the most dominant treatment in North America to tap the unconventional resources. Large scale analysis of the completion size in terms of the amount of proppant, water, and lateral length trends over the period of 2011 and 2018 is presented in this paper. The objective of this study is to elucidate the completion trends over time and to summarize the average values of the completion and stimulation parameters. It will put the readers' mind in perspective of how much proppant and water have been utilized over the years and shine a light on the progression of the industry growing demand on water and proppant.

Data from FracFocus website, which serves as an official chemical disclosure registry for many of the oil and gas states, were utilized in this study. For each stimulated well, it reports water volume and mass percentage of proppant and all other chemical ingredients. The raw data were scraped, parsed, and cleaned using advanced statistical and validation approaches to extract useful information out of the noisy data. A database of

more than 80,000 wells was built and integrated into this study. Proppant total mass and the concentration of other chemical ingredients were calculated. The obtained new variables were coupled with other completion parameters such as the horizontal length, perforated lateral length, and well types.

The data of each parameter were subjected to rigorous quality control and inspected for outliers. After passing all the data quality and validation procedures, an advanced data visualization technique was selected to reflect the trends over time. The overall trends of the whole aggregated United States are presented in this paper. Investigating the created visualizations showed that there is an undeniable increasing trend in the amount of the proppant and water being consumed over the investigated period.

This paper exploits a large number of wells and their associated parameters. Such a large dataset will provide a comprehensive and practical representation of the completion and stimulation trends. It will also establish a baseline and reference values for the investigated parameters by benchmarking the mean, median, and other statistical expressions for each time period.

1. INTRODUCTION

Over the last decade, the unconventional resources development has grown tremendously as a result of the advancement in horizontal drilling technology coupled with hydraulic fracturing stimulation. Large data have been accumulated over the years as more wells are drilled and stimulated. The gathered data of reservoir, drilling, completion, and stimulation parameters can be utilized to properly optimize the productivity of the wells and can lead to more economic benefits from the unconventional plays. Data utilization have been implemented by large industry sectors. The oil and gas field data analytics have been the most dynamically growing area for the industry (Khvostichenko and Makarychev-Mikhailov, 2018). Realizing the importance of data-driven decisions enticed the operators and service contractors to dedicate many of their resources to the booming field of data science. Although this field of utilizing data analytics is relatively new, many studies have been conducted and the outputs were very insightful (Arthur et al., 2014; Olaoye and Zakhour, 2018; Srinivasan et al., 2018; Myers et al., 2017; Luo and Zhang, 2018; Yang et al., 2013; Romero and Poston, 2016; Carman and Wheeler, 2018).

Implementing data science in the oil and gas industry has many challenges and pitfalls to be overcome before reaping any benefits. The first obstacle is data availability and accessibility issues. For many years in the past, the completion and stimulation data accessibility was limited to the operators and service companies and data sharing culture was not prevailing. The success of any data-driven analysis is based on the data wide range availability. Any missing part of the data will hinder such analyses. If the data are not shared across the board or utilized publicly, the operators and service companies will end up only possessing a piece of the puzzle. On the other hand, collaboration on creating a joint database will help everyone to see the big picture. The second challenge is the data quality and ingestion form, i.e., how good and reliable the reported data and how the data can be accessed. For example, is the data available on paper, uneditable electronic file, or saved on an easy to copy excel file format. The third challenge is the availability of data handling and processing software. The good old way of spreadsheet calculations might reach a limit where Excel software cannot accept more rows of data as it reaches the maximum allowable number of cells (Excel maximum rows = 1,048,576). Handling large datasets, such as FracFocus, demands a higher data processing software and advanced workflows.

In this study, data from the FracFocus (2019) website were processed and combined with other parameters from a commercial source (DrillingInfo) to produce a database for North America to serve as a source for the stimulation, completion, and production data. The developed database covers all the data from 2011 to 2018 and will be utilized to track trends and activities on any levels of details.

As the unconventional resources in the United States became the industry focus over the years and as the operators and the service companies started to try many drilling and completion techniques, the industry objective was to maximize the production or to find the best applications to deliver this purpose. Gathering completion and stimulation data over an extended time period will help in determining the general trends that the industry is pursuing in their journey of experimentation and refinement for the aspiration of optimizing production. The advancement in drilling technology led to an increase in the horizontal lateral length and consequently the number of the stimulated stages. This induced an increase in the amount of the proppant and fluid volumes being pumped downhole.

In the early stages of the unconventional resources development, operators started their stimulation designs with simple engineered procedures to establish their treatment baseline. Those designs were usually built to accommodate one or more of the following conditions such as the availability of equipment and logistics, the cost of the materials, and the targeted initial production. Some operators simply copied the designs that the offset operators are implementing. The industry is going through a steep learning curve trying to improve early baseline designs and still making every effort to come up with the best combination that fits their productivity objectives. The developments did not only consider the well completion and stimulation operational designs, but also extended to the types of products such as proppant types, additives, and fluid types.

Many publications in the literature tried to document the history and the trends of the shale best practices. The early studies did not contain representative formation descriptive data that can be utilized to generalize and extrapolate trends to field level. Early studies were usually conducted on data from a limited number of localized wells and reflected a very area specific trend (Carman and Wheeler, 2018). King (2010) enriched the published literature with a published study that provided a comprehensive review of trends in the shale plays and best practices. At that time, his manuscript was considered as one of the best summary studies and it is still regarded as a very valuable reference. As more data are shared, and more comprehensive databases are built, presenting case history studies can be very helpful in tracking the industry progression towards the best and optimum completion/stimulation strategies. Having high data quality of the various number of parameters which describes the formation, drilling, and completion, will serve the purpose of finding meaningful trends and correlations by using data visualization and descriptive analysis techniques.

This study is investigating the general trends of the completion and stimulation in the United States over the time frame of 2011 and 2018. The trends of the proppant and water consumption over time were presented along with the distribution histograms and statistical values of each parameter. The normalized values of stimulation parameters by the perforated lateral length were also discussed. The effects of the amount of proppant and water on initial production were introduced in the results and discussions section to show the response of well productivity to the variation of different completion parameters.

1.1. DATA OF THE STUDIED AREA

Having the ability to access data from many wells as presented in this analysis enhances the quality of the investigation trends. It helps in providing many insights into the overall use of proppant and water volumes across the United States. It also gives intuition about the time progression of trends within each basin, shale play, operators, fields, counties, etc.

Figure 1 shows a map of the United States with all the wells integrated as part of this study. The wells' marker was color-coded based on the type of produced fluid (Oil, Gas, O&G, CBM: Coal Bed Methane). The shape of the markers was also identified based on the well drilling type (D: Directional, H: Horizontal, and V: Vertical).

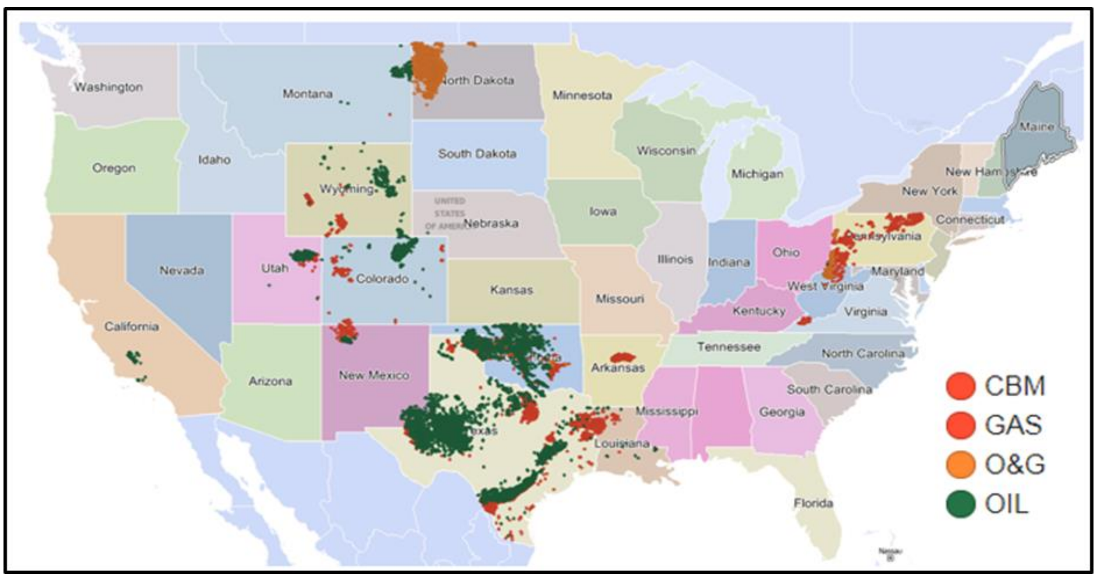


Figure 1. The Wells Location on the U.S. Map

Figure 2 shows the data coverage in this study in terms of geological basins. The pie chart sectors are sized based on the number of wells in each basin and all the sectors were color-coded by basins. The basins with less than 2% wells count were grouped into one sector named "Others", which represents 5.6% of the data presented in this paper.

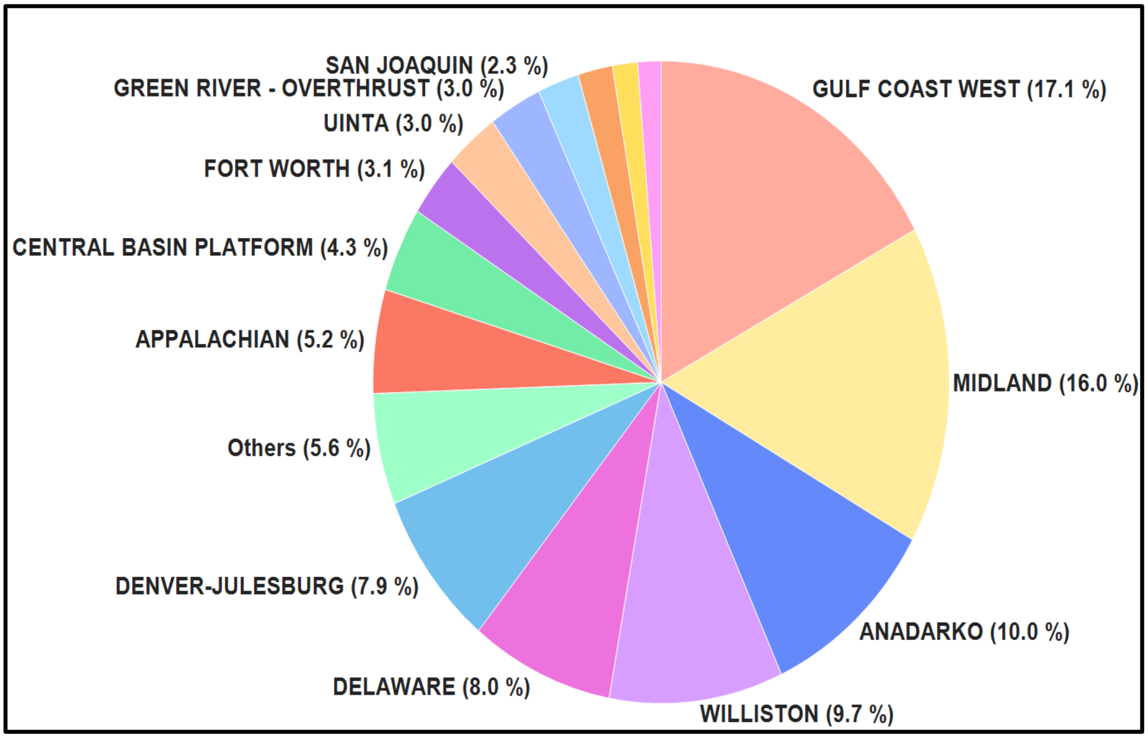


Figure 2. Data Coverage of Major Basins in the United States

The basins data can be subdivided into different shale plays and further analysis of completion and production trends can be presented. Figure 3 shows a pie chart of the percentage distribution of the major shale plays participating in this study. Eagle Ford and Midland represents the largest shale plays in the United States and they are represented very well in this study with more than 19,000 wells to cover 33% of the wells in this dataset.

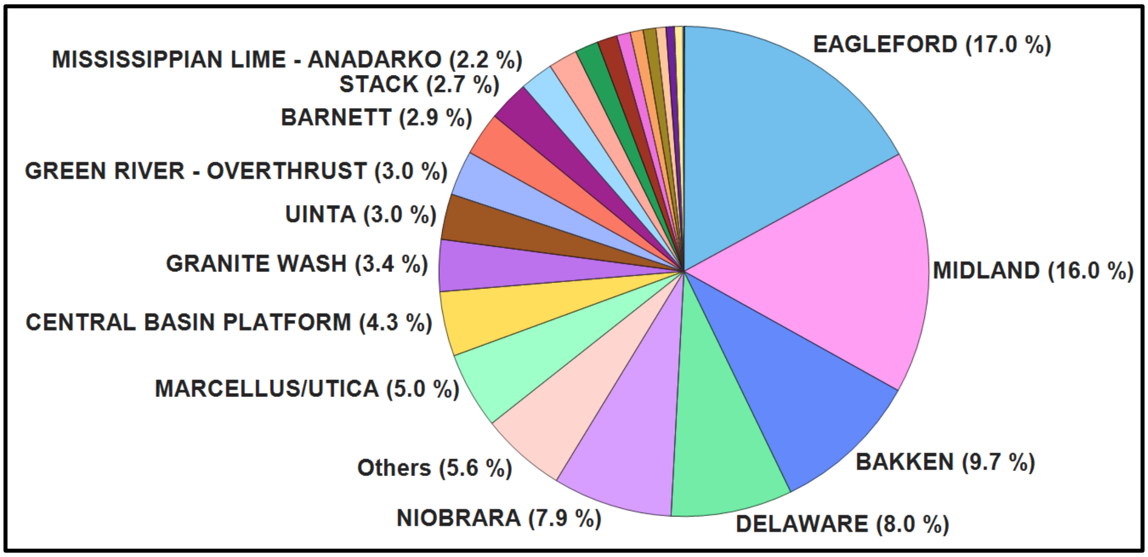


Figure 3. Data Coverage of Major Shale Plays in the United States

2. METHODOLOGY

The first step in this research was to build a comprehensive database that contains the chemical data, the volume of pumped fracturing fluids, proppant mass, completion parameters, and production records. The Fracfocus website and DrillingInfo for production and completion commercial database were selected to be the two major databases to be downloaded and merged.

The FracFocus website is a public source of the types and percentages of all fluids and chemicals pumped as part of hydraulic fracturing stimulation. The data of each well is reported separately on a pdf format. The website allows searching for the wells by state, county, etc. down to the unique API number. It is managed by the Ground Water Protection Council and Interstate Oil and Gas Compact Commission. Since 2011, more than 127,000 well sites have been registered into the website as of March 2019. Due to the limitation of the maximum number of rows that each Excel sheet can handle, the FracFocus database was downloaded in several separate excel spreadsheets. Each spreadsheet contains the well identifiers, chemicals percentage, and description data. Examples of the parameters are API number, longitude, latitude, job data, state, county, depth, water volume used, and mass percentage of most of the chemical additives including water and proppant. The data quality obtained from the raw fields of FracFocus is challenging and requires a considerable amount of time to be spent on processing and cleansing procedures to eliminate the typos and errors and to group the chemicals based on the intended purpose of use. Data processing algorithms were developed in this study. Many data workflows were also integrated to eliminate errors and repeated fields to come up with an accurate calculated mass value for each chemical ingredient and to group the additives based on their intended purpose of use. The workflows were also developed to detect re-fractured wells and many built-in validation processes were integrated to confirm the accuracy of the outcomes. The concentration in pound/gallon of all the chemical additives was also determined in the newly created database. More than 80,000 wells with validated data quality were extracted from FracFocus. Figure 4 shows a data processing workflow example.

Production and completion data were extracted from a commercial database source (DrillingInfo, 2019). It compiles and authenticates data retrieved from the States' public records and third-party sources. Since the data quality is already verified by DrillingInfo (D.I.), routine data quality check and outliers' detection procedures were applied to D.I. parameters. Practical initial production was calculated as the second-month production divided by the number of produced days (BBL/Day, MCF/Day). The perforated lateral length was used as a proxy to stage count and as a normalizing parameter. It was calculated by subtracting the upper perforation depth from the lower perforation depth. Proppant density (Proppant Loading) was calculated by dividing the total amount of the pumped proppant mass (lbs) by the total volume of water (gal). Proppant mass, water volume, and practical IP of gas and oil were also normalized to the length of the perforated horizontal lateral length (pound/ft, gallon/ft, (MCF/Day)/ft, and (BBL/Day)/ft). Normalizing by the perforated lateral length will eliminate the difference of the lateral length between the wells.

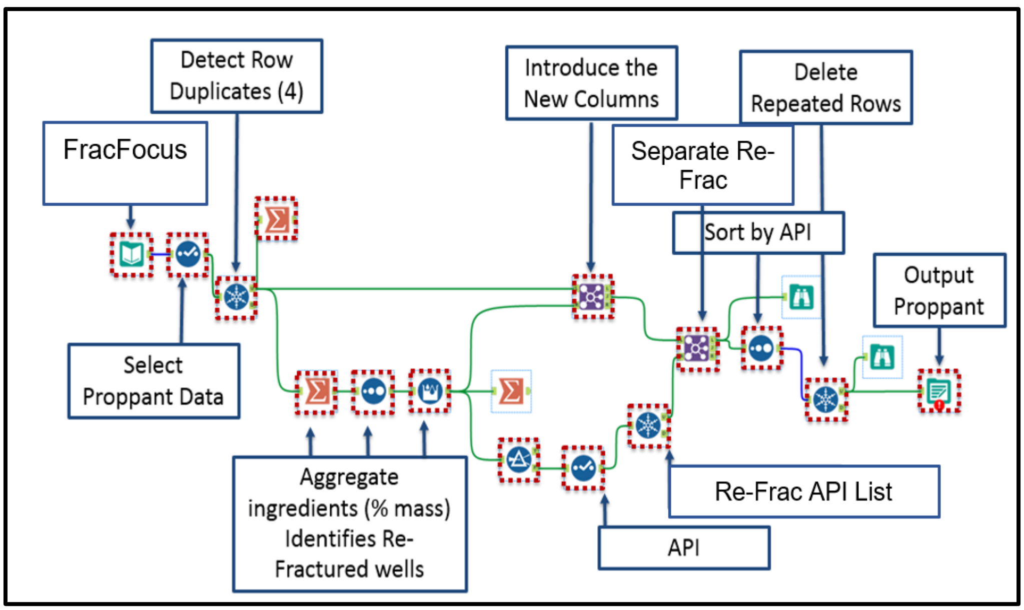


Figure 4. Example of Workflow to Process Data Extracted from FracFocus

3. RESULTS AND DISCUSSION

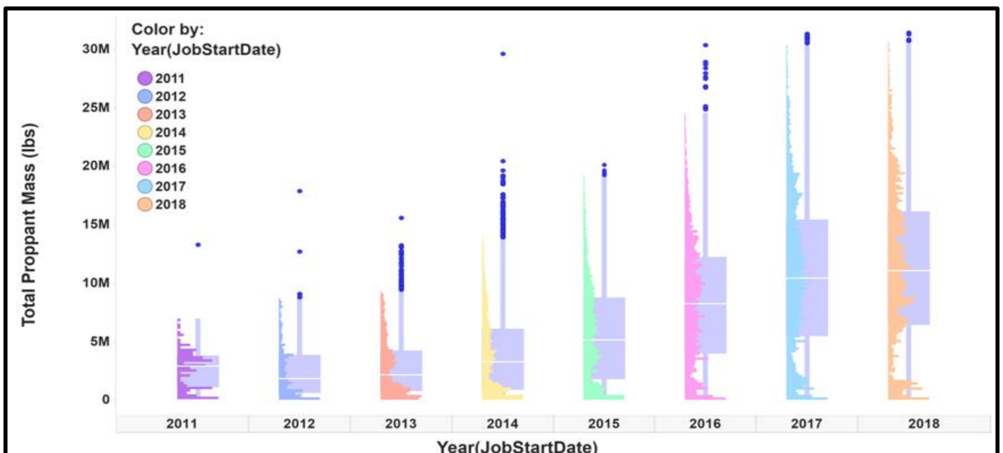
Completion trends across the United States have been showing a continuous upswing in the amount of proppant and fluid used in well stimulation. Data from all the major U.S. unconventional basins were utilized to visualize the growing trends over time. Although the conventional wisdom in the unconventional resources calls for considering each formation separately due to the geological differences; the completion trends in all the formations show the same correlation. This section will be broken into subsections to crystallize the main aspects and parameters discussed in this study.

3.1. PROPPANT DATA DISTRIBUTION AND TRENDS

Proppant (sand) plays a fundamental role in hydraulic fracturing stimulation. It is mixed with water and other different chemicals to hydraulically fracture low and ultra-low permeability shale formations. Proppant have been utilized in hydraulic fracturing for almost 70 years. The relatively recent advancement of long lateral horizontal drilling and multi-stage fracturing in addition to the growing demand for energy sources were the main factors of the growing demand on proppant. The escalation of oil and gas prices provides extra impetus to drilling, completion, and stimulation activities and hence proppant demand.

Figure 5 shows a combined box plot and distribution histogram of the proppant data for the whole United States hydraulic fracturing stimulation applications. The plots are also associated with a statistical table that shows the count of the participating wells; average value of the proppant mass pumped for each particular year; median value of the proppant which are used to describe the variables with wide range; summation which represents the total mass of proppant consumed for the particular year; maximum and minimum values to show the range of the data, standard deviation, the first and third quartile values of the data distribution; interquartile range; and the number of suspected outliers. Each box plot was color-coded based on the stimulation year. Comparing the median and the average values of the proppant mass over the years shows an undeniable trend that the operators have been utilizing more proppant as time progress.

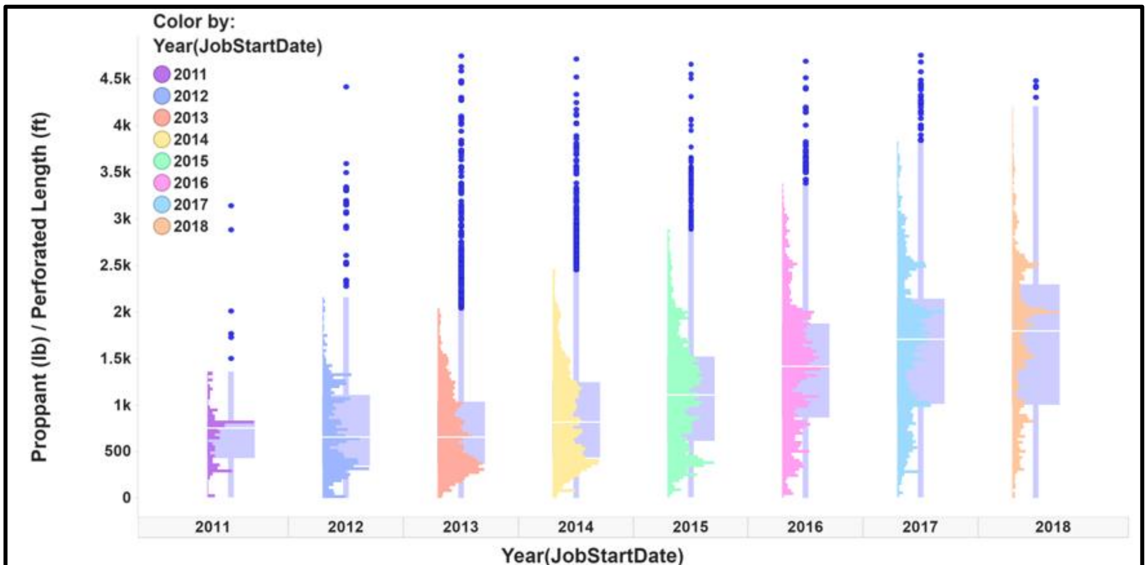
The years 2011 and 2012 had the least number of wells because the fact that FracFocus was released during the second quarter of 2011 and reporting the stimulation data to the website was voluntary, after 2013 more operators started to participate in data reporting and many states started to require all operators to report the stimulation data as part of the regulations. The number of wells in 2018 may not be representative to all of the wells stimulated during that year because the data was collected during the last quarter of 2018 and it takes few months to update FracFocus after the stimulation job is conducted. The average proppant mass per well started around 2 million pounds in 2011 and as the time progressed and the lateral length became longer, more proppant was pumped surging over 10 million pounds during 2017 and 2018.



	2011	2012	2013	2014	2015	2016	2017	2018
Count	119	824	11951	17076	10521	6178	8849	3986
Avg	2.64924e+06	2.53556e+06	2.8137e+06	3.92145e+06	5.7177e+06	8.38732e+06	1.0715e+07	1.12693e+07
Median	2.87361e+06	1.80315e+06	2.15673e+06	3.26852e+06	5.14971e+06	8.24703e+06	1.0421e+07	1.10612e+07
Sum	3.15259e+08	2.0893e+09	3.36265e+10	6.69626e+10	6.01559e+10	5.18169e+10	9.4817e+10	4.49193e+10
Max	1.32892e+07	1.78819e+07	1.56023e+07	2.96148e+07	2.01341e+07	3.03988e+07	3.1332e+07	3.14451e+07
Min	59175.5	23930.5	10327.3	10963.6	11974.3	13566.7	11334.3	10596.5
StdDev	1.93227e+06	2.36883e+06	2.43677e+06	3.4401e+06	4.3537e+06	5.49869e+06	6.81429e+06	6.90031e+06
Q1	1.06611e+06	588153	772135	888545	1.75078e+06	3.96728e+06	5.46775e+06	6.39634e+06
Q3	3.77372e+06	3.86917e+06	4.23373e+06	6.09239e+06	8.74472e+06	1.22105e+07	1.54281e+07	1.61188e+07
IQR	2.70762e+06	3.28102e+06	3.46159e+06	5.20384e+06	6.99394e+06	8.24323e+06	9.96037e+06	9.72247e+06
Outliers	1	6	53	67	6	13	11	5

Figure 5. Proppant Mass Box Plot and Data Distribution

To account for the increase in lateral length over time a box plot and distribution diagrams were created for the normalized value of proppant mass over the perforated lateral length (lbs/ft) as shown in Figure 5 It is still evident that the well average and median value of the net proppant pumped per each foot of the stimulated lateral was increasing with time. It can also be inferred that the data distribution range started to get wider during 2015 onward. Over the last 3 years, the operators started to push the proppant usage range further by trying to pump more proppant per foot as demonstrated in the data distribution histograms. The suspected outliers were kept in all of the box plots show the trials that some operators have conducted to test the operational limits. Only the extreme outliers were removed for the sake of this analysis.



	2011	2012	2013	2014	2015	2016	2017	2018
Count	113	744	10718	14682	8820	5228	7569	3054
Avg	748.742	801.359	754.591	907.854	1138.8	1424.09	1659.05	1710.31
Median	758.282	657.286	660.15	821.215	1109.93	1419.41	1703.36	1797.17
Sum	84607.8	596211	8.08771e+06	1.33291e+07	1.00442e+07	7.44514e+06	1.25574e+07	5.2233e+06
Max	3143.38	4419.83	4748.88	4718.66	4665.62	4697.89	4759.15	4487.18
Min	12.4244	6.04181	3.90575	6.13146	7.9183	3.39789	13.0529	4.45045
StdDev	465.127	628.709	510.284	570.668	637.912	726.271	804.283	821.189
Q1	429.423	353.203	366.332	437.255	615.582	869.054	1014.87	1007.69
Q3	834.325	1112.9	1035.56	1244.32	1523.55	1875.99	2145.01	2300.41
IQR	404.902	759.693	669.231	807.061	907.967	1006.94	1130.14	1292.72
Outliers	6	29	185	206	96	48	30	5

Figure 6. Proppant/Perforated Lateral Box Plot and Data Distribution

Figure 7 and Figure 8 clearly illustrate that between 2012 and 2018 there is an undeniable trend in the average and median values of total proppant and total proppant per perforated foot, respectively. Those two figures are combined charts of average and median values for each quarter of the studied period. They confirm the increase in average total proppant mass and proppant mass per foot especially during 2017 and 2018 which was more than 4 times the amount in 2013 for proppant average mass and more than double the net proppant per foot when compared to 2013.

Figure 9 shows the average value of proppant further sub-divided based on the types of drilled wells (H: Horizontal, D: Deviated, and V: Vertical). The figure also shows the corresponding number of stimulated wells for each type over time (the black bold line curve). The curve shows that the horizontal wells consume most of the proppant followed by the deviated then the vertical wells. It also shows how the horizontal wells' consumption of proppant increased with time. This incremental over time is attributed to the increased in the lateral length and the number of stimulated stages over time as well as to the operators' willingness to perform bigger jobs by pumping more proppant and fluids.

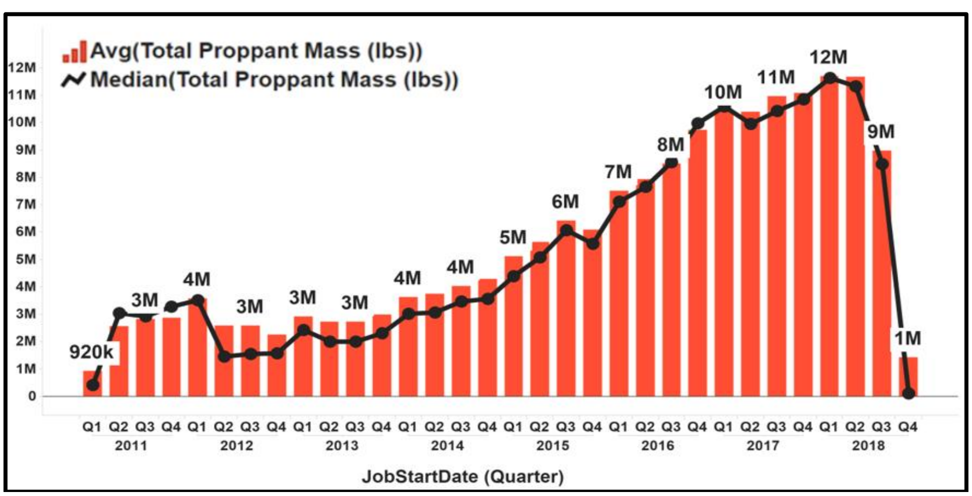


Figure 7. Average and Median of Total Proppant Over Years in the U.S.

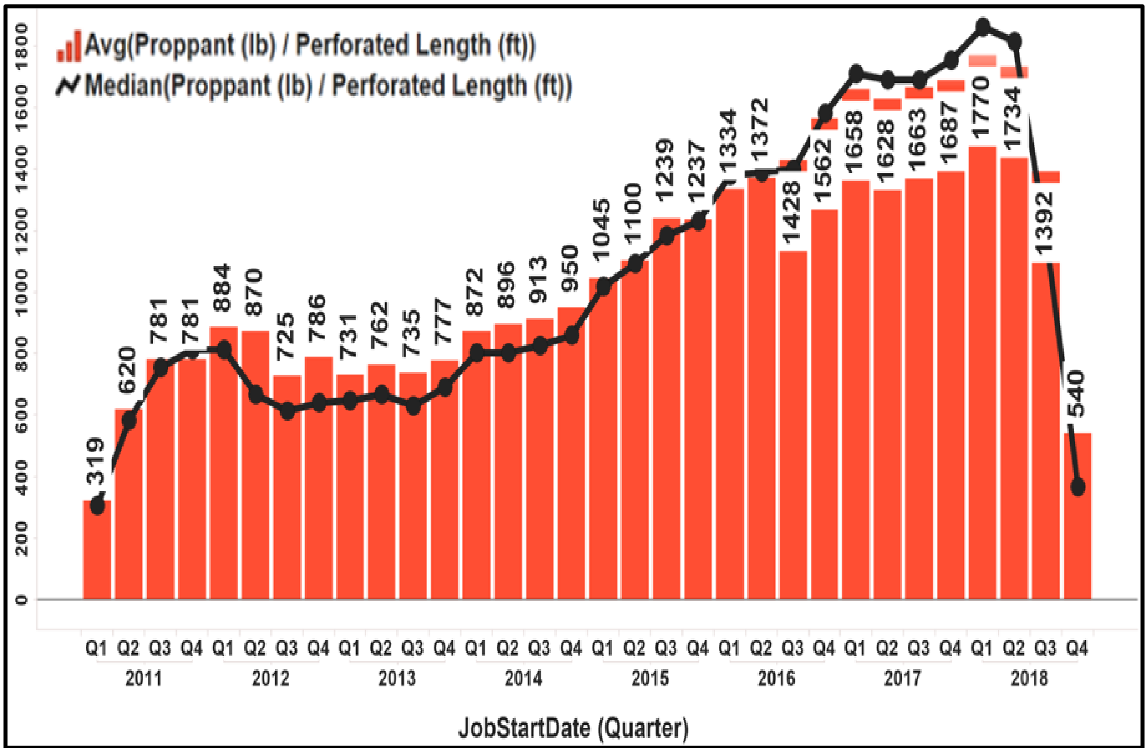


Figure 8. Average and Median Proppant Mass / Perforated Later Length over Time

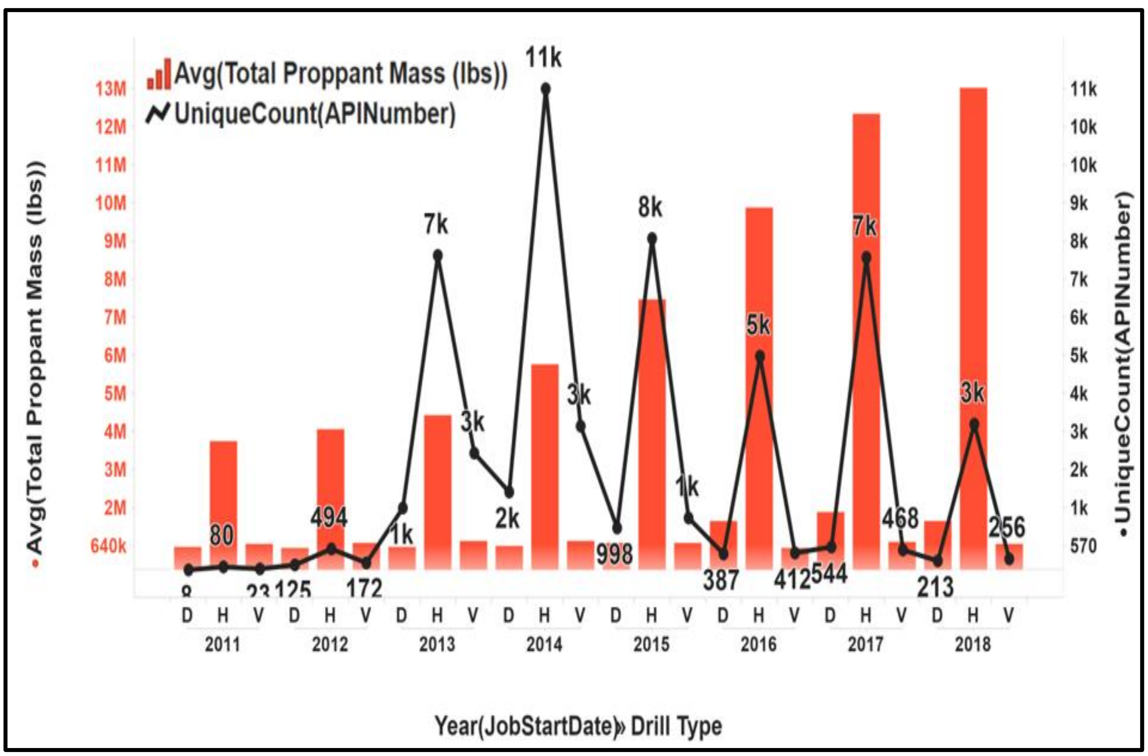


Figure 9. Average Total Proppant and Number of Wells over Time for Each Type

Figure 10 shows an example of the states' consumption of proppant. Texas has the lead with more than 50% of all the proppant used in the United States followed by Oklahoma and Colorado with more than 10% each.

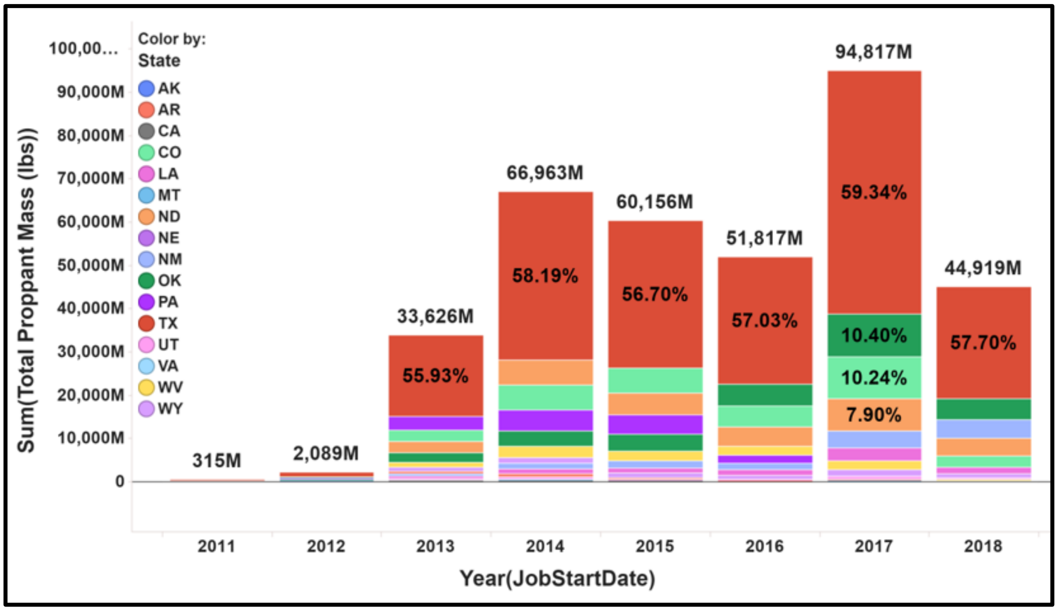


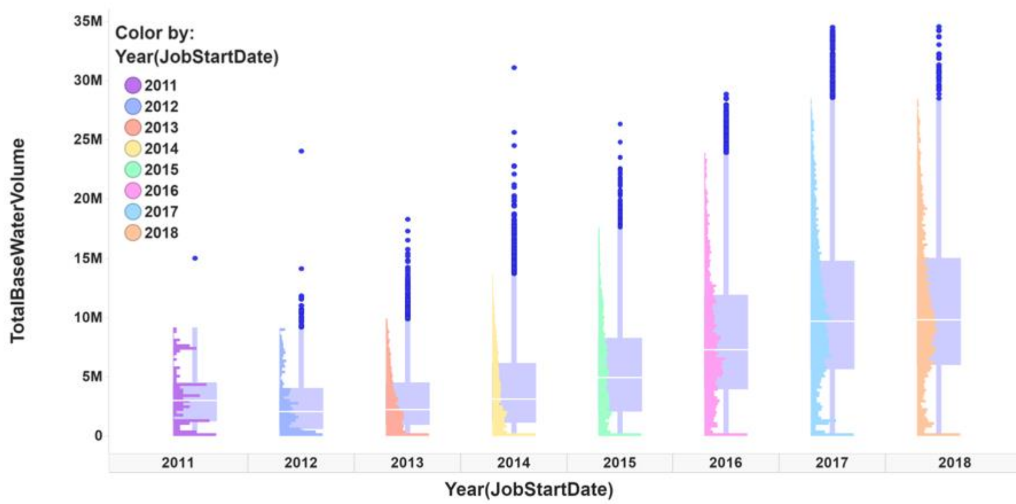
Figure 10. Proppant Consumption by State

3.2. WATER DATA DISTRIBUTION AND TRENDS

Figure 11 shows the volume of water (Million Gallons) usage box plot and distribution histograms. The average and median values showed a progression in the water volume from 2 to over 10 million gallons. Over the past three years, the water volume distribution range got wider which reflects the wider operational range that the operators started to implement by experimenting with bigger pumped fracturing jobs.

Figure 12 shows the normalized value of water volume per the perforated lateral length. It was created to show the pure trends of the water volume over time by taking the effects of the increased lateral length with time out of the equation. Between 2012 and

2017, the volume of water gallons per stimulated foot increased from year to year. The last two years showed similar trends, this could be attributed to reaching the optimized designs that the operators have been developing over the past years as they continued in trying different combinations of stimulation designs.



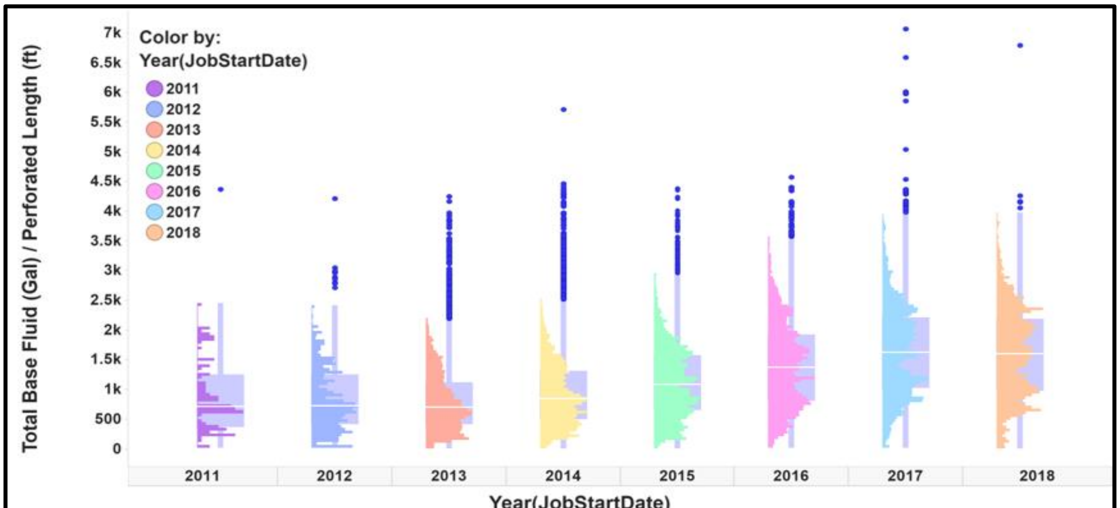
	2011	2012	2013	2014	2015	2016	2017	2018
Count	119	824	11951	17076	10521	6178	8849	3986
Avg	3.39259e+06	2.90966e+06	3.01013e+06	3.97926e+06	5.61678e+06	8.36119e+06	1.05704e+07	1.07183e+07
Median	3.01094e+06	2.06893e+06	2.25821e+06	3.13489e+06	4.9517e+06	7.26155e+06	9.68815e+06	9.8038e+06
Sum	4.03718e+08	2.39756e+09	3.5974e+10	6.79498e+10	5.90941e+10	5.16554e+10	9.3537e+10	4.27232e+10
Max	15015462	24067313	18321745	31119522	26349792	28912884	34544328	34587126
Min	27090	11257	5605	3798	8721	8867	9273	9581
StdDev	2.72435e+06	2.84219e+06	2.62906e+06	3.45275e+06	4.384e+06	5.90773e+06	6.92047e+06	6.83028e+06
Q1	1.23026e+06	581801	924684	1.14817e+06	2.04791e+06	3.91022e+06	5.66636e+06	6.00294e+06
Q3	4.49662e+06	4.02732e+06	4.51888e+06	6.16851e+06	8.30008e+06	1.1919e+07	1.48162e+07	1.50085e+07
IQR	3.26636e+06	3.44552e+06	3.5942e+06	5.02034e+06	6.25217e+06	8.00875e+06	9.14987e+06	9.0056e+06
Outliers	1	29	175	192	112	80	102	35

Figure 11. Water Volume Box Plot and Data Distribution

Figure 13 and Figure 14 clearly illustrate that between 2012 and 2018 there is an undeniable trend in the average and median values of total water volume and total water volume per perforated foot, respectively. Those two figures are combined charts of average and median values for each quarter of the studied period. They confirm the increase in average total water volume and water volume per foot especially during 2017 and 2018 which was more than 3 times the amount in 2013 for water average volume and more than double the net water volume per stimulated foot when compared to 2013.

Figure 15 shows the average value of water volume further sub-divided based on the types of the drilled wells (H: Horizontal, D: Deviated, and V: Vertical). The figure also shows that the corresponding number of wells for each type over time (the black bold line curve). The curve shows that the horizontal wells consume most of the water followed by the deviated then the vertical wells. It also shows how the horizontal wells' consumption of water increased with time.

Figure 16 shows the states consumptions of water. Texas has the lead with more than 55% of all the water consumed in the United States followed by Oklahoma and Colorado with more than 11% each.



	2011	2012	2013	2014	2015	2016	2017	2018
Count	113	744	10718	14682	8820	5228	7569	3054
Avg	915.936	847.403	814.968	943.113	1133.02	1416.95	1647.78	1613
Median	710.866	723.267	702.43	844.878	1087.29	1378.37	1620.37	1606.61
Sum	103501	630468	8.73483e+06	1.38468e+07	9.99325e+06	7.40783e+06	1.24721e+07	4.9261e+06
Max	4375.43	4214.3	4255.05	5711.89	4386.01	4577.73	7070.59	6789.83
Min	16.1437	17.4285	10.0839	8.11425	10.2244	13.3318	20.6685	6.6605
StdDev	663.287	579.235	532.565	581.725	626.585	742.311	786.553	777.729
Q1	363.668	417.808	412.454	499.495	648.536	811.299	1027.21	981.183
Q3	1259.13	1248.69	1126.87	1308.96	1574.96	1919.68	2209.89	2183.63
IQR	895.46	830.877	714.414	809.463	926.421	1108.38	1182.68	1202.44
Outliers	1	10	162	213	80	38	27	5

Figure 12. Water Volume/ Perforated Lateral Box Plot and Data Distribution

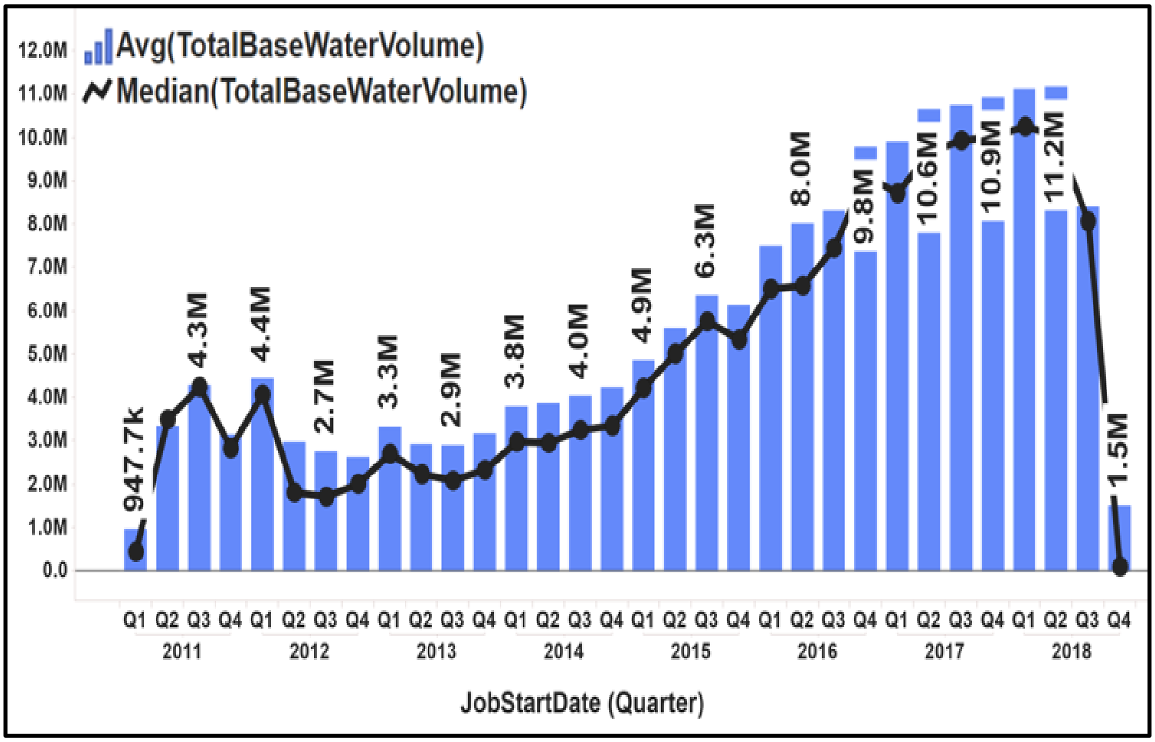


Figure 13. Average and Median Values of Total Water Over Years in the U.S.

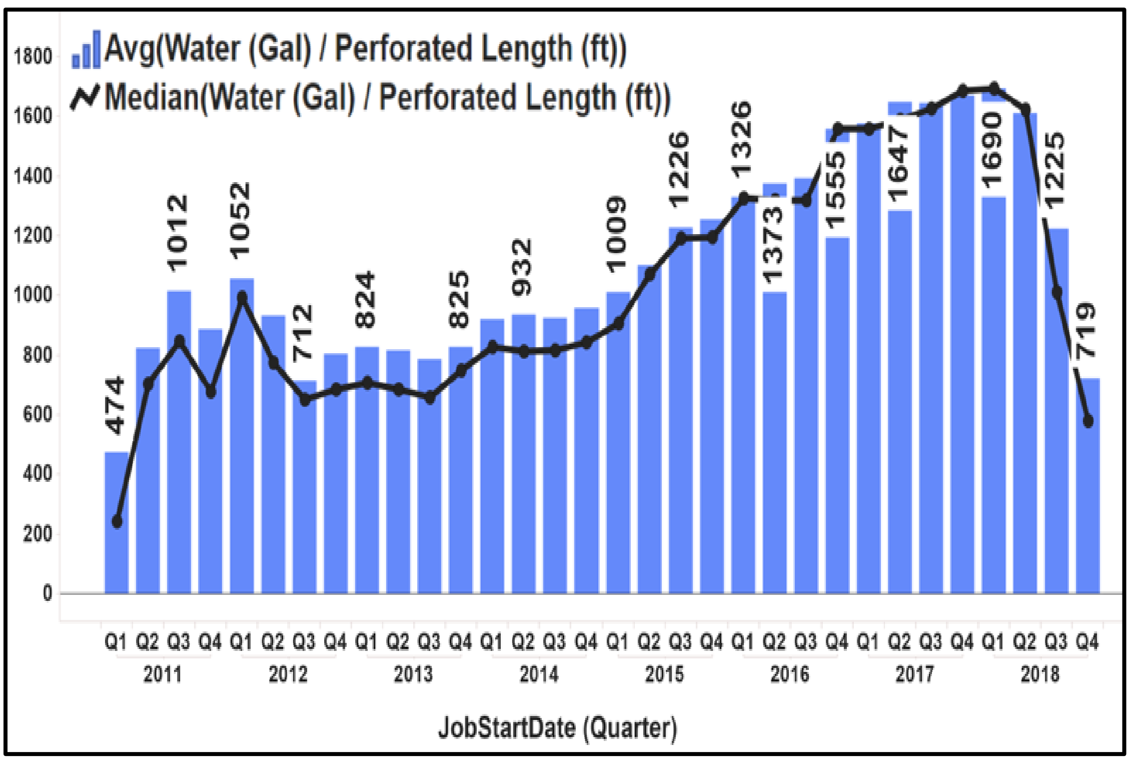


Figure 14. Average and Median Water Volume / Perforated Later Length over Time

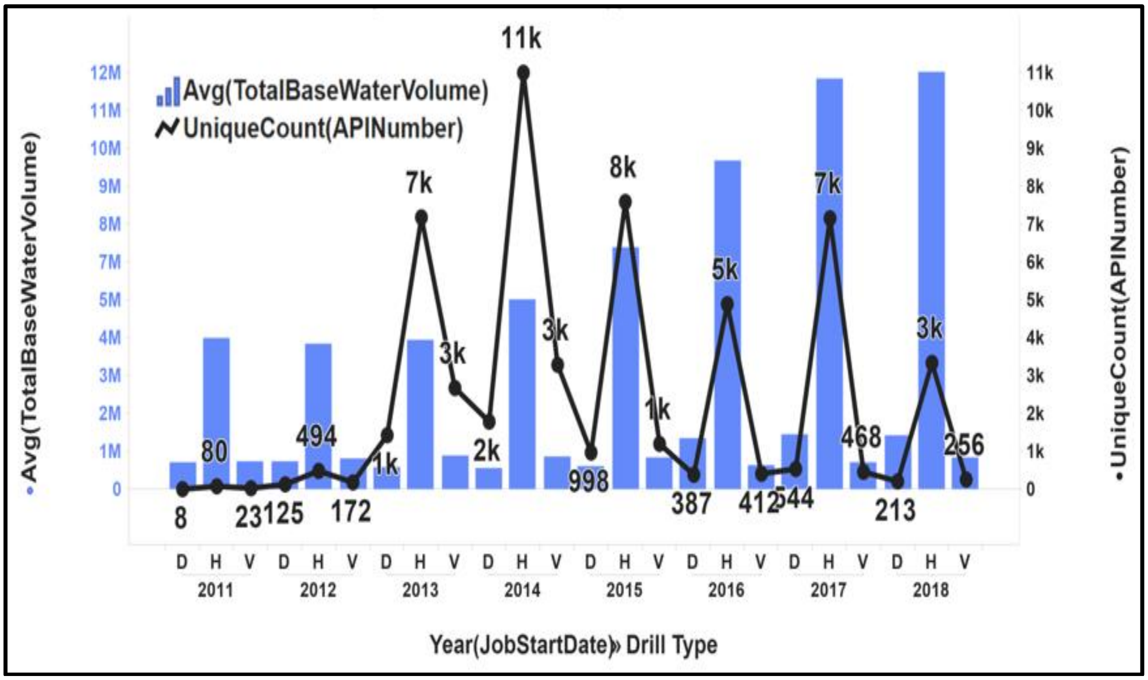


Figure 15. Average Water Use and Number of Wells over Time for Each Well Type

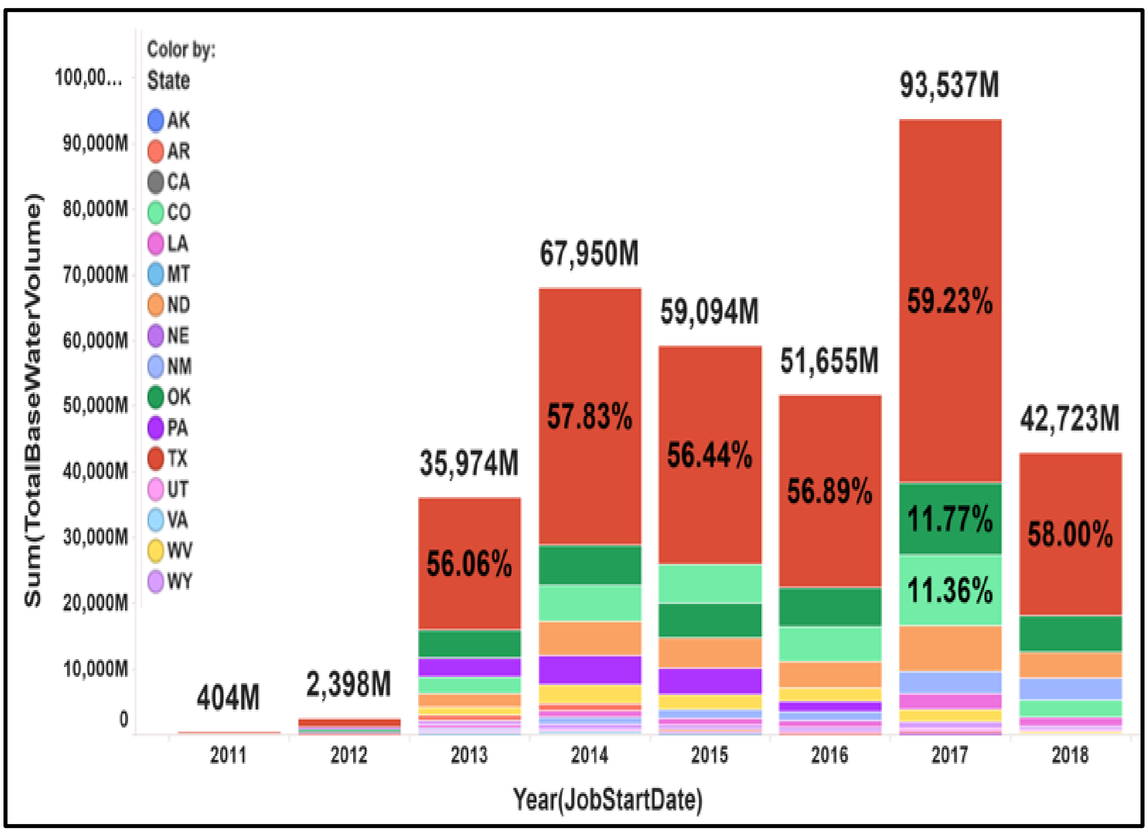
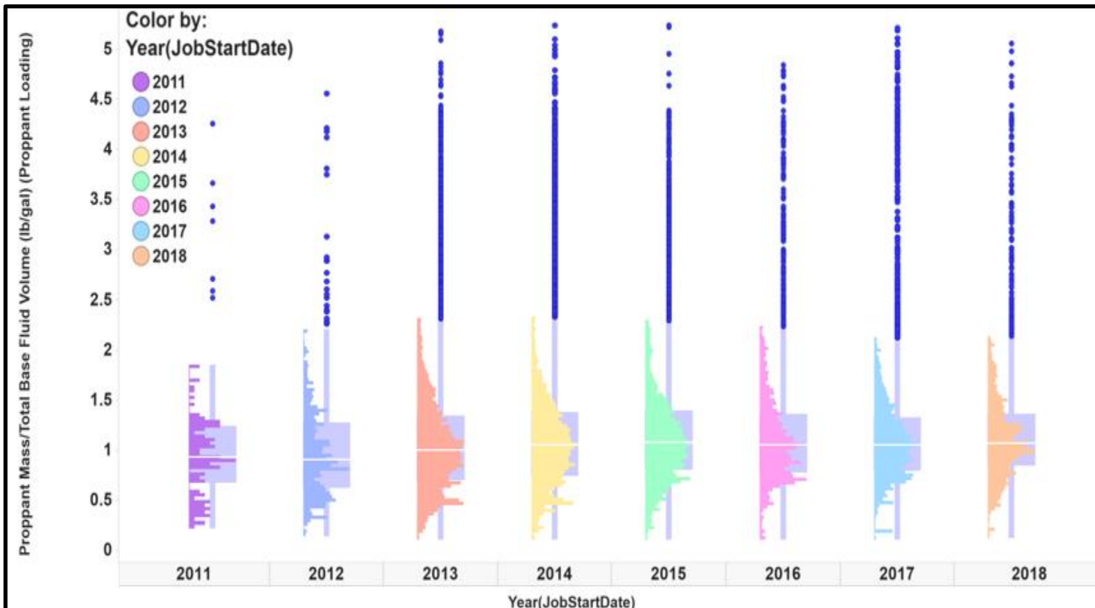


Figure 16. Water Consumption by State

3.3. PROPPANT LOADING DATA DISTRIBUTION AND TRENDS

Figure 17 shows the proppant loading or proppant concentration in pound/gallon (ppg), the plots showed a range of practical values from 0.1 to 2 ppg. Some wells had proppant concentrations exceeding 4 ppg but the distribution histogram and the box plots suggest that those values may be considered as statistical outliers and they were kept in this trend study to show the wide range of the data and to reflect the field applications. Figure 18 shows the combined bar and line chart of the average and median values of the proppant loading for each quarter of the studied period. The average and median values over time plateaued between 1.1 and 1.2 ppg.



	2011	2012	2013	2014	2015	2016	2017	2018
Count	119	824	11951	17076	10521	6178	8849	3986
Avg	1.04429	1.0055	1.10506	1.1406	1.18861	1.14055	1.11766	1.15912
Median	0.923904	0.903132	0.998756	1.04669	1.07133	1.04474	1.04832	1.06545
Sum	124.271	828.533	13206.6	19476.9	12505.4	7046.3	9890.14	4620.25
Max	4.25915	4.56553	5.18848	5.24192	5.24478	4.84805	5.21683	5.06418
Min	0.216076	0.133378	0.100804	0.100188	0.101987	0.100886	0.102911	0.117647
StdDev	0.669412	0.549761	0.625485	0.632435	0.634058	0.560005	0.541478	0.546685
Q1	0.671254	0.613176	0.694138	0.739701	0.795011	0.775549	0.78721	0.83713
Q3	1.23398	1.26963	1.34246	1.37682	1.39547	1.36109	1.32174	1.35765
IQR	0.562731	0.656455	0.648323	0.637123	0.60046	0.585543	0.534526	0.520519
Outliers	7	25	543	848	626	201	271	160

Figure 17. Proppant Mass/ Water Volume Box Plot and Data Distribution

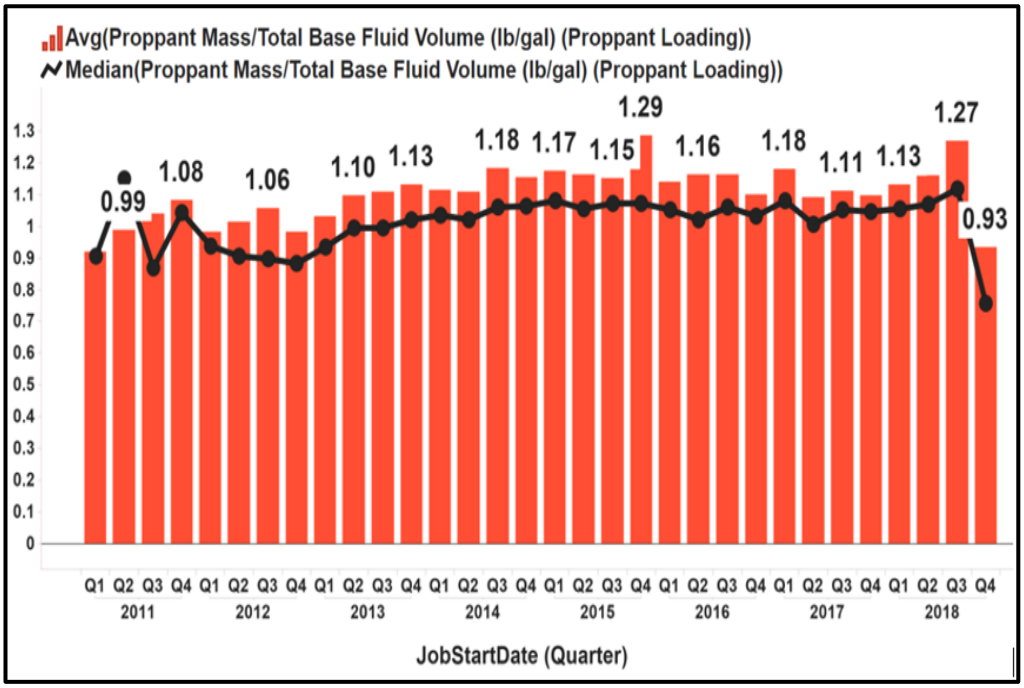
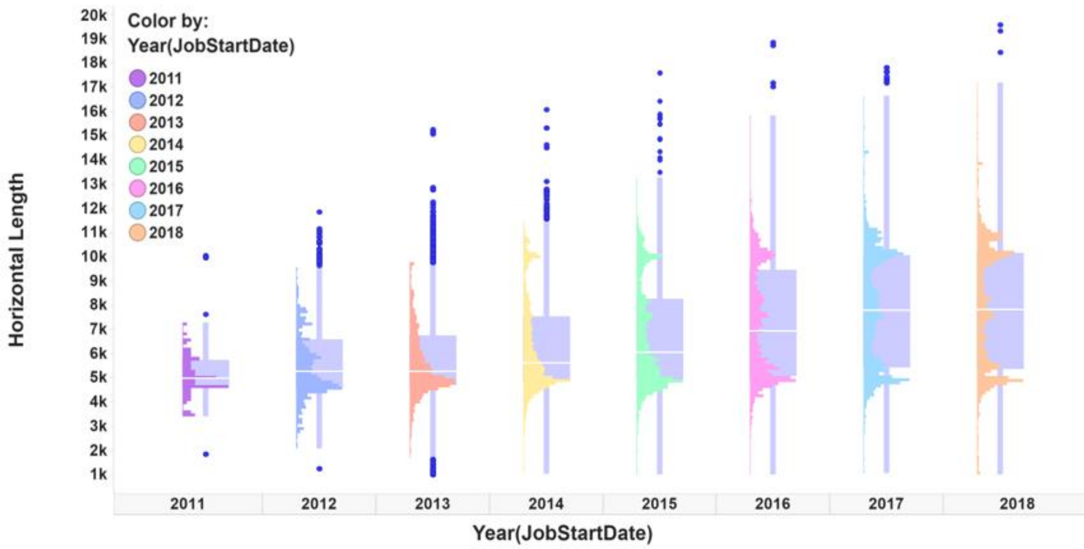


Figure 18. Average and Median Values of Proppant Loading

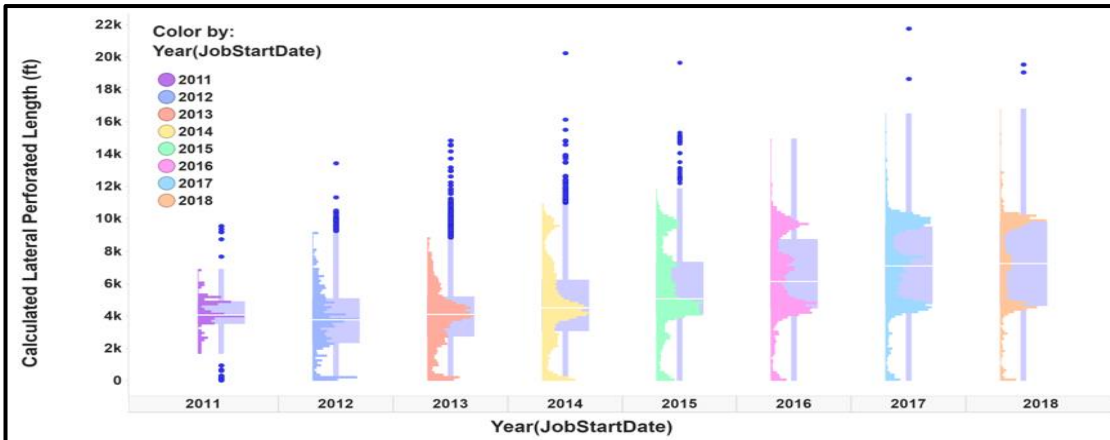
3.4. LATERAL LENGTH AND PERFORATED INTERVALS

Figure 19 and Figure 20 show the distribution histograms and box plots for the horizontal and the perforated lateral length, respectively. The plots show that the average horizontal length has been increased from 5000 to 7800 ft. The median values also corroborate this increase in horizontal length. Over the past two years, the data indicate that the horizontal length has been optimized and the median, average, and the distribution histograms confirm the similarity in the trend. Analyzing the perforated lateral length data showed that at earlier years, the perforated lateral length was approximately 70% of the drilled horizontal lateral. As time progressed, and especially over the last 3 years, the percentage has been pushed to 90%. Perforated lateral length can be used as a proxy to the number of stimulated stages which means over the time the number of stages have increased.



	2011	2012	2013	2014	2015	2016	2017	2018
Count	72	508	7482	11470	8134	5304	7667	3432
Avg	5203.39	5763.13	5910.7	6313.48	6726.53	7219.53	7820.2	7870.88
Median	4978	5255.06	5276.39	5609.47	6064.14	6937.75	7778.8	7812.77
Sum	374644	2.92767e+06	4.42239e+07	7.24156e+07	5.47136e+07	3.82924e+07	5.99575e+07	2.70129e+07
Max	10067	11861.4	15273.3	16080.6	17608.7	18860.3	17833.8	19590.6
Min	1865.38	1255.54	1011.73	1024.62	1006.47	1014.34	1064.61	1021.83
StdDev	1424.12	1870.31	1919.5	2052.09	2192.33	2362.26	2519.34	2740.16
Q1	4625.26	4572.58	4704.11	4854.41	4927.81	5081.27	5421.28	5358.63
Q3	5742.65	6579.42	6735.71	7541.91	8268.44	9471.18	10055.6	10154
IQR	1117.4	2006.84	2031.6	2687.5	3340.63	4389.91	4634.3	4795.4
Outliers	6	44	718	50	17	4	10	3

Figure 19. Horizontal Length Box Plot and Data Distribution



	2011	2012	2013	2014	2015	2016	2017	2018
Count	113	744	10720	14683	8820	5228	7569	3056
Avg	4101.76	3882.58	4214.37	4732.75	5603.91	6329.05	6934.1	7039.05
Median	4080	3775	4110	4492	5067.5	6144	7109	7246.5
Sum	463499	2.88864e+06	4.51781e+07	6.9491e+07	4.94265e+07	3.30882e+07	5.24842e+07	2.15113e+07
Max	9584	13441	14864	20267	19683	14953	21772	19563
Min	40	10	0	0	6	4	4	0
StdDev	2050.99	2375.99	2402.39	2631.9	2634.32	2699.93	2844.67	3131.81
Q1	3517	2321	2747.75	3085.5	4108	4482.5	4724	4604.75
Q3	4901	5086.25	5199.25	6244.5	7353	8746.25	9554	9823
IQR	1384	2765.25	2451.5	3159	3245	4263.75	4830	5218.25
Outliers	21	34	805	62	24	0	2	2

Figure 20. Perforated Lateral Length Box Plot and Data Distribution

Figure 21 shows a combined bar chart that represents the average horizontal length and two line-charts which represent the median horizontal length and median perforated lateral length over each quarter of the studied period. As time progressed, the horizontal and perforated length have been increasing while the gap between the median horizontal length and the median perforated lateral length have been getting narrower. This confirms that with time progression, more clusters have been added to the stimulated lateral length. In 2017 and 2018, the average well-stimulated length has reached 7 to 8 thousand feet.

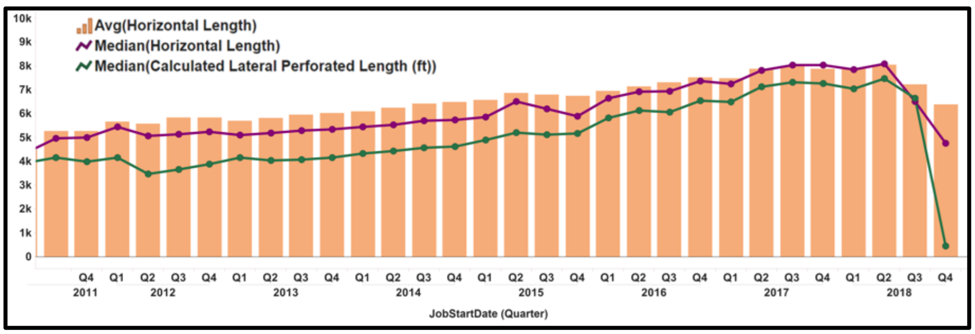


Figure 21. Average and Median Horizontal Length, Median Perforated Lateral Length over Time

3.5. DRILLING WELL TYPES AND ACTIVITIES OVER TIME

Figure 22 shows a bar chart of the number of completed wells over time (quarterly). The bars are also color-coded based on the type of the completed well (Red: Horizontal, Green: Vertical, and Blue: Deviated). The figure shows that as time progressed, more horizontal wells and less vertical and deviated wells were drilled. It also depicts the industry recent downturn which indicated by the drop in the number of completed wells starting in the first quarter of 2015. The number of completed wells started to pick up during the fourth quarter of 2016.The third and the fourth quarters of 2018 in this figure are not representative of the actual wells completed over this time period due to the fact that the data was collected during the last quarter of 2018 and the lag in FracFocus reporting system.

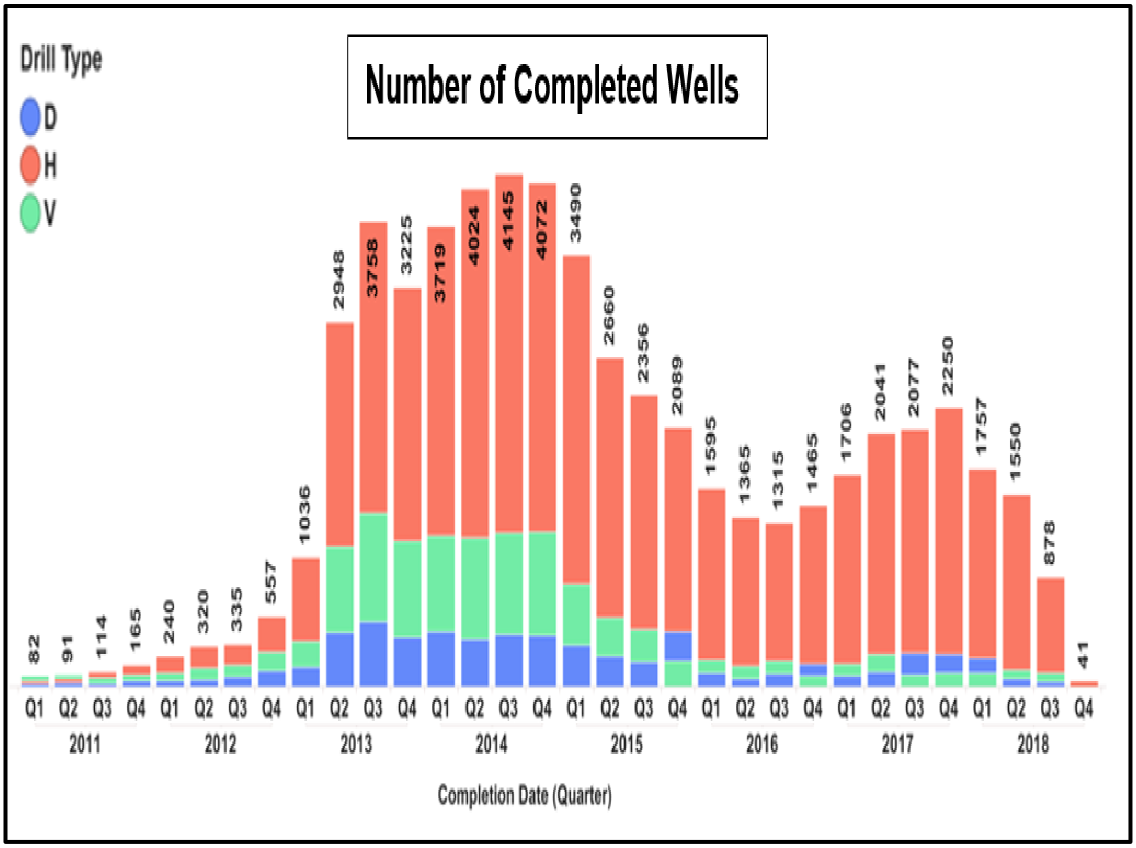


Figure 22. Data Sample of Number of Wells Completed over Time Colored by Completion Type

3.6. INITIAL PRODUCTION RESPONSE TO STIMULATION AND COMPLETION VARIATIONS

Figure 23 and Figure 24 show sensitivity analyses of the water gal/stimulated foot and proppant lbs/stimulated foot on average gas production, respectively. Both figures indicated that over 4000 gal/ft of water and 4000 lbs/ft of proppant, the overall average gas production starts to decrease. This value should be considered as the upper bound when designing stimulation for gas wells. Figure 25 and Figure 26 confirm the limits of 4000 gal/ft and lbs/ft versus the gas production per stimulated foot (MCF/Day)/ft.

Figure 27 and Figure 28 show sensitivity analyses of the water gal/stimulated foot and proppant lbs/stimulated foot on average oil production, respectively. Both figures showed that above 3000 gal/ft of water and 3500 lbs/ft of proppant, the overall average oil production is decreasing. This value should be considered as the upper bound when designing stimulation for oil wells. Figure 29 and Figure 30 confirm that the limit is 3500 for both gal/ft and lbs/ft versus the oil production per stimulated foot (BBL/Day)/ft.

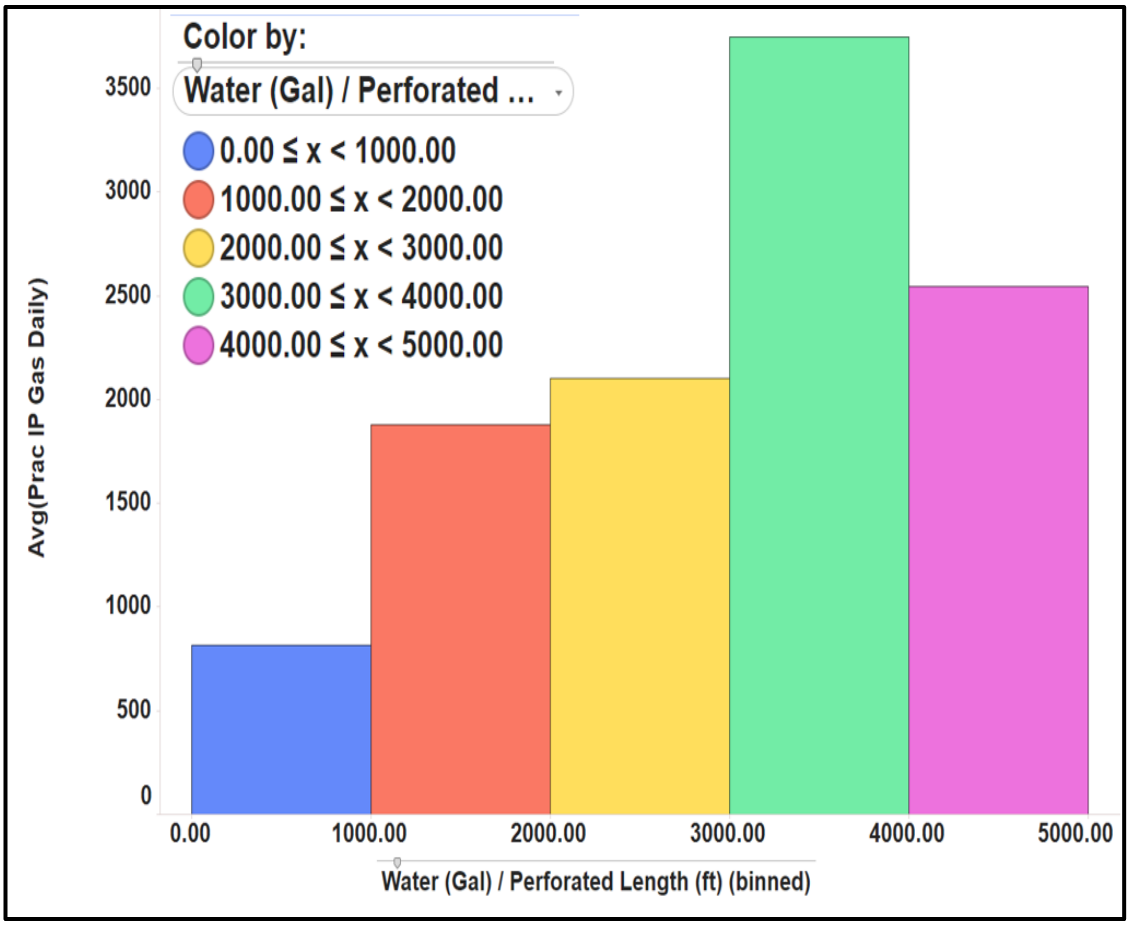


Figure 23. Water/ft Effect on Average Gas IP

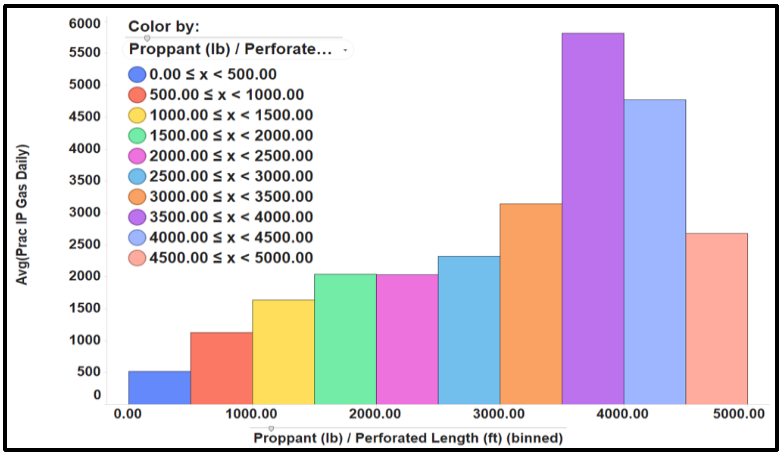


Figure 24. Proppant/ft Effect on Average Gas IP (MCF/Day)

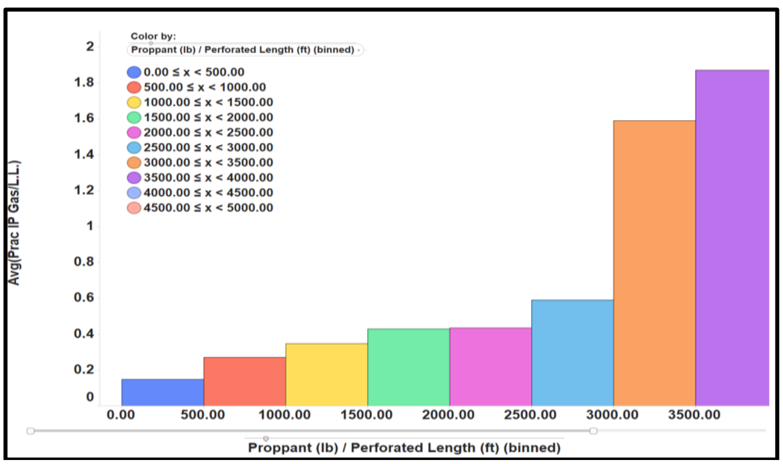


Figure 25. Proppant/ft Effect on Average Gas IP/Perforated Length (MCF/Day)/ft

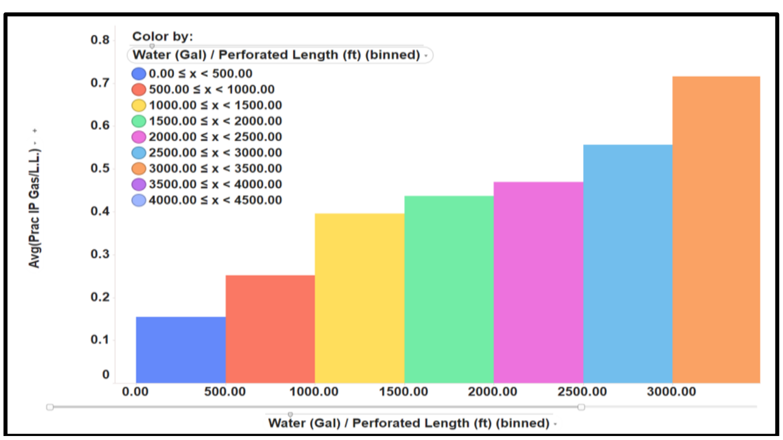


Figure 26. Water/ft Effect on Average Gas IP/Perforated Length

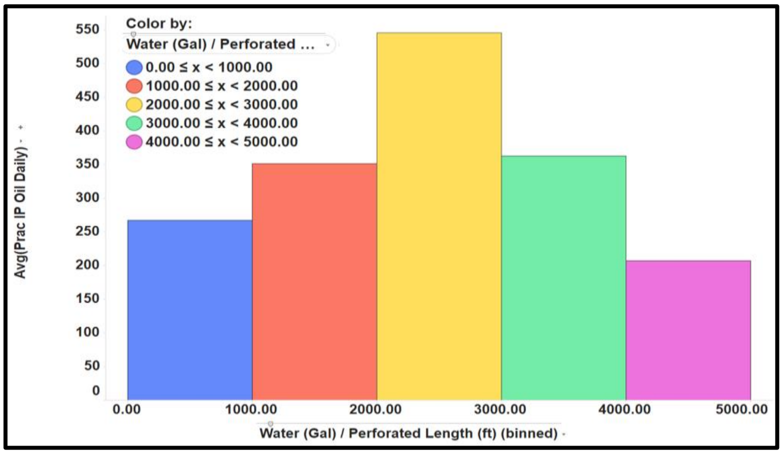


Figure 27. Water/ft Effect on Average Oil IP

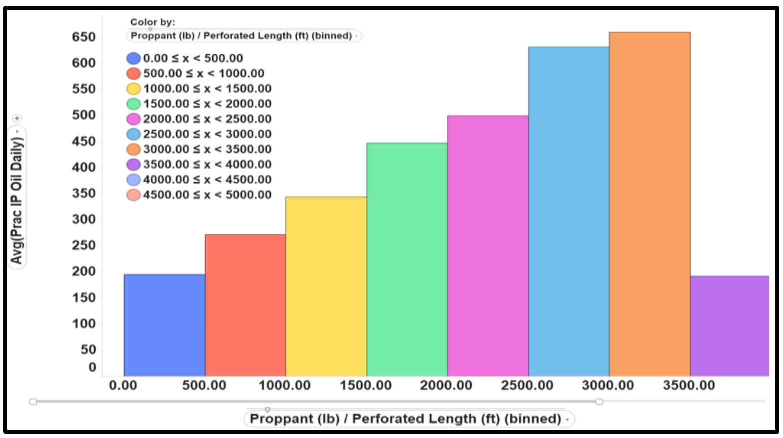


Figure 28. Proppant/ft Effect on Average Oil IP

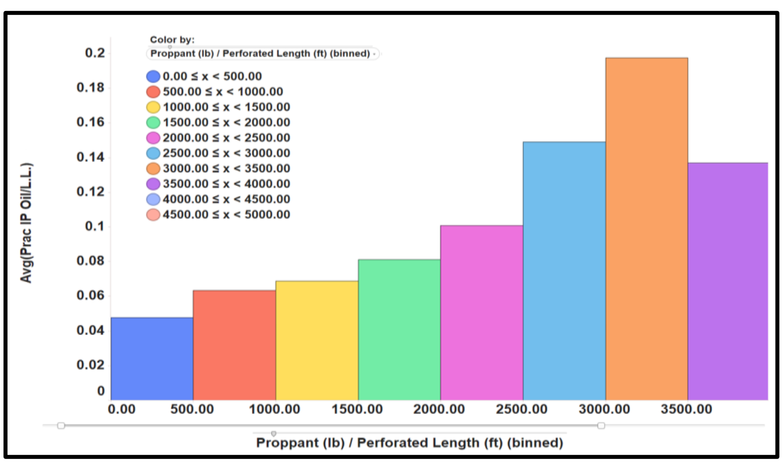


Figure 29. Proppant/ft Effect on Average Oil IP/Perforated Length

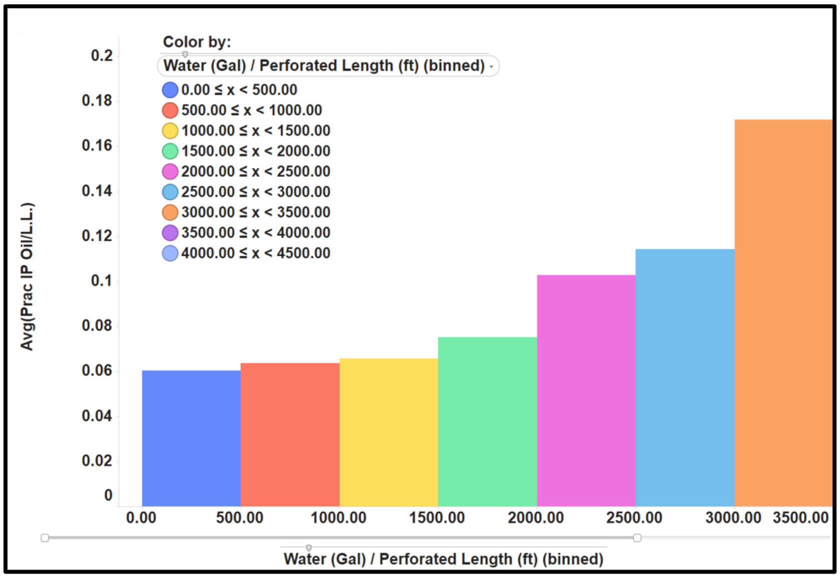


Figure 30. Water/ft Effect on Average Oil IP/Perforated Length

4. CONCLUSIONS

Initially, a comprehensive database was built from combining processed stimulation data from FracFocus website as well as production and completion data from DrillingInfo database. The database was used to demonstrate the overall trends of water, proppant and completion activities across the United States. Practical IP was considered as the initial production rate for the comparison purpose in this study. The focal point of using practical IP was to avoid the late production start date during the first month of production, and it was calculated by dividing the second-month cumulative production by the number of produced days during that month. Moreover, the perforated lateral length was used as a proxy to stage count. Consequently, the following conclusions are reached from the study:

- The number of wells in 2011 and 2012 was less than the following years because the FracFocus reporting website was established during the second quarter of 2011 and it took some time for the operators to cooperate and report all their wells as part of the states adopted regulations.

- The average proppant mass increased over time from 2 million to over 10 million pounds by 2018. The same was true for the average water volume which ranged from 2 to 10 million gallons.
- For 2017 and 2018 water average and median volume was almost the same which indicates that the operators have reached to the optimized values of water volumes.
- The data distribution histograms over the past two years showed a wider range in the stimulation parameters ranges which reflect the operators' increased operational range.
- The practical ranges of proppant loading ranged from 0.1 to 2 pound/gallon.
- Water and proppant values were normalized to the perforated lateral length to eliminate the effect of the difference in the perforated lateral length on the analysis.
- The average water per foot and proppant per foot have more than doubled between 2013 and 2017 onward.
- The average horizontal length has increased from 5000 to 7800 ft over the studied period.
- The perforated lateral length has increased from 70% of the horizontal length to 90% by 2017.
- As time progressed, the operators tend to drill more horizontal wells and less vertical and deviated wells.

- Plotting the number of completed wells versus completion time reflected the recent industry downturn which started during the first quarter of 2015 and started to recover in the fourth quarter of 2016.
- Sensitivity curves of proppant/ft and water/ft versus the practical initial production showed that 4000 gal/ft and lbs/ft can be the limits for gas wells while 3500 gal/ft and lbs/ft can be the limits for oil wells.
- Analyzing the proppant and water total consumption per state showed that Texas has taken the lead with more than 50% of the total consumed water and proppant nationwide.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to DrillingInfo team and the co-founder and director of university outreach Mr. Mark Nibbelink for their generosity in providing access to the U.S. production and completion data.

REFERENCES

Carman, P., Li, C., & Wheeler, R. (2018, September 24). Unorthodox Fluids for Unconventional Completions: Maximizing Production and Avoiding Pitfalls. Society of Petroleum Engineers. doi:10.2118/191751-MS

Carman, P., Li, C., & Wheeler, R. (2018, September 24). Unorthodox Fluids for Unconventional Completions: Maximizing Production and Avoiding Pitfalls. Society of Petroleum Engineers. doi:10.2118/191751-MS

FracFocus website. Retrieved on February/02/2019, from http://fracfocus.org/

DrillingInfo website. Retrieved on February/07/2019, from https://info.drillinginfo.com/

Khvostichenko, D., & Makarychev-Mikhailov, S. (2018, August 9). Oilfield Data Analytics: Linking Fracturing Chemistry and Well Productivity. Unconventional Resources Technology Conference. doi:10.15530/URTEC-2018-2903086

King, G. E. (2010, January 1). Thirty Years of Gas Shale Fracturing: What Have We Learned? Society of Petroleum Engineers. doi:10.2118/133456-MS

Luo, H., & Zhang, Y. (2018, October 5). Mining FracFocus and Production Data for Efficacy of Fracturing Fluid Formulations. Society of Petroleum Engineers. doi:10.2118/191787-18ERM-MS

Myers, R. R., Knobloch, T. S., Jacot, R. H., & Ayers, K. L. (2017, October 4). Production Analysis of Marcellus Shale Well Populations Using Public Data: What Can We Learn from The Past? Society of Petroleum Engineers. doi:10.2118/187505-MS

Olaoye, O., & Zakhour, N. (2018, September 24). Utilizing a Data Analytics Workflow to Optimize Completion Practices While Leveraging Public Data: A Permian Basin Case Study. Society of Petroleum Engineers. doi:10.2118/191416-MS

Romero, D. H., & Poston, S. W. (2016, May 10). Data Mining and Analysis - Eagle Ford Formation, South Texas. Society of Petroleum Engineers. doi:10.2118/179994-MS

Srinivasan, K., Ajisafe, F., Alimahomed, F., Panjaitan, M., Makarychev-Mikhailov, S., & Mackay, B. (2018, August 28). Is There Anything Called Too Much Proppant? Society of Petroleum Engineers. doi:10.2118/191800-MS

Yang, Y., Robart, C. J., & Ruegamer, M. (2013, February 4). Analysis of U.S. Hydraulic Fracturing Design Trends. Society of Petroleum Engineers. doi:10.2118/163875-MS

VI. DESCRIPTIVE DATA ANALYTICS FOR THE STIMULATION, COMPLETION ACTIVITIES, AND WELL'S PRODUCTIVITY IN THE MARCELLUS SHALE PLAY

Mustafa A. Al-Alwani, Shari Dunn-Norman, Larry K Britt, Husam H. Alkinani, Abo Taleb T. Al-Hameedi, Atheer M. Al-Attar, Hector A. Trevino, Waleed Al-Bazzaz

Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA

ABSTRACT

Drilling, completion, and stimulation designs have changed over time as a result of the oil and gas industry's ongoing efforts to increase well productivity. Over the last five years hydraulic fracturing treatments, represented by the volume of pumped water and the amount of proppant utilized, have increased significantly, along with the lengths of horizontal wells. This work represents a large-scale descriptive analysis study to illustrate the trends and the range of completion, stimulation and production parameters in the Marcellus Shale play of the Appalachian Basin between 2012 and the last quarter of 2017 (2012-2018).

A database was created by combing stimulation fluids and proppant data from the FracFocus 3.0 chemical registry, with completion and production data from the DrillingInfo database. More than 2000 Marcellus Shale wells were utilized in this study. The data were processed and cleaned from outliers. Box plots and distribution bar charts are presented for most of the parameters in this study, to show the range in values for each parameter and its frequency of use. The stimulation parameters were normalized to perforated lateral length in order to compare productivity between the wells.

Trends identified in this study show how operators in the Marcellus have increased the use of hybrid fracturing fluids, in addition to increasing water and proppant volumes over time. The work also illustrates the point at which increasing fracture treatment volumes no longer increases production rate.

This paper demonstrates the utility of integrating publicly available databases to examine well completion trends in the Marcellus. The work also provides a summary of well response as a function of treatment volume over the five-year study period.

1. INTRODUCTION

Over the last decade, the United States experienced tremendous growth in unconventional oil and gas resource development, primarily in the shale plays. This growth is mainly attributed to the success in coupling horizontal drilling with multi-stage hydraulic fracturing to produce oil and gas from ultra-low permeability, or 'tight' resources. Energy information administration (EIA) (2018) reports over 120,000 active horizontal wells completed since the year 2000. A large amount of data from drilling, completion, stimulation, and production of these wells is available, often in different databases. Analysis of such data can provide insights into completion trends and their impact on well productivity, methods of gaining operational efficiency, cost saving, as well as health, safety and environmental considerations.

Shale development history, trends, and best practices have been documented by many studies in the literature. Some earlier studies lacked formation descriptive data which made it difficult to utilize the findings and generalize it to the field level because the data were limited to small number of localized wells representing a small, area specific trend (Carman et al, 2018). However, with data being shared more readily and more accessible databases available, more recent studies have provided reviews of shale development trends (Yang et al., 2013; Arthur et al., 2014; Romero and Poston, 2016; Luo and Zhang, 2018; Olaoye and Zakhour, 2018; Srinivasan et al., 2018; Carman and Wheeler, 2018; Weijers et al., 2019). Myers et al. (2017) provides a production analysis summary of the Marcellus Shale. Other recent case studies of various shale development trends provide useful insight to maximize well productivity (Al-Alwani et al., 2019a; Al-Alwani et al., 2019b; Al-Alwani et al., 2019c; Al-Alwani et al., 2019d; Al-Alwani et al., 2019e).

As with many other industries, the oil and gas industry has begun utilizing descriptive and predictive data analytics in processes optimization, future planning, and to help in making data-based decisions. The use of data analytics in the oil and gas industry has been described by Khvostichenko and Makarychev-Mikhailov (2018) as the most dynamically growing fields. Many of the leading operators and service companies have realized the power of data-driven decisions and have developed specialized teams to review their accumulated data and analyze trends, in an effort to improve performance. Although large scale data analytics are relatively new to the oil and gas industry, this field has been advanced by many of the studies noted previously. Mohaghegh et al. (2017) demonstrates optimizing Marcellus well production through data analysis using artificial intelligence (AI) methods. Shahkarami et al. (2018) presents a Marcellus data analysis using machine learning algorithms. Data science and machine learning approaches are also used extensively by the Department of Energy oil and gas research to process the complex data streams generated by drilling, stimulation, and production to increase the productivity of the oil and gas wells (Oil & Gas Journal, 2019).

In this study, data were collected from FracFocus 3.0 (2019), the publically available chemical registry containing chemical and proppant data for stimulated wells. These data were combined with drilling, completion and production data available through DrillingInfo (2019) and processed to create a final database that integrates stimulation and completion parameters with production data. This database was analyzed to identify trends of completion and stimulation over a five-year period (2012-2017) and to quantify the stimulation and completion parameters effects on the wells' productivity. Proppant loads, water volumes, lateral length increases, types of stimulation fluids used, the response of combined stimulation and completion parameters on production, and other normalized parameters are all discussed in this work.

2. MARCELLUS SHALE AND ITS CONTRIBUTION TO ENERGY SUPPLY

The Marcellus Shale is a gas play located in the Appalachian Basin. It consists mainly of black shale interbedded by some layers of limestone. The formation was known to contain hydrocarbons for many years but was uneconomic to produce with vertical well completions. In 2003, Range Resources drilled a well using horizontal well technology combined with multistage hydraulic fracturing that had been found to be efficient in the Barnett Shale in Texas (Harper, 2008). This well proved successful and, as of November, 2017, EIA reported more than 11,300 wells have now been drilled in the Marcellus Shale as shown in Figure 1 (EIA, 2017).

The Marcellus Shale is now one of the largest shale plays in the U.S. and contributes significantly to U.S. gas production. Over the past decade, the Appalachian Basin has contributed 85% of all the growth in U.S. gas production, and it is now equal to approximately one-third of current U.S. gas production. Predictions suggest that Marcellus gas production will account for about 45% of the U.S. entire gas production by 2040 (Oil & Gas Journal, 2019).

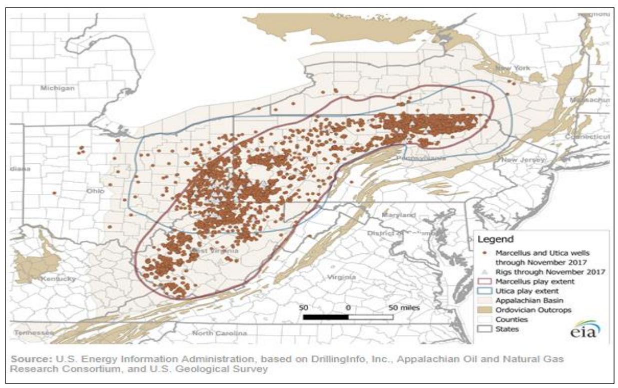


Figure 1. Marcellus producing wells as of November 2017 (EIA, 2017)

3. DATA AND METHODS

Data for this study were collected from FracFocus as well as DrillingInfo. A comprehensive database was built by extracting the chemical ingredients data for every Marcellus well stimulated and reported in FracFocus. Chemcial volumes, trades names, and use were obtained, along with all chemical mass percentages. FracFocus data were processed then combined with completion and production data for the Marcellus Shale wells obtained from DrillingInfo. The data from FracFocus are available to the public but require rigorous cleaning and processing for use.

The main objective in processing the data from FracFocus was to identify and categorize the stimulation fluid types based on the individual chemical ingredients that are reported in the list of treatment ingredients.

After extracting the data from FracFocus a series of data processing and formatting was applied to combine all the data in one file to be ready for the next phase of data analysis. Chemical ingredients were reported by several identifiers such as the chemical ingredient name, commercial name, and the chemical abstract service number (CAS#). There is inconsistency in the way fracturing fluids and their ingredients are named and reported in FracFocus, making it difficult to use chemical names in the analysis. Many operators simply report the names based on their own standards. In addition, there are many chemical naming abbreviations and typos. For this study and for the purpose of classifying the chemical ingredients into several groups (e.g. gelling agents, cross-linkers, surfactants), the CAS# was chosen to be the main identifier for each chemical ingredient. A data analysis technique, referred to as TreeMap charts, was applied to sort the chemical data into 19 prospective groupings based on their CAS# and the number of wells sharing the same chemical ingredient. Environmental protection agency (EPA) safer chemical ingredients list and a study conducted by Helmholtz Zentrum München, Institute of groundwater ecology (which provides a complete list of all hydraulic fracturing chemicals extracted from the Waxman list and the FracFocus database) were utilized in identifying the prospective chemical groups (Elsner & Hoelzer, 2016; EPA, 2019).

After the 19 chemical groups were generated, the groups were individually processed to eliminate any potential duplication in the reporting process from FracFocus. Once duplication issues were resolved, re-fractured wells were identified in the data processing workflow and grouped separately to eliminate aggregating several frac-jobs' ingredients into one well rather than separately assigning the appropriate number of fracturing jobs per well. Once all the ingredients mass percentage for the specific chemical group were summed up and assigned for each well, the following step was to combine all the chemical groups together into one dataset representing the processed FracFocus. Following that, multiple new parameters were calculated based on the processed data such as the calculated total proppant mass and the ingredients' concentrations. Figure 2 shows a workflow that combines all the processed chemical groups obtained from FracFocus. The final database was created by matching the production and completion data from DrillingInfo with the chemical components for each well. This combined database was quality checked. Several normalized parameters were also introduced to the database to help in the analysis phase and to generate production, completion, and stimulation insights for the Marcellus Shale stimulated wells. Al-Alwani et al. (2019c) presents a more detailed discussion of the procedures used to create the final database.

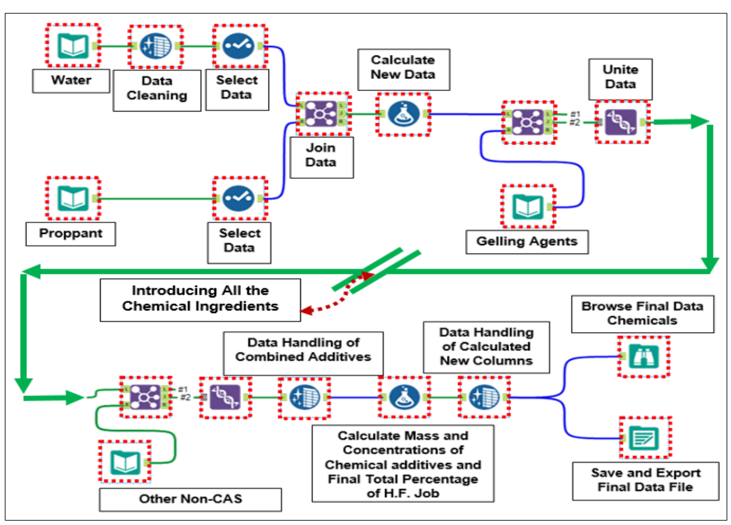


Figure 2. Data Processing Workflow to Combine All of the FracFocus Processed Data into One Database File

More than 2000 Marcellus wells were identified with sufficient data for trend analysis and to provide insights regarding completions in the Marcellus Shale. The wells, and their type of stimulation fluid used, are distributed widely across Pennsylvania and West Virginia as illustrated in Figure 3 and Figure 4 shows the number of wells per each stimulation fluid type between 2012 and 2018. Water stimulation fluid treatments dominate all other fluid treatment over the entire period. Beginning in 2015, hybrid fluid treatments increase and in 2017 are only slightly less than water treatments. Both linear gel and cross linked gel treatments have grown in use in 2017, which may be attributed to longer lateral lengths.

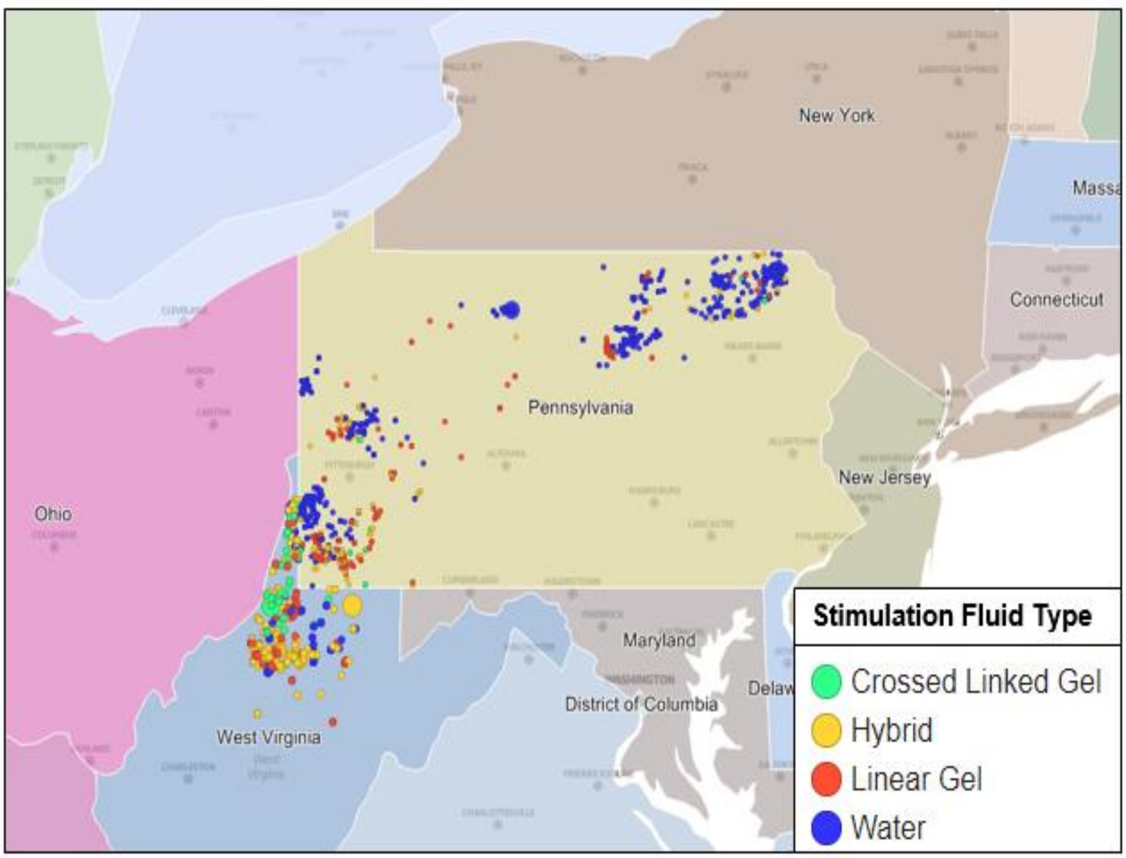


Figure 3. Appalachian Basin Wells Distribution on the Map

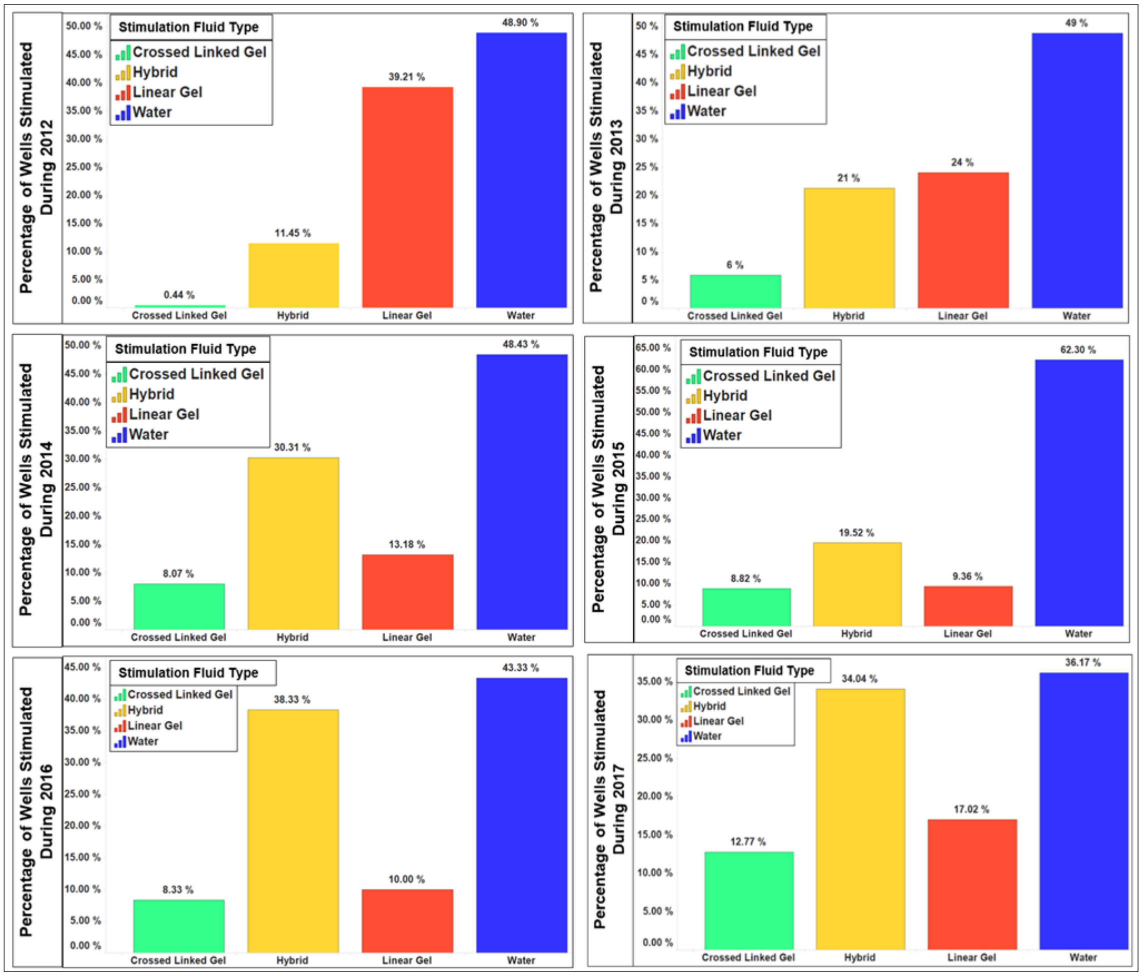


Figure 4. Number of Stimulated Wells for Each Stimulation Fluid Type in the Marcellus Shale (2012-2018)

4. RESULTS AND DISCUSSION

This section illustrates temporal changes in the types of fracturing fluids used, median water volume pumped per well and the median proppant loading per well. The section also provides the distribution values for different stimulation and completion parameters such as proppant mass pumped, water volume injected, horizontal length of the wells' lateral and the perforated segments. The perforated lateral length is used to normalize treatment volumes. Short- and long-term productivity response to the normalized water and proppant injected per perforated foot of horizontal lateral is presented.

4.1. HORIZONTAL LATERAL AND PERFORATED LENGTH TRENDS IN THE MARCELLUS SHALE

The wells in this study were all horizontal wells with multi-stage hydraulically fracturing and producing from the Marcellus Shale. The lateral length data and the gross perforated interval (which was calculated by subtracting the upper perforation measured depth from the lower perforation measured depth) were collected from DrillingInfo. The lateral length and perforated length data were cleaned from any extreme outliers and validated using the box plots charts as indicated in Figure 5 and Figure 6 Both figures show box plots superimposed by distribution diagrams of the horizontal lateral length and the gross perforated interval length, respectively. The box plots and the distribution diagrams were generated for each stimulation fluid type to show the application range of the lateral length as well as to identify and eliminate the outliers from skewing the average values of lateral length for any further analyses. Figure 7 shows box plots and distribution charts for the ratio between the perforated lateral length to the horizontal lateral length. The distribution bars indicate that most of the horizontal laterals were 70 % to 100 % perforated. In some cases, the ratio indicated more than 100% which means that the upper perforation length was extended to the build section of the well. Figure 8 shows the value of the average lateral length for each stimulation fluid type. Hybrid fluid treatments were associated with the longest average lateral lengths while water frac fluids were associated with the shortest average lateral length. Figure 9 shows the temporal change in Marcellus lateral well lengths, noted by both the average and median values of horizontal lateral and perforated lengths. The bar chart represents the average value while the line chart represents the median value of the lateral and perforated lengths. This figure shows that both the average and median values of the lateral and perforated lengths have increased every year over the time period of 2012 to 2017. The median and the average values are approximately within the same value which also indicates that the data distribution is uniform and clean of outliers which results in accurate analyses. The figure also indicates that the 88-90 % of the horizontal laterals were perforated and hydraulically fractured in the Marcellus.

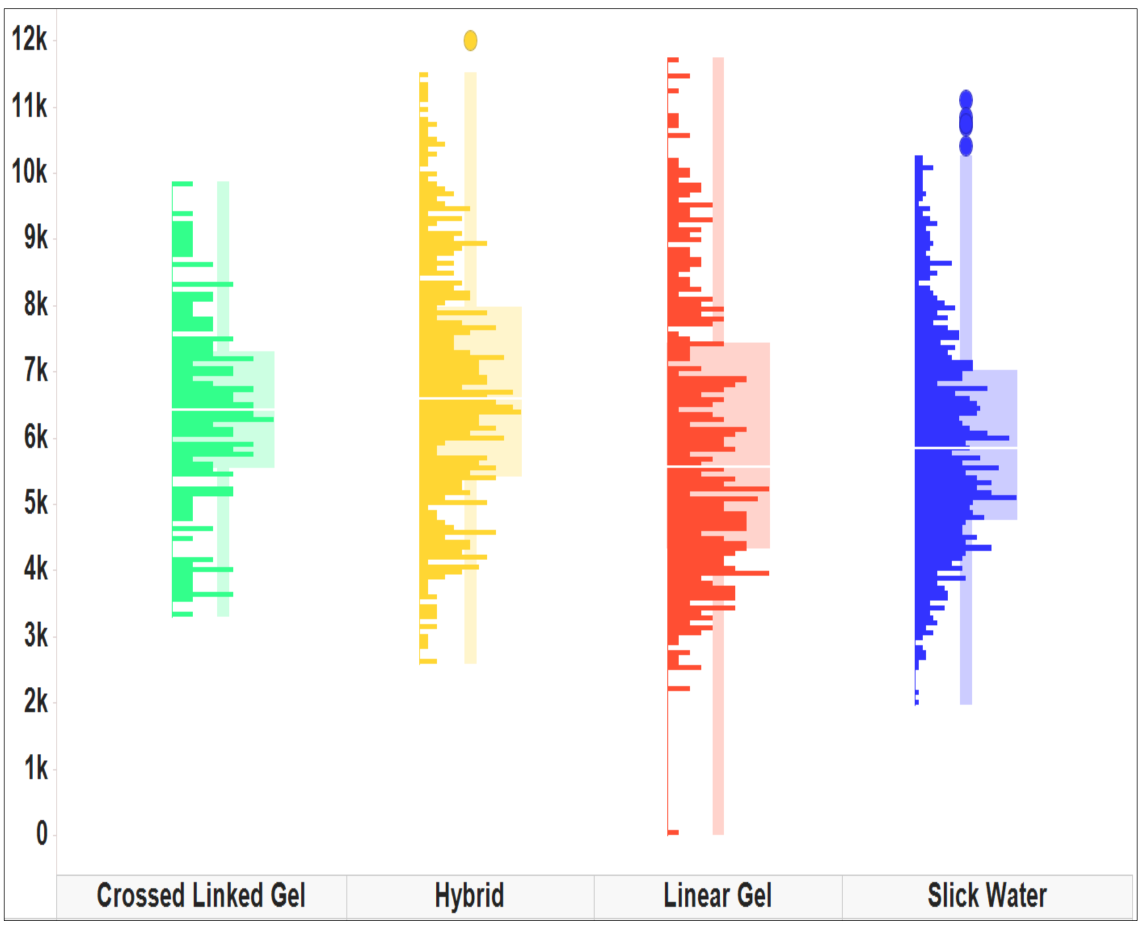


Figure 5. Box Plots and Distribution Charts of Horizontal Lateral Length of the Stimulated Wells for Each Stimulation Fluid Type in the Marcellus

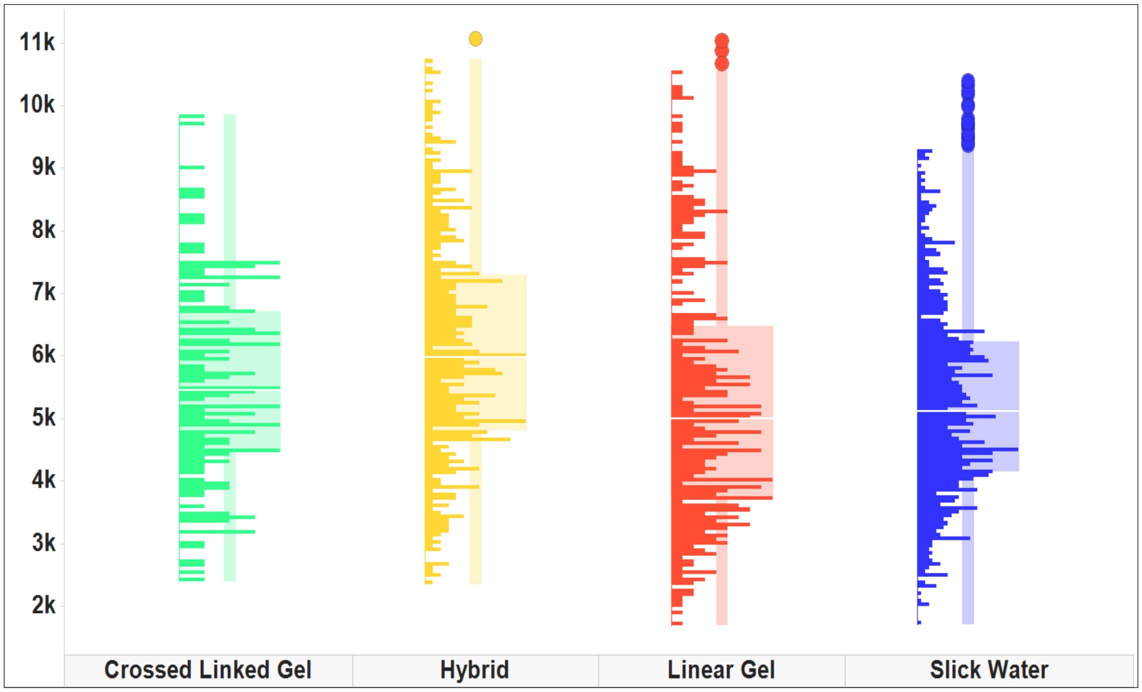


Figure 6. Box Plots and Distribution Charts of the Perforated Horizontal Lateral Length of the Stimulated Wells for Each Stimulation Fluid Type in the Marcellus

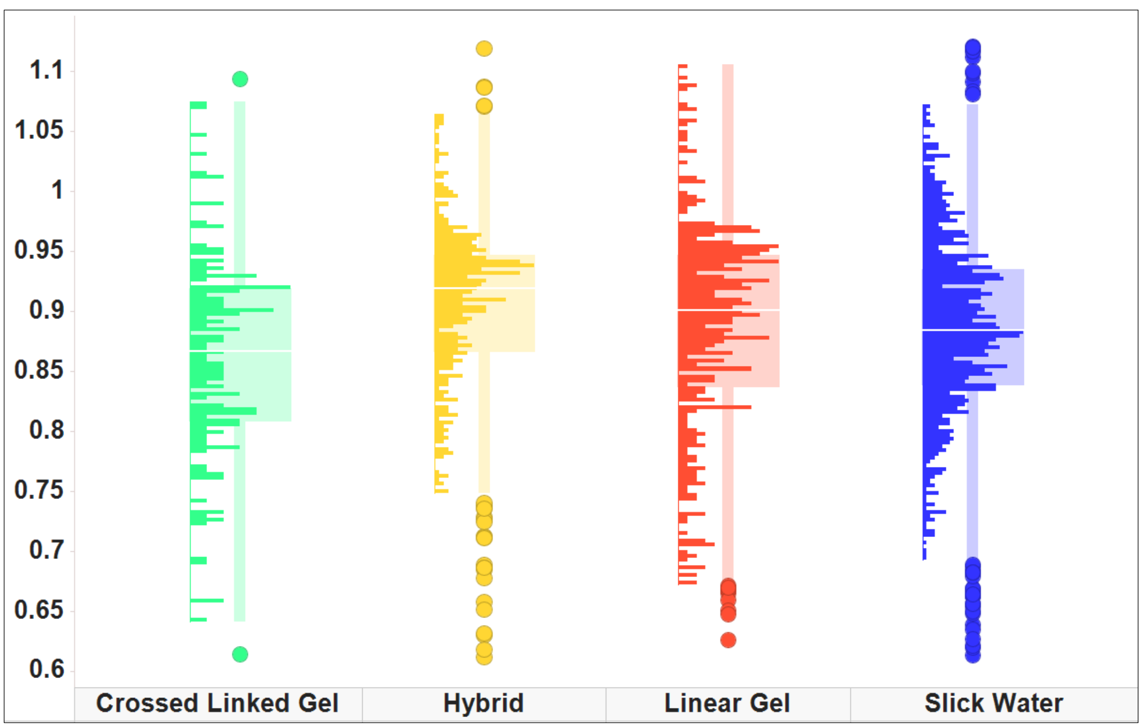


Figure 7. Box Plots and Distribution Charts of the Ratio between Perforated Lengths to the Lateral Length for Each Stimulation Fluid Type in the Marcellus

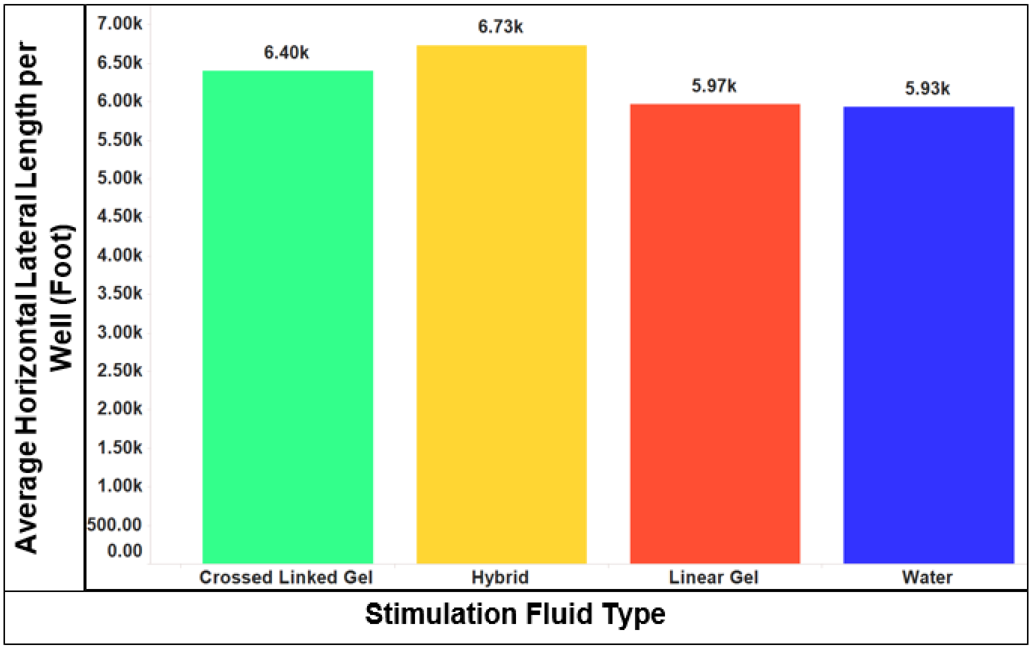


Figure 8. Bar Chart Illustrating the Average Horizontal Lateral Length for Each Stimulation Fluid Type in the Marcellus Shale

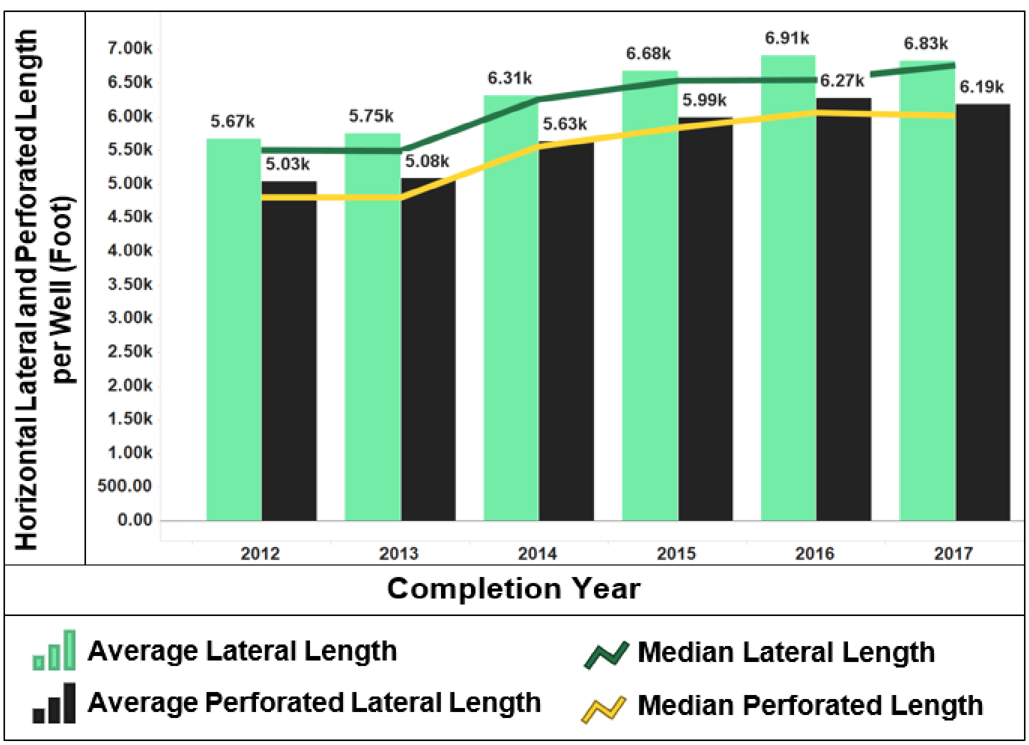


Figure 9. Combined Bar and Line Charts of Average and Median Horizontal Lateral Length and Perforated Lateral Length per Well in the Marcellus Shale

4.2. PROPPANT UTILIZATION TRENDS IN THE MARCELLUS

To better understand the amount of pumped proppant per well in the Marcellus Shale, Figure 10 and Figure 11 were prepared to show the application range of the amount of proppant pumped per well as well as the amount of pumped proppant per stimulated lateral length for each well, respectively. The box plots show the four quartiles ranges of proppant amount and the distribution charts illustrate the number of wells that utilize the same range of pumped proppant mass. Figure 12 and Figure 13 illustrate the average and median values of the amount of proppant pumped per well over time and the normalized proppant pumped to the perforated lateral length per well over time, respectively. It can be seen that both the average and the mean values of pumped proppant per well, and the amount of proppant pumped per stimulated foot, increased throughout 2012 to 2018.

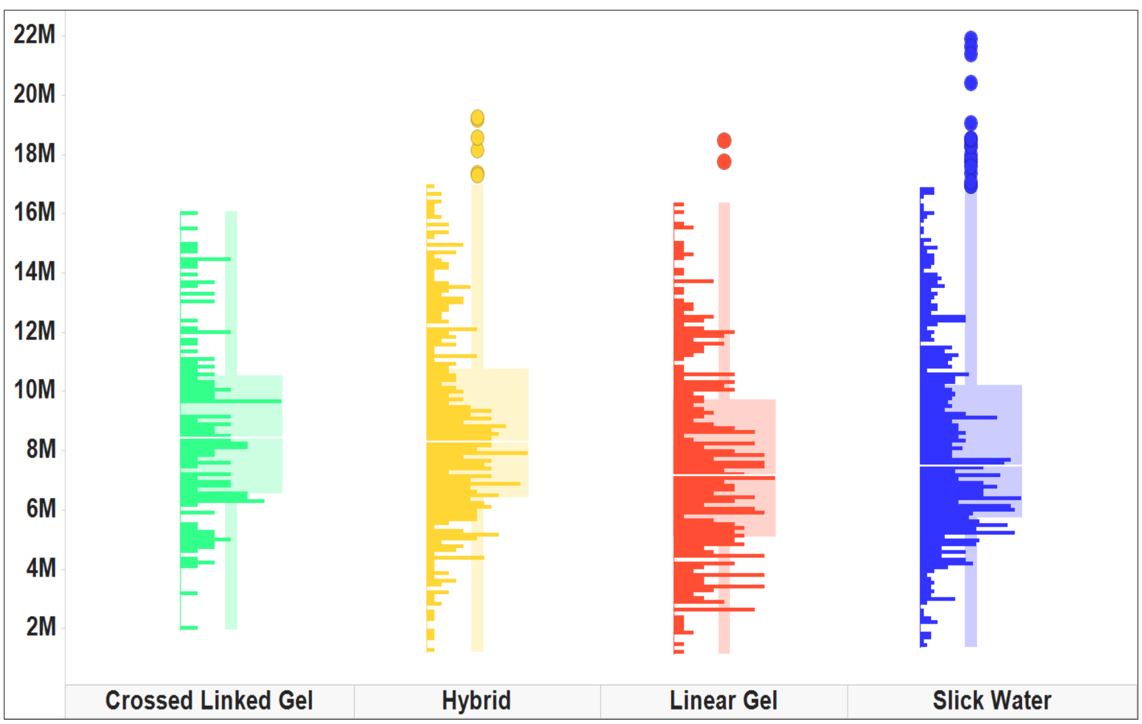


Figure 10. Box Plots and Distribution Charts of Proppant Pumped per Well for Each Stimulation Fluid Type in the Marcellus

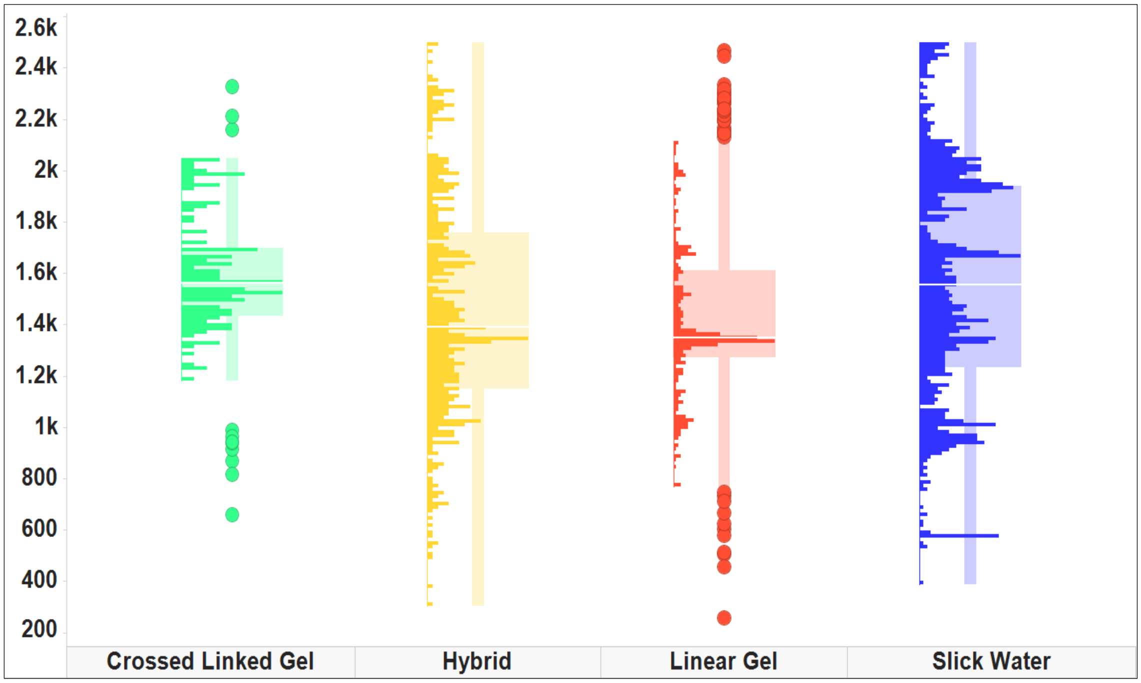


Figure 11. Box Plots and Distribution Charts of Proppant per Perforated Length for Each Stimulation Fluid Type in the Marcellus

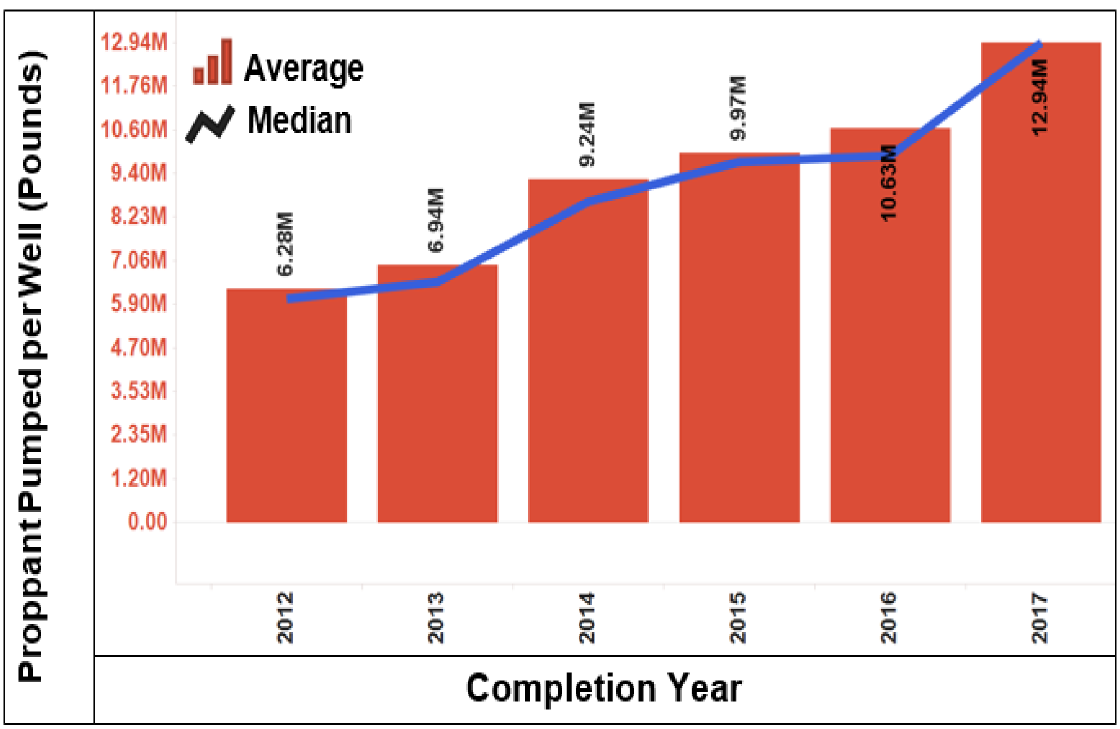


Figure 12. Combined Bar and Line Charts of Average and Median Proppant Pumped per Well in the Marcellus

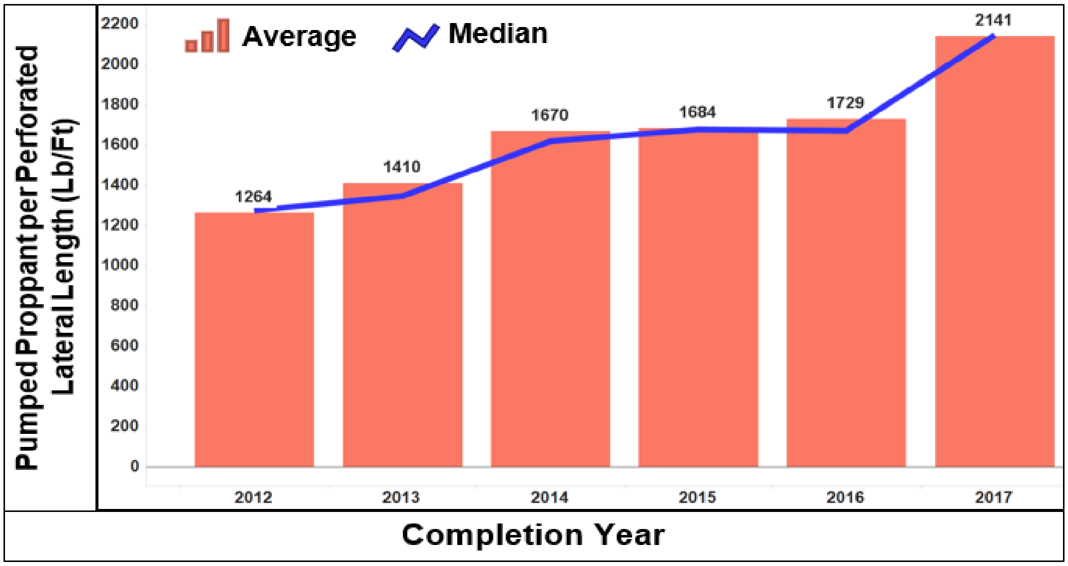


Figure 13. Combined Bar and Line Charts of Average and Median Pumped Proppant per Perforated Lateral Length per Well in the Marcellus

4.3. WATER UTILIZATION TRENDS IN THE MARCELLUS

To better understand the volume of pumped water per well in the Marcellus Shale Figure 14 and Figure 15 were prepared to show the application range of the volume of water pumped per well as well as the volume of pumped water per stimulated lateral length for each well, respectively. The box plots show the four quartiles ranges of water volume and the distribution charts illustrate the number of wells that utilize the same value range of pumped water volume. Figure 16 and Figure 17 illustrate the average and median values of the volume of water pumped per well over time and the normalized water pumped to the perforated lateral length (gallon/foot or gal/ft) per well over time, respectively. Over the years from 2012 to 2018, both the average and the mean values showed a successive increase in the volume of pumped water per well and the volume of water pumped per stimulated foot. Figure 18 compares the average use of water and proppant per well between the wells in Pennsylvania and West Virginia. The Marcellus Shale wells in West Virginia consumed on average more water and proppant than the wells in Pennsylvania.

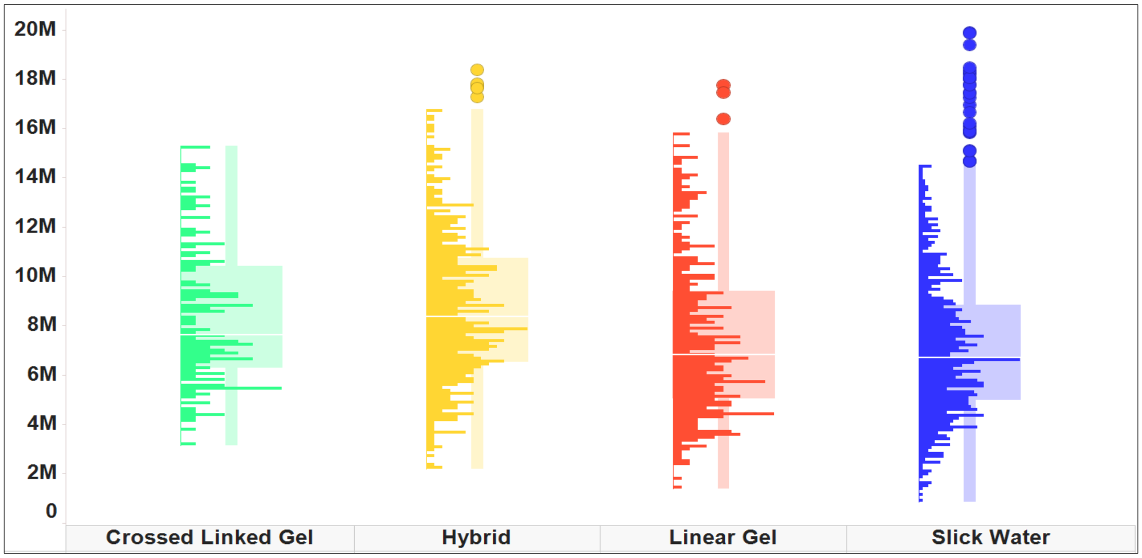


Figure 14. Box Plots and Distribution Charts of Water Pumped per Well for Each Stimulation Fluid Type in the Marcellus

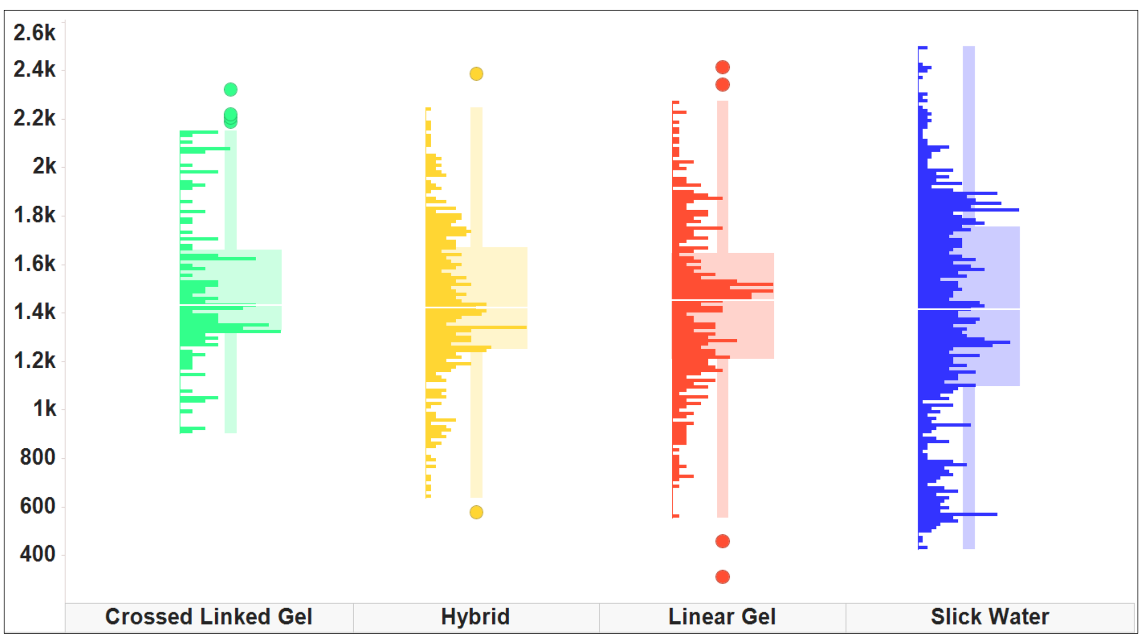


Figure 15. Box Plots and Distribution Charts of Water per Perforated Length for Each Stimulation Fluid Type in the Marcellus

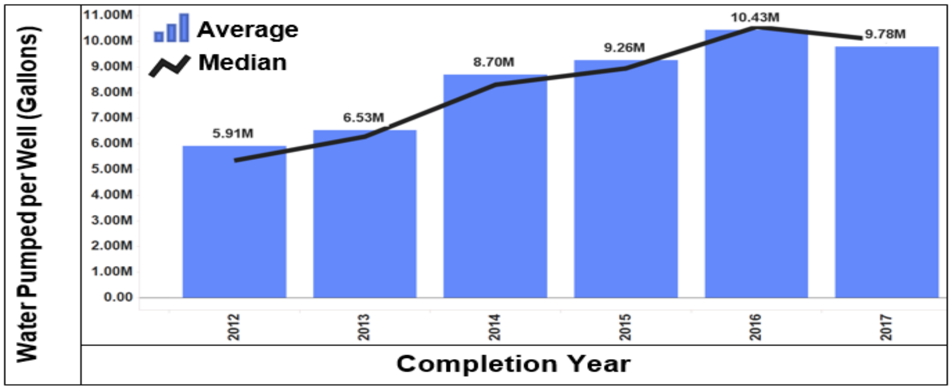


Figure 16. Combined Bar and Line Charts of Average and Median Water Pumped per Well in the Marcellus

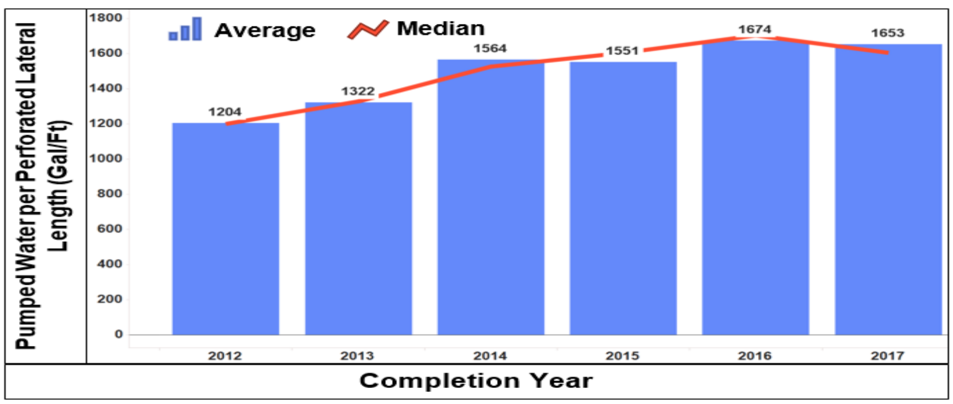


Figure 17. Combined Bar and Line Charts of Average and Median Pumped Water per Perforated Lateral Length per Well in the Marcellus

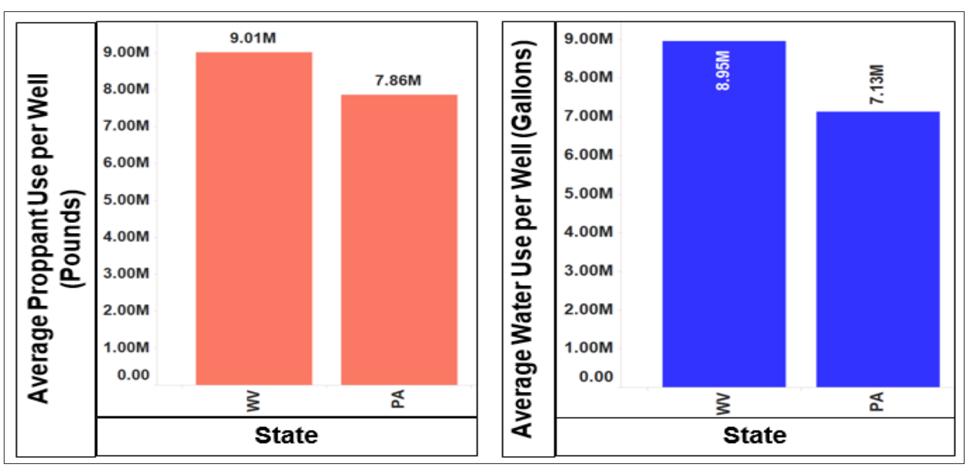


Figure 18. Water and Proppant Average Usage per the States in the Marcellus

4.4. PROPPANT LOADING TRENDS IN THE MARCELLUS

Proppant loading in pounds per gallon (lb/gal) is obtained by dividing the total proppant pumped mass by the total water volume pumped per well. Figure 19 shows the box plots and distribution bars of the proppant loading ranges for each stimulation fluid type. Figure 20 shows a combined bar and line charts of the average and median values of proppant loading in the Marcellus wells.

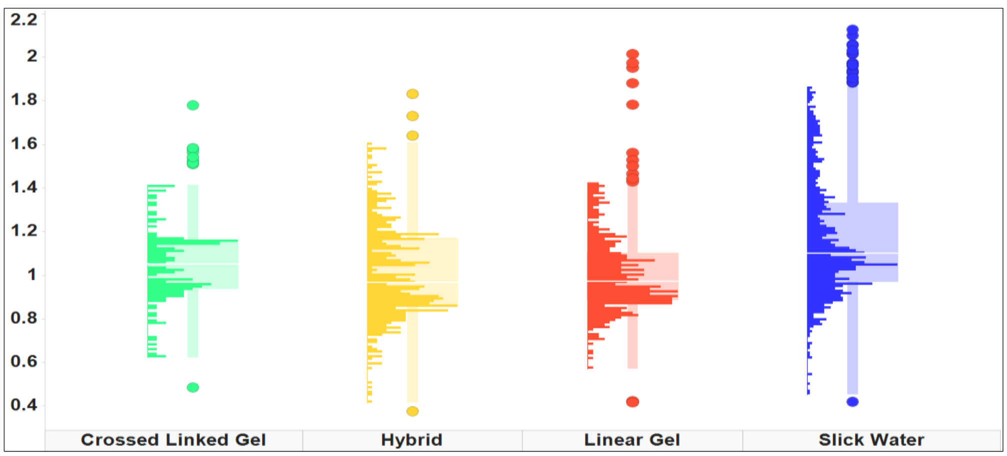


Figure 19. Box Plots and Distribution Charts of Proppant Loading for Each Stimulation Fluid Type in the Marcellus

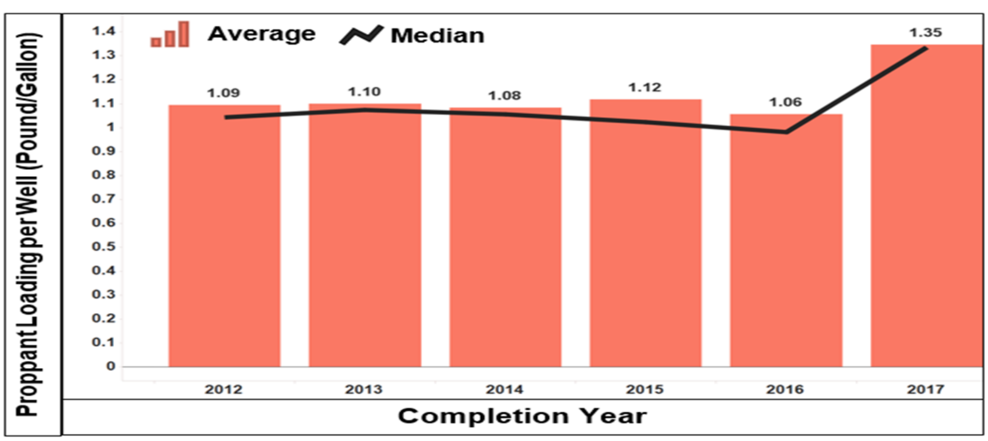


Figure 20. Combined Bar and Line Charts of Average and Median Proppant Loading per Well in the Marcellus

4.5. PRODUCTIVITY OF THE MARCELLUS SHALE WELLS

Marcellus Shale production is mainly gas but to account for any liquid hydrocarbon production associated with the gas production, all production volumes have been converted to barrel of oil equivalent (BOE) using the following conversion, where each 6 MCF of gas equal 1 BOE. Figure 21 and Figure 22 illustrate the range of production in the Marcellus Shale wells in box plot charts superimposed by distribution bars for the practical initial production (BOE/day). The practical initial production was calculated by dividing the cumulative production of the second month by the number of produced days. The practical initial production is considered more representative of initial production rate because it averages the entire month of early production and avoids starting late during the first month when the well is put on production. The figures also show the cumulative BOE production for the first year, first two years, and the five years cumulative production.

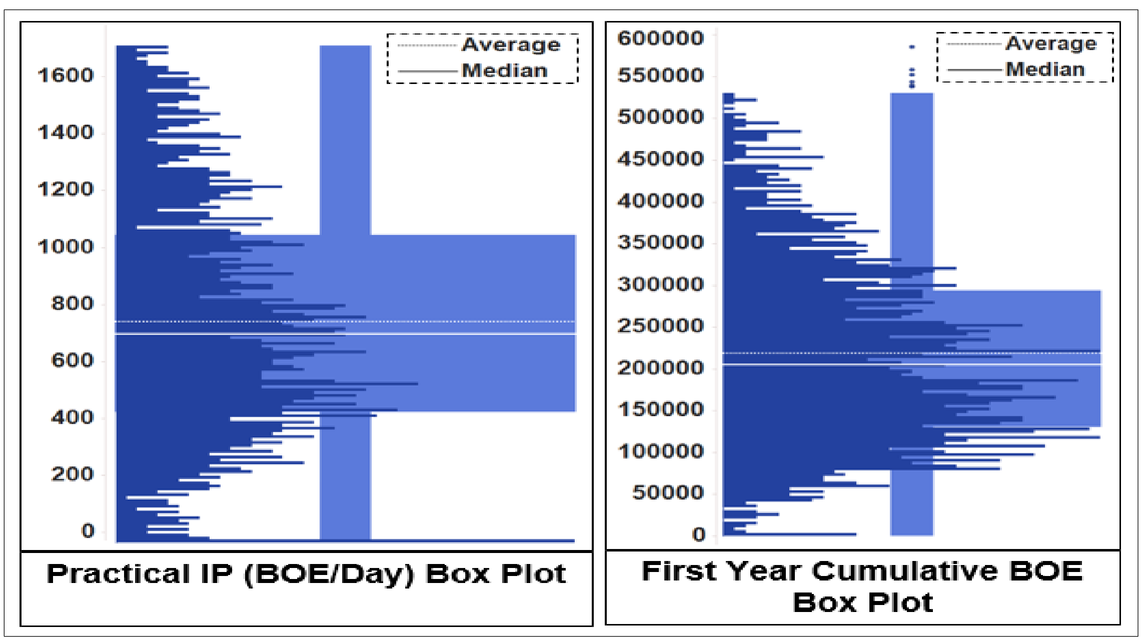


Figure 21. Box Plots and Distribution Charts of the Initial Production and First Year Cumulative Production for the Marcellus Shale Wells

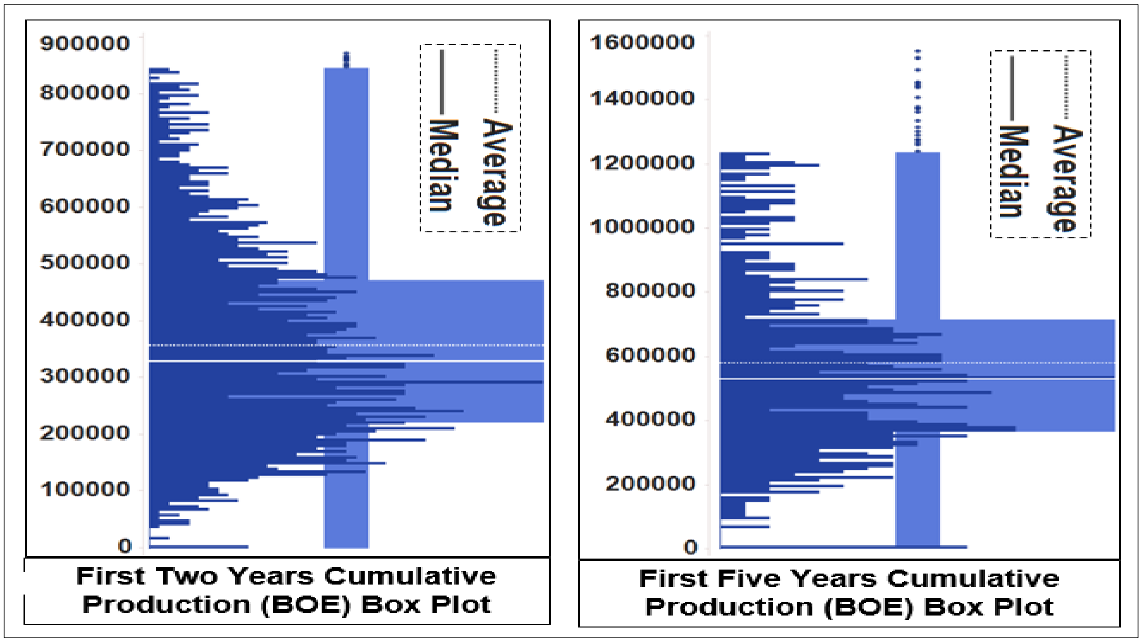


Figure 22. Box Plots and Distribution Charts of the First Two and Five Years Cumulative Production for the Marcellus Shale Wells

4.6. PRODUCTION RESPONSE TO THE AMPOUNT OF PROPPANT PER STIMULATED LATERAL LENGTH

The production performance of stimulated Marcellus wells was tested against the amount of pumped proppant per foot. The data of total proppant mass pumped per well were divided by the gross perforated lateral length (lb/ft) and grouped into five groups with increment of 500 (lb/ft). The average production performance was then plotted in a bar chart for practical initial production, first year cumulative BOE, two and five years cumulative BOE as illustrated in Figure 23, Figure 24, Figure 25, and Figure 26, respectively. The initial production response to the pumped proppant per foot showed a higher rate at the lower range of 0 – 1000 lb/ft. However, the long-term cumulative production consistently showed an improvement in average annual cumulative production with higher ranges of pumped water per foot up to 2000 gallon per foot. For the high range of gal/ft (2000-2500) the average cumulative production showed a slight drop in performance which indicates that the best range of water per foot in the Marcellus Shale was 1500 – 2000 gal/ft.

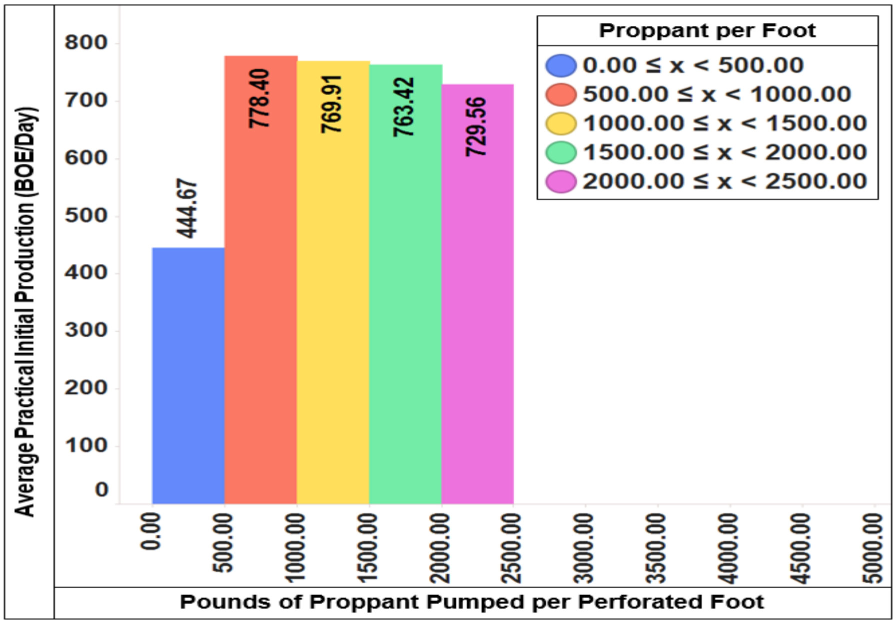


Figure 23. Proppant per Foot Effect on Initial Production in the Marcellus

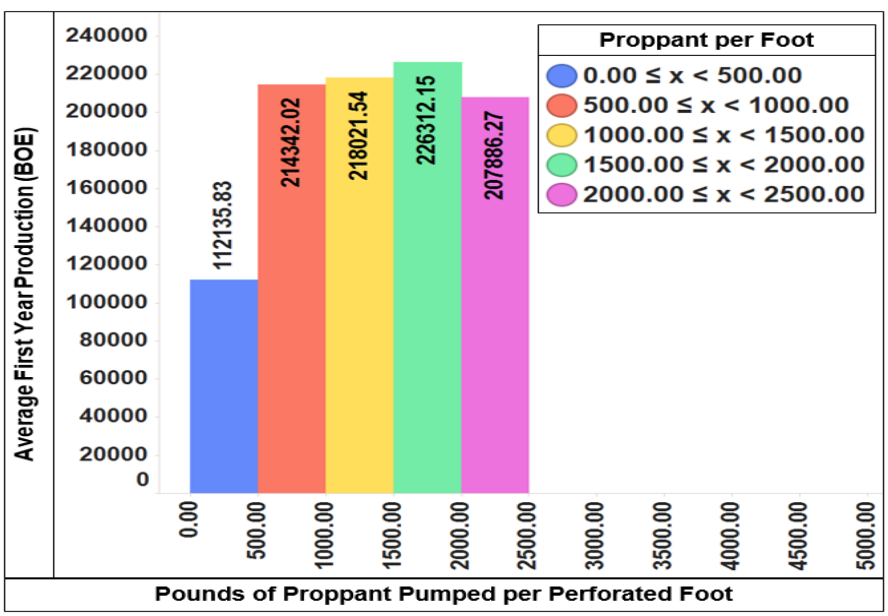


Figure 24. Proppant per Perforated Foot Effect on First Year Equivalent Production in the Marcellus

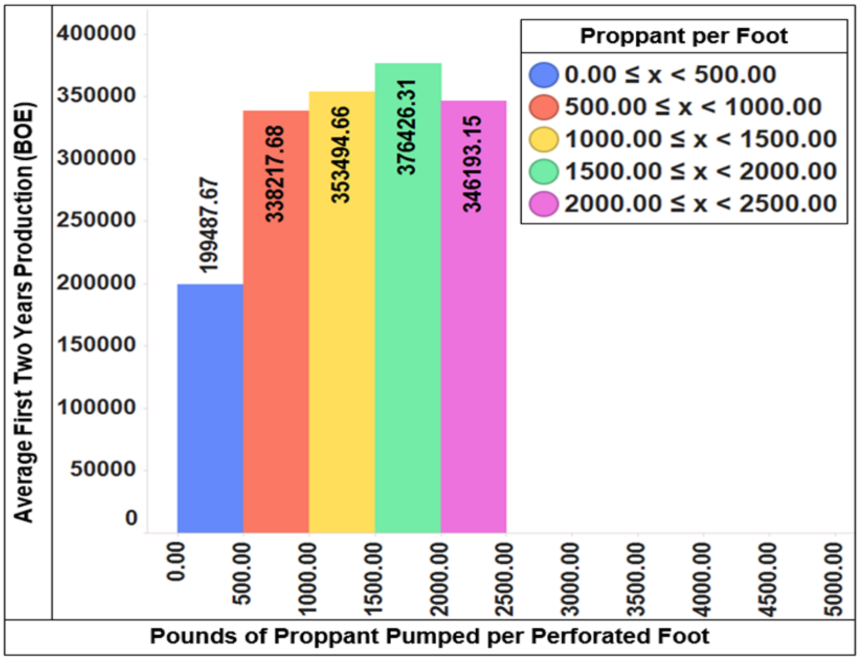


Figure 25. Proppant per Perforated Foot Effect on First Two Years Equivalent Production in the Marcellus

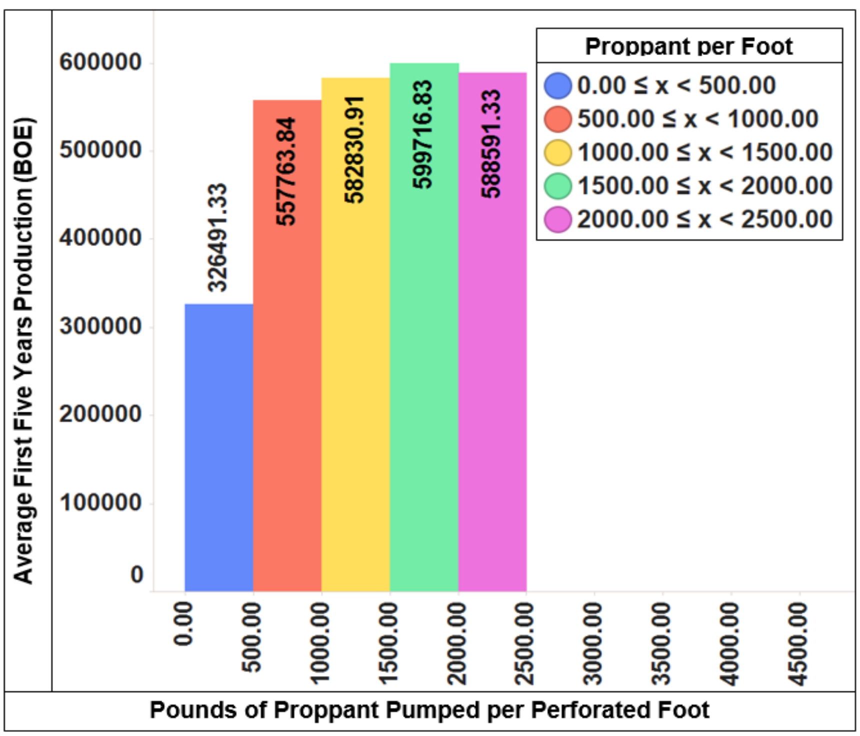


Figure 26. Proppant per Perforated Foot Effect on First Five Years Equivalent Production in the Marcellus

4.7. PRODUCTION RESPONSE TO THE VOLUME OF WATER PUMPERD PER STIMULATED LATERAL LENGTH

The production performance of the stimulated wells in the Marcellus Shale was also tested against the size of the stimulation fluid represented by the volume of pumped water per foot. The data of total water volume pumped per well were divided by the gross perforated lateral length (gal/ft) and grouped into five groups with increment of 500 (gal/ft). The average production performance was then plotted in a bar chart for practical initial production, first year cumulative BOE, two and five years cumulative BOE as illustrated in Figure 27, Figure 28, Figure 29, and Figure 30, respectively. The initial production response to the pumped water per foot showed a higher rate at the lower range of 0 – 1500 gal/ft. However, the long-term cumulative production consistently showed an improvement in average annual cumulative production with higher ranges of pumped water per foot up to 2000 gal/ft. For the high range of gal/ft (2000-2500) the average cumulative production showed a slight drop in performance which indicates that the best range of water per foot in the Marcellus Shale was 1500 – 2000 gal/ft.

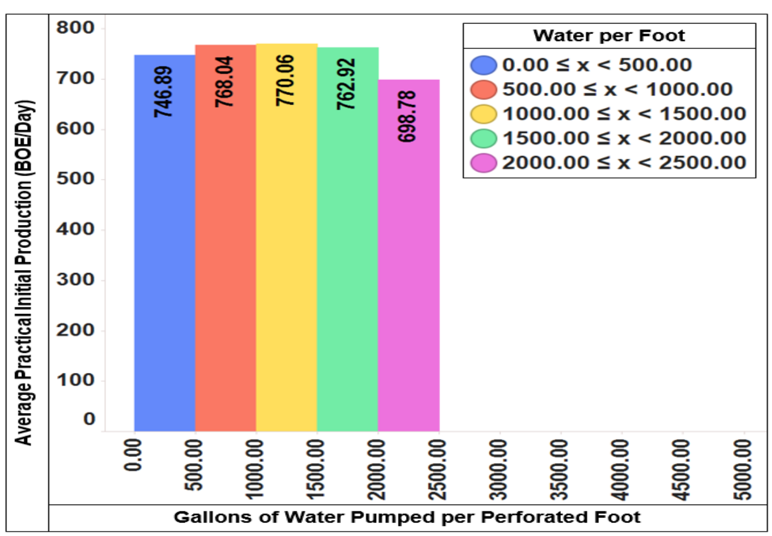


Figure 27. Water per Perforated Foot Effect on Initial Production in the Marcellus

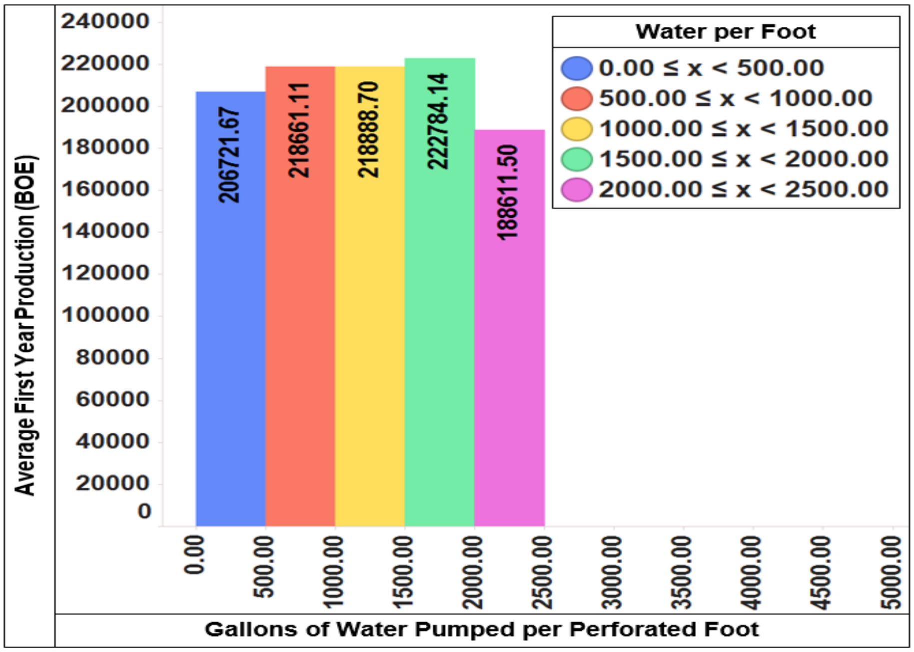


Figure 28. Water per Perforated Foot Effect on First Year Equivalent Production in the Marcellus

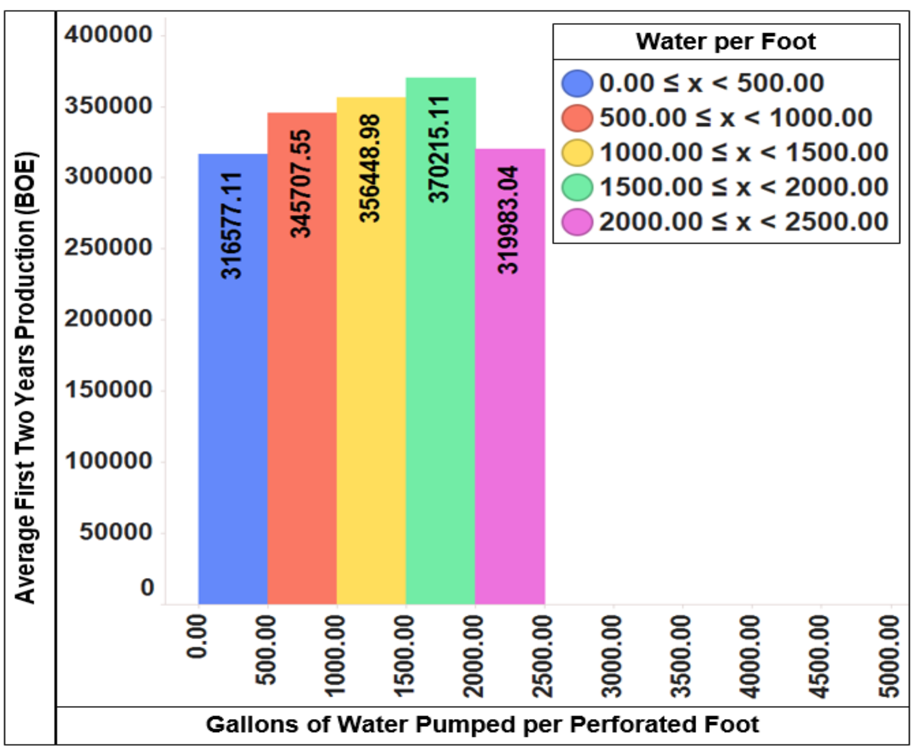


Figure 29. Water per Perforated Foot Effect on First Two Years Equivalent Production in the Marcellus

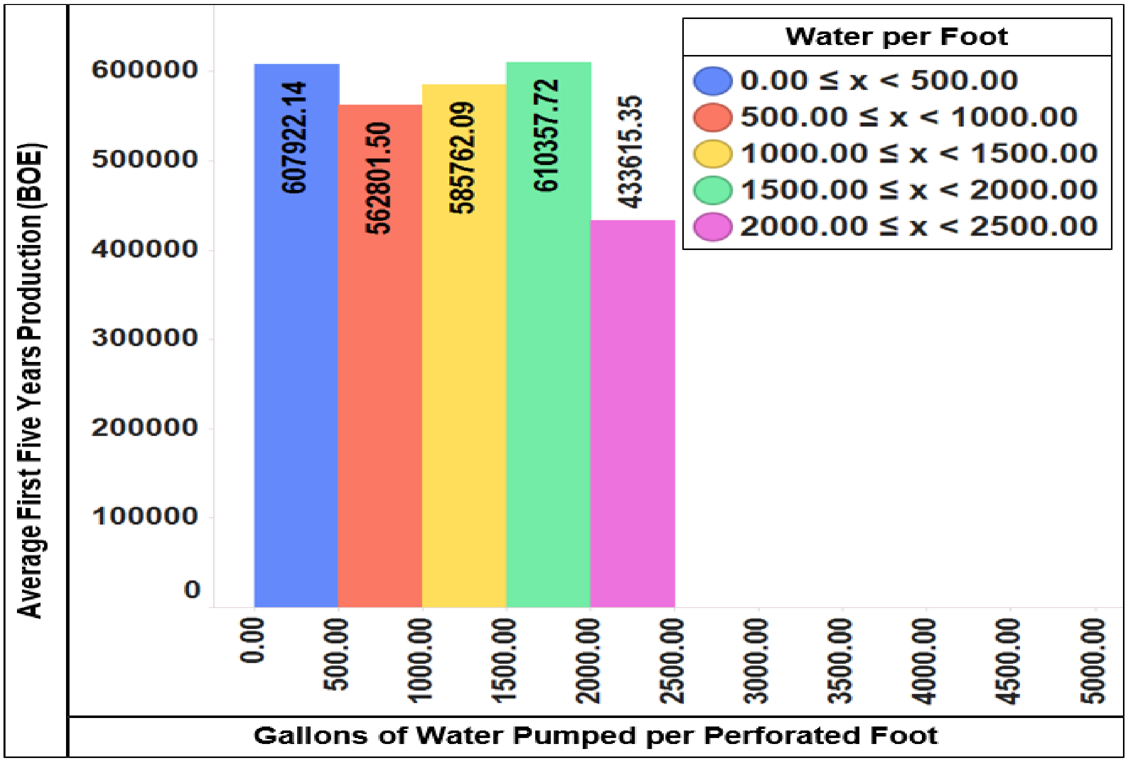


Figure 30. Water per Perforated Foot Effect on First Five Years Equivalent Production in the Marcellus

5. CONCLUSIONS

In this study, a comprehensive database for the Marcellus Shale play was created by extracting and processing fracturing fluids chemical ingredients from FracFocus and then integrated with production and completion data acquired from DrillingInfo. The database was used to create a descriptive analysis approach to gain insights and understand the trends of stimulation parameters of proppant mass, water volume, proppant to water ratio (proppant loading), horizontal lateral length, and the perforated interval length of the laterals. The production performance of the wells was also introduced in this study by comparing the size and the amount of pumped water and proppant per the stimulated lateral foot. The main conclusions of this study are summarized in the points below:

- The chemical registry of FracFocus 3.0 can be utilized as a practical source of chemical ingredients for most of the wells stimulated after 2012. The raw data from FracFocus need to go through rigorous cleaning and processing workflows to generated usable and insightful parameters to be used in descriptive and predictive analytical approaches.
- The hydraulic fracturing process and chemical ingredients subject matter expertise are paramount to successfully utilize FracFocus data and to present insights into the stimulation and completion trends in any unconventional shale play development.
- In the Marcellus Shale wells, more than 50% of the stimulated wells were stimulated with water fracturing fluids.
- About 90% of the horizontal lateral length was perforated and stimulated in the Marcellus wells.
- Studying the proppant and water quantities trends over time showed that every year the operators in the Marcellus have increased the average size of the fracturing fluids water base volume and the amount of the associated proppant.
- Testing the productivity of the wells indicated that the long term cumulative BOE production in the Marcellus Shale wells positively increased as the amount of pumped water and proppant per foot was increased. Pumping more than 2000 pound per foot of proppant or 2000 gallon per foot of water indicated a negative trend in term of cumulative production.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to DrillingInfo team and the co-founder and director of university outreach Mr. Mark Nibbelink for their generosity in providing access to the U.S. production and completion data.

REFERENCES

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar, A. M. "Review of Stimulation and Completion Activities in the United States Shale Plays". Paper ARMA 19-2003, presented at the 53rd US Rock Mechanics / Geomechanics Symposium held in New York City, New York, USA, 23-26 June 2019a

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar, A. M. "Production Performance Estimation from Stimulation and Completion Parameters Using Machine Learning Approach in the Marcellus Shale". Paper ARMA 19-2034, presented at the 53rd US Rock Mechanics / Geomechanics Symposium held in New York City, New York, USA, 23-26 June 2019b

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar, A. M. "Long-Term Comparison of Gel and Water Fracture Stimulation in Marcellus Shale Play". Paper SPE-195990-MS, Accepted to be presented at SPE Annual Technical Conference and Exhibition (ATCE) in Calgary, Alberta, Canada, 30 September-2 October 2019c

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar A. M., M. M. Alkhamis, Al-Bazzaz, W. H. "From Data Collection to Data Analytics: How to Successfully Extract Useful Information from Big Data in the Oil & Gas Industry?" Paper SPE-196428-MS, Accepted to be presented at SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition in Calgary, Bali, Indonesia, 29 - 31 Oct 2019d

Arthur, J. D., Layne, M. A., Hochheiser, H. W., & Arthur, R. (2014, February 4). Spatial and Statistical Analysis of Hydraulic Fracturing Activities in US Shale Plays and the Effectiveness of the FracFocus Chemical Disclosure System. Society of Petroleum Engineers. doi:10.2118/168640-MS

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Cedeno M. A., Al-Attar A. M., Al-Bazzaz, W. H. “Short and Long-Term Productivity Comparison of Hydraulic Fracturing Fluid Systems in East Texas Cotton Valley Formation” Paper SPE-196515-MS, Accepted to be presented at SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition in Calgary, Bali, Indonesia, 29 - 31 Oct 2019e.

Carman, P., Li, C., & Wheeler, R. (2018, September 24). Unorthodox Fluids for Unconventional Completions: Maximizing Production and Avoiding Pitfalls. Society of Petroleum Engineers. doi:10.2118/191751-MS

DrillingInfo website. Retrieved on March/01/2019, from https://info.drillinginfo.com/

EIA, 2017. Appalachia Region Drives Growth in U.S. Natural Gas Production since 2012. Retrieved on August/01/2019, from https://www.eia.gov/todayinenergy/detail.php?id=33972

EIA, 2018. U.S. Oil and natural Gas Wells by Production Rate Retrieved on August/20/2019, from https://www.eia.gov/petroleum/wells/

Elsner, M., & Hoelzer, K. (2016). Quantitative Survey and Structural Classification of Hydraulic Fracturing Chemicals Reported in Unconventional Gas Production. Environmental Science & Technology, 50(7), 3290-3314. doi:10.1021/acs.est.5b02818

FracFocus website. Retrieved on May/01/2019, from http://fracfocus.org/

Harper, John A. (2008). The Marcellus Shale - An Old "New" Gas Reservoir in Pennsylvania. Pennsylvania Geology, Volume 38, Number 1. Pennsylvania Bureau of Topographic and Geologic Survey

Khvostichenko, D., & Makarychev-Mikhailov, S. (2018, August 9). Oilfield Data Analytics: Linking Fracturing Chemistry and Well Productivity. Unconventional Resources Technology Conference. doi:10.15530/URTEC-2018-2903086

King, G. E. (2010, January 1). Thirty Years of Gas Shale Fracturing: What Have We Learned? Society of Petroleum Engineers. doi:10.2118/133456-MS

Luo, H., & Zhang, Y. (2018, October 5). Mining FracFocus and Production Data for Efficacy of Fracturing Fluid Formulations. Society of Petroleum Engineers. doi:10.2118/191787-18ERM-MS

Mohaghegh, S. D., Gaskari, R., & Maysami, M. (2017, January 24). Shale Analytics: Making Production and Operational Decisions Based on Facts: A Case Study in Marcellus Shale. Society of Petroleum Engineers. doi:10.2118/184822-MS

Myers, R. R., Knobloch, T. S., Jacot, R. H., & Ayers, K. L. (2017, October 4). Production Analysis of Marcellus Shale Well Populations Using Public Data: What Can We Learn from The Past? Society of Petroleum Engineers. doi:10.2118/187505-MS

Oil & Gas Journal article entitled "DOE official highlights Appalachia's role in US energy renaissance" Retrieved on March/01/2019, Link https://www.ogj.com/general-interest/economics-markets/article/14038028/doe-official-highlights-appalachias-role-in-us-energy-renaissance

Olaoye, O., & Zakhour, N. (2018, September 24). Utilizing a Data Analytics Workflow to Optimize Completion Practices While Leveraging Public Data: A Permian Basin Case Study. Society of Petroleum Engineers. doi:10.2118/191416-MS

Romero, D. H., & Poston, S. W. (2016, May 10). Data Mining and Analysis - Eagle Ford Formation, South Texas. Society of Petroleum Engineers. doi:10.2118/179994-MS

Shahkarami, A., Ayers, K., Wang, G., & Ayers, A. (2018, October 5). Application of Machine Learning Algorithms for Optimizing Future Production in Marcellus Shale, Case Study of Southwestern Pennsylvania. Society of Petroleum Engineers. doi:10.2118/191827-18ERM-MS

Srinivasan, K., Ajisafe, F., Alimahomed, F., Panjaitan, M., Makarychev-Mikhailov, S., & Mackay, B. (2018, August 28). Is There Anything Called Too Much Proppant? Society of Petroleum Engineers. doi:10.2118/191800-MS

Weijers, L., Wright, C., Mayerhofer, M., Pearson, M., Griffin, L., & Weddle, P. (2019, January 29). Trends in the North American Frac Industry: Invention through the Shale Revolution. Society of Petroleum Engineers. doi:10.2118/194345-MS

Yang, Y., Robart, C. J., & Ruegamer, M. (2013, February 4). Analysis of U.S. Hydraulic Fracturing Design Trends. Society of Petroleum Engineers. doi:10.2118/163875-MS

VII. PRODUCTION PERFORMANCE EVALUATION FROM STIMULATION AND COMPLETION PARAMETERS IN THE PERMIAN BASIN: DATA MINING APPROACH

Mustafa A. Al-Alwani, Shari Dunn-Norman, Larry K Britt, Husam H. Alkinani, Abo Taleb T. Al-Hameedi, Atheer M. Al-Attar, Hector A. Trevino, Waleed Al-Bazzaz

Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA

ABSTRACT

Over the last decade, there have been numerous advancements in horizontal drilling applications and in combination with hydraulic fracturing there has been a plethora of growth in producing unconventional resources. Scouring the literature shows a close correlation between the oil and gas prices and the operators' willingness to pump more or less proppant in their wells. In correspondence to the change in trends, the operators are constantly assessing the optimization of completions design to find the optimum stimulation and completion applications. This paper utilizes 3,782 unconventional horizontal wells to analyze the impact of proppant volume and the length of the perforated lateral on short and long-term well productivity across the Permian (Midland) Basin.

The raw stimulation data were collected from FracFocus website. Rigorous data management techniques were utilized to build a comprehensive dataset with all the volumes of proppant, water, perforated lateral length, and all other stimulation chemicals. Quality control and data validations were applied to the dataset and all outliers were removed. In this study, the amount of the proppant pumped is compared to the cumulative production of equivalent barrel of oil (BOE) for the first year, 2 years, and five years. The

effect of normalized proppant per perforated foot on the cumulative production is also investigated.

Tying cumulative production to completion and stimulation practices, showed that increasing the pumped proppant per well from 5 million pounds to less than 10 million pounds, yielded a 34% increase in the five years cumulative average BOE. Stepping up the pumped proppant per well to 10-15 Million pounds and 15-20 Million pounds increased the 5 years BOE cumulative from the previous proppant range group to 27% and 18.5%, respectively. The cumulative production per foot versus the amount of pumped proppant, lateral length versus production, cumulative production per foot versus lateral length ranges, and finally the proppant per foot versus cumulative production per foot are all discussed in this study.

This study represents a descriptive analysis approach to investigate the impact of different stimulation and completion designs on short and long-term cumulative production.

1. INTRODUCTION

Unconventional resource development has become one of the primary sources of both oil and gas production in the United States. This was made possible by both advancing horizontal drilling technologies and the capability of economically placing hydraulic fracturing stimulations along horizontal wells. As drilling and stimulation of unconventional resources have drastically increased, the data obtained from these actions have also increased. Data gathered from these processes include drilling, completion, stimulation, and production parameters. Analyzing the immense amount of data supports the advancement of well productivity and can help in identifying economic exploitation methods essential to drilling and completion on future wells.

Hydraulic fracturing consists of fluids, composed of a base fluid, proppant, and a variety of chemicals, being injected under the influence of high pressures into the wellbore. When the fluid is pumped into the well, fractures are created providing pathways for oil and gas to flow into the well and finally be pumped to the surface. This process consists of four stages; the spearhead or acid stage, the pad stage, the proppant or slurry stage, and the flush stage. The spearhead stage is completed by pumping water mixed with diluted acid. This mixture cleans the perforations and areas around the wellbore containing debris. The pad stage requires a mixture of water and chemicals being pumped under specific pressures to initiate fractures in the formation. The proppant stage (slurry) consists of a proppant and fracturing fluid mixture being pumped into the well to prop and keep the fractures created in the pad stage open. The final stage, the flush stage, involves a clean base fluid being pumped into the wellbore to clean out any leftover proppant in the wellbore.

Industry focus has shifted towards unconventional resources in the United States over the past few years. The shift in focus has resulted in implementation of different drilling and completion strategies. By consistently challenging design and operating boundaries, operators and service providers are quickly finding the best methods to improve well productivity.

To analyze the industry’s tendencies, it is essential to collect drilling, completion, and stimulation parameters data. Based on said collected data, it is evident that strategic evolution has been occurring over the years. The target of these strategies is to enhance well productivity and reduce the footprint. This can clearly be seen by the adoption of multiple well pads instead of single well pads. Advancements in drilling technology has allowed for longer horizontal laterals which calls for more stimulated stages. More proppant and water are required so that the reservoir is stimulated effectively. Historically, operators in the unconventional resources have started designs with basic engineering procedures in unison with establishing baseline treatments. This occurred during the early development stage of drilling and completion plans.

Recently, one of the quickest growing disciplines within the oil and gas industry is data analytics. By applying data descriptive and predictive methods, many industries can optimize their operations as well as better their decision making and future planning (Khvostichenko and Makarychev-Mikhailov, 2018). The vital importance of data-driven decision making has recently started being appreciated by many operator and service companies. Because of this, these companies have begun to allocate many of their resources to develop data use to drive their everyday decisions.

There is a growing interest in data analysis studies and the literature contains a copious number of articles that illustrate this interest. Early studies seem to have a lack of formation descriptive data. Without sufficient data, it is difficult to extrapolate trends to the field level. Early data studies were normally executed on a limited amount of localized well data and only described trends from very small areas (Carman and Wheeler, 2018). With the addition of a published study which included comprehensive reviews of trends for shale plays best practices, King (2010) was able to add sustenance to the literature. This study has since then remained a very valuable reference. Historical case studies are very helpful in seeing the industry's progression towards ideal completion strategies. Those publications have proven the power of data analyses in comprehending trends and to deduce specific insights (Yang et al., 2013; Arthur et al., 2014; Romero and Poston, 2016; Myers et al., 2017; Luo and Zhang, 2018; Olaoye and Zakhour, 2018; Srinivasan et al., 2018; Carman and Wheeler, 2018; Weijers et al., 2019).

Performing data science projects are very challenging in the oil and gas industry. One of the biggest, if not the biggest, challenges is finding available comprehensive data. Drilling, completion, and stimulation data has for a long period of time been difficult to come by. The lack of data was caused by the data being available only to service companies and operators and data sharing between the two and the public was limited. To make data analyses successful, a wide range of available data is essential. Any missing data will cause hindrances in any study and will result in decisions that may not be entirely correct. To fully execute comprehensive data analyses, a readily available and complete database that is a combination of both operator and service provider data is needed. When some data is not made public, researchers would run analyses on incomplete data sets. The lack of data when completing data analytics may render inefficient analyses. Another challenge that is presented in data analytics is the quality of data and how the data is formatted when being used for analyses. Examples include the accuracy of the data and how the data is saved. Datasets can be saved as scanned paper reports, unchangeable electronic files, or formats that make it easy to transform and use the data. The last obstacle for large scale data analytics is finding large scale data processing tools and software. Microsoft Excel, one major data processing tool, is very frequently utilized. Even so, a single spreadsheet can only contain 1,048,576 rows of data which may not be large enough for certain datasets. This limitation could possibly hinder analysts from properly analyzing large databases. For example, the data gathered from FracFocus contains several million rows of stimulation chemical data and Microsoft Excel is not able to handle datasets that large. To perform such studies, more advanced processing software is needed.

In the current days, with more data are being shared and more accessible databases are becoming available, studying the historical cases and presenting case studies of the trends over the recent period of time will serve as a useful and insightful tools to track the progress that the industry is making towards the optimum strategies to maximize productivity of the wells (Al-Alwani et al., 2019a; Al-Alwani et al., 2019b; Al-Alwani et al., 2019c; Al-Alwani et al., 2019d; Al-Alwani et al., 2019e).

For this study, the Permian (Midland) Basin wells' stimulation chemical data were downloaded from FracFocus for all the horizontal wells that were completed and stimulated between 2012 and 2018. The data then subjected to rigorous cleaning and processing and then combined with DrillingInfo completion and production parameters. Combining this data provides ample parameters for stimulation, completion and production data. The objective of this study is to investigate the production performance of the Permian Basin (Midland) wells as a result to different ranges of stimulation and completion parameters.

2. DATA AND METHODS

Data for this study were collected from FracFocus as well as DrillingInfo. A comprehensive database was built by extracting the chemical ingredients data for every Permian (Midland) Basin well stimulated and reported in FracFocus. Chemical volumes, trades names, and use were obtained, along with all chemical mass percentages. FracFocus data were processed then combined with completion and production data for the Permian (Midland) Basin wells obtained from DrillingInfo. The data from FracFocus are available to the public but require rigorous cleaning and processing for use.

The main objective in processing the data from FracFocus was to identify and categorize the stimulation fluid types based on the individual chemical ingredients that are reported in the list of treatment ingredients. It also helps in calculating the total mass of proppant pumper per each well.

After extracting the data from FracFocus a series of data processing and formatting was applied to combine all the data in one file to be ready for the next phase of data analysis. Chemical ingredients were reported by several identifiers such as the chemical ingredient name, commercial name, and the chemical abstract service number (CAS#). There is inconsistency in the way fracturing fluids and their ingredients are named and reported in FracFocus, making it difficult to use chemical names in the analysis. Many operators simply report the names based on their own standards. In addition, there are many chemical naming abbreviations and typos. For this study and for the purpose of classifying the chemical ingredients into several groups (e.g. propping agents, gelling agents, cross-linkers, surfactants), the CAS# was chosen to be the main identifier for each chemical ingredient. A data analysis technique, referred to as TreeMap charts, was applied to sort the chemical data into 19 prospective groupings based on their CAS# and the number of wells sharing the same chemical ingredient. Environmental protection agency (EPA) safer chemical ingredients list and a study conducted by Helmholtz Zentrum München, Institute of groundwater ecology (which provides a complete list of all hydraulic fracturing chemicals extracted from the Waxman list and the FracFocus database) were utilized in identifying the prospective chemical groups (Elsner & Hoelzer, 2016; EPA, 2019).

After the 19 chemical groups were generated, the groups were individually processed to eliminate any potential duplication in the reporting process from FracFocus. Once duplication issues were resolved, re-fractured wells were identified in the data processing workflow and grouped separately to eliminate aggregating several frac-jobs' ingredients into one well rather than separately assigning the appropriate number of fracturing jobs per well. Once all the ingredients mass percentage for the specific chemical group were summed up and assigned for each well, the following step was to combine all the chemical groups together into one dataset representing the processed FracFocus. Following that, multiple new parameters were calculated based on the processed data such as the calculated total proppant mass and the ingredients' concentrations. Figure 1 shows a workflow that combines all the processed chemical groups obtained from FracFocus. The final database was created by matching the production and completion data from DrillingInfo with the chemical components for each well. This combined database was quality checked. Several normalized parameters were also introduced to the database to help in the analysis phase and to generate production, completion, and stimulation insights for the Permian (Midland) Basin stimulated wells. Al-Alwani et al. (2019c) presents a more detailed discussion of the procedures used to create the final database. The database representation of the major counties in the Permian (Midland) Basin with the number of wells in each county is illustrated in Figure 2. A map of Texas with all the wells that are presented in this study is shown in Figure 3.

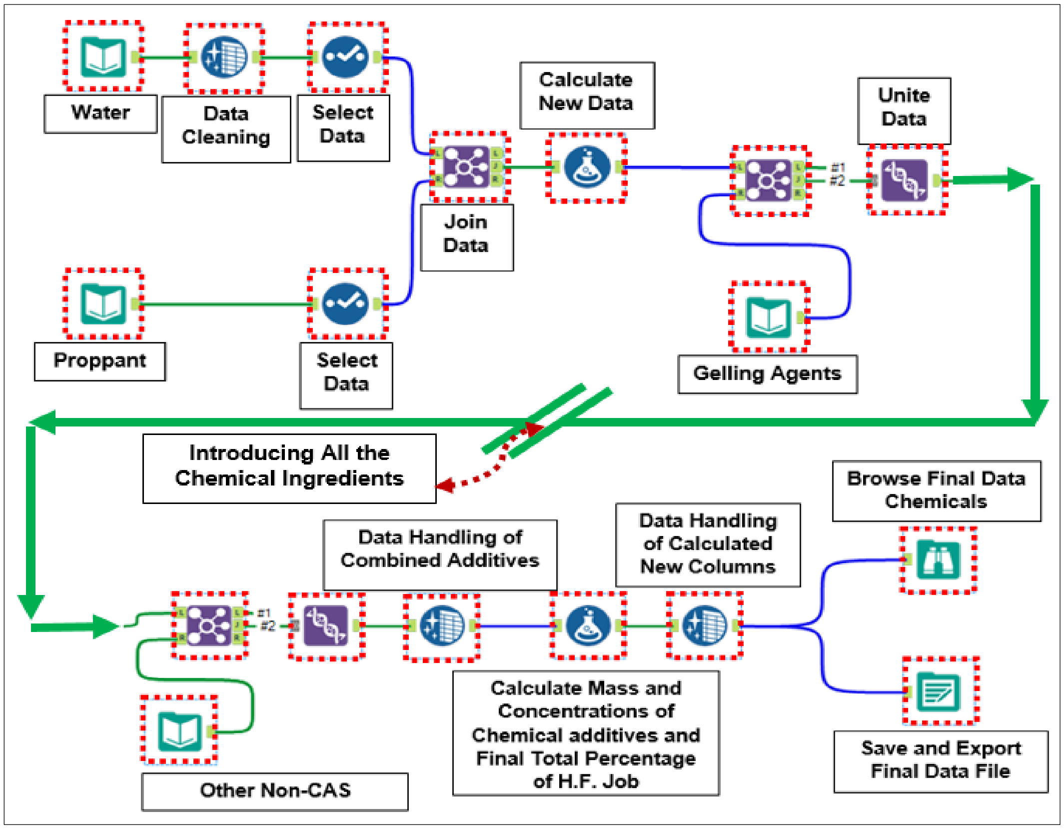


Figure 1. Data Processing Workflow to Combine All of the FracFocus Processed Data into One Database File (Al-Alwani et al., 2019c)

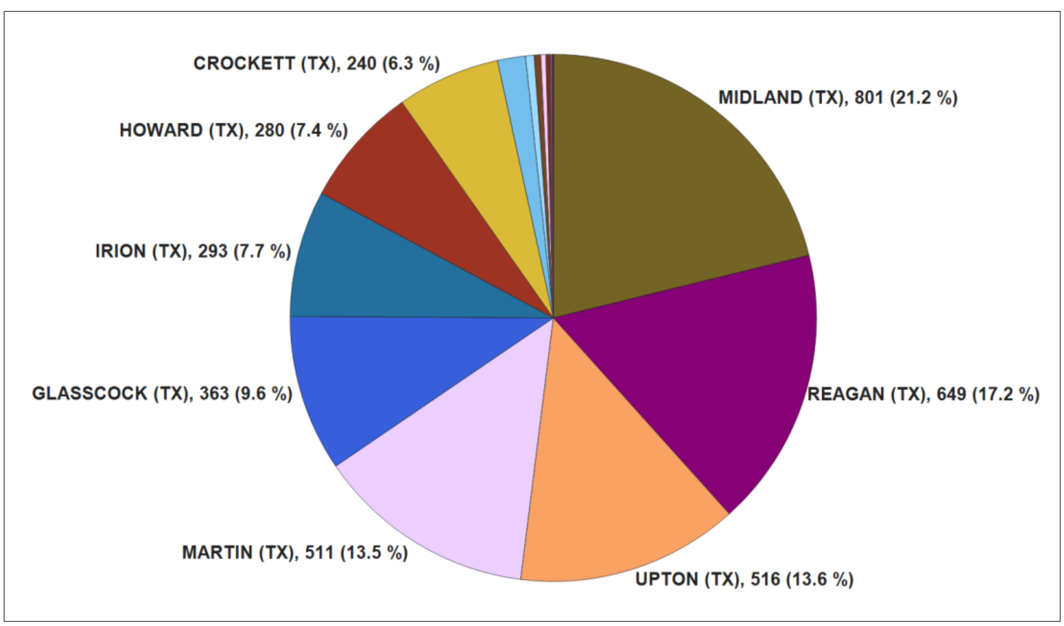


Figure 2. Pie Chart Represents the Number of Wells in the Major Counties of Midland Basin that were Utilized in this Study

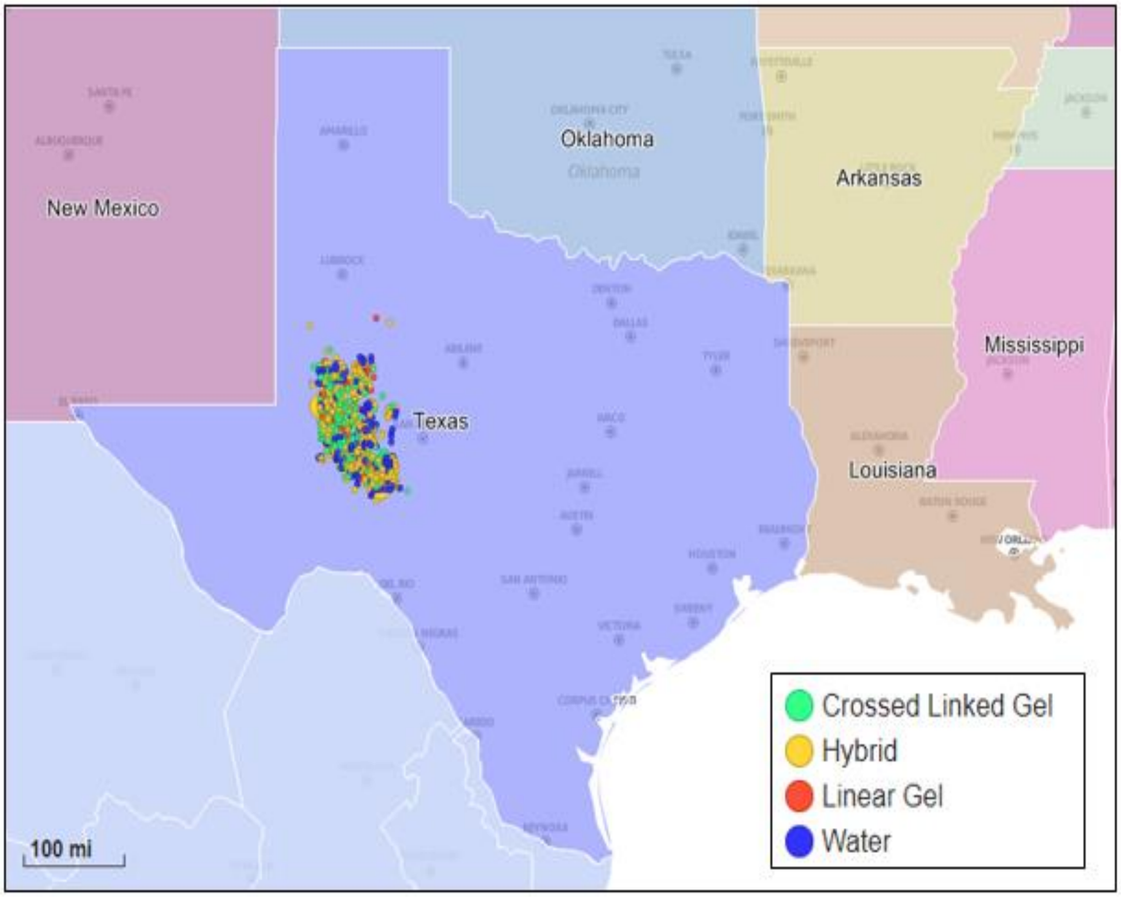


Figure 3. Map of Texas (Permian- Midland Basin Area) showing the Wells Distribution Color Coded based on the Type of Treatment Fluids

3. RESULTS AND DISCUSSION

This section will discuss the production performance for all the stimulated unconventional horizontal wells in the Permian Basin (Midland) To substitute for any quantities of produced gas, all the production data of oil and gas have been converted to equivalent barrel of oil (BOE) by using the conversion factor of 1 BOE = 6 MCF. The amount of proppant being pumped and the length of the perforated lateral length have been selected to represent the stimulation size and the completion magnitude, respectively.

3.1. AVERAGE PRODUCTION PERFORMANCE IN RESPONSE TO DIFFERENT RANGES OF TOTAL PROPPANT PUMPED PER WELL

Figure 4 illustrates the short and long-term production of the Permian (Midland) unconventional wells for four different ranges of total proppant pumped per well. The first group includes the wells that are stimulated with an average total pumped proppant mass of less than 5 million pounds per well, the second through fourth groups include the wells that have been stimulated with an average total proppant mass per well of 5 to 10 Million, 10 to 15 Million, and 15 to 20 Million pounds, respectively. The figure shows that average cumulative production in BOE per well versus time in years for each proppant range group. It is evident that increasing the amount of the proppant being pumped per well from 5 million to 10 million have positively increase the productivity of the well. The same was true when the amount of proppant range has increased from 10 million to 15 million pounds, the productivity of the wells in that group has increased as a result to the larger stimulation volume of pumped proppant. Upscaling the amount of proppant to more than 15 million pounds and up to 20 million pounds per well did not improve the productivity especially during the first three years of production compared to the 10-15 million pounds group of wells. However, the five years productivity have improved compared to 10-15 million pounds wells' group which indicates that the wells continued to produce at a higher rate.

Figure 5 demonstrate the percentage of average productivity change per well (BOE) when upscaling from one pumped proppant group to the following higher proppant range. The blue line represents the productivity per well (BOE) change as a result to increasing the proppant pumped from less than 5 million pounds wells' group to less than 10 million pounds wells' group.

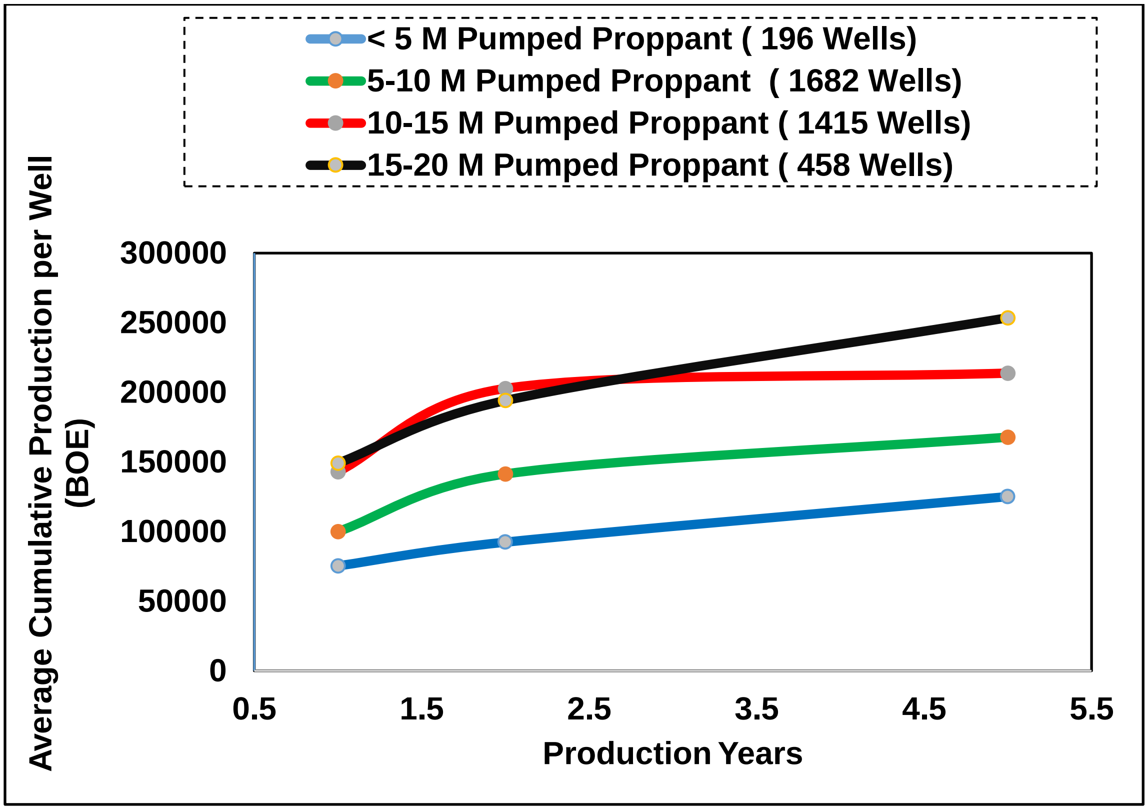


Figure 4. Cumulative Production versus Production Years for Different Groups of Pumped Proppant

The productivity has increased with reference to the productivity of wells stimulated with proppant range of < 5 million pounds per well by +32%, +52%, and +34% for the first, second, and the fifth year's average cumulative production per well (BOE), respectively. Elevating the proppant wells' group from the range of (5-10) million pounds per well to the range of (10-15) million pounds per well (green line) improved the average cumulative well's productivity over time (BOE) with reference to the productivity of wells' group stimulated with proppant range of 5-10 million pounds per well by +43%, +43.4%, and +27% for the first, second, and the fifth year's average cumulative production per well (BOE), respectively. Further elevating the proppant ranges from (10-15) million pounds per well to (15-20) million pounds per well (red line) changed the cumulative well's productivity with reference to the productivity of wells' group stimulated with proppant range of 10-15 million pounds by +4%, -4.2%, and +18.5% for the first, second, and the fifth year's average cumulative production per well (BOE), respectively. This change in productivity need to be evaluated against the cost of the stimulation and completion versus the short- and long-term productivity improvement.

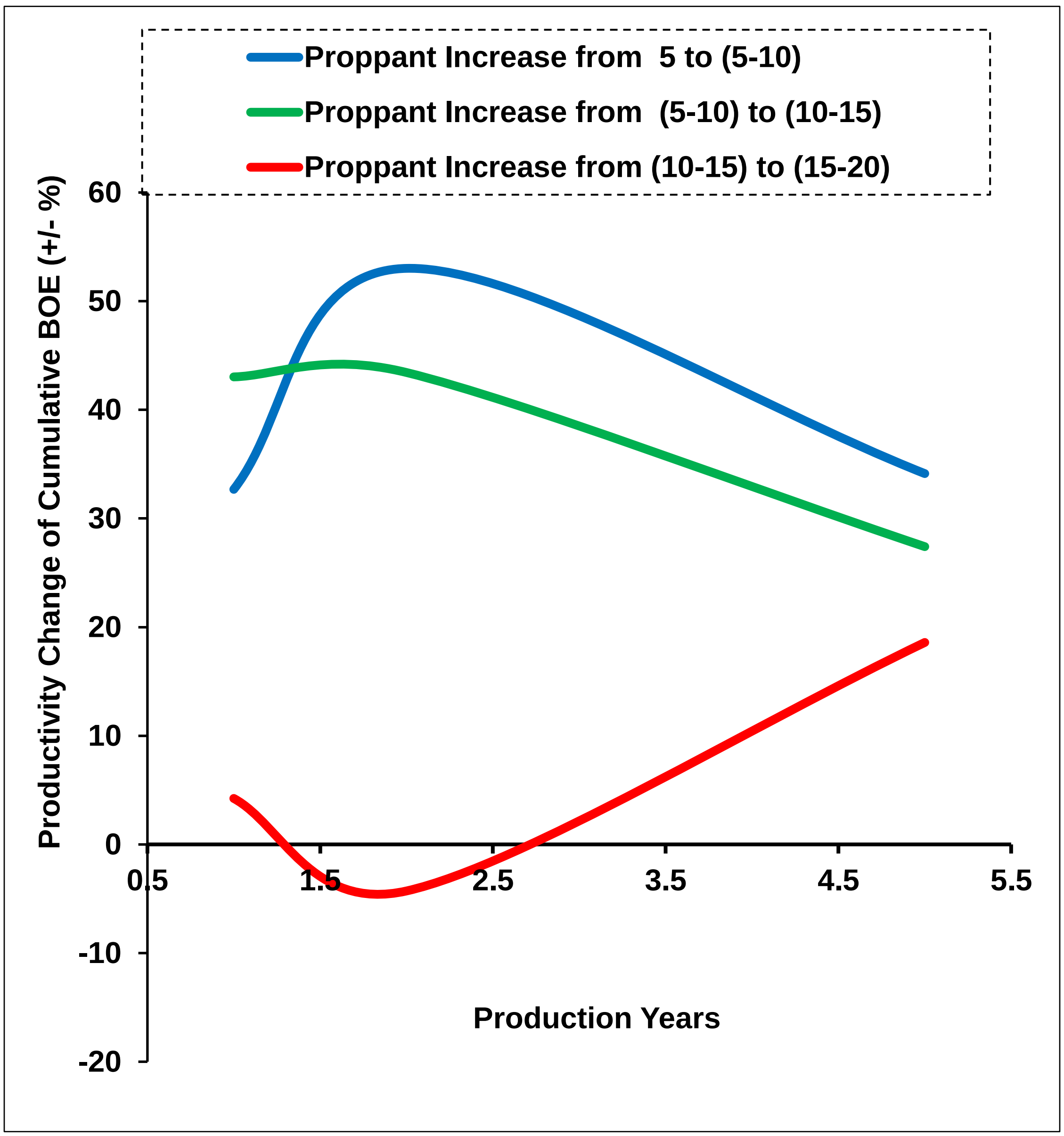


Figure 5. Productivity Change over Time in Response to Increasing Proppant

3.2. AVERAGE PRODUCTION PER FOOT (BOE/FT) PERFORMANCE IN RESPONSE TO DIFFERENT RANGES OF TOTAL PROPPANT PUMPED PER WELL

In order to account for the differences in lateral length among the groups of wells, Figure 6 has been constructed to illustrate the short and long-term production per each stimulated foot of lateral length for the wells hydraulically fractured in the Permian (Midland) Basin. The productivity per foot is calculated for four different ranges of total proppant pumped per well (<5, 5-10, 10-15, and 15-20 million pounds per well). The figure shows that average cumulative production in BOE/Ft per well versus time in years for each proppant range wells' group. It is evident that increasing the amount of the proppant being pumped per well from 5 million to 10 million have positively increase the productivity of each stimulated foot of the well. The same was true when the amount of proppant range has increased from 10 million to 15 million pounds, the productivity per stimulate foot of the wells in that group has increased as a result to the larger stimulation volume of pumped proppant. Upscaling the amount of proppant to more than 15 million pounds and up to 20 million pounds per well has negatively reduced the productivity per stimulated foot of the lateral length, the performance per foot has dropped to almost match the 5-10 million pounds group of wells but for the long-term productivity (5 years cumulative production) it showed a slight improvement than the wells stimulated with 10-15 million pounds of proppants.

Figure 7 demonstrate the percentage change of the average productivity of each stimulated foot per well (BOE/Ft) when upscaling from one pumped proppant group to the following higher proppant range. The blue line represents the productivity per well (BOE/Ft) change as a result to increasing the proppant pumped from less than 5 million pounds wells' group to less than 10 million pounds wells' group. The productivity/foot has increased with reference to the productivity/foot of wells stimulated with proppant range of < 5 million pounds per well by +26%, +32%, and +3.6% for the first, second, and the fifth year's average cumulative production/foot per well (BOE/Ft), respectively. Elevating the proppant wells' group from the range of (5-10) million pounds per well to the range of (10-15) million pounds per well (green line) improved the cumulative well's productivity/foot over time (BOE/Ft) with reference to the productivity/foot of wells' group stimulated with proppant range of 5-10 million pounds per well by +17.4%, +16.5%, and +6.7% for the first, second, and the fifth year's average cumulative production/foot per well (BOE/Ft), respectively. Further elevating the proppant ranges from (10-15) million pounds per well to (15-20) million pounds per well (red line) changed the cumulative well's productivity/foot with reference to the productivity/foot of wells' group stimulated with proppant range of 10-15 million pounds by -8.8%, -13.2%, and +3.5% for the first, second, and the fifth year's average cumulative production/foot per well (BOE/Ft), respectively.

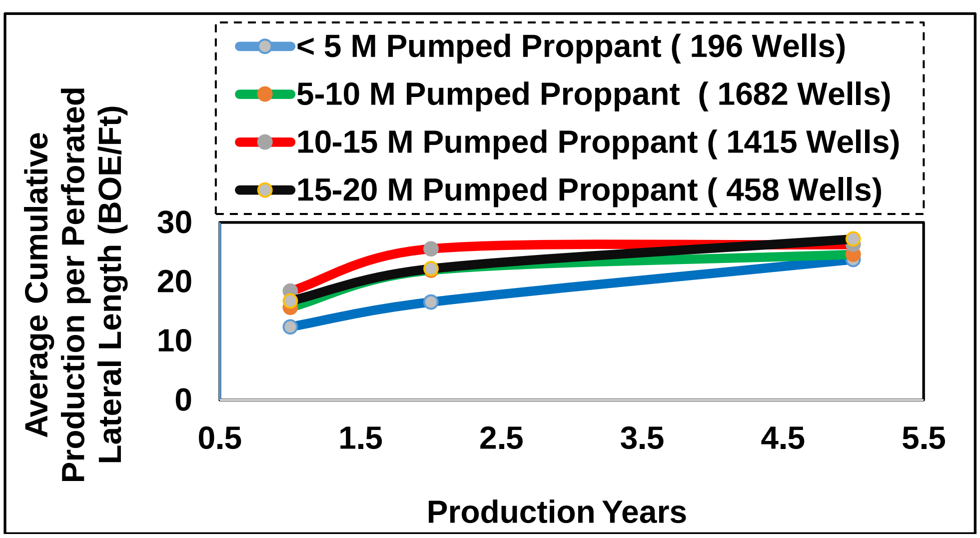


Figure 6. Cumulative Production per Foot versus Production Years for Different Groups of Pumped Proppant

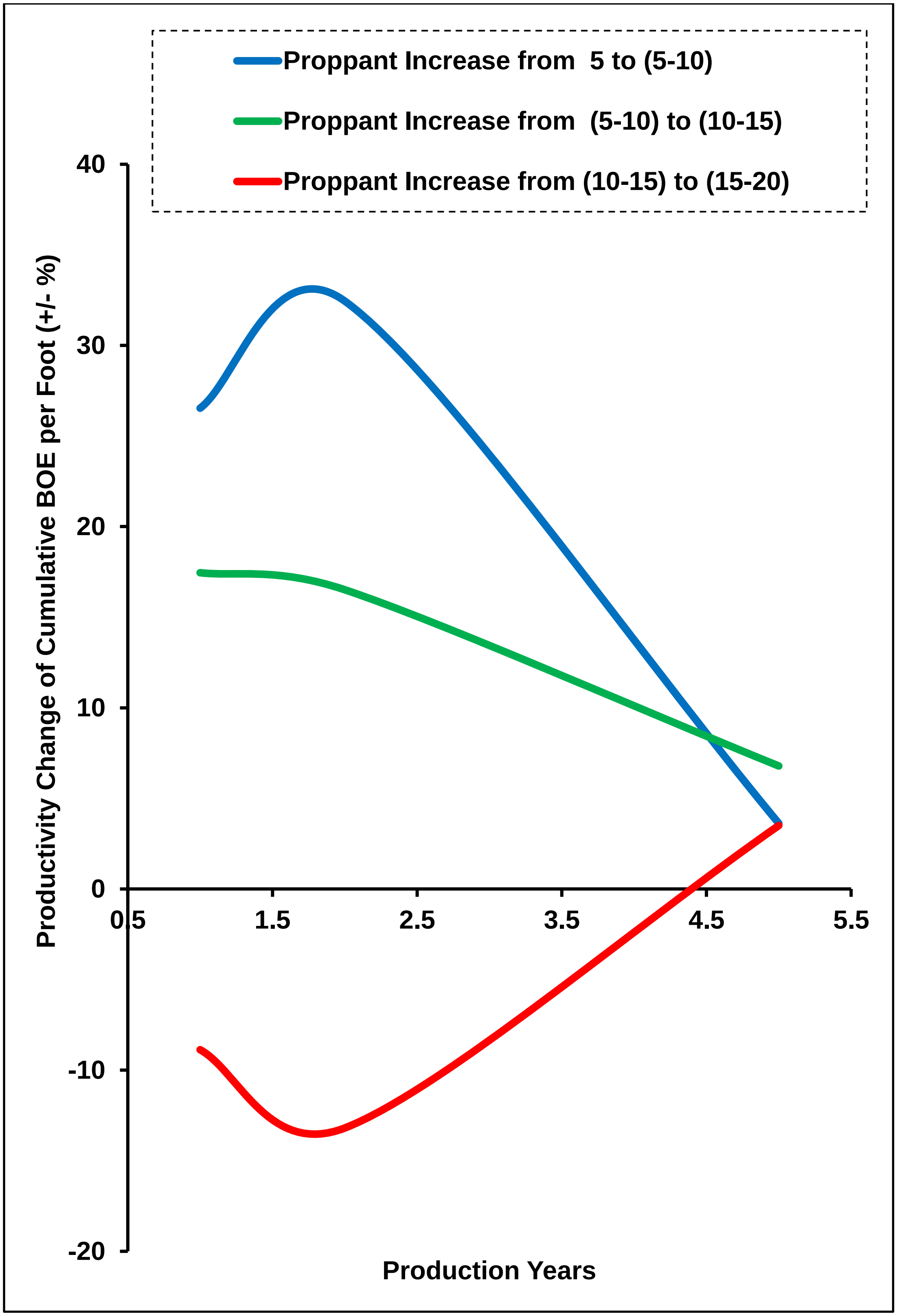


Figure 7. Productivity per Foot Change over Time in Response to Increasing Proppant Amount

3.3. AVERAGE PRODUCTION PERFORMANCE RESPONSE TO DIFFERENT RANGES OF PERFORATED LATERAL LENGTH

To consider the completion variation (perforated lateral length) on wells average cumulative production, Figure 8 has been constructed to illustrate the short and long-term average cumulative production per well for the wells hydraulically fractured in the Permian (Midland) Basin. The average cumulative well productivity is calculated for three different ranges of perforated lateral length of the stimulated horizontal well (<5,000, 5,000 -10,000, and 10,000 -15,000, feet per well). The figure shows that average cumulative production (BOE) per well versus time in years for each perforated lateral length wells' group. It is evident that increasing the length of the perforated lateral per well from 5,000 to 10,000 feet have positively increased the average cumulative productivity per well. The same was especially true when the perforated lateral length was increased from 10,000 to 15,000 feet, the average productivity per well (BOE) was tremendously increased.

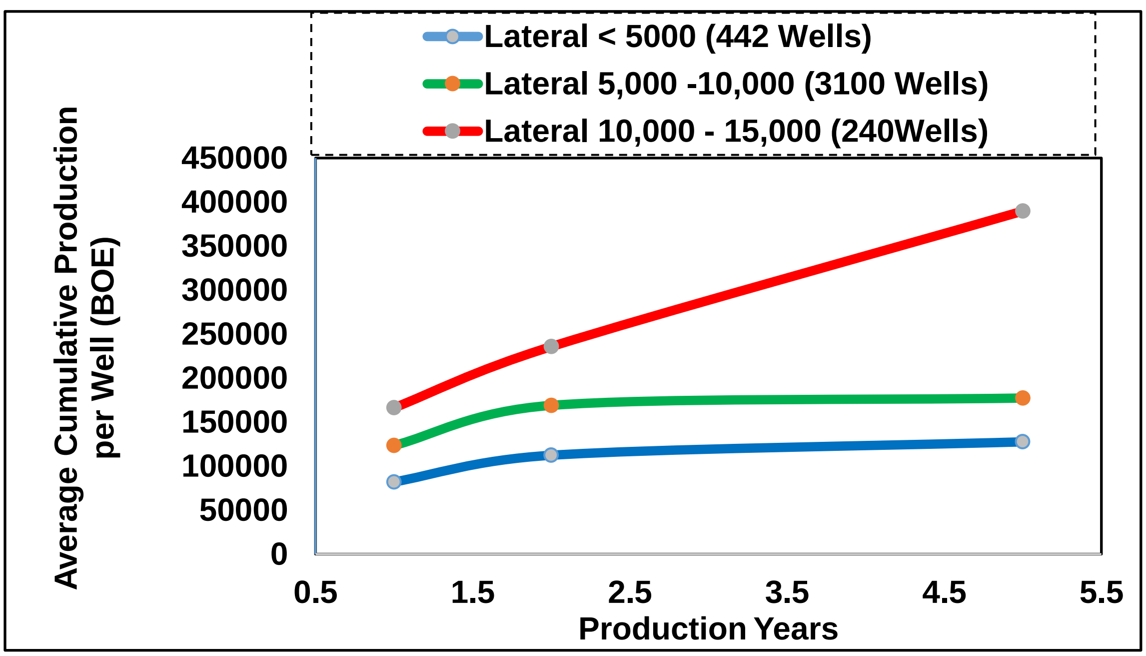


Figure 8. Cumulative Production versus Production Years for Different Groups of Perforated Lateral Length

Figure 9 demonstrate the percentage change of the average cumulative BOE per well in response to drilling and stimulating longer laterals. The blue line represents the cumulative productivity per well (BOE) change as a result to increasing the perforated lateral length from less than 5,000 feet wells' group to less than 10,000 feet wells' group. The productivity has increased with reference to the productivity of wells stimulated with perforated lateral length of < 5,000 feet by +50.5%, +50.4%, and +39% for the first, second, and the fifth year's average cumulative production per well (BOE), respectively. Increasing the perforated lateral length from the range of (5,000-10,000) feet per well to the range of (10,000-15,000) feet per well (green line) improved the average cumulative well's productivity over time (BOE) with reference to the productivity of wells' group stimulated with perforated lateral length of 5,000-10,000 feet per well by +34.6%, +39.5%, and +120% for the first, second, and the fifth year's average cumulative production per well (BOE), respectively.

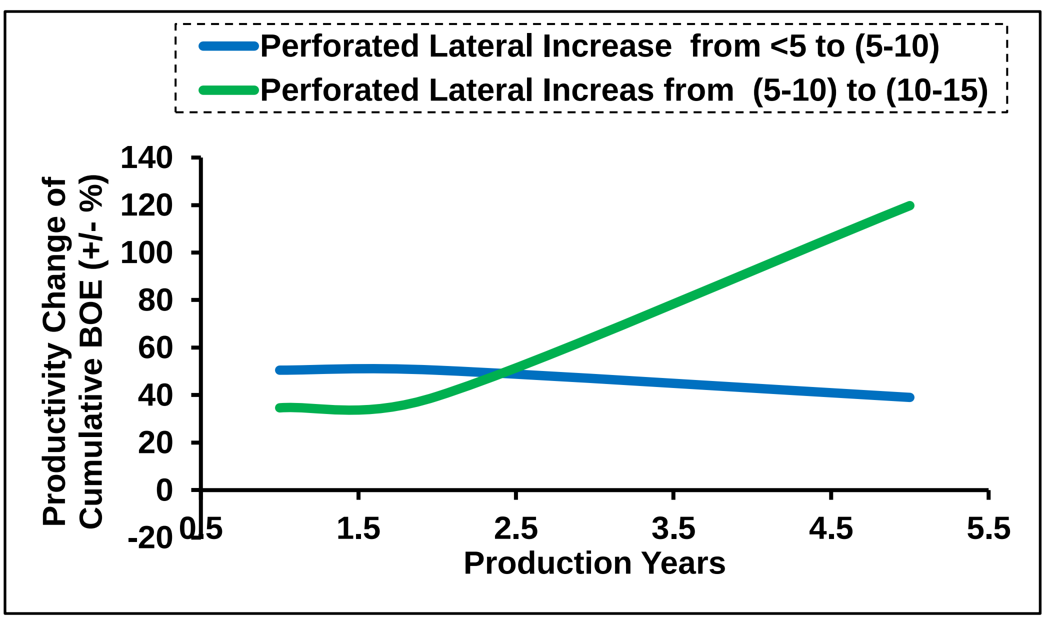


Figure 9. Productivity Change over Time in Response to Increasing Perforated Lateral Length

3.4. AVERAGE PRODUCTION PER FOOT PERFORMANCE IN RESPONSE TO DIFFERENT RANGES OF PERFORATED LATERAL LENGTH

Figure 10 shows cumulative production per stimulated foot versus time for two groups of perforated lateral lengths. The figure shows that the shorter perforated lateral outperformed the longer lateral range in terms of the amount of BOE per stimulated foot. This is generally true when not considering the size of the stimulation job (proppant pounds, mesh size, volume of base fluid). This is demonstrating the importance of properly sizing the fracturing stimulation treatment to comply with the fracture lateral length requirements to efficiently stimulate the well. Figure 11 shows the percentage in the drop in cumulative production per foot (BOE/Ft) as a result to increasing the perforated lateral length from < 5,000 feet to (5,000 – 10,000) feet. The first-year production/foot showed -7.7% drop, the second and the fifth year's production/foot showed 7-.31% and -14.8%, respectively.

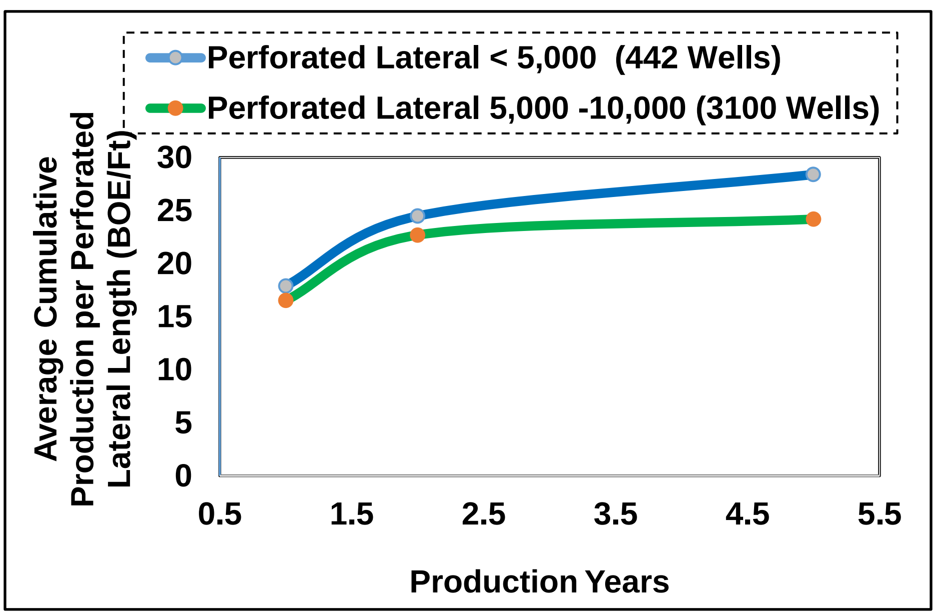


Figure 10. Cumulative Production per Foot versus Production Years for Different Groups of Perforated Lateral Length

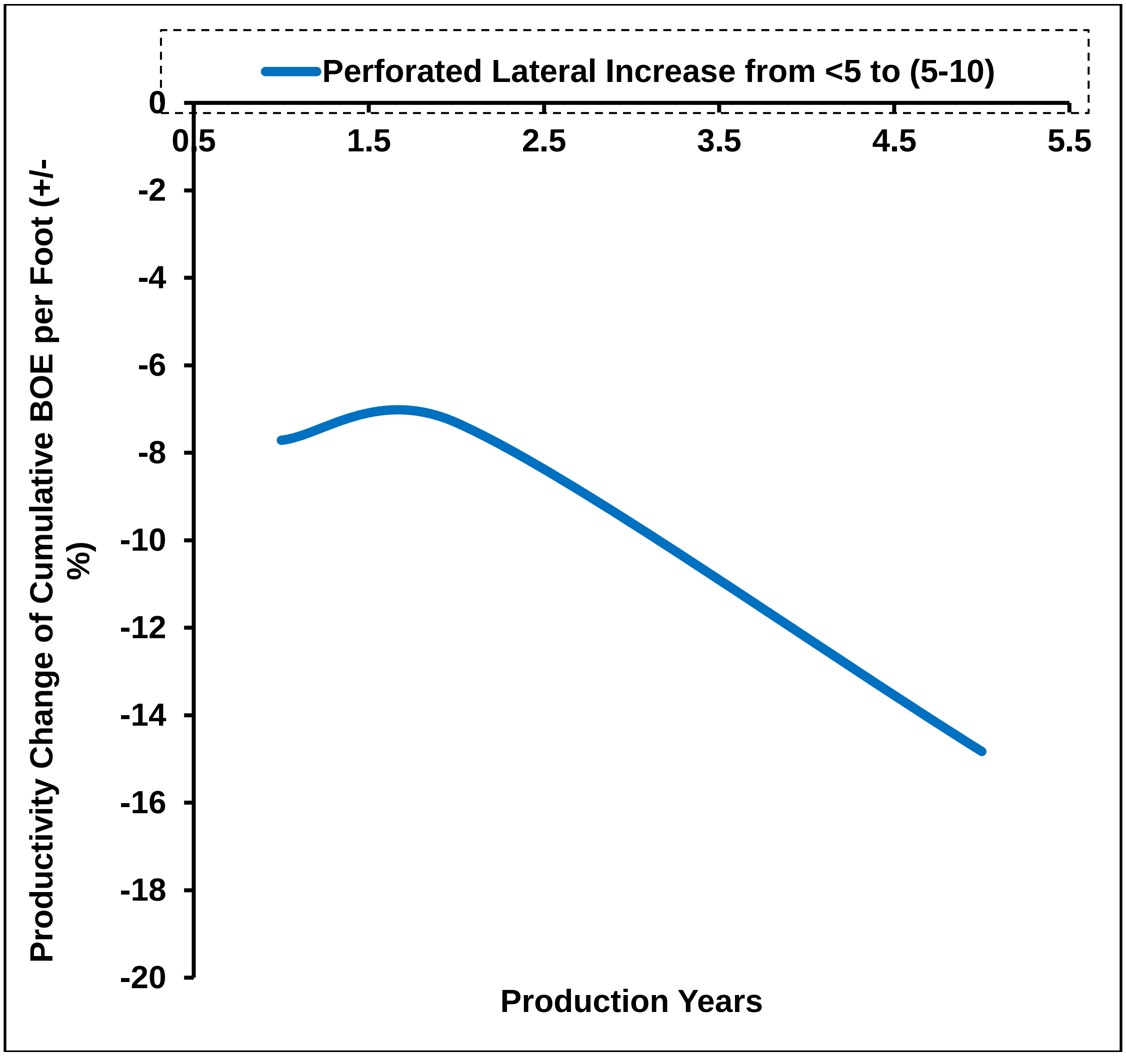


Figure 11. Productivity per Foot Change over Time in Response to Increasing Perforated Lateral Length

3.5. AVERAGE PRODUCTION PER FOOT PERFORMANCE IN RESPONSE TO DIFFERENT RANGES OF PROPPANT MASS PER PERFORATED LATERAL LENGTH

Figure 12 shows the average production per foot versus time for four different ranges of the amount of total pumped proppant per perforated lateral length. The figure shows that the high amount of proppant pumped per stimulated lateral resulted in higher average production from each stimulated lateral foot unit. This proves that the wells in the Permian responded very well to big size fracture stimulation jobs.

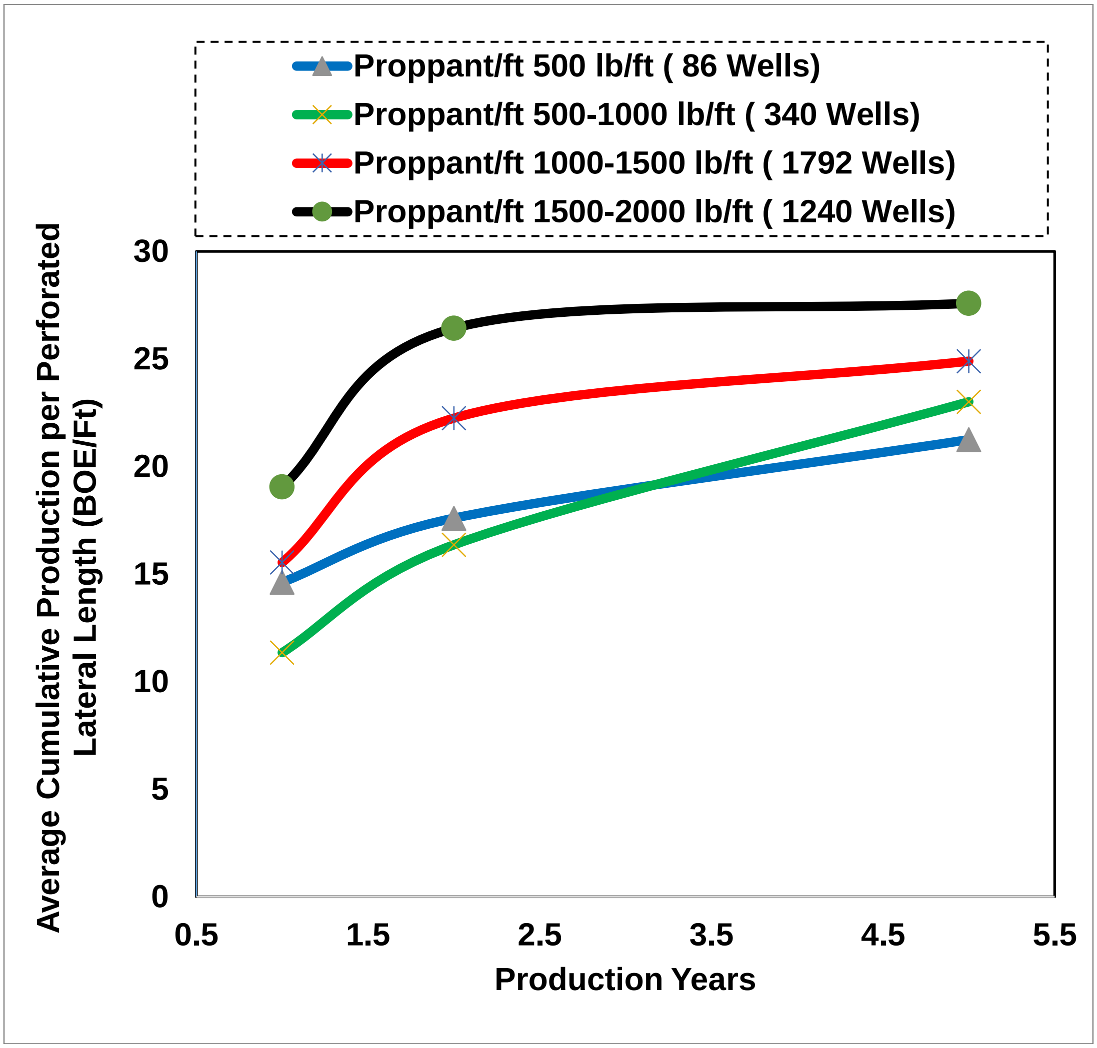


Figure 12. Cumulative Production per foot versus Production Years for Different Groups of Proppant Pounds per Perforated Lateral Length

4. CONCLUSIONS

- Pumping 10-15 million pounds per well had the best performance in term of production per foot and overall cumulative production. It also produced by the second year 94% of the total BOE produced / 5 years.

- The wells stimulated with 15-20 million pounds per well have shown similar performance to the wells stimulated with 10-15 million pounds per well in term of cumulative BOE production for almost the first three years of production, the only difference is that the wells continued to perform better after year 3 and produced 18.5 % more cumulative production at the end of year 5 than the wells stimulated with 10-15 million pounds.
- The average cumulative production per foot for the wells stimulated with different ranges of total proppant pumped has shown an improvement in the amount of BOE produced per stimulated foot as a result to increasing the total proppant pumped. The exception to that was with the very high amount of total proppant pumped (15-20 Million pounds) the productivity/foot for the first two years was lower than the productivity/foot for the wells stimulated with 10-15 million pounds of proppant. At year five the productivity/foot have shown an improvement which indicates that the excess amount of proppant has improved the long-term productivity/foot.
- Wells with longer stimulated laterals have produced more average cumulative BOE per well.
- Comparing the productivity per foot of different lateral lengths showed that the shorter lateral wells produced better BOE/foot than the long lateral wells which indicates that the longer laterals need to be stimulated with higher amount of pumped proppants.
- The wells that were appropriately stimulated with larger amounts of proppants per foot are the wells who produced the best in the Permian (Midland) Basin.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to DrillingInfo team and the co-founder and director of university outreach Mr. Mark Nibbelink for their generosity in providing access to the U.S. production and completion data.

REFERENCES

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar, A. M. "Review of Stimulation and Completion Activities in the United States Shale Plays". Paper ARMA 19-2003, presented at the 53rd US Rock Mechanics / Geomechanics Symposium held in New York City, New York, USA, 23-26 June 2019a

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar, A. M. "Production Performance Estimation from Stimulation and Completion Parameters Using Machine Learning Approach in the Marcellus Shale". Paper ARMA 19-2034, presented at the 53rd US Rock Mechanics / Geomechanics Symposium held in New York City, New York, USA, 23-26 June 2019b

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar, A. M. "Long-Term Comparison of Gel and Water Fracture Stimulation in Marcellus Shale Play". Paper SPE-195990-MS, accepted to be presented at SPE Annual Technical Conference and Exhibition (ATCE) in Calgary, Alberta, Canada, 30 September-2 October 2019c

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Al-Attar A. M., M. M. Alkhamis, Al-Bazzaz, W. H. "From Data Collection to Data Analytics: How to Successfully Extract Useful Information from Big Data in the Oil & Gas Industry?" Paper SPE-196428-MS, accepted to be presented at SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition in Calgary, Bali, Indonesia, 29 - 31 Oct 2019d

Arthur, J. D., Layne, M. A., Hochheiser, H. W., & Arthur, R. (2014, February 4). Spatial and Statistical Analysis of Hydraulic Fracturing Activities in US Shale Plays and the Effectiveness of the FracFocus Chemical Disclosure System. Society of Petroleum Engineers. doi:10.2118/168640-MS

Al-Alwani, M. A., Britt, L.K., Dunn-Norman, S., Alkinani, H. H., Al-Hameedi, A. T., Cedeno M. A., Al-Attar A. M., Al-Bazzaz, W. H. “Short and Long-Term Productivity Comparison of Hydraulic Fracturing Fluid Systems in East Texas Cotton Valley Formation” Paper SPE-196515-MS, Accepted to be presented at SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition in Calgary, Bali, Indonesia, 29 - 31 Oct 2019e

Carman, P., Li, C., & Wheeler, R. (2018, September 24). Unorthodox Fluids for Unconventional Completions: Maximizing Production and Avoiding Pitfalls. Society of Petroleum Engineers. doi:10.2118/191751-MS

DrillingInfo website. Retrieved on March/01/2019, from https://info.drillinginfo.com/

Elsner, M., & Hoelzer, K. (2016). Quantitative Survey and Structural Classification of Hydraulic Fracturing Chemicals Reported in Unconventional Gas Production. Environmental Science & Technology, 50(7), 3290-3314. doi:10.1021/acs.est.5b02818

FracFocus website. Retrieved on May/01/2019, from http://fracfocus.org/

Harper, John A. (2008). The Marcellus Shale - An Old "New" Gas Reservoir in Pennsylvania. Pennsylvania Geology, Volume 38, Number 1. Pennsylvania Bureau of Topographic and Geologic Survey

Khvostichenko, D., & Makarychev-Mikhailov, S. (2018, August 9). Oilfield Data Analytics: Linking Fracturing Chemistry and Well Productivity. Unconventional Resources Technology Conference. doi:10.15530/URTEC-2018-2903086

King, G. E. (2010, January 1). Thirty Years of Gas Shale Fracturing: What Have We Learned? Society of Petroleum Engineers. doi:10.2118/133456-MS

Luo, H., & Zhang, Y. (2018, October 5). Mining FracFocus and Production Data for Efficacy of Fracturing Fluid Formulations. Society of Petroleum Engineers. doi:10.2118/191787-18ERM-MS

Myers, R. R., Knobloch, T. S., Jacot, R. H., & Ayers, K. L. (2017, October 4). Production Analysis of Marcellus Shale Well Populations Using Public Data: What Can We Learn from The Past? Society of Petroleum Engineers. doi:10.2118/187505-MS

Olaoye, O., & Zakhour, N. (2018, September 24). Utilizing a Data Analytics Workflow to Optimize Completion Practices While Leveraging Public Data: A Permian Basin Case Study. Society of Petroleum Engineers. doi:10.2118/191416-MS

Romero, D. H., & Poston, S. W. (2016, May 10). Data Mining and Analysis - Eagle Ford Formation, South Texas. Society of Petroleum Engineers. doi:10.2118/179994-MS

Srinivasan, K., Ajisafe, F., Alimahomed, F., Panjaitan, M., Makarychev-Mikhailov, S., & Mackay, B. (2018, August 28). Is There Anything Called Too Much Proppant? Society of Petroleum Engineers. doi:10.2118/191800-MS

United States Environmental Protection Agency website (EPA). " Safer Chemical Ingredients List" Retrieved on August/01/2019, from https://www.epa.gov/saferchoice/safer-ingredients#scil

Weijers, L., Wright, C., Mayerhofer, M., Pearson, M., Griffin, L., & Weddle, P. (2019, January 29). Trends in the North American Frac Industry: Invention through the Shale Revolution. Society of Petroleum Engineers. doi:10.2118/194345-MS

Yang, Y., Robart, C. J., & Ruegamer, M. (2013, February 4). Analysis of U.S. Hydraulic Fracturing Design Trends. Society of Petroleum Engineers. doi:10.2118/163875-MS

VIII. FROM DATA COLLECTION TO DATA ANALYTICS: HOW TO SUCCESSFULLY EXTRACT USEFUL INFORMATION FROM BIG DATA IN THE OIL AND GAS INDUSTRY?

Mustafa A. Al-Alwani, Shari Dunn-Norman, Larry K Britt, Husam H. Alkinani, Abo Taleb T. Al-Hameedi, Atheer M. Al-Attar, Mohammed Alkhamis, Waleed Al-Bazzaz

Department of Geosciences and Geological and Petroleum Engineering, Missouri University of Science and Technology, Rolla, MO 65401, USA

ABSTRACT

Big data has become a major topic in many industries. Most recently, the oil and gas industry adopted a special interest in data science as a result to the increasing availability of public domains and commercial databases. Utilizing and processing such data can help in making better future decisions. The aim of this work is to provide an example and demonstrate methodologies on how to collect and utilize big data to help in making better future decisions in the oils and gas industry.

After reading a good number of papers and books about the applications of data analysis in the oil and gas industry in addition to other industries, and given that data analysis is the area of expertise of the authors, this paper was written to demonstrate real examples of data processing and validation workflows. This work is intended to cover the gap in the literature were many of the publications only discuss the importance of data-driven analytics.

This paper provides an overview of the diverse and bulk data generating sources in the oil and gas industry, starting from the exploration phase to the end of the lifecycle of the well. It provides an example of utilizing a public domain database (FracFocus) and demonstrates a step by step workflow on how to collect and process the data based on the

objective of the analytics. Two real examples of descriptive and predictive analytics are also demonstrated in this paper to show the power of having a diverse and multiple resources databases. A framework of data validation and preparation is also shown to illustrate data quality checks combined with best practices of data cleansing and outlier detection methodologies.

This paper provides a clear methodology on how to successfully apply data analysis which can serve as a guide for some future data analysis applications in the oil and gas industry.

1. INTRODUCTION

The anecdotal saying about data is knowledge is not necessarily true. Data are generally recorded events as they take place. Processing and analyzing the data yields knowledge. With gained knowledge, an understanding of why things are happening can be gained and then utilized in the desired direction to optimize the outcomes. The route data analysis studies must go through several phases starting with the collection and understanding the data parameters, passing through data visualization and descriptive analysis, especially for large and complex datasets. Visualization of the data will help in conveying insights and overall comprehending of the common trends buried within the data. Good data visualization will also help the non-specialist or non-technical individuals to be able to understand technical and specialized data. After understanding the data and investigating the trends, modeling and testing hypotheses follow through as the phase of predictive analytics which leads to prescriptive analytics. The goal of data analytics is to gain useful insights and extract valuable parameters that can help in producing informed decisions.

Big Data can be defined as a collection of datasets that are large and complex in nature. The structured and unstructured datasets grow large and fast with time as more data are generated and added to the original database. The fast-growing rate of such data makes the traditional relational database systems and other conventional statistical tools fail to manage this kind of databases.

Staring with big and complex dataset then subjecting it to rigorous analysis and interpretation phases can generate educated solutions for businesses or any operations success. To gain more knowledge, the descriptive analytics is taken a step further to investigate the trends and to understand the driving hypotheses behind the data by building independent models and formulating assumptions on why the trends and patterns exist. The built complex mathematical models can be run to validate the assumptions and theories behind the trends and generate artificial datasets. The generated models are checked against other sampled data from the original data set that have never been used to generate the model.

The type of the data is characterized by the speed they are generated, the volume of represented items, the variety of the type of the data they contain, and the degree of veracity they own. Big data analytics comes to play to find answers from the data by utilizing a combination of high technology systems and mathematics which together are capable of processing all the information and providing valuable insights.

The data analytics starts with collect the data, store, process, analyze, and find patterns. Data analytics can be classified into three major types. Descriptive data analytics describes what happened in the past by presenting the data through graphics and reports, descriptive analysis is not necessarily capable of explaining why the patterns exist or what will be the future trend. Predictive analytics utilizes the data to predict what could happen in the future or what are the expected trends of certain parameters. Prescriptive analytics evaluates the outcomes of predictive analytics and decides on how to proceed with future decisions or alter previous designs. Big data analytics and the revolution of datafication help companies and public administrations to better understand the data, find previously unnoticeable patterns and provide better solutions for existing and future operations.

The difference between the traditional datasets and big datasets are characterized by many attributes. The volume of the data in the traditional databases may reach to gigabytes or even terabytes while in big datasets the volume can go up to petabytes and zettabytes. The data structure is mostly structured (such as tables, columns, rows) in traditional and semi-structured or/and unstructured (no specific formats such as emails, text, video, audio files) in big datasets. The storage organization of the data is stored centrally in the traditional data and distributed in big data. The data model is based on the strict schema in the traditional and flat schema in big data. Finally, the data relationship is complex interrelationships in traditional while almost flat with few relationships in big data.

This study provides an overview of the data sources in the oil and gas industry with an example from FracFocus database to demonstrate the process of collecting, processing, and analyzing the data based on the objective of the analysis. In addition, a framework of data validation and preparation is shown to formulate data quality rules combined with best practices of data cleansing and outlier detection methodologies. To show the power of having multiple sources of data, two real examples of descriptive and predictive analytics are also shown.

2. OIL AND GAS INDUSTRY DATA SOURCES

The oil and gas industry generates and stores an enormous amount of data. In the past, the industry has been tormented by accumulating too much data without the capabilities of deducting insightful outcomes from the collected data (Feblowitz et al., 2013). The industry has recently started to transition from the data collection mode to proactive use of the data mode. The industry generates a massive amount of data during all phases from exploration to production then abandonment. The industry is going through an exponential increase in the number of sensors installed in upstream, midstream, and downstream operations. The ultimate objectives of the utilization of big data and analytics are to reduce operational cost, increase safety, and enhance productivity. The following sub-section briefly demonstrate the source of data being generated in different phases during the lifecycle of oil and gas operations.

2.1. EXPLORATION AND RESERVOIR MODELING PHASE

In the exploration phase, big data approach is very important to process the seismic data. Seismic data processing requires high-speed computational power and clustered and parallel high-performance data storage (Vega-gorgojo et al., 2016). Such infrastructure is deemed necessary to create geological models in 3D to explain the complex geological structures underground. The data obtained from seismic usually coupled and boosted by other data sources obtained from offset wells such as rock types and logs. As technology advances, an additional vast amount of data are being generated. For example, in offshore exploration, the old technology was called narrow azimuth towed streaming (NATS) while the recent technology is called wide-azimuth (WAZ). WAZ generates more than six times the data being generated by NATS. Seismic data recording has not only been used for the exploration phase but also recently has been utilized in monitoring wells that contain permanent geophones. Those geophones are used to collect data to investigate fluid front movement, monitor carbon capture sequestration, and detect microseismic events due to fracture stimulation in nearby wells to map the fracture growth and dimensions. Seismic data centers collect and store data which reaches to 20 petabytes of information and to put that in perspective, it represents 926 times the size of the data of the U.S. Library of Congress (Beckwith, 2011). The new techniques of machine learning and data analytics can be utilized as part of the seismic interpretation methods to find new discoveries.

2.2. DRILLING PHASE

Most of the modern drilling rigs used in drilling oil and gas wells are normally equipped with many sensors to continuously record all the operation from the time the well is spud till the well is completed and ready for the production phase. The safety and the performance are the two main key performance indicators that the drilling engineers and the operators are actively trying to achieve. Data such as the weight on bit (WOB), the rotary speed of the bit (RPM), pumping pressure (psi), torque (lb.ft), etc., are constantly monitored and studied to optimize the drilling time and reduce the non-productive time (NPT). Multiple sources of data are also recorded and collected by service companies on location and all the data gathered and stored for each drilled well. Well logs are run several times during the drilling phase to collect formation and fluid information with open hole logs that can be performed using wireline logs or logging while drilling techniques. The cement bond logs are run to check for the quality of the cement around the casing to make sure zonal isolation is achieved. Stress and geomechanical tests generate a large amount of data associated with each well. Well control and safety monitoring data are also an important part of the data collection and monitoring. Data such as the percentage of the gas in the mud while drilling through an active reservoir formation and the type of the gases coming to surface with drilling mud (gas chromatography) are constantly monitored and modeled to prevent accidents from happening. A large amount of the drilling data that is generated, stored, and interpreted are dramatically increasing with time and as the regulations require close monitoring and capturing more data, data analytics are becoming more important as the time progresses. Drilling data can be utilized to minimize the cost and NPT to achieve an optimum drilling operation (Alkinani et al., 2018a; Alkinani et al., 2018b; Alkinani et al., 2018c; Alkinani et al., 2019; Al-Hameedi et al., 2017a; Al-Hameedi et al., 2017b; Al-Hameedi et al., 2017c; Al-Hameedi et al., 2018).

2.3. PRODUCTION PHASE

Production data are very important and always coupled with other phases. The objective in many of the reservoir management and all drilling and completion techniques is to improve the productivity of the wells. Production data are used in data science as a gauging parameter (response) to other parameters that the engineers are trying to optimize. Production data are also utilized in reservoir modeling by trying to history matching the production to fine-tune the geological and simulation models. Water production data can be compared to oil and gas data to avoid completing near aquifers as well as to help in taking decisions in production monitoring by adjusting the artificial lift devices' settings (e.g. control ESP, plunger, rod pump). Some operators compare the production data and the geographic distribution of the wells to select the best investment areas to drill more wells by identifying the best reservoir properties and oil and gas pockets. The development of the concept of smart oilfields has flourished and enriched big data analytics. For every well with intelligent completion, a large amount of data are being generated and stored. Hence, artificial intelligence applications are being implemented to take actions based on pre-set algorithms. Different sensors are installed as part of the completion downhole and the data are transmitted by fiber optic cables. Examples of the data being measured are flow rate for every phase of the production fluids, water and gas oil ratio, methane emission during hydraulic fracturing fluids flowback, pressure and temperature along the well all the way to the wellhead, electrical submersible pumps operating parameters and power usage, etc.

2.4. OPERATION PHASE

Operations encompass a big stream of structured and unstructured data. The data are varied and consists of a large range of formats from complex 3D models to sensors data. The speed accumulated is also challenging as a result of the enormous number of sensors that are incorporated in most of the operations. Applying big data in the oil and gas industry drives the reduction in wells' shutdown or productivity impairment. Big data analytics also helps in extending the operational lifetime of equipment by predicting the failure rates and suggesting condition-based preventative maintenance. Real-time data streaming enables worldwide operational support which improved the quality of the jobs while minimizing crew in remote locations such as offshore environments. Service providers analyze the operational conditions of their downhole tools such as mud motor and logging while drilling (LWD) by continuously analyzing data from several sensors that measures the temperature, pressure, and vibration to predict failure and provide surface warning to change the operational parameters and hence prevent potential tool failure and consequential NPT and loss of money. Overall, in all phases of the oil and gas industry, data analytics are used to leverage data-based decision making in all operations and processes.

In all the E&P phases, the dramatic growth in data generation is not useful by itself. The ability to utilize and integrate the diverse data sources to seek useful insights and provide data-driven actions and decisions is the main target of using big data approaches in the oil and gas industry.

3. DATA COLLECTION AND FORMATTING

Data collection and preparation considered the first step in any data analytics. Data collection can be a challenging task because all the subsequent analyses will depend on data quality and reliability. Before embarking in any data-driven studies, a well-defined objective statement should be formulated to decide on what type of data should be collected to give a meaningful output. The results of data analytics are the end-product, insights will be generated and decisions will be made based on those outcomes.

Understanding and organizing the data at the early phase of data collection saves the time of the data analysts when they try to sift through the data to extract useful insights for a better-informed decision that can save time and cost. The engineers in the oil and gas industry spending a large portion of their time trying to search and assemble the data needed for their projects (BruléGroup, 2015). A recent survey conducted in 2018 by

General Electric and Accenture among oil and gas executives and 81% of the participants indicated that big data and data acquisitions are on their top three priorities for 2018 (Mehta, 2018)

To start a data collection project, the first step is to identify the scope of the project and search for the potential sources. The state and the local governments in the United States regulate the oil and gas industry and require the operators to reports their data to the state which are also made publicly available for the community. Texas railroad commission website (2019) is one example of public data source that contains production and wells' data files. Some other websites combine most of the states and the regulations require the operators to report specific data to that website for the purpose of public information. An example of that is the FracFocus website (2019) that is managed by the groundwater protection council. FracFocus contains all the chemicals types and percentages that are pumped in the hydraulically fractured wells. Some commercial databases such as IHS Market (2019) and DrillingInfo (2019) have been developed to collect and process the oil and gas data and provide a service of validated, standardized and easy to access organized data. Those databases provide a large number of data elements on millions of wells' records which can go back to wells drilled and produced since 1859. The wells data in the United States are obtained from the regulatory agencies or directly from the operators. The following subsections demonstrate examples of data collection and processing steps.

3.1. COMBINING FILES FROM THE SAME SOURCE

Once the objective of the study has been identified, the first step is to collect the data files from the source then combine them into one database to make the data ready for the analysis. In this paper, an example will be shown of how to utilize the public records of FracFocus to collect hydraulic fracturing chemical data and integrate it with production and completion database obtained from a commercial database (DrillingInfo). Figure 1 shows a workflow of stacking more than one file to generate one final file to process the data for the analysis phase. In this example, all the data on FracFocus were downloaded and tabulated in an Excel file. To stack data files together and merge them into one file, all data columns of each file must be exactly the same in terms of the column header name and the format of the data type. Each chemical component of the wells from the FracFocus records was represented in one Excel row with multiple columns that cover all the reported data. Due to the limitations of how many rows each Excel sheet can contain (1,048,576 rows) five Excel sheets were used to contain all the 4,146,279 chemicals (rows) that were found in 142,978 wells. The workflow was generated using self-service data analytics platform (Alteryx) and the final combined file was saved in a txt format which can be used for the next step of data cleaning and format checking.

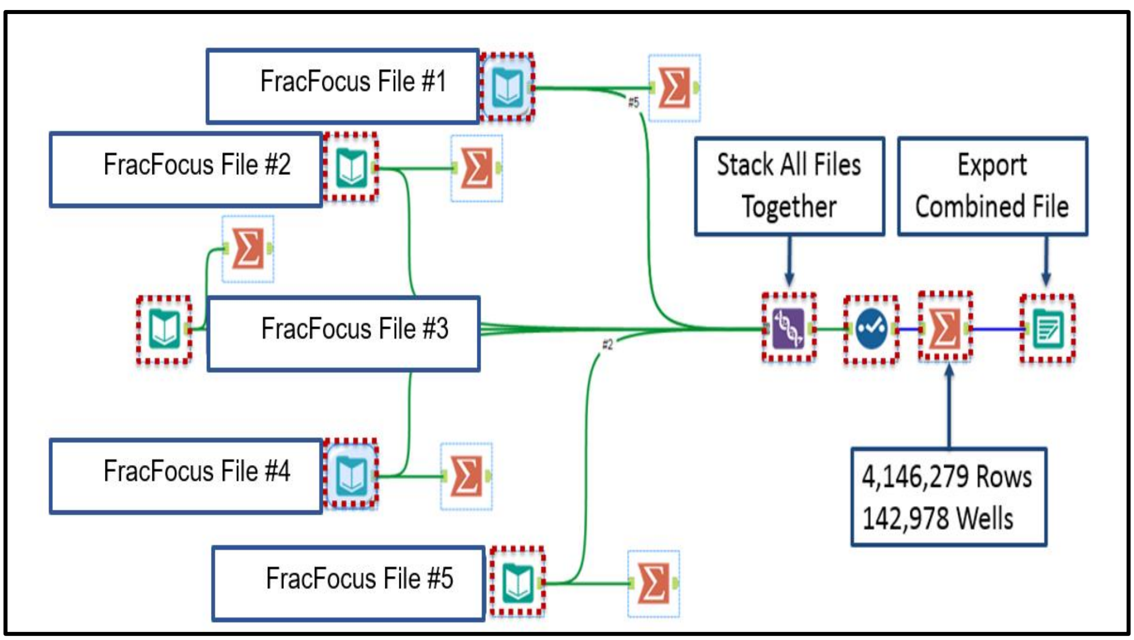


Figure 1. Example of Data Collecting and Stacking into One Data File

3.2. DATA CLEANING AND FORMATTING

Once all the extracted data are combined in one file, the second step is to format and select the columns needed for the analysis. Figure 2 shows a workflow example data formatting and columns selection process, in this particular example, the dates columns of the hydraulic fracturing start and end dates were in a long format which includes the hours and seconds, a date formatting function was chosen to change the data formats to mm/dd/yyyy. Data cleaning tool was used to remove and handle whitespace, capitalization issues, and nulls. The geographical tool was also used to convert the longitude and latitude data into map points to plot the wells on the map for the descriptive analytics. The column selector function was used to eliminate the data that have no meaning to the analysis and only retain the columns with useful parameters. Finally, the cleaned and formatted file is exported to a txt file to be used as an input to the next data processing step.

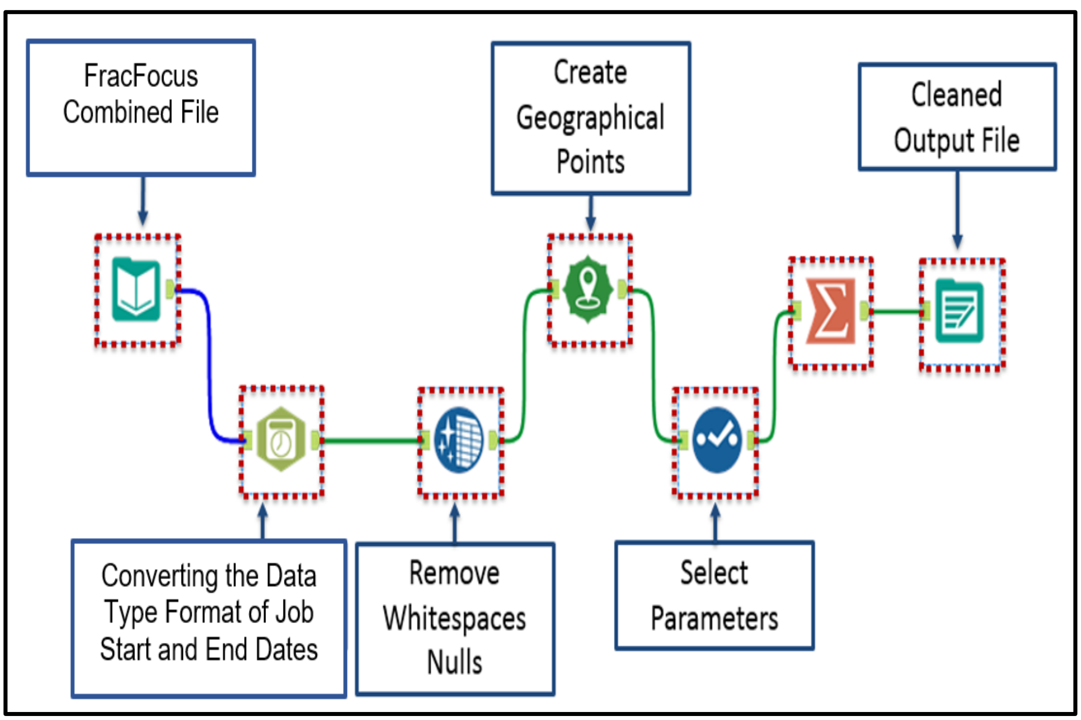


Figure 2. Example of Data Formatting and Columns Selection

3.3. DATA PROCESSING TO SATISFY THE OBJECTIVES

Once the data is stacked, formatted, cleaned, and selected, the next step is to process the data to be organized in a format that is useful for the analysis purpose. In the following example which is depicted in Figure 3, the objective was to evaluate the quantities of proppant (sand) used in each fracturing job per well in the U.S. Since FracFocus reported data only provides the weight percentage of the chemicals (including proppant) and for each well there are several types of proppant used with different mass percentages, data processing workflow was created to process the proppant data. The workflow starts by selecting the proppant related columns and the unique well number (API Number). The workflow is designed to detect and remove any data duplications by comparing all the parameters and flag the duplicated rows. The following step is to aggregate the proppant data for each well as one grouped value in the same time inspect if the well has been refractured by comparing the API number and the job start and end date. The wells that have been refractured will be grouped separately and the final proppant summed up value for each well will be determined and then saved in a file that contains all the proppant related data. The same procedure and workflow will be created to handle and process the other chemical ingredients separately. Once all the chemical ingredients and the water data files are created, a joining process will be started to create the final FracFocus database that includes all the water and chemical data processed and summed up for each well.

3.4. COMBINING DIFFERENT DATABASES

To enlarge the database and include more parameters such as the well completion and production data, another database has to be joined with the created chemicals database. In this database example, DrillingInfo which contains the completion and production data of the wells was chosen to be integrated with the stimulation chemicals database. Figure 4 shows the workflow of joining two different databases. The two files are imported from DrillingInfo and FracFocus then the column selector tool was used after each database to select the parameters to be joined. The join tool is used in this workflow to join all columns (completion, production, and stimulation chemicals). In order to use the join tool, a common entity must exist between the two databases to be identified as the joining parameter and in this case, the API number was used to join the parameters from both databases.

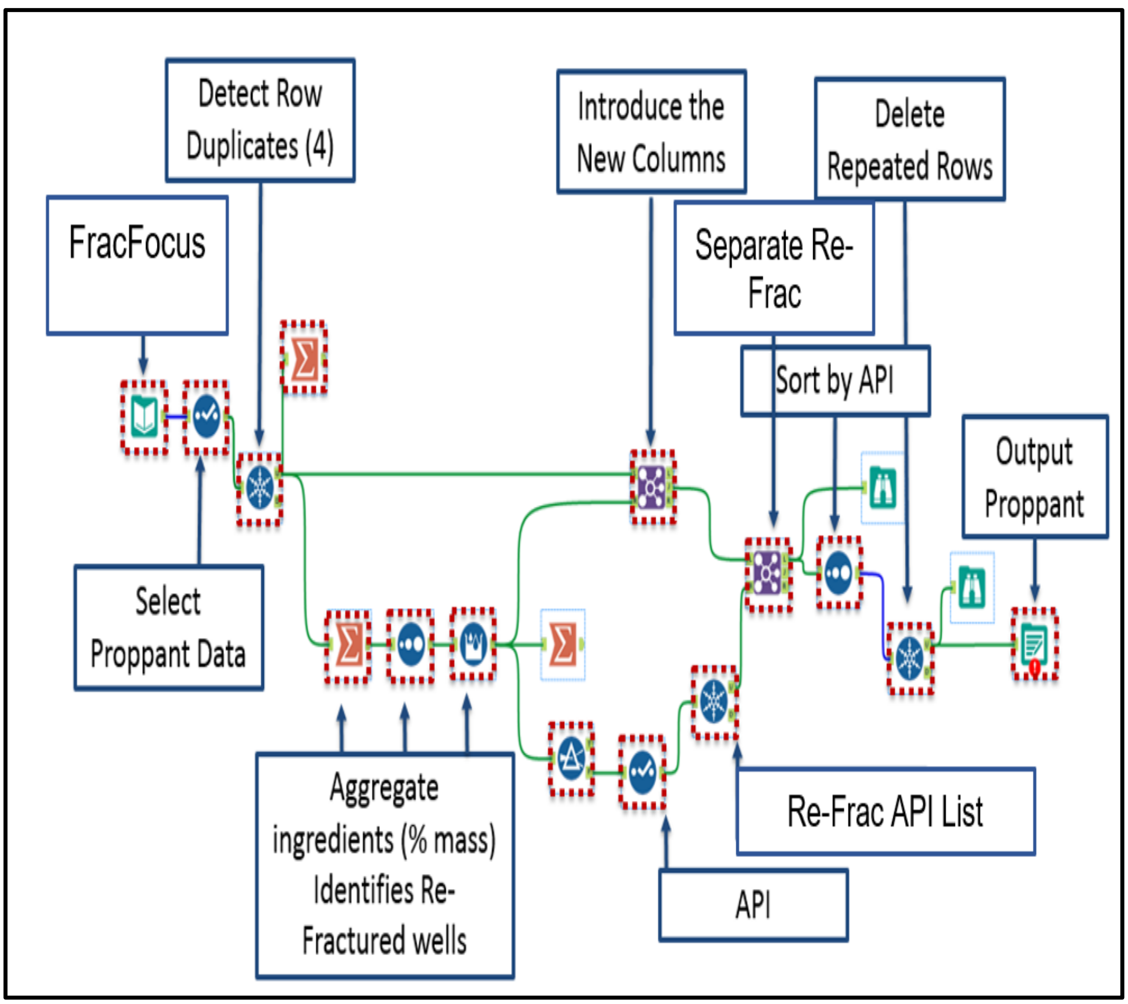


Figure 3. Proppant Data Processing Example

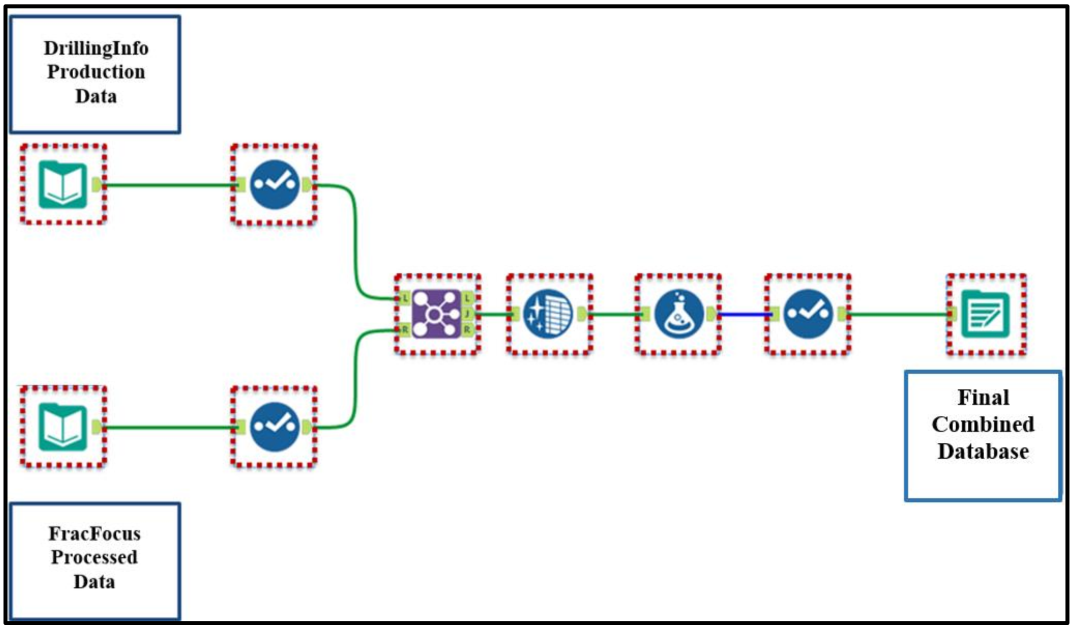


Figure 4. Example of Combining Two Different Databases

4. DATA VALIDATION AND QUALITY ASSURANCE

In this section, a non-case dependent framework is presented to deal with data validation and preparation. The data point passes quality check when it can serve the planned purpose in an operation or when it can lead to a clear and correct decision making (Dai et al., 2018). The data coming from the oil and gas industry are not always accurate and sometimes have some missing values. Corrupted data cost the industry about $60 billion per year (Nobakht and Mattar, 2009). Most datasets are expected to contain from 1-5% error. The more caution and validation applied during the data collection process, the less the error percentage. The most common error sources in the data are human errors, wrong measurement set up, or malfunction in measuring equipment. Dealing with missing data or suspected data requires a deep understanding of the dataset and its behavior. A good understanding of the dataset leads to a better decision when it comes to remedy and substitute for the corrupted and missing data (Al Attar et al., 2016). The enforcement of data quality best practices will not yield instantaneous results, but it will at least ensure a large signal/noise ratio that accumulates as the process of data collection progress. Data quality and validation tests vary from being simple to structures generalized, but in both cases, a wide knowledge in the data domain is required. The data validation for the oil and gas datasets is significantly different in context from a validity check for a healthcare dataset, with this contrast the validity is still looking for similar patterns to evaluate the data quality and fit for operations decision-making.

4.1. DATA VALIDATION

Data validation defined as the process of trying to validate and verify if the value of a data point comes from a known finite or infinite set of defined acceptable values (UNECE, 2013). Data validation is also defined as the process of ensuring that the final dataset complies with several predetermined quality characteristics (Simon, 2013). Di Zio et al. (2013) adopted a definition that considers the communication between the data records on the variable level and on the field domain level which defines data validation as the process of verifying that the dataset combination of values whether or not belongs to a set of acceptable combinations.

4.2. ACCURACY AND PRECISION

Accuracy and precision are two terms often used without referring to their real meaning as they define two different aspects of data measures. Accuracy refers to the correlation between an expected (theoretical) result and an actual value. For example, the predicted value of the temperature of the reservoir and the measured value. The percentage of the difference between these two values (Equation 1) describes the accuracy of that prediction.

$$\% \, Difference = \frac{Theoretical\ or\ Experimental - True\ Value}{True\ Value} * 100\% \qquad (1)$$

Precision, in summary, is how the experimental data are reproducible. Yielding low differences between sequential realizations or experimental trials pronounces a high precision experiment. Precision can be a good indicator of how good and solid the experimental setup taking in consideration that the domain is highly reproducible because some domains have high variability that yields low precision. For example, if a density measurement tool measures the reservoir fluid, assuming the tool was run three times to measure the density of the same fluid under the same conditions. The larger the difference between the three measurements, the lower the precision of the tool. The precision can be quantified by utilizing the descriptive statistical measures such as range and standard deviation.

4.3. ERROR SOURCES IN THE DATA

Some examples of sources of errors in the data are as follow:

1. Human error: these errors include the misreading and interpretation of the data that has been recorded by field personnel.

2. Data collection setup error: the way that data are collected might have a probability of error to propagate. An example of that is when the single well flow rate of four producing wells is calculated by dividing the total flow rate by four. This assumes that the rates

from the four producers are equal, and in fact, they most likely are not. Nobakht and Mattar (2009) mentioned some error sources related to production and injection data include:

(i) Averaging the rates of a specific well from an adjacent group of wells.

(ii) Incorrect assumptions (single phase flow while GOR is increasing for example).

(iii) Incorrect location of the pressure gauge (i.e. it is downstream while the choke is not fully opened).

(iv) Incorrect synchronization (i.e. taking the injection and production data with a time shift between them and treating them as being measured at the same time).

3. Device malfunction which results in corrupted data: The missing or corrupted data can have an impact on the decisions that need to be made. The error in one dataset may propagate to parameters in another dataset.

5. OUTLIERS DETECTION METHODOLOGIES AND TREATMENTS

Hawkins (1980) defined the outlier as an observation in the dataset that significantly deviates from other observations which raise the suspension that the observation was generated by different mechanism or approach. The complexity of dealing with outliers come from the method used to decide if the data point is corrupted or is normal behavior and that comes from the understanding of that specific dataset. Three methods to identify the outliers which are box plots, descriptive statistics method, and the Local Outlier Factor (LOF) will be disused in the following sub-sections.

5.1. BOX PLOT

Box plots are a useful and easy visualization tool to detect outliers. Three are five elements of box plots which are maximum, minimum, median, first and third quartiles as shown in Figure 5 The difference between the first and the third quartiles is called the interquartile range (IQR). Data points fall outsides the maximum or the minimum of the box plot whiskers can be potential outliers (Kirkman, 1992).

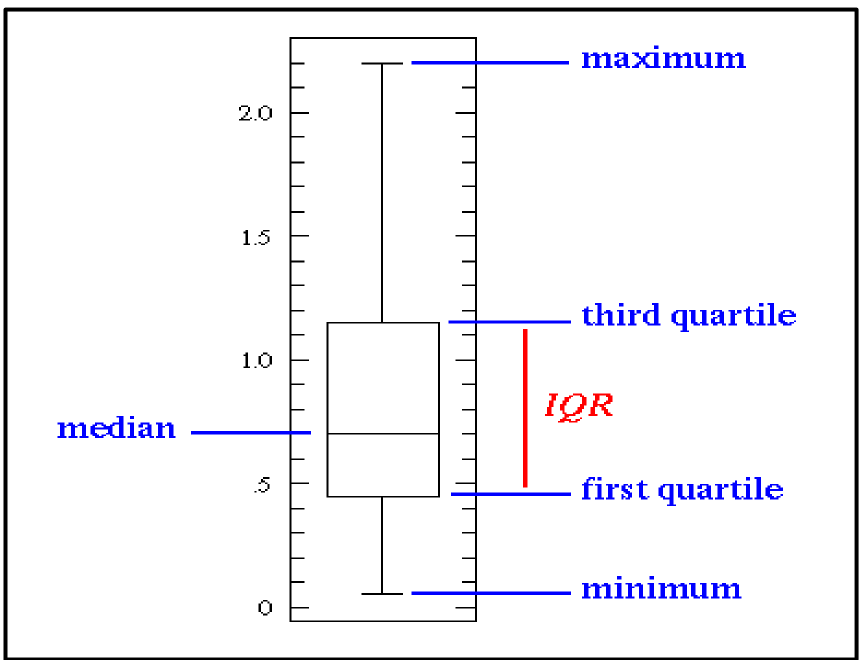


Figure 5. Box Plot Elements (Kirkman, 1992)

5.2. DESCRIPTIVE STATISTICS METHOD

The assumption here is that the data is normally distributed according to the Central Limit Theory (Franklin and Brodeur, 1997). This methods flags any data point falls outside the six standard deviations from the mean, the data point will be considered as a potential outlier. Example of flow rate outlier detection using this method is shown in Figure 6.

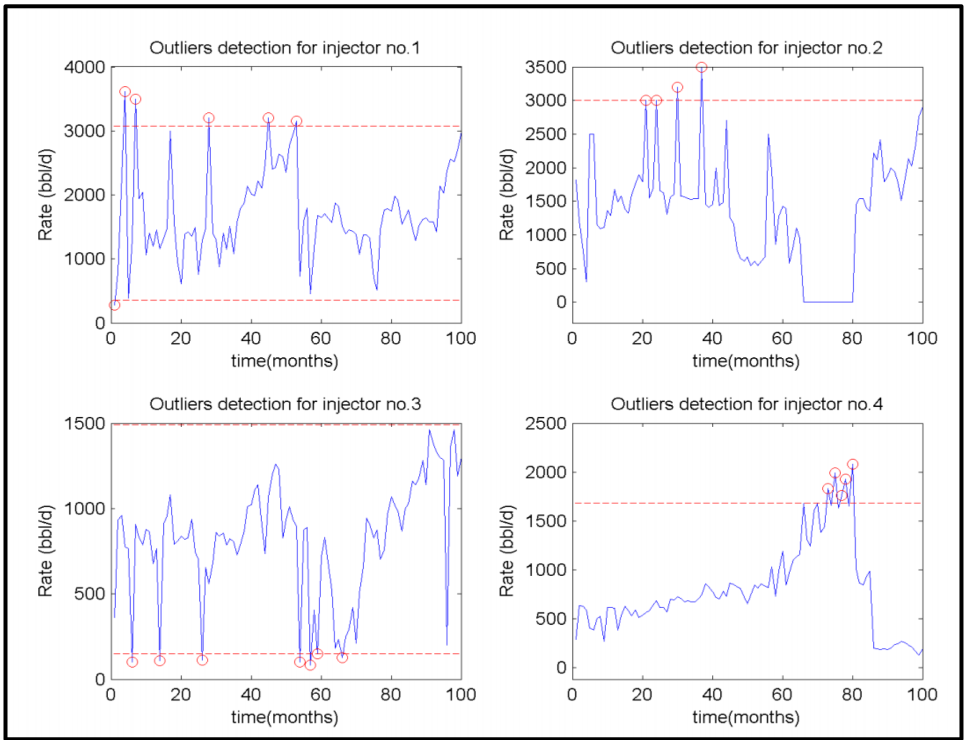


Figure 6. Descriptive Statistics Method to Detect Outliers (Al Attar et al., 2016)

5.3. LOCAL OUTLIER FACTOR (LOF)

LOF was developed by Breunig et al. (2000). The k-distance (A) can be defined as the distance from the kth nearest neighbor to point A as shown in Figure 7. Nk(A) is the set of k points, the reachability distance between the group of points Nk(A) and A can be defined as:

$$Reachability\ distance\ (A, B) = max\{k - distance(A), d(A, B)\} \qquad (2)$$

Where:

d (A, B) is the distance between A and B i.e. $|Y(A) - Y(B)|$

B ∈ Nk (A). i.e. the reachability distance between two points A and B is the true distance and at least k-distance (B).

The local reachability density lrd is defined as:

$$lrd(A) = \frac{\sum B \in N_k\ (A) Reachability\ distance\ k\ (A,B)}{n(N_K(A))} \tag{3}$$

The LOF factor is given by:

$$LOF_k = \frac{\frac{\sum B \in N_k\ (A) \frac{lrd(B)}{lrd(A)}}{n\left(N_k(A)\right)}}{lrd(A)} \tag{4}$$

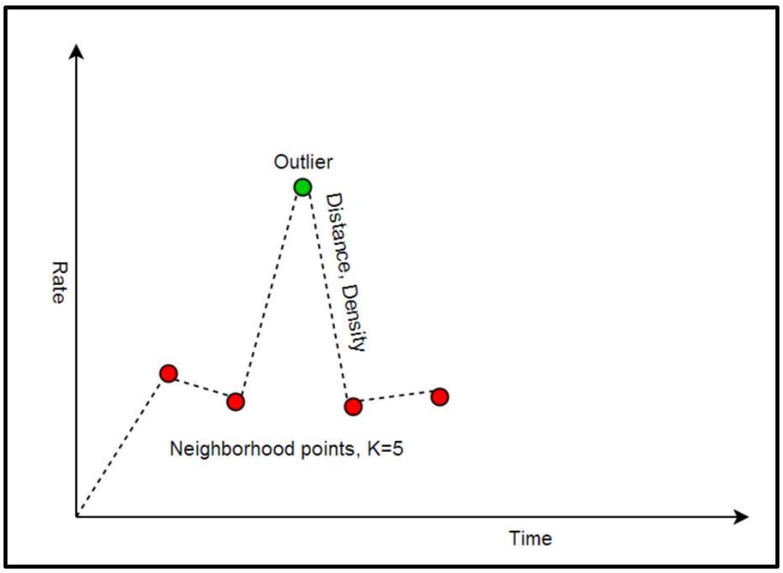
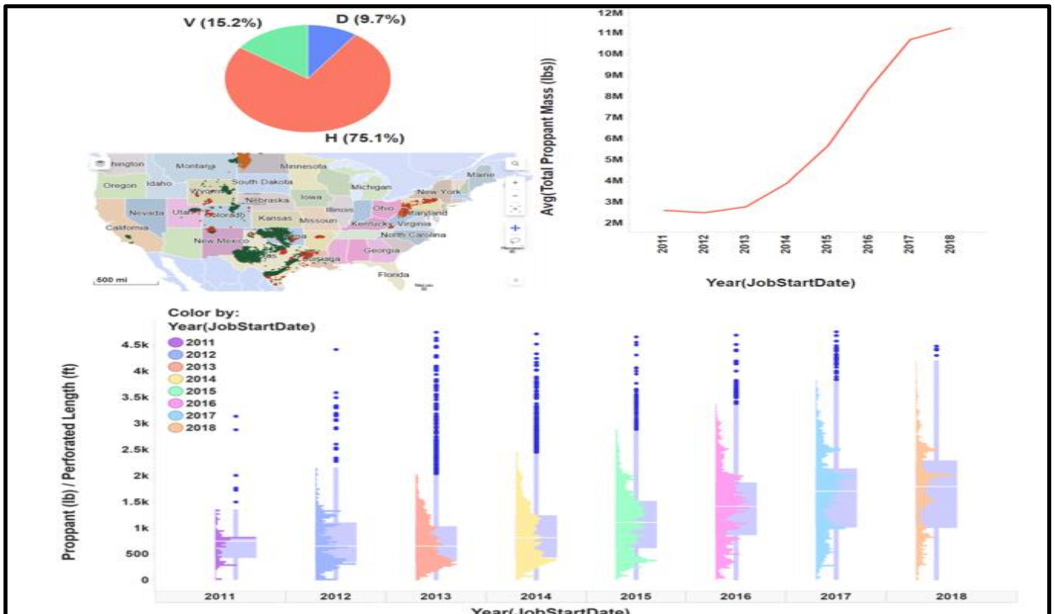


Figure 7. Local Outlier Factor (LOF) Method Definition (Al Attar et al., 2016)

6. DESCRIPTIVE DATA ANALYTICS

The descriptive analysis uses the data to investigate and understand hidden trends and insights by utilizing statistics and data visualization. It has become a growing topic in the oil and gas industry especially after the availability of public and commercial databases and the advancement in statistical and visualization software. Many of the commercial databases have their own version of dashboards-based analytics which helps the users to drill down the data and select the parameters of interest that the user would like to investigate and observe their trends. The descriptive analysis should not be done without an overall knowledge about the parameters and the process being analyzed. Misleading insights and decisions might be inferred because of the inability to justify some of the events within the data. Figure 8 illustrates an example of a descriptive dashboard and statistical parameters for proppant used in the United States over the past 8 years.



	2011	2012	2013	2014	2015	2016	2017	2018
Count	113	744	10718	14682	8820	5228	7569	3054
Avg	748.742	801.359	754.591	907.854	1138.8	1424.09	1659.05	1710.31
Median	758.282	657.286	660.15	821.215	1109.93	1419.41	1703.36	1797.17
Sum	84607.8	596211	8.08771e+06	1.33291e+07	1.00442e+07	7.44514e+06	1.25574e+07	5.2233e+06
Max	3143.38	4419.83	4748.88	4718.66	4665.62	4697.89	4759.15	4487.18
Min	12.4244	6.04181	3.90575	6.13146	7.9183	3.39789	13.0529	4.45045
StdDev	465.127	628.709	510.284	570.668	637.912	726.271	804.283	821.189
Q1	429.423	353.203	366.332	437.255	615.582	869.054	1014.87	1007.69
Q3	834.325	1112.9	1035.56	1244.32	1523.55	1875.99	2145.01	2300.41
IQR	404.902	759.693	669.231	807.061	907.967	1006.94	1130.14	1292.72
Outliers	6	29	185	206	96	48	30	5

Figure 8. Example of Descriptive Analyses using Dashboards and Statistics

7. PREDICTIVE DATA ANALYTICS

The oil and gas industry deals with many variables which are having complex relationships among each other. Like in all industries, investment in oil and gas involves a great deal of risk. In order to understand the intensity of those risks and to make informed decisions, the industry uses data. Data that are generated at every stage of the investment from seismic exploration and reservoir identification to drilling and production then refining operations and more. The bulk of data being generated and stored every second from multiple systems and sensors need to be analyzed to help in optimizing the operations and predict productivity based on the most influential parameters. Having the data without the proper tools for processing and modeling is just like have unrefined crude oil as it is not very useful without proper refining and processing to get to the useful end products. Predictive analytics provides an optimized prediction by investigating the relationships between the response parameters and gauge their influence on the predicted parameters. It converts raw data into mathematical models to generate actionable insights by combining the investigated domain deep knowledge with advanced data science analytics after filtering the noise and undesirable faulty data points. The objective behind utilizing predictive analytics is to improve the operations and reduce the cost and time.

Machine learning and intelligent systems are very important to solve problems that can't be modeled with analytical tools. The utilization of machine learning and intelligent systems can help to minimize NPT and cost. Examples of intelligence systems, including but not limited to, regression, fuzzy logic, genetic algorithms, probabilistic reasoning, recurrent neural networks (RNN), and evolutionary computing (Mohaghegh, 2000; Ertekin, 2005).

An example of that would be utilizing data to predict well loading before it occurs and take actions to eliminate the production shut down as a result to water level getting high in the well which exerts hydrostatic pressure in the well higher than the reservoir pressure that halts the production. Another example is to use machine learning approaches such as partial least square (PLS) method to predict the initial gas production in the Marcellus shale based on some completion and stimulation parameters such as lateral length, total proppant mass, and the concentration of other stimulation chemicals such as friction reducer, biocide, corrosion inhibitor, surfactant, and clay control. Figure 9 shows the prediction model versus actual initial gas production in Marcellus using PLS.

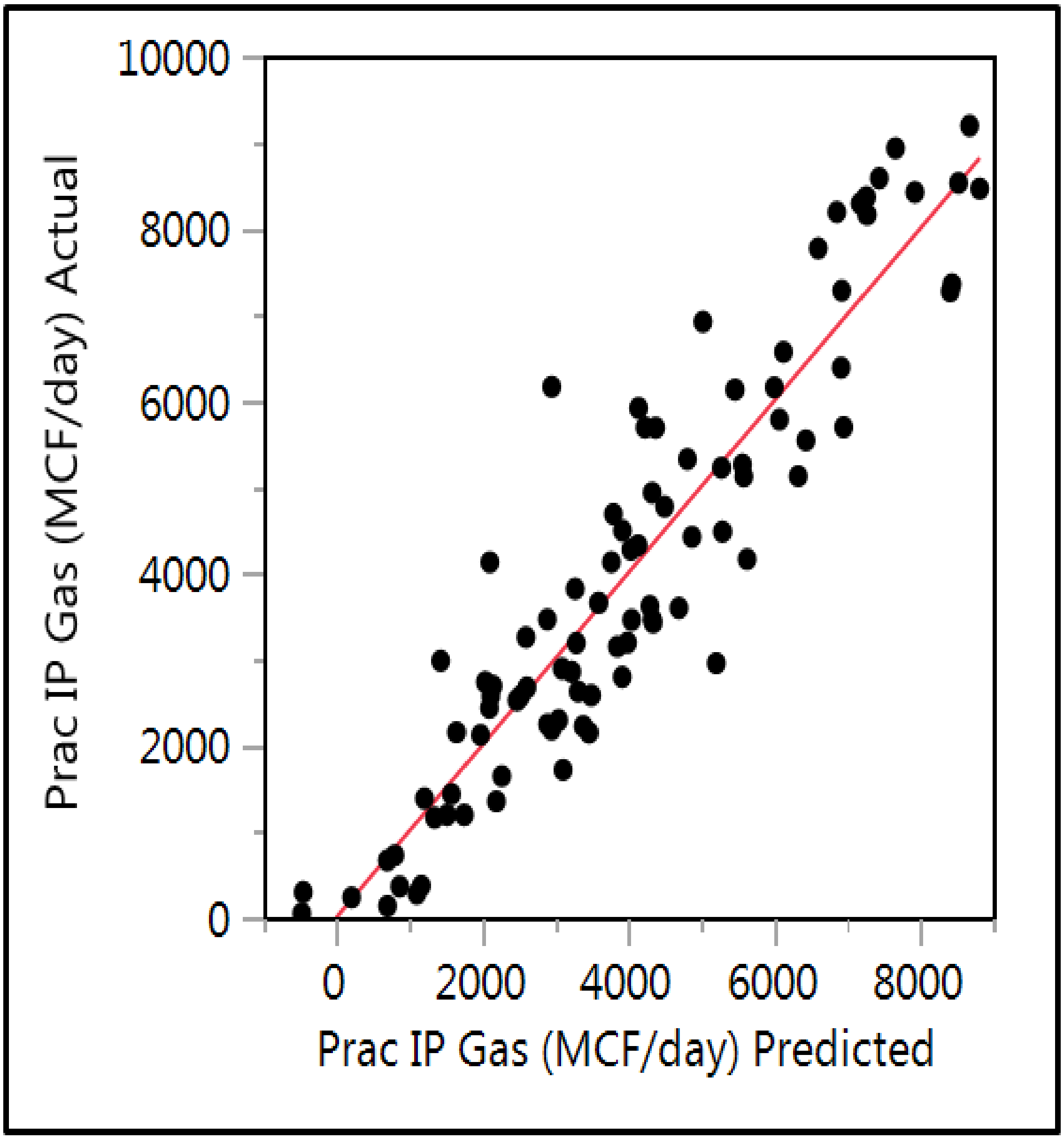


Figure 9. Predicted and Actual Initial Gas Production in the Marcellus

8. CONCLUSIONS

Data analytics has become important for the oil and gas industry due to the large data that are available in the exploration, drilling, production, and operations. Utilizing the available data will help to make better future decisions. However, there are many challenges in the process of collection, formatting, validation, managing, and analyzing the data that require close attention from the people who work on the data. The following conclusions were made based on this study:

- Big data analytics and the revolution of datafication helped companies and public administrations to better understand the data, find previously unnoticeable patterns, and provide better solutions for existing and future operations.
- There is a substantial transition in the oil and gas industry towards data-driven operations.
- Many operators have incorporated data science as part of their organizational structure and are training the next generation of engineers to be hybrid engineers that are expert in their area of specialization and data science.
- In all the E&P phases, the dramatic growth in data generation is not useful by itself. The ability to utilize and integrate the diverse data sources to seek useful insights and provide data-driven actions and decisions is the main target of using big data approaches in the oil and gas industry.
- To join two different databases, a common entity must exist between the two databases to be identified as the joining parameter.

- Datasets, especially oil and gas data are case-specific when it comes to data cleaning and transformation.
- Although the process of the data validation seems very ad-hoc process, there are many aspects of data validation remain valid and applicable across all the types of the datasets in oil and gas. It is also recommended that a walkthrough framework is implemented especially for oil and gas applications.

REFERENCES

Al-Hameedi AT, Alkinani HH, Dunn-Norman S, Flori RE, Hilgedick SA, Amer AS (2017a) Limiting Key Drilling Parameters to Avoid or Mitigate Mud Losses in the Hartha Formation, Rumaila Field, Iraq. J Pet Environ Biotechnol 8: 345. doi:10.4172/2157-7463.1000345

Al-Hameedi, A. T. T., Alkinani, H. H., Dunn-Norman, S., Flori, R. E., Hilgedick, S. A., Alkhamis, M. M., Alsaba, M. T. (2018, August 16). Predictive Data Mining Techniques for Mud Losses Mitigation. Society of Petroleum Engineers. doi:10.2118/192182-MS

Al-Hameedi, A. T., Dunn-Norman, S., Alkinani, H. H., Flori, R. E., & Hilgedick, S. A. (2017b, August 28). Limiting Drilling Parameters to Control Mud Losses in the Dammam Formation, South Rumaila Field, Iraq. American Rock Mechanics Association

Al-Hameedi, A.T., Dunn-Norman, S., Alkinani, H.H., Flori, R.E., and Hilgedick, S.A. (2017c). Limiting Drilling Parameters to Control Mud Losses in the Shuaiba Formation, South Rumaila Field, Iraq. Paper AADE-17- NTCE- 45 accepted, and it was presented at the 2017 AADE National Technical Conference and Exhibition held at the Hilton Houston

Alkinani, H. H., Al-Hameedi, A. T. T., Dunn-Norman, S., Flori, R. E., Hilgedick, S. A., Amer, A. S., & Alsaba, M. T. (2018c). Economic Evaluation and Uncertainty Assessment of Lost Circulation Treatments and Materials in the Hartha Formation, Southern Iraq. Society of Petroleum Engineers. doi:10.2118/192097-MS

Alkinani, H. H., Al-Hameedi, A. T., Flori, R. E., Dunn-Norman, S., Hilgedick, S. A., & Alsaba, M. T. (2018a). Updated Classification of Lost Circulation Treatments and Materials with an Integrated Analysis and their Applications. Society of Petroleum Engineers. doi:10.2118/190118-MS

Alkinani, H.H., Al-Hameedi, A.T.T., Dunn-Norman, S., Flori, R.E., Alsaba, M.T., Amer, A.S., and Hilgedick, S.A. 2019. "Using Data Mining to Stop or Mitigate Lost Circulation." Journal of Petroleum Science and Engineering, vol. 173 (February 2019), 1097–1108. https://doi.org/10.1016/j.petrol.2018.10.078

Alkinani, H.H., Al-Hameedi, A.T.T., Dunn-Norman, S., Flori, R.E., Hilgedick, S.A., Al-Maliki, M.A., Alshawi, Y.Q., Alsaba, M.T., and Amer, A.S. 2018b. "Examination of the Relationship Between Rate of Penetration and Mud Weight Based on Unconfined Compressive Strength of the Rock." Journal of King Saud University - Science. https://doi.org/10.1016/j.jksus.2018.07.020

Attar, A. A., Hughes, R. G., & Hassan, O. F. (2016). Handling Missing and Corrupted Data in Waterflood Surveillance, Using Reservoir Linear Characterization Models. Society of Petroleum Engineers. doi:10.2118/182207-MS

Beckwith, R. 2011. Managing Big Data: Cloud Computing and Co-Location Centers. J. Pet. Tech. 63 (10): 42–45

Breunig, Markus M, Hans-Peter Kriegel, Raymond T Ng, and Jorg Sander. 2000. LOF: identifying density-based local outliers. ACM Sigmod Record, Volume 29. ACM, 93–104

Dai, W., Cheng, X., Jiao, K., & Vigh, D. (2018). Least-squares reverse time migration with dynamic time warping. Society of Exploration Geophysicists

Di Zio, M., Fursova, N., Gelsema, T., Gießing, S., Guarnera, U., Petrauskienė, J., Quensel-von Kalben, L., Scanu, M., ten Bosch, K.O., van der Loo, M., and Walsdorfer, K. (2016). Methodology for data validation 1.0. Retrieved March 24, 2019

DrillingInfo, website link: https://info.drillinginfo.com/ , retrieved on March/20/2019

Ertekin, T. (2005). Virtual Intelligence - A Panacea or Hype for Long-standing Reservoir Engineering Issues. Society of Petroleum Engineers

Feblowitz, J. (2013). Analytics in Oil and Gas: The Big Deal About Big Data. Society of Petroleum Engineers. doi:10.2118/163717-MS

FracFocus, website link: https://fracfocus.org/, retrieved on March/01/2019

Franklin, S., Brodeur, M. 1997. "A practical application of a robust multivariate outlier detection method. Proceedings of the Survey Research Methods Section, American Statistical Association. 186–191

Hawkins, Douglas M. 1980. Identification of outliers. Volume 11. Springer

IHS Market, website link: https://ihsmarkit.com/index.html, retrieved on March/20/2019

Kirkman, T. W. (1992). Statistics to Use. Retrieved March 24, 2019, from http://www.physics.csbsju.edu/stats/

M.R. BruléGroup IBMS, The Data Reservoir : How Big Data Technologies Advance Data Management and Analytics in E & P Introduction – General Data Reservoir Concepts Data, Reservoir for E & P, 2015

Mehta, A. (2016). Tapping the Value From Big Data Analytics. Society of Petroleum Engineers. doi:10.2118/1216-0040-JPT

Mohaghegh, S. (2000). Virtual-Intelligence Applications in Petroleum Engineering: Part 1—Artificial Neural Networks. Society of Petroleum Engineers. doi:10.2118/58046-JPT

Nobakht, M., & Mattar, L. (2009). Diagnostics of Data Quality for Analysis of Production Data. Petroleum Society of Canada. doi:10.2118/2009-137

Simon A. (2013). Definition of validation levels and other related concepts v01307. Working document. Available from https://webgate.ec.europa.eu/fpfis/mwikis/essvalidserv/images/3/30/Eurostat__definition_validation_levels_and_other_related_concepts_v01307.doc

Texas Railroad Commission, website link: https://www.rrc.state.tx.us/, retrieved on March/02/2019

UNECE (2013). Glossary of terms on statistical data editing. Retrieved March 24, 2019, from http://www1.unece.org/stat/platform/display/kbase/Glossary

Vega-gorgojo, Guillermo & Fjellheim, Roar & Roman, Dumitru & Akerkar, Rajendra & Waaler, Arild. (2016). Big Data in the Oil & Gas Upstream Industry - A Case Study on the Norwegian Continental Shelf. Oil Gas European Magazine. 42. 67-7

SECTION

2. CONCLUSIONS

The main conclusions from this research is summarized in the followingpoints

- Big data analytics and the revolution of datafication helped companies and public administrations to better understand the data, find previously unnoticeable patterns, and provide better solutions for existing and future operations.
- FracFocus chemical registry is a valuable source of hydraulic fracturing fluid, chemical and proppant data for wells stimulated after 2012. The data are available for download to Excel. The raw data must be reformatted for practical use, which may require some fundamental understanding of hydraulic fracturing treatments.
- This work provides detailed methodologies to utilize and process FracFocus public chemical registry database of stimulated wells in the United States. The workflows include raw data manipulation, grouping chemicals, calculating and verifying wells mass percent treatments, and merging final FracFocus data with DrillingInfo.
- A workflow was developed to classify the types of the hydraulic fracturing fluid based on the chemical ingredients of the treatment fluid.
- Over time, the operators in the U.S. have increased the total proppant mass and water volume pumped per well as well as the proppant mass per stimulated foot and water volume per stimulated foot.
- When the oil and gas prices crashed in Q4 2014, the wells count dropped and the industry focus was shifted towards drilling more horizontal wells compared to

vertical drilling and the lateral length and stimulation volumes got bigger to increase the productivity while reducing the wells count.

- In the U.S., the average proppant mass increased over time from 2 million to over 10 million pounds by 2018. The same was true for the average water volume which ranged from 2 to 10 million gallons.
- The perforated lateral length has increased from 70% of the horizontal length to 90% by 2017. This indicates that more stages have been fractured along the lateral length to execute massive hydraulic fracturing treatments.
- In the Marcellus Shale Play, the completions vary as a function of hydraulic fracturing fluid used. The greatest average treatment fluid volume and proppant are associated with hybrid fluid treatments.
- In the Marcellus Shale wells, more than 50% of the stimulated wells were stimulated with water fracturing fluids
- In Cotton Valley formation, the completions vary as a function of hydraulic fracturing fluid and proppant used. The greatest average treatment fluid volume and proppant are associated with cross-linked gel fluid treatments.
- The wells that were appropriately stimulated with larger amounts of proppants per foot are the wells who produced the best in the Permian (Midland) Basin.
- In the Permian Basin, comparing the productivity per foot of different lateral lengths showed that the shorter lateral wells produced better BOE/foot than the long lateral wells which indicates that the longer laterals need to be stimulated with higher amount of pumped proppants.

3. RECOMMENDATIONS AND FUTURE WORK

The developed databse requires continuous updating as FracFocus continues to add new stimulation reports every month. The database can also be improved by integrating it with other commercial databases such as HIS in addition to DrillingInfo.

More studies are planned to be excuted in the near future to assist the Health Effect Institute with a study to track the confidential ingredients in hydraulic fracturing chemical components.

More journal papers are planned to be published that describes the stimulation chemicals and classify them based on the purpose for use and the CAS#.

Another study is intended to be published which will discuss the pad drilling surface density distribution as well as the wells proximity downhole and the drainage radius.

REFERENCES

Joe Leimkuhler, S. a. (2012). Unconventional Resources. Retrieved 9 21, 2018, from https://www.spe.org/en/print-article/?art=619

U.S. Energy Information Administration, retrieved on Sep/03/201, from https://www.eia.gov/petroleum/production/#oil-tab

Government of Western Australia Department of Mines, Industry Regulation and Safety, Hydraulic Fracturing Estimation, retrieved on Sep/03/201, from https://www.dmp.wa.gov.au/Petroleum/Hydraulic-fracture-stimulation-20018.aspx

VITA

Mustafa Adil Al-Alwani was born and grow up in Iraq. He received his bachelor degree in petroleum engineering from Baghdad University, Baghdad, Iraq in 2010, he was the valedictorian of the petroleum engineering class 2006-2010 and graduated with a perfect 4.0 GPA.

After graduation, Mustafa joined BP-SOC, and worked as a night drilling supervisor then he was promoted to be a day drilling supervisor in Rumaila oil field. In January 2013, he was awarded Fulbright Scholarship to study Master's degree in petroleum engineering funded by the U.S. Department of States (2012-2014). He joined Missouri University of Science and Technology in Spring 2013 to work under the supervision of Dr. Shari Dunn-Norman as a graduate research and teaching assistant. He graduated with Master of Science degree in petroleum engineering in December 2014 with 4.0 GPA.

In 2015, Mustafa worked with Halliburton as a drilling supervisor in an ExxonMobil operated field then joined BP-SOC in 2016 to work as a wellsite leader south of Iraq. As part of his civic duties, and to honor his Fulbright commitment he volunteered to work in Alumni University program funded by the U.S. Embassy in Baghdad in cooperation with University of Cincinnati to serve as a Fulbright alumni trainer.

In Spring 2017, Mustafa started his PhD program. He worked with Dr. Shari Dunn-Norman and Professor Larry Britt as a research and teaching assistant. In Spring 2019, Mustafa was the recipient of the College of Engineering and Computing Dean's Graduate Educator Award. He received his Ph.D. degree in petroleum engineering from Missouri University of Science and Technology in December 2019.