University of Alberta

A Study on System Identification and Input Design for SAGD Reservoirs

by

Song Yao

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

in

Petroleum Engineering

Department of Civil and Environmental Engineering

©Song Yao
Spring 2014
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Abstract

Physics-based large-scale reservoir models are routinely employed in the prediction of the SAGD (Steam Assisted Gravity Drainage) process under different operation situations. However, due to the uncertainty of the reservoir and the limitations of the commercial reservoir simulators, the computational time is highly associate with impractical simulated results for all locations, especially when uncertainty and unexpected operational parameters are included. This thesis develops a system identification proxy model to forecast the SAGD reservoir production. Several combinations of system identification model structures and input datasets are tested for short-term predictions to understand the impact of model structures and input selection on the proxy model performance. Then recursive proxy model estimations are performed to increase the accuracy for long-term production prediction. After trying to improve the model fits for a multiple-step proxy model prediction, a set of input design cases using a simulator-built model are re-identified to test the validity of open-loop simulation by a data-driven proxy model. The data-driven proxy model could be used in reservoir management and optimization or to reduce the computing load.

Acknowledgements

First and foremost I offer my sincerest gratitude to my supervisors, Dr. Japan Trivedi and Dr. Vinay Prasad, who have supported me throughout my thesis with their patience and knowledge. I want to thank them for their constant guidance, support, motivation and ongoing help during the course of my Master's program. One simply could not wish for better or friendlier supervisors.

I would like to thank Siavash Nejadi for his valuable guidance in the use of Petrel, ECLIPSE and MATLAB software. His critical remarks and suggestions have always been very helpful in improving my reservoir engineering skills. Special thanks to Shiv who offered tremendous help at the beginning of the research. I also want to thank Ali, Ankit, Min, Mingxiang, Yessica, Qing and Yanming, who are my friends and have always been helpful.

I would like to express my love and gratitude to my family back home: my parents, my elder sister and Yanbo for their unconditional love and help. Without them I would never have made it. I would also like to extend my thanks to

my friends Kaiqi, Jindong, Bo, Pu, Zhongguo, Cui, Di, Jingwen, Xinming and Weitian for constant help.

The financial support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and software support by Schlumberger and CMG are also greatly appreciated.

Table of Contents

List of Tables

List of Figures

List of Abbreviations, Symbols and Nomenclature

ANN	*Artifcial Neural Networks*
AR	*Autoregressive*
ARX	*Autoregressive with Exogenous Input*
ARMAX	*Autoregressive Moving Average with Exogenous Input*
BJ	*Box-Jenkins*
DFN	*Discrete Fracture Network*
EOR	*Enhanced Oil Recovery*
MISO	*Multiple Input Single Output*
MPC	*Model Predictive Controller*
NARX	*Nonlinear ARX*
NPV	*Net Present Value*
OE	*Output Error*

PRBS	*Pseudo Random Binary Sequence*
RTO	*Real Time Optimization*
SAGD	*Steam Assisted Gravity Drainage*
SISO	*Single Input Single Output*
SOR	*Steam Oil Ratio*
SRM	*Surrogate Reservoir Model*
UTF	*Underground Test Facility*
a	Output Coefficients
A	Polynomial Filter
b	Input Coefficients
B	Polynomial Filter
e	Noise
z	Shift Operator in the Frequency Domain
k	Time Instant
n	Model Order
$\mathbf{u}$	Input Vector
x	State Vector
$\mathbf{y}$	Output Vector

q_w	Steam Injection Rate
p_{wf}	Bottom Hole Pressure of Producer
q_{wp}	Water Production Rate
$\hat{y}$	Simulated Output
$\bar{y}$	Mean of Measured Output
$\hat{\theta}(t)$	Parameter Estimated at Time t
$K(t)$	Prediction Error Multiplier
$\Psi(t)$	Gradient of Model Output
λ	Forgetting Factor
τ	Memory Horizon

Chapter 1

Introduction

The application of Steam Assisted Gravity Drainage (SAGD) is of great importance because of the vast reserves accessible with this production mechanism. The Athabasca Oil Sands deposit (also known as the McMurray formation) is located in northern Alberta, spanning 40,000 square kilometers and containing 140 billion cubic meters of original bitumen-in-place. However, only 10% is economically recoverable by surface mining techniques currently used. SAGD can provide significantly greater production rates, high recoveries, and a lower steam-oil ratio (SOR), as compared to conventional surface mining extraction techniques and other thermal recovery methods. An estimated additional 25% of the Athabasca Oil Sands is economically recoverable using SAGD extraction.

To better understand the SAGD process and perform tasks such as assisted history matching, production optimization and forecasting, there has been a lot of research about reservoir modeling. In recent years, the development of

commercial simulators has blossomed, which opened a new frontier for reservoir modeling. However, the major limiting factor in reservoir modeling or optimization is computing ability. Reservoir engineers are always finding ways to reduce the simulation time. Nonetheless the simulation time, according to Van Essen et al. (2012):

1. Reservoir simulation models usually fail to represent near-well bore reservoir dynamics such as gas or water coning.

2. Random operational activities such as breakdown maintenance or well interventions cannot be taken account into the reservoir simulator.

Therefore, the proxy models of the reservoir are gaining more and more attention these days. Proxy models are widely used in different areas of numerical modeling such as sensitivity analysis of uncertainty variables, probabilistic forecasting and reservoir management and optimization.

Jalali et al. (2010) proposed a technique that uses Artificial Neural Networks (ANN) in order to build a Surrogate Reservoir Model (SRM) to perform an uncertainty analysis on a coalbed methane reservoir. An SRM is a replica of the full-field reservoir model that mimics the behavior of the reservoir. A small number of realizations of the reservoir are required to develop the SRM, which will reduce the simulation time significantly. Yeten et al. (2005) performed a traditional one-parameter-at-a-time approach for linear sensitivity analysis. Slotte et al. (2008) used polynomials and multi-dimension kriging

to construct a proxy model for assisted history matching and uncertainty assessment. Venegas et al. (2008) built a proxy model based on Butlers theory to perform an uncertainty assessment of SAGD performance. Fedutenko et al. (2012) proposed a method for predicting production performance during the SAGD process using a data-driven proxy model. Azad et al. (2013) recommended an analytical proxy model to mimic the essential feature of a SAGD field for history matching. Ghasemi et al. (2013) described an alternative proxy approach to model SAGD with an isothermal black-oil reservoir simulator to reduce the computing time.

Proxy models are also used in probabilistic forecasting as an input to the Monte Carlo sampling process (Fishman, 1996). Proxy models exhaustive sampling rates can be achieved due to high computational efficiency.

Reservoir management and optimization is another field where proxy model is very useful. Reservoir management is the recurring process in which an oilfield operator uses mathematical modeling, data and expertise to optimize some stated objective about oilfield performance, such as net present value (NPV). Mohajer et al. (2010) applied the proxy reservoir model in real time optimization (RTO) to shortcut a complex numerical model because the proxy model can provide a sufficiently accurate output value that describes the system in a very short time. However, the restricted frequency of optimization in this approach results in the shortcomings, which include the limited utiliza-

tion of the dynamic degrees of freedom in the plant as well as the discrepancy of the steady-state prediction and the dynamic behavior of the plant. To overcome the shortcomings of the steady-state approach, RTO and the model predictive controller (MPC) are combined into a single entity, which is also called dynamic real time optimization (DRTO). Control technology has been of particular interest for optimizing reservoir performance and displacement efficiency. Ramirez (1987), Sudaryanto (1999) and Brouwer et al. (2001, 2004) have proposed optimal control-theory strategies (Lee et al., 1967) for enhancing oil recovery (EOR) in projects. Nowadays MPC is widely used in the control process and a data-driven proxy model is needed in the MPC algorithm. The linear MPC has been becoming more and more popular due to the theoretical results provided by the academic community and the successful installations in the industry (Borgesen, 2009). Nikolaou et al. (2001) performed an appropriate control action through online optimization of a cost objective function or production rate over a future horizon using a proxy model together with MPC with respect to various constraints. Another advantage of the MPC method, according to Genceli and Nikolaou (1996), Eker and Nikolaou (1999), Schwarm et al (1998) or Schwarm and Nikolaou (1999), is that it can perform a robust identification and control of linear and nonlinear systems at the same time. Saputelli et al. (2005) presented an industrial automation framework for control and optimization of hydrocarbon-producing fields while satisfying business and physical constraints. Awasthi et al. (2007) built a proxy model to simulate the reservoir dynamics, then applied it to the control process to

optimize the production operation. Ljung (1987) has presented a system identification technique which creates a model solely based on measured data. The quality of the system identification model determines the results in optimization when using the MPC method.

1.1 Problem Definition

Nowadays, reservoir management plays an significant role in the oil and gas industry. Historically, reservoir management used to be identified with production engineering, and as time passed, it shared the same meaning with reservoir simulation. Traditional reservoir simulation is time consuming and requires copious CPU ability. What's more, in order to reflect the real reservoir dynamics through the simulation, history matching, which makes the process more CPU-demanded, must be performed. In life-cycle optimization process, short-term optimizations have recently attracted more attention (Van Essen, 2011). Short-term optimizations need a dynamic model to capture the fast dynamics of the reservoir. To decrease the simulation time and make better predictions for the short-term time period, a simple reservoir model should be developed and studied. System identification uses statistical methods to build mathematical models of dynamical systems from measured data. It has been widely used in the control process (Czop and Wszotek, 2010). The characteristics of system identification make it perfect to build a simple proxy model of the reservoirs. Several studies about production optimization using a sys-

tem identification model have been presented. However, none of these studies elucidate how to build a promising system identification model itself.

1.2 Objective

The principal of this thesis is to develop a methodology using system identification to build proxy model of a SAGD reservoir. This work investigates not only a short-term prediction of the production rate, but also a long-term prediction using system identification, which is usually generated by a physics-based model. To apply system identification to various kinds of reservoirs, several cases studies, including cases involving a naturally fractured gas reservoir and a channel reservoir, are compared to a real reservoir model together with commercial simulators.

1.3 Outline

This thesis consists of 8 chapters. Chapter 1 gives a brief overview of the SAGD process and also states the challenges with it. This chapter also explains the aim of the research and the methodology adopted to address the problem statement. Chapter 2 is devoted to the introduction of system identification and model predictive control. Chapter 3 proposes the methodology for developing a proxy reservoir model. Chapter 4 is comprised of implementation details of system identification carried out on real SAGD reservoir data. The

aim of this chapter is to demonstrate how to construct a dynamic model based on pure data. Chapter 5 discusses the possibility of using system identification to predict long-term production. Chapter 6 presents some method to improve the performance of the multiple-step prediction of system identification. Chapter 7 compares the system identification results combining different kinds of reservoirs and input signals. Chapter 8 states the summary and conclusion of the conducted study and discusses possible modifications of the proposed algorithm.

Chapter 2

System Identification and Model Predictive Control

Model predictive control (MPC) has gained significant attention in the control of dynamic systems. It is an advanced method of process control that has been been used in process industries in chemical plants and oil refineries since the 1980s. In recent years more and more reservoir management cases have been implementing the MPC method. Model predictive controllers rely on dynamic models of the process, most often linear empirical models obtained by system identification. System identification uses statistical methods to build mathematical models of dynamical systems from measured data. As a result, not only should model predictive control be investigated, much attention should be paid to system identification.

2.1 System Identification

Although system identification is considered a still-maturing subject, the first International Federation of Automatic Control symposium on system identification was held as long ago as 1967 and some classic textbooks and papers on this topic were published during the 1970s and 1980s. Engineering problems can be broadly categorized as direct or inverse problems. To clarify this distinction, consider the system illustrated in Figure 2.1. If the equations describing the system in the figure are known, the direct problem consists of finding the response to a specific input or excitation: for example, using Newton's second law to calculate the object's acceleration. However, in most cases, the input and the response to the particular input are known, but the equations describing the process are unknown, which gives us a more difficult inverse problem. System identification aims at constructing mathematical models from prior

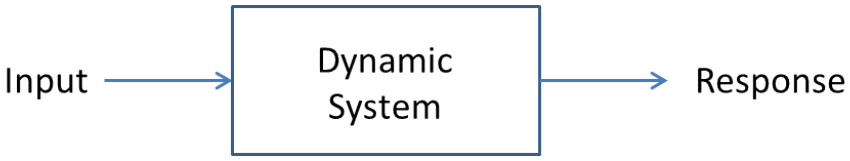


Figure 2.1: Engineering problem overview

knowledge of the system under study and noisy time series data. Therefore system identification is a typical inverse problem. For example, the SAGD process can be considered as a system identification problem, since we can take the well bottom hole pressure as the known input and the oil production rate as the known output, but the system, the reservoir itself, is unknown.

In a dynamic system, the values of output signals are decided by both the instantaneous values of the input signals and by the past behavior of the system. For example, a car seat is a dynamic system - the seat shape (settling position) depends on both the current weight of the passenger (instantaneous value) and how long this passenger has been riding in the car (past behavior). A model is a mathematical relationship between a system's input and output variables. Models of dynamic systems are typically described by differential or difference equations, transfer functions, state-space equations, and pole-zero-gain models. In system identification we need to specify the model structure, then apply the data to the model to obtain the unknown parameters in the model. There are various kinds of model structures and obviously we could not use all of them. Next we will discuss the widely used model structures, such as the common linear model, state-space model and nonlinear model, in system identification.

1. Common linear model structures

A transfer function model is parameterized with $G(z^{-1})$, input-to-output, and $H(z^{-1})$, which are rational functions of the operator z^{-1} of the form:

$$G(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})F(z^{-1})} \tag{2.1.1}$$

$$H(z^{-1}) = \frac{C(z^{-1})}{A(z^{-1})D(z^{-1})} \tag{2.1.2}$$

z is the counter part of time domain forward shift operator q, i.e.,

$q^{-1}x(k) = x(k-1)$. And the overall model structure is shown as follows:

$$y(k) = G(z^{-1})u(k) + H(z^{-1})e(k) \tag{2.1.3}$$

Where

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{nA} z^{-nA} \\ B(z^{-1}) &= 1 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_{nB} z^{-nB} \\ C(z^{-1}) &= 1 + c_1 z^{-1} + c_2 z^{-2} + \ldots + c_{nC} z^{-nC} \\ D(z^{-1}) &= 1 + d_1 z^{-1} + d_2 z^{-2} + \ldots + d_{nD} z^{-nD} \\ F(z^{-1}) &= 1 + f_1 z^{-1} + f_2 z^{-2} + \ldots + f_{nF} z^{-nF} \end{aligned} \tag{2.1.4}$$

are polynomials used for model estimation. y stands for output, u stands for input and e stands for noise. k here denotes the time step. Special cases are:

$$\text{AR}(nD),$$

$$\text{ARX}(nA, nB, k),$$

$$\text{ARMAX}(nA, nB, nC, k),$$

$$\text{BJ}(nB, nC, nD, nF, k),$$

$$\text{OE}(nB, nF, k).$$

In the bracket, A, B, C, D and F are polynomials that are widely used in linear models. n stands for the order of the polynomials. k is the delay of the input, which is the number of samples before the input affects the output of the system.

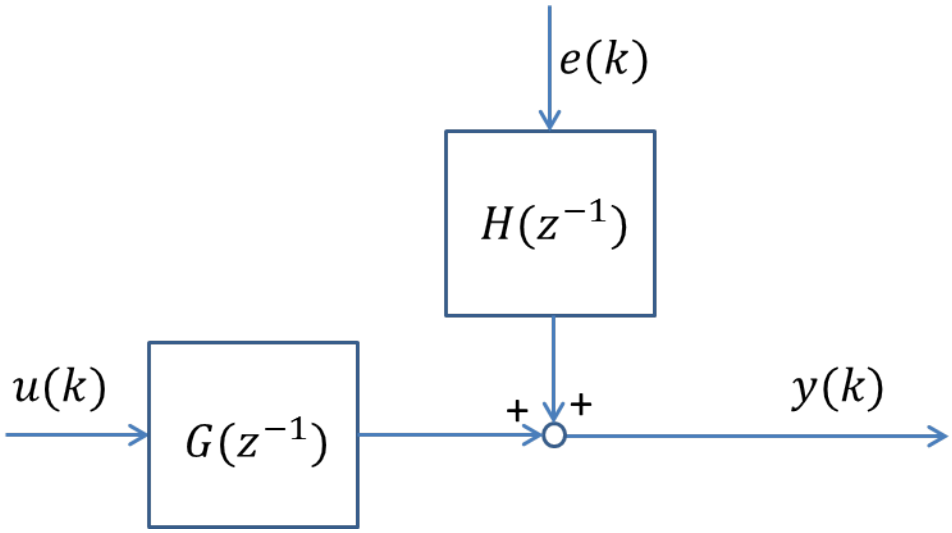


Figure 2.2: Overall linear model structure

The linear model structures are shown in Figure 2.2.

2. State space model structure

Instead of the input/output model, the state space model can also be considered. The state space model structure can be described using the following formula:

$$x(k+1) = Ax(k) + Bu(k+1) + Ke(k) \tag{2.1.5}$$

$$y(k) = Cx(k) + Du(k) + e(k) \tag{2.1.6}$$

Where A,B,C,D are the system matrices, K is the noise matrix, x is the state vector and y,u,e,k have the same meaning as in the previous section.

3. Nonlinear model structure

For the sake of simplicity, the following equations are formulated for

single input single output (SISO) systems. The optimal predictor of an m^{th} order ARX model is:

$$y(k) = b_1u(k-1)+\ldots+b_mu(k-m)-a_1y(k-1)-\ldots-a_my(k-m) \quad (2.1.7)$$

This can be extended to a nonlinear ARX (NARX) model in a straightforward manner by replacing the linear relationship with some unknown nonlinear function $f(\cdot)$, that is,

$$y(k) = f(u(k-1),\ldots,u(k-m),y(k-1),\ldots,y(k-m)) \quad (2.1.8)$$

The nonlinear ARX model structure is depicted in Figure 2.3.

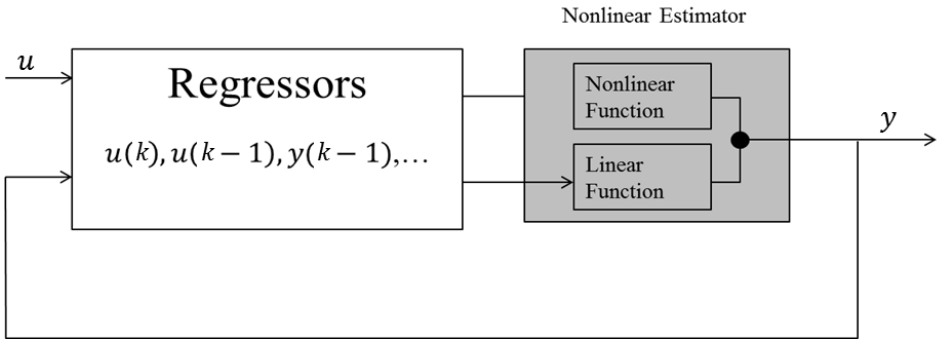


Figure 2.3: Nonlinear ARX model structure

2.2 Model Predictive Control

The MPC can be summarized as follows (Camacho and Bordons, 2004), (Maciejovski, 2002), (Rossiter, 2003):

1. Predict the future behavior of the process state/output over the finite time horizon.

2. Compute the future input signals on line at each step by minimizing a cost function under inequality constraints on the manipulated (control) and/or controlled variables.

3. Apply on the controlled plant only the first of vector control variable and repeat the previous step with new measured input/state/output variables.

Take the SAGD process, for instance. As usual, the life-cycle optimization aims at maximizing NPV defined as an economic objective function J, which, for our example, becomes:

$$J = \sum_{K=1}^{K} \left[\frac{\sum_{i=1}^{n_1} r_{wi}(q_{wi,i})_k + \sum_{i=1}^{n_2} [r_{wp}(q_{wp,j}) + r_o(q_{o,j})_k]}{(1+h)^{\frac{t_k}{\tau_t}}} \triangle t_k \right] \tag{2.2.1}$$

where $k = 1, 2, ..., K$ is are the time steps in the model, n_1 is the number of injection wells, n_2 is the production wells, $q_{wi,i}$ is the water injection rate of well i, $q_{wp,j}$ is the water production rate of well j, $q_{o,j}$ is the oil production rate of well j, r_{wj} is the water injection cost, r_{wp} is the water production cost, r_o is the oil revenue, $\triangle t_k$ is the time interval of time step k, and h is the discount rate for reference time τ_t. During the optimization procedure, the input vectors $\mathbf{u}$ are iteratively changed until a maximum value of J is obtained. For most of the time, the optimal input vector $\hat{\mathbf{u}}$ is the desired life-cycle optimization procedure. In the MPC algorithm, we will choose the corresponding optimal output $\hat{\mathbf{y}}$ as the reference value. The controller now attempts to find the corrected inputs $\bar{\mathbf{u}}$ such that the difference between the actual measured outputs $\mathbf{y}$ and the optimal outputs $\hat{\mathbf{y}}$ is minimal. This is

achieved by minimizing the following objective function over a relatively short time horizon:

$$Q = \min \left[\sum_{i=1}^{N} (y_{k+i} - y_{k+i}^{sp})^T W_y (y_{k+i} - y_{k+i}^{sp}) + \sum_{i=0}^{M-1} (u_{k+i} - u_{k+i-1}^{sp})^T W_u (u_{k+i} - u_{k+i-1}^{sp}) \right] \tag{2.2.2}$$

$$u_{min} \leq u_{k+i} \leq u_{max} \tag{2.2.3}$$

$$y_{min} \leq y_{k+i} \leq y_{max} \tag{2.2.4}$$

where N is the predict horizon, y_{k+i}^{sp} is the output targets, y_{k+i} is the measured outputs, u_{k+i} is the i step ahead inputs and W_y and W_u are the weight coefficients on output and input deviations respectively. Only the first step of the control strategy is implemented, then the plant state is sampled again and the calculations are repeated starting from the now current state, yielding a new control and new predicted state path. The basic MPC structure is illustrated in Figure 2.4. The plant in the figure stands for the system identification data-driven model. The setpoint is the optimal output vector $\hat{\mathbf{y}}$ we want to obtain. At each time step, an input vector $\mathbf{u}$ is inserted until Q is minimal.

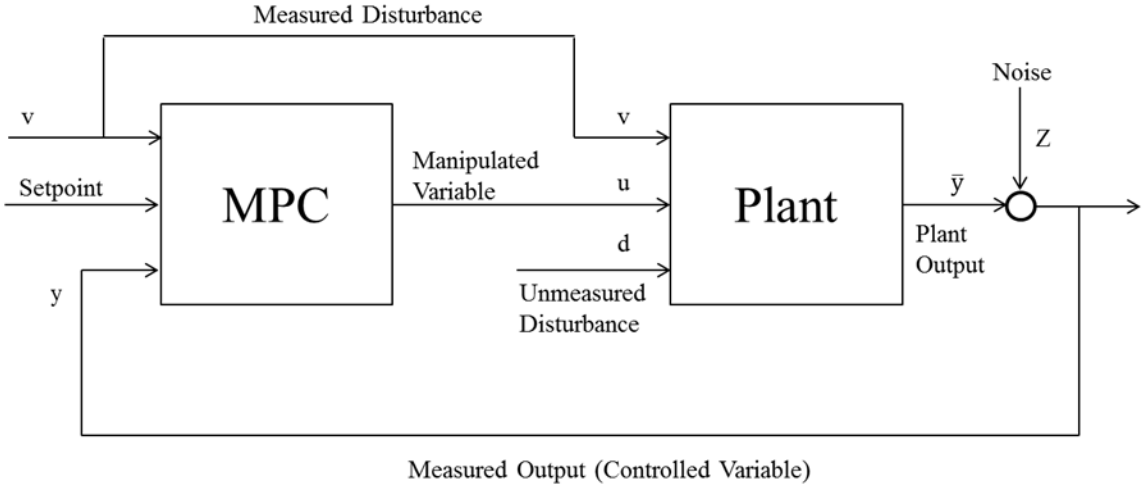


Figure 2.4: Model predictive controller structure

Chapter 3

Workflow of Building Proxy Model Using System Identification

Different from the physics-based reservoir model, the quality of a proxy model built with system identification is determined by the underlying algorithm and dataset used to create the model. It is obvious that we cannot obtain a dataset of infinite size to capture all reservoir dynamics. The different input variables and model structures can lead to completely different simulation results. In order to make precise predictions on the production profiles, various combinations of datasets, input variables and model structures should be tested. The good model should not only capture the main reservoir dynamics but also be simple to build considering the computing efficiency. Once the validated model

is built, it can be used in short-term or long-term production prediction. With input design and the CMG flexwell model, the feasibility of the proxy model can be tested. A typical workflow of steps of proxy-modeling is depicted in Figure 3.1.

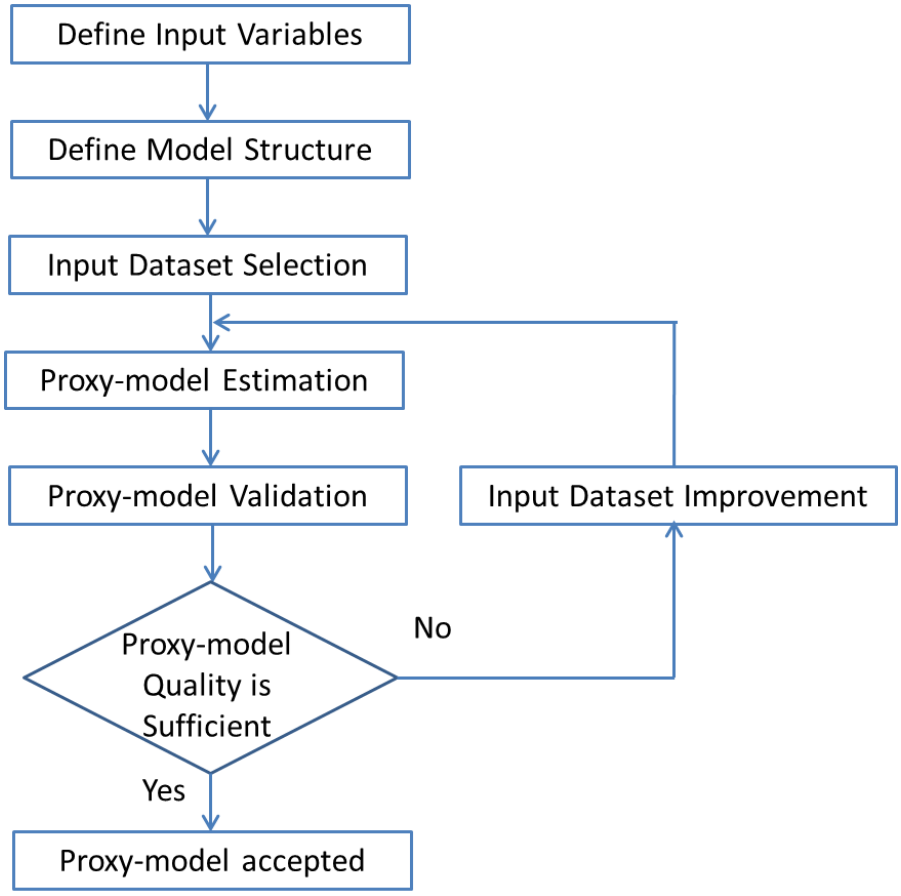


Figure 3.1: Proxy-modeling workflow

1. **Input variables**

   In proxy reservoir modeling using the system identification method, only a few variables are needed to describe the reservoir dynamics. There is no restriction to determine which variables should be used. However, it

is always better to include the variables that people are interested in or variables which will be applied to the control process in the future. Also, the manipulated variables should be considered first.

2. Model structure

The model structures are very significant in the system identification process. There are various kinds of model structures which could be applied to the reservoir: for example, the autoregressive (AR) model and output error (OE) model. A good model should capture the reservoir dynamics and be explicit. Based on the performances of different model structures, the model structures that can predict the production data most precisely will be chosen to predict the production profiles.

3. Input data selection

Considering that the data collected by the sensors is usually of great quantity, it is impossible to build the proxy model with all of them. And in real operation, the shut-in well event will give zero injection or zero production. This kind of data can't reflect the reservoir dynamics. As a result, a reasonable dataset must be chosen. There are two problems need to be addressed:

(a) How much data is needed to build a validated model?

(b) When should start building the model?

Taking these two questions into consideration, a specific time period

and time instant of the dataset will be selected to build the final system identification model.

4. **Proxy model estimation and validation**

   Once the dataset is fixed, it will be divided into two equal parts. The first part will be used for model estimation and the second part for model validation. A good model should not only capture the reservoir dynamics in the validation part but also fit the trend of the dataset in the estimation part, meaning it can make predictions about the future. However, it should be noted that over-fitting one part will lead to bad performance in the other part; therefore a balance point should always be found and the standard should not be fixed.

5. **Input dataset improvement**

   The initial dataset selection is not always sufficient for construction of a reliable proxy model. If the model validation part reveals a high error in predicting the production data, a new dataset should be selected, one with either new data or starting from another time instant. The new dataset should be used to identify a new model until the proxy model quality is sufficient.

Chapter 4

System Identification Using Raw Field Data

In this chapter we will build proxy reservoir models step by step, using system identification. Except for common linear model structures, we will also test the performance of state space models. The data used here is obtained from 15-year operation of SAGD reservoir located in northern Alberta. All the production rates presented throughout the thesis are subtracted by the linear trends. First, short term prediction will be introduced, considering the various combinations of input/output parameters, model structures and the time when the modeling should start. Then we will try to build models which are able to predict for longer time periods. When building the model, the data comes from the same well pair. After the model is validated, we can apply the same model to another well pair in the same reservoir to test whether the model can reflect the reservoir dynamics or whether it will be significantly in-

fluenced by external factors. Some geological specifications are listed in Table 4.1.

Table 4.1: Reservoir geological parameters

Parameters	Value
Depth from surface	125 to 175 m
Net pay thickness	15 to 20 m
Sand porosity	35%
Sand permeability	5 to 12 Darcy
Bitumen saturation	85%
Bitumen viscosity (70 Celsius)	5×10^6 cp
Weight of bitumen	16%
API	8°

4.1 Input and Output Variables Selection

The first step toward developing a proxy model for a SAGD reservoir is to assign input and output variables. The continuously obtained field data can be used as input variables (L.Saputelli, 2005). Also, in order to apply the model to the control process in the future, some manipulated variables need to be chosen. In this thesis, the data is obtained from a reservoir with three well pairs, and one certain well pair will be considered. According to van Essen

et al. (2012), the input variables could be:

$$\mathbf{u} = [u_1 \quad u_2]^T = [q_w \quad p_{wf}]^T \tag{4.1.1}$$

Where q_w represents the steam injection rate in the injector, and p_{wf} is the bottom hole pressure in the producer. To represent the reservoir dynamics, the water and bitumen production rates will be chosen as output variables. In order to eliminate the cross correlation and increase the accuracy of the model, the water rate and bitumen rate will be simulated separately. That is,

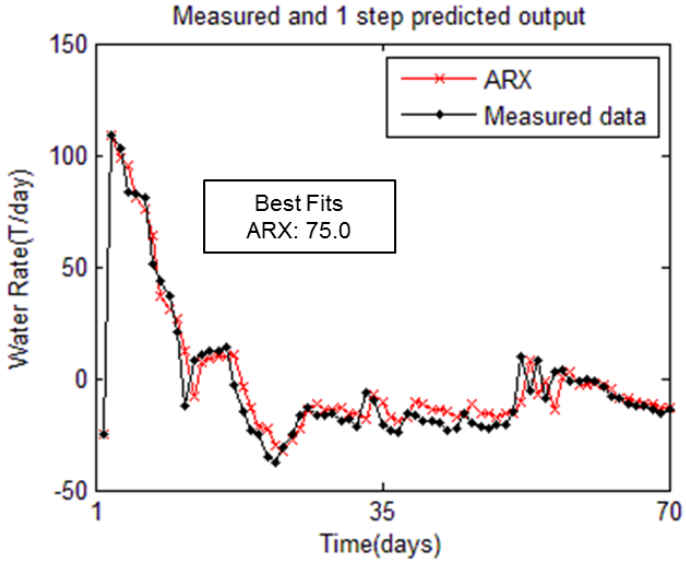
$$y_o = [y_1]^T = [q_o]^T \tag{4.1.2}$$

$$y_w = [y_2]^T = [q_{wp}]^T \tag{4.1.3}$$

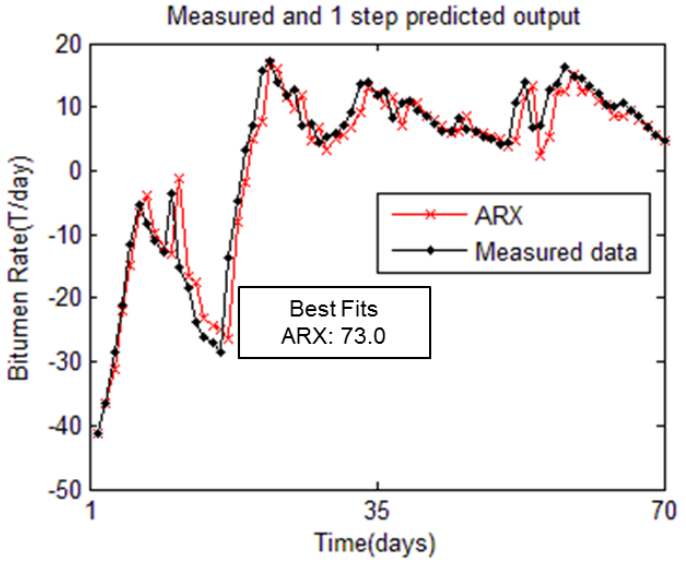
Where q_o is the oil production rate in the producer and q_{wp} is the water production rate in the producer.

The model structure here is the default as the ARX model, since the purpose here is to test the validity of the input and output variables. The result is shown in Figure 4.1 and Figure 4.2. Figure 4.1 shows the model estimation at the beginning, which means the model is built based on the data shown in the figure. Figure 4.2 shows the model validation, which means using the model built in Figure 4.1 to predict the production data. It is obvious that the input and output variables used here can capture the reservoir dynamics. The model predicted production rates are almost the same as the real ones. The reason why the production rates have a negative value is that the dataset

used here is subtracted by the linear trends.

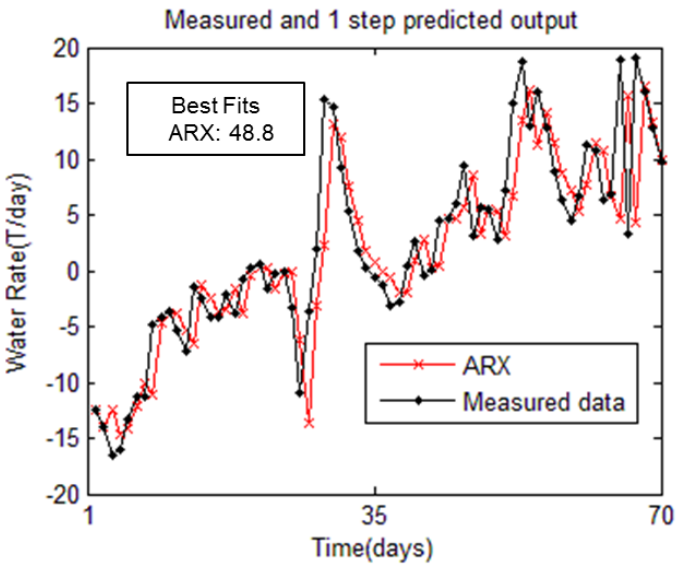


(a) Water production rate

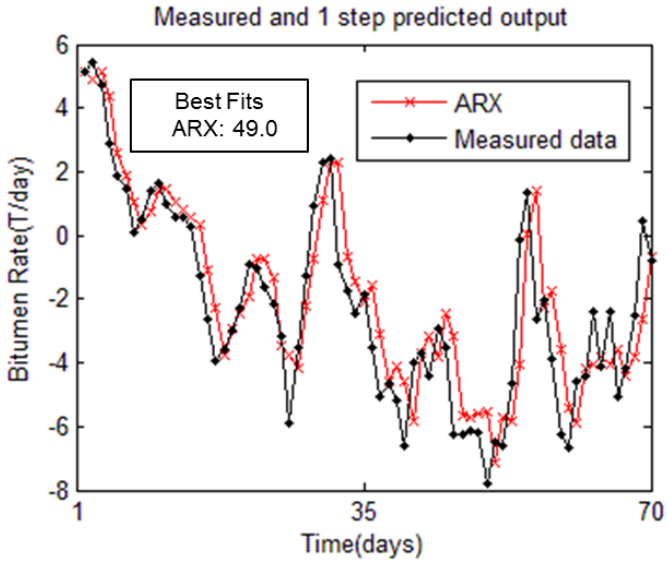


(b) Bitumen production rate

Figure 4.1: Input and output variables' selection: model estimation



(a) Water production rate



(b) Bitumen production rate

Figure 4.2: Input and output variables' selection: model validation

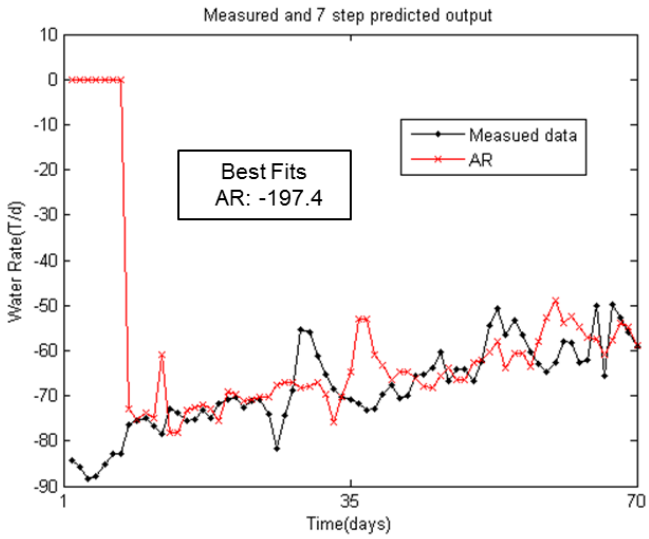
From the figures we notice an index called "best fits," which is an important standard of choosing the best model. Best fits can be described in the following

formula:

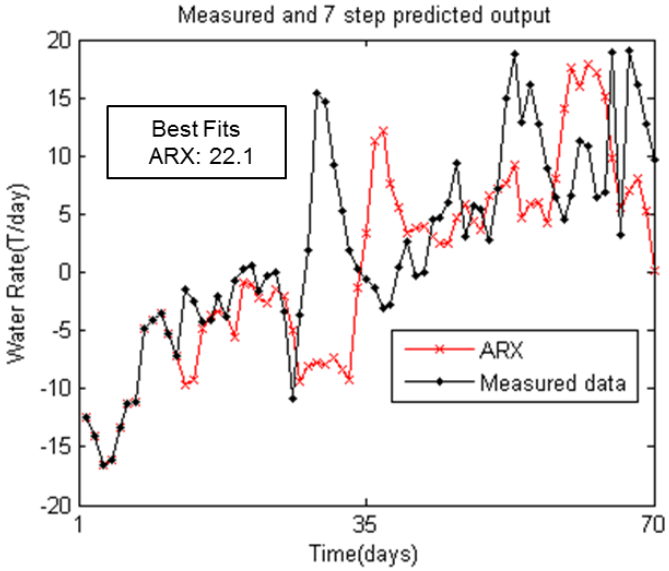
$$\text{Best Fits} = \left(1 - \frac{|y - \hat{y}|}{|y - \bar{y}|}\right) \times 100\% \tag{4.1.4}$$

In Equation 4.1.4, y is the measured output, $\hat{y}$ is the simulated or predicted model output, and $\bar{y}$ is the mean of y. One hundred percent corresponds to a perfect fit, and zero percent indicates that the fit is no better than guessing the output to be a constant($\hat{y} = \bar{y}$). Because of the definition of best fit, it is possible for this value to be negative. A negative value is worse than zero percent.

To further investigate the importance of input, we will build an AR model using the same data. As shown in Chapter 2, the AR model has only D polynomials, which means there is no input value contributing to the model's output. If the model's best fit is acceptable, it means the input signal is not significant for the field data we used. The results are depicted in Figure 4.3. From the comparison, we can see clearly that the ARX model has a better best fit compared to the AR model, which indicates that the proper input selection can surely improve the model's performance. And one thing needed to clarify is these two models are based on seven-step-ahead prediction, which will be introduced in Chapter 6.



(a) AR model without input channel

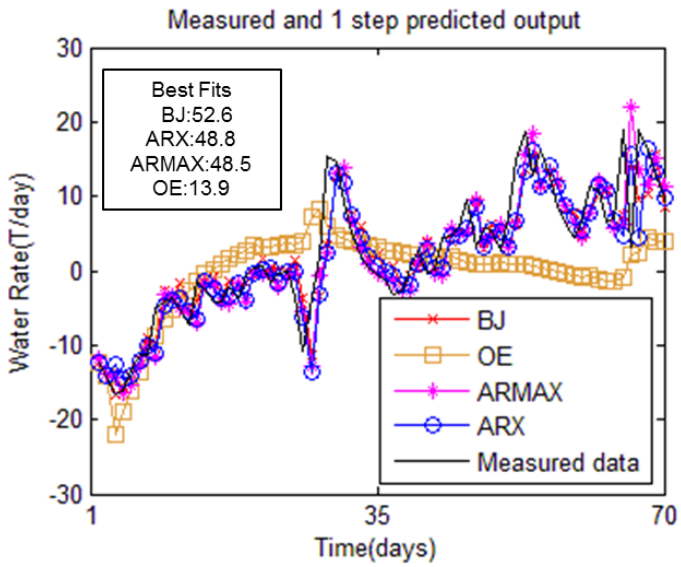


(b) ARX model with input channel

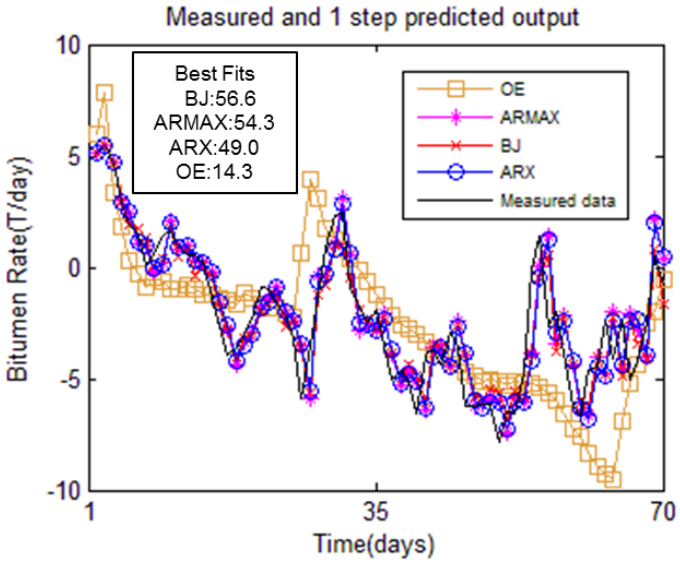
Figure 4.3: The impact of input on the model's quality

4.2 Model Structures Selection

Once the input and output variables are decided, the next step is to find the suitable model structure which can describe the reservoir dynamics best. In general, the good model should be not only accurate, but implicit. TThe model structures used here are the linear models used in Chapter 3. The results are depicted in Figure 4.4, and all of the models except for the OE can capture the reservoir dynamics. The model orders are listed in Table 4.2. Of all the linear models discussed so far, the BJ model is the most general and flexible. It allows one to estimate separate transfer functions with arbitrary numerators and denominators from the input to the output and from the disturbance to the output. However, the flexibility of the BJ model requires one to estimate a large number of parameters. From Table 4.2 and Table 4.3, we can easily note that BJ model has larger orders than the others. As a result, the BJ model yields better best fits. Considering all the models used here are easily identified using the MATLAB system identification toolbox, we will choose the BJ model as the target model structure.



(a) Water production rate



(b) Bitumen production rate

Figure 4.4: Model structure selection

Table 4.2: Water rate model orders

Model Structure	Orders
ARX	$nA = 2, nB = [1 \quad 1], k = [1 \quad 1]$
ARMAX	$nA = 2, nB = [2 \quad 2], nC = 2, k = [1 \quad 1]$
BJ	$nB = [1 \quad 1], nC = 1, nD = 1, nF = [2 \quad 2], k = [1 \quad 1]$
OE	$nB = [3 \quad 1], nF = [1 \quad 1], k = [1 \quad 1]$

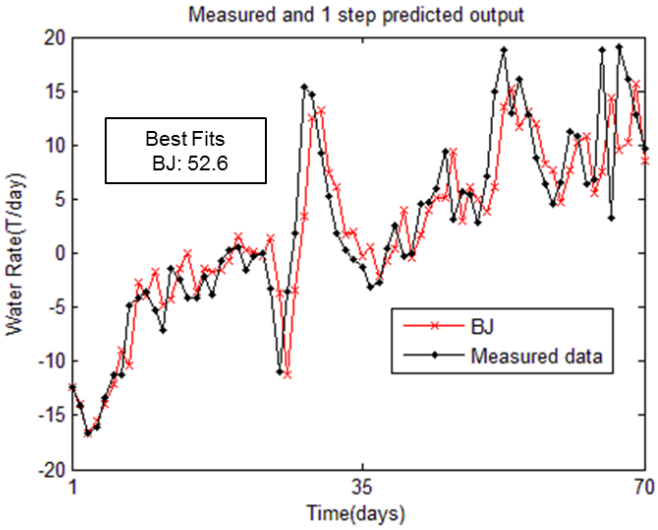
Table 4.3: Bitumen rate model orders

Model Structure	Orders
ARX	$nA = 1, nB = [1 \quad 1], k = [1 \quad 1]$
ARMAX	$nA = 1, nB = [4 \quad 1], nC = 1, k = [3 \quad 1]$
BJ	$nB = [6 \quad 1], nC = 1, nD = 1, nF = [1 \quad 1], k = [3 \quad 3]$
OE	$nB = [2 \quad 2], nF = [2 \quad 2], k = [1 \quad 1]$

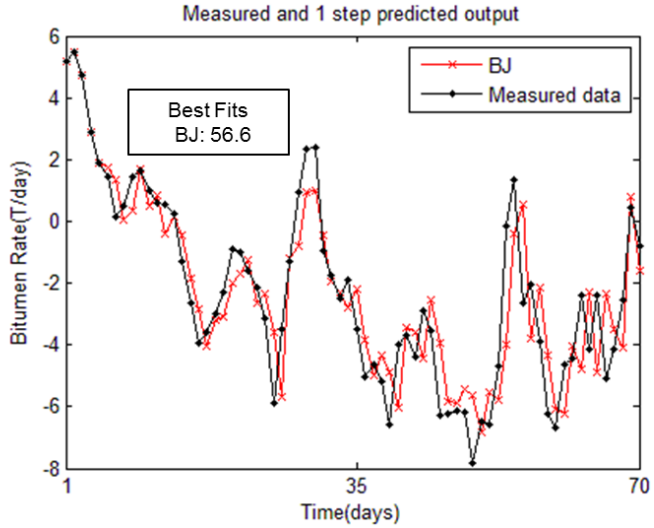
4.3 Model Construction Based on Different Time Periods

As mentioned before, the data used here is 15-year SAGD reservoir production profile based on the characteristics of system identification. Thus it is impossible to include all the data in one model, for example, using the first seven years' worth of data to build the model, then predicting for the next

eight years' production profiles. A relatively short time period, that is, two months, six months and one year of dataset, will be used to build the model. The purpose here is to find out how much data to include to give the best performance. The results can be observed in Figure 4.5 - Figure 4.7. All models built can capture the reservoir dynamics. The models built based on six months and one year of data have even better performances than the one with two months of data. However, it is obvious that when the real production rate is zero (the horizontal line in Figure 4.6 and Figure 4.7, there are big jumps in production profile and around this time period a lot of model errors appear. It is hard to identify this kind of data in system identification.

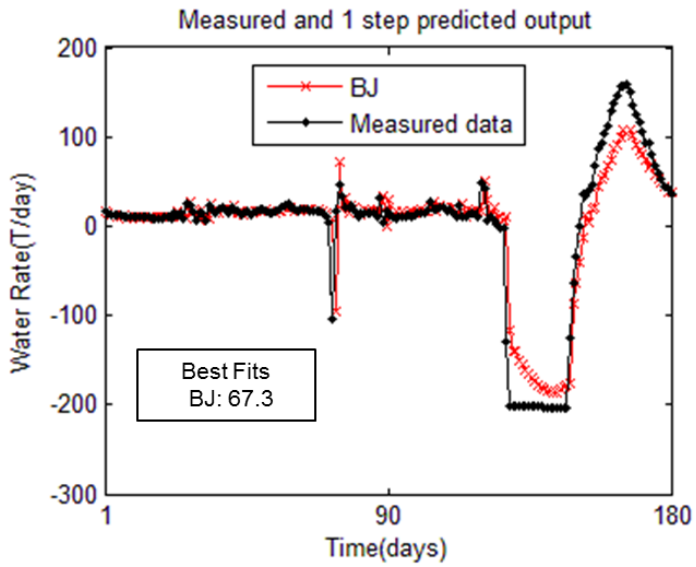


(a) Water production rate

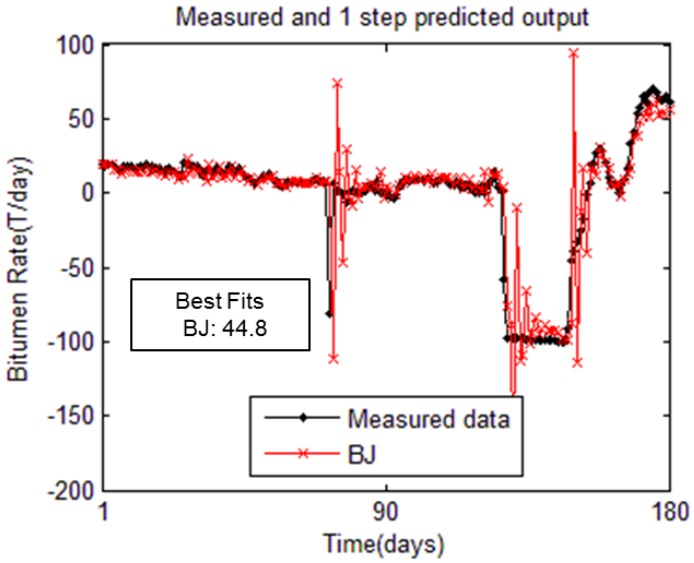


(b) Bitumen production rate

Figure 4.5: Model construction based on two months of data

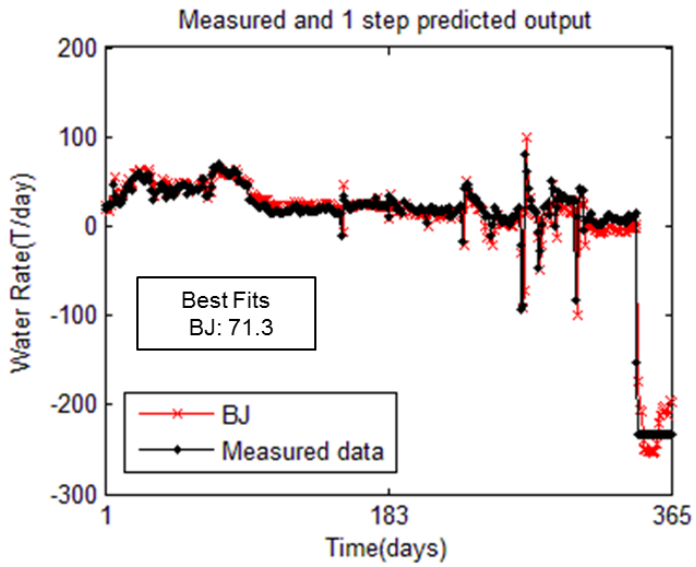


(a) Water production rate

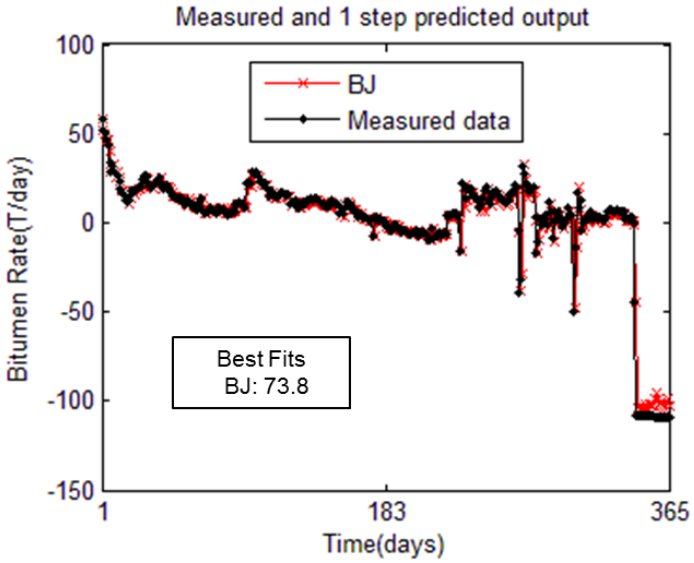


(b) Bitumen production rate

Figure 4.6: Model construction based on six months of data



(a) Water production rate

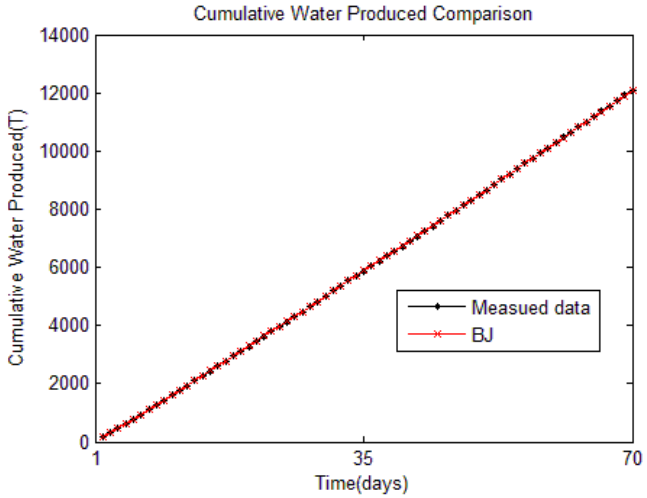


(b) Bitumen production rate

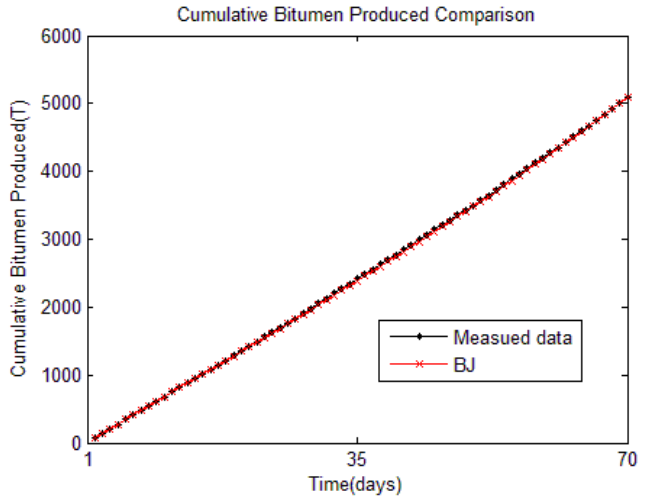
Figure 4.7: Model construction based on one year of data

From the results, it seems that all three models are validated and satisfied. In order to distinguish them, we will bring in another standard for model selection. Figure 4.8 - Figure 4.9 show the cumulative production rate comparison.

From the figures we notice that the models built with two months worth of data and one year worth of data have a better fit in cumulative production comparison, while the fit of model built with 6 month's data has some gaps at the end. Considering the cumulative production is easier to match, the model built with six months worth of data will be ignored.

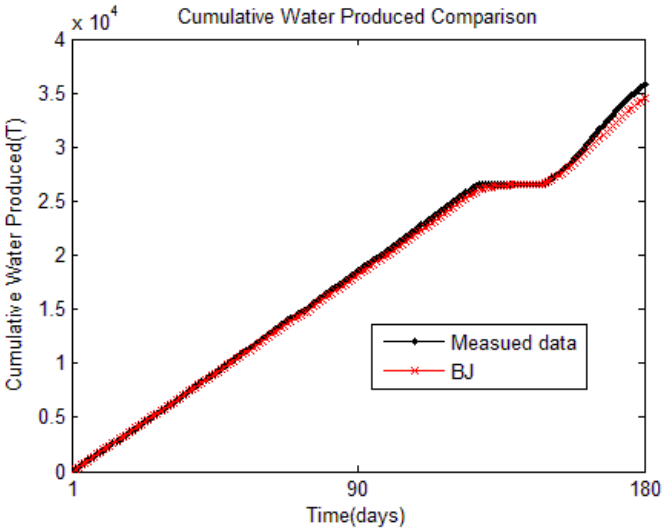


(a) Cumulative water

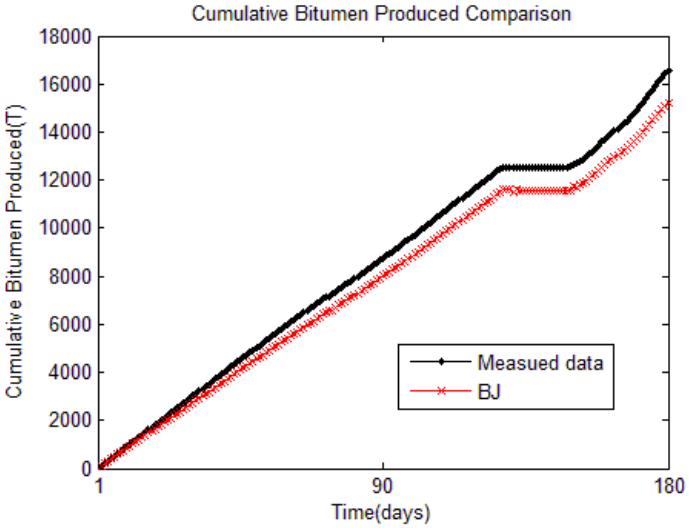


(b) Cumulative bitumen

Figure 4.8: Two months cumulative production comparison

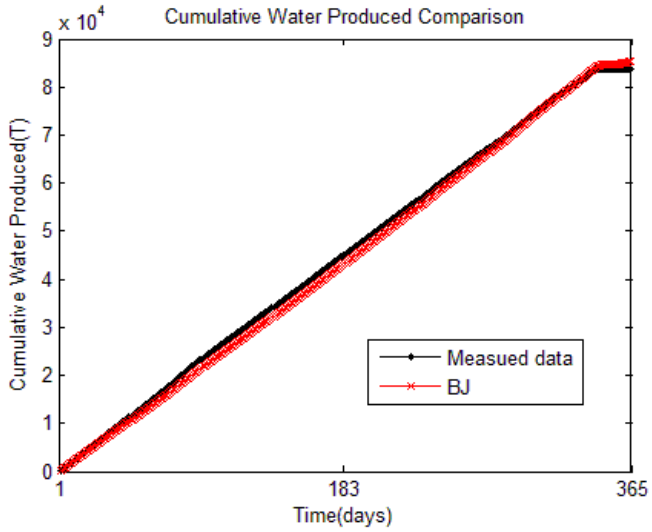


(a) Cumulative water

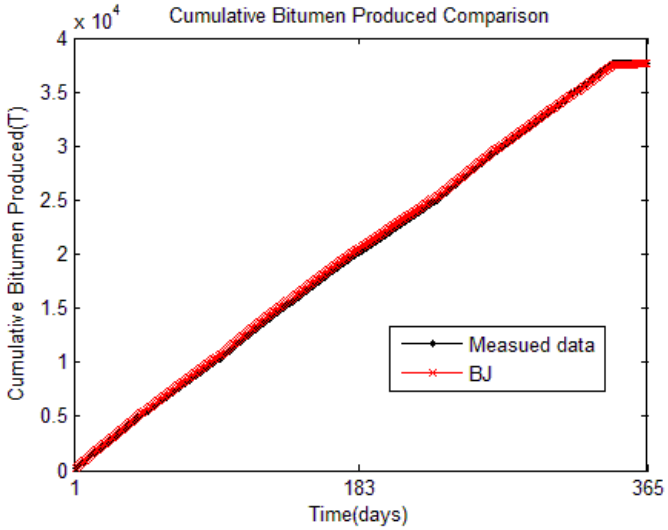


(b) Cumulative bitumen

Figure 4.9: Six months cumulative production comparison



(a) Cumulative water

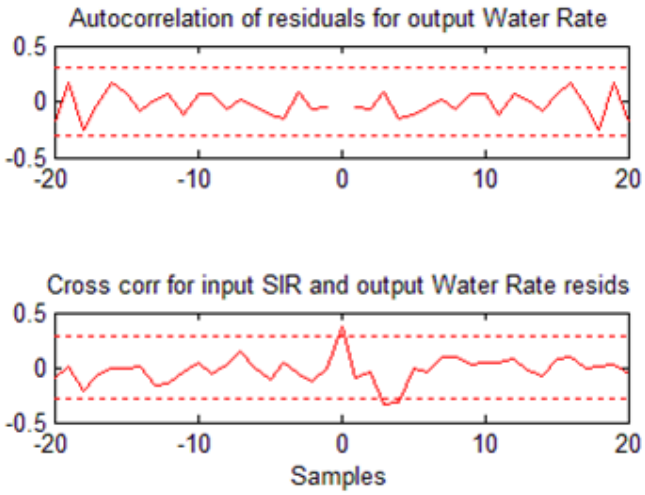


(b) Cumulative bitumen

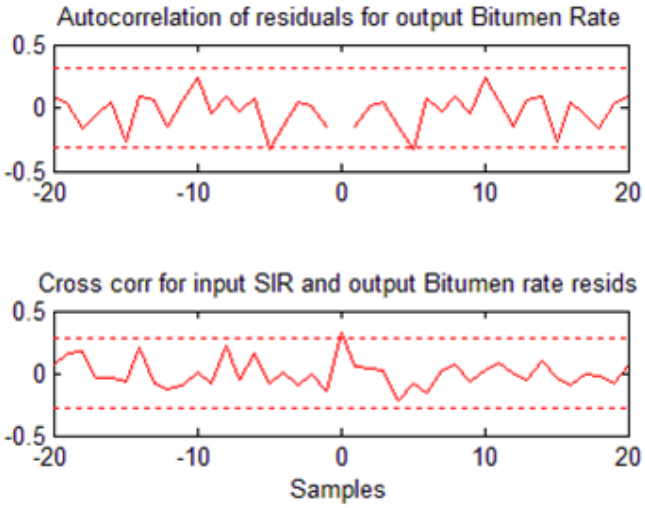
Figure 4.10: One year cumulative production comparison

Another way to test the model's performance is the residual test. The results are depicted in Figure 4.11 - Figure 4.13. The top axes show the autocorrelation of residuals for the output (whiteness test). The horizontal scale is the number of lags, which is the time difference (in samples) between the signals

at which the correlation is estimated. The horizontal dashed lines on the plot represent the confidence interval of the corresponding estimates. Any fluctuations within the confidence interval are considered to be insignificant. A good model should have a residual autocorrelation function within the confidence interval, indicating that the residuals are uncorrelated. The bottom axes show the cross-correlation of the residuals with the input. A good model should have residuals uncorrelated with past inputs (independence test). Evidence of correlation indicates that the model does not describe how the output is formed from the corresponding input. For example, when there is a peak outside the confidence interval for lag k, this means that the contribution to the output $y(t)$ that originates from the input $u(t-k)$ is not properly described by the model. From these three figures it is obvious that only the model with two months of data passed the residuals test. Taking all these factors into consideration, it is better to build the model based on two months worth of data.

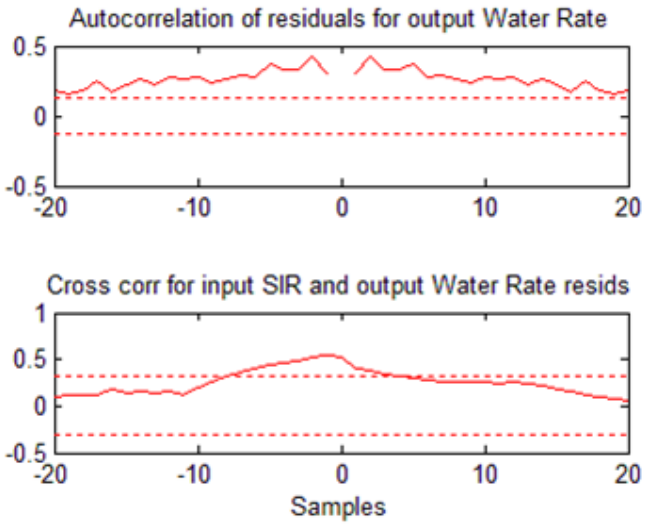


(a) Water channel

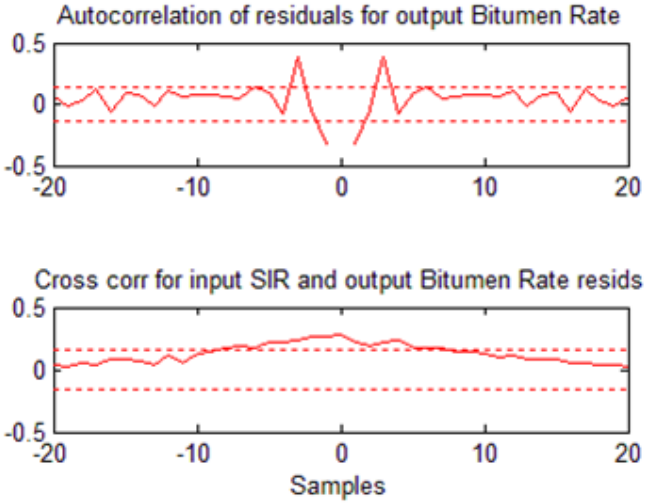


(b) Bitumen channel

Figure 4.11: Model residuals test of two-month model

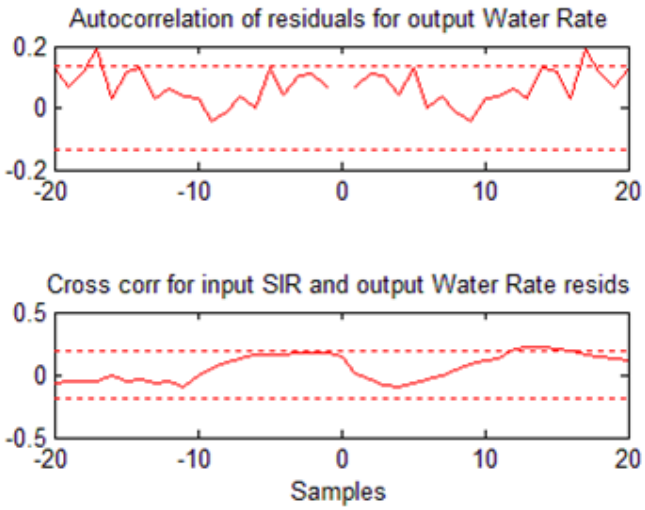


(a) Water channel

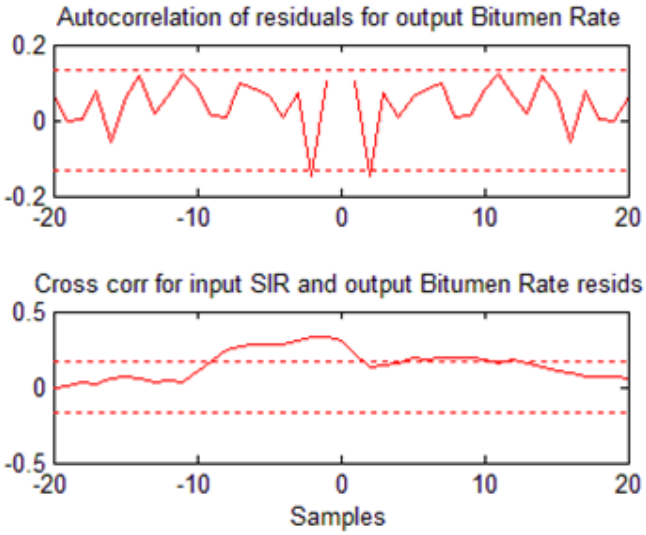


(b) Bitumen channel

Figure 4.12: Model residuals test of six-month model



(a) Water channel

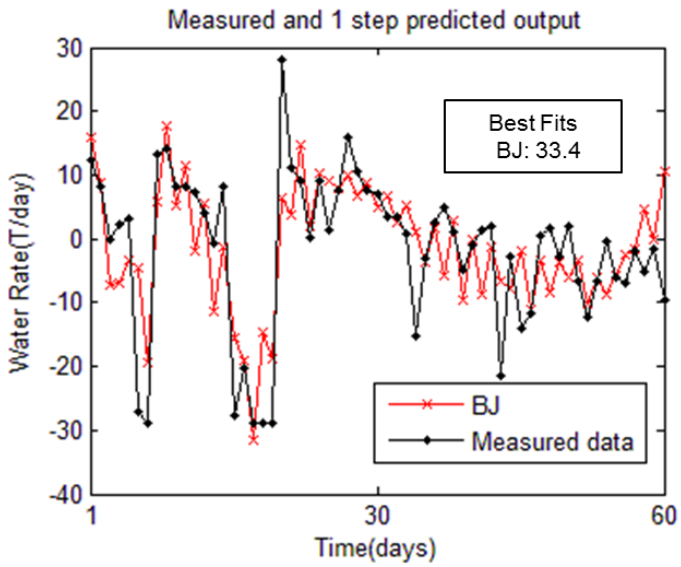


(b) Bitumen channel

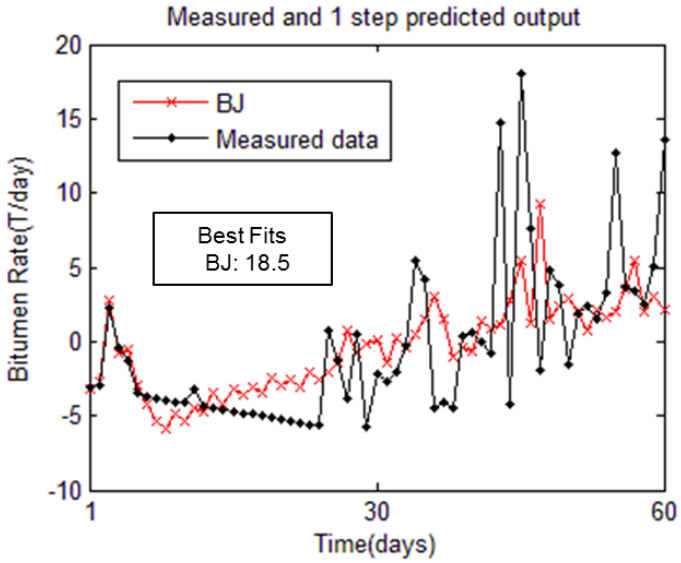
Figure 4.13: Model residuals test of one-year model

4.4 Model Construction Start From Different Time Spots

In real reservoir life, the well may be shut in for many reasons, which will create zero values in production profiles and the production and make it hard to identify the production. In addition, the reservoir dynamics are unstable at the very early and late stages of production. Therefore, in order to avoid the zero value and unstable dynamics, a reasonable time instant must be found to start to build the proxy reservoir model. In this research we chose four typical time periods in which to build the proxy model. The first is the beginning of the reservoir production history, staring from year one. The second is when the reservoir comes to a stable time period, starting from year three. The third is also when the reservoir comes to a stable time period, starting from year eight. The fourth is the end of the reservoir production history, staring from year 13. The results are shown in Figure 4.14 - Figure 4.17. At the beginning and the end of the reservoir production history, the model fits are not good because the reservoir dynamics are not stable at that time. Even when the reservoir comes to a relatively stable time period, the case starting from year eight, the model fit is not good enough to represent reservoir dynamics. Only the one staring from year three gives the best model fit and it captures all trends of the production profile. Taking all these into consideration, the model will be built starting from year three.

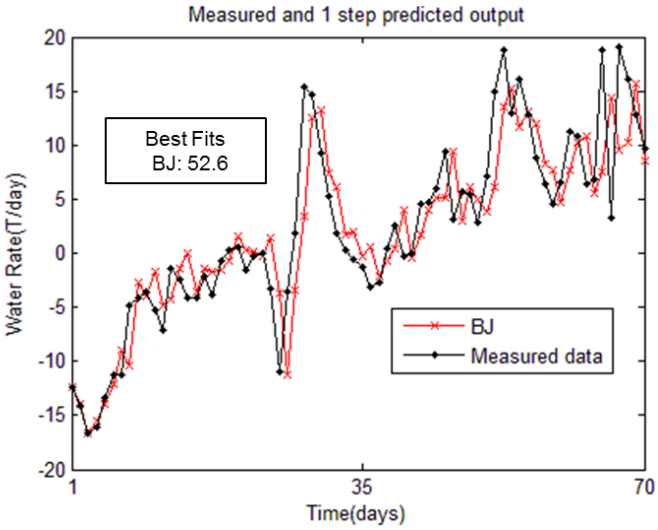


(a) Water production rate

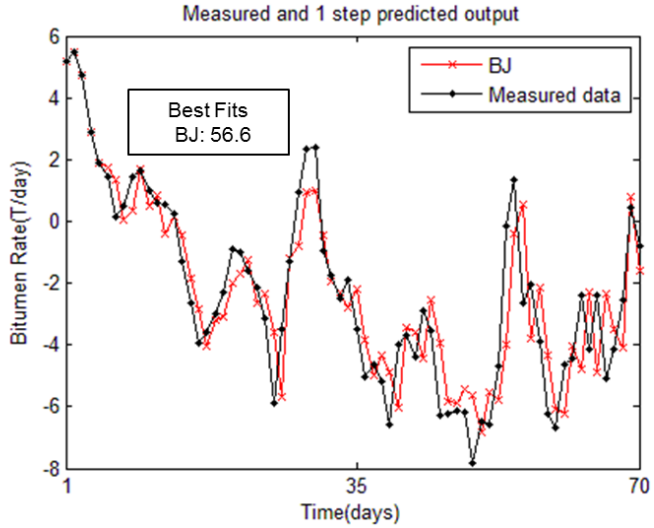


(b) Bitumen production rate

Figure 4.14: Model construction starts from year one

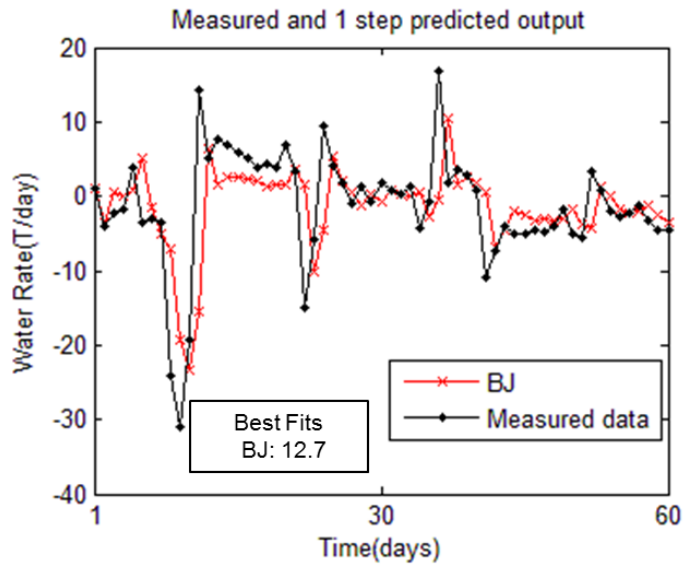


(a) Water production rate

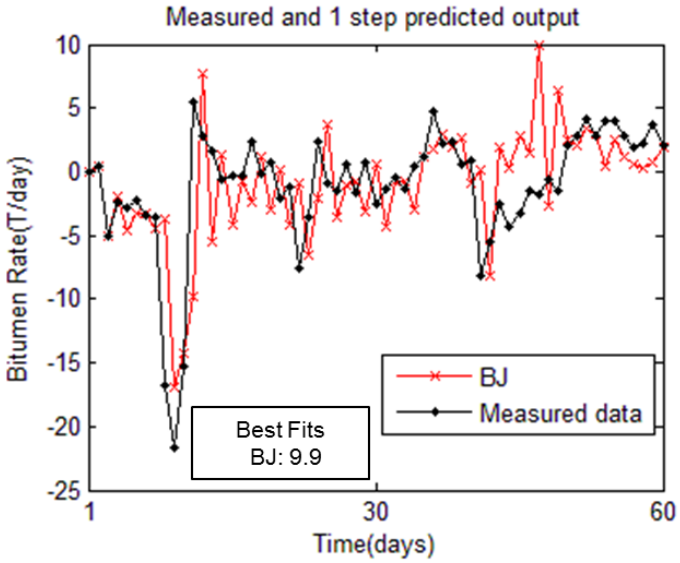


(b) Bitumen production rate

Figure 4.15: Model construction starts from year three

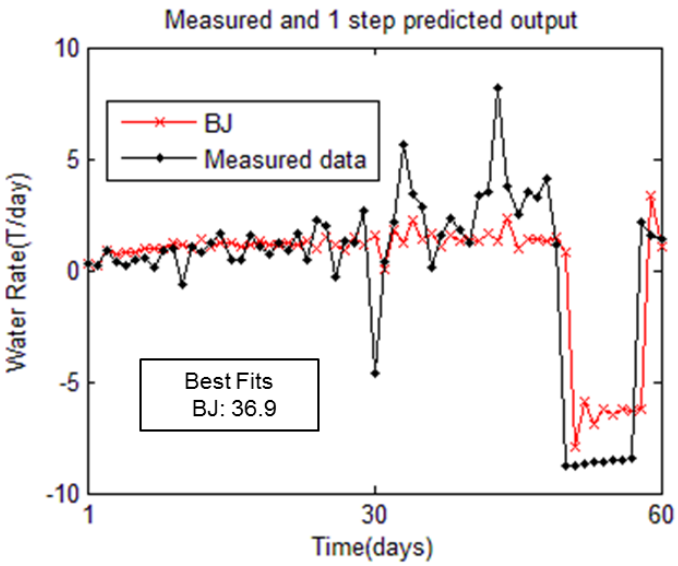


(a) Water production rate

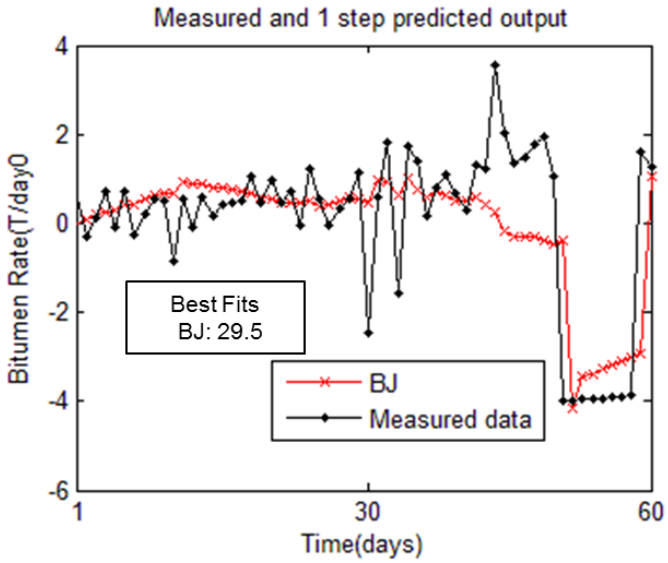


(b) Bitumen production rate

Figure 4.16: Model construction starts from year eight



(a) Water production rate

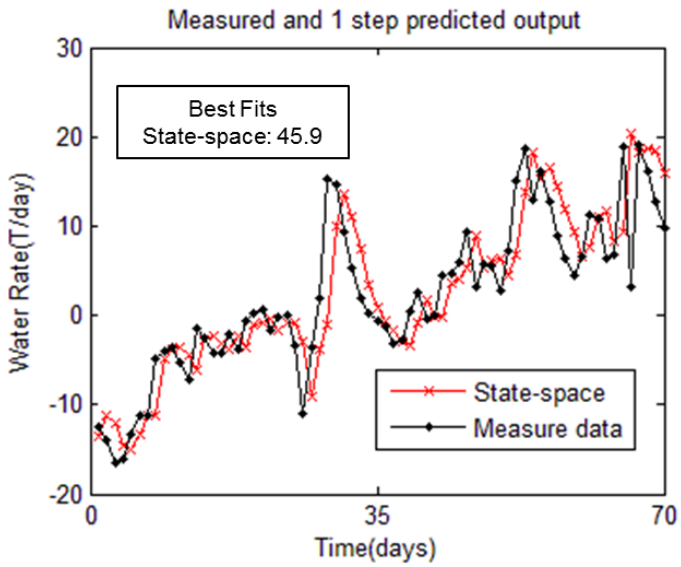


(b) Bitumen production rate

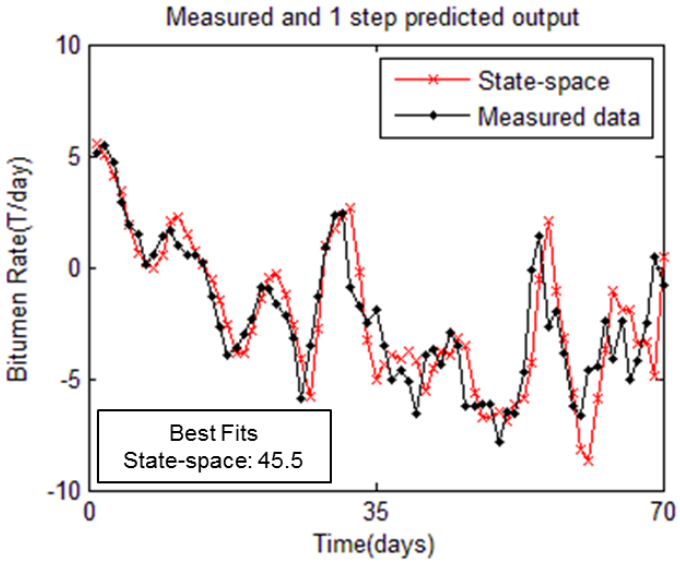
Figure 4.17: Model construction starts from year 13

4.5 Check the Performance of State-space Model Using the Subspace Method

In multivariable dynamical system identification, the subspace-based state-space system identification method is always applied. For this method, the only parameter that needs to be specified is the model order. The major advantage of a state-space representation is that prior knowledge from first principles can be incorporated in the form of equations and can be utilized to prestructure the model. Furthermore, the number of regressors is usually smaller in state space models than in input/output models. For a system of m^{th} order a state-space model possesses $m+1$ regressors $(x_1(k), \ldots, x_m(k) \text{and} u(k))$ while an input/output model requires $2m$ regressors $(u(k-1), \ldots, u(k-m) \text{and} y(k-1), \ldots, y(k-m))$. After choosing the input, output, time period and time spot to start building the model, a state-space model is constructed. Figure 4.18 shows the result. From the figure, it is obvious that the reservoir dynamics are captured, which means the state-space model using subspace method is validated for building the dynamic model.



(a) Water production rate



(b) Bitumen production rate

Figure 4.18: State-space model using subspace method

Chapter 5

Recursive Estimation for Long-term Prediction

Until now, the common system identification method was only used to predict relatively short time periods, because in real-field operations, there are a lot of emergencies that cannot be predicted by reservoir dynamics. In order to make a long-term prediction, a new technique: recursive estimation is introduced (Wellstead and Zarrop, 1996). Many real-world applications, such as adaptive control, adaptive filtering, and adaptive prediction, require that a model of the system be available online while the system is in operation. Therefore, recursive estimation can be used to make online predictions, which will increase the performance of long-term predictions. Every identification model structure has its own coefficient applied to input and output data, such as a and b parameters. These coefficients are correlated with the reservoir condition and fixed once the model is built. For that reason, the model structure is validated

just for the time when the model is built and cannot be used for predicting a longer time horizon. However, recursive estimation can update the coefficient every time step instead of just fixing it, which will give us a new model for every time step corresponding to the current reservoir condition. Therefore, using recursive estimation is a good choice for long-term prediction.

5.1 Recursive Estimation Algorithm

The general recursive identification algorithm is given by the following equation:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)[y(t) - \hat{y}(t-1)] \tag{5.1.1}$$

where $\hat{\theta}(t)$ is the parameter estimate at time t. $y(t)$ is the measured output at time t and $\hat{y}(t)$ is the prediction of output based on observations up to time $t-1$. $K(t)$ is the multiplier that determines how much the current prediction error $y(t) - y(t-1)$ affects the update of the parameter estimation. The estimation algorithms minimize the prediction error $y(t) - \hat{y}(t-1)$. The multiplier usually has the following form:

$$K(t) = Q(t)\Psi(t) \tag{5.1.2}$$

The recursive algorithm supported by the System Identification Toolbox™ differs based on different approaches for choosing the form of $Q(t)$ and computing $\Psi(t)$, where $\Psi(t)$ stands for the gradient of the predicted model output $\hat{y}(t|\theta)$ with respect to the parameters θ.

Considering the model with a linear-regression form:

$$\hat{y}(t) = \Psi^T(t)\theta_0(t) + e(t) \tag{5.1.3}$$

In this equation, $\Psi(t)$ is the regression vector that is computed based on previous values of measured inputs and outputs. $\theta(t)$ represents true parameters. $e(t)$ is the noise term which is assumed to be white noise. The specific form of $\Psi(t)$ depends on the structure of the polynomial model. For linear regression equations, the predicted output is given by the following:

$$\hat{y}(t) = \Psi^T(t)\hat{\theta}(t-1) \tag{5.1.4}$$

There are three main algorithms used for recursive estimation: Kalman filter algorithm, forgetting factor algorithm and unnormalized and normalized gradient algorithm. In the thesis, we will use the forgetting factor algorithm. Equation 5.1.1, 5.1.2 and 5.1.4, together with the following set of equations, summarized the forgetting factor adaptation algorithm:

$$Q(t) = P(t) = \frac{P(t-1)}{\lambda + \Psi(t)^T P(t-1)\Psi(t)} \tag{5.1.5}$$

$$P(t) = \frac{1}{\lambda}\left[P(t-1) - \frac{P(t-1)\Psi(t)\Psi(t)^T P(t-1)}{\lambda + \Psi(t)^T P(t-1)\Psi(t)}\right] \tag{5.1.6}$$

To obtain $Q(t)$, the following function is minimized at time t:

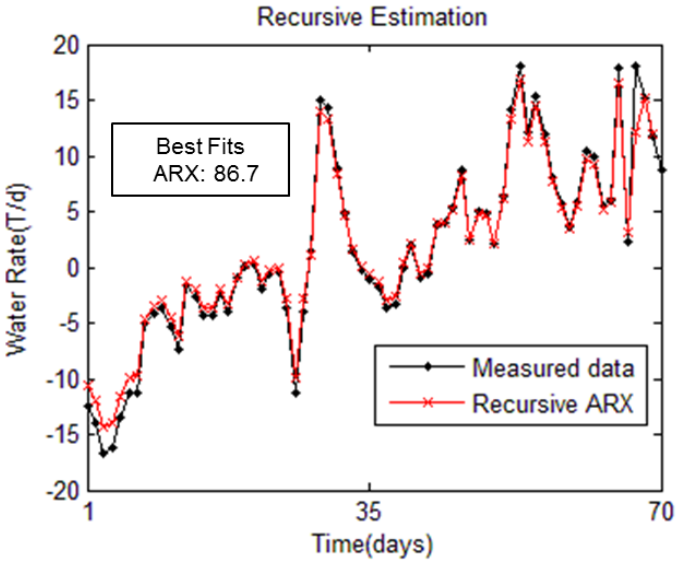
$$\sum_{k=1}^{t} \lambda^{t-k} e^2(k) \tag{5.1.7}$$

This approach discounts old measurements exponentially such that an observation that is τ samples old carries a weight that is equal to λ^τ times the weight of the most recent observation. $\tau = \frac{1}{1-\lambda}$ represents the memory horizon of this

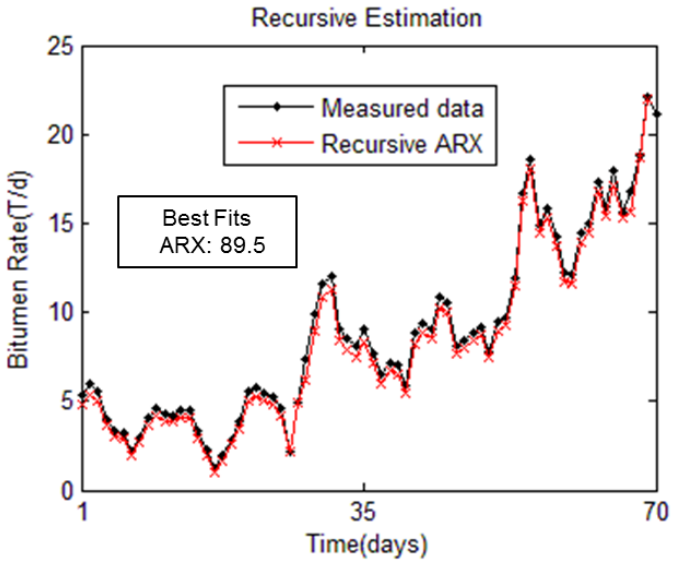
algorithm. Measurements older than $\tau = \frac{1}{1-\lambda}$ typically carry a weight that is less than about 0.3. λ is called the forgetting factor and typically has a positive value between 0.97 and 0.995.

5.2 Recursive Estimation for Two Months

First, the performance of recursive estimation for the short term, two months, will be tested. The basic model structure was built in the model structure selection section in Chapter 4 and the ARX model is chosen because the BJ model has too many parameters to be estimated, which will decrease the recursive estimation accuracy. The results are shown in Figure 4.1. The model fits improve a lot to the non-recursive estimation and the coefficients are slightly changed every time step, which can be observed in Figure 5.2. Because the order of the ARX model for the water production rate is $nA = 2$ and $nB = [1 \quad 1]$, there are four free coefficients changing with time. The order of the ARX model for the bitumen production rate is $nA = 1$ and $nB = [1 \quad 1]$ — there are three free coefficients. It is obvious that changing the coefficients due to the reservoir condition will result in a better performance than just fixing them. Using the recursive estimation, the model will update every time step. We also noticed that for a relatively short time period, the coefficients are almost unchanged, which proves that the non-recursive estimation has a acceptable performance for short-term prediction.

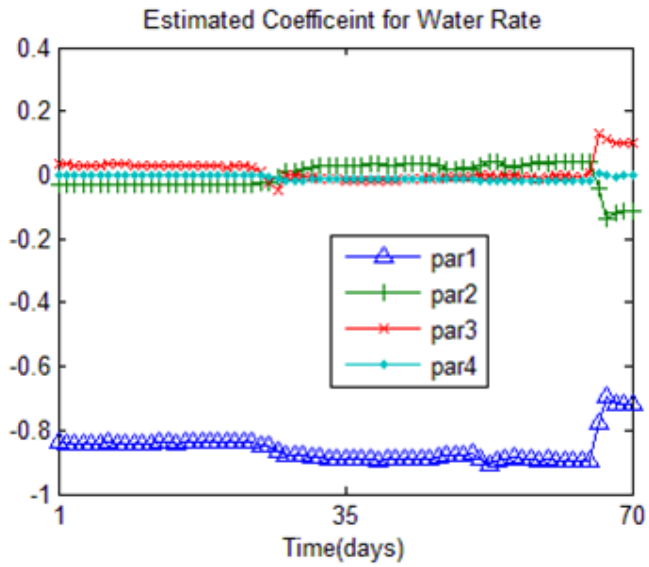


(a) Water production rate

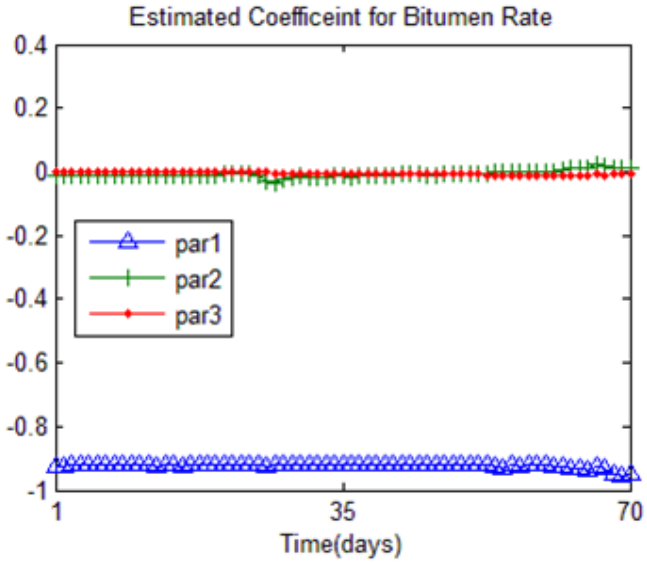


(b) Bitumen production rate

Figure 5.1: Recursive estimation for two months



(a) Water channel

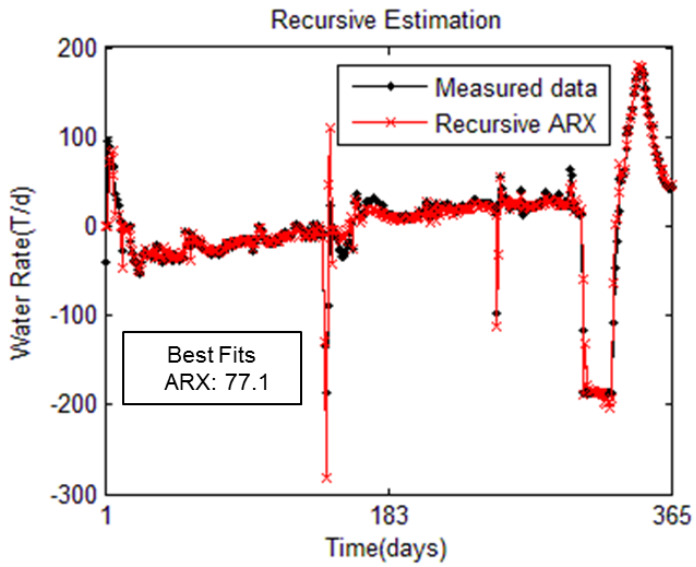


(b) Bitumen channel

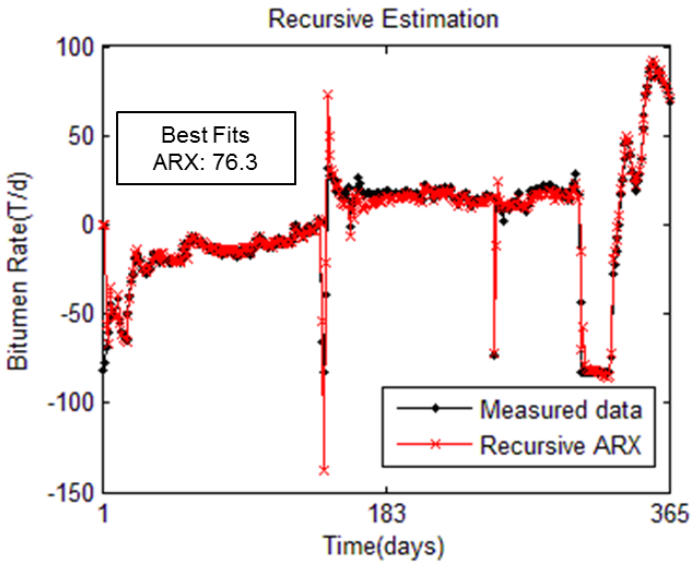
Figure 5.2: Coefficients changing with time for two-month recursive estimation

5.3 Recursive Estimation for One Year

Once the recursive estimation has been completed for the two-month model, a longer time period, for example, one year, can be used to test the validity of the recursive estimation. The results are depicted in Figure 5.3. The coefficients changing with time are shown in Figure 5.4. In Figure 5.3, recursive estimation captured nearly all reservoir dynamics even when there was no production. There are some jumps in coefficients due to big fluctuations in the production profile, which can explain why non-recursive estimation cannot be used to predict long time periods. The coefficients of non-recursive estimation is fixed. When there are no big changes in the production profiles, coefficients won't be changed significantly. As a result, the non-recursive model could capture the main dynamics of the system. However, obviously the coefficients need to be changed either when the production is stable or when it is fluctuating.

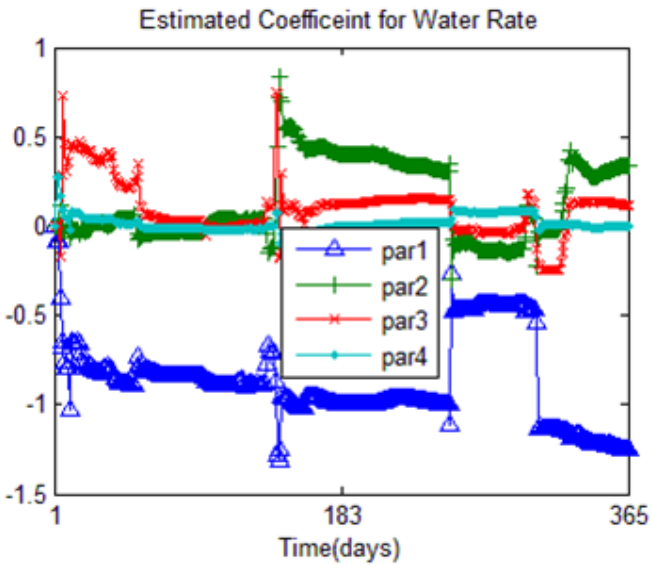


(a) Water production rate

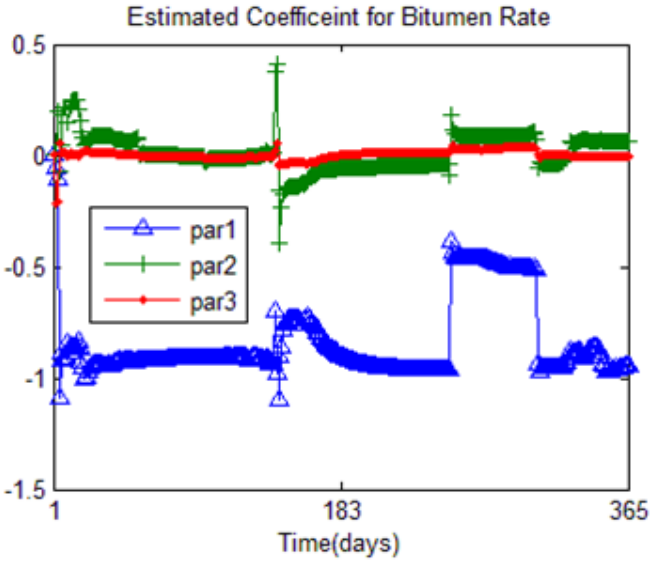


(b) Bitumen production rate

Figure 5.3: Recursive estimation for one year



(a) Water channel

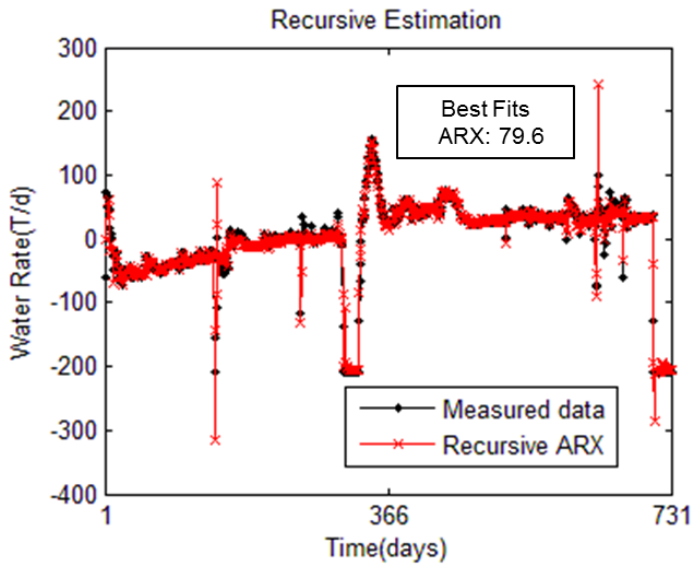


(b) Bitumen channel

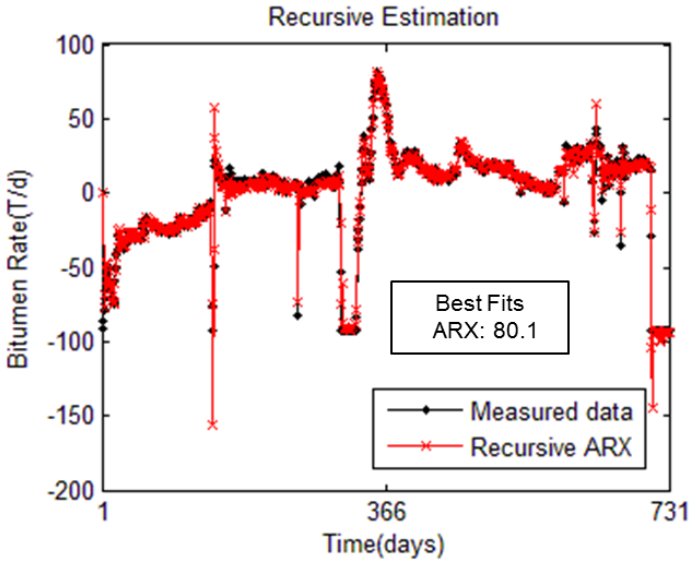
Figure 5.4: Coefficients changing with time for one-year recursive estimation

5.4 Recursive Estimation for Two and Five Years

Based on the previous case, we could infer that recursive estimation is validated for an even longer time period. Another two cases, recursive estimation for two years and five years, will be tested here. Figure 5.5 - Figure 5.8 show the results. From the figures we can see that the predicted production rates are almost perfectly matched with the real ones. The model fits are even better than those of recursive estimation for one year. That is because when the time period is longer, the weight of absolute error is lower than that of the short time case. The recursive estimation captures all the reservoir dynamics. Through all these investigation, it is safe to say that recursive estimation can be used for long-term prediction.

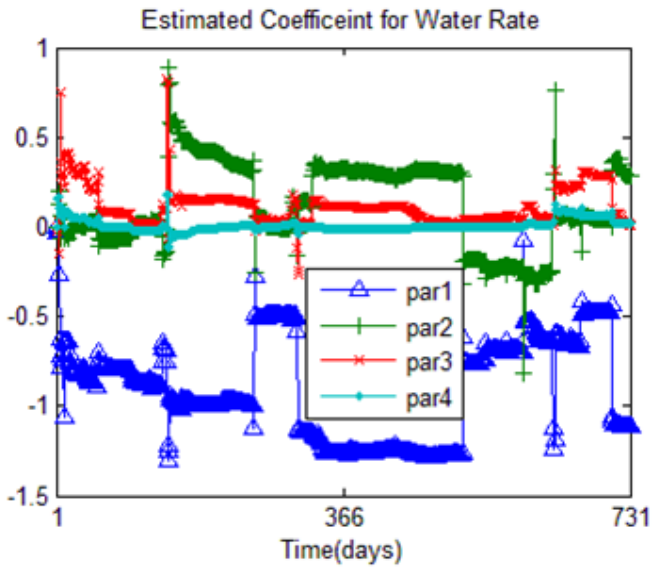


(a) Water production rate

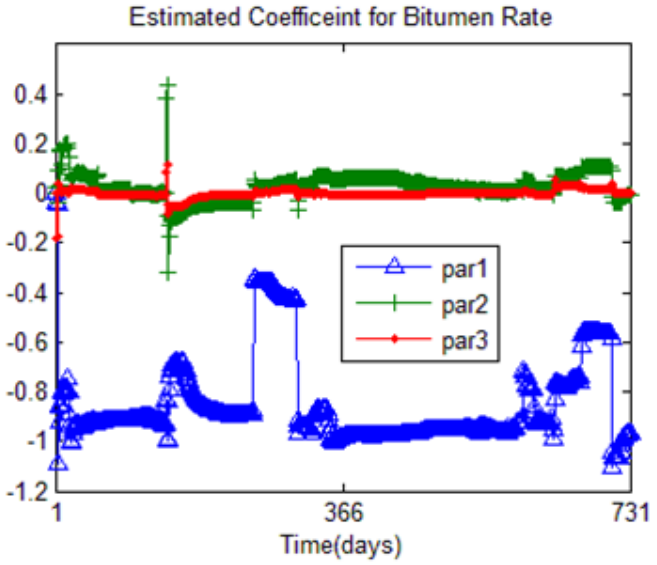


(b) Bitumen production rate

Figure 5.5: Recursive estimation for two years

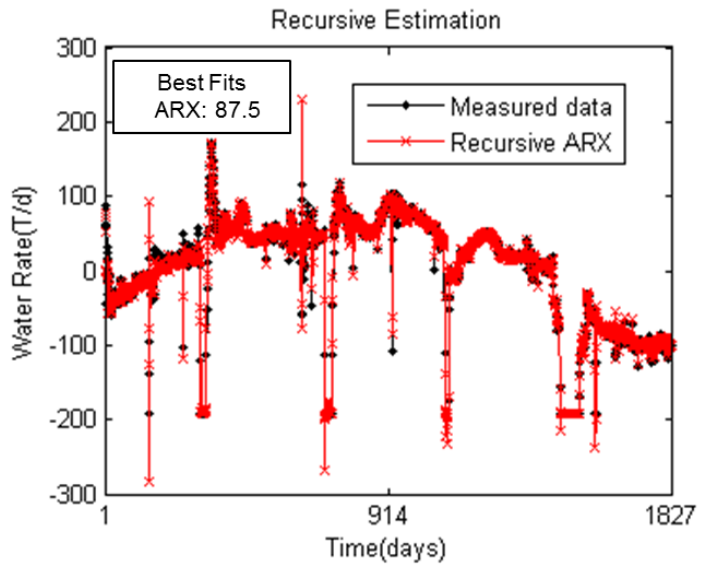


(a) Water channel

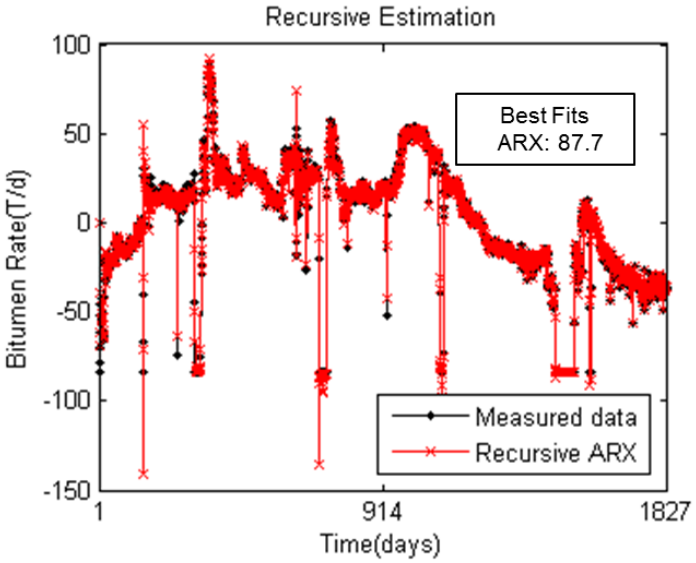


(b) Bitumen channel

Figure 5.6: Coefficients changing with time for two-year recursive estimation

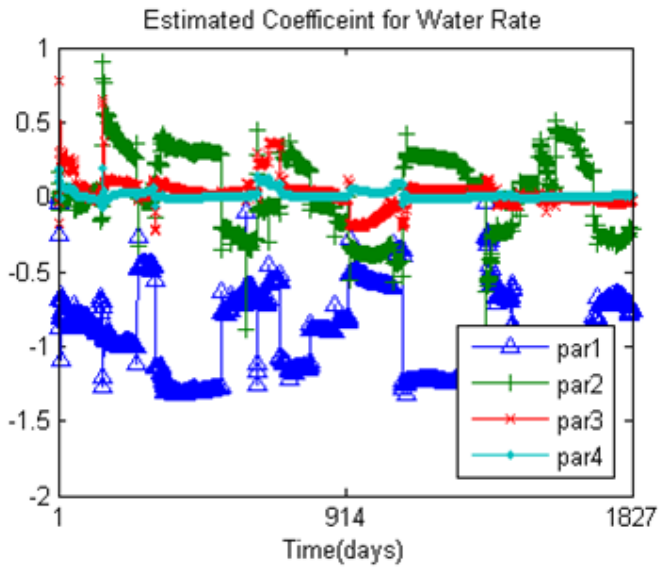


(a) Water production rate

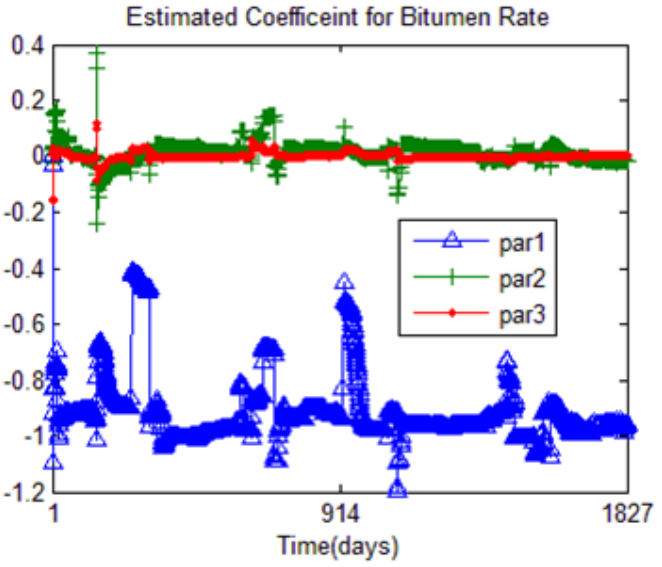


(b) Bitumen production rate

Figure 5.7: Recursive estimation for five years



(a) Water channel

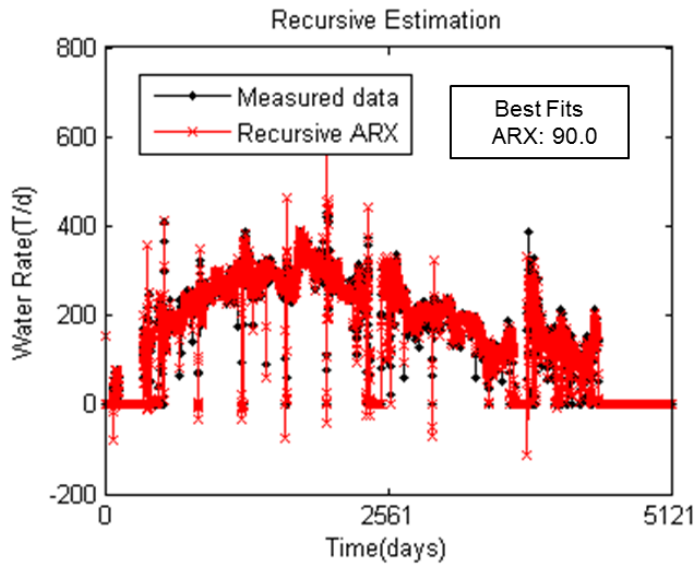


(b) Bitumen channel

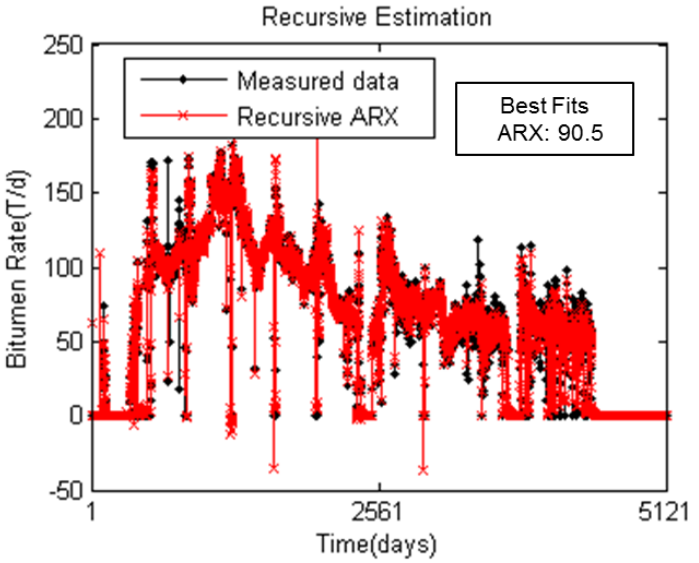
Figure 5.8: Coefficients changing with time for five-year recursive estimation

5.5 Using the Same Model Structure to Predict in Another Well Pair

The data used here is recorded through a single well pair while there are three well pairs in total in the reservoir. Another way to test the effectiveness of model is to apply the same model structure in another well pair. If the model can successfully predict the water and bitumen rate in another well pair in spite of some difference in the operation and well conditions, it means the model can really capture the reservoir dynamics instead of merely fitting into a single well pair. As shown in Figure 5.9, the model output fits the measured data perfectly. The coefficients are depicted in Figure 5.10. The coefficients are different from those in Figure 5.8 because of different well event. Supported by this case, the model built before can represent reservoir dynamics.

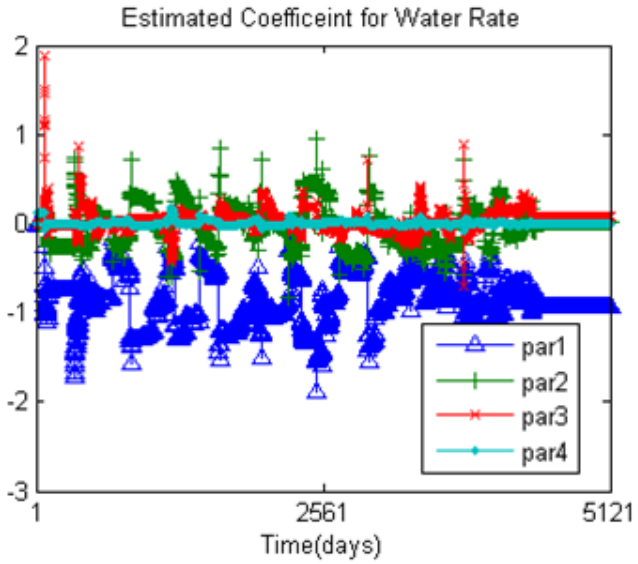


(a) Water production rate

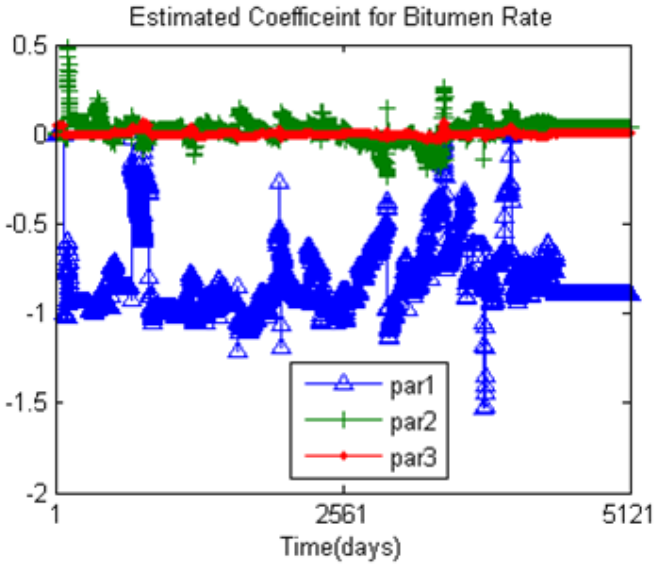


(b) Bitumen production rate

Figure 5.9: Recursive estimation for 15 years in another well pair using the same model structure



(a) Water channel



(b) Bitumen channel

Figure 5.10: Coefficients changing with time for 15 years recursive estimation in another well pair using the same model structure

Chapter 6

Strategy of Improving Multiple-step Predictions

Before we talk about the multiple-step prediction, two cases have to be distinguished: simulation and prediction. If the responses of the model to an input sequence have to be calculated while the process outputs are unknown, that is called simulation. However, if the process outputs are known to some time instant, say, k, and it is asked for the model output l steps in the future, that is called prediction. Until now all the models' forecasts are based on one step predictions, and in this case $l = 1$. Figure 6.1 and Figure 6.2 explain the meaning of these two scenarios (Nelles, 2001). From Figure 6.1, it is obvious that the simulation is fully deterministic:

$$\hat{y}(k) = G(z^{-1})u(k) \tag{6.0.1}$$

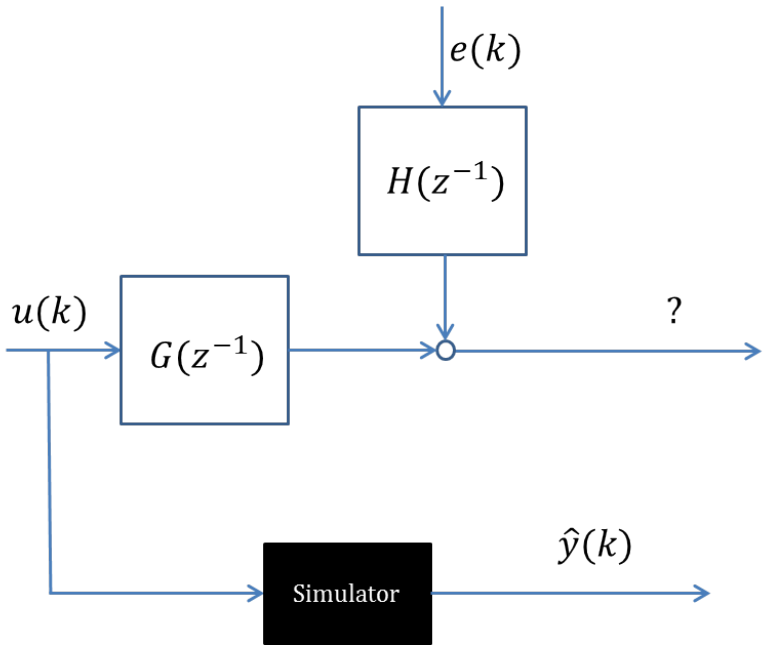
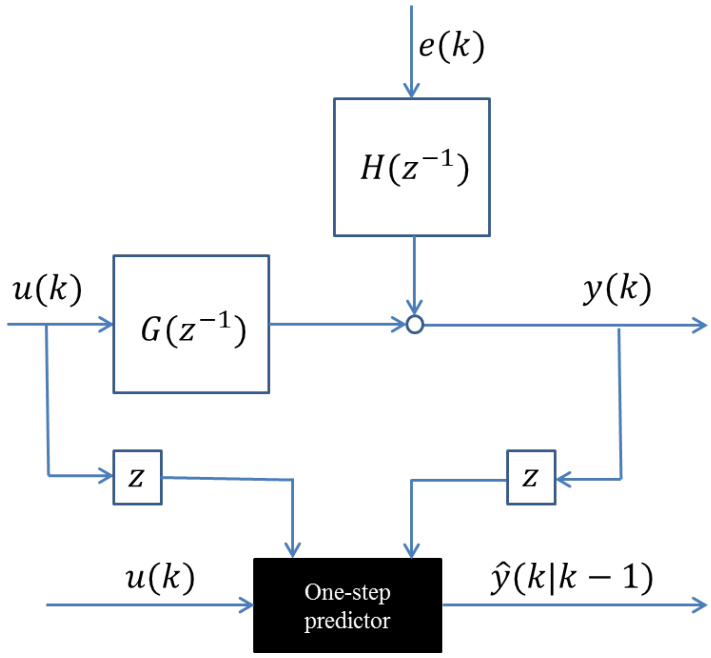


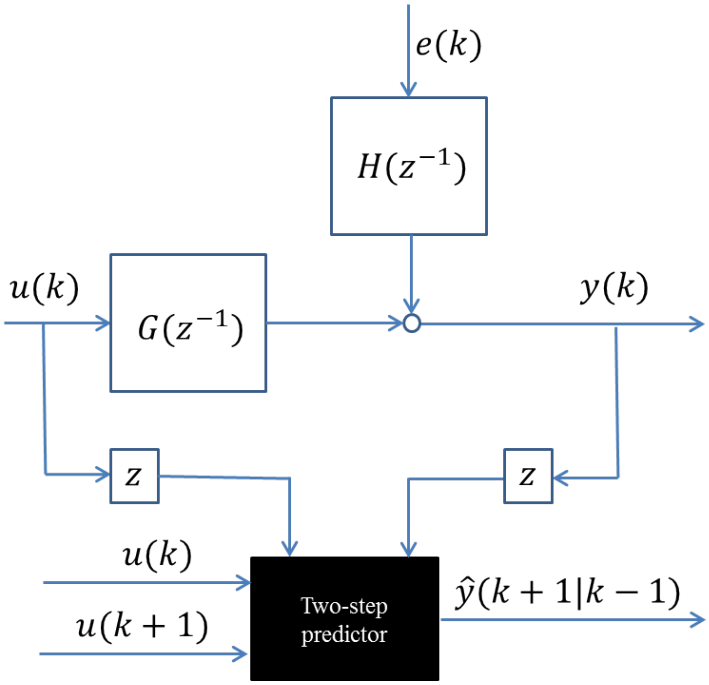
Figure 6.1: Simulation structure

Thus, the noise model $H(z^{-1})$ seems irrelevant for simulation. But actually $H(z^{-1})$ influences the estimation of the parameters in $G(z^{-1})$, so it does have an impact on the simulation results even if it doesn't appear in Equation 6.0.1.

In contrast to simulation, the information about the previous process output can be used for prediction. In Figure 6.2, the expression "$|k-1$" means "on the information available at time instant $k-1$." For prediction, besides the inputs, the previous process outputs are known. Note that if the prediction horizon l becomes very large the information about the previous outputs becomes less important. As a result, as $l \to \infty$ prediction approaches simulation.



(a) One-step prediction

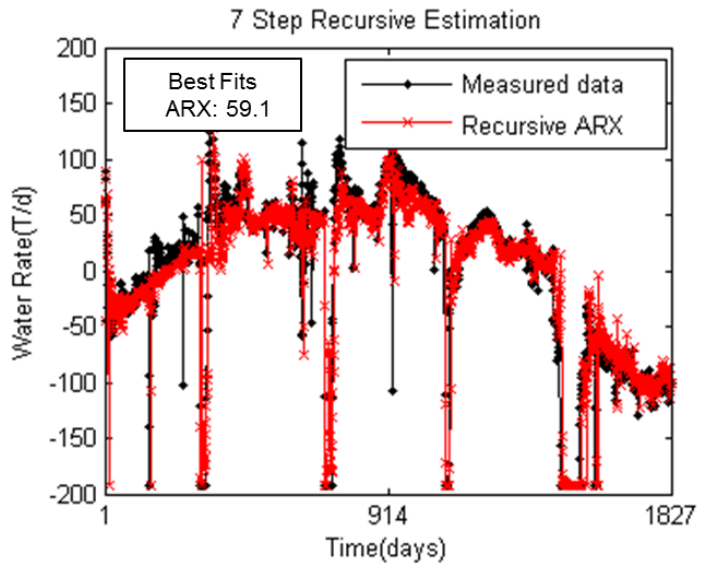


(b) Two-step prediction

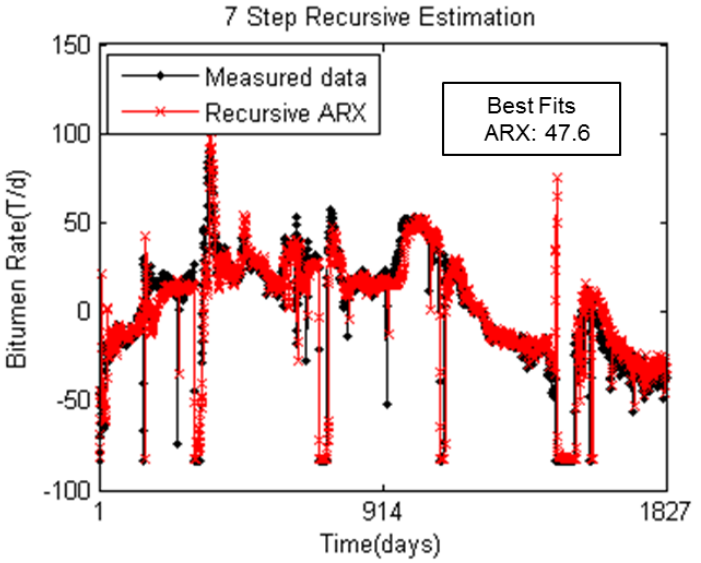
Figure 6.2: Prediction structure

6.1 Seven-step and Ten-step Recursive Estimations

At each time step the recursive model has one set of fixed coefficients. In the seven-step recursive estimation, the coefficients set obtained at time k will be used to predict the water and bitumen rate at time $k + 7$. Here,e a five years recursive estimation is chosen directly to test the validity of the long-term prediction. The results are depicted in Figure 6.3. Because the seven-step prediction uses data until the current time instant to predict for the time instant which is seven time steps away from it, the accuracy of the model fits decrease a little bit. However, it is fair to say that the model basically captures reservoir dynamics. All trends of production profiles are fitted ideally. It is also important to clarify that a small modification was made by setting all the predicted negative and too high production value to zero. This was done because the production value cannot be zero or bigger than the valve value in reality. As discussed previously, if the production value is wrong, errors will result. All in all, the seven-step recursive estimation is valid for long-term prediction. We also test the longer prediction horizon, ten steps, to see the effect of increasing the prediction horizon. The results are shown in Figure 6.4. From the figure we can see that the water rate model is still promising; however, there is a big decrease of best fit in the bitumen rate model, which could be caused by the initial sets of coefficients, which are not that precise.

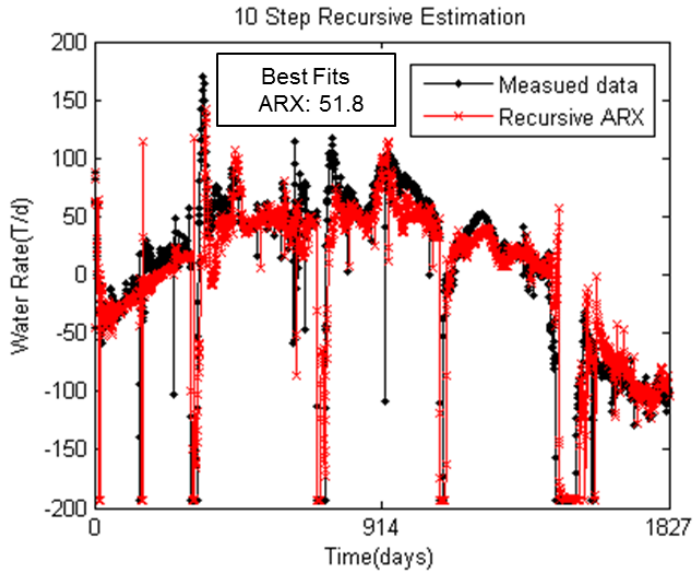


(a) Water channel

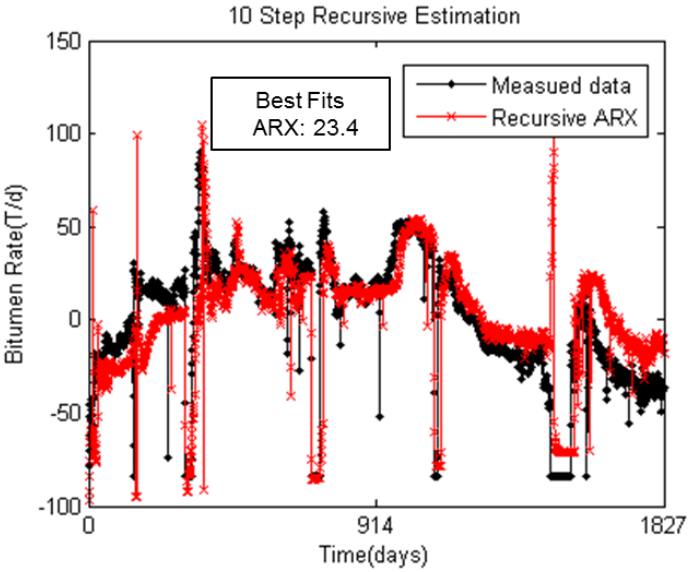


(b) Bitumen channel

Figure 6.3: Seven-step recursive estimation for five years



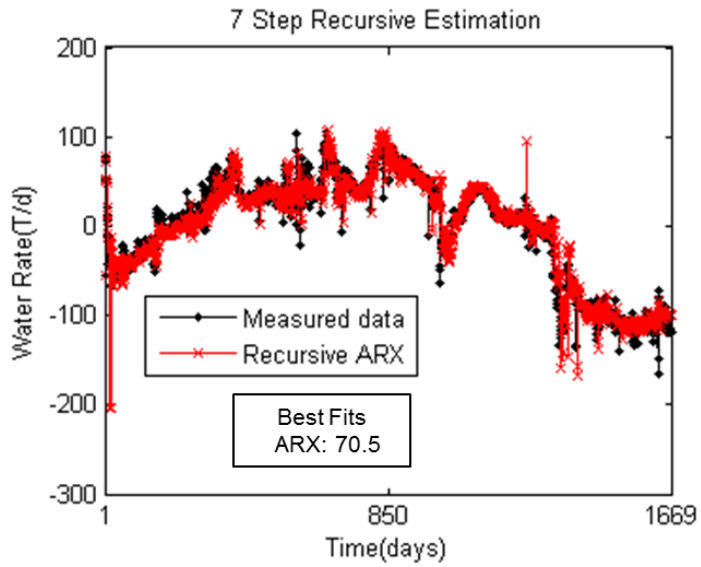
(a) Water channel



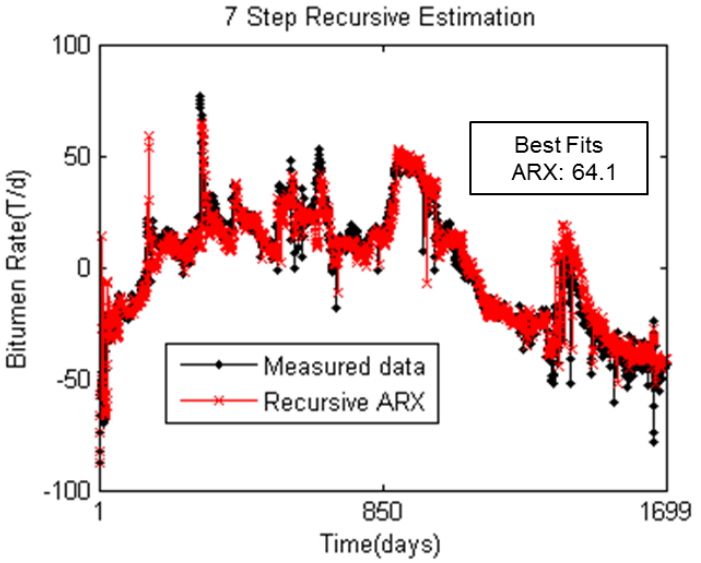
(b) Bitumen channel

Figure 6.4: Ten-step recursive estimation for five years

When there are production shutdowns in production profiles, the model fits are never as good as before. In order to test the effect of big changes in production profiles, we can remove them and then rebuild the model. In Figure 6.5 and Figure 6.6, the model fits improved a lot compared to the model using unprocessed data. The reason why one-step recursive estimation will give better results compared to seven-step and ten-step recursive estimations is that one-step recursive estimation will update the model structure every time step so the model itself "knows" what happened at each time step, while production shutdowns in production profiles are too abrupt for seven-step and ten-step recursive prediction and it will take some time for the model to find the proper coefficients to reflect the real reservoir dynamics. Basically, when the production rate is stable or not changing too much, the coefficients of the model won't change too much either. In that case, when the production shutdowns of the data is removed, the coefficients also tend to be stable and easier to identify. Thus, in reality, if there are production shutdowns in the long-term dataset, one should remove the them or restart the identification from a new time instant when the production is stable.

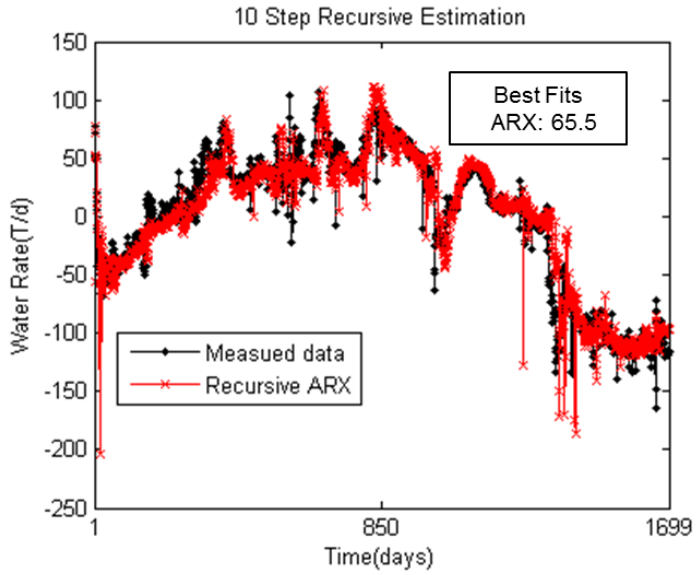


(a) Water production rate

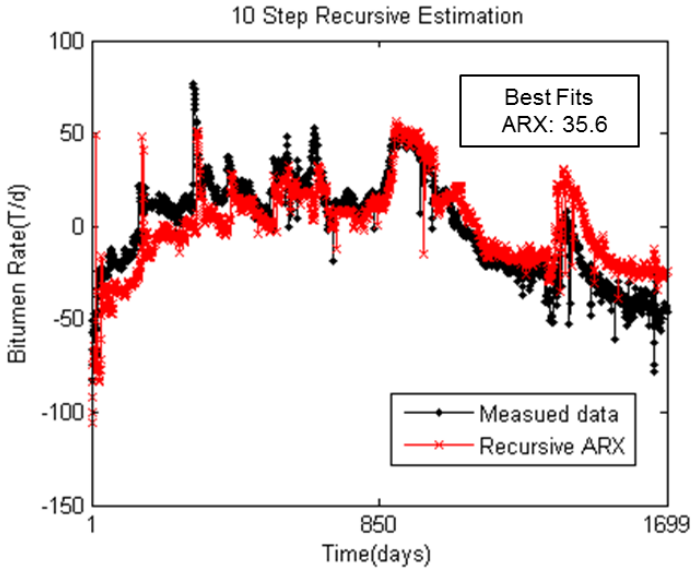


(b) Bitumen production rate

Figure 6.5: Seven-step recursive estimation without production shutdown for five years



(a) Water channel

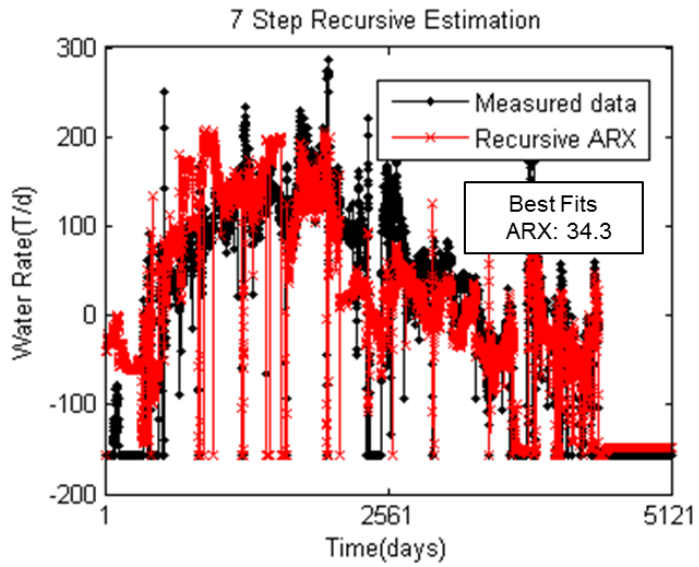


(b) Bitumen channel

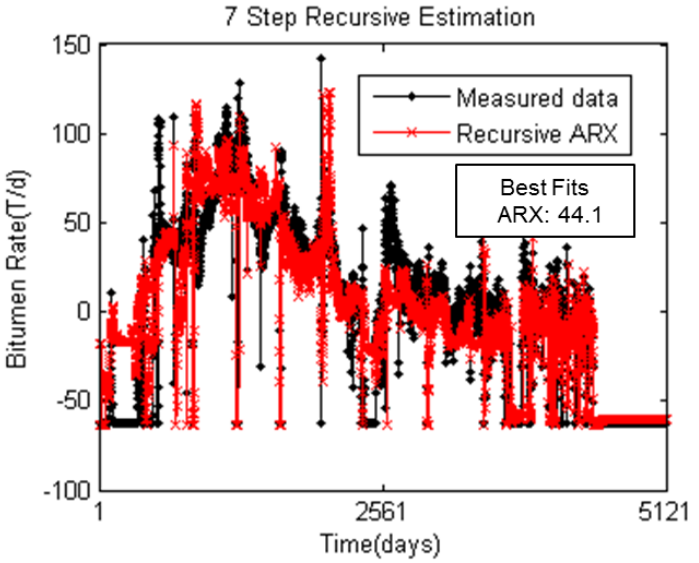
Figure 6.6: Ten-step recursive estimation without production shut-down for five years

6.2 Seven-step Recursive Estimation in Another Well Pair

In order to show that a seven-step recursive estimation can be used to describe the reservoir dynamics, the same model structure will be applied to another well pair to check the model performance. The data used here are the 15-year operation profiles. One thing that needs to be clarified is that in this part the coefficients will also be the same as in the previous seven-step recursive estimation in Section 6.1. Thus, it means we just use a completely built model to make the prediction under a different operating situation. The only same part of these two cases is these two well pairs come from the same reservoir, which means the reservoir dynamic is the same. The results are depicted in Figure 6.7. Compared to the previous results, the model fits are not satisfactory, but they capture the main trends of the production profiles. Considering that this is a seven-step recursive estimation in another well pair, it is still safe to say that the results are encouraging.



(a) Water production rate

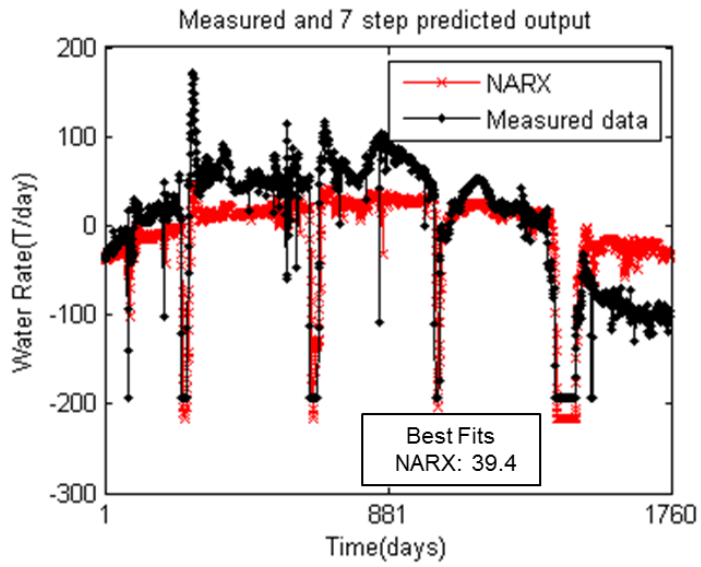


(b) Bitumen production rate

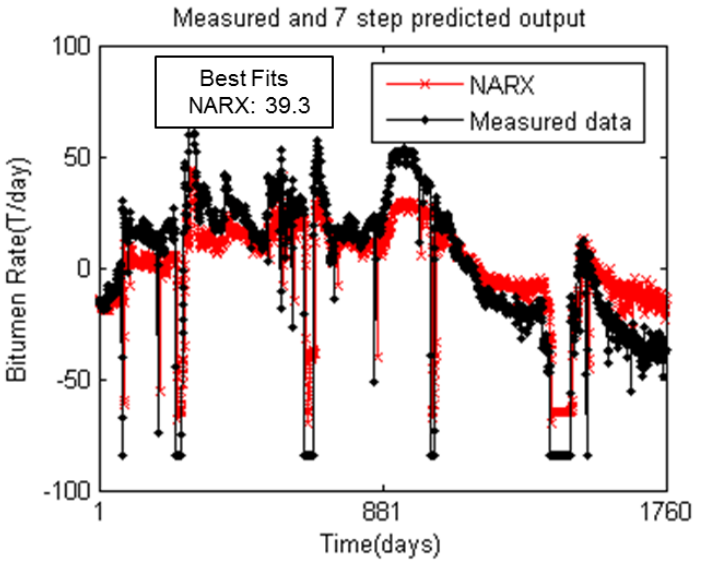
Figure 6.7: Seven-step recursive estimation for 15 years in another well pair using the same model structure

6.3 Seven-step Nonlinear ARX Estimation

After trying all kinds of linear models, a set of nonlinear models should also be checked. The most widely used nonlinear model structure - nonlinear ARX model will be used. The input and output variables will be kept the same as before. The time period will be the same as the seven-step recursive estimation. The results are shown in Figure 6.8. The model can just reflect the basic reservoir dynamics, and in order to fit the production shutdown, the model sacrifices accuracy in the nearby region. The same assumption made here is that the error of seven-step nonlinear estimation comes from the production shutdown. The big jumps are removed as well to verify the assumption. The results are depicted in Figure 6.9. The model performances improve a lot, and all reservoir dynamics are captured basically. Once again, we proved that the big error comes from the big fluctuation from the production profile, which should be avoided from the identification process.

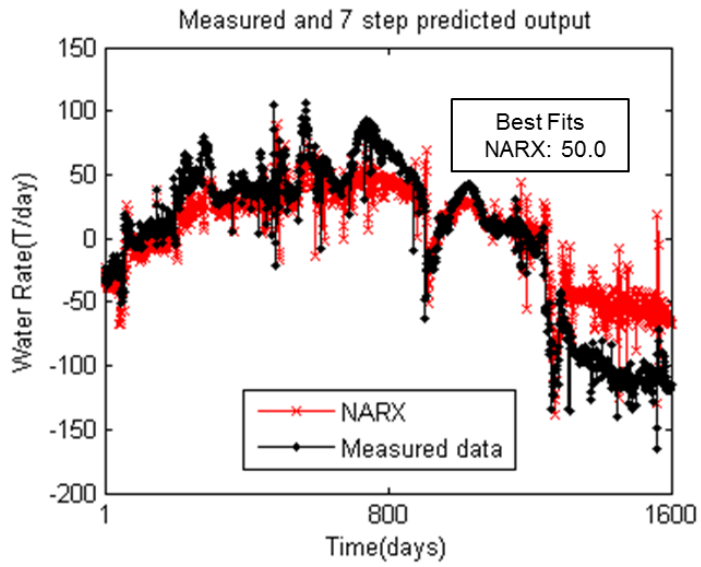


(a) Water production rate

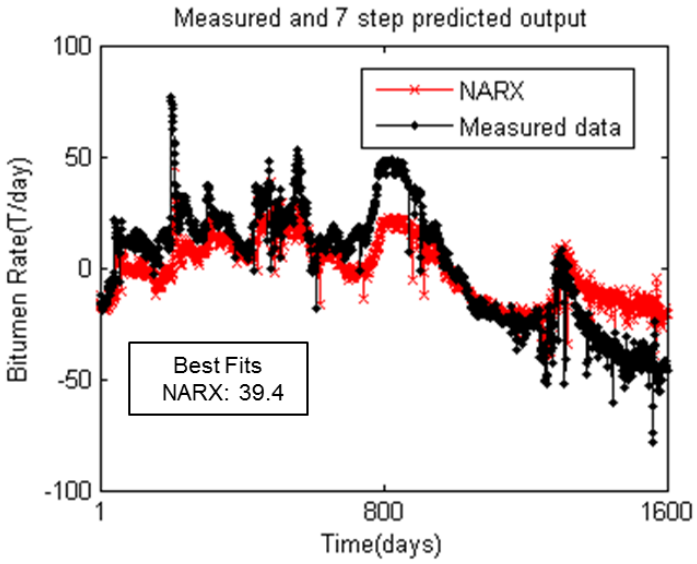


(b) Bitumen production rate

Figure 6.8: Seven-step nonlinear ARX estimation for five years



(a) Water production rate

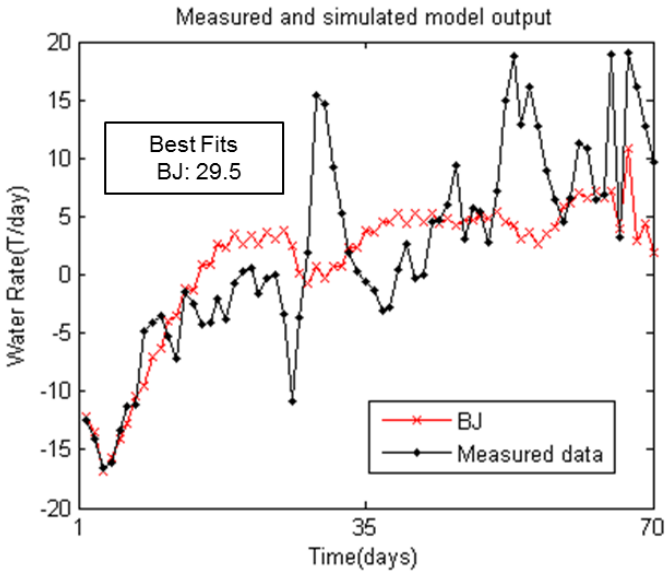


(b) Bitumen production rate

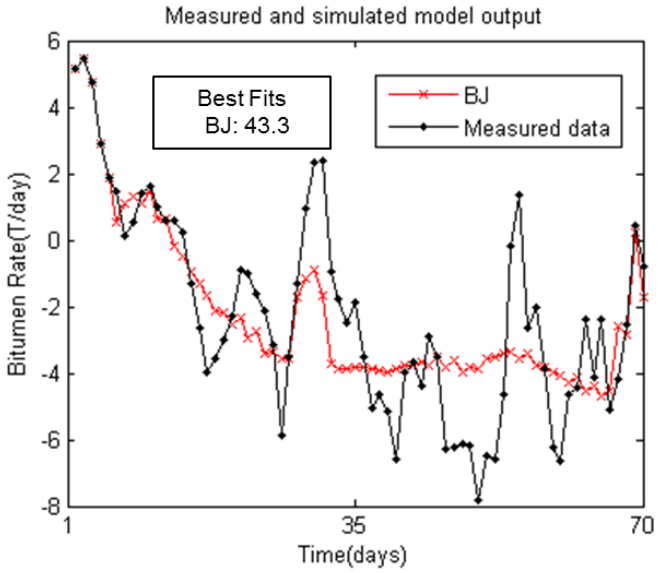
Figure 6.9: Seven-step nonlinear ARX estimation without production shutdowns for five years

6.4 Pure Simulation

After many cases of multiple-step recursive estimation and nonlinear estimation, we also want to test the performance of linear non-recursive pure simulation, which means there will be infinite-step predictions. Because the BJ model is the best linear model that could capture the reservoir dynamics, we use BJ model to test the pure simulation case. The results are shown in Figure 6.10. The BJ model here is employed to simulate the output for the next two months. From the figure we can see that the models capture the main trend in the measured data. We also use the model to simulate the output for the next four months. The results are depicted in Figure 6.11. When there is a production shutdown, the fix-coefficients BJ models fail to capture the abrupt change and, overall, the models' performance is insufficient. This is another example showing that the non-recursive model should be used in short-term predictions.

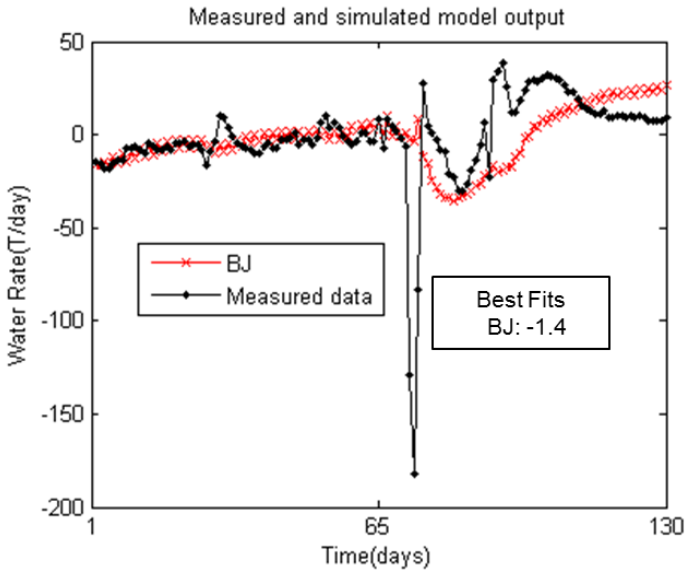


(a) Water production rate

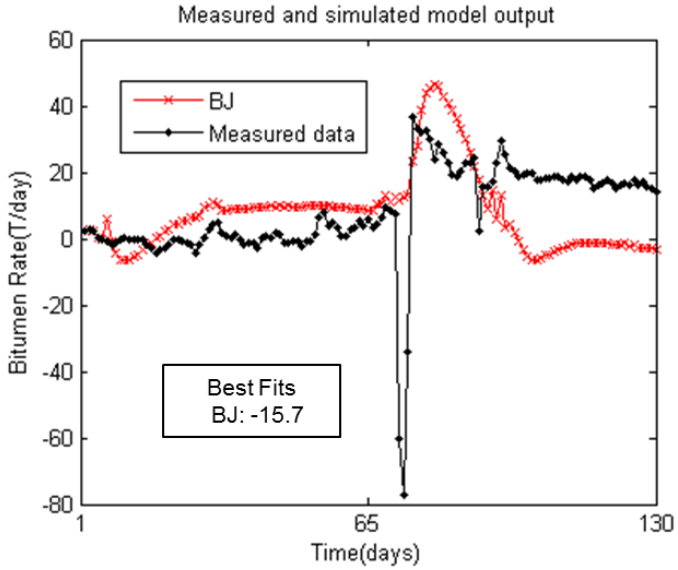


(b) Bitumen production rate

Figure 6.10: Open loop simulation for two months



(a) Water production rate



(b) Bitumen production rate

Figure 6.11: Open loop simulation for four months

Chapter 7

Importance of Input Design for System Identification of Reservoirs

The previous discussions show that the model error comes mainly from the production shutdowns in production profiles. Production rate is the output variable in the model and the output is actually determined by the input. If the input variables have enough excitation to be identified, the accuracy of the model will be improved significantly. However, the real reservoir cannot be accessed. As a result, simulation reservoir models will be built in this section to realize the input design. In this section, we will combine different reservoir models and different kinds of input signals. The reservoir models are a homogeneous SAGD reservoir, naturally fractured gas reservoir, channelled

water-flooding reservoir and heterogeneous SAGD reservoir. The different input signals are pseudo random binary sequence (PRBS) signal, multiple level PRBS signal and sine wave signal. We are trying to determine two things: if the input signal is well-designed, will the system identification performance be improved, and what kind of reservoir will be suitable for system identification.

7.1 Homogeneous SAGD Reservoir

A 3-D SAGD reservoir model, with $11 \times 10 \times 10$ grid blocks with model dimensions of 44 meters $\times$ 1400 meters $\times$ 30 meters, has been used in this case study. A pair of injector and producer wells is used in this simulation model. As we discussed before, steam injection rate and production pressure will be used as in the input signals, and the output signals are water rate and bitumen rate in the producer. The permeability is 2200 md in the i and j direction, and the permeability in teh k direction is 1760 md. The other reservoir properties are presented in Table 7.1.

Table 7.1: Homogeneous SAGD reservoir properties

Parameters	Value
Depth from surface	280 m
Porosity	34%
Reservoir temperature	10° C
Formation heat capacity	2.35×10^6 J/($m^3 \times$°C)
Rock thermal conductivity	6.6×10^5 J/($m \times$ day$\times$°C)
Fluid injected	Steam
Steam quality	0.9
Steam temperature	224°C

The reservoir model is depicted in Figure 7.1. Figure 7.2 shows part of the PRBS input signal applied to this reservoir. The bottom axes are the designed PRBS input signal. The values change between two levels and the changing frequency is random. The input signal is applied to the CMG flexwell model and the top axes show the output value obtained from the CMG simulator. All these values will be sent to system identification to build a new proxy model. The blue line in the figure is for model construction and the purple one is for model validation. The results are depicted in Figure 7.3. From the figure we can see that the model can capture the reservoir dynamics perfectly, and the focus here is pure simulation, as discussed in Chapter 6.

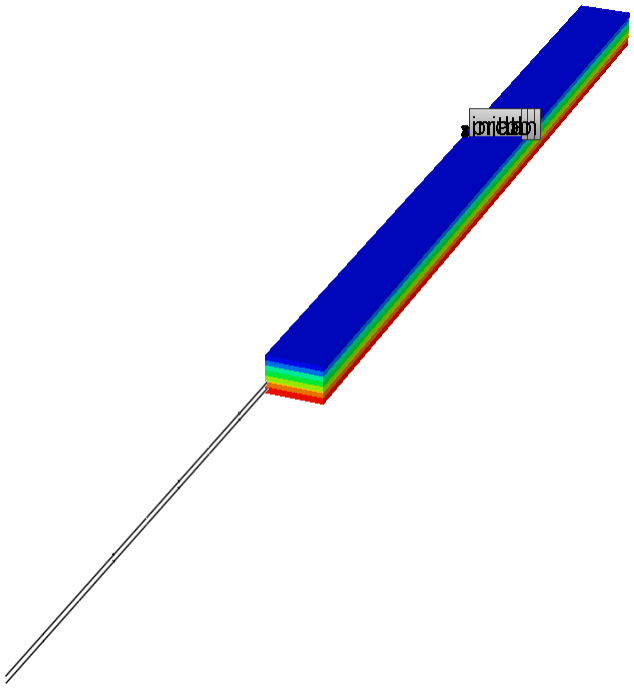


Figure 7.1: Homogeneous SAGD reservoir

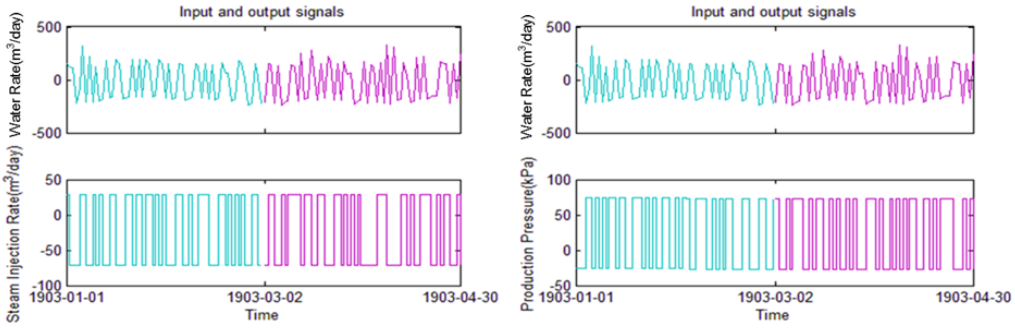
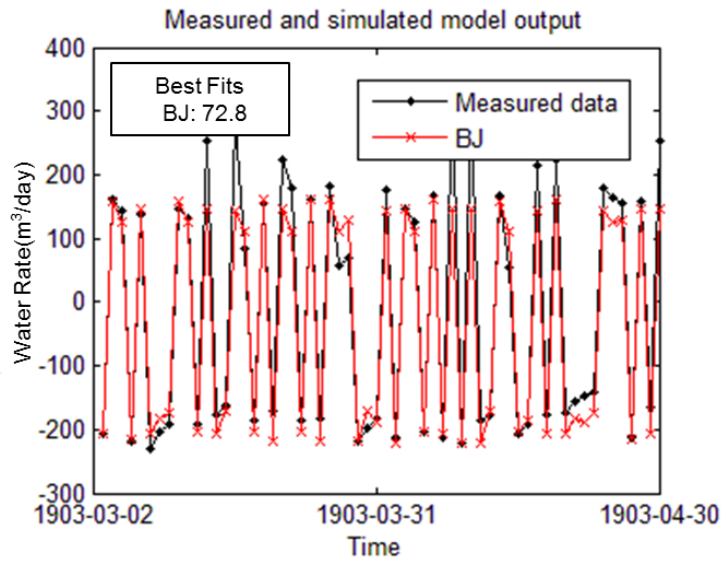
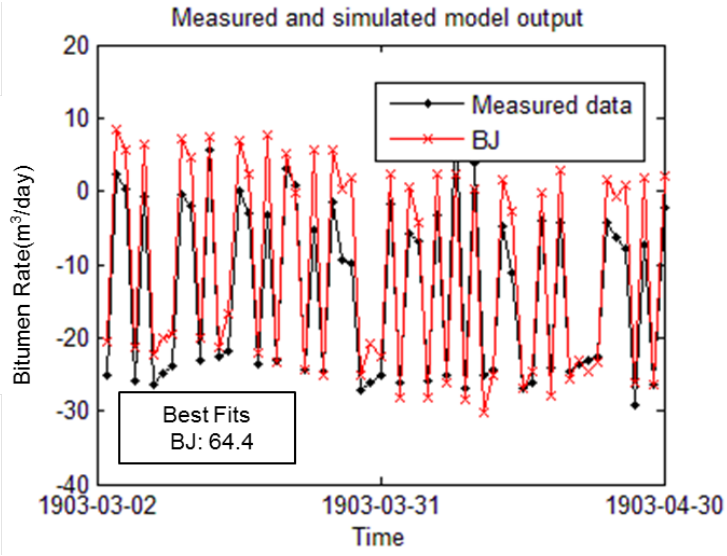


Figure 7.2: PRBS input signal



(a) Water production rate



(b) Bitumen production rate

Figure 7.3: Open loop simulation of homogeneous SAGD reservoir for PRBS input design

Although the results in the previous section are very encouraging, the input values just have two levels. In order to add complexities to the case, we arranged a multiple-level PRBS input signal as shown in Figure 7.4. The results are depicted in Figure 7.5. From the figure we can see that the model fits are still promising. The quality of the fits is slightly less than that of the fits generated by the PRBS signal. That is because the multiple-level PRBS signal has more uncertainties than that of PRBS signal. All in all, this is a good simulation that captures basically all reservoir dynamics. One thing that needs to be clarified is the selection of the frequency band of PRBS and multiple level PRBS signal. The lower and upper frequencies are expressed in fractions of the Nyquist frequency.

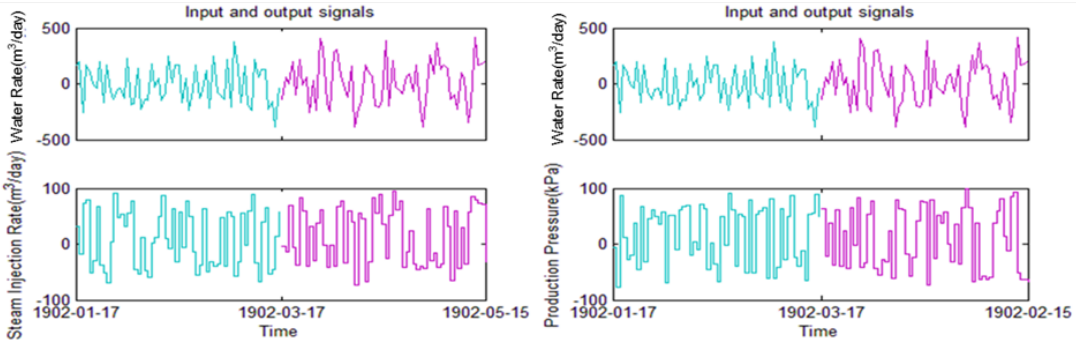
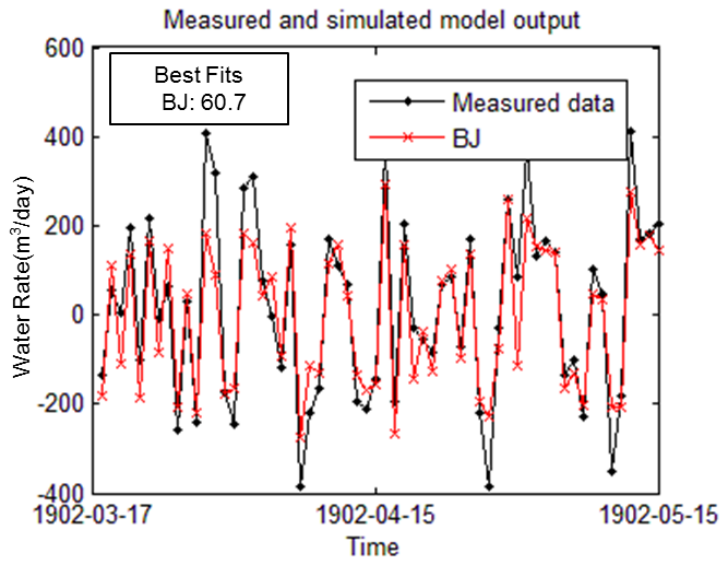
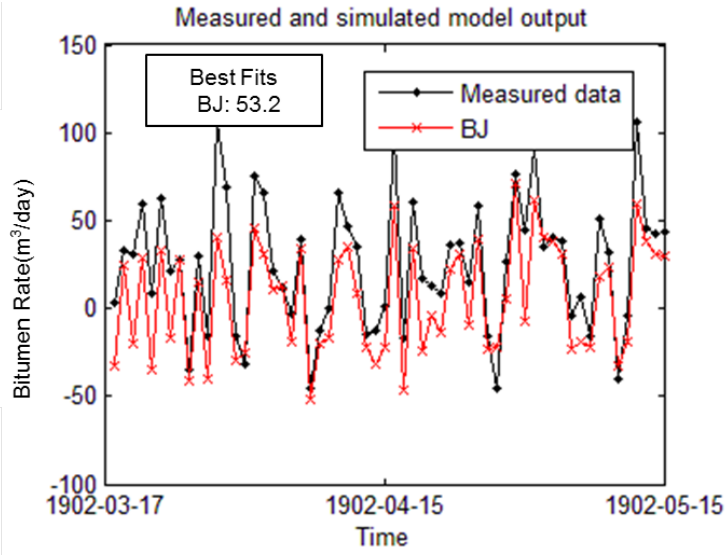


Figure 7.4: Multiple-level PRBS input signal



(a) Water production rate



(b) Bitumen production rate

Figure 7.5: Open loop simulation of homogeneous SAGD reservoir for multiple-level PRBS input design

Some other types of input designs, such as the sine wave input signals shown in Figure 7.6, can be tested. In order to increase the model validity, the data used for model construction has a different amplitude and frequency than the data used for model validation. The results are shown in Figure 7.7. The models capture the reservoir dynamics as well.

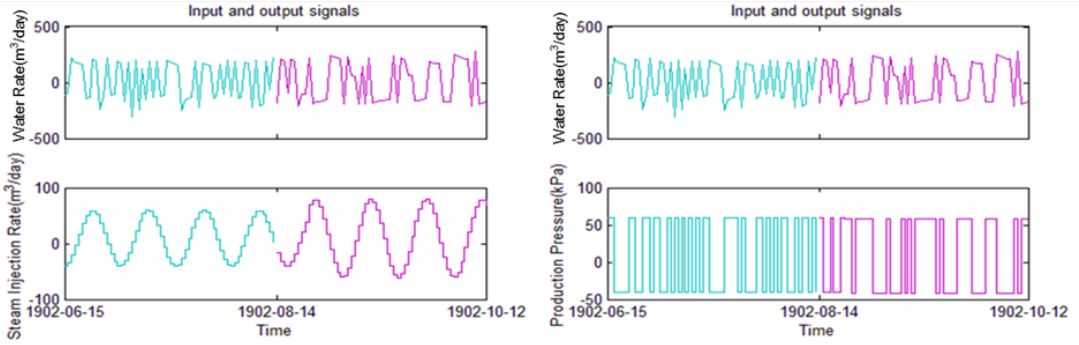
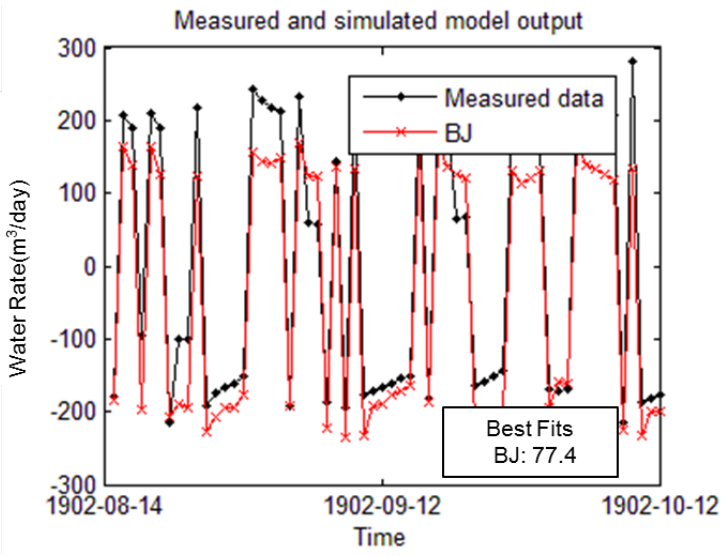
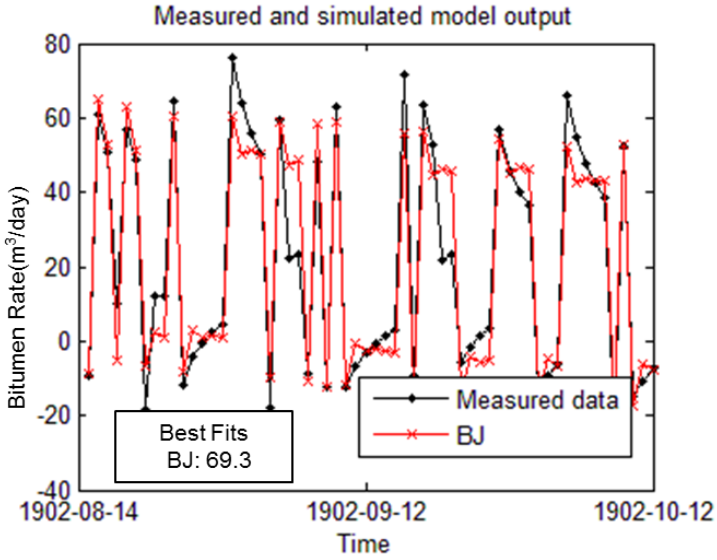


Figure 7.6: Sine wave input signal



(a) Water production rate



(b) Bitumen production rate

Figure 7.7: Open loop simulation of homogeneous SAGD reservoir for sine wave input design

7.2 Naturally Fractured Gas Reservoir

In this section a dual porosity 2-D naturally fractured gas reservoir will be used to test for the input design. The model has $25 \times 25 \times 1$ gird blocks with a dimension of 6000 feet $\times$ 7000 feet $\times$ 250 feet. Other parameters are presented in Table 7.2.

Table 7.2: Naturally fractured gas reservoir properties

Parameters	**Value**
Depth from surface	4000 ft
Matrix porosity	20%
Matrix permeability	1 mD
Fracture porosity	100%
Fracture permeability	0 to 11.34 mD
Initial reservoir pressure	3000 psia
Initial gas saturation	100%
No. of wells	3 producers

The fracture's structure is shown in Figure 7.8. In this model we didn't activate dual permeability, which means there is no direct flow between matrix cells. The gas will flow from the matrix cells to the fractures, and then to the producers. The permeability distribution of the fracture system is depicted in Figure 7.9.

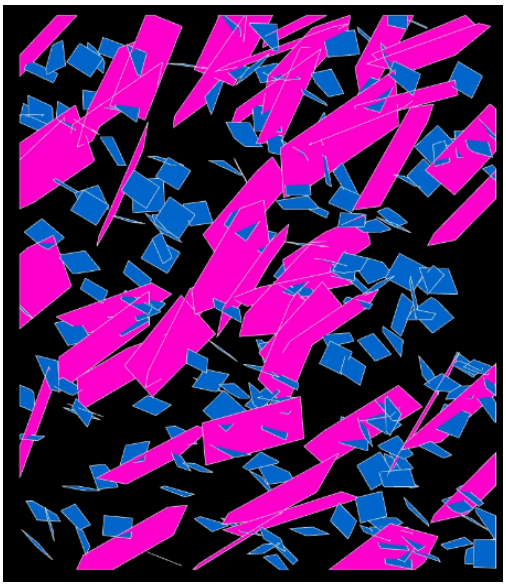


Figure 7.8: Discrete Fracture Network (DFN) model. Five percent of the fractures are displayed in the figure

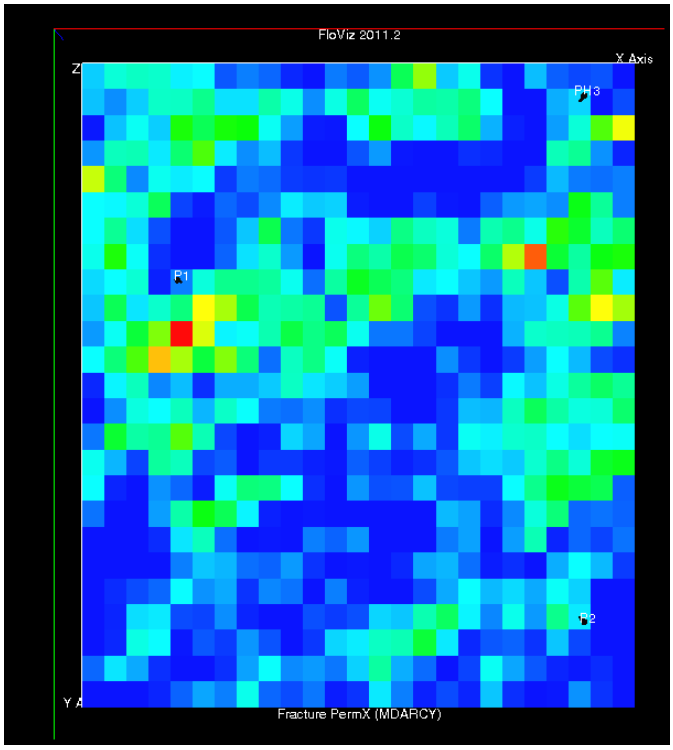


Figure 7.9: Permeability of fracture system

Since the model is a naturally fractured gas reservoir, we don't need to add injection wells. Only production wells are needed, which means that the steam injection rate isn't an input signal any more. In fact, in this case study, we have just one input, production pressure, and one output, gas production rate. We have chosen P2 well as our data source. All types of input signals are used such as PRBS signal, multiple level PRBS signal and sine wave signal. The only differences are the value and frequency. The procedure of building a proxy model is the same as that in the previous section. The results are shown in Figure 7.10 - Figure 7.12. From the figures, we can see that every input design case captures the reservoir dynamics and the prediction is almost perfect. That is because this is a relatively simple model compared to the SAGD reservoir model. And in case, we can build a Single Input Single Output (SISO) model for identification.

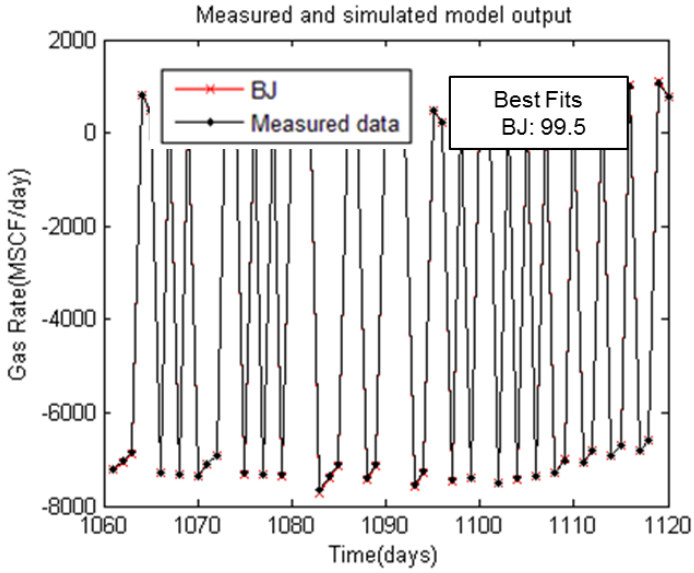


Figure 7.10: Open loop simulation of naturally fractured gas reservoir for PRBS input design

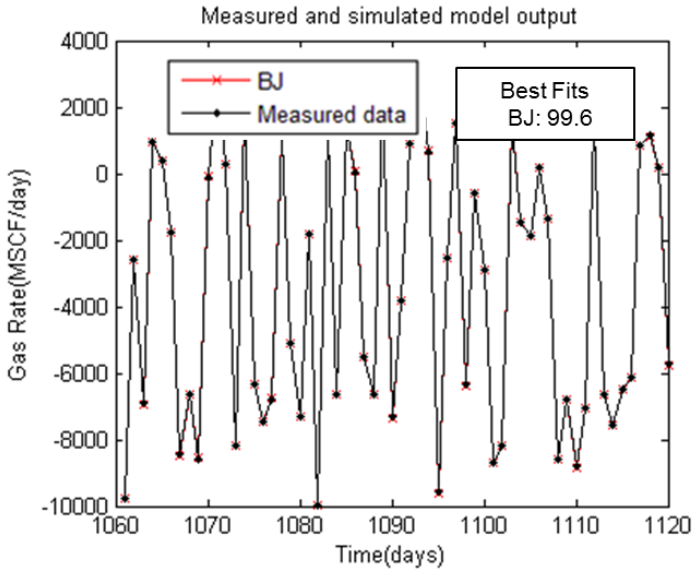


Figure 7.11: Open loop simulation of naturally fractured gas reservoir for multiple PRBS input design

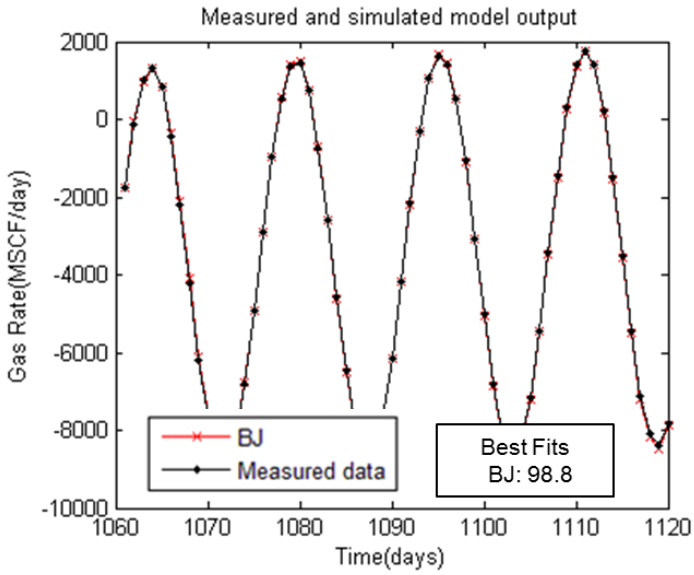


Figure 7.12: Open loop simulation of naturally fractured gas reservoir for sine wave input design

7.3 Channel Reservoir

After testing the performance of system identifications applied to a homogeneous SAGD reservoir and a naturally fractured gas reservoir, a 2-D channel reservoir will be built to proceed with the input design again, in this section. This is a 64 × 64 × 1 reservoir with the dimension of 2100 ft × 2100 ft × 32.8084 ft. The highly permeable sand channel is surrounded by low permeable shale. The permeability of the sand is 10,000 mD, and for the shale it is 500 mD. The other parameters are shown in Table 7.3. The visual shape of the channel is shown in Figure 7.13. I1 is a horizontal injector and P1 is a horizontal producer.

Table 7.3: Channel reservoir properties

Parameters	Value
Depth from surface	3000 ft
Porosity	20%
Initial reservoir pressure	5000 psia
Initial oil saturation	98%
No. of wells	1 injector & 1 producer
Fluid injected	Water

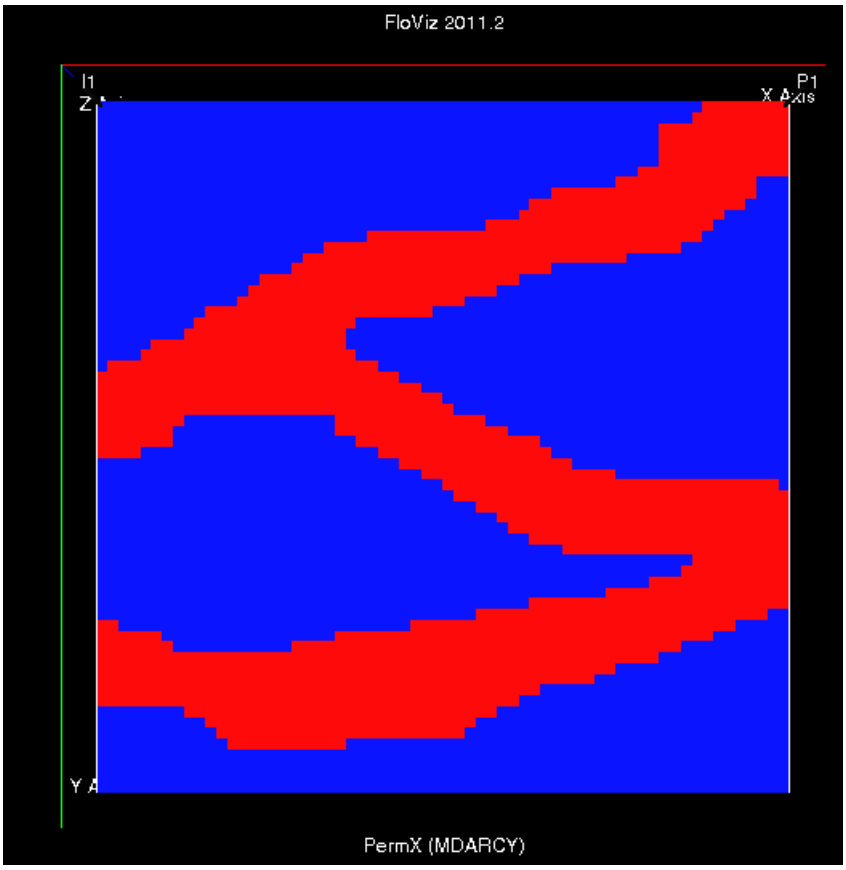
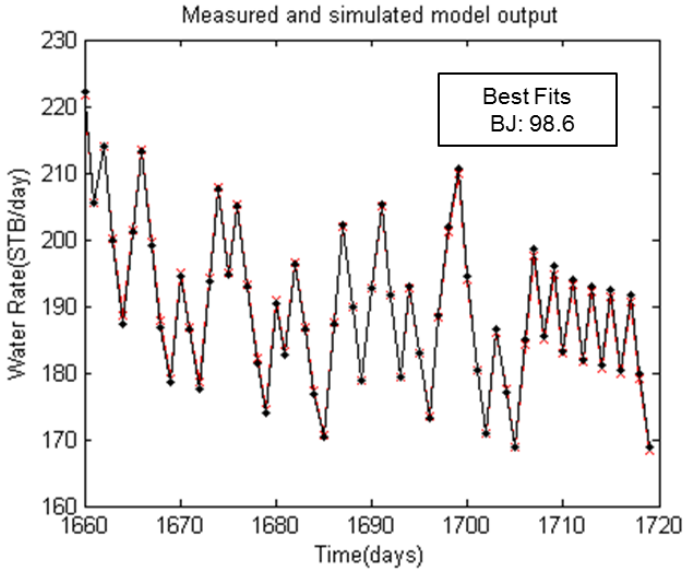
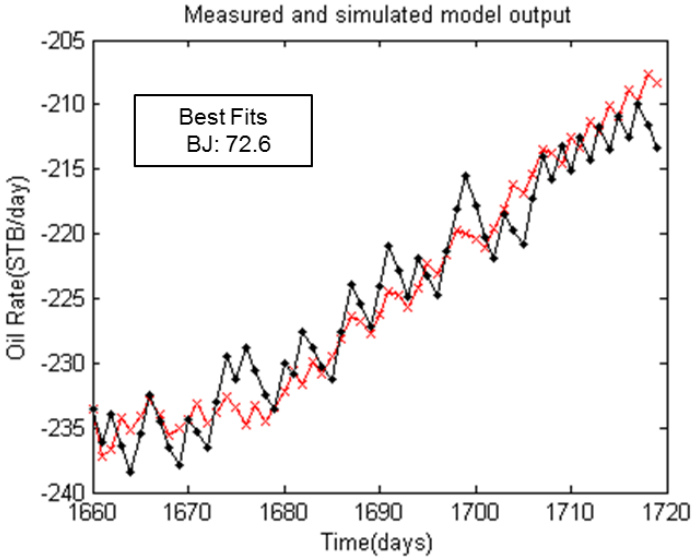


Figure 7.13: Permeability distribution of channel reservoir

The method for EOR here is water flooding. As a result, the water injection rate and production pressure will be used as input signals. The output signals are the water and oil production rates. The model is switched to a Multiple Input Single Output (MISO) model again. The types of the input signal are kept the same as in the previous section. The results are presented in Figure 7.14 - Figure 7.16. Once again, the models make accurate predictions for the production rate. In Figure 7.14 and Figure 7.15 we can see that the fit of the water production rate model is better than that of the oil rate model. That is because the reservoir model is a water-flooding model, and the water production rate is more affected by the input signal - water injection rate. The sine wave signal seems appropriate for both the water and oil production rate models.

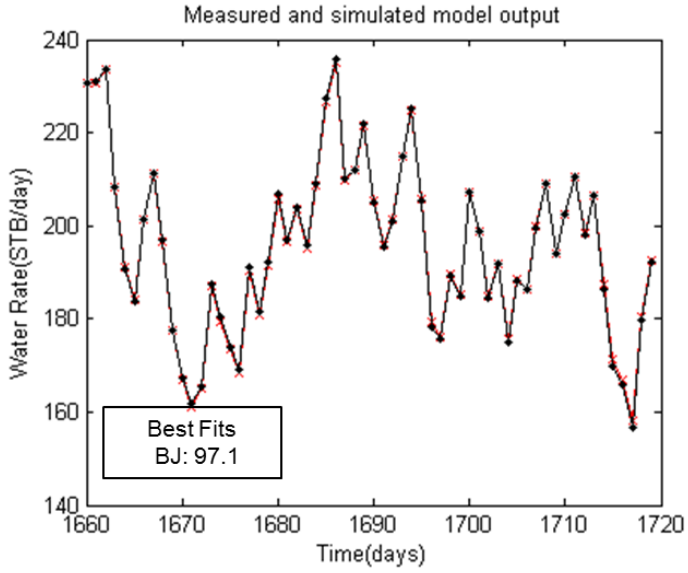


(a) Water production rate

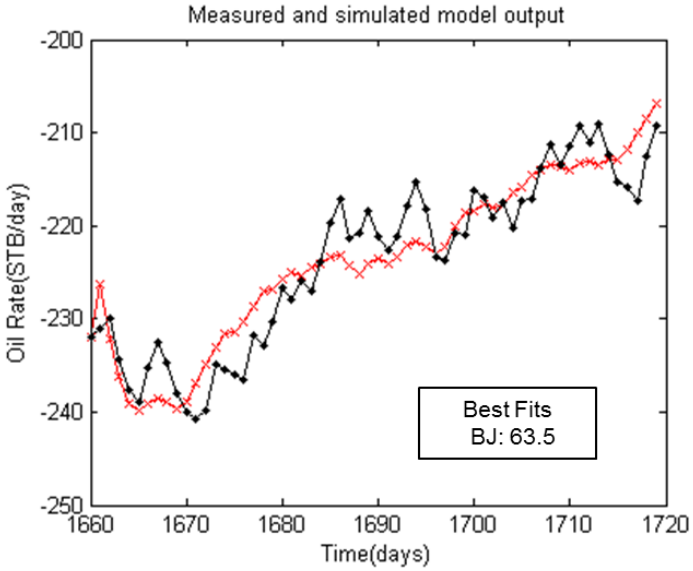


(b) Oil production rate

Figure 7.14: Open loop simulation of channel reservoir for PRBS input design

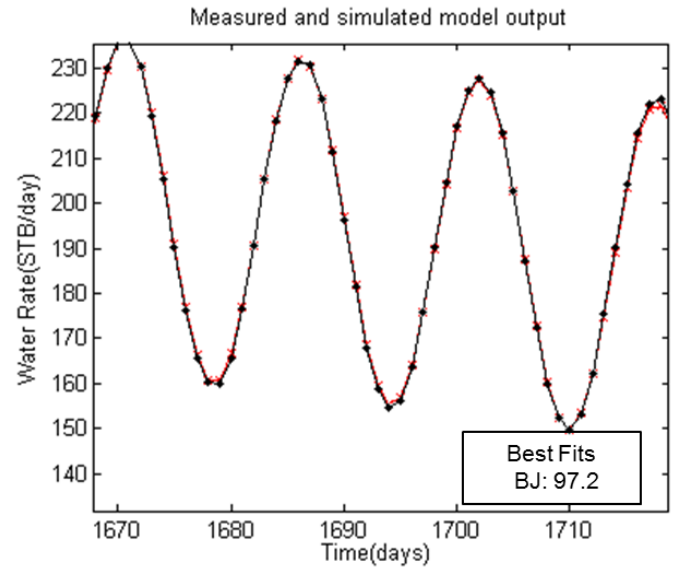


(a) Water production rate

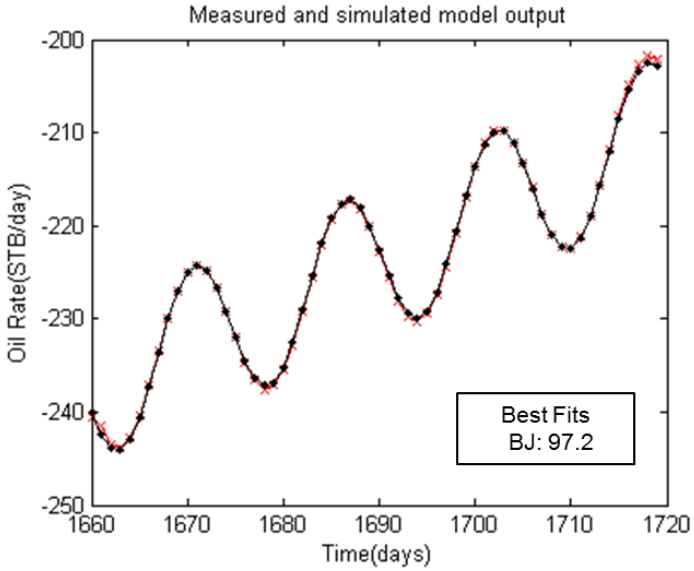


(b) Oil production rate

Figure 7.15: Open loop simulation of channel reservoir for multiple level PRBS input design



(a) Water production rate



(b) Oil production rate

Figure 7.16: Open loop simulation of channel reservoir for sine wave input design

7.4 Heterogeneous SAGD Reservoir

In Section 7.1 we discussed the homogeneous SAGD reservoir case. What if we change the homogeneous case to a heterogeneous one? In this section, we replicate the procedures in the reservoir model built in section 7.1 but change the permeability. The permeability ranges from 1982 mD to 4834 mD. The permeability distribution of one layer is shown in Figure 7.17.

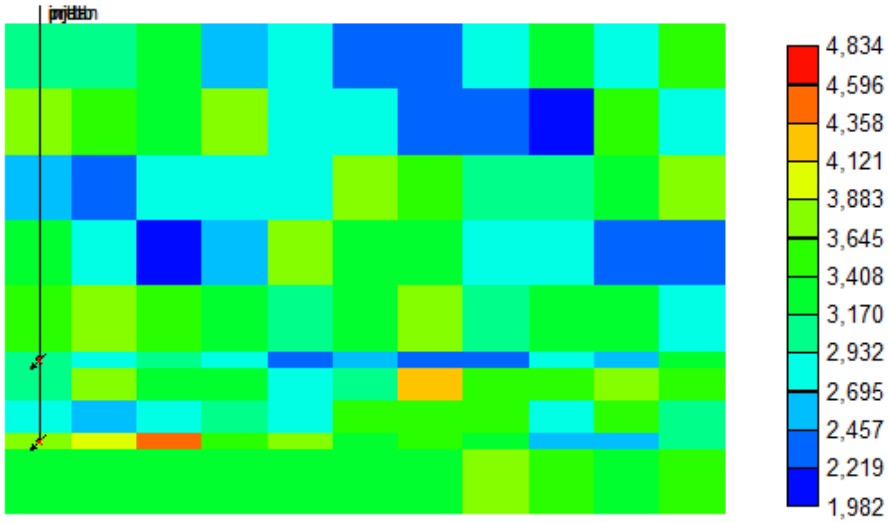
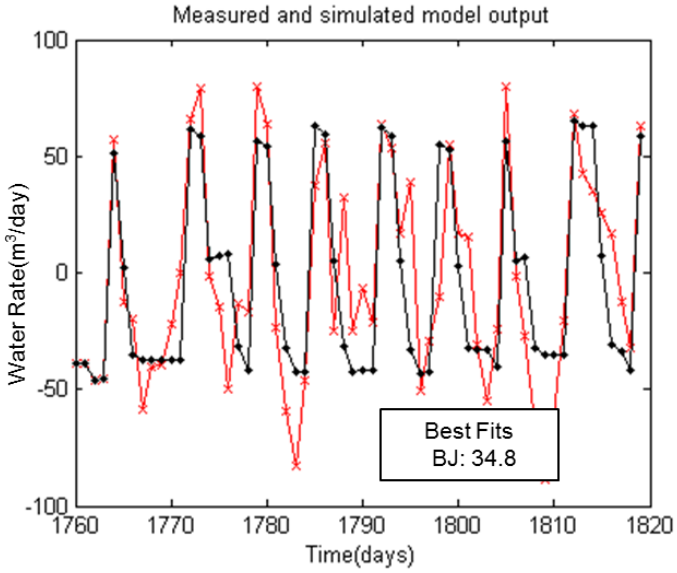


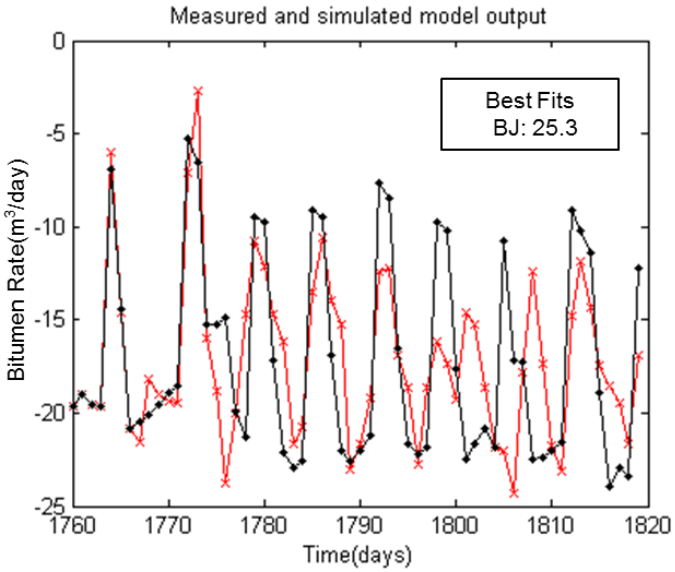
Figure 7.17: Permeability distribution of heterogeneous SAGD reservoir

After repeating the system identification procedure, the results are depicted in Figure 7.18 - Figure 7.20. We can see that the model fits decrease a lot although the model can still capture the main trends of the production profiles. Needless to say, random permeability acts as an important factor in system identification. System identification is a technique to find the relationship

between the input and output. The simpler the system is, the better results the system identification yields. From the figures we can also see that the production rate is more abrupt than that of other cases. But all in all, these are still good models and can direct the decision-making of production schedule.

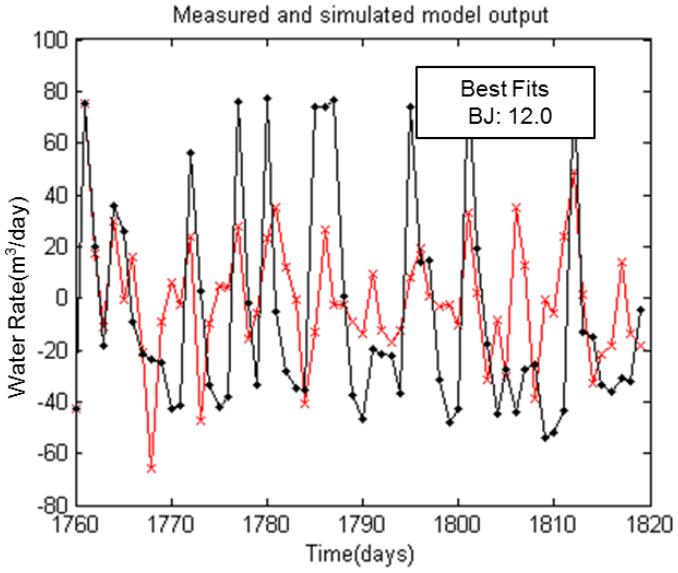


(a) Water production rate

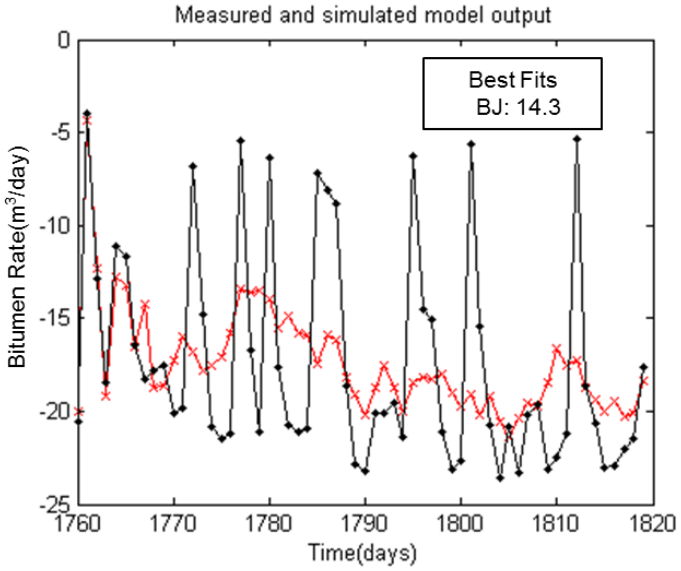


(b) Oil production rate

Figure 7.18: Open loop simulation of heterogeneous SAGD reservoir for PRBS input design

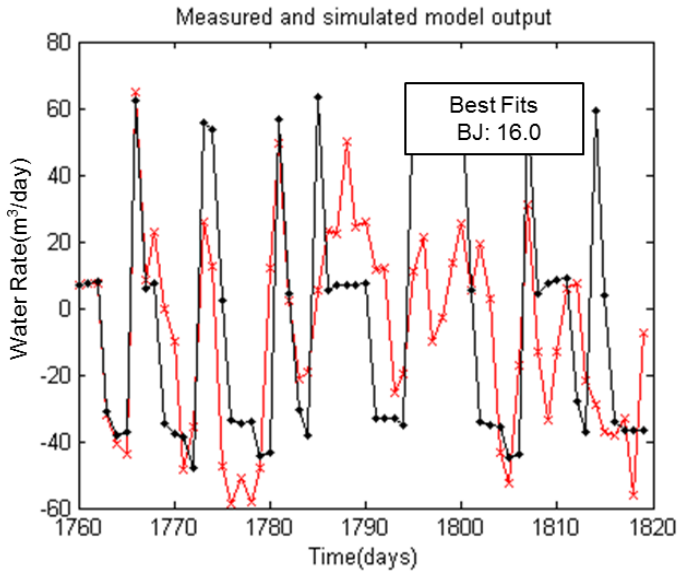


(a) Water production rate

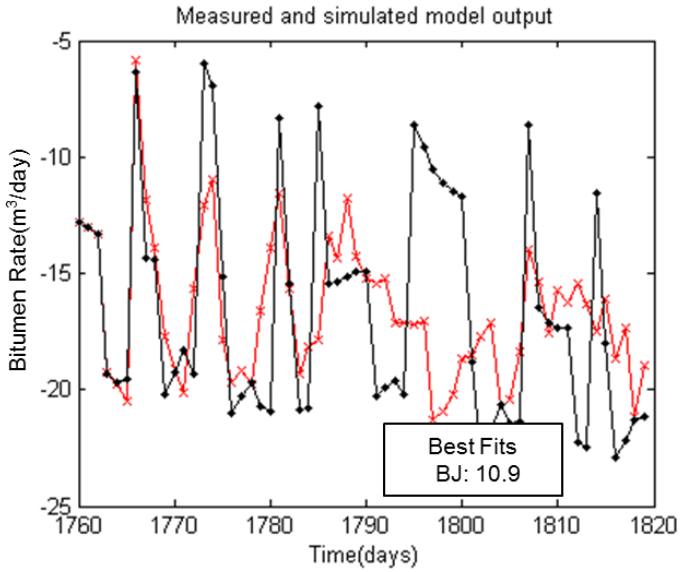


(b) Oil production rate

Figure 7.19: Open loop simulation of heterogeneous SAGD reservoir for multiple-level PRBS input design



(a) Water production rate



(b) Oil production rate

Figure 7.20: Open loop simulation of heterogeneous SAGD reservoir for sine wave input design

Chapter 8

Discussions, Conclusions and Future Work

8.1 Discussions

System identification techniques, which are widely used in process control, seem to have promising results in forecasting production profiles of various kinds of reservoirs. Compared to physics-based reservoir models, system identification models are easier to build and calculate. Traditional models can run for days; however, system identification models need just a few seconds. Advanced production measurement devices increase the measurement frequency to a quite high level, i.e., on the order of seconds to minutes. This efficiency promotes the development of system identification; the time step in most physics-based models takes on the order of days to months. As a result, before the high-frequency data are assimilated into these physics modes, they

are post-processed to match the simulator time step size. However, system identification can capture the fast dynamics such that short-term predictions are better, compared to those of physics-based reservoir models. In this thesis we applied system identification on real SAGD reservoir data and discussed how to build system identification models in detail. After accomplishing model construction for short-term prediction, we presented self-tuning models used for long-term prediction. To investigate what kinds of reservoirs and data are suitable for system identification, we designed different kinds of input for different kinds of reservoirs. The case studies shown in this thesis and their conclusions can be summarized as follows:

1. In the first case study we built a system identification proxy model using real data. The data used here is 15 years worth of SAGD operation data. Production pressure and the steam injection rate are used as the input signal, and the water production rate and bitumen production rate are used, separately, for the output signal. This creates two MISO models. The models built here are utilized for the two-month short-term prediction. The model structure are fixed as BJ model due to the error residuals between the model output and the real data. A time instant of when to start the effective identification is also specified to better reflect the reservoir dynamics. From this work we can say that in order to build a validated system identification model, the factors listed must be considered and a perfect combination of them should be chosen.

2. The second case study was also system identification on real SAGD reservoir data. However, in this case we tried to make a long-term prediction. To achieve this goal, we introduced recursive estimation, which can automatically update the system identification model every time step. In real data, there are big jumps which are difficult for non-recursive estimation to detect. Once the model is built, the coefficients of the models are fixed. Based on pure data, if there are significant changes, the system identification's performance is barely satisfied. With recursive estimation, new data can come into estimation, leading to updating coefficients to adjust to every situation. Therefore recursive estimation can not only predict for long time period, but can also lead to a better model fit.

3. In order to increase the prediction horizon, another seven-step recursive estimation case was built. As expected, the fit of the models decreased a little, but the results were still promising because the estimation case captured basically all the reservoir dynamics. After removing big jumps of the data, the performance was better. A non-linear model was also tested here. However, non-linear models are not main topics of this thesis.

4. To address the question of whether the system identification performance will be better if there aren't big jumps in the data, we designed three different kinds of input signals — the PRBS signal, multiple-level PRBS signal, and sine wave signal — for different kinds of reservoirs. The

reservoirs are a homogeneous SAGD reservoir, a naturally fractured gas reservoir, a water-flooding channel reservoir and a heterogeneous SAGD reservoir. The results showed that enough excitations in input are really important for system identification, and that system identification can be used for all kinds of reservoirs.

8.2 Conclusions

System identification using MATLAB System Identification Toolbox coupled with CMG STARS and Eclipse E100 simulators built a proxy reservoir model and forecasted the production profile. Different identification methods were compared to increase the model validity. The following conclusions can be drawn from this thesis:

1. System identification is a very efficient and robust technique for forecasting production of all kinds of reservoirs. The quality of system identification is determined by various factors such as input and output parameter selection, model structure selection, identification time period, time instant, and where to start the identification and excitation level of the input signal.

2. System identification can not only be used in short-term prediction but can also be used in long-term prediction once the recursive estimation technique is added. If the big jumps of data can be removed, the model fit will increase significantly.

3. Although most reservoir production problems are non-linear ones, linear system identification models can also be applied to capture the reservoir dynamics. Among all linear models, BJ models have the best fit due to their flexibility.

4. For short-term prediction, although the rule of thumb for the chosen experiment duration is that the length should be minimal give times the largest time constant. We demonstrated that too much data would also lead to a poor identification and a proper amount of data samples should be chosen to build the system identification model.

5. System identification of real data is restricted by the quality of the data. Once the model is built, it can only reflect the reservoir dynamics over a short period of time. However recursive estimation could update the model coefficients every time step to obtain better results.

6. Multiple step estimation will decrease the model fit but can still capture the reservoir dynamics, which is good news for control process.

7. With input design, the model fit of system identification increases because there is enough excitation in input signal. The SAGD process is harder to identify than other cases because it is a more sophisticated process. For easy case, the SISO model yields better results than that of MISO model.

8.3 Recommendations for Future Work

As for future work, the system identification can be applied to a history-matched model. Nonlinear model structures can be tested for different reservoir cases. System identification models are also important in dynamic real-time optimization. One could combine system identification with a model predictive controller to investigate production optimization.

Bibliography

[1] Awasthi A., Sankaran, S., Nikolaou, M., Saputelli, L. and Mijares, G.: "Closing the Gap Between Reservoir Simulation and Production Optimization," paper SPE107463 presented at the 2007 SPE Digital Energy Conference and Exhibition in Houston, USA, 11-12 April.

[2] Azad, A. and Chalaturnyk, R. J.: "Application of Analytical Proxy Models in Reservoir Estimation for SAGD Process: UTF-Project Case Study," Journal of Canadian Petroleum Technology (2013).

[3] Bempoard, A., Morari, M. and Ricker, N. L.: *Model Predictive Control Toolbox: Users Guide*, MathWorks, Natick (2004).

[4] Borgesen, J. F.: "Efficient optimization for Model Predictive Control in reservoir models," MSc dissertation, Norwegian University of Science and Technology, Trondheim (2009).

[5] Brouwer, D.R. and Jansen, J.D.: "Dynamic Optimization of Water flooding With Smart Wells Using Optimal Control Theory," SPEJ (December 2004) 391.

[6] Brouwer, D.R. et al.: "Recovery Increase Through Water Flooding With Smart Well Technology," paper SPE 68979 presented at the 2001 SPE European Formation Damage Conference, The Hague, 2122 May.

[7] Camacho, E.F & Bordons, C.: *Model predictive control*, Springer-Verlag London Limited (2004).

[8] Computer Modelling Group STARS Version 2012.10 User's Guide (2012). Computer Modelling Group, Calgary, Alberta, Canada.

[9] ECLIPSE, *Reference manual and technical description*. Schlumberger GeoQuest, 2009.

[10] Eker, S.A. and Nikolaou, M.: "Adaptive Control Through On-Line Optimization: The MPCI Paradigm," Appl. Math. And Comp. Sci. (1999) 9, No. 1, 101.

[11] Fedutenko, E., Yang, C. and Nghiem, L.: "Forecasting SAGD Process Under Geological Uncertainties Using Data-Driven Proxy Model, paper SPE1157942 presented at the 2012 SPE Heavy Oil Conference in Calgary, Canada, 12-14 June.

[12] Fishman, G. S.: *Monte Carlo: Concepts, Algorithms and Applicatoins*. Springer-Verlag, New York (1996).

[13] Genceli, H. and Nikolaou, M.: "A New Approach to Constrained Predictive Control with Simultaneous Model Identification, AIChE J. (1996) 42, No. 10, 2857.

[14] Ghasemi, M. and Whitson, C. H.: "Modeling Steam-Assisted Gravity Drainage with a Black-Oil Proxy, SPE Reservoir Evaluation & Engineering (2013).

[15] Jalali, J. Mohaghegh, Sh. D. and Gaskari, R.: "Coalbed Methane Reservoir Simulation and Uncertainty Analysis with Artificial Neural Networks," Transactions C: Chemistry and Chemical Engineering. (2010) 17, No. 1, pp. 65-76.

[16] Lee, E.B. and Markus, L.: *Foundations of Optimal Control Theory*, John Wiley and Sons, New York City (1967).

[17] Ljung, L.: *System Identification: Theory for the Users*, Prentice Hall, Englewood Cliffs, New Jersey (1987).

[18] Maciejovski, J.M.: *Predictive Control with Constraints*, Prentice Hall (2002).

[19] Mohajer, M., Damas, C., Silva, A. J. B. and Al-Kinani, A.: "An Integrated Framework for SAGD Real-Time Optimization," paper SPE128426 presented at the 2010 SPE Intelligent Energy Conference and Exhibition in Utrecht, Netherlands, 23-25 March.

[20] Nells, O.: *Nonlinear system identification: From Classical Approaches to Neural Networks and Fuzzy Models*, Springer, Berlin (2001).

[21] Nikolaou, M.: "Model Predictive Controllers: A Critical Synthesis of Recent Developments and Industrial Needs," Chemical Engineering Series, Academic Press, Burlington, Massachusetts (2001).

[22] Nikolaou, M.: "Model Predictive Controllers: A Critical Synthesis of Recent Developments and Industrial Needs," Chemical Engineering Series, Academic Press, Burlington, Massachusetts (2001).

[23] Ramirez, W.F.: *Application of Optimal Control Theory to Enhanced Oil Recovery*, Elsevier, New York City (1987).

[24] Rossiter, J.A.: *Model Based Predictive Control: A Practical Approach*, Control Series (1987).

[25] Saputelli, L., Nikonlaou, M. and Economides, M.J.: "Self-Learning Reservoir Management," Reservoir Evaluation & Engineering J. (2005).

[26] Schwarm, A., Eker, S.A., and Nikolaou, M.: "MPCI: A New Approach to Closed-Loop Identification and Adaptive Control," FOCAPO, Snowbird, Utah (1998).

[27] Schwarm, A. and Nikolaou, M.: "Chance Constrained Model Predictive Control," AIChE J. (1999) 45, No. 8, 1605.

[28] Slotte, P. A. and Smorgrav, E.: "Responese Surface Methodology Approach for History Matching and Uncertainty Assessment of Reservoir Simulation Models," paper SPE113390 presented at the 2008 SPE Eu-

ropec/EAGE Annual Conference and Exhibition in Rome, Italy, 9-12 June.

[29] Sudaryanto, B.: "Optimization of Displacement Efficiency of Oil Recovery in Porous Media Using Optimal Control Theory," PhD dissertation, U. of Southern California, Los Angeles (1999).

[30] Van Essen, G.M., Van den Hof, P.M.J., and Jansen, J.D.: "A Two Level Strategy to Realize Life-Cycle Production Optimization in an Operation Setting," paper SPE149736 presented at the 2012 SPE Intelligent Energy International in Utrecht, Netherlands, 27-29 March.

[31] Van Essen, G.M., Van den Hof, P.M.J., and Jansen, J.D.: "Hierarchical Long-Term and Short-Term Production Optimization," SPE Journal. (2011) pp. 191-199.

[32] Venegas, J. W., Deutsch, C. V. and Cunha, L. B.: "Uncertainty Assessment of SAGD Performance Using a Proxy Model Based on Butlers Theory," paper SPE115662 presented at the 2008 SPE Annual Technical Conference and Exhibition held in Denver, USA, 21-24 September.

[33] Wellstead, P.E. and Zarrop, M.B.: *Self-Tuning Systems: Control and Signal Precessing*, Wiley, Chichester (1996).

[34] Yeten, B., Castellini, A., Guyaguler, B., and Chen, W. H.: "A Comparison Study on Experimental Design and Response Surface Methodolo-

gies," paper SPE93347 presented at the 2005 SPE Reservoir Simulation Symposium in Houston, USA, 31 January 2 February.