APPLICATION OF MACHINE LEARNING AND DATA ANALYTICS METHODS TO

DETECT INTERFERENCE EFFECTS FROM OFFSET WELLS

A Thesis

by

ZINYAT AGHARZAYEVA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Eduardo Gildin
Co-Chair of Committee,	Thomas Blasingame
Committee Member,	Satish Bukkapatnam
Head of Department,	Jeffery Spath

August 2018

Major Subject: Petroleum Engineering

ABSTRACT

The goal of this thesis is to demonstrate that linear-based data-driven models are innovative and robust. They have the potential to forecast well bottom-hole pressure and identify interference effects between wells.

Permanent Downhole Gauges (PDGs) provide a continuous real-time record of pressure and temperature in the downhole environment. These real-time downhole measurements of pressure contain information about the reservoir properties and interactions with offset wells.

This work presents a methodology to reproduce well bottom-hole pressure behavior quickly and to forecast future behavior using those measurements. It also identifies the influence of offset wells based on flowrate-pressure measurements using linear data analysis methods.

In this methodology, we chose linear-based machine learning methods as they are much faster, more robust, and more easily interpreted. Furthermore, we formulate the functional relationship between flowrate and bottom-hole pressure into linear relationships using superposition techniques and physical flow behavior assumptions. Then, without making any further physical assumptions, we regulate process into two stages — training and testing. Training is the regression phase where the flowrates and pressures are correlated using linear machine learning algorithms. Testing is the extrapolation, or forecasting, of the training model to predict well pressure behavior based on a flowrate history.

First, to identify offset well interference effects for a selected well, we reproduce the well's bottom-hole pressure response using only flowrate and time data for that well. Subsequently, we test the influence of offset wells on the selected well's bottom-hole pressure response by considering the selected well and offset well's flowrate history one at a time, until we have examined all possible offset wells. By systematically studying the effects of offset wells on the selected well's bottom-hole pressure, we are able to determine the interference of offset wells using only flowrate histories for the considered wells.

We validate the methodology by using a synthetic reservoir model whose behavior (connectivity) is known. We reproduce and forecast the pressure behavior of a selected well and determine the influence of offset wells. Then, we compare the identified interference wells

with known answers. We note that there is an agreement between the algorithm's results and synthetic model. Also, we test the methodology on the actual field cases. We observe agreement between identified interference effects from offset wells using linear-based data analytics method and those determined from the interpretation of multi-well tests and dynamic observations.

DEDICATION

This work is dedicated to:

My family — my mother, Guler Agharzayeva, my father, Yusif Agharzayev, my sister,

Nurlana Agharzayeva, and my brother, Vasif Agharzayev, for their unconditional love and

support;

My friends — Seljan Gurbanova and Sebastien Penou for their trust and support;

And all people who supported and helped me during my masters pursuit years.

''A dream doesn't become a reality through magic;

It takes sweat, determination, and hard work.''

- Colin Powell

iv

ACKNOWLEDGEMENTS

CONTRIBUTORS AND FUNDING SOURCES

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Real-time downhole pressure and temperature acquisition is vitally crucial for well management and reservoir characterization. It enables the reservoir engineer to monitor both the behavior of offshore wells, where operations are costly, and onshore wells in unconventional reservoirs, where it is complicated to assess/evaluate reservoir performance without reservoir pressure data. In particular, long-term downhole measurements of pressures and temperatures contain information about reservoir properties, architecture, and pressure interactions between individual and groups of wells.

A permanent downhole gauge (PDG) is a pressure and temperature measurement device permanently installed in a well to provide a continuous record of pressures and temperatures in real-time. PDGs began to be installed in wells as early as 1963 (Nestlerode 1963). The installation of PDGs in the North Sea and Saudi Arabia during the 1990s significantly increased their worldwide usage (Kuchuk et al. 2010). Baker et al. (1995) also noted that the reliability, longevity, and accuracy of the pressure gauges were the main reasons for a rapid increase in their installations in the oil industry. In addition to downhole pressure and temperature measurement systems, there are also PDG systems that record downhole flowrate. However, these systems are quite expensive, and therefore most production flowrates continue to be measured only at the surface facilities.

Data from PDGs are generally used to monitor the well performance and to optimize production rates of wells with gas lift or with electrical submersible pump (ESP) systems. Various applications of PDGs in the upstream oil and gas industry are listed below (Frota and Destro 2006), (Oliveira and Kato 2004), (Zheng and Li 2007), (Kuchuk et al. 2010):

- Reservoir Pressure Monitoring — Continuous monitoring of the reservoir pressure can help minimize subsidence and sand production, delay gas or water break-through, and contribute to avoiding producing below a bubble point (or dew point) in a reservoir.
- Well Production Optimization — Real-time downhole measurements can help to evaluate well completion performance, analyze and optimize gas lift behavior, and monitor and optimize electrical submersible pump (ESP) behavior. In addition,

hydraulic fracturing operations can be monitored to evaluate and optimize the treatment process.

- Reservoir Performance — Continuous high-frequency pressure measurements coupled with flow-rate and temperature data can be used to estimate the depletion performance of a given well at any time. This analysis process is known as "Rate Transient Analysis" (RTA) and uses the entire production history acquired by the gauge. Accurate measurement of pressure at bottom-hole conditions provides a mechanism for analysis and interpretation of reservoir performance behavior and can yield reservoir properties, completion properties, and estimated ultimate recovery (EUR).
- Pressure Transient Analysis — A pressure transient "event" is generally defined by an intentional or unintentional shut-in of production at a given well. The corresponding analysis/interpretation of these data can yield reservoir properties, as well as evidence of offset well interference. There are also specialized scenarios where production from one or more offset wells is systematically turned off or turned on in order to observe the effect of these actions at offset wells, which is known as a well interference test.

Many PDG systems are set to acquire pressure and temperature measurements every second, but this sampling is often reduced to a point every 5-10 seconds or even every minute for data archiving and retrieval. In short, the data volume captured by PDG systems can be enormous (*e.g.*, 10s of millions of data points per year) (Horne 2007). Such large amounts of data represent challenges in storage capacity and data retrieval infrastructure. Another challenge with data captured by PDG systems is the minor fluctuations, which can create noise and reduce the accuracy of data analysis.

In addition, PDG systems are just like any other measurement device, and can experience system malfunctions such as gauge failures, broken connections or interfaces, manifold failures, and umbilical failures. Frota and Destro (2006) evaluated five years of reliability data for a given PDG system and concluded that the reliability of PDGs in each case could vary between 50% and 70% depending on the type of gauges.

As a final comment, data processing issues with so many data points are not trivial — as an example, the workflow proposed by Athichanagorn et al. (2002) provides guidance on

preprocessing of PDG data so that analyses and interpretations can be effectively performed on a reasonable and representative dataset.

1.1 Problem Statement

Data-Driven Pressure Forecast

In the petroleum industry, it is vitally essential to forecast future behavior of a reservoir and its penetrating wells to be able to make timely management decisions. The traditional approach for forecasting reservoir or well pressure behaviors is first to construct a static geological model, then subsequently create a dynamic reservoir model that can be used to perform history-matching comply with the production and pressure histories. After history-matching, a forecast of future reservoir and well performance can be generated.

It is important to note that the generated reservoir model is an approximation of reality and may not capture reality in its full extent. There are inevitable errors and wrong approximations during the constructing of the model. So, there are high uncertainties in model parameters, and those uncertainties are almost always underestimated. Besides, during history-matching, many uncertain parameters of a numerical reservoir model are calibrated to obtain an acceptable match between simulated and historically measured production data (Oliver et al. 2008). As history-matching is an inverse problem, the calibration of a model suffers from non-uniqueness. Another disadvantage of the traditional approach is that it is time consuming and in cases where fast decision making is required, traditional methods are not suitable.

One of the targets of this research is to create a data-driven model which will accurately and quickly reproduce and forecast the well pressure performance based on pressure-rate measurements. This model is useful in the cases where fast pressure forecast is needed to make decisions about reservoir and well management. This approach does not require a conventional numerical reservoir simulation model or analytical reservoir model(s). Instead it relies on a data driven approach, whereby the model will be trained with a historical dataset and the algorithm will try to capture or "learn" reservoir/well behavior for a selected well. During the testing process, the algorithm will forecast bottom-hole pressure from the obtained relationship.

We observe that the trained data-driven model can accurately predict the bottom-hole pressure for an arbitrary sequence of future flowrates.

First, we reproduce bottom-hole pressure response using only the flowrate history of the selected well. By doing so, we exclude the interference effects from offset wells. In order to add interference effects and accurately reproduce bottom-pressure responses, we systematically add the offset wells to the linear based machine learning algorithm, which leads us to the next phase of this study.

Multi-Well Tests — Estimation of Inter-Well Communication/Well Interference Effects

From the reservoir engineering point of view, pressure communication between wells provides essential information about reservoir characteristics and heterogeneities. Identifying offset well interference effects is particularly significant in areas where high geological uncertainties exist. As such, reservoir pressure data can be useful in determining the conductivity of faults, the conductivity of the natural fractures, barriers to flow, and heterogeneities — all of which improve our understanding of reservoir performance behavior and can aid in determining the well spacing and the placement of future wells.

Historically, interference and pulse tests have been used to establish and evaluate communication between wells. Interference and pulse tests are conducted by production or injection from an active well while monitoring pressure response from an observation well which is generally shut-in.

Interpretation of interference tests can be challenging as the magnitude of the pressure response in the observation well(s) may be quite small (Kamal 1983). Besides, different wells can potentially communicate with a single observation well during multi-well tests, leading to significant challenges in determining which offset well dominates the pressure response at the observation well. Moreover, performing interference and pulse tests can be expensive because the tests require shutting in the observation well for the duration of the test.

This research also seeks to determine the influence of offset wells on an observation well using data-driven linear models without performing traditional interference or pulse tests. This

methodology requires continuous flowrate and bottom-hole pressure measurements at the well of interest and the possible offset wells.

First, the data-driven linear model reproduces the selected well's bottom-hole pressure behavior by using its flowrate history. If one or more offset wells influence the well's response, some of the pressure trends will come from flowrate histories of offset wells. These pressure trends will not be reproduced by the algorithm when we did not consider the influence of interference wells to predict pressure behavior. Then, we add flowrate histories of possible offset wells one by one to more accurately reproduce bottom-hole pressure measurements. Communicating offset wells are then determined by the quality of the match of the pressure trend that results from taking in to account the offset wells.

1.2 Primary Objectives

The primary objectives of this work are:

- To create a practical and robust workflow to process pressure and flowrate data from Permanent Downhole Gauges (PDGs). This workflow will include: data cleaning, data pre-processing, addressing missing data, and applying the concepts of linearity (*i.e.*, superposition).
- To forecast <u>future</u> bottom-hole pressure response of a well, in a time efficient way, based on pressure-rate measurements using linear machine learning algorithms. First, the algorithm will accurately reproduce the existing bottom-hole pressure data by learning reservoir/well behavior using flowrate and bottom-hole pressure measurements. Then, it will forecast bottom-hole pressure responses for given flowrates.
- To create a methodology using the linear machine learning techniques to identify interference effects of offset wells using the flowrate and bottom-hole pressure histories at the well of interest, as well as the production and injection histories of the offset wells.

1.3 Organization of Thesis

In the second chapter, we present the literature review about used techniques for this study. First, we present multi-well tests which are used to estimate well connectivity among wells. After, we explain data analytics methods which we use for this work.

The third chapter demonstrates steps of the methodology to identify interference effects from offset wells and forecast bottom-hole pressure behavior with linear based machine learning algorithms using only flowrate and bottom-hole pressure measurements.

Forth chapter describes two fields which we use to evaluate the proposed methodology in this work. First, we describe the Guler field which is the synthetic field. We use synthetic case to validate the results of the methodology as we know connectivity between wells from the simulation model. Then, Yusif field is described which is used to evaluate the applicability of the methodology to the field data.

In the fifth chapter, we show applications of the methodology to synthetic and field data. In this chapter, we explain how interference effects from offset wells can be identified based on flowrate and pressure measurements. After we compare identified interference effects by the methodology to known interferences.

The six chapter provides summary, conclusions and recommendations for this work.

LITERATURE REVIEW

In this chapter, we present a survey of approaches to the estimation of well interference effects in the petroleum engineering literature and an introduction to data analytic methods.

To start with, we will summarize traditional and novel multi-well tests approaches. Next, we will explain statistical methods which have been examined in this study.

2.1 Multi-Well Tests

A multi-well transient test has been used to establish and evaluate communication between wells. In a multi-well test, flowrate is changed in at least one active (producing or injecting) well, and the pressure response is measured in an observation well. The traditional examples of multi-well tests are interference and pulse tests (Earlougher 1977).

Interference tests are conducted by long-duration production or injection from an active well while monitoring pressure from an observation well. The observation well is generally shut-in because of practical considerations. **Figure 1** shows a schematic illustration of flowrate at the active well and bottom-hole pressure responses at the active and observation wells. During the interference test, the active well is producing until t_1 time after which flow rate set to zero. From the pressure responses, we see that after the shut-in of active well, the observation well's bottom-hole pressure increases. There is a time lag between the end flowrate at the active well and the pressure response observed at the observation well. One can conclude that there is pressure communication between observation and active wells using interference tests.
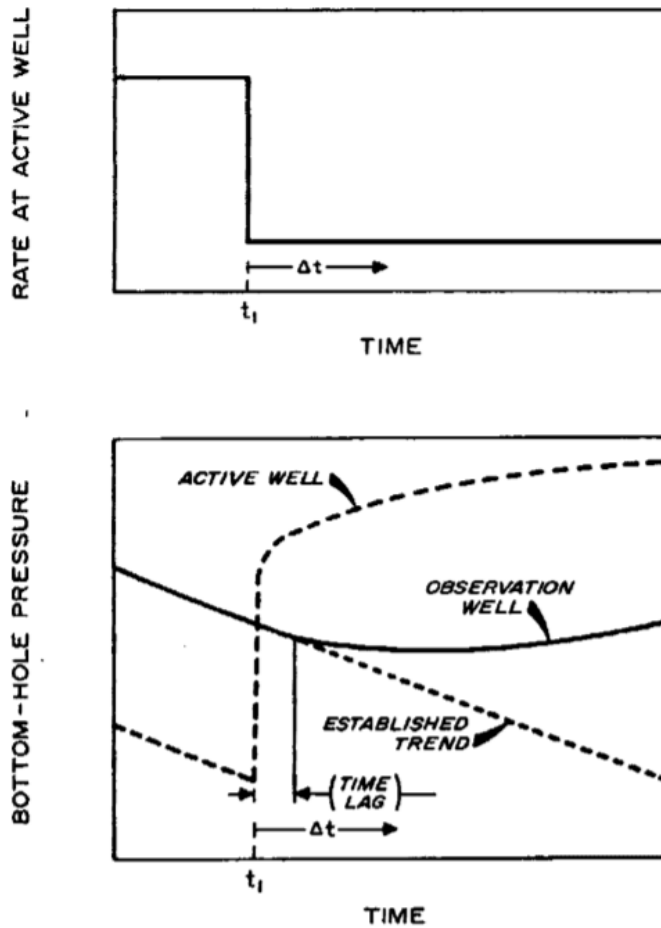
Figure 1 — Schematic illustration of rate at the active well and bottom-hole pressure responses at the active and observation wells for an interference test (Earlougher 1977).

Pulse tests are a special form of multi-well testing, which was introduced in 1966 by Johnson, Greenkorn, and Woods (Kamal 1983). They have the same objectives as interference tests — establishing pressure communication between wells and estimating reservoir properties. For pulse tests, a controlled sequence of short-term production or injection pulses is used at the active (production) well. Generally, pulses are periods of production and shut-in at the active well. If communication exists between wells, then these "pulses" will yield cyclical patterns in the pressure response at the observation well. Because of the tight control required for pulse tests, data noise is generally less than for interference tests. However, short duration rates lead to smaller pressure responses, sometimes less than 0.01 psi (Earlougher 1977).

8

Figure 2 illustrates rate history at the active well and bottom-hole pressure response at the observation well for "pulses". The lower portion of the figure illustrates pressure pulse responses at the observation well.
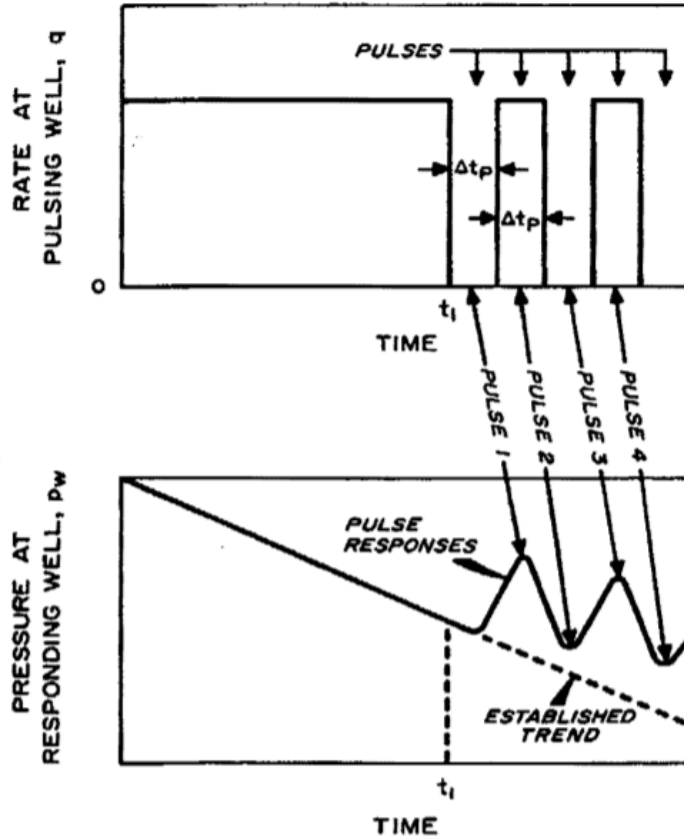
Figure 2 — Schematic illustration of rate (pulse) history at the active well and bottom-hole pressure responses at the observation well for a pulse test (Earlougher 1977).

Besides traditional interference and pulse tests, Tian (2015) performed a multi-well test using machine learning techniques. Tian (2015) used a feature-coefficient target model with mathematically transformed flowrate histories. **Eq. 1** shows the features to which every flowrate history of all wells was converted. $q^{(i)}$ and $t^{(i)}$ are the flowrate and time values at the time i-th data point, and n is the number of observations. The goal was to reproduce pressure response using these features with linear-based algorithms. **Eq. 2** represents linear-coefficient target model where $x^{(i)}$ is features, $y^{(i)}$ is pressure response and w is selected well. The coefficients (θ_w^T) were calculated by minimizing the mean-square error which will be explained

9

in the following parts of this chapter. It was assumed that computed coefficients would contain the information about reservoir behavior and well interferences.

$$x^{(i)} = \begin{bmatrix} \sum_{j=1}^{i-1}(q^j - q^{j-1}) \\ \sum_{j=1}^{i-1}(q^j - q^{j-1})\log(t^i - t^j) \\ \sum_{j=1}^{i-1}(q^j - q^{j-1})(t^i - t^j) \\ \sum_{j=1}^{i-1}\frac{(q^j - q^{j-1})}{(t^i - t^j)} \end{bmatrix} . \quad i = 1, \cdots, n \tag{1}$$

$$y_w^{(i)} = \theta_w^T x^{(i)}, \quad w = 1, \cdots, N, \quad i = 1, \cdots, n \tag{2}$$

The Tian (2015) case examined a homogeneous reservoir with two wells. The flowrate histories of both wells added as an input to the algorithm. The intuition was that as model coefficients contain information about the well interactions, statistical parameters of coefficients should reveal interference wells during the algorithm training. Tian illustrated an example, which showed that the algorithm predictions became more accurate by including the interference well's flowrate history.

2.2 Data Analytics and Machine Learning Techniques

Data analytics methodologies seek to extract patterns or signatures in a dataset as well as to capture the relationship between input and output. The input, which can also be named as predictors, features, variables or independent variables, are denoted using the symbol X. The output can be called response or dependent variables which in this study are indicated by symbol Y. It is assumed that there is some unknown relationship (f) between quantitative response and predictors, which can be generalized in **Eq. 3**.

$$Y = f(X) + \epsilon, \tag{3}$$

10

where f is a true relationship between response and predictor and ϵ is a random error term with zero mean. In statistical learning, we are seeking to determine f, true and unknown relationships between predictor and response. To do that, data analytics tries to estimate the true relationship using different data analytics techniques. In **Eq. 4**, $\hat{f}$ is the estimated relationship or data analytics technique and $\hat{Y}$ is the estimated response:

$$\hat{Y} = \hat{f}(X). \qquad\qquad (\ 4\ )$$

Machine learning is a small portion of the rich area of the data analytics. Over the last four decades, machine learning has grown from rule-based coding, which was part of computer science, to the discipline of algorithms which can learn without being explicitly programmed. Combining computer science and statistics gives an opportunity to create self-programed algorithms. Statistics focus on finding patterns from data, and the computer science discipline tries to find optimal computational algorithms to most effectively capture, store, index, retrieve and merge these data (Mitchell 1997). In machine learning, computer program learns from experience 'E' to perform a particular task 'T' with performance 'P'. The system reliability improves its performance 'P' at task 'T' with following experience 'E' (Mitchell 1997).

There are four types of machine learning techniques as illustrated in **Figure 3** and the ones used most frequently are supervised and unsupervised learning methods.
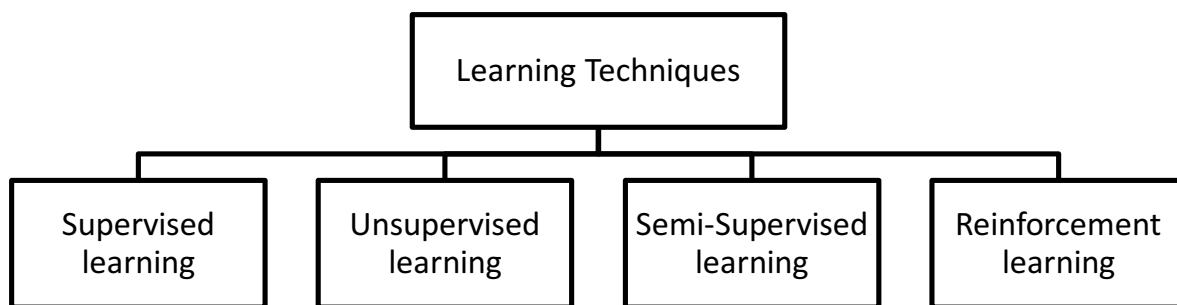
Figure 3 — Different machine learning techniques.

In supervised learning, for each observation of the predictor measurements, there is an associated response measurement (James 2013). By contrast, in unsupervised learning for each observation, there is no measured response for measured predictors. Cases which fall between supervised learning and unsupervised learning are called semi-supervised learning where not all measured predictors have measured responses.

There are two types of statistical learning methods: parametric methods and non-parametric methods. Parametric methods make assumptions about the structure of the model or the functional form of f which is a true relationship between response and predictor. Non-parametric methods do not make any explicit assumptions about the functional form of f. Although non-parametric methods have more flexibility and no bias included to force a model structure, they require a significant amount of data and have less interpretability. In comparison, parametric methods introduce bias by making assumptions about the model structure but do not need a large number of observations and have higher interpretability.

Variables can be characterized as quantitative or qualitative variables. Quantitative variables are taking numerical values while qualitative variables are taking categorical values. If a response is qualitative, it is called a regression problem; if a response is quantitative, it is called a classification problem. The focus of this study is regression problems.

The primary purpose of the data analytics model is not to find the best fit for all data points with which models are trained, but rather to forecast outputs for unseen and untrained data correctly. To validate model accuracy, the data is divided into two parts: training and testing.

First, the model is run with training data to best capture the relationship between inputs and output. During the training process, the average error that results from using a data analytic method to make predictions is called training error.

After the training process, the captured relationship in data is tested with the testing data set to validate the prediction accuracy. The testing error is the error which is computed using data analytics methods to forecast unseen data.

There are two primary approaches to estimating testing and training errors. The first approach is the validation set approach where we divide each data set into training and testing data sets

and then calculate error for each set. Another approach is k-fold cross-validation. In k-fold cross validation, the original data sample is partitioned into *k* equal size subsamples. From the *k* samples, a single subsample is chosen as the testing set and then *k* 1 subsamples become the training set. The train and testing errors are calculated for each data set. This process is repeated *k* times, each time with a different testing set. Final training and testing errors are estimated by averaging *k* number of training and testing errors respectively.

As mentioned, the primary interest is to find the best approximation for the unseen data. The question is how one can find a data analytics method which approximates the real relationship between data and predicts the best fit for a testing data set. The following chapters attempt attempts to address this question.

In **Figure 4.a**, the black dots are the real observations and the black line is the true relationship between x and Y. Orange, blue and green curves illustrate three different data analytical methods with an increasing flexibility level. The flexibility of the model is the ability of the model to capture complex relationships in data. The higher the flexibility of the model, the more complex relationship it can capture.

The orange line is a linear regression fit, which has the least amount of flexibility. If true relationships between x and Y are not linear, the linear regression is not able to approximate the true relationship well. The green curve has the highest flexibility and matches observed points well. However, the green line also poorly captures the true relationship of the black curve as it is too wiggly. The closest to the true relationship is the blue curve which has an intermediate level of flexibility for the given data.

Figure 4.b demonstrates training and testing error changes with flexibility. The orange, blue and green dots are representations of the curves in Figure 4.a so the orange dot has the lowest amount of flexibility and the green dot has the highest flexibility. The training error always decreases with increasing flexibility as the training data better fits the model. But the increase in flexibility of the model does not guarantee a rise in the prediction of accuracy of the testing data. The highly flexible models can cause the overfitting of data. The overfitting is observed when the training error decreases while the testing error increases. Therefore, the highly flexible models are able to capture more complex boundary relationships and fit training data more

accurately. However, it does not mean that the higher the flexibility of the model the better it will perform. The nature of the data needs to be considered for choosing the flexibility of the model (James 2013).
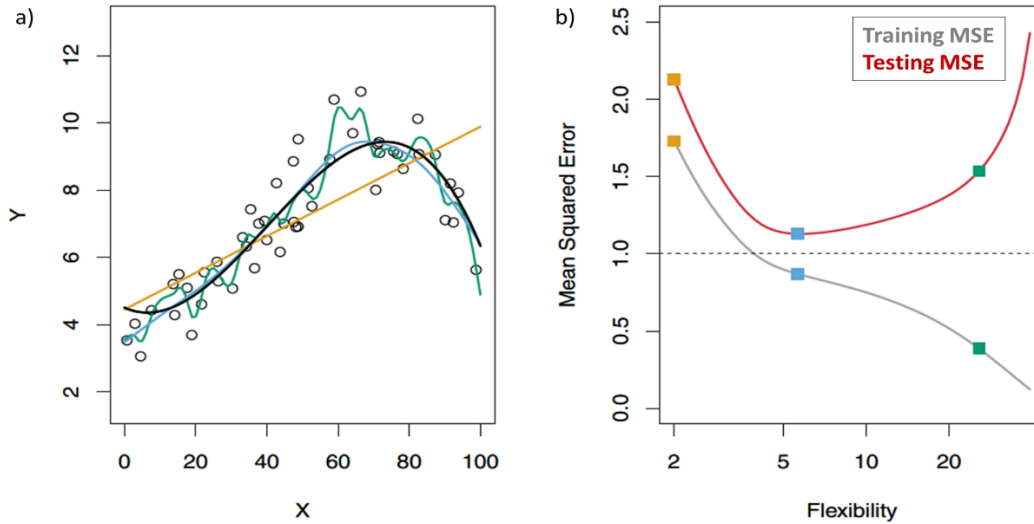
Figure 4 — Testing and training error change with flexibility (James 2013).

It is a known fact that as the flexibility of the model increases, its interpretability decreases, so the inverse relationship between flexibility and interpretability needs to be considered as well while choosing the model. **Figure 5** demonstrates the tradeoff between interpretability and flexibility for different statistical learning methods.
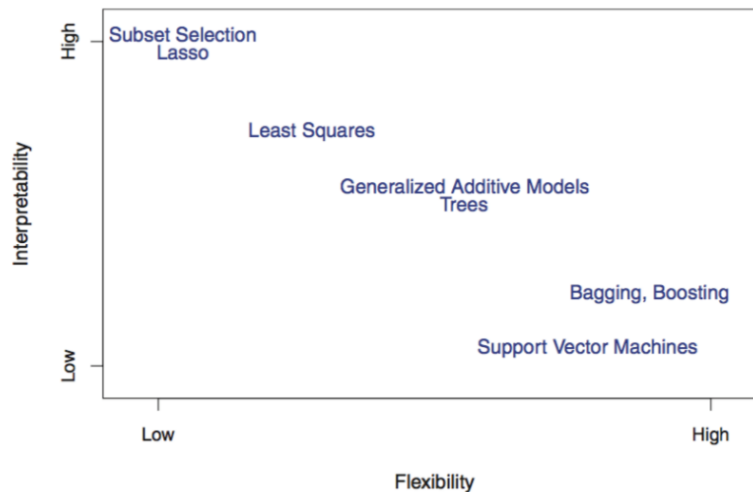
Figure 5 — Tradeoff between interpretability and flexibility for different statistical learning methods (James 2013).

14

After choosing the data analytics method, next step is to calculate accuracy of the model. The difference between predicted response ($\hat{Y}$) and real response (Y) defines the accuracy of the model. $\hat{Y}$ is never equal to Y because of two quantities: reducible error and irreducible error.

$$E\left[(Y - \hat{Y})^2\right] = E\left[(f(X) + \epsilon - \hat{f}(X))^2\right] \qquad (5)$$
$$= var\left[\hat{f}(X)\right] + Bias\left[\hat{f}\right]^2 + var(\epsilon).$$

In **Eq. 5**, $var\left[\hat{f}(X)\right]$ represents the variance error which is a change in prediction accuracy by changing the training set. $Bias\left[\hat{f}\right]^2$ shows bias error which measures the deviation of the predicted values from the real data. $var(\epsilon)$ is the irreducible error from pure noise which cannot be reduced. Bias and variance are called reducible errors as they can be changed by changing the flexibility of the model. With increasing flexibility of the model, bias decreases while variance increases. So, the tradeoff between bias and variance needs to be considered to decrease the overall error and improve the model's accuracy.

There are some standardized measurements to assess the accuracy of a model. All those measurements can be done for both training and testing data. It is worth noting that the objective of data analytical methods is to increase the accuracy of testing predictions rather than with overfitting data and getting a perfect fit for training data. In regression setting, the most commonly used measurement is mean square error *(MSE):*

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2, \qquad (6)$$

where y_i is measured data, $\hat{f}(x_i)$ is a prediction for the i-th observation. If true response is close to predictive response, then *MSE* is going to be small. If the responses are substantially different from each other, *MSE* is going to be high. As model complexity and flexibility increase, *MSE* for training data tends to decrease but testing *MSE* may not. As mentioned above, the overfitting is taking place when with increased flexibility of model training *MSE* decreases but testing MSE increases.

Residual sum square *(RSS)* measures the amount of variability that is left unexplained after performing the data analytical methods. A total sum of square *(TSS)* identifies the total variance in the response before the model was implemented:

$$RSS = \sum_{i=1}^{n} \left(y_i - \hat{f}(x_i)\right)^2,$$ (7)

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2,$$ (8)

where *n* is number of observations and $\bar{y}$ is average value of responses.

R^2 is another alternative measurement of fitting:

$$R^2 = 1 - \frac{RSS}{TSS}.$$ (9)

R^2 measures proportion of variability in response which can be explained by predictors. It is in the form of a fraction which makes it easy to interpret compare to *MSE*. R^2 equal to 1 indicates that the variation in response is fully explained by predictors with the used technique.

In data analysis, there are three main steps which are illustrated in **Figure 6**:

- Preparing Data
- Conducting Data Analysis
- Making Decisions
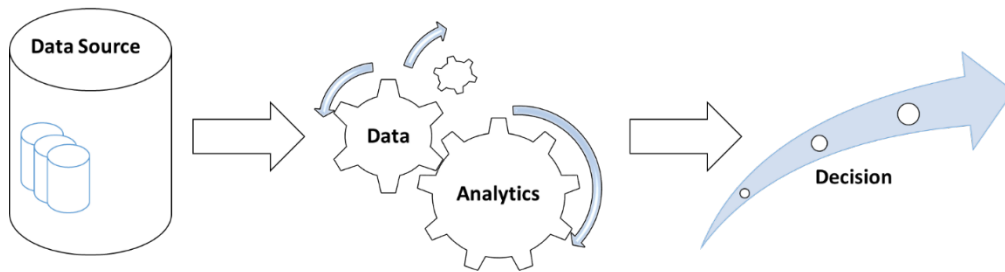
Figure 6 — Steps for data analytical methods. First step is gathering data after which data needs to be cleaned and reprocessed. Next step require conducing the data analysis using algorithms. After analysis, decisions can be takes.

2.3 Linear Regression

Linear regression is a simple approach which is widely used in statistical methods for predicting a quantitative response. As can be understood from the name, linear regression assumes that the real relationship between predictors and response is linear as shown in **Eq. 10**.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon. \qquad (10)$$

$$\text{Let} \quad \hat{Y} = \begin{bmatrix} \widehat{y_1} \\ \widehat{y_2} \\ \vdots \\ \widehat{y_n} \end{bmatrix}, \; X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \; \hat{\beta} = \begin{bmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \\ \vdots \\ \widehat{\beta_n} \end{bmatrix}, \; e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \qquad (11)$$

then a mathematical representation of linear regression model is:

$$\hat{Y} = X\hat{\beta}. \qquad (12)$$

In **Eq. 11** n is the number of observations while p is the number of features in a linear regression model. $\widehat{y}$ indicates a response on the bases of $x_{11}, x_{12} \cdots x_{1p}$ which are predictors. $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3 \cdots \hat{\beta}_p$ are called coefficients or parameters. e is called residual which is the difference between real observation and response from the model.

In linear regression, parameters are estimated using ordinary least square (OLS) approach which chooses a coefficient to minimize the sum of squared residuals (RSS), subject to having at least 's' non-zero coefficients as shown in **Eq. 13**. Linear regression is the linear model, but non-linearity can be added by adding non-linear and interaction terms as features.

$$minimize \left\{ \sum_{(i=1)}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_i x_{ij} \right)^2 \right\}, \quad \sum_{j=1}^{p} I(\beta_j \neq 0) \leq s \qquad (13)$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad (14)$$

There are statistical parameters (standard error, t-statistics, and p-value) which quantify both the extent to which the model fits data and the importance of features. The standard error is the average amount that the response deviates from the true regression line (James 2013). James goes on to state that T-statistics and a p-value provide information about whether each feature

17

is related to the response. He declares that there is a relationship between every feature and every response if the p-value is small enough and t-statistics are larger than 1. Typical p-value cutoffs are between 5% and 1%.

The question is if p-values are sufficient to determine an association between features and response. James (2013) considered a case where there are hundreds of features (*X*) and none of them has an association with response (*Y*). James noted that in this situation, about 5% of the p-values will be below 0.05 by chance. Therefore, he concluded that there is a 5% chance that a p-value will incorrectly determine that there is a relationship.

2.4 Shrinkage Methods

There two main criteria to evaluate the quality of the model:

- Accuracy of prediction of the testing data
- Interpretability of the model

It is well known that ordinary least square (OLS) method can perform poorly for both of these criteria. Penalizing techniques have been proposed to improve OLS. Shrinkage methods introduced new approaches to OLS to increase accuracy and interpretability of regression models. Ridge regression, lasso, and elastic net are the best-known shrinkage techniques for linear regression. These methods improve the accuracy of linear regression by significantly reducing the variance. Instead of using the OLS approach, ridge regression and lasso are introduced penalty term to OLS for estimating coefficients in linear regression.

Ridge regression minimizing **Eq. 15** in order to estimate coefficients

$$minimize \left\{ \sum_{(i=1)}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_i x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}. \tag{15}$$

The coefficients are calculated using criteria in **Eq. 16** for lasso method

$$minimize \left\{ \sum_{(i=1)}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_i x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}. \tag{16}$$

18

In Eq. 15 and Eq. 16, λ is a tuning parameter. $\lambda \sum_{j=1}^{p} \beta_j^2$ and $\lambda \sum_{j=1}^{p} |\beta_j|$ are called the shrinkage penalty terms and they have ability to shrink the estimates of coefficients towards or to zero. For the ridge regression and lasso fit, two criterion are considered to determine coefficients. As with ordinary least square, the first criterion is to the fit data well by finding the smallest RSS. The second criterion is minimizing the shrinkage penalty parameter.

Lambda (λ) serves to control coefficient values. When λ equals to zero, the penalty term has no effect and the ridge regression and lasso methods perform as the linear regression model. Hence, the model coefficients are estimated with OLS (James 2013). When $\lambda \rightarrow \infty$, the influence of tuning parameter increases which causes estimated coefficients to approach zero. For each value of λ, ridge regression and lasso estimate different coefficient values. So, selecting a good lambda value is essential.

Generally, lambda value is computed using k-fold cross-validation. The k-fold cross-validation error is calculated for each value of lambda. The selected lambda value is the one which gives the smallest cross-validation error value. **Figure 7** shows cross-validation error (left) and estimated coefficient (right) changes for different lambda values (James 2013). The left part of Figure 7 displays the vertical dash line which indicates the selected lambda values for the smallest cross-validation error. The right part of the figure provides the illustration of coefficients for the selected lambda value.

For the case presented by James, the selected lambda value is less than 5×10^{-1} which is a relatively small value. This indicates that the fit involves a small amount of shrinkage relative to the least square solution (James 2013). As lambda values approach zero, the lasso method performs similarly to the linear regression method. In these types of cases, the lasso model is going to perform as the linear regression model using the simple OLS approach.
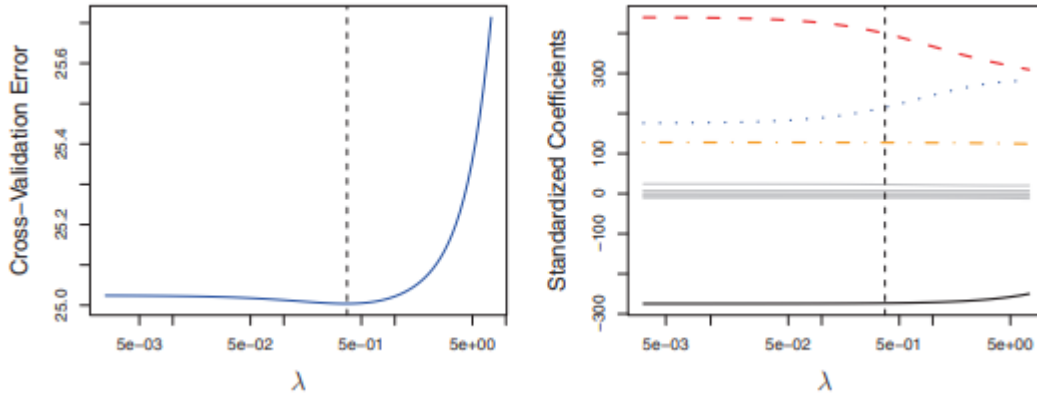
Figure 7 — Cross-validation error and estimated coefficients for different values of lambda. The vertical dashed line indicates the lambda value which gives smallest cross-validation error (James 2013). In the right graph, colored lines indicate features which are statistically important. The gray color represents features which are statistically unimportant so the algorithms force their coefficients toward zero.

When we compare the ridge regression method with the lasso method, ridge regression has one disadvantage. Even if ridge regression approximates coefficients that are close to zero, they are never equal to zero exactly. The lasso method overcomes this problem. Because of its penalty term, the lasso method can force coefficients to zero. The lasso method is used to identify important predictors, reduce the number of predictors in linear regression, and introduce an alternative method to estimate coefficients with increased accuracy and interpretability of the model.

Although the lasso method is a successful method, it has some limitations:

- Presence of group of variables with very high pairwise correlation (Hui Zou and Trevor Hastie 2003)
- Presence of high correlation between predictors when $n>p$ (Tibshirani 1996).

Empirical studies have shown that elastic net can over-perform lasso when the limitations stated above are present in data (https://www.mathworks.com/help/stats/lasso-and-elastic-net.html). Elastic net is a hybrid of ridge regression and lasso. Like lasso method, the elastic net can also reduce coefficients to be precisely zero. **Eq. 17** and **18** demonstrate criteria for the elastic net to estimate coefficients.

20

$$minimize\left\{\frac{1}{2n}\sum_{(i=1)}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_i x_{ij}\right)^2+\lambda\,P_\alpha(\beta)\right\},\qquad(17)$$

$$P_\alpha(\beta)=\frac{1-\alpha}{2}\|\beta\|_2^2+\alpha\|\beta\|_1=\sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2+\alpha|\beta_j|\right).\qquad(18)$$

When $\alpha=1$, the elastic net method is the same as lasso method. When α reduces towards zero, the elastic net method approaches the ridge regression method.

2.5 Tree Based Model

A decision tree is a machine learning technique which learns by a partition of predictor space. The decision tree can be both a regression tree and a classification tree. Building decision tree requires two steps:

- Dividing the predictor space into distinct and non-overlapping regions.
- Assigning the response into each region.

The first step is dividing predictor space into different distinct boxes such that the lost function will be minimized. A recursive binary splitting approach is believed to find the best split. First, one of the predictors is considered (like X_1) and the cut-point ($s_{1)}$ is chosen to achieve highest possible reduction in lost function. In the same manner, repeatedly all the predictors and the cut-points are considered until the stopping criteria is met.

After regions are created, value is assigned from the training observations for each region. Assigned values are the mean values of observation inside of each region for regression tree model. Regression tree assumes a model form as:

$$f(X)=\sum_{m=1}^{N}c_m*1_{X\in R_m},\qquad(19)$$

where m is the number of partitioned.

21

The decision tree has high interpretability; it is both easy to explain, and it can be displayed graphically **(Figure 8)**. It requires little data preparation. Decision trees perform better than linear regression when a relationship between predictors and response are highly nonlinear. However, decision trees suffer from high variance, and they over-fit data. Decision trees are not robust and not as accurate as some advanced data analytical methods.
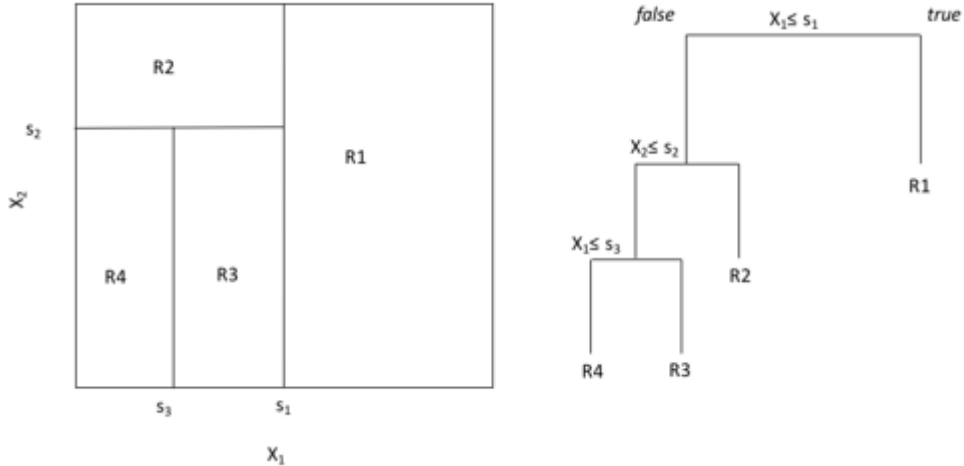
Figure 8   — Division of predictor space for decision tree

These limitations are met with the random forest method. The random forest is an ensemble learning method that constructs multiple decision trees from training data and then outputting mean of responses from individual trees for regression models. The random forest uses bagging which is a combination of bootstrap and aggregating. Bootstrapping is the resampling of the observed dataset with random sampling with replacement. Aggregation is the averaging of the set of observations which reduces variance. The random forest splits each tree with a different number of predictors as illustrated in **Figure 9**. By using the subset of predictors, random forest avoids using the same strong predictors in every tree, and this decreases variance even more. Hence, random forest builds numerous decision trees from bootstrapping training samples and uses a subset of predictors for split candidates.

Random forest is also used to rank the importance of predictors both in regression and classification problems. The first step in measuring the importance of a variable is to fit random forest mode with training data. For the individual feature to calculate the importance, each feature is excluded and new error is calculated without that feature. The importance of the

feature is calculated from the difference of error with and without of that features. The features which have a large increase in error value when they are excluded from predictor space, ranked as the more important predictors.
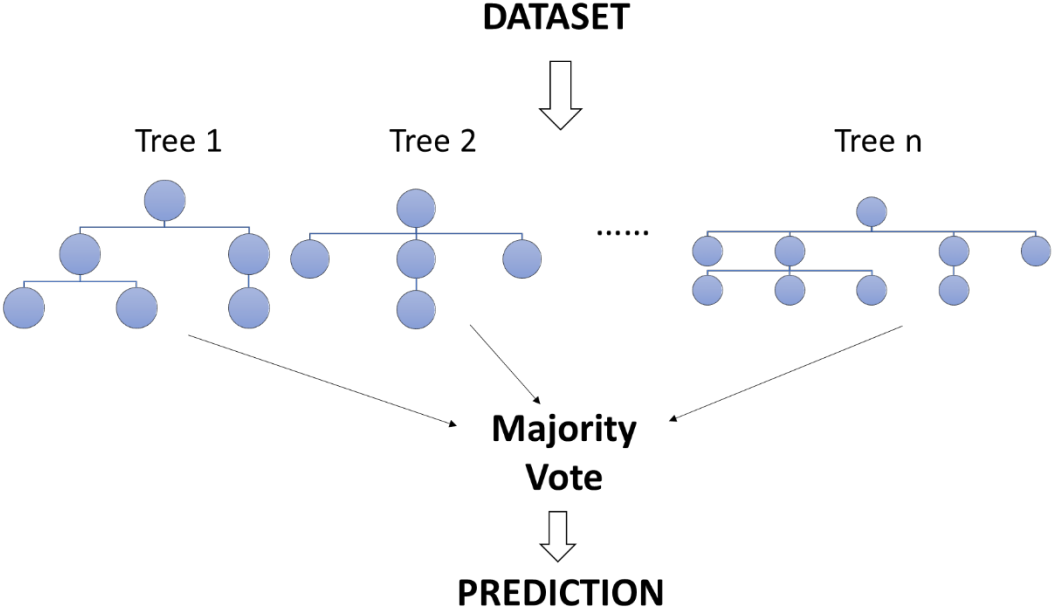
Figure 9 — Random Forest Model.

METHODOLOGY

In this section, we explain the methodology and steps to reproduce bottom-hole pressure responses and identify interference effects from offsets wells using linear machine learning techniques. In this work, we examine linear machine learning techniques such as linear regression, ridge regression, lasso, and elastic net. Among all examined linear machine learning techniques, lasso performs better to reproduce the bottom-hole pressure responses and to identify interference effects. **Figure 10** demonstrates steps of the methodology which will be explained in more detail in this chapter.

Input	Data Cleaning	Data Pre-processing	Single Well Tests	Multi-Well Tests	Identify Interference Wells
• Flowrate • Bottom-hole pressure	• Data Reduction • Outlier Removal • Behavior Filtering • Filling Missing Data	• Linearization relationship between flowrate and pressure	• Pressure prediction without including offset well interference effects using the lasso algorithm	• Estimation of inter-well communication/ well interference effects using the lasso algorithm	

Figure 10 — Steps of the methodology to forecast bottom-hole pressure and identify interference wells using the lasso algorithm.

3.1 Data Processing

Long-term data from the permanent downhole gauges (PDGs) contains noise and outliers. Noise or outliers in downhole measurements can be caused by the permanent gauge data-acquisition system which operates in an uncontrolled and extreme subsurface environment (Veneruso 1992). Another challenging aspect of PDG record comes from dynamic changes in reservoir properties and gauge mechanical deterioration (Horne 2007). As a result, one of the important parts of analyzing PDG data is data cleaning and preparation.

The workflow proposed by Athichanagorn et al. (2002) provides guidance to process PDG data. In the workflow, Athichanagorn et al. pointed out three steps to clean data:

- Data reduction
- Outlier removal

● Behavioral filtering

First, we reduce the number of data to a representative set by taking mean pressure value per day. The next step is outlier removal with threshold approach for which we use Ecrin software by Kappa (Ecrin 2013). First, we remove obvious outliers, and then we remove the less obvious ones by iteratively setting a threshold of delta signal to 20psi, 15psia, 10psia and 5 psi respectively. **Figure 11.a** presents results after eliminating outliers from bottom-hole pressure data. Although the de-noising step was included in the Athichanagorn et al. (2002) workflow, we skip that step as we assume that de-noising can lead to information loss regarding interference effects. In the behavioral filtering step, aberrant transients are removed as suggested in Athichanagorn et al. (2002) methodology.

Missing data poses challenges in accurately determination of the functional relationship between flowrates and bottom-ole pressure. In particular, missing bottom-hole pressure values were common in field data for this work. We use the random forest algorithm to provide estimates for missing bottom-hole pressure data points. Input for estimating pressure data is liquid flowrate history. **Figure 11.b** illustrates predicted bottom-hole pressure values for missing pressure values using the random forest algorithm. The random forest algorithm is found to be effective in predicting short-term missing bottom-hole pressure values, but it is not able to forecast bottom-hole pressure values which are outside of training range.
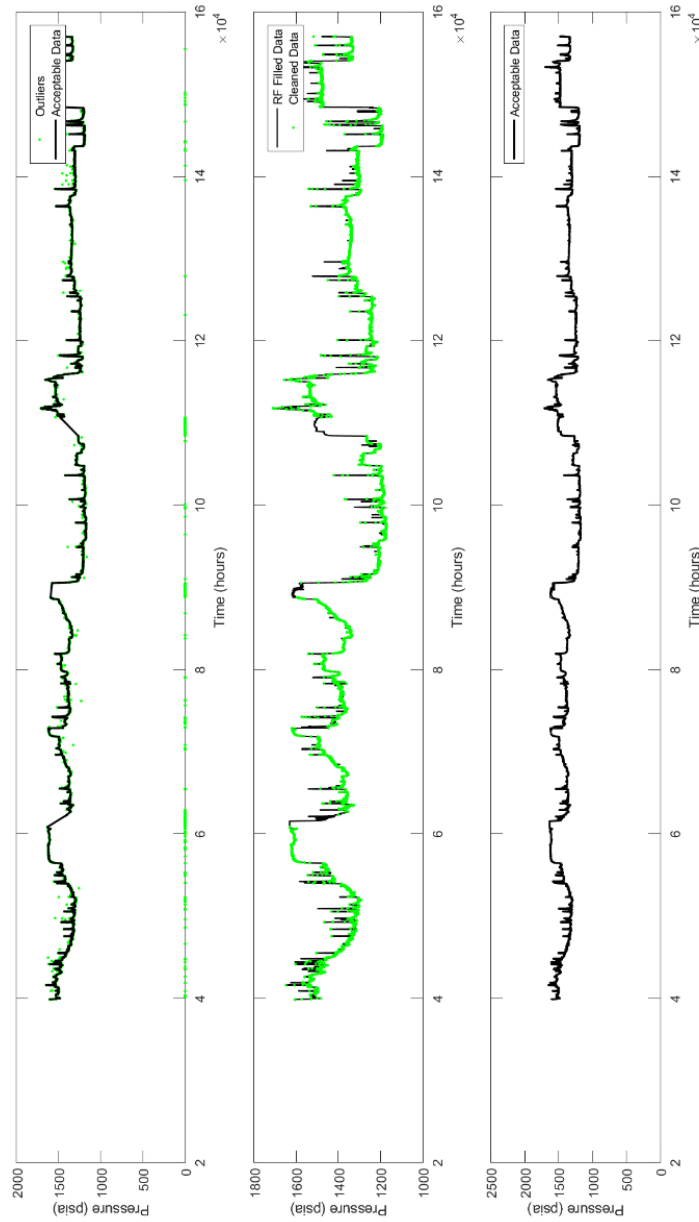
Figure 11 — (a) outlier removed data, (b) missing data filled with Random Forest algorithm, (c) acceptable data for analysis.

3.2 Relationship between Flowrate and Pressure

The focus of this research is to reproduce and forecast bottom-hole pressure behavior and identify the influence of offset wells using linear based data analytics methods. Before performing such analysis, it is vital to understand the physical relationship between flow rate

and pressure. Therefore, in this section, we present equations showing relationships between flowrate and pressure.

Darcy's empirical flow law describes the fluid flow through a porous medium under the influence of a pressure gradient. For a given permeability (k), cross-sectional area (A), length of a core (L) and viscosity (μ) of the fluid, flowrate (q) and delta pressure (Δp) are linearly related:

$$q = \frac{kA}{\mu}\frac{\Delta p}{L}$$

$$(\ 20 \ )$$

Mass conservation principle and Darcy law need to be considered in order to derive the basic differential equation for radial flow in a porous medium, as shown in **Eq. 21** (Dake 1978). Derivation assumptions are stated below:

- Reservoir is a homogeneous reservoir in all rock properties and isotropic with respect to permeability
- Producing well is completed in the entire formation thickness
- Formation is saturated with only one fluid
- Total compressibility of the system is small and constant, permeability and viscosity are constant, $\left(\frac{\partial p}{\partial r}\right)^2$ can be ignored.

$$\frac{\partial^2 p}{\partial r^2} + \frac{1}{r}\frac{\partial p}{\partial r} = \frac{\phi\mu c}{k}\frac{\partial p}{\partial t}.$$

$$(\ 21 \ )$$

Eq. 22 presents the line-source solution for the differential equation:

$$p_i - p_{wf} = \frac{q\,B\,\mu}{k\,h}\,E_i\left(\frac{\phi\mu c_t r^2}{kt}\right),$$

$$(\ 22 \ )$$

where

$$E_i(-x) = -\int_x^\infty \frac{e^{-u}}{u}\,du.$$

$$(\ 23 \ )$$

27

Eq. 24 presents infinity acting radial flow behavior for vertical well:

$$p_i - p_{wf} = \frac{162.6 \, q \, B \, \mu}{k \, h} \left[\log\left(\frac{k \, t}{\phi \mu c_t r_w^2}\right) - 3.23 + 0.869s \right]. \tag{24}$$

Eq. 25 presents the wellbore storage effect for constant flowrate:

$$p_i - p_{wf} = \frac{q \, B \, \Delta t}{24 \, C}, \qquad C = c_{wb} V_{wb}. \tag{25}$$

We will use the physical relationship presented in Eq. 22, 24 and 25 to mathematically linearize relationship between flowrate and pressure, which will be presented in the next section.

3.3 Creating a Linear Relationship between Pressure and Flowrate

Liu (2013) and Tian (2014) used data mining and data analytics methods for reservoir analysis. Their research showed that nonlinear relationships between flow rate and pressure could be formulated as linear relationships between features corresponding reservoir behavior and pressure. Then, those features can effectively be used as inputs for linear machine learning techniques.

Liu (2013) tried to match the reservoir model using the convolution kernel method (Laskov et al. 2012, Collins et al. 2002) which is a data mining method. Liu (2013) used discrete data as a basis, rather than the traditional approach of a numerical or analytical model. Two deliverables were generated in this study:

- Detection of the reservoir pressure signature from noisy PDG data.
- Forecasting of downhole pressure behavior for arbitrary flowrate based on a correlation of time, flowrate, and pressure data (i.e., a training set).

To achieve these deliverables, Liu (2013) used a non-parametric data mining method that does not require an *a priori* model, and the algorithm learns from the given data during the training

process. Features of the convolution kernel base algorithm were constructed considering the use of superposition. As can be seen in **Figure 12** (left) (after Liu 2013), if we have two independent cases of flowrates q_1 and q_2, then we will have the corresponding pressure behaviors shown by p_1 and p_2 respectively. In this schematic example, q_1 represents a constant production rate, and q_2 represents a constant injection rate. When flowrate is positive, pressure decreases — and when flowrate is negative, pressure increases. Assuming a linear combination of q_1 and q_2 as shown in **Figure 12** (right), then the pressure response also will be combined (or "convolved") — specifically, the pressure will first decrease as this represents production and will be followed by an increase in pressure as this represents injection. For this case, we have the particular case where the injection cancels the production and we have a zero flowrate (shut-in).

To use convolution as part of the convolution kernel algorithm, Liu (2013) presents the input variables as shown in **Table 1.** These variables represent the control functions for the superposition relationship to be used to "convolve" or "superimpose" the pressure and flowrate histories. Liu describes the terms in his approach as follows: (1) Δq represents the pressure signal as a function of the flowrate changes, (2) $\Delta q \log \Delta t$ is the dominant term during infinite-acting radial flow, (3) $\Delta q \times \Delta t$ can describe both reservoir boundary and wellbore storage effects, and (4) $\Delta q / \Delta t$ is the primary function for the second order approximation of the exponential integral function and captures reservoir behavior which decay with time.

Tian (2014) proposed linear regression models to interpret flowrate and pressure data that should use significantly less computational time and human intervention compared to the convolution kernel method (Liu 2013). Tian showed that the proposed linear regression models have equal learning quality as compared to the convolution kernel model and that these linear models are less computationally expensive.
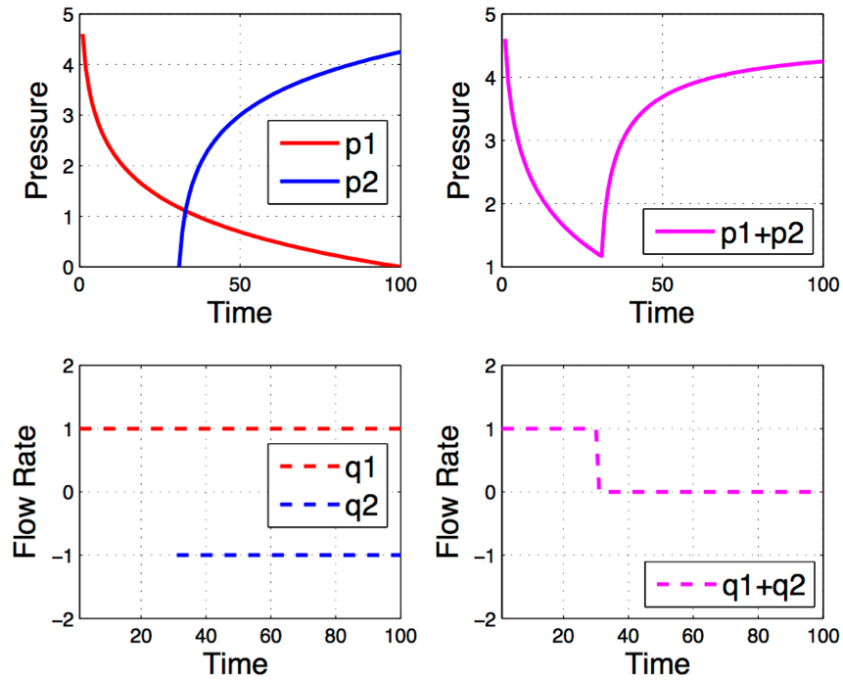
Figure 12 — Schematic illustration of superposition (Liu 2013). q1 represents a constant production rate where pressure decreases, and q2 represents a constant injection rate where pressure increases. Assuming a linear combination of q1 and q2, shown in right, the pressure response is combined. Therefore, the pressure first decreases as this represents production and is followed by an increase in pressure as this represents injection or zero flowrate

Table 1   — Reservoir behavior and corresponding input features.

Relationships for Different Regimes

Superposition:	Δq
Infinite-acting radial flow:	$\Delta q \log[\Delta t]$
Closed boundary:	$\Delta q \times \Delta t$
Constant pressure boundary:	$\Delta q \times \Delta t$
Wellbore storage effect:	$\Delta q \times \Delta t$
Decay effect:	$\Delta q / \Delta t$

3.4 Superposition in Pressure

In this research, we use superposition principles in pressure to identify interference effects of the offset wells. Superposition is a mathematical technique based on the property that solutions to *linear* partial equations can be added to provide another solution (*e.g.*, a variable-rate or variable pressure drop solution). Recall that in the case of a *linear* partial differential equation, the coefficients of the partial differential equation are *not* functions of the dependent variable. In our case, this would be pressure. As such, superposition is only strictly valid for "slightly compressible liquid" cases. Adjustments can be made for compressible liquids and gases, but such cases are beyond the scope of this study.

The principle of superposition indicates that total pressure drop at any point in the reservoir is the sum of the pressure drops from the flow of all wells in the reservoir (Lee 1982). We assume that we have a homogeneous, infinite-acting reservoir with four wells (Well A, B, C, and D) as shown in **Figure 13.** Wells start production at different times: t_A, t_B, t_C, t_D.
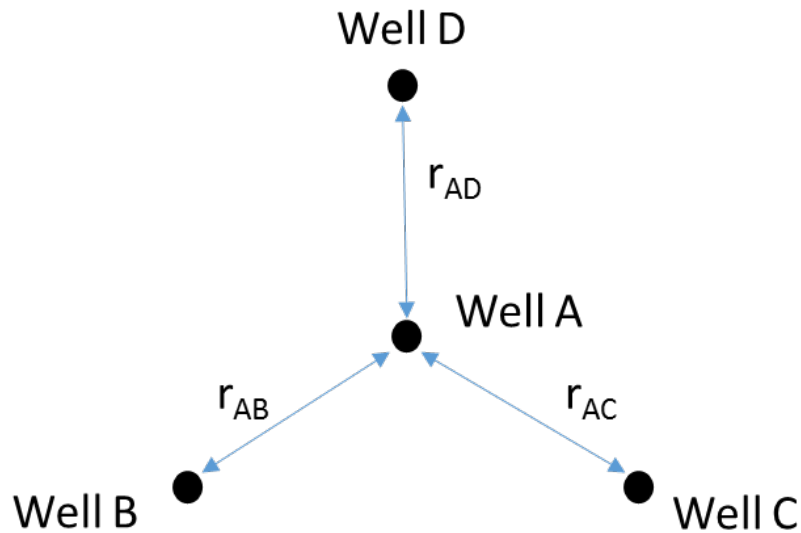
Figure 13 — Schematic illustration of multiple well locations in an infinite reservoir. All wells are connected with each other (after Lee 1982).

The superposition principle suggests that the pressure drop at Well A is a function of pressure drops caused by flow at Well A, Well B, Well C and Well D. **Eq. 26** is a mathematical representation of the overall pressure drop at Well A which has pressure connectivity with Well B, Well C and Well D. In Eq.26, p_i is initial pressure, p_{wf} is bottom-hole pressure and p is current pressure at a specified location in the reservoir.

$$(p_i - p_{wf})_{Total\ \Delta p\ at\ well\ A} = (p_i - p)_{due\ to\ Well\ A} + (p_i - p)_{due\ to\ Well\ B} \quad (26)$$

$$+ (p_i - p)_{due\ to\ Well\ C} + (p_i - p)_{due\ to\ Well\ D}$$

For the infinite-acting reservoir case, the solution of pressure drop includes logarithmic approximations for Ei (Exponential integral) function as well as the Ei function itself for the offset wells. This result is given by **Eq. 27**:

$$(p_i - p_{wf})_{Total\ \Delta p\ at\ well\ A} = -\frac{70.6(q_A B\mu)}{kh}\left[\ln\left(\frac{(1.688\phi)\mu c_t r_{wA}^2}{k(t - t_A)}\right) - 2s_A\right] \quad (27)$$

$$-\frac{70.6(q_B B\mu)}{kh}\left[E_i\left(\frac{-948\phi\mu c_t r_{AB}^2}{k(t - t_B)}\right)\right] - \frac{70.6(q_C B\mu)}{kh}\left[E_i\left(\frac{-948\phi\mu c_t r_{AC}^2}{k(t - t_C)}\right)\right]$$

$$-\frac{70.6(q_D B\mu)}{kh}\left[E_i\left(\frac{-948\phi\mu c_t r_{AD}^2}{k(t - t_D)}\right)\right]$$

where q_A, q_B, q_C, and q_D are the flowrates at which Wells A, B, C, and D are producing, respectively, and p_i and p_{wf} represent the initial and bottom-hole pressure values.

In **Eq. 27**, r_{wA} is the wellbore radius of Well A, while r_{AB}, r_{AC}, and r_{AD} are distances (respectively) for the offset wells surrounding Well A. B is formation volume factor, μ is viscosity, k is permeability, h is formation thickness, ϕ is porosity, c_t is total compressibility, and t is elapsed time. s_A is the skin factor in Well A which indicates permeability change near the wellbore due to foreign-fluid invasion into the reservoir rock. The skin factors for the other wells are not included as these parameters do not affect the pressure measurements for Well A. Considering the terms in Eq. 27, it can be seen that pressure drop depends on the production behavior of the wells, as well as the reservoir and fluid properties.

One of the most important applications of superposition is to model variable-rate production in a given well. **Figure 14** (after Lee 1982) illustrates variable flowrate history for Well A. Well A is producing with flowrate of q_1 between time 0 and t_1 and at t_1 the flowrate increases to q_2; and time from t_1 to t_2 flowrate stays q_2, then at time t_2 the flowrate decreases to q_3. To find sandface pressure at time $t > t_2$, the superposition principles are used. Well A's flowrate history will be divided into three fictitious "wells" with different flowrate histories and different starting times. It is assumed that the fictitious Wells 1, 2 and 3 are producing from the same "location" as Well A. The first contribution to the total drawdown is Well 1 which produces from $t=0$ with q_1 flowrate. Starting at t_2, Well 2 will produce with a production rate of q_2-q_1. Well 3 will behave as an injector (as q_3 is smaller than q_2) and Well 3 will start "injection" at time t_3. The total pressure drop for Well A, is going to be the sum of pressure changes at Wells 1, 2, and 3 caused by production or injection.
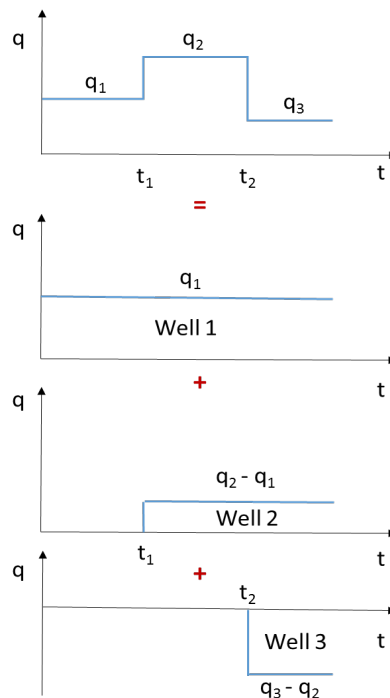
Figure 14 — Variable-rate production schedule for Well A (after Lee 1982). The well is producing with flowrate of q_1 between time 0 and t_1 and at t_1 the flowrate increases to q_2; and time from t_1 to t_2 flowrate stays q_2, then at time t_2 the flowrate decreases to q_3. For the superposition principles Well A's flowrate history is divided into three fictitious "wells" with different flowrate histories of and different starting times

33

3.5 Data-Driven Pressure Forecast

We present the workflow to correlate and forecast bottom-hole pressure behavior and determine interference influence of offset wells. First, we present how single wells tests are conducted, and then we show steps for multi-well tests.

3.5.1 Single Well Tests — Pressure Forecast without Including Offset Well Interference Effects

The lasso method is applied to both flowrate and bottom-hole pressure data to establish the correlation model and to predict future bottom-hole pressure behavior for an arbitrary flowrate history. We apply 5-fold cross-validation to the lasso fit to estimate lambda value which gives lowest cross-validation error. In 5-fold cross-validation, we divide data into 5 equal size subsamples. From the 5 subsamples, a single subsample is chosen as a testing set and 4 subsamples became a training set. The train and testing errors are calculated for each data set. This process is repeated 5 times, each time we select different testing subset. Final training and testing error are estimated by averaging 5 training and testing errors. The cross-validation error is calculated for different values of lambda. We chose the lambda which gives the smallest cross-validation error to use in lasso algorithm.

The workflow procedure is given as follows:

- Data Processing: We visually examine flowrate and bottom-hole pressure data and apply "cleaning" to these data. The flowrate and pressure data are then processed using the Athichanagorn et al. (2002) methodology. First, we reduce permanent downhole gauge (PDG) data by averaging daily, and then identify and eliminate outliers. Next, we apply behavioral filtering to the bottom-hole pressure and flowrate data and remove nonphysical behaviors, such as negative or zero pressure values, or negative production rates, or positive injection rates. We use the random forest algorithm to provide estimates for missing pressure data point by generating a pressure estimate where a null value exists. After cleaning the data, we mathematically transformed flowrate history into features which are discussed in Chapter 3.3.

- Training: We use 80% of flowrate and pressure data as *training data*. The mathe-matically transformed flowrate and time history (features), and bottom-hole pressure

34

data are then input into the lasso correlation method to learn or correlate relationship between features and bottom-hole pressure.

- Testing: After the training process, we use the remaining 20% of flowrate and pressure data as *testing data*. In the testing stage, we forecast the bottom-hole pressure behavior using only the time and flowrate data from learned relationship during training stage. Then we compare the forecasted pressures to the actual bottom-hole pressure data to examine accuracy of forecast using the lasso algorithm.

3.5.2. Multi-Well Tests — Estimation of Inter-Well Communication/Well Interference Effects

We include offset well interference effects to accurately reproduce the observed bottom-hole pressure history and estimate well interference effects. We examine the rate histories of the possible offset wells. We systematically include the rate histories of the offset wells, one by one to identify the influence of each offset well. We follow a similar workflow for a single well case as we consider the rate histories of the offset wells.

- Training: We train the lasso algorithm with mathematically transformed flowrates for the well of interest *and* all possible combinations of interference wells.

- Testing: As before, during the testing period, the lasso algorithm forecasts the future bottom-hole pressure behavior for the well of interest. Any improvements in the predictions of bottom-hole pressure we attribute to the use of the flowrates for the offset wells. The intuition is that these flowrates capture the influence of the offset well interference effects. As a practical consideration, the lasso algorithm could be said to have learned the impact of interference from the offset wells based on the flowrates for these offset wells.

FIELD DESCRIPTION

In this section, we describe the Guler field which is the synthetic simulated reservoir and the Yusif field which is the real field. We apply the methodology to the wells from both fields. The application and results of the methodology explained in the Chapter 3 are presented in the Chapter 5. We use the Guler field to validate the methodology as it is a synthetic example and connectivity between wells is known. We use the Yusif field to show the applicability of the methodology to real field examples as purpose of the study is create methodology for the petroleum industry which can identify interference effects without using traditional multi-well tests.

4.1 Guler Field Description

The synthetic data of the Guler field was generated using an Eclipse 100 simulation module. This Eclipse module was set up as a 3-phase (black oil, dissolved gas, and water) case in a 3D reservoir model consisting of five layers in the vertical (z) direction. The model has three producers and two injectors. **Figure 15** and **Figure 16** illustrate the oil saturation and permeability maps with the labeled well locations. Injection wells are designated with "I" and production wells designated with "P". For reference, the production and injection wells were constrained to the liquid rate for history-matching. As shown in **Figure 17**, a sealing fault divides the reservoir into two separate compartments. Wells P1 and I2 are located in the first compartment, and Wells P2, P3, and I1 are located in the second compartment. There is poor connectivity between Wells P2, P3 and I1, and Wells P1 and I2, which is indicated by the red arrows in **Figure 17**. Wells have good connectivity within each compartment, which is indicated by the green arrows.

Figure 18 shows the vertical placement of the wells and the location of oil-water contact (OWC) in the Guler field's cross-section. The production wells (Wells P1, P2, and P3) were completed in the oil column; the first injection well (Well I1) was completed in the aquifer zone, and the second injection well (Well I2) was completed in both the oil and water columns.
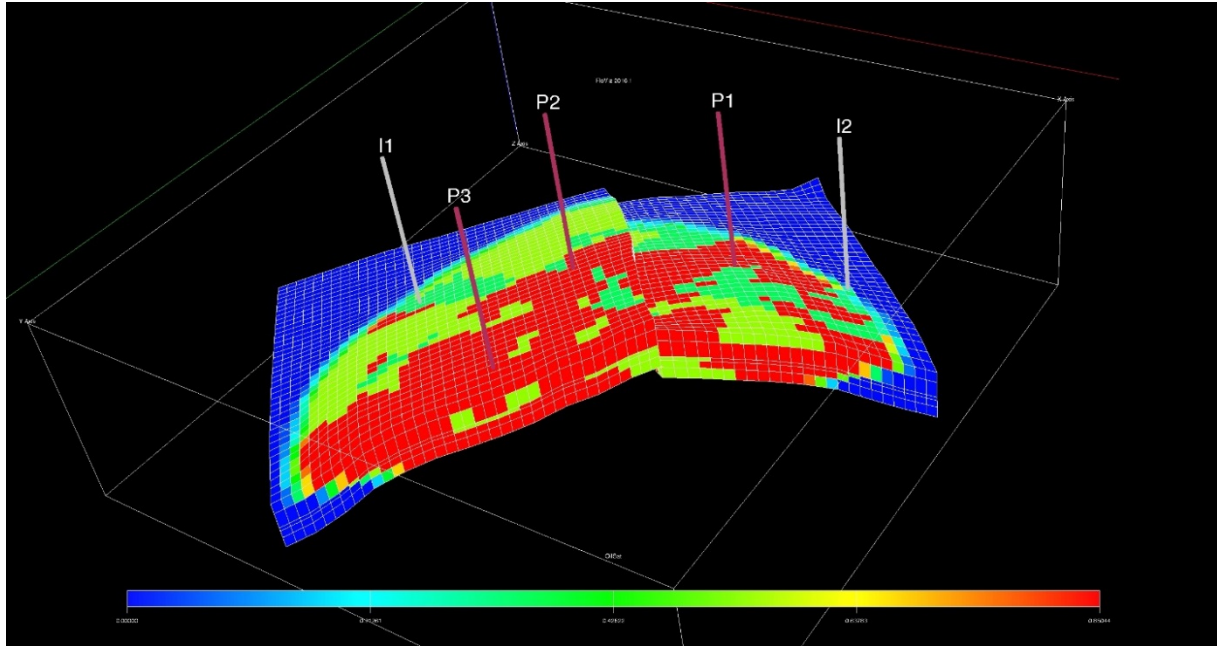
Figure 15 — 3D reservoir model of oil saturation with well locations. Producers are indicated with P, and injectors are indicated with I.
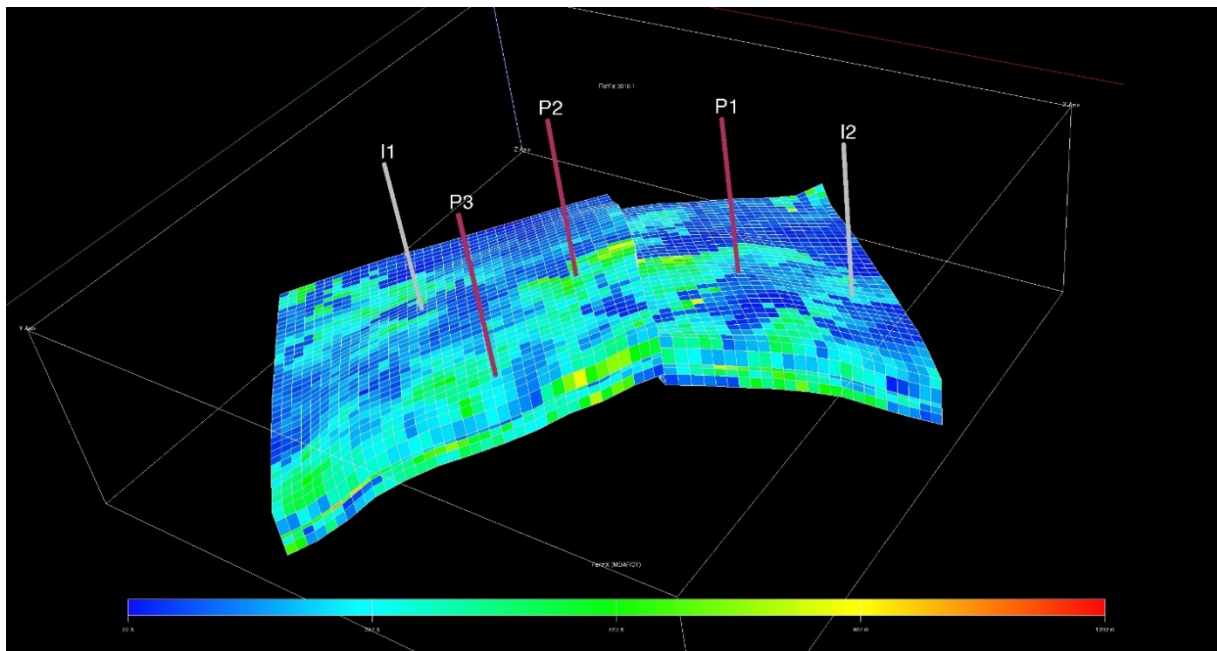
Figure 16 — Map of the permeability to the x direction of the field with well locations. Producers are indicated with P, and injectors are indicated with I.
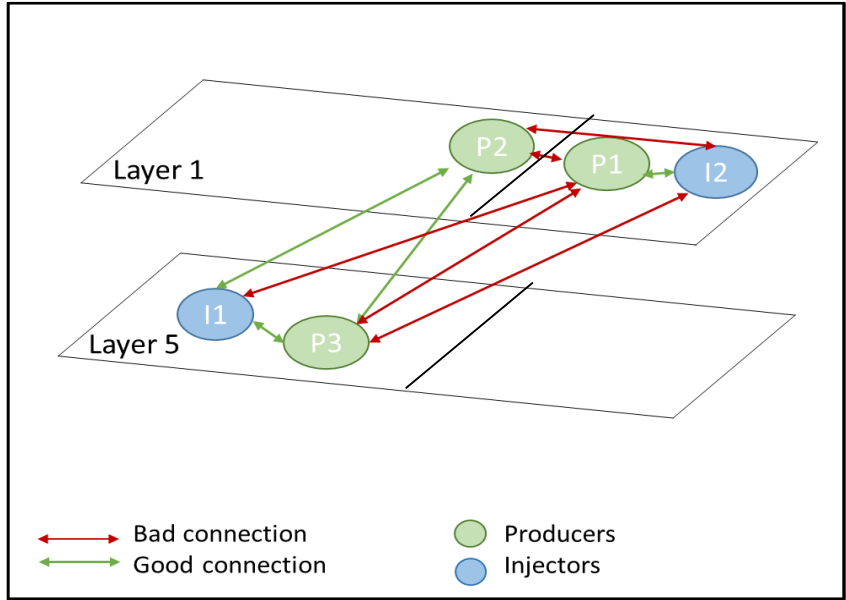
Figure 17 — Guler field connections schematics. The sealed fault illustrated by black line divides the reservoir into two separate compartments. Injectors are designated with blue color and producers designated with green color. The red arrows indicate the poor connectivity between wells. The green arrows indicate the good connectivity between wells.
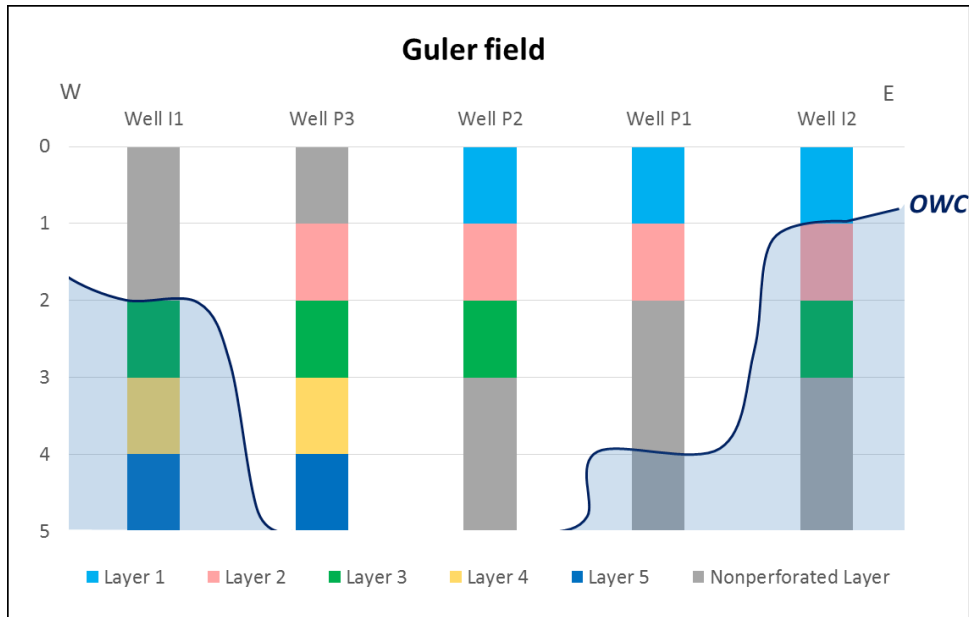
Figure 18 — Perforation layers for all wells in Guler field. The blue color represents water zone.

Figure 19 shows oil and water relative permeability curves which we use in the construction of the synthetic Guler field case. In **Figure 20**, we present average reservoir pressure change with time. Also, we present production rates and bottom-hole pressures for producer wells (Wells P1, P2 and P3), and injection rates and bottom-hole pressures for injectors (Wells I1 and I2) in **Figure 21, Figure 22,Figure 23, Figure 24 and Figure 25.**
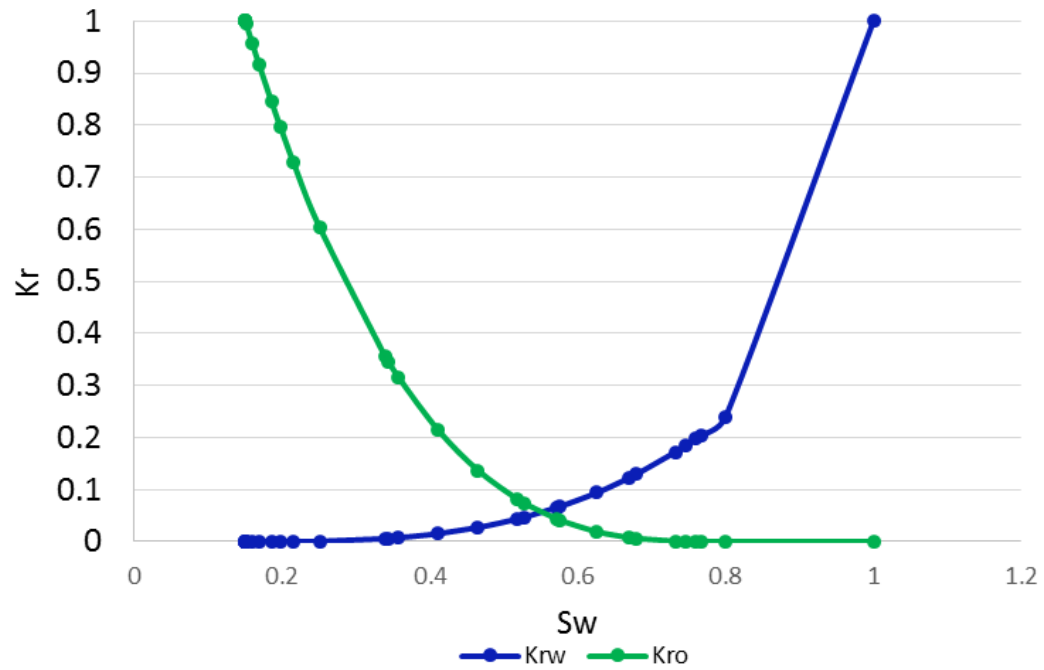
Figure 19 — Oil and water relative permeability curves used for the synthetic Guler field case.
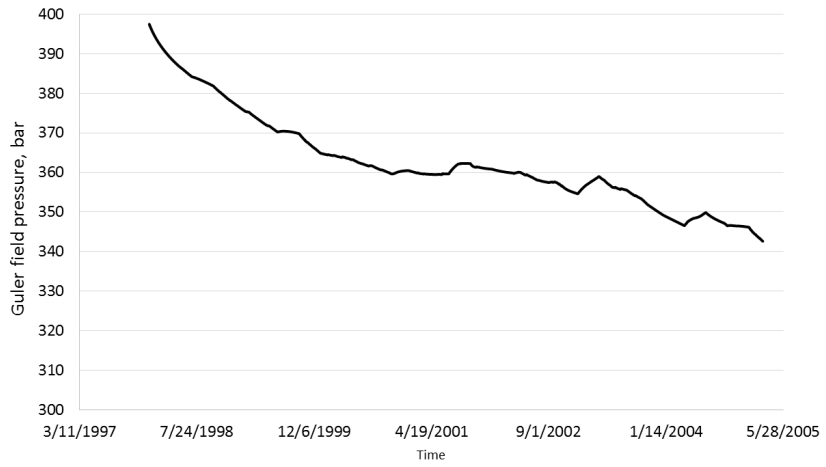
Figure 20 — Reservoir pressure change with time for the synthetic Guler field.

40
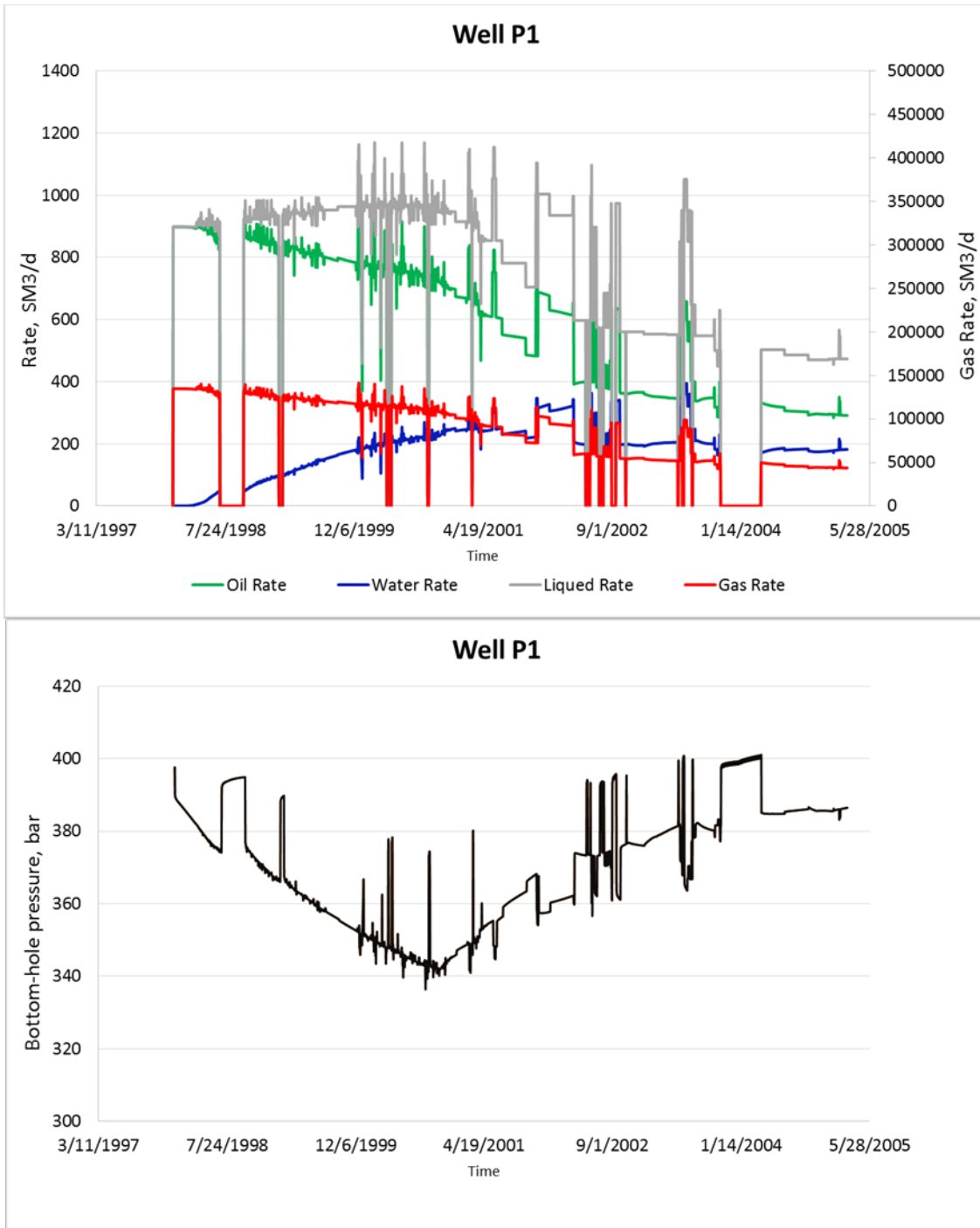
Figure 21 — Upper - Oil, water, gas and liquid rate of Well P1, Bottom-hole pressure
change with time of Well P1.
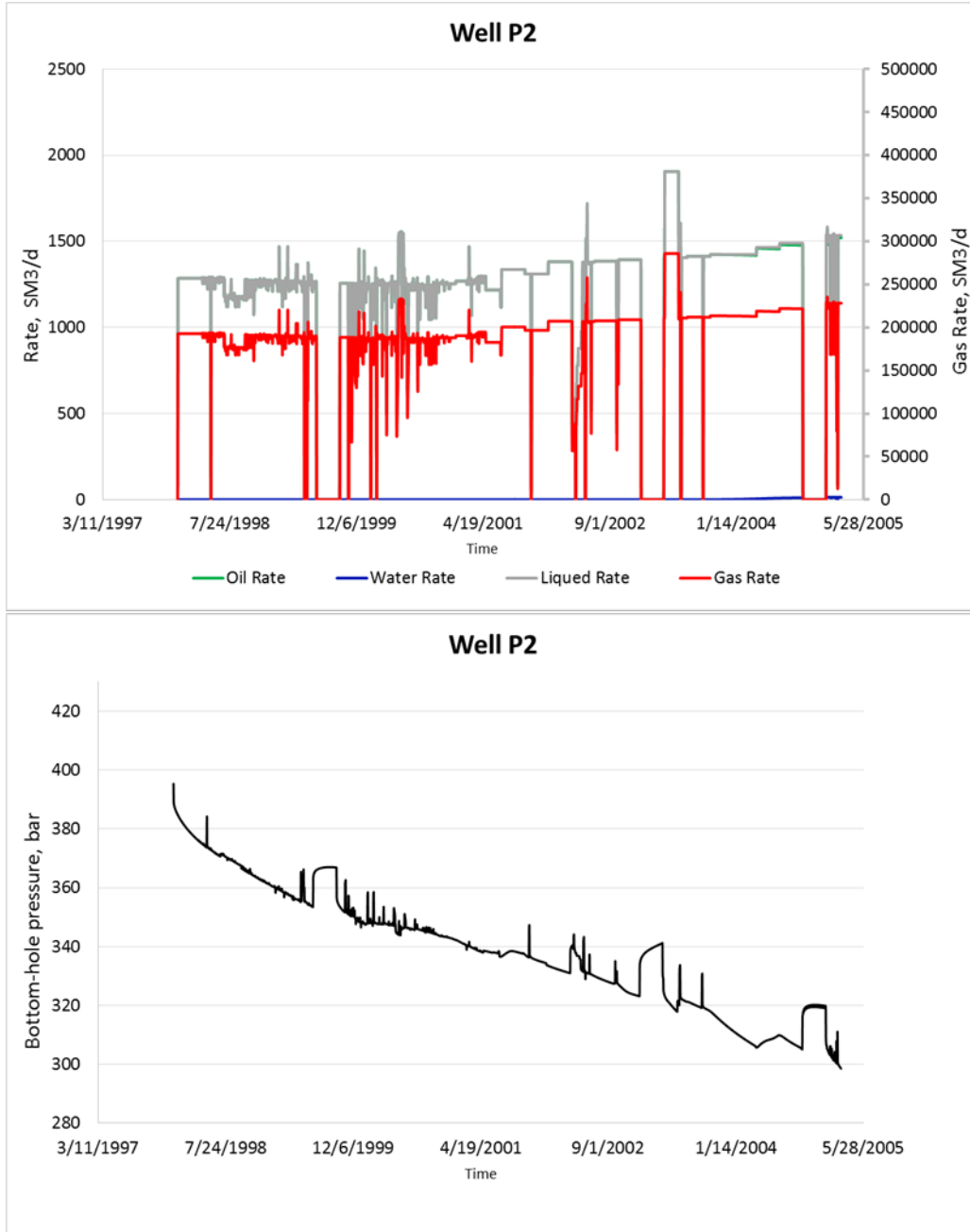
Figure 22 — Upper - Oil, water, gas and liquid rate of Well P2, Bottom-hole pressure change with time of Well P2.
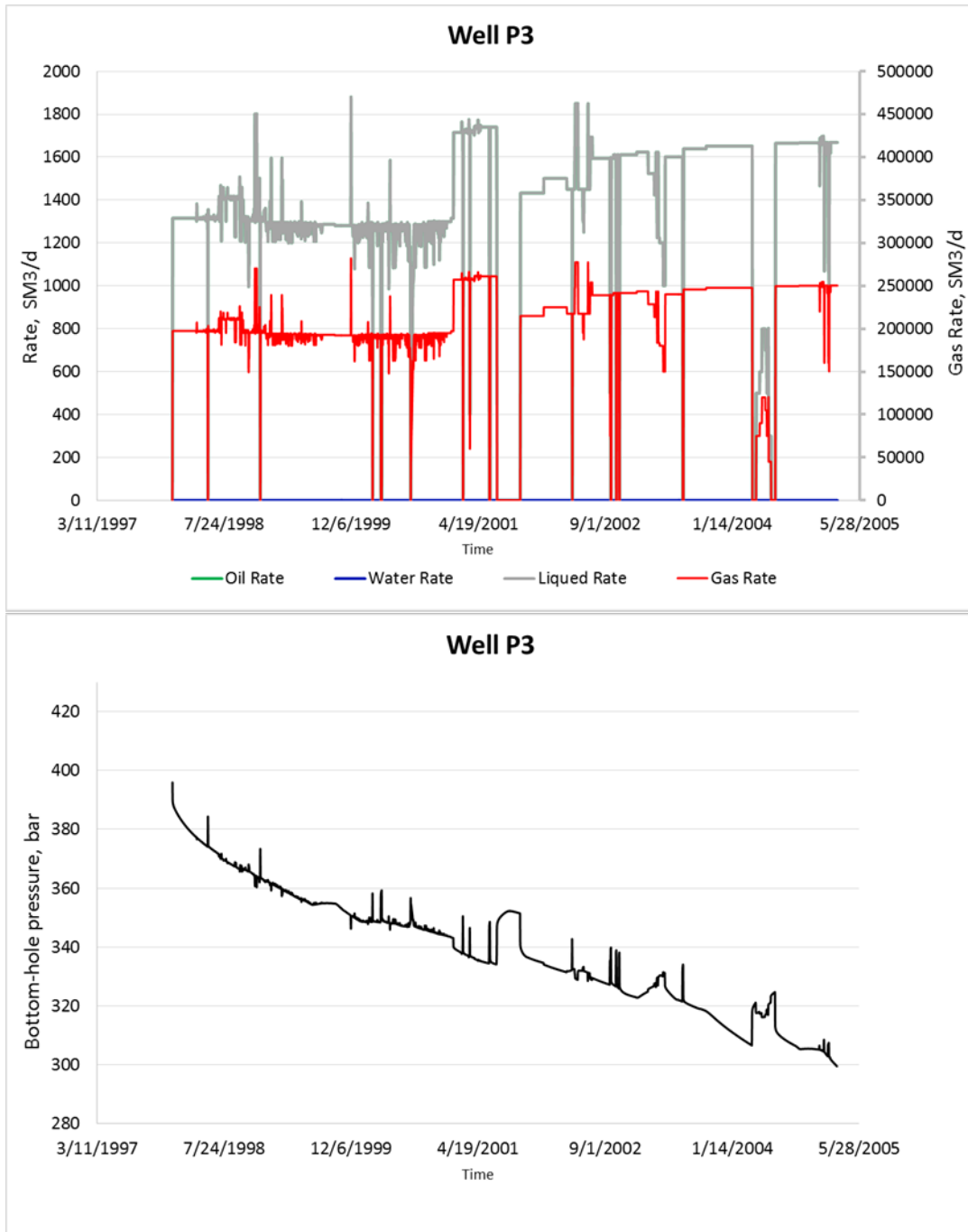
Figure 23 — Upper - Oil, water, gas and liquid rate of Well P3, Bottom-hole pressure change with time of Well P3.
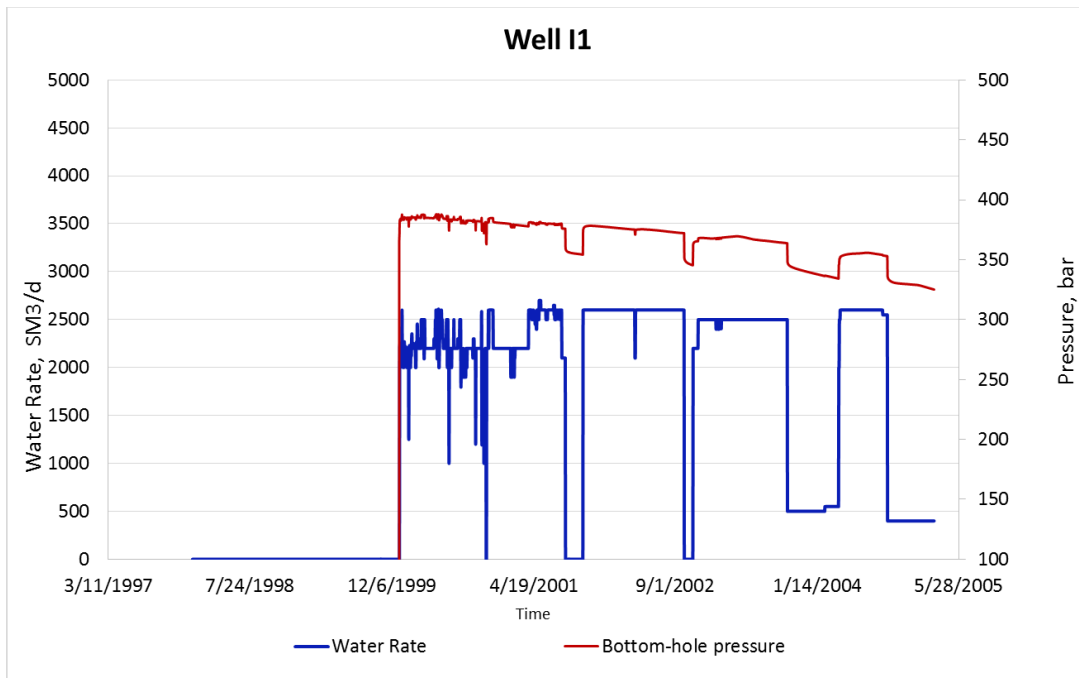
Figure 24　—　Injection rate and bottom-hole pressure versus time for Well I1.
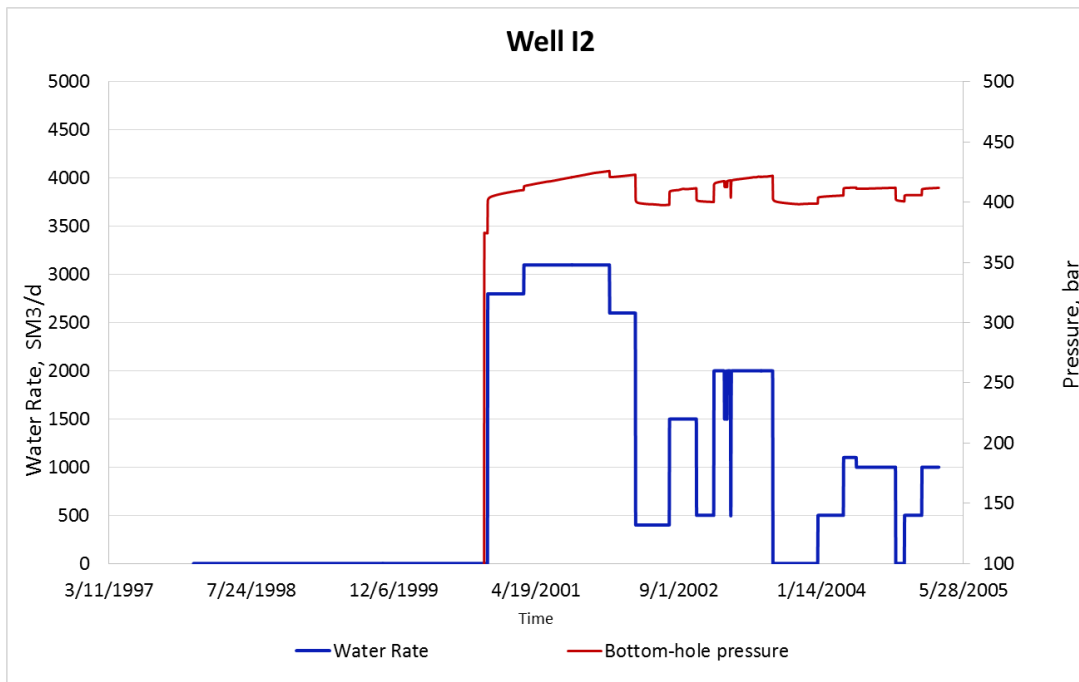
Figure 25　—　Injection rate and bottom-hole pressure versus time for Well I2.

44

4.2 Yusif Field Description

The Yusif is an offshore oil field, which was discovered in 1991 and put into production after six years. The wells were drilled horizontally and drained two reservoir layers. All wells have electrical submersible pump (ESP) systems installed.

The Yusif field covers an area of 210 km^2. It is a carbonate reservoir with a monocline stratigraphic closure. The oil is trapped in the upper part of a relatively float monocline. The structure of the field is a gentle dipping to southeast where it is eroded and capped by overlaying shales. The Yusif reservoir has a shallow water marine depositional environment. The reservoir facies range from lagoonal muddy deposits to rudist shoal facies. It is almost 100% limestone. Two main diagenetic processes were observed in the field:

- Early diagenesis due to meteoric waters or transient paleosols occurrences
- Strong erosion and karstification during eustatic fall of the sea level and regional uplift

Wellbore image logs and formation pressure measurements showed the presence of fractures in the Yusif field. The reservoir consists of thin reservoir beds which resulted in a succession of high (100mD to a few Darcy) and low permeability (0.1 to 50mD) layers. Among the 20 layers defined within the reservoir, the six layers are main produced layers which are L7, L5, L3, L0b1, L0a2, and L0a4. The reservoir lies within the capillary transition zone and consists of a succession of high and low conductive oil-bearing layers.

The burial depth of the Yusif field is approximately 1100 m. It produces a 28 ˚API oil and the maximum oil column thickness is 60 to 70 m, which is mostly located in the transition zone. A 50-m aquifer underlies the field. The free-water-level (FWL) is considered to be at 1225m; however, detailed studies confirm that a varying fluid level exists through the field.

APPLICATION OF THE METHODOLOGY

We present applications of the methodology to correlate and forecast bottom-hole pressure behavior and determine interference influence of offset wells for both the synthetic Guler field and the real Yusif field cases. We validate the methodology by testing the synthetic Guler field. Then, we show applications of methodology with the real field examples. Three of the real field cases are illustrated in this chapter, and one of the real well cases is presented in Appendix.

The purpose of these cases is to demonstrate:

- How the reservoir and well behavior are learned using the statistical algorithm.
- How accurately the bottom-hole pressure can be forecasted.
- How we can detect pressure communication between offset wells and the well of interest.

5.1 Single Well Test

In this section of study, we perform single well tests to reproduce a selected well's bottom-hole pressure responses using only its flowrate history. In single well test, we exclude offset wells interference effects. The importance of this section is to identify which bottom-hole pressure responses are not caused by selected well's flowrate history.

5.1.1 Synthetic Field — Guler field

In this section, we present synthetic demonstration examples that are generated using the reservoir model for the Guler field, the fictitious field case. We believe it is best to demonstrate the methodology for a synthetic case where all inputs and outputs are known and deemed as accurate.

Case 1 —Guler Well P3

In case 1, our well of interest is Well P3 from the Guler field which is located in the second compartment. Well P3 has been producing for 7.5 years.

Figure 26 presents the flowrate history and pressure response versus time for Well P3. We note that the vertical dashed line at approximately 5×10^4 hours denotes the boundary between the training and testing data — the data to the left of this line are used for training and data to the right of this line are used for testing.

Figure 26 — Flowrate and bottom-hole pressure history for Well P3. The data to the left side of the dashed lines are training data, and the data to the right side of the dashed line are testing data.

Figure 27 presents the bottom-hole pressure prediction using the lasso method. The blue data trend is the actual bottom-hole pressure values. The red and green trends are produced by the lasso algorithm. The red data trend is the pressure prediction achieved during the training process. Lastly, the green data trend is the pressure forecast for a given flowrate during the testing period.

 As can be seen from Figure 27, the lasso algorithm generally captures pressure behavior from the flowrate history during both the training and the testing periods. R^2_{train} is equal to 0.9720 and R^2_{test} is equal to 0.6929, which are calculated with Eq. 9. There are some discrepancies between predicted and true pressure values which are indicated by the black circles.

From Figure 26 it can be seen that these discrepancies occur when flowrates of the selected well are relatively constant, but pressure behavior is changing — two pressure increase trends

47

occur at approximately 1.5 x10⁴ and 6 x10⁴ hours when the flowrate of Well P3 is fairly constant. Besides, there is a visible decline in real pressure values (blue curve) during a shut-in which occurs at about 3.3 x10⁴ hours, indicated by the black-dashed circle. That decline in the pressure buildup trend coupled with pressure changes while flowrates are constant suggest that these data might be affected by offset wells.

As the Guler field is synthetic case, we know that Well P3 has communication with offset wells (Wells P2 and I1).

Figure 27 — Pressure prediction with using the lasso technique for Well P3. The blue curve represents the true bottom-hole pressure history. The red curve indicates pressure predictions for the training data while the green curve indicates pressure predictions for the testing data. Only rates from Well P3 were used for training and prediction. Black circles indicate pressure trends that were not captured by linear machine learning algorithm when only Well P3's flowrates were considered. R^2_{train} is equal to 0.9720 and R^2_{test} is equal to 0.6929.

Eq. 28 presents a linear model for reproducing the bottom-hole pressure history of Well P3 using Well P3's flowrate history only. As explained in Figure 7 (Chapter 2.3), the lambda value

is estimated as 0.0026 with 5-fold cross-validation. The coefficients are calculated using criteria in Eq. 16 for the selected lambda value.

As mentioned, the lasso algorithm forces the coefficients of statistically unimportant features to zero. So, by examining the lasso coefficient we can understand the statistical importance of each feature. From the coefficients in Eq. 28, it can be seen that $\Delta q \, \Delta t$ term has less statistical significance for predicting the bottom-hole pressure of Well P3.

$$P_{bh\,WellP3} = 393.0370 + 0.0116 \, \Delta q_{Well\,P3} - 0.0076 \, \Delta q \, log\Delta t_{Well\,P3} \qquad (\; 28\;)$$
$$+ 5.21x10^{-07} \, \Delta q \, \Delta t_{Well\,P3} - 0.0490 \, \Delta q / \Delta t_{Well\,P3}.$$

For case 1, the estimated lambda value is $2.6x10^{-3}$ which is a relatively small value. A small lambda value indicates that the fit involves a small amount of shrinkage relative to the least square solution (James 2013). So, the lasso regression estimates coefficients as the linear regression model. This means we can use the linear regression model to analyze the statistical parameters.

In **Table 2** we present the statistical parameters of the least square approach for linear regression model when relationships of different flow regimes (features) are used to predict bottom-hole pressure (response). We use the linear regression model to analyze the statistical parameters (standard error, t-statistics, and p-value). These statistical parameters quantify to what the extent the model fits the data and the importance of features. The standard error indicates if there is relationship between feature and target — a relatively small value may provide evidence that there is a relationship. The t-statistics and p-value provide information about whether each feature is related to the response. If the p-value is smaller than 1%, we assume that there is an association between feature and target. It is worth recalling that there is a 5% chance that a p-value will incorrectly identify a relationship between features and response.

As shown in Table 2, small p-values indicate that there is an association between the features and the bottom-hole pressure. Therefore, all flow regime features are statistically important in predicting bottom-hole pressure when we interpret p-values.

We will not discuss statistical parameters in detail for following cases. However, statistical parameter tables are shown in order to indicate which feature has statistical association with bottom-hole pressures.

Table 2 — The statistical parameters of the least square model for predicting Well P3's bottom-hole pressure. The small p-values indicate that all flow regimes (features) are statistically significant to forecast bottom-hole pressure.

	Coefficient	Std. error	t-statistic	p-value
Intercept	393.1	0.34842	1128.2	<0.01
q	0.011818	0.00061533	19.206	<0.01
qlogt	-0.0076364	0.00021304	-35.845	<0.01
qt	-6.012e-07	6.7886e-09	-88.56	<0.01
q/t	-0.052307	0.012558	-4.1653	<0.01

Case 2 — Guler Well P1

In case 2, we selected Well P1 to demonstrate the accuracy of the pressure prediction without including the interference effects of the offset wells.

In **Figure 28**, we present flowrate and bottom-hole pressure history of Well P1 for more than 7 years of production. The vertical dashed line at 5×10^4 hours divides data into training and testing datasets.

Figure 28 — Flowrate and bottom-hole pressure history for Well P1. The data to the left side of the dashed lines are training data, and the data to the right side of the dashed line are testing data.

Figure 29 presents bottom-hole pressure prediction for Well P1 using the lasso method. There is general agreement between pressure prediction by the lasso algorithm and real pressure values. R^2_{train} is equal to 0.8307 and R^2_{test} is equal to -0.5188. The negative testing R^2 value indicates that mean values of testing data are closer to the testing dataset than forecasted values with the algorithm. However, there is an obvious pressure discrepancy, indicated by the black oval.

Figure 28 makes it clear that despite the reasonably constant flowrate between 2.3×10^4 and 3×10^4 hours, we observe a continuous increase in real pressure data. This can be a signature of offset well interference effects on bottom-hole pressure measurements. As Guler Well P1 is a synthetic well, we know that there is communication between nearby well (Well P1).

51

$R^2_{train} = 0.8307$ $R^2_{test} = -0.5188$
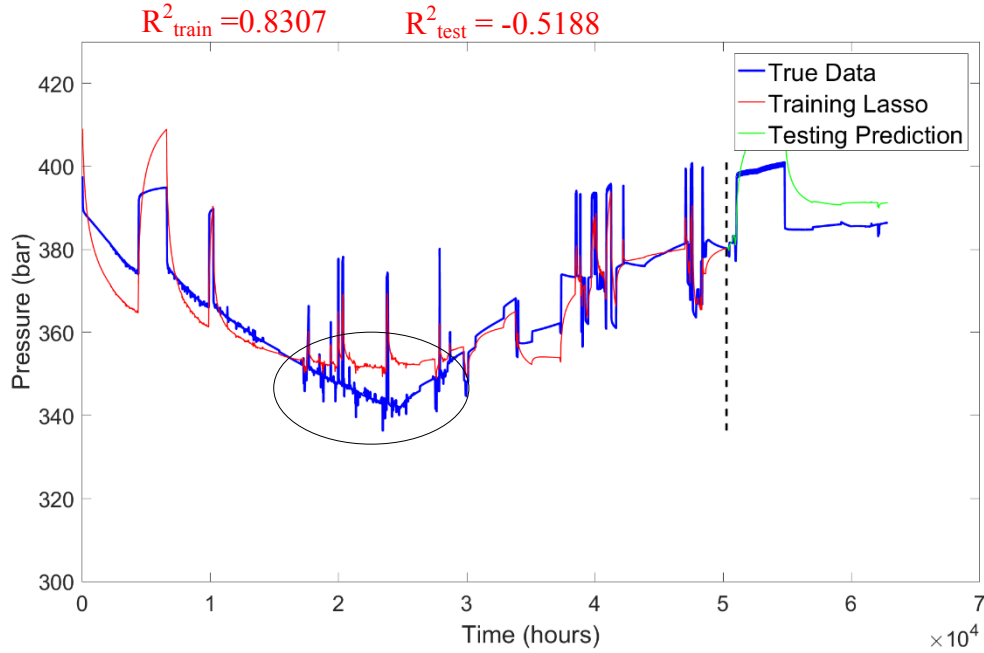
Figure 29 — Pressure prediction for Well P1 using the lasso algorithm. The blue curve represents the true bottom-hole pressure history. The red curve indicates pressure predictions for the training data while the green curve indicates pressure predictions for the testing data. Only rates from Well P1 were used for training and prediction. Black oval indicates pressure trends that were not captured by linear machine learning algorithm. R^2_{train} is equal to 0.8307 and R^2_{test} is equal to -0.5188.

Eq. 29 presents a linear model for reproducing the bottom-hole pressure history of Well P1, using only its flowrate history using the lasso fit where the lambda value is 0.0032. From the coefficients of the lasso algorithm, it can be seen that $\Delta q \, \Delta t$ term has less statistical significance for predicting the bottom-hole pressure of Well P1.

$$P_{bh\,WellP3} = 418.7138 + 0.0376 \, \Delta q_{Well\,P1} - 0.0272 \, \Delta q \, log\Delta t_{Well\,P1} \quad (\ 29\ )$$
$$+ 3.28x10^{-07} \, \Delta q \, \Delta t_{Well\,P1} - 0.2583 \, \Delta q / \Delta t_{Well\,P1}.$$

As the estimated lambda value is $3.2x10^{-3}$ which is a relatively small value, the linear regression model is going give similar results as the lasso method.

We present the statistical parameters of the least square model for regression in **Table 3.** The p-values, which are smaller than 0.01, indicate that there is a relationship between different flow regime features and bottom-hole pressure predictions.

52

Table 3 — Statistical parameters of the least square model for predicting Well P1's bottom-hole pressure. The small p-values indicate that all flow regimes (features) are statistically significant to forecast bottom-hole pressure.

	Coefficient	Std. error	t-statistic	p-value
Intercept	418.73	0.58551	715.15	<0.01
Δq	0.037805	0.0017702	21.357	<0.01
ΔqlogΔt	-0.027304	0.0005402	-50.545	<0.01
ΔqΔt	4.56e-07	1.7211e-08	26.494	<0.01
Δq/Δt	-0.26197	0.041177	-6.3621	<0.01

5.1.2 Field Example — Yusif Field

In this section, we present the bottom-hole pressure correlation and forecast for real field examples. We believe that the applicability of the methodology in real field data is essential as the purpose of this work is to make the methodology which can be used by the oil industry. In this section, we chose three wells (Wells 12, 10, 110) to demonstrate the applications of the methodology to the real field data. One real well results are presented in Appendix.

In the Yusif field, daily liquid flowrates were reported. Bottom-hole pressure data was acquired from PDG measurements. After visualizing the data, we cleaned and preprocessed pressure data to facilitate the correlation captured by the lasso algorithm.

Case 3 — Yusif Well 12

In case 3, our well of interest is Well 12. Well 12 has been in production for 7 years.

In **Figure 30**, we present flowrate and bottom-hole pressure data versus time for Well 12. **Figure 31** shows pressure predictions using the lasso correlation algorithm. There is not any obvious pressure discrepancy when we are reproducing bottom-hole pressure response of Well 12 using only its flowrate history. R^2_{train} is equal to 0.8996 and R^2_{test} is equal to -2.6256. The

negative testing R^2 value indicates that mean values of the testing data are closer to the testing dataset than forecasted values with the algorithm.
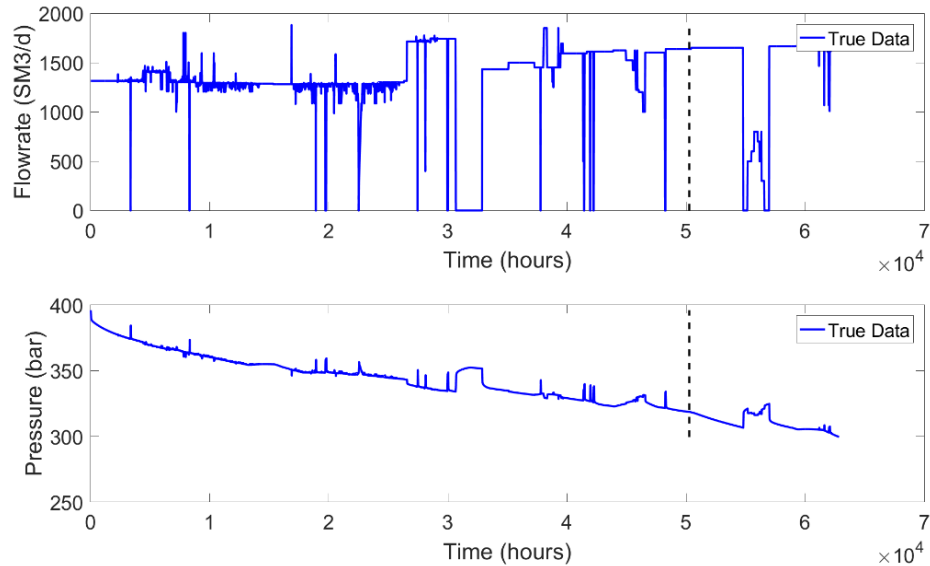

Figure 30 — Flowrate and bottom-hole pressure history for Well 12. The data to the left side of the dashed lines are training data, and the data to the right side of the dashed line are testing data.
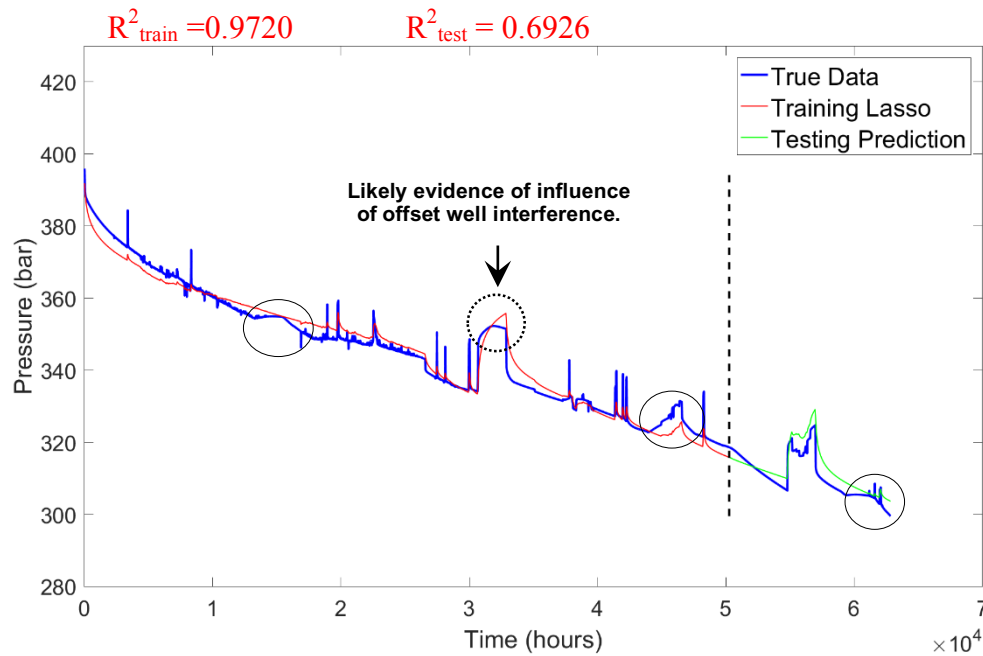
Figure 31 — Pressure prediction with using the lasso algorithm for Well 12. The blue curve represents the true bottom-hole pressure history. The red trend indicates pressure predictions for the training data while the green trend indicates pressure predictions for the testing data. Only rates from Well 12 were used for training and prediction process. Pressure trends were accurately captured by the lasso algorithm. R^2_{train} is equal to 0.8996 and R^2_{test} is equal to -2.6256.

Eq. 30 presents a linear model for reproducing the bottom-hole pressure response of Well 12 using its flowrate history only. The coefficients are calculated using the lasso algorithm when lambda is 0.009. From the coefficients, it can be seen that $\Delta q\ \Delta t$ term has less statistical significance for predicting the bottom-hole pressure of Well 12.

$$P_{bh\ WellP3} = 2085.442 + 0.051\ \Delta q_{Well\ 12} - 0.0472\ \Delta q\ log\Delta t_{Well\ 12} \qquad (\ \mathbf{30}\ )$$
$$+ 8.28x10^{-07}\ \Delta q\ \Delta t_{Well\ 12} - 0.7238\ \Delta q/\Delta t_{Well\ 12}.$$

The estimated lambda value is $0.9x10^{-2}$ which is a relatively small value. Hence, the linear regression model is going give similar results as the lasso method. We use the linear regression model, to present the statistical parameters.

Table 4 provides information about the statistical importance of features to predict and forecast pressure behavior (target) for Well 12 using the least square approach. All the p-values are smaller than 0.01 meaning that all flow regimes are statistically significant.

Table 4   — The statistical parameters of the least square model for predicting Well 12's bottom-hole pressure. The small p-values indicate that all flow regimes (features) are statistically significant to forecast bottom-hole pressure.

	Coefficient	Std. error	t-statistic	p-value
Intercept	2086.3	15.693	132.94	<0.01
Δq	0.05131	0.0024508	20.936	<0.01
ΔqlogΔt	-0.047359	0.00093075	-50.883	<0.01
ΔqΔt	8.3013e-07	1.5244e-08	54.455	<0.01
Δq/Δt	-0.72795	0.038434	-18.94	<0.01

Case 4 — Yusif Well 10

In case 4, we selected Well 10 to demonstrate the accuracy of the pressure prediction with excluding the interference effects of the offset wells. Well 10 is a laterally drilled horizontal well which has been in production for more than 6.5 years.

In **Figure 32,** we demonstrate flowrate and pressure history of Well 10.

Figure 33 presents bottom-hole pressure predictions for Well 10 for training and testing datasets. As can be seen from Figure 33, the lasso algorithm is capable of capturing general pressure behavior. However, we note that there are obvious discrepancies between pressure predictions and real measurements, indicated by black ovals. R^2_{train} is equal to 0.8000 and R^2_{test} is equal to 0.3982.

Figure 32 makes it clear that these discrepancies occur between 1.03×10^5 and 1.04×10^5 hours, when the flowrates of Well 10 are relatively constant but pressure trend changes. We observe the same behavior from time 1.23×10^5 hours to 1.3×10^5 hours, where the flowrate fairly

56

constant while the pressure responses increase. These discrepancies suggest that the pressure responses might be dominated by offset wells.

In general, we note for this particular case, flowrate history of Well 10 can-not accurately reproduce its bottom-hole pressure responses.
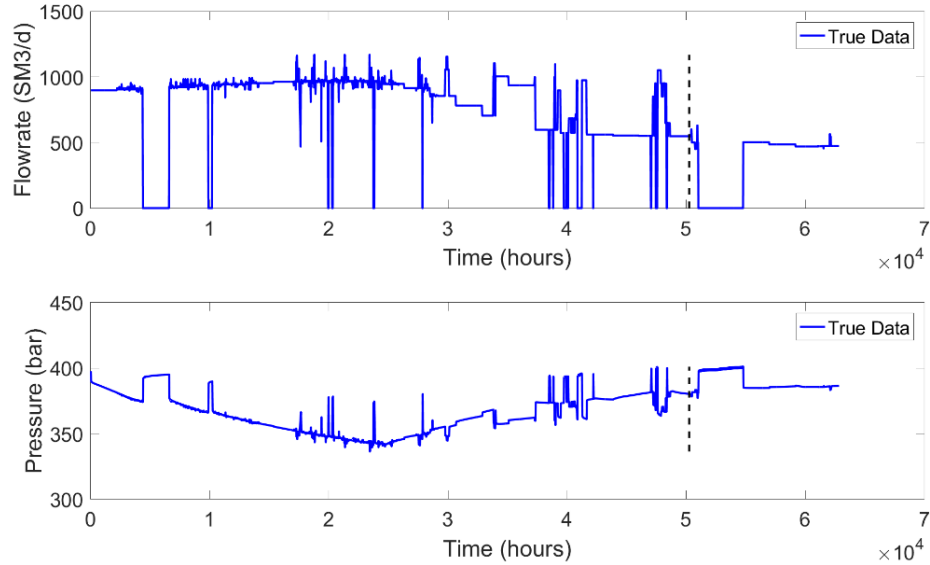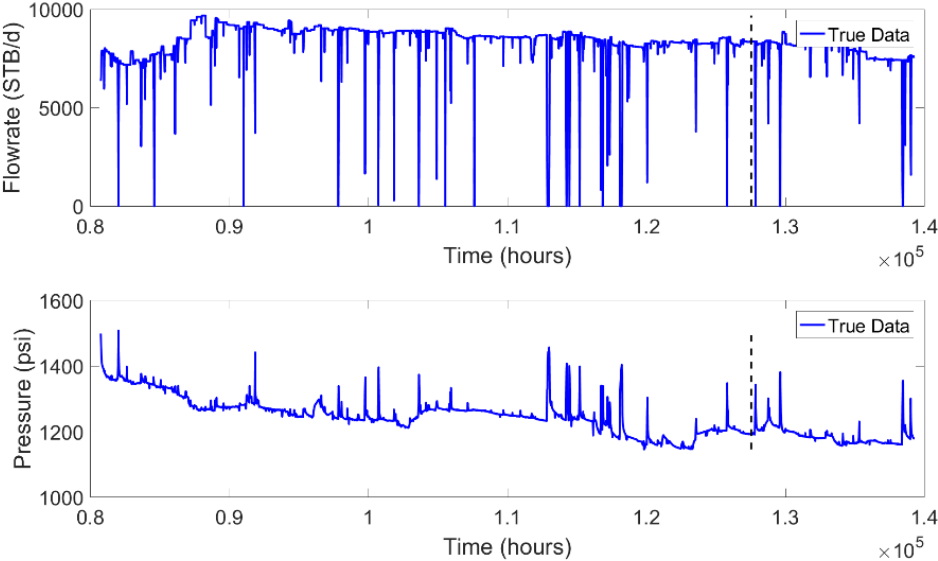

Figure 32 — Flowrate and bottom-hole pressure history for Well 10. The data to the left side of the dashed lines are training data, and the data to the right side of the dashed line are the testing data.
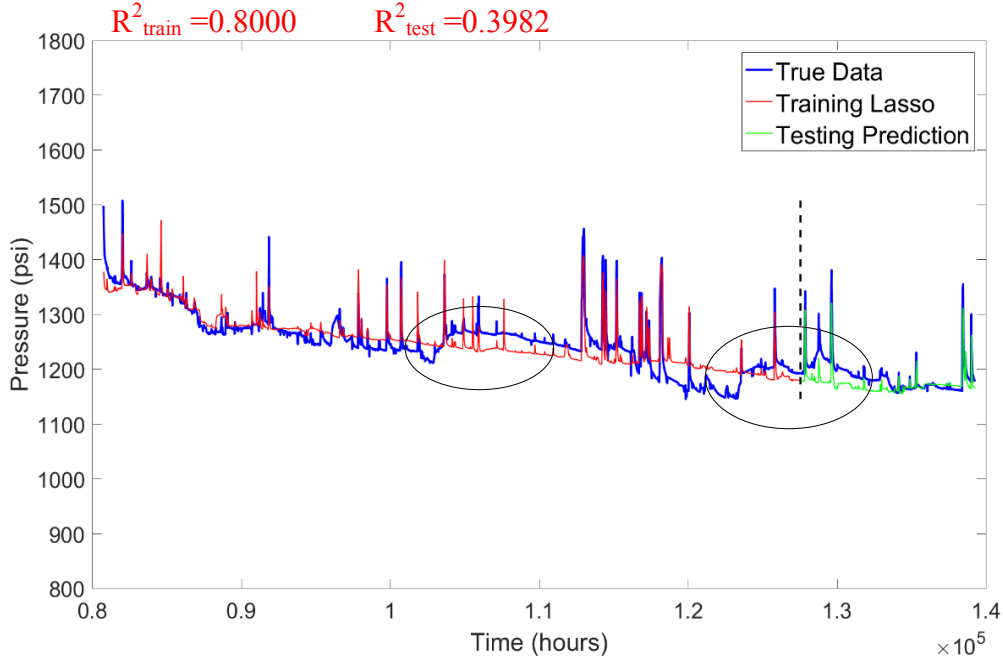
Figure 33 — Pressure prediction using the lasso algorithm for Well 10. The blue trend represents the true bottom-hole pressure history. The red trend indicates pressure predictions for the training data while the green trend indicates pressure predictions for the testing data. Only rates from Well 10 were used for training and prediction. Black circles indicate pressure trends that were not captured by linear machine learning algorithm. R^2_{train} is equal to 0.8000 and R^2_{test} is equal to 0.3982.

Eq. 31 presents a linear model for reproducing bottom-hole pressure history of Well 10 using the lasso algorithm when the lambda value is 0.03. From the coefficients, it can be seen that $\Delta q\ \Delta t$ term has less statistical significance for predicting the bottom-hole pressure of Well 10.

$$P_{bh\ WellP3} = 1732.86 - 0.023\ \Delta q_{Well\ 10} - 0.0004\ \Delta q\ log\Delta t_{Well\ 10} \qquad (\ 31\ )$$
$$- 3.827x10^{-07}\ \Delta q\ \Delta t_{\ Well\ 10} + 0.3057\ \Delta q/\Delta t_{Well\ 10}.$$

The estimated lambda value is $0.3x10^{-1}$ which is a relatively small value. Hence, the linear regression model is going give similar results as the lasso model. We use the linear regression model to present the statistical parameters.

As shown in **Table 5**, we present the statistical parameters of the least square model. The high p-value (0.617) infers that there is not a strong association between infinite-acting radial flow feature and Well 10's bottom-hole pressure measurements. We note that Well 10 is close to the

58

boundary of the field, so it exhibits infinite-acting flow for a short period. This can explain why infinite-acting radial flow is statistically unimportant.

Table 5 — The statistical parameters of the least square model for predicting Well 10's bottom-hole pressure. The small p-values indicate that all flow regimes (features) are statistically significant to forecast bottom-hole pressure except infinite-acting radial flow feature.

	Coefficient	Std. error	t-statistic	p-value
Intercept	1732.5	8.8986	194.7	<0.01
Δq	-0.023364	0.0016454	-14.199	<0.01
$\Delta q \log \Delta t$	-0.00028789	0.00057562	-0.50014	0.617>0.01
$\Delta q \Delta t$	-3.8405e-07	7.1716e-09	-53.552	<0.01
$\Delta q / \Delta t$	0.31148	0.024783	12.568	<0.01

Case 5 — Yusif Well 110

In case 5, we selected Well 110 to demonstrate the accuracy of the pressure prediction without including the interference effects from the offset wells. Well 110 has been in production for more than 7 years. We note that, dynamic observation indicate that Well 110 has strong interference effects with nearby injector.

Figure 34 presents the flowrate and pressure history of Well 110. The data at the left part of the vertical dash line (at approximately 1.44×10^5 hours) is the training dataset, and right part of this line is the testing dataset.

In **Figure 35**, we present the bottom-hole pressure prediction for *training* and forecast for *testing* datasets using the lasso algorithm. We acknowledge that in case 5, the results are similar to previous cases but this case differs because there are significant discrepancies between real measurements and pressure predictions. For this specific case, the lasso algorithm does not accurately reproduce pressure behavior using only Well 110's flowrate history. R^2_{train} is equal to 0.3872 and R^2_{test} is equal to -6.1772.

As can be seen from Figure 34, despite the approximately the constant flowrate history of Well 110, there are noticeable variations in pressure responses. For example, drastic pressure decrease between 1.05×10^5 hours and 1.18×10^5 hours accompany by flowrate decrease. This might be indication of strong interference effects from offset wells.
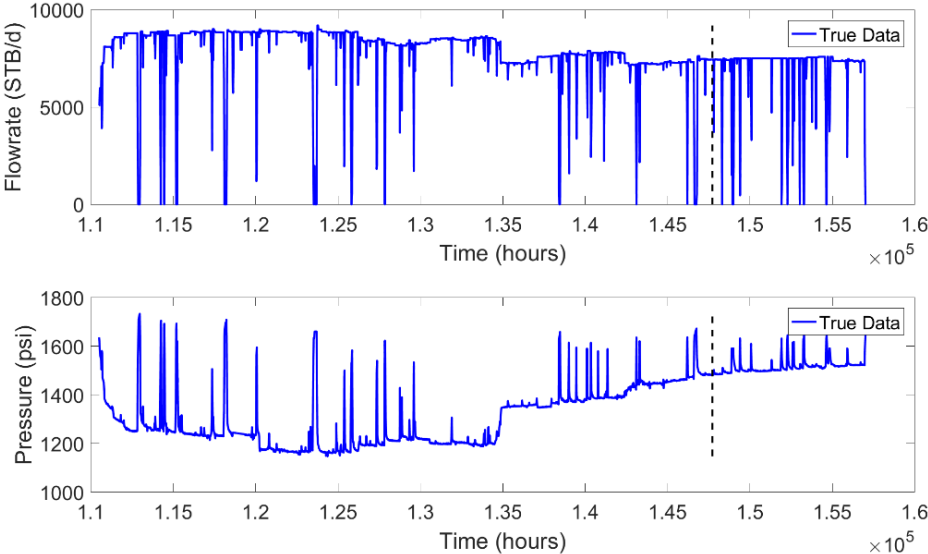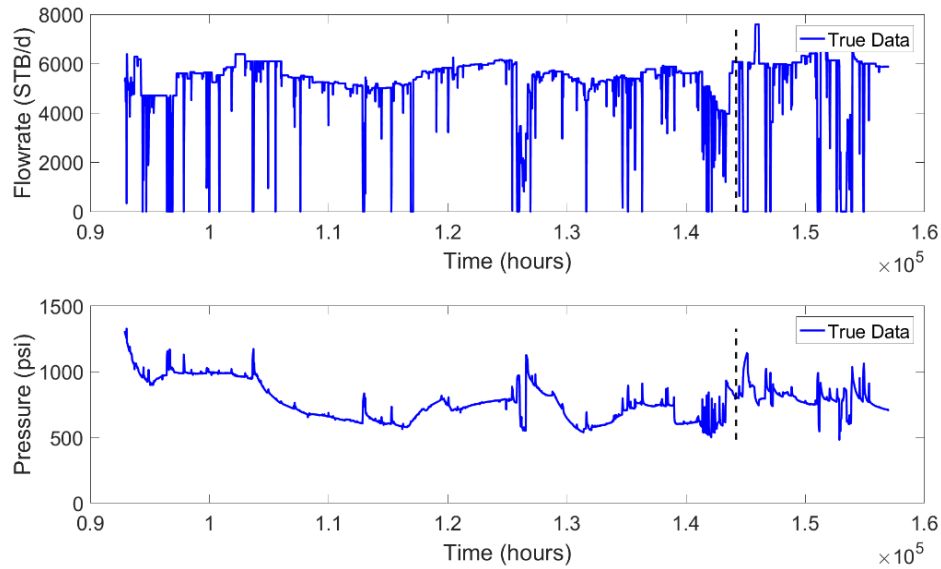

Figure 34 — Flowrate and bottom-hole pressure history for Well 110. The data to the left side of the dashed lines are training data, and the data to the right side of the dashed line are the testing data.
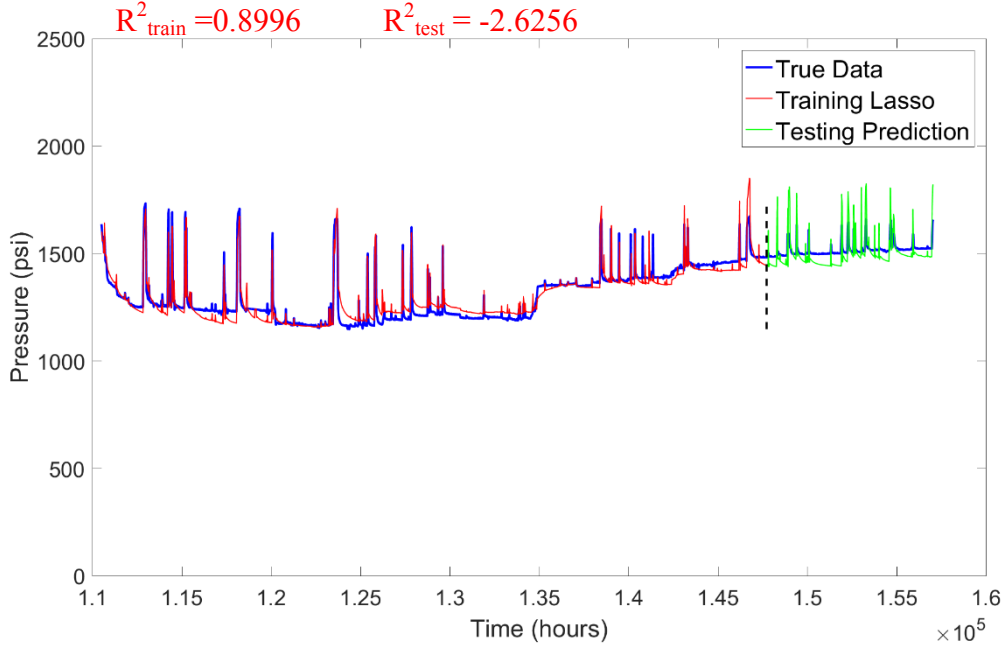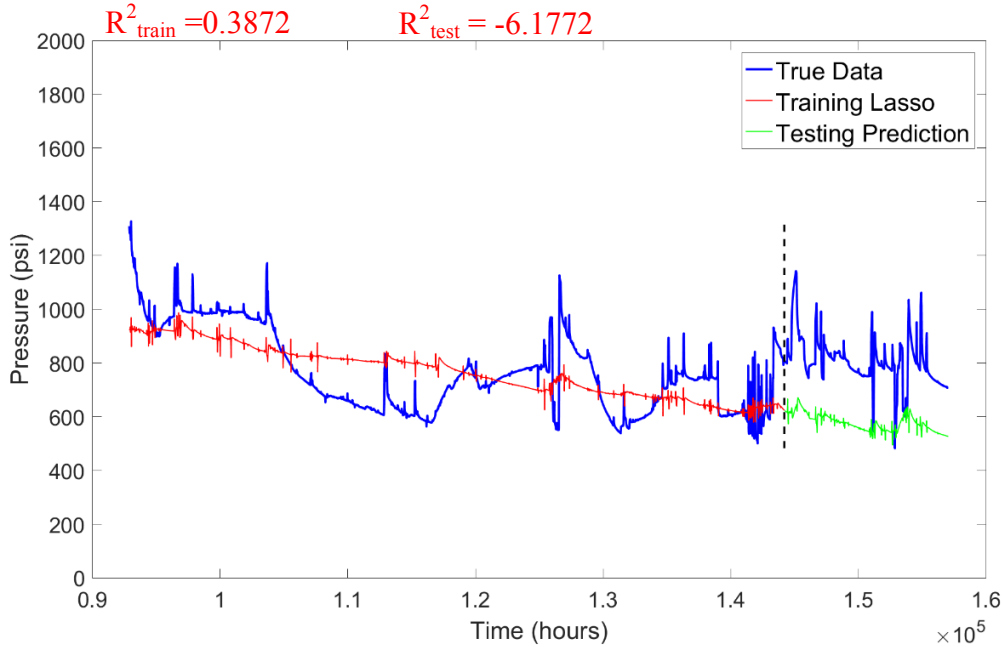
Figure 35 — Pressure prediction using the lasso algorithm without considering interference effects for Well 110. The blue trend represents the true bottom-hole pressure history. The red trend indicates pressure predictions for the training data while the green trend indicates pressure predictions for the testing data. Only rates from Well 110 were used for training and prediction. The lasso algorithm is not able to capture pressure trends using only well's flowrate history for Well 110. R^2_{train} is equal to 0.3872 and R^2_{test} is equal to -6.1772.

Eq. 32 presents linear model for reproducing bottom-hole pressure history of Well 110 using its flowrate history only. The coefficients are calculated using the lasso algorithm when the estimated lambda is 0.037. As for the other well cases, from the coefficients, it is obvious that $\Delta q \, \Delta t$ term has less statistical significance for predicting the bottom-hole pressure of Well 110.

$$P_{bh\,WellP3} = 2085.442 + 0.051 \, \Delta q_{Well\,110} - 0.0472 \, \Delta q \, log\Delta t_{Well\,110} \quad (\ 32\ )$$
$$+ \ 8.28x10^{-07} \, \Delta q \, \Delta t_{\,Well\,110} - 0.7238 \, \Delta q / \Delta t_{Well\,110}.$$

The estimated lambda value is $0.37x10^{-1}$ which is a relatively small value. Hence, the linear regression model is going give similar results as the lasso method. We use the linear regression model, to present the statistical parameters.

61

We present the statistical parameters of the least square regression model in **Table 6**. All the p-values are smaller than 0.01, therefore, all flow regimes are statistically significant.

Table 6   —   The statistical parameters of the least square model for predicting Well 110's bottom-hole pressure. The small p-values indicate that all flow regimes (features) are statistically significant to forecast bottom-hole pressure.

	Coefficient	Std. error	t-statistic	p-value
Intercept	1815.7	53.829	33.731	<0.01
Δq	0.064775	0.01173	5.5222	<0.01
ΔqlogΔt	-0.026643	0.0044553	-5.98	<0.01
ΔqΔt	-1.0974e-06	3.7568e-08	-29.212	<0.01
Δq/Δt	-0.90044	0.16355	-5.5055	<0.01

5.1.3. Discussion of Results

The objective of predicting the pressure without including offset well interference effects is to determine which pressure signals are not caused by the selected well's flowrate. By using only the selected well's flowrate, we excluded the influence of offset wells. We note that wells having strong interference effects, do not reproduce accurate the pressure responses. This evidence leads us to the next phase of this study where we systematically add the offset wells to the lasso algorithm to detect from which well interference effects are coming.

5.2 Pressure Prediction Including Offset Well Interference Effects

The goal of this section is to explain how we identify the interference effects of offset wells using the lasso algorithm. As mentioned in the discussion of the results of the previous section, there are cases where the interference effects must be included to accurately reproduce the observed bottom-hole pressure history. In this section, we systematically include the rate histories of the offset wells one by one to capture the influence effects from the offset wells.

We use the mathematically transformed flowrate histories systematically for all offset wells, identifying which wells provide responses. If there is more than one identified interference effects from offset wells, then we create a new scenario with combination of those offset well. The idea is that by doing this, we visualize and identify the influence of each offset well on the bottom-hole pressure. We establish the "most likely" well interference scenarios using the R^2 value and the visual fit. While doing it, we compare results of base cases or single-well tests with multi-well tests.

5.2.1 Synthetic Field —Guler field

As mentioned in Chapter 3, the Guler field is a synthetic field which has three producers (Wells P1, P2, and P3) and two injectors (Wells I1 and I2). A sealing fault divides reservoir into two compartments. We note that there is no horizontal pressure barrier between layers; however vertical permeability is lower than horizontal permeability. With the Guler field, we validate the methodology to identify interference wells by comparing the results with a known reservoir model.

Case 1 — Guler Well P3

For case 1, we demonstrated pressure predictions for Well P3 and in this part we include offset wells interference effects. From the synthetic reservoir model, we know that there is communication between Wells P2 and I1 and Well P3. Moreover, the sealed fault disconnects communication between Well P3 and Wells P1 and I2.

This section presents a detailed explanation of how we identify interference effects from offset wells using the lasso algorithm. We present the scenarios or realizations generated by the lasso algorithm by considering the combinations of Well P3 and its offsets one by one. We compare the improvements in pressure predictions for different scenarios with the base case where pressure history is reproduced using only Well P3's flowrate history. As mentioned, we compare the visual fit and the R^2 values for different scenarios to identify interference effects. In general, we consider the visual fit more critical than the R^2 value.

Figure 36 presents pressure predictions for the Well P3/Well I1 configuration. The R^2_{train} is equal to 0.9935 and R^2_{test} is equal to 0.8564.

Compared with the base case, there are visual improvements gained in pressure prediction trends between 1.8×10^4 and 3×10^4 hours by adding Well I1. The R^2 values for both training and testing increase when we add Well I1 flowrate history. Therefore, we assume that there might be interference effects between these wells.
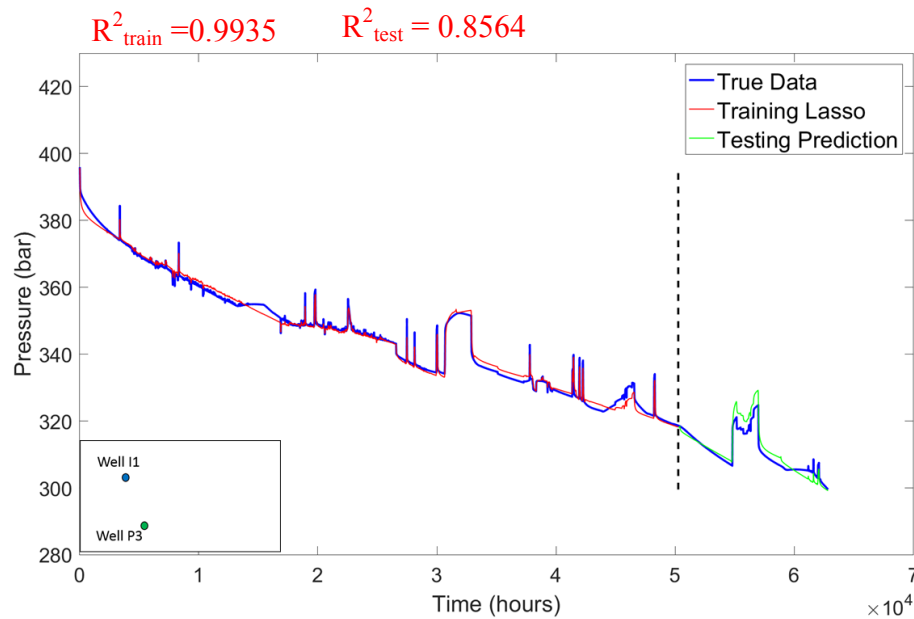
Figure 36 — Pressure predictions for the Wells P3 and I1 configuration using the lasso algorithm. Compare to the base case, there are visual improvements in pressure prediction trends between 1.8×10^4 and 3×10^4 hours. R^2_{train} is equal to 0.9935 and R^2_{test} is equal to 0.8564.

Eq. 33 presents the linear model for this configuration. Coefficients are calculated using the lasso algorithm.

$$P_{bh\,Well\,P3} = 396.4172 - 0.0059\,\Delta q_{Well\,P3} - 0.0018\,\Delta q\,log\Delta t_{Well\,P3} \quad (\ 33\ )$$
$$- 1.339x10^{-06}\Delta q\,\Delta t_{Well\,P3} + 0.1899\,\Delta q/\Delta t_{Well\,P3}$$
$$+ 0.0004\Delta q_{Well\,I1} + 0.0001\,\Delta q\,log\Delta t_{Well\,I1}$$
$$+ 4.5333x10^{-07}\,\Delta q\,\Delta t_{Well\,I1} - 0.0145\,\Delta q/\Delta t_{Well\,I1}\,.$$

Figure 37 presents pressure predictions for the Well P3/Well P2 configuration. The R^2_{train} is equal to 0.9819 and R^2_{test} is equal to 0.653.

Although the Well P3/Well P2 configuration has a slightly low R^2_{test} value compared to the base case, the pressure increase and decrease trends between $1.5x10^4$ and $1.8x10^4$ hours and $6x10^4$ and $6.2x10^4$ hours are captured in pressure predictions by adding Well P2 flowrate history. The configuration does not have the highest R^2 value but has the best visual fit. Therefore, we assume that there might be interference effects between these wells.
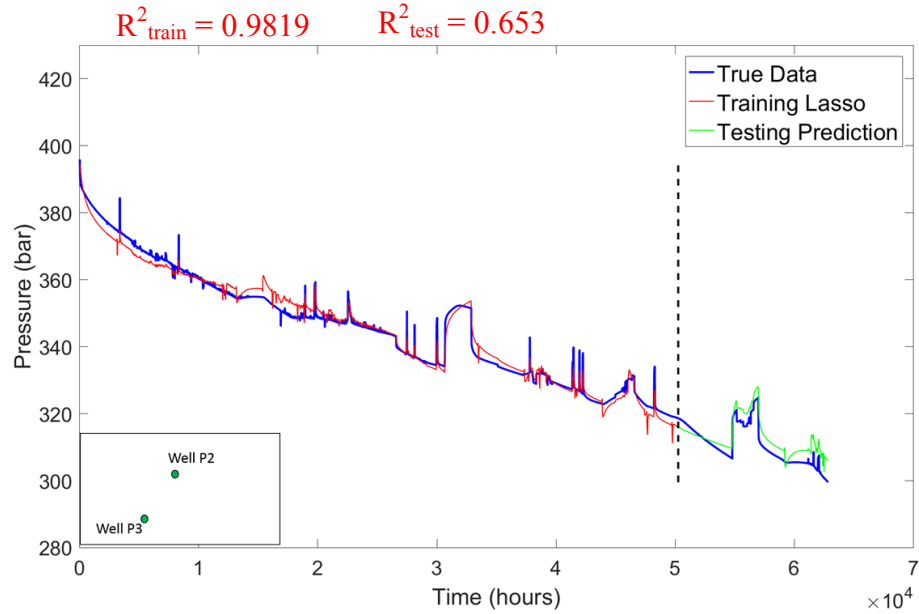
Figure 37 — Pressure prediction for the Wells P3 and P2 configuration using the lasso algorithm. The pressure increase and decrease between $1.5\,x10^4$ and $1.8x10^4$ hours and $6\,x10^4$ and $6.2x10^4$ hours in real bottom-hole pressure values are captured in pressure predictions by adding Well P2 flowrate history. R^2_{train} is equal to 0.9819 and R^2_{test} is equal to 0.653.

Eq. 34 presents the linear model for the Well P3/Well P2 configuration. Coefficients are calculated using the lasso algorithm.

$$P_{bh\,Well\,P3} = 395.791 + 0.0014\,\Delta q_{Well\,P3} - 0.0039\,\Delta q\,log\Delta t_{Well\,P3} \quad (\ 34\ )$$
$$- 1.185x10^{-06}\,\Delta q\,\Delta t_{Well\,P3} + 0.0767\Delta q/\Delta t_{Well\,P3}$$
$$+ 0.0191\Delta q_{Well\,P2} - 0.0068\,\Delta q\,log\Delta t_{WellP\,2}$$
$$+ 7.205x10^{-07}\,\Delta q\,\Delta t_{Well\,P2} - 0.2477\Delta q\Delta t_{Well\,P2}$$

Figure 38 presents pressure predictions for the Well P3/Well P1 configuration. R^2_{train} is equal to 0.9875 and R^2_{test} is equal to -0.2108.

The Well P3/Well P1 configuration has a low R^2_{test} value and no pressure prediction trends are captured by adding Well P1 compare to base case. On the contrary, the lasso algorithm reproduces a pressure increase and a decrease between $0.5x10^4$ and $0.8x10^4$ hours and a pressure increase at $5.1x10^4$ hours by adding Well P1 flowrate history which is not observed in real pressure values. Therefore, we cannot assume that there are interference effects between these wells.

$R^2_{train} = 0.9875$ $R^2_{test} = -0.2108$
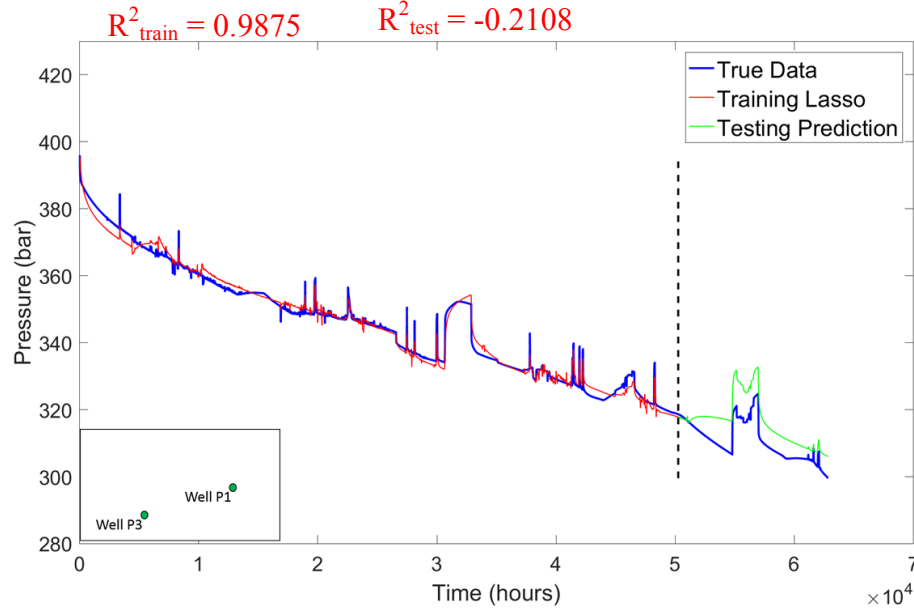
Figure 38 — Pressure prediction for the Wells P3 and P1 configuration using the lasso algorithm. Compare to the base case, there are not pressure trends which are captured by adding Well P1. The lasso algorithm accurately reproduce pressure increase and decrease trend between $0.5x10^4$ and $0.8x10^4$ hours and pressure increase at $5.1 \ x10^4$ hours by adding Well P1 flowrate history. R^2_{train} is equal to 0.9875 and R^2_{test} is equal to -0.2108.

Eq. 35 presents the linear model for the Well P3/Well P1 configuration. Coefficients are calculated using the lasso algorithm.

$$P_{bh \ Well \ P3} = 398.6474 - 0.0004 \ \Delta q_{Well \ P3} - 0.0036 \ \Delta q \ log\Delta t_{Well \ P3} \quad (\ 35 \ )$$
$$- 8.074x10^{-06} \ \Delta q \ \Delta t_{Well \ P3} + 0.1092 \ \Delta q/\Delta t_{Well \ P3}$$
$$+ \ 0.0219 \ \Delta q_{Well \ P1} - 0.0087 \ \Delta q \ log\Delta t_{Well \ P1}$$
$$+ 3.35x10^{-07} \ \Delta q \ \Delta t_{Well \ P1} - 0.2890 \ \Delta q/\Delta t_{Well \ P1} \ .$$

Figure 39 presents pressure predictions for the Well P3/Well I2 configuration. R^2_{train} is equal to 0.9864 and R^2_{test} is equal to 0.5924.

The Well P3/Well I2 configuration has a low R^2_{test} value and no visual improvements in pressure prediction trends are captured by adding Well I2 compare to base case. Therefore, there is no obvious evidence of interference effects between these wells.

67
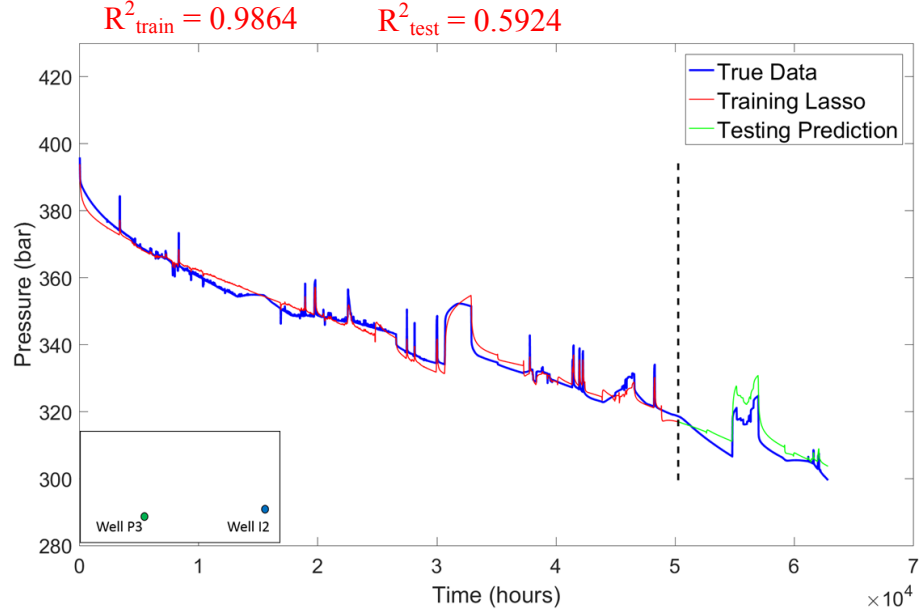
Figure 39 — Pressure prediction for the Wells P3 and I2 configuration using the lasso algorithm. R^2_{train} is equal to 0.9864 and R^2_{test} is equal to 0.5924. Compare to the base case, there are not visual improvements in pressure trend predictions and in R^2 values by adding Well I2.

Eq. 36 presents the linear model for the Well P3/Well I2 configuration. Coefficients are calculated using the lasso algorithm.

$$
\begin{aligned}
P_{bh\,Well\,P3} = {}& 394.9968 + 0.0004\,\Delta q_{Well\,P3} - 0.004\,\Delta q\,log\Delta t_{Well\,P3} \qquad (\ 36\ ) \\
& - 9.679x10^{-07}\,\Delta q\,\Delta t_{Well\,P3} + 0.1047\,\Delta q/\Delta t_{Well\,P3} \\
& + 0.0063\,\Delta q_{Well\,I2} - 0.0019\,\Delta q\,log\Delta t_{Well\,I2} \\
& + 3.899x10^{-07}\,\Delta q\,\Delta t_{Well\,I2} - 0.1056\,\Delta q/\Delta t_{Well\,I2}\,.
\end{aligned}
$$

During the generated scenarios, we assume that Wells P2 and I1 might interfere with Well P3. Therefore, we generate the Well P3/Well I1/Well P2 configuration scenario to examine pressure predictions. For this scenario, R^2_{train} is equal to 0.9988 and R^2_{test} is equal to 0.953.

Figure 40 presents pressure predictions for this scenario. The Well P3/Well I1/Well P2 configuration has the highest training and testing R^2 values relative to the other scenarios and provides an almost perfect visual fit. Therefore, we assume that the flowrate histories for Wells P3, P2, and I1 are required to accurately reproduce the bottom-hole pressure of Well P3.

$$R^2_{train} = 0.9988 \qquad R^2_{test} = 0.953$$
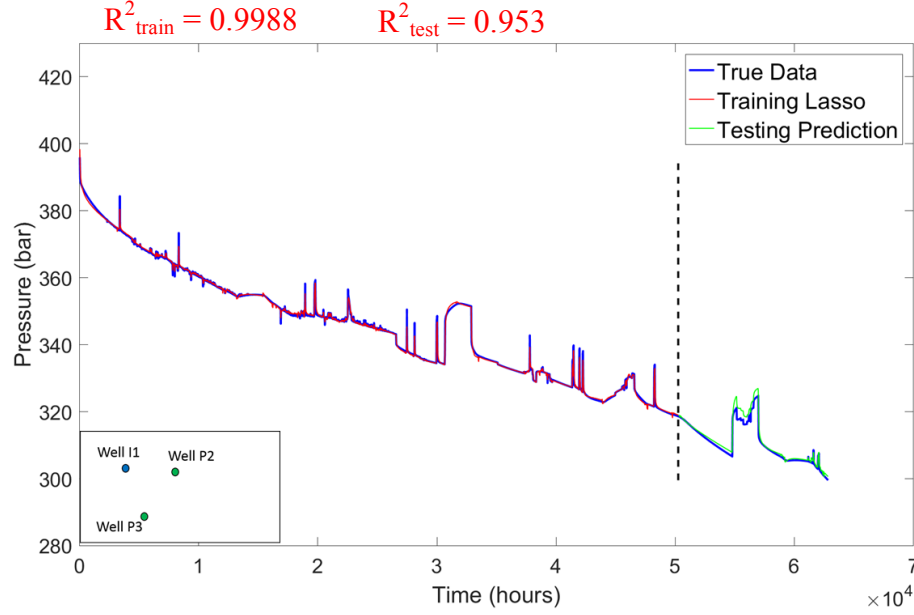
Figure 40 — Pressure predictions for Wells P3, P2 and I1 configuration using the lasso algorithm. R^2_{train} is equal to 0.9988 and R^2_{test} is equal to 0.953. There is almost perfect match between real bottom-hole pressure values and pressure predictions for Wells P3, P2 and I1 configuration. This configuration has highest R^2 values compare to other scenarios.

Eq. 37 presents the linear model for the Well P3/Well P2/ Well I1 configuration using the lasso algorithm.

$$
\begin{aligned}
P_{bh\,Well\,P3} = {} & 399.1643 - 0.0065\,\Delta q_{Well\,P3} - 0.0015\,\Delta q\,log\Delta t_{Well\,P3} \qquad (\;37\;) \\
& - 0.7485x10^{-06}\,\Delta q\,\Delta t_{Well\,P3} + 0.1969\,\Delta q/\Delta t_{Well\,P3} \\
& - 0.0017\,\Delta q_{Well\,I1} - 0.0007\,\Delta q\,log\Delta t_{Well\,I1} \\
& + 0.4417x10^{-06}\,\Delta q\,\Delta t_{Well\,I1} + 0.0218\,\Delta q/\Delta t_{Well\,I1} \\
& + 0.0064\,\Delta q_{Well\,P2} - 0.0025\,\Delta q\,log\Delta t_{Well\,P2} \\
& - 0.6403x10^{-06}\,\Delta q\,\Delta t_{Well\,P2} - 0.0742\,\Delta q/\Delta t_{Well\,P2}\,.
\end{aligned}
$$

In **Figure 41.a,** we present an enlarged view of Well P3, where we have reproduced and forecasted pressure responses using only the flowrates for Well P3. In **Figure 41.b** we repeat it for the Well P3/Well I1/Well P2 configuration. Visually and statistically, the Well P3/Well I1/Well P2 configuration is substantially better than the Well P3 configuration.

69
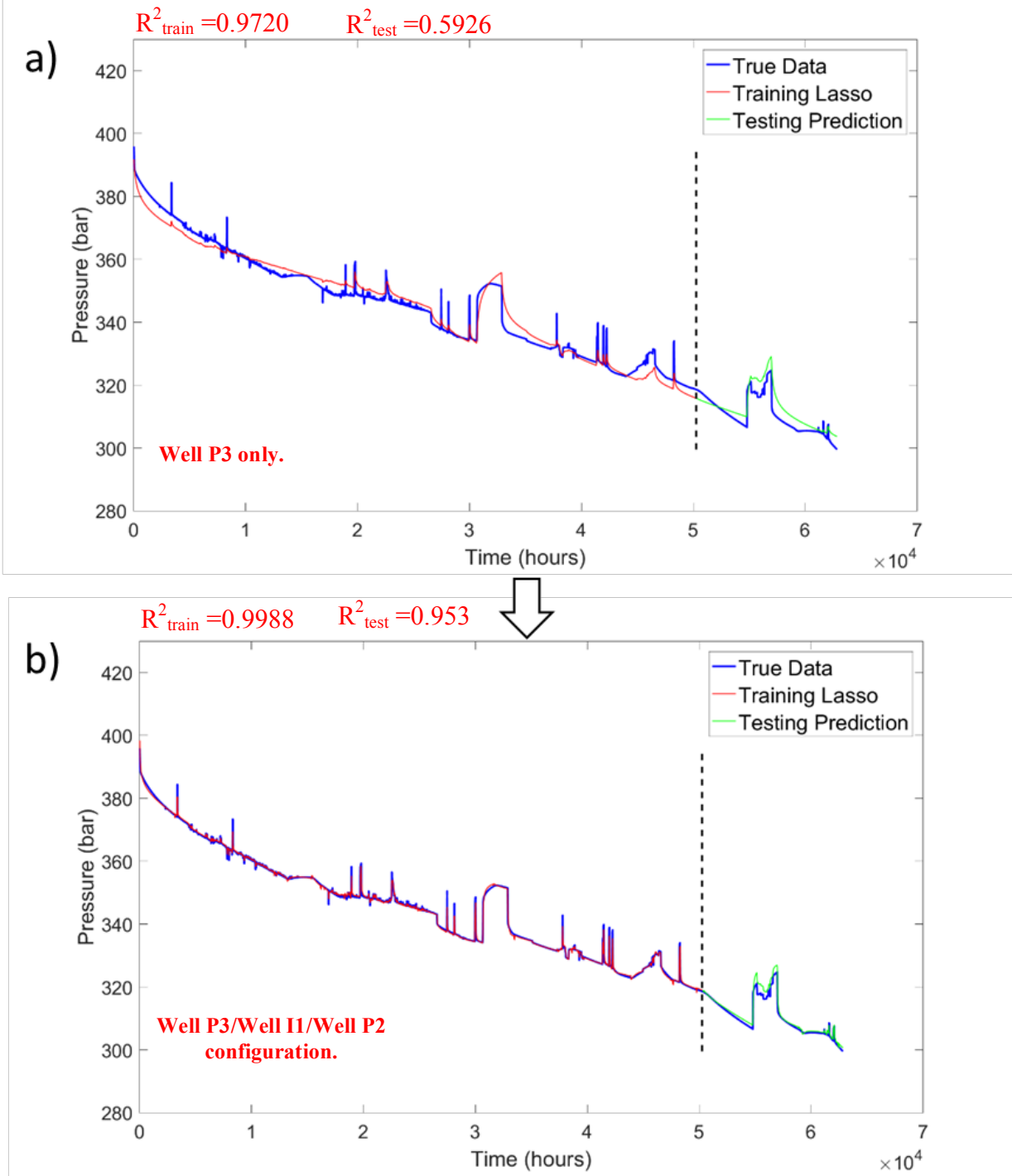
Figure 41 — Figure 41.a illustrates the pressure prediction obtained using only the flowrate history for Well P3. Figure 41.b illustrates the improved pressure prediction obtained using the flowrate histories for Wells P3, I1, and P2 configuration.

Table 7 presents the coefficients of lasso model for predicting Well P3's bottom-hole pressure when all possible interference wells are considered. The estimated lambda value for the lasso algorithm is 1.6×10^{-3}. As mentioned, the coefficients are calculated using criteria in Eq. 16 which is the OLS approach with penalty term. Because of its penalty term, the lasso can assign coefficients of statistically unimportant features to zero. By doing so, the lasso can identify statistically important features.

In **Table 7**, features whose coefficients are smaller value than 0.0001 are considered as statistically unimportant.

It can be seen that $\Delta q \, \Delta t$ terms for all wells have less statistical significance for predicting the bottom-hole pressure of Well P3. The lasso coefficients indicate that only Well I2's flowrate features are not statistically important when reproducing Well P3's bottom-hole pressure history. For Well P1, two features out of four are statistically unimportant.

From the synthetic model, we know that there is no communication between Well P3 and Well P1. As two features of Well P1 are considered statistically important, we cannot conclude that lasso coefficients are fully able to determine which flowrate features have importance in reproducing the bottom-hole pressure response of Well P3.

Table 7 — The coefficients of the lasso model for predicting Well P3's bottom-hole pressure when all possible interference wells are considered. The features having coefficients less than 0.0001 are considered as statistically unimportant, indicated by red color.

	Coefficient	Statistical Importance
Intercept	398.892	
Δq_{P3}	-0.007	
$\Delta q log \Delta t_{P3}$	-0.001	
$\Delta q \Delta t_{P3}$	-8.193e-07	■■■
$\Delta q / \Delta t_{P3}$	0.204	
Δq_{I1}	-0.0011	
$\Delta q log \Delta t_{I1}$	0.0004	
$\Delta q \Delta t_{I1}$	4.7647e-07	■■■
$\Delta q / \Delta t_{I1}$	0.014	
Δq_{P2}	0.006	
$\Delta q log \Delta t_{P2}$	-0.0025	
$\Delta q \Delta t_{P2}$	-6.332e-07	■■■
$\Delta q / \Delta t_{P2}$	-0.076	
Δq_{P1}	0.00029	
$\Delta q log \Delta t_{P1}$	-1.381e-05	■■■
$\Delta q \Delta t_{P1}$	8.389e-08	■■■
$\Delta q / \Delta t_{P1}$	-0.0068	
Δq_{I2}	4.893e-05	■■■
$\Delta q log \Delta t_{I2}$	0	■■■
$\Delta q \Delta t_{I2}$	-2.132e-08	■■■
$\Delta q / \Delta t_{I2}$	0	■■■

Recall that in Case 1 when we include all possible offset wells, the estimated lambda value for lasso algorithm is 1.6×10^{-3} which is a small value. As mentioned in Chapter 2.4, a small lambda value indicates that the lasso model performs like the linear regression model. Therefore, we can ignore the penalty term in Eq. 16 and use only the OLS approach.

We use the linear regression model to analyze the statistical parameters and to see if we can determine interference effects using statistical parameters such as standard error, t-statistics, and p-value. First, we remove all statistically unimportant parameters which are identified by the lasso algorithm in Table 7. Then we input the remaining features into the linear regression model to determine if p-values can identify which flowrate features do not affect the bottom-hole pressure response of Well P3.

Figure 8 presents statistical parameters of the least squares regression model. The statistical parameters such as standard error, t-statistics, and p-value are explained in Chapter 2.3. As all the p-values are smaller than 0.01, all flowrate features have association with Well P3's bottom-hole pressure.

However, we know from the synthetic model that Well P1's flowrate history does not affect Well P3's bottom-hole pressure response. Therefore, the p-values are not able to fully determine which well's flowrate history does not impact the bottom-hole pressure response of Well P3.

We note that there is a 5% chance that a p-value will incorrectly determine that there is a relationship (James 2013).

Table 8   —   The statistical parameters of the least square model for predicting Well P3's bottom-hole pressure when all statistical important features are included. The small p-values indicate that flowrate features of Wells P3, I1, P2 and P1 are statistically significant to forecast bottom-hole pressure.

	Coefficient	**Std. error**	**t-statistic**	**p-value**
Intercept	404.06	0.71345	566.35	<0.01
Δq_{P3}	0.016784	0.016784	19.35	<0.01
$\Delta q log \Delta t_{P3}$	-0.010381	0.00028851	-35.983	<0.01
$\Delta q / \Delta t_{P3}$	-0.11768	-0.11768	-6.8157	<0.01
Δq_{I1}	0.011116	0.00042781	25.984	<0.01
$\Delta q log \Delta t_{I1}$	-0.003845	0.00011893	-32.33	<0.01
$\Delta q / \Delta t_{I1}$	-0.23435	0.014301	0.014301	<0.01
Δq_{P2}	0.013265	0.0009288	14.281	<0.01
$\Delta q log \Delta t_{P2}$	-0.0045304	0.00032113	-14.108	<0.01
$\Delta q / \Delta t_{P2}$	-0.1922	0.0018874	12.183	<0.01
Δq_{P1}	-0.0056011	0.00049933	-14.349	<0.01
$\Delta q / \Delta t_{P1}$	-0.24054	0.0249469	-9.8624	<0.01

Case 2 — Guler Well P1

In case 2, we know from the simulated reservoir model that Well P1 and Well I2 are pressure communicating wells. We note that there is no communication between Well P1 and Wells P2, P3, and I1 because of the sealing fault.

This this case, we present the scenarios or realizations generated by the lasso algorithm by considering the combinations of Well P1 and its offsets. As in Case 1, we compare the improvements in pressure predictions for different scenarios with the base case using the R^2 values and visual fit.

Figure 42 presents pressure predictions for the Well P1/Well I2 configuration. The R^2_{train} is equal to 0.9655 and R^2_{test} is equal to 0.8857.

Visually and statistically, the Well P1/Well I2 configuration performs noticeably better than the Well P1 configuration. We note that the Well P1/Well I2 configuration captures particularly well the start of the pressure increase, which occurs at approximately 2.4×10^4 hours. Also, there are visual improvements in pressure prediction trends between 2×10^4 and 3×10^4 hours. The R^2 values for both training and testing increase significantly by adding Well I2's flowrate history. Therefore, we assume that there might be interference effects between these wells.
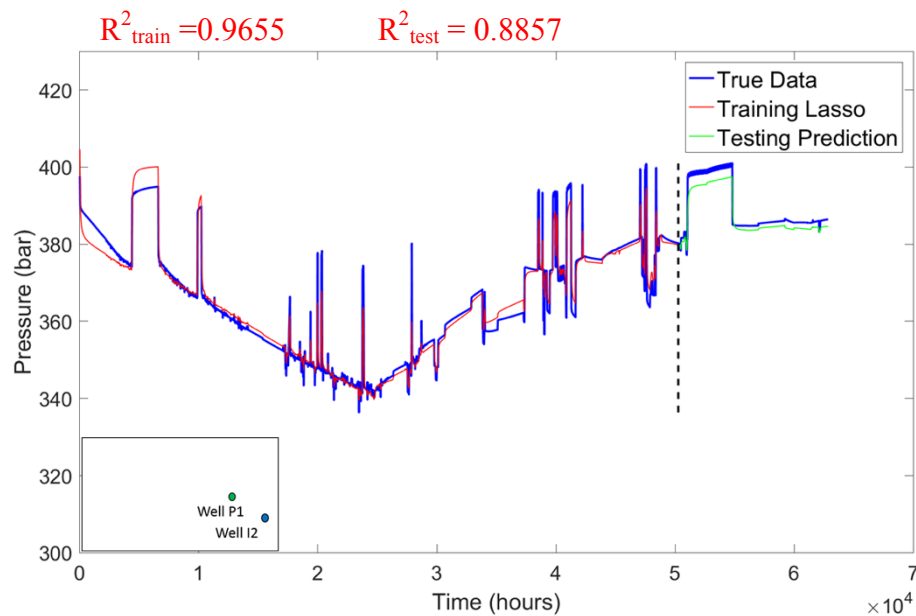
Figure 42 — Pressure predictions for the Well P1/Well I2 configuration using the lasso algorithm. R^2_{train} is equal to 0.9655 and R^2_{test} is equal to 0.8857. Compare to the base case, there are visual improvements in pressure prediction trends between 1.8×10^4 and 3×10^4 hours. New configuration captures particularly well the start of the pressure increase, which occurs at approximately 2.4×10^4 hours.

Eq. 38 presents linear model for this configuration using the lasso algorithm

$$P_{bh\,Well\,P1} = 407.7766 - 0.0281\,\Delta q_{Well\,P1} - 0.0006\,\Delta q\,log\Delta t_{Well\,P1} \quad (\ 38\ )$$
$$- 1.852x10^{-06}\,\Delta q\,\Delta t_{Well\,P1} + 0.6088\,\Delta q/\Delta t_{Well\,P1}$$
$$+ 0.0019\Delta q_{Well\,I2} + 0.0004\,\Delta q\,log\Delta t_{Well\,I2}$$
$$+ 1.2117x10^{-06}\,\Delta q\,\Delta t_{Well\,I2} - 0.0350\,\Delta q/\Delta t_{Well\,I2}\,.$$

Figure 43 presents pressure predictions for the Well P1/Well P3 configuration. The R^2_{train} is equal to 0.8451 and R^2_{test} *is* equal to -1.7832**.**

The Well P1/Well P3 configuration has decrease in testing and training R^2 values compare to base case. Besides, there are not any visual improvements on pressure trends by adding Well P1's flowrate history. Therefore, we do not have evidence to conclude that Well P3's flowrate history affect Well P1's bottom-hole pressure response.
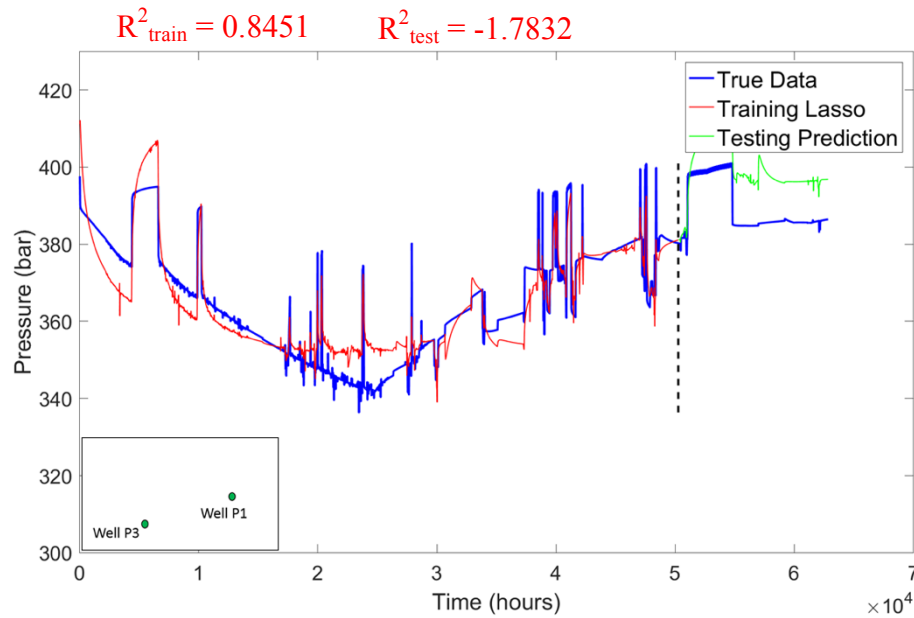
Figure 43 — Pressure prediction for the Well P1/Well P3 configuration using the lasso algorithm. R^2_{train} is equal to 0.8451 and R^2_{test} is equal to -1.7832. There is not improvements in match and R^2_{test} by adding Well P3's flowrate history.

Eq. 39 presents linear model for the Well P1/Well P3 configuration. Coefficients are calculated using the lasso algorithm

$$P_{bh\,Well\,P1} = 422.4716 + 0.0229\,\Delta q_{Well\,P1} - 0.0220\,\Delta q\,log\Delta t_{Well\,P1} \quad (\,39\,)$$
$$+\,0\,\Delta q\,\Delta t_{Well\,P1} - 0.0661\,\Delta q/\Delta t_{Well\,P1}$$
$$+\,0.0223\Delta q_{Well\,P3} - 0.0079\,\Delta q\,log\Delta t_{Well\,P3}$$
$$+\,4.1286x10^{-07}\,\Delta q\,\Delta t_{Well\,P3} - 0.3155\,\Delta q/\Delta t_{Well\,P3}\,.$$

Figure 44 presents pressure predictions for the Well P1/Well P2 configuration. R^2_{train} is equal to 0.8463 and R^2_{test} is equal to -5.8755.

The Well P1/Well P2 configuration has significantly low R^2_{train} and R^2_{test} values compare to base case. Moreover, there are no improvements in the visual fit. Therefore, we assume that there we do not have evidence of interference effects between these wells.
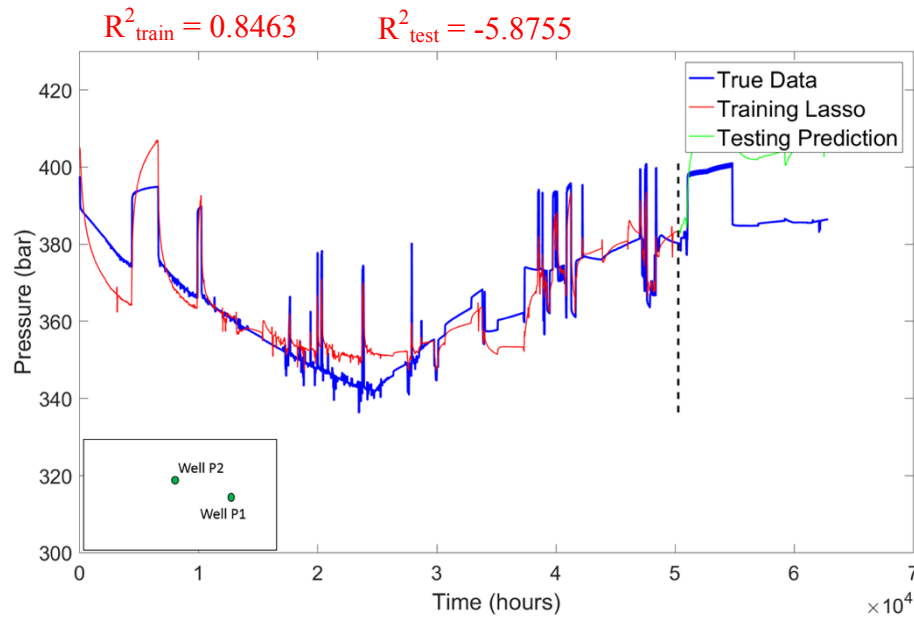
Figure 44 — Pressure prediction for the Well P1/Well P2 configuration using the lasso algorithm. R^2_{train} is equal to 0.8463 and R^2_{test} is equal to -5.8755. There are not improvements on R^2 values and in visual fit of pressure predictions by adding Well P2's flowrate history.

Eq. 40 presents the linear model for the P1/Well P2 configuration. Coefficients are calculated using the lasso algorithm.

$$P_{bh\,Well\,P1} = 407.7766 - 0.0281\,\Delta q_{Well\,P1} - 0.0006\,\Delta q\,log\Delta t_{Well\,P1} \quad (\;40\;)$$
$$- 1.623x10^{-06}\,\Delta q\,\Delta t_{Well\,P1} + 0.6088\,\Delta q/\Delta t_{Well\,P1}$$
$$+ 0.0019\Delta q_{Well\,P2} + 0.0004\,\Delta q\,log\Delta t_{Well\,P2}$$
$$+ 1.5237x10^{-06}\,\Delta q\,\Delta t_{Well\,P2} - 0.0350\,\Delta q/\Delta t_{Well\,P2}\,.$$

Figure 45 presents pressure predictions for the Well P1/Well I1 configuration. R^2_{train} is equal to 0.9329 and R^2_{test} *is* equal to -3.7455.

The Well P3/Well I1 configuration has a low R^2_{test} value and no visual improvements in pressure prediction trends are captured by adding Well I1. Therefore, there is not obvious evidence of interference effects between these wells.
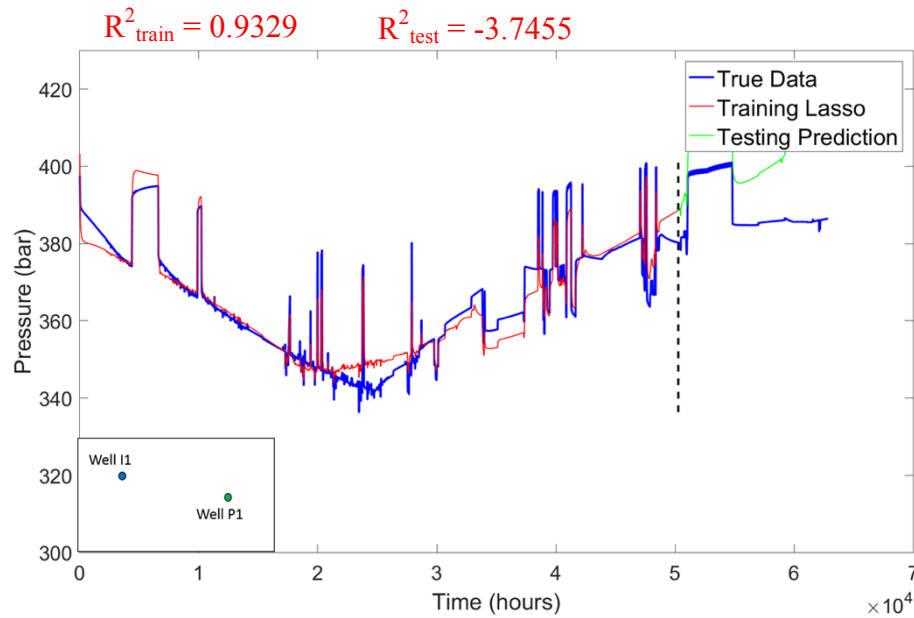
Figure 45 — Pressure prediction for the Well P1/Well I1 configuration using the lasso algorithm. R^2_{train} is equal to 0.9329 and R^2_{test} is equal to -3.7455. Compare to the base case, there are not visual and R^2_{test} value improvements in pressure trend predictions by adding Well I1.

Eq. 41 presents the linear model for the Well P3/Well I2 configuration. Coefficients are calculated using the lasso algorithm.

$$P_{bh\,Well\,P1} = 407.7766 - 0.0281\,\Delta q_{Well\,P1} \;-\; 0.0006\,\Delta q\,log\Delta t_{Well\,P1} \qquad (\;41\;)$$
$$- 2.38x10^{-06}\,\Delta q\,\Delta t_{Well\,P1} + 0.6088\,\Delta q/\Delta t_{Well\,P1}$$
$$+ \;0.0019\Delta q_{Well\,I1}\; + 0.0004\,\Delta q\,log\Delta t_{Well\,I1}$$
$$+ \;1.351x10^{-06}\,\Delta q\,\Delta t_{Well\,12} - 0.0350\,\Delta q/\Delta t_{Well\,I1}\,.$$

As mentioned, the lasso algorithm forces statistically unimportant features to zero. We assume the features whose coefficients are smaller value than 0.0001 are statistically unimportant to reproduce the bottom-hole pressure of Well P1.

From **Table 9**, it can be seen that $\Delta q\,\Delta t$ terms for all wells are statistical insignificance for predicting the bottom-hole pressure of Well P1. The lasso coefficients indicate that three flowrate features of Well I1 are statistically unimportant when reproducing Well P1's bottom-hole pressure history. While, the three flowrate features of Wells P2 and P3 are found to be statistically important.

From simulated model we know that Well P1 does not have connectivity with Wells P2, P3 and I1. Well P1 has connectivity only with Well I2.

Therefore, the lasso coefficients we can-not accurately identify which flowrate history affects the bottom-hole pressure of Well P1 when all wells are examined at the same time. Hence, the coefficient are not fully reliable to identify interference effects.

Table 9 — The coefficients of the lasso model for predicting Well P1's bottom-hole pressure when all possible interference wells are considered. The features having coefficients less than 0.0001 are considered as statistically unimportant, indicated by red color.

	Coefficient	Statistical Importance
Intercept	410.852	
Δq_{P1}	-0.047	
$\Delta qlog\Delta t_{P1}$	0.0069	
$\Delta q\Delta t_{P1}$	-2.342e-06	🟥
$\Delta q/\Delta t_{P1}$	0.858	
Δq_{I2}	-0.0003	
$\Delta qlog\Delta t_{I2}$	0.0002	
$\Delta q\Delta t_{I2}$	8.650e-07	🟥
$\Delta q/\Delta t_{I2}$	0.0041	
Δq_{P2}	0.0065	
$\Delta qlog\Delta t_{P2}$	-0.0024	
$\Delta q\Delta t_{P2}$	1.9594e-07	🟥
$\Delta q/\Delta t_{P2}$	-0.0838	
Δq_{P3}	0.0060	
$\Delta qlog\Delta t_{P3}$	-0.0023	
$\Delta q\Delta t_{P3}$	-1.8994e-07	🟥
$\Delta q/\Delta t_{P3}$	-0.0787	
Δq_{I1}	-3.924e-05	🟥
$\Delta qlog\Delta t_{I1}$	0	🟥
$\Delta q\Delta t_{I1}$	5.3527e-07	🟥
$\Delta q/\Delta t_{I1}$	0.0079	

Recall that in Case 2, for the lasso algorithm, the estimated lambda value is 4.7×10^{-3} which is relatively a small value. As mentioned in Chapter 2.4, a small lambda value indicates that we can ignore penalty term in Eq. 16 and use only the OLS approach.

Therefore, we use the linear regression model to analyze the statistical parameters and to see if we can determine interference effects using these parameters. As in case 1, first, we remove all statistically unimportant parameters which are identified by lasso algorithm in Table 9. Then we input the remaining features into the linear regression model to determine if p-values can identify flowrate features which do not affect bottom-hole pressure response.

Table 10 presents statistical parameters of the least squares regression model. When p-values are smaller than 0.01, we assume that there is an association with Well P1's bottom-hole pressure and features.

According to p-values in Table 10, three flowrate features of Well P2 and two flowrate features of Well P3 are statistically important to reproduce bottom-hole pressure of Well P1.

However, we know from the synthetic model that flowrate histories for Wells P3, P1 and I1 do not affect Well P1's bottom-hole pressure response. Therefore, for Case 2, the p-values are not fully able to determine which well's flowrate features do not impact the bottom-hole pressure response of Well P1.

As can be seen from results and discussion of Case 1 and Case 2, the lasso coefficients and the statistical parameters, particularly p-values, are unreliable for identifying interference effects. Therefore, we will not rely on statistical parameters for the other cases.

Table 10 — The statistical parameters of the least square model for predicting Well P1's bottom-hole pressure when all possible interference wells are included. The small p-values indicate that almost all flow regimes (features) for different wells are statistically significant to forecast bottom-hole pressure.

	Coefficient	Std. error	t-statistic	p-value
Intercept	411.03	0.77165	544.5	<0.01
Δq_{P1}	0.022435	0.001526	14.702	<0.01
$\Delta qlog\Delta t_{P1}$	-0.021133	0.00048962	-43.163	<0.01
$\Delta q/\Delta t_{P1}$	-0.049772	0.032902	-1.5128	0.13> 0.01
Δq_{I2}	-0.009332	0.00047816	-40.323	<0.01
$\Delta qlog\Delta t_{I2}$	0.29037681	0.00014772	44.41	<0.01
$\Delta q/\Delta t_{I2}$	0.023592	0.013275	13.02	<0.01
Δq_{P2}	0.0085376	0.00086194	7.34	<0.01
$\Delta qlog\Delta t_{P2}$	-0.0032415	0.00031135	-8.411	<0.01
$\Delta q/\Delta t_{P2}$	-0.13883	0.013846	-5.8602	<0.01
Δq_{P3}	0.0049963	0.00078115	3.6762	<0.01
$\Delta qlog\Delta t_{P3}$	-0.00083476	0.00027648	-2.4913	0.13> 0.01
$\Delta q/\Delta t_{P3}$	-0.078535	0.012537	-4.2643	<0.01
$\Delta q/\Delta t_{I1}$	-0.025524	0.011494	-1.62088	0.07 > 0.01

5.2.2 Field Example — Yusif Field

The Yusif field is the large and geologically complex field as stated in Chapter 4. The Yusif field consists of an alternative sequence of thin impermeable and permeable layers. In this section, we show the application of the methodology to real field data. We compare identified interference effects from offset wells using the methodology with observed results from the

field. Among all wells in the field, we test only possible interference wells which are selected considering well locations and perforation layers.

Case 3 — Yusif Well 12

In case 3, we chose Wells 38I, 24, and 33 as possible interference wells for Well 12 and analyzed their impact to bottom-hole pressure measurements. Wells 12 and 24 were completed in L5 layer while Well 38I was completed in L3 and L5; Well 33 was completed in L5, L7 and L8. From the dynamic field observations, we know that there are no interference effects from Wells 24 and 33.

From the dynamic observations of field, there are not reported interference effects from offset wells for Well 12. In fact, from the field reports it is indicated that there is not communication between Well 12 and Wells 24 and 33.

Although only flowrate history of Well 12 can accurately reproduce its bottom-hole pressure response using the lasso algorithm and no discrepancy is observed, in this section we add possible offset wells systematically to examine the algorithm behavior.

In **Figure 46** we present the bottom-hole pressure response for Well 12 and flowrate histories for Wells 12, 38I, 24, and 33.
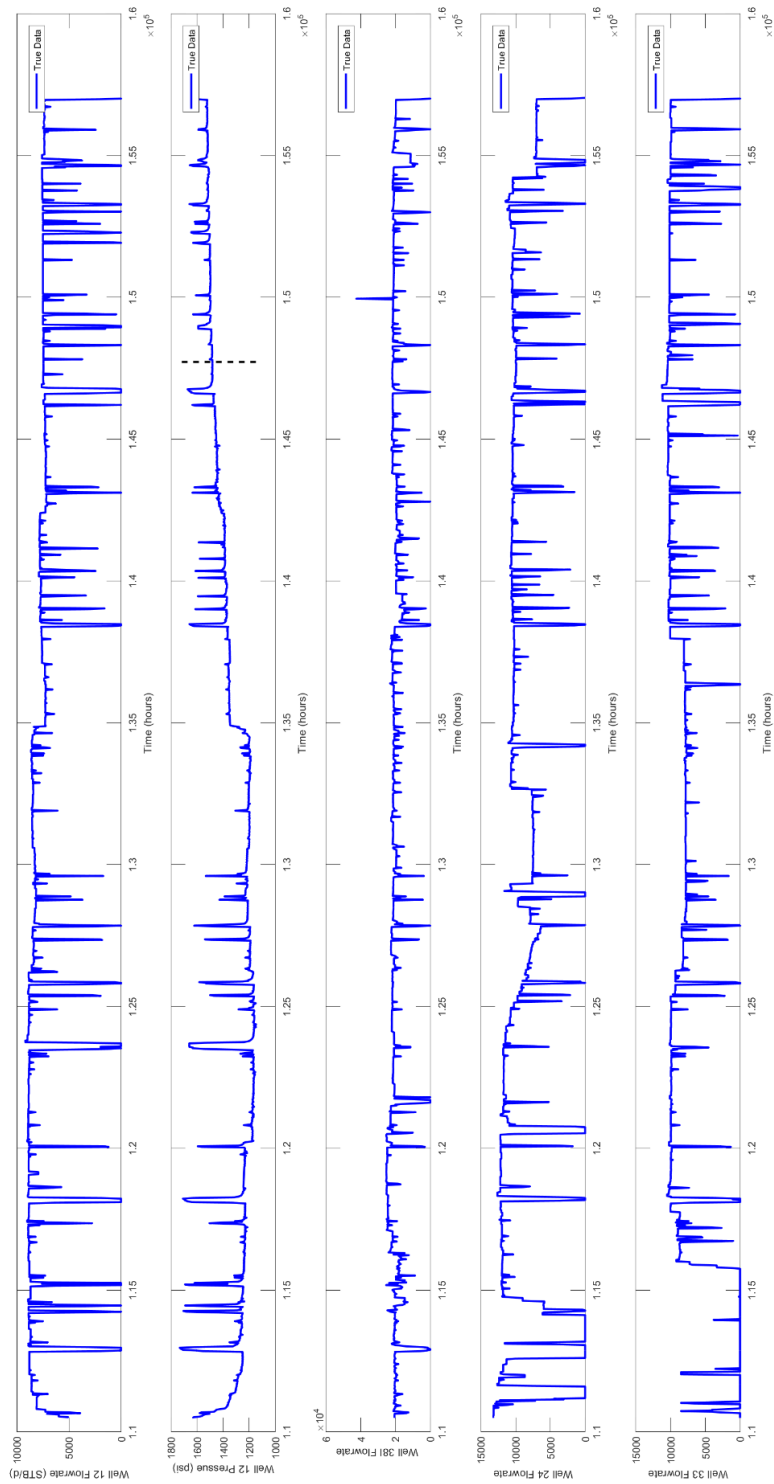
Figure 46 — Flowrate and bottom-hole pressure history for Well 12 and flowrate histories for Wells 38I, 24, and 33.

Figure 47 presents pressure predictions for the Well 12/Well 38I configuration. R^2_{train} is equal to 0.9036 and R^2_{test} is equal to -3.2595.

The Well 12/Well 38I configuration has a lower R^2_{test} value compare to base case where only Well 12 flowrate's is used to predict its bottom-hole pressure response. Besides, there are not any improvements in capturing pressure trends by adding Well 38I's flowrate history. Therefore, we do not have evidence to conclude that Well 38I's flowrate history affect Well 12's bottom-hole pressure response.
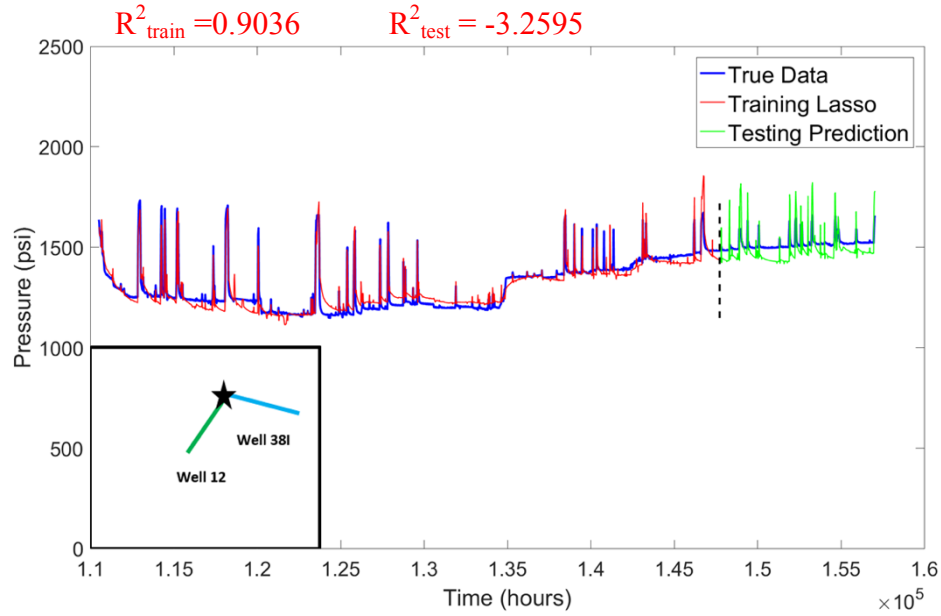
Figure 47 — Pressure predictions for the Well 12/Well 38I configuration using the lasso algorithm. R^2_{train} is equal to 0.9036 and R^2_{test} is equal to -3.2595. Compare to the base case, there are not visual improvements in pressure predictions.

Eq. 42 presents linear model for the Well 12/Well 38I configuration. Coefficients are calculated using the lasso algorithm.

$$
\begin{aligned}
P_{bh\,WellP3} = {}& 2647.546 - 0.0448\,\Delta q_{Well\,12} - 0.0455\,\Delta q\,log\Delta t_{Well\,12} \quad\quad (\ 42\ ) \\
& + 2.1720x10^{-06}\,\Delta q\,\Delta t_{Well\,12} - 0.63496\,\Delta q/\Delta t_{Well\,12} \\
& + 0.0097\Delta q_{Well\,38I} - 0.0032\,\Delta q\,log\Delta t_{Well\,38I} \\
& - 5.2468x10^{-07}\,\Delta q\,\Delta t_{Well\,38I} - 0.1231\Delta q/\Delta t_{Well\,38I}\,.
\end{aligned}
$$

85

Figure 48 presents pressure predictions for the Well 12/Well 24 configuration. R^2_{train} is equal to 0.9053 and R^2_{test} is equal to -3.3079.

The Well 12/Well 24 configuration has a decrease in testing R^2 value compare to base case. Also there are not any improvements in capturing the pressure trends by adding Well 24's flowrate history. Therefore, we do not have evidence to conclude that Well 24's flowrate history affect Well 12's bottom-hole pressure response.

Figure 48 — Pressure prediction for the Well 12/Well 24 configuration using the lasso algorithm. R^2_{train} is equal to 0.9053 and R^2_{test} is equal to -3.3079. There is not improvements in pressure prediction by adding Well 24's flowrate history.

Eq. 43 presents linear model for the Well 12/Well 24 configuration using the lasso algorithm.

$$P_{bh\,WellP3} = 2776.5770 + 0.0542\,\Delta q_{Well12} - 0.04901\,\Delta q\,log\Delta t_{Well\,12} \quad (\ 43\ )$$
$$+ 1.9479x10^{-06}\,\Delta q\,\Delta t_{Well\,12} - 0.7365\,\Delta q/\Delta t_{Well\,12}$$
$$- 0.0023\,\Delta q_{Well\,24} + 0.0019\Delta q\,log\Delta t_{Well\,24}$$
$$- 9.351x10^{-07}\,\Delta q\,\Delta t_{Well\,24} - 0.0169\,\Delta q/\Delta t_{Well\,24}.$$

Figure 49 presents pressure predictions for the Well 12/Well 33 configuration. R^2_{train} is equal to 0.9053 and R^2_{test} is equal to -3.3079.

The Well 12/Well 33 configuration has a slightly low R^2_{test} value when we are comparing with base case. Moreover, there are no improvements in visual fit. Therefore, we would not assume that there are not interference effects between these wells.

Figure 49 — Pressure prediction for the Well 12/Well 33 configuration using the lasso algorithm. R^2_{train} is equal to 0.9053 and R^2_{test} is equal to -3.3079. There are not improvements in visual fit of pressure predictions by adding Well 33's flowrate history.

Eq. 44 presents linear model for the P1/Well P2 configuration. Coefficients are calculated using the lasso algorithm.

$$
\begin{aligned}
P_{bh\,Well\,P3} = {} & 2376.6503 - 0.0009\,\Delta q_{Well\,12} - 0.0271\,\Delta q\,log\Delta t_{Well\,12} \quad\quad (\ \mathbf{44}\ ) \\
& - 1.732x10^{-07}\,\Delta q\,\Delta t_{Well\,12} - 0.1224\Delta q/\Delta t_{Well\,12} \\
& + 0.0395\,\Delta q_{Well\,33} - 0.0125\,\Delta q\,log\Delta t_{Well\,33} \\
& + 1.3274x10^{-06}\,\Delta q\,\Delta t_{Well\,33} - 0.5018\Delta q/\Delta t_{Well\,33}\,.
\end{aligned}
$$

In general, there are slight increase in the R^2_{train} values and decrease in the R^2_{test} values with additional well. The coefficients of the models gives statistical importance of flowrate histories

of nearby wells even if most probably there is no interference effects between wells. The most importantly, there are not any visual improvements in pressure prediction trends by adding possible offset wells. So, we assume that only Well 12's flowrate history is needed to accurately reproduce bottom-hole pressure behavior.

Case 4 — Yusif Well 10

In case 4, we chose Wells 22, 11, and 23 as possible interference wells for Well 10 and analyzed their impact to bottom-hole pressure measurements. Well 10 and Well 11 were completed in L5 layer while Well 22 was completed in L5 and L3 layers; Well 23 was completed in L7, L6, and L5 layers.

From dynamic field observation, we known that there are strong interference effects between Well 10 and Wells 22 and 11.

In **Figure 50** we illustrate the bottom-hole pressure responses for Well 10 and flowrate histories for Wells 10, 22, 11 and 23.
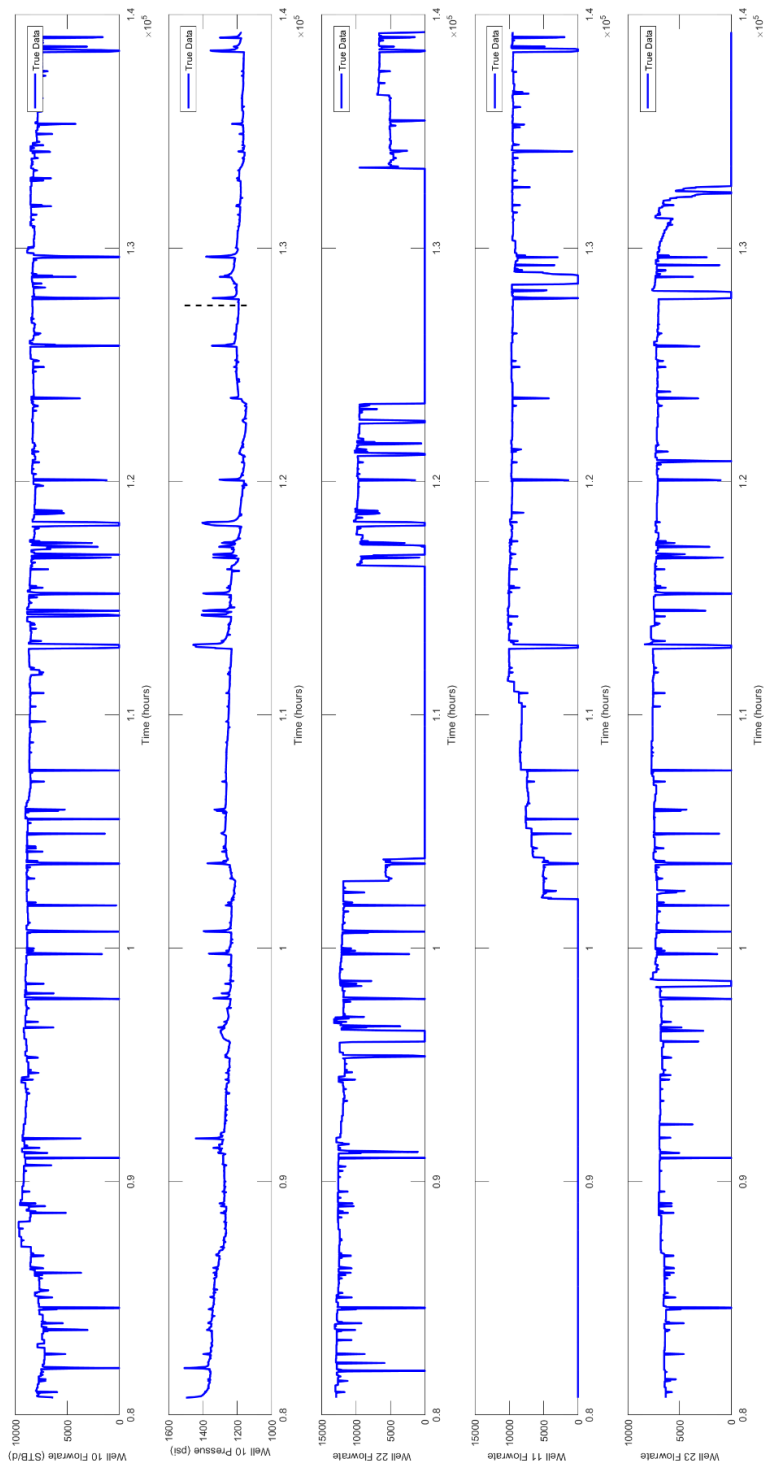
Figure 50 — Flowrate and bottom-hole pressure history for Well 10 and flowrate histories for Wells 22, 11, and 23.

Figure 51 presents pressure predictions for the Well 10/Well 22 configuration. R^2_{train} is equal to 0.9376 and R^2_{test} is equal to -0.3459.

With adding Well 22 flowrate history to predict bottom-hole pressure of Well 10, the R^2_{train} value increase significantly while the R^2_{test} value slightly decrease compare to base case. We observe significant improvement in visual fit when we add Well 22 flowrate history. Therefore, we assume that both Well 10 and Well 22 flowrate histories might affect bottom-hole pressure of Well 10. In another word, both wells flowrate history is required to accurately reproduce bottom-hole pressure behavior.
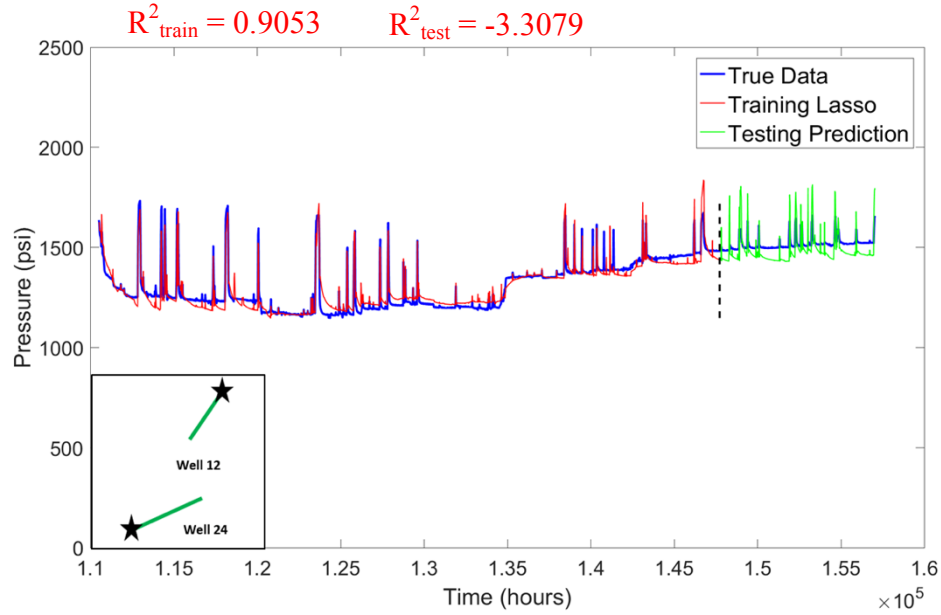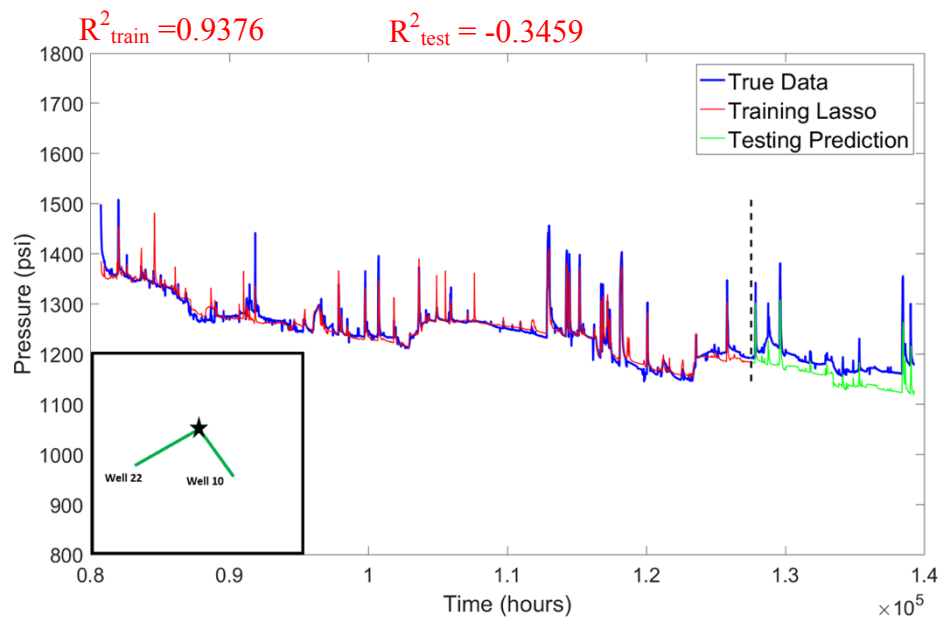

Figure 51 — Pressure predictions for the Well 10/Well 22 configuration using the lasso algorithm. R^2_{train} is equal to 0.9376 and R^2_{test} is equal to -0.3459. Compare to the base case, there are visual improvements in pressure prediction trends between 1×10^4 and 1.3×10^4 hours. This configuration captures particularly well increase in bottom-hole pressure at 1.01×10^4 and 1.2×10^4 hours.

Eq. 45 presents linear model for the Well 10/Well 22 configuration using the lasso algorithm.

$$P_{bh\,WellP3} = 2647.546 - 0.0448\,\Delta q_{Well\,12} - 0.0455\,\Delta q\,log\Delta t_{Well\,12} \qquad (\ 45\ )$$
$$+ 2.1720 x 10^{-06}\,\Delta q\,\Delta t_{Well\,12} - 0.63496\,\Delta q/\Delta t_{Well\,12}$$
$$+ 0.0097\,\Delta q_{Well\,38I} - 0.0032\,\Delta q\,log\Delta t_{Well\,38I}$$
$$- 5.2468 x 10^{-07}\,\Delta q\,\Delta t_{Well\,38I} - 0.1231\Delta q/\Delta t_{Well\,38I}\,.$$

In **Figure 52**, we present pressure predictions for the Well 10/Well 11 configuration. R^2_{train} is equal to 0.8437 and R^2_{test} is equal to -2.7581.

The Well P1/Well P3 configuration has low R^2_{test} and slightly high R^2_{train} values compare to base case. We are not observing any improvements pressure prediction trends by adding Well 11's flowrate history. Although it is reported from the field observations that there is communication between wells, we do not observe that Well 11's flowrate history affects Well 10's bottom-hole pressure responses using the lasso algorithm.
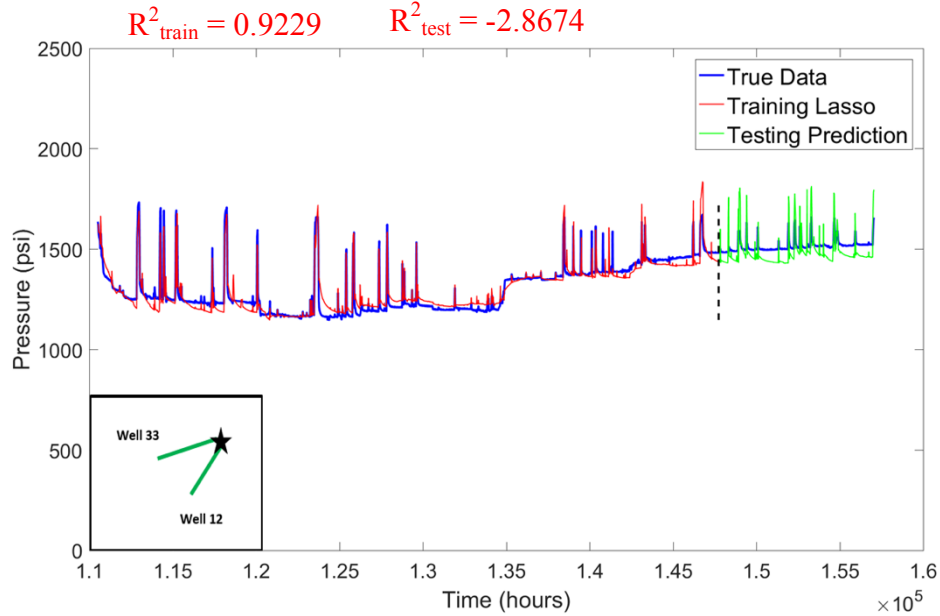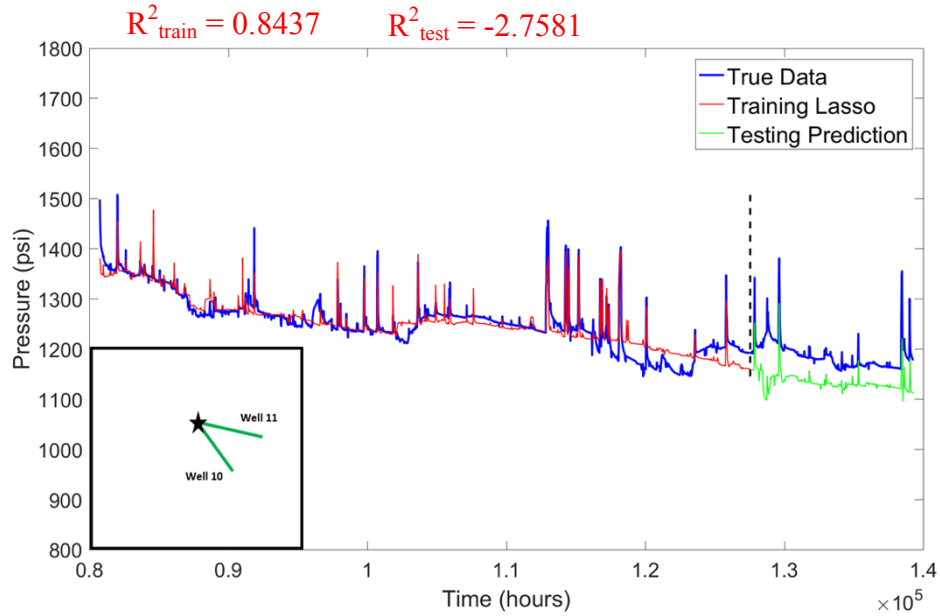

Figure 52 — Pressure prediction for the Well 10/Well 11 configuration using the lasso algorithm. R^2_{train} is equal to 0.8437 and R^2_{test} is equal to -2.7581. There is not improvements in match by adding Well P3's flowrate history.

Eq. 46 presents linear model for the Well 10/Well 11 configuration. Coefficients are calculated using the lasso algorithm.

$$P_{bh\,WellP3} = 2776.5770 + 0.0542\,\Delta q_{Well12} - 0.04901\,\Delta q\,log\Delta t_{Well\,12} \quad (\;46\;)$$
$$+\;1.9479x10^{-06}\,\Delta q\,\Delta t_{Well\,12} - 0.7365\,\Delta q/\Delta t_{Well\,12}$$
$$-\;0.0023\,\Delta q_{Well\,24} + 0.0019\Delta q\,log\Delta t_{Well\,24}$$
$$-\;9.351x10^{-07}\,\Delta q\,\Delta t_{Well\,24} - 0.0169\,\Delta q/\Delta t_{Well\,24}\,.$$

Figure 53 presents pressure predictions for the Well 10/Well 23 configuration. R^2_{train} is equal to 0.8419 and R^2_{test} is equal to -14.3543.

This configuration has a low R^2_{test} value and decrease in prediction accuracy with adding Well 23 flowrate history. Therefore, we do not have enough evidence to assume that there are interference effects between these wells.

Figure 53 — Pressure prediction for the Well 10/Well 23 configuration using the lasso algorithm. R^2_{train} is equal to 0.8419 and R^2_{test} is equal to -14.3543. There are significant mismatch between forecasted pressure and real measurements during $1.31x10^5$ and $1.39x10^5$ hours by adding Well 23 flowrate history. Overall, there are not improvements in visual fit of pressure predictions by adding Well 23's flowrate history.

Eq. 47 presents linear model for the Well 10/Well 23 configuration using the lasso algorithm.

$$P_{bh\,Well\,12} = 1833.788 - 0.0024\,\Delta q_{Well\,12} - 0.0093\,\Delta q\,log\Delta t_{Well\,12} \quad (\ 47\ )$$
$$+ 1.6233x10^{-06}\,\Delta q\,\Delta t_{Well\,12} + 0.0554\,\Delta q/\Delta t_{Well\,12}$$
$$- 0.0371\,\Delta q_{Well\,33} + 0.0180\,\Delta q\,log\Delta t_{Well\,33}$$
$$- 2.523\,x10^{-06}\Delta q\,\Delta t_{Well\,33} + 0.4033\,\Delta q/\Delta t_{Well\,33}.$$

Although the algorithm detect that there are interference effects between Well 10 and Well 22, it does not identify that the flowrate of Well 11 has impact on Well 10's bottom-hole pressure. From the dynamic observation of field, we know that both Wells 22 and 11 have communication with Well 10. Hence, in this part, we examine the Well 10/Well 11/Well 22 configuration.

In **Figure 54.a,** we present an enlarged view of Well 10 pressure prediction using only the flowrates for Well 10 and in **Figure 54**.b, we have reproduced and forecasted pressure responses for the Well 10/Well 11/Well 22 configuration. Visually and statistically, the Well 10/Well 11/Well 22 configuration performs noticeably better than the Well 10 configuration or any other presented scenarios.
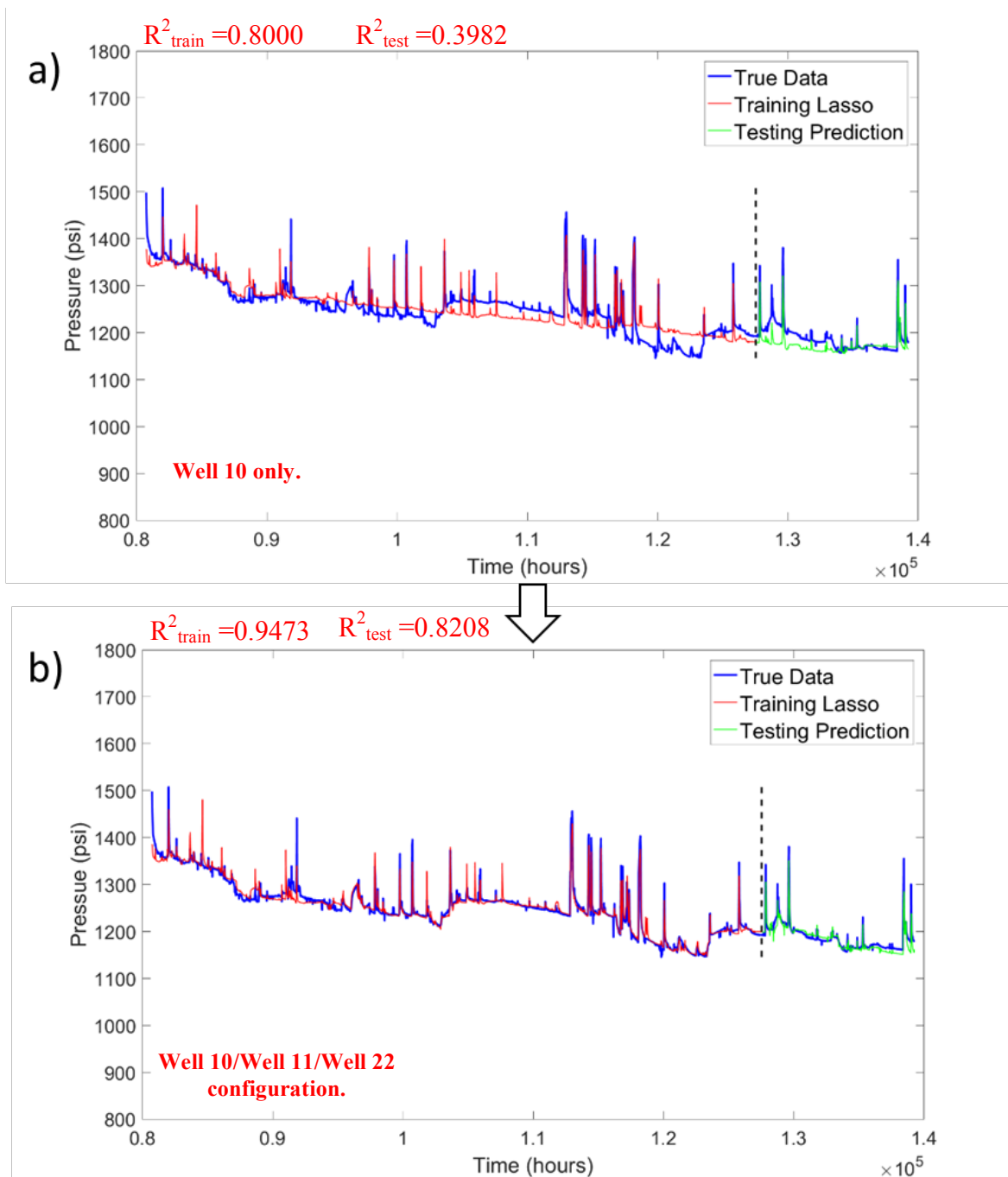
Figure 54 — Pressure prediction for Well 10 using the influence of Wells 11 and 22. Figure 54.a illustrates the pressure prediction obtained using only the flowrate history for Well P3. R^2_{train} is equal to 0.8000 and R^2_{test} is equal to 0.3982. Figure 54.b illustrates the improved pressure prediction obtained using the flowrate histories for Wells 10, 11, and 22. R^2_{train} is equal to 0.9473 and R^2_{test} is equal to 0.38208. Statistically and visually, this configuration gives the best match compare to other scenarios.

Case 5 — Yusif Well 110

In case 5, we demonstrate pressure predictions for Well 110 with including offset wells interference effects. We chose Wells 16I, 02, and 13 as possible interference wells for Well 110 and analyzed their impact to bottom-hole pressure measurements. Well 110 and Well 13 were completed in L5 and L3 layers while Well 16I was completed in L5, L4, L3, Lob1 and Loa2 layers; Well 02 was completed in L3 layer. From the dynamic field observation, we know that there is strong pressure communication between Well 110 and Well 16I.

From the dynamic observations of field, it was reported that Well 110 has strong interference effects with nearby injector well which is Well 16I.

In **Figure 55** we present bottom-hole pressure history for Well 110 and flowrates for Well 110, 16I, 02, and 13.
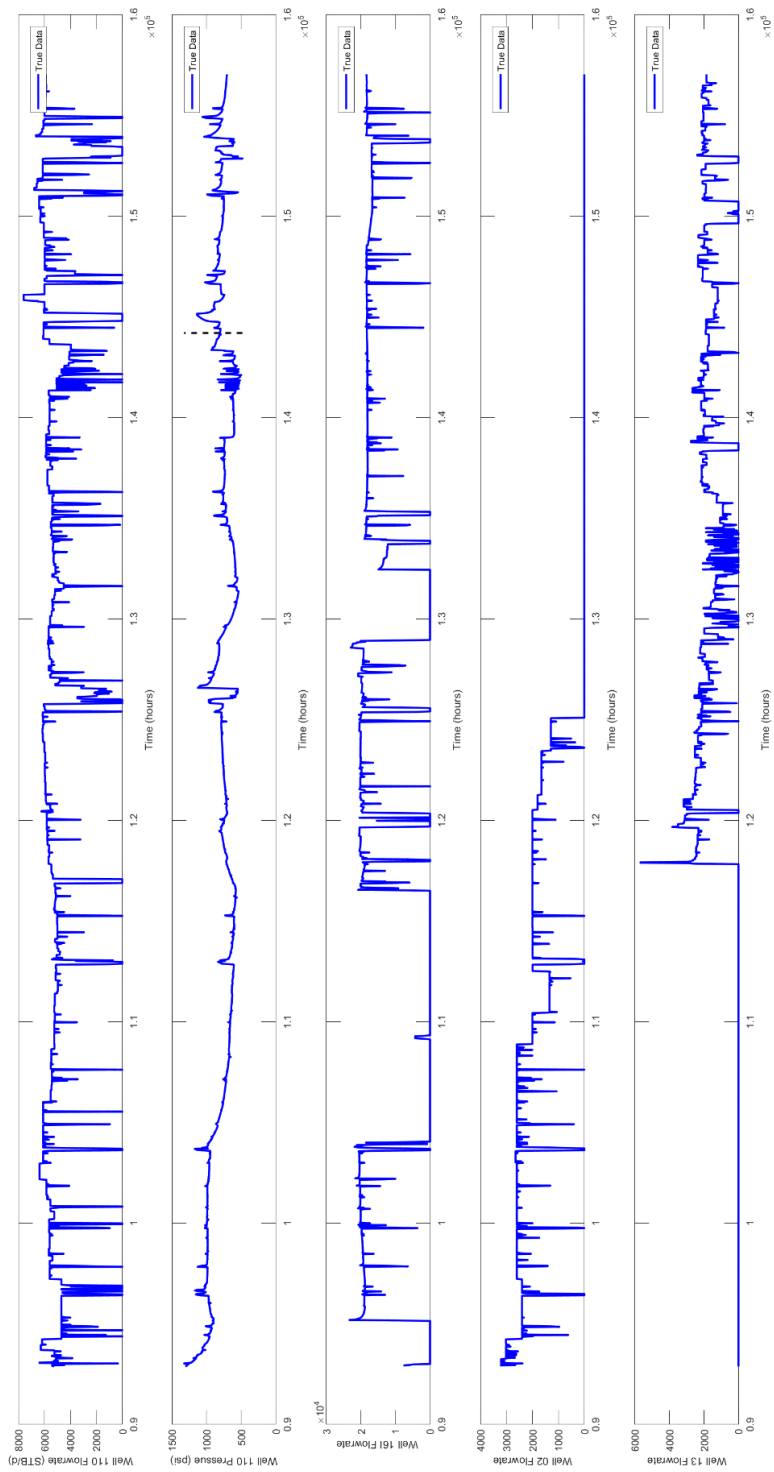
Figure 55 — Flowrate and bottom-hole pressure history for Well 110 and flowrate histories for Wells 16I, 02, and 13.

Figure 56 presents pressure predictions for the Well 110/ Well 16I configuration. R^2_{train} is equal to 0.7655 and R^2_{test} is equal to -3.9949.

The R^2 value and visual fit improve dramatically by adding Well 16I's flowrate history as a feature. The Well 110/ Well 16I configuration captures all pressure increase and decrease times for Well 110's bottom-hole pressure response. Therefore, we can conclude that the lasso algorithm requires both Well110 and Well 16I's flowrate histories to accurately reproduce the bottom-hole pressure response of Well 110.
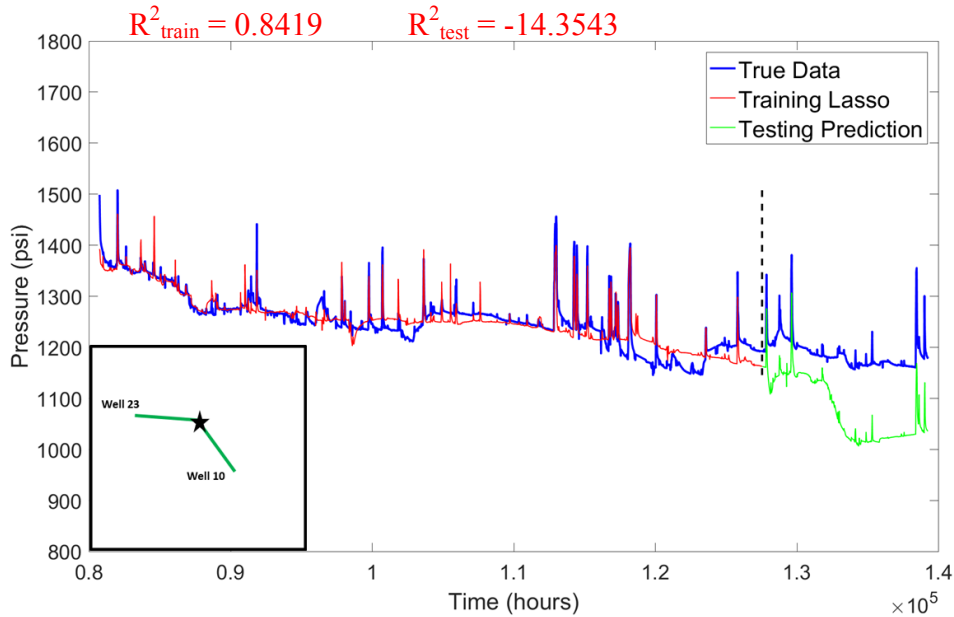

Figure 56 — Pressure predictions for the Well 110/ Well 16I configuration using the lasso algorithm. R^2_{train} is equal to 0.7655 and R^2_{test} is equal to -3.9949. By adding flowrate of Well 16I, the algorithm captures pressure trend variations, pressure decrease and increate times accurately.

Eq. 48 presents linear model for the Well 110/ Well 16I configuration using the lasso algorithm.

$$P_{bh\ WellP3} = 1950.9957 + 0.0575\ \Delta q_{Well\ 12} \quad\quad\quad\quad\quad (\ 48\ )$$

$$- 0.0293\ \Delta q\ log\Delta t_{Well\ 12} - 1.434x10^{-06}\ \Delta q\ \Delta t_{Well\ 12}$$

$$- 0.7196\ \Delta q/\Delta t_{Well\ 12} - 0.02746\ \Delta q_{Well\ 38I}$$

$$+ 0.01137\ \Delta q\ log\Delta t_{Well\ 38I} - 1.2203x10^{-07}\ \Delta q\ \Delta t_{Well\ 38I}$$

$$+ 0.2855\ \Delta q/\Delta t_{Well\ 38I}\ .$$

Figure 57 presents pressure predictions for the Well 110/ Well 13 configuration. R^2_{train} is equal to 0.9053 and R^2_{test} is equal to -3.3079.

The R^2_{train} and R^2_{test} values increase with adding Well 13 flowrate history. But, there are not obvious improvements in pressure trends by adding Well 13's flowrate history as we observe by adding Well 16I. Therefore, we do not have evidence to conclude that Well 13's flowrate history affect Well 110's bottom-hole pressure response.
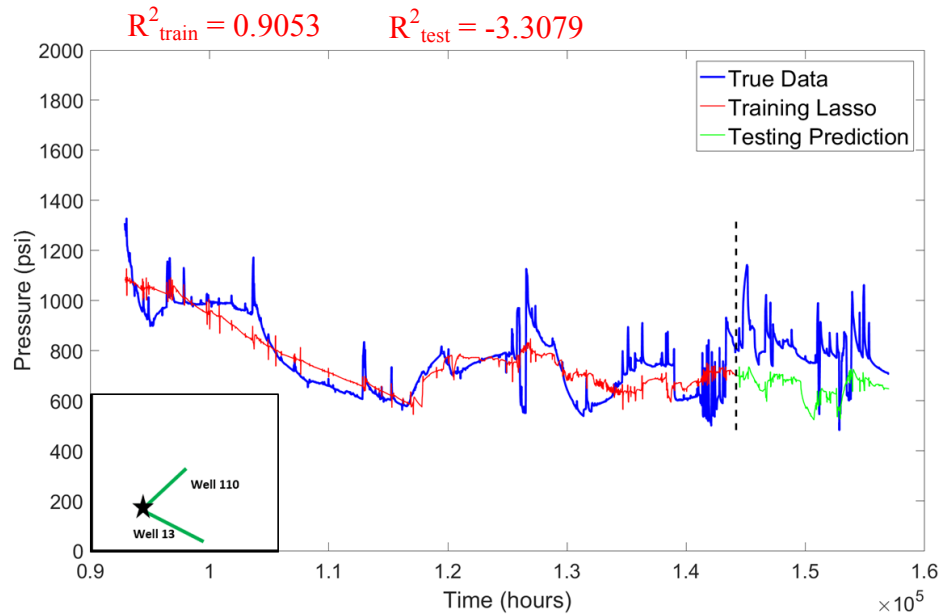
Figure 57 — Pressure prediction for the Well 110/ Well 13 configuration using the lasso algorithm. R^2_{train} is equal to 0.9053 and R^2_{test} is equal to -3.3079. Pressure trend changes are not captured accurately by adding Well 13's flowrate as they are captured by adding Well 16I's flowrate history.

Eq. 49 presents linear model for the Well 110/ Well 13 configuration. Coefficients are calculated using the lasso algorithm.

$$
\begin{aligned}
P_{bh\,WellP3} = {} & 2776.5770 + 0.0542\,\Delta q_{Well12} - 0.04901\,\Delta q\,log\Delta t_{Well\,12} \quad (\ 49\ ) \\
& + 1.9479x10^{-06}\,\Delta q\,\Delta t_{Well\,12} - 0.7365\,\Delta q/\Delta t_{Well\,12} \\
& - 0.0023\,\Delta q_{Well\,24} + 0.0019\Delta q\,log\Delta t_{Well\,24} \\
& - 9.351x10^{-07}\,\Delta q\,\Delta t_{Well\,24} - 0.0169\,\Delta q/\Delta t_{Well\,24}\,.
\end{aligned}
$$

Figure 58 presents pressure predictions for the Well 110/Well 23 configuration. R^2_{train} is equal to 0.9229 and R^2_{test} is equal to -2.8674.

Although R^2_{train} and R^2_{test} values increase by adding Well 23 flowrate history, visually there are not improvements in pressure trends. The algorithm is not able to capture bottom-hole pressure variations. Therefore, we assume that most probably there are not interference effects between these wells.
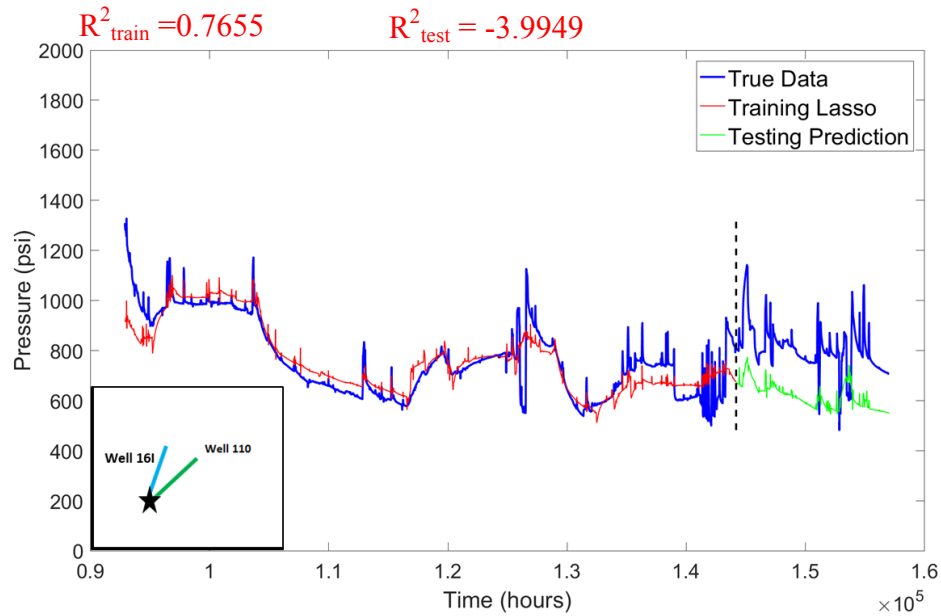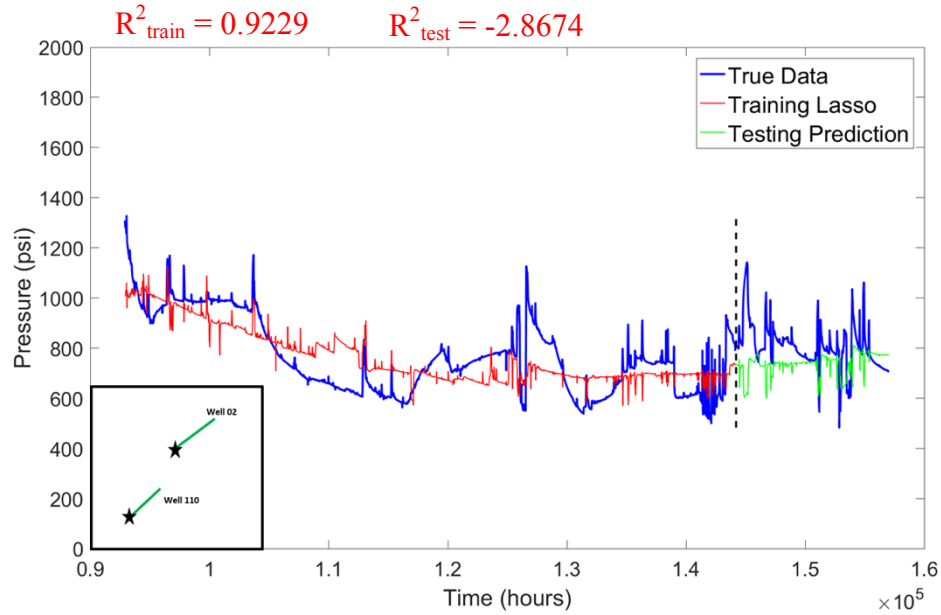

Figure 58 — Pressure prediction for the Well 110/Well 23 configuration using the lasso algorithm. R^2_{train} is equal to 0.9229 and R^2_{test} is equal to -2.8674. There are not visual improvements in pressure predictions by adding Well 23's flowrate history.

Eq. 50 presents linear model for the Well 110/Well 23 configuration using the lasso algorithm.

$$P_{bh\,Well\,P3} = 2376.6503 - 0.0009\,\Delta q_{Well\,12} - 0.0271\,\Delta q\,log\Delta t_{Well\,12} \quad (\ 50\ )$$
$$- 1.732x10^{-07}\,\Delta q\,\Delta t_{Well\,12} - 0.1224\Delta q/\Delta t_{Well\,12}$$
$$+ 0.0395\,\Delta q_{Well\,33} - 0.0125\,\Delta q\,log\Delta t_{Well\,33}$$
$$+ 1.3274x10^{-06}\,\Delta q\,\Delta t_{Well\,33} - 0.5018\Delta q/\Delta t_{Well\,33}\,.$$

99

5.2.3. Validation of Results

Our analysis shows that by including the offset wells interference effects, we increase the accuracy of the bottom-hole pressure predictions. Those improvements in accuracy can be examined visually and with R^2 values. However, the statistical parameters are not reliable to identify interference effects.

From the applications of the methodology, we prepared tables for the synthetic (the Guler field) and the field (the Yusif field) cases to validate the results.

Table 11 presents identified interference effects for the Guler field using the methodology — the dark green color indicates strong interference effects and the light green color indicates medium interference effects captured by the algorithm. Known well interference effects from the synthetic model are indicated with as "yes." In **Table 11**, we observe good agreement between identified interference and known interference effects.

Table 11   —   Interference table for the Guler field. The dark green color indicates strong interference effects and the light green color indicates medium to interference effects identified using the lasso algorithm. "Yes" designates proven connectivity between wells.

	P1	P2	P3
P1			
P2			yes
P3		yes	
I1		yes	yes
I2	yes		

Table 12 presents pressure communicating offset wells identified by the lasso algorithm and dynamically observed ones for the Yusif field. The dark green color indicates a strong interference effects identified by lasso algorithm. The white color indicates the absence of interference effects identified by lasso algorithm. The gray color indicate well configurations

which we do not test. The pressure interference effects dynamically observed in the field are designated as "Seen interference" and dynamically observed non-communicating wells are designated as "Not seen." In Table 12, we observe good agreement between identified interference and known interference effects

Table 12 — Interference table for Yusif field. The dark green color indicates good connectivity between wells. The gray color indicate well configuration which we did not test. "Seen interference" indicates observed interference effects in the field and "Not seen" indicates unobserved interference effects.

	Well 12	Well 10	Well 110
Well 38I			
Well 24	Not seen		
Well 33	Not seen		
Well 22		Seen interference	
Well 11		Seen interference	
Well 23			
Well 16I			Seen interference
Well 02			
Well 13			

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FUTURE WORK

6.1 Summary

In this thesis, our primary goals are to reproduce and forecast pressure behavior of a well and to create a methodology to identify the interference influence of offset wells using linear-based machine-learning techniques. Both synthetic and field cases data are used to evaluate the proposed methodology. An E&P company provided the data for the field case on the provision that the data remain anonymous (as a requirement of confidentiality).

Specifically, the work in this thesis is conducted using commercial software such as Eclipse 100 (Schaumberg 2008), Kappa Workstation (Kappa 2018), and MATLAB (MATLAB 2018). As a preliminary scenario, we evaluate a single well case using only the performance data for that well (*i.e.,* time, flowrate, and bottom-hole pressure data). In subsequent applications, we use multi-well cases where we systematically test the interference influence of offset wells by adding one well at a time until all well are considered.

The purpose of this work is to establish that the interference influence of offset wells can be qualified and quantified using statistical methods (*i.e.*, machine learning), but more importantly, to demonstrate that offset well interference influence can be detected without using a numerical or analytical reservoir model.

6.2 Conclusions

- It is essential to linearize relationships between flowrate and pressure data by introducing physically-based features prior to data analysis using linear machine-learning algorithms.
- Linear machine learning algorithms are capable of identifying/capturing the reservoir and well behavior for a selected well without any explicit or prior physical assumptions.
- Once trained (*i.e.*, calibrated), linear machine learning algorithms can provide forecasts of bottom-hole pressure or a given well (or set of wells) using only time and flowrate data.
- Influence of interference from offset wells can be detected using the lasso algorithm (dominant interference effects are captured more easily).

● As a caveat, artifacts in flowrate histories can be incorrectly interpreted as an offset well interference effects.

6.3 Limitations

● Lasso algorithm needs a relatively long and accurate bottom-hole pressure and flowrate training data to learn pressure behavior.
● Human intervention (opening the new perforation layers), changes in a reservoir and well performance (change from infinite-acting flow to boundary dominated flow, in well skin, etc.) affect prediction accuracy.
● Similar flowrate histories among different wells could lead to wrong interpretations.

6.4 Recommendations for Future Work:

● Introducing pseudo-pressure concept as feature to be able to include gas flow.
● Introducing a pressure propagation time lag from offset wells should improve interference effects detection for more distant wells. It should also help to prevent spurious correlations due to platform maintenance operations where several wells are shut down at the same time.
● Dividing trained flowrate and bottom-hole pressure history into two parts — early times (infinite-acting flow) and late times (boundary dominated flow) — should increase pressure prediction accuracy.

NOMENCLATURE

p_i = Initial reservoir pressure, psia

p_{wf} = Flowing bottom-hole pressure, psia

p = Pressure, psia

t = Time, hours

s = Skin factor, dimensionless

r_w = Wellbore radius, ft

B = Formation volume factor, reservoir vol./surface vol.

μ = Viscosity, cp

k = Permeability, md

h = Formation thickness, ft

ϕ = Porosity

c_t = Total compressibility, psia-1

REFERENCES

Athichanagorn, S., Horne, R. N., and Kikani, J. (2002). Processing and Interpretation of Long-Term Data Acquired From Permanent Pressure Gauges. Society of Petroleum Engineers. doi:10.2118/80287-PA.

Baker, A., Gaskell, J., Jeffery, J., Thomas, A., Veneruso, T. and Unneland, T., 1995, Permanent monitoring - looking at lifetime reservoir dynamics, Oilfield Rev, 7(4): 32-46.

Collins, M. and Duffy, N. (2002). Convolution kernels for natural language. Advances in Neural Information Processing Systems, 14(1):625–632.

Dake, L. P., 1978. *Fundamentals of Reservoir Engineering*, first edition. Amsterdan, The Netherlands: Elsevier.

De Oliveira Silva, M. I., and Kato, E. T. (2004). Reservoir Management Optimization Using Permanent Downhole Gauge Data. Society of Petroleum Engineers. doi:10.2118/90973-MS.

Earlougher, R. C. Jr., 1977. *Advances in Well Test Analysis*, first edition. Dallas, USA: Society of Petroleum Engineers of AIME.

Ecrine Computer Software. Kappa. 2013. http://www.kappaeng.com

Frota, H. M., and Destro, W. (2006). Reliability Evolution of Permanent Downhole Gauges for Campos Basin Subsea Wells: A 10-Year Case Study. Society of Petroleum Engineers. doi:10.2118/102700-MS.

Horne, R. N. (2007). Listening to the Reservoir—Interpreting Data from Permanent Downhole Gauges. Society of Petroleum Engineers. doi:10.2118/103513-JPT http://www.jstor.org/stable/2346178

James, G., Witten, D., Hastie T., and Tibshirani R., 2013. *An introduction to Statistical Learning*, Springer.

Kamal, M. M. (1983). Interference and Pulse Testing — A Review. Society of Petroleum Engineers. doi:10.2118/10042-PA.

KAPPA-Workstation 2018, Kappa Engineering, Sophia Antipolis, France.

Kuchuk, F. J., Onur, M. and Hollaender, F., 2010. *Pressure Transient Formation and Well Testing*, first edition. Oxford, UK: Elsevier.

Laskov, P. and Nelson, B. (2012). Theory of kernel functions. University Tubingen, Germany. Lecture Notes for Advanced Topics in Machine Learning.

Lee, J. W., 1982. *Well Testing*, first edition. Dallas, USA: Society of Petroleum Engineers of AIME.

Liu, Y., 2013. Interpreting Pressure and Flowrate Data from Permanent Downhole Gauges Using Data Mining Approaches. PhD dissertation, Stanford University.

MATLAB 2018, the MathWorks, Inc., Natick, Massachusetts, United States.

Mitchell, M. T., 1997. *Machine Learning*, first edition. Maidenhead, U.K.: McGraw-Hill Science/Engineering/Math.

Nestlerode, W. A. (1963). The Use of Pressure Data from Permanently Installed Bottom Hole Pressure Gauges. Society of Petroleum Engineers. doi:10.2118/590-MS.

Oliver, D. S., Reynolds, A. C., and Liu, N., 2008. *Inverse theory for petroleum reservoir characterization and history matching,* Cambridge University Press.

Schlumberger. 2008. *Eclipse Reference Manual*, Houston, Texas

Tian, C., (2015). Machine Learning Applied to Multiwell Test Analysis and Flow Rate Reconstruction. Society of Petroleum Engineers. doi:10.2118/175059-MS.

Tian, C., 2014. Applying Machine Learning and Data Mining Techniques To Interpret Flowrate, Pressure And Temperature Data From Permanent Downhole Gauges. MS Thesis, Stanford University.

Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1), 267-288.

Veneruso, A., Economides, C., and Akmansoy, A. (1992). Computer based downhole data acquisition and transmission in well testing. Society of Petroleum Engineers. doi:10.2118/24728-MS.

Zheng, S., & Li, X. (2007). Analyzing Transient Pressure from Permanent Downhole Gauges (PDG) Using Wavelet Method. Society of Petroleum Engineers. doi:10.2118/107521-MS.

Zou, H., & Hastie, T. (2003). Regression shrinkage and selection via the elastic net, with applications to microarrays. Journal of the Royal Statistical Society: Series B. v67, 301-320.

APPENDIX A.

This appendix presents further application of the methodology to identify interference effects from offset wells using the lasso algorithm. One extra synthetic well example and one real well cases are presented.

Well P2

Well P2 section presents the scenarios generated by the lasso algorithm considering the flowrate history for Well P2, as well as all of the combinations of Well P2 and its possible offsets.

Figure A. 1 presents the flowrate history and pressure response versus time for Well P2. We note that the vertical dashed line at approximately $5x10^4$ hours denotes the boundary between the training and testing data — the data to the left of this line are used for training and data to the right of this line are used for testing.
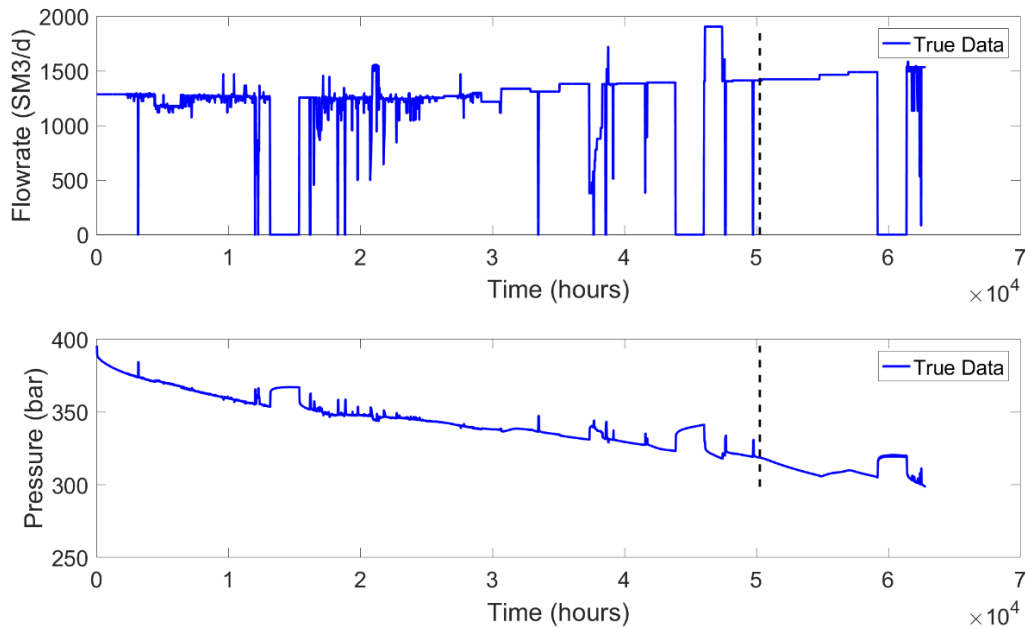

Figure A. 1 — Flowrate and (bottom) bottom-hole pressure history for Well P2. The data to the left side of the dashed lines are training data, and the data to the right side of the dashed line are the testing data.

Figure A. 2 presents the bottom-hole pressure prediction of Well P2 using the lasso method. The lasso algorithm generally captures pressure behavior from the flowrate history during both the training and the testing periods with noted discrepancies, indicated by the black circles.

From **Figure A. 1** it can be seen that these discrepancies occur when flowrates of selected well are relatively constant, but pressure behavior is changing — two pressure increase trends occur at approximately 3.1×10^4 and 5.5×10^4 hours when the flowrates of Well P3 slightly increase. This pressure increase with flowrate increase, might indicate interference effects from offset wells. As the Guler field is synthetic case, we know that Well P2 have connectivity with Wells P3 and I1.

Eq. A. 1 presents linear model for reproducing bottom-hole pressure history of Well P2 using Well P2's flowrate history only. The coefficients are calculated using the lasso algorithm. From the coefficients, it can be seen that $\Delta q \, \Delta t$ term has less statistical significance for predicting the bottom-hole pressure of Well P2.
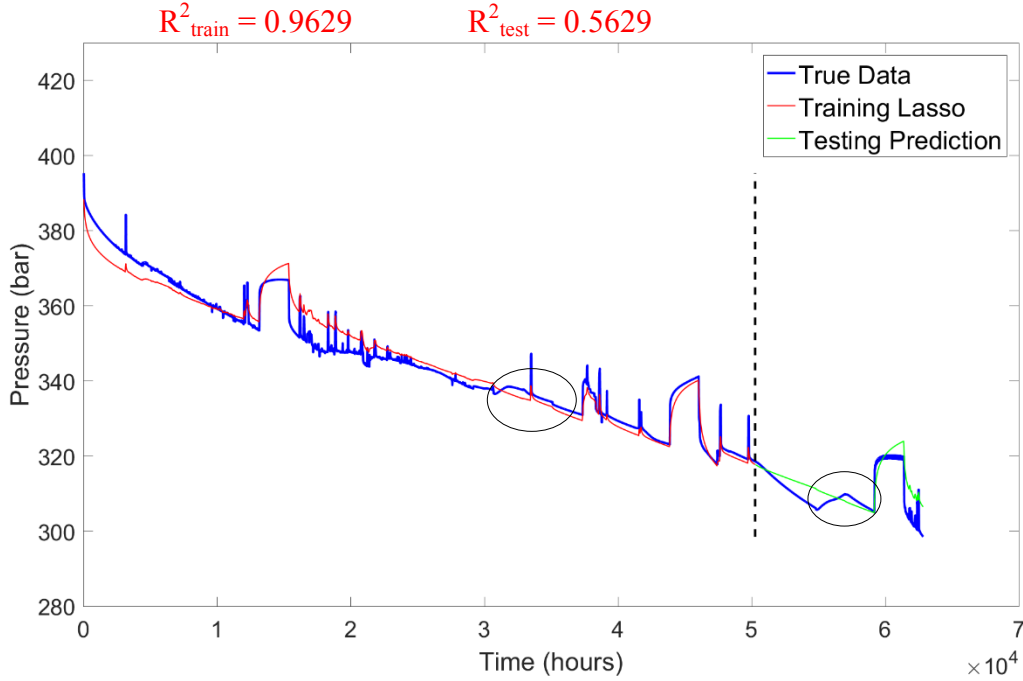
Figure A. 2   — Pressure prediction with using linear machine learning techniques for Well P2. The blue trend represents the true bottom-hole pressure history; the red trend indicates training pressure predictions; the green trend indicates testing pressure predictions. Only rates from Well P2 were used for training and prediction. Black circles indicate pressure trends that were not captured by the lasso algorithm.

$$P_{bh\,Well\,P2} = 389.9052 + 0.0071\,\Delta q_{Well\,P2} - 0.006\,\Delta q\,log\Delta t_{Well\,P2} \qquad (\text{ A. 1 })$$
$$- 7.645x10^{-07}\,\Delta q\,\Delta t_{Well\,P2} - 0.0004\,\Delta q/\Delta t_{Well\,P2}\,.$$

Figure A. 3 presents pressure predictions for the Well P2/Well I1 configuration. The Well P2/Well I1 configuration has high R^2_{train} and R^2_{test} values and there are visual improvements in pressure prediction trends which are captured by adding Well I1. Improvement in visual fit can be particularly well observed between $1.5x10^4$ and $2.5x10^4$ hours. Therefore, there might be interference effects between these wells.

Eq. A. 2 presents linear model for the Well P3/Well I2 configuration. Coefficients are calculated using the lasso algorithm. From the coefficients, it can be seen that $\Delta q\,\Delta t$ terms for Well P2 and I1 have less statistical significance for predicting the bottom-hole pressure of Well P2.
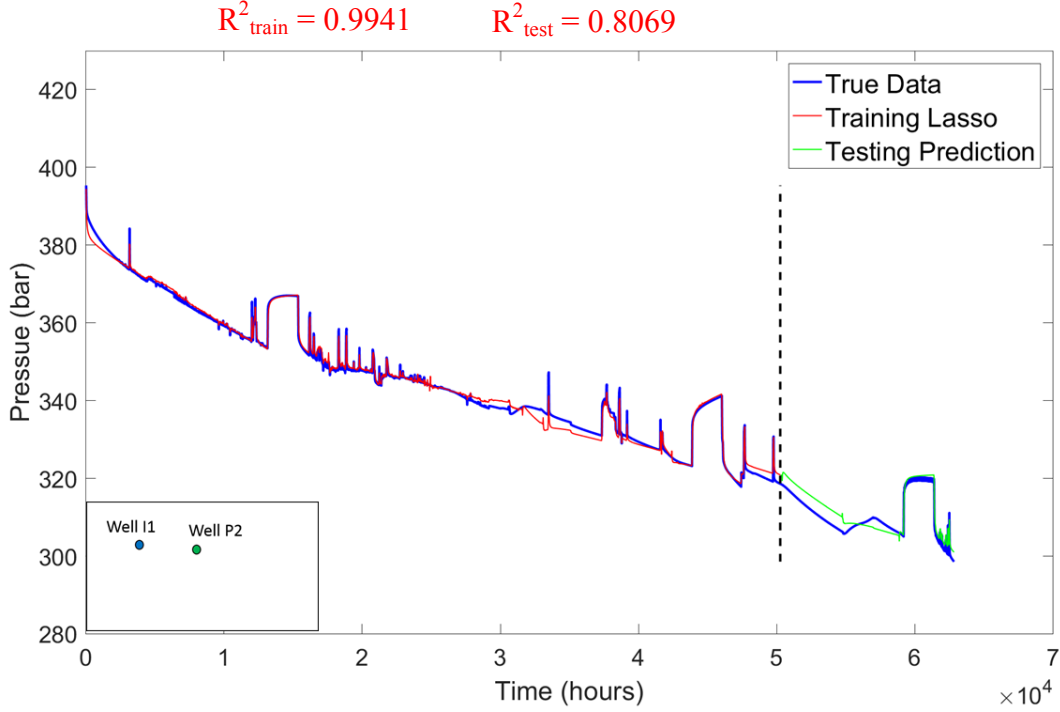
$R^2_{train} = 0.9941$　　　$R^2_{test} = 0.8069$

Figure A. 3　— Pressure prediction for the Wells P2 and I1 configuration using the lasso algorithm. Compare to the base case, there are visual improvements in pressure trend predictions between 1.5x104 and 2.5x104 hours by adding Well I2.

$$P_{bh\,Well\,P2} = 395.2071 - 0.0087\,\Delta q_{Well\,P2} - 0.0007\,\Delta q\,log\Delta t_{Well\,P2} \quad (\mathbf{51})$$
$$- 1.675x10^{-06}\,\Delta q\,\Delta t_{Well\,P2} + 0.2214\,\Delta q\,log\Delta t_{Well\,P2}$$
$$- 0.0033\,\Delta q_{Well\,I1} + 0.0011\,\Delta q\,log\Delta t_{Well\,I1}$$
$$+ 4.999x10^{-07}\,\Delta q\,\Delta t_{Well\,I1} + 0.0581\,\Delta q/\Delta t_{Well\,I1}.$$

Figure A. 4 presents pressure predictions for the Well P2/Well P3 configuration. Compare with **Error! Reference source not found.**, the lasso algorithm captures the pressure increase etween $3x10^4$ and $3.5x10^4$, the start of pressure increase at $5.5x10^4$ hours, and start of pressure decrease at $5.8x10^4$ hours by adding Well I1. Although R^2 testing value decrease compare to base case, there are visual improvements in pressure prediction trends. Therefore, we assume that there might be interference effects between these wells.

111

Eq. A. 3 presents linear model for this configuration. Coefficients are calculated using the lasso algorithm. From the coefficients, it can be seen that $\Delta q \, \Delta t$ terms for Well P2 and P3 have less statistical significance for predicting the bottom-hole pressure of Well P2.
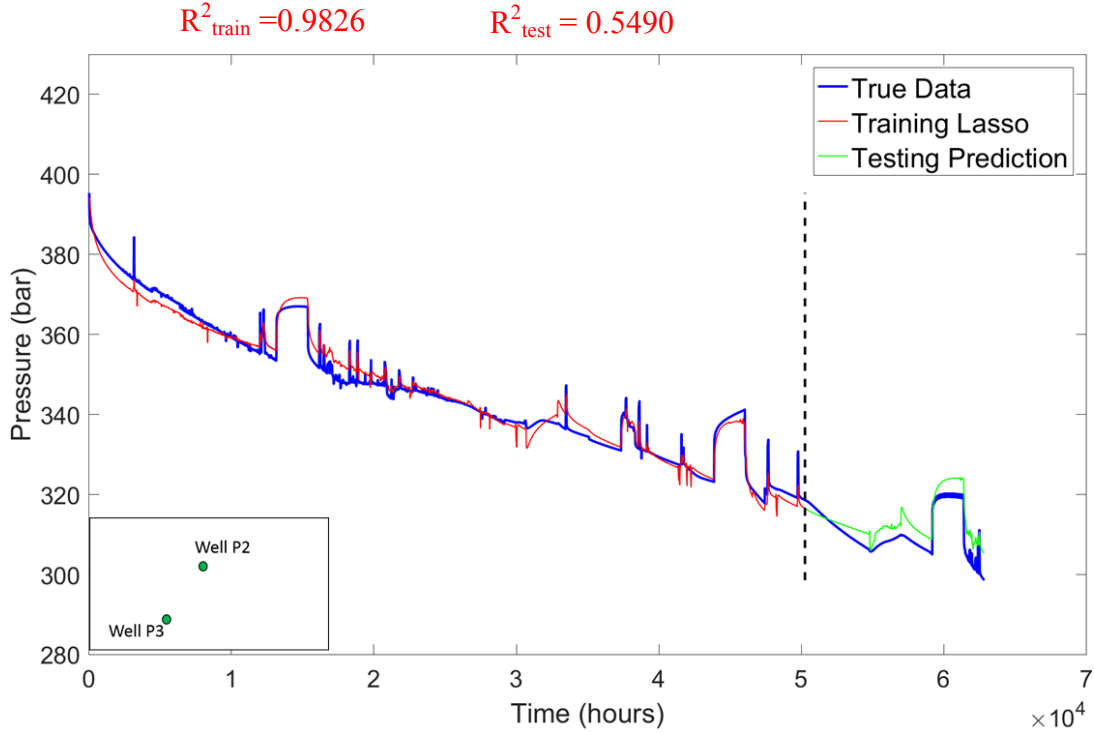
$R^2_{train} = 0.9826 \qquad R^2_{test} = 0.5490$

Figure A. 4 — Pressure predictions for the Wells P2 and P3 configuration using the lasso algorithm. Compare to the base case, the lasso algorithm captures time of the pressure increase between 3x104 and 3.5x104, the start of pressure increase at 5.5x104 hours, and start of pressure decrease at 5.8x104 hours.

$$
\begin{aligned}
P_{bh\,Well\,P2} = {}& 396.0083 + 0.0048\,\Delta q_{Well\,P2} - 0.0055\,\Delta q\,log\Delta t_{Well\,P2} \qquad (\;52\;) \\
& + 4.99x10^{-07}\,\Delta q\,\Delta t_{Well\,P2} + 0.0429\,\Delta q/\Delta t_{Well\,P2} \\
& + \; 0.0147\,\Delta q_{Well\,P3} - 0.0052\,\Delta q\,log\Delta t_{Well\,P3} \\
& - 9.927810^{-07}\Delta q\,\Delta t_{Well\,P3} - 0.1914\,\Delta q/\Delta t_{Well\,P3}\,.
\end{aligned}
$$

Figure A. 5 presents pressure predictions for the Well P2/Well P1 configuration. The R^2_{test} value decreases by adding Well P1 flowrate history. Moreover, there are pressure prediction increase trends approximately at $0.5x10^4$ and $5x10^4$ hours which we do not observe in real

pressure data. Therefore, we can-not assume that there is interference effects between these wells.

Eq. A. 4 presents linear model for the Well P3/Well P2 configuration using the lasso algorithm. $\Delta q \ \Delta t$ terms for Well P2 and P1 have less statistical significance for predicting the bottom-hole pressure of Well P2.
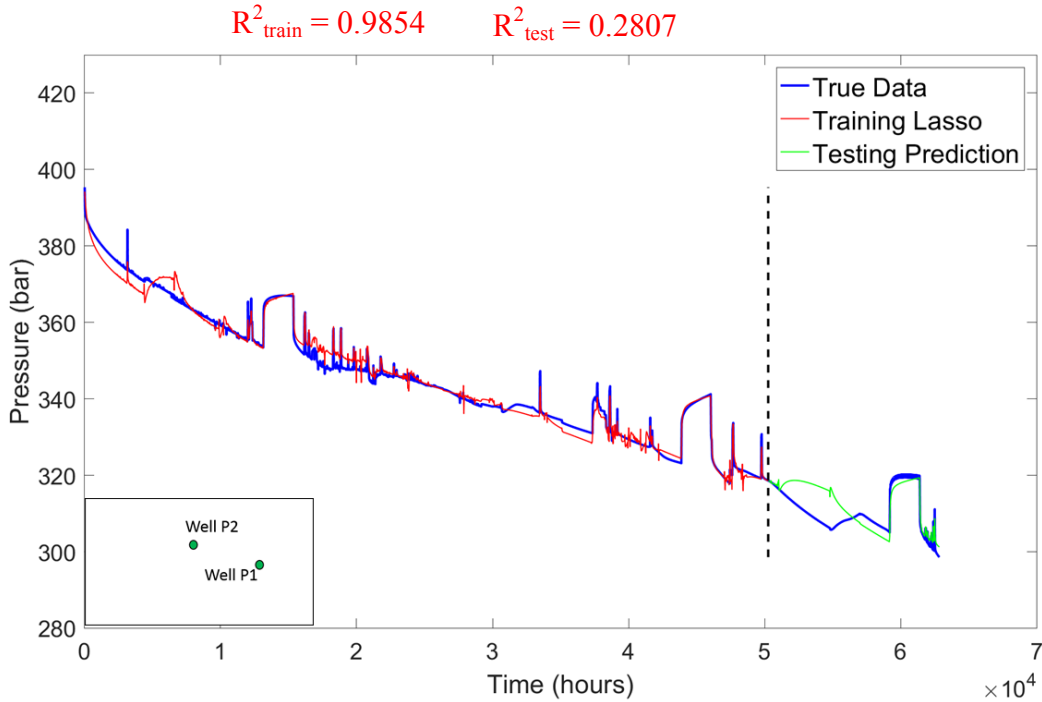

Figure A. 5 — Pressure prediction for the Well P2/Well P1 configuration using the lasso algorithm. There are pressure prediction increase trends approximately at 0.5x104 and 5x104 hours which we do not observe in real pressure data when we add Well P2 flowrate history.

$$P_{bh \ Well \ P2} = 397.9619 - 0.0083 \ \Delta q_{Well \ P2} - 0.0006 \ \Delta q \ log\Delta t_{Well \ P2} \qquad (\ A. 4 \ )$$
$$- \ 1.669x10^{-06} \ \Delta q \ \Delta t_{Well \ P2} + 0.2009\Delta q/\Delta t_{Well \ P2}$$
$$+ \ 0.0313\Delta q_{Well \ P1} - 0.01246 \ \Delta q \ log\Delta t_{Well \ P1}$$
$$+ \ 1.322x10^{-06} \ \Delta q \ \Delta t_{Well \ P1} - 0.4088 \ \Delta q\Delta t_{Well \ P1}$$

Figure A. 6 presents pressure predictions for the Well P2/Well I2 configuration. Although the Well P2/Well I2 configuration has a slightly high R^2_{test} value, there are not any visual

improvements in pressure prediction trends. Therefore, there is not obvious evidence of interference effects between these wells.

Eq. A. 5 presents linear model for the Well P3/Well P1 configuration. Coefficients are calculated using the lasso algorithm. $\Delta q \, \Delta t$ terms for both wells have less statistical significance for predicting the bottom-hole pressure of Well P2.
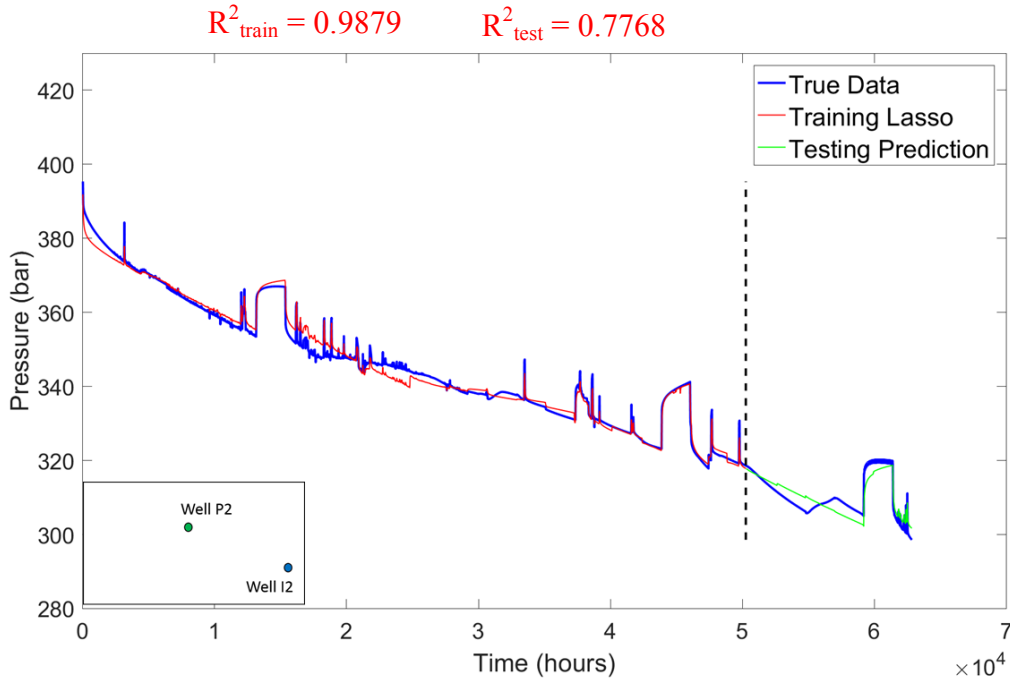
$$R^2_{train} = 0.9879 \qquad R^2_{test} = 0.7768$$

Figure A. 6 — Pressure prediction for the Wells P2 and I2 configuration using the lasso algorithm. Compare to the base case, there are not visual improvements in pressure trend predictions by adding Well I2.

$$
\begin{aligned}
P_{bh \, Well \, P2} = {} & 392.4923 - 0.0045 \, \Delta q_{Well \, P2} - 0.00184 \, \Delta q \, log\Delta t_{Well \, P2} \quad (\ 53 \ ) \\
& - 1.341x10^{-06} \, \Delta q \, \Delta t_{Well \, P2} + 0.162 \, \Delta q/\Delta t_{Well \, P2} \\
& + \ 0.00236 \, \Delta q_{Well \, I2} - 0.0005 \, \Delta q \, log\Delta t_{Well \, I2} \\
& + 4.493x10^{-07} \, \Delta q \, \Delta t_{Well \, I2} - 0.0151 \, \Delta q/\Delta t_{Well \, I2} \, .
\end{aligned}
$$

When we add Wells P3 and I1 as possible offset wells to predict the bottom-hole pressure, the visual fit improves considerable. So, we assume that Wells P3 and I1 might interfere with Well P2. Now, we generate the Well P2/Well I1/Well P3 configuration scenario to examine pressure predictions. **Figure A. 7** presents pressure predictions for this scenario. The Well P2/Well

I1/Well P3 configuration has a highest training and testing R^2 values relative to the other scenarios and an almost perfect visual fit. As can be seen, Wells P3, P2, and I1's flowrate history are required to accurately reproduce the bottom-hole pressure of Well P3.

Eq. A. 6 presents linear model for the Well P2/Well P3/ Well I1 configuration using the lasso algorithm. $\Delta q\ \Delta t$ terms for both wells have less statistical significance for predicting the bottom-hole pressure of Well P2.

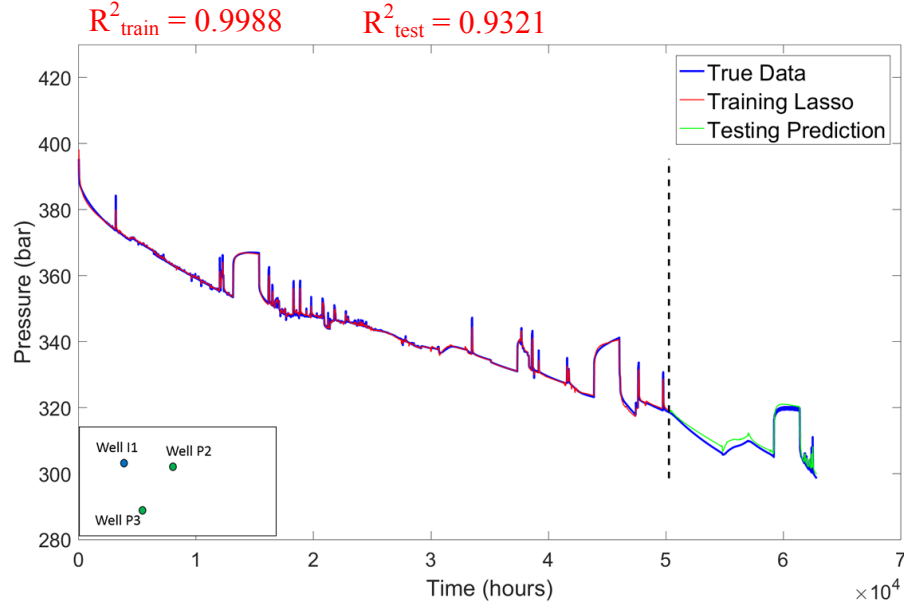$$R^2_{train} = 0.9988 \qquad R^2_{test} = 0.9321$$

Figure A. 7 — Pressure predictions for Wells P2, P3 and I1 configuration using the lasso algorithm. There is almost perfect match between real bottom-hole pressure values and pressure predictions for Wells P3, P2 and I1 configuration.

$$
\begin{aligned}
P_{bh\,Well\,P2} = {} & 399.239 - 0.0069\,\Delta q_{Well\,P2} - 0.0015\,\Delta q\,log\Delta t_{Well\,P2} \qquad (\ 54\ ) \\
& - 7.606x10^{-07}\,\Delta q\,\Delta t_{Well\,P2} + 0.2020\Delta q/\Delta t_{Well\,P2} \\
& - 0.0019\,\Delta q_{Well\,I1} + 0.0007\,\Delta q\,log\Delta t_{Well\,I1} \\
& + 4.222x10^{-07}\,\Delta q\,\Delta t_{Well\,I1} + 0.0250\Delta q/\Delta t_{Well\,I1} \\
& +\ 0.0066\,\Delta q_{Well\,P3} - 0.0027\,\Delta q\,log\Delta t_{Well\,P3} \\
& - 6.2037x10^{-07}\,\Delta q\,\Delta t_{Well\,P3} - 0.06943\,\Delta q/\Delta t_{Well\,P3}\,.
\end{aligned}
$$

Well 34

Well 34 section presents the scenarios generated by the lasso algorithm considering the flowrate history just for Well 34, as well as all of the combinations of Well 34 and its possible offsets. When we add Well 39 as possible offset wells to predict the bottom-hole pressure, the visual fit improves considerable.

Figure A. 8 presents the flowrate history and pressure response versus time for Well 34. We note that the vertical dashed line at approximately $11.3\text{x}10^4$ hours denotes the boundary between the training and testing data — the data to the left of this line are used for training and data to the right of this line are used for testing.
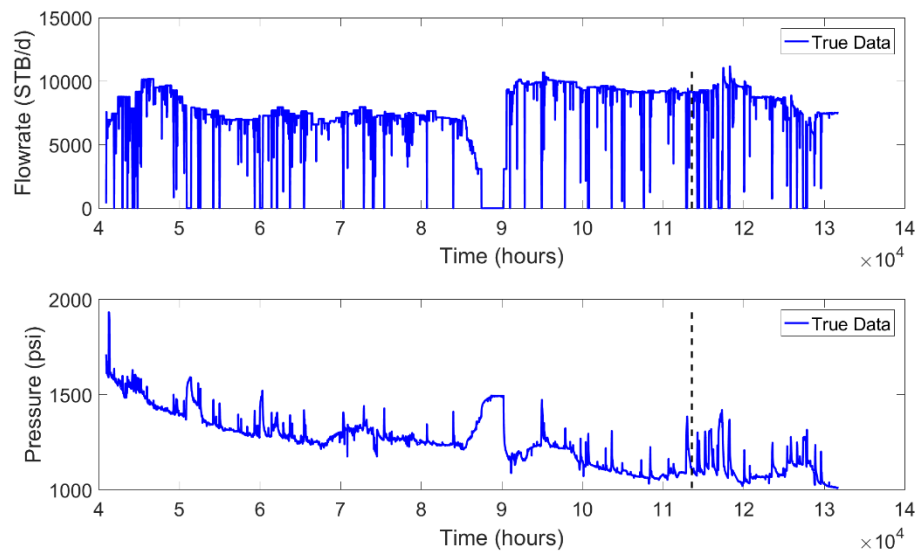

Figure A. 8 — Flowrate and (bottom) bottom-hole pressure history for Well 34. The data to the left side of the dashed lines are training data, and the data to the right side of the dashed line are the testing data.

Figure A. 9 presents the bottom-hole pressure prediction of Well 34 using the lasso method. The lasso algorithm generally captures pressure behavior from the flowrate history during both the training and the testing periods with noted discrepancies, indicated by the black circles.

From Figure A. 8 it can be seen that one of the discrepancies occur when flowrates and pressure of selected well slightly increase at $6.9\text{x}10^4$ hours. Another discrepancy occurs from time $9.1\text{x}10^4$ to $10\text{x}10^4$ hours where flowrate of Well 34 is fairly constant but pressure response is

116

increasing and then decreasing. This discrepancies might indicate interference effects from offset wells.

Eq. A. 7 presents linear model for reproducing bottom-hole pressure history of Well 34 using its flowrate history only.
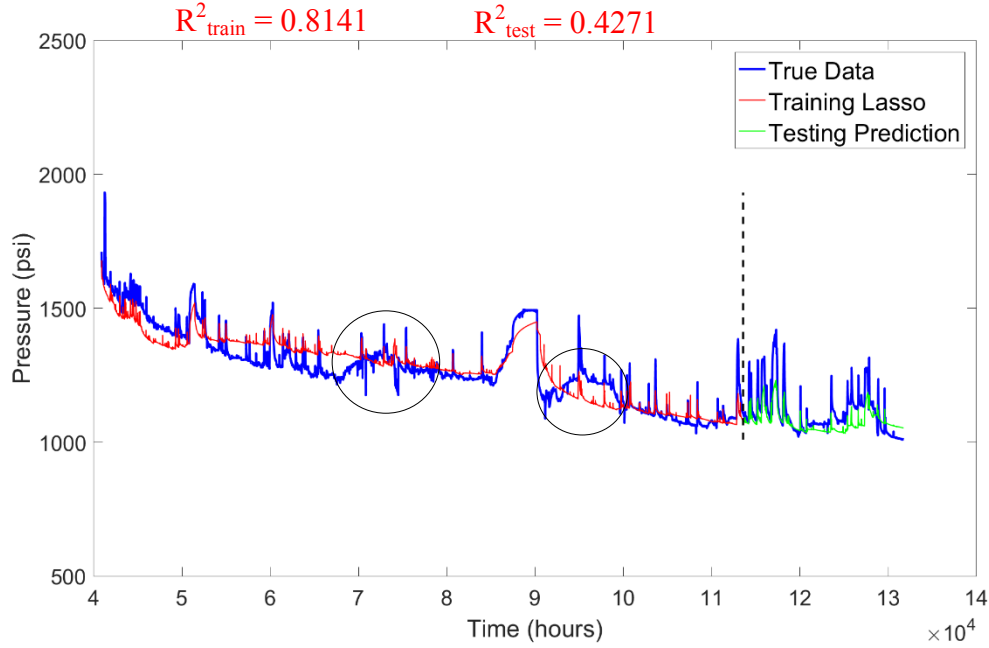

Figure A. 9　　— Pressure prediction with using linear machine learning techniques for Well 34. The blue trend represents the true bottom-hole pressure history; the red trend indicates training pressure predictions; the green trend indicates testing pressure predictions. Only rates from Well 34 are used for training and prediction.　Black circles indicate pressure trends that were not captured by linear machine learning algorithm.

$$P_{bh\ Well\ 34} = 1702.5296 + 0.0180\ \Delta q_{Well\ 34} - 0.0133\ \Delta q\ log\Delta t_{Well\ 34} \quad (\ 55\ )$$
$$- 4.219x10^{-07}\ \Delta q\ \Delta t_{Well\ 34} - 0.2056\ \Delta q/\Delta t_{Well\ 34}\ .$$
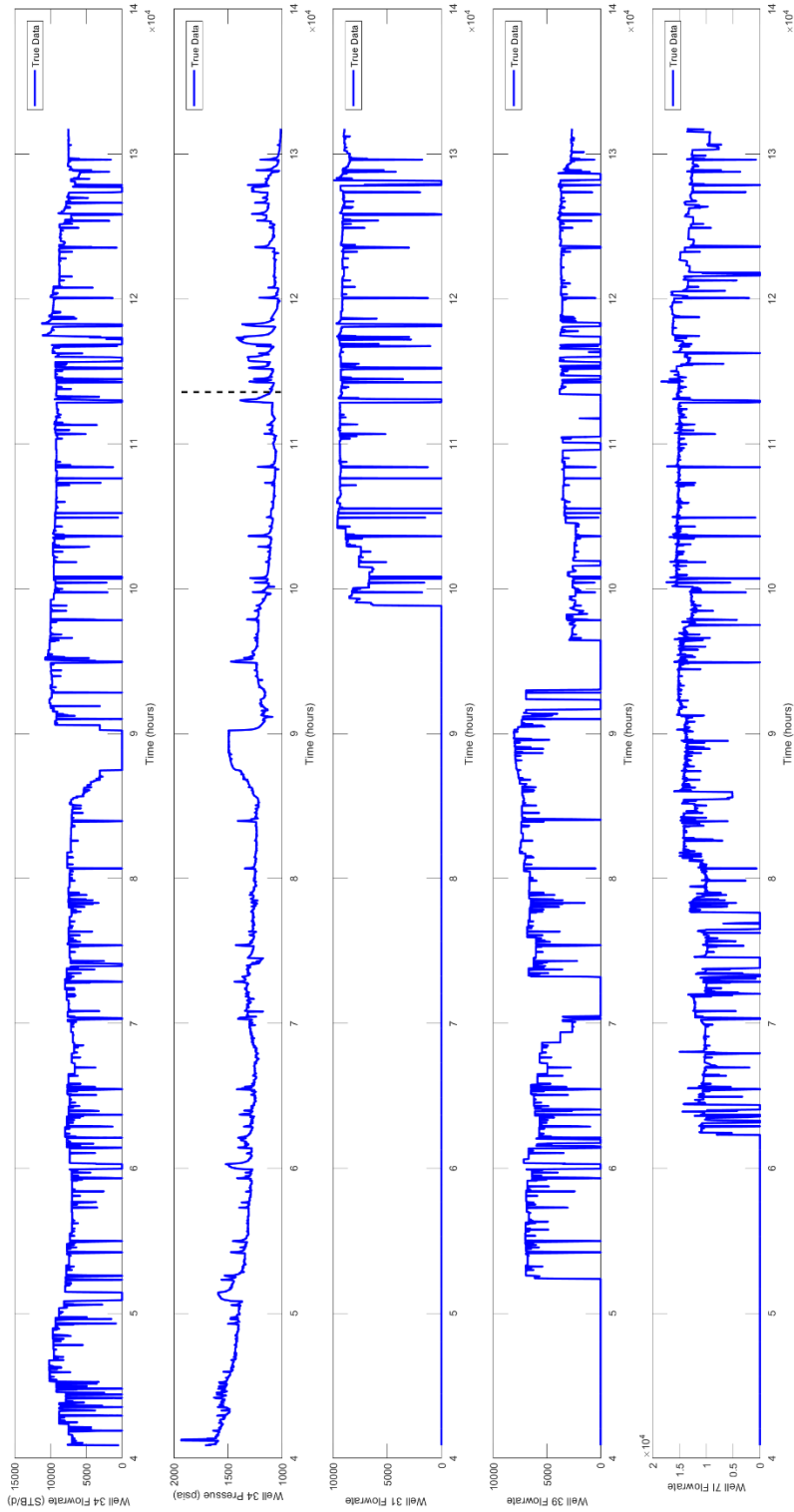
Figure A. 10 — Flowrate and bottom-hole pressure history for Well 34 and flowrate histories for Wells 31, 39, and 7I.

118

Figure A. 11 presents pressure predictions for the Well 34/Well 31 configuration. There are not increase in R^2 values by adding Well 31 flowrate history. Also, there is no visual improvements. Therefore, we do not have evidence that Well 31's flowrate history affect the bottom-hole pressure response of Well 34 by using the lasso algorithm.

Eq. A. 8 presents linear model for the Well P3/Well I2 configuration. Coefficients are calculated using the lasso algorithm. From the coefficients, it can be seen that $\Delta q \, \Delta t$ terms for Well P2 and I1 have less statistical significance for predicting the bottom-hole pressure of Well P2.
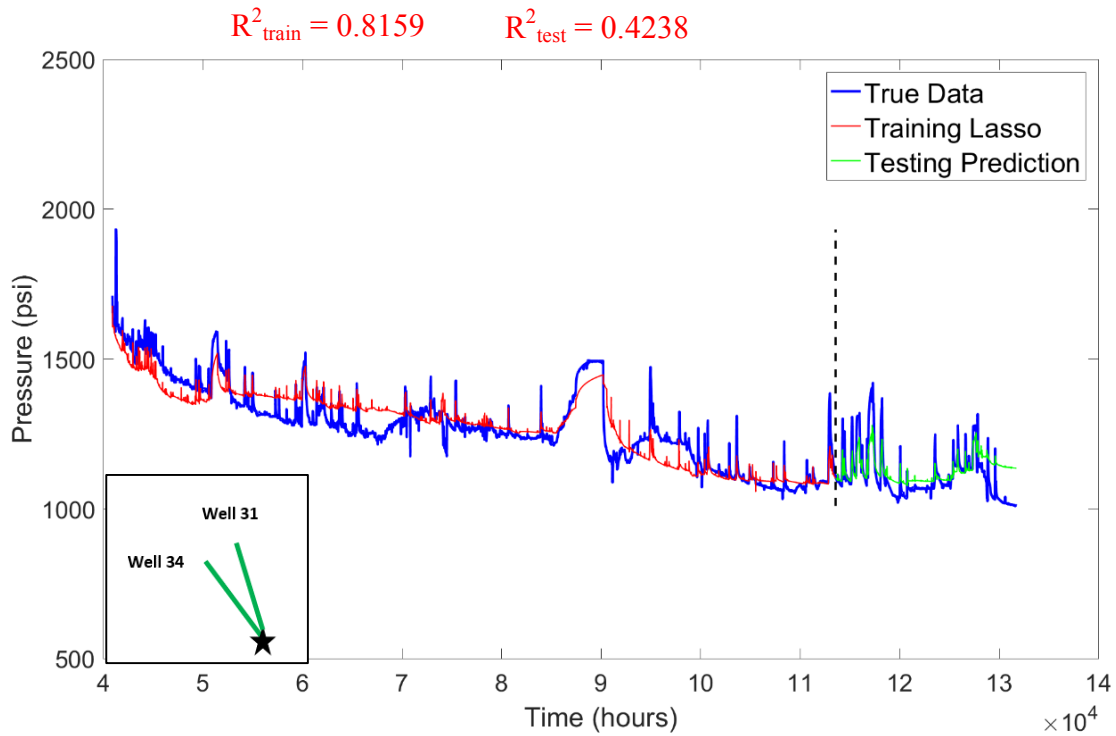

Figure A. 11   —   Pressure prediction for the Wells 34 and 31 configuration using the lasso algorithm. Compare to the base case, there are not visual improvements in pressure trend predictions by adding Well 31.

$$P_{bh \, Well \, 34} = 1701.339 + 0.0182 \, \Delta q_{Well \, 34} - 0.0133 \, \Delta q \, log\Delta t_{Well \, 34} \quad (\ 56 \ )$$
$$- 4.2623 x 10^{-07} \, \Delta q \, \Delta t_{Well \, 34} - 0.2316 \Delta q \, log\Delta t_{Well \, 34}$$
$$+ 0 \, \Delta q_{Well \, 31} - 0.0008 \, \Delta q \, log\Delta t_{Well \, 31}$$
$$+ 4.1738 x 10^{-07} \, \Delta q \, \Delta t_{Well \, 31} + 0.1278 \, \Delta q / \Delta t_{Well \, 31} \, .$$

119

Figure A. 12 presents pressure predictions for the Well 34/Well 39 configuration. Compare with base case, the lasso algorithm captures the pressure increase between 6.4×10^4 and 7.2×10^4, the start of pressure increase at 9×10^4 hours, and start of pressure decrease at 9.2×10^4 hours by adding Well 39's flowrate history. Although R^2 testing value decrease compare to base case, there are visual improvements in pressure prediction trends. Therefore, we assume that there might be interference effects between these wells.

Eq. A. 9 presents linear model for this configuration. Coefficients are calculated using the lasso algorithm. .
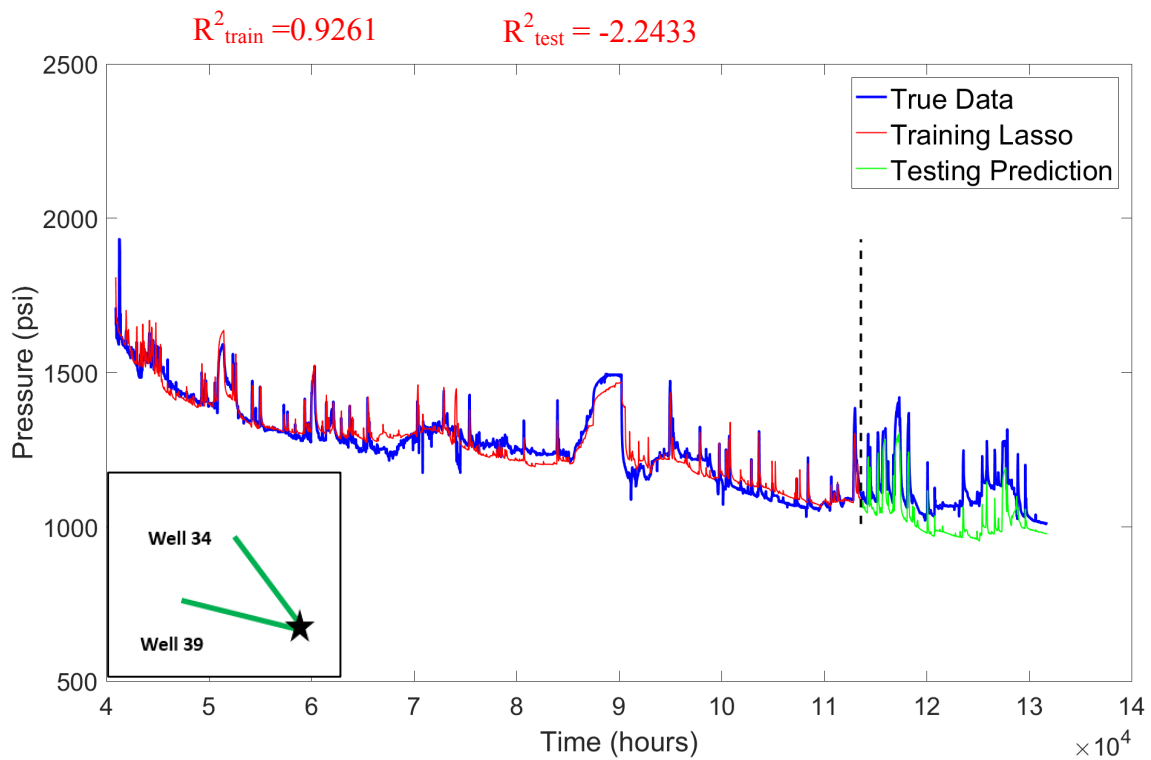

Figure A. 12 — Pressure predictions for the Wells 34 and 39 configuration using the lasso algorithm. Compare with base case, the lasso algorithm captures the pressure increase between 6.4x104 and 7.2x104, the start of pressure increase at 9x104 hours, and start of pressure decrease at 9.2x104 hours by adding Well 39's flowrate history.

$$P_{bh\,Well\,34} = 1848.025 - 0.0138\,\Delta q_{Well\,34} \qquad\qquad (\,57\,)$$
$$- 0.0055\,\Delta q\,log\Delta t_{Well\,34} - 1.511x10^{-06}\,\Delta q\,\Delta t_{Well\,34}$$
$$+ 0.2534\,\Delta q/\Delta t_{Well\,34} + 0.0272\,\Delta q_{Well\,39}$$
$$- 0.0124\,\Delta q\,log\Delta t_{Well\,39} - 1.807x10^{-06}\Delta q\,\Delta t_{Well\,39}$$
$$- 0.2147\,\Delta q/\Delta t_{Well\,39}\,.$$

Figure A. 13 presents pressure predictions for the Well 34/Well 7I configuration. The R^2_{train} *and* R^2_{test} values increase by adding Well 7I flowrate history. However, there are not improvements visually in pressure predictions. There is pressure increase at $6.1x10^4$ in pressure predictions which is not observed in real measurements. Therefore, we can not assume that there is interference effects between these wells.

Eq. A. 10 presents linear model for the Well 34/Well 7I configuration using the lasso algorithm.
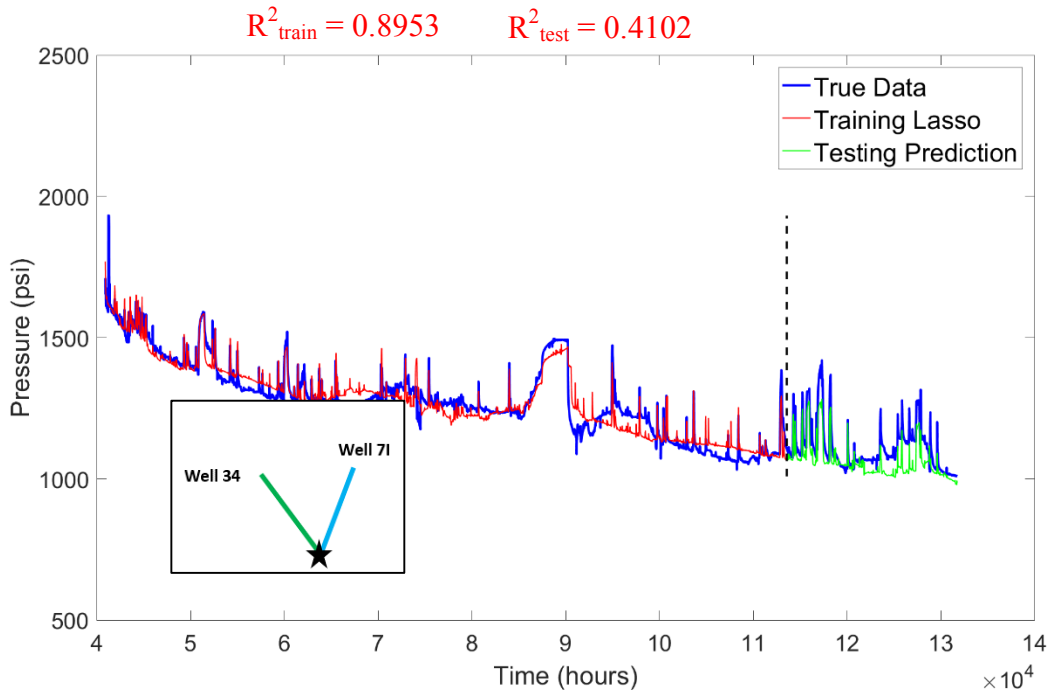

Figure A. 13 — Pressure prediction for the Wells 34 and 7I configuration using the lasso algorithm. There are not any improvements in pressure prediction trends visually by adding Well 7I's flowrate.

121

$$P_{bh\,Well\,34} = 1811.709 - 0.0157\,\Delta q_{Well\,34} - 0.0031\,\Delta q\,log\Delta t_{Well\,34} \qquad (\,58\,)$$
$$- 1.928x10^{-06}\,\Delta q\,\Delta t_{Well\,34} + 0.2413\Delta q/\Delta t_{Well\,34}$$
$$- 0.0184\Delta q_{Well\,7I} + 0.0071\,\Delta q\,log\Delta t_{Well\,7I}$$
$$+ 6.387x10^{-07}\,\Delta q\,\Delta t_{Well\,7I} + 0.2422\,\Delta q\Delta t_{Well\,7I}$$