The Pennsylvania State University

The Graduate School

Department or Energy and Mineral Engineering

DEVELOPMENT OF AN EXPERT SYSTEM TO IDENTIFY PHASE EQUILIBRIA AND ENHANCED OIL RECOVERY CHARACTERISTICS OF CRUDE OILS

A Thesis in

Energy and Mineral Engineering

by

Luoyi Hua

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2012

The thesis of Luoyi Hua was reviewed and approved* by the following:

Turgay Ertekin
Professor of Petroleum and Natural Gas Engineering
Thesis Advisor

Zuleima T. Karpyn
Associate Professor of Petroleum and Natural Gas Engineering

Li Li
Assistant Professor of Petroleum and Natural Gas Engineering

R. Larry Grayson
Professor of Energy and Mineral Engineering
Graduate Program Officer of Energy and Mineral Engineering

*Signatures are on file in the Graduate School

ABSTRACT

With the increasing demand of oil and gas in the past decades, great endeavors in the oil industry have been devoted to develop and incorporate better techniques for production. Enhanced oil recovery (EOR) is put forward as an effective way to maintain the reservoir pressure and increase oil production. Because of the complexity of reservoir conditions and interpretation of case dependent production history, it is time consuming to run the reservoir numerical simulation applying EOR method such as gas flooding, thermal flooding and chemical flooding. In the literature, one can find tools which are developed to provide screening criteria for oil recovery methods and assessment of production scenarios. Some of those tools are qualitative utilizing experimental field applications knowledge and some are based on models which are quantitative.

Artificial neural network(ANN) based tool-box by Claudia Parada(2008) provides an alternative way for screening and designing enhanced oil recovery methods with four fluid types included to investigate EOR processes. The evaluation of EOR production scenarios or performance forecast for crude oils may not be easily obtained with high accuracy because in reality there are many crude oils with various physical and thermodynamic properties. In this research, four artificial neural networks are developed to represent four fluid types, respectively. Networks predictions are interpreted to categorize crude oils and the classification results are validated by numerical simulations. An expert system integrating all networks by Graphical User Interface is created to categorize a crude oil in a visualized manner.

Once the classification is achieved for a particular crude oil, investigation of appropriate recovery techniques and adequate design guidelines can be applied to a certain fluid type representing this crude oil in Parada’s tool-box with specific reservoir characteristics. Thus, the developed expert system assists in narrowing down the selection of a proper fluid type combined

with feasible EOR processes, which subsequently can help to reduce time-consuming experiments.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Turgay Ertekin for being my master research advisor and academic mentor during my study in Pennsylvania State University. His continuous, illuminated, patient and erudite guidance is essential in my three years knowledge growth. I would like to express my sincere gratitude to him because he leads me to the world of petroleum engineering and introduces a fancy idea of artificial neural network. I truly appreciate his contribution at every step of my work. I would like to thank my committee members Dr. Zuleima Karpyn and Dr. Li Li.

I would also like to thank doctoral students Yogesh Bansal and Ihsan Burak Kulga and master student Phani Bhushan Chintalapati for their advice and opinion during my master research. I thank the Department of Energy and Mineral Engineering for acceptance made and providing scholarship to me.

I would like to thank my friends Xinqian Li, Hang Li and David Lin for their encouragement and help when I am having a hard time.

Finally, I would like to show my deepest gratitude to my mother Yanwen Chen. I would not complete my master study and thesis without her support and love. This thesis is dedicated to her and my father Yongshen Hua.

Chapter 1

Introduction

The increasing exploitations of crude oil in the past decades make the industry to develop and apply better techniques during production for the existing and depleting oil fields. The conventional oil recovery after natural flow and secondary maintenance such as water flood is usually below 40%. Considering the progressively developing imbalance of the global demand and supply of oil, Enhanced oil recovery (EOR) is put forward as an effective way to maintain the reservoir pressure and improve oil production. EOR techniques employ fundamental physical and chemical rock and fluid interactions to improve reservoir sweep and reduce residual oil saturation. [Regiten, 2010] There are several EOR processes, while thermal injection, gas injection and chemical are among the most basic.

Because of the complexity of reservoir conditions and case by case production history, it’s time consuming and demanding to run the numerical reservoir simulation applied with EOR techniques such as gas injection, steam and water flooding. It usually takes longer time to determine the most proper EOR method for a particular reservoir. The technology of artificial neural networks (ANN) is widely used in the academic research and engineering industry for its capability of capturing and computing nonlinear relationships. ANN, inspired from basic biology, is working as an expert system. It is utilized to overcome inefficiencies and uncertainties of complicated reservoir models .It provides an alternative approach to run compositional simulation and work out reasonable results. With its efficiency, the production profiles of reservoir fluid are generated in a rapid manner for a large number of scenarios. Many ANN dependent researches have demonstrated several successful applications of the technology in the petroleum industry.

Using artificial intelligence, Claudia Helena Parada Minakowski [Parada, 2008] has developed an expert system which provides EOR project screening and design tool-box. Two black reservoir fluids, one volatile reservoir fluid and one real sample of heavy oil represent the conventional crude oils in this tool-box. Four recovery techniques are applied and investigated, which are N_2, CO_2, water flood and steam injection. Steam injection has been the most important EOR recovery process during the past decades for its field test and commercial implementation.

This study is focusing on developing a fast and accurate expert system to categorize a crude oil to one of the four reservoir fluid types by identification to phase equilibrium and crude oil characteristics. A commercial simulation CMG®STARS® is used to generate and validate the compositional simulation and ANN prediction. Winprob® phase behavior software is applied to determine the thermodynamic phase equilibrium of crude oils and then imported into STARS® to complete simulation.

The reservoir conditions follow the general data from several literature references [Victor M, 1987][Sarkar et al., 1994][DOFMaster]. The well configuration is a simplified inverted 5-spot well pattern with one steam injector and one oil producer, as shown in Figure 1-1. It's a square reservoir with uniform grid distribution.

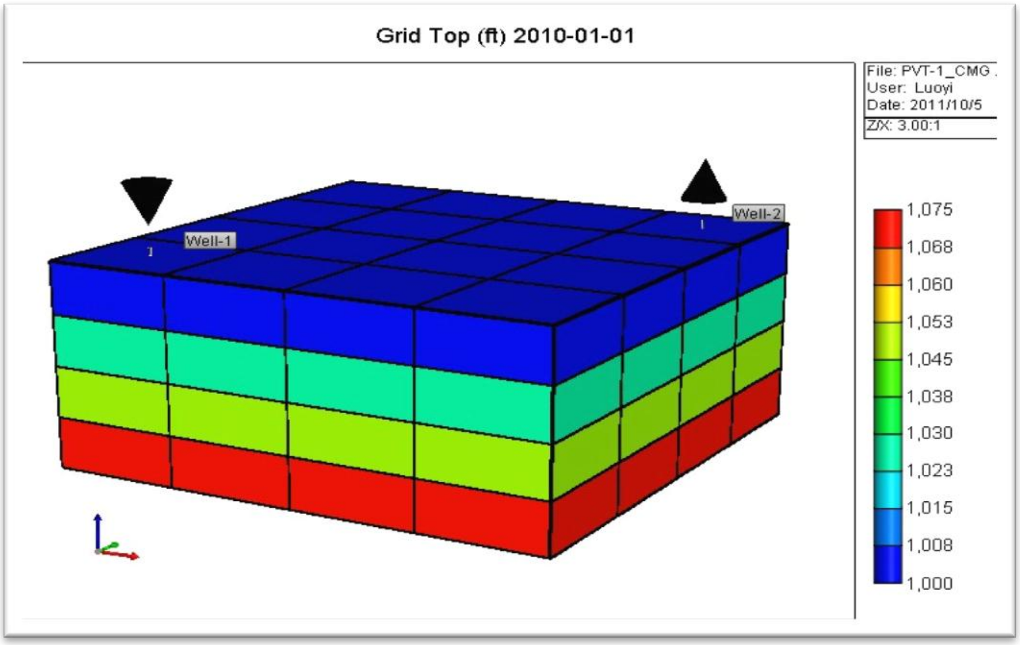


Figure 1-1 Simplified Well Pattern with One Injector and One Oil Producer

Chapter 2

Literature Review

To train and test the expert system by Artificial neural network needs ample simulation data sets. This project is focusing on the steam injection as one type of thermal processes to black oil, volatile oil and heavy oil. The reservoir model employs the compositional simulator to create the production data set. This chapter discusses the reservoir model for fluid and rock physical properties and thermal recovery screening criteria.

2.1 Reservoir Model

2.1.1 Reservoir Fluid Properties

The four base reservoir fluid properties considered in this study are from Parada's screen-tool box: two black oil fluids, one volatile oil and one heavy oil fluid from field which is undergoing steam injection (Table 2-1). Figure 2-1 shows the two phase envelopes which display the typical characteristics of four base crude oils in screening tool box sequentially. All the other hydrocarbon properties including composition, viscosity, density, C_{7+} properties etc. are within the ranges so that the categorization of reservoir fluids to one of the four base models is accurate and reasonable. The hydrocarbon fluid compositions for numerical simulation and network training are selected from literature.

Steam flood is primarily applicable to viscous oils but some lighter components will help for the steam to distillate the heavier components. [Lyons and Plisga, 2004]

Table 2-1 Molar Compositional Range of PVT Data

	Max.	Min.
CO_2	3.2	0.11
N_2	1.8	0.03
C_1	46.8	10.78
C_2	9.67	0.12
C_3	10.91	0.42
iC_4	4.26	0.3
nC_4	6.86	0.32
iC_5	3.71	0.29
nC_5	3.81	0.26
C_6	4.73	0.64
C_{7+}	86.07	8.52
C_{7+} MW	532	156
C_{7+} s.g.	0.925	0.782

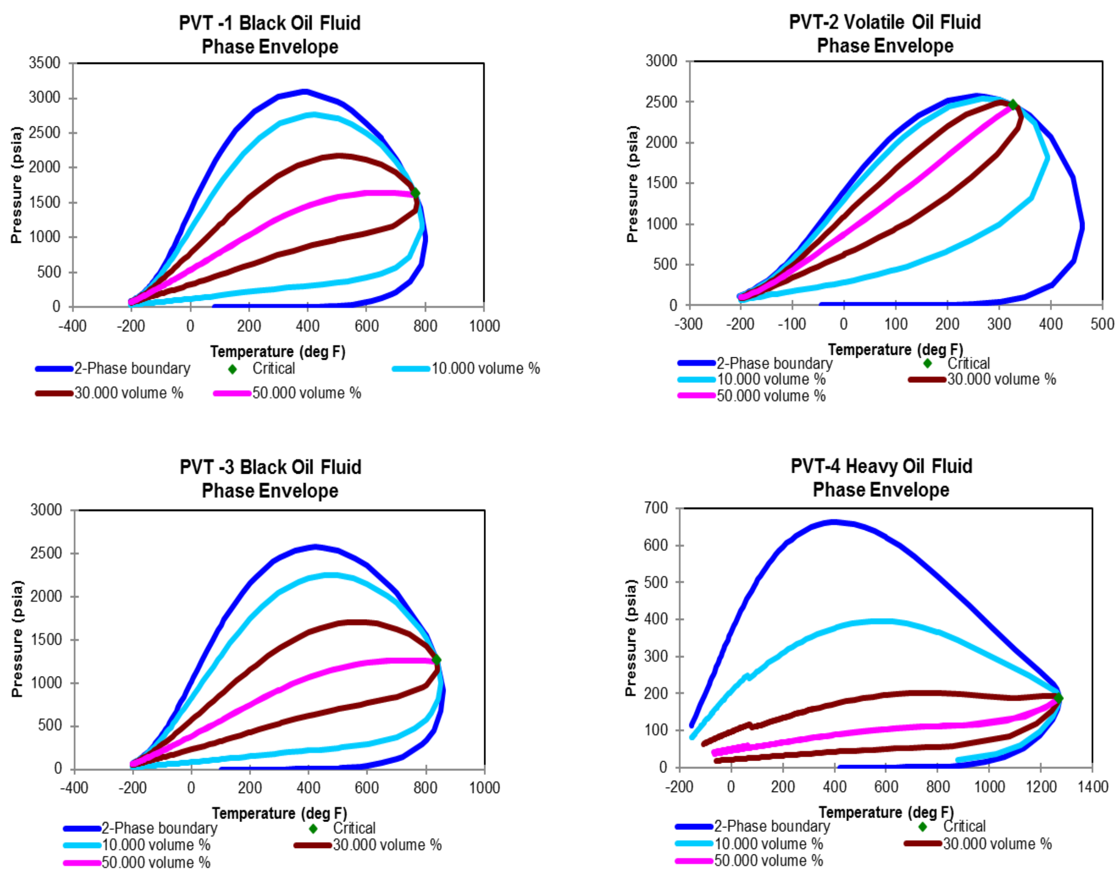


Figure 2-1 PVT Phase Envelopes (Generated by WinProb)

The fourth PVT is a real sample from the field which represents heavy oil. Heavy oils have API gravity usually less than 20° and a viscosity between 100-100,000 cp. Oil with less than 10° API gravity can be considered as extra-heavy oil. Regarding the steam recovery technical screening guides, the reservoir model is applicable for heavy crude oils. An example will be discussed later to demonstrate that the fourth trained network is capable to represent the heavy oil.

2.1.2 Reservoir Model Conditions Determination

Generally, compositional simulation is utilized to thermal process to the reservoir fluid. In order to build a reservoir model in STARS®, reservoir properties need to be determined and keep uniform while varying the compositions of the reservoir fluid components and C_{7+} physical properties. The steam injection process has its limitations in permeability, depth, reservoir pressure and net pay thickness.

In order to maintain the quality of steam in the wellbore and propagate a steam zone in the reservoir, permeability plays the important role to compensate the heat loss. Medium to high permeability or fine transmissibility provides favorable conditions for the steam injection processes. Table 2-2 summarizes the general screening criteria of reservoir candidates for steam flooding processes. [Green & Willhite, 1998]

Table 2-2 Screening Parameters for Steam Injection Processes[reproduced from Green & Willhite,1998]

Screening Parameters	Steam Injection
Oil gravity, API	10 - 34
In-situ oil viscosity, cp	20 ≤ 15,000
Depth, ft	≤ 3,000
Pay-zone thickness, ft	≥ 20
Reservoir Temperature, F	-
Porosity, ϕ	≥ 0.2
Average permeability, md	250
Transmissibility, md-ft/cp	≥ 5
Reservoir pressure, psi	≤ 1,500
Minimum oil content initial, $S0X\phi$	≥ 0.1
Total dissolved solids(TDS), ppm	-
Rock type	Sandstone or Unconsolidated sands

The steam injection is applicable to comparable shallow oil (usually less than 3280 ft) reservoirs so that the desire quality of steam can be injected to the wellbore. To minimize heat losses to adjacent formation, the thickness is usually more than 20 ft with medium to high porosity. Formation with higher porosity helps the heating efficiency, which indicates that sandstone or unconsolidated sands are the main target rock type.

Reservoir pressure is another criterion for steam flooding. The maximum injection pressure cannot exceed the critical pressure of steam, which is 3206.2 psia. Most successful steam injection projects perform more economically injection of 1500 psia or lower [Green & Willhite, 1998]. All the reservoirs models set the initial pressure at 500 psia to develop production performance.

This is a three-dimensional simulation reservoir model with uniform grid distribution. It is homogeneous and anisotropic in Cartesian coordinate system. The vertical permeability takes 1/10th of the value in the lateral directions. The initial reservoir pressure is 500 psia and depth is 1000 ft. The pay zone is 100 ft. and divided into four layers. The injector well and producing well has the completed perforation. No gas cap in the pay zone and oil-water contact surface is located below the reservoir layer. Wellbore storage and skin effect are neglected.

Relative shallow reservoirs with medium to high viscosity of the reservoir fluid are usually potential candidates for the steam injections. Shallow reservoirs are most likely associates with comparable low initial temperature. The goal for this project is to create an expert system based on characteristics of crude oil. In order to encompass the reservoirs fluids categorization applicable to screening and design EOR tool-box, it is also necessary to refer to screening guidelines such as temperature. The developed screening tool-box incorporates CO_2 and N_2 injection along with steam and water flooding. Taber and Martin [1983] concluded that a minimum pressure for miscible gas flooding to ensure miscibility. Developed miscibility only can be achieved at high pressures. Depth and pressure are two essential criteria to indicate a higher reservoir temperature. The simulation model refers to a set of field projects in the Malay Basin. Figure 2-2 shows the distribution of reservoir temperatures in the Malay Basins. Therefore, the reservoir temperature in this research is set at 220 °F for it is the highest frequency.

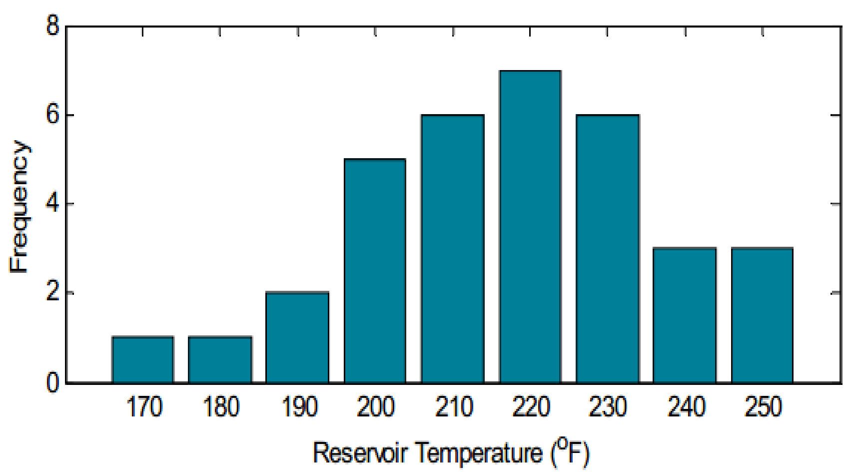


Figure 2-2 Reservoir Temperatures in the Malay Basin (from Teletzke,et al., 2005)

2.1.3 Rock Properties

Usually, the three-phase relative permeability curves are calculated from two sets data which are oil-water relative permeably data and gas-oil relative permeability data. Assuming that it is water wet reservoir. Stone' second model is performed to generate the two-phase relative permeability curve as Figure 2-3 and Figure 2-4. Baker's linear interpolation scheme [Baker, 1988] calculates the middle phase relative permeability. The points of equal oil relative permeability form a series of k_{ro} contours. Figure 2-5 is a schematic of ternary diagram of saturations with isoperm family curves of k_{ro}.

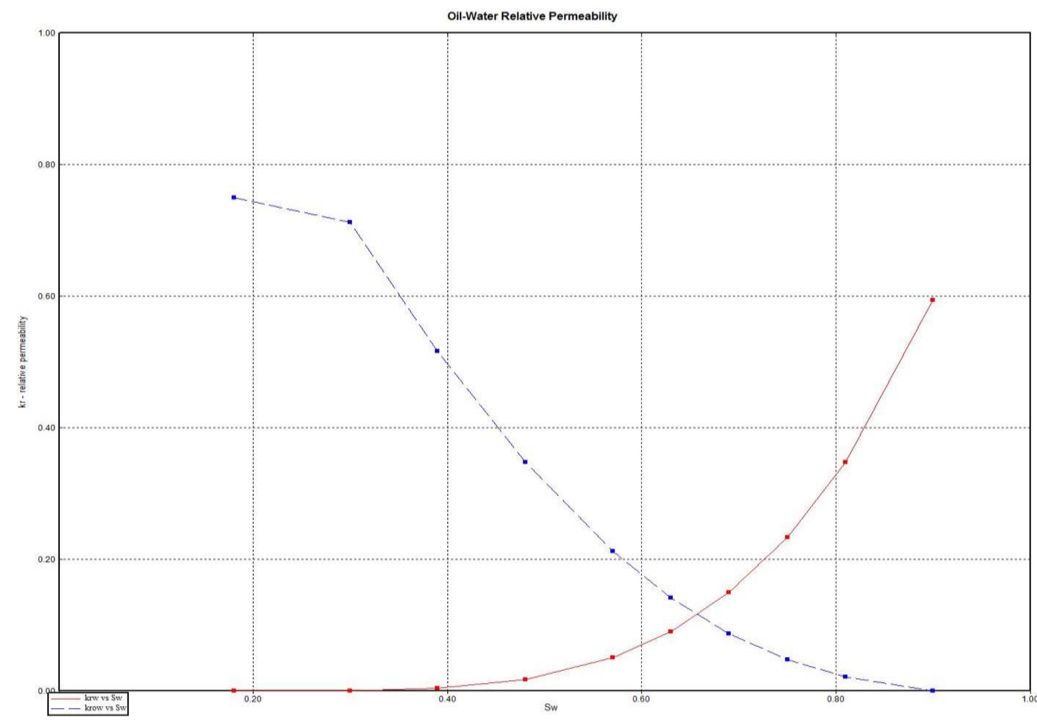


Figure 2-3 Relative Permeability of Water/Oil System.

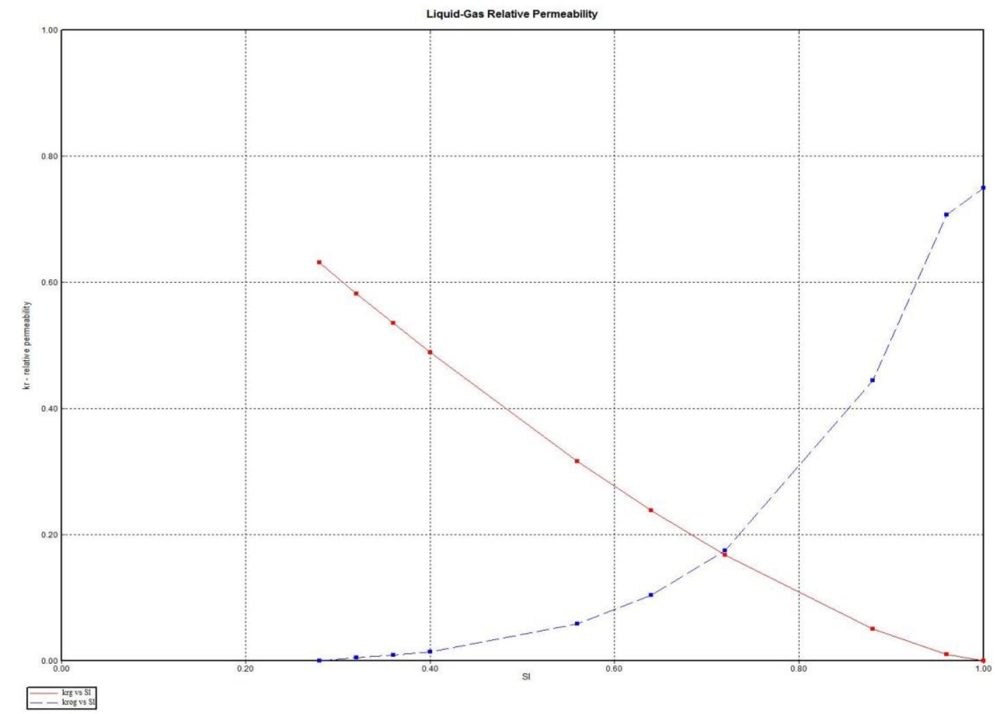


Figure 2-4 Relative Permeability of Liquid/Gas System.

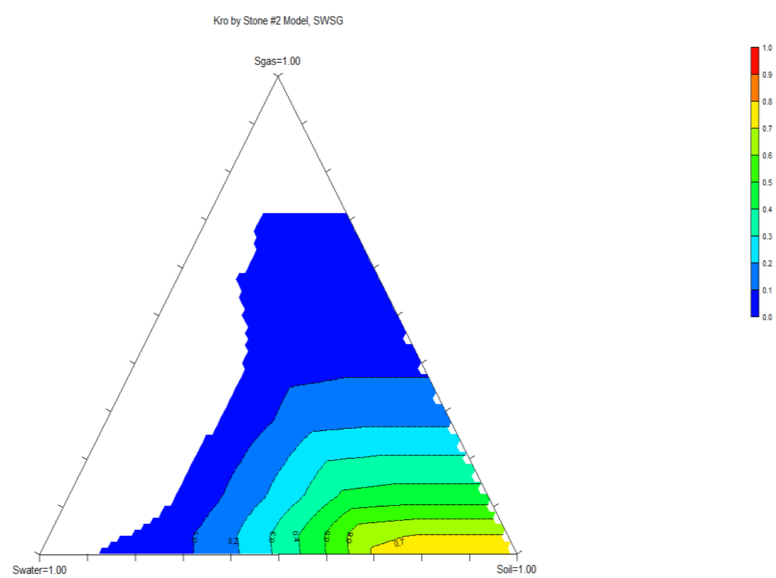


Figure 2-5 Ternary Diagrams of Saturations for k_{ro} Contours.

In thermal reservoir simulations, it is necessary to consider the heat loss from the reservoir to the surrounding environment, usually to the cap and base rock [Vinsome and Westerveld, 1980]. The heat will be carried into the reservoir fluid and the adjacent cap and base rock after steam is injected.

That's a key reason that compositional simulation is applied in the thermal recovery. Energy balance equation takes into account of the heat loss by using the semi-analytical method for the heat losses due to overburden and underburden. The only properties to determine heat losses to rock are thermal conductivity and heat capacity of the adjacent base and cap rock.

Another significant source of heart losses in thermal recovery processes is heat loss to fluid. Heat loss to fluids is taken place by diffusion of energy from a region of high temperature to a region of low temperature. [Parada, 2008] As the steam moves away from injector towards production well, the interface temperature is decreasing.

2.1.4 Well Patten

Well spacing is often smaller than for more conventional methods of exploitation in order to reduce the relative importance of heat loss.

If existing wells are drilled in a square pattern, 5-spots and 9-spots are common and both yield similar oil recovery and water oil ration performance. If the injected fluid is more mobile than the displacing fluid (which is often the case, especially when oil viscosity is high), a pattern having more producers than injectors may be desired to balance the injection and production rate. In cases where the injected fluid is less mobile or when the formation permeability is low, a pattern having more injectors than producers may be desired. [Lyons & Plisga, 2004]

For either the normal (or regular) 5-spot or the inverted 5-spot (inverted means one injector per pattern), space ratio of producers to injectors is 1:1. The distinction between normal and inverted is only important if a few patterns are involved, such as for a small pilot flood.

2.2 Reservoir Simulation via Soft Computing (ANN)

Reservoir simulators can be classified in several ways. The most common criteria for classifying reservoir simulator are the types of reservoir and reservoir fluids to be simulated and the recovery processes. Based on reservoir and fluid descriptions, reservoir modeling can been categorized as two types: black oil simulation and compositional simulations.

Black oil simulators are used in situations where recovery processes are insensitive to the compositional changes in the reservoir. [Ertekin et al] Usually reservoir temperature is assumed to be constant. Thus, the mass balance equation is not applied. The reservoir fluid is represented in three phase: oil, water and gas and they are pressure dependent.

Compositional simulators are employed when recovery mechanism is sensitive to compositional changes. Among different recovery processes, thermal recovery processes use the compositional simulator because these simulators work on the energy-balance and mass-balance equations at time. Besides, vapor-liquid flash calculation determines the phase equilibrium along the pressure depletion during the hydrocarbon productions. WinProb is applied for every crude oil to compositional changes.

The evaluation and optimization of EOR method for every particular reservoir is time and energy demanding using numerical simulation. Different design schemes with proper recovery mechanism can be only decided after numerous simulation trials. Additionally, k-value calculation is repeated at every time step to check the compositional change for its effect on production. With all these complexities, a reservoir simulation study requires high computational time. The artificial neural networks used in combination with a reservoir simulator demonstrate the great advantages in rapid and accurate prediction and performance as a soft simulation computing approach.

Chapter 3

Statement of the Problem

Enhanced oil recovery has been widely studied to interpret the technical screening criteria and optimize design for a specified reservoir production assessment. Numerical simulation can help to get the answer but it is time and extensive man power demanding as numerous runs varying parameters in small steps through trial and error. The complexity of each reservoir characteristics along with the field exploitation plans may not provide the most optimized design. Furthermore, there are overlapped regions of screening guidelines among different enhanced oil recovery processes, which add difficulty in assessments of EOR field design for a specific crude oil reservoir.

Artificial neural network proves to be a robust and efficient approach to mimic the key reservoir properties and predict good simulation results. Claudia Parada's[2008] EOR screening and design tool-box combines both reservoir characteristics and EOR operation scheme to provide recommended field development and design guideline by narrowing the possible EOR schedules. However, there are only four reservoir fluid types inside the tool-box where in reality there are many crude oils with various physical and thermodynamic characteristics. To utilize the tool-box, there comes up a complexity in operation as we do not know which fluid type should be investigated to stand for a particular case. Thus, the evaluation of EOR production scenarios or performance forecast under operating conditions for this crude oil may not be obtained with great accuracy.

In this study, the objective is to identify one of the four fluid types to best represent new crude oils using artificial neural technology. The categorization needs to be verified with numerical simulation. Once validation is achieved, an expert system will be developed integrating all networks and by Graphical User Interface (GUI) to facilitate in a user-friendly way to rapidly interpret a crude oil from thermodynamic properties to result in an accurate categorization so that a suitable EOR plan in Claudia Parada’s tool-box can be implemented.

Chapter 4

Artificial Neural Network (ANN) Development

4.1 Introduction

Artificial neural network (ANN) is a mathematical model inspired from the biological neural networks as they have a key similarity: highly connected neuromas.

A simplified biological neural network consists of three principal components: dendrite, axon and cell body. A simple schematic is expressed in Figure 4-1. The dendrites are tree-like nerve fiber that receive and transport the electrical signals into a cell body. The cell body effectively sums the signal and then carries the signal/information to another neuron through axon. The incoming signal is transmitted across synapses by means of a chemical process, which modifies the signal by scaling the frequency. The structure of the neurons and strengths of synapse affect the construction of a biological neural network. Accordingly, Figure 4-2 shows a single input neuron. The input p is multiplied by the weight w and then is sent forward. The bias b times its input of value 1 and also is transmitted forward. The transfer function is expressed by f and it produces the actual output from the net input n. In the manner of a biological neuron, the weight w corresponds to the strength of synapse. The cell body is the transfer function and the summations. Finally, a represents the signal transported to another neuron [Hagan, M. et al, 1996]. Obviously, the output depends on the transfer function and weights. Bias is another form of the weigh but has the input value of 1.

Artificial neural networks share other important characteristics with biological neural systems such as fault tolerance and damage tolerance. These robust features differentiate ANN from conventional computational schemes.

ANN is capable of dealing with nonlinear information processing to learn and understand the complex relations. It been studied and demonstrated successfully in many applications such as general mapping, forecasting, classification, pattern recognition and optimization with a high degree of fault tolerance and fast computation speed. [Hagan, M. et al, 1996] It is not a novel technique in the petroleum engineering. In recent years, an increasing number of studies have been observed from literature review. A number of previous studies from Pennsylvania State University students has exhibited its application in screening tool-box to EOR projects, CBM performance predictions [Srinivasan,2008],hydraulic fractured horizontal wells productions behavior[Kulga,2009], field development for natural gas [Adewale,2010],optimization of in-situ combustion process etc.[Shihab,2011]. Successful applications have demonstrated the use of artificial neural networks in petroleum engineering such as permeability prediction, well completion, seismic pattern recognition and analysis of gas well production. Neural networks technology could also contribute significantly to the analysis, prediction and optimization of well performance, integrated reservoir characterization and portfolio management. [Ail, 1994]

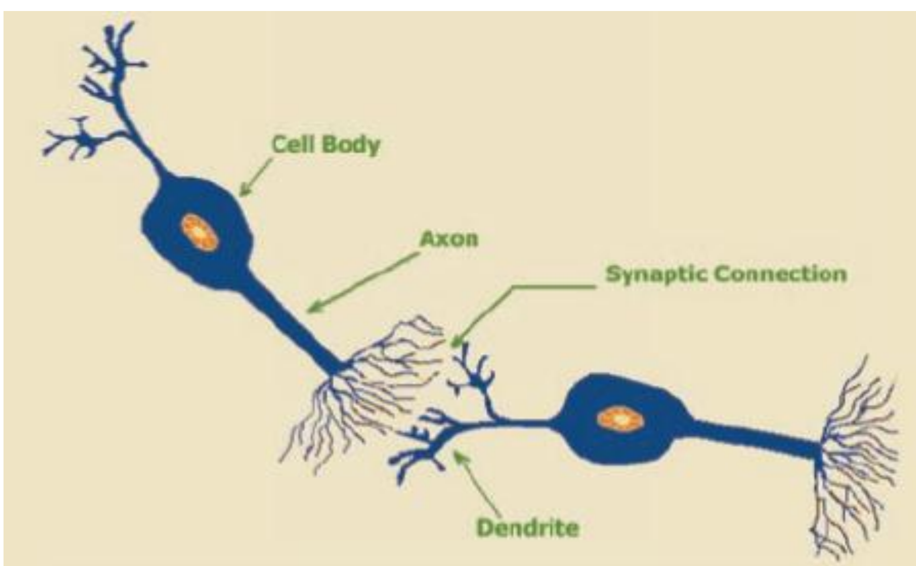
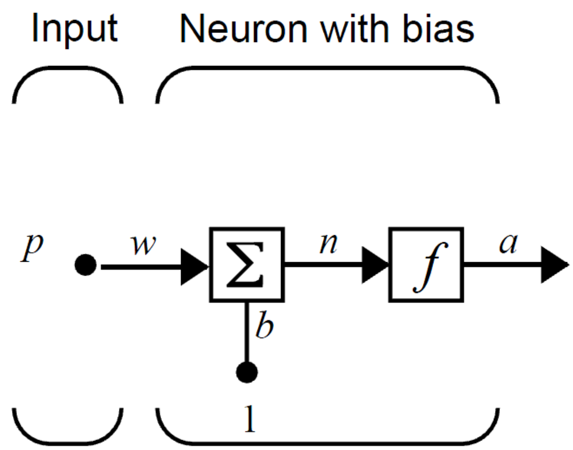


Figure 4-1 Schematic of Simple Biological Neural (Mohaghegh, 2000)



$$a = f(wp + b)$$

Figure 4-2 Single Input Neuron [Reproduced from Hagan, 1996]

Typically, multilayer networks are widely used due to their capability to be trained to approximate most functions. The neurons of hidden layer(s) are not predictable to get the optimal numbers but require trial and error in their development. With adequate numbers of neurons in the hidden layer(s), it would almost approximate most types of function to a certain degree of accuracy.

The beginning of the artificial neurons is considered as the introductions by Warren McCulloch and Walter Pitts in 1943. However, due to the limitation and incapability of its implementation, ANN technology has been becoming widely publicized and accepted until 1980s by improvement of multilayer perceptron networks associated with development of learning rules.

4.2 Feedforward Backpropagation Network

The multiplayer perceptron, trained by the backpropagation algorithm (BP) is the most widely used neural network. The basic BP algorithm is a generalization of a single-layer linear network and advances in dealing with the nonlinear inputs and outputs for multiplayer network.

Backpropagation learning algorithm is an approximate steepest descent algorithm using the performance index: mean square error. The overall procedures of BP are as follows:

1. Propagate the inputs feedforward through the network
2. Starting from the last layer, calculate and backpropagate the error value backward through the network layer by layer. This is how the BP algorithm got its name. The adjustments of weights are made in the backward direction.
3. Finally, weights and biases are updated using the approximate steepest descent rule.

Using BP, the hidden layer weights are updated using the error from the subsequent layer. Thus, the errors computed from outputs and target outputs are used to update the weights between the output layer and last hidden layer in the network structure. Likewise, the weights corrections are propagated in the backward directions until the update between the input layer and first hidden layer is achieved in an iterative manner. The process is repeated in the training until the total output error converges to a minimum or training limits are reached, the training is then terminated. [Patterson, 1996]

Network training using BP has three default processes: re-encode the unknown inputs; remove inputs vector with constant values and normalize the inputs values. Normalization rescales and standardizes the numerical inputs between a small range (usually between -1 and 1). It improves the fairness of training by preventing an input parameter with large value from swamping out another input that is equally significant but with smaller values. [AI-Fattah et al, 2001].

Functional links are usually a favorable pre-process to the network inputs. It provides additional connections to the network of input and output parameters. Hence, the network correspondingly enhances its predictive capability. An eigenvalue is a mathematical functional link and characterizes the data sets. It amplifies subtle difference between certain data sets or enhances the connections. The maximum eigenvalue is found to be the most effective. [Ramgulam & Erterkin & Flemings, 2007].

4.2.1 The Limitations and Variations of Backpropagation

The key drawback of the BP is that the mean square error for the multiplayer network is general much more complex and has many local minima. While the performance surface for a single-layer network has only one single minimum point and one constant curvature. The error surface varies dramatically over the parameters, and the determinations of the appropriate learning rate are difficult to achieve. The standard backpropagation algorithm modifies weight in the direction of most rapid decrease of the error surface for the current weights. This does not always move the weights directly toward the optimal weigh vector. That is to say when the BP algorithm converges we cannot be sure that we have an optimum solution [Hagan, M. et al, 1996].

The proper way to start the network parameters is using nonzero values for weights and biases. Small random values are favorable as they can keep away from the very flat region of performance surface. The very flat region usually refers distance from the optimum point. On the other hand, if the initial weights are too small, the gradient is close to zero, which may take longer time to train the network. Another key point is to try several different initial conditions to ensure that an optimum solution/global minimum point have been obtained [Hagan, M. et al, 1996].

The basic BP algorithm using the steepest gradient descent means it takes longer time to learn the network. There are plenty of researches focusing on the acceleration of the BP convergence speed. Overall, they fall into two categories. [Hagan, M. et al, 1996]

1. Development of heuristic techniques, arise out a study of techniques include varying learning rate, using momentum and rescaling variables.
2. Standard numerical optimization techniques: training feedforwad neural network to minimize squared error is a simply numerical optimization problem. Two successfully applied to the training of multiplayer perceptron are the conjugate gradient algorithm and Levenberg-Marquarde algorithm (a variation of Newton's method).

One way to improve the speed of training for backpropagation is by changing the learning rate. [Fausett.1994]Whether the learning rate is too high or too low the ANN architecture will suffer. If the learning rate is too low, the synaptic weights of each link will change slowly over time, which will make the learning process much too slow. On the other hand, if the learning rate is too high, the system will become unstable especially when the algorithm begins to diverge into a narrow error surface. It is oscillating back and force instead of smoothing out. [Haykin et al, 2009]

Adding momentum is one of the possibilities to update the weights that could still use a small learning rate and maintain a fairly rapid training speed. One good news for using momentum is that it's likely to reduce the weights updated at local but not global minimum. Because, the network training process is not in the direction of gradient only but in the direction of current gradient combined with the previous weight update correction (recent trends in the error surface). On the contrary to these benefits, learning rate thus limits the upper amount to update the weights by inadequate effectiveness. The combination of gradient to previous weights may lead to the wrong direction, which means to increase the error.

The basic backpropagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient) while in the conjugate gradient algorithm a search is performed so that the learning rate to the weight update is adjusted at each iteration. A search is made along the conjugate gradient directions to determine the step size that minimizes the performance function along that line. Scaled Conjugate Gradient *(trainscg*) is the one algorithm that utilizes the conjugate gradient approach but avoid the time-consuming line search. The *trainscg* routine can require more iteration numbers to converge than the other conjugate gradient algorithms, but the number of computations in each iteration is significantly reduced because no line search is performed.

Generally speaking, it is very tough to know which training function performs best for a particular problem. It depends on too many factors that it is even difficult to know the key factors in a given problem. There are some general conclusions after a couple of experiments carried on using some most prevailing training functions. The default backpropagation algorithm is Levenberg-Marquard. This is the fastest method yet requires large memory. Besides, the advantages of LM algorithm decrease as the number of network parameters increases. The conjugate gradient algorithms, in particular *trainscg* is applicable over a wide variety of problems without at the cost of lowering speed. Furthermore it requires relatively modest memory requirements.

4.3 Improving generalization

Overfitting is one of the problems during the neural network training. The larger the architecture of the network, the more complex problem the network can solve. The ideal network for a given problem is large enough so that it is able to train the network powerfully and will not create the overfitting problem. Unfortunately, the proper network is hard to tell and reach. Overfitting or memorization result in the bad prediction as the network does not give the corresponding outputs by new inputs. It forecasts the accurate results for particular data sets only. Actually, the network begins to associate the predicted output with the outputs belonging to training set. As the result, the network did not train appropriately and cannot represent the system. Several methods are developed to improve the network generalization.

Early stopping is a common method for improving generalization. All the data sets in this technique are divided into three parts: training, validation and testing. The training set, as indicated by its name, is employed to train the network. The weights and bias adjustment are computing by this set of data. Validation set is applied in the training network and the error is recorded. At the beginning, the error of the validation set shares the decreasing trend as the training set. At a certain point, the error of the validation set becomes roughly constant and starts to increase. It shows that the network begins to experience overfitting and memorizes the inputs parameters instead of training. The users are able to set up a number of iterations when the validation error increases. The training is stopped after it reaches the specified number, which is defined as maximum validation checks.

The testing set does not take part in the network training. It is applied to the trained network as new data set to analyze the performance of the network. The plot of testing data error in the company of the training and validation errors are useful. If the iteration number to obtain minimum error of testing set is dramatically different from the number of validation set, the network starts memorization and is necessary to be adjusted.

4.4 Data Generation for Artificial Neural Network

Typically, the technical screening criteria for enhanced oil recovery processes consider the reservoir fluid and rock properties. In addition, the initial reservoir conditions, well pattern, grid refinement contribute to the production as well. Since this project is to identify the characteristics of crude oil, each crude oil production history is compared with the four basic PVT models and records the differences of production rate at specific time period. The reservoir characteristics such as fluid and rock properties, steam injection schedule, production schedules are keeping uniform for all crude oil to train, test and validate networks.

A commercial simulator is utilized to predict the reservoir performances for fixed production plans. After collecting the production history (production differences), four neural networks are fed with certain inputs and outputs parameters. The successful networks are capable to forecasts the minimum production history differences compared with one of the basic PVT model. Several case studies are carried out to verify the accuracy of four network prediction as the last stage. The developed expert systems using artificial neural technology operate on black oil, volatile oil and heavy oil. In order to be more applicable to most of these three type reservoirs, the maximum and minimum values for some of the components have been extended to train and test four networks. There are eight extra hydrocarbon fluid samples for validation of the network after trained, which all compositions fall within the spectrum of specified data. Among the composition values, C_{7+} properties are the most vital data to distinguish each hydrocarbon fluid as well as the lighter compositions.

Table 4-1 lists the components' molar fraction range for basic PVT models and the network training samples referred to couple of published paper. Figure 4-3 to Figure 4-15 represents the compositional values of each component and they are rationally distributed along with the frequency. The figures show that the values are selected randomly and typical physical properties of black oil, volatile oil and heavy oil.

Table 4-1 Compositions range from Parada's research and forward ANN

Input	PVT Data (From Parada)		Network Training Data	
	Max.	Min.	Max.	Min.
CO2	3.2	0.11	3.15	0.038
N2	1.8	0.03	1.94	0.01
C1	46.8	10.78	63.49	6.04
C2	9.67	0.12	17.24	0.01
C3	10.91	0.42	12.07	0.021
iC4	4.26	0.3	5.67	0.068
nC4	6.86	0.32	12.27	0.01
iC5	3.71	0.29	11	0.26
nC5	3.81	0.26	5.95	0.01
C6	4.73	0.64	10.76	0.15
C7+	86.07	8.52	85.12	8.91
C7+ MW	532	156	368.9	160
C7+ s.g.	0.925	0.782	0.9594	0.7978

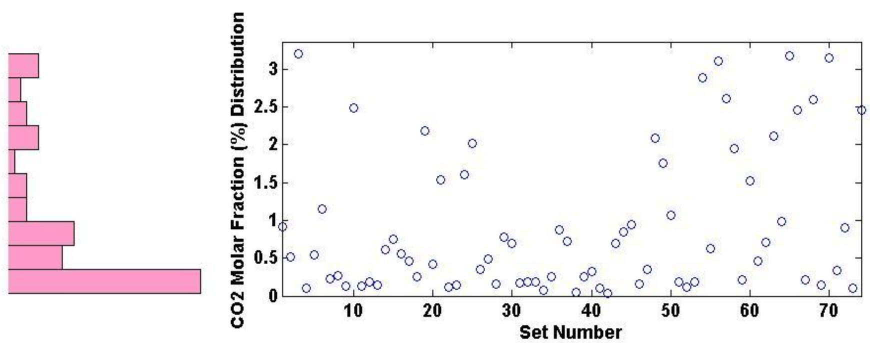


Figure 4-3 CO_2 Molar Fraction Distribution

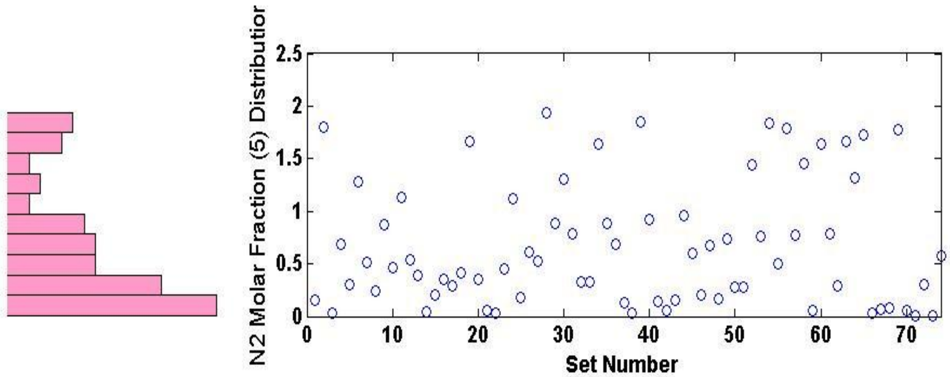


Figure 4-4 N_2 Molar Fraction Distribution

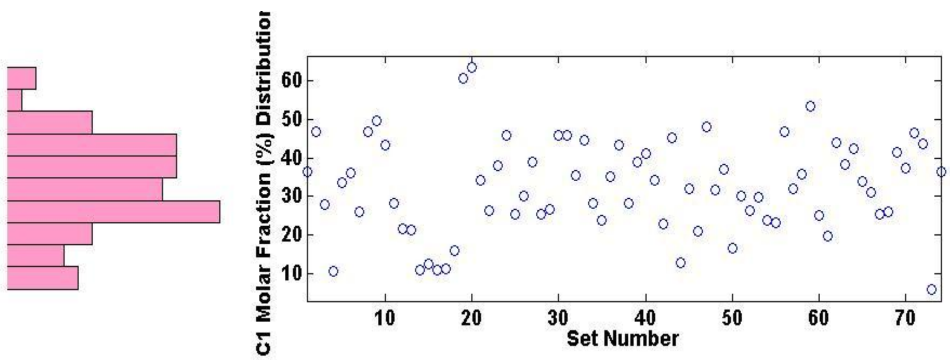


Figure 4-5 C_1 Molar Fraction Distribution

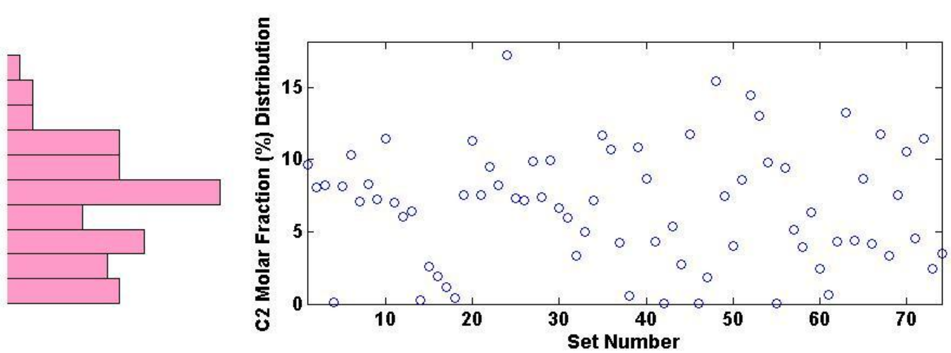


Figure 4-6 C_2 Molar Fraction Distribution

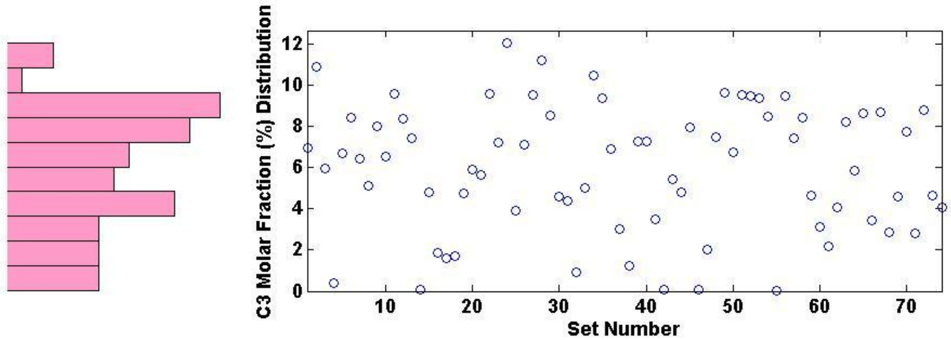


Figure 4-7 C_3 Molar Fraction Distribution

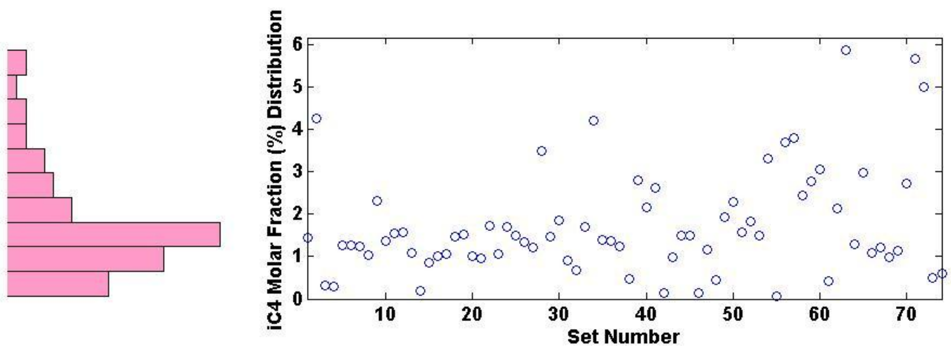


Figure 4-8 iC_4 Molar Fraction Distribution

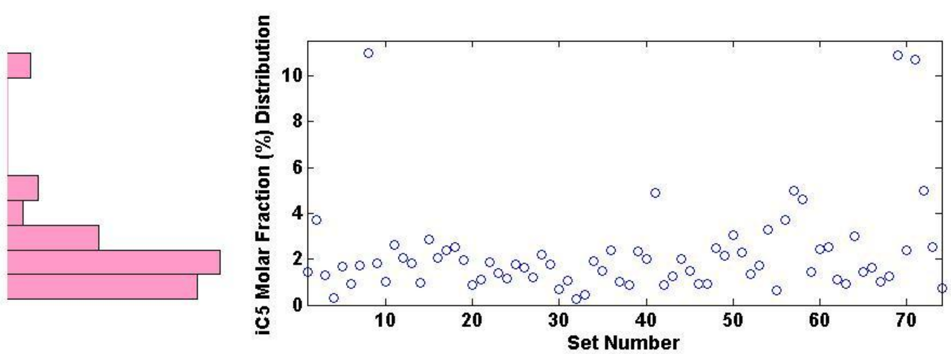


Figure 4-9 nC_4 Molar Fraction Distribution

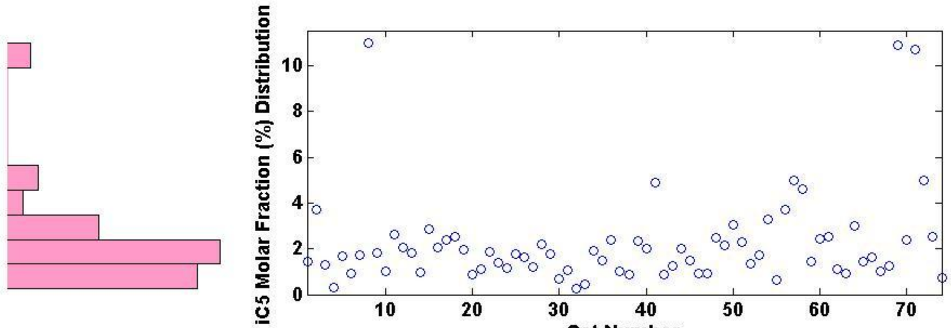


Figure 4-10 iC_5 Molar Fraction Distribution

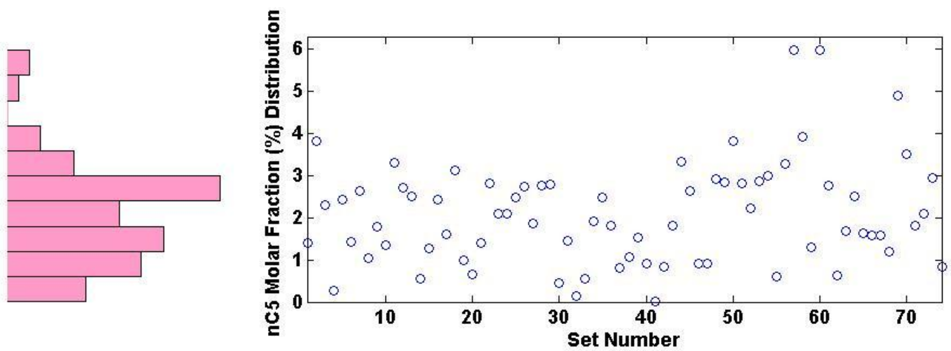


Figure 4-11 nC_5 Molar Fraction Distribution

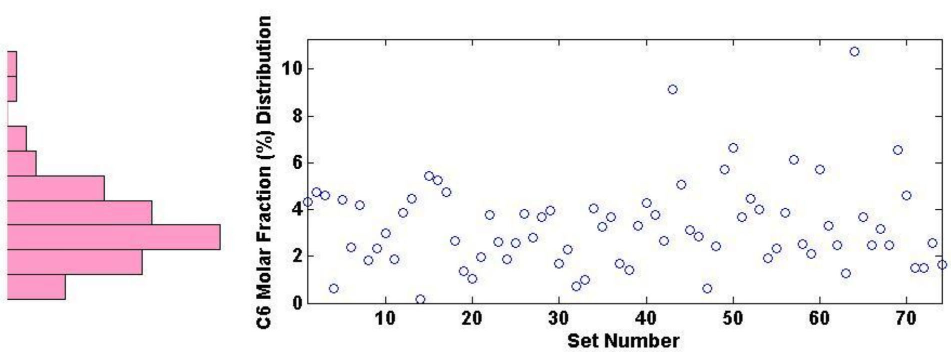


Figure 4-12 C_6 Molar Fraction Distribution

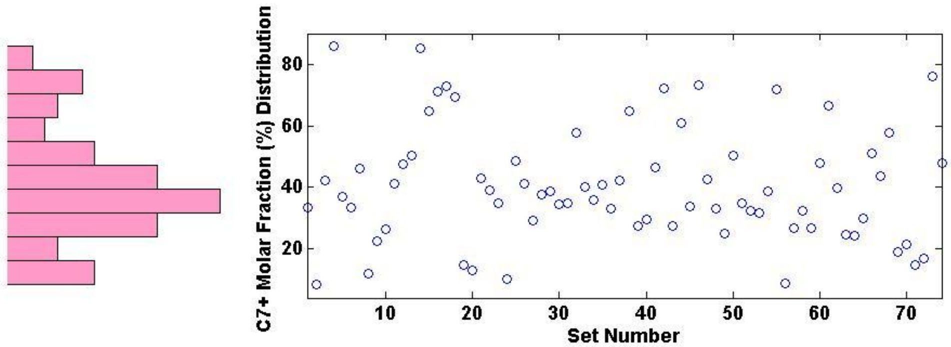


Figure 4-13 C_{7+} Molar Fraction Distribution

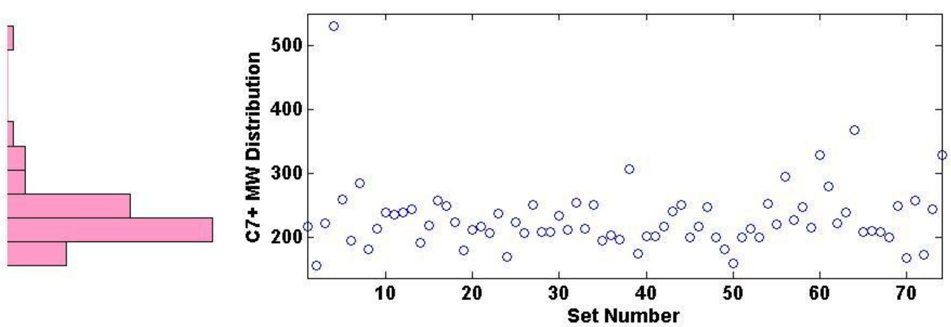


Figure 4-14 C_{7+} MW Distribution

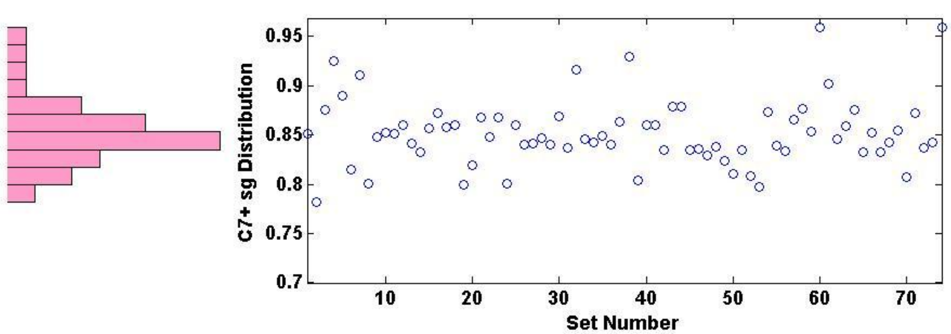


Figure 4-15 C_{7+} SG Distribution

4.4.1 Inputs Parameters Generation and Collection

Usually, the number of data sets is crucial to the development of network training. Too many data sets create more complexity to the network structure, which requires more time until a network is trained with accurate predictions and acceptable errors. On the other side, few numbers of sets may not enough for the network to understand and represent the inherent relationship between inputs and targets. Basically, there is no pre-determined numbers of data sets. In this research, 63 data sets used to train the networks. There are 8 extra data sets to test and validate the accuracy. The workflow of the training data sets is presented in Figure 4-16.

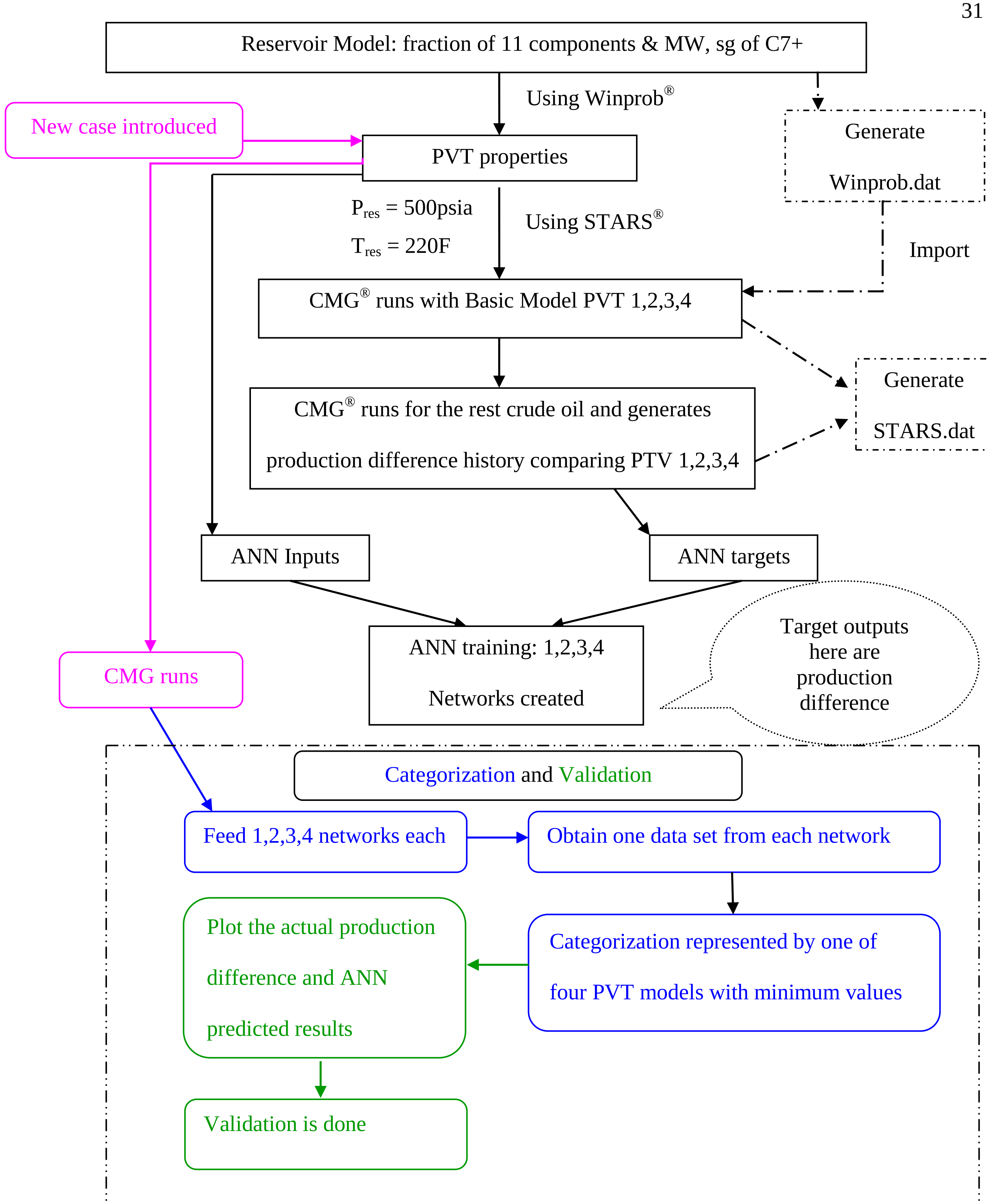


Figure 4-16 Workflow chart for ANN inputs and outputs generated

In this study, input data used in training are generated to identify the characteristics and phase equilibrium so that the numerical simulation is capable to obtain the pertinent physical properties and track the compositional change through steam injection and hydrocarbon production.

The viscosity of the hydrocarbon fluid is calculated by using Pedersen Correlation [1987]. The steam injection processes many times contain the nonlinearity in the flow equation of the viscosity temperature relationship. Hence, the temperature dependent μ_i is necessary to be represented adequately. The Pedersen Correlation Model [1987] has modified to improved viscosity results for mixture. It uses the principle of corresponding states to calculate the viscosity of each component first with coefficients *AVISC (cp)* and *BVISC* (temperature difference) using the following equation. The viscosity of the oil phase is obtained by the linear mixing rule as the second equation.

$$VISO(i) = AVISC(i) \times e^{\left[\frac{BVISC(i)}{T}\right]}$$

$$\ln(visoil) = \sum_{i=1}^{i=11} \ln[VISO(i) \times x(i)]$$

where *x(i)* is liquid phase mole fraction.

Generally speaking, there are no fixed rules to compile the networks for input and output parameters. A preliminary input data set in this study is the compositions of 11 components. However, the composition data sets as input data sets did not result in accurately trained network. This can be explained by stating that the combinations of the composition value did not represent unique characteristics of a certain crude oil. In other words, several distributions of the compositions can lead to the same or similar reservoir fluid properties such as critical pressure, critical temperature. In such case, there is possibility that the network will memorize the values rather than training the network properly.

Theoretically, artificial neural network development is selecting the most representative inputs and the outputs to construct the network. The inputs selection is a key point to the network. The ideal case is to pick inputs which are able to capture the main feature of the problem with fewer numbers of parameters. The critical observations and reservoir engineering judgment are essential to an effective network structure. The development is considered to be experimental and heuristic process. Fluid and rock properties, reservoir characteristics and operation strategies affect the productions explicitly. In a complex and multiple wells production problem for artificial neural training, spatial information of well, inter-well interactions along with individual well property is observed to contribute to the main features to reservoirs. By appropriate input and output entries and proper constructed network, a specific petroleum engineering problem can be well represented by an expert system.

Improved inputs for the network consist of thermodynamics properties of crude oil such as critical point temperature, cricondenbar, cricondentherm, physical properties at reservoir condition such as compressibility, viscosity, oil formation value factor and interfacial tension. However, the determination of inputs is through trial-and-error. Since the different properties of crude oil depend mainly on the methane and pseudo C_{7+} component, the molecular weight and specific gravity are critical to the network in terms of the identification of phase equilibrium. [Abass, 2009] Functional links are implemented as mathematical representations in forms of eigenvalue to scale critical point between cricondenbar point and criconertherm point using following equations:

$$E(CP,CB) = \max(eig\begin{bmatrix} Tc & Tcb \\ Pcb & Pc \end{bmatrix})$$

$$E(CP,CT) = \max(eig\begin{bmatrix} Tc & Tct \\ Pct & Pc \end{bmatrix})$$

Intuitively an analyst would like to include only those features that make a significant contribution to the network [Belue and Bauer, 1995]. However, in practice, many variables are noisy, irrelevant or redundant and provide little information. By elimination these least influencing parameters, it can reduce the size of the network and often result in a more accurate and faster training. The inputs selection becomes even effective for multilayer networks. In this study, a method reviewed by Redondo and Espinosa is applied to optimize the network inputs entries using:

$$S_i = \sum_{j=1}^{N} |w_{ij}|$$

where S_i represents the relevance of input *i* and w_{ij} is the weight of input *i* to the neuron *j* in the first hidden layer. The symbol *n* is the number of neurons. The lager relevance implies the higher influence of the input to the network.

Using the command *net.IW{i,j},* the weights of inputs to hidden layer can be extracted. *i* and *j* are the number of neuron and inputs respectively.

After many trials of the different input parameters collection and order, 15 inputs are used as inputs to train the four networks. Table 4-2 lists the 15 inputs for the ANN training of 53 data sets and extra 8 data sets.

The cricondenbar corresponds to the maximum pressure on the PVT phase envelope whereas the cricondentherm is for the maximum temperature. It illustrates the shape of the two-phase envelop and gives a brief idea of the reservoir fluid type.

At reservoir temperature of 220°F, saturation pressure of each crude oil is calculated. There are two saturation pressures obtained, the upper and lower values. The upper values can be either bubble point or dew point fluid (if it’s a retrograde gas condensate). The lower value is the dew point fluid. The input parameter for the ANN is taking the upper value as one of the input parameters. Compressibility refers to oil and liquid compressibility at initial conditions, while a set of compressibility values with the pressure increment is imported to STARS® model to track the compositional changes. Solution GOR is of unit SCF/STB at 14.70 psia and 60.0°F. Oil formation value factor (Oil FVF) calculates the ratio between volume of oil and dissolved gas at initial reservoir condition and volume of oil at standard condition.

Table 4-2 Network training and validation input parameters

	Input Parameters for ANN		Input Parameters for New Data Sets	
	Max	Min	Max	Min
Tc (°F)	1124.895	344.574	840.127	711.682
Pc (psia)	4281.37	344.7	2210.36	1020.05
Tcb (°F)	633.314	179.608	421.406	324.695
Pcb (psia)	5802.718	537.083	4020.673	2091.109
Tct (°F)	1131.968	528.517	874.629	754.228
Pct (psia)	1350.588	384.817	1079.11	783.198
Saturation Pressure (psia)	5784.938	349.688	3777.899	1787.975
Compressibility ($psia^{-1}$)	2.77E-05	4.72E-06	1.528E-05	1.1E-05
IFT (lbf/ft)	0.008167	0.0027844	0.0068785	0.005322
Solution GOR (SCF/STB)	320.12	35.5003	172.7741	78.8897
Oil FVF	1.3256	1.0429	1.1715	1.0873
C_{7+} MW	368.9	160	248.8	208.7
C_{7+} sg	0.9594	0.7978	0.877	0.832
E(CP,CB)	4539.927	1130.375	2885.7131	1874.138
E(CP,CT)	4508.905	1217.818	2670.9692	1737.733

4.4.2 ANN Outputs Generation and Collection

Artificial Neural Network Outputs Generation:

As discussed earlier, initial reservoir condition, well pattern, grid refinement, and thermal recovery scheme contribute to the oil production. In addition, reservoir settings and well pattern are two critical factors. When running the numerical simulations, these technical data specified respectively and listed in Table 4-3 and Table 4-4 according to several published SPE papers.

Table 4-3 Technical Reservoir Settings

Reservoir Settings	
Reservoir Initial Pressure, psia	550
Reservoir Initial Temperature, °F	220
Reservoir Depth, ft.	1000
Reservoir Thickness, ft.	100
Water-Oil Contact Depth, ft.	1100
Gas-Oil Contact Depth, ft.	1000
Formation Compressibility, psia^{-1}	0.0005
Thermal Expansion Coefficient, °F^{-1}	0.0004
Overburden Volumetric Heat Capacity, Btu/(ft^3*°F)	35
Underburden Volumetric Heat Capacity, Btu/(ft^3*°F)	35
Overburden Thermal Conductivity, Btu/(ft*day*°F)	24
Underburden Thermal Conductivity, Btu/(ft*day*°F)	24
Water Phase Thermal Conductivity, Btu/(ft*day*°F)	24
Oil Phase Thermal Conductivity, Btu/(ft*day*°F)	24
Gas Phase Thermal Conductivity, Btu/(ft*day*°F)	24

Table 4-4 Wells Specifications

Injection Well	Injection time, day	25
	Soaking, day	15
	Injection rate, bbl/day	1000
	Steam Temperature, °F	518.2 [DORMaster]
	Steam Quality	0.8
	Wellbore Radius, ft.	0.5
	Perforation Completion	Yes
Production Well	Production period, year	2
	Production Constrains	Min Bottom Hole Pressure= 20 psia Max Steam Production(if happens) = 10 bbl/day
	Wellbore Radius, ft.	0.28
	Perforation Completion	Yes

The production histories of all reservoir fluid (all together 74) are generated with respect to oil rate daily and cumulative oil at standard conditions. Figure 4-17 and Figure 4-18 show the four basic PVT model oil production rates and cumulative oil production.

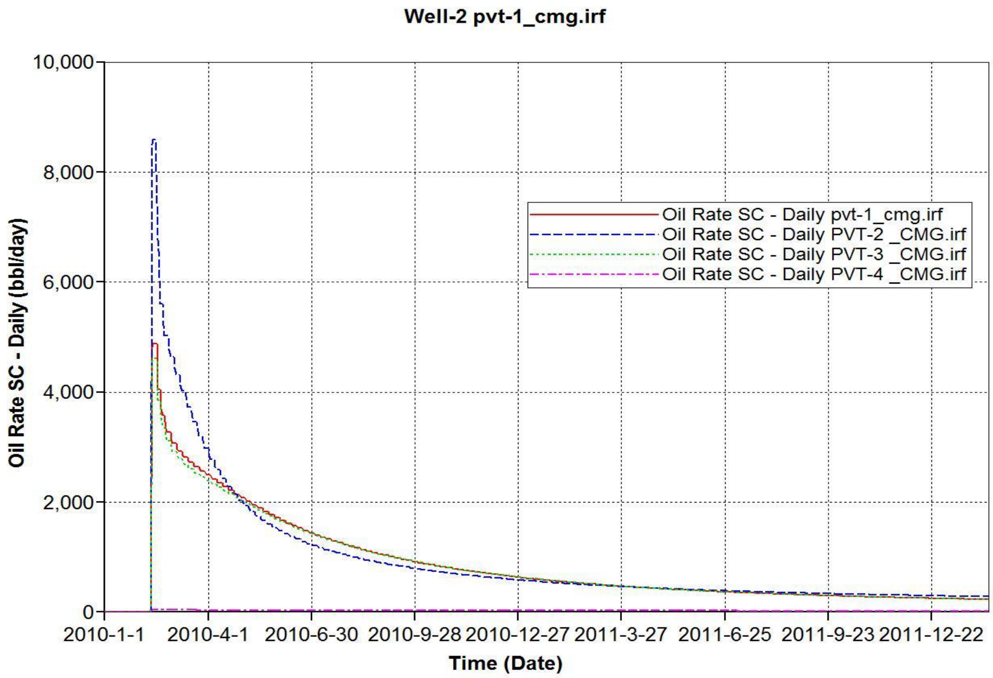


Figure 4-17 Production well (Well 2) History of Four Basic PVT Models

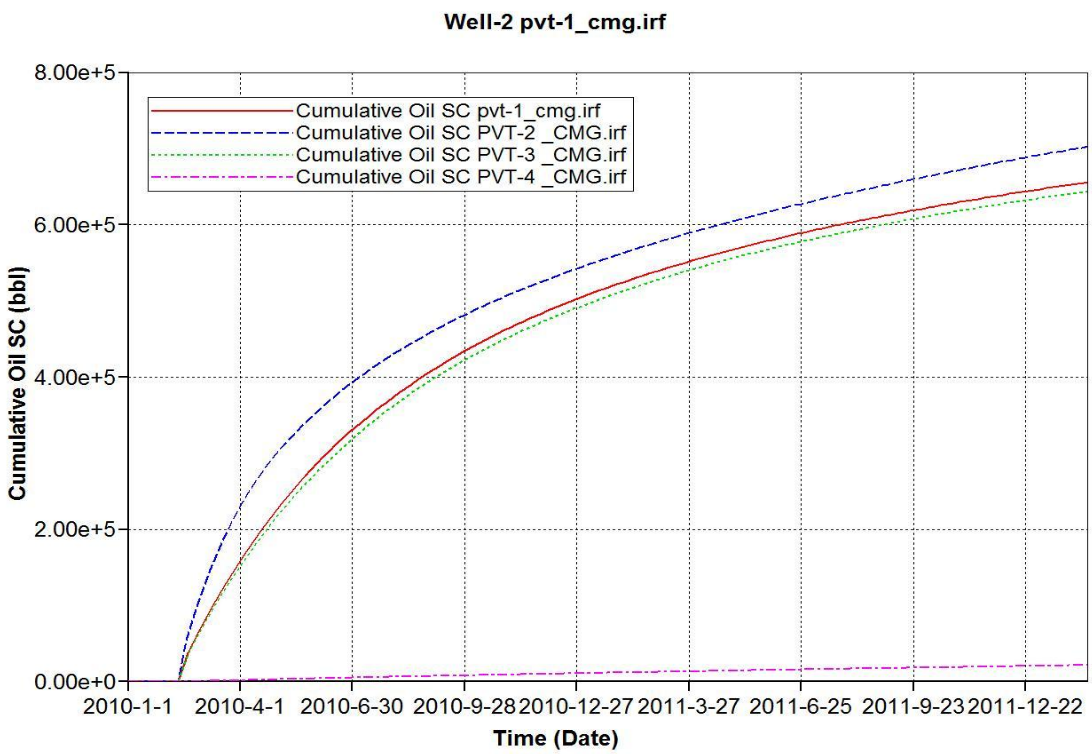


Figure 4-18 Cumulative Production well (Well 2) History of Four Basic PVT Models

It's obvious that the production history of PVT-1 and PVT-3 are very similar as they are typical black oils. PVT-2 exhibits the larger production rate and cumulative production because it is composed of more light components. PVT-4 however shows quite small but constant production rate. Generally speaking, heavy oil with large viscosity is good candidates to thermal recovery processes. Nevertheless, viscous oil without optimized thermal operation and reservoir model construction may not result in economic production rate. The uniform reservoir settings and well operation schedule for the PVT-4 do not stimulate the oil production. But it is still able to differentiate the production from other three PVT models.

The ultimate goal of this study is to create the networks and categorize crude oil to one of the four PVT models. To establish the criteria to the classifications without personal bias, the production difference at certain time period between the crude oil and one of the four PVT models are utilized as the outputs/targets to train the networks.

Artificial Neural Network Outputs Collection:

Initially, outputs were composed of two sorts of production data, which were flow rates at specific time period and cumulative oil production at certain time period. Table 4-5 lists the time schedule for network. The production life was determined for 365 days. One abnormal point can be seen in Figure 4-19. There was a peak production rate normally at the first beginning and then the flow rate started to decrease. The network prediction did not get improved by adjustment of flow rate and cumulative production at different time period.

Table 4-5 Example of Unsuccessful Network Selection of Outputs

Testing Data No 3.	
Network Inputs	Network Outputs
Tc,°F	q@1m
Pc, psia	q@2m
Tcb, °F	q@3m
Pcb, psia	q@6m
Tct, °F	q@9m
Pct, psia	q@1year
MW @ Reservoir Condition	cumulative@3m
Saturation Pressure, psia	cumulative@6m
Compressibility	cumulative@9m
Interfacial Tension, lbf/ft	cumulative@12m
Solution GOR	
Oil FVF	
Viscosity , cp	
Density, lb/ft^3	
Eigenvalue (CP,CB)	
Eigenvalue (CP,CT)	

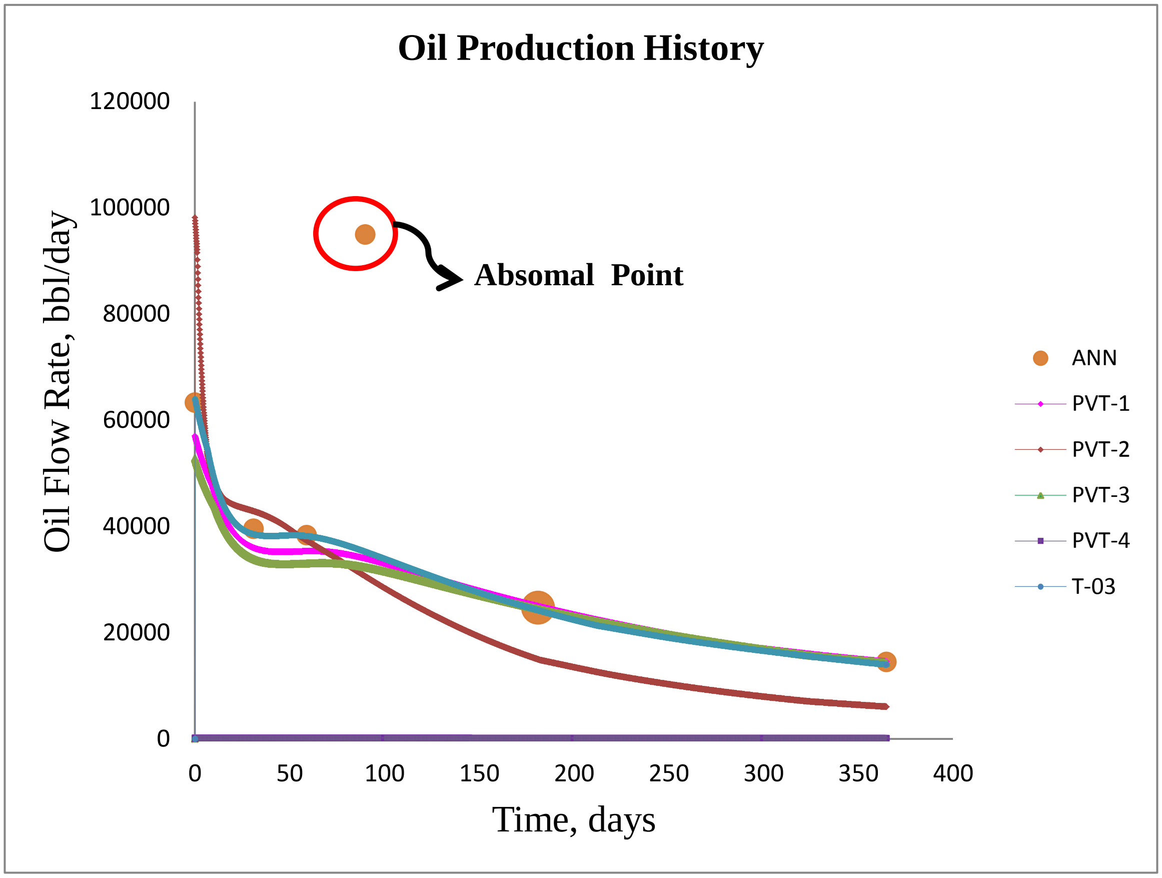


Figure 4-19 Network Production Prediction using Inefficient Inputs and Outputs

The trial finally came out with good network prediction through following changes. The initial reservoir pressure of 2000 psia from literature review was substituted by pressure gradient for sandstone reservoir formation. Production life had been extended to two years from the very beginning of steam injection. Production profile was extracted from the numerical simulation of crude oil under new specified reservoir conditions. The most important change of the network structure is elimination of the cumulative production data from network outputs, which will be discussed in the error analysis part.

Sensitivity of Training Data Sets to Networks:

The Performance of a network is not fixed as every time network training is launched, the training subset picks randomly 51 data sets. As the entry spectrum for each input is slightly changing, the prediction results of the testing subset can be not the same. It is better to train the network several times at a fixed network structure. Then it can tell whether the network is trained properly.

Finally, networks successfully trained using 20 outputs shows in Table 4-6.

Table 4-6 Networks Output Parameters

Output Parameter	q@40(start),q@peak,q@45,q@50, q@55, q@60, q@70, q@80, q@90, q@100, q@120, q@140, q@160, q@180, q@250, q@300, q@400, q@500, q@600, q@770

4.5 Network Architecture Design

MATLAB™ neural network toolbox is applied to train the network. Matlab tool-box '*nftool*' defines as neural network fitting wizard. It allows the users to solve a fitting problem by a two-layer feed forward network with a sigmoid hidden layer and linear output layer in an interface. This method was first employed to test whether Levenberg-Marquardt (*trainlm*) was applicable to four networks in this study. If the weights numbers are fewer than one hundred, *trainlm* may be best suited in an approximation problem with fastest convergence. Network PVT-1 was studied using this fitting tool. From the observation of regression and mean square error, this network structure did not work properly. The neuron number for hidden layer had been tested from 5 up to 50. The network is poorly trained and an example was represented in Figure 4-20.

Increasing hidden layer neuron number resulted in even worse training. Either the number of hidden layers or the training function did not function well.

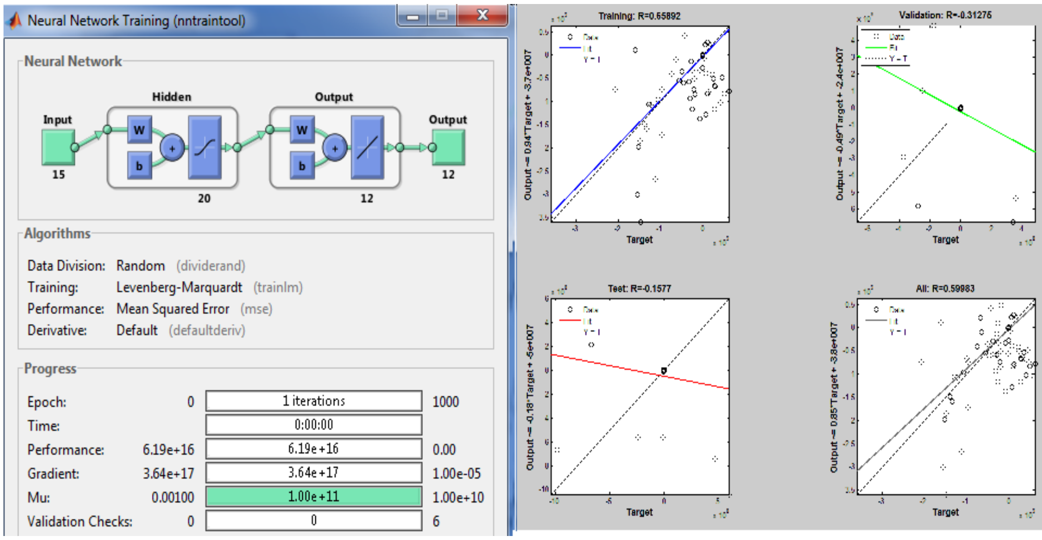


Figure 4-20 Network PVT-1 with One Hidden Layer in Levenberg-Marquardt Algorithm

A network with two hidden layer was then attempted using *trainlm* algorithm to figure out the possible network structure. The network structure of network PVT-1 is expressed in Figure 4-21.

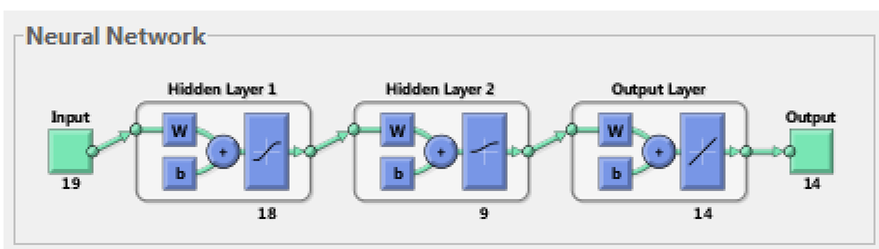


Figure 4-21 Network PVT-1 with Two Hidden Layer in Levenberg-Marquardt Algorithm

Experiments were undertaken for transfer functions of two hidden layers, a most properly trained network delivered the following results with regression for training, validation and testing. (Figure 4-22) The training result indicated that with smaller number of epochs (iterations), the regression value approached to unit value. But it may not guarantee the good testing data set prediction. Using Levenberg-Marquardt (*trainlm*) always has fewer validation checks, always less than 25. The network was not adequately well trained to understand the nonlinear interconnections of inputs and outputs.

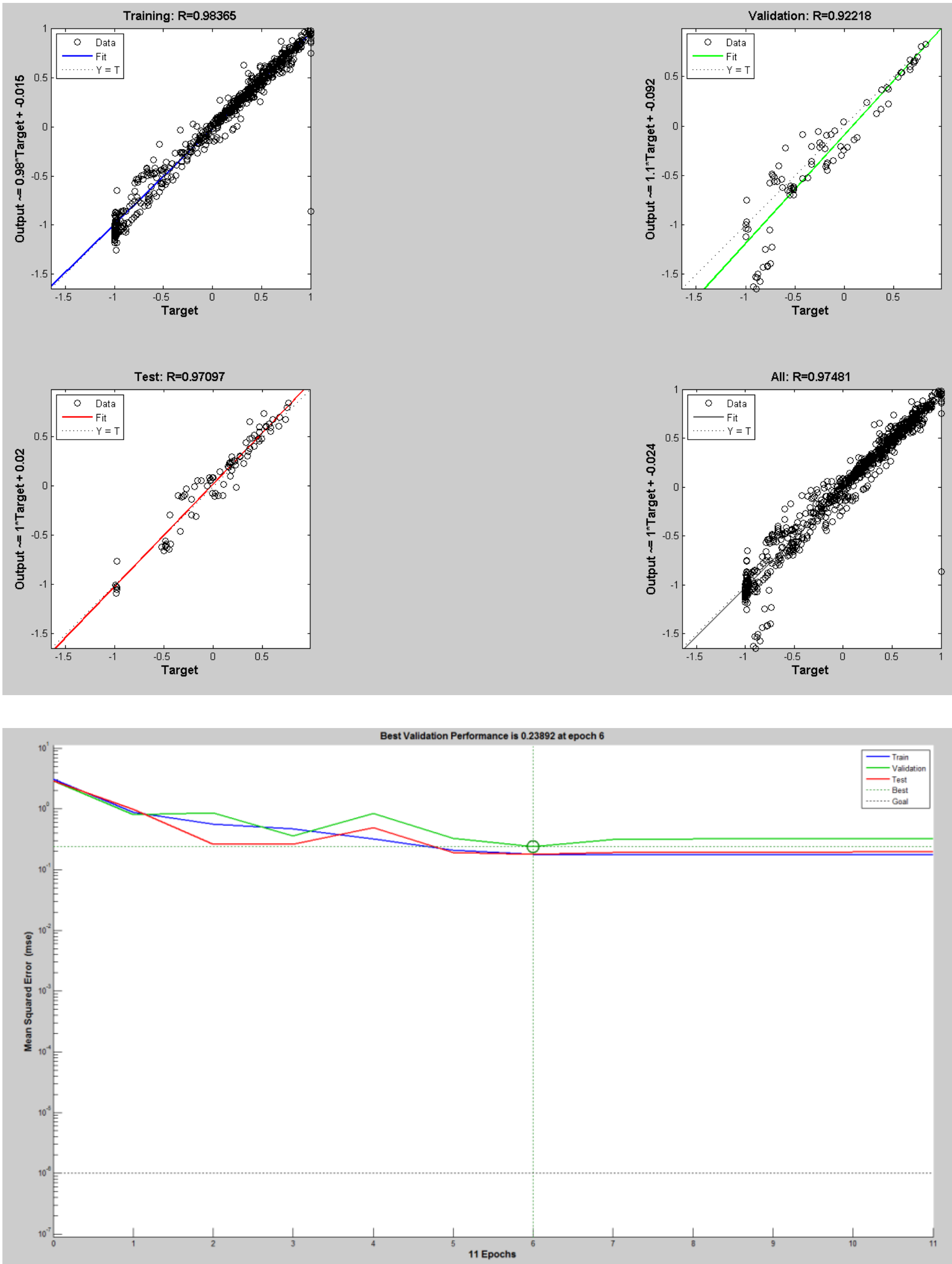


Figure 4-22 Performance of Trained network PVT-1

Two hidden layer structures using Levenberg-Marquardt algorithm (*trainlm*) were proved not able to represent the network. With larger number of data sets, more hidden layers may be an option to testify the possibility of Levenberg-Marquardt algorithm. Nevertheless, only 63 crude oil samples using numerical simulation were employed as training data set, more hidden layers with many neurons can complicate the problem in such a manner.

Scaled conjugate gradient backpropagation (*trainscg*) in general performs well in a wide range of problems with moderate memory required. It became another training function option in this study to be investigated for a two hidden layer neural architecture. The inputs and outputs were kept same as used in *trainlm* algorithm. Figure 4-23 to Figure 4-26 are appropriate finalized network structures after a number of trials of transfer functions.

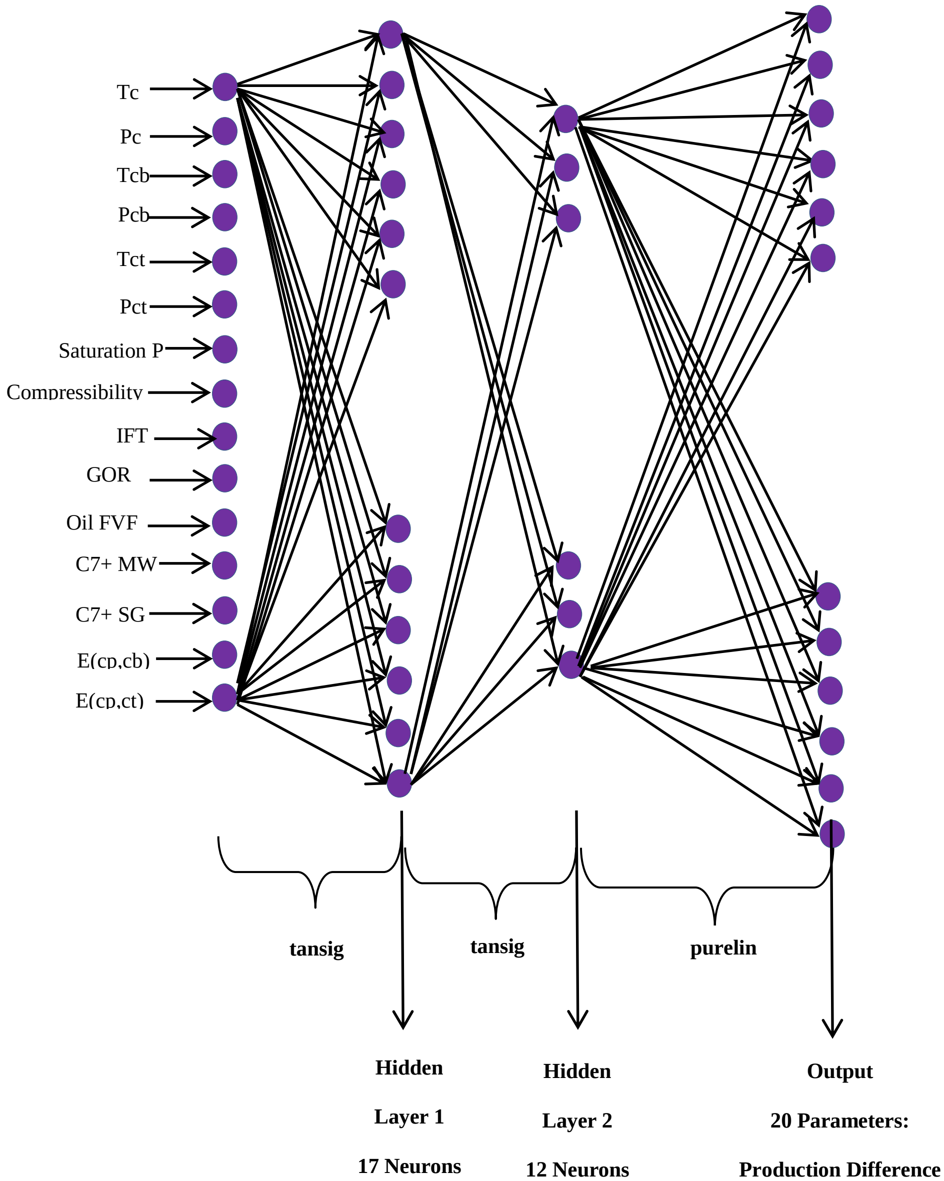


Figure 4-23 Network Architecture for PVT-1 Model

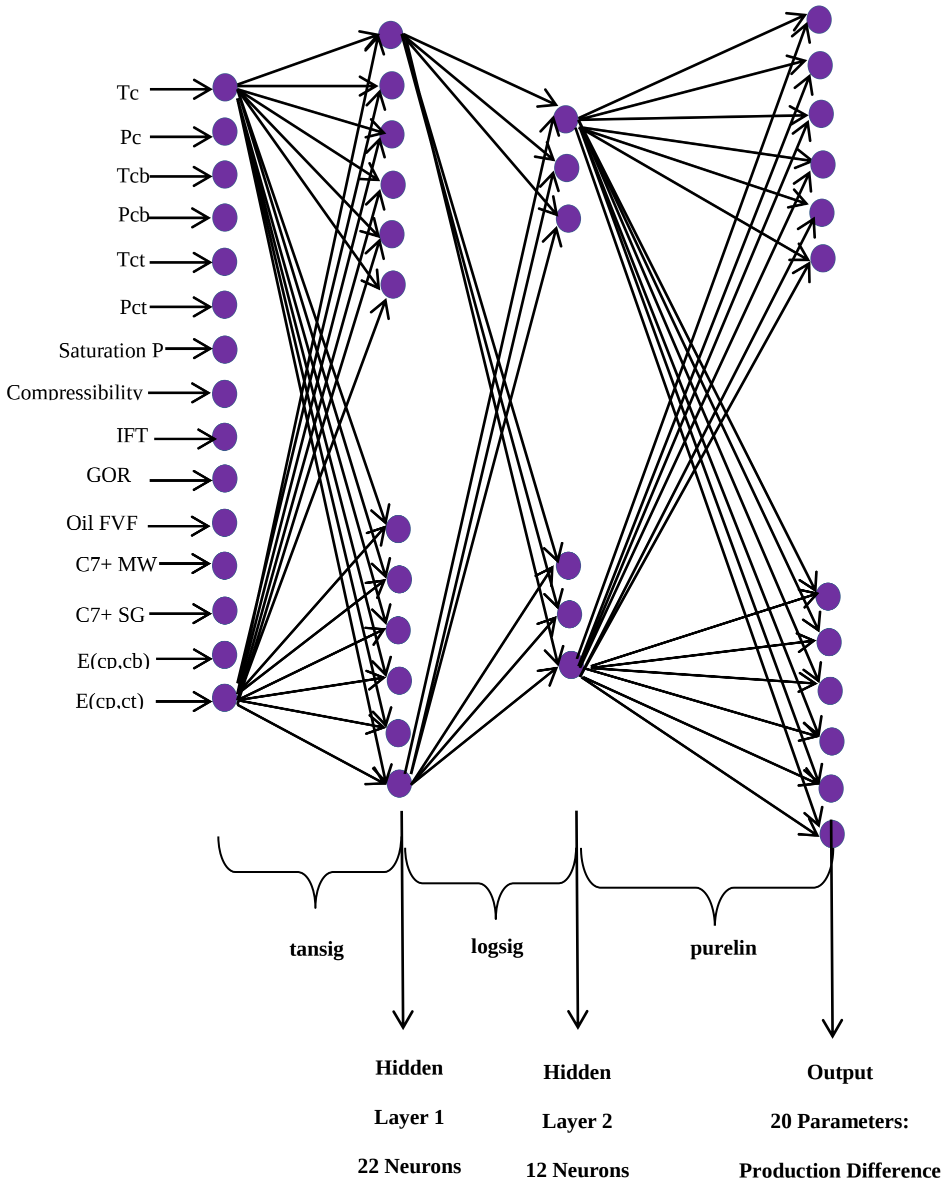


Figure 4-24 Network Architecture for PVT-2 Model

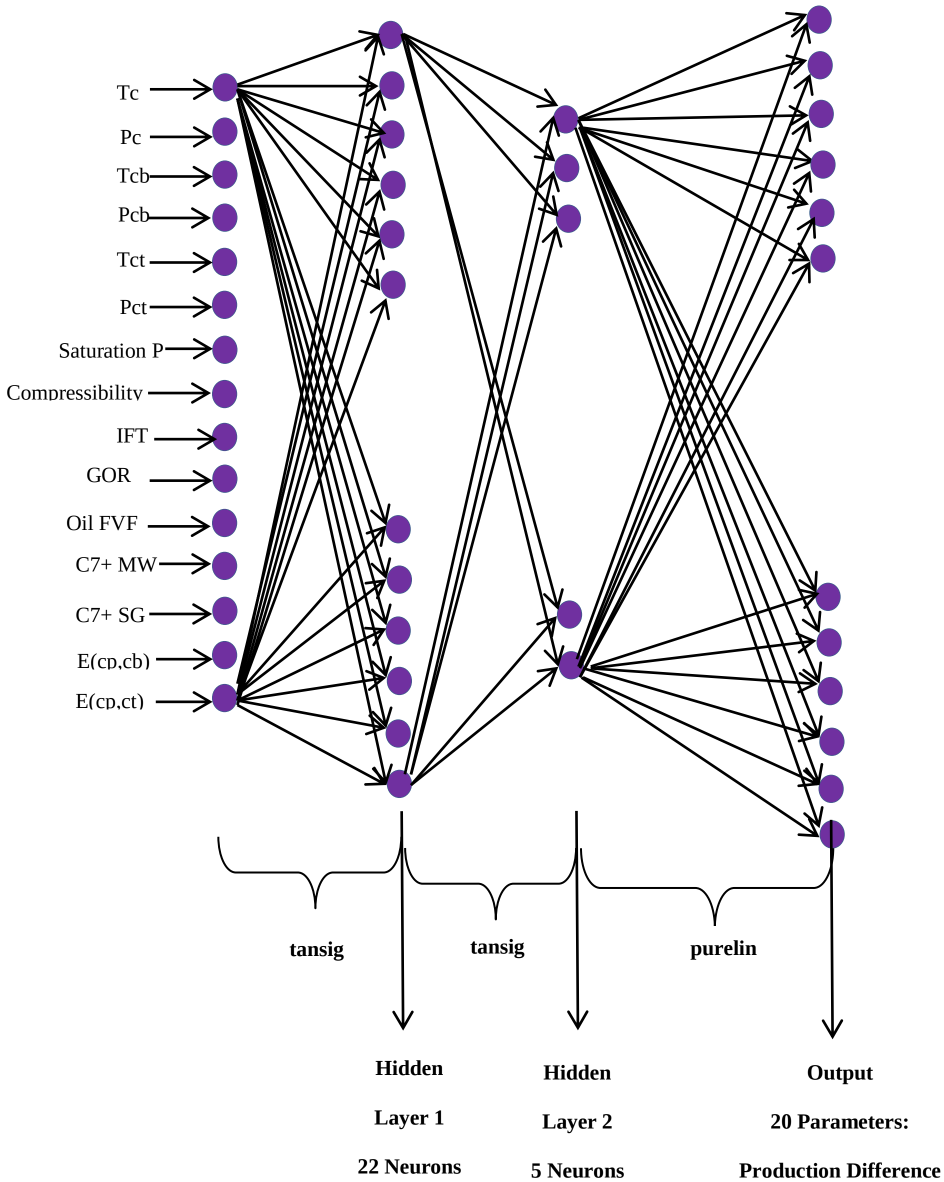


Figure 4-25 Network Architecture for PVT-3 Model

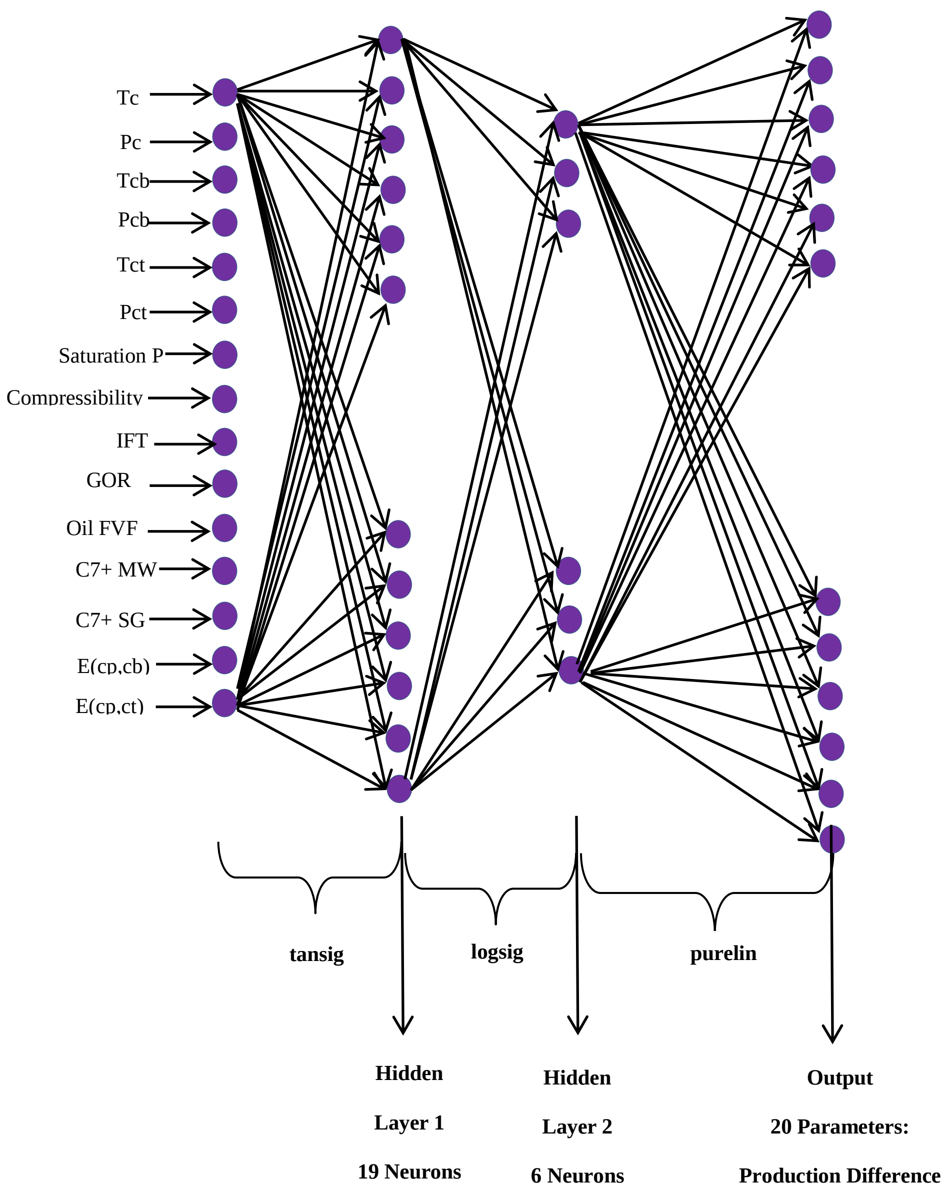


Figure 4-26 Network Architecture for PVT-4 Model

Chapter 5

Results and Discussion

This chapter discusses the identification of a crude oil to one of the PVT fluid types by using artificial neural technology and how the four built neural networks assist in categorization.

The oil production rate differences at specific times are the outputs from networks, which each crude oil have four data sets from four networks. The values of data sets are analyzed with numerical simulation results. After validation checks for each network, the next stage is to work on several case studies including heavy oil. The utilization of ANN in this study is starting from a simplified 3-D, 2-phase homogenous reservoir model with same steam injection scheme to develop a compiled expert system to categorize a specific crude oil to one of the four PVT models. Once the inputs are collected and entered to Graphical User Interface (GUI), the classification is accomplished.

5.1 Performance Analysis of Artificial Neural Networks

5.1.1 Crude Oils Represented by Network PVT-1

Once the network is trained with regression closer to unit such as Figure 5-1 for Network of PVT-1 and the performance plot of the trained network in Figure 5-2, it is the time to move forward to check the accuracy of the network prediction.

Good prediction of the network requires longer time of training which may cause overtraining. When the validation error check is over 5000 epochs or the mean squared error is smaller than 10E6, the training procedure is stopped to avoid overtraining. In network PVT-1, the mean error cannot obtain the desired goal of 10^6, overall performance error is 0.0096. The validation data set inputs and predicted absolute error are calculated and tabulated from Table 5-1 to Table 5-3.

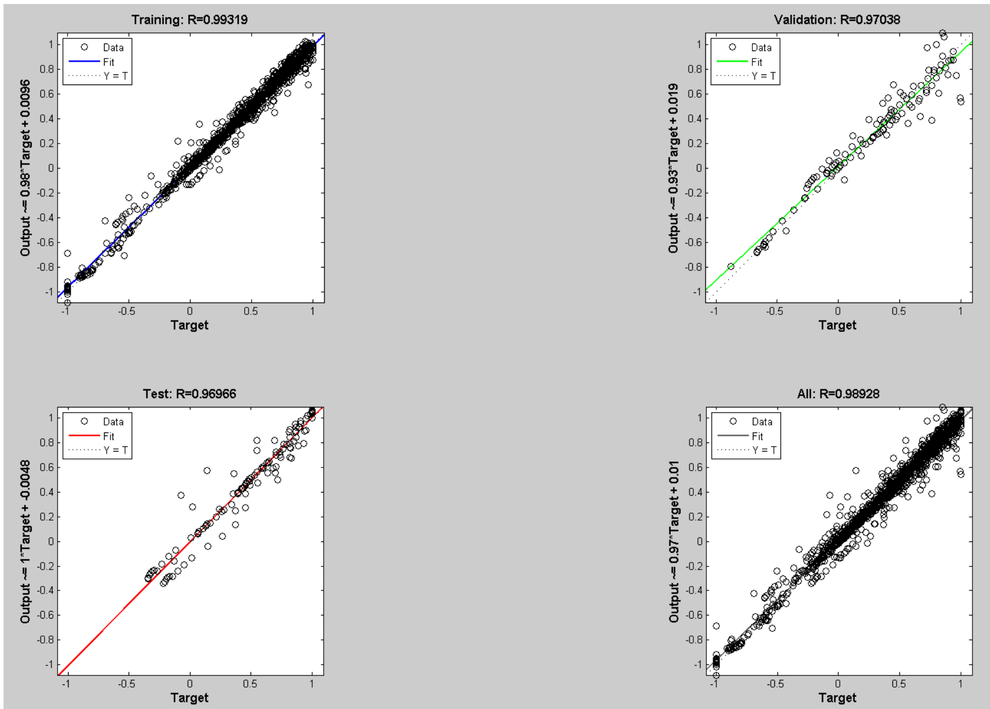


Figure 5-1 Network Regression Values of Network PVT-1

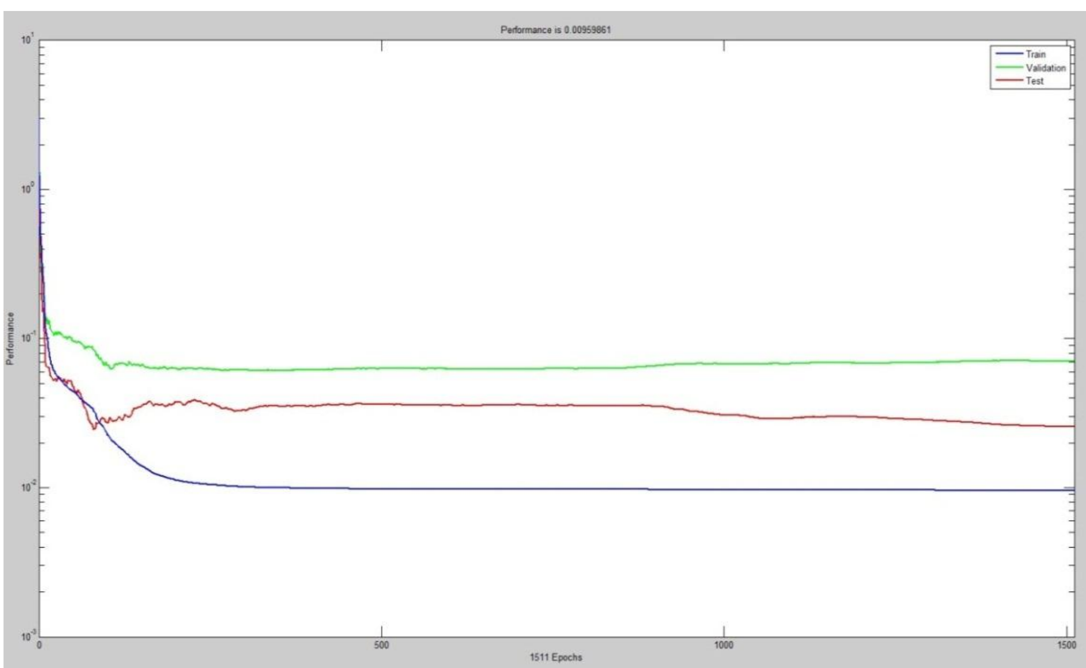


Figure 5-2 Network PVT-1 Performance at 0.0096 and 1151 epochs

Table 5-1 Composition Data for the Validation Sets

No.	3*	11	12	31	44	59
CO_2	0.23	0.75	0.56	0.25	2.09	3.15
N_2	0.51	0.21	0.35	0.88	0.17	0.05
C_1	26.03	12.56	10.9	23.94	31.72	37.45
C_2	7.12	2.59	1.95	11.67	15.41	10.52
C_3	6.43	4.83	1.87	9.36	7.49	7.78
iC_4	1.24	0.86	1	1.39	0.45	2.73
nC_4	3.96	3.83	2.64	4.61	1.81	6.5
iC_5	1.73	2.88	2.05	1.5	2.49	2.41
nC_5	2.62	1.26	2.42	2.48	2.904	3.5
C_6	4.17	5.42	5.23	3.26	2.436	4.61
C_{7+}	45.96	64.81	71.03	40.66	33.03	21.3
C_{7+} MW	285	220	258	196	201	168
C_{7+} s.g.	0.91	0.857	0.8722	0.8494	0.8383	0.8067

*: Worst Case

Table 5-2 Inputs of Validation Sets for Network PVT-1

No.	3*	11	12	31	44	59
Tc (°F)	985.822	863.584	948.481	747.742	716.762	528.522
Pc (psia)	1006.843	601.67	467.65	1284.29	1645.24	2089.32
Tcb (°F)	455.723	508.009	633.314	425.402	386.973	338.05
Pcb (psia)	2590.625	1010.08	703.305	2197.713	2835.968	2659.559
Tct (°F)	998.77	868.132	943.139	771.848	753.763	601.771
Pct (psia)	732.963	524.177	449.399	858.847	969.19	1032.384
Saturation Pressure (psia)	1833.33	799.09	687.03	1871.316	2528.856	2464.022
Compressibility ($psia^{-1}$)	1.26E-05	1.15E-05	9.11E-06	1.47E-05	1.45E-05	2.35E-05
IFT (lbf/ft)	6.20E-03	6.56E-03	5.71E-03	6.78E-03	6.21E-03	5.95E-03
Solution GOR (SCF/STB)	117.2314	88.2114	58.6263	176.5931	157.2086	250.95
Oil FVF	1.1186	1.0955	1.0684	1.1644	1.1493	1.2481
C_{7+} MW	285	220	258	196	201	168
C_{7+} sg	0.91	0.857	0.8722	0.8494	0.8383	0.8067
E(CP,CB)	2082.941	1460.83	1417.44	2019.451	2326.846	2536.963
E(CP,CT)	1852.003	1419.798	1402.072	1873.262	2153.656	2418.101

Table 5-3 Error Analysis for the Six Validation Data Sets Using Network PVT-1

Error in %	3*	11	12	31	44	59	Max	Min
q@40(start day)	12.29%	0.21%	4.85%	0.82%	0.11%	2.11%	12.29%	0.11%
q@MAX	12.14%	0.07%	5.55%	0.75%	0.30%	2.47%	12.14%	0.07%
q@45	11.82%	0.39%	6.97%	0.52%	0.16%	0.10%	11.82%	0.10%
q@50	10.43%	0.23%	7.03%	0.60%	0.19%	2.65%	10.43%	0.19%
q@55	10.06%	0.26%	6.47%	0.78%	0.03%	2.46%	10.06%	0.03%
q@60	10.69%	0.20%	6.14%	1.46%	0.86%	2.74%	10.69%	0.20%
q@70	12.34%	1.33%	3.02%	0.87%	0.36%	2.60%	12.34%	0.36%
q@80	11.66%	1.14%	4.14%	0.64%	0.49%	1.78%	11.66%	0.49%
q@90	11.62%	1.41%	4.77%	0.34%	0.43%	2.49%	11.62%	0.34%
q@100	11.20%	2.65%	4.31%	0.11%	0.48%	1.99%	11.20%	0.11%
q@120	12.20%	3.67%	1.82%	0.14%	1.00%	3.04%	12.20%	0.14%
q@140	13.08%	3.92%	0.16%	0.63%	1.14%	4.06%	13.08%	0.16%
q@160	12.27%	3.75%	0.72%	0.89%	1.24%	3.83%	12.27%	0.72%
q@180	8.68%	4.19%	1.77%	1.49%	1.09%	2.51%	8.68%	1.09%
q@250	8.00%	3.82%	0.22%	1.56%	1.52%	1.91%	8.00%	0.22%
q@300	4.89%	4.06%	0.04%	1.56%	1.63%	0.91%	4.89%	0.04%
q@400	1.34%	3.29%	0.48%	1.58%	1.86%	0.16%	3.29%	0.16%
q@500	1.59%	2.47%	0.61%	1.13%	2.25%	0.59%	2.47%	0.59%
q@600	4.28%	2.02%	3.05%	0.59%	3.18%	0.09%	4.28%	0.09%
q@770	5.60%	0.36%	5.90%	0.32%	3.38%	1.04%	5.90%	0.32%
Max	13.08%	4.19%	7.03%	1.58%	3.38%	4.06%		
Min	1.34%	0.07%	0.04%	0.11%	0.03%	0.09%		

Figure 5-3 illustrates prediction results by using one of the validation data sets. Training, validation and testing data sets are randomly divided. The numbers of these six data sets are 3rd, 11th, 12th, 31st, 44th and 59th. Comparisons of simulation results and ANN predictions for other five data sets are represented in Appendix A in the form of plots.

Error analysis for network performance is available through several methods. One way is simply comparing the flow rate differences generated by the simulator and ANN predicted oil flow rate differences. Sometimes, the oil rate difference error can be misleading as the distinguished magnitude of rate difference and actual flow rate values. This will lead to an analysis misunderstanding problem when the flow rate of a new testing crude oil is close to one of the four PVT models. Taking an extreme example, if the simulator generates flow rate difference at a certain time period is 20 bbl/day while ANN predicts 40 bbl/day. It gives out the error result at 100% and looks like the ANN forecasting performance is bad. However, if the flow rate different is restored to the actual oil flow rate. The magnitude of the flow result in the good predications of the network and it is very likely that PVT model represents that new reservoir fluid best among the four.

Another method of error analysis is to consider the cumulative oil production of CMG run and ANN prediction as described in the following equation.

$$error\_Cum. = \frac{abs(Cum._{CMG} - Cmu._{ANN})}{Cum._{CMG}} \times 100\%$$

In the oil rate production history, the area in between the flow rate curve and x-axis using the integral calculation represent the cumulative production. Again, this analyzed method is considered not able to give the most consistent analytical results. First of all, the areas of two curves may be cancelled out like the situation shown in Figure 5-4.

Considering the goal of the research, error analysis is applied to compare the resumed flow rate as following equation and Table 5-3 compiles the error of oil rate at each specified time period.

$$error\_q_{time} = \frac{abs(q_{t,CMG} - q_{t,ANN})}{q_{t.CMG}} \times 100\%$$

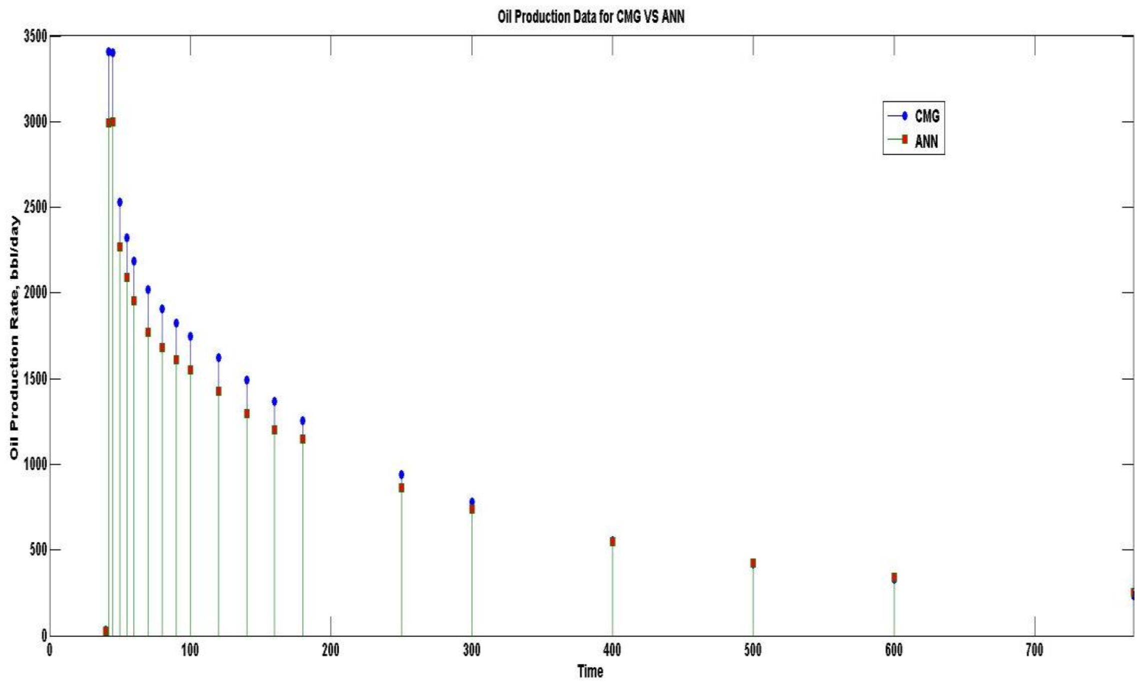


Figure 5-3 Validation Data #3 Prediction Performance of Network PVT-1

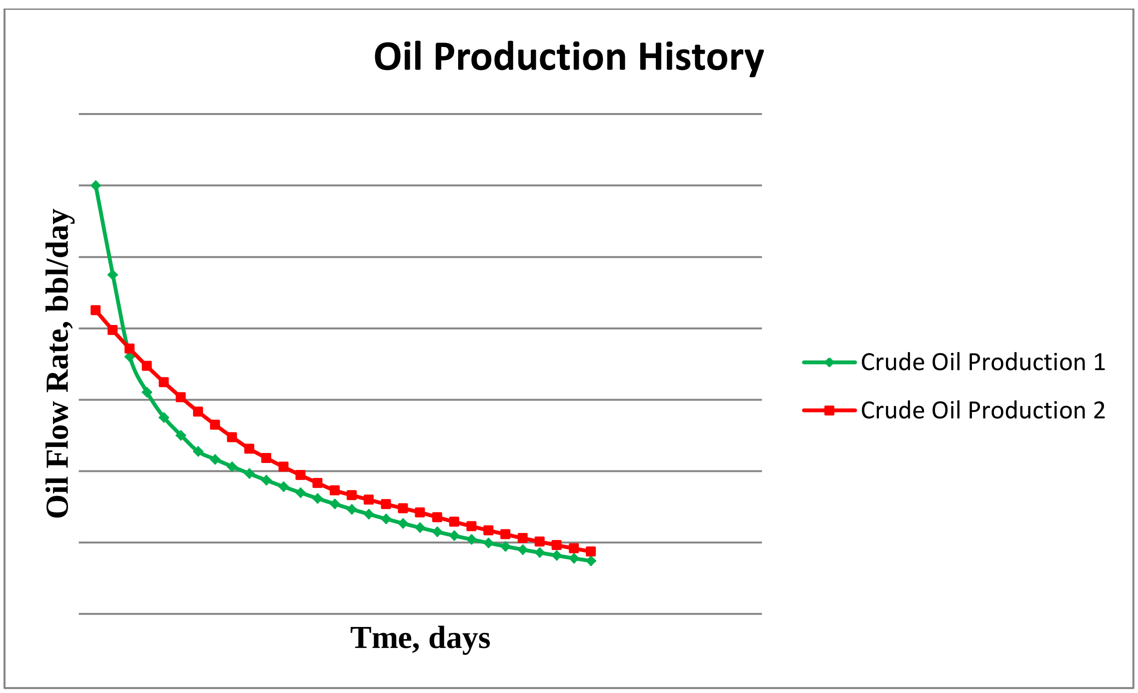


Figure 5-4 A Possible Infeasible Method by Cumulative Error Analysis

The evaluated performances of the network presents that among all six validation data set (represent six crude oils), the maximum error in terms of the flow rate is not fixed at a certain time period. The third data set has a maximum error of 13.08% at time period of 140 days, and minimum error of 1.34% at 400 days. The maximum error of other five data sets is decreased to 1.58%-7.03%. The minimum error is within 0.03% to 0.11%, which indicates the good network performance.

The simulation and ANN predictions plots show that the ANN values are not always larger than simulation results. It indicates that the input parameters (network training data set) cover a reasonable spectrum of the values and the validations parameters are not touching the limits or extreme cases. As the production is approaching the end of the specified time period, the flow rate from simulation and ANN almost overlaps on each other. The error for each time period is showed Table 5-3. It is clear to see that before the time of 180 days, the error at each time period is relatively large. Usually, they are found to be 10.06% and 13.08% and maximum analyzed absolutely error comes from 3rd data set.

The technique to avoid overfitting is applied to train the network approaching to the global minimum. Testing data sets participate in the network learning. There are 6 data sets are divided into the testing group. It illustrates a range of the predicted error values for oil flow rate at specified time period. The results are tabulated in Table 5-4. The errors from validation and testing data sets are relatively close as to prove that the network is trained properly and the error ranges of the network are predictable as well.

Table 5-4 Network PVT-1 Prediction Errors for Testing Data Sets

	5	13*	14	25	28	38	max	min
q@40(start day)	2.18%	10.49%	5.08%	0.12%	2.75%	0.04%	10.49%	0.04%
q@MAX	2.34%	10.51%	5.25%	0.12%	1.72%	0.06%	10.51%	0.06%
q@45	2.27%	12.19%*	6.11%	0.27%	0.78%	0.50%	12.19%	0.27%
q@50	2.62%	12.05%	5.28%	0.20%	5.66%	0.20%	12.05%	0.20%
q@55	2.92%	11.45%	5.06%	0.46%	1.22%	0.08%	11.45%	0.08%
q@60	2.92%	11.80%	5.40%	1.10%	2.81%	0.42%	11.80%	0.42%
q@70	3.25%	9.57%	4.09%	0.63%	0.23%	0.45%	9.57%	0.23%
q@80	3.63%	10.87%	4.42%	0.61%	0.24%	0.32%	10.87%	0.24%
q@90	4.23%	11.25%	4.52%	0.42%	0.41%	0.91%	11.25%	0.41%
q@100	4.11%	10.80%	3.59%	0.05%	2.54%	1.88%	10.80%	0.05%
q@120	5.09%	10.05%	2.97%	0.07%	3.30%	2.72%	10.05%	0.07%
q@140	4.66%	8.83%	2.55%	0.15%	3.03%	3.07%	8.83%	0.15%
q@160	4.63%	8.75%	3.00%	0.15%	2.41%	2.90%	8.75%	0.15%
q@180	3.87%	11.12%	2.49%	1.35%	0.30%	5.10%	11.12%	0.30%
q@250	2.05%	9.18%	2.08%	1.32%	3.96%	4.67%	9.18%	1.32%
q@300	0.81%	7.60%	0.41%	1.77%	4.53%	6.28%	7.60%	0.41%
q@400	0.36%	4.90%	0.13%	1.64%	6.25%	5.16%	6.25%	0.13%
q@500	1.00%	3.49%	0.07%	1.48%	7.07%	3.71%	7.07%	0.07%
q@600	1.80%	2.75%	0.67%	0.76%	6.79%	2.42%	6.79%	0.67%
q@770	0.64%	0.46%	1.10%	0.02%	4.49%	1.19%	4.49%	0.02%
max	5.09%	12.19%	6.11%	1.77%	7.07%	6.28%		
min	0.36%	0.46%	0.07%	0.02%	0.23%	0.04%		

*: Worst Case

Comparing the testing data input parameters with the network inputs spectrum is as below in Table 5-5. The inputs values of these six data sets do not close to the limits. Both testing and validation data sets network inputs distribution plots are demonstrated in Appendix B.

Table 5-5 Network PVT-1 Inputs Range with Six Testing Data Sets Comparison

No.	5	13*	14	25	28	38	Testing Set Max.	Testing Set Min.	Network Max.	Network Min.
Tc (°F)	680.943	924.355	886.993	759.257	995.049	858.756	995.049	680.943	1124.895	344.574
Pc (psia)	2548	445.341	552.81	1281.01	895.48	544.75	2548	445.341	4281.37	344.7
Tcb (°F)	308.072	487.645	428.905	415.635	338.033	353.64	487.645	308.072	633.314	179.608
Pcb (psia)	4318.681	785.243	1081.866	2299.222	3176.08	1312.239	4318.681	785.243	5802.718	537.083
Tct (°F)	753.667	924.837	889.671	784.609	999.116	860.293	999.116	753.667	1131.968	528.517
Pct (psia)	1168.517	438.075	497.781	836.842	752.61	550.723	1168.517	438.075	1350.588	384.817
Saturation Pressure (psia)	4211.416	653.93	957.03	1992.596	3063.02	1234.422	4211.416	653.93	5784.938	349.688
Compressibility (psia-1)	1.32E-05	9.71E-06	1.02E-05	1.47E-05	6.36E-06	1.05E-05	1.47E-05	6.36E-06	2.77E-05	4.72E-06
IFT (lbf/ft)	6.05E-03	5.51E-03	6.53E-03	5.87E-03	8.07E-03	6.15E-03	0.008073	0.005506	0.008167	0.002784
Solution GOR (SCF/STB)	138.3813	57.2011	60.7121	157.1952	45.245	50.9576	157.1952	45.245	320.12	35.5003
Oil FVF	1.1382	1.0689	1.0738	1.1519	1.0518	1.0664	1.1519	1.0518	1.3256	1.0429
C_{7+} MW	215	250	225	210	255	217.5	255	210	368.9	160
C_{7+} sg	0.8479	0.858	0.86	0.8399	0.9165	0.8351	0.9165	0.8351	0.9594	0.7978
E(CP,CB)	3098.365	1348.387	1421.284	2031.911	1982.616	1400.831	3098.365	1348.387	4539.927	1130.375
E(CP,CT)	2938.159	1364.93	1406.036	1871.397	1813.84	1407.751	2938.159	1364.93	4508.905	1217.818

Curve fitting and two-phase envelopes are two other methods to better interpret the ANN prediction performance. Hyperbolic equation can best present production history using the equation:

$$q = \frac{a}{(1 + b * t)^c}$$

Coefficients of a, b, c and R^2 are recorded for both simulation and ANN predictions. The relative error of three coefficients a, b and c are not able to be a new protocol to identify the categorization of crude oils.

Phase envelope of testing and validation data sets is studied for four development networks as well to see if a crude oil with largest ANN predicted error shares the least overlapping area to the corresponding base reservoir model. The created two-phase envelope provides a general direction of crude oil thermodynamic characteristics but it is not a determinant guideline.

5.1.2 Crude Oils Represented by Network PVT-2

PVT-2 represents a volatile reservoir fluid which has a slightly higher molar fraction of lighter components. Table 5-6 collects validation data set network input entries. Figure 5-5 displays the network prediction performance of one validation example. The mean squared error in this network is 0.0197686 and the validation check ends at 603 epochs. $5^{th,}$ 13^{th}, 17^{th}, 25^{th}, 41^{st} and 48^{th} in this network are the validation data sets. Again, this is a randomly distributed action taking up 20% of complete training data sets. The relative flow rate error is compiled in Table 5-7

Table 5-6 Inputs of Validation Data Sets for Network PVT-2

No.	5	13*	17	25	41	47
Tc (°F)	680.943	924.355	828.976	759.257	717.66	716.691
Pc (psia)	2548	445.341	1304.65	1281.01	1551.54	1432.97
Tcb (°F)	308.072	487.645	379.264	415.635	386.241	404.11
Pcb (psia)	4318.681	785.243	2951.568	2299.222	2686.765	2384.35
Tct (°F)	753.667	924.837	846.505	784.609	752.068	749.414
Pct (psia)	1168.517	438.075	903.646	836.842	934.53	885.159
Saturation Pressure (psia)	4211.416	653.93	2711.813	1992.596	2407.34	2084.536
Compressibility ($psia^{-1}$)	1.32E-05	9.71E-06	1.06E-05	1.47E-05	1.52E-05	1.62E-05
IFT (lbf/ft)	6.05E-03	5.51E-03	7.10E-03	5.87E-03	6.05E-03	6.00E-03
Solution GOR (SCF/STB)	138.3813	57.2011	100.2	157.1952	163.5677	179.3157
Oil FVF	1.1382	1.0689	1.1005	1.1519	1.1576	1.1734
C_{7+} MW	215	250	217	210	201	200
C_{7+} sg	0.8479	0.858	0.868	0.8399	0.835	0.835
E(CP,CB)	3098.365	1348.387	2151.244	2031.911	2235.317	2119.725
E(CP,CT)	2938.159	1364.93	1973.183	1871.397	2070.906	1964.558

*: Worst Case

Table 5-7 Error Analysis for Six Validation Data Sets Using Network PVT-2

Error in %	5	13*	17	25	41	47	Max	Min
q@40(start day)	3.06%	6.36%	4.03%	1.36%	1.38%	2.97%	6.36%	1.36%
q@MAX	1.77%	5.72%	3.15%	1.09%	0.77%	2.05%	5.72%	0.77%
q@45	2.77%	6.26%	3.21%	1.41%	0.84%	2.26%	6.26%	0.84%
q@50	3.15%	5.37%	3.77%	1.46%	1.27%	2.48%	5.37%	1.27%
q@55	2.86%	6.05%	3.91%	1.95%	1.32%	2.91%	6.05%	1.32%
q@60	2.39%	6.83%	4.28%	1.14%	0.80%	2.63%	6.83%	0.80%
q@70	2.41%	7.41%	4.18%	1.25%	0.65%	2.60%	7.41%	0.65%
q@80	1.62%	5.51%	3.67%	0.78%	0.24%	1.49%	5.51%	0.24%
q@90	2.33%	4.92%	3.74%	1.68%	0.61%	2.67%	4.92%	0.61%
q@100	1.59%	5.98%	3.38%	0.85%	0.47%	1.28%	5.98%	0.47%
q@120	0.75%	2.60%	3.74%	1.45%	0.06%	1.95%	3.74%	0.06%
q@140	1.26%	1.84%	3.37%	1.50%	0.24%	1.63%	3.37%	0.24%
q@160	1.29%	2.90%	3.25%	0.98%	0.71%	1.20%	3.25%	0.71%
q@180	1.70%	1.66%	3.39%	0.95%	0.56%	1.55%	3.39%	0.56%
q@250	1.74%	2.46%	2.50%	0.73%	0.01%	0.93%	2.50%	0.01%
q@300	2.39%	1.17%	2.94%	0.50%	0.40%	0.35%	2.94%	0.35%
q@400	1.03%	0.68%	1.23%	1.14%	0.61%	0.45%	1.23%	0.45%
q@500	0.42%	1.18%	0.70%	1.66%	0.28%	0.48%	1.66%	0.28%
q@600	0.12%	1.95%	0.25%	1.45%	0.06%	1.17%	1.95%	0.06%
q@770	1.73%	2.25%	1.89%	1.71%	0.47%	0.93%	2.25%	0.47%
Max	3.15%	7.41%	4.28%	1.95%	1.38%	2.97%		
Min	0.12%	0.68%	0.25%	0.50%	0.01%	0.35%		

*: Worst Case

The validation data sets tell a well-trained network as the maximum error 7.41%. The minimum error is dropped to 0.01%. For sure, the early time and model time flow rate are more valuable than the late time prediction. This network requires more time-consuming experiments. Apart from several base training parameters, the learning rate and ratio to decrease learning rate are incorporated to assist in improving network accuracy. If the network successfully predicts oil rate at early times, the middle time and the overall relative error for that crude oil are low as well.

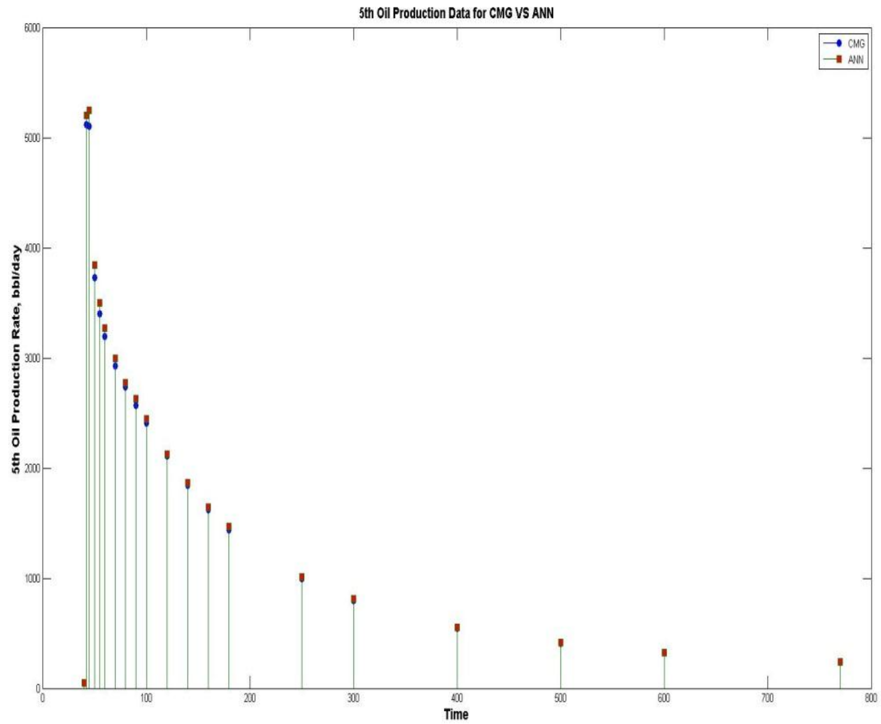


Figure 5-5 Validation Data #5 Prediction Performance of Network PVT-2

The testing data set are studied to verify the network in Table 5-8. The initial guess of PVT-2 network performance was not as good as the other three because the peak production rate nearly doubles the value of PVT-1 and production of PVT-2 overlapped in late time, which indicates a dramatic decline in a very short time. However, the relative error percentages values prove that with a well-trained network, the complicated and implicit relationship of nonlinear inputs and outputs can be properly understand.

Table 5-8 Network PVT-2 Prediction Errors for Testing Data Sets

	2	8	20	37	42	62*	max	min
q@40(start day)	4.92%	1.40%	4.43%	9.77%	1.70%	10.47%	10.47%	1.40%
q@MAX	3.76%	1.77%	4.32%	8.46%	2.95%	10.41%	10.41%	1.77%
q@45	5.58%	2.29%	7.17%	8.92%	2.51%	10.40%	10.40%	2.29%
q@50	5.13%	1.62%	2.92%	9.95%	0.86%	7.16%	9.95%	0.86%
q@55	5.51%	2.05%	3.95%	9.18%	0.88%	9.48%	9.48%	0.88%
q@60	5.36%	1.11%	2.33%	9.30%	2.23%	10.51%	10.51%	1.11%
q@70	5.92%	1.45%	4.05%	9.13%	2.43%	9.92%	9.92%	1.45%
q@80	6.49%	1.68%	5.67%	8.79%	1.48%	6.74%	8.79%	1.48%
q@90	6.35%	2.11%	6.58%	9.16%	0.36%	6.83%	9.16%	0.36%
q@100	7.05%	1.78%	7.81%	8.87%	0.02%	8.85%	8.87%	0.02%
q@120	6.21%	1.54%	2.73%	7.28%	1.60%	6.82%	7.28%	1.54%
q@140	6.19%	1.64%	3.49%	6.69%	1.57%	4.68%	6.69%	1.57%
q@160	6.07%	1.51%	2.58%	6.49%	1.41%	3.04%	6.49%	1.41%
q@180	6.10%	1.48%	4.47%	6.50%	2.79%	0.60%	6.50%	0.60%
q@250	3.61%	1.26%	5.94%	4.82%	2.92%	3.22%	5.94%	1.26%
q@300	3.47%	0.22%	5.60%	5.44%	4.72%	4.66%	5.60%	0.22%
q@400	1.65%	0.04%	2.09%	3.57%	3.00%	4.39%	4.39%	0.04%
q@500	0.19%	0.43%	0.08%	2.74%	0.87%	3.35%	3.35%	0.08%
q@600	0.87%	1.02%	0.95%	2.12%	0.94%	3.26%	3.26%	0.87%
q@770	2.24%	1.49%	2.63%	0.48%	4.50%	2.84%	4.50%	0.48%
max	7.05%	2.29%	7.81%	9.95%	4.72%	10.51%		
min	0.19%	0.04%	0.08%	0.48%	0.02%	0.60%		

*: Worst Case

Table 5-9 Network PVT-2 Crude Oil Properties Range for Training, Testing and Validation

	Training		Testing		Validation	
	max	min	max	min	max	min
Tc(°F)	1124.895	432.914	902.388	344.574	924.355	680.943
Pc (psia)	4281.37	344.7	3120.89	392.35	2548	445.341
Tcb (°F)	633.314	179.608	611.044	280.404	487.645	308.072
Pcb (psia)	5802.718	632.187	3212.821	537.083	4318.681	785.243
Tct (°F)	1131.968	546.59	902.891	528.517	924.837	749.414
Pct (psia)	1350.588	420.606	1102.901	384.817	1168.517	438.075
Saturation Pressure (psia)	5784.938	576.22	3132.452	349.688	4211.416	653.93
Compressibility ($psia^{-1}$)	2.77E-05	4.72E-06	2.24E-05	1.03E-05	1.62E-05	9.71E-06
IFT (lbf/ft)	0.00817	0.00278	0.00758	0.00537	0.0071	0.00551
Solution GOR (SCF/STB)	320.12	35.5003	313.9514	51.0338	179.3157	57.2011
Oil FVF	1.3256	1.0429	1.292	1.0665	1.1734	1.0689
C_{7+} MW	368.9	160	245	171	250	200
C_{7+} sg	0.9594	0.7978	0.86	0.801	0.868	0.835
E(CP,CB)	4539.927	1130.375	3414.359	1274.438	3098.365	1348.387
E(CP,CT)	4508.905	1217.818	3316.994	1289.617	2938.159	1364.93

5.1.3 Crude Oils Represented by Network PVT-3

PVT-3 is another typical black oil reservoir. Figure 5-6 displays the network training condition and prediction performance of a validation example. The mean squared error in this network is 0.0272088 and the validation check ends at 1361 epochs. 12^{th}, 15^{th}, 37^{th}, 40^{th}, 53^{rd} and 61^{st} in this network are the validation data sets. Table 5-10 and Table 5-11 are the composition molar fraction and their input parameters. The summary of flow rate prediction performance is showed in Table 5-12.

Table 5-10 Composition Data for the Validation Sets

No.	12*	15	37	40	53	61
CO_2	0.56	2.18	0.1	0.84	0.22	0.9
N_2	0.35	1.67	0.139	0.96	0.054	0.3
C_1	10.9	60.51	34.208	12.92	53.45	43.47
C_2	1.95	7.52	4.304	2.74	6.36	11.46
C_3	1.87	4.74	3.486	4.82	4.66	8.79
iC_4	1	1.52	2.633	1.49	2.78	5
nC_4	2.64	2.6	0.01	5.04	1.01	4.56
iC_5	2.05	1.97	4.875	2.01	1.44	5
nC_5	2.42	1	0.01	3.33	1.3	2.09
C_6	5.23	1.38	3.771	5.07	2.116	1.51
C_{7+}	71.03	14.91	46.464	60.78	26.61	16.92
C_{7+} MW	258	181	202	251	215.9	173
C_{7+} s.g.	0.8722	0.799	0.86	0.8783	0.853	0.8364

*: Worst Case

Table 5-11 Inputs of Validation Data Sets for Network PVT-3

No.	12*	15	37	40	53	61
Tc (°F)	948.481	469.688	801.389	927.355	734.158	494.5
Pc (psia)	467.65	3680.29	1181.18	632.4	2322.38	2728.34
Tcb (°F)	633.314	227.599	367.209	515.732	282.403	332.609
Pcb (psia)	703.305	4765.515	2575.149	1213.11	4703.994	3238.12
Tct (°F)	943.139	605.464	815.828	933.235	778.12	610.299
Pct (psia)	449.399	1307.756	857.939	532.393	1189.143	1136.004
Saturation Pressure (psia)	687.03	4761.397	2395.184	962.155	4652.822	3060.367
Compressibility ($psia^{-1}$)	9.11E-06	1.71E-05	1.17E-05	9.94E-06	1.16E-05	2.00E-05
IFT (lbf/ft)	5.71E-03	5.84E-03	7.58E-03	6.08E-03	6.49E-03	6.80E-03
Solution GOR (SCF/STB)	58.6263	138.0833	91.4305	81.7476	97.745	260.93
Oil FVF	1.0684	1.1481	1.0997	1.0875	1.1035	1.2491
C_{7+} MW	258	181	202	251	215.9	173
C_{7+} sg	0.8722	0.799	0.86	0.8783	0.853	0.8364
E(CP,CB)	1417.44	3988.524	1982.081	1584.483	2927.924	3136.064
E(CP,CT)	1402.072	3910.416	1849.184	1500.015	2775.63	3004.55

Table 5-12 Error Analysis for Six Validation Data Sets Using Network PVT-3

Error in %	12*	15	37	40	53	61	Max	Min
q@40(start day)	8.89%	3.31%	4.86%	1.53%	9.25%	3.30%	9.25%	1.53%
q@MAX	10.09%	4.45%	3.92%	2.55%	10.13%	2.78%	10.13%	2.55%
q@45	9.77%	5.11%	3.37%	2.59%	9.56%	1.64%	9.77%	1.64%
q@50	12.19%	5.75%	3.71%	3.80%	9.15%	3.23%	12.19%	3.23%
q@55	11.55%	5.85%	3.78%	3.33%	9.64%	3.98%	11.55%	3.33%
q@60	9.86%	4.97%	3.50%	2.44%	8.19%	3.89%	9.86%	2.44%
q@70	10.56%	5.58%	4.19%	2.83%	8.91%	3.38%	10.56%	2.83%
q@80	11.18%	5.36%	6.38%	2.82%	9.91%	4.53%	11.18%	2.82%
q@90	10.24%	6.23%	3.89%	2.98%	9.50%	3.81%	10.24%	2.98%
q@100	8.32%	5.33%	5.16%	1.42%	8.44%	3.69%	8.44%	1.42%
q@120	6.99%	5.59%	4.47%	0.72%	7.59%	5.65%	7.59%	0.72%
q@140	3.25%	5.28%	3.53%	1.13%	5.76%	6.72%	6.72%	1.13%
q@160	0.59%	4.77%	2.49%	1.98%	4.38%	6.59%	6.59%	0.59%
q@180	0.05%	5.08%	3.70%	2.08%	5.45%	6.70%	6.70%	0.05%
q@250	2.03%	4.48%	0.55%	2.30%	2.62%	0.99%	4.48%	0.55%
q@300	3.86%	3.55%	0.85%	3.48%	1.61%	0.96%	3.86%	0.85%
q@400	4.94%	1.85%	0.21%	3.30%	0.25%	0.14%	4.94%	0.14%
q@500	4.72%	1.02%	0.43%	2.52%	0.80%	1.45%	4.72%	0.43%
q@600	2.88%	0.28%	0.40%	2.42%	0.41%	1.08%	2.88%	0.28%
q@770	0.51%	0.70%	0.49%	1.92%	0.14%	1.45%	1.92%	0.14%
Max	12.19%	6.23%	6.38%	3.80%	10.13%	6.72%		
Min	0.05%	0.28%	0.21%	0.72%	0.14%	0.14%		

The evaluated performances of the network present that for each validation data set. 12th data set results in a maximum error of 12.19% at a time period of 50 days, and a minimum error of 0.05% at 180 days. The maximum error of other five data sets is decreased to 3.80% - 10.13%. The minimum error is within 0.14% and 0.72%, which indicates a good network performance. The mean square error after training is not lowered to 10E-6 but still predicts the favorable results. By comparing ANN prediction and simulation the accuracy of the network is verified.

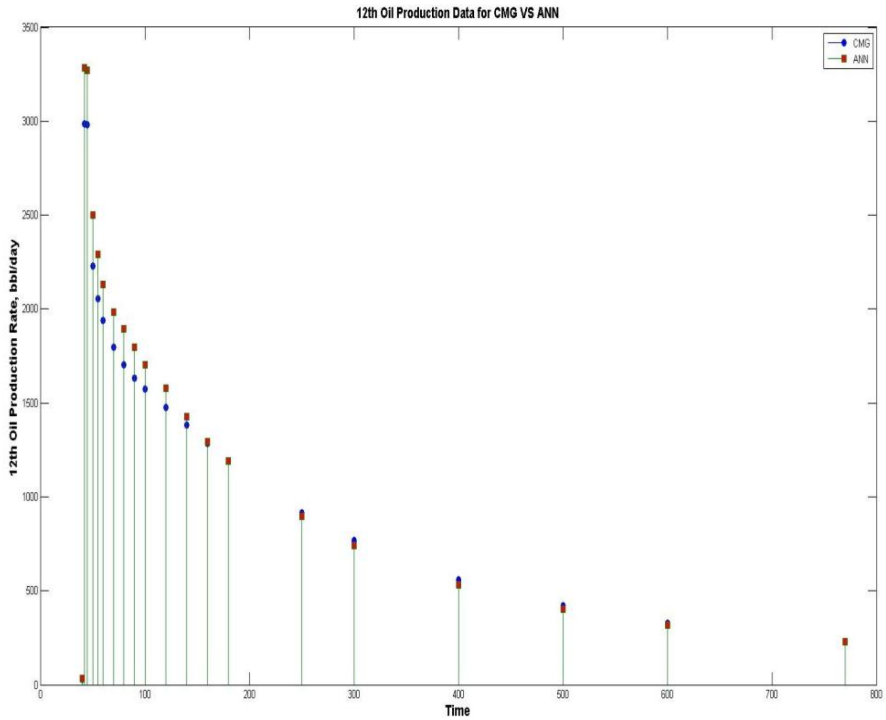


Figure 5-6 Validation Data #12 Prediction Performance of Network PVT-3

It is clear to see that the maximum error for each validation data set is taken place at early time of production. The reason can be demonstrated in Figure 5-7 by the simulator run of PVT-3. The significantly decrease of oil production rate at the first days brings the difficulty to the network to precisely track the oil rate change. Every time to initiate and train the same network, it can produce some different results since they choose different data sets randomly for the purpose of training, validation and training. But it is obvious that overall performance of each run is similar and the prediction error value ranges are close.

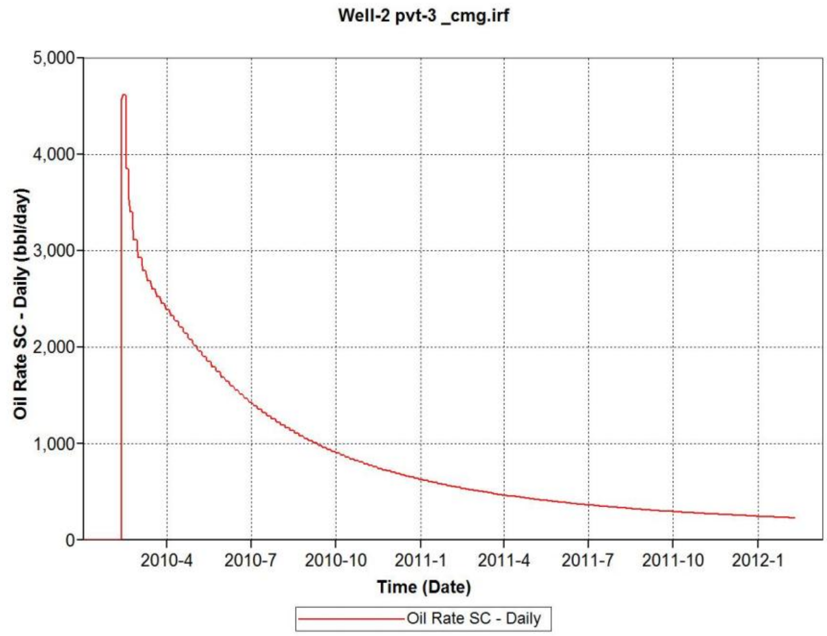


Figure 5-7 Oil Production Rate of PVT-3 by Numerical Simulation

Six testing data sets are fed into the network to indicate the prediction of errors results. Table 5-12 and Table 5-13 list and summarize the maximum and minimum error values for testing data inputs parameters. Network PVT-1 and Network PVT-3 both represent black oil fluid. The prediction error is acceptable in general. The deviation of network predicted oil rate values can be lower or higher than the simulation run results. ANN functions better interpolation than extrapolation as long as the testing data are within the input parameters values.

Table 5-12 Network PVT-3 Prediction Errors for Testing Data Sets

	7	19	21	44	51	58*	max	min
q@40(start day)	0.99%	2.67%	1.64%	3.26%	4.66%	9.37%	9.37%	0.99%
q@MAX	0.16%	3.20%	1.43%	3.09%	2.76%	10.37%	10.37%	0.16%
q@45	1.07%	2.78%	1.06%	2.71%	1.68%	11.02%	11.02%	1.06%
q@50	1.02%	2.28%	1.28%	2.29%	1.28%	10.37%	10.37%	1.02%
q@55	1.22%	2.70%	1.46%	2.11%	1.54%	10.48%	10.48%	1.22%
q@60	1.10%	1.94%	1.28%	2.12%	1.58%	10.26%	10.26%	1.10%
q@70	1.35%	2.16%	1.46%	1.31%	1.79%	10.28%	10.28%	1.31%
q@80	1.66%	2.49%	2.18%	0.26%	4.26%	9.10%	9.10%	0.26%
q@90	2.19%	2.73%	1.75%	0.25%	0.23%	10.79%	10.79%	0.23%
q@100	2.78%	2.43%	2.46%	1.25%	1.87%	9.96%	9.96%	1.25%
q@120	2.85%	2.64%	2.69%	2.44%	1.44%	8.52%	8.52%	1.44%
q@140	2.68%	2.13%	2.34%	2.83%	1.29%	7.83%	7.83%	1.29%
q@160	2.44%	1.85%	1.86%	2.74%	0.94%	7.10%	7.10%	0.94%
q@180	2.56%	2.56%	2.36%	3.47%	2.44%	5.53%	5.53%	2.36%
q@250	1.29%	1.75%	1.00%	2.16%	0.87%	3.86%	3.86%	0.87%
q@300	0.86%	1.55%	0.99%	2.38%	2.37%	2.50%	2.50%	0.86%
q@400	0.06%	0.42%	0.33%	1.43%	1.72%	1.58%	1.72%	0.06%
q@500	0.47%	0.27%	0.90%	0.30%	0.69%	2.36%	2.36%	0.27%
q@600	0.23%	0.11%	1.24%	1.10%	1.76%	3.39%	3.39%	0.11%
q@770	0.16%	0.38%	1.44%	1.64%	5.97%	7.10%	7.10%	0.16%
max	2.85%	3.20%	2.69%	3.47%	5.97%	11.02%		
min	0.06%	0.11%	0.33%	0.25%	0.23%	1.58%		

*: Worst Case

Table 5-13 Network PVT-3 Crude Oil Fluids Properties Range for Training, Testing and Validation

	Training		Testing		Validation	
	max	min	max	min	max	min
Tc(°F)	1124.895	344.574	868.961	702.842	948.481	469.688
Pc(psia)	4281.37	344.7	2290.18	552.13	3680.29	467.65
Tcb(°F)	611.044	179.608	425.87	347.181	633.314	227.599
Pcb(psia)	5802.718	537.083	3517.436	1410.531	4765.515	703.305
Tct(°F)	1131.968	528.517	870.412	753.763	943.139	605.464
Pct(psia)	1350.588	384.817	992.792	507.77	1307.756	449.399
Saturation P	5784.938	349.688	3246.645	1335.598	4761.397	687.03
Compres.	2.77E-05	4.72E-06	1.62E-05	1.01E-05	2.00E-05	9.11E-06
IFT(lbf/ft)	0.00817	0.00278	0.00621	0.00496	0.00758	0.00571
Solution GOR	320.12	35.5003	157.2086	49.2563	260.93	58.6263
Oil FVF	1.3256	1.0429	1.1581	1.064	1.2491	1.0684
C_{7+} MW	368.9	160	250	201	258	173
C_{7+}SG	0.9594	0.7978	0.8681	0.8383	0.8783	0.799
E(CP,CB)	4539.927	1130.375	2888.116	1428.045	3988.524	1417.44
E(CP,CT)	4508.905	1217.818	2685.566	1393.967	3910.416	1402.072

5.1.4 Crude Oils Represented by Network PVT-4

PVT-4 is a real crude oil sample collected from the field which is undergoing stream injecting. The steam injection schedule does not contribute to enhance the oil rate and cumulative recovery apparently. Since all networks are taking the same input parameters, doubt may arise that whether the trained network is able to present the characteristics of PVT -4 production history. The randomly selected sets are validated along the training. The properties are collected in the Table 5-12 and Table 5-13, while the inputs and output errors are calculated in Table 5-14.

Table 5-14 Composition Data for the Validation Sets

No.	11	31	38	44	45	53*
CO_2	0.75	0.25	0.038	2.09	1.75	0.22
N_2	0.21	0.88	0.051	0.17	0.74	0.054
C_1	12.56	23.94	22.984	31.72	37.15	53.45
C_2	2.59	11.67	0.015	15.41	7.45	6.36
C_3	4.83	9.36	0.083	7.49	9.64	4.66
iC_4	0.86	1.39	0.131	0.45	1.93	2.78
nC_4	3.83	4.61	0.38	1.81	5.65	1.01
iC_5	2.88	1.5	0.854	2.49	2.13	1.44
nC_5	1.26	2.48	0.823	2.904	2.83	1.3
C_6	5.42	3.26	2.647	2.436	5.72	2.116
C_{7+}	64.81	40.66	71.994	33.03	25.01	26.61
C_{7+} MW	220	196	217.5	201	182	215.9
C_{7+} s.g.	0.857	0.8494	0.8351	0.8383	0.8236	0.853

*: Worst Case

Table 5-15 Inputs for Network PVT-4

No.	11	31	38	44	45	53*
Tc (°F)	863.584	747.742	858.756	716.762	606.302	734.158
Pc (psia)	601.67	1284.29	544.75	1645.24	1957.44	2322.38
Tcb (°F)	508.009	425.402	353.64	386.973	352.906	282.403
Pcb (psia)	1010.08	2197.713	1312.239	2835.968	2801.52	4703.994
Tct (°F)	868.132	771.848	860.293	753.763	665.837	778.12
Pct (psia)	524.177	858.847	550.723	969.19	1018.91	1189.143
Saturation Pressure (psia)	799.09	1871.316	1234.422	2528.856	2587.899	4652.822
Compressibility ($psia^{-1}$)	1.15E-05	1.47E-05	1.05E-05	1.45E-05	1.96E-05	1.16E-05
IFT (lbf/ft)	6.56E-03	6.78E-03	6.15E-03	6.21E-03	6.13E-03	6.49E-03
Solution GOR (SCF/STB)	88.2114	176.5931	50.9576	157.2086	216.5194	97.745
Oil FVF	1.0955	1.1644	1.0664	1.1493	1.2124	1.1035
C_{7+} MW	220	196	217.5	201	182	215.9
C_{7+} sg	0.857	0.8494	0.8351	0.8383	0.8236	0.853
E(CP,CB)	1460.83	2019.451	1400.831	2326.846	2483.98	2927.924
E(CP,CT)	1419.798	1873.262	1407.751	2153.656	2347.151	2775.63

Table 5-16 Error Analysis for Six Validation Data Sets Using Network PVT-4

Error in %	11	31	38	44	45	53*	Max	Min
q@40(start day)	0.55%	5.93%	0.09%	1.82%	2.28%	7.62%	7.62%	0.09%
q@MAX	0.30%	5.85%	0.00%	1.83%	1.91%	7.80%	7.80%	0.00%
q@45	0.36%	5.65%	0.83%	1.77%	2.46%	7.54%	7.54%	0.36%
q@50	2.09%	5.65%	1.73%	1.48%	3.44%	7.45%	7.45%	1.48%
q@55	1.85%	5.54%	1.00%	1.06%	2.65%	8.47%	8.47%	1.00%
q@60	1.93%	4.87%	1.76%	1.10%	3.89%	6.62%	6.62%	1.10%
q@70	1.57%	5.10%	1.01%	0.78%	2.24%	8.40%	8.40%	0.78%
q@80	0.39%	4.00%	0.23%	0.44%	3.61%	6.68%	6.68%	0.23%
q@90	2.54%	2.76%	2.53%	0.36%	2.45%	6.87%	6.87%	0.36%
q@100	3.84%	1.63%	3.47%	1.11%	1.96%	7.09%	7.09%	1.11%
q@120	0.45%	3.22%	1.42%	0.82%	2.09%	7.57%	7.57%	0.45%
q@140	2.07%	0.23%	2.40%	2.17%	3.96%	5.71%	5.71%	0.23%
q@160	0.56%	1.28%	0.35%	1.74%	3.01%	6.19%	6.19%	0.35%
q@180	2.28%	0.40%	3.26%	2.81%	2.20%	6.05%	6.05%	0.40%
q@250	0.85%	0.59%	1.76%	1.88%	2.04%	3.55%	3.55%	0.59%
q@300	0.48%	1.54%	1.53%	2.35%	2.16%	2.65%	2.65%	0.48%
q@400	0.23%	1.77%	0.65%	1.82%	1.46%	1.71%	1.82%	0.23%
q@500	0.54%	2.17%	2.46%	0.87%	1.45%	0.00%	2.46%	0.00%
q@600	0.50%	1.72%	5.24%	0.53%	1.65%	0.20%	5.24%	0.20%
q@770	1.49%	1.72%	8.65%	1.25%	1.95%	1.37%	8.65%	1.25%
Max	3.84%	5.93%	8.65%	2.81%	3.96%	8.47%		
Min	0.23%	0.23%	0.00%	0.36%	1.45%	0.00%		

The 11th data set is illustrated in this section for the error analysis in Figure 5-8. Table 5-12 clearly demonstrates the good validation results for this network. With same input parameter, the network representing heavy oil has better predictions. The 11th has the maximum error of 3.84% which is pretty low and the minimum error is even smaller at 0.23%. The maximum prediction relative error at each time period is at 8.65%.

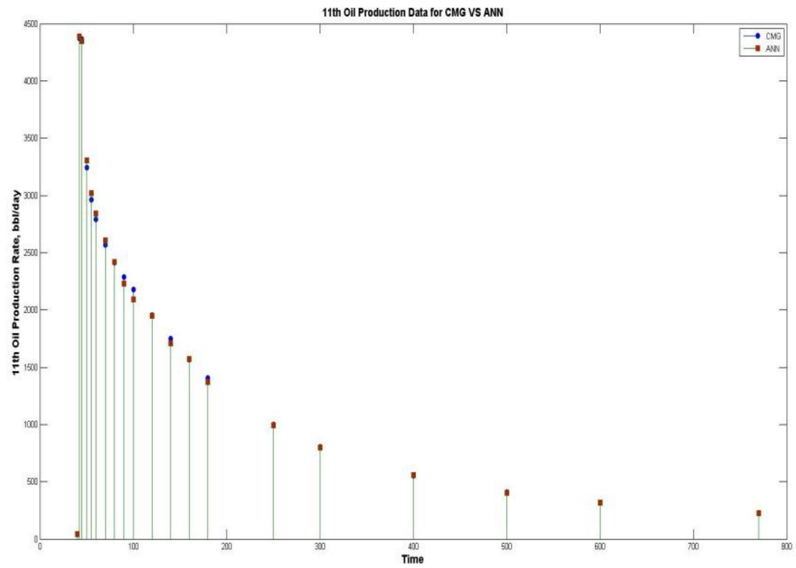


Figure 5-8 Validation Data #11 Prediction Performance of Network PVT-4

ANN cannot train the network itself. It trains and learns the data only from users' inputs. The oil production of PVT-4 has relatively constant and smooth values even they are significantly small and negligible comparing with other three PVT values. Difference of production rate after deduction is generally falling into an easier trend to track in this network. While the previous three PVT models, they themselves experience sharp drop at early time. The network outputs are likely to suffer more oscillations. This complexity increases the difficulty for the networks to understand the nonlinear relationships between inputs and targets.

Again, the six testing data sets randomly selected to check this heavy oil network with the prediction errors in Table 5-17. Table 5-18 displays the input parameters spectrum for training, testing and validation samples.

Table 5-17 Network PVT-4 Prediction Errors for Testing Data Sets

	5	6	17	30	56*	59	max	min
q@40(start day)	2.95%	7.46%	1.34%	1.44%	36.72%	7.37%	36.72%	1.34%
q@MAX	2.93%	7.89%	0.66%	1.89%	38.23%	6.79%	38.23%	0.66%
q@45	2.69%	7.52%	0.74%	1.74%	35.24%	7.59%	35.24%	0.74%
q@50	3.40%	7.32%	1.95%	0.88%	34.78%	7.15%	34.78%	0.88%
q@55	4.19%	8.45%	0.96%	1.74%	31.88%	5.89%	31.88%	0.96%
q@60	3.13%	6.46%	2.33%	0.34%	34.18%	7.62%	34.18%	0.34%
q@70	4.31%	8.67%	0.34%	2.13%	36.63%	5.05%	36.63%	0.34%
q@80	3.32%	6.43%	1.54%	0.77%	34.29%	6.41%	34.29%	0.77%
q@90	3.01%	6.57%	0.21%	1.76%	30.47%	4.64%	30.47%	0.21%
q@100	3.22%	6.89%	1.41%	2.55%	26.22%	3.71%	26.22%	1.41%
q@120	4.78%	7.57%	0.72%	2.16%	35.59%	0.75%	35.59%	0.72%
q@140	3.49%	5.43%	1.01%	1.38%	27.79%	1.21%	27.79%	1.01%
q@160	4.46%	6.14%	0.94%	1.26%	32.95%	0.13%	32.95%	0.13%
q@180	3.91%	5.70%	2.54%	1.75%	26.68%	0.69%	26.68%	0.69%
q@250	2.38%	3.32%	1.94%	0.05%	23.53%	0.95%	23.53%	0.05%
q@300	2.09%	2.44%	1.96%	0.10%	20.27%	0.73%	20.27%	0.10%
q@400	1.99%	1.70%	1.18%	0.13%	14.13%	0.15%	14.13%	0.13%
q@500	1.21%	0.16%	0.88%	0.06%	9.76%	0.88%	9.76%	0.06%
q@600	1.28%	0.06%	0.06%	1.01%	5.57%	0.75%	5.57%	0.06%
q@770	1.77%	0.77%	1.39%	1.96%	1.96%	2.52%	2.52%	0.77%
max	4.78%	8.67%	2.54%	2.55%	38.23%	7.62%		
min	1.21%	0.06%	0.06%	0.05%	1.96%	0.13%		

*: Worst Case

Table 5-18 Network PVT-4 Crude Oil Properties Range for Training, Testing and Validation

No.	5	6	17	30	56*	59	Validation Set Max.	Validation Set Min.	Network Max.	Network Min.
Tc (°F)	680.943	770.553	828.976	823.977	952.209	528.522	863.584	606.302	1124.895	344.574
Pc (psia)	2548	2079.09	1304.65	1244.93	1618.51	2089.32	2322.38	544.75	4281.37	344.7
Tcb (°F)	308.072	339.89	379.264	421.505	379.67	338.05	508.009	282.403	633.314	179.608
Pcb (psia)	4318.681	4120.912	2951.568	2510.918	4071.709	2659.559	4703.994	1010.08	5802.718	537.083
Tct (°F)	753.667	820.522	846.505	850.377	995.189	601.771	868.132	665.837	1131.968	528.517
Pct (psia)	1168.517	1079.969	903.646	785.254	839.427	1032.384	1189.143	524.177	1350.588	384.817
Saturation Pressure (psia)	4211.416	3938.92	2711.813	2185.911	3842.152	2464.022	4652.822	799.09	5784.938	349.688
Compressibility (psia-1)	1.32E-05	1.15E-05	1.06E-05	1.31E-05	9.92E-06	2.35E-05	1.96E-05	1.05E-05	2.77E-05	4.72E-06
IFT (lbf/ft)	6.05E-03	5.29E-03	7.10E-03	4.44E-03	2.78E-03	5.95E-03	0.00678	0.00613	0.00817	0.00313
Solution GOR (SCF/STB)	138.3813	113.112	100.2	140.9302	78.2185	250.95	216.5194	50.9576	320.12	35.5003
Oil FVF	1.1382	1.1122	1.1005	1.1381	1.0864	1.2481	1.2124	1.0664	1.3256	1.0429
C_{7+} MW	215	240	217	252	***368.9***	168	220	182	329	160
C_{7+} sg	0.8479	0.8518	0.868	0.8424	0.875	0.8067	0.857	0.8236	0.9594	0.7978
E(CP,CB)	3098.365	2777.125	2151.244	2084.532	2572.564	2536.963	2927.924	1400.831	4539.927	1130.375
E(CP,CT)	2938.159	2571.21	1973.183	1878.291	2258.179	2418.101	2775.63	1407.751	4508.905	1217.818

The validation data sets in this network prove earlier that the network is well learned with small error comparing numerical simulation and ANN predictions. However, there is one data sample among the six testing data sets has high error in Table 5-17. The 56th sample experiences high deviation of the predicted production rate throughout the production history. Tracing back to the inputs and compositions for the 56th, it proves that molecular weight of C_{7+} is crucial to the network training. It also demonstrates the nature of an artificial neural network that extrapolation many times does not work well.

5.2 Case Study

Once networks' prediction performance is confirmed by a couple of experiments and validations in the previous section, new crude oils as case studies are desirable to be discussed comparing the numerical simulation results in details. After the categorization of a crude oil, the developed expert system can be utilized to assist in reducing EOR simulation tests in Claudia Parada's screening and design tool-box. Based on four fluid types, crude oils are analyzed accordingly to testify how it supports to narrowing down the selection of proper EOR process. The reservoir model keeps consistent and same steam injection process is applied for enhanced oil production.

5.2.1 Crude Oil Represented by Network PVT-1 and Categorization Performance

A crude oil sample from literature is selected to initiate the categorization. Table 5-19 and Table 5-20 give the molar fraction and thermodynamic properties as network inputs. Two-phase envelop is displayed in Figure 5-9.

Table 5-19 Crude Oil T03 Molar Fraction

Crude Oil T03 Composition	
CO_2	2.61
N_2	0.77
C_1	31.88
C_2	5.14
C_3	7.43
iC_4	3.8
nC_4	4.41
iC_5	4.99
nC_5	5.96
C_6	6.13
C_{7+}	26.88
C_{7+} MW	228.1
C_{7+} s.g.	0.865

Table 5-20 Crude Oil T03 Thermodynamic Properties for Network

Network Inputs	**T03**
Tc (°F)	742.857
Pc (psia)	1761.92
Tcb (°F)	420.128
Pcb (psia)	2810.167
Tct (°F)	797.421
Pct (psia)	932.313
Saturation Pressure (psia)	2423.915
Compressibility ($psia^{-1}$)	1.53E-05
IFT (lbf/ft)	5.92E-03
Solution GOR (SCF/STB)	172.7741
Oil FVF	1.1715
C_{7+} MW	228.1
C_{7+} sg	0.865
E(CP,CB)	2452.4936
E(CP,CT)	2253.9215

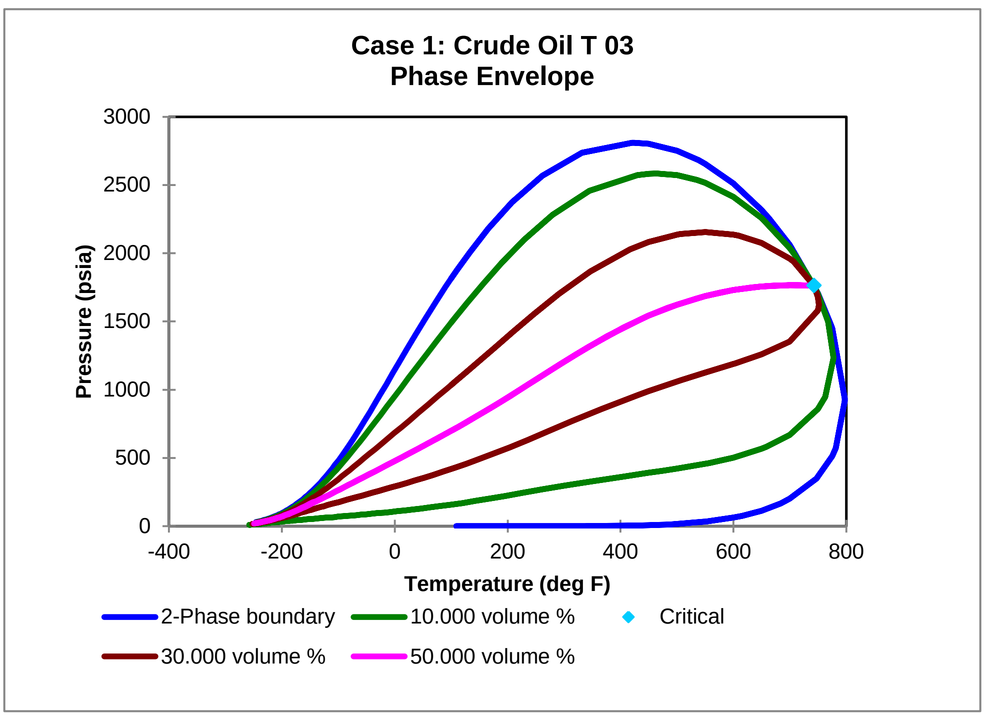


Figure 5-9 Crude Oil T03 Two-phase Envelope

The artificial neural intelligence based expert system identifies the categorization of this crude oil T03 can be best represented by base model PVT-1(fluid type PVT-1 in Claudia Parada's tool-box). Classification is determined from four output data sets and the results are plotted in Figure 5-10.

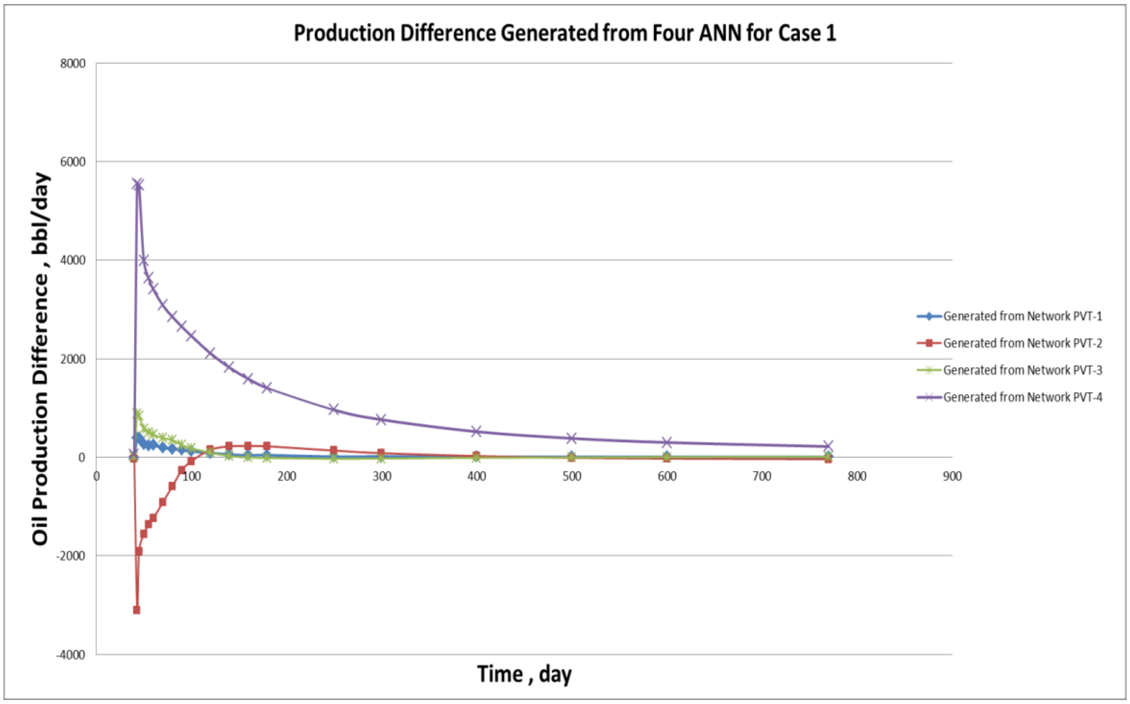


Figure 5-10 Case study Crude Oil T03 Results From ANN Based Expert System

There are four curves in Figure 5-10 and each curve indicates predicted production differences at specified times by four networks. One feature observed from this plot is that apparent production difference comparing with four base PVT models happens till approximately 200 days. Results from network PVT-1, PVT-2 and PVT-3 almost overlap on each other in late times. Doubt may arise that PVT-3 in green curve can also represent the crude oil since there is crossover with blue curve and overall two curves are very close. However, looking into the values at earlier times one sees that the peak oil flow rate difference in green curve almost doubles the number as in the blue one.

The criterion to categorize a crude oil is to find out the minimum summation of network outputs among four data sets, which means a new crude oil production is closest to that fluid type. According to the production history decline trend and the ANN prediction, results are divided into two parts (demarcated by specified time period) and the summation of each part is calculated and a final value comes out for comparison. Adding the absolute value of network prediction until time of 180 days together and the total stands for the summation of first part. The former part in Figure 5-10 determines the categorization to a great extent. Either the network prediction value is positive or negative, it quantifies the production difference. Since the production difference from all networks begin with comparable large numbers, the total summation before 180 days has a major influence. The latter part takes into account of the production difference from 180 days to the end. In this part, the predicted results are added together before taking the absolute value. The reason is that oil production of four base models tends to converge as shown in Figure 4-17 and thus production difference values of testing crude oils are becoming similar. Figure 5-10 shows the oil rate difference without considering the effect of cumulative oil production on categorization. The absolute value of the latter production difference summation considers the factor of total oil production because the positive and negative production differences imply this crude oil total production is likely to approximate one of fluid types included in Parada’s tool-box. The network having the minimum summation from two parts is the best representable candidate.

The classification predicted by ANN is compared to numerical simulation results for validation. Figure 5-11 expresses the production history of case study numbered T03 along with four base PVT models. The red curve represents T03 production history and from observation, the black dash curve (PVT-1) best captures the feature of peak oil rate and decline tendency of this testing crude oil. The crossover of T03 and green dash curve (PVT-2) is not able to compensate for the total production difference summation. The numerical simulation result verifies the accuracy of ANN categorization prediction in this case.

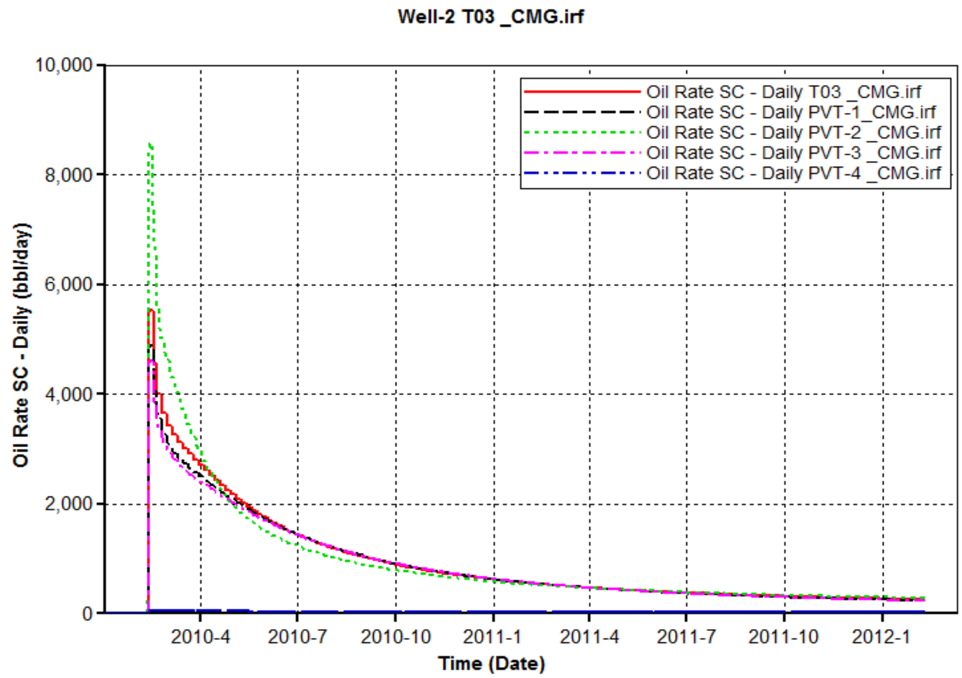


Figure 5-11 Numerical Simulation Results of Crude Oil T03 and Four Base Crude Oils

5.2.2 Crude Oil Represented by Network PVT-2 and Categorization Performance

Similarly, a crude oil molar composition is required to start up a new case study. Table 5-21 lists the molar fraction of crude oil T24, which composition of each component is falling within the limit. Table 5-22 shows network entries. Two-phase envelop in Figure 5-12 is provided for reference.

Table 5-21 Crude Oil T24 Molar Fraction

Crude Oil T24 Composition	
CO_2	1.6
N_2	1.12
C_1	45.78
C_2	17.24
C_3	12.07
iC_4	1.7
nC_4	5.02
iC_5	1.18
nC_5	2.09
C_6	1.86
C_{7+}	10.34
C_{7+} MW	171
C_{7+} s.g.	0.801

Table 5-22 Crude Oil T24 Thermodynamic Properties for Network

Network Inputs	**T24**
Tc (°F)	344.574
Pc (psia)	3120.89
Tcb (°F)	280.404
Pcb (psia)	3212.821
Tct (°F)	528.517
Pct (psia)	1102.901
Saturation Pressure (psia)	3132.452
Compressibility ($psia^{-1}$)	2.24E-05
IFT (lbf/ft)	5.39E-03
Solution GOR (SCF/STB)	313.9514
Oil FVF	1.292
C_{7+} MW	171
C_{7+} sg	0.801
E(CP,CB)	3414.359
E(CP,CT)	3316.994

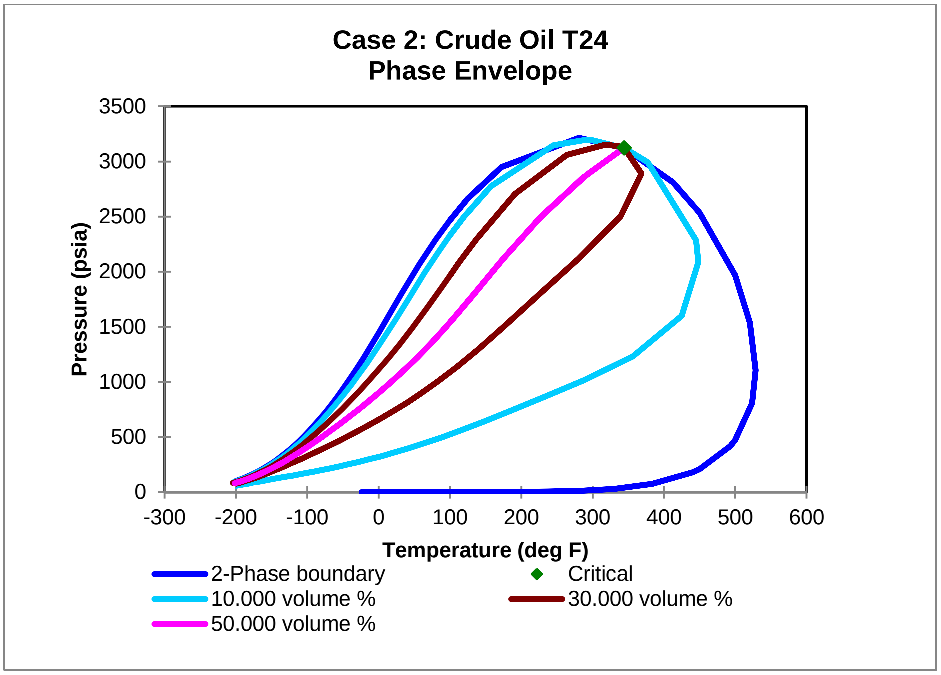


Figure 5-12 Crude Oil T24 Two-phase Envelope

The networks predict the classification that PVT-2(red curve) is the fluid type to represent this crude oil T24 as demonstrated in Figure 5-13. It can be clearly seen that production differences from production starting around 180 days are distinguished from each other. According to the summation criteria, red curve (PVT-2 predicted) has the minimum summation after taking the absolute value along the curve before 180 days. The final result is adjusted by adding the absolute value of latter part total production differences. It shows again that network predictions at first 180 days have a dominant impact on the classification.

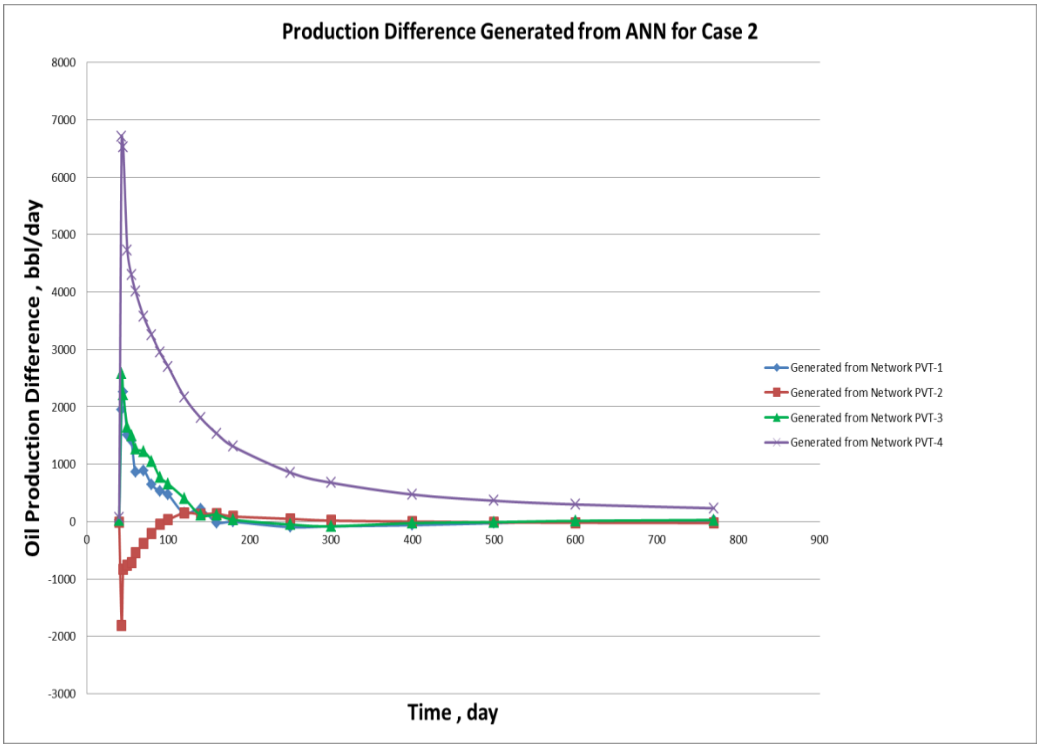


Figure 5-13 Case study Crude Oil T24 Results from ANN Based Expert System

Figure 5-14 gives the compositional simulation results of oil production rate and yields the same categorization. In this graph, it is obvious to find out crude oil T24 is closest to the green dash curve (PVT-2 oil production), especially the oil rate in early production times. Cumulative oil production in Figure 5-15 supports the accurate classification as well and shows the feasibility of criterion used to interpret expect system predicted data sets. The crossover of green dashed and red curve at early time in Figure 5-14 is manifested in the cumulative oil production shown in Figure 5-15. The late time production of crude oil T24 is not easy to understand from Figure 5-14 but with high oil rate at the beginning and less steep decline production behavior than PVT-2, the cumulative oil production is the highest and much closer to the PVT-2 production history. Conclusions drawn from Figure 5-15 are consistent with the ANN predictions.

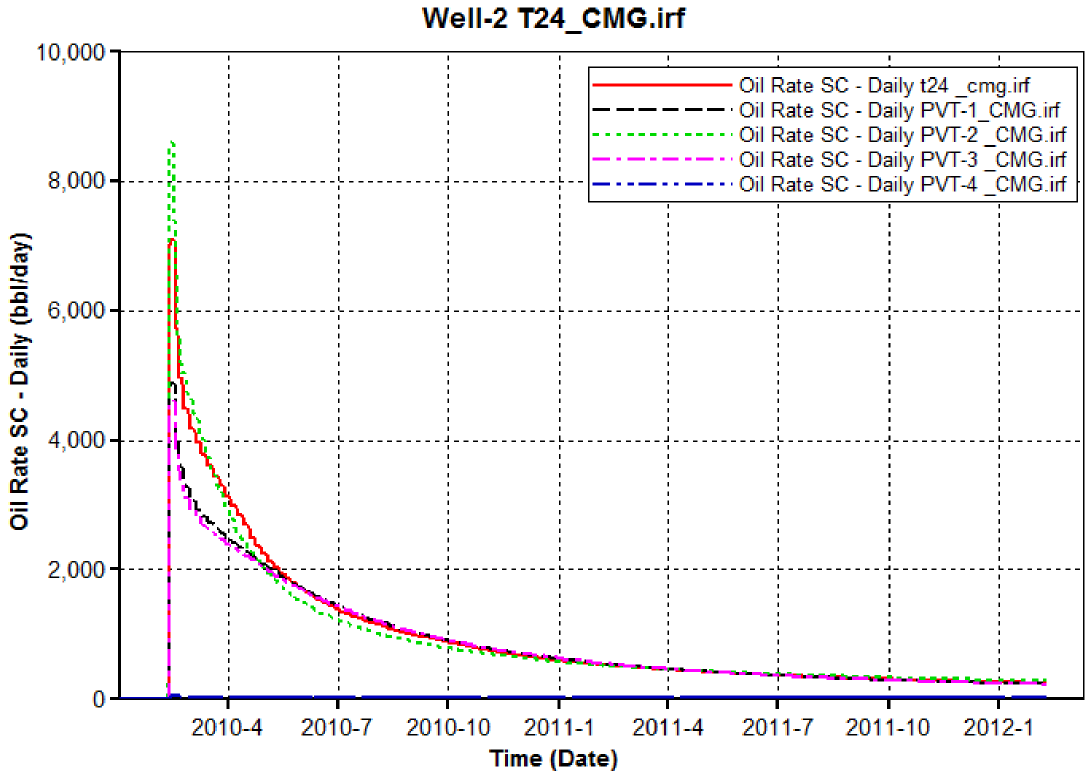


Figure 5-14 Numerical Simulation Results of Crude Oil T24 and Four Base Crude Oils

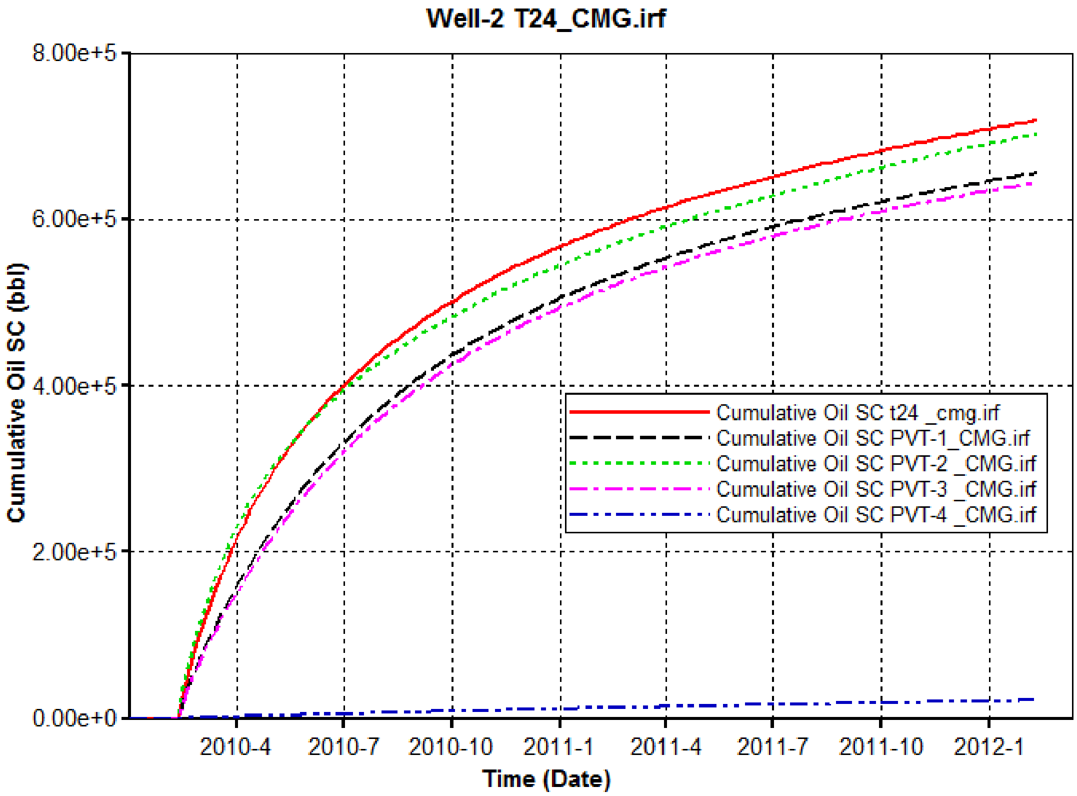


Figure 5-15 Cumulative Oil Production of Crude Oil T24

5.2.3 Crude Oil Represented by Network PVT-3 and Categorization Performance

PVT-1 and PVT-3 share the similar production history trends as can be seen in oil production history and cumulative oil production plots in Figure 4-17 and Figure 4-18. A case study is necessary to see if PVT-3 is able to be differentiated from PVT-1 and accurately represent crude oils. Crude oil T08 has the molar compositions and thermodynamics properties in Table 5-23 and Table 5-24, followed with its two-phase envelope in Figure 5-16.

Table 5-23 Crude Oil T08 Molar Fraction

Crude Oil T08 Composition	
CO_2	0.71
N_2	0.29
C_1	44.04
C_2	4.32
C_3	4.05
iC_4	2.14
nC_4	0.5
iC_5	1.11
nC_5	0.63
C_6	2.47
C_{7+}	39.74
C_{7+} MW	223.2
C_{7+} s.g.	0.846

Table 5-24 Crude Oil T08 Thermodynamic Properties for Network

Network Inputs	**T08**
Tc (°F)	813.213
Pc (psia)	1396.65
Tcb (°F)	324.695
Pcb (psia)	3483.607
Tct (°F)	831.749
Pct (psia)	937.139
Saturation Pressure (psia)	3367.132
Compressibility ($psia^{-1}$)	1.10E-05
IFT (lbf/ft)	6.03E-03
Solution GOR (SCF/STB)	78.8897
Oil FVF	1.0873
C_{7+} MW	223.2
C_{7+} sg	0.846
E(CP,CB)	2207.7504
E(CP,CT)	2034.7509

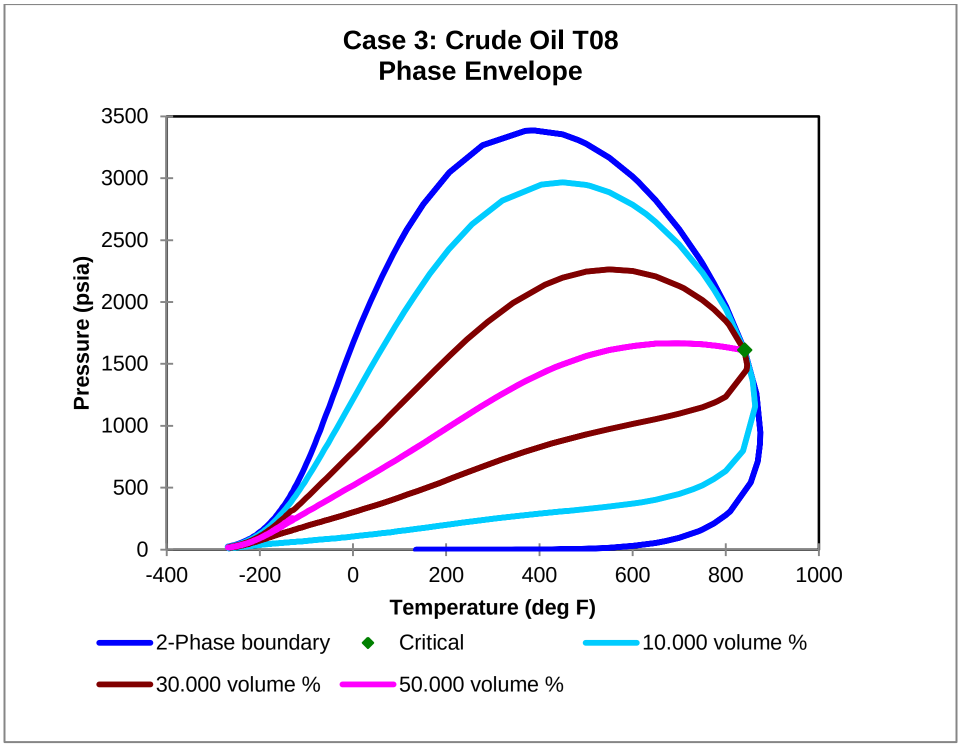


Figure 5-16 Crude Oil T08 Two-phase Envelope

It is observed that PVT-3 best represents this crude oil applying the interpretation criteria and four networks predicted outputs are shown in Figure 5-17. It is expected to see the production difference generated from network PVT-3 has the minimum value as well. Two curves standing for PVT-1 and PVT-3 will be zoomed in by adjusting y-axis limit and are displayed in Figure 5-18. The predicted production difference results in Figure 5-18 are illustrated from the initial production until 180 days, which is discussed earlier composing of the former part of summation. The green curve representing the network PVT-3 prediction and it can be seen that, the absolute values at each specified time step is smaller than the results on the blue curve. The calculation shows that average production difference generated by network PVT-1 is 20% greater than the difference from network PVT-3 at first 180 days. After adding the latter part summation, the expert system categorizes this crude oil T08 production history closer to PVT-3 than other three.

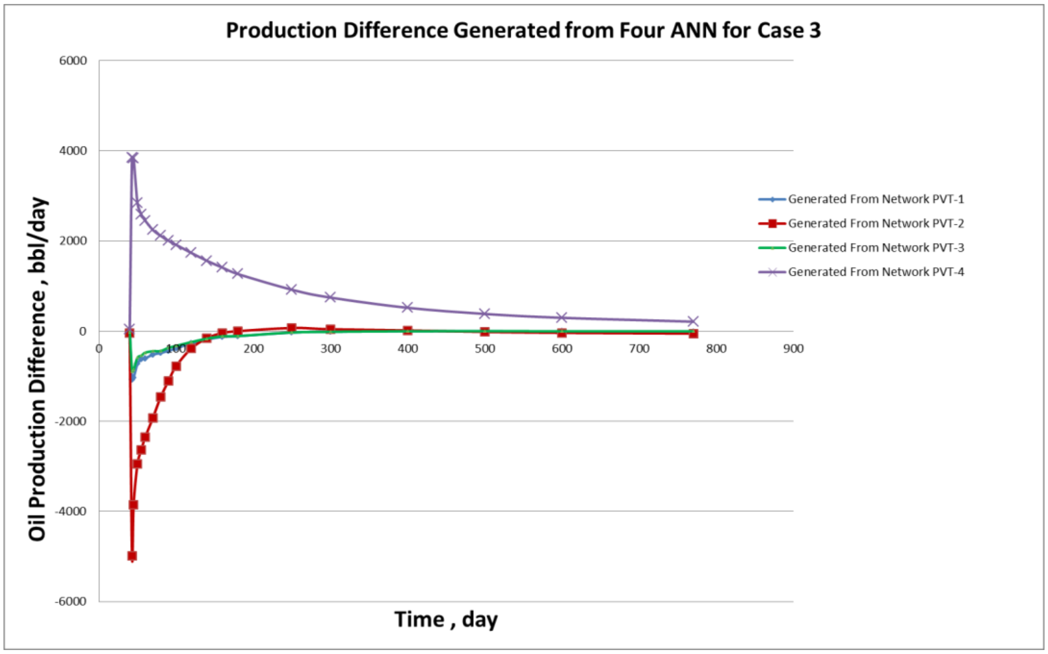


Figure 5-17 Case study Crude Oil T08 Results from ANN Based Expert System

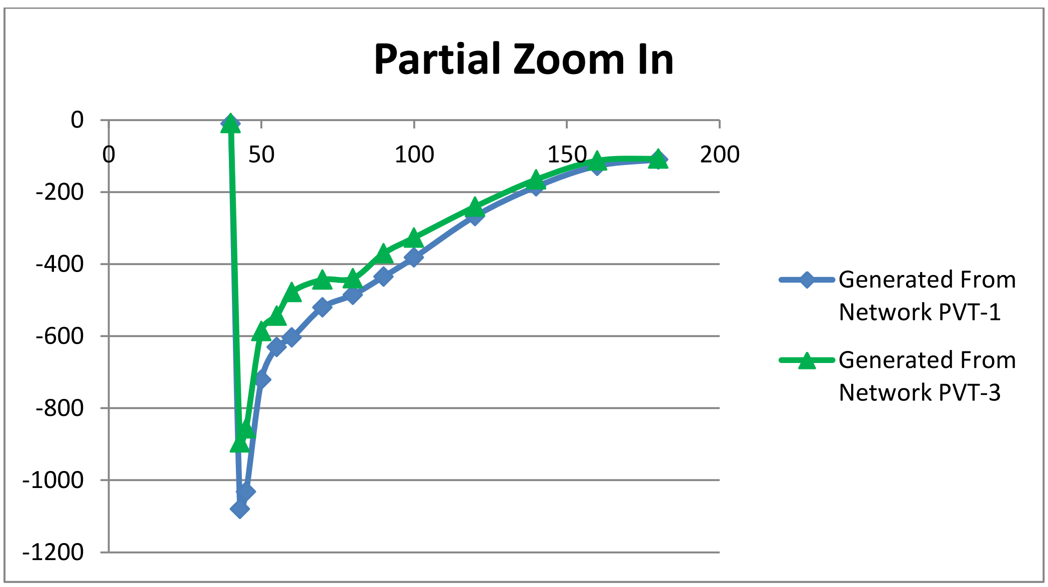


Figure 5-18 Zoomed in Production Difference Profile before 180 days

If the system classification is correct, the numerical simulation results should be able to reflect it both in production history in Figure 5-19 and cumulative oil production in Figure 5-20.

There is the least discrepancy between crude oil T08 in red curve and base fluid type PVT-3 expressed in magenta dash curve.

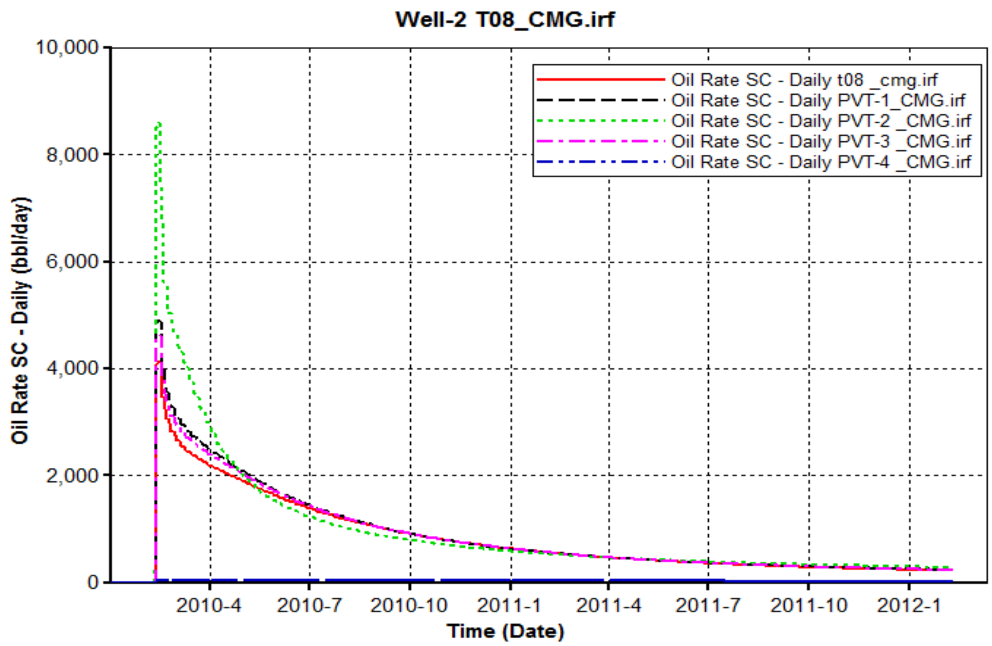


Figure 5-19 Numerical Simulation Results of Crude Oil T08 and Four Base Crude Oils

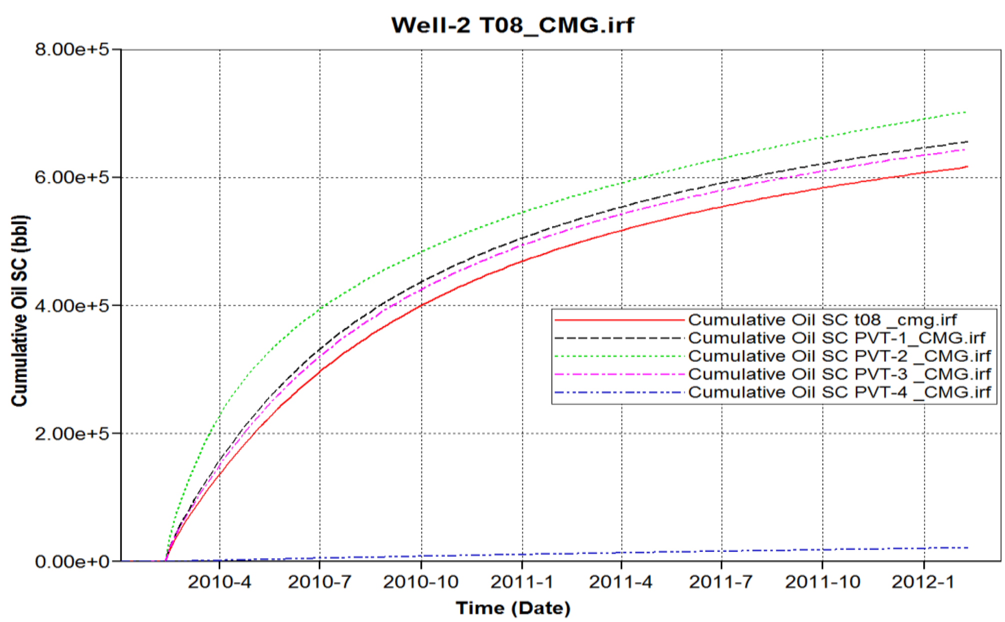


Figure 5-20 Cumulative Oil Production of Crude Oil T08

5.2.4 Crude Oil Represented by Network PVT-4 and Categorization Performance

PVT-4 shows relatively small but constant production rate in Figure 4-17. Generally speaking, heavy oil with large viscosity is a good candidate for thermal recovery processes. But viscous oil without optimized thermal operation may not result in apparent production enhancement. The uniform reservoir settings and well operation schedule for the PVT-4 do not stimulate the oil production significantly. But it is still able to differentiate the production from other three PVT models. Actually, the cumulative production for PVT-4 is twice as much as that of without steam injection.

A crude oil sample is collected from paper for case study. The molecular weight and specific gravity of C_{7+} are the greatest among four case studies but within system high limits. Composition data and properties are found inTable 5-25 and Table 5-26.

Table 5-25 Crude Oil Heavy Oil Molar Fraction

Heavy Oil Composition	
CO_2	0.15
N_2	0.5
C_1	12.1
C_2	0.7
C_3	0.35
iC_4	0.33
nC_4	0.74
iC_5	1.88
nC_5	2.98
C_6	1.79
C_{7+}	78.48
C_{7+} MW	510
C_{7+} s.g.	0.92

Table 5-26 Crude Oil Heavy Oil Thermodynamic Properties for Network

Network Inputs	**Heavy Oil**
Tc (°F)	1243.753
Pc (psia)	235.49
Tcb (°F)	478.892
Pcb (psia)	791.472
Tct (°F)	1242.843
Pct (psia)	232.671
Saturation Pressure (psia)	671.151
Compressibility ($psia^{-1}$)	4.85E-06
IFT (lbf/ft)	1.75E-03
Solution GOR (SCF/STB)	22.2518
Oil FVF	1.0296
C_{7+} MW	510
C_{7+} sg	0.92
E(CP,CB)	1535.3
E(CP,CT)	1476.7

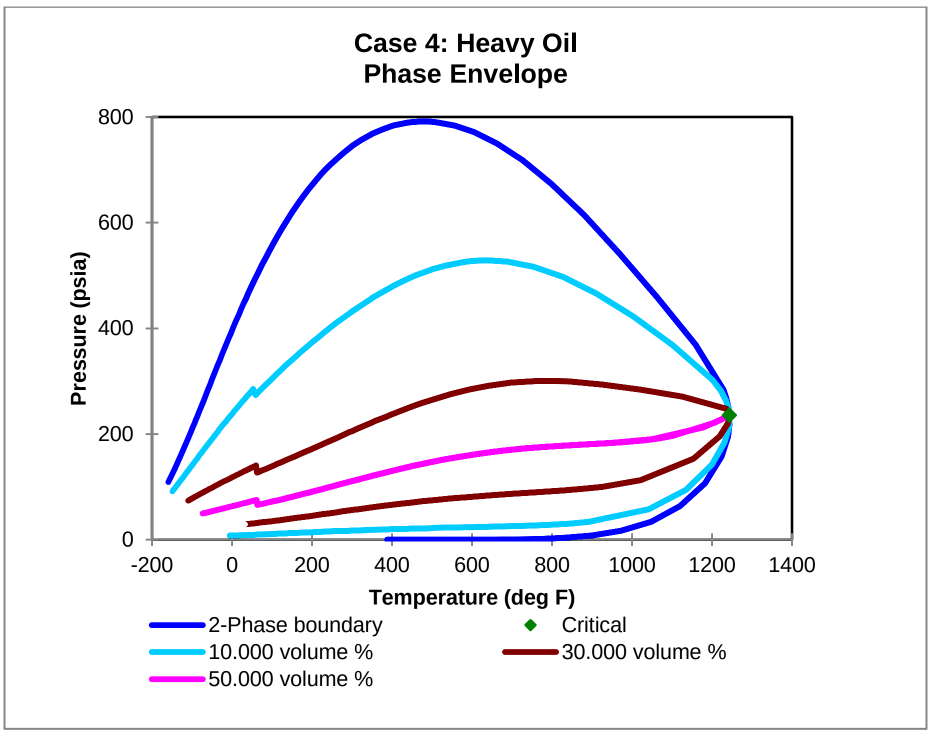


Figure 5-21 Crude Oil Two-phase Envelope

System generates four data sets as illustrated in Figure 5-22. It is clearly seen that network PVT-4 best represents this testing crude oil. Total summation from two parts of network predicted production difference comes out in the same categorization. Preceding three cases demonstrated that the established ANN based system is able to perform trustful categorization. Same result is expected to see from numerical simulation, which is plotted in Figure 5-23 and Figure 5-24.

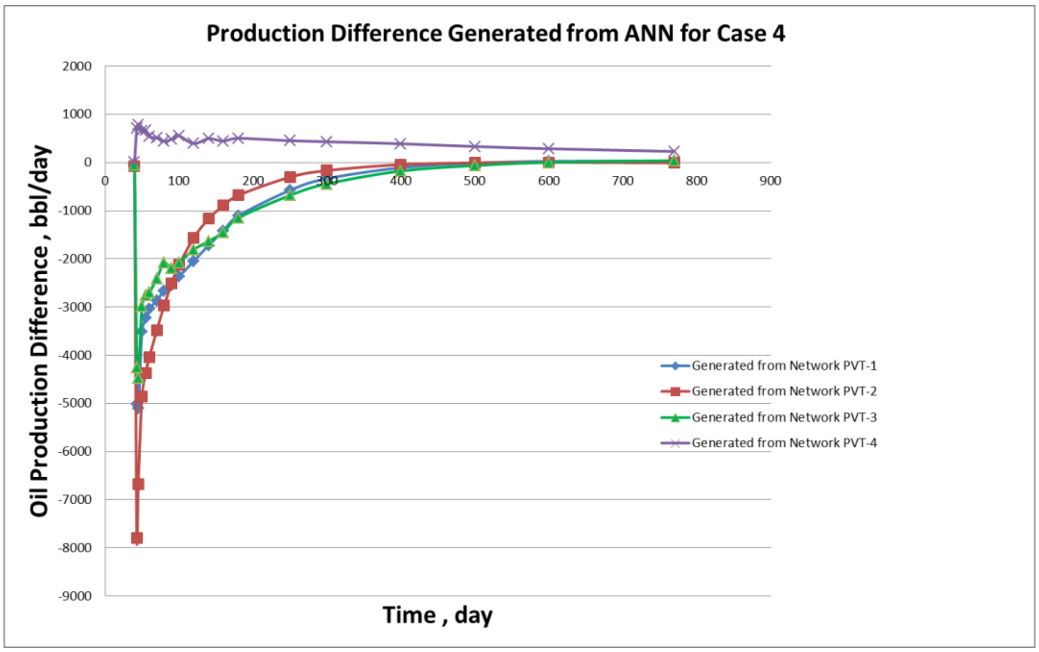


Figure 5-22 Case study Crude Oil Results from ANN Based Expert System

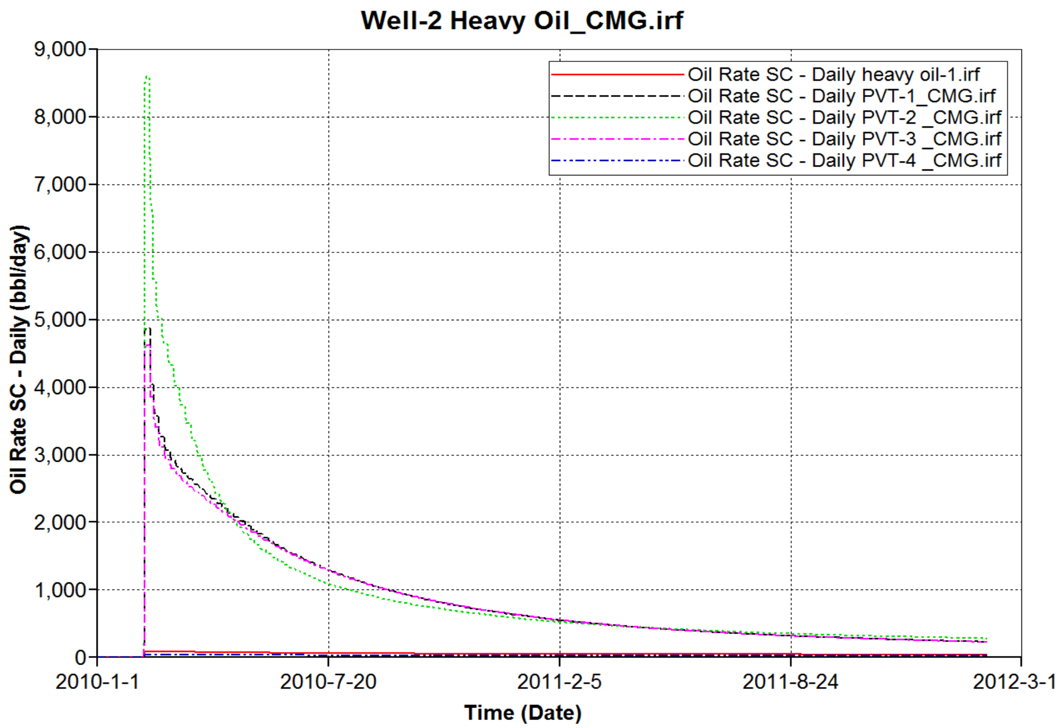


Figure 5-23 Numerical Simulation Results of Crude Oil and Four Base Crude Oils

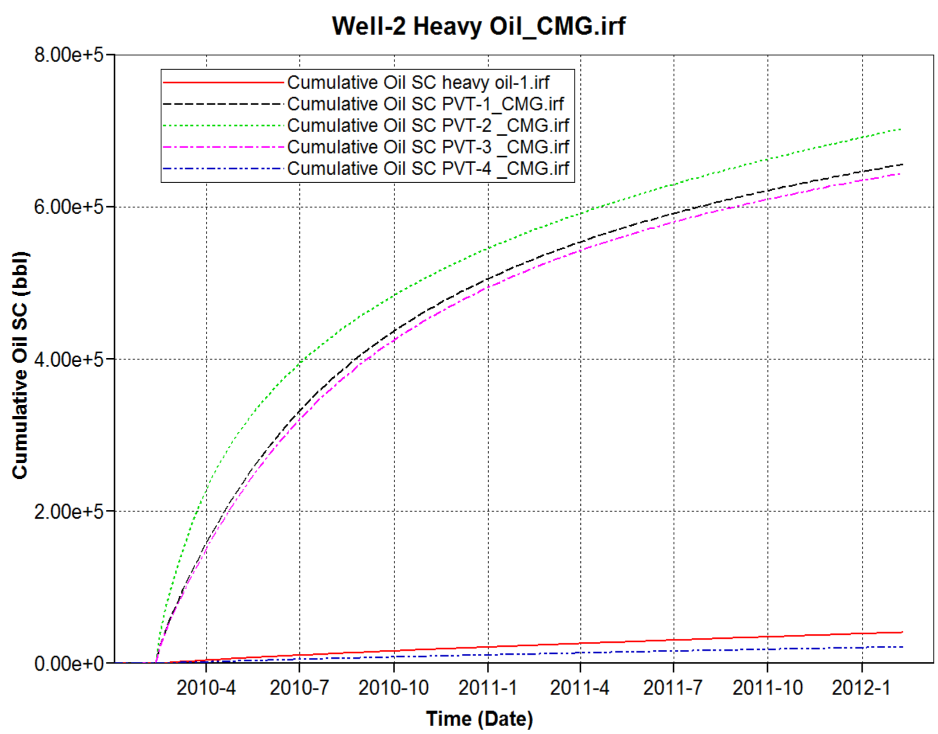


Figure 5-24 Cumulative Oil Production of Heavy Oil

Validation proves that a crude oil if given certain properties, it can rapidly determine a fluid type in tool-box to represent and thus appropriate recovery techniques and adequate design guidelines can be applied with its specific characteristics by time-saving.

5.3 Graphical User Interface Expert System

The preceding case studies demonstrated that categorization of a crude oil to one of four base fluid types is achieved utilizing artificial neural intelligence.

ANN base expert system is created in a single panel using Graphical User Interphase (GUI) shown in Figure 5-25. This user-friendly interface integrates four developed artificial neural network as background processing and provides a visual classification in one main window.

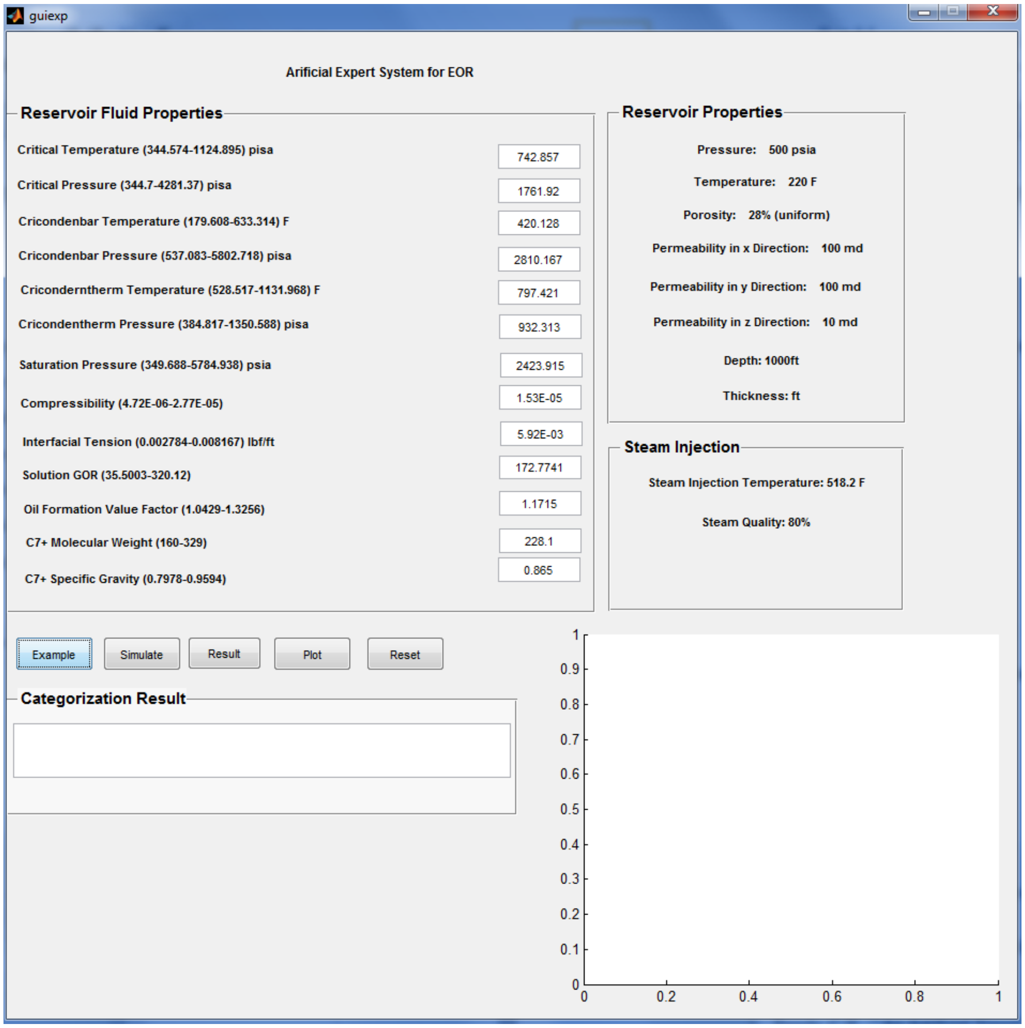


Figure 5-25 Graphical User Interphase to New Crude Oil with 15 Inputs

Once fifteen inputs are entered into the interface, each network predicts the production difference. The background processing can be depicted in Figure 5-26 and Figure 5-27. Figure 5-26 outlines four networks and each predicted result can be sketched in Figure 5-27.

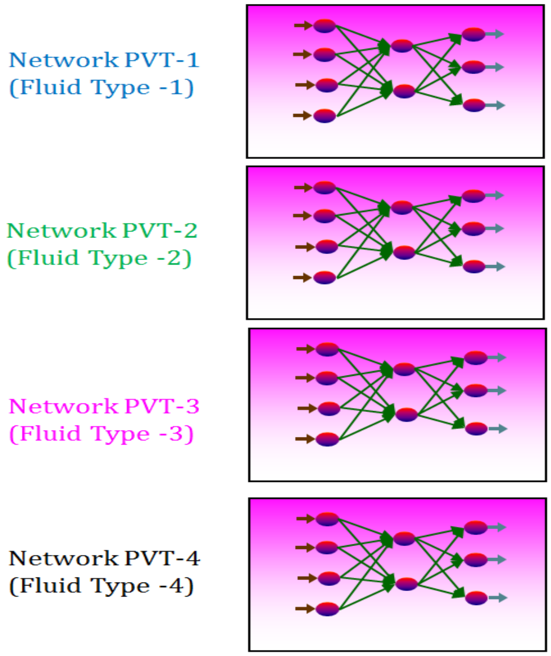


Figure 5-26 Schematics of Background Processing Networks

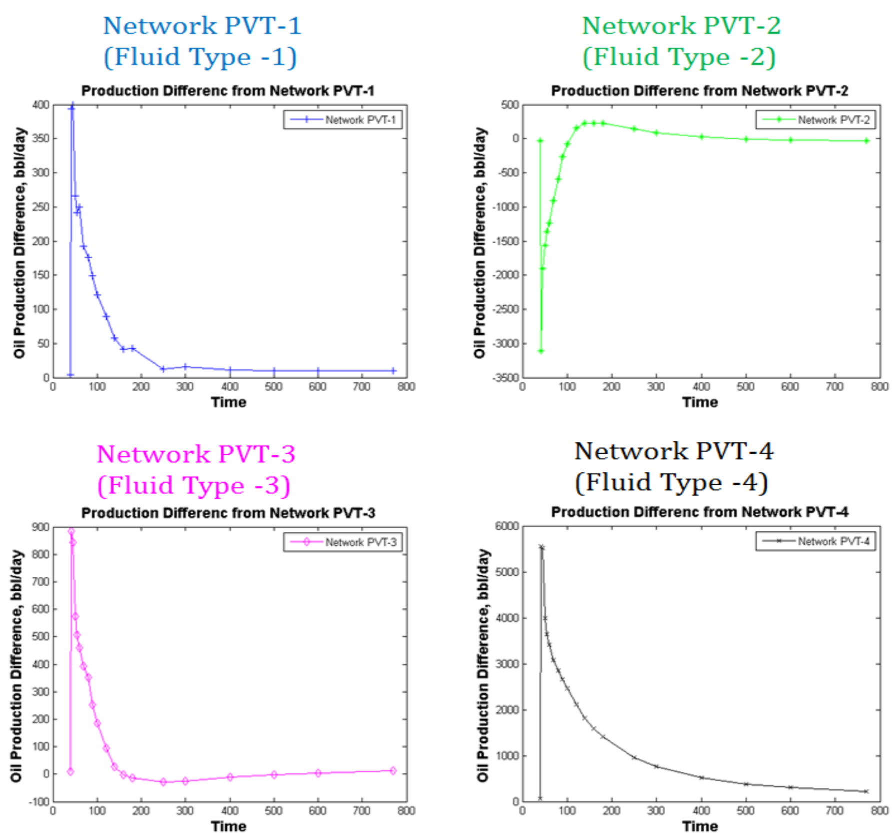


Figure 5-27 Networks Prediction Results in Background Processing

The reservoir properties and steam injection scheme is displayer in the window. And the upper left space is for entering inputs to networks. As long as the inputs are within the given range for each property, it is able to launch the categorization of system. Once clicking the 'Simulate' button, the inputs are adopted into four networks separately. 'Result' button is to display the categorization result in text to tell the user the final classification in this panel illustrated in Figure 5-28.

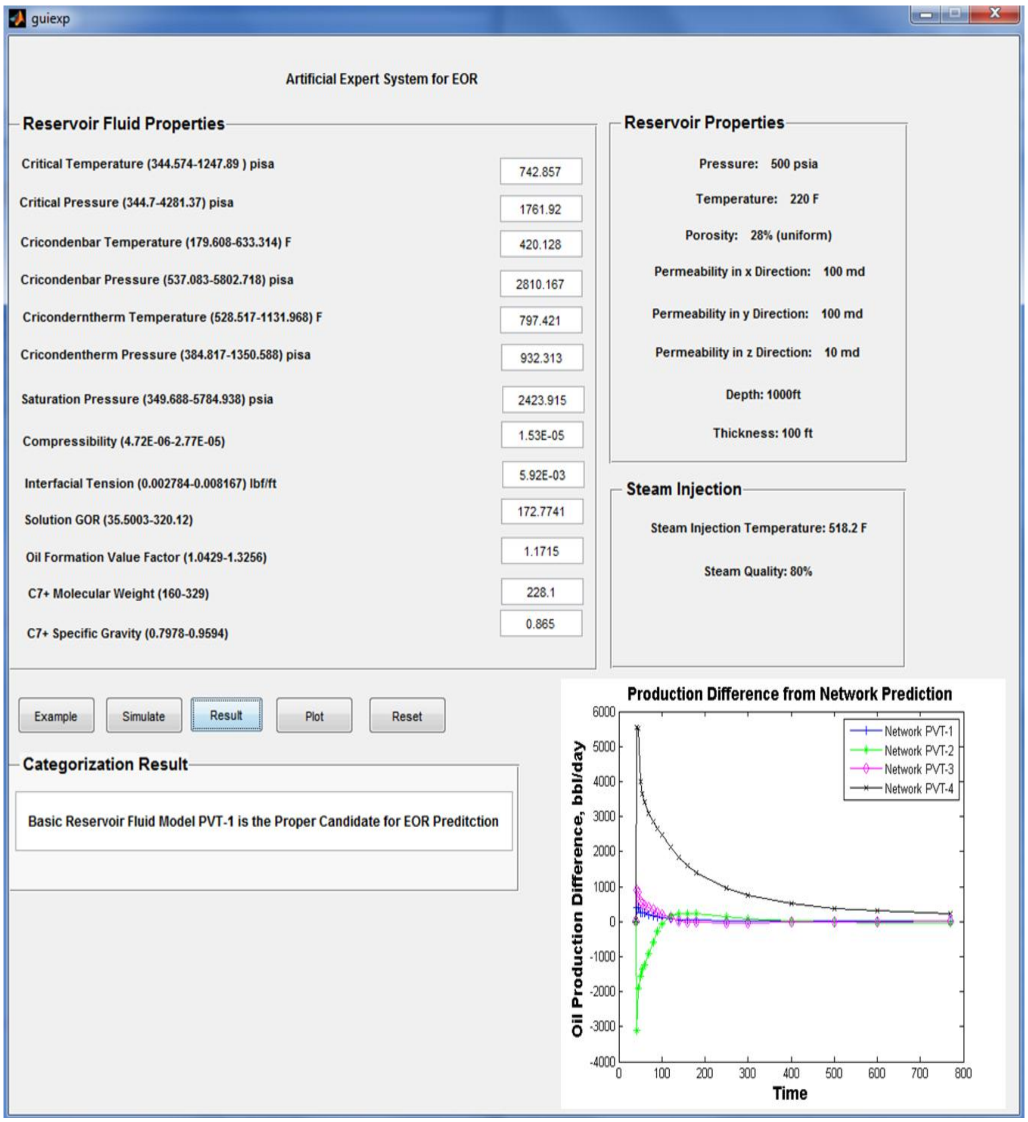


Figure 5-28 Example Using GUI and Results Shown in Main Window

Last but not least, the lower right space is prepared to check whether the classification is accurate. It integrates the prediction shown in Figure 5-27, which otherwise is running background. Users are at ease to visualize the categorization result comply with the plot at right hand side.

Chapter 6

Conclusions

Pertinent conclusions of the study are summarized below.

- Artificial intelligence technology was successfully utilized in this study.
- The categorization of a given crude oil to one of four fluid types in Claudia Parada's [2008] screening and design tool-box could be accurately identified.
- Integrating all trained networks to an expert system in Graphical User Interface was able to fast determine the classification in a visualized manner.
- Once the categorization was achieved, appropriate recovery techniques and adequate design guidelines can be applied to a crude oil reservoir with its specific characteristics.
- The developed expert system assisted in narrowing down the selection of a proper fluid type combined with feasible EOR processes by saving time-consuming and investigation.
- The inputs of the expert system are thermodynamic properties calculated under specified initial conditions. But the categorization is valid and applicable to all crude oils because simulation in this study is using fixed reservoir properties and well specification.

Here are some observations regarding to the heuristic network training process.

1) Functional links as mathematical representations amplified the importance of certain inputs and improved the network training.
2) The trial-and –error of the network inputs selection could be more efficiently with the help of relevance study.

3) Among all inputs, six variables having the most dominant impact on the prediction performance were: Compressibility, C_{7+} M.W., C_{7+} S.G., P_{ct}, Saturation Pressure and Eigenvalue scaling critical point and cricondenbar point.

References

- Adewale Olatunji:" Optimization of Natural Gas Field Development Using Artificial Neural Networks." Master's Thesis, The Pennsylvania State University, 2010.
- Asha Ramgulam, Turgay Ertekin, and Peter B. Flemings: "Utilization of Artificial Neural Networks in the Optimization of History Matching", SPE 107468, presented at the 2007 SPE Latin American and Caribbean Petroleum Engineering Conference held in Buenos Aires, Argentina, 15-15 April, 2007.
- Dan W. Patterson: "Artificial Neural Networks: theory and application.", Singapore, New York, Prentice Hall, ISBN 0132953536, 1996
- Don W. Green and G. Paul Willhite: *Enhanced Oil Recovery*. SPE Textbook Series, 1998.
- Ertekin, T., Abou-Kassem, J.H., and King, G.R.: *Basic Applied Reservoir Simulation*.SPE Textbook Series Volume 7, Henry L. Doherty Memorial Fund of AIME, Richardson, TX, ISBN 1-55563-089-8, 2001.
- Fausett, L.: *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice-Hall, Englewood Cliffs, NJ, 1994.
- Hagan, M. T., Demuth, H. B., and Beale, M.: *Neural Network Design*. PWS Publishing, Boston, MA, 1996.
- Jeroen Regiten: "Enhanced Oil Recovery: Here to Stay." Oilfield Review. 2010/2011.
- J. K. Ali: "Neural Networks: A New Tool for the Petroleum Industry?", SPE 27561, presented at the European Petroleum Computer Conference held in Aberdeen, U.K.,15-17 March, 1994.

- Karthik Srinivasan: “Development and Testing of an Expert System for Coalbed Methane Reservoirs Using Artificial Neural Networks.”, Master’s Thesis, The Pennsylvania State University, 2008,
- Kulga, I. B.: “Development of an Artificial Neural Network for Hydraulically Fractured Horizontal Wells in Tight Gas Sands.”, Master’s Thesis, The Pennsylvania State University, 2009.
- L. E. Baker: “Three-Phase Relative Permeability Correlations.”, SPE/DOE 17369, prepared for presentation at the SPE/DOE Enhanced Oil Recovery Symposium held in Tulsa, Oklahoma, April 17-20, 1998.
- Lisa M. Belue and Kenneth W. Bauer Jr.: “Determining input features for multilayer perceptron, OH 45433, USA, 2000.
- M. F. Redondo, Castellon Espnosa: “Inputs Selection by Multilayer Feedforward Trained Netwokrs.” Vol 3, 1823-1839, 1999.
- Parada, C.H.:” An Artificial Neural Network Based Tool-Box For Screening and Designing Improved Oil Recovery Methods.” PhD thesis, The Penn State, 2008.
- Pedersen and Fredenslund: “An improved corresponding States Model for the Prediction of Oil and Gas Viscosities and Thermal Conductivities.”, Chemical Engineering Science, Vol.42, No.1,182-186,Pergamon Journals L.td, 1987
- P. K. W. Vinsome, J. Westerveld: “A Simple Method for Prediction Cap and Base Rock Heat Losses in Thermal Reservoir Simulators.” JCPT80-03-04, 1980.
- Rizvi Shihab:” Development and Testing of an Expert System Using Artificial Neural Networks for a Forward in-situ Combustion Process.” Master’s Thesis, The Pennsylvania State University, 2011.

- Sarkar et al.: “Use of Horizontal Wells for Improving Steamflood Performance of a Thin, Low-Permeability Heavy-Oil Reservoir.” SPE/DOE Improved Oil Recovery Symposium, 17-20, Tulsa, Oklahoma April, 1994.
- Shahab Mohaghegh: “Virtual-Intelligence Applications in Petroleum Engineering: Part I-Artificial Neural Networks.” SPE 58046 September 2000.
- S. M. AI-Fattah et al: “Predicting Natural gas Production Using Artificial Neural Network.”, , SPE 68593 ,presented at the 2001 SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas, 2-3 April, 2001
- STARS Advanced Process and Thermal Reservoir Simulator, Version 2008 User Guide. 2008. Calgary, Alberta: CMG.
- Steam Tables Calculator - Water and Steam Properties with HEI Condenser Performance Calculator, http://www.dofmaster.com/steam.html.
- Taber, J.J.: “Technical Screening Guides for the Enhanced Recovery of Oil, SPE 12069, SPE Annual Technical Conference and Exhibition, 1983.
- William C. Lyons and Gary J. Plisga: “Standard Handbook of Petroleum and Natural Gas Engineering.” Elsevier Science & Technology, 2004.
- Yunusa A Abass:” Determination of Cricondentherm, cricondenbar and Critical Points of Natural Gases Using Artificial Neural Networks.”,Master’s Thesis, The Pennsylvania State University, 2011.
- Ziegler, Victor M., Chevron Oil Field Research Co.: “A Comparison of Steamflood Strategies: Five-Spot Pattern vs. Inverted Nine-Spot Pattern”, Journal of SPE Reservoir Engineering, Vol 2(4), p.549-558, November, 1987.

- Zhong He, Liuyu Yang et al:" Neural –Network Approach to Predict Well Performance Using Available Field Data." SPE 68801, presented at the SPE Western Retinal Meeting held in Bakersfield, California, 26-30, March, 2001.

Appendix A

Comparisons of Simulator Runs and ANN Predictions

The validations of four developed networks are studied using the randomly distributed 20% of the samples to understand their accuracy and put forward an error range for each output parameter. Thus, six input data sets are applied for each network. The performance discussion on the networks prediction is analyzed in the Chapter 5 by one validated data set. Since the other 5 validation data sets in most cases have the similar trend and conclusion within each network, this appendix is to show their results and provide a visualized idea of each network forecast.

A.1 Network Validation Performance of PVT Model-1

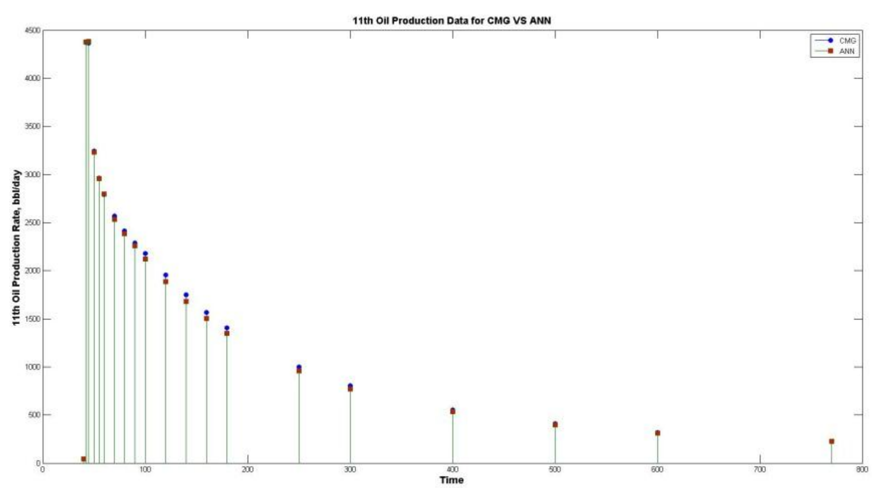


Figure A-1 CMG vs ANN Oil Production 11th Data Set

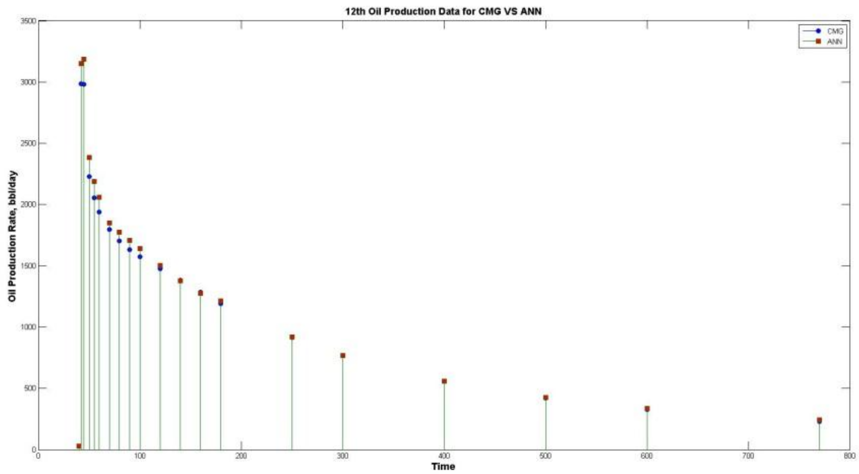


Figure A-2 CMG vs ANN Oil Production 12th Data Set

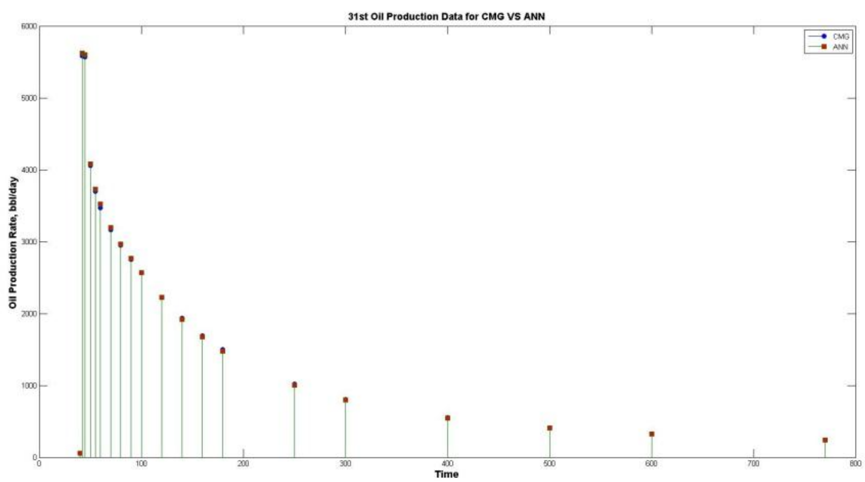


Figure A-3 CMG vs ANN Oil Production 31st Data Set

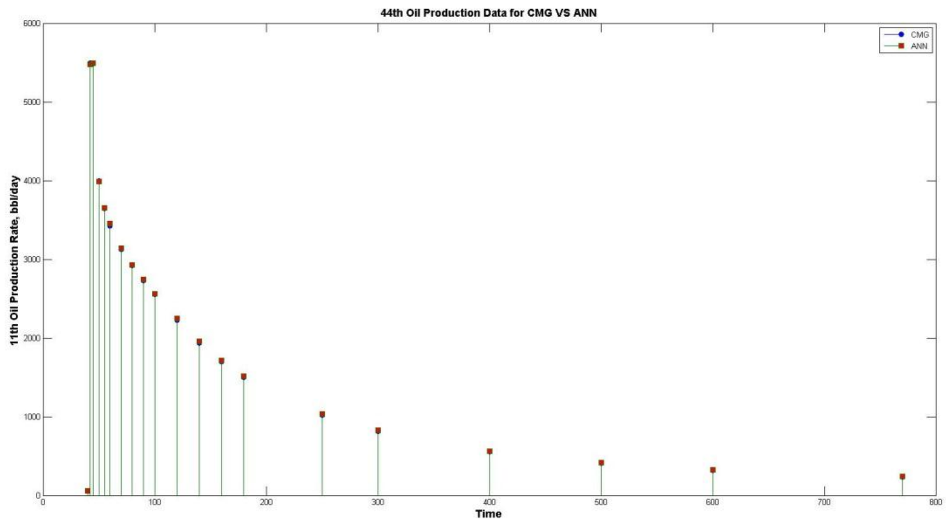


Figure A-4 CMG vs ANN Oil Production 44th Data Set

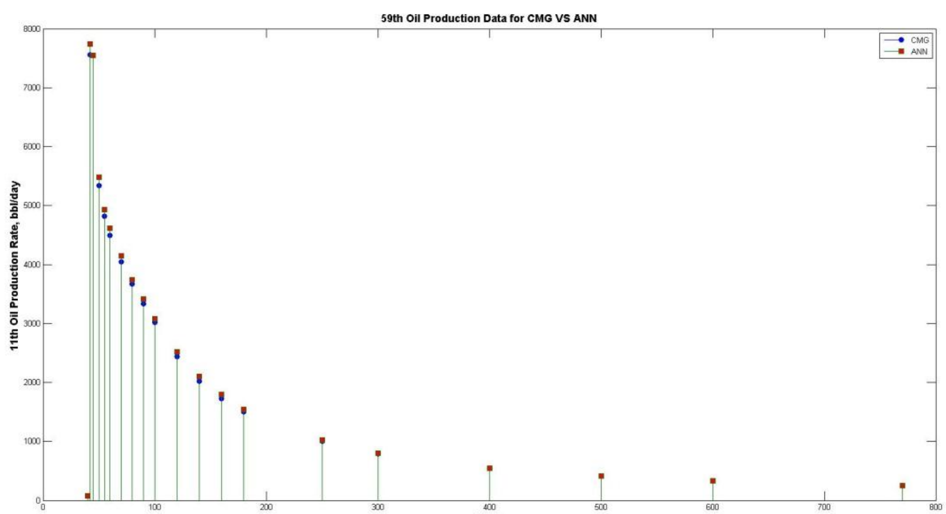


Figure A-5 CMG vs ANN Oil Production 59th Data Set

A.2 Network Validation Performance of PVT Model-2

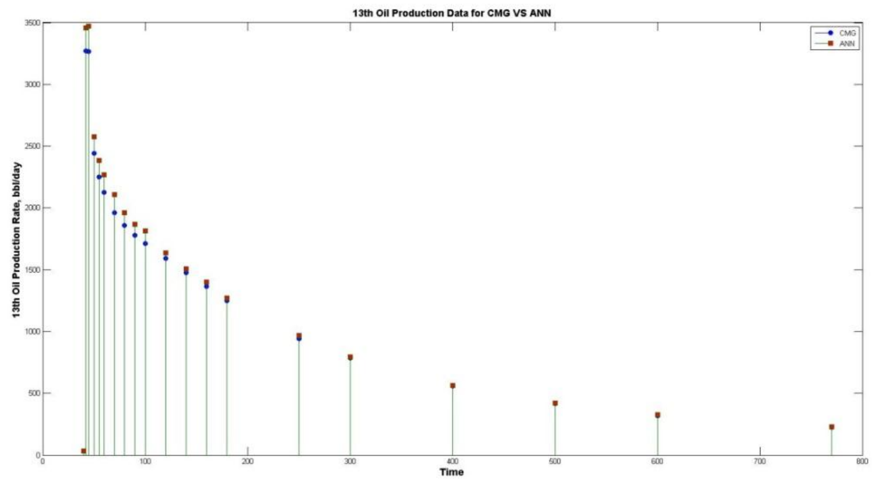


Figure A-6 CMG vs ANN Oil Production 13th Data Set

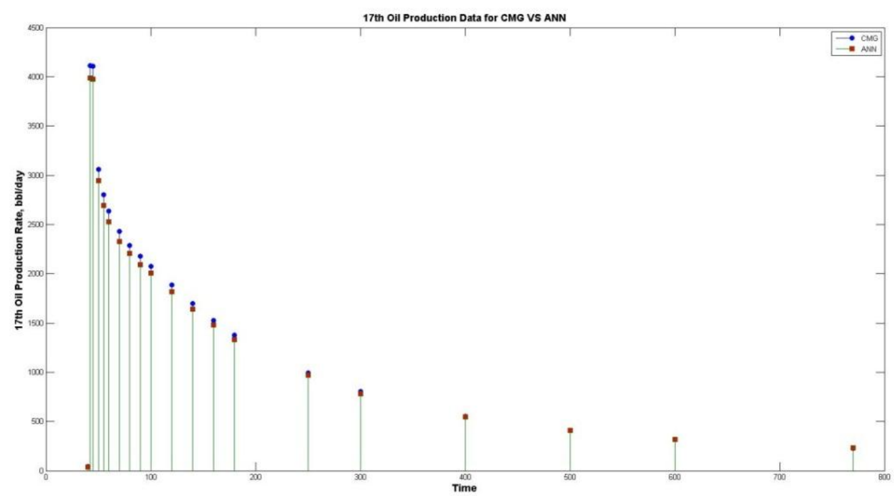


Figure A-7 CMG vs ANN Oil Production 17th Data Set

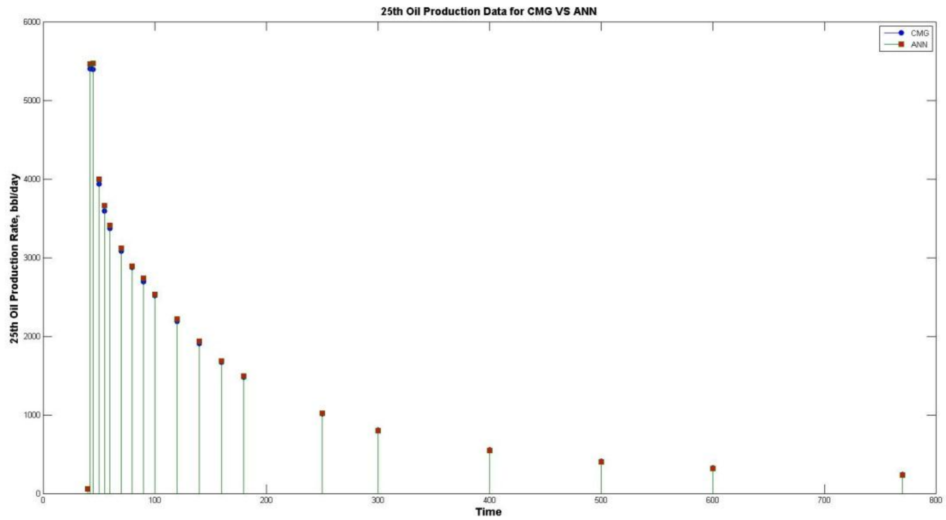


Figure A-8 CMG vs ANN Oil Production 25th Data Set

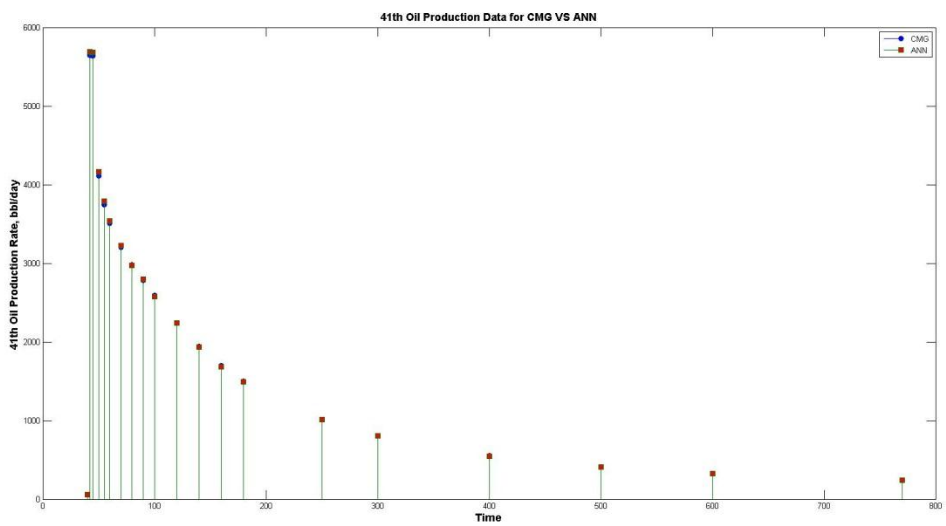


Figure A-9 CMG vs ANN Oil Production 41st Data Set

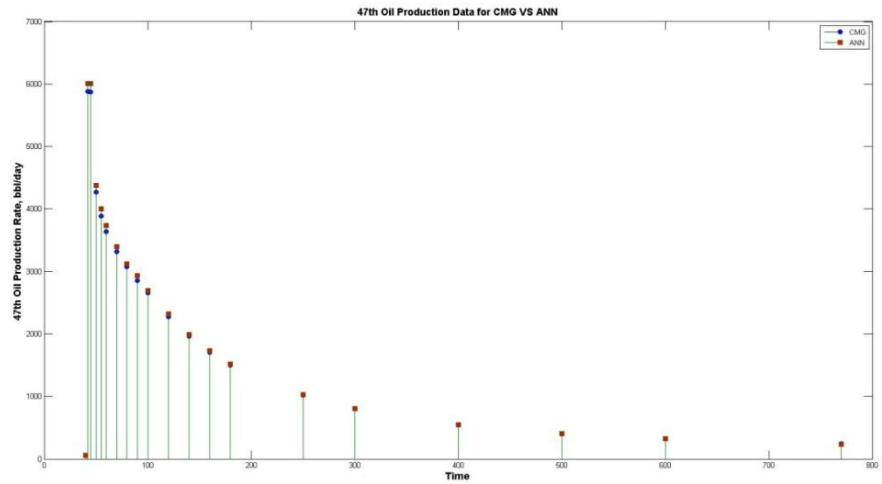


Figure A-10 CMG vs ANN Oil Production 5th Data Set

A.3 Network Validation Performance of PVT Model-3

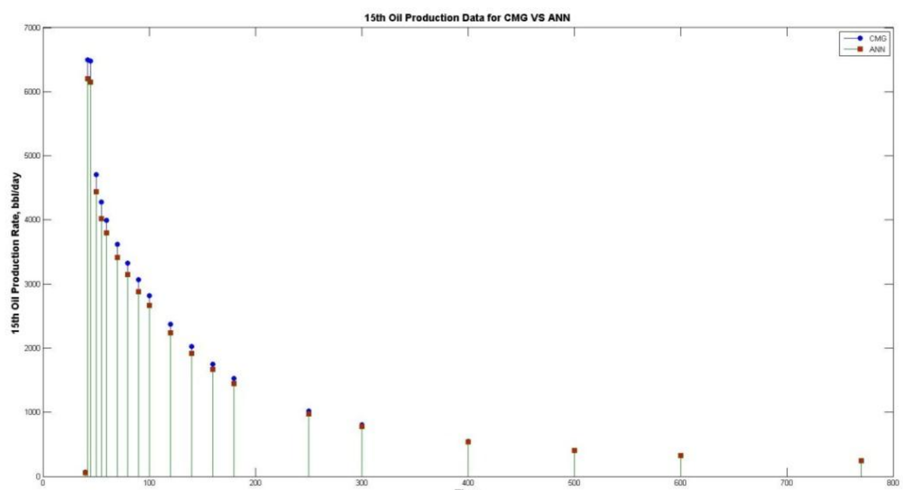


Figure A-11 CMG vs ANN Oil Production 15th Data Set

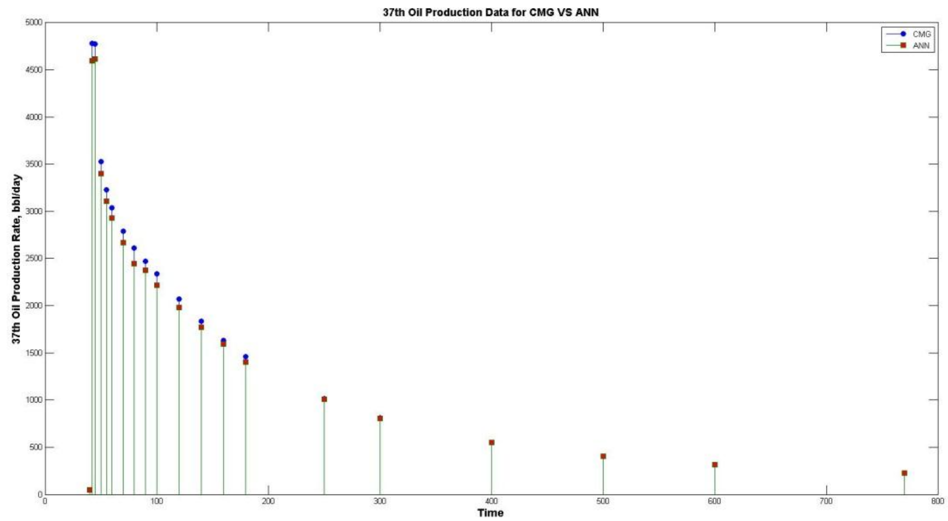


Figure A-12 CMG vs ANN Oil Production 37th Data Set

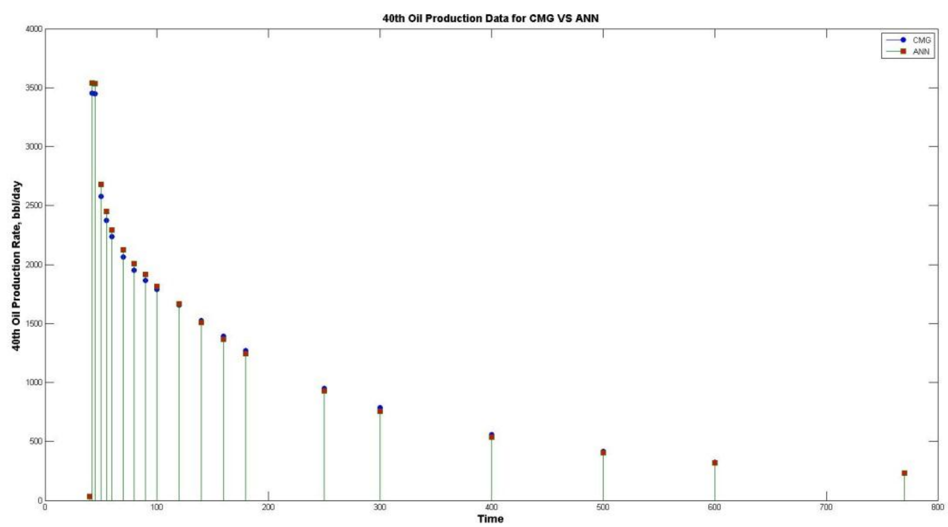


Figure A-13 CMG vs ANN Oil Production 40th Data Set

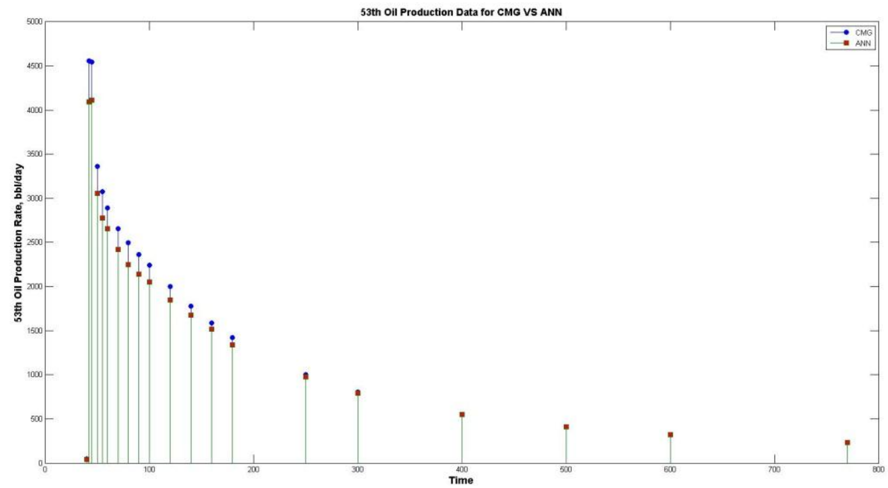


Figure A-14 CMG vs ANN Oil Production 53rd Data Set

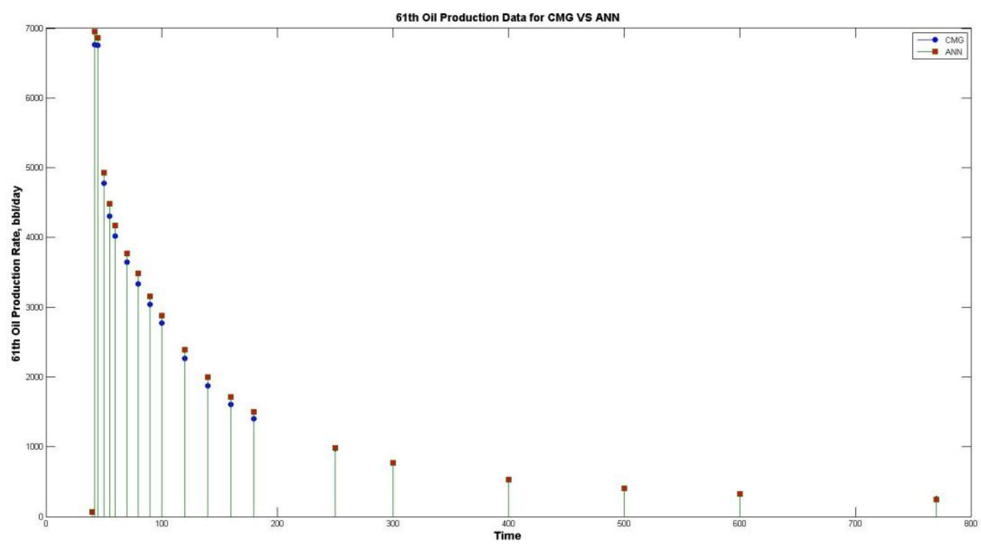


Figure A-15 CMG vs ANN Oil Production 61st Data Set

A.3 Network Validation Performance of PVT Model-4

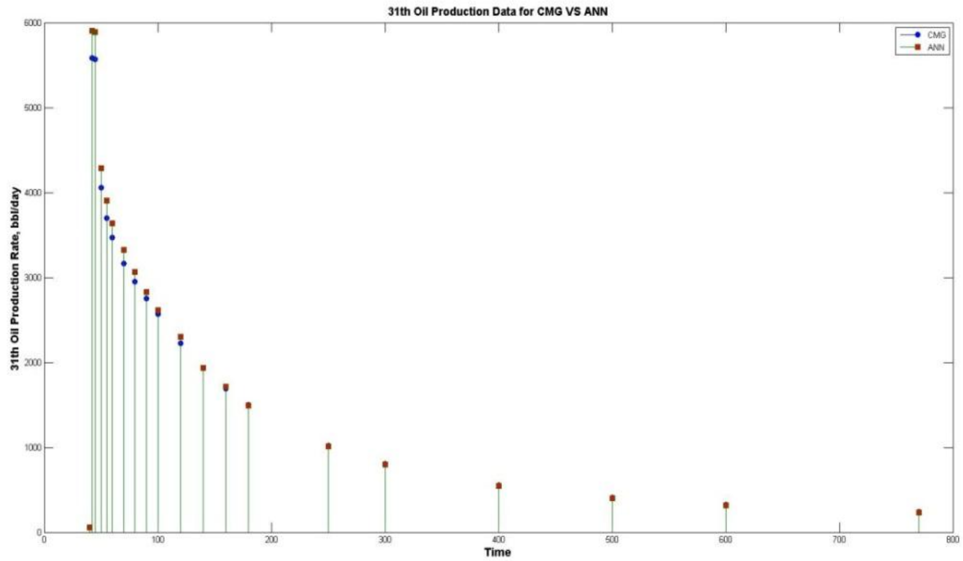


Figure A-16 CMG vs ANN Oil Production 31th Data Set

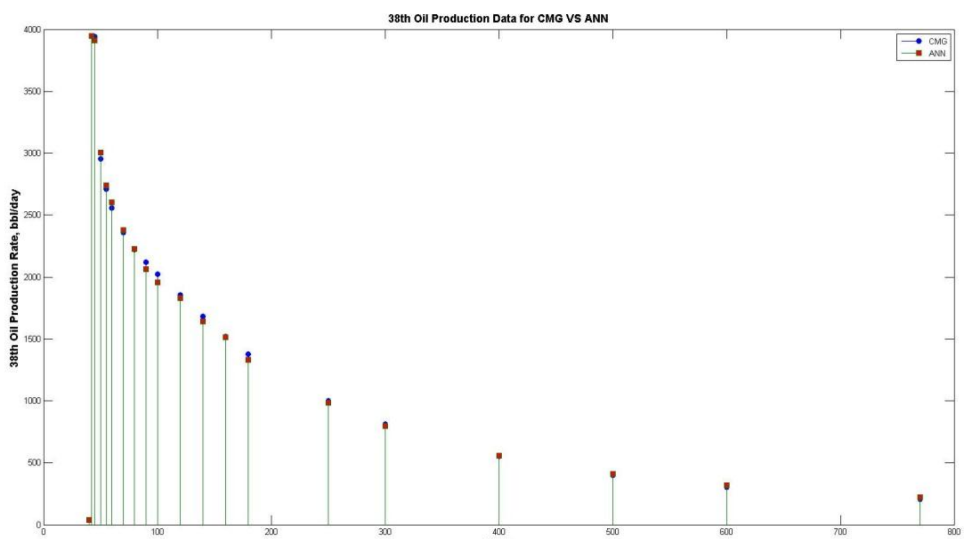


Figure A-17 CMG vs ANN Oil Production 38th Data Set

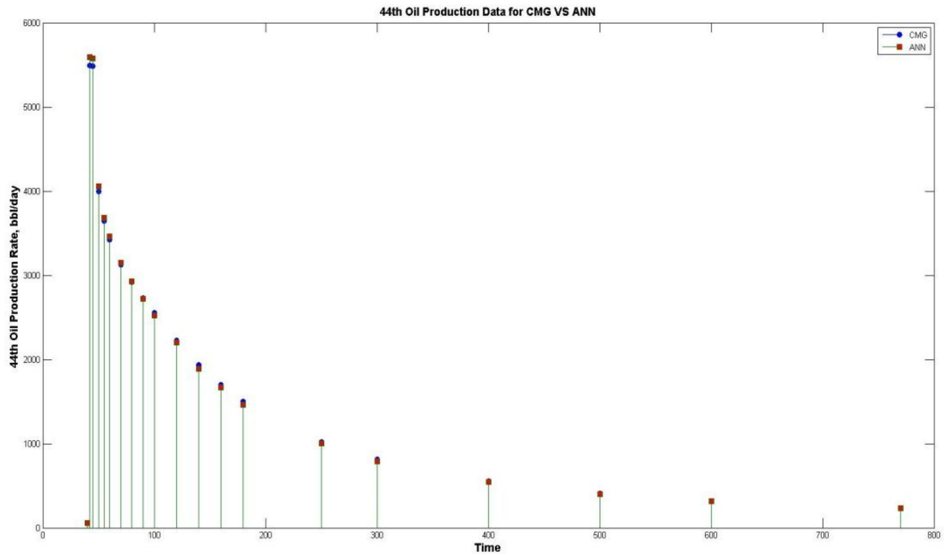


Figure A-18 CMG vs ANN Oil Production 44th Data Set

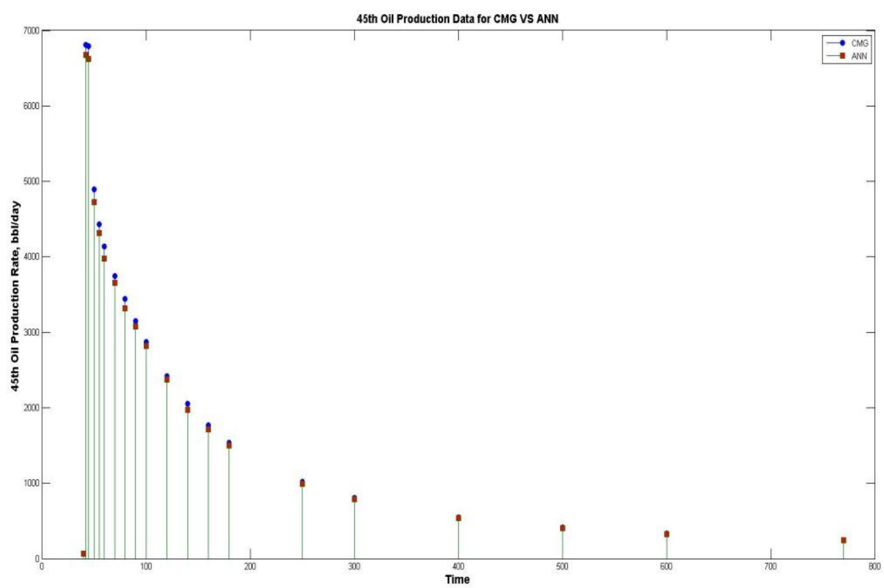


Figure A-19 CMG vs ANN Oil Production 45th Data Set

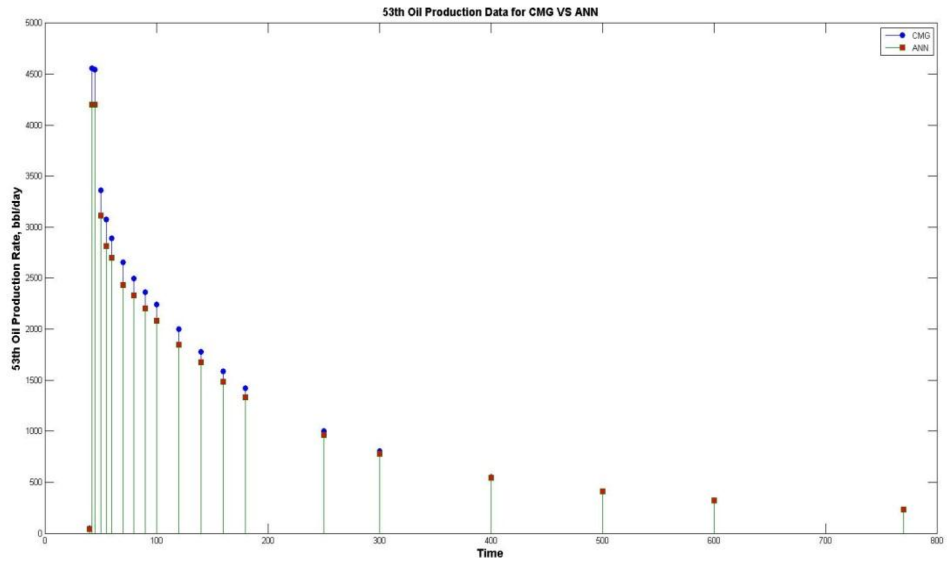


Figure A-20 CMG vs ANN Oil Production 53rd Data Set

Appendix B

Networks Testing and Validation Inputs Distribution

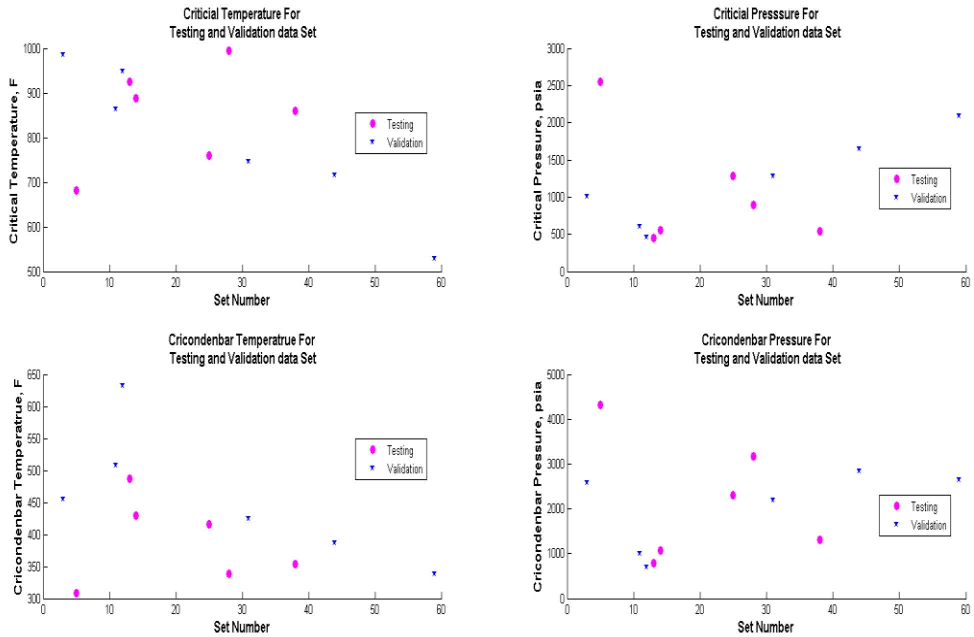


Figure B-1 Network PVT-1 Inputs Testing and Validation Distribution-1

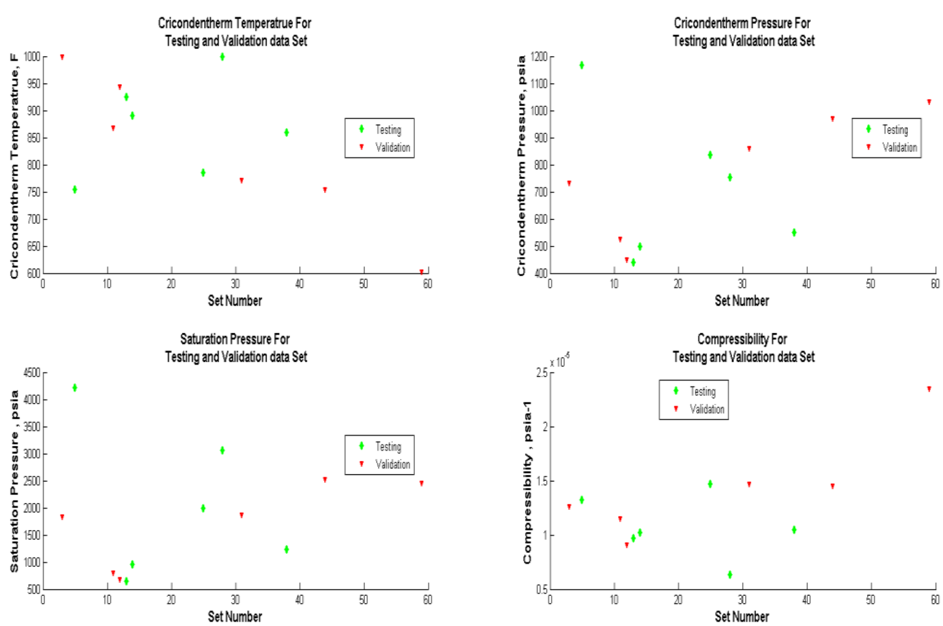


Figure B-2 Network PVT-1 Inputs Testing and Validation Distribution-2

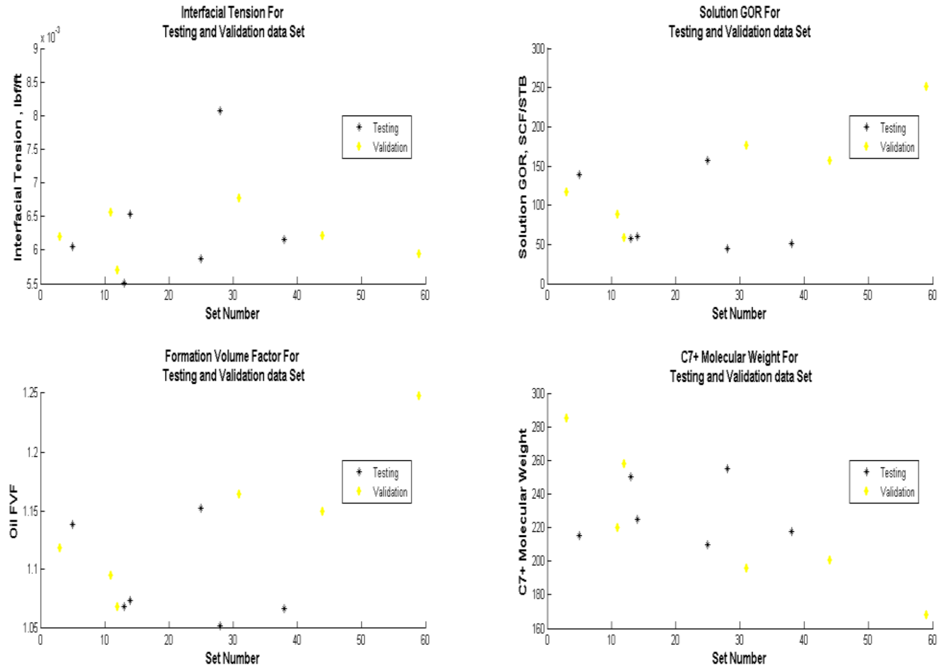


Figure B-3 Network PVT-1 Inputs Testing and Validation Distribution-3

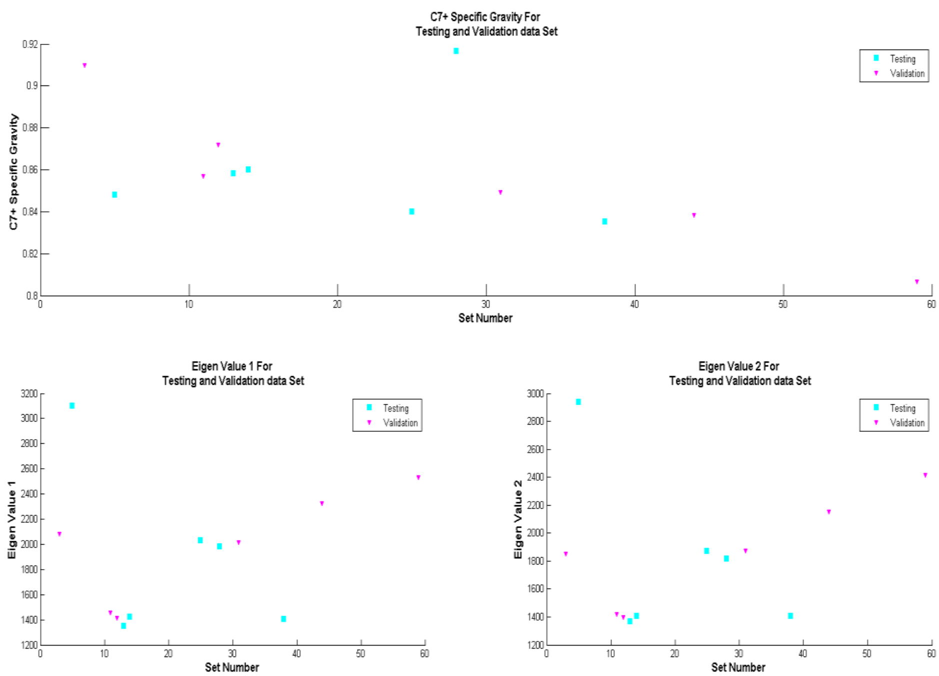


Figure B-4 Network PVT-1 Inputs Testing and Validation Distribution-4

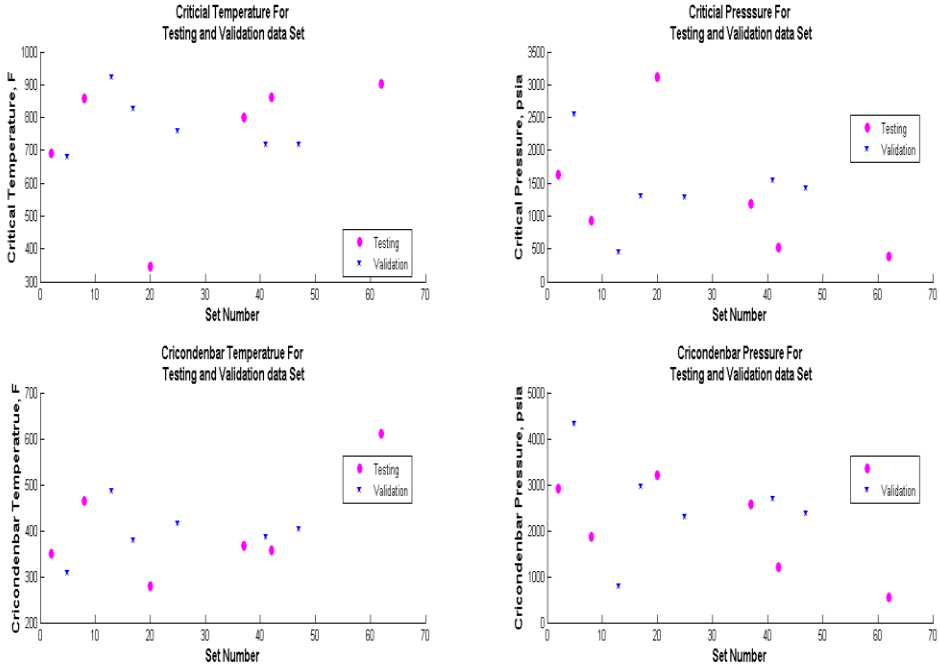


Figure B-5 Network PVT-2 Inputs Testing and Validation Distribution-1

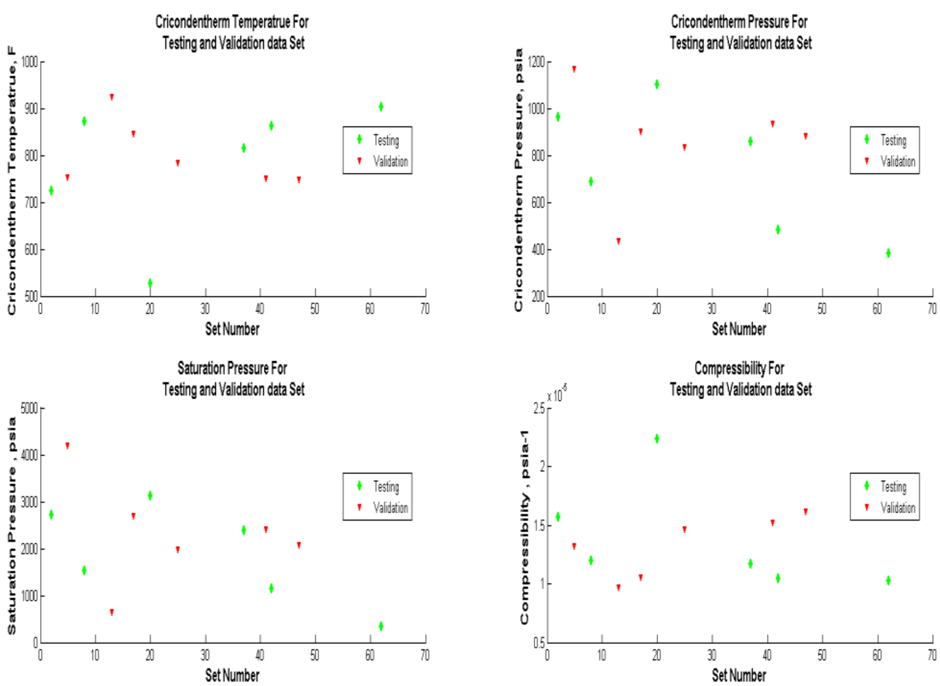


Figure B-6 Network PVT-2 Inputs Testing and Validation Distribution-2

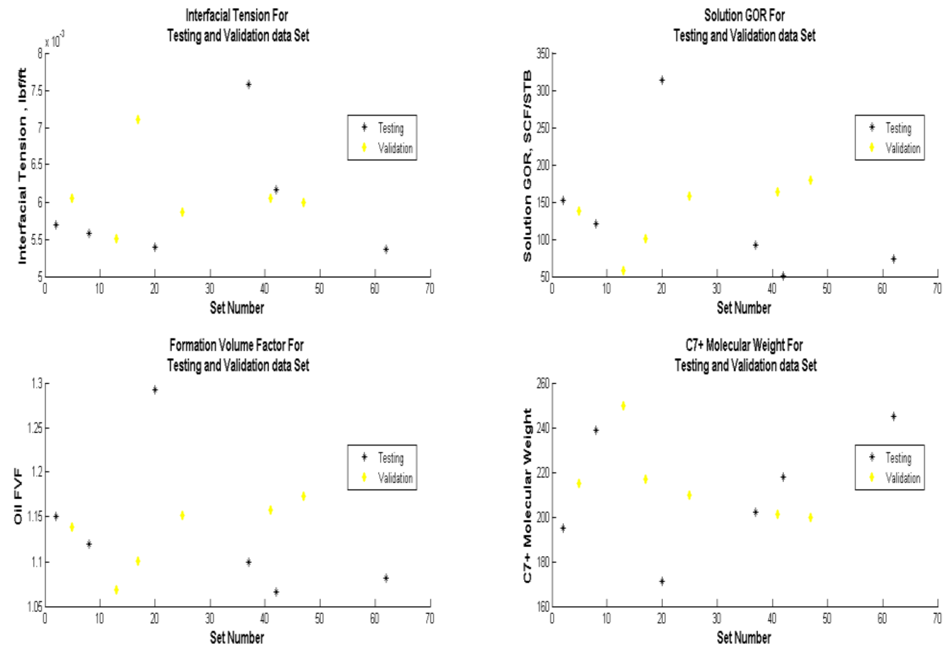


Figure B-7 Network PVT-2 Inputs Testing and Validation Distribution-3

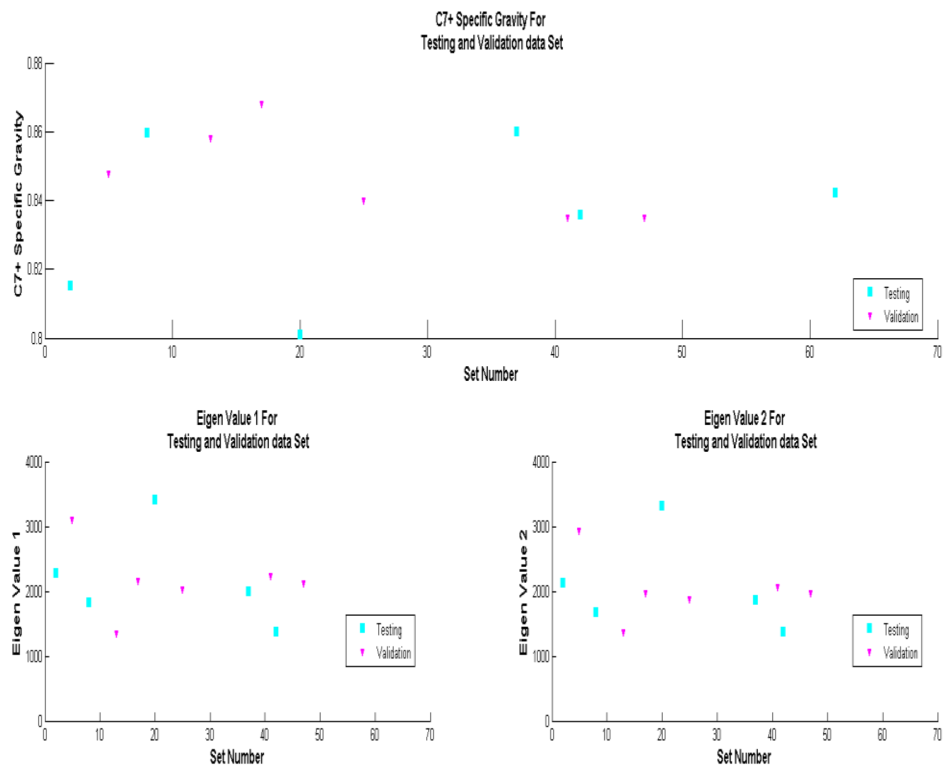


Figure B-8 Network PVT-2 Inputs Testing and Validation Distribution-4

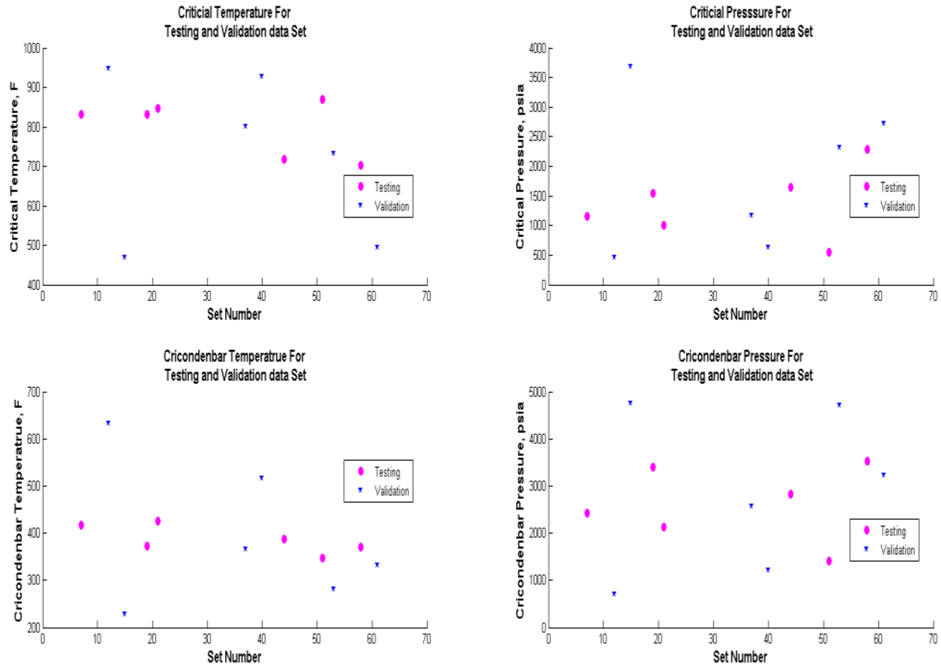


Figure B-9 Network PVT-3 Inputs Testing and Validation Distribution-1

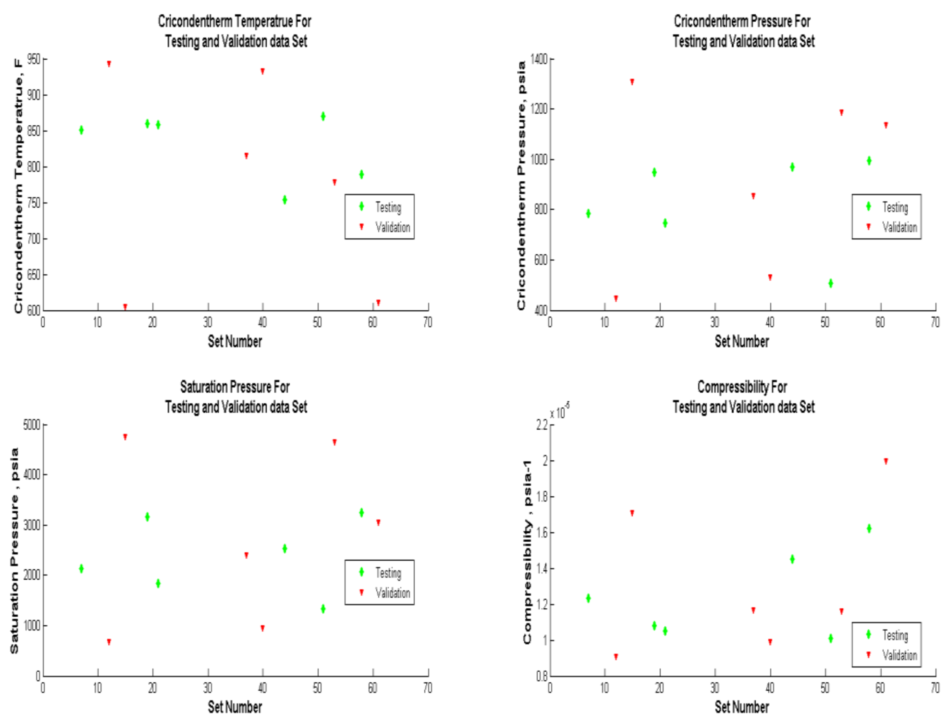


Figure B-10 Network PVT-3 Inputs Testing and Validation Distribution-2

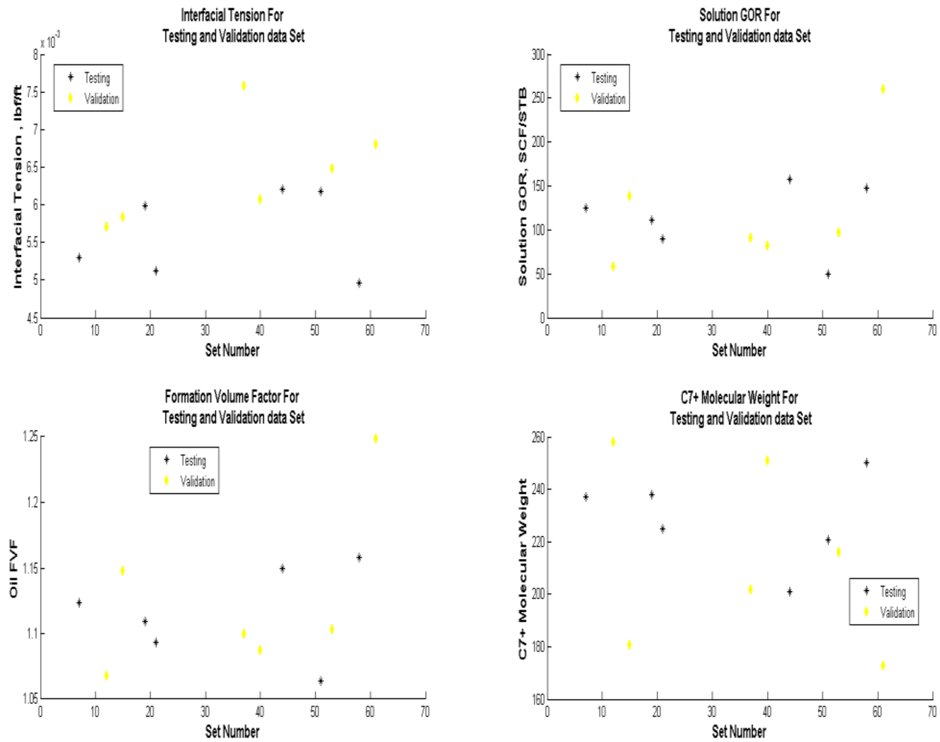


Figure B-11 Network PVT-3 Inputs Testing and Validation Distribution-3

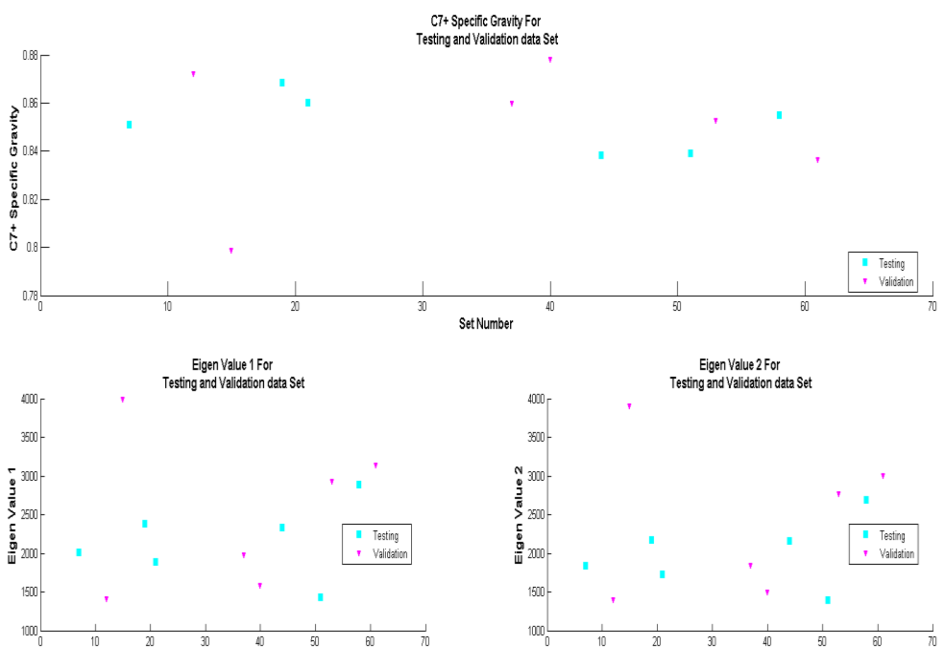


Figure B-12 Network PVT-3 Inputs Testing and Validation Distribution-4

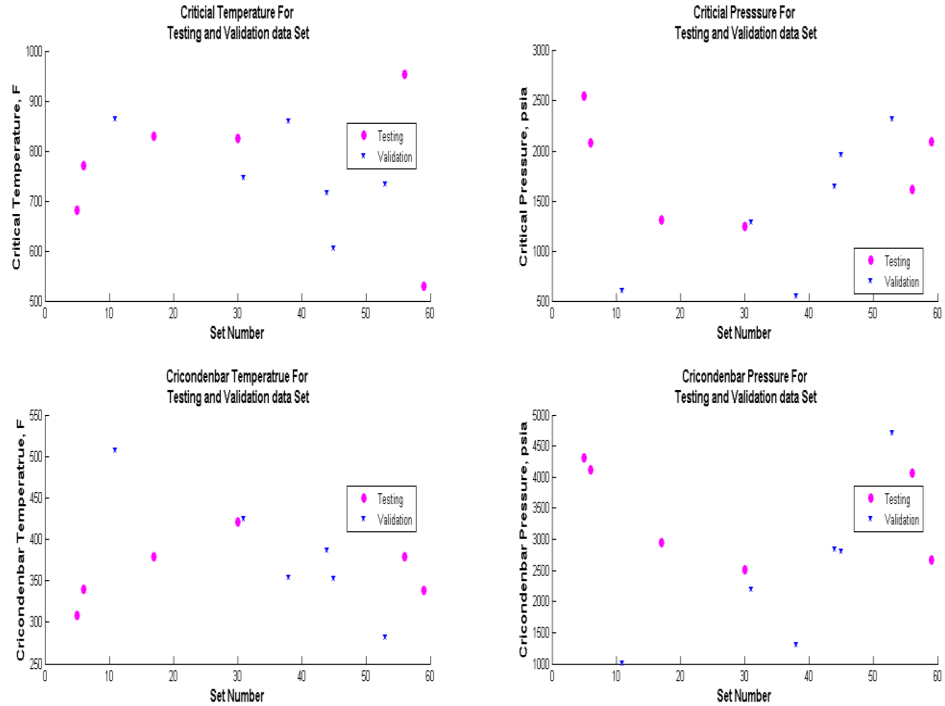


Figure B-13 Network PVT-4 Inputs Testing and Validation Distribution-1

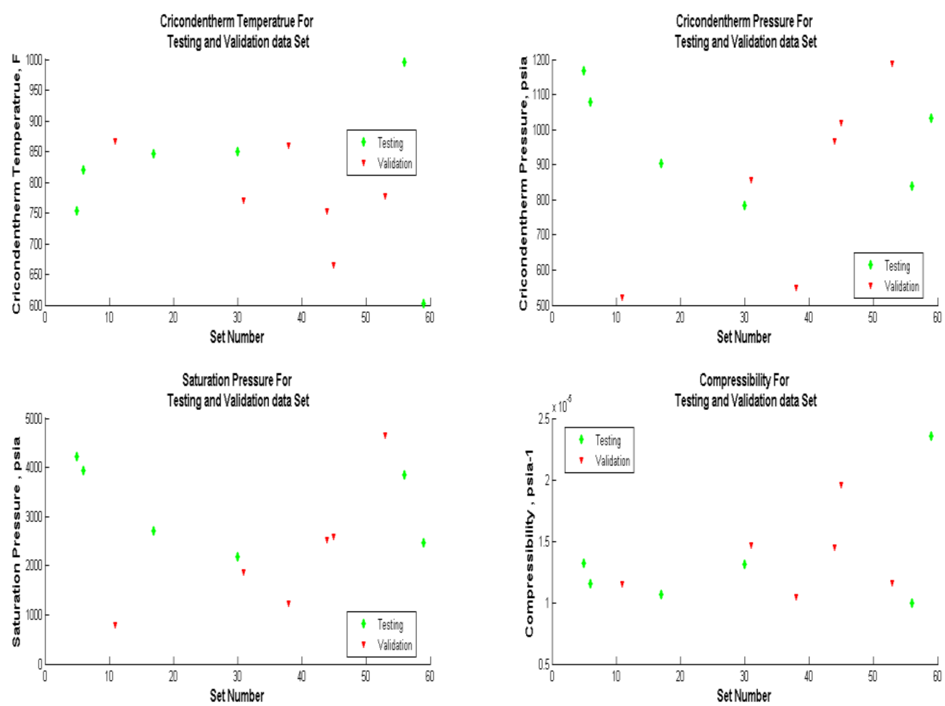


Figure B-14 Network PVT-4 Inputs Testing and Validation Distribution-2

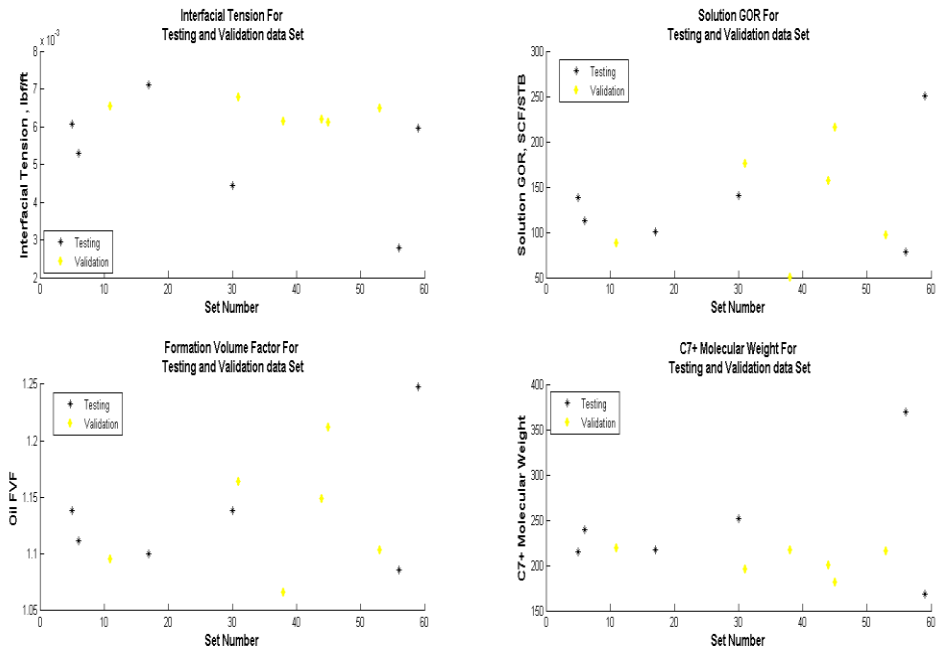


Figure B-15 Network PVT-5 Inputs Testing and Validation Distribution-3

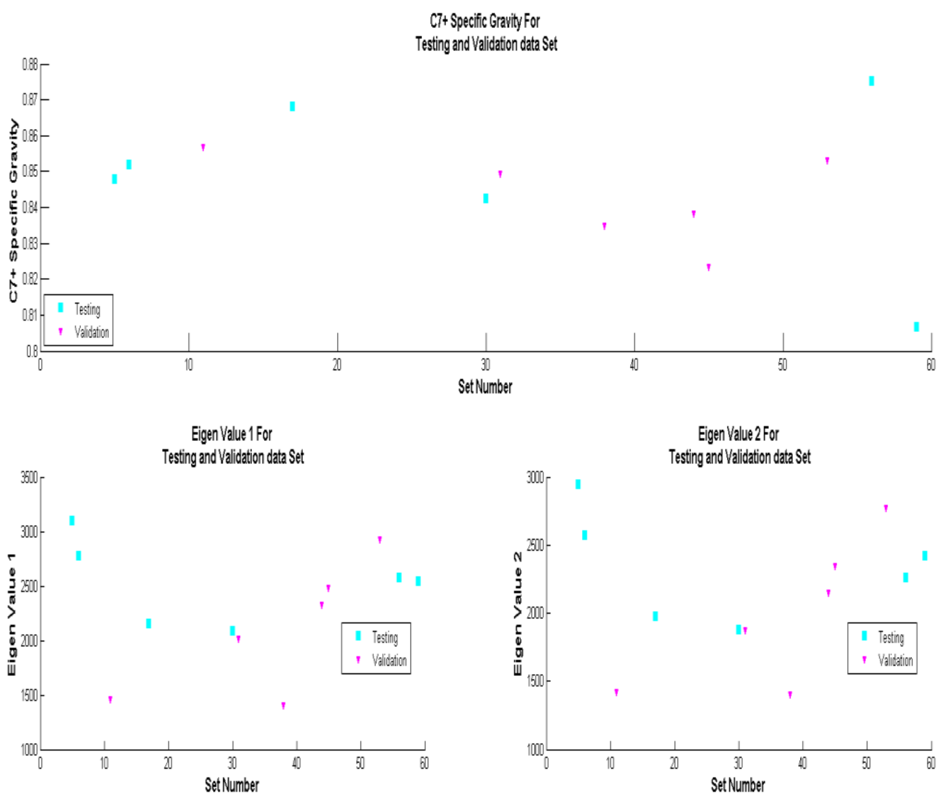


Figure B-16 Network PVT-4 Inputs Testing and Validation Distribution-4