UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

CASE STUDIES IN CAUSAL INFERENCE AND ANOMALY DETECTION

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

ROHAN SAKHARDANDE
Norman, Oklahoma
2020

CASE STUDIES IN CAUSAL INFERENCE AND ANOMALY DETECTION

A THESIS APPROVED FOR THE
MEWBOURNE SCHOOL OF PETROLEUM AND GEOLOGICAL ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Deepak Devegowda, Chair

Dr. Chandra Rai

Dr. Hamidreza Karami

Acknowledgements

This work would not have been possible without the help and support of many individuals. I take this opportunity first to thank my advisor Dr. Deepak Devegowda for his constructive feedback, insights and leadership. Dr. Deepak has not only taught me Reservoir Engineering and Data Analytics but also has been a great mentor to me throughout my study at the University of Oklahoma. His suggestions and guidance from time to time have been instrumental to this work. I am also very grateful to my committee members Dr. Chandra Rai and Dr. Hamidreza Karami for their insights, feedback and valuable time.

I express my gratitude to all members in our IC^3 research group especially Dr. Carl Sondergeld and Mikki Langevin for their feedback during this work. Also, I am thankful to all the staff members in the Petroleum Engineering department including Danika Hines, Sonia Grant and Haley Pearson for their help throughout this journey.

Finally, I thank my wife, parents and friends in Norman, Oklahoma for their constant support and unconditional love.

Table of Contents

List of Tables

List of Figures

Appendix

Abstract

The study of parent-child well interactions in unconventional shales has generated high interest both in the industry and academia over the last decade. This is largely because of the growing number of child wells and their immediate impact on the parent well production owing to several dynamic factors, one of them, including well spacing. Evaluating the impact of well spacing on parent and child well production performance is challenging. Several studies have resorted to geomechanical stress and fracture modeling combined with dynamic simulation techniques while a few operators have chosen field trials to evaluate optimal well spacing. Several data-driven approaches to address the well-spacing problem have also become popular. One such commonly used data-driven approach simply calculates the difference in cumulative production over a specified period of time for parent and child wells grouped by spacing. This approach has been the method of choice for several different recent analyses of well spacing; however, given that the method of simple averages does not account for formation properties or completion design, the results may be compromised and can lead to counterintuitive results.

In this thesis, I introduce a new data-driven approach leveraging the power of causal inference as seen in clinical trials for multivariate observational studies. The causal approach addresses the problem behind the routinely used simple averages approach by providing a formalism to control for reservoir and completion variables when evaluating the impact of well spacing on production performance. I apply the causal inference workflow to a dataset from a prolific oil basin in Texas with over 700 wells in the analyses. It includes the formation properties, fluid volume, proppant weight, landing zones and the downhole locations of the wells. Using the causal inference workflow, I evaluate the effect of well spacing on well performance at different parent-child spacing ranges. The optimal well spacing is then estimated for this shale play based on the magnitude of the causal effects. These estimates are then compared with the simple averages approach to demonstrate the power and utility of causality.

In the second part of the thesis, I transition into a discussion on anomaly detection approaches applied in the oil and gas industry. I discuss current anomaly detection methods for a widely used artificial lift method – the Sucker Rod Pump (SRP). Today, there is a growing need for fast and accurate anomaly detection systems given the emergence of Internet of Things (IoT) and access to Big Data. Anomaly detection using human operators can be expensive, is often subject to bias and experience-levels and does not scale very well with the need to monitor more than a few tens of wells. With SRPs, the problem of anomaly detection becomes a problem of image classification where dynamometer cards are evaluated for signatures of failure. While this has been the mainstay of anomaly detection for pumpjacks, in this thesis, I automate this task of monitoring and detecting the anomalies from the SRP pump cards.

Several thousand synthetic pump cards specific to pump failures modes are generated from the literature and fed to a deep learning model. This deep learning model is a Convolutional Neural Network (CNN) which is commonly used in image classification tasks, speech recognition tasks as well as many other modern-day technology applications including smart phones, self-driving cars, aerospace etc. The CNN used in this work offers a very high accuracy for detecting a variety of pump failures modes thereby offering the potential to save costs, time and unnecessary workovers for the operator.

Chapter 1: Introduction

The cumulative US production per day from shale plays as of August 2020 is approximately 7 MMbbl of crude oil per day and 68 Bcf of gas per day (EIA 2020). The number of wells drilled and completed in the US since 2014 is over 100,000 and continues to rise despite the recent downturn and a global health pandemic clouding the industry's near future, signifying the importance of unconventional shales in securing the nation's energy supply.

The success of shale plays owes a lot to advancements in horizontal drilling and hydraulic fracturing. Shales are typically tight rock formations with very low permeability and porosity. In order to extract hydrocarbons from these formations, wells are first drilled vertically up to a certain depth and then deviated to drill a horizontal section which is then perforated in stages from the toe to the heel along the lateral length. Fluid mixtures mainly comprising of water, sand and chemical additives are then pumped under high pressure to create additional fractures in the formation. The sand or the proppant keep the fractures open, allowing oil and natural gas trapped inside the tight rocks to flow into wellbore.

As oil and gas operators continue to exploit these hydrocarbon rich resources through horizontal drilling and hydraulic fracturing, the density of the wells per acre also increases with time. The wells which are drilled, completed and put on production following lease acquisition are called the "Parent" wells. After a period of time, the operator drills more wells in order to continue to legally hold the lease economically by production. These newer infill wells are called the "Child" wells. Of course, with time factored in, some Child wells can become Parent wells to other newer infill wells.

Miller et al. (2016) and Xu et al. (2019) compare the number of parent and child wells brought on production between years 2005 to 2018 for 5 different unconventional basins. The trends observed in **Fig. 1.1** obtained from their analysis show a definite increase in the number of child wells with respect to the number of parent wells. A larger number of child wells compared to the number of parent wells indicates decreasing well spacing with time. Well spacing cannot decrease indefinitely. This leads to the question of optimal spacing for maximizing recovery and what are the optimal number of wells per section.

The answer to the above question is certainly not easy. Closely spaced wells tend to recover more hydrocarbons but suffer from well interference at some point in the future. If the well spacing is too large, operators end up with stranded resources that will require additional infill wells at some point in the future. Therefore, if the goal is to maximize the oil recovery, then it is imperative that we find the optimal well spacing.

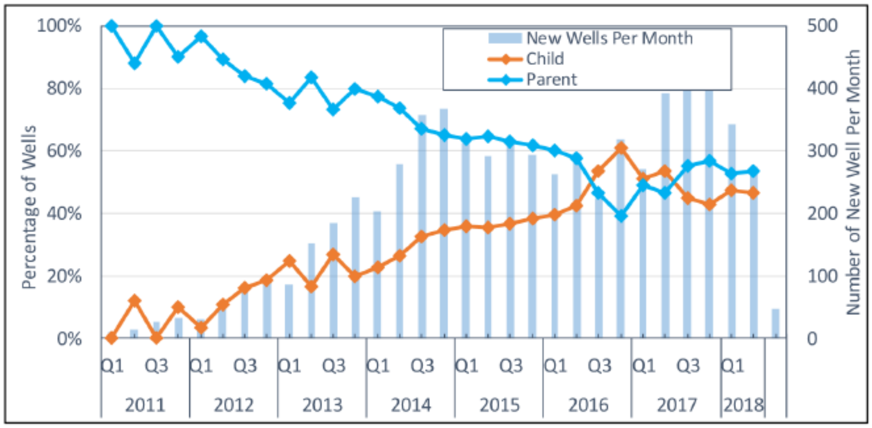


Fig. 1. 1—Percentage of child and parent wells between years 2011 and 2018 (Xu et al. 2019).

There are multiple ways to determine optimal well spacing. Studies which consider well spacing optimization can largely be classified into two groups. The first group is the physics-driven approach. This approach is the classical approach combining fracture modeling integrated with dynamic reservoir simulation modeling. The second group is the data-driven approach. This approach has seen more applications in recent times as operators hope to leverage the information content of the data they have acquired. Powerful machine learning and artificial intelligence algorithms have now become a part of this data-driven approach. In the next few paragraphs, I provide a comprehensive literature review on well-to-well spacing considerations using both approaches as mentioned above.

Well spacing considerations are closely tied to well-to-well interference. Closely spaced wells suffer increased pressure and production interference. Ajani and Kelkar (2012) were among the first to assess production interference in shale wells. They use a data-driven approach to analyze fracture and initial 60-day production normalized to lateral length for gas wells in the Woodford shale play. They quantify the production interference seen in the parent wells as newer child wells are put on production. If there is any change in the production trends for the parent well as a result of addition of new child wells within a certain distance from the parent, then that is considered as interference impact. This is done through manual inspection of production data changes for the parent well when the child well is brought on production. In general, they observe that child wells closer to the parent well have a significant negative impact on parent well performance compared to wells placed farther away. The negative impact was also seen to be proportional to the age of the parent well because of higher levels of depletion. Since there may be multiple impacts on parent wells as more child wells come online, the quantification of optimal spacing becomes very difficult using this manual inspection method.

Well interference and frac-hits are both technically different phenomena but are used interchangeably for shale wells. Well interference is seen when there is an overlap between the drainage volume of the wells through the matrix but there is no overlap between the fracture networks of the wells. The term frac-hit or fracture driven interaction (FDI) is a parent-child intra well communication when the child well is hydraulically fractured in the vicinity of a pre-existing parent well. Frac-hits usually occur when there is an intersection between the fracture-networks of the parent-child wells which is sometimes aided through pre-existing natural fracture systems between the wells (Daneshy and King 2019; Yu et al. 2017). Sometimes even if the new child well is placed at more than 3000 ft apart from the parent, there may be production interference. In this case, the parent-child wells are primarily connected through the pre-existing natural fractures or faults and not the hydraulic generated fractures (Yaich et al. 2014).

The above-mentioned forms of inter-well communication through hydraulic fractures, matrix and natural fractures systems in unconventional shales are shown through a simple cartoon in **Fig. 1.2** (Gupta et al. 2020). It is important to note that there could be combinations of two or all three forms of inter-well communication occurring at the same time. Gupta et al. (2020) provide a detailed literature review of factors controlling frac-hits in unconventional shale plays. They divide these factors into controllable and uncontrollable factors from an operator's perspective. Uncontrollable factors are related to geological features, in situ stress environment, petrophysical and geomechanical properties. Controllable factors include choice of well placement or well spacing, selection of completion design parameters and control over the well depletion.

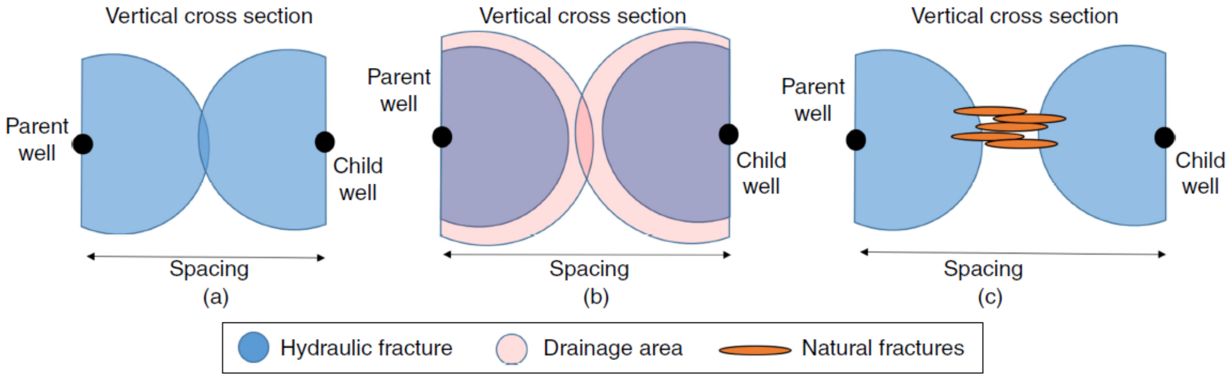


Fig. 1. 2—(a) Inter-well communication in shales through hydraulic fracture network (b) Inter-well communication in shales through the matrix with no connected hydraulic fractures (c) Inter-well communication between parent and child wells through a natural fracture system (Gupta et al. 2020).

In another study, Esquivel and Blasingame (2017) study frac-hit events in the Haynesville shale. They report that parent well depletion creates pressure sinks or pressure depleted zones that are identified as a cause as well as a necessary condition for a frac-hit to occur. Typically, during the hydraulic fracture treatment of the child well, the parent well is shut down. The authors show that the after the parent well shut down period, child to parent well frac-hits are correlated with the inter-well spacing, in-situ principal stress re-orientations and completion design size (proppant per foot, frac fluid per foot). **Fig. 1.3** illustrates a case of a frac-hit between a parent well and child well in Haynesville shale. In this case, the parent well production is boosted after the child well is completed.

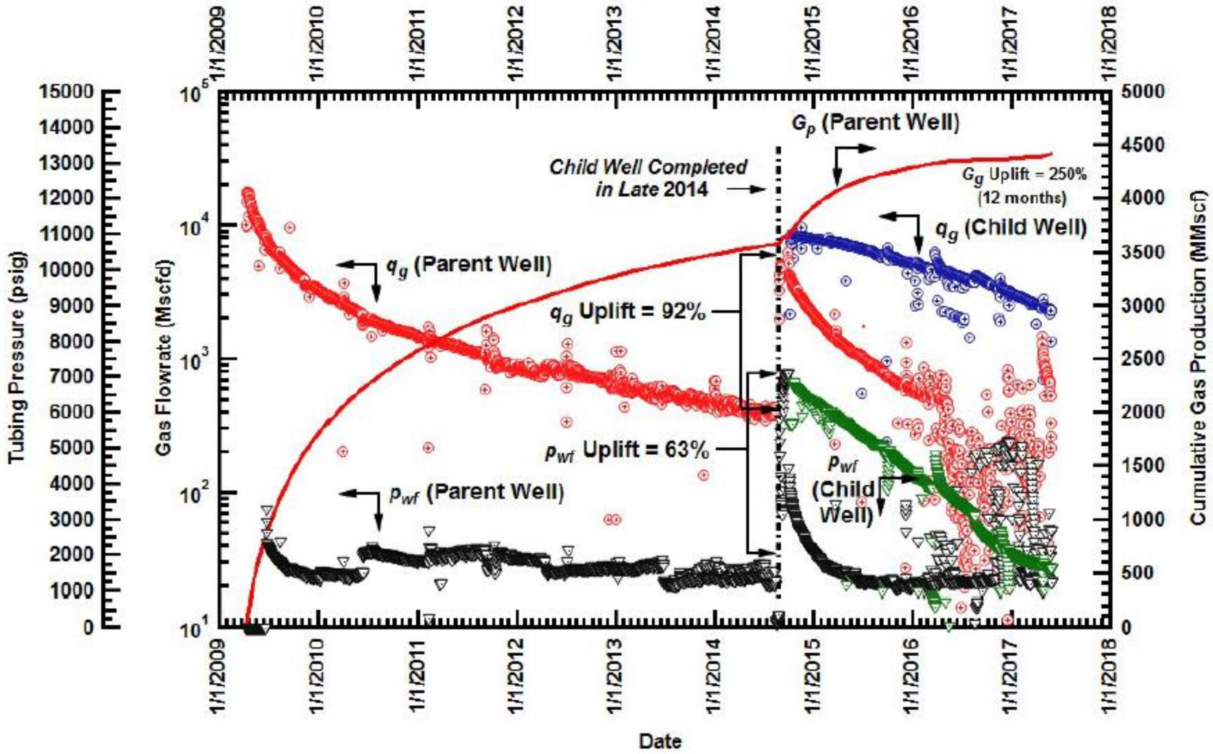
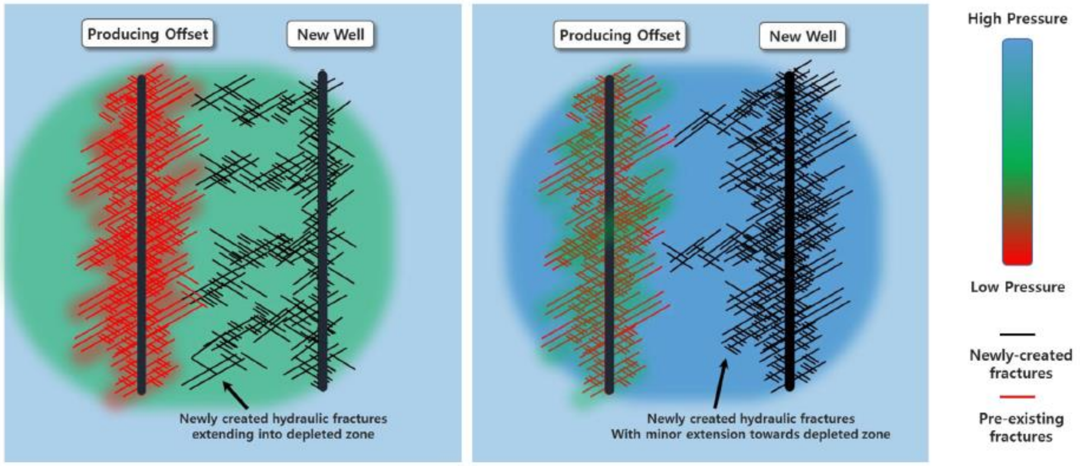


Fig. 1. 3—Example showing the effect of a positive frac-hit (dotted black line) in Haynesville Shale. The parent well gets an uplift in its gas production rate (shown in red) after the child well gets completed in late 2014 (Esquivel and Blasingame 2017).

There have been other approaches to diagnose frac-hits. Sani et al. (2015) and Schertz et al. (2019) interpret chemical tracer data to diagnose fracture propagation and analyze frac-hits in two different unconventional shale plays. Schertz et al. (2019) report that depletion plays an important role in fracture driven interactions between the parent and child wells. Delays in bringing child wells on production are reported to increase the chances of fracture communication between the wells. This can potentially lead to asymmetric fracture growth from the child well towards the parent well. The above observations are corroborated by Xu et al. (2019) who show that the child well placed nearest to a parent well underperforms. The production performance of the remaining child wells positioned farther away from the parent is seen to be less compromised. They also report that the time period between start of production for the parent and the child well is critical and longer intervals worsens child well performance. This again is due to the asymmetric fracture growth from the child well towards the parent wells into its depleted drainage area. Kumar et al. (2020) report similar observations in their fracture and geomechanics simulation study stating that the child well fracture asymmetry is directly proportional to reservoir diffusivity.

Portis et al. (2013) provide a case study from the Eagle Ford where they test well spacing strategies using chemical and radioactive tracers along with pressure interference tests for 3 single pad child wells spaced 500 ft from each other with a parent well located 1500 ft away from the nearest child well. They report that field tests coupled with seismic and well log data can provide a better understanding of complex fracture geometry and pressure communication between the wells. Use of chemical tracers and pressure monitoring throughout the field development cycle give an idea about the fracture conductivity, pressure communication and mobility of fluids transported across large distances through the reservoir. All this information along with production responses aid in reaching the optimal well spacing.

Kurtoglu and Salman (2015) apply decline curve analysis (DCA) to get the difference in estimated ultimate recovery (EUR) before and after well-to-well frac interference. The authors mention that geological features like pre-existing natural fractures, heterogeneity in reservoir properties and size of completion designs play an important role in this interference. High internal energy reservoirs possess a high GOR and initial reservoir pressure therefore any interference due to fracture hits for parent-child wells will likely be positive. However, the reservoir which has low internal energy and is marked through a low GOR and initial pore pressure, any interference due to fracture hits is most likely to be negative. This phenomenon is also seen in **Fig. 1.4a.** where the older parent well on left is depleted with low internal reservoir energy. The child well fractures on the right fractures grow in a preferentially asymmetric manner towards the pressure depleted region of the parent well on the left resulting in a negative frac-hit. This case is worsened by closely spaced wells and bigger child well completion sizes.



(a) **(b)**

Fig. 1. 4—(a) Negative frac-hit from the child well on right with fractures growing in preferential manner towards the pressure depleted region of the parent well on left (b) Positive frac-hit between the child well on right and the parent well on left. Here the parent well has higher reservoir pressure and has not depleted as much in comparison (Kurtoglu and Salman 2015).

Similar behavior is observed in a field study of wells in the Midland basin by McDowell et al. (2019). According to their study, horizontal parent wells experience higher fracture driven interference than vertical parent wells. They divide the wells into five categories based on fracture driven interaction (FDI) intensity and perform a qualitative analysis of the frac-hits in the study area. Also, they find a negative correlation between the intensity of the fracture driven interference and a 2-dimensional well spacing. The parent wells received a frac-hit after the child well was hydraulically fractured for all well spacings at less than 1000 ft. Water hits in general were observed to be stronger for tighter well spacing configurations. Cao et al. (2017), however, report that FDIs directly indicate the required well spacing configuration. If the operator does not see any frac-hits, the wells are spaced farther apart than optimal leaving behind unrecovered hydrocarbon. As indicated in the data analyses conducted by Miller et al. (2016) and Xu et al. (2019), frac-hits can create positive or negative impacts for the parent wells. However, it is seen that the child wells usually perform poorly whenever there are frac-hits.

Ajisafe et al. (2017) also study the impact of well spacings (660 ft and 1320 ft) on production interference in parent-child wells for the Avalon shale in the Delaware Basin. They build a discrete fracture network (DFN) model integrated with seismic data and geophysical logs. At the 600 ft well spacing scenario, the parent well after 1 year of depletion develops a pressure sink around the wellbore. This pressure sink then alters the stress regime between wells and helps the fractures from the child well navigate towards the parent well. Ma and Zoback (2016) report that this is because of the in-situ poroelastic stress changes between the parent and child wells induced due to depletion of the parent well. This in turn had a negative impact on the child well production performance at the 660 ft spacing. Daneshy (2018) explains this in terms of reduction in the magnitude of the principal stresses for the parent well due to production which then attracts the hydraulic fractures from the child wells. The magnitude of principal stresses on the two sides of the child lateral are not equal and tend to slowly decrease in the direction of the parent well and increase away from it (Daneshy 2018).

In another work by Gupta et al. (2020), a data-driven approach is used to quantify the impact (positive or negative) of frac-hits on the affected parent-child wells in the Woodford, Wolfcamp and Meramec formations. Parent wells are shut down prior to the child well hydraulic fracturing. The authors identify frac-hit wells that are characterized by a specific signature in oil, gas and water production rates post shut-down period of the parent well. This signature is marked by an increase in the gas-oil ratio (GOR), initial decrease in oil production rate and high water cut percentage. The difference in production as a function of time between the estimated decline pre- and post frac-hits is quantified as the overall impact of the frac-hit. Also, based on predictive multilinear regression models for EUR regressed on several production data and completion design parameters, they propose a recommendation chart for minimum well spacing with zero impact on EUR for a given completion size and a cumulative production of the parent well.

Rafiee and Grover (2017) also study the well spacing problem in the Eagle Ford shale. Their study combines fracture modeling, dynamic reservoir simulation and data analytics. They report that wells spaced at a wider distance with a sparse cluster spacing are characterized by a 30 percent more EUR compared to a densely spaced configuration.

Chu et al. (2020) describe the use of downhole pressure gauges which measure well bottomhole pressures continuously throughout the life of the well. At different well spacings, these bottomhole pressure measurements are then used to quantify the measured pressure interference (MPI). MPI is derived from the Chow Pressure Group (CPG) (Engle and Buswell 1952). Their methodology can be also used as a guide to determine the optimal well spacing needed. They also observed that tighter spacings yielded stronger MPI values between the wells. As many have observed, stronger interference between the wells lead to temporary or permanent loss in production (King et al. 2017).

Along with different approaches discussed so far for finding optimal spacing, there are a few mitigation procedures also suggested in the literature for minimizing losses when the spacing is not optimal. Whitfield et al. (2018) through field trials discuss the success of pre-loading or re-pressurizing the parent well through water injection prior to and during the hydraulic fracture job at an adjacent well. Another option is refracturing the parent well to increase the magnitude of the surrounding principal stresses. Manchanda et al. (2018) report that refracturing the parent well alters the stress regime around it inhibiting the asymmetric fracture growth from the child well to the parent. Sani et al. (2015) suggest the use of a "stress barrier" to stimulate the child well such that it does not lose its deliverability to a nearby parent well. Another proven field trial conducted by Zhang et al. (2020) pumps far-field diverters during the child well hydraulic fracture treatment. These field diverters as their name suggests divert the fluids at the fracture tips. An application of the far-field diverter is illustrated in **Fig. 1.5**. This stops the flow of proppant and hydraulic fracturing fluid heading in the direction of the preferential planes of connected fractures thus offering less resistance near the parent well depleted zones (Esquivel and Blasingame 2017).

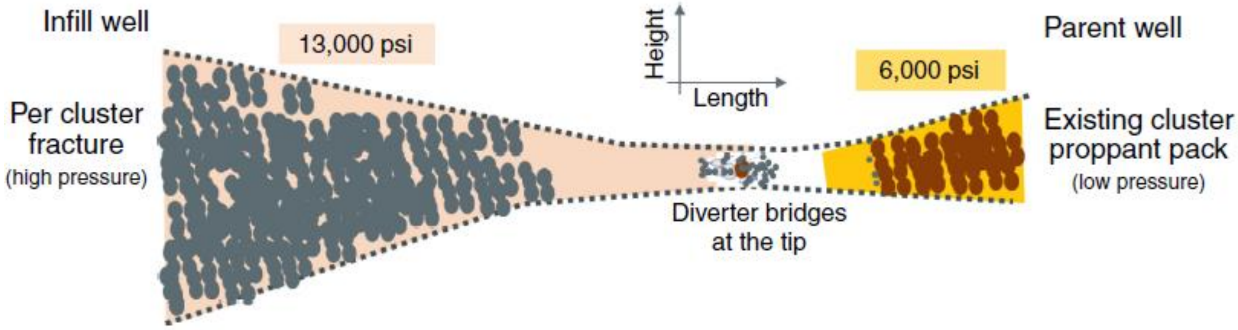


Fig. 1. 5—Example showing the application of far-field diverters bridging the gap between a parent and child hydraulic fracture. This is used for preventing a frac-hit (Zhang et al. 2020).

With proper field development and planning, frac-hits can be prevented ahead of time. For this to happen, one would need to optimize the controllable factors like completion design size, schedule of drilling, fracture treatments, production of child wells and more importantly inter-well spacing. Almost all the physics-driven modelling approaches to solve the well spacing problem are associated with a certain degree of uncertainty for each of modeling parameters. Modelling inputs for permeability, fracture half length, number of fractures, fracture conductivity, cluster spacing etc. have a direct influence on the results. Field well spacing trials offer a more reliable solution for optimizing well spacing, but these methods are capital intensive and may not be always practical given the current market conditions and economy (Yaich et al. 2014).

On the other hand, data-driven approaches are gaining increasing acceptance. Access to reliable and good quality field data can allow meaningful interpretation. However, the traditional and routine method of comparing simple averages of cumulative normalized well production numbers for parent wells and child wells, while useful for a preliminary analysis, can provide inaccurate results because there is little to no consideration for the year of completion, completion design and reservoir and petrophysical properties.

In this thesis, I describe a more sophisticated data-driven approach using causal inference applied to solving well spacing problems. To the best of my knowledge, this is the first application of formal causal inference in petroleum engineering and the geosciences.

This study evaluates the impact of well spacing on 180-day and 360-day cumulative oil production normalized to lateral length (bbl/ft) for four different ranges of parent-child well spacing. By controlling for the parent-child well completion and formation properties and establishing a statistical balance between the wells, the causal impact of well spacing on production is estimated. This effect is compared with the routinely used simple averages approach. For all the four spacing ranges considered, the causal effect quantifies the overall production loss per lateral length between the parent and the child wells. Based on the magnitude of this effect, I determine an optimal well spacing strategy for the parent and child wells.

For the second portion of this thesis, I introduce the reader to the concept of a process anomaly, anomaly detection and some of its applications using different methods in the rapidly developing field of data analytics. Although the anomaly detection algorithms are widely used in aerospace, healthcare, manufacturing and finance related applications, most of the applications as seen in the literature that are discussed in the thesis are viewed from the perspective of the oil and gas industry.

Any complex process, irrespective of the industry, with measurable parameters over a time period is subject to certain specification limits beyond which the process goes out of control. The data points which lie outside the process specification limits are called anomalies. The mechanism to find such process anomalies or data patterns which tend to deviate from the normal or standard expected value is called as Anomaly Detection. Sometimes an anomaly may have an overall positive impact but most often, it has negative connotations which are indicators of an impending process failure. Therefore, many researchers in industry and academia have utilized the power of anomaly detection algorithms to identify these abnormalities to save potential downtime, enhance safety and increase productivity.

The use of anomaly detection algorithms spans upstream, midstream and downstream oil and gas operations. One of the biggest benefits of automated anomaly detection is the elimination of the need for continuous human monitoring. However, a process expert is definitely warranted to ensure that the detected anomalies are not merely false positives.

In this section, I discuss a few algorithms used for anomaly detection. One of the most promising approaches is detecting change points in time-series data. Change points can be defined as abrupt variations in a time series of sequential data (Adams and MacKay 2007). These change points are anomalies if they represent a deviation from the normal patterns. Truong et al. (2020) provide an extensive review of offline change point detection algorithms for multivariate time series data. The methods discussed are mainly applicable to offline change detection. Adams and MacKay (2007) introduce a Bayesian method for online change detection which is an improvement from other retrospective Bayesian change detections proposed earlier. This online change detection method can be used on a variety of time series data observed in any oil and gas industry applications. One such example of this method is shown in **Fig. 1.6** where change points are detected in the nuclear magnetic response of a well during drilling.

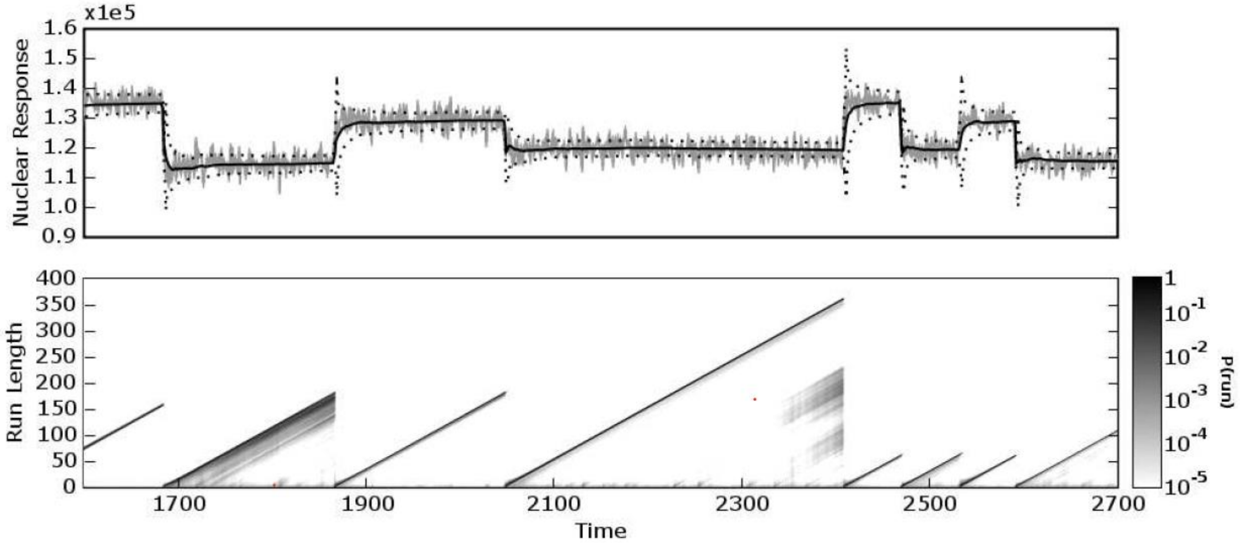


Fig. 1. 6—(Top) Change point detection applied to a nuclear magnetic response during drilling of a well. (Bottom) Linear increase in run length and its reset to zero indicates the time until a change point occurs. Darker regions show high probability of a change (Adams and MacKay 2007).

In machine learning, there are two main methods of anomaly detection known as Supervised and Unsupervised classification. The Supervised method is based on labelling univariate or multivariate observations with normal and abnormal tags. This method necessitates careful labelling of normal operations and dysfunctional modes by the subject matter expert. The training process requires huge amounts of data and therefore supervised methods can be time-consuming and laborious. Some examples of popular supervised machine learning methods for anomaly detection are Decision Trees, Support Vector Machines, Random Forest, Linear Regression and Ensemble methods (Liu et al. 2013; Pennel et al. 2018; Boguslawski et al. 2018).

The unsupervised method works with unlabeled data. Anomalies are reported for data points that are far away using some distance-metric from the majority of the points. However, this method still necessitates human intervention to avoid reporting false positives. Some examples of such unsupervised methods for anomaly detection are K-means clustering, Isolation Forests and Principal Component Analysis (Gupta et al. 2016; Cadei et al. 2020).

Deep learning methods are extensions of machine learning algorithms which are designed to discover complex data structures or patterns which can then be used for anomaly detection. Deep learning algorithms have proven to be popular for anomaly detection in the oil industry (Yan and Yu 2015). Here, I discuss some of recent applications of anomaly detection specific to the oil and gas industry.

An application of anomaly detection for power plants flags abnormal operating modes for gas turbines. Yan and Yu (2015) build an unsupervised deep learning model designed to detect dysfunction in the gas combustion process using a stacked denoising autoencoder (SDAE) for feature learning and a Neural Network classifier called Extreme Learning Machine (ELM) for detecting anomalies.

Marti et al. (2015) discuss anomaly detection for oil platform turbomachines used at the rig site. High speed turbomachines which generate power near the well head are equipped with multiple sensors to monitor performance. The authors also describe this as a big data problem because the sensors provide more than 43 million measurements daily. They use a low computational segmentation method along with an unsupervised one-class super vector machine classifier (SVM) to detect anomalies in the turbomachines. One-class SVM classifier assumes that all the training data belongs to one label or class. The model uses the data to learn a boundary such that the normal data points are separated from the abnormal ones. Any data point which lies outside the boundary is considered an anomaly (Marti et al. 2015).

Sidahmed and Bailey (2016) develop an anomaly detection approach based on sliding windows to detect sand entry events into the wellbore. They take acoustic signal measurements and reconstruct them into a new signal using an unsupervised deep-learning model. The reconstructed signal error rate is then used with a selected threshold to detect anomalous events in the system. They also examine these events using a probabilistic Finite Gaussian Mixture model for clustering the observations into normal and abnormal events. Romanenkova et al. (2019) use a combination of machine learning and change point detection algorithm to detect rock type changes in North Western Siberian oil wells. They build a supervised classifier using Gradient Boosting Decision Trees to predict probabilities for the rock type corresponding to each drill bit position. This information is fed to a change point detection algorithm to detect changes in 2 rock types – the productive sand zone and non-productive shale zone.

Gurina et al. (2019) demonstrate an anomaly detection method applied to directional drilling data accidents like stuck pipe, drill string wash outs, mud loss and fluid shows. They compare measuring while drilling (MWD) data in a 2-hour time windows of a test well with another 2-hour time window of a pre-existing well from a database of MWD drilling accidents history records. For each time window, sample statistics are calculated. These aggregated statistics are then fed to a gradient boosting classification model which computes a similarity score for the two classes of wells. If the similarity score is higher than a certain assigned threshold value, then it is classified as a drilling accident.

Kuesters et al. (2020) apply another change point detection algorithm for drill string washout prevention. Typically, a pressure change is observed when there is a drill string washout. If the washout condition is not identified early, then there could be potential twist-off in the drill string. Therefore, detection of the start to a washout is necessary so the drill string can be retrieved without severe damage. The variables used here were pump pressure and flow rate. In this study, the authors successfully used a change point algorithm based on the difference between the Z scores. First Z score is computed for a smaller window of data points leading up to the change and a second Z score is computed for a wider window of data points leading up to the change point. If the difference in the two Z scores goes beyond a fixed threshold, then it is regarded as an anomaly.

Saghir et al. (2019) use time series data from Coal Seam wells and convert it into images to gauge abnormal Progressive Cavity Pump behaviors due to production of coal fines. This is accomplished through a Symbolic Aggregation Approximation (SAX) applied to the PCP time series for flow, speed and torque data. Conversion in SAX is the generation of symbols corresponding to the time series followed by transformation of the symbols to heat maps. Each heat map corresponding to each variable (flow, speed and torque) is converted into a multivariate heat map with a 1 day-time window. A change in the shape and color of heat map corresponds to an anomaly for the PCP. An example of SAX heat maps is shown in **Fig. 1.7** (Saghir et al. 2019). They further use K-mean clustering to cluster and label the anomaly heat maps for each time series variable considered. However, they could have easily accomplished anomaly detection using K-means directly. The motivation for using SAX is not presented.

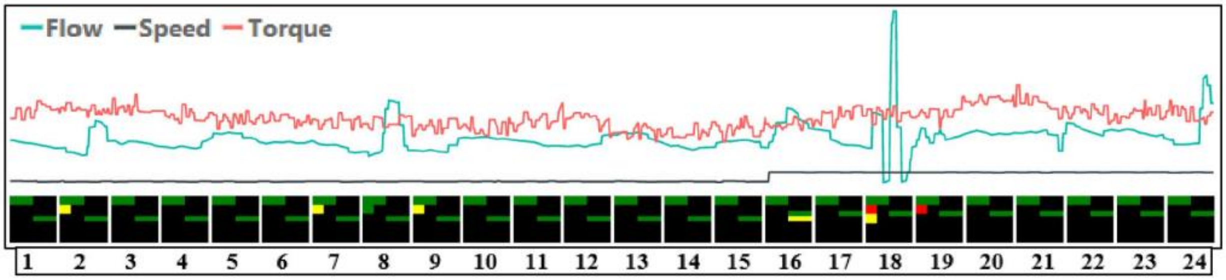


Fig. 1. 7—Example of a multivariate heat map for PCP flow, speed and torque plotted over a period of 24 hours. Heat maps are shown for each 1-hour time window (Saghir et al. 2019).

For an oil and gas central processing facility which serves five production lines of multiphase flow, Cadei et al. (2020) develop an anomaly detection method using plant sensor data. **Fig. 1.8** shows the architecture of their digital oil field connecting field sensors to the data center for monitoring and reporting. Plant sensors are flowmeters, pressure and temperature transmitters fed to an unsupervised Isolation forest algorithm and a trained Ridge regression model to predict anomalies in the future. Akinsete and Oshingbesan (2019) detect natural gas pipeline leaks using 5 different models including Random Forests, Decision Trees, Support Vector Machines, Gradient Boosting and Artificial Neural Networks. Each of the model predicts the flow rate based on operational parameters. The authors define a leak index formula which is dependent on the number of flowrate residuals. There is a pipe leak if the leak index is over a predefined threshold based on Mean Absolute Errors from the model. They observe that Random Forests and Decision Trees models are the most sensitive to detect leaks in pipelines.

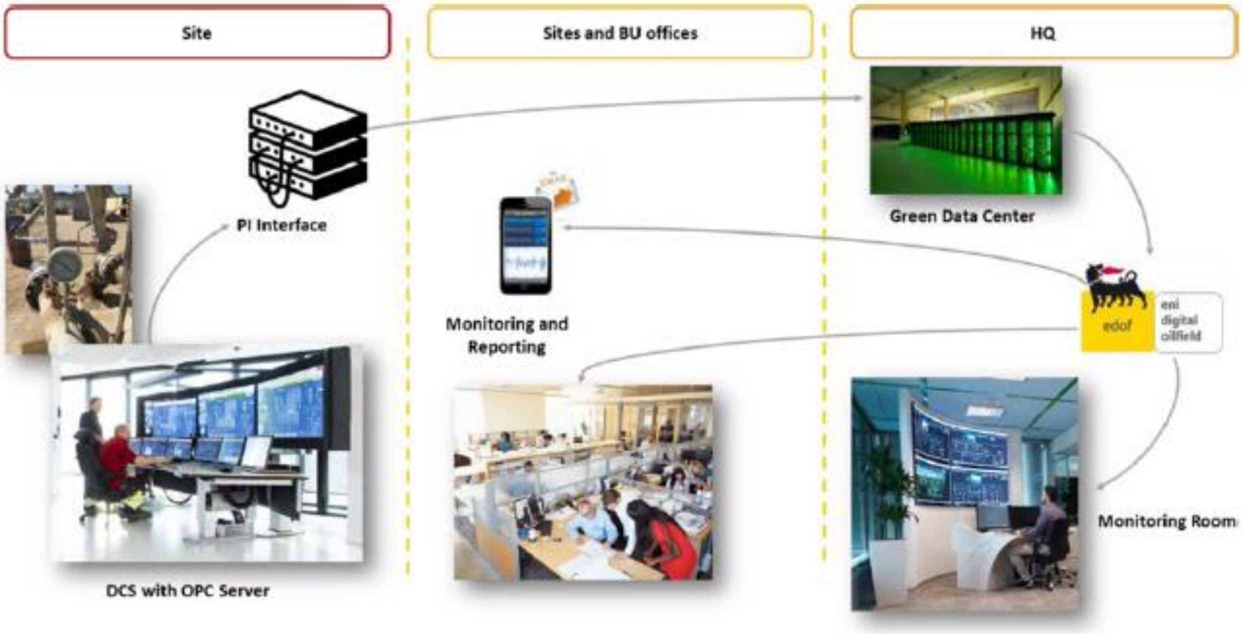


Fig. 1. 8—Example of the architecture in a digital oil field (Cadei et al. 2020).

Bello et al. (2018) use a distributed temperature sensing (DTS) data analytics system for online anomaly detection for gas lift valves. The authors use a combination of statistical feature extraction based on the position of gas lift valve, unsupervised K-means clustering and a supervised Support Vector Machines classifier. Snyder et al. (2019) provide an anomaly detection solution for plunger lifts. They take over 50 input features including gas rate, pressure, plunger arrival times etc. Their methodology is a self-adjusting detection system which comprises of a model statistic and a dynamic threshold. The authors, however, do not explain the model statistic as well as the dynamic threshold used in any detail. An example of this system on plunger lifted gas well is shown in **Fig. 1.9**. This advanced detection system according to the authors not only reduces the overall downtime due to plunger lift failures but also saves more than $2 million per year overall for the operator.

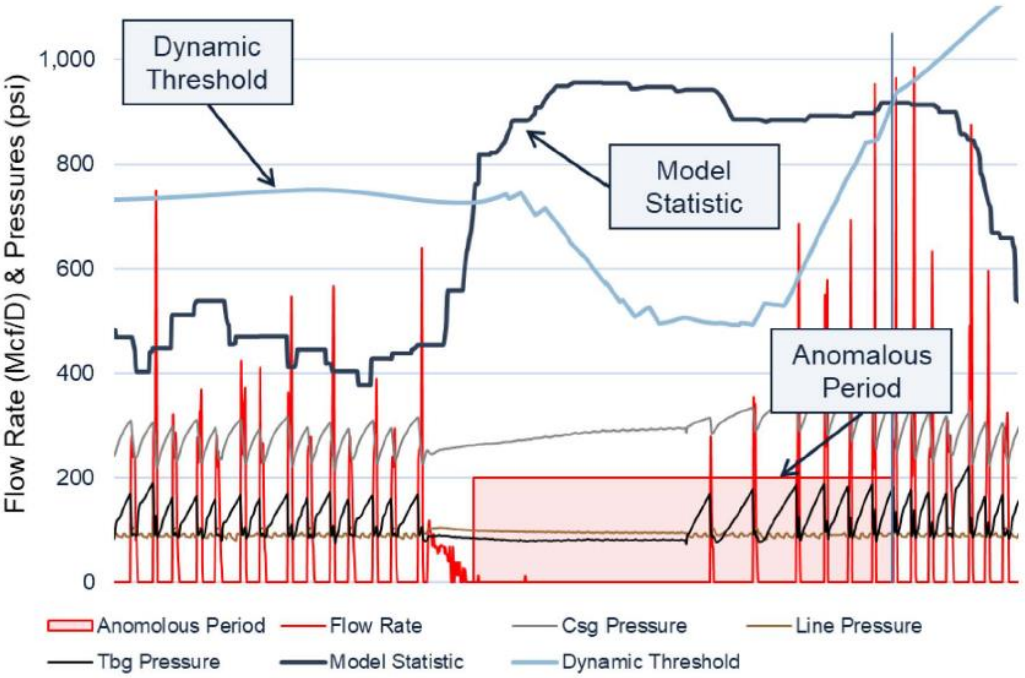


Fig. 1. 9—Self-adjusting detection system using a dynamic threshold and a model statistic (Snyder et al. 2019).

An application of natural language processing and artificial intelligence is seen in anomaly detection from drilling and completion reports. Zhang et al. (2020) use natural language processing to convert words from drilling and completion reports into vectors. Neural Networks then use the numerical vectors as inputs for a classification task between normal and abnormal. Jansen van Rensburg (2019) uses deep neural networks to detect anomalous behavior in Electrical Submersible Pumps (ESP). Several ESP's in an oilfield are connected remotely to a cloud server through a Supervisory Control and Data Acquisition (SCADA) system. Neural networks hosted on the SCADA system build an autonomous field surveillance system which identify severe anomalies and prevent occurrence of critical ESP failures.

In this thesis, I also apply a deep learning network called as the Convolutional Neural Network (CNN) for anomaly detection in sucker rod pumps.

This thesis is organized as follows: Chapter 2 discusses the application of a data-driven approach using causal inference applied to solving well spacing problems. This chapter explains the history behind causality, its assumptions and a detailed workflow for estimating the effect of well spacing on well performance. The causal inference workflow is applied to a dataset from a prolific oil basin in Texas with over 700 wells in the analyses. It includes the formation properties, fluid volume, proppant weight, landing zones and the downhole locations of the wells. The results of the case study show the optimal well spacing decision for this specific shale play. Chapter 3 discusses applications of anomaly detection for the oil and gas industry and how machine learning methods have been applied to solved problems related to the most widely used artificial lift in the world – sucker rod pump. Then, I explain the dataset used for the case study and discuss the chosen CNN model's architecture more generally. I dive into how the CNN works for a supervised classification task. Finally, the chapter finishes with the results and conclusions for the sucker rod pump failure anomaly detection.

Chapter 2: Causality and Well Spacing Decisions

2.1 History of Causal Inference

Causal inference is a growing field of study and has increased relevance today. Questions in many disciplines such as the efficacy of new vaccines against COVID-19, the impact of a new corporate tax law on the growth of GDP or quantifying the impact of a new marketing campaign has on the sales of a product can be answered through causal inference (Pearl and Mackenzie 2017).

In statistics, one is taught "Correlation is not Causation" with very good convincing examples but it does not tell us what causation really is. There are, however, spurious correlations prevalent in real data where two completely unrelated variables are highly correlated with each other. In **Fig. 2.1**, the per capita consumption of Margarine (black curve) correlates with the divorce rate (red curve) in Maine, but we know that consuming margarine does not cause a divorce.

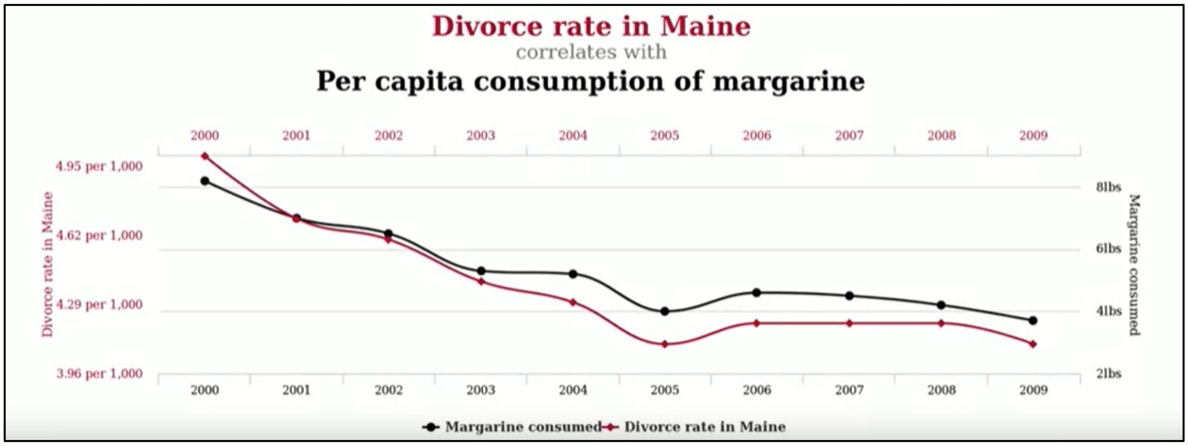


Fig. 2. 1— Divorce rate in Maine and per capita consumption of margarine plotted versus time between years 2000 and 2009 https://blogs.ams.org/blogonmathblogs/2017/04/10/divorce-and-margarine/ **(accessed 11 October 2020) (Vigen 2014).**

In **Fig. 2.2**, a rooster's crow at sunrise but does it cause the sun to rise? (Pearl and Mackenzie 2017). These are cases with obvious correlation with a lack of causation.

Fig. 2. 2—Rooster's crow heralds the sunrise https://www.trialrun.us/roosters-crow-causes-sun-rise-eat-chocolate-win-nobel-prize-statements-not-make-sense-blog-will/ (accessed 11 October 2020) (Trial Run 2020).

However, in some cases there might be a correlation statistically and the existence of any causality is not that obvious. In the 1950's, such an example was the effects of smoking on lung cancer. Sir Ronald Fischer who created and designed the randomized experiment in 1935 believed the association between smoking and lung cancer was spurious. Fischer (1958) claimed that in order to understand the effects of cigarette smoking on lung cancer, other common causes such as genotypes should be investigated before implying causation. Interestingly here, the doctors could not hold randomized control trials to ascertain causation simply because they were unable to randomly select individuals and ask them to smoke cigarettes in order to gauge the effect on their health decades later. An advertisement using doctors as an advertising aid for cigarettes is shown in **Fig. 2.3**. Eventually, the US surgeon general in 1964 declared that "Cigarette Smoking is causally related to lung cancer in men" which then resulted in decline of cigarette smoking among the US population in the years to follow and has increased life expectancy (Pearl and Mackenzie 2017).

"I'll Be Right Over!"

...24 hours a day your doctor is "on duty"... guarding health... protecting and prolonging life...

• Plays...novels...motion pictures...have been written about the "man in white." But in his daily routine he lives more drama, and displays more devotion to the oath he has taken, than the most imaginative mind could ever invent. And he asks no special credit. When there's a job to do, he does it. A few winks of sleep . . . a few puffs of a cigarette . . . and he's back at that job again . . .

According to a recent independent nationwide survey:

More Doctors Smoke Camels

than any other cigarette

Fig. 2. 3—Controversial advertisement in 1948 reassures the public that smoking is not injurious to health (Pearl and Mackenzie 2017).

In the field of medicine where randomized clinical trials are used to determine the effect of a treatment on the prevention of a disease, a large group of people given the drug are allocated to the "Treatment" group and another large group of people who are not given the treatment or given a placebo are allocated to the "Control" group. However, just like the smoking-lung cancer example, randomized control trials which are considered the gold standard in determining casual relationships are not always feasible for designing a statistical experiment. Fortunately, if we have a lot of observational data available to us, then one could still identify a causal effect if there is one. Cochran (1972) writes “Observational studies are an interesting and challenging field which demands a good deal of humility, since we can claim only to be groping toward the truth”.

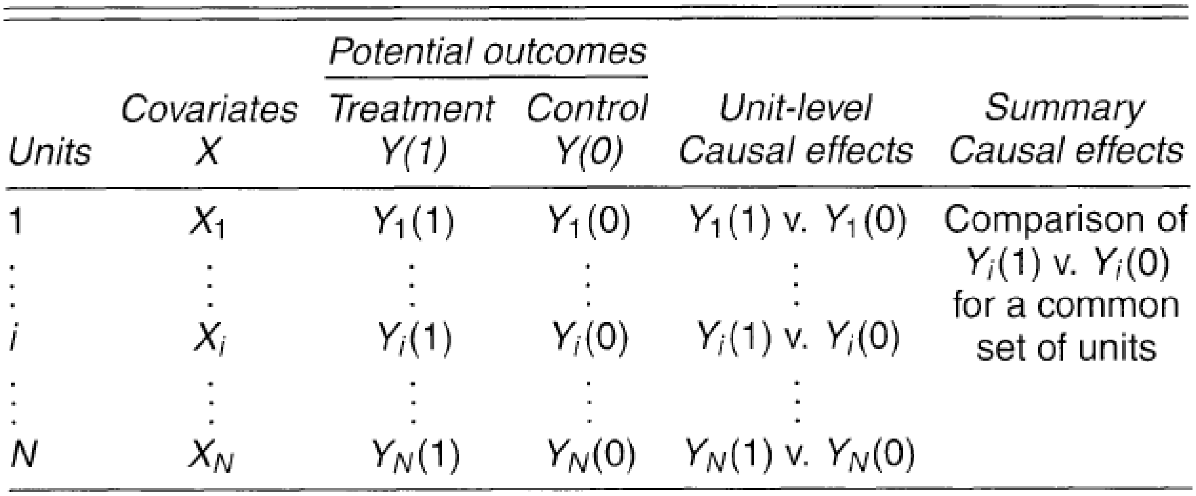
There is a clear distinction between randomized control trials (RCT) and observational studies which are both aimed at finding causal effects. A randomized control trial is an experiment where the study is under the complete control of the investigator in the sense that the people or subjects and the treatments assigned to those subjects are selected by the investigator. In the observational studies, the selection of the subjects to treatment groups are outside the controls of the investigator (Cox and Reid 2000; Morgan and Winship 2014). Therefore, the main design criteria which separates randomized control trials from the observational studies is the randomized allocation of subjects to treatment and control groups.

Although the English statistician, Sir Ronald Fischer was the first to propose the randomized control trials in 1925, Jerzy Neyman conducted a completely randomized control trial for his Ph.D. dissertation in Mathematics in 1923. More importantly, he used a different notation for a randomized agricultural experiment called as potential outcomes. Here his subjects were units of land, the treatment was type of manure used and potential outcomes were the crop yields from the land units. Neyman’s notation is given by the expression in Eq. 2.1. Here the causal effect in a randomized controlled experiment is the difference of sample means between the treatment group $\overline{Y}_1$ and the control group $\overline{Y}_0$ (Rubin 2005).

$$\overline{Y}_1 - \overline{Y}_0 = \sum_{i=1}^{N} \frac{Y_i(1) - Y_i(0)}{N} \qquad (2.1)$$

Almost half a century later in 1974, Donald Rubin expanded Neyman's potential outcome notations for defining casual effects into non-randomized or observational studies. Hence his model of potential outcomes framework is popularly known as the Neyman-Rubin model. The potential outcomes framework today is used in many fields including social sciences, economics, epidemiology, medicine, computer science and mathematics.

Suppose we are interested in the causal effect of treatment, T on some outcome, Y. **Table 2.1** shows the potential outcomes Y (1) (treatment), which is the outcome that would be observed when the treatment status, T=1 at time t and Y (0) (control) which is the outcome that would be observed when the treatment, T=0 at time t, for both treatment and control groups for all units ranging from 1 to N. These potential outcomes can be compared at the unit-level and at the group level. Covariates given by X are those variables which cannot be affected by the treatment status or take a fixed value before the treatment assignment. I refer to these outcomes as potential outcomes because one cannot observe both the outcomes when treatment status, T=0 or T=1 for a unit i at time t (Rubin 2005).



		Potential outcomes			
Units	*Covariates X*	*Treatment Y(1)*	*Control Y(0)*	*Unit-level Causal effects*	*Summary Causal effects*
1	X_1	$Y_1(1)$	$Y_1(0)$	$Y_1(1)$ v. $Y_1(0)$	Comparison of $Y_i(1)$ v. $Y_i(0)$ for a common set of units
⋮	⋮	⋮	⋮	⋮	
i	X_i	$Y_i(1)$	$Y_i(0)$	$Y_i(1)$ v. $Y_i(0)$	
⋮	⋮	⋮	⋮	⋮	
N	X_N	$Y_N(1)$	$Y_N(0)$	$Y_N(1)$ v. $Y_N(0)$	

Table 2. 1— Potential Outcomes (Rubin 2005).

2.2 Potential Outcomes and the Fundamental Problem of Causal Inference

The assignment of the treatment to the subjects in RCTs is essentially random. Therefore, from **Table 2.1**. assuming we have conducted the RCT, we can compute the unbiased estimate of the average treatment effect or the causal effect directly from the experimental data. This estimate is given by the expression by Eq. 2.2. As one can notice, this is same as the important observation from Neyman's seminal work in 1923. For RCTs, we can take the difference between the means of the treatment group and the control group to arrive at the unbiased causal effect (Austin 2011).

$$E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] \qquad (2.2)$$

This, however, is not the case for the observational studies. The subjects in observational studies in the treatment group and control group are different from each other as the assignment of treatment itself is not random to begin with. Hence, we have the expressions in Eq. 2.3 and Eq. 2.4 considering T as the treatment status (1=given, 0=not given). So, we cannot obtain an unbiased estimate for the causal effect by taking the difference in the means of the treatment and control group (Austin 2011).

$$E[Y|\,T = 1] \neq E[Y(1)] \qquad (2.3)$$

$$E[Y|\,T = 0] \neq E[Y(0)] \qquad (2.4)$$

Therefore, $E[Y|\,T = 1] - E[Y|\,T = 0]$ is not the average causal effect here because we are considering two different sub-sets of the subject population with different treatments and covariates. Unfortunately, this has been the basis of several published papers (Miller et al. 2016; Lindsay et al. 2018; Xu et al. 2019) that discuss parent-child well performance.

$E[Y(1)] - E[Y(0)]$ is the true causal effect that we seek to find since it is comparing and taking the difference in potential outcomes for the same population (Roy 2019).

This leads us to the discussion of the fundamental problem in causal inference. In order to illustrate this problem, I will take the following example from day-to-day life. In **Fig. 2.4**, suppose a person has a strong headache and as a quick remedy or treatment he or she takes 2 pills of Ibuprofen. Let us assume the headache is gone in an hour. So, can we say that treatment- Ibuprofen for the headache worked and that was the causal effect of taking Ibuprofen on the headache? That claim is incorrect because we do not know what would have happened had the person not taken Ibuprofen at all.

This is the fundamental problem of causal inference that we can never ever really observe a causal effect because we cannot have both outcomes at the same time for one person. Such outcomes as aforementioned are called “Potential Outcomes”. In this case, where we do not know the outcome when the person has not taken Ibuprofen, is considered as a potential outcome. Instead, we should ask the question, “What would be the rate of headache remission be if everyone took Ibuprofen when they had a headache versus if no one did?”.

We deal with the fundamental problem of causal inference by looking at a group level rather than on an individual level. Similarly, in multivariate observational studies like the one discussed in this thesis; I conduct the causal inference analysis at the group level. Only when there is a difference between the two potential outcomes taken at the group level, there is a causal effect of the treatment on the outcome (Morgan and Winship 2014).

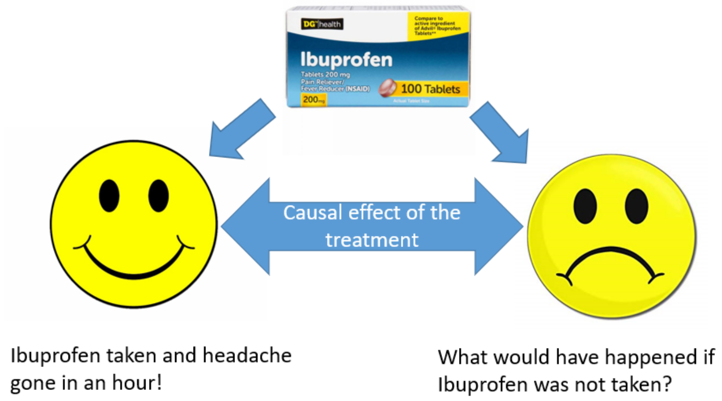


Fig. 2. 4—Example of potential outcomes illustrating the fundamental problem of causal inference (Sakhardande and Devegowda 2021).

2.3 Assumptions behind Causal Inference

Since we do not always have the capability to conduct a randomized control trial in most cases and are often left with data to infer about causality, observational studies must be accompanied with certain assumptions. These assumptions are called causal assumptions. The causal assumptions will be made regarding the observed outcome, Y, some treatment, T and a set of pre-treatment covariates, X (Roy 2019). Here I am listing the casual assumptions here as follows:

1. Stable Unit Treatment Value Assumption (SUTVA)
    a. SUTVA is simply the a priori assumption that the value of outcome, Y for a unit u when exposed to treatment, T will be the same no matter what mechanism is used to assign treatment, T to unit u and no matter what treatments the other units receive (Rubin 1986 ; Morgan and Winship 2014).
    b. There is only one version of the treatment (Morgan and Winship 2014).
2. Ignorability

Given the pre-treatment covariates, the treatment assignment is said to be ignorable or independent from the potential outcomes. In other words, the knowledge of a unit's treatment assignment yields no information whatsoever about the unit's potential outcomes conditional on its covariates (Rosenbuam and Rubin 1983a; Morgan and Winship 2014). This can be expressed as follows in Eq. 2.5:

$$Y(1), Y(0) \perp\!\!\!\perp T|\boldsymbol{X} \quad \text{(2.5)}$$

Thus, the ignorability assumption is satisfied if some units across the treatment and the control group have similar covariates which means that their treatment assignment is essentially random. The ignorability is achieved using a method called propensity score matching which I will discuss later in the methodology section.

3. Positivity

Every unit has a non-zero probability of being in the treatment or the control group conditional on the unit's covariates, **X** (Rosenbuam and Rubin 1983a; Austin 2011). This is given by the expression in Eq. 2.6 where I assume that treatment, T is binary:

$$0 < P(T = t|\boldsymbol{X} = \boldsymbol{x}) < 1 \ \forall \ t, \boldsymbol{x} \quad \text{(2. 6)}$$

2.4 Confounding Bias and Directed Acyclic Graphs

To explain the bias due to confounding, first I define the term “Confounders”. Confounders are those variables or a set of pre-treatment covariates which affect both the outcome and the treatment assignment. To understand the variables which can be considered as confounders and which cannot be considered as confounders, I will give a couple of examples:

1. Let us assume if the treatment assignment is random and it is based on a coin toss. Then the toss of a coin affects the treatment status, but it does not affect the outcome which is a head or a tail. Therefore here, the coin toss cannot be considered as a confounder.
2. Let us assume that people with high levels of cholesterol have a high chance of developing a heart disease. If cholesterol levels in blood is considered as a factor which influences the

type of treatment given, then cholesterol levels in this population would affect the treatment status as well as the outcomes. Therefore here, the blood cholesterol level is a confounder.

We want to be able to find such confounders or covariates, X such that the ignorability assumption holds true. If we find such covariates or set of pre-treatment variables, then it said to be sufficient for controlling the bias due to confounding. This bias stems from the fact that the confounder or the covariates, X have an association with the treatment variable, T and the outcome, Y. In other words, we cannot observe the true causal effect of the treatment, T on the outcome, Y unless we have controlled for this bias due to the presence of common cause variables, that are, covariates, X. This is also known as the bias due to confounding. In order to achieve the goal of controlling this bias, I now introduce the concept of directed acyclic graphs (DAGs) which will help us in finding those set of covariates, X such that ignorability assumption is achieved (Pearl 2009; Roy 2019).

We can express our assumed causal relationships in the form of directed acyclic graphs (DAGs). Specially in the case, where there is confounding (there are common cause variables/covariates present), DAGs are useful in explaining how we can achieve ignorability. To start off, let us consider this simple directed graph in **Fig. 2.5**.

Fig. 2. 5— An example of a simple directed graph

Here, I am assuming that there is a treatment, T which affects the outcome, Y. There is a belief that T causes Y and the assumed cause and effect relationship is expressed by the blue arrow directed from T to Y. Now, let's consider another directed graph in **Fig. 2.6**.

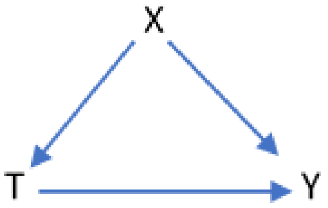


Fig. 2. 6— An example of a directed acyclic graph (DAG)

Here I am assuming that T affects Y and X affects T and Y. This graph is also called as a directed acyclic graph because every variable in it has a direction or an arrow associated with it and it has no cycles present ,that is, if we start at one variable and follow the direction of the arrows we won't end up back on the same variable. There are more complicated DAGs that we can draw based upon the complexity of relationships between the variables in the study. I will mention the DAG from Fig. 2.6 later when I discuss the design of the experiment.

In Fig. 2.6., there is also a path going from Y to T indirectly through X. Now in **Fig. 2.7**, if we condition on levels or some values of X, then this would block the path carrying the association between X and Y and X and T. This would essentially mean that we are able to remove all other association between T and Y (through X) which is not through a direct effect of T on Y. This path blocking by conditioning on a common cause in a directed acyclic graph eliminates the bias due to confounding.

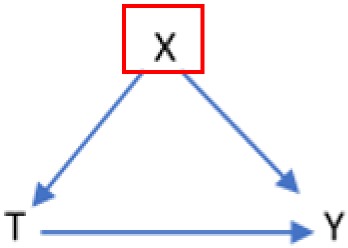


Fig. 2. 7— Conditioning on a confounder X.

For example, let us assume that treatment, T is a heart treatment and outcome, Y was heart attack and there is only one common cause blood pressure, X. If we think that there are people in the population with abnormal blood pressure and they have a higher chance of getting the heart treatment, then the treatment assignment, T is not random to begin. Conversely, we can say that people with normal blood pressure are less likely to get the treatment. Thus, the ignorability assumption does not hold here.

In order to achieve the ignorability assumption then, we must condition on a certain sub-set of population with only abnormal blood pressure or only normal blood pressure. Then within that sub-set of population, everyone has an equal chance of getting treatment thereby making the treatment assignment mechanism random satisfying the ignorability assumption. So, if we condition on levels of population with a similar range of abnormal blood pressure, X as per Fig. 2.7., we block any association between heart treatment, T and heart attack, Y which is not due to direct effect of heart treatment, T on heart attack, Y. This eliminates the confounding bias.

With this brief history and some background in causality, I will now transition to the problem at hand which is quantifying the causal effect of well spacing on well performance. Before I discuss the methodology used, first I will describe the dataset used for the causal inference workflow.

2.5 Dataset Description

For this study, a multivariate observational data from 725 wells is used. These wells are in the Permian Basin, West Texas. The dataset includes location of these wells in latitudes and longitudes for the well head as well as the bottomhole locations, true vertical depths (TVD) and reservoir and completion design variables. The reservoir properties, completion design variables and TVD are defined as covariates because together they impact the well spacing decision and well performance. The dataset has over 400 covariates, which are known to impact the well performance and the well spacing. There are 2 completion variables which are the fracture fluid volume and proppant weight. The TVD was another variable and the remaining were petrophysical/reservoir properties.

As explained earlier by Gupta et al. (2020) and Rafiee and Grover (2017), completion size and reservoir properties have a direct influence on the long term well EUR. Also, any lease development requires careful study of the complex interplay between reservoir properties, the choice of completion design, landing zones and geomechanical properties (Shahkarami and Wang 2017) as they all influence the well placement decision and economics behind it.

In order to understand the importance of the completion variables in the causal analyses, in **Fig. 2.8 and 2.9** , I plot the actual cumulative 360-day production normalized to lateral length from the dataset for all wells versus the proppant weight normalized to lateral length, fracture fluid volume normalized to lateral length and TVD respectively. Fig. 2.8 and 2.9 show that well performance has an overall weak but positive relationship with respect to both completion variables, proppant weight and fracture fluid volume normalized to well lateral length. This relationship makes assessing the impact of well spacing on performance quite challenging.

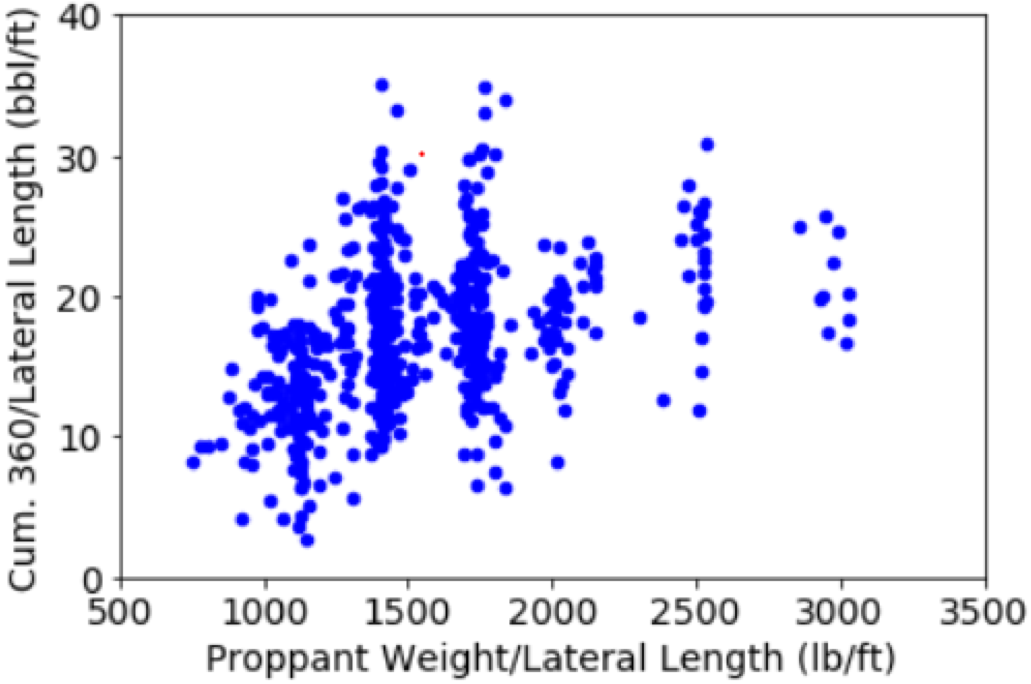


Fig. 2. 8— A plot of actual 360-day cumulative production normalized to lateral length (bbl/ft) for all wells versus proppant weight normalized to lateral length (lb/ft).

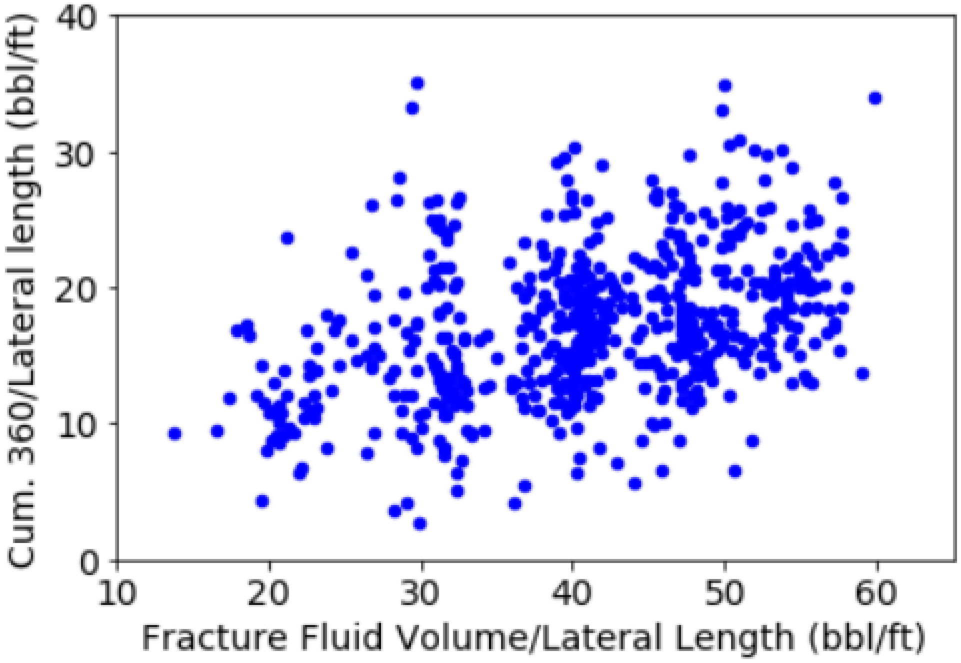


Fig. 2. 9— A plot of actual 360-day cumulative production normalized to lateral length (bbl/ft) for all wells versus fracture fluid volume normalized to lateral length (bbl/ft).

As mentioned earlier, this dataset has over 400 covariates. In order to minimize the number of covariates, I use a dimension reduction technique called Principal Component Analysis (PCA).

PCA is used to combine the highly correlated variables together to form a smaller number of uncorrelated variables which are called "Principal Components" (Pearson 1901; Hotelling 1933; Jolliffe 2002). These principal components account for or explain most of the variance from the original data. After conducting principal component analysis on the reservoir variables, I choose the first 5 principal components which explain about 80% of variance in the data as shown in **Fig. 2.10**. These 5 principal components become the new set of covariates which represent the original 400+ reservoir variables.

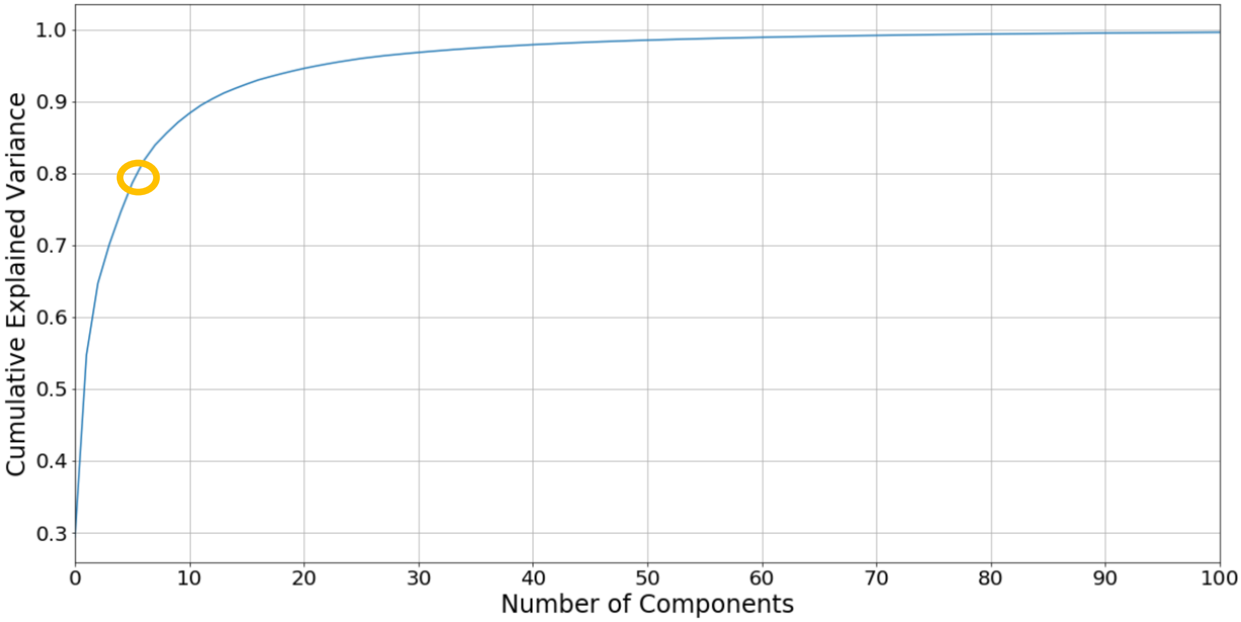


Fig. 2. 10— A Scree plot showing that 80% of cumulative variance (y-axis) is represented by 5 Principal components (x-axis).

Using TVD and bottomhole latitudes and longitudes, a 3-dimensional well distance for each well with respect to the rest of the wells is computed. Then taking the 3D well distance and time criteria based on well production start dates, the wells were divided into parent wells and child wells groups.

In this case study, four well spacing options for parent-child wells groups are considered. These were 500 ft, 600 ft, 700 ft and 800 ft. The distance and time criteria used for separating the wells into a parent well and a child well group is explained in the next section. Well spacing options outside this range, from 500 ft to 800 ft, were not considered because the parent-child well search criteria led to insufficient number of wells to carry out the causal inference workflow. The well performance was estimated by normalizing cumulative 180-day and 360-day well production numbers by their lateral length for each well in the dataset.

2.6 Design of the Experiment

For this dataset, the treatment variable is chosen to be well spacing to evaluate the impact of spacing decisions on well performance. The child wells are placed in the treatment group and the parent wells are placed in the control group. The outcome is set to either cumulative 180-day or 360-day production normalized to lateral length. If the outcome is cumulative 360-day production normalized to lateral length, then we are interested in seeing the difference in the potential outcomes at the group level. As mentioned earlier, 5 principal components representing the reservoir variables, 2 completion variables (proppant and fracture fluid volumes normalized to lateral length) and true vertical depths form the set of covariates or the confounding variables.

For each well pair, a 3-dimensional distance is computed from the true vertical depth, bottomhole latitude, and longitude. The wells get assigned to either the parent wells group or child wells group based on the 3-dimensional distance and time criteria. Using an example, I describe the criteria to assign wells to a treatment (child) group or control (parent) group. Let us consider the group with 700 ft spacing and an outcome of 360-day cumulative oil normalized to lateral length:

1. If the 3-dimensional distance between a well pair is greater than 700 ft, then both wells are placed in parent well group. The assumption is that over a period of one year, the impact of interference between the wells is negligible.
2. If the 3-dimensional distance is less than 700 ft, then the well production start dates for the well pair is considered. If well 2 is put on production within 1-year of well 1 start date, both wells are placed in the child well group. The assumption is that within a period of one year due to their proximity to each other, there is well interference impacting production of both wells.
3. If well 2 is put on production more than 12 months of well 1 being on production, then well 1 goes into the parent well group and well 2 goes into child well group. The assumption is over a period of 1 year, the impact of interference is negligible for well 1 but not for well 2 as well 2 may not have access to virgin reservoir pressure and undrained volume unlike well 1 when it started production.
4. Using the same logic, wells are allocated into the parent and child well groups at all other well spacing options considered, that are, 500 ft, 600 ft and 800 ft.

Also, the well spacing status is binary with parent wells denoted by “0” and child wells by “1”. As mentioned earlier, other spacing options outside the 500 ft to 800 ft range were not considered because the parent-child well search led to insufficient number of wells to carry out the causal inference workflow.

In **Fig. 2.11**, I use a directed acyclic graph just as shown in Fig. 2.6 to explain the association between the covariates, the well spacing decision and the outcome, that is, well performance. Here the covariates are denoted by **X**, the well spacing decision is denoted by T and the well performance denoted by Y. The direction of the arrows indicates the assumed cause and effect relationship between the variables. In order to estimate the impact of well spacing on well performance, it is necessary to decouple the effect of well spacing on the well performance from the variations that we see in the heterogenous completion designs, reservoir properties and well landing zones.

If we select candidate wells in the parent and child wells groups with covariates that lie in a similar range like Fig. 2.6 shown earlier, we block any association between well spacing and well performance arising due to the covariates, or we eliminate the impact of reservoir properties or completion design on well spacing decisions. This eliminates the confounding bias. In other words, the aim is to search for child and parent wells in both groups that have similar covariates before comparing their performance over a certain period. The workflow that would enable such a parent-child well production performance comparison over time is described in the next section.

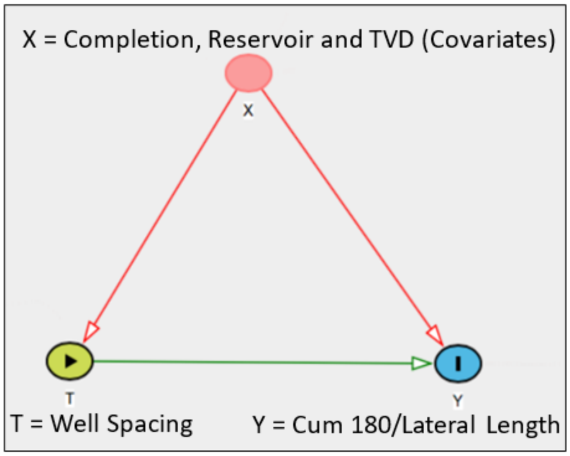


Fig. 2. 11—Association between well spacing, covariates and well performance (Sakhardande and Devegowda 2021).

2.7 Methodology

Here I explain the causal inference workflow in five steps as follows:

1. *Propensity score estimation*

The first step here is to estimate a probability score on each well if it is placed in either of parent or child well groups given their covariates. I will refer to this probability estimate as the propensity score. The assumption for the well spacing status here is binary, that is, after the search for the parent and child wells is complete. Parent wells are placed in the group assigned status “0” and the child wells are placed in the group assigned status “1”.

The estimation of the propensity score is carried out through a logistic regression model of the treatment variable (well spacing) on the covariates (reservoir properties, completion design and TVD). Here I assign these variables as letters for simplicity. The treatment variable ,that is, well spacing is denoted by T, the set of covariates denoted by **X** and the outcome ,that is, 180-day or 360-day cumulative production normalized to lateral length denoted by Y. Mathematical expression given in Eq. 2.7 shows the probability of getting treatment π_i and the binary regression model is given by Eq. 2.8. This is the propensity score estimate for each child well and parent well given their covariates.

$$\pi_i = P(T = 1|\boldsymbol{X}_i) \quad \ldots\ldots (2.7)$$

The binary logistic regression model for $\pi_i$is given by the expression:

$$\pi_i = P(T = 1|\boldsymbol{X}_i = \boldsymbol{x_i}) = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \quad \ldots\ldots (2.8)$$

where **X** is the set of j discrete and/or continuous covariates such as $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3 \ldots, \boldsymbol{X}_j)$. Each $\boldsymbol{x}_i$ is the observed value for each observation i covariate in set **X**. β_0 and β_1 are model coefficients which are estimated using a maximum likelihood estimator.

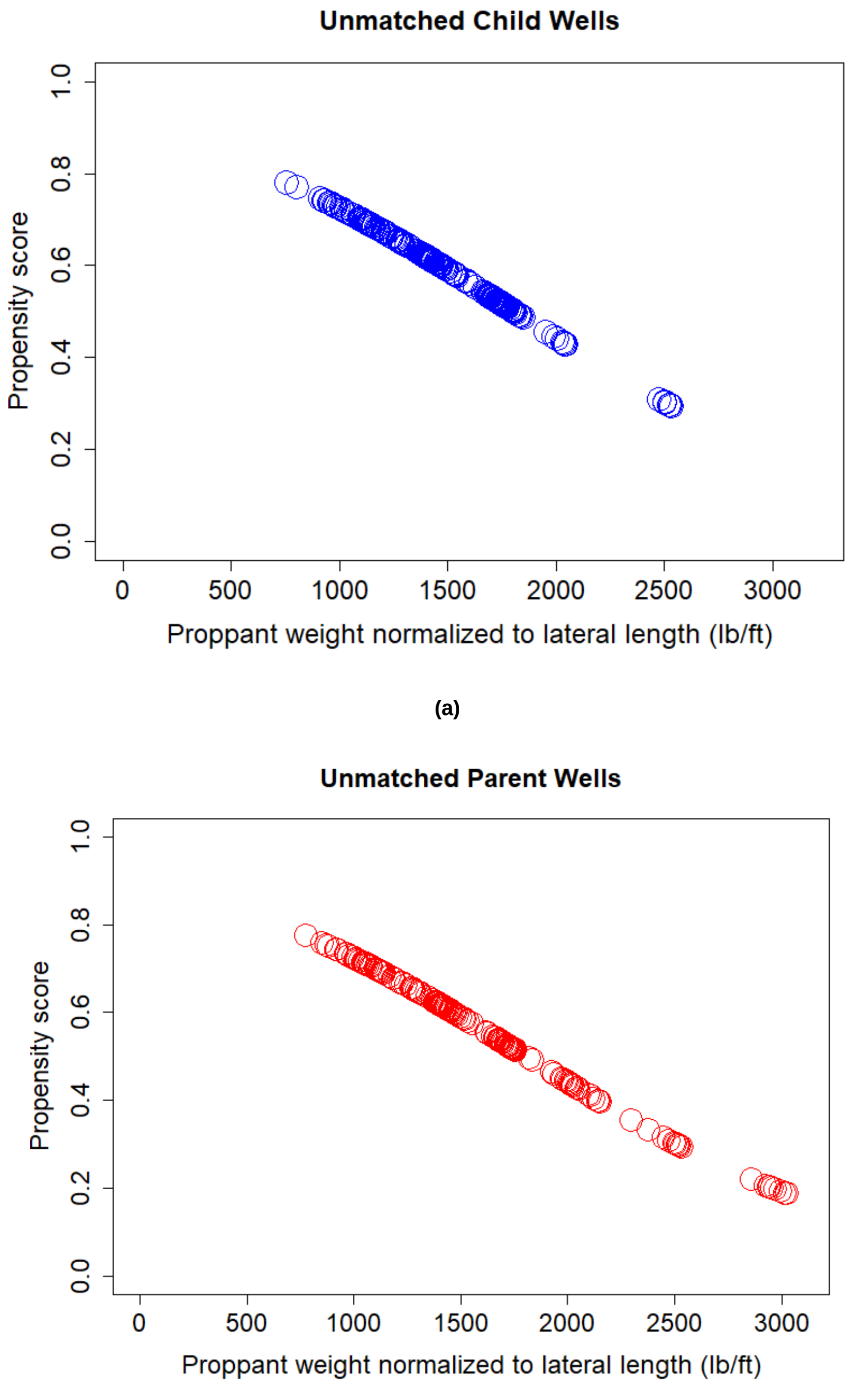
Table 2.2 shows the analysis of variance output for the logistic regression considering all the covariates. The last column shows the level of significance based on the P-value for each covariate. P-values of less than 0.05 are considered as statistically significant.

	Degrees of freedom	Deviance	Residual Degrees of Freedom	Residual Deviation	Pr(>Chi)
NULL			604	824.36	
P1	1	7.4	603	816.95	6.00E-03
P2	1	59.6	602	757.29	1.12E-14
P3	1	1.14	601	756.14	2.85E-01
P4	1	6	600	750.14	1.40E-02
P5	1	31.56	599	718.58	1.93E-08
Proppant weight normalized to lateral length (lb/ft)	1	34.9	598	683.68	3.46E-09
Fracture fluid volume normalized to lateral length (bbl/ft)	1	6.28	597	677.4	1.20E-02
TVD (ft)	1	7.39	596	670	6.00E-03

Table 2. 2—Analysis of Variance (ANOVA) for the logistic regression model with respect to all covariates.

As an example, if proppant weight normalized to lateral length is the only covariate, then we can determine the probability of a well belonging to a parent well or child well group for a given normalized proppant weight. In **Fig. 2.12**, we can see the range of the propensity scores on the y-axis for parent wells and child wells plotted versus their proppant weight normalized to lateral length on the x-axis. Fig. 2.12a shows child well scores in blue and Fig. 2.12b shows parent well scores in red.

For the child and parent wells, we can notice that the value of the propensity scores is lower for higher proppant weight (lb/ft). Comparing both graphs together, we also see that few parent wells have proppant weight higher than 2700 lb/ft whereas none of the child wells have proppant weight over this value. This indicates the variability in the normalized proppant weight in both child and parent groups.



(a)

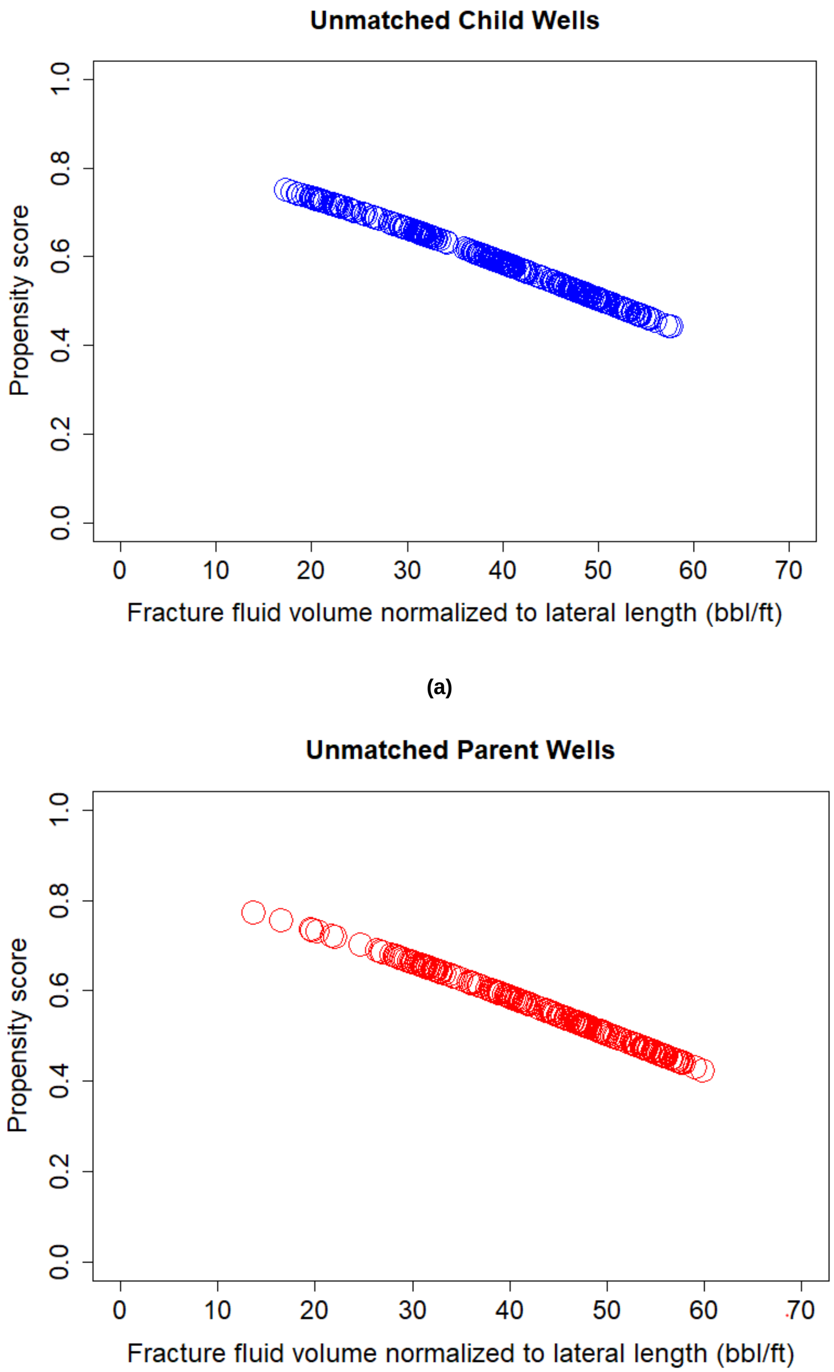
(b)

Fig. 2. 12—(a) Plot of propensity score versus proppant weight normalized to lateral length (lb/ft) for unmatched child wells (b) Plot of propensity score versus proppant weight normalized to lateral length (lb/ft) for unmatched parent wells (Sakhardande and Devegowda 2021).

Similarly, if we had only normalized fracture fluid volumes (bbl/ft) as the only covariate, we can determine the propensity scores for the child and parent well groups. This is shown in **Fig. 2.13**. Fewer parent wells in Fig. 2.13b. have propensity score between 0.7 and 0.8 (shown in red) whereas for the child wells in Fig. 2.13a. propensity score density (shown in blue) is higher in that specific range. This again indicates the variability in the range of the propensity scores between the parent and child wells with respect to the covariate, fracture fluid volume normalized to lateral length (bbl/ft).

If we consider TVD as the only covariate in **Fig. 2.14**, we see a similar trend of lower propensity scores for wells with deeper landing zones. In general, by looking at the density of the propensity scores in Fig. 2.14b, we notice that a greater number of parent wells have TVD values exceeding 9750 ft compared to the child wells in Fig. 2.14a.

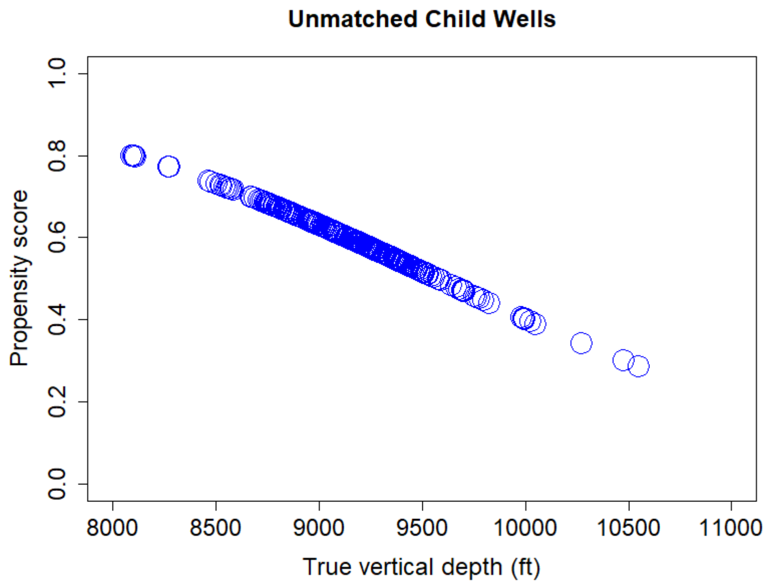
The main takeaway from Fig. 2.12, 2.13 and 2.14 is that we still need to find wells with similar propensity scores density considering both parent and child wells groups which is not the case currently. This dissimilarity can be also verified when all the covariates are considered together by looking at the balance between the parent-child groups. This leads us into the second step of workflow which is the balance computation.



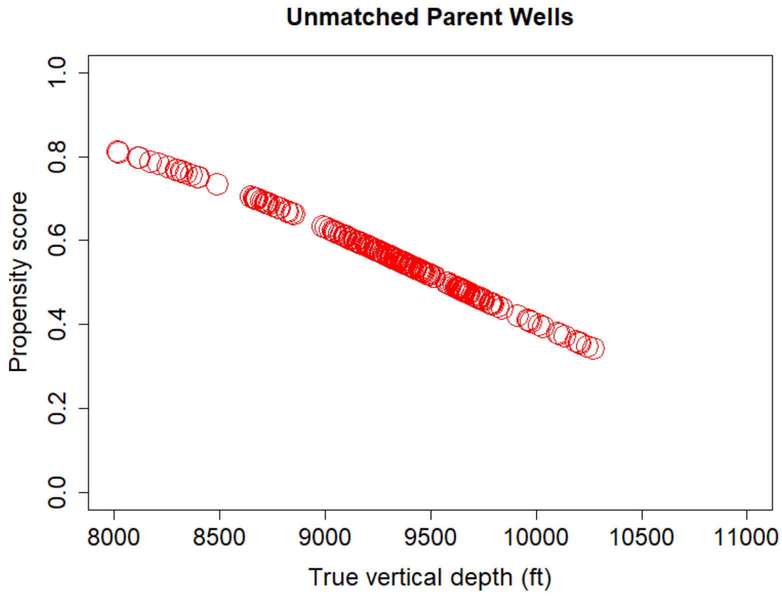
(a)

(b)

Fig. 2. 13—(a) Plot of propensity score versus fracture fluid volume normalized to lateral length (bbl/ft) for unmatched child wells (b) Plot of propensity score versus fracture fluid volume normalized to lateral length (bbl/ft) for unmatched parent wells.



(a)



(b)

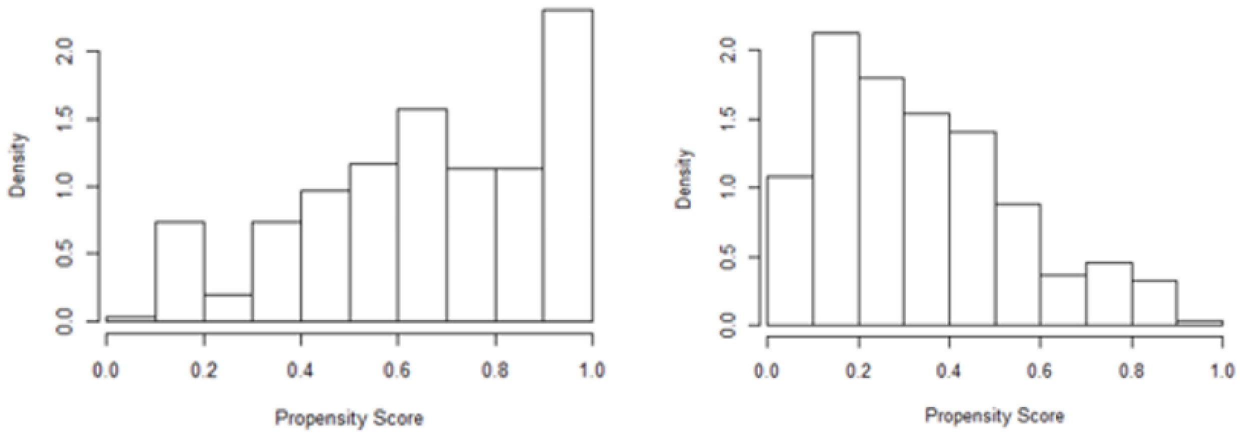
Fig. 2. 14—(a) Plot of propensity score versus true vertical depth (ft) for unmatched child wells (b) Plot of propensity score versus true vertical depth (ft) for unmatched parent wells.

2.Balance computation

The second step in this workflow is computing the statistical balance between child or treatment and parent or control well groups. This is to be done at every spacing option (500 ft, 600 ft, 700 ft and 800 ft) and outcome (180-day or 360-day cumulative production normalized to lateral length) considered. To find if there is a statistical balance between the parent and child well groups, we use the distribution of propensity score as a measure. Eq. 2.9 states that probability distribution of being in the child or treatment wells group should be the same as probability distribution of being in the parent or control wells group if we condition on a similar set of covariates X for each group (Hirano et al. 2003; Imbens and Rubin 2015)

$$P(\boldsymbol{X} = \boldsymbol{x} | \pi(\boldsymbol{X}) = p, T = 1) = P(\boldsymbol{X} = \boldsymbol{x} | \pi(\boldsymbol{X}) = p, T = 0) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2.9)}$$

The expression in Eq. 2.9 holds true only if we take the propensity score as a balance measure. By using this score, we can restrict the analysis to a smaller subset of the wells in both groups. Our goal is to search for the parent and wells where the difference in their propensity scores is very small or it is almost close to zero (Rosenbaum and Rubin 2006; Roy 2019). At the 600 ft well spacing option and the outcomes set to normalized Cumulative 360-day production, **Fig. 2.15** shows the propensity score distribution for both child Fig 2.15a and parent well groups Fig. 2.15b. I observe that the of the propensity scores are different and therefore for a given set of covariates (5 principal components, proppant weight normalized to lateral length, fracture fluid volume normalized to lateral length and TVD), the probability of being in the parent or child group is unequal. As a result, wells in parent and child groups currently have statistically dissimilar completion designs, reservoir and landing zone properties.



(a) (b)

Fig. 2. 15—(a) Distribution of propensity score for unmatched child wells (b) Distribution of propensity score for unmatched parent wells (Sakhardande and Devegowda 2021).

Fig. 2.16, 2.17 and 2.18 show box plots for TVD, proppant weight normalized to lateral length (lb/ft) and fracture fluid volume normalized to lateral length (bbl/ft) respectively. From the box plots, one can easily spot that there is variation in the range of the landing zones and completion size properties between parent and child well groups.

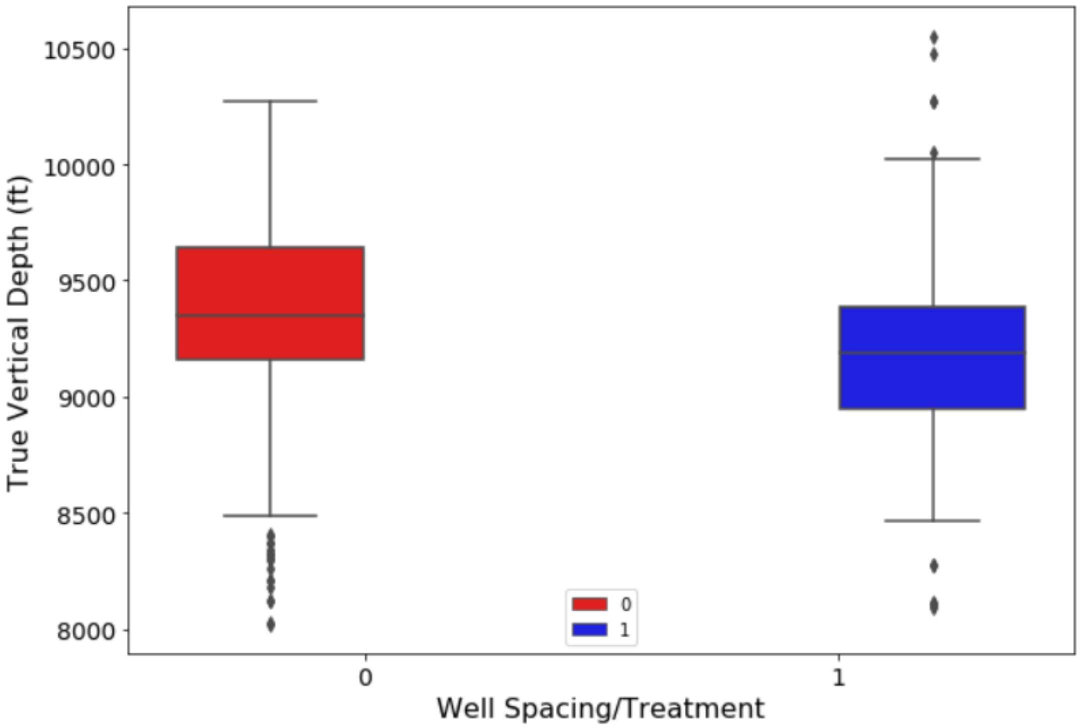


Fig. 2. 16—True vertical depth (ft) box plots for parent (0) and child (1) well group.

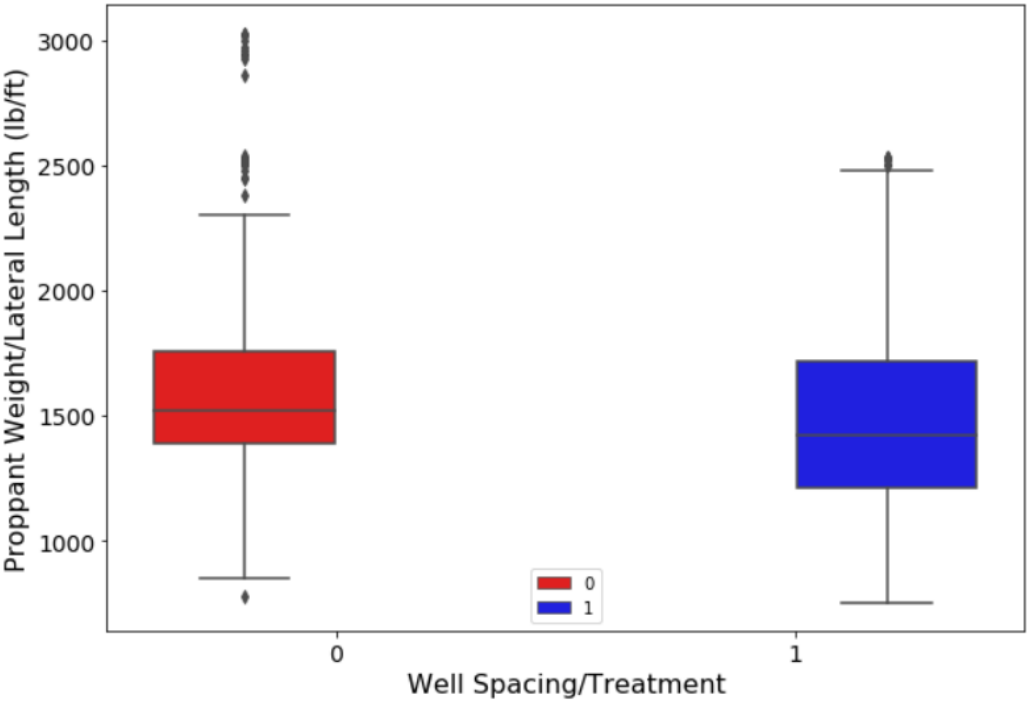


Fig. 2. 17—Proppant weight normalized to lateral length (lb/ft) box plots for parent (0) and child (1) well group.

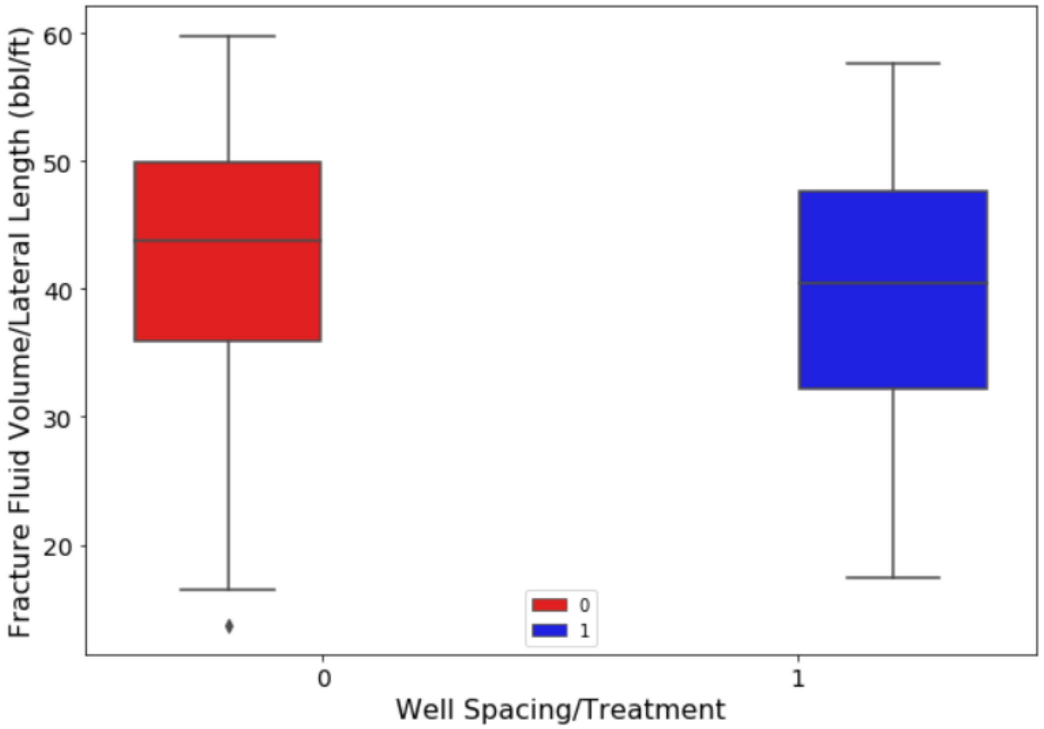


Fig. 2. 18—Fracture fluid volume normalized to lateral length (bbl/ft) box plots for parent (0) and child (1) well group.

Another way one can check for the balance is by computing the standardized mean differences (SMD) for all the individual covariates. Standardized mean differences are given by the expression in Eq. 2.10:

$$SMD = \frac{\bar{X}_{treatment} - \bar{X}_{control}}{\sqrt{(s^2_{treatment} - s^2_{control})}} \quad \text{(2.10)}$$

Here we are simply taking the difference in the covariate means ($\bar{\boldsymbol{X}}_{treatment}$, $\bar{\boldsymbol{X}}_{control}$) between child/treatment and parent/control groups and dividing them by the square root of difference in the squared standard deviations of the covariates ($s^2_{treatment}, s^2_{control}$) between the treatment and control groups.

In the **Table. 2.3**, the standardized mean differences between the two groups (child wells denoted by “1” and parent wells denoted by “0”) are shown in the last column for all the covariates used in this study at well spacing of 700 ft. SMD is measure of how different the covariates are between both groups. As rule of thumb from the literature (Lanza et al. 2013; Stuart et al. 2013), SMD values over 0.2 are considered bad, between 0.1 and 0.2 are good and anything less than 0.1 is great. In the first column of the Table 2.3, the five principal components are denoted from P1 through P5 along with the 3 other covariates, that are, proppant weight, fracture fluid volumes both normalized to lateral length and true vertical depth. The 2nd and 3rd columns show the mean and the standard deviation for the parent well group and 4th and 5th column show the mean and standard deviation for child well group for each covariate. SMD values are shown in the last column.

As highlighted by red rectangles, we see that the SMD values for P2 and the proppant weight are over 0.4 which indicate significant differences at the individual covariate level between the child and parent wells groups. Therefore, there is a statistical dissimilarity seen not only at the individual covariates level but also in the overall propensity scores distribution considering all covariates together as seen in Fig. 2.15. In order to improve from this current state, I perform one to one matching which is explained in the next step.

Covariates	256 Parent Wells (0)		349 Child Wells (1)		SMD
	Mean	Standard Deviation	Mean	Standard Deviation	
P1	1.5	12.92	-1.1	10.4	0.222
P2	3.78	8.6	-2.77	11.32	0.651
P3	0.3	7.6	-0.22	6.05	0.077
P4	0.5	4.48	-0.37	5.29	0.178
P5	1.08	4.74	-0.79	4.22	0.417
Proppant weight normalized to lateral length (lb/ft)	1639.98	462.95	1466	307.6	0.444
Fracture fluid volume normalized to lateral length (bbl/ft)	42.65	9.55	39.6	0.74	0.317
TVD (ft)	9311.9	433.85	9171	357.03	0.355

Table 2. 3— Standardized mean differences (SMD) between parent (0) and child wells (1) groups at individual covariate levels (before matching).

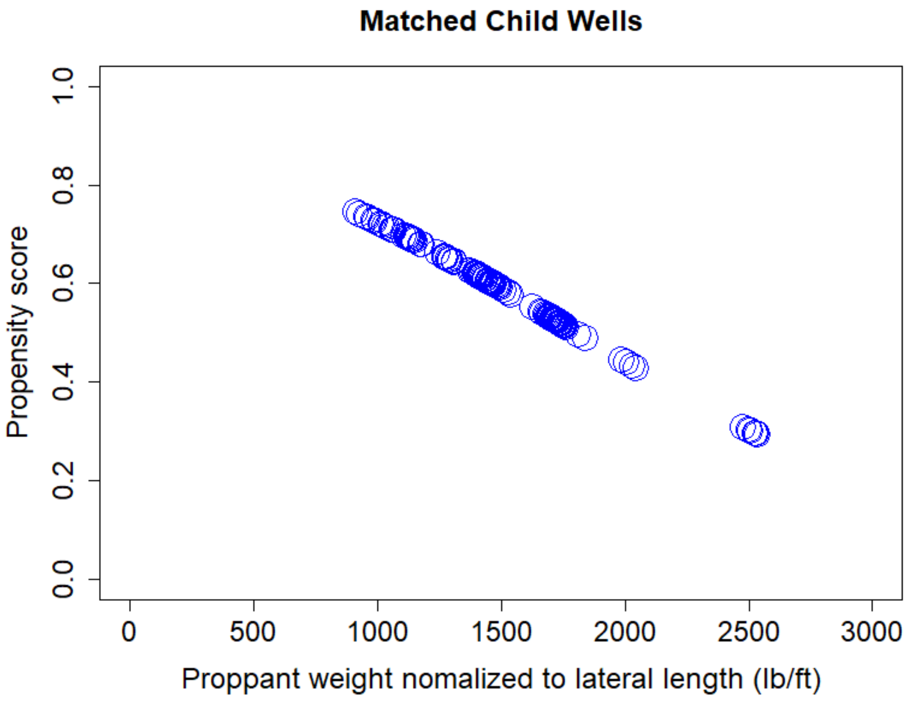
3. One to One Matching

The third step is to find the way to improve the statistical balance shown in Fig. 2.15. I can do this by limiting the analysis to a smaller subset of parent and child wells with similar propensity scores. This process is known as matching. Here the aim is to discard wells from both groups that do not have statistically similar covariates. In order to do this, a one-to-one match is conducted on the propensity score based upon the covariates in **X**.

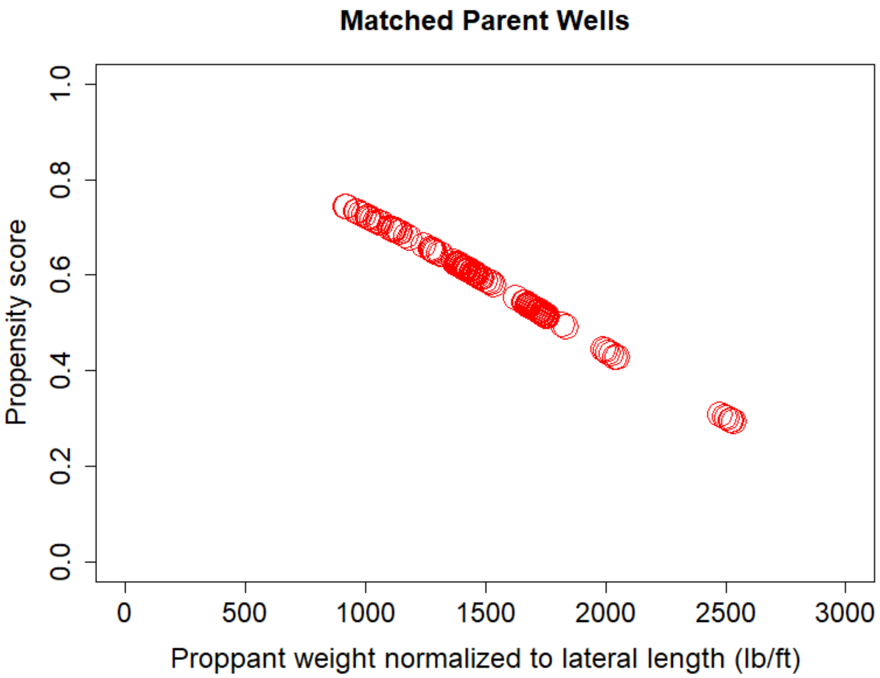
The expression in Eq. 2.11 signifies that we are taking the difference in the propensity score as a measure of distance between the wells used for one-to-one matching. The wells which do not find a match in this iterative matching process are eliminated. We start this process by taking the difference in propensity scores given by $\|(\pi(\boldsymbol{X}))_j - (\pi(\boldsymbol{X}))_i\|$ between a parent well j (T=0) and a child well i (T=1) for all wells in the parent and child well group (Olmos and Govindasamy 2015).Pairs with the smallest difference are considered a good match and retained for subsequent causal analyses. Once the wells are matched to each other, they are not considered for subsequent iterations in the matching process.

$$\text{argmin}_{j:Tj\neq Ti}\ \|(\pi(\boldsymbol{X}))_j - (\pi(\boldsymbol{X}))_i\|,\ j \in T = 0, i \in T = 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.11)$$

Let us now look at the plots of propensity scores versus a single covariate such as proppant weight, fracture fluid volume normalized to lateral length or true vertical depth, after we have performed the 1:1 matching process. In **Figs. 2.19**, **2.20** and **2.21**, we can see that the density of the propensity scores between the parent and child well groups look very similar or almost identical with regards to each of 3 covariates. The reason behind this is that we have discarded the wells in either group that do not have a matching well in the other group leading to improved balance between the two groups. In order to verify this finding, we move to step 4 in the workflow to reassess the balance.

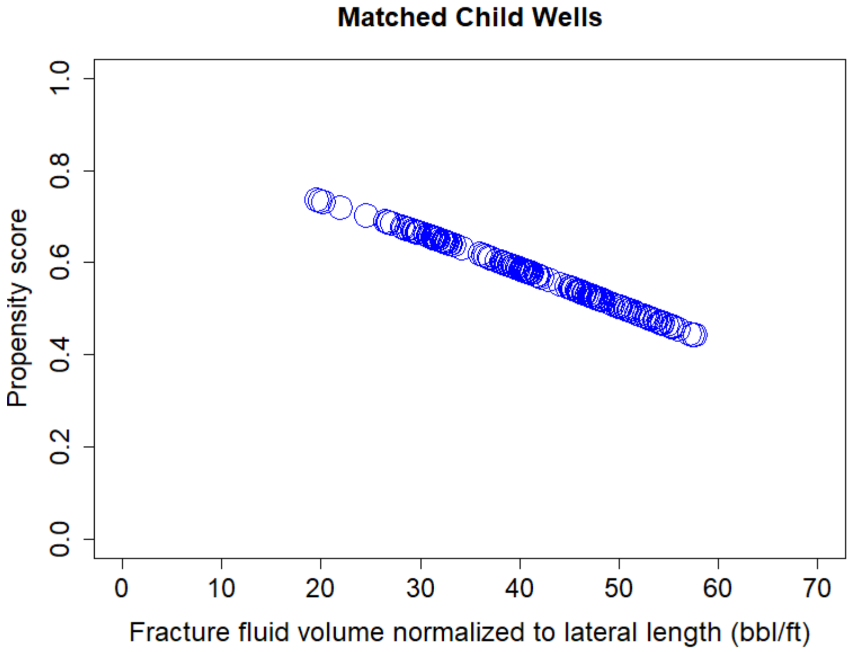


(a)

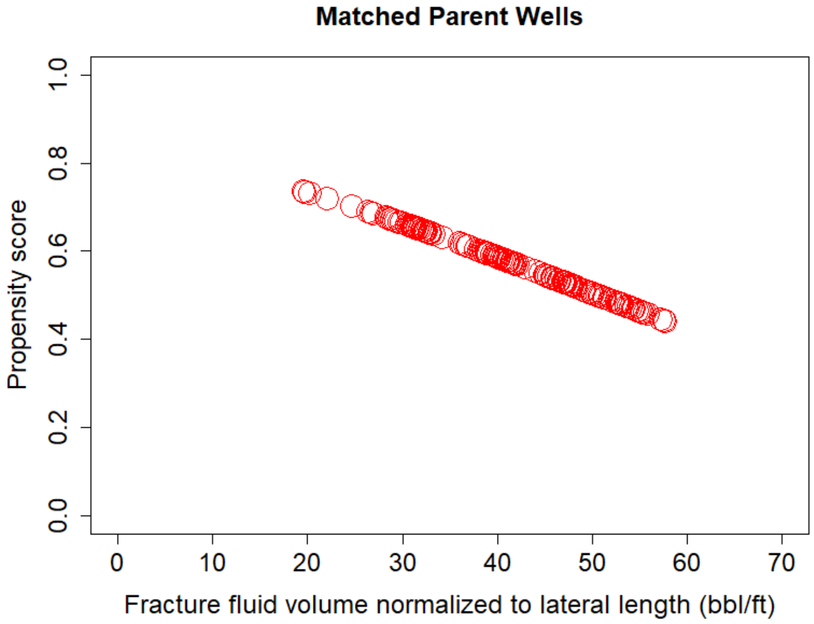


(b)

Fig. 2. 19—(a) Plot of propensity score versus proppant weight normalized to lateral length (lb/ft) for matched child wells (b) Plot of propensity score versus proppant weight normalized to lateral length (lb/ft) for matched parent wells.


Matched Child Wells

(a)


Matched Parent Wells

(b)

Fig. 2. 20—(a) Plot of propensity score versus fracture fluid volume normalized to lateral length (bbl/ft) for matched child wells (b) Plot of propensity score versus fracture fluid volume normalized to lateral length (bbl/ft) for matched parent wells.

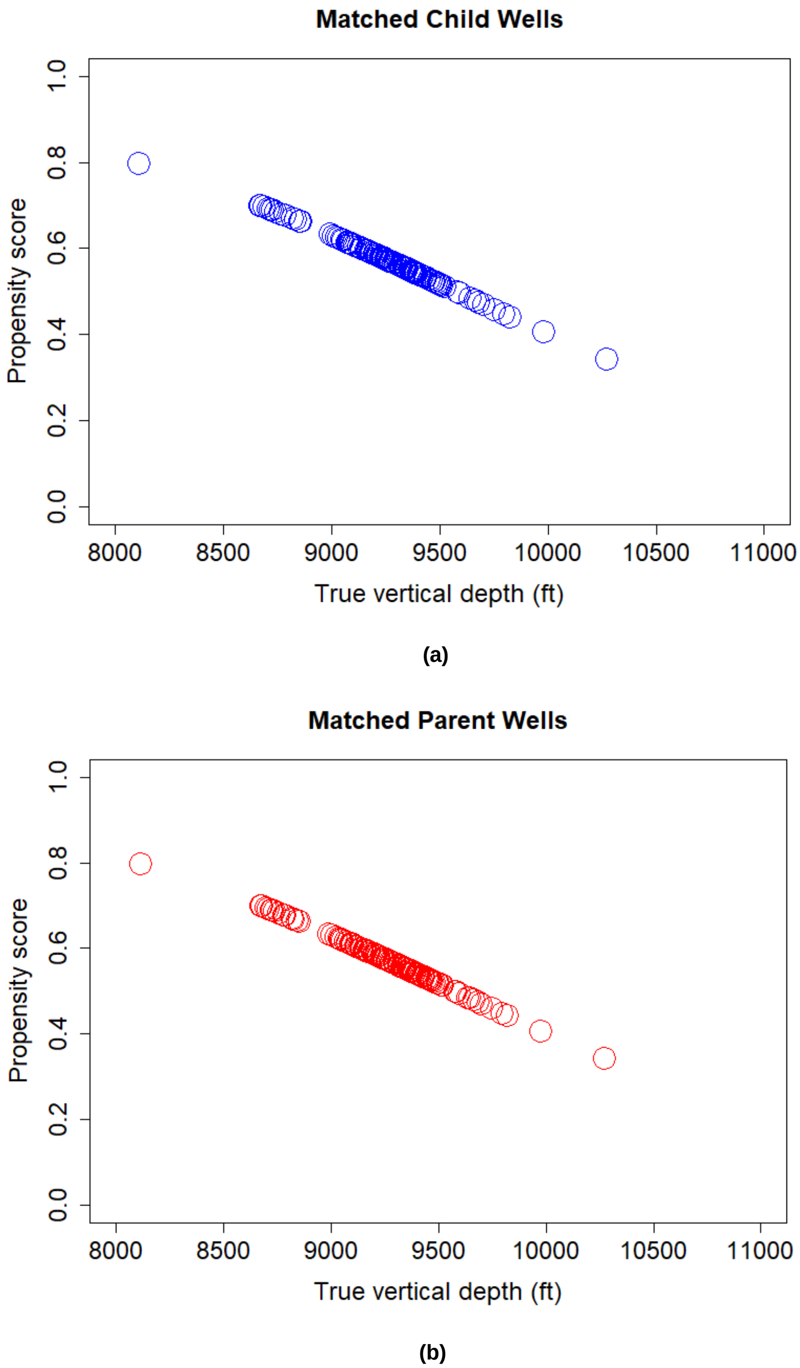


Fig. 2. 21 —(a) Plot of propensity score versus true vertical depth (ft) for matched child wells (b) Plot of propensity score versus true vertical depth (ft) for matched parent wells.

4. Reassessing the Balance

The fourth step is to compare the propensity score distributions for the parent and child wells groups after the process of one-to-one matching is complete. **Fig. 2.22** shows the density of the propensity scores between the matched and unmatched parent and child well groups at 600 ft spacing. As the propensity score is a probability estimate, the density clusters are bounded between 0 and 1. We can easily observe that the propensity score cluster density, in the center of Fig. 2.22, for the matched parent and child well groups look similar to each other. The wells for which a match was not obtained are called unmatched treatment units or control units at the top and bottom of Fig. 2.22. These wells are the ones to be discarded from the causal inference workflow.

In **Fig. 2.23**, after the matching is performed, we can check and compare the propensity score distributions for the matched child wells group as seen in Fig. 2.23a and matched parent wells group as seen in Fig. 2.23b. Since both distributions look the same or are identical, there is a great balance between the parent and child well groups with statistically similar parent and child well pairs. Thus, I have controlled for the variations in heterogenous reservoir properties, different fluid volume, proppant weight and landing zones for both well groups.

The propensity density plots and propensity score distributions for 500 ft, 600 ft, 700 ft and 800 ft spacing options, before and after matching, are shown in the Appendix from **Fig. A1 to Fig. A16.**

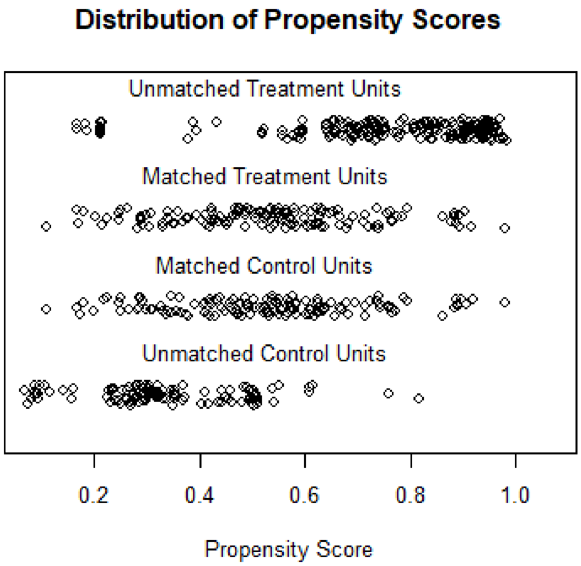
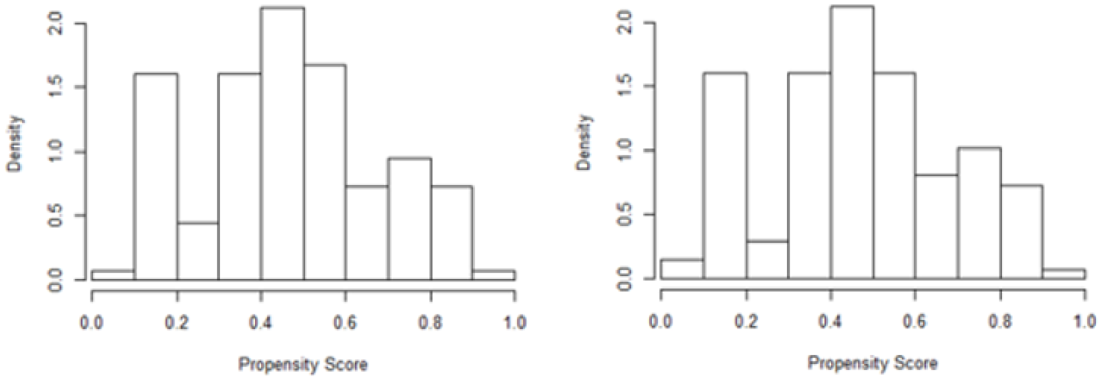


Fig. 2. 22—Plot illustrating the density of propensity scores for wells after matching.



(a) (b)

Fig. 2. 23—(a) Distribution of propensity score for matched child wells (b) Distribution of propensity score for matched parent wells (Sakhardande and Devegowda 2021).

Here again, we can check for the balance after matching by computing the standardized mean differences (SMD) for all the individual covariates for both child and parent well groups. In the **Table. 2.4**, the standardized mean differences between the two groups (Child wells denoted by "1" and Parent wells denoted by "0") after matching are shown in the last column for all the covariates used in this study at well spacing of 700 ft.

As we can see in Table 2.4, the SMD values are below 0.1 which indicates that we have achieved a much better balance within the covariates of both groups at the individual covariate level. Another point to notice is that the number of wells in both groups at the 700 ft well spacing option is now 139 which shows the fact that we have discarded some statistically dissimilar wells from both groups through the matching process.

	139 Parent Wells (0)		139 Child Wells (1)		
Covariates	Mean	Standard Deviation	Mean	Standard Deviation	SMD
P1	1.77	12.53	1.17	11.54	0.049
P2	2.92	8.77	2.31	8.42	0.071
P3	0.4	7.6	0.37	7.02	0.004
P4	0.06	4.66	0.2	5.77	0.028
P5	0.01	4.66	-0.07	4.59	0.015
Proppant weight normalized to lateral length (lb/ft)	1513.92	349.39	1529	317.94	0.045
Fracture fluid volume normalized to lateral length (bbl/ft)	40.49	9.4	41.4	9.33	0.101
TVD (ft)	9226.38	434.44	9254	402.07	0.065

Table 2. 4— Standardized mean differences (SMD) between parent (0) and child wells (1) groups at individual covariate levels (after matching).

The number of wells with 180-day and 360-day of production history in the parent well and the child well groups before and after matching can be seen in **Table 2.5 and 2.6**. Before matching, the number of wells in each group is decided based on the 3-dimensional distance and time criteria explained earlier. For the outcome of 180-day cumulative production normalized to lateral length, I had 725 wells initially and 605 wells that had 360-day cumulative production data normalized to lateral length.

As we can see in Table 2.5 and 2.6 after matching, a few wells in both parent and child groups are discarded or eliminated through the matching process. Lunt (2014) showed that the use of a "caliper" for the matching process is essential for achieving great balance between treatment and control groups. A caliper is a measure of tolerance that can be allowed for the propensity score distance measure. The value of a caliper or the maximum permitted difference between the matches is typically user-defined and often between 0.01 to 0.1. Therefore, tighter the tolerance, smaller is the value of the caliper which leads to closeness in the matches obtained. A tighter tolerance can however significantly reduce the number of wells. Here, I iteratively check the quality of the balance as seen in Fig. 2.23 by adjusting the caliper such that the best possible balance is obtained between the parent and child wells groups. Only when the best possible balance is achieved between the two groups, we can proceed to the final step in this workflow which is estimating the causal effect.

Well Spacing Options considered	Number of Parent Wells (Before Matching)	Number of Child Wells (Before Matching)	Number of Parent Wells (After Matching)	Number of Child Wells (After Matching)	Choice of the Caliper value selected for Matching
500 ft	501	224	157	157	0.02
600 ft	346	379	168	168	0.0165
700 ft	302	423	193	193	0.015
800 ft	183	542	136	136	0.0175

Table 2. 5— The number of parent wells and child wells before and after matching along with the choice of the caliper used for obtaining the best possible balance in the matching process for the 4 well spacing options considered with outcome cumulative 180-day production normalized to lateral length.

Well Spacing Options considered	Number of Parent Wells (Before Matching)	Number of Child Wells (Before Matching)	Number of Parent Wells (After Matching)	Number of Child Wells (After Matching)	Choice of the Caliper value selected for Matching
500 ft	384	221	113	113	0.0125
600 ft	306	299	137	137	0.05
700 ft	256	349	139	139	0.0175
800 ft	166	439	105	105	0.015

Table 2. 6—The number of parent wells and child wells before and after matching along with the choice of the caliper used for obtaining the best possible balance in the matching process for the 4 well spacing options considered with outcome cumulative 360-day production normalized to lateral length.

5.Causal Effect Estimation

The final step in the causal inference workflow is the estimation of the causal effect well spacing on well performance. As mentioned earlier, causal effects are differences in the means of potential outcomes for both groups (Morgan and Winship 2014; Imbens and Rubin 2015). As a reminder, the potential outcomes are the calculated responses for treatment subjects if they did not receive the treatment and the calculated responses for control subjects if they received the treatment.

In order to predict the potential outcomes, I build a random forest regression model (Ho 1995; Breiman 2001) between the outcomes: 180-day and 360-day cumulative production numbers normalized to lateral length for the matched child wells and their covariates. Then, I predict the potential outcome using matched parent well covariates. In a similar manner, I build another random forest regression model between the outcomes: 180-day and 360-day cumulative production numbers normalized to lateral length for the matched parent wells and their covariates. Then, I fit this model with child wells (matched) covariates. This model gives the predicted potential outcomes for the parent wells. The causal effect of well spacing on well performance is given by the group mean difference between these two predicted potential outcomes for the child and parent well groups (Ho et al. 2011).

Fig. 2.24 shows the predicted potential outcomes of the matched child wells at 700 ft well versus their actual outcome.

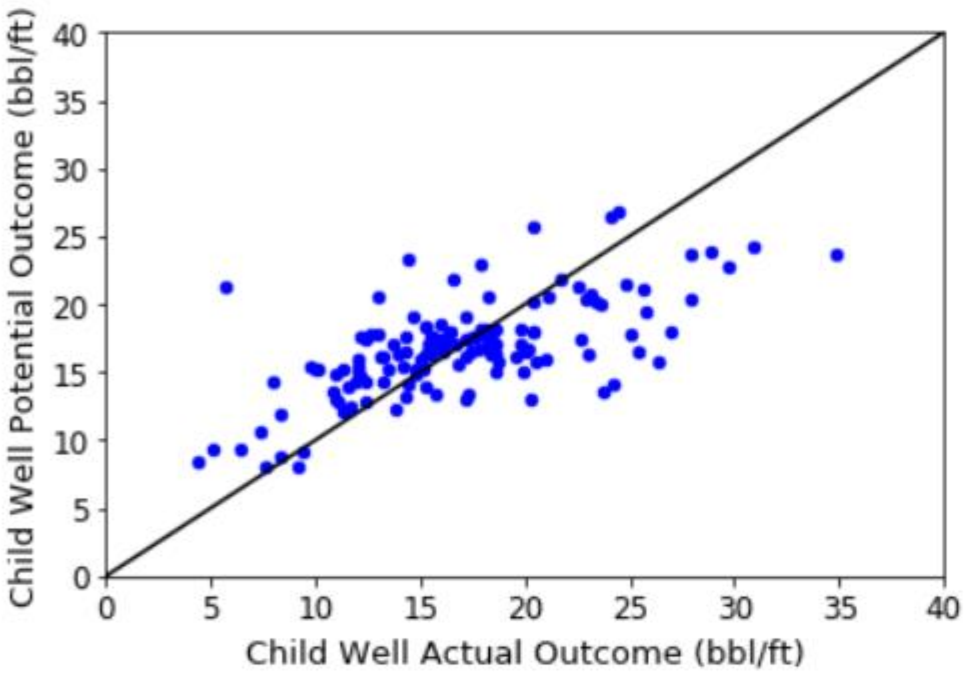


Fig. 2. 24— Plot of child well potential outcome (bbl/ft) versus child well actual outcome (bbl/ft).

Fig. 2.25 shows the predicted potential outcomes of the matched parent wells at 700 ft well versus their actual outcome.

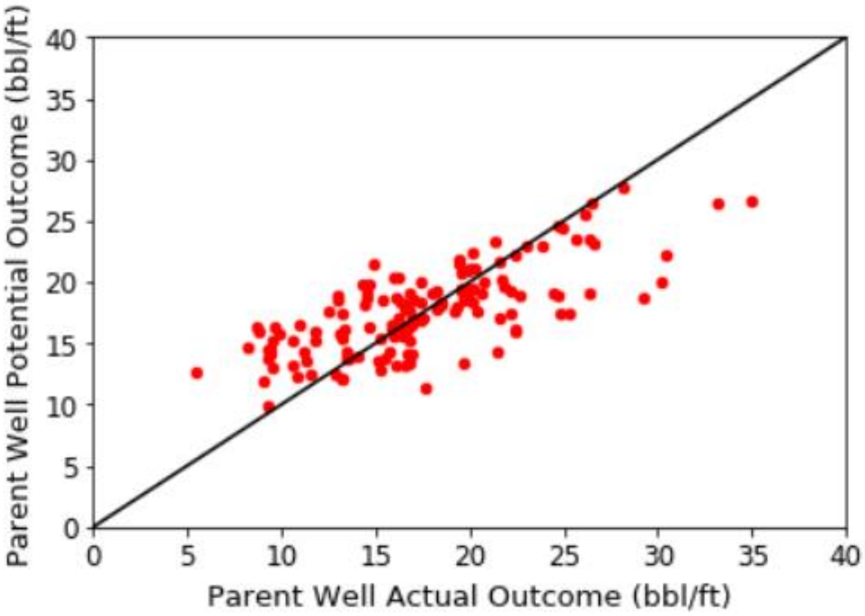


Fig. 2. 25—Plot of parent well potential outcome (bbl/ft) versus parent well actual outcome (bbl/ft).

In the next section, I discuss the results from the causal inference workflow comparing it to the naïve, simple averages approach at the four well spacing options considered in this study.

2.8 Results and Discussion

In this section, results with the simple averages approach or naïve approach of finding the well spacing effects are compared with the results from the causal inference workflow.

Simple Averages Approach for 360-Day Cumulative Production Normalized to Lateral Length (bbl/ft)

Fig. 2.26 shows the results from the simple averages approach where the outcomes are computed as the average of the 360-day cumulative production normalized to lateral length for each of the parent and child well groups shown in **Table 2.7**. Fig. 2.26 shows the outcomes on the primary y-axis and the difference in outcomes for parent and child well groups on the secondary y-axis and the four well spacing options considered in this study are shown on the x-axis. Child well group average production in bbl/ft is shown by blue bars and the parent well group average production in bbl/ft is shown by red bars. The black error bars show their standard deviations at each well spacing (Sakhardande and Devegowda 2021).

Well Spacing Options considered	Number of Parent Wells (Unmatched)	Number of Child Wells (Unmatched)
500 ft	384	221
600 ft	306	299
700 ft	256	349
800 ft	166	439

Table 2. 7—The number of unmatched parent wells and child wells at the 4 well spacing options used in the simple averages approach for outcome cumulative 180-day and 360-day production normalized to lateral length.

Before discussing further, I make a note here emphasizing that an overlap in error bars does not let us conclude whether difference between the means of groups is statistically significant or not. In order to check the significance, one needs to conduct a paired t-test (Lanzante 2005).

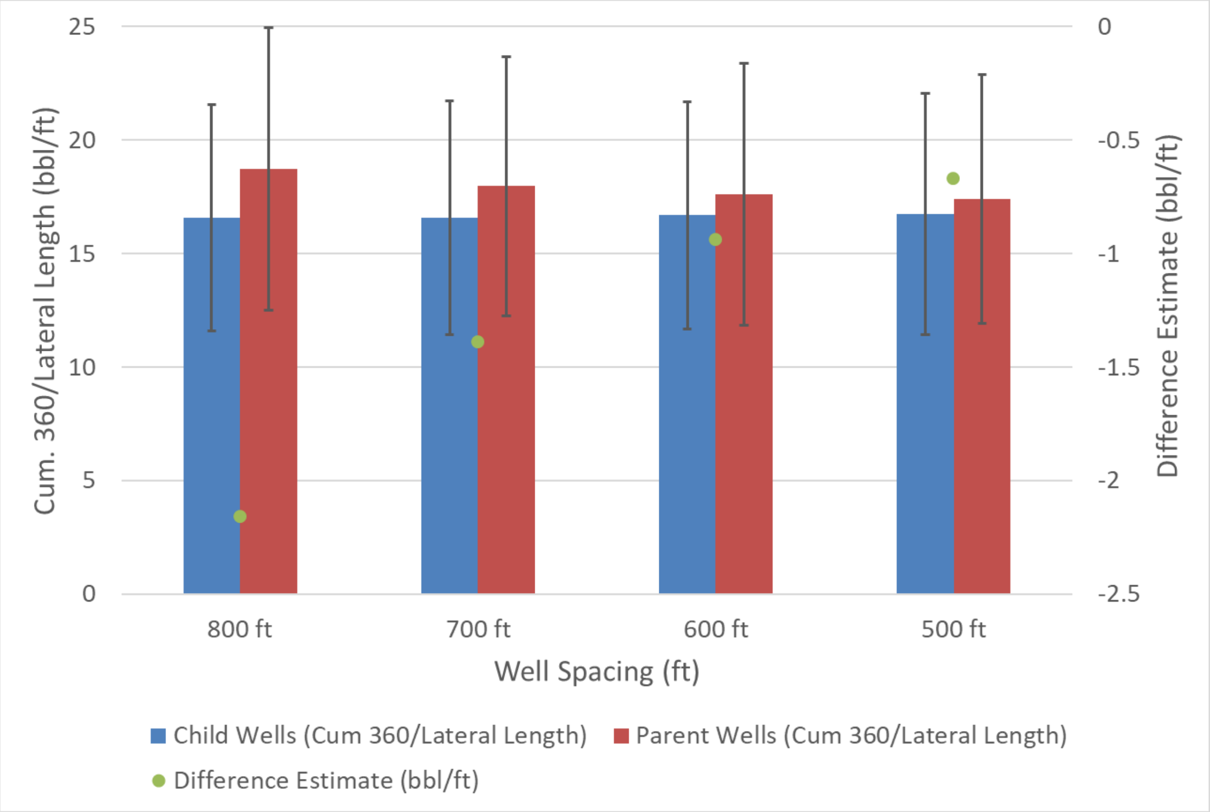


Fig. 2. 26— Bar chart comparing 360-day cumulative production normalized to lateral length of parent-child wells and difference estimates for each well spacing option from the routinely used simple average approach.

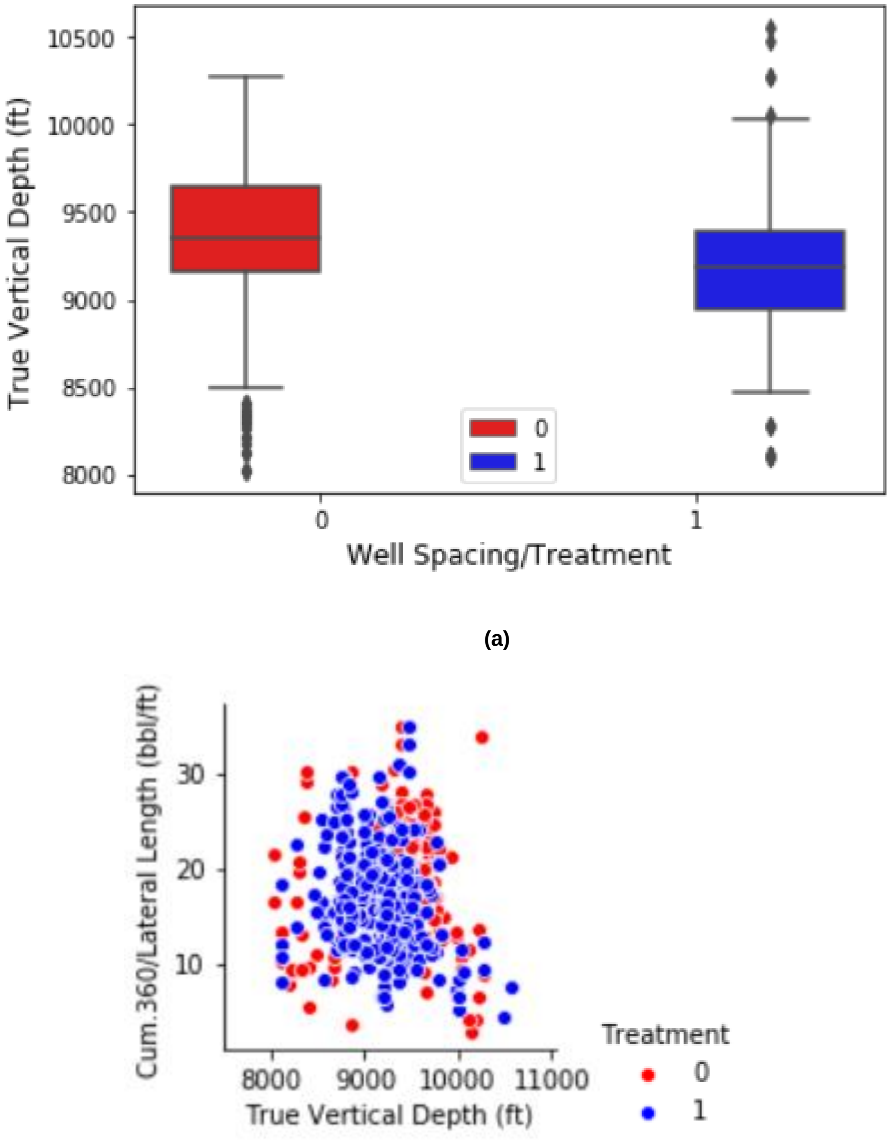
We can observe that at all spacing values, parent wells appear to be doing better than the child wells. This is shown by the difference in these simple averages of 360-day cumulative production normalized to lateral length (bbl/ft) (the green dots corresponding to the secondary y-axis) in Fig.2.26. However, there is no appreciable difference between the performance of the parent wells at different well spacings nor is there a difference between child well performance at varying spacings. As an example, in **Table 2.8**, I show the output of a paired t-test conducted between a parent well at 500 ft well spacing and another parent well at 700 ft well spacing. The results show a p-value of more than 0.05 indicating that a simple averages approach does not yield a statistically significant difference between parent well performance at different spacing levels. There is also no statistically meaningful difference in child well performance at less than 500 ft spacing and less than 700 ft spacing either.

Well groups at 360-day cumulative production normalized to lateral length (bbl/ft)	Degrees of Freedom	P-value	Mean of 360-day cumulative production normalized to lateral length (bbl/ft)	Mean of 360-day cumulative production normalized to lateral length (bbl/ft)	95 % confidence interval
500 ft parent and 700 ft parent	638	0.21	17.4	17.9	(-1.42, 0.3)
500 ft child and 700 ft child	568	0.72	16.73	16.57	(-0.71,1.3)

Table 2. 8—Paired t-test output comparing the actual outcomes for 500 ft parent with 700 ft parent well group and 500 ft child well with 700 ft child well group using simple averages approach.

These results show that the simple averages approach provides counter-intuitive results. We would expect that as well spacing increases, well performance metrics should show a corresponding increase because of the access to more undrained reservoir volume.

To again illustrate that the naïve, simple averages approach is not an ideal way of calculating well spacing effects, consider **Fig. 2.27**. Boxplots for TVD and a scatter plot of 360-day cumulative production normalized to lateral length for the parent (shown in red) and child (shown in blue) wells are shown in Fig. 2.27a and b. respectively. The range of the landing zones for the parent wells is seen to be considerably different from the range for the child wells. Using the simple averages approach, we are considering wells completed in different landing zones. This can be significant in multi-bench plays such as the Permian and the Meramec (Gaurav et al. 2017; Price et al. 2017). This is the motivation behind taking a more sophisticated causal inference approach where we would control for reservoir and completion variables before estimating the spacing impact.



(a)

(b)

Fig. 2. 27—(a) TVD boxplots of parent (0) and child wells (1) (b) Scatter plot of Cum-360-day production normalized to lateral length (bbl/ft) versus TVD (ft) for parent (0) and child wells (1) at 500 ft spacing (Sakhardande and Devegowda 2021).

Also, in Fig. 2.27, we observe parent wells associated with a much wider TVD range and child wells in a narrower range. This, again, is to be expected. As we gain more knowledge of the reservoir, our well placement becomes more focused and optimized. However, if we were to use a simple-averages approach, we would be comparing wells from different benches/horizons and this can impact our spacing analyses. Although we are controlling for reservoir properties, it is also important to separately consider TVD. Different landing zones may be associated with different frac-barriers. Even if the reservoir properties and completions are similar for a parent and child well in 2 different benches, there may be different levels of fracture growth in both of them, based on the thickness of the zone and presence/absence of frac-barriers.

Causal Inference Approach for 360-Day Cumulative Production Normalized to Lateral Length (bbl/ft)

Given this background and the potentially compromised interpretations from a simple averages approach, I now show the results from the causal inference workflow in **Fig. 2.28.** The primary y-axis and the x-axis in Fig. 2.28 are same as Fig. 26 shown earlier. The causal estimate is shown on the secondary y-axis in bbl/ft. The casual estimate is the difference between the potential outcomes of parent and child well groups. Blue bars and red bars show the potential outcome mean values of cumulative 360-day production numbers normalized to lateral length for child wells and parent wells respectively. Standard deviation is shown by black error bars.

Considering the 800 ft well spacing in Fig. 2.28, the mean of the potential outcomes for the parent wells is around 19 bbl/ft and 16 bbl/ft for the child wells. The difference of 2.6 bbl/ft is the causal estimate which is shown by the green circles with the green error bars representing the 95% confidence interval. This implies that the parent well at 800 ft produces 2.6 bbl/ft of oil more than the child well. The results for all paired t-tests between the parent and child well group potential outcomes, that is, the predicted cumulative 360-day production numbers normalized to lateral length (bbl/ft), at all well spacings (500 ft, 600 ft,700 ft and 800 ft) are summarized in the Appendix **Table A2.** All the differences can be seen to be statistically significant.

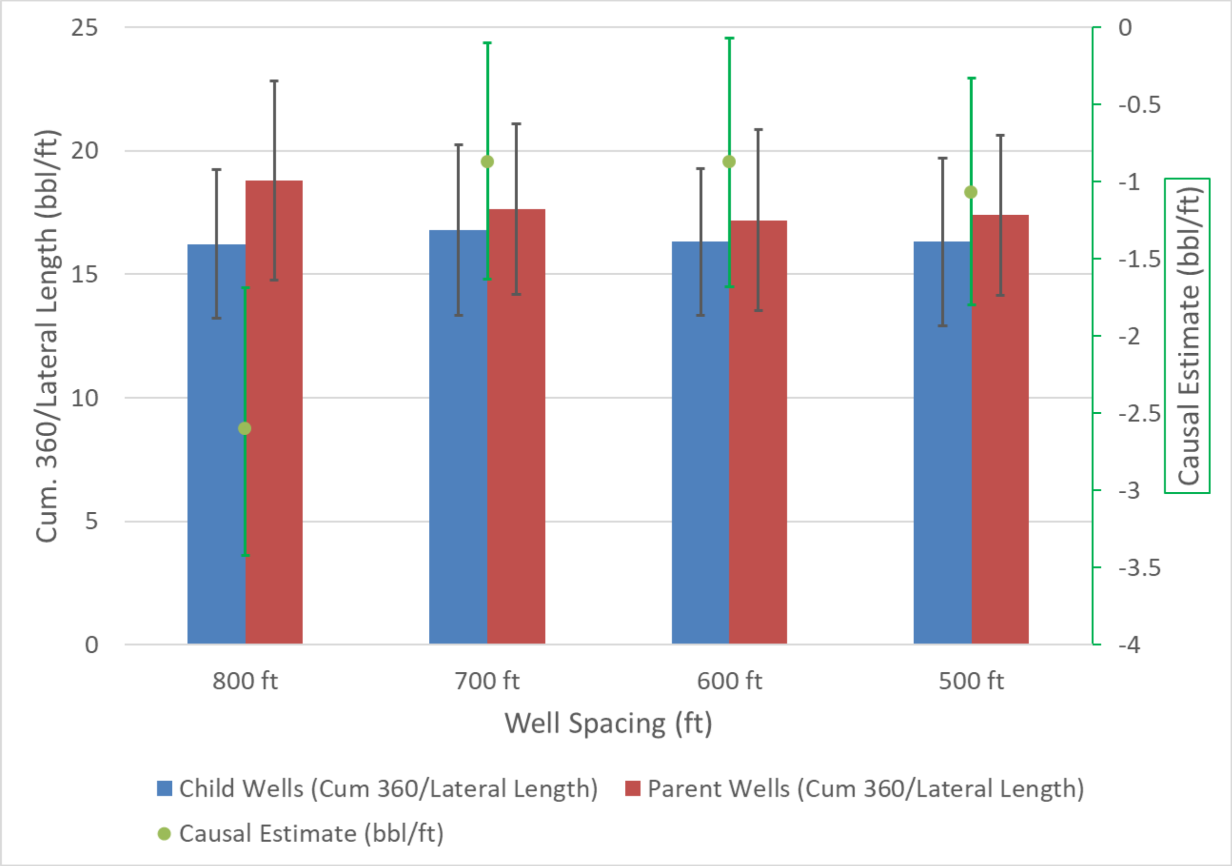


Fig. 2. 28—Bar chart comparing potential outcomes for 360-day cumulative production normalized to lateral length (bbl/ft) of parent-child well groups and causal estimates for each well spacing option using causal inference (Sakhardande and Devegowda 2021).

The parent well at 800 ft well spacing in Fig. 2.28 performs the best relative to the child well given that the magnitude of causal estimate is the largest. The difference in the causal estimates at the 500 ft, 600 ft and 700 ft well spacing levels are not as large as 2.6 bbl/ft. This result is expected because wells spaced farther than 800 ft will likely have more access to virgin reservoir pressure and undrained reservoir volume compared to wells with spacing more than just 500 ft. This indicates that at the much tighter spacing levels of 500 ft, 600 ft and 700 ft, the inter-well interference problem is unavoidable even for the parent wells.

Another paired t-test confirms that the 800 ft parent well performs significantly better than the 500 ft parent well. An example output of this paired t-test comparing the potential outcomes for the parent-parent well groups is shown in **Table 2.9**. Here we can see that the difference is statistically significant based on the p-value and 95% confidence interval at the 360-day mark potential outcome. Other comparisons coupled with paired t-test outputs between the 800 ft spacing parent well group with the rest of well spacing parent well groups were also found to be statistically significant and are shown in the Appendix **Table A4**. Given the data in the study, 800 ft well spacing would then be considered to be the optimal spacing where production interference is the least.

Parent Well groups at 360-days of production	Degrees of Freedom	P-value	Mean of Parent Wells Potential Outcome at 500 ft spacing (bbl/ft)	Mean of Parent Wells Potential Outcome at 500 ft spacing (bbl/ft)	95 % confidence interval
500 ft and 800 ft	216	0.0084	17.39	18.7	(-2.29, -0.34)

Table 2. 9— An example output of a paired t-test between the 500 ft parent well group and 800 ft parent well group with potential outcomes for 360-day cumulative production normalized to lateral length (bbl/ft).

This is an important result of this work. A simple average-based approach does not work because an analysis of Fig. 2.26 shows no difference in parent well performance across all 4 spacing groups indicating that there is no optimal spacing, which is an extremely counter-intuitive result. Well performance should be better at larger spacing levels (up to a point) given the access to larger, undrained reservoir volumes. This reiterates my assertion in Chapter 1 that the simple, average-based approach can potentially compromise quantitative interpretation.

Causal Inference Approach for 180-Day Cumulative Production Normalized to Lateral Length (bbl/ft)

The same causal inference workflow was carried out with 180-day cumulative production normalized to the lateral length (bbl/ft) as the potential outcome and the results are shown in **Fig. 2.29.** The observation that parent wells do better than child wells is borne out even with the first 180 days of production. The causal comparison of parent to child well performance is shown in green circles and are statistically significant at all 4 spacing options considered. The results from all the paired t-tests conducted for the causal estimates of parent-child groups with outcome as 180-day cumulative production normalized to the lateral length (bbl/ft) are shown in the Appendix **Table A1.**

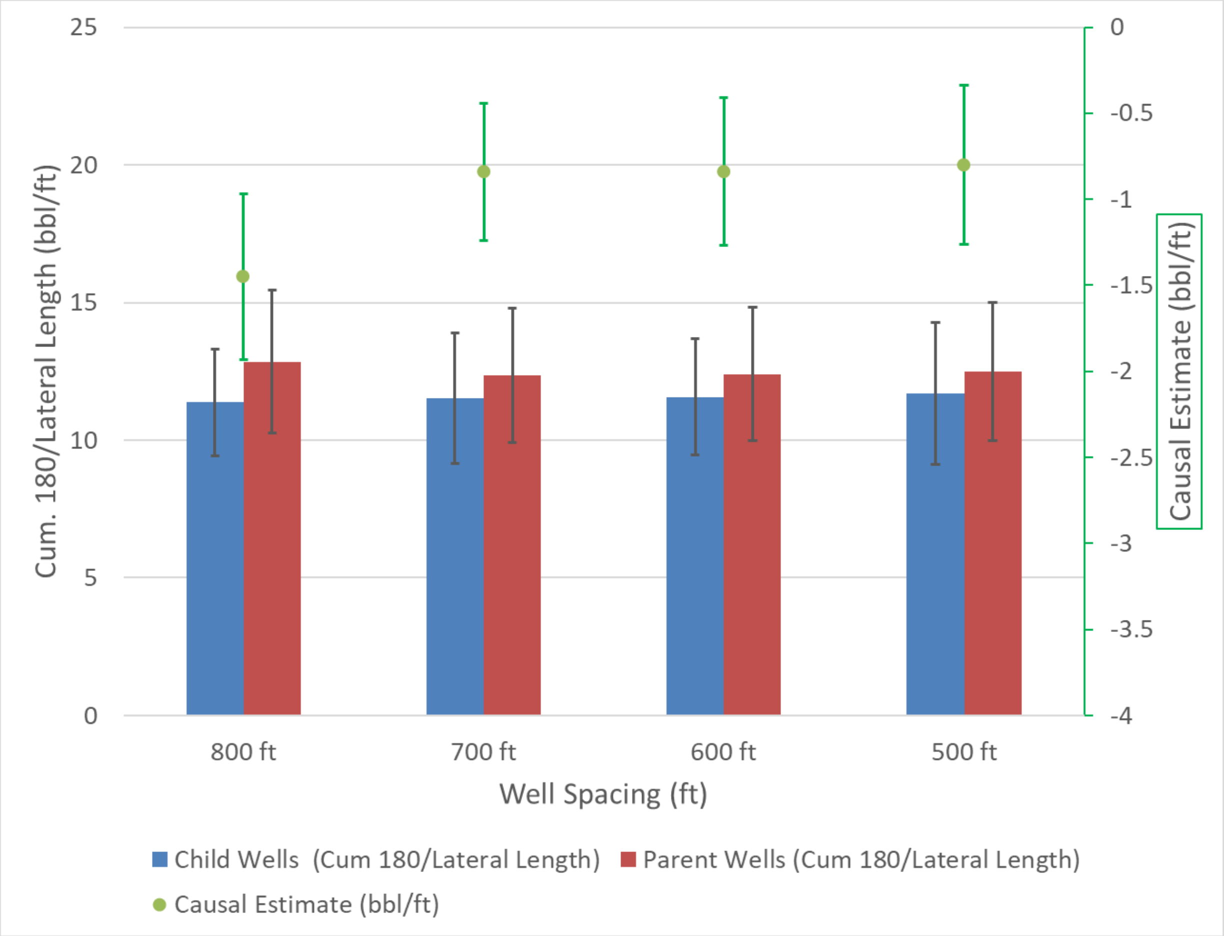


Fig. 2. 29—Bar chart comparing potential outcomes for 180-day cumulative production normalized to lateral length (bbl/ft) of parent-child well groups and causal estimates for each well spacing option using causal inference (Sakhardande and Devegowda 2021).

In Fig. 2.29, it again appears that 800 ft spacing is the optimal spacing value. However, a paired t-test comparison between the 800 ft parent well performance with the 500 ft, 600 ft and 700 ft parent well performance does not show a statistically meaningful difference. **Table 2.10** shows the results of the paired t-test for the 700 ft and 800 ft parent groups. Paired t-tests for all parent-parent well group 180-day performance comparisons are shown in the Appendix **Table A3.**

Parent well groups at 180-day production	Degrees of Freedom	P-value	Mean of Parent Wells at 700 ft spacing (bbl/ft)	Mean of Parent Wells at 800 ft spacing (bbl/ft)	95 % confidence interval
700 ft and 800 ft	327	0.078	12.35	12.84	(-1.04,0.055)

Table 2. 10— Output from a paired t-test comparing 700 ft well spacing parent wells with 800 ft spacing parent wells for the outcome 180-day cumulative production normalized to lateral length (bbl/ft).

This is another important conclusion. If the parent well 180-day cumulative production normalized to lateral length at the 500 ft, 600 ft, 700 ft and 800 ft spacing levels are not statistically different, then it implies that at the 180-day point, wells spaced 500 ft apart are not seeing any appreciable interference over and above the interference for wells spaced at 800 ft apart. Because we now know that the 500 ft spacing is the least optimal when evaluating 360-day cumulative oil production, the interference is likely to begin only after the 6-month point, on average. It also indicates that the assumptions behind placing wells in parent or child well groups are validated by the data.

Therefore, I cannot confirm that 800 ft spacing is optimal at 180-day mark of production. However, at the 360-day mark, the individual paired t-tests comparing the potential outcomes of the 800 ft parent well and the 500 ft, 600 ft and 700 ft parent wells show that the difference in the potential outcome means is statistically significant as mentioned earlier. Thus, from this case study, we can say that 800 ft well spacing indeed is optimal at the 360-day mark of production among parent and child wells (Sakhardande and Devegowda 2021). Although there may not be enough child wells on production at the same time when a parent well is on production at the 360-day mark but for a parent well, there are enough other parent wells on production. So, the casual inference results with outcome 180-day and 360-day production are valid for the parent wells.

2.9 Conclusions

In this case study, I have shown how the routinely used spreadsheet-based approaches of simple averages can lead to counter-intuitive results for estimating well spacing effects. The main reason is that the simple averages approach control for the complex interplay of reservoir properties, well completion designs and well landing zones. This study introduces a new, promising workflow known as causal inference which provides a more appropriate method to assess the well spacing effect on well performance. Through the casual inference workflow, one can control for the complex variations seen in all the measured variables known as covariates which influence the spacing decision as well as impact the parent and child well production performance. The conclusions from this chapter can be summarized as follows:

1. The parent well at 800 ft spacing is considered as optimal. This is confirmed by paired t-tests production performance comparisons between 800 ft spacing parent well and the rest of the spacing options considered.
2. The statistically significant causal estimates at 180-day mark and 360-day mark of production confirm that parent wells perform better than the child wells with respect to normalized cumulative production numbers at all four well spacing options considered.
3. The search and selection of wells, into the parent and child well groups, is subjective yet very crucial for the causal inference workflow (Sakhardande and Devegowda 2021).
4. True vertical depth is found to be very important as a covariate for the finding causal effects of well spacing on well performance for this specific multi-bench shale play.

Chapter 3: Anomaly Detection in Sucker Rod Pumps

3.1 Introduction

Wells tend to flow naturally when the reservoir pressure is sufficient to overcome hydrostatic and frictional losses in the production piping. As reservoir pressure declines due to depletion, the natural drive of the well must be supplemented with additional energy to sustain production and/or prevent liquid loading. Artificial lift systems are designed to provide this additional energy in order to lift the fluids to the surface and maintain economically viable production rates.

Of the over 2 million producing oil wells around the world, more than 50% are operated on artificial lifts (Lea 2007). This underscores the significance of artificial lift in the industry. The oldest and most widely used artificial lift method is the Sucker Rod Pumps (SRP). SRP's are typically used for onshore wells with depths ranging from 1000 ft to 16000 ft with a moderate lifting capacity ranging anywhere between 1 to 1000 bbl/day. This chapter is focused on discussion of SRP's basic working principle along with its major components, use of dynamometer pump cards and the application of various anomaly detection techniques to SRP failure modes.

Fig. 3.1 shows a schematic of the basic components of a Sucker Rod Pump system. The surface equipment consists of a prime mover system, gear reducers, pumping unit and a polished rod. The prime mover system provides the driving force through an electric motor or combustion gas engine. The gear reducers enable the reduction of rotational motor speed of the prime mover and simultaneously increase the torque at the shaft. The pitman arm coverts the rotational energy from the gear reducers into a reciprocating movement needed for the pump operation.

The polished rod connects the walking beam, which is a mechanical lever of the pumping unit, to the sucker rod string and provides a sealing surface at the wellhead. The equipment downhole are the rod string and plunger pump. The rod string which is enclosed inside the tubing transmits the reciprocating motion from the polished rod to the pump. Main components of the subsurface pump are travelling valve, standing valve, working barrel (stationary part) and the plunger (moving part). The positive displacement pump allows fluids to move from the well to the working barrel which is connected to the bottom end of the rod string.

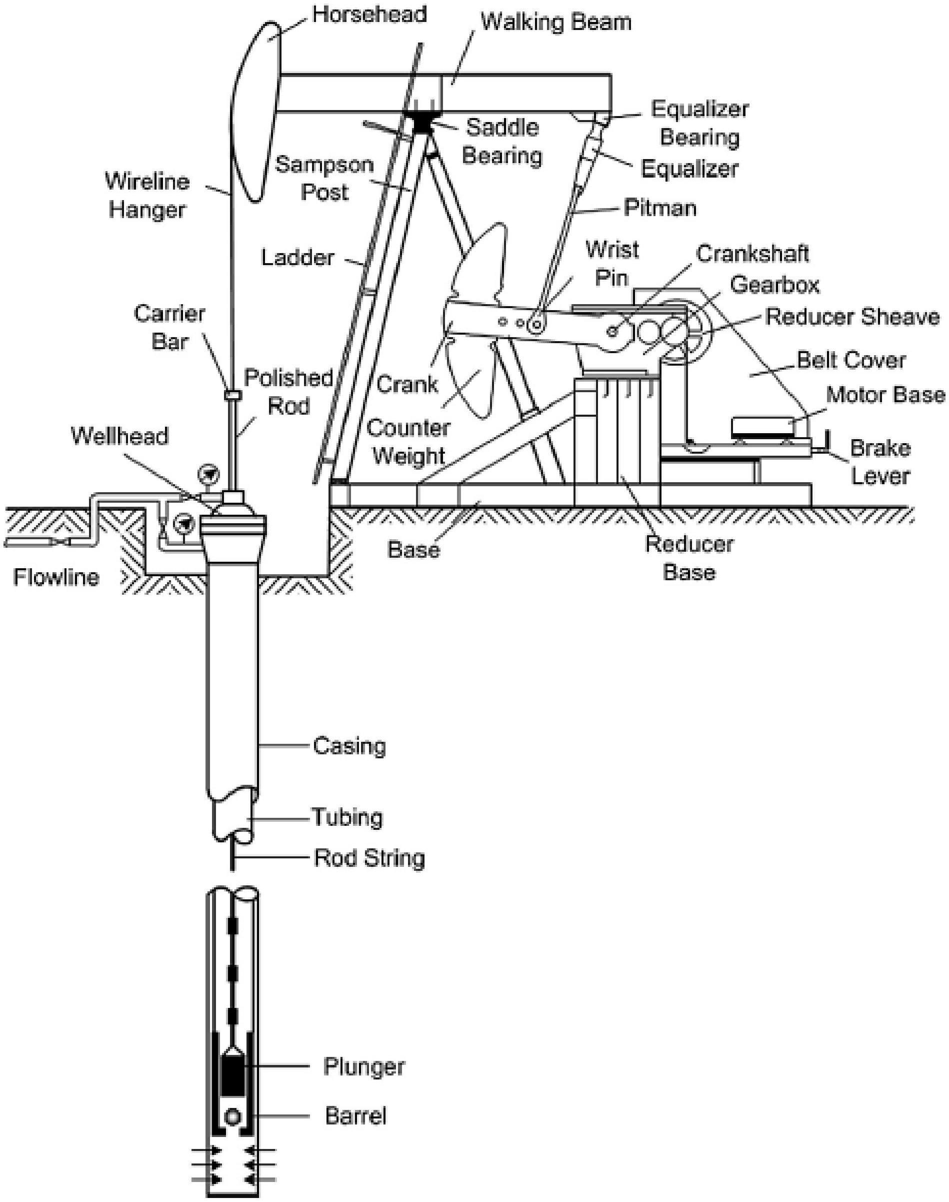


Fig. 3. 1— Components of a Sucker Rod Pump (Golan and Whitson 1991; Takacs 2015).

A typical downhole pump cycle is as shown in **Fig. 3.2**. This cycle consists of an upstroke and downstroke corresponding to the direction of rod string movement and the position of the plunger. The traveling and standing valve are two ball valves and the alternating opening and closing events of these valves along with the reciprocating movement of the plunger enable the lifting of reservoir fluids up to the surface.

Assuming an incompressible fluid, the plunger which is the moving part of the pump assembly reaches the bottom of the pump and the upstroke cycle begins. Owing to the hydrostatic pressure in the tubing above the plunger, the traveling valve closes and the reservoir fluids in the tubing above the traveling valve are carried up to the surface with the plunger moving up. There is a pressure drop between the standing and the traveling valve positions at this stage which keeps the standing valve open as the plunger goes to its topmost position. Due to the pressure differential between the bottomhole flowing pressure and area between the two valves, fluids from the formation enter through the standing valve into the working barrel. The formation fluids fill up the barrel until the upstroke cycle is completed. The fluid column weight is carried by the plunger and the rod string when the traveling valve is closed. This creates tension in the rod string causing it to stretch (Takacs 2015).

At the beginning of downstroke, the plunger is at the top of the pump. As the rod string begins its movement downwards, the traveling valve opens and standing valve closes. The weight of the fluid column is transferred to the standing valve from the plunger. This means the load shifts to the tubing string during the downstroke while the traveling valve is still open. During the downstroke cycle, the pump barrel is filled with the formation fluid and as soon as the cycle ends, the direction of motion of the rod string is reversed. The fluid column weight is again transferred back to the plunger. This stretches the rod string and releases the tension on the tubing string marking the beginning of another upstroke cycle (Takacs 2015).

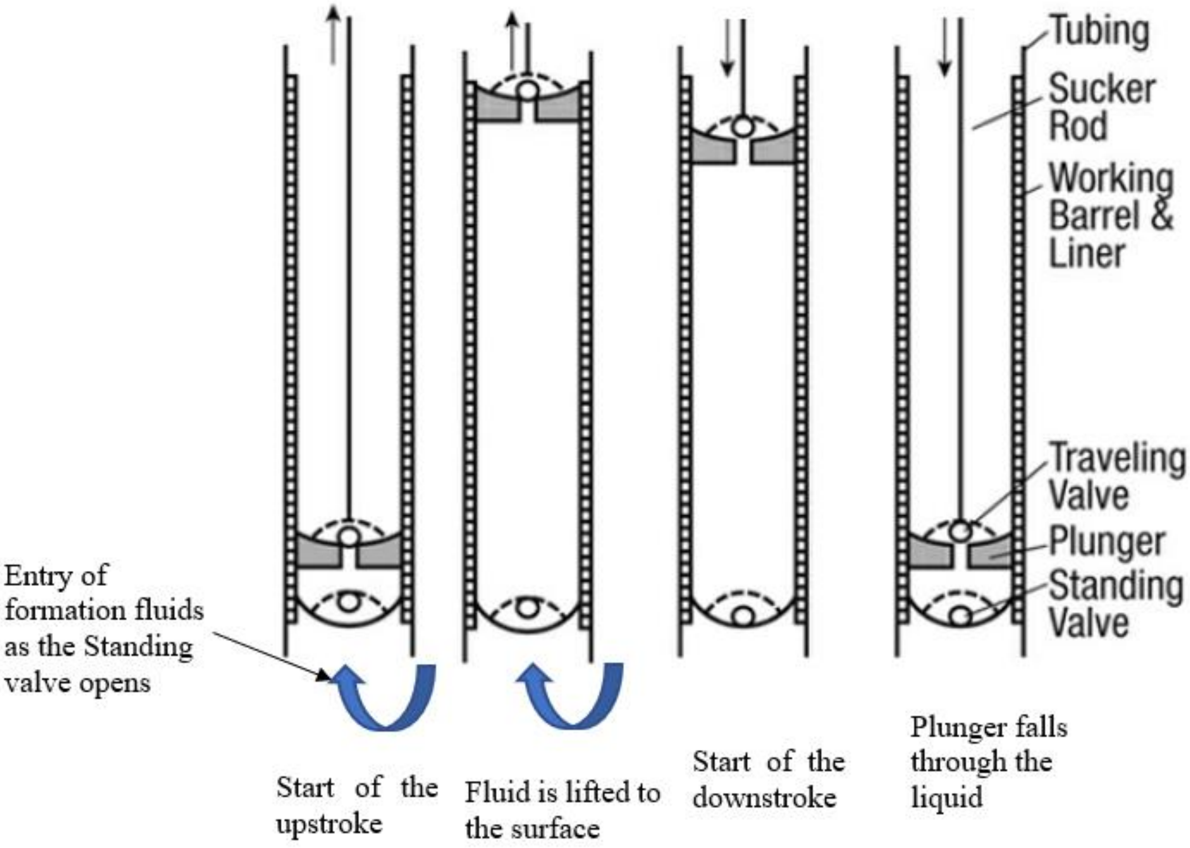


Fig. 3. 2—Typical Sucker Rod Pump cycle (Brown 1980; Apergy 2019).

Complete filling of the pump barrel for a single-phase fluid during the upstroke signifies an ideal pump condition. In the ideal condition, the plunger travels its full length and valves open and close instantaneously. The pump dynagraph or dynamometer card is derived from the surface dynamometer card. The surface dynamometer, which are basically load cells or strain gauges, record the variations in the polished rod loads during a pump cycle (Guo 2017). The surface card generated then is a plot of the polished rod load versus its position. However, the polished load is a function of all system components in the sucker rod pump system and the surface plots generated are complex due to its non-standardized behavior. Therefore, it becomes very difficult to interpret for the production engineers.

Fortunately, this problem was solved by Gibbs (1963). Dr. Sam Gibbs provided a linear second order hyperbolic partial differential equation which when solved numerically, using finite difference equations, mathematically models the elastic nature of the rod string and the forces that act on it during a pump cycle. Gibb's equation shown in Eq. 3.1 is popularly known as the Wave Equation. It is used to convert the surface conditions to downhole pump conditions. Hence, these downhole dynamometer cards are also referred as pump cards (Takacs 2015). The pump cards exhibit a standardized behavior for each downhole pump condition and are used for pump diagnostics.

The expression in Eq. 3.1 is the wave equation.

$$\frac{\partial^2}{\partial t^2}u(x,t) = {v_s}^2\frac{\partial^2}{\partial x^2}u(x,t) - c\frac{\partial u(x,t)}{\partial t} \qquad 3.1$$

where $u(x,t)$= rod displacement, ft.

c=viscous damping factor, 1/s.

$v_s = \sqrt{\frac{144 g_c E}{\rho}}$= velocity of sound in the rod material, ft/s.

x = position of rods, ft.

t = time, s.

An example of a pump card and surface card at normal operating conditions is shown in **Fig. 3.3.** The pump card plot shows the rod string load (lbs) on the y-axis plotted versus the rod displacement (inch) on the x-axis.

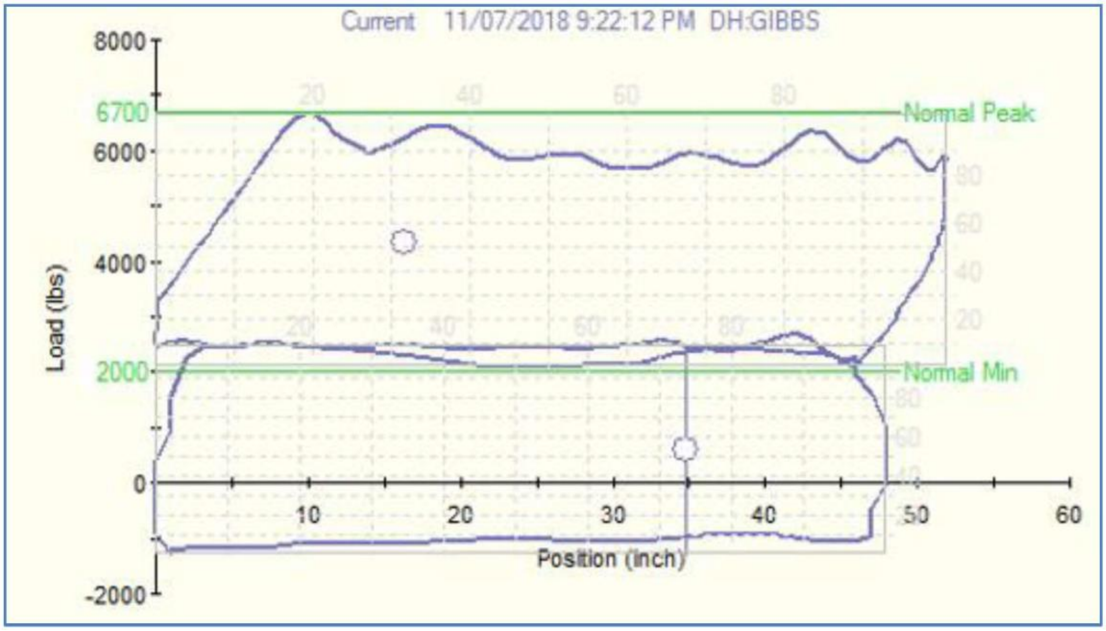


Fig. 3. 3—Pump card (bottom) and surface card (top) for a normal operating condition (Sharaf et al. 2019).

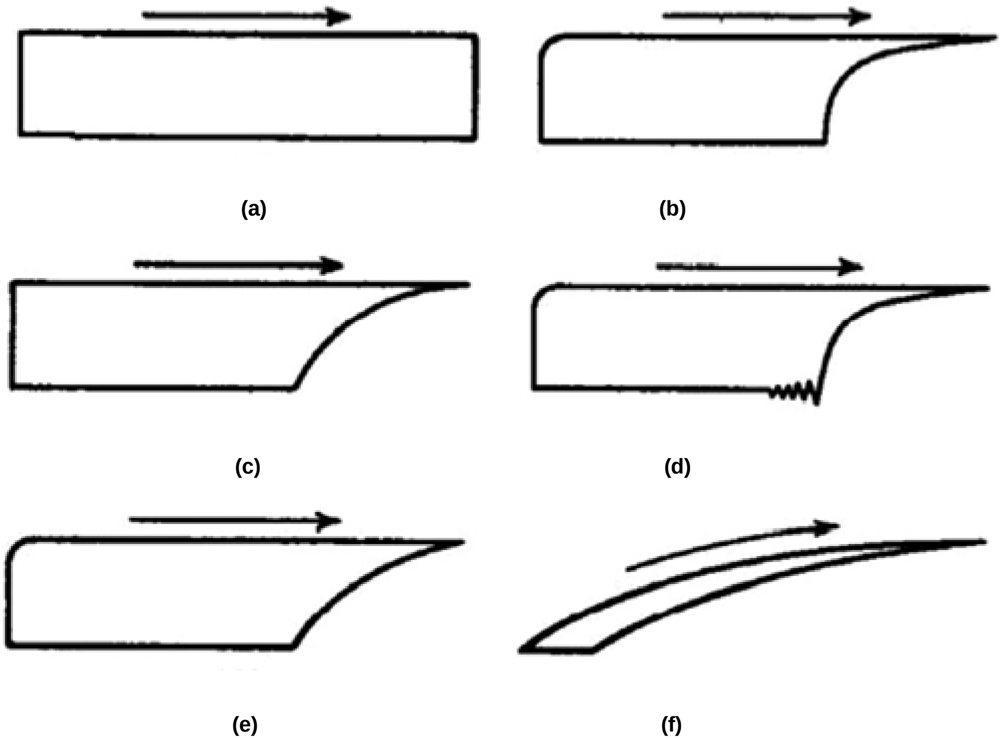


Fig. 3. 4—Pump Dynamometer cards (a) Ideal Pump Card (b) Fluid Pound (c) Gas Compression on Downstroke (d) Vibration during Fluid Pound (e) Gas Expansion on Upstroke (f) Gas Lock (Nind 1964).

A normal pump cycle is indicated by a pump dynamometer card in **Fig. 3.4a**. The rest of the figures in Fig. 3.4 highlight a few of the non-ideal downhole pump conditions. For example, Fig. 3.4c shows a case of gas compression during the downstroke of the plunger. This typically occurs when there is free gas present in the pump during the upstroke. During the downstroke, this free gas undergoes compression and the traveling valve is closed temporarily. This is indicated by slight drop in the rod displacement. Here the pump performance stays erratic until the gas is completely displaced out to the surface.

A much more severe case of gas interference is shown in Fig. 3.4f which is called a gas lock. Here the pump efficiency drops significantly as most of the energy is spent in compressing and expanding the gas phase and hardly any liquid is pumped up to the surface. Another non-ideal case is indicated in Fig. 3.4b where there is a pumped-off condition when the pump displacement is greater than the bottomhole fluid intake. This may result in the rod string hitting the bottom of the pump barrel and damaging the traveling valve, rod string and the tubing due to the collision impact (Takacs 2015).

There are several other undesirable downhole pump operating conditions. These cases require immediate attention, diagnosis and remedial actions to prevent equipment damage or loss in production. These malfunctions can be viewed as anomalies in the SRP system. It is imperative that the production engineer immediately identify these anomalies in day-to-day operations. With hundreds or even thousands of wells, tracking or monitoring dyno cards in real-time becomes tedious and even impossible. Additionally, human-based anomaly detection owes a lot to the experience of the operator and therefore can be subjective. This underscores the need for efficient and automated anomaly detection algorithms. Here, I discuss some of the work done by researchers in terms of anomaly detection specifically applied to SRP systems.

Bangert (2019) discusses and groups the methods for sucker rod pump failure identification into three categories – library-based, model-based and segment-based methods. Library-based methods use a set of pump cards labelled by experts. Then a distance measure is computed between the card to be classified and the labelled pump cards. The choice of distance metric can vary. For example, Keating et al. (1991) compare two cards taking the sum of differences between the rod string load and displacement measurement. Another metric given by Dickinson and Jennings (1990) describe a pattern matching technique using differences in Fourier series representation between the pump cards.

Model-based methods are used when simple features of the card are extracted using geometry like finding centroids from the card load and displacement coordinates, area over segments of the pump card etc. to diagnose failures. Examples of this method are given by Bezerra et al. (2009) and Gao et al. (2015).

Segment-based methods rely on the ability to detect changes within segments of the pump card. A good example of segmenting method is given by Reges et al. (2015) who use downhole dynamometer cards (DDC) from real field data and artificially generated DDCs from the literature to form a total of 16 pumping downhole conditions as classification labels. They identify anomalies in the pumping conditions using segmentation features related to the points between the opening and closing of traveling and standing valves on the DDC. Extracted features statistics from each segment of the DDC are given as input to a fuzzy logic model for the classification of failure modes on the new, incoming DDCs.

Several of the more recent approaches in literature have adopted either a machine learning approach which have the capability of handling large real-time datasets combined with accurate classification of pump failures modes. Liu et al. (2011) describe a supervised machine learning model using 14 different measured parameters from Sucker Rod Pump surface data available from 12 wells. Their model is based on two classification algorithms known as AdaBNet (AdaBoost and Bayesian Network) and AdaDT (AdaBoost and Decision Trees) which learn from the training data. This model is then tested on 456 new wells and is shown to provide more than 90% accuracy in terms of detecting failures.

In another work by Liu et al. (2013), well data from 2000 rod pump pump-off controllers in 5 different oil fields is used. The data includes surface card area, maximum surface load, minimum surface load and daily runtime ratio. Using historical data, the normal and failure modes of the pump are labelled using an automated clustering and rule-based filtering process. This labelled data is then fed to a Support Vector Machines classifier for training and then to predict failures for new well data from other oil fields.

Boguslawski et al. (2018) build an ensemble model comprising of 4 sub-models including supervised Convolutional Neural Network (CNN), supervised Siamese Neural Network (SNN), a self-supervised Autoencoder Neural Network and a histogram of oriented gradients (HOG). Images from the downhole pump cards serve as inputs for each sub-model. The authors mention a data augmentation technique to generate more labeled input card data from few manually labelled ones. The ensemble model employs a stacking technique that takes the weighted outputs from each sub-model as its input to combine the probability distributions from all the model output predictions to correctly identify failure mode. **Fig. 3.5** shows the process behind their ensemble model.

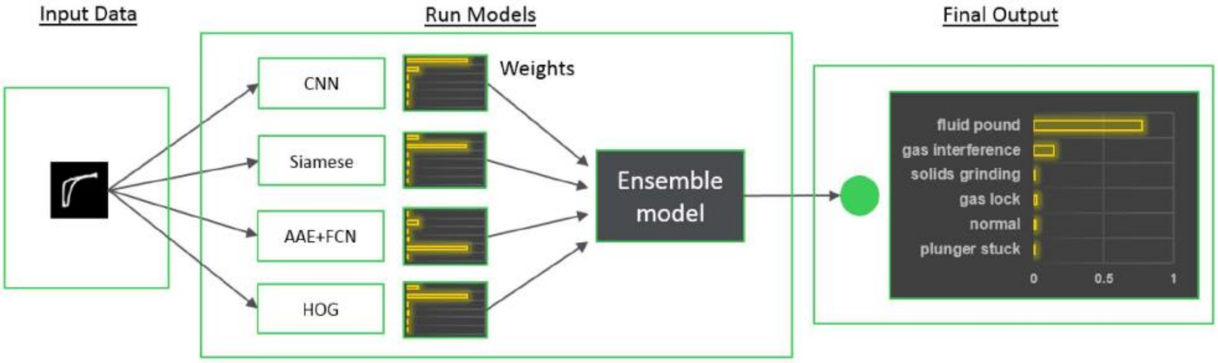


Fig. 3. 5—Ensemble of 4 machine learning models (Boguslawski et al. 2018).

Pennel et al. (2018) use signal data from 800 wells in the Bakken where subject matter experts assign rule-based labels on pump failures. Signal data includes time-series pressure data in the tubing and casing, fluid load, minimum and peak loads, stroke cycle time and others. Labels are attached to those instances where pump dysfunction is seen. The multivariate, time-series data along with the labels given as training data to machine learning models such as random forests, gradient boosted trees and neural networks are shown to provide high degree of accuracy in classification of failure.

Another pump card classification study by Sharaf et al. (2019) use data from a Supervisory Control and Data Acquisition (SCADA) system to acquire 5.3 million downhole pump cards from 297 wells in a Bahrain oilfield. **Fig. 3.6** shows an example of the instrumentation needed for a sucker rod pump. For this study, around 35,000 cards were manually labelled into 12 different pump conditions corresponding to normal as well as abnormal cases. 80% of the data is used for training 8 different machine learning algorithms and the remaining 20% of the data is used for the testing. They also report high accuracies of over 99% in the classification of failure modes.

In another work, Abdalla et al. (2020) train artificial neural networks (ANN) along with a genetic algorithm using over 4400 dynamometer cards. 30% of these cards are operating in the normal pump conditions and 70% showing abnormal pump conditions. The model is tested on 1915 new cards not seen by the ANN and achieves more than 99% accuracy.

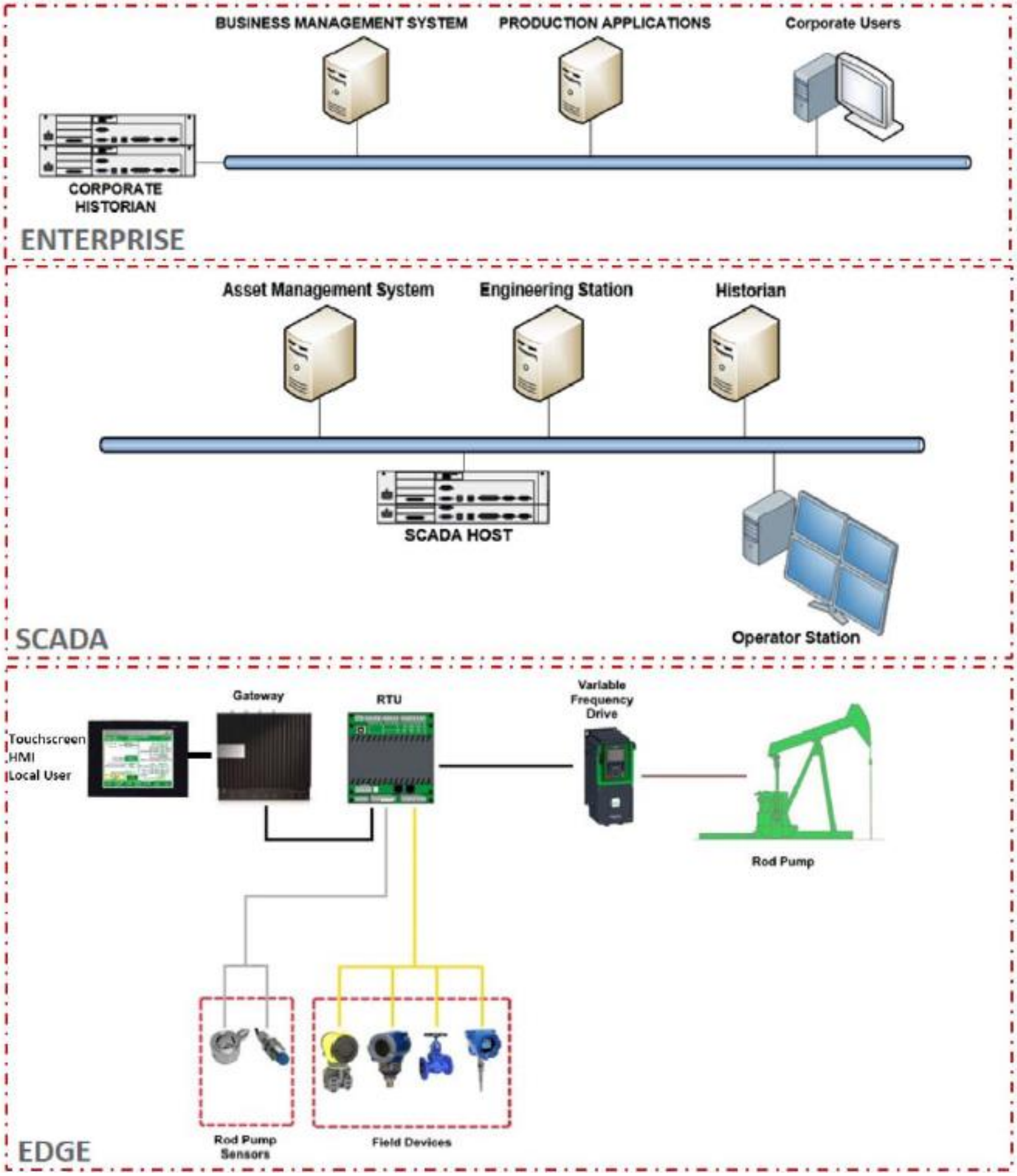


Fig. 3. 6— Instrumentation for data acquisition and control on the Sucker Rod Pump (Boguslawski et al. 2018).

It is evident that the oil industry and academia are leveraging the power of high-level computing coupled with machine learning and artificial intelligence models. This has enabled solving complex problems where human intervention is not always feasible. Every well which is supported or lifted by a sucker rod pump generates hundreds of pump cards daily. Assuming there are more than 100 wells in an oil field, the number of pump cards generated can easily become overwhelming for manual inspection. Therefore, the need for machine learning models, which can process a large number of cards in an efficient manner and correctly identify failure modes with low rates of false positives and false negatives.

In this chapter, I describe the use of Convolutional Neural Networks (CNN's) to predict pump malfunctions from pump dynamometer cards. CNNs are a relatively new development and belong to the class of deep learning algorithms because of their multi-layer architecture. The first reported work using CNN's was done by Lecun et al. (1998) at Bell Labs to classify hand-written digits from 0 to 9 from the MNIST database. Their network is shown in **Fig. 3.7** and it was used by the US Postal Services for recognizing digits in zip codes. Since then, the use of CNNs has rapidly grown to include visual image or face recognition, classification problems, self-driving cars and anomaly detection.

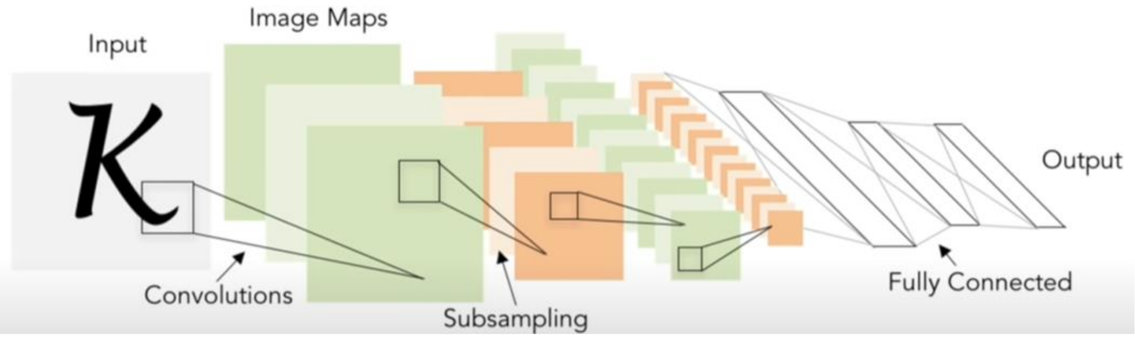


Fig. 3. 7—Illustration of a Convolutional Neural Network (Lecun et al. 1998).

In 2012, image recognition in very large datasets experienced a breakthrough. AlexNET from Krizhevsky et al. (2012) was the first large scale CNN model to outperform all other non-deep learning models in the ImageNet Large Scale Visual Recognition Challenge competition (ILSVRC). ImageNet is an image database with over 14 million images from 22,000 categories labelled by humans. For the competition, a subset of ImageNet was used and AlexNet correctly classified 1.2 million images with a lower error rate compared to other models. After 2012, there has been ever growing interest in the area of deep learning CNN models for image recognition and many high-performance CNN models have won the ImageNet challenge (Russakovsky et al. 2015). **Fig. 3.8** show the winners of ILSVRC comparing their percent error rates (shown by numbers on the % error rate bars) by the corresponding number of layers. The figure shows that CNNs have dramatically increased in complexity with a corresponding decrease in error rates.

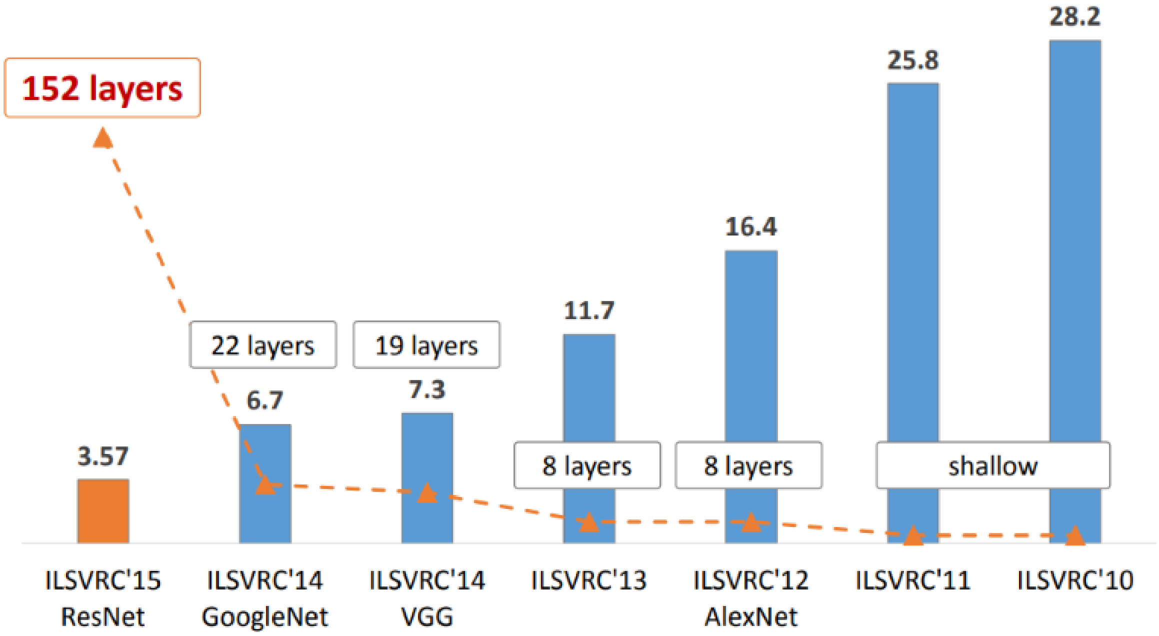


Fig. 3. 8— Revolution of depth in CNN models between years 2010 and 2015 (He et al. 2016).

In this thesis, I use a convolutional neural network (CNN) model known as VGG16, which won the 2014 ILSVRC, to identify and diagnose various cases of sucker rod pump failures. In the next section, I describe the architecture of VGG16 and explain its use for classification problems.

3.2 Architecture of the VGG16 model

The VGG16 model is a CNN model by Simonyan and Zisserman (2015) which won the first place in object localization and the second place for image classification for the 2014 ILSVRC. Object localization is a task of detecting objects inside images, while image classification is the task of identifying unseen test images. VGG16 achieved 92.7% top-5 image classification test accuracy in ImageNet, an improvement over AlexNet in 2012. Two of the key features of the model are its relatively smaller filter sizes and the depth of the network. The number 16 in VGG16 indicates the number of layers in the model as shown in **Fig. 3.9**.

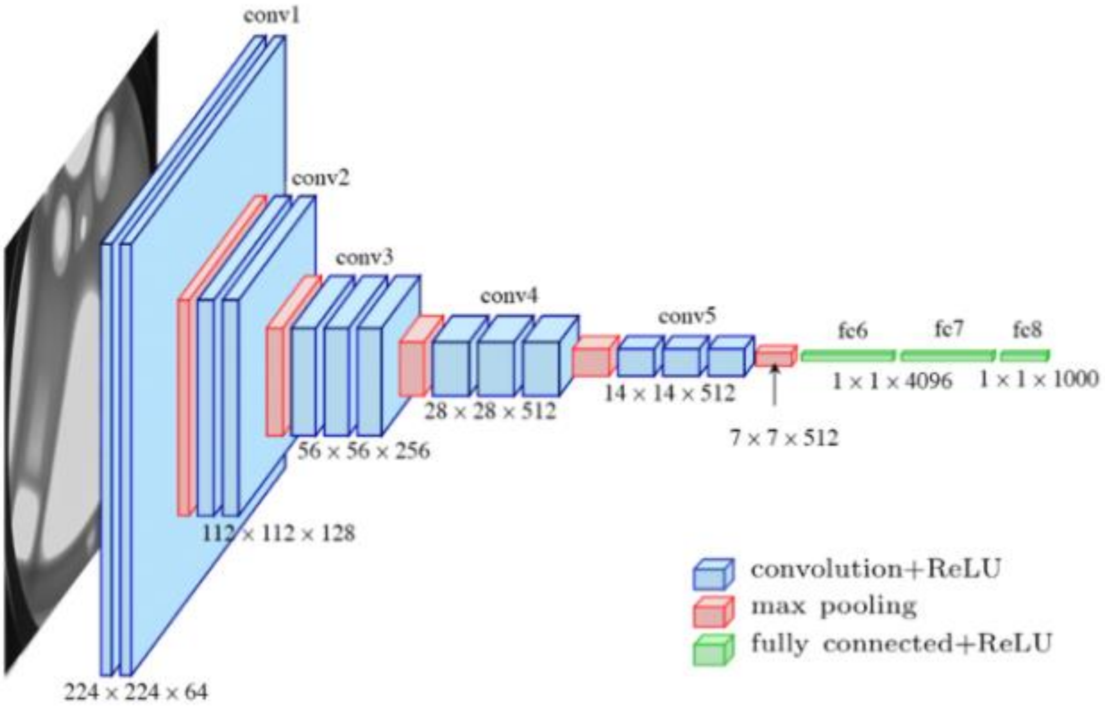


Fig. 3. 9 —VGG16 Architecture https://medium.com/towards-artificial-intelligence/the-architecture-and-implementation-of-vgg-16-b050e5a5920b (accessed 15 September 2020) (Khandelwal 2020).

1. Convolution Layers

VGG16 has 13 convolution layers as shown by the blue cuboids in the Fig 3.9. The idea behind convolution is much like multiplying two signals elementwise (signal and filter) and then summing the result. For VGG16, we have the input image (signal) that needs to be of dimensions 224 x 224 x 3 pixels.

Fig. 3.10 shows an example of convolution. If the input image has dimensions 32 x 32 x 3 and we apply a filter of dimension 5 x 5 x 3, then convolving the filter over the image is the action of sliding the filter over every possible spatial location on the image, taking the dot product between the filter and the image where they overlap. Every convolution over the entire image then results in an activation map. Here in this example, the dimensions of the activation map become 28 x 28 x 1.

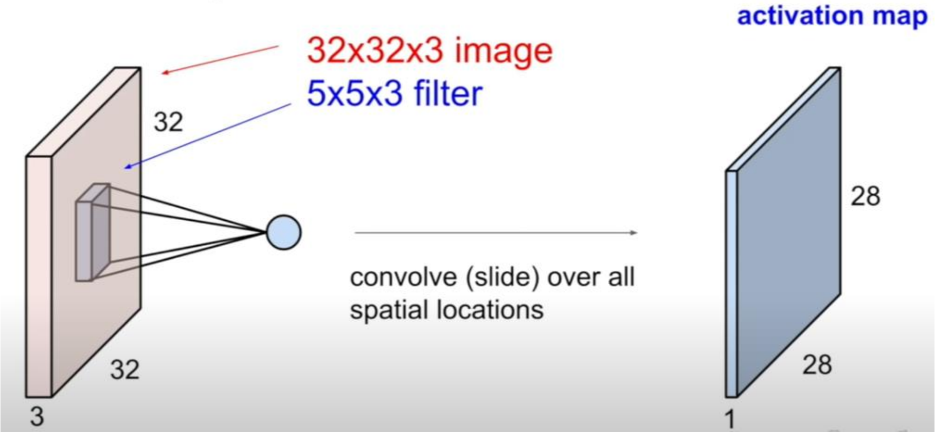


Fig. 3. 10— Example of Convolution https://cs231n.github.io/convolutional-networks (accessed 21 September 2020) (Li et al. 2017).

There are however a few other details. **Fig. 3.11** shows convolution with a stride. The stride refers to the number of locations skipped at each successive translation of the filter. Strides typically happen in the column- and row-direction. In **Fig. 3.12**, the stride length is seen to be 2. The formula for the output is given by (N-F+P)/S +1 where N is the dimension of input, F is size of filter, S is the number of strides used and P is the number of rows or columns added after padding. It is common practice to pad the first row, first column, last row and last column of the input with zeros during convolution. Padding allows us to capture the details surrounding the edges or corners of an input. In this case, the output size by taking the dot products between the input and the filter without any padding is 3 x 3.

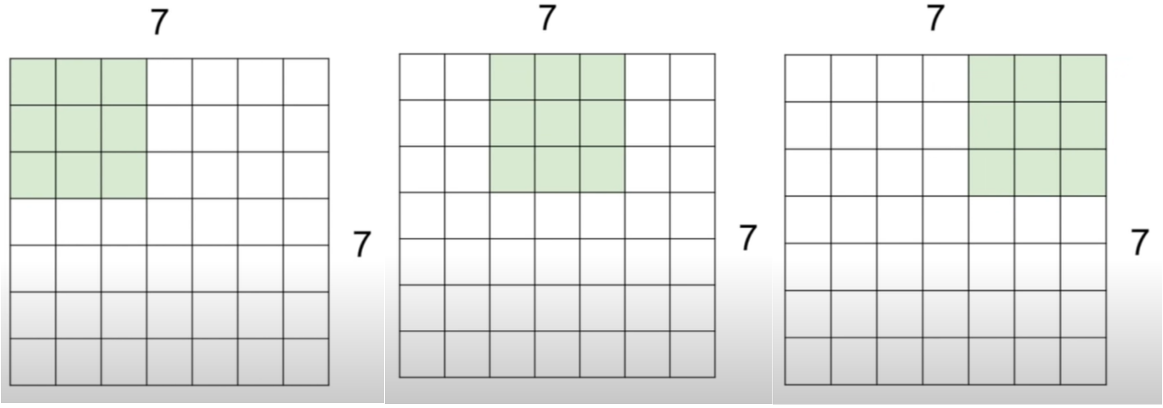


Fig. 3. 11—An illustration of a 2-dimensional input convolved with a 3 x 3 filter with a stride of 2 https://cs231n.github.io/convolutional-networks (accessed 21 September 2020) (Li et al. 2017).

Because VGG16 takes a 3-dimensional input, the filters used are extended to the full breadth of the input image and are also 3-dimensional in size. There are 64 filters in the 1st convolutional layer with dimensions 3 x 3 with stride and padding of 1. Output size = (N-F+P)/S +1 = (224-3+2)/1+1=224. Therefore, the output dimension from the 1st convolution is again 224 x 224 x 3. The output with all 3 channels is then 224 x 224 x 64.

Fig. 3.12 shows another example of two 3 x 3 x 3 filters working through a convolution on a 7 x 7 x 3 size image. In terms of image recognition, the earlier convolution layers learn the low-level features like edges or curved areas on the image and as we go deeper into the network, the complexity of abstraction increases thereby enabling the model to learn high-level features within the image.

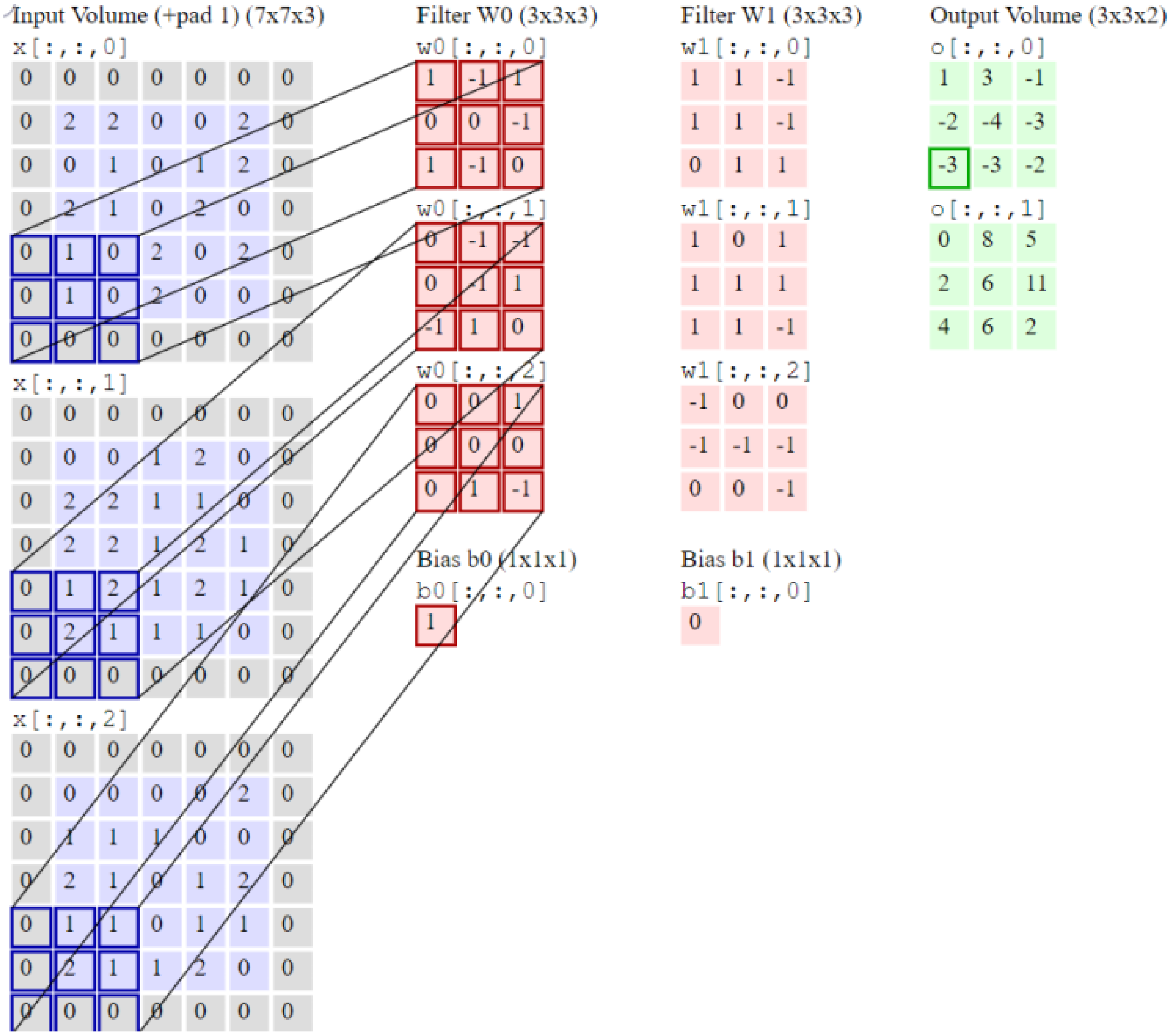


Fig. 3. 12—Example of a 3-dimensional input image convolved with 3-dimensional filter https://cs231n.github.io/convolutional-networks (accessed 21 September 2020) (Li et al. 2017).

2. Activation or the Rectified Linear Unit (ReLU) Layer

Every convolution layer in VGG16 is supplemented with a Rectified Linear Unit whose main function is to introduce non-linearity in the convolved output. The ReLU function is given by Eq. 3.2 and an example is shown in **Fig. 3.13**.

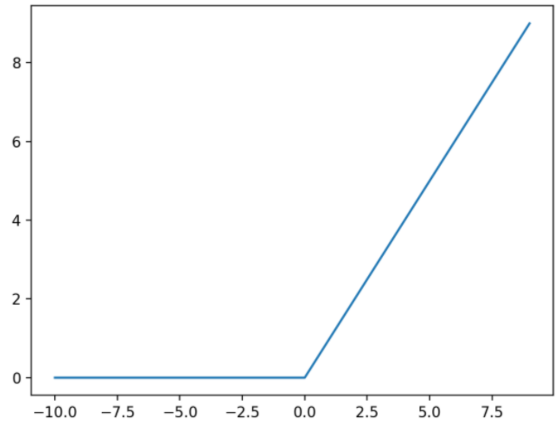
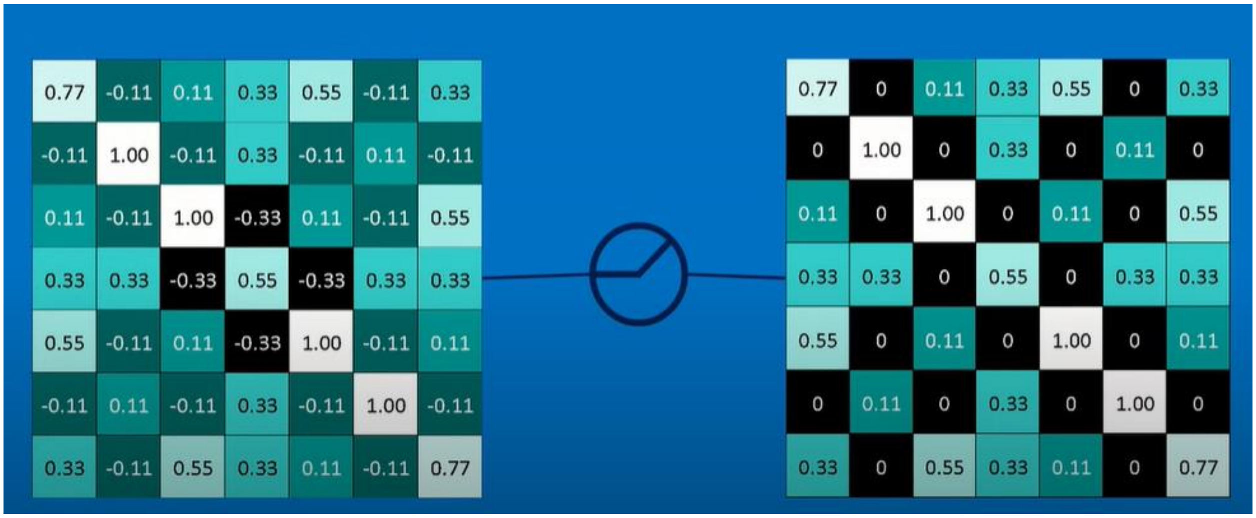


Fig. 3. 13—Example of Rectified Linear Activation Function https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks (accessed September 21 2020) (Brownlee 2017).

f(z)= max (0, z) ………………………………………………………………………...3.2

where z is the input image or in our case, z is the convolved output from the convolution layer. In other words, this is a normalization step where each negative value on the activation map generated after convolution becomes zero. This operation is illustrated in **Fig. 3.14**. This helps to keep the memory usage down while still maintaining and taking the vital information from the initial image forward.



0.77	-0.11	0.11	0.33	0.55	-0.11	0.33
-0.11	1.00	-0.11	0.33	-0.11	0.11	-0.11
0.11	-0.11	1.00	-0.33	0.11	-0.11	0.55
0.33	0.33	-0.33	0.55	-0.33	0.33	0.33
0.55	-0.11	0.11	-0.33	1.00	-0.11	0.11
-0.11	0.11	-0.11	0.33	-0.11	1.00	-0.11
0.33	-0.11	0.55	0.33	0.11	-0.11	0.77

0.77	0	0.11	0.33	0.55	0	0.33
0	1.00	0	0.33	0	0.11	0
0.11	0	1.00	0	0.11	0	0.55
0.33	0.33	0	0.55	0	0.33	0.33
0.55	0	0.11	0	1.00	0	0.11
0	0.11	0	0.33	0	1.00	0
0.33	0	0.55	0.33	0.11	0	0.77

Fig.3.14—Example of ReLU function indicated by the circled operator applied on an activation map

https://e2eml.school/how_convolutional_neural_networks_work.html (accessed 15 September 2020) (Rohrer 2019).

3. Pooling Layer

The output from the ReLU activation maps are now shrunk down to a smaller dimension of maps using max pooling. VGG16 uses max pooling of size 2 x 2 with a stride of 2. Max pooling takes the maximum value within the region of filter overlap on the image. This filter window is spatially moved along the output from the previous layer and a maximum value of each window is stored. **Fig. 3.15** shows an example of an image of size 224 x 224 x 64 which is spatially downsampled to a smaller size of 112 x 112 x 64 using a pooling layer. **Fig. 3.16** shows the max pooling filter in action. The pooling layer therefore helps in shrinking down the size of stack of filtered images and yet manages to maintain the same of information at a high level.

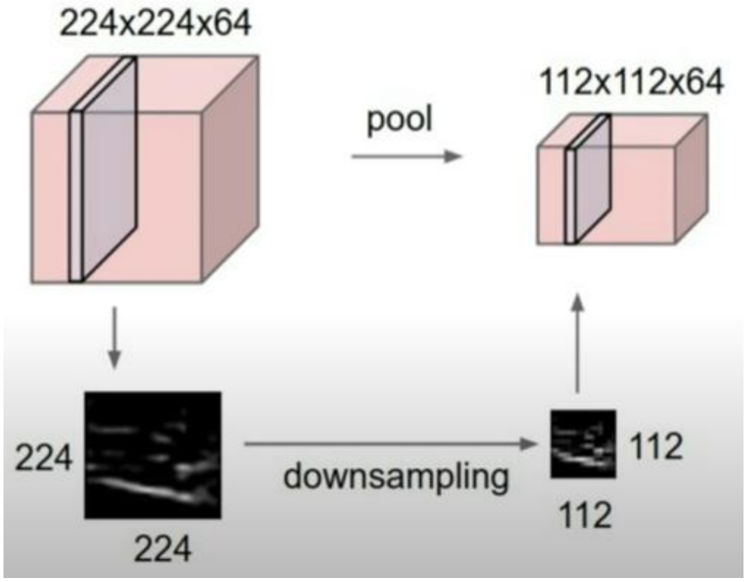


Fig. 3. 15—Example of downsampling using the Pooling Layer https://cs231n.github.io/convolutional-networks (accessed 21 September 2020) (Li et al. 2017).

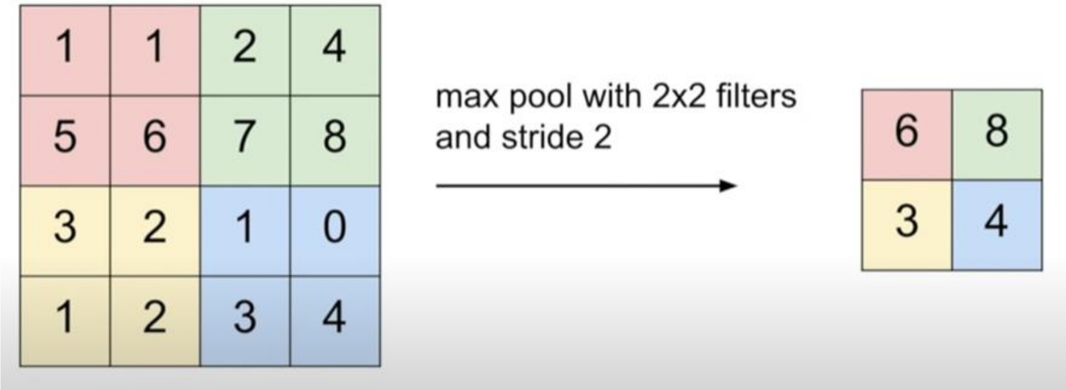


Fig. 3. 16—Example of Max Pooling using a stride of 2 https://cs231n.github.io/convolutional-networks (accessed 21 September 2020) (Li et al. 2017).

This process of convolution, ReLU activation and pooling occurs several times within the VGG architecture as shown in Fig. 3.9.

4. Fully Connected Layer (FC)

VGG16 has 3 fully connected layers at the end after the last pooling layer. Once the 16 convolutional layers of VGG16 have learned about the high-level features of the input image, the fully connected layer just stretches out the input from the last pooling layer into a 1-dimensional array. For example, if the input image is 32 x 32 x 3 in dimensions, then FC layer flattens it into a 1 x 3072. As shown in **Fig. 3.17**, this array is multiplied with the weights of FC layer represented by W with the values from the last input layer.

Essentially, a fully connected layer accepts all the high-level features learned from previous layers and computes probabilities for the class or label to be identified.

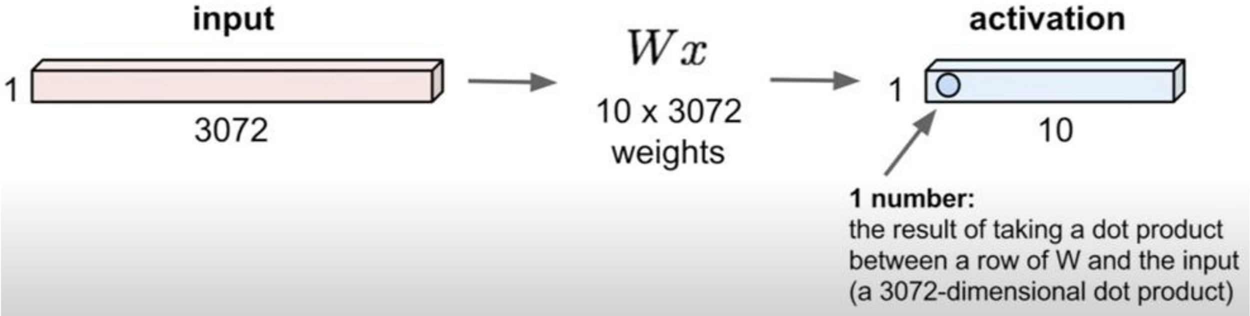


Fig. 3. 17—Example of a Fully Connected Layer output https://cs231n.github.io/convolutional-networks (accessed 21 September 2020) (Li et al. 2017).

3.3 Training the CNN

The CNN initially has randomized representation of the weights that the filters carry and hence after the first training image, it is likely that the prediction will be inaccurate. It is through a process called back propagation that every initial weight associated with each layer of the CNN is updated. For VGG16, the number of layers which carry weights are the 13 convolutional layers with filters and 3 fully connected layers.

The process of backpropagation can be explained in its 4 stages. The first stage is called a forward pass. For example, the first training input image of size 32 x 32 x 3 goes through the entire VGG16 network. As mentioned earlier, the initial weights associated with each of the layers are randomized. With this limited information, the output from the CNN model can neither identify nor differentiate between the low-level and high-level features present in the input image.

The second stage is the computation of the loss function. To begin with, since we have a training image and its label, we can define a suitable loss function associated with the error in the prediction. Examples of some commonly used loss functions are root mean squared error (L2 loss), categorical cross entropy loss, mean absolute error (L1 loss) etc. In this work, for VGG16 I have used the cross-entropy loss function which is mathematically represented as follows in Eq. 3.3:

$$Categorical\ Cross\ Entropy\ Loss\ (L) = -\sum_{i}^{C} t_i \log\left(f(s)_i\right) \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.3$$

where t is the target class ground truth, C is the number of outputs in the last layer of the CNN, f(s) is a softmax activation function which takes the scores s for each class from the CNN model and converts them between 0 and 1. The softmax activation function is computed as follows in Eq. 3.4.

$$f(s)_i = \frac{e^{s_i}}{\sum_{j}^{C} e^{s_j}} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.4$$

The elements s_i in the vector s represents the ith class output score and s_j are the scores computed by the CNN for each class in C.

Generally, the loss is high for initial part of the training process. In order to improve the next prediction, the aim is to minimize the cross-entropy loss function value.

To minimize the loss, we perform the third stage called the backward pass. Here we compute the derivative of the loss function with respect to the computed weights W, $\frac{dL}{dW}$ .By performing a backward pass, we find the contribution of each weight corresponding to each layer to the loss function and adjust them through Eq. 3.5:

$$w_{i+1} = w_i - \alpha \frac{dL}{dW} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.5$$

where w_i are initial weights, w_{i+1} is the newly computed weight for the next forward pass and α is the learning rate which is user-defined. This is the last step in the backpropagation algorithm where the weights are continually updated through the forward-backward pass process and recorded for each training image of every class such that we locate the global minimum of the loss function. After all iterations are completed, the weights of the layers in the CNN are now tuned or in the other words, our model has been trained for classification task.

The process of tuning for the model weights is called gradient descent optimization. There are many variants of the gradient descent algorithms. Here in this work, I have used a gradient descent algorithm known as Adaptive Moment Estimation or Adam by Kingma and Ba (2015).

Transfer Learning

The previous section was essentially about how the convolutional neural network is trained from scratch using a variant of the gradient descent optimization and backpropagation. The CNN model chosen here, VGG16, was previously trained on a portion of the ImageNet database to recognize 1000 classes of high-resolution objects. I retain the weights of the initial 15 layers and only modify the final fully connected layer to suit our specific application of downhole dynamometer cards classification. This process of leveraging the power of a pre-trained CNN and customizing the last layer of the network specific to our application is called Transfer Learning. In this work, I have modified the final classifier layer to identify the 13 classes of sucker rod pump failure modes considered in this study.

3.4 Dataset Description and Data Augmentation

In this work, I generate synthetic dynamometer cards corresponding to different failure modes. With a large number of training images and transfer learning, the hypothesis is that our customized model should be able to correctly classify and detect the various modes of sucker rod pump failures. There are 17 different cases of pump card failures considered and a few of them, belonging to the same class, were combined based on the levels of severity. For example, the different levels of severity of gas interference are grouped into one class. Following this reduction, the total number of failure modes to be detected is 13. **Fig. 3.18** shows few of cases of pump failure modes.

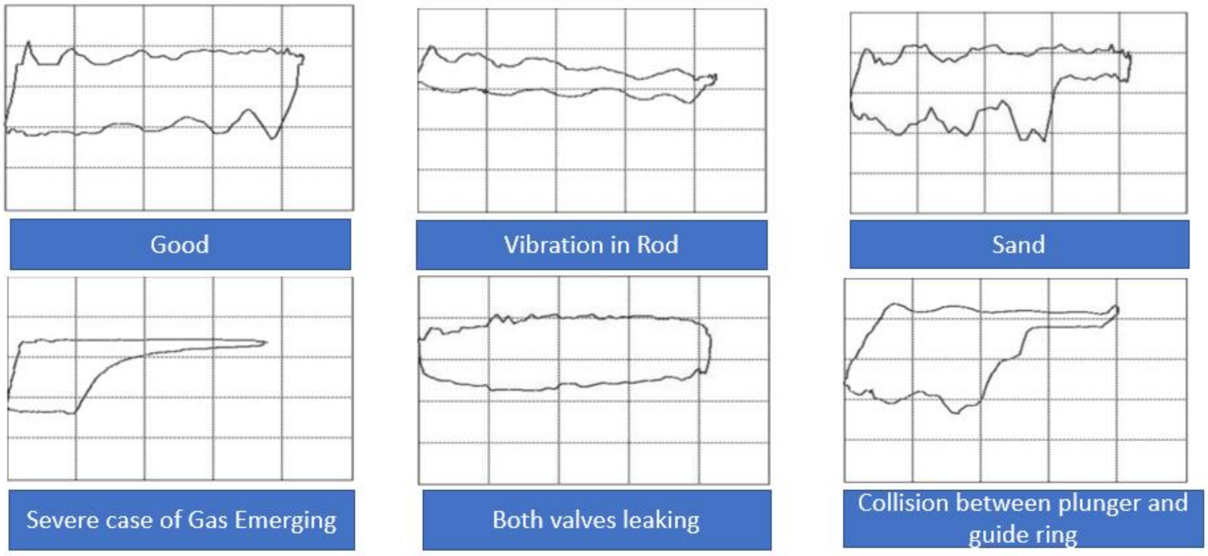


Fig. 3. 18— Example of Rod Pump Failures modes through downhole dynamometer cards (Tan et al. 2015).

In order to expand the diversity of images seen by the CNN, I rely on a process called 'Data Augmentation' for each dynamometer card. This was done by adding Gaussian white noise on both load and displacement data, as well as vertical and horizontal translation and shrinking and enlargement. These operations were performed on the base images from each case and a grand total of 8500 images belonging to 13 classes were generated. **Table 3.1** list the 13 classes with their description and number of training/test images in each class. The base images pertaining to each class are shown in the Appendix from **Fig. A17** to **A29.**

Pump Failure Type Label	Failure Description	Number of Images generated (Training, Test)
Normal	Two different cases of a good working pump condition.	1000, 200
Light Severe Fluid Pound	Light and severe cases of fluid pounding. Light case of plunger not being filled completely, and severe case of plunger hardly being filled	1000, 200
Thick Oil	Pump is working okay with high viscosity oil	500, 100
Sand	Presence of sand inside the rod pump assembly	500, 100
Light Severe Gas	Light and severe cases of gas interference	1000, 200
Leaking Inlet Valve	Leaking standing valve	500, 100
Leaking Outlet Valve	Leaking traveling valve	500, 100
Both Valves Leaking	Standing and traveling valves leaking	500, 100
Delay Closing Valve	Shorter and longer delays in closing of Standing valve	1000, 200
Collision Plunger Guide Ring	Collision between the plunger and guide ring	500, 100
Resistance Flow Oil	Resistance to flow of oil due to presence of paraffin wax	500, 100

Table 3. 1—13 different cases of pump cards comprising of normal and abnormal conditions (Tan et al. 2015).

Training and Testing Setup

The 8500 images belonging to 13 different classes of pump failures were split into 2 groups. 80% of the data was kept for training the CNN model and 20% were kept aside for the testing the model. The data selected for training and testing is fed to the network in batches and the weights are updated after each batch.

There are 3 methods of feeding images to the CNN. One option is to have the CNN process each image one-by-one sequentially with the weights updated after every training image is processed through the network. Unfortunately, this contributes to prohibitively high processing time. The second option is to process the entire training set in one step, but this option necessitates a high memory usage. The third option trains the network in mini-batches where the batch size is of the order 2^n where n typically ranges from 4 to 8. This method provides a good balance between memory use and computation time.

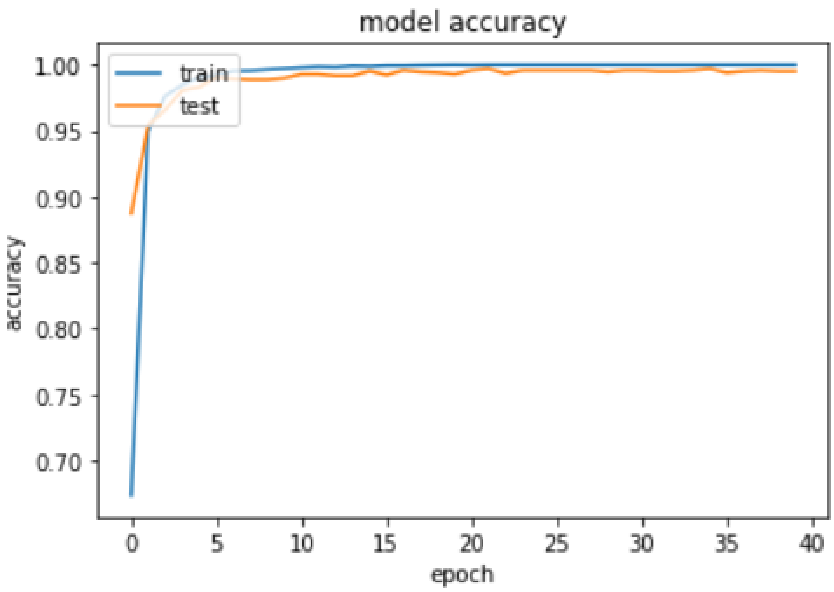
The time for one cycle of training and testing is measured in epochs. One epoch is said to be complete when all the batches are processed for the entire training and test dataset. For example, in this work I have selected the batch size to be 32 (2^5) and fit the model for 40 epochs. For each epoch, the batches are reconstructed randomly from the entire image set.

3.5 Results and Conclusions

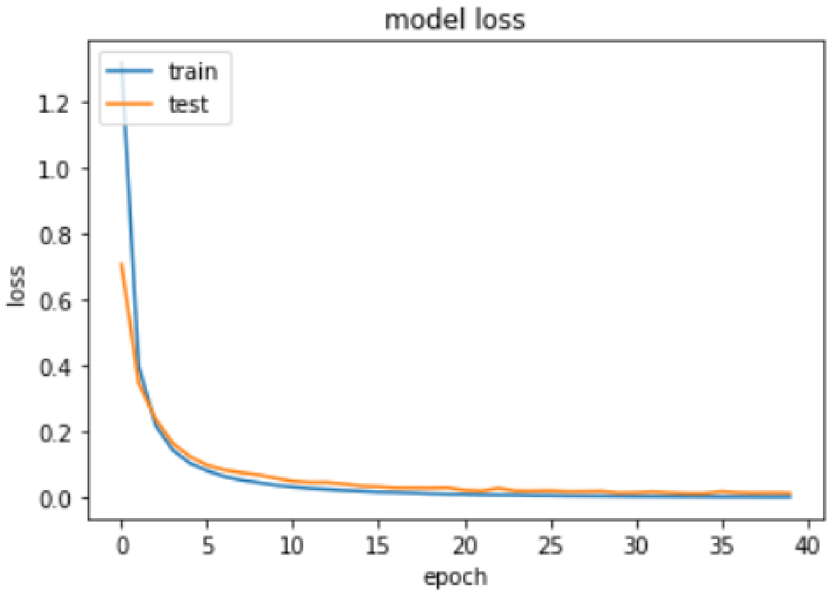
In this study, the hyperparameters used during training the CNN are as follows:

Learning rate = 0.0001, filter size = 3 x 3, number of filters used = 2944, batch size = 32, epoch step size = 212 and number of epochs=40. Epoch step size is calculated by dividing the total number of images in the dataset by the number of epochs. This step size is used to monitor the progress of training and test phases of the CNN. The model is run on a Google Colab Graphics Processing Unit (GPU) to reduce the computation time.

Fig. 3.19a. **and 3.19b**. show the plot of model accuracy and loss on the y-axes versus epochs on x-axes respectively during training and testing.



(a)



(b)

Fig. 3. 19—(a) Plot of VGG16 model accuracy and (b) model loss during training and testing.

In the **Fig. 3.20**., I plot the confusion matrix for the test data. The y-axis represents the true, but unseen, labels and the x-axis represents the predicted labels. The colors indicate the number of images associated with each class.

There are a few off-diagonal elements which denote either false positives or false negatives, but they are too few to be seen in Fig. 3.20. I show the numerical values within the confusion matrix in **Fig. 3.21**.

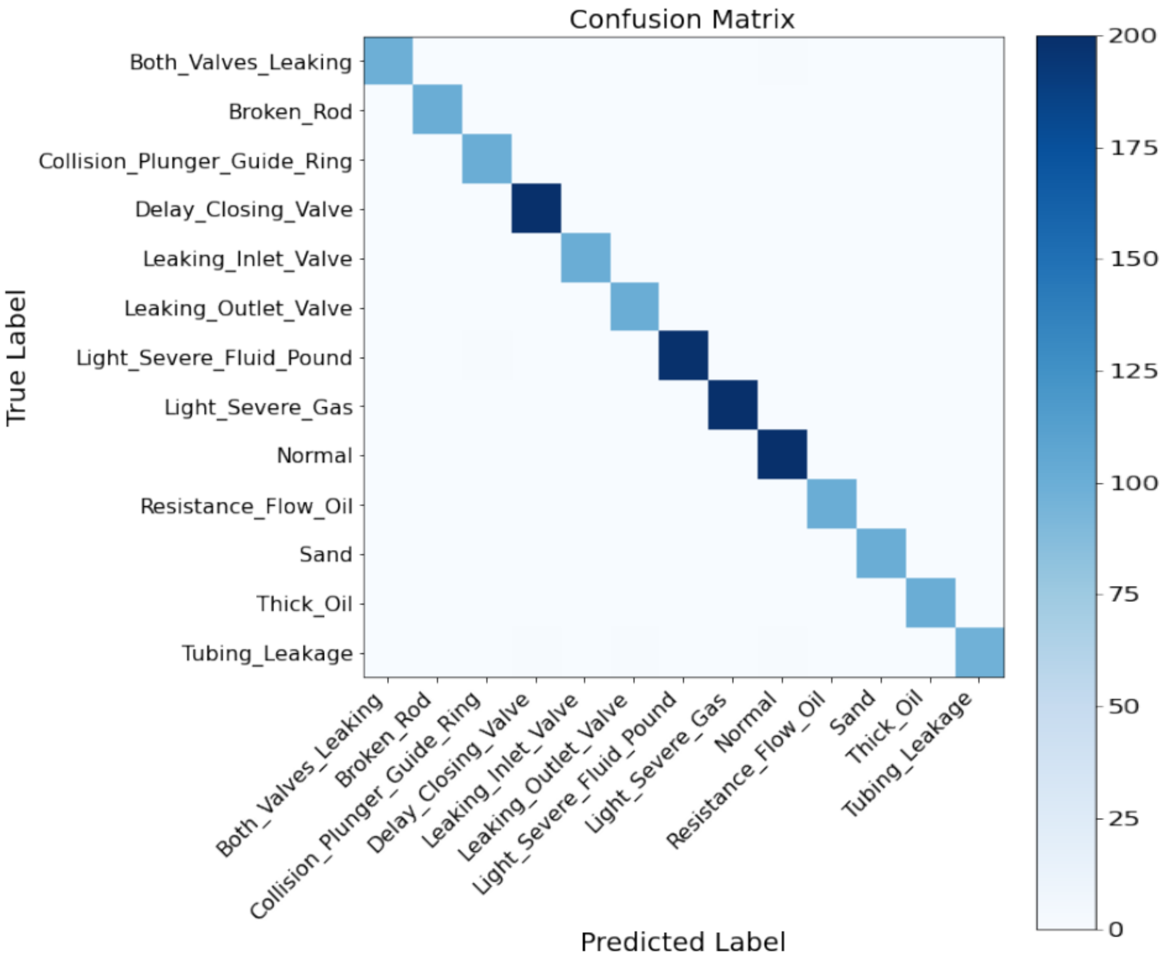
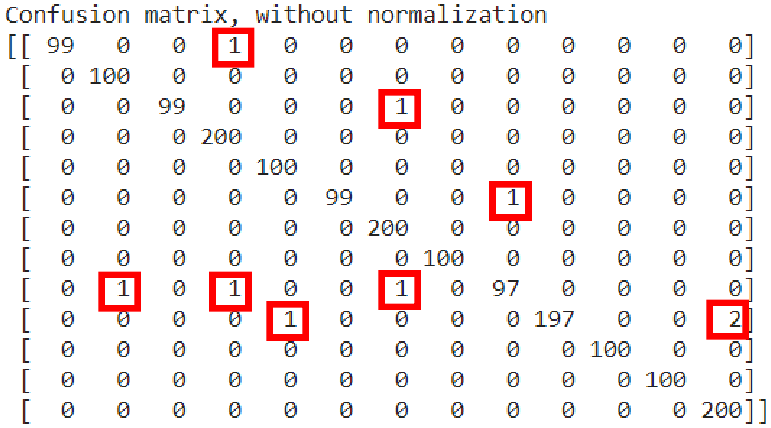


Fig. 3. 20—Confusion Matrix showing True labels versus Predicted Labels.

In Fig. 3.21, there are a few misclassified labels, but the confusion matrix is strongly diagonally dominant.

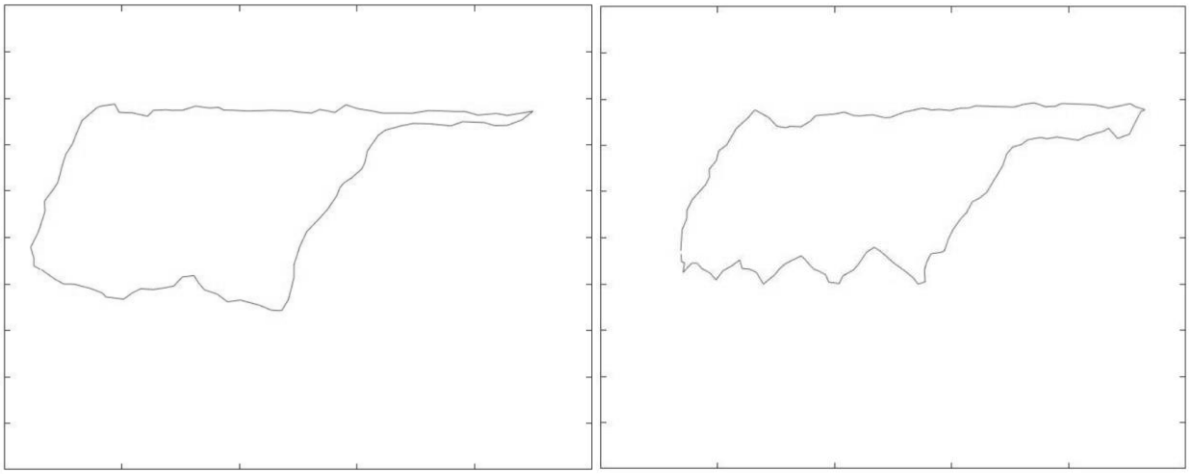


```
Confusion matrix, without normalization
[[ 99   0   0   1   0   0   0   0   0   0   0   0   0]
 [  0 100   0   0   0   0   0   0   0   0   0   0   0]
 [  0   0  99   0   0   0   1   0   0   0   0   0   0]
 [  0   0   0 200   0   0   0   0   0   0   0   0   0]
 [  0   0   0   0 100   0   0   0   0   0   0   0   0]
 [  0   0   0   0   0  99   0   0   1   0   0   0   0]
 [  0   0   0   0   0   0 200   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0 100   0   0   0   0   0]
 [  0   1   0   1   0   0   1   0  97   0   0   0   0]
 [  0   0   0   0   1   0   0   0   0 197   0   0   2]
 [  0   0   0   0   0   0   0   0   0   0 100   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0 100   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0 200]]
```

Fig. 3. 21— Confusion Matrix without normalization showing True labels versus Predicted Labels.

Fig. 3.22 shows an example of a misclassified mode of failure. Fig. 3.22a shows the correct label on the left and Fig. 3.22b shows the incorrectly predicted label on the right by the model.

The dyno card to be classified is from a failure mode indicating poor pump fillage, where the plunger is not being completely filled. The CNN model however predicts this one card as belonging to a moderate case of gas interference. This card can easily be seen as gas interference and perhaps need an expert to examine it with more information for proper prediction.

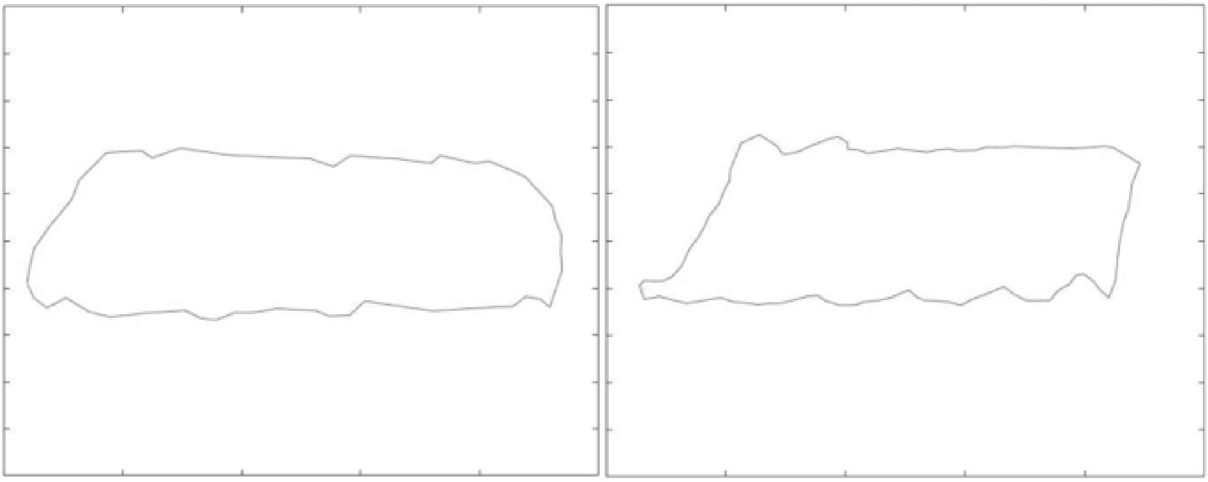


(a) (b)

Fig. 3. 22—An example of model's misclassification (a) True label related to the severe mode of pump fillage problem (b) Model Predicted label related to the moderate case of gas interference (Tan et al. 2015).

Fig. 3.23 shows another example of a misclassified mode of failure. Fig. 3.23a shows the correct label on the left and Fig. 3.23b shows the incorrectly predicted label on the right by the model.

The dyno card to be classified is from a failure mode indicating the problem of leaking inlet or standing. The CNN model however predicts this one card as belonging to a case of time delay in the closing of the standing valve.



(a)

(b)

Fig. 3. 23— An example of model's misclassification (a) True label related to the leaking standing valve (b) Model Predicted label related to the delay in closing of the standing valve (Tan et al. 2015).

In **Fig. 3.24 and 3.25**, I show two different test pump cards which the model had no problems predicting for all samples belonging to that class or label. Fig. 3.23a shows the correct label on the left and Fig. 3.23b shows the correctly predicted label on the right by the model. The dyno card to be classified is from a failure mode indicating the presence of sand inside the rod pump assembly. The CNN model correctly predicts this card as belonging to a case of sand interference.

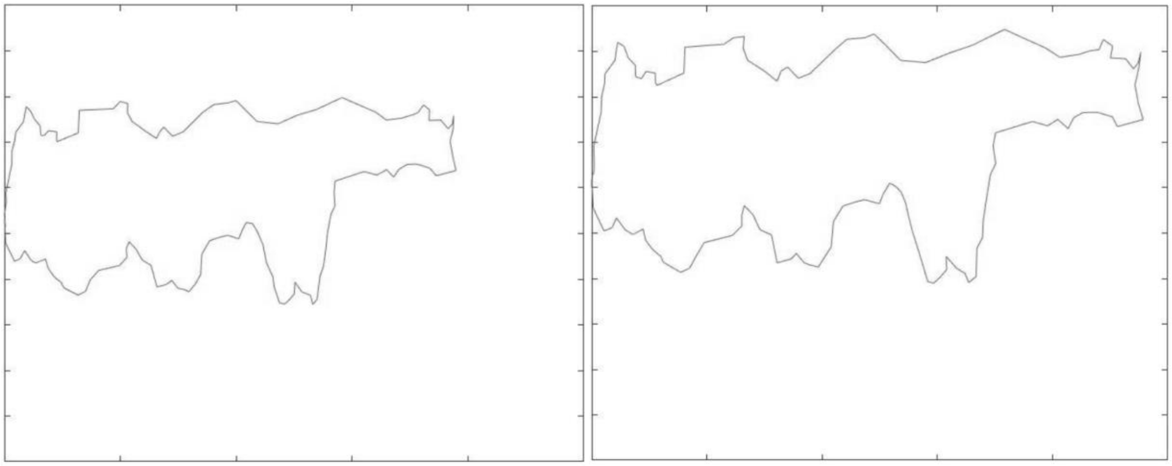

(a) (b)

Fig. 3. 24— An example of model's accurate classification (a) True label related to presence of sand (b) Model predicted label related to the presence of sand (Tan et al. 2015).

Fig. 3.25a. shows the correct label on the left and Fig. 3.25b shows the correctly predicted label on the right by the model. The dyno card to be classified is from a failure mode indicating a severe case of gas interference inside the rod pump assembly. The CNN model correctly predicts this card as belonging to a case of severe gas interference.

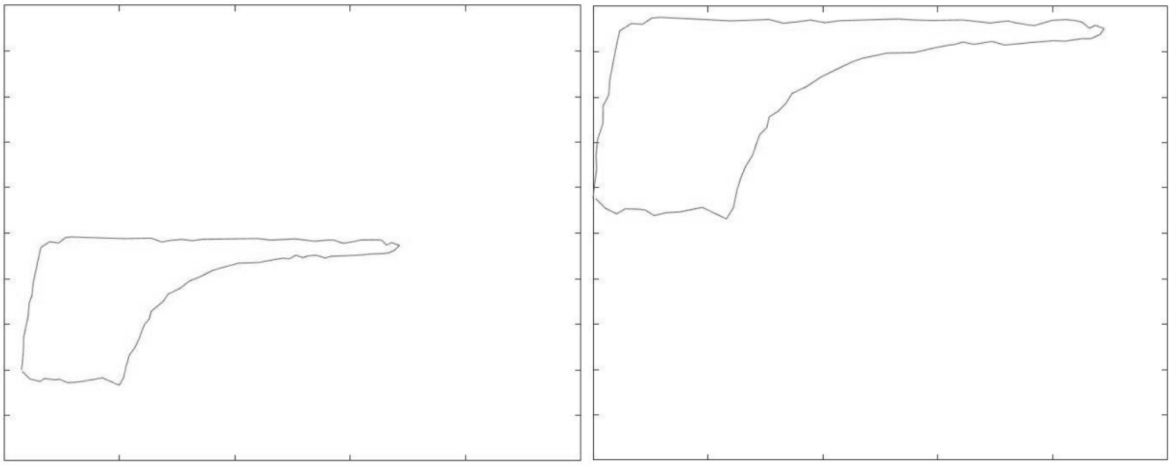

(a) (b)

Fig. 3. 25— An example of model's accurate classification (a) True label related to severe case of gas interference (b) Model predicted label related to the severe gas of gas interference (Tan et al. 2015).

Thus, we can conclude that using transfer learning, a CNN model like VGG16 can be customized as per the application required. Here the application was to identify and classify the rod pump failure cases. The model currently uses synthetic data for pump cards generated from the literature and the results shown can serve as a proof of concept. One can now extend this concept and apply it to real world oil field pump data. By training this model on several thousand downhole pump cards, it can be put to test for classifying online real time pump cards. This can potentially serve as a continuous surveillance system with pre-set alarms conditioned for abnormal failure modes. With the power of Internet of Things (IoT), machine learning algorithms and big data available from downhole sensors, a suitable time series anomaly detection model could be used in conjunction with the CNN model.

A future work to extend this study would be to predict anomalies in real-time. This will prevent sucker rod pump failures in most cases, increase their efficiency and thus save the operator millions of dollars in potential downtime spent for root cause analysis, equipment repair and costly workovers at the well site.

References

Abdalla, R., El Ela, M. A., & El-Banbi, A. (2020). Identification of Downhole Conditions in Sucker Rod Pumped Wells Using Deep Neural Networks and Genetic Algorithms. SPE Production and Operations, 35(2), 435–447. https://doi.org/10.2118/200494-PA

Adams, R. P., & MacKay, D. J. C. (2007). Bayesian Online Changepoint Detection. http://arxiv.org/abs/0710.3742

Ajani, A., & Kelkar, M. (2012). Interference study in shale plays. Society of Petroleum Engineers - SPE Hydraulic Fracturing Technology Conference 2012, February, 36–50. https://doi.org/10.2118/151045-ms

Ajisafe, F., Solovyeva, I., Morales, A., Ejofodomi, E., & Porcu, M. M. (2017). Impact of well spacing and interference on production performance in unconventional reservoirs, Permian Basin. SPE/AAPG/SEG Unconventional Resources Technology Conference 2017. https://doi.org/10.15530/urtec-2017-2690466

Akinsete, O., & Oshingbesan, A. (2019). Leak detection in natural gas pipelines using intelligent models. Society of Petroleum Engineers - SPE Nigeria Annual International Conference and Exhibition 2019, NAIC 2019, August, 5–7. https://doi.org/10.2118/198738-MS

Apergy Artificial Lift Academy, University of Oklahoma – Artificial Lift (2019). October.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research, 46(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Bangert, P. (2019). Diagnosing and Predicting Problems with Rod Pumps using Machine Learning. SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings, 2019-March(c), 1–13. https://doi.org/10.2118/194993-ms

Bello, O., Bale, D. S., Yang, L., Yang, D., Kb, A., Lajith, M., & Lazarus, S. (2018). Integrating downhole temperature sensing datasets and visual analytics for improved gas lift well surveillance. Proceedings - SPE Annual Technical Conference and Exhibition, 2018-Septe. https://doi.org/10.2118/191626-ms

Bezerra, M. A. D., Schnitman, L., Barreto Filho, M. D. A., & Felippe De Souza, J. A. M. (2009). Pattern recognition for downhole dynamometer card in oil rod pump system using artificial neural networks. ICEIS 2009 - 11th International Conference on Enterprise Information Systems, Proceedings, AIDSS, 351–355. https://doi.org/10.5220/0002000403510355

Boguslawski, B., Boujonnier, M., Bissuel-Beauvais, L., & Saghir, F. (2019). Edge analytics at the wellhead: Designing robust machine learning models for artificial lift failure detection. Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 2018, ADIPEC 2018. https://doi.org/10.2118/192886-ms

Boguslawski, B., Boujonnier, M., Bissuel-Beauvais, L., Saghir, F., & Sharma, R. D. (2018). IIoT edge analytics: Deploying machine learning at the wellhead to identify rod pump failure. Society of Petroleum Engineers - SPE Middle East Artificial Lift Conference and Exhibition 2018, MEAL 2018. https://doi.org/10.2118/192513-ms

Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

Brown, K.E., 1980. The Technology of Artificial Lift Methods, vol. 2a. Petroleum Publishing Co., Tulsa, OK.

Brown, M. J., Hutchinson, L. A., Rainbow, M. J., Deluzio, K. J., & De Asha, A. R. (2017). A comparison of self-selected walking speeds and walking speed variability when data are collected during repeated discrete trials and during continuous walking. Journal of Applied Biomechanics, 33(5), 384–387. https://doi.org/10.1123/jab.2016-0355

Brownlee (2017) https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks (accessed September 21 2020).

Cadei, L., Rossi, G., Montini, M., Fier, P., Milana, D., Corneo, A., Lancia, L., Loffreno, D., Purlalli, E., Carducci, F., Nizzolo, R., & Sophia, G. (2020). Achieving digital twin through advanced analytics support: A novelty detection framework to highlight real-time anomalies in time series. International Petroleum Technology Conference 2020, IPTC 2020. https://doi.org/10.2523/iptc-19834-ms

Cao, R., Li, R., Girardi, A., Chowdhury, N., & Chen, C. (2017). Well interference and optimum well spacing for Wolfcamp Development at Permian Basin. SPE/AAPG/SEG Unconventional Resources Technology Conference 2017, 1–11. https://doi.org/10.15530/urtec-2017-2691962

Cox, D. & Reid, N. (2000). The Theory of Design of Experiments.

Cochran, W. G. (1972). The planning of observational studies of human populations (with Discussion). Journal of the Royal Statistical Society A, 234-266.

Chu, W.-C., Scott, K. D., Flumerfelt, R., Chen, C., & Zuber, M. D. (2020, February 1). A New Technique for Quantifying Pressure Interference in Fractured Horizontal Shale Wells. Society of Petroleum Engineers. doi:10.2118/191407-PA

Daneshy, A. (2018). Analysis of horizontal well fracture interactions, and completion steps for reducing the resulting production interference. Proceedings - SPE Annual Technical Conference and Exhibition, 2018-Septe. https://doi.org/10.2118/191671-ms

Daneshy, A., & King, G. (2019). Horizontal Well Frac-Driven Interactions: Types, Consequences, and Damage Mitigation. Journal of Petroleum Technology, 71(06), 45–47. https://doi.org/10.2118/0619-0045-jpt

Dickinson, R. R., & Jennings, J. W. (1990). Use of pattern-recognition techniques in analyzing downhole dynamometer cards. SPE Production Engineering, 5(2), 187–192. https://doi.org/10.2118/17313-pa

Economides, M., Hill, D., Economides, C., & Zhu, D. (2012). Petroleum production systems Second edition. In Dairy Science & Technology, CRC Taylor & Francis Group.

EIA (2020). https://www.eia.gov/petroleum/drilling/

Engle, C. H. and A. M. Buswell. "On The Determination of Transmissibility and Storage Coefficients from Pumping Test Data." (1952).

Esquivel, R., & Blasingame, T. A. (2017). Optimizing the development of the haynesville shale - Lessons-learned from well-to-well hydraulic fracture interference. SPE/AAPG/SEG Unconventional Resources Technology Conference 2017, 1, 1–22. https://doi.org/10.15530/urtec-2017-2670079

Fisher, R. (1958). Cigarettes, cancer, and statistics. In The Centennial Review of Arts & Science (Vol. 2, pp. 151–166).

Li, F., Krishna, R., Xu, D., (2017) https://cs231n.github.io/convolutional-networks.

Gao, Q.; Sun, S.; Liu, J. (2015): Fault Diagnosis of Suck Rod Pumping System via Extreme Learning Machines. Proceedings of the 5th IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, June 8-12 2015, Shenyang, China. 503–507.

Gaurav, A., Gibbon, E. J., & Roberson, T. M. (2017, February 15). Asset Evaluation Utilizing Multi-Variate Statistics Integrating Data-Mining, Completion Optimization, and Geology Focused on Multi-Bench Shale Plays. Society of Petroleum Engineers. doi:10.2118/185018-MS

Gibbs, S.G. (1963): Predicting the Behavior of Sucker-Rod Pumping Systems. SPE 588. SPE Rocky Mountain Regional Meeting, Denver, May 27-28, 1963.

Golan, M, and Whitson, C H. Well test performance. United States: N. p., 1991. Web.

Guo, B., Liu, X, Tan, X. (2017) Petroleum Production Engineering (Second Edition), Gulf Professional Publishing.

Guo, X., Wu, K., Killough, J., & Tang, J. (2018). Understanding the mechanism of interwell fracturing interference based on reservoir-geomechanics-fracturing modeling in Eagle Ford Shale. SPE/AAPG/SEG Unconventional Resources Technology Conference 2018, URTC 2018, 2017. https://doi.org/10.15530/urtec-2018-2874464

Gupta, I., Rai, C., Devegowda, D., & Sondergeld, C. (2020). A Data-Driven Approach to Detect and Quantify the Impact of Frac-Hits on Parent and Child Wells in Unconventional Formations. 71–85. https://doi.org/10.15530/urtec-2020-2190

Gupta, I., Rai, C., Devegowda, D., & Sondergeld, C. H. (2020). Fracture Hits in Unconventional Reservoirs: A Critical Review. SPE Journal, February 1–23. https://doi.org/10.2118/203839-pa

Gupta, S., Nikolaou, M., Saputelli, L., & Bravo, C. (2016, September 6). ESP Health Monitoring KPI: A Real-Time Predictive Analytics Application. Society of Petroleum Engineers. doi:10.2118/181009-MS

Gurina, E., Klyuchnikov, N., Zaytsev, A., Romanenkova, E., Antipova, K., Simon, I., Makarov, V., & Koroteev, D. (2019). Failures detection at directional drilling using real-time analogues search. 1–17. http://arxiv.org/abs/1906.02667

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. Econometrica, 71(4), 1161–1189. https://doi.org/10.1111/1468-0262.00442.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. Journal of Statistical Software, 42(8), 1–28. https://doi.org/10.18637/jss.v042.i08

Ho, Tin Kam (1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282.

Hotelling, H.,1933. Analysis of a Complex of Statistical Variables Into Principal Components, Journal of Educational Psychology, volume 24, pages 417-441 and 498-520.

Huang, J., Morris, J. P., Fu, P., Settgast, R. R., Sherman, C. S., & Ryerson, F. J. (2019, February 1). Hydraulic-Fracture-Height Growth Under the Combined Influence of Stress Barriers and Natural Fractures. Society of Petroleum Engineers. doi:10.2118/189861-PA

Imbens, G., & Rubin, D. (2015). Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139025751

Jansen van Rensburg, N. (2019). AI4ESP - Autonomous well surveillance for ESP pumps using artificial intelligence. Society of Petroleum Engineers - SPE Oil and Gas India Conference and Exhibition 2019, OGIC 2019. https://doi.org/10.2118/194587-ms

Jolliffe.I.T., 2002. Principal Component Analysis, second edition, New York: Springer-Verlag New York, Inc.

Keating, J.F.; Laine, R.E.; Jennings, J.W. (1991): Application of a Pattern-Matching Expert System to Sucker-Rod, Dynamometer-Card Pattern Recognition. SPE 21666.

Khandelwal (2020) VGG16 Architecture https://medium.com/towards-artificial-intelligence/the-architecture-and-implementation-of-vgg-16-b050e5a5920b (accessed 15 September 2020).

King, G. E., Rainbolt, M. F., & Swanson, C. (2017). Frac hit induced production losses: Evaluating root causes, damage location, possible prevention methods and success of Remedial treatments. Proceedings - SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/187192-ms

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–15.

Krizhevsky, A. & Sutskever, I. & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

Kuesters, A., Mason, C., Gomes, P., Cockburn, C., & Lodhi, H. (2020). Drillstring failure prevention - A data-driven approach to early washout detection. SPE/IADC Drilling Conference, Proceedings, 2020-March 1–12. https://doi.org/10.2118/199610-ms

Kumar, A., Shrivastava, K., Elliott, B., & Sharma, M. (2020). Effect of parent well production on child well stimulation and productivity. Society of Petroleum Engineers - SPE Hydraulic Fracturing Technology Conference and Exhibition 2020, HFTC 2020, 2016. https://doi.org/10.2118/199700-ms

Kurtoglu, B., & Salman, A. (2015). How to utilize hydraulic fracture interference to improve unconventional development. Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference, ADIPEC 2015. https://doi.org/10.2118/177953-ms

Lanza, S. T., Coffman, D. L., & Xu, S. (2013). Causal Inference in Latent Class Analysis. Structural equation modeling: a multidisciplinary journal, 20(3), 361–383. https://doi.org/10.1080/10705511.2013.797816

Lanzante. A Cautionary Note on the Use of Error Bars. Journal of Climate (2005) vol. 18 pp. 3699-3703.

Lea JF (2007) Artificial lift selection. Chapter 10 in SPE petroleum engineering handbook, Vol. IV. Society of Petroleum Engineers, Dallas

Lecun, Y., Bottou, L,,Bengio Y., and Haffner, P. "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.

Lindsay, G., Miller, G., Xu, T., Shan, D., & Baihly, J. (2018). Production performance of infill horizontal wells vs. Pre-existing wells in the major US unconventional basins. Society of Petroleum Engineers - SPE Hydraulic Fracturing Technology Conference and Exhibition 2018, HFTC 2018. https://doi.org/10.2118/189875-ms

Liu, S., Liu, Y., Usc, K. Y., & Sciences, I. (2011). SPE 146038 Automatic Early Fault Detection for Rod Pump Systems.

Liu, Y., Yao, K. T., Raghavenda, C. S., Wu, A., Guo, D., Zheng, J., Olabinjo, L., Balogun, O., & Ershaghi, I. (2013). Global model for failure prediction for rod pump artificial lift systems. Society of Petroleum Engineers - SPE Western Regional / Pacific Section AAPG Joint Technical Conference 2013: Energy and the Environment Working Together for the Future, 687–696.

Lunt, M. (2014). Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. American Journal of Epidemiology, 179(2), 226–235. https://doi.org/10.1093/aje/kwt212

Ma, X., & Zoback, M. D. (2016). Laboratory investigation on effective stress in Middle Bakken: Implications on poroelastic stress changes due to depletion and injection. 50th US Rock Mechanics / Geomechanics Symposium 2016, 2, 1014–1019.

Manchanda, R., Bhardwaj, P., Hwang, J., & Sharma, M. M. (2018). Parent-child fracture interference: Explanation and mitigation of child well underperformance. Society of Petroleum Engineers - SPE Hydraulic Fracturing Technology Conference and Exhibition 2018, HFTC 2018. https://doi.org/10.2118/189849-ms

Martí, L., Sanchez-Pi, N., Molina, J. M., & Garcia, A. C. B. (2015). Anomaly detection based on sensor data in petroleum industry applications. Sensors (Switzerland), 15(2), 2774–2797. https://doi.org/10.3390/s150202774

McDowell, B., Yoelin, A., & Pottebaum, B. (2019, July 25). Production Effects From Frac-Driven Interactions in the Southeastern Midland Basin, Reagan County, Texas. Unconventional Resources Technology Conference. doi:10.15530/urtec-2019-125

Miller, G., Lindsay, G., Baihly, J., & Xu, T. (2016). Parent well refracturing: Economic safety nets in an uneconomic market. Society of Petroleum Engineers - SPE Low Perm Symposium, May 5–6. https://doi.org/10.2118/180200-ms

Morgan, S., & Winship, C. (2014). Counterfactuals and Causal inference: Methods and Principles for Social Research (2nd ed., Analytical Methods for Social Research). Cambridge: Cambridge University Press. doi:10.1017/CBO9781107587991.

Nielsen, M. (2019). Neural Networks and Deep_Learning.

Nind, T.E.W., 1964. Principles of Oil Well Production. McGraw-Hill Book Co, New York.

Olmos, A., & Govindasamy, P. (2015). Propensity Scores: A Practical Introduction Using R. Journal of MultiDisciplinary Evaluation, 11(25), 68–88.

Pearl, J. 2009. Causality: Models, Reasoning and Inference (2nd. ed.). Cambridge University Press, USA.

Pearl, J. and Mackenzie, D. 2017. The Book of Why: The New Science of Cause and Effect (1st. ed.). Basic Books, Inc., USA.

Pearson, Karl, 1901. On lines and planes of closest fit to systems of points in space, Philosophical Magazine, Series 6, vol. 2, no. 11, pp. 559-572.

Peng, Y., & Riped, P. (2019). SPE-196159-MS Artificial Intelligence Applied in Sucker Rod Pumping Wells: Intelligent Dynamometer Card Generation, Diagnosis, and Failure Detection Using Deep Neural Networks. 2, 196159.

Pennel, M., Hsiung, J., & Putcha, V. B. (2018). Detecting failures and optimizing performance in artificial lift using machine learning models. SPE Western Regional Meeting Proceedings, 2018-April. https://doi.org/10.2118/190090-ms

Portis, D. H., Bello, H., Murray, M., Barzola, G., Clarke, P., & Canan, K. (2013). Searching for the optimal well spacing in the eagle ford shale: A practical toolkit. Unconventional Resources Technology Conference 2013, URTC 2013, 1–8. https://doi.org/10.1190/urtec2013-027

Price, B., Haustveit, K., & Lamb, A. (2017, July 24). Influence of Stratigraphy on Barriers to Fracture Growth and Completion Optimization in the Meramec Stack Play, Anadarko Basin, Oklahoma. Unconventional Resources Technology Conference. doi:10.15530/URTEC-2017-2697585

Rafiee, M., & Grover, T. (2017). Well spacing optimization in Eagle Ford shale: An operator’s experience. SPE/AAPG/SEG Unconventional Resources Technology Conference 2017. https://doi.org/10.15530/urtec-2017-2695433.

Rainbolt, M. F., & Esco, J. (2018). Paper title: Frac hit induced production losses: Evaluating root causes, damage location, possible prevention methods and success of remediation treatments, part II. Society of Petroleum Engineers - SPE Hydraulic Fracturing Technology Conference and Exhibition 2018, HFTC 2018, 187192(part I).

Reges, G. D., Schnitman, L., Reis, R., & Mota, F. (2015). A new approach to diagnosis of Sucker Rod Pump Systems by analyzing segments of Downhole Dynamometer Cards. 2015 SPE Artificial Lift Conference - Latin America and Caribbean, 1936(Figure 1), 414–426. https://doi.org/10.2118/173964-ms

Romanenkova, E., Zaytsev, A., Klyuchnikov, N., Gruzdev, A., Antipova, K., Ismailova, L., Burnaev, E., Semenikhin, A., Koryabkin, V., Simon, I., & Koroteev, D. (2019). Real-Time Data-Driven Detection of the Rock-Type Alteration During a Directional Drilling. IEEE Geoscience and Remote Sensing Letters, 1–5. https://doi.org/10.1109/lgrs.2019.2959845

Rosenbaum, P. R., & Rubin, D. B. (1983a). The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41–55.

Rosenbaum, P. R., & Rubin, D. B. (2006). The central role of the propensity score in observational studies for causal effects. Matched Sampling for Causal Effects, 1083, 170–184. https://doi.org/10.1017/CBO9780511810725.016.

Roy, J., (2019). A Crash Course in Causality: Inferring Causal Effects from Observational Data, University of Pennsylvania.

Rohrer (2019). https://e2eml.school/how_convolutional_neural_networks_work.html

Rubin, D. B (1986) "Which Ifs Have Causal Answers (Comment on 'Statistics and Causal inference' by Paul W. Holland)." Journal of the American Statistical Association 81:961–2.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100(469), 322–331. https://doi.org/10.1198/016214504000001880

Russakovsky, O., Deng, J., Su, H. et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 115, 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y

Saghir, F., Gonzalez Perdomo, M. E., & Behrenbruch, P. (2019). Converting time series data into images: An innovative approach to detect abnormal behavior of progressive cavity pumps deployed in coal seam gas wells. Proceedings - SPE Annual Technical Conference and Exhibition, 2019-Septe. https://doi.org/10.2118/195905-ms

Sakhardande, R., Devegowda, D. (2021). Data-Driven Causal Analyses of Parent-child Well Interactions for Well Spacing Decisions. Society of Petroleum Engineers - SPE Hydraulic Fracturing Technology Conference and Exhibition 2021, HFTC 2021.

Sani, A. M., Podhoretz, S. B., & Chambers, B. D. (2015). The use of completion diagnostics in Haynesville shale horizontal wells to monitor fracture propagation, well communication, and production impact. Society of Petroleum Engineers - SPE/CSUR Unconventional Resources Conference, 1–35. https://doi.org/10.2118/175917-ms

Scherz, R. Y., Resources, E. E., Rainbolt, M. F., Corporation, A. C., Pradhan, Y., Resources, E. E., & Tian, W. (2019). Parent, Child, and Vertical Wells in the Midland Basin Lower Spraberry and Wolfcamp Reservoirs.

Shahkarami, A., & Wang, G. (2017). Horizontal Well Spacing and Hydraulic Fracturing Design Optimization: A Case Study on Utica-Point Pleasant Shale Play. Journal of Sustainable Energy Engineering, 5(2), 148–162.https://doi.org/10.7569/jsee.2017.629508.

Sharaf, S. A., Bangert, P., Fardan, M., Alqassab, K., Abubakr, M., & Ahmed, M. (2019). Beam pump dynamometer card classification using machine learning. SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings, 2019-March. https://doi.org/10.2118/194949-ms

Sidahmed, M., & Bailey, R. (2016). Machine learning approach for irregularity detection in dynamic operating conditions. Proceedings - SPE Annual Technical Conference and Exhibition, 2016-Janua. https://doi.org/10.2118/181435-ms

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–14.

Snyder, J., Scott, S., & Kassim, R. (2019). Self-adjusting anomaly detection model for well operation and production in real-time. Society of Petroleum Engineers - SPE Oklahoma City Oil and Gas Symposium 2019, OKOG 2019. https://doi.org/10.2118/195234-ms

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. Journal of clinical epidemiology, 66(8 Suppl), S84–S90.e1. https://doi.org/10.1016/j.jclinepi.2013.01.013

Takacs, G., (2015): Sucker-Rod Pumping Handbook. Gulf Professional Publishing.

Tan, C., Li, G., & No, P. (n.d.). Predicting the Dynamometer Card of a Rod Pump. 2015.

Trial Run (2020). Rooster's crow heralds the sunrise https://www.trialrun.us/roosters-crow-causes-sun-rise-eat-chocolate-win-nobel-prize-statements-not-make-sense-blog-will/ (accessed 11 October 2020).

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, 167.

Vigen (2014). Divorce rate in Maine and per capita consumption of margarine plotted versus time between years 2000 and 2009 https://blogs.ams.org/blogonmathblogs/2017/04/10/divorce-and-margarine/ (accessed 11 October 2020).

Whitfield, T., Hunter Watkins, M., & James Dickinson, L. (2018). Pre-loads: Successful mitigation of damaging frac hits in the eagle ford. Proceedings - SPE Annual Technical Conference and Exhibition, 2018-(September), 24–26. https://doi.org/10.2118/191712-ms

Xu, W., Prioul, R., Berard, T., Weng, X., & Kresse, O. (2019, January 29). Barriers to Hydraulic Fracture Height Growth: A New Model for Sliding Interfaces. Society of Petroleum Engineers. doi:10.2118/194327-MS

Xu, T., Lindsay, G., Zheng, W., Yan, Q., Patron, K. E., Alimahomed, F., Panjaitan, M. L., & Malpani, R. (2018). Advanced modeling of production induced pressure depletion and well spacing impact on infill wells in spraberry, Permian basin. Proceedings - SPE Annual Technical Conference and Exhibition, 2018-Septe. https://doi.org/10.2118/191696-ms

Xu, T., Zheng, W., Baihly, J., Dwivedi, P., Shan, D., Utech, R., & Miller, G. (2019). Permian basin production performance comparison over time and the parent-child well study. Society of Petroleum Engineers - SPE Hydraulic Fracturing Technology Conference and Exhibition 2019, HFTC 2019, 1–38.

Yaich, E., De Souza, O. C. D., Foster, R. A., & Abou-Sayed, I. (2014). A methodology to quantify the impact of well interference and optimize well spacing in the marcellus shale. Society of Petroleum Engineers - SPE Canadian Unconventional Resources Conference 2014, 1, 107–117. https://doi.org/10.2118/171578-ms

Yan, W., & Yu, L. (2015). On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach. Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM, 440–447.

Yazdani A, Boerwinkle E. Causal inference in the Age of Decision Medicine. J Data Mining Genomics Proteomics. 2015;6(1):163. doi:10.4172/2153-0602.1000163.

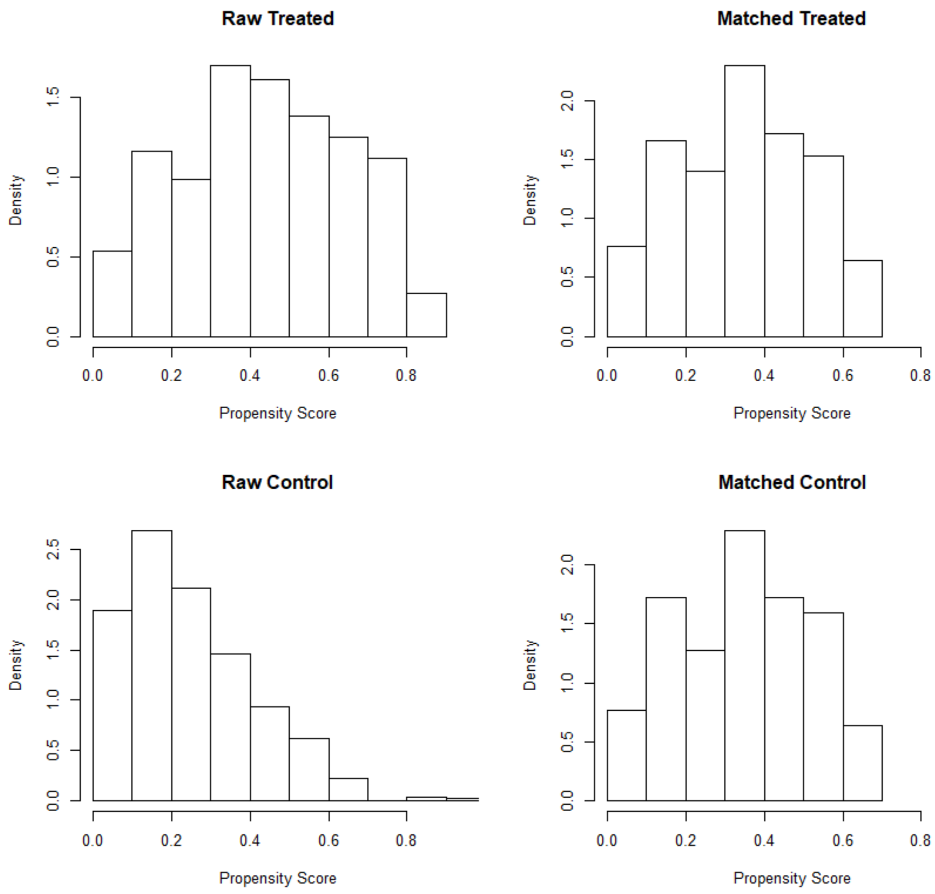
Yu, W., Weijermars, R., Xu, Y., Wu, K., & Sepehrnoori, K. (2017). Impact of well interference on shale oil production performance: A numerical model for analyzing pressure response of fracture hits with complex geometries. Society of Petroleum Engineers - SPE Hydraulic Fracturing Technology Conference and Exhibition 2017, 196–220. https://doi.org/10.2118/184825-ms

Zhang, H., Zeng, Y., Bao, H., Liao, L., Song, J., Huang, Z., Chen, X., Wang, Z., Xu, Y., & Jin, X. (2020). Drilling and Completion Anomaly Detection in Daily Reports by Deep Learning and Natural Language Processing Techniques. https://doi.org/10.15530/urtec-2020-2885

Zhang, J., Cramer, D. D., McEwen, J., White, M., & Bjornen, K. (2020, May 1). Use of Far-Field Diverters To Mitigate Parent- and Infill-Well-Fracture Interactions in Shale Formations. Society of Petroleum Engineers. doi:10.2118/194329-PA

Zoback, M.D, Kohli, A.H. (2019) Unconventional reservoir geomechanics: shale gas, tight oil and induced seismicity. Mark D. Zoback and Arjun H. Kohli. Cambridge, United Kingdom: Cambridge University Press.

Appendix


Raw Treated

Matched Treated

Raw Control

Matched Control

Fig. A1—Propensity Score Distribution before and after matching at 500 ft well spacing with outcome 180-day cumulative production normalized to lateral length (bbl/ft).

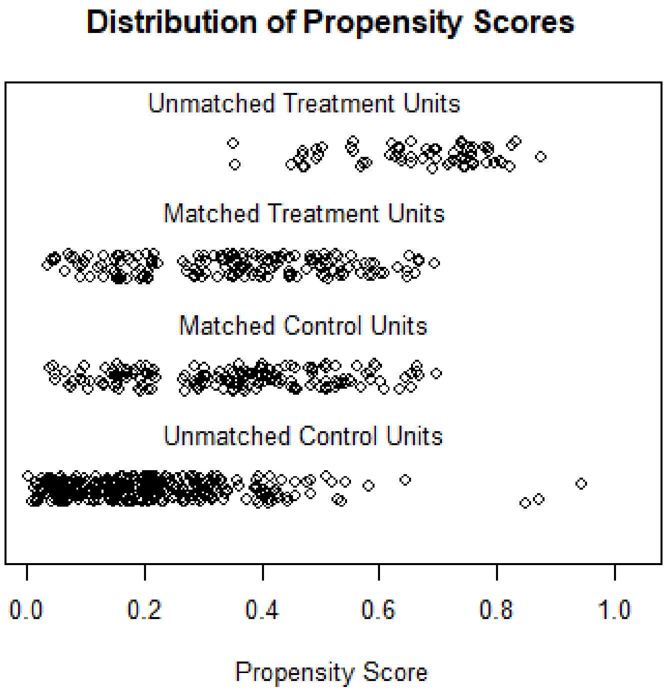


Fig. A2—Density of the propensity scores at 500 ft well spacing across matched and unmatched parent-child well groups with outcome 180-day cumulative production normalized to lateral length (bbl/ft).

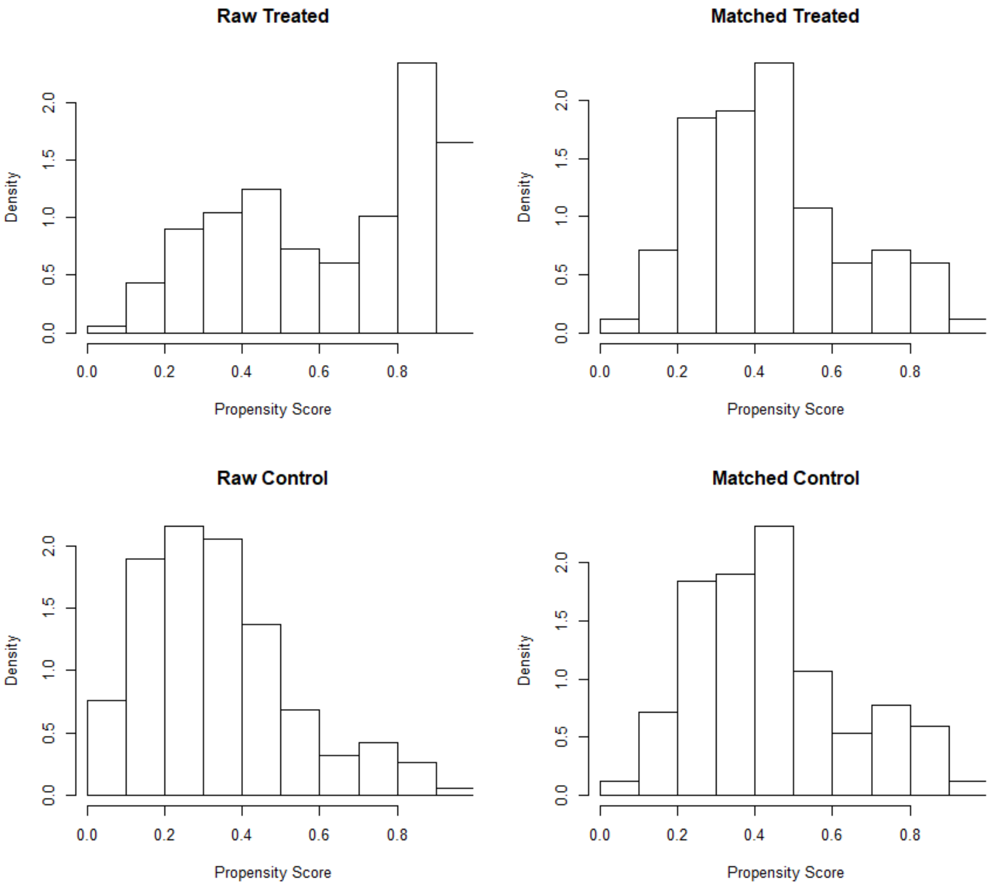


Fig. A3—Propensity Score Distribution before and after matching at 600 ft well spacing with outcome 180-day cumulative production normalized to lateral length (bbl/ft).

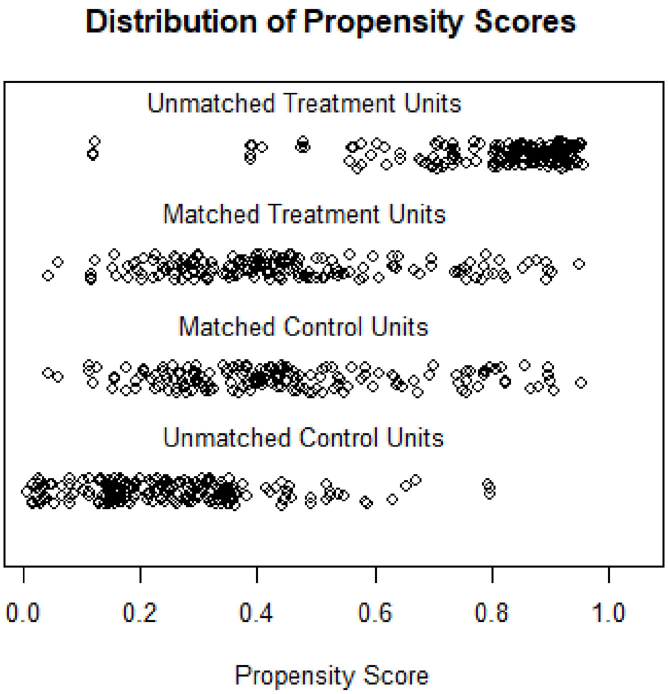
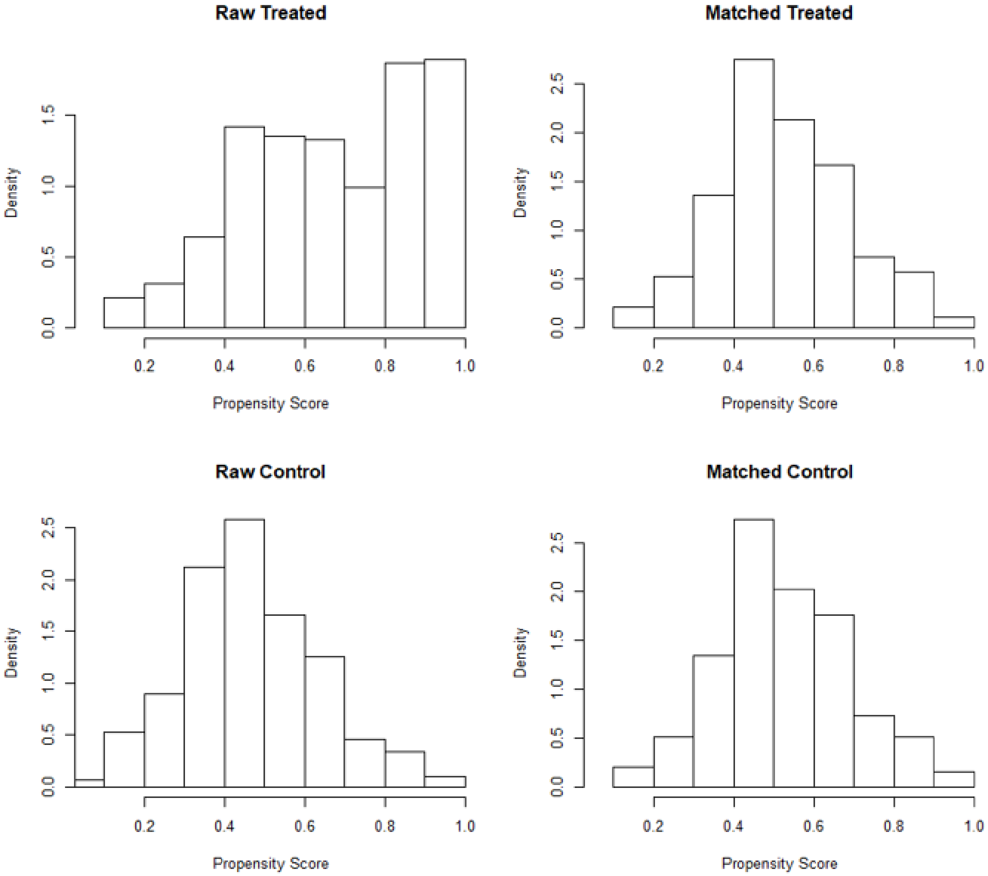


Fig. A4—Density of the propensity scores at 600 ft well spacing across matched and unmatched parent-child well groups with outcome 180-day cumulative production normalized to lateral length (bbl/ft).


Raw Treated

Matched Treated

Raw Control

Matched Control

Fig. A5—Propensity Score Distribution before and after matching at 700 ft well spacing with outcome 180-day cumulative production normalized to lateral length (bbl/ft).

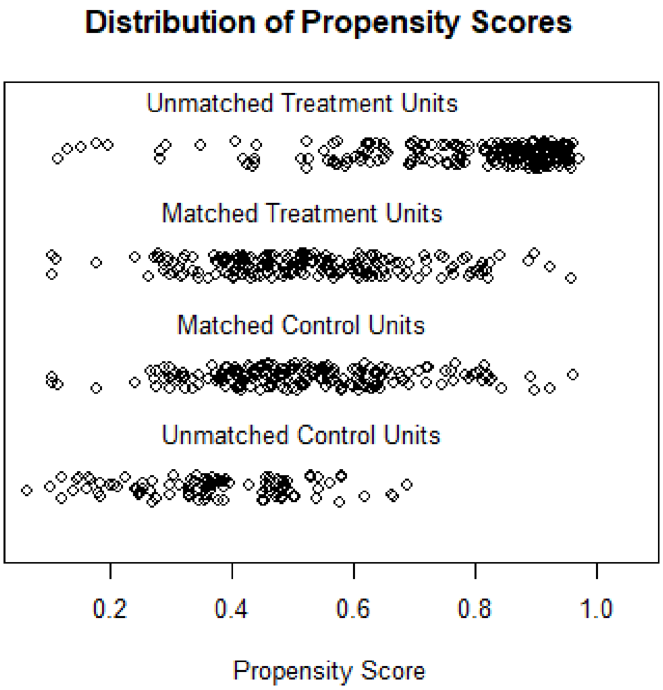


Fig. A 6—Density of the propensity scores at 700 ft well spacing across matched and unmatched parent-child well groups with outcome 180-day cumulative production normalized to lateral length (bbl/ft).

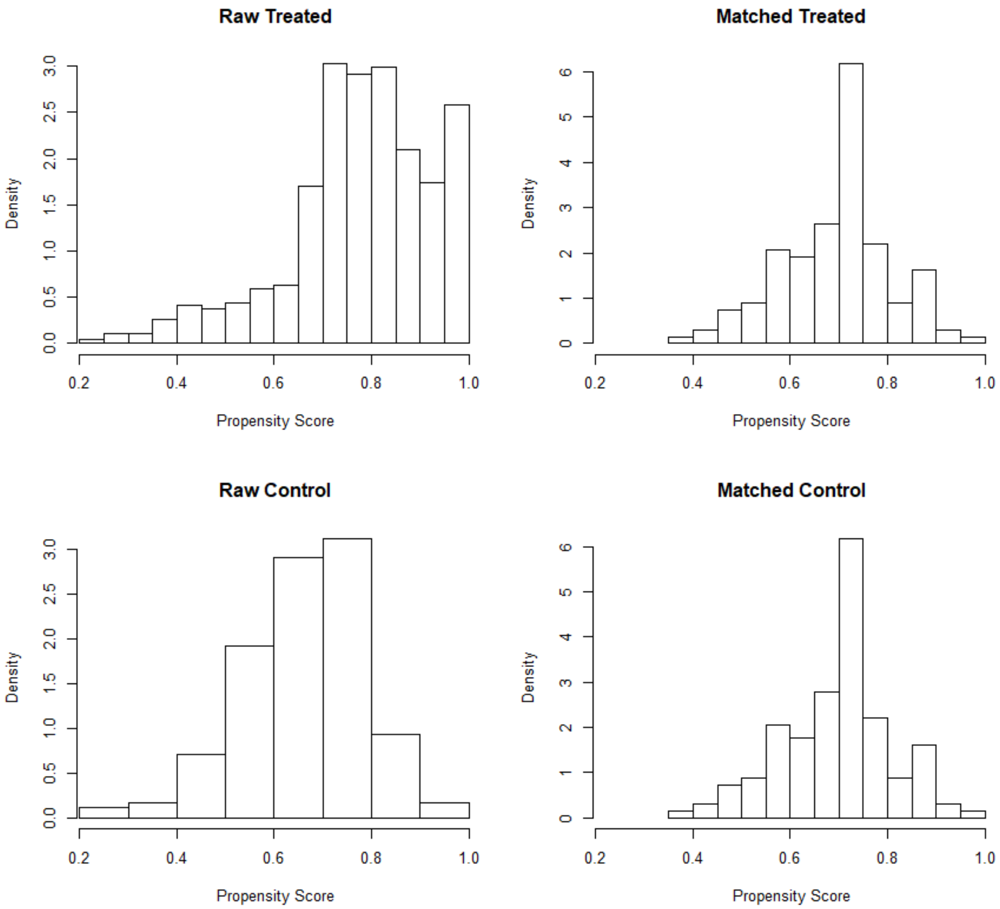


Fig. A 7—Propensity Score Distribution before and after matching at 800 ft well spacing with outcome 180-day cumulative production normalized to lateral length (bbl/ft).

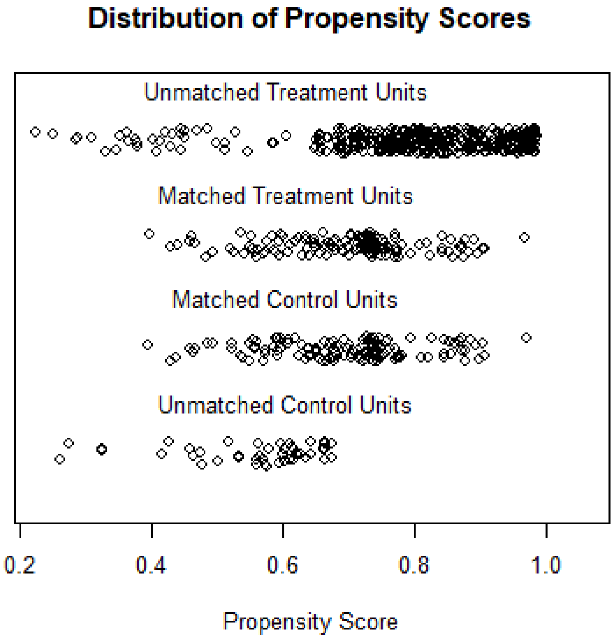
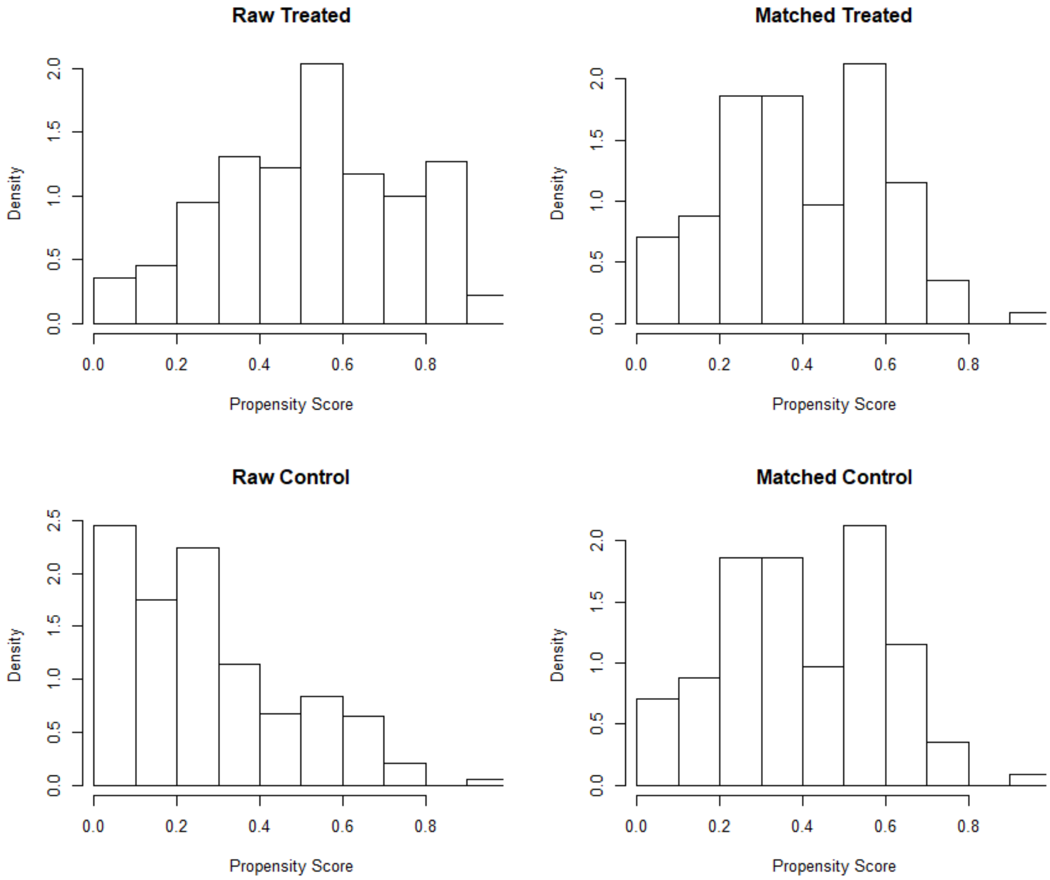


Fig. A 8—Density of the propensity scores at 800 ft well spacing across matched and unmatched parent-child well groups with outcome 180-day cumulative production normalized to lateral length (bbl/ft).


Raw Treated

Matched Treated

Raw Control

Matched Control

Fig. A 9—Propensity score distribution before and after matching at 500 ft well spacing with outcome 360-day cumulative production normalized to lateral length (bbl/ft).

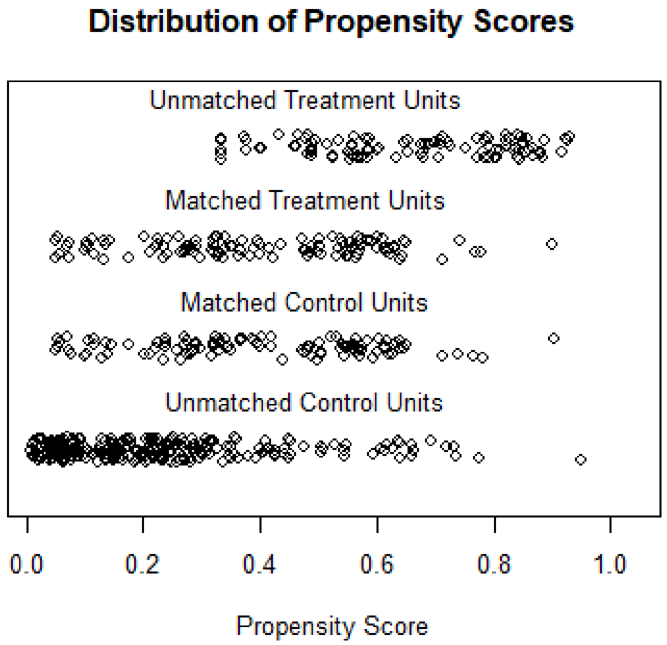


Fig. A 10—Density of the propensity scores at 500 ft well spacing across matched and unmatched parent-child well groups with outcome 360-day cumulative production normalized to lateral length (bbl/ft).

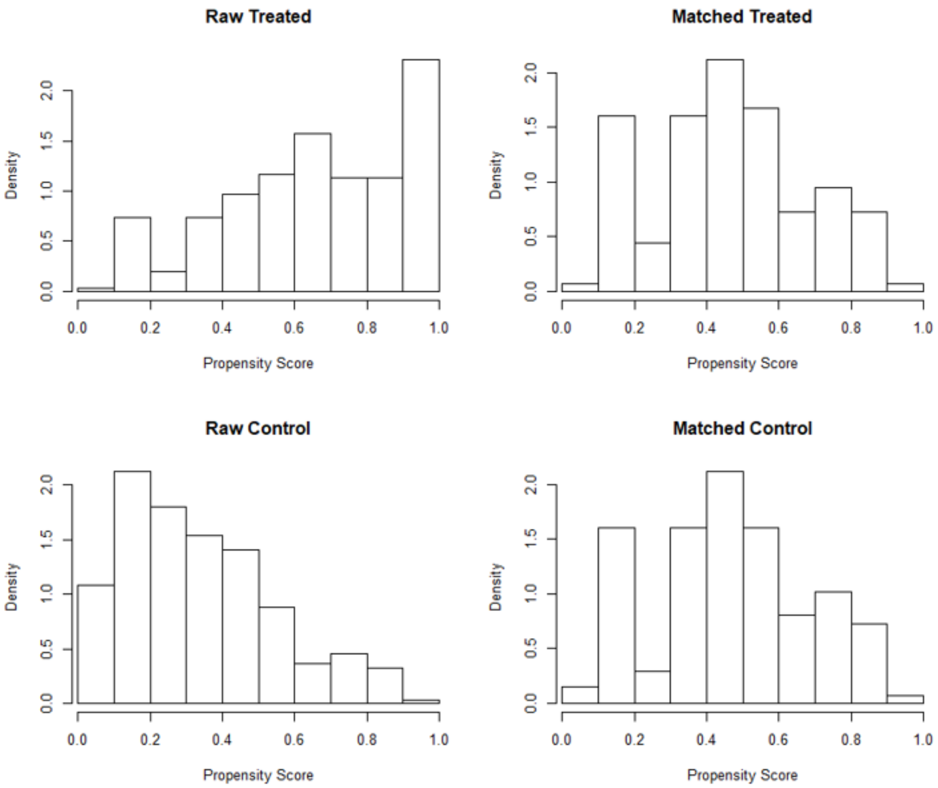


Fig. A 11—Propensity Score Distribution before and after matching at 600 ft well spacing with outcome 360-day cumulative production normalized to lateral length (bbl/ft).

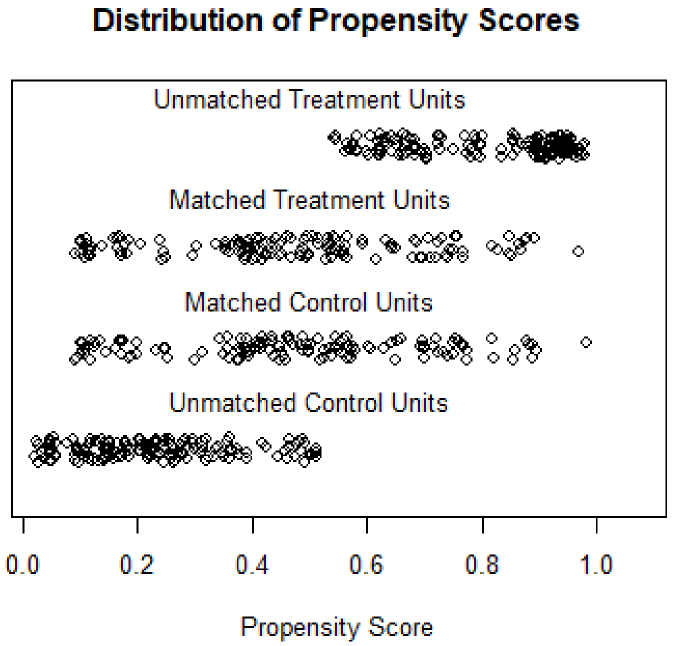


Fig. A 12—Density of the propensity scores at 600 ft well spacing across matched and unmatched parent-child well groups with outcome 360-day cumulative production normalized to lateral length (bbl/ft).

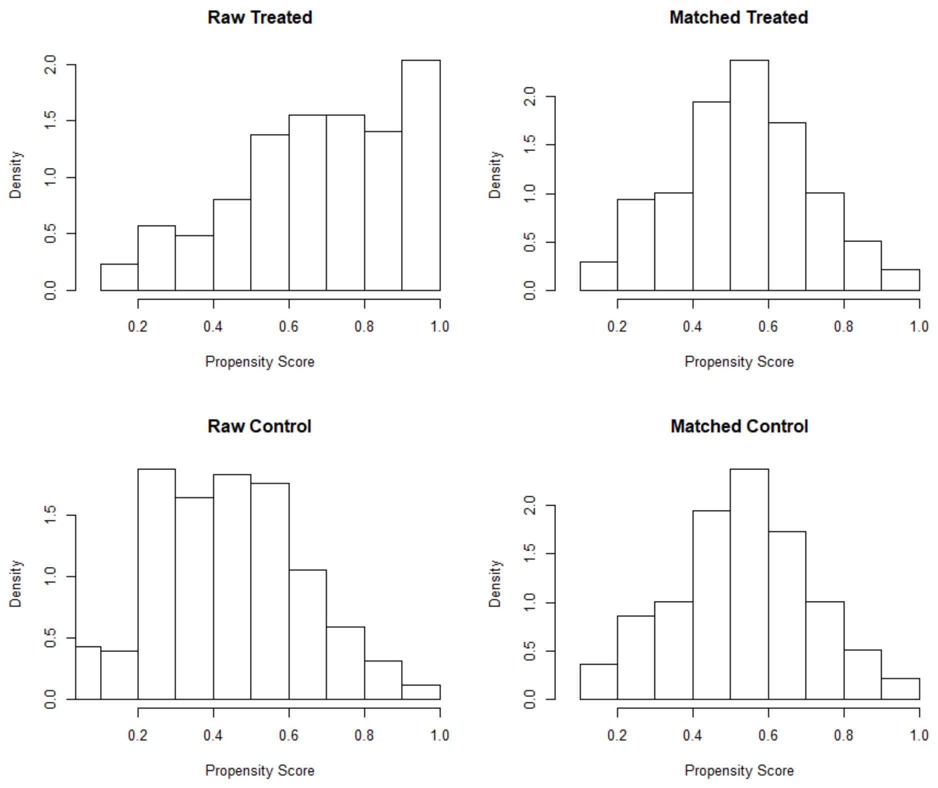


Fig. A 13—Propensity Score Distribution before and after matching at 700 ft well spacing with outcome 360-day cumulative production normalized to lateral length (bbl/ft).

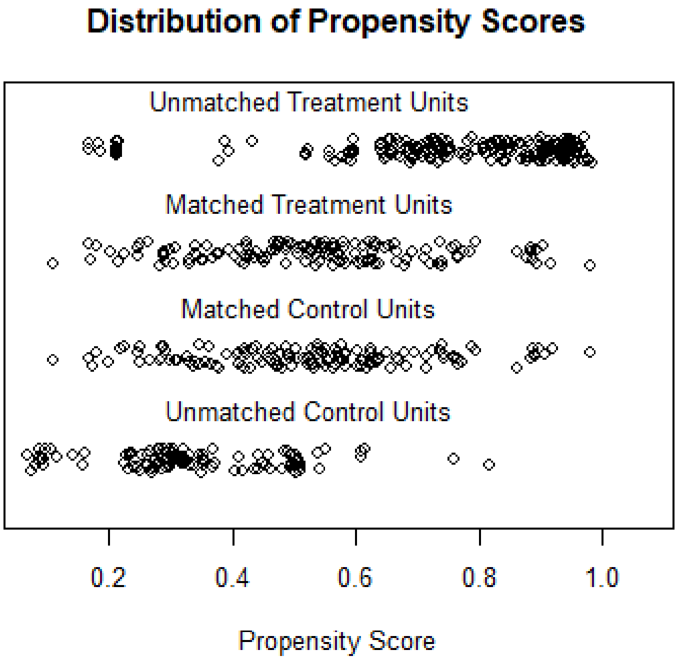


Fig. A 14—Density of the propensity scores at 700 ft well spacing across matched and unmatched parent-child well groups with outcome 360-day cumulative production normalized to lateral length (bbl/ft).

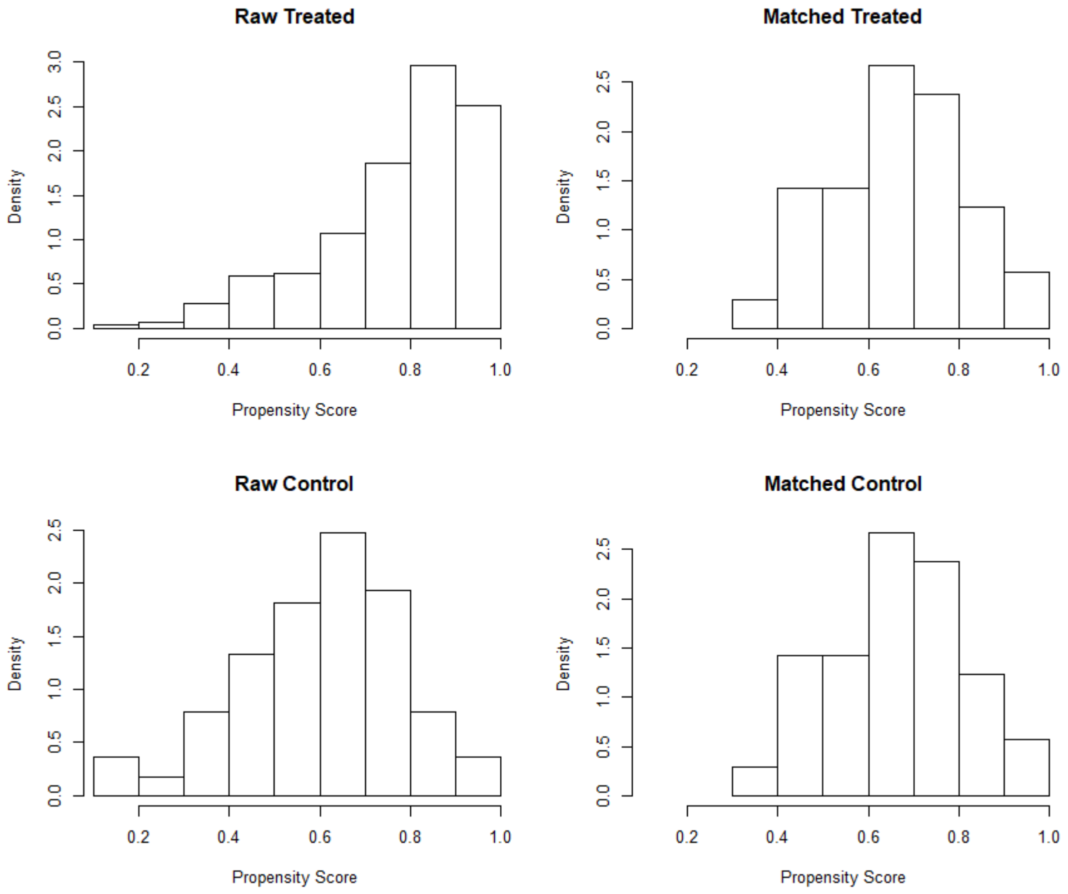


Fig. A 15—Propensity Score Distribution before and after matching at 800 ft well spacing with outcome 360-day cumulative production normalized to lateral length (bbl/ft).

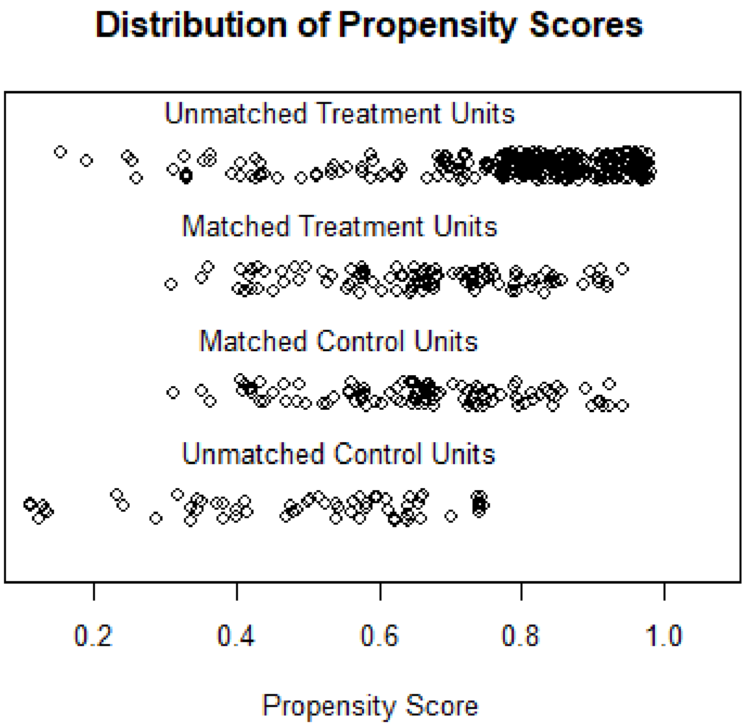


Fig. A 16—Density of the propensity scores at 800 ft well spacing across matched and unmatched parent-child well groups with outcome 360-day cumulative production normalized to lateral length (bbl/ft).

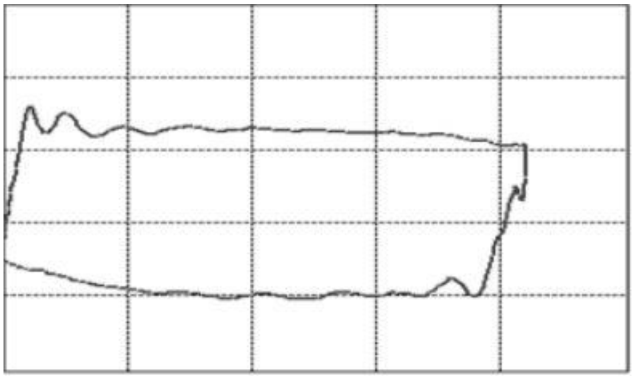


Fig. A 17—Normal Pump card (Tan et al. 2015).

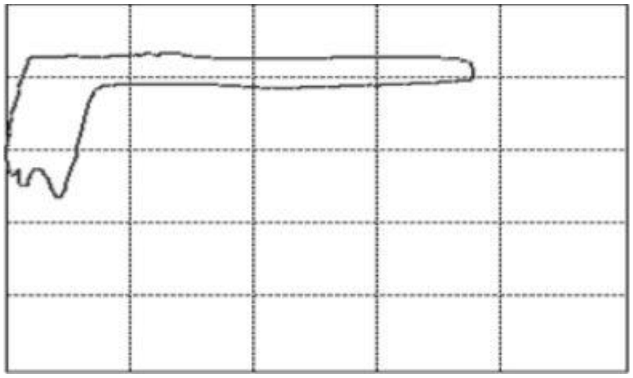


Fig. A 18—Severe case of fluid pound (Tan et al. 2015).

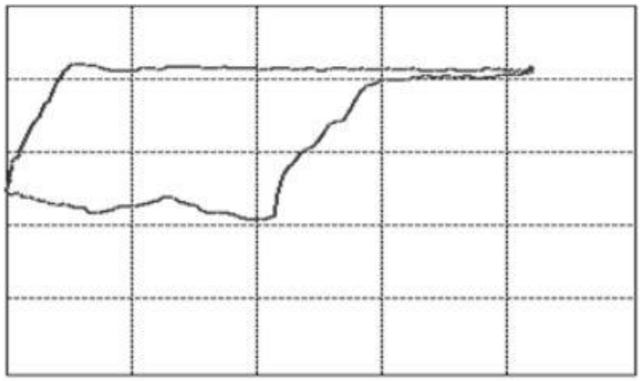


Fig. A 19—Light case of fluid pound (Tan et al. 2015).

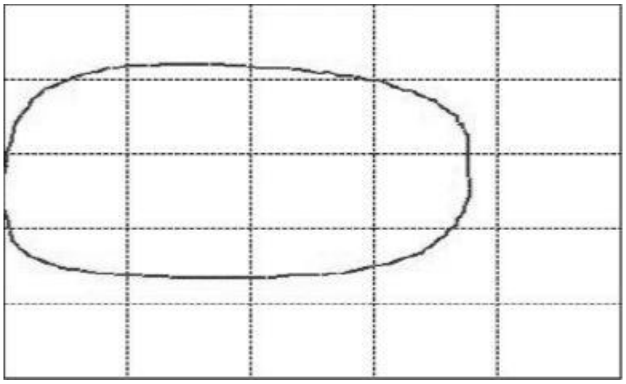


Fig. A 20—Pump with high viscosity oil (Tan et al. 2015).

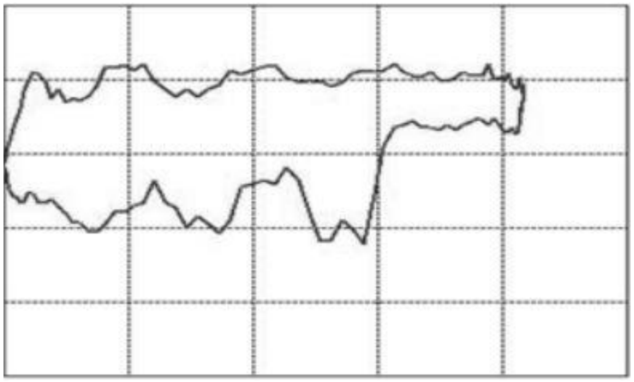


Fig. A 21—Sand interference inside the pump (Tan et al. 2015).

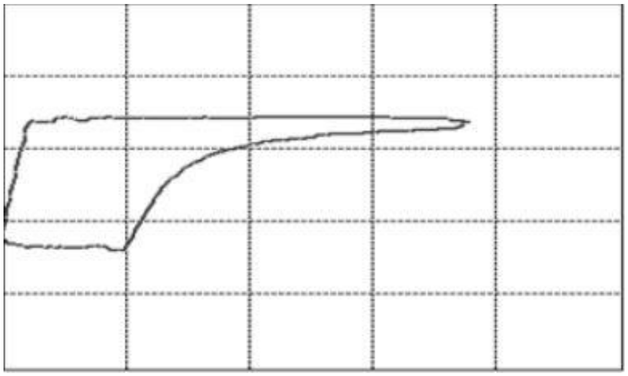


Fig. A 22—Severe gas interference (Tan et al. 2015).

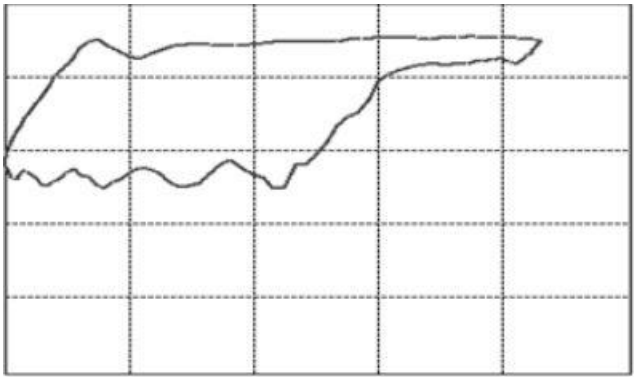


Fig. A 23—Light case of gas interference (Tan et al. 2015).

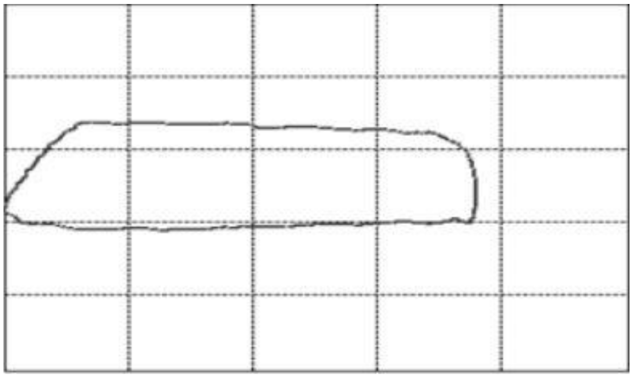


Fig. A 24—Leaking inlet valve (Tan et al. 2015).

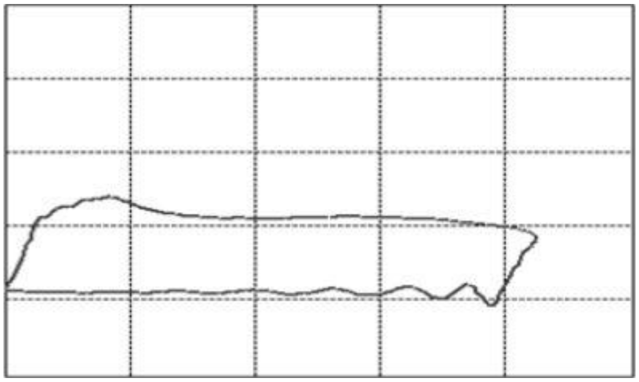


Fig. A 25—Leaking outlet valve (Tan et al. 2015).

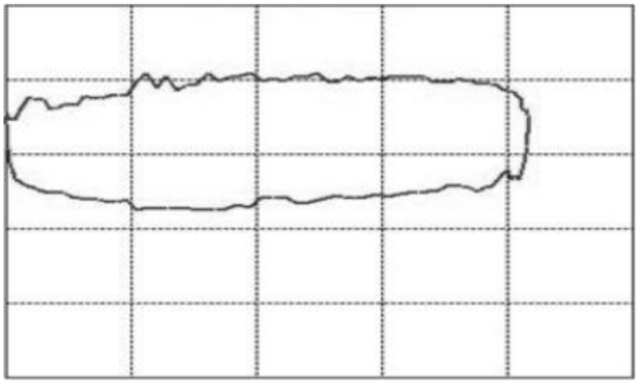


Fig. A 26—Both valves leaking (Tan et al. 2015).

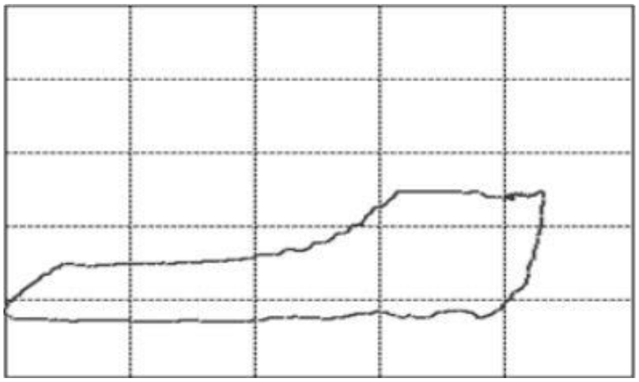


Fig. A 27—Delay in closing of Standing valve (Tan et al. 2015).

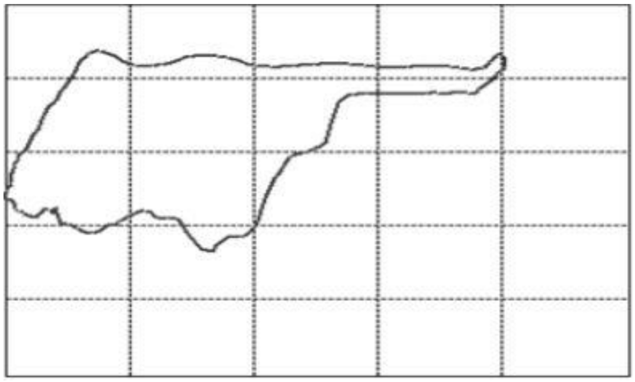


Fig. A 28—Collision between plunger and guide ring (Tan et al. 2015).

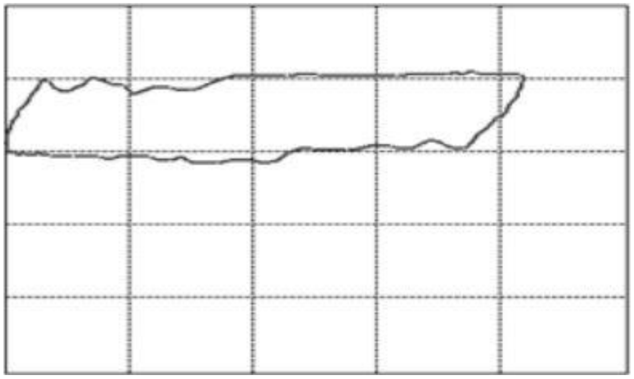


Fig. A 29—Resistance to oil flow (Tan et al. 2015).

Well Spacing (ft)	Child Wells Mean (Cum. 180/Lateral Length) (bbl/ft)	Parent Wells Mean (Cum. 180/Lateral Length) (bbl/ft)	Causal Estimate (Cum. 180/Lateral Length) (bbl/ft)	Lower 95 % CI	Upper 95 % CI	P-value	Degrees of Freedom
800 ft	11.37	12.85	-1.45	-1.93	-0.97	1.9E-9	135
700 ft	11.52	12.36	-0.84	-1.23	-0.44	4.2E-5	192
600 ft	11.57	12.4	-0.84	-1.25	-0.41	0.00013	167
500 ft	11.7	12.5	-0.8	-1.25	-0.34	0.00074	156

Table A1—Causal estimates for all well spacing options with 95% confidence intervals. Outcome is 180-day cumulative production normalized to lateral length (bbl/ft).

Well Spacing (ft)	Child Wells Mean (Cum. 360/Lateral Length) (bbl/ft)	Parent Wells Mean (Cum. 360/Lateral Length) (bbl/ft)	Causal Estimate (Cum. 360/Lateral Length) (bbl/ft)	Lower 95 % CI	Upper 95 % CI	P-value	Degrees of Freedom
800 ft	16.23	18.8	-2.51	-3.4	-1.63	1.5E-7	104
700 ft	16.78	17.64	-0.87	-1.64	-0.1	0.026	138
600 ft	16.31	17.19	-0.87	-1.67	-0.069	0.033	136
500 ft	16.32	17.39	-1.07	-1.8	-0.33	0.004	112

Table A2—Causal estimates for all well spacing options with 95% confidence intervals. Outcome is 360-day cumulative production normalized to lateral length production (bbl/ft).

Parent Well Group Comparison at different spacings.	Parent Wells Group #1 Mean (Cum. 180/Lateral Length) (bbl/ft)	Parent Wells Group #2 Mean (Cum. 180/Lateral Length) (bbl/ft)	Lower 95 % Confidence Interval	Upper 95 % Confidence Interval	P-value	Degrees of Freedom
700 ft and 800 ft	12.36	12.85	-1.04	0.055	0.078	327
600 ft and 800 ft	12.4	12.85	-1	0.12	0.125	302
500 ft and 800 ft	12.5	12.85	-0.92	0.24	0.25	291

Table A3 - Paired t-test outputs conducted between the parent well groups at 800 ft spacing and the parent well groups at the rest of spacings with outcome 180-day cumulative production normalized to lateral length production (bbl/ft).

Parent Well Group Comparison at different spacings.	Parent Wells Group #1 Mean (Cum. 360/Lateral Length) (bbl/ft)	Parent Wells Group #2 Mean (Cum. 360/Lateral Length) (bbl/ft)	Lower 95 % Confidence Interval	Upper 95 % Confidence Interval	P-value	Degrees of Freedom
700 ft and 800 ft	17.65	18.7	-1.04	-2	0.029	242
600 ft and 800 ft	17.19	18.7	-2.49	-0.53	0.0025	240
500 ft and 800 ft	17.39	18.7	-2.39	-0.34	0.008	216

Table A 4—Paired t-test outputs conducted between the parent well groups at 800 ft spacing and the parent well groups at the rest of spacings with outcome 360-day cumulative production normalized to lateral length production (bbl/ft).