Fall 2015

Survey and data analysis of polymer flooding pilot and field applications in China

Yandong Zhang

Department:

SURVEY AND DATA ANALYSIS OF POLYMER FLOODING PILOT AND FIELD APPLICATIONS IN CHINA

by

YANDONG ZHANG

A THESIS

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

IN

PETROLEUM ENGINEERING

2015

Approved by
Dr. Mingzhen Wei, Advisor
Dr. Baojun Bai, Co-advisor
Dr. Ralph Flori

ABSTRACT

Enhanced oil recovery (EOR) processes are regarded as important methods to recovery remaining oil after primary and secondary recovery. It is significant to select the most appropriate EOR process among the possible techniques for a candidate reservoir. EOR screening criteria has been created using available EOR datasets and served as the first step to compare the suitability of each EOR method for a particular reservoir. Most of these datasets are collected from EOR surveys published by Oil & Gas Journal. This study proposes a comprehensive study of a dataset including 55 pilot and field polymer flooding applications in China. Statistical analysis has been used to analyze the data collected. Histograms and box plots combined with violin plots are used to show the distribution of each parameter and present the range of the data. Scatter plots are constructed to compare relationships between different polymer properties and reservoir properties. Screening criteria for polymer flooding has been updated by real pilot and field polymer flooding data. Multiple imputation method is also proposed and implemented on the original dataset and a predicting model to predict incremental oil recovery using reservoir and polymer properties is constructed in steps.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my advisor Dr. Mingzhen Wei for her valuable guidance throughout the process of this thesis.

I would like to appreciate Dr. Baojun Bai for his kindly assistance for providing professional advice and critical reviews.

I'm grateful to my committee member Dr. Ralph Flori for his support and valuable comments.

I also would like to thank my group members Munqith Aldhaheri, Na Zhang, Hao Sun, Yue Qiu, Mingfei Yin for being such wonderful partners and their help and support in timely discussion on the work.

Last but not least, I want to thank my parents for their trust, support, caring and understanding all the time.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

Page

LIST OF TABLES

Page

Page

NOMENCLATURE

Symbol	Description
EOR	Enhance Oil Recovery
IOR	Improved Oil Recovery
OOIP	Original Oil In Place
MI	Multiple Imputation
M	Mobility ratio
μ	Oil viscosity, (cP)
T	Formation temperature (°F)
f	Porosity (%)
k	Reservoir permeability (md)
q	Flow rate
A	Cross-section area
L	Length of the sample
ΔP	Pressure drop
h_{app}	Apparent viscosity
K	Flow consistency index
n	Flow behavior index
$\dot{\gamma}$	Shear rate
HPAM	Hydrolyzed Polyacrylamides
PV	Polymer volume
MW	Molecular Weight
R^2	Coefficient of determination

1. INTRODUCTION

Since oil and gas is in an increasing need nowadays and it's a challenge to recovery oil from existing mature oilfields or hydrocarbon reservoirs because of high water cut and complicated geological settings, it's a necessity to improve oil recovery in the high water cut reservoirs using enhanced oil recovery methods to improve oil recovery. Among them, polymer flooding method is proved to be a successful method in many oilfields and had very significant results. However, whether or not should an oilfield fit the polymer flooding method still needs more investigation.

Screening criteria for polymer flooding is allowed for evaluation and selection of a particular reservoir. Nearly all recently published screening criteria regarding polymer flooding were based on the data collected from the bi-annual EOR surveys of the Oil & Gas journal or some specific fields. The survey missed significant polymer flooding parameters such as formation water salinity, polymer type and molecular weight, polymer concentration, reservoir heterogeneity, and so on. To overcome this issue, a new dataset with important reservoir information and polymer properties from pilot and field tests is necessary to establish for polymer flooding. Thus, this study proposes to develop a new dataset of screening criteria for polymer flooding based on real significant data from pilot and field polymer flooding applications.

Polymer flooding has been widely applied in China for over 20 years and a large number of pilot and field projects have been conducted. These projects include important information to quantify the development of polymer flooding as an EOR method. Nevertheless, most of them have been published in Chinese, and are not accessible to

worldwide research community due to language barrier. Thus, this work is made to collect all relevant information of polymer flooding from available Chinese publications.

The primary objective of this study is to summarize polymer flooding applications in China and analyze the data collected from pilot and field polymer flooding applications using statistical method to provide an updated guidance of screening for polymer flooding. Not all applications are published due to some reasons, thus the data collected are only available in publications in China. This project collected 55 polymer flooding projects after reviewing nearly 200 publications in Chinese, including 31 pilot projects and 24 field projects from 1991 to 2012. Also based on the data collected, a multiple linear regression analysis is conducted to generate a predicting model for oil recovery increased, which is useful for prediction in the future study. Since there are a lot of missing values in different parameters due to the availability from publications, multiple imputation method is used to assist doing the multiple linear regression analysis.

This thesis is organized into six sections. The first section presents the overall introduction and objectives of this study. The second section is literature review about the basic theories and knowledge for polymer flooding methods. EOR screening methods, multiple imputation method and EOR prediction methods are also reviewed in this section. The third section is an introduction for polymer flooding, in which polymer flooding mechanisms and polymer flooding processes are explained in detail. The fourth section describes and analyzes the data collected. In this section, parameters that affect polymer flooding have been discussed and data ranges and distributions of them are also been observed and analyzed. The fifth section presents the multiple imputation method and multiple linear regression analysis method and results. In this section, the multiple

imputation method has been proposed and used to impute the missing data and generate a complete dataset to assist finishing the predicting model using the multiple linear regression. The last section covers the overall summary, conclusions and recommendations for further study.

2. LITERATURE REVIEW

This section includes literature review of overall mechanisms of enhanced oil recovery methods. Also, EOR screening methods and data analysis method including multiple imputation and prediction method are reviewed.

2.1. EOR INTRODUCTION

During the oil recovery process, three mechanisms are included that are primary, secondary and enhanced recovery. EOR, which means enhanced oil recovery, is the implementation of various techniques for increasing the amount of crude oil and gas that can be extracted from an oil field. Enhanced oil recovery is also called tertiary recovery as opposed to primary and secondary recovery. Figure 2.1 shows the three stages oil production by Willhite (1998).

Primary recover is recovery by natural drive energy initially available in the reservoir including rock and fluid expansion, solution gas, water influx, gas cap and gravity drainage. It does not require the injection of any external fluids or heat as driving energies and can produce 5 to15% of the original oil in place (OOIP) (Tzimas, et al., 2005). Secondary recovery is recovery gained by injecting external fluids like water and/or gas, which is mainly for the purpose of pressure maintenance and improved volumetric sweep efficiency. Primary and secondary recovery methods can totally produce about one third of OOIP and leave behind two thirds of OOIP. The remaining oil located in reservoirs that are difficult to access and in pores as a result of capillary pressure and wettability. Also, the interfacial forces, high oil viscosity and reservoir heterogeneity are the factors that lead to high remaining oil saturation after primary and secondary recovery. Enhanced oil recovery is oil recovery to displace and recover the remaining oil by injection of gases or chemicals

and/or thermal energy into the reservoir. Applications of EOR are not restricted to a particular phase in the producing life of the reservoir.

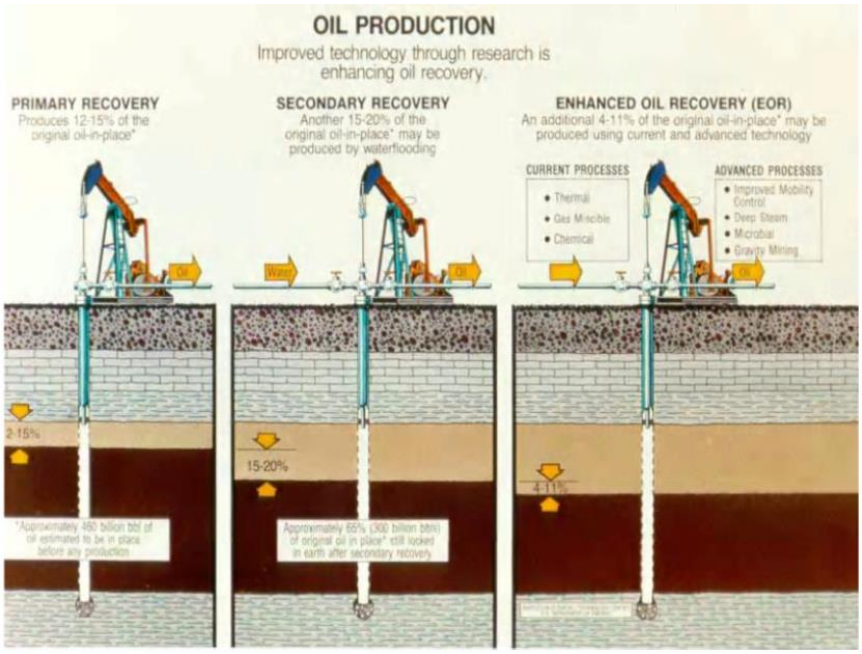

Figure 2.1. Three stages of oil production (Willhite, 1998)

Polymer flooding is one of chemical EOR methods. Chemical methods commonly deal with the injection of interfacial-active components such as surfactants and alkalis, polymers, and their blends. There are several categories of surfactants for foam flooding, including those intended for deep conformance in solvent flooding. Traditionally, the target of chemical methods is to increase the capillary number (Lake, 1989; Thomas, 2008). The best-known method is micellar-polymer (Lake, 1989). After significant technical successes in field trials, the process gave way to new alternatives, such as alkaline-surfactant-polymer

(ASP) flooding, and a renewed interest in surfactant-polymer (SP) flooding. Straight polymer flooding has been a sustained production method in many areas, China, especially in Daqing, has been the most successful case (Satter et al., 2008). In ASP flooding, the polymer acts as a mobility control agent, while the alkali and surfactant act synergistically to widen the range of ultra low interfacial tension (10-3mN/m). In SP flooding, which is a combination of two surfactant (a surfactant and a co-surfactant) co-solvents, no caustic agent is used.

2.2. EOR SCREENING METHODS

In the past decades, reservoir screening criteria for polymer flooding were adopted from the 1984 National Petroleum Council report (Bailey, 1984) and revised EOR screening criteria by Taber et al., (1997). Technology advances will update these criteria. For example, oil viscosity should be low. But people started to inject polymer in high-viscous oil reservoirs (Moe Soe Let et al., 2012; Wassmuth et al., 2009). In addition, more laboratory research has been done (Seright, 2010; Wassmuth et al., 2007). Among the reservoir properties, several aspects should be of concern when selecting the reservoir candidate for polymer flooding, such as reservoir type, reservoir temperature, reservoir viscosity, reservoir permeability, and formation water salinity.

2.3. MULTIPLE IMPUTATION METHOD

Data imputation, which is the practice of 'filling in' missing data with plausible values, is an attractive approach to analyzing incomplete data. It apparently solves the missing data problem at the beginning of the analysis. However, a naive or unprincipled imputation method may create more problems than it solves, distorting estimates, standard errors and hypothesis tests, as documented by Little and Rubin (1987) and others.

Rubin (1987) addressed the question of how to obtain valid inferences from imputed data. MI is a Monte Carlo technique in which the missing values are replaced by $m>1$ simulated values, where m is typically small (e.g. 3-10) (Rubin, 1987). In Rubin's method for 'repeated (m) imputation' inference, each of the simulated complete datasets is analyzed by standard methods, and the results are combined to produce estimates and confidence intervals that incorporate missing-data uncertainty.

Potential uses of MI primarily for large public-use data files from sample surveys and censuses are also addressed by Rubin (1987). With the advent of new computational methods and software for creating multiple imputaions, however, the technique has become increasingly attractive for researchers in the biomedical (Mackinnon, A., 2010), behavioral (Van Ginkel, Joost R., 2010), and social sciences (Saunders, Jeanne A., et al.,2006) whose investigations are hindered by missing data. These methods are documented in a recent book by Schafer (1997) on incomplete multivariate data.

However, few publications were found that multiple imputation been used in oil industry.

2.4. EOR PREDICTION METHODS

It is very important to evaluate an EOR project if is successful and very helpful to predict future performance. One of the most important objectives for prediction is to determine the amount of oil that can be recovered or incremental oil recovery after EOR methods are applied. EOR prediction methods can be separated into three methods, analytical methods, empirical methods and numerical methods.

Analytical methods depend on the reservoir's actual characteristics and most of them were derived from theoretical calculation based on fractional flow theory (Buckley and Leverett, 1942; Welge, 1952; Welge et al., 1961).

Empirical methods were based on the actual available data from laboratory experiments and/or fields. Most of the empirical methods in publications were used for water flooding performance (Wayhan et al., Khan, 1971). However, no empirical methods were found to predict polymer flooding performance.

Numerical methods are the most popular method used to predict the recovery and performance of EOR processes. The advantage is it can predict an EOR performance for a complex reservoir and operation conditions accurately. Numerical models can be used to develop a correlation for predicting oil recovery or incremental oil recovery for different EOR processes (Paul et al., 1982).

Statistical methods are the methods of collecting, summarizing, analyzing, and interpreting variable numerical data. Data collection involves deciding what to observe in order to obtain information relevant to the questions whose answers are required, and then making the observations. Data summarization is the calculation of appropriate statistics and the display of such information in the form of Tables, graphs, or charts.

Statistical analysis relates observed statistical data to theoretical models, such as probability distributions or models used in regression analysis. By estimating parameters in the proposed model and testing hypotheses about rival models, one can assess the value of the information collected and the extent to which the information can be applied to similar situations. Statistical prediction is the application of

the model thought to be most appropriate, using the estimated values of the parameters. Saleh et al., (2014) used statistical methods to analyze data for polymer flooding from lab, pilot and field projects including data processing and different graphical observations.

3. POLYMER FLOODING

Polymer flooding is one of the most important enhanced oil recovery methods and has been used since 1960s (Sandiford, 1964). Polymer flooding is a water-based method used to improve the efficiency of water flooding by reducing the mobility of the brine. In water flooding, fingering problem is the major problem that causes the water break through to the production wells. Polymer flooding can improve the mobility ratio between oil and water so that the sweep efficiency is improved which in turn increases the oil recovery. Polymer flooding has improved oil recovery by 5 to 15% of original oil in place (OOIP) (Zaitoun et al.1998; Wang et al., 2002). In Daqing oilfield in China, the incremental oil recovery is 12% higher than water flooding when using polymer flooding and 120 tons for every ton of polymer injected (Yabin et al., 2001).

3.1. MECHANISMS OF POLYMER FLOODING

There are several displacement mechanisms of polymer flooding. One obvious mechanism in polymer flooding is the reduced mobility ratio of displacing fluid to the displaced fluid so that viscous fingering is reduced which in turn the sweep efficiency is improved. When polymer is injected in vertical heterogeneous layers, cross flow between layers improves polymer allocation in the vertical layers so that vertical sweep efficiency is improved. This mechanism is detailed in (Sorbie, 1991). One economic reason of polymer is the reduced amount of water injected and produced compared with water flooding. Because polymer improves mobility ratio and sweep efficiency so that less water is injected and less water is produced. In some situations such as offshore environments and desert areas, water and the treatment of water could be costly.

In polymer and gel treatment, another mechanism is called disproportionate permeability reduction (DPR). Polymer is also used to shut off water channeling through high-permeability layers and water coning from bottom aquifers. In these kinds of applications, if the injected polymer volume is not large or a large volume may not be injected because of high injection pressure constraints or short gelation time, blocking water channeling or water coning is temporary. Thus eventually, water will bypass the injected polymer zone and cross flow to high permeability zones or bypass the polymer zone to the producing wellbores. To avoid this kind of problem, a weak gel that has high resistance to flow but is still able to flow can be injected deep into reservoir. Thus, a large volume or large area of polymer zone is formed to block water thief zones or channels. Through the use of this mechanism, polymer and gel can reduce water permeability much more than oil permeability.

In a very heterogeneous reservoir, an injected viscous polymer solution may still break through producers early. An idea similar to weak gel was proposed to deal with this problem (Yang and Ni, 1998). Cationic polymer is injected through producers instead of crosslinkers. The injected cationic polymer has high adsorption on the rock and can form a water-insoluble gel to block water channeling or fingering when meet the anionic polymer injected through an injection well.

Another mechanism is related to polymer viscoelastic behavior. The interfacial viscosity between polymer and oil is higher than that between oil and water. The shear stress is proportional to the interfacial viscosity. Because of polymer's viscoelastic properties, there is normal stress between oil and the polymer solution, in addition to shear stress. Thus, polymer exerts a larger pull force on oil droplets or oil films. Oil therefore

can be "pushed and pulled" out of dead-end pores. Thus, residual oil saturation is decreased (Sheng, J, 2010).

3.2. POLYMER PROPERTIES

Below are the concepts of polymer properties.

3.2.1. Polymer Viscosity. Viscosity is the most important parameter for polymer solution. Darcy's law describes the flow of fluid through porous media as

$$\mu = \frac{kA\Delta P}{qL} \tag{1}$$

where q is the flow rate, A is the cross-section area, L is the length of the sample, ΔP is the pressure drop, and μ is the Newtonian viscosity of the flowing fluid.

Polymer solution is non-Newtonian fluid that the apparent viscosity (η_{app}) is not a constant value, which is defined as

$$\eta_{app} = \frac{kA\Delta P}{qL}\mu_p = K\dot{\gamma}^{(n-1)} \tag{2}$$

Some of the factors that affect polymer viscosity including salinity, concentration, shear effect, pH effect, and temperature effect. Below are the discussions about the factors that affect the polymer viscosity and about their relationships using data collected from pilot and field cases from China.

Polymer viscosity is affected by water salinity and divalent ions like calcium (Ca^{+}) and magnesium (Mg^{2+}), which can decrease the viscosity of the polymer solution. The distance between the polymer chain and the molecules decreases as the water salinity increases. Thus, the polymer viscosity decreases as the water salinity increases as Figure 3.1 shows.

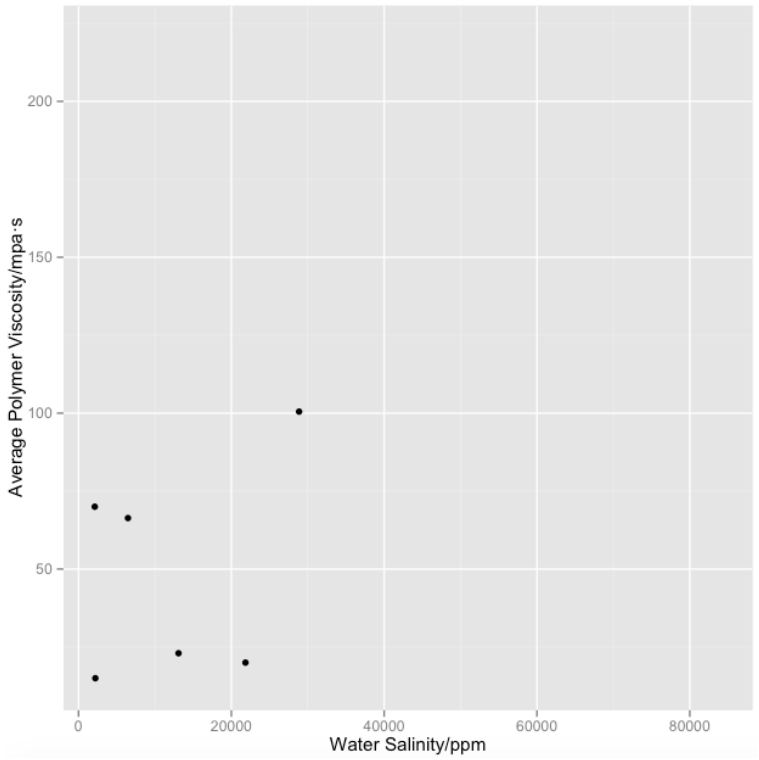

Figure 3.1. Relationship between water salinity and average polymer viscosity

The polymer concentration has a direct relationship with the polymer viscosity. The polymer viscosity increases when the polymer concentration increases as shown in Figure 3.2.

Generally, the viscosity of a polymer solution decreases as the temperature increases as shown in Figure 3.3. However, it is not always the same relationship for different patterns. It depends also on other factors, such as polymer concentration, molecular weight, salt and hydrolyzation (Nouri & Root, 1971).

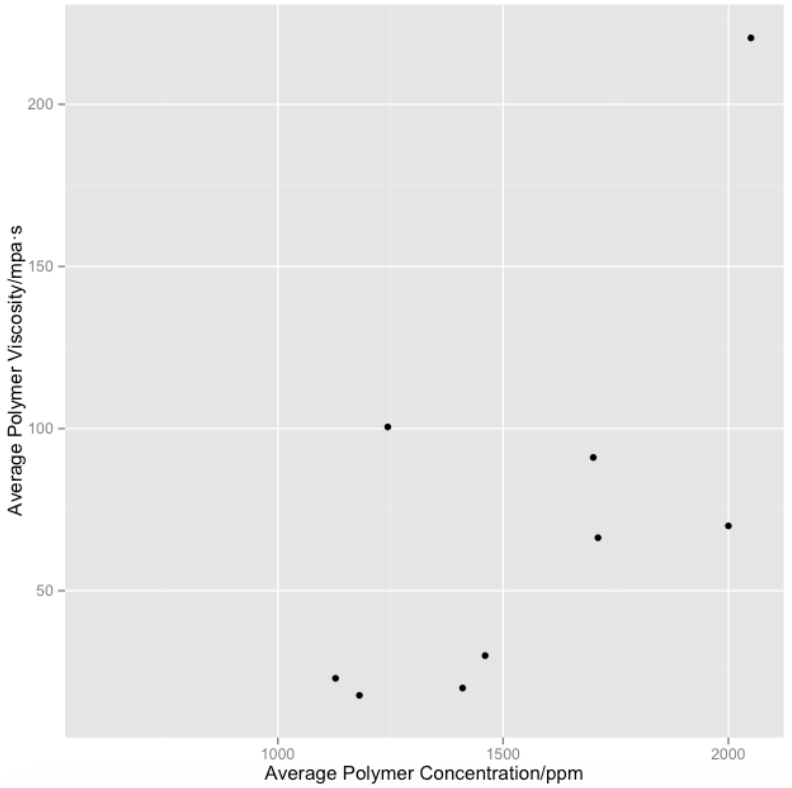

Figure 3.2. Relationship between average polymer concentration and average polymer viscosity

The molecular weight of polymer is related to its molecular size that means polymers with a larger molecular size tend to have a higher molecular weight. Polymer with a higher molecular weight provides higher viscosity as shown in Figure 3.4.

The viscosity of a polymer solution is strongly shear dependent. A power-law model can be used to describe a polymer solution as follow

$$\mu_p = K\dot{\gamma}^{(n-1)} \tag{3}$$

where K is the flow consistency index, n is the flow behavior index and $\dot{\gamma}$ is the

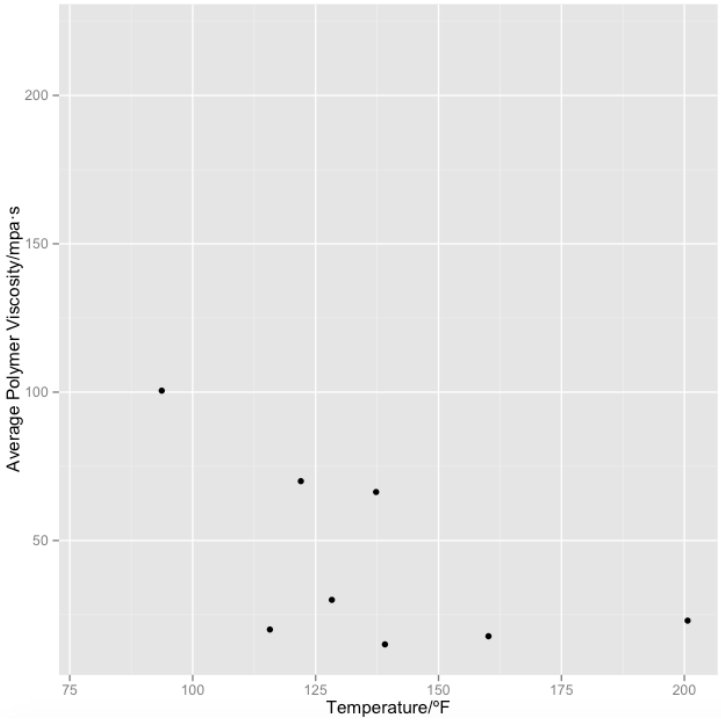

Figure 3.3. Relationship between temperature and average polymer viscosity

shear rate. In the pseudoplastic region, n is less than 1, typically 0.4-0.7 and have little change in different concentrations while K changes.

Initially, polymer viscosity increases with the degree of hydrolysis. A high degree of hydrolysis about 35% increases the polymer viscosity in fresh water and sodium chloride brines (Martin & Sherwood, 1975). However, pH will increase when alkali is added that could decrease the viscosity of polymer because the salt effect of the alkali is dominant compared with the pH effect on hydrolysis.

3.2.2. Polymer Stability. Polymer degradation refers to any processed that break down the molecular structure of macromolecules. The main degradation ways in oil recovery applications are chemical, mechanical and biological. Sorbie (1991) summarizes the research on polymer stability from mid-1970s to late-1980s.

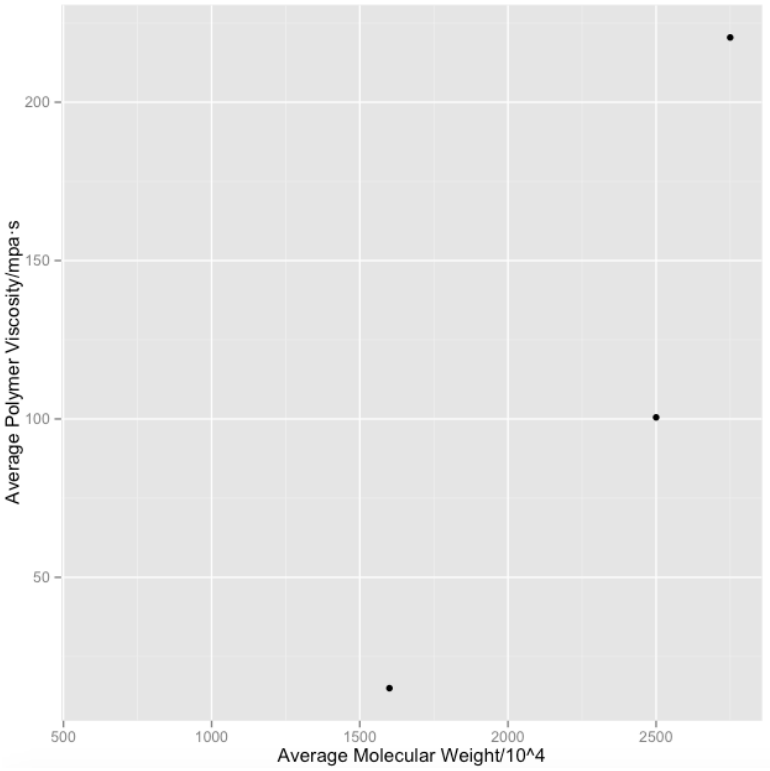

Figure 3.4. Relationship between average polymer molecular weight and average polymer viscosity

3.2.2.1. Chemical degradation. Chemical degradation refers to the breakdown of polymer molecules, either through short-term attack by contaminants like oxygen and iron, or through long-term attack to the molecular backbone by processes such as hydrolysis. Hydrolysis is caused by the intrinsic instability of molecules even in the absence of oxygen or other attacking species. Generally, polymer chemical stability is mainly controlled by oxidation-reduction reactions and hydrolysis.

The presence of oxygen virtually always leads to oxidative degradation of the polyacrylamide polymer. However, the effect of dissolved oxygen on polymer solution viscosity is not significant at a low temperature, and the polymer solution could be sTable

for a long time. As the temperature increases, polymer solution viscosity quickly decreases with time even if a small amount of oxygen exists. As the oxygen concentration increases, the viscosity decreases faster (Luo et al., 2006).

Also, the polymer viscosity could lose by the salinity effect. For example, when the Fe^{3+} concentration was high enough, it will crosslink with polymer to form insoluble gel so that the viscosity loss is significant. Levitt et al. (2010) reported that sodium carbonate and bicarbonate are demonstrated to play a key role in stabilizing polymer against multiple reported sources of degradation, and it seems likely that this is due to their effect on iron stability.

Overall, in a reservoir of low temperature and low hardness, as hydrolysis increases gradually, polymer viscosity may not change within some period of time. Sometimes, the viscosity may increase initially. In a reservoir of low temperature and high hardness, polymer viscosity decreases slowly as hydrolysis increases gradually. Finally, precipitation may occur. In a reservoir of high temperature and low hardness, polymer viscosity decreases sharply as hydrolysis increases rapidly due to the strong temperature effect, but precipitation may not occur. In a reservoir of high temperature and high salinity, polymer viscosity decreases sharply as hydrolysis increases rapidly, and precipitation may occur.

3.2.2.2. Mechanical degradation. Mechanical degradation describes the breakdown of molecules in the high flow rate region close to a well as a result of high mechanical stresses on the macromolecules. Shear degradation reduces the size of the molecules and causes the polymer to lose the viscosity needed for mobility control (Maerker, 1975). Polymer degradation occurs at high shear rates when polymer molecules begin to degrade due to high fluid stresses (viscoelastic stresses) that are generated by elongational

deformation (Maerker, 1975). Mechanical degradation first occurs when the polymer passes from the well bore to the porous media (Mareker, 1975; Seright, 2010). Zaitoun et al., (2011) found that shear degradation could occur at different stages, such as during flow through downhole valves and chokes under high pressure, flow through perforations at a high rate, a use of a shearing device, and recirculation with a centrifugal pump. However, not all polymers exhibit mechanical degradation. For instance, biopolymers like xanthan whose molecules are more rigid than polyacrylamide and have greater resistance to shear degradation thus are much sTable at a high shear rate (Seright, 2008; Stanislav & Kabir, 1977)

3.2.2.3. Biological degradation. Biological degradation refers to the microbial breakdown of macromolecules of polymers by bacteria during storage or in the reservoir. Although the problem is more prevalent for biopolymers, biological attack may also occur for synthetic polymers. It has been found that HPAM can provide nutrition to sulfate-reducing bacteria (SRB). As the number of SRB increases, HPAM viscosity decreases. For example, when the number of SRB reaches 36000/mL, the viscosity loss of HPAM of 1000 mg/L is 19.6% (Luo et al., 2006). Biological degradation is critical only at low temperatures or without using effective biocides. The use of a biocide is the almost the most prevalent answer to biological degradation. Perhaps the most common biocide used in oilfield applications in the past was formaldehyde (HCHO) diluted in aqueous solution (O'Leary et al., 1985; Luo et al., 2006). Because of the toxic of formaldehyde, applications are limited these days.

Also, if such a biocide is used, it may affect other chemicals in the package that are used to protect the polymer; for example, it may interact with the oxygen scavengers.

Bacterial attack has been observed in at least two field tests (van Horn, 1981; Bragg et al., 1982).

3.3 POLYMER FLOW BEHAVIOR IN POROUS MEDIA

The polymer flow behavior in porous media can be categorized as polymer retention and inaccessible polymer volume.

3.3.1. Polymer Retention. Typically, some polymer is retained when a polymer solution flows through porous media. Polymer retention includes adsorption, mechanical trapping, and hydrodynamic retention. Willhite and Domin-guez (1977) discussed these three different mechanisms. Mechanical entrapment and hydrodynamic retention are related and occur only in flow-through porous media. Mechanical entrapment is viewed as occurring only when larger polymer molecules become lodged in narrow flow channels (Willhite and Dominguez, 1977). The significance of the mechanical entrapment depends on the pore size distribution. It is more like a mechanism for polymer retention in low-permeability formation (Szabo, 1975; Dominguez and Willhite, 1977). If the entrapment process acts on polymer molecules about the average size in the distribution, it will inevitably lead to a buildup of material close to the injection well, which gives an approximately exponential penetration profile into the formation. This will eventually lead to pore blocking and well plugging that is indeed unsatisfactory. This is one reason that the polymer flooding should be implemented in a high permeability formation.

Adsorption refers to the interaction between polymer molecules and the solid surface. This interaction causes polymer molecules to be attached to the surface of the solid, mainly by physical adsorption, van der Waals forces, and hydrogen bonding. In fact, the polymer

occupies surface adsorption sites. Adsorption generally depends on the surface area exposed to the polymer solution.

Chauveteau and Kohler (1974) did the experiment that the total level of retention increases when the fluid flow rate increased after a steady state reached in a polymer retention experiment in a core. This type of rate-dependent retention called hydrodynamic retention and is not understood as well. Fortunately, it is generally thought to give a small contribution to the total retained material (Sorbie, 1991).

For these three mechanisms of polymer retention, adsorption is a fundamental property of the polymer-rock surface-solvent system and is the most important mechanism. Because it is difficult to distinguish these three mechanisms in dynamic flood test, the term 'retention' is simply used to describe the polymer loss and sometimes just use the term 'adsorption', which is discussed more often in literature related to this topic.

3.3.2. Inaccessible Pore Volume. Inaccessible pore volume is a general characteristic of polymer flow in porous media. The polymer molecules cannot flow through those pores when polymer molecular sizes are larger than some pores in a porous medium. The volume of those pores that cannot be accessed by polymer molecules is called inaccessible pore volume (IPV). In an aqueous polymer solution with tracer, the polymer molecules will run faster than the tracer because they flow only through the pores that are larger than their sizes. This results in earlier polymer breakthrough in the effluent end. On the other hand, because of polymer retention, the polymer breakthrough is delayed. In other words, if only polymer retention is considered, the polymer will arrive in the effluent later than the tracer.

Another fact is that both polymer molecules and pores have a wide range of size distribution. Some small polymer molecules can flow through small pores, which tends to

help the polymer flow with the tracer. However, IPV has been observed in all types of porous media for both synthetic polymers and biopolymers. Several models have been offered to explain why IPV occurs (DiMarzio and Guttman, 1970; Chauvetean, 1982; Chauveteau and Kohler, 1984; Kolodziej, 1987), but none has gained universal acceptance (Green and Willhite, 1998). Laboratory data indicate that inaccessible pore volume is usually greater than adsorption loss for polymers following a micellar solution. The inaccessible pore volume in laboratory cores typically is 20% (Trushenski et al., 1974).

4. DATA COLLECTION AND ANALYSIS

4.1. DATA PREPARATION

This study starts from collecting polymer flooding field and pilot projects data. 55 projects are found from different oil fields and blocks in China from publications as shown in Table 4.1. The data preparation part includes two steps below: data collection and statistical analysis.

4.1.1. Data Collection. A dataset was created by collecting polymer flooding pilot and field projects data from the published report from China year 1991 to 2012.

Dataset in this study are classified into four categories. As shown in Table 4.1, the first category is ‘reservoir properties’, which includes general field information as like formation type, net thickness, temperature, etc. The second category is ‘polymer properties’ which includes polymer concentration, polymer molecular weight, etc. The third category is ‘well information’, which includes well pattern, well spacing, injection rate, etc. The fourth category is ‘evaluation’, which includes water cut before polymer flooding, oil recovery increased after polymer flooding, etc.

Table 4.2 shows the numbers of blocks that from different oil fields in China.

4.1.2. Types of Projects. All the projects are pilot projects and field projects from China. There are total 55 projects in this study; among them are 24 field projects and 31 pilot projects as Figure 4.1 shows.

4.1.3. Projects Start Year. As Figure 4.2 shows, projects in this study started from 1991 to 2012, most of the field projects which is about 7 started in 1996 in Daqing oil field while most of the pilot projects which is about 6 started in 2004 in Shengli oil field and other oil fields.

Table 4.1. Four different data categories

Reservoir Properties	Basic information: depth, net thickness, formation type Reservoir rock properties: reservoir temperature, porosity, average permeability, Dykstra-Parsons Coefficient, reservoir bubble point pressure, original formation pressure, reservoir present pressure Reservoir fluid properties: oil viscosity, oil gravity, formation water salinity
Polymer Properties	Polymer molecular weight, polymer concentration, polymer viscosity
Well Information	Injection well numbers, production well numbers, injecting pressure, injecting rate, well pattern, well spacing
Evaluation	Polymer volume injected, water cut before polymer flooding, water cut after polymer flooding, water cut decreased after polymer flooding, oil recovery increased after polymer flooding

Table 4.2. Blocks in different oil fields

Oil Field name	Blocks in each oil field
Daqing	21
Henan	15
Shengli	5
Huabei	3
Changqing	3
Liaohe	2
Bohai	2
Yanchang	1
Xinjiang	1
Zhongyuan	1
Jidong	1

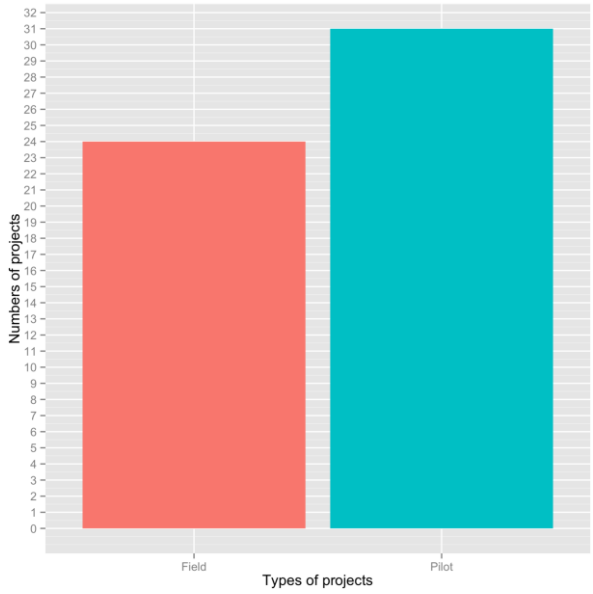

Figure 4.1. Numbers of different types of projects

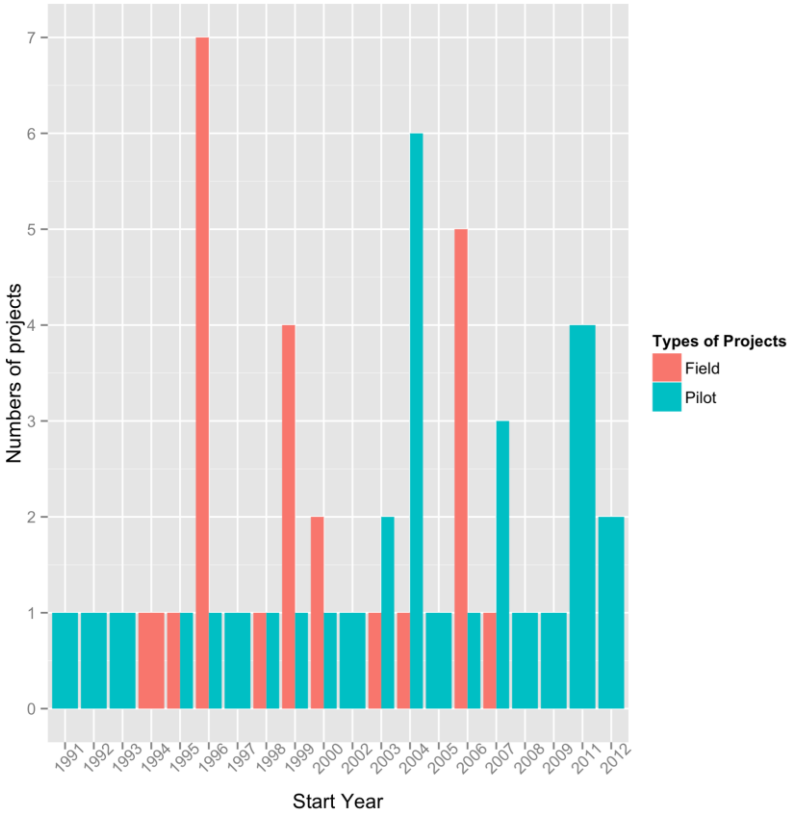

Figure 4.2. Start year of polymer flooding projects in China

4.1.4. Formation Type. Most polymer flooding projects were conducted in sandstone reservoirs. But many polymer projects were carried out in carbonate reservoirs as well according to (Manrique et al., 2007) survey. So far, most successful polymer projects have been implemented in sandstone reservoirs. The typical example is the large-scale applications of polymer flooding in Daqing oilfield in China, where 10-12% average incremental oil recoveries were obtained between 1996-2010. The typical polymer projects demonstrate that sandstone reservoir is still preferable target for polymer flooding project.

As Figure 4.3 shows, almost all the formation type in this study is sandstone; only one formation type of conglomerate is in Xinjiang oil field.

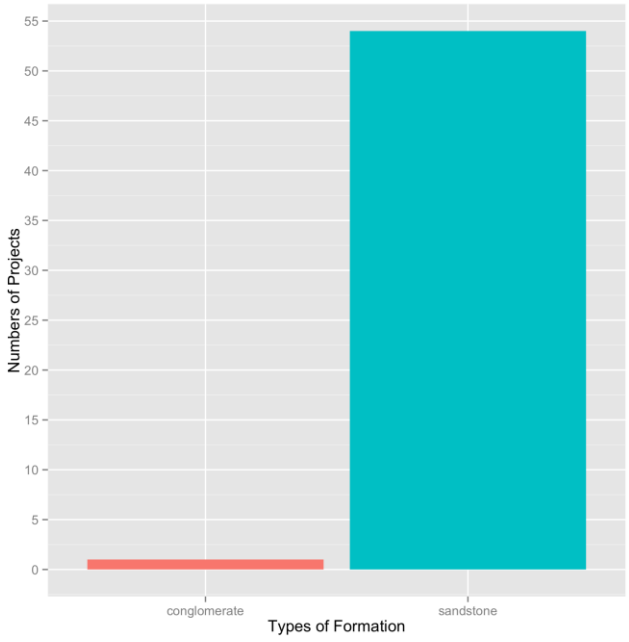

Figure 4.3. Types of formation of different reservoirs

4.1.5. Types of Polymers. Basically, two types of polymers are used in enhancing oil recovery, synthetic polymers like partially hydrolyzed polyacrylamide (HPAM) and biopolymers like xanthan. Derivatives and variations of them are developed to fit specific needs. HPAM polymers are much more widely used than biopolymers, because HPAM has advantages in price and large-scale production. Wang et al. (2006a) believes that HPAM solutions significantly exhibit greater viscoelasticity than xanthan solutions. Associating polymer (AP) is a new kind of polymer being used for polymer flooding. Studies have reported that oil recovery increased 6% more than HPAM when using AP solution in core flooding experiments (Yabin et al., 2001;Reichenbach et al., 2011). However, it is still not been popular used in field applications for polymer flooding.

Polymers that been used in this study are almost all the HPAM which is 53 projects while only two projects used AP to process the test as Figure 4.4 shown.

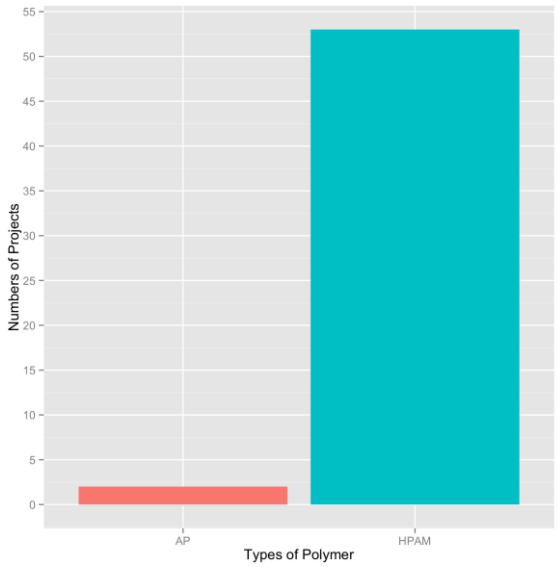

Figure 4.4. Types of polymer that used in different projects

4.2. STATISTICAL ANALYSIS

- **Histogram**

Frequency histograms show the distributions of the project numbers of different ranges of different parameters. From frequency histograms, numbers of projects in each range could be observed and analyzed. Thus, frequency distribution of projects is shown in each parameter.

- **Box plot and violin plot**

Box plots are used during numerical analysis for dataset. Box plots show minimum, first quartile, median, third quartile, maximum, and mean values for each parameter. Also lower limit and upper limit could be calculated from first quartile and third quartile. Data points are regarded as outliers if their values are below the lower limit or above the upper limit. Violin plots are plots that look like a violin that are used to show the distribution of different parameters.

A box plot combined with a violin plot is created to describe the following five values of a dataset: minimum, 1st quartile, median, the 3rd quartile, and maximum. The upper limit is defined as Q3 (3rd quartile) +1.5*(Q3-Q1) and the lower limit is defined as Q1 (1st quartile) -1.5*(Q3-Q1). The outliers are the values that higher than the maximum value or lower the minimum value. The maximum observation is the maximum value in the dataset besides the outliers, same with the minimum value. A schematic of a box plot combined with a violin plot is shown in Figure 4.5 below.

- **Scatter plot**

Scatter plots are used to describe a specialized relationship that compares two related parameters from reservoir. In this study, scatter plots are mainly used to show the

relationship between polymer properties and reservoir properties like polymer concentration versus formation water salinity and between evaluation parameters like incremental oil recovery versus polymer volume injected.

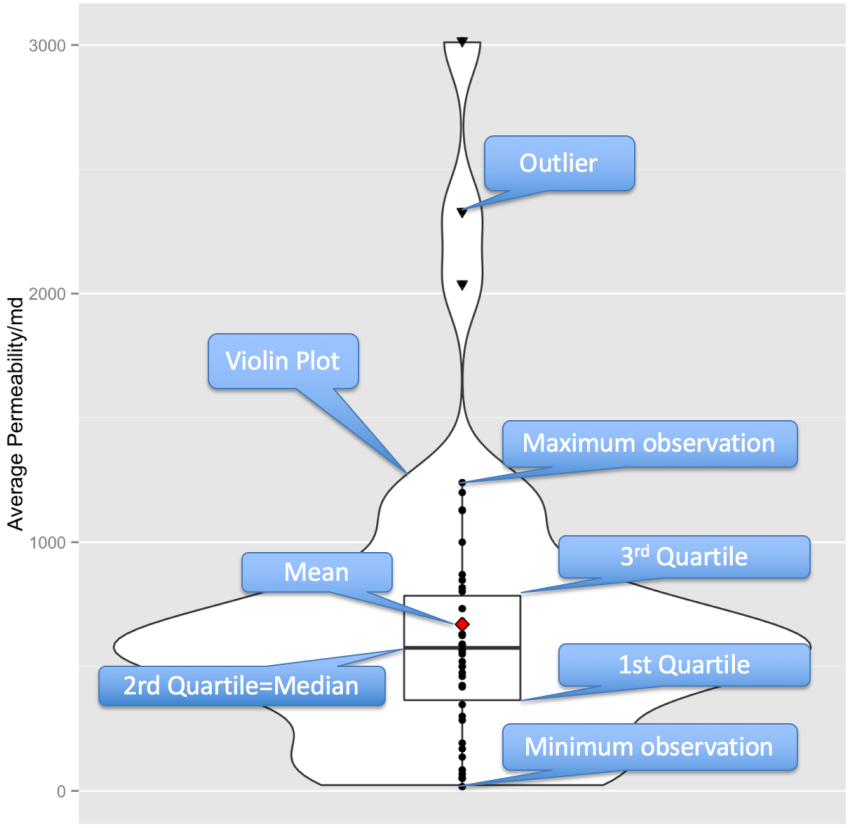

Figure 4.5. Schematic of a boxplot combined with a violin plot

4.2.1. Reservoir Properties. Below are the statistical analysis of reservoir properties.

4.2.1.1. Depth. As shown in Figure 4.6, for field cases, the minimum and maximum value are 3215ft and 5139ft and the mean value is 4106ft. For pilot cases, the minimum

and maximum value are 1558ft and 8186ft and the mean value is 5135ft. Most of depth values of field cases fall into the range of 3000ft to 6000ft as shown in Figure 4.6.

Combining field and pilot cases together as shown in Figure 4.7, depth values formed a formal distribution and most of values fall into the range of 3000ft to 6000ft.

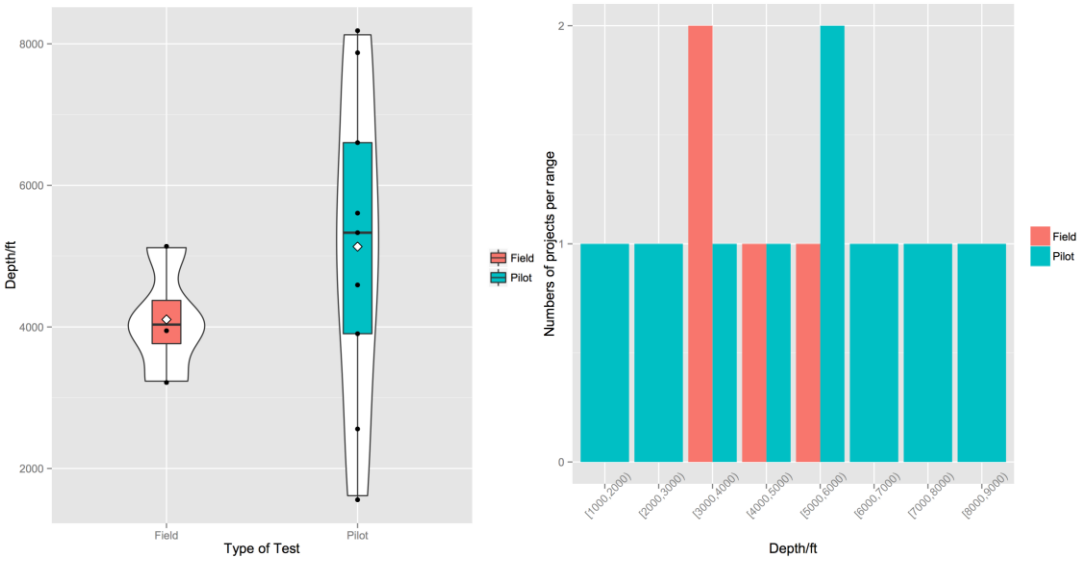

Figure 4.6. Pilot and field depth information (A) box and violin plot (B) histogram

4.2.1.2. Net thickness. Out of the 55 cases studied, 27 cases reported their net thickness information including 10 field cases and 17 pilot cases. For field cases, as Figure 4.8 shows, the minimum value is 13.2ft, the maximum value is 54.8ft and the mean value is 36ft. Figure 4.8 shows a normal distribution for field cases that most values fall into the range of 40-50ft.

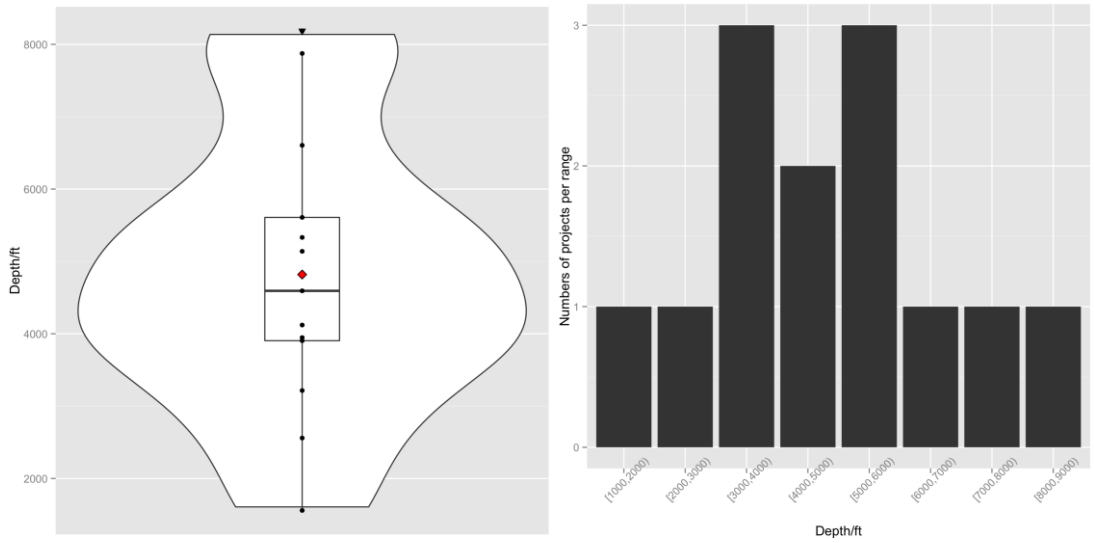

Figure 4.7. Total depth information (A) box and violin plot (B) histogram

For pilot cases, as Figure 4.8 shows, the minimum value is 12.1ft, the maximum value is 53.8ft and the mean value is 33.46ft. There's not too much variance between two test types. Figure 4.8 also shows a multimodal distribution for pilot cases with two peaks, one is the range of 10-20ft and the other is the range of 40-50ft. There are 5 cases and 6 cases fall into the ranges respectively.

Combing two tests type together, as Figure 4.9 shows, the minimum value is 12.1ft, the maximum value is 54.8ft and the mean value is 34.4ft. Figure 4.9 also shows a multimodal distribution with two peak ranges of 10-20ft and 40-50ft. 7 cases and 9 cases fall into the two ranges respectively.

4.2.1.3. Reservoir temperature. Oxidative degradation, hydrolysis and precipitation with divalent cations are some of the key factors affecting polymer stability. As the temperature increases, polymer degradation becomes more severe, especially above 158ºF. Hydrolyzed polyacrylamide (HPAM) can be reasonably sTable if there are no

dissolved oxygen or divalent cations present (Seright, 2010). For field cases as shown in Figure 4.10, the minimum value is 93.7ºF, the maximum value is 176.5ºF and the mean value is 145.7ºF. And as Figure 4.10 also shows, most of values fall into the range of 150-160ºF.

For pilot cases, Figure 4.10 shows that the minimum value is 78.89ºF, the maximum value is 200.66ºF and the mean value is 155.94ºF. Figure 4.10 also shows that most values fall into the ranges of 140-150ºF and 160-170ºF.

Combining field and pilot cases together, Figure 4.11 shows that the minimum value is 78.89ºF, the maximum value is 200.66ºF and the mean value is 151.42ºF. Figure 4.11 also shows a normal distribution and most of cases of reservoir temperature values fall into the range of 140-180ºF.

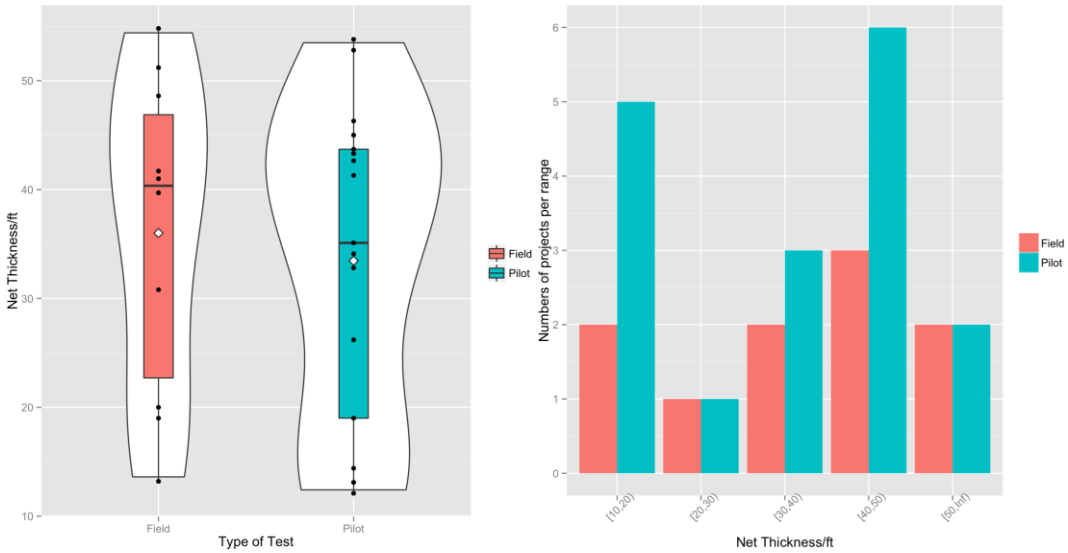

Figure 4.8. Pilot and field net thickness information (A) box and violin plot (B) histogram

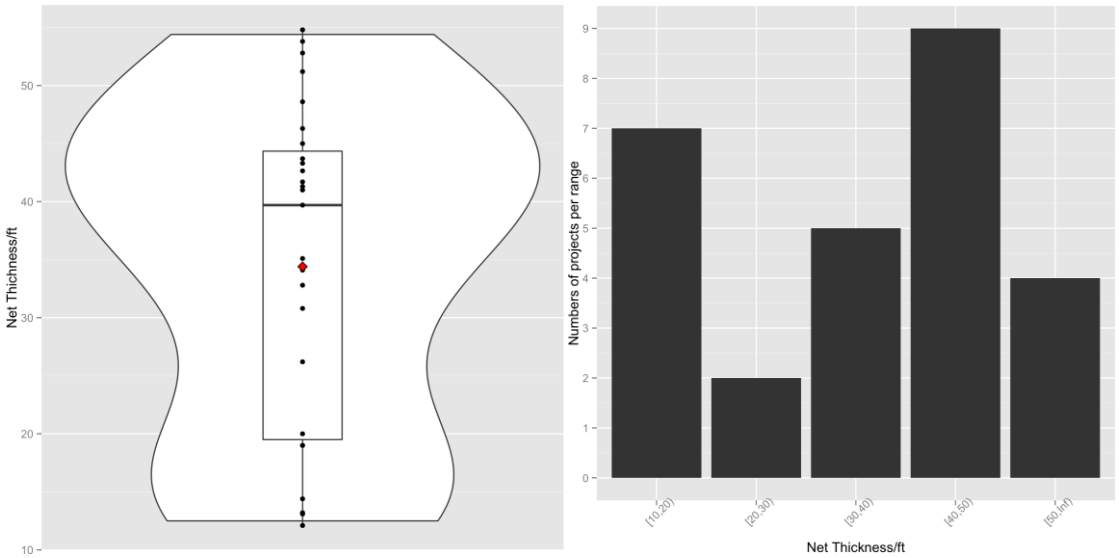

Figure 4.9. Total net thickness information (A) box and violin plot (B) histogram

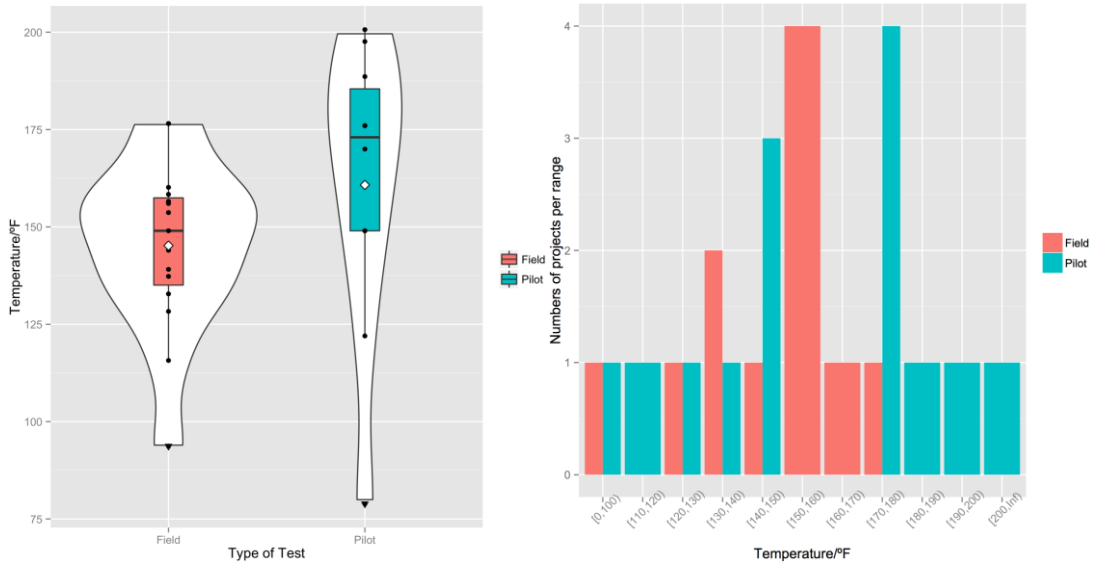

Figure 4.10. Pilot and field reservoir temperature information (A) box and violin plot (B) histogram

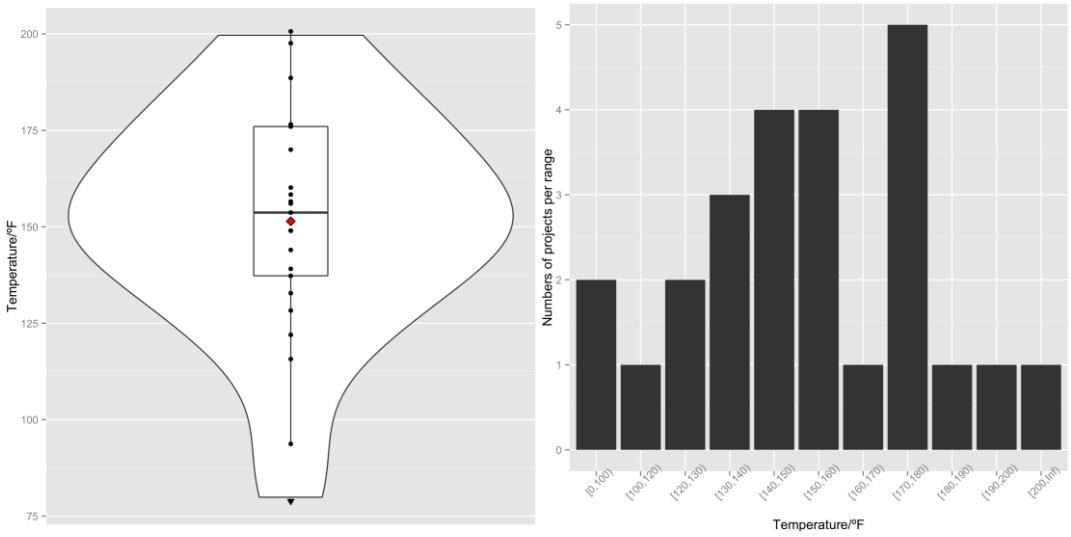

Figure 4.11. Total reservoir temperature information (A) box and violin plot (B) histogram

4.2.1.4. Porosity. Out of the 55 cases studied, 20 cases reported their porosity values. Among 46 cases there are 5 field cases and 15 pilot cases.

For field cases, as Figure 4.12 shows, the minimum value is 18.2%, the maximum value is 34.8% and the mean value is 23.47%. Most of values fall into the range of 15-20% as shown in Figure 4.12.

For pilot cases, as Figure 4.12 shows, the minimum value is 8.3%, the maximum value is 32% and the mean value is 20.79%. Most of values fall into the range of 15-30% as shown in Figure 4.12.

Combining field and pilot cases together, as shown in Figure 4.13, the minimum value is 8.3%, the maximum value is 34.8% and the mean value is 21.46%. Figure 4.13 shows a normal distribution that most of values fall into the range of 15-30%.

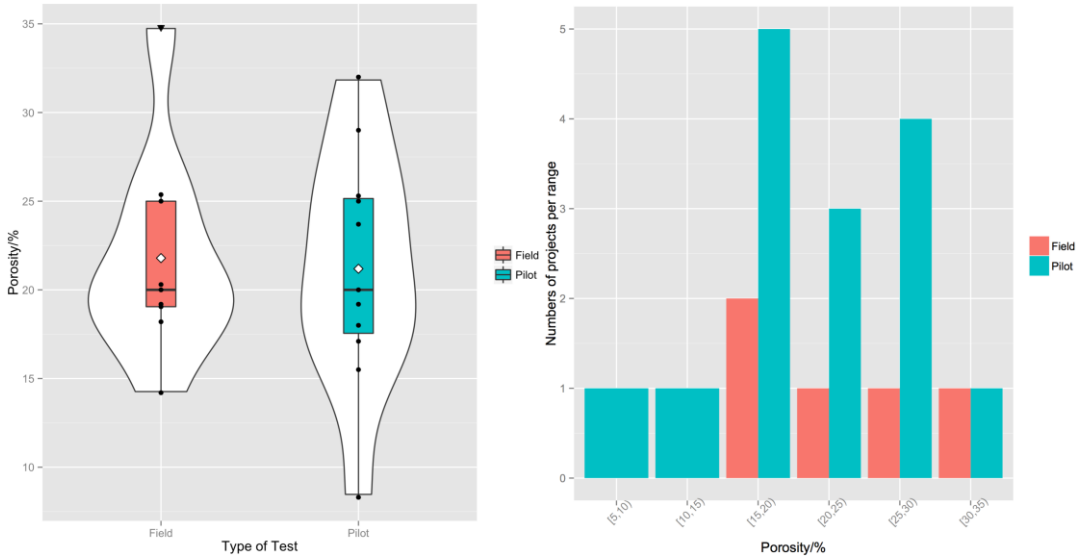

Figure 4.12. Pilot and field porosity information (A) box and violin plot (B) histogram

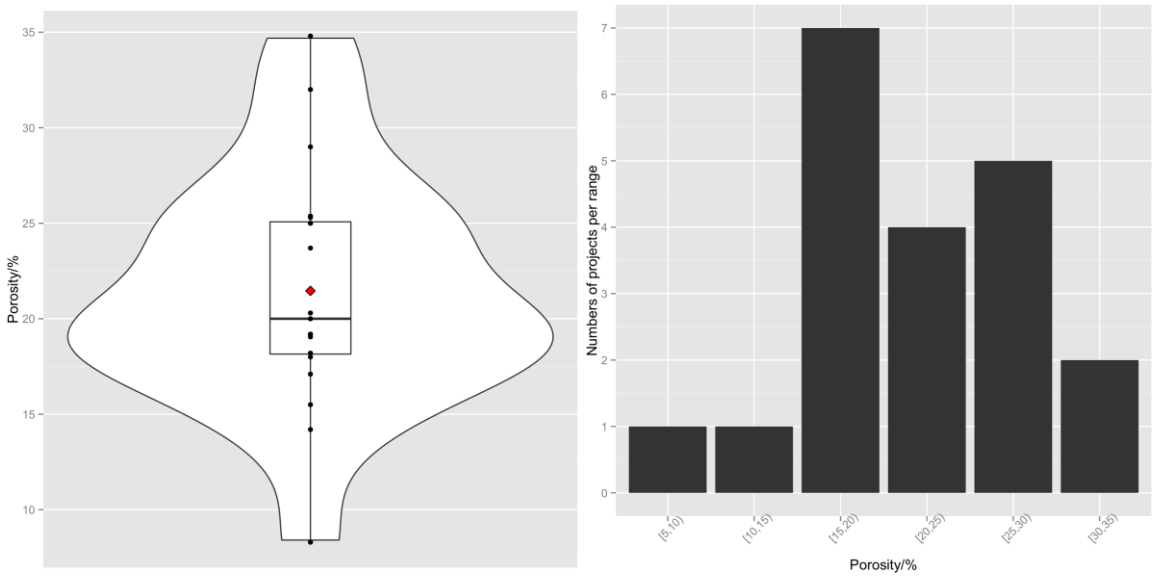

Figure 4.13. Total porosity information (A) box and violin plot (B) histogram

4.2.1.5. Average permeability. Reservoir permeability is a key factor that affects the propagation of a polymer solution.

The polymer MW affects the effectiveness of polymer flooding. That is to say, MW must be small enough so that the polymer can enter and propagate effectively through the reservoir rock. For a given rock permeability and pore throat size, a threshold MW exists, above which polymers exhibit difficulty with propagation.

Out of the 55 cases studied, 46 cases reported their reservoir average permeability values. Among 46 cases there are 28 pilot cases and 18 field cases.

For field tests, as shown in Figure 4.14, average permeability ranges from 192md to 3017md and mean value is 775.7md. The only outlier that is 3017md comes from Shengli Oil Field.

For pilot tests, as shown in Figure 4.14, average permeability ranges from 17md to 2330md and mean value is 601.5md. The two outliers that are 2039md and 2330md come from Shengli Oil Field and Henan Oil Field. Figure 4.14 also shows a multimodal distribution. For pilot tests, there are 6 cases of average permeability fall into the range of 0-1000md and others mostly fall into the range of 300-600md. For field tests, most values fall into the range of 500-700md.

Combining field and pilot tests together, most average permeability values fall into the range of 500-700md as shown in Figure 4.15.

4.2.1.6. Dykstra parsons coefficient. "Dykstra-Parsons coefficient of permeability variation" V is defined as:

$$V_k = \frac{V_{50} - V_s}{V_{50}} \tag{4}$$

where V_k is the permeability variation, k_{50} is the permeability value at the 50th percentile, and k_s is the permeability at the 84.1 percentile.

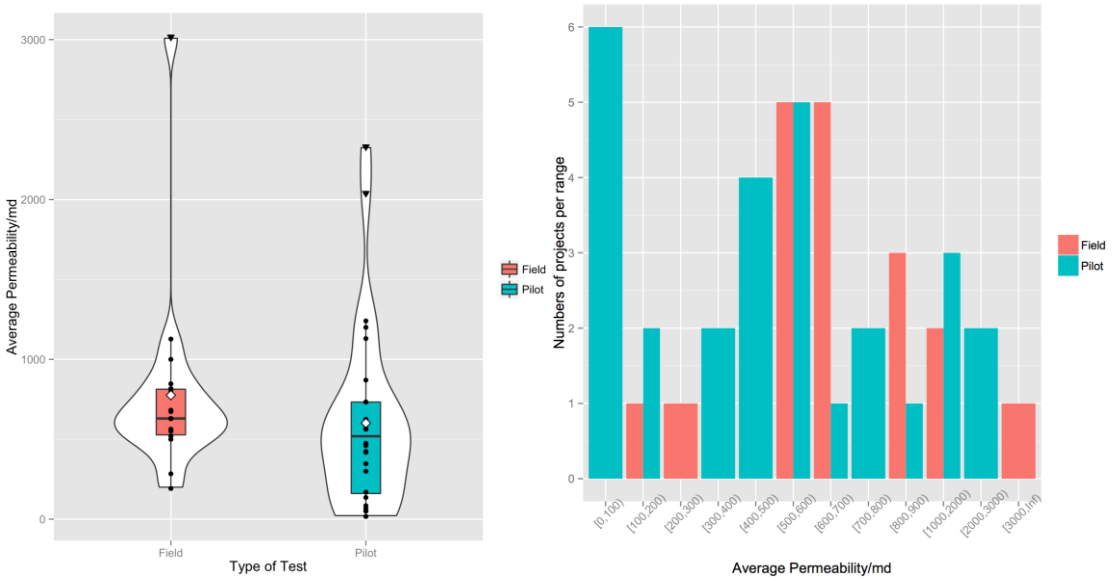

Figure 4.14. Pilot and field average permeability information (A) box and violin plot (B) histogram

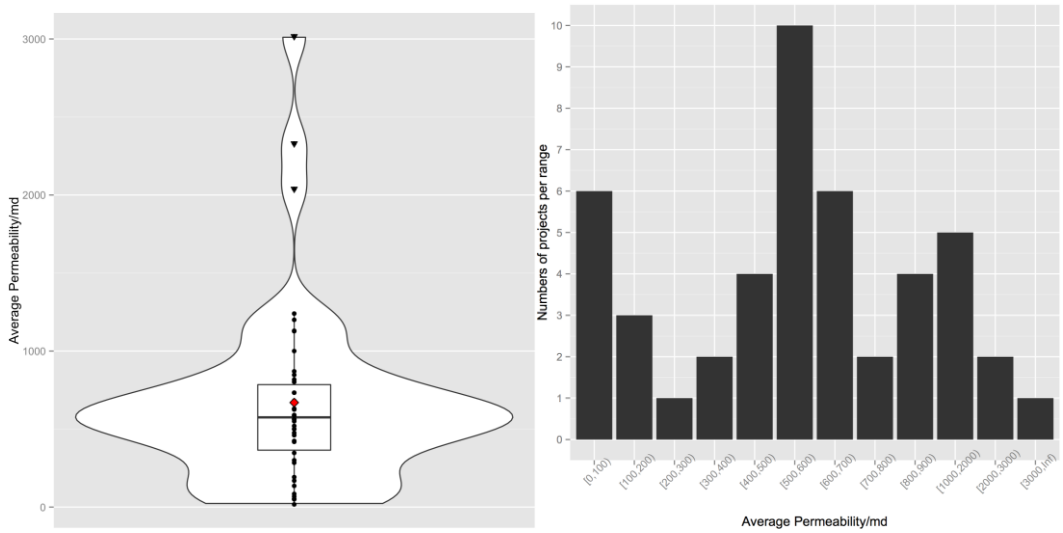

Figure 4.15. Total average permeability information (A) box and violin plot (B) histogram

The dispersion or scatter of permeability values measures reservoir heterogeneity. A homogeneous reservoir has a permeability variation that approaches zero, while an

extremely heterogeneous reservoir has a permeability variation that approaches one. Polymer floods can improve the heterogeneity between layers or within layers.

Out of the 55 cases studied, 18 cases including 6 field cases and 12 pilot cases reported their Dykstra Parsons Coefficient values.

For field test cases, as shown in Figure 4.16, the minimum value is 0.7, the maximum value is 0.87 and the mean value is 0.7717. Figure 4.16 also shows that most values fall into the range of 0.7-0.8.

For pilot cases, as shown in Figure 4.16, the minimum value is 0.4, the maximum value is 0.87 and the mean value is 0.6761. Figure 4.16 shows a nearly average distribution.

Combining field and pilot cases together, as shown in Figure 4.17, the minimum value is 0.4, the maximum value is 0.87 and the mean value is 0.7079. Figure 4.17 also shows that most values fall into the range of 0.7-0.9.

4.2.1.7. Reservoir bubble point pressure. Out of the 55 cases studied, 10 cases including 3 field cases and 7 pilot cases reported their reservoir bubble point pressure values.

For field test cases, as shown in Figure 4.18, the minimum value is 7.1Mpa, the maximum value is 14.8Mpa and the mean value is 10.27Mpa.

For pilot cases, as shown in Figure 4.18, the minimum value is 1.4Mpa, the maximum value is 14.95Mpa and the mean value is 8.55Mpa. Figure 4.18 also shows a peak distribution that most values fall into the range of 10-15Mpa.

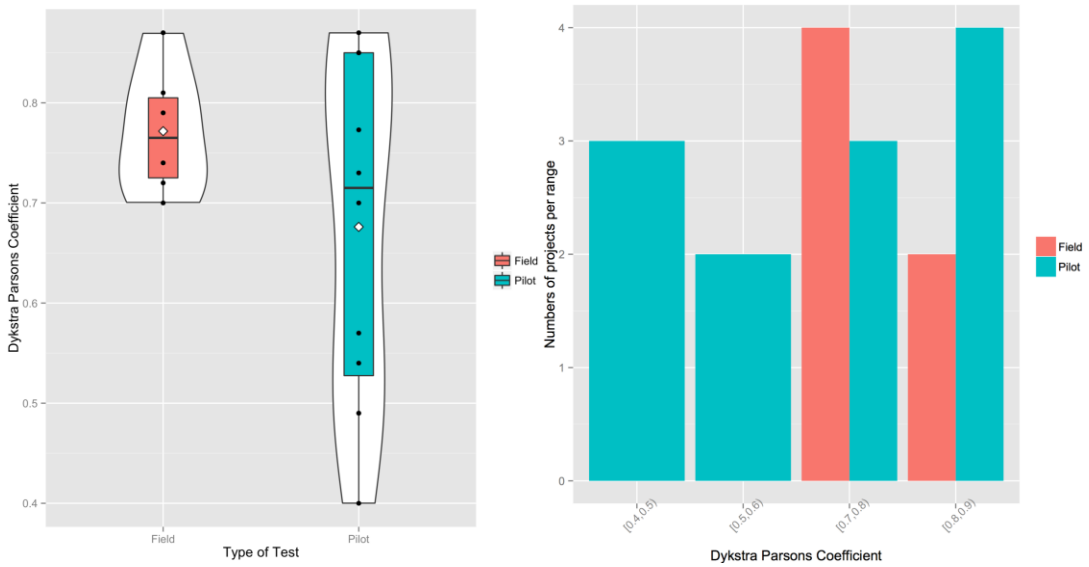

Figure 4.16. Pilot and field dykstra parsons coefficient information (A) box and violin plot (B) histogram

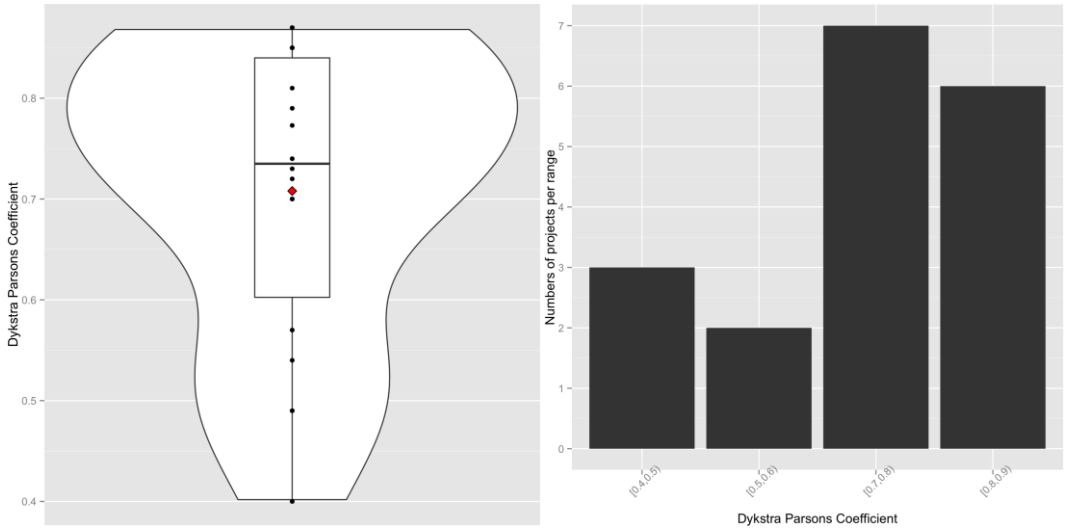

Figure 4.17. Total dykstra parsons coefficient information (A) box and violin plot (B) histogram

Combining field and pilot cases together, as shown in Figure 4.19, the minimum value is 1.4Mpa, the maximum value is 14.95Mpa and the mean value is 8.55Mpa. Figure 4.19 also shows that most values fall into the range of 10-15Mpa.

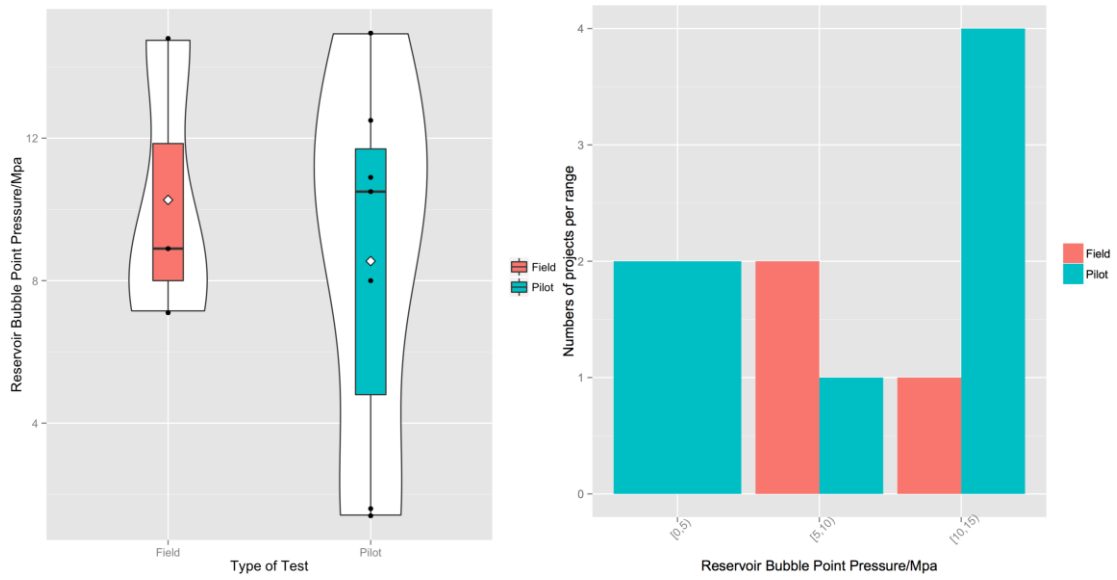

Figure 4.18. Pilot and field reservoir bubble point pressure information (A) box and violin plot (B) histogram

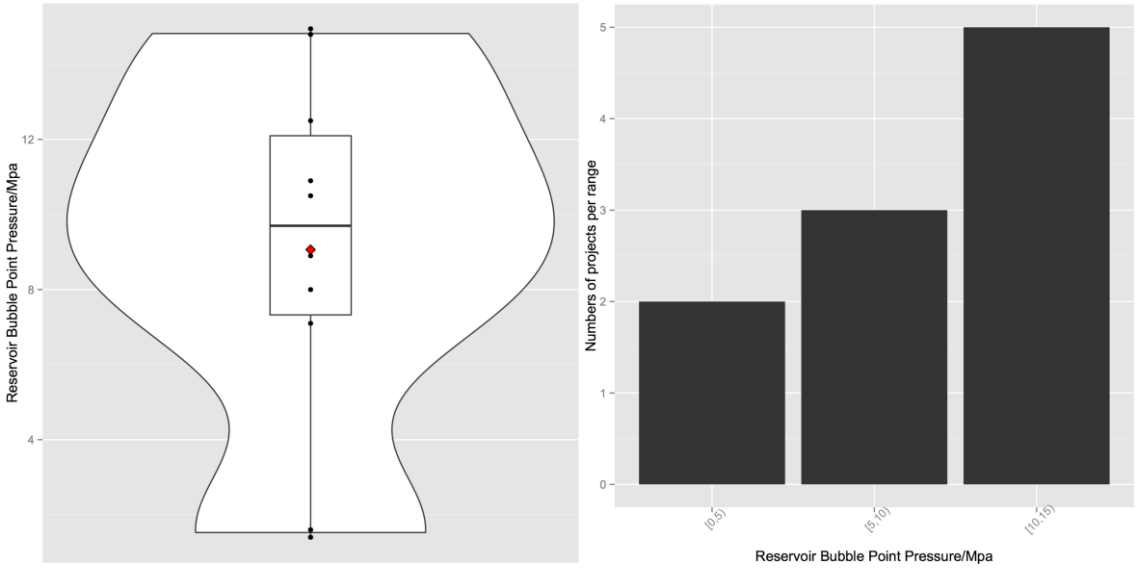

Figure 4.19. Total reservoir bubble point pressure information (A) box and violin plot (B) histogram

4.2.1.8. Original formation pressure. Out of the 55 cases studied, 10 cases including 4 field cases and 6 pilot cases reported their original formation pressure values.

For field test cases, as shown in Figure 4.20, the minimum value is 7.19Mpa, the maximum value is 17.1Mpa and the mean value is 12.12Mpa.

For pilot cases, as shown in Figure 4.20, the minimum value is 4.914Mpa, the maximum value is 20.2Mpa and the mean value is 13.476Mpa.

Combining field and pilot cases together, as shown in Figure 4.21, the minimum value is 4.914Mpa, the maximum value is 20.2Mpa and the mean value is 12.932Mpa. Figure 4.21 also shows that most values fall into the range of 10-20Mpa.

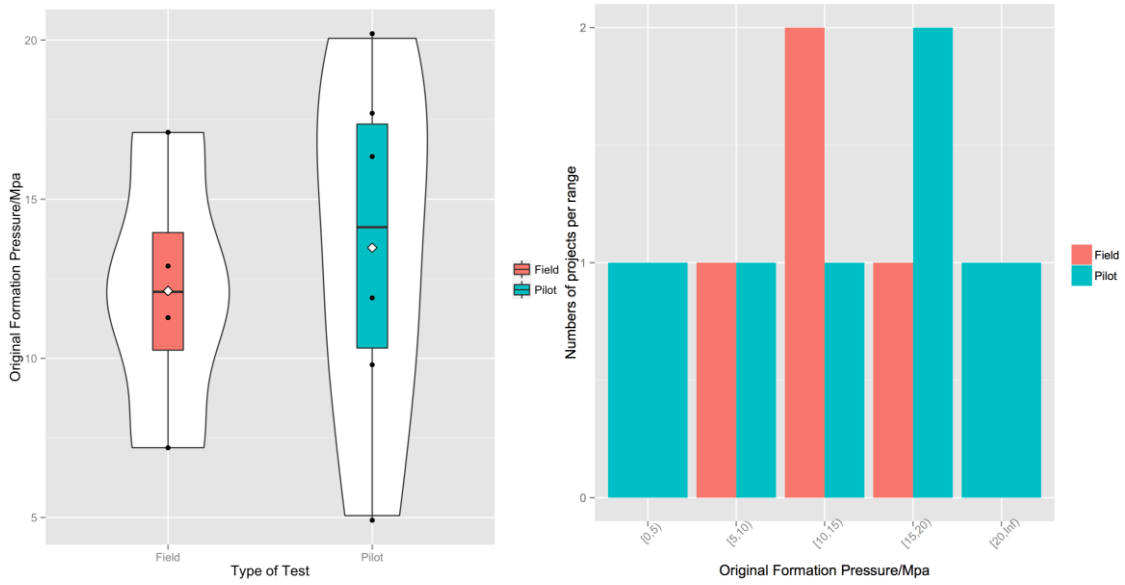

Figure 4.20. Pilot and field original formation pressure information (A) box and violin plot (B) histogram

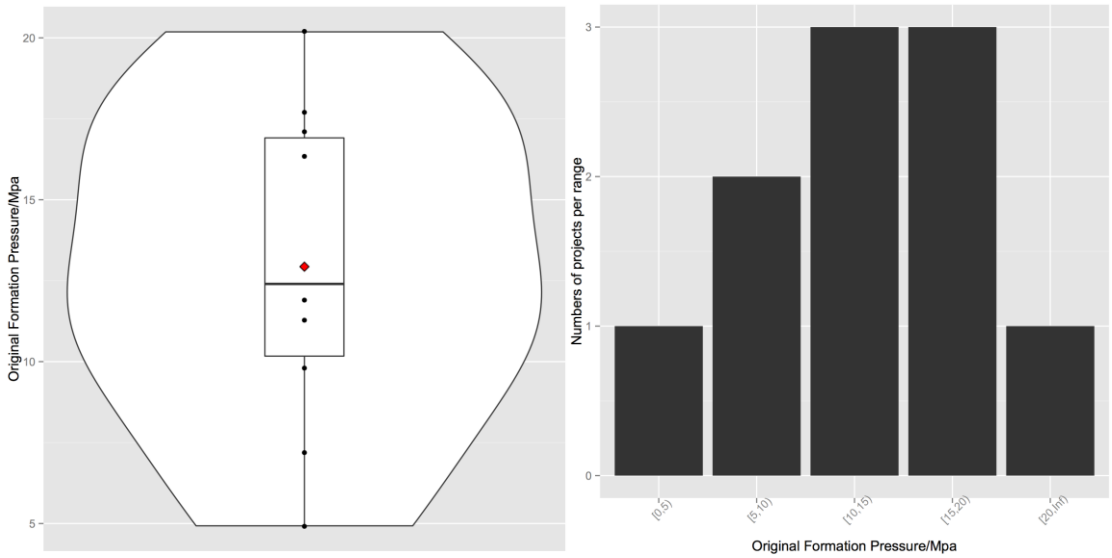

Figure 4.21. Total original formation pressure information (A) box and violin plot (B) histogram

4.2.1.9. Reservoir present pressure. Out of the 55 cases studied, 10 cases including 3 field cases and 7 pilot cases reported their original formation pressure values.

For field test cases, as shown in Figure 4.22, the minimum value is 9.8Mpa, the maximum value is 12Mpa and the mean value is 11.13Mpa.

For pilot cases, as shown in Figure 4.22, the minimum value is 10Mpa, the maximum value is 14.8Mpa and the mean value is 12.44Mpa. Figure 4.22 also shows that all of values fall into the range of 10-15Mpa.

Combining field and pilot cases together, as shown in Figure 4.23, the minimum value is 9.8Mpa, the maximum value is 14.8Mpa and the mean value is 12.05Mpa. Figure 4.23 also shows that most values fall into the range of 10-15Mpa.

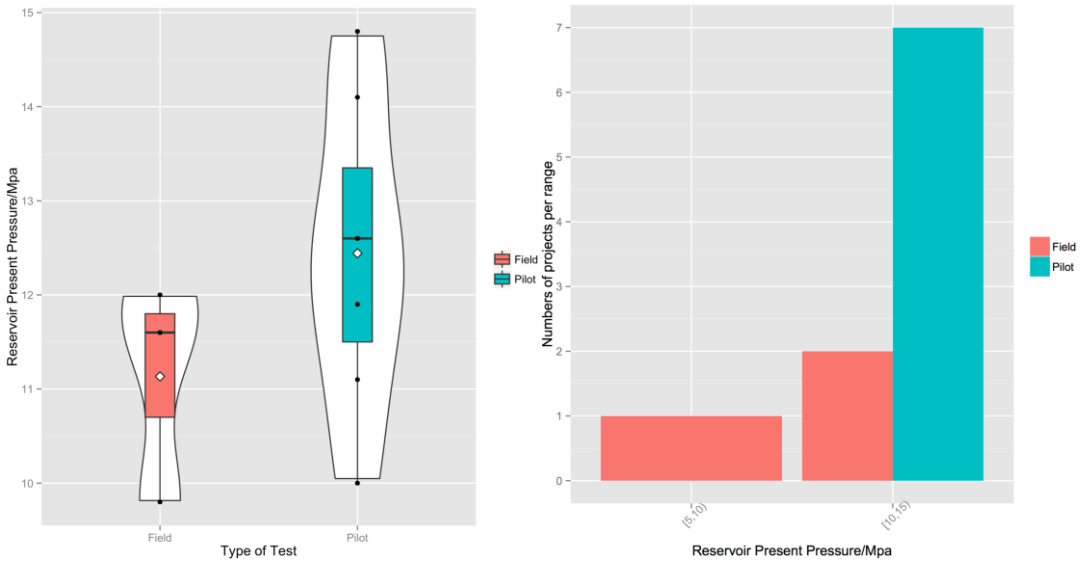

Figure 4.22. Pilot and field reservoir present pressure information (A) box and violin plot (B) histogram

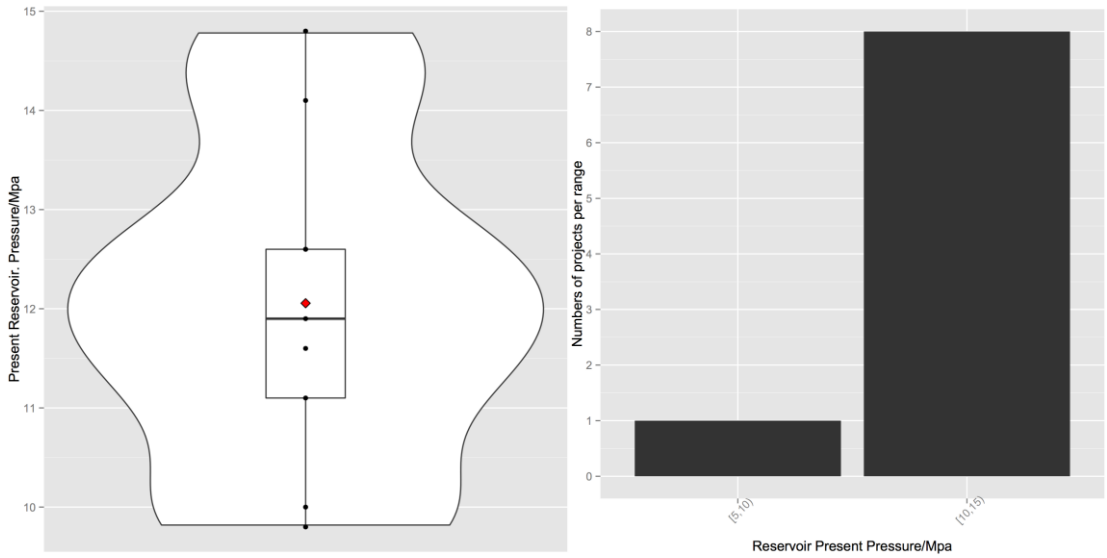

Figure 4.23. Total reservoir present pressure information (A) box and violin plot (B) histogram

4.2.1.10. Oil viscosity. Out of the 55 cases studied, 25 cases including 11 field cases and 14 pilot cases reported their oil viscosity values.

For field test cases, as shown in Figure 4.24, the minimum value is 2.6cP, the maximum value is 76.96cP and the mean value is 29.31cP. Figure 4.24 also shows that most of values fall into the range of 0-10cP.

For pilot cases, as shown in Figure 4.24, the minimum value is 2.3cP, the maximum value is 285.2cP and the mean value is 61.987cP. Figure 4.24 also shows that most of values fall into the range of 0-10cP.

Combining field and pilot cases together, as shown in Figure 4.25, the minimum value is 2.3cP, the maximum value is 285.2cP and the mean value is 47.61cP. Figure 4.25 also shows that most values fall into the range of 0-10cP.

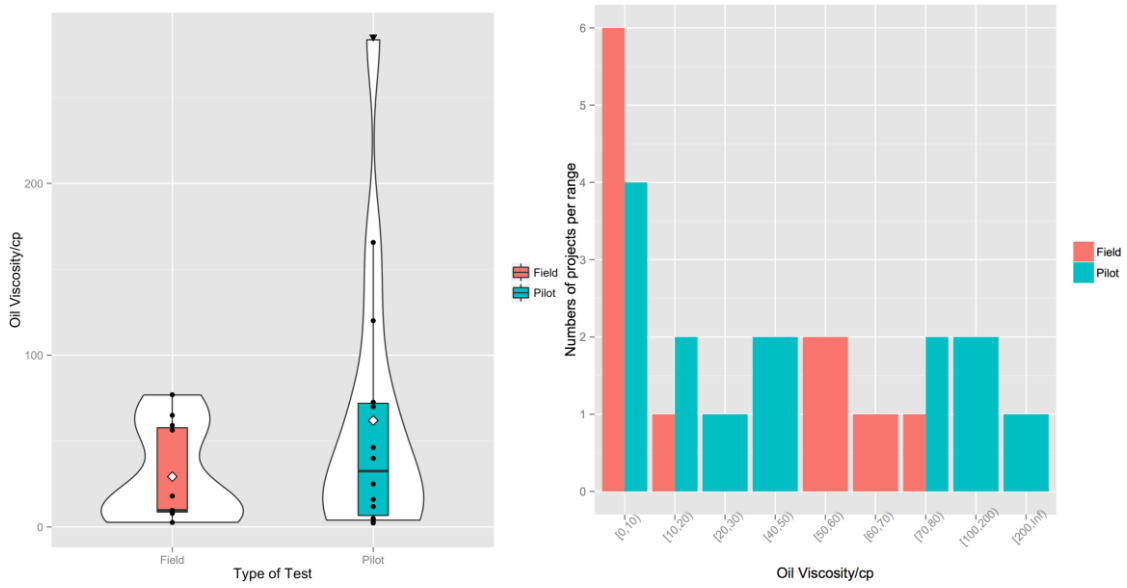

Figure 4.24. Pilot and field oil viscosity information (A) box and violin plot (B) histogram

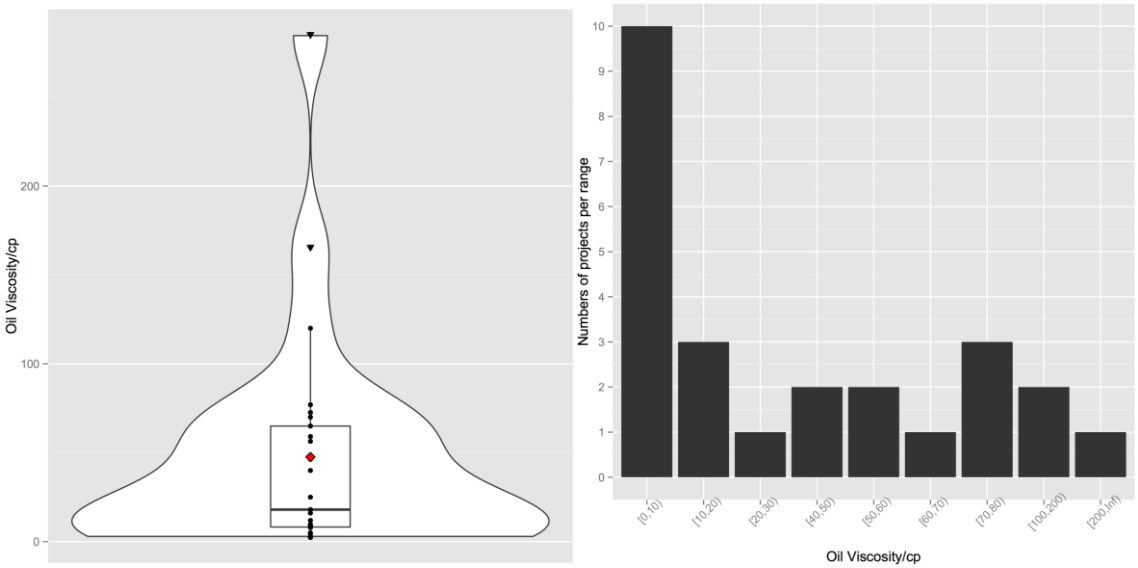

Figure 4.25. Total oil viscosity information (A) box and violin plot (B) histogram

4.2.1.11. Oil gravity. Out of the 55 cases studied, 15 cases including 9 field cases and 6 pilot cases reported their oil gravity values.

For field test cases, as shown in Figure 4.26, the minimum value is 16.2° API, the maximum value is 53.2° API and the mean value is 36.89° API. Figure 4.26 also shows that most of values fall into the range of 40-50° API.

For pilot cases, as shown in Figure 4.26, the minimum value is 14.96° API, the maximum value is 51.1° API and the mean value is 24.81° API. Figure 4.26 also shows that most of values fall into the range of 10-20° API.

Combining field and pilot cases together, as shown in Figure 4.27, the minimum value is 14.96° API, the maximum value is 53.2° API and the mean value is 32.06° API. Figure 4.27 also shows that most values fall into the range of 10-50° API.

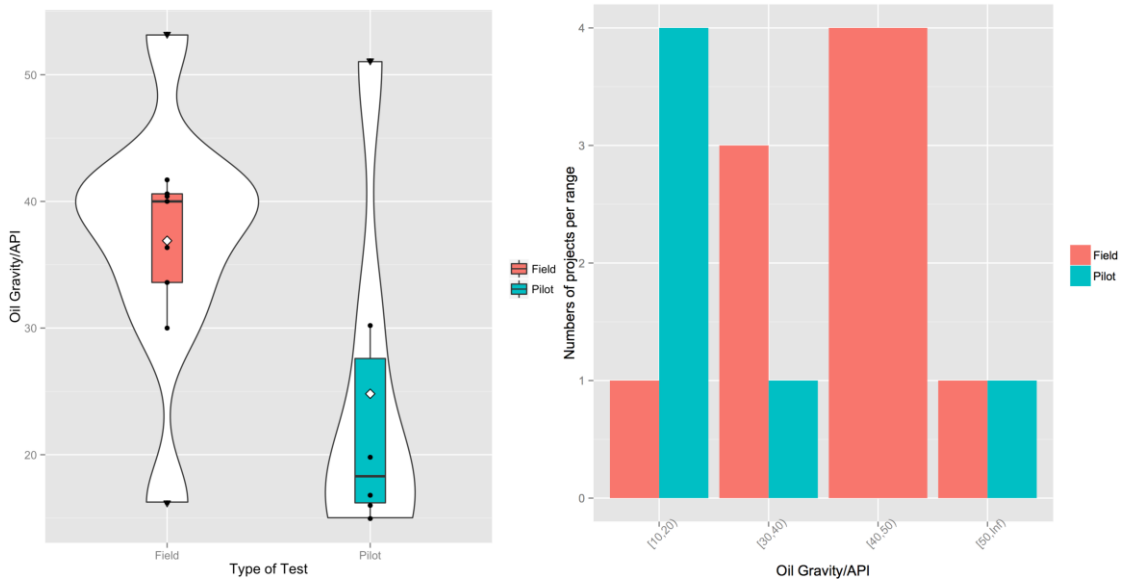

Figure 4.26. Pilot and field oil gravity information (A) box and violin plot (B) histogram

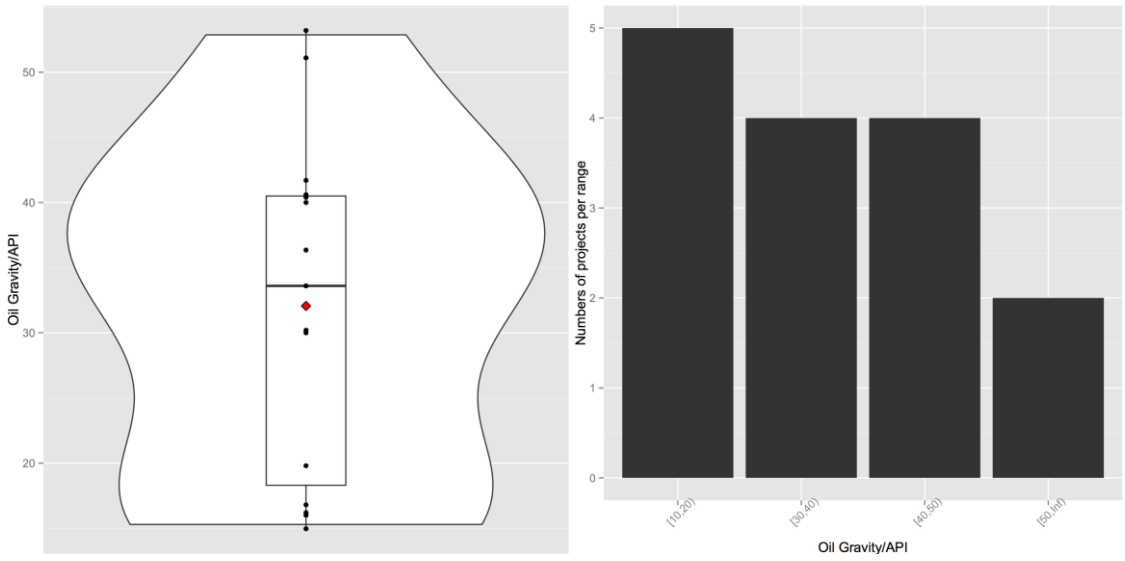

Figure 4.27. Total oil gravity information (A) box and violin plot (B) histogram

4.2.1.12. Water salinity. Formation water salinity has a strong effect on polymer viscosity, especially for HPAM. Polymer solution viscosity decreases with salinity. Polymer viscosity is sensitive to the cation content of water solution: Ca^{2+}, Mg^{2+}, Fe^{3+},

etc., far more than K^+, Na^+. High divalent or trivalent content in the formation water may cause polymer participation. Lower polymer viscosity will lead to poor mobility control by polymer processes.

Out of the 55 cases studied, 25 cases including 10 field cases and 15 pilot cases reported their water salinity values.

For field test cases, as shown in Figure 4.28, the minimum value is 3580ppm, the maximum value is 28868ppm and the mean value is 8087ppm. Figure 4.28 also shows that most of values fall into the range of 6000-8000ppm.

For pilot cases, as shown in Figure 4.28, the minimum value is 2127ppm, the maximum value is 84128ppm and the mean value is 23135ppm. Figure 4.28 also shows that most of values are above 40000ppm.

Combining field and pilot cases together, as shown in Figure 4.29, the minimum value is 2127ppm, the maximum value is 84128ppm and the mean value is 17116ppm. Figure 4.29 also shows a multimodal distribution that one peak range is 4000-8000ppm and the other peak range is above 20000ppm.

4.2.2. Polymer Properties

4.2.2.1. Polymer molecular weight. Molecular weight is a key factor that affects polymer flooding effectiveness. Polymers with higher molecular weight can provide greater viscosity and thus leads to high oil recovery. The reason is simply that for a given solution viscosity and sweep efficiency increase with increased polymer MW. In other words, to recover a given volume of oil, less polymer is needed using a high MW polymer than a low MW polymer.

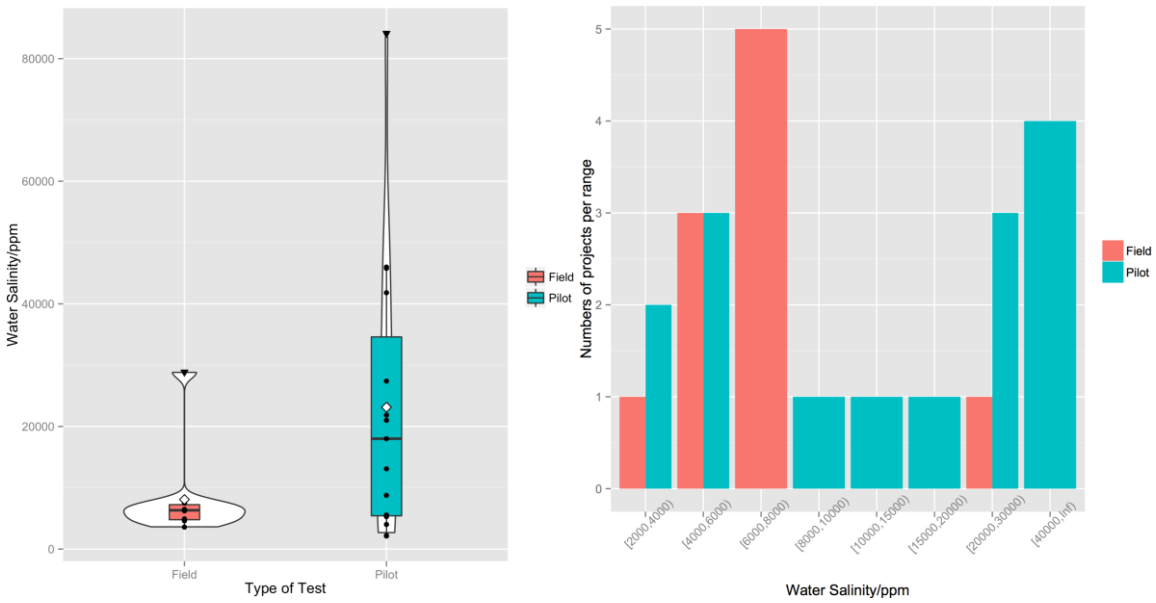

Figure 4.28. Pilot and field water salinity information (A) box and violin plot (B) histogram

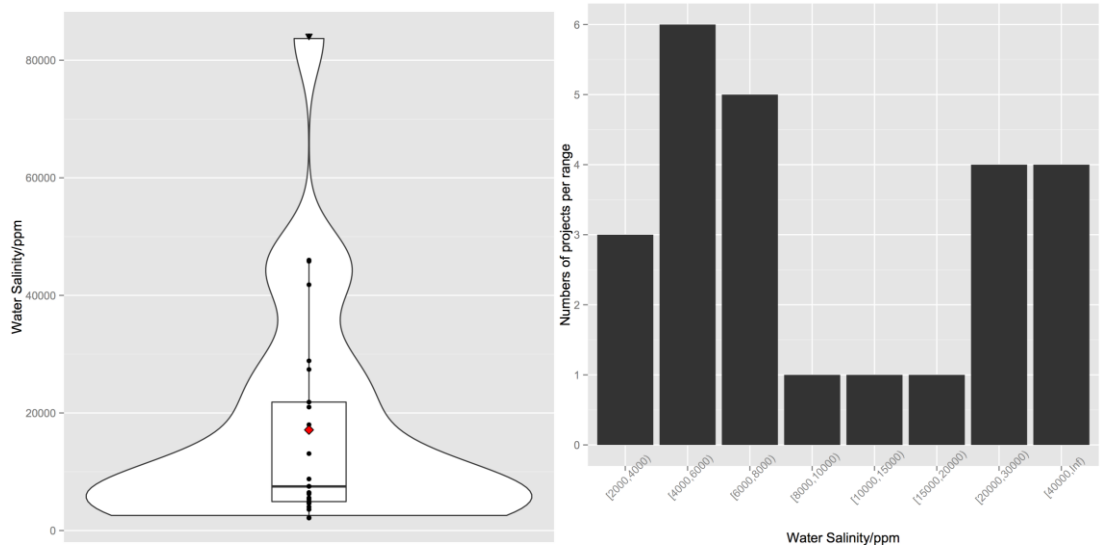

Figure 4.29. Total water salinity information (A) box and violin plot (B) histogram

Two factors should be considered when choosing the polymer MW. On one hand, choose the polymer with the highest MW practical to minimize the polymer cost. On the

other hand, the MW must be small enough so that the polymer can enter and propagate effectively through the reservoir rock. For a given rock permeability and pore throat size, a threshold MW exists, above which polymers exhibit difficulty in propagation. Mechanical entrapment can significantly retard polymer propagation if the pore throat size and permeability are too small. Thus, depending on MW and permeability differential, this effect can reduce sweep efficiency. A trade-off must be made in choosing the highest MW polymer that will not exhibit pore plugging or significant mechanical entrapment in the less permeable zones.

Figure 4.30 shows the relationship between reservoir average permeability and polymer molecular weight. It also indicates that a medium polymer weight (12-16 million daltons) is applicable for oil zones with average permeability greater than 100md and a high molecular weight (17-25 million daltons) is appropriate for oil zones with the average permeability greater than 500md.

Out of the 55 cases studied, 20 cases including 6 field cases and 14 pilot cases reported their polymer molecular values.

For field test cases, as shown in Figure 4.31, the minimum value is 1075(10^4), the maximum value is 2750(10^4) and the mean value is 1896(10^4).

For pilot cases, as shown in Figure 4.31, the minimum value is 600(10^4), the maximum value is 2500(10^4), and the mean value is 1436(10^4). Figure 4.31 also shows that most of values are above 1200-2000(10^4).

Combining field and pilot cases together, as shown in Figure 4.32, the minimum value is 600(10^4), the maximum value is 2750(10^4), and the mean value is 1574(10^4). Figure 4.32 also shows a peak range that is 1600-2000(10^4).

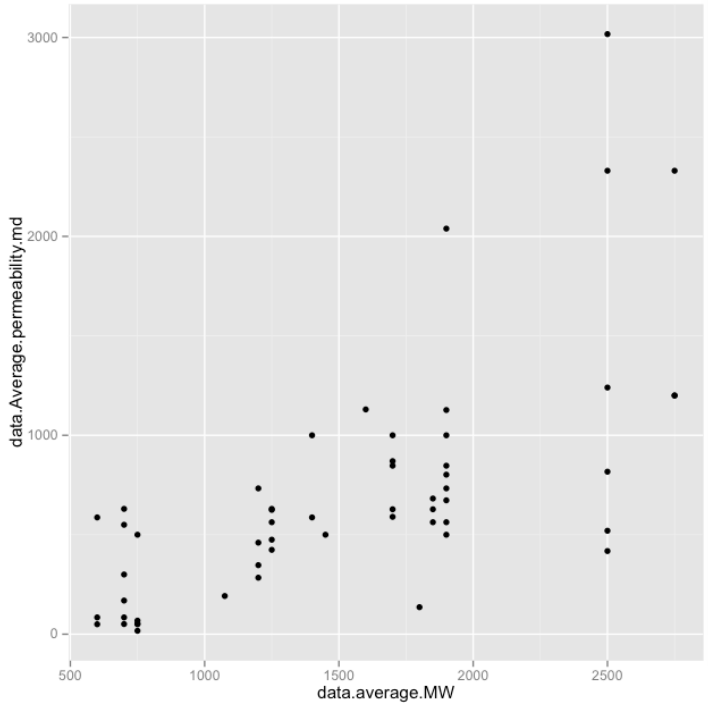

Figure 4.30. Relationship between reservoir average permeability and polymer molecular weight

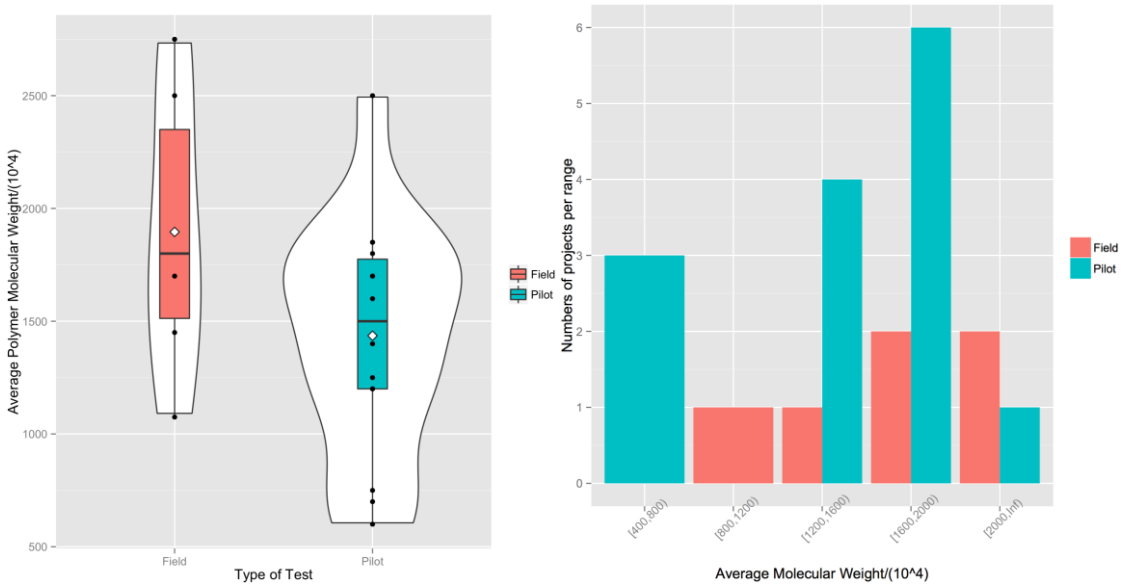

Figure 4.31. Pilot and field average polymer molecular weight information (A) box and violin plot (B) histogram

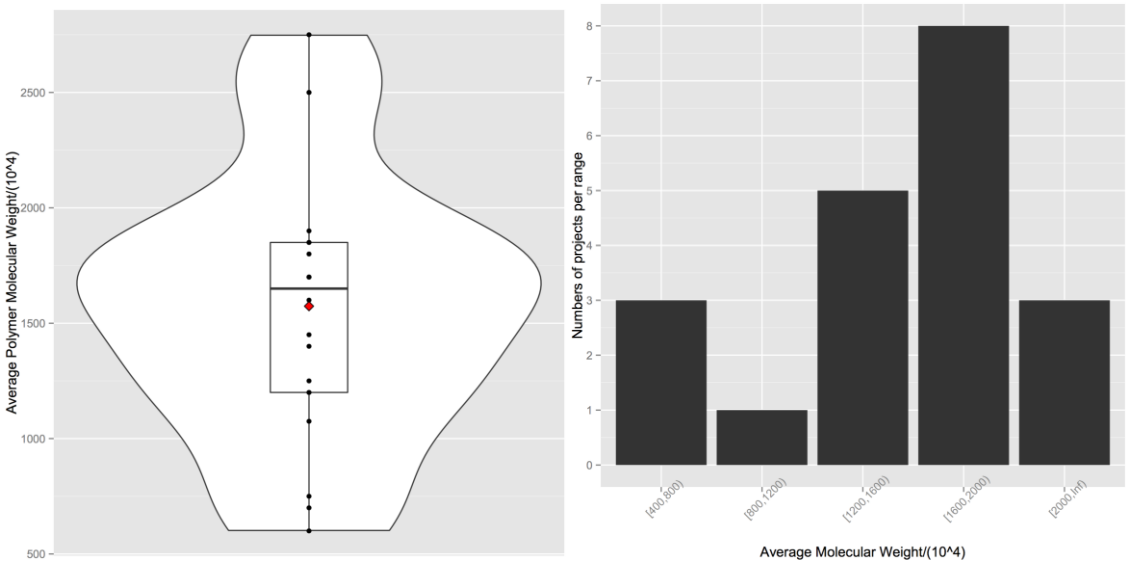

Figure 4.32. Total average molecular weight information (A) box and violin plot (B) histogram

4.2.2.2. Polymer concentration. Polymer concentration is a key factor that higher polymer concentrations can cause greater reductions in water cut and can shorten the time required for polymer flooding. High polymer concentration can also lead to a faster decrease in water cut, an earlier response in the production wells, a greater decrease in water cut, less required pore volumes of polymer, and less required volume of water injected during the overall period of polymer flooding. As polymer concentration increases, EOR increases and the water cut during polymer flooding decreases. However, higher concentrations will cause higher injection pressures and lower injectivity.

Out of the 55 cases studied, 44 cases including 17 field cases and 27 pilot cases reported their polymer concentration values.

For field test cases, as shown in Figure 4.33, the minimum value is 650ppm, the maximum value is 2050ppm and the mean value is 1250ppm. Figure 4.33 also shows that most values fall into the range of 1000-1400ppm.

For pilot cases, as shown in Figure 4.33, the minimum value is 600ppm, the maximum value is 2000ppm, and the mean value is 1325ppm. Figure 4.33 also shows a multimodal distribution that one peak range is 0-1400ppm and the other peak range is 1600-1800ppm.

Combining field and pilot cases together, as shown in Figure 4.34, the minimum value is 600ppm, the maximum value is 2050ppm and the mean value is 1296ppm. Figure 4.34 also shows a multimodal distribution that one peak range is 1000-1200ppm and the other peak range is 1600-1800ppm.

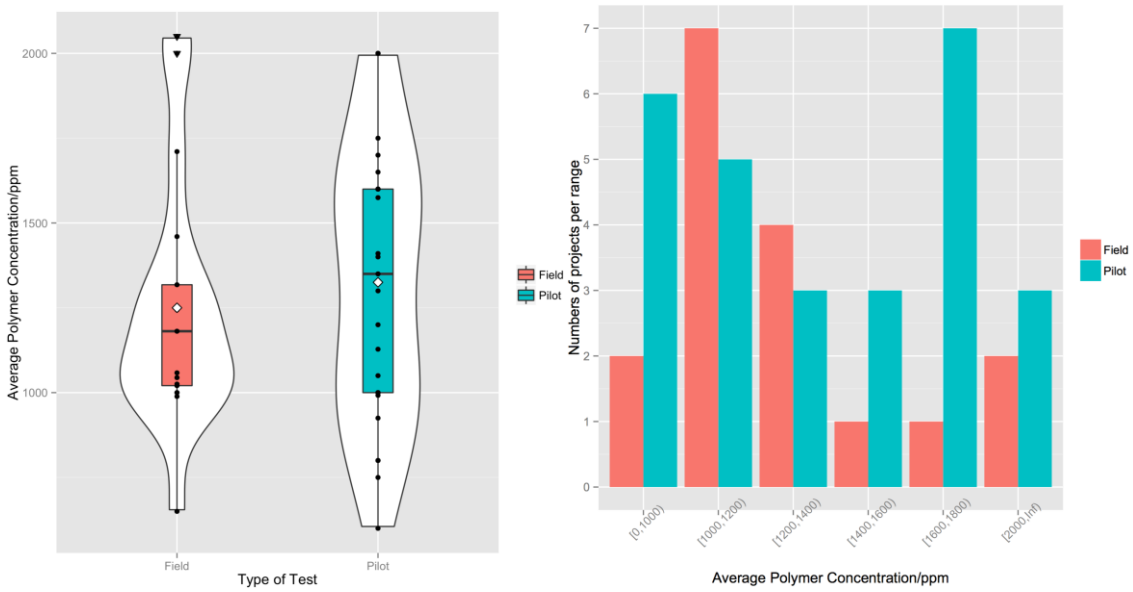

Figure 4.33. Pilot and field average polymer concentration information (A) box and violin plot (B) histogram

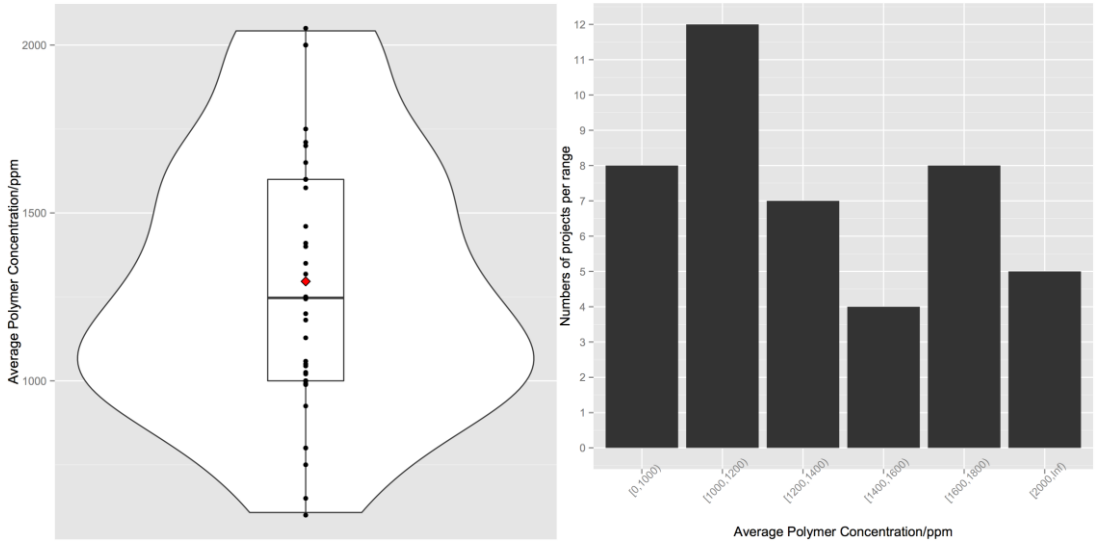

Figure 4.34. Total average polymer concentration information (A) box and violin plot (B) histogram

4.2.2.3. Polymer viscosity. The polymer solution viscosity is a key parameter to improve the mobility ratio between oil and water. As injection viscosity increases, the effective-ness of polymer flooding increases. A number of factors such as polymer MW, polymer concentration, and degree of HPAM hydrolysis, temperature, salinity, and hardness can affect the viscosity. The effectiveness of a polymer flood is directly determined by the magnitude of the polymer viscosity.

Out of the 55 cases studied, 10 cases including 5 field cases and 5 pilot cases reported their polymer viscosity values.

For field test cases, as shown in Figure 4.35, the minimum value is 17.75cP, the maximum value is 220.45cP and the mean value is 87.01cP.

For pilot cases, as shown in Figure 4.35, the minimum value is 15cP, the maximum value is 91.1cP, and the mean value is 43.82cP.

Combining field and pilot cases together, as shown in Figure 4.36, the minimum value is 15cP, the maximum value is 220.45cP and the mean value is 65.42cP. Figure 4.36 also shows that most values fall into the range of 20-40cP.

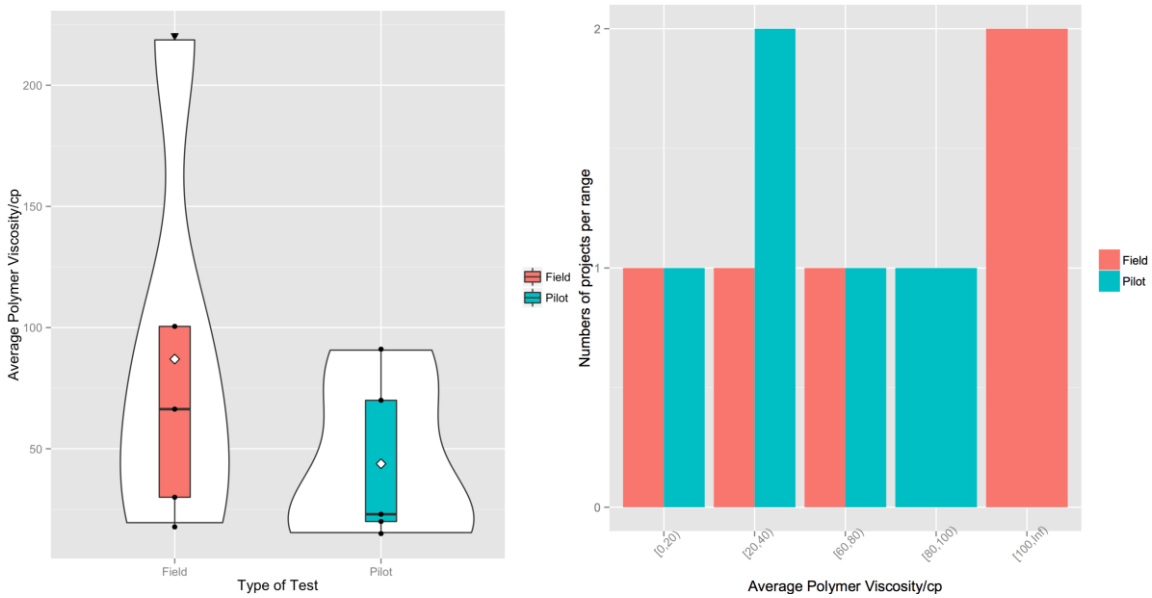

Figure 4.35. Pilot and field average polymer viscosity information (A) box and violin plot (B) histogram

4.2.3. Well Information

4.2.3.1. Injection well numbers. Out of the 55 cases studied, 24 cases including 8 field cases and 16 pilot cases reported their injection well numbers values.

For field test cases, as shown in Figure 4.37, the minimum injection well numbers are 9 wells, the maximum well numbers are 99 wells and the mean well numbers are 43 wells. Figure 4.37 also shows that most cases have above 50 injection wells.

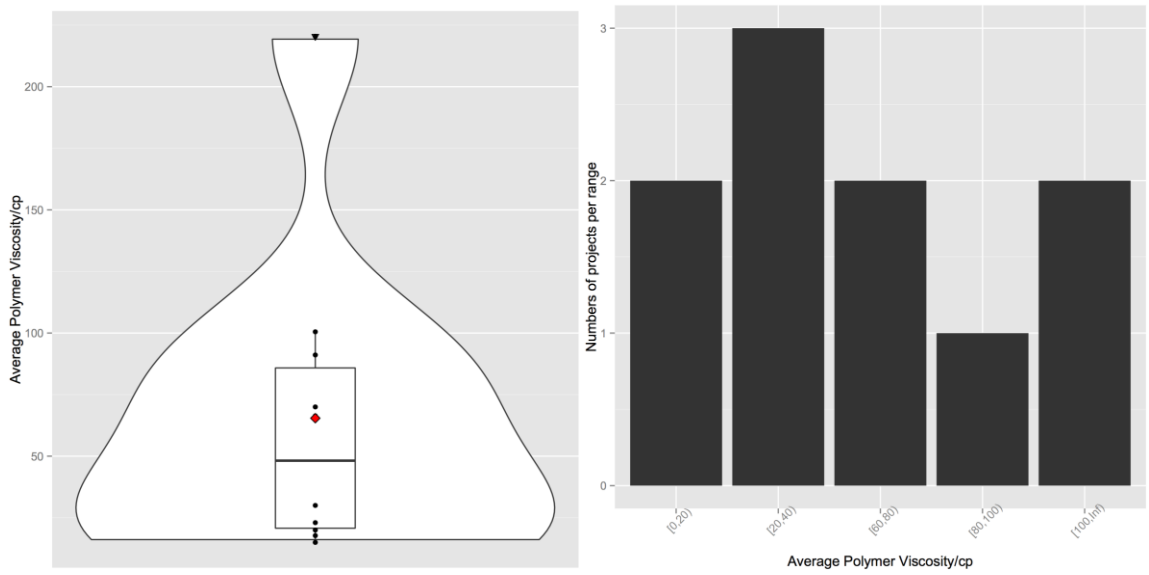

Figure 4.36. Total average polymer viscosity information (A) box and violin plot (B) histogram

For pilot cases, as shown in Figure 4.37, the minimum well numbers are 4 wells, the maximum well numbers are 49 wells, and the mean well numbers are 17 wells. Figure 4.37 also shows that most cases have no more than 20 injection wells.

Combining field and pilot cases together, as shown in Figure 4.38, the minimum injection well numbers are 4 wells, the maximum well numbers are 99 wells and the mean well numbers are 25 wells. Figure 4.38 also shows that both field and pilot cases have most injection well numbers of no more than 20 wells.

4.2.3.2. Production well numbers. Out of the 55 cases studied, 25 cases including 8 field cases and 17 pilot cases reported their production well numbers values.

For field test cases, as shown in Figure 4.39, the minimum production well numbers are 16 wells, the maximum well numbers are 109 wells and the mean well numbers are 64 wells. Figure 4.39 also shows that most cases have above 50 production wells.

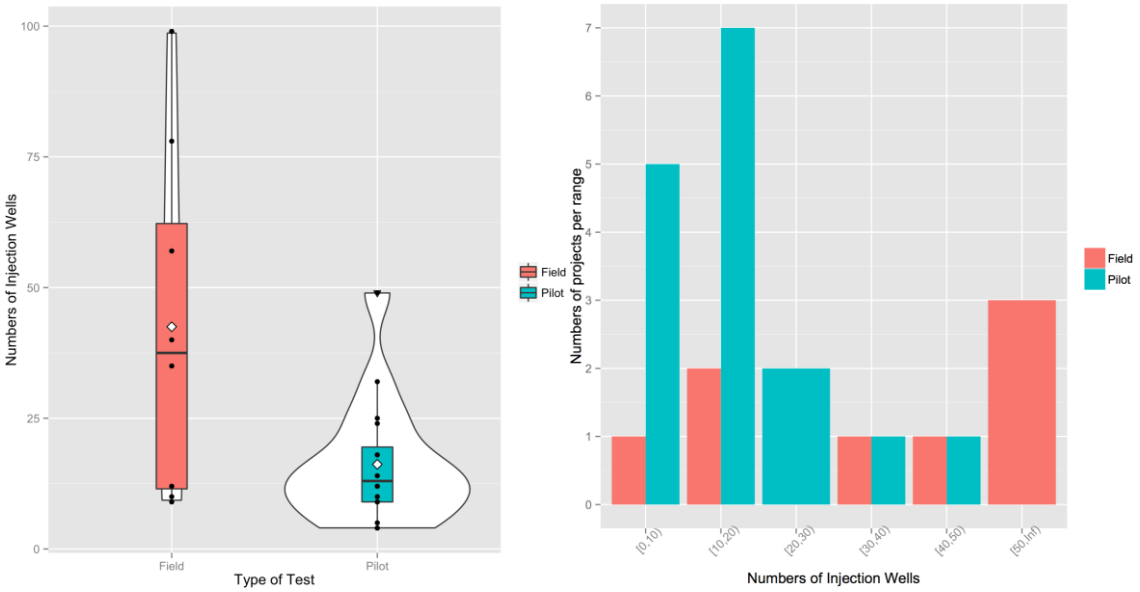

Figure 4.37. Pilot and field injection well numbers information (A) box and violin plot (B) histogram

For pilot cases, as shown in Figure 4.39, the minimum well numbers are 9 wells, the maximum well numbers are 63 wells, and the mean well numbers are 25 wells. Figure 4.39 also shows that most cases have 10-20 production wells.

Combining field and pilot cases together, as shown in Figure 4.40, the minimum production well numbers are 9 wells, the maximum well numbers are 109 wells and the mean well numbers are 38 wells. Figure 4.40 also shows a multimodal distribution that one peak range of is 10-30 wells and the other peak range is above 50 wells.

4.2.3.3. Injecting pressure. Out of the 55 cases studied, 12 cases including4 field cases and 8 pilot cases reported their injecting pressure values.

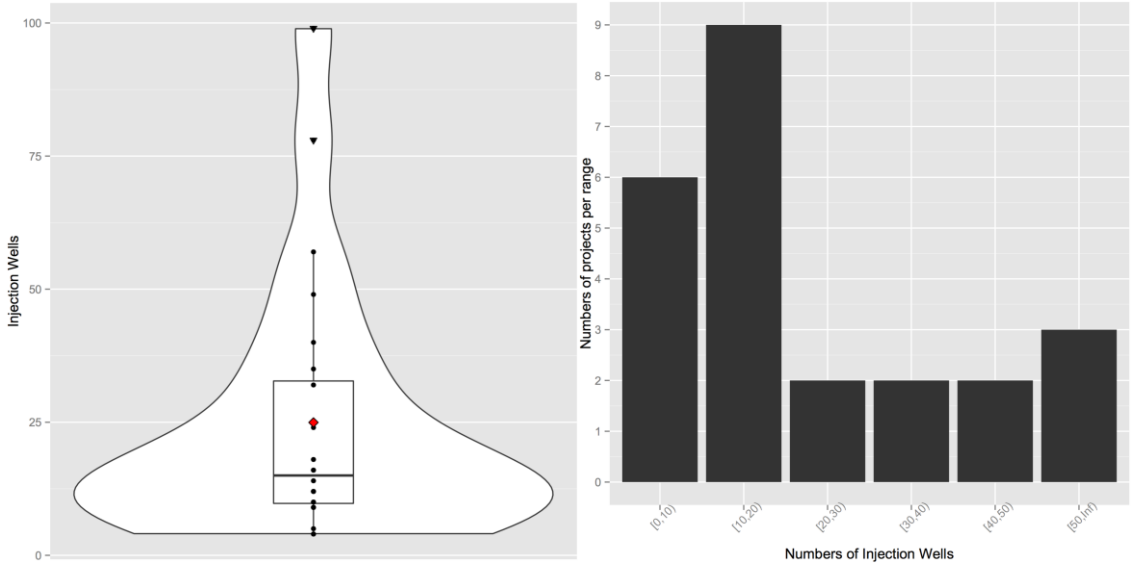

Figure 4.38. Total injection well numbers information (A) box and violin plot (B) histogram

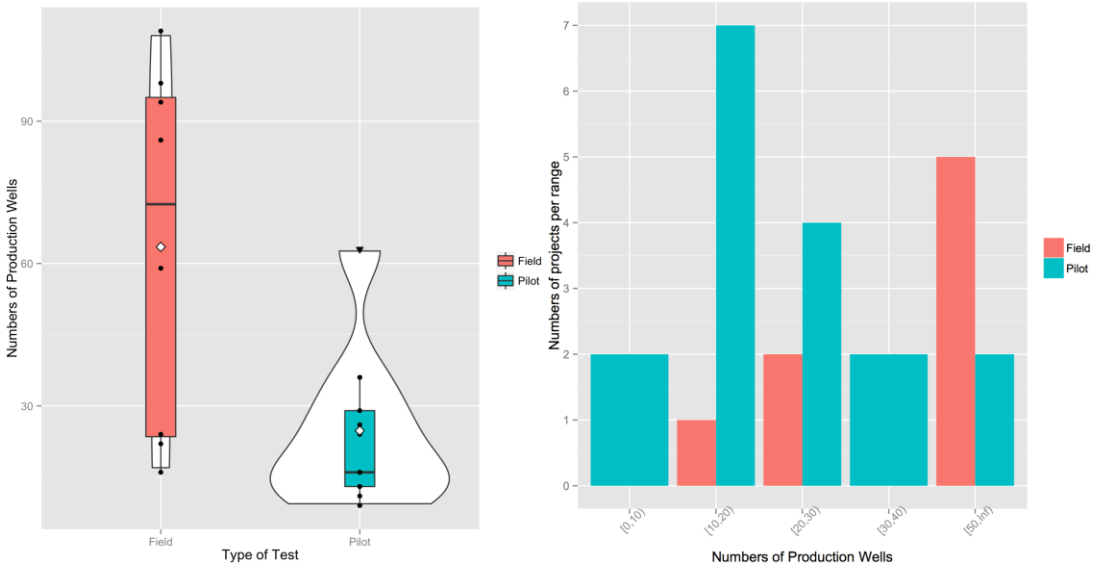

Figure 4.39. Pilot and field production well numbers information (A) box and violin plot (B) histogram

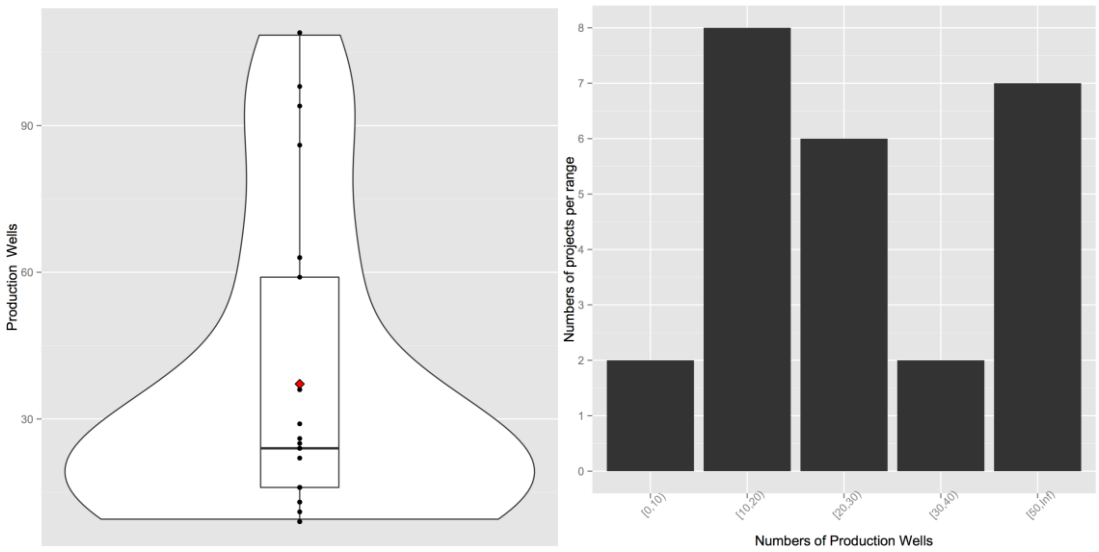

Figure 4.40. Total production well numbers information (A) box and violin plot (B) histogram

For field test cases, as shown in Figure 4.41, the minimum value is 12Mpa, the maximum value is 12.3Mpa and the mean value is 12.15Mpa. Figure 4.41 also shows that most values fall into the range of 10-15Mpa.

For pilot cases, as shown in Figure 4.41, the minimum value is 10Mpa, the maximum value is 20.5Mpa, and the mean value is 13.84Mpa. Figure 4.41 also shows that most values fall into the range of 10-15Mpa.

Combining field and pilot cases together, as shown in Figure 4.42, the minimum value is 10Mpa, the maximum value is 20.5Mpa and the mean value is 13.28Mpa. Figure 4.42 also shows that most values fall into the range of 10-15Mpa.

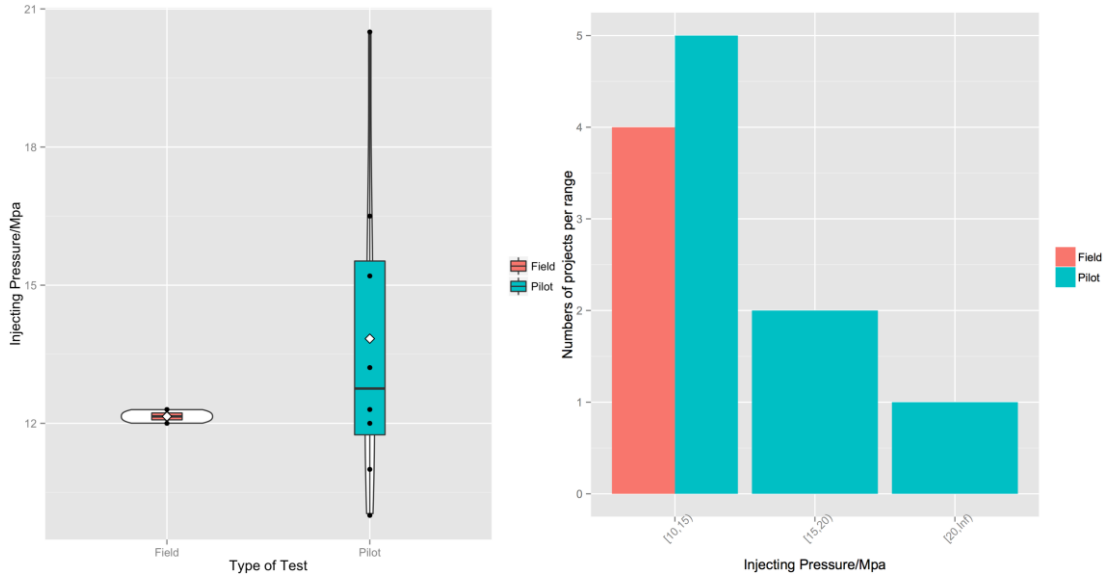

Figure 4.41. Pilot and field injecting pressure information (A) box and violin plot (B) histogram

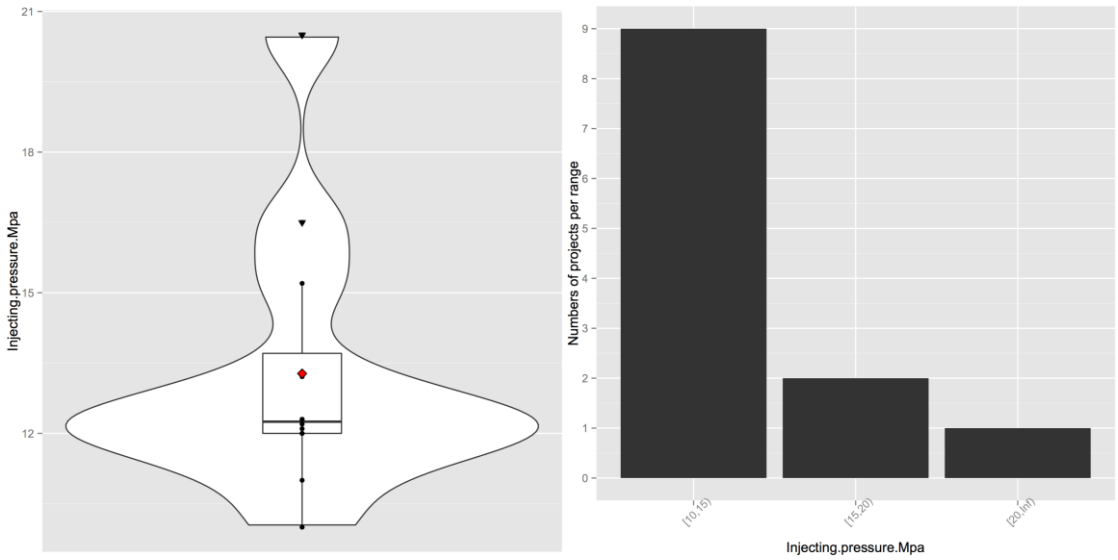

Figure 4.42. Total injecting pressure information (A) box and violin plot (B) histogram

4.2.3.4. Injection rate. The injection rate of polymer solution is also a key factor in the project and affects the whole development and effectiveness of polymer flooding. The injection rate has significant effect on the cumulative production time. Lower injection rates lead to longer production times and higher rates may increase shear degradation of the polymer. Injection rates must be controlled not too high to minimize polymer flow out of the pattern or out of the target zones. Also, the injection rate should not exceed the reservoir fracture pressure. To maximize the oil production, the injection rate should be maintained from 0.14 to 0.16PV/year for well spacing of 250m and 0.16-0.20 PV/year for well spacing of 150-175m (James J, 2013). Figure 4.43 shows the relationship between well spacing and injection rate that conform to previous research.

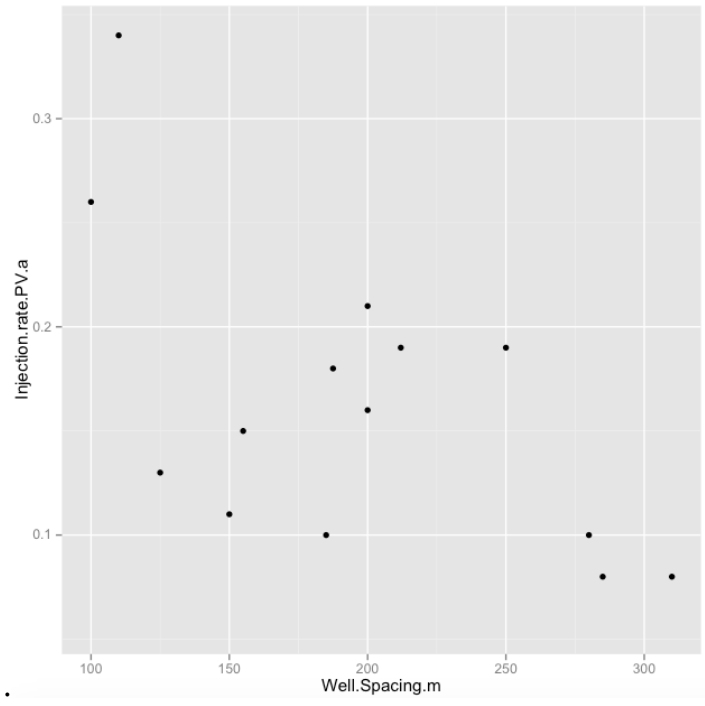

Figure 4.43. The relationship between well spacing and injection rate

Out of the 55 cases studied, 23 cases including 5 field cases and 18 pilot cases reported their injection rate values.

For field test cases, as shown in Figure 4.44, the minimum value is 0.1PV/a, the maximum value is 0.16PV/a and the mean value is 0.118PV/a. Figure 4.44 also shows that most values fall into the range of 0.3-0.4PV/a.

For pilot cases, as shown in Figure 4.44, the minimum value is 0.057PV/a, the maximum value is 0.34PV/a, and the mean value is 0.146PV/a. Figure 4.44 also shows that most values fall into the range of 0.2-0.6PV/a.

Combining field and pilot cases together, as shown in Figure 4.45, the minimum value is 0.057PV/a, the maximum value is 0.34PV/a and the mean value is 0.1399PV/a. Figure 4.45 also shows a normal distribution that the peak range is 0.3-04PV/a.

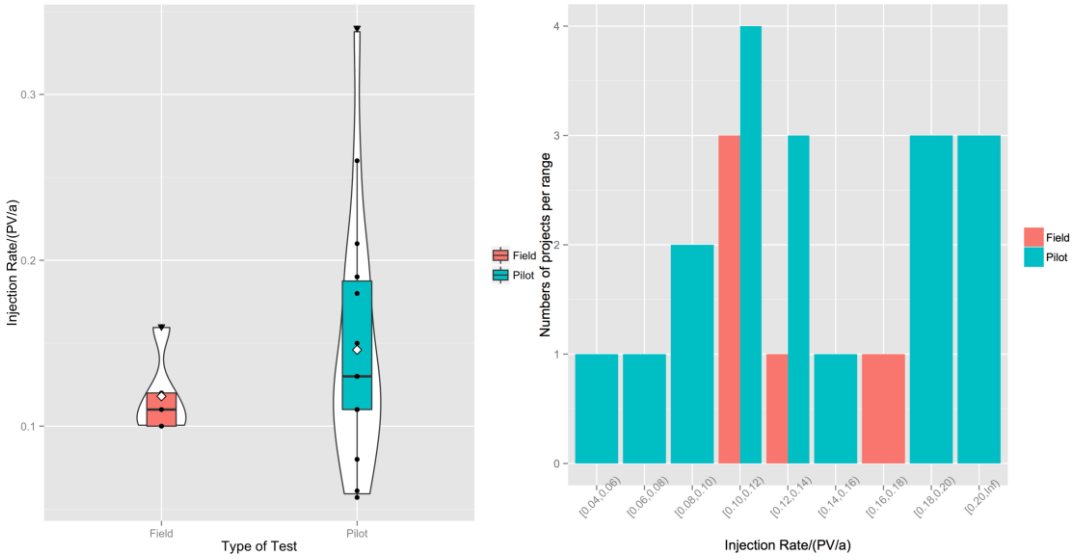

Figure 4.44. Pilot and field injection rate information (A) box and violin plot (B) histogram

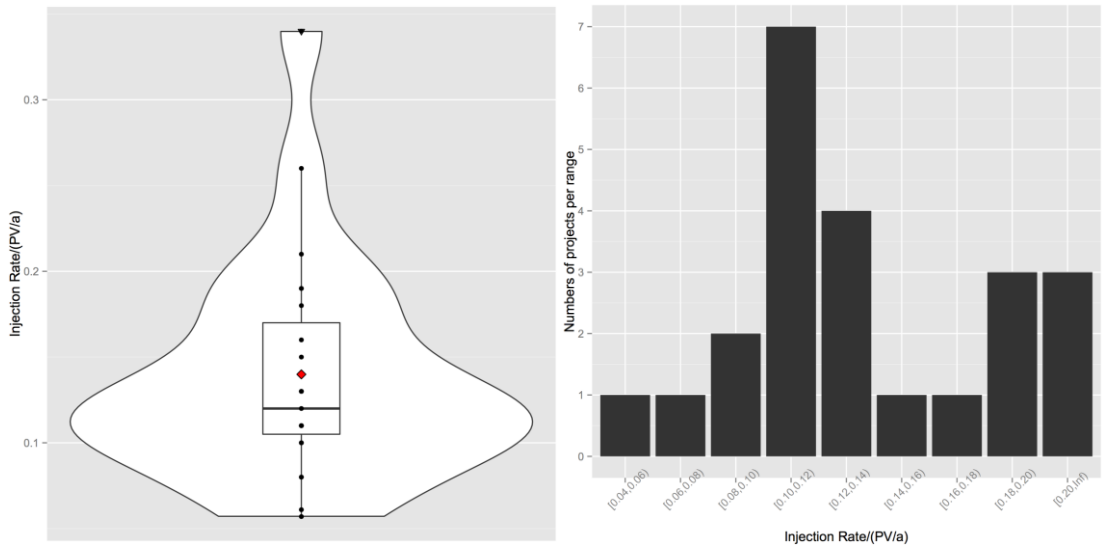

Figure 4.45. Total injection rate information (A) box and violin plot (B) histogram

4.2.3.5. Well pattern. According to (Li and Chen, 1995), the well pattern has a relatively small effect on the incremental oil recovery by polymer flooding as others. Table 4.3 provides an EOR comparison of various well patterns based on (Li and Chen, 2005) numerical simulations. The results indicate that the incremental recovery is10.9% for a line-drive pattern and 10.6% for an inverted 9-spot. For a 5-spot, the incremental oil recovery is 10.3%. However, the injection volume will be three times more for the inverted 9-spot than for the 5-spot which leads to a temptation to inject above the fracture pressure when using the inverted 9-spot pattern. Also, the connectivity factor will be much smaller with a line pattern than with the 5-spot. Therefore, the 5-spot pattern appears to be more attractive (Wang et al., 2009).

Table 4.3. Comparisons of various well patterns for polymer flooding (Li and Chen, 2005)

Well Pattern	Δη—EOR/%
Line in positive	**10.6**
Line in diagonal	10.9
5-spot	10.3
4-spot	10.1
9-spot	10.0
Inverted 9-spot	10.6
Parameters: 5 layers, net pay=39 ft, Vk=0.70, φ=0.26, k=101, 260, 491, 938, and $3207*10^{-3}\mu m^2$.	

In this study, as Figure 4.46 shows, there are 15 cases reported well pattern information. There are 10 5-spot patterns, 1 inverted 5-spot pattern, 2 inverted 9-spot patterns and 3 irregular well patterns. This validates that 5-spot well pattern is more popular in use than other well patterns.

4.2.3.6. Well spacing. Considered the interwell continuity, the well spacing is suggested to be from 200 to 250 m for oil zones with average permeability above 300-$400*10^{-3}\mu m^2$and net pay above 5 m. For oil zones with the average permeability above 100-$200*10^{-3}\mu m^2$ and the net pay of 1-5, 150-175 m is an ideal well spacing (Wang et al., 2009). Figure 4.47 shows the relationship between well spacing and reservoir average permeability.

Out of the 55 cases studied, 26 cases including 12 field cases and 14 pilot cases reported their well spacing values.

For field test cases, as shown in Figure 4.48, the minimum value is 150m, the maximum value is 280m and the mean value is 230m. Figure 4.48 also shows that most values fall into the range of 250-300m.

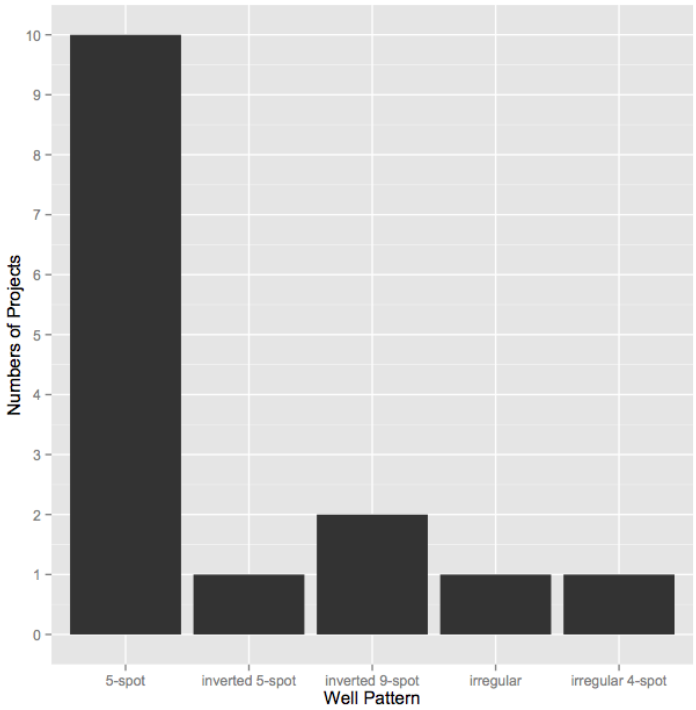

Figure 4.46. Well pattern information in this study

For pilot cases, as shown in Figure 4.48, the minimum value is 100m, the maximum value is 310m, and the mean value is 192m. Figure 4.48 also shows that most values fall into the range of 100-150m and 250-300m.

Combining field and pilot cases together, as shown in Figure 4.49, the minimum value is 100m, the maximum value is 310m and the mean value is 210m. Figure 4.49 also shows the peak range is 250-300m.

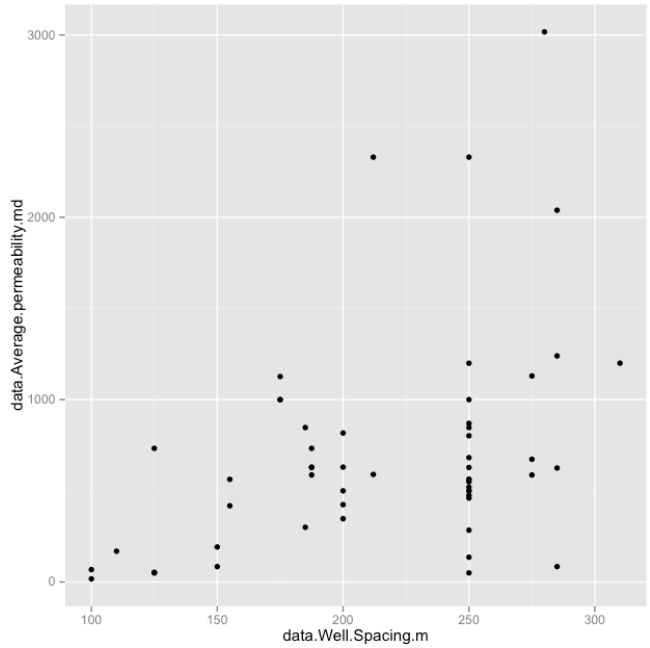

Figure 4.47. Relationship between well spacing and reservoir average permeability in this study

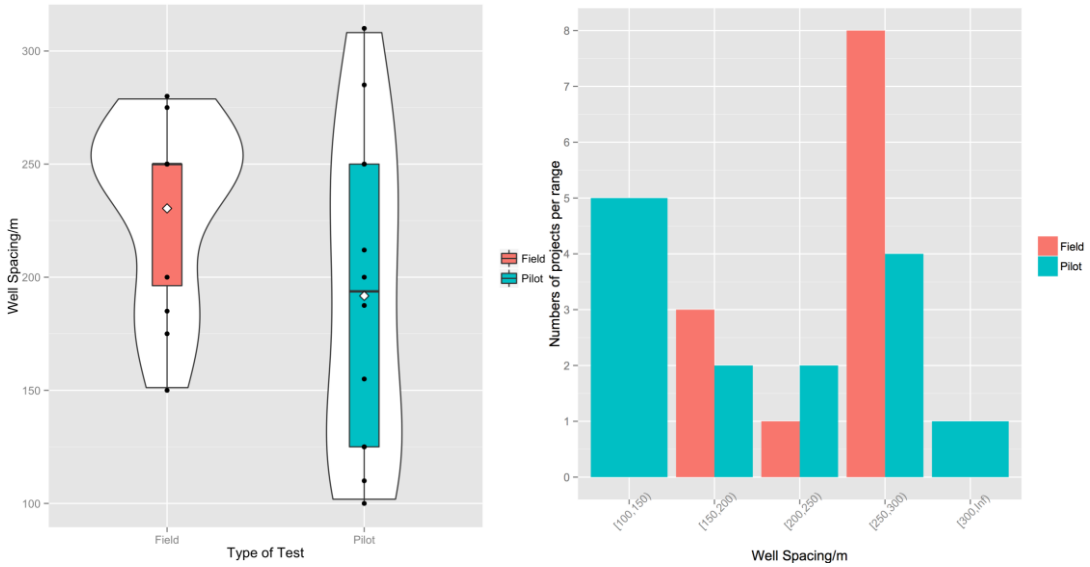

Figure 4.48. Pilot and field well spacing information (A) box and violin plot (B) histogram

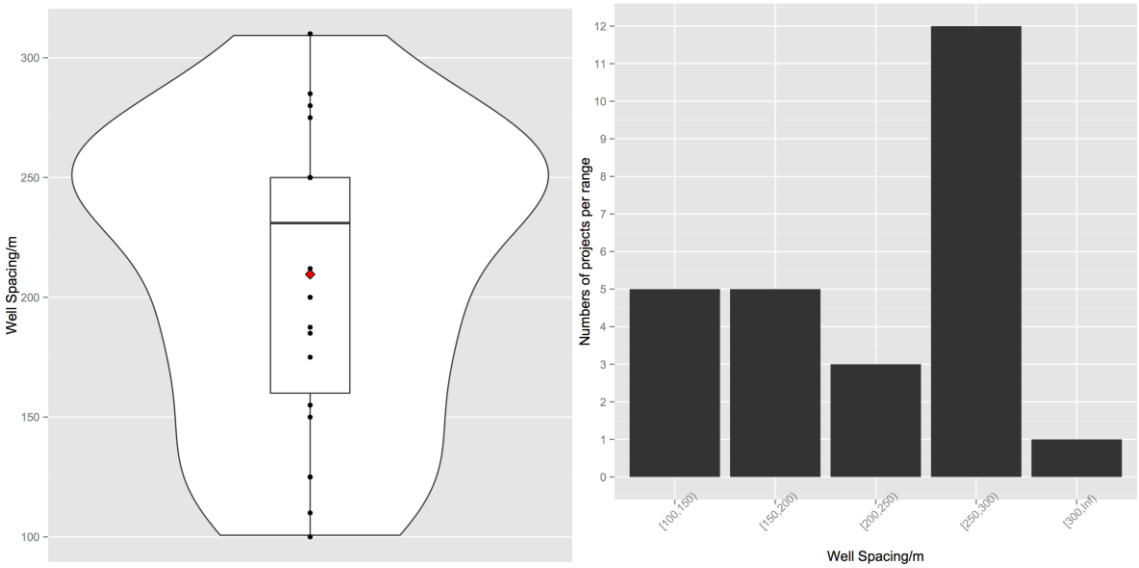

Figure 4.49. Total well spacing information (A) box and violin plot (B) histogram

4.2.4. Evaluation. Below are the parameters for evaluation information of polymer flooding.

4.2.4.1. Polymer volume injected. Based on theoretical research and practical experiences (Shao et al., 2001; Wang et al., 2009), the polymer volume should be determined by the gross water cut of the flooding unit. Generally, when the gross water cut achieves 92-94%, the polymer injection should be stopped.

Out of the 55 cases studied, 32 cases including 13 field cases and 19 pilot cases reported their polymer volume injected values.

For field test cases, as shown in Figure 4.50, the minimum value is 0.15, the maximum value is 0.875 and the mean value is 0.5068. Figure 4.50 also shows that most values fall into the range of 0.3-0.4.

For pilot cases, as shown in Figure 4.50, the minimum value is 0.033, the maximum value is 0.8, and the mean value is 0.4119. Figure 4.50 also shows that most values fall into the range of 0.2-0.6.

Combining field and pilot cases together, as shown in Figure 4.51, the minimum value is 0.033, the maximum value is 0.876 and the mean value is 0.4505. Figure 4.51 also shows the peak range is 0.3-0.4.

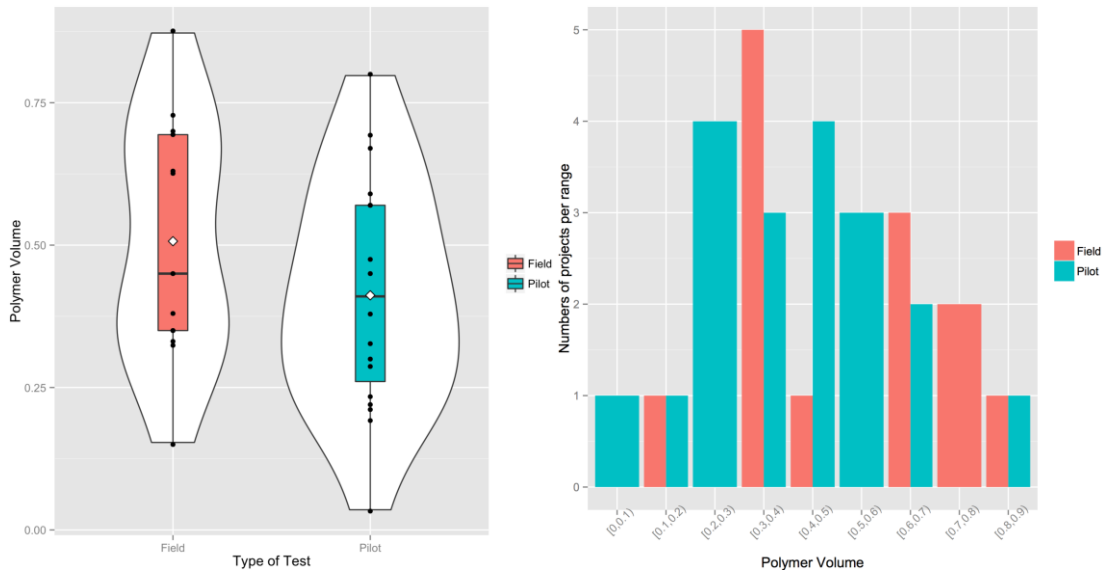

Figure 4.50. Pilot and field polymer slug size information (A) box and violin plot (B) histogram

4.2.4.2. Water cut before polymer flooding. Out of the 55 cases studied, 36 cases including 14 field cases and 22 pilot cases reported their water cut values before polymer flooding.

For field test cases, as shown in Figure 4.52, the minimum value is 86.1%, the maximum value is 96.49% and the mean value is 93.87%. Figure 4.52 also shows that most values fall into the range of 94%-98%.

For pilot cases, as shown in Figure 4.52, the minimum value is 78.51%, the maximum value is 0.8, and the mean value is 98.14%. Figure 4.52 also shows that most values fall into the range of 96%-98%.

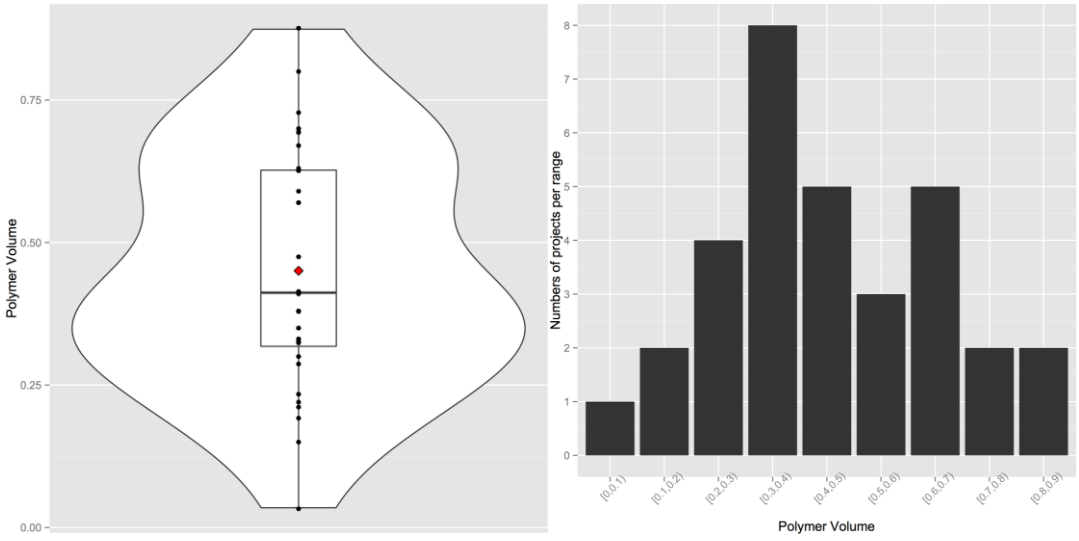

Figure 4.51. Total polymer slug size information (A) box and violin plot (B) histogram

Combining field and pilot cases together, as shown in Figure 4.53, the minimum value is 78.51%, the maximum value is 98.14% and the mean value is 93.13%. Figure 4.53 also shows the peak range is 94%-98%.

4.2.4.3. Water cut after polymer flooding. Out of the 55 cases studied, 26 cases including 9 field cases and 17 pilot cases reported their water cut values after polymer flooding.

For field test cases, as shown in Figure 4.54, the minimum value is 67.9%, the maximum value is 92.5% and the mean value is 84.2%.

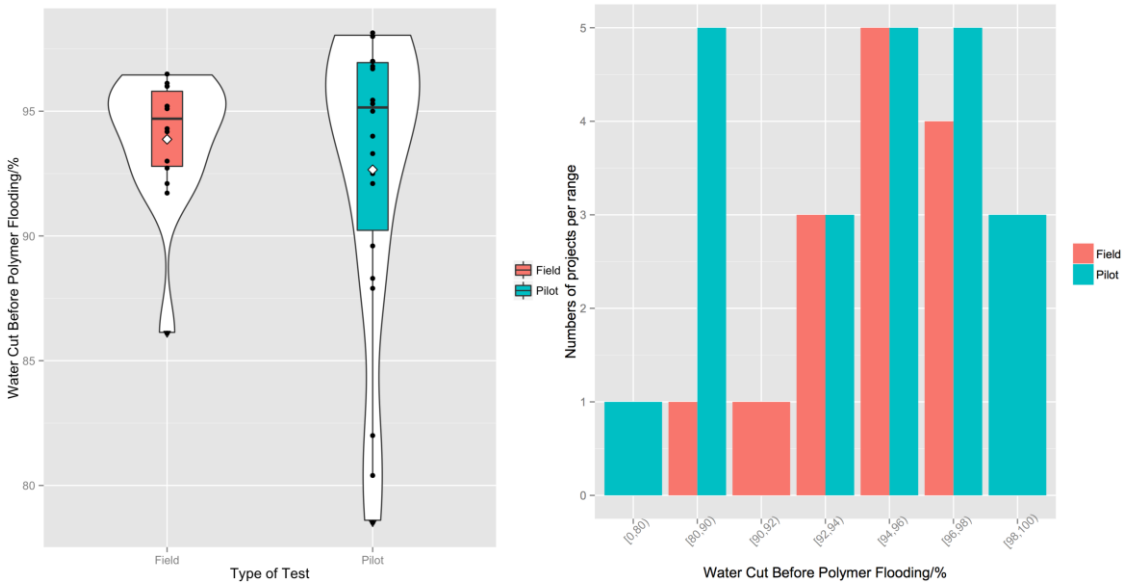

Figure 4.52. Pilot and field water cut before polymer flooding information (A) box and violin plot (B) histogram

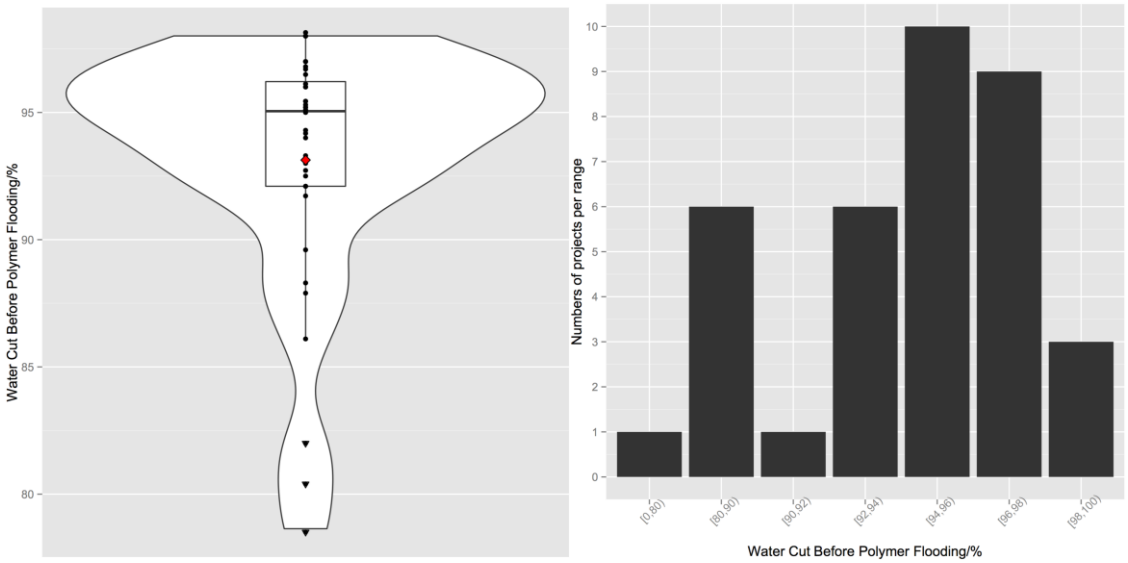

Figure 4.53. Total water cut before polymer flooding information (A) box and violin plot (B) histogram

Figure 4.54 also shows that most values fall into the range of 85%-90%.

For pilot cases, as shown in Figure 4.54, the minimum value is 69.16%, the maximum value is 95%, and the mean value is 84.6%. Figure 4.54 also shows that most values fall into the range of 90-95%.

Combining field and pilot cases together, as shown in Figure 4.55, the minimum value is 67.9%, the maximum value is 95% and the mean value is 84.46%. Figure 4.55 also shows the peak range is 85%-95%.

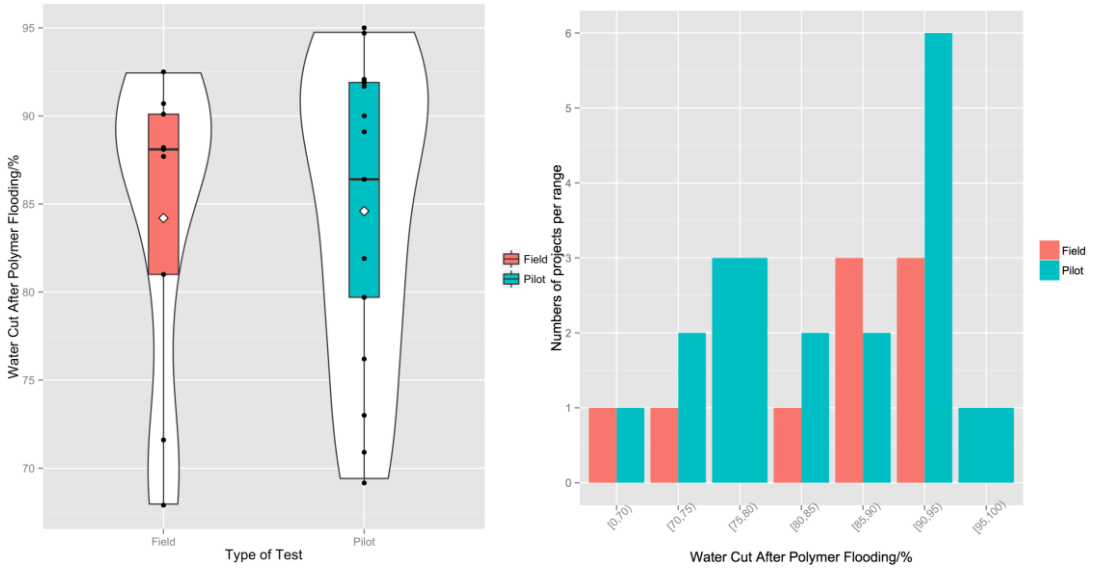

Figure 4.54. Pilot and field water cut after polymer flooding information (A) box and violin plot (B) histogram

4.2.4.4. Water cut decreased after polymer flooding. Out of the 55 cases studied, 26 cases including 8 field cases and 18 pilot cases reported their water cut values decreased after polymer flooding.

For field test cases, as shown in Figure 4.56, the minimum value is 2%, the maximum value is 25.1% and the mean value is 10.32%. most values fall into the range of 5%-10%. Figure 4.56 also shows that most values fall into the range of 5%-10%.

For pilot cases, as shown in Figure 4.56, the minimum value is 2.1%, the maximum value is 21.2%, and the mean value is 8.289%. Figure 4.56 also shows that most values fall into the range of 5%-10%.

Combining field and pilot cases together, as shown in Figure 4.57, the minimum value is 2%, the maximum value is 25.1% and the mean value is 8.916%. Figure 4.57 shows the peak range is 5%-10%.

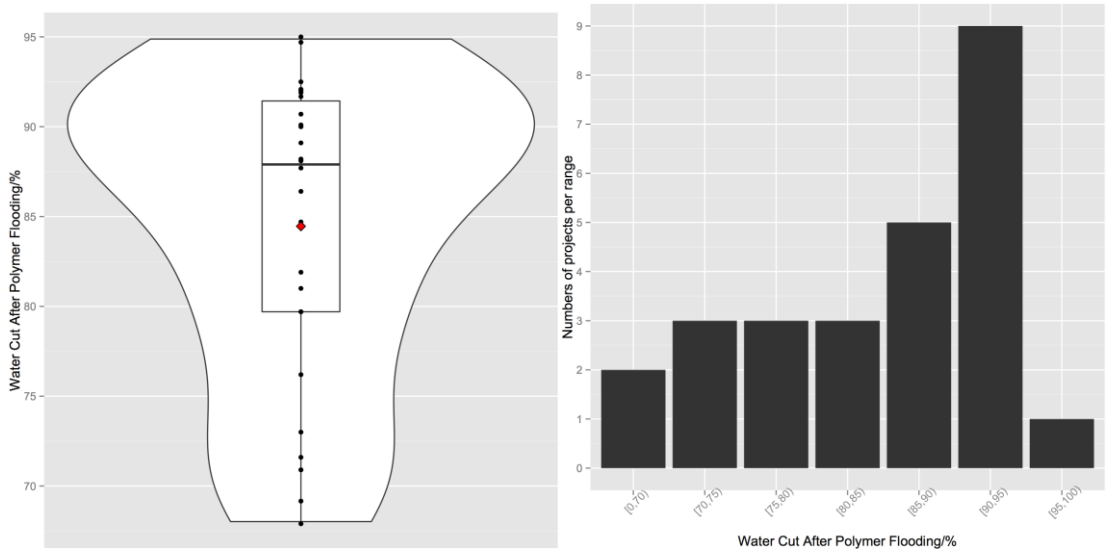

Figure 4.55. Total water cut after polymer flooding information (A) box and violin plot (B) histogram

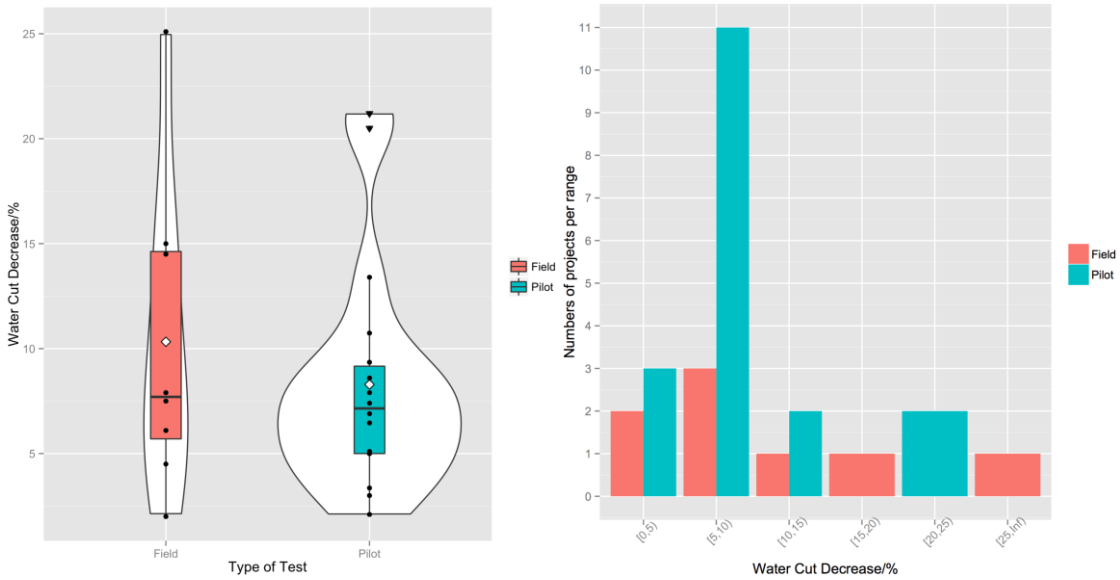

Figure 4.56. Pilot and field water cut decreased after polymer flooding information (A) box and violin plot (B) histogram

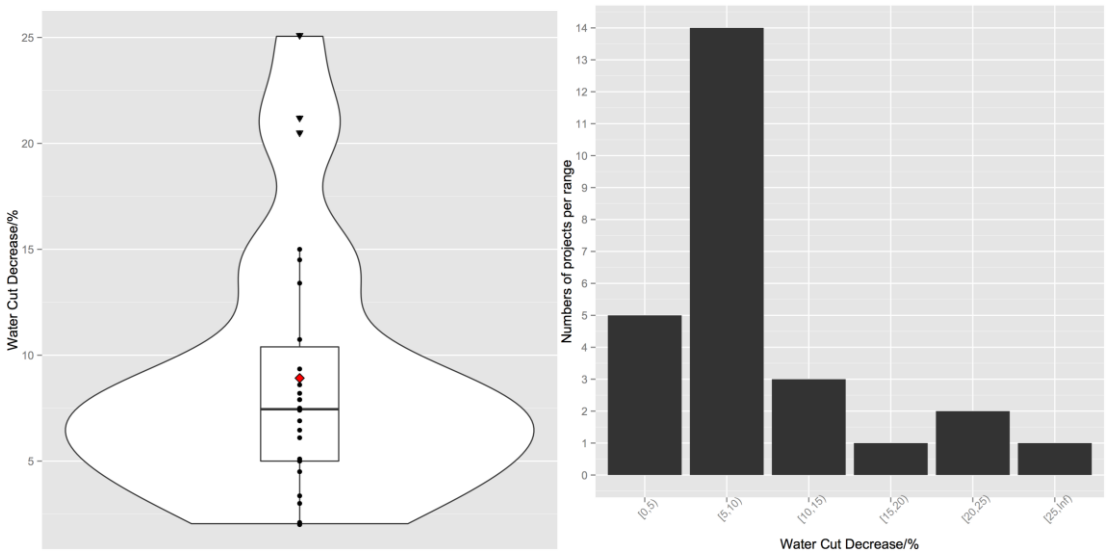

Figure 4.57. Total water cut decreased after polymer flooding information (A) box and violin plot (B) histogram

4.2.4.5. Oil recovery increased after polymer flooding. Out of the 55 cases studied, 29 cases including 14 field cases and 15 pilot cases reported their oil recovery increased values after polymer flooding.

For field test cases, as shown in Figure 4.58, the minimum value is 5.23%, the maximum value is 19.5% and the mean value is 9.449%. Figure 4.58 also shows that most values fall into the range of 5%-10%.

For pilot cases, as shown in Figure 4.58, the minimum value is 1.58%, the maximum value is 14.15%, and the mean value is 7.986%. Figure 4.58 also shows that most values fall into the range of 5%-10%.

Combining field and pilot cases together, as shown in Figure 4.59, the minimum value is 1.58%, the maximum value is 19.5% and the mean value is 8.692%. Figure 4.59 also shows the peak range is 5%-15%.

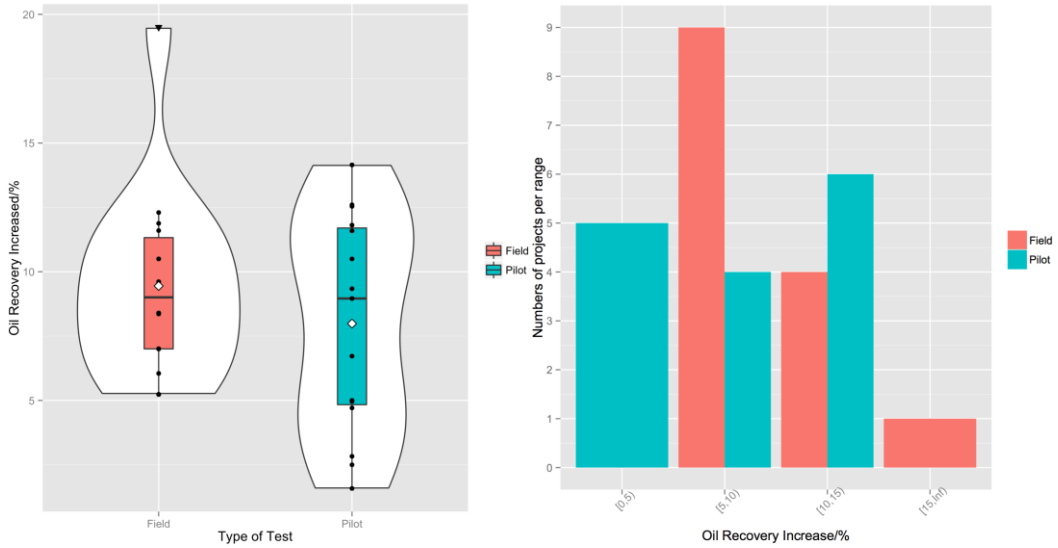

Figure 4.58. Pilot and field oil recovery increased information (A) box and violin plot (B) histogram

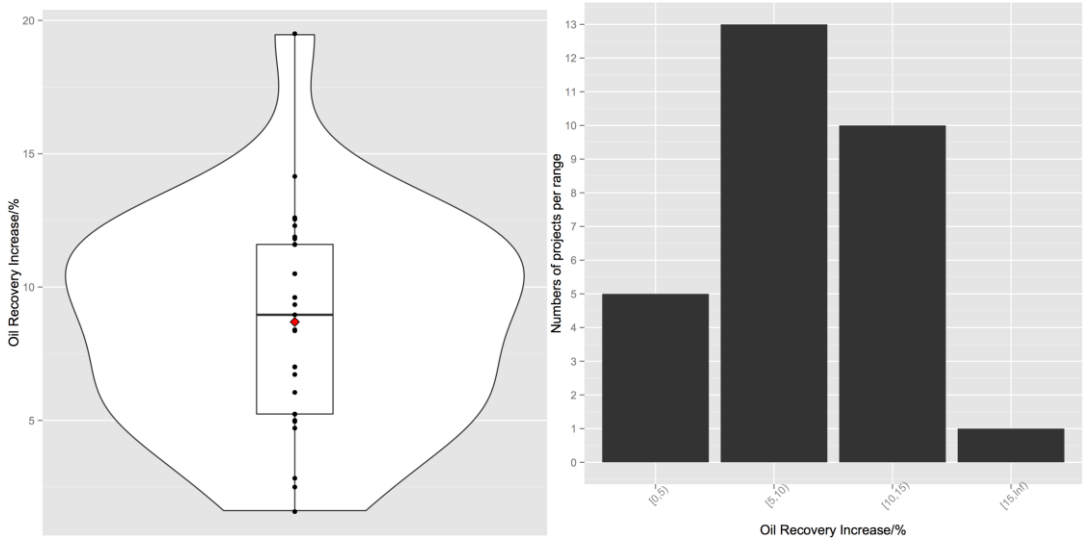

Figure 4.59. Total oil recovery increased information (A) box and violin plot (B) histogram

4.3. SUMMARIZING AND DISCUSSION ON SCREENING RANGE

Table 4.4, 4.5, 4.6 provides a summary of polymer flooding screening range for pilot, field and total dataset based on the above statistical analysis of the dataset in this study. This summary includes the parameters that contribute to the success of a polymer flooding project, including reservoir properties, polymer properties and evaluations. The minimum and maximum observations mean and median of dataset values are the standard statistics used to describe the screening range.

Table 4.7 shows a comparison between the screening range for polymer flooding in this work and previous work. Beyond this, Table 4.8 shows the proposed screening range for other parameters in this work that previous researches did not include.

Table 4.4. Summary of polymer flooding screening range for pilot dataset

Screening range for pilot dataset				
Parameters	Statistics			
	Mean	Median	Minimum	Maximum
Depth/ft	5135	5331	1558	8186
Net thickness/ft	33.46	35.1	12.1	53.8
Temperature/ºF	155.9	159.5	78.89	200.7
Oil gravity/° API	24.81	18.3	14.96	51.1
Oil viscosity/cP	61.99	32.5	2.3	285.2
Water salinity/ppm	23140	18000	2127	84130
Average permeability/md	601.5	519	17	2330
Dykstra Parsons Coefficient	0.6761	0.715	0.4	0.87
Porosity/%	20.79	20	8.3	32
Average molecular weight/10^4	1436	1500	600	2500
Average polymer viscosity/cP	43.82	23	15	91.1
Average polymer concentration/ppm	1325	1350	600	2000
Injecting pressure/Mpa	13.84	12.76	10	20.5
Injection rate/(PV/a)	0.146	0.13	0.057	0.34
Well spacing/m	191.8	193.8	100	310
Polymer slug size/PV	0.4119	0.41	0.033	0.8
Water cut before polymer flooding/%	92.66	95.15	78.51	98.14

Table 4.5. Summary of polymer flooding screening range for field dataset

Screening range for field dataset				
Parameters	Statistics			
	Mean	Median	Minimum	Maximum
Depth/ft	4101	4034	3215	5139
Net thickness/ft	36	40.35	13.2	54.8
Temperature/ºF	145.7	153.7	93.7	176.5
Oil gravity/° API	36.89	40	16.2	53.2
Oil viscosity/cP	29.31	9.7	2.6	76.96
Water salinity/ppm	8087	6326	3580	28870
Average permeability/md	775.7	630	192	3017
Dykstra Parsons Coefficient	0.7717	0.765	0.7	0.87
Porosity/%	23.47	20.3	18.2	34.8
Average molecular weight/10^4	1896	1800	1075	2750
Average polymer viscosity/cP	87.01	66.35	17.75	220.4
Average polymer concentration/ppm	1250	1181	650	2050
Injecting pressure/Mpa	12.15	12.15	12	12.3
Injection rate/(PV/a)	0.118	0.11	0.1	0.16
Well spacing/m	230.4	250	150	280
Polymer slug size/PV	0.5068	0.45	0.15	0.876
Water cut before polymer flooding/%	93.87	94.7	86.1	96.49

Table 4.6. Summary of polymer flooding screening range for combined pilot and field dataset

Screening range for combined pilot and field dataset				
Parameters	Statistics			
	Mean	Median	Minimum	Maximum
Depth/ft	4819	4593	1558	8186
Net thickness/ft	34.4	39.7	12.1	54.8
Temperature/ºF	151.4	153.7	78.89	200.7
Oil gravity/° API	32.06	33.6	14.96	53.2
Oil viscosity/cP	47.61	18	2.3	285.2
Water salinity/ppm	17120	7500	2127	84130
Average permeability/md	669.7	575	17	3017
Dykstra Parsons Coefficient	0.7079	0.7350	0.4	0.87
Porosity/%	21.46	20	8.3	34.8
Average molecular weight/10^4	1574	1650	600	2750
Average polymer viscosity/cP	65.41	48.18	15	220.4
Average polymer concentration/ppm	1296	1247	600	2050

Table 4.6. Summary of polymer flooding screening range for combined pilot and field dataset (cont.)

Injecting pressure/Mpa	13.28	12.25	10	20.5
Injection rate/(PV/a)	0.1399	0.12	0.057	0.34
Well spacing/m	209.6	231	100	310
Polymer slug size/PV	0.45	0.412	0.033	0.876
Water cut before polymer flooding/%	93.13	95.05	78.51	98.14

Table 4.7. Comparison between the screening range for polymer flooding in this work and previous work

Proposed by	Permea Bility/md	Tempe rature/°F	Water salinity /ppm	Oil viscosity/ cP	Oil gravity/° API	Depth/ft
NPC (1976)	≥20	≤93		≤200		
Brashear and Kuuskraa (1978)	＞20	＜93	＜ 50,000	＜20	＞15	
Chang (1978)	＞20	＜93		＜200		
Carcoana (1982)	＞50	＜80	Low	50-80	25-35	＜6561
NPC (1984)	＞10	＜121	＜ 200,00 0	＜150		
Goodlett et al. (1986)	＞20	＜93.3	＜ 100,00 0	＜100	＞25	＜9000
Taber et al. (1997a, b)	＞10	＜93.3		10＜Oil viscosity ＜150	＞15	＜9000
Al-Bahar et al. (2004)	＞50	＜70	＜ 100,00 0	＜150		
Aladasani and Bai (2011)	2-5500	74-237		1-4000	13-43	700-9460
Saleh et al. 2014a	＞10	＜99		＜5000	＞12	
Saleh et al. 2014b	1-5500	65-210		0.3-130		
Sheng, J et al. (2015)	50	＜93	＜ 50,000	＜150		
This work	≥17	≤200	2127-74130	≤285	＞15	1158-8186

Table 4.8. Screening range for other parameters in addition to previous researches

Parameters	Suggest Range
Net thickness/ft	**＞12**
Dykstra Parsons Coefficient	**＜0.87**
Porosity/%	**＞8.3**
Average molecular weight/10^4	**600-2750**
Average polymer viscosity/cp	**＜220**
Average polymer concentration/ppm	**600-2050**
Injecting pressure/Mpa	**10-20.5**
Injection rate/(PV/a)	**0.12-0.14**
Well spacing/m	**200-310**
Polymer slug size/PV	**＞0.45**
Water cut before polymer flooding/%	**93-98**

5. MULTIPLE IMPUTATION AND MULTIPLE LINEAR REGRESSION

Regression analysis is one of the statistical methods to estimate the relationships among different variables. Regression analysis specifically helps one understand how the typical value of the dependent variable varies when one or more of the independent variables are changed. Regression analysis is widely used for prediction and forecasting by revealing forms of relationships between dependent variables and independent variables. Regression analysis has been widely used in petroleum industries. Hawkins (1994) used regression analysis for integrated formation evaluation. Jablonowski (2011) used regression analysis to identify HSE leading indicators. Bandyopadhyay (2011) used regression analysis to model the improved estimation of bubble point pressure of crude oils. However, rare applications of regression analysis in oil recovery prediction using the same or similar set of reservoir and fluid parameters were found. This study investigates the complex relationships among oil recovery increased after polymer flooding and other related reservoir and fluid properties using multiple linear regression method. The final result of regression analysis can be used as a guided prediction model in further data analysis of enhanced oil recovery projects.

5.1. DESCRIPTION OF THE MODEL DATASET

By collecting reservoir and fluid information from publications of polymer flooding from China, the dataset that used in this study was created. The dataset contains almost all the pilot and field applications of polymer flooding from different blocks of different oil fields. The dataset in this study includes 55 polymer flooding projects from Daqing, Henan, Shengli, Liaohe, Huabei, Changqing, Yanchang, Xinjiang, Bohai, Zhongyuan, Jidong oil fields.

Some parameters that less significant to polymer flooding or in case of colinearity between independent variables were not chosen in order to build a reliable predictive model. The parameters that chosen include reservoir properties of water salinity (ppm), average permeability (md), dykstra parsons coefficient and polymer properties of polymer concentration (ppm), polymer average molecular weight (10^4) and well information of well spacing (m) and evaluation information of polymer volume injected (PV), water cut before polymer flooding (%), oil recovery increased after polymer flooding (%).

However, not all the data that required can be found in publications. Thus, There's a problem that affected the dataset's quality that the dataset contains quite a lot missing values for each parameter. Each oil field has missing values for one or more parameters. Each case in the dataset has at least one missing value for one parameter. Thus, there are no complete cases in this work. Since regression analysis process cannot go through with a case that has missing values even one, methods should be found dealing with missing values in order to make the process successful.

5.2. MISSING DATA PATTERNS

Missing data or missing values appear when there is no value stored for a variable in a dataset. Missing data are common in data analysis and can have a significant effect on the conclusion that can be drawn from the dataset. There are three types of missing data. The missing data type of this dataset should be known before doing further analysis in case of bias.

5.2.1. Missing Completely At Random. Values in a dataset are missing completely at random (MCAR) only if the events or projects that cause any particular data value being missing which is independent both of observable variables and of unobservable parameters of interest and appear at random entirely. The analyses performed on the datasets that have missing values completely at random are unbiased. However, data are rarely MCAR (Polit, 2012).

5.2.2. Missing At Random. Missing at random (MAR) is an alternative condition and occurs then the missing values are related to a particular variable but not related to the values of the variable that has missing data (Polit, 2012).

5.2.3. Missing Not At Random. Missing not at random (MNAR) is the value is missing for a specific or particular reason that is different with other common values (Polit, 2012). For example, if the permeability value in one area is missing because this area doesn't have permeability value. It's impossible for this study, thus it can be ignored.

Based on the above, the type of missingness of this study is missing completely random since the value that is missing is not given by the author of publication thus it has nothing to do with other values within the variable and also other variables. Therefore, there should be none or less bias when dealing with missing data using statistical methods.

5.3. IMPUTATION METHODS

Missing data reduce the representativeness of the sample and thus can distort the inferences from the dataset. While facing the occurrence of missing data, it's often advised for the researcher to plan to use methods of data analysis methods that are proper and robust to missingness. An analysis is regarded as proper and robust when mild to

moderate violations of the method's key assumptions will produce little or no bias or distortion in the conclusions.

5.3.1. Case Deletion/Data Deletion. The most common method of dealing with missing data by far is listwise deletion that all cases with a missing value are deleted. This method is simple but it may introduce bias or affect the representativeness of the results or decrease power of the analysis by decreasing the effective sample size. The dataset in this study is from publications that has not very much data as from EOR survey thus if delete the cases with missing values it will not have enough data points to do data analysis and build a model. Therefore, this method shouldn't be used for this dataset.

5.3.2. Single Imputation. Single imputation includes methods of hot-deck, cold-deck, mean, regression, stochastic regression and other single imputation methods. Although single imputation has been widely used, it cannot reflect the full uncertainty created by missing data. Therefore, a more reliable method should be found for this dataset.

5.3.3. Multiple Imputation. In case of increased noise due to single imputation, Rubin (1987) developed a method called multiple imputation by averaging the outcomes across multiple imputed data sets to account for this. The imputation processes is similar to stochastic regression are run on the same dataset multiple times and the imputed datasets are saved for later analysis. Each of the imputed datasets is analyzed separately and the results are averaged except for the standard error. The standard error is constructed by the variance within each data set and the variance between imputed values on each dataset. Thus, the noises due to imputation process as well as the residual variance are introduced to the regression model because of the standard error determined by the square root of these two variances added together (Rubin, 1987).

Multiple imputation draws values of parameters from a posterior distribution that reflects the noise associated with the uncertainty surrounding the parameters of the distribution that generates the data. Therefore, multiple imputation simulate both the date generating process and the uncertainty associated with the parameters of the probability distribution of the data. More traditional imputation methods fail to give a complete simulation of the uncertainty associated with missing data.

Figure 5.1 shows the general three steps for application of multiple imputation. First step is imputation that imputes the missing entries of the incomplete datasets not once but several m times (here m=3 in the Figure). The imputed values are drawn for a posterior distribution that can be different for each missing entry. At last of the first step, three complete datasets with impuations are created. Second step is to analyze that each of the m completed datasets with regression analysis for a result. The third step is pooling process that integrate the m analysis results into a final result.

In this study, multiple imputation is used to impute the dataset in order to build a complete dataset to build the prediction model. There are also many softwares available to use to implement multiple imputation as R, SAS, SPSS, etc. R is chosen to do the multiple imputation in this study.

The package “MICE (Multiple Imputation by Chained Equations)” is used within R for its compatibility with the dataset in this study, for example, the dataset in this study is not multivariate normal which violates the assumption of some other packages. The reason its name is “chained equations” is it assumes that for each incomplete variable the user specifies a conditional distribution for the missing data given by the other data.

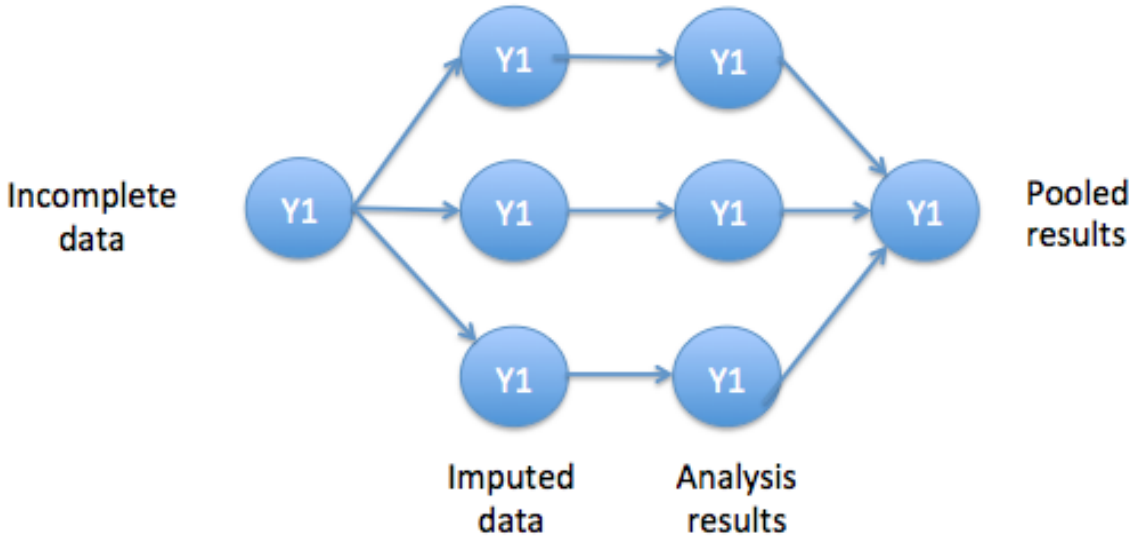

Figure 5.1. Three steps for application of multiple imputation

For instance, incomplete binary variables could use logistic regression, categorical data could use polytomous regression, and numerical data could use linear regression. Under this assumption that a multivariate distribution exists, from which these conditional distributions can be derived. MICE constructs a Gibbs sampler from the specified conditions, which is used to generate multiple imputaions. A number of publications document the method (Van Buuren et al. 1999; Brand 1999). Thus this MICE package is useful and convenient to generate multivariate multiple imputations, it also suits the dataset in this study that only has numerical data.

The default method in MICE for imputating numerical data is predictive mean matching (PMM) method. Predictive Mean Matching (PMM) is a semi-parametric imputation approach. It is similar to the regression method except that for each missing value, it fills in a value randomly from among the a observed donor values from an

observation whose regression-predicted values are closest to the regression-predicted value for the missing value from the simulated regression model (Heitjan and Little 1991;

Schenker and Taylor 1996). The PMM method ensures that imputed values are plausible and it might be more appropriate than the regression method (which assumes a joint multivariate normal distribution) if the normality assumption is violated (Horton and Lipsitz 2001, p. 246). Since the dataset in this study is not multivariate normal distribution, it can fit this method without violating the assumptions as other methods.

In the traditional regression method, a regression model is fitted for a continuous variable with the covariates constructed from a set of effects. Based on the fitted regression model, a new regression model is simulated from the posterior predictive distribution of the parameters and is used to impute the missing values for each variable (Rubin 1987, pp. 166–167). That is to say, for a continuous variable Y_j with missing values, a model

$$Y_j = b_0 + b_1 X_2 + b_2 X_2 + ... + b_k X_k \quad (5)$$

is fitted using observations with observed values for the variable Y_j and its covariates $X_1, X_2, ..., X_k$. The fitted model includes the regression parameter estimates $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k)$ and the associated covariance matrix $\hat{s}_j^2 \mathbf{V}_j$, where $\mathbf{V}_j$ is the usual $\mathbf{X}'\mathbf{X}$ inverse matrix derived from the intercept and covariates $X_1, X_2, ..., X_k$.

The following steps are used to generate imputed values for each imputation:

New parameters $\boldsymbol{\beta}_* = (\beta_{*0}, \beta_{*1}, ..., \beta_{*(k)})$ and $\hat{s}_{*j}^2$ are drawn from the posterior predictive distribution of the parameters.

That is to say, they are simulated from $(\hat{b}_0, \hat{b}_1, ..., \hat{b}_k)$, $\hat{s}_j^2$, and $\mathbf{V}_j$. The variance is drawn as

$$\hat{s}_{*j}^2 = \hat{s}_j^2 (n_j - k - 1) / g$$

(6)

where g is a $x_{nj}^2 - k - 1$ random variate and nj is the number of nonmissing observations for Y_j. The regression coefficients are drawn as

$$\beta_* = \hat{\beta} + \sigma_{*j} \mathbf{V}'_{hj} \mathbf{Z}$$

(7)

where $\mathbf{V}_{hj}$¢ is the upper triangular matrix in the Cholesky decomposition, $\mathbf{V}_j = \mathbf{V}_{hj}$¢$\mathbf{V}_{hj}$, and $\mathbf{Z}$ is a vector of $k+1$ independent random normal variates.

The missing values are then replaced by

$$b_{*0} + b_{*1x1} + b_{*2x2} + ... + b_{*(k)xk} + z_i s_{*j} \quad (8)$$

where $x_1, x_2, ..., x_k$ are the values of the covariates and z_i is a simulated normal deviate.

Following the description of the model above about the traditional regression method for missing data below is how predictive mean matching method generates imputed values.

For each missing value in the dataset, a predicted value

$$y_{i*} = b_{*0} + b_{*1x1} + b_{*2x2} + ... + b_{*(k)xk} \quad (9)$$

is computed with the covariate values $x_1, x_2, ..., x_k$.

Then a set of k observations whose corresponding predicted values are closest to y_{i*} is generated. Finally, the missing value is replaced by a random value drawn from these k observed values. The predictive mean matching method requires the number of closest observations k to be specified. The default in MICE package is $k = 5$. That is to say, each missing case is matched to the 5 cases that have the closest predicted values. One of the 5 cases is chosen at random and its value is assigned to the case with missing data. In SPSS and Strata, the k is only 1 which is a serious error because it's not enough to produce proper imputations and the estimated standard errors tend to be too much low which leads to inflated test statistics and confidence intervals that are much too narrow (Morris et al. 2014). Schenker and Taylor (1996) did simulations with k=3 and k=10 finding there was small difference in performance but there was less bias and more sampling variation with k=3. Based on previous simulations, Morris et al. (2014) recommended k=10 is good for most situations but a lot depends on sample size. $_k$=10 is probably the better choice with large samples. Otherwise, it will probably include too many cases that are rather unlike the case to which they are matched with smaller samples. Thus, MICE in which k=5 is good for most of samples and is popular in multiple imputation researches. Also, it's fit with the dataset in this study to generate plausible results.

5.4. PROCESS OF MULTIPLE IMPUTATION

5.4.1. Missing Data Inspection. Before the imputation process, a data inspection is given below to show the distribution of missing data in the dataset. Thus a big picture of missing data can be seen to do further review and deep analysis.

Figure 5.2 shows a margin plot of polymer volume injected versus oil recovery increased as an example to express the situation of the incomplete data. The blue dots in

the big area indicate the data points that observed both in PV and oil recovery increased. The red dots at left indicate that observed in oil recovery increased but missed in PV. The red boxplot at left correspond to the red dots and the blue boxplot at left indicates the distribution of observed data in oil recovery increased. On the opposite, the red dots at bottom indicate observed in PV but missed in oil recovery increased. The orange dot at the left bottom corner indicates that there are records both oil recovery increased and PV are missing and the number is 17 as shows. Also, 26 records in which 26 out of 55 data points of oil recovery increased are missing and 23 out of 55 data points of PV are missing.

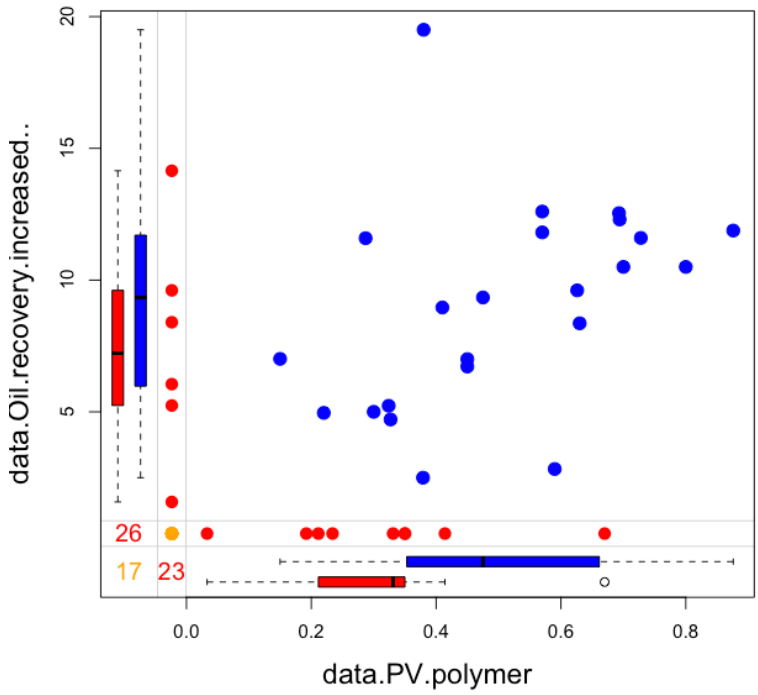

Figure 5.2. Margin plot about PV vs. oil recovery increased.

5.4.2. Create Imputations. Creating imputations can be down with code to generate. In this study, the multiple imputation times m=5 is chosen. Rubin (1996) claims that unless the rate of missing information is very high, there is nearly little advantage to producing and analyzing more than a few imputed datasets in most situations. Thus, m=5 is quiet fit this dataset in this study.

5.4.3. Diagnostic Checking. After imputations were done, diagnostic checking is an important step in multiple imputation to evaluate whether the imputations are plausible or not. Imputaitons should be values that can have been obtained had they not been missing and should be close to the data available. Data values that are obviously impossible (e.g. porosity more than 1, negative PV) should not appear in the imputed data. Imputaitons should respect relations between each two variables and reflect the appropriate amount of uncertainty about their 'true' values.

Figure 5.3 shows an imputation example of PV and each row corresponds to a missing case in PV till total of 23 cases out of 55 cases. The column at most left is the cases number of dataset and the rest columns are the results of multiple imputaions after five iterations. The actual results are different due to different value randomly drawn during the imputation process. Also, the complete data combine both observed and imputed data can be viewed. It is often useful and meaningful to inspect the distributions of original and the imputed data.

Figure 5.4 shows a strip plot of an example of distributions of the observed data and imputed data of PV. The number 0 represents the original dataset and the numbers 1 to 5 represent 5 complete datasets with observed data and imputed data. The observed data points are plotted in blue and the imputed data points are plotted in red.

	1	2	3	4	5
9	0.6930	0.6700	0.6260	0.5700	0.2870
11	0.8000	0.3790	0.3800	0.1500	0.4141
16	0.4100	0.2340	0.4750	0.5700	0.4500
19	0.8760	0.4750	0.6930	0.3310	0.0330
20	0.6300	0.1920	0.3500	0.4750	0.2870
21	0.8000	0.5900	0.3500	0.6300	0.4500
23	0.2113	0.5900	0.3500	0.2113	0.5700
24	0.7000	0.3790	0.6300	0.6700	0.1920
25	0.4750	0.1920	0.2870	0.1500	0.5900
26	0.3500	0.3500	0.1500	0.7000	0.3240
27	0.3000	0.2113	0.3000	0.8760	0.5900
28	0.3240	0.1920	0.3240	0.4500	0.8000
29	0.3500	0.8000	0.3240	0.2113	0.3240
37	0.5700	0.3240	0.4500	0.2870	0.5700
38	0.6930	0.6930	0.4141	0.2870	0.6260
43	0.1500	0.5900	0.2113	0.2113	0.1500
44	0.3310	0.3000	0.3790	0.6260	0.4100
45	0.6930	0.2200	0.6700	0.3310	0.0330
46	0.3000	0.6940	0.3000	0.4100	0.6700
49	0.0330	0.6300	0.5700	0.3790	0.1920
50	0.7000	0.3270	0.3500	0.2340	0.4100
52	0.3500	0.5900	0.3000	0.3000	0.3800
53	0.5700	0.1500	0.6300	0.2870	0.8760

Figure 5.3. Example of imputation of PV (m=5)

Since the predictive mean matching method draws imputations from the observed data, the imputed values have the same gaps as in the observed data and are always within the range of the observed data. Also, the distributions of the observed data and the imputed data are similar thus the imputed data could reflect the feature like clusters around PV equals 0.4 and 0.6 closely.

Figure 5.5 shows a scatter plot group of polymer volume injected versus oil recovery increased after the 5 imputations. The 0 plot is the original scatter plot without the imputation and number 1 to 5 are scatter plots after imputaions in which the red dots represent the imputed values and the blues represent the observed values.

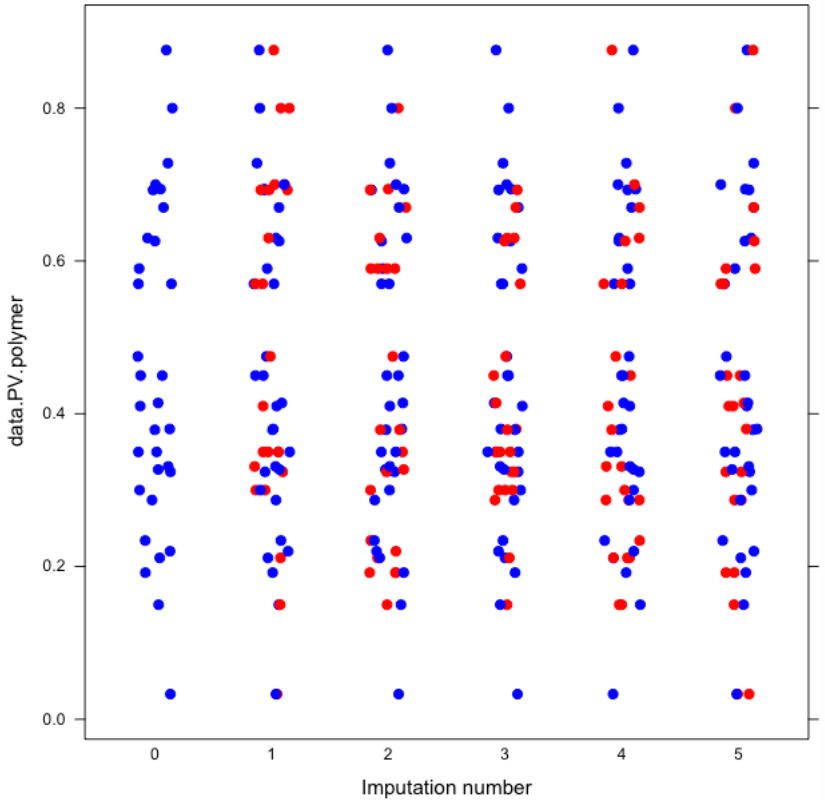

Figure 5.4. Strip plot of distributions of compared original and imputed dataset of PV

The blue dots are the same in different plots but the red ones vary. The red dots have more or less the same shape as blue dots that indicates that they could have been plausible measurements if they had not been missing. The differences between the red points represent the uncertainty about the true but unknown values.

Figure 5.6 shows a density plot of comparison of observed and imputed data of polymer volume injected. It indicates that the imputaions are reasonable because nearly all 5 imputaions have nearly the same density as the observed data except number 3 in which imputed values are a little bit higher than the observed data.

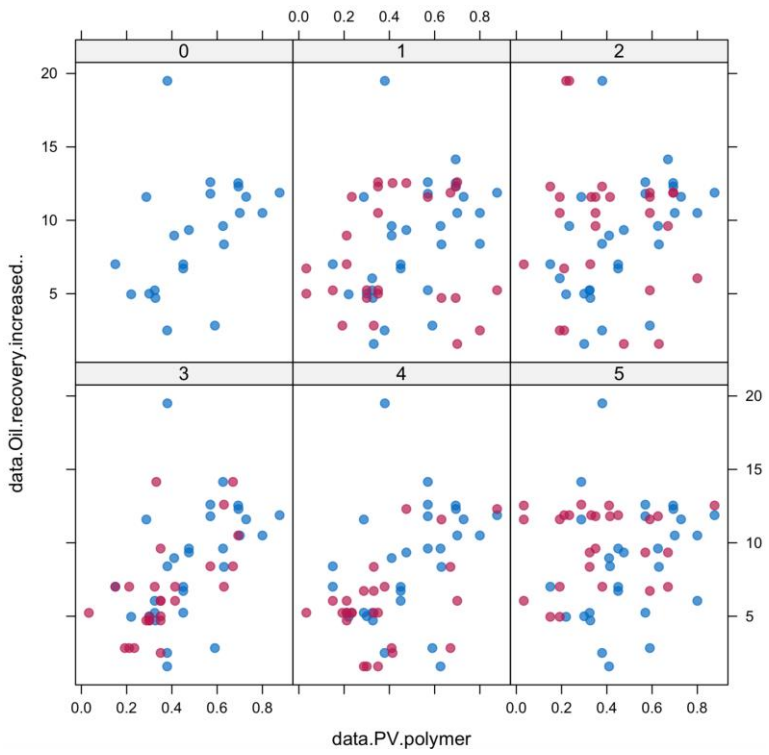

Figure 5.5. Scatter plot of PV and oil recovery increased with imputed data

Differences in the densities between the observed and imputed data may suggest a problem that needs to be further checked. The reason for number 3 in which is that the probability that polymer volume injected is missing in small values like 0.2-0.3 is larger since most of the oil fields have higher polymer volume injected like 0.7. Plots like this are very useful to detect interesting difference between the observed and imputed data.

5.4.4. Analysis of Imputed Data. After multiple imputation, the dataset now is complete and ready to do the multiple regression analysis. Multiple linear regressions are used to evaluate the relationship between a single response (Y) and more than one predictor variable ($x_1, x_2, ..., x_k$).

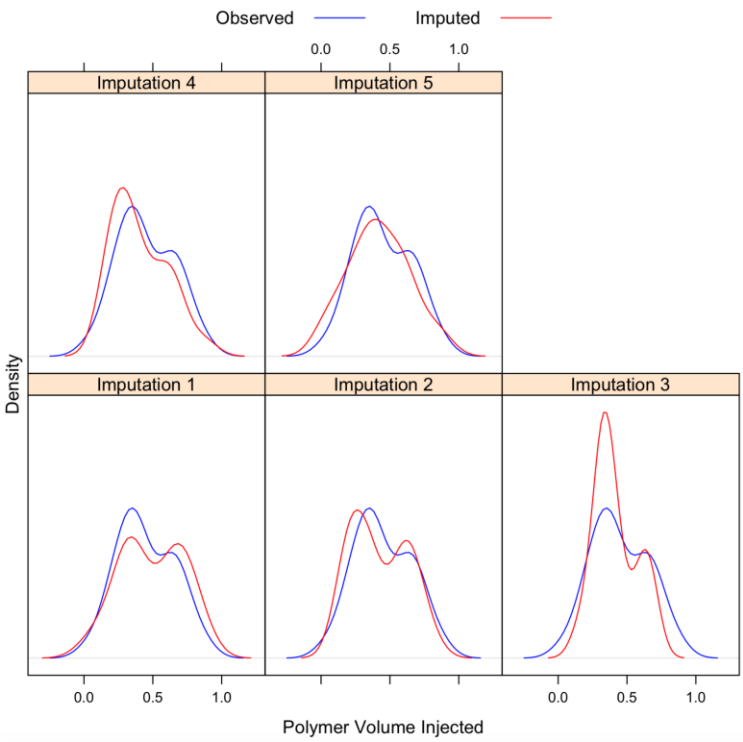

Figure 5.6. Density plot of comparison of observed and imputed data of PV

The general form of the multiple linear regression equation is given by

$$Y_j = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + e_i \quad (10)$$

The bs are the regression coefficients (unknown parameters).

In this study, the dependent Y is oil recovery increased after polymer flooding and independent variables are water salinity (ppm), average permeability (md), dykstra parsons coefficient, polymer concentration (ppm), polymer average molecular weight (10^4), well spacing (m), polymer volume injected (PV), water cut before polymer flooding (%).

The results of multiple linear regression of five imputed datasets are as follow Table 5.1 to Table 5.5 shows.

Table 5.1. Multiple linear regression result of imputation dataset 1

Independents	Estimate(β)	P	Significance
Intercept	-9.154	0.47726	
Water salinity	$-5.763*10^{-5}$	0.00115	**
Permeability	$-2.039*10^{-3}$	0.00305	**
DPC	-13.24	0.000011	***
polymer concentration	$-1.966*10^{-3}$	0.07832	.
polymer molecular weight	$2.537*10^{-3}$	0.00121	**
well spacing	$3.166*10^{-2}$	0.00066	***
PV	3.486	0.04394	*
Water cut before polymer flooding	0.2081	0.05629	.
R-squared	0.7834		

In statistics, the coefficient of determination, also known as R^2, measures the proportion of the total variation in response y is that is explained by a linear model. This value is always between 0 and 1 as a fraction or 0 and 100 as a percent and the most fitting model will be with R^2 equals to one, which means that the predictor's values (x) allow

perfect prediction of response y. Similarly, the adjusted R^2 is an alternative to the R^2. The adjusted R^2 is considered better than R^2 for comparing models. A model that has a good value of adjusted R^2 which is close to one indicates a good fit of the data.

Table 5.2. Multiple linear regression result of imputation dataset 2

Independents	Estimate(β)	P	Significance
Intercept	-22.86	0.05758	.
Water salinity	$4.703*10^{-5}$	0.08252	.
Permeability	$1.220*10^{-3}$	0.20071	
DPC	-4.611	0.27539	
polymer concentration	$2.694*10^{-4}$	0.86957	
polymer molecular weight	$-7.942*10^{-4}$	0.41527	
well spacing	$4.578*10^{-2}$	0.00000167	***
PV	6.165	0.00803	**
Water cut before polymer flooding	0.2435	0.01652	*
R-squared	0.6467		

The p-value for each term tests the null hypothesis that the coefficient is equal to zero, which means no effect. A low p-value (< 0.05) indicates that the null hypothesis can be rejected. In other words, a predictor that has a low p-value indicates this independent variable is a meaningful addition to the predict model because changes in the predictor's

values are related to changes in the response value variable. On the contrary, a larger p-value indicated that the predictor is insignificant, which means that changes in the predictor are not associated with changes in the response.

Table 5.3. Multiple linear regression result of imputation dataset 3

Independents	Estimate(β)	P	Significance
Intercept	20.04	0.01875	*
Water salinity	$-2.981*10^{-5}$	0.19859	
Permeability	$4.229*10^{-3}$	0.00000523	***
DPC	-15.69	0.0000183	***
polymer concentration	$-3.469*10^{-3}$	0.00743	**
polymer molecular weight	$-2.448*10^{-4}$	0.74816	
well spacing	$-2.353*10^{-2}$	0.01088	*
PV	9.327	0.0000352	***
Water cut before polymer flooding	$3.132*10^{-2}$	0.65713	.
		0.6871	

From Table 5.1 to Table 5.5, five multiple linear regression have different results due to their differences in imputation values. Some are good and some are bad even worse. The results of number 1 and 3 seem quite satisfying since the adjusted R^2 of them (74.57% for number 1 and 63.27% for number 3) are much higher than other three results.

Based on five regression analyses above, Rubin (1987) developed a set of rules for combing the separate estimates (intercepts) and standard errors from each of the m (m=5 in this study) imputed datasets into an overall estimate with standard error, confidence intervals and P-values. These rules are based on asymptotic theory on the normal distribution and the pooling process could be implemented in function in MICE.

Table 5.4. Multiple linear regression result of imputation dataset 4

Independents	Estimate(β)	P	Significance
Intercept	22.07	0.13559	
Water salinity	$-4.077*10^{-5}$	0.27779	
Permeability	$2.434*10^{-3}$	0.02262	*
DPC	-8.576	0.00947	**
polymer concentration	$-3.392*10^{-3}$	0.03571	*
polymer molecular weight	$2.169*10^{-5}$	0.98458	
well spacing	$-7.838*10^{-3}$	0.50018	
PV	8.293	0.00361	**
Water cut before polymer flooding	$-7.527*10^{-2}$	0.58518	
R-squared	0.4054		

Table 5.5. Multiple linear regression result of imputation dataset 5

Independents	Estimate(β)	P	Significance
Intercept	-11.19	0.3148	
Water salinity	$5.204*10^{-5}$	0.2342	
Permeability	$1.356*10^{-3}$	0.1988	
DPC	-2.943	0.4262	
polymer concentration	$-8.87*10^{-4}$	0.6558	
polymer molecular weight	$1.363*10^{-3}$	0.3896	
well spacing	$-1.144*10^{-3}$	0.9217	
PV	1.081	0.7048	
Water cut before polymer flooding	0.2124	0.0594	.
R-squared	0.2321		

Table 5.6. shows the pooling results of the five multiple linear regression results.

From the Table 5.6 of pooling results above, it shows that the p-values are very poor that all the p-values for each predictor are over 0.05. The reason is some of the imputaions are not very good imputaions thus the values are not good for regression analysis which in turn causes the bad overall pooling result from five regression results. Otherwise, the pooling result is better than each of those five results if those five results all have good fittings, like have high R^2 s. Although it is not always a good fitting model is with higher R^2, those three results except number 1 and 3 still seem too bad.

Table 5.6. Pooling results for five multiple regression analysis results

Independents	Estimate(β)	P	Significance
Intercept	-0.2181047	0.9934662	
Water salinity	$-5.827571*10^{-6}$	0.9321054	
Permeability	$1.440177*10^{-3}$	0.6323410	
DPC	-9.013298	0.2563224	
polymer concentration	$-1.888974*10^{-3}$	0.4446328	
polymer molecular weight	$5.765398*10^{-4}$	0.7638077	
well spacing	$8.987941*10^{-3}$	0.8092299	
PV	5.670414	0.2563152	
Water cut before polymer flooding	0.1239951	0.5317838	

Based on discussed above, the number 1 regression analysis that has the highest R^2 (74.57%) is chosen to do further analysis. From the number 1 regression results, there are two insignificant independent variables ‘average polymer concentration’ and ‘water cut before polymer flooding’ for their p-values are over 0.05. Thus, another multiple linear regression analysis without these two predictors is encouraged to process.

Table 5.7 shows the result of multiple regression analysis result after removing two insignificant variables ‘average polymer concentration’ and ‘water cut before polymer flooding’. In fact, polymer concentration also plays an important role for polymer flooding.

In this dataset, however, the values are mostly fall into the range of low concentration. Thus, there is a bias occurs which in turn misleads the result of multiple linear regression. The p-values are all less than 0.05 now and the most significant variables are formation water salinity, average permeability, dykstra parsons coefficient and polymer molecular weight. Although the overall fit based on R^2 of the model, which is 67.76%, is lower than the previous fit that is 74.57%, this model is still regarded as the most fitting model for predicting.

Table 5.7. Multiple linear regression result with removed insignificant variables

Independents	Estimate(β)	P	Significance
Intercept	10.36	0.006650	**
Water salinity	$-6.512*10^{-5}$	0.000682	***
Permeability	$-2.634*10^{-3}$	0.000567	***
DPC	-13.73	0.00000291	***
polymer molecular weight	$2.792*10^{-3}$	0.000113	***
well spacing	$1.954*10^{-2}$	0.011549	*
PV	4.401	0.022193	*
R-squared		0.7134	

Thus, this model

$$Y = 10.36 - 0.00006512x_1 - 0.002634x_2 - 13.73x_3 + 0.002792x_4 + 0.01954x_5 + 4.401x_6$$

(11)

where Y is oil recovery increased after polymer flooding (%), x_1 is the water salinity (ppm), x_2 is average permeability (md), x_3 is the dykstra parsons coefficient, x_4 is average molecular weight (10^4), x_5 is the well spacing (m), x_6 is the polymer volume injected is chosen for the final multiple regression analysis predicting model for polymer flooding.

5.4.5. Model Validation. In order to validate the predicting model constructed, a validation process is encouraged to show whether the model is plausible or not. Since the original dataset has no enough data and unable to do linear regression using this model, the number 3 dataset with imputaions is chosen to validate the model.

Table 5.8 shows the comparison of original dataset and predicted dataset of dykstra parsons coefficient versus oil recovery increased. The values in black are the original values of oil recovery increased from number 3 imputed dataset and the red values are the predicted values of oil recovery increased using predicting model built by number 1 imputed dataset using same independent variables of water salinity, average permeability, dykstra parsons coefficient, average molecular weight, well spacing and polymer volume injected.

Figure 5.7 shows a scatterplot of the comparison for original dataset of number 3 and predicted dataset of dykstra parsons coefficient versus oil recovery increased. The black points are the original dykstra parsons coefficient values versus oil recovery increased values in number 3 dataset and the red points are the original dykstra parsons coefficient values newly predicted oil recovery increased values. From Figure 5.7, it can be seen that

the predicted values have similar locations as the original ones. That is to say, the predicted values can also reflect the linear relationship between dykstra parsons coefficient and oil recovery increased just as the original data points do.

Table 5.8. Comparison of original dataset and predicted dataset of dykstra parsons coefficient versus oil recovery increased

original	prediction	original	prediction
13	10	7	8
10	10	7	7
12	9	7	8
12	11	7	5
8	7	5	5
12	14	5	6
13	12	6	14
9	9	3	5
14	12	5	3
20	13	7	11
8	11	3	11
12	9	12	9
9	7	14	6
5	7	10	13
8	11	7	7
10	9	2	4
5	5	14	5
3	2	5	6
11	5	5	7
6	6	3	1
5	3	8	8
3	5	3	5
6	10	11	8
7	14	5	7
5	11	13	15
7	12	5	4
5	6	11	14
6	7		

Based on the validation process stated, the predicted model is good and plausible for predicting oil recovery increased.

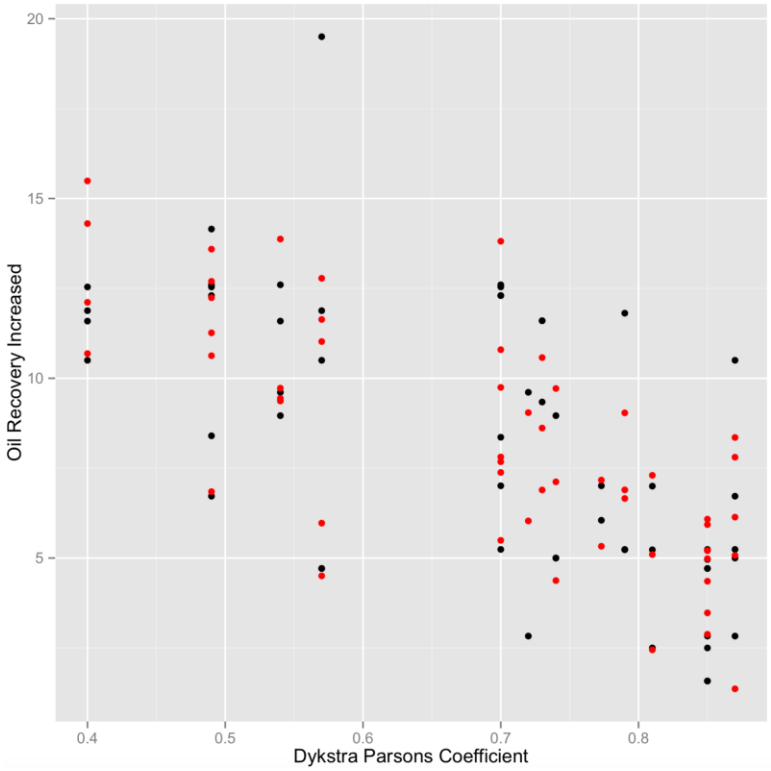

Figure 5.7. Comparison for original dataset and predicted dataset of dykstra parsons coefficient versus oil recovery increased

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. CONCLUSIONS

This study constructed a comprehensive dataset for polymer flooding including pilot and field applications in China. Statistical analysis using histograms, boxplots, violin plots and scatter plots is presented in this study and gave overall updated screening range for polymer flooding. In addition, multiple imputation method is used to impute the missing data after collecting the original data. At last, a predicting model to predict incremental oil recovery based on reservoir properties and polymer properties using multiple linear regression method is constructed.

- 55 polymer flooding including 31 pilot and 24 field projects in China were analyzed based on the reservoir properties, polymer properties, and production data.
- Each parameter was displayed graphically using box plots, frequency histograms to analyze the range of the parameters that influence polymer flooding.
- Screening range for polymer flooding was summarized and updated compared to previous research. New screening range with more parameters for polymer flooding was generated based on statistical analysis.
- Multiple imputation method was introduced and used to impute missing data.
- Multiple linear regression analysis was processed and a model for predicting incremental oil recovery was built based on reservoir and polymer properties.

6.2. RECOMMENDATIONS

- This study only collects 55 pilot and field projects from China only and have missing data problem. Thus, it is necessary and encouraged to enlarge the dataset from other country or sources to have a more satisfying result with more data.
- PMM has been around for a long time (Rubin 1986, Little 1988), but only recently has it become widely available and practical to use. Some organizations have made considerable investments to develop procedures for imputing key variables, like income or family size, whose values are subject to all kinds of subtle constraints. Using one of the built-in imputation methods like PMM could be a waste of this investment, and may fail to produce what is needed. Thus it is somehow inaccurate using this method in oil industry because of the constraints in different specific situations. Therefore, It is possible to write elementary imputation function instead of PMM specifically for oil industry or even an oilfield. This may be an encouraging work in the future.
- Some advanced screening methods like big data methods can be used to build better predicting models instead of numerical analysis.

BIBLIOGRAPHY

Al-Bahar, Mohammad A., et al. "Evaluation of IOR potential within Kuwait." *Abu Dhabi International Conference and Exhibition*. Society of Petroleum Engineers, 2004.

Al Adasani, Ahmad, and Baojun Bai. "Analysis of EOR projects and updated screening criteria." *Journal of Petroleum Science and Engineering* 79.1 (2011): 10-24.

Buckley, Se E., and MCi Leverett. "Mechanism of fluid displacement in sands."*Transactions of the AIME* 146.01 (1942): 107-116.

Brashear, J. P., and V. A. Kuuskraa. "The potential and economics of enhanced oil recovery." *Journal of Petroleum Technology* 30.09 (1978): 1-231.

Bailey, Ralph E., and L. B. Curtis. "Enhanced oil recovery." *National Petroleum Council, Washington, DC* (1984).

Bandyopadhyay, Parag. "Improved Estimation of Bubble Point Pressure of Crude Oils: Modelling by Regression Analysis." *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2011.

Chauveteau, G., and N. Kohler. "Polymer flooding: The essential elements for laboratory evaluation." *SPE Improved Oil Recovery Symposium*. Society of Petroleum Engineers, 1974.

Chang, Harry L. "Polymer flooding technology yesterday, today, and tomorrow."*Journal of Petroleum Technology* 30.08 (1978): 1-113.

Carcoana, Aurel N. "Enhanced oil recovery in Rumania." *SPE Enhanced Oil Recovery Symposium*. Society of Petroleum Engineers, 1982.

DiMarzio, E. A., and C. M. Guttman. "Separation by flow." *Macromolecules* 3.2 (1970): 131-146.

Dickson, Jasper Lane, Alana Leahy-Dios, and Philip L. Wylie. "Development of improved hydrocarbon recovery screening methodologies." *SPE Improved Oil Recovery Symposium*. Society of Petroleum Engineers, 2010.

Goodlett, G. O., et al. "The role of screening and laboratory flow studies in EOR process evaluation." *SPE Rocky Mountain Regional Meeting*. Society of Petroleum Engineers, 1986.

Green, Don W., and G. Paul Willhite. "Enhanced oil recovery; Henry L. Doherty Memorial Fund of AIME." *Society of Petroleum Engineers: Richardson, TX*(1998).

Hawkins, J. M. "Integrated formation evaluation with regression analysis."*Petroleum Computer Conference*. Society of Petroleum Engineers, 1994.

Heitjan, Daniel F., and Roderick JA Little. "Multiple imputation for the fatal accident reporting system." *Applied Statistics* (1991): 13-29.

Horton, Nicholas J., and Stuart R. Lipsitz. "Multiple imputation in practice: comparison of software packages for regression models with missing variables." *The American Statistician* 55.3 (2001): 244-254.

Jablonowski, Christopher J. "Identification of HSE Leading Indicators Using Regression Analysis." *SPE Americas E&P Health, Safety, Security, and Environmental Conference*. Society of Petroleum Engineers, 2011.

Khan, A. M. "An Empirical Approach to Waterflood Predictions." *Journal of Petroleum Technology* 23.05 (1971): 565-573.

Kolodziej, E. J. "Mechanism of microgel formation in xanthan biopolymer solutions." *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 1987.

Little, Roderick JA. and Donald B. Rubin. (1987). "Statistical Analysis withMissing Data."

Lake, Larry W. "Enhanced oil recovery." (1989).

Li, R., and F. Chen. "Reasonable Well Pattern and Spacing for Polymer Flooding in Daqing." *Yearly report* 11 (1995).

Luo, J. H., Y. Z. Liu, and P. Zhu. "Polymer solution properties and displacement mechanisms." *Enhanced oil recovery–polymer flooding. Petroleum Industry Press, Beijing* (2006): 1-72.

Let, Moe Soe. "RN, and Seright, RS 2012. Polymer Flooding a $500-cP Oil."*SPE Improved Oil Recovery Symposium, Tulsa*.

Mackinnon, A. "The use and reporting of multiple imputation in medical research–a review." Journal of internal medicine 268.6 (2010): 586-593.

Maerker, J. M. "Shear degradation of partially hydrolyzed polyacrylamide solutions." *Society of Petroleum Engineers Journal* 15.04 (1975): 311-322.

Nouri, Hossein H., and Paul J. Root. "A Study of Polymer Solution Rheology, Flow Behavior, and Oil Displacement Processes." *Fall Meeting of the Society of Petroleum Engineers of AIME*. Society of Petroleum Engineers, 1971.

NPC. Enhanced Oil Recovery. The National Petroleum Council, Washington, D.C. 1984.

O'Leary, W. B., et al. "Biocide evaluation against sessile xanthan polymer-degrading bacteria." *SPE Reservoir Engineering* 2.04 (1987): 647-652.

Paul, George W., et al. "A simplified predictive model for micellar-polymer flooding." *SPE California Regional Meeting*. Society of Petroleum Engineers, 1982.
Polit, Denise F., and Cheryl Tatano Beck. *Nursing research: Generating and assessing evidence for nursing practice*. Lippincott Williams & Wilkins, 2008.

Rubin, Donald B. *Multiple imputation for nonresponse in surveys*. Vol. 81. John Wiley & Sons, 2004.

Sandiford, B. B. "Laboratory and field studies of water floods using polymer solutions to increase oil recoveries." *Journal of Petroleum Technology* 16.08 (1964): 917-922.

Szabo, M. T. "Some Aspects of Polymer Retention in Porous Media Using a C14 Tagged Polyacrylamide." *Polyelectrolytes and their Applications*. Springer Netherlands, 1975. 287-337.

Stanislav, J. F., and C. S. Kabir. "Polymer Flow Behaviour As Applied To Enhanced Recovery." *Journal of Canadian Petroleum Technology* 16.02 (1977).

Schenker, Nathaniel, and Jeremy MG Taylor. "Partially parametric techniques for multiple imputation." *Computational Statistics & Data Analysis* 22.4 (1996): 425-446.

Schafer, Joseph L. *Analysis of incomplete multivariate data*. CRC press, 1997.

Seright, Randall Scott, James M. Seheult, and Todd Talashek. "Injectivity characteristics of EOR polymers." *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2008.

Seright, Randall. "Potential for polymer flooding reservoirs with viscous oils."*SPE Reservoir Evaluation & Engineering* 13.04 (2010): 730-740.

Saunders, Jeanne A., et al. "Imputing missing data: A comparison of methods for social work researchers." Social work research 30.1 (2006): 19-31.

Sheng, James. *Modern chemical enhanced oil recovery: theory and practice*. Gulf Professional Publishing, 2010.

Sheng, James J., Bernd Leonhardt, and Nasser Azri. "Status of Polymer-Flooding Technology." *Journal of Canadian Petroleum Technology* 54.02 (2015): 116-126.

Sorbie, Kenneth S. *Polymer-improved oil recovery*. Springer Science & Business Media, 2013.

Saleh, Laila, Mingzhen Wei, and Baojun Bai. "Data Analysis and Novel Screening Criteria for Polymer Flooding Based on a Comprehensive Database."*SPE Improved Oil Recovery Symposium*. Society of Petroleum Engineers, 2014a.

Saleh, Laila Dao, Mingzhen Wei, and Baojun Bai. "Data Analysis and Updated Screening Criteria for Polymer Flooding Based on Oilfield Data." *SPE Reservoir Evaluation & Engineering* 17.01 (2014b): 15-25.

Trushenski, Scott P. "Micellar flooding: sulfonate-polymer interaction." *Improved Oil Recovery by Surfactant and Polymer Flooding* (1977): 555-75.

Taber, Joseph John, F. D. Martin, and R. S. Seright. "EOR screening criteria revisited-Part 1: Introduction to screening criteria and enhanced recovery field projects." *SPE Reservoir Engineering* 12.03 (1997): 189-198.

Taber, J. J., F. D. Martin, and R. S. Seright. "EOR screening criteria revisited—Part 2: Applications and impact of oil prices." *SPE Reservoir Engineering* 12.03 (1997): 199-206.

Tzimas, E., et al. "Enhanced Oil Recovery using Carbon Dioxide in the European Energy System." (2005).

Thomas, S. "Enhanced oil recovery-an overview." *Oil & Gas Science and Technology-Revue de l'IFP* 63.1 (2008): 9-19.

Van Horn, L. E. "El Dorado Micellar-Polymer Demonstration Project." *Eighth Annual Report* (1983): 13070-92.

Van Buuren, Stef, Hendriek C. Boshuizen, and Dick L. Knook. "Multiple imputation of missing blood pressure covariates in survival analysis." *Statistics in medicine* 18.6 (1999): 681-694.

Van Ginkel, Joost R. "Investigation of multiple imputation in low-quality questionnaire data." Multivariate Behavioral Research 45.3 (2010): 574-598.

Welge, Henry J. "A simplified method for computing oil recovery by gas or water drive." *Journal of Petroleum Technology* 4.04 (1952): 91-98.

Welge, H. J., et al. "The linear displacement of oil from porous media by enriched gas." *Journal of Petroleum Technology* 13.08 (1961): 787-796.

Willhite, G. P., and J. G. Dominguez. "Mechanisms of polymer retention in porous media." *Improved oil recovery by surfactant and polymer flooding* (1977): 511-553.

Willhite, G. Paul, and Don W. Green. "Enhanced Oil Recovery." *Society of* (1998).

Wang, Demin, et al. "Producing by Polymer Flooding more than 300 Million Barrels of Oil, What Experiences Have Been Learnt?" *SPE Asia Pacific Oil and Gas Conference and Exhibition*. Society of Petroleum Engineers, 2002.

Wang, Dongmei, et al. "Sweep improvement options for the Daqing oil field."*SPE/DOE symposium on improved oil recovery*. Society of Petroleum Engineers, 2006.

Wang, Dongmei, et al. "Review of practical experience by polymer flooding at Daqing." *SPE Reservoir Evaluation & Engineering* 12.03 (2009): 470-476.

Wassmuth, F. R., et al. "Polymer flood technology for heavy oil recovery."*Canadian International Petroleum Conference*. Petroleum Society of Canada, 2007.

Wassmuth, F. R., et al. "Polymer flood application to improve heavy oil recovery at East Bodo." *Journal of Canadian Petroleum Technology* 48.02 (2009): 55-61.

Martin, F. D. "The effect of hyrolysis of polyacrylamide on solution viscosity, polymer retention and flow resistance properties." *SPE Rocky Mountain Regional Meeting*. Society of Petroleum Engineers, 1975.

Niu, Yabin, et al. "Research on Hydrophobically Associating water-soluble polymer used for EOR." *SPE International Symposium on Oilfield Chemistry*. Society of Petroleum Engineers, 2001.

Zaitoun, A. L. A. I. N., et al. "Implementing a Heavy-Oil Horizontal Well Polymer Flood in Western Canada." *Manuscript* 191 (1998): 27-30.

VITA

Yandong Zhang was born on January 26th, 1991 in Karamay, Xinjiang, China. He received his Bachelor of Geological Engineering degree from China University of Petroleum (Beijing) in 2013. After that in August 2013, he started his Master of Science at Missouri University of Science and Technology. He received his Master's degree in Petroleum Engineering from Missouri University of Science and Technology, Rolla, Missouri in December 2015.