University of Stavanger

Faculty of Science and Technology

MASTER'S THESIS

Study program/ Specialization: Computer Science	Spring semester, 2014 Open / Restricted access
Writer: Reza Asadollahi	……………………………………… (Writer's signature)
Faculty supervisor: Rui Esteves Chunming Rong External supervisor(s): Oliver Barreto (ConocoPhillips)	
Thesis title: **Predict the flow of well fluids; A big data approach**	
Credits (ECTS): 30	
Key words: Big Data, Sensor Data, Well Test, Well Fluids(Oil, Water, Gas), Downhole Pressure and Temperature, Wellhead Pressure, Gas Lift, Choke Artificial Neural Network	Pages: 90 + enclosure: Program on CD Stavanger,26.06.2014 Date/year

Predict the flow of well fluids

A big data approach

Reza Asadollahi

Faculty of Science and Technology

University of Stavanger

June 2014

Acknowledgements

I would like to express the deepest appreciation to:

Dr. Rui Esteves from University of Stavanger;

Prof. Dr. Chunming Rong from University of Stavanger;

Dr. Tatyana Kholodna from ConocoPhillips;

Eng. Henning Lillejord from ConocoPhillips;

Eng. Oliver Barreto from ConocoPhillips;

Dr. Homam Nikpey from University of Stavanger;

Eng. Vadood Azadegan, Petroleum Engineer;

Abstract

In the oil and gas industry, millions of records of data are registered every day. The data is composed by a mix of structured and unstructured sources. For example downhole gauges and wellhead sensors are logging pressure and temperature in different places of the well. Emerging technologies such as fiber-optic, wireless communication, allow the sensors to be digital, more accurate and reliable. The number of sensors as well as their resolution increase, which imposes new challenges in terms of amount of data that is necessary to be processed. Moreover, engineers in oil and gas industry store well operations, interventions logs and their interpretation in different files and formats. Combining and analyzing these different types of data is a worthy challenge for each company to obtain valuable information.

The production/performance engineers need the production rates to monitor the well situation and optimize the well performance. However, in a production well, the well production rates are not determined in real time. To measure the fluid rates Well Test operation is carried out periodically, normally once a month. However, these rates might change dramatically within two Well Test. For this reason, a model that can predict the fluid rates, gives great advantages to the production engineers to optimize the well performance in real time.

In this thesis a real usecase of an oil well was studied and an approach was proposed to process and analysis a large amount of structured and unstructured data using Exploratory Data Analysis (EDA) and design an Artificial Neural Network (ANN) to predict the flow of the well fluids. The usefulness of this method has been proved by predicting the rates of the well fluids for the usecase with reasonable low errors. Moreover, we proposed a second approach that increased even more the network accuracy[lower errors].

Table of Contents

List of Tables

List of Figures

1. Introduction

In oil and gas industry, millions of dollars are spent on improving measurement instruments such as downhole gauges technologies and wireless surface sensors, installing these instruments at the downhole or wellhead, and data acquisition and transmission solutions to log pressure and temperature permanently in different location of the well and the reservoir. Moreover, using the new technologies such as fiber-optic and digital sensors have recently increased the data accuracy and also allows the engineers to collect pressure and temperature in different points of the wellbore. These all produce more and more data.

Additionally, logging different kind of data by engineers at offshore and onshore is a daily-based job. Some information regarding well intervention and operation, interpreting of these operations are logged in Well-view and Well-history files in unstructured format.

In other hands, you may be surprised to know that the engineers do not know the production rates of each well during the production! Because normally, a group of wells are connected to one separator and the total flow rates of these wells are measured not the rate of fluids of each well. For experts and production engineers it is important to know the rates of fluids of each well in order to optimize the production, manage the reservoir and well monitoring. For this reason, normally a test separator is installed in the field and each well for example once a month is connected directly to this test separator to determine the well flow rates.

In ConocoPhillips, while this raw data is analyzed by experts to realize the well situation and its behavior there is no any dynamic model to predict the fluids flow of the well real-time based by using this data.

The purpose of this thesis is to study how this large amount of structured and unstructured data could be process and analysis to find the variables related to the well fluid rates. In addition, design a model to predict the fluid rates of the well based on these variables and data.

1.1. Thesis Outline

This thesis report is organized in 7 chapters. Chapter 2 gives an overview about the case study. Section 2.1 introduce the Ekofisk area followed by section 2.2 that give some information about Eldfisk area where our target well, well 2/7-B-19 is located. Section 2.3 talk about well specification. In section 2.4 we talk about the data we have received for this work form ConocoPhillips. Section 2.4.1 presents an introduction about well test operation and well test data while section 2.4.2 gives some information about different kind of sensor data we have received. The last section give more information about the structured and semi-structured data have been analyzed during this work.
Chapter 3 represents the theoretical background of this thesis. First in section 3.1 we talk about data processing.

In this section (3.1) we briefly talk about data processing steps that starts with preprocessing in section 3.1.1. Data cleaning, data integration and data reduction techniques are described in the sections 3.1.1.1,3.1.1.2 and 3.1.1.3 respectively. Then we will discuss data analysis in section 3.1.2 and a specific approach called Exploratory Data Analysis(EDA) is described in detail in the section 3.1.2.1. Using EDA in a big data approach is discussed in the section 3.1.2.2 followed by a small section about EDA techniques.

Artificial Neural Network (ANN) is the next subject that is discussed in chapter 3, section 3.2. First an introduction about ANN is given. A small discussion about how the human brain learns will be given in section 3.2.1 . In section 3.2.2 we present the structure of an artificial neuron model. Then in section 3.2.3 The architecture of a neural network will be discussed. This section has two subsection to discuss feedforward neural network, present in section 3.2.3.1 and the learning process will be present in section 3.2.4.2. This chapter ends with a section that discuss backpropagation learning algorithm.

In chapter 4 we present our methodology.

Chapter 5 the model is presented. The chapter is started by analysis the data in section 5.1. In this section first we discuss the output of the model in section 5.1.1. The Oil rate is discussed in section 5.1.1.1 then in the next section we discuss Water rate followed by Gas rate and then liquid rate.

In the section 5.1.2 the possible inputs of the model are discussed. This section starts with an introduction about downhole pressure in section 5.1.2.1 then the relation between this variable and production rates, Oil, water, Gas and Liquid are analyzed. Section 5.1.2.2 presents the downhole temperature and then its relation with production rates are discussed. In Section 5.1.2.3 we present wellhead pressure and its relation with production rates will be discussed. Gas lift and its relationship with the production rates is discussed in the section 5.1.2.4. And finally in section 5.1.2.5 we give a brief description about choke and its role in the well operations and its relationship with production rates.

The chapter 5 ended with section 5.2 when we discuss the design of the model for our problem after we have analyzed the model in section 5.1. The basic specification of the model is presents in section 5.2.1. Input layer, output layer and finally hidden layer are discussed in this section. The section ended with a brief talk about Matlab neural network toolbox that we used as our tools to design this model.

Chapter 6 presents experimental results and discussion. Scenario A is discussed in section 6.1. In section 6.1.1 we will discuss an experiment to choose the number of neurons in the hidden layer. In the next section, 6.1.2, we analysis the performance of the network and then in section 6.1.3 we test the performance of the network with some extra test. In the Section 6.2 we present another scenario called scenario B. Same as scenario A, we discuss the hidden nodes in section 6.2.1 then the network performance in section 6.2.2 and finally test the performance of this network with some extra data in section 6.2.3. Scenario C is discussed in the section 6.3. Same as other scenarios we discuss the hidden nodes, network performance and test the network performance in section 6.3.1, 6.3.2, 6.3.3 respectively. This chapter ends with the section 6.4, where we talk about few other scenarios which we have tested.

Chapter 7 finalizes the report with the conclusions in section 7.1 and some ideas for the future works in section 7.2.

2. Case Study

This chapter pretends to give a short overview of the reservoir and the well that we have been studied during this work. It starts with a brief introduction on Ekofisk Area and the Eldfisk filed followed by general information about the wells B18 and B19 of Eldfisk.

2.1. Ekofisk Area

The Ekofisk area lies in 70-75 meters of water at the southern end of Norway's North Sea sector, about 280 kilometers south-west of Stavanger. The reservoir is located at 2900 to 3250 meters below sea level with more than 300 meters oil column. In addition to Ekofisk, it embraces Eldfisk, Embla and Tor. All four fields are operated by ConocoPhillips on behalf of the Ekofisk licensees. [1][3]

In 1967, Balder and in 1968 Cod fields were discovered in Norwegian side of North Sea. Although these fields were not commercial that time but it showed that Norwegian side of the North Sea has hydrocarbons as well. In 1969, Philips group discovered Ekosfisk that seemed to be a huge filed and in summer of 1971, production started from this field. [1][3]

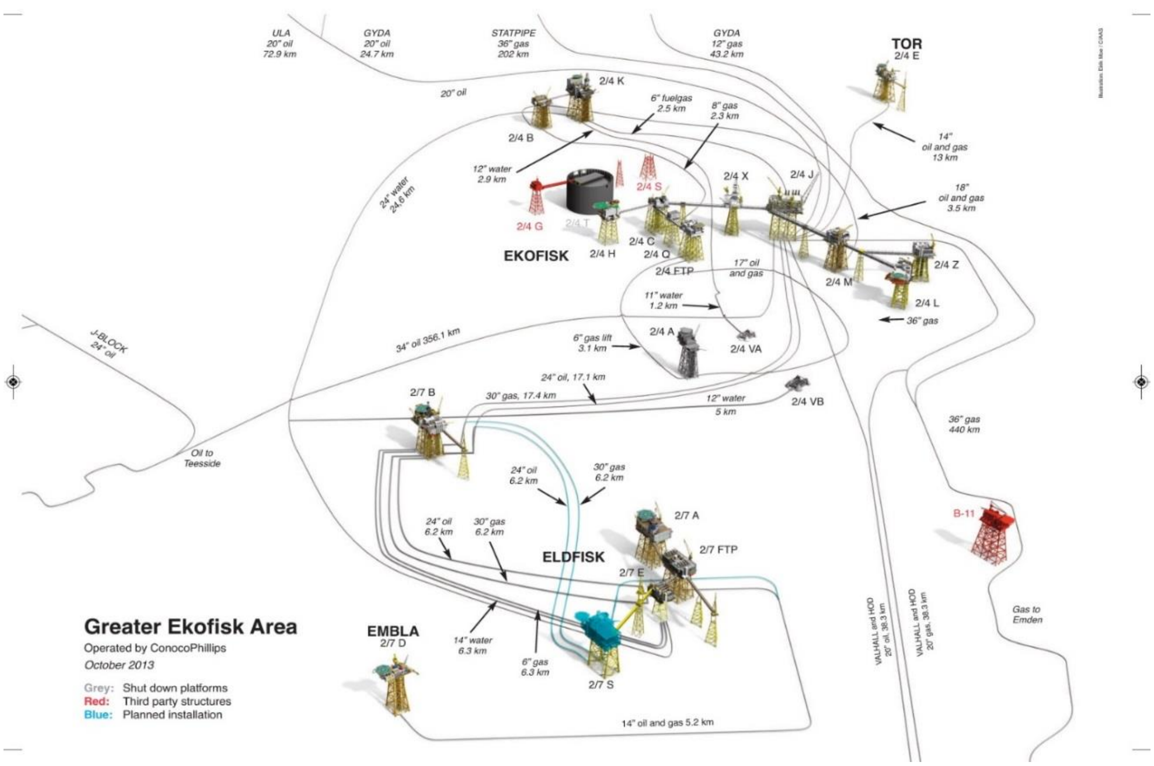

Figure 2.1 Ekofisk Area (Source: ConocoPhillips)

2.2. Eldfisk Filed

Eldfisk is the second largest of four producing fields in the Greater Ekofisk Area and one of the largest on the Norwegian continental shelf. The field is located in block 2/7, about 16 kilometers south of Ekofisk, not far from the UK and Danish shelves. The water depth in the area is just under 70 meters. The field was discovered in 1970, approved for development in 1975 and started production in 1979. [1][2]

The filed produce from chalk layer of Ekofisk that has been formed at late Cretaceous age and early Paleocene. The reservoir has rocks with fine-grained but it has a high porosity and the fluids of the reservoir can flow easily because of the natural fracturing and the reservoir in this field is located at the depth of 2700-2900 meters.

Figure 2.2:(From left) Eldfisk 2/7 A,Eldfisk FTP,Eldfisk E and the flare stack.Photo:Kjetil Alsvik/ConocoPhilips

The Eldfisk has four platforms; Eldfisk 2/7 A, 2/7 FTP and 2/7 E, are linked together by bridges [Image1] and the Eldfisk 2/7 B that a few kilometers are away from these three platforms [Image2]. Eldfisk 2/7 B is a combined drilling and production platform with a living quarter and it has its own flare stack connected with gangway. This platform was installed in 1975 and put in production in 1979. Eldfisk Bravo(2/7 B) has 19 wells produce both oil and gas, the oil separated and the gas is dewatered and compacted.[1][3]

Figure 2.3: Eldfisk2/7 B Photo:Kjetil Alsvik/ConocoPhilips

2.3. Well 2/7-B-19

Platform 2/7-B has 19 wells. In this work, we have been studying two wells, well 2/7-B-18 and well 2/7-B-19. As each well in a reservoir has its own specification and goes through different kind of operations, a decision to choose well 2/7-B-19 as main subject of study and model its production was made. We also used well 2/7-B-19 data as a supplementary subject in order to compare and analysis the data.

The original 2/7-B-19 (deviated Tor & Ekofisk producer), located in the northern part of the Bravo structure, was drilled in 1979, recompleted in 1988 and abandoned in January 1993. The well was put on production on April 2006. [37]

Wellbore name	2/7-B-19
Main area	Noth Sea
Filed	Eldfisk
Well name	2/7-B-19
Production facility	Eldfisk B
Well type	Deviated production well
Target formation	Well is a commingled Ekofisk – and Tor formation producer
Well length	TD ~13,346 ft MD RKB
Perforations	11 intervals, 6 shots each, ~1” spacing, 0° phasing (only down) 66 perforation holes total
Stimulation technique	acid stimulation with ball sealers for diversion
First production	6 April 2006
Expected well life	11 years
Slimhole desing well	No
Tubing size	11,797 ft 5½" (17#)
Liner size	6 5/8"(66#)
Gaslift	GLV installed in lowest SPM (to avoid trapped pressure in annulus if ASV close), no gaslift commenced.
PDHG	11,554 ft MD RKB (8,937 ft TVD RKB)
Stimulated	April 2006
Scale squeezed	None performed
Unexploited potential	None performed
Known tubing or liner restrictions	None

Table 2.1 Well 2/7-B 19 Specification[5][37]

2.4. Data

In this section, all the received data from ConocoPhillips is listed and for each one we give a brief information for those who are not familiar with the domain and the data.

The data has been divided into two categories, structured and unstructured data. The structured data is mostly the sensor data from PI that we have received in the spreadsheets and based on the key-value model and unstructured data is the data that have been collected by the human being such as well history or some data from SAP such as RTL (real time logging) files. The period of the data is between May 2011 and December 2013. In some cases, especially in unstructured data we have some data since 2006 when the well 2/7-B-19 was put on production again.

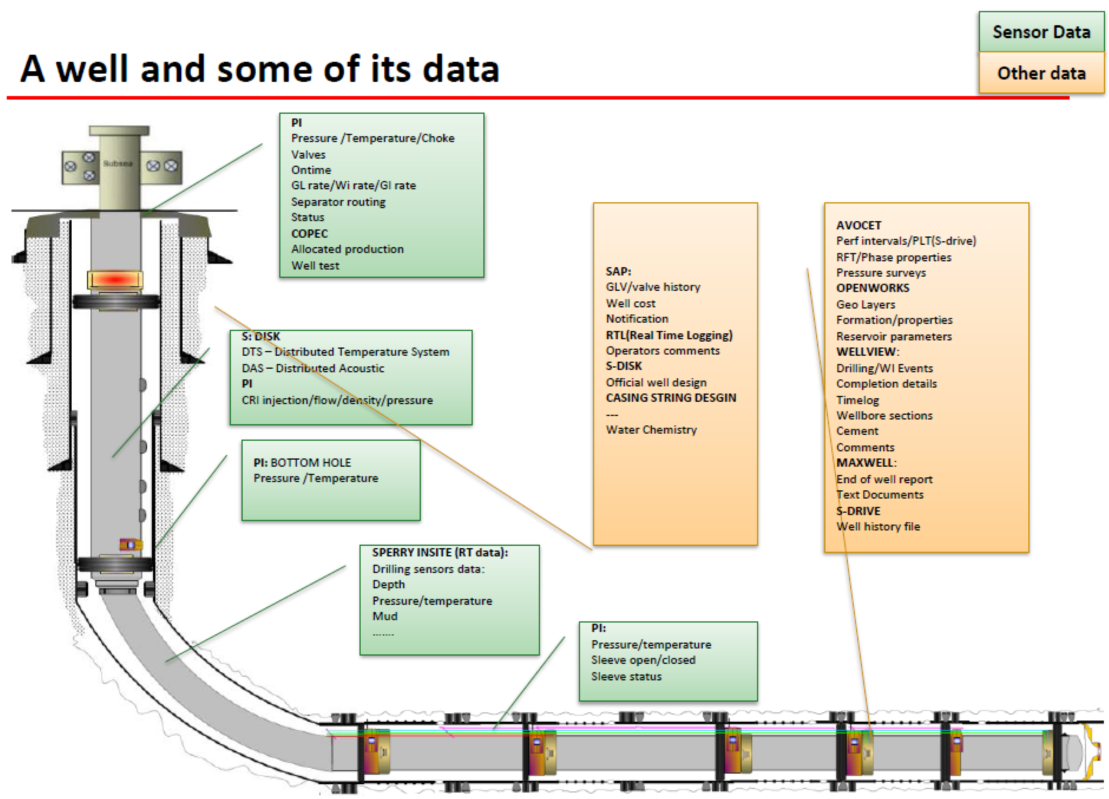

Figure 2.4: A well and its data(Source: ConocoPhillips)

2.4.1. Well Test Data

In fact, the crude oil are flowing from wellbore mixed with water, gas, condensates and different kinds of contaminants. This fluid needs to be separated into oil, gas and water to be individually measured and depending on the application be transferred for collection, controlled disposal or burning/flaring. This separation is a necessary process that must be carry out in a separator. [4]

A separator is a pressure vessel used to separate well fluids into their constituent components of oil, water and gas and can be installed in onshore well pads or offshore platforms. It can be in horizontal, vertical or even a sphere design and is used in upstream oil and gas applications as a production separator or for periodic well testing as a test separator. [4][6]

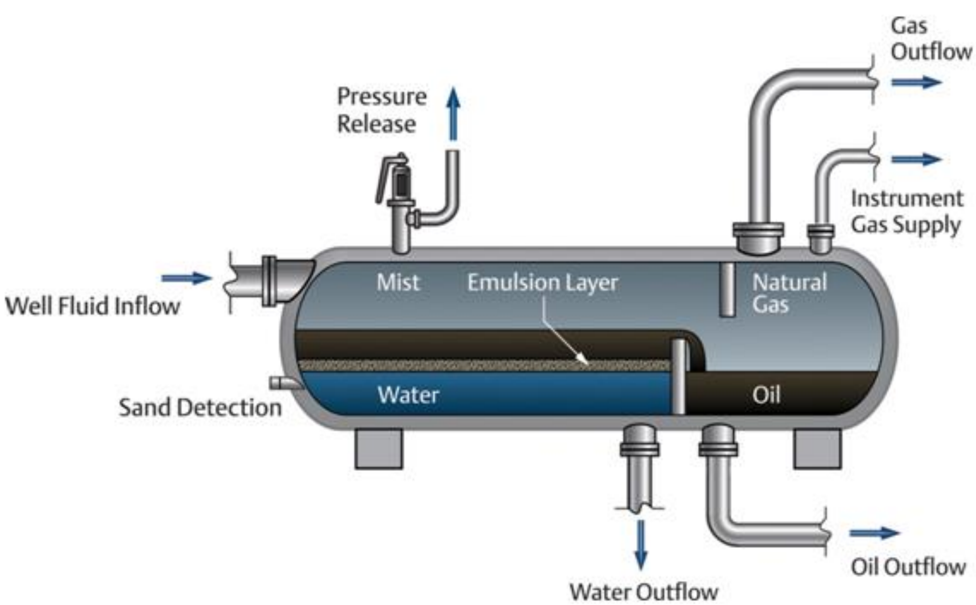

Figure 2.5: Well fluid separator (Source: 6)

In offshore fields and some onshore fields, a group of wells are connected to one separator and flows from several separators are goes into a pipeline to be transport for sale. In this case, the total flow rate of all wells are measured and the rate of fluids of each well is unknown. In other hand, it is important to know the rates of fluids of each well in order to optimize the production, manage the reservoir and well monitoring. For this reason, normally a smaller separator as a test separator is installed in the field and each well for example once a month is connected directly to this test separator to determine the well flow rates. [4]

The Separator divides the well fluid into oil, water and gas to obtain their rates and make it possible to measure their rate for that specific well.

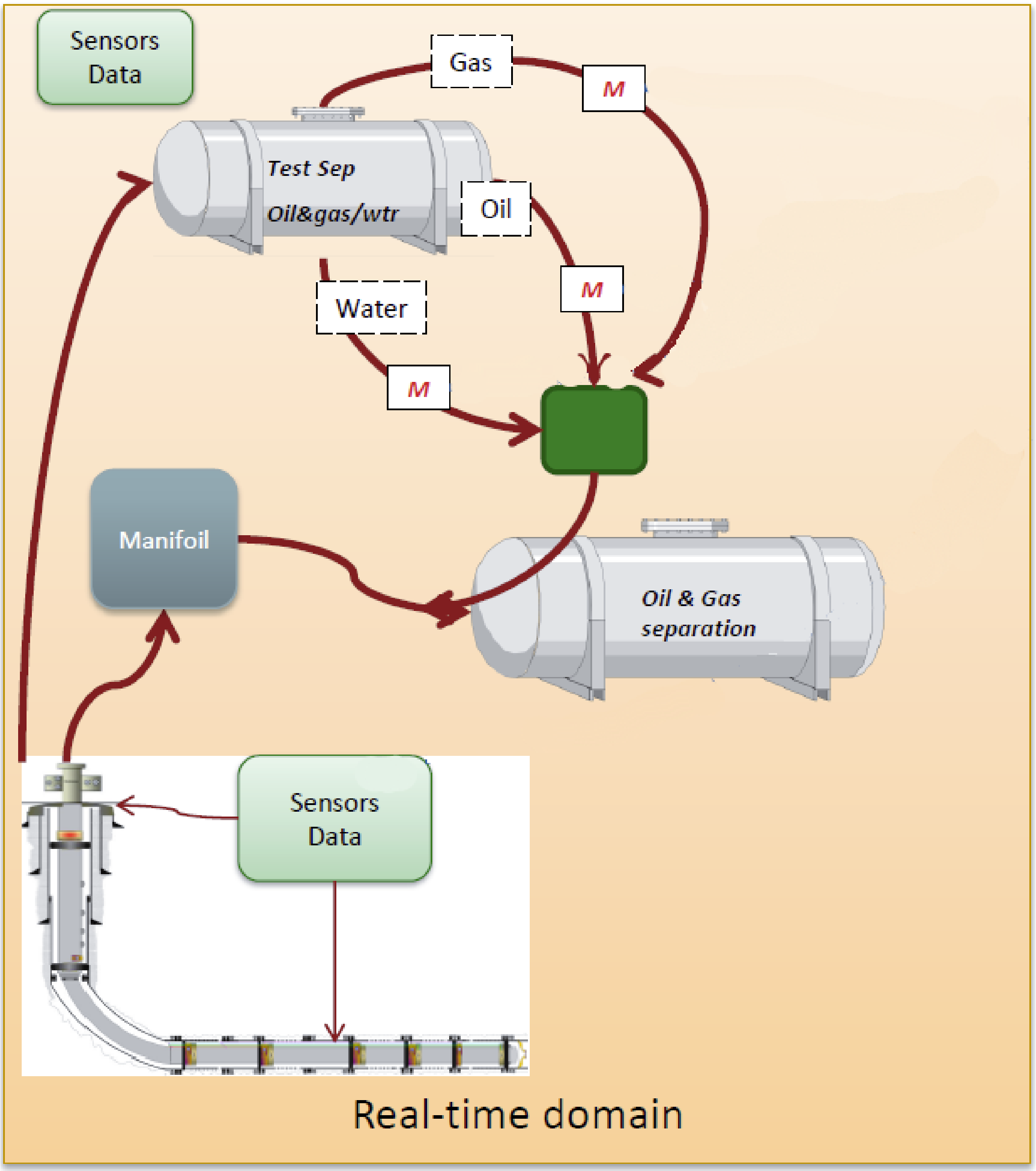

Figure 2.6: Well Test Separator (source: ConocoPhillips)

The well test normally carries out once a month for each well in a period of 4 or 8 hours with sampling interval of 30 minutes. The well is connected to the test separator for more than 8 hours but the data at the beginning and the end of the test are not included in the test data because of the changes in the pressure, temperature, and well condition are not stable.

2.4.2. Sensor Data

The permanent measurement of pressures and temperatures of a well is essential to monitor the productivity of a flowing well, monitor pressure and temperature changes in the reservoir or for example monitor performance of completions equipment, optimize their performance and so on. In other hand permanent downhole monitoring helps the engineers to manage the production more accurately and in real time.

Different type of sensors that cover different range of pressure and temperature are used in the well. They enable to transfer the data in digital or analog form. Different sensors at multiple locations in the wellbore, such as wellhead, downhole both in tubing and annuli are installed; and by using real time monitoring systems such as PI, these continues data are logged in the system. This raw data is analyzed by experts to realize the well situation and its behavior. There is another set of sensors in the test separator to measure the temperature, pressure and production rates.

The sensor data we have been working on for this project includes:

Location	**Data**
Test Separator	Pressure, Temperature, and Oil, Water and Gas rates.
Well gas lift rate	The rate and temperature of the gas lift system
Well head data	Temperature and pressure of the fluid at the well head
Down hole tubing data	Temperature and pressure of the fluid in the down hole of the well

Table 2.2 Sensor data

2.4.3. Unstructured and Semi Structured Data

Unstructured files are another part of the data that are logged for a reservoir and the wells. Well-view file is one of these unstructured data that is written by the people from well intervention and give some information about the operation that has been done on a well. Another unstructured file is Well-History file that is written by the offshore engineers and gives more detail about the operation on the well, history of the well and interpretation of the operations and data. Real Time Log (RTL) is the observation of the well for 24/7.

File	Description
Well views	Written by intervention operators, information about the operations
Well Histories	Written by the offshore engineers, history and interpretation of the operations
Real time log(RTL)	Observation of the well for 24/7

Table 2.3 Unstructured & Semi Structured Data

3. Theoretical background

3.1. Data Processing

Exploring and studying the data to reveal the pattern and relation between the variables, identifying noisy, incomplete and duplicate data and finally specifying the possibly inputs and outputs of the model is the next step in this work.

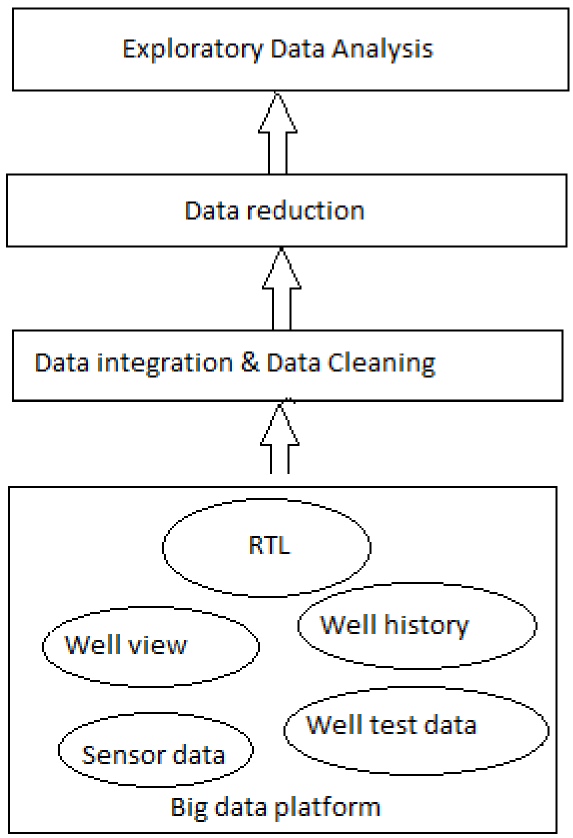

Figure 3.1: Data processing steps

3.1.1. Data Pre-Processing

Data pre-processing is a data mining technique that helps us to transfer the raw data into an understandable format, as we needed. In this project, the understandable format for us is to prepare an input and output data set from the millions of records of raw data to be used in the model.

Real world data is often incomplete, inconsistent and/or lacking in certain behaviors or trends and is likely to contain many errors and data pre-processing is a proven method of resolving such issues. Analyzing data that has not been carefully screened for such problem can produce misleading result. It is important to understand that the quality of data is a key issue when we are going to mining from it [7]. Nearly 80% of mining efforts often spend to improve the quality of data [8].

3.1.1.1. Data Cleaning

Incorrect and inconsistent data can lead us to a wrong conclusion and wrong design and subsequently wrong model. Especially in the modeling techniques such as Artificial Neural Network (ANN) that is essentially based on the data and data has the main role in the training and validation the model data accuracy is very important. Data Cleaning is a technique that has been applied to detect and correct (or remove) noisy, inaccurate, irrelevant part of the data, correct inconsistency in the data, and remove the data for the situation that may mislead the model, such as the data for the period that the well has been closed or the data before and after the well test operation. [7][10][11]

To apply this technique we have been using Exploratory Data Analysis (EDA) that will be discussed in the section 3.2.1

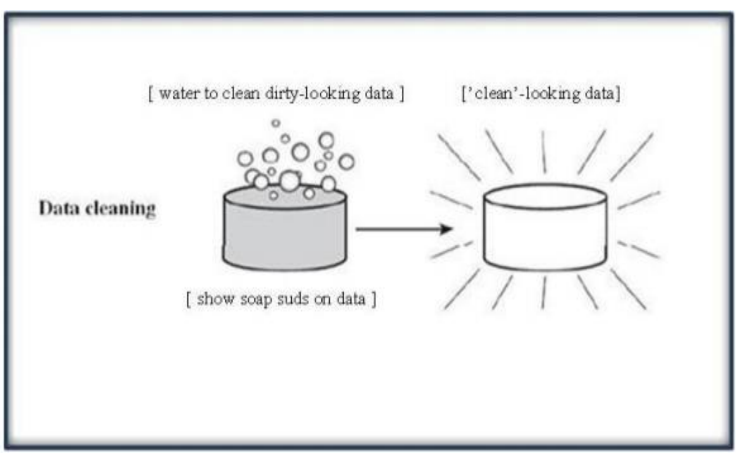

Figure 3.2: Data cleaning (Source: 10)

3.1.1.2. Data Integration

After cleaning data, the next step to prepare a dataset for the model is data integration. In this phase data from different sources such as sensor data, from multiple sensors and files, well test data etc. are merged into a single data source in order to be used in the model. [7][11]

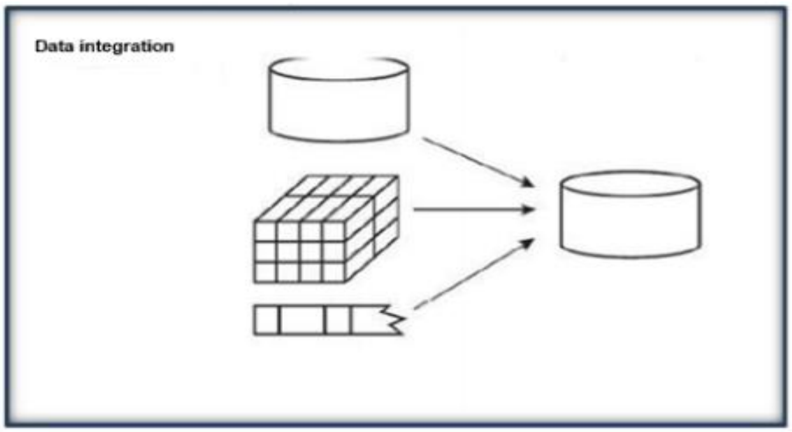

Figure 3.3: Data integration (Source: 10)

3.1.1.3. Data Reduction

Data reduction is the reduction of multitudinous amounts of data down to the meaningful parts. Further speaking, data reduction is the transformation of numerical or alphabetical digital information derived empirically or experimentally into a corrected, ordered, and simplified form. Editing, scaling, sorting, coding, collating and producing tabular summaries are some typical actions in data reduction. In this work, we have millions of records of sensor data for about two years for the target well, by using these techniques; we could manage to have an average of each tag per day to have an overview of the data and analysis the data. In other hand the well test results that are valid and trustable are about 8 hours per month for each well with time interval of 30 minutes, in this case we reduced the data to that specific test periods with average values per 30 minutes for each tag. [7][8][10][11]

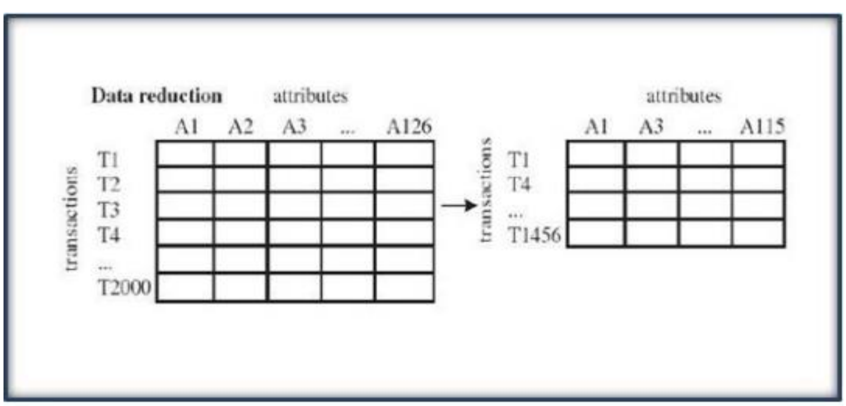

Figure 3.4: Data reduction (source: 10)

3.1.2. Data Analysis

Data analysis is the science of examining raw data with the purpose of drawing conclusions about that information [12]. Data analysis is divided into two parts: Exploratory data analysis (EDA) that focuses on the discovering new features in the data, and confirmatory data analysis (CDA) with focus on existing hypothesis to be proved as true or false. In this work, we are dealing with exploratory data analysis (EDA) part as we are working on the data to find a new method to predict the production of the well based on the existing data(discovering new features).[9][12]

3.1.2.1. Exploratory Data Analysis(EDA)

According to Ratner, B. [13]: In 1962, John Tukey in his influential paper, “The future of statistics”, claimed that the field of statistics was not enough advanced. He believed that there was too much focus on the mathematics part of statistics and not enough focus on the analysis of the data, and he predicted a movement in statistics to change the inflexibility of the discipline. He then took the first step in revolutionizing statistics by referring himself as a data analyst not a statistician. However in 1977, Tukey lead statistics into a new area known as EDA, by publishing his main masterpiece Exploratory Data Analysis. EDA offered a fresh, assumption-free, nonparametric approach to problem solving, in which the analysis is guided by the data itself, and utilize self-educated techniques, such as iteratively testing and modifying the analysis as the evolution of the feedback, to improve the final analysis for reliable results.

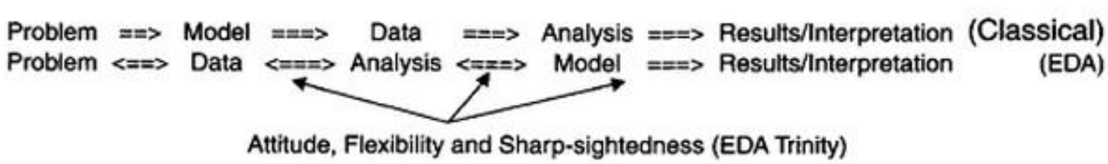

Figure 3.5: Classical and EDA data analysis

1. John Wilder Tukey (1915-2000) was an American mathematician best known for development of the FFT algorithm and plot box.

Ratner, B. [13] claimed that the essence of EDA is best described in Tukey's own words in the Tukey's seminal book *Exploratory Data Analysis(EDA).* "Exploratory data analysis is detective work, numerical detective work, counting detective work, or graphical detective work…It is about looking at data to see what it seems to say. It concentrate on simple arithmetic and easy-to-draw pictures. It regards whatever appearances we have recognized as partial description, and tries to look beneath them for new insights."[13].

Characteristics of EDA:

1. Flexibility: The techniques used in the EDA has more flexibility to dig into the data
2. Practicality: Some procedures to analysis the data
3. Innovation: Some techniques to interpret the results
4. Universality: Using statistics that can be used to analysis the data
5. Simplicity : Simplicity is the golden rule

Exploratory data analysis (EDA) is an approach/philosophy for data analysis, which employs a variety of techniques (mostly graphical) to

1. Maximize insight into a dataset;
2. Uncover underlying structure;
3. Extract important variables;
4. Detect outliers and anomalies;
5. Test underlying assumptions;
6. Develop parsimonious models; and
7. Determine optimal factor settings [8].

EDA is not about set of techniques but it is a philosophy and attitude that tells us how we can analysis the data. It is different from statistical graphics that postpone the normal assumption that what kind of model the data may follow and it mostly focus on the data itself and the model and underlying structure that data will reveal. As we mentioned before EDA is not a collection of techniques to tell us how we can analysis the data, it is a philosophy of how analysis the data, what we look for and how.

3.1.2.2. EDA and Big Data

Nowadays small data normally refers to the few sheets of paper or small files in excel files for example; fill by rows of observation of variables or features. Another feature of small data is about their properties. In addition to the small size they have, they are neat and tidy. In other hand they are “Clean” means that they contain no unexpected values and they are in the “ready-to-run” condition that was a requirement in classical statistic.[13]

EDA was originally developed for small data and there is a misunderstanding that EDA does not work as well with big data. Certainly some of the EDA techniques that are used in small data cannot be used with large amount of data. Certainly some counting and numerical method such as folding and binning and some graphical method such as the steam-and-leaf plots will crash by using the large amount of data but most of the EDA methodology are unaffected by data size. By working on big data not only the reliability of the results is not changed even the manner in which the methods carried out will not be affected. In fact, some of the most powerful EDA techniques such as smoothing, re-expressing and ladder of power are very useful techniques for large amount of data. [9][13]

Pre-processing techniques we have discussed in chapter 3.1, such as Cleaning , Integration and specially Reduction have been applied on the big data in this work to prepare the data to be used in the EDA techniques in order to analysis the data in better manner.

3.1.2.3. EDA Techniques

Most techniques that are used in EDA are graphical techniques. The reason is that EDA is all about exploring the data in the best possible way, and graphical techniques that displays the data in some sort of pictorial form gives us an open mindedly exploration of data and make the data to reveal its structural secrets and give use some new understating of the data.[9][40]

The graphical techniques that are used in EDA are normally quite simple such as:

- Plotting the raw data
    - Data traces
    - Histograms
    - Scatter charts
    - Run charts
    - Stem-and-leaf plots
    - Bihistograms
    - Probability plots
    - Youden plots
    - …

- Plotting simple statistics
    - Mean plots
    - Standard deviation plots
    - Box plots
    - Main effects plots
    - …
- Increase the probability of pattern recognition by positioning these plots by using multiple plots per page.

3.2. Artificial Neural Network

According to Haykin, S. [25], a neural network is a massively parallel distributed processor made up of simple processing units that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through a learning process.
2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.

When we are talking about a neural network, we normally talk about Artificial Neural Network (ANN). An Artificial Neural Network (ANN) is a system that process the information based on biologically nervous systems such as brain and the way they process information. The main idea of this paradigm is the structure of biologically nervous systems, where a large number of interconnected processing elements work together to solve a problem. These elements called neurons. Same as human being ANNs learn by example and the procedure that is used to perform the learning process is called learning algorithm. The learning algorithm adjust the synaptic weights of the network to create a required design objective. [25][26]

3.2.1. Why Neural Network?

During the last few years the neural networks have got a great attention. The main reason is that they offer a different approach to problem solving than the conventional approach. Conventional approach is applied when we know the algorithm and teach the computer to follow a set of instructions to solve the problem. In the conventional approach we need to understand the problem completely and clarify the steps that the computer needs to follow to solve the problem otherwise the computer cannot solve the problem. These requirements has restricted conventional approach to the problems that we already understand and know how to solve. [26][28]

The neural networks, unlike the conventional approach, are used when we do not exactly know who to solve a problem. We cannot program a neural network to perform a specific task or to follow a set of instructions. They learn by example and data has an essential role in neural network. However, there are some problems regarding this approach; First the examples must be selected very carefully otherwise the network will function incorrectly. Another problem is that in neural network there is no way to recognize if the system is faulty or not.[26][28]

3.2.2. How The Human Brain Learns?

Since antiquity, we have known that the brain is the center of learning. But for neuroscientists it has only been possible during the last decade to go inside the brain and observe the learning process. In the human brain, neurons are the fundamental processing elements. They receive their inputs from other sources, combine them perform a nonlinear operation on the result and finally send the result to the other neurons. [28][29]

Neurons collect signals from other neurons through dendrites[1]. The majority of neurons send their outputs as a series of brief voltage pulses through a very long and thin stand known as axon[2]. Each axon connects to thousands of other structure called synapse[3]. Synapses convert the signals from axon into electrical effects that can excite activity in the neurons connected to them. These neurons send a spike of electrical through their axon if the received voltage is enough large compared to their inhibitory input. Learning occurs by changing the effectiveness of synapses. [26][28]

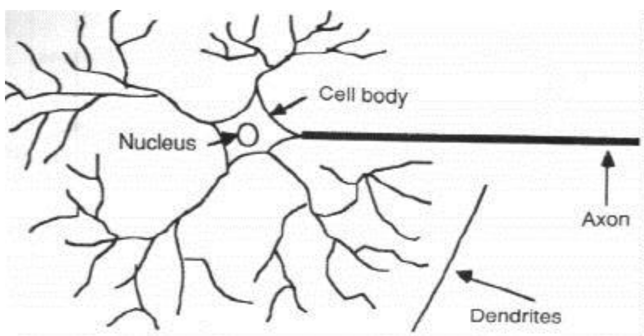

Figure 3.6: Components of a neuron (Source:29)

1. A short branched extension of a nerve cell, along which impulses received from other cells at synapses are transmitted to the cell body. [27]
2. The long threadlike part of a nerve cell along which impulses are conducted from the cell.[27]
3. A junction between two nerve cells, consisting of a minute gap across which impulses pass by diffusion of a neurotransmitter.[27]

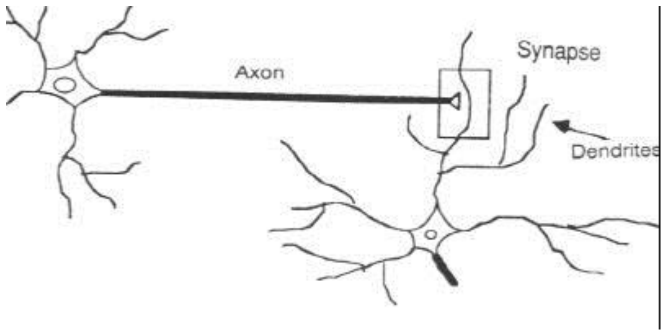

Figure 3.7: The synapse (Source:29)

3.2.3. Artificial Neuron Model

A biological neuron has some important functionality. First receives inputs from other sources, then processes these inputs and finally outputs the result. While the biological neuron is significantly more complicated than this structure, the artificial neuron is designed based on this basic structure of the biological neuron and simulate its basic functions. Figure 3.8 shows the structure of an artificial neuron. The inputs represented by x(n) symbol multiplied by corresponding connection weight w(n). Then a transfer function applied to the sum of the products to generate a result as output.[26][28][29]

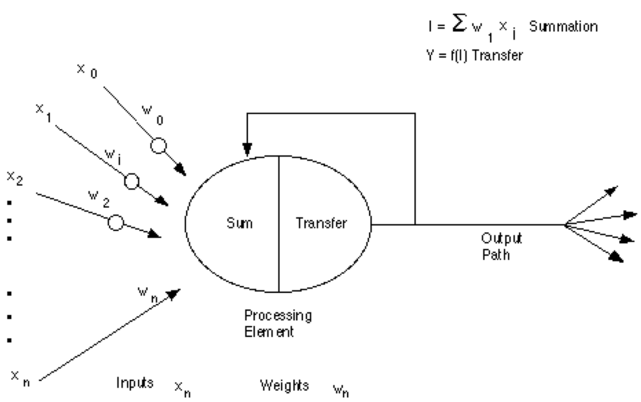

Figure 3.8: Basic Artificial Neuron (Source: 29)

3.2.4. Artificial Neural Network Architecture

In artificial neural network, same as human brain, (artificial) neurons are the fundamental processing elements. An ANN consists of interconnected neurons organized in layers.

- An Input layer: This layer represents the raw information for the network.
- Hidden layer(s): One or more hidden layers to transfer the inputs into something that can be used by output layers.
- An output layer: Send the generated result as output to the outside world.[28][29]

3.2.4.1. Feedforward Neural Network

Feedforward networks, known as multilayer perceptrons are the most popular neural networks. They have been applied successfully to different type of problems such as image recognition, financial prediction, speech recognition, medical diagnosis and many other applications. In these networks, the information only moves in forward direction. Figure 3.9 illustrate a feedforward neural network with one hidden layer and shows how information moves from input layer to the hidden layer and then from the hidden layer to the output. Recurrent neural networks are another type of networks with feedback connections that enable the nodes to fed the information back to the previous layers.[30][31][32]

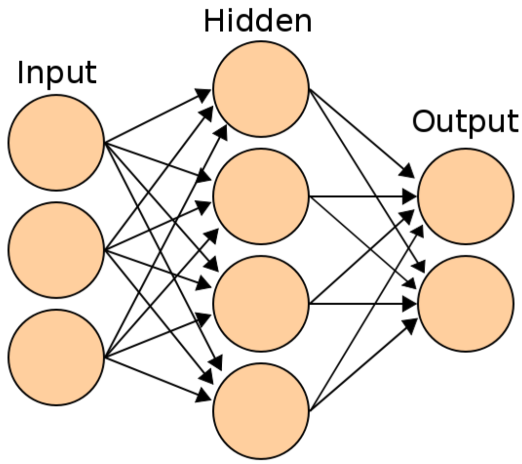

Figure 3.9: Feedforward neural network with one hidden layer

3.2.4.2. The Learning Process

The knowledge of a neural network is represented by the values of the connection weights between the neurons and the neurons thresholds. The learning process is the determination of these values.

The learning processes are classified into two main categories:

- **Supervised learning**

In supervised learning, the training data consists of pairs of input and desired output values. In this process, a supervised learning algorithm is used to produce an inferred function. This function is called classifier is the output is discrete or regression function if the output is continuous. The weight matrix is adjusted during the learning process and the network will be able to predict the correct output for unseen valid input. [33]

- **Unsupervised learning**

In unsupervised learning, the training data consists of input values only. This method is used to find the hidden structures in the data. It can cluster the data into different classes. Some common example of unsupervised learning are the Hebbian learning rule, and the competitive learning rule. [33]

In this work, we are using supervised learning process. Therefore, we will not discuss unsupervised learning process further. In next section, we discuss one of the most popular supervised learning algorithm that we are using in the network for this work.

3.2.4.3. The Back-Propagation Algorithm

In a supervised learning, a learning algorithm is used to adjust the weights of the units in a way that error between the output of the model and supervised outputs is reduced. To fulfill this process, the neural network compute the error derivative of the weights. The algorithm increase or decrease the weights of the units slightly and then recalculate the error again to see how the error changes. [25][26]

The back propagation algorithm is one of the most common supervised learning algorithm to train the network. First the weight matrix initiated by some small random values. Then the input data is applied to the network and the output is calculated. This phase is called *feed forward*. In the next step that is called *backpropagation*, the error of each neuron is calculated in the backward direction. It means these errors are calculated from the output layer and then the hidden layer just before the output layer and so on. This is what gives back propagation its name. Then these errors are used to change the weights of each neuron in such a way that errors get smaller. This process will be repeated until the error is minimal.[25][26][33]

Back propagation steps for a three-layer network:

1. Initiate the weights matrix with random values
2. Predict the output for the input example(feed forward)
3. Compute errors at the output layer for output neurons
4. Compute errors for the weights between the hidden layer and the output layer(backpropagation)
5. Compute errors for the weights between input layer and hidden layer(backpropagation)
6. Update the weights based on the calculated errors
7. Repeat steps 2 to 6 until reach the minimum error(based on the defined conditions)[34][38]

4. Methodology

To fulfill this work, the present methodology was followed:

- Preprocessing the data (Cleaning, integrating and reducing the data);
- Analysis the data using Exploratory Data Analysis(EDA) approach and techniques;
- Analysis and study the outputs of the model using EDA;
- Determination of the model inputs using EDA;
- Design a proper model using Artificial Neural Network;
- Design several scenario and test the networks performance;

5. Model

5.1. Data Analysis

Working with Big Data in this work is not only about the different type of data and dealing with a large amount of data, it is also about finding possible variables that may affect the model. In this section, we are exploring the data first to find the variables that could possibly have any relationship with the fluid rates of the well such as Oil, Water and Gas. In the next step, we are exploring the data for those variables and the possible relationship between the variables and the fluid rates. For each variable we apply preprocessing techniques to clean, integrate and especially reduce the data, then apply EDA techniques to analyze the data.

5.1.1. Outputs of the model (Production Rates)

The fluids that flow from the reservoir to the surface of an oil well is a mixture of oil, water and gas that is called *production fluid* [1]. The characteristics and composition of this fluid vary. As we described in the section 2.1, a separator (production or test separator) is used to separate this fluid into their constituent components of oil, water and gas.

Therefore, the outputs of the model are almost clear even though it is possible we do some changes during the modeling to find the best outputs fit the model. As we mentioned in section 2.1, during the production operation the total flow rate of all wells are measured and the flow rate each well is unknown and to specify the fluid rates of each well we have to connect this well to a test separator and carry out the well test operation. This operation normally applies once a month per well in a period of 8 hours. The data normally are available per 30 minutes for the rates. Therefore, this situation restricted us to concentrate on test periods for inputs and outputs such as sensor data, gas lift and so on to design the model properly. However, as we will describe in the next section, we are using the whole continues data, especially sensor data for pressures and temperatures to reveal the relationship between these variables.

---

1. Production fluid: A fluid mixture of oil, water and gas in formation fluid that flows to the surface of an oil well from reservoir. Its consistency and composition varies (Wikipedia).

5.1.1.1. Oil Rate

The oil rate is measured during well test operation. This operation normally performs once a month for each well. The unit that we are using for this rate is barrel oil per hour (BBLS/H).

A simple run-sequence plot can graphically summarize the oil rate during the period of the study. The period of study is from January 2011 until December 2013. We must mention that in some periods the well has been shut in for some operation.

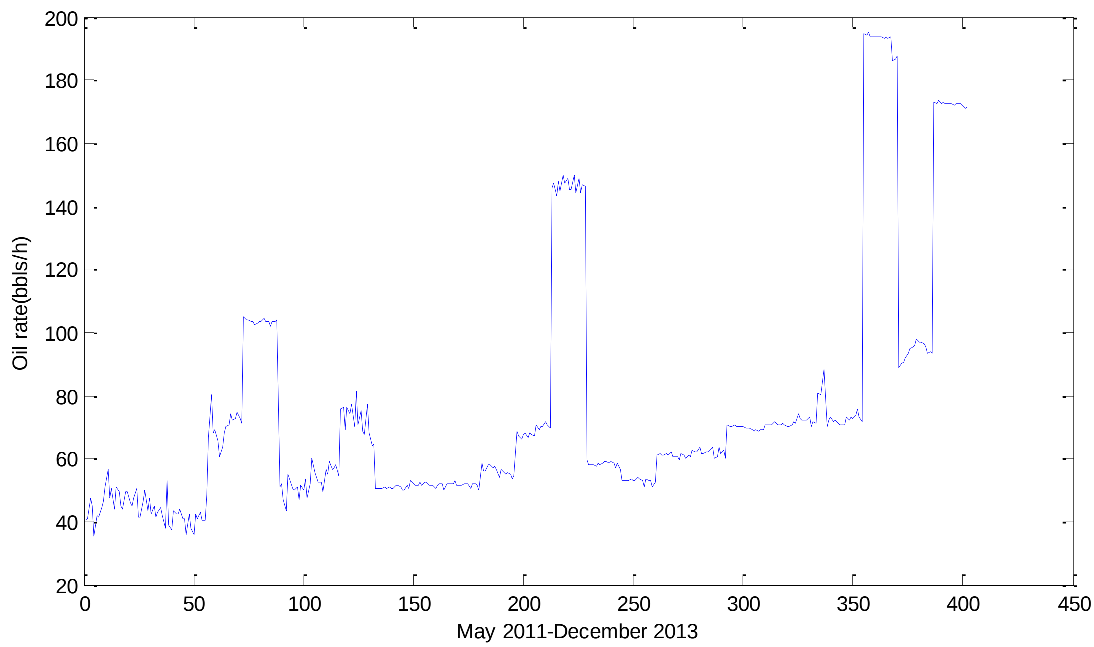

Figure 5.1: Oil rate

As you can see, in Figure 5.1, the oil rate run-sequence plot reveals that in some periods, the rate has increased dramatically and then it has decreased afterwards. Generally, if we ignore the exceptional changes in the rate the slope of the graph is positive that shows a general increase in the oil production for this well during this period.

5.1.1.2. Water Rate

In oil and gas reservoirs, normally water is mixed with the hydrocarbons and brought along with oil and gas during the production. This water is separated from oil and gas known as produced water[1]. This water may contain some chemicals, oil and metal that considered as industrial waste and must be treated before disposal. Part of this water may be re-injected back to the reservoir to increase the production and the rest must be treated. Water treatment is an important issue in oil industry as the amount of produced water increases eventually in all oil wells. *The Produced Water Society* (PWS)[2] was formed to study and improve water treatment for both industry and environment. [14][15]

The flow rate of the water is in barrel water per hour (BBLS/H). To study the water rate we use a sequence run chart to summarize the water rate during the study period.

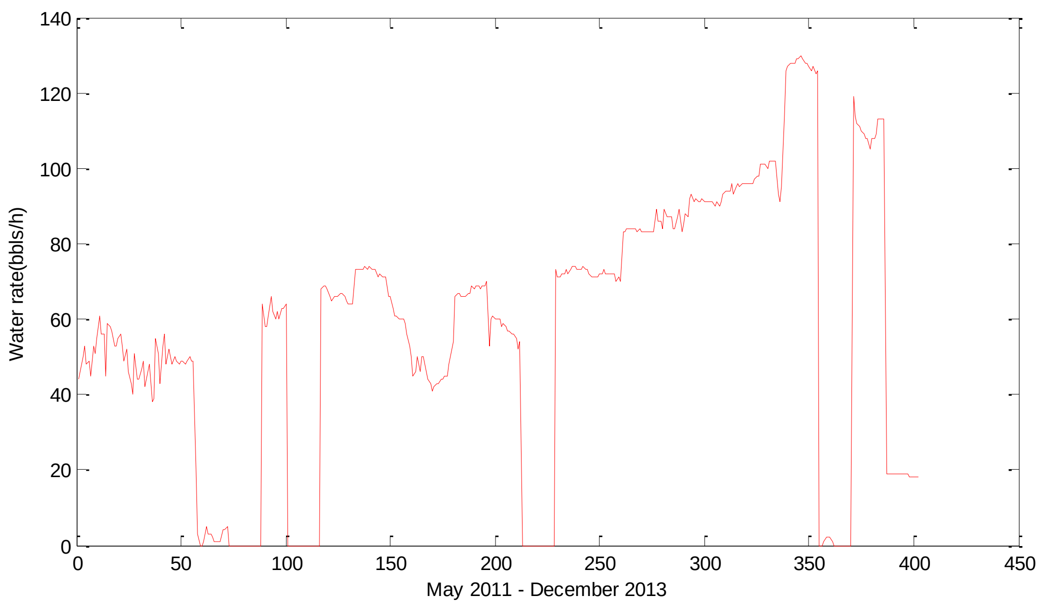

Figure 5.2: Water rate

Figure 5.2 shows that the water rate for the target well has increased steadily in the past few years. There are few exception that water rate is zero. Nevertheless, after a short period the water rate increased to the previous rate again.

----

1. Produced water: Is a term used in oil industry to describe water that is produced as a byproduct along with the oil and gas. (Wikipedia).
2. http://www.producedwatersociety.com/

As oil and water are both part of the liquid rate of the well. We decided to study their changes in the one graph. The result are shown by two run sequence graphs in Figure 5.3. By using these two plots, we can see an interesting relationship between water and oil rates. In the exact period, that oil rate has increased dramatically; the water rate has decreased actually to zero. In other periods, water rate is slightly more than oil rate and seems the water rate and oil rate are diverging by time. This situation could be explained as we mentioned before that all oil wells eventually produce more and more water. We can also see that the slope of water rate is a bit steeper than the slope of oil rate.

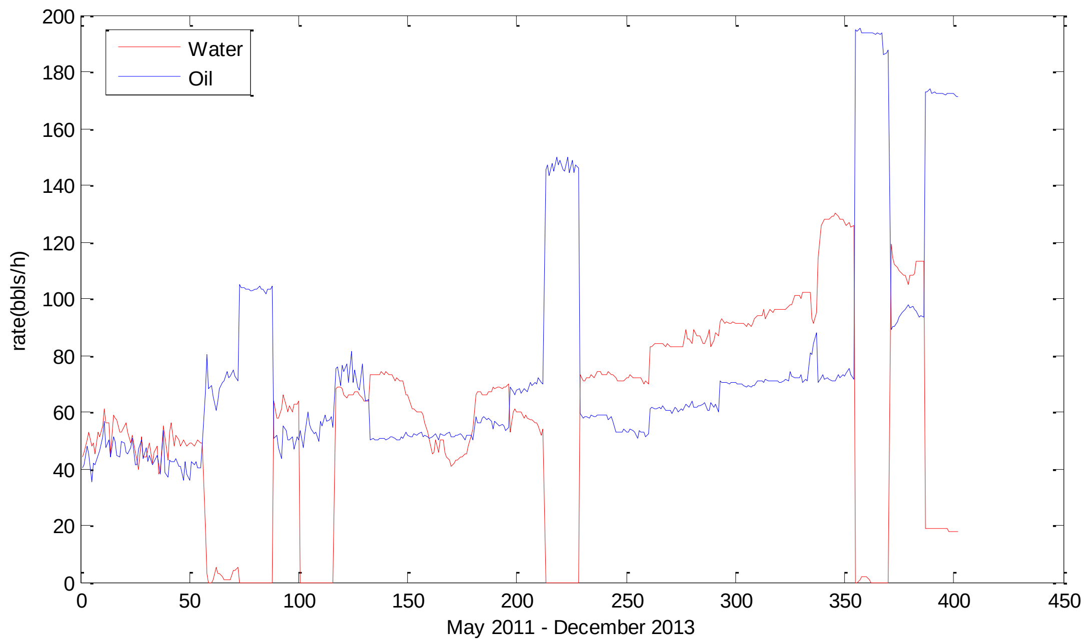

Figure 5.3: Oil and Water rates

We will discuss the relation between oil and water rates and the input variables such as downhole pressure, downhole temperature, wellhead pressure and gas lift rate in the section 5.1.2 in details.

5.1.1.3. Gas Rate

In the petroleum industry, we have oil reservoir and gas reservoir. In the gas reservoirs, the wells are mainly designed to produce gas and engineers refer these wells as gas wells. The other wells that are designed to produce mainly oil called oil wells but usually some natural gas produced during production in these oil wells. The natural gas as same as oil is produced, measured and transferred to be used in the market. The natural gas not only can be used in the homes, business and other places to heat the places, dry clothes, cook foods and so on, it is also used in industries to provide energy for different manufacturing and can be used to produce electricity and so on.[40][41]

The well that we are studying is an oil well that also produce considerable amount of gas. The flow rate of the gas is standard cubic feet per hour(SCF/H). We use a sequence run chart to summarize the gas rate during the study period.

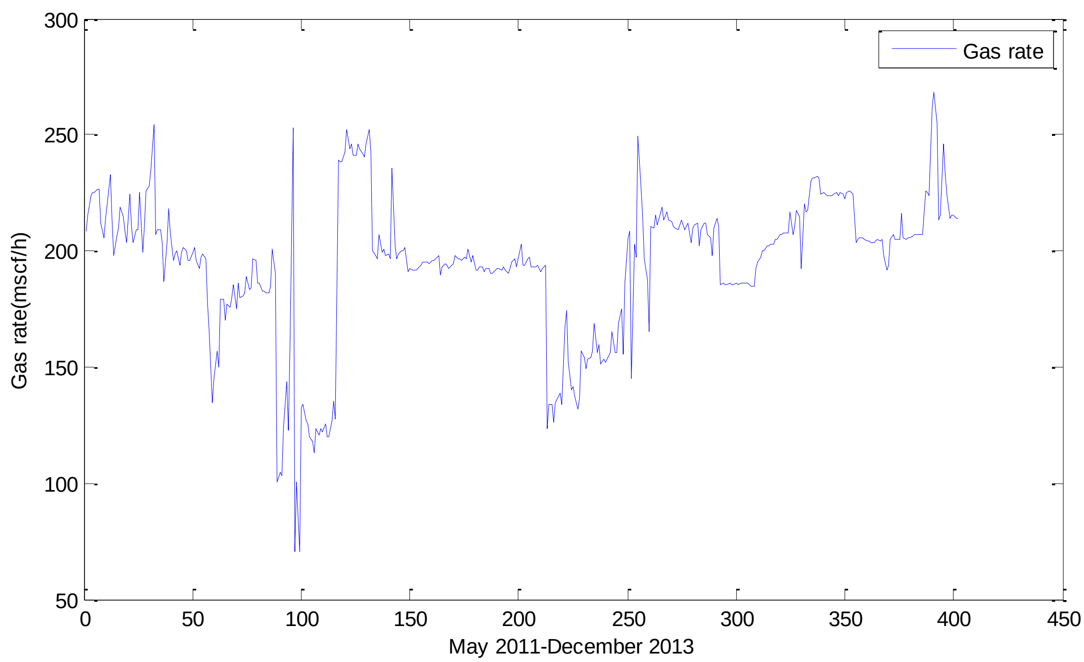

Figure 5.4 Gas rate

As you can see in the graph we have some changes during the production, we will discuss these changes in section 5.1.2 to see what has happened in the well and reservoir that caused these changes. Generally we can say that the average gas rate is constant and around 200,000 SCF/H or 200 MSCF/H.

The fluid rates of a wellbore consist of these three main production, *Oil, Water and Gas*. Figure 5.5 illustrates these three fluid rates during the study period, from May 2011 until December 2013. We should mention that these period is not continues as in some periods the well has been shut in for few months.

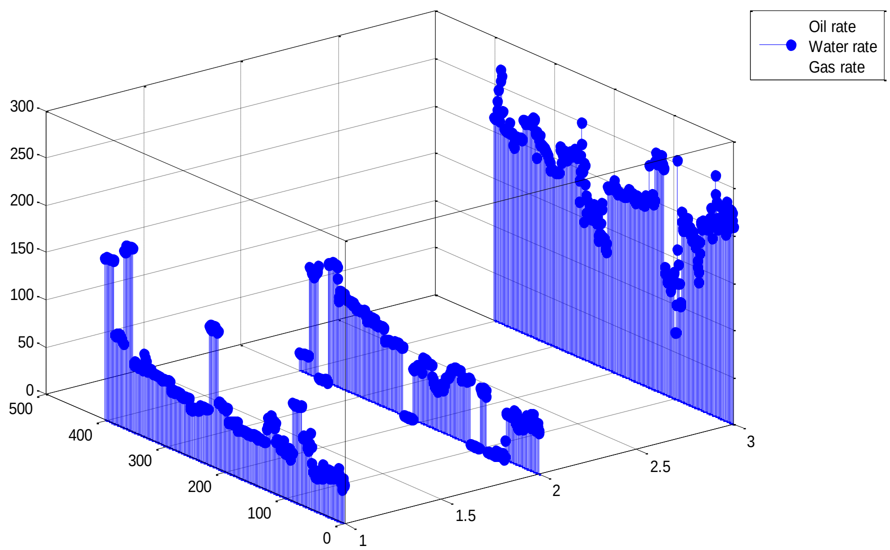

Figure 5.5: Oil, water and gas rates(May 2011 – December 2013)

Figure 5.5 shows a sequence diagram for Oil, Water and Gas rates in a 3D graph together. This graph give us a better overview of the rates and their changes during the time. If we ignore the exceptions, we can conclude that the oil rate has increased slightly while the water rate has increased more than oil, there are also some sudden decreasing in the water rate but at the same time sudden increasing in the oil rate, that shows somehow the fluid rate are more stable as we will see in the next section. The graph also shows that the gas rate has been almost constant.

5.1.1.4. Liquid Rate

Liquid rate is sum of the liquids, oil and water, of the wellbore during the production. Liquid rate is a playing factor for reservoir monitoring and simulation purposes. It gives an indication about the well productivity and reservoir inflow performance, and used mainly in reservoir simulation and voidage replacement studies. [41] In Figure 5.6 we illustrate oil, water and liquid rates together to have a better overview of their changes.

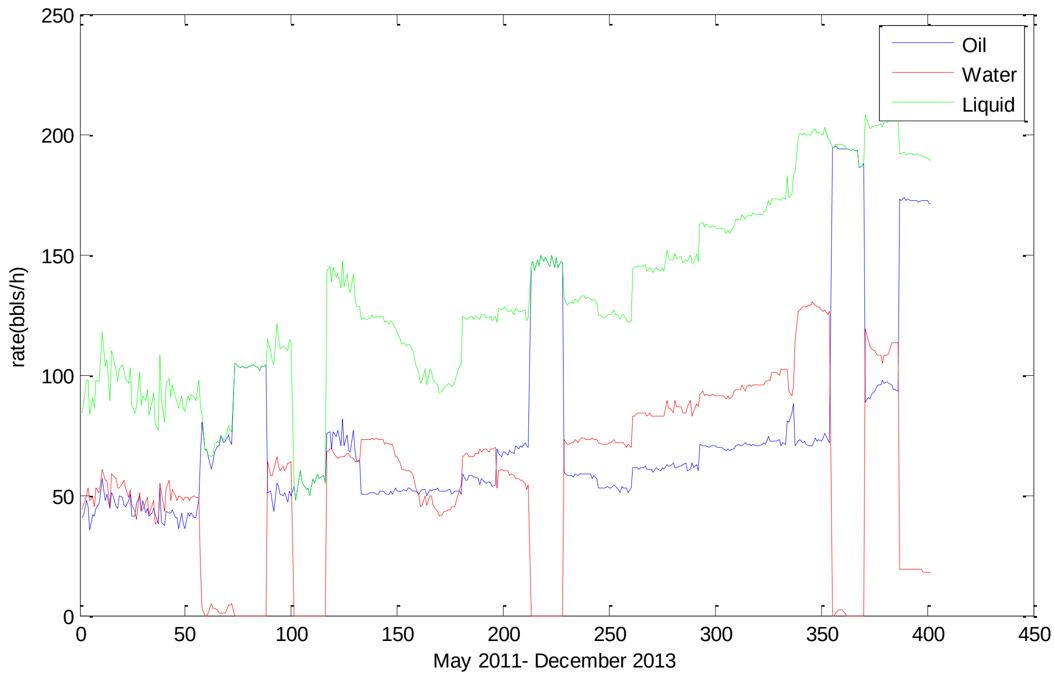

Figure 5.6: Oil, water and liquid rates

Figure 5.6 shows that the liquid rate, as we expected, has a positive slope same as oil and water. As you can see, the liquid rate is more stable than oil and gas rates because sudden changes in one of the liquid is compensate by the other liquid therefore the sum of the liquids is more stable. This situation gives us an indicator that liquid rate could be a possible output for the model. We will discuss this possibility in the section 6.1 as scenario A.

5.1.2. Possible Model Inputs

In this section, we study possible variables that could have any relationship with the fluid rates. In each section, we give a brief information about the variable, then we will have an overview of its changes during the study period, and finally we discuss its relationship with the fluid rates.

5.1.2.1. Downhole Pressure

In petroleum engineering, there is nothing more important than knowing the pressure at the bottom of the oil well (downhole) during the operations such as drilling or production. The downhole pressure has an essential role in reservoir management and production optimization. The engineers have been using bottom-hole pressure since 1930 to increase production performance and solve the problems of the reservoir. During the past two decades, measurement tools have improved significantly and produce more and more reliable, real-time and accurate data about down-hole conditions that have made the engineers to rely on this data and use it more and more for various purposes. This data is continuously collected by monitoring devices called permanent *downhole gauges* (PDG). [16][41]][44]

While we could collect data for fluid rates from May 2011 until December 2013, the sensors data for the target well were available from January 2011 until end of December 2013. To catch a better view of the downhole pressure changes we preferred to use the whole data from 2011 until 2013 instead of study the same period of the fluid rates.

Millions of records of downhole pressure has been registered during the past few years for the target well. To have an overall view of the downhole pressure, we started with calculation of the average of the downhole pressure per day for the period of study. Then we illustrate the result in a run-sequence plot. Figure 5.7 shows a run sequence plot of downhole pressure.

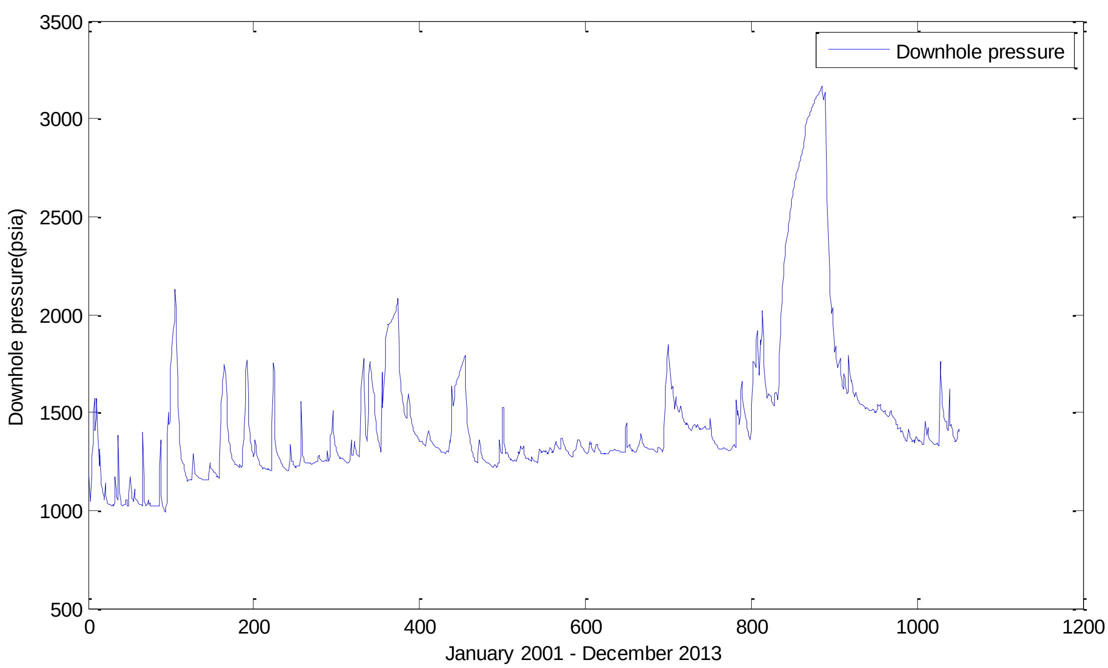

Figure 5.7: Downhole pressure

To analysis this variable, we need more information about the well situation and well status during this period. Well-log and Well-view (Section 2.3) give us general information about the operations and well interventions that have happened on this well during the past few years. Another variable that gives us a better and accurate information about the well conditions is called Flow-Status and available in our data we have collected for this well. Each well can have different status during the operation. Table 5.1 shows the possible status of a well.

No	Description
0	Shut-in
1	Natural Flowing
2	Gas Injection
3	Water Injection
4	Gas lift
5	In Test Separator
6	Slurry Injection

Table 5.1 Flow status parameters

Figure 5.8 shows an overview of the flow status for the target well. We realized that the state of well could change several times during some days. As Figure 5.8 shows, in some periods, these changes mostly have happened between Gas lift and Test Separator. We realized the well have been set to natural flowing sometimes but not for long time as it seems the well needs gas lift operation to continue flowing.

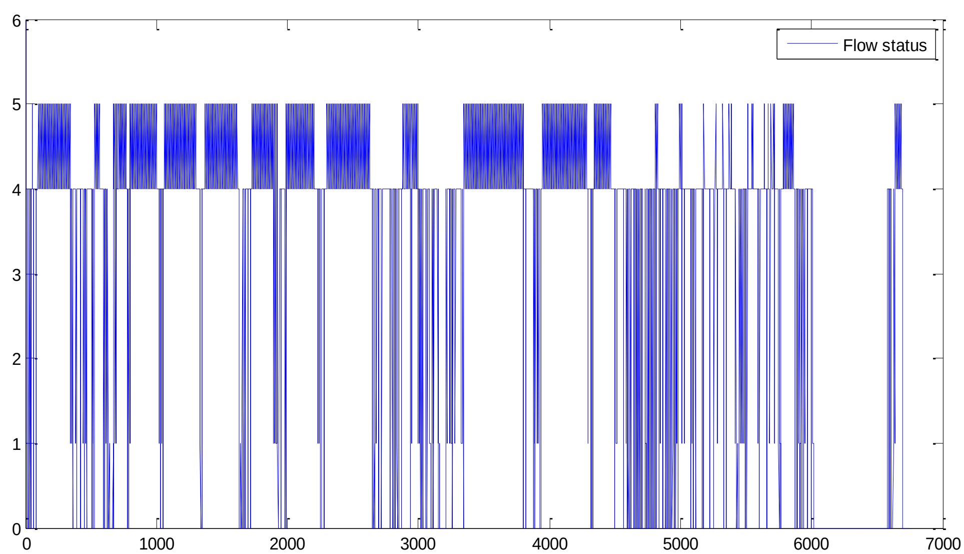

Figure 5.8: Flow status

Studying both downhole pressure and flow-status, shows that shutting in the well is the main reason of dramatic changes that have happened in the downhole pressure. For example, as you can see in Figure 5.7 during a period of about 100 days the pressure build up to more than 3000 psi. The status flow shows that during this period the well has been shut in for some reason that the reason is not in our interest in this work. We also realized that in other instances that you can see the pressure has increased suddenly the well has been shut in for few days. The study also shows that whenever the well test has applied after few days of well shut-in the rates are higher than normal rate due to the higher pressure. However, this pressure is decreasing during the time when the well has been opened again until the pressure almost has got back to the normal pressure and in some cases a bit higher than the previous pressure. As the Figure 5.7 shows, the normal pressure for this well is between 1200 and 1400 psi.

To understand the relation between the down-hole pressure and the production rate, first we need to calculate the average pressure for each 30 minutes of the down hole during the well test periods. A simple scatter plot can reveal possible relation between down-hole pressure and the production rates.

5.1.2.1.1. Downhole Pressure and Oil Rate

We start this analysis by using a scatter plot for the oil rate against the downhole pressure during the period that both rates and pressure are available, from May 2011 until December 2013, to analyze the relation between these two variables(Figure 5.9).

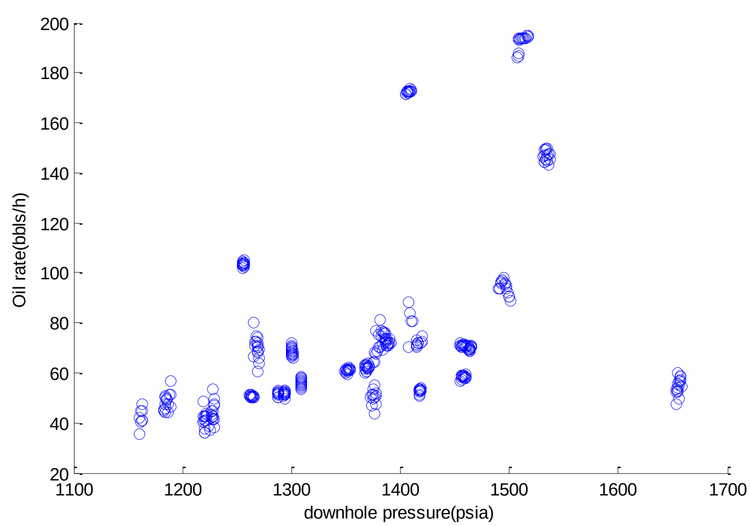

Figure 5.9: Oil rate against downhole pressure

The scatter plot in Figure 5.9 shows that how the oil rate may vary with downhole pressure. The plot reveals a direct relation between the oil rate and the pressure. The oil rate is increasing by increasing the downhole pressure. The plot shows that at pressure between 1500 to 1600 psi the rate are too high compared to other rates. Our study shows that this happens because the water rate actually in this period is zero. In other hands, this lead us to the fact that we can somehow study the relation between the liquid rate and the pressure as well.

5.1.2.1.2. Downhole Pressure and Water Rate

The scatter plot for water rate against the downhole pressure reveals that water rate as same as oil rate has a direct relation with downhole pressure. This relation here seems to be stronger than oil. In the other hand, we can claim that by increasing the pressure at downhole in this well the water rate has increased more and more as the imaginary line has a steeper slope compared to oil rate line.

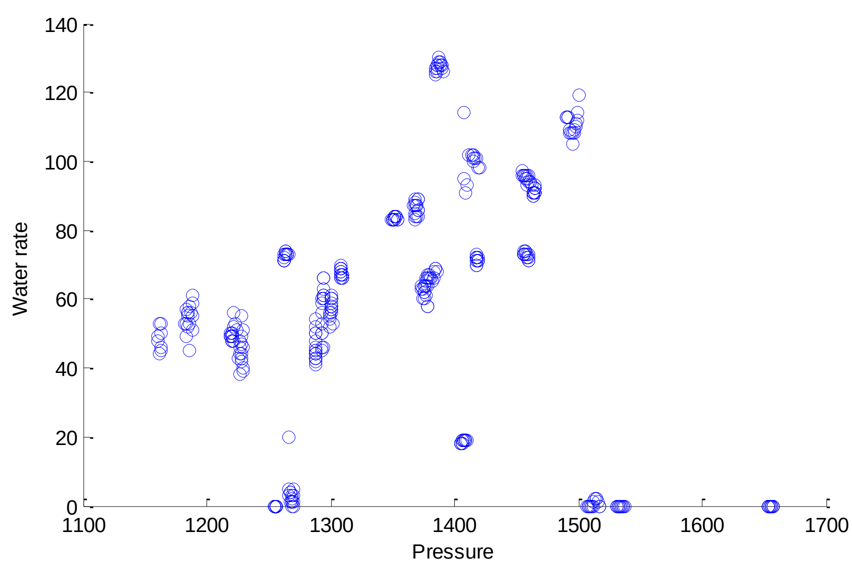

Figure 5.10 water rate against downhole pressure

5.1.2.1.3. Downhole Pressure and Liquid Rate

We use liquid rate, sum of oil and water rates, to have a better overview of the relation between downhole pressure and the well productions. As graph X shows, this relation is a direct and more obvious than the relation between oil and water against down-hole pressure separately.

The graph also shows that the slope of the imaginary line of liquid rate against downhole pressure is steeper than oil and water slope and clearly shows that the liquid rate increases when downhole pressure increases.

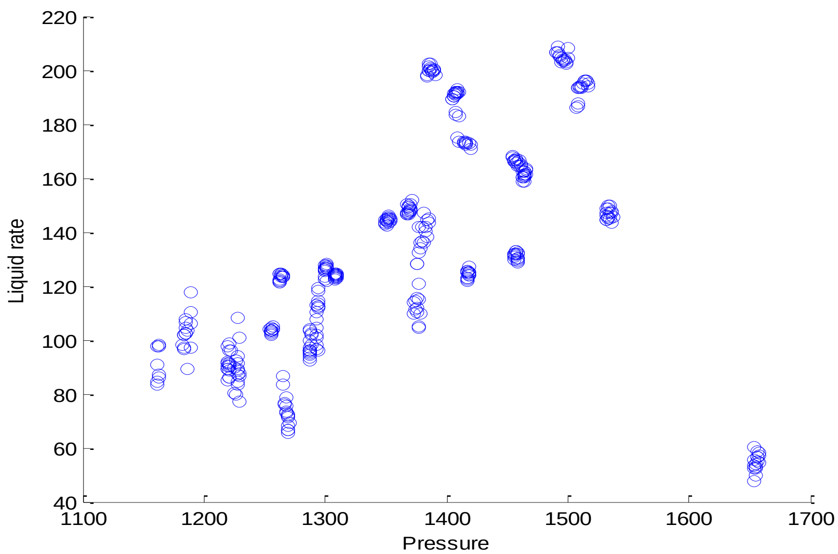

Figure 5.11: liquid rate against downhole pressure

Finally, we can conclude that there is a direct relationship between downhole pressure and the liquid rates as both oil rate and water rate are increased by downhole pressure. So downhole pressure could be considered as an important input on the model.

5.1.2.1.4. Downhole Pressure and Gas Rate

In this section, we discuss the possible relation between the downhole pressure and the gas rate.

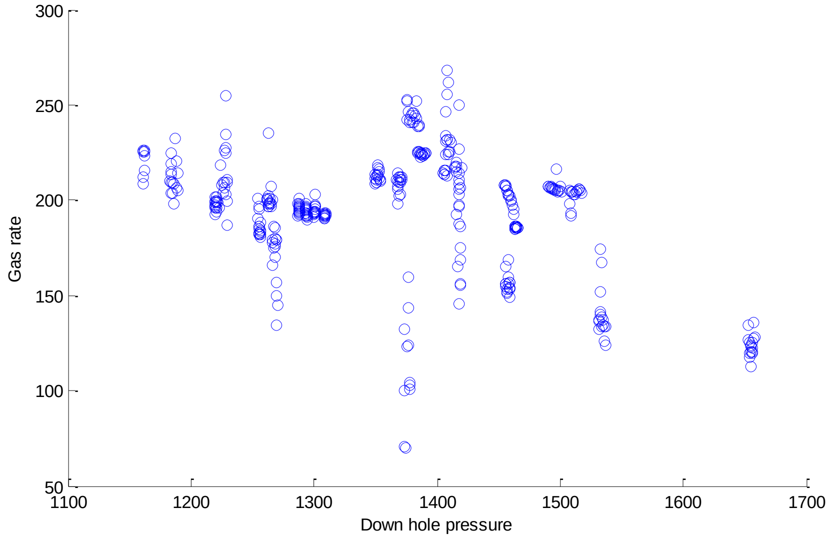

Figure 5.12: Gas rate against downhole pressure

As scatter plot in figure 5.12 shows, we can say that the gas rate has indirect relation with downhole pressure even though we can see some increasing in gas rate by increasing the pressure but the general overview shows decreasing in gas rate by increasing down hole pressure. However, we are not rely on this relationship too much because of the some increasing in the rate and unclear relation.

5.1.2.2. Downhole Temperature

As same as downhole pressure, downhole temperature plays an important role in the well operations such as drilling and production. The downhole temperature is used to validate and monitor the performance of the well and help them to understand what is happening downhole. Indeed, since the 1930s, engineers have used wellbore-temperature data for calculating flow contributions, evaluating water-injection profiles, diagnosing the effectiveness of fracture jobs, finding cements top behind casing and identifying cross flow between zones.[17]. The sensor data used to be gathered from a single point of wellbore but development in the sensor tools and using fiber-optic and digital sensors allows the engineers to collect the temperature in different points of the wellbore.[16][17]

Same as downhole pressure, the data for downhole temperature is available from beginning of 2011 until end of 2013. During this period millions of records of downhole temperature have been collected by downhole gauges for the this well. For this variable same as downhole pressure we calculate the average of the downhole temperature per day for the period of study to have an overview of the changes during this period. In the next step, we use a run-sequence plot to summarize the data in a graphical view.

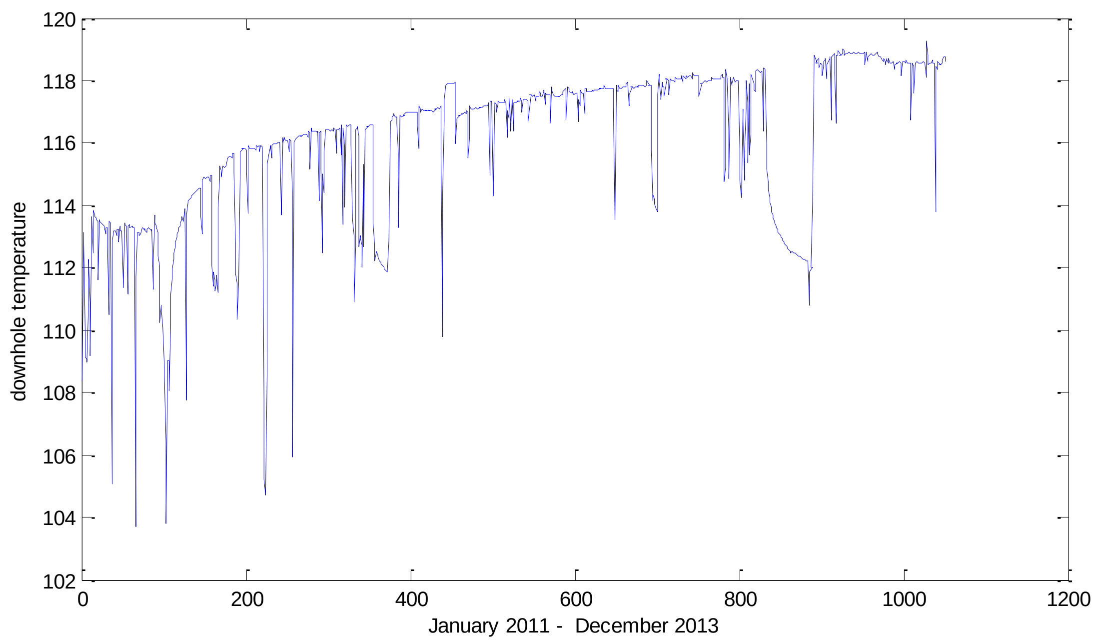

Figure 5.13: Downhole temperature

According to the Figure 5.13, we can claim that the temperature has steadily increased during the past few years. From 113 F in 2011, the temperature has increased to 118F at the end of 2013. Moreover, there are some sudden changes in the temperature, mostly decreasing; by analyzing the changes, we realized that these changes mostly have happened when the well has been shut in for few hours to few days or more. For example, as you can see in the graph(points 780-880) the temperature has decreased slowly from 117F to 112F for a period of 100 days. Analyzing the other data such as well status shows that the well has been shut in in this period for about three months. Another issue is that we can clearly see that the domain of the changes in temperature is very limited. Five degrees during almost three years and even in sudden changes are not more than ten degrees.

In the next few sections, first we do a statistical analysis on the downhole temperature to get more information about this variable and then we investigate the possible relationship between this variable and the production rates.

5.1.2.2.1. Downhole Temperature and Production Rates

To understand the relation between the downhole temperature and the production rates, as same as downhole pressure we calculate the average of the downhole temperature for 30 minutes during the well test periods. We also calculated the mean of the values and the variance to see the changes during this period (table 5.2).

Mean	117.5293
Variance	1.1865

Table 5.2: Downhole temperature statistics

The small variance of 1.1865 shows that the temperatures tend to be very close to the mean value and hence to each other.

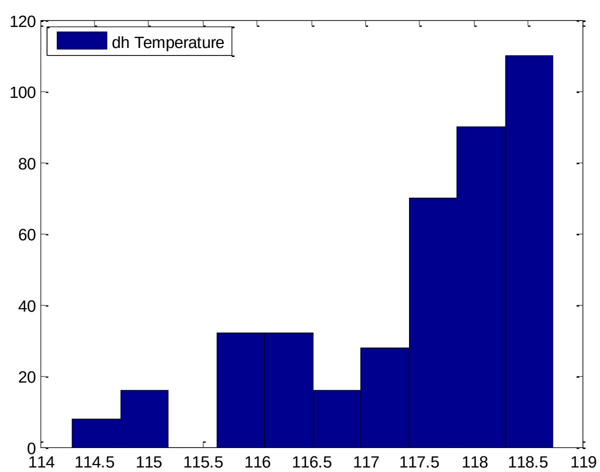

Figure 5.14: Downhole temperature (Histogram)

A histogram of the downhole temperature can show a better distribution of the data. The histogram in Figure 5.17 shows that about 70% of the data are between 117.5 and 118.5. It means just about one degree changes in the downhole temperature during almost 3 years for more than 70% of the period. Even though the changes in the temperature are small but it is possible that these small changes have a big effect on the flow contributions or flow rate.

5.1.2.2.2. Oil, Water, and Liquid Rates

A scatter graph can reveal the relation between downhole temperature and the production rates. Figure 5.15, shows that the rates of the oil and water and consequently liquid are increasing by increasing the temperature. The relation, specially between the water rate and the temperature is almost a curve. We can conclude that the water rate increases more and more by temperature while for the oil rate the relation seems not very clear or atlest we can say it has a positive slope but not very steep. The liquid rate against temperature is more clear and shows the direct relation between downhole temperature and the liquid rate and we can see a sharper curve between liquid rate and temperature.

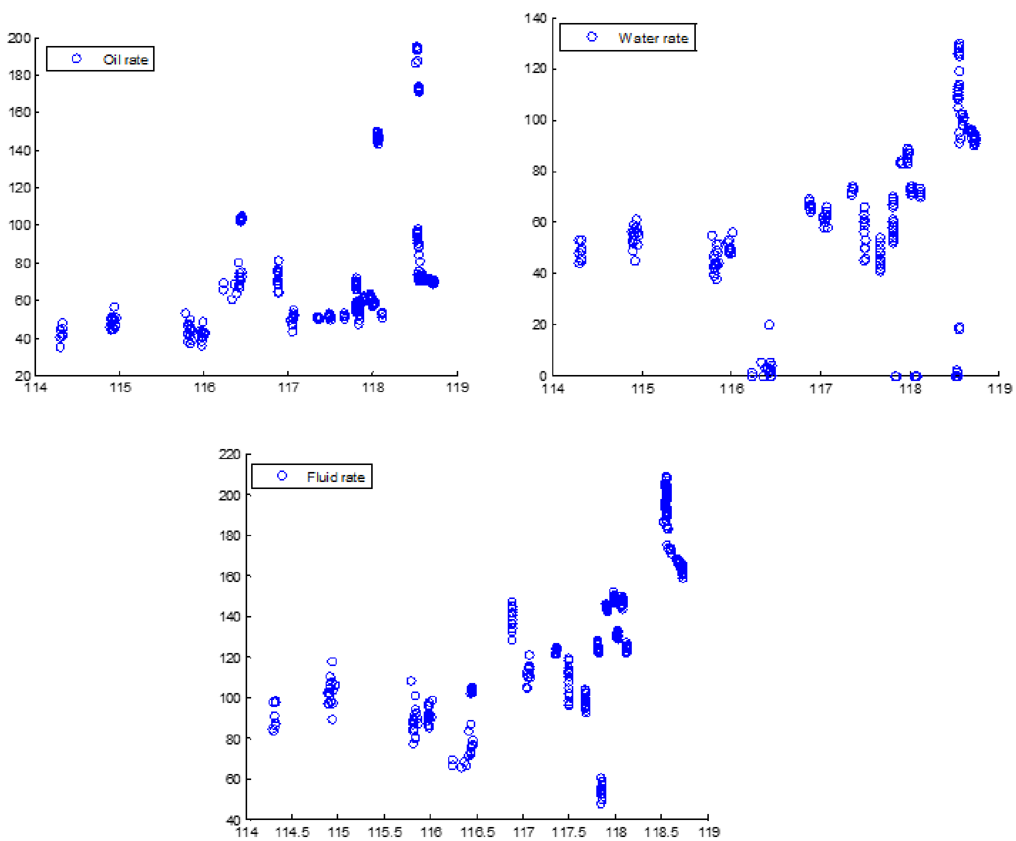

Figure 5.15: Oil, water, and liquid rates against downhole temperature

5.1.2.2.3. Gas Rate

The relation between gas rate and downhole temperature has been shown in figure 5.16. There is no clear relation between these two variables even though there are some decreasing in the gas rate by increasing the temperature or at least we can say that the gas rate has not increased by increasing temperature in down hole.

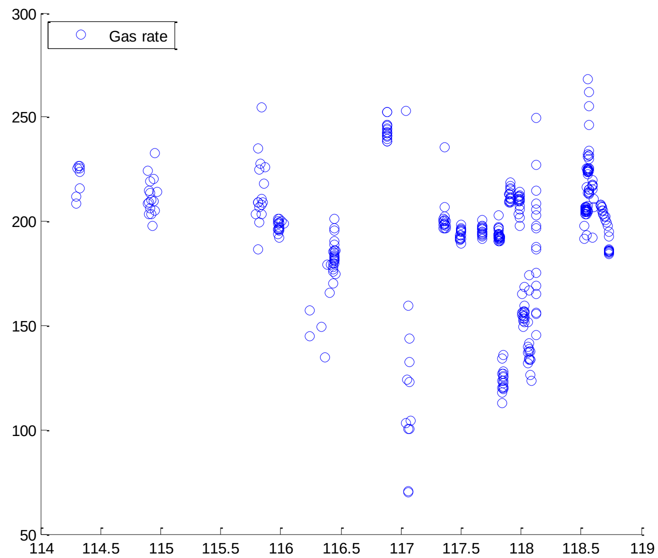

Figure 5.16: Gas rate against downhole temperature

At the end, we can conclude that even though there is a relationship between the downhole temperature and the water rate but at least in this case we cannot see any relation between this variable and oil and gas rates. Another issue about this variable is that this variable has not changed too much during the three years of the study. In other hands, we should mention that the experts in this fields claim that the downhole temperature is a good indicator to study the composition of the fluid. Finally, we can say that there is a possibility that we exclude this variable from the model. We will discuss this issue in the chapter 6.

5.1.2.3. Wellhead Pressure

So far, we have been discussing downhole variables that could have a relationship with the well production rates. On the surface of the well there are some equipment such as wellhead, Christmas tree and Choke that are installed during the well completion. A wellhead is the component at the surface of oil or gas well that provides the structural and pressure-containing interface for the drilling and production equipment[45]. Wellhead has different functionality during drilling or production or even when the well shut in temporary. Wellhead is used as pressure termination and when the wellbore is completed a Christmas tree is connected to the wellhead. The Christmas tree is used to control the pressure inside the well and to control the flow of the fluids from the well as well as an entry point to inject the water or gas into the well. [45][46][47]

The wellhead pressure has a very important role in the fluid flow control and production. The pressure of the wellhead is regulated by Christmas tree valves and Choke. Sensors at the wellhead as same as downhole, logs millions of records of pressure at the wellhead. To study this variable first we calculate the average pressure rate per day for the study period to have an overview of the wellhead pressure. A simple run sequence plot can give us an overview of the this variable during these period(Figure 5.17).

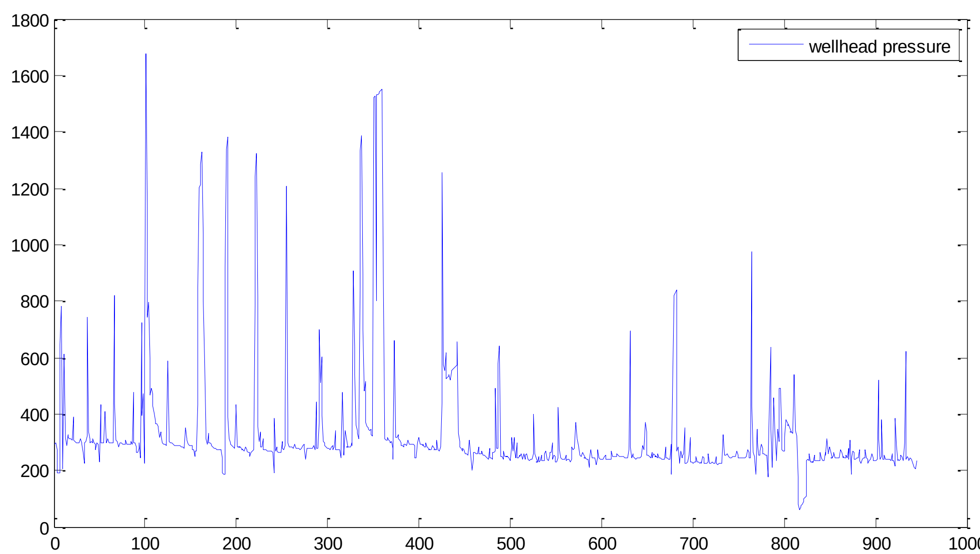

Figure 5.17: Wellhead pressure (01.01.2011- 31.12.213)

As the plot shows there are some sudden changes in the wellhead pressure, studying the days with sudden changes shows that these increasing in pressure happens due to changes in the well flow status specially when the well get shut in even for few hours. By shutting in the well for few hours, the wellhead pressure starts to increase rapidly and by opening the well it start to decrease again. The graph also shows that the sudden changes are mostly increasing in the wellhead pressure, we rarely see any sudden decreasing, and the main reason is shutting in the well in that day even for few hours.

Figure 5.20 also shows that the graph has a steadily gentle negative slope that shows the wellhead pressure has decreased slightly during this period.

5.1.2.3.1. Wellhead Pressure and Production Rates

To understand the relation between wellhead pressure and the production rates, we need to compare wellhead pressure and production rates at the same time. Therefore, we calculated the average of wellhead pressure per 30 minutes during the well test periods then by using a scatter chart we can study the relation between wellhead pressure and the rates.

5.1.2.3.2. Oil, Water, and Liquid Rates

Figure 5.18, shows that the oil and water rates and consequently the liquid rate are decreasing by increasing wellhead pressure. The oil rate descreasing very slowly while the water rate decreased faster and then more stable.The liquid rate against temperature is more clear as usuall and shows the indirect relation between wellhead pressure and the liquid rate.

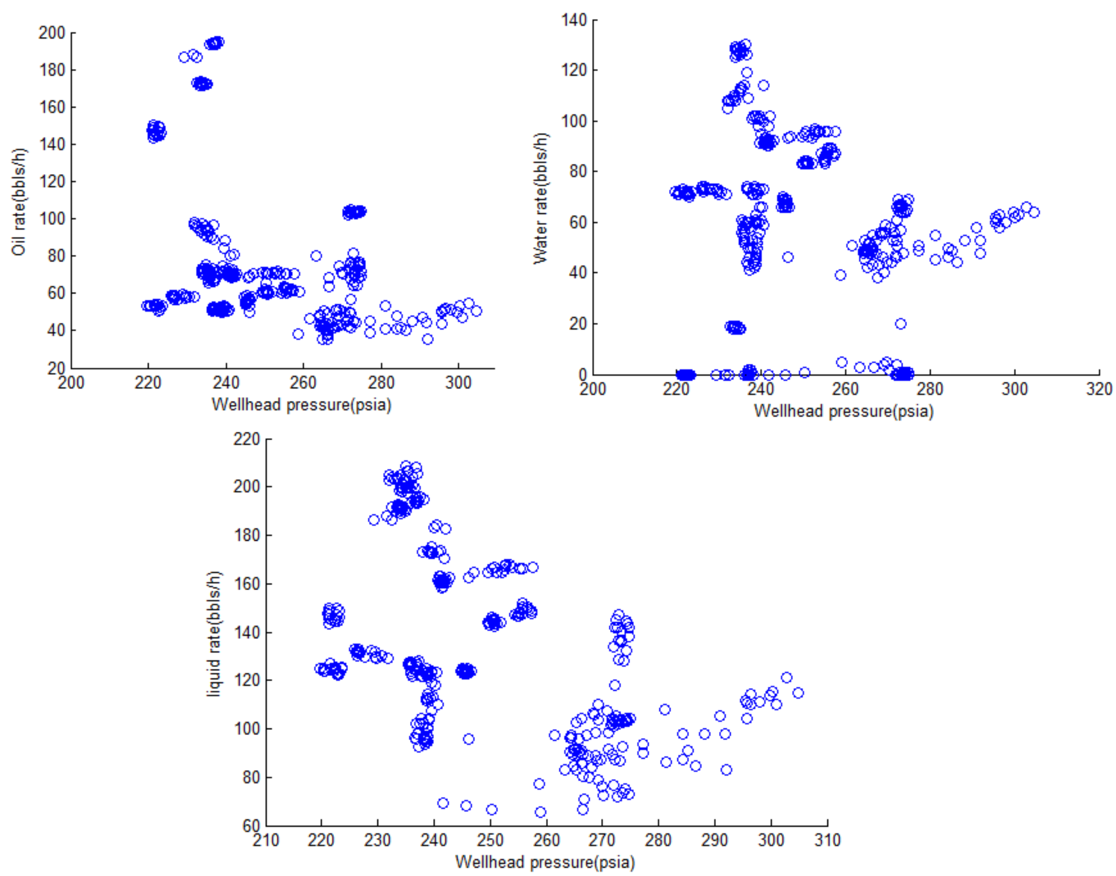

Figure 5.18: Oil, water, liquid rates against wellhead pressure

5.1.2.3.3. Gas Rate

Figure 5.19, the scatter plot between the gas rate and the wellhead pressure, shows that there is no relationship between the gas rate and the wellhead pressure. At the left side of the plot, we can see some increasing in the gas rate but then while the wellhead pressure has increased from 230 psi to 280 psi the gas rate has almost been constant. There are some decreasing in the 290 psi to 310 psi but unfortunately, we do not have enough example in this pressure to be sure about the changes.

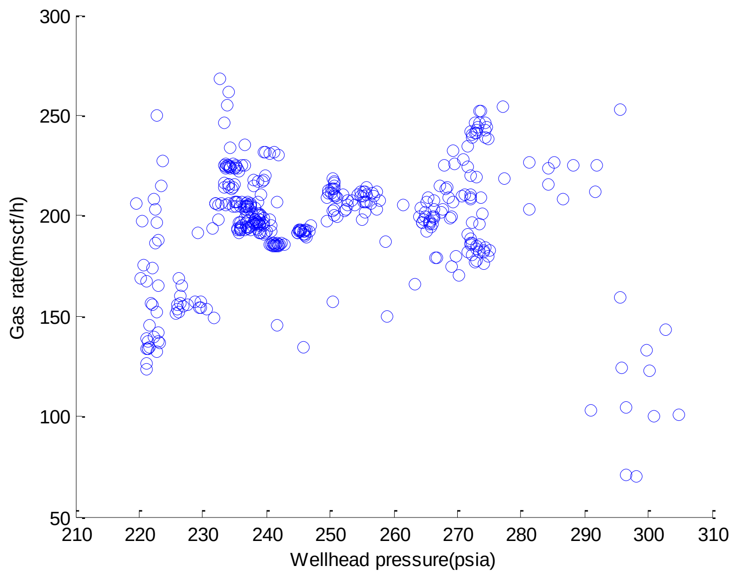

Figure 5.19 Gas rate against wellhead pressure

Finally, according to the analysis we have done in this section, we can conclude that there is a direct relationship between the wellhead pressure and the liquid rates while its relationship with the gas rate is not clear. Therefore, the wellhead pressure could be an important variable in the model.

5.1.2.4. Gas Lift

Gas lift is a method of artificial lift[1] that uses an external source of high-pressure gas for supplementing formation gas to lift the well fluids[2]. Gas lift is used in the well with insufficient reservoir pressure, in the other hand as we can see in our production rates for the current well, water that are much heavier than gas and heavier than oil encroach more and more into production by passing time and eventually makes the well to stop flowing. Because of heavier density in water-producing wells, sometimes even from the very beginning of the production, gas lift is applied to the well. Injected gas reduces the density of the fluids subsequently the weight of the hydrostatic column is reduced that lead to lower backpressure that allows the reservoir to push more fluids to the surface. [18][48]

In the gas lift operation, choosing the correct injection gas pressure is critical. Different factors can affect the choice but there is one critical factor that must be considered and any compromise with this factor can lead to a less efficient gas injection. The injection gas pressure at injection depth must be greater than flowing pressure at that depth. The rate of the injected gas is another factor that must be chosen correctly. Injecting gas only reduced the fluid density above the depth of gas injection so how much gas at which depth must be injected is important. A smaller volume at lower depth is more efficient than a higher volume at the upper part of the fluid column.[18]

5.1.2.4.1. Gas Lift Pressure

Unfortunately, gas lift pressure data is not available for some periods of the study and we have to exclude this variable from the model.

---

1. Artificial lift is a method used to lower the producing bottom-hole pressure on the formation to obtain a higher production rate from the well (Petorwiki).
2. Wikipedia

5.1.2.4.2. Gas Lift Rate

The well that we are studying is a water producing well and the engineers inject gas into the well continuously during the production. The flow rate of the injected gas is standard cubic feet per hour (SCF/H). We use a sequence run chart to summarize the average gas rate per day during the study period same as other variables (Figure 5.20).

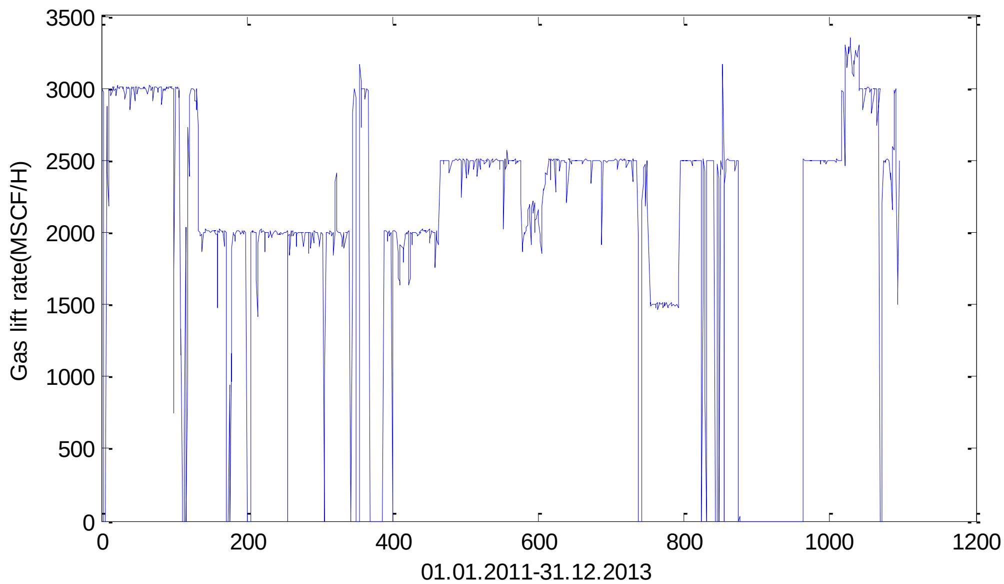

Figure 5.20: Gas lift rate

5.1.2.4.3. Gas Lift and Production Rates

The main goal of gas lift operation and injecting gas into the well is increasing the production rates and in some wells, it is a mandatory operation that must be carried out to produce oil and gas. Therefore, the relation between gas lift and production rates are clear. The Figure 5.24 shows the scatter plots between oil, water and gas against gas lift.

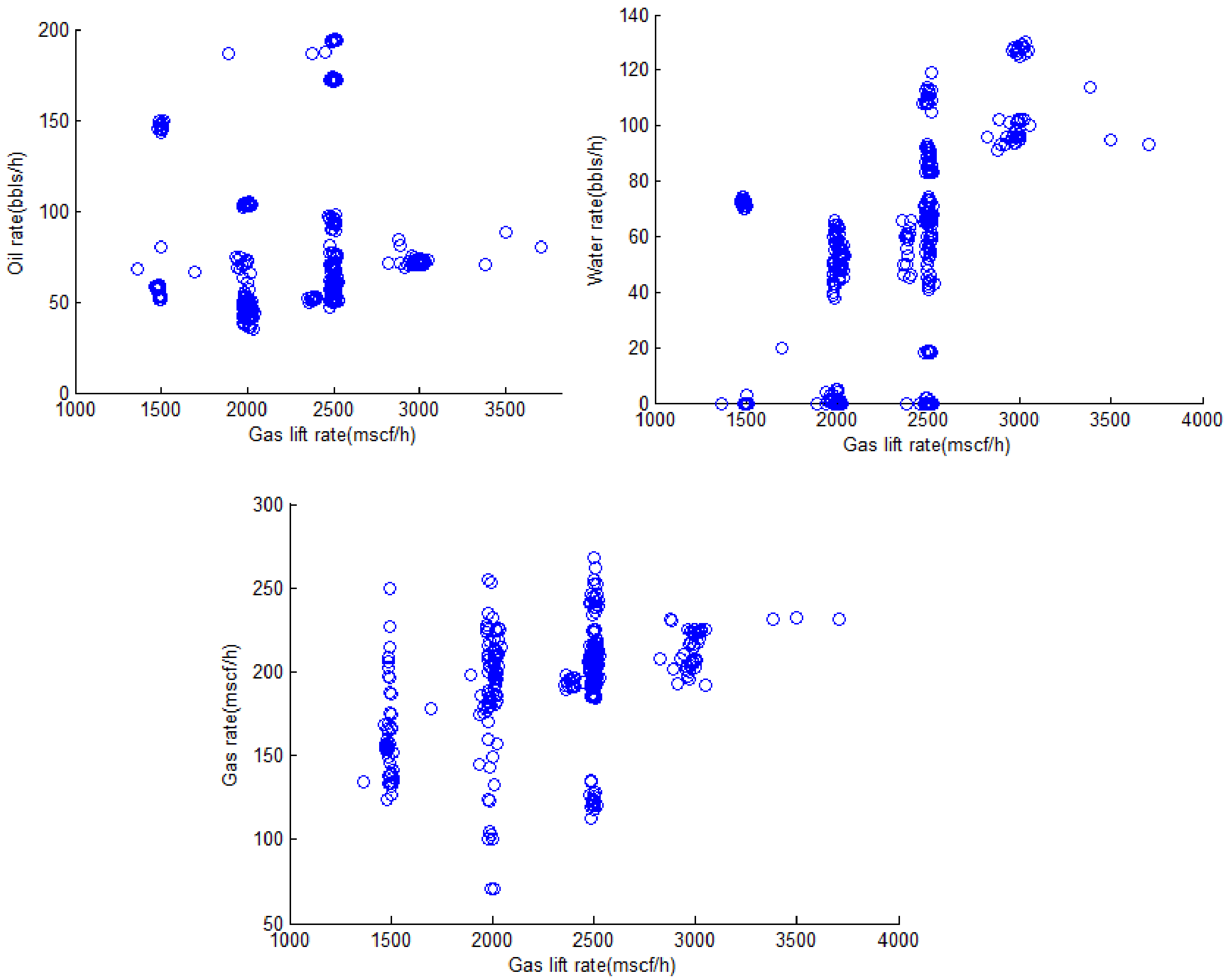

Figure 5.21: Gas lift rate and production rates (oil, water and gas)

Figure 5.24 as we expected shows that there is a direct relation between the gas lift and the production rates. The production rates are increasing by increasing in the gas lift injection. Changes in the production rates while the gas lift rate is constant shows the other parameters such as downhole pressure or wellhead pressure may have been changed in these periods.

The analysis shows that this variable also could have an important role in the model same as downhole pressure and wellhead pressure. Especially to predict the gas rate as you can see so far this variable is the only variable with a clear relation with the gas rate.

5.1.2.5. Choke

Choke valve is a type of control valves, mostly used in oil and gas production wells to control the flow of the well fluids being produced. There are different types of chokes. We can divide them into two general categories: Fixed choke and adjustable choke. Fixed chokes are used to kill the reservoir pressure when the value is open completely and adjustable chokes are used regulate the pressure and adjust the production level of the well. [19] [20][21]

We are studying adjustable chokes as they are using to adjust the production rates. The choke opening percentage and choke pressure are two variables we could work on. We couldn't collect enough data of the choke pressure because the data mostly were not available therefore we ignore this variable. We studied choke opening percentage that is used to adjust pressure and consequently production rates.

The choke has a close relation with wellhead pressure. In order to study this relation we start with a scatter plot of wellhead pressure against choke opening percentage. Figure 5.22 shows the result. The plot shows an indirect relation between wellhead pressure and the opening percentage of the choke. As you can see the data has a negative slope. We can say that these two variables could be redundant in the model therefore there is a possibility to remove the choke opening percentage from the model. We will study this possibility in the chapter 6.

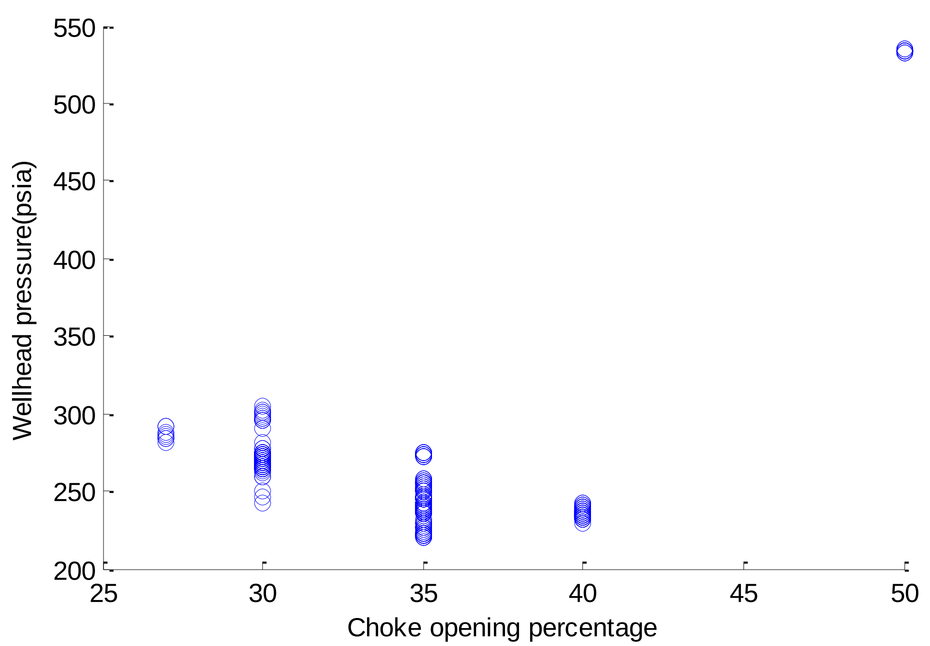

Figure 5.22: Wellhead pressure against choke opening percentage

As we expected and figure 5.25 shows, the choke opening percentage has a direct relation with the all production rates.

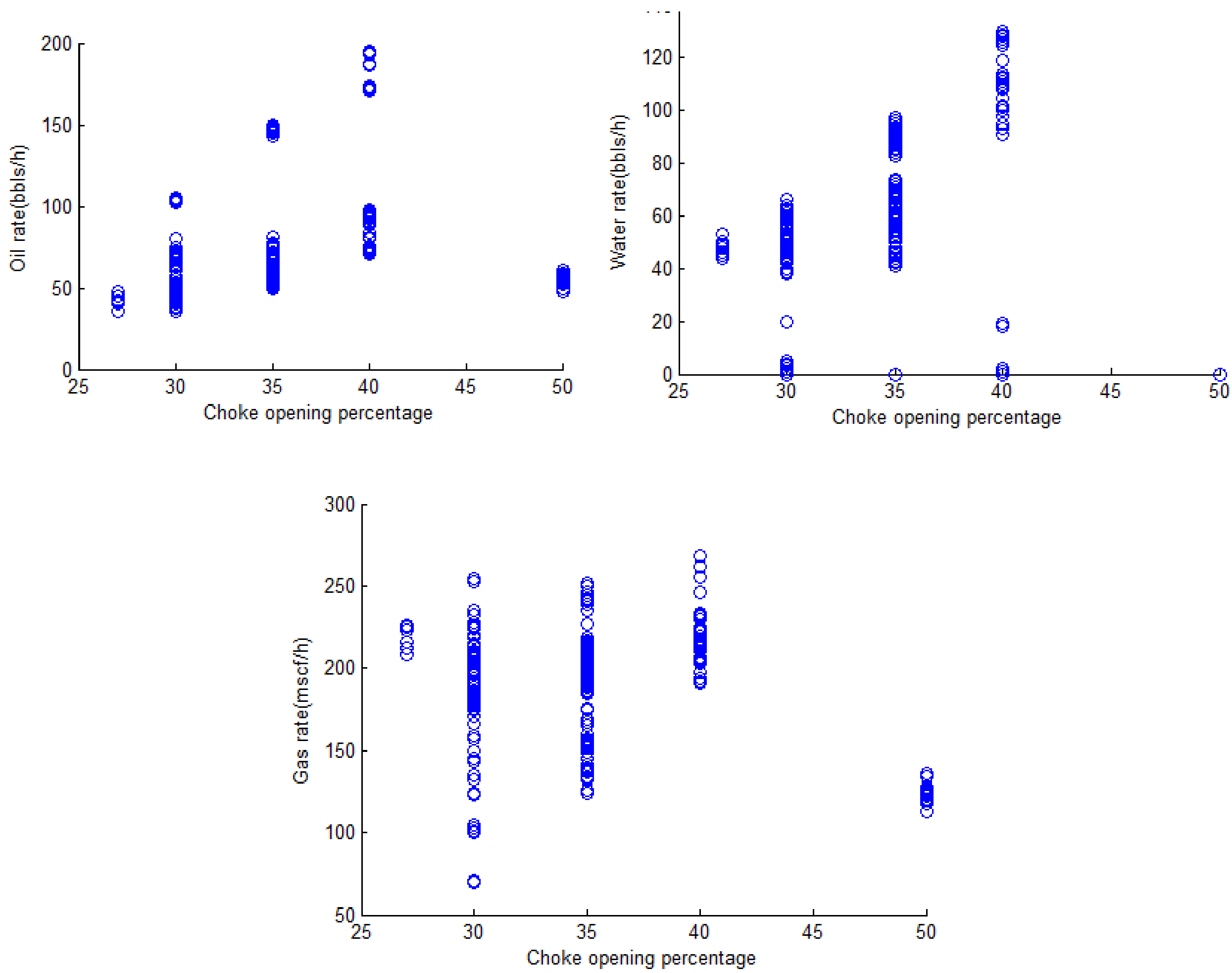

Figure 5.23: Production rates against choke opening percentage

Another variable related to the choke is *choke pressure*. Unfortunately, there was not enough data for this variable and we had to exclude it from the model.

5.2. Design The Model

Analyzing the data enable us to understand the relation between the variables and identify the possible input variables of the model. This knowledge leads us to design a proper model to predict the outcomes of the model.

In this section, we design a model based on the analyzed data in the section 5.1. We start by the basic specification of the model. Different scenarios have been designed to experiment our hypothesis. These scenarios are different in details such as inputs, outputs and hidden nodes of the model.

5.2.1. The Basic Specification Of The Model

The model is a multilayer perceptron artificial neural network (ANN). Back-propagation algorithm is used to train the network.

Unfortunately, we could not collect large amount of data to train the model. The fact that production rates are just available during the well test periods, which carried out once a month, restricted us to these periods. In other hands, engineers have rejected some well test results for different reasons, and for some well test periods, the sensor data at downhole or wellhead were not available. In the end, a dataset of more than 400 rows is prepared for the model but in order to test the model with some data out of the training dataset we exclude one/two test period (16 rows) from the dataset. This period is used to compare different scenarios result. The rest of the prepared dataset is used to train, validate and test the network.

5.2.1.1. Input Layer

The inputs nodes of the model are chosen from five possible inputs:

- Downhole pressure
- Downhole temperature
- Wellhead pressure
- Gas lift rate
- Choke opening percentage

5.2.1.2. Output Layer

The outputs are chosen from four possible outputs:

- Oil rate
- Water rate
- Liquid rate
- Gas rate

5.2.1.3. Hidden Layer

To design the hidden layer there are two questions that must be answered:

First:

How many hidden layer do we need in our neural network?

Second:

How many neurons in each layer we need?

To answer the first question we should know that most neural networks would have only one hidden layer when having more than two layers is very rare. In fact, many practical problems can be solved by using just one hidden layer. [22][23]

Number of hidden layers	Result
None	Only capable of representing linear separable functions or decisions.
1	Can approximate any function that contains a continuous mapping from one finite space to another.
2	Can represent an arbitrary decision boundary to arbitrary accuracy with rational activation functions and can approximate any smooth mapping to any accuracy.

Table 5.3: Number of hidden layers (Source 22)

Based on what we described about the number of hidden layers we have decided to choose one layer for our neural network.

Determining the number of hidden layer is almost an easy decision as most of the problems can just use one layer. Determining the number of neurons in this layer is a more important question that must be answered during designing a neural network architecture.

There are many rule-of-thumb to specify the number of nodes in the hidden layers, the most common rule is that the number of nodes in the hidden layer is the mean of the input and output nodes.

There is another general rule to calculate the number of nodes in the hidden layer:

Number of hidden neurons = (Number of inputs + outputs nodes) * 2/3

This is another rule-of-thumb says the number of hidden neurons must be less than twice of the neurons in the input layer.[22]

All these rules are empirically-derived rules-of-thumbs. To choose the correct number of neurons in the hidden layer we use these rules as a starting point and then by trial and error determine the number of the nodes.

5.2.2. The Modeling Tools

To model our network we used Neural Network Toolbox in Matlab. This toolbox has the flexibility to choose an inputs, outputs and hidden neurons. There is the possibility to easily change the hidden neurons and test the result. In other hands the there are some different methods to divided the dataset such as random or by block. You can easily change the percentage of training, validation and test subsets. There are different tools such as performance plot, regression plot and histogram to analysis the performance of the network.

6. Experimental Results and Discussion

In this chapter, we will design different scenarios to test the model and will present the results of our experiments with brief explanations.

6.1. Scenario A

This scenario considers liquid rate as output, and downhole pressure, downhole temperature, wellhead pressure, choke opening percentage, and gas lift rate as inputs of the network.

The logic behind this scenario is that while oil and water rates may change suddenly, the liquid rate is more stable and predictable. In the other hand, we have a small dataset to train the model. Having a model that could predict the liquid rate could still be a great achievement while the ideal prediction would be oil, water and gas prediction.

6.1.1. Hidden Nodes

As we described in the section 4.2.1.3, we use current rules-of-thumb as start point and do our trial and error experiment to determine optimal hidden node for our neural network.

In this experiment, we start from one node less than the mean of the inputs and outputs nodes.

Mean = 5 +1 /2 = 3.

Start point = 2 nodes.

For each node, we trained the model 3 times and registered the regression between the target and the output of the model for training, validation and testing results. Finally, we calculated the averages (table 6.1).

Nodes	Training	Validation	Testing
2	0,973667	0,963333	0,968667
3	0,962667	0,966333	0,967667
4	0,974333	0,972667	0,968667
5	**0,99**	**0,986533**	**0,983533**
6	0,9898	0,985833	0,987667
7	0,991333	0,988	0,986667
8	0,987767	0,981	0,982

Table 6.1: Hidden nodes regression results (Scenario A)

We illustrate the results in a graph to have a better understanding of the results. As the Figure 6.1 shows even though the result for 3 nodes is still acceptable but from 5 to 7 nodes we have a better regression (more than 0,985). As we have a small dataset to train the model we prefer to choose 5 nodes for the hidden layer to prevent overfitting[1] our neural network.

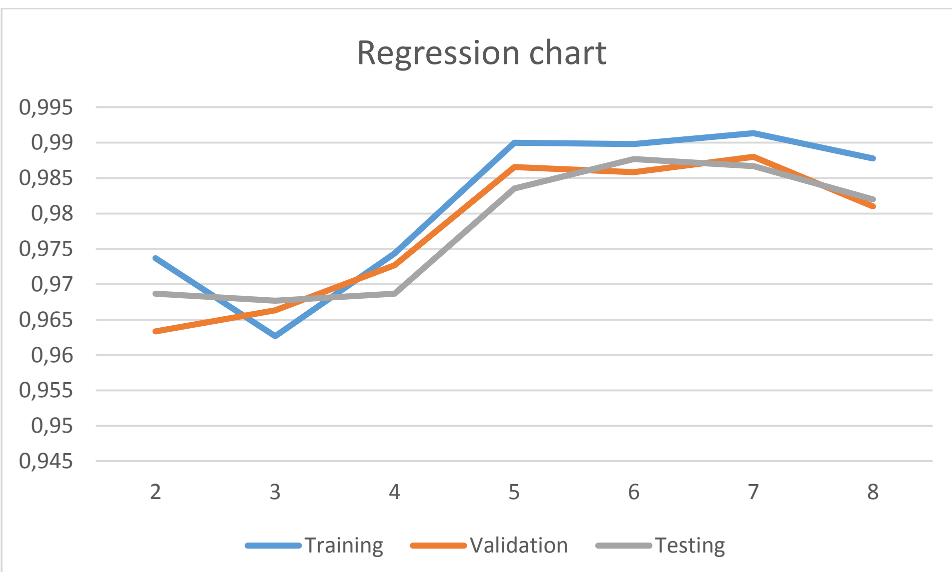

Figure 6.1: Regressions for different nodes in hidden layer

1. Overfitting occurs when the neural network has so much information processing capacity that the limited amount of information contained in the training set is not enough to train all of the neurons in the hidden layers.

6.1.2. The Network Performance

After the specifying the inputs, outputs and the hidden neurons for the neural network, the next step is training the model.

70 percent of our dataset is used to train the model while 20 percent is used to validate the training and 10 percent to test the model.

Figure 6.2: Network performance (Scenario A)

Figure 6.2 shows the performance plot of the training. The validation performance reached the minimum at iteration 140. The training continued to iteration 148 before it stops. The figure does not show any major problems during the training.

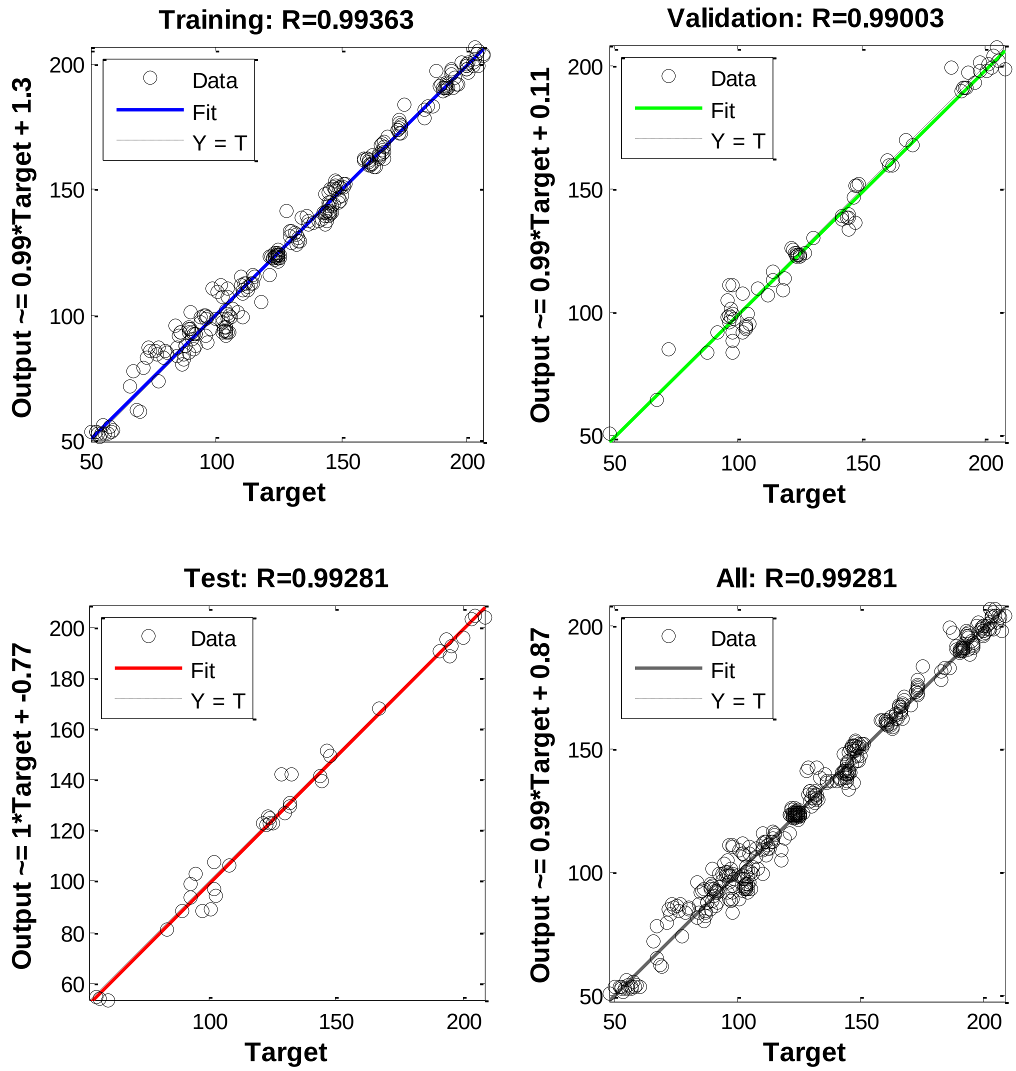

Figure 6.3: Regression plots (scenario A)

Figure 6.3 shows the regression plots, which shows the relation between the outputs of the model, generated values by the neural network, and the targets, the liquid rates from the dataset. If the training was perfect, the outputs of the network and the target would be the same, but in practice, this rarely happens.

In each plot, the dashed line represents the perfect result – outputs = targets and the solid line represents the best fit linear regression line between outputs and targets. The relation between the outputs of the model and the targets are represented by R value. R = 1 represents an exact linear regression between the outputs and the targets. For R close to zero, there is no linear relation between outputs of the model and the targets.

For all the subsets, R-values are greater than 0.99 that indicates a good fit of the data.

Error histogram is another plot that help us to get more information about the network. This plot shows the distribution of the errors = Targets – Outputs. As you can see the distribution of the training, validation and test subsets are acceptable as the error distribution shows. The plot also gives us a good indicator of outliers. While most errors fall between -6 and 6, we have some points with error more than 10. Because of the nature of our data having this kind of the error distribution will not satisfy us.

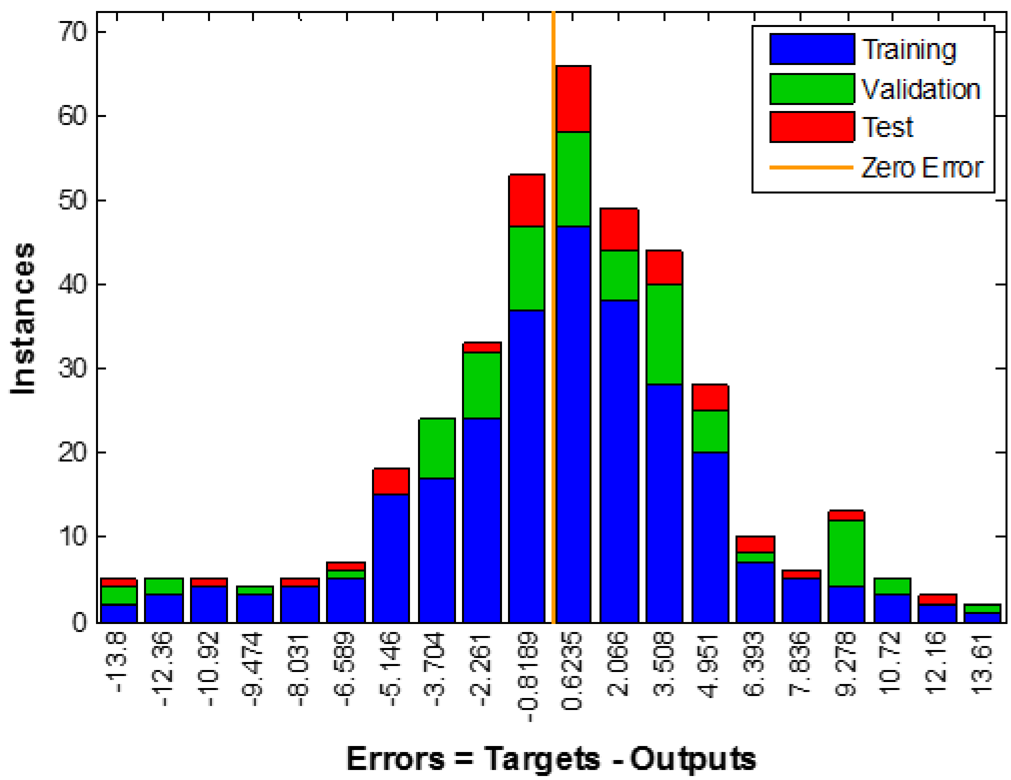

Figure 6.4: Error histogram (Scenario A)

6.1.3. Test The Network

The regression plots and error histogram illustrate the network behavior with training data. But the best indicator to see if the network performance is good enough or not is testing the network with more data outside of the training dataset. For this purpose, we exclude one Well Test period from the prepared dataset. In this section, we test the network with this data and evaluate the results:

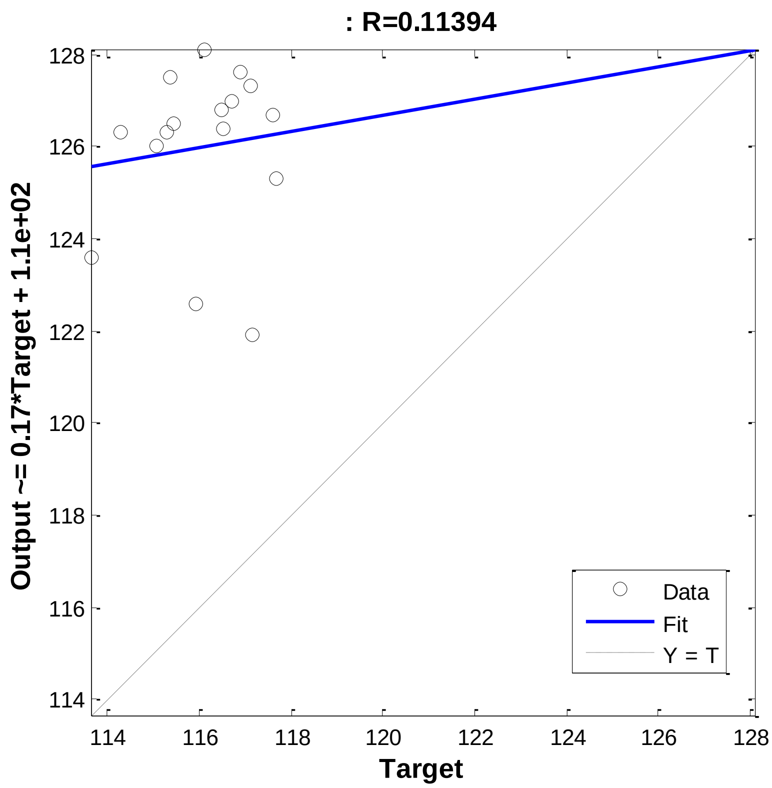

Figure 6.5: Regression plot for extra test (Scenario A)

Unfortunately, the results are disappointing, not only the outputs and the targets are far away; even the R value is about 0.11 that shows almost no relation between the outputs of the network and the targets.

6.2. Scenario B

This scenario considers all production rates as outputs, and downhole pressure and temperature, wellhead pressure, choke opening percentage, and gas lift rate as inputs of the network.

The logic behind this scenario is that, the best network for us is a model that predict all production rates.

6.2.1. Hidden Nodes

Same as Scenario A, First we do some trial and error experiment to determine the hidden neurons for the hidden layer. We start or experiment by one node less than the mean of the inputs and the outputs of the network.

Mean = (5 +3) /2 = 4

Start point = 3 nodes.

For each node, we trained the model 3 times and registered the regression between the target and the output of the model for training, validation and testing results. Finally, we calculated the averages (table 6.2).

Nodes	**Training**	**Validation**	**Testing**
3	0,9597	0,96265	0,94495
4	0,96805	0,9643	0,9722
5	0,97795	0,9704	0,9859
6	0,9837	0,979533	0,983433
7	0,983933	0,980833	0,977667
8	0,988067	0,975767	0,986767
9	0,989	0,987	0,987333
10	**0,991333**	**0,991333**	**0,991667**
11	0,991	0,979	0,981667
12	0,986667	0,987667	0,985

Table 6.2: Hidden nodes regression results (Scenario B)

Figure 6.6 shows by having more than 5 nodes the R values for all subsets are more than 0, 97 and 10 nodes has the best R values for all subsets.

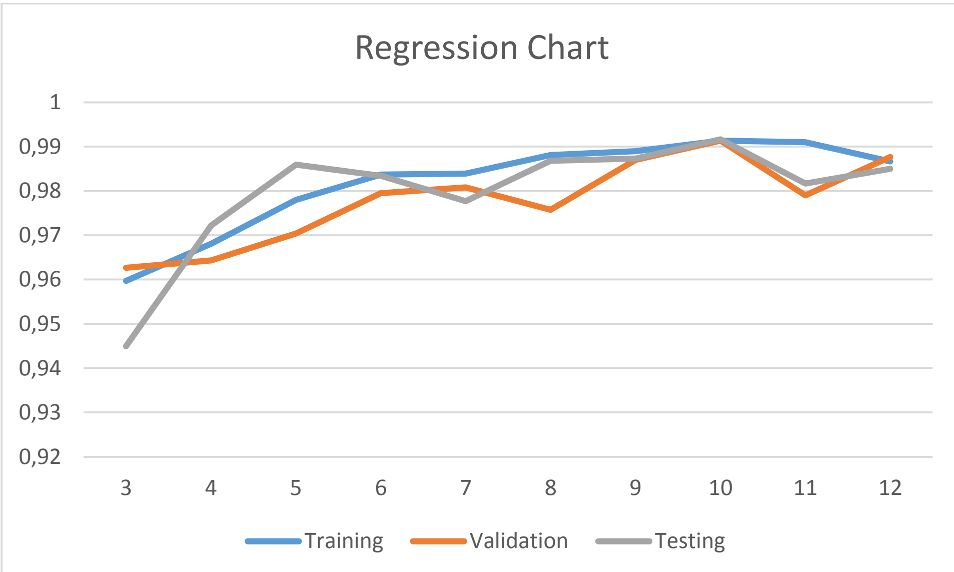

Figure 6.6: Regressions & number of nodes in hidden layer (Scenario B)

Based on our experiment, for this scenario we chose 10 nodes for the hidden layer. Same as scenario A, 70 percent of our dataset is used for training, 20 percent for validation and the final 10 percent is used to test the model.

6.2.2. Network Performance

Figure 6.7 shows the performance plot for this network. The validation performance reached a minimum at iteration 78 and the training stops at iteration 84. We cannot see any major problems with the training. The validation curve and the testing cure are almost close. If the test curve increased significantly, more than validation curve then it is possible some overfitting might have occurred. [24]

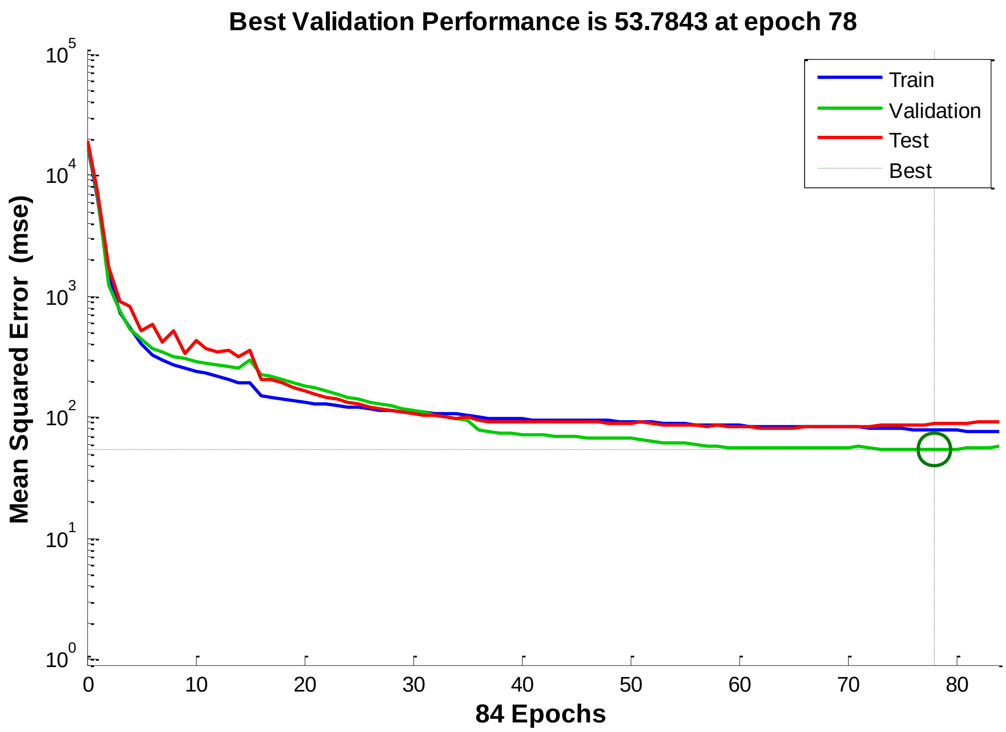

Figure 6.7: Network performance (Scenario B)

The next step to validate the network is to create the regression plots of network outputs with respect to targets for training, validation and testing subsets.

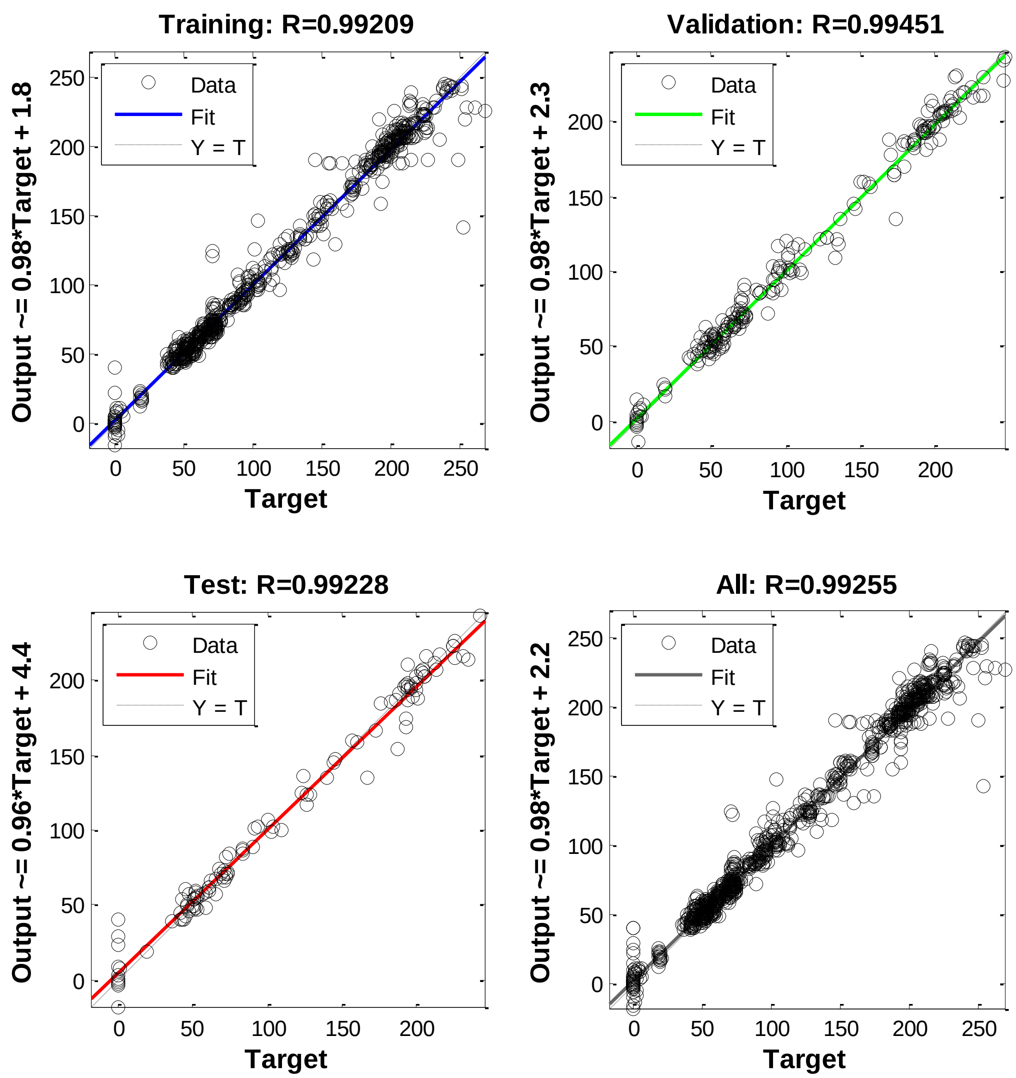

Figure 6.8: Regression plots (scenario B)

In this scenario, the fit is reasonably good for all subsets with R values above 0.992 that the outputs of the network track the targets very well for all subsets.

To obtain additional verification of the network performance we analysis the error histogram of the data. In this scenario, the histogram shows that while most errors fall between 1 and 2 we have some errors less than 20 but still there are some validation points with errors of 120 , -53,-44. These outliers are also visible in training regression plot. There are some points with outputs of 150 for target 200, outputs of 140 and 180 for the target of 250.

In this situation, either the data is bad or the data points are different from the rest of the data. In our case, the data has been chosen during a process of cleaning and verification so we conclude that the data points are different and the network is extrapolating for these points. In this case, we should collect more data close to outliers data and retrain the network again to gain a better results.

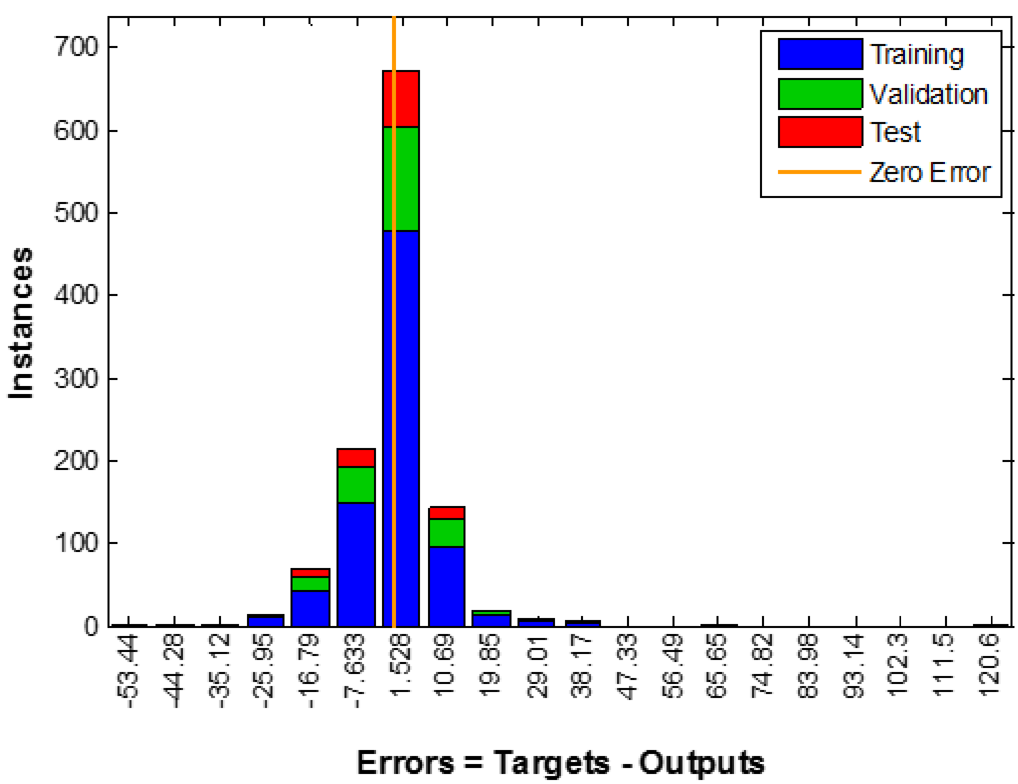

Figure 6.9: Error histogram (Scenario B)

6.2.3. Test The Network

The regression plots and error histogram illustrate the network behavior with training data. Nevertheless, the best indicator to see if the network performance is good enough or not is testing the network with more data outside of the training dataset. For this purpose, we exclude one Well Test period from our dataset. In this section, we test the network with this data and evaluate the result:

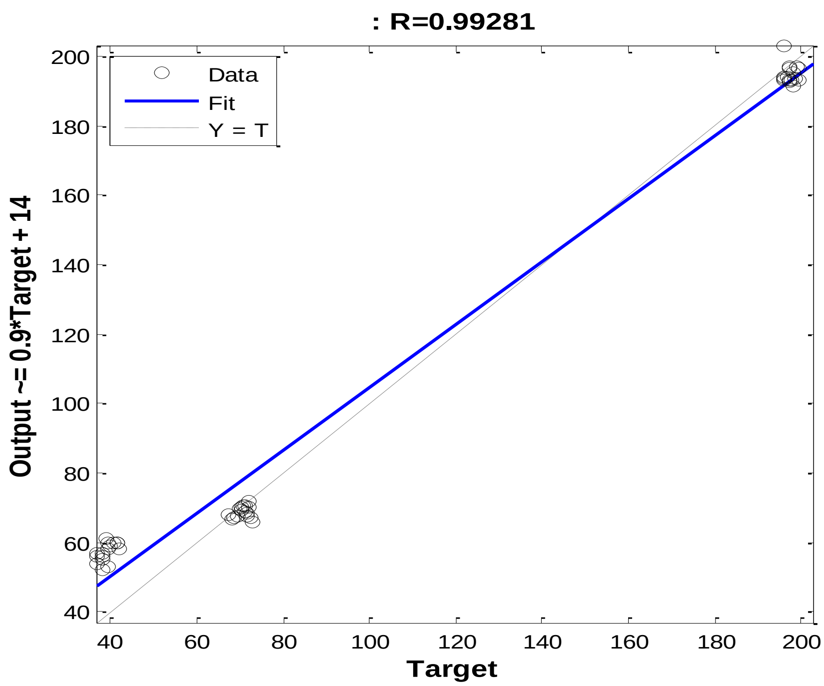

Figure 6.10: Regression plot for extra test (Scenario B)

Figure 6.10 shows the regression plot for outputs of the network with respect to the targets for the extra test. The fit line of the mode is very close to the actual line of the targets but they have slightly different slope. The outputs of the network are reasonably close to the targets. There are some points with targets about 40 while the outputs are between 50 and 60. Generally, we can claim that the results are acceptable with reasonable errors and very close to the targets.

6.3. Scenario C

As we discussed in section 5.2, the network for scenario B has a good performance and the result for the extra test was reasonably good. However, we still were not totally satisfy with the result as we mentioned in the previous section. We tried several different scenario, such as a scenario with outputs of fluid and gas rate, a scenario with outputs of oil and water and finally a novel scenario with all production rates, Oil, Water, and Gas rates plus Liquid rate as an extra output. The result for this scenario was interesting. We will discuss the network and the result in this section.

6.3.1. Hidden Nodes

Same as Scenario A, and B, we determined the number of the hidden nodes for the hidden layer. For more details, we refer to sections 5.1.1 and 5.2.1.

Nodes	Training	Validation	Testing
3	0,95685	0,95475	0,96055
4	0,9646	0,9639	0,95835
5	0,981367	0,982233	0,976433
6	0,986333	0,982	0,975367
7	0,989367	0,9887	0,986233
8	0,9864	0,9842	0,985467
9	0,988267	0,988067	0,9886
10	0,991733	0,9906	0,9916
11	**0,993767**	**0,986767**	**0,9918**
12	0,993767	0,990267	0,9863
13	0,9888	0,9814	0,9848

Table 6.3: Hidden nodes regression results (Scenario C)

Figure 6.11 shows having more than 6 nodes, the R values for all subsets are more than 0,98 and having 10 tor 12 nodes the network has the best R values for all subsets.

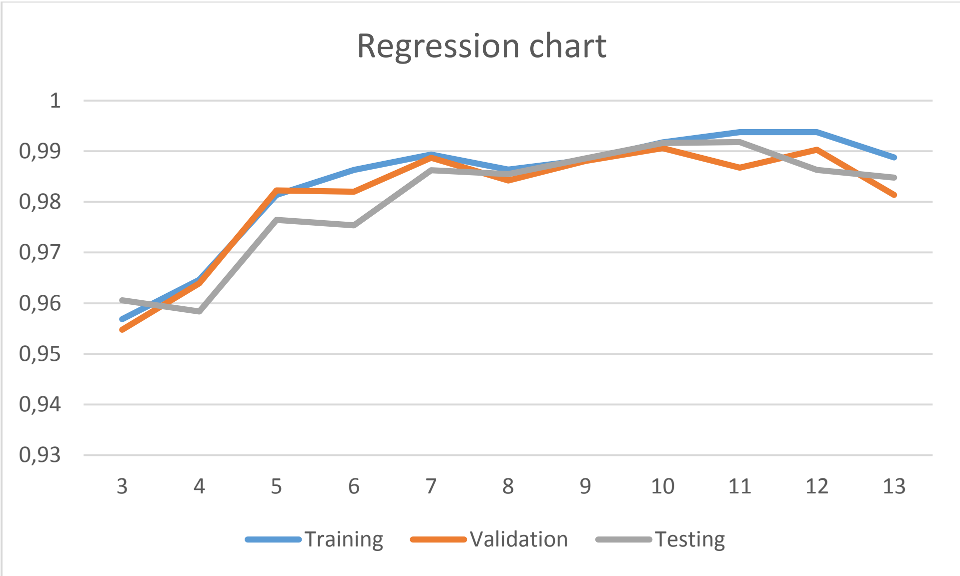

Figure 6.11: Regressions & number of nodes in hidden layer (Scenario C)

Based on this experiment, we chose 11 nodes for the hidden layer in this network.

6.3.2. Network Performance

Figure 6.12 shows the performance plot for the network. The validation performance reached a minimum at iteration 38 and the training continues until 44. There is no big problem in the training again but there is a small distance between validation and testing curves. The distance is not diverging so there is no overfitting in the network. The best validation performance is about 75, while it was 54 for scenario B. This shows a better performance for scenario B than the current scenario.

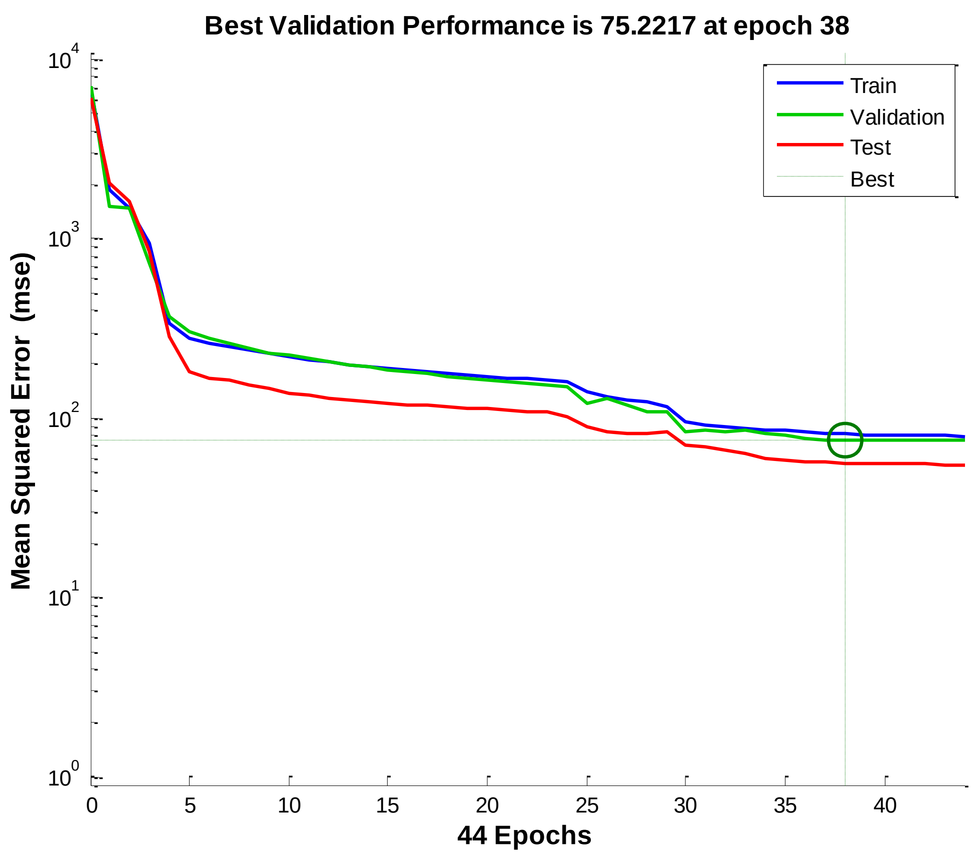

Figure 6.12: Performance plot (Scenario C)

Figure 6.13 shows the regression plots of the outputs of the network with respect to the targets for training, validation and test subsets.

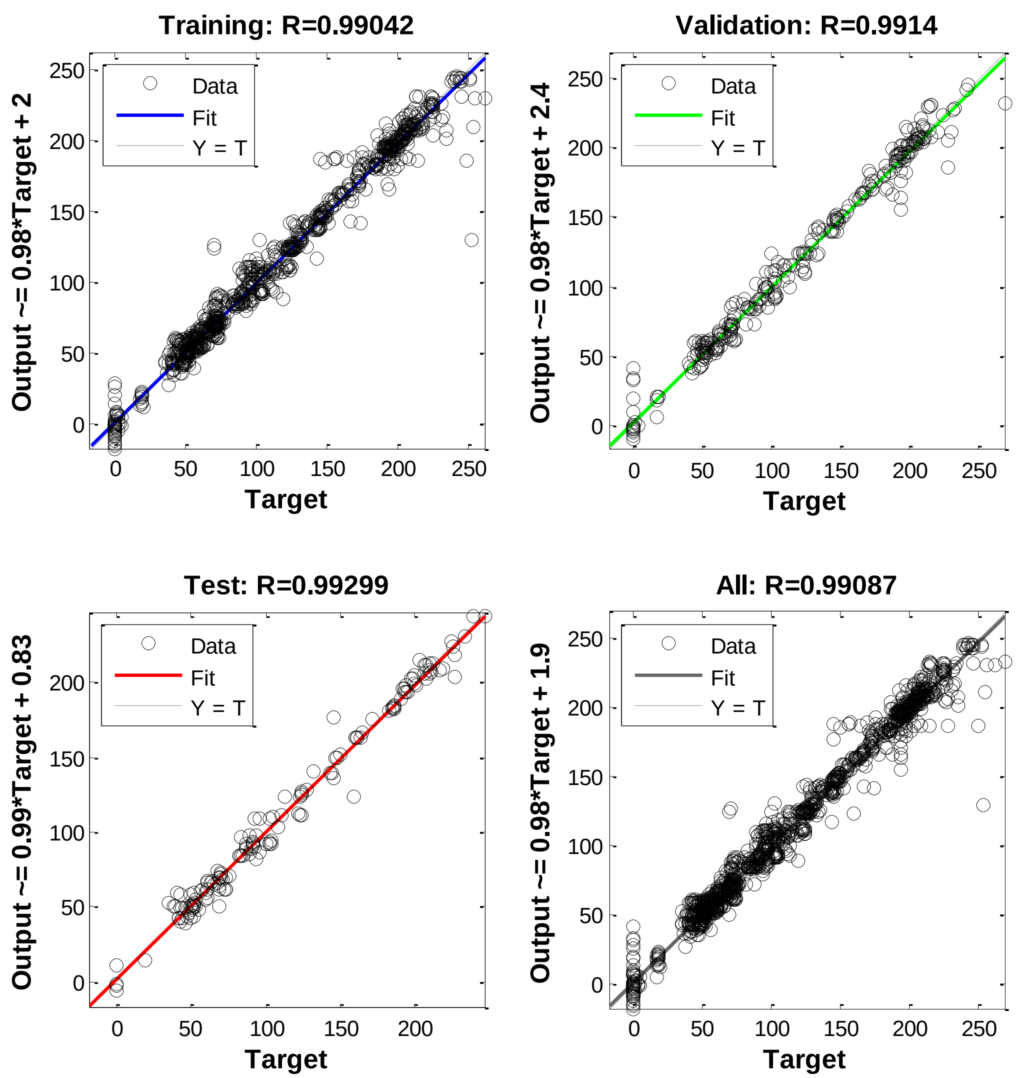

Figure 6.13: Regression plots (Scenario C)

Same as the scenario B, the fit is reasonably good for all subsets. The R values for this scenario are still more than 0.99 but they are slightly, about 0,001, smaller than R values in scenario B. It is noteworthy to mention that, having more data may reduce the regression values but may make the network more robust. In this scenario, we have increased the outputs about 25 percent by adding liquid rate to the outputs and we still have almost same R values.

Figure 6.14 shows the error histogram of the network for scenario C. As you can see, the pattern of distribution of the errors is similar to scenario B. The plot shows that this scenario has a better distribution, about 60% of points with 1.122 in this scenario compared to 60% in scenario B with 1.528 errors. This happens despite the 25% increasing in the outputs in this scenario. Same as scenario B we still have some validation points with errors of 118 , -53,-44. As we discussed for the scenario B. we should collect more data close to outliers data and retrain the network again to gain a better results.

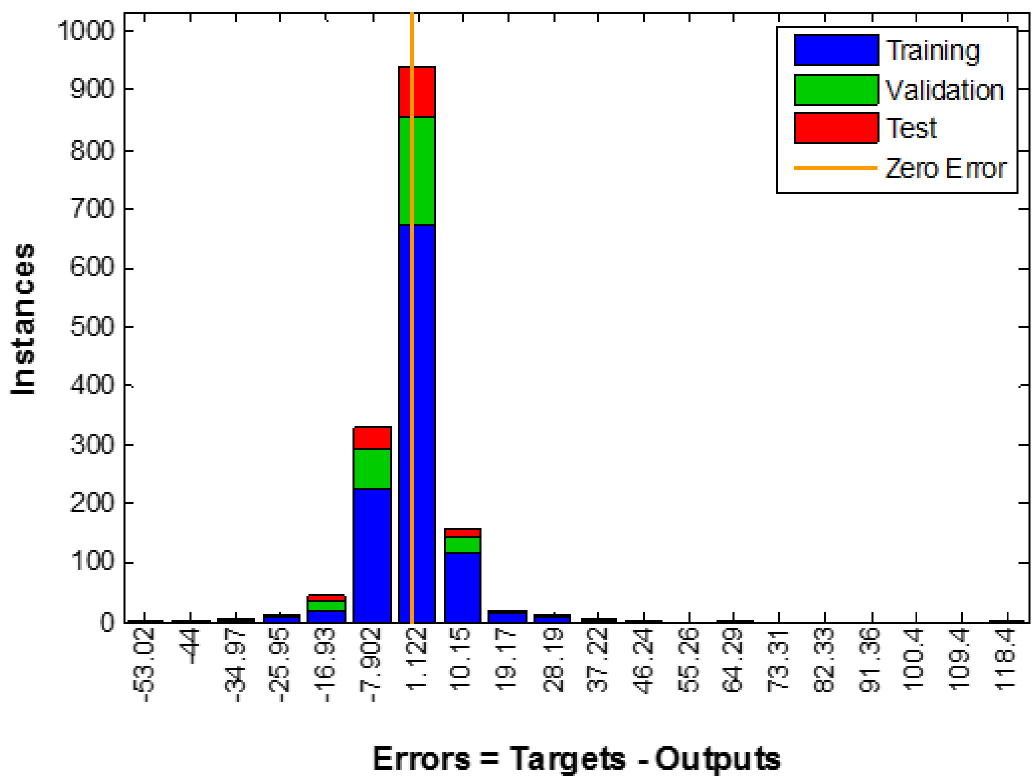

Figure 6.14 : Error histogram (Scenario C)

6.3.3. Test The Network

While R values and error distributions of both scenarios B and C are very similar and the differences are negligible a better indicator to see the difference between the networks could be testing the model with same data out of the training dataset.

Same as scenario B we tested the network with the same extra data we excluded from the dataset. Figure 6.15 shows the regression plot of the outputs of the network with respect to the targets. Comparing this regression plot with the same regression plot for network B in section X shows a better performance for this network. The slope of the fitting line is as same as the slope of the perfect line, 45 degree, and very close to the perfect line while in scenario B the fitting line has a different slope as we described in section X. The outputs of the network are closer to the targets compare to scenario B.

The formula of the output make the difference very clear:

Scenario B: Output = 0.9 * Target + 14 (Figure 6.10).

Scenario C: Output = 1* Target + 4

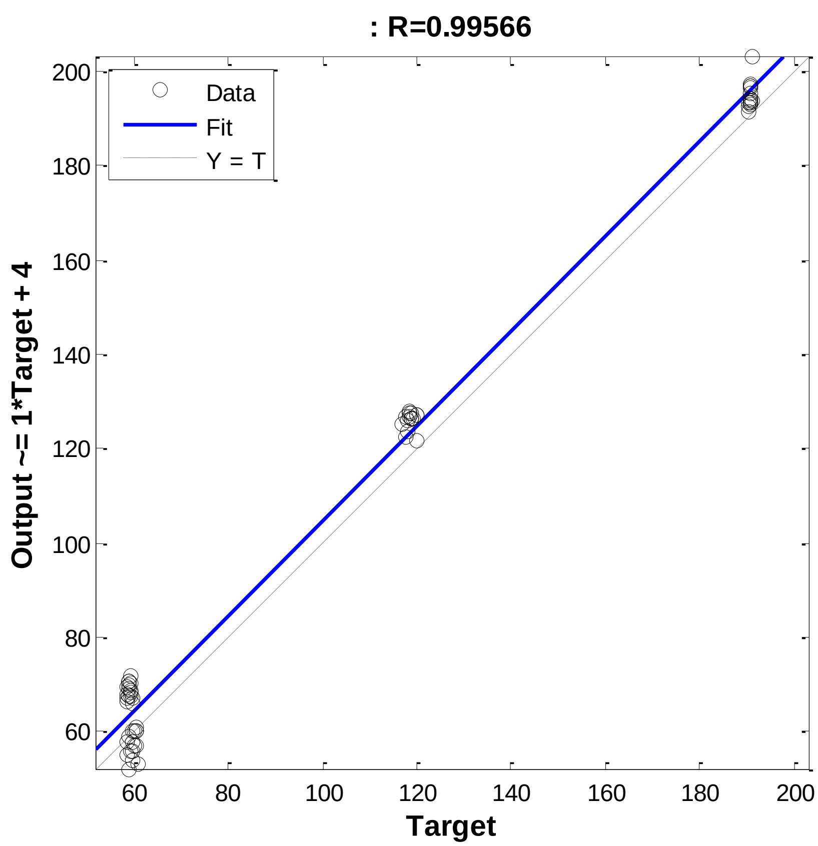

Figure 6.15: Regression plot for extra test (Scenario C)

Why better performance?

A good explanation for the better performance in scenario C compared to the scenario B, could be the fact that we do not have a big dataset to train the network, so adding a new output to the network while this new output has a mathematical (sum) relation with the two of the other outputs can increase the complexity of the model and consequently a better training and performance for the network.

6.4. Other Scenarios

In this work, we have tested scenario A, B and C with different inputs and outputs but the result was not promising or the prefromance has not changed as we expected so we will not discuss these scenarios and here we just briefly mention them. For the scenario A , we added gas rate as output, the perfromance has increased and the result for the extra test was better compare to the basic scenario but as we had two better scenario B and C later so we will not mention the result.

Another scenario that we have tested is removing the downhole temeprature from the inputs. As we discussed in the chapter 5, this variable could be removed from the inputs due to the fact that it has just relation with water rate and we couldn’t find any clear relation between this variable and the other rates. In this scenario the performance of the network didn’t change too much but the result of the extra test was better when we had this variables as input. Same situation has happened for choke opening percentage. The reason that we kept there variables as input is the fact that the training dataset is small and any little information can improve the network performance. But the result fo the network haven’t changed that much and it shows that with a biger dataset for training we can remove these variables from the model.

7. Conclusions

In this chapter, we present the conclusions of this work. Some considerations regarding the data analysis policy and model performance are given. It ends with some suggestion regarding feature work are further development.

7.1. Conclusions

With this work, we could conclude that, the fluid rates of a well can be predicted by using Artificial Neural Network (ANN). First, we approached the problem with a simple scenario to predict the liquid rate that sounds reasonable and more predictable. The result for this approach was not satisfactory. Then we approach the problem with another scenario to predict all production rates means oil, water and gas. The predicted results were very close to the actual rates with reasonable low errors. Finally, we slightly changed the current scenario by adding the sum of oil and water called liquid rate as an extra output for the network. This approach could increase the network accuracy by predicting results that are more accurate.

We also could show that, Exploratory Data Analysis (EDA) approach and graphical technique tools can be used to analysis the well data. With this approach, we could easily reveal the relationship between the variables and identify the possible input variables for the model. However, for large-scale dataset some preprocessing techniques such as cleaning, integration and particularly reduction must be applied.

We also discovered some problems regarding the dataset we could prepare to train the network. Despite analyzing millions of records, the final result that we could prepare for the network is a small dataset. This happens due to the nature of the problem we are dealing with that the well test operations that provide us production rates, are carried out about once a month for each well. We can take advantages of this situation, that we do not need to analysis all the data but the data of the periods that well test operations have carried out. This gives us an opportunity to prepare a bigger dataset with less time and effort.

There are some exceptions in the prepared dataset with no enough similar data to train the model for these exceptions. This kind of data decreases the performance of the model. Regarding this problem, we suggest to collect more data close to these exceptions and retrain the network.

More suggestions regarding feature works and development will be discussed in the next section.

7.2. Future Developments

7.2.1. Analysis more data

Considering Big Data approach, more data could be included in this work, the data that may affect the behavior of the well or reservoir directly or indirectly such as weather data (temperature on the surface, wind and so on), some other information written by engineers who are in touch with the well operations such as emails, letters and notes. It could be also interesting to study if problems such as earthquake, hurricanes or tsunami which have happened around the reservoir have affected the production and the reservoir specification or not.

7.2.2. Add previous production rates as inputs

Studying well performance[1] and productivity index[2] show that oil rate, calculated in well test operation, is an important parameter to calculate the productivity index. The idea is that, moreover than current inputs we have chosen for the network, we use oil, water and gas rates of the previous well test operation as inputs for the next well test operation. In this case, actually not only we can predict the next production rates but also we can train and update the model with the last rates.

7.2.3. Model the reservoir

Another idea is to make a model of reservoir not a well. Since the serveral wells are connected to the same reservoir and the specification of the reservoir such as reservoir pressure,temeprature and hydrocarbons ingredients for all these wells are the same it would be interesting to model the reservoir.We can create this model for the reservoir by collecting all the data for all the wells in the reservoir. In this case we can make a larger dataset of the inputs and outputs for the model and the currenct problem for the exceptional data will be automatically solved.

1. http://petrowiki.org/Oil_well_performance
2. http://petmuh.blogspot.no/2008/05/productivity-index.html

References

1. Oil and Gas Fields in Norway, Industrial heritage plan, can be found at web address: http://www.norskolje.museum.no/stream_file.asp?iEntityId=1900
2. ConocoPhillips Norway, Norway Operations, Greater Ekofisk Area, Eldfisk; From: http://www.conocophillips.no/EN/our-norway-operations/greater-ekofisk-area/eldfisk/Pages/eldfisk-2-7b.aspx
3. http://www.npd.no/Global/Engelsk/3-Publications/Facts/Facts2012/Figures/Chapter-10/Ekofisk.pdf
4. Devold, H.; Oil and Gas Production Handbook; And Introduction to oil and gas production, transport, refining and petrochemical industry.
5. Norwegian Petroleum Directorate from : http://factpages.npd.no/factpages/
6. Rumbles, L.; Importance of flow measurement for Separators available at web address : http://www.emersonprocessxperts.com/2014/01/importance-of-flow-measurement-for-separators/#.U44Fc_mSzgk
7. Kumar, S. and Gupta, s; Design of an Efficient Clustering using GNG and SOM.
8. N Tyagi, A. Solanki and S. Tyagi, 2010. “An algorithmic Approach to Data Pre-processing in Web Usage Mining”, Int. Journal of Information Technology and Knowledge Management, Volume 2, No. 2, pp. 279-283.
9. NIST/SEMATCH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/
10. Ronald Cody, Ed.D., Robert Wood Johnson Medical School, Piscataway, NJ; Data Cleaning.
11. Jinchuan CHEN, Yueguo CHEN, Xiaoyong DU, Cuiping LI, Jiaheng LU ,Suyun ZHAO, Xuan ZHOU; Big Data Challenge: A Data Management Perspective.
12. Rose, M. Data Analytics; The web address is : http://searchdatamanagement.techtarget.com/definition/data-analytics
13. Ratner, B., Statistical Modeling and Analysis for Database Marketing: Effective Techniques for Mining Big Data.
14. Society of Petroleum Engineers,Challenges in Reusing Produced Water; available at address : http://www.spe.org/industry/docs/reusingwater.pdf
15. Katie Guerra, Katharine Dahm, Steve Dundorf, Oil and Gas Produced Water Management and Beneficial Use in the Western United States. Available at address: http://www.usbr.gov/research/AWT/reportpdfs/report157.pdf
16. Duru, O., Modeling of Reservoir Temperature Transitions, and Parameter Estimation Constrained to a Reservoir Temperature Model.
17. Brown, G. Downhole Temperature from Optical Fiber. Available at address: http://www.slb.com/~/media/Files/resources/oilfield_review/ors08/win08/downhole_temperatures_from_optical_fiber.pdf
18. Petrowiki, Gas Lift System Design; Available at address: http://petrowiki.org/Gas_lift_system_design

19. Engineering Design Encyclopedia, Choke Valves, Can be found at address: http://www.enggcyclopedia.com/2012/03/choke-valves/
20. Schelumberger, Oilfield Glossary; Available at address : http://www.glossary.oilfield.slb.com/en/Terms/c/choke.aspx
21. Experience in Oil and Gas Industry, About Chokes in Oil and Gas well ; The web address : http://flyinggame.blogspot.no/2009/01/well-is-normally-producing-fluid-with.html
22. The number of hidden layers, http://www.heatonresearch.com/node/707
23. Xu, S and Chen, L , A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining. http://ecite.utas.edu.au/54013
24. http://www.mathworks.se/help/nnet/ug/analyze-neural-network-performance-after-training.html
25. Haykin, Simon. Neural Networks and Learning Machines(Third Edition).
26. Stergiou, Ch. And Siganos,D. Neural Networks. http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
27. Nervous system vocab http://quizlet.com/12470767/nervous-system-vocab-flash-cards/
28. Neural Network, the Human Brain and Learning http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol2/cs11/article2.html
29. Artificial Neural Network Technology from website : http://www.psych.utoronto.ca/users/reingold/courses/ai/cache/neural2.html
30. Artificial Neural Network http://en.wikibooks.org/
31. Wolfram mathematica documentation center, Forward and radial basis Function Networks. Url: http://reference.wolfram.com/
32. Shepherd, A. , Protein Secondary Structure Prediction with Nueral Networks. Url : http://www.biochem.ucl.ac.uk/
33. Supervised learning ,Url: http://www.saylor.org/site/wp-content/uploads/2011/11/CS405-6.2.1.2-WIKIPEDIA.pdf
34. BackPropagation url : http://en.wikipedia.org/wiki/Backpropagation
35. Kjela, G. Short-term Forecasting of Electricity Consumption using Gaussian Processes.
36. Wendy L. Martinez and Angel R. Martinez; Exploratory Data Analysis with MATLAB, CHAPMAN & HALL/CRC.
37. ConocoPhillips, Well History 2-7-B-19.
38. Information integration and Governance ; Web address :http://www-01.ibm.com/software/data/integration/
39. Pyle, D. ; Data Preparation for Data Mining; Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©1999.
40. Petro Strategies, Inc. Learning Center. Natural Gas Operation; Production. The web address : http://www.petrostrategies.org/Learning_Center/production.htm
41. S. A. Al-Yahya and B. D. Al-Anazi; Reservoir monitoring and performance using Simbest II Black Oil Simulator for Middle East reservoir case study.
42. Ray Mireault, P. Eng., Lisa Dean, P. Geol. Reservoir Engineering for Geologists.

43. C.V.Millikan, Amerada Petroleum Corp. Botton-Hole Pressure in Current Oil-Well Operations.
44. PetroWiki , Reservoir Pressure and Temperature available at web address: http://petrowiki.org/Reservoir_pressure_and_temperature
45. Wellhead definition. Available at address : http://en.wikipedia.org/wiki/Wellhead
46. Wellhead systems. Available at address : http://petrowiki.org/Wellhead_systems
47. Stream Flo Industries LTD. Wellhead and Gas Valves. Available at address : http://www.streamflo.com/media/uploads/wh_gv_brochure.pdf
48. Rigzone web site, How Does Artificial Lift Works? Available at address: https://www.rigzone.com/training/insight.asp?insight_id=315&c_id=4