Conceptual framework for using system identification in reservoir production forecasting

by

Shodiq Khoirur Rofieq

17019

Dissertation submitted in partial fulfillment of
the requirements for the
Bachelor of Engineering (Hons)
(Petroleum)

MAY 2015

Universiti Teknologi PETRONAS

Bandar Seri Iskandar

31750 Tronoh

Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Conceptual framework for using system identification in reservoir production forecasting

by

Shodiq Khoirur Rofieq
17019

A project dissertation submitted to the
Petroleum Engineering Programme
Universiti Teknologi PETRONAS
in partial fulfillment of the requirement for the
BACHELOR OF ENGINEERING (Hons)
(PETROLEUM)

Approved by,

____

(Mr. Berihun Mamo Negash)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
MAY 2015

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

SHODIQ KHOIRUR ROFIEQ

ABSTRACT

System identification is the concept of utilizing statistical models to describe dynamic systems. System Identification modelling techniques are already popular in many science fields such as control & signal processing, process control, GPS tracking and economics. However, the complexity of a petroleum reservoir system and the availability of numerous model structures in system identification make it challenging to adapt this method effectively for petroleum engineering purposes.

In this thesis, a conceptual framework for using system identification is proposed. Based on a reservoir's recovery mechanism, the conceptual framework will help to systematically select an appropriate model structure from the various model structures available in system identification. This model can then be used to identify the reservoir for the purpose of forecasting fluid production. Only linear system identification models will be considered for identification in this study and special emphasis will be put on polynomial models. Only primary and secondary drive mechanisms will be investigated in this study.

For each recovery mechanism, a synthetic reservoir simulation model is made and run to generate input and output data for the system identification process. Next, for each recovery mechanism, MATLAB software is used to identify the system identification model(s) that can best forecast three important production parameters based on the input and output data. These parameters are field oil production rate, field water cut and field gas-oil-ratio. Lastly, a framework is created by analyzing and matching each recovery mechanism to their best system identification models.

The results show that System identification polynomial models can provide very accurate models to predict oil rate, water cut and GOR curves for reservoirs under the drive mechanisms listed in this study. System identification based reservoir models can be established as a practical, cost-effective and robust tool for forecasting reservoir fluid production. The procedures described in this thesis as well as the final conceptual framework can serve as a guide to reservoir engineers who wish to use system identification for forecasting production.

ACKNOWLEDGEMENTS

All praise to God, the Most Gracious for giving me the means to complete my final year project successfully under the guidance of my supervisor, Mr. Berihun. I would like to express my highest gratitude and appreciation to my supervisor for always being supportive of my work and for always being patient with me. Thanks to him, I was able to solve all the problems I encountered in the process of proposing a new methodology for production forecasting.

Also, with Mr. Berihun's assistance, I was able to enter my project into the 35th Science Engineering Design and EXhibition (SEDEX) and win the gold medal award under the Geoscience and Petroleum Engineering category. This award highlights that we have the best project in the faculty and it also recognizes all the hard work we have put into the project.

I also want to thank the coordinators of FYP1 and FYP2, Mr. Titus and Mr. Asif respectively, for their continuous support, understanding and for granting us final year students deadline extensions when necessary. Lastly, I wish to thank my internal examiners, Dr. Mahbubur and Dr. Maqsood, for all their feedback, advice and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

Chapter 1: Introduction

1.1 Importance of forecasting

One of the most important jobs of reservoir engineers, in principal collaboration with production engineers, is to predict or forecast future fluid production rates. Forecasting is an integral part of reservoir management as it allows us to estimate the upcoming production profile. This in turn allows us to meet numerous objectives, some of which include:

- Evaluating the economics of developing the reservoir: The economic feasibility of any project in the oil and gas industry will undoubtedly depend on the amount of hydrocarbon fluid production among other things. Hence, a forecast of the production rates will allow the oil company to evaluate whether or not a certain reservoir is a suitable candidate for development. Even after preliminary forecasts show that reservoir rates would be economic, forecasting production during development will still be useful as it enables the company to constantly re-assess the project economics and help plan suitable recovery techniques that will optimize net present value (NPV). Moreover, prediction will help to estimate project life (Spencer & Morgan, 1998).

- Planning the required equipment and facilities: The expected production profile will determine the design of the well (casings, tubings, perforations). It will also determine the design of surface facilities (pipelines, separators, storage and transport) required to handle and process the amounts and types of fluids produced (Hickman, 1995).

- Planning each well's completions and the regularity of work-over process: Implementing certain geometries of completions can help to optimize the production rates by reducing the damage (skin) to the formation near the wellbore. Similarly, the production tubing and the casings may need to be replaced at times to be able to cater to an optimized production rate (Spencer & Morgan, 1998).

- Evaluating strategies to boost production: If the forecast shows that rates will decline or become uneconomical under current production methods, then alternative recovery strategies may be taken into consideration to improve the rates. Some of these strategies may include implementing well stimulation techniques such as hydraulic fracturing, or others such as secondary and tertiary recovery mechanisms.

- Evaluate well performance and effectiveness of operations: Assuming that we have a reliable forecast, by comparing actual future production performance against the estimated production profile, we can evaluate whether or not the well is producing to its full potential (Lockwood & Cannon, 1982). Additionally, the effectiveness of well operations (work-over, well stimulation, injection for enhanced recovery) can be evaluated. For example, if forecasted rates after stimulation are lower than expected, this might be an indication that the skin has not been sufficiently reduced and that the stimulation operation was not successful. Similarly, comparison of expected and actual rates can help to evaluate the effectiveness and success of EOR processes such as water or gas injection.
- To help understand the reservoir behavior better: In normal practice, a computer reservoir model would be built at the beginning of field development by considering all the static and dynamic data available (reservoir characterization & simulation) and by implementation of history matching techniques to obtain uncertain parameters. The forecast obtained from any of the forecasting techniques (discussed in literature review) would have either made use of the computer model itself, made use of some of its parameters (examples: porosity, permeability) or would have made use of assumptions derived from analysis of the computer model. This is because the computer model is considered the best mathematical representation of the true system (reservoir). Hence, if the actual production performance differs from our forecast, this can be an indication that something is wrong with our computer model. The reasons for any discrepancies can be studied and the computer model updated accordingly for the purpose of reservoir characterization.

From the points above, it is clear that production forecasting has many important uses and that there is a need to get good forecasts simply because they would be the basis for important decisions (Mannon, 1964). Therefore, the primary purpose of this project is to introduce a novel method for reservoir production forecasting called system identification that has been widely used in many other engineering fields.

1.2. Background of system identification

System identification is the concept of utilizing statistics to describe a dynamic system. This is done by inferring a statistical model (SI model) based on the observations (the inputs and outputs) of the dynamic system and/or based on a prior knowledge of the system (Keesman, 2011). There are numerous SI models available and selecting the most suitable model to describe the system would require experimentation and engineering knowledge of the system. System identification has proven to be a valuable tool in many fields of science, including electrical and electronics engineering, chemical engineering, civil engineering and economics fields. One of the most well-known applications of system identification includes the forecast of future outputs of the system once a suitable SI model has been selected.

System identification is still a new concept in petroleum engineering. Literature review shows that it has previously been used by researchers to predict water cuts (Renard, Dembele, Lessi & Mari, 1998) as well as optimize production rates (Elgsaeter, Slupphaug & Johansen, 2008). However, these studies only focused on a certain recovery mechanism (reservoir state). This is because there is a limitation to system identification, which is that the system has to be in the in the same state during the course of the observations, i.e. no variations (Renard, et al. 1998). Additionally, in direct contrast to reservoir simulation, system identification treats the system under study as either a grey box model or a black box model. A possible input signal for reservoirs is injection rate of displacing fluid while a possible output signal could be fluid production rate or any other production parameter. Typically, inputs are linked to outputs through functions and not by considering physical phenomena. Even though the true physics of the system is not being considered, system identification can be an efficient method for prediction, especially when bearing in mind that it takes considerably shorter time to implement compared to history matching & simulation.

1.3. Problem Statement

The most established methods for reservoir production forecasting are decline curve analysis (DCA), reservoir simulation and material balance. Each method has its own merits and its own limitations (Olominu & Sulaimon, 2014). It is a usual practice to implement more than one method to reduce uncertainty and increase efficacy. Hence, any new proposed method would be complementary to the existing methods and is not meant to replace them.

Currently, there is still a need for a set of simple, quick and flexible modelling techniques that can also be used when the reservoir description is limited. The value of such modelling techniques would be in the exploration and early field development stage, where reservoir data is still inadequate and computer models are still unreliable. Hence, it is the interest of this project to investigate a prediction method that has been proven to be very effective outside the petroleum engineering field.

This is a new area of application for system identification. A thorough study is required to identify its potential use in reservoir performance prediction. Moreover, because there are many possible drive mechanisms (system states) that a reservoir can be under, it is expected that no one SI model would be capable of adequately describing all the mechanisms. In other words, different mechanisms might be best described by different SI models. Therefore, there needs to be a framework to which petroleum engineers can refer to when required to fit an SI model to their reservoir system.

1.4. Objectives

The major objectives of this project are:

I. To design and develop a conceptual framework for implementation of system identification techniques.

II. To describe and classify reservoirs into distinguishable recovery mechanisms and to associate system identification models to each drive mechanism.

III. To evaluate the efficacy of the proposed method (SI).

1.5. Scope of Study

This study will be limited by the following considerations due to time constraints:

1) System identification will be implemented on primary and secondary reservoir recovery mechanisms only.
2) Only linear SI models will be considered in this study and special focus will be put on polynomial models.
3) Working data will come from synthetic models unless real field data can be obtained. Synthetic models will be limited in size to less than 10,000 grid-blocks due to limitations placed on university licenses for commercial simulator software.
4) The end product will be a visual diagram (framework) linking drive mechanisms to their respective verified SI models (Equations).

Chapter 2: Literature Review and Theory

2.1. System identification

A dynamic system can be defined as an object in which certain variables interact together to produce outputs. Moreover, the current output value should depend on several things, namely the inputs to the system, the disturbances to the system that can't be measured and the values of previous outputs (Ljung, 1987). A petroleum reservoir can be classified as such a system.

A description of the system identification process is as follows:

1. Conducting an experimental design. Before carrying out any experiment on the system to obtain input and output data, there needs to be careful planning of the process to collect information (Peirce, 1983). Inherent in this process would no doubt be the determination of suitable inputs and outputs for the identification process. Planning the experiment is needed to be able to make sure that the data observed would give maximum information on the system (Ljung, 1987).

2. Carrying out the experiment and obtaining the observed data (input and output data).

3. Since there are too many models available in system identification, they need to be narrowed down to a suitable number of models that has the potential to describe the relationship between observed data. This process requires involvement and judgment from the engineers who will have to use their knowledge of the system to cut out any models that would not be a good representation of the system (Ljung, 1987). Some decisions that come in this process include determining whether the model will be parametric or non-parametric and whether or not a linear or nonlinear model is most suitable. Parametric models (grey box models) contain parameters that have a direct connection to a physical quantity in the system (such as porosity) while a non-parametric models (black box models) has infinite parameters that does not relate to variables in the real system (Nelles, 2001). A linear system is one which follows the superposition principle; hence its outputs and inputs are directly proportionate to each other. Meanwhile, non-linear systems do not follow the superposition principle (Billings, 2013).

4. Training: The shortlisted model from step 3 should be provided with training data and then, using a suitable algorithm (or a toolkit such as MATLAB), the data is used to adjust the free coefficients in the model. There is also a need to decide on the model order number that will give the highest accuracy. Analysis of the best order model is usually conducted before moving to the prediction stage (Mathworks, 2015).

5. Prediction and Validation. The model is let to predict using the coefficients already tuned during training stage – step 4 (*Figure 1*). This is the process of testing the accuracy of the SI model by seeing how well it predicts the system's output signal. The SI model's predicted values are compared to observed data and the accuracy is calculated using a criterion, such as the Normalized Root Mean Squared Error (NRMSE) criterion.

   From this step, it is evident that there would be two sets of real output data required, one for step 4 and another for step 5. Hence, the output data obtained in step 2 needs to be divided into data for training (for step 4) and data for validation (for step 5). This method of validation is often referred to as cross validation (Browne, 2000).

6. Looping back to step 3. System identification is a natural looping process (Ljung, 1987). There is a big chance that the model does not pass validation at the first try because it cannot produce satisfactory forecasting accuracy. Hence, there needs to be additional iterations until a satisfactory model is obtained. In each new iteration, the model is either modified by changing model order number or completely replaced with a different model type (structure). This is done in the hopes that the next iteration will bring better forecasting accuracy than the previous ones. *Figure 2* shows this process.

7. Once a model or set of models have been tested and validated, those models that provide the highest accuracy can be deemed worthy for prediction.

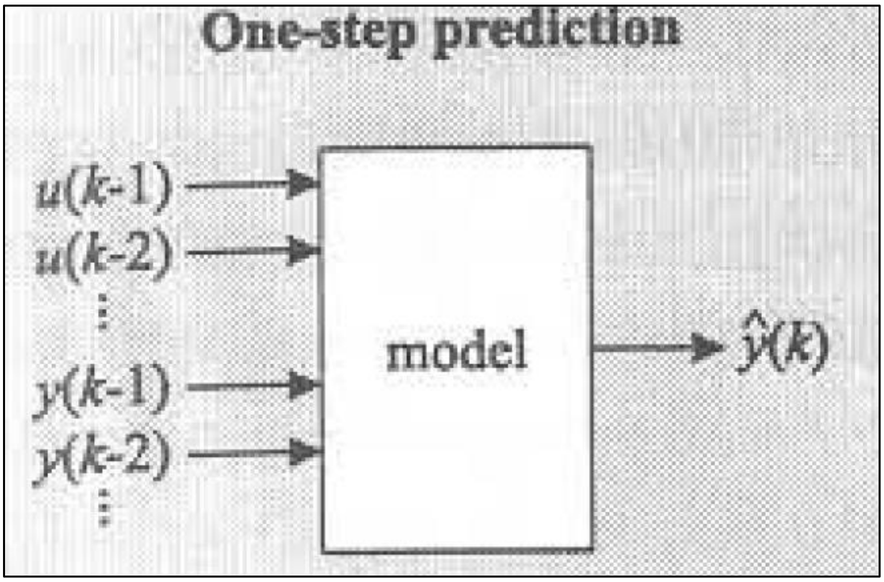

Figure 1: Graphical representation of the prediction process after training stage

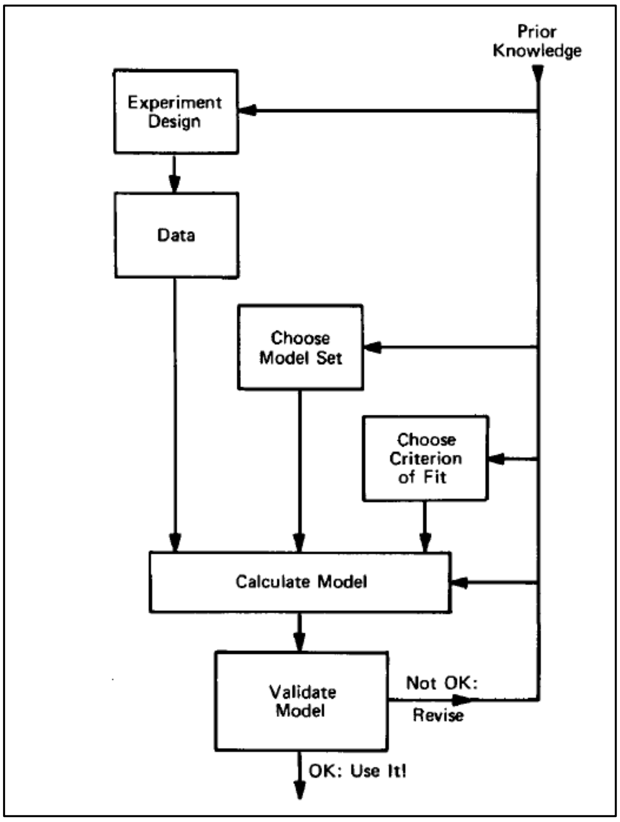

Figure 2: Graphical representation of the system identification process

2.2. Examples of system identification being used in many different industries

Some common uses of system identification include modelling the following physical systems (Instruments, 2010):

- Power systems - amplifier systems, circuits and others
- Electromechanical systems - robot arms, motor models, hydraulic systems and others
- Civil systems (structures) – bridges, buildings and others
- Process systems - chemical reactions, thermal processes and others

Examples of other uses for system identification

- ARIMA models are currently used to predict solar irradiance (Brabec, et al. 2015).
- Novel black- box prediction techniques are being proposed to predict the energy consumption and performance of storage devices (Prada, et al. 2013).
- Fuzzy black-box models are being used to predict indoor illuminance inside buildings for the purpose of energy conservation (Logar, et al. 2014).
- Statistical modelling techniques are being proposed used to model power distribution transformers (Papadopoulos, et al. 2015)
- Black-box models are being used to provide predictive control for non-linear models (Grancharova, et al. 2011).
- System identification is being used to control the process of cooking corn snacks products (Haley, et al. 2000)
- Finite impulse response models are being proposed to assess cerebral autoregulation for the purpose of maintaining stable flow of blood (Angarita-Jaimes, et al. 2014).
- System identification models is being proposed to quantify the influence of several economic parameters to trading activity (Criner, 2008)

Some examples of system identification applications in oil and gas industry are listed below:

- Artificial neural networks were used to predict the integrity of downhole casings when corrosion logging data is missing for the purpose of well integrity surveillance (AlAjmi, et al. 2015).

- ARIMA models were used to predict reservoir production (Olominu & Sulaimon, 2014).
- Artificial intelligence and data mining techniques were proposed for the purpose of history matching (Shahkarami, et al. 2014).
- Artificial neural networks were used to forecast the production of advanced well structures and designs (Enyioha & Ertekin, 2014).
- Artificial neural networks were used to obtain permeability predictions by making use of log measurements (Anifowose, et al. 2013).
- Black box models were used as an interpolation technique to obtain initial guesses of pressure solutions for the purpose of speeding up reservoir simulations (Chen, et al. 2013).

Currently, as shown by the examples above, artificial neural network is the most investigated SI technique for prediction in the oil and gas industry. However, system identification has many modelling techniques under it and there has been no research done yet to map the large variety of SI modelling techniques to the numerous reservoir drive mechanisms for the purpose of production forecasting. Hence, it can be concluded that system identification has not been investigated as much in the oil and gas industry as it is for other industries such as process industry.

2.3. System identification models – Polynomial models (Mathworks, 2015).

A polynomial model uses a generalized notion of transfer functions to express the relationship between the input, u(t), the output y(t), and the noise e(t) using the equation:

$$A(q)y(t) = \sum_{i=1}^{nu} \frac{B_i(q)}{F_i(q)} u_i\left(t - nk_i\right) + \frac{C(q)}{D(q)} e(t)$$

Equation 1: General polynomial equation

The variables A, B, C, D, and F are polynomials expressed in the time-shift operator q^-1. u_i is the ith input, nu is the total number of inputs, and nk_i is the ith input delay that characterizes the transport delay.

In practice, not all the polynomials are simultaneously active. Often, simpler forms, such as ARX, ARMAX, Output-Error, and Box-Jenkins are employed. Scientists also have the option of introducing an integrator in the noise source so that the general model takes the form:

$$A(q)y(t) = \sum_{i=1}^{nu} \frac{B_i(q)}{F_i(q)} u_i\left(t - nk_i\right) + \frac{C(q)}{D(q)} \frac{1}{1 - q^{-1}} e(t)$$

Equation 2: Adding noise source integrator to the general equation

For estimation, scientists must specify the model order as a set of integers that represent the number of coefficients for each polynomial to include in their selected structure—na for A, nb for B, nc for C, nd for D, and nf for F. Scientists must also specify the number of samples nk corresponding to the input delay—dead time—given by the number of samples before the output responds to the input. The number of coefficients in denominator polynomials is equal to the number of poles, and the number of coefficients in the numerator polynomials is equal to the number of zeros plus 1. When the dynamics from u(t) to y(t) contain a delay of nk samples, then the first nk coefficients of B are zero.

The model structures differ by how many of these polynomials are included in the structure. Thus, different model structures provide varying levels of flexibility for modeling the dynamics and noise characteristics. *Table 1* summarizes common linear polynomial model structures supported by the System Identification Toolbox in MATLAB. If scientists have a specific structure in mind for their application, they can decide whether the dynamics and the noise have common or different poles.

A(q) corresponds to poles that are common for the dynamic model and the noise model. Using common poles for dynamics and noise is useful when the disturbances enter the system at the input. F_i determines the poles unique to the system dynamics, and D determines the poles unique to the disturbances.

Model structure	Equation
ARX	$A(q)y(t)=\sum_{i=1}^{nu} B_i(q)u_i\left(t-nk_i\right)+e(t)$ **Equation 3: ARX Equation**
ARIX	$Ay=Bu+\frac{1}{1-q^{-1}}e$ **Equation 4: ARIX Equation**
ARMAX	$A(q)y(t)=\sum_{i=1}^{nu} B_i(q)u_i\left(t-nk_i\right)+C(q)e(t)$ **Equation 5: ARMAX Equation**
ARIMAX	$Ay=Bu+C\frac{1}{1-q^{-1}}e$ **Equation 6: ARIMAX Equation**
Box-Jenkins (BJ)	$y(t)=\sum_{i=1}^{nu} \frac{B_i(q)}{F_i(q)}u_i\left(t-nk_i\right)+\frac{C(q)}{D(q)}e(t)$ **Equation 7: Box-Jenkins Equation**

Table 1: Polynomial model structures

2.4. Criterion of fit - NRMSE

The criterion is used to measure how good the fit is between the observed output from simulation and the predicted values from the SI models. The Equation for Normalized Root Mean Squared Error is given below:

$$fit(i) = 1 - \frac{\left\| xref(:,i) - x(:,i) \right\|}{\left\| xref(:,i) - mean(xref(:,i)) \right\|}$$

Equation 8: NRSME

Where:

- ‖ indicates the 2-norm of a vector.
- fit is a row vector of length N
- i = 1,...,N, where N is the number of channels.

2.5. Reservoir recovery mechanisms - different reservoir states:

A) There are five key primary drive mechanisms (Tarek, 2010) for pushing fluids to the production well:

- Expansion of the reservoir rock and the reservoir fluids
- Expansion of solution gas escaping out of the oil phase
- The pressure exerted from a gas cap
- The pressure exerted from an aquifer
- Gravity causing a segregation effect that separates liquids of different densities. Oil tends to travel downwards and gas travels upwards.

A reservoir under primary drive can be represented by the block diagram in *Figure 3* below. It is immediately obvious that since there are no injection wells, the reservoir does not have any input. Outputs from systems of this kind are called time series and identification falls under a special branch of system identification called time series analysis (Nelles, 2001).

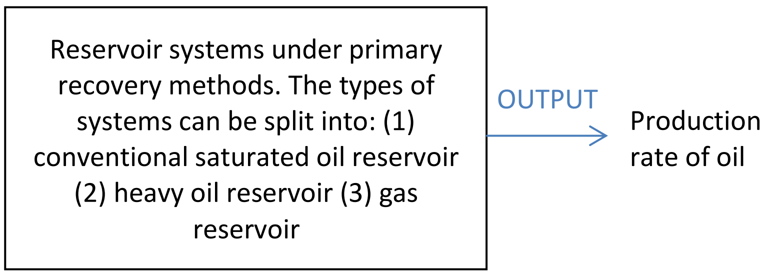

Figure 3: Block Diagram of primary recovery

B) Secondary drive mechanisms involve injection of fluids to maintain or increase the reservoir pressure in addition to displacing the reservoir fluids with the injected fluid (Satter & Iqbal, 2008). The main injection fluids are water and gas that is not miscible with the reservoir hydrocarbons. The block diagram for this drive mechanism can be seen in *Figure 4* below.

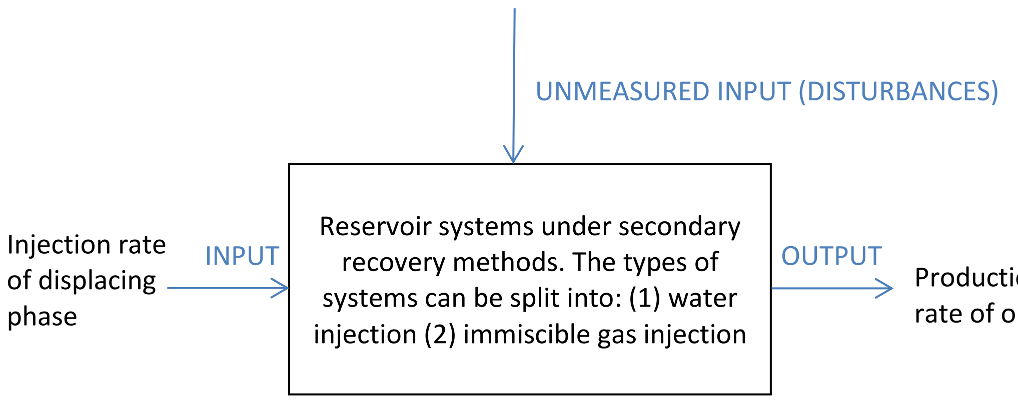

Figure 4: Block Diagram of secondary recovery

C) Tertiary drive mechanisms. This involves the use of special injection fluids and materials that can alter the properties of the reservoir fluids to achieve a combination of the following objectives: (1) make it easier for oil to flow, (2) to limit water flow and to cause a more effective displacement by the injection fluid compared to secondary recovery methods. The diagram for this recovery method can be seen in figure 5 below. It should be noted that, although the inputs and outputs of tertiary drive seem similar to the secondary drive mechanism, the injection fluids are affecting the system in a different way (alters reservoir properties). This may cause the relationship between inputs and outputs to become non-linear and hence producing a nonlinear system.

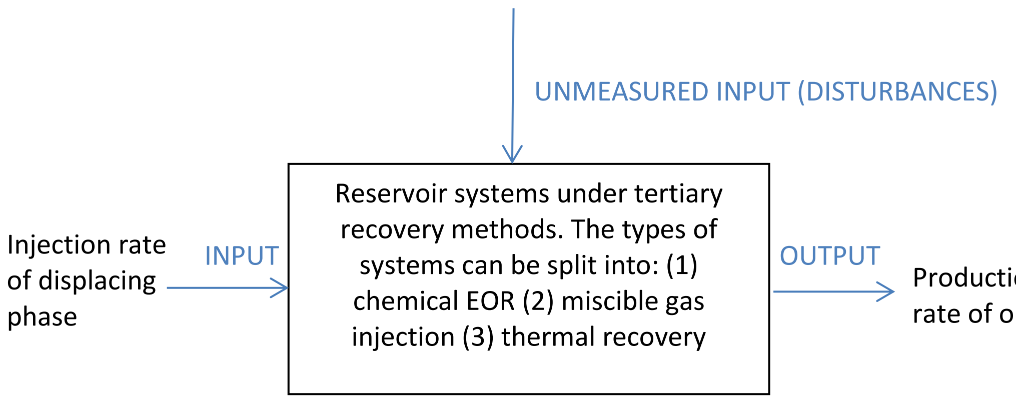

Figure 5: Block Diagram of tertiary recovery

2.6. Limitations of other forecasting methods – reasons for investigating system identification

- Reservoir modelling and simulation (history matching)
    1) Long period of time required for the building of models and running of simulations.
    2) Need as much information on the reservoir as can be obtained in order to get good model. Information from logs, cores, well tests and other sources are expensive to obtain and not available at early stage of development.
    3) Elaborate and complex process due to many reservoir variables involved (Olominu & Sulaimon, 2014).
    4) Models have inherent uncertainty because there is never enough available data.
    5) There isn't a unique model that satisfies the field data. Many different models with different variables can produce same results. (Tomomi, 2000).
    6) Un-pragmatic approach: imposing rigorous mathematical descriptions to nature (Ljung, 1987).

- Decline Curve Analysis
    1) Only applicable to forecast production that is declining.
    2) Due to it being an extrapolation method, it has a tendency to underestimate and overestimate (Li & Horne, 2005). Hence accuracy can be very poor.

- Material balance
    1) Does not take into account heterogeneity of the reservoir (Tarek, 2010). Hence the method does not take into account the importance of well locations.
    2) Treats the reservoir as a tank (Tarek, 2010): Ignores pressure distributions in reservoir and does not take into account fluid front movement during water influx or injection scenarios.

Chapter 3: Methodology/Project Work

3.1. Literature review

Examples of system identification application in petroleum engineering field are still very scarce, and the author would need to refer to other science fields to learn about how to properly implement system identification. Thus, required information will be taken from:

- Books on system identification by renowned world experts (such as Lennart Ljung)
- Highly credible internet sources: Society of Petroleum Engineers (SPE) papers (One Petro website), Sciencedirect papers, Petrowiki by SPE, etc.
- Papers/journals from other engineering fields

3.2. Data Gathering/ Collection

Using simulation software to obtain input and output data: First of all, suitable test data sets (inputs and outputs) need to be obtained for each recovery scenario for the purpose of system identification. Inputs come from injection wells in the form of total field injection rate vs. time. Meanwhile, outputs come from producer wells in the form of (a) total field oil production rate vs. time, (b) total field water cut vs. time and (c) total field Gas-Oil-Ratio (GOR) vs. time.

Because real field data is hard to obtain, the next best alternative would be to create and run reservoir simulation models using commercial simulators and then to extract suitable test data sets from the simulation results. The reason for obtaining working data this way is that reservoir simulation uses first principles (proven mathematical and physical laws) to build a white model of the real reservoir. Hence, in this project it is assumed that the simulation model is the true mathematical representation of the real reservoir system. The simulators to be used will be those from computer modelling group (CMG).

3.3. System identification process and forecasting

MATLAB system identification tool will be used to find suitable SI models for each recovery scenario. The model would be obtained by using statistics and the process will be a combination of trial and error process as well as applying reservoir engineering judgment of the system (Mathworks, 2015). Once a model has been deemed to simulate the recovery mechanism well, its forecasting performance can then be evaluated by seeing how closely the predicted outputs compare to those outputs from the reservoir simulators. The criterion used will be Normalized Root Mean Squared Error (NRMSE). Embedded in this process are steps 4, 5 & 6 of the system identification process discussed in the literature review.

3.4. Summary of procedure

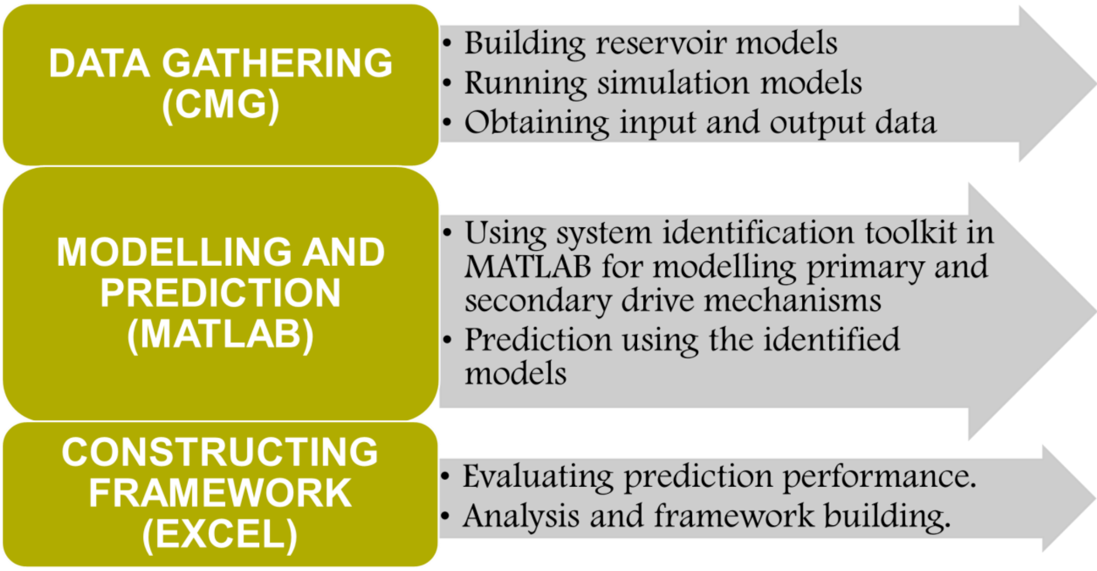

Figure 6: Summary of procedure

3.5. Flow chart of methodology used in FYP 1

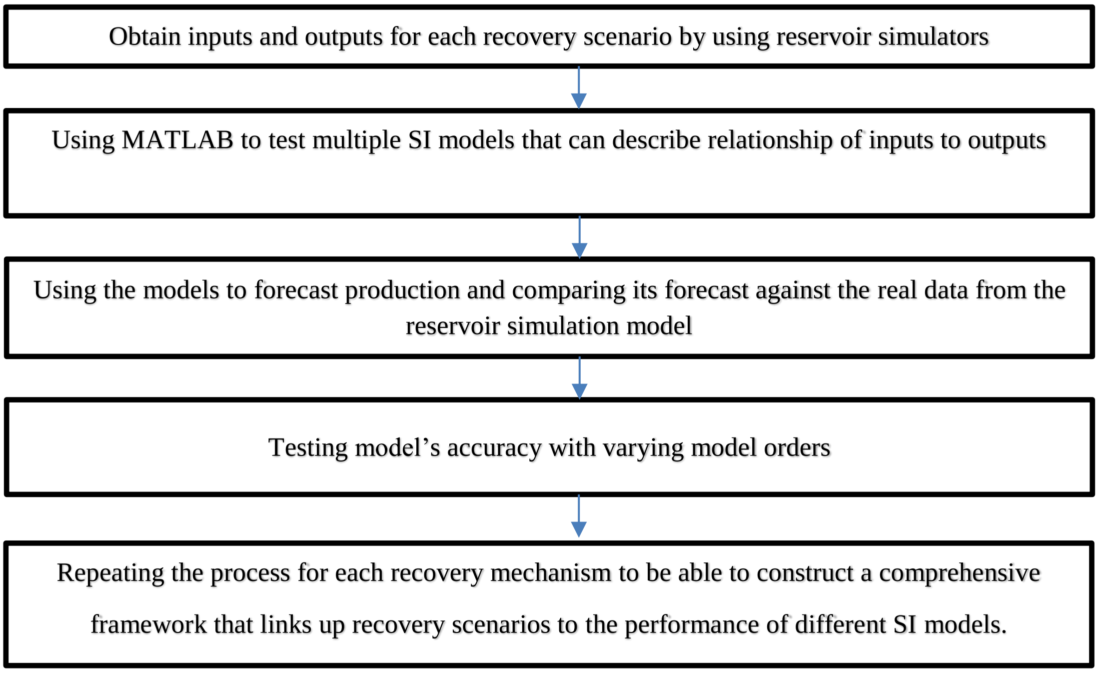

Figure 7: Flow chart of project

3.6. Gantt Chart

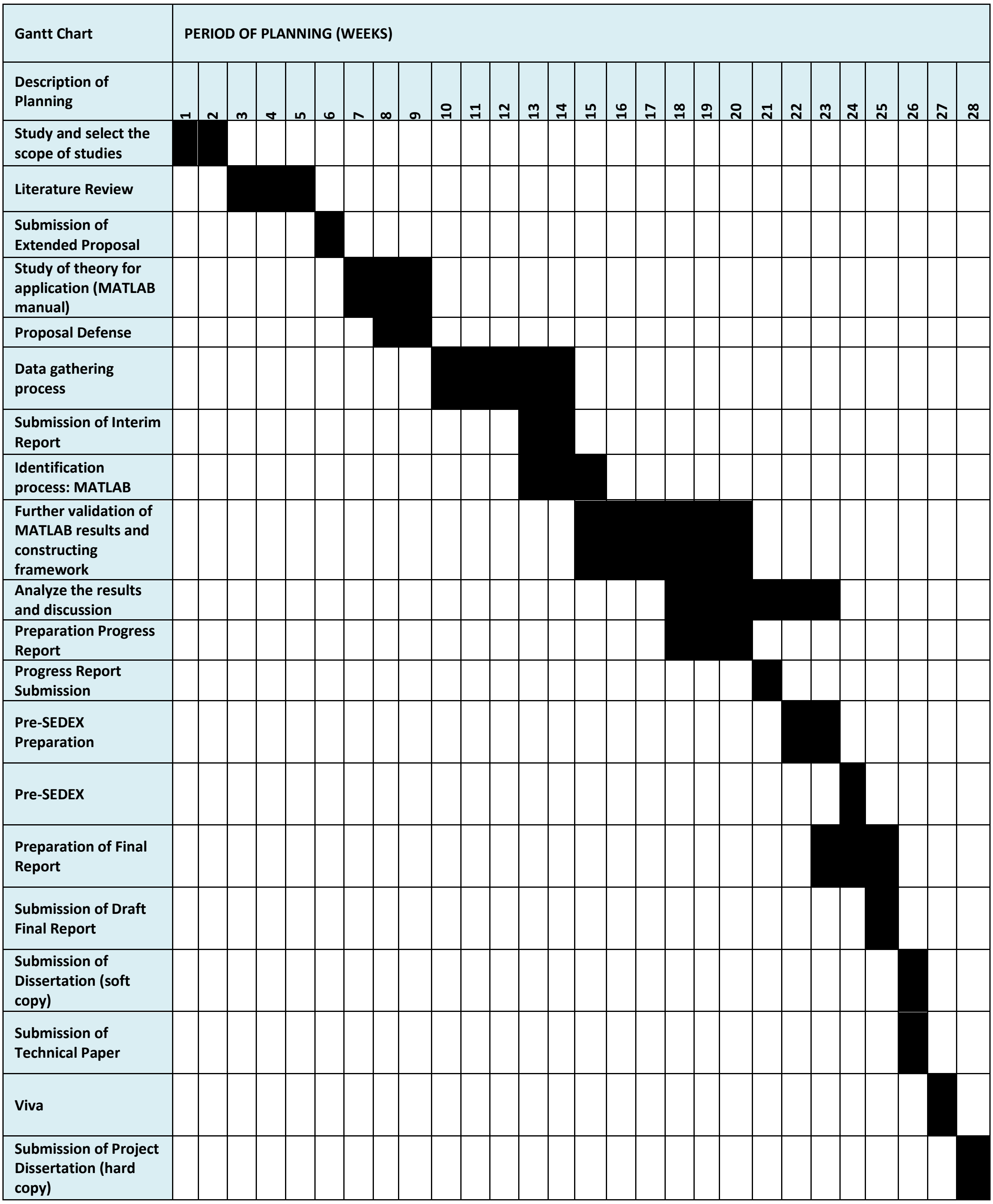

Figure 8: Gantt chart of project

Chapter 4: Results and Discussion

4.1. Data Gathering: Reservoirs descriptions

The process of collecting simulation models for each recovery scenario involves creating suitable models from existing template models provided by CMG Company. These template models are heavily modified by the author and are only considered suitable once they clearly reflect the drive mechanism under investigation. The final models should also have enough complexity, which is measured through the following criteria:

1) Total number of grid-blocks in the model. Generally, the higher the number, the more complex the model due to longer runtimes. However, a variety of big and small models are also taken to be able to test the effectiveness of SI models in forecasting different sized reservoirs.
2) Heterogeneity of rock properties, such as permeability and porosity. Large heterogeneities are required.
3) The presence of faults make the reservoir more complex.

These models are also modified to run for a total period of 10 years and the wells in each model are set to operate on an optimized constant bottom-hole pressure (BHP) constraint. BHP is optimized by selecting one that results in highest oil recovery but also taking into account the restrictions of the reservoir, i.e. fracture pressure of the reservoir rock (for injection wells) and the lowest BHP that can support production without having to resort to artificial lift methods (for production wells). The CMG simulator (IMEX) is set to record data every day for 10 years. This means that for each drive mechanism there are around 3653 data points for each input and output. Furthermore, a black oil fluid model is used in all cases.

4.1.1. Rock & Liquid Expansion Drive (RLD)

Figure 9 shows that the obtained reservoir model follows the typical trend of reservoirs under expansion drive (Tarek, 2010). Average reservoir pressure rapidly declines in the first two years of production and the total producing GOR is constant.

This model is modified from mxgro002.dat template. *Figure 10* shows plots of the production parameters (outputs) we wish to predict and *Figure 11* shows the 3D view of the reservoir. The key characteristics of the reservoir are:

- Total number of grid-blocks: 286
- Rock properties: Porosity varies in each block, from 0 to 0.17. Absolute permeability in I & J direction varies in each block from 0 to greater than 300 mD, while Kv/Kh ratio is 0.5. Sw also varies from block to block.
- Relative permeability curves are provided in *Figure 12*.
- Initial average reservoir pressure is 4000 psi and bubble point (Pb) is 2000 psi.
- There are 10 producer wells, all operating at constant BHP of 2050 psi only in order to deplete the reservoir not below the bubble point. This allows for rock and fluid expansion to be the only drive mechanism.
- No aquifer support.

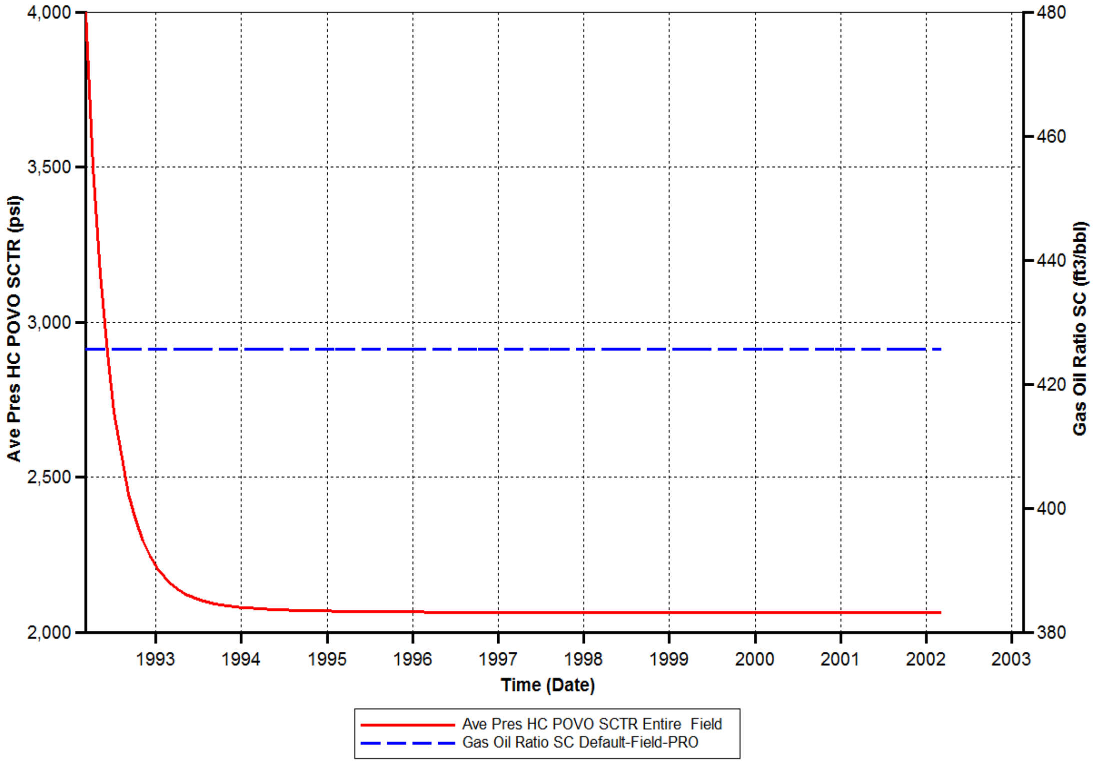

Figure 9: Identification of drive mechanism – RLD

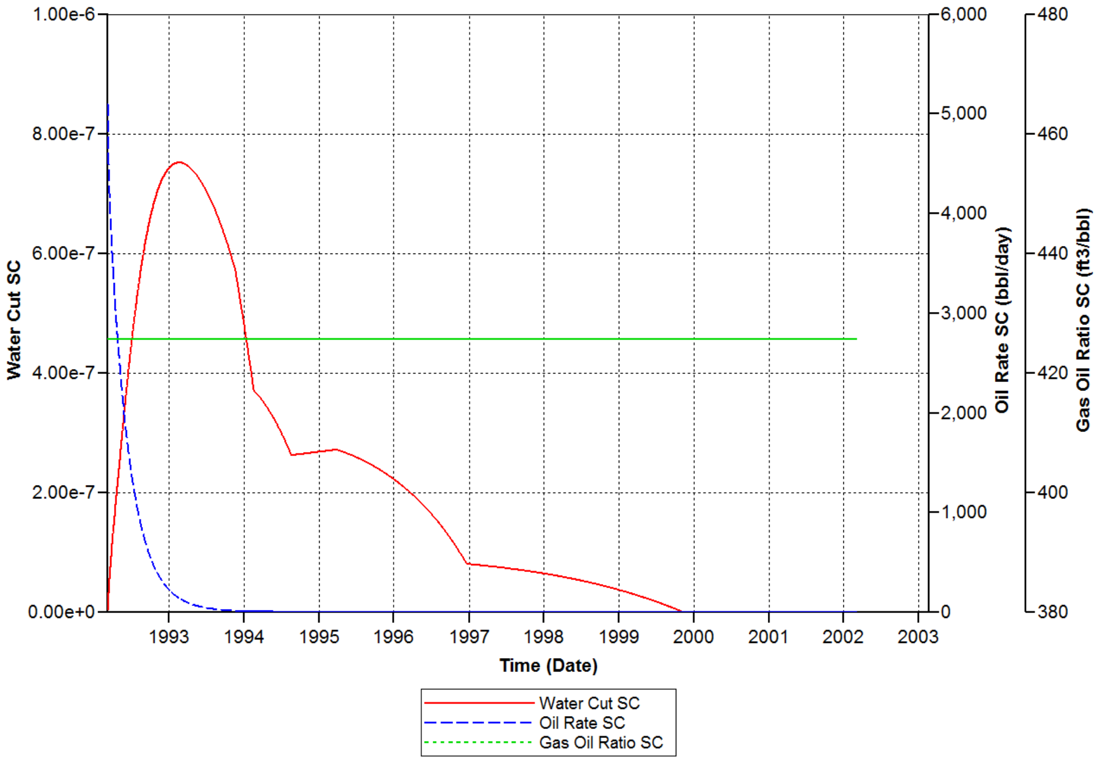

Figure 10: Production performance parameters (Outputs) – RLD

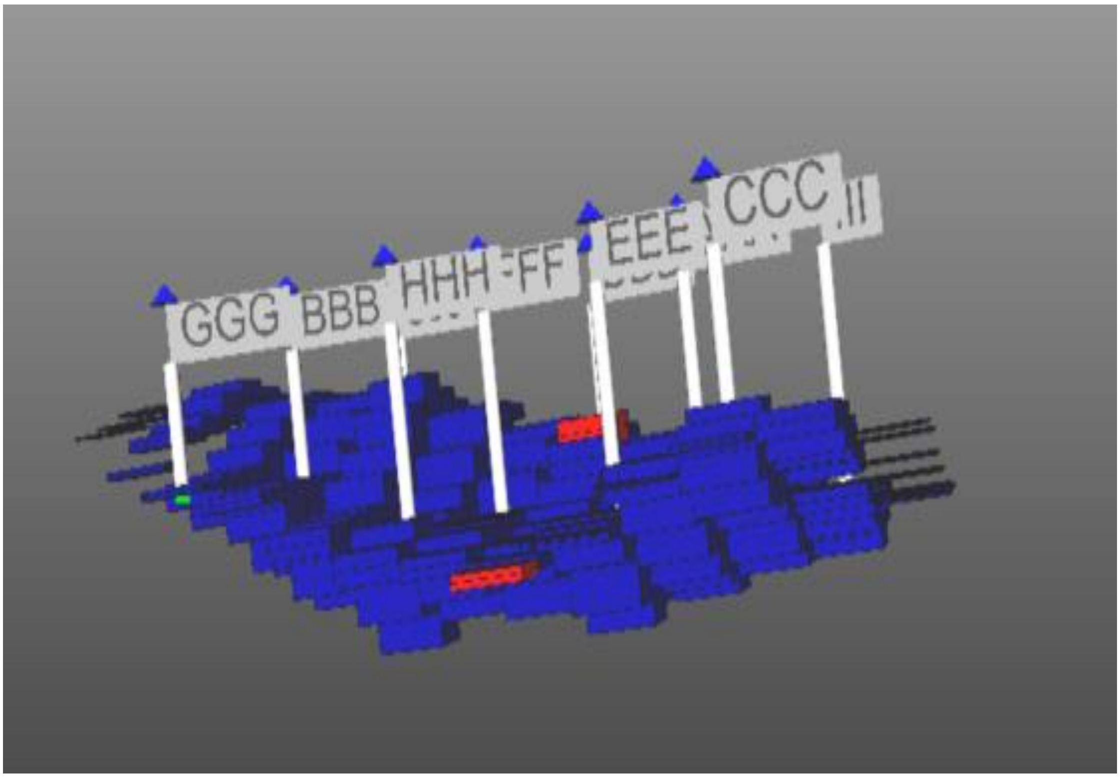

Figure 11: 3D view of reservoir– RLD

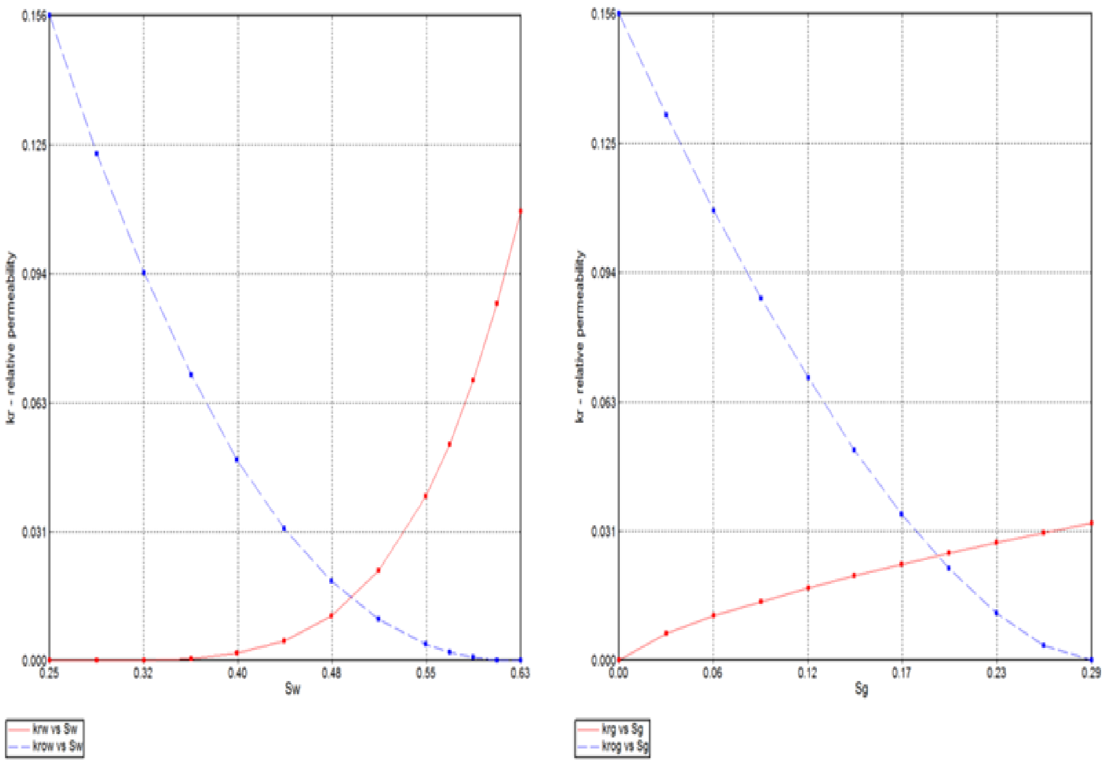

Figure 12: Relative permeability curves – RLD

4.1.2. Solution Gas Drive (SGD)

Figure 13 shows that the obtained reservoir model follows the typical trend of reservoirs under solution gas drive (Tarek, 2010). Similarly, average reservoir pressure also rapidly declines, though not as fast as the previous drive mechanism because gas is more compressible than live oil. The total producing GOR rose rapidly at first due to gas coming out of the oil below bubble point and getting produced in large amounts very quickly due to it being more mobile. Later on, GOR drops because the gas production rate eventually reduces as there is less and less gas in the reservoir. As for the field cumulative water production, it is very negligible (less than one barrel), which shows that the main drive mechanism is solution gas drive.

This model is also modified from mxgro002.dat template and is based on the expansion drive model obtained previously. *Figure 14* shows plots of the production parameters (outputs) we wish to predict. The key characteristics of the reservoir are similar to the previous drive mechanism with only the following differences:

- Initial average reservoir pressure is 2561 psi and Pb is 2000 psi.

- There are 10 producer wells, all operating at constant BHP of 500 psi in order to deplete the reservoir below the bubble point.

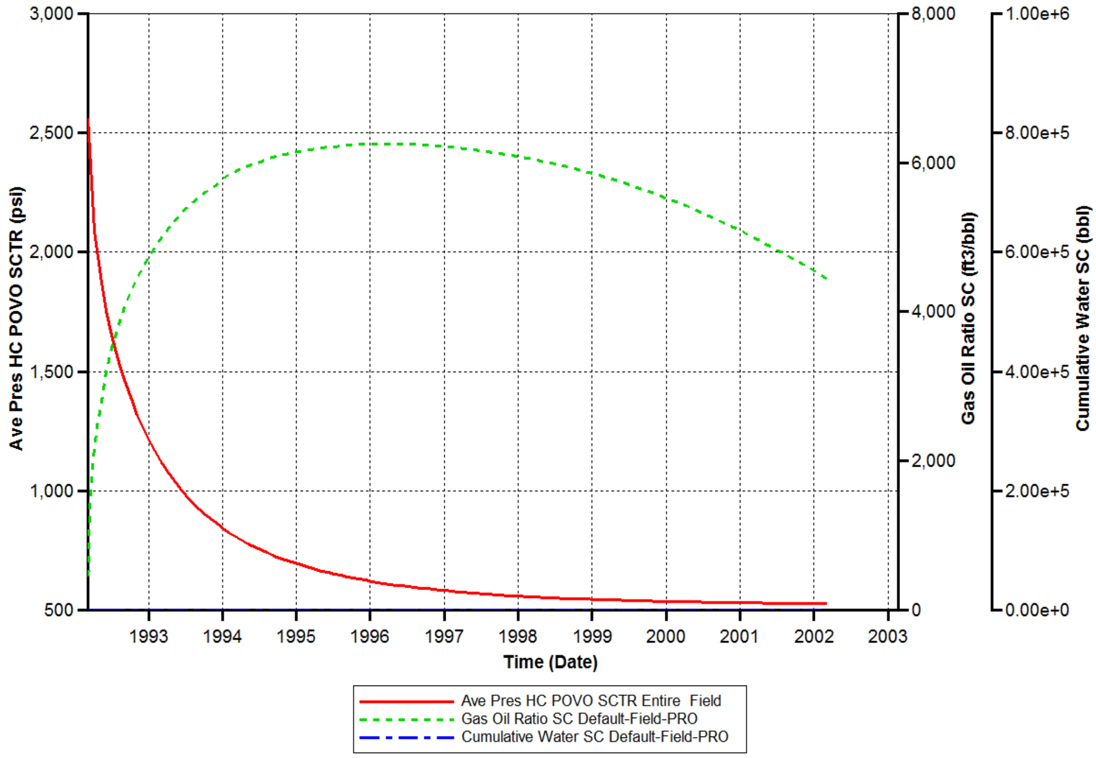

Figure 13: Identification of drive mechanism – SGD

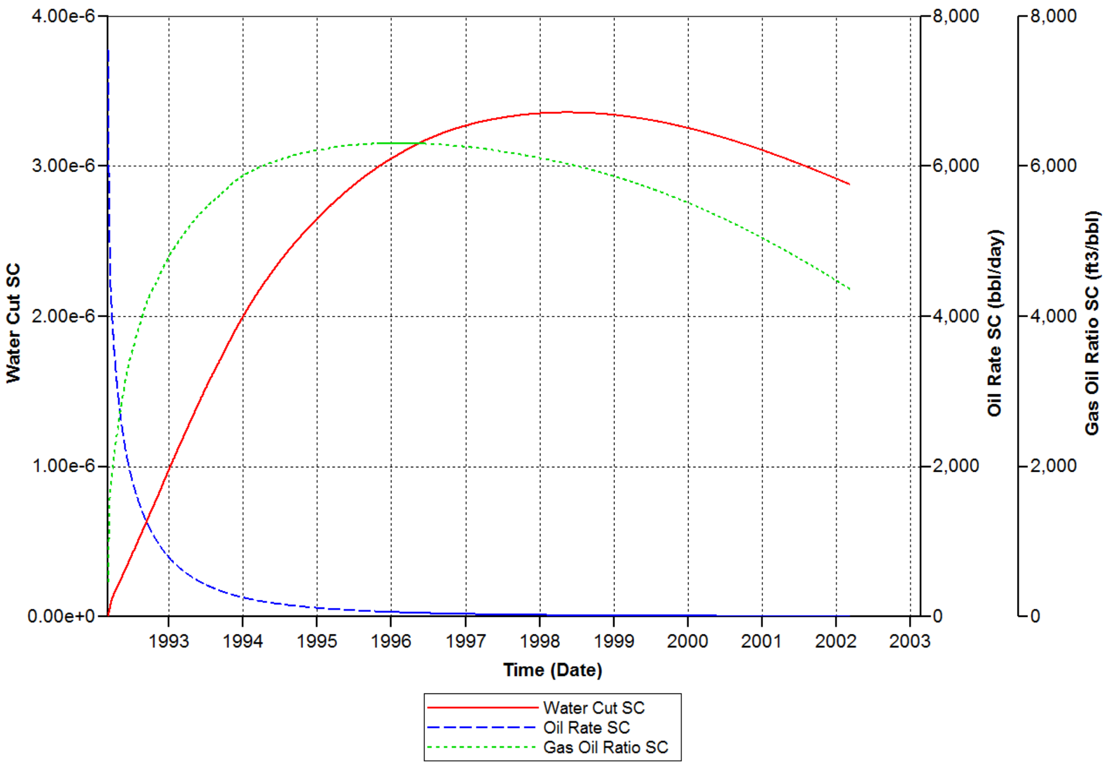

Figure 14: Production performance parameters (Outputs) – SGD

4.1.3. Gas Cap drive (GCD)

Figure 15 shows that the obtained reservoir model follows the typical trend of reservoirs under gas cap drive (Tarek, 2010). Average reservoir pressure declines continuously but much slower than that of solution gas drive or expansion drive. Water production is very small, as shown by the water rate almost being zero at all times, and hence water production can be ignored when compared to oil production rates and cumulative volumes produced. Lastly, GOR will slowly increase with time due to the expanding gas cap.

This model is modified from mxdrm003.dat template. *Figure 16* shows plots of the production parameters (outputs) we wish to predict and *Figure 17* shows the 3D view of the reservoir. The key characteristics of the reservoir are:

- Total number of grid-blocks: 1400
- Rock properties: Porosity varies in the vertical direction with values ranging between 0.15 to 0.27. Absolute permeability in I & J direction varies in the vertical direction with values ranging between 45 to 350 mD, while Kv/Kh ratio is 1. Kv/Kh ratio is set to a high value in order to promote gravity segregation effect, so that the solution gas that comes out of the oil will go upwards to the gas cap and the oil downwards towards the producers. Also, Sw varies from block to block.
- Relative permeability curves are provided in *Figure 18.*
- Initial average reservoir pressure is around 10576 kPa and Pb is 9570 kPa.
- There are 5 producer wells, all operating at constant BHP of 7000 kPa. The reason for the high BHP is that gas cap drive is very sensitive to the production rate and hence lower rates is better in the long run because it will allow the gas cap to displace the oil more evenly (piston like manner) and result in higher recovery.
- No aquifer support.

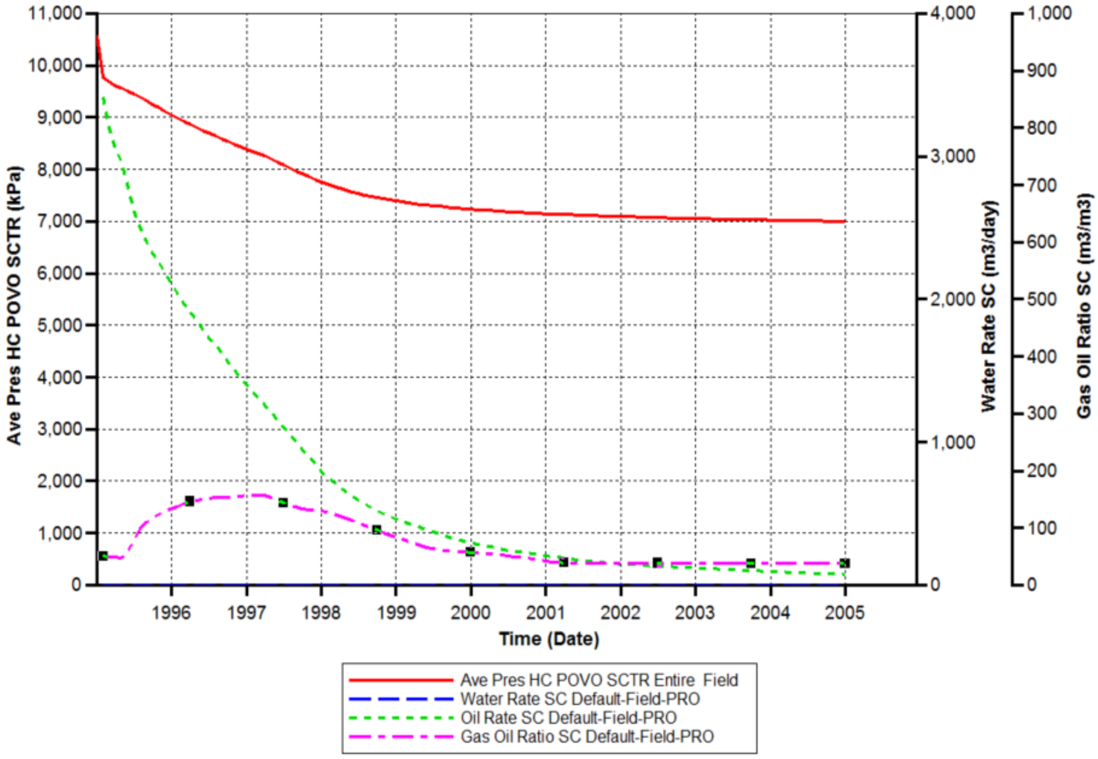

Figure 15: Identification of drive mechanism – GCD

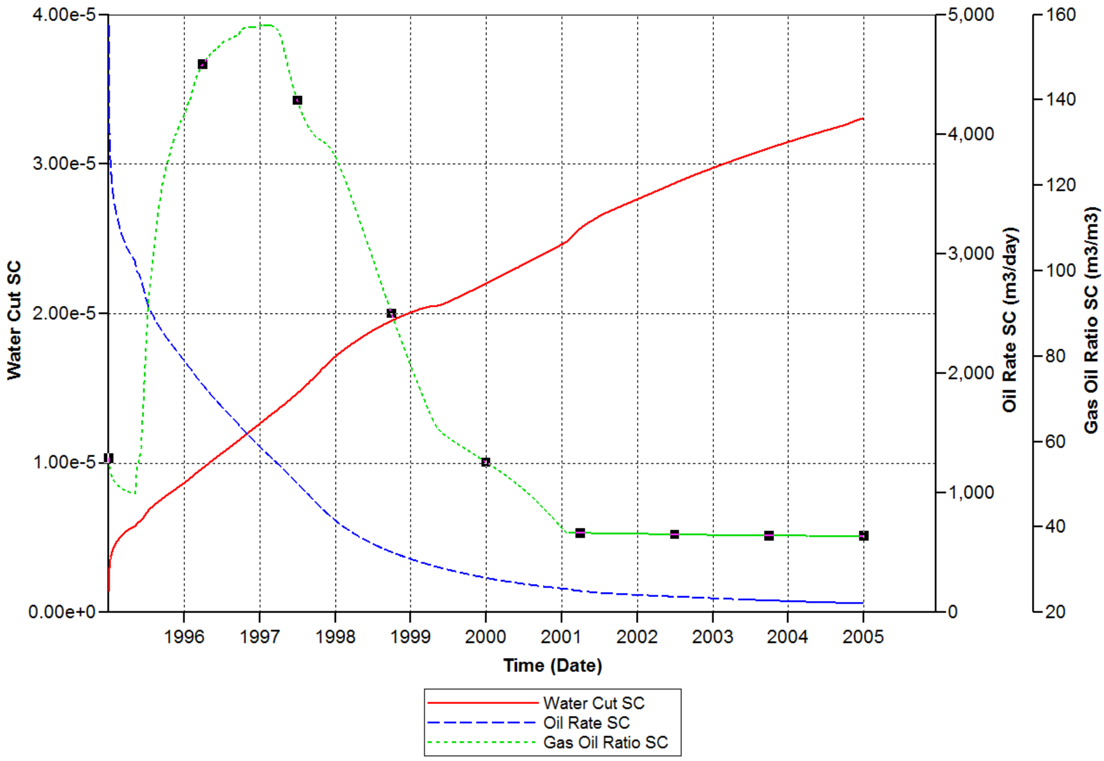

Figure 16: Production performance parameters (Outputs) – GCD

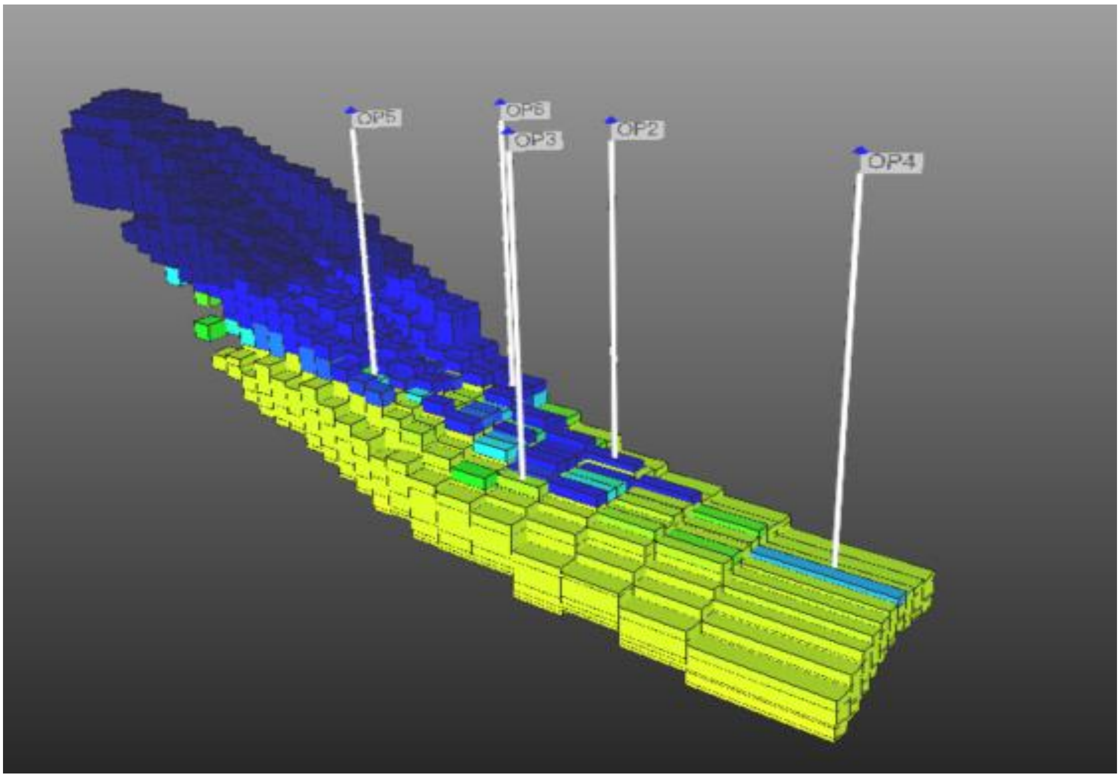

Figure 17: 3D view of reservoir – GCD

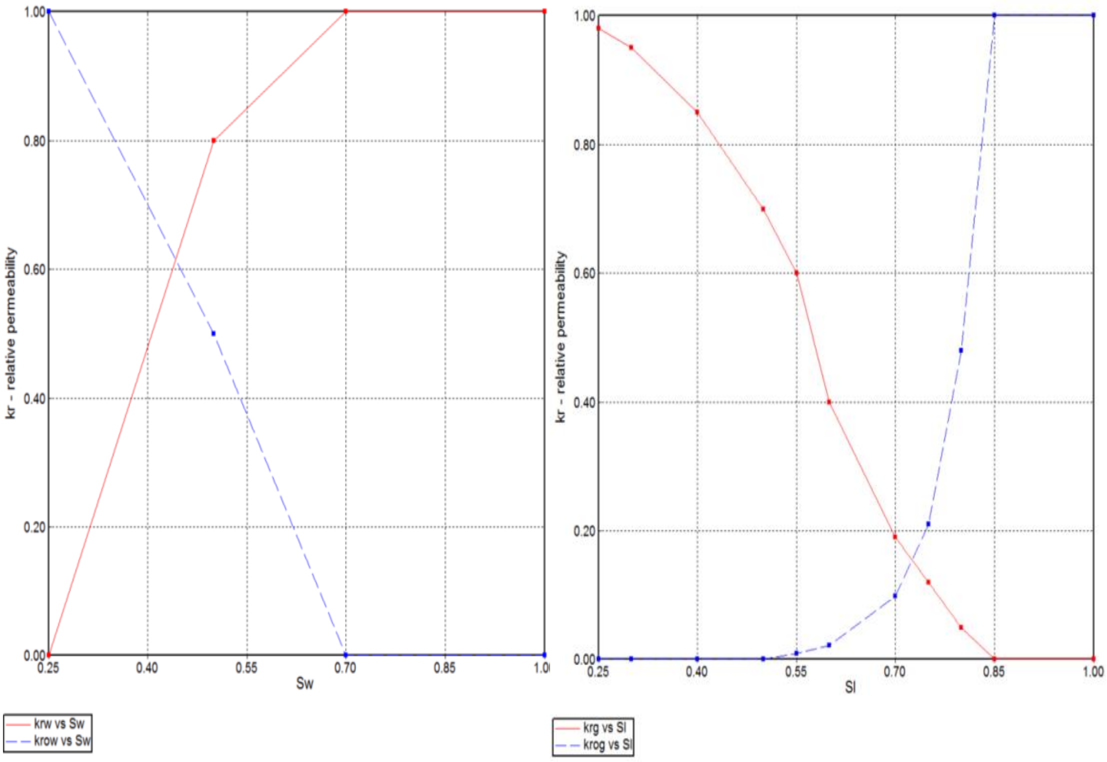

Figure 18: Relative permeability curves – GCD

4.1.4. Aquifer drive (AQD)

Figure 19 shows that the obtained reservoir model follows the typical trend of reservoirs under aquifer drive (Tarek, 2010). This particular reservoir has a bottom aquifer drive support. As can be seen from the average reservoir pressure curve, the decline is very gradual or almost non-existent after just a brief period of steep decline. Water production increases quite rapidly due the water encroaching into the oil zone, but this is expected of an aquifer drive. Lastly, GOR stays roughly constant for most of the production period due to the reservoir pressure being maintained.

This model is modified from mxgeo004.dat template. *Figure 20* shows plots of the production parameters (outputs) we wish to predict and *Figure 21* shows the 3D view of the reservoir. The key characteristics of the reservoir are:

- Total number of grid-blocks: 2500
- Rock properties: Porosity varies in each block, ranging mainly between 0.2 to 0.3. Absolute permeability in I & K direction varies in each block from 30 to 300 mD, while absolute J permeability=I permeability. Sw also varies from block to block.
- The relative permeability curves are provided in *Figure 22.*
- Initial average reservoir pressure is around 5490 kPa and Pb is 5570 kPa.
- There are 6 producer wells, all operating at constant BHP of 1000 kPa.

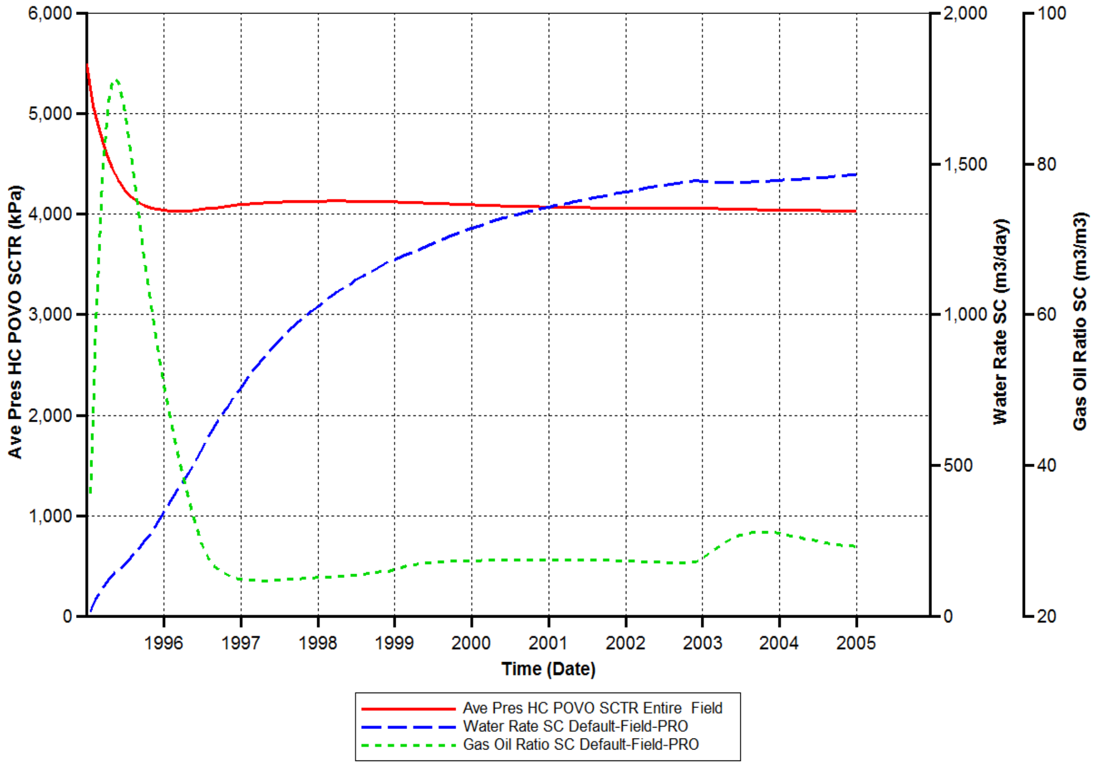

Figure 19: Identification of drive mechanism – AQD

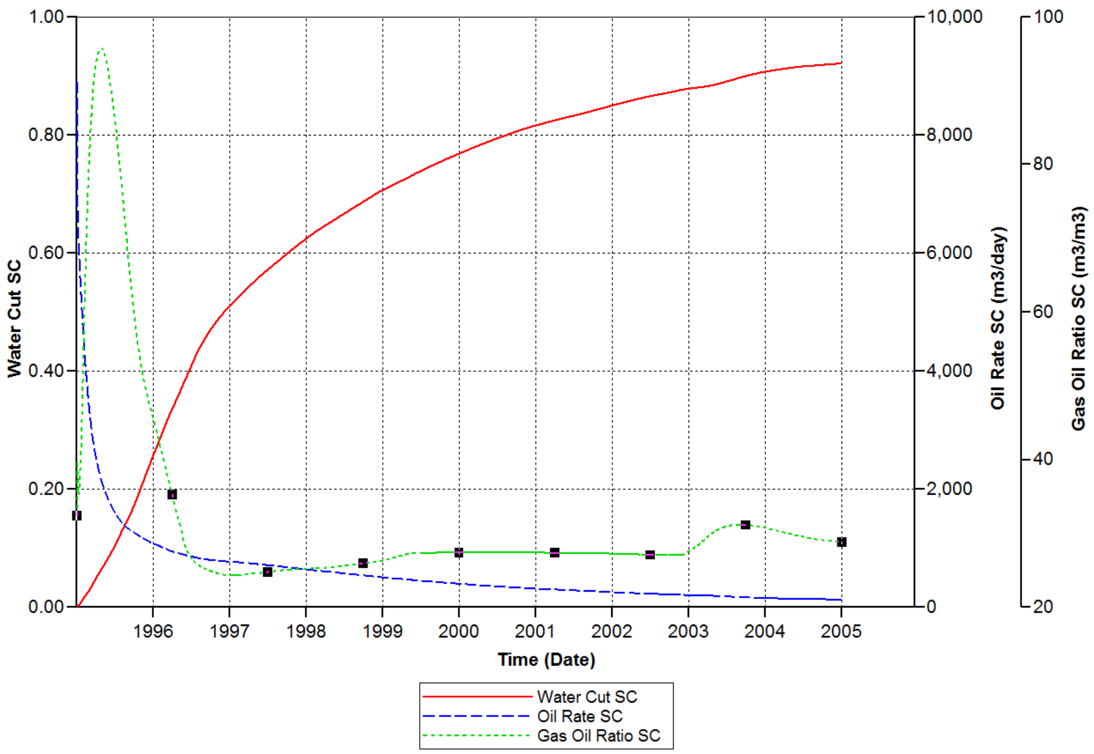

Figure 20: Production performance parameters (Outputs) – AQD

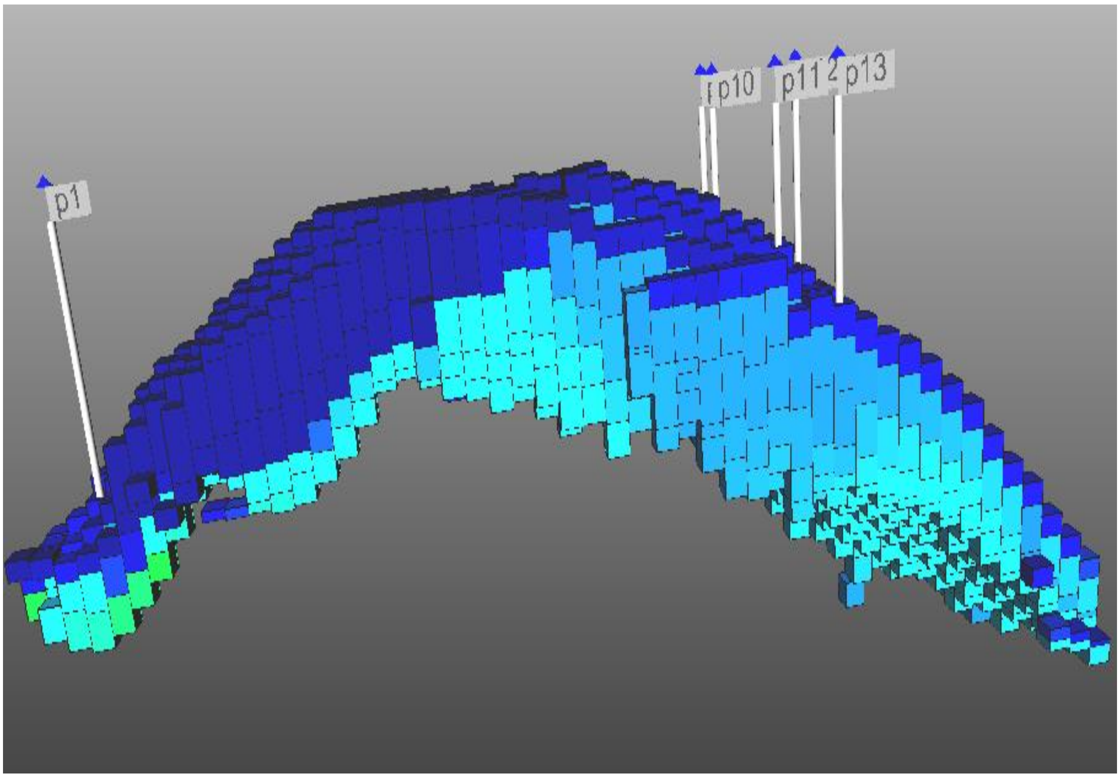

Figure 21: 3D view of reservoir – AQD

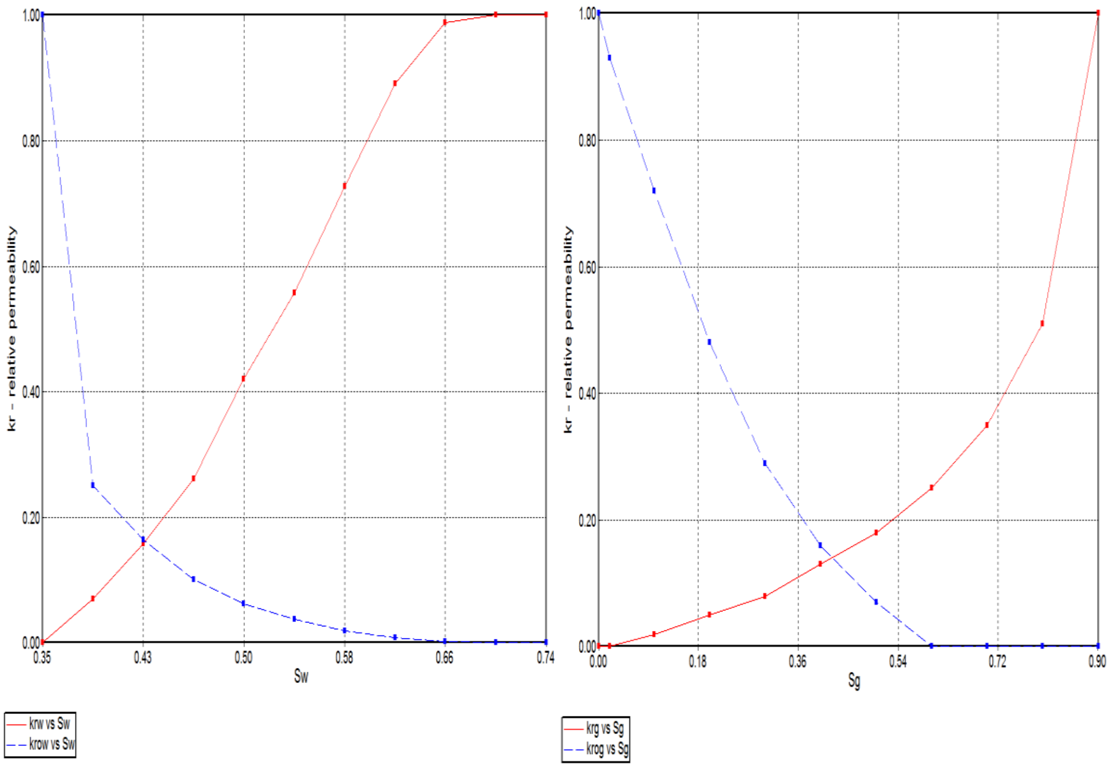

Figure 22: Relative permeability curves – AQD

4.1.5. Combined drive (COD)

Figure 23 shows that the obtained reservoir model follows the typical trend of reservoirs under combined drive (Tarek, 2010). This particular reservoir has a combination of solution gas drive, gas cap drive and aquifer drive. As can be seen from the average reservoir pressure curve, the decline is very fast because the pressure support from both the gas cap and aquifer is not strong. This is also shown from the very slow encroachment of water, which result in only very little water production (the cumulative water production curve looks like it is constantly very close zero). Lastly, the GOR curve is continually increasing as the gas cap continues to expand and more solution gas comes out of the oil.

This model is modified from mxgeo003.dat template. *Figure 24* shows plots of the production parameters (outputs) we wish to predict and *Figure 25* shows the 3D view of the reservoir. The key characteristics of the reservoir are:

- Total number of grid-blocks: 3888. This reservoir has several faults.
- Rock properties: Porosity varies in each block, ranging between 0.156 to 0.17. Absolute permeability in I & J direction varies in each block from 4 to 10 mD,

while Kv/Kh ratio is 0.3. Also, Sw and Net-to-gross ratio varies from block to block.

- There are 4 different rock types in the reservoir, each with their own set of relative permeability curves. One of these relative permeability sets are provided in *Figure 26*.
- Initial average reservoir pressure is around 10716 psi and Pb is 30000 psi.
- There are 21 producer wells, all operating at constant BHP of 1500 psi

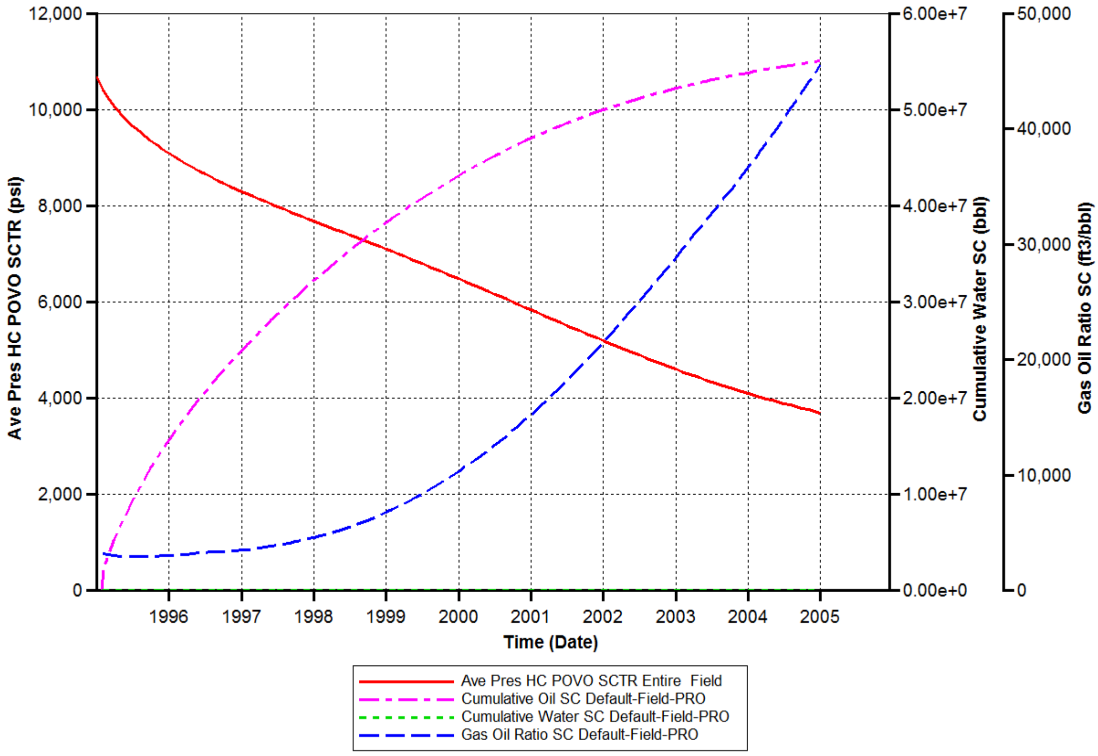

Figure 23: Identification of drive mechanism – COD

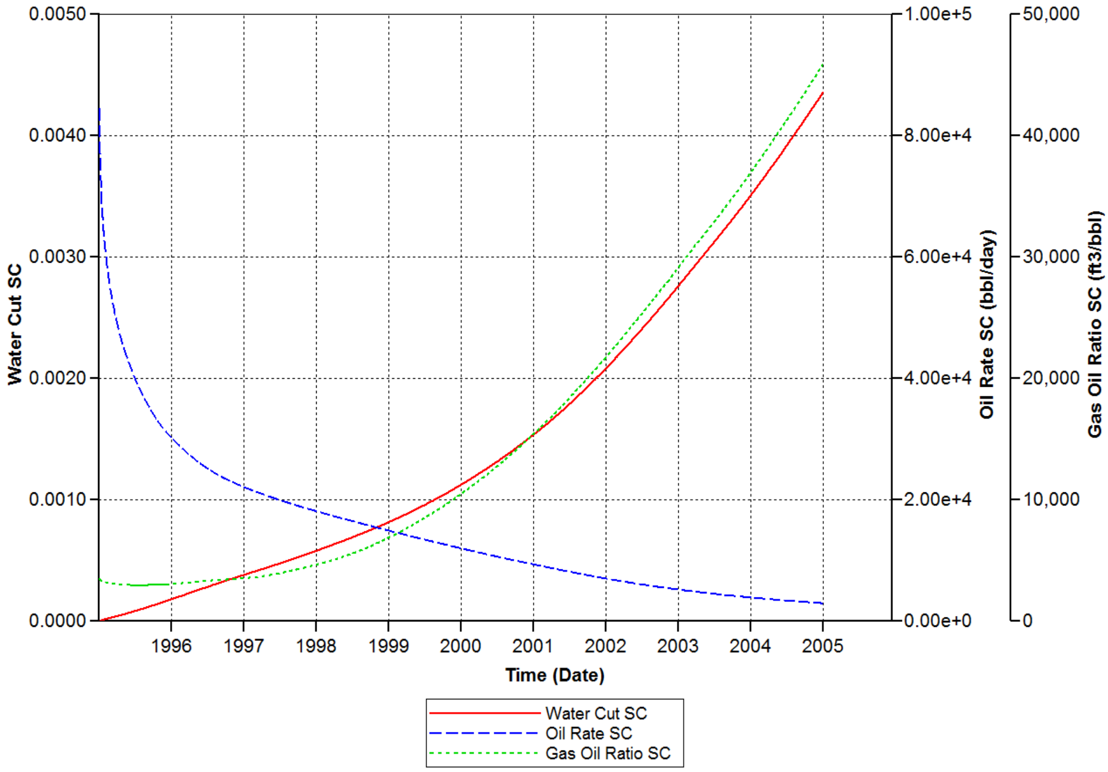

Figure 24: Production performance parameters (Outputs) – COD

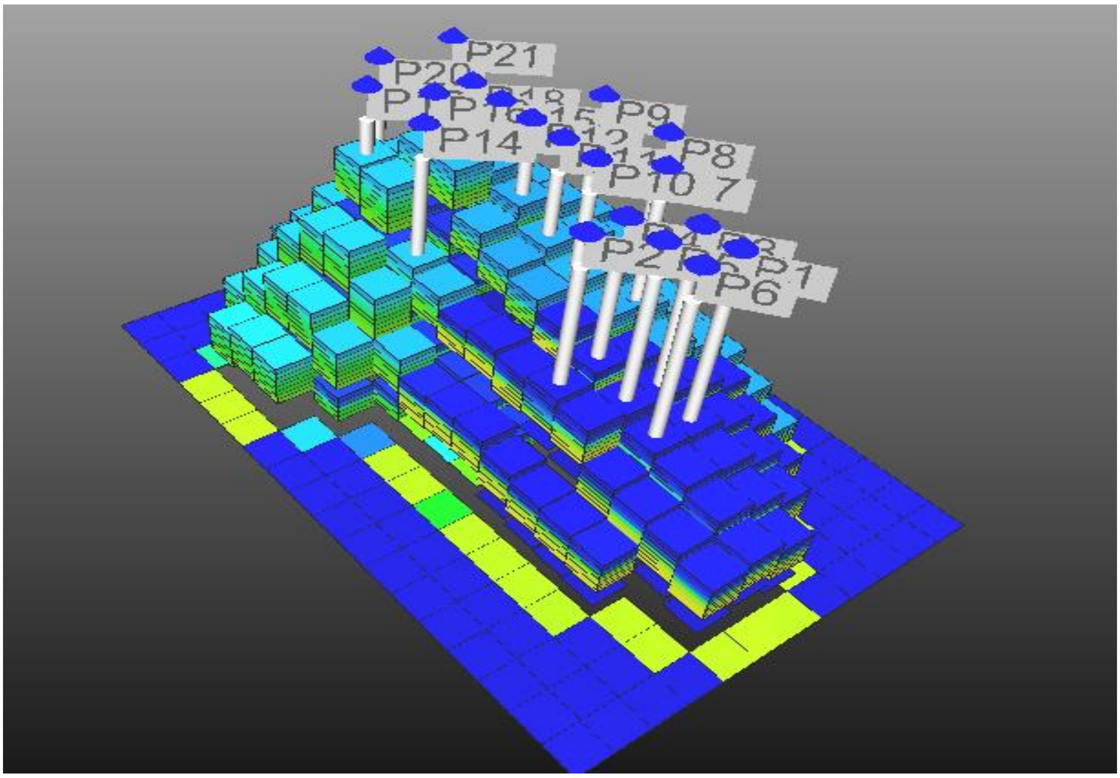

Figure 25: 3D view of reservoir – COD

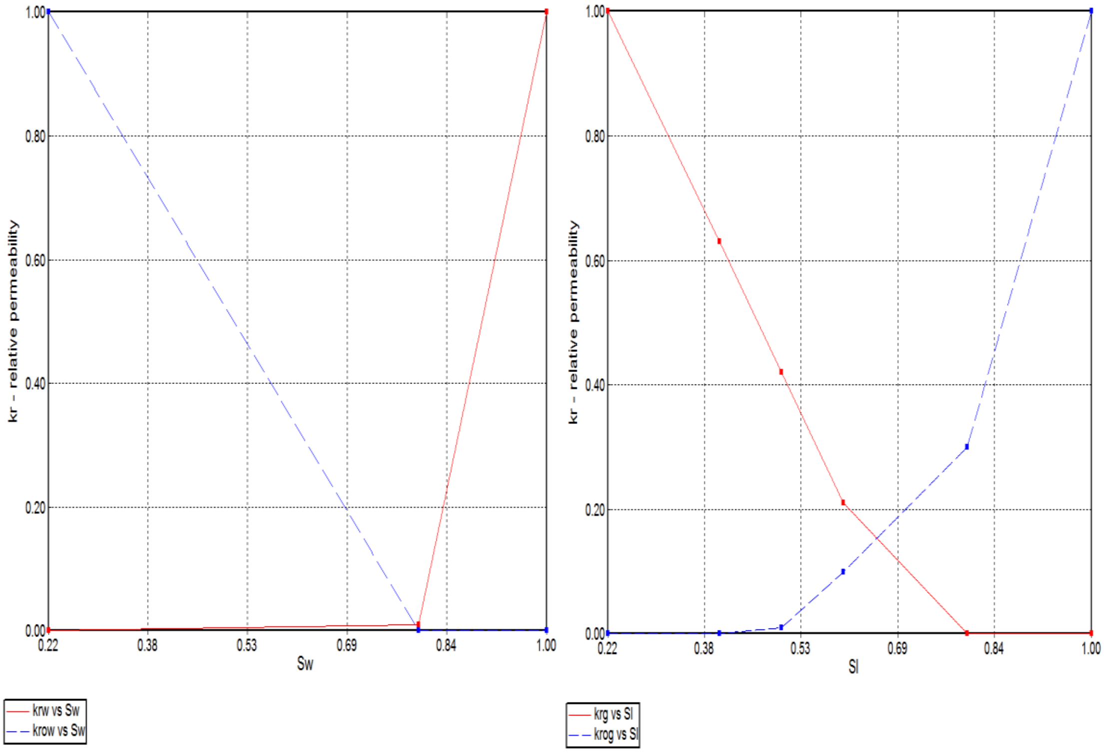

Figure 26: Relative permeability curves – COD

4.1.6. Water Injection (WAI)

Figures 27 and 28 show two cases of the same reservoir:

i) The reservoir is only being depleted by 25 producer wells.

ii) The reservoir has 25 producer wells and 10 injector well injecting water into the aquifer.

As can be seen from the average reservoir pressure curves of the two cases, the water injection case provides very good pressure support and maintains the reservoir pressure at a much higher pressure than the case with no injection. Moreover, the case with water injection provides higher levels of both oil and water recoveries. Cumulative oil recovery increases by around 25% due to the higher reservoir pressure. Understandably the water production also increases significantly due to injected water bypassing the oil. This proves that case (ii) is a good representation of secondary recovery by water injection and it can used to provide input-output data for the system identification process.

This model is modified from mxspe009.dat template. *Figure 29* shows plots of the production parameters (outputs) we wish to predict and the corresponding injection rate of water (input). Also, *Figure 30* shows the 3D view of the reservoir. The key characteristics of the reservoir are:

- Total number of grid-blocks: 9000
- Rock properties: Porosity varies for each block in the K direction, ranging between 0.8 to 0.17. Absolute permeability in I & J direction varies in each block and their distributions were made using geo-statistical techniques, varying mainly in the range between 20 mD to 700 mD. Meanwhile Kv/Kh ratio is 0.01. Also, Sw varies from block to block.
- The relative permeability sets are provided in *Figure 31*.
- Initial average reservoir pressure is around 4566 psi and Pb is 3600 psi.
- There are 25 producer wells, all operating at constant BHP of 2000 psi. There are also 10 injector wells, all operating at constant BHP of 4543.39 psi.

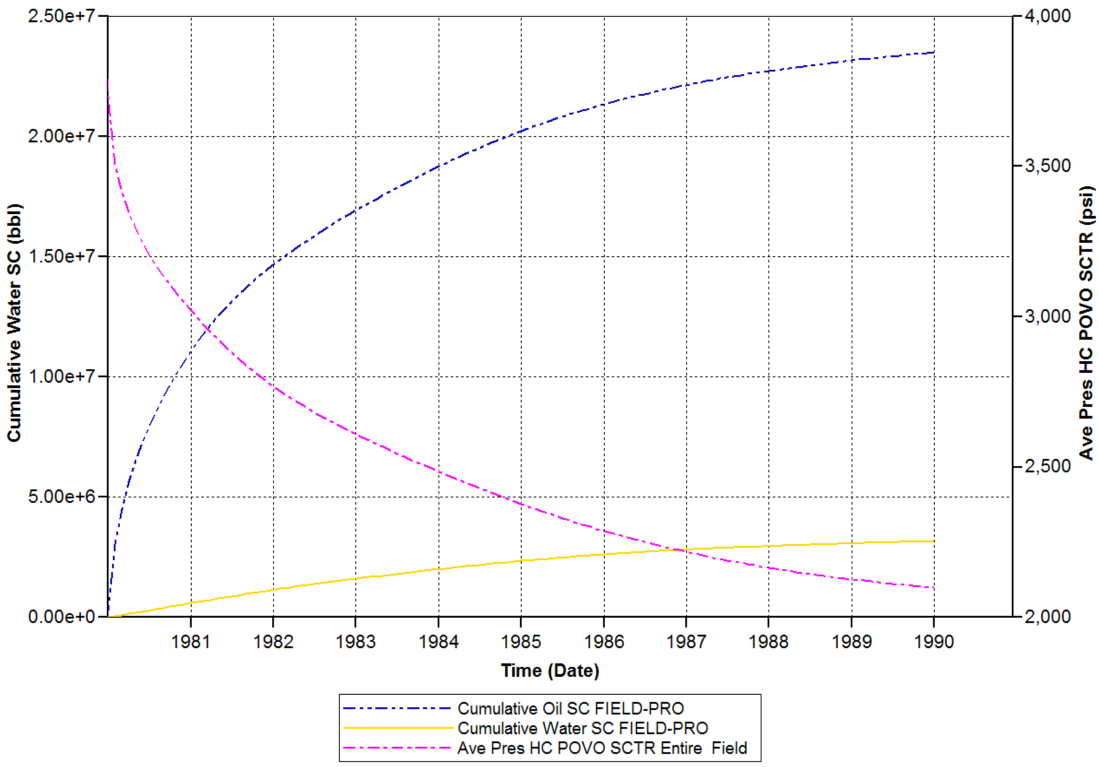

Figure 27: Identification of drive mechanism – WAI Case (i)

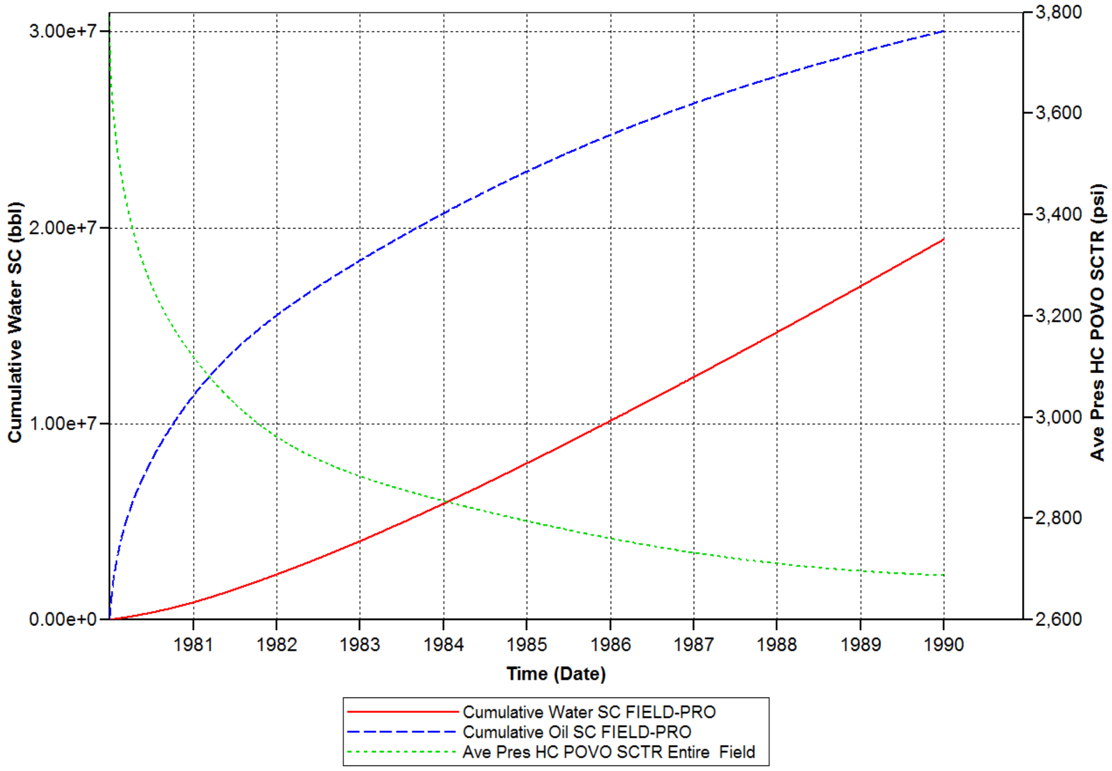

Figure 28: Identification of drive mechanism – WAI Case (ii)

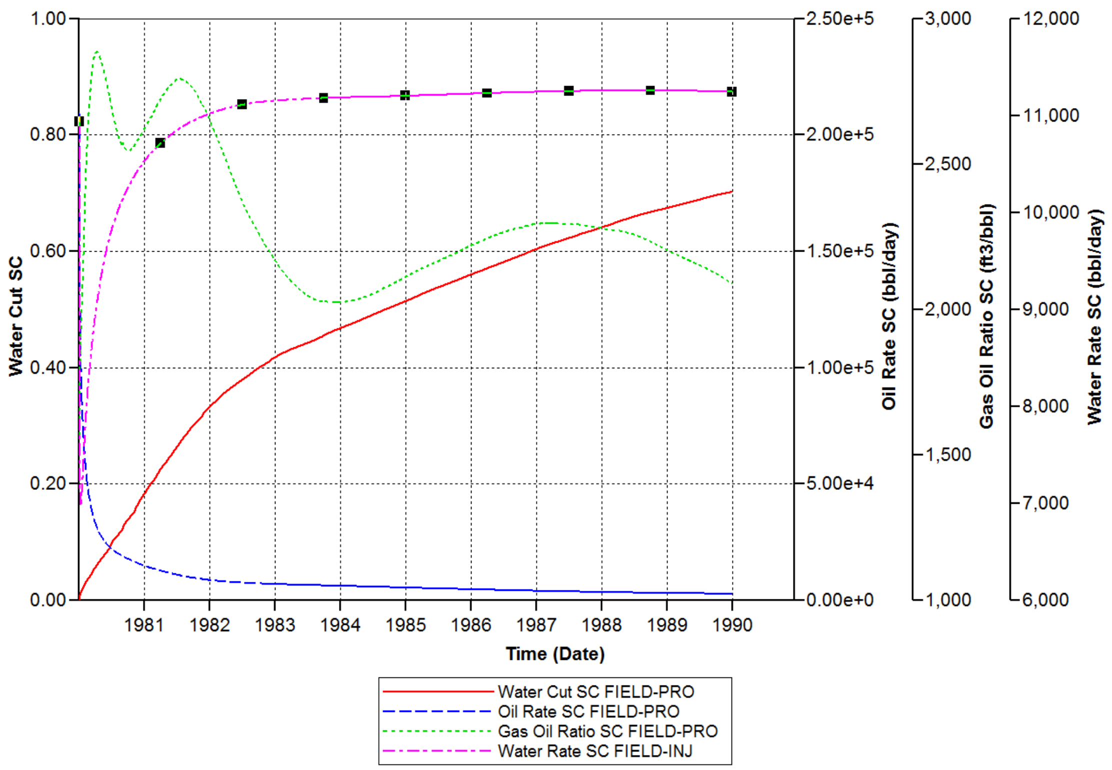

Figure 29: Production and injection parameters (Outputs and Input) – WAI

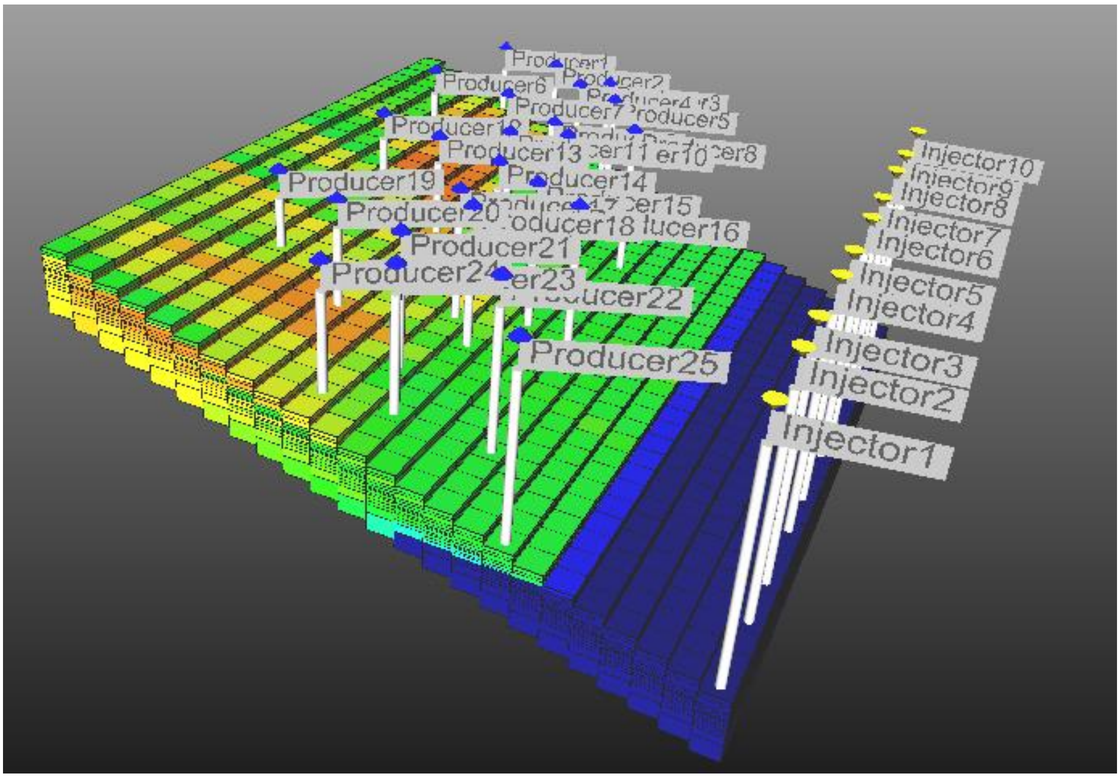

Figure 30: 3D view of reservoir – WAI

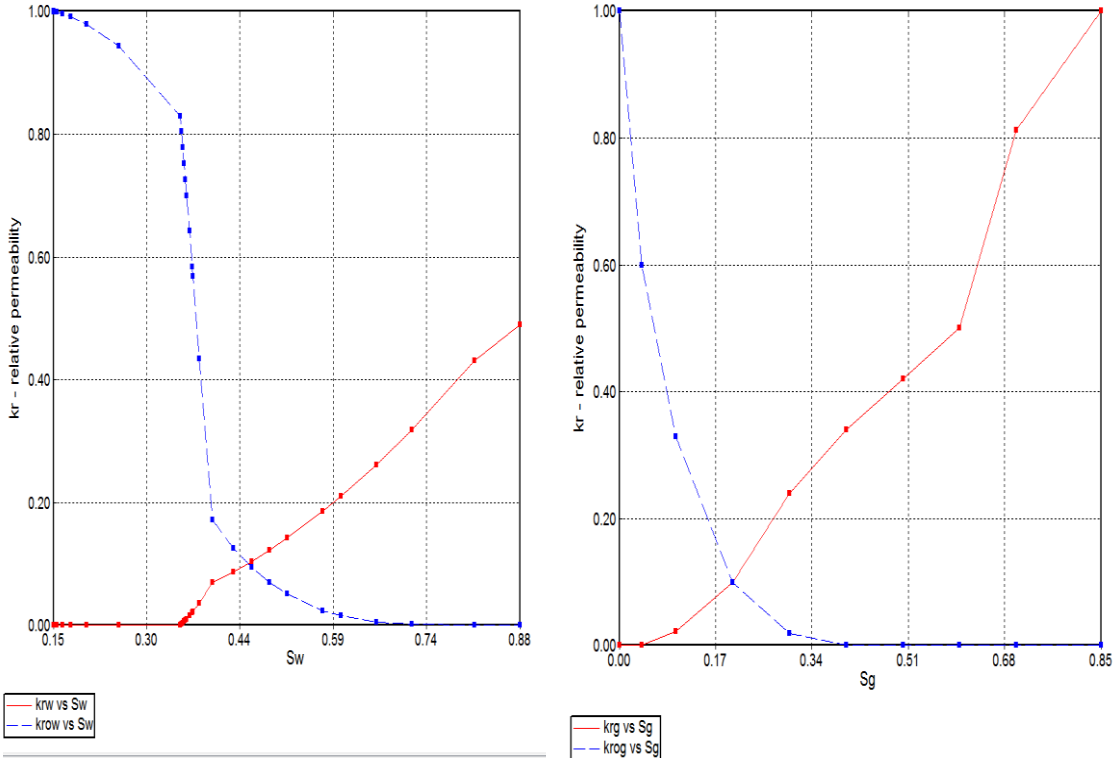

Figure 31: Relative permeability curves – WAI

4.1.7. Gas Injection (GAI)

Figures 32 and 33 show two cases of the same reservoir:

i) The reservoir is only being depleted by 5 producer wells.
ii) The reservoir has 5 producer wells and 1 injector wells injecting gas into the gas cap.

As can be seen from the average reservoir pressure curves of the two cases, the gas injection case provides good pressure support and maintains the reservoir pressure at a much higher pressure than the case with no injection. Moreover, the case with gas injection provides higher levels of both oil and gas recoveries. Cumulative oil recovery increases by around 86% due to the higher reservoir pressure. Understandably the gas production also increases significantly due to injected gas bypassing the oil. This proves that case (ii) is a good representation of secondary recovery by gas injection and it can used to provide input-output data for the system identification process.

This model is modified from mxdrm003.dat template and is actually based on the gas cap reservoir model obtained previously. *Figure 34* shows plots of the production parameters (outputs) we wish to predict and the corresponding injection rate of gas (input). The only addition in this model compared to the gas cap drive model is the addition of one injector well, operating at constant BHP of 25,000 kPa.

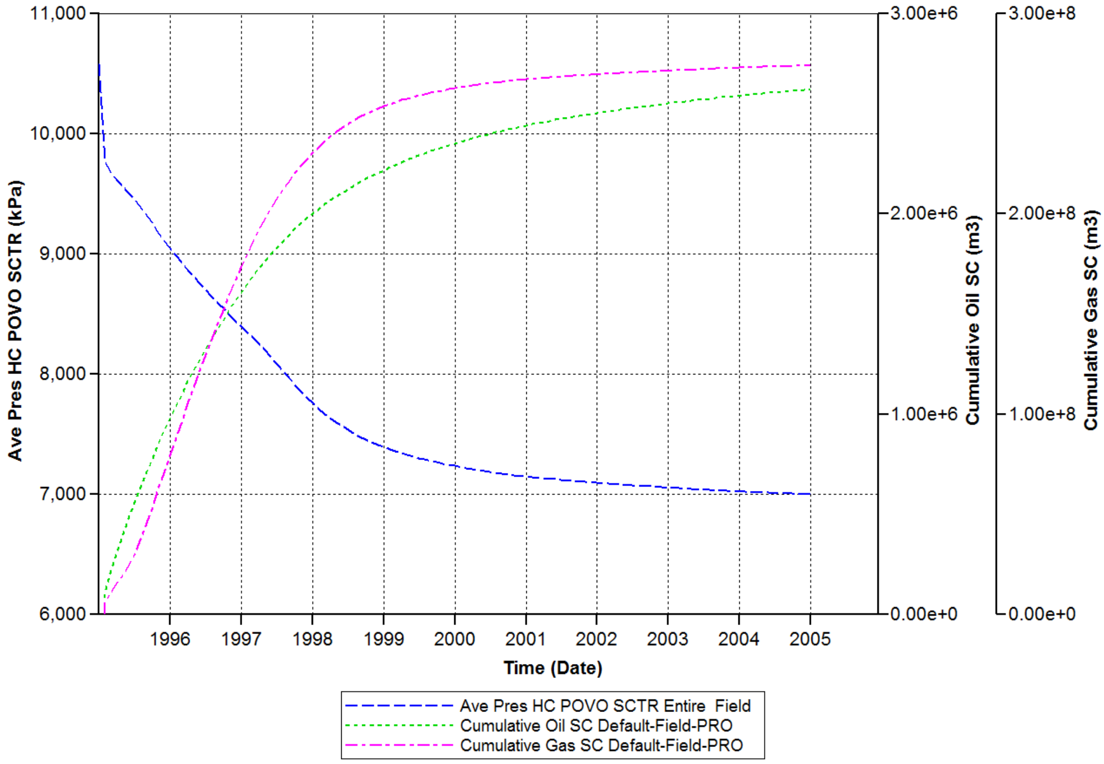

Figure 32: Identification of drive mechanism – GAI Case (i)

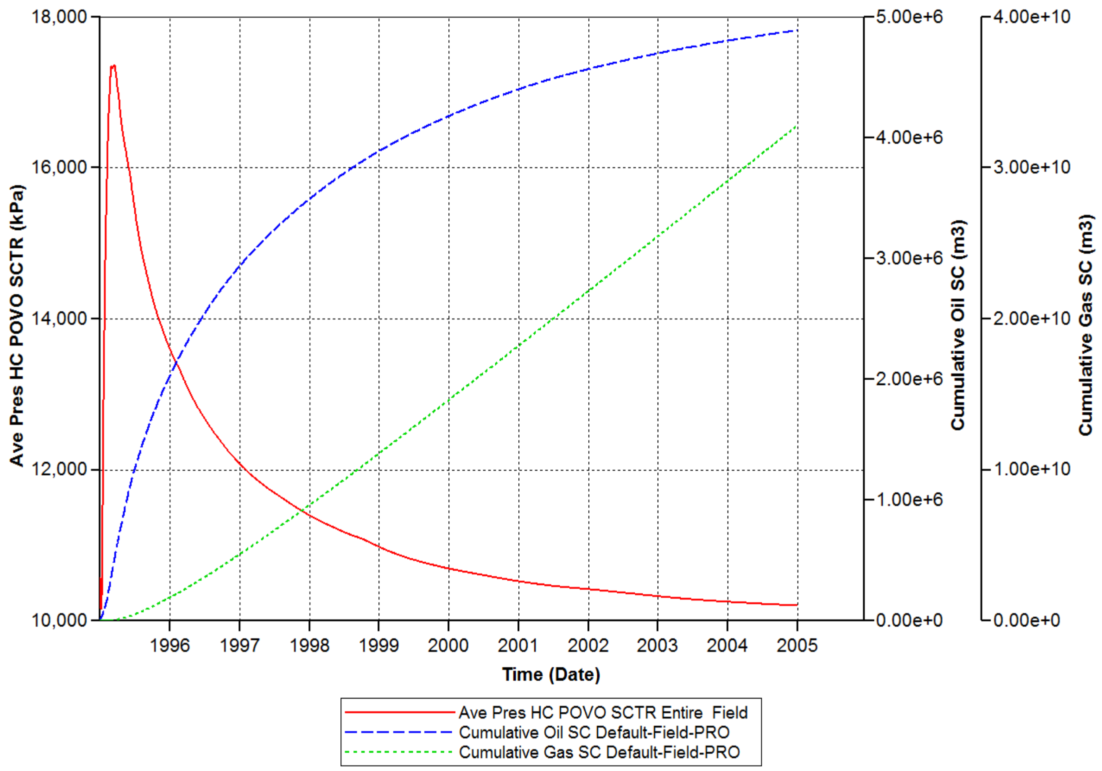

Figure 33: Identification of drive mechanism – GAI Case (ii)

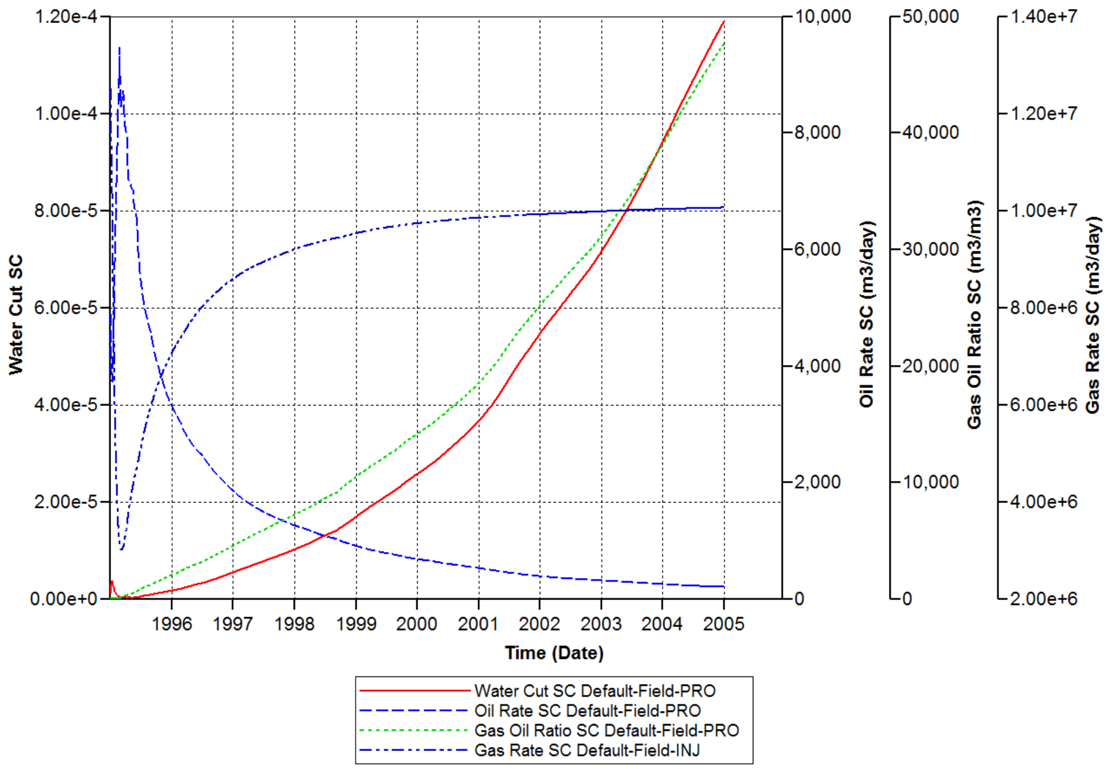

Figure 34: Production and injection parameters (Outputs and Input) – GAI

4.2. Detailed procedure for System identification and forecasting

The data sets (inputs and outputs) obtained from the data gathering process (section 4.1) were then used in the system identification and forecasting processes. The typical workflow used is shown as follows (water injection recovery mechanism used as example):

i. MATLAB system identification toolkit will be used to generate the polynomial SI models. However, pre-processing of data needs to take place beforehand. The data sets (*Figure 35*) are first inserted into MATLAB using SI commands, as shown in *Figure 36*. For each drive mechanism, there are three outputs under investigation, namely oil rate, water cut and GOR. These three curves will be treated separately with their own test and validation samples, meaning that each curve will be fitted to its own polynomial model. However, the input for each curve is the same, which is the corresponding water injection rate. Hence, as shown in *Figure 37*, each data set is split equally in half into a test data sample (starting with 't_') and a validation sample (starting with 'v_'). For the cases of primary drive mechanism, there is no input stream.

	A	B	C	D	E	F	G	H
1				Y1	Y2	Y3	Y4	
2				FIELD-PRO	FIELD-PRO	FIELD-PRO	FIELD-INJ	
3		TIME	DATE	Water Cut SC	Oil Rate SC	Gas Oil Ratio SC	Water Rate SC	
4		(day)		()	(bbl/day)	(ft3/bbl)	(bbl/day)	
5		1	1/2/1980	0.002048907	209502.4688	1564.181274	10934.10742	
6		2	1/3/1980	0.003416463	177679.7656	1578.222534	8973.727539	
7		3	1/4/1980	0.005393418	157394.0781	1674.689697	8039.971191	
8		4	1/5/1980	0.007199169	143095.5156	1740.460449	7552.12793	
9		5	1/6/1980	0.008789328	132890.4375	1780.214111	7282.821289	
10		6	1/7/1980	0.009953482	124821.8203	1814.606079	7130.952148	
11		7	1/8/1980	0.010917888	118263.5859	1843.669312	7046.686523	
12		8	1/9/1980	0.011781259	112631.2344	1871.103394	6999.083008	
13		9	1/10/1980	0.012661402	107731.7891	1894.941406	6978.487305	
14		10	1/11/1980	0.013300875	103509.5625	1914.447022	6975.972656	
15		11	1/12/1980	0.013968245	99779.05469	1930.896729	6985.940918	
16		12	1/13/1980	0.014686272	96439.48438	1946.056519	7004.501953	
17		13	1/14/1980	0.015422417	93413.75781	1961.150147	7029.156738	
18		14	1/15/1980	0.016163321	90655.40625	1976.625855	7058.159668	

Figure 35: Sample - data set for water injection recovery

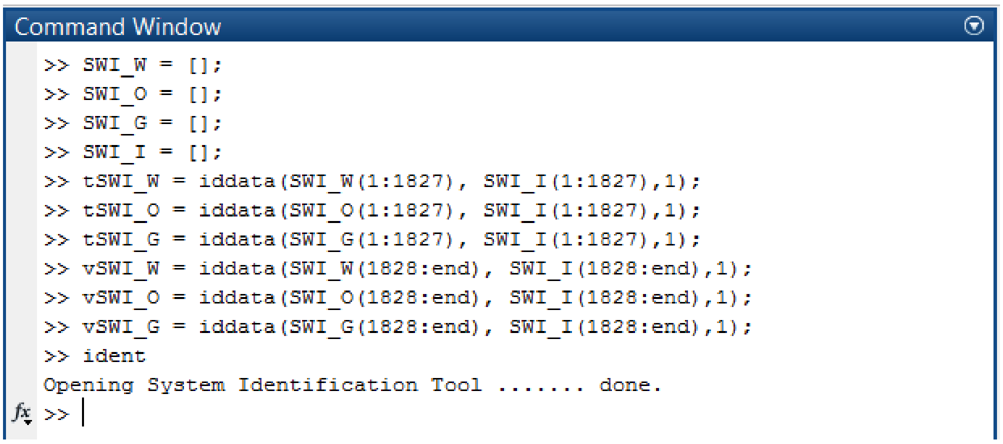

Figure 36: Sample - MATLAB Commands for water injection recovery

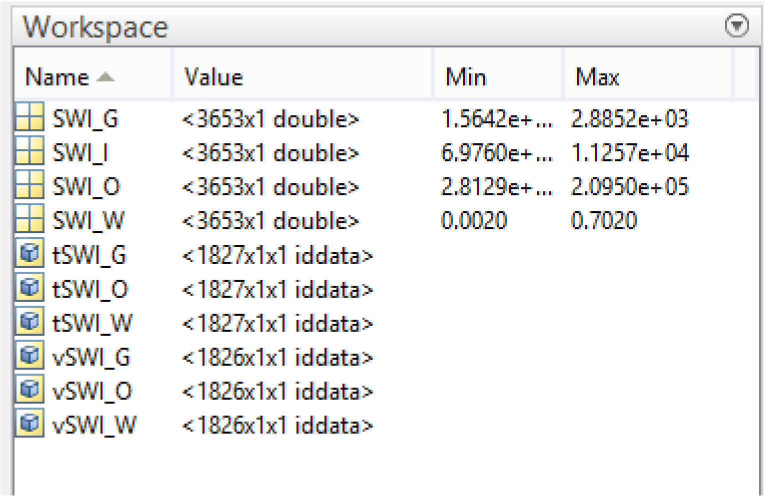

Figure 37: Sample - MATLAB objects created by MATLAB commands

As stated previously, the cross-validation method split used is 50:50. This means that the first half of data will be used for training and second half used for validation. The author decided on this ratio after looking at MATLAB manual examples and several papers that have references to cross-validation. A prominent paper (Browne, 2000) established that this is the classic way of splitting the data. The author also believes that this is the easiest method to visualize for the readers. This is because, considering whatever the period of data we have (let's say 1 year), we can make prediction of the exact same period (another one year) using the SI models.

ii. After data preparation, the SI toolkit is then opened and the data samples are loaded into the program, as shown in *Figure 38*. The polynomial model builder (*Figure 39*) will then be used to create different structures of polynomial models. Moreover, for each structure, there will be 10 models built with orders ranging from order 1 to order 10. An important note is that the author has decided to simplify the study and the analysis of results by setting the order of all the poles, zeros and delays to be the same (all changed together from order 1 to 10). Hence, they were not allowed to vary independently.

The model structures available for primary drive mechanisms (time series models) are:

- AutoRegressive (AR)
- AutoRegressive Integrated (ARI)
- AutoRegressive Moving Average (ARMA)
- AutoRegressive Integrated Moving Average (ARIMA)

The model structures available for the secondary drive mechanism are:

- AutoRegressive with eXogenous inputs (ARX)
- AutoRegressive Integrated with eXogenous inputs (ARIX)
- AutoRegressive Moving Average with eXogenous inputs (ARMAX)
- AutoRegressive Integrated Moving Average with eXogenous inputs (ARIMAX)
- Box-Jenkins (BJ)
- Box Jenkins Integrated (BJI).

Figure 38: Sample - System identification tool interface

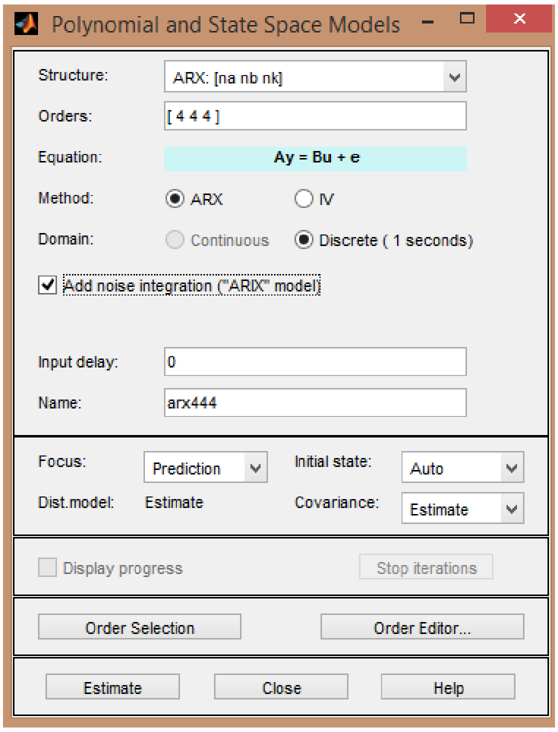

Figure 39: Sample - Polynomial model builder

iii. The MATLAB toolkit calculates the best fit of the models using Normalized Root Mean Square (NRMSE). NRMSE is the criterion used to generate the fit % number. It is a measure of how much better the model is in reproducing the observed data relative to the mean of the data. A percentage of zero indicates that the model does not predict values better than the mean value of the data.

iv. For each model structure, the results are plotted and ranked (*Figure 40*). A graph of Fit result vs. model order is then plotted (*Figure 41*) in order to be able to pick the best order for a given model structure. The best model is chosen by looking at the graph and seeing where the increase in accuracy with increasing order number starts to plateau. In other words, the best model order is the lowest order after which there is no more significant increase in accuracy when order is increased. For this example case the best order is 4. The author always tries to get the lowest possible order in order to decrease the complexity of the final polynomial model.

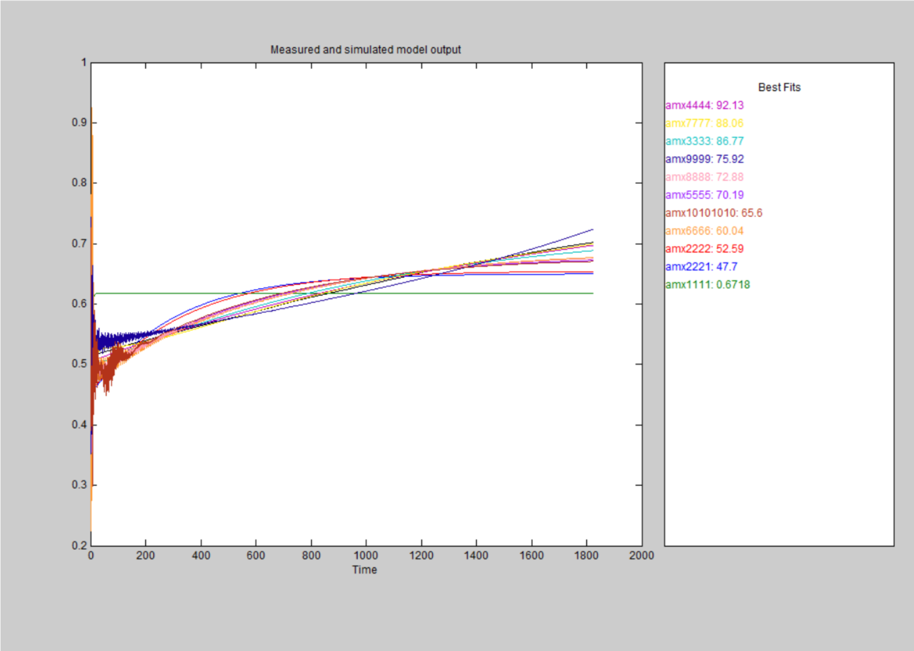

Figure 40: Sample - Fit results for ARIMAX water cut model (water injection case)

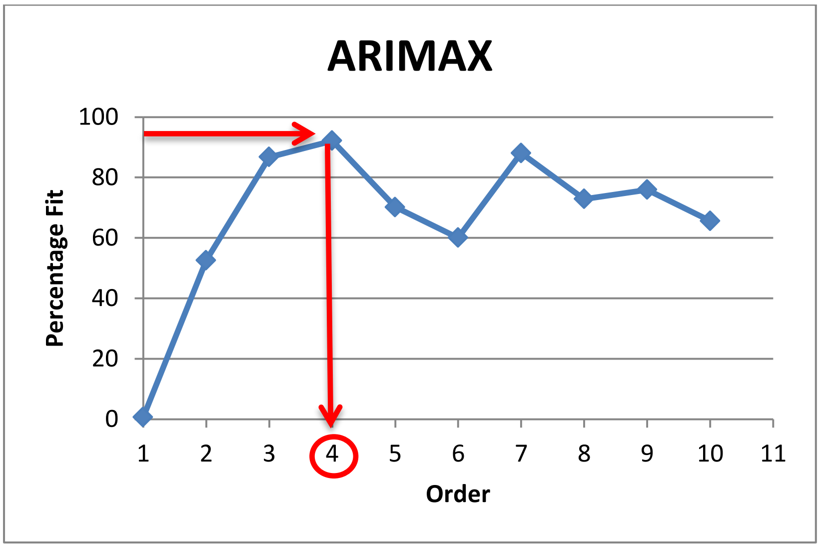

Figure 41: Sample - Fit results vs. model order for ARIMAX water cut model (water injection case)

v. Lastly, a graph is made for each output of each drive mechanism (*Figure 42*) to compare the fit results vs. order number for all model structures. This graph summarizes the prediction performance of all the polynomial model structures for that output. Note that results of percentage fits (NRSME) below 0% are taken out because it means that the model predict the observed data worse than the mean of the observed data.

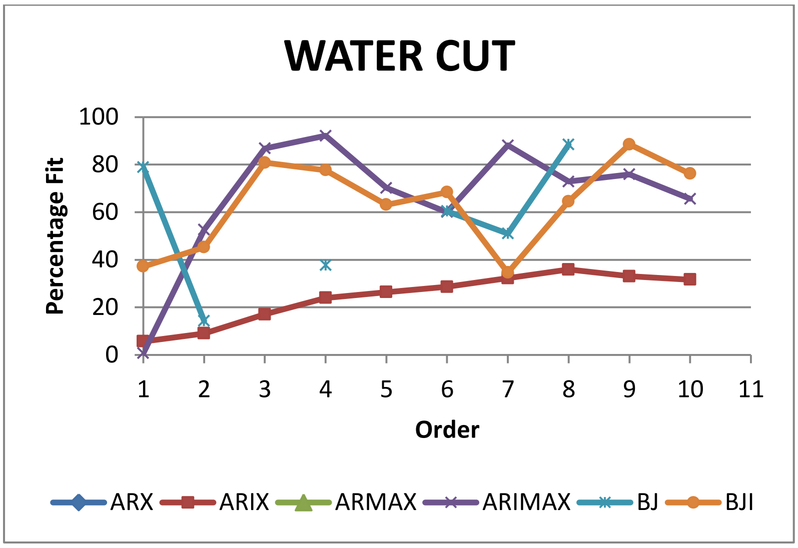

Figure 42: Sample - Fit results vs. model order for all water cut models (water injection case)

4.3. Analyzing prediction performance

Figures 43 to 63 show the line graphs of fit result vs model order. There is a curve for all tested model structures for each output of each drive mechanism. The best model orders for each model structure were chosen from these graphs according to the method stated in section 4.2.iii. *Table 2* is a table summarizing the best order numbers as well as the fit percentage for the different structures of models for each drive mechanism.

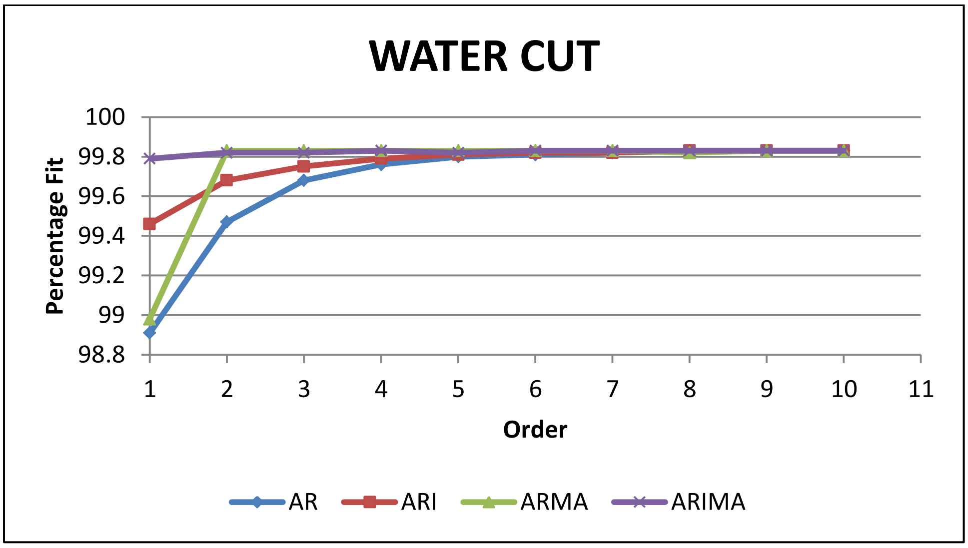

Figure 43: Results - Rock & Liquid Expansion Drive – Water Cut

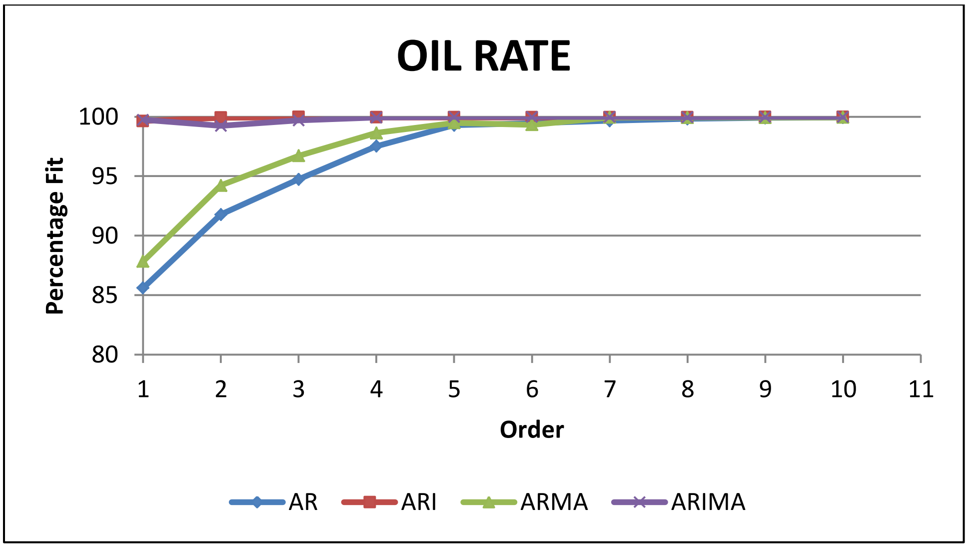

Figure 44: Results - Rock & Liquid Expansion Drive – Oil Rate

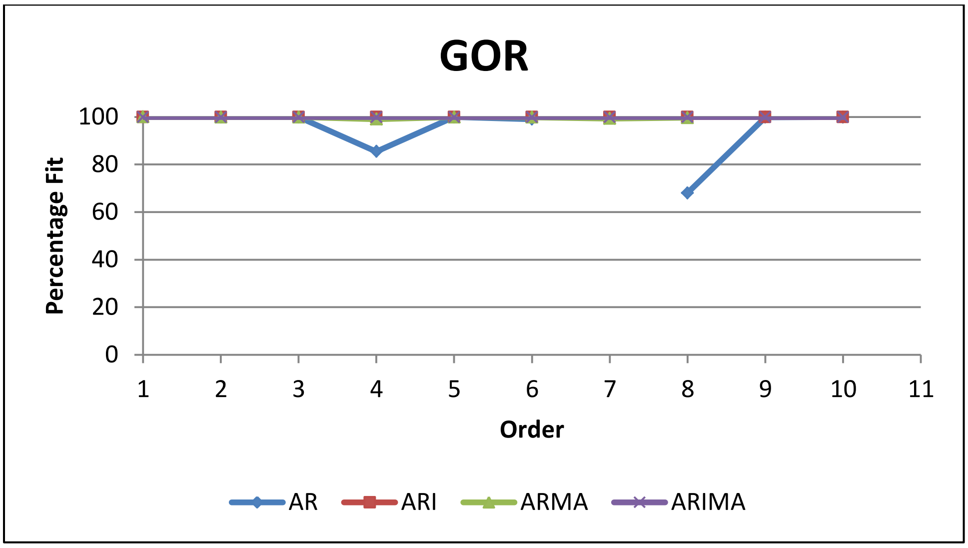

Figure 45: Results - Rock & Liquid Expansion Drive – Gas Oil Ratio

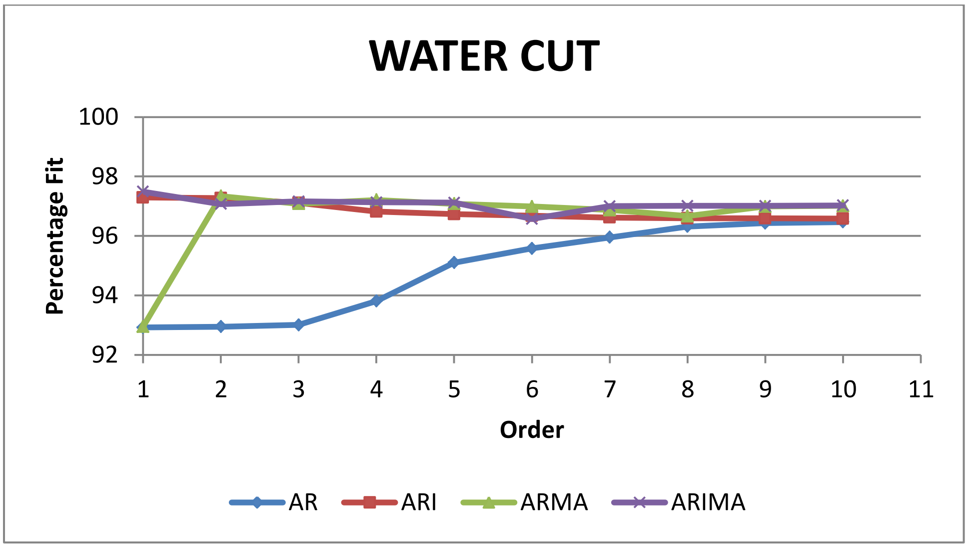

Figure 46: Results - Solution Gas Drive – Water Cut

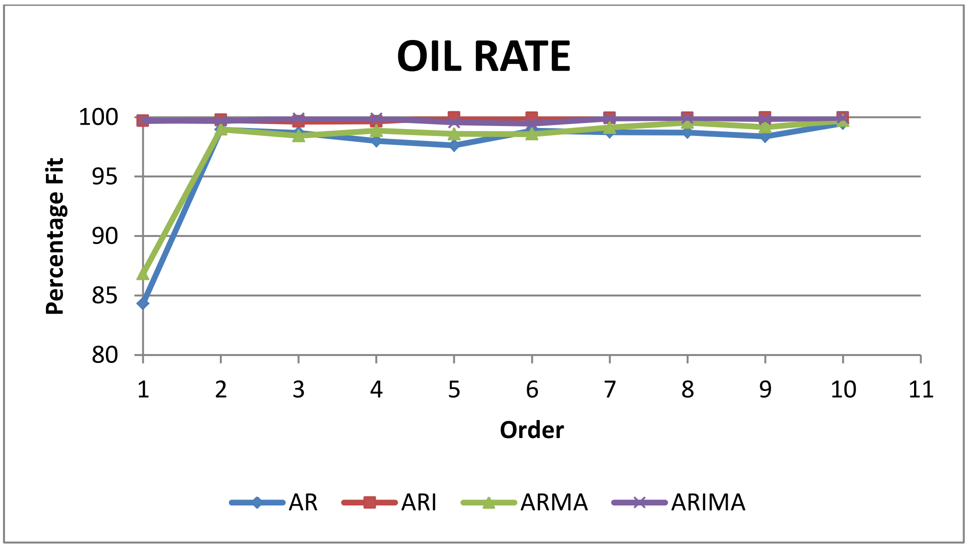

Figure 47: Results - Solution Gas Drive – Oil Rate

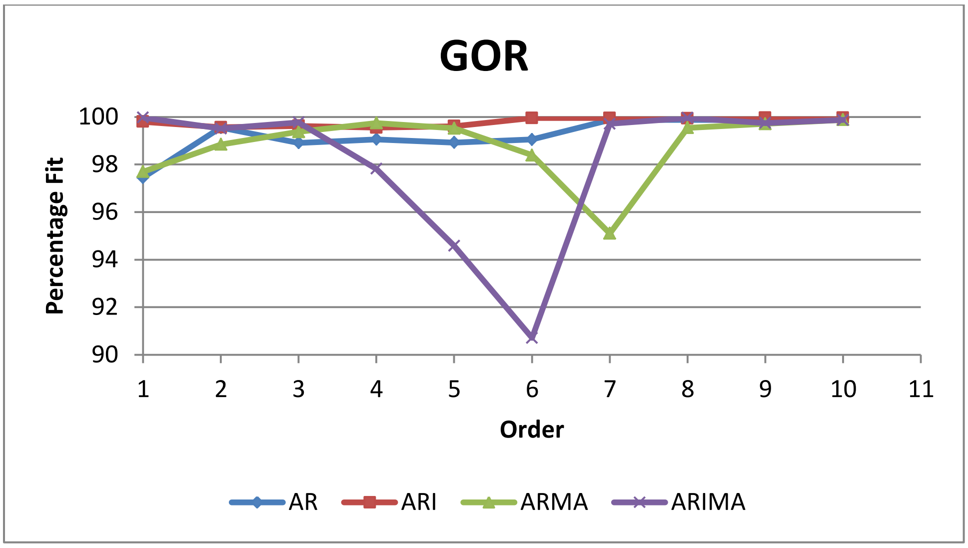

Figure 48: Results - Solution Gas Drive – Gas Oil Ratio

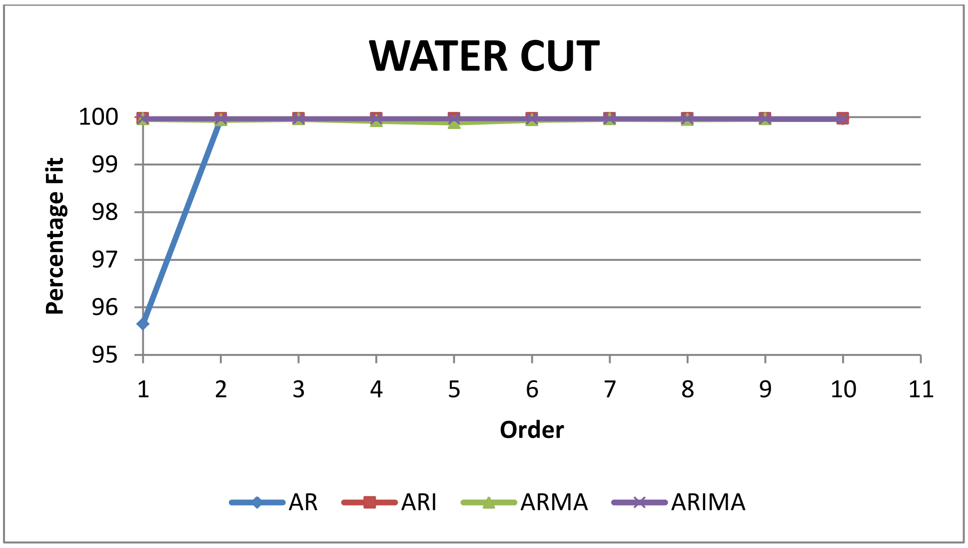

Figure 49: Results - Aquifer Drive – Water Cut

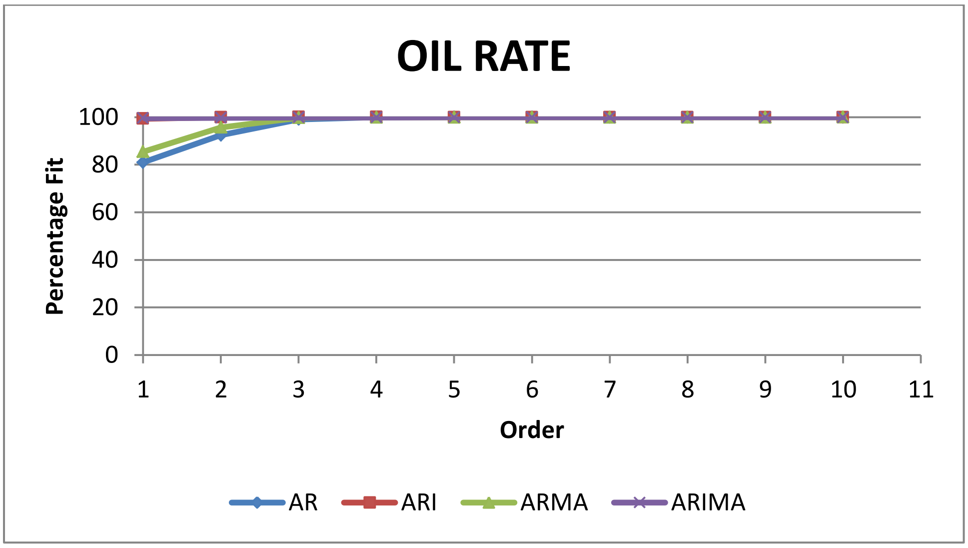

Figure 50: Results - Aquifer Drive – Oil Rate

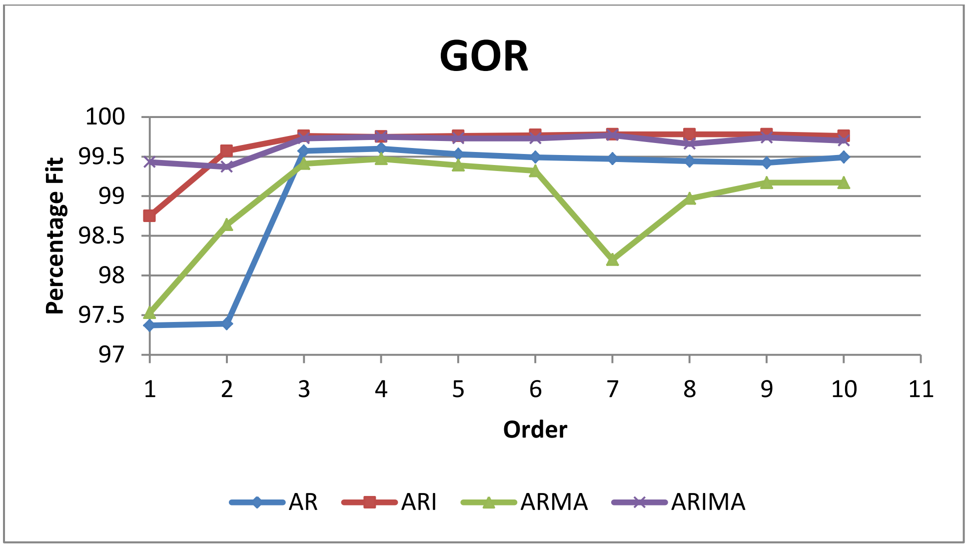

Figure 51: Results - Aquifer Drive – Gas Oil Ratio

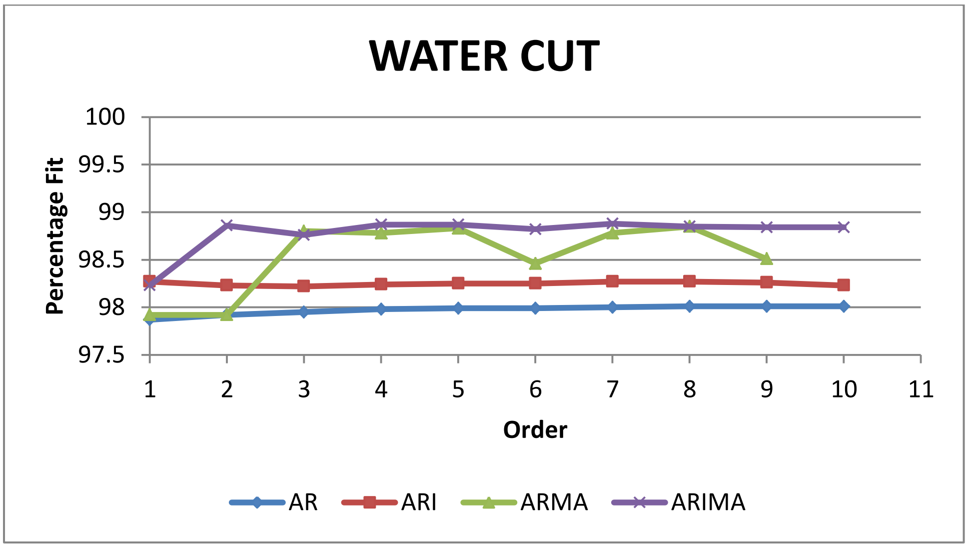

Figure 52: Results - Gas Cap Drive – Water Cut

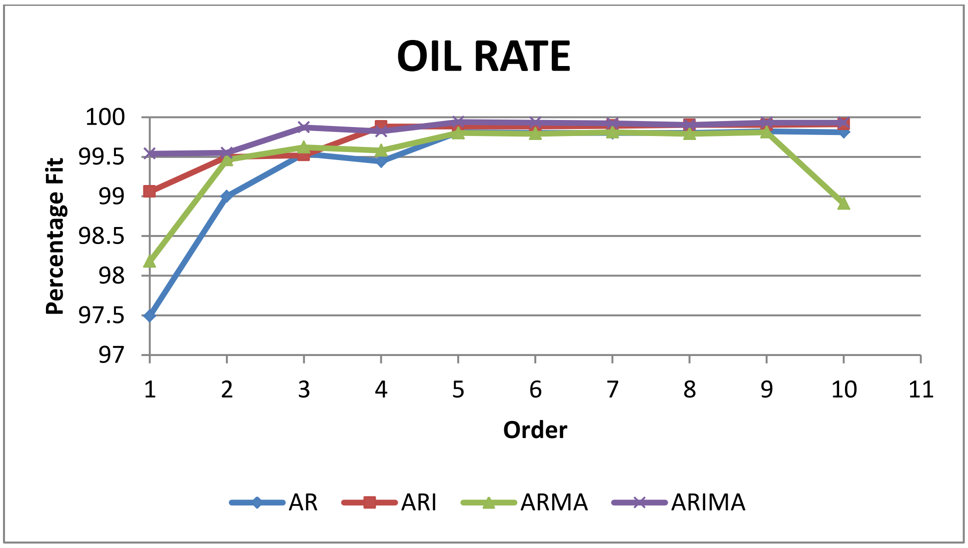

Figure 53: Results - Gas Cap Drive – Oil Rate

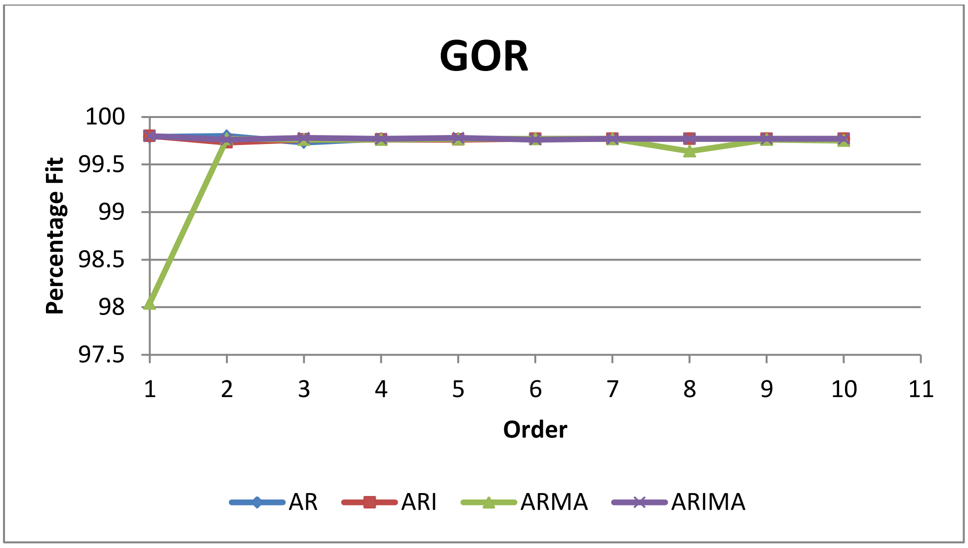

Figure 54: Results - Gas Cap Drive – Gas Oil Ratio

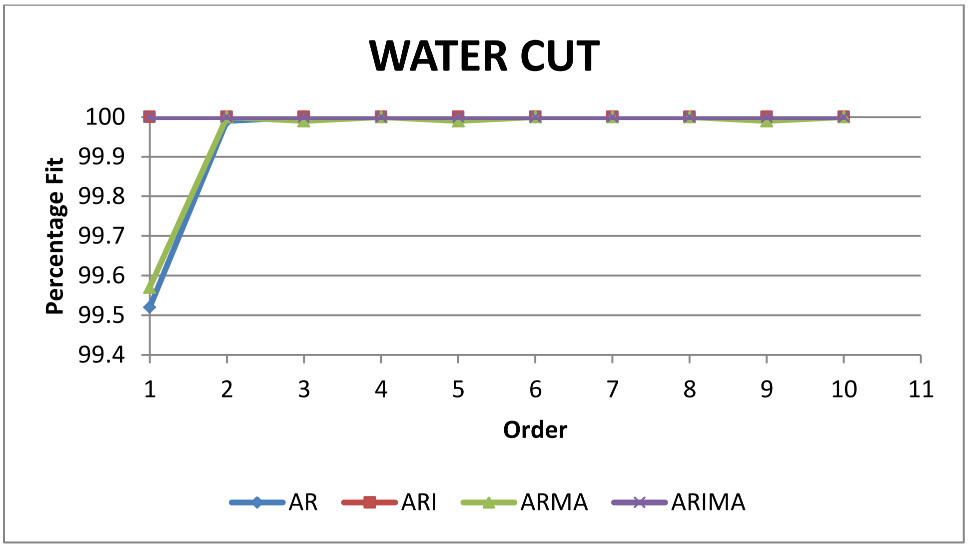

Figure 55: Results - Combined Primary Drive – Water Cut

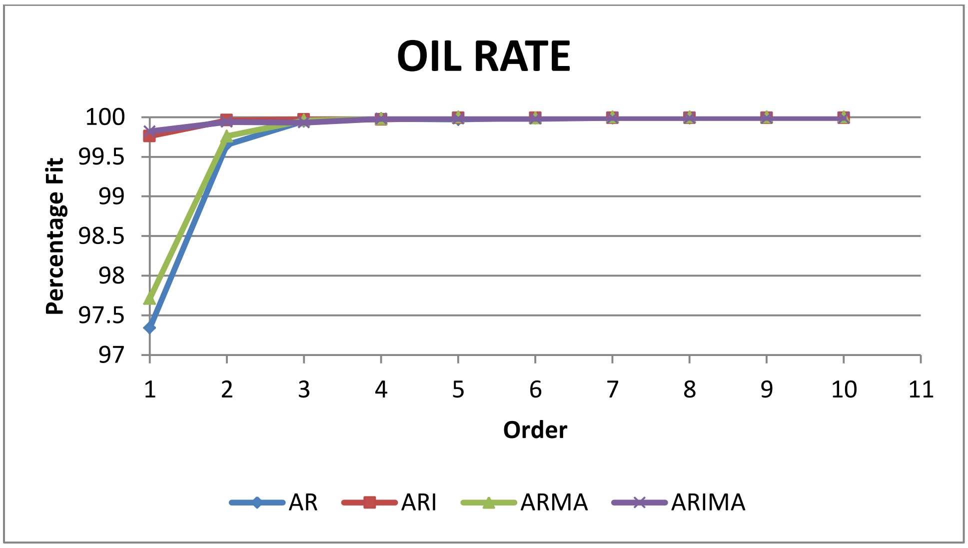

Figure 56: Results - Combined Primary Drive – Oil Rate

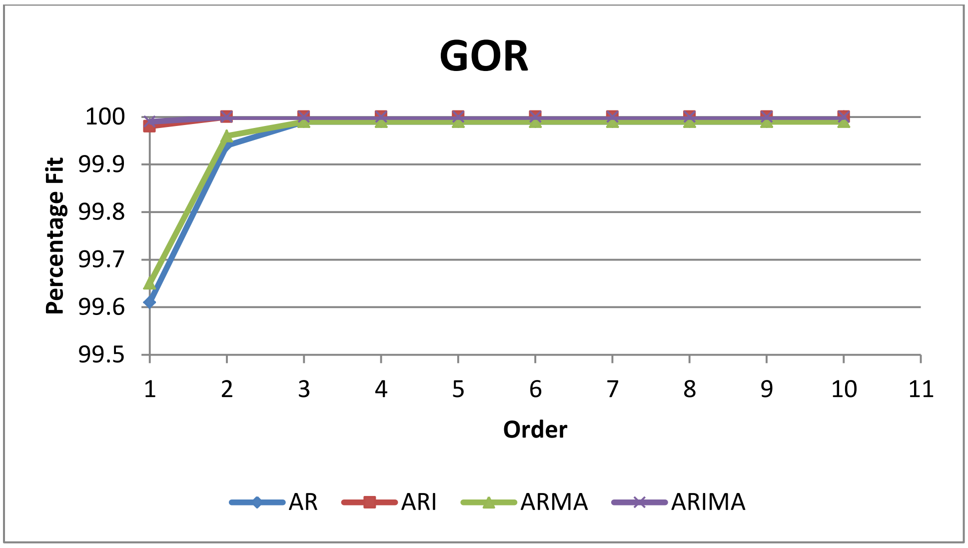

Figure 57: Results - Combined Primary Drive – Gas Oil Ratio

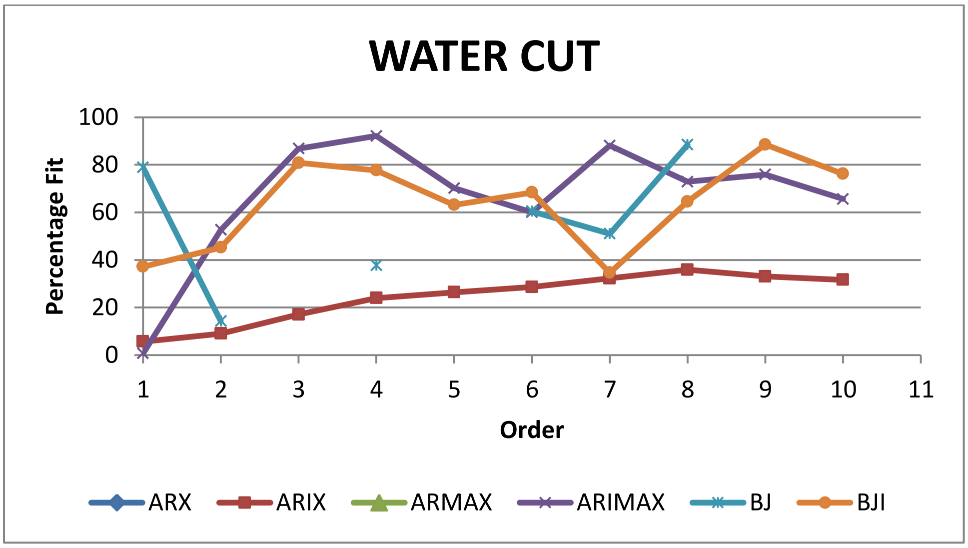

Figure 58: Results - Water Injection – Water Cut

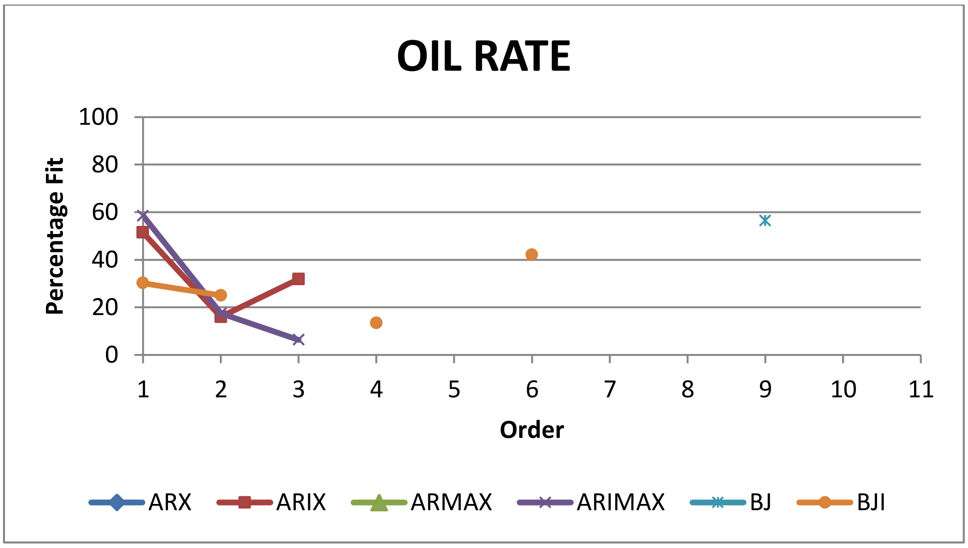

Figure 59: Results - Water Injection – Oil Rate

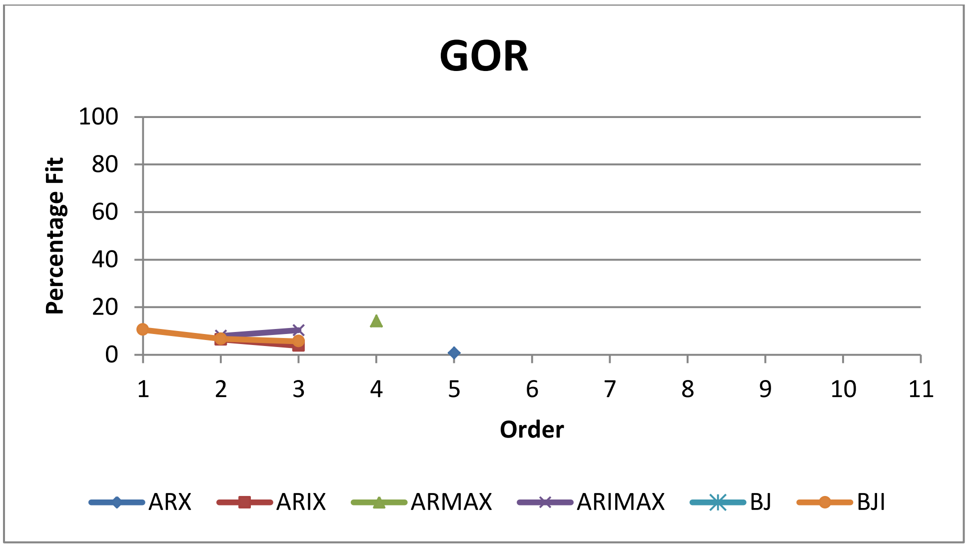

Figure 60: Results - Water Injection – Gas Oil Ratio

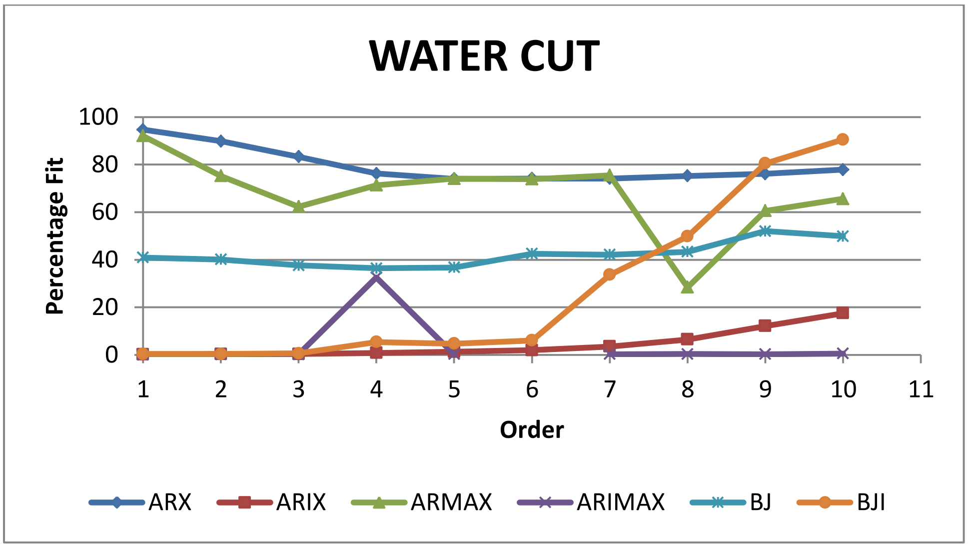

Figure 61: Results - Gas Injection – Water Cut

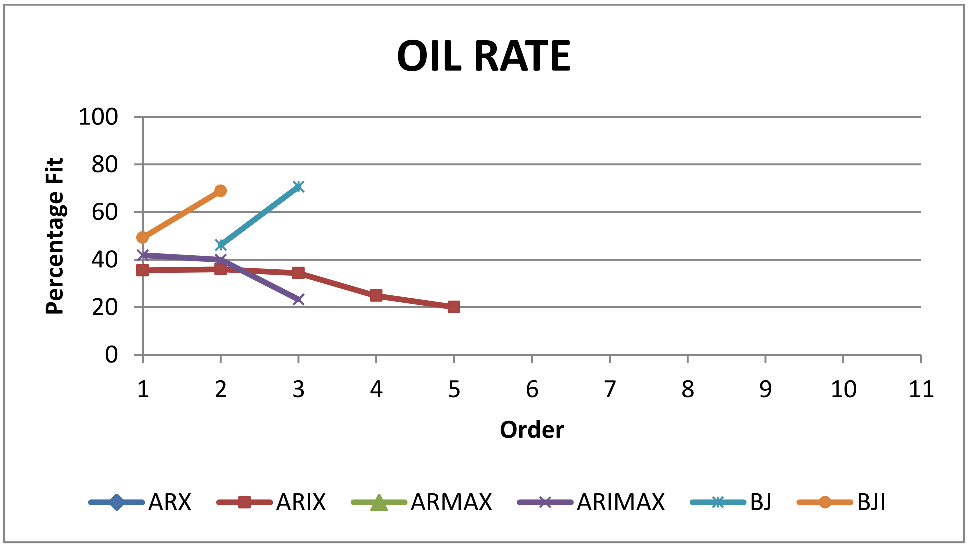

Figure 62: Results - Gas Injection – Oil Rate

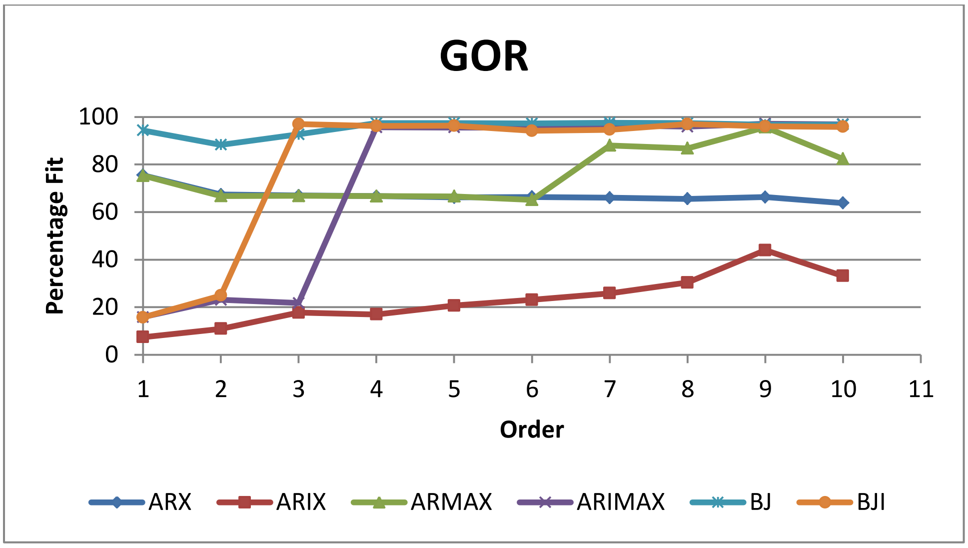

Figure 63: Results - Gas Injection – Gas Oil Ratio

From *Table 2* it is very clear to see that for all five primary drive mechanisms, all polynomial model structures have managed to predict the validation data sets (5 years production profiles) extremely well. The fit percentage is well over 97% for all the best model orders. These are excellent results because rarely are predictions this accurate, even using reservoir simulation. This shows that time series analysis can be a very reliable forecasting tool and that it can establish itself among the established forecasting methods. However, it should be noted that the data sets used in this thesis assume that there is no noise in the data. Noise here refers to measurement errors due sampling method as well as accuracy limitations of measurement devices. Hence, research should be done into investigating the effect of noise on the prediction accuracy of these time series models.

Furthermore, for most of the model structures, the model order for the best fit is of order 3 or less. This is also a good result because this means that we do not need overly complex polynomial models that have large number of parameters in order to get more than satisfactory levels of prediction.

RESULTS TABLE				
Recovery Mechanism	Model structure	Best Model Order : Fit percentage		
		OIL RATE	WATER CUT	GAS:OIL RATIO
Rock & Liquid Expansion Drive	AR	5 : 99.28	3 : 99.68	2 : 99.92
	ARI	2 : 99.92	1 : 99.46	1 : 99.93
	ARMA	5 : 99.49	2 : 99.83	1 : 99.96
	ARIMA	1 : 99.74	1 : 99.79	1 : 99.93
Solution Gas Drive	AR	3 : 99.54	1 : 97.87	1 : 99.79
	ARI	1 : 99.06	1 : 98.27	1 : 99.8
	ARMA	2 : 99.46	1 : 97.89	2 : 99.77
	ARIMA	1 : 99.54	2 : 98.86	1 : 99.8
Gas Cap Drive	AR	2 : 98.95	7 : 95.95	2 : 99.55
	ARI	2 : 99.75	1 : 97.29	1 : 99.8
	ARMA	2 : 98.97	2 : 97.34	4 : 99.74
	ARIMA	1 : 99.72	1 : 97.49	1 : 99.97
Aquifer Drive	AR	3 : 99.06	2 : 99.96	3 : 99.57
	ARI	3 : 99.96	1 : 99.97	3 : 99.76
	ARMA	3 : 99.72	2 : 99.96	3 : 99.41
	ARIMA	3 : 99.88	1 : 99.97	3 : 99.73
Combined Drive	AR	3 : 99.95	3 : 100	3 : 99.99
	ARI	2 : 99.96	1 : 100	2 : 100
	ARMA	3 : 99.96	2 : 100	3 : 99.99
	ARIMA	2 : 99.94	1 : 100	2 : 100
Water Injection	ARX	N/A	N/A	5 : 0.765
	ARIX	1 : 51.42	8 : 35.84	2 : 6.427
	ARMAX	N/A	N/A	4 : 14.14
	ARIMAX	1 : 58.46	4 : 92.13	3 : 10.33
	BJ	9 : 56.36	8 : 88.41	N/A
	BJI	6 : 42.04	9 : 88.5	1 : 10.5
Gas Injection	ARX	N/A	1 : 94.68	1 : 75.51
	ARIX	2 : 35.95	10 : 17.5	9 : 43.95
	ARMAX	N/A	1 : 92.03	9 : 95.71
	ARIMAX	1 : 41.79	4 : 32.57	4 : 95.61
	BJ	3 : 70.61	9 : 51.99	4 : 97.38
	BJI	2 : 68.87	10 : 90.48	3 : 96.96

Table 2: Results Table

However, for the secondary recovery cases, the prediction performances of most the model structures do not show results that are as good as the results for primary production. What is most evident is that for the oil rate and GOR curves of water injection case as well as the oil rate curve of the gas injection case, there are no model structures that could predict with a fit percentage greater than 90%. This could indicate that the relationship between the input (displacing phase injection rate) and these output curves are not linear and this may be why these linear models cannot adequately model the relationship. However, for the remaining three parameters, the results show that good prediction (above 90% fit) can be obtained from one or more model structures. It seems that more studies need to be done for the cases of gas and water injection. Research could be done into investigating if changing the orders of the poles, zeros and delays of the model independently of each other can yield better fit results. Research could also be done to investigate if single input-multiple output (SIMO) models can yield better fit results because the output parameters would have each other to benchmark themselves to. If all that does not help to improve prediction accuracy, then research can be done into using non-linear SI models for the purpose of modelling gas and water injection.

Tables 3 and 4 in the next few pages are two versions of the final conceptual framework, which is derived from the results of *Table 2*. Reservoir engineers can refer to any of the two versions. The way they should use it is by first choosing what production parameter they want to predict and then looking under the drive mechanism which corresponds to their reservoir. The numbers, 1 to 4 for primary drive and 1 to 6 for secondary drive, show the accuracy of the model structures, with 1 being most accurate and increasing number being less accurate. This framework will provide a good starting point for engineers so that they do not have to test so many different model structures with many different order numbers. Rather, they would have a guideline of recommended models based on decreasing accuracy, all of which are established from the results of this thesis.

Table 3: Conceptual framework version 1

OIL RATE

SI model

Drive Mechanism	AR	ARI	ARMA	ARIMA	ARX	ARIX	ARMAX	ARIMAX	BJ	BJI
Rock & Liquid Expansion Drive	4	1	3	2						
Solution Gas Drive	2	4	3	1						
Gas Cap Drive	4	1	3	2						
Aquifer Drive	4	1	3	2						
Combined Drive	3	1	2	4						
Water Injection					5	3	6	1	2	4
Gas Injection					5	4	6	3	1	2

WATER CUT

SI model

Drive Mechanism	AR	ARI	ARMA	ARIMA	ARX	ARIX	ARMAX	ARIMAX	BJ	BJI
Rock & Liquid Expansion Drive	3	4	1	2						
Solution Gas Drive	4	2	3	1						
Gas Cap Drive	4	3	2	1						
Aquifer Drive	3	1	4	2						
Combined Drive	1	2	3	4						
Water Injection					5	4	6	1	3	2
Gas Injection					1	6	2	5	4	3

GAS-OIL-RATIO

SI model

Drive Mechanism	AR	ARI	ARMA	ARIMA	ARX	ARIX	ARMAX	ARIMAX	BJ	BJI
Rock & Liquid Expansion Drive	4	3	1	2						
Solution Gas Drive	3	1	4	2						
Gas Cap Drive	4	2	3	1						
Aquifer Drive	3	1	4	2						
Combined Drive	3	1	4	2						
Water Injection					5	4	1	3	6	2
Gas Injection					5	6	3	4	1	2

Table 4: Conceptual framework version 2

Chapter 5: Conclusion and recommendation for future work

In conclusion, system identification is a very promising production forecasting method that deserves much further investigation. In addition to just investigating one recovery method, the purpose of this project is to create a framework that connects various drive mechanism its most suitable forecasting model. This framework will serve as a reference to reservoir engineers help to speed up the identification process when modelling their own reservoir using system identification.

The results show that SI polynomial models can provide an excellent set of tools to predict oil rate, water cut and GOR for reservoirs under the drive mechanisms listed in this study. Time series models can predict production parameters of reservoirs under primary drive mechanisms with up to 100% accuracy NRSME. Meanwhile, reservoirs under secondary drive mechanisms can also make use of system identification models, with some models having prediction accuracy well above 90%. However, more research needs to be done to improve the prediction accuracy for secondary drive mechanisms. This is due to the increased complexity of the models and the presence of input data.

System identification based reservoir models can be established as a practical, cost-effective and robust tool for forecasting reservoir fluid production. The procedures described in this thesis as well as the final conceptual model can serve as a framework or guide to reservoir engineers if they wish to implement system identification for production forecasting.

The author recommends that more study be done to increase our understanding of how system identification can be turned into a proven forecasting method in petroleum engineering. The recommended research areas are:

1) To use these algorithms on real reservoir data and to investigate the effect of measurement noise.
2) Investigate if changing the orders of the poles, zeros and delays of the model independently of each other can yield better fit results, especially for the secondary recovery mechanisms.
3) Investigate if single input- multiple output models can yield better fit results for the secondary recovery mechanisms.

4) Investigate if multivariate time series can provide better fit results for primary drive mechanisms when taking into account how much training data is available as compared to single variable time series.
5) Investigate if non-linear models can provide better forecasting accuracy than linear models for the secondary drive mechanisms.

References

- Nelles, O. (2001). Linear Dynamic System Identification. In Nonlinear System Identification (p. 786). Springer.
- Ljung, L. (1987). Introduction. In System Identification: Theory for the user (p. 238). P T R Prentice Hall.
- Keesman, K. (2011). System identification an introduction. London: Springer.
- Billings, S. (2013). Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains.
- Tarek, A. (2010). Oil recovery mechanisms and the material balance equation. In Reservoir Engineering Handbook (4th ed., p. 1454). Gulf Professional Publishing.
- Olominu, O., & Sulaimon, A. A. (2014, August 5). Application of Time Series Analysis to Predict Reservoir Production Performance. Society of Petroleum Engineers. doi:10.2118/172395-MS
- Elgsaeter, S., Slupphaug, O., & Johansen, T. A. (2008, January 1). Production optimization; System Identification and Uncertainty Estimation. Society of Petroleum Engineers. doi:10.2118/112186-MS
- Renard, G., Dembele, D., Lessi, J., & Mari, J. L. (1998, January 1). System Identification Approach to Watercut Analysis in Waterflooded Layered. Society of Petroleum Engineers. doi:10.2118/39606-MS
- Mannon, R. W. (1964, January 1). Some Aspects Of Production Forecasting. Society of Petroleum Engineers. doi:10.2118/832-MS
- Lockwood, G. W., & Cannon, D. E. (1982, February 1). Production Forecasting. Society of Petroleum Engineers. doi:10.2118/9946-PA
- Tomomi, Y. (2000, January 1). Non-Uniqueness of History Matching. Society of Petroleum Engineers. doi:10.2118/59434-MS
- Li, K., & Horne, R. N. (2005, January 1). Verification of Decline Curve Analysis Models for Production Prediction. Society of Petroleum Engineers. doi:10.2118/93878-MS
- Hickman, T. S. (1995, October 1). A Rationale for Reservoir Management Economics. Society of Petroleum Engineers. doi:10.2118/26411-PA
- Spencer, J. A., & Morgan, D. T. K. (1998, January 1). Application of Forecasting and Uncertainty Methods to Production. Society of Petroleum Engineers. doi:10.2118/49092-MS

- Satter, A., & Iqbal, G. (2008). Practical enhanced reservoir engineering: Assisted with simulation software. Tulsa, Okla.: PennWell.
- Peirce, C. (1983). Studies in logic. Amsterdam: John Benjamins Pub.
- Browne, M. (2000). Cross-Validation Methods. Journal of Mathematical Psychology, 108-132.
- Mathworks. (2015). System Identification Toolbox: User's Guide (r2015a). Retrieved April 1, 2015 from http://www.mathworks.com/help/pdf_doc/ident/ident.pdf
- Instruments, N. (2010, June 21). When and How Should I Use System Identification. Retrieved August 11, 2015, from http://www.ni.com/white-paper/11646/en/
- AlAjmi, M. D., Abdulraheem, A., Mishkhes, A. T., & Al-Shammari, M. J. (2015, March 3). Profiling Downhole Casing Integrity Using Artificial Intelligence. Society of Petroleum Engineers. doi:173422-MS
- Shahkarami, A., Mohaghegh, S. D., Gholami, V., & Haghighat, S. A. (2014, April 17). Artificial Intelligence (AI) Assisted History Matching. Society of Petroleum Engineers. doi:10.2118/169507-MS
- Enyioha, C., & Ertekin, T. (2014, April 1). Advanced Well Structures: An Artificial Intelligence Approach to Field Deployment and Performance Prediction. Society of Petroleum Engineers. doi:10.2118/167870-MS
- Anifowose, F. A., Abdulraheem, A., Al-Shuhail, A. A., & Schmitt, D. P. (2013, March 10). Improved Permeability Prediction From Seismic and Log Data using Artificial Intelligence Techniques. Society of Petroleum Engineers. doi:10.2118/164465-MS
- Chen, H., Klie, H., & Wang, Q. (2013, February 18). A Black-Box Interpolation Method To Accelerate Reservoir Simulation Solutions. Society of Petroleum Engineers. doi:10.2118/163614-MS
- Marek Brabec, Marius Paulescu, Viorel Badescu, Tailored vs black-box models for forecasting hourly average solar irradiance, Solar Energy, Volume 111, January 2015, Pages 320-331, ISSN 0038-092X, http://dx.doi.org/10.1016/j.solener.2014.11.003.
- Laura Prada, Javier Garcia, Alejandro Calderon, J. Daniel Garcia, Jesus Carretero, A novel black-box simulation model methodology for predicting performance and energy consumption in commodity storage devices, Simulation Modelling

Practice and Theory, Volume 34, May 2013, Pages 48-63, ISSN 1569-190X, http://dx.doi.org/10.1016/j.simpat.2013.01.006.

- Vito Logar, Živa Kristl, Igor Škrjanc, Using a fuzzy black-box model to estimate the indoor illuminance in buildings, Energy and Buildings, Volume 70, February 2014, Pages 343-351, ISSN 0378-7788, http://dx.doi.org/10.1016/j.enbuild.2013.11.082.
- Theofilos A. Papadopoulos, Andreas I. Chrysochos, Angelos I. Nousdilis, Grigoris K. Papagiannis, Simplified measurement-based black-box modeling of distribution transformers using transfer functions, Electric Power Systems Research, Volume 121, April 2015, Pages 77-88, ISSN 0378-7796, http://dx.doi.org/10.1016/j.epsr.2014.12.003.
- Alexandra Grancharova, Juš Kocijan, Tor A. Johansen, Explicit output-feedback nonlinear predictive control based on black-box models, Engineering Applications of Artificial Intelligence, Volume 24, Issue 2, March 2011, Pages 388-397, ISSN 0952-1976, http://dx.doi.org/10.1016/j.engappai.2010.10.009.
- Timothy A Haley, Steven J Mulvaney, On-line system identification and control design of an extrusion cooking process: Part I. System identification, Food Control, Volume 11, Issue 2, April 2000, Pages 103-120, ISSN 0956-7135, http://dx.doi.org/10.1016/S0956-7135(99)00068-7.
- N. Angarita-Jaimes, H. Kouchakpour, J. Liu, R.B. Panerai, D.M. Simpson, Optimising the assessment of cerebral autoregulation from black box models, Medical Engineering & Physics, Volume 36, Issue 5, May 2014, Pages 607-612, ISSN 1350-4533, http://dx.doi.org/10.1016/j.medengphy.2013.12.012.
- O. Criner, Control Systems Identification In Finance And Economics, 2008, ISSN 1743-3517, doi:10.2495/CF080011.
- "Documentation." Goodness of Fit between Test and Reference Data. Web. 11 Aug. 2015. <http://www.mathworks.com/help/ident/ref/goodnessoffit.html>.

Appendices

Sample MATLAB code for primary drive - combined drive (COD).

The following code is used for entering test data, creating objects, splitting them into training and validation sets and opening MATLAB SI toolkit:

```
COD_W = [];
COD_O = [];
COD_G = [];
tCOD_W = iddata(COD_W(1:1827),[],1);
tCOD_O = iddata(COD_O(1:1827),[],1);
tCOD_G = iddata(COD_G(1:1827),[],1);
vCOD_W = iddata(COD_W(1828:end),[],1);
vCOD_O = iddata(COD_O(1828:end),[],1);
vCOD_G = iddata(COD_G(1828:end),[],1);
```

Sample MATLAB code for secondary drive - water injection (SWI).

The following code is used for entering test data, creating objects, splitting them into training and validation sets and opening MATLAB SI toolkit:

```
SWI_W = [];
SWI_O = [];
SWI_G = [];
SWI_I = [];
tSWI_W = iddata(SWI_W(1:1827), SWI_I(1:1827),1);
tSWI_O = iddata(SWI_O(1:1827), SWI_I(1:1827),1);
tSWI_G = iddata(SWI_G(1:1827), SWI_I(1:1827),1);
vSWI_W = iddata(SWI_W(1828:end), SWI_I(1828:end),1);
vSWI_O = iddata(SWI_O(1828:end), SWI_I(1828:end),1);
vSWI_G = iddata(SWI_G(1828:end), SWI_I(1828:end),1);
```