Graduate Theses, Dissertations, and Problem Reports

2017

A Data-Driven Smart Proxy Model for a Comprehensive Reservoir Simulation

Faisal Alenezi

Follow this and additional works at: https://researchrepository.wvu.edu/etd

A DATA-DRIVEN SMART PROXY MODEL FOR A COMPREHENSIVE RESERVOIR SIMULATION

Faisal Alenezi

Dissertation submitted to the
Benjamin M. Statler College of Engineering and Mineral Resources
at West Virginia University

in partial fullfillment of the requirements for the degree of

Doctor of Philosophy
in
Petroleum and Natural Gas Engineering

Shahab D. Mohaghegh, Chair Ph.D.
Samuel Ameri, Professor
Ebrahim Fathi, Ph.D.
Ali T. Borujeni, Ph.D.
Dengliang Gao, Ph.D.

Department of Petroleum and Natural Gas Engineering

Morgantown, West Virginia
2017

Keywords: Smart Proxy, Reservoir Simulation, SACROC Field, Artificial Intelligence, Data Mining

ABSTRACT

A Data-Driven Smart Proxy Model for a Comprehensive Reservoir Simulation

Faisal Alenezi

The preferred common tool to estimate the performance of oil and gas fields under different production scenarios is numerical reservoir simulation. A comprehensive numerical reservoir model has tens of millions of grid blocks. The massive potential of existing numerical reservoir simulation models have gone unrealized because they are computationally expensive and time-consuming. Therefore, an effective alternative tool is required for fast and reliable decision making. To reduce the required computational time, proxy models have been developed. Traditional proxy models are either statistical or reduced-order models (ROM). They were developed to substitute complex numerical simulation with producing a representation of the system at a lower computational cost. However, there are shortcomings associated with these approaches when applied to complex systems.

In this study, a novel proxy-model approach is presented in order to overcome the computational size and the traditional proxy-model challenges. The smart proxy model presented is based on artificial intelligence and data-mining techniques. The objective of this study was to develop two types of smart proxy models at each grid block. The first smart proxy model was generated to identify dynamic reservoir properties (pressure and saturations). The other proxy model was created to determine the production profile of a well. The two smart proxy models can be coupled in order to examine field production performance under different operational and geological realization.

The field of study in this work is the SACROC unit. It is a depleted oil field located in Scurry County, Texas. The production history of this field began back in the late 1940s. Based on the long period of production and the different drive mechanisms employed throughout the fields exploitation, its performance history was divided into three phases in this study. Each phase was investigated and smart proxy models were applied to each.

To develop a smart proxy model, multiple reservoir simulation scenarios are designed for different operational constraints and geological realizations. The geological parameters along with the results from the designed simulation runs are collected to build the spatial-temporal database. The parameters in the database are studied and key performance indicators are measured to select the required data to build the smart proxy model. The smart proxy is trained, calibrated, and validated using a series of neural networks. To validate a smart proxy model, it is deployed to replicate a blind numerical simulation run.

The developed smart proxy models are capable of supplying reservoir properties along with production profiles very quickly (seconds) and with an acceptable range of error compared to numerical reservoir models.

*To my mother **Moneerah** for her love, encouragement, and prays day and night make me able to get such success. To my brother **Abdullah** for being the guardian and support for my sisters and brothers during my educational journey.*

Acknowledgments

In the name of Allah, the Most Gracious and the Most Merciful Alhamdulillah, all praises to Allah for the strengths and His blessing in completing this dissertation.

Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from many people.

First and Foremost, I would like to express my sincere gratitude to my advisor **Professor Shahab Mohaghegh** for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

My deep appreciation goes out to my other committee members, **Professor Sam Amiri**, **Dr. Ebrahim Fathi**, **Dr. Ali Takbiri**, and **Dr. Dengliang Gao**. Thank you for investing time and providing interesting and valuable feedback. I feel proud and honored that you have accepted to be on my committee.

I also want to take a moment to thank the Laboratory for Engineering Application of Data Science (LEADS) graduate team members: Amir Ansari, Maher Alabooodi, and Sina Hosseini Boosari for their endless support, timely motivation, sympathetic attitude and unfailing help during the course of entire study. I am also grateful to them for their friendship and the warmth they extended to me during my time in the lab.

I am grateful to the Department of Petroleum and Natural Engineering faculty and staff who have been kind enough to advise and help in their respective roles. Also, I want to extend my appreciation to all my friends that I have known in the department, thank you for the great times that we have shared. I am also grateful to all my best friends in Morgantown. Your love and care are highly appreciated.

I want to acknowledge Computing Modeling Group (CMG) and Intelligent Solutions Inc. (ISI)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Return on Investments in Reservoir Simulation

The petroleum industry strives to find oil and gas reserves, develop these resources, meet the worlds energy demand, and maximize profits. One of the most important tools for oil and gas reservoir development and management is reservoir simulation. It is a necessary tool for reservoir engineering strategy plans. The key goal of reservoir simulation is to predict future performance of the reservoir and to find ways and means of optimizing the recovery of some of the hydrocarbons under different operating conditions. The simulation involves four main interrelated modeling stages, establishment of physical models, development of mathematical models, discretization of these models, and design of computer algorithms. It requires a combination of skills of physicists, mathematicians, reservoir engineers, and computer scientists[1].

With the recent advances in computers and simulation software capabilities, the development of algorithms, and the amount of data that can be collected from the reservoir, it is now possible to build a comprehensive reservoir simulation model. On the other hand, these models require very sophisticated performance platforms to run. As the size of the reservoir models grows, the time required to run different scenarios also increases. Therefore, it is impractical to use a comprehensive simulation model on a routine basis.

Oil and gas organizations invest heavily in reservoir simulation, often spending millions of dollars. Every organization seeks to maximize the value of the capital, so good management means

evaluating the return on this investment. Investment decisions are made in uncertain circumstances and require a proper cost analysis. The time spent to develop the reservoir simulation models, the cost of the platforms used, and the labor involved are the main parameters for calculating the return on investment in reservoir simulation.

1.2 Problem Statement and Research Objective

The reservoir engineers main duty is to take critical decisions for the oil and gas reservoirs future strategies. Drilling new wells, shutting-in some of the production wells, adding injection wells, and many other development plans should be based on intelligent decisions. Reservoir engineers usually take these critical decisions based on numerical simulation models, in which it is possible to study all possible production and injection scenarios. Using a comprehensive numerical model, these different scenarios could lead to a better development plan.

Unfortunately, to build a comprehensive numerical model, hundreds and sometimes thousands of simulation runs should be conducted, which is computationally intensive and time consuming. Moreover, with a new drive towards smart fields (smart completions and smart wells) in the petroleum industry that, the need to process the data in real time has become more prominent. Therefore, reservoir engineers are looking for ways to reduce the computational size of the simulation models, keeping the model as complex as possible with acceptable output accuracy.

Efforts have been made to reduce the computational size of the model by using proxy models. Proxy models act as a substitute for the complex numerical simulation, since they produce a meaningful representation of the complex system using fewer simulation runs.

In petroleum simulation models, engineers and scientists have developed different types of proxy models based on different techniques. These proxy models are generally based on statistical or mathematical approaches. The main proxy models can be cited as:

- Response Surface Proxy Models

- Reduced Physics Proxy Models

- Reduced Order Proxy Models

- Data Driven Models

Although these methods work well in many cases, there are some drawbacks and limitations, which are discussed in the following sections. The current dissertation develops a smart proxy model

using a state-of-the-art data-driven technique. The smart proxy model must be able to produce comprehensive reservoir properties (pressure and saturation) along with the wells production profiles in a significantly shorter simulation time compared to the numerical reservoir models.

1.3   Dissertation Outline

This dissertation is divided into nine chapters and they are organized as follows:

Chapter 1 introduces the problem statement and the objective of this study.

Chapter 2 is literature review. In this chapter, the related topics to this study are reviewed. The data mining and machine learning techniques are described. The concept of proxy model and the general types of smart proxy models is also briefed.

Chapter 3 Starts with a background on the field of study and describes the details of the generated numerical reservoir model.

Chapter 4 Elaborates on the procedure of developing a smart proxy model. It shows the general work flow of this study.

Chapter 5 presents the development of the smart proxy model to replicate the numerical simulator results. It includes the steps of training and validating the smart proxy developed for this goal.

Chapter 6 discusses the development of the smart proxy for geological and operational reservoir field realization. In this chapter, the ability of the smart proxy model to perform geological and operational realization is proofed. However, this chapter covers the non-cascading feature only.

Chapter 7 extends the work on developing the smart proxy for geological and operational realization to include the cascading feature.

Chapter 8 covers the development of smart proxy for production profiles. In this chapter, details of generating smart proxy for oil rate, gas rate, and water rate are presented along with deployment results.

Chapter 9 summarizes the research on form of concluding remarks. Also, recommendations for future work are made.

Chapter 2

Literature Review

2.1 Machine learning

Over the past three decades, machine learning has become one of the pillars of information technology. With ever-increasing amounts of data, the world is entering the era of big data. Due to advances in computer technology, it is possible to store and process large amounts of data and to access it from physically distant locations over a computer network. There is therefore a good reason to call for automated methods of data analysis, which is what machine learning provides [2].

In particular, machine learning can be defined as a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision-making under uncertainty. Mitchell [3] outlined the main elements of machine learning by defining machine learning as follows A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. For difficult tasks, programs can be written and designed to carry them out. In machine learning, tasks are how to process an example using machine learning. The performance measurement is simply the way machine-learning technique is evaluated.

The fundamental goal of machine learning is to generalize beyond the examples in the training set. This is because no matter how much data there is available, it is very unlikely that those exact examples will reoccur at test time [4]. In traditional programming, the data and the program are

run on the computer to produce the output. In machine learning, the data and the output are run on the computer to create a program.
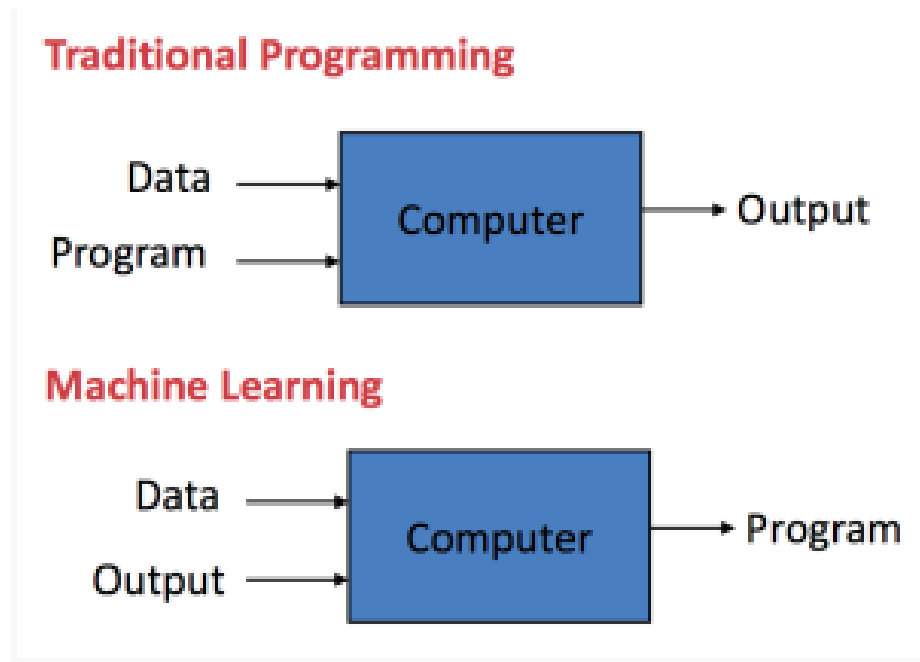
Figure 2.1: Traditional Programming and Machine Learning

There are three main types of machine learning, supervised, unsupervised, and reinforcement learning. The supervised approach is also known as the predictive approach; the objective of the machine learning is to learn the relationship between inputs to outputs, given a labeled set of input-output pairs. The inputs and outputs are called features and responses. The second main type of machine learning is the descriptive or unsupervised learning approach. There are only inputs in this type of machine learning. In this type of learning, the machine learns how to find patterns in the given data. Reinforcement learning is used to solve problems of decision making (usually a sequence of decisions) [5].

The machine learning model is constructed from three elements [6]:

- Representation: In supervised learning, a model must be represented as a conditional probability distribution (usually called classifier) a decision function. The set of classifiers is called the hypothesis space of the model. Choosing a representation for a model is equivalent to choosing the hypothesis space that it can possibly learn.

- Evaluation: In the hypothesis space, an evaluation function is needed to distinguish good

classifiers from bad ones.

- Optimization: Finally, we need a training algorithm (also called learning algorithm) to search among the classifiers in the hypothesis space for the highest-scoring one.

Most recently, machine learning techniques are being widely used to solve real-world problems by storing, manipulating, extracting and retrieving data from large sources. This success is also apparent in petroleum engineering applications as it will be shown in the following sections.
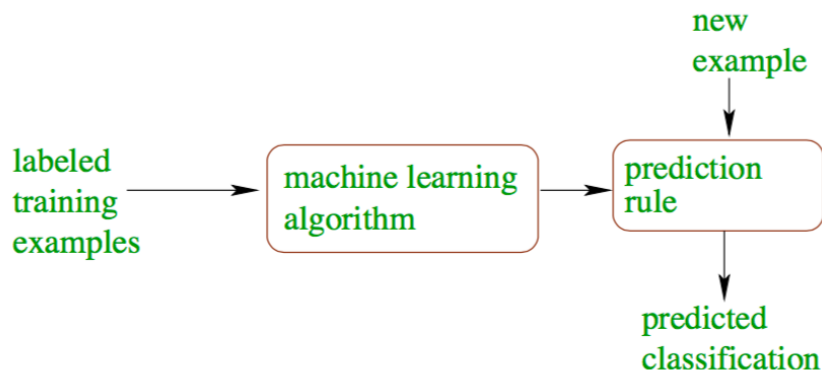
Figure 2.2: Typical Structure of Machine Learning Model

2.1.1 Artificial Neural Networks

Artificial neural networks (ANN) are biologically inspired methods of computing and thought to be the next major advancement in the computing industry. They are self-learning mechanisms which do not require the traditional skills of a programmer. Artificial neural networks are relatively crude electronic models based on the neural structure of the brain. These artificial neurons are biologically inspired. The brain learns from experience.

The best and simple definition of the artificial neural networks is provided by Nielsen [7]:

"A computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs."

The basic part of the human brain is the cell. It provides humans with the ability to remember, think, and use previous experience for future decisions [8].The human brain includes 100 billion cells,

7

which are known as neurons. In addition to that, there are up to 200,000 connections between these neurons.
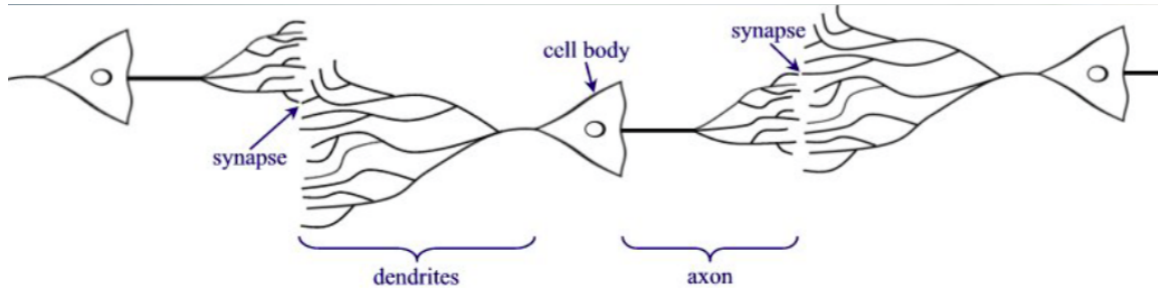
Figure 2.3: Biological Neuron
[9]

The history of ANN goes back several decades. In 1943, a neurophysiologist, Warren McCulloch, and a mathematician, Walter Pitts, developed a neural network with electrical circuits called threshold logic. The first book on neural networks was written in 1949 by Donald Hebb. Hebb built a learning mechanism based on neural plasticity. With the advances in computers in the 1950s, IBM researcher Nathanial Rochester started the early efforts to simulate a neural network. Dartmouth Summer Research worked on artificial intelligence in 1956. They developed a process to stimulate research in artificial intelligence. In 1959, Bernard Widrow and Marcian Hoff developed the first neural network to be applied to a real world problem. After that, however, due to the limitations of computer capabilities available at the time, ANN research grew only slowly, until 1982.

In 1982, ANN gained renewed interest of researchers. A conference on neural networks was held in Kyoto, Japan, in 1982. The outcomes of this conference brought new funding for neural networks research.

Currently, neural networks are used in several applications. The fundamental idea behind the nature of neural networks is that if it works in nature, it has to be able to work in computers. The future of neural networks, though, lies in the development of hardware.

Biologically, neurons receive information known as inputs from another source, connect these inputs in some way, and finally perform a nonlinear operation to generate outputs. The process is summarized in following steps:

- A signal reaches a synapse: Certain chemicals called neurotransmitters are released.

- Process of learning: The synapse effectiveness can be adjusted by signal passing through.

Information are not transmitted directly, but stored in interconnections. The term Connections model initiated from this idea.

From biological neuron to artificial neuron, the model elements can be defined as following:

Human neuron   **The model neuron**
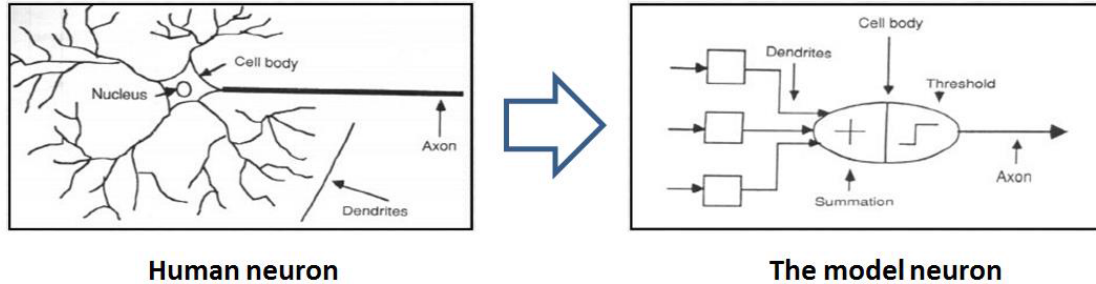
Figure 2.4: From Human neuron to Artificial neuron

- Wires : axon and dendrites

- Connection weights: Synapse

- Threshold function: activity in cell body

The Model of the Artificial Neural Network

The ANN is represented by a mathematical model. In the network, the number of inputs is modeled using the symbol x. Each of these inputs is multiplied by a connection weight w, and the product of this operation is summed. After that, the summation is fed through a transfer function to generate the outputs. The information processing of a neuron can be divided into three basic elements:

1. A group of connecting links. The main characterization of this set is how strength is associated with each link. The strength is also known as synaptic weight w. Usually the weight is written with two subscripts k and j. The neuron is represented by k, and k is reverse to the input end of the synapse. The term wkj is either positive or negative.

2. The summation of the inputs (signals).

3. An activation function. The goal of this function is to limit the amplitude of the neuron output.

The neuron model can be described mathematically with the following equation [10]:

$$y = F(\sum_{i=0}^{m} w_i * x_i + b) \tag{2.1}$$
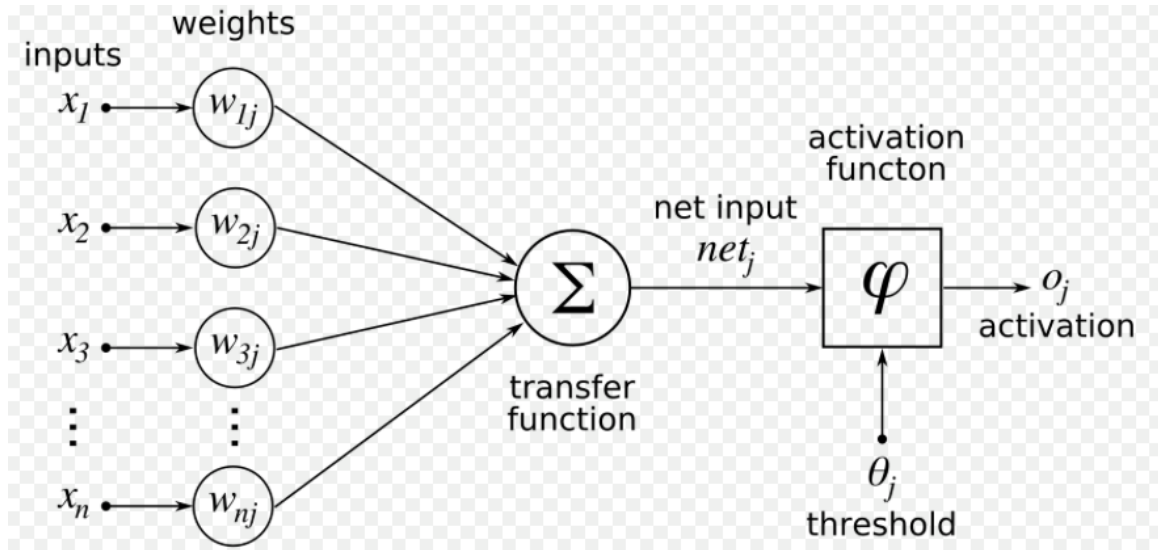
Figure 2.5: Artificial Neural Network

Where:

x_i is input value,

w_i is the weight value,

b is bias,

F is a transfer function,

y_i is the output value

The activation function defines the relation between the inputs and the outputs. Figure 2.6 shows different types of activation functions.

Artificial Neural Network Architectures

Based on the type of the training algorithm, the ANN architecture is build. Generally, the most popular ANN architecture used is the multilayer feed-forward networks. In this architecture, the neurons are organized in the form of layers; input, hidden, and output. A layer of input units is connected to a layer of hidden units, which is connected to a layer of output units.

The function of these three layers is defined as the following:

- Input layer: Represents the raw information that is fed into the network.

- Hidden layer: Is determined by the activities of the input units and the weights on the con-nections between the input and the hidden units.

- Output layer: Depends on the activity of the hidden units and the weights between the hidden

10

Name	Input/Output Relation	Icon
Hard Limit	$a = 0 \quad n < 0$ $a = 1 \quad n \geq 0$	
Symmetrical Hard Limit	$a = -1 \quad n < 0$ $a = +1 \quad n \geq 0$	
Linear	$a = n$	
Saturating Linear	$a = 0 \quad n < 0$ $a = n \quad 0 \leq n \leq 1$ $a = 1 \quad n > 1$	
Symmetric Saturating Linear	$a = -1 \quad n < -1$ $a = n \quad -1 \leq n \leq 1$ $a = 1 \quad n > 1$	
Log-Sigmoid	$a = \dfrac{1}{1 + e^{-n}}$	
Hyperbolic Tangent Sigmoid	$a = \dfrac{e^{n} - e^{-n}}{e^{n} + e^{-n}}$	
Positive Linear	$a = 0 \quad n < 0$ $a = n \quad 0 \leq n$	

Figure 2.6: Activation Functions of Artificial Neural Network
[11]
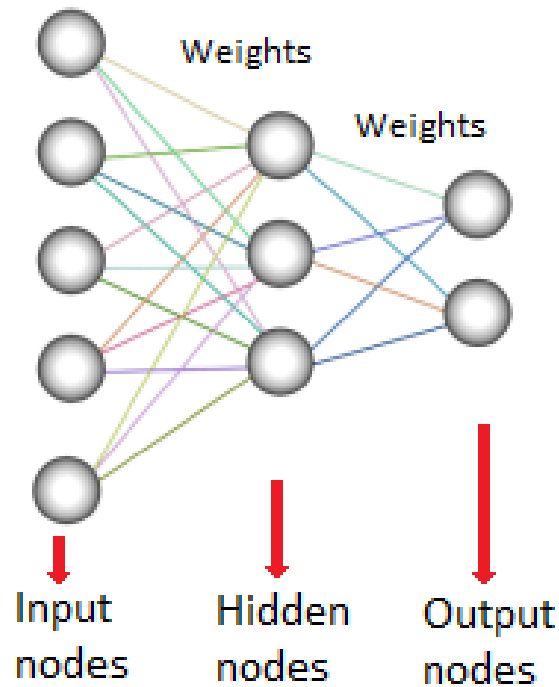
Figure 2.7: Multilayer ANN Architecture
[12]

and output units.

Back-propagation Algorithm

A widely used algorithm with multilayer ANN is back-propagation. It has been successfully applied to solve problems in different industries. It uses the error correction technique. The process of this algorithm has two main phases. First, in the forward passes, the inputs are fed into the nodes of the network and they propagate from layer to layer until the output layer has been reached with a network response. In the forward pass, the weights are calculated. Second, in the backward pass, errors between neuron output and target output are measured and the weights are adjusted accordingly. Moreover, the error signal is propagated backward through the network. This process continues until a minimum amount of errors has been achieved.

The algorithm can be summarized in the following four steps:

1. Feed-forward computation

2. Back propagation to the output layer

3. Back propagation to the hidden layer

4. Weight updates

The above algorithm steps are repeated until an acceptable minimum number of errors has been achieved. The error is defined as the difference between the actual activation of an output and the desired activation. The total error is the sum of these errors and it is the measurement tool for neural network performance.

$$E = (1/2) \sum_{j=1}^{j} (target_i - actual_j)^2 \qquad (2.2)$$

Function signals

Error signals
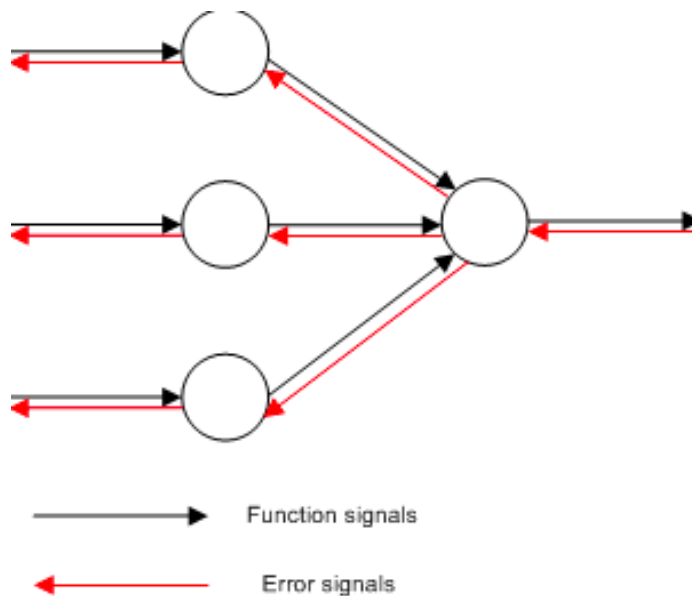
Figure 2.8: Back-propagation Algorithm

2.1.2 Data Mining

The application of machine learning methods to large databases is called data mining. It is a process involving analyzing and using the data for predictive modeling.

Data mining is about performing data analysis to explore and discover significant patterns in the database. Because the data mining is performed on large quantities of data, the process needs to be automatic or semi-automatic. In order to ensure that the data is suitable for automated processing,

the raw data is collected, cleaned, and transformed to a standardized format [13]. It is well known that the biggest part of the work on data mining is data preparation. The typical work flow of data mining is shown in Figure 2.9.
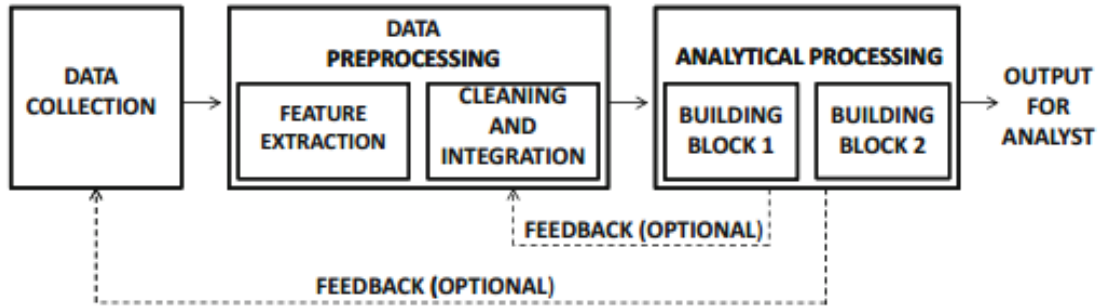
Figure 2.9: Data Mining Work flow
[13]

There are different activities can be generated using data mining, a short definition will be given on these activities in the following paragraphs [14].

Classification and Estimation

classification is performed by a well-defined classes. Moreover, classification is a process to examine features in a newly presented database. This process involves building a model that can used for classification. classification is dealing with discrete outcomes. On other hand, estimation is dealing with continuous outcomes.

Clustering

Clustering is a process of segmenting a set of data (features) into a meaningful subgroups.Clustering is different from classification is that there are no predefined examples.

Visualization

From the Oxford dictionary, data visualization is defined as:

"The representation of information in the form of a chart, diagram, picture, etc."

Data visualization is a powerful data mining technique. A good representation of the data will provide a good start for data analysis. Therefore, it is very important to choose the right visualization tool in order to obtain enough and useful information from the database.

2.1.3 Machine Learning in Petroleum Engineering

The petroleum industry has been interested in machine learning since the 1990s. A quick search in the database of the Society of Petroleum Engineers (SPE) turns up hundreds of papers about this subject. These papers show a successful implementation of machine learning techniques in both; upstream operations and downstream operations. Moreover, like other industries, the petroleum industry is challenged by an increasing volume of data collected every day.

In a series of articles in the SPEs journal, Mohaghegh has presented three types of virtual intelligence (neural networks, genetic algorithms, and fuzzy logic) and their applications in the oil and gas industry [15, 16, 17] . These articles show the ability and the potential of artificial intelligence to solve complex problems in petroleum reservoir engineering.

As part of a study on recovery efficiency, Brown used machine learning techniques to extract valuable information from a database. He concluded that through machine learning, objective predictions can be achieved [18]. Al-Baiyat and Heinze implemented ANN in stuck pipe prediction. They show that machine learning techniques can predict stuck pipe with acceptable accuracy [19]. Anifowose reviewed machine learning and its successful application in reservoir characterization [20]. Subrahmanya et al. explored machine learning methods for production pattern recognition. They reviewed the effectiveness of these methods in a real-world asset management scenario. They also explored the use of active learning to intelligently identify data points with the highest value of information and semi-supervised learning to combine the information from labeled and unlabeled sources [21]. In 2006, Mohaghegh introduced an alternative reservoir simulation tool that is based on data mining and artificial intelligence. In the early work, the developed smart proxy primarily dealt with the production profile at the well level. Since then, the smart proxy model has been applied to different reservoir simulation studies for obtaining dynamic reservoir properties at the grid block level and to perform history matching [22].

Tian and Ronald studied machine learning algorithms to interpret flow rate, pressure, and temperature from downhole permanent gauges. They showed the potential of using machine learning to interpret the data recorded by permanent downhole gauges with complex physical models [23]. Shadravan et al. applied machine learning algorithms to fracturing and cementing operations. They

used ANN models to reduce the cost by eliminating personnel supervision. Moreover, the developed machine learning model offers comprehensive insights for fluid design [24]. Shadravan and Tarrahi discussed the benefits of using machine learning algorithms for fluid design. With intelligent fluid design, thousands of hours of laboratory testing are saved. The developed prediction model is able to successfully extract and relate input parameters to output variables [25]. Cao et al. introduced machine learning algorithms to forecast production for existing and new wells in unconventional assets using inputs (e.g. geological maps, production history, pressure data, and operational constraints). The proposed technique requires less work for model building. [26].

The aforementioned work on the use of machine learning is an example of its successful application in the field of petroleum engineering. Machine learning is now attracting more attention from petroleum engineers, especially since the inception of the intelligent field era.

2.2   Proxy Models

An accurate reservoir simulation involves a comprehensive description of the reservoirs properties. Until recently, computational science, discovering numerical solutions to complex, multi-physic, non-linear, and partial differential equations, has been at the leading edge of engineering problem solving and optimization [27]. . Lately and with the progress of technology and software capabilities, the developed models for simulating different processes in different science and engineering areas have become more complex and computationally costly to complete. Often, these complex models need to be performed hundreds or thousands of times, which requires a large amount of time and computational effort.

To resolve this challenge, many efforts have been made to develop proxy models that can be used as a substitute for a complex reservoir simulation model. Proxy-modeling (also known as surrogate modeling) is a computationally cheap alternative to full numerical simulation in assisted history matching, production optimization and forecasting. A proxy model is defined as a mathematically, statistically, or data-driven defined function that replicates the simulation models output for selected input parameters [28]. The proxy models results cannot mimic the numerical simulation results with 100% accuracy, but within the amount of time required to run these models, the outputs generated have a very acceptable range of error. Reducing the computational time from hours and sometimes days to seconds makes these models significantly competent and attractive to the reservoir engineers

[29].

During the last decades, different techniques of the proxy models have been developed and introduced to the petroleum industry. The computational techniques of these proxy models range from the statistical approach (e.g. response surface) and the mathematical approach (e.g. reduced order models), to data mining and the artificial intelligence approach (e.g. smart proxy models).

The following sections will discuss these techniques in more details. In addition to that, their abilities and limitation will be considered.

2.2.1 Response Surface Methodology

The modeling of experimental responses was developed by Box and Draper in 1987 [30]. A response surface, also called a proxy model, is a representation of a system simulation. This type of proxy modeling is very useful when dealing with a large number of complex system parameters. Response surface methodology (RSM) consists of a group of mathematical and statistical techniques used in the development of an sufficient functional relationship between a response of interest, y, and a number of associated input variables denoted by X_1, X_2, .., X_k [31]. This type of proxy modeling is very useful when dealing with huge number of complex system parameters [32]. In a reservoir simulation and based on the number of selected runs, a response surface is generated that can be used predict other input parameters. The response surface technique is used to find the relationship between reservoir parameters and the simulation responses (production or pressure). To find the optimum response, responses generated from a high number of simulation runs are collected to evaluate the relation between the input and output variables.

To show a simple example of response surface methodology, the liquid flow rate depends on many variables, the reservoir pressure, X_1, and the bottom hole pressure, X_2 are the main parameters. The flow rate can vary under any combination of the values of these parameters. Therefore, reservoir pressure and bottom hole pressure can vary continuously. The relationship between the flow rate, reservoir pressure, and bottom hole pressure can be generated using the following form of equation:

$$y = f'(x_1, x_2)\beta + \epsilon \tag{2.3}$$

Where ϵ is the experimental error term and β is a vector of unknown constant coefficient referred to as parameters [33].

There are two main models in response surface methodology, the first order which can be expressed

using the following equation:

$$y = \sum_{i=1}^{k} \beta_i x_i + \epsilon \tag{2.4}$$

And the second degree:

$$y = \sum_{i=1}^{k} \beta_i x_i + \epsilon + \sum_{i<j} \sum \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \epsilon \tag{2.5}$$

In order to predict the desired response values, it is necessary to develop a relationship between the input parameters and to construct the optimum setting of these input parameters. To establish this relationship, a number of experiments are conducted and in every experiment the response is calculated under specified ranges of the control variables [31].

Response surface methodology is a method that describes the different settings of input variables in a series of simulation runs [34]. The goal of this method is to create these settings in order to get the maximum information in a reasonable number of simulation runs.

The response surface methodology has a wide range of applications, commonly in Industrial, Biological and Clinical Science, and in Engineering Science.

Over the past few years, the literature in the petroleum industry has shown how the response surface methodology can be applied to reduce the reservoir simulations complexity. Ahmadloo [35] has developed a simple approach for the prediction of medium to heavy oil viscosity by using Response Surface Methodology. Beceera (2012)) work combines history matching with a probabilistic analysis of a number of scenarios to minimize the range of uncertainty in the input parameters. Dejea [36]has presented a procedure using response surface methodology to manage the uncertainties in production prediction. Purwar [37]has built a workflow for the optimization of the gas flood using response surface methodology. Rezo (2008) study proposed the use of response surface methodology to secure a reliable initial guess for nonlinear inversion and for understanding the separate contributions of the various measurements of the formation tester to specific inversion parameters. Risso [38] concluded that response surface methods are a good alternative to risk assessments in oil fields. Silva [39] has proposed a methodology using surface response to eliminate non-significant parameters, which reduces the computational effort and the time spent by the engineer in the oil fields developments. Vecente (2010) presented an example of how static and dynamic uncertainties were combined using a scenario-based approach and surface response methodology. He has developed

models on the variables that control the recovery: reservoir quality, distribution and connectivity, porosity, permeability and fluid contacts in order to best assist the oil fields development plans.

All things considered, there are some limits of using surface response modeling. If more than one response needs to be designed, more response surface models are required. Furthermore, with complex systems, a large number of simulation runs are required in order to perform statistical analysis. Another limitation is the underlying assumption that the response varies smoothly depending on the parameters. Also, in the petroleum industry, companies are dealing with discrete parameters, such as faults, different depositional environments, and other geological features. The response surface technique, on the other hand, is most suitable to continuous parameters.

2.2.2 Reduced Physics Models

To interpret reservoir behavior, a simulation model is built using different physical scenarios. If more physical parameters are introduced to the model, the measured results are more reliable, but the model will be more complex.

One way to reduce the complexity and computational time of the full reservoir model is by simplifying the model physics. The reduced physics method works by reducing the number of physical properties or grid blocks. The reduced physics parameters can be determined using the tuning procedure, keeping the full physics model as the reference solution. A certain amount of numerical experimentation must be performed to determine the level of physics and grid resolution required for the reduced physics model.

Wilson and Durlofsky [40]have presented a general workflow for applying optimization to the development of shale gas reservoirs using the reduced physics approach. They started with a full physics simulation model with dual porosity, dual permeability, and gas desorption. Then, a much simpler reduced physics surrogate model, with single porosity and single permeability, was developed. The computational time to run the simulation model is far less.

The main theoretical premise behind the reduced physics model is the use of the tuning procedure. The tuning procedure is performed by matching the production data (such as gas production) of the full physics model. The comprehensive model is run for many reservoir scenarios based on the field development plans and the parameters are calculated for the reduced physics model. The surrogate model is confirmed when the error between the full physics and the reduced physics model is minimized over the designed reservoir scenarios.

In conclusion, the available studies on the reduced physics models show the efficiency of this approach, but the application of the approach to the petroleum industry is limited. The issue of whether the reduced physics model is applicable for a complex reservoir is clouded by the fact that this approach has never been applied to a comprehensive, three-dimensional reservoir.

2.2.3 Reduced Order Models

There is an increasing amount of literature on reduced order modeling approaches in petroleum reservoir engineering. Reduced order modeling is a technique in which the full, high-dimensional models are replaced with models that are smaller in dimension. The reduced model still characterizes the main features of the system dynamics. The behavior of high-order reservoir models is generally determined by only a part of the information it contains. Therefore low-order models are often satisfactorily accurate for describing reservoir dynamics [41].

Most subsurface techniques for high-order reduction are Krylov subspace, balanced truncation, and proper orthogonal decomposition. Proper orthogonal decomposition is the most commonly used method for reduced order models in reservoir engineering.

Reduced order models have been applied to reservoir simulation with acceptable results. However, there are limitations to the application of this approach. Because of the nature of the developed reduced order model, the time needed to develop this type of model is sometimes close to the time needed for the numerical simulation run. Therefore, reduced order models are sometimes computationally expensive.

2.2.4 Data Driven Models

The amount of data in the world is increasing dramatically. Data mining is about solving problems by analyzing and discovering the patterns already present in databases. Artificial Intelligence is a powerful technique that teaches the machines how to process data [42].

As the name suggests, Data Driven Modeling is focused on analyzing the total data within a system and find a connection between the system variables (inputs and outputs).

The main fields of the DDM are:

- artificial intelligence (AI), which is the overarching study of how human intelligence can be incorporated into computers.

- computational intelligence (CI), which includes neural networks, fuzzy systems and evolutionary computing as well as other areas within AI and machine learning.

- soft computing (SC), which is close to CI, but with special emphasis on fuzzy rule-based systems induced from data.

- machine learning (ML), which was once a sub-area of AI that concentrates on the theoretical foundations used by CI and SC.

- data mining (DM) and knowledge discovery in databases (KDD) are focused often at very large databases and are associated with applications in banking, financial services and customer resources management. DM is seen as a part of a wider KDD. Methods used are mainly from statistics and ML.

Data-driven modeling techniques can be used to determine the relationship between systems inputs and outputs using a training set that represents the systems behavior. Once the relationship is generated through the training set, another data set that is not included in the training process is used to test the ability of the model for predication (generalization) [43].

2.3   Smart Proxy Models

The main goal of developing proxy models is to reduce the computational size of complex reservoir models. The reduced order models and reduced physics models simplify the original problem as discussed in the previous sections. Recently, real-time measurements using intelligent fields have produced a large amount of at short intervals. The reservoir models then need to be updated accordingly. On the other hand, there is an increasing need for a simulation technique that can perform the tasks in a shorter period of time. The smart proxy model is introduced as a technique for fast reservoir simulation with acceptable accuracy. This introduced proxy model has the ability to preserve the complexity of the reservoir model.

The numerical reservoir simulation is still the leading technique in the oil and gas industry and has been for many decades. The smart proxy model does not replace the conventional numerical simulation, but it is a reliable alternative technique for fast assessments. The construction of the smart proxy model is different from other types of proxy models. It is based on data mining and artificial intelligence methods. The advantage of applying this proxy technique is that it provides

an accurate replica of a conventional reservoir simulation using a very short computational time. Based on the objective of the smart proxy model, it can be divided into two classes: a well-based smart proxy model and a grid-based smart proxy model. To understand well-production injection behavior, a well-based smart proxy model is developed to simulate the reservoir characterization at the well level. The grid-based smart proxy model is developed for reservoir information such as pressure and saturation at the model grid level. In special cases, a coupled well-based and grid-based smart proxy model is generated.

The concept of the smart proxy model was introduced by Mohaghegh in 2006 [34]. He demonstrated the ability of the developed smart proxy model (surrogate reservoir model) to mimic a one million grid-block reservoir model. The power of the smart proxy model to predict fluid flow in different designed geological scenarios with reasonable accuracy has also been demonstrated. After that, Mohaghegh et al. presented a smart proxy model for fast-track analysis of complex reservoirs [44]. They discussed the new simulation technique and analyzed the impact of the input parameters uncertainties. The proposed model was validated with data that had not been used in the development and the results were satisfactory.

Since then, the smart proxy model has been developed for different field and reservoir problems. Jalalai used the smart proxy model for reservoir simulation and uncertainty analysis of coal bed methane production [45]. He performed a Monte Carlo Simulation, which requires thousands of simulation runs for uncertainty analysis. The developed smart proxy model helped to generate these simulation runs in a shorter time. Also, the number of simulation runs used to build the response surface was very small compared to conventional statistical techniques. Kalantari introduced the new simulation technique to a shale gas reservoir [46]. The smart proxy model was able to predict the production profile of the shale reservoir production with acceptable accuracy.

The aforementioned work on smart proxy models applies to the well level. In 2012, Mohaghegh et al. presented the first smart proxy model away from the well, at the grit-block level [47]. The grid-level smart proxy model was applied to a massive oil field with large number of wells. The grid-based smart proxy model was able to capture the pressure and saturation changes throughout the reservoir with high accuracy compared to conventional simulation techniques. Amini presented a grid-based smart proxy model for a CO_2 sequestration project [29]. It was the first work considering a CO_2 injection scenario using this new approach. With a CO_2 injection in the reservoir, the

developed smart proxy model is used to generate the output of gas saturation at the grid level with minimum error compared to the numerical simulator. Gholami was able to build a smart proxy model for reservoir injection optimization [48]. Gholami concluded that reservoir simulation using proxy modeling is more practical tool that numerical reservoir simulation.

The smart proxy model is extended to another reservoir simulation application, namely history matching. Soodabeh used the smart proxy technique to predict the future well/reservoir performance and to analyze the behavior of the newly drilled wells in a shale reservoir [49]. She performed a history match for every single well with very high accuracy. Shakarami developed a smart proxy model to history match a synthetic reservoir model. This model contains 24 production wells with 30 years of production performance [50]. The history matching method was performed using the smart proxy model and by modifying the geological parameter (permeability). The technique resulted in an accurate replica of numerical simulator history matching in less time. He et al. presented a case study using the smart proxy model to automate the history matching process in the SACROC (Scurry Area Canyon Reef Operators Committee) unit [51]. In the case study, the optimal value of 40 uncertain reservoir parameters could be determined for each well at any given time. The developed smart proxy model was able to successfully perform automatic history matching in a very short time compared to conventional history matching techniques.

Hagighat used a smart proxy model to detect CO_2 leakage [52]. In his work, the smart proxy model was examined under different conditions. It was tested for multiple well leakage, the availability of additional monitoring wells, and CO_2 leakage through cap rock. Shahkarami et al. presented work on assisted history matching using smart proxy technology [53]. The developed technology showed a high accuracy in reproducing the reservoir simulation behavior. Also, the best cases for history matching were generated to perform future predictions.

The main advantages of using the smart proxy model against other proxy models in reservoir simulation are:

1. There is no limitation in reservoir complexity.

2. There is no simplification in the reservoir physics.

3. The time to run the smart proxy model is very short.

Chapter 3

Reservoir Modeling

In order to analyze the fluid behavior of a hydrocarbon reservoir, a reservoir simulation that uses geological and petrophysical parameters should be developed [54]. The following sections describe the reservoir model development of the SACROC unit.

3.1   Field Background

The SACROC Unit, which constitutes the major part of the Kelly-Snyder Field, is located in Scurry County, Texas (Figure 3.1). It is one of the largest unitized fields in the world, covers approximately 50,000 acres, and contains 1,259 wells [55]. The Kelly-Snyder Field was discovered in 1948 and is one of the major oil reservoirs in the United States, originally containing in excess of 2.73 billion barrels of oil.

In 1953, due to a rapid pressure decline in the field, the Texas Railroad Commission approved the creation of SACROC to facilitate a coordinated pressure maintenance program for the field [56].

3.1.1   Reservoir Description

The Scurry Area Canyon Reef Operations Committee Unit is located on a geologic feature named the Horseshoe Atoll. The Horseshoe Atoll is an isolated carbonate platform accumulated during the Middle Pennsylvanian through Early Permian age. It is a 175-mile limestone chain in western Texas. Moreover, the Horseshoe Atoll includes an area of approximately 6,000 mi^2 occupying most of the northern end of the Midland basin, that has the thickest Middle Pennsylvanian to Early Permian isolated limestone reef sequence known in North America [57].
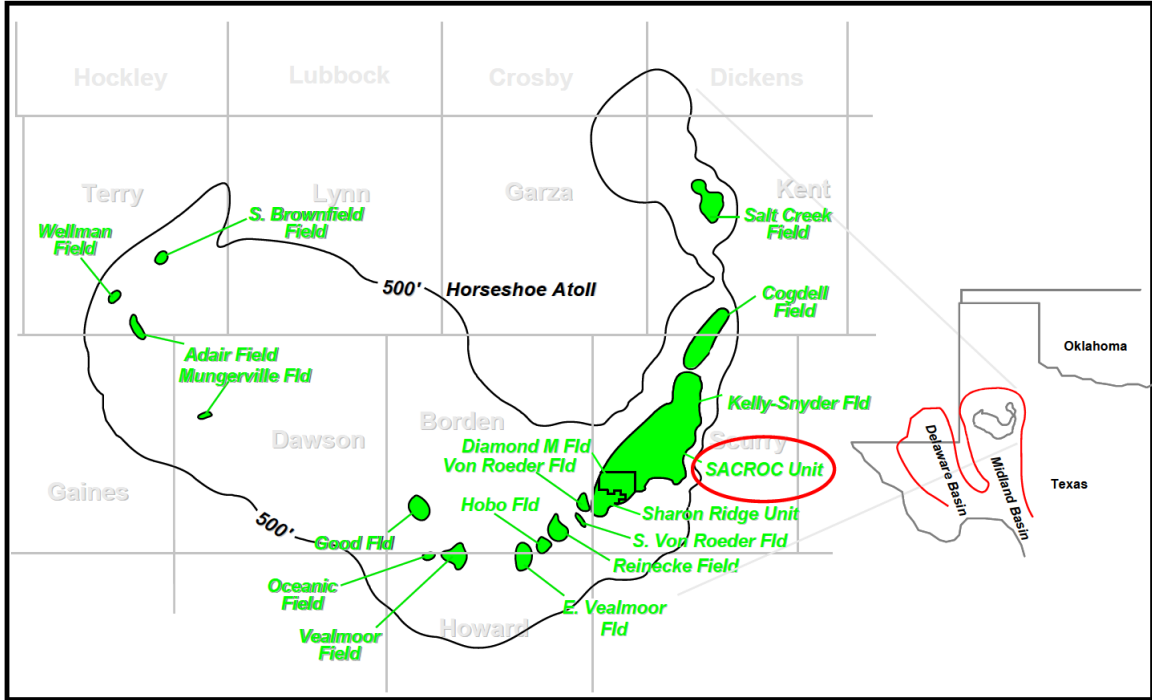
Figure 3.1: Location of SACROC Unit

The SACROC Unit is located in the southeastern segment of the Horseshoe Atoll. It lies within the Midland basin in western Texas. The SACROC Unit contains an area of 141 mi^2 with an approximate north-south length of 25 miles and an east-west length between 2 and 9 miles [58]. The northern half of the unit contains the thickest portion of the reservoir. Also, it is structurally the highest and dips steeply to the west and east [59].

The reservoir of the SACROC Unit is directed north-south trending carbonate buildup with a slight dogleg to the west. The geology of the SACROC Unit is described as a massive bio-clastic limestone with thin shale beds. The analysis of the core samples shows that the SACROC Unit is mostly calcite with low quantities of quartz and clay lenses [60].The seal of the reservoir is formed of a thick sequence of dark black, organic-rich basinal shale and is the source of the trapped hydrocarbons.

Based on the geological description of the available cores and logging information, the main geological beds of the SACROC Unit are divided into the Canyon and Cisco units [58]. The Canyon unit is divided into three formations, the Lower Canyon, the Middle Canyon, and the Upper Canyon. The Lower Canyon is placed below the reservoir. The Upper Canyon is a highly porous unit. It

includes erosional fragments resulting from extended low sea levels. The Middle and Lower Canyon contain high porosity layers and are separated by low porosity carbonate muds [58].
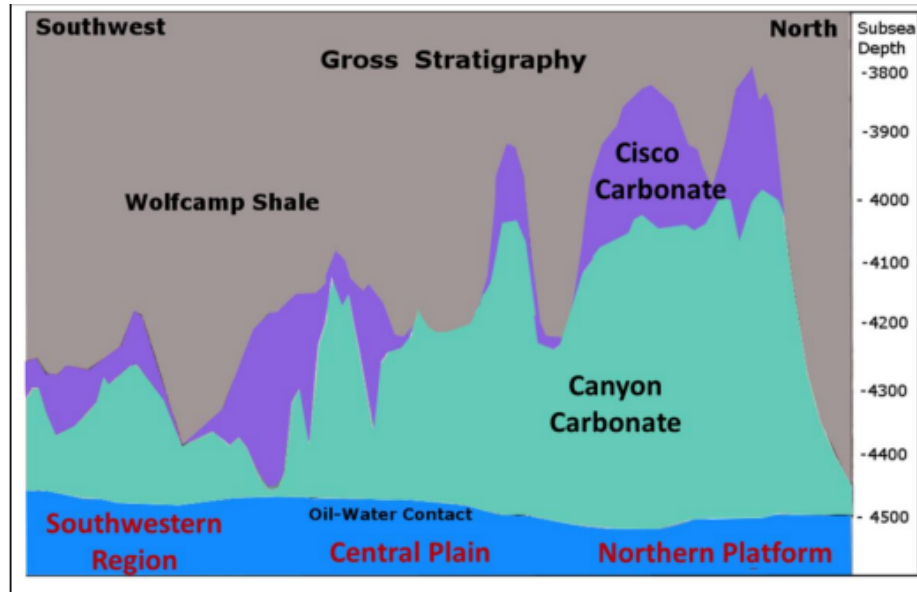
Figure 3.2: SACROC Unit Lithology
[61]

Due to the prepositional environment changes, both porosity and permeability show significant variation. Different studies on the Cisco and Canyon units show porosity values from 0 to 30% and permeability values from 0 to 1000 mD.

In summary, the average depth of the SACROC Unit formation is 6,700 ft. The formation thickness ranges from 800 ft on the crest to less than 50 ft on the flanks. The average thickness is 213 ft. On the basis of drill-stem tests and production data, it has been determined that the first occurrence of water was approximately 4,500 ft subsea.

The initial reservoir pressure at original BHP is 3,122 psig. The solution gas is under 1000 scf/STB and the bubble point pressure is 1,805 psig [62].

3.1.2 Field Performance

The initial reservoir pressure of the Kelly-Snyder field was 3,122 psig. The primary reservoir mechanism was natural solution gas drive. The field was discovered in 1948 and at the end of 1951; more than 1,200 wells had been drilled in the field. More than 80 companies operated in the field and
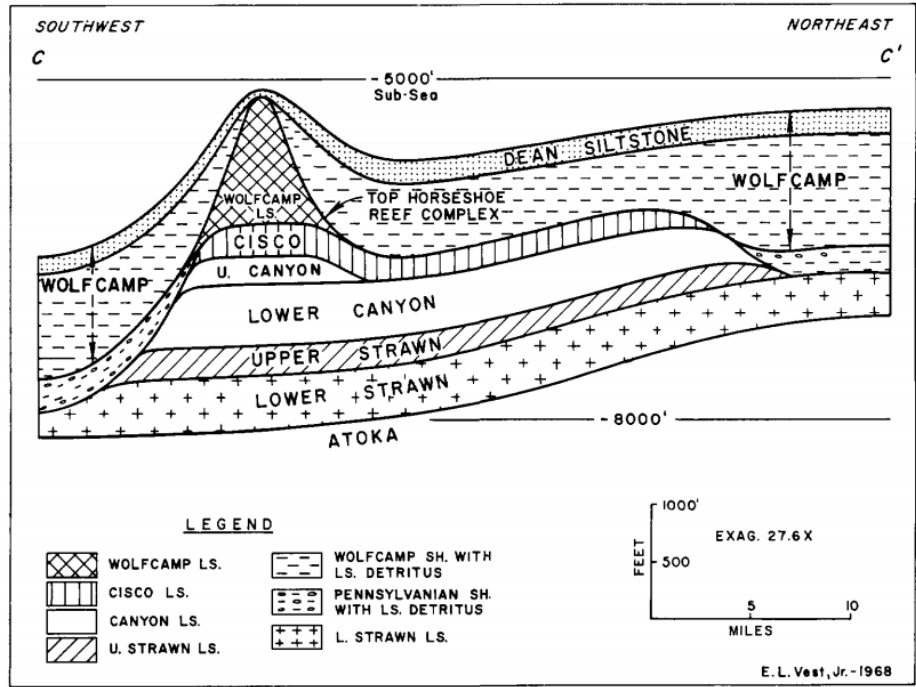
Figure 3.3: Southwest-northeast schematic cross-section C-C through thickest known part of Horseshoe atoll

[58]

SACROC Field Characteristics	
Average Porosity	7.6%
Average Permeability	19.4 mD
Formation Depth	6700-7000 ft
Initial Reservoir Pressure	3122 psig
Reservoir Temperature	130 °F
Bubble Point Pressure	1820 psig

Table 3.1: Basic Reservoir Data of SACROC Unit

due to this competitive environment, the reservoir pressure had dropped to 1,650 psig with only 5% of the reserved oil produced. With the primary mechanism, it was estimated that only 19% of the original amount of oil in place could be recovered [63].

Due to the rapid pressure decline, a pressure maintenance program became essential.
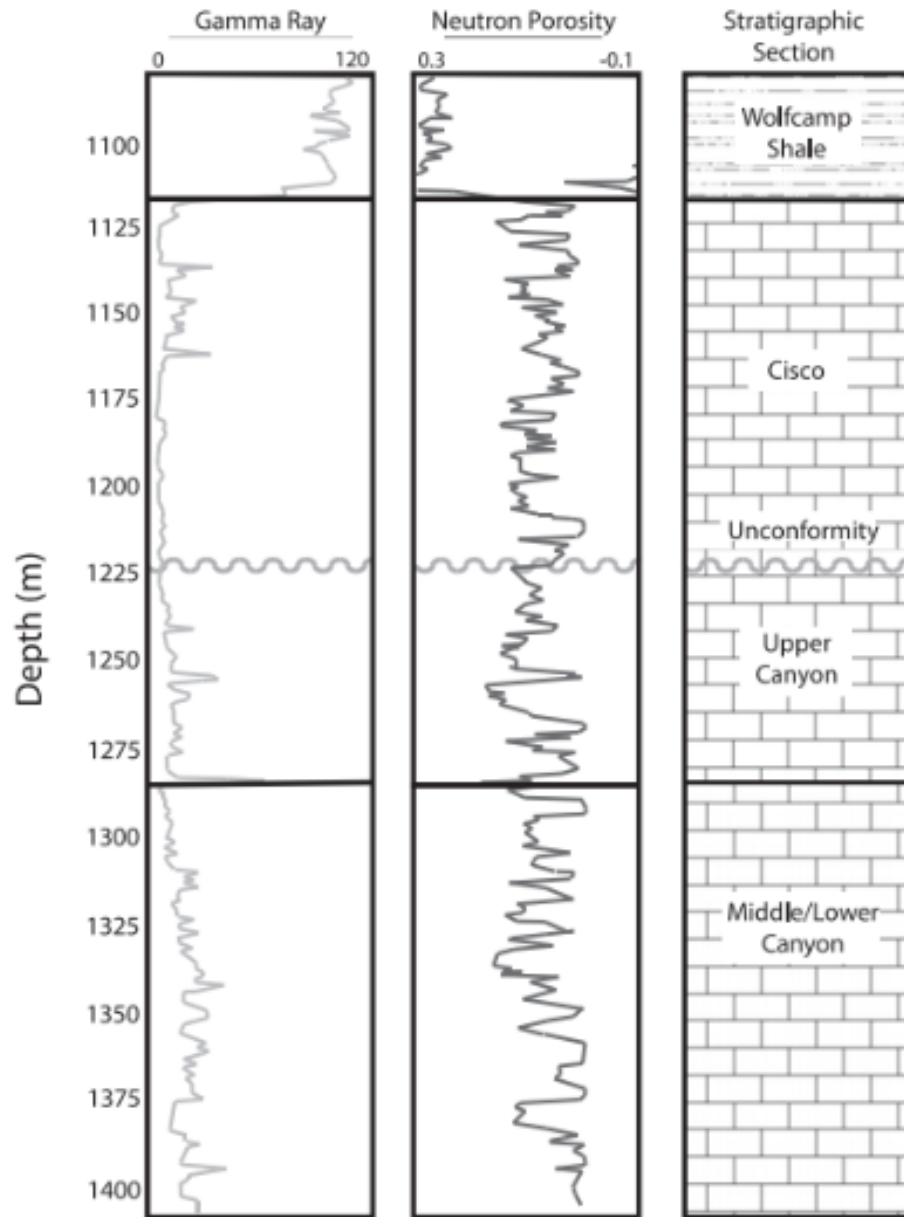
Figure 3.4: Typical Well Logs in SACROC Unit
[58]

3.1.3 Water-Flood Program

Due to the rapid decline in pressure, the SACROC Unit was put under a full-scale pressure maintenance program in 1954, using a water injection system. The water injection wells were placed along the longitudinal crest of the structure (Figure 3.5). The pressure maintenance program was very effective for many years. The water injection program resulted in a restoration of the average reservoir pressure from a 1546 psi in 1954 to nearly 2300 psi by 1967 [55].
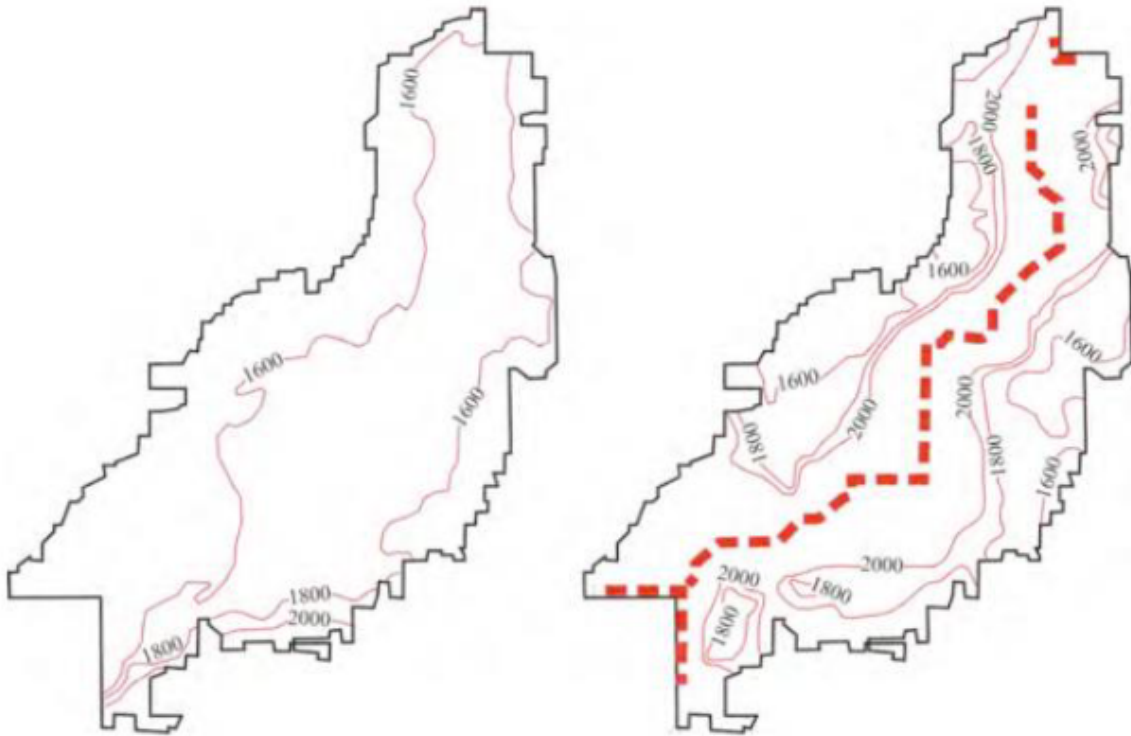
Figure 3.5: SACROC unit pressure contour map. Left side shows the pressure contour map in 1954 where the right side of the figure shows the pressure contour map in 1970.
[48]

Before the water injection, 99% of the reservoir was below the bubble-point pressure (bubble point pressure is 1820 psig). Two years after the injection, 45% of the reservoir was above the bubble-point pressure. In 1961, 80% of the reservoir volume was above the bubble-point pressure and by 1972 the average reservoir pressure was 2356 psig [64].
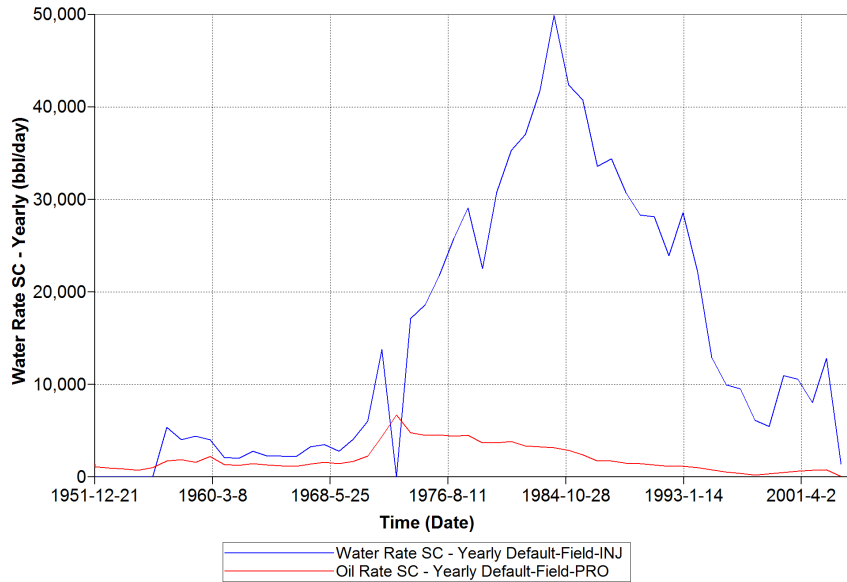
Figure 3.6: Performance history of SACROC Unit

3.1.4 CO2 Injection

The pressure maintenance program using water injection was successfully performed to increase the pressure. However, there were still challenges with pressure maintenance on the eastern and western margins of the reservoir. To overcome these challenges and after studies showing that the SACROC Unit properties are ideal for miscible CO_2 recovery, a CO_2 injection recovery technique (water alternating gas) was applied in 1972 [65].

The CO_2 injection was carried out in three consecutive phases; central, north and south WAG flooding. In the central SACROC Unit, the WAG injection was started in 1972 and lasted until the second quarter of 1973. During this phase, the maximum oil rate increased from 30,000 barrels per day to 100,000 barrels per day. The second WAG injection was performed in the north part of the field in June 1973. The oil rate reached 80,000 barrels per day from 40,000 barrels per day before the WAG process. The south part of the SACROC field was chosen for phase three of the WAG process in April 1973. The oil rate was increased from 30,000 barrels per day to 40,000 barrels per day [66]. After two years of WAG injections, oil production peaked at 211,000 barrels per day.
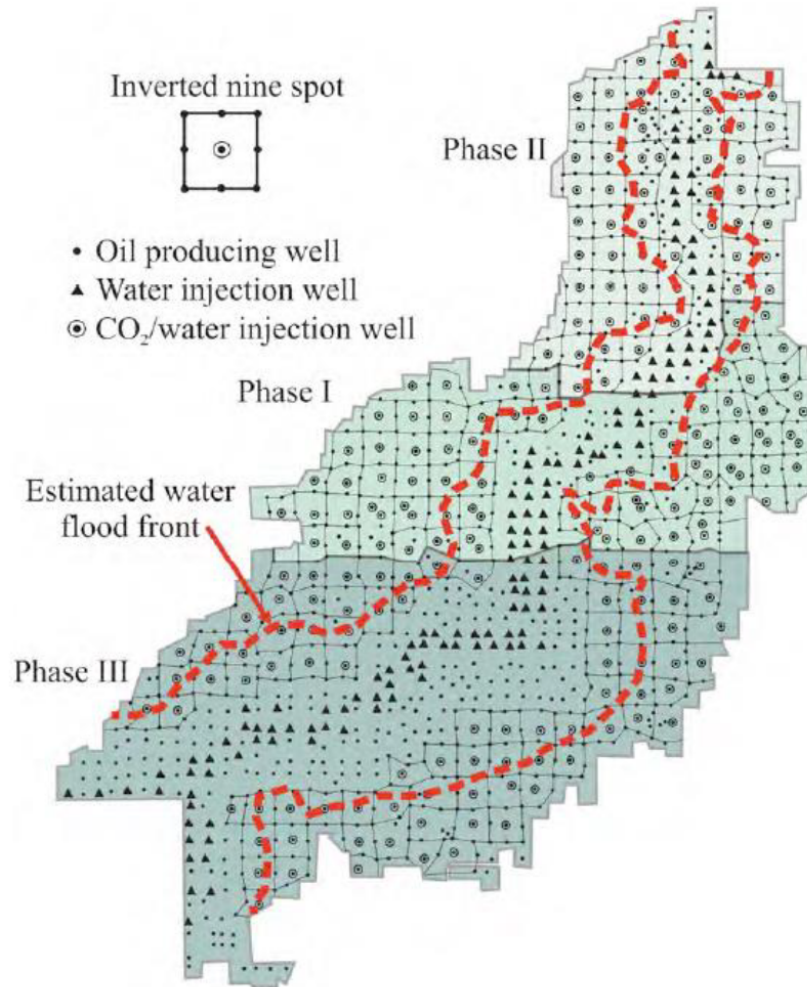
Figure 3.7: WAG Phase Locations of SACROC Unit
[67]

3.2 Geo-Cellular Model Development

In order to analyze the fluid behavior of a hydrocarbon reservoir, a reservoir simulation model that uses geological and petrophysical parameters is developed.

The Texas Bureau of Economic Geology has developed a geo-cellular model for the northern platform of the SACROC field. It is a high-resolution model with a dimension of 149X287 and 221 layers. The size of the geo-cellular model is 4,000 m wide and 10,000 m long. The following sections describe the reservoir model development of the SACROC field.
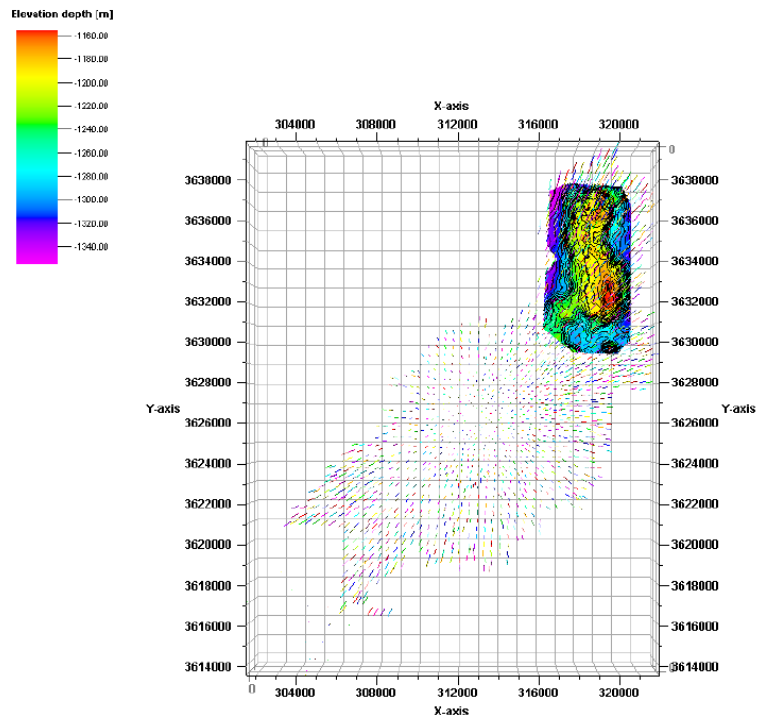
Figure 3.8: Reservoir Model of Northern Platform in SACROC Unit
[66]

3.2.1 Estimation of Porosity and Permeability

A very important step in building the geo-cellular model is developing an accurate porosity-permeability relationship. This step becomes more challenging when dealing with complex reservoir descriptions. Typically, the porosity measurements are easier to obtain because of the available density and sonic logging tools. On the other hand, directed permeability techniques are expensive and difficult to get [68]. Therefore, to calculate the permeability, porosity-permeability correlations are generated

using the available core permeability data.

A number of techniques to develop porosity-permeability correlation have been discussed in the literature. The one used for the SACROC geo-cellular model is based on the rock-fabric approach introduced by Lucia in 1995 [69]. Lucia related the permeability distribution in carbonate formations to the rock pore size. Moreover, the pore size distribution is related to rock fiber. In his work, the pore size (porosity) is divided into two classes, inter-particle porosity and vuggy porosity. Inter-particle porosity includes inter-grain and inter-crystal pore spaces. The separate and connected vugs belong to the vuggy porosity class.

Three types of rock fabrics are presented by Lucia: first class, inter-particle size from 100 to over 500 m dominant in grain-stones, dolograinstones, and large crystalline dolostones; second class, inter-particle size from 20 to 100 m, dominant in grain-dominated pack-stones, fine and medium crystalline, grain-dominated dolopackstones, medium crystalline, and mud-dominated dolostones; third class, inter-particle sizes from 5 to 20 m, and is dominant in mud-dominated lime-stones and fine crystalline, mud-dominated dolostones.

The following equation developed by Lucia [69] shows the rock-fabric permeability calculation method:

$$log(k) = (9.7982 + 12.0838 log(\lambda)) + (8.6711 + 8.2965 log(\lambda))\phi \tag{3.1}$$

Where:

λ : rock-fabric number

k : permeability

ϕ : inter-particle porosity

The data plots of porosity-permeability in the SACROC field, show the high heterogeneity of this field. The porosity values ranges between 0 to 30 percent, and the permeability values ranges between o md to 2000 md (figures 3.9 and 3.10).

The porosity used in the equation is calculated from 368 wireline logs available and three-dimensional seismic data [70]. Because of the heterogeneity of the Cisco formation and the low porosity intervals, wireline logs were not enough to build a high quality geological model. Therefore, seismic data was added to the wire-line logs in order to build a high quality porosity distribution

Table 3.2: Permeability Calculations Using Porosity and Fabric Number

	Formation/Sequence	Rock Fabric Number	Transformation Equation
[69]	Cisco/Late		$K = 2.1625 \times 10^6 \times \varphi^{3.8844}$
	Cisco/Late	1.7	$K = 1.031 \times 10^7 \times \varphi^{6.7592}$
	Cisco/Early	1.9	$K = 2.96 \times 10^6 \times \varphi^{6.3584}$
	Canyon/1	2.5	$K = 97,628 \times \varphi^{5.396}$
	Canyon/2	1.75	$K = 38,520 \times \varphi^{5.0923}$
	Canyon/3		

in the Cisco formation. In the Canyon formation, on other hand, due to the continuous and better porosity, the wire-line logs alone were able to develop a reasonable porosity disturbance for the geological model [48].

Based on the aforementioned rock fabric equation coupled with generated porosity distribution, permeability is calculated for each layer and complete porosity-permeability is developed for the SACROC Unit.

3.2.2 Geo-Cellular Model Upscaling

Because the high level of detail of the SACROC geo-cellular model (9,450,623 grid blocks) significantly increases the computational cost of this model, procedures have been carried out to coarsen the reservoir description to reduce the computational cost. Such procedures are referred to as "upscaling" or "scale up" techniques [71]. The upscaling process is an averaging procedure in which the static characteristics are up scaled from properties on a fine-scale grid [72].

In the high-resolution model, there are 221 geological layers. All these layers belong to five geological formations (Canyon 1, Canyon 2, Canyon 3, Canyon 4, and Cisco). The main aim of the upscaling is to reduces the geological layers from 221 layers to 16 layers in the five main geological formations. Reducing the number of layers means reducing the number of grid cells. In other words, the static parameter values in the upscaled cells need to be averaged. The averaging method used for porosity and permeability are harmonic averaging and geometric averaging respectively.
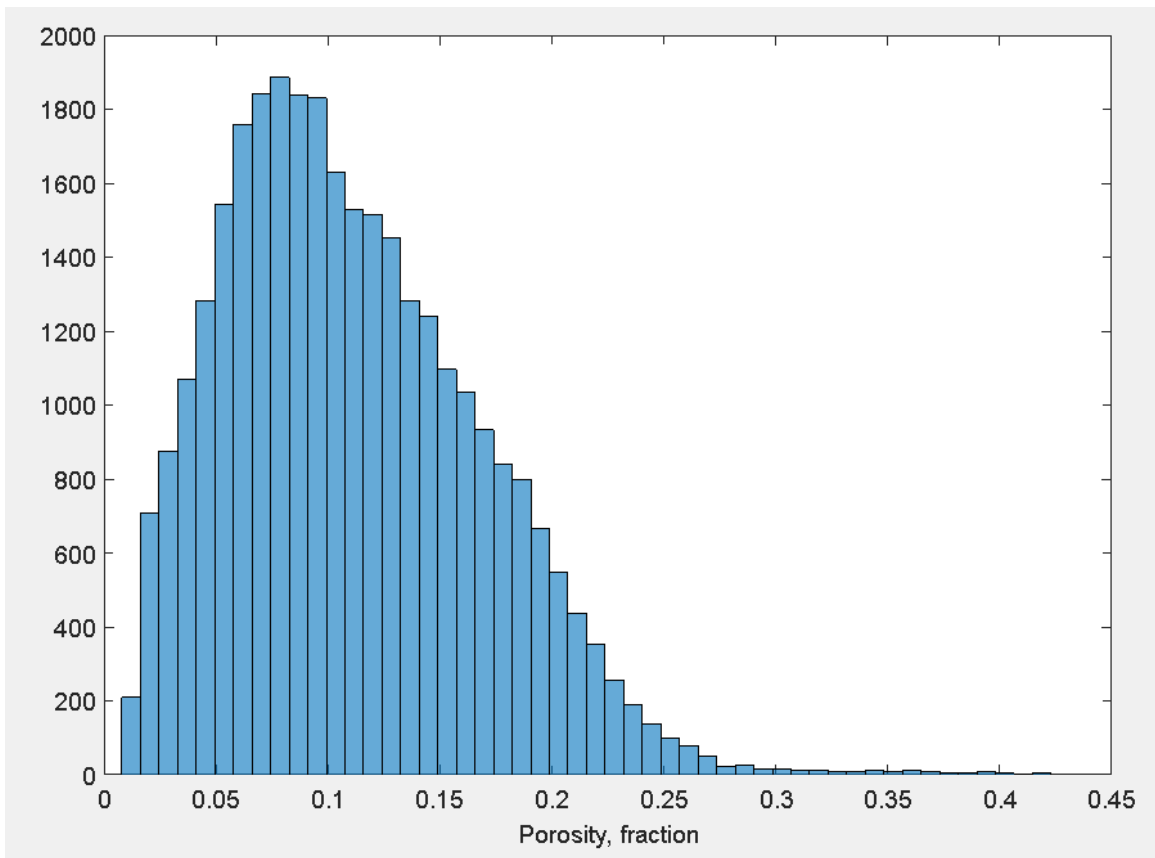
Figure 3.9: Porosity histogram for the selected study area.
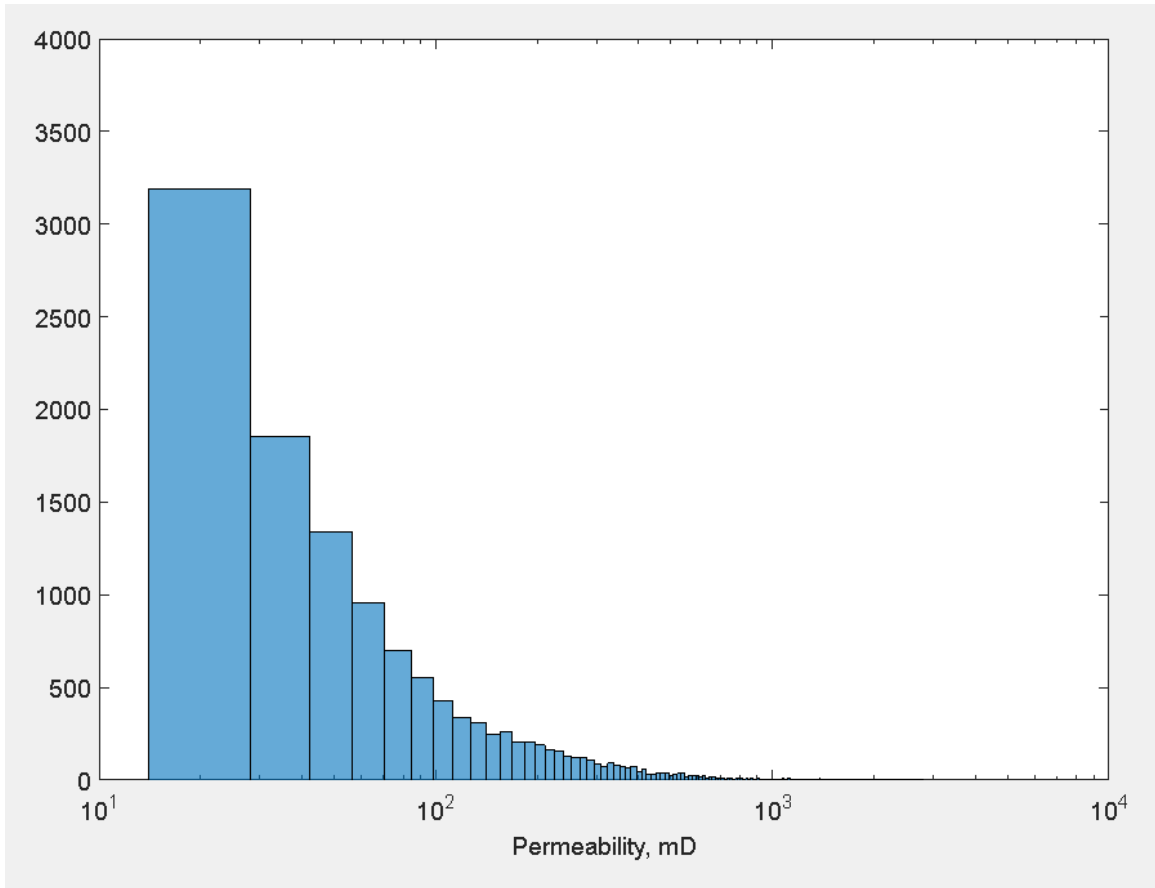
Figure 3.10: Permeability histogram for the selected study area.
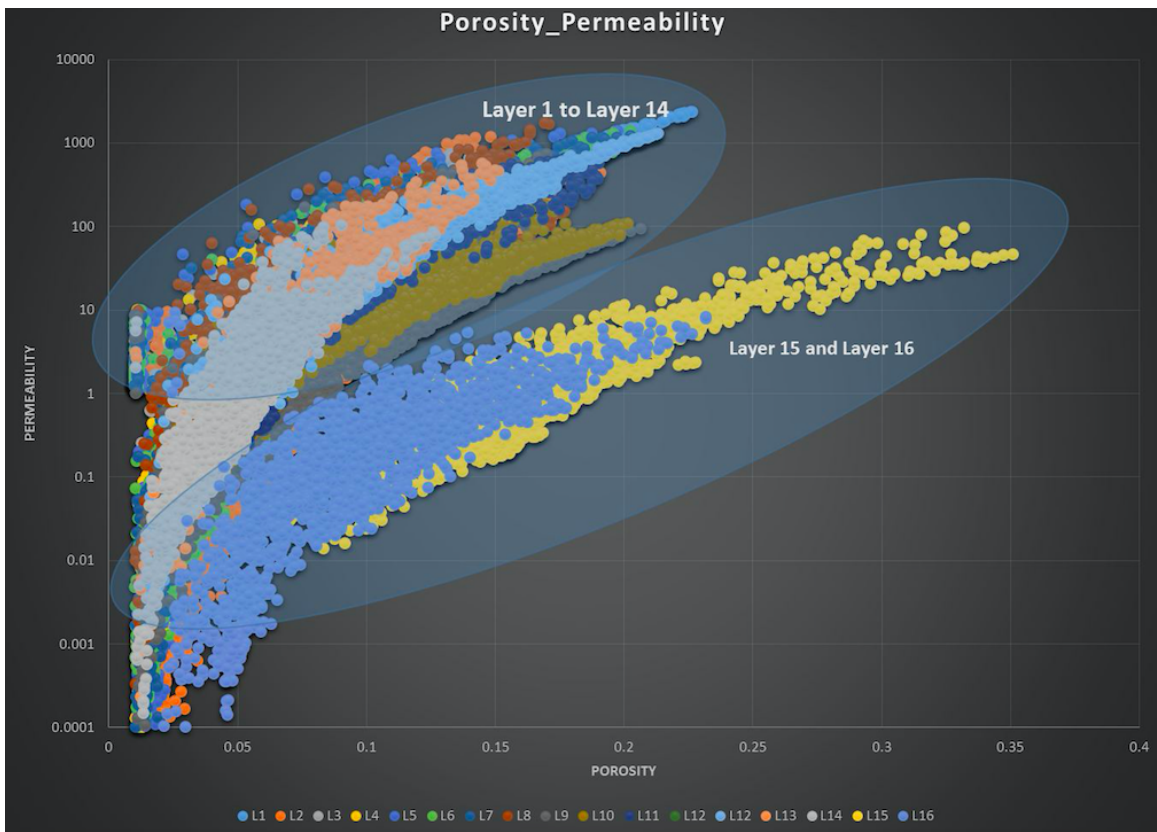
Figure 3.11: Porosity-Permeability Relationship.

The upscaling procedure leads to a reduced model resolution. That said, efforts were made to pre-serve the heterogeneity as much as possible.

In addition to the upscaling process, to fit the model for the purpose of this study, a northern part of the SACROC field was selected for this study. The selected part of the fields dimension was 51X39X16 with a total of 31,824 grid blocks.
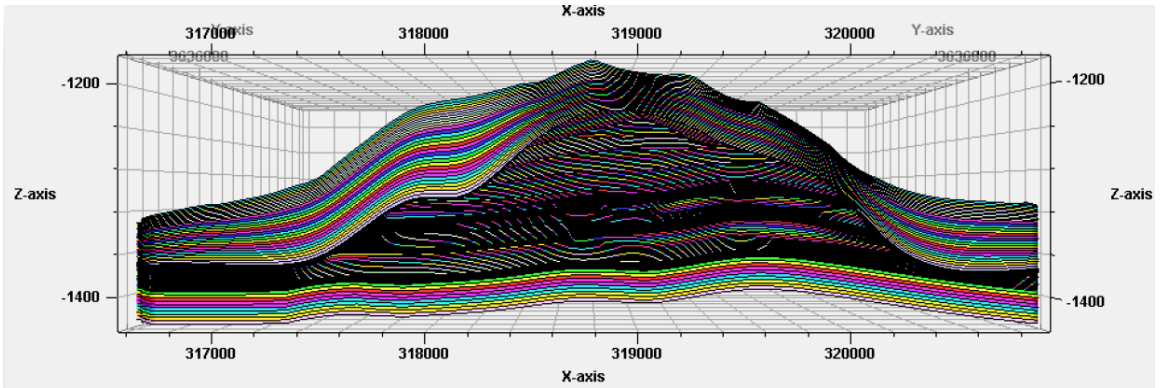
Figure 3.12: Side view of the high resolution model of SACROC field with 221 layers [48]
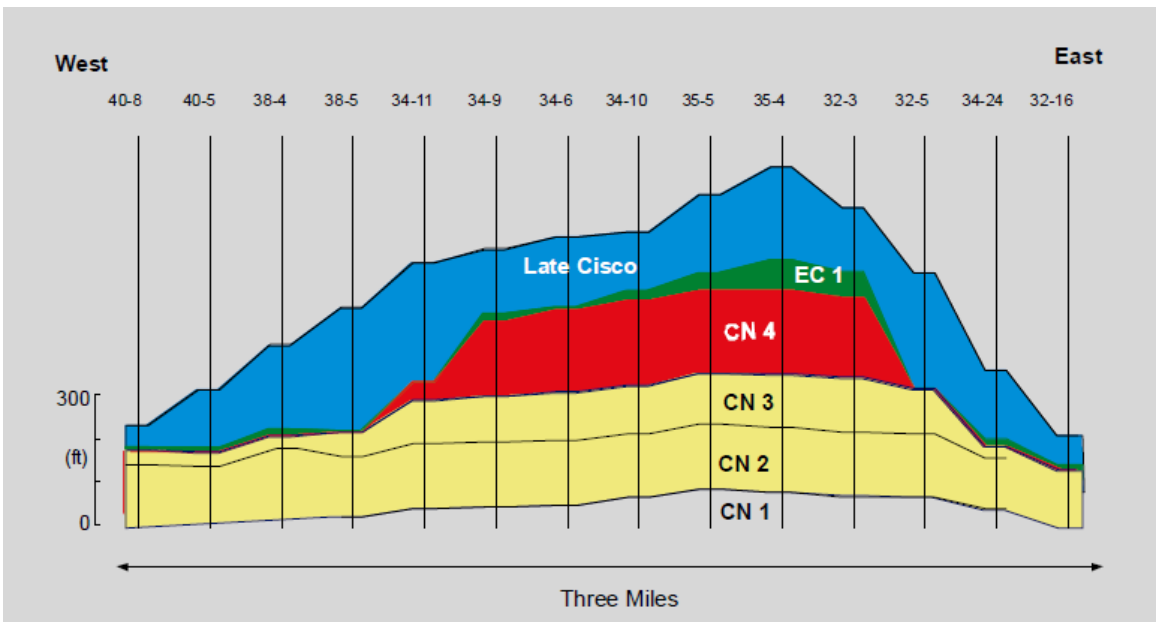
Figure 3.13: The five main formation where the model upscaled from 221 layers to 16 layers (Canyon 1 (2 layers), Canyon 2 (2 layers), Canyon 3 (3 layers), Canyon 4 (2 layers), Cisco (7 layers)).
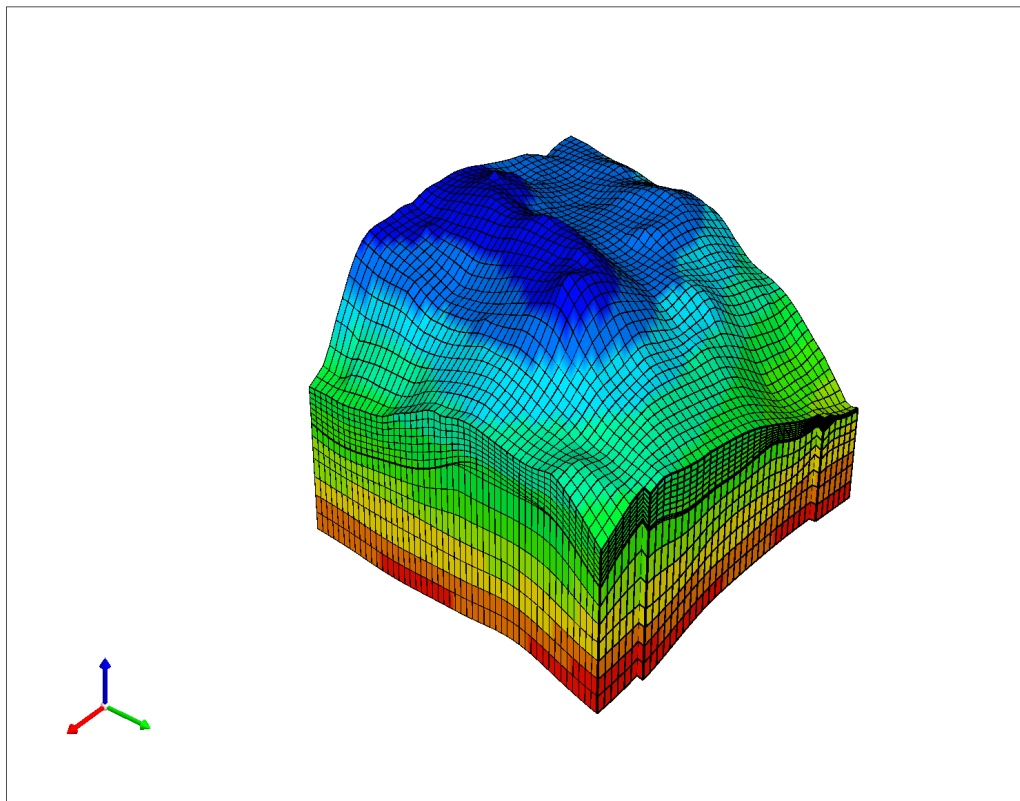
Figure 3.14: 3-D Grid Top of the selected area of study.

3.2.3   Rock Fluids Properties

In oil and gas reservoirs, two or sometimes three phases are present. Generally, permeability measurements are made with a single fluid filling the pore space. Therefore, the measured permeability in a multi-phase reservoir is expected to be lower than that in a reservoir with a single fluid. This is because it occupies only part of the pore space and may also be affected by interaction with other phases. Moreover, when a wetting and a non-wetting phase flow together in a reservoir rock, each phase follows separate and distinct paths. This behavior is called relative permeability [73].

Usually, the permeability of the chosen base is measured at the start of the experiment. For instance, the experiment starts with measurements of the permeability of oil in the presence of irreducible oil in the core. Water is then injected into the core. The permeability of both oil and water are measured as water replaces the oil. In this case, the base permeability is the initial permeability of the oil at Swi [74].
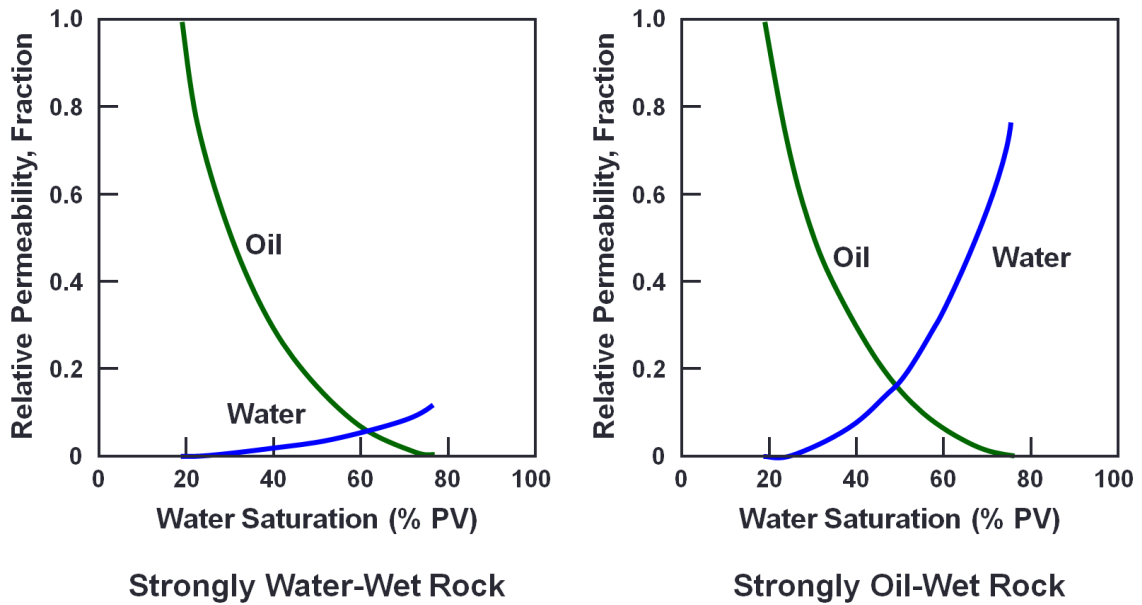
Figure 3.15: Typical Relative Permeability Curves

In the SACROC Unit, the relative permeability function has been developed from a similar carbonate rock. The developed function was extrapolated from relative permeability data collected from Wabamun Lake (Figure 3.16 a). Then, the relative permeability of SACROC core Core128V at 50C and 1.4 MPa [75] was used to calibrate the oil permeability curve from Wabmun Lake (figure 3.15 b).
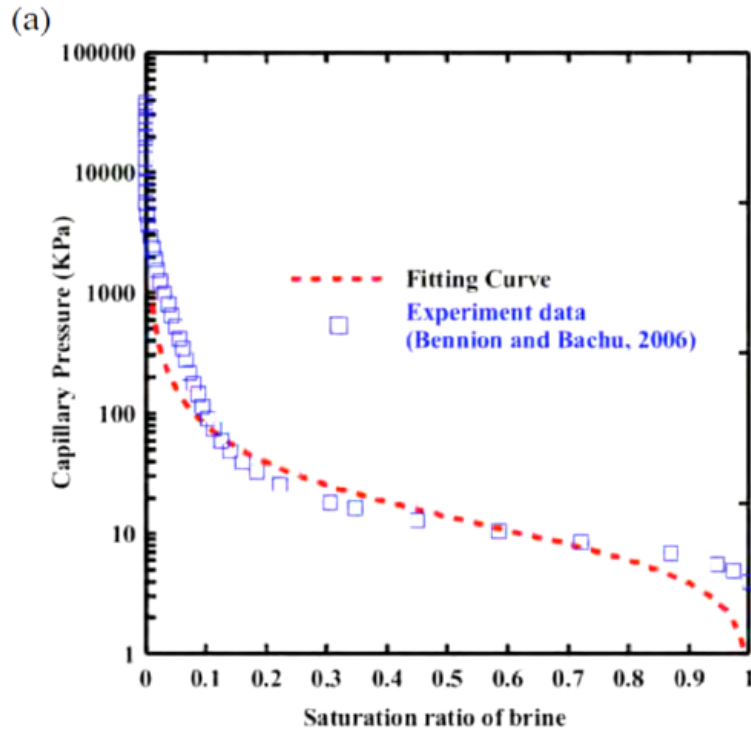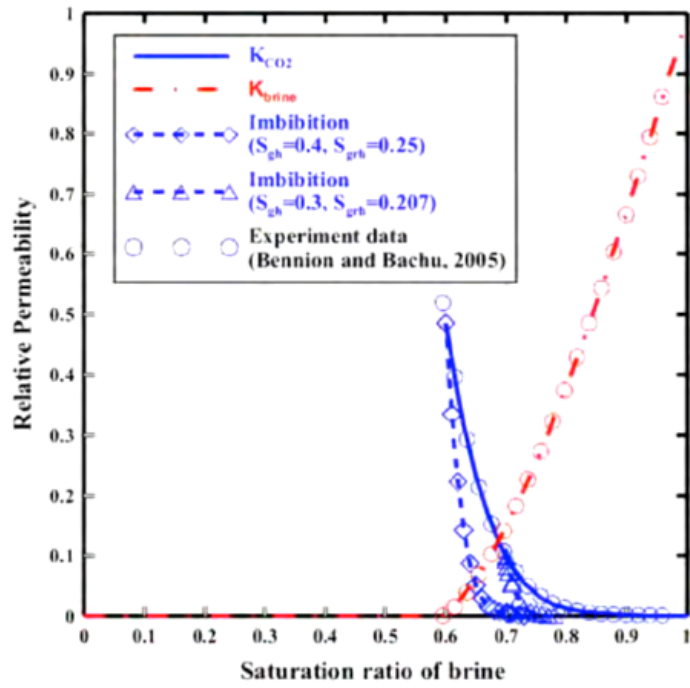
Figure 3.16: (a) Relative permeability curve with experimental data from Wabamun Lake (b) Fitted relative permeability curve with experimental data from SACROC core test [76]

The relative permeability curves used for this study numerical model are shown in Figure 3.17.
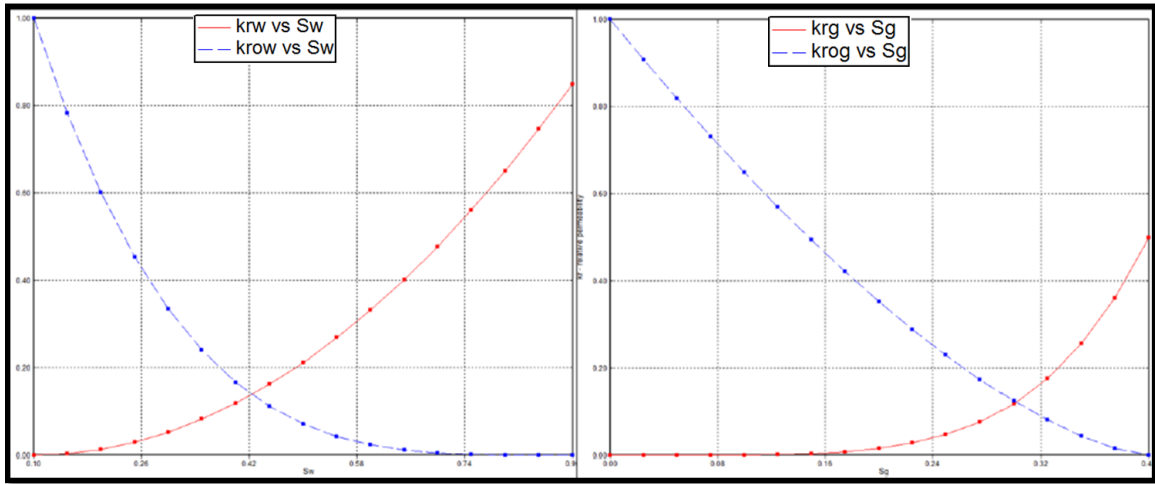
Figure 3.17: Relative Permeability Curves for The Area of Study

3.2.4 Initial and Boundary Conditions

The SACROC Unit reservoir was initially saturated with oil and gas. The initial reservoir pressure was assumed to be 3,122 psig at oil-water contact of -4,300 ft. The reservoir temperature used for the developed model was 130 F and the water saturation below the oil-water contact was 36% [64]. The initial oil composition contained 11 gas components summarized in table no 3.3 [58].

All reservoir boundaries were set as no-flow boundaries. Han [70] has determined the Wolf-camp shale located above the Cisco and Canyon formations as an effective seal. Therefore, the upper boundary was treated as a no-flow boundary. The description of the wolf-camp shale formation as a carbonate reef complex and the prism shape of the Cisco and Canyon formations mean that the eastern, western, and northern boundaries could be set as no-flow conditions. Finally, the bottom boundary was set as a no-flow boundary due to the low permeability in the Strawn formation [58].

For this study, part of the northern part of the SACROC field was chosen. To minimize the effect of the wells beyond the selected area and to replicate the open-flow boundary, six hypothetical production wells are placed in the extended section of the model.
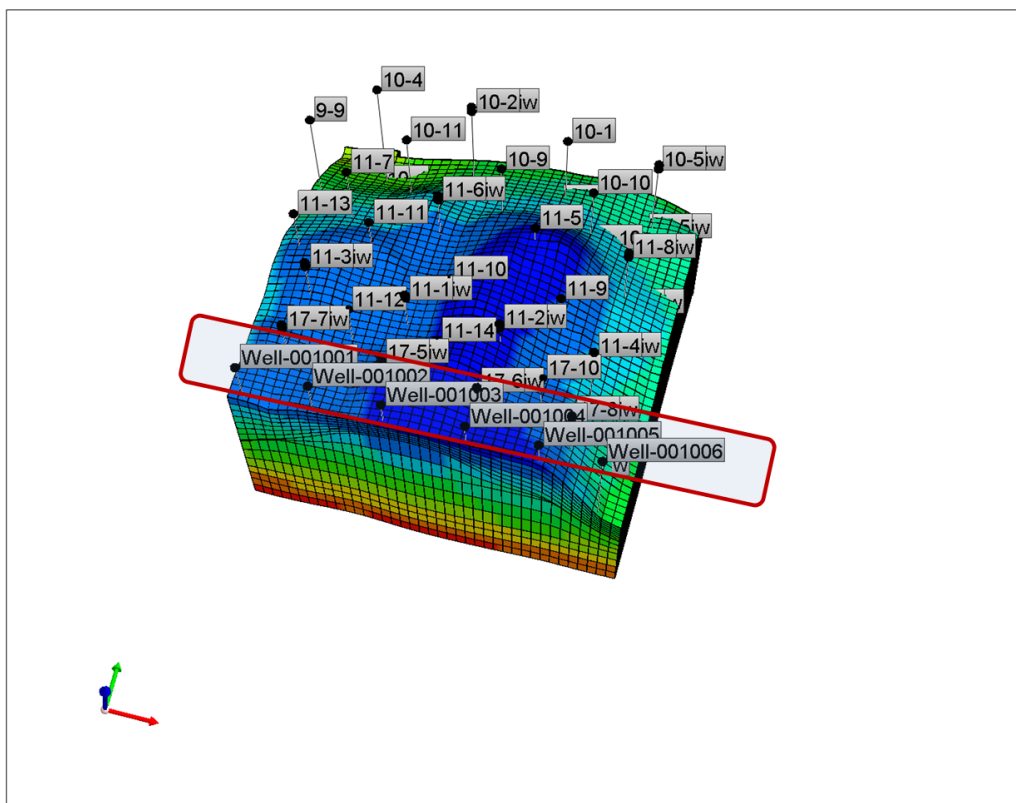
Figure 3.18: Hypothetical wells to mimic open flow conditions.

Table 3.3: Oil Composition for SACROC Unit Model

Oil Composition	Mole	Mocular Weight
CO2	0.0032	44.01
N2	0.0083	28.01
C1	0.2865	16.04
C2	0.1129	30.07
C3	0.1239	44.1
I-C4	0.0136	58.12
N-C4	0.0646	58.12
I-C5	0.0198	72.15
N-C5	0.0251	72.15
F-C6	0.0406	86
C7+	0.3015	197.4

[58]

3.2.5 Wells and Completions

This section mainly concerns the wells in the selected area of study of the SACROC Unit.

In this selected area, there are 27 production wells. Twelve of these were converted to injection wells at a later stage. The number of production/injection wells varies throughout the fields history. Figure 3.19 shows the number of production/injection wells over the field history from 1949 to 2003 and Figure 3.20 shows the production/injection wells history of the selected area of study. For instance, in 1980, there were 18 production wells and nine injection wells. However, the numbers changed in 2000 to nine production wells and 11 injection wells. This figure is conclusive evidence of the production/injection complexity of the studied area. In addition to the production/injection wells in this section of the field, a number of hypothetical production wells was added to the model to mimic the flow boundary.

In the simulation model, the operational constrains used are the liquid rates. For production wells, the minimum bottom-hole pressure (BHP) was set to 28 psi. To ensure that the entire required amount of injection can be achieved through the low permeability reservoir, the bottom-hole

injection pressure was set to 5,000 psi for injection wells.

All the production/injection wells are drilled vertically and completed as open hole wells across the 16 geological layers.
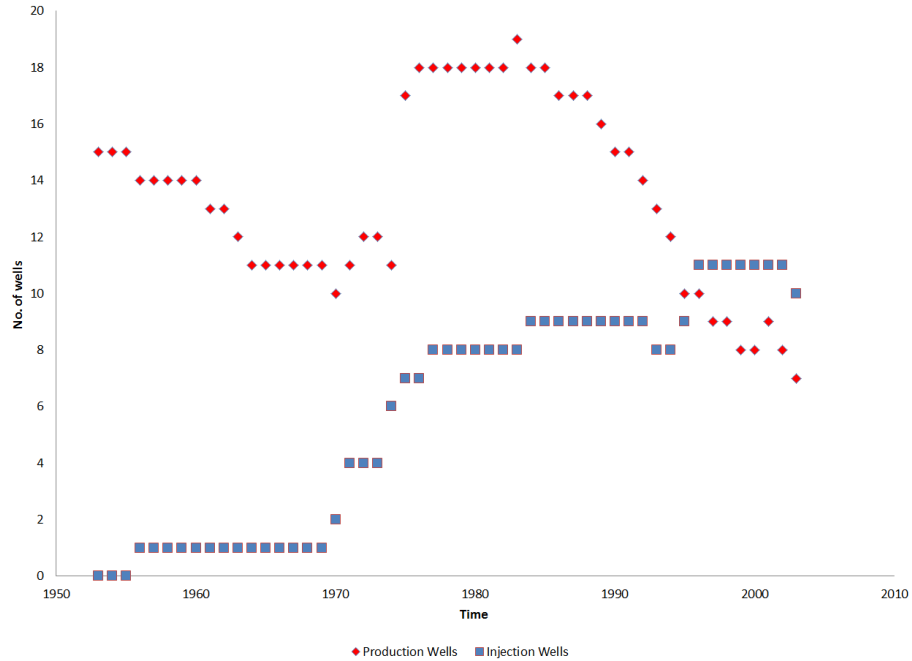
Figure 3.19: Production/Injection wells versus the history of the studied area

3.2.6 History Matching

History matching is a very important process that validates the built reservoir model for future field plans.

As the name implies, history matching is the process of adjusting the reservoir model until the closest match to previous reservoir performance has been achieved [77]. The availability and quality of historical production and pressure data play a fundamental rule in history matching. In this process, some of the reservoir parameters are modified with reservoir engineering experience until the reservoir model has been validated.

The history-matched reservoir model provides a higher degree of confidence for simulating different field production scenarios. It helps the reservoir engineer to better understand reservoir behavior,
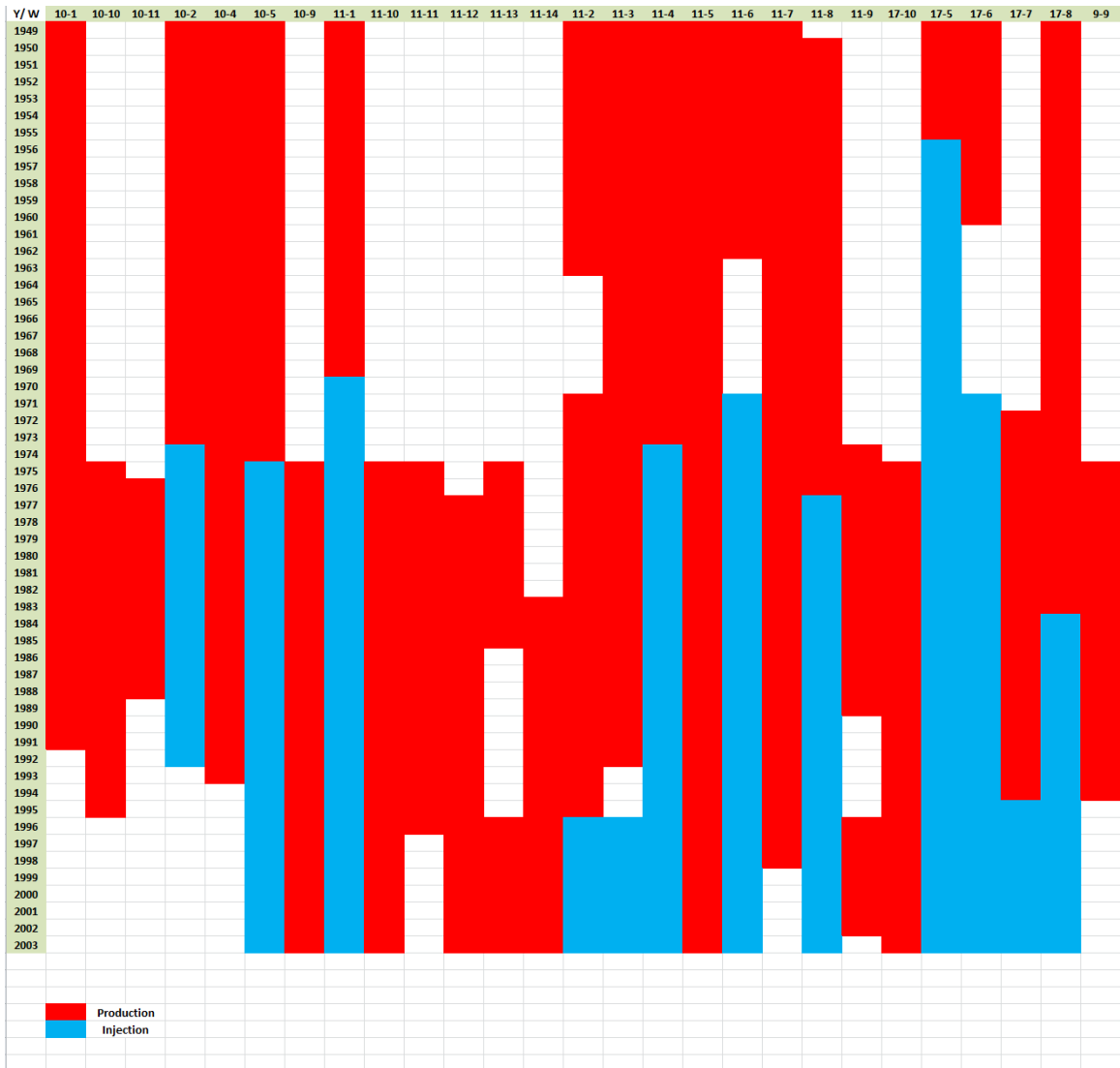
Figure 3.20: Production/Injection history of the Selected area of study.

to improve reservoir description, and to identify unusual operating conditions [78].

Due to the heterogeneity of the reservoir properties, performing history matching for the SACROC Unit is challenging. The area of study of the SACROC Unit has been history matched by He [66] as part of her graduate study research project. In her work, all reservoir properties were considered trustworthy, except for porosity, permeability, and relative permeability [66].

In the process of reservoir model adjustments, the porosity and permeability were modified using the multiplier technique. Both reservoir properties were decreased by 50% and increased up to 400%. However, only one relative permeability curve was used due to the limit of information availability [66].

Based on different geological scenarios, 80 simulation runs have been developed for history matching. He [66] was able to match water production, oil production, and average reservoir pressure.
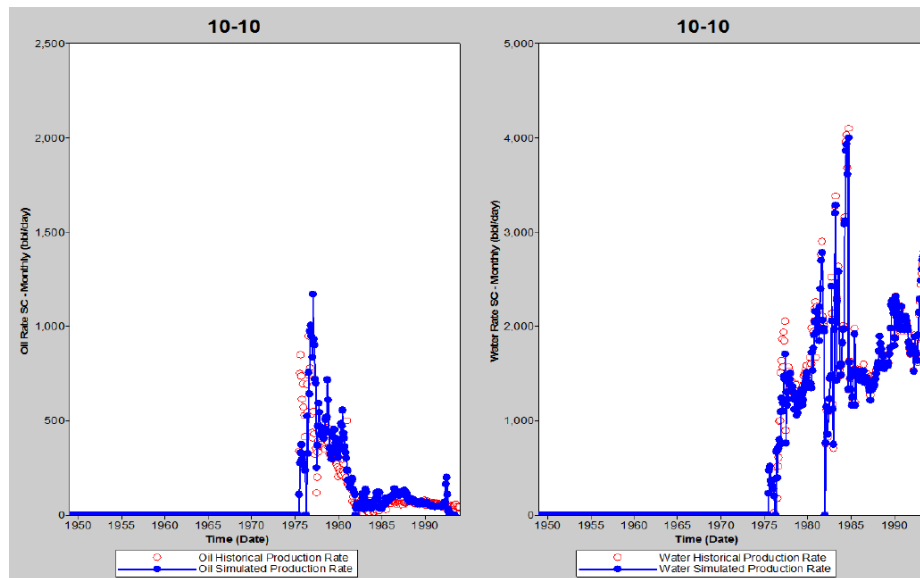
Figure 3.21: Water and oil production history match results of well 10-10
[66]
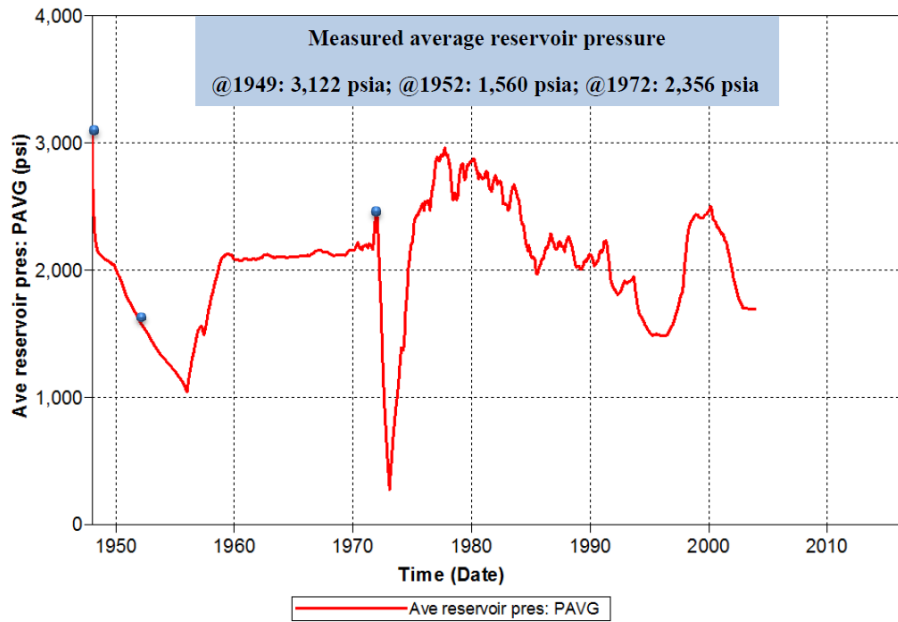
Figure 3.22: Average reservoir pressure history match results of three known points
[66]

3.3 Numerical Reservoir Development

Once the geo-cellular model has been developed, it is transformed to a fluid flow simulator. The commercial simulator utilized in this study came from the Computer Modeling Group. It is a multi-dimensional compositional simulator. [79].

Chapter 4

Smart Proxy Model Development Methodology

The smart proxy model is a data mining and artificial intelligence approach. The main step of developing the smart proxy model is to define the objective of the smart proxy. Based on the defined objective, the type of smart proxy and the number of numerical simulations can be determined. Moreover, the nature of the reservoir realization is identified based on the objective of the smart proxy. For example, a geological realization would be generated if the objective is history matching. The operational constraints would be modified for field optimization. The technique of the smart proxy is to learn from provided examples, so it is necessary to have a sufficient quantity and quality of information to build a successful smart proxy.

Generally, there are six main steps involved in developing the smart proxy model. First, define the objective of the smart proxy. Second, design the full field numerical simulation model. Third, extract the static and dynamic data from the simulation run to generate a spatiotemporal database. Fourth, select the required parameters from the database based on the objective of the smart proxy. Fifth, using neural networks, train and validate data for the targeted reservoir property (the trained and validated neural network is the developed smart proxy model). Finally, apply the smart proxy model to a blind simulation run which was not included in the training set for verification [80]. The following sections discuss these steps in more detail..
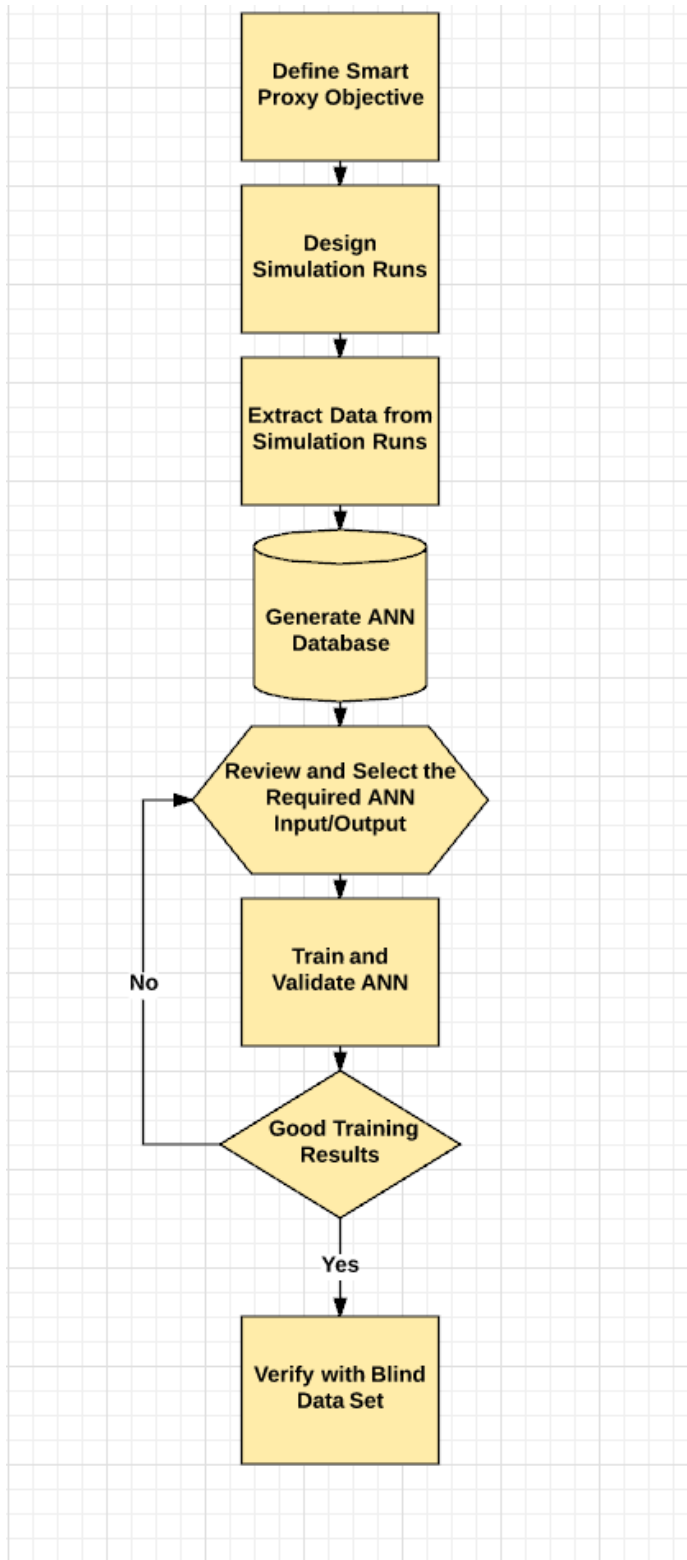
Figure 4.1: Smart Proxy Model Work Flow

4.1   Numerical Simulation Design

The purpose of designing and running the numerical simulation model is to provide the sample space of model input-output relationships required to train the neural network. The simulation run design must take the objective of the smart proxy model into consideration.

The aim of the current study is to replicate and predict the reservoir dynamic properties at grid-block level under different geological and operational realizations. In reservoir simulators, the hydrocarbon flow is calculated in every grid block but presented in the final output as one well production value at each time. In this study, the final output is the flow calculations at each grid block. This means that for each well there will be 16 production data points that represent 16 geological layers. Obtaining the production data at every grid block of the wells generates the required data heterogeneity for the training of the neural network (Figure 4.2 shows the oil flow rate heterogeneity).
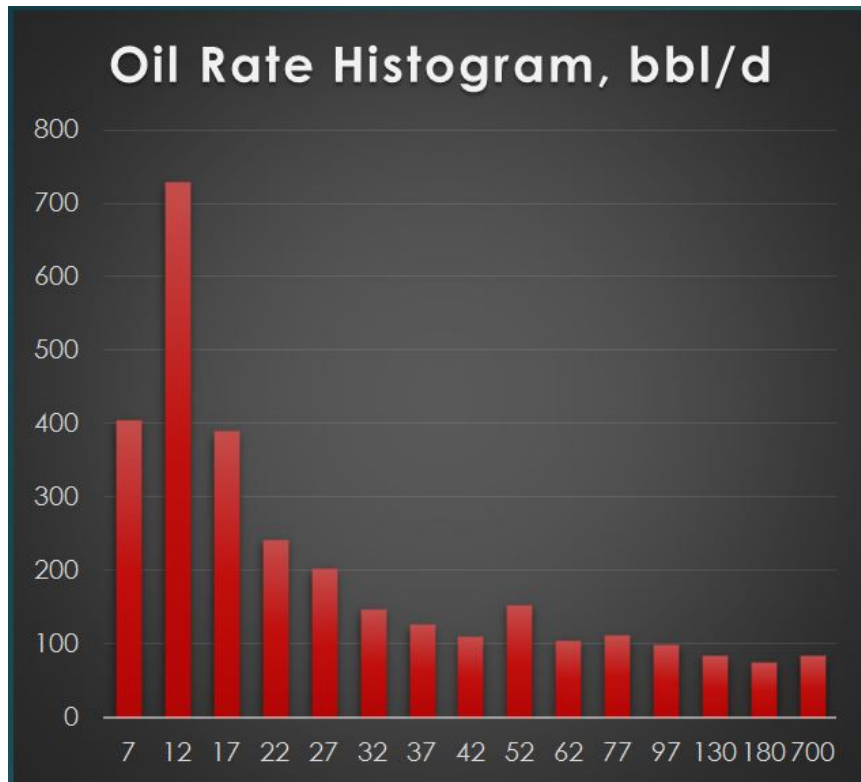
Figure 4.2: Field Oil Rate Histogram

4.2 Database Generation

A database that includes time-related parameters is called a temporal database. The parameters have different values at different times. The database stores a sequence of events over time. Spatial databases are databases that contain spatial-related information (e.g., geographic (map)) [81].

The majority of time spent developing the smart proxy model is consumed by database generation. The smart proxy model is constructed to represent the principles of the reservoir physics. Therefore, reservoir engineer knowledge and inputs are essential to develop the smart proxy model. The data input in database comes from two sources, the geo-cellular model (static data) and the fluid-flow model (dynamic data) at each grid block.

The static data include reservoir structure and grid blocks information. The dynamic data are the properties that change with time and come from two domains. The first domain is the well domain data such as production/injection rates and well BHP. The grid block domain data are the system state variables such as pressure and saturation.

In order to monitor the pressure and saturation movement, the data from neighboring grid blocks are collected using a process called tier system. In this study, the tier system is consisting of four tier systems as follows:

- Tier 1 represents data of immediate block in the same layer

- Tier 2 represents data from the immediate block in the above layer

- Tier 3 represents data from the immediate block in the bottom layer

- Tier 4 represents data from the surrounding blocks in the same layer

4.3 Data Sampling

Mining data from 31200 grid blocks generates an exceedingly large data set of 1.5 million records. It is difficult to handle such a large data set with limited resources (time and computational tools). Therefore, a data sampling technique was applied in this study.

Table 4.1: Data Selected to Develop the Database

Data Type	
Static Data	Dynamic Data
Grid Location	Grid Injection Rate
Grid Top	Grid Injection Cumulative
Grid Thickness	Grid Production Rate
Grid Porosity	Grid Production Cumulative
Grid Permeability	Grid Pressure
Distance to Injection and Boundaries	Grid Saturation
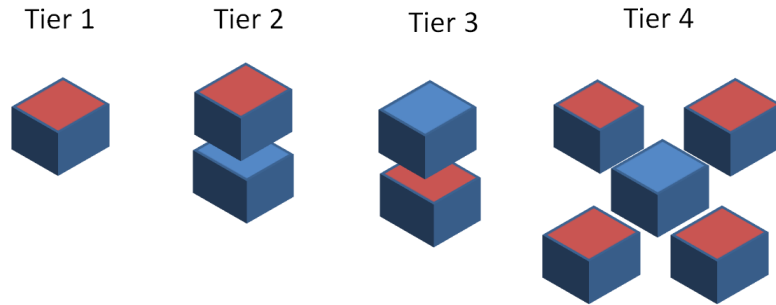
Figure 4.3: Tier System

In detention, sampling involves using a small data set out of a larger data set in an attempt to draw conclusions about the whole data set [82]. In the literature, there are different techniques for performing data sampling. In this study, two sampling approaches were used. The first approach was random sampling. In this technique, population members have an equal chance of being selected. Each subject is selected independently of the other members of the population [83]. The other approach is called the smart sampling technique. In this sampling approach, a histogram of the targeted output is plotted. Then the output distribution is divided based on the range of the values. In this data sampling process, the output value represented by a high number of data points will have a lower percentage of data sampling. On the other hand, the values with less data will have a higher proportion of the data sampled. This sampling process will give the proxy model the required data heterogeneity for better training. Figure 4.4 explains the smart sampling procedure.
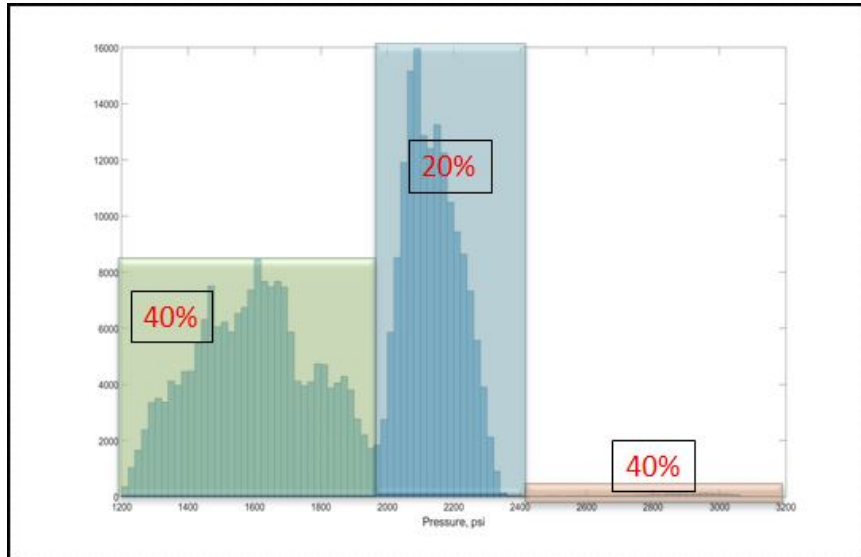
Figure 4.4: Smart Sampling Technique for Pressure Data
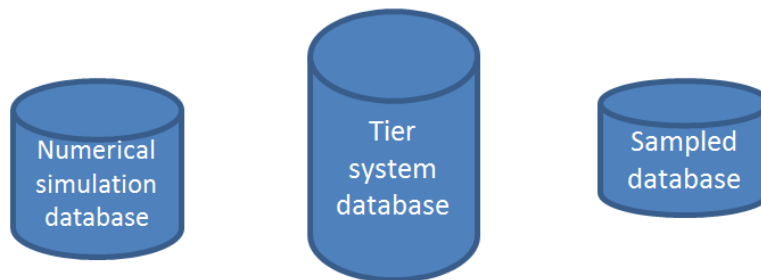
Figure 4.5: Database Generation Represented by the Size of Data

4.4   Data Partitioning

As mentioned above, the database was generated to be used for neural network training in order to develop the smart proxy model. When training artificial neural networks, the database is divided into three main data sets. The first set is the training set, which is used to train the network. The second set is the test set, which is used to stop the training when overfitting takes place. The third set is the validation set, which measures the accuracy of the trained network [84].

It is very important to study the database set in order to determine which partitioning strategy should be performed. For the current studys database, a random partitioning technique was used. Random partitioning divided the database so that 70% of the database was assigned to the training set, 15% to the test set, and 15% to the validation set.

4.5   Artificial Neural Network Construction

An ANN consists of a network of neurons (also known as nodes) that are connected to each other. There are three types of nodes in an ANN, input nodes (input layer), hidden nodes (hidden layer), and output nodes (output layer). The input layer corresponds to the input features, so they have as many nodes as there are selected inputs. The output layer consists of one node, denoting the property being modeled. The hidden layer entails the nodes in which the computations happen. The number of the hidden layers in this study was set to double the number of input nodes.

A neural network was designed for each of the three reservoir properties investigated in this work (pressure, oil saturation, and water saturation). The algorithm used to construct the neural networks was backpropagation. In this algorithm, the error for each output is propagated backwards to the input in order to adjust the weights in each layer of the neural network [85]. Once the network has been constructed, the training process starts and the network performance can be monitored using several visualization plots in the software [79].

4.6   Validation with a Blind Data Set

Once the neural network has been trained to satisfaction, the last step in smart proxy development is to verify the model with blind data sets. A blind data set is defined as a numerical simulation of

input/output data which has not been used in the training of the neural network.

4.7 Error Measurement

The developed smart proxy model is validated by applying the model to a blind set. It is important to determine the precision of the model to the blind set. In this study, the precision is determined by measuring the error between the numerical simulator output and the smart proxy model output. The developed smart proxy is generating output at each grid block, so the error should be calculated at each grid block.

Different error calculations formulas are used based on the nature of the output data. For the pressure output, the following error formula is used:

$$\text{Absolute Error Percentage} = [(|\text{Artificial Neural Network Output} - \text{Numerical Simulator Output}|)/$$
$$\text{Numerical Simulator Output}] * 100$$

$$(4.1)$$

For the reservoir saturation data, the nature of data is different. The values are between 0 and 1, therefore the following error formula is used:

$$\text{Absolute Error Percentage} = (|\text{Artificial Neural Network Output} - \text{Numerical Simulator Output}|) * 100$$

$$(4.2)$$

Chapter 5

Smart Proxy to Replicate Numerical Simulator

The objective of this chapter is to test the ability of the smart proxy model technique to replicate the numerical simulation results. The geological heterogeneity and flow performance heterogeneity of the SACROC field create an ideal situation for testing whether this proxy modeling technique is able to mimic the results of the numerical simulation.

As discussed in chapter 4, it is the smart proxy objective that defines the type and the scale of the proxy model. Because the objective in this chapter is to test the ability of the smart proxy to mimic the results from the numerical simulator, only one numerical simulation run was designed for each smart proxy.

Generally, the production/injection performance of the SACROC field in the studied area is divided into two phases, the first 20 years of the production/injection when there was only one injection well in the field and the second stage, from 1972, when a number of production wells was converted into injection wells. Therefore, a smart proxy was generated for each phase.

5.1 Phase One Smart Proxy

The SACROC Unit initially produced under natural solution gas drive. When the reservoir pressure declined dramatically, the pressure maintenance program using water injection was introduced to

the field in 1953. In the study area, only one well had been converted to a water injection well (well 17-5) between 1955 and 1972. The water injected into the reservoir brought the reservoir pressure from 1300 psi to 2000 psi by the end of 1960. The total number of producing wells during this stage was 13.

5.1.1 Training Model Construction

The static and dynamic reservoir data collected to develop the smart proxy model came from one simulation run (history-matched model). The database was generated using the geo-cellular model data and the data from flow performance between 1953 and 1963. Given the reservoir model dimension, there are 312,000 data records (31,200 grids X 10 time steps). The software used for training is not capable of handling this amount of data. Therefore, data sampling was applied to the input data. As mentioned above, it is very important to make sure that the sampled data is representative of the full data set. For instance, Figure 5.1 shows all data pressure points compared with the sampled pressure data points. It is clear from the histograms that the sampled data is an excellent representation of the complete data set.
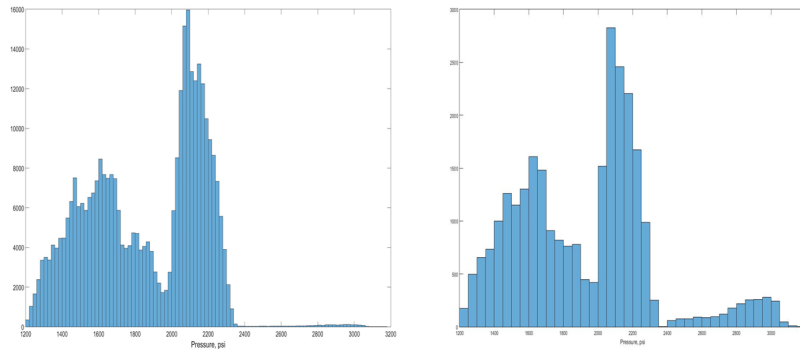
Figure 5.1: All Pressure points versus sampled points histograms, for years from 1953 to 1963

From this production phase, only 28,000 data points out of 312,000 were used for training. In other words, only 9% of the data was used in the training. The training data was divided into three parts, 70% for training (19,600 data records), 15% for calibration (4,200 data records), and 15% for validation (4,200 data records).

The ANN was constructed in three layers, an input layer, a hidden layer, and an output layer. The input layer contains 69 selected parameters. The hidden layer has 120 nodes and there was one output in the output layer.
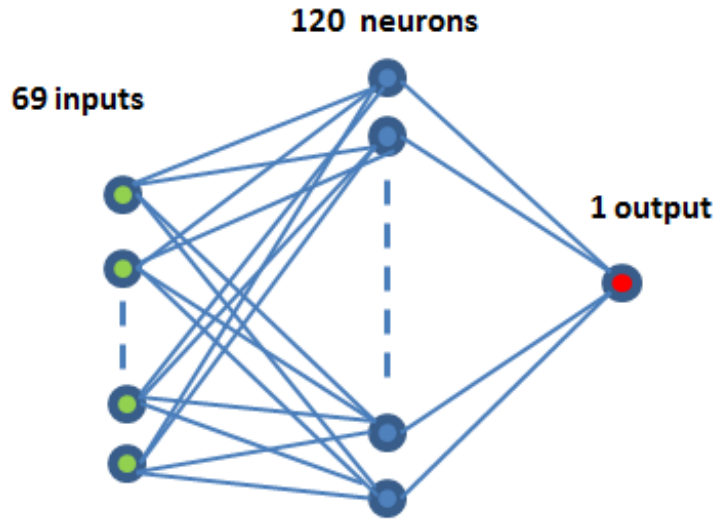
Figure 5.2: The ANN architecture for training the reservoir pressure and saturation

5.1.2 Training Results

As mentioned in the previous chapter, the ANN was designed based on the smart proxy development procedure. Three ANN were designed for the three targeted reservoir properties, grid pressure, oil saturation, and water saturation. In the ANN, the input layer contains 69 selected parameters, 120 nodes in the hidden layer, and one output node in the output layer. The network was then trained, calibrated, and tested for the targeted reservoir property. Table 5.1 shows the training results of the targeted reservoir properties.

Reservoir Property	R-squared coefficient
Grid Pressure	99%
Oil Saturation	99%
Water Saturation	99%

Table 5.1: Phase One Smart Proxy Training Results

Parameters	
D1- Distance to boundary	Oil_Saturation T4(t-1)
D2- Distance to boundary	Vertical Permeability_T1
D3- Distance to boundary	Vertical Permeability_T2
D4- Distance to boundary	Vertical Permeability_T3
Distance to Injection 1	Vertical Permeability_T4
Distance to Injection 2	Permeability_I T1
Distance to Production 1	Permeability_I T2
Distance to Production 2	Permeability_I T3
First Injection Well Cumulative Water SC (bbl)	Permeability_I T4
First Injection Well Cumulative Water SC (bbl)(T-1)	Porosity T1
First Injection Well Water Rate SC (bbl/day)	Porosity T2
First Injection Well Water Rate SC (bbl/day)(T-1)	Porosity T3
First Production Well Cumulative Gas SC (ft3)(T-1)	Porosity T4
First Production Well Cumulative Oil SC (bbl)(T-1)	Pressure T1(t-1)
First Production Well Cumulative Water SC (bbl)(T-1)	Pressure T2(t-1)
First Production Well Gas Rate SC (ft3/day)(T-1)	Pressure T3(t-1)
First Production Well Oil Rate SC (bbl/day)(T-1)	Pressure T4(t-1)
First Production Well Pressure (psi)(T-1)	Second Injection Well Cumulative Water SC (bbl)
First Production Well Water Rate SC (bbl/day)(T-1)	Second Injection Well Cumulative Water SC (bbl)(T-1)
Grid No	Second Injection Well Water Rate SC (bbl/day)
Grid_Thickness T1	Second Injection Well Water Rate SC (bbl/day)(T-1)
Grid_Thickness T2	Second Production Well Cumulative Gas SC (ft3)(T-1)
Grid_Thickness T3	Second Production Well Cumulative Oil SC (bbl)(T-1)
Grid_Thickness T4	Second Production Well Cumulative Water SC (bbl)(T-1)
Grid_Top T1	Second Production Well Gas Rate SC (ft3/day)(T-1)
Grid_Top T2	Second Production Well Oil Rate SC (bbl/day)(T-1)
Grid_Top T3	Second Production Well Pressure (psi)(T-1)
Grid_Top T4	Second Production Well Water Rate SC (bbl/day)(T-1)
I	Water_Saturation T1(t-1)
J	Water_Saturation T2(t-1)
K	Water_Saturation T3(t-1)
Oil_Saturation T1(t-1)	Water_Saturation T4(t-1)
Oil_Saturation T2(t-1)	X- Coordinates
Oil_Saturation T3(t-1)	Y- Coordinates
t: time	
T: Tier	

Figure 5.3: Selected Parameters to Develop the Smart Proxy Model

5.1.3 Validation with a Blind Data Set

The smart proxy model was developed for the years from 1953 to 1963 for the targeted reservoir properties (pressure, oil saturation, and water saturation). The studied area of the field had the same production performance for the 10 years following that (up to 1970). The ideal way to test the ability of the developed smart proxy model is to apply it to a blind data set that has never been seen by the trained model.

The model was applied to selected blind data sets. Static and dynamic reservoir data from the years 1964, 1965, and 1968 were used as blind data sets for the smart proxy model. The results showed an excellent match between the numerical simulation model and smart proxy model. The absolute average error between the two models was 1%. Some of the blind sets results are shown as data maps images in Figures 5.4 to 5.12.
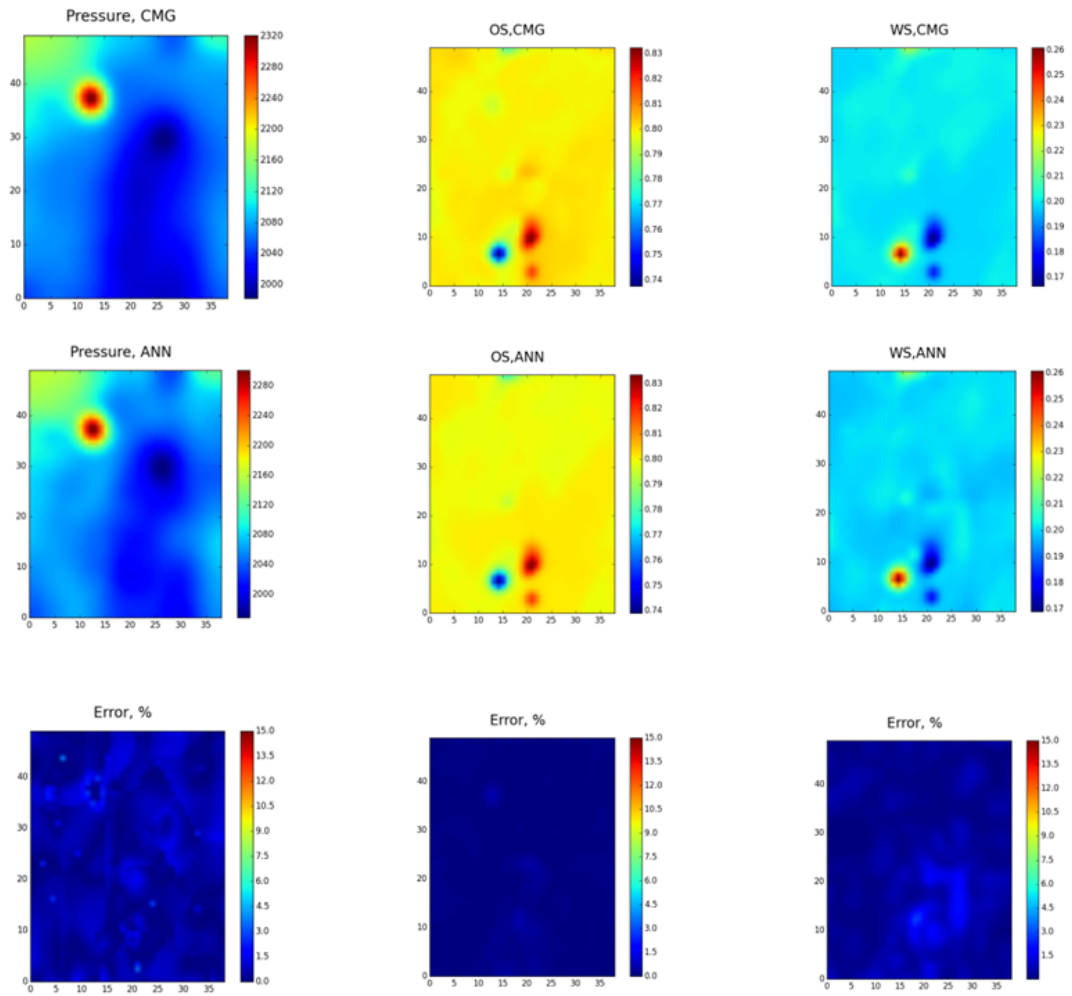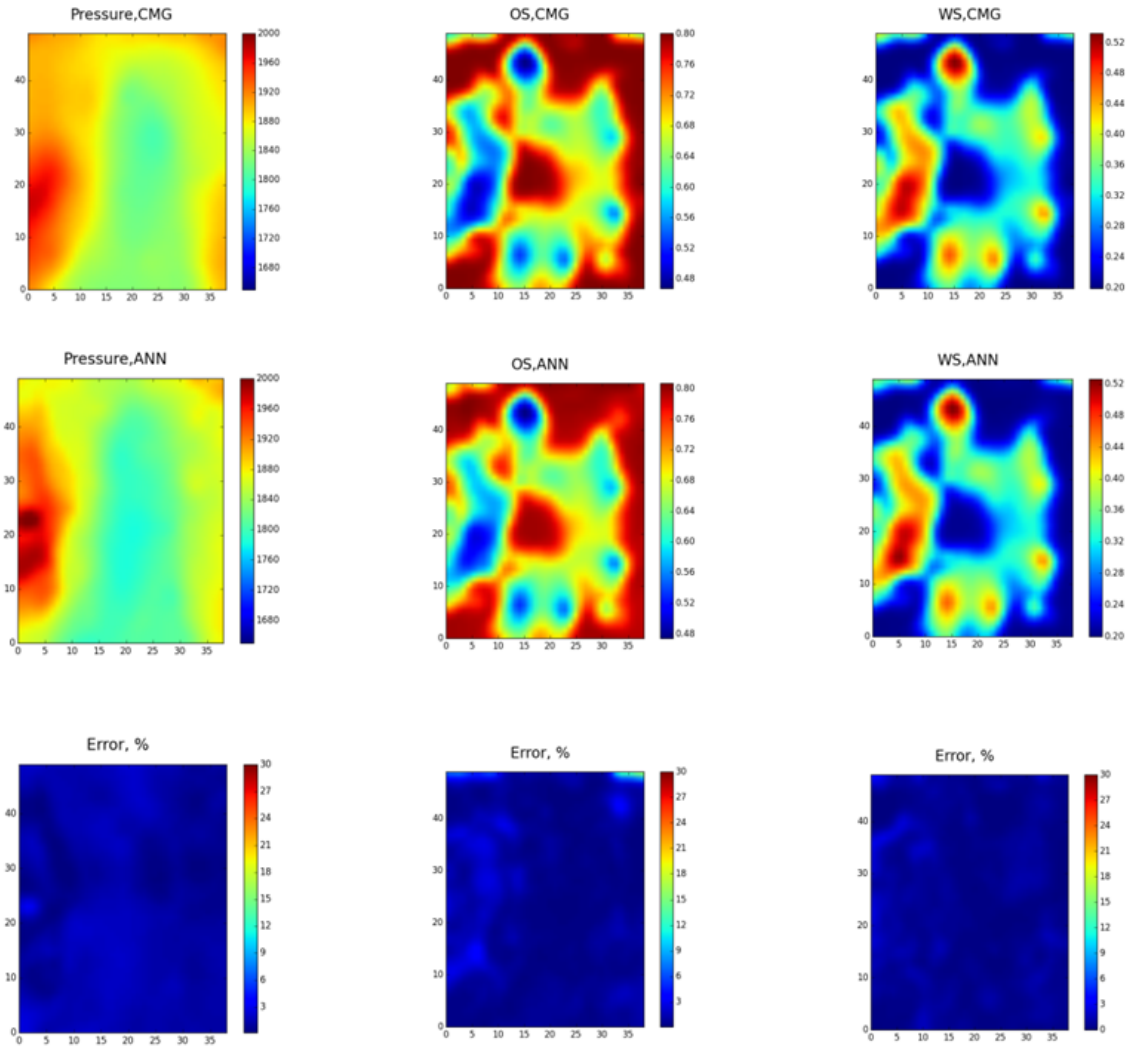
Figure 5.4: Year 1964 Layer 1 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
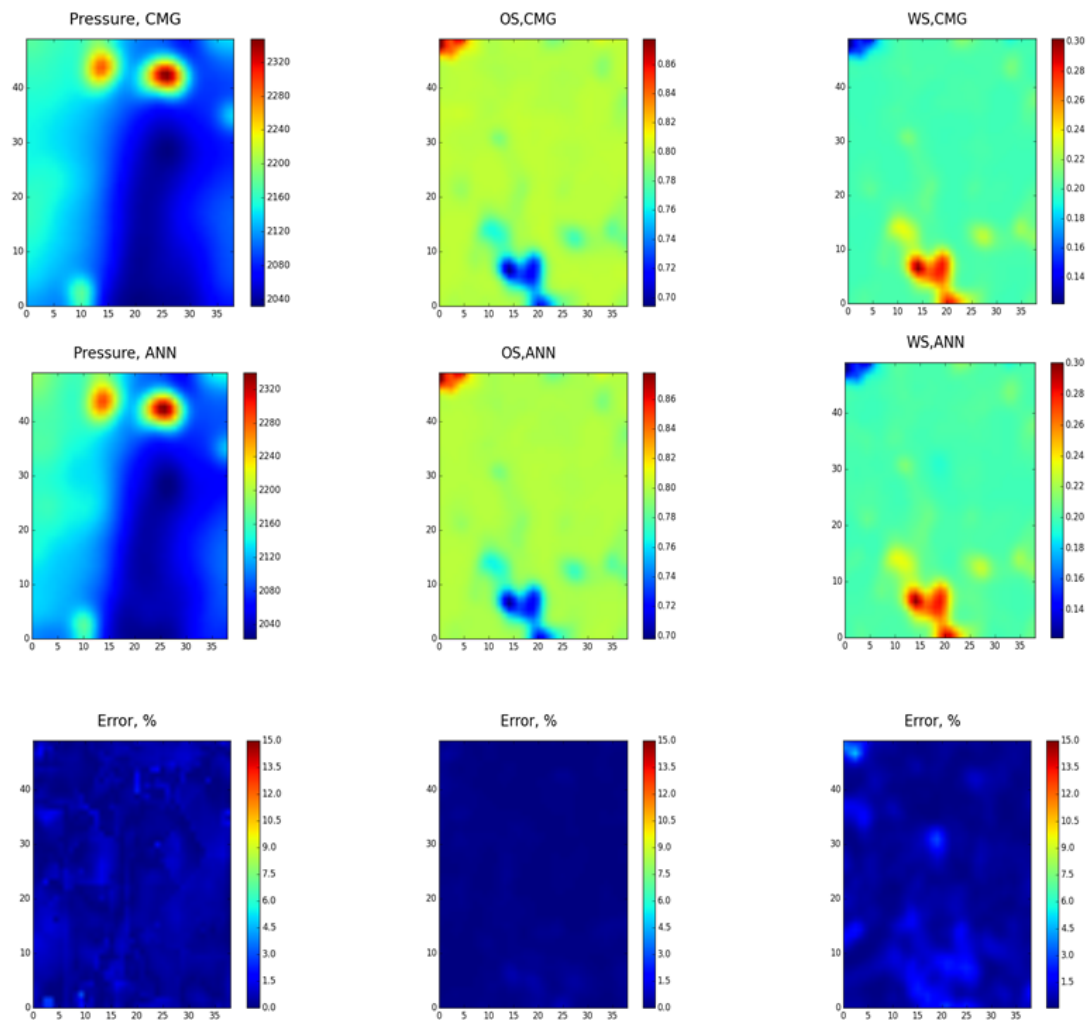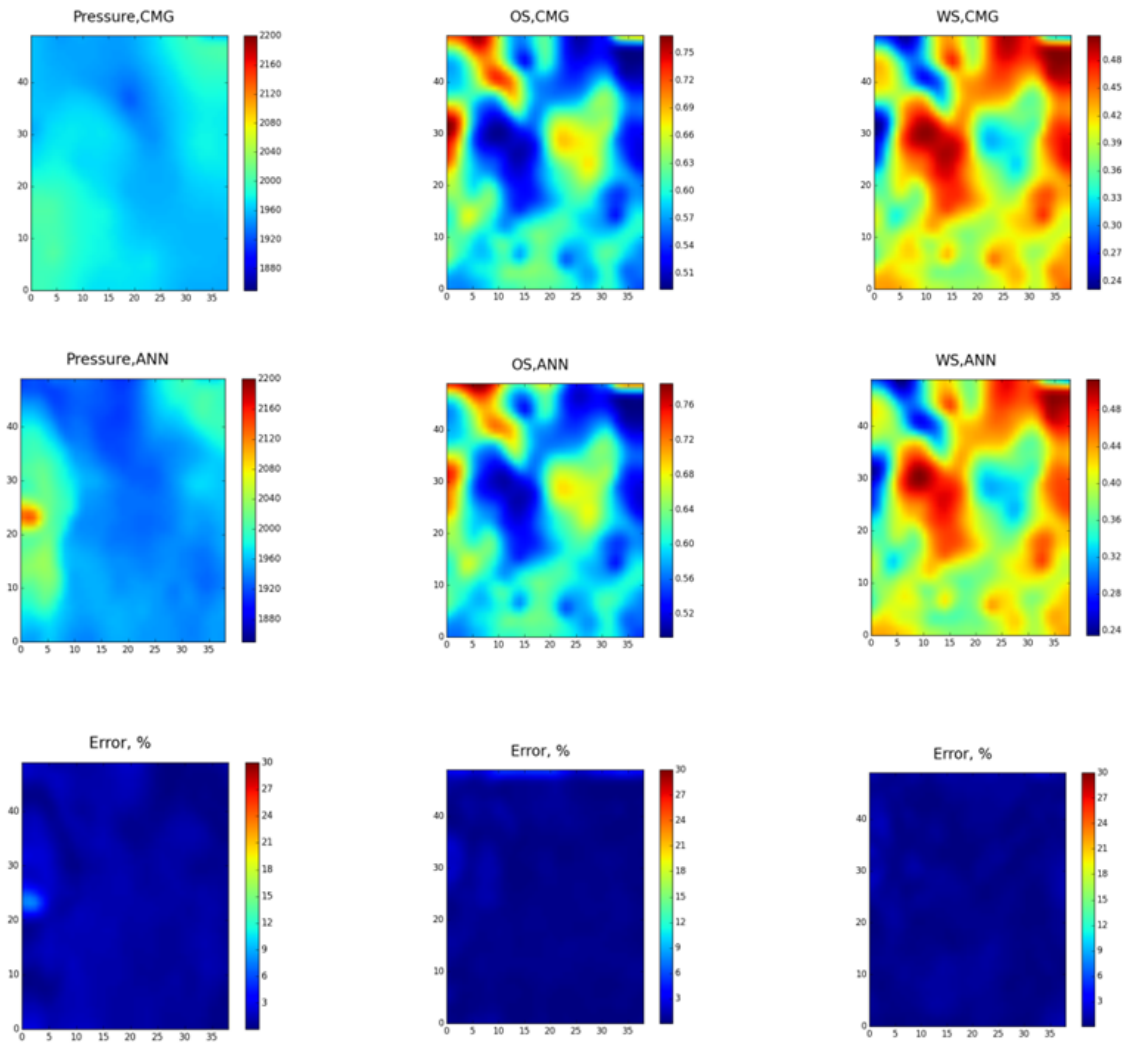
Figure 5.5: Year 1964 Layer 5 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
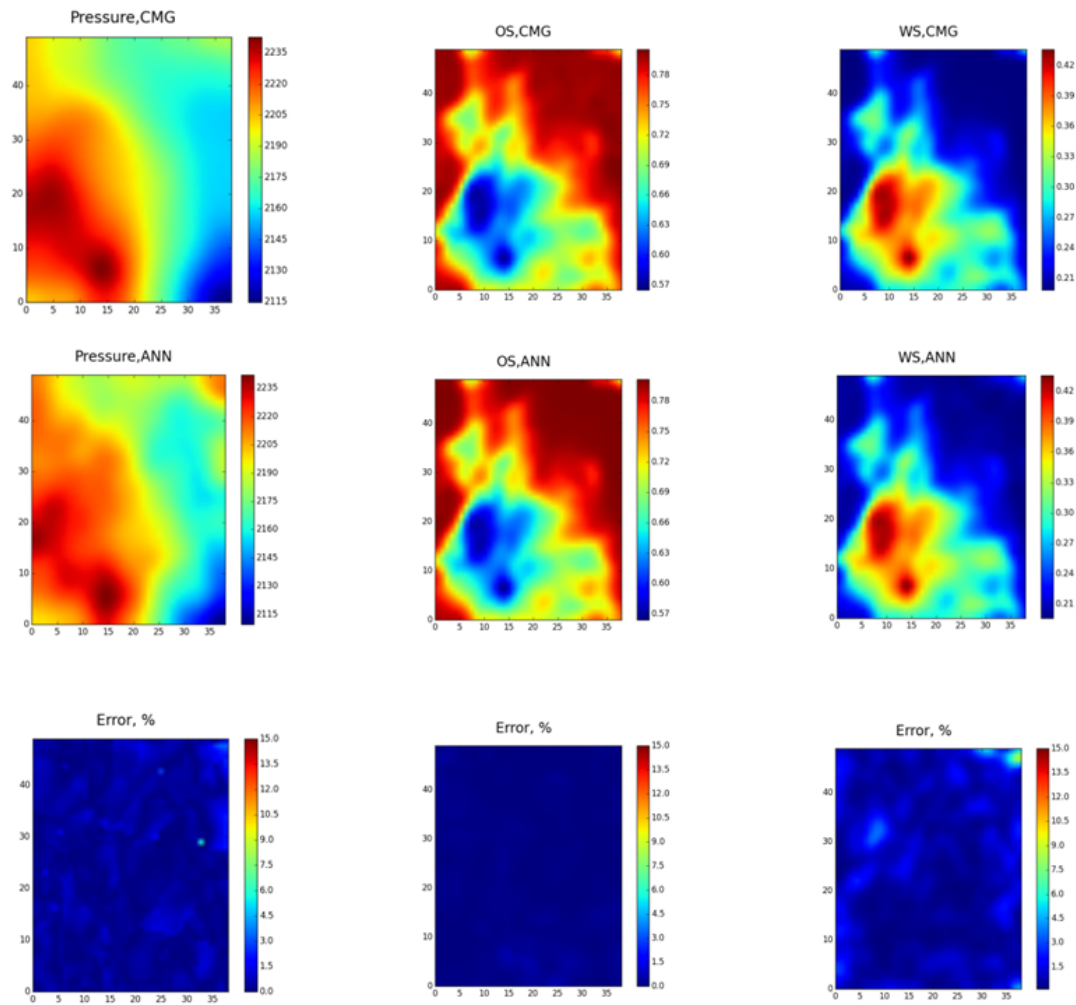
Figure 5.6: Year 1964 Layer 12 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
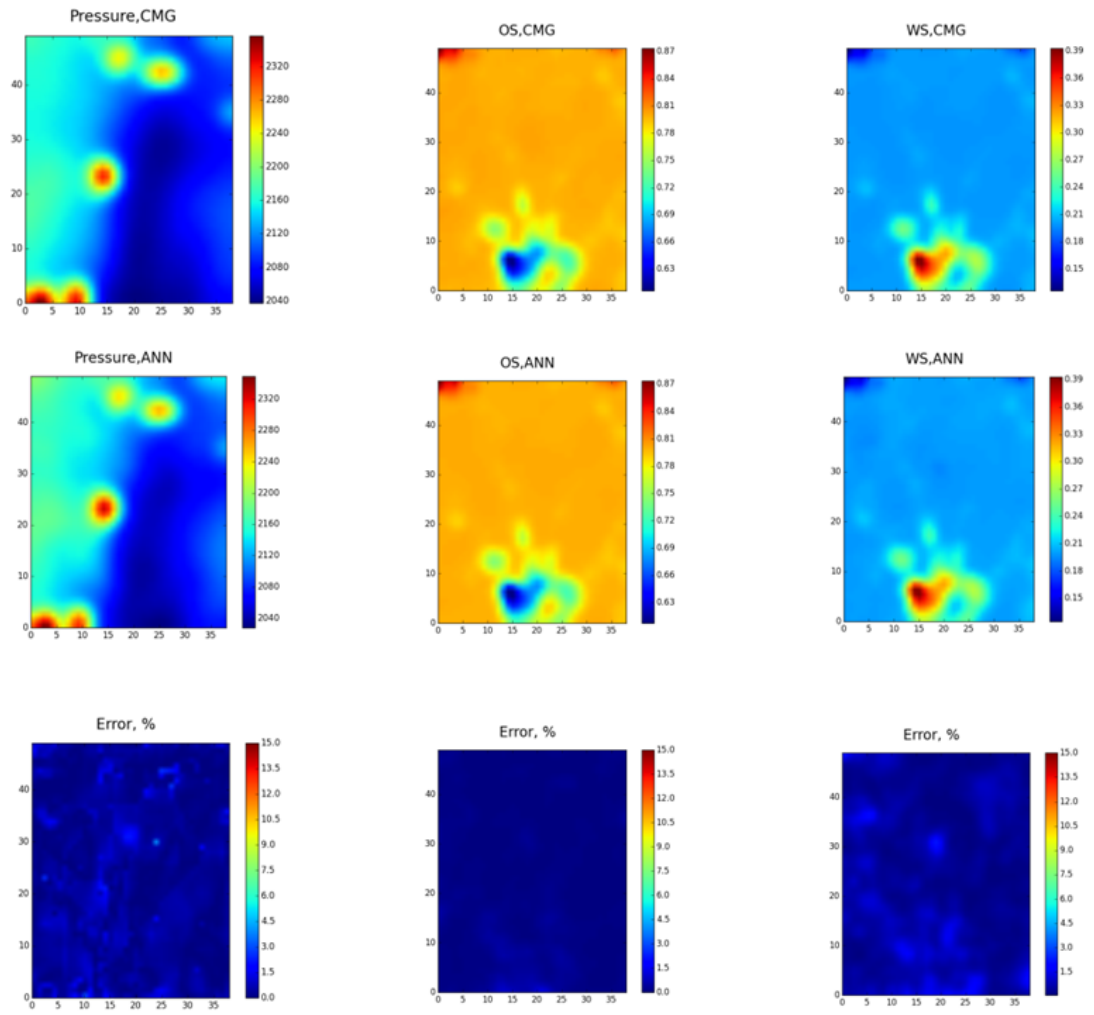
Figure 5.7: Year 1965 Layer 6 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
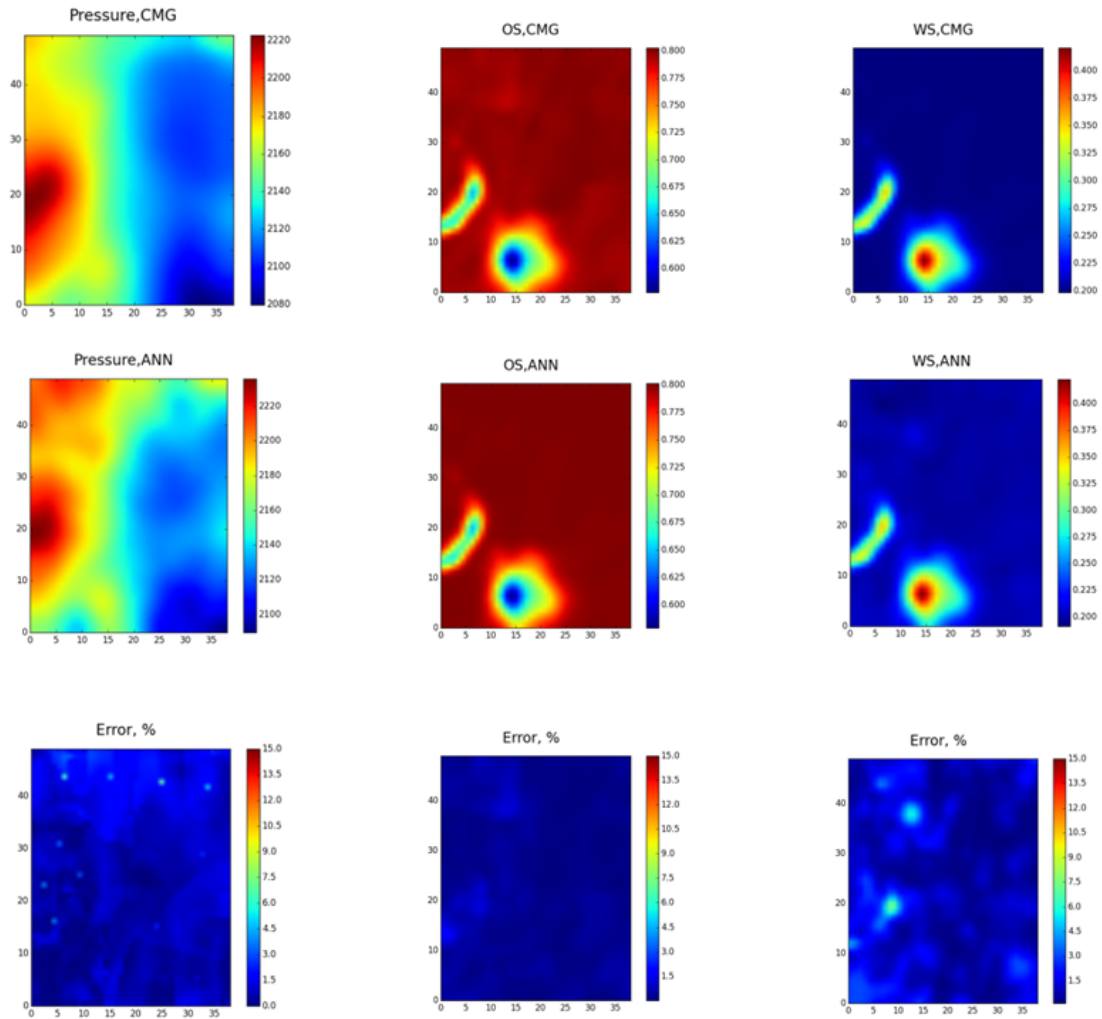
Figure 5.8: Year 1965 Layer 10 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
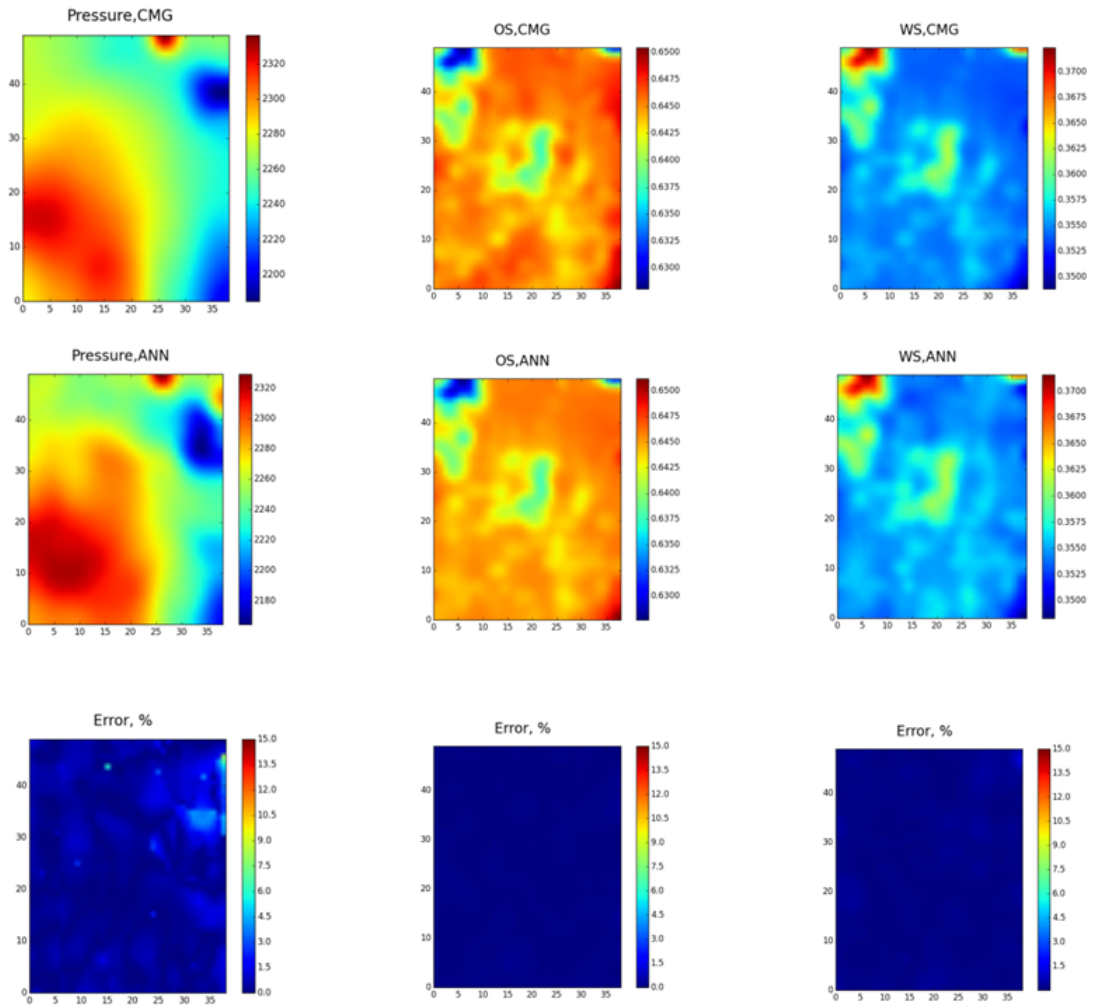
Figure 5.9: Year 1965 Layer 16 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
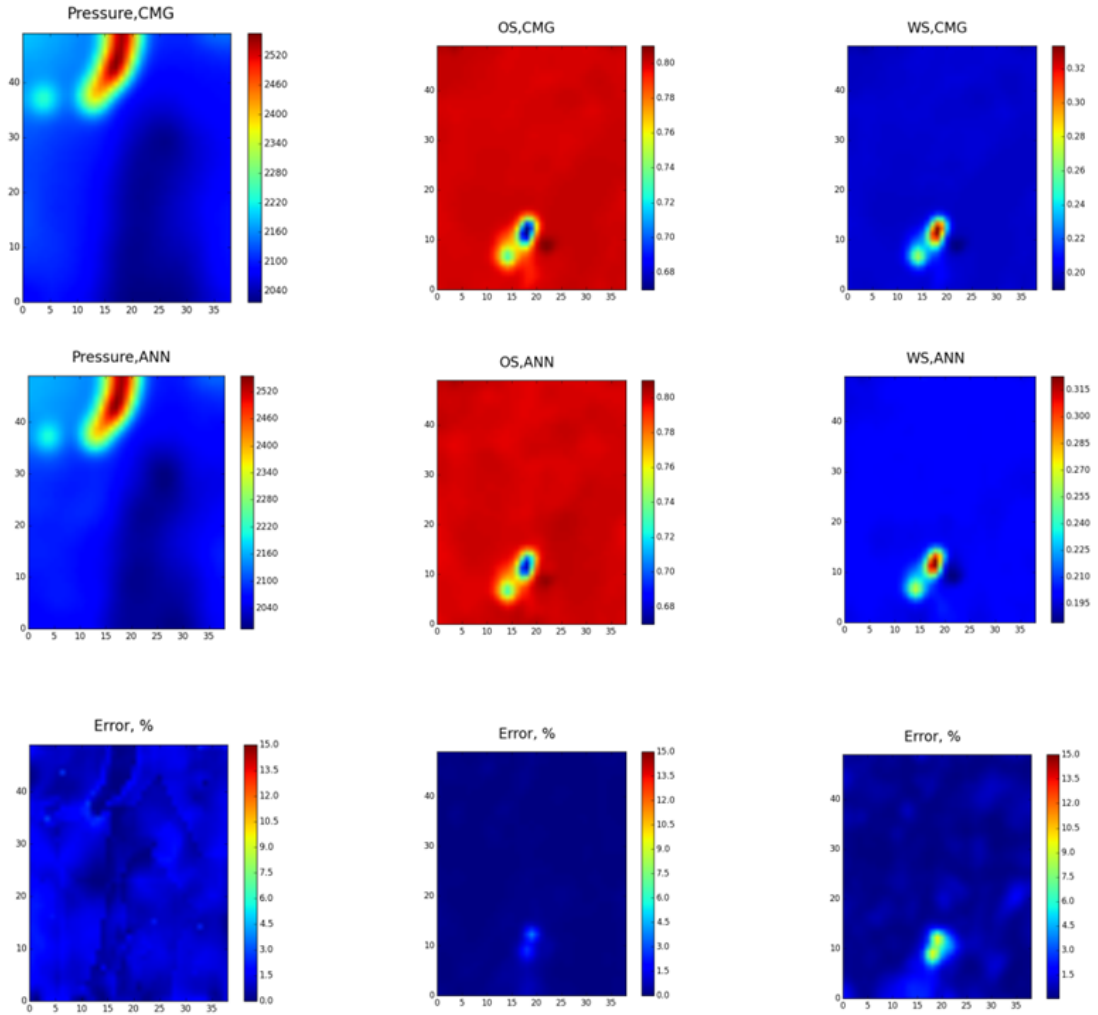
Figure 5.10: Year 1968 Layer 2 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
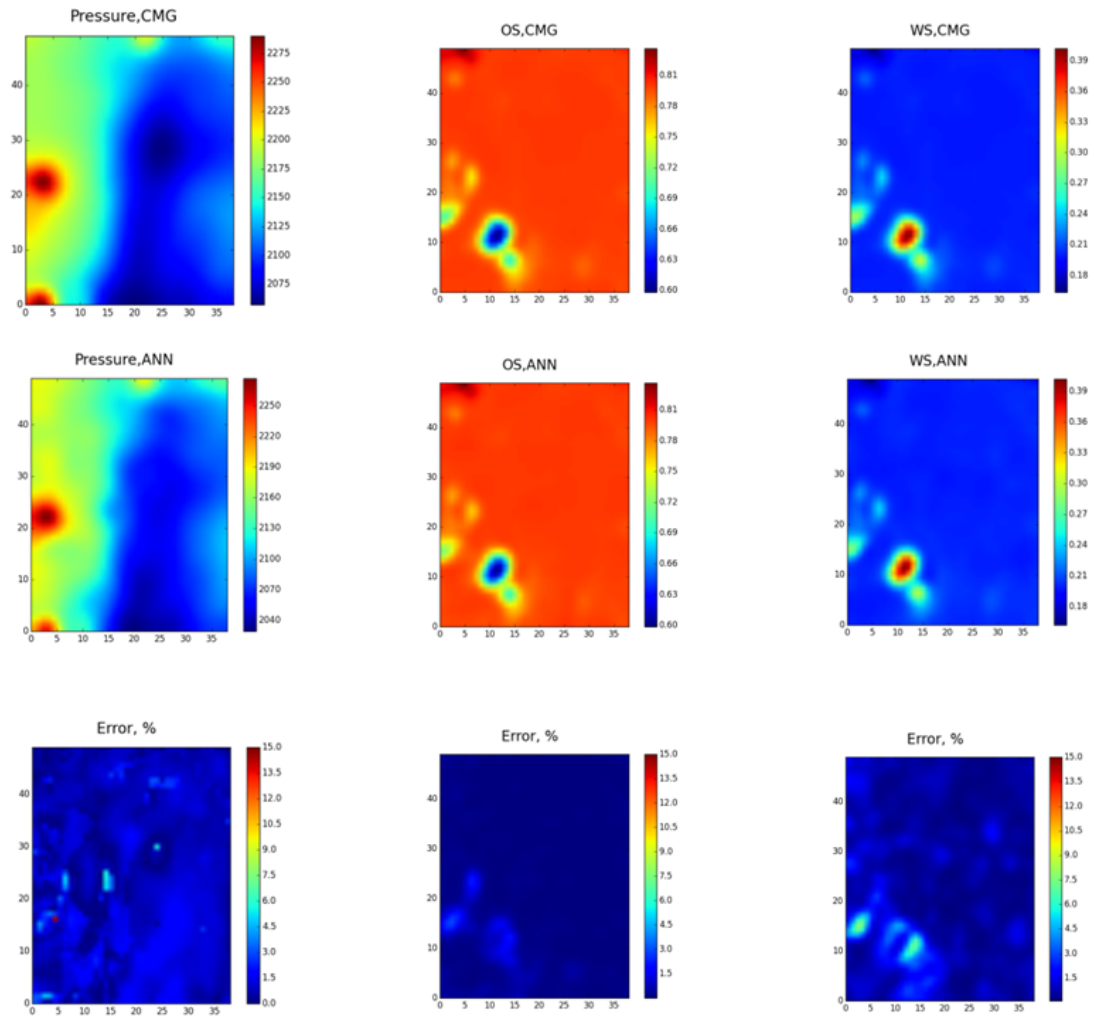
Figure 5.11: Year 1968 Layer 7 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
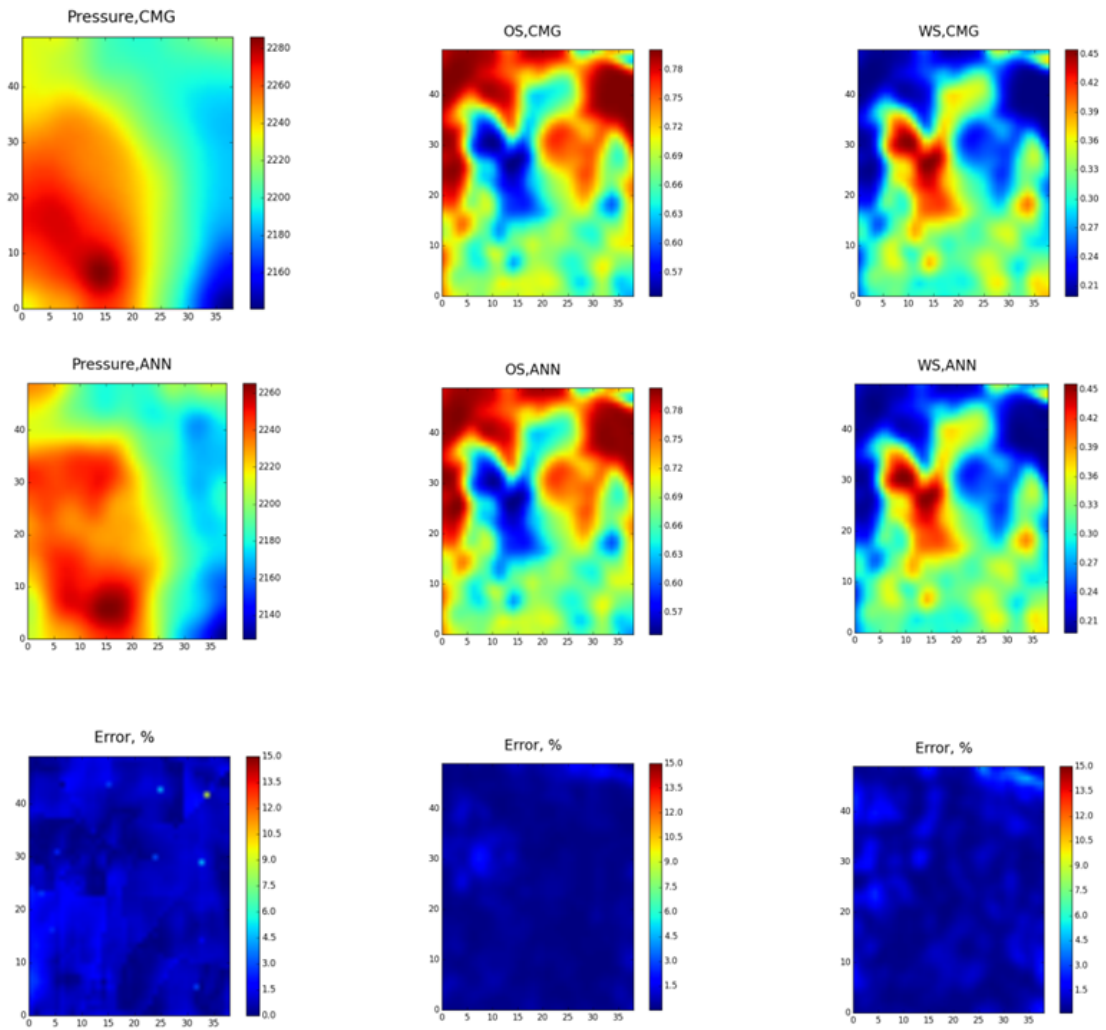
Figure 5.12: Year 1968 Layer 13 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).

Reservoir Property	R-squared coefficient
Grid Pressure	99%
Oil Saturation	99%
Water Saturation	99%

Table 5.2: Phase Two Smart Proxy Training Results

5.2 Phase Two Smart Proxy

The selected second phase was from 1975 to 2003. In this phase, the number of production and injection wells introduced to the field varied each year. This type of behavior makes the production performance very heterogeneous and challenging for smart proxy development.

Two water injection wells were added to the smart proxy model database as in order to monitor the water injection affect at each grid block. Indeed, adding another injection well provides more opportunities for monitoring the reservoirs dynamic property changes due to the injection/production process at every time step.

In this phase, the database generated was exceedingly large and used data from 28 years of injection/production performance (from 1975 to 2003). The database had 900,000 data records. Again, sampling was applied and only 90,000 data points were selected for the ANN training and smart proxy development. Figure 5.13 shows the complete data set versus the sampled data points.

5.2.1 Training Results

In the development of the ANN at this phase, the same neural network structure as in the first production stage was followed. The input layer contained 69 selected parameters, 120 nodes in the hidden layer, and one output node in the output layer.

The neural network was then trained, calibrated, and tested. Table 5.2 shows the training results of the targeted reservoir properties.
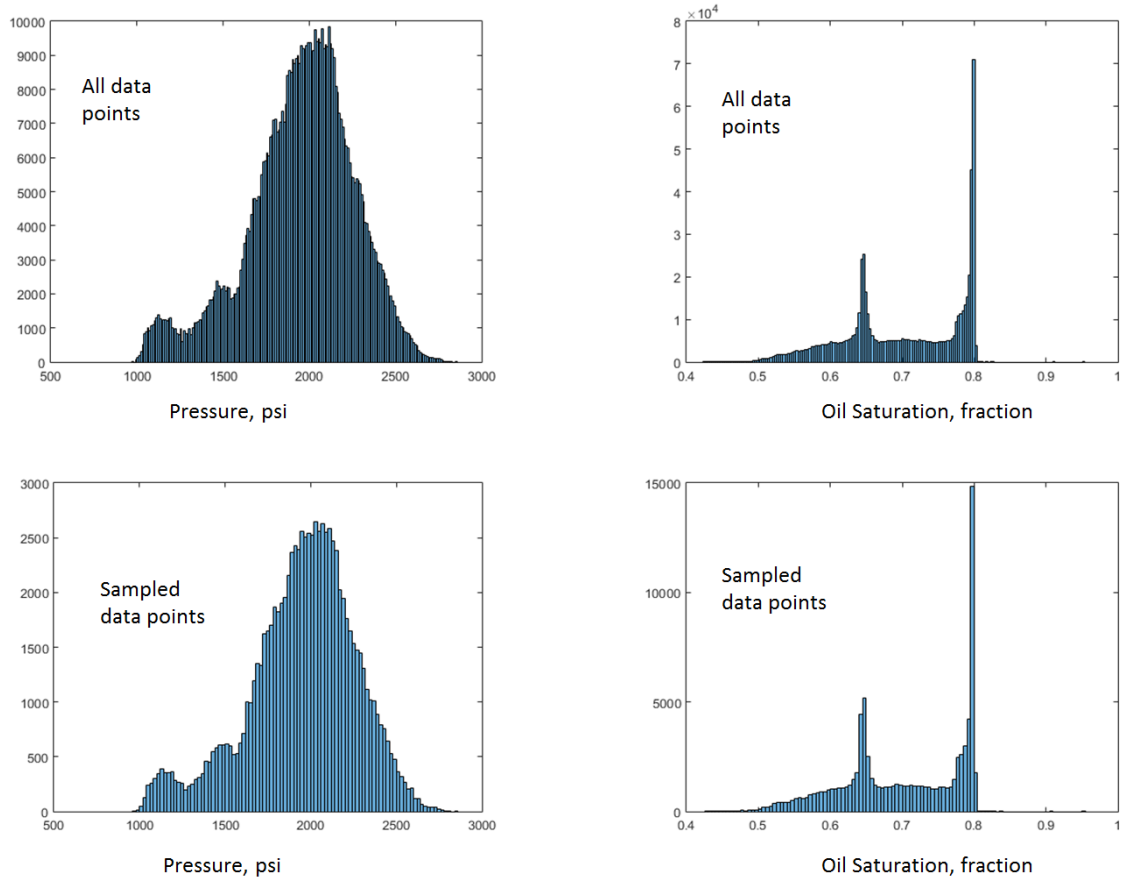
Figure 5.13: All points versus sampled points histograms, Pressure on left side and the Oil Saturation on right side

5.2.2   Validation Results

The smart proxy model was developed using sampled data, which is less than 10% of the entire database. So, 90% of the database had not been seen by the trained model. To ensure the robustness of the developed model, it has to be validated by introducing it to different data sets. As explained in the previous section, the production time period examined in this phase was from 1975 to 2003. In the validation process followed here, the model was applied to all production years and compared to the results from the numerical simulation. The smart proxy validation showed a very vigorous model in which the results are an excellent match with the numerical simulation results.

Due to the long period of production (28 years), only results from three years were selected to be presented in this section. These years were carefully selected based on the production performance of the field during the studied period of time. 1980 is when the field had a maximum number of injection/production wells. In 1992, the number of production wells became lower, and in 2000 the field had the highest number of injection wells.
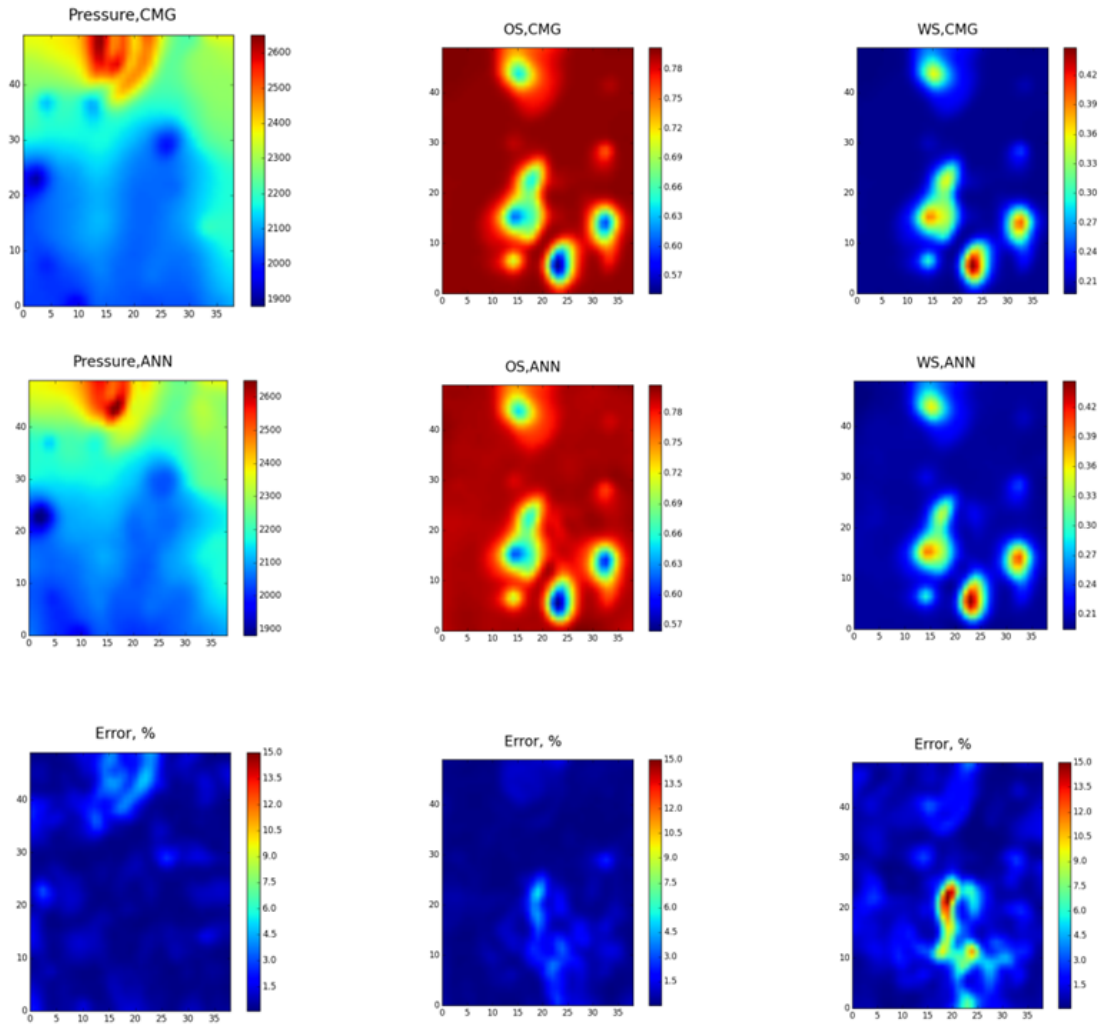
Figure 5.14: Year 1980 Layer 1 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
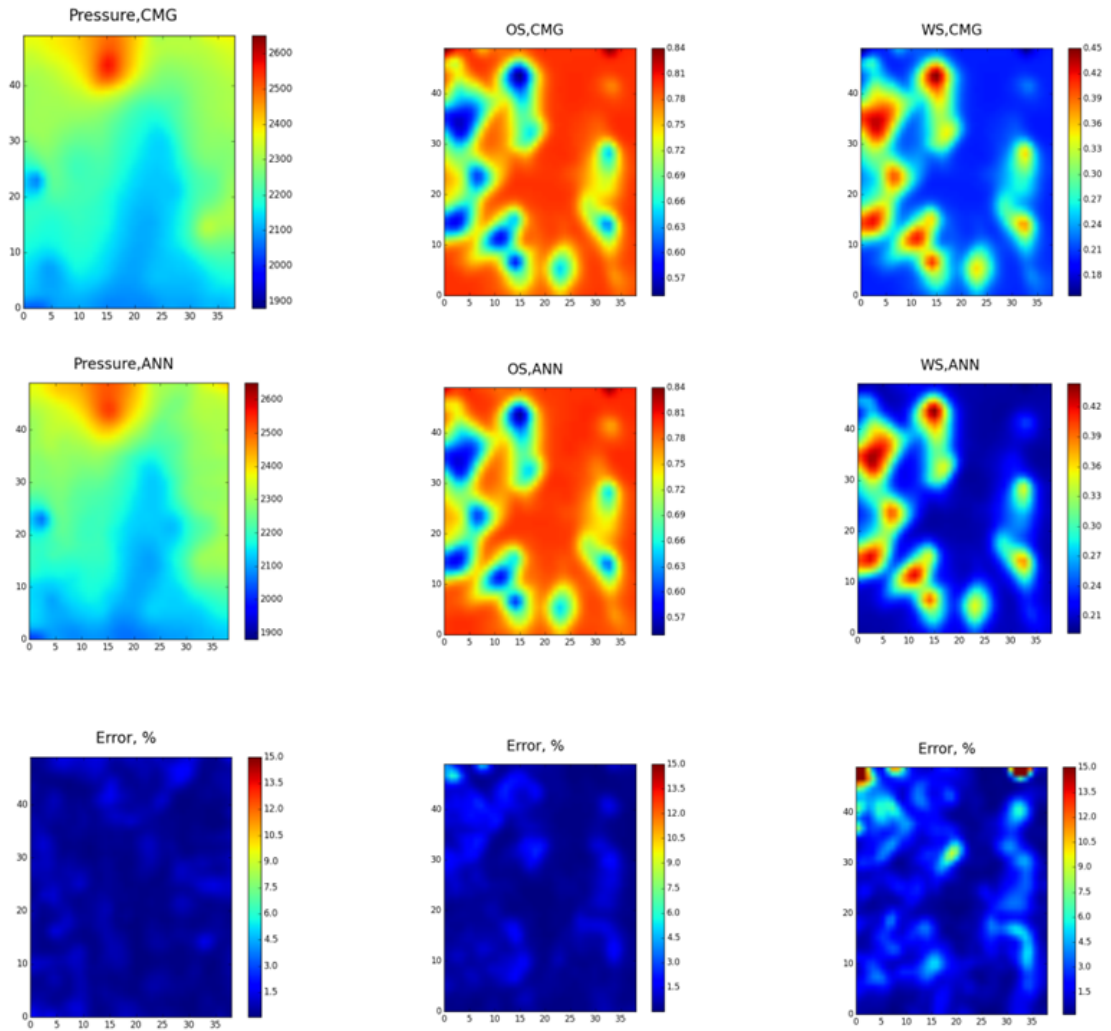
Figure 5.15: Year 1980 Layer 7 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
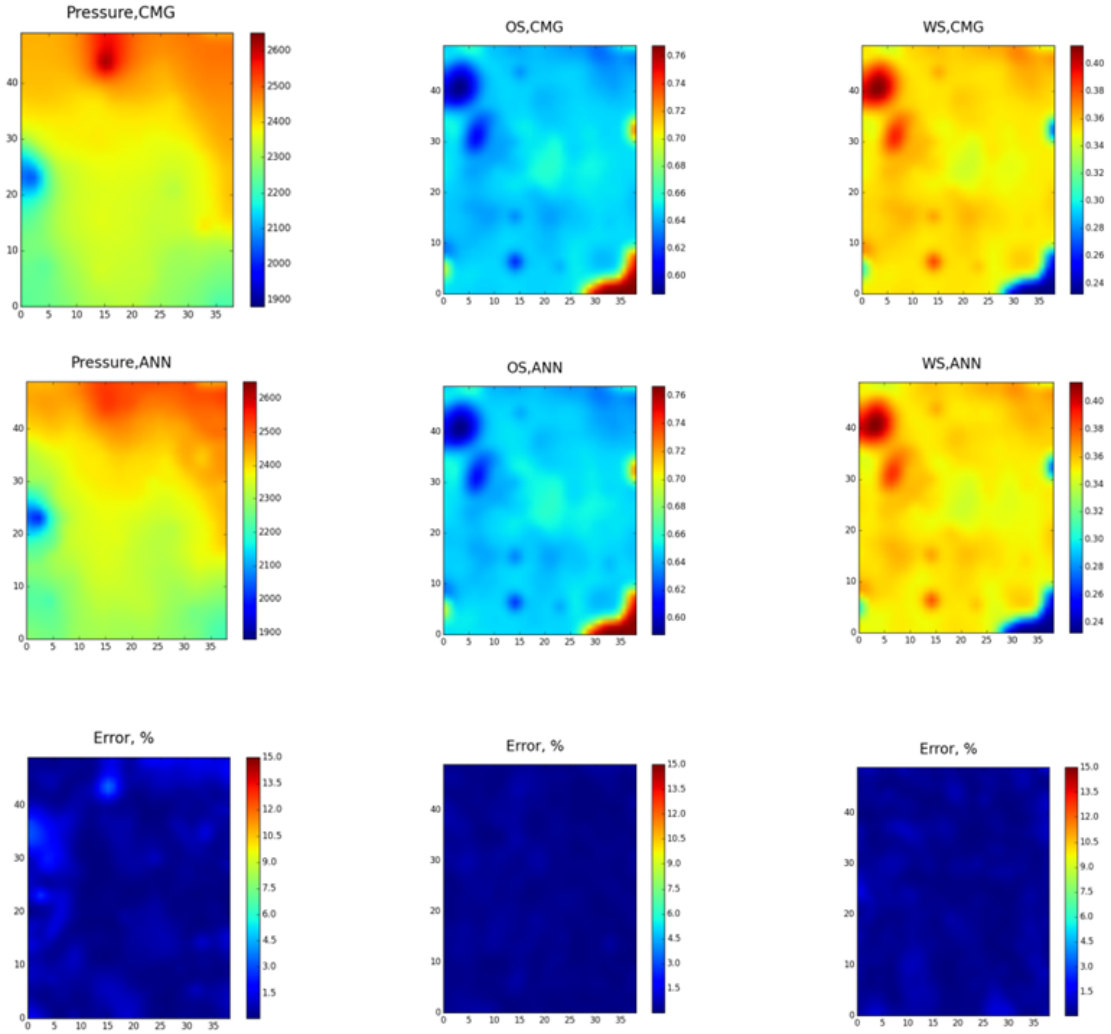
Figure 5.16: Year 1980 Layer 15 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
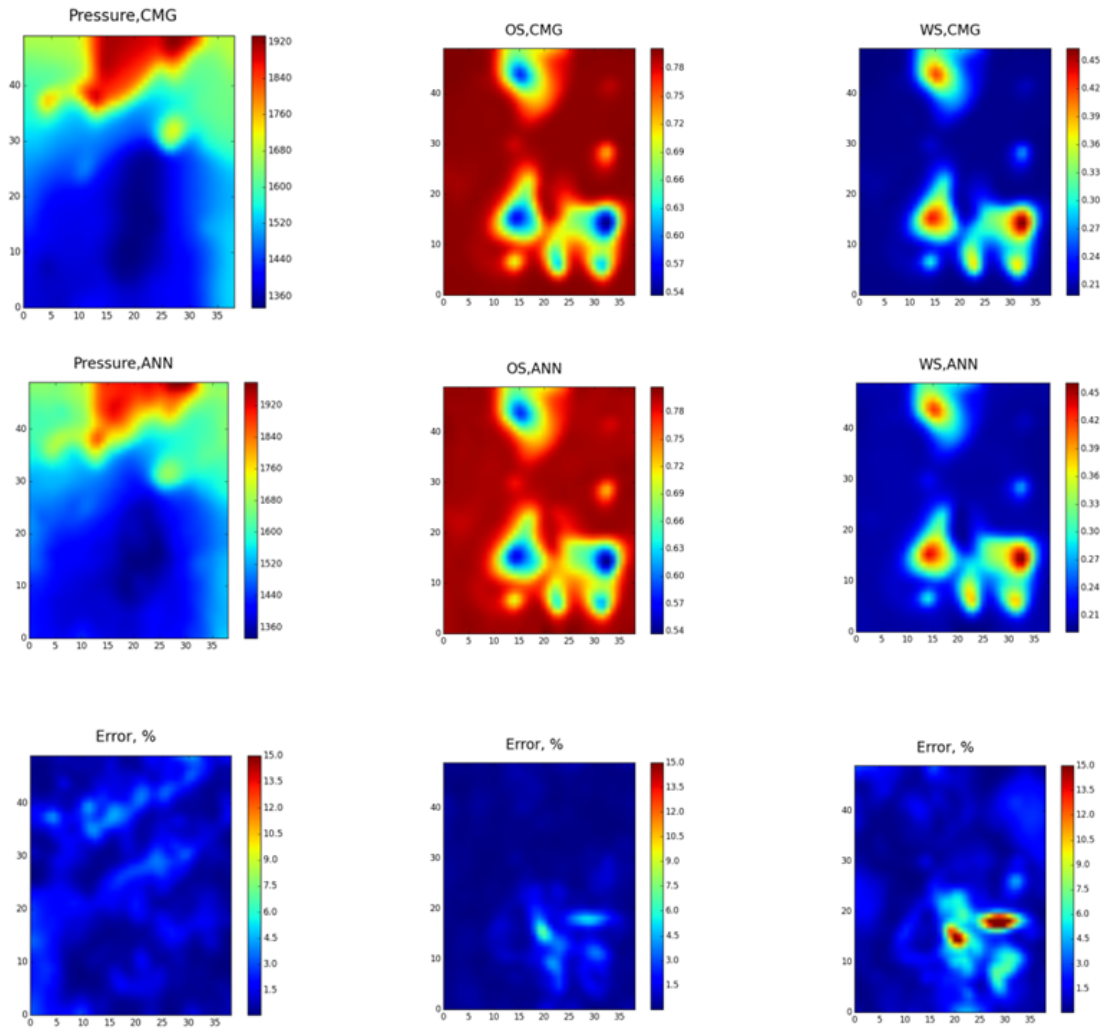
Figure 5.17: Year 1992 Layer 1 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
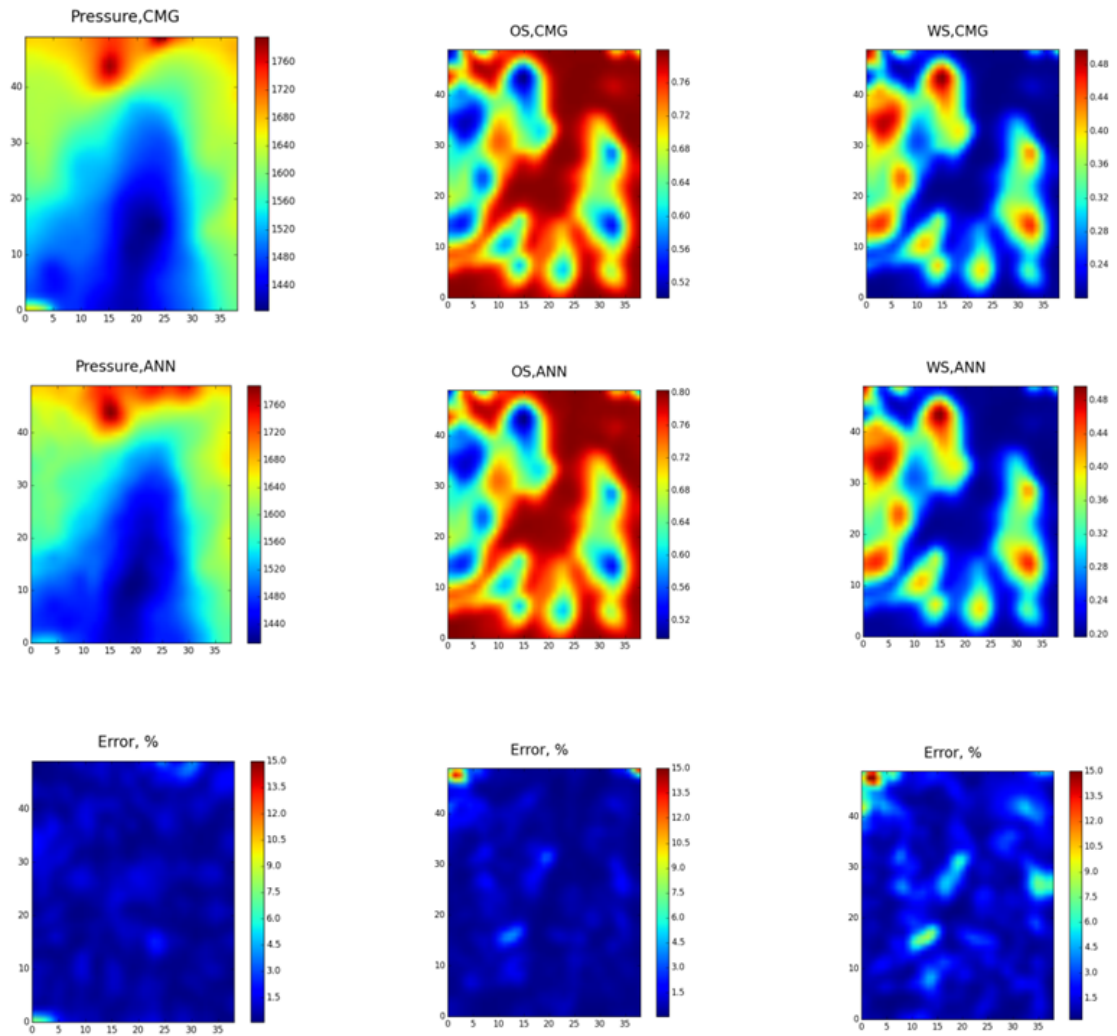
Figure 5.18: Year 1992 Layer 7 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
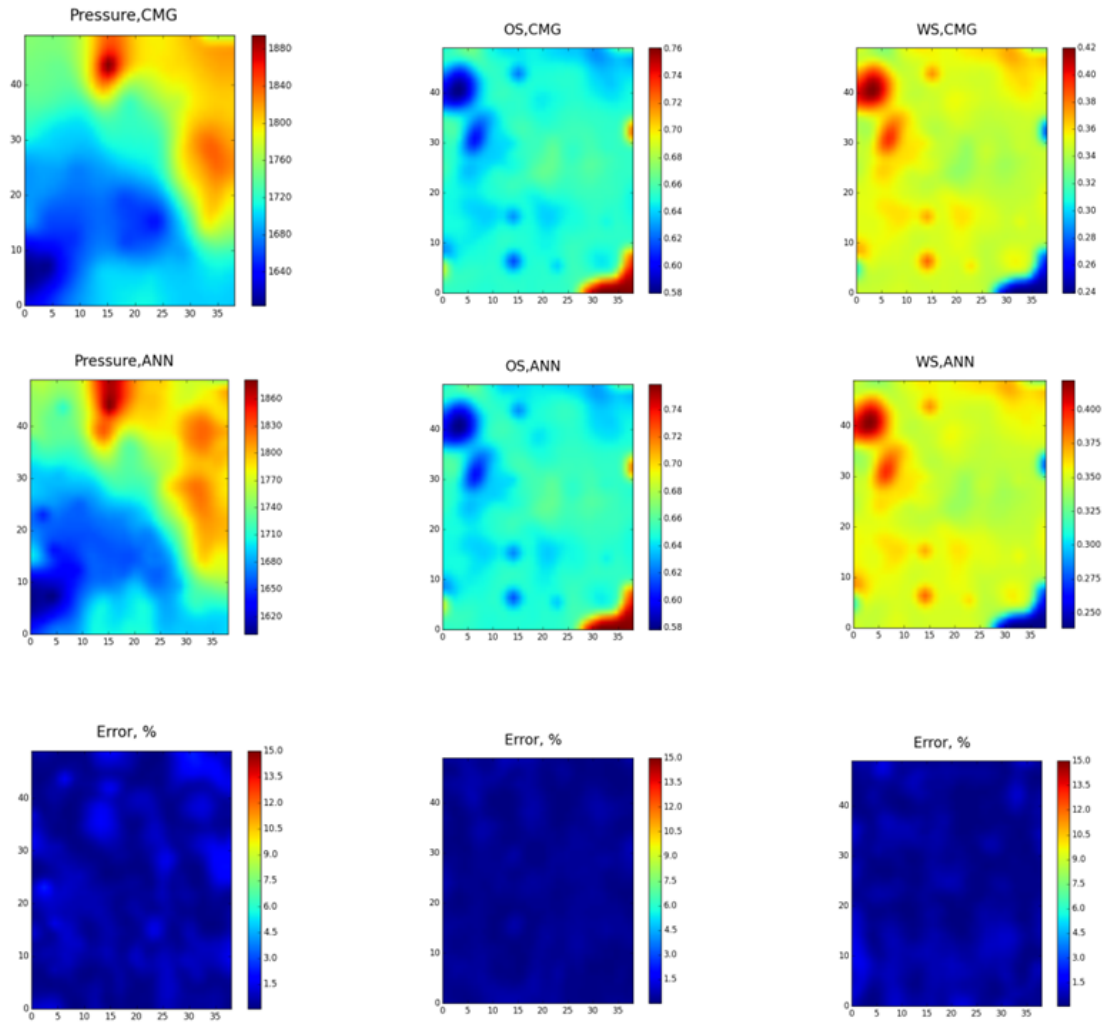
Figure 5.19: Year 1992 Layer 15 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
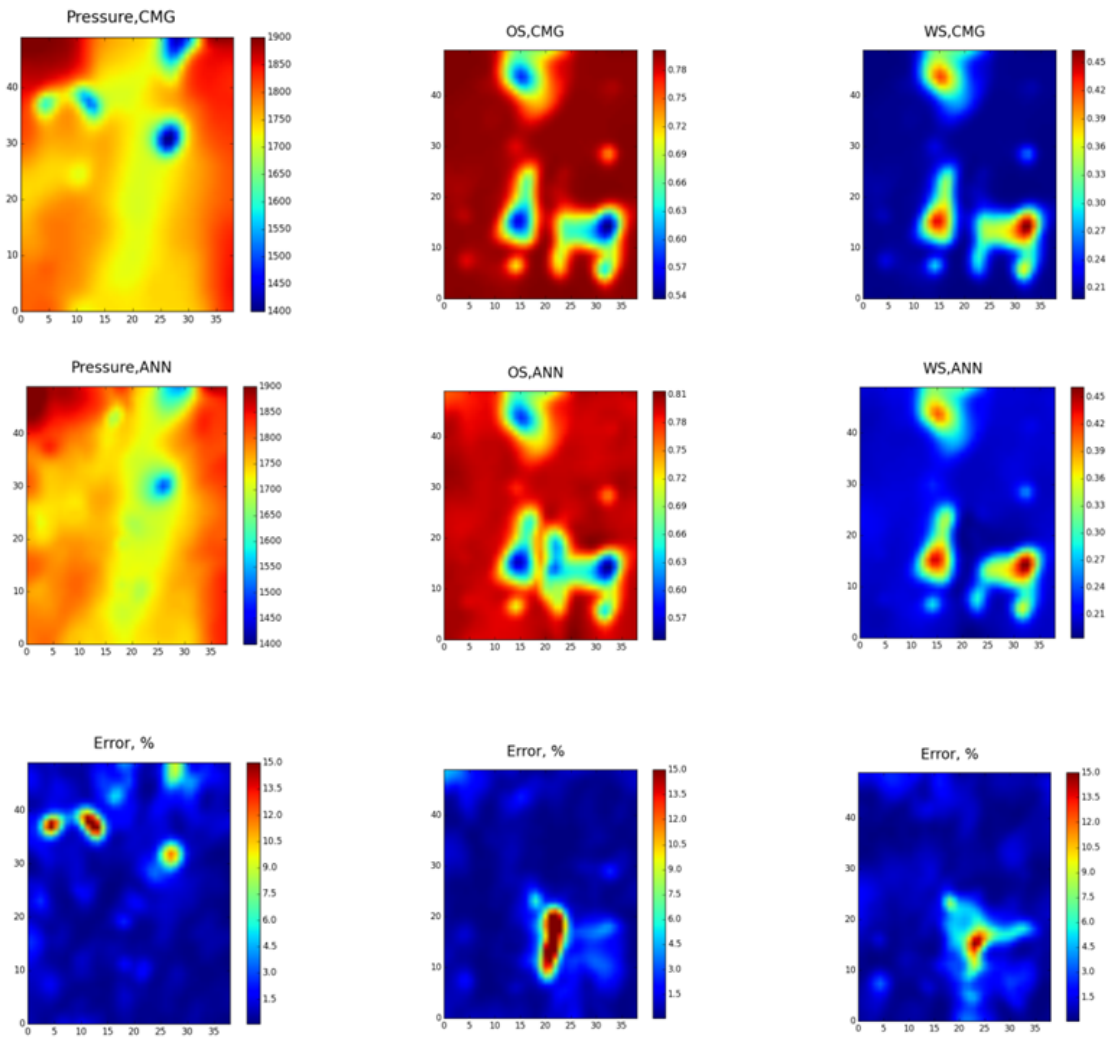
Figure 5.20: Year 2000 Layer 1 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
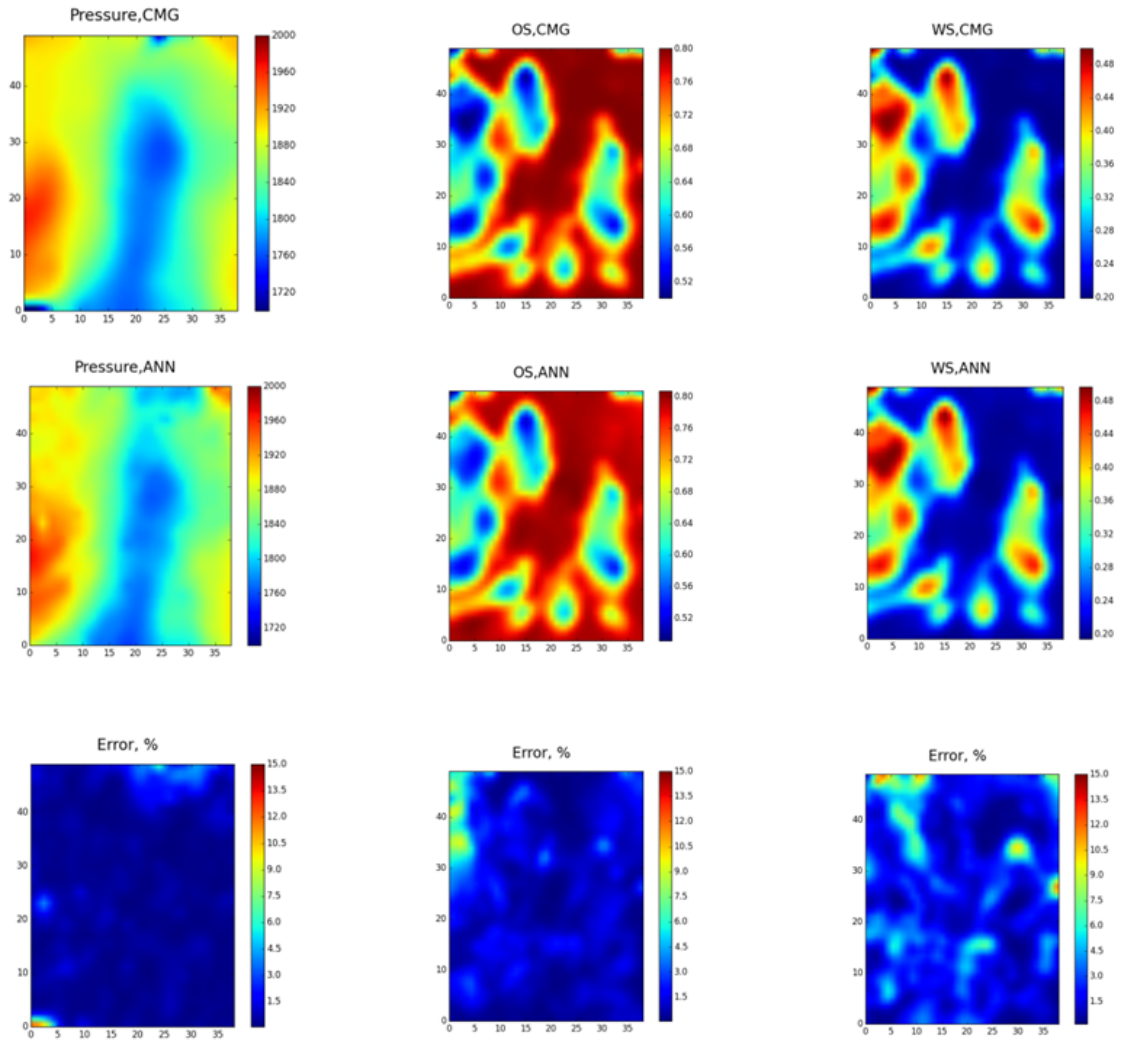
Figure 5.21: Year 2000 Layer 7 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).
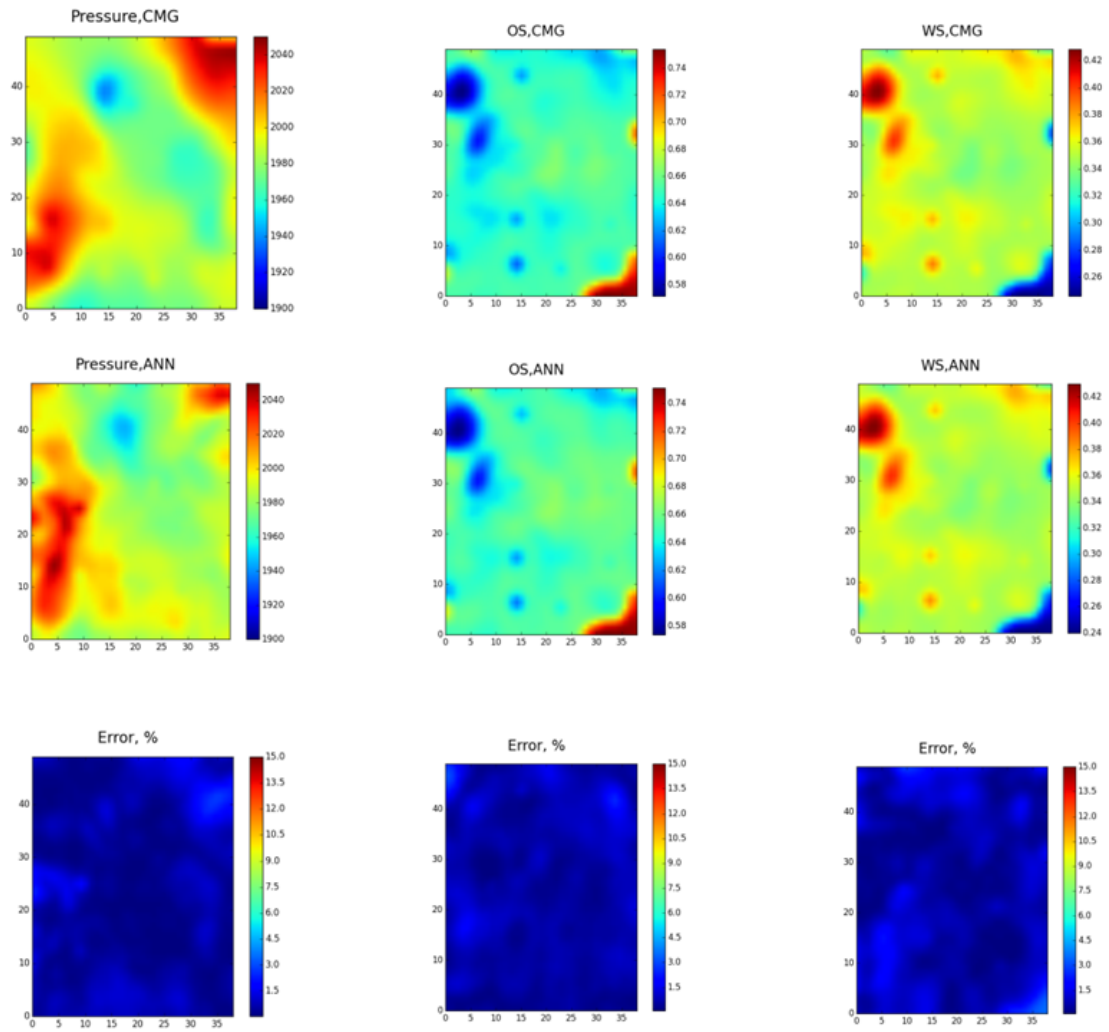
Figure 5.22: Year 2000 Layer 15 Pressure, Oil Saturation, and Water Saturation Distribution From Numerical Simulation and Smart Proxy Model (ANN).

Chapter 6

Smart Proxy for Geological and Operational Realization - Non Cascading Process

In Chapter 5, the developed smart proxy was able to replicate the numerical reservoir simulation results with acceptable range of error. The main objective of this Chapter is to prove that the smart proxy model is a reliable tool for reservoir simulation realization. The smart proxy model should be able to supply the reservoir properties (pressure and saturation) when changing the static (geological) and dynamic (operational constraints) data.

To develop the smart proxy for this objective, a non-cascading process was performed for the application of the smart proxy model. In non-cascading, the input data is entered to the smart proxy from the numerical simulation model rather than from the output of the smart proxy at the previous time step.

To design the numerical simulation run, the permeability distribution in the history-matched model was modified for each layer. It is important to keep the modification of the permeability values between the maximum and minimum values at each layer. This process meant a change in the locations of high and low permeability at each layer.

To develop the smart proxy model for this case, one of the static (geological) parameters was selected for modification. Due to the significant impact of permeability in the geological model and in the dynamic model, this was the parameter used for this case.

Similar to the procedure for changing the static data, the BHP for every production well was modified from the history-matched model. The wells BHP profiles were plotted and the high/low points were identified. The new simulation run was designed based on flipping the BHP profile for every well. The process can clearly be seen in plots for some wells.

Once the static model was built using the new data, the model was fed to a numerical simulator. Then, the inputs and outputs of the numerical simulator were used to generate the smart proxy database. The procedure used to develop the smart proxy model has already been discussed in Chapter 4.

Permeability (md)- Layer 10

(a)   (b)
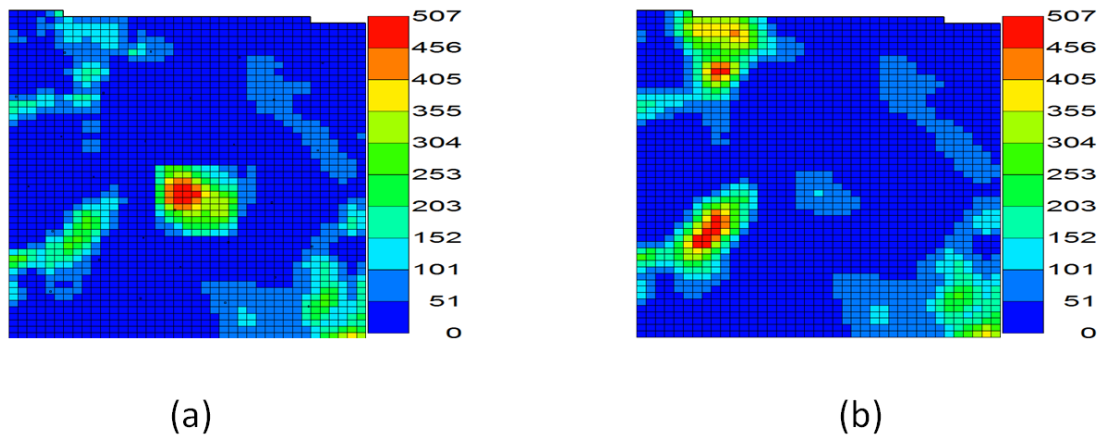
Figure 6.1: The permeability (md) data changing technique, left side of the figure shows the HM model where the right side shows the changed permeability distribution model (Layer 10).

6.1 Smart Proxy Deployment Results

The developed smart proxy was deployed for a blind run. The results showed a very robust model when changing static and dynamic data. Compared to the numerical simulator for the reservoir prop-
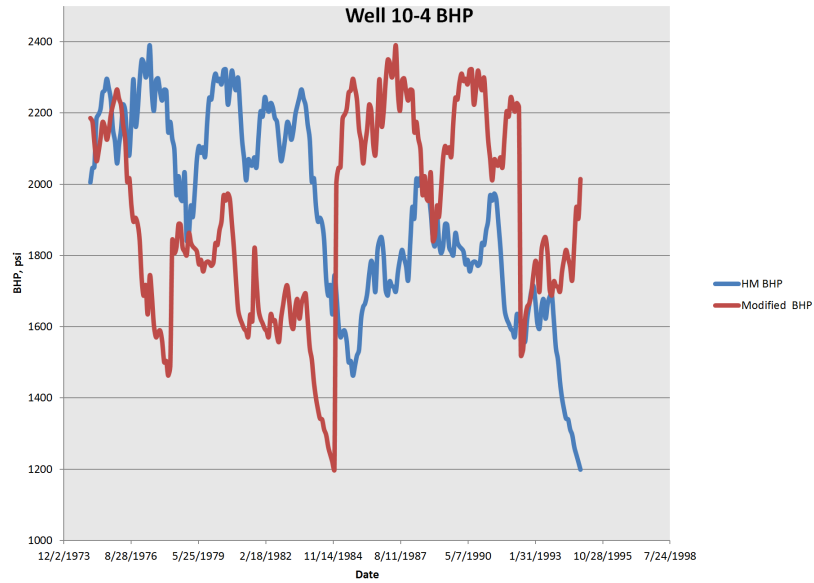
Figure 6.2: Example of the BHP modification technique for well no. 10-4. The blue curve is the HM BHP where the red one is the modified BHP used in designing the numerical simulation run to develop the smart proxy model.
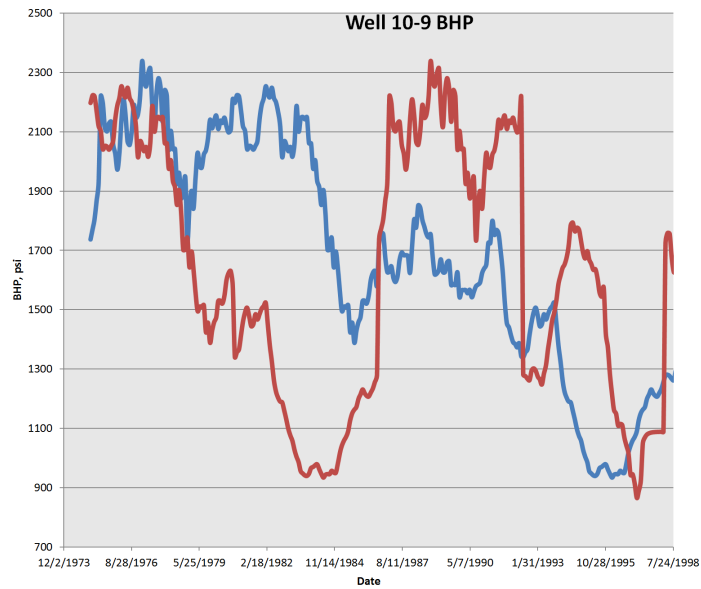
Figure 6.3: Another Example of the BHP modification technique for well no. 10-9.

erties (pressure and saturation), the smart proxy results were within an acceptable range of accuracy.

The absolute error percentage of the reservoir pressure over about 25 time steps was 5%. On the other hand, the absolute error percentage of the saturation for the same time period was about 1%.
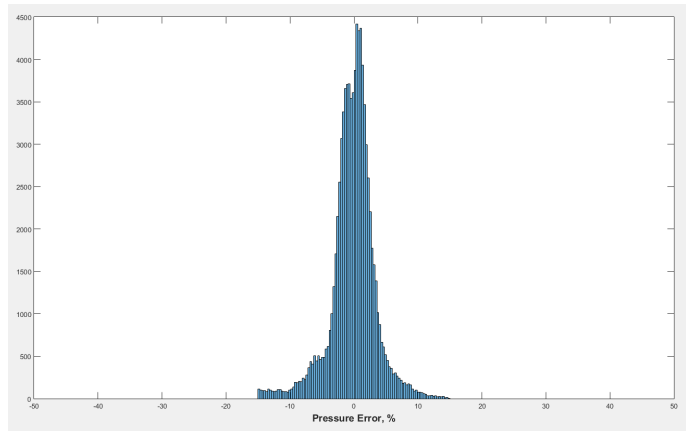
Figure 6.4: Pressure error histogram of randomly selected grids from 1975 to 2003 (smart proxy error compared to numerical simulator).
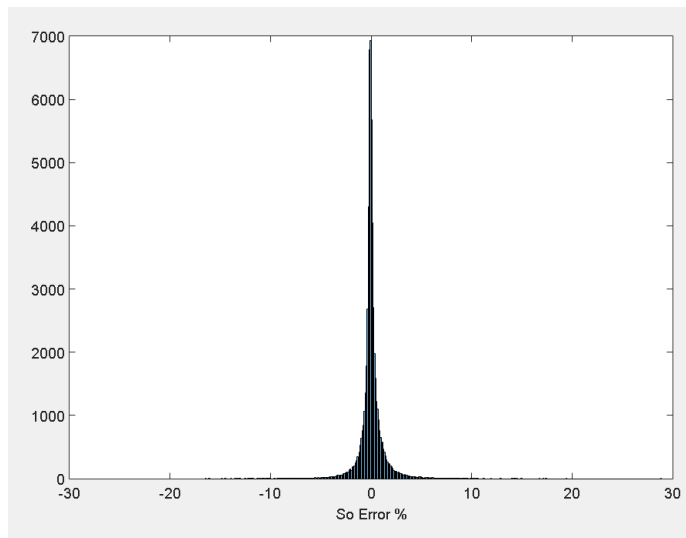
Figure 6.5: Oil saturation error histogram of randomly selected grids from 1975 to 2003 (smart proxy error compared to numerical simulator).
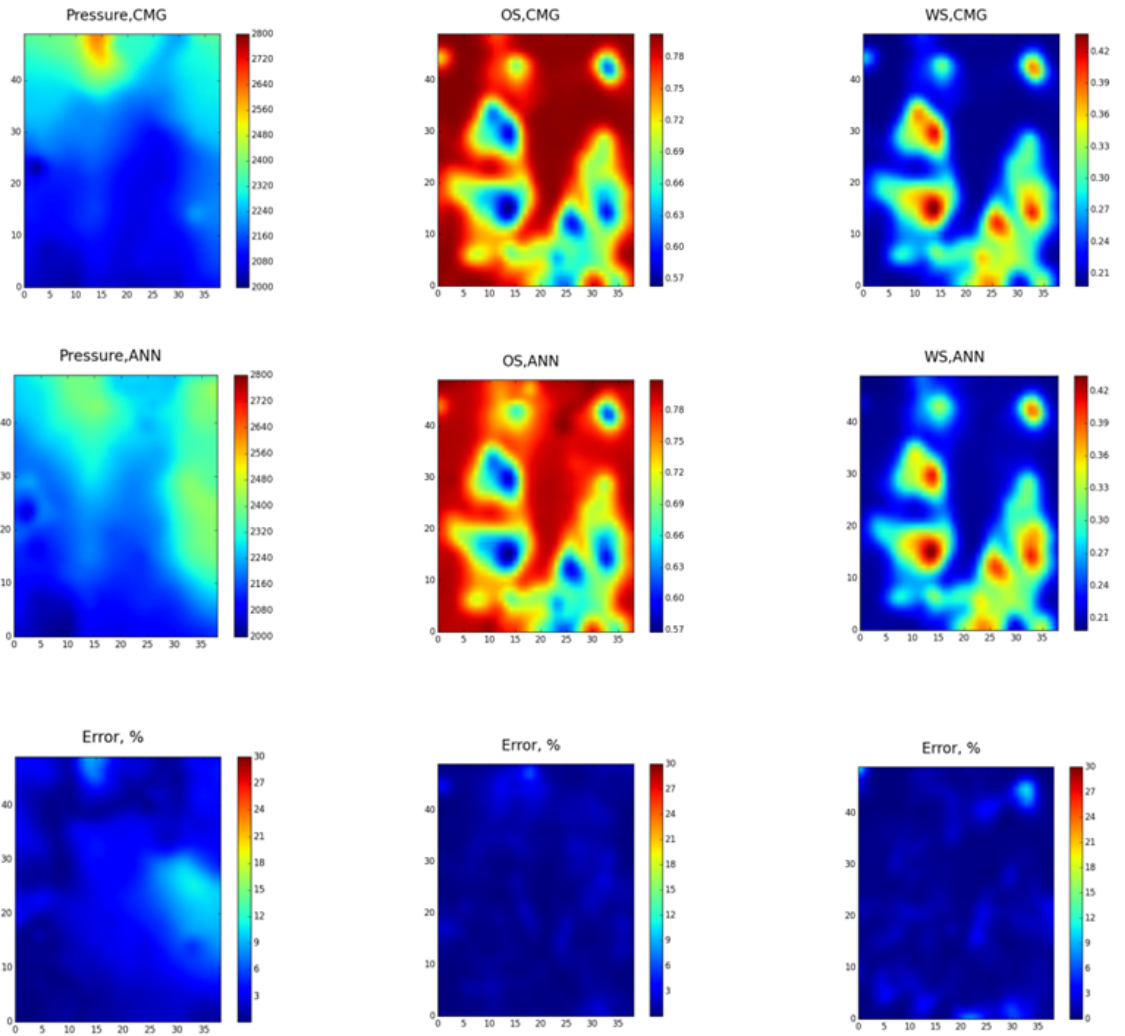
Figure 6.6: Pressure, Oil Saturation, and Water Saturation Distributions From Numerical Simulation and Smart Proxy Model (ANN) In year 1980 Layer 4.
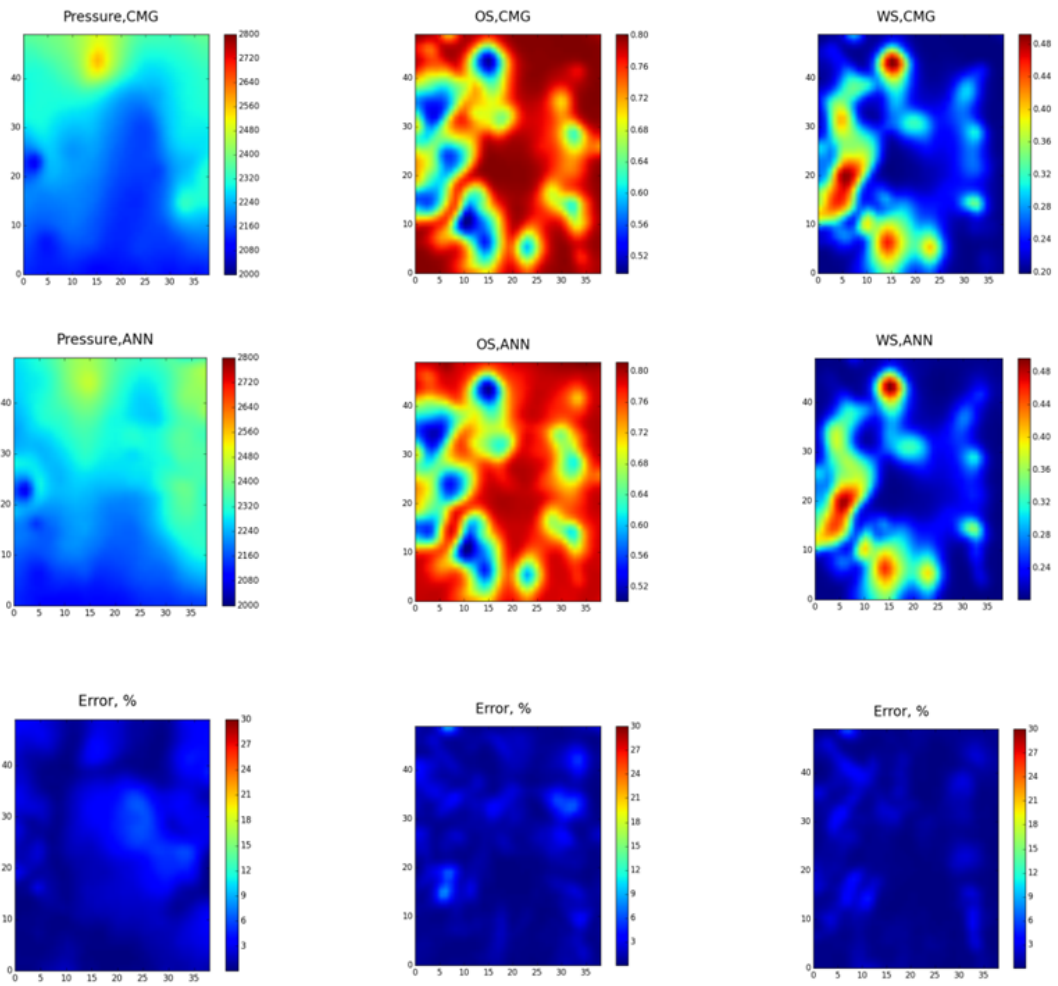
Figure 6.7: Pressure, Oil Saturation, and Water Saturation Distributions From Numerical Simulation and Smart Proxy Model (ANN) In year 1980 Layer 8.
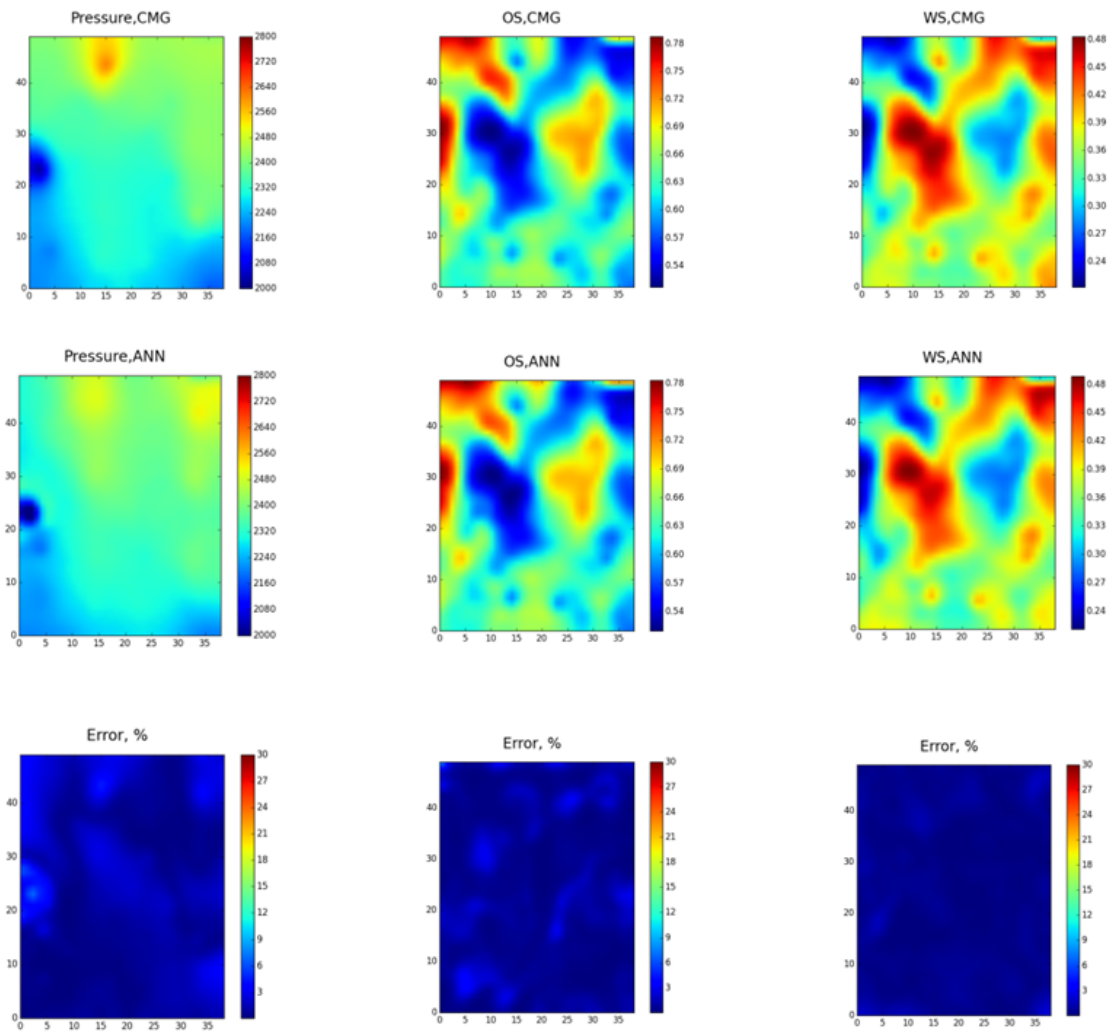
Figure 6.8: Pressure, Oil Saturation, and Water Saturation Distributions From Numerical Simulation and Smart Proxy Model (ANN) In year 1980 Layer 13.
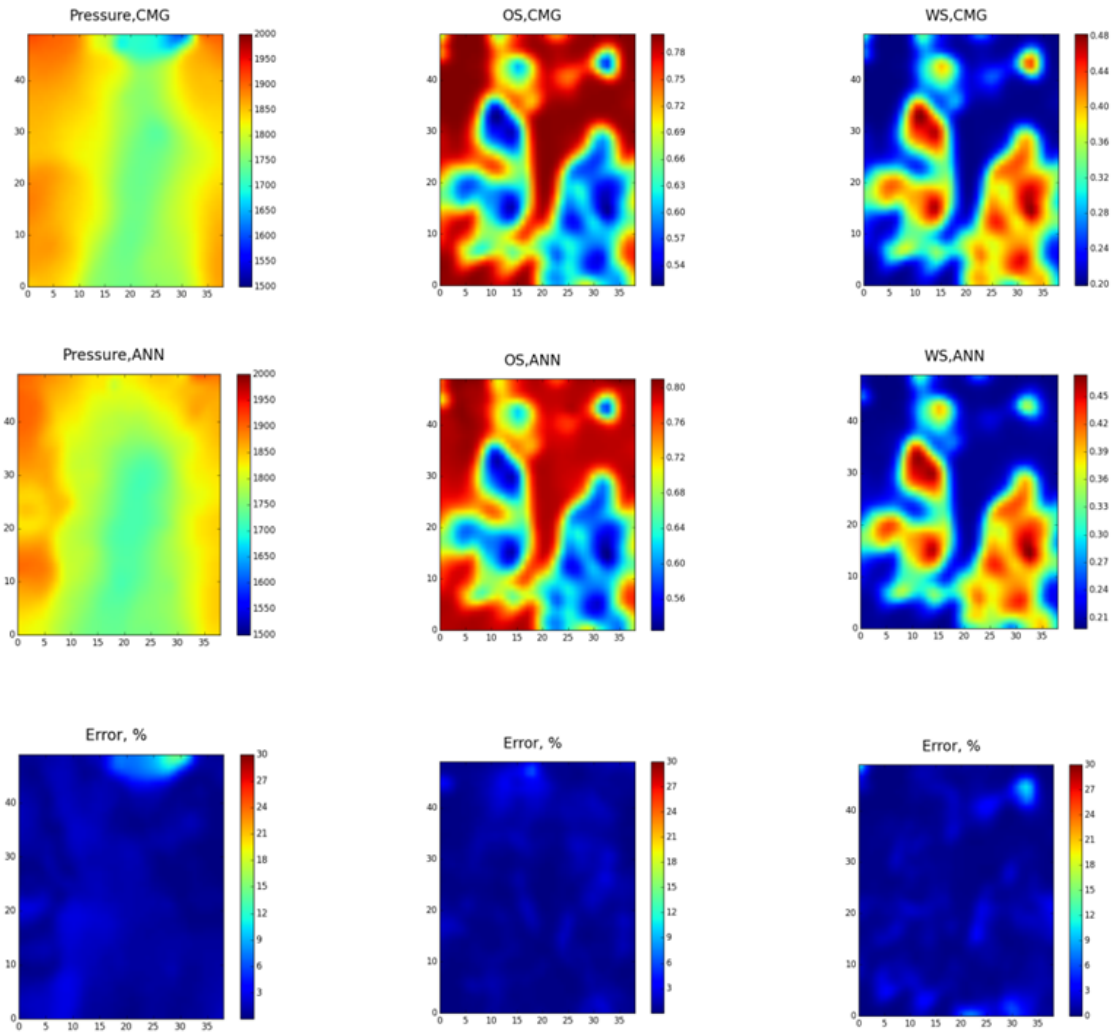
Figure 6.9: Pressure, Oil Saturation, and Water Saturation Distributions From Numerical Simulation and Smart Proxy Model (ANN) In year 2000 Layer 4.

Figure 6.10: Pressure, Oil Saturation, and Water Saturation Distributions From Numerical Simulation and Smart Proxy Model (ANN) In year 2000 Layer 8.

Figure 6.11: Pressure, Oil Saturation, and Water Saturation Distributions From Numerical Simulation and Smart Proxy Model (ANN) In year 2000 Layer 13.

6.2 Non-Cascading Discussion

The developed smart proxy model results show strong agreement with the results from the numerical simulator using a non-cascading process. The non-cascading was applied in the deployment of the smart proxy to test the concept of using data mining and artificial intelligence in a reservoir simulation application. In a non-cascading process, the inputs always come from the numerical simulator at each targeted time step. Therefore, according to the current author, non-cascading is impractical for predicting the reservoirs performance over many times steps (days, months, or years).

It is more practical to have a smart proxy model that is independent and can make predictions with a minimum of communication with the numerical simulators inputs/outputs. In order to achieve this objective, the cascading process was used when building the smart proxy model. In this process, the smart proxy model used the previous time step output as an input for the next time step until the last time step was reached. To put it another way, in this process, the smart proxy model was almost automated. The cascading process and smart proxy models are elaborated upon in the next sections of this dissertation.

Chapter 7

Smart Proxy Model for Geological and Operational Realization-Cascading Processes

As stated in Chapter 4, the most important step in developing the smart proxy is to define the objective. In this section, the objective is to develop a smart proxy model for geological and operational realization of the SACROC Unit using the cascading feature.

In cascading, the initial database is generated at the beginning. At the first time step, the ANNs models were trained and validated for the targeted reservoir properties (pressure and saturations). The models outputs with a calculated tier system were imported into the database of the next time step. This procedure was continued until the last time step was reached. The cascading feature was used in the training and deployment of the smart proxy development.

It is very important to design the best training technique in order to achieve the assigned objective. As mentioned in the previous sections, the majority of the time spent in developing the smart proxy model is taken up with studying the system and preparing the data. With cascading feature, the technique used to generate the smart proxy with a cascading feature was modified from the technique used in the non-cascading process.

The smart proxy models aim to supply the reservoir properties (pressure and saturations) for the SACROC Unit during the history of the field performance. The targeted period ranges from 1951 to 1995. For a better smart proxy performance, the field performance was divided into three phases for two reasons. Firstly, the training and validation of the proxy model is performed at each time step and since there are about 40 time steps, error accumulation would make this impractical to do in one phase. Second, there were different reservoir drive mechanisms during the development of the SACROC field (solution gas drive at the early production, then pressure maintenance using water injection).
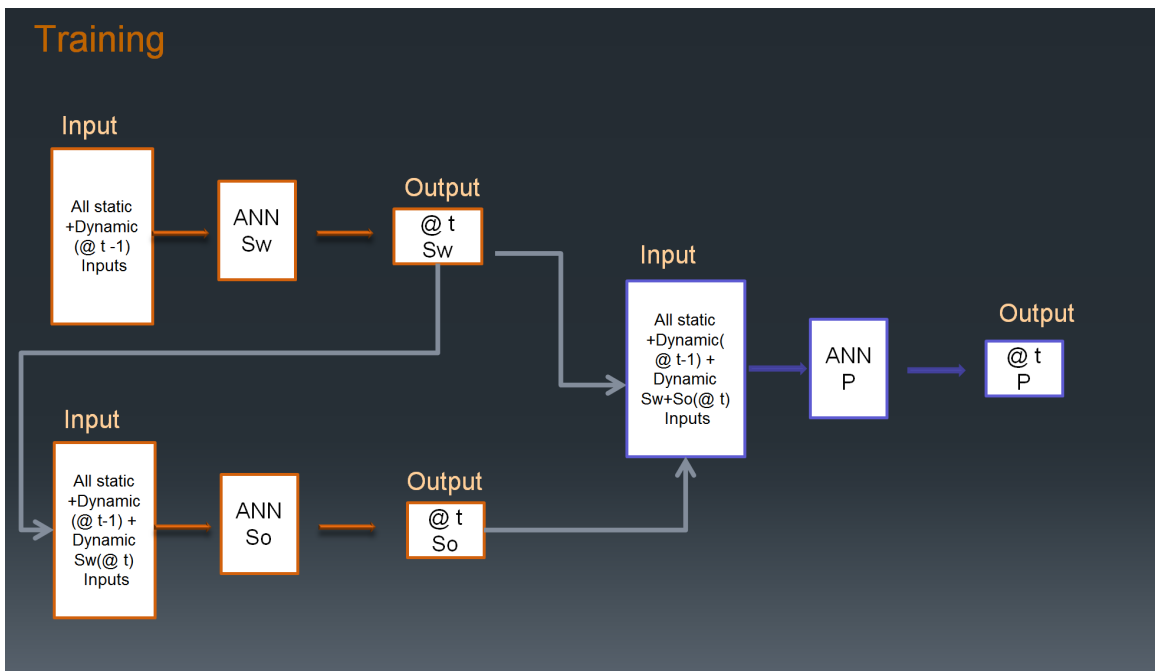
Figure 7.1: Cascading Training Flow-chart

7.1   Numerical Simulation Runs Design

The objective of the smart proxy is to examine the field performance under different geological and operational realizations. To do this, a handful of numerical simulation runs were designed to build the spatiotemporal database. The numerical simulation runs were built by changing the geological parameters (permeability and porosity) and the operational constraint (BHP).

For the geological parameters, permeability and porosity, two geological scenarios were designed.
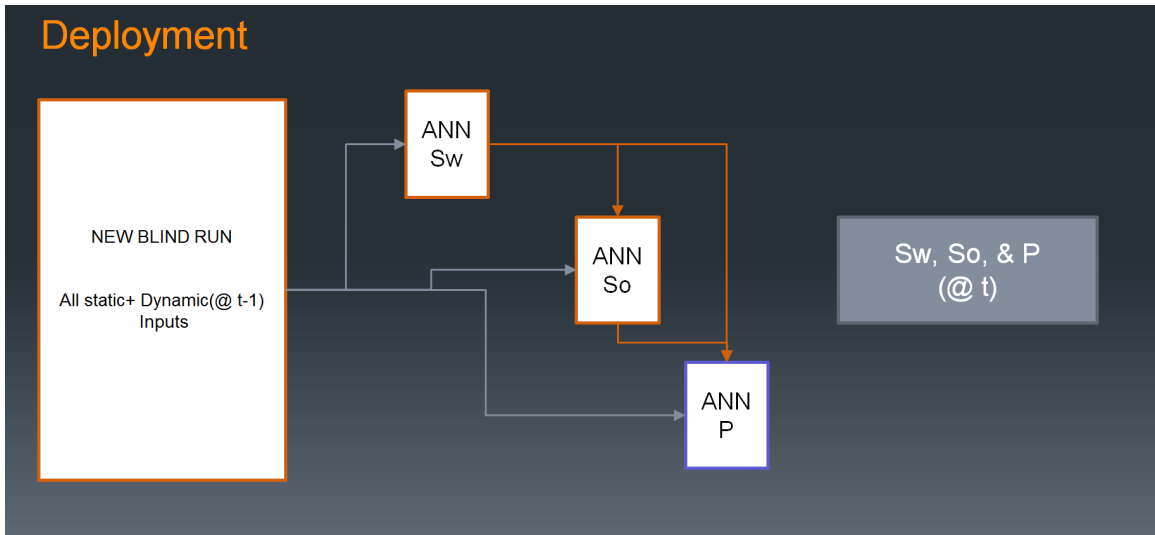
Figure 7.2: Cascading Deployment Flow-chart
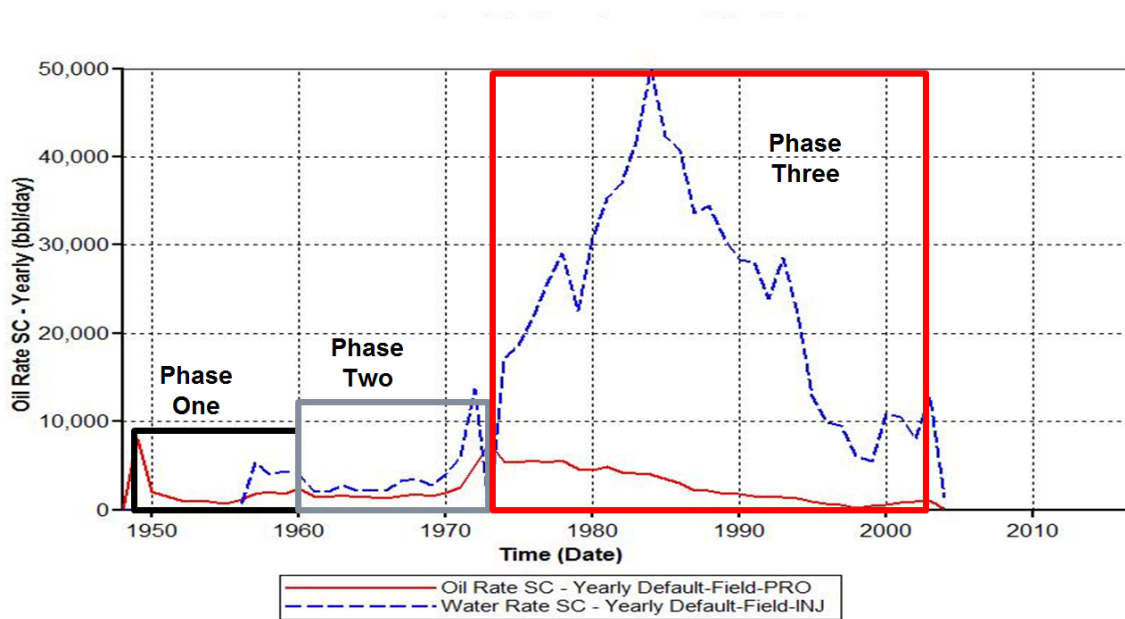
Figure 7.3: The divided three production phases of SACROC unit

The first scenario was generated by multiplying the history-matched model porosity and permeability by 1.5. The second geological scenario was performed by multiplying the permeability and porosity by 0.5. In other words, two general scenarios were designed, a high porosity-permeability scenario and a low porosity-permeability one.

The operational constraint examined in this study is BHP. The operational constraint numerical simulation run designs were based on the profile of the BHP for every production well. To have sufficient data for the model to be trained and generalized for different BHP profiles, 10 numerical simulation runs were generated by modifying each wells BHP profile. To elaborate, the BHP of the wells in the history-matched model was plotted. The maximum and minimum BHP of all wells was identified. After that, for each well, 10 BHP profiles were designed between the maximum and minimum values of BHP.

7.2   Database Generation, Input Selection, and Neural Network Training

The designed numerical simulations were executed to collect the static and dynamic data of these runs to generate a database that would represent the entire reservoir in terms of data. Once the database had been generated, the tier system was calculated (tier system details are provided in Chapter 4). Subsequently, inputs for neural network training were selected for the reservoir properties (pressure and saturation).

The neural network was constructed to develop a smart proxy model for each reservoir property. The training was performed at each time step for each property. This means that the number of smart proxy models developed was equal to the number of the time steps for each reservoir property.

7.3   Phase One Smart Proxy Model

Phase one represents the early time production of the field. The duration of this phase is 10 years (from 1950 to 1960). During this phase, the field started producing under the gas solution drive. The reservoir pressure decreased to the point where a pressure maintenance program was needed.
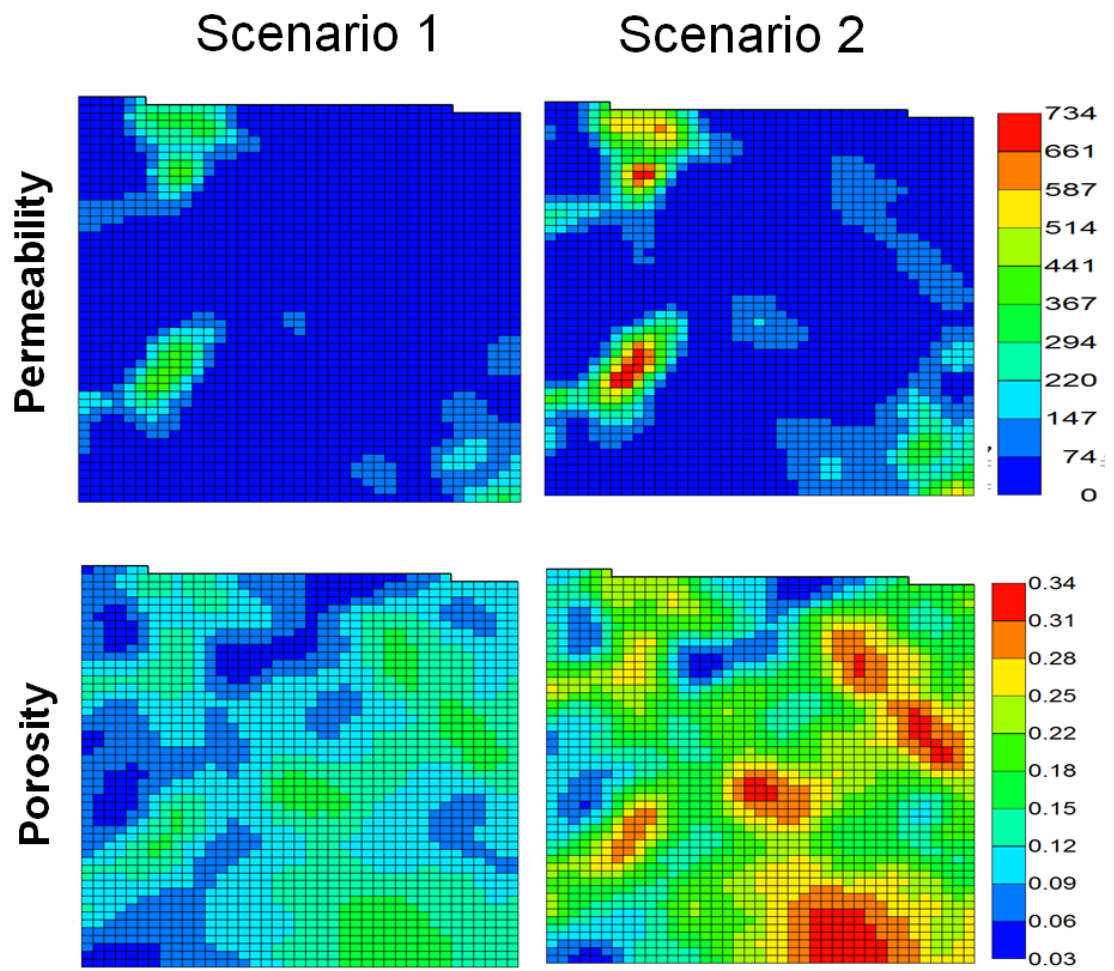
Figure 7.4: The high and low porosity-permeability scenarios at layer 10 of the geological model
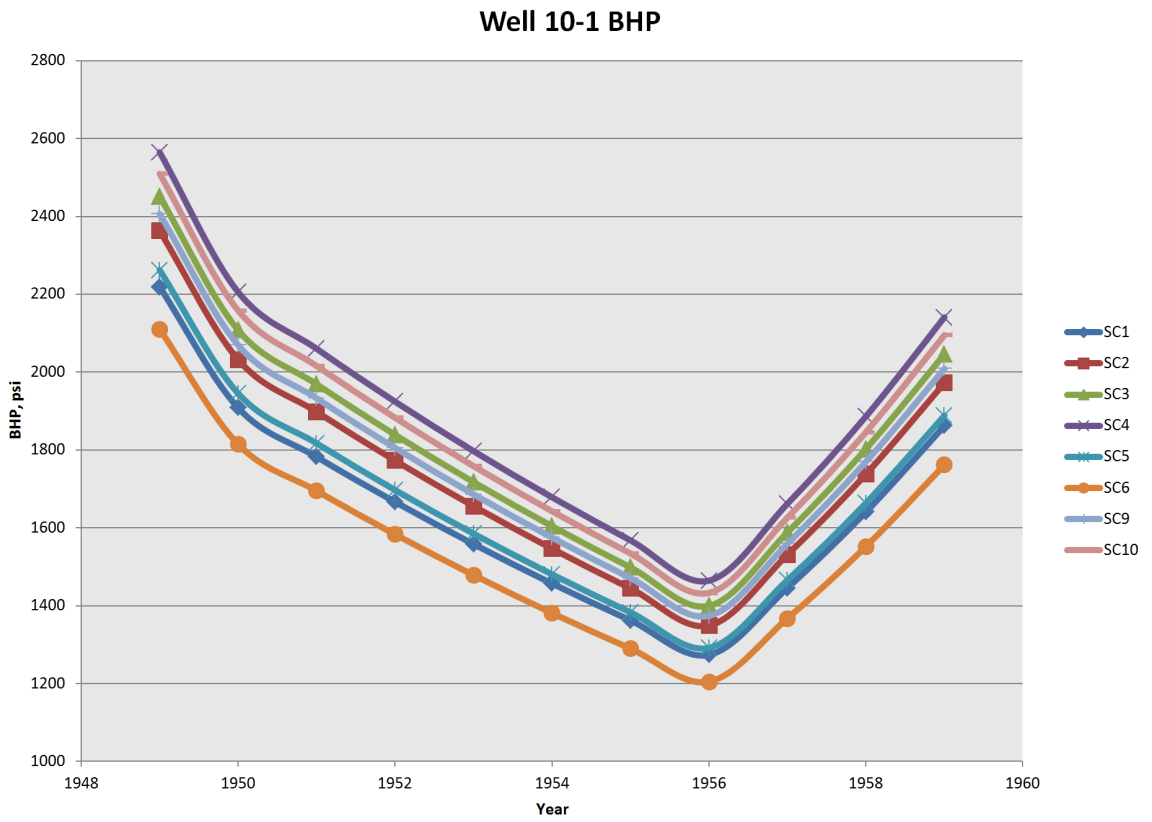
Figure 7.5: BHP scenarios for well number 10-1. Same technique was applied for all other production wells.

Parameters			
Grid Based		**Well Based**	
Static	**Dynamic**	**Static**	**Dynamic**
Porosity	Pressure	Distances	Injection
Permeability (x,y,z)	Saturation		Production
Thickness			Operational Constraints
Grid Top			
Tier System	**Distances**	**Operation**	
Above	Producers	FBHP	
Below	Injectors	Injection rates	
East	Boundaries	Oil production	
West		Water Production	
South		Gas Production	
North			

Figure 7.6: Cascading Process Input Training Parameters.

In 1956, the first water injection well was introduced to the field.

The number of time steps in phase one is 10 (10 years of production). This means 10 smart proxy models were developed for every reservoir property. Every model was trained and validated and the output of each model at each time step was imported to the next time step database for training until the last time step was reached.

7.3.1 Phase One Training Results

The construction of the neural networks as well as the data input selection was based on the smart proxy development explained earlier in this study. The neural networks hidden layer and the nodes in the hidden layers could be assigned based on the number of inputs and the data records size. The data was divided for the training, with 80% in the training set and 20% for the calibration and validation sets. Ten neural networks were trained for every reservoir property. The calibration data sets performance was monitored for network training validation.

The networks training showed an excellent performance for all targeted reservoir properties. Table 7.1 shows the average R-squared coefficient for each property (calculated average based on 10 trained

neural networks).

Reservoir Property	Average R-squared coefficient
Grid Pressure	99%
Oil Saturation	99%
Water Saturation	99%

Table 7.1: Phase one Training Results

7.3.2 Phase One Blind Run Results

In order to validate the developed smart proxy model and to test that it was able to provide the reservoir properties at each grid block, it was applied to a blind simulation run scenario that was designed with different geological and operational realizations.

The deployment results showed that the smart proxy model was capable to produce the reservoir properties (pressure and saturation) with an acceptable range of error compared to the numerical simulator. The absolute average error for the pressure was 1.5% and the absolute average error of the saturation was 1%.

The results of the deployment were shown using data distribution maps. Since the reservoir has 16 layers and since there were 10 smart proxy models for each reservoir property, 480 images were generated from these results. Therefore, it is difficult to show all results in this section. Selected layers of the reservoir are shown in the following images, based on the productivity of each layer. For highly productive layers, layers 3, 7, and 10 were selected. On the other hand, layers 1, 4, and 16 were selected as less productive ones.
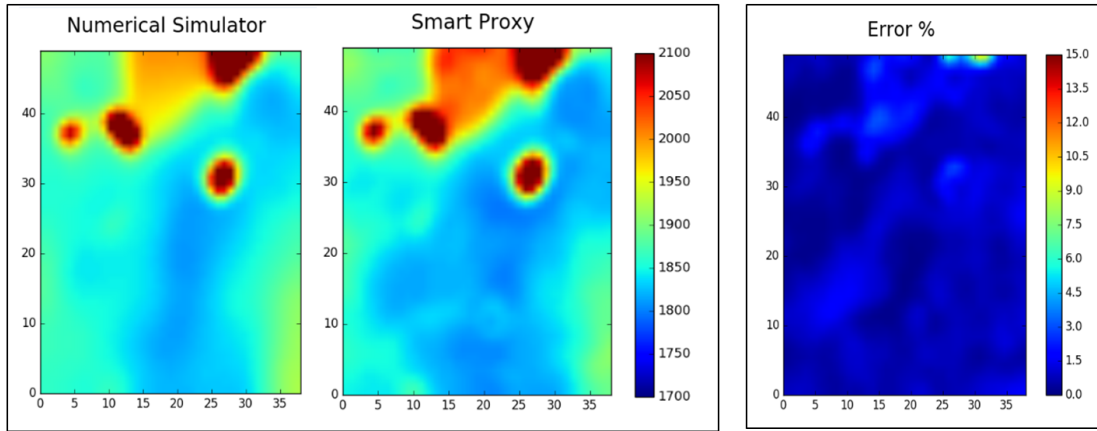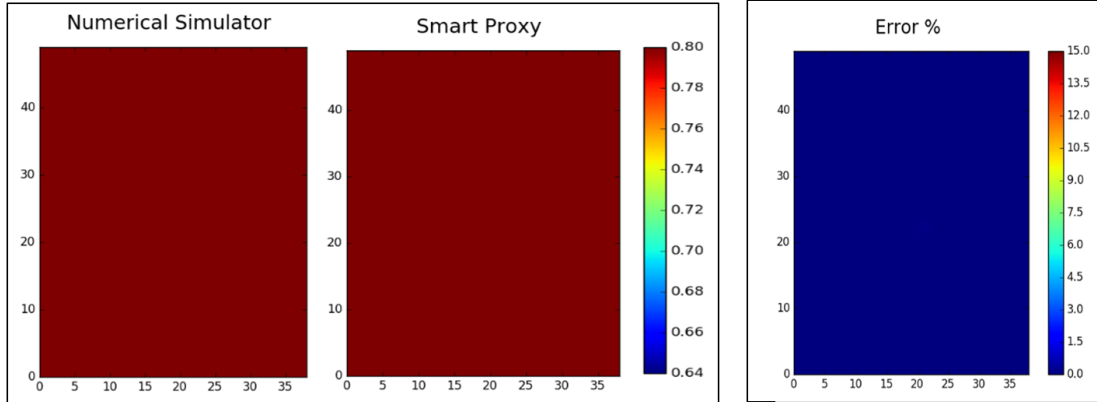
Figure 7.7: Phase one Layer-1 in 1951

Reservoir Pressure – Layer 3 on 01/01/1951
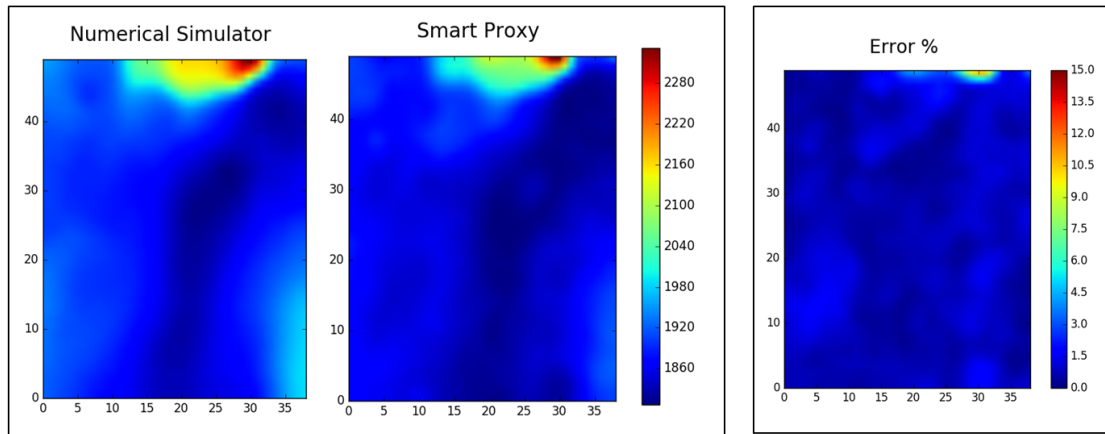
Oil Saturation– Layer 3 on 01/01/1951
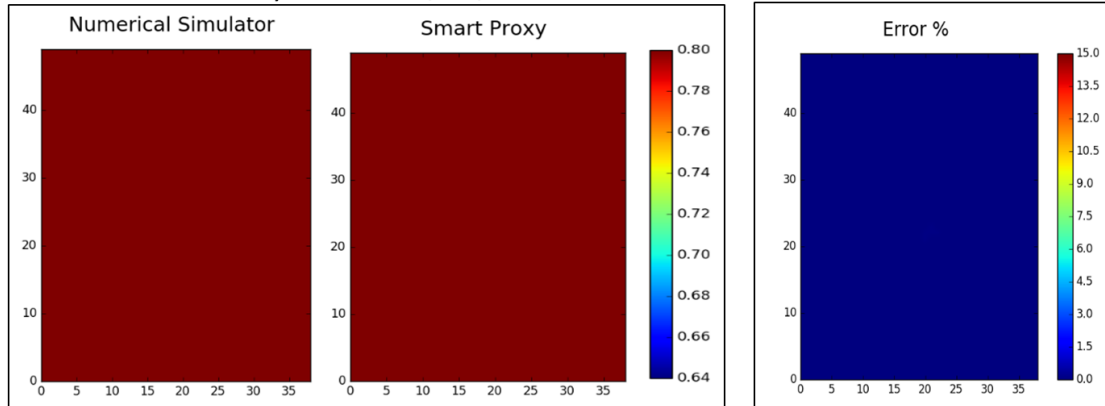
Figure 7.8: Phase one Layer-3 in 1951
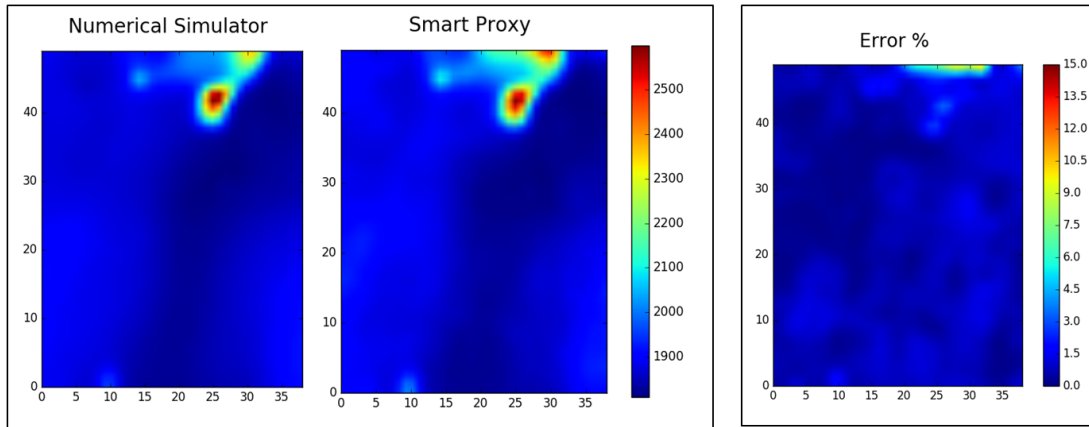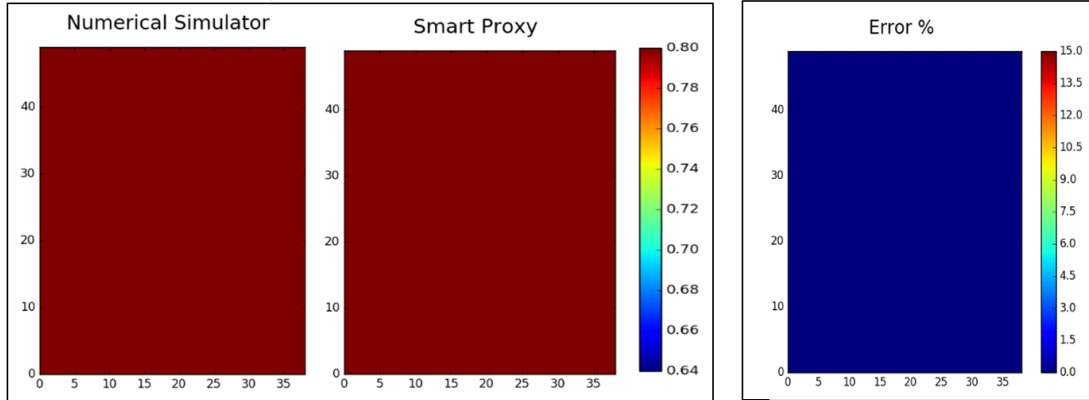
Figure 7.9: Phase one Layer-5 in 1951

Reservoir Pressure – Layer 7 on 01/01/1951
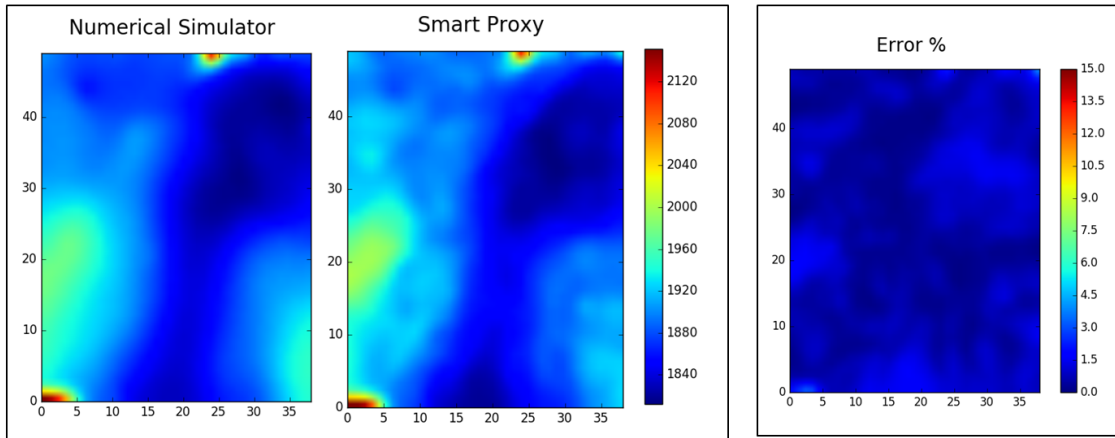
Oil Saturation– Layer 7 on 01/01/1951
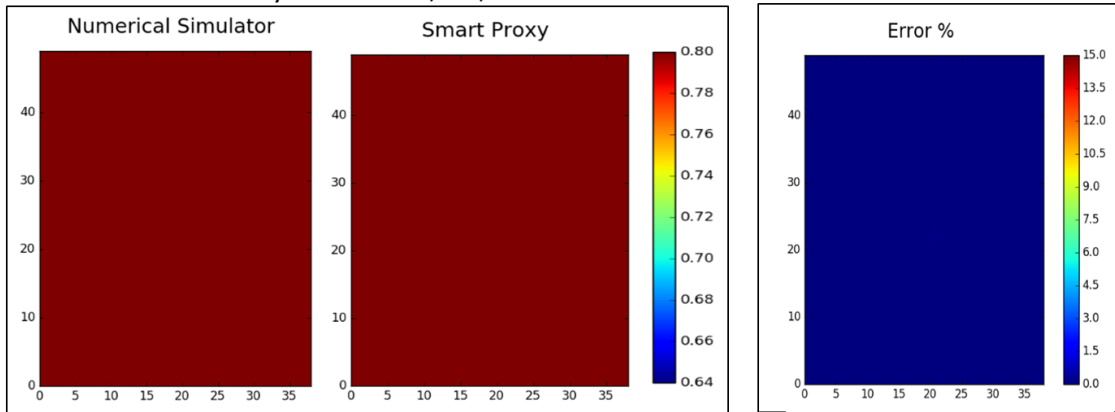

Figure 7.10: Phase one Layer-7 in 1951
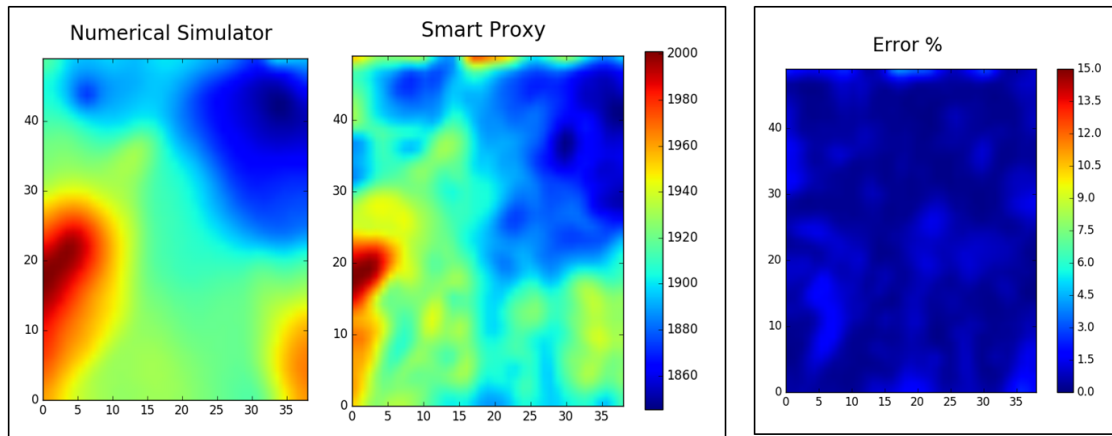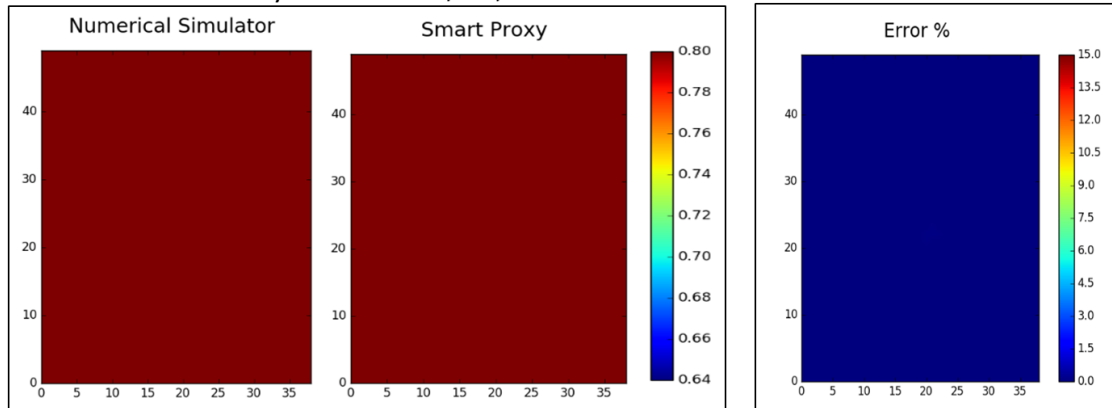
Figure 7.11: Phase one Layer-10 in 1951

Reservoir Pressure – Layer 16 on 01/01/1951
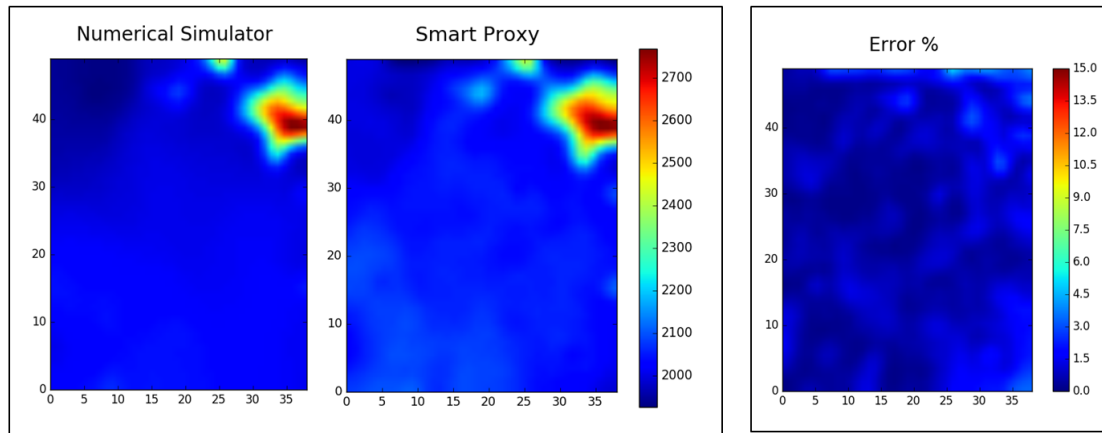
Oil Saturation– Layer 16 on 01/01/1951
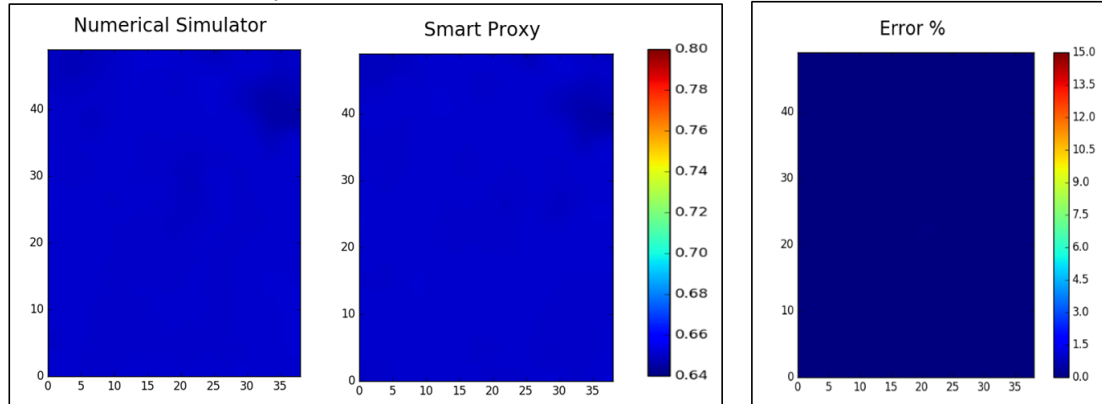

Figure 7.12: Phase one Layer-16 in 1951
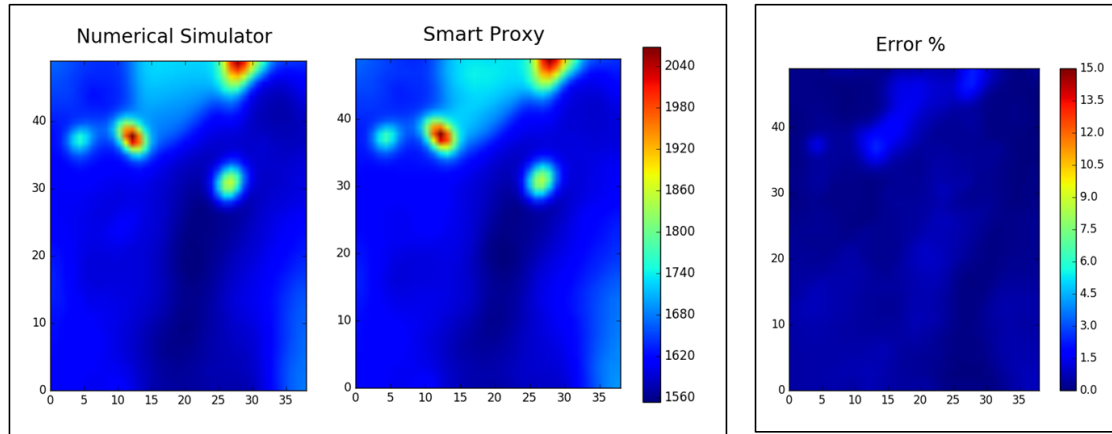
Figure 7.13: Phase one Layer-1 in 1953

Reservoir Pressure – Layer 3 on 01/01/1953
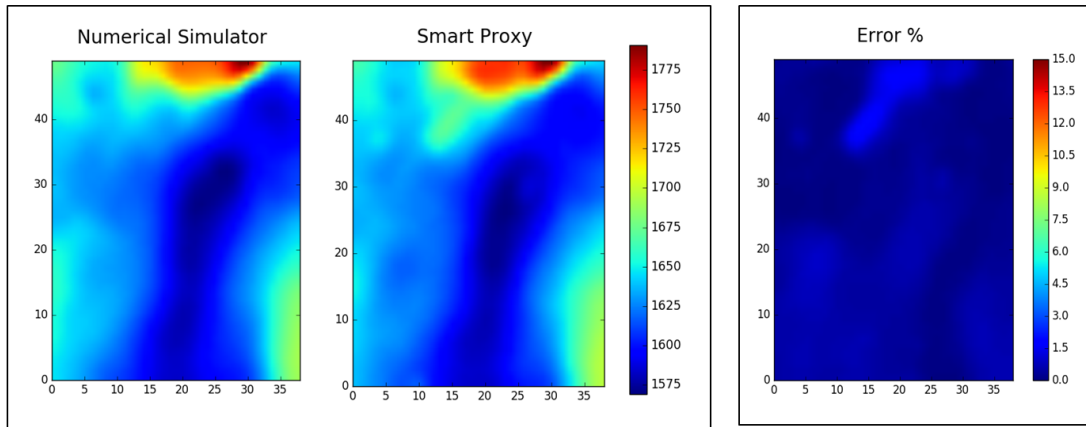
Oil Saturation– Layer 3 on 01/01/1953
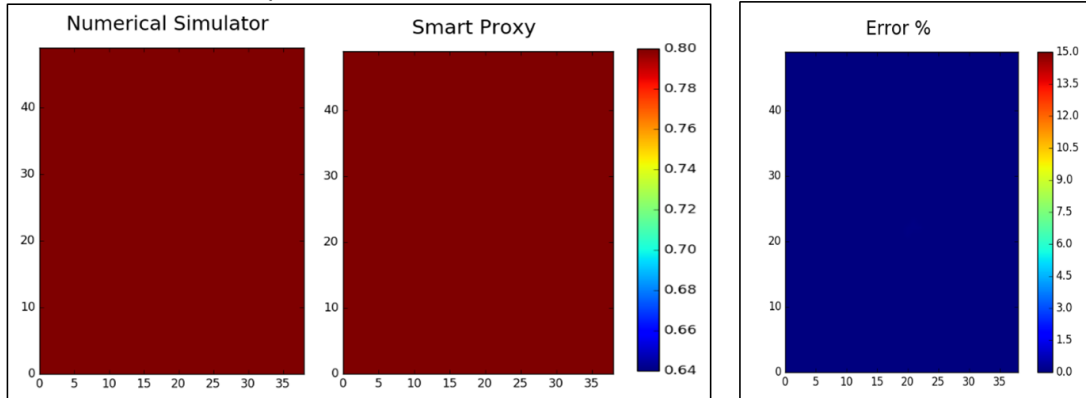
Figure 7.14: Phase one Layer-3 in 1953
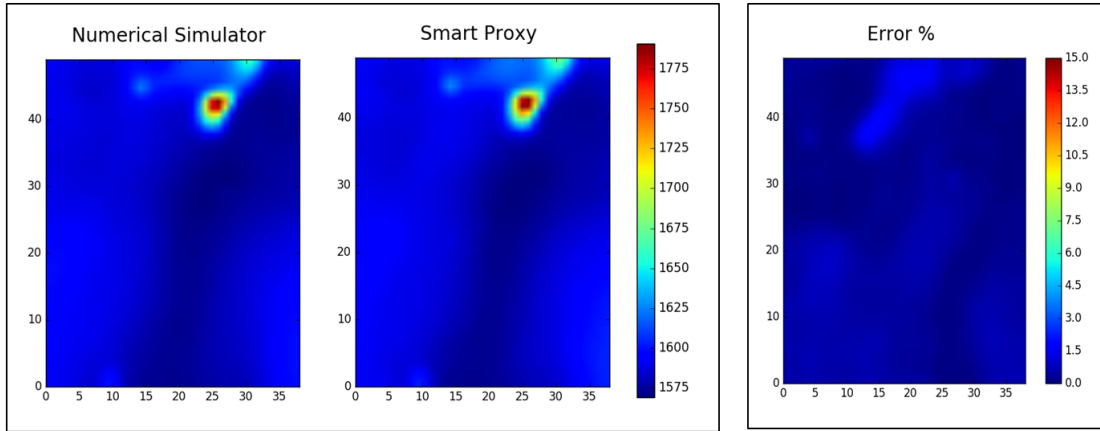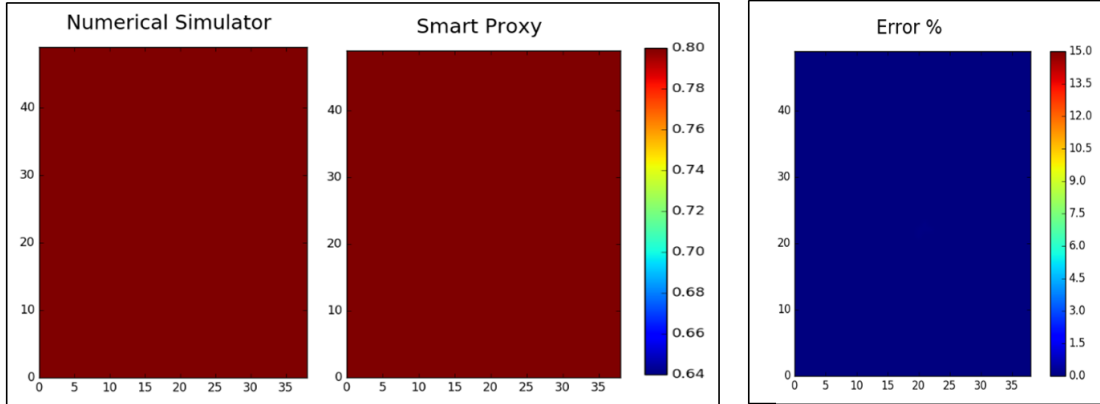
Figure 7.15: Phase one Layer-5 in 1953
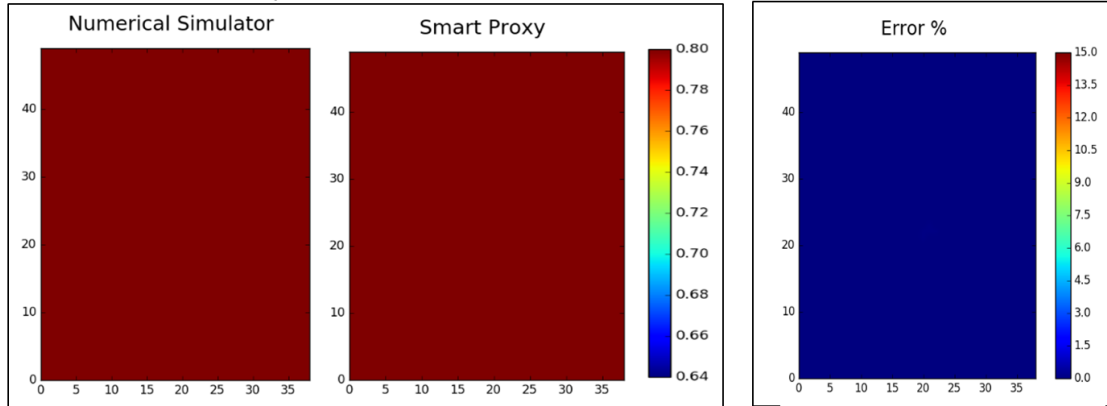
Figure 7.16: Phase one Layer-7 in 1953
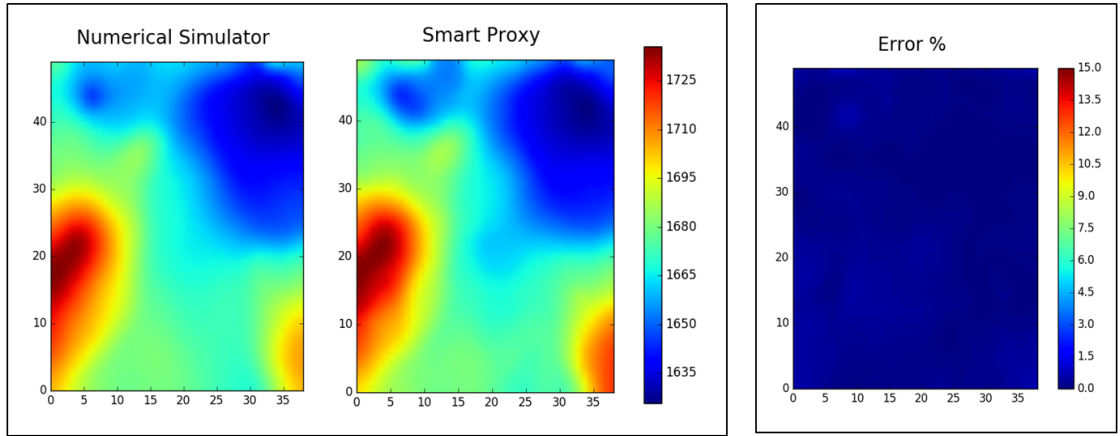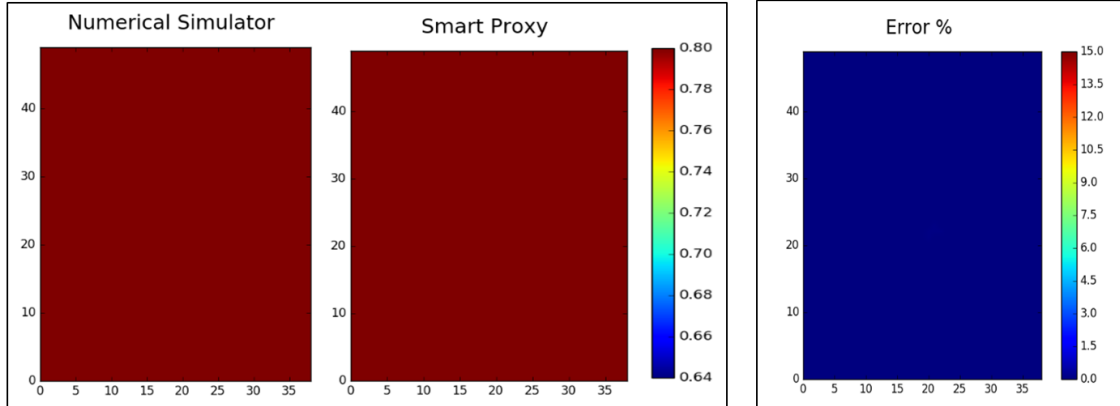
Figure 7.17: Phase one Layer-10 in 1953
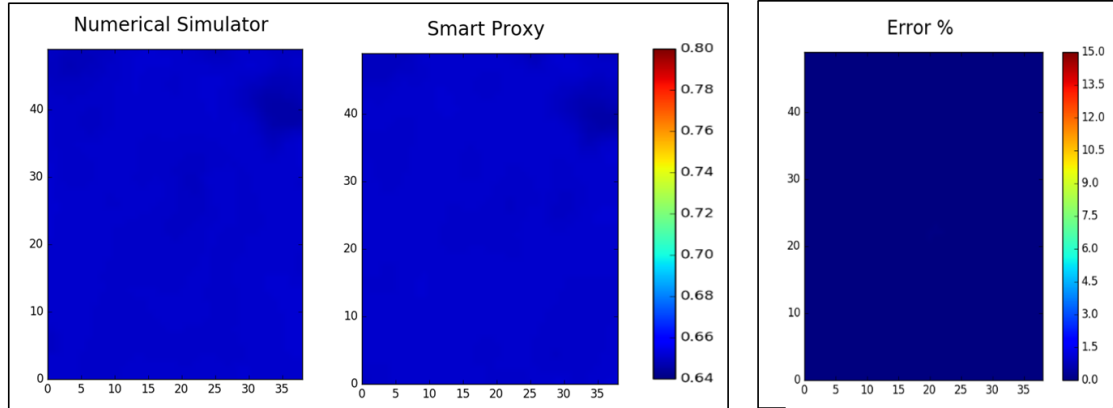
Figure 7.18: Phase one Layer-16 in 1953

Reservoir Pressure – Layer 1 on 01/01/1955
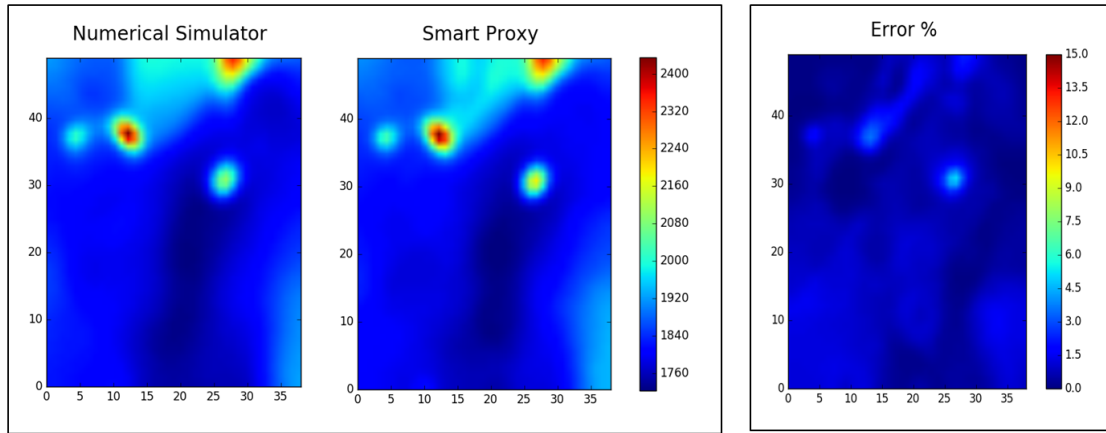
Oil Saturation– Layer 1 on 01/01/1955
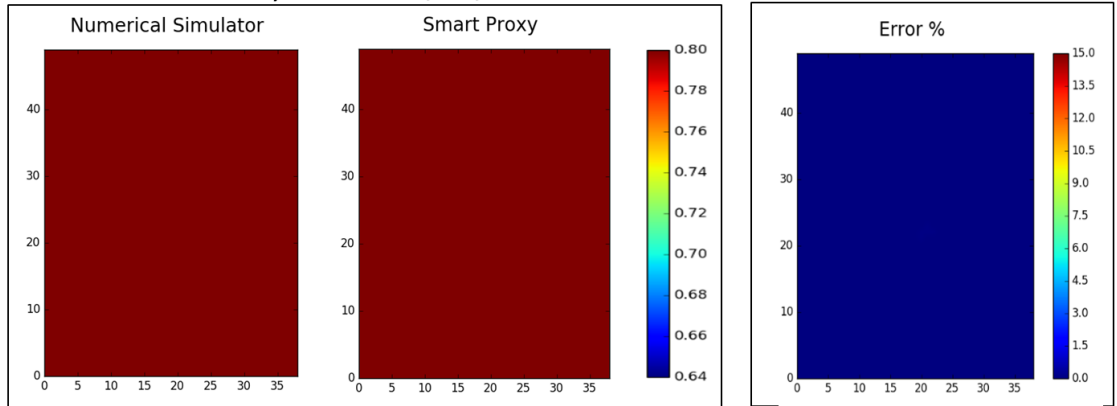
Figure 7.19: Phase one Layer-1 in 1955
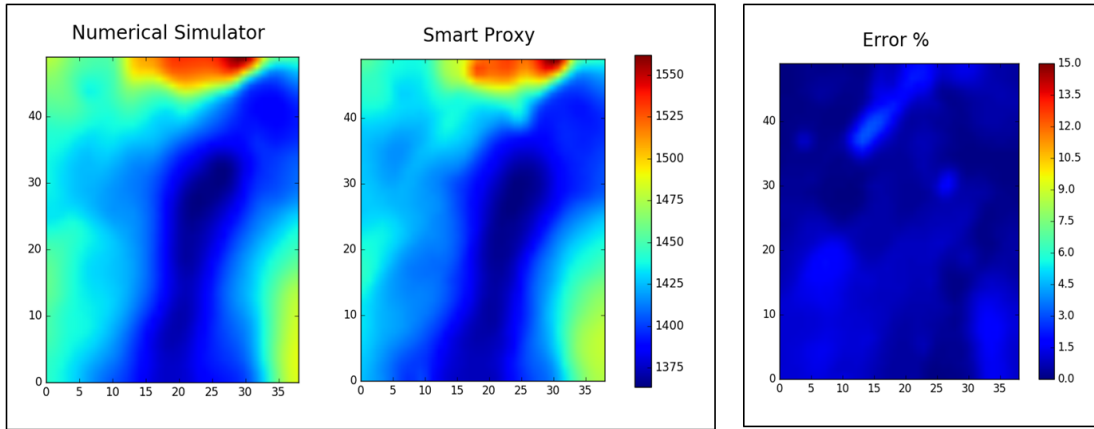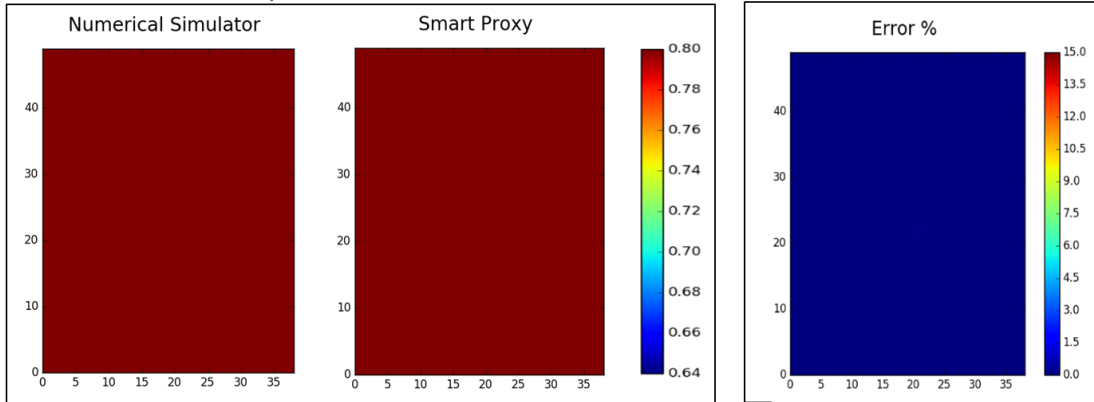
Figure 7.20: Phase one Layer-3 in 1955
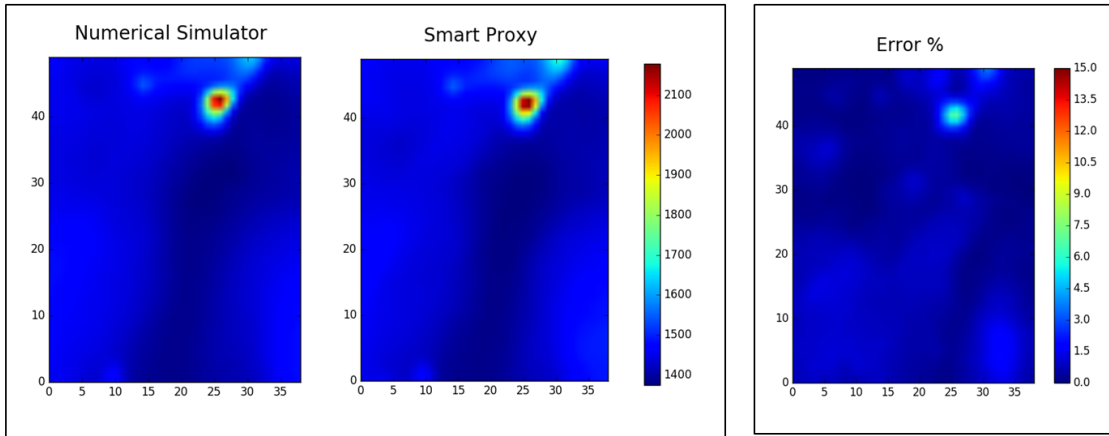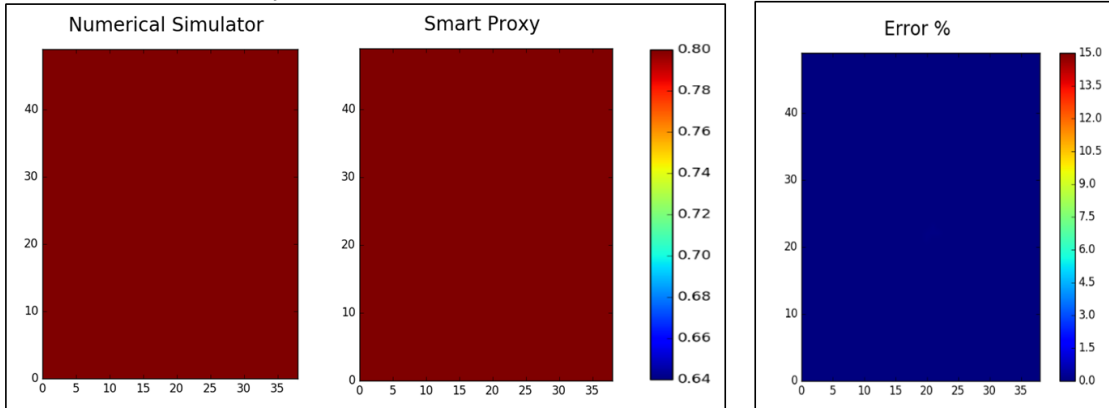
Figure 7.21: Phase one Layer-5 in 1955

Reservoir Pressure – Layer 7 on 01/01/1955
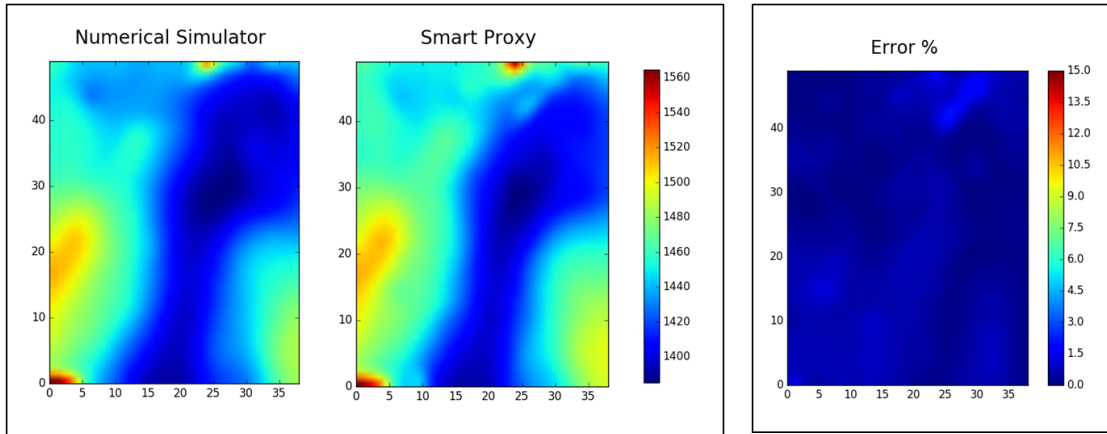
Oil Saturation– Layer 7 on 01/01/1955
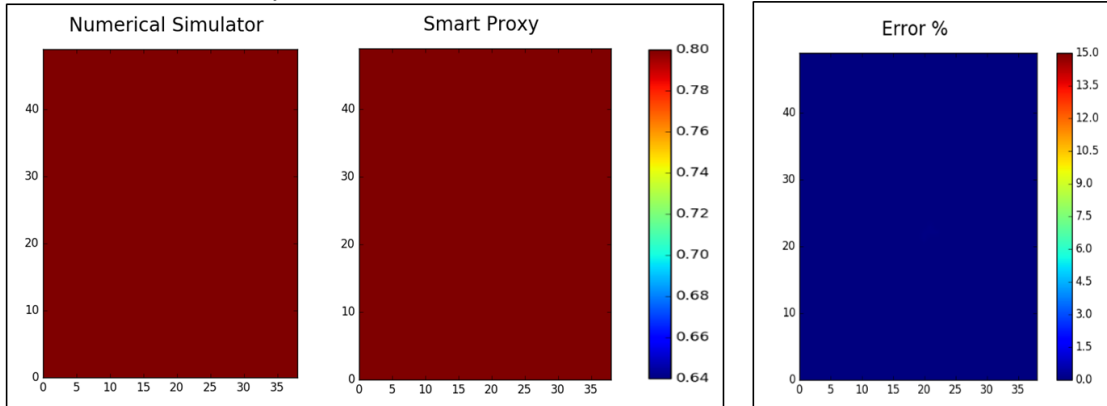
Figure 7.22: Phase one Layer-7 in 1955
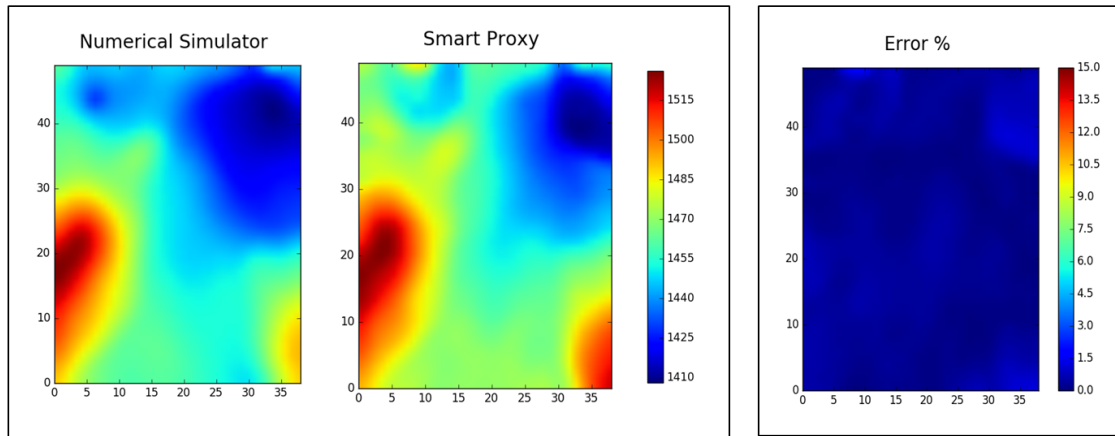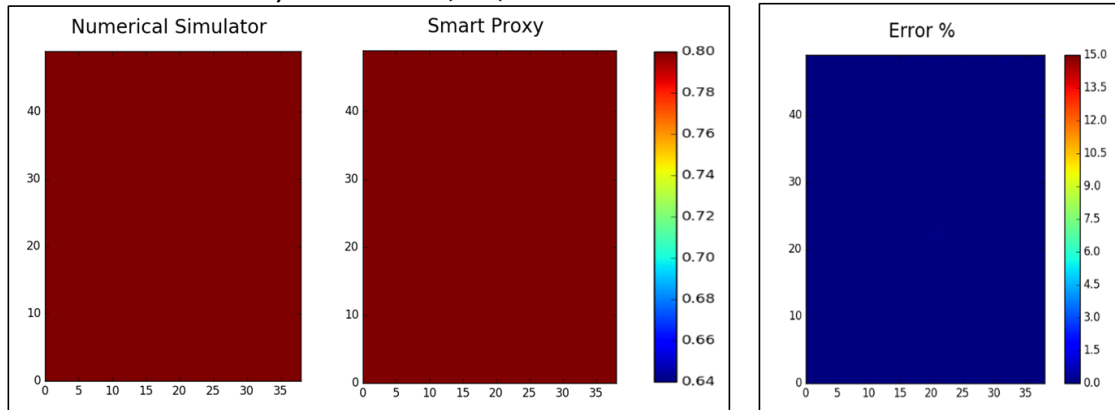
Figure 7.23: Phase one Layer-10 in 1955

Reservoir Pressure – Layer 16 on 01/01/1955
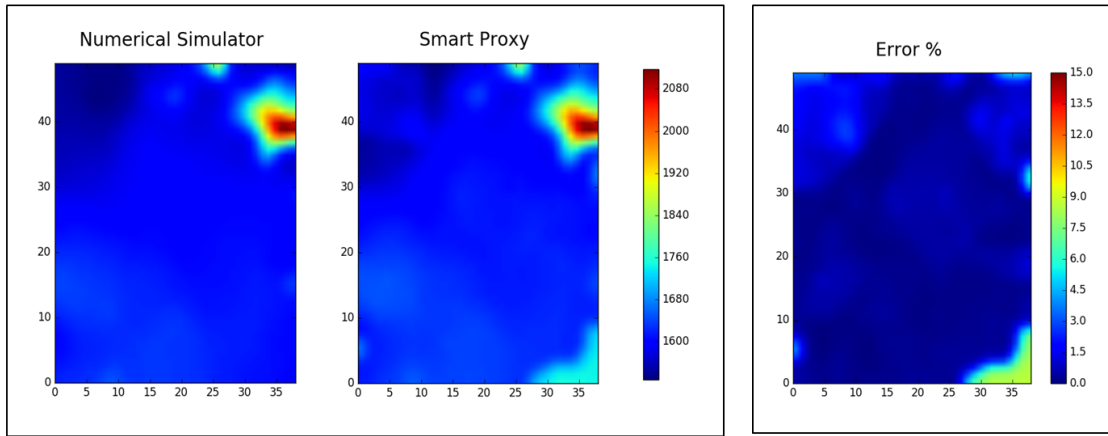
Oil Saturation– Layer 16 on 01/01/1955
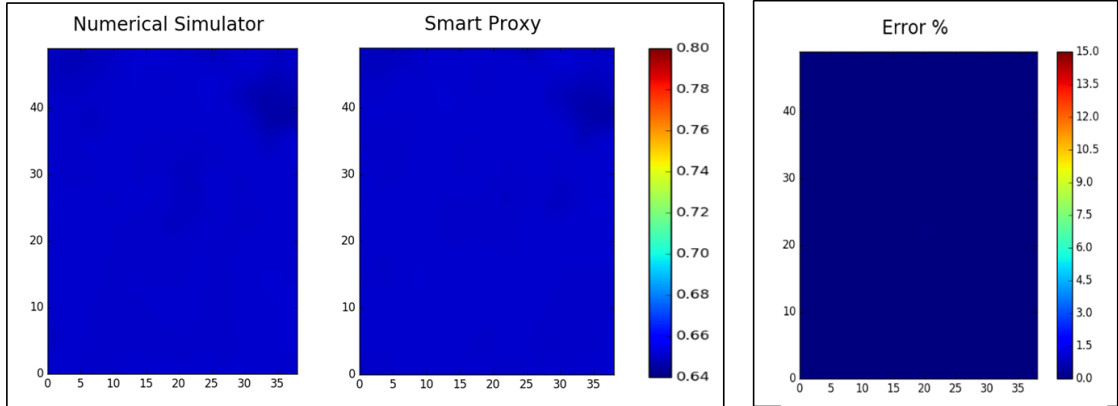
Figure 7.24: Phase one Layer-16 in 1955

Reservoir Pressure – Layer 1 on 01/01/1957
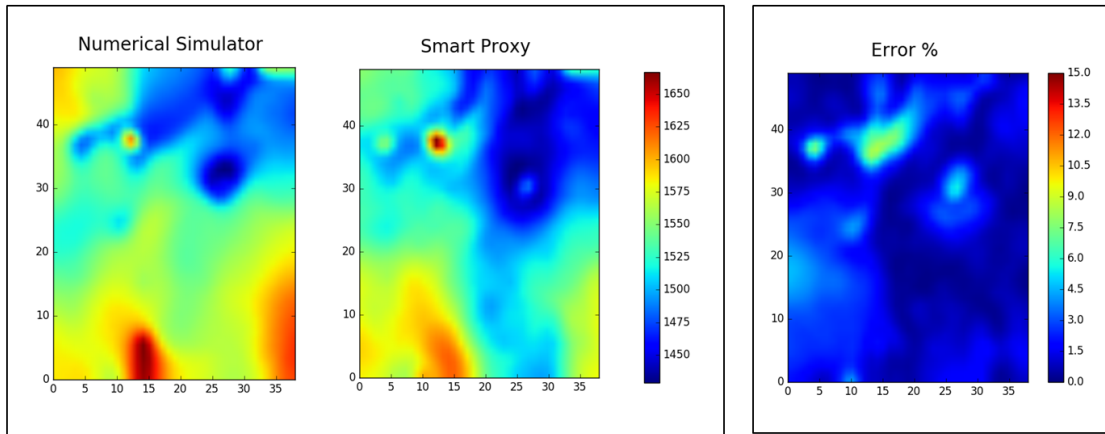
Oil Saturation– Layer 1 on 01/01/1957
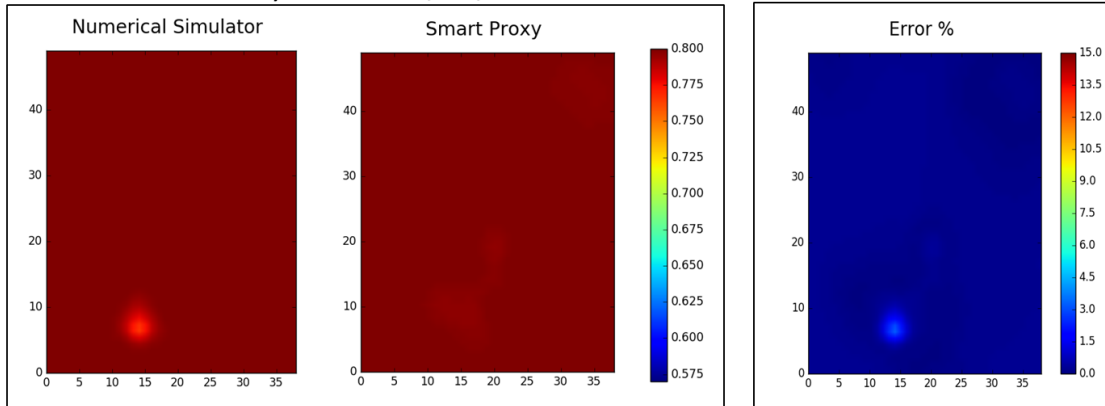
Figure 7.25: Phase one Layer-1 in 1957

Reservoir Pressure – Layer 3 on 01/01/1957
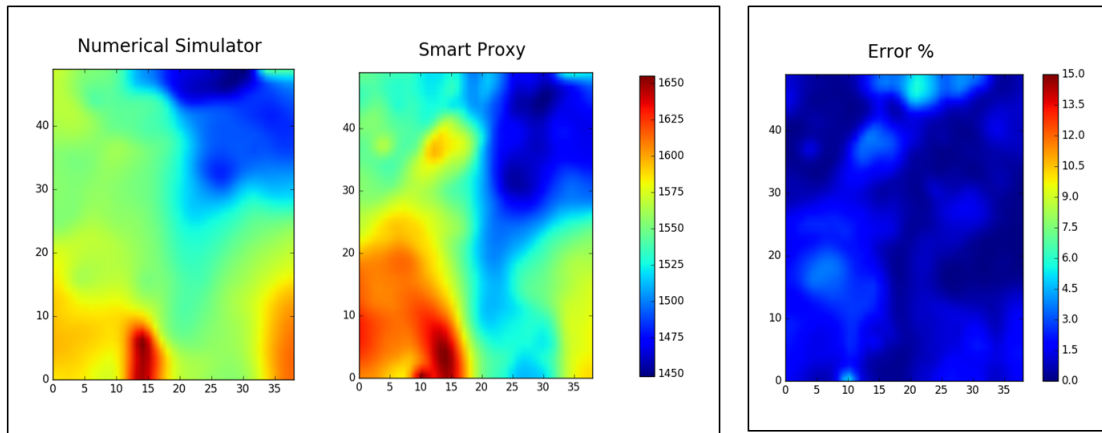
Oil Saturation– Layer 3 on 01/01/1957
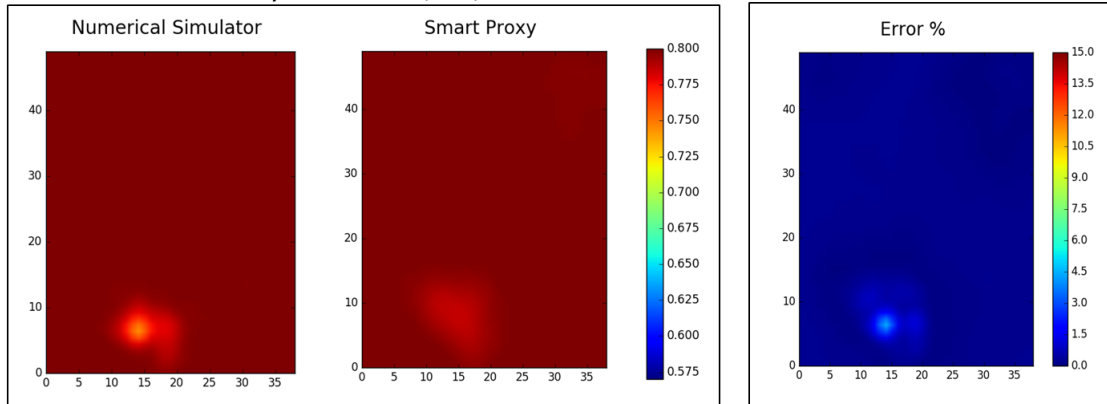
Figure 7.26: Phase one Layer-3 in 1957

Reservoir Pressure – Layer 5 on 01/01/1957
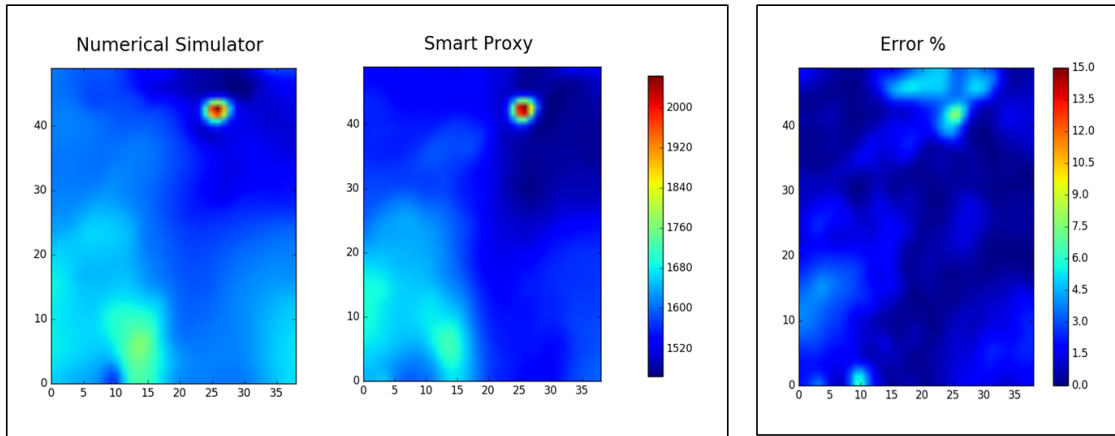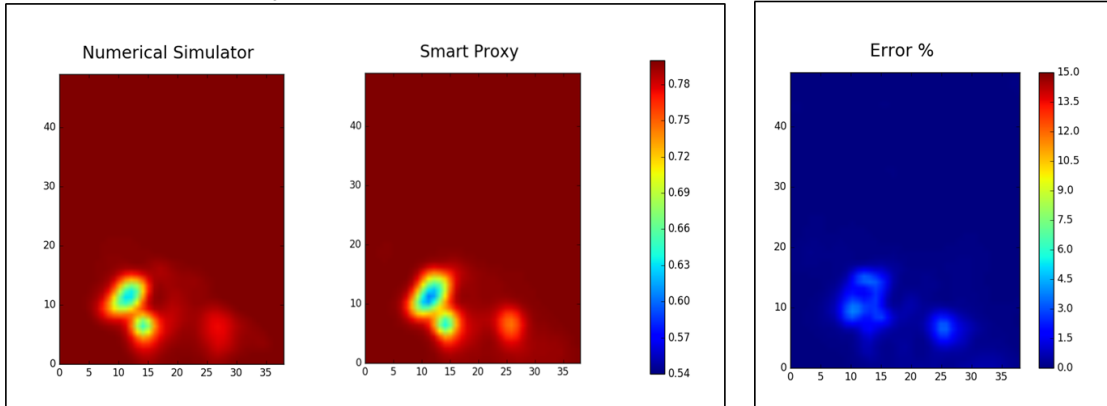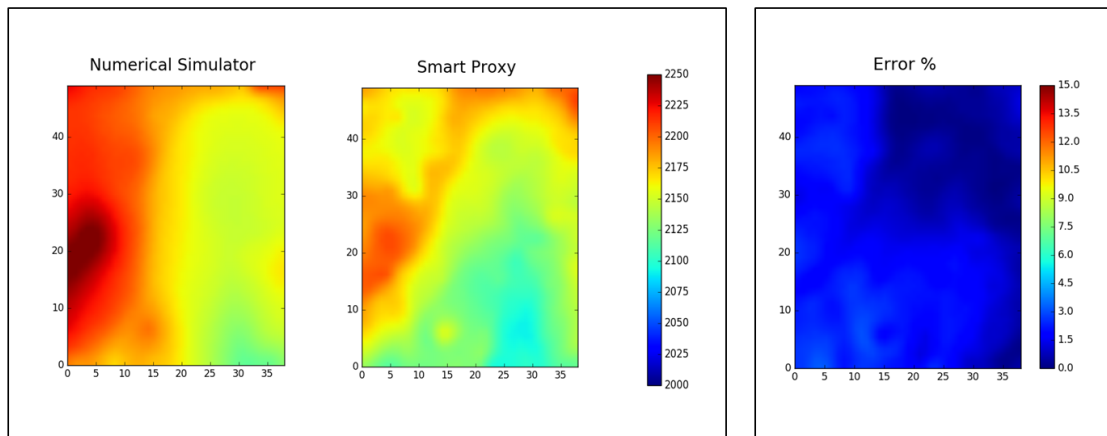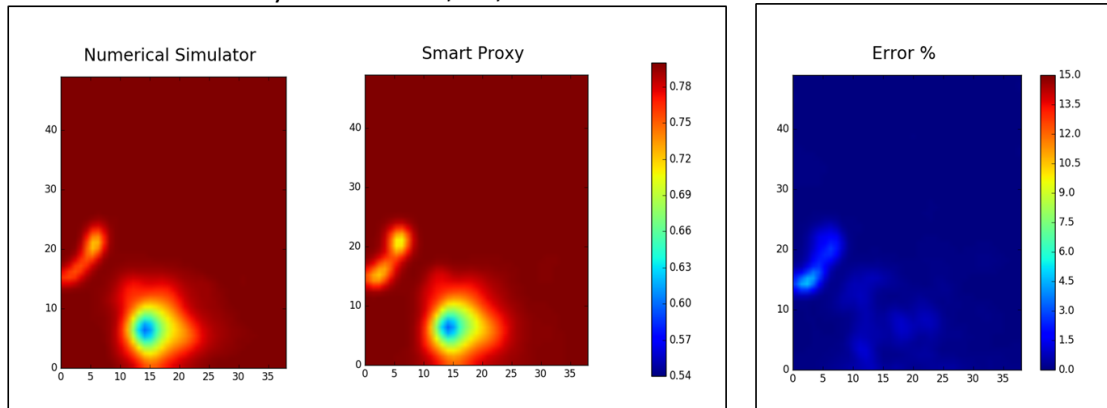

Oil Saturation– Layer 5 on 01/01/1957
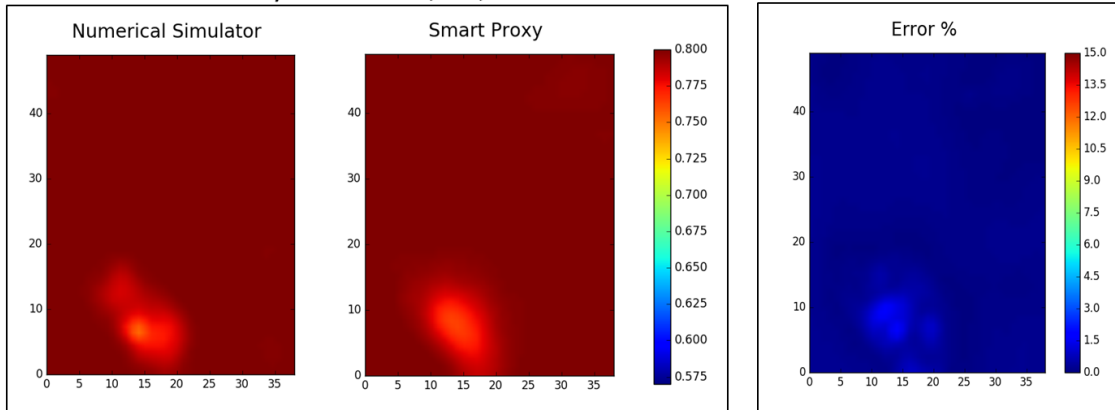
Figure 7.27: Phase one Layer-5 in 1957
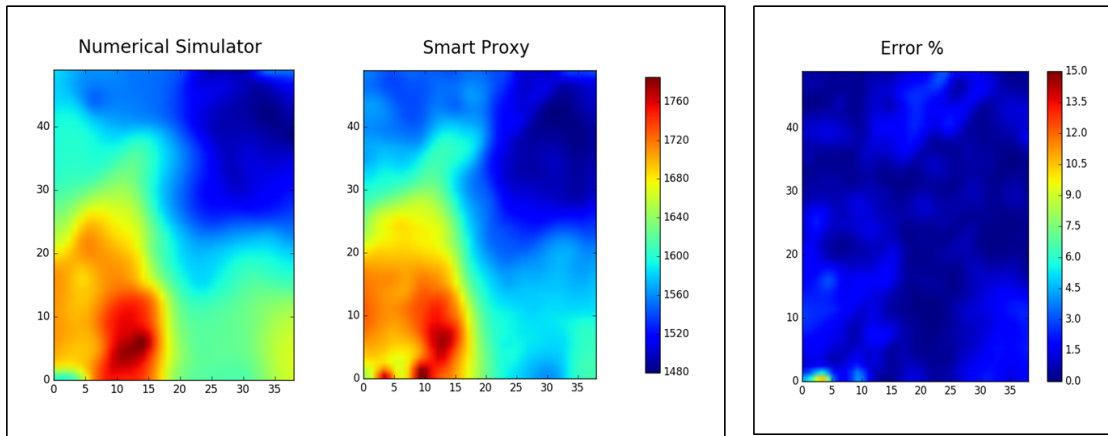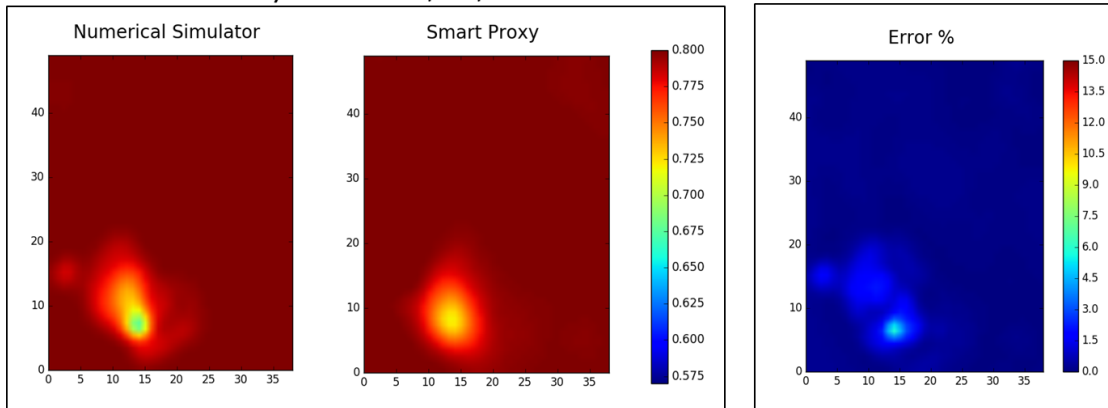
Figure 7.28: Phase one Layer-7 in 1957

Reservoir Pressure – Layer 10 on 01/01/1957
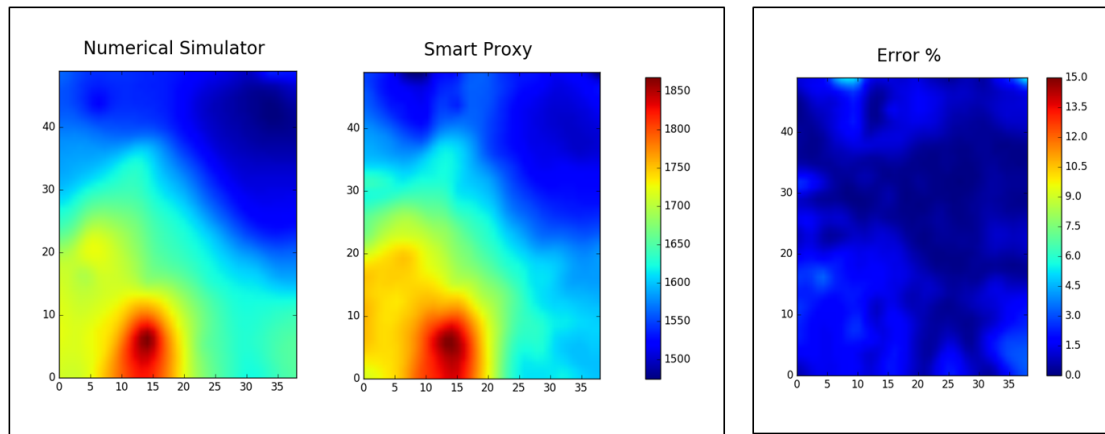
Oil Saturation– Layer 10 on 01/01/1957
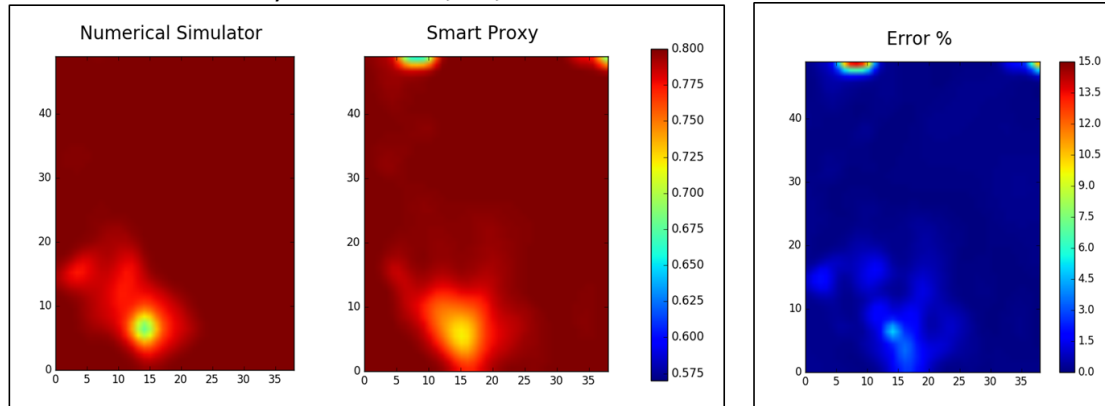
Figure 7.29: Phase one Layer-10 in 1957

Reservoir Pressure – Layer 16 on 01/01/1957
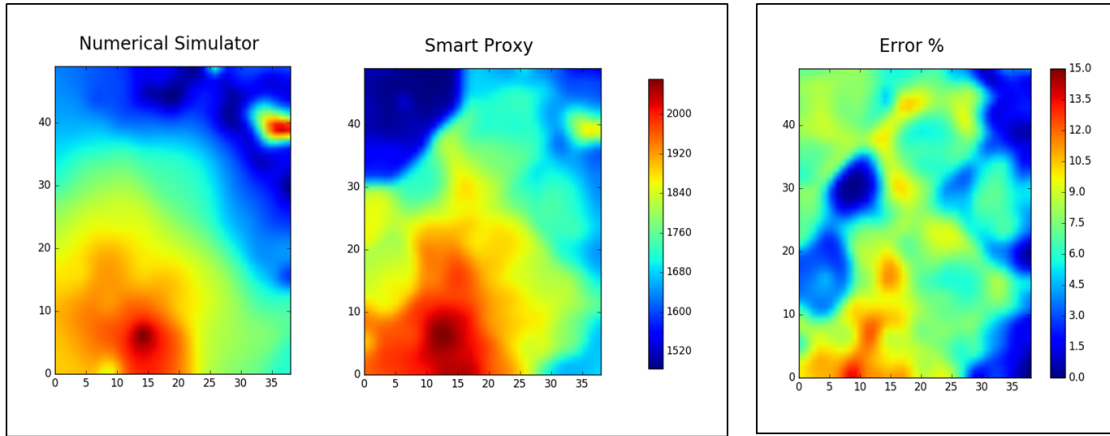
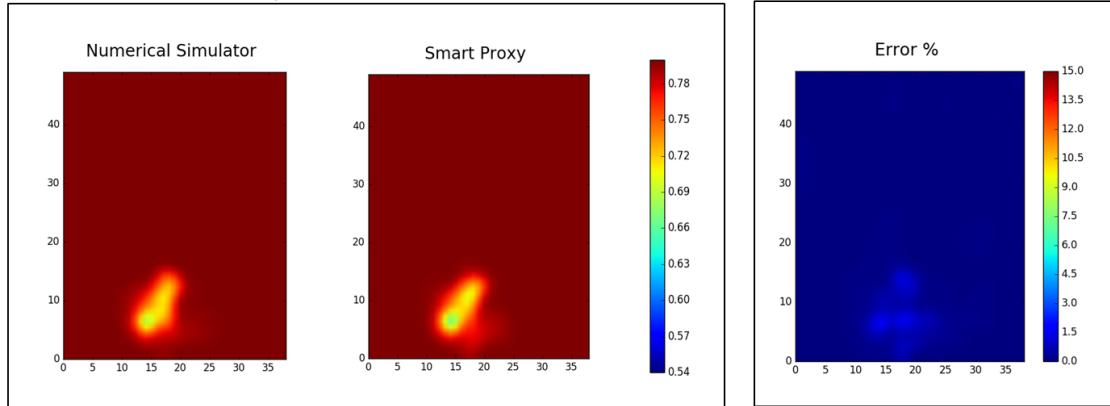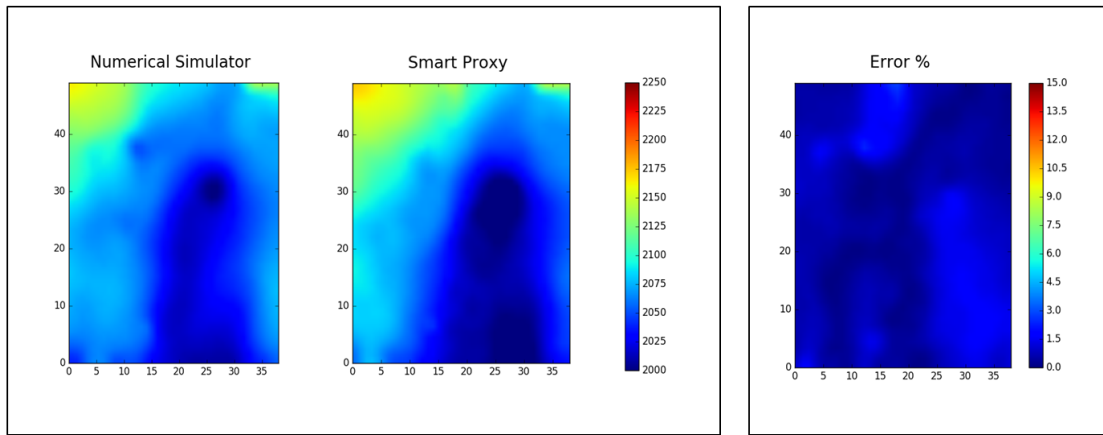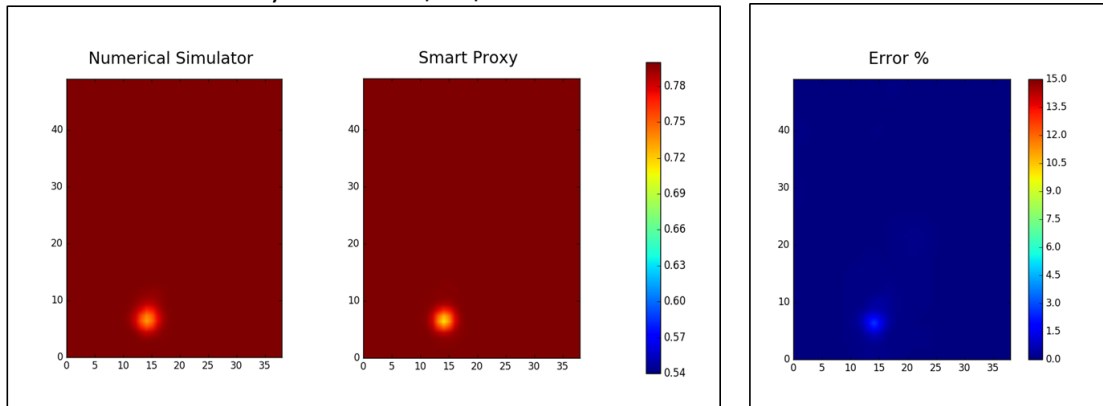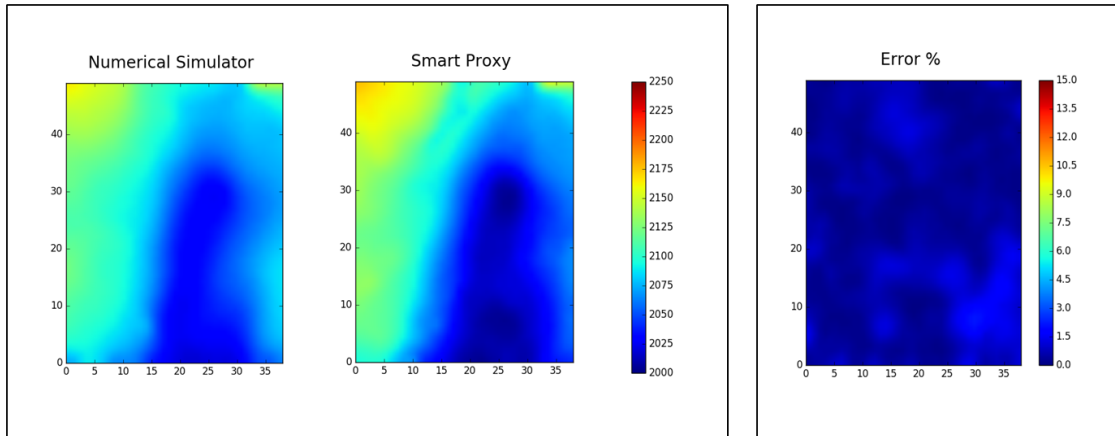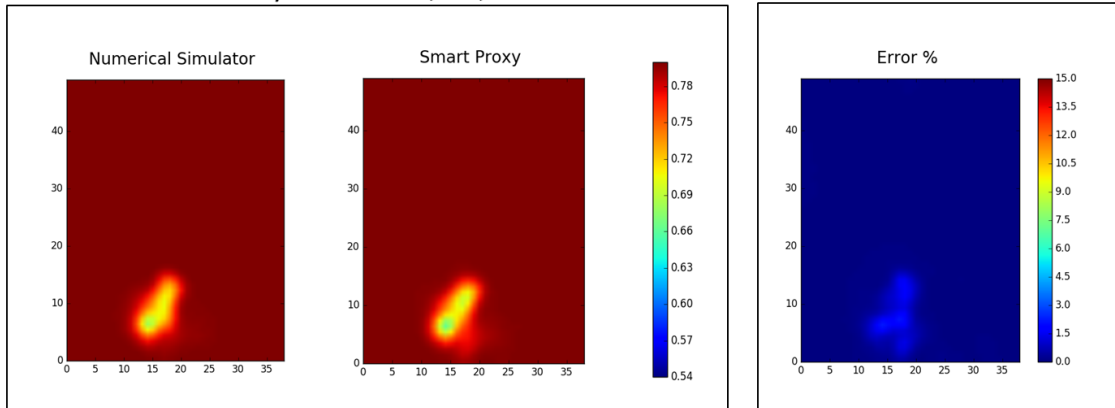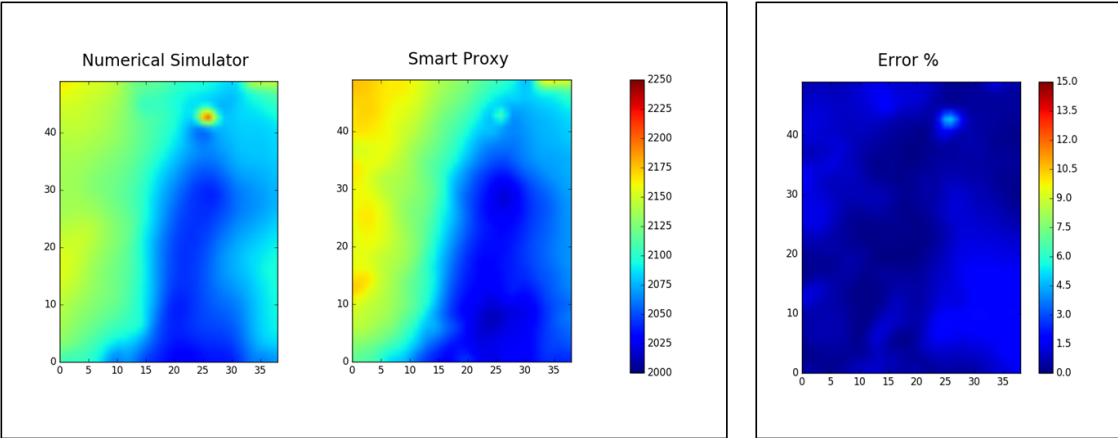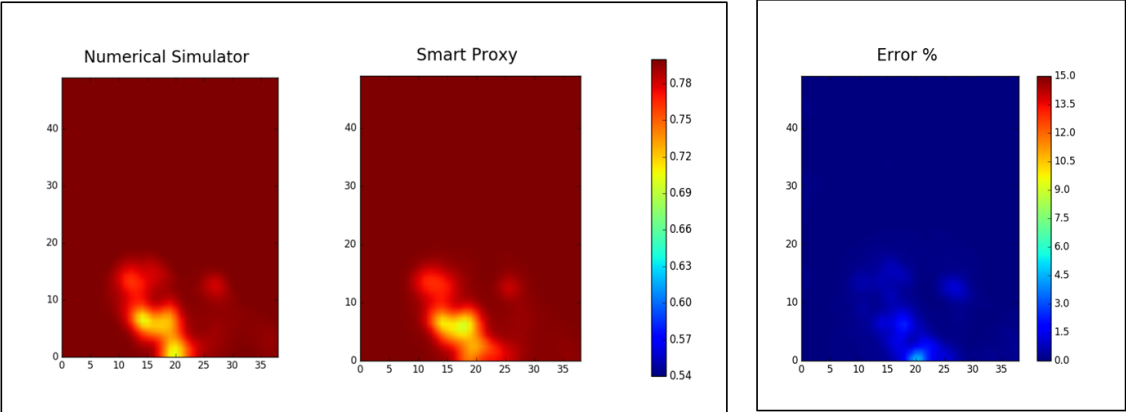## Oil Saturation– Layer 16 on 01/01/1957
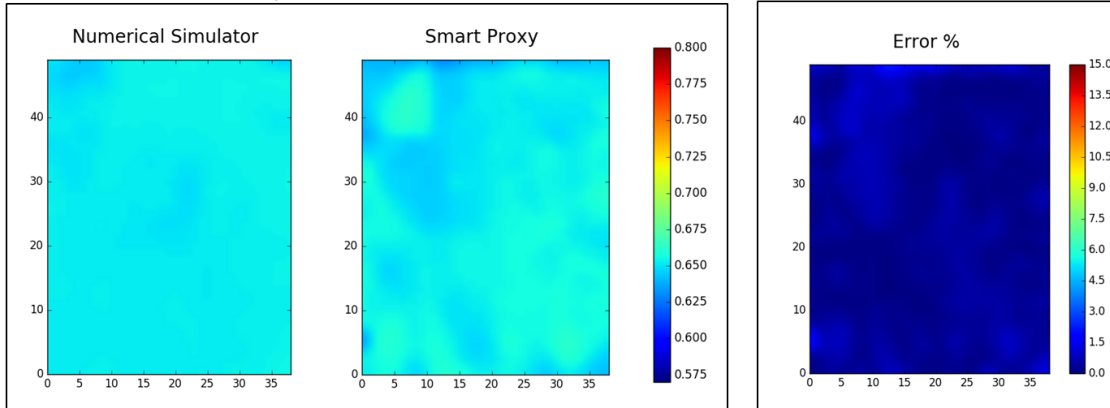
Figure 7.30: Phase one Layer-16 in 1957

7.4 Phase Two Smart Proxy Model

Phase two production performance started in 1960 and ended in 1970. In this phase, only one water injection well was added in the field. Also, the variation in the wells BHP during the 10 years of production was relatively small. The BHP for every production well was between 1950 psi to 2100 psi. The designed numerical simulation runs that generated the database and the construction of the neural networks were carried out following the same workflow as that used for phase one.
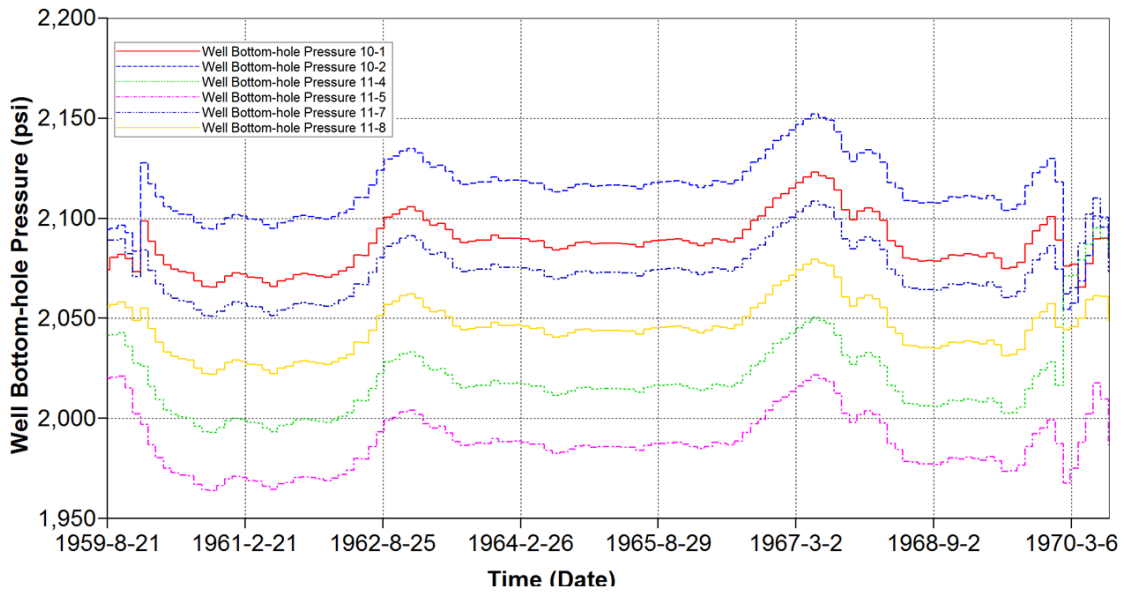
Figure 7.31: Phase Two BHP for Production Wells

7.4.1 Phase Two Training Results

The construction of the trained neural networks in this phase was similar to that used for phase one. There was a neural network trained for each reservoir property at each time step. In order to verify the training, the calibration data set was monitored for a better network performance.

Reservoir Property	Average R-squared coefficient
Grid Pressure	99%
Oil Saturation	99%
Water Saturation	99%

Table 7.2: Phase Two Training Results

7.4.2 Phase Two Blind Run Results

The validated smart proxy models were used to mimic the results produced by a blind numerical simulation run. The deployment process was implemented using a cascading feature. The smart proxy models showed robust performance with an acceptable range of error. The absolute average error at the first time step was under 1% and it reached about 2% by the last time step. For saturation, the absolute average error was 1% at the last time step.

The following figures show the smart proxy model results compared to the numerical simulator in form of distribution maps. Like for the phase one result, only selected time steps and geological layers are shown in this section.

Reservoir Pressure – Layer 1 on 01/01/1961
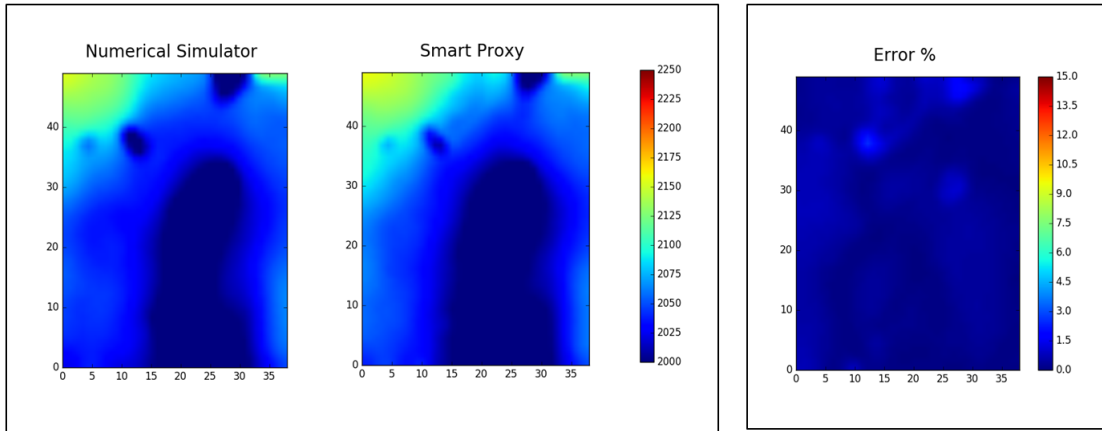
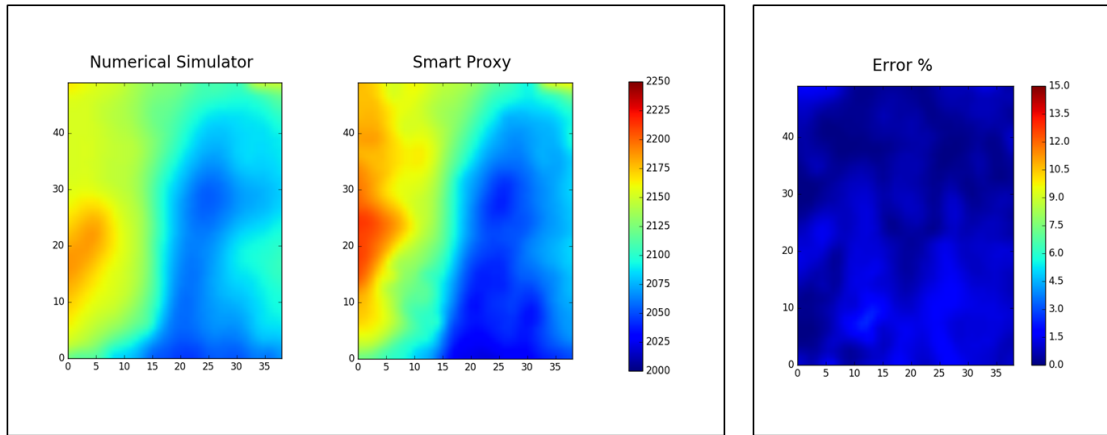Oil Saturation– Layer 1 on 01/01/1961
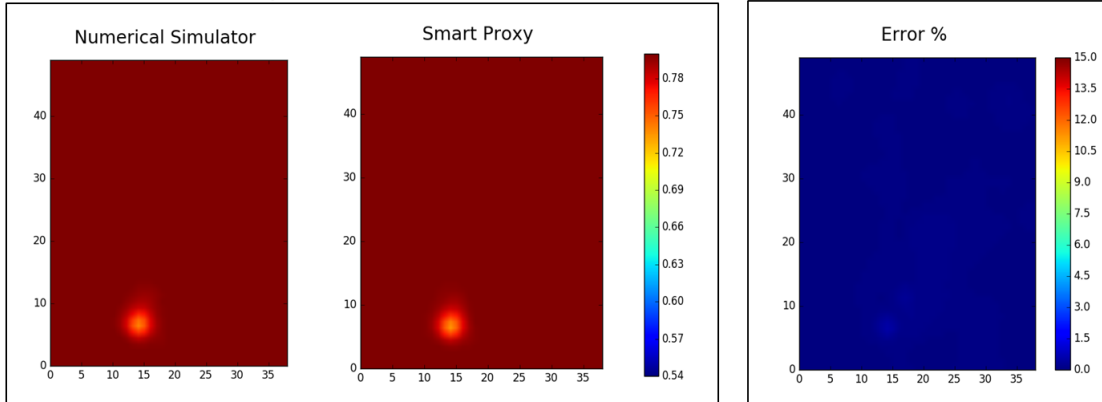
Figure 7.32: Phase two Layer-1 in 1961

Reservoir Pressure – Layer 3 on 01/01/1961
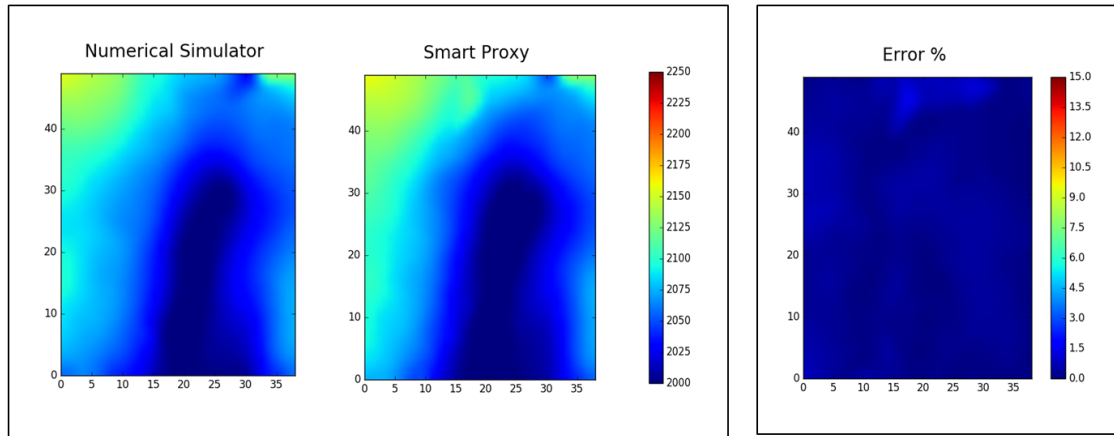
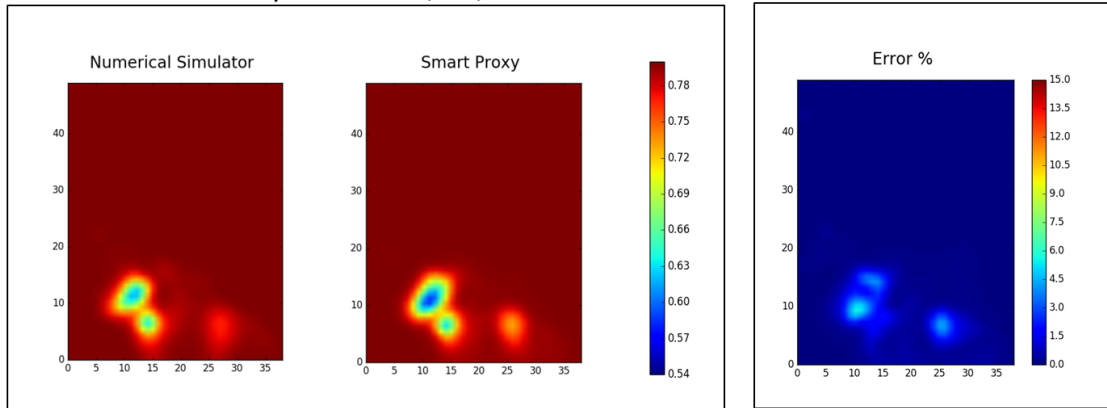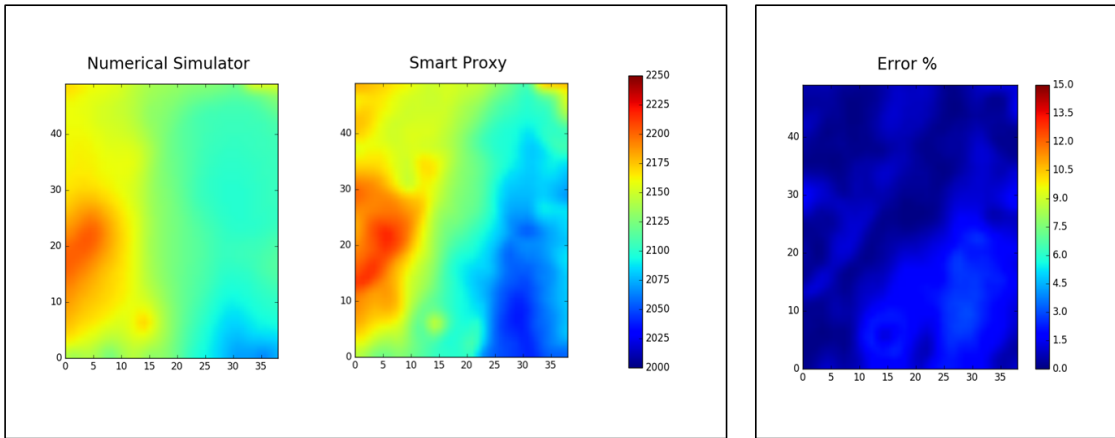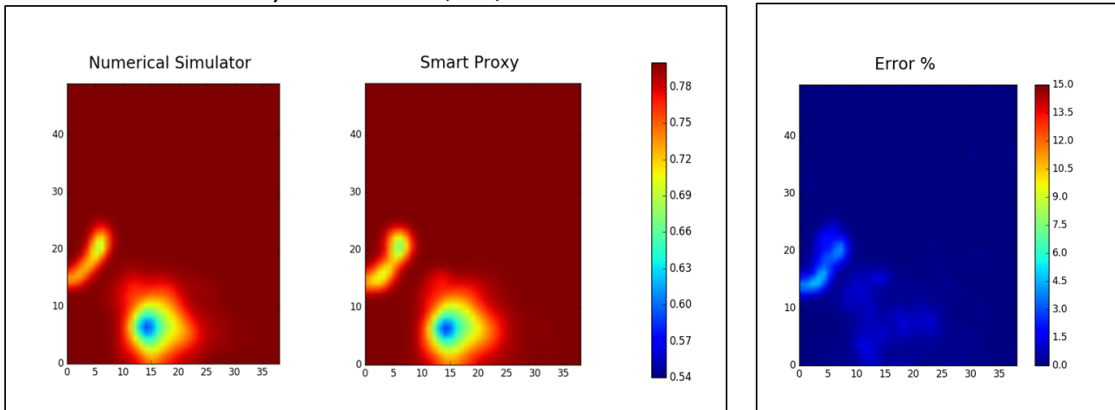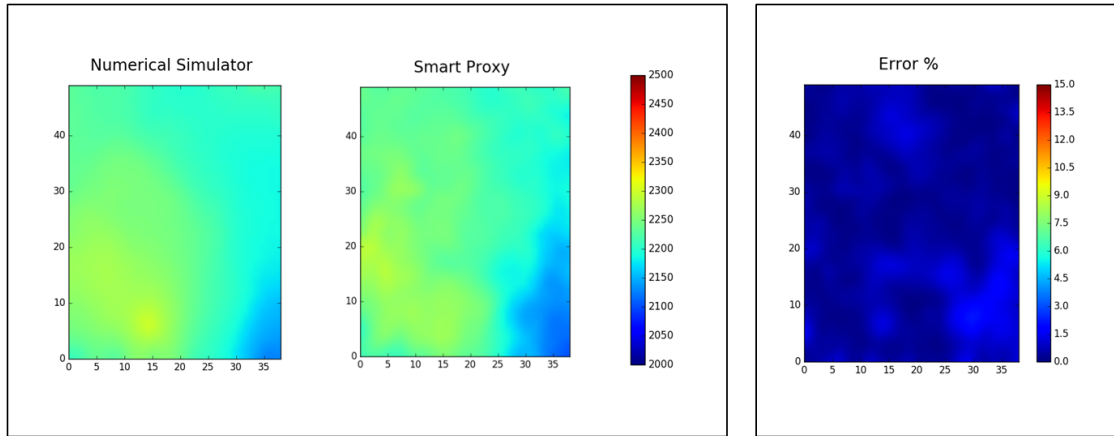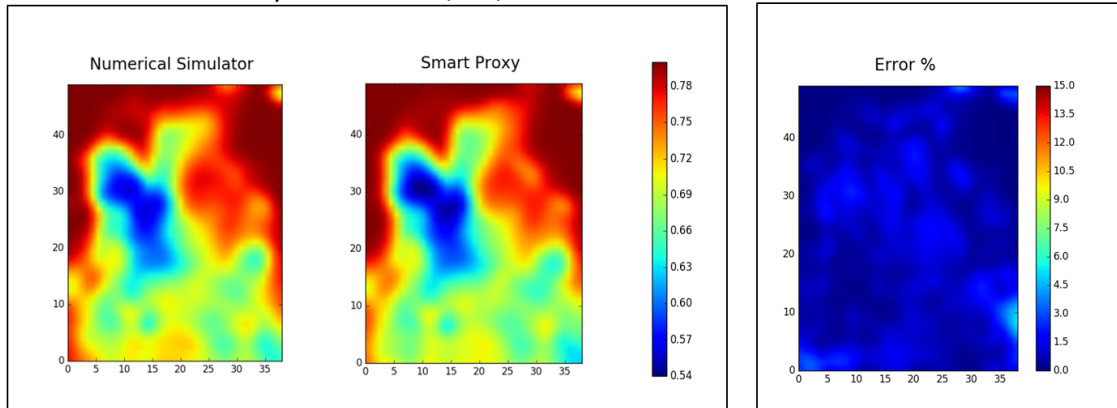# Oil Saturation– Layer 3 on 01/01/1961
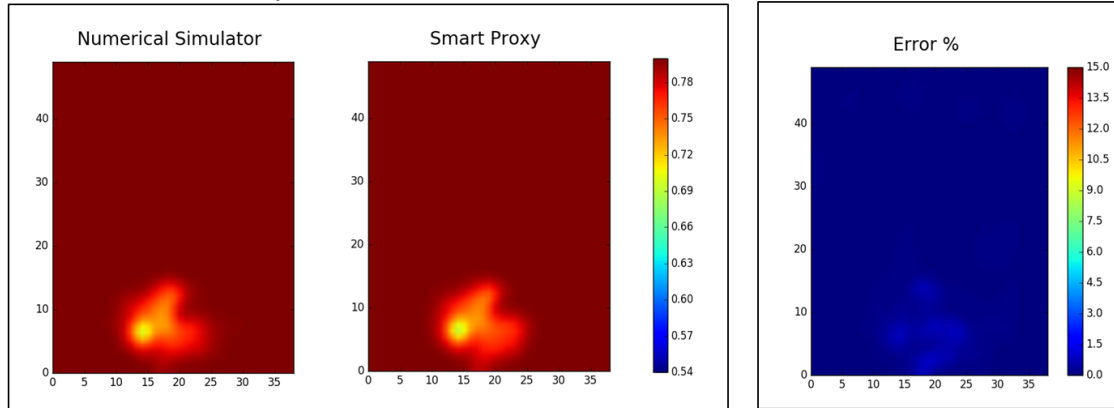
Figure 7.33: Phase two Layer-3 in 1961

Reservoir Pressure – Layer 5 on 01/01/1961
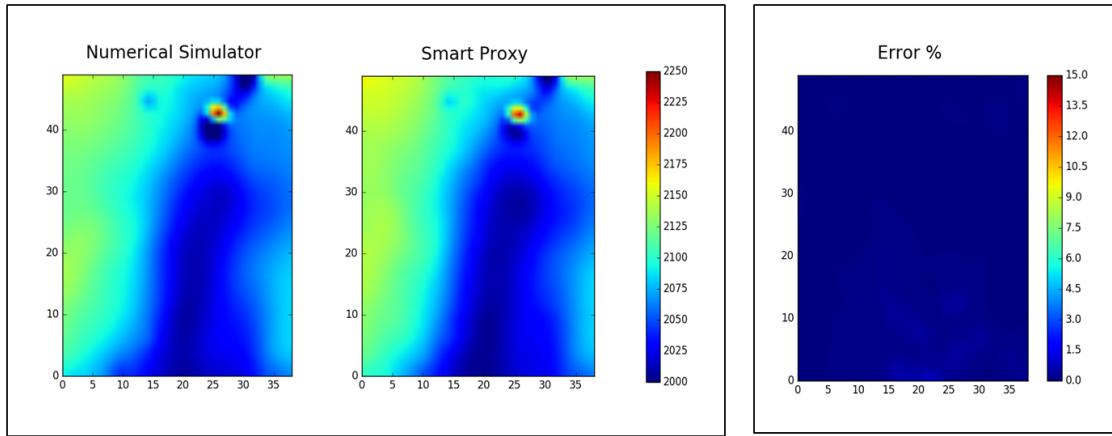
Oil Saturation– Layer 5 on 01/01/1961
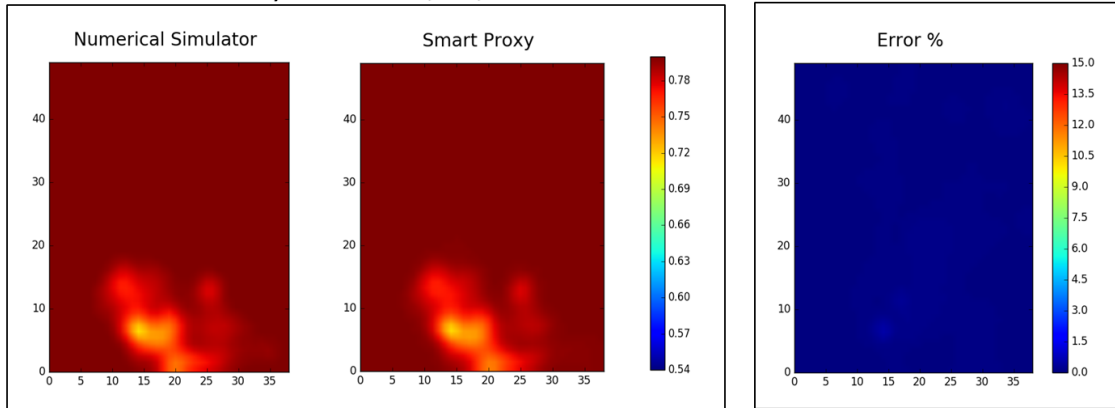
Figure 7.34: Phase two Layer-5 in 1961
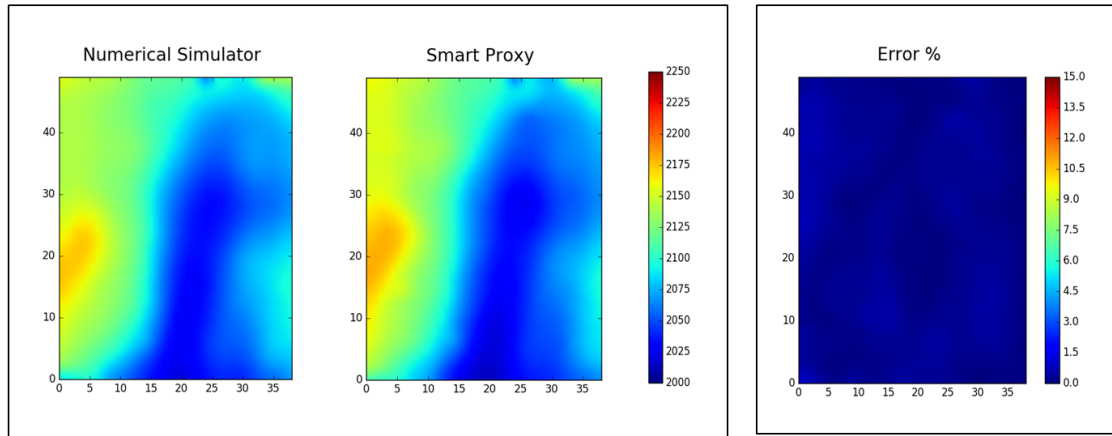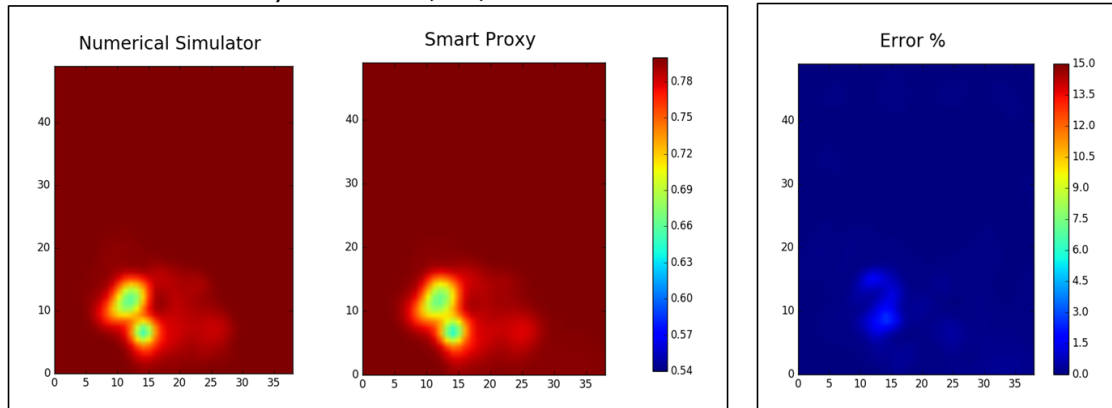
Figure 7.35: Phase two Layer-7 in 1961
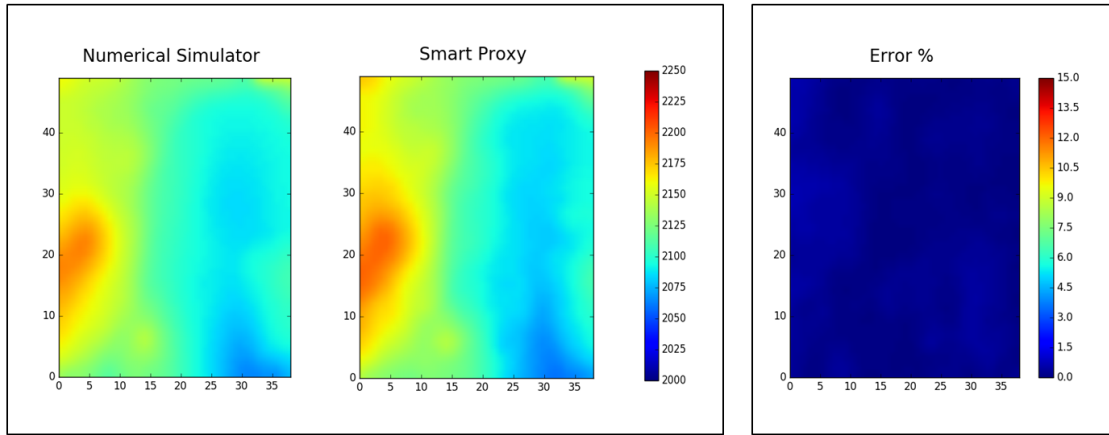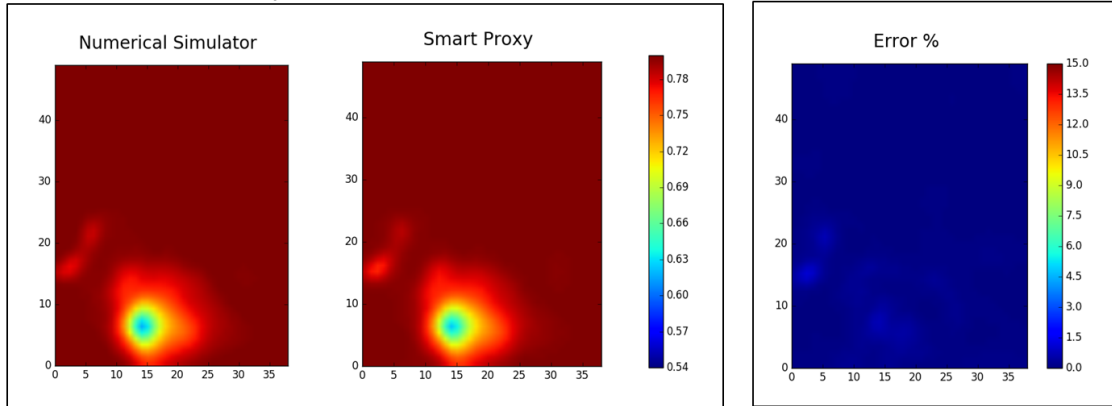
Figure 7.36: Phase two Layer-10 in 1961

Figure 7.37: Phase two Layer-13 in 1961

Reservoir Pressure – Layer 1 on 01/01/1963
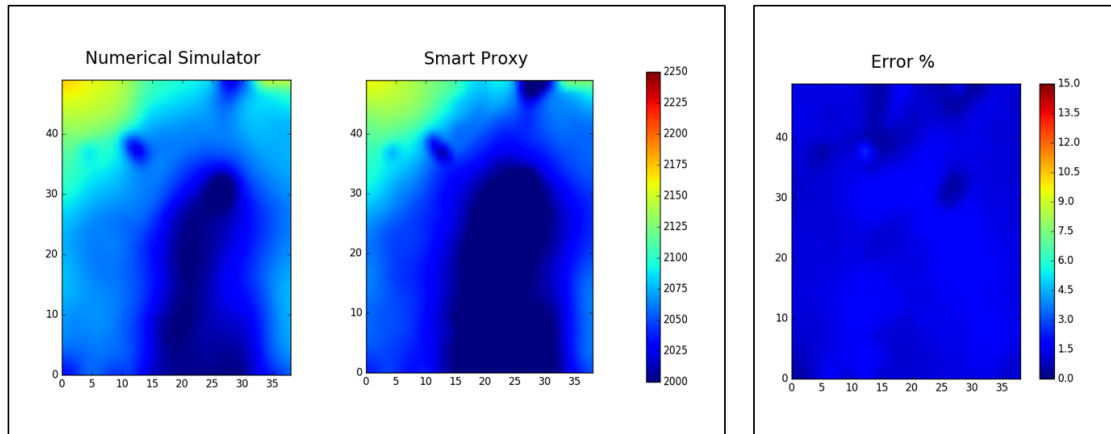
Oil Saturation– Layer 1 on 01/01/1963
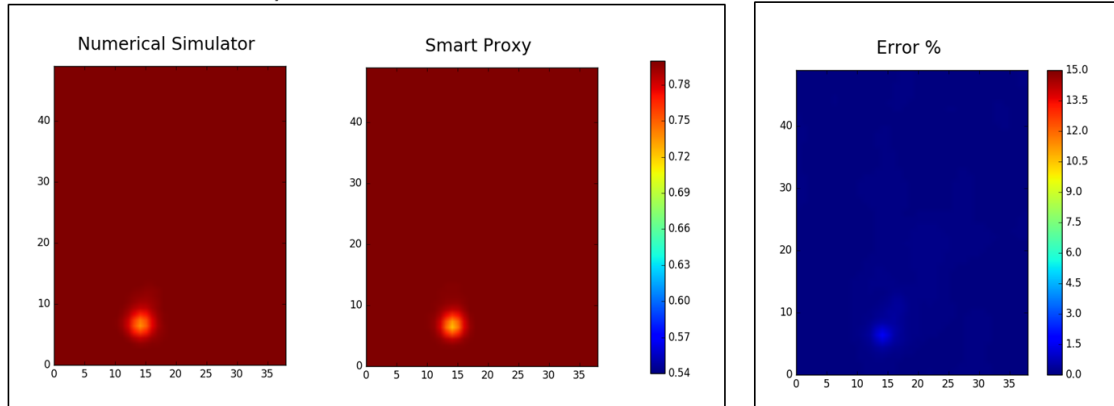
Figure 7.38: Phase two Layer-1 in 1963

Reservoir Pressure – Layer 3 on 01/01/1963
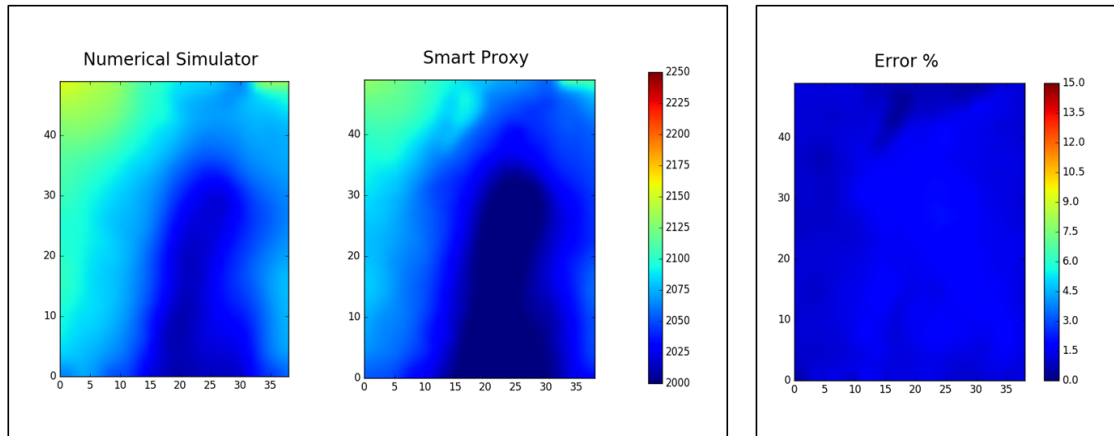
Oil Saturation– Layer 3 on 01/01/1963
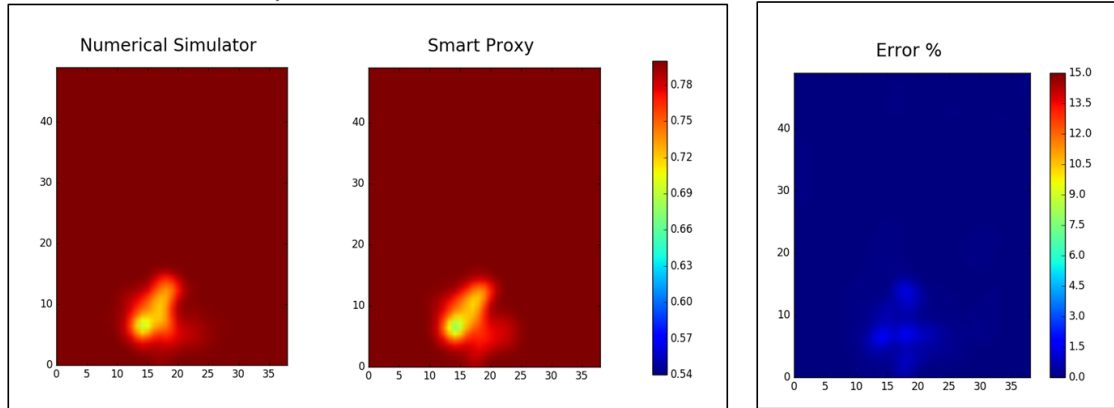
Figure 7.39: Phase two Layer-3 in 1963

Reservoir Pressure – Layer 5 on 01/01/1963
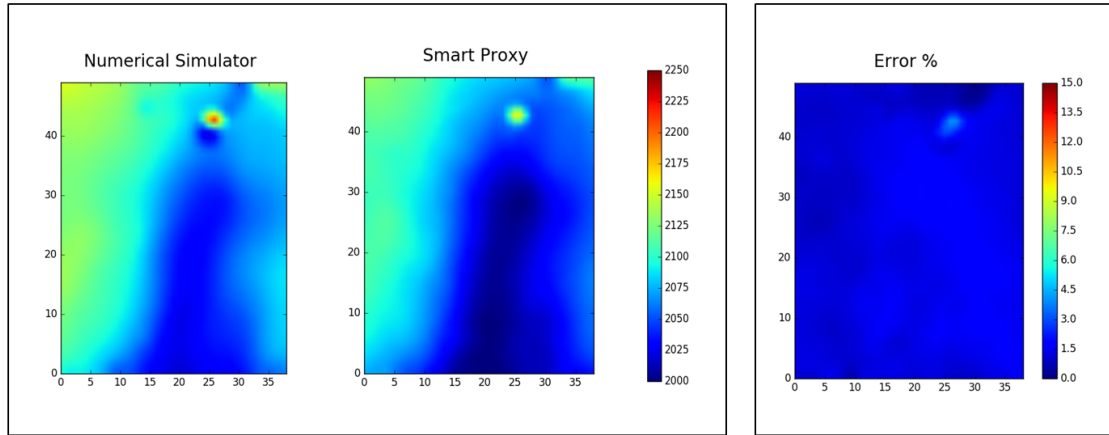
Oil Saturation– Layer 5 on 01/01/1963
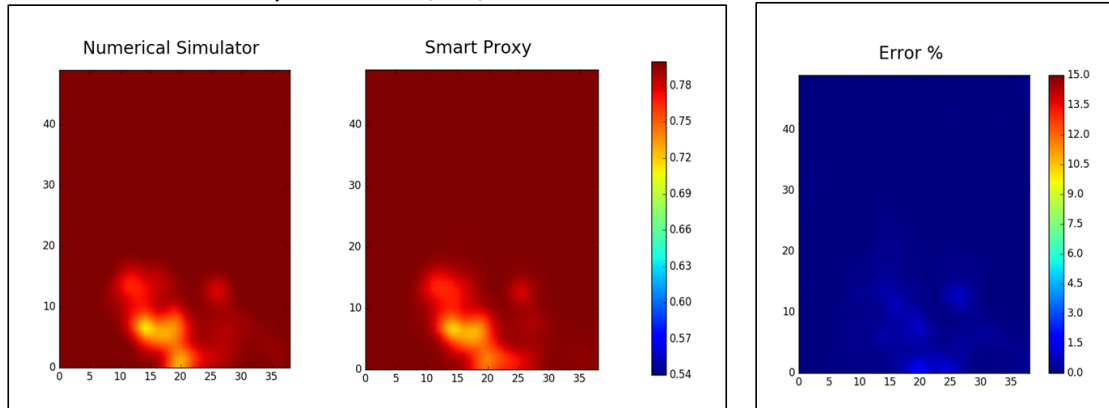
Figure 7.40: Phase two Layer-5 in 1963

Reservoir Pressure – Layer 7 on 01/01/1963
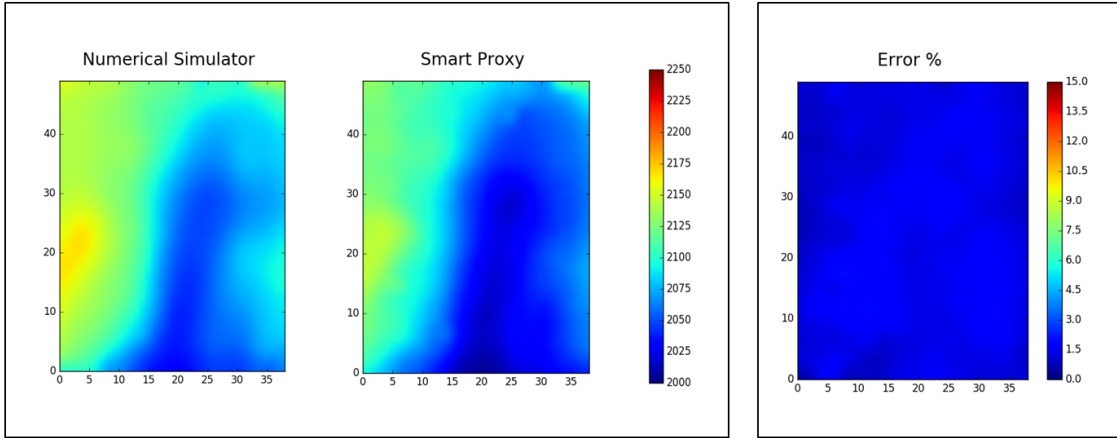
Oil Saturation– Layer 7 on 01/01/1963
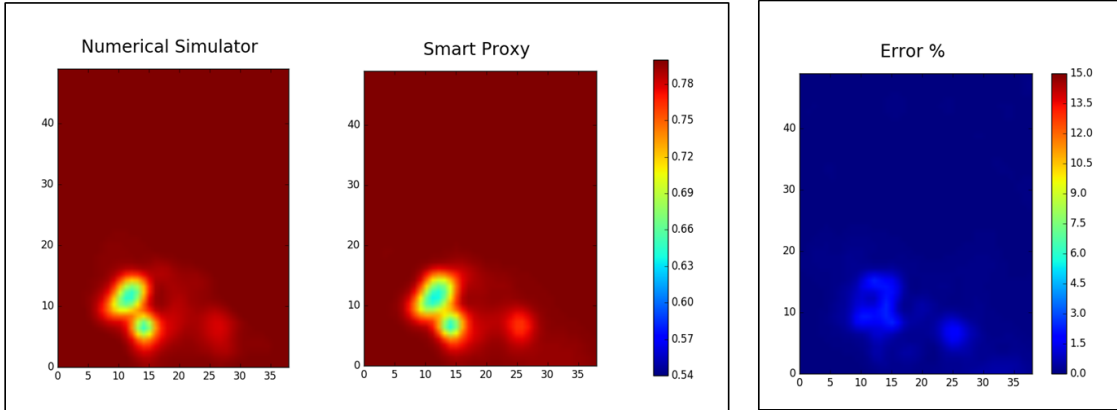
Figure 7.41: Phase two Layer-7 in 1963

Reservoir Pressure – Layer 10 on 01/01/1963
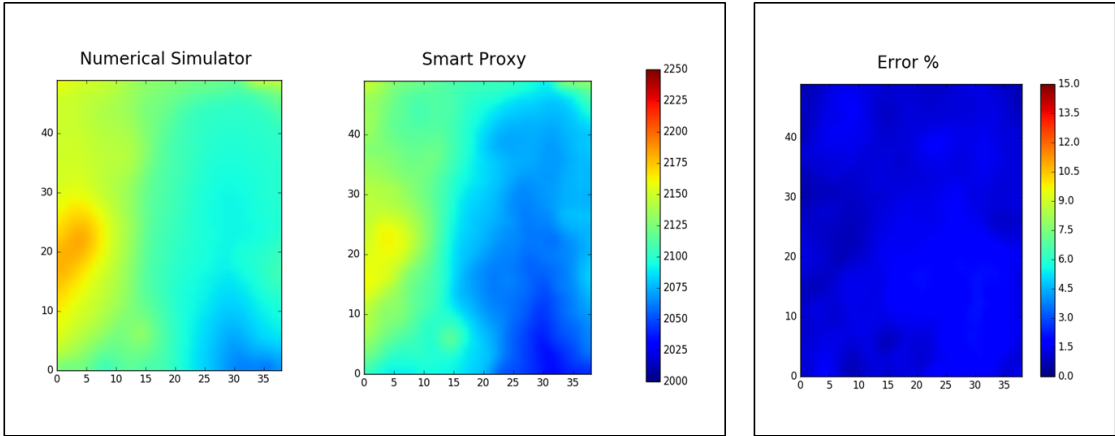
Oil Saturation– Layer 10 on 01/01/1963
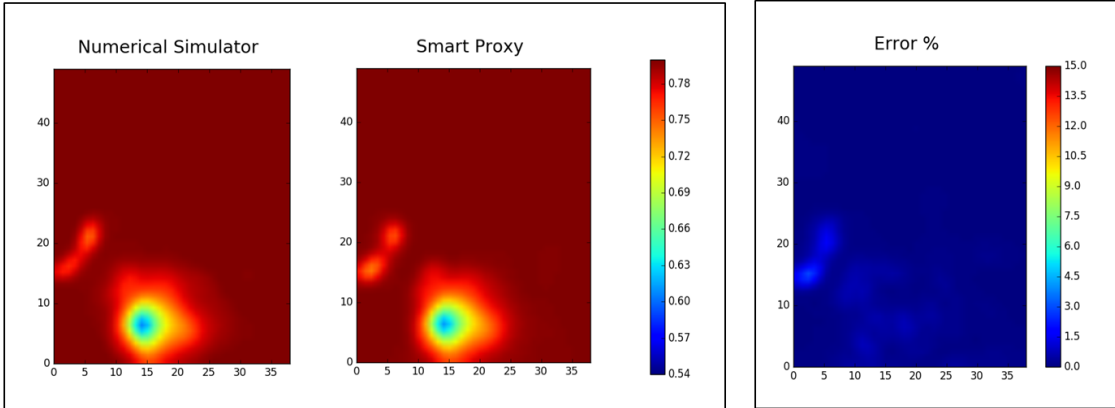
Figure 7.42: Phase two Layer-10 in 1963

138

Reservoir Pressure – Layer 13 on 01/01/1963
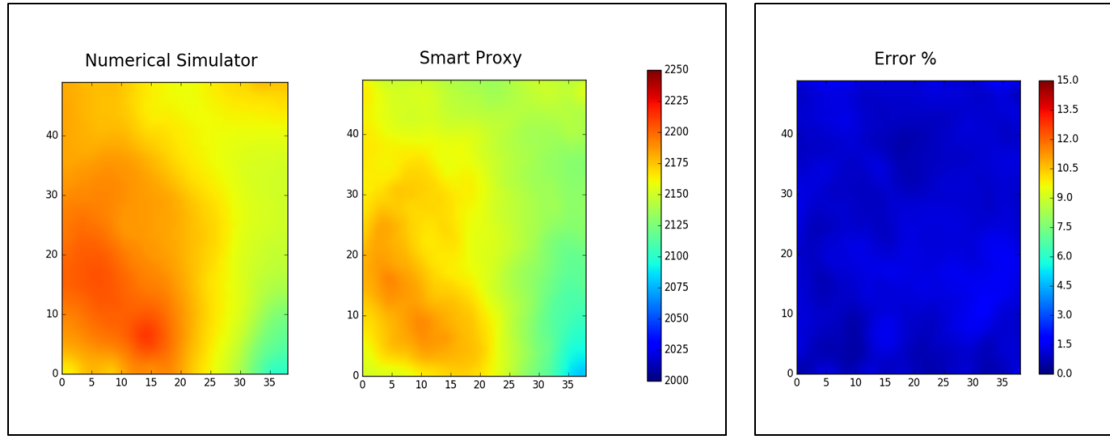
Oil Saturation– Layer 13 on 01/01/1963
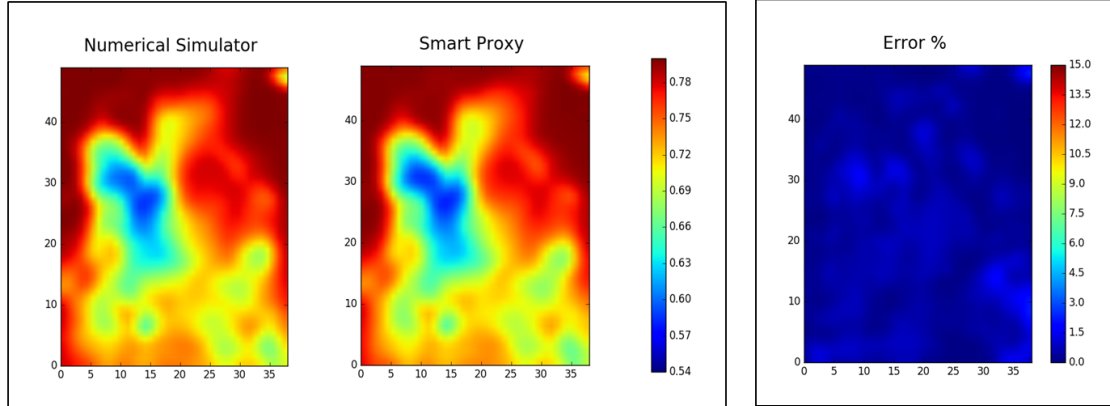
Figure 7.43: Phase two Layer-13 in 1963
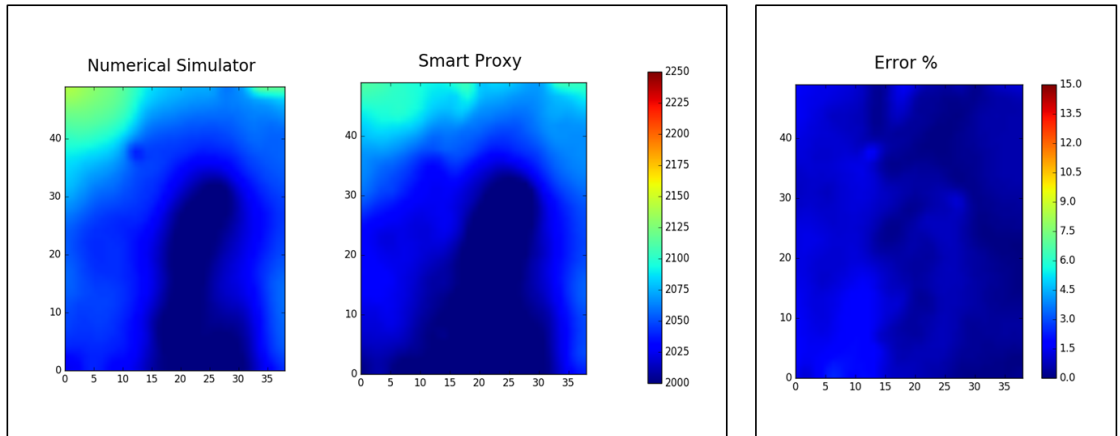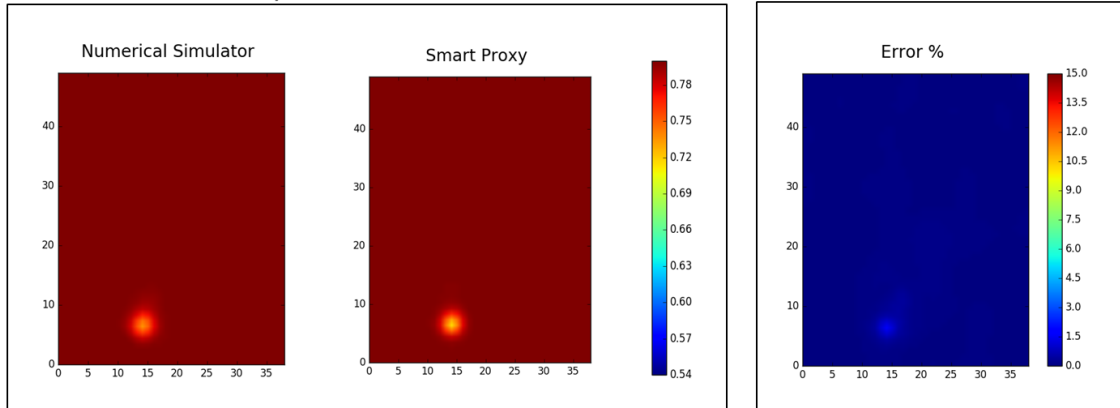
Figure 7.44: Phase two Layer-1 in 1965

Reservoir Pressure – Layer 3 on 01/01/1965
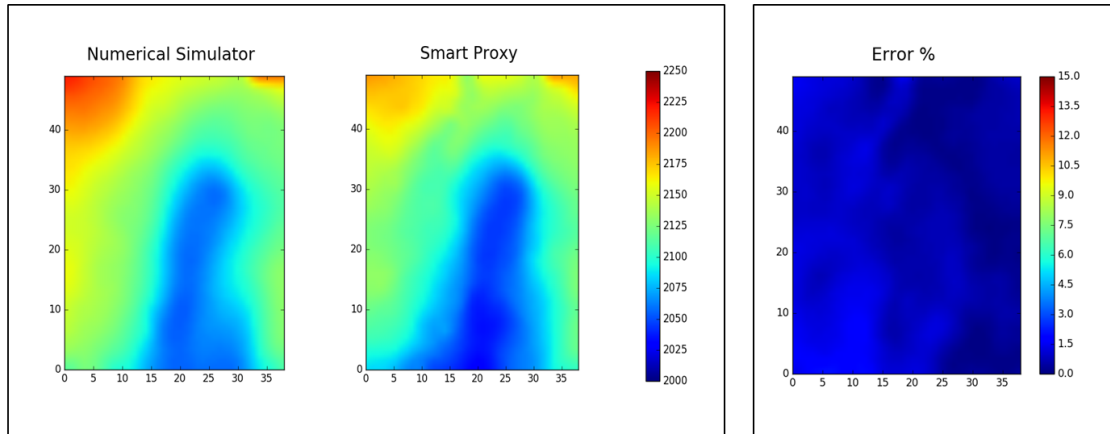
Oil Saturation– Layer 3 on 01/01/1965
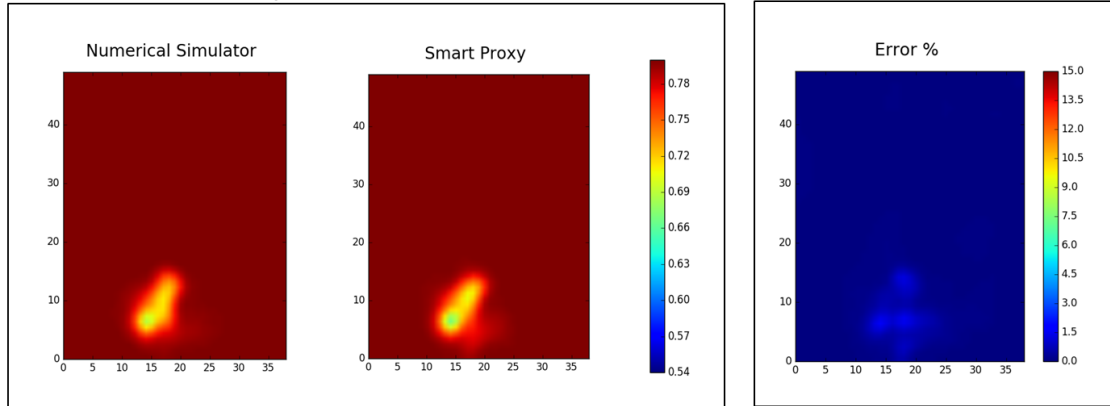
Figure 7.45: Phase two Layer-3 in 1965

Reservoir Pressure – Layer 5 on 01/01/1965
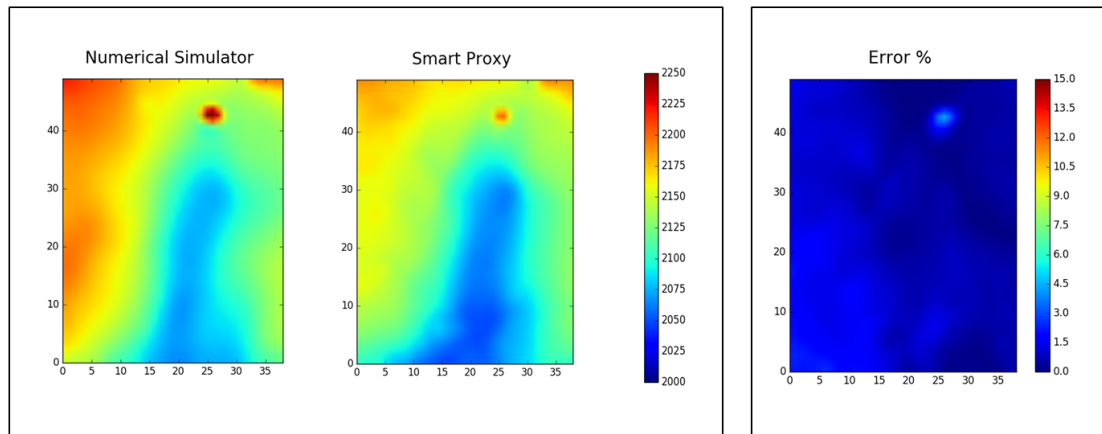
Oil Saturation– Layer 5 on 01/01/1965
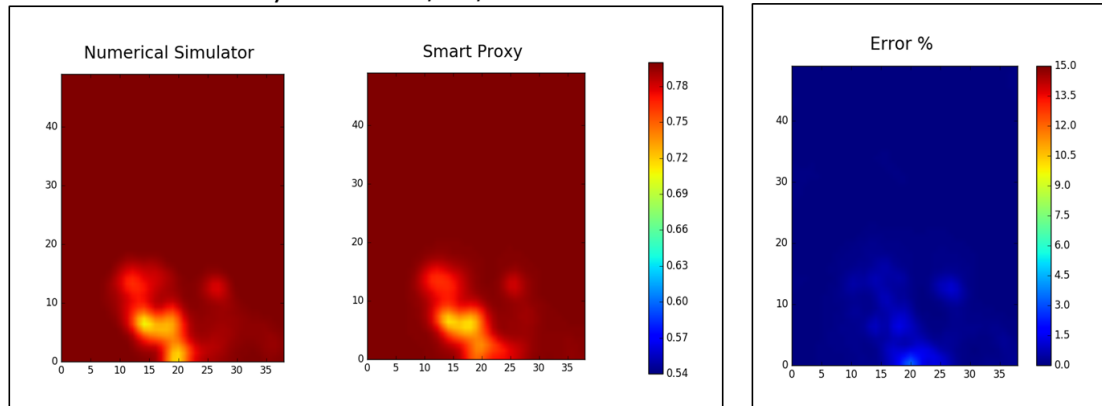
Figure 7.46: Phase two Layer-5 in 1965

Reservoir Pressure – Layer 7 on 01/01/1965
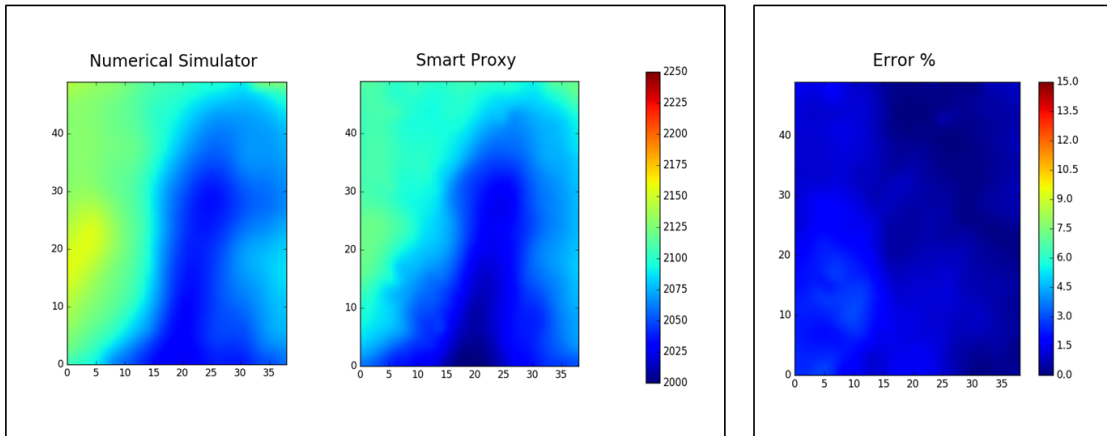
Oil Saturation– Layer 7 on 01/01/1965

Figure 7.47: Phase two Layer-7 in 1965

143

Figure 7.48: Phase two Layer-10 in 1965

Reservoir Pressure – Layer 3 on 01/01/1965
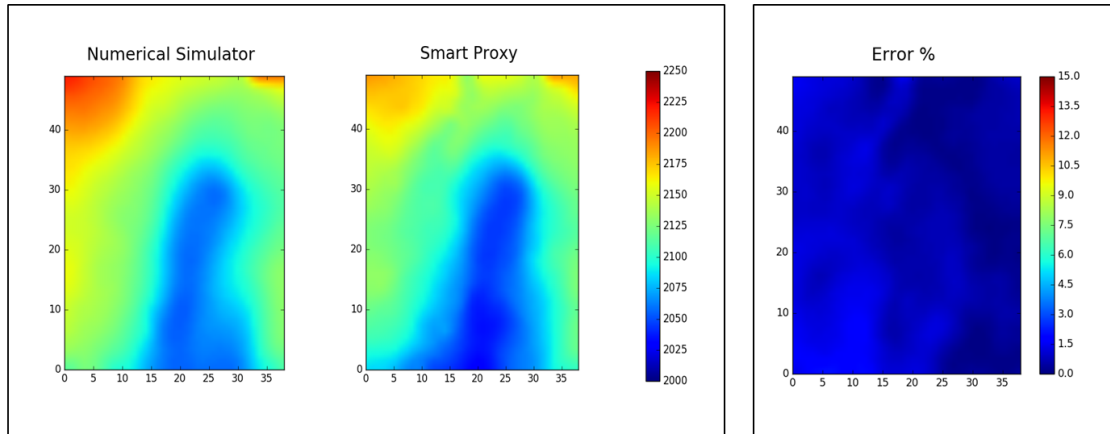
Oil Saturation– Layer 3 on 01/01/1965

Figure 7.49: Phase two Layer-13 in 1965

Reservoir Pressure – Layer 1 on 01/01/1967

Oil Saturation– Layer 1 on 01/01/1967

Figure 7.50: Phase two Layer-1 in 1967

Reservoir Pressure – Layer 3 on 01/01/1967

Oil Saturation– Layer 3 on 01/01/1967

Figure 7.51: Phase two Layer-3 in 1967

Reservoir Pressure – Layer 5 on 01/01/1967

Oil Saturation– Layer 5 on 01/01/1967



Figure 7.52: Phase two Layer-5 in 1967

Figure 7.53: Phase two Layer-7 in 1967

Reservoir Pressure – Layer 10 on 01/01/1967

Oil Saturation– Layer 10 on 01/01/1967

Figure 7.54: Phase two Layer-10 in 1967

Figure 7.55: Phase two Layer-13 in 1967

7.5   Phase Three Smart Proxy Model

Phase three of the field production profile is the longest in duration and the most active production phase. In this phase, there were up to 15 production wells and up to 12 water injection wells. The production duration of this phase was about 20 years.

Again, in this phase, the smart proxy development procedure was the same as that used for the first and second phases. However, in this phase, the training and validation was particularly challenging, due to the number of time steps. There were 20 time steps, requiring the development of 20 smart proxies.

Using the cascading feature, it is necessary to have a very well trained model at each time step in order to keep the validation and deployment error low. The error propagation needs to be considered with this type of training. Therefore, the training should be performed carefully to minimize the error at each time step.
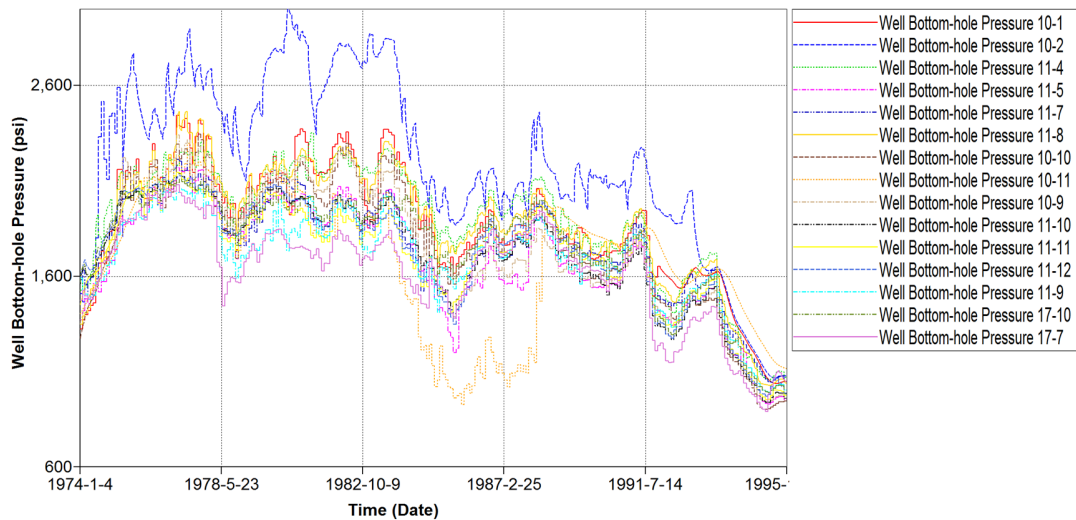
Figure 7.56: Phase Three BHP for Production Wells

7.5.1   Phase Three Training Results

Similar to phase one and phase two, neural networks were trained for reservoir pressure and saturation. For each reservoir property, 20 neural networks were trained and validated. The selected neural network inputs along with carefully designed simulation runs resulted in the required training results. The following table summarizes the neural networks training in this phase.

Reservoir Property	Average R-squared coefficient
Grid Pressure	99%
Oil Saturation	99%
Water Saturation	99%

Table 7.3: Phase Three Training Results

7.5.2   Phase Three Blind Run Results

Once the smart proxy had been developed and trained, its capability for generalization was tested by applying it to a blind data set. The blind data set was generated using a new numerical simulation run that had not been used in the training process. In this phase, the same blind simulation run as that used in the previous phases was used to verify the phase three smart proxy model.

The developed proxy deployment performed with a cascading feature. In other words, 20 proxy models were used in this process. The initial time step used the reservoir pressure and saturation from the numerical simulator. However, the next time step used the reservoir pressure and saturation from the output of the initial time step deployment. This process was continued until the last time step had been reached.

The deployment results showed a low average absolute error between the numerical simulator and the smart proxy at most of the time steps. However, due to the error propagation, the average absolute error at the last time step for reservoir pressure was 11%. On the other hand, the saturation average absolute error was under 2% at all-time steps. The following data distribution maps show the comparison between the numerical simulator and smart proxy for reservoir pressure and saturation. Because of substantial amount of data maps that were generated from the results of this

phase, only selected time steps (years) and layers are presented in this section.

Reservoir Pressure – Layer 1 on 01/01/1978
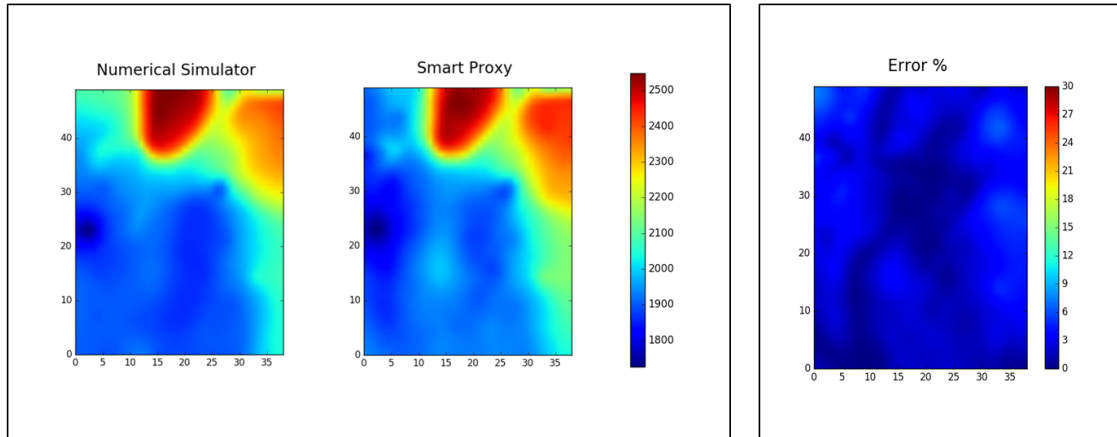
Oil Saturation– Layer 1 on 01/01/1978
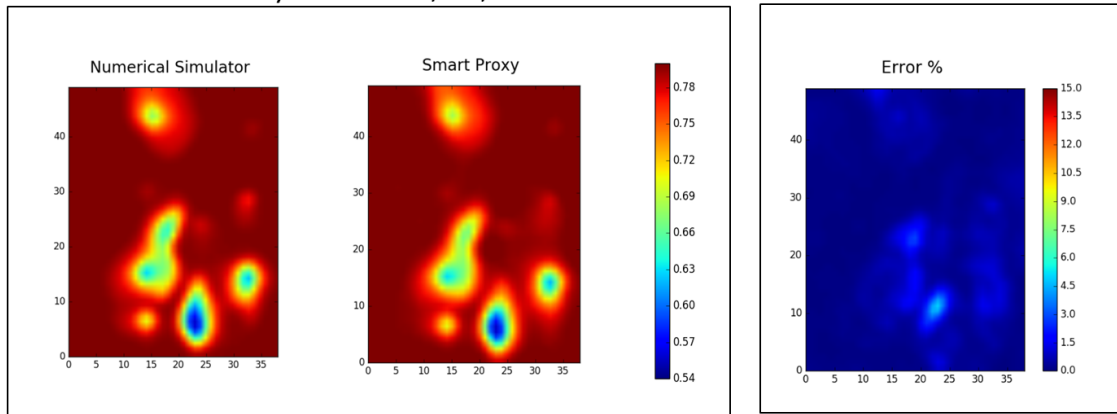
Figure 7.57: Phase three Layer-1 in 1978

Reservoir Pressure – Layer 4 on 01/01/1978
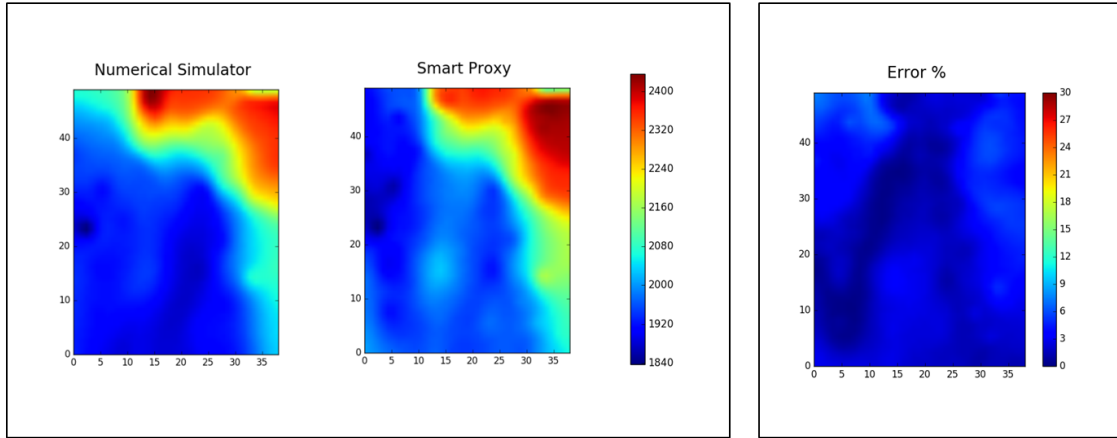
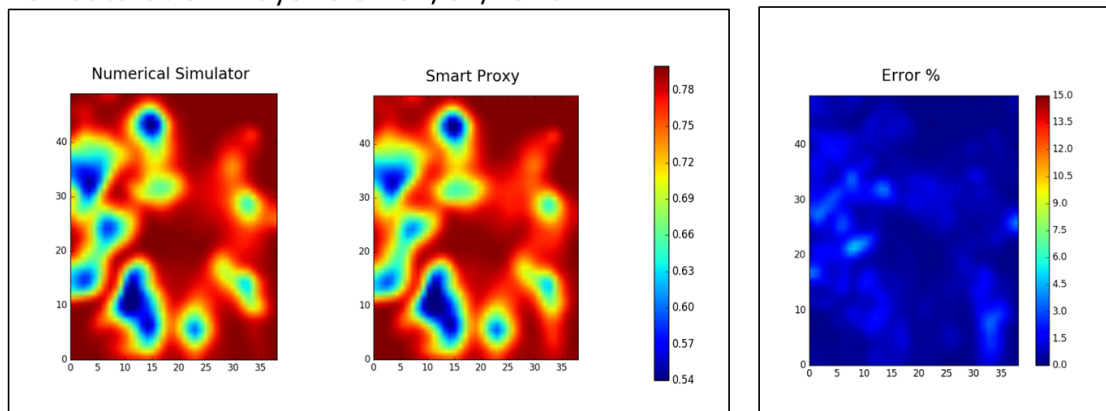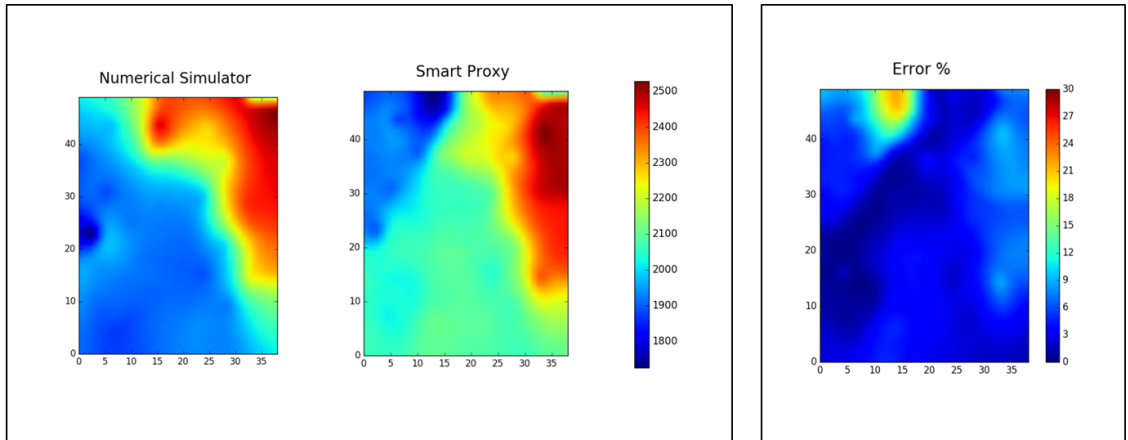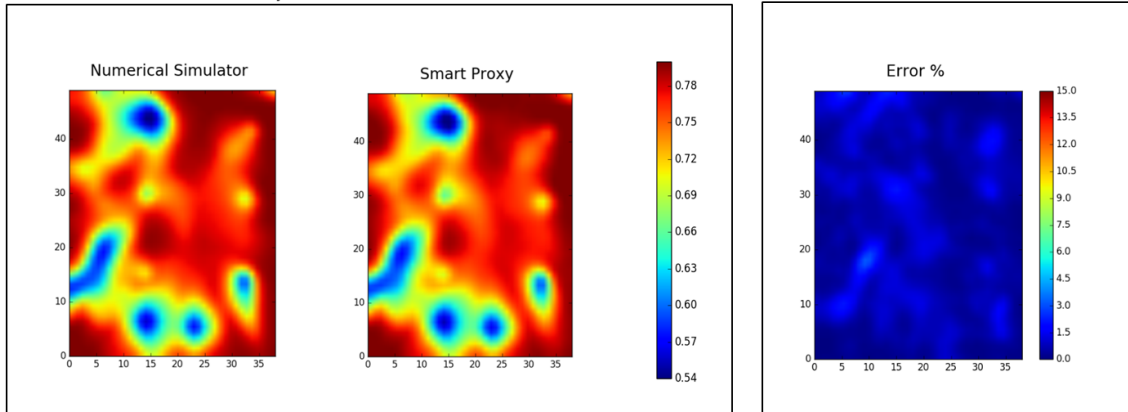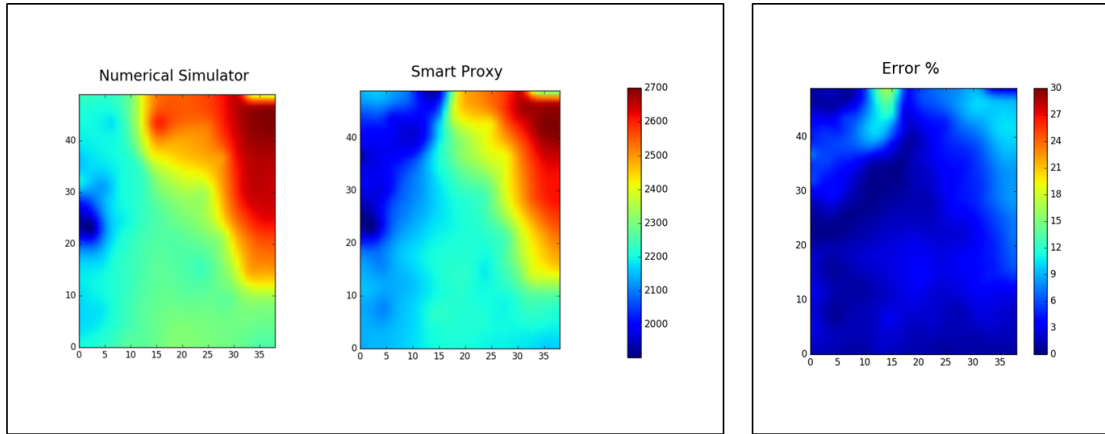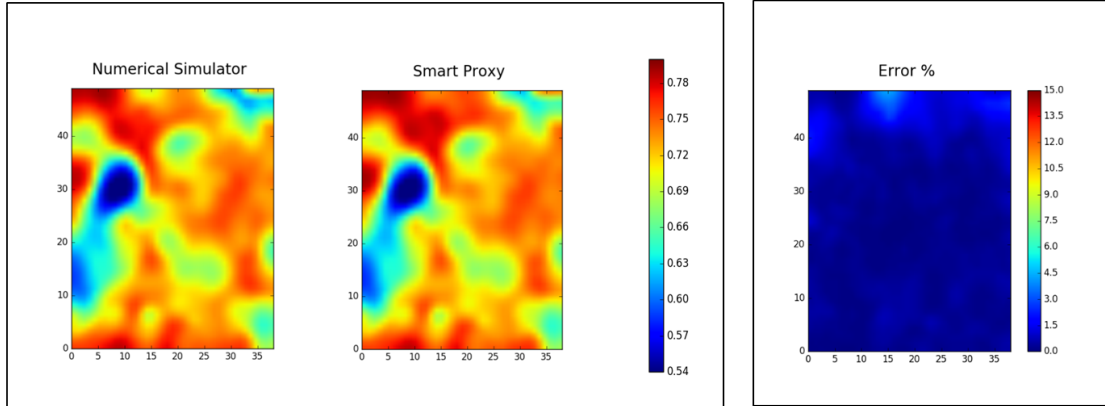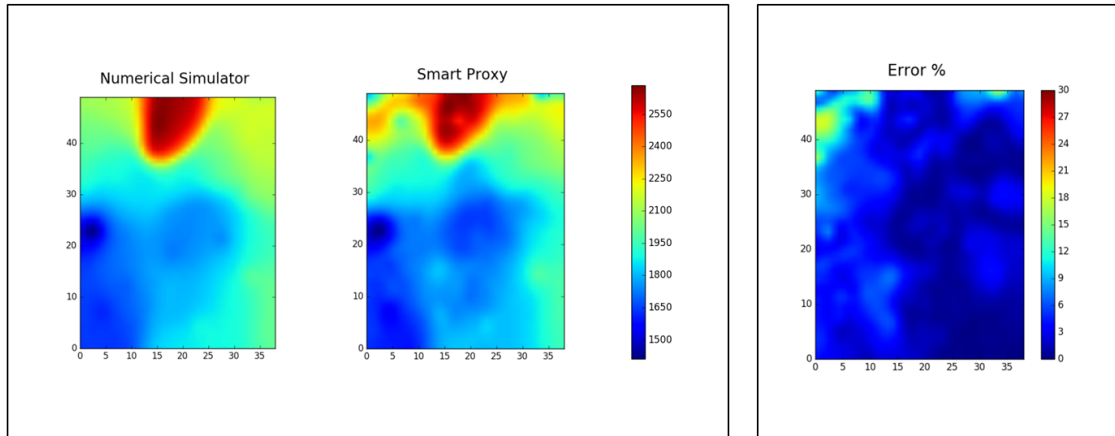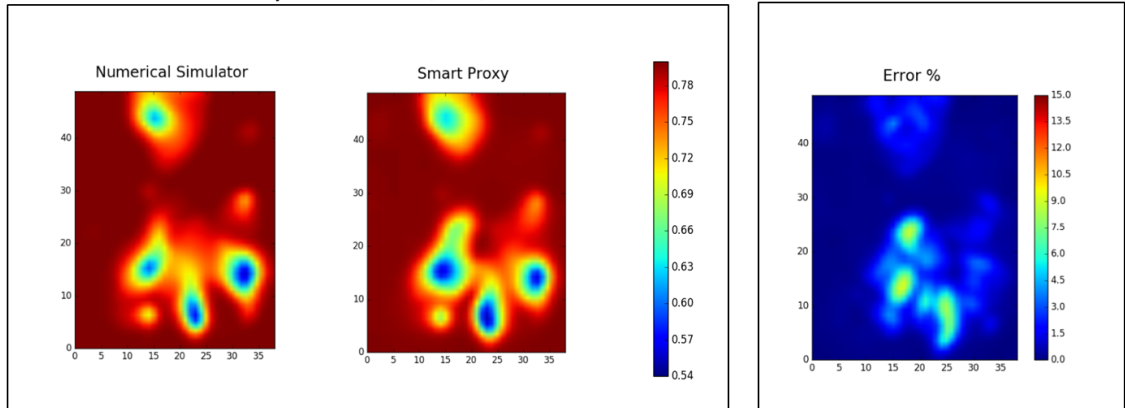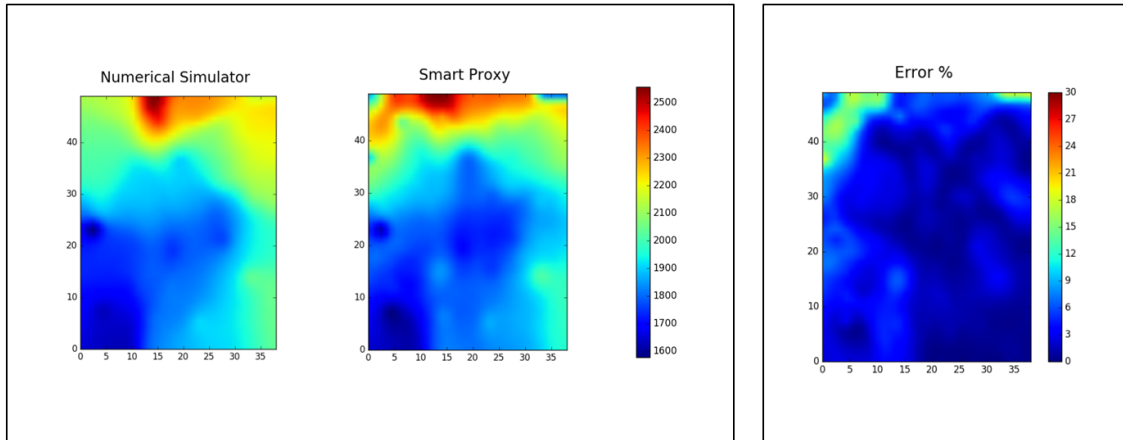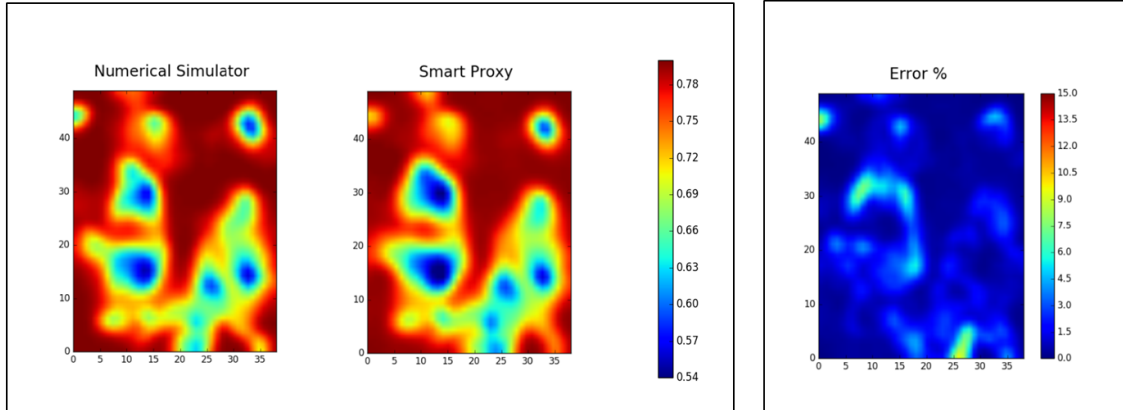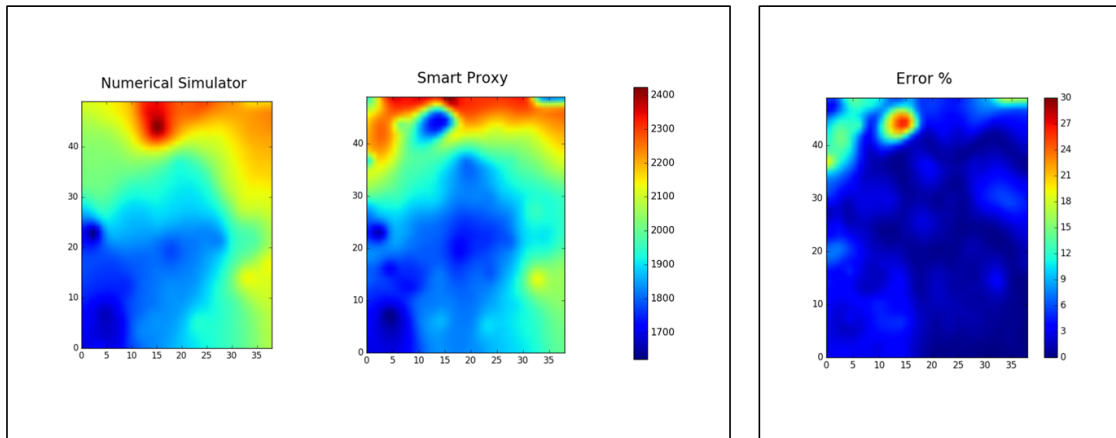## Oil Saturation– Layer 4 on 01/01/1978
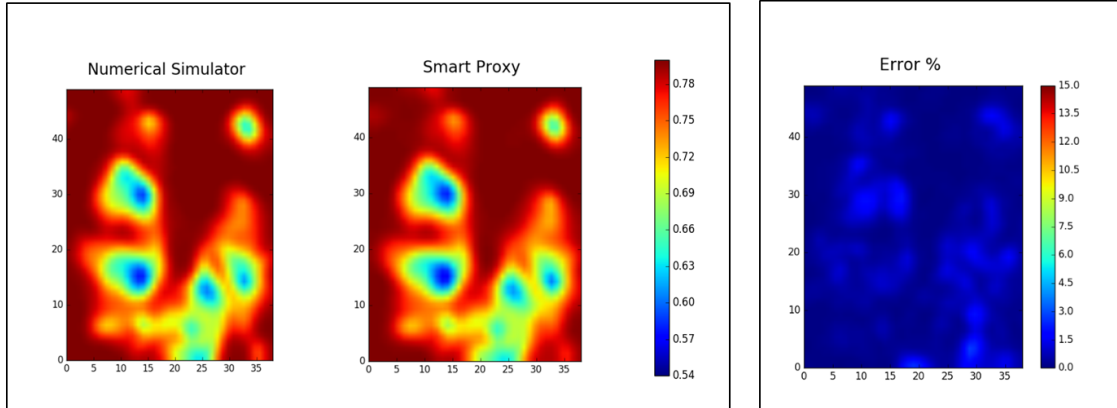
Figure 7.58: Phase three Layer-4 in 1978

155

Reservoir Pressure – Layer 6 on 01/01/1978
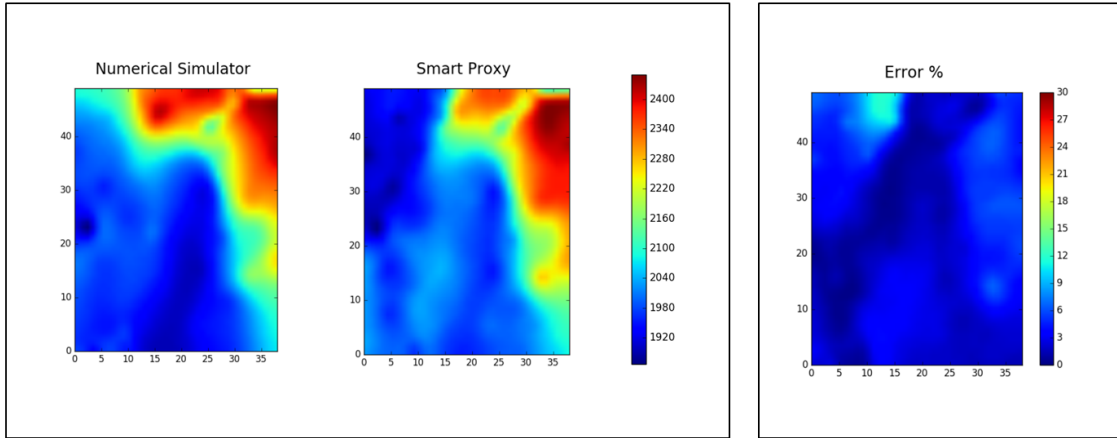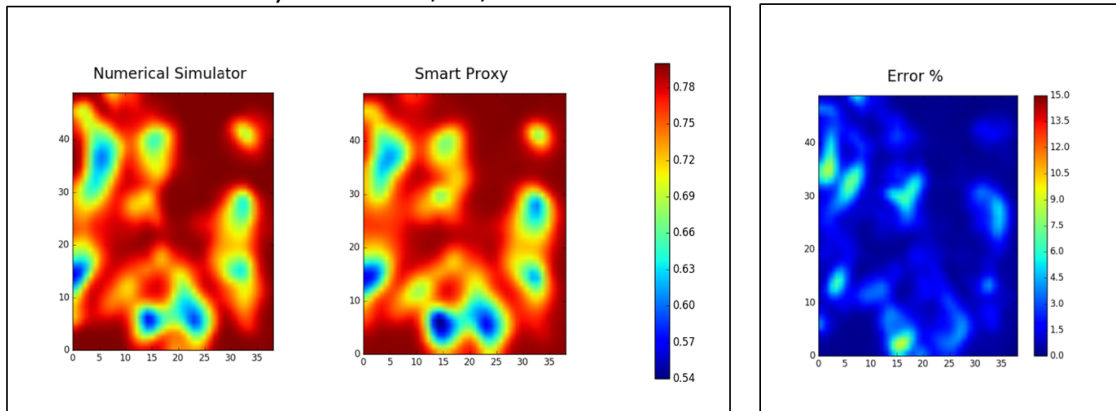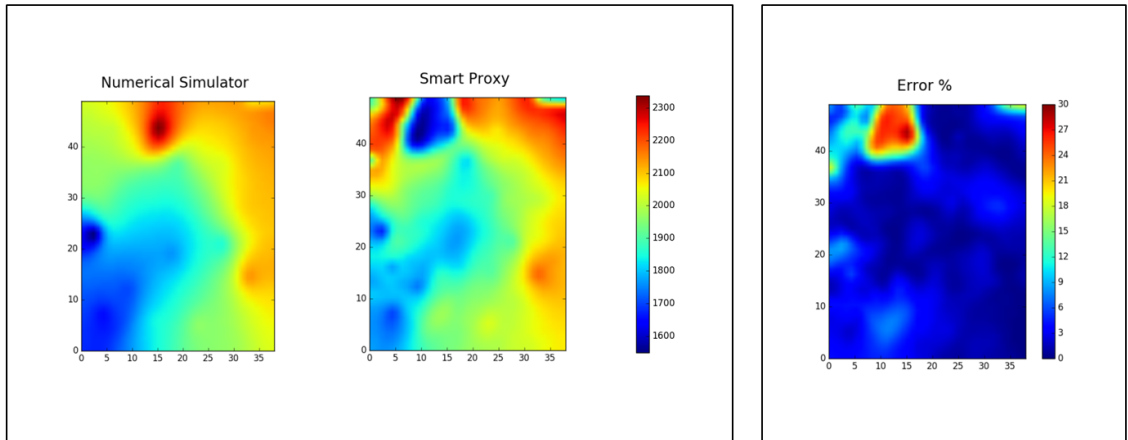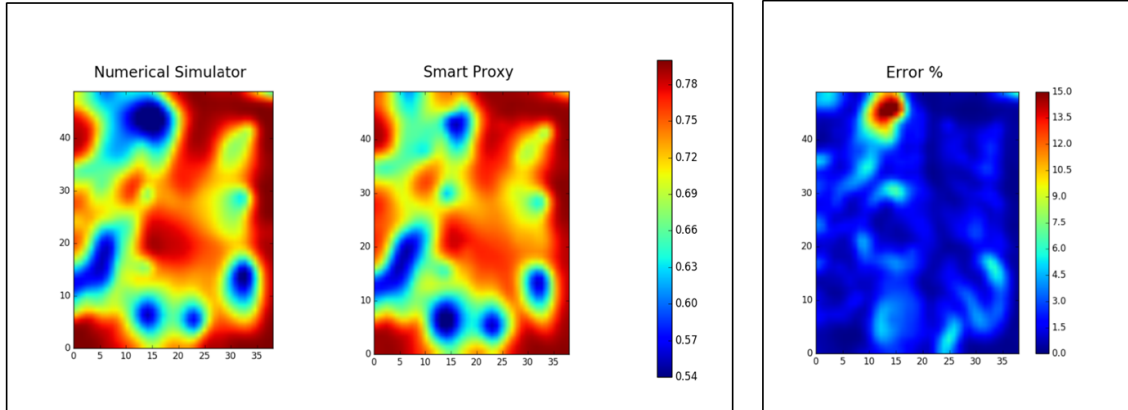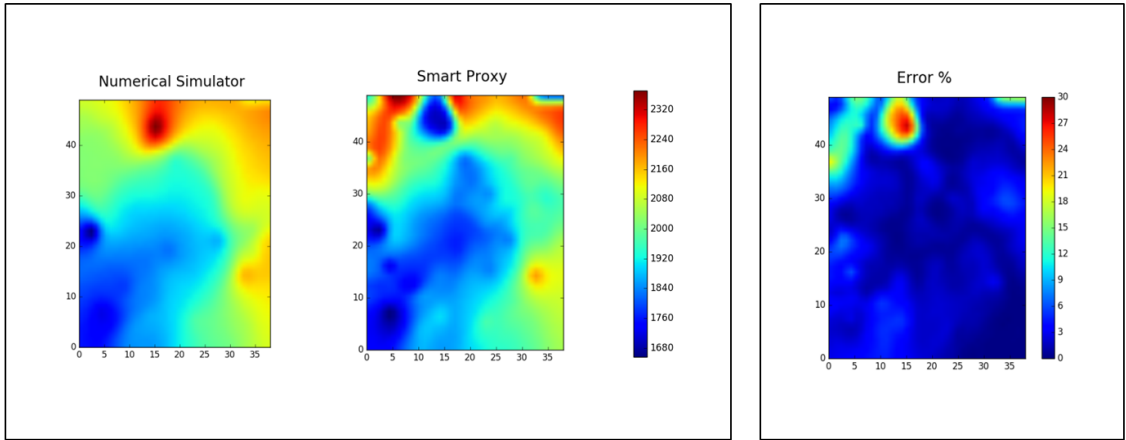

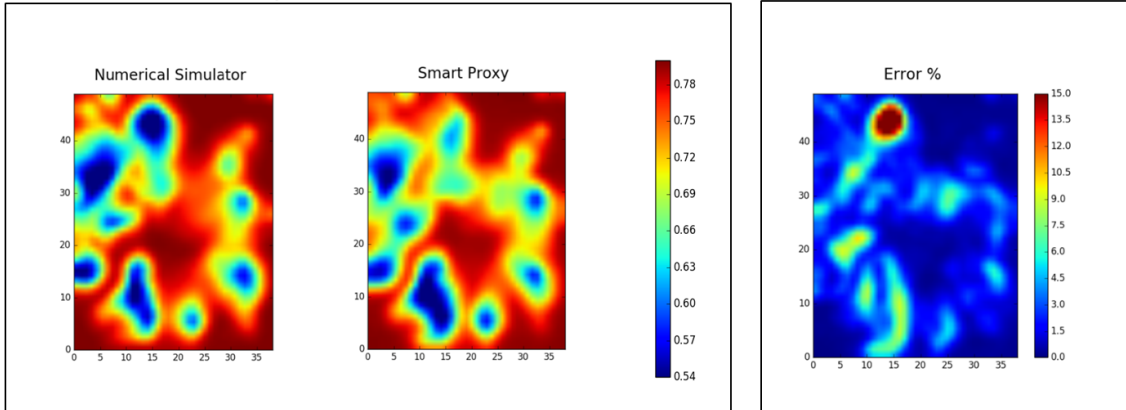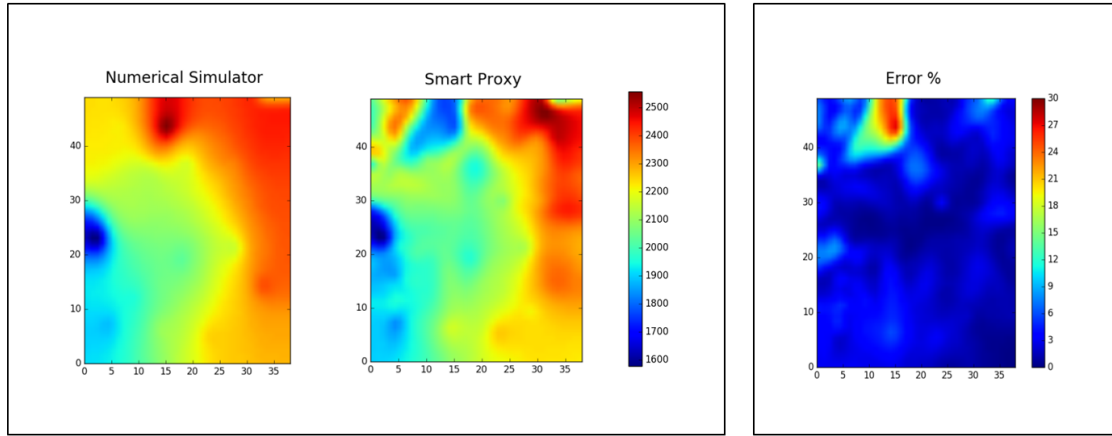# Oil Saturation– Layer 6 on 01/01/1978
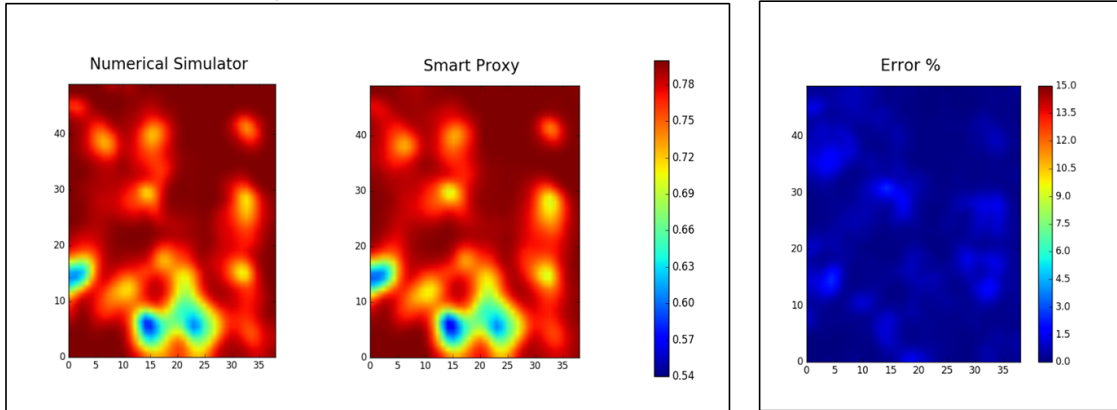
Figure 7.59: Phase three Layer-6 in 1978

Reservoir Pressure – Layer 8 on 01/01/1978
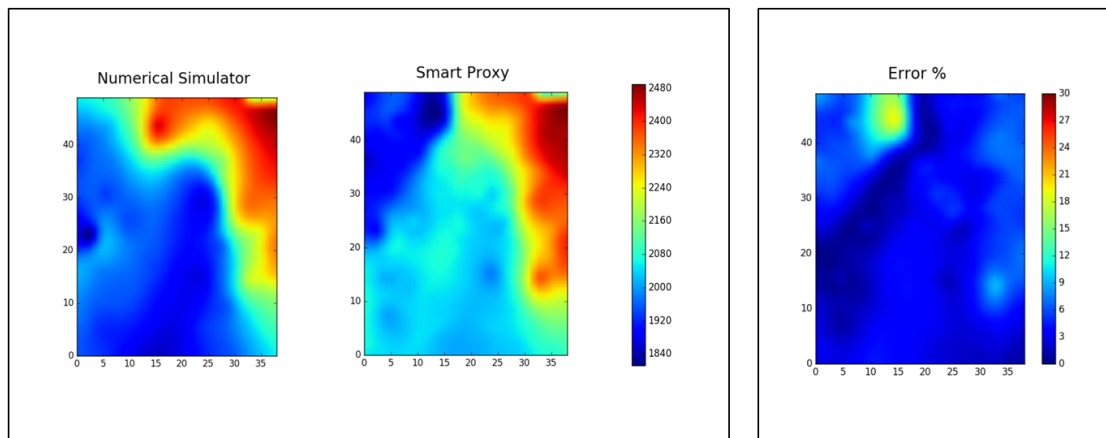
Oil Saturation– Layer 8 on 01/01/1978

Figure 7.60: Phase three Layer-8 in 1978

Reservoir Pressure – Layer 10 on 01/01/1978

Oil Saturation– Layer 10 on 01/01/1978

Figure 7.61: Phase three Layer-10 in 1978

Reservoir Pressure – Layer 14 on 01/01/1978
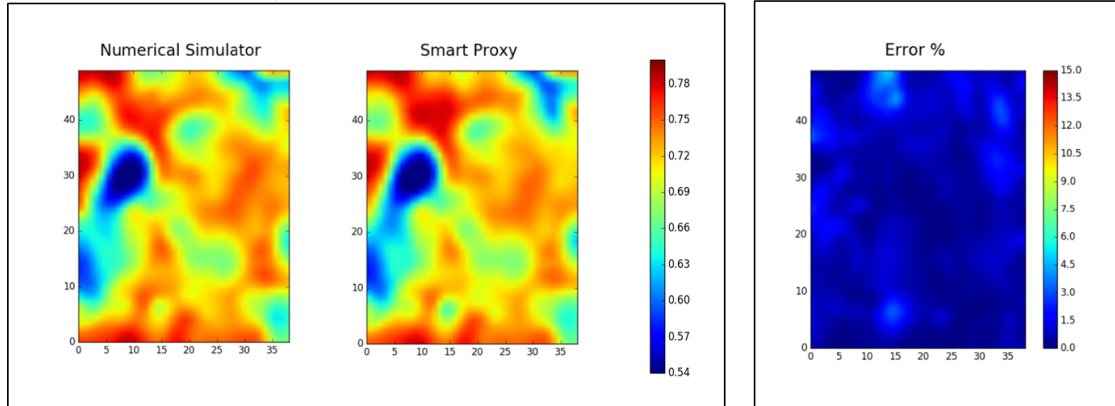

Oil Saturation– Layer 14 on 01/01/1978

Figure 7.62: Phase three Layer-14 in 1978

159

Reservoir Pressure – Layer 1 on 01/01/1983
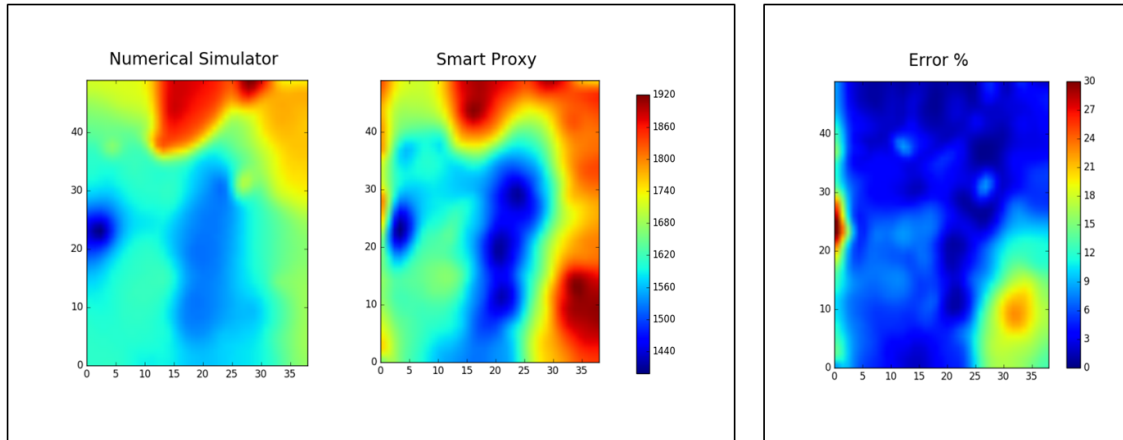

Oil Saturation– Layer 1 on 01/01/1983

Figure 7.63: Phase three Layer-1 in 1983

Reservoir Pressure – Layer 4 on 01/01/1983

Oil Saturation– Layer 4 on 01/01/1983

Figure 7.64: Phase three Layer-4 in 1983

Reservoir Pressure – Layer 6 on 01/01/1983
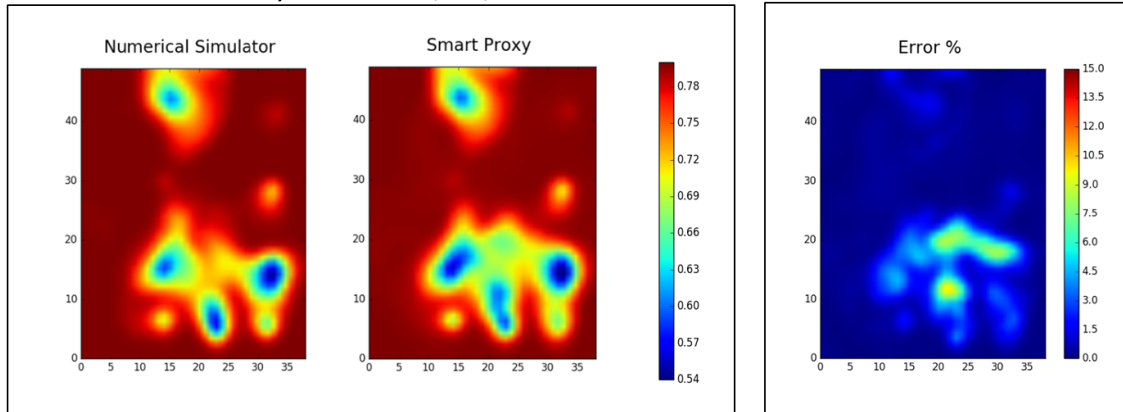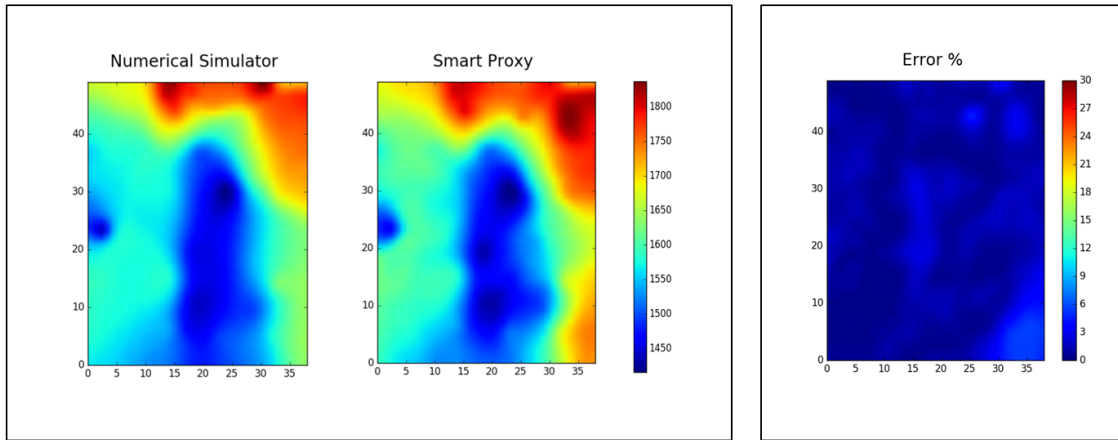

Oil Saturation– Layer 6 on 01/01/1983

Figure 7.65: Phase three Layer-6 in 1983

Reservoir Pressure – Layer 8 on 01/01/1983
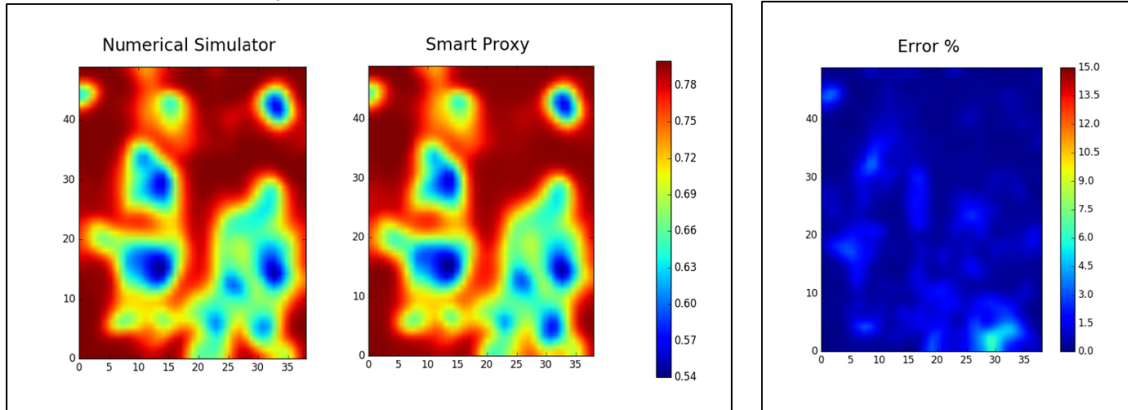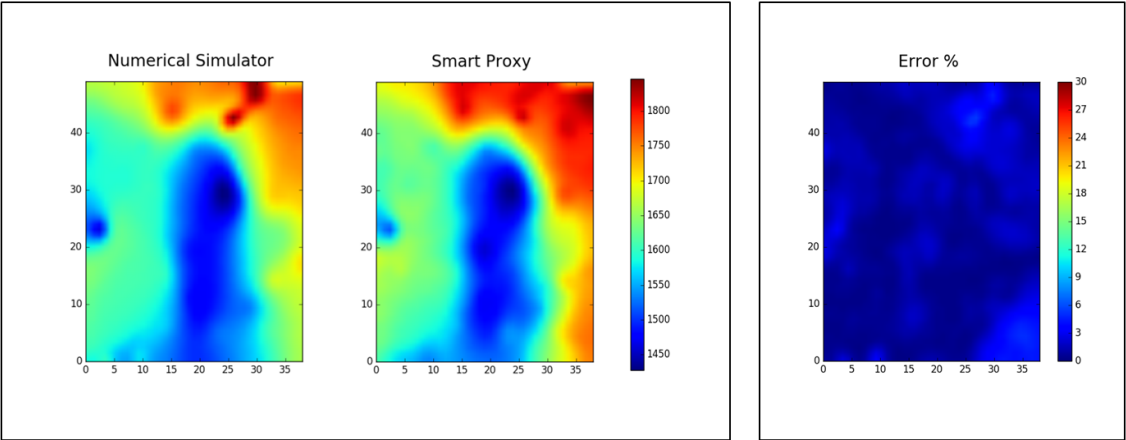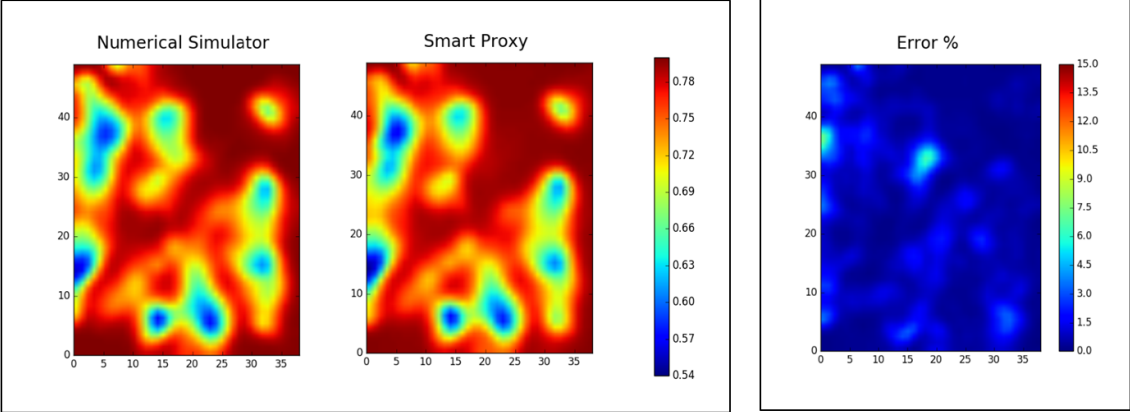

Oil Saturation– Layer 8 on 01/01/1983

Figure 7.66: Phase three Layer-8 in 1983

Reservoir Pressure – Layer 10 on 01/01/1983

Oil Saturation– Layer 10 on 01/01/1983

Figure 7.67: Phase three Layer-10 in 1983

Reservoir Pressure – Layer 14 on 01/01/1983

Oil Saturation– Layer 14 on 01/01/1983

Figure 7.68: Phase three Layer-14 in 1983

Reservoir Pressure – Layer 1 on 01/01/1986

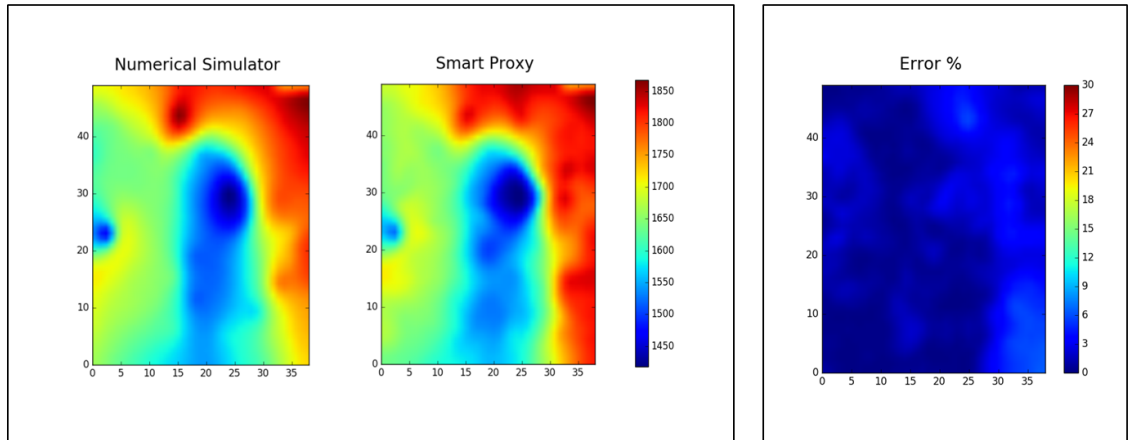## Oil Saturation– Layer 1 on 01/01/1986

Figure 7.69: Phase three Layer-1 in 1986

Reservoir Pressure – Layer 4 on 01/01/1986

Oil Saturation– Layer 4 on 01/01/1986

Figure 7.70: Phase three Layer-4 in 1986

167

Reservoir Pressure – Layer 6 on 01/01/1986

Oil Saturation– Layer 6 on 01/01/1986

Figure 7.71: Phase three Layer-6 in 1986

Reservoir Pressure – Layer 8 on 01/01/1986

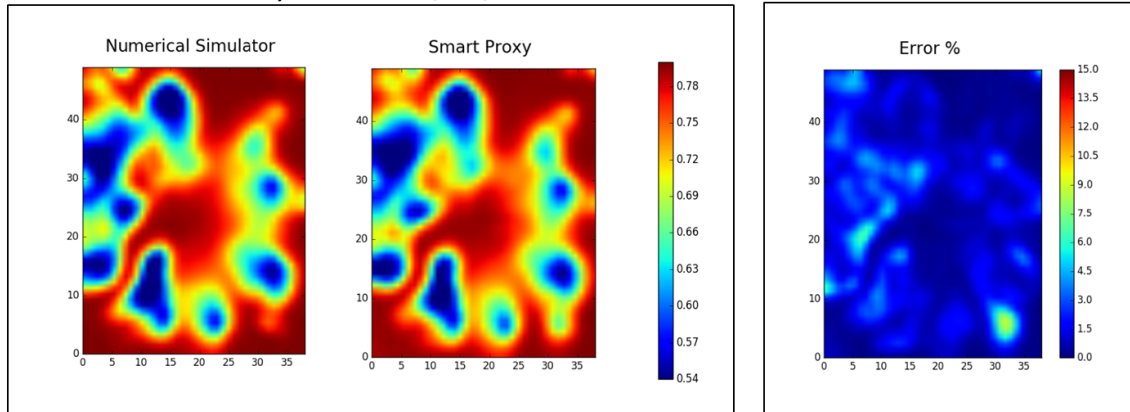
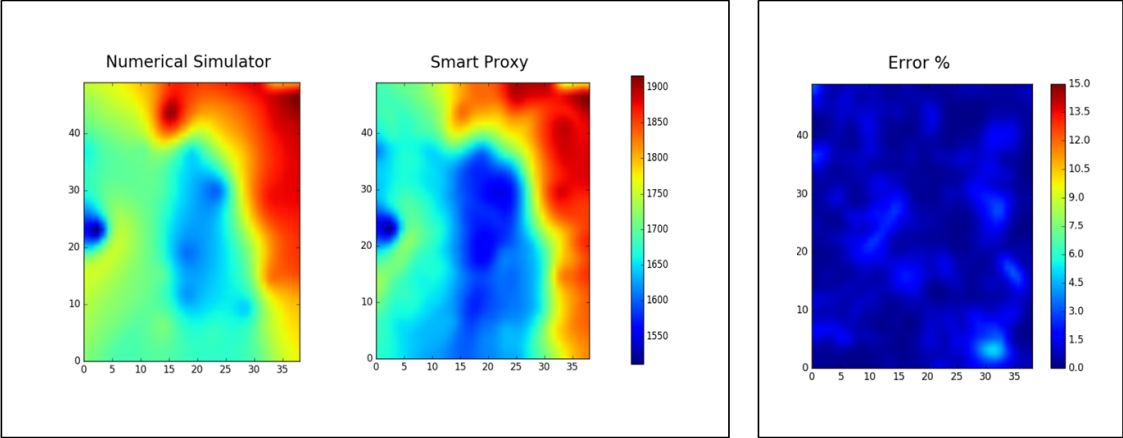## Oil Saturation– Layer 8 on 01/01/1986

Figure 7.72: Phase three Layer-8 in 1986

Reservoir Pressure – Layer 10 on 01/01/1986

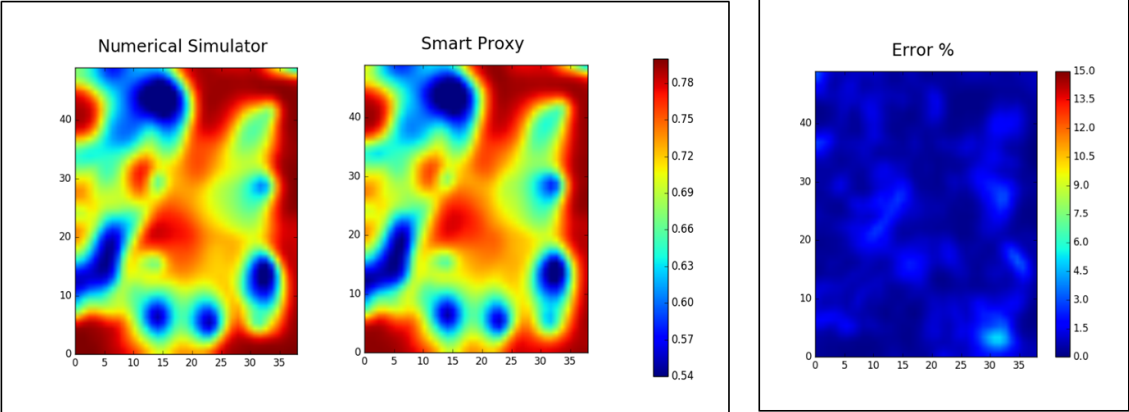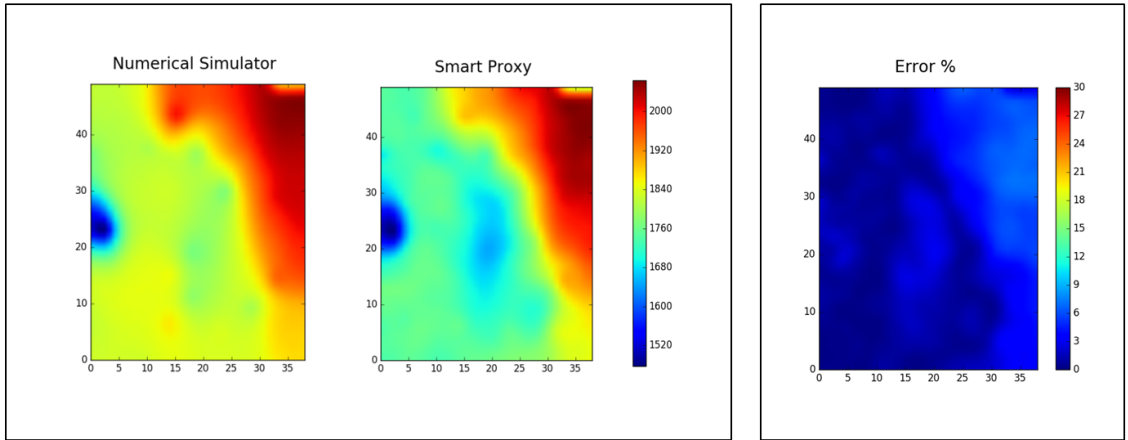# Oil Saturation– Layer 10 on 01/01/1986

Figure 7.73: Phase three Layer-10 in 1986

Reservoir Pressure – Layer 14 on 01/01/1986
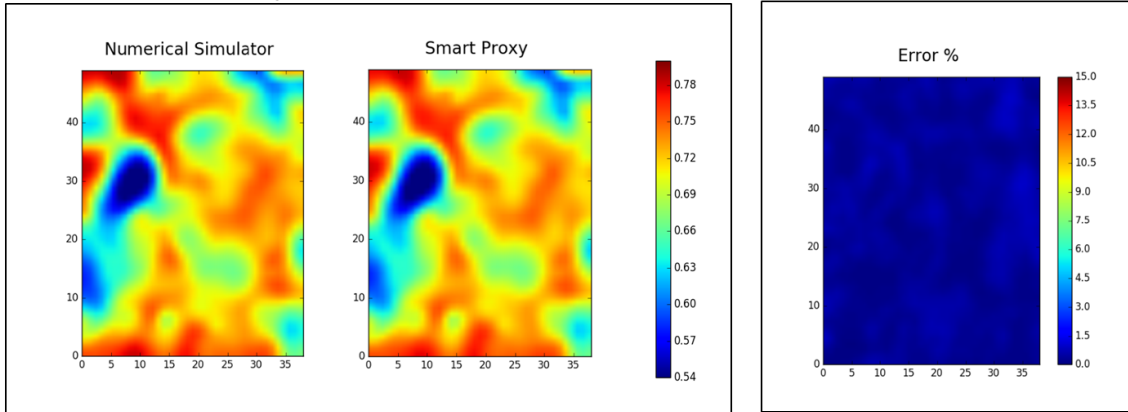

Oil Saturation– Layer 14 on 01/01/1986

Figure 7.74: Phase three Layer-14 in 1986

Reservoir Pressure – Layer 1 on 01/01/1989

Oil Saturation– Layer 1 on 01/01/1989

Figure 7.75: Phase three Layer-1 in 1989

Reservoir Pressure – Layer 4 on 01/01/1989

Oil Saturation– Layer 4 on 01/01/1989

Figure 7.76: Phase three Layer-4 in 1989

Reservoir Pressure – Layer 6 on 01/01/1989

Oil Saturation– Layer 6 on 01/01/1989

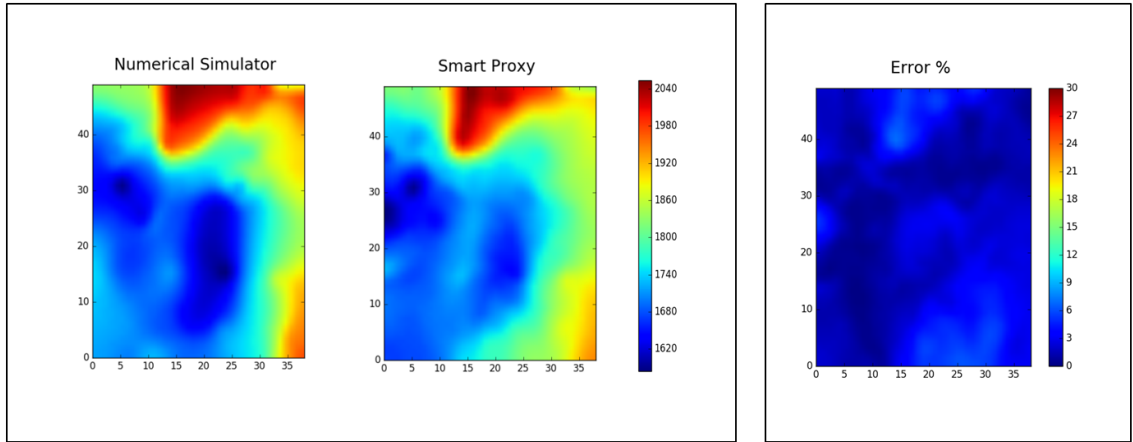Figure 7.77: Phase three Layer-6 in 1989

Figure 7.78: Phase three Layer-8 in 1989

Reservoir Pressure – Layer 10 on 01/01/1989

Oil Saturation– Layer 10 on 01/01/1989

Figure 7.79: Phase three Layer-10 in 1989

Reservoir Pressure – Layer 14 on 01/01/1989

Oil Saturation– Layer 14 on 01/01/1989

Figure 7.80: Phase three Layer-14 in 1989

Reservoir Pressure – Layer 1 on 01/01/1991
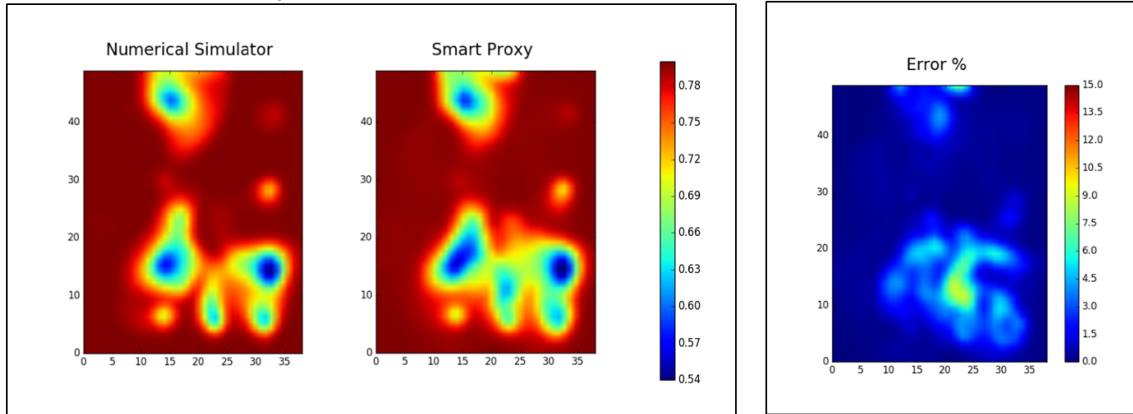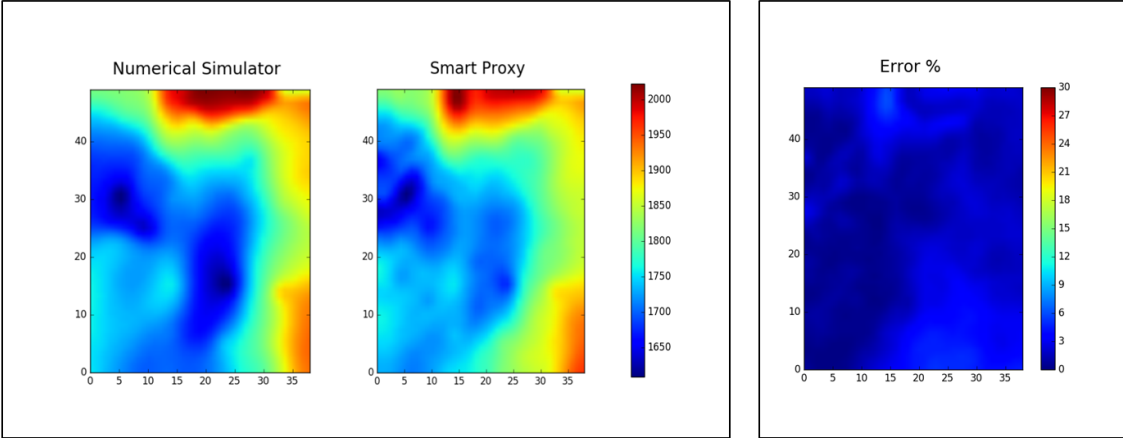

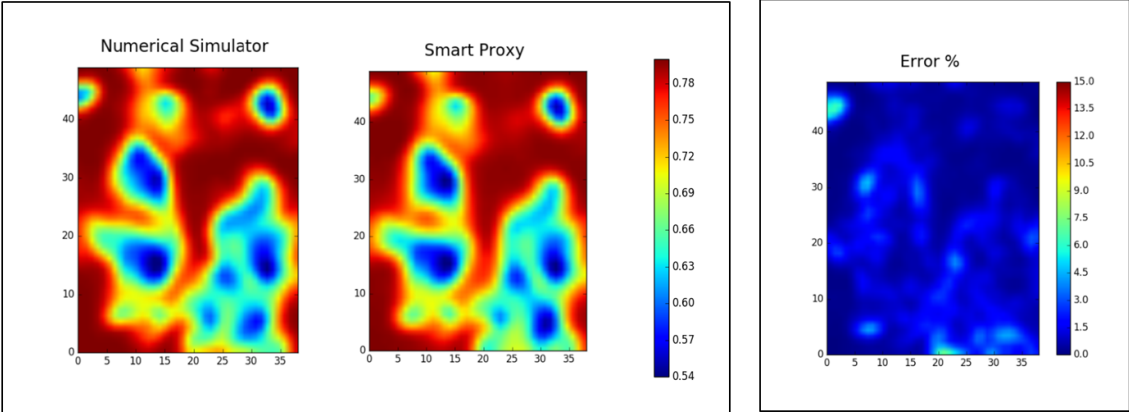Oil Saturation– Layer 1 on 01/01/1991

Figure 7.81: Phase three Layer-1 in 1991

Reservoir Pressure – Layer 4 on 01/01/1991

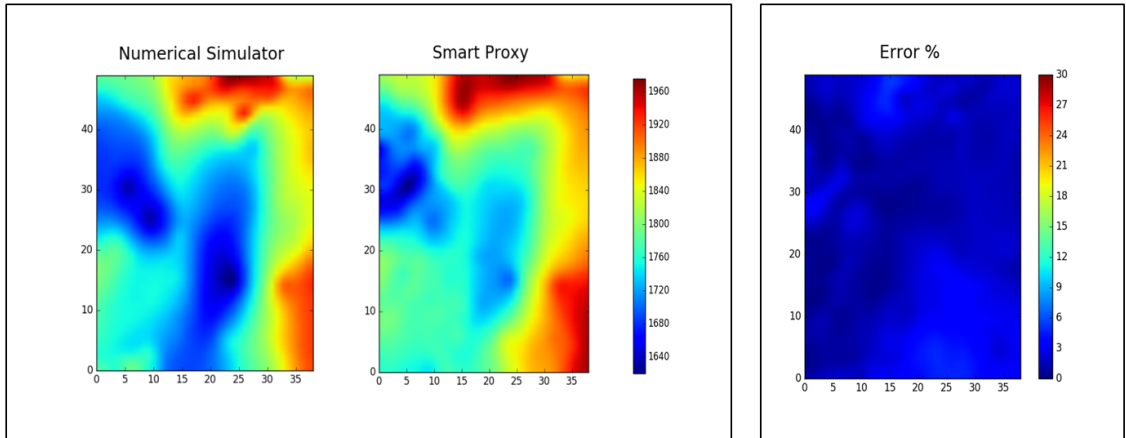# Oil Saturation– Layer 4 on 01/01/1991

Figure 7.82: Phase three Layer-4 in 1991

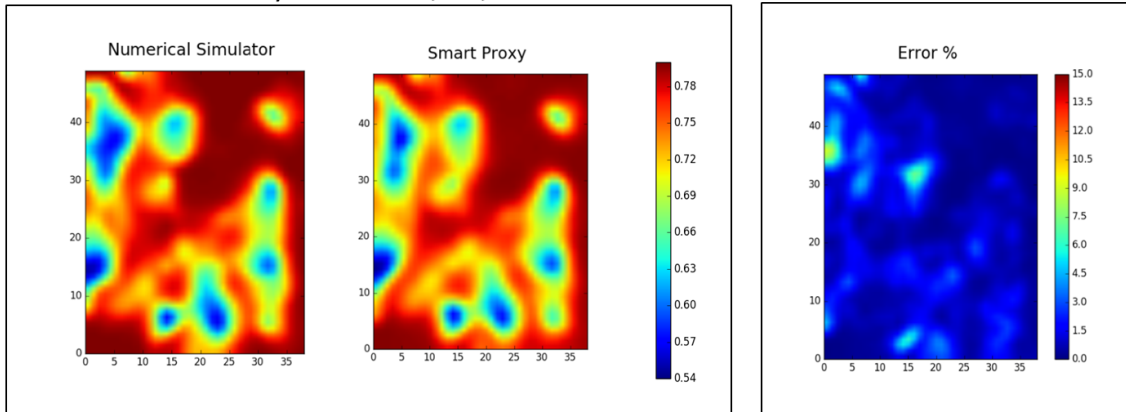Reservoir Pressure – Layer 6 on 01/01/1991

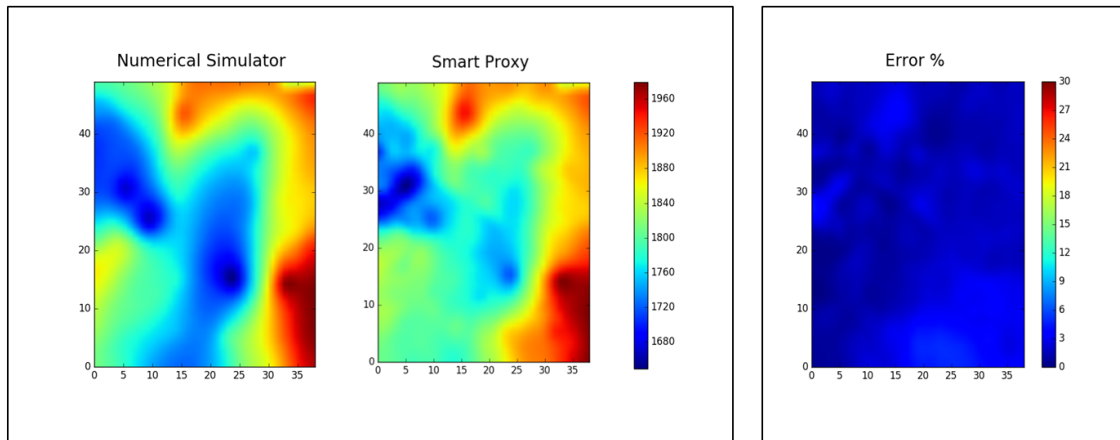Oil Saturation– Layer 6 on 01/01/1991

Figure 7.83: Phase three Layer-6 in 1991

Reservoir Pressure – Layer 8 on 01/01/1991

Oil Saturation– Layer 8 on 01/01/1991

Figure 7.84: Phase three Layer-8 in 1991

Reservoir Pressure – Layer 10 on 01/01/1991

Oil Saturation– Layer 10 on 01/01/1991

Figure 7.85: Phase three Layer-10 in 1991

Reservoir Pressure – Layer 14 on 01/01/1991

Oil Saturation– Layer 14 on 01/01/1991
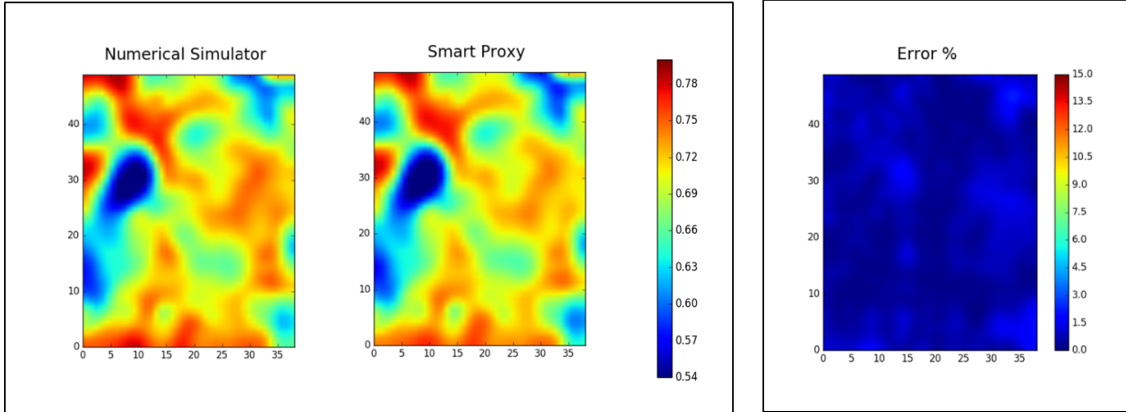
Figure 7.86: Phase three Layer-14 in 1991
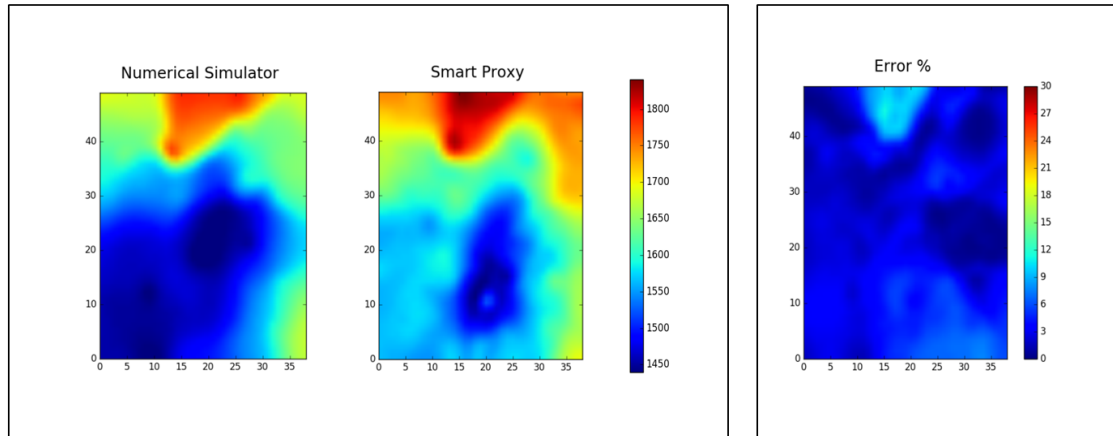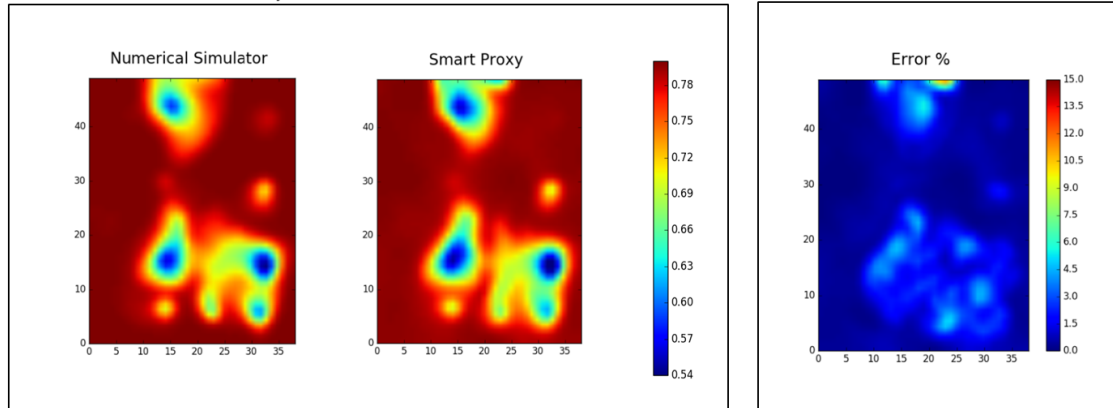
Figure 7.87: Phase three Layer-1 in 1993

Reservoir Pressure – Layer 4 on 01/01/1993
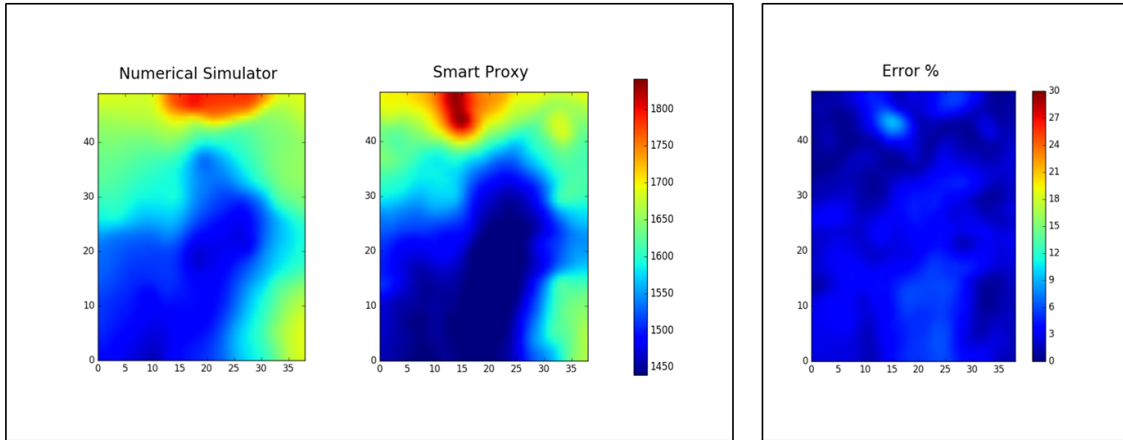
Oil Saturation– Layer 4 on 01/01/1993
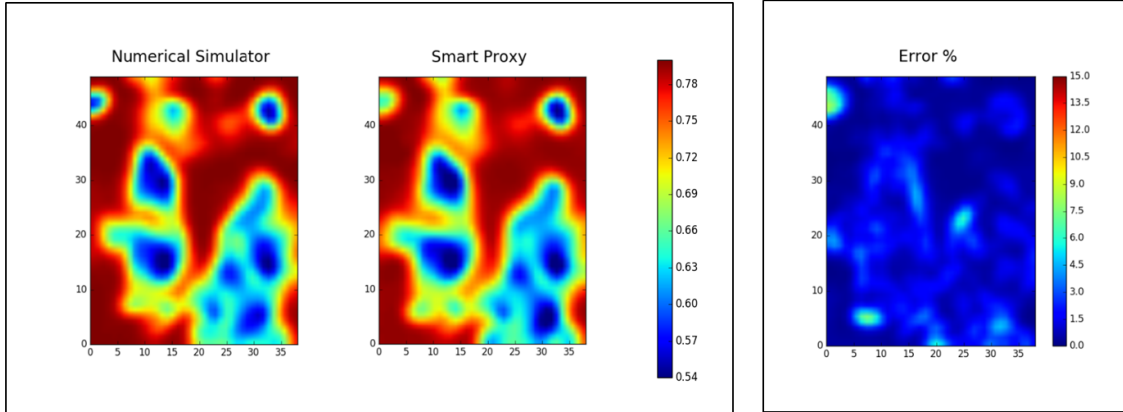
Figure 7.88: Phase three Layer-4 in 1993

Reservoir Pressure – Layer 6 on 01/01/1993
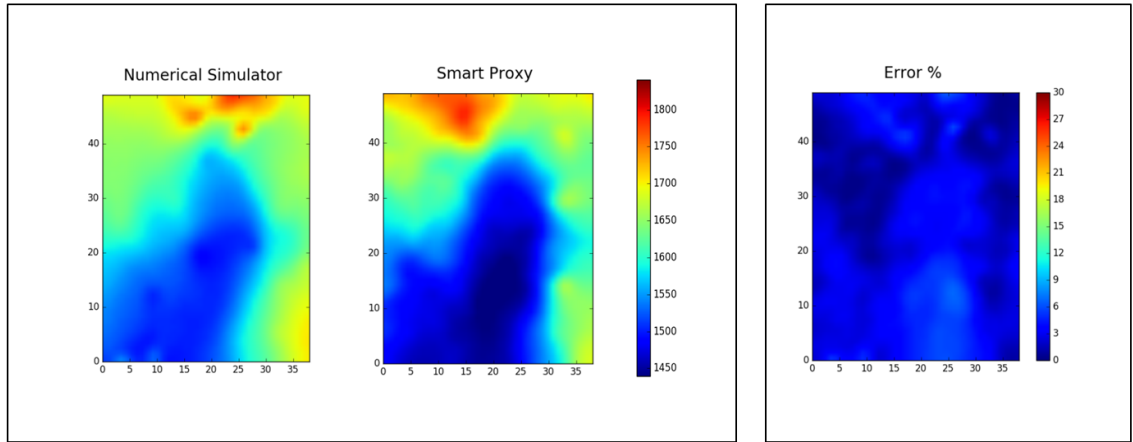
Oil Saturation– Layer 6 on 01/01/1993
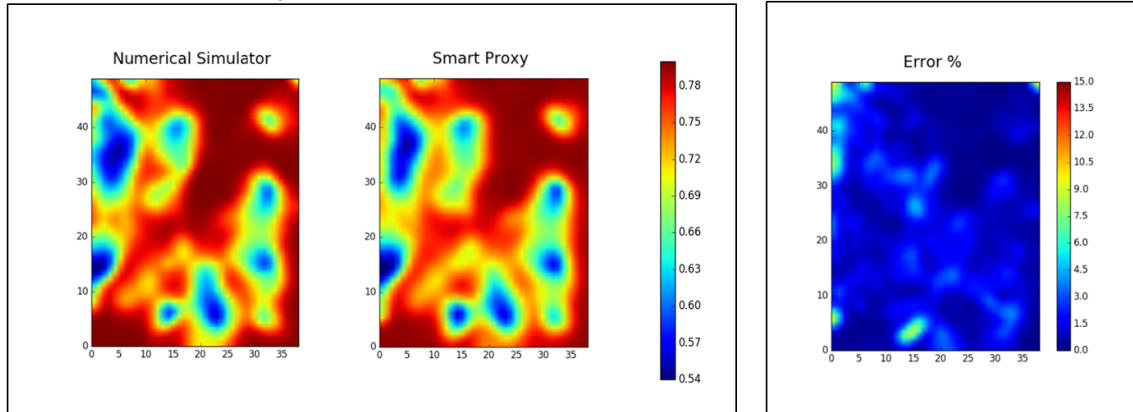
Figure 7.89: Phase three Layer-6 in 1993

186

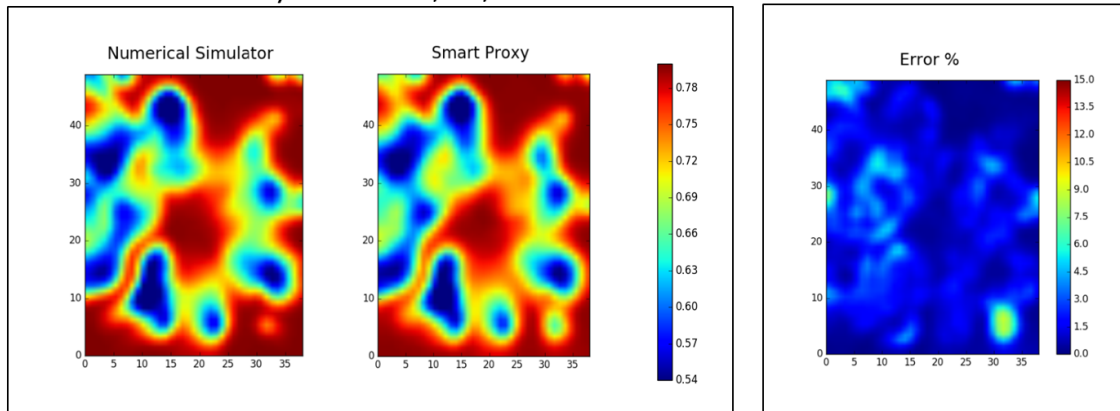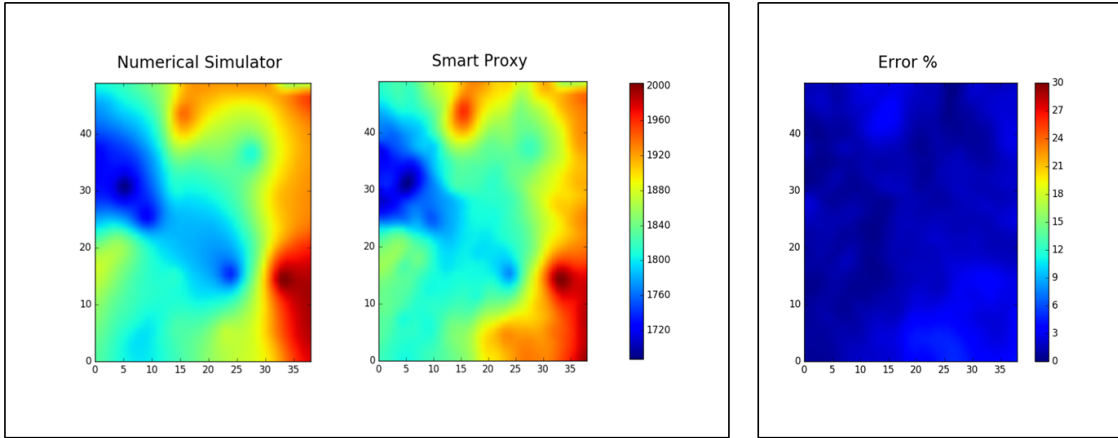Reservoir Pressure – Layer 8 on 01/01/1993
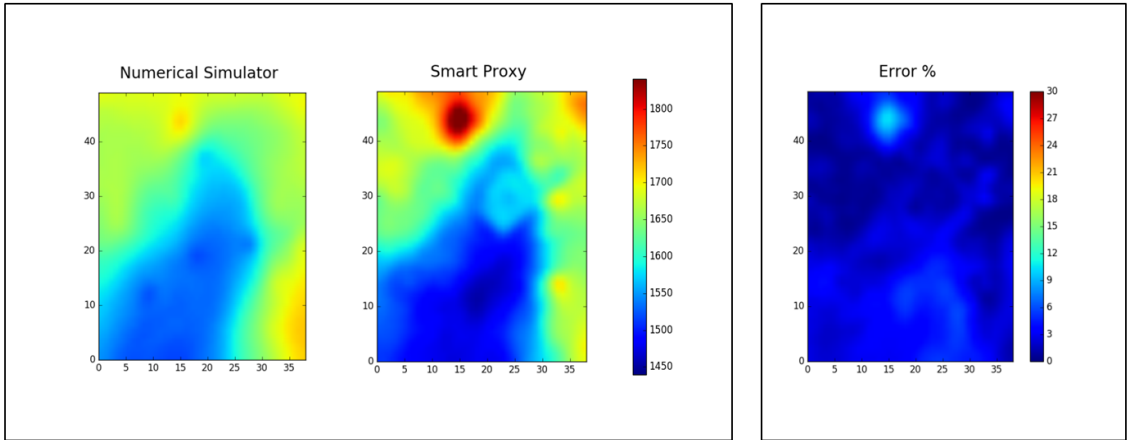
Oil Saturation– Layer 8 on 01/01/1993
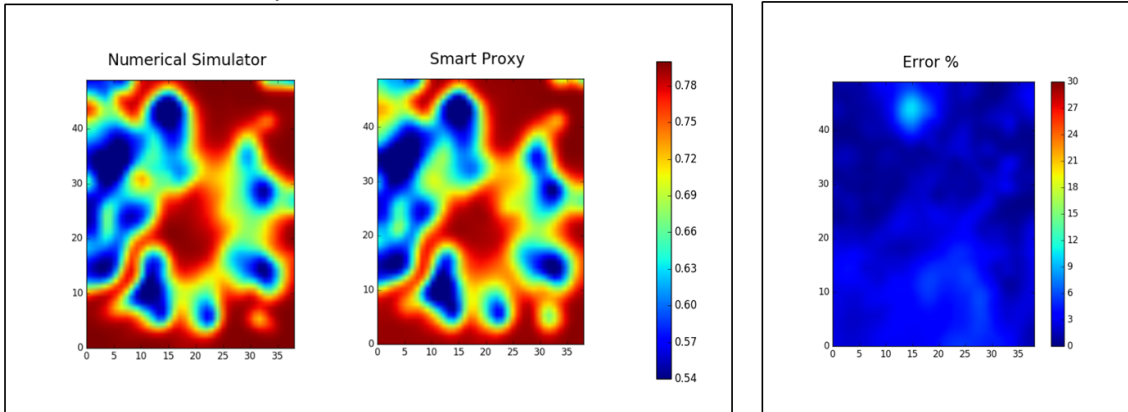
Figure 7.90: Phase three Layer-8 in 1993

Reservoir Pressure – Layer 10 on 01/01/1993
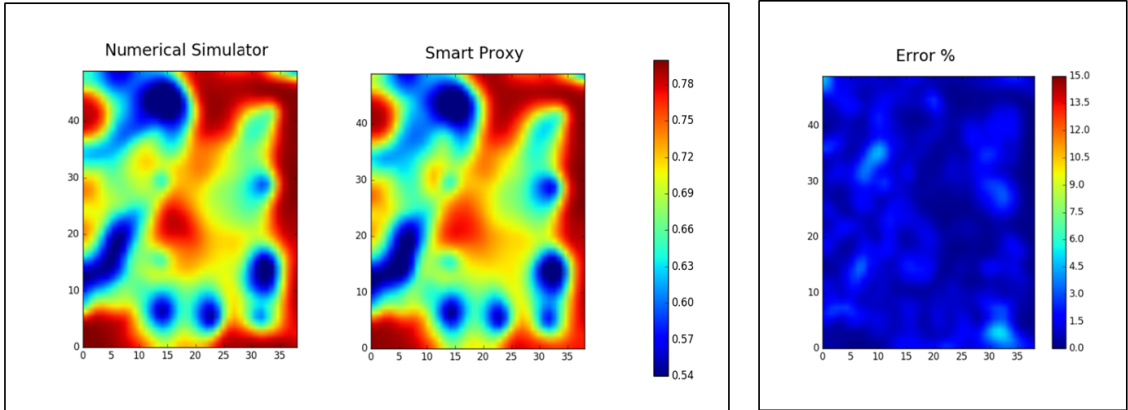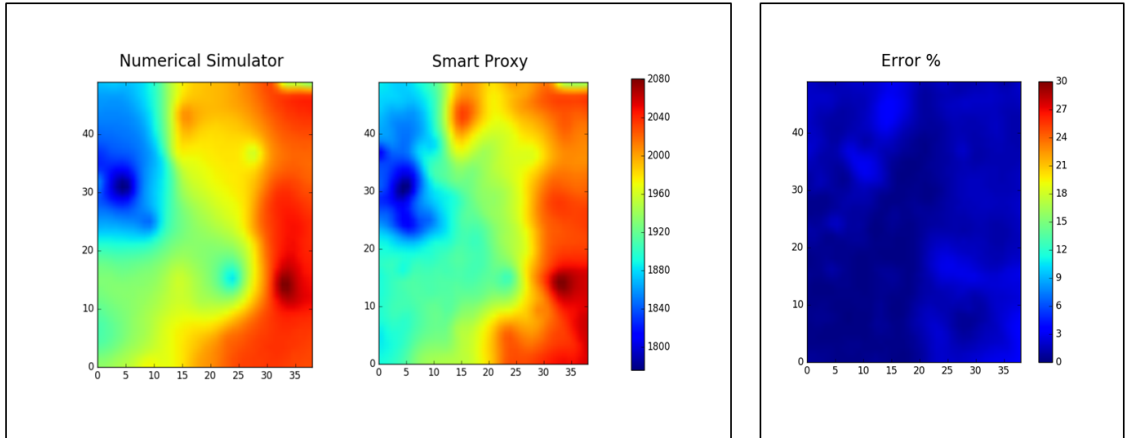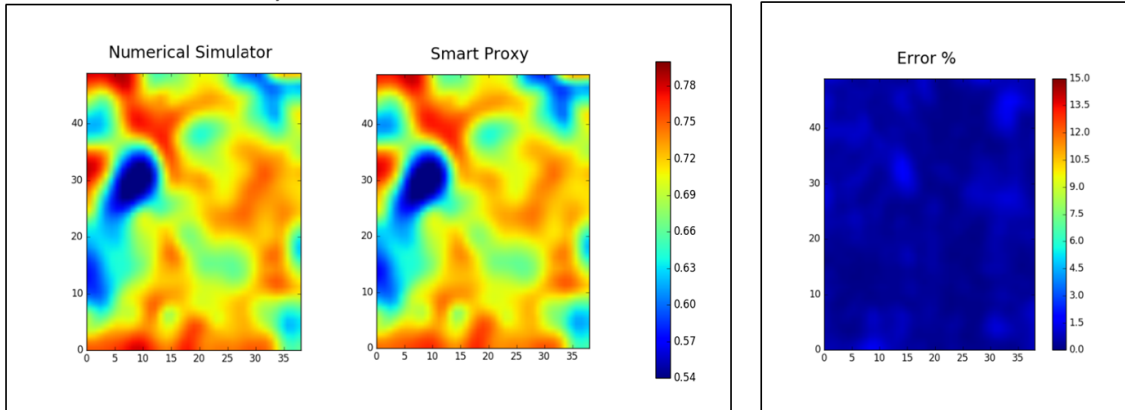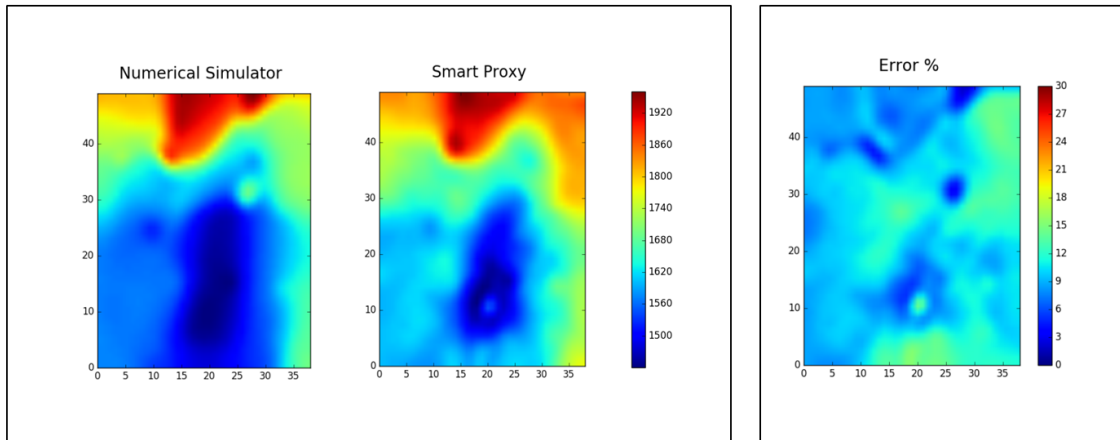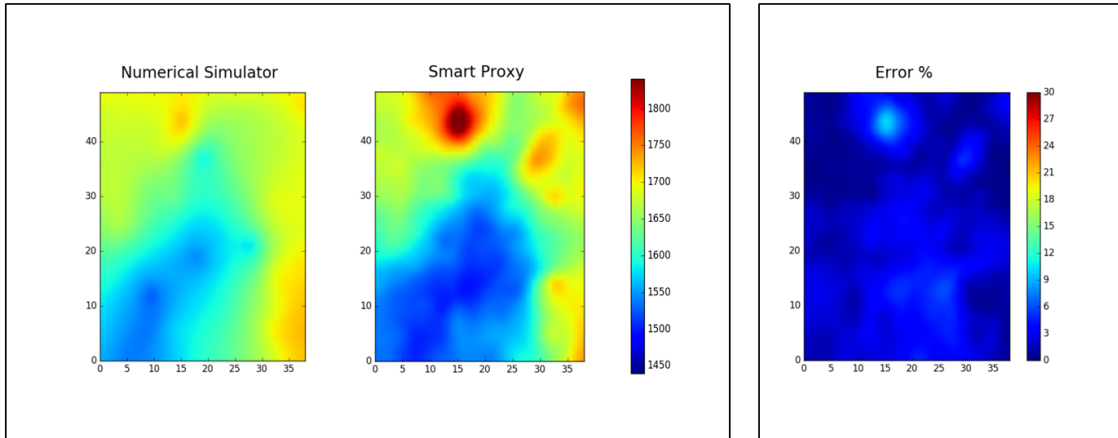

Oil Saturation– Layer 10 on 01/01/1993
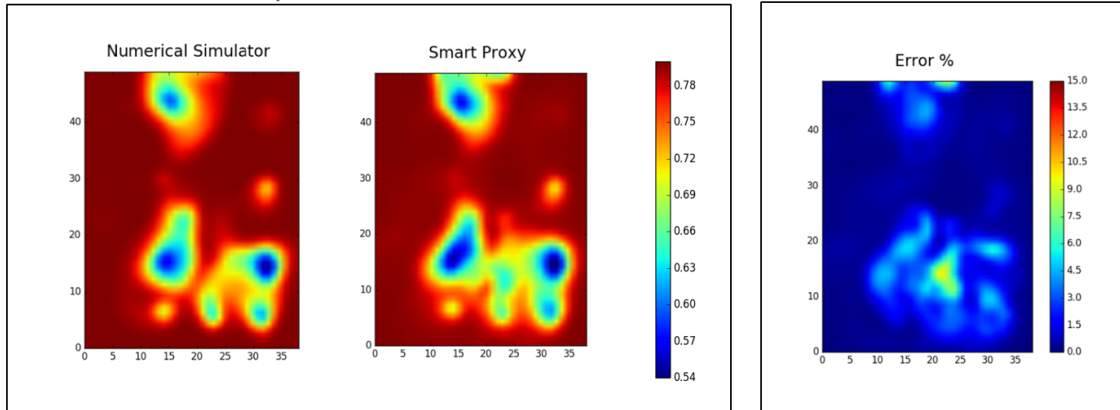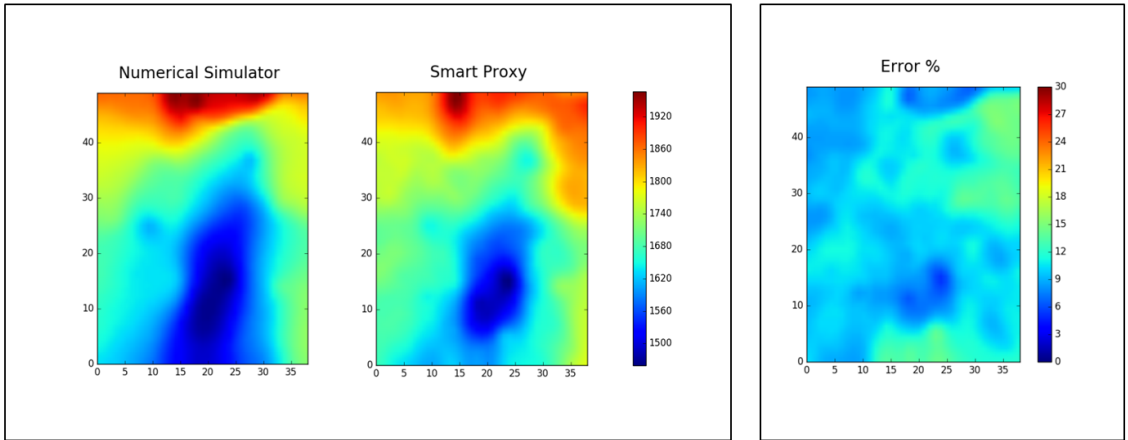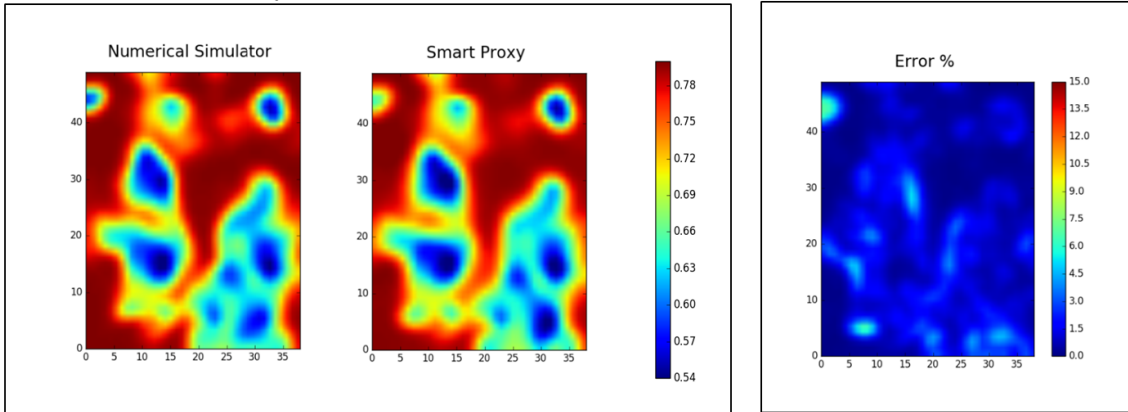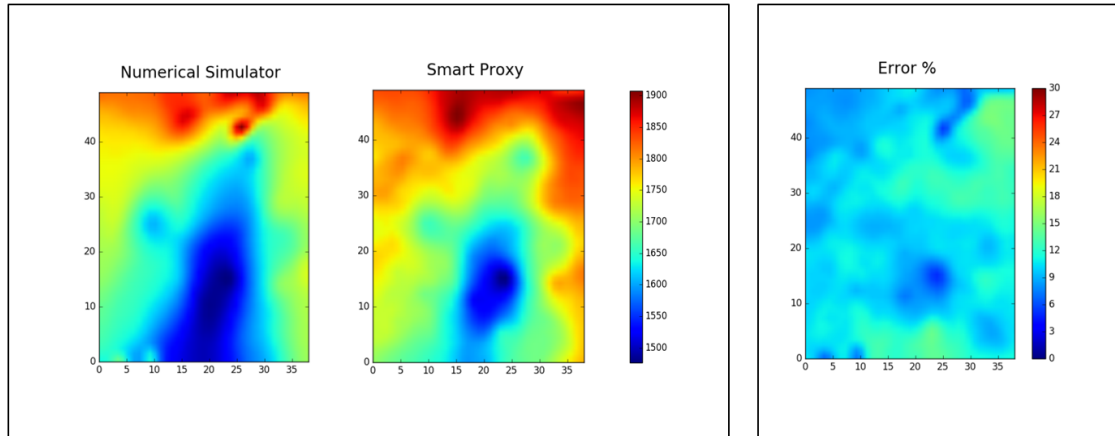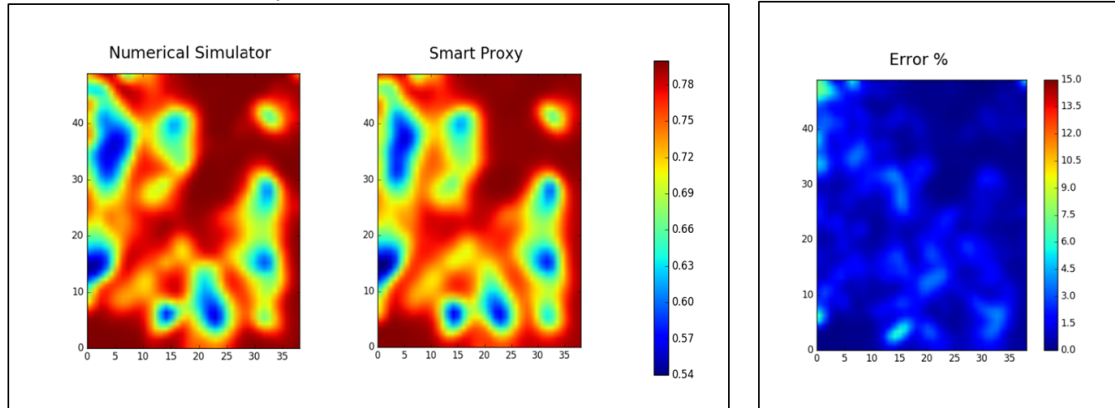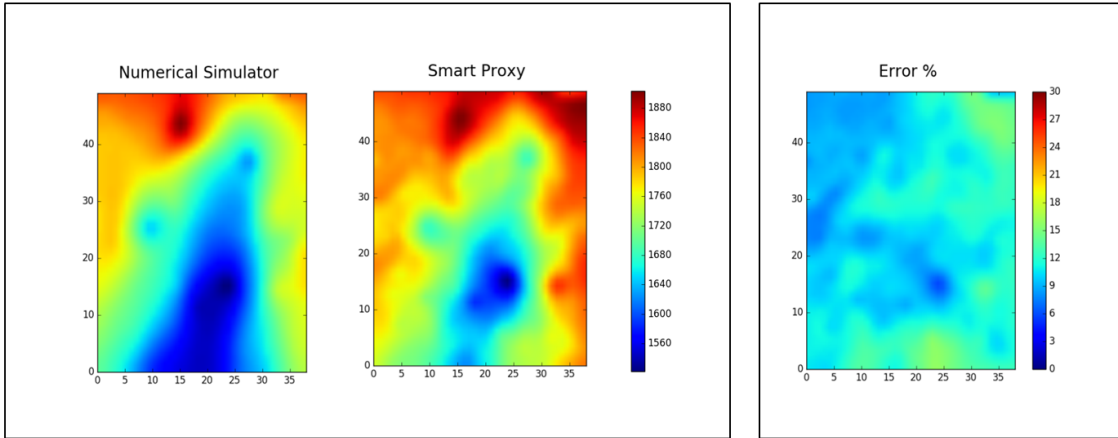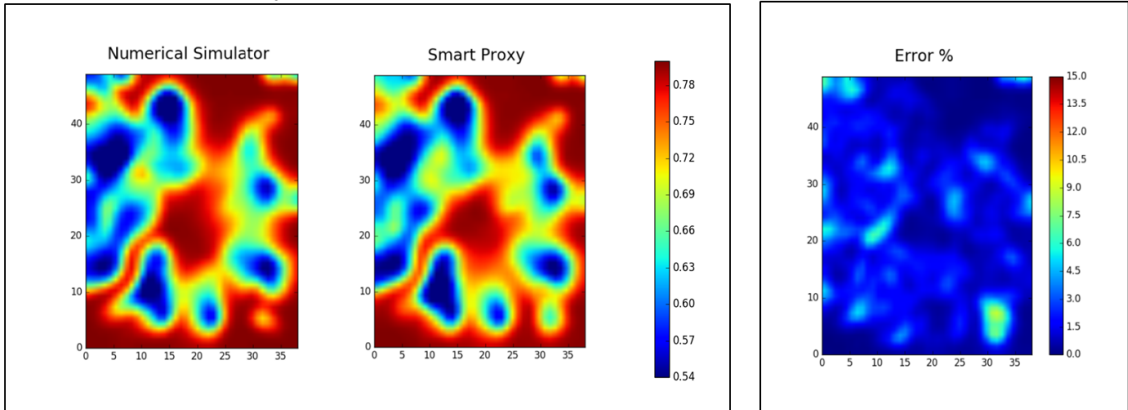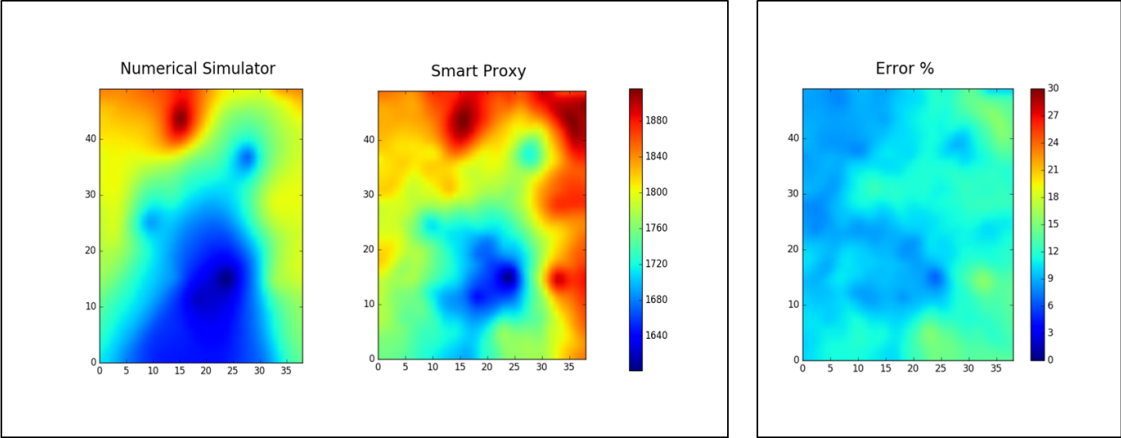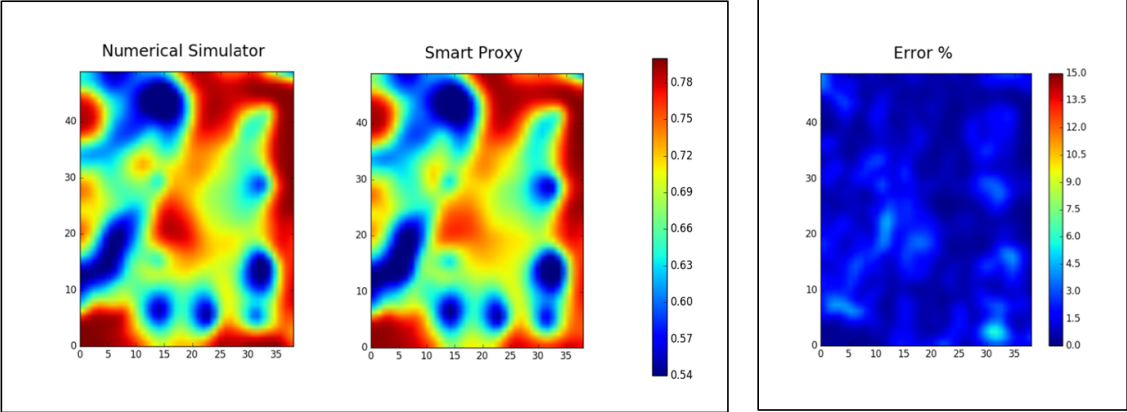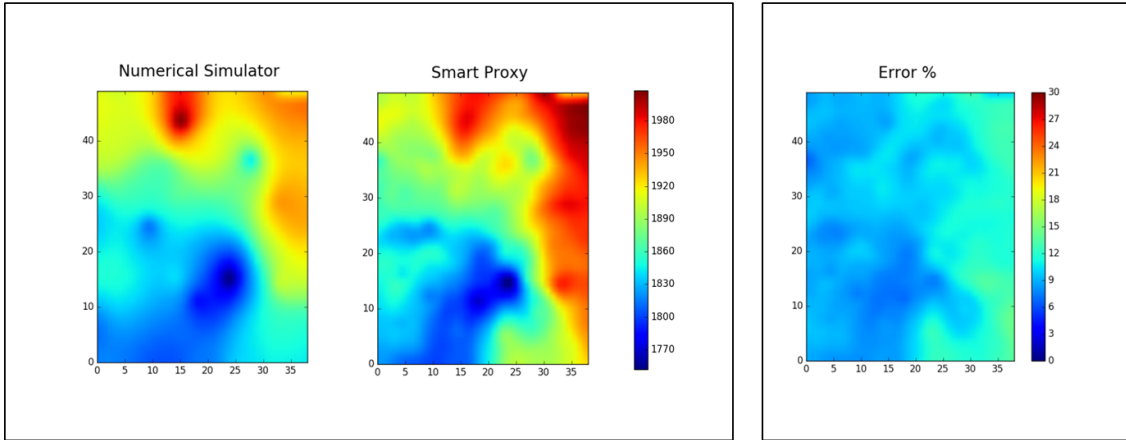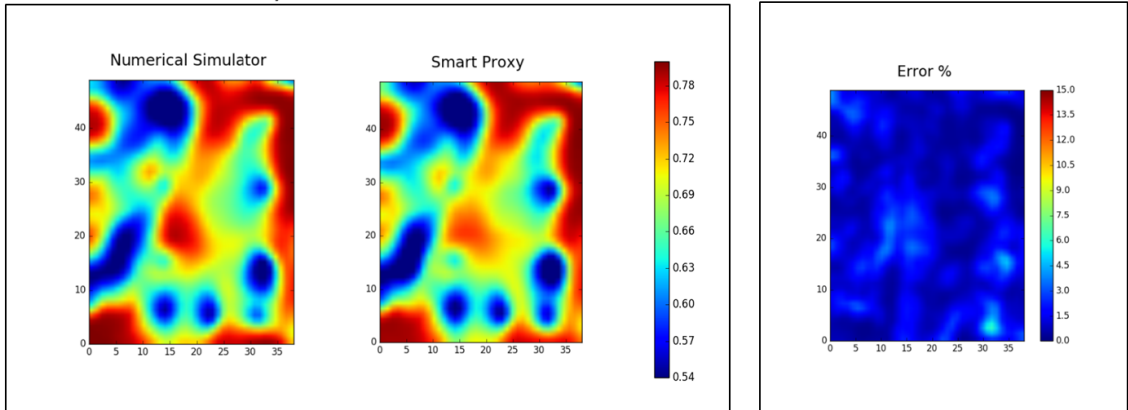

Figure 7.91: Phase three Layer-10 in 1993

Reservoir Pressure – Layer 14 on 01/01/1993
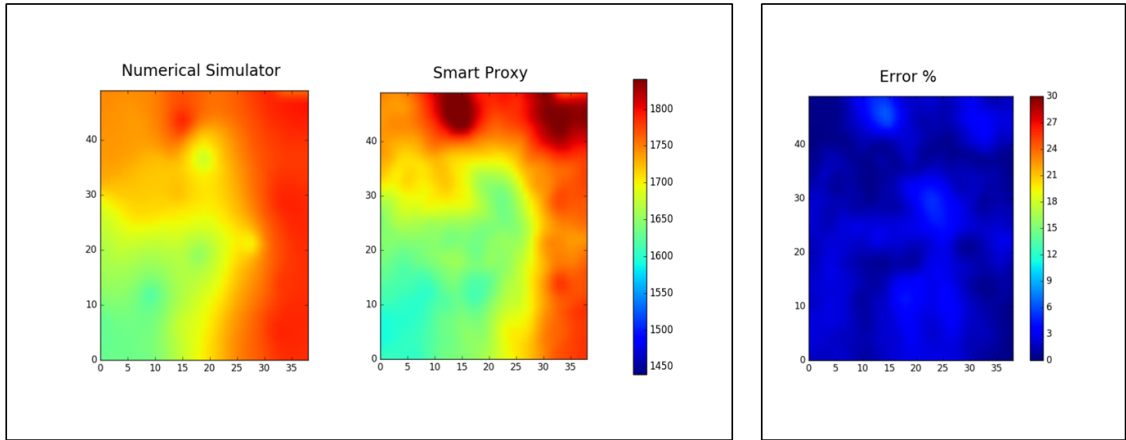
Oil Saturation– Layer 14 on 01/01/1993
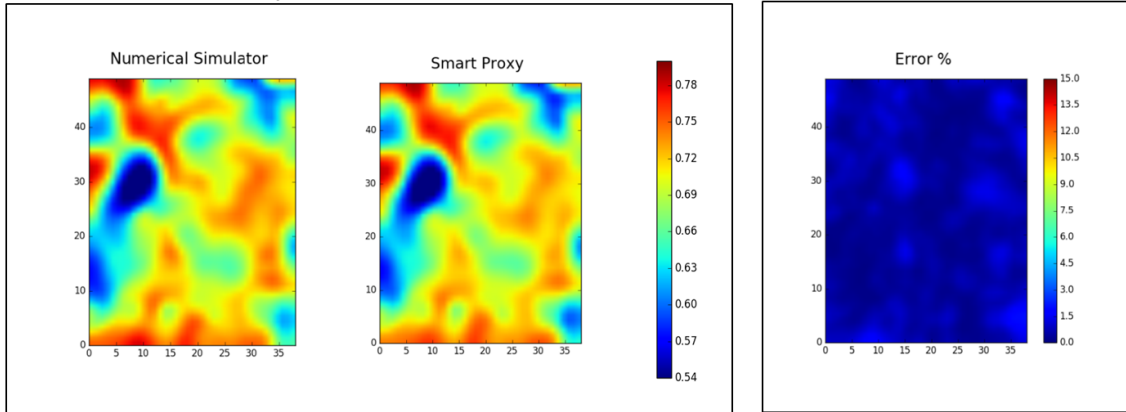
Figure 7.92: Phase three Layer-14 in 1993

Chapter 8

Smart Proxy for Field Production

In the previous Chapters, smart proxy models were developed for dynamic reservoir properties (pressure and saturation) at grid-block level. Since the main objective of this study was to build a smart proxy for field geological and operational realization, it was important to design a smart proxy that is able to predict the production profile of the field.

The objective of the production profile smart proxy model is to imitate the response of the production wells in the form of production rates to different geological and operational realizations. This type of smart proxy model and dynamic reservoir properties smart proxy models are great tools for decision making about field development plans.

To build a smart proxy for field production profile, the general workflow for a dynamic reservoir properties smart proxy model in a cascading process was followed. The field performance history from 1950 to 1995 was divided into three phases and for each phase, smart proxy models were developed. For each phase, three smart proxy models were generated for three production profiles; oil rate, gas rate, and water rate. In the following section, the production profile smart proxy development is explained.

8.1   Numerical Simulation Design

To make the smart proxy model more practical in reservoir simulation, it is essential that it is capable of producing the reservoir properties as well as the production profiles (production rates). Thus, the dynamic reservoir smart proxy model (grid based) and the production profile (well based)

should be run under the same geological and operational realizations. To achieve this, the designed numerical simulation runs for the production profile smart proxy model were the same numerical simulation scenarios used for the reservoir properties smart proxy models training and validation at each production phase.

8.2   Spatiotemporal Database Generation

Spatiotemporal database generation is an essential step in developing the smart proxy model. It has the reservoir physics and phenomenon in form of data. The required parameters are imported and organized from the designed numerical reservoir simulation runs to generate the database. The selection of the input parameters is based on the impact of these parameters on the reservoir performance (production profile in this case). Also, the domain knowledge (reservoir engineering) in this study plays a significant rule in the selection of the input parameters in order to deliver the desired results out of the smart proxy model.

The parameters selected for production profile smart proxy were directly impact the production profile performance. Geological parameters such as reservoir thickness, grid top, porosity, and permeability are selected for the database. Reservoir thickness and porosity represent the pore volume of the reservoir and how much fluids can be produced. The permeability affects the ability of the fluids to flow from the reservoir. In addition to that, the location of producers is defined by the grid block location (I, J, K) in the geological model. In addition to the geological parameters, which do not change with time, wells production rates were added to the database at each previous time step to teach the network efficiently about the production profiles.

8.3   Data Partitioning

Different data partition techniques were examined for this case. The software used to develop the production profile smart proxy model is IDEA. In this software, a novel technique for data portioning is available. This technique is called intelligent data portioning. It ensures that data distribution is honored in all three categories, training, calibration, and validation.

Since the number of records in the database for the well base model (production profile) is much smaller than the grid base model, 80% of the data was assigned for the training set and 20% for

Static Data	Dynamic Data
I	Cumulative Oil
J	Cumulative Gas
K	Cumulative Water
Thickness	GOR
Top	Oil Rate
Permeability	Gas Rate
Porosity	Water Rate
Well Index	Pressure

Figure 8.1: Smart Proxy for Production Profile Input Parameters

calibration and validation sets (10% for each).

8.4   Neural Network Construction and Training

Once the spatiotemporal database had been generated and the data portioning had been selected, neural network construction and training could be implemented. Three smart proxy models for each production phase were developed for the production profiles, oil production, gas production, and water production.

The training and validation workflow include a cascading feature. The neural network was trained and validated at each time step and the outputs of the trained model were used for the next time step. The backpropagation algorithm was used to train the networks. The network architecture (input layer, hidden layer, and output layer) was designed based on the number of inputs/outputs of the training set. The success of training and validation was measured by using the calibration data set performance.

8.5   Verification with a Blind Data Set

As discussed in other sections of this study, strong training results alone do not ensure the generalization ability of the developed neural network. The trained ANN has to be applied to a blind data set for verification. Therefore, the developed proxy models were examined with a designed simulation case that had not been used in the training process.

The designed blind simulation run was the same one used for the smart proxy model developed for reservoir properties in order to verify both smart proxies (reservoir property model and production profile model).
The smart proxy model deployment used a cascading feature. The proxy model was applied at each time step and the output was used for the next time step. In other words, the smart proxy production inputs from the previous time step were imported from the numerical simulation results only at the initial time step.

In the following sections, the training and deployment results are demonstrated for each production phase.

8.6   Phase One Production Smart Proxy Model

8.6.1   Training Results

In phase one of production, there were only three wells producing. These were well 10-4, well 10-5, and well 11-3. Network training was challenging due to a low number of data records. However, the right input selection provided the desired training results.
The network training results of the three production profiles (oil rate, gas rate, and water rate) are shown in the following figures.

Figure 8.2: Phase One Oil Production Training Cross Plot

Figure 8.3: Phase One Gas Production Training Cross Plot
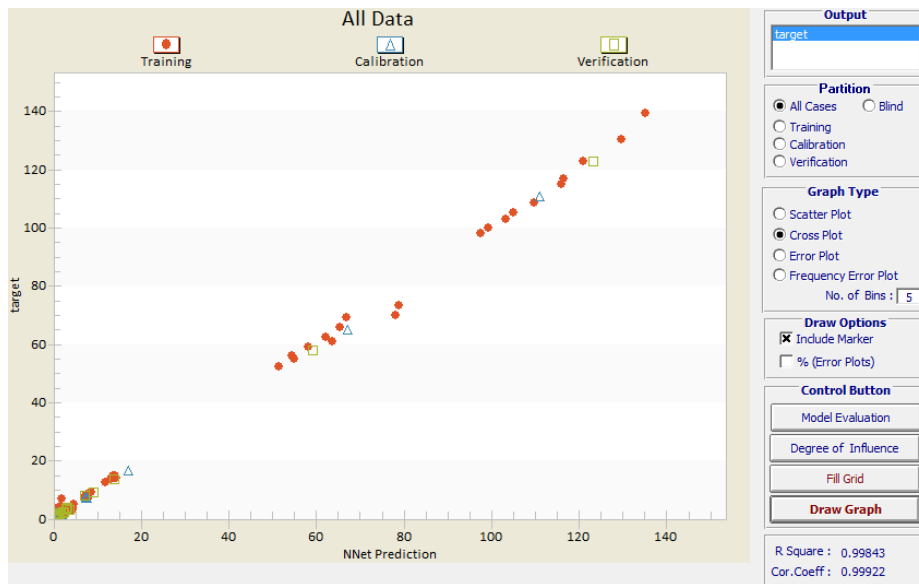
Figure 8.4: Phase One Water Production Training Cross Plot

8.6.2 Blind Run Results

The smart proxy deployment results are shown in the form of data maps for the oil, gas, and water production profiles. In general, compared to the numerical simulator, the results show an acceptable performance of the smart proxy model.
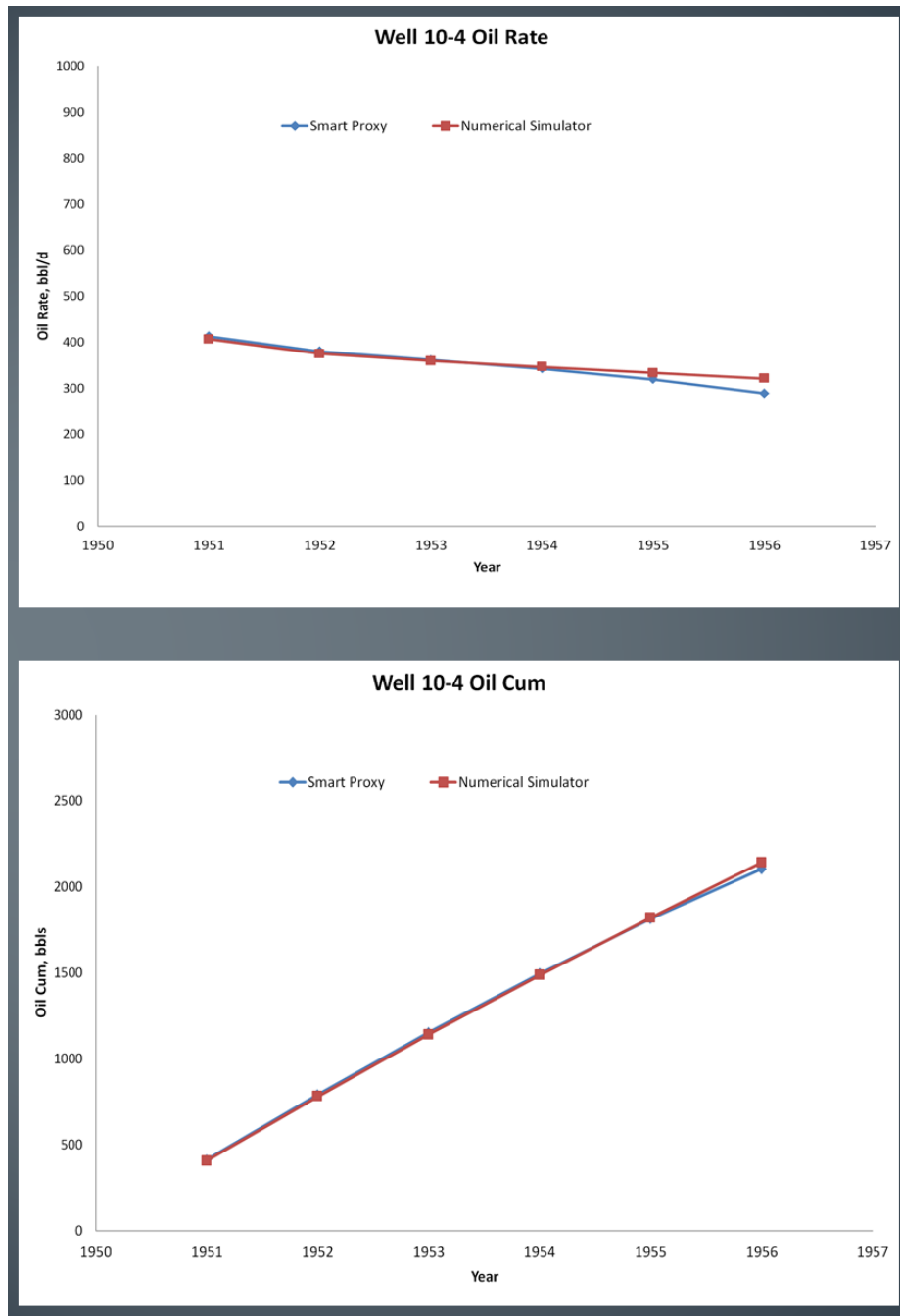
Figure 8.5: Phase one well 10-4 oil production profile
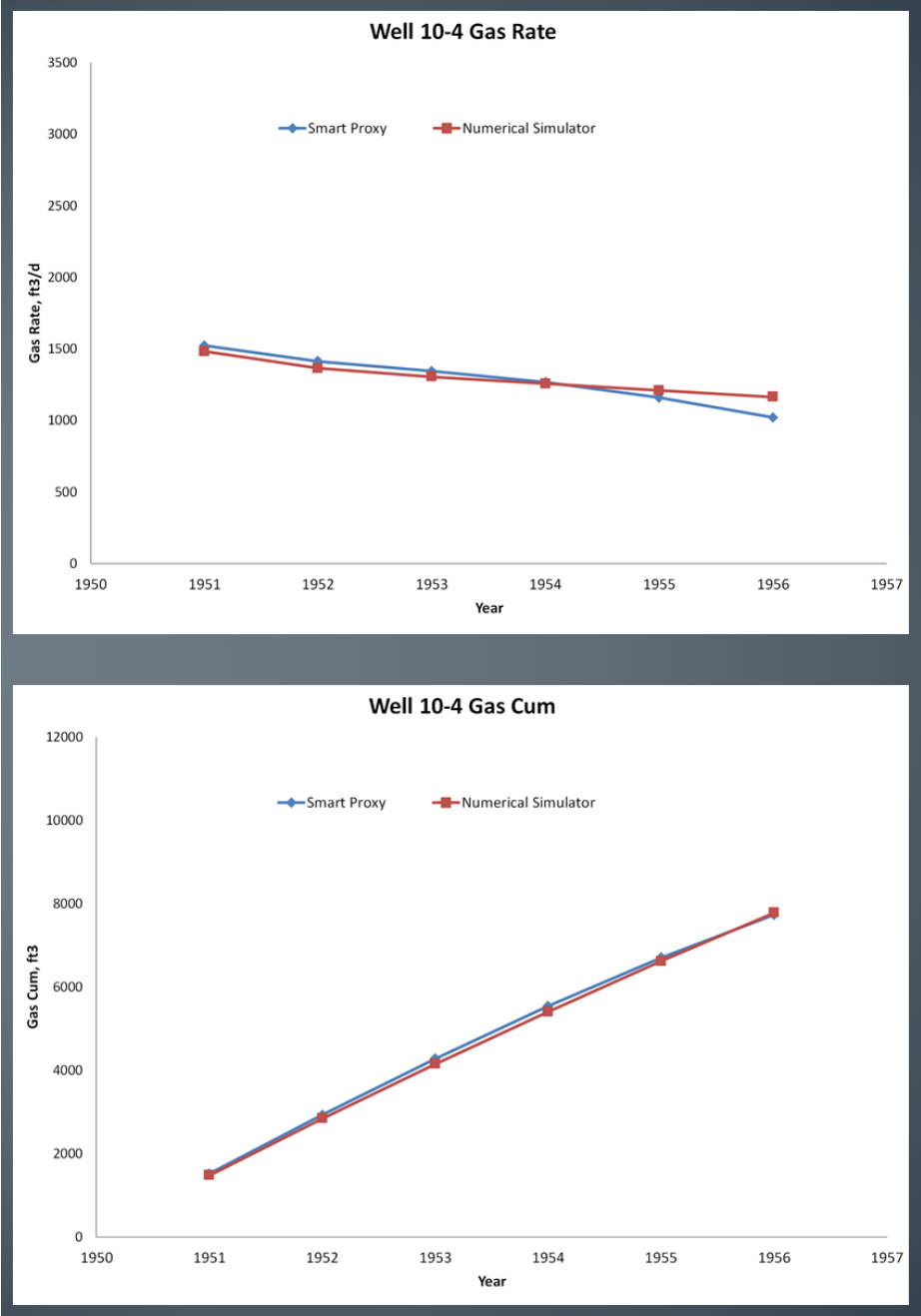
Figure 8.6: Phase one well 10-4 gas rate profile
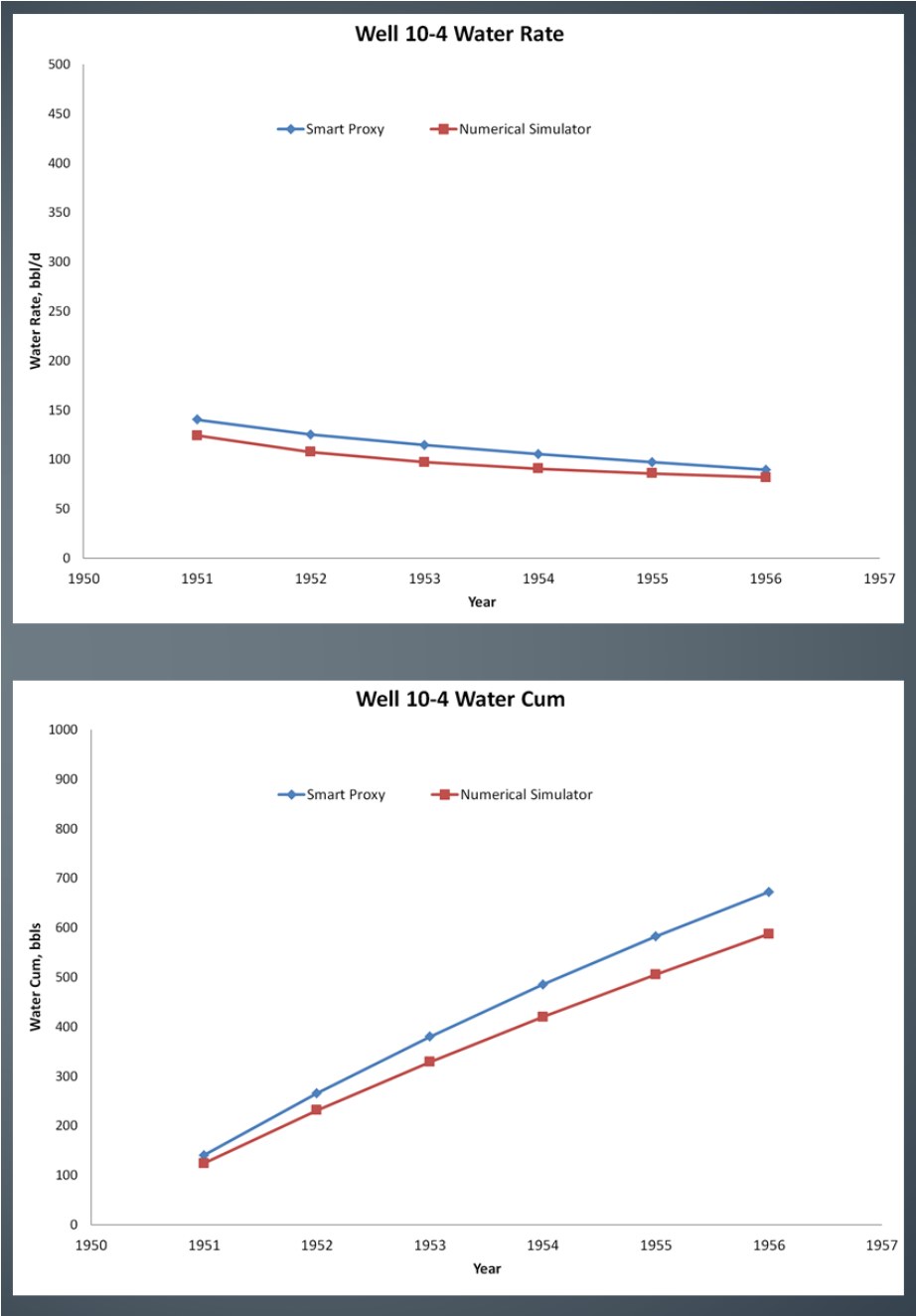
Figure 8.7: Phase one well 10-4 water rate profile
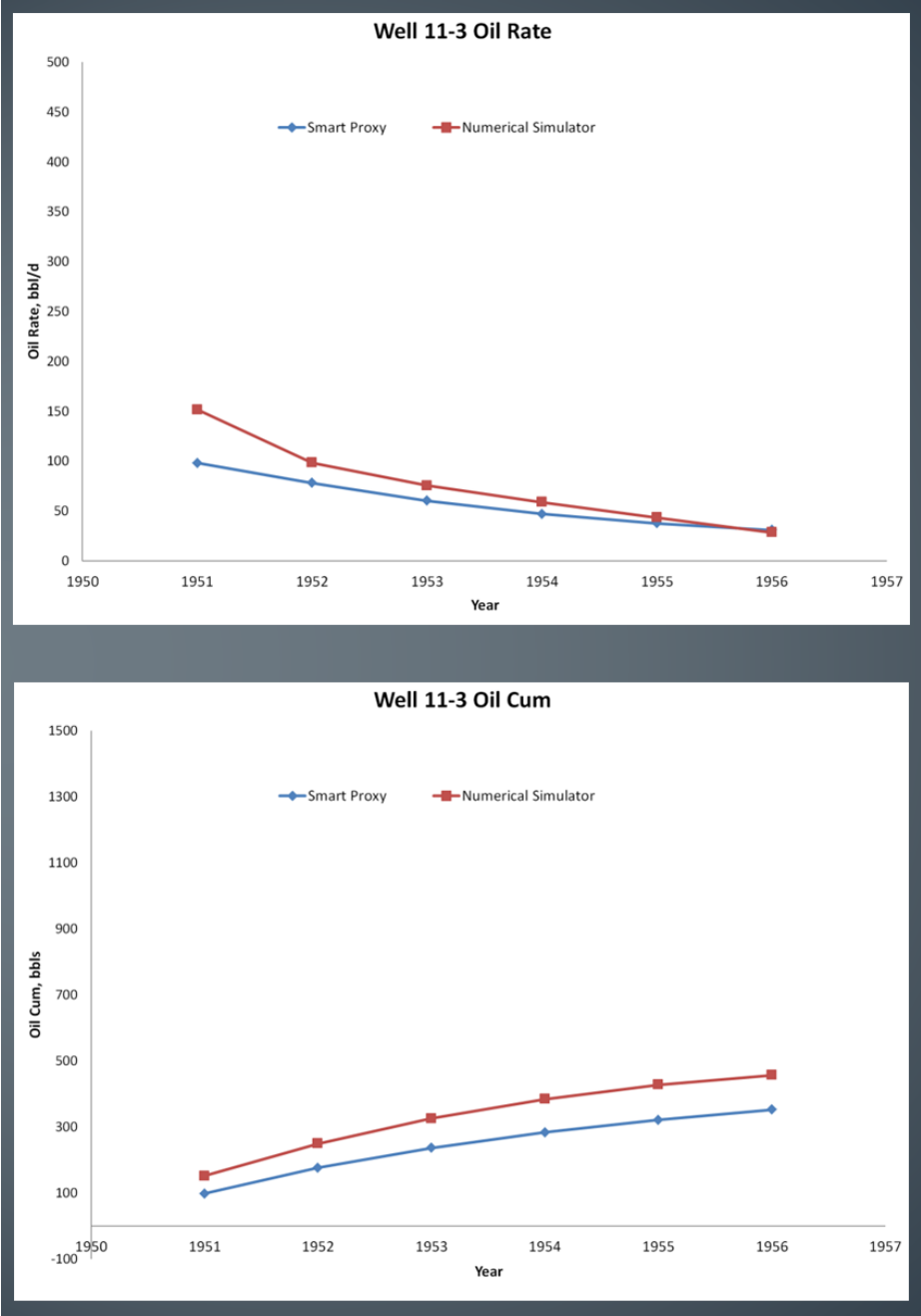
Figure 8.8: Phase one well 11-3 oil rate profile
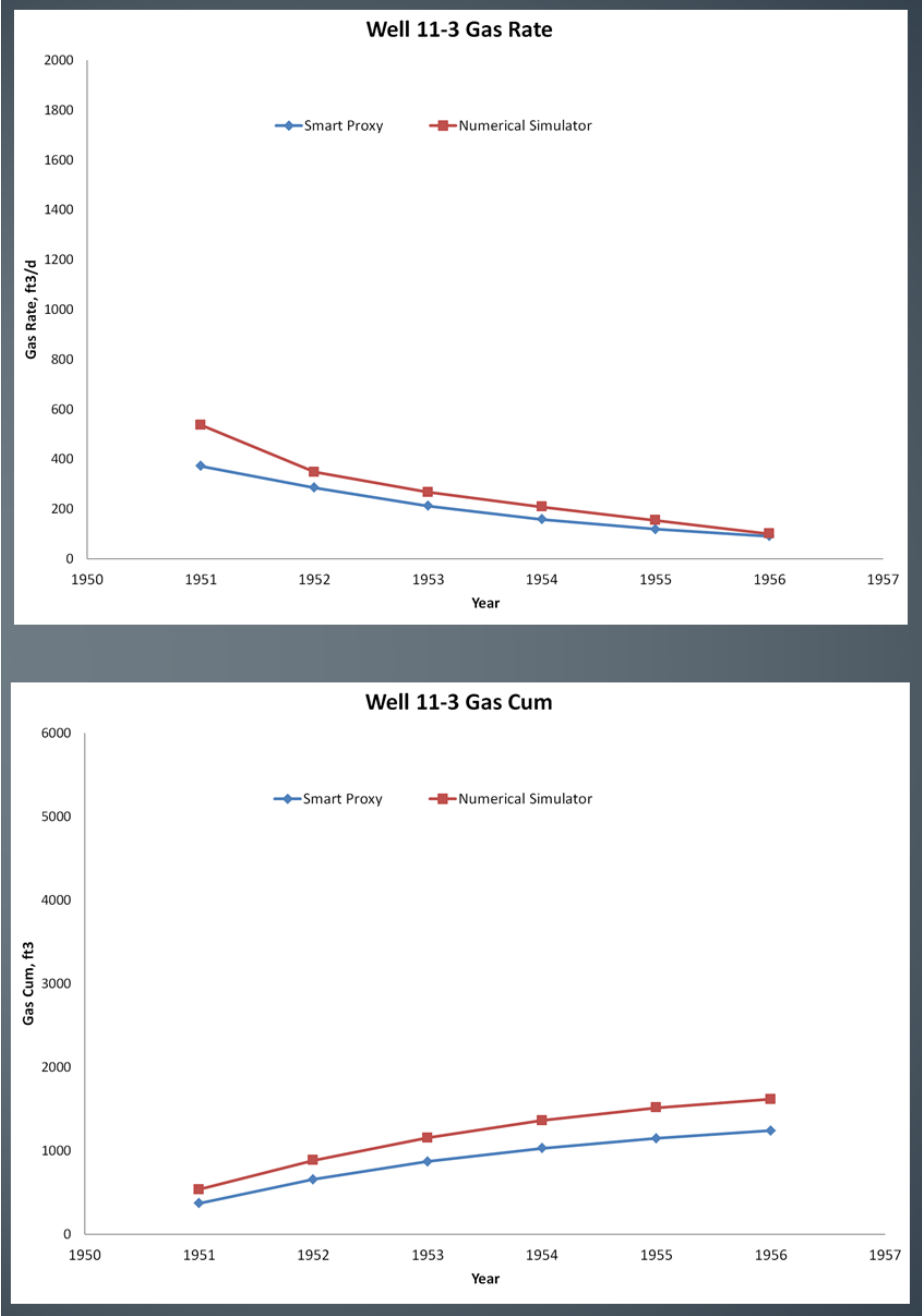
Figure 8.9: Phase one well 11-3 gas rate profile
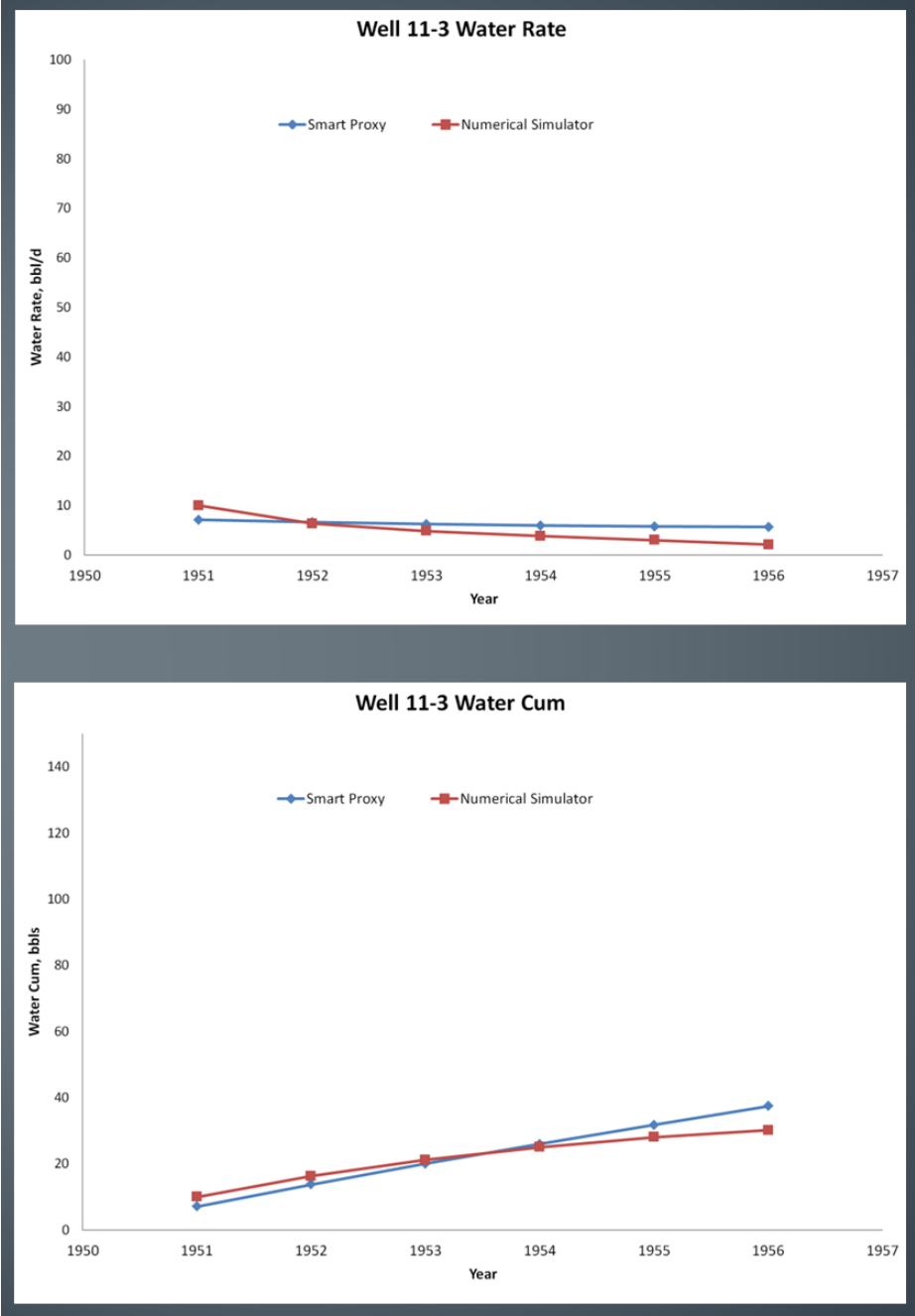
Figure 8.10: Phase one well 11-3 water rate profile

Figure 8.11: Phase one well 10-5 oil rate profile
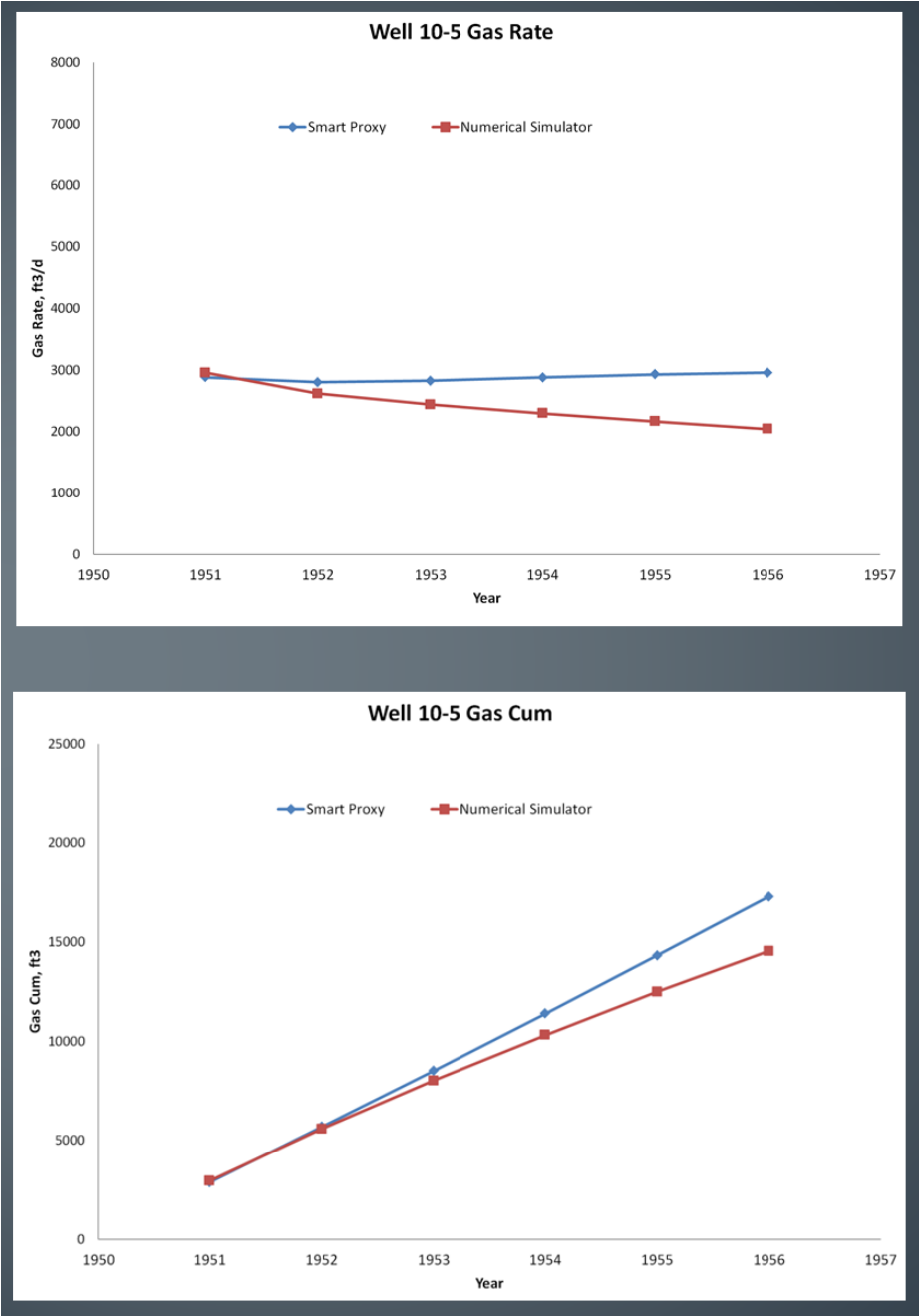
Figure 8.12: Phase one well 10-5 gas rate profile
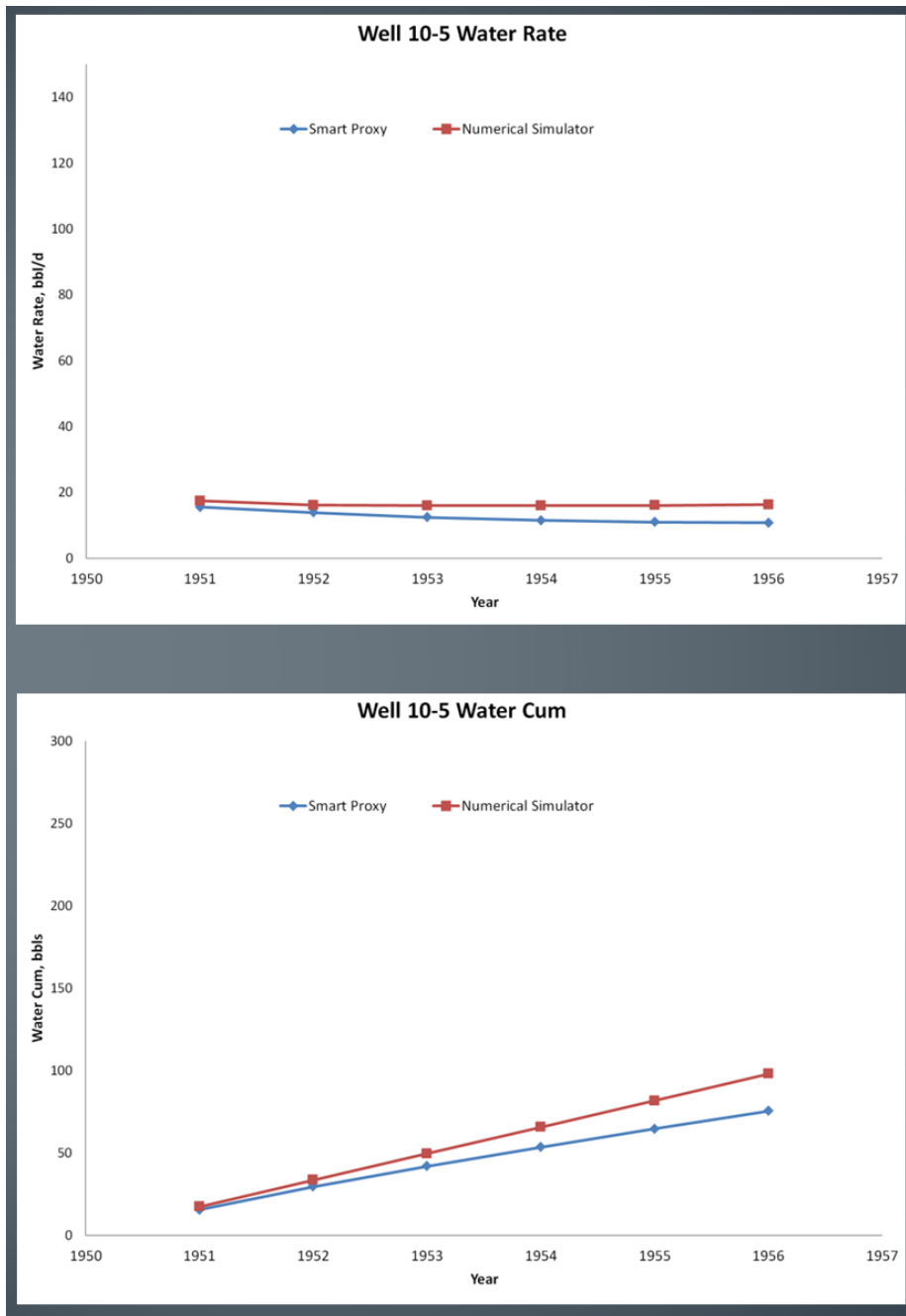
Figure 8.13: Phase one well 10-5 water rate profile

8.7 Phase Two Production Smart Proxy Model

8.7.1 Training Results

Phase two production was similar to phase one with regards to the low number of data records for training. In this phase, there were four active production wells. These wells were 10-1, 10-2, 11-4, and 11-5. Again, the domain knowledge allowed for selecting the required input data for training. The network training is shown in the following cross plots.
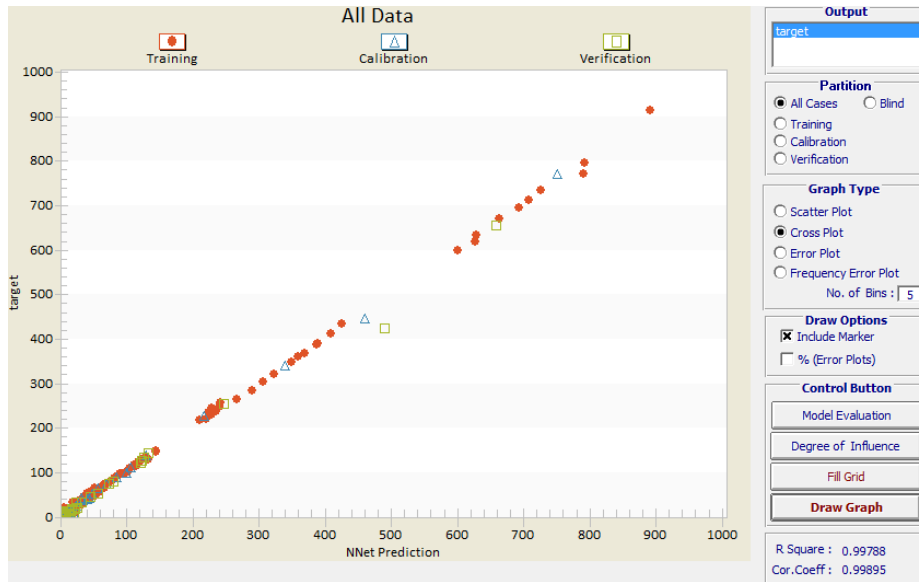

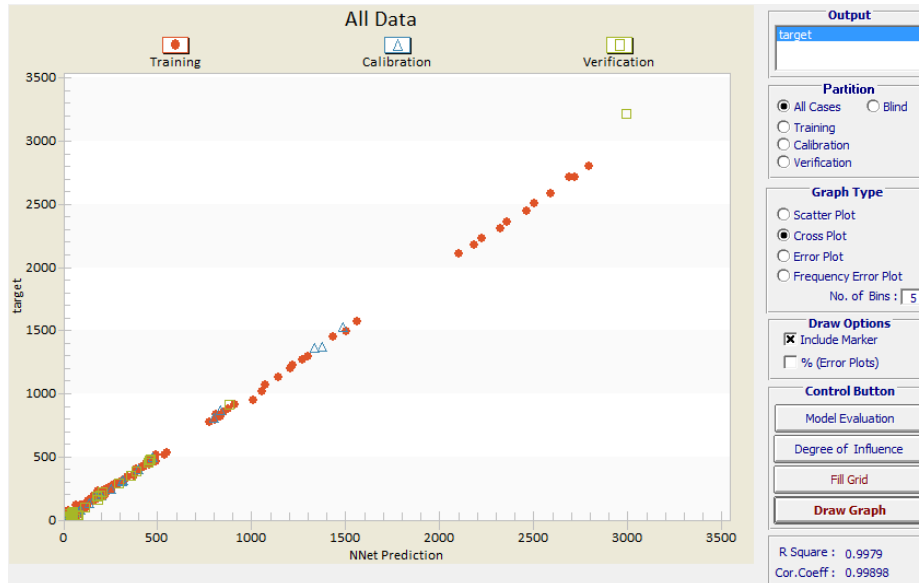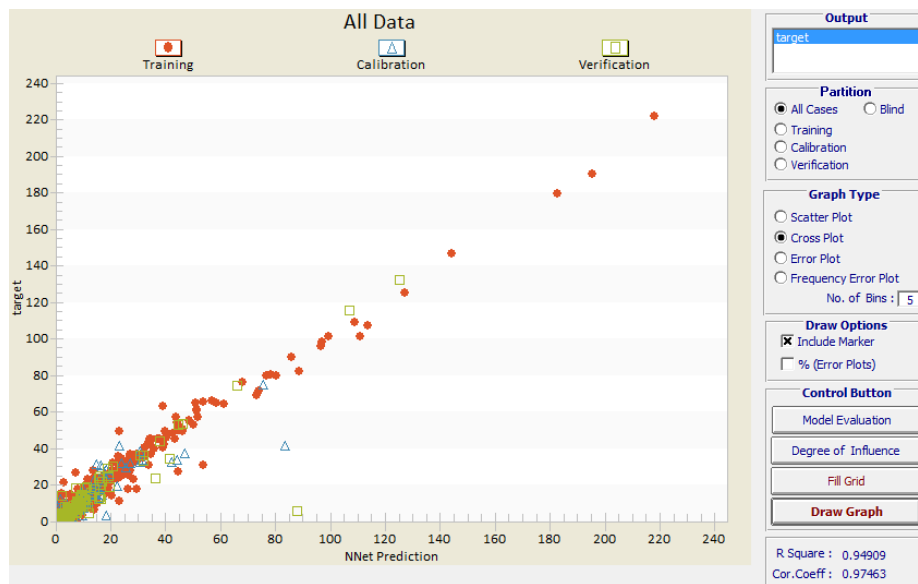Figure 8.14: Phase two Oil Production Training Cross Plot
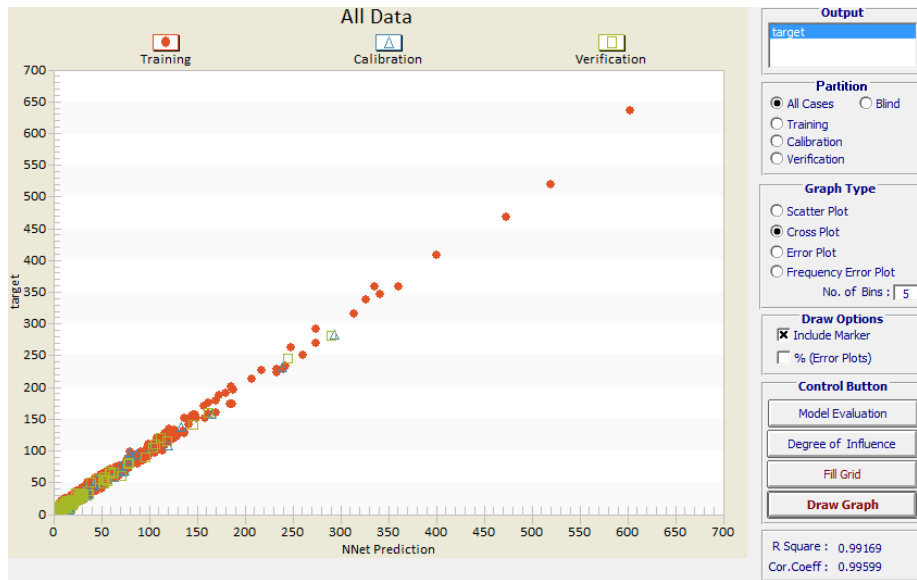
Figure 8.15: Phase two Gas Production Training Cross Plot
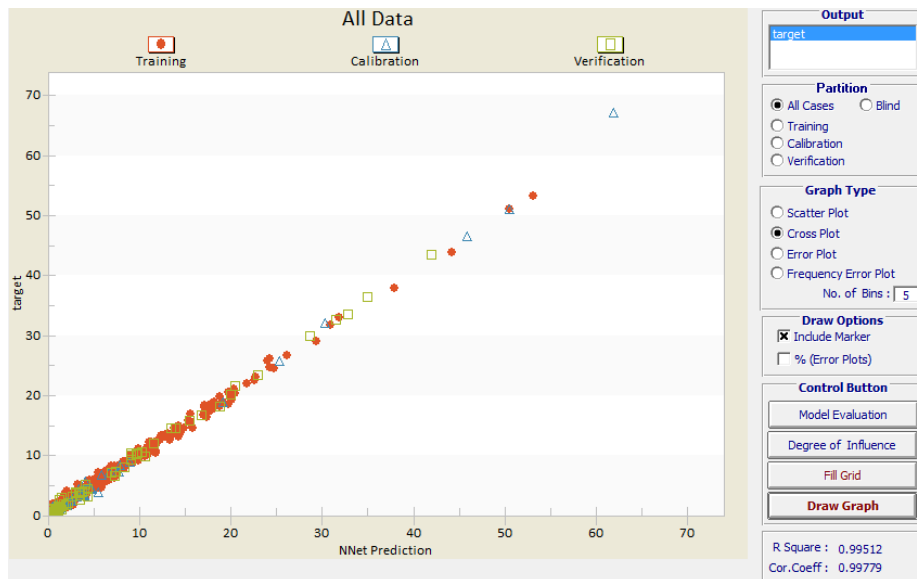
Figure 8.16: Phase two Water Production Training Cross Plot

8.7.2 Blind Run Results

The application of the smart proxy model results compared to the numerical simulation results are shown in the following figures. However, the results from well 11-4 differ significantly from the numerical simulation and are therefore not displayed in this section.
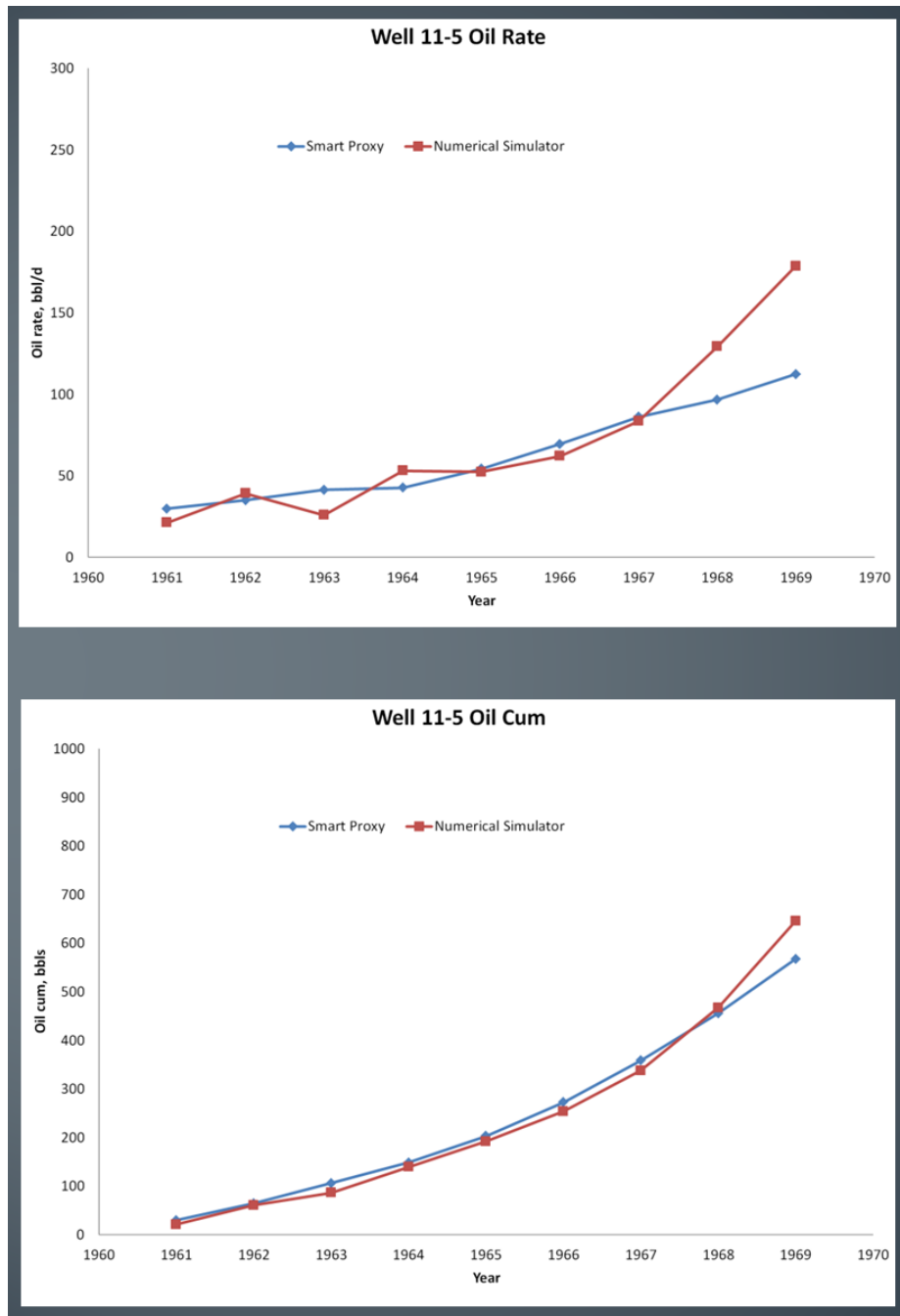
Figure 8.17: Phase two well 11-5 oil production profile

Figure 8.18: Phase two well 11-5 gas production profile
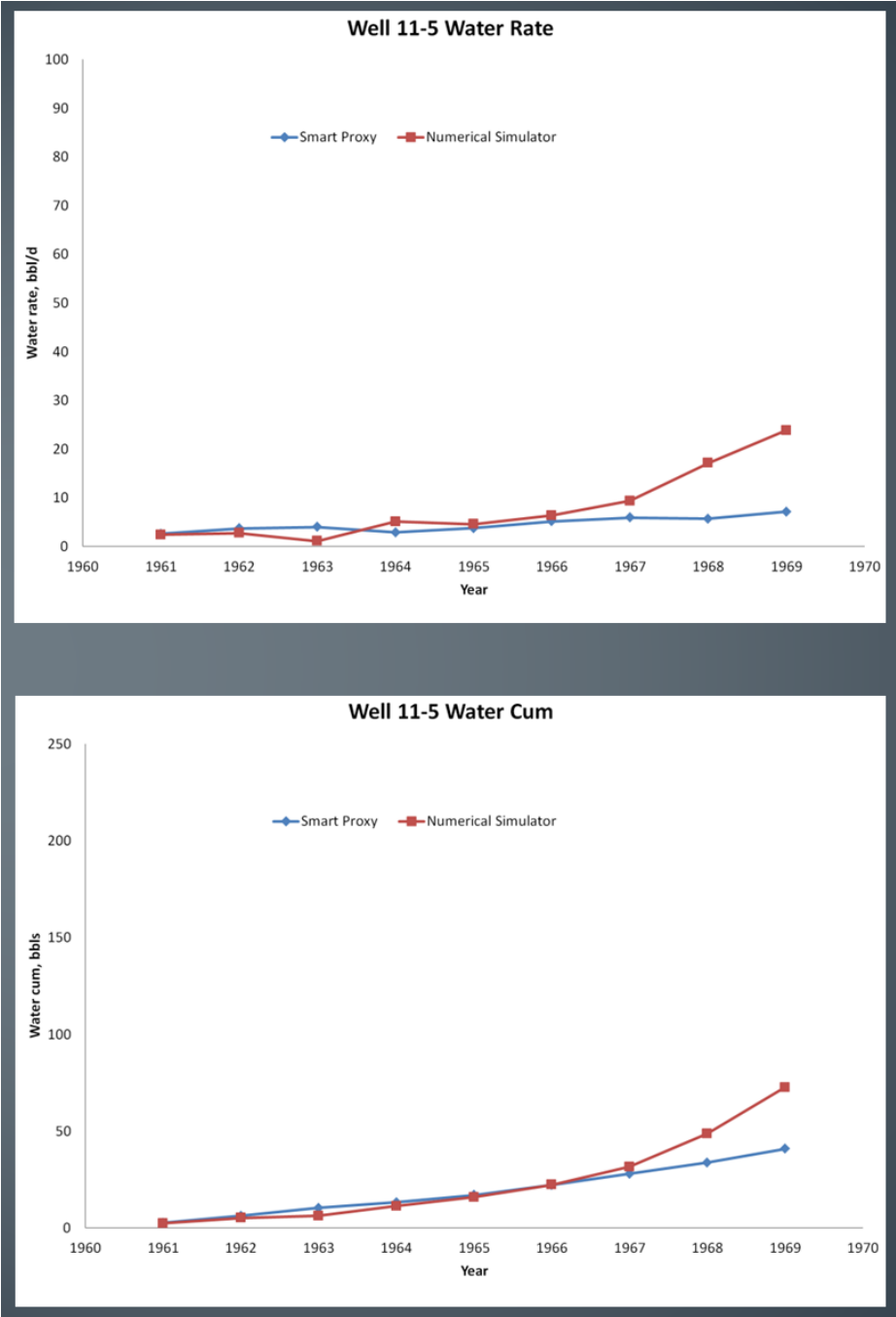
Figure 8.19: Phase two well 11-5 water production profile
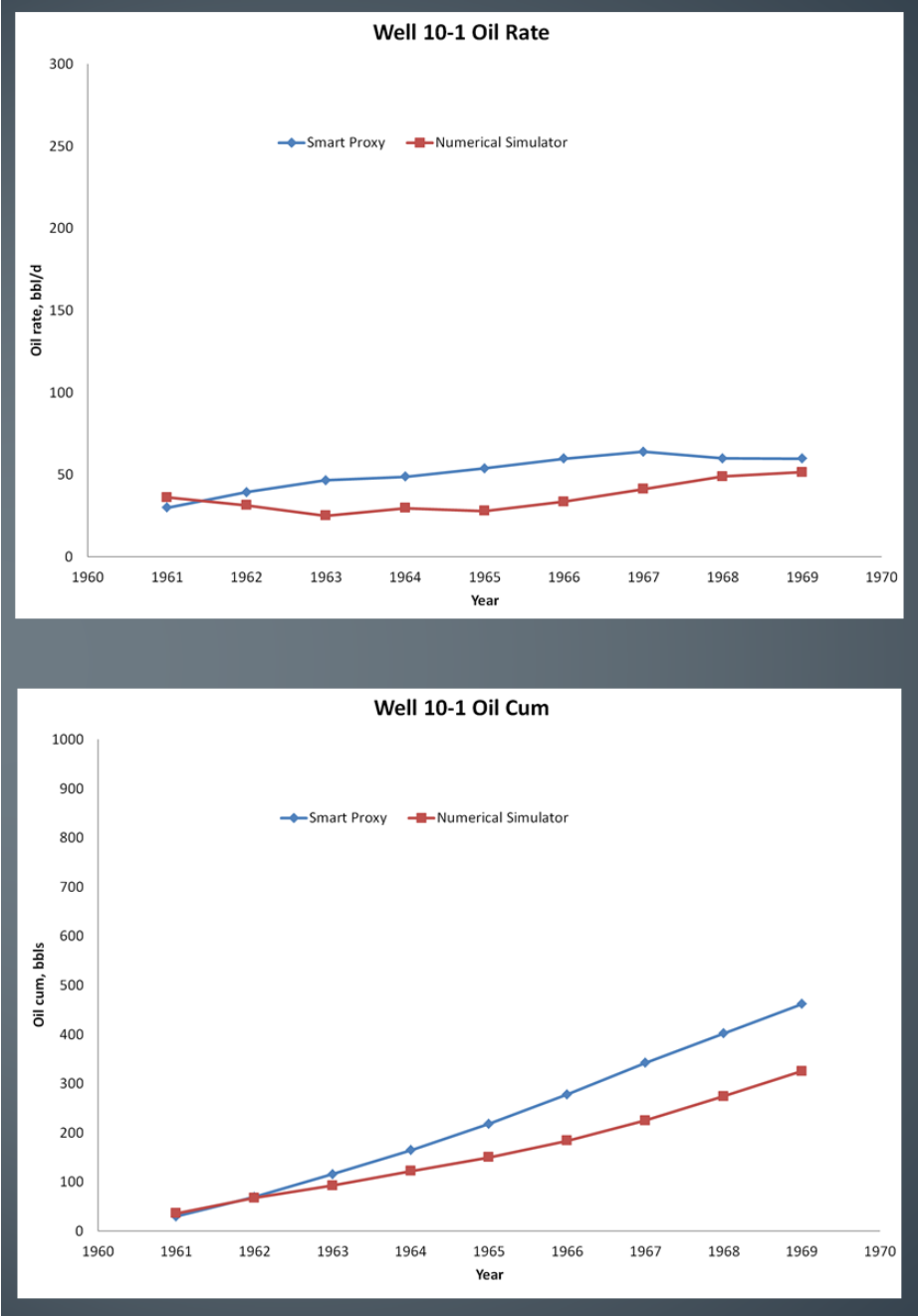
Figure 8.20: Phase two well 10-1 oil production profile
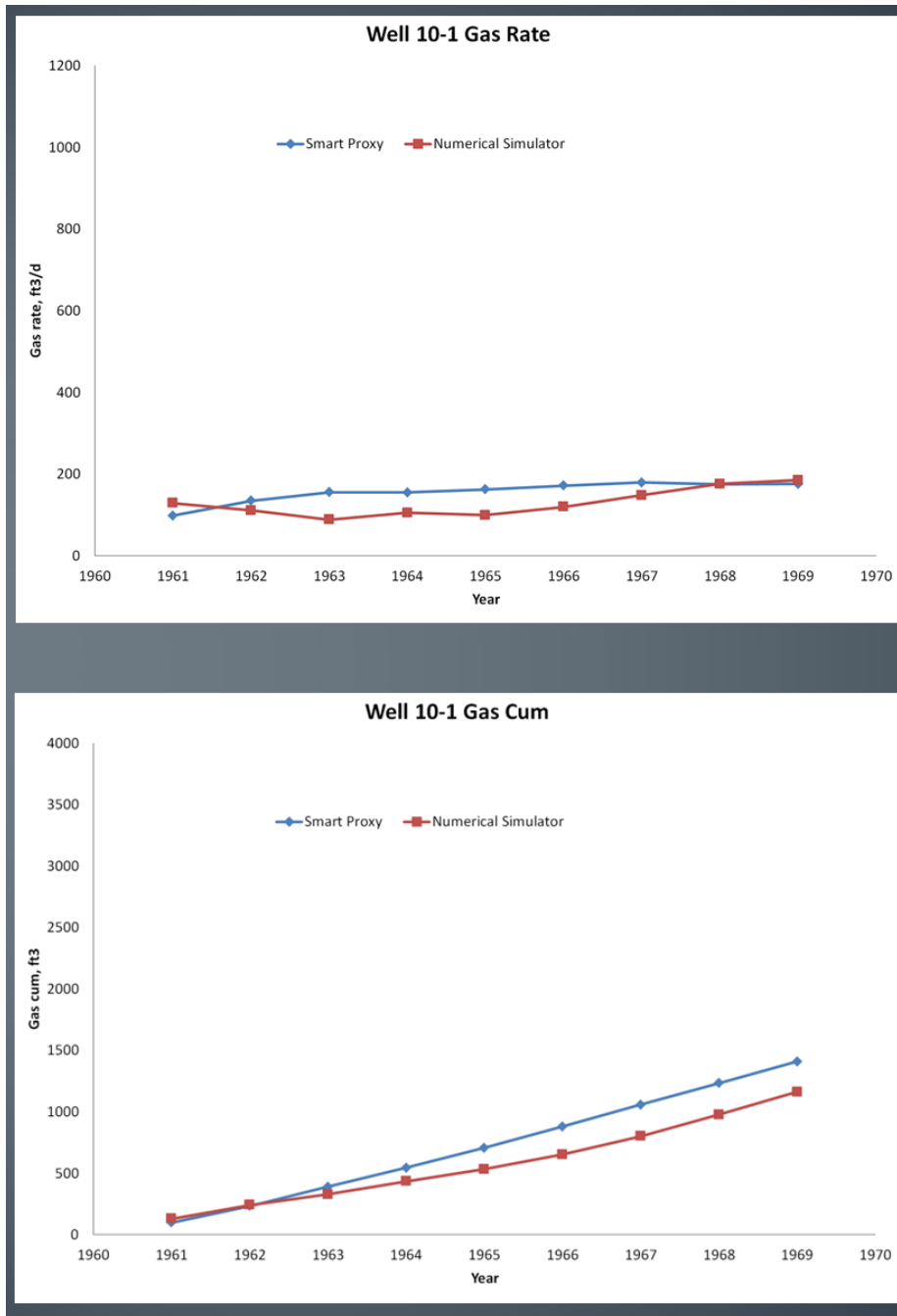
Figure 8.21: Phase two well 10-1 gas production profile

Figure 8.22: Phase two well 10-1 water production profile

Figure 8.23: Phase two well 10-1 oil production profile
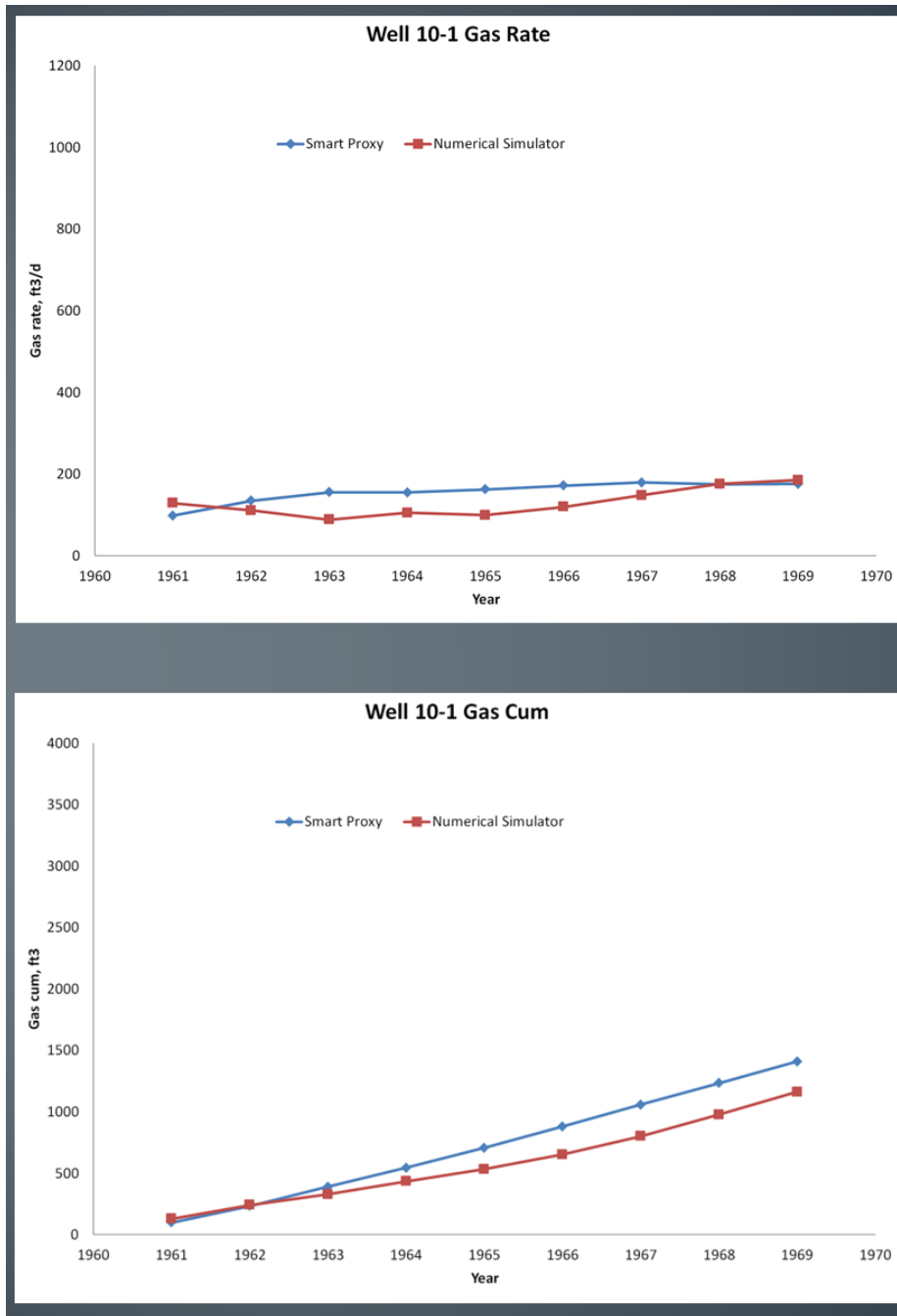
Figure 8.24: Phase two well 10-1 gas production profile
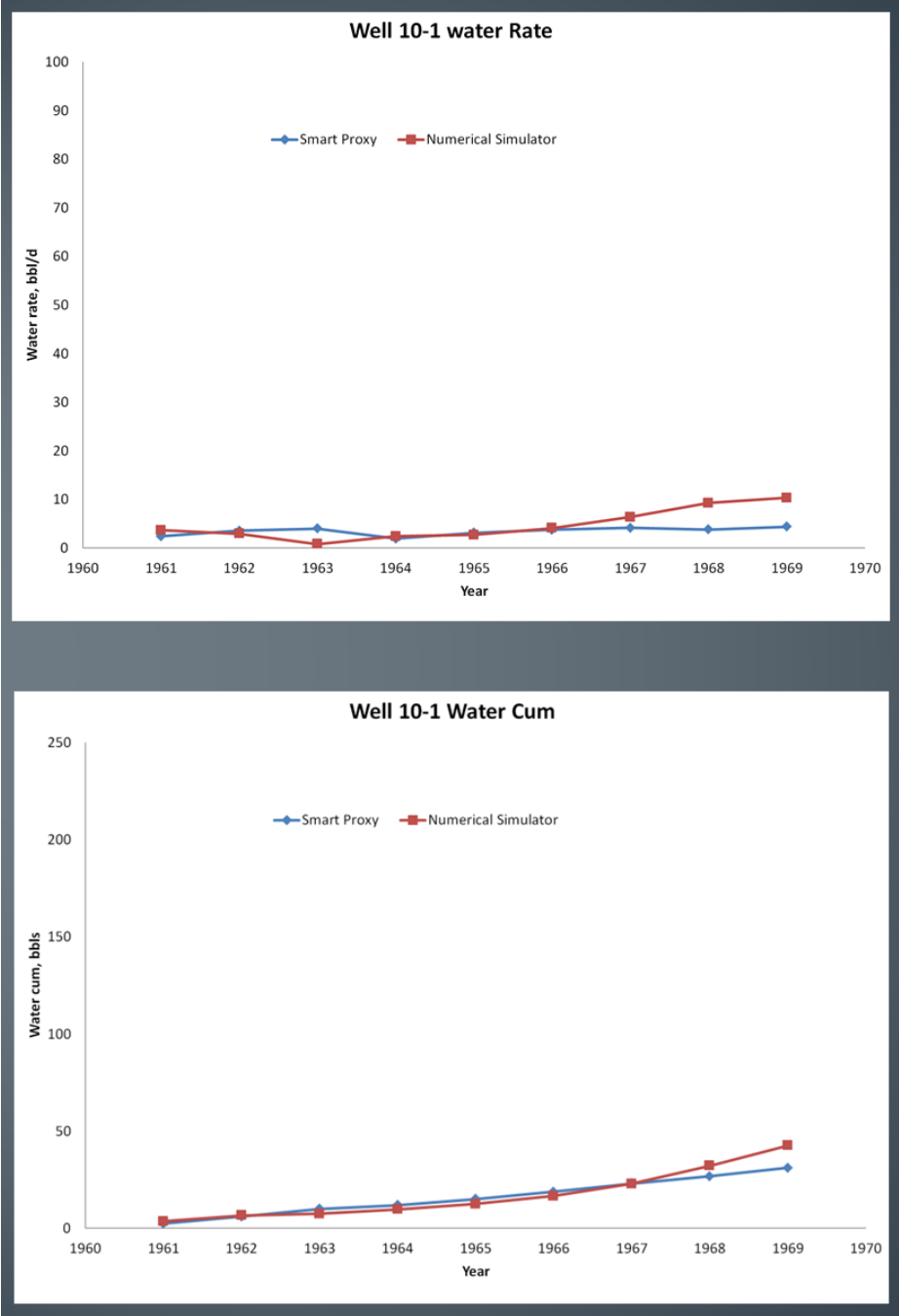
Figure 8.25: Phase two well 10-1 water production profile

8.8 Phase Three Production Smart Proxy Model

8.8.1 Training Results

Phase three production performance was the most active with regards to the number of wells and the production time. There were 15 production wells in this phase. Moreover, the production duration was 20 years. Therefore, the number of data records available for training was sufficient for a robust proxy development. In addition, the variation of production profiles coming from these wells provides the required data heterogeneity for network training and generalization.

The following cross plots are the training results of oil rate, gas rate, and water rate.
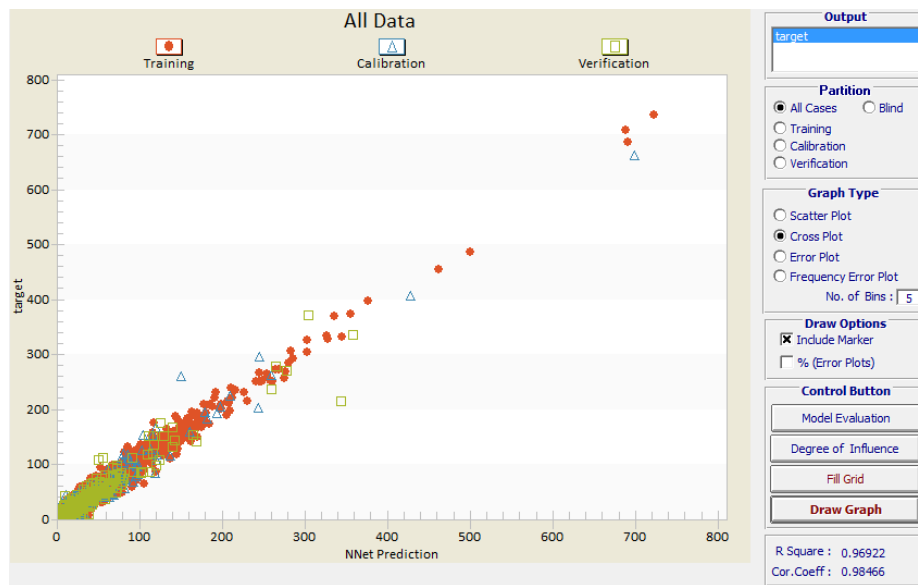
Figure 8.26: Phase three Oil Production Training Cross Plot
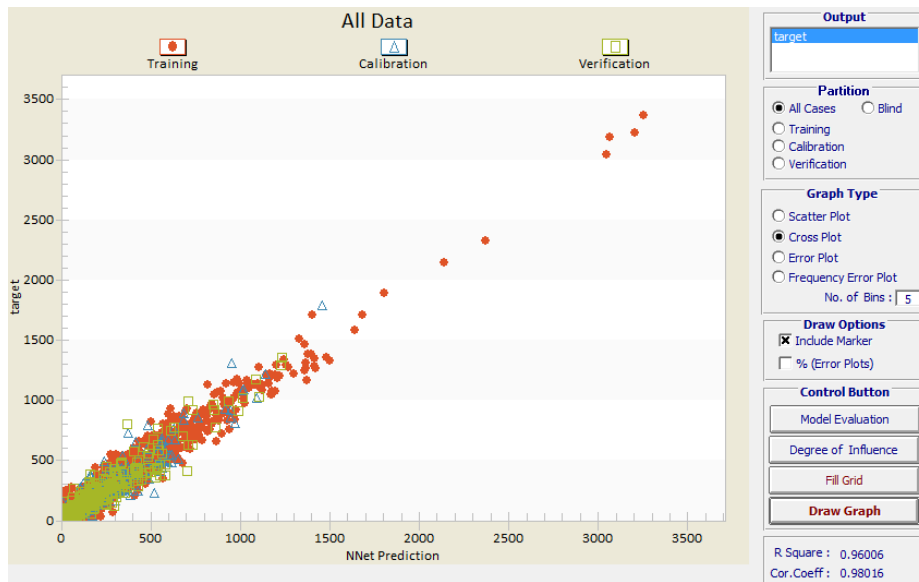
Figure 8.27: Phase three Gas Production Training Cross Plot
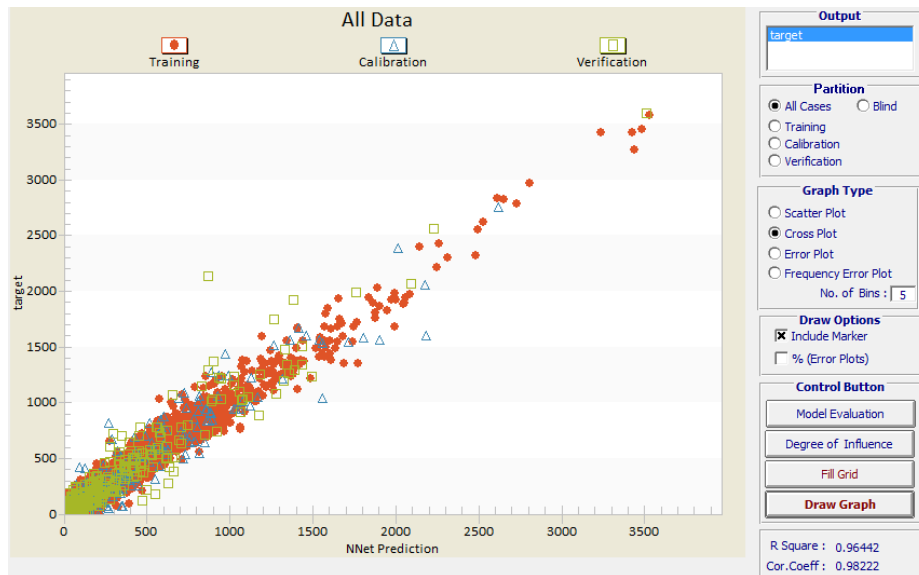
Figure 8.28: Phase three Water Production Training Cross Plot

8.8.2 Blind Run Results

The deployment results of the smart proxy model in phase three showed a very close match to the production results of the numerical simulator for all wells. The error between the proxy results compared to the numerical simulator is within the acceptable range. The following figures show the deployment results compared to the numerical simulator for the oil, gas, and water rates profiles.
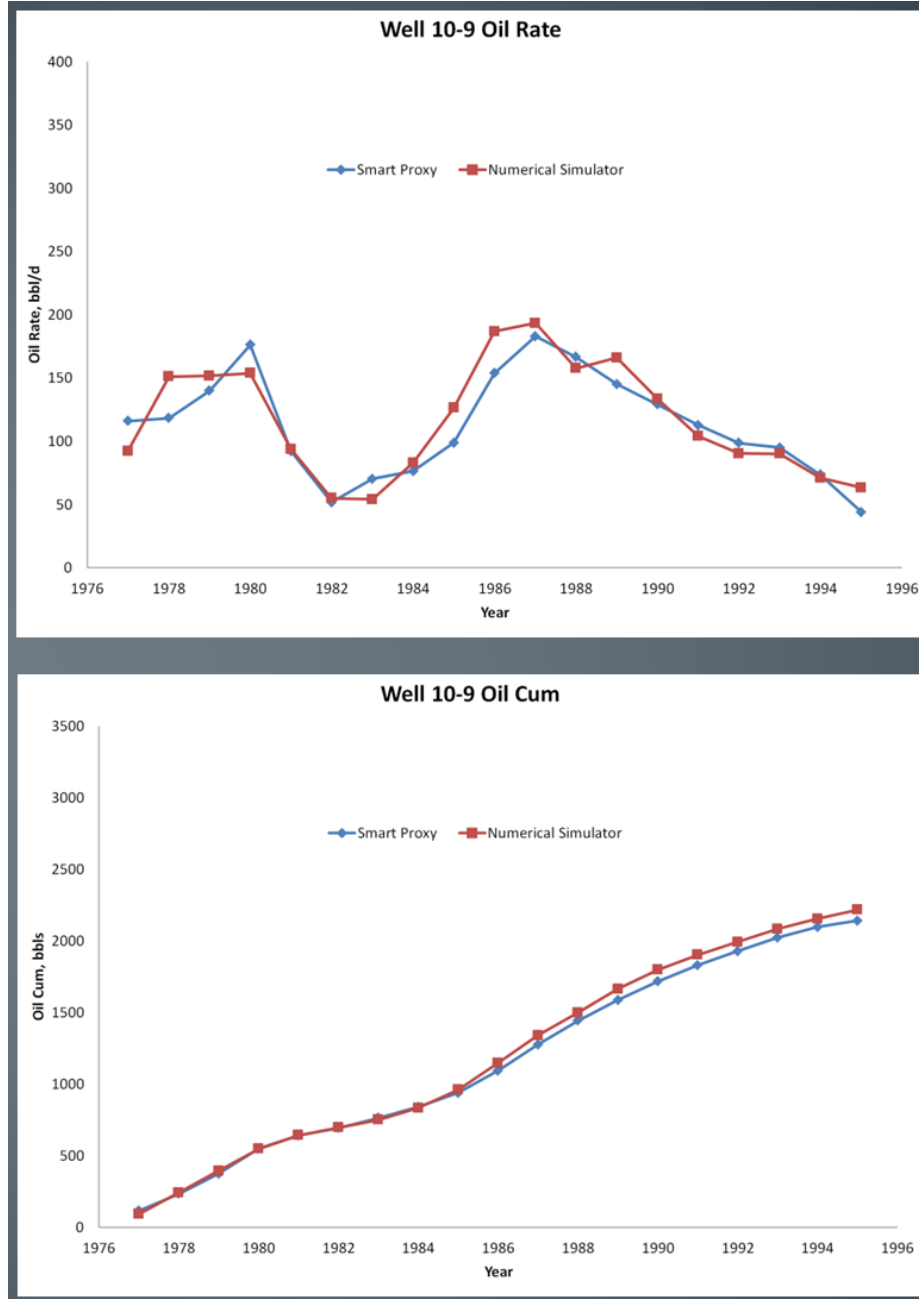
Figure 8.29: Phase three well 10-9 oil production profile
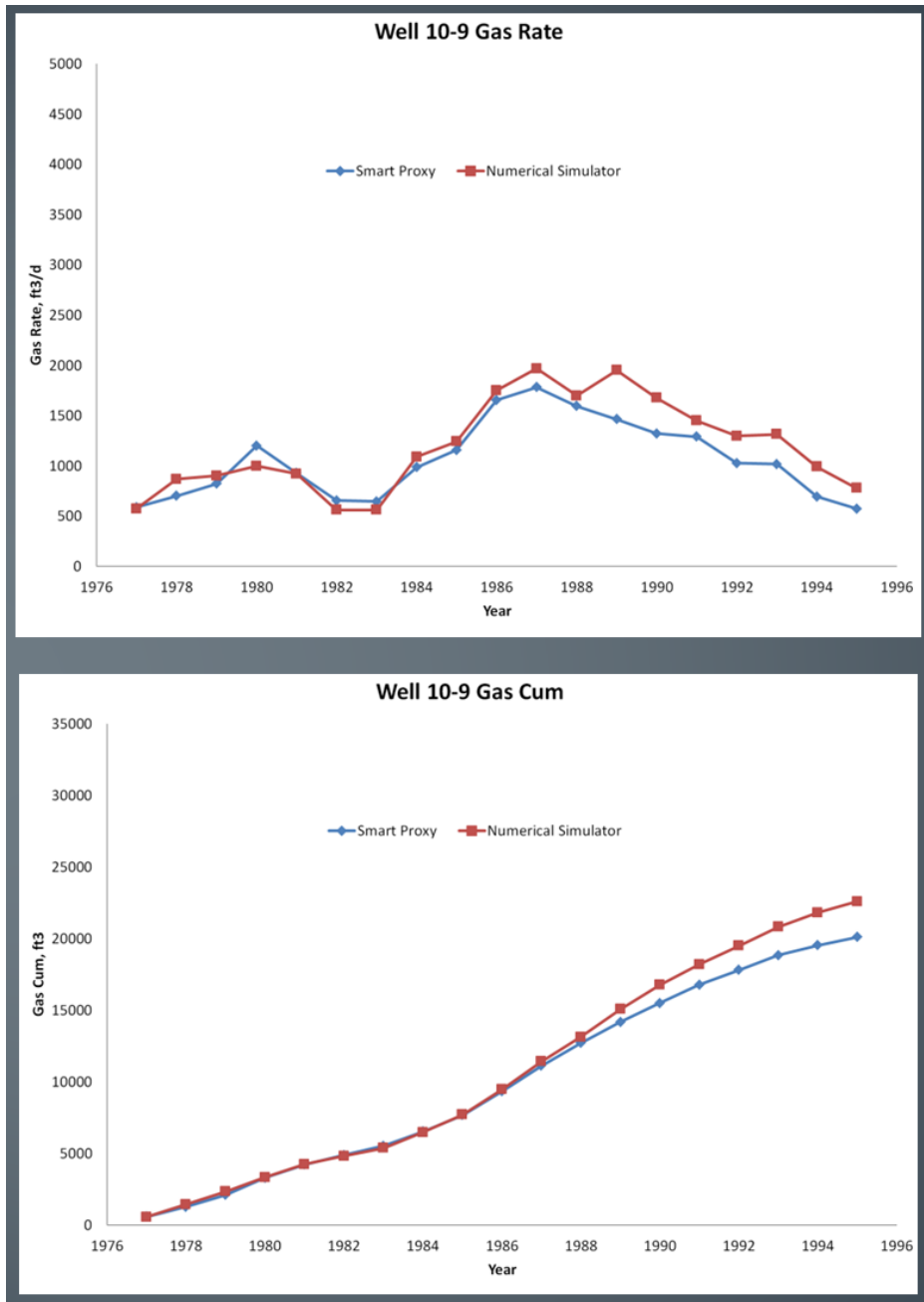
Figure 8.30: Phase three well 10-9 gas production profile
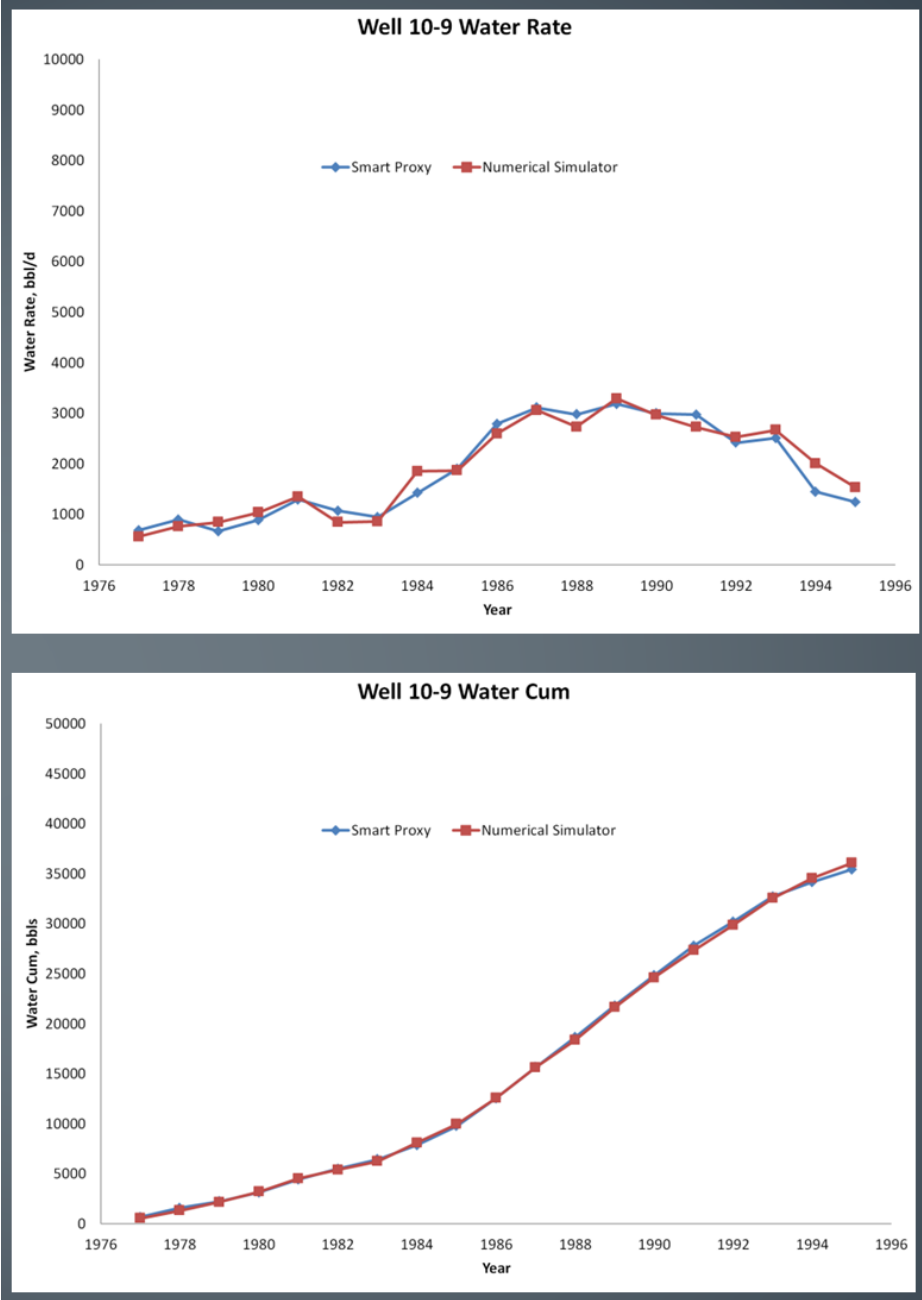
Figure 8.31: Phase three well 10-9 water production profile
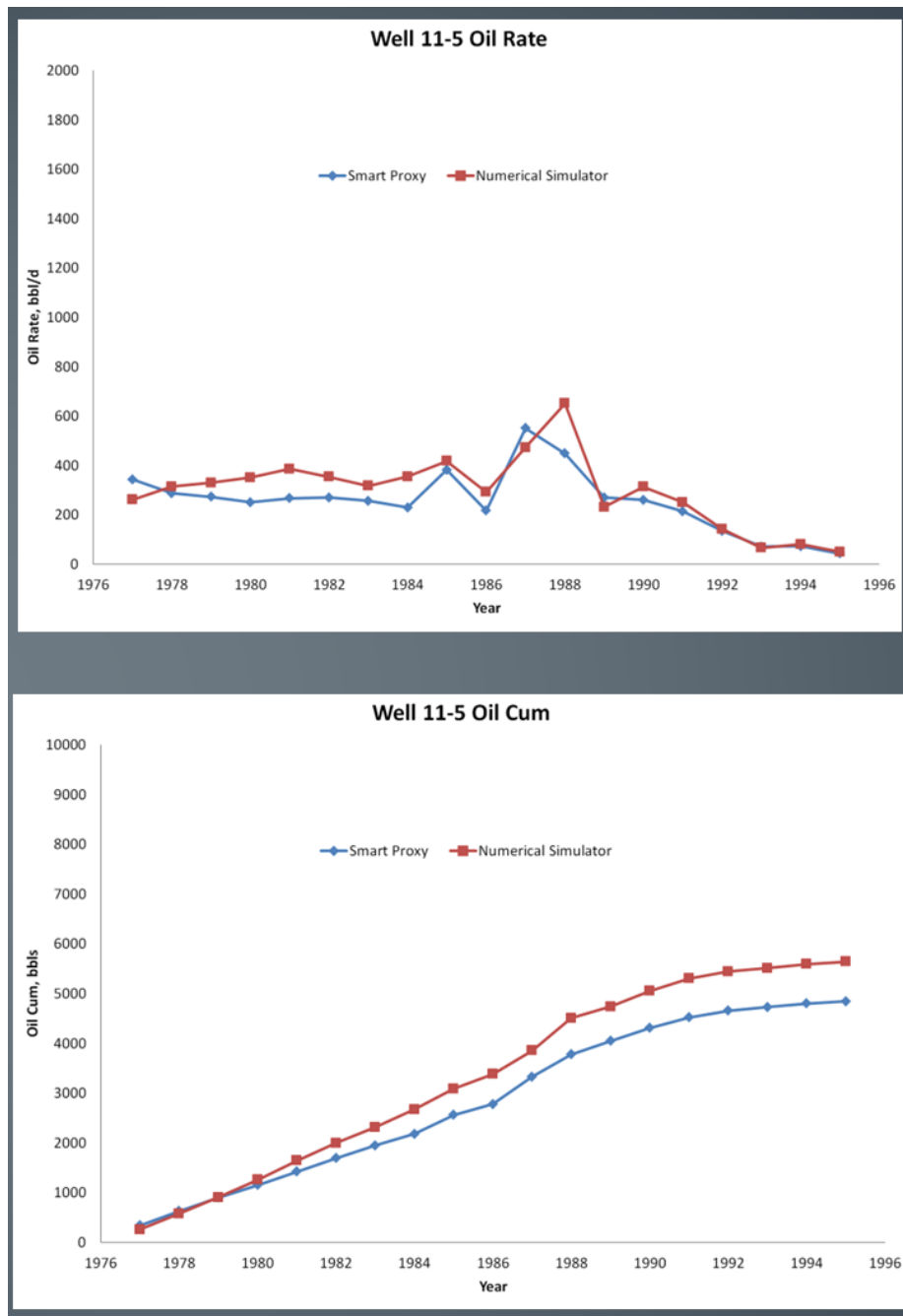
Figure 8.32: Phase three well 11-5 oil production profile

Figure 8.33: Phase three well 11-5 gas production profile
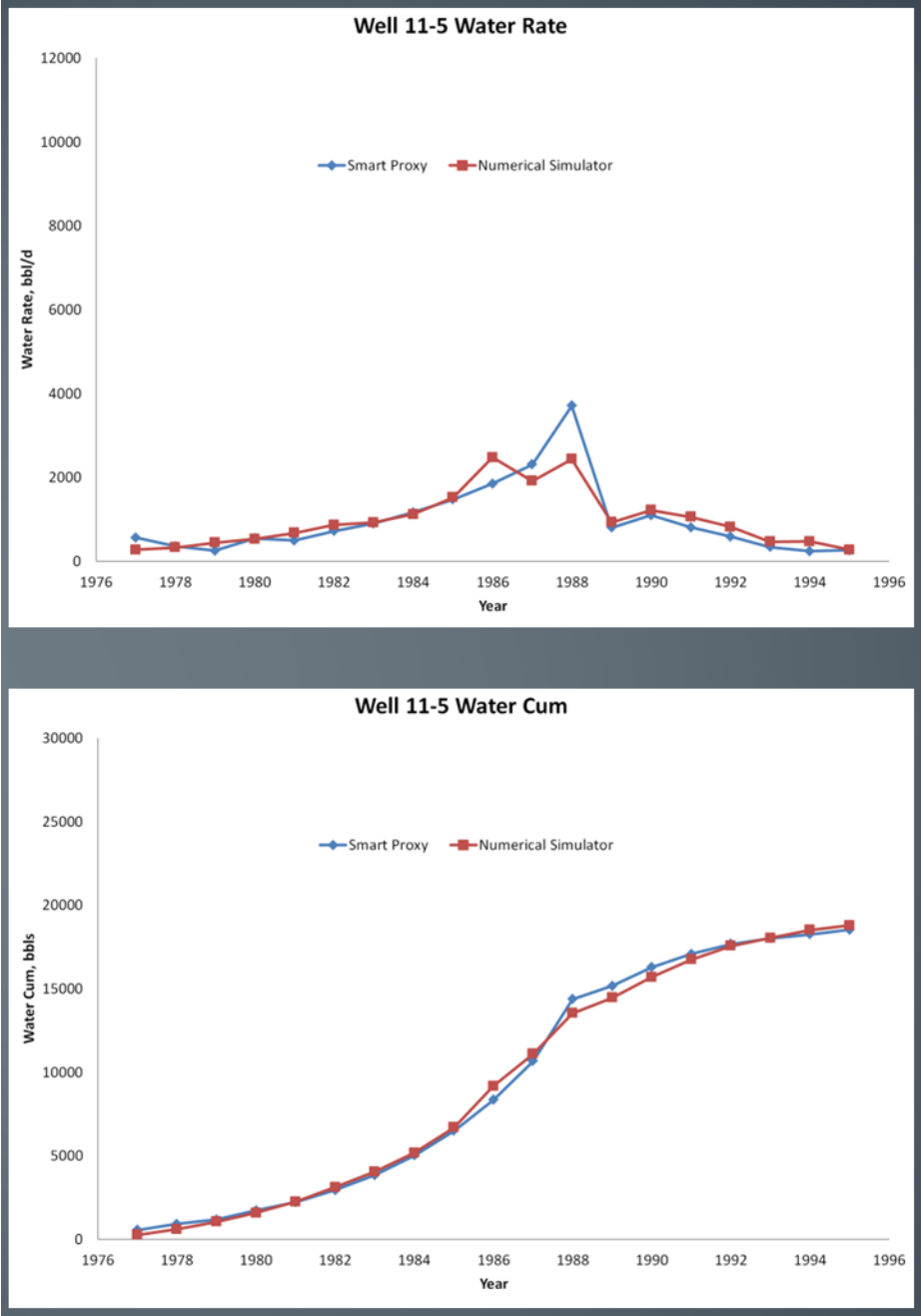
Figure 8.34: Phase three well 11-5 water production profile
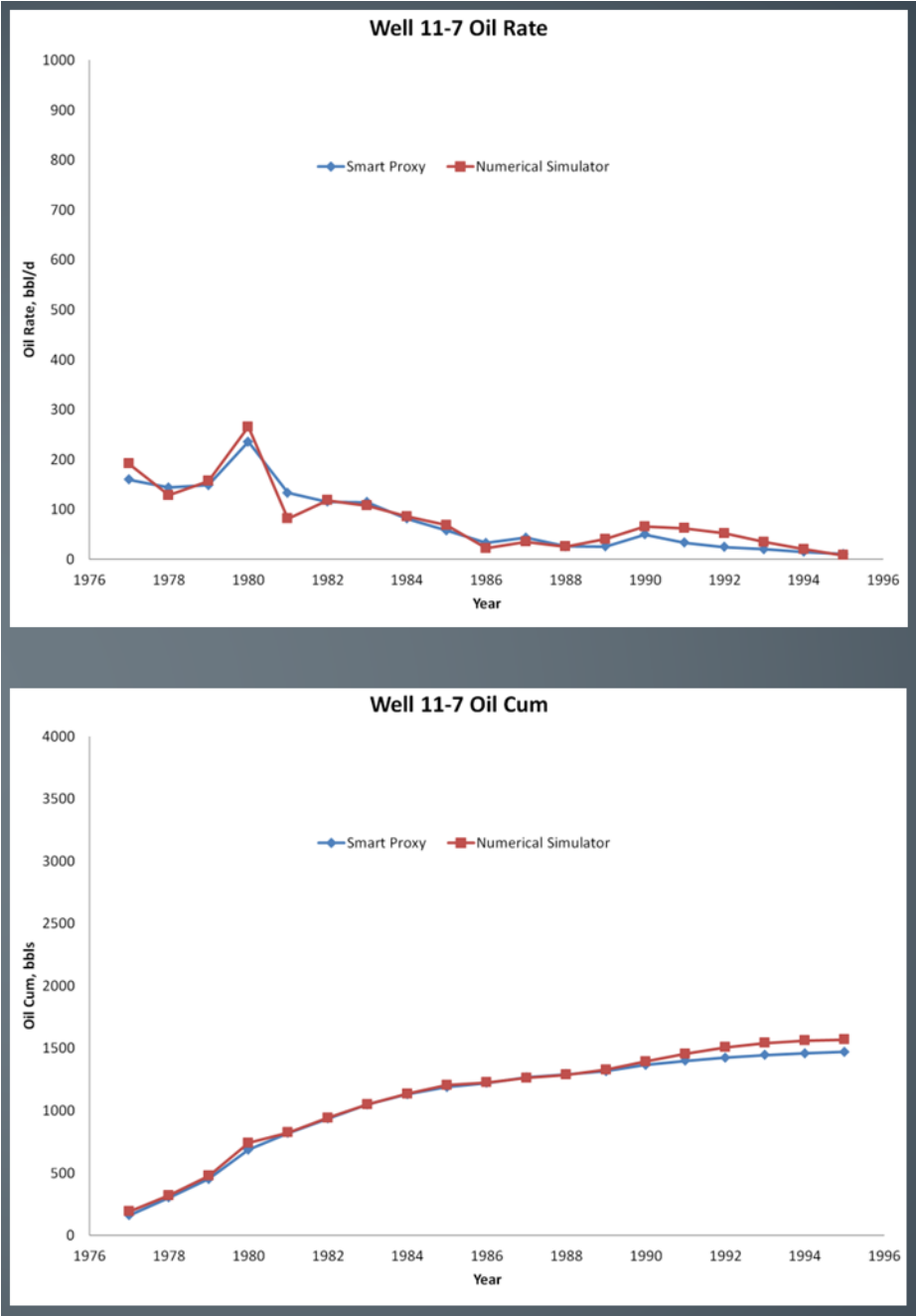
Figure 8.35: Phase three well 11-7 oil production profile
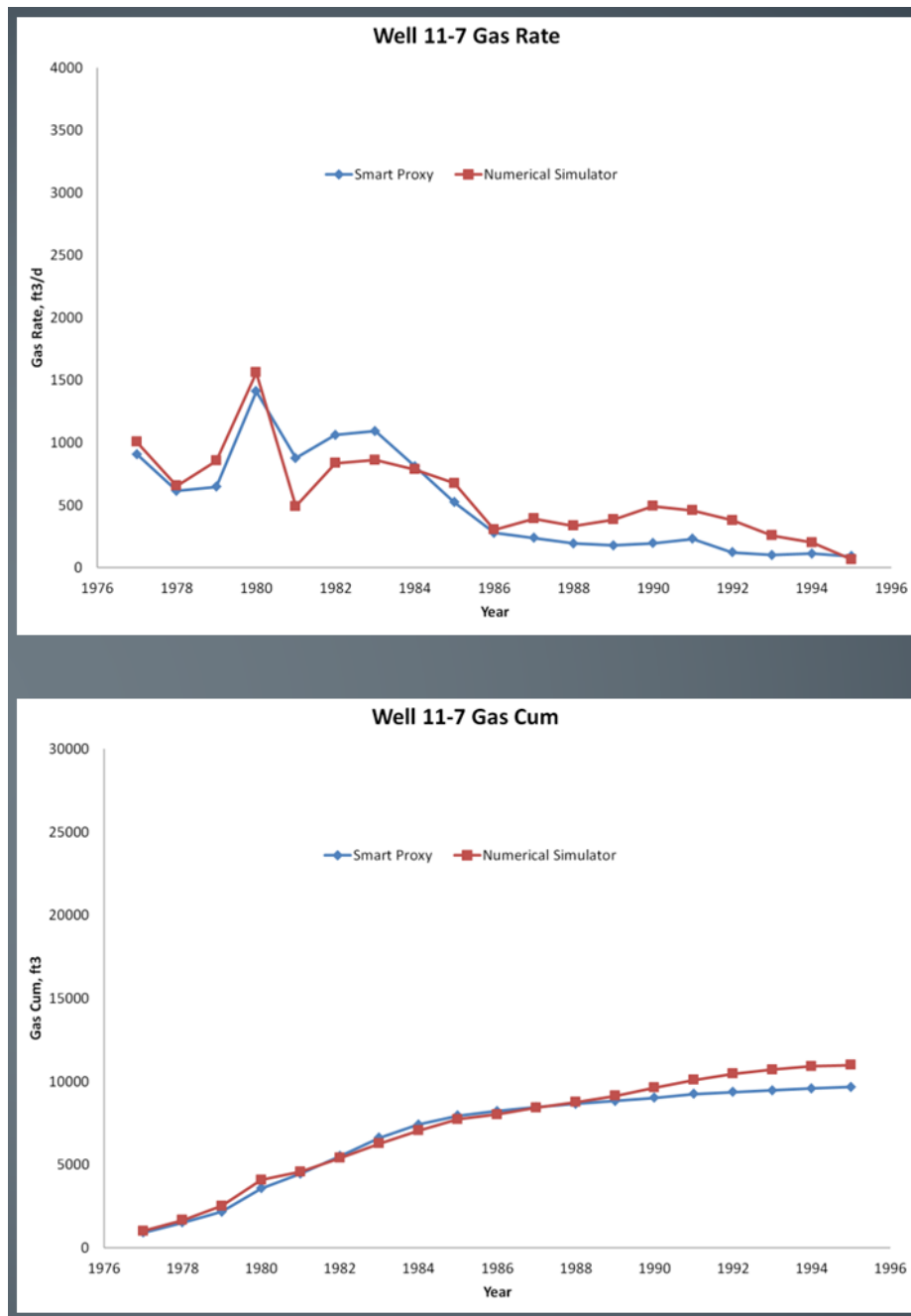
Figure 8.36: Phase three well 11-7 gas production profile
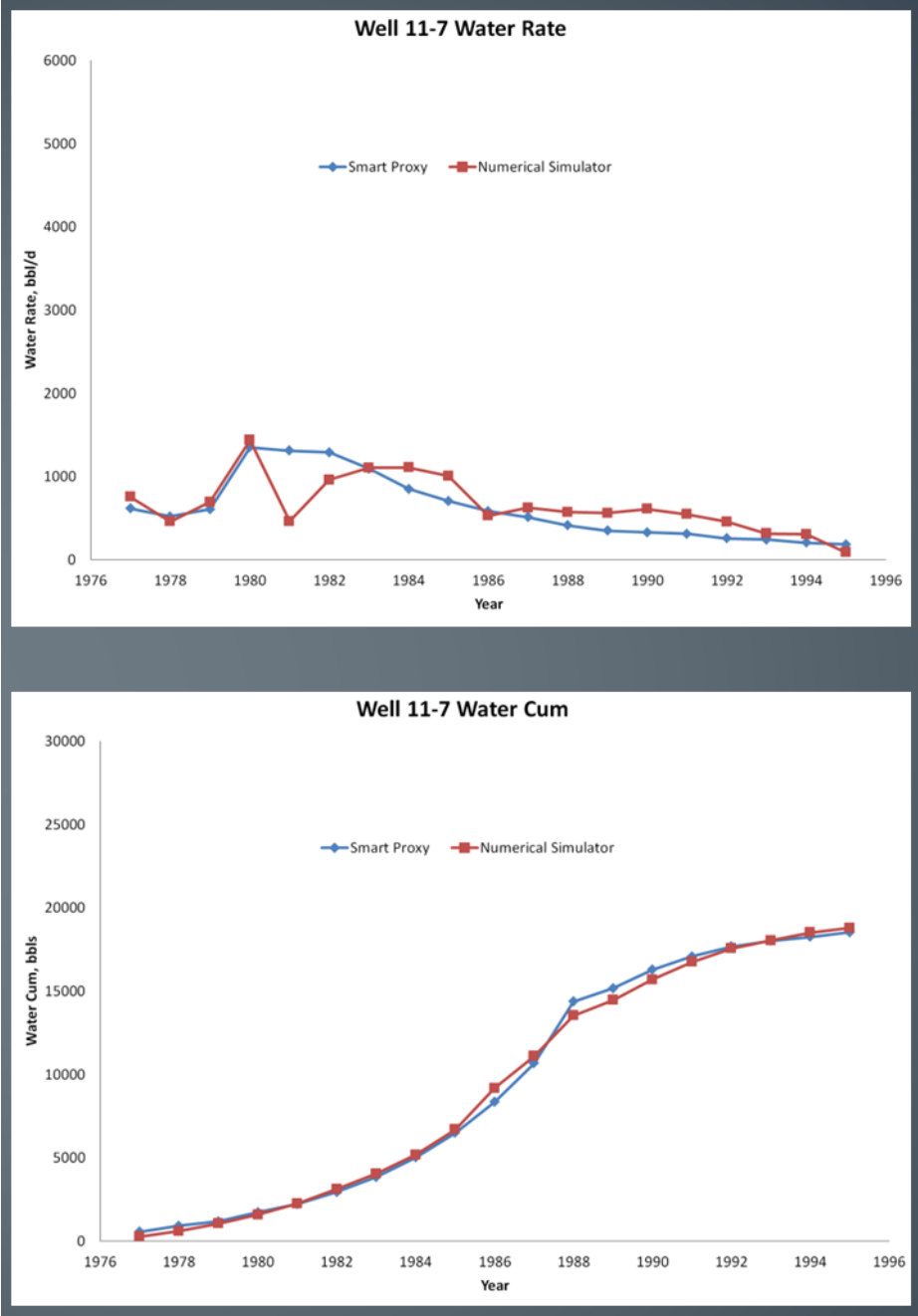
Figure 8.37: Phase three well 11-7 water production profile

Figure 8.38: Phase three well 11-9 oil production profile

Figure 8.39: Phase three well 11-9 gas production profile
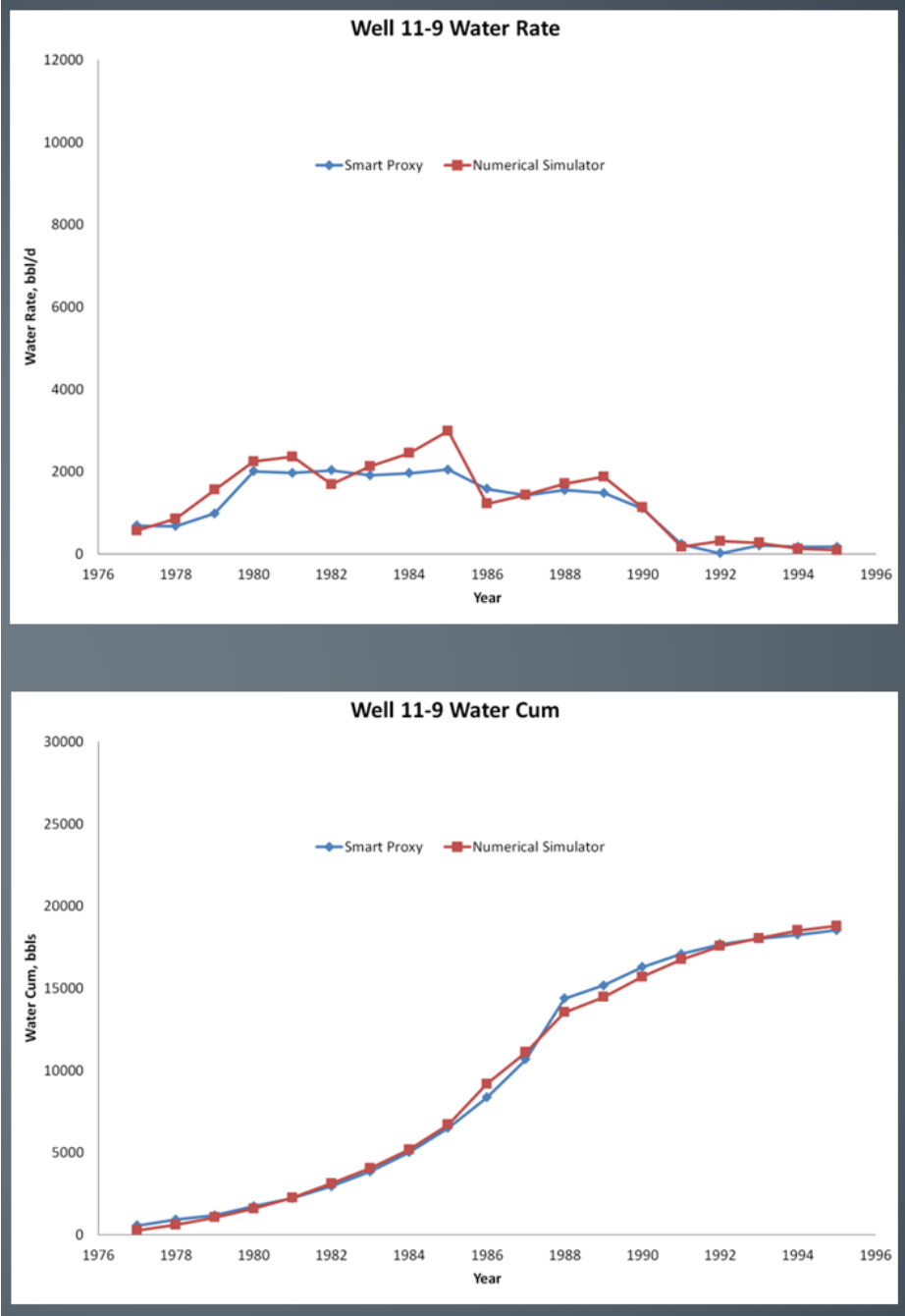
Figure 8.40: Phase three well 11-9 water production profile

Figure 8.41: Phase three well 11-10 oil production profile

Figure 8.42: Phase three well 11-10 gas production profile
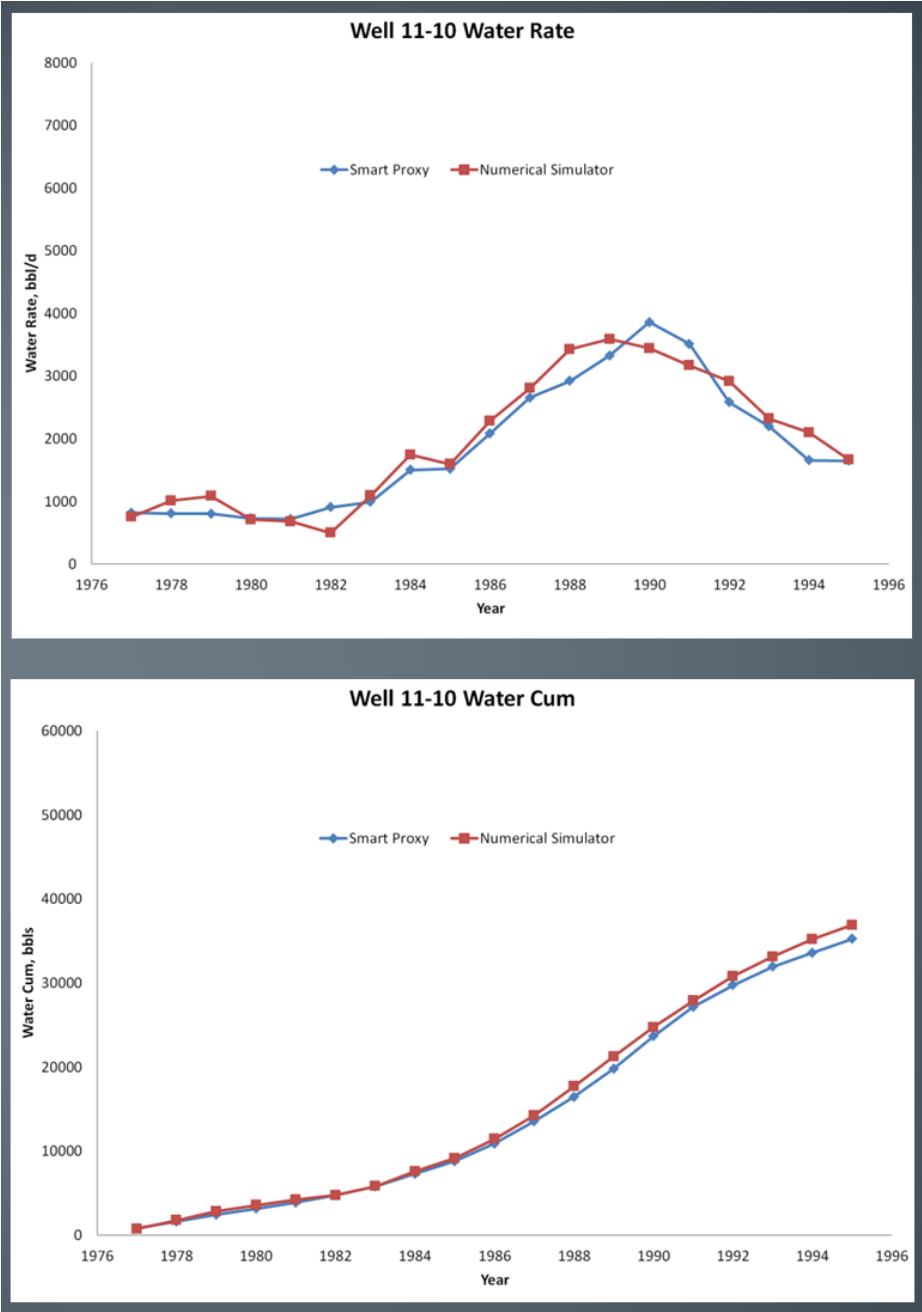
Figure 8.43: Phase three well 11-10 water production profile

Figure 8.44: Phase three well 17-7 oil production profile

Figure 8.45: Phase three well 17-7 gas production profile
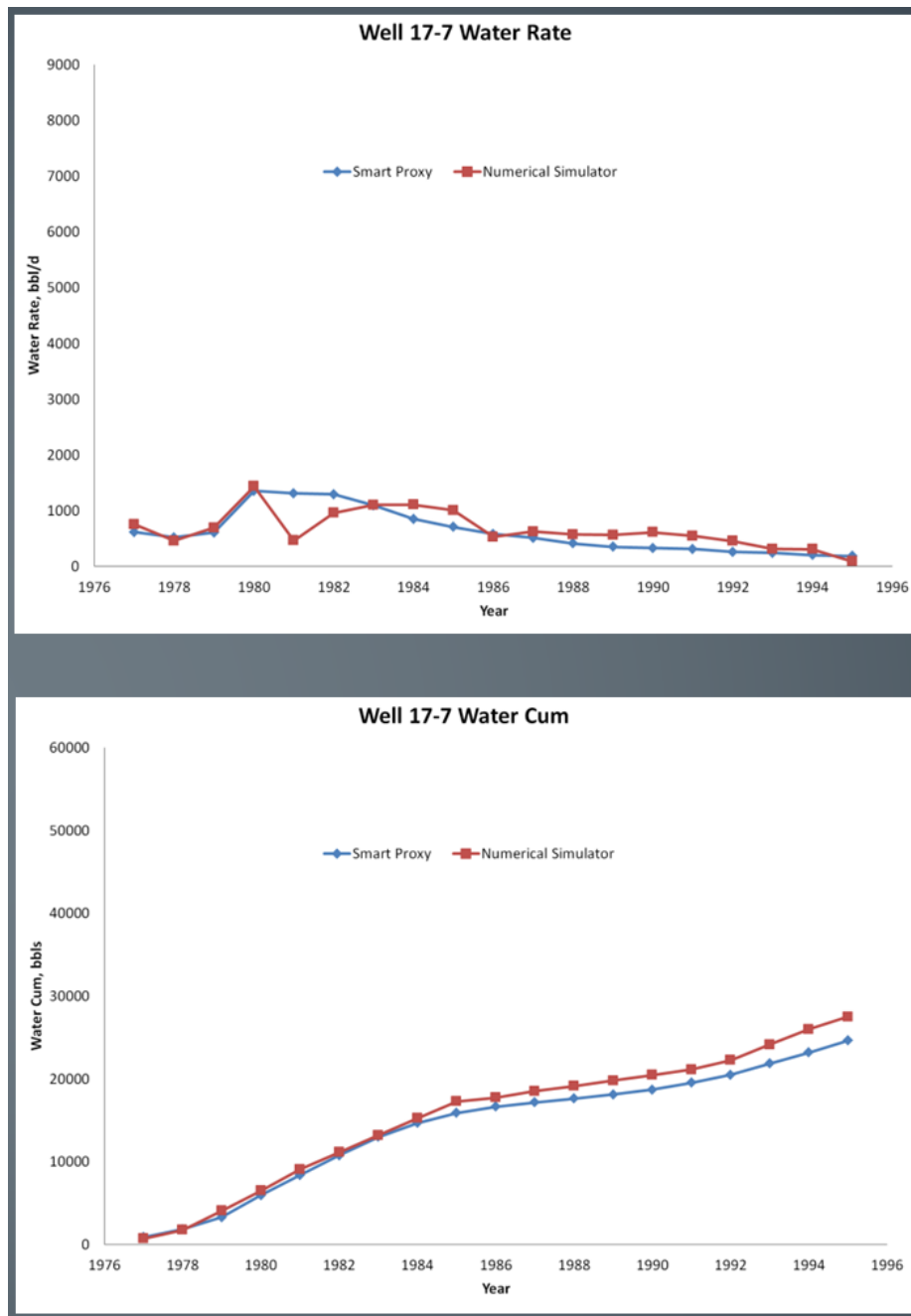
Figure 8.46: Phase three well 17-7 water production profile
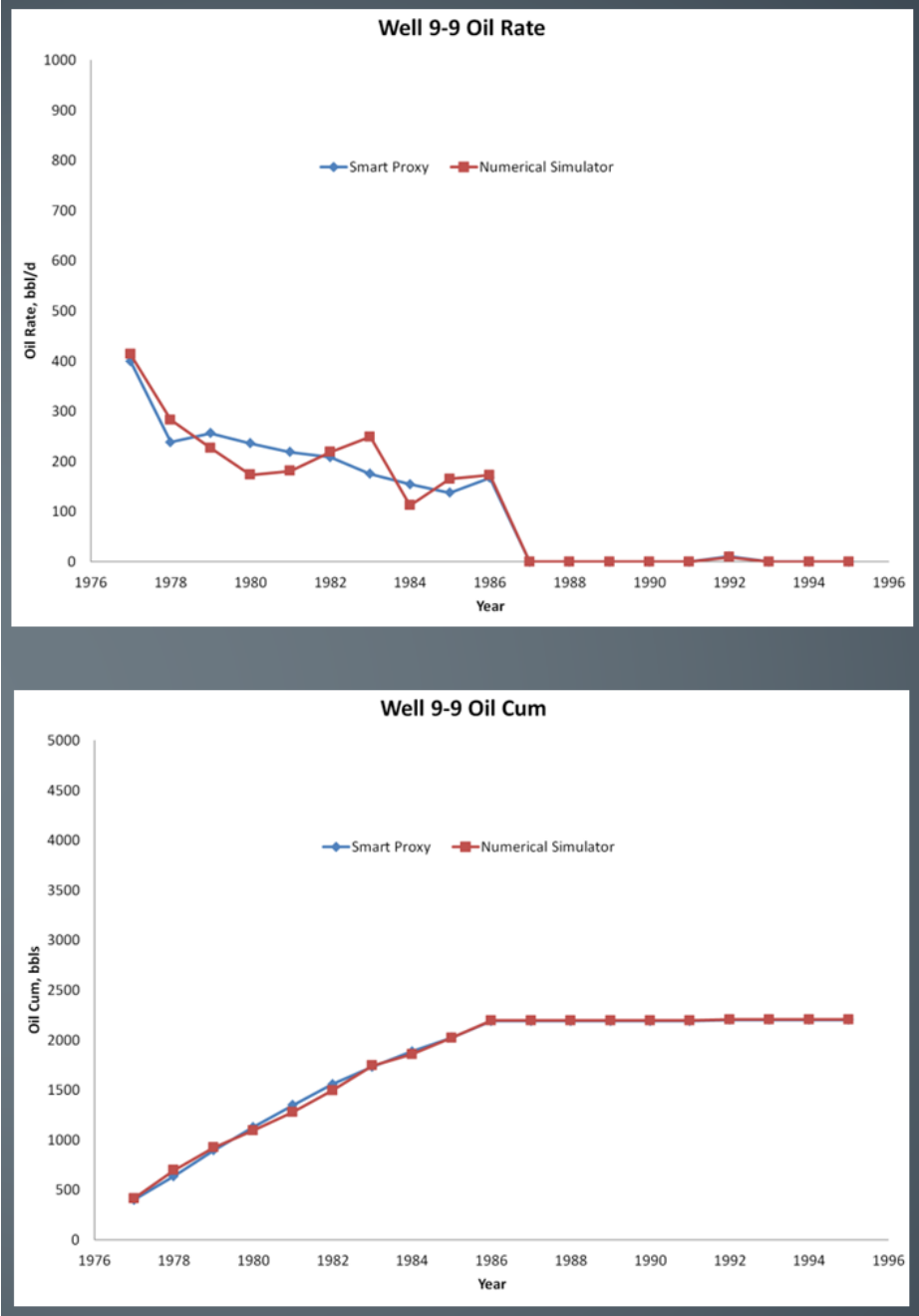
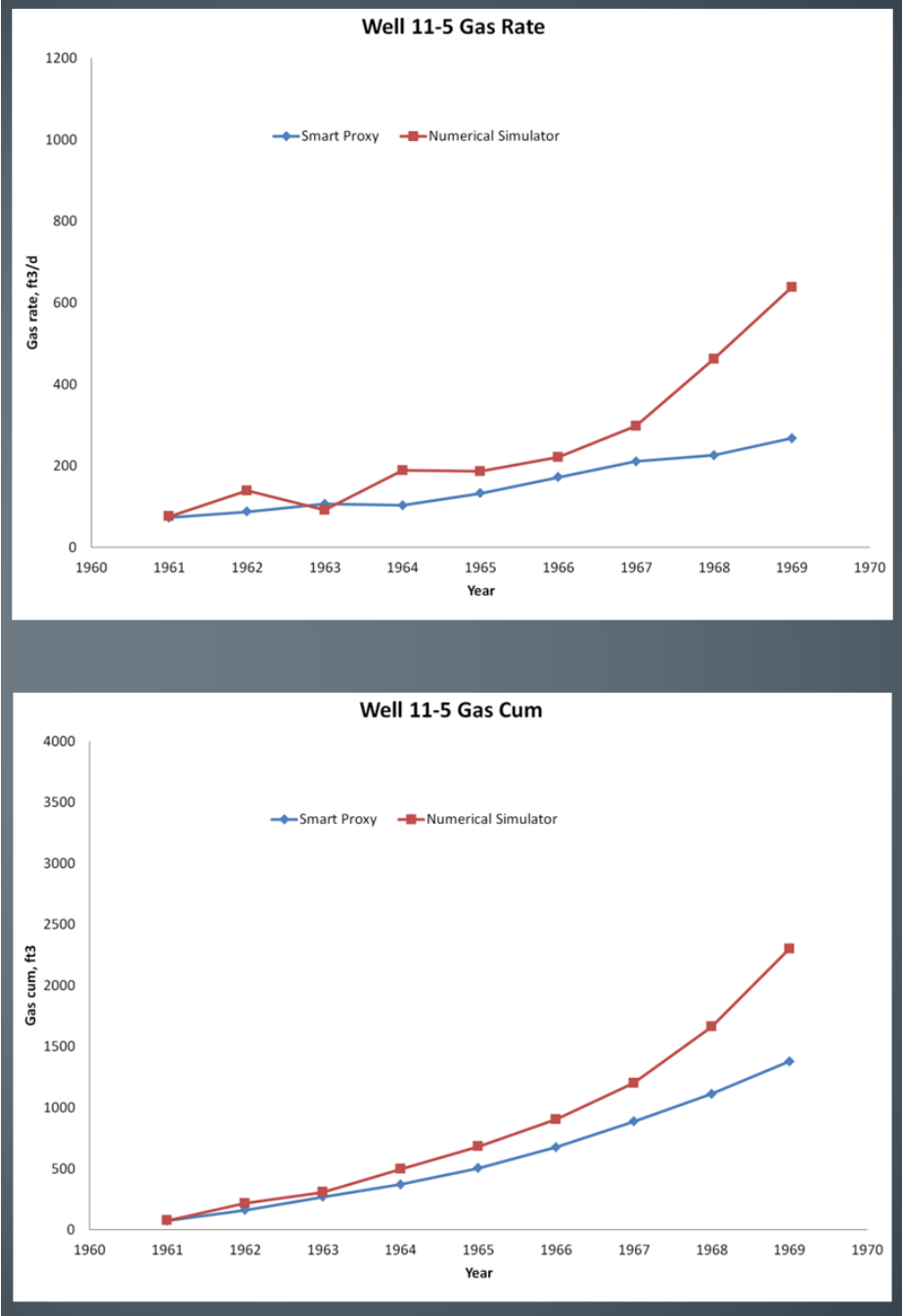Figure 8.47: Phase three well 9-9 oil production profile
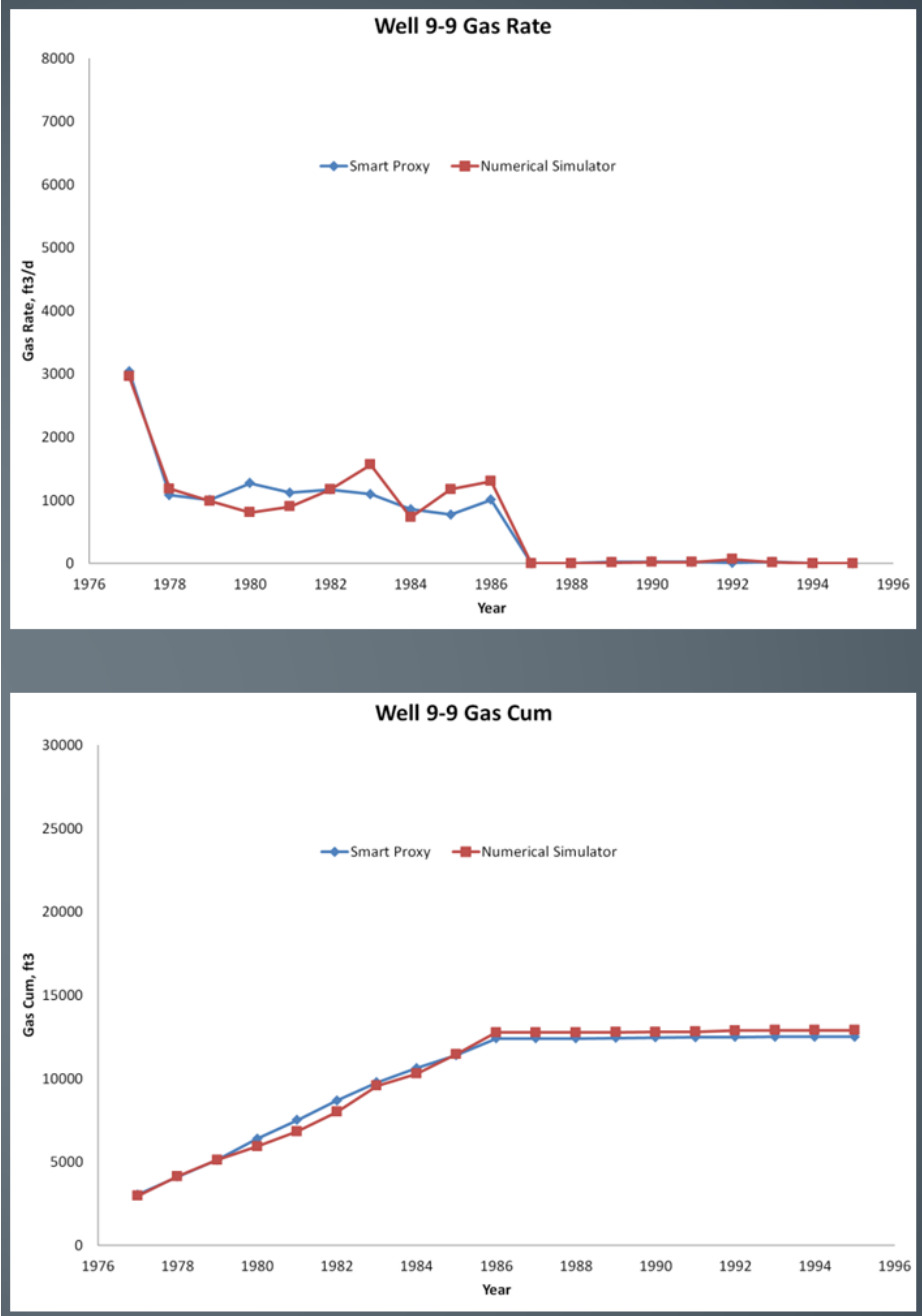
Figure 8.48: Phase three well 9-9 gas production profile
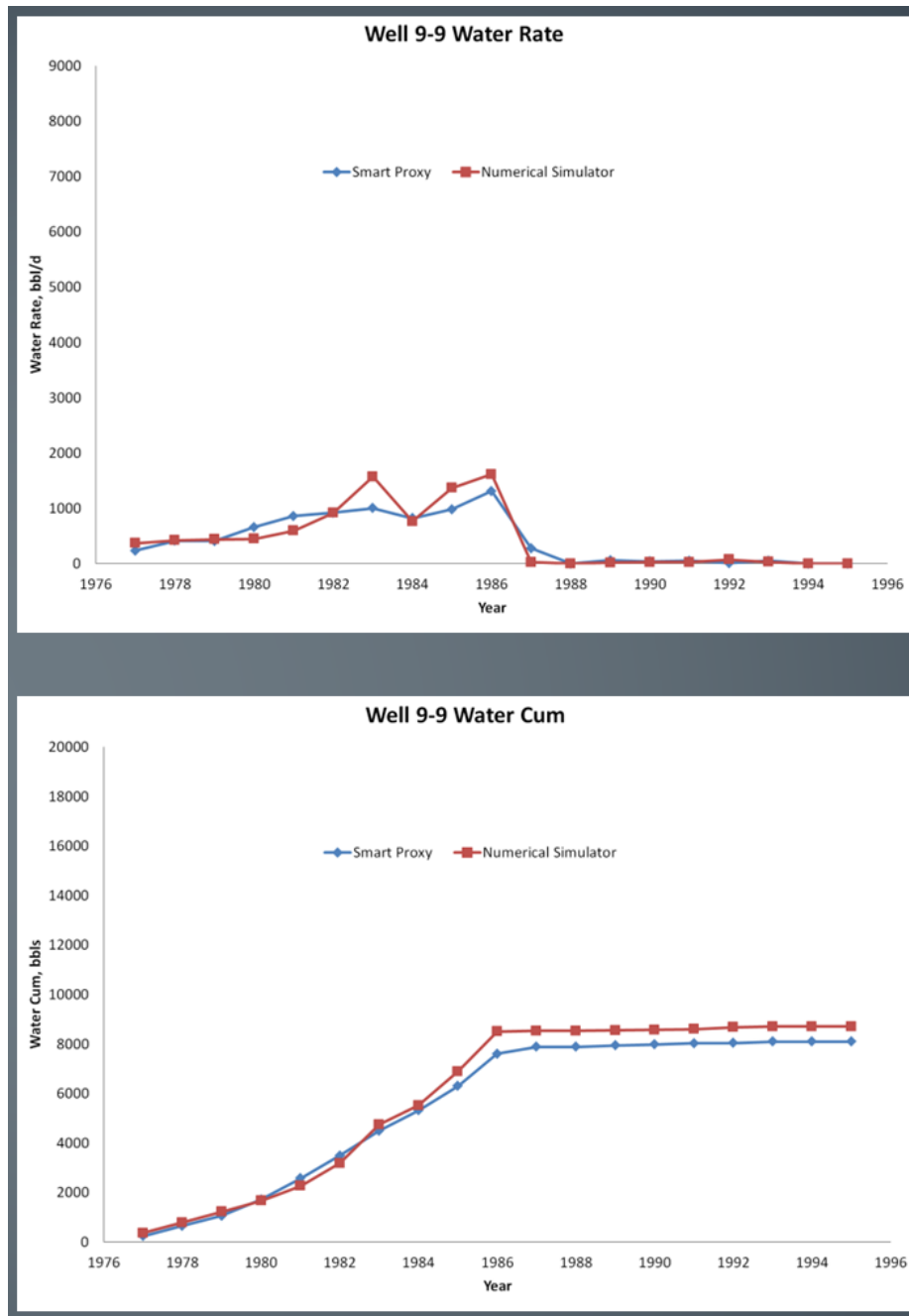
Figure 8.49: Phase three well 9-9 water production profile
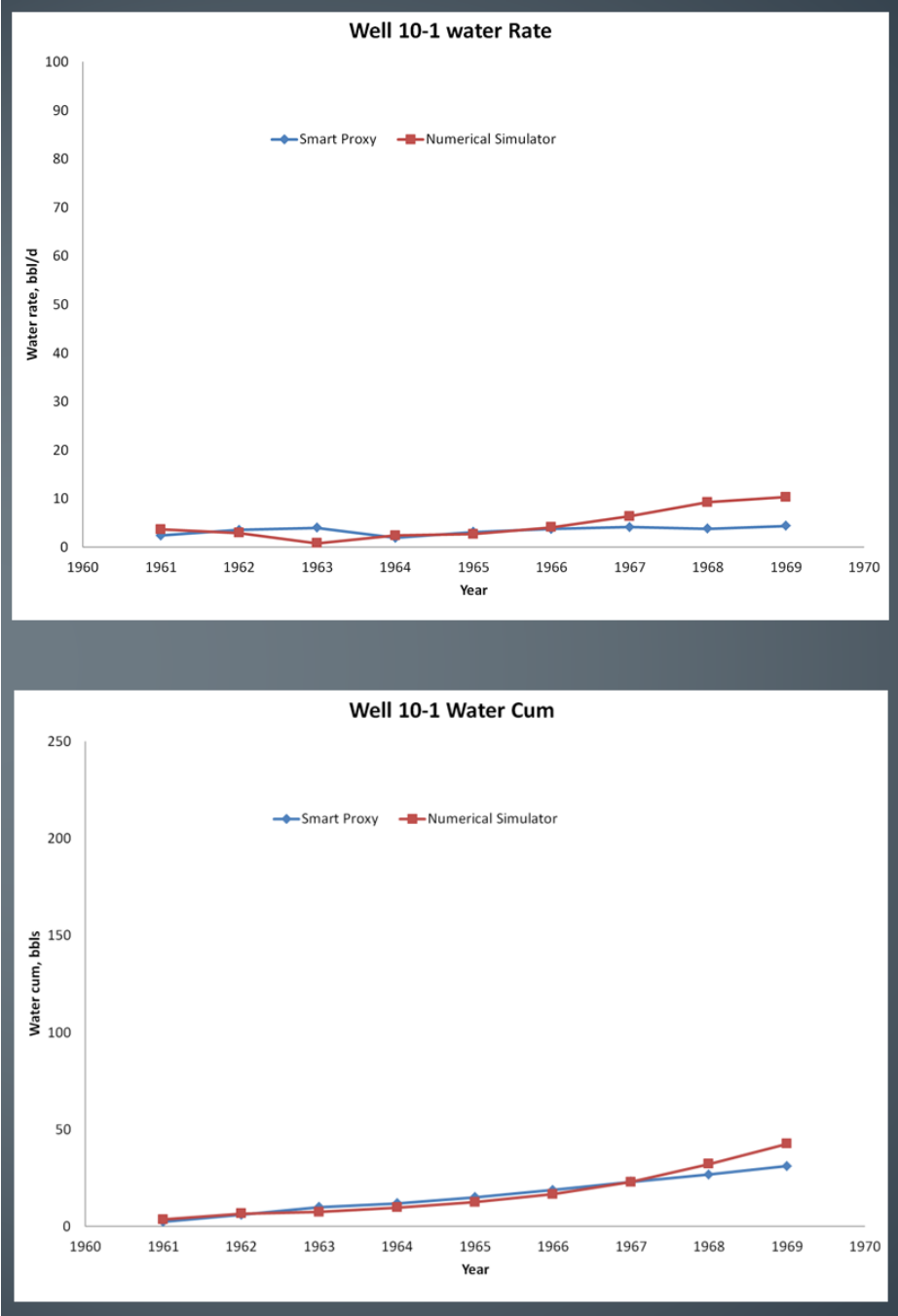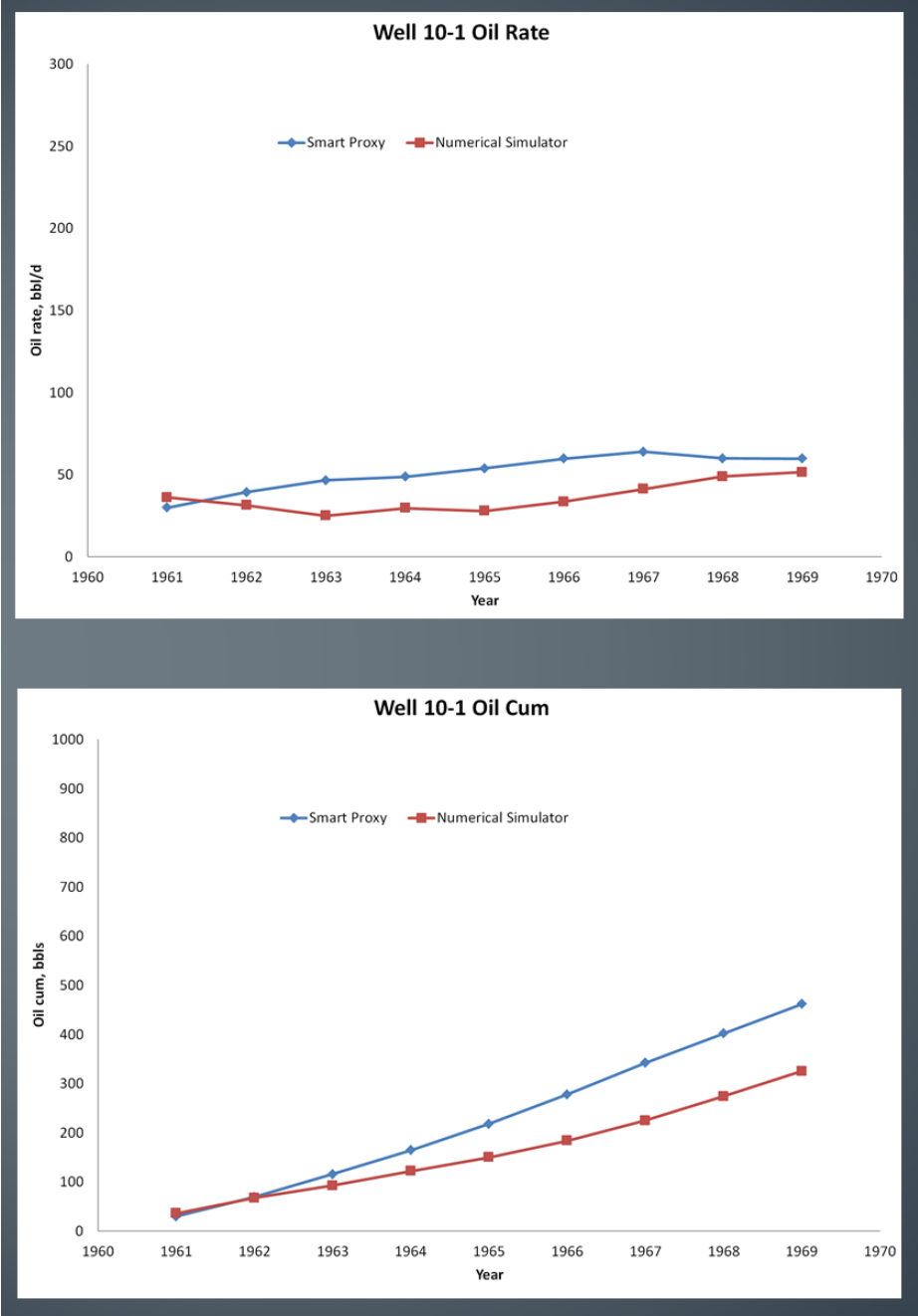
Figure 8.50: Phase three field oil production profile
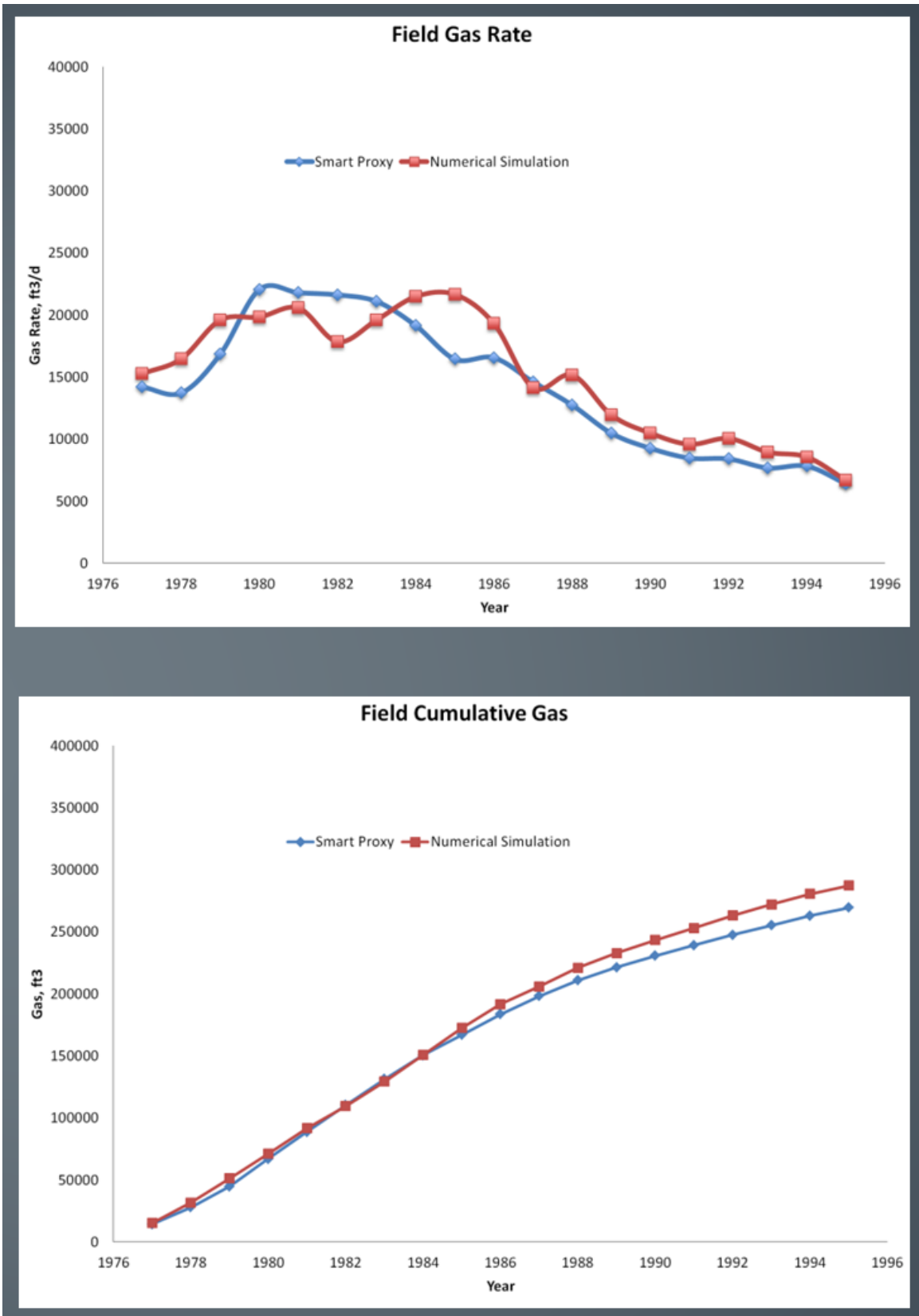
Figure 8.51: Phase three field gas production profile
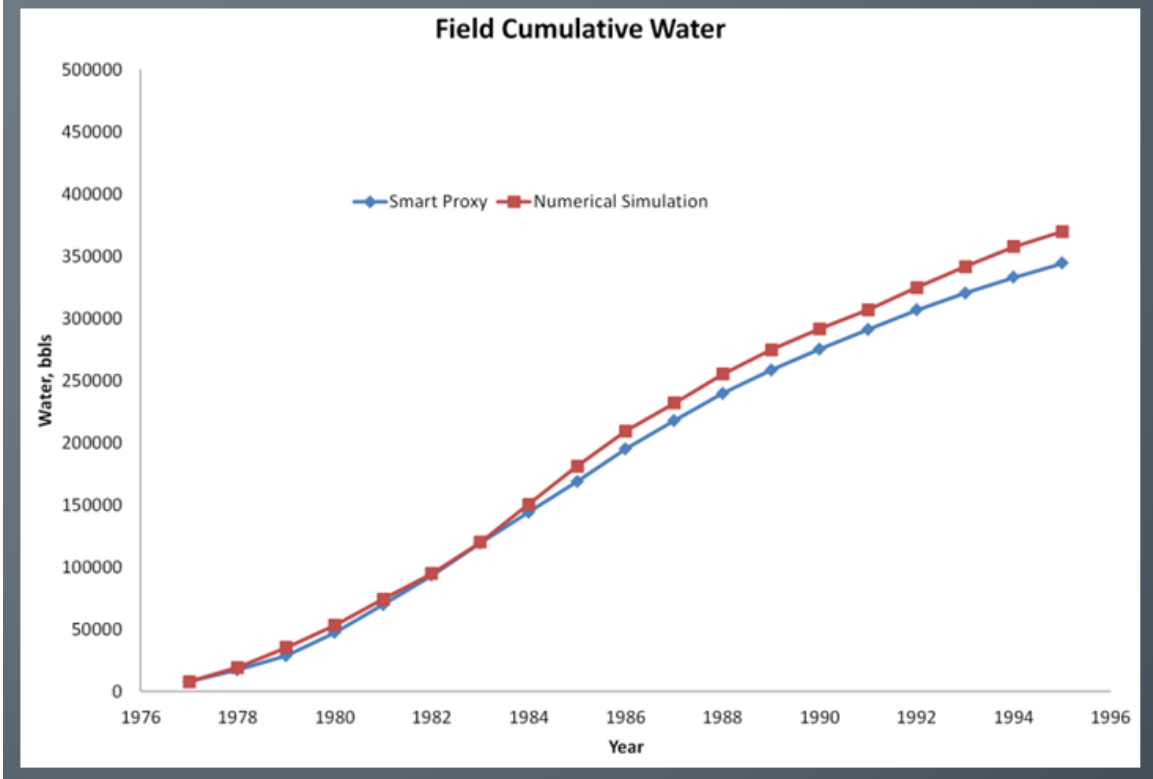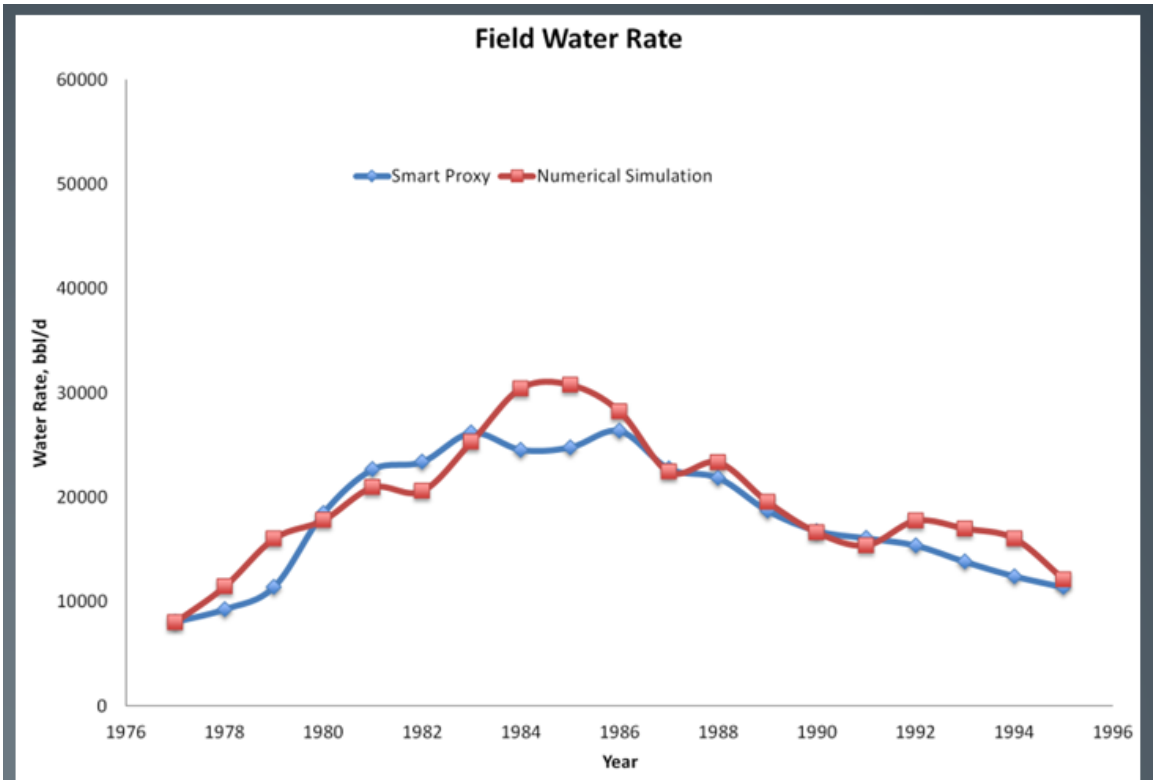
Figure 8.52: Phase three field water production profile

Chapter 9

Concluding Remarks and Recommendations

9.1   Concluding Remarks

In the oil and gas industry, as in any other industry, the ultimate goal is to increase the revenues and decrease the operational costs. Therefore, the industry strives to find new technologies to help achieve this goal. Reservoir simulation helps oil and gas companies plan and set future strategies for their fields. Reservoir simulation can be used in many tasks, such as history matching, evaluation, optimization, and other necessary field-development tasks . With the new geological description tools, geological models can be significantly complex, which greatly increases the computational size of these models. This of course increases the costs associated with running complex models.

To reduce the cost of reservoir simulation and increase the rate of investment, a smart proxy model was introduced. As defined earlier in this dissertation, smart proxy modeling is a technique that is based on data mining and artificial intelligence. Many recent studies conclude that this type of proxy model can be utilized as an alternative to traditional reservoir-simulation techniques. Smart proxy models can perform reservoir simulation with very low computational cost compared to conventional reservoir simulators.

All new technologies take some time to be accepted by the industry and require comprehensive studies and applications to be welcomed. Upon the announcement of the smart proxy model, there

was resistance from the petroleum industry. This can be expected for two main reasons. First, industry reservoir engineers are used to the traditional reservoir techniques, and it is hard for them to shift and change their habits. Second, in some cases, the new technology was misused and applied without considering its limitations. It is very important to understand the limitations and conditions of smart proxy applications. First and foremost, this technology is an objective-oriented technology. In other words, the smart proxy model that is built for CO2 injection cannot be used in a water-injection simulation scenario. Applying the right smart proxy model for the right simulation scenario is key to convincing the industry to make this paradigm shift in reservoir simulation.

Two types of smart proxy model are developed in this study for the SACROC unit. This field has been producing for a long time with deferent production-drive mechanisms. The complexity of the production performance and the geological characterization make it an ideal candidate for examining the smart proxy performance. The first type of smart proxy model was built to generate reservoir properties (pressure and saturation) at grid block level under different geological (porosity and permeability) and operational scenarios (BHP). To gain a larger picture of field performance, another proxy model was developed to generate the well-production profile.

Because the numerical reservoir simulator is the technique most widely accepted and used for reservoir simulation, the results of the smart proxy models were compared to the numerical simulator under the same scenarios to determine the smart proxy performance. Both smart proxy models yielded very acceptable absolute average errors compared to the numerical simulator. The development of the smart proxy model is explained comprehensively in this study. The results used for comparison came from a totally blind simulation run and were not used in the training and developing of the proxy model. They demonstrate the ability of the smart proxy to generate accurate reservoir simulation outputs under any designed simulation scenario within the range of the trained geological and operational parameters.

The development of the smart proxy is based on the solution objective, so the development technique is based on the type of problem in hand. This study tests two types of training and validation techniques. The non-cascading method was using the inputs at each time step from the numerical simulator. Although this method yields an acceptable range of error, it is not practical for reservoir simulation forecasting. Therefore, this study was mainly using the cascading feature. In the cascading feature, the inputs at each time step come from the output of the previous time step until

the last time step is reached (except for the first time step). This feature allows the smart proxy model to feed itself with selected dynamic data inputs. The real challenge of using cascading is error propagation. In non-cascading, the error comes mainly from the neural-network training process. On other hand, the error in the cascading method comes mainly from two sources. The first source is the error generated from the neural-network training; the other source is the error generated from the deployment of the previous time step of the smart proxy model. The error from the cascading process can be significant if the neural network is not trained properly especially with a large number of time steps. The error propagation of the pressure and oil saturation for this study is illustrated in figure 9.1 .

The main purpose of the study is to draw attention to the time and computational size savings that can be achieved using the smart-proxy-model technique. In addition to simulation output accuracy, the smart proxy model is competitive with the conventional simulation techniques because of its low computational size and faster running time. To run a smart proxy model, no expensive processing computers are needed; it can be run using a regular personal computer. The smart proxy database can be very large, but it can be sampled to reduce the size if necessary.

It has been found that this technique constitutes an effective way to improve the speed of the simulation run. In most cases, smart proxy simulations can be completed in fractions of a second. To illustrate, the run time of one simulation scenario in this study takes seven hours with the convectional numerical simulator. The smart proxy model using the cascading feature for 30 time steps takes up to 30 seconds. To run 100 simulation scenarios with a conventional numerical simulator could take up to one month. On the other hand, it would take only a couple of minutes using smart proxy.

The question of whether the physics is ignored has recently caused much debate over the use of this technique. In the smart proxy model, by studying all the corresponding reservoir data, the network learns from the data that carries information about the physics of the system. On these grounds, it can be argued that the physics of the system is not disregarded.
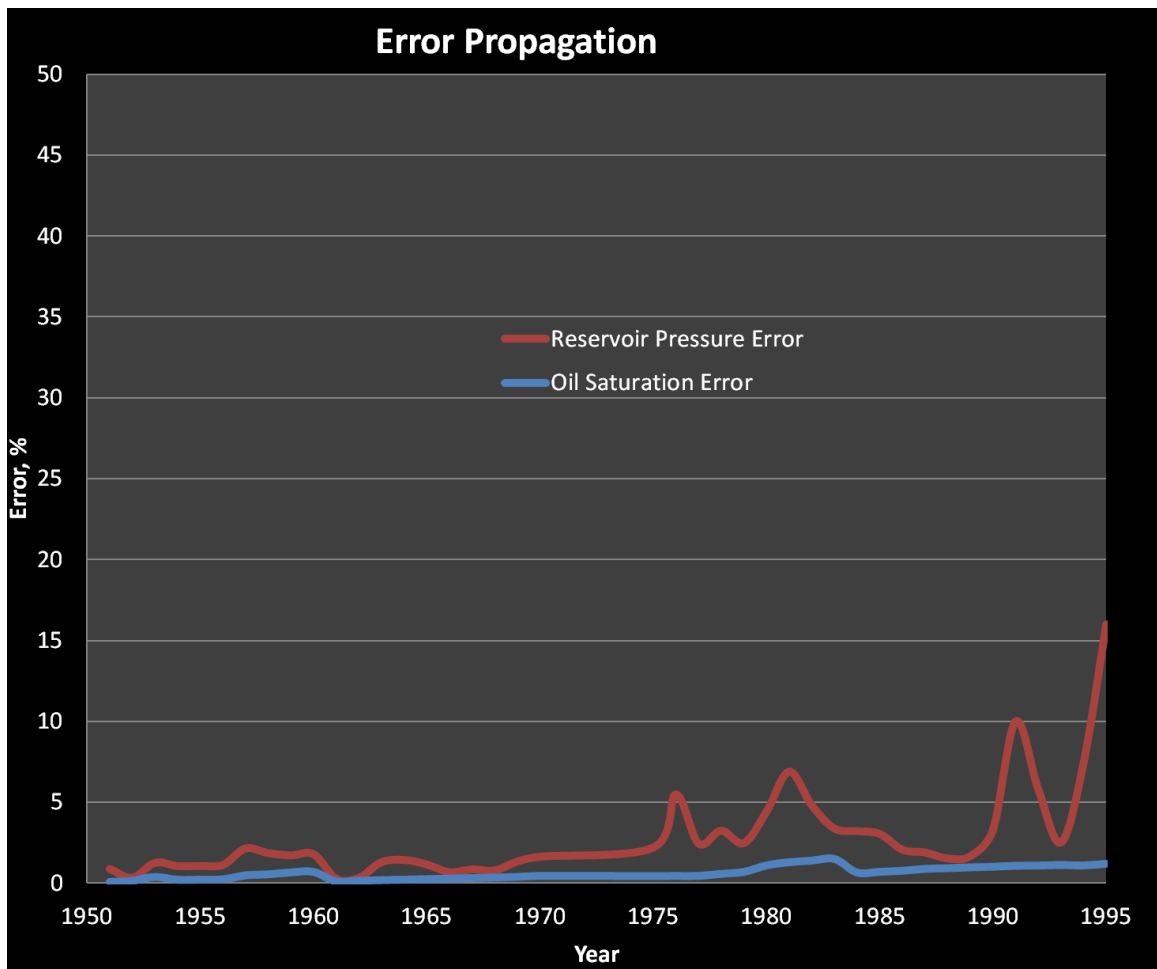
Figure 9.1: Reservoir Pressure and Oil Saturation Error Propagation

9.2 Recommendations

The following recommendations are offered for related research in the field of smart proxy development for reservoir simulation.

1. To build a smart proxy for reservoir simulation applications, the number of time steps and the time resolution should be chosen carefully. Time resolution is a key feature for planning how much data needs to be collected out of the field performance. A finer time resolution means more information can be gathered in order to understand the fields behavior. In this study, an annual time resolution was selected. It is highly recommended to consider obtaining a monthly time resolution in order to monitor the reservoir pressure and saturation along with their effect on the production performance on a month-by-month basis. In this way, the smart proxy would be able to discover any missing information during the production year. .

2. Designing numerical simulation runs plays a significant role in smart proxy development. It helps to train the ANN with different field scenarios. For the operational scenarios in this study, only the BHP was used to design numerical simulation runs. It is recommended to include multiple injection scenarios along with BHP scenarios. In this way, the smart proxy would be able to obtain the optimum production/injection strategy of the field.

3. To make the smart proxy more practical in a reservoir simulation, it is recommended to investigate the feasibility of developing a smart proxy that is able to produce more than one reservoir property. In this study, one smart proxy was developed for each reservoir property (pressure and saturation). It would be very useful to build one smart proxy that can supply both reservoir properties.

Bibliography

[1] Z Chen. Reservoir simulation: mathematical techniques in oil recovery 2007. *SIAM, USA*.

[2] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2014.

[3] Tom M Mitchell. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45(37):870–877, 1997.

[4] Pedro Domingos. A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78–87, 2012.

[5] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[6] Tsinghua University. *Machine Learning Cheat Sheet*. Soul Machine, 2014.

[7] Maureen Caudill. Neural nets primer, part vi. *AI Expert*, 4(2):61–67, 1989.

[8] Dave Anderson and George McNeill. Artificial neural networks technology. *Kaman Sciences Corporation*, 258(6):1–83, 1992.

[9] Vincent Cheung and Kevin Cannons. An introduction to neural networks. *Signal & Data Compression Laboratory, Electrical & Computer Engineering University of Manitoba, Winnipeg, Manitoba, Canada*, 2002.

[10] Andrej Krenker, Andrej Kos, and Janez Bešter. *Introduction to the artificial neural networks*. INTECH Open Access Publisher, 2011.

[11] Howard B Demuth, Mark H Beale, Orlando De Jess, and Martin T Hagan. *Neural network design*. Martin Hagan, 2014.

[12] Saedsayad. Artificial neural network, 2016.

[13] Charu C Aggarwal. *Data mining: the textbook*. Springer, 2015.

[14] Michael JA Berry and Gordon Linoff. Mastering data mining, 2000.

[15] Shahab Mohaghegh et al. Virtual-intelligence applications in petroleum engineering: Part 1artificial neural networks. *Journal of Petroleum Technology*, 52(09):64–73, 2000.

[16] Shahab Mohaghegh et al. Virtual-intelligence applications in petroleum engineering: Part 2evolutionary computing. *Journal of Petroleum Technology*, 52(10):40–46, 2000.

[17] Shahab Mohaghegh et al. Virtual-intelligence applications in petroleum engineering: Part 3fuzzy logic. *Journal of petroleum technology*, 52(11):82–87, 2000.

[18] JP Brown et al. A machine learning approach to studies of recovery efficiency. In *Petroleum Computer Conference*. Society of Petroleum Engineers, 1991.

[19] Lloyd Heinze, Islam Ali Al-Baiyat, et al. Implementing artificial neural networks and support vector machines in stuck pipe prediction. In *SPE Kuwait International Petroleum Conference and Exhibition*. Society of Petroleum Engineers, 2012.

[20] Fatai A Anifowose et al. Ensemble machine learning: The latest development in computational intelligence for petroleum reservoir characterization. In *SPE Saudi Arabia Section Technical Symposium and Exhibition*. Society of Petroleum Engineers, 2013.

[21] Niranjan Subrahmanya, Peng Xu, Amr El-Bakry, Carmon Reynolds, et al. Advanced machine learning methods for production data pattern recognition. In *SPE Intelligent Energy Conference & Exhibition*. Society of Petroleum Engineers, 2014.

[22] Shahab D Mohaghegh et al. Quantifying uncertainties associated with reservoir simulation studies using a surrogate reservoir model. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2006.

[23] Chuan Tian, Roland N Horne, et al. Applying machine learning techniques to interpret flow rate, pressure and temperature data from permanent downhole gauges. In *SPE Western Regional Meeting*. Society of Petroleum Engineers, 2015.

[24] Arash Shadravan, Mohammadali Tarrahi, Mahmood Amani, et al. Intelligent tool to design fracturing, drilling, spacer and cement slurry fluids using machine learning algorithms. In *SPE Kuwait Oil and Gas Show and Conference*. Society of Petroleum Engineers, 2015.

[25] Arash Shadravan, Mohammadali Tarrahi, et al. Machine learning leads cost effective intelligent fluid design: Fluid engineering perspective. In *SPE Bergen One Day Seminar*. Society of Petroleum Engineers, 2016.

[26] Q Cao, R Banerjee, S Gupta, J Li, W Zhou, B Jeyachandra, et al. Data driven production forecasting using machine learning. In *SPE Argentina Exploration and Production of Unconventional Resources Symposium*. Society of Petroleum Engineers, 2016.

[27] Shahab Mohaghegh. Smart proxy modeling for numerical reservoir simulations - big data analytic in e and p, 2015.

[28] Arash Mirzabozorg. *Incorporation of Engineering Knowledge in History Matching, Optimization, and Uncertainty Assessment Frameworks with Application to the SAGD Process*. PhD thesis, University of Calgary, 2015.

[29] S Amini, SD Mohaghegh, R Gaskari, GS Bromhal, et al. Pattern recognition and data-driven analytics for fast and accurate replication of complex numerical reservoir models at the grid block level. In *SPE Intelligent Energy Conference & Exhibition*. Society of Petroleum Engineers, 2014.

[30] George EP Box, Norman Richard Draper, et al. *Empirical model-building and response surfaces*, volume 424. Wiley New York, 1987.

[31] André I Khuri and Siuli Mukhopadhyay. Response surface methodology. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2):128–149, 2010.

[32] Burak Yeten, Alexandre Castellini, Baris Guyaguler, WH Chen, et al. A comparison study on experimental design and response surface methodologies. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2005.

[33] Nuran Bradley. *The response surface methodology*. PhD thesis, Indiana University South Bend, 2007.

[34] P Vanegas, JC Cunha, LB Cunha, et al. Uncertainty assessment of production performance for a heavy oil offshore field by using the experimental design technique. In *Canadian International Petroleum Conference*. Petroleum Society of Canada, 2006.

[35] F Ahmadloo, K Asghari, M Masehi Araghi, et al. Heavy oil viscosity prediction using surface response methodology. In *Canadian International Petroleum Conference*. Petroleum Society of Canada, 2009.

[36] JP Dejean, G Blanc, et al. Managing uncertainties on production predictions using integrated statistical methods. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 1999.

[37] Suryansh Purwar, Christopher J Jablonowski, Quoc Phuc Nguyen, et al. A method for integrating response surfaces into optimization models with real options: A case in gas flooding. In *SPE Hydrocarbon Economics and Evaluation Symposium*. Society of Petroleum Engineers, 2010.

[38] Fernanda Vaz Alves Risso, FF Risso, Denis José Schiozer, et al. Risk assessment of oil fields using proxy models: a case study. *Journal of Canadian Petroleum Technology*, 47(08), 2008.

[39] Genilson Antnio Da Silva, Andre M Oliveira Guimaraes, Gislene Aparecida da Silva, Marcos Aurelio Lucas, et al. Integrated risk analysis of uncertainties in geological model and dynamic properties flow for the development of oil fields. In *SPE Latin America and Caribbean Petroleum Engineering Conference*. Society of Petroleum Engineers, 2012.

[40] Kurt Wilson, Louis J Durlofsky, et al. Computational optimization of shale resource development using reduced-physics surrogate models. In *SPE Western Regional Meeting*. Society of Petroleum Engineers, 2012.

[41] T Heijn, R Markovinovic, JD Jansen, et al. Generation of low-order reservoir models using system-theoretical concepts. In *SPE Reservoir Simulation Symposium*. Society of Petroleum Engineers, 2003.

[42] Terry Ngo. Data mining: practical machine learning tools and technique, by ian h. witten, eibe frank, mark a. hell. *ACM Sigsoft Software Engineering Notes*, 36(5):51–52, 2011.

[43] D Solomatine, LM See, and RJ Abrahart. Data-driven modelling: concepts, approaches and experiences. In *Practical hydroinformatics*, pages 17–30. Springer, 2009.

[44] Shahab D Mohaghegh, Abi Modavi, Hafez Hafez, Masoud Haajizadeh, and Srikant Guruswamy. Development of surrogate reservoir model (srm) for fast track analysis of a complex reservoir. *International Journal of Oil, Gas and Coal Technology*, 2(1):2–23, 2009.

[45] Jalal Jalali, Shahab D Mohaghegh, et al. Reservoir simulation and uncertainty analysis of enhanced cbm production using artificial neural networks. In *SPE Eastern Regional Meeting*. Society of Petroleum Engineers, 2009.

[46] Amirmasoud Kalantari Dahaghi, Shahab D Mohaghegh, et al. Numerical simulation and multiple realizations for sensitivity study of shale gas reservoirs. In *SPE Production and Operations Symposium*. Society of Petroleum Engineers, 2011.

[47] Shahab D Mohaghegh, Shohreh Amini, Vida Gholami, Razi Gaskari, Grant S Bromhal, et al. Grid-based surrogate reservoir modeling (srm) for fast track analysis of numerical reservoir simulation models at the gridblock level. In *SPE Western Regional Meeting*. Society of Petroleum Engineers, 2012.

[48] Vida Gholami. *On the optimization of CO2-EOR process using surrogate reservoir model*. WEST VIRGINIA UNIVERSITY, 2014.

[49] Soodabeh Esmaili. *Production History Matching and Forecasting of Shale Assets Using Pattern Recognition*. West Virginia University, 2013.

[50] Alireza Shahkarami, Shahab D Mohaghegh, Vida Gholami, and Sayed Alireza Haghighat. Spe-169507-ms. 2014.

[51] Qin He, Shahab D Mohaghegh, Zhikun Liu, et al. Reservoir simulation using smart proxy in sacroc unit-case study. In *SPE Eastern Regional Meeting*. Society of Petroleum Engineers, 2016.

[52] S Alireza Haghighat, Shahab D Mohaghegh, et al. Interpretation of real-time pressure measurements to detect co 2 leakage. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2015.

[53] Alireza Shahkarami, Shahab D Mohaghegh, Yasin Hajizadeh, et al. Assisted history matching using pattern recognition technology. In *SPE Digital Energy Conference and Exhibition*. Society of Petroleum Engineers, 2015.

[54] KNUT-ANDREAS LIE and BRADLEY T MALLISON. Mathematical models for oil reservoir simulation. *Encyclopedia of Applied and Computational Mathematics*, pages 1–8, 2013.

[55] A Spivak, TL Perryman, RA Norris, et al. Pd 13 (1) a compositional simulation study of the sacroc unit co2 project. In *9th World Petroleum Congress*. World Petroleum Congress, 1975.

[56] TD Crameik, JA Plassey, et al. Carbon dioxide injection project sacroc unit, scurry county, texas. In *Annual Meeting Papers, Division of Production*. American Petroleum Institute, 1972.

[57] EL Vest Jr. Oil fields of pennsylvanian-permian horseshoe atoll, west texas. 1970.

[58] Weon Shik Han, Brian J McPherson, Peter C Lichtner, and Fred P Wang. Evaluation of trapping mechanisms in geologic co2 sequestration: Case study of sacroc northern platform, a 35-year co2 injection site. *American Journal of Science*, 310(4):282–324, 2010.

[59] JT Hawkins, AJ Benvegnu, TP Wingate, JD McKamie, CD Pickard, JT Altum, et al. Sacroc unit co2 flood: Multidisciplinary team improves reservoir management and decreases operating costs. *SPE Reservoir Engineering*, 11(03):141–148, 1996.

[60] J William Carey, Marcus Wigand, Steve J Chipera, Giday WoldeGabriel, Rajesh Pawar, Peter C Lichtner, Scott C Wehner, Michael A Raines, and George D Guthrie. Analysis and performance of oil well cement with 30 years of co 2 exposure from the sacroc unit, west texas, usa. *International Journal of Greenhouse Gas Control*, 1(1):75–85, 2007.

[61] Batur Isdiken. Integrated geological and petrophysical investigation on carbonate rocks of the middle early to late early canyon high frequency sequence in the northern platform area of the sacroc unit. 2013.

[62] WM Brummett Jr, AS Emanuel, JD Ronquille, et al. Reservoir description by simulation at sacroc-a case history. *Journal of Petroleum Technology*, 28(10):1–241, 1976.

[63] HH Allen, CR LaRue, et al. Sacroc unit operations. *Journal of Petroleum Technology*, 9(07):21–25, 1957.

[64] Roy M Dicharry, TL Perryman, JD Ronquille, et al. Evaluation and design of a co2 miscible flood project-sacroc unit, kelly-snyder field. *Journal of Petroleum Technology*, 25(11):1–309, 1973.

[65] Rebecca J Larkin, Prentice George Creel, et al. Methodologies and solutions to remediate inner-well communication problems on the sacroc co2 eor project: A case study. In *SPE Symposium on Improved Oil Recovery*. Society of Petroleum Engineers, 2008.

[66] Qin He. *Investigating Continuously Updated History Matching Using Smart Proxy (Surrogate Reservoir Model)*. PhD thesis, WEST VIRGINIA UNIVERSITY, 2016.

[67] AV Kane et al. Performance review of a large-scale co2-wag enhanced recovery project, sacroc unit kelly-snyder field. *Journal of Petroleum Technology*, 31(02):217–231, 1979.

[68] Martin O Saar. *The relationship between permeability, porosity, and microstructure in vesicular basalts*. PhD thesis, University of Oregon, 1998.

[69] F Jerry Lucia. Rock-fabric/petrophysical classification of carbonate pore space for reservoir characterization. *AAPG bulletin*, 79(9):1275–1300, 1995.

[70] Weon Shik Han. *Evaluation of CO2 trapping mechanisms at the SACROC northern platform: Site of 35 years of CO2 injection.* PhD thesis, Citeseer, 2008.

[71] Louis J Durlofsky. Upscaling of geological models for reservoir simulation: Issues and approaches-introduction to the special issue, 2002.

[72] Michael J King et al. Upgridding and upscaling: current trends and future directions. 2007.

[73] Apostolos Kantzas, Jonathan Bryan, and Saeed Taheri. Fundamentals of fluid flow in porous media. *Pore size distribution*, 2012.

[74] Paul Glover. Formation evaluation msc course notes. *Aberdeen University*, pages 84–94, 2001.

[75] Carlon S Land et al. Calculation of imbibition relative permeability for two-and three-phase flow from rock properties. *Society of Petroleum Engineers Journal*, 8(02):149–156, 1968.

[76] Brant Bennion, Stefan Bachu, et al. Relative permeability characteristics for supercritical co2 displacing water in a variety of potential sequestration zones. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2005.

[77] James R Gilman and Chet Ozgen. *Reservoir simulation: history matching and forecasting.* Soc. of Petroleum Engineers, 2013.

[78] Petya Ivanova Vakova and Muhammad Adeel Nasser Sohal. Reservoir simulation history matching and forecasting.

[79] The Mathworks, Inc., Calgary, Alberta, T2M 3Y7 Canada. *CMG version 2015.1)*, 2015.

[80] Faisal Alenezi and Shahab Mohaghegh. A data-driven smart proxy model for a comprehensive reservoir simulation. Manuscript submitted for publication, 2016.

[81] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques.* Elsevier, 2011.

[82] Weng Marc Lim and Ding Hooi Ting. Research methodology: a toolkit of sampling and data analysis techniques for quantitative research. 2012.

[83] LB Paula and P Justo. Population and sample. sampling techniques. *Management Mathematics for European Schools*, 2001.

[84] Patricia S Crowther and Robert J Cox. A method for optimal division of data sets for use in neural networks. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pages 1–7. Springer, 2005.

[85] RC Chakraborty. Back propagation network, 2010.

[86] Shahab D Mohaghegh. Shale analytics. In *Shale Analytics*, pages 29–81. Springer, 2017.
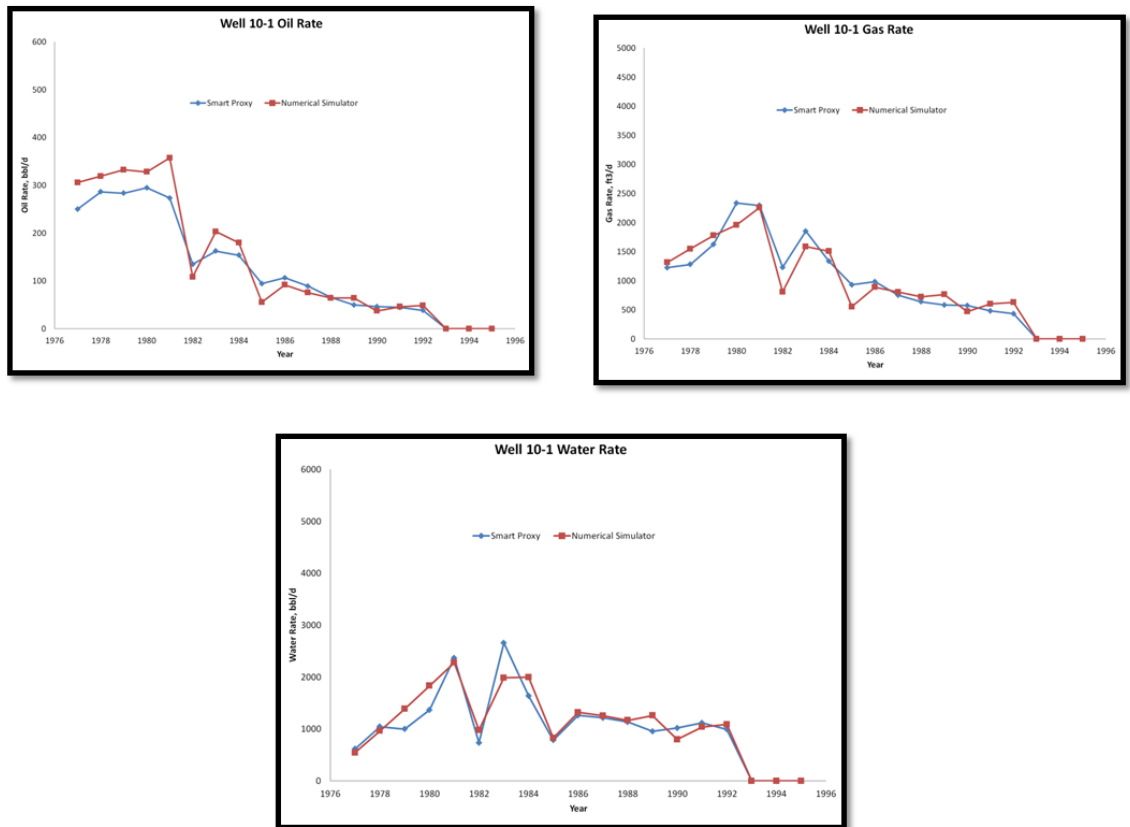
Appendix
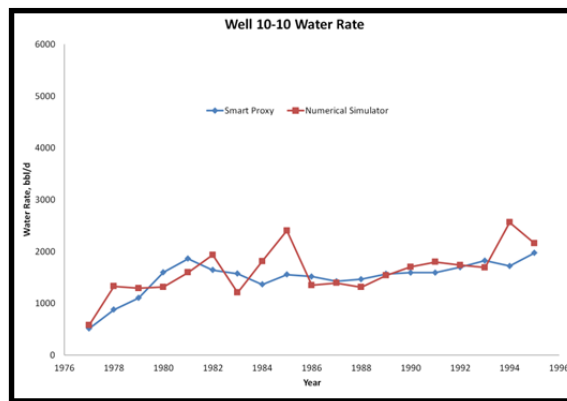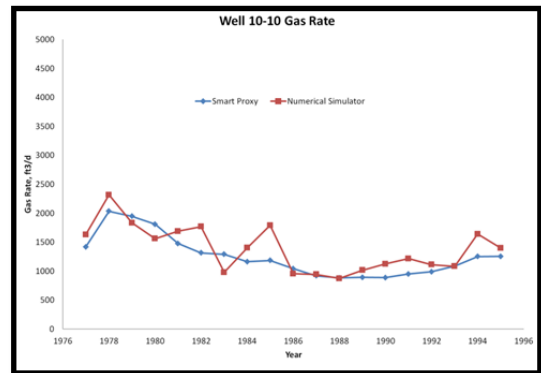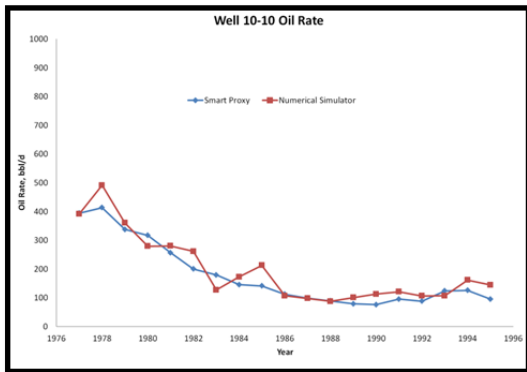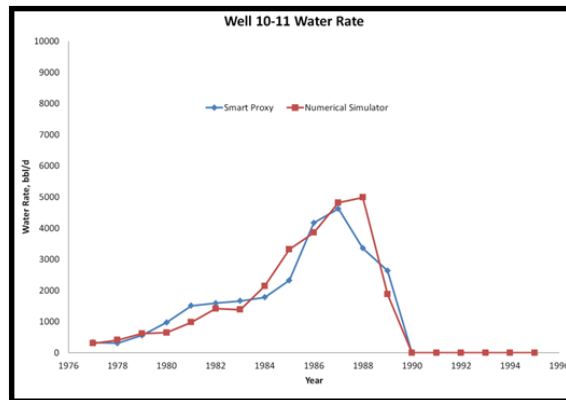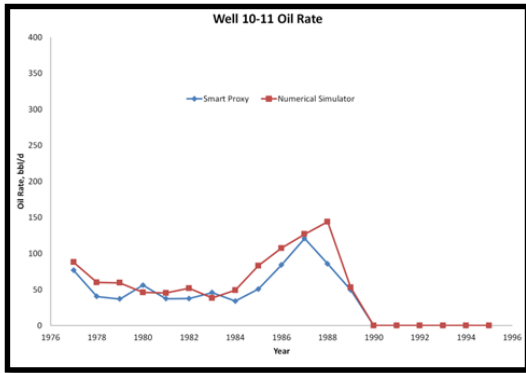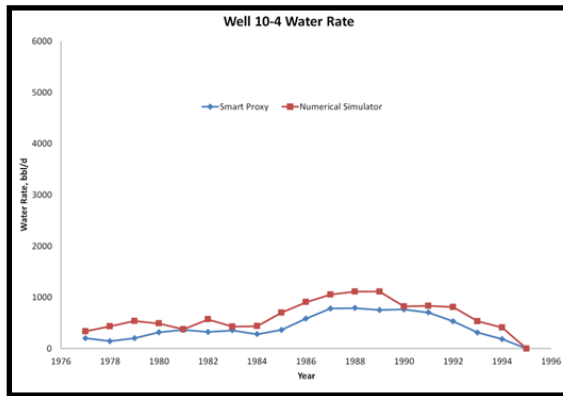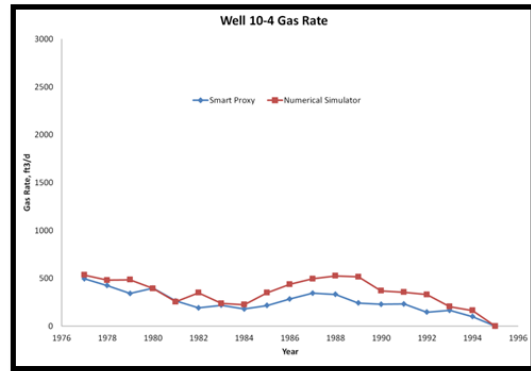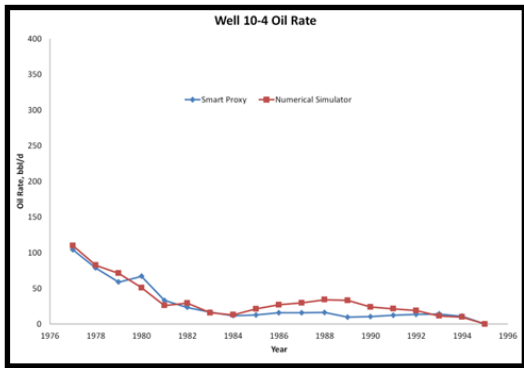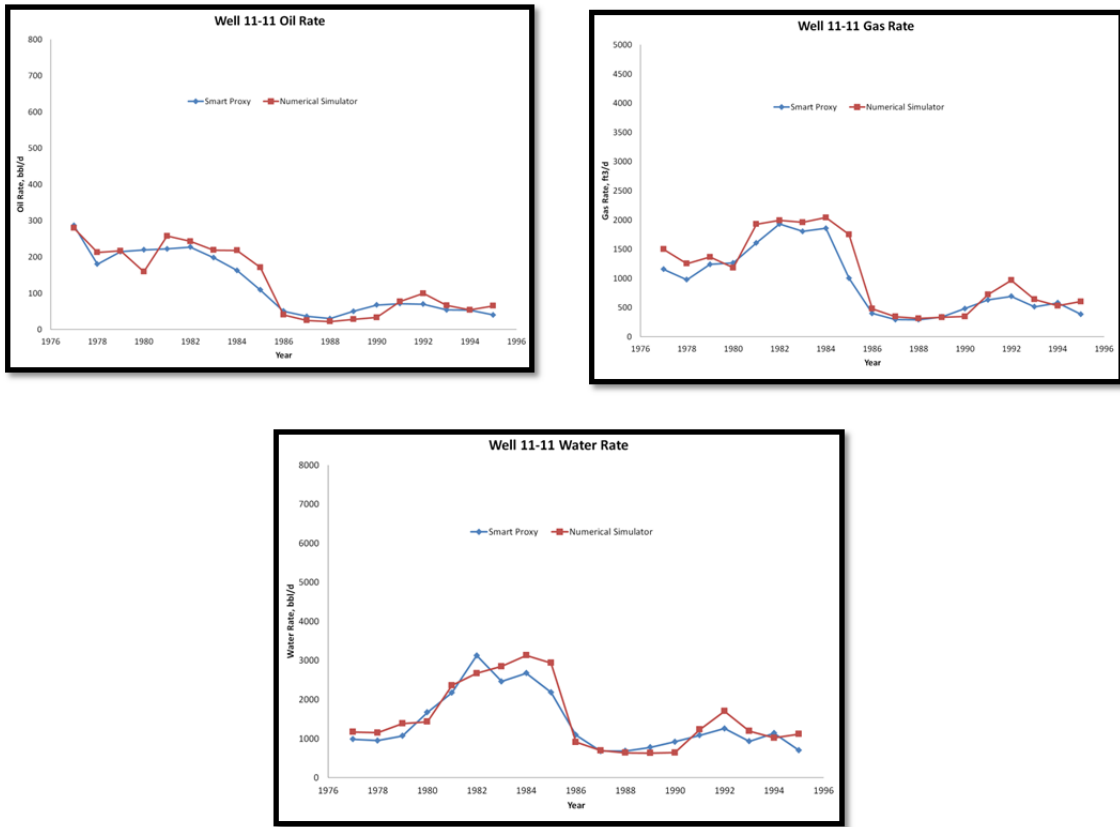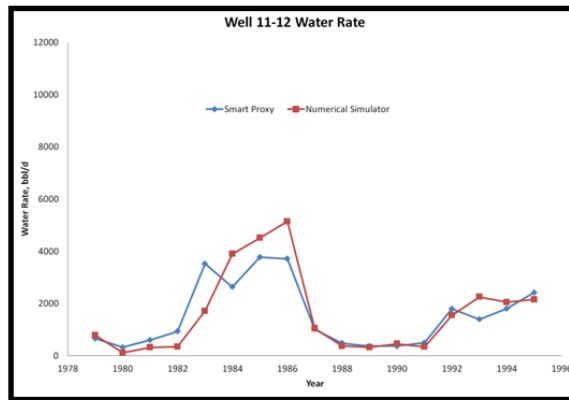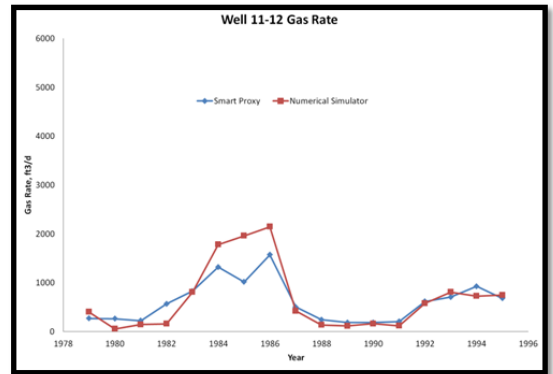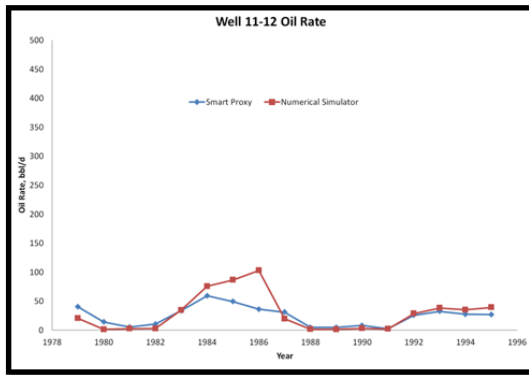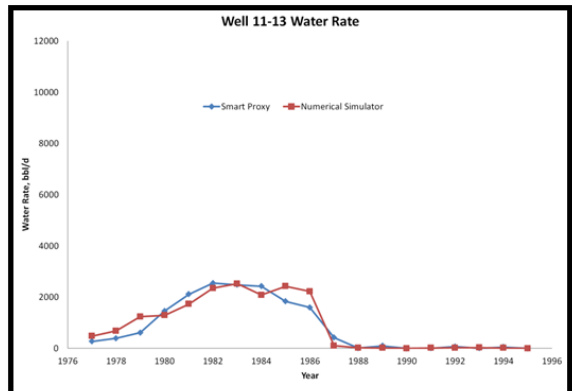
Phase 3 wells production profile.

Figure 9.2: Well 10 1

Figure 9.3: Well 10 1
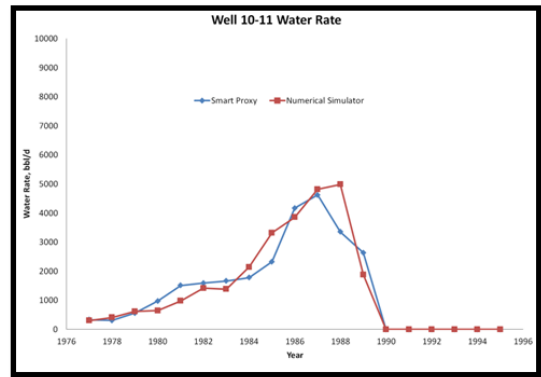
Figure 9.4: Well 10 11
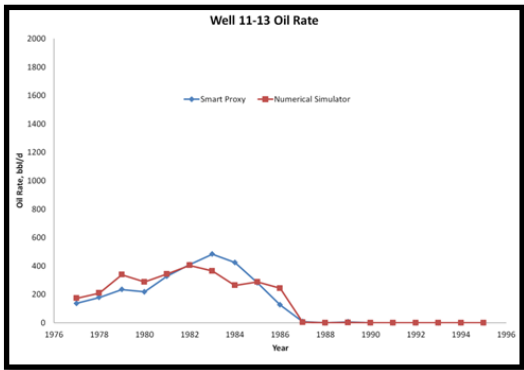
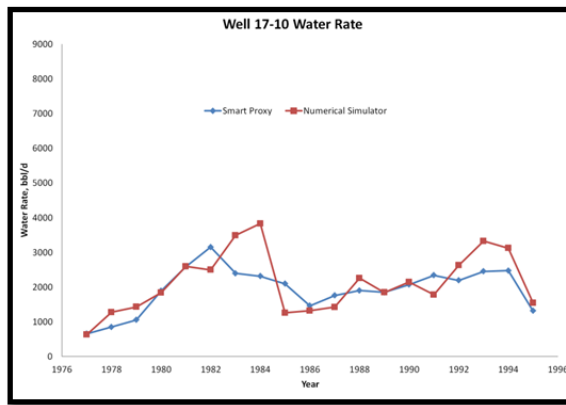Figure 9.5: Well 10 4

Figure 9.6: Well 11 11

Figure 9.7: Well 11 12
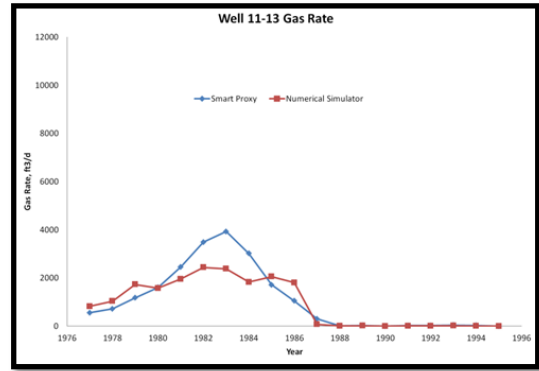
Figure 9.8: Well 11 13

Figure 9.9: Well 17 10