Data-Driven Approaches to Estimate the Impact and Presence of Shale Barriers in SAGD Reservoirs

by

Jingwen Zheng

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

PETROLEUM ENGINEERING

Department of Civil and Environmental Engineering
University of Alberta

ABSTRACT

Steam-assisted gravity drainage (SAGD) is one of the most successful in-situ techniques and has been widely adopted for heavy oil and bitumen recovery. Steam chamber development and SAGD production performance are highly sensitive to the reservoir heterogeneity, thus characterizing shale heterogeneities and quantifying their influence is essential for analysis and optimization of SAGD operations. This thesis presents a novel workflow with data-driven techniques to quantify the uncertain influences of heterogeneous shale barrier configurations and infer potential shale barrier configurations from the SAGD production histories. The workflow is trained and tested on a set of 2D and 3D synthetic models, and it is subsequently applied to a field case study.

The data employed for establishing the workflow is derived from a set of synthetic cases based on petrophysical properties and operational constraints representative of the Athabasca oil sands reservoirs. Reservoir heterogeneities are simulated by superimposing sets of idealized shale barrier configurations on the homogeneous model; each shale configuration is parameterized by a unique set of indices that represent the location and geometry of shale barriers. The workflow consists of three major steps: (1) multidimensional scaling and cluster analysis are performed to classify the cases in the training dataset into multiple groups sharing similar production characteristics; (2) for each cluster, an AI-based forward model is constructed to correlate the reservoir simulation predictions (outputs) with the associated shale barrier configurations (inputs); (3) to infer the unknown shale barrier configuration corresponding to a new SAGD production history, its production profile is first analyzed and assigned to one of the

clusters identified in step 2, and a hybrid inverse modeling scheme, which integrates the genetic algorithm and previously-trained forward model, is adopted.

The workflow is initially tested using several cases with arbitrary shale barrier configurations. Good agreement between the predicted and target shale barrier configurations is observed. The workflow is later applied to examine the actual field production histories extracted from Firebag. The inferred shale barrier features are consistent with those interpreted from the petrophysical log. This work introduces a novel systematic workflow for applying machine learning techniques to analyze the impact and distribution of heterogeneous shale barriers from SAGD field data. It presents an innovative 3D shale barrier parameterization scheme and a cluster-based approach for visualizing and inferring internal structures among many realizations of shale barrier configurations. It offers a potentially efficient way for identifying a reasonable ensemble of initial configurations that can be subjected to further (more detailed) history matching and facilitate the optimization of operational strategies for mitigating the impact of shale barriers.

PREFACE

This thesis is an original work by Jingwen Zheng. Parts of the research project have been previously published, or are ready for the journal submission.

Chapter 3 and 4 are composed in part by Zheng, J., Leung, J. Y., Sawatzky, R. P., & Alvarez, J. M. (2018a), "A proxy model for predicting SAGD production from reservoirs containing shale barriers", *Journal of Energy Resources Technology*, *140*(12), 122903; and Zheng, J., Leung, J.Y., Sawatzky, R.P., and Alvarez, J.M. (2018b), "An AI-based workflow for estimating shale barrier configurations from SAGD production histories", *Neural Computing and Applications*, https://doi.org/10.1007/s00521-018-3365-9. I was responsible for the data collection, model construction, and analysis as well as the manuscript composition. Leung, J. Y., Sawatzky, R. P. and Alvarez, J. M. were the supervisory authors and were involved in concept formation and manuscript composition.

Chapter 5 is composed in part by Zheng, J., Leung, J.Y., Sawatzky, R.P., and Alvarez, J.M. (2018c), "A cluster-based approach for visualizing and quantifying the uncertainty in the impacts of uncertain shale barrier configurations on SAGD production", paper presented at *SPE Canada Heavy Oil Technical Conference*, Calgary, AB, Canada. I was responsible for the data collection, model construction, and analysis as well as the manuscript composition. Leung, J. Y., Sawatzky, R. P. and Alvarez, J. M. were the supervisory authors and were involved in concept formation and manuscript composition.

Chapter 6 and 7 of this thesis are originally written by Jingwen Zheng and have not been published before.

DEDICATION

Dedicated to my family and my teachers for their love, endless support and mentorship.

ACKNOWLEDGEMENT

First and foremost, I would like to thank my supervisors, Dr. Juliana Y. Leung, Dr. Ronald P. Sawatzky and Dr. Jose M. Alvarez, for their continuous support, encouragement and invaluable advice throughout my studies and research. This thesis would not have been possible without their guidance. I am also grateful to my supervisory and examining committee, Dr. Clayton V. Deutsch, Dr. David Potter, Dr. Xin Wang and Dr. Japan Trivedi, for their attendance to my final defense and their insightful comments.

I would like to thank all my colleagues in Dr. Leung's research group for providing an enjoyable research environment and all the inspiring discussions. I would also like to express my thanks to my partner Yun and my dear friends Bo, Min, Mingxiang, Xinming and Yixiong for their company and help during my PhD studies.

My appreciation also goes to InnoTech Alberta for their financial support and Computer Modeling Group (CMG) for providing academic licenses for STARS.

Finally, I would like to give my eternal gratitude to my parents. I could not have gotten this far without their endless love, support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

a_j^l	=	Input at neuron j in layer l in MLP network
a_k^{l-1}	=	Output from neuron k in layer l-1 in MLP network
$\boldsymbol{A}_\mathrm{n}$	=	Diagonal matrix with n largest eigenvalues
b_j^l	=	Bias for the neuron j in layer l in MLP network
B	=	Bias parameter in SVR
C	=	Objective function that compares the similarity between two production profiles
C_m	=	Soft margin parameter in SVR
d	=	Distance function that compares the similarity between two reservoir cases
$\boldsymbol{D}$	=	$N \times N$ dissimilarity matrix
$\boldsymbol{D'}$	=	$N' \times N'$ dissimilarity matrix
$\mathbf{e}$	=	The error matrix of the preceding layer in MLP network
$\boldsymbol{E}_\mathrm{n}$	=	Matrix with n largest eigen vectors
f	=	Transfer function of MLP network
i	=	Single data point (monthly) along the production profile
$\boldsymbol{I}$	=	Identity matrix
$\mathbf{J}$	=	The Jacobian matrix of the preceding layer in MLP network
$\boldsymbol{J_n}$	=	N-dimensional all-ones matrix
k	=	Total number of neurons in a layer of MLP network
L	=	Total number of layers of MLP network
m	=	Total number of monthly production records for one production profile
m_L	=	Iteration step of the Levenberg-Marquardt method
n	=	Dimension of the Euclidean space $\boldsymbol{R}$
n_t	=	Total number of training samples
N_{2D}	=	Number of reservoir cases in the 2D data set
N_{3D}	=	Number of reservoir cases in the 3D data set
N_{2D}'	=	Number of reservoir cases in the reduced 2D data set
N_{3D}'	=	Number of reservoir cases in the 3D data set after adding additional cases
N_a	=	Number of new cases added to the clusters with limited members
N_c	=	Number of clusters for K-means clustering analysis

N_{locj} = Number of shale barriers presenting at Shale Location j in a 3D reservoir model

N_{seg} = Number of 2D segments in a 3D reservoir model

$\boldsymbol{P}$ = Square of dissimilarity matrix $\boldsymbol{D}$

$\boldsymbol{P'}$ = Square of dissimilarity matrix $\boldsymbol{D'}$

q_{oi} , q_{oi}^* = Oil production rate at time i

q_{si} , q_{si}^* = Steam injection rate at time i

Q_{of}, Q_{ot} = Cumulative oil production

Q_{sf}, Q_{st} = Cumulative steam injection

$\boldsymbol{R_{2D}}$ = Euclidean feature space after MDS with N_{2D} 2D cases

$\boldsymbol{R_{2D}}'$ = Euclidean feature space after MDS with N_{2D}' 2D cases

$\boldsymbol{R_{3D}}$ = Euclidean feature space after MDS with N_{3D} 3D cases

$\boldsymbol{R_{3D}}'$ = Euclidean feature space after MDS with N_{3D}' 3D cases

S_o , S_s = Allocation ratios

w_{jk}^{l} = The weight connecting between neuron k in in layer l-1 and neuron j in layer l

$\mathbf{w}$ = The weight matrix of the preceding layer in MLP network

$\boldsymbol{w}$ = A set of coefficients in SVR

$\boldsymbol{x}$ = Input variables

y = Target output

Δx = Grid size in the x direction

Δy = Grid size in the y direction

Δz = Grid size in the z direction

***Greek letters*:**

μ = Combination coefficient in Levenberg-Marquardt method

α_i, α_i^* = Nonnegative Lagrange multipliers in SVR

ε = Error parameter in SVR

ξ_i, ξ_i^* = Slack variables in SVR

Acronyms:

AI	=	Artificial intelligence
ANN	=	Artificial neural network
CHV	=	Connected hydrocarbon volume
CSS	=	Cyclic steam stimulation
DOE	=	Design of experiments
EnKF	=	Ensemble Kalman filter
EOR	=	Enhanced oil recovery
GA	=	Genetic algorithm
GOR	=	Gas-oil ratio
GP	=	Genetic programming
IHS	=	Inclined heterolithic strata
LI	=	Lean zone indicator
MDS	=	Multidimensional scaling
MLP	=	Multi-layer perceptron
NMSE	=	Normalized mean squared error
OFAT	=	One-factor-at-a-time
RBF	=	Radial basis function
PCA	=	Principle component analysis
PCE	=	Polynomial chaos expansion
PVT	=	Pressure-volume-temperature
SAGD	=	Steam-assisted gravity drainage
SA	=	Simulated annealing
SI	=	Shale indicator
SPSA	=	Simultaneous perturbation stochastic approximation
SOR	=	Steam-oil ratio
SVM	=	Support vector machine
SVR	=	Support vector machine regression
WOR	=	Water-oil ratio
UTF	=	Underground test facility

CHAPTER 1: INTRODUCTION

1.1 Background and Motivations

Canada's oil sands and bitumen reserves are among world's largest petroleum deposits, most of which are located in the province of Alberta and Saskatchewan. The density and viscosity of heavy oil or bitumen are always higher than that of light crude oil, which inhibits it from flowing easily to the production wells under normal reservoir conditions. Therefore, conventional production methods become ineffective and many thermal enhanced oil recovery (EOR) techniques have been invented for bitumen exploration and production.

Steam-assisted gravity drainage (SAGD) is a widely-adopted in-situ thermal production technique for heavy oil and bitumen recovery. SAGD production performance is highly sensitive to the reservoir heterogeneity. Low-permeability shale barriers obstruct the communication between the injected steam and the in-situ bitumen, impeding the lateral and vertical expansion of the steam chamber. Therefore, characterizing shale heterogeneities and quantifying their influences is essential for the success of SAGD project management. Analytical and semi-analytical models, such as those originally proposed by Butler (1981, 1985) and many subsequent improvements have been proposed over the past two decades (Reis, 1992; Akin, 2005; Sharma and Gates, 2010). However, capturing the dynamic influence of reservoir heterogeneity is difficult with these analytical techniques. Experimental analysis reveals the impacts of shale barriers with lab-scale models; however, such experiments cannot reproduce all the operational conditions and field-scale heterogeneities. Studies that are based on numerical modeling offer insights about the impacts of field-scale heterogeneity, and numerical simulation based history-matching process is usually applied for heterogeneity characterization. These approaches are effective but usually computationally expensive.

On the other hand, the SAGD field data has received increasing attentions as it provides fundamental information of reservoir properties and production characters. Data-driven modeling techniques with the implementation of artificial intelligence (AI) methods offer feasible choices for capturing internal correlations between field data and reservoir heterogeneities. Although AI techniques have been successfully adopted in various petroleum engineering applications, they have not been widely implemented in SAGD characterization workflows.

In this thesis, a dataset is gathered from the public domain which contains both geological and production information of in-operation SAGD projects. An AI-based workflow is proposed based on this field data set. The workflow describes a novel feature parameterization approach for shale barrier configurations, and data-driven approaches for correlating production time-series with shale heterogeneities and estimating shale barrier configurations from SAGD production profiles. The goal of this project is to explore the potential of applying data-driven models as feasible alternatives for capturing internal structures and non-linear relationships.

1.2 Problem Statement

Reservoir heterogeneity, especially the presence of low-permeability shale barriers, is an essential consideration for any SAGD project. Despite the successes of integrating machine learning and data-driven modelling techniques in various petroleum engineering applications, several challenges pertinent to the quantification and inference of the distribution of heterogeneous shale barriers in SAGD reservoirs still remain.

The first challenge is related to the multifaceted nature of the dataset. SAGD field data is usually sparse and noisy. For most of the fields, only monthly records are available from the public databases, yet the corresponding on-site operations are not clear. On the other hand, the field data comes from diverse sources, such as oil and water production time series, log and core data, temperature measurements at observation wells, operational constraints and etc.; their data recording frequencies and formats are different. Therefore, it is not trivial to directly apply existing machine learning algorithms to SAGD field data; instead, customization of existing techniques by integrating domain knowledge is necessary.

Another challenge is related to the proper parameterization of reservoir heterogeneity for model training. Although the influences of heterogeneity were explored in previous studies, descriptive variables of shale configurations are difficult to define; therefore, this thesis intends to propose a parameterization scheme for representing a particular shale barrier configuration and suitable for subsequent data-driven modeling.

The key problems to be addressed in this thesis can be categorized into one of three groups:

(1) Regression: predicting SAGD production time-series with shale heterogeneities;
(2) Optimization: inferring probable shale barrier configurations from target SAGD production history;
(3) Clustering and classification: investigating the internal structures and quantifying the dissimilarities among different groups of shale configurations.

These are usually high-dimensional and non-linear problems. Thus, the data-driven models established for these problems should address the following aspects: (1) selecting and formulating proper AI algorithms; (2) identifying and parameterizing relevant features as input and output variables; (3) building representative training and testing datasets. Although AI techniques have been successfully adopted in various petroleum engineering applications, there is no systematic and established workflows for their applications for heterogeneity inference in SAGD projects.

Proposed Thesis Statement:

The application of artificial intelligence and data-driven techniques could facilitate the distribution of shale barriers and their impacts on SAGD recovery performance to be inferred from production time-series data.

1.3 Research Objectives

The theme of this research is to demonstrate the feasibility of data mining and artificial intelligence algorithms in correlating reservoir heterogeneities with SAGD production profiles in a practical workflow suitable for field applications. This theme can be divided into the following objectives:

(1) Explore the methodologies for SAGD field data analysis and feature selection. Establish datasets that are representative of SAGD production histories, reservoir petrophysical properties and operational constraints. Generate 2D and 3D synthetic datasets from numerical simulation predictions.

(2) Propose a parameterization approach for representing heterogeneous characteristics of shale barrier configurations.

(3) Construct a series of data-driven model for predicting SAGD production time series based on the shale barrier configuration parameters.

(4) Construct an AI-based workflow for inferring the presence and distribution of shale barriers from SAGD production time-series data.

(5) Investigate the internal structures among different realizations of shale configuration with clustering techniques and quantify the dissimilarities between different realizations.

(6) Demonstrate the feasibility of proposed workflow in field applications, including:

    1) Assembling a representative field production dataset from an existing SAGD project (i.e., Suncor's Firebag);

    2) Infer probable shale barrier configurations from the SAGD field production histories using the developed workflows and validate the results with well log interpretations.

1.4 Thesis Outline

This thesis consists of 8 chapters, and it is organized as follows:

Chapter 1 presents a general introduction of this thesis including background information and research motivations, problem statement, research objectives and thesis outline.

Chapter 2 presents the literature review including the fundamentals of SAGD, impacts of reservoir heterogeneity on SAGD performance, application of AI and data-driven techniques for heterogeneity characterization.

Chapter 3 presents the workflow for preparing the training/testing datasets including SAGD field data collection, synthetic model set-up, heterogeneity parameterization, and sensitivity analysis.

Chapter 4 presents the methodology for applying data-driven approaches to correlate reservoir heterogeneities with SAGD production profiles. The workflow consists of two parts: a forward regression model that estimates SAGD production time series from reservoir heterogeneities; and an inverse optimization scheme that infer unknown shale barrier configurations from a given SAGD production profile. The testing results on a set of 2D synthetic cases are also discussed.

Chapter 5 presents a workflow that applies multidimensional scaling (MDS) and cluster analysis techniques to represent the uncertain influences of different shale barrier configurations on SAGD production and to quantify the dissimilarities between realizations. The proposed workflow offers a systematic framework for optimizing the spanning of model parameter space in a given data set, which is particularly important for constructing representative training data sets for 3D proxy modeling.

Chapter 6 presents the application of proposed methodologies on 3D synthetic cases. The workflow introduced in Chapter 3 is slighted modified to accommodate for the increasing complexity of 3D problem, where internal structures of 3D shale configurations are investigated and multiple forward regression models are trained for different clusters.

Chapter 7 presents the application of proposed methodologies on a set of SAGD field production profiles assembled from Suncor's Firebag Project.

Chapter 8 presents the conclusions and contributions of this study. Limitations and recommendations for future work are also discussed.

CHAPTER 2: LITERATURE REVIEW

2.1 Fundamentals of the SAGD Process

Oil sands are mixtures of sands, water, clay and bitumen. Canada's oil sands and bitumen reserves are the third largest petroleum deposits in the world, next to Venezuela and Saudi Aribia, and most of which are located in the province of Alberta. According to Alberta Energy Regulator (2018), the remaining established crude bitumen reserves in Alberta is 164 billion barrels and the annual production is 1.036 billion barrels as of 2017. The density (API gravity 7° to 10°) and viscosity (up to millions of cP) of bitumen are much higher than that of conventional crude oil, inhibiting it from flowing easily to the production wells under normal reservoir conditions. Therefore, conventional production methods become ineffective and many thermal enhanced oil recovery (EOR) techniques have been invented for bitumen exploration and production. In Alberta, about 80% of the oil sands reserves are recoverable by in-situ methods such as steam assisted gravity drainage (SAGD) (Government of Alberta, 2017).

SAGD is a widely-adopted in-situ thermal production technique for heavy oil recovery. The SAGD process was first introduced by Butler (1981). A pair of parallel wells is drilled horizontally into the reservoir, where an injector is placed approximately 5 to 10 meters above a producer. A steam chamber is created and expands upwards and laterally, while the viscosity of the heavy oil decreases, as high-temperature and high-pressure steam is continuously injected. The heated oil then drains along the edge of the steam chamber to the producer by gravitational force (Butler et al. 1981). In practice, the SAGD recovery process usually contains several different stages:

(1) [**Start-up Operations**] This stage aims to establish thermal and hydraulic communication between the wells and improve initial SAGD response (Saltuklaroglu et al., 1999; Elliot and Kovscek, 2001). Inter-well communication is established by circulating steam through both the injector and producer. Conduction is the main heating mechanism. During the process, high temperature steam is injected through a tubing string to the toe of both injector and producer. Then the steam is condensed and heat is released into the reservoir, the condensed water will flow back through the casing-tubing annuls (Cenovus Energy, 2011). Typical duration of steam circulation is about 60 to 120 days.

(2) [**Ramp-up**] After the inter-well communication is established, steam is injected into the reservoir through the upper injector while mobilized water and oil are produced through the lower producer. During this stage, the entire length of the well pair will be heated, and the steam chamber will grow vertically. Once the steam chamber reaches the top of the reservoir, both oil production rate and steam injection rate reach a maximum level (Cenovus Energy, 2011).

(3) [**Conventional SAGD Operations**] After the ramp-up stage, the steam chamber has reached the maximum height and lateral development becomes the dominant mechanism for oil recovery. In this stage, the steam injection rate is usually controlled to maintain a target steam chamber pressure and over-burden heat loss will cause declining oil rate at steady steam injection rate (Cenovus Energy, 2011). During the SAGD operation, the bottom-hole temperature of the producing fluids is controlled to be below the steam temperature (usually 5°C to 40°C). This temperature differential is known as subcool, and this steam trap control setting would help liquids to accumulate above the producer and reduce the possibility of

steam flowing into the producing well directly (Butler, 1991; Edmunds, 2000; Ito and Suzuki, 1999).

(4) [**Winddown and Blowdown Operations**] The conventional SAGD operations will last for years until the optimal amount of steam has been injected into the reservoir. Then the steam injection will be terminated, and SAGD operation will proceed to a blowdown phase where non-condensable gas will be injected into the reservoir to maintain pressure (Butler, 2004; Yee and Strotch, 2004; Zhao et al., 2005; Cenovus Energy, 2011). During this phase, the bitumen production continues until the production drops to an uneconomic rate. This stage aims to increase the thermal efficiency by reducing the usage of steam (Gu et al., 2013).

2.2 Impact of Reservoir Heterogeneity

The first SAGD project was tested at the Underground Test Facility (UTF) in Fort McMurray, Canada (Edmunds et al. 1988). The test has revealed a number of issues associated with field applications. One of the primary challenges with the SAGD process is that steam chamber development and production performance are highly sensitive to the reservoir heterogeneity. Low-permeability shale barriers obstruct the communication between the injected steam and the in-situ bitumen, impeding the lateral and vertical expansion of the steam chamber. Learning from the UTF (Birrell et al. 2000) and 4D seismic surveillance over the Christina Lake SAGD project (Zhang et al. 2005) also highlights the negative effects of geological heterogeneity on SAGD performance. It was observed that the drainage and flow of hot fluid in the near-well region is highly sensitive to shale distribution, while the expansion of steam chamber away from the well pair is compromised by the presence of long, continuous shale or high shale proportion. Similar observations can also be found in Dang et al. (2013).

Many analytical and semi-analytical models have been proposed to represent the SAGD process (Butler, 1985, 1991; Rose, 1993; Pooladi-Darvish, 1994); attempts were made to improve the original Butler's model: Reis (1992) proposed a linear steam chamber profile to account for heat loss to the surroundings; Akin (2005) modeled the impacts of asphaltene; Sharma and Gates (2010) modeled the effects of multi-phase flow and relative permeability, while Mojarad and Dehghanpour (2016) incorporated emulsion flow. However, the influences of reservoir heterogeneity are generally overly simplified in these models. Yang and Butler (1992) conducted a series of experiments with a two-dimensional model, where shale barriers with varying permeability and length are incorporated. Their results showed that shale barriers with extensive lateral continuity would reduce oil production rate, although the impact on ultimate recovery is less pronounced. A major deficiency with experimental analysis is that lab-scale models could not reproduce all the operational conditions and field-scale heterogeneities.

Studies that are based on numerical modeling could offer some insights about the impacts of field-scale heterogeneity. Kisman and Yeung (1995) indicated that the effect of shale barriers on SAGD performance is minor, unless the shale is both continuous and extensive. Chen et al. (2008) presented a stochastic model of shale distribution. It was observed that fluid flow in the near-well region is highly sensitive to the distribution of shale barriers, while steam chamber development in the region further away from the well pair is affected only by long/continuous shale. Similar conclusions were presented by Le Ravalec et al. (2009); results from their sensitivity analysis reiterated that impacts on SAGD performance due to shale barriers depend strongly on their sizes and relative locations to the well pairs, and the shale barriers located between injector and producer are the most detrimental ones. Amirian et al. (2015) and Wang & Leung (2015) demonstrated that as the distance between a shale barrier and the horizontal well pair decreases, or as the volume or continuity of a shale barrier increases, SAGD recovery efficiency would decrease. They parameterized the impacts of a given shale barrier with an index, which is the distance between a shale barrier and the well pair divided by the shale volume. Similar conclusions were also reported by Lee et al. (2015); 400 SAGD cases corresponding to different correlation length and proportions of shale barriers were simulated numerically, and the results confirmed that recovery factor is negatively related to the correlation length or proportions of shale barriers. It should be noted that computational costs associated with numerical simulations with fine (high-resolution) meshes for capturing all the relevant reservoir heterogeneities could be high, limiting its application in real-time optimization.

2.3 Assessing the Distribution of Reservoir Heterogeneity

Given the detrimental effect of reservoir heterogeneities, it is crucial to effectively infer the presence and distribution of heterogeneous reservoir features (e.g. low permeability shale barriers) for optimizing operating strategies and reservoir management.

Assessing the impact and distribution of reservoir heterogeneity for a particular recovery process often involves numerical simulation: the initial multivariate probability distribution of petrophysical variables (porosity, permeability and facie distribution) is inferred from static data, from which numerous realizations are sampled and subjected to flow simulation. This probability distribution would be updated to incorporate the dynamic data (time series of oil/water/steam rates, bottom-hole pressure, and 4D seismic) during history matching. This process of dynamic data integration is an inverse mathematical problem, where the solutions could be non-unique: multiple cases of petrophysical properties could potentially yield a similar set of dynamic data. Since some of the earlier works by Kruger (1960) and Jacquard (1965), in which permeability was adjusted, many advancements in assisted history matching were made over the past few decades such as colony optimization (Hajizadeh et al., 2011), ensemble Kalman filter (EnKF) (Liu and Oliver, 2005; Jafarpour and Tarrahi, 2011; Patel et al., 2015), genetic algorithm (Ballester and Carter, 2007; Maschio et al., 2008), gradient-based approaches (Brun et al., 2004), Markov chain Monte Carlo (Liu and Oliver, 2003), stimulated annealing (Ouenes, 1993), scatter/tabu search (de Sousa, 2007; Yang et al., 2007), and simultaneous perturbation stochastic approximation (SPSA) (Gao et al., 2007).

Some of these techniques have also been adopted for SAGD reservoir characterizations. Jia et al. (2009) implemented a SPSA algorithm to estimate parameters such as horizontal permeability, porosity and initial oil saturation of a synthetic SAGD model. Ito and Chen (2010) performed history matching with field data collected from a well-pair in Burnt Lake. Their model was used to understand the interplay between steam injection, operating pressure and oil production; their model confirmed that those factors had significant impacts on steam chamber growth, migration of non-condensable gas and oil drainage pattern. Hiebert et al. (2013) incorporated 4D seismic data to history match the production data (water injection rate, oil/water production rate), as well as temperature data (which is indicative of the location or top of the steam chamber, its thickness and average pressure) for one of the well pair in Pad 101 of the Suncor's Firebag SAGD project. A few recent history-matching studies have been applied for a number of SAGD reservoirs, including the MacKay River (Liu et al., 2013) and the Long Lake (Zhang et al., 2014 and Feizabadi et al., 2014) deposits.

It is important to note that the history-matching procedure usually requires a large number of forward simulations, which could be quite computationally intensive (Lee et al., 2017). In the case of SAGD production time-series data, a large number of shale configurations would be subjected to flow simulations repeatedly until a production match is obtained. Therefore to optimize the workflow and improve the efficiency of conventional SAGD history-matching routine, it is in great demand to: (1) develop fast and reliable proxies to numerical simulation for capturing the effects of complex heterogeneities and generating time series SAGD production profiles; (2) develop an efficient and reliably workflow for quickly inferring a set of probable shale barrier configurations that are consistent with the target production history; these models can be regarded as an initial estimate of input production profiles and subjected to further history matching for a more accurate final match.

2.4 Artificial Intelligence and Data-Driven Approaches

Artificial intelligence and data mining techniques have been successfully adopted in various petroleum engineering applications, such as well log interpretation, production optimization, CO2 flooding, pump failure prediction and history matching (Wu and Nyland, 1986; Popa et al. 2005; Liu et al. 2013; Shahkarami et al. 2014; Elkatatny 2018; Le Van and Chon, 2018; Moussa et al. 2018).

In particular, the use of proxy models by incorporating AI and data mining techniques to replace expensive forward simulations is often adopted to lower the computational cost. Yu et al. (2008) trained a proxy model using the genetic programming (GP) symbolic regression algorithm; the training data set consisted of over 800 forward simulations, and the model was used to predict oil production for a field in West Africa. For SAGD applications, Fedutenko et al. (2014) formulated an artificial neural network (ANN) based proxy to predict the field oil production and steam injection rates with time-varying operating conditions; the limitation, however, was that the training data set consisted of only 50 cases, limiting the model's validity to a narrow set of heterogeneity distribution and reservoir conditions. Klie (2015) proposed a hybrid proxy by combining an analytical chamber growth model with a data analytics component. The model was effective for predicting cumulative oil production and steam-oil ratio. Jain et al. (2017) employed the polynomial chaos expansion (PCE) method to construct a proxy based on 10 synthetic simulation models. Similarly, Sun & Ertekin (2017) presented an ANN-based expert system that was trained with 2921 2D synthetic simulation cases to model the cyclic steam stimulation (CSS) process and to predict the producing oil rate. Various input parameters, including petrophysical properties, initial conditions and operational constraints, were included. The main drawback of these studies is that the permeability distribution was always assumed to be Gaussian; however, shale barriers would likely lead to abrupt transition in the model parameters. In the end, the proper parameterization of specific shale barrier configurations in a proxy model is not trivial.

Amirian et al. (2015), Ma et al. (2015, 2018) and Wang & Leung (2015) have incorporated a variety of data-mining techniques, including ANN, cluster analysis, time series analysis, and principal component analysis (PCA), in an attempt to correlate shale barrier characteristics with oil production. A number of dimensionless variables such as shale indicator (SI) and lean zone indicator (LI) were implemented. However, these parameters capture only limited characteristics of the nearest heterogeneous features; information regarding the overall shale distribution is not represented.

Another important factor for building data-driven proxy models is design of experiments (DOE). Although many AI-based proxy models produced reasonable results, there was a huge difference with respect to the size of the training data employed in their studies. One-factor-at-a-time (OFAT) is a basic method of DOE where a single variable is varied at a time while keeping others fixed. Despite its ease of implementation, OFAT usually requires a large number of experiments and ignores the interactions among multiple factors (Qu & Wu, 2004; Wang & Wan, 2008). Fisher (1936, 1937) proposed a series of DOE fundamental principles and methods, including randomization, statistical replication, blocking, Latin squares (orthogonal arrays), confounding, and factorial experiments. Factorial design is a widely applied method (Box, 1980; Preece, 1990). It can be classified as either full factorial design or fractional factorial design. Full factorial design investigates every combination of the factor levels; hence, the number of experiments increases geometrically with the number of factors. On the other hand, fractional factorial design utilizes only a fraction of the full design (Wang & Wan, 2008).

In the context of SAGD data-driven modeling, a particular challenge is assembling a representative training data set. Implementing a full factorial design that involves all possible shale barrier configurations would be formidable. A representative training data set should be large enough to sufficiently span the feature (input and output parameter) space without exhaustively sampling realizations with similar production characteristics. However, this step is usually achieved via trial-and-error, and there is little discussion in the reviewed literature regarding the methodology for quantifying similarities among different realizations and optimizing training dataset to efficiently span the feature parameter space.

In this research, artificial neural network and support vector machine regression (SVR) are employed to establish the relationship between reservoir heterogeneities and SAGD production performance. Genetic algorithm (GA) is applied to solve the inverse problem which is inferring shale barrier configurations from SAGD production history. Other AI techniques such as multidimensional scaling (MDS), K-means clustering and decision tree are also applied for the purpose of feature selection, cluster analysis and model performance improvement. The detailed formulas of these methods will be discussed in the following chapters.

CHAPTER 3: MODEL SETUP AND PARAMETERIZATION OF SHALE HETEROGENETIES[1]

In this research, the production data employed for constructing data-driven models is derived from a set of synthetic SAGD reservoir simulations based on petrophysical properties and operational constraints representative of Athabasca oil sands reservoirs. One primary challenge is related to proper parameterization of reservoir heterogeneity for model training. Although the influences of heterogeneity were explored in previous studies, descriptive variables of shale configurations are rarely defined. A number of dimensionless variables such as shale indicator (SI) and lean zone indicator (LI) were implemented by Amirian et al. (2015), Ma et al. (2015, 2017) and Wang & Leung (2015) to describe reservoir heterogeneity; however, these parameters capture only limited characteristics of the nearest heterogeneous features; information regarding the overall shale distribution is not represented.

The primary objectives of this chapter are to (1) summarize the petrophysical properties and operational constraints representative of Athabasca oil sands reservoirs for synthetic model construction and (2) propose a feature parameterization approach for shale barrier configurations.

3.1 SAGD Field Data Collection and Analysis

A set of SAGD field data is gathered for Suncor's Firebag project from public domain. The dataset consists of field data from both horizontal production/injection well pairs and vertical observation wells. Data including production profiles, well logs, core measurements, reservoir

[1] The content in Chapter 3 of this thesis is derived from the following papers:

Zheng, J., Leung, J. Y., Sawatzky, R. P., & Alvarez, J. M. (2018a). A proxy model for predicting SAGD production from reservoirs containing shale barriers. *Journal of Energy Resources Technology*, *140*(12), 122903.

Zheng, J., Leung, J.Y., Sawatzky, R.P., and Alvarez, J.M. (2018b). An AI-based workflow for estimating shale barrier configurations from SAGD production histories. *Neural Computing and Applications*.

pressures and steam chamber temperatures are assembled from AccuMap (IHS Energy, 2015) and TOP analysis software (TOP Analysis, 2015), and representative properties are interpreted and extracted. These values are also compared against average reservoir properties recorded in Suncor's Firebag annual reports (Alberta Energy Regulator, 2012, 2013, 2014, 2015). Certain operational constraints, such as steam quality and injection pressures are obtained from the annual reports.

The petrophysical logs are interpreted to calculate parameters such as shale volume, water saturation, reservoir porosity and permeability, etc. The temperature files of observation wells are also retrieved from TOP database. The temperature files record the monthly change of temperature at different depths, which can be seen as indicators of the development of steam chamber. When observation well temperature reaches steam temperature, it usually means that the steam chamber has grown to that depth. Oppositely, lower temperature at a depth may indicate the steam chamber has not fully heated this region or shale barriers nearby block the development. In this way, the temperature files are combined with well logs to identify the location and thickness of shale barriers, as well as net pay thickness and depth of reservoir.

The production data and injection data are assembled for each well-pair to calculate key parameters such as steam-oil ratio (SOR), gas-oil ratio (GOR), water-oil ratio (WOR) and etc. Every logging well is assigned to the nearest well-pair, information like distance and incline angle between injector and the observation well are then calculated. The detailed data records for individual well-pair and pad are sorted for the usage of data-driven models, the representative properties for each pad and the whole reservoir are also summarized.

3.2 Homogeneous Base Model

A homogeneous base model consisting of only sand is created ($\Delta x = \Delta z = 1$ m). The model is 50 m along the *x*-direction, representing half the distance between two neighboring well pairs, and 40 m along the *z*-direction, which is the reservoir thickness. The major pay zone is located in the McMurray formation. Its petrophysical properties and certain operational constraints have been taken from representative values for several pads (pads 104-108, as shown in **Fig. 3.1**) in Suncor's Firebag project. A set of horizontal wells are placed at the left edge of the model (y-direction).

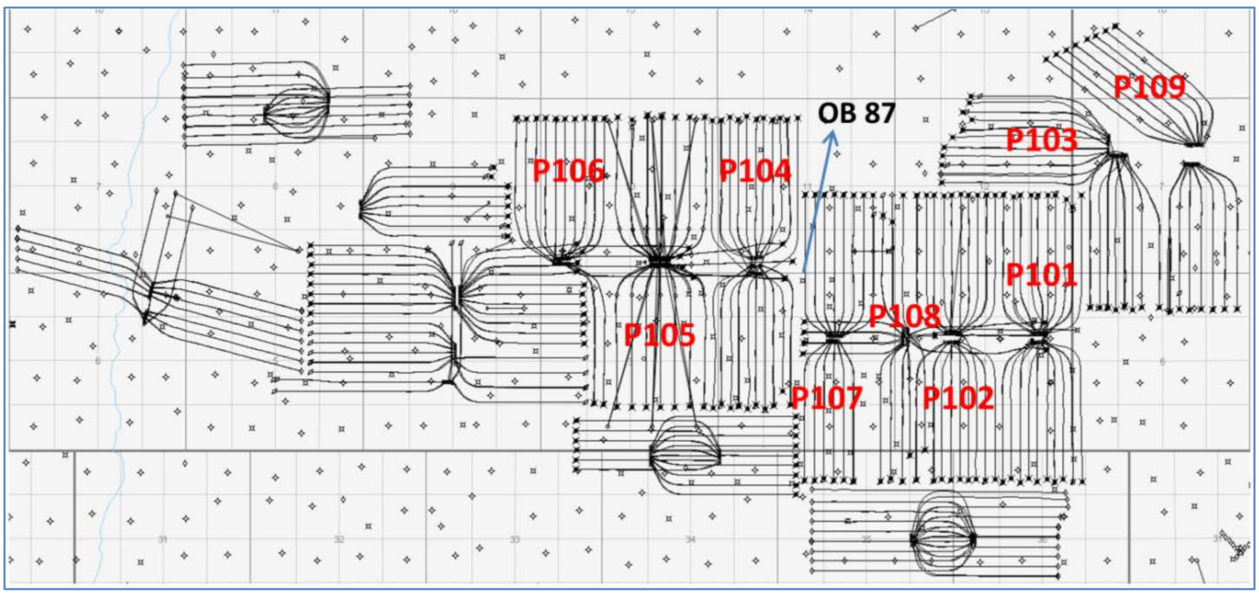


Fig. 3.1 – Relative pad locations for the Firebag project – Image extracted using AccuMap (IHS Energy, 2015).

For the 2D cases, the dimension along the *y*-direction is 20 m. To construct the 3D cases, a series of four 2D cases are stacked along the *y*-direction, resulting in a total length of 80 m (Δy = 20 m; total length = 4 × 20 m). Since only a portion of the horizontal well length is simulated in these 2D and 3D cases, the production volumes obtained with these synthetic cases must be scaled up proportionally to reflect actual field volumes. For instance, the typical length of a horizontal producer in the field is approximately 1000 m; therefore, the simulated oil and steam volumes are multiplied by factors of 1000/20 = 50 and 1000/80 = 12.5, respectively, for the 2D and 3D cases.

Table 3.1 - Reservoir properties and operating constraints for the base and heterogeneous models. Data are derived from Pads 104-108 in the Firebag Project.

Initial Reservoir Pressure [1]	800 kPa	Initial Reservoir Temperature [1]	8 ℃
Net Pay Thickness [1]	40 m	Steam Injected Pressure [1]	1500 kPa – 2423 kPa
Reservoir Top [1]	Depth: 285 m	Well Spacing [1]	100 m
Reservoir Bottom [1]	Depth: 325 m	Well Length [1]	1000 m
Injected Steam Quality [1]	95%	Steam Temperature [1]	225 ℃
Sand Porosity [1]	0.32	Shale Porosity [1]	0.25
Sand Permeability [1]	3 Darcy	Shale Permeability	3×10^{-8} Darcy
Oil Saturation in Sand [1]	0.85	Oil Saturation in Shale	0
Water Saturation in Sand [1]	0.15	Water Saturation in Shale	1
Formation Compressibility [2]	2e-6 /kPa	Formation Thermal Expansion Coefficient [2]	6e-5 /℃
Volumetric heat capacity of formation [2]	2.35e6 J/m^3-°C	Thermal conductivity of matrix [2]	1.5e5 J/m-day-°C
Thermal conductivity of water [2]	5.35e4 J/m-day-°C	Thermal conductivity of oil [2]	1.15e4 J/m-day-°C

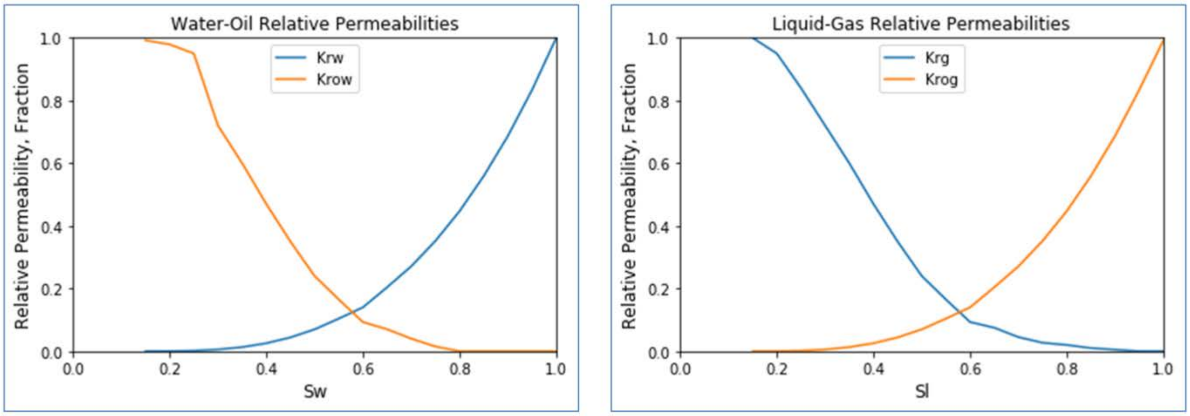
Ref: [1] Alberta Energy Regulator (2012).
[2] Li, P. (2006).

Relevant reservoir properties and operating conditions are presented in **Table 3.1**. Since there is limited information pertinent to the reservoir fluid properties in the Firebag annual reports, fluid properties representative of typical Athabasca bitumen are assumed. A pseudo "oil" component is defined to represent the dead oil properties, and it is combined with 5% methane to reproduce the live oil. The relevant PVT properties used for the simulation are listed in **Table 3.2**. The relative permeability (Good et al., 1997) is presented in **Fig. 3.2**; only the effective range of water saturation in sand (S_w>0.15) is important to fluid displacements. Viscosity-temperature correlation (Good et al., 1997) is also presented in **Fig 3.3**. Temperature data are assembled from the public domain to understand the steam chamber development.

The base model is subjected to numerical simulation (CMG, 2015). A comparison of oil production profiles between the base case and a randomly-selected producing well pair from Firebag Pad 104 is shown in **Fig. 3.4**. It is observed that production trends from the base case are in reasonable agreement with the actual field observations, considering that reservoir heterogeneities have been excluded in the base model.

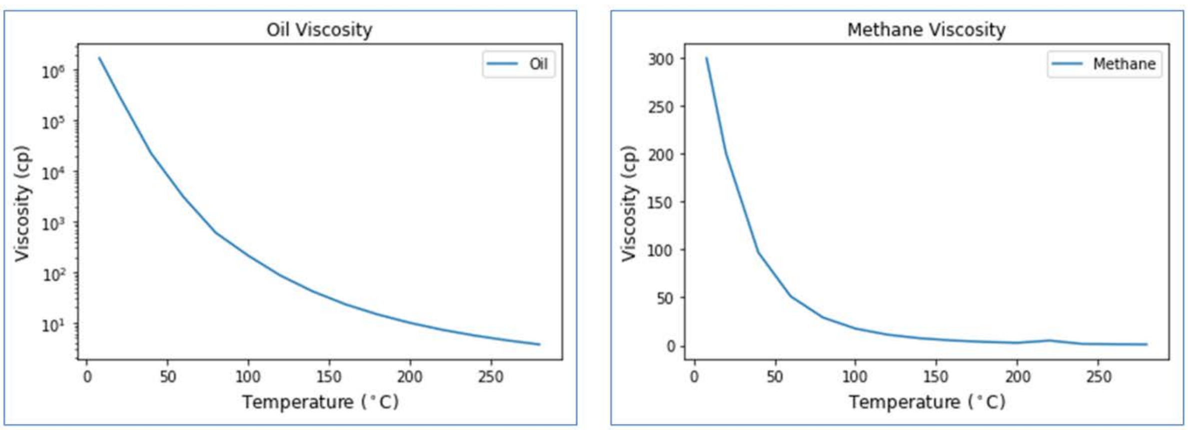
Table 3.2 – PVT properties of the fluids in the simulation model.

Component	Water	Oil	Methane
Molecular mass (kg/gmol)	0.01802	0.511	0.01604
Critical pressure (kPa)	22048	1360	4600
Critical temperature (°C)	374.15	624.65	-82.55
KV1	0	0	31914
KV2	0	0	0
KV3	0	0	0
KV4	0	0	-33.07
KV5	0	0	-27.71
Liquid density (kg/m^3)	1000	1010	421
Liquid compressibility (1/kPa)	5.70e-7	4.50e-7	1.45e-6
Coefficient of thermal expansion (1/°C)	2.14e-4	6.80e-4	8.00e-4



(a) (b)

Fig. 3.2 – Relative permeability functions: (a) water-oil; (b) liquid-gas.



(a) (b)

Fig. 3.3 – Viscosity functions with temperature: (a) oil; (b) methane.

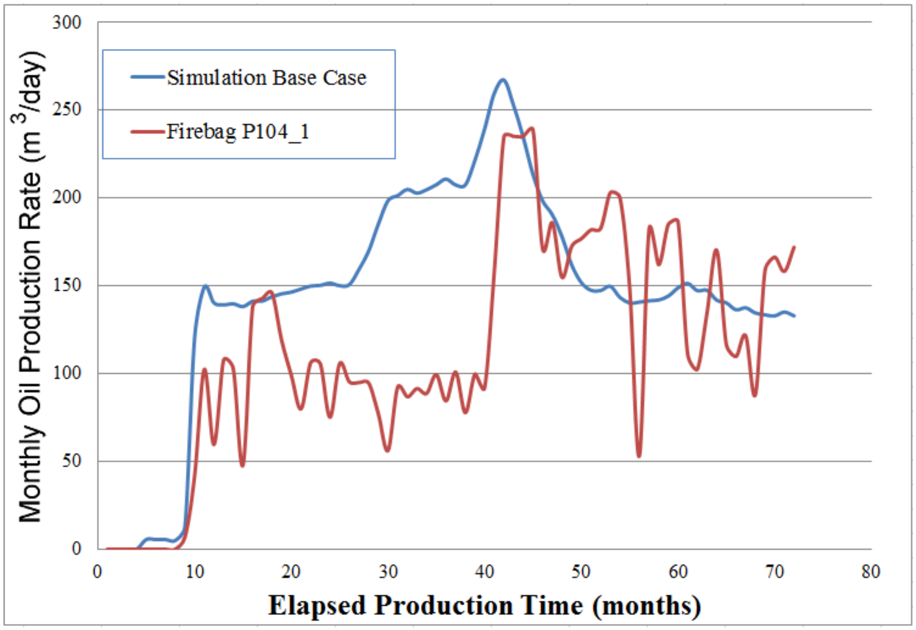


Fig. 3.4 – Comparison of monthly oil production between the base model and field observations.

3.3 Modeling and Parameterization of Shale Heterogeneities

3.3.1 Strategy and Hypotheses

To model the influence of shale barriers on SAGD production performance, various configurations of idealized shale barriers are randomly superimposed on the base model. The permeability of each shale barrier (approximately 10^{-7} to 10^{-8} Darcy) is several orders of magnitude smaller than the permeability of the oil sand (approximately 3 Darcy) in the simulation model. The individual shale barriers are categorized by their location relative to the SAGD well pair (vertical and lateral), and by their geometry (thickness and lateral extent).

In order to model shale barriers that are representative of field conditions, logging data and temperature files were collected from observation wells located on the north side of pad 104 and 107. Although many shale layers can be interpreted from the logs, not all of them would act as barriers to steam chamber development. After reviewing the temperature profiles from numerous observation wells, it is noted that shale layers of minimal thickness do not typically cause any detectable signature in the temperature measurement. Therefore, it is concluded that shale barriers that are less than 30 cm in thickness would not hinder steam chamber advancement and production performance, and, hence, they are not regarded as barriers. Average porosity and saturation of only those shale layers that serve as barriers are extracted and used as input parameters of the idealized shale barriers (**Table 3.1**). Relevant statistics of these extracted shale barriers are summarized in **Table 3.3**.

Table 3.3 - Summary of shale barriers interpreted from petrophysical information and temperature data obtained from observation wells located in pads 104N and 107N.

Count of observation wells	7
Count of shale barriers	19
Maximum shale barriers thickness	3.48 m
Minimum shale barriers thickness	0.31 m
Average shale barriers thickness	1.22 m

Parameterization of individual shale barriers is facilitated via the principle of superposition, and the key assumptions or hypotheses are:

(1) The net impact on production performance caused by a large shale barrier is the sum of individual impacts caused by a set of smaller (basic) shale units.

(2) A basic shale unit denotes the smallest shale barrier that could instigate an impact on the steam chamber development and recovery performance; in other words, the smallest allowable shale barrier is that of a basic shale unit.

(3) Any shale barrier of arbitrary size and shape can be represented by a combination of one or more of these basic shale units.

Therefore, the goal is to divide the 2D reservoir domain into a number of non-overlapping zones and assign one basic shale unit to each zone. And the shale distribution in a 3D reservoir model will be represented by stacking a set of 2D heterogeneous segments along the *y*-direction.

3.3.2 Sensitivity Analysis of Shale Barriers

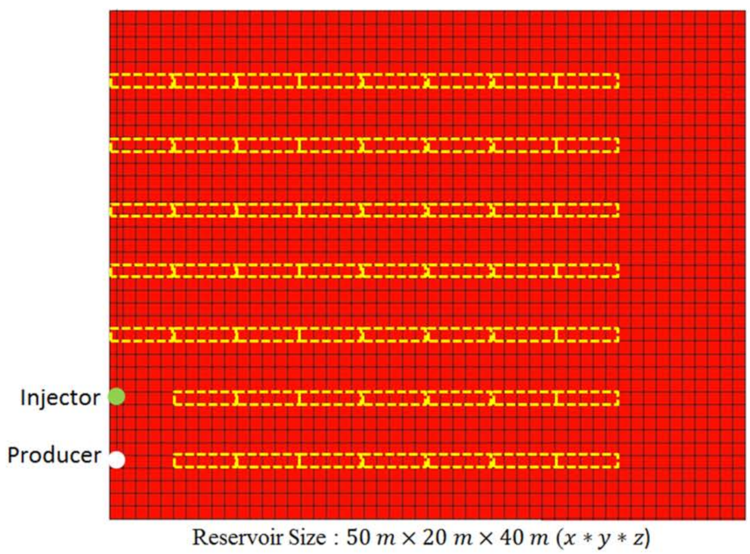
For the synthetic models, the smallest possible basic shale unit would be 1 m × 1 m, which is equal to the size of a single grid block. This choice offers the best model resolution, but the associated computational cost would be intractable: number of possible combinations = 2^N, where N = number of grid blocks. In addition, numerical simulation suggests that, in 2D reservoir model, shale barriers with limited lateral extent (~ 5 m) do not introduce an observable change in the corresponding production profiles. Thus, the choice of an appropriate basic shale unit should balance between model accuracy and computational cost. To this end, a series of sensitivity experiments was conducted for 2D models.

(1) Location of shale barrier – a single shale barrier of 5 m × 1 m is randomly placed in the domain. A length of 5 m represents approximately 1/10 of the reservoir length, while a 1 m thickness corresponds to the average thickness observed in the field data (**Table 3.3**). A total of 54 cases are created (**Fig. 3.5**).

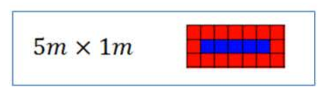
(2) Length of shale barrier – a single shale barrier with length ranging between 5 m to 35 m (thickness = 1 m) is placed at one of the 7 randomly-selected locations 8 × 7 = 56 cases are constructed (**Fig. 3.6**).

(3) Thickness of shale barrier – a single shale barrier with thickness ranging between 1 m to 5 m (length = 5 m) is placed at one of the 11 randomly-selected locations. A total of 5 × 11 = 55 cases are constructed (**Fig. 3.7**).

(4) Aspect ratio of shale barrier – a single shale barrier with aspect ratio of either 10 m × 1 m or 5 m × 2 m is placed at one of the 15 randomly-selected locations. A total of 2 × 15 = 30 cases are constructed (**Fig. 3.8**).

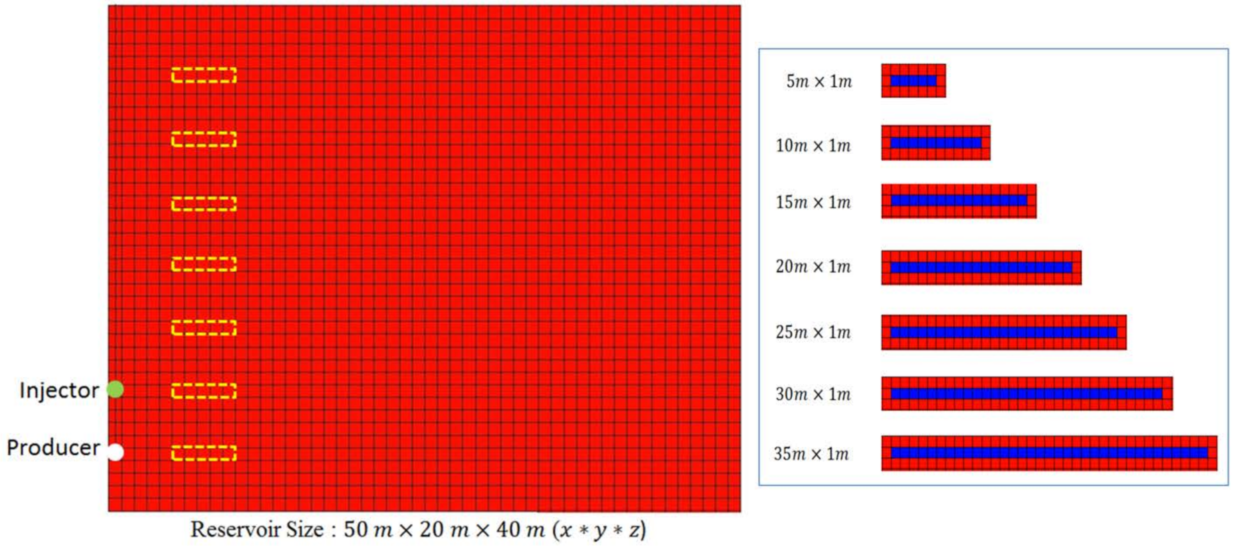


Shale locations for sensitivity experiments (1)



Applied shale barrier for sensitivity experiments (1)

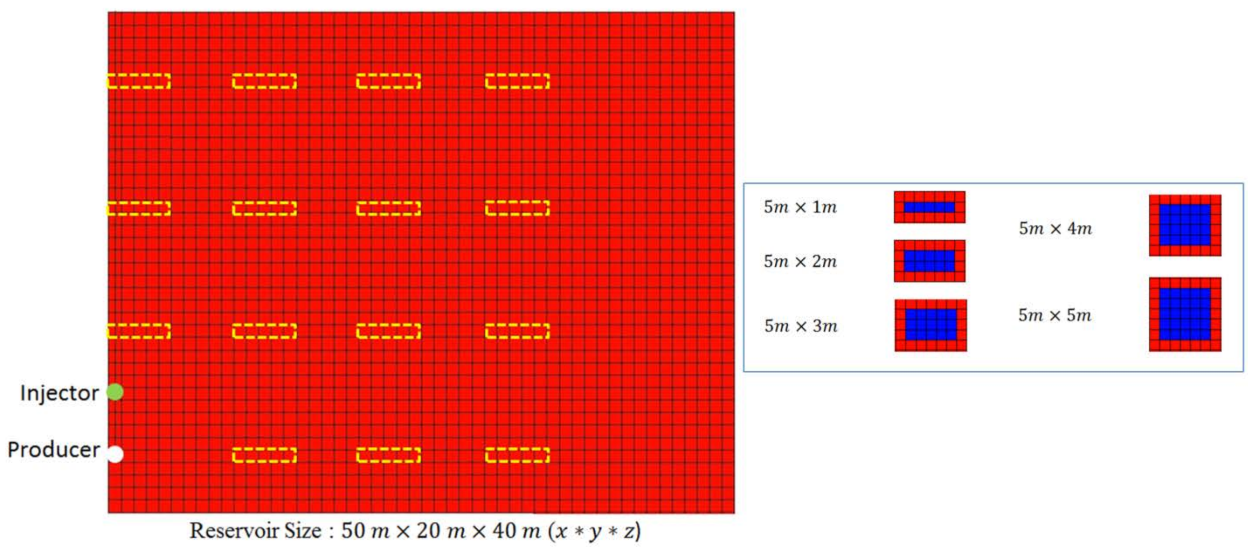
Fig. 3.5 – Sensitivity experiments on location of shale barriers.



Shale locations for sensitivity experiments (2)

Applied shale barriers for sensitivity experiments (2)

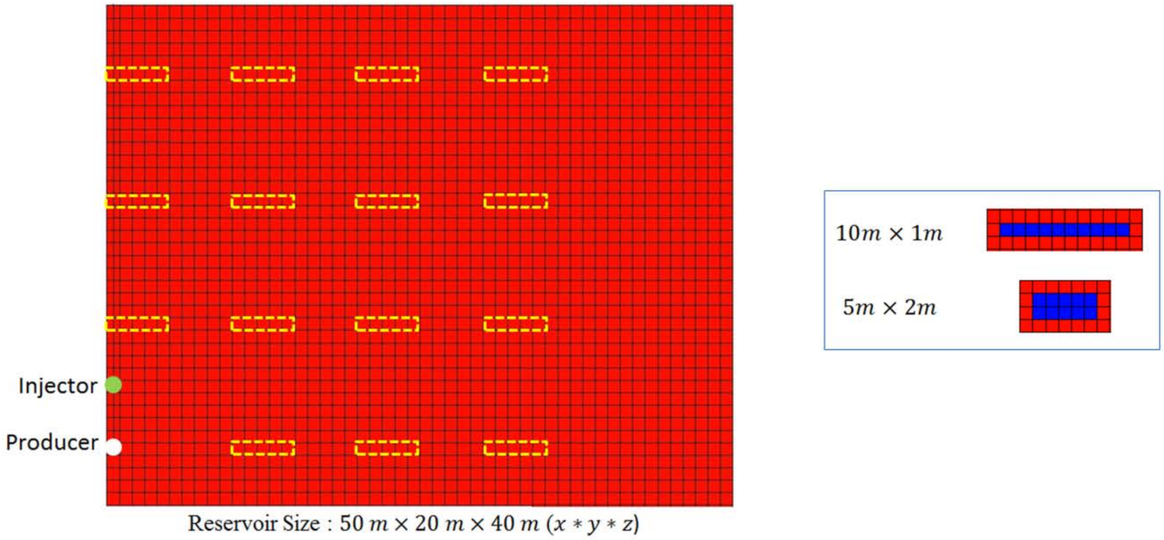
Fig. 3.6 – Sensitivity experiments on length of shale barriers.



Shale locations for sensitivity experiments (3)

Applied shale barriers for sensitivity experiments (3)

Fig. 3.7 – Sensitivity experiments on thickness of shale barriers.



Shale locations for sensitivity experiments (4)

Applied shale barriers for sensitivity experiments (4)

Fig. 3.8 – Sensitivity experiments on aspect ratio of shale barrier.

The corresponding production profiles are compared to that of the homogeneous base case in **section 3.2**. To quantify the mismatch in the production profile (time-series data) between the homogeneous base case (q_o^b) and the cases with shale barriers (q_o^s), an objective function (C) is defined as half of the mean squared error:

$$C = \frac{1}{2m}\sum_{i=1}^{m}\left(q_{oi}^{b} - q_{oi}^{s}\right)^2 \quad \text{......} \quad (3.1)$$

Where m denotes the total number of monthly data records in a production profile (oil rate); i denotes a single time point of the time series; q_{oi}^b is the oil production history of homogeneous base case at time i; and q_{oi}^s is the production history of heterogeneous case at time i. All profiles corresponding to a particular model are normalized between their minimum and maximum values on a scale of [0, 1]. Therefore the maximum value for *C* is 0.5; this would only occur if the target profile is producing constantly at the maximum oil rate, while the other profile is producing constantly at the minimum oil rate. On the opposite side, a value of zero for *C* denotes a complete match between the two profiles. In this way, the absolute difference between two production histories are compared and calculated as *C*.

The *C* values for all the sensitivity cases are calculated and analyzed (**Fig. 3.9** and **Table 3.4**). Since all the sensitivity cases are compared against the homogeneous base case, the difference in *C* values would indicate the different impacts of corresponding shale barriers. A few important observations can be made:

(1) Impact of a shale barrier of a given size decreases with the distance to the well pair;

(2) Impact of a shale barrier at a given location increases with its length and thickness;

(3) In this 2D reservoir model, production (and chamber advancement) is more sensitive to the lateral extent of a shale barrier than to its vertical extent. For a shale barrier of a fixed volume (e.g., 5 m × 2 m or 10 m × 1 m), the length increment has a bigger impact than the thickness increment.

Table 3.4 – Average *C* values for shale barriers with different aspect ratio.

Shale Barrier (length × thickness)	Average *C* value (10^{-3})
5 m × 2 m	3.71
10 m × 1 m	31.32

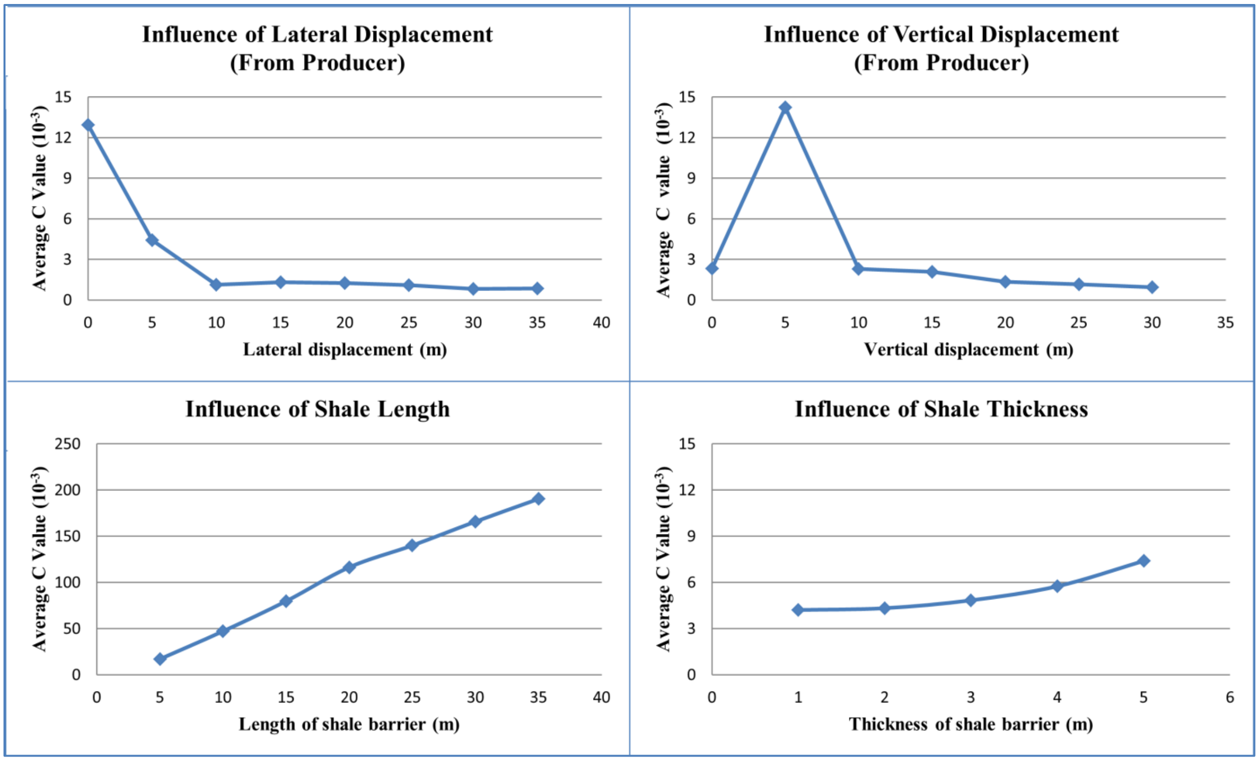


Fig. 3.9 – Average *C* values for different sensitivity experiments.

3.3.3 Shale Configuration Setup for 2D and 3D models

Based on above observations, it is proposed that the entire 2D reservoir be divided into four zones or domains based on their relative positions to the well pair. The method is designed such that smaller shale units are placed in zones that are close to the well pair and larger shale units are placed in zones that are far away; for example, the basic shale unit in zone 1 is multiple times smaller than the basic shale unit in zone 4. The underlying rationale is that the influence of any shale barrier with a particular size diminishes as the distance from the well pair increases; recovery performance is sensitive to a small shale barrier, if it is located in zone 1, but it is sensitive only to a much larger shale barrier, if it is located in zone 4.

Next, each domain is further divided into numerous sub-domains. The presence of a shale barrier within a particular sub-domain is represented by a binary index (1 = a shale barrier is present; 0 = shale barrier is absent). In this way, all possible shale barrier configurations could be parameterized with a set of binary indices.

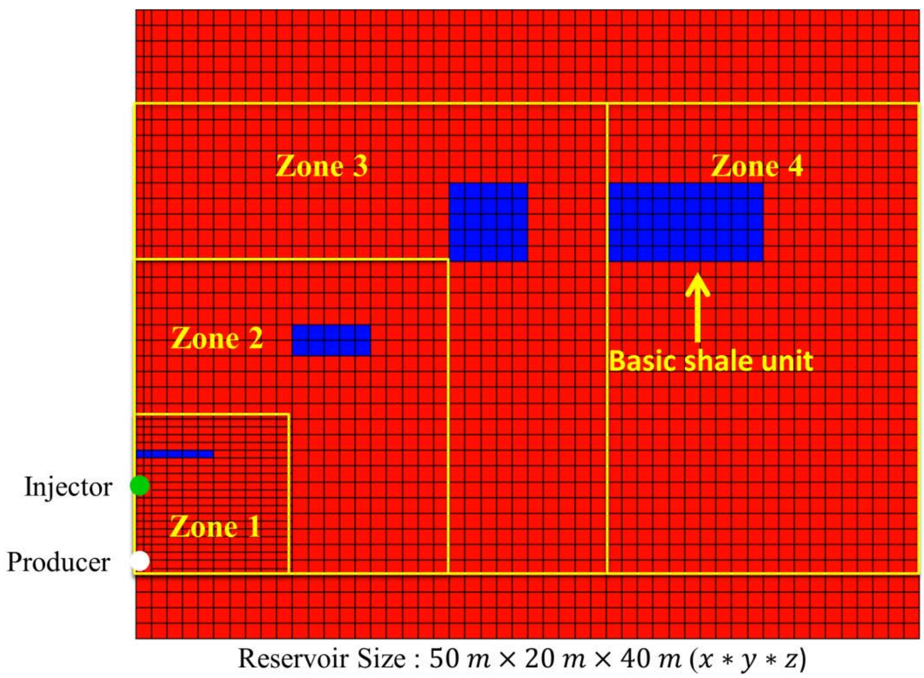
It is true that a very small basic shale unit may reproduce the reservoir heterogeneity with higher precision, but the possible scenarios of different shale configurations will also increase exponentially, requiring a more extensive training dataset and larger computational time. As a result, the size of an individual basic shale unit should be determined considering a balance between model detail and computational efficiency.

The 3D reservoir models employed in this study are constructed by stacking together four 2D reservoir models (segments) along *y*-direction (**Fig. 3.11**). Two shale configuration setups of the 2D models (segments) are applied in this research:

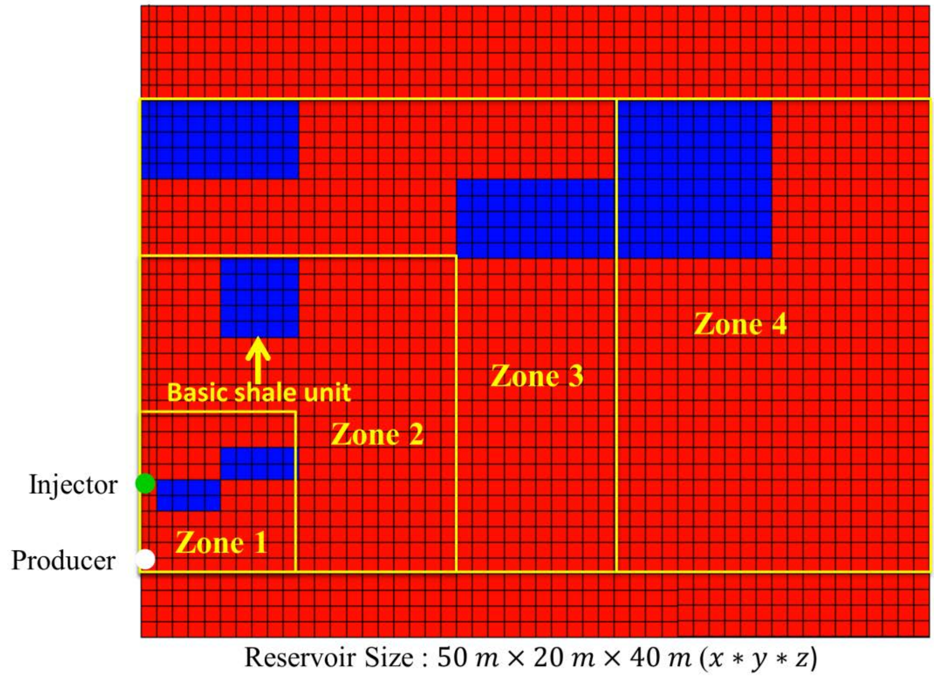
(1) The first setup is applied in Chapters 4 and 5, where 2D cases are considered. Local grid refinement is implemented in zone 1 to highlight the impact of fine-scaled heterogeneous features at the near-well region. A different basic shale unit is defined in each of the four domains, and there are a total of 102 sub-domains, as illustrated in **Table 3.5** and **Fig. 3.10(a)**. A large shale barrier spanning over more than one sub-domain would be represented by multiple basic shale units across various sub-domains.

(2) The second setup is applied in Chapters 6 and 7, where 3D cases and field case study are considered. The reason for adopting a different setup is that, with the dramatic increase in model parameter space in 3D (more binary shale indices due the stacking of numerous 2D models along the *y*-direction), the second setup is proposed to achieve some balance between maintaining model complexity and reducing computational costs. The basic shale units within each zone are enlarged, while the division of the four zones is the same as (1). An

illustration of the zone division in 2D *x-z* cross section is depicted in **Table 3.6** and **Fig. 3.10(b)**. There are 38 sub-domains for each 2D segment, resulting in a total of 152 (38 × 4) binary indexes to parameterize all possible shale barrier configurations in 3D. Due to slight reduction in model resolution in comparison to setup 1, the total number of model parameters (i.e., binary indices) has only increased by approximately 50%. It should be noted that the overall shale proportion for both 2D and 3D cases are assumed to be less than 25%.



(a)



(b)

Fig. 3.10 – Illustration of the four zones and the corresponding basic shale unit for 2D models: (a) Setup 1 (b) Setup 2.

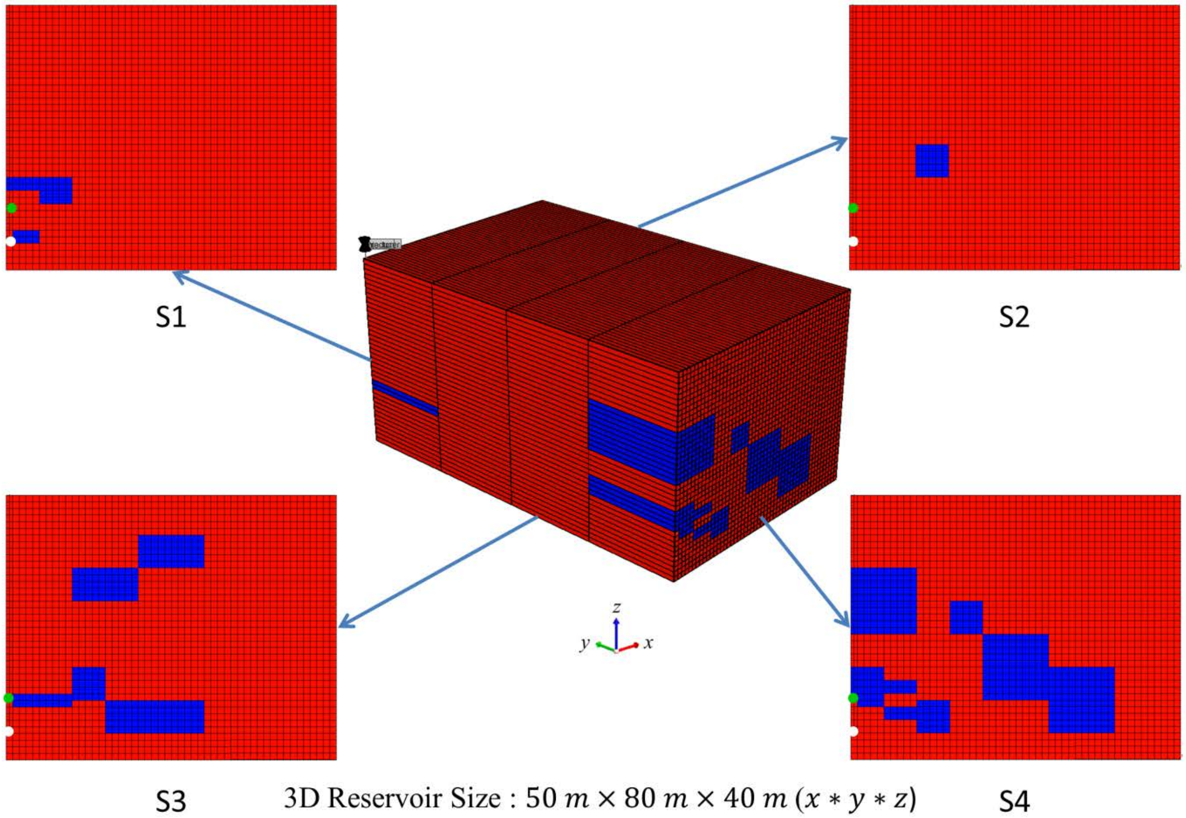


Fig. 3.11 – Illustration of 3D model and constitutive 2D segments.

Table 3.5 – Dimensions of basic shale unit and the number of sub-domains in each zone for 2D models (Setup 1).

	Dimensions of Basic Shale Unit (m^2)	Number of Sub-Domains
Zone 1	5 × 0.5	40
Zone 2	5 × 2	30
Zone 3	5 × 5	20
Zone 4	10 × 5	12
Total		102

Table 3.6 – Dimensions of basic shale unit and the number of sub-domains in each zone for 2D models (Setup 2).

	Dimensions of Basic Shale Unit (m^2)	Number of Sub-Domains
Zone 1	5 × 2	10
Zone 2	5 × 5	12
Zone 3	5 × 10	10
Zone 4	10 × 10	6
Total		38

It should be emphasized that shale barriers below the resolution of an individual basic shale unit and sub-domain are not included in this idealized representation of shale heterogeneity. This assumption seems reasonable because temperature data collected from the field observation wells would also suggest that chamber development is not overly sensitive to shale barriers that are smaller than the basic shale units. An example is shown in **Fig. 3.12**, where with gamma ray and temperature profiles of an observation well near Pad 107 are examined. Two shale barriers could be inferred from the logs: a thick one (> 2 m) at depth = 312 m and another thin one (< 1 m) at depth = 304 m. It is clear from the temperature profiles that only the thick shale has obstructed the steam chamber development for an extended period of time. Similar observations at other locations would support the decision of ignoring any shale barrier less than 0.5 m in thickness. Besides, inclined heterolithic strata (IHS) is not specifically modelled in this research, all shale barrier configurations are constructed by stacking several basic shale units laterally or vertically. It should be noted that the divisions of zones and subdomains may depend on the domain size and the specific shale barrier characteristics for a given SAGD reservoir. Sensitivity analysis that is similar to the one presented here should be conducted to assess the appropriate parameterization criteria.

It is observed that in some cases the steam chamber will reach the boundary on *x*-direction after about 10 years. Therefore, to avoid the impact of impermeable flow boundaries, only the first 10 years' simulation data of all synthetic cases are used in this research.

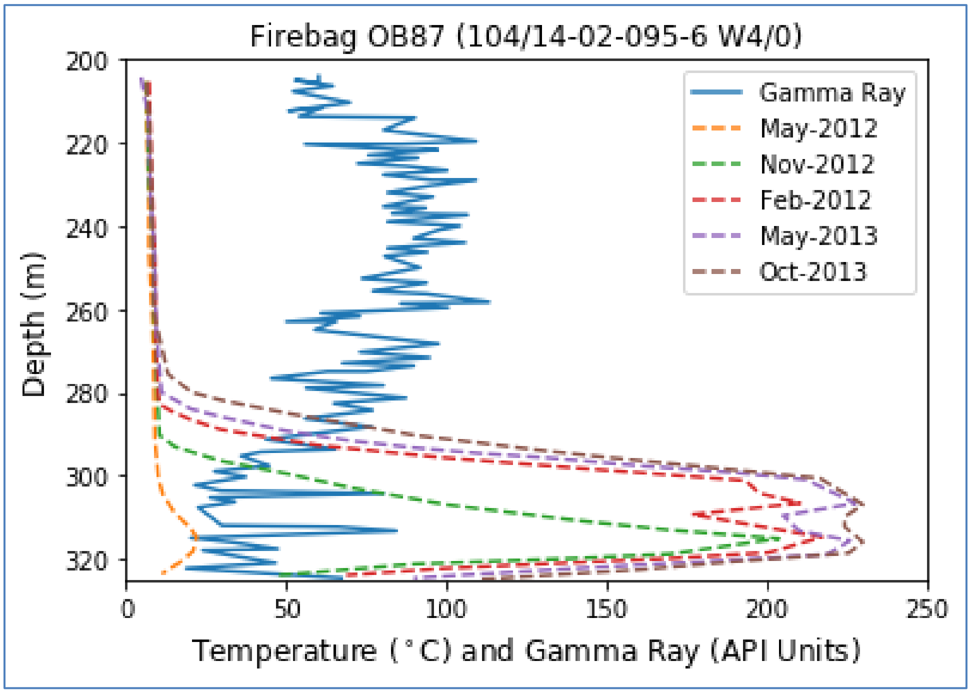
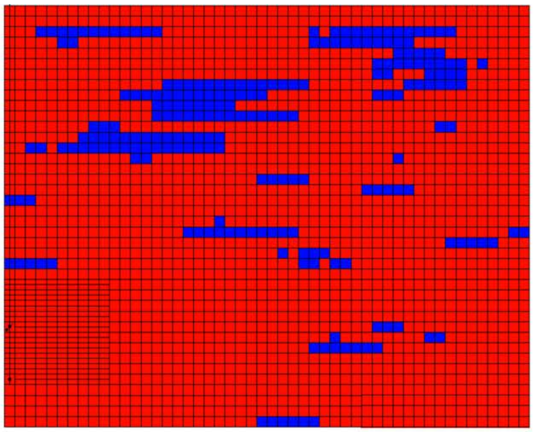


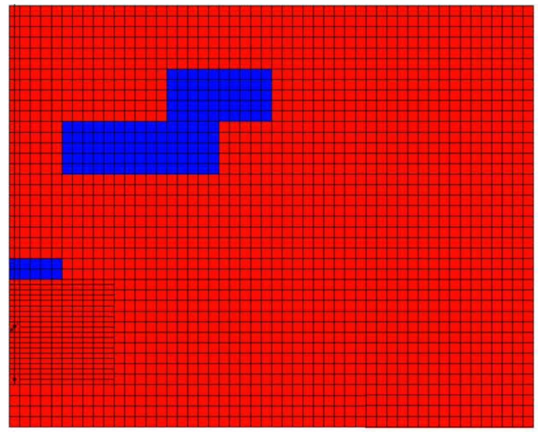
Fig. 3.12 – Temperature and gamma ray profiles for the observation well 87 in Firebag – data is extracted from Suncor's Firebag annual reports (Suncor Energy, 2012, 2013).

3.4 Validation of Parameterization Approach

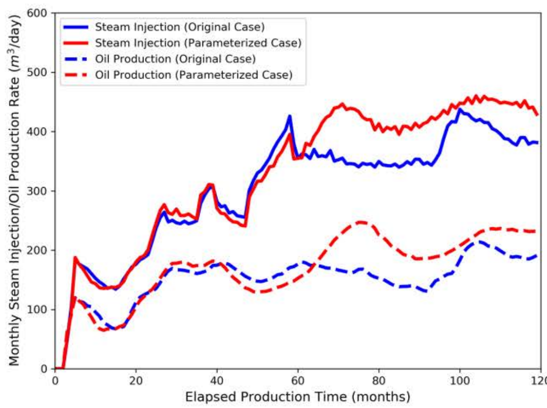
To illustrate the utility of the proposed parameterization approach for representing realistic reservoir shale distribution, two stochastic realizations of 2D shale configurations are generated via sequential indicator simulation (Deutsch, 1998). The shale barriers in these two realizations are subsequently categorized with the proposed parameterization approach. The parameterized cases are subjected to numerical simulation, which are then compared to the flow simulation results of the original stochastic realizations (**Fig. 3.13**). It is observed that the production trends from the parameterized cases are in good agreement with the original stochastic realizations. For Validation Case 1, the oil production profiles simulated using the original shale configuration and the parameterized shale configuration are similar for the initial 60 months. A slight discrepancy begins to emerge after 60 months. To understand this discrepancy, the oil saturation distributions for the two set-ups after 60 months and 72 months are compared in **Fig. 3.14**. It is clear that the steam chamber development for both scenarios are similar up to 60 months; however, at 72 months, the front of the steam chamber in the parameterized scenario develops better than the original configuration; this is due to the missing of certain small shale barriers located at the middle and the top right corner of the modeling domain after parameterization. This would explain why there is a slight drop in production rate for the original shale configuration after 60 months, while the opposite is observed for the parameterized case.



Validation Case-1 (a)



Validation Case-1 (b)



Validation Case-1 (c)

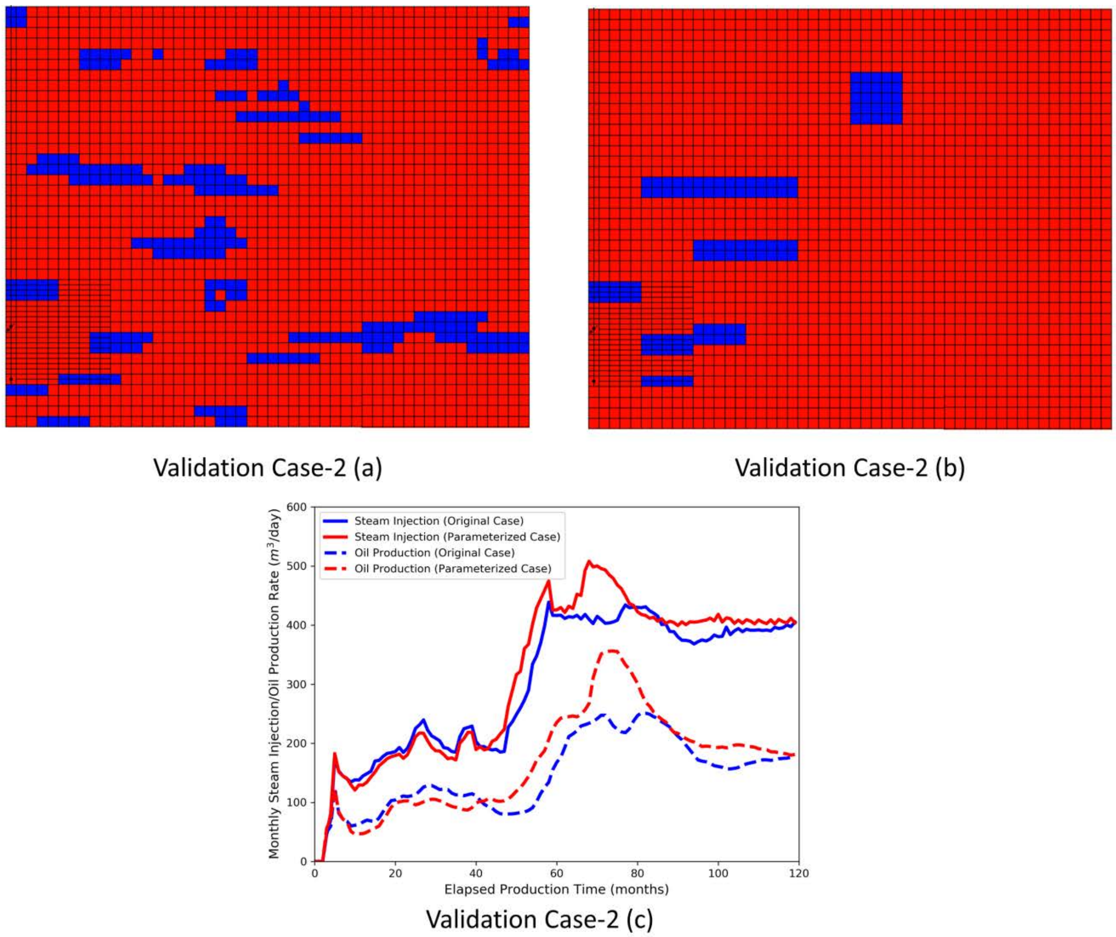


Fig. 3.13 – Validation Cases: (a) original shale configuration; (b) parameterized shale configuration; (c) numerical simulation results.

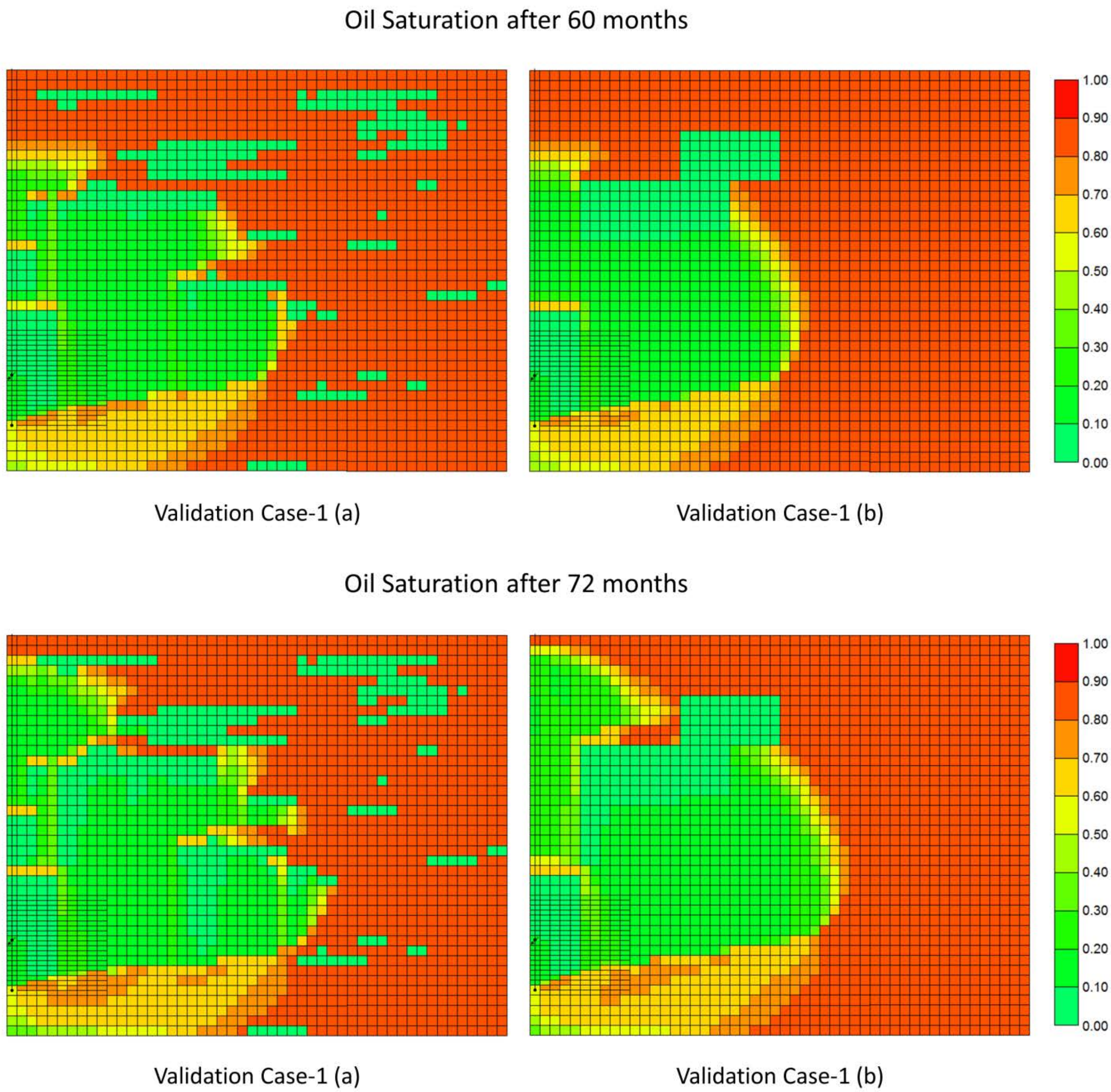


Fig. 3.14 – Oil saturation distribution of Validation Case-1: (a) original shale configuration; (b) parameterized shale configuration.

It should be acknowledged that a limitation of the current parameterization scheme is that it may not properly represent certain shale barriers smaller the size of a basic shale unit; however, it is anticipated that further optimization of the resolution of the basic shale units could potentially resolve this issue. Overall, the differences in cumulative production for all set-ups are minor. Despite of these minor deviations, the results support the idea that the proposed approach is effective in capturing the primary and aggregated influences of groups of shale barriers and generating reliable estimates of the recovery performance.

The results demonstrate the capability and flexibility of the proposed parameterization technique in capturing the complex heterogeneity exhibited by the shale barriers. The current model assumes that the reservoir heterogeneity can be sufficiently parameterized using the idealized basic shale units. In order to represent more complex distribution of irregularly-shaped shale barriers, refining the division of various domains and reducing the size of each basic shale unit may be required.

3.5 Summary

In this chapter, a set of SAGD field data is gathered and analyzed for Suncor's Firebag project. A homogeneous model is created, the petrophysical properties and certain operational constraints have been taken from representative values for several pads in Firebag.

A novel parameterization scheme is introduced to characterize shale heterogeneity. This is facilitated by dividing the entire reservoir domain into a number of sub-domains and formulating a set of basic shale units. Sensitivity analysis of the basic shale units (size, geometry, position) is conducted. Different parameterization setups for 2D and 3D models are proposed. The proposed approach will be used to construct synthetic cases employed in this thesis, and synthetic datasets containing shale heterogeneity and production time-series will be assembled from the corresponding numerical simulations. These datasets will be used in the following chapters to establish a set of AI-based data-driven models for exploring the correlation between shale heterogeneities and SAGD production.

CHAPTER 4: ESTIMATING SHALE BARRIER CONFIGURATIONS IN 2D EXAMPLES[2]

To estimate the effects of complex heterogeneities on SAGD production and infer the presence and distribution of heterogeneous shale barriers from SAGD production time-series data, a methodology is developed and tested on 2D synthetic cases. The methodology consists of two parts: (1) a (forward) forecasting regression model is trained to estimate SAGD production time series (outputs) from reservoir heterogeneities that are parameterized in terms of shale barrier configurations (inputs); and (2) an (inverse) optimization scheme, where the GA approach is adopted to infer unknown shale barrier configurations from a given SAGD production profile. During the GA procedure, the regression model trained in part (1) is used to generate a forward prediction, which is used to evaluate the fitness or objective function. The proposed workflow is depicted in a data flow diagram (**Fig. 4.1**), and details are presented in the following sections.

[2] The content in Chapter 4 of this thesis is derived from the following papers:

Zheng, J., Leung, J. Y., Sawatzky, R. P., & Alvarez, J. M. (2018a). A proxy model for predicting SAGD production from reservoirs containing shale barriers. *Journal of Energy Resources Technology*, *140*(12), 122903.

Zheng, J., Leung, J.Y., Sawatzky, R.P., and Alvarez, J.M. (2018b). An AI-based workflow for estimating shale barrier configurations from SAGD production histories. *Neural Computing and Applications*.

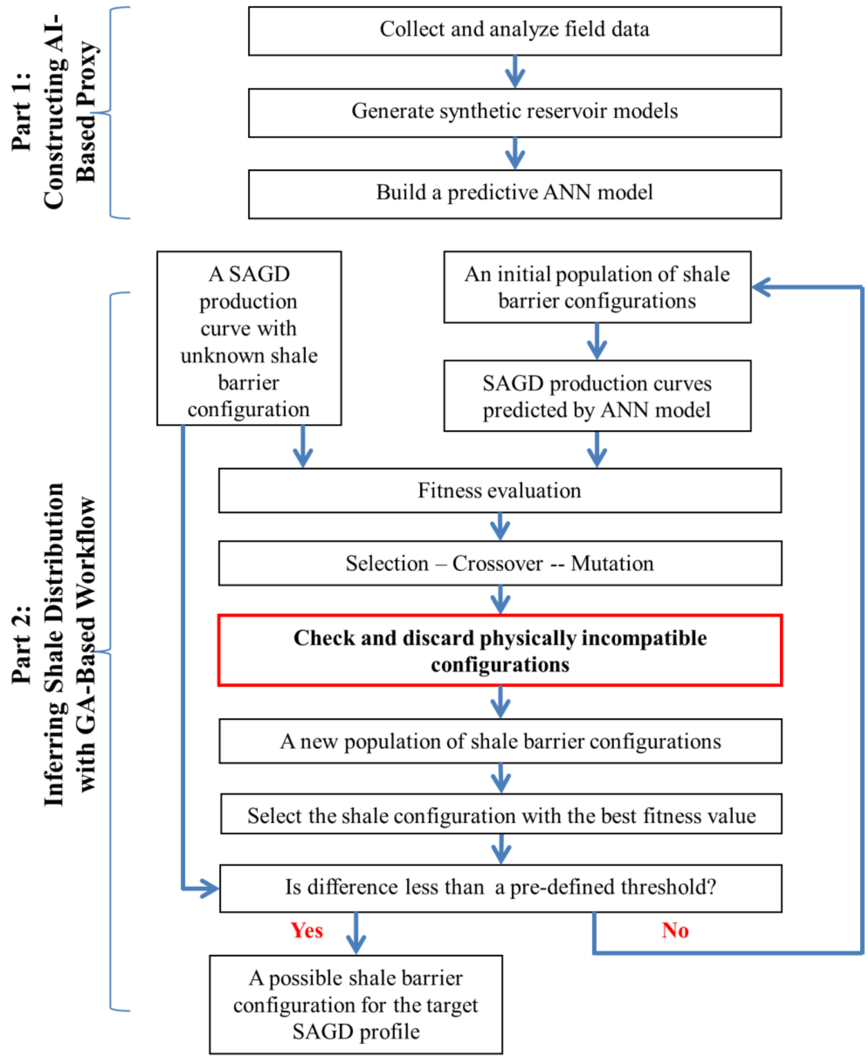


Fig. 4.1 – Proposed workflow for correlating SAGD production profiles with shale barrier configurations.

4.1 Forecasting SAGD Production from Reservoirs Containing Shale Barriers

4.1.1 Experimental Design

The construction of forward regression model requires a training dataset that sufficiently span the model parameter space. In this chapter, a set of 2D cases are constructed by varying the locations, intensity and dimensions of shale barriers. As shown in **Table 3.5**, there are a total of 102 sub-domains, rendering a total of $2^{102} \approx 5 \times 10^{30}$ combinations of different basic shale units. Therefore, experimental design is carried out to identify a representative subset of these shale configurations, which will form the dataset for the proxy modeling in the next stage. This subset should encompass sufficient characteristics regarding different configurations (i.e., size, geometry and position) of shale barriers, such that it can be used to train a proxy model capable of predicting SAGD production profiles for other sets of randomly-distributed shale barriers. A training dataset is constructed by considering four particular configurations of shale distribution:

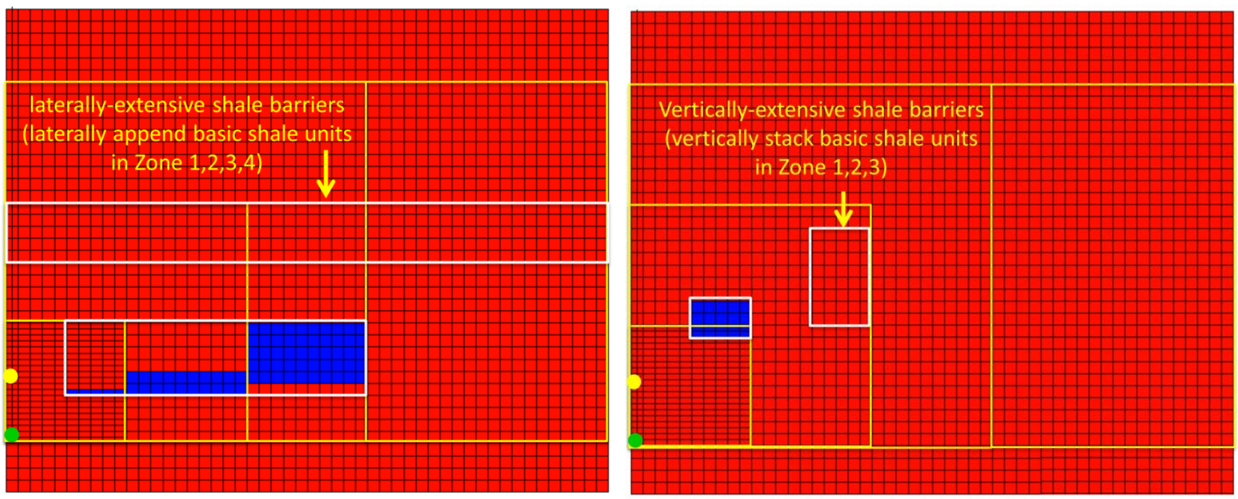
(1) Single basic shale unit – Considering each basic shale unit as the fundamental building block that constitutes the reservoir heterogeneity, 102 cases with only one basic shale unit are constructed.

(2) Laterally-extensive shale barriers – A laterally-extensive shale barrier is modeled by combining numerous basic shale units along the *x*-direction. This type of shale barrier typically hinders steam chamber advancement; oil rate would decline as the steam chamber encounters a shale barrier, and this decline continues until the steam chamber has advanced beyond the shale barrier. In order to capture the influences of shale barriers with varying lateral extent, a total of 267 cases are constructed by combining basic shale units from three representative regions in zones 1 to 4 to form laterally-extensive shale barriers. These three representative regions, as shown in **Fig. 4.2(a)**, encompass all possible lateral combinations of basic shale units located among zones 1, 2, 3 and 4.

(3) Vertically-extensive shale barriers – A vertically-extensive shale barrier is modeled by stacking numerous basic shale units along the *z*-direction. A total of 94 cases are constructed

by combining basic shale units from three representative regions in zones 1 to 3, as shown in **Fig. 4.2(b)**, to form vertically-extensive shale barriers. These three representative regions encompass all possible lateral combinations of basic shale units located among zones 1, 2 and 3.

(4) Multiple disjoint shale barriers – A total of 337 cases are constructed by randomly placing numerous disjoint shale barriers of different sizes in the reservoir domain.



(a) (b)

Fig. 4.2 – Representative regions for the models: (a) laterally-extensive shale barriers; (b) vertically-extensive shale barriers.

A training dataset consisting of a total of 800 2D cases of shale barrier configurations is prepared according to the aforementioned criteria. For the testing dataset, an additional 40 cases with randomly-distributed shale barriers are generated. The shale barriers in the testing cases are subsequently categorized according to the same parameterization scheme. They are used to test both the forward and inverse modeling workflows. A summary of the training and testing datasets is presented in **Table 4.1**.

Numerical simulation (CMG, 2015) is performed with each model, and the corresponding oil production profiles are recorded. The training and testing datasets are assembled such that the input variables correspond to the shale heterogeneities (i.e., a total of 102 binary location indices) and the operating constraints (i.e., steam injection rate and months on production, *t*), while the output variable is monthly oil production rate at *t*.

Table 4.1 - Summary of training and testing datasets.

<table>
<tr><td></td><td>Training Dataset</td><td>Testing Dataset</td></tr>
<tr><td>Number of SAGD Scenarios</td><td>800</td><td>40</td></tr>
<tr><td>Number of Data Records for Each SAGD Scenario</td><td>120 (monthly data)</td><td>120 (monthly data)</td></tr>
<tr><td>Total Data Records</td><td>96000</td><td>4800</td></tr>
<tr><td>Input Variables</td><td colspan="2">Month of production, Steam injection rate, 102 Location variables</td></tr>
<tr><td>Output Variable</td><td colspan="2">Oil production rate</td></tr>
</table>

4.1.2 Artificial Neural Network Modeling

The forward model should capture the high dimensional non-linear correlations between the input and output variables. For 2D datasets, there are 104 input variables, with an extensive training dataset of 96000 records. ANN is one of the most widely-adopted machine-learning techniques due to its versatility in generalization and mapping of non-linear correlations, provided that the training dataset is large enough; hence, it has also been successfully applied in many subsurface-related applications.

In this chapter, a type of artificial neural network (ANN) known as Multilayer Perceptron (MLP) is employed to approximate the non-linear relationship between the input and output variables. Once the ANN is trained, it can be used as a proxy model to predict oil production profiles corresponding to any arbitrary shale barrier configuration. The training dataset is used in a supervised learning procedure to estimate the unknown network parameters. An optimal structure is identified using the *n*-folds cross-validation method.

A MLP is a feedforward ANN that consists of an input layer, an output layer and at least one hidden layer. Each layer is fully connected to the next one, but connections within a single layer are not allowed. MLP is capable of approximating any non-linear functions, and, hence, it serves as a versatile regression model. The structure of the applied MLP is presented in **Fig. 4.3**. The input signal a_j^l at a neuron *j* in layer *l* is calculated according to **Eq. 4.1** (Nielsen, 2015):

$$a_j^l = f(\sum_k w_{jk}^l a_k^{l-1} + b_j^l) \quad (4.1)$$

The summation is performed over all *k* neurons in layer *l-1*; a_k^{l-1}denotes the output from neuron *k* in layer *l-1*; w_{jk}^l is the weight connecting between neuron *k* in in layer *l-1* and neuron *j* in layer *l*; and b_j^l is the bias for the neuron *j* in layer *l*. The output of neuron *j* in layer *l* is obtained by subjecting a_j^l to an activation function (*f*), which is the commonly-adopted hyperbolic tangent function, in **Eq. 4.2**:

$$f(a_j^l) = \tanh(a_j^l) \quad (4.2)$$

The hyperbolic tangent function scales between -1 and 1, it is differentiable everywhere which allows gradient-based optimization methods to be adopted for estimating the weights. All input and output variables are normalized to vary between -1 and 1 prior to ANN modeling. This step is needed to reduce the large disparity in scales of different data sources and to alleviate bias in the minimized solution as a result of values with overwhelmingly large magnitude. The objective function of the multilayer perceptron is defined as:

$$C = \frac{1}{2n_t}\sum_{\boldsymbol{x}} \left\| y(\boldsymbol{x}) - a^L(\boldsymbol{x}) \right\|^2 \quad \text{.......(4.3)}$$

The total number of training samples is denoted by n_t; the summation is performed over all training samples $\boldsymbol{x}$, and $y(\boldsymbol{x})$ denotes the target output corresponding to the input $\boldsymbol{x}$; L denotes the total number of layers in the network, and $a^L(\boldsymbol{x})$ denotes the network approximation corresponding to the input $\boldsymbol{x}$. This objective function is also known as the quadratic cost function or mean square error (Nielsen, 2015).

The backpropagation algorithm, which is a gradient-based supervised learning technique, is applied to adjust the weight and biases during the learning stage by minimizing the objective function. The network is first initialized with a set of weights; **Eq. 4.1** and **Eq. 4.2** are used to propagate the function signals from the input layer to the output layer. An error signal, which is computed as the mismatch between target and prediction at the output layer, is transmitted in the reverse direction, and the weights of each layer are updated. This entire updating procedure must be repeated for many epochs until the error is reduced below a predetermined threshold. Since backpropagation is a gradient descent method, the transfer functions must be differentiable.

Various optimization schemes, including the Newton's method and the gradient descent, are available. The Newton's method may converge rapidly to the local or global minimum, but its convergence is not guaranteed; on the other hand, the gradient descent technique is guaranteed to converge to the local or global minimum with proper parameters, but it is computationally demanding (Haykin, 2008). The Levenberg-Marquardt method was proposed as a compromise, and it is a reliable and efficient technique for updating the network parameters (Haykin, 2008). The updating equation is presented as follows:

$$\mathbf{w}_{m_L+1} = \mathbf{w}_{m_L} + (\mathbf{J}_{m_L}^T \mathbf{J}_{m_L} + \mu\mathbf{I})\mathbf{J}_{m_L}\mathbf{e}_{m_L} \quad (4.4)$$

Where **w** is the weight matrix, **J** is the Jacobian matrix, **I** is the identity matrix, μ is the combination coefficient, **e** is the error matrix, and m_L is the iteration step (Yu and Wilamowski, 2011).

To determine the optimal network structure, the *n*-fold cross validation procedure is implemented (Ma et al. 2015). In particular, 10 folds are used to assess the performance of eight network structures (i.e., the number of neurons in the hidden layer). Results of the 10-fold cross-validation are compared in **Table 4.2**. As shown in **Fig. 4.3**, the architecture of two hidden layers, each with 7 neurons, is selected due to its lowest corresponding normalized mean square error (*NMSE*). The MLP network with the optimal architecture is subjected to one final learning stage using the entire training dataset.

Table 4.2 - Average performance of the 10-folds cross-validation for different network structures.

Number of Neurons in Hidden Layer(s)	8	10	12	14	3-3
NMSE	0.0322	0.0323	0.0308	0.0328	0.0311
Number of Neurons in Hidden Layer(s)	4-4	5-5	6-6	7-7	8-8
NMSE	0.0341	0.0236	0.021	0.0175	0.0178

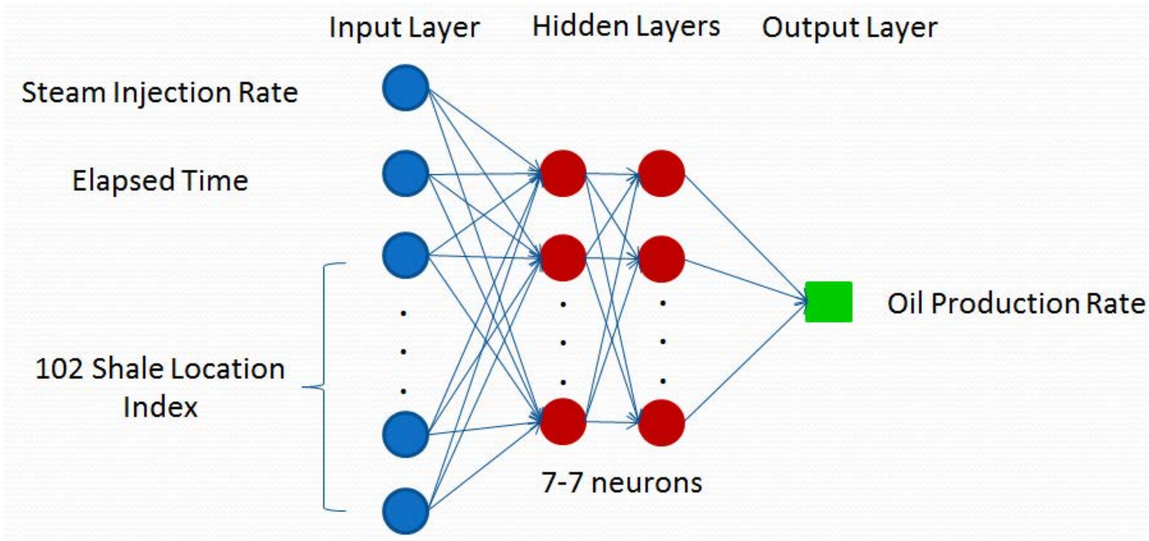


Fig. 4.3 – Schematic of the applied MLP network architecture with two hidden layers.

4.2 Estimating Shale Barrier Configurations from SAGD Production Histories

Once the forward AI-based proxy for correlating SAGD production with shale heterogeneities is established, it is integrated into a history-matching workflow for inferring shale barrier configurations for SAGD production histories.

4.2.1 Formulation of Objective Function from Production Time-Series Data

To quantify the mismatch in the production profile (time-series data) between the actual or target history (q_o^t) and the ANN model forecast (q_o^f), an objective function (C) is defined as half of the mean squared error (similar to **Eq. 3.1**):

$$C = \frac{1}{2m}\sum_{i=1}^{m}\left(q_{oi}^{t} - q_{oi}^{f}\right)^2 \quad (4.5)$$

Where m denotes the total number of monthly data records in a production profile (e.g., oil rate, SOR); i denotes a single time point of the time series; q_{oi}^{t} is the target production history at time i; and q_{oi}^{f} is the ANN model forecast at time i. In this way, the absolute difference between two production histories are compared and calculated as C.

4.2.2 Genetic Algorithm Based Optimization Scheme

As discussed in Chapter 3, reservoir heterogeneity in terms of shale barrier configuration is parameterized by a set of binary location indices. For a given SAGD production history with an unknown shale barrier configuration, the goal is to identify a shale barrier configuration, whose corresponding production profiles would match those of the target production history.

The inverse problem can be solved using a number of popular optimization algorithms, such as genetic algorithm (GA), simulated annealing (SA), and various gradient-based minimization algorithms. The unknowns for this optimization problem are the 102 binary location indexes; the ease of binary encoding renders the GA an ideal solution approach. It is implemented to perturb these location indices and propose a shale barrier configuration, which is subjected to the AI-based proxy to generate a model forecast.

A typical genetic algorithm requires a representation of the solution domain with binary strings (Whitley, 1994). Hence, the binary coding of shale barrier configuration is well suited for genetic representation in this study. For a minimization problem, a fitness function can be defined as the inverse of the objective function (1/C). The following GA-based workflow has been adopted using the Global Optimization Toolbox in Matlab® (The MathWorks Inc., 2014) for this problem:

(1) **[Initialization]** For a target SAGD production curve, a value of *C* is computed for each of the 800 cases in the training dataset. An initial population is generated by selecting *N* training cases with the smallest *C* value.

(2) **[Fitness]** The solution (i.e., shale barrier configuration) is encoded into binary strings, and the fitness of each solution is evaluated.

(3) **[New population]** The following steps are repeated until a new population is created:

   1) **[Selection]** Select two parents from the initial population according to their fitness. The solutions with better fitness values will have a bigger chance to be selected. A user-defined constraint is also applied at this step. In our case, any scenario with a shale volume proportion (i.e., total volume of shale barriers/reservoir volume) that is considered to be too large (> 25%) will not be selected.

   2) **[Crossover]** A crossover point along the binary string is randomly selected, and the values of two strings are exchanged up to this point to form new individuals (children).

   3) **[Mutation]** According to a pre-defined mutation probability, a portion of the binary strings in the new individuals will be changed randomly (0→1 or 1→0).

   4) The new individuals are placed into the new population. Repeat the selection-crossover-mutation steps until enough new individuals are created.

(4) **[Replace]** Replace the previous population with the new one and calculate its fitness value.

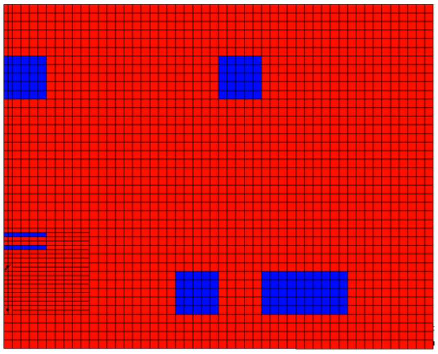
(5) **[Test]** If the stop criterion (e.g. maximum iterations, best fitness below a threshold) is encountered, select the solution with the highest fitness from the latest population to be the optimized solution. Otherwise, repeat from Step 2.

Up to this point, the entire workflow is established: the first part is to construct an AI-based neural network proxy for correlating SAGD production profiles with reservoir shale barrier configurations; the training dataset is derived from a set of synthetic numerical simulations in 2D; the second part involves a GA-based history-matching step for inferring shale barrier configuration from SAGD production history; the AI-based proxy developed in the first part is used to quickly generate a model forecast. A flowchart of this workflow is shown in **Fig. 4.1**.

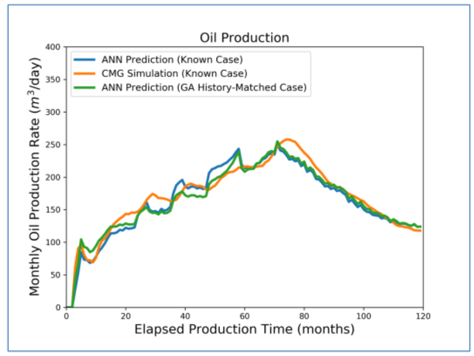
4.3 Testing Results and Discussions

The testing dataset constructed in **section 4.1.1** is used to assess the predictive performance of the ANN model and the accuracy of the GA workflow. These cases represent a diverse range of shale barrier configurations: the shale volume proportion ranges from 0.5% to 20%; the lateral and vertical extent of the shale barrier varies from 5 m to 40 m and 0.5 m to 20 m, respectively.

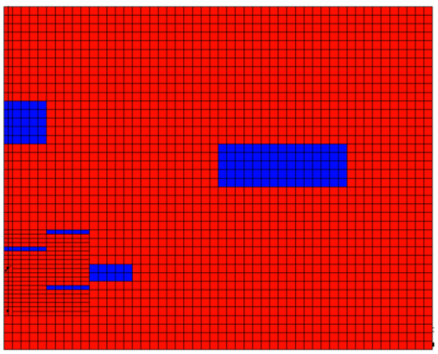
For the sake of brevity, only results for the 20 out of 40 shale barrier models in the testing dataset are presented in **Fig. 4.4**. These 20 cases are selected to illustrate 3 distinct groups of behavior, which will be discussed later. First, the known shale barrier configuration is shown in subplot (a); its corresponding oil production profile from numerical simulation is compared with that predicted using the AI-based proxy (or ANN model) in subplot (c). This comparison helps to illustrate the accuracy of the ANN prediction. Next, assuming that the oil production profile generated from numerical simulation in subplot (c) represents the actual history, the GA-based workflow can be employed to infer the shale barrier configuration that corresponds to this target production history. The final GA history-matched shale barrier configuration is shown in subplot (b), which can be compared with that of the known configuration; since the AI-based proxy is used to generate the model forecast, the corresponding production profile from ANN prediction is also shown in subplot (c). The idea is to assess how well the GA-based workflow could identify a possible shale barrier configuration by minimizing the mismatch between model forecast and target history. Finally, the GA history-matched model is subjected to numerical simulation to assess its true production response. The corresponding oil production and steam injection profiles are compared with the target production history in subplot (d). The idea is to evaluate the reliability of the entire workflow and the forecast predictability of the history-matched models.



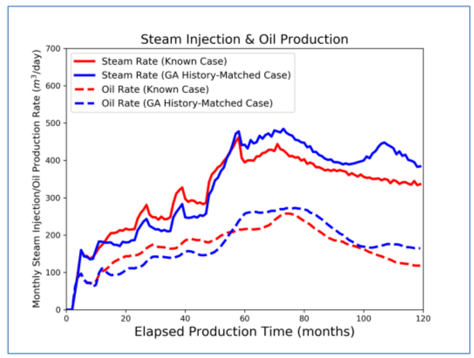
Case-1 (a)



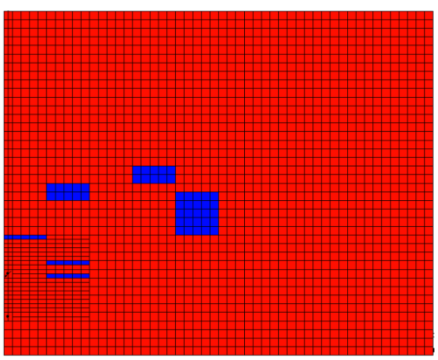
Case-1 (c)



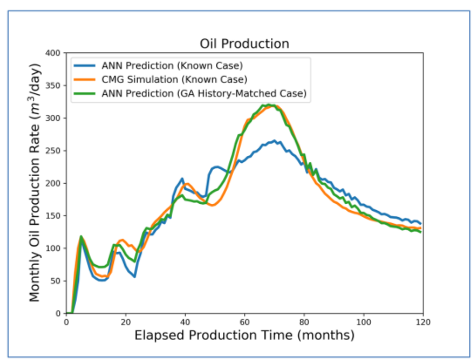
Case-1 (b)



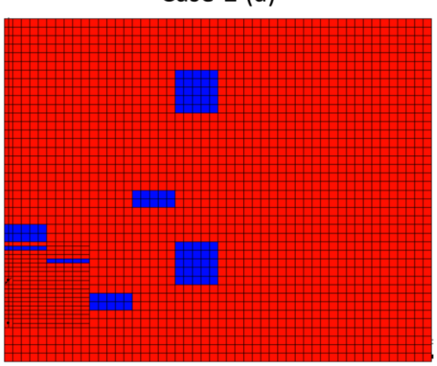
Case-1 (d)



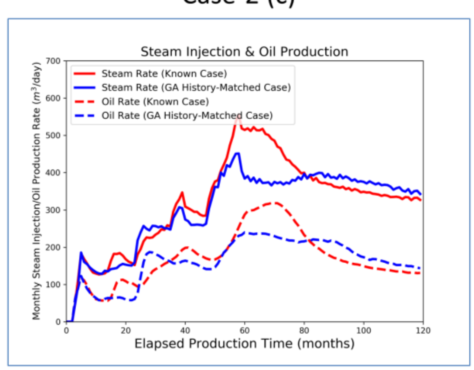
Case-2 (a)



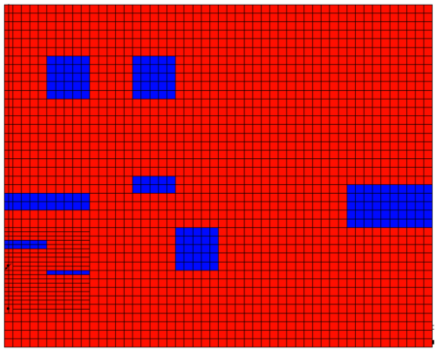
Case-2 (c)



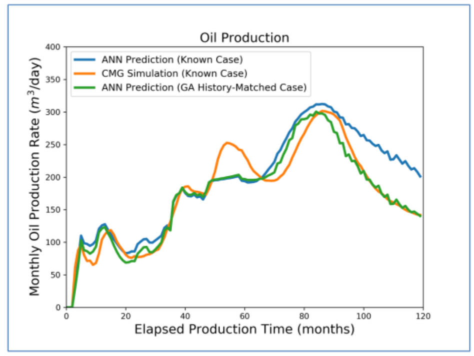
Case-2 (b)



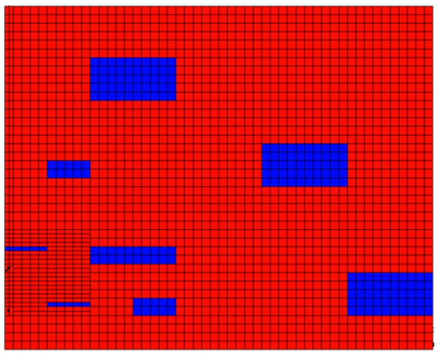
Case-2 (d)



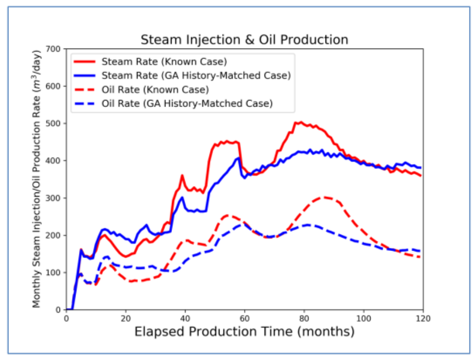
Case-3 (a)



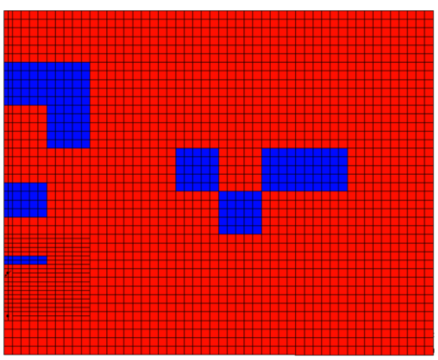
Case-3 (c)



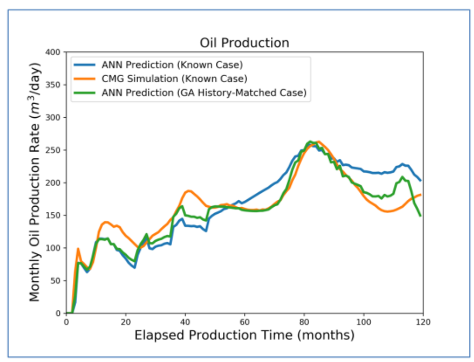
Case-3 (b)



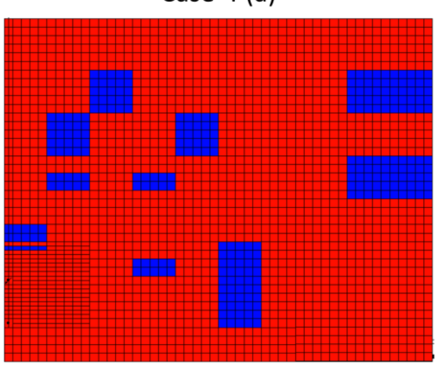
Case-3 (d)



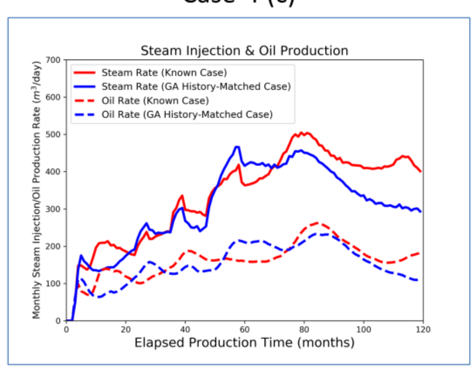
Case-4 (a)



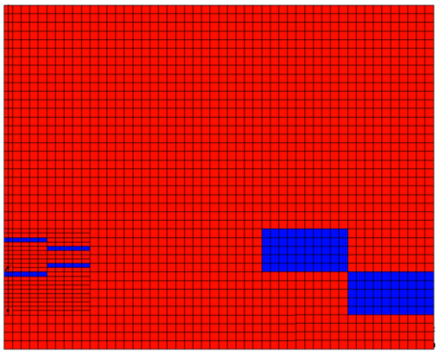
Case-4 (c)



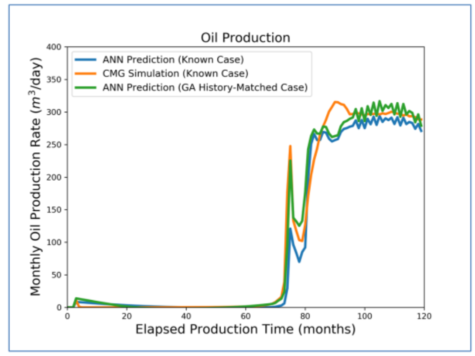
Case-4 (b)



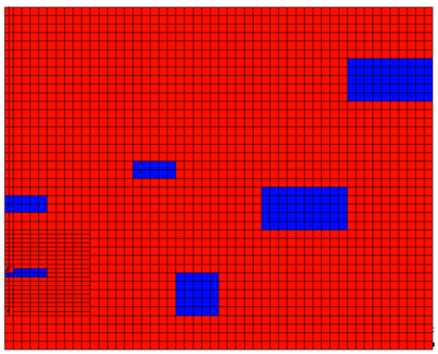
Case-4 (d)



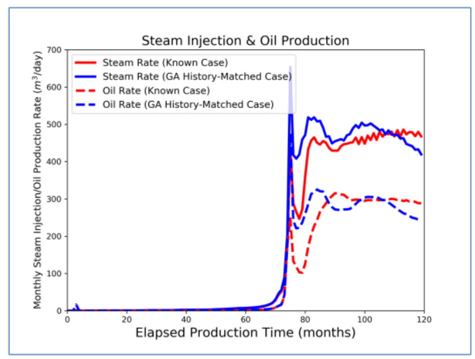
Case-5 (a)



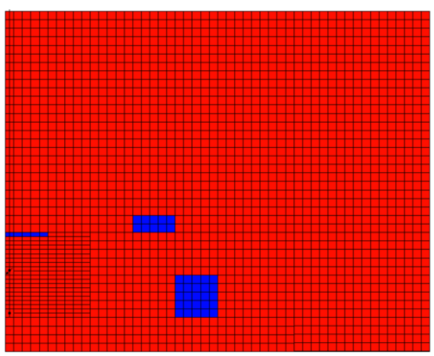
Case-5 (c)



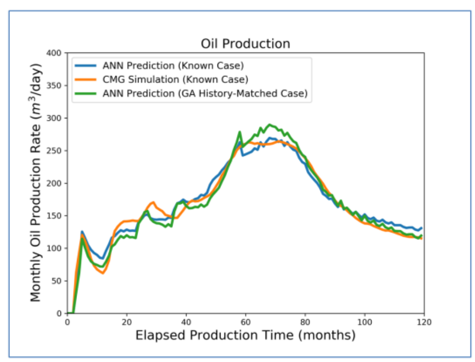
Case-5 (b)



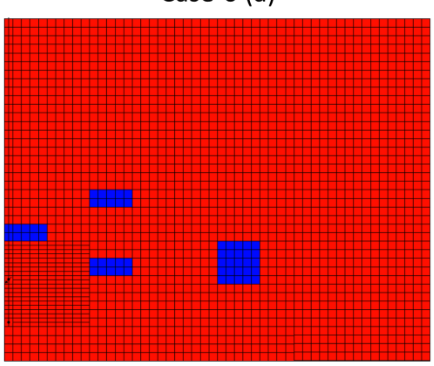
Case-5 (d)



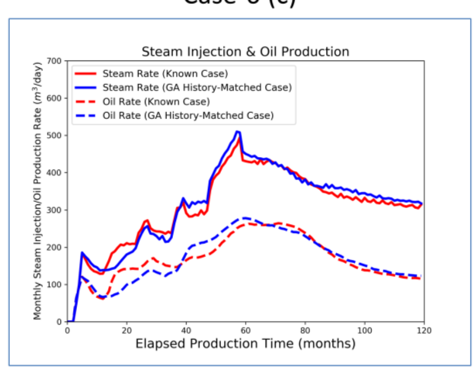
Case-6 (a)



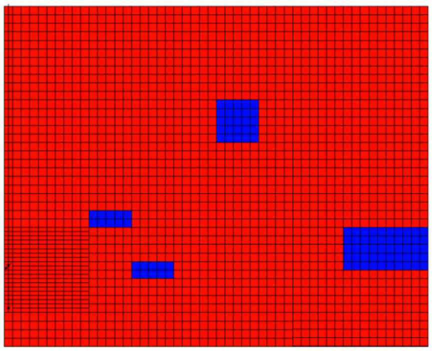
Case-6 (c)

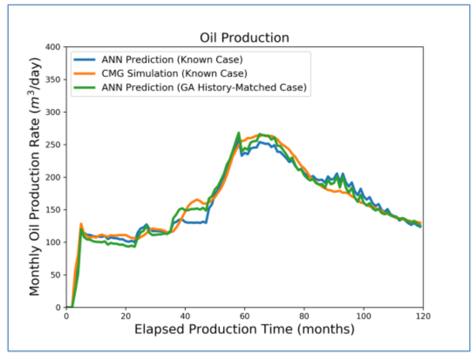


Case-6 (b)

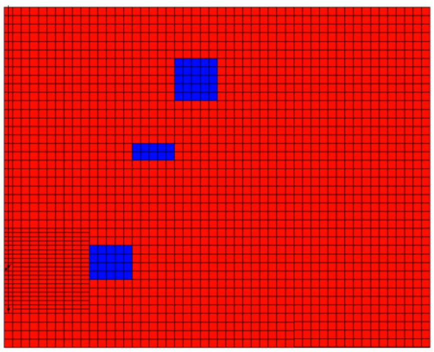


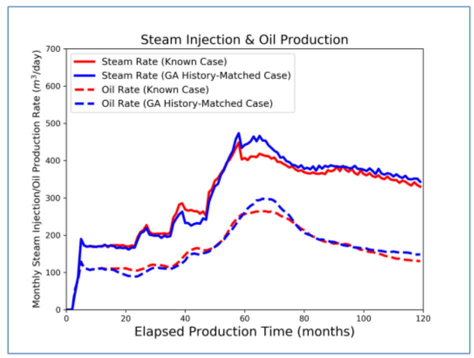
Case-6 (d)


Case-7 (a)

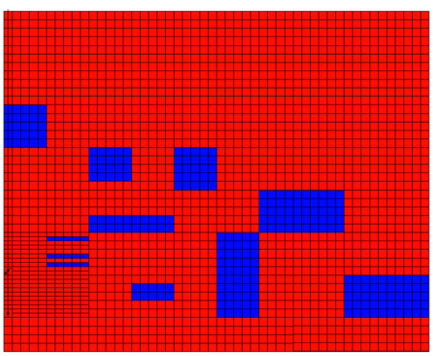


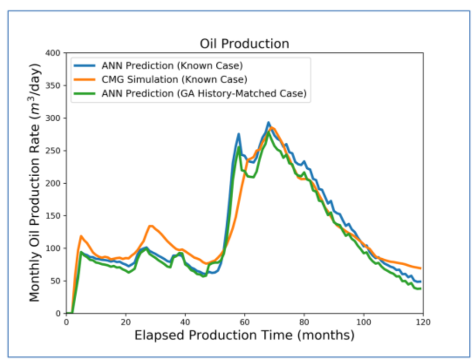
Case-7 (c)


Case-7 (b)

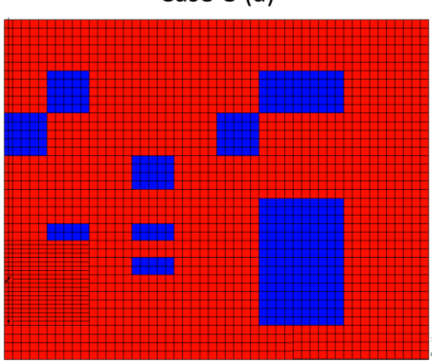


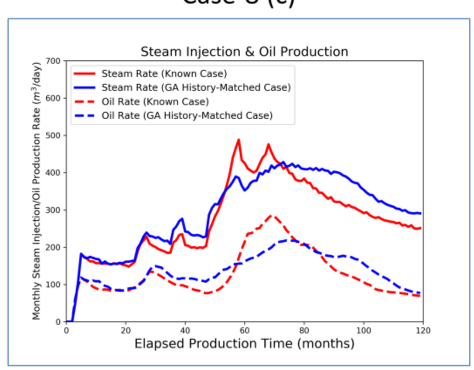
Case-7 (d)


Case-8 (a)

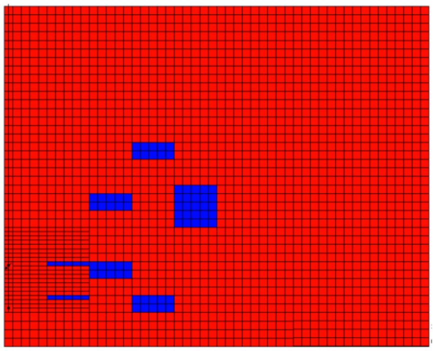


Case-8 (c)

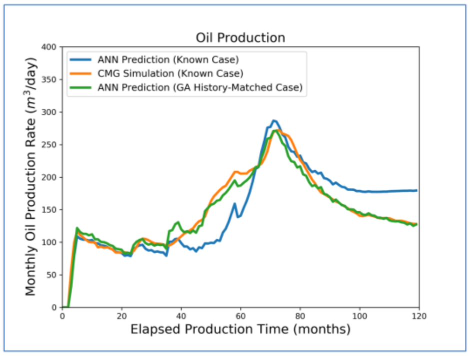

Case-8 (b)



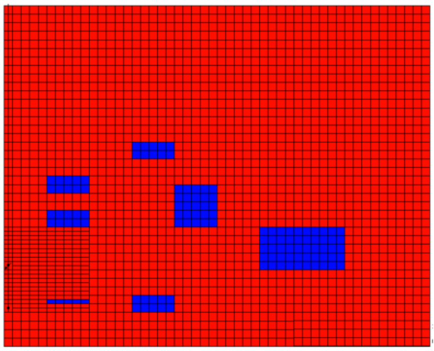
Case-8 (d)



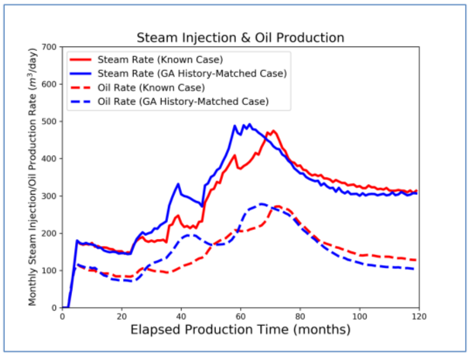
Case-9 (a)



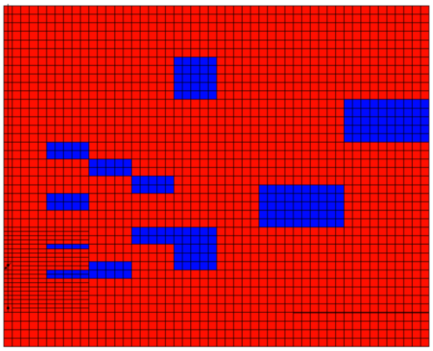
Case-9 (c)



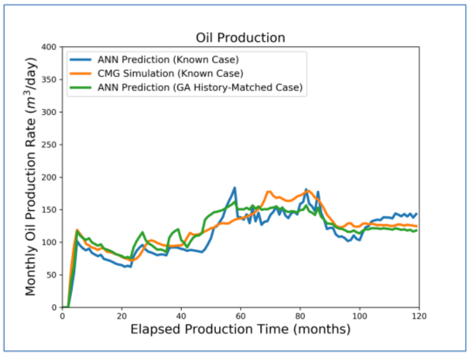
Case-9 (b)



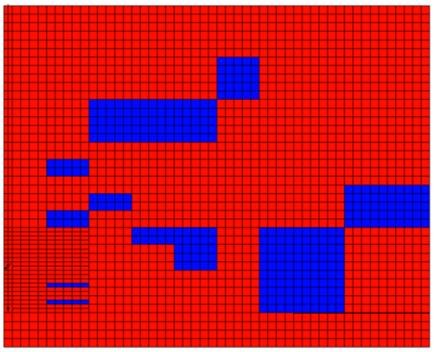
Case-9 (d)



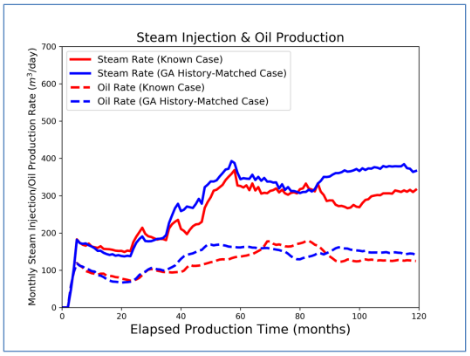
Case-10 (a)



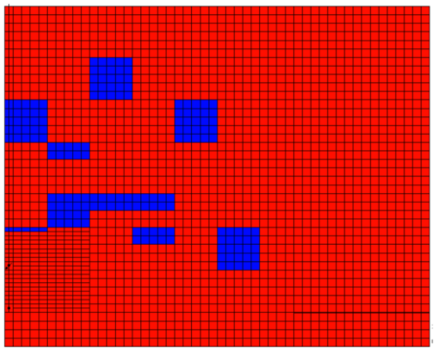
Case-10 (c)



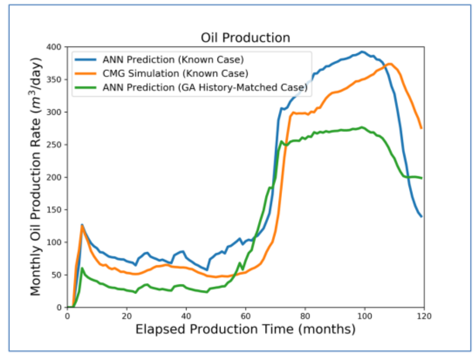
Case-10 (b)



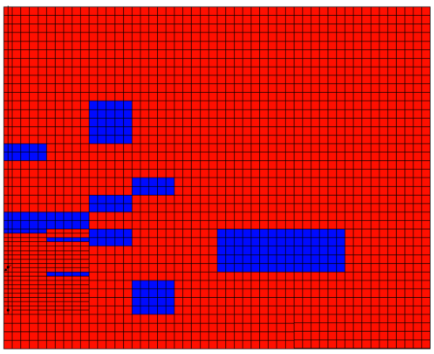
Case-10 (d)



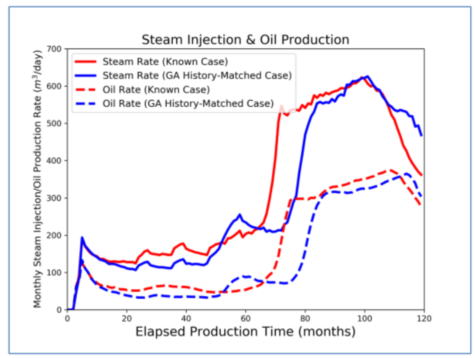
Case-11 (a)



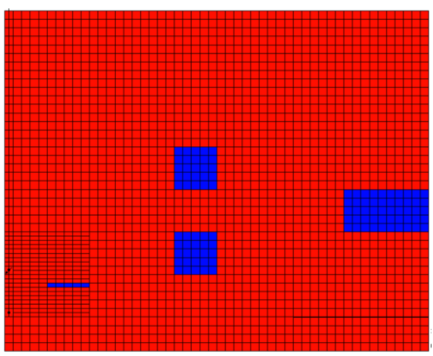
Case-11 (c)



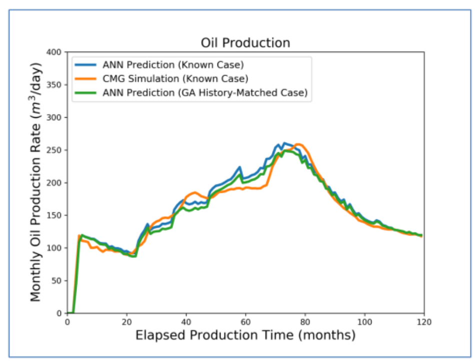
Case-11 (b)



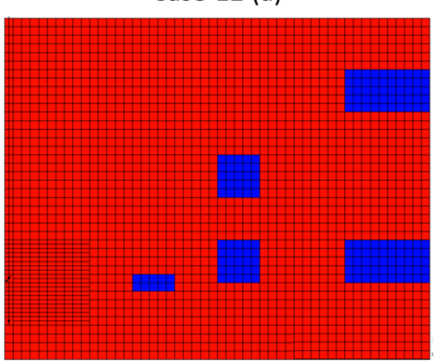
Case-11 (d)



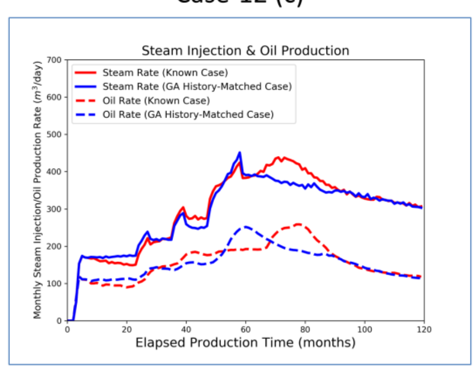
Case-12 (a)



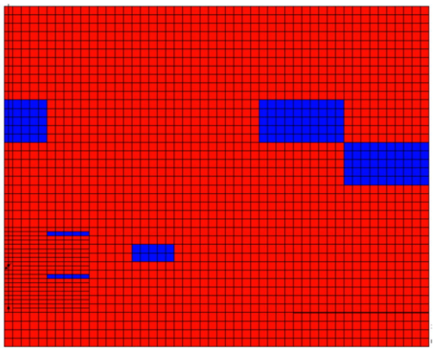
Case-12 (c)



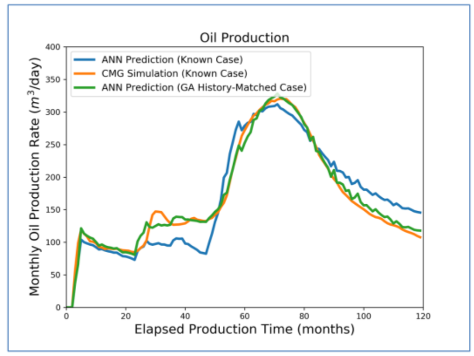
Case-12 (b)



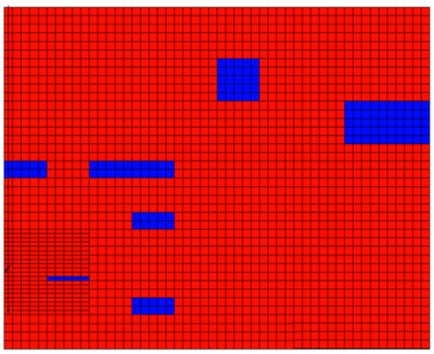
Case-12 (d)



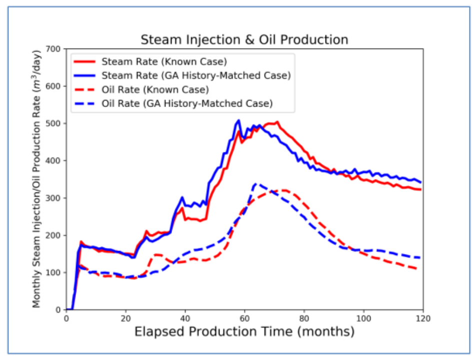
Case-13 (a)



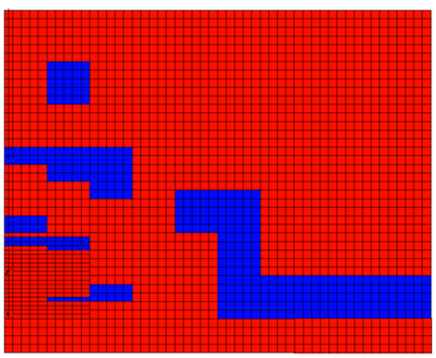
Case-13 (c)



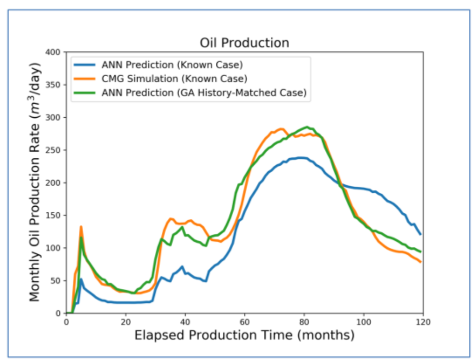
Case-13 (b)



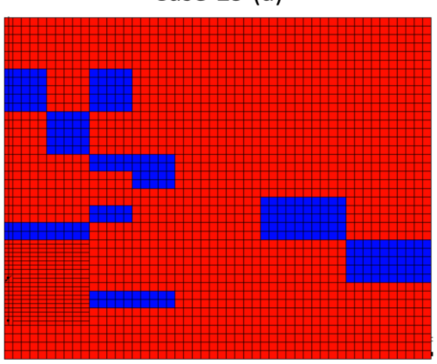
Case-13 (d)



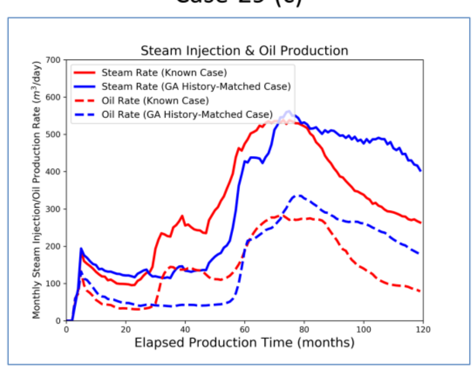
Case-29 (a)



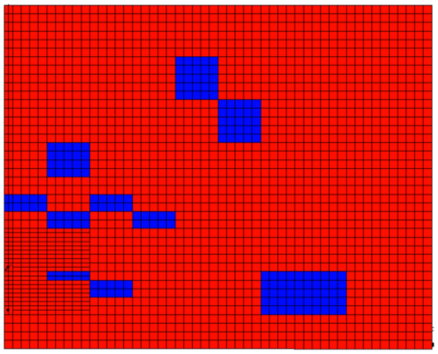
Case-29 (c)



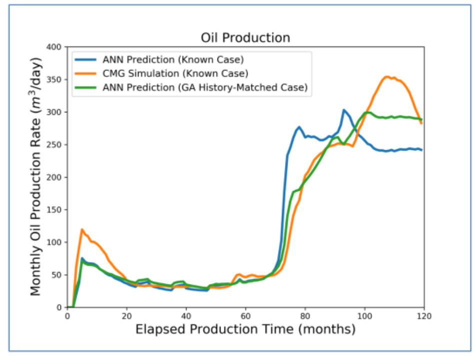
Case-29 (b)



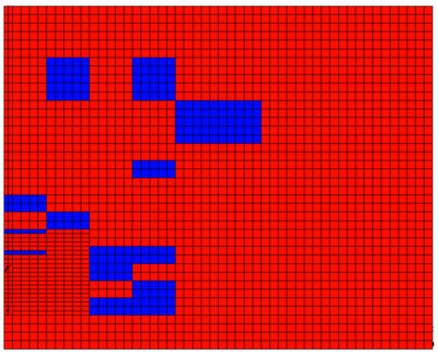
Case-29 (d)



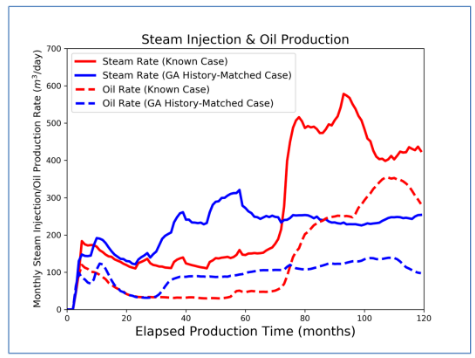
Case-30 (a)



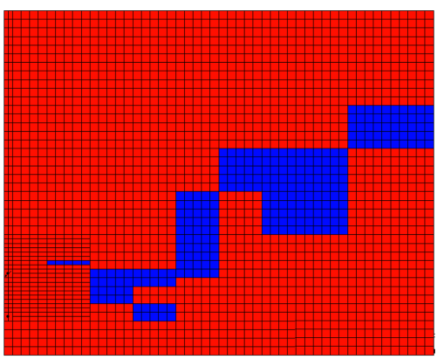
Case-30 (c)



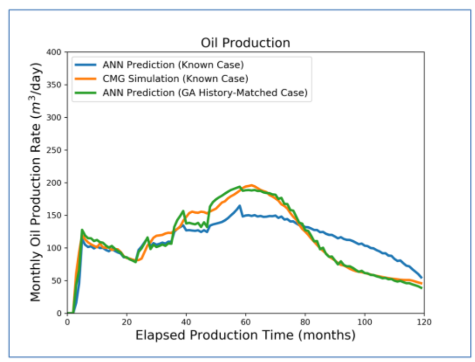
Case-30 (b)



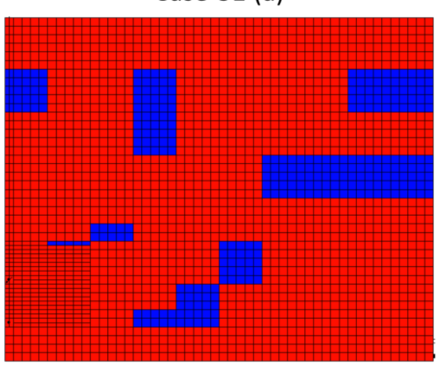
Case-30 (d)



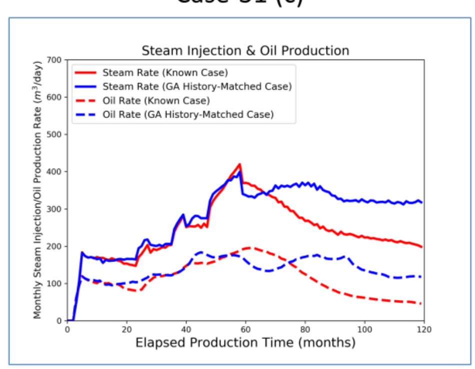
Case-31 (a)



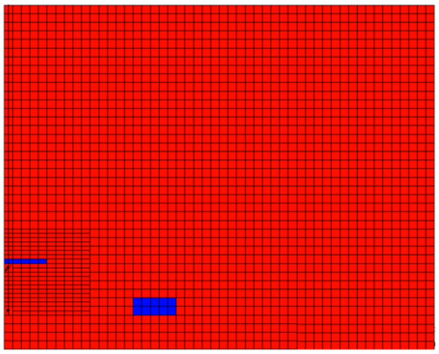
Case-31 (c)



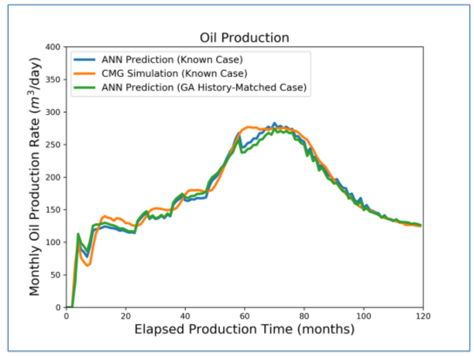
Case-31 (b)



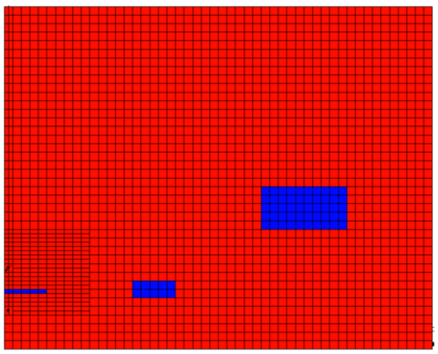
Case-31 (d)



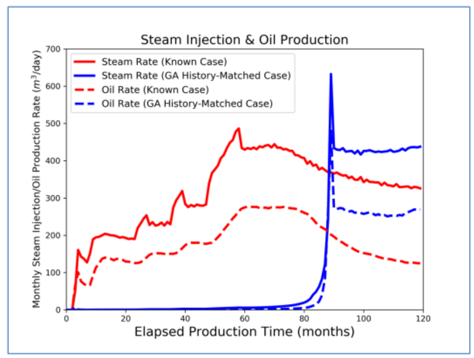
Case-34 (a)



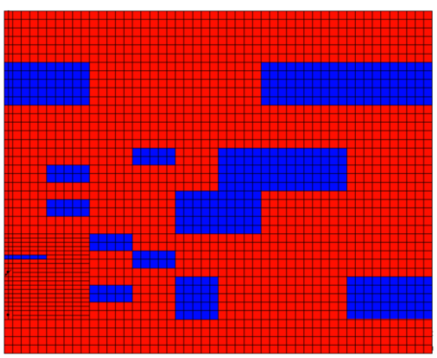
Case-34 (c)



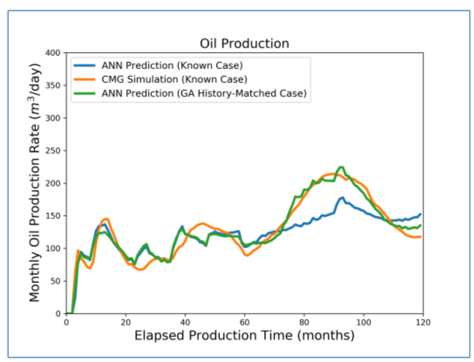
Case-34 (b)



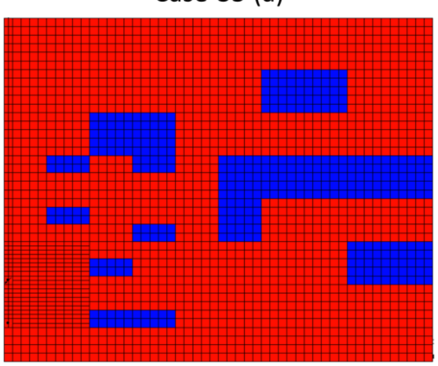
Case-34 (d)



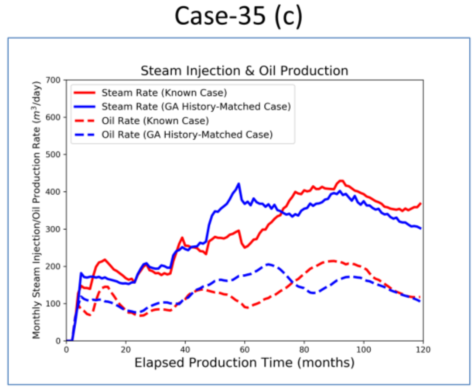
Case-35 (a)



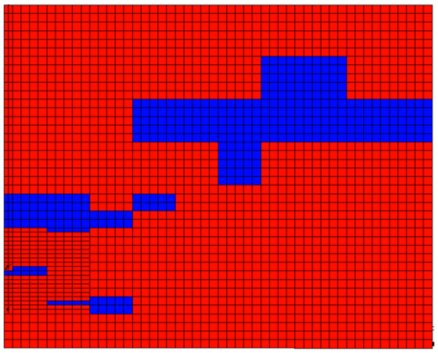
Case-35 (c)



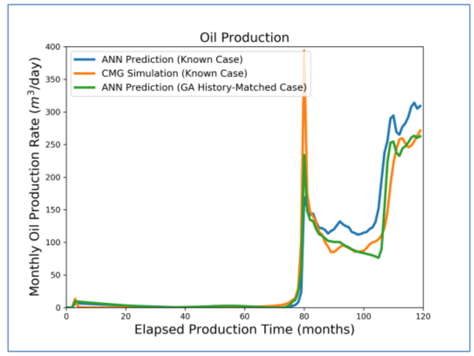
Case-35 (b)



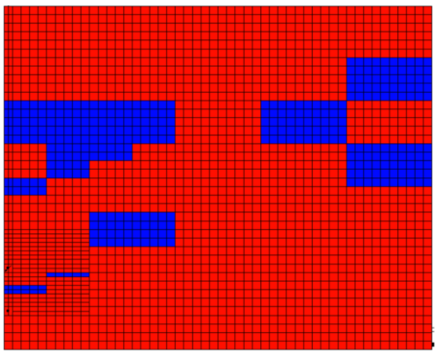
Case-35 (d)



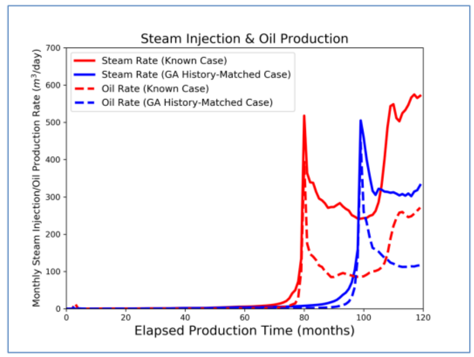
Case-36 (a)



Case-36 (c)



Case-36 (b)



Case-36 (d)

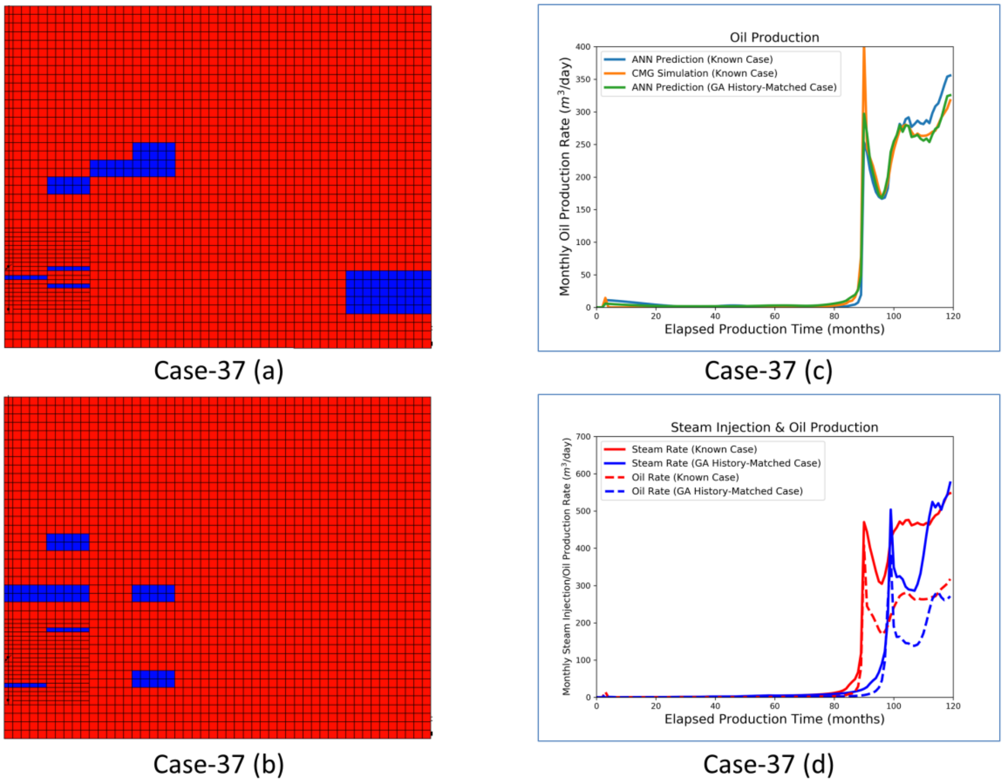


Fig. 4.4 – Testing results (20 out of a total of 40 cases are shown): (a) shale barrier configuration of the known case; (b) shale barrier configuration of the GA history-matched case; (c) oil production profiles from numerical simulation for the known case and from ANN predictions for both the known and GA history-matched cases; (d) oil production and steam injection profiles from numerical simulation for the known and GA history-matched cases.

To compare the shale barrier configuration of the known case and that of the GA history-matched case, the corresponding number of shale barriers and shale volume proportion are presented in in **Table 4.3**. To compare the similarity between the production time series (e.g., oil production and steam injection), all profiles corresponding to a particular model are normalized between their minimum and maximum values on a scale of [0, 1]. The similarity of any two normalized production profiles can be assessed by calculating their *C* values (as defined in **Eq. 4.5**). It is evident from **Eq. 4.5** that the maximum value for *C* is 0.5. This would only occur if the target profile is producing constantly at the maximum oil rate, while the other profile is producing constantly at the minimum oil rate (as determined over the entire set of forecast profiles). On the opposite side, a value of zero for C denotes a complete match between the two profiles. A set of C values have been calculated for each case, and the results are presented in **Table 4.3**.

(1) $C1$: comparing the oil production profiles between the numerical simulation and the ANN prediction of the known case [subplot (c)].

(2) $C2$: comparing the oil production profiles between the numerical simulation of the known case (target history) and the ANN prediction of the GA history-matched model [subplot (c)].

(3) $C3$: comparing the oil production profiles between the numerical simulation of the known case (target history) and the numerical simulation of the GA history-matched model [subplot (d)].

(4) $C4$: comparing the steam injection profiles between the numerical simulation of the known case (target history) and the numerical simulation of the GA history-matched model [subplot (d)].

Table 4.3 - Performance of proposed workflow for the testing dataset.

Case	Known Case		GA Forecasted Case		$C1$	$C2$	$C3$	$C4$
	# of shale barriers	Proportion of shale	# of shale barriers	Proportion of shale				
1	6	6.50%	7	5.88%	0.0019	0.0013	0.0077	0.0051
2	6	2.63%	7	4.25%	0.0044	0.0008	0.0084	0.0052
3	10	8.13%	10	9.75%	0.0084	0.0024	0.0088	0.0045
4	10	10.00%	12	13.38%	0.0080	0.0027	0.0104	0.0084
5	6	5.50%	7	7.50%	0.0013	0.0007	0.0037	0.0021
6	3	1.88%	4	2.75%	0.0011	0.0033	0.0033	0.0008
7	4	4.75%	4	2.75%	0.0011	0.0019	0.0011	0.0009
8	14	12.88%	12	16.25%	0.0042	0.0102	0.0084	0.0045

9	7	3.50%	7	5.88%	0.0081	0.0028	0.0071	0.0038
10	13	10.88%	14	16.00%	0.0066	0.0120	0.0115	0.0090
11	11	8.13%	14	9.38%	0.0089	0.0132	0.0108	0.0105
12	4	5.13%	5	8.00%	0.0016	0.0055	0.0085	0.0017
13	6	7.00%	8	6.38%	0.0039	0.0023	0.0028	0.0013
14	5	1.75%	5	5.25%	0.0022	0.0025	0.0035	0.0013
15	4	4.75%	3	2.25%	0.0018	0.0031	0.0033	0.0012
16	6	5.00%	6	4.50%	0.0030	0.0035	0.0035	0.0013
17	5	4.13%	6	4.63%	0.0027	0.0080	0.0100	0.0034
18	3	3.13%	3	3.50%	0.0013	0.0022	0.0028	0.0011
19	5	6.13%	7	8.63%	0.0051	0.0178	0.0053	0.0052
20	11	11.00%	15	12.13%	0.0010	0.0658	0.0625	0.0379
21	4	4.25%	6	6.00%	0.0022	0.0030	0.0062	0.0033
22	5	3.25%	5	5.63%	0.0030	0.0062	0.0124	0.0053
23	7	9.50%	7	5.00%	0.0195	0.0089	0.0090	0.0045
24	7	5.13%	10	11.75%	0.0048	0.0050	0.0076	0.0049
25	6	5.00%	8	9.13%	0.0011	0.0046	0.0041	0.0024
26	5	6.88%	5	5.25%	0.0008	0.0110	0.0105	0.0040
27	6	6.50%	6	6.50%	0.0022	0.0069	0.0028	0.0016
28	6	4.63%	7	9.75%	0.0035	0.0084	0.0078	0.0034
29	21	16.00%	13	12.75%	0.0120	0.0013	0.0252	0.0196
30	12	8.75%	15	9.75%	0.0115	0.0025	0.0487	0.0479
31	11	13.38%	11	14.88%	0.0112	0.0013	0.0339	0.0145
32	15	16.50%	18	14.75%	0.0179	0.0023	0.0415	0.0266
33	14	11.00%	13	15.13%	0.0123	0.0062	0.0155	0.0105
34	2	0.63%	3	3.13%	0.0003	0.0004	0.0706	0.0979
34b	2	0.63%	1	0.13%	0.0003	0.0003	2.90E-05	4.80E-05
35	17	20.63%	15	18.50%	0.0084	0.0021	0.0211	0.0078
36	18	16.50%	18	16.88%	0.0030	0.0010	0.0159	0.0296
37	8	4.88%	8	4.75%	0.0011	0.0004	0.0148	0.0147
38	8	5.88%	8	7.13%	0.0058	0.0010	0.0139	0.0111
39	10	7.00%	10	8.38%	0.0065	0.0025	0.0213	0.0101
40	10	12.13%	10	12.00%	0.0063	0.0234	0.0255	0.0201

As mentioned earlier, the performance of all 40 test cases can be divided into three distinct groups. For Cases 1-28, the results in **Fig. 4.4** and **Table 4.3** suggest that (1) the AI-based proxy is accurate in correlating the shale barrier configuration and the corresponding production time series (as illustrated by the low value of $C1$); (2) the GA-based workflow is useful for minimizing the mismatch between model forecast and target history (as illustrated by the low value of $C2$); (3) as a result of (1) and (2), the profiles obtained from numerical simulation for the history-matched model closely resemble those derived from ANN prediction, demonstrating the reliability of the entire workflow. It is observed that the shale barrier configurations proposed by the GA workflow have successfully captured many essential features exhibited by the known ones, as confirmed by the comparison in the shale volume proportion, and the lateral and vertical distribution of the shale barriers between the two sets of configurations. A visual comparison of the GA history-matched model with the known configuration would reveal some discrepancies: the precise location and distribution of various shale barriers appear to be slightly different between the two models. This deviation serves to illustrate the ill-posed nature of this inverse problem (Kabanikhin, 2008). According to Hadamard (1902), a problem is well posed if a unique solution that depends continuously on the data and model parameters exists. In this case, it is probable that, among all the possible shale barrier configurations ($2^{102} \approx 5\times10^{30}$), many would introduce a similar influence on the SAGD production performance. As a result, the final solution would likely be dependent on the initial population of training cases selected in the initialization procedure for the GA-based workflow.

For Cases 29-33, some discrepancy in the production profiles between the known case and the GA history-matched case ($C3 > 0.01$) can be detected. The prediction accuracy of the ANN model is also relatively low ($C1 > 0.01$). It is anticipated that improving the predictive capability of the ANN model would enhance the overall accuracy of the workflow, and this can be achieved by expanding the spanning of the model parameter space in the training dataset.

For the next group of Cases 34-40, despite the high prediction accuracy of the ANN model ($C1 < 0.01$ and $C2 < 0.01$), the performance of the GA-based workflow is poor ($C3 > 0.01$ and $C4 > 0.01$, except for Case 35). This implies that, despite the ability of the GA algorithm to propose a shale barrier configuration that matches the production history predicted by the AI-based proxy (ANN model), the numerically-simulated production profiles corresponding to this history-matched model deviate from the actual history. Upon further inspection, a common feature among these four cases is that there are shale barriers located closely above the injector or in between the injector-producer well pair in the known configuration; however, the GA history-matched model could capture only approximately or partially the exact location of these shale barriers, especially along the vertical direction. As discussed in Chapter 3, steam chamber development is highly sensitive to shale barriers in the near-well region. A slight variation in shale location in the area close to the well pair should have exacted a significant change to the injection and production profiles.

For Case 35, the final history-matched model has not captured the shale barrier that is located right above the injector, which is responsible for the poor performance, as evidenced by the high $C3$ value. For Cases 36-40, the high values of $C3$ and $C4$ can be explained by the minor mismatch in the arrangement of shale barriers in between the well pair. Except for a moderate shift in the production profile between the known case and the history-matched case, both profiles in the subplot (d) are exhibiting a similar trend. Considering that the values of $C3$ and $C4$ are only slightly larger than those for Cases 1-33, the history-matching results for all three cases are deemed satisfactory.

However, the same conclusion cannot be drawn for Case 34. It is the only case for which the history-matching process is incapable of accurately inferring the presence or absence of a shale barrier in between the well pair. A shale barrier was incorrectly placed right above the producer in the history-matched model, resulting in the high values of $C3$ and $C4$. However, why would the corresponding ANN prediction be similar to the target history? When proposing new configurations during the GA algorithm, it is the steam injection profile for the known case (history) that is imposed as input. In reality, though, the steam injection profile for the proposed configuration should be much different. In Case 34, the steam injection rate for the history-matched configuration, which consists of a shale barrier located in between the well pair, would have been much lower than the target history where the shale barrier is located above the injector (compare the numerical simulation results for Case 4 and Case 5). Clearly, a high steam injection rate is physically incompatible with this particular shale barrier configuration; its corresponding ANN prediction turns out to be extrapolating from a portion of the model parameter space, as none of the cases in the training dataset contains this type of shale barrier configuration under a high steam injection rate (particularly during early time). This explains why the ANN predicted profiles of these GA history-matched cases are similar to the target history (small $C2$).

A potential solution is to implement, at the end of Step 3 in the GA workflow in **section 4.2.2**, a constraint/check-point to discard proposed configurations that are physically incompatible with the injection profile (or have not been represented in the training dataset). To this end, clustering analysis is implemented with following steps to ensure compatibility between the imposed steam injection profile and a proposed shale barrier configuration:

(1) **[Feature extraction] Eq. 4.5** is used to compute the dissimilarity in steam injection rate between each training case and the homogeneous base case (with m = 12) for each of the 10 years. A positive C value is assigned, if the cumulative steam injection of training case is

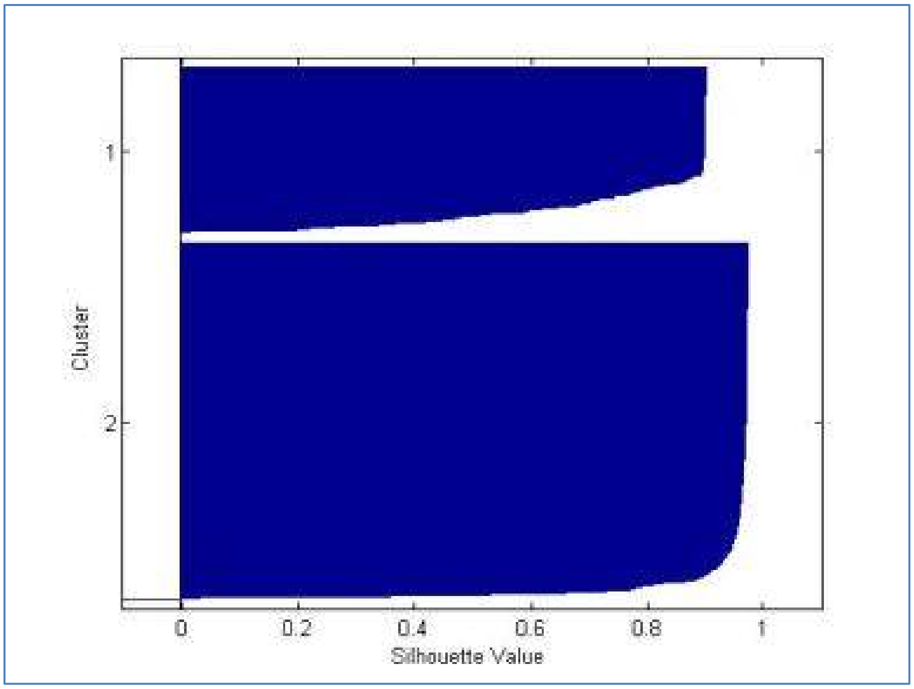
higher than homogeneous base case for that particular year; otherwise, a negative C value is assigned. In the end, a feature vector of 10 C values is computed for each training case.

(2) **[Clustering analysis]** K-means clustering is applied to the feature vectors computed in step 1. Two clusters could be discovered with high confidence (as evidenced by the high silhouette value in **Fig. 4.5**). All cases in Cluster 1 experience low steam injection for an extended period upon start-up due to the presence of shale barrier(s) in between the well pair.

(3) **[Supervised classification]** The clustering results are used to train another AI-based model using the decision trees (Kingsford & Salzberg, 2008) to classify any given shale barrier configurations (represented by 102 binary location indexes) into one of the two categories (i.e., with or without shale barrier(s) in between the well pair). The accuracy of the classification corresponding to the trained decision-tree model is summarized by the confusion matrix in **Fig. 4.6**. Given the highly distinct features that are observable with each category, the ensuing classification performance is essentially perfect. Alternative approaches, such as support vector machine (SVM) (Nobel, 2006), logistic regression (McDonald, 2009), could also be adopted for formulating the classification model.

(4) **[Integration with the GA workflow]** At the end of Step 3 in the GA workflow in **section 4.2.2**, the classification model developed in Step #3 is integrated to discard any configuration in the proposed population that is physically incompatible with the injection profile. This update is shown as the red box in **Fig. 4.1**.

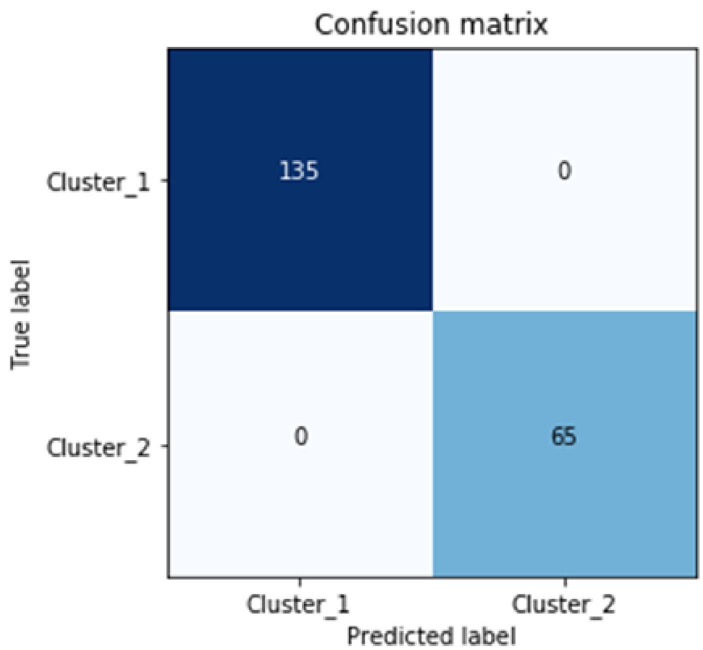
The analysis of Case 34 is repeated using this revised workflow, and the results are presented in **Table 4.3** and **Fig. 4.7** as Case 34(b). Dramatic improvement in terms of $C3$ and $C4$ values can be observed.



Cluster	# of Cases	Feature
1	253	Long period of low injection rates
2	547	Normal injection rates

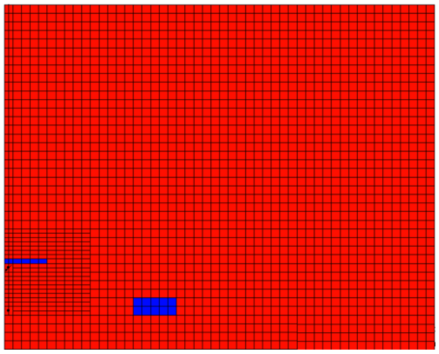
Mean Silhouette Value: 0.9003

Fig. 4.5 – K-means cluster analysis of the steam injection profiles of 800 training cases.

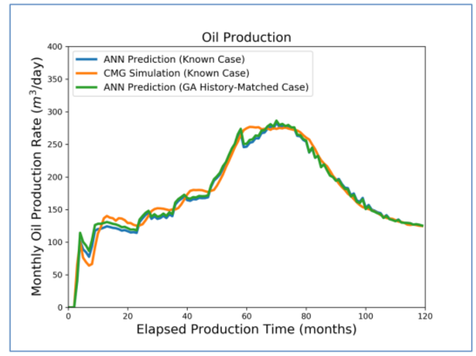


# of Training Cases	600	Precision	1.0
		Recall	1.0
# of Testing Cases	200	f-score	1.0

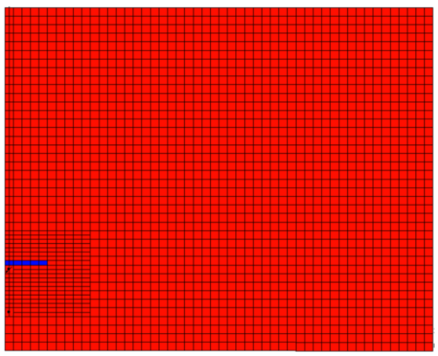
Fig. 4.6 – Confusion matrix of the decision-tree model of 200 testing cases.



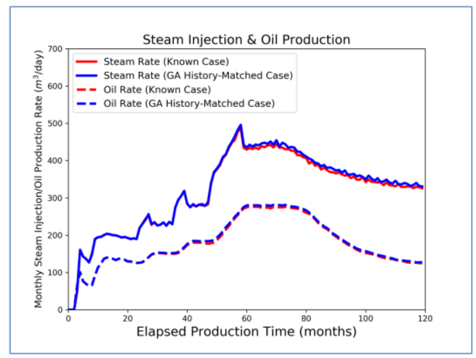
Case-34 (a)



Case-34 (c)



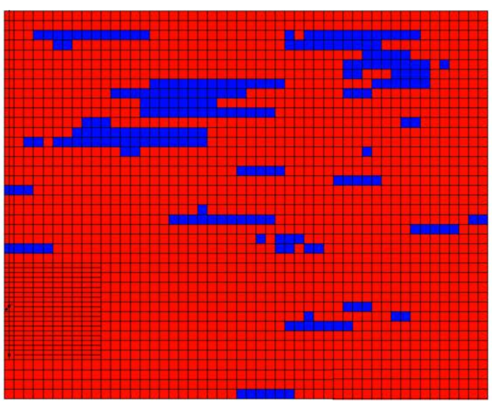
Case-34 (b)



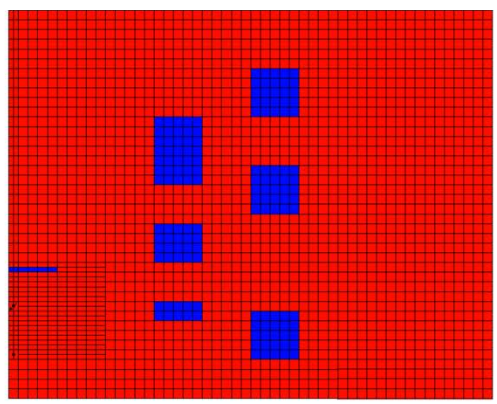
Case-34 (d)

Fig. 4.7 – Testing results for case 34 based on the revised workflow: (a) shale barrier configuration of the known case; (b) shale barrier configuration of the GA history-matched case; (c) oil production profiles from numerical simulation for the known case and from ANN predictions for both the known and GA history-matched cases; (d) oil production and steam injection profiles from numerical simulation for the known and GA history-matched cases.

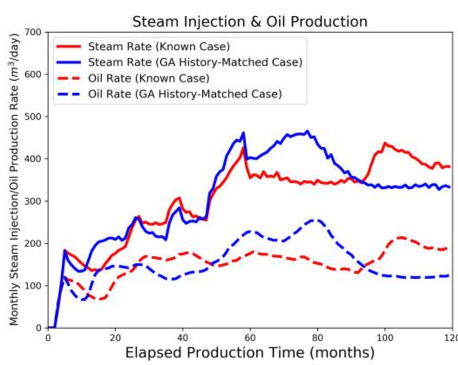
In order to examine the utility of the proposed parameterization approach for representing realistic reservoir shale distribution, two additional testing cases, which consist of stochastically-distributed shale barriers as shown in subplot (a) of **Fig. 4.8**, are constructed via sequential indicator simulation (Deutsch, 1998). Once again, the actual history is obtained from numerical simulation, and the GA-based workflow is used to infer the shale barrier configuration that corresponds to this target production history. The final GA history-matched shale barrier configuration is shown in subplot (b). The corresponding oil production and steam injection profiles for both the known and GA history-matched models are compared in subplot (c). Given that the proposed workflow can infer shale barrier configurations that consist of only pre-defined basic shale units, it is expected that the history-matched models would not represent all irregularities associated with individual shale barriers; however, it is encouraging to note that the aggregations of nearby shale barriers are detected. This observation also reflects the non-uniqueness of an ill-posed inverse problem: production data alone is often insufficient to completely resolve the high-frequency information associated with the model parameters. The results support the idea that, though minor deviations (mostly in the late times) are observed, the proposed approach is effective in capturing the influences of groups of shale barriers and generating reliable estimates of the recovery performance.



Validation Case-1 (a)



Validation Case-1 (b)



Validation Case-1 (c)

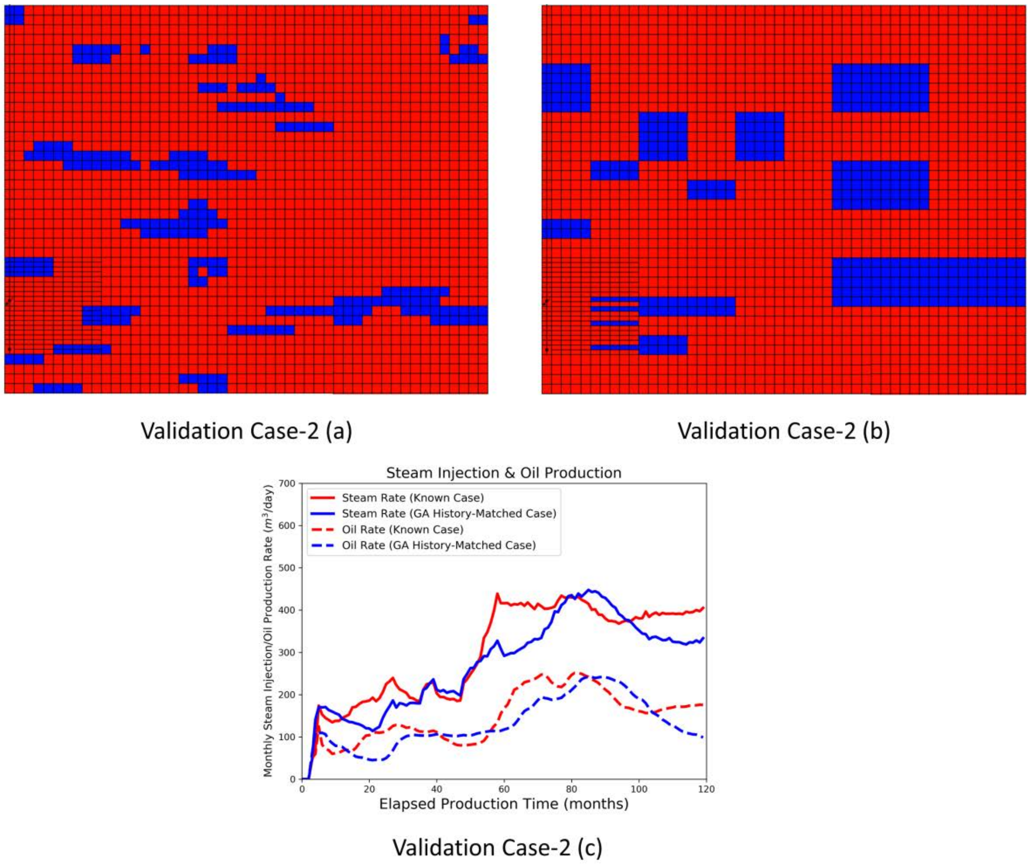


Fig. 4.8 – Testing cases with more realistic shale barrier configurations: (a) shale barrier configuration of the known case; (b) shale barrier configuration of the GA history-matched case; (c) oil production and steam injection profiles from numerical simulation for the known and GA history-matched cases.

Overall, the results demonstrate the capability and flexibility of the proposed AI-based workflow in inferring the presence and distribution of heterogeneous shale barriers from SAGD production profiles (time-series data). The method offers a tool to quickly identify a set of shale barrier configurations that are consistent with the production history. Once the ANN model is trained, the computational requirement for the forward ANN prediction would take only a few seconds, while the inverse GA history-match routine would take less than 3 minutes. The main computational cost lies in performing the numerical simulation for all the cases in the training dataset.

4.4 Summary

In this chapter, a two-part workflow is adopted to infer the presence and distribution of heterogeneous shale barriers from SAGD production profiles (time-series data). First, an artificial neural network (ANN) model is constructed to correlate SAGD production profiles with heterogeneous shale barrier configuration. Next, a genetic algorithm (GA) framework is applied for inferring shale barrier configuration by matching the target production time-series data. The performance of the workflow on 2D synthetic cases is shown to be both reliable and efficient in capturing most salient features of the production profiles and shale barrier configurations.

Analysis of the testing results reveals that screening out the shale configurations that are physically incompatible with the input production profiles could help improve the workflow performance. The following chapters would address two particular aspects. The first one is to improve the accuracy and robustness of the developed workflow by quantifying internal structures in the training dataset by incorporating clustering/classification techniques such as multidimensional scaling or support vector machine. The goal is to refine the selection of training cases and to establish multiple AI models corresponding to different characteristics in the model parameters. The second area is to extend the technique for 3D models and field applications, which would be more representative of the field conditions.

CHAPTER 5: VISUALIZING AND QUANTIFYING THE UNCERTAINTY IN THE IMPACTS OF UNCERTAIN SHALE CONFIGURATIONS[3]

In previous chapters, a training data set consisting of many different 2D realizations of reservoir heterogeneity is created and an AI-based modeling approach is successfully established to correlate SAGD production profiles with heterogeneous shale barrier configuration.

It is expected that incorporating additional cluster/classification techniques to analyze internal structures on the shale configurations could further improve the accuracy of the developed models. This could also help answer the key question when proceeding to 3D examples: "how to assemble a representative training data set?" The impact of shale barrier configurations on 3D cases will be more complex than that on 2D cases, and the computational cost for running 3D numerical simulation is higher. Therefore, implementing a full factorial design that involves all possible shale barrier configurations would be formidable. A representative training data set should be large enough to sufficiently span the feature (input and output parameter) space without exhaustively sampling realizations with similar production characteristics. However, this step is usually achieved via trial-and-error, and there is little discussion in the reviewed literature regarding: (1) How big should the data set be to span the feature parameter space? (2) How to optimize a training data set by maximizing spanning and quantifying similarities among different realizations?

The main objective of this chapter is to develop a workflow based on multidimensional scaling (MDS) and K-means clustering to assess and visualize similarities among a set of

[3] The content in Chapter 5 of this thesis is derived from the following paper:

Zheng, J., Leung, J.Y., Sawatzky, R.P., and Alvarez, J.M. (2018c). A cluster-based approach for visualizing and quantifying the uncertainty in the impacts of uncertain shale barrier configurations on SAGD production, Paper presented at *SPE Canada Heavy Oil Technical Conference*, Calgary, AB, Canada.

realizations. The results form the basis for selecting representative cases from individual clusters, as well as guiding the addition of new cases to under-sampled clusters.

5.1 Methodology

The proposed methodology is illustrated in **Fig. 5.1**. It is designed to facilitate the selection of an optimal subset of reservoir realizations that efficiently spans the feature space. An initial set of N synthetic cases corresponding to a range of shale barrier configurations is subjected to flow simulation. A distance function is defined, and an $N\times N$ dissimilarity matrix is constructed. MDS is applied to transform the distribution of all N cases into an n-dimensional Euclidean space based on the dissimilarity matrix. Cases that are located close to others in this transformed space are likely to be redundant and, hence, only a subset of these cases will be retained. MDS is applied again and several clusters with distinct features pertinent to the shale barrier configuration are discovered via K-means clustering. Additional cases of specific shale barrier features are generated for individual clusters with limited members. The new cases are subjected to flow simulation to verify their membership in the assigned clusters. Good consistency in the results has been observed.

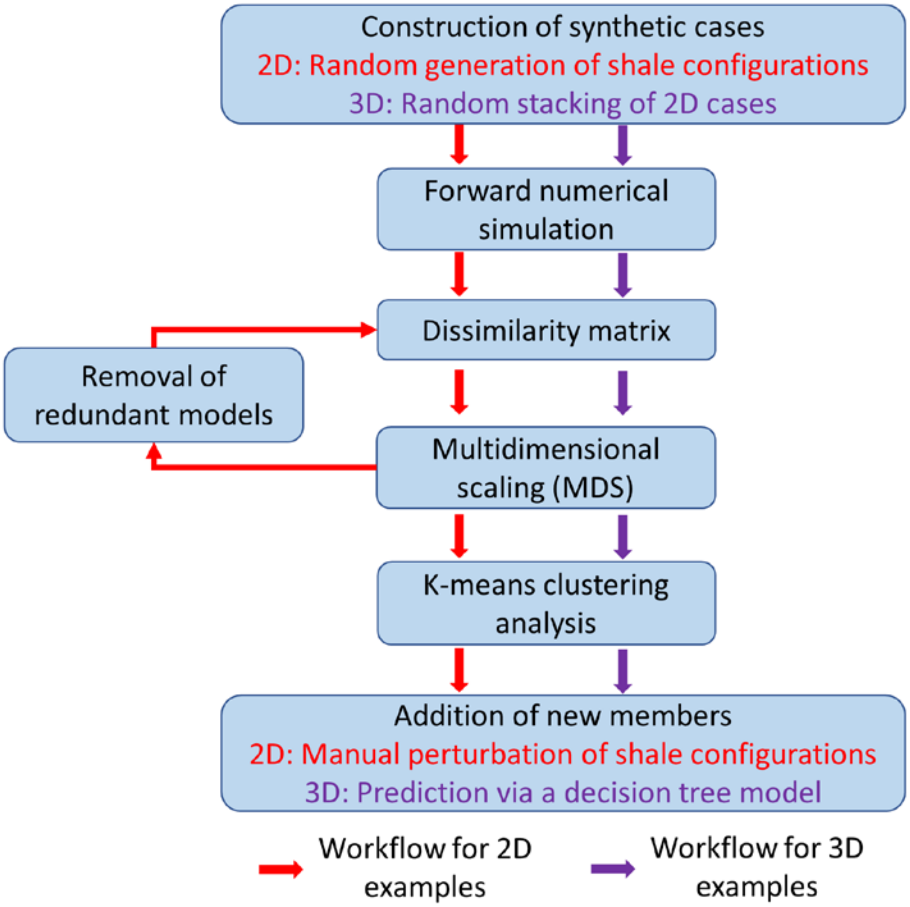


Fig. 5.1 – Proposed workflow for visualization and analysis of heterogeneous shale configurations.

5.1.1 Formulation of Distance Function

The use of a metric or distance function (*d*) to measure the similarity between a pair of geological models was first introduced by Arpat (2005) and Suzuki and Caers (2008). In mathematics, a metric is used to define the distance between two elements of a set. In this context, *d* measures the (dis)similarity of two models in terms of their spatial properties and/or transfer function response (Scheidt & Caers, 2009). For example, the Hausdorff distance measures the geometrical difference (Dubussion & Jain 1994), and many studies have assumed that the Hasudorff distance is correlated directly with the difference in flow response (Suzuki and

Caers, 2008; Lee et al., 2017). However, the formulation of Hausdorff distance does not explicitly take into account the proximity of a given shale barrier to the injector-producer well pair, which has a dramatic impact on the SAGD performance. Alternative formulations may include the Manhattan distance (Honarkhah & Caers, 2010), flow-based distance (Scheidt & Caers, 2009), or connectivity-based distance (Park & Caers, 2007). In general, the choice of the distance function would depend on the flow or transport process, as well as the types of reservoir heterogeneity considered (Suzuki & Caers, 2008). In theory, the distance function can be defined in any fashion, as long as it reflects the difference in production response corresponding to the given pair of models (Scheidt & Caers, 2009).

In this study, a flow-based distance function, *d*, is formulated to fully capture the differences in SAGD production profiles. Inspired by the cost function used in linear regression algorithms, which is half of the mean squared difference between the predicted and actual values, *d* is defined to quantify the dissimilarity in the time series of oil production and steam injection:

$$d = \frac{1}{2m}\left(\sum_{i=1}^{m} \left(q_{oi} - q_{oi}^{*} \right)^{2} + \sum_{i=1}^{m} \left(q_{si} - q_{si}^{*} \right)^{2} \right) \quad (5.1)$$

where *m* is the number of monthly data records of oil production rate and steam injection rate; *i* denotes a single record; q_{oi} and q_{oi}^{*} refer to the i^{th} oil production rates corresponding to the individual members for a given pair of shale configurations; q_{si} , q_{si}^{*} are defined similarly for the steam injection rates. To enhance the visualization of *d* after MDS, each of the squared differences in **Eq. 5.1** is divided by an arbitrary constant for normalization purposes; a value of 100, which is approximately the mean of q_s and q_o, is chosen. Values of the distance function between every pair of synthetic cases are calculated, and an $N \times N$ dissimilarity matrix, $\boldsymbol{D}$, is assembled.

5.1.2 Multidimensional Scaling (MDS)

MDS is a set of mathematical operations, where the dissimilarity matrix is transformed into a configuration of points in an n-dimensional Euclidean space (Cox & Cox, 2001). It computes a set of vectors in the n-dimensional space such that the matrix of Euclidean distances among them reproduces as closely as possible the features of the dissimilarity matrix. A high correlation between the Euclidean distance and the dissimilarity distance is expected. Each point in the transformed (new feature) space, $\boldsymbol{R}$, represents one of the N cases, and the spatial distribution of all the points is such that the Euclidean distances among them would reflect the distances computed in **section 5.1.1** (Scheidt & Caers, 2009).

MDS is useful for representing high-dimensional data in a low-dimensional space that facilitates visual inspection and exploration (Borg & Groenen, 2005). This process may reveal internal structures that are hidden in the high-dimensional space (Honarkhah & Caers, 2010). In this case, MDS offers a visual representation of the distances (or proximities) among a set of objects, which are cases of shale barrier configurations.

The algorithm of classical MDS can be illustrated as follows (Wickelmaier, 2003):

(1) Compute a matrix of squared dissimilarities $\boldsymbol{P}=[\boldsymbol{D}^2]$;

(2) Perform double centering:

1) Calculate the matrix $\boldsymbol{J}$=$\boldsymbol{I}$-(1/N)$\boldsymbol{J}_n$, where $\boldsymbol{I}$ is the identity matrix, N is the number of objects, and $\boldsymbol{J}_n$ is an N-dimensional all-ones matrix;
2) Apply $\boldsymbol{J}$ to calculate the double centered matrix $\boldsymbol{B}$=(1/2)$\boldsymbol{JPJ}$;

(3) Extract the n largest eigenvalues and eigenvectors, where n is the dimension in the new feature space. Assemble a matrix $\boldsymbol{E}_n$ of n eigenvectors and a diagonal matrix $\boldsymbol{A}_n$ of n eigenvalues;

(4) Transform the N objects to an n-dimensional Euclidean space with the matrix $X=\boldsymbol{E}_n\sqrt{\mathbf{A}_n}$.

The production histories of all cases will be used to compute the matrix $\boldsymbol{D}$. In most cases, a two dimensional feature space (n = 2) is sufficient to show the distribution of objects.

5.1.3 Removal of Redundant Cases

Given that the initial N cases are generated randomly, it is expected that some of these cases would exhibit similar production profiles, even though the respective shale barrier configurations are different. Visualizing and inspecting these cases in the new feature space $\boldsymbol{R}$ facilitates the identification of some closely-distributed cases; removing these redundant cases could reduce the size of the data set while preserving the variability of the retained ensemble. The following procedure is adopted in this study:

(1) Populate a dummy set (initially empty) by randomly selecting one case from the original N cases and adding it sequentially to the dummy set;

(2) Pick a second case from the original data set and compute the Euclidean distances in $\boldsymbol{R}$ between the second case and the first case in the dummy data set (from step #1). If the

distance between the first case in the dummy set and the second case exceeds a certain user-defined threshold, the second case is added to the dummy set; otherwise, it is not;

(3) Select a third case from the original data set and repeat step #2 for all cases in the dummy set. Iterate over all N cases.

(4) The cases that were included in the dummy set from a reduced data set consisting of N' records; a new dissimilarity matrix $\boldsymbol{D'}$ should be computed with this reduced set (**section 5.2.1**). Applying MDS to $\boldsymbol{D'}$ would yield a new feature space, which is denoted as $\boldsymbol{R'}$.

Given that the results of MDS depend on the relative dissimilarity among elements of a given set, it is expected that the spatial arrangement of the elements corresponding to the new dissimilarity matrix $\boldsymbol{D'}$ in $\boldsymbol{R'}$ will be different from the spatial arrangement of the elements obtained using the original $\boldsymbol{D}$ in $\boldsymbol{R}$.

5.1.4 K-means Clustering

K-means clustering is then applied to $\boldsymbol{D'}$ in the new feature space $\boldsymbol{R'}$. It is one of the most widely applied clustering techniques for discovering internal structures or groupings among data records. The idea is to identify N_c clusters that would minimize the mean squared Euclidean distance from each data point to its nearest centroid (Kanungo et al., 2002). The classical K-means algorithm is described as follows:

(1) Define the number of total clusters N_c;

(2) Initialize randomly N_c cluster centroids in the feature space $\boldsymbol{R'}$;

(3) Assign each data point in $\boldsymbol{R'}$ to one of the N_c cluster centroids with the shortest distance;

(4) Calculate a new centroid for each cluster as the arithmetic mean of all the points that belong to that particular cluster;

(5) Repeat steps (3) and (4) iteratively until no further changes are made to the cluster centroids.

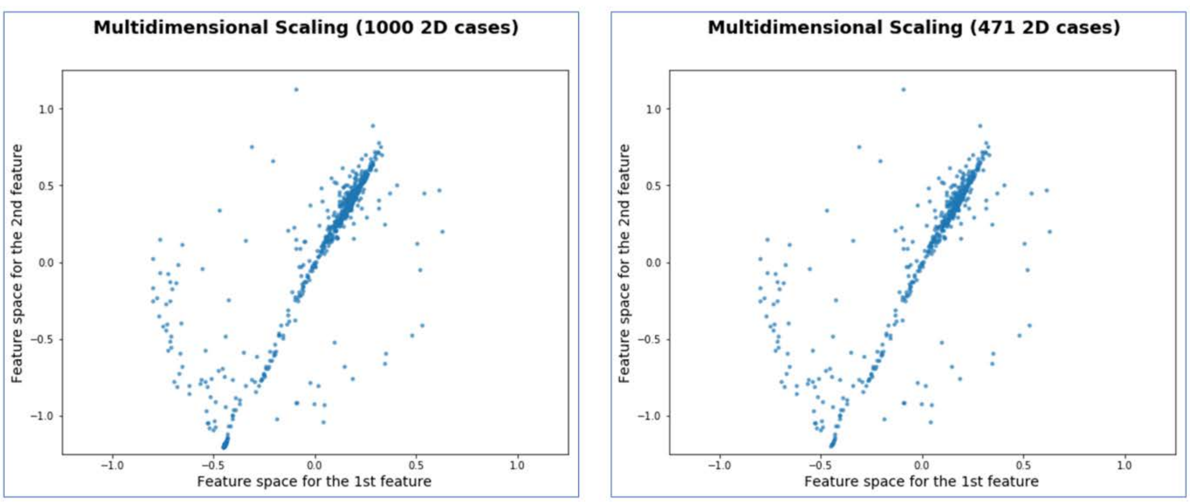
Given that this procedure is sensitive to the initial guess for the cluster of centroids, steps (2) to (5) should be repeated with different initializations of the cluster centroids, and at the end the scenario that gives the maximum separation between clusters should be selected. The overall performance of the clustering analysis is usually evaluated by the average silhouette value. Silhouette measures the similarity between a sample and its own cluster (i.e., cohesiveness) in comparison to other clusters (i.e., separation). The silhouette has a range of [-1, 1]. A value close to 1 means that an object within a given cluster is more similar to other objects in that same cluster than to the rest of the set, reflecting a desirable outcome where a distinct separation exists between the given cluster and the others. A value of 0 suggests that the object can be assigned to more than one cluster. Finally, a silhouette value close to -1 implies that the object is misclassified. The average silhouette value over the entire dataset provides an overall assessment of the cluster analysis, and it is often used as a measure for optimizing the number of clusters (Rousseeuw, 1987).

5.2 2D Examples

5.2.1 Synthetic Model Setup and MDS Analysis

A total of N_{2D} = 1000 shale configurations are constructed randomly and parameterized according to **section 3.3.3**. All 1000 2D cases are subjected to numerical simulation, and their production histories are used to compute the matrix $\boldsymbol{D}$. The distribution of all cases in the 2D

Euclidean space ***R*** after applying MDS is shown in **Fig. 5.2(a).** Each point in the figure corresponds to one of the 1000 two-dimensional cases. As discussed in section **5.1.2**, the goal of MDS is to reproduce the dissimilarity matrix in a lower dimensional space; it is much easier to visualize the distribution of all cases in a two-dimensional map than a higher dimensional map. The n^{th} feature in the transformed Euclidean space ***R*** is calculated from the n^{th} largest eigenvector and eigenvalue of matrix ***D***. A linear distribution in space ***R*** might indicate that the original dataset is linearly separable, because the Euclidean distance of any two points in space ***R*** is highly correlated to the pairwise dissimilarity distance (difference of production profiles). Inspection of **Fig. 5.2(a)** reveals that many points are overlapping with one another; analysis of their corresponding shale configurations and production histories would further confirm that many of these cases are indeed redundant.



(a) Original dataset with 1000 2D cases (b) Screened dataset with 471 2D cases

Fig. 5.2 – MDS results in the features space ($\boldsymbol{R_{2D}}$) calculated based on the original N_{2D} = 1000 2D cases.

To remove these redundant cases, procedures as described in **section 5.1.3** are followed. A threshold of 0.005 is arbitrarily set, which yields a total of 471 cases in the reduced data set. The distribution of all 471 cases in the original MDS feature space $\boldsymbol{R_{2D}}$ is shown in **Fig. 5.2(b)**. It can be observed that, although more than half of the original cases have been removed, the remaining 471 cases are still effectively spanning the feature space, as compared with **Fig. 5.2(a)**. This observation seems to support the hypothesis that many cases in the original data set are exhibiting similar production characteristics. The distribution of all N_{2D}'=471 cases in the new MDS feature space $\boldsymbol{R_{2D}'}$ is also shown in **Fig.5.3**.

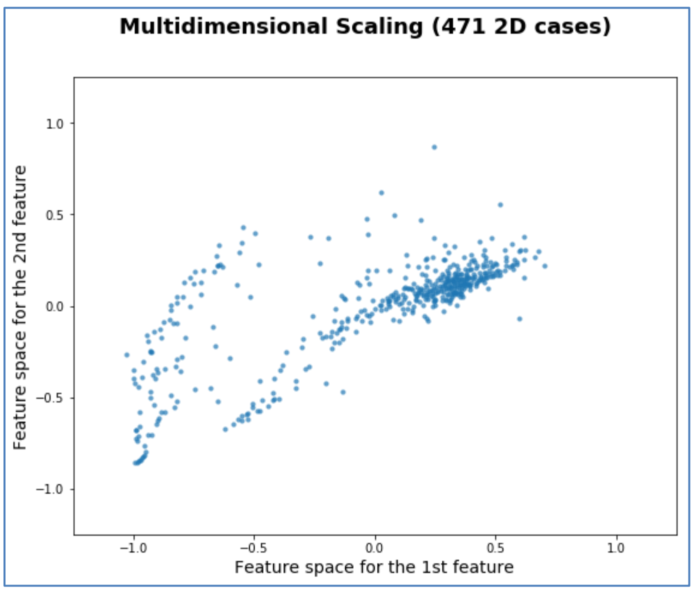


Fig. 5.3 – MDS results in the features space ($\boldsymbol{R_{2D}'}$) calculated based on the original N_{2D}' = 471 2D cases.

5.2.2 Cluster Analysis

K-means clustering is then applied to analyze the dissimilarity and internal structures of these 471 cases. To identify the optimal value for N_c, clustering is repeated for different numbers of clusters (ranging between 4 and 10). The corresponding average silhouette values of the N_{2D}' cases (number of 2D cases in $\boldsymbol{D'}$) are listed in **Table 5.1**. It is observed that the average silhouette value stays approximately constant at 0.49 for N_c = 4, 5, and 6, but it starts to decline if N_c keeps increasing, indicating that further dividing the entire dataset into more clusters is not ideal: more similarities in terms of production profiles among the clusters would be observed, while additional clusters are unlikely to capture sufficient objects with unique features. As a result, N_c = 6 is selected as the optimal number of clusters for the 2D cases. This choice offers a reasonable degree of coherence within individual clusters, while maintaining sufficient distinction among different clusters. The K-means results are shown in **Fig. 5.4**.

Table 5.1 – Average silhouette value for different number of clusters (471 2D cases).

Number of Clusters	4	5	6	7	8	9	10
Average Silhouette Value	0.498	0.49	0.481	0.433	0.448	0.439	0.433

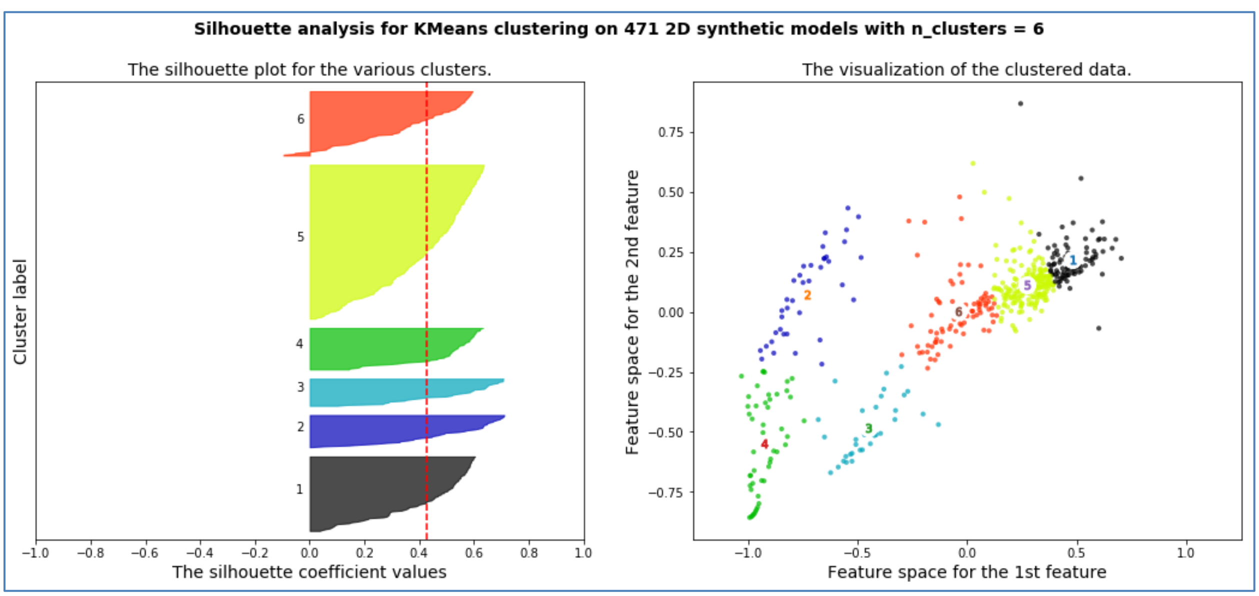


Fig. 5.4 – K-means clustering results corresponding to 6 2D clusters.

The oil rate profiles for all clusters corresponding to N_c = 6 are shown in **Fig. 5.5**. It is observed that most of production curves within same cluster have similar shapes. One particular outlier is the light blue curve with a spike in cluster 3. Further examination of **Fig. 5.4** shows that this case is located at the boundaries in between clusters 2, 3 and 4, it might be more reasonable to classify this case into either cluster 2 or 4. To investigate the characteristics of shale barrier configurations corresponding to these clusters, cases that are located closest to the cluster centroids are selected and the salient characteristics of their shale barrier configurations are examined in **Fig. 5.6**. The key features of each cluster are summarized in **Table 5.2**.

In summary, shale barriers that are located between the injector and producer would have a more pronounced impact on steam chamber development and oil production (clusters 2 and 4); these cases should be considered as the extreme scenarios, especially when considering 2D cases, as it would be highly unlikely for an operator to keep injecting steam for 80 months with nearly zero production. Laterally-extensive shale barriers that are located right above the injector would impact oil production to a certain extent (clusters 1 and 5). Finally, complex shale configurations in the near-well region are most detrimental to SAGD production, as they would dramatically obstruct the steam chamber growth (clusters 3 and 6).

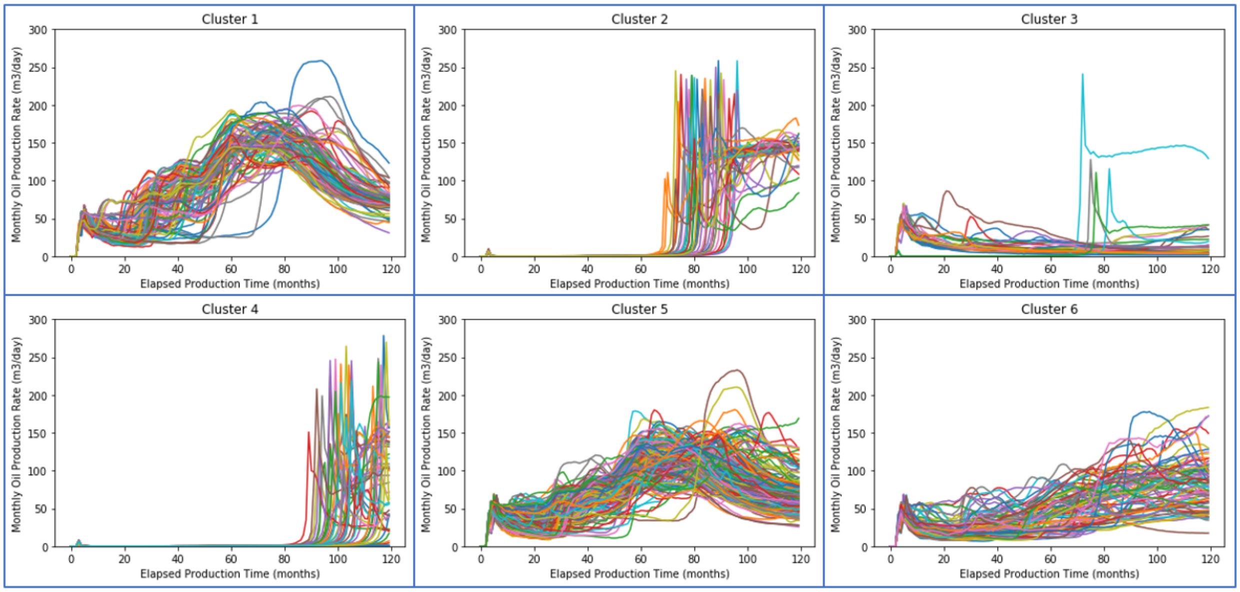
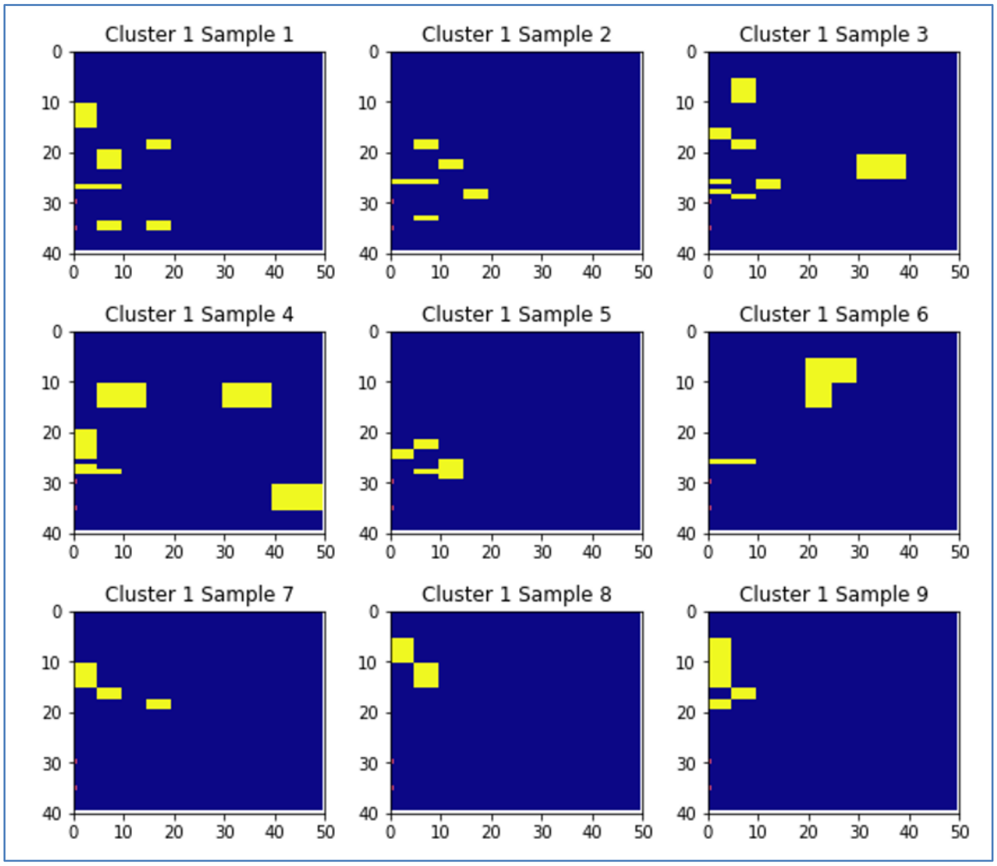


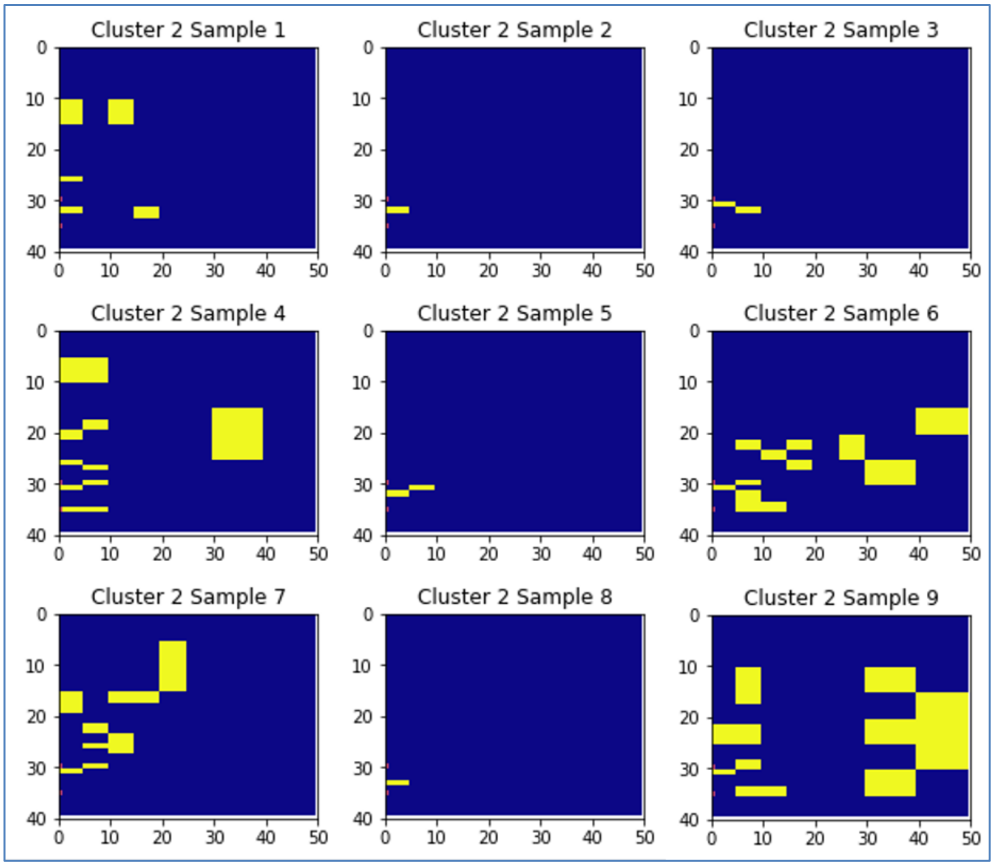
Fig. 5.5 – Production profiles for each of the 6 clusters of 471 2D cases.

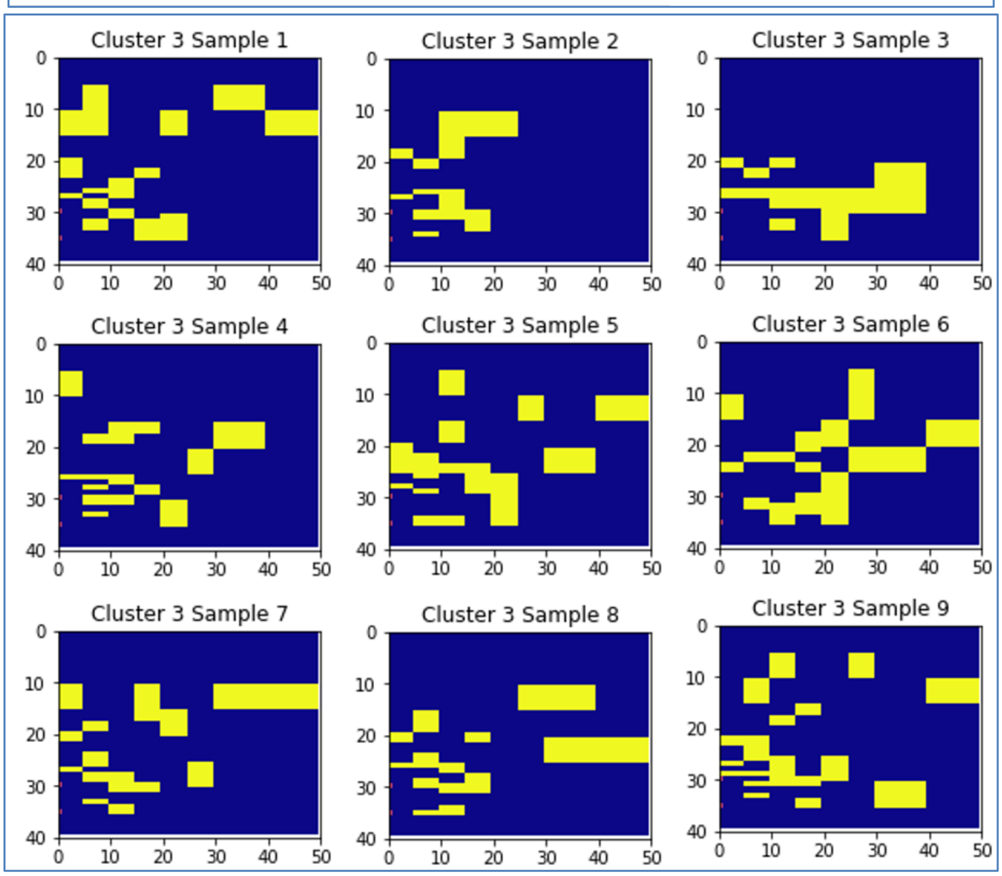
Table 5.2 – Production characteristics and shale configuration features of all 6 clusters for the 2D examples.

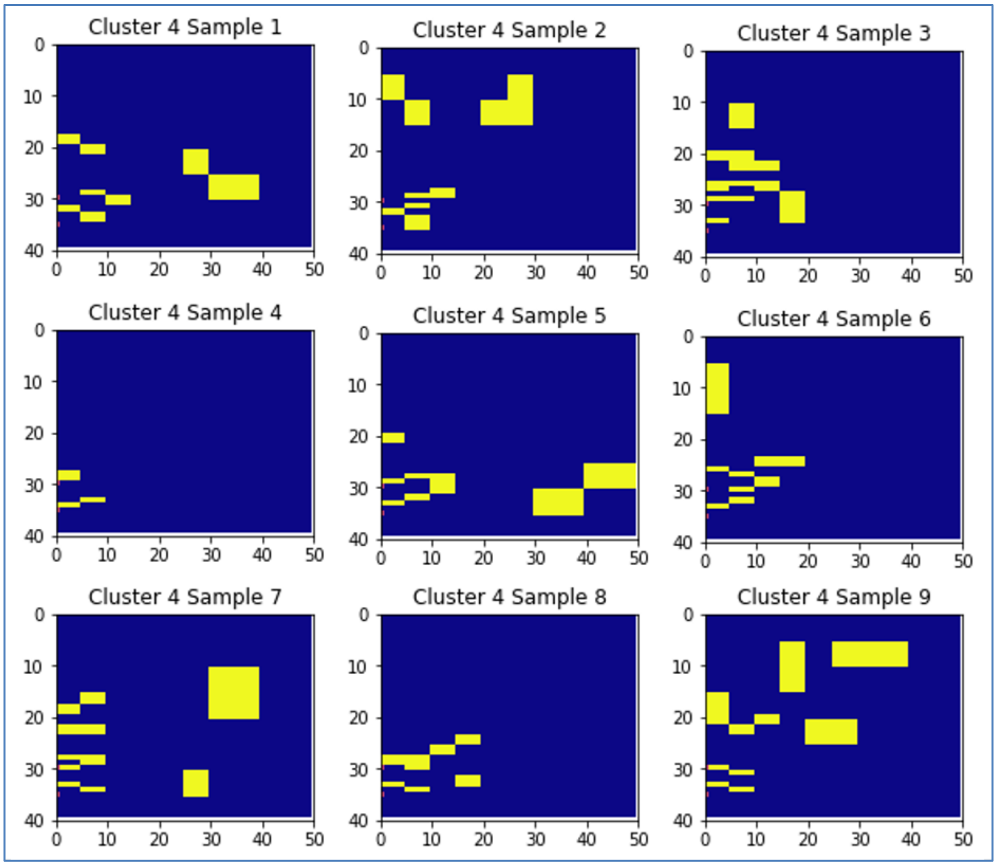
Cluster	Production Characteristics	Shale Configuration Features
1	q_o gradually increases to a maximum value of approximately 200 m^3/day and starts to decline after 60 to 80 months.	At least one shale barrier with a lateral extent of approximately 10 m is located right above the injector. However, very few other shale barriers are located in the near-well region to block the steam chamber growth. Therefore, the overall impact of all shale barriers is moderate.
2	There is no oil production ($q_o \sim 0$ m^3/day) during the initial 80 months. Subsequently, the steam chamber begins to grow; hence, q_o rapidly increases to a maximum value of approximately 250 m^3/day and then stabilizes to roughly 150 m^3/day afterwards.	The shale barriers have severely hindered the steam chamber development. There is at least one shale barrier with a lateral extent of 5 to 10 m located between the injector and producer. Due to their proximity to the injector, there is no oil production until the steam chamber has forced its way beyond the near-well region.

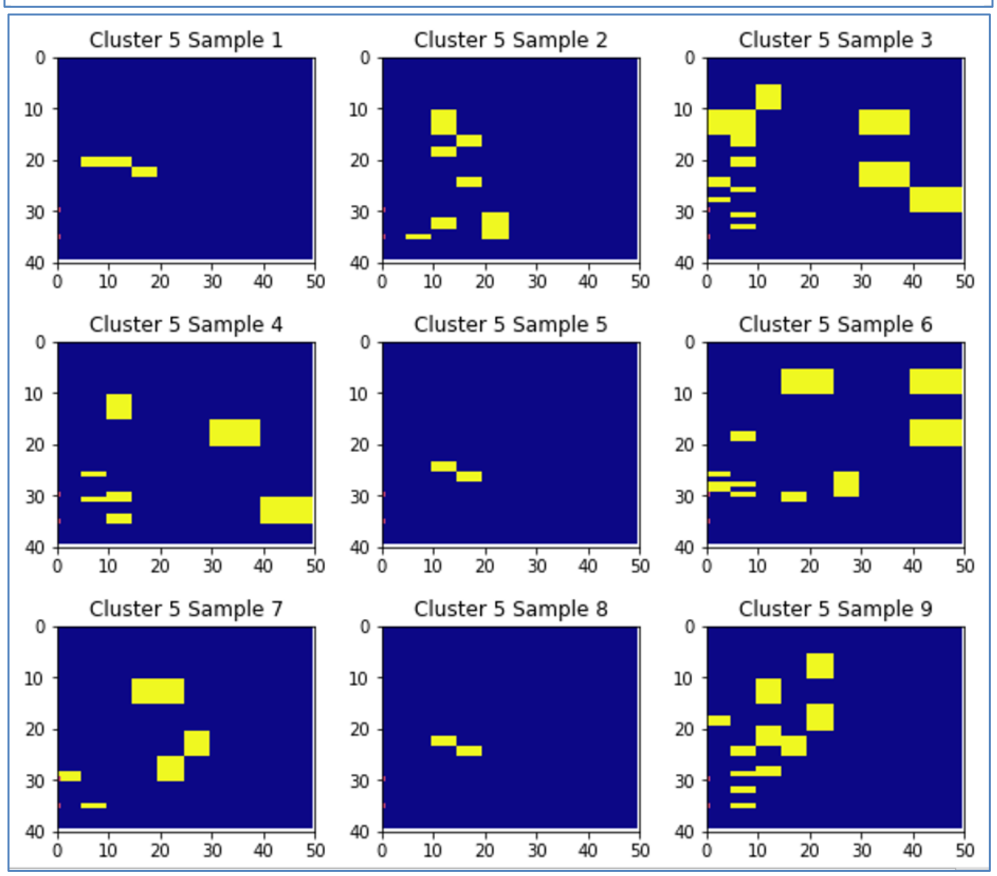
3	q_o is low ($\leq$ 50 m^3/day). There is a short initial increase to a maximum value of approximately 80 m^3/day and then it starts to decline after 5 months.	The shale barrier configurations are complex, especially in the near-well region. They are mostly extensive along both the lateral and vertical directions, and they are mainly located around the well-pair. Although their impacts are not as severe as those in cluster 2, they could potentially isolate the near-well region from the rest of the reservoir and, hence, prevent an effective steam chamber development.
4	Production characteristics are similar to those in cluster 2; the initial production is further delayed to 80 to 120 months.	The shale barriers have a more severe impact on the steam chamber development than those in cluster 2, since they are located closer to the producer in most of the cases. The cluster analysis presented in **Fig. 5.4** reveals that clusters 2 and 4 are linearly separable from the other clusters (implying that they share some unique characteristics that are distinct from the remaining clusters).
5	Production characteristics are similar to those in cluster 1; the maximum rate is lower (approximately 150 m^3/day).	The overall impact of all shale barriers is moderate. In comparison to cluster 1, the shale barriers are slightly longer (> 10 m) and thicker (> 5 m) in the near-well region. There are several instances where a 10-m shale barrier is located right above the injector, but it does not impede on the flow path.
6	Production characteristics during the early stage (initial 20 months) are similar to those in cluster 3; however, q_o increases slowly after the initial decline period.	The shale configurations are complex, especially in the near-well region. There are some laterally-extensive shale barriers that are located right above the injector or close to the well pair. However, unlike cluster 3, the steam chamber development is not fully obstructed. Therefore, q_o increases slowly, as more bitumen is contacted by the steam.


Cluster 1 Sample 1
Cluster 1 Sample 2
Cluster 1 Sample 3
Cluster 1 Sample 4
Cluster 1 Sample 5
Cluster 1 Sample 6
Cluster 1 Sample 7
Cluster 1 Sample 8
Cluster 1 Sample 9
0
10
20
30
40
50


Cluster 2 Sample 1
Cluster 2 Sample 2
Cluster 2 Sample 3
Cluster 2 Sample 4
Cluster 2 Sample 5
Cluster 2 Sample 6
Cluster 2 Sample 7
Cluster 2 Sample 8
Cluster 2 Sample 9


Cluster 3 Sample 1
Cluster 3 Sample 2
Cluster 3 Sample 3
Cluster 3 Sample 4
Cluster 3 Sample 5
Cluster 3 Sample 6
Cluster 3 Sample 7
Cluster 3 Sample 8
Cluster 3 Sample 9


Cluster 4 Sample 1
Cluster 4 Sample 2
Cluster 4 Sample 3
Cluster 4 Sample 4
Cluster 4 Sample 5
Cluster 4 Sample 6
Cluster 4 Sample 7
Cluster 4 Sample 8
Cluster 4 Sample 9


Cluster 5 Sample 1
Cluster 5 Sample 2
Cluster 5 Sample 3
Cluster 5 Sample 4
Cluster 5 Sample 5
Cluster 5 Sample 6
Cluster 5 Sample 7
Cluster 5 Sample 8
Cluster 5 Sample 9

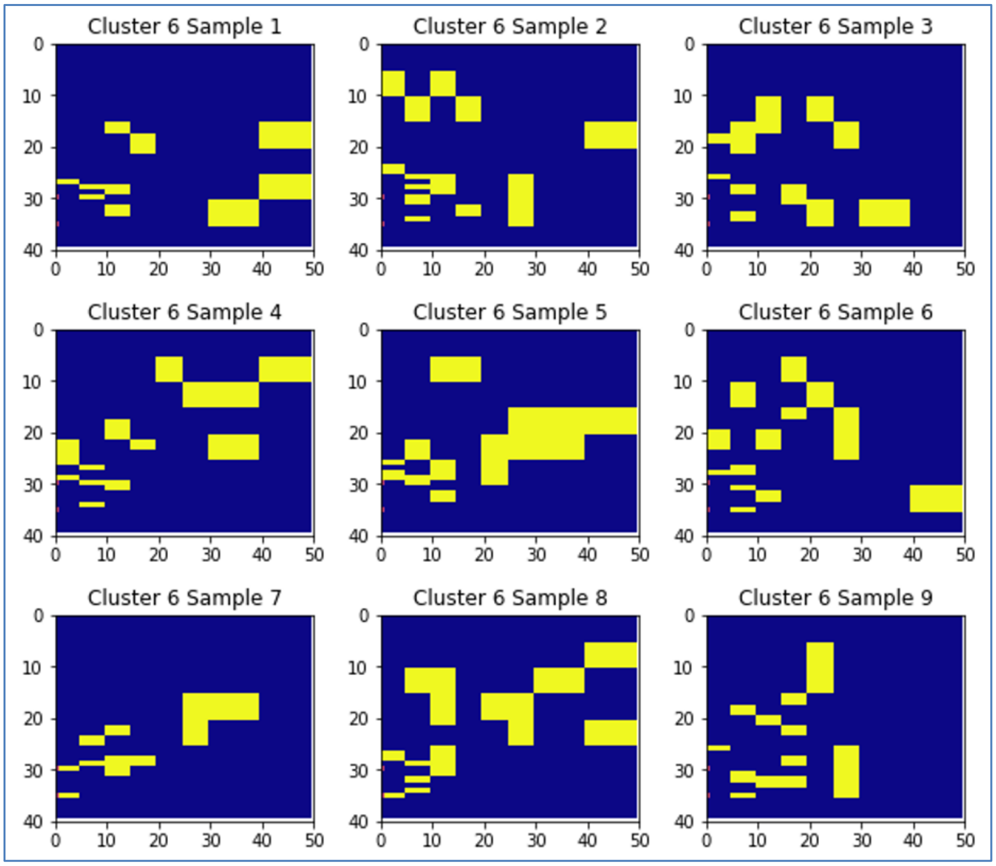


Fig. 5.6 – Representative shale configurations for all 6 clusters.

5.2.3 Adding New Cases with Cluster Features

For those clusters with limited cases, it would be desirable to populate them with additional cases sharing similar cluster features, which could improve spanning of model parameters in those under-sampled clusters. **Table 5.3** summarizes the number of cases belonging to each cluster in **section 5.2.2**, and clusters 2, 3 and 4 have the least members in comparison to others. The following procedure is implemented to generate a set of additional cases:

(1) Select a case from the target cluster;

(2) Construct a new case based on the case selected from the target cluster: recognizing that shale barriers in the near-well region (i.e., zones 1 and 2) are highly effective in obstructing the steam chamber development, these shale barriers should remain; however, shale barriers that are located further away (i.e., zones 3 and 4) can be randomly removed or altered, generating a case that is different from the original case selected;

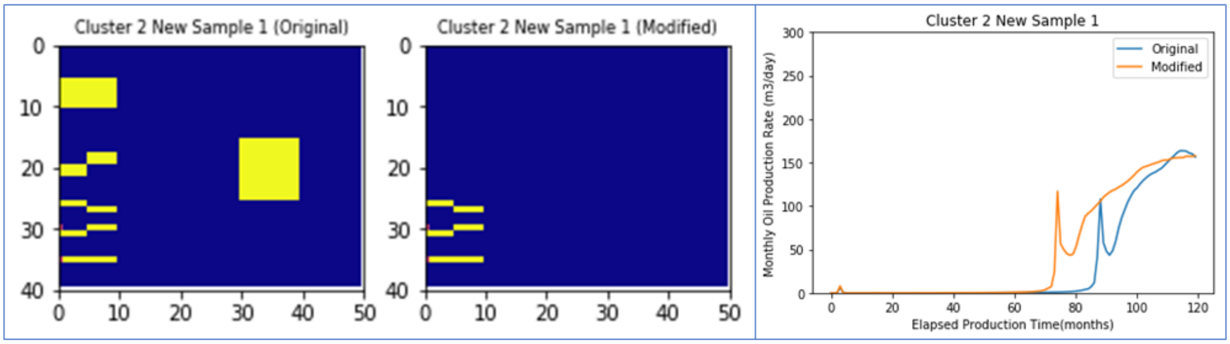
(3) Subject the new case to flow simulation to verify its membership in the target cluster;

(4) Repeat step (1) - (3) to generate as many new cases for each target cluster as desired. The results of applying this procedure are presented in **Fig. 5.7**.

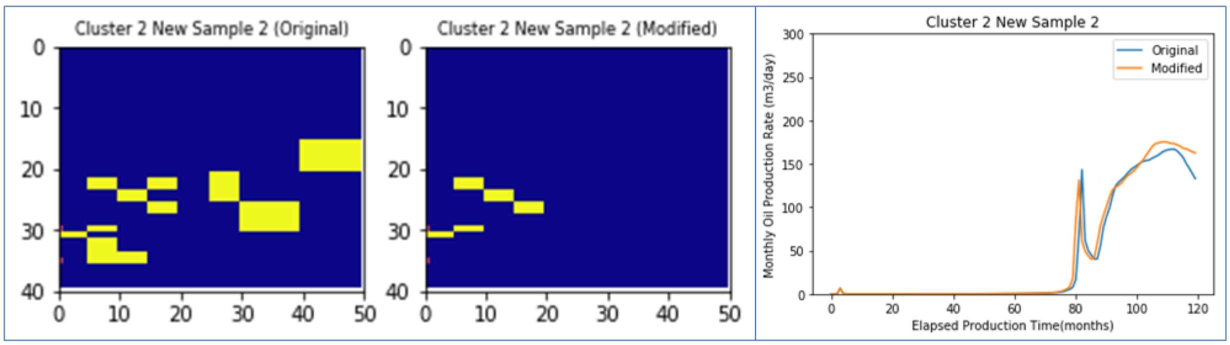
Table 5.3 – Number of cases in different clusters (2D examples).

Cluster	1	2	3	4	5	6	Total
Number of models	89	39	33	50	183	77	471

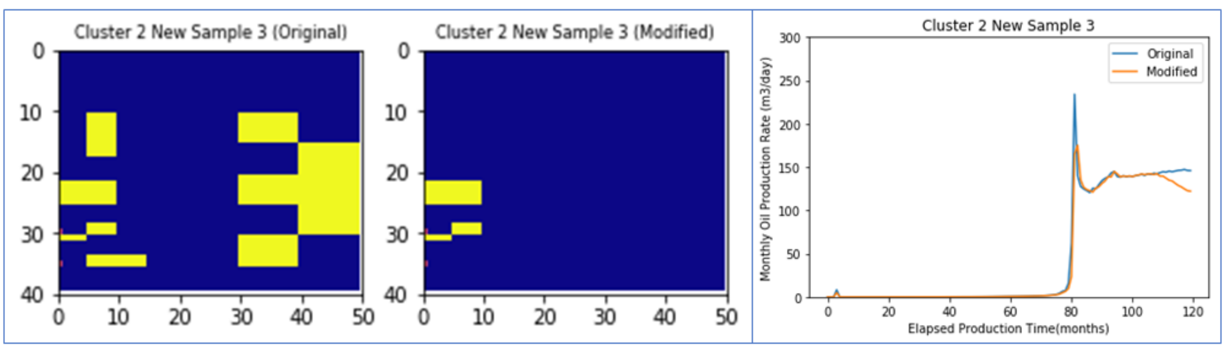
The flow simulation results confirm that all the modified cases exhibit similar oil production profiles as those from the original target cluster. This result also reaffirms the observed features for the different clusters, as described in **section 5.2.2**. The comparison results in **Fig. 5.7** support the hypothesis that SAGD production is most sensitive to the shale barriers in the near-well region.



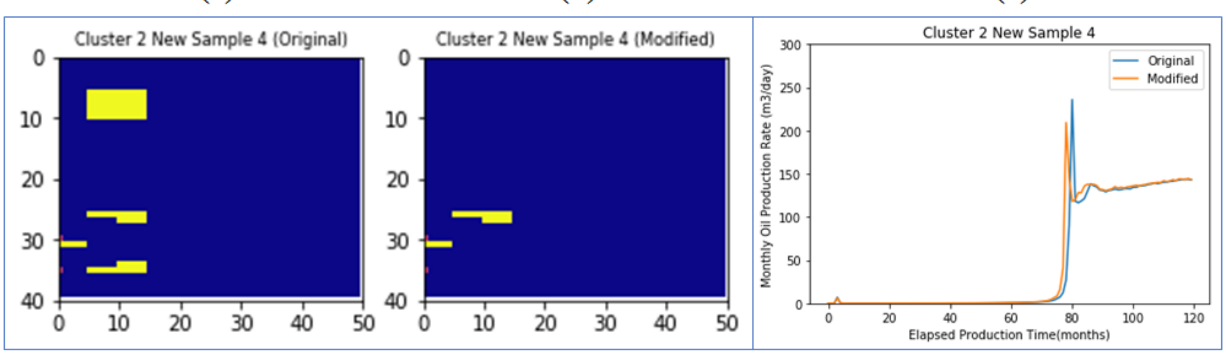
(a) (b) (c)



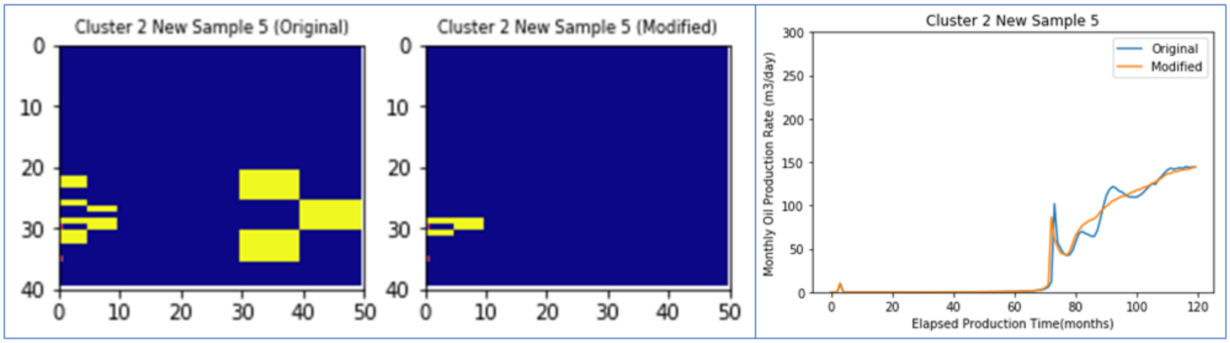
(a) (b) (c)



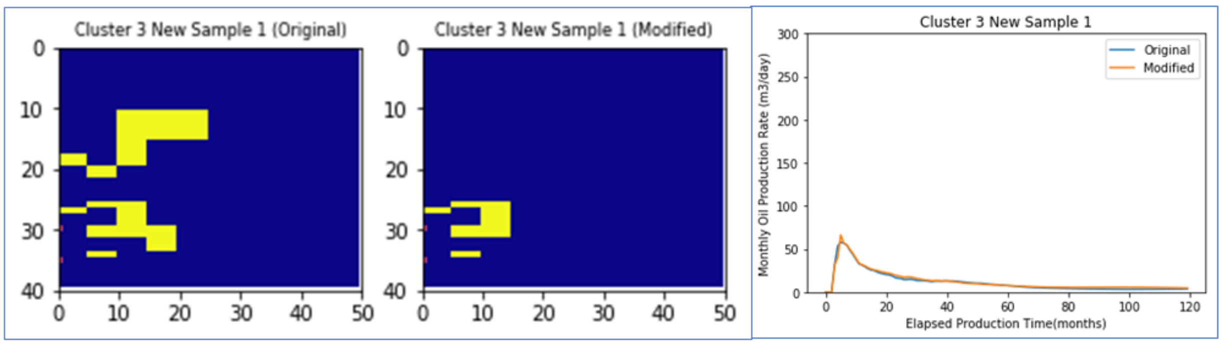
(a) (b) (c)



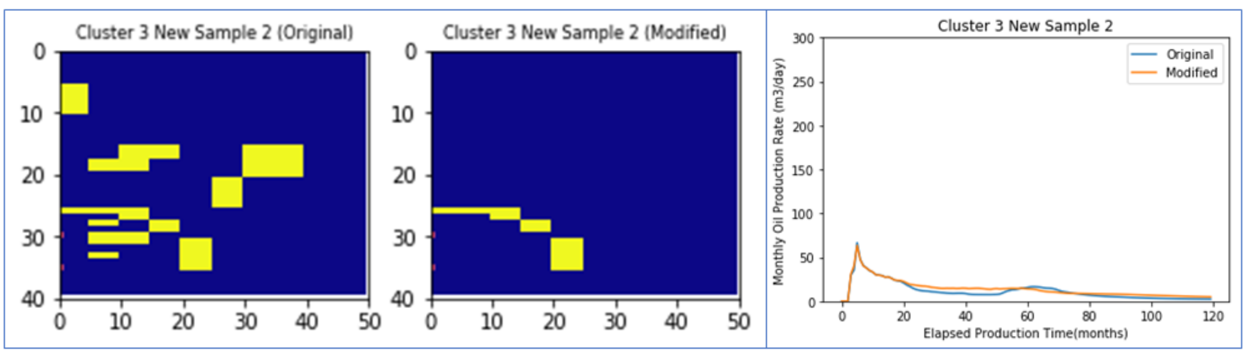
(a) (b) (c)



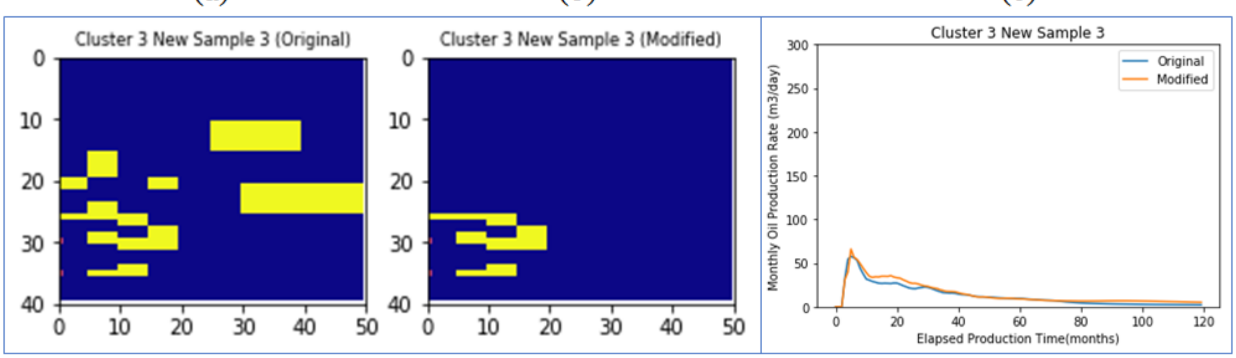
(a) (b) (c)



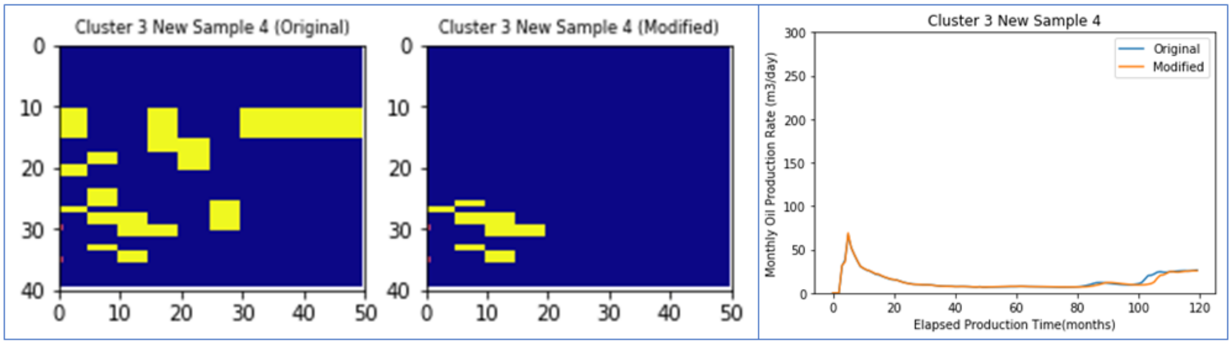
(a) (b) (c)



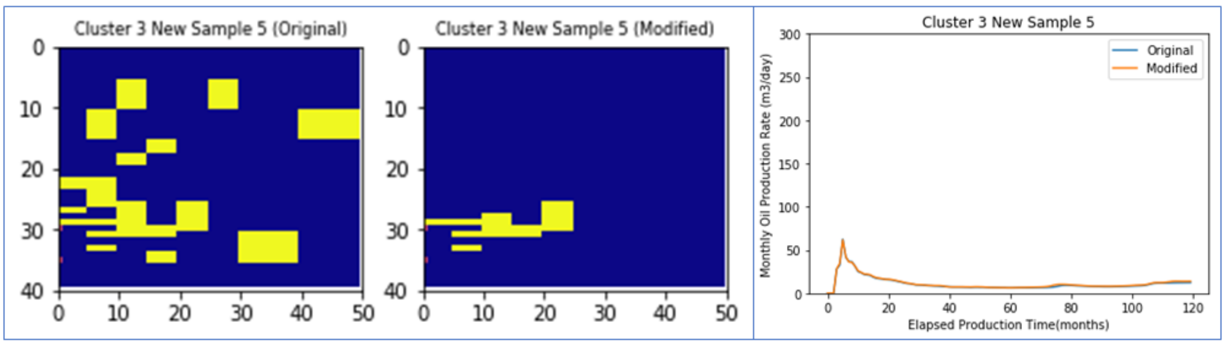
(a) (b) (c)



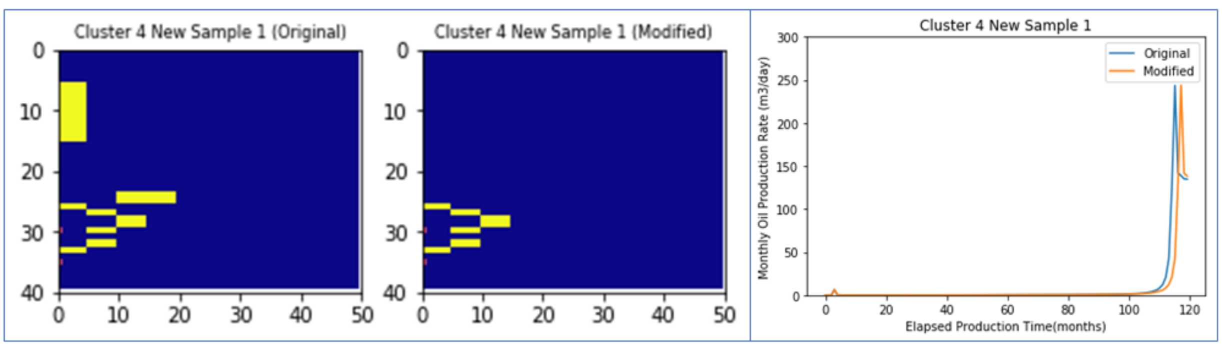
(a) (b) (c)


Cluster 3 New Sample 4 (Original)
Cluster 3 New Sample 4 (Modified)
Cluster 3 New Sample 4
Original
Modified
Monthly Oil Production Rate (m3/day)
Elapsed Production Time(months)

(a) (b) (c)


Cluster 3 New Sample 5 (Original)
Cluster 3 New Sample 5 (Modified)
Cluster 3 New Sample 5
Original
Modified
Monthly Oil Production Rate (m3/day)
Elapsed Production Time(months)

(a) (b) (c)


Cluster 4 New Sample 1 (Original)
Cluster 4 New Sample 1 (Modified)
Cluster 4 New Sample 1
Original
Modified
Monthly Oil Production Rate (m3/day)
Elapsed Production Time(months)

(a) (b) (c)

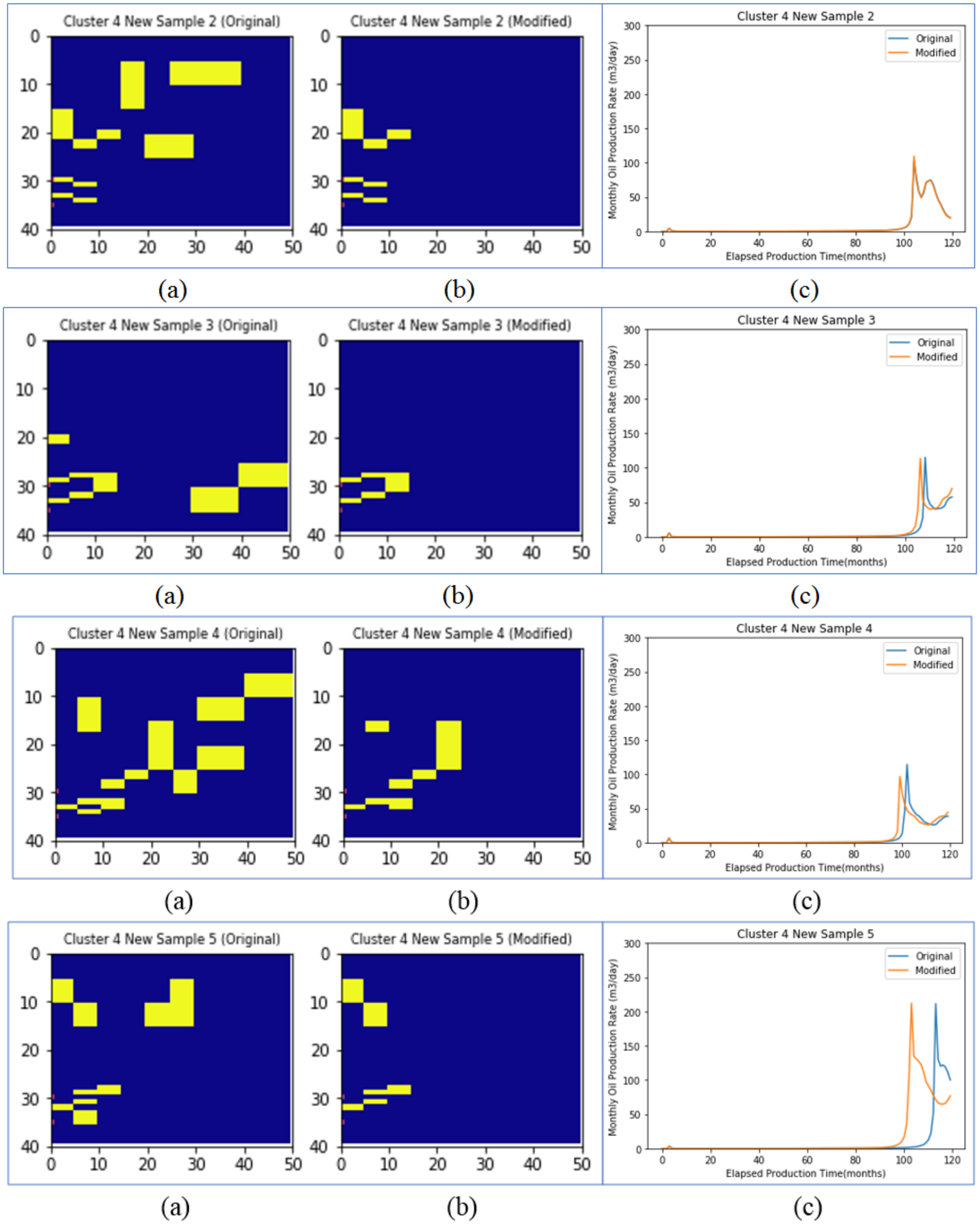


Fig. 5.7 – Additional 2D cases for Clusters 2, 3 and 4: (a) original shale barrier configuration of an existing case in the cluster; (b) a new shale barrier configuration modified from (a) by preserving certain key features; (c) comparison of oil production profiles obtained from numerical simulation.

5.3 3D Examples

There are several practical challenges associated with assembling a representative data set consisting of 3D cases. First, the simulation run time for a 3D simulation model is substantially increased in comparison to its 2D counterpart (e.g., ~ 60 minutes for a 3D model). Second, the model parameter space in 3D is large. As a result, implementing a full factorial design that involves all possible shale barrier configurations would be formidable. Therefore, a few modifications, as compared to the 2D examples, are introduced and explained in the flow chart in **Fig. 5.1**.

5.3.1 Synthetic Model Setup and MDS Analysis

Instead of generating a total of N_{3D} = 1000 shale configurations randomly (as in the 2D examples), a total of N_{3D} = 300 3D cases are constructed by stacking 4 random 2D segments from the 471 screened cases. There are two key justifications for this procedure. Generally speaking, it would be an inefficient use of computational effort to run flow simulations for 1000 3D cases and then discard a significant portion of these cases at a later step. Therefore, it is recommended to start with a smaller number of initial cases, and add more cases systematically after performing the MDS and clustering analyses.

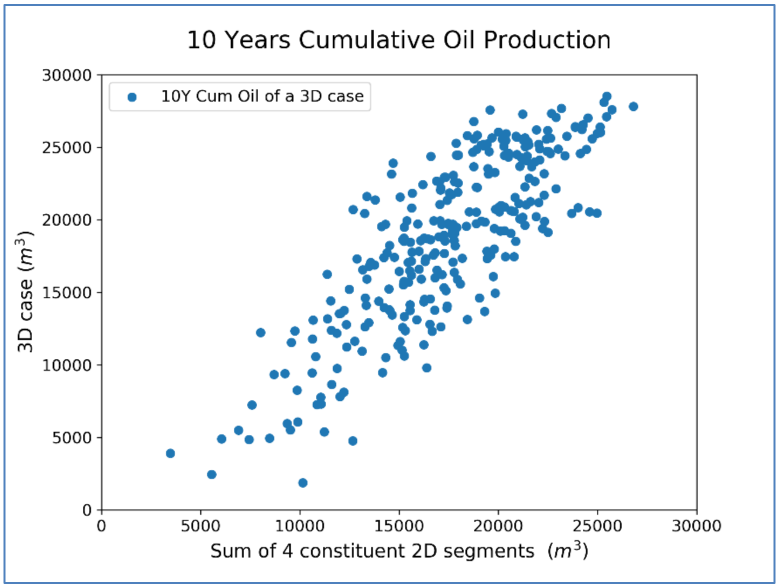
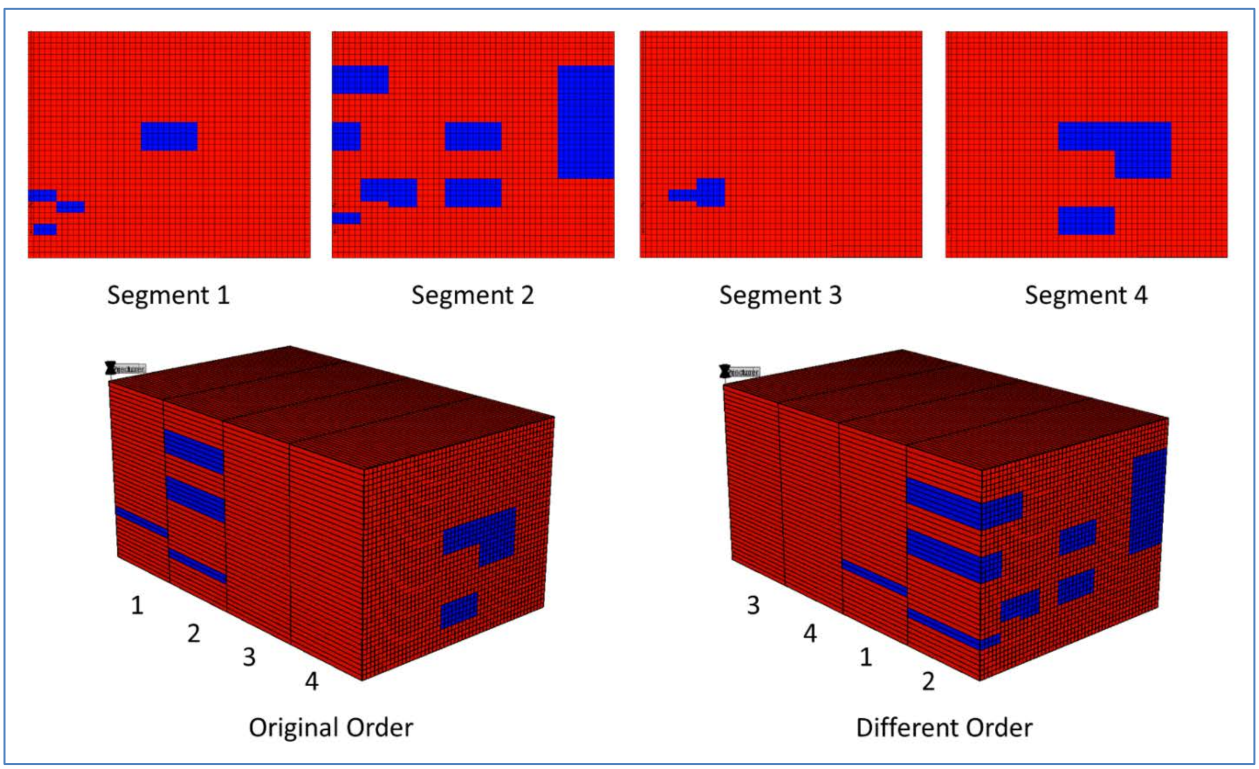


Fig. 5.8 – Comparison of cumulative oil production after 10 years between a certain 3D case and the sum of production from its constituent 2D segments.

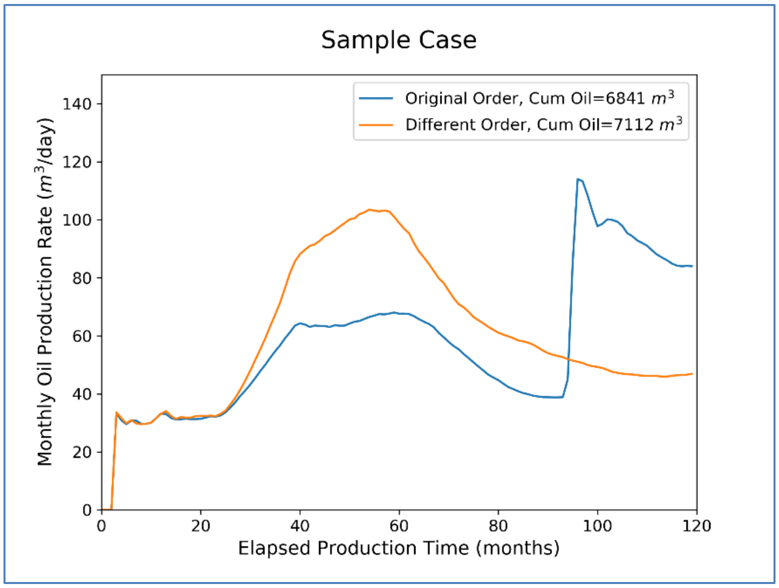
Based on the recovery behavior, there is a strong correlation between each 3D case and its constituent 2D segments. For example, cumulative oil production after 10 years corresponding to a given 3D case is highly consistent with the direct sum of production from its 4 constituent 2D segments. As illustrated in **Fig. 5.8**, among the 300 original cases, the maximum total difference is only 7%. However, a different ordering of the 2D segments can have an impact on the production profile. An example of two different realizations of a 3D case constructed by re-arranging the same 2D segments in different order is shown in **Fig. 5.9**. The results of the simulation indicate that the specific arrangement of the 2D segments may lead to different flow responses (i.e., production profiles). Therefore, the stacking order should be considered when analyzing the 3D behavior.

It is hypothesized that a decision tree model (or other supervised-training technique) can be trained to capture the correlation between the behavior of each 3D case and the corresponding behavior of its 2D segments, and generate additional cases corresponding to individual target clusters. Specifically, utilizing the clustering results of 2D cases in **section 5.2**, a decision tree model will be integrated to systematically propose different stacking of 2D segments extracted from certain clusters.

By stacking a series of 2D cases, the spatial continuity of the shale distribution along *y*-direction across adjacent segments is not captured. As mentioned previously, the 3D simulation model is 50 m × 80 m × 40 m (4 2D segments, each with Δy = 20 m, along the y-direction). All 300 cases are subjected to numerical simulation, and their production histories are used to compute the matrix $\boldsymbol{D}$. The distribution of all cases in the two-dimensional Euclidean space $\boldsymbol{R_{3D}}$ after applying MDS is shown in **Fig. 5.10.**



(a)



(b)

Fig. 5.9 – (a) Different realizations of a 3D case constructed by re-arranging of the same 2D segments in different order; (b) oil production profiles corresponding to the two 3D realizations in (a).

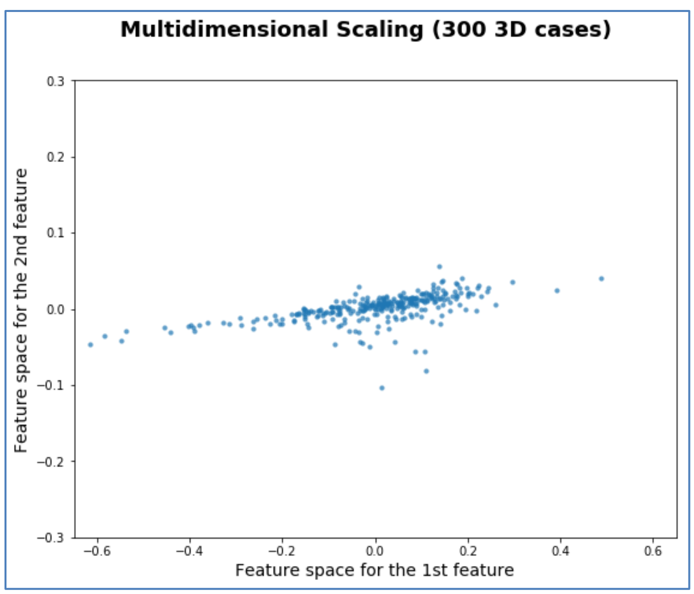


Fig. 5.10 – MDS results in the features space ($\boldsymbol{R_{3D}}$) calculated based on the original N_{3D} = 300 3D cases.

5.3.2 Cluster Analysis and Addition of New Cases

K-means clustering was applied to analyze the dissimilarity and internal structures of these 300 cases. Different number of clusters (ranging between 4 and 10) were tested to identify the optimal value for N_c. The corresponding average silhouette values for the N_{3D} cases are listed in **Table 5.4.** It is evident that the optimal number of clusters for 3D cases is N_c = 4 (with the highest average silhouette value). The K-means results are shown in **Fig. 5.11**; the corresponding oil rate profiles for all clusters are shown in **Fig. 5.12**, and the number of cases within each cluster is compared in row 1 of **Table 5.5.**

Table 5.4 – Average silhouette value for different number of clusters (original 300 3D cases).

Number of Clusters	4	5	6	7	8	9	10
Average Silhouette Value	0.505	0.479	0.461	0.469	0.467	0.478	0.468

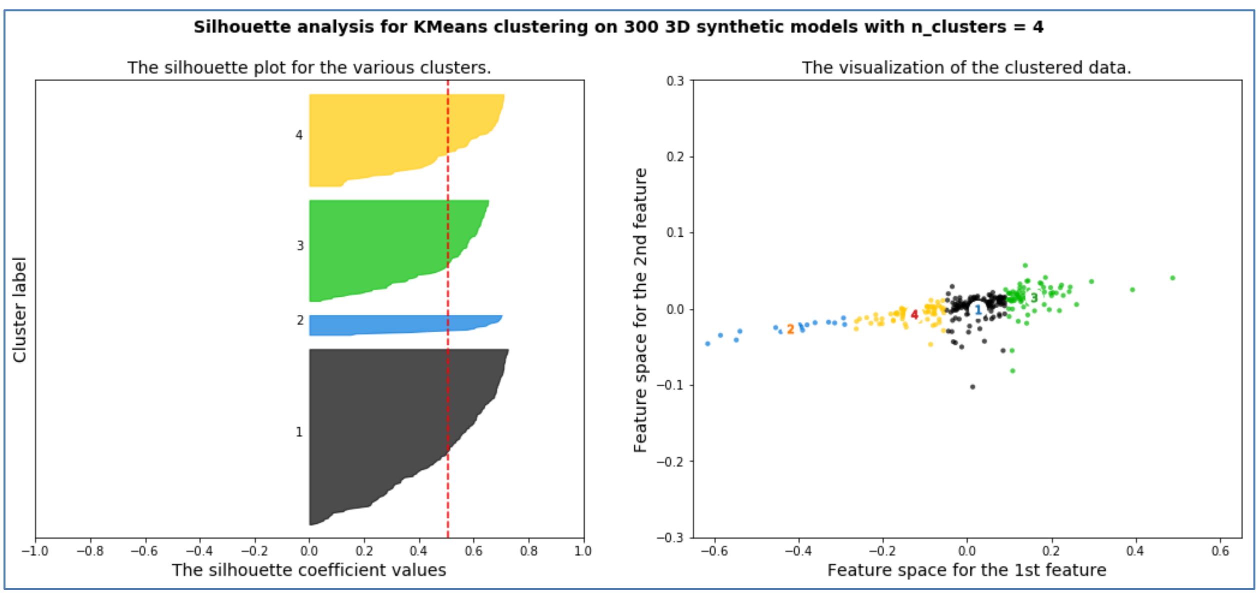


Fig. 5.11 – K-means clustering results corresponding to 4 clusters of the original 300 3D cases.

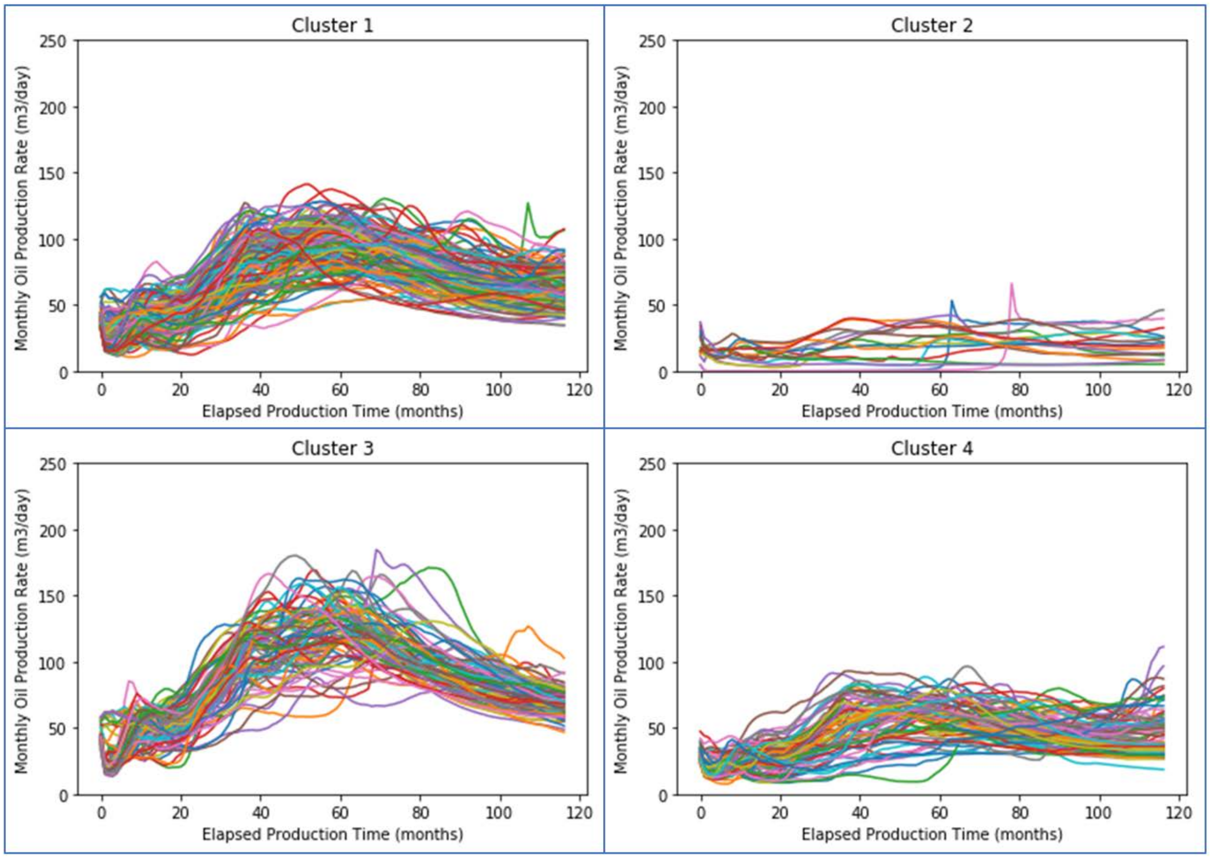


Fig. 5.12 – Production profiles for each of the 4 clusters of the original 300 3D cases.

Similar to the 2D cases, these 300 3D original cases are not optimally distributed among the 4 clusters; for example, there are many fewer cases in cluster 2 than there are in cluster 1. To improve the spanning of model parameters in under-sampled clusters, it would be desirable to create additional cases sharing similar shale barrier characteristics as other members of the cluster. However, the shale barrier configurations in 3D are more complex; hence, it would not be straightforward to visualize and summarize features of the shale barrier configurations for every cluster and directly propose a set of new (modified) cases, as described in step 2 of **section 5.2.3**. An alternative approach would be to assume that the performance of a 3D case is directly correlated to the specific arrangement or sequence of its corresponding 2D segments and their flow patterns. In particular, it is hypothesized that the cluster to which a given 3D case would belong to can be predicted from the cluster labels of the corresponding sequence of 2D segments. The following procedure is implemented:

(1) Assemble a training dataset for the decision tree model based on the original N_{3D} cases. Each training record consists of 4 input variables [c_1, c_2, c_3, c_4], which are the cluster labels of the 2D segments; the subscript refers to the position of a particular 2D segment along the well trajectory (e.g., 1 – closest to the heel; 4 – closest to the toe). There is a single output consisting of the corresponding 3D cluster label c_{3D};
(2) Build a decision tree model to capture the correlation between [c_1, c_2, c_3, c_4] and c_{3D};
(3) To propose a new 3D case to be added to the entire data set, randomly sample (with replacement) a series of 4 out of the 471 2D cases obtained in **section 5.2.** Parameterize these 4 cases into [c_1, c_2, c_3, c_4] using the 2D cluster labels identified in **section 5.2.2**;
(4) Predict the value of c_{3D} for the proposed 3D case in (3) using the decision tree trained in step (2);
(5) Repeat steps (3) and (4) until desired number of cases are generated for all clusters.

A total of 150 new cases were generated following the aforementioned procedure. The training performance of the decision tree model is shown in **Fig. 5.13**. The predicted cluster labels of these 150 additional cases, as well as the labels of entire data set consisting of 450 cases, are listed in **Table 5.5**. It is intended to populate the under-sampled Clusters 2, 3 and 4 with more cases. Therefore, as illustrated in row 2 of **Table 5.5** , steps (3) and (4) of the procedure were carried out such that, according to the prediction of c_{3D}, 30% (135 cases) of the entire data set would belong to each of Clusters 1, 3 and 4, respectively, while the remaining 10% (45 cases) would be assigned to Cluster 2. This is reflected in row 4 of **Table 5.5**. The justification for assigning fewer cases to Cluster 2 is that most cases in that cluster have relatively low production rates (peak rate < 50 m^3/day), since their constituent 2D segments are mostly extracted from Clusters 2 and 4 (**Table 5.2** and **Fig. 5.5**), where laterally-extensive shale barriers are located in between the injector and producer. Therefore, it was decided that these cases represented extreme scenarios and hence, should contribute to only a small portion of the entire data set.

Table 5.5 – Number of cases in all clusters (3D examples).

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
[1] Original 300 Cases (Actual Label)	135	16	78	71	300
[2] Additional 150 Cases (Predicted Label Based on the Decision Tree Model)	0	29	57	64	150
[3] Additional 150 Cases (Actual Label after MDS + K-Means Analysis)	34	33	44	39	150
Total 450 Cases ([1] + [2])	135	45	135	135	450
Total 450 Cases ([1] + [3])	169	49	122	110	450

To test the accuracy of the proposed procedure for generating these new cases, the 150 additional cases were subjected to numerical simulation. The production histories of all 450 cases were subjected to MDS and K-means analysis once again, and 4 clusters were identified (**Fig. 5.14**). Despite of some discrepancies in the predicted and target cluster labels, as shown in rows 2 and 3 of **Table 5.5**, the procedure was broadly effective in accomplishing its objectives: (1) the number of cases assigned to Cluster 2 was tripled; (2) many additional cases were correctly assigned to Clusters 3 and 4; due to the proximity of Cluster 1 to Clusters 3 and 4 (see **Figs. 5.11** and **5.14**), several additional cases were assigned to Cluster 1 eventually, instead of Clusters 3 and 4**.** The distribution of all cases in the new feature space $\boldsymbol{R_{3D}}'$ after applying MDS is shown in **Fig. 5.15**, where the green dots denote the original 300 cases and the red dots represent the additional 150 cases. It is encouraging to observe that the 150 new cases have successfully enhanced the spanning across the feature space. The key advantage of this procedure is that it enables us to effectively propose a set of new cases that would correspond to certain specific cluster features, prior to subjecting them to flow simulation.

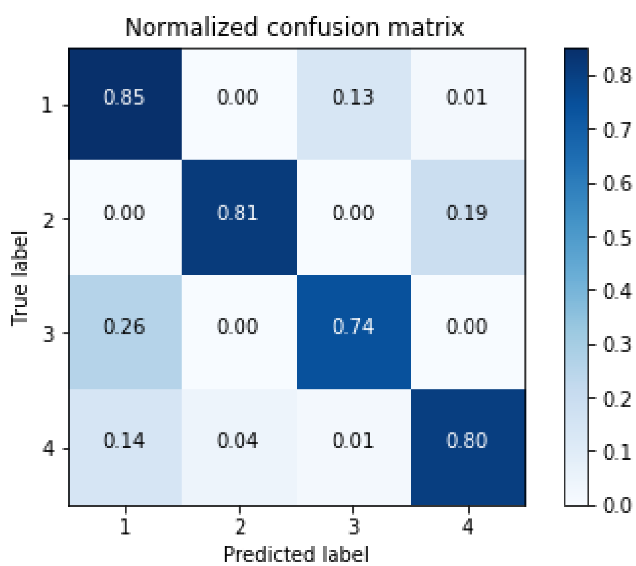


Fig. 5.13 – Training performance of the decision tree model for prediction of 3D cluster label from the cluster labels of its corresponding 2D segments.

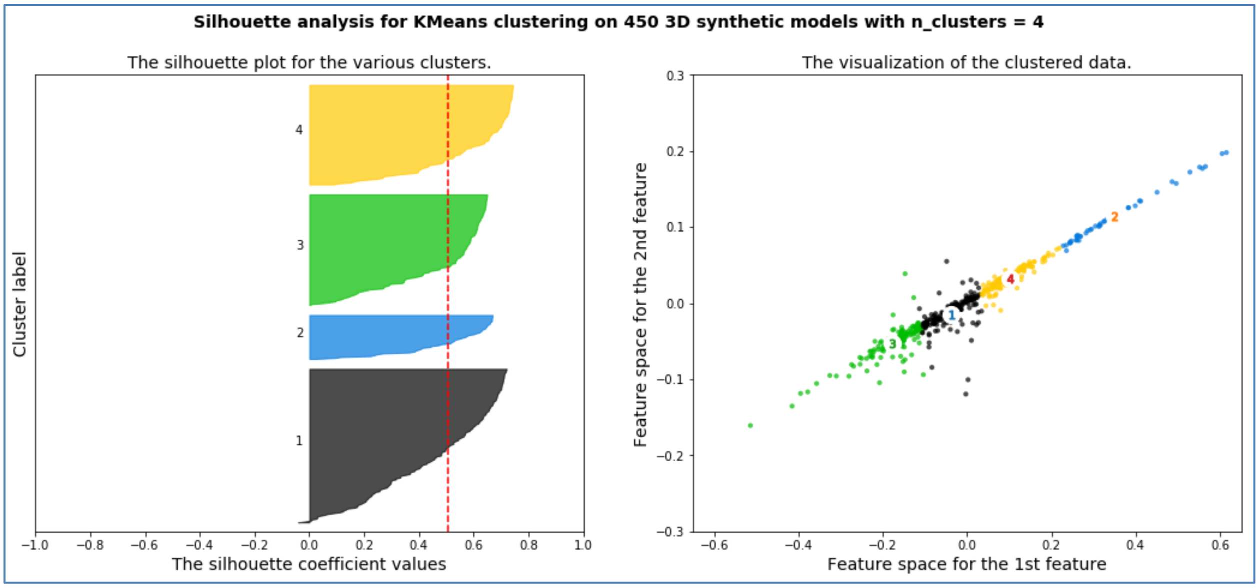


Fig. 5.14 – K-means clustering results corresponding to 4 clusters of the total 450 3D cases.

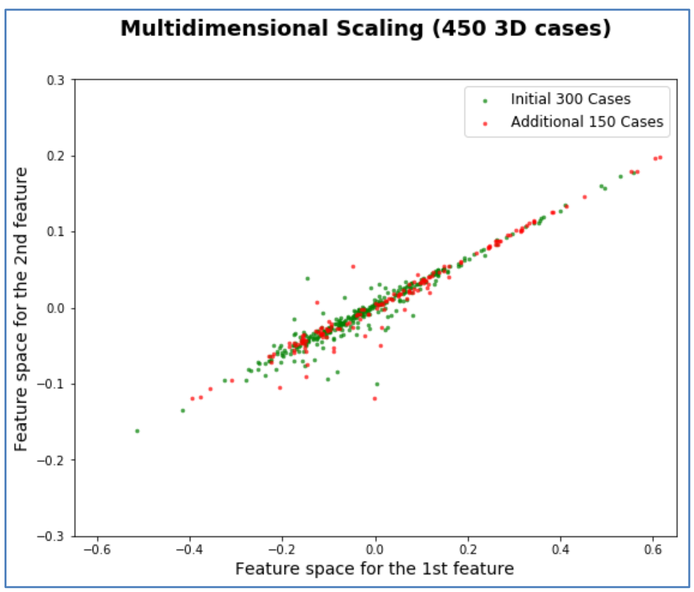


Fig. 5.15 – MDS results in the revised features space ($\boldsymbol{R_{3D}'}$) calculated based on the total N_{3D}' = 450 3D cases.

5.4 Summary

In this chapter, a new workflow is proposed to visualize (dis)similarities among realizations of shale barrier configurations and to quantify their influences on SAGD production. The method is tested using a set of 2D and 3D synthetic cases using representative petrophysical properties and operating constraints extracted from Suncor's Firebag project. A flow-based distance function is formulated, and multidimensional scaling (MDS) is applied to project the dissimilarity matrix to an *n*-dimensional Euclidean space, where redundancy in the original data set can be identified and cases that are too close to one another can be discarded. The reduced data set is subjected to MDS again, and K-means clustering is conducted to identify internal groupings among the data set.

For 2D cases, clustering results confirm that the number of shale barriers in the near-well region, particularly those that are located in between the well pair, and the lateral extent of individual shale barriers are important factors impacting SAGD performance. New cases are added to several clusters with limited members, where only the shale barriers in the near-well region (zones 1 and 2) are kept. The new cases are subjected to flow simulation to verify their membership to the assigned clusters. The comparison of flow simulation results between the original cluster samples and the new (modified) cases confirms that consistent cluster features are shared between both sets of cases, further supporting the hypothesis that shale barriers that are located outside of the near-well region would have minimal impact on production performance.

For 3D cases, a decision-tree model is incorporated to propose new cases to reflect specific cluster features. This is based on the assumption that the performance of a 3D case is directly correlated to the specific arrangement or sequence of its corresponding 2D segments and their flow patterns. The decision tree model is used to predict the cluster to which a given 3D case would belong to from the cluster labels of the corresponding sequence of 2D segments. The new cases are subjected to flow simulation to verify their membership in the assigned clusters.

Despite some discrepancies in the predicted and target cluster labels, the implemented procedure is effective in proposing new cases that would significantly enhance the spanning of the model feature space.

The proposed workflow will be used in next chapter to identify internal structures among 3D shale configurations and select representative cases for establishing separate proxy models corresponding to different characteristics.

CHAPTER 6: ESTIMATING SHALE BARRIER CONFIGURATIONS IN 3D EXAMPLES

A two-step workflow is successfully implemented in Chapter 4 to infer the presence and distribution of heterogeneous shale barriers in 2D synthetic cases. Chapter 5 proposed a workflow consisting of multidimensional scaling (MDS) and cluster analysis technique to examine and visualize the (dis)similarities among these shale barrier configurations in terms of production behavior. The results identify internal structures in both 2D and 3D cases, and show that cases belonging to the same cluster would exhibit similar shale configuration characteristics and production features. Due to the complex relationship between 3D shale barrier configurations and the production time-series, it is anticipated that quantifying the internal structures in 3D cases and constructing separate forward models for individual clusters would improve the performance of the workflow.

In this chapter, a three-part workflow incorporating cluster analysis and multiple forward models is proposed for inferring 3D shale barrier configurations from SAGD production histories:

(1) Multidimensional scaling (MDS) and cluster analysis are applied to identify the internal structures of the synthetic cases in the training dataset based on (dis)similarities in their production profiles;

(2) For each cluster, a forward forecasting model (e.g., neural network or support vector regression) is trained to estimate SAGD production time series (outputs) as a function of time-dependent steam injection constraint and the shale barrier configuration parameters;

(3) To infer the unknown shale configuration corresponding to a new SAGD production history, which was not involved during the training stage, its production profile is first analyzed and

assigned to one of the clusters identified in step 2. Next, a hybrid inverse modeling scheme that combines the genetic algorithm (GA) and previously-trained forward model is adopted. The proposed workflow is depicted in **Fig. 6.1**.

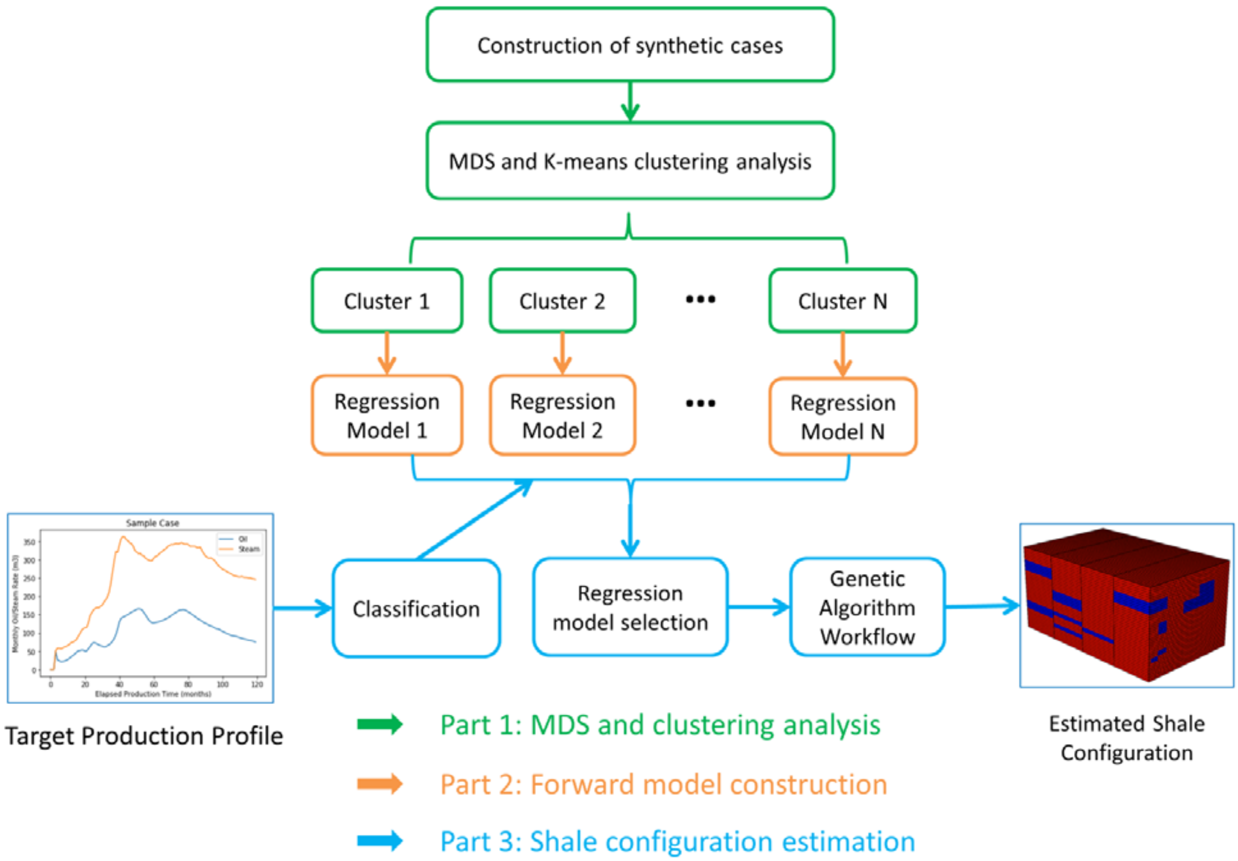


Fig. 6.1 – Schematic of the proposed workflow.

6.1 Multidimensional Scaling and K-means Clustering

In Chapter 5, MDS and K-means cluster analysis were successfully implemented to quantify the dissimilarities between 2D SAGD realizations, where <500 2D cases were selected to represent the cluster features exhibited by 1000 2D cases in the original space. The results reveal that the vast feature space of reservoir heterogeneity can be effectively represented by a small set of selected shale barrier configurations. Based on previous study, a set of 3D synthetic cases are generated with the following procedure:

(1) 1000 2D cases with different shale configurations are constructed and parameterized according to **section 3.3.3**. Experimental design is carried out following the methodology described in Chapter 4, where various features of single basic shale unit, lateral-extensive shale barriers, vertical-extensive shale barriers and multiple disjoint shale barriers are considered.

(2) MDS analysis is implemented on the 1000 2D cases (step 1) and redundancy is reduced by removing closely located cases. As a result, 425 2D cases are retained;

(3) K-means clustering analysis is applied on the 425 2D cases and 5 different clusters are identified (**Fig. 6.2**);

(4) Each 3D case is constructed by randomly sampling four of the 425 2D segments (step 3) and stacking them arbitrarily along the y-direction. This strategy ensures that more populated clusters in the 2D feature space would be more abundantly represented in the resultant 3D cases.

All N_{3D} = 1200 cases are subjected to flow simulation (CMG STARS, 2015), and the simulated production results are recorded. MDS analysis is performed on the 3D cases, and redundancy is reduced by retaining only a subset of 585 cases. The distributions of the original 1200 3D cases and the reduced set of 585 cases in the original MDS feature space are compared in **Fig. 6.3**. Each point represents one 3D case and the distance between any two cases is correlated to distance *d* (**Eq. 5.1**).

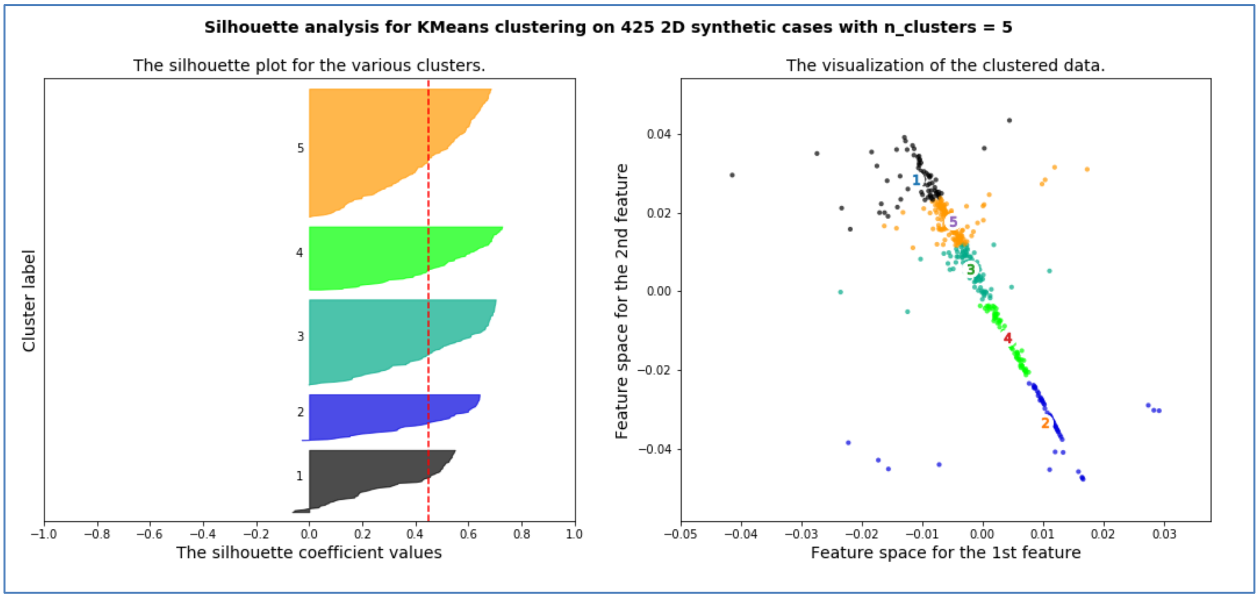
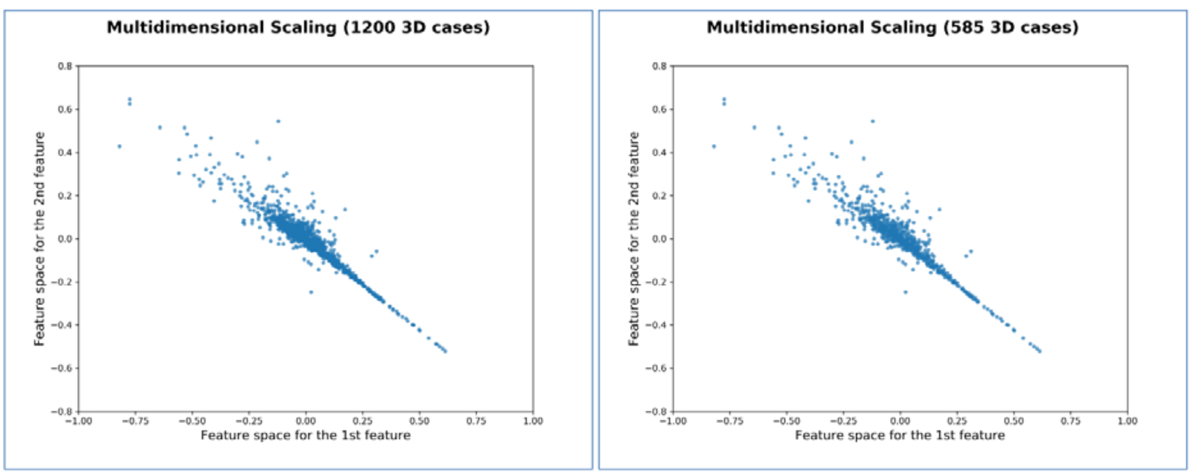


Fig. 6.2 – K-means clustering and MDS analysis corresponding to the 2D dataset.



(a) Original dataset with 1200 3D cases (b) Screened dataset with 585 3D cases

Fig. 6.3 – MDS analysis based on N_{3D} = 1200 3D cases.

A set of N_c (ranging between 4 to 10) clusters can be identified from these 585 cases, and the corresponding average silhouette values are listed in **Table 6.1**. The average silhouette value is relatively constant for N_c = 4, 5, 6, but it starts to decline as N_c increases, indicating that dividing the dataset into more than 6 clusters may lead to overfitting and is not ideal. Further examination reveals that by increasing N_c from 5 to 6, the additional cluster would consist of only 6 cases (~ 1% of the entire dataset). Therefore, in order to maintain a reasonable degree of cluster coherence, while achieving sufficient distinction among different clusters, N_c = 5 is selected as the optimal number of clusters. The MDS and K-means clustering results are shown in **Fig. 6.4**, and corresponding number of cases in each cluster are listed in **Table 6.2**.

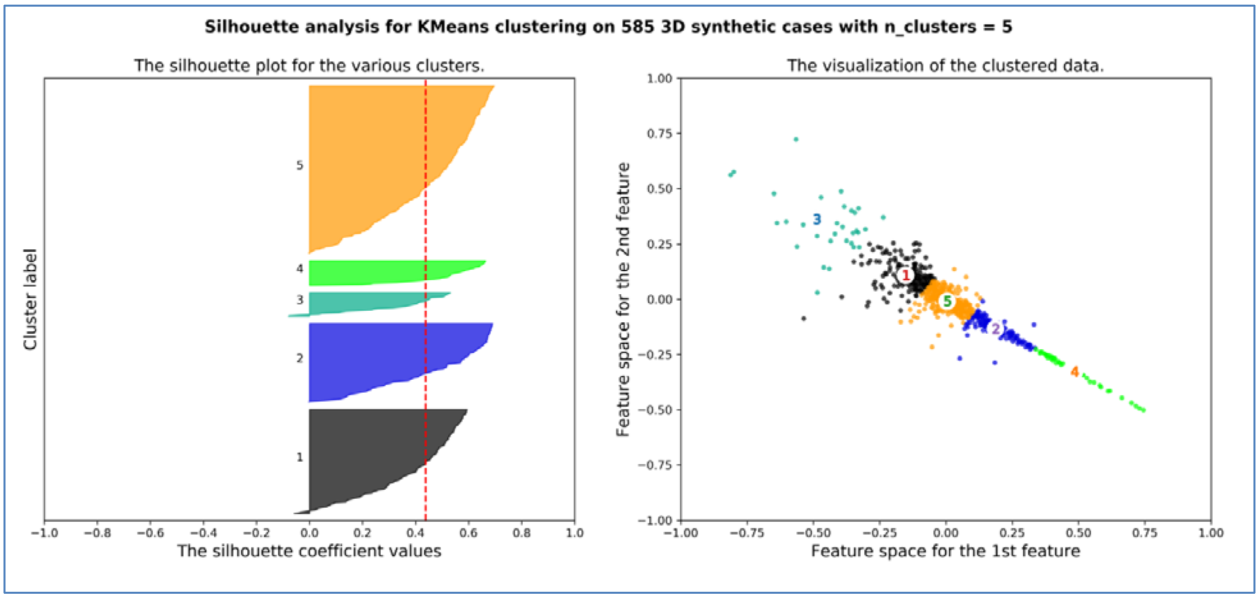


Fig. 6.4 – K-means clustering and MDS analysis corresponding to the 3D dataset.

Table 6.1 – Average silhouette value for different number of clusters (3D cases).

Number of Clusters	4	5	6	7	8	9	10
Average Silhouette Value	0.464	0.46	0.46	0.45	0.442	0.418	0.405

Table 6.2 – Number of cases in different clusters (3D cases).

Cluster	1	2	3	4	5	Total
Number of cases	153	115	35	37	245	585

In each cluster, a set of average production profiles are computed and denoted as “type curves” (**Fig. 6.5**). Further examination is conducted to understand the characteristic shale barrier configuration pertinent to specific clusters. It is widely acknowledged that SAGD production is highly sensitive to the distribution of shale barriers in the near well region (Kisman and Yeung 1995; Chen et al., 2008; Le Ravalec et al., 2009; Wang & Leung, 2015; Zheng et al., 2018a). Therefore, it is postulated that the distinguishing characteristics among these clusters would likely be their shale barrier configurations in the near-well region. To that end, potential shale barriers in four particular near-well locations are examined, and they are marked as “Shale Location 1” to “Shale Location 4” in **Fig. 6.6**; the frequency of occurrence corresponding to each of these four near-well shale barriers are analyzed. For example, in a given 3D case consisting of four 2D segments, the maximum frequency corresponding to a specific near-well shale barrier is 4, implying that the same near-well shale barrier (e.g. Shale Location 1) is present in all four 2D segments. The histograms of these four near-well shale barriers are shown in **Fig. 6.7**. For each cluster, values of P95 and P5 are extracted from the histograms, as listed in **Table 6.3**. These values in **Table 6.3** describes the most impactful shale barrier characteristics associated with individual clusters, which can be interpreted as constraints on the shale barrier configurations for each cluster.

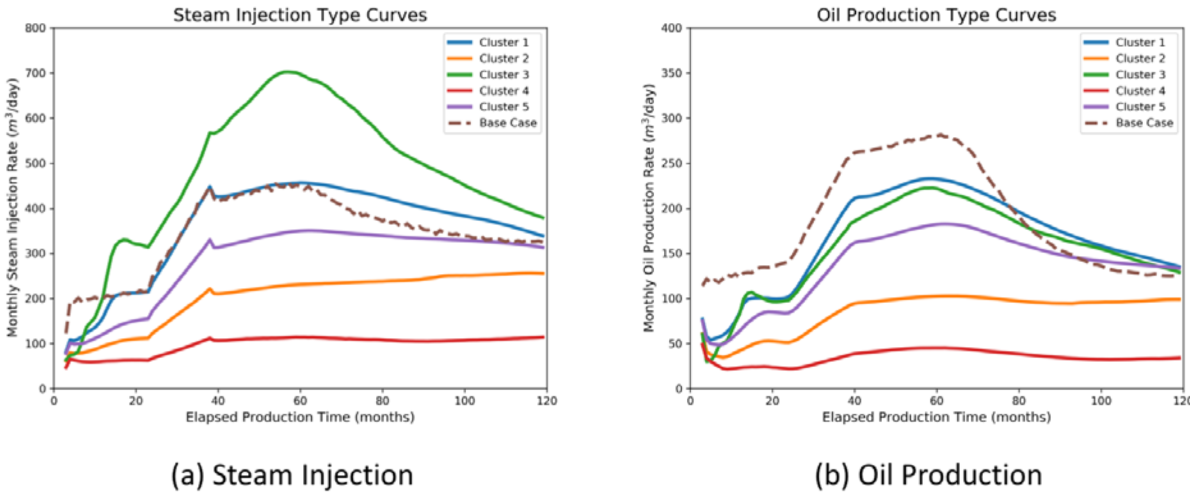


Fig. 6.5 – Average steam injection and oil production profiles (type curves) corresponding to the 5 clusters of 3D cases.

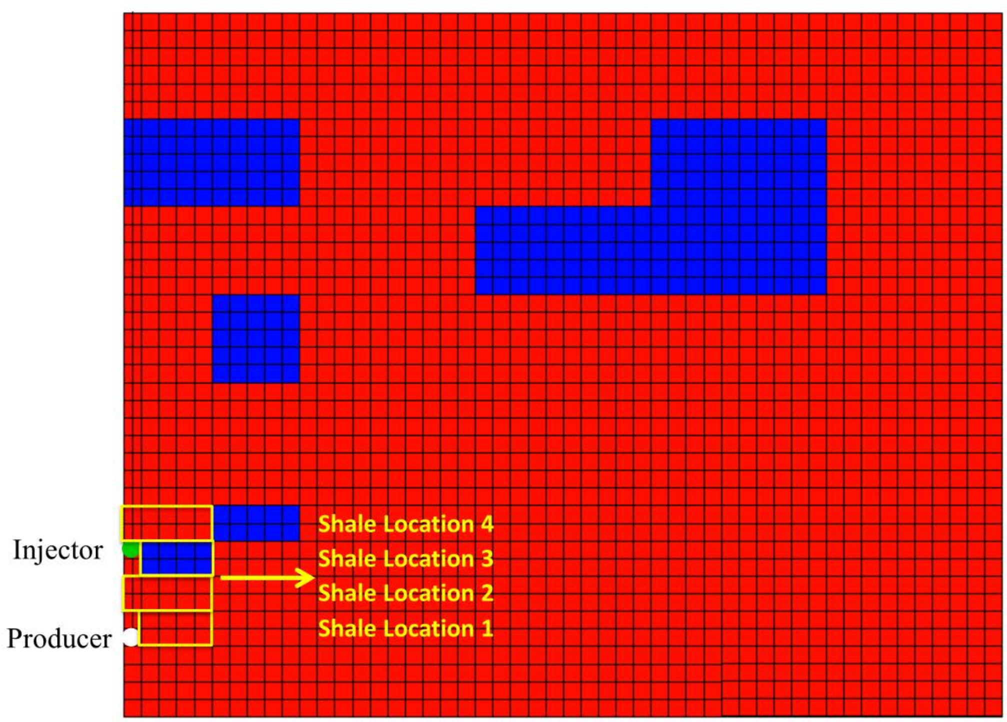
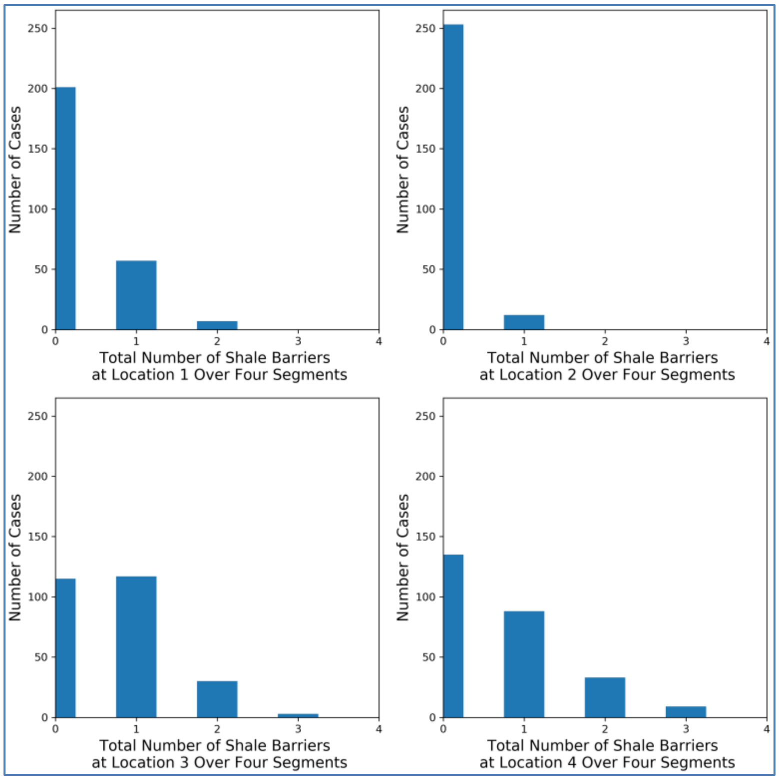
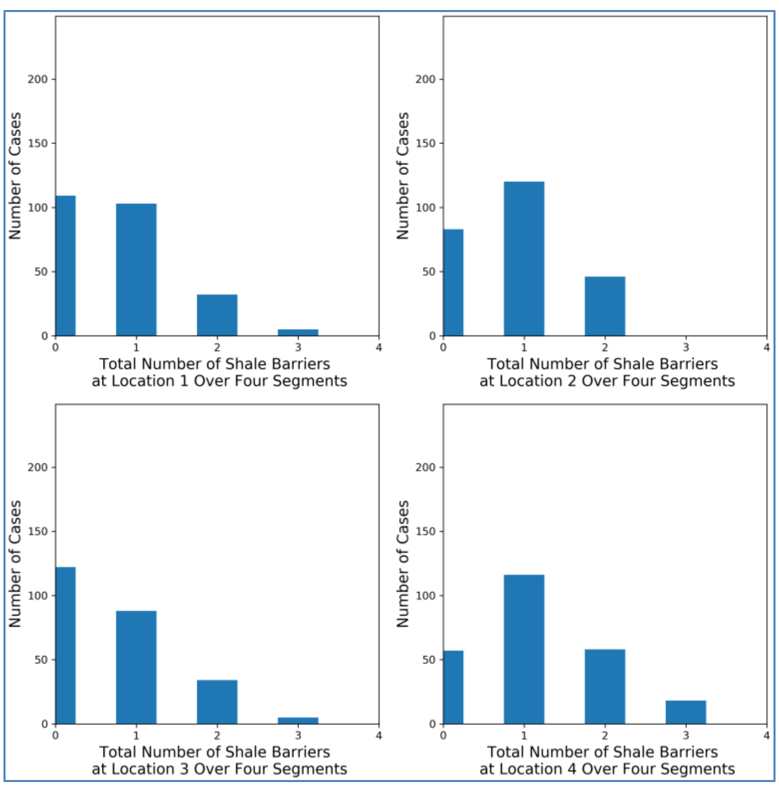


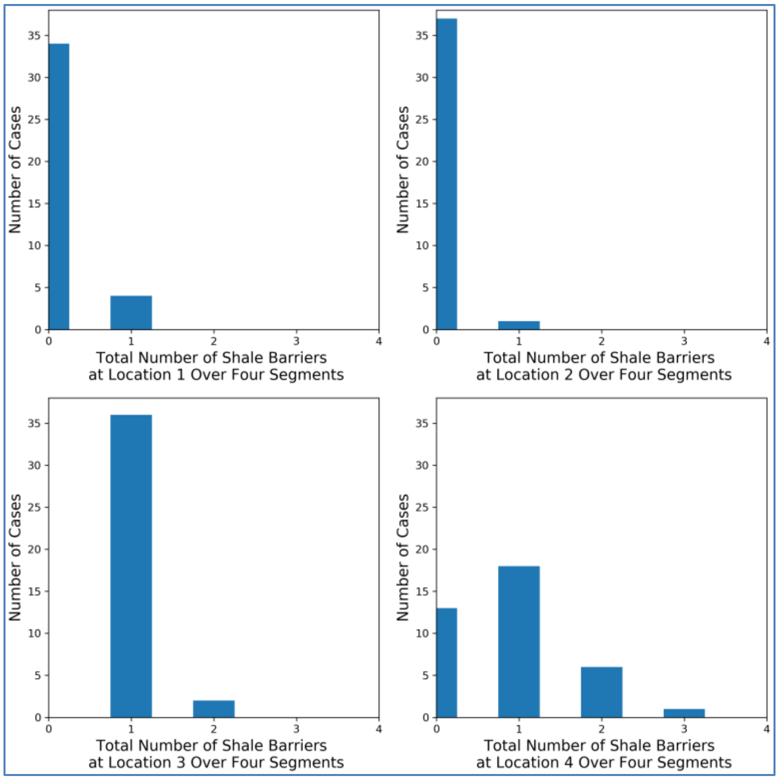
Fig. 6.6 – Illustration of four near-well shale locations used in cluster feature analysis.



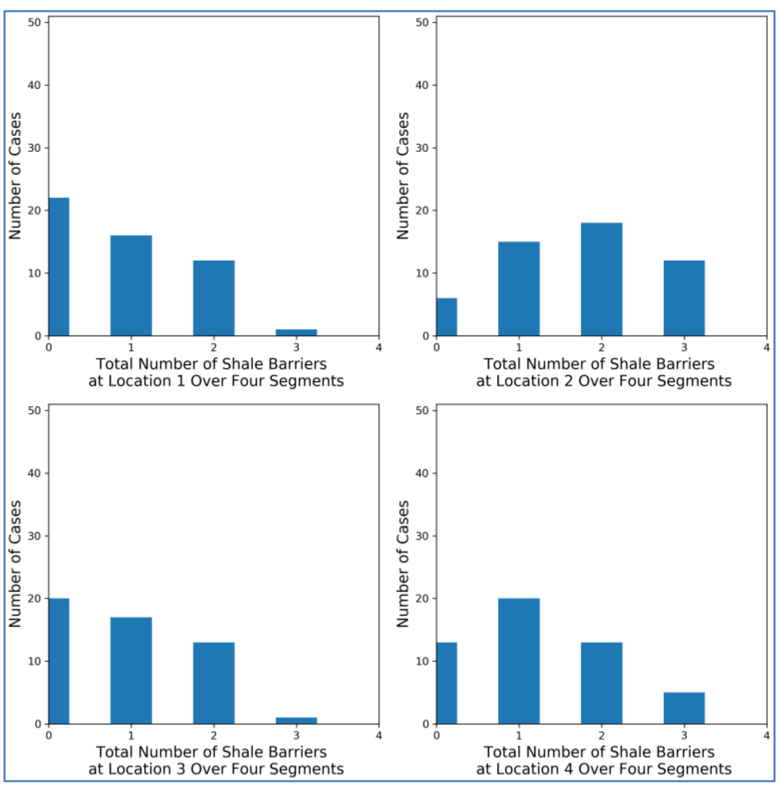
Cluster 1



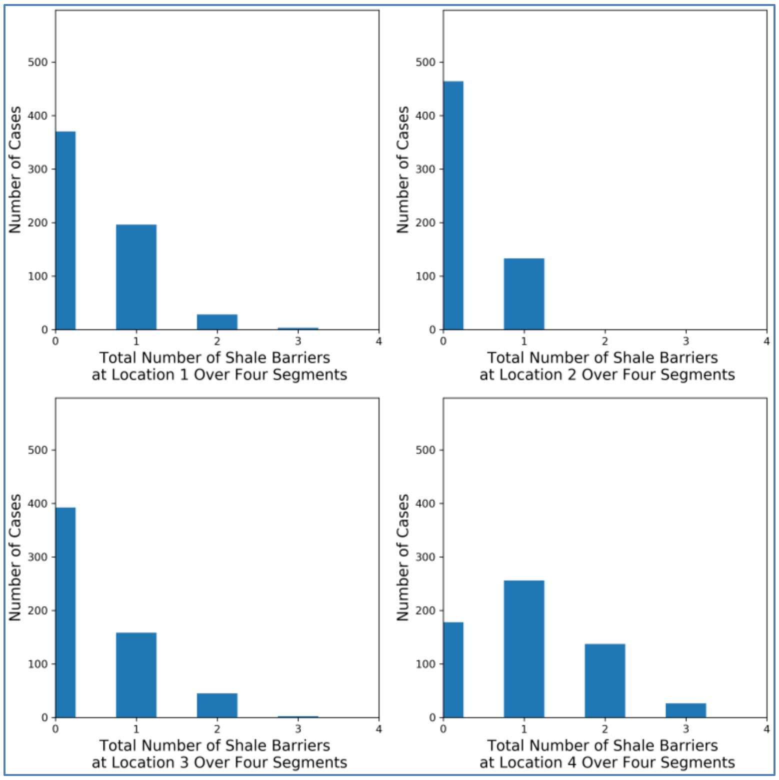
Cluster 2



Cluster 3



Cluster 4



Cluster 5

Fig. 6.7 – Histograms of total number of shale barriers at four near-well locations for different clusters.

Table 6.3 – 5^{th} and 95^{th} percentiles from the shale frequency distributions of shale barriers at the four near-well locations.

	Location_1		Location_2		Location_3		Location_4		Sum	
	5^{th}	95^{th}	5^{th}	95^{th}	5^{th}	95^{th}	5^{th}	95^{th}	5^{th}	95^{th}
Cluster 1	0	1	0	0	0	2	0	2	0	4
Cluster 2	0	2	0	2	0	2	0	3	1	6
Cluster 3	0	1	0	0	1	1	0	2	0	3
Cluster 4	0	2	0	3	0	2	0	3	2	8
Cluster 5	0	2	0	1	0	2	0	2	0	5

If the P5 values from all four selected locations are combined to infer an estimate of the total number of shale barriers in the near-well region, the following relation can be inferred in terms of shale content: Cluster 3 ≤ Cluster 1 ≤ Cluster 5 ≤ Cluster 2 ≤ Cluster 4. An inverse relation is observed in the corresponding injection type curves (**Fig. 6.5**): the average steam injection rate follows an opposite trend: Cluster 3 ≥ Cluster 1 ≥ Cluster 5 ≥ Cluster 2 ≥ Cluster 4. A similar trend can be observed for the oil production profiles, except for that the oil production rate for Cluster 1 is slightly higher than Cluster 3. Another interesting observation is that Clusters 3, 1, 5, 2, and 4 are sequentially positioned along a straight line in the Euclidean space after MDS analysis (**Fig. 6.4**). The distance in the transformed Euclidean space is supposed to reflect the dissimilarities among realizations. These observations corroborate the expectation that shale barriers located in between the injector and producer would likely to hamper the steam chamber development, as illustrated by Clusters 2 and 4.

As mentioned in the previous paragraph, there are certain peculiar features associated with Cluster 3: although the steam injection rate is higher than that of Cluster 1, the corresponding oil production rate is lower. This observation suggests that the incremental steam injection does not contribute to an increase in oil production. Further examination of the distribution of shale barriers in the near-well region may offer a possible explanation. The results in **Fig. 6.7** and **Table 6.3** suggest that, for most cases in Cluster 3, there is a shale barrier at Shale Location 3 (adjacent to the injector) in one of the segments, but none at Shale Locations 1 and 2. To better understand the impacts of this particular shale barrier at Location 3, a case from Cluster 3 is randomly selected, and a comparative case is created by removing the shale barrier at Location 3 in segment 1, as shown in **Fig. 6.8**. Their corresponding steam injection and water production profiles are compared in **Fig. 6.9**. For the original case, the shale barrier at Location 3 in segment 1 seem to have a dramatic impact on the steam injection and water production in all 4 segments: in segment 1, water production is higher than the injection, implying that the steam injected in segment 2 is migrating to segment 1; the steam injection and water production in segments 2 and 3 are also much higher than those in segment 1. The temperature and pressure distributions in segment 1 of the two cases are compared in **Fig. 6.10** and **Fig. 6.11**, respectively. The flow vectors of the gas phase are also depicted. It is clear that the pressure in segment 1 is lower near the producer in the original case; this is because the injectivity in segment 1 is limited by the shale barrier at Location 3, while fluid production is constrained with a constant bottom-hole pressure in all segments, inadvertently creating a pressure gradient between segment 1 and others. The pressure distribution (cross-sectional view in the *y*-*z* plane) of original case in **Fig. 6.12** substantiates the observation that the injected steam is quickly produced where there is a shale barrier at Location 3, while pressure and temperature (**Fig. 6.10**) are not building up as expected in other segments. This analysis substantiates the hypothesis that any shale barriers in the near-well locations are particularly important and should be considered as key distinguishing features associated with each clusters. It will be illustrated later in **section 6.3** that the ranges in **Table 6.3** can be used as screening criteria for cluster membership.

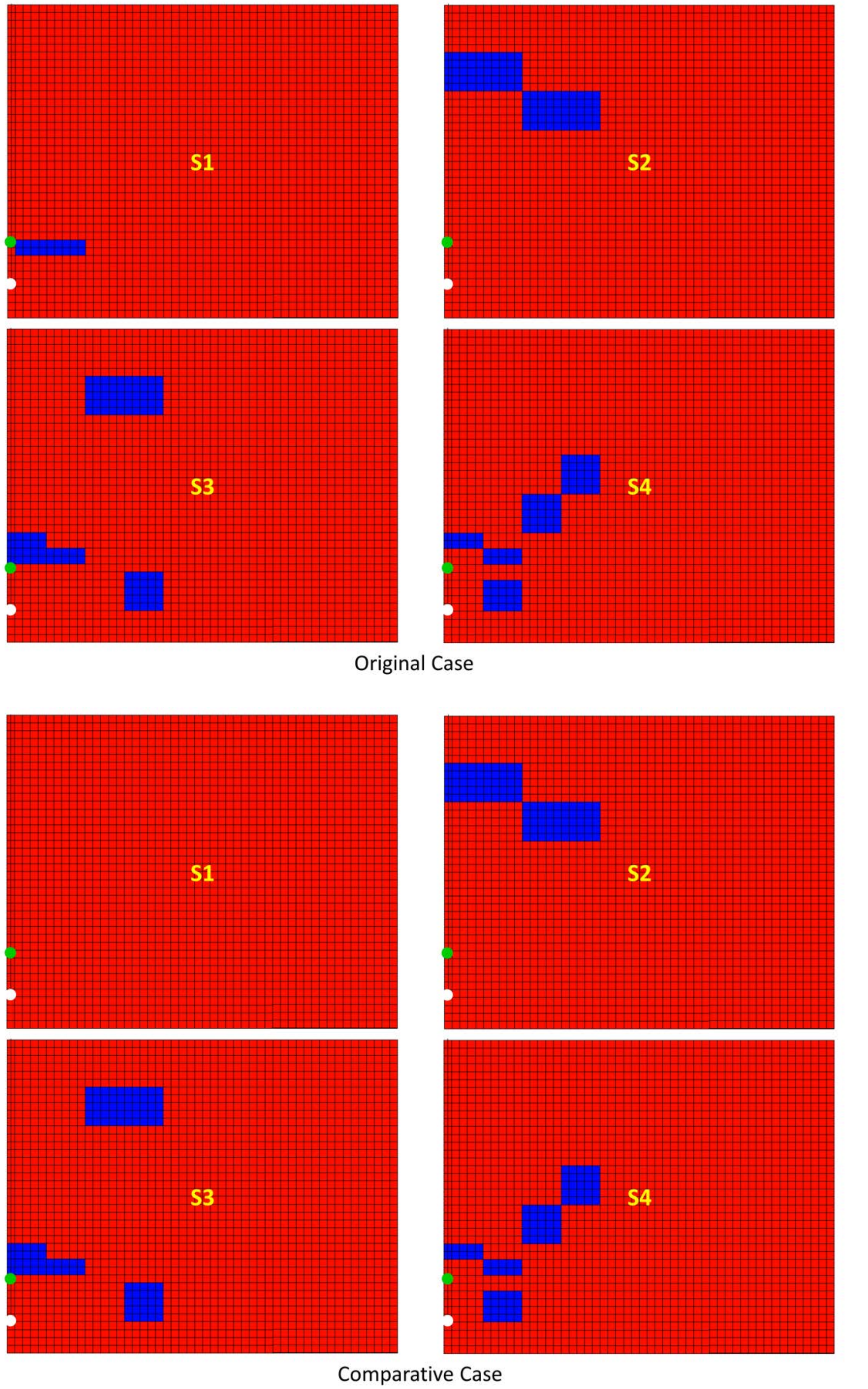
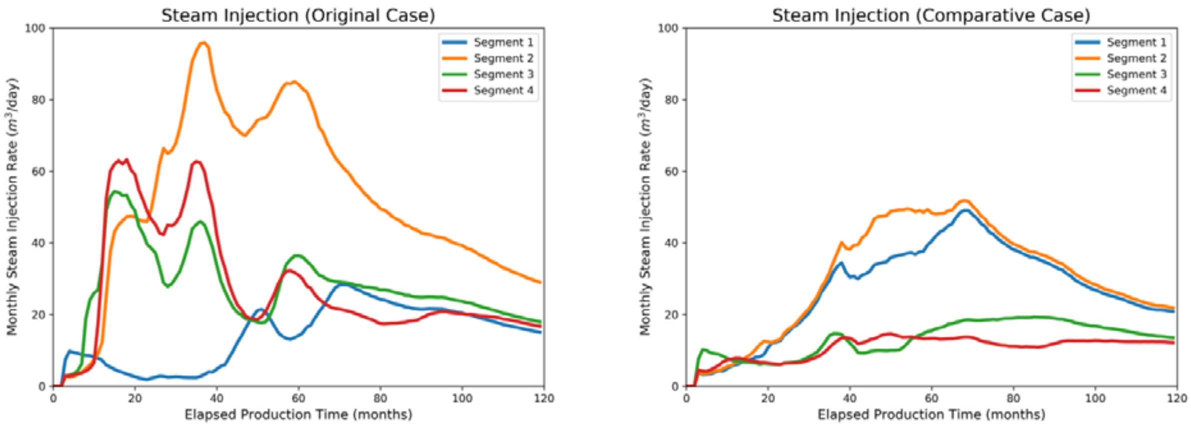
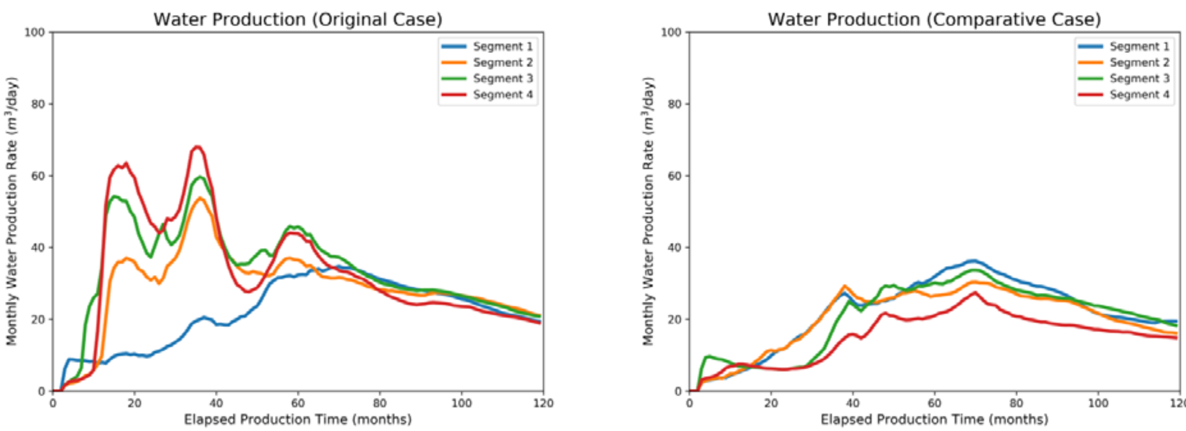


Fig. 6.8 – Shale configurations of an original case from Cluster 3 and a comparative case, where the shale barrier at Location 3 in segment 1 is removed.



(a)



(b)

Fig. 6.9 – Production profiles of the original and comparative cases: (a) steam injection; (b) water production.

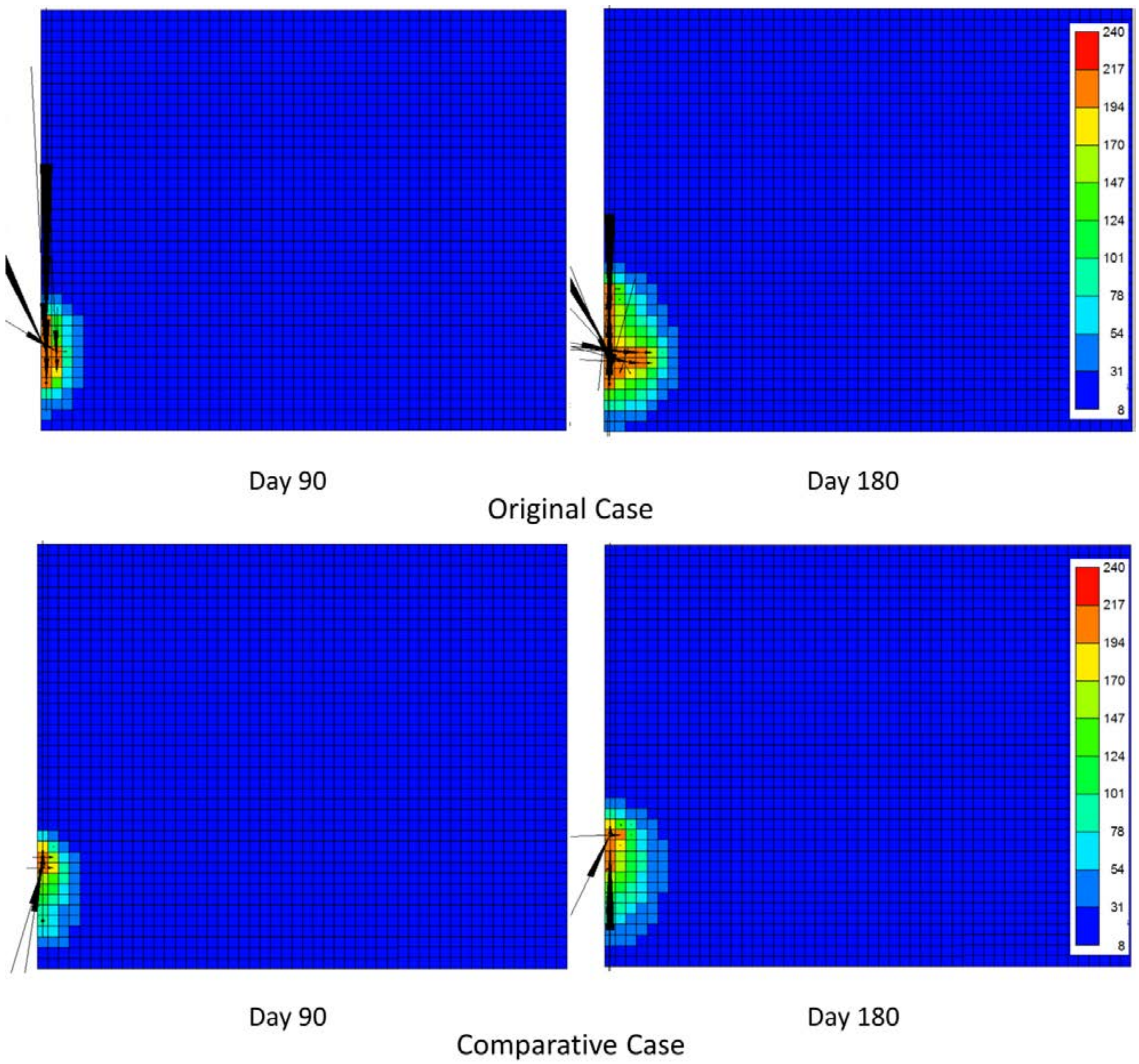


Fig. 6.10 – Temperature (℃) distribution (cross-sectional view in the *x-z* plane) in segment 1. The arrows represent flow vectors of the gas phase.

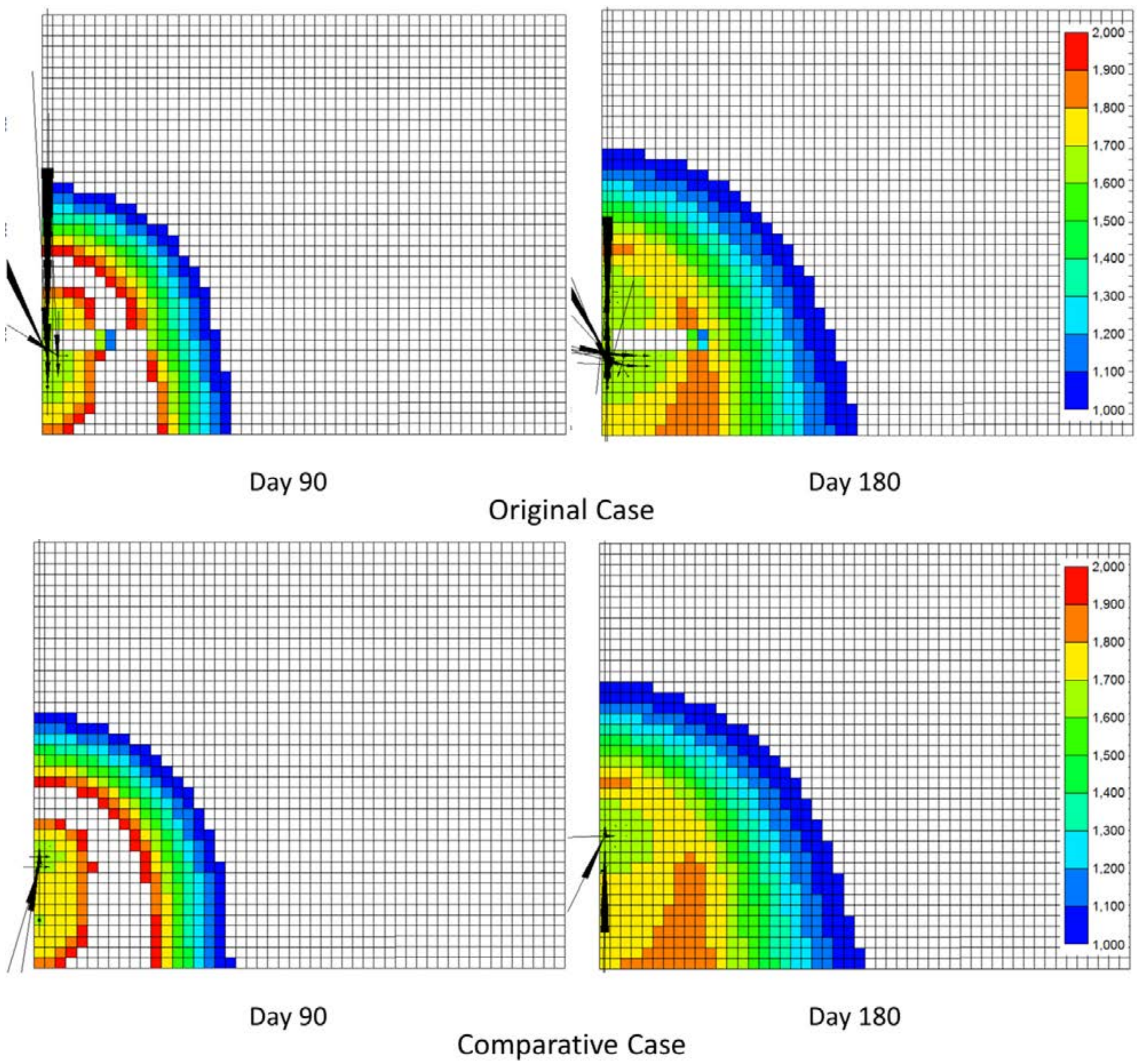


Fig. 6.11 – Pressure (kPa) distribution (cross-sectional view in the *x-z* plane) in segment 1. The arrows represent flow vectors of the gas phase.

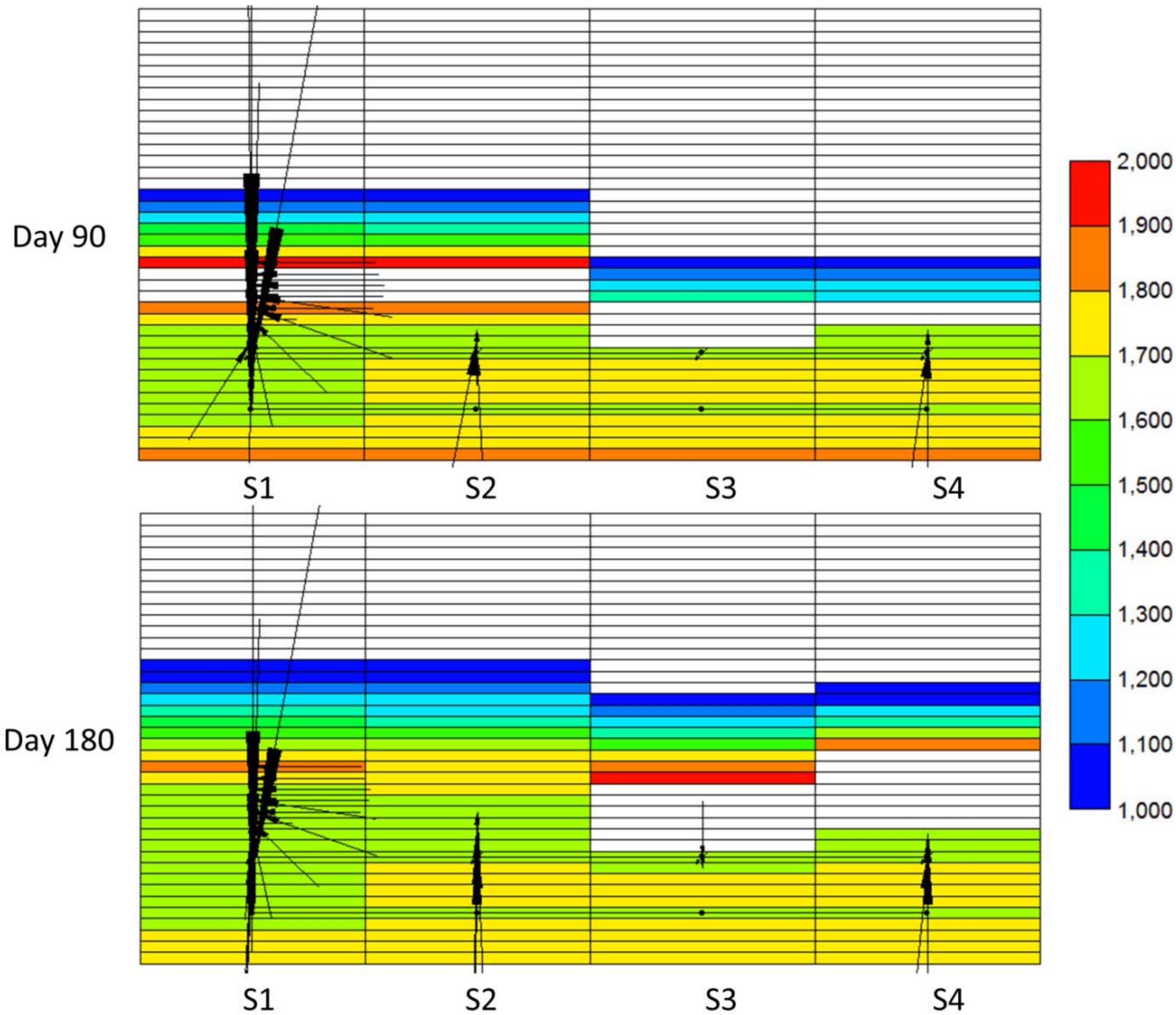


Fig. 6.12 – Pressure (kPa) distribution (cross-sectional view in the *y-z* plane) for the original case.

6.2 Forecasting SAGD Production from Reservoirs Containing Shale Barriers

Forward models are formulated to correlate SAGD production time-series with shale heterogeneities. The input parameters are elapsed production time in months, monthly steam injection rate and the 152 binary shale location indices, while the output parameter is the monthly oil production rate. An artificial neural network (ANN) model was previously adopted in the case of 2D shale barrier configurations in Chapter 4. ANN typically shows good capability in generalization and mapping of nonlinear relationships for large training datasets. It worked well in the 2D cases, where there were a total of 96000 records (corresponding to a training dataset of 800 synthetic cases) and 104 input parameters.

In the case of 3D configurations, the number of input parameters are increased to 154, while the number of records is 70200 (corresponding to a training dataset of 585 synthetic cases). However, due to the complex relationship between shale barrier configurations in 3D and the production time-series, it is proposed that different forward models should be calibrated for individual clusters to better capture the internal structures among the 3D cases. The drawback, though, is that the dataset for training each forward model is smaller (e.g. Cluster 3 only has 35 cases). To handle a smaller training dataset, another popular machine-learning technique, support vector machine regression (SVR), is also examined. The performances corresponding to ANN and SVR are compared.

Support vector machine was first introduced in 1992 by Vladimir Vapnik and his colleagues; it is a nonlinear generalized portrait method, where kernel functions are applied to create nonlinear classifiers (Vapnik and Lerner, 1963; Boser et al., 1992; Guyon et al., 1993; Cortes and Vapnik, 1995). A version of the SVM for regression problems was later introduced (Vapnik, 1995; Drucker et al., 1997). Many applications in petroleum-related problems can be found in the literature (Al-Aanzi and Gates, 2010; Liu and Horne 2012; Liu et al. 2015; Liu et al. 2016; Oloso et al. 2018).

The algorithm for support vector regression (SVR) is summarized here. For a set of N multivariate observations $\boldsymbol{x}$ (input variables) and y (output variable) that are linearly related, the goal is to find a set of coefficients $\boldsymbol{w}$ and a bias b, in **Eq. 6.1** , such that the sum of residuals between observations and predictions are less than an error ε (Smola and Schölkopf, 2004).

$$f(\boldsymbol{x}) = \boldsymbol{w}'\boldsymbol{x} + b \quad \text{..........(6.1)}$$

It can be formulated as a minimization problem with the following objective function:

$$\frac{1}{2}\|\boldsymbol{w}\|^2 + C_m \sum_{i=1}^{N} \left(\xi_i + \xi_i^* \right) \quad \text{..........(6.2)}$$

subjected to the following constraints:

$$\begin{aligned} & y - f(\boldsymbol{x}) \le \varepsilon + \xi_i \\ & f(\boldsymbol{x}) - y \le \varepsilon + \xi_i^* \\ & \xi_i \ , \ \xi_i^* \ge 0 \end{aligned} \quad \text{..........(6.3)}$$

where ξ_i, ξ_i^* are slack variables that allow the errors to be slightly larger than ε; and C_m is a soft margin parameter that controls the penalty imposed on the outliers (i.e., deviations > ε); it is needed to avoid overfitting by offering a trade-off between the training accuracy and model generalization capability (Crone et al., 2006; Vapnik, 1995).

The problem definition in **Eqs. 6.2-6.3** is also known as the primal problem; the minimization can be simplified by solving its Lagrange dual formulation (Vapnik, 1995; Smola and Schölkopf, 2004). The dual formulation introduces two non-negative Lagrange multipliers, α_i and α_i^*, and minimizes the following objective function:

$$\frac{1}{2}\sum_{i,j=1}^{N} \left(\alpha_i - \alpha_i^* \right)\left(\alpha_j - \alpha_j^* \right)\left(\boldsymbol{x}_i'\boldsymbol{x}_j \right) - \varepsilon \sum_{i=1}^{N} \left(\alpha_i + \alpha_i^* \right) - \sum_{i=1}^{N} y_i \left(\alpha_i - \alpha_i^* \right) \quad \text{..........(6.4)}$$

subjected to the following constraints:

$$\sum_{i=1}^{N}\left(\alpha_i-\alpha_i^*\right)=0 \qquad \text{...(6.5)}$$
$$0\le\alpha_i,\alpha_i^*\le C_m$$

In the case where $\boldsymbol{x}$ and y are nonlinearly related, kernel functions can be applied to map the data into a high-dimensional feature space. For example, a radial basis function (RBF) kernel can be employed.

$$k\left(\boldsymbol{x_i},\boldsymbol{x_j}\right)=\exp\left(\frac{-\left\|\boldsymbol{x_i}-\boldsymbol{x_j}\right\|^2}{2\sigma^2}\right) \qquad \text{...(6.6)}$$

Substituting the inner product in **Eq. 6.4** with **Eq. 6.6** gives the objective function for a nonlinear SVR model:

$$\frac{1}{2}\sum_{i,j=1}^{N}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)k\left(\boldsymbol{x_i},\boldsymbol{x_j}\right)-\varepsilon\sum_{i=1}^{N}\left(\alpha_i+\alpha_i^*\right)-\sum_{i=1}^{N}y_i\left(\alpha_i-\alpha_i^*\right) \qquad \text{...(6.7)}$$

The constraints are the same as **Eq. 6.5**. Once α_i and α_i^* are solved, the output of a new observation can be predicted by:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \left(a_i - a_i^*\right) k\left(\boldsymbol{x_i}, \boldsymbol{x}\right) + b \qquad (6.8)$$

To train a SVR model with a dataset of *N* observations, *N* pairs of α_i and α_i^* are tuned. Typically, only for some of the coefficients that their differences $(\alpha_i - \alpha_i^*)$ would be non-zero, and their corresponding input vector $\boldsymbol{x}_i$ are called support vectors. In other words, the number of parameters and support vectors is not determined a priori; it depends on the training dataset and could vary from 1 to *N*. On the other hand, the number of turning parameters in a multilayer perceptron neural network (MLP) model – a common ANN technique – is fixed for a given network architecture, depending on the number of input features, number of hidden layers, number of neurons, and the bias terms. The optimal architecture is generally determined via processes such as n-fold cross validation (Ma et al., 2015). The applicability of SVR and MLP can also be compared by examining the degrees of freedom in the adjustable parameters.

Table 6.4 – Average normalized mean squared error (NMSE) of the MLP and SVR models.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
# of cases	153	115	35	37	245
SVR NMSE	0.0352	0.2005	0.0389	0.7253	0.0507
MLP NMSE	0.2492	0.6991	0.6038	0.7321	0.1608

Both SVR and MLP are employed to train a forward model for each of the 5 clusters identified in **section 6.1**. For each cluster, 10-fold cross validation is applied for the training of both models and to optimize the MLP architectures. The number of 3D cases in each cluster and the prediction errors are compared in **Table 6.4**. The architectures and numbers of adjustable parameters for the final models are listed in **Table 6.5**. It is observed that SVR gives a superior performance with lower normalized mean squared error (NMSE). Therefore, the SVR forward models are selected and integrated in the next step for inferring an unknown shale barrier configuration corresponding to a new SAGD production history.

Table 6.5 – Numbers of parameters tuned in MLP and SVR models.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
MLP architecture (neurons in hidden layers)	13-13	11-11	9-9	9-9	19-19
# of parameters tuned (MLP)	2211	1849	1495	1495	3345
# of parameters tuned (SVR)	7305	3740	1427	550	12036

6.3 Estimating Shale Barrier Configurations from SAGD Production Histories

These forward regression models are then integrated in a GA scheme for inferring the unknown shale barrier configuration corresponding to a given set of SAGD production histories: steam injection and oil production time series are given, and 152 binary location indexes are inferred. This inverse problem consists of two steps: (1) cluster identification; (2) genetic algorithm (GA) based optimization.

6.3.1 Cluster Identification for the Target SAGD Production History

The target cluster is identified with the following procedure:

(1) Calculate the values of d between the actual production history and the other 585 training cases (prepared in **section 6.1**) with **Eq. 5.1**.

(2) Identify from the 585 training cases the most similar one with the smallest distance to the target case.

(3) Assign the target case to the same cluster as the most similar training case.

6.3.2 Genetic Algorithm (GA) Based Optimization

The goal is to identify a particular shale barrier configuration for which the oil production and steam injection profiles are consistent with the known production histories. Given that each shale barrier configuration is parameterized by 152 binary indices, they represent the unknowns in this optimization problem, where the objective function is defined as the mismatch between the predicted and known oil production time series. The fact that each shale barrier configuration is uniquely parameterized by a set of binary location indices renders the GA an ideal approach, as compared to other gradient-based minimization algorithms. To apply GA, a set of binary strings (chromosomes) is utilized as genetic representation of the solution domain (Whitley, 1994). Details of the general GA method can be found in the reference.

A similar GA-based optimization scheme was successfully adopted in Chapter 4 for inferring shale barrier configurations in 2D SAGD cases. The workflow is implemented using the Python DEAP library (Fortin et al., 2012) here for the 3D cases.

(1) An initial population is generated by selecting a set of most similar cases from the target cluster.

(2) The shale barrier configurations of the initial solutions are encoded into binary strings, and the values of their fitness functions are computed. The fitness function is defined as $C=1/d$, where d is the flow-based distance (**Eq. 5.1**) between the actual production history and a particular initial case.

(3) Two cases (parents) are randomly selected from the initial population (the selection probability is inversely related to the fitness value). Two new cases (children) are generated by swapping parts of the parents' binary strings and mutating a portion of the binary values ($0\rightarrow1$ or $1\rightarrow0$).

(4) Step 3 is repeated until a new population is created.

(5) The oil production profiles for the new population are estimated using the forward SVR model trained for the specific target cluster.

(6)Steps 2-5 are iterated multiple times until certain user-defined stop criterion is encountered (e.g. maximum iterations, highest fitness above a threshold), and the individual case with the highest fitness is identified as the shale barrier configuration that best matches the production history.

A possible drawback of the aforementioned scheme is that, as GA randomly perturbs the shale barrier configurations, there is no guarantee that the resultant shale barrier configurations would actually belong to the same target cluster. If we blindly subject these configurations to the forward SRV model corresponding to the target cluster, the predicted oil production profiles would be incorrect. In addition, since the steam injection profile is imposed as an input, there is also a distinct probability that the steam injection profile is physically incompatible with a shale barrier configuration that should have been classified into a different cluster. It was observed in a

Chapter 4 that it is possible to achieve a match in the oil production profile by imposing a physically incompatible steam injection input. In such instances, the inferred shale barrier configurations would exhibit strikingly different characteristics from the "true" model. Therefore, it is necessary to screen each proposed shale barrier configuration and ensures that its membership to the same target cluster. In particular, the P5 and P95 values of the four near-well shale barriers (**section 6.1**) can be used as simple screening criteria.

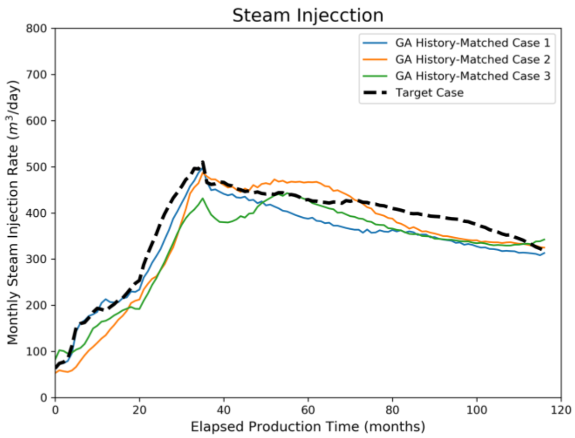
(1) At the end of the GA workflow, each inferred shale barrier configuration is screened for the presence of shale barriers at the four near-well locations. The numbers of near-well shale barriers in all four segments are compared to the P5 and P95 values of the target cluster, as listed in **Table 6.3**. For example, assuming the GA proposed shale barrier configuration has N_{locj} shale barriers at Shale Location j (j = 1, 2, 3, 4), if N_{locj} is bounded by the P5 and P95 values of the target cluster, the proposed shale barrier configuration is retained; otherwise, it is eliminated, and the GA scheme (**section 6.3.2**) is repeated. It is important to note that, in this study, only ranges of N_{locj} are used to screen the GA proposed configurations; however, if additional conditioning information is available (e.g., petrophysical logs along the wellbore), it can be used further constrain the specific positions of shale barriers in the near-well region among those GA proposed configurations.
(2) The overall shale proportion for all cases is set to be less than 25% (same as the training dataset).

It is worth mentioning that estimating an unknown shale barrier configuration from a given SAGD production is an ill-posed inverse problem. It is expected that multiple shale barrier configurations would yield similar production behavior (Zheng et al., 2018b). Therefore, this GA workflow should be repeated numerous times to generate a suite of plausible shale barrier configurations for a specific production history.

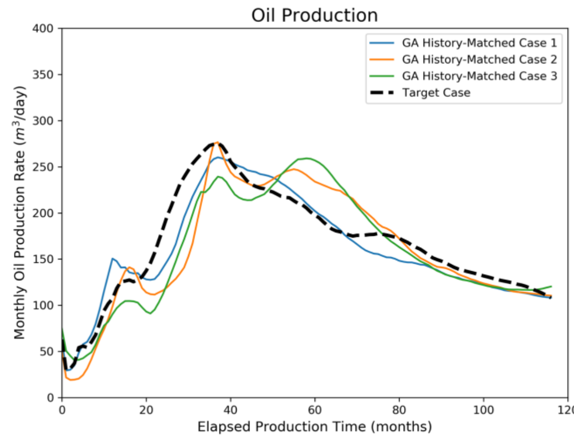
A testing dataset of 30 3D cases is employed to assess the performance of the forward SVR models, as well as the inverse GA workflow. These 30 cases are constructed by randomly assigning shale barriers in all four segments and ensuring that all five clusters are included. They are not part of the original 1200 3D cases, and the total shale volume proportion is less than 25%. Three different shale barrier configurations are inferred for each of the 30 testing cases following the prescribed procedure. To validate the model predictions, each of these shale barrier configurations are also subjected to flow simulation to obtain the correct oil production and steam injection profiles.

6.4 Testing Results and Discussions

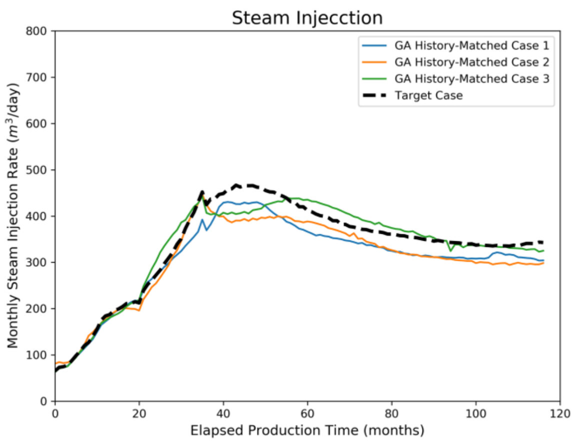
For the sake of brevity, only two shale barrier configurations from each cluster (i.e., a total of 2 × 5 = 10 cases, out of 30) in the testing dataset are presented in **Fig. 6.13.** The flow simulation predictions of both steam injection and oil production profiles are compared. To compute a measure of dissimilarity between target case and 3 proposed history-matched cases, all four profiles are normalized to a scale of 0 and 1, and an average *C* value is calculated (**Eq. 5.1**) and presented in **Table 6.6.** The shale barrier features of the known case and the GA history-matched cases are also compared in **Table 6.6**.



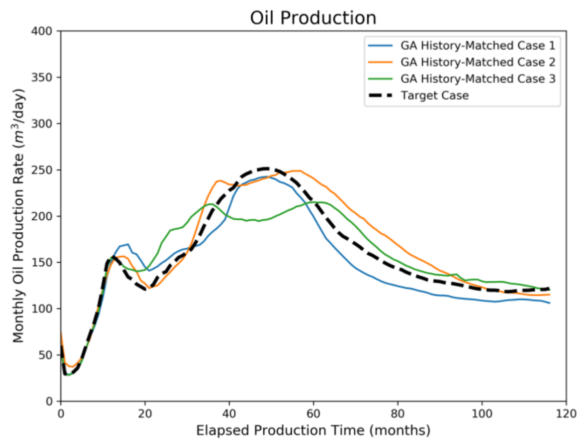
Case-1 (a)



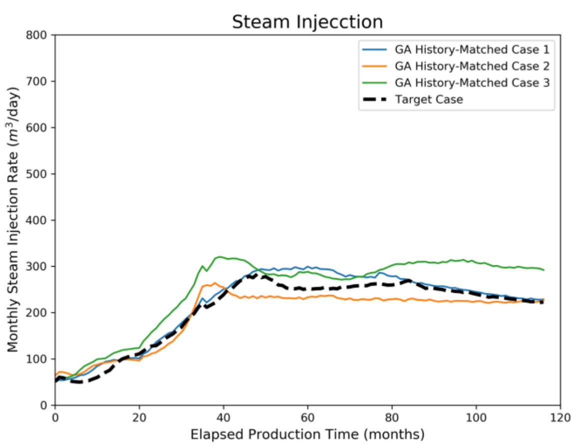
Case-1 (b)



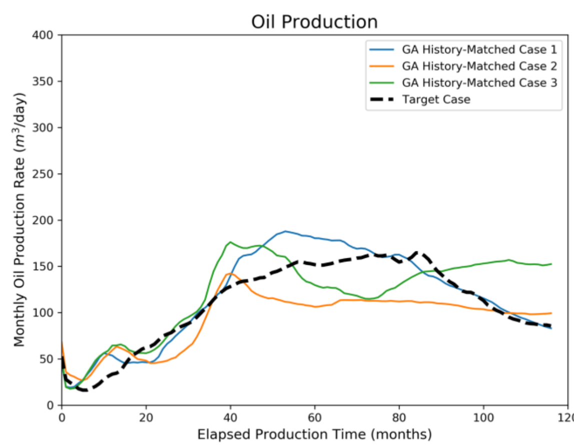
Case-2 (a)



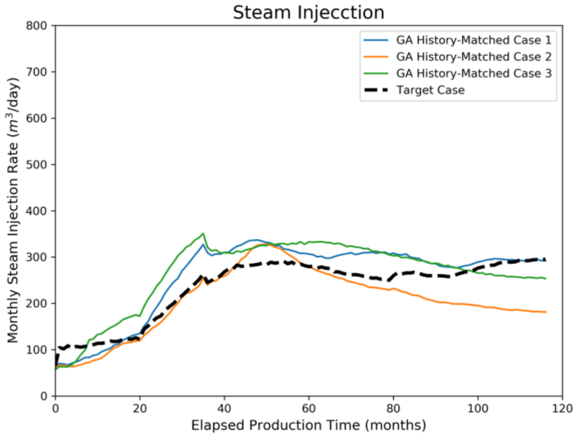
Case-2 (b)



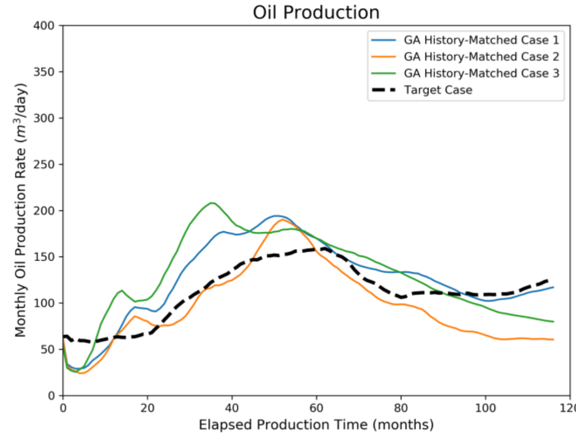
Case-3 (a)



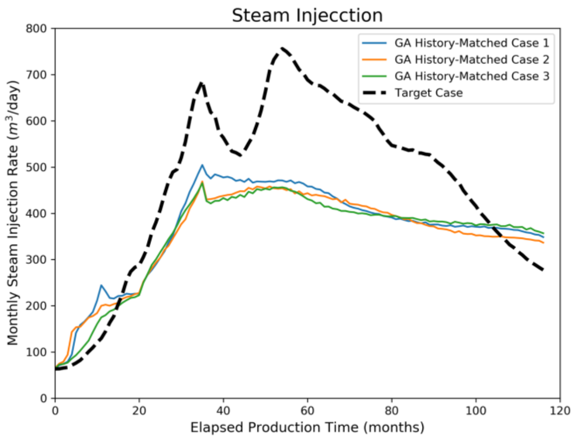
Case-3 (b)



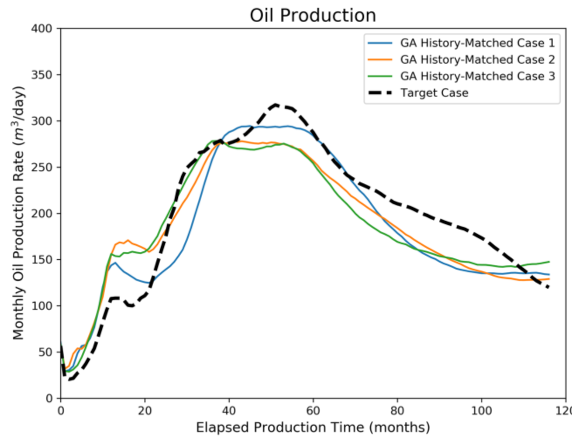
Case-4 (a)



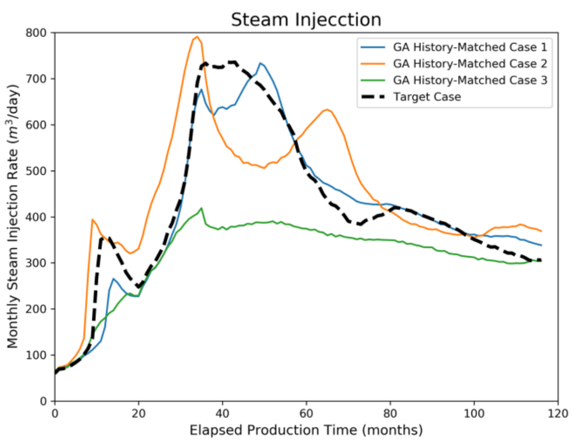
Case-4 (b)



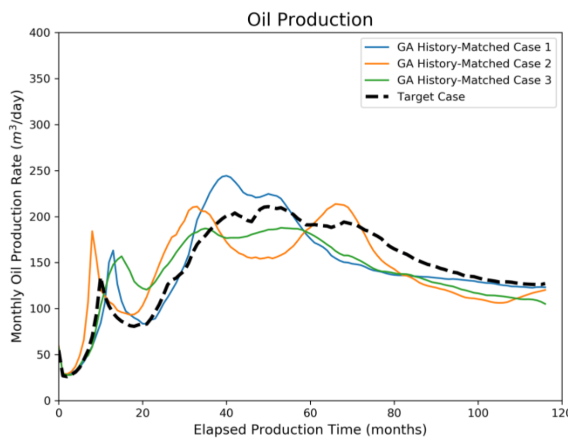
Case-5 (a)



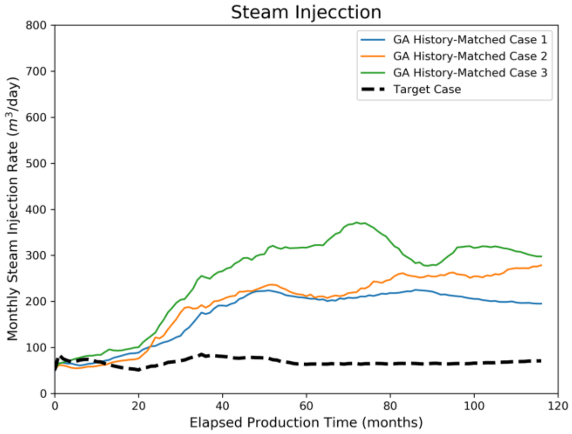
Case-5 (b)



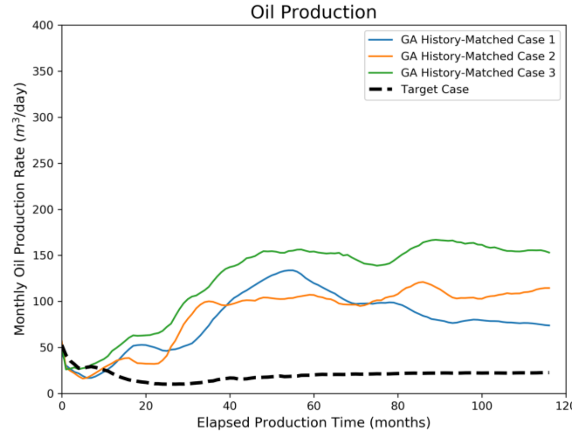
Case-6 (a)



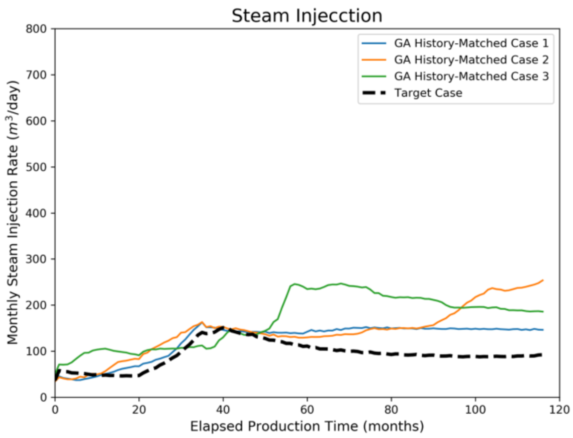
Case-6 (b)



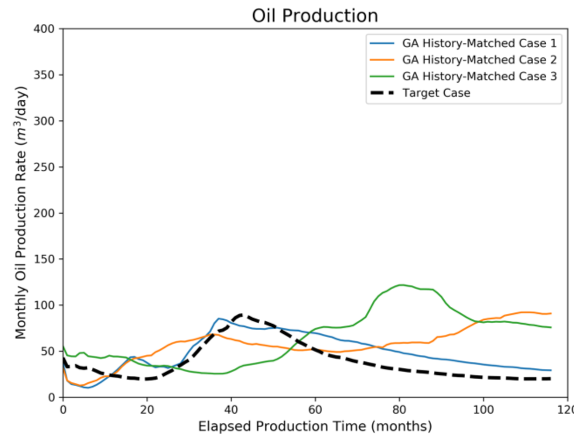
Case-7 (a)



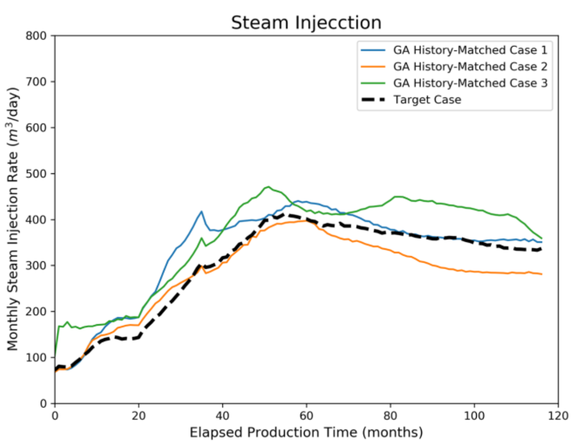
Case-7 (b)



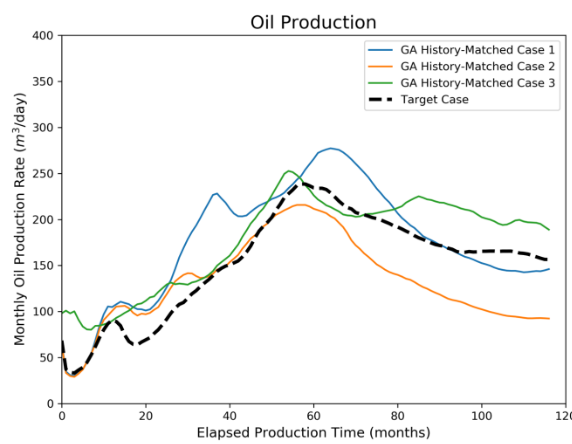
Case-8 (a)



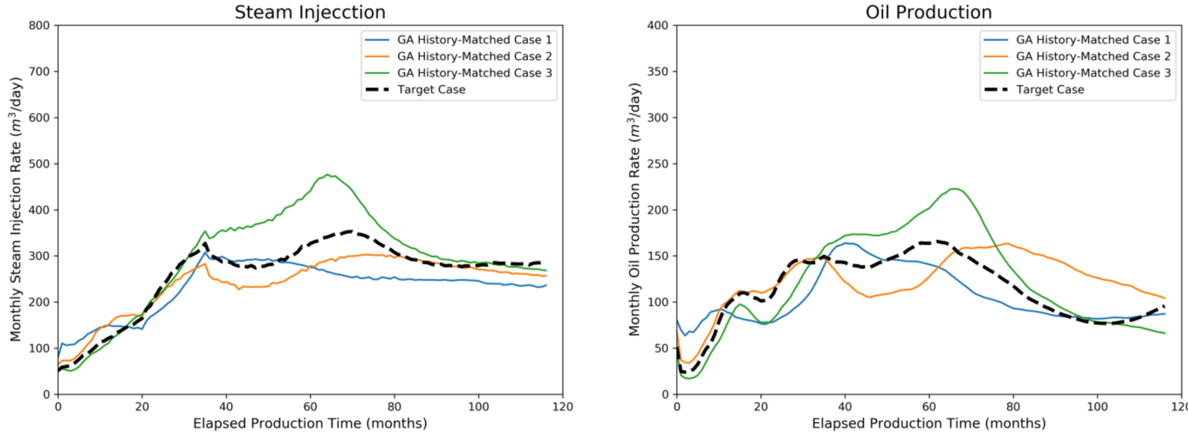
Case-8 (b)



Case-9 (a)



Case-9 (b)



Case-10 (a) Case-10 (b)

Fig. 6.13 – Results of 10 (out of 30) cases in the testing dataset: (a) steam injection profiles from numerical simulation for the known case and GA history-matched cases; (b) oil production profiles from numerical simulation for the known case and GA history-matched cases.

Table 6.6 – Comparison of shale barrier configuration features between the target case and the GA inferred cases.

Test Case	Cluster	Near Well Shales (Target)				Near Well Shales (GA Estimation)(mean of 3 cases)				Dissimilarity
		Loc1	Loc2	Loc3	Loc4	Loc1	Loc2	Loc3	Loc4	Target vs GA Estimation (mean of 3 cases)
1	1	0	0	1	1	0.3	0	1	1.3	0.0077
2	1	0	0	1	0	0	0	0.7	0.3	0.0048
3	2	1	1	0	1	1.3	0.3	1	1.3	0.0223
4	2	1	1	1	0	0.3	1	1.3	0	0.0142
5	3	1	0	1	1	0	0	1	0.7	0.0243
6	3	1	0	2	0	1	0	1	0	0.0158
7	4	2	1	1	0	0.7	0.7	1.3	1.3	0.0633
8	4	1	2	0	0	1	1.7	1	0.7	0.0275
9	5	1	0	1	1	1	0	0.3	1	0.0097
10	5	0	1	1	0	0.3	0.7	0.3	0.7	0.0100

A visual inspection of **Fig. 6.13** reveals that, for most testing cases, the simulated production profiles of the GA history-matched (inferred) cases are consistent with the target (actual) production history, where only minor dissimilarities (average $C < 0.03$) are observed in **Table 6.6**. Interestingly, cases that belong to Clusters 1 and 5 seem to exhibit the least dissimilarities, while cases that belong to Cluster 4 would yield the largest dissimilarities. To some extent, this difference in prediction quality may be related to the forward model errors (**Table 6.4**), where the forward models for Cluster 4 have the largest training error. It is anticipated that improving the predictive capability of the forward model for Cluster 4 may enhance the accuracy of the entire workflow.

It is also observed that the GA history-matched cases share many key features of the target case, especially with respect to the distributions of near-well shale barriers. Deviations in the total shale proportion are noted for some cases (e.g. case 4, 6); however, these discrepancies do not necessarily lead to poor outcomes (or large C values). This is because the GA history-matched cases contain more shale barriers in areas that are far away from the well pair, which do not instigate any observable impacts on the production behavior.

6.5 Summary

In this chapter, a three-part workflow is proposed for inferring 3D shale barrier configurations from SAGD production histories. First, identifying internal clusters among the training dataset based on MDS and K-means analysis. Second, constructing forward regression models for individual clusters to correlate SAGD production profiles with heterogeneous shale barrier configurations. Finally, implementing a GA framework to infer unknown shale barrier configuration from a known (target) production history. The cluster feature analysis reveals that shale barriers at the near-well region would have the most observable impact on the SAGD production. The workflow is capable of capturing most salient features of the production profiles and near-well shale barrier distributions in a reliable and efficient manner.

CHAPTER 7: FIELD CASE STUDY

Chapter 6 has demonstrated the capability and flexibility of the proposed AI-based workflow in inferring the presence and distribution of heterogeneous shale barriers for 3D synthetic SAGD cases. In this chapter, the utility of the proposed approach for field application is tested; in particular, a case study on the Suncor's Firebag Project is presented. Given that the training dataset is constructed based on representative petrophysical properties and operational constraints from Firebag, the previously established cluster features and forward models in Chapter 6 are retained. A few well pairs from Pads 104 and 107 are selected, and their production histories, as well as gamma-ray (GR) logs, are extracted.

7.1 Field Production Histories

With access to limited public data, two major challenges are encountered: production/steam allocation and steam leakage (injection efficiency). Regarding the first issue, the model dimension is only 80 m along the *y*-direction for all the synthetic cases, while the average horizontal well length is approximately 1000 m in the field. Therefore, a 80-m segment from the well pair should be selected. However, due to reservoir heterogeneities, not all 80-m segments of a given well pair in the field would contribute equally to the total production. Therefore, it is necessary to devise a scheme to properly allocate an estimated annual contribution.

The second issue is that the field steam-oil ratio (SOR) is consistently higher than the predictions by all the synthetic cases. As an example, the oil production profile of the homogeneous base case is compared against that of a producing well pair extracted from Pad 104 in **Fig. 7.1**. The oil production profiles of the homogeneous base case would follow similar trends as the field data; however, steam injection rates in the field are much higher than the homogeneous case. Similar observations can be made for other well pairs in Pads 104 and 107. It is hypothesized that higher SOR in the field might be attributed to other factors, such as leaking, which are not considered in the synthetic models.

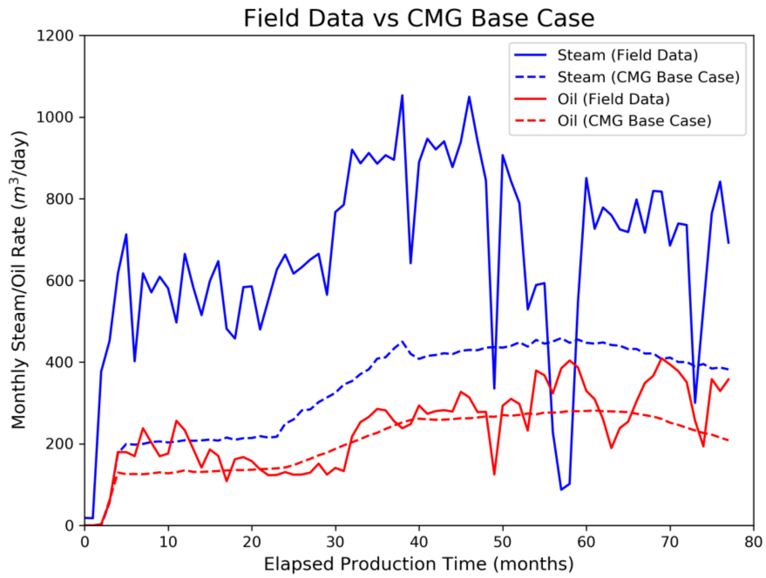


Fig. 7.1 – Comparison of production profiles between numerical simulation (CMG) homogeneous base case and a randomly selected well pair in Firebag Pad 104.

Given that the entire workflow, including the forward models, is trained based on synthetic simulation cases, the following procedures are proposed to ensure the applicability of the developed workflow for analyzing the field data:

(1) **[Steam Chamber Analysis]** It is assumed that production/injection over a particular horizontal segment can be allocated in accordance to the nearby steam chamber thickness.

Steam chamber thickness map, which is inferred from 4D seismic data, of Pad 104 is extracted from Suncor's Firebag annual AER performance presentation (Suncor Energy, 2015). The yearly development of the steam chamber in Pad 104 from 2010 to 2013 is present with the same color scale. To be consistent with the annual reporting requirements, it is assumed that the corresponding steam chamber map reflects the change within the same reporting period (e.g. from March in previous year to Feb next year). In Pad 104, the average well length is about 1000 m along the horizontal portion, and the average well spacing is approximately 155 m (i.e., infill wells are not considered). Each steam chamber image can be digitized into a 258 × 40 pixel matrix; this resolution means that the producing area for Pad 104 is divided into 258 portions along the 1000 m horizontal well length and 40 portions in between two well pairs; in other words, each pixel is representing an area of approximately 4 m × 4 m. The RGB color code for a given pixel is translated into an equivalent steam chamber thickness (in meters) by linearly scaling from the RGB scale of [0, 255] to thickness [0, 48 m]. The maximum thickness of 48 m is obtained from the AER reports. The total steam chamber thickness at each of the 258 positions along the well is obtained by summing all 40 thickness values in the direction perpendicular to the wellbore.

(2) **[Petrophysical Log Analysis and Segment Selection]** The gamma ray (GR) logs are used to infer the presence of shale near the wellbore. The GR logs and the steam chamber thickness maps are analyzed collectively to identify several 80-m intervals along the horizontal portion that exhibit different shale characteristics. A cut-off of 60 is applied on the GR logs (as per the Suncor's Firebag annual AER reports). The shale characteristics inferred from the GR

logs will also be used to check the final configurations proposed by the GA workflow for consistency.

(3) **[Production/Injection Allocation]** Production/injection over a particular 80-m segment is allocated in accordance to the total steam chamber thickness computed in (1). As mentioned earlier, the field steam-oil ratio (SOR) is consistently higher than the predictions by all the synthetic cases; therefore, to establish some level of consistency between the synthetic cases and the field data, a calibration scheme is proposed. The actual field history is scaled to one of the type curves in **Fig. 6.5**. The choice of a type curve will be explained later, but, first, the scaling procedure is discussed here. There is a total of 4 years of production history from 2009 to 2013. For this production period, two allocation ratios for oil production (S_o) and steam injection (S_s) are calculated as:

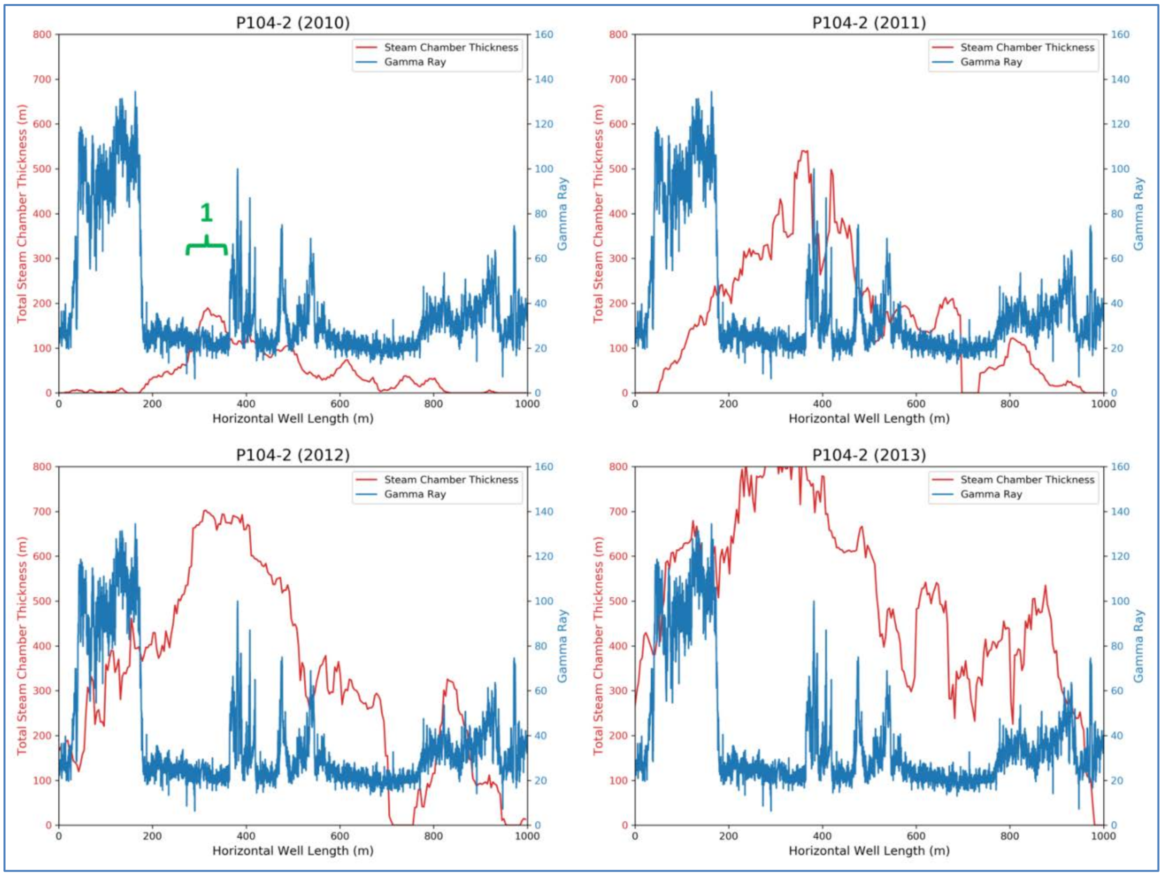
$$S_o = \frac{Q_{of}}{Q_{ot}}$$
$$S_s = \frac{Q_{sf}}{Q_{st}} \quad \text{(7.1)}$$

where Q_o and Q_s denote the cumulative oil production and steam injection, respectively. The subscripts *f* and *t* refer to the field history and the selected type curve, respectively. The field data of production time series are subsequently scaled by the raw values by S_o and S_s, such that the shapes of the oil and steam profiles remain unchanged, while some level of consistency is achieved between the synthetic cases and the field data.

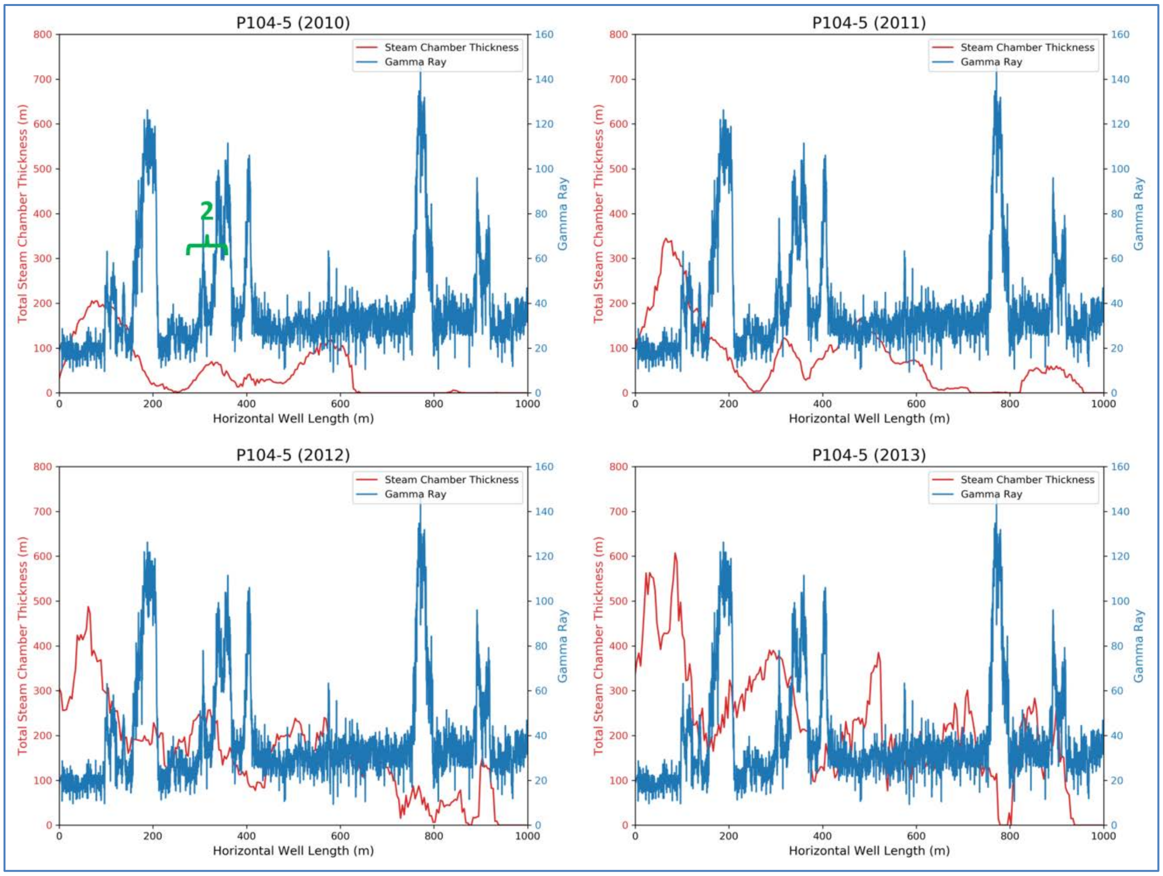
The choice of an appropriate type curve is discussed next. Two particular aspects can be considered: the overall production performance of the field history and the GR logs. The selection could be subjective, but the principles applied in this study are illustrated here:

- If the production data indicates a uniform and steady steam chamber development, while relatively low shale content is interpreted from the GR log, it seems reasonable to select from Cluster 1 or the homogeneous case for the reference type curve.
- If there is some indication of uneven steam chamber development and high shale content is interpreted from the GR logs over a portion of the 80-m interval, the type curve from Cluster 5 may be an appropriate choice.
- If the production data indicates the steam chamber development has been highly hampered and high shale content can be inferred from the GR logs in most parts of the 80-m interval, a type curve from either Clusters 2 or 4 seem to be a reasonable choice.

Two specific well pairs in Pad 104 (P104-2 and P104-5) are analyzed. The steam chamber thickness within the vicinity of each well pair, as well as the GR logs, between 2010 and 2013 are shown in **Fig. 7.2**. It is clear that the steam chamber development is more extensive in P104-2 than P104-5, and this is corroborated by the higher production rate in P104-2.


P104-2 (2010)
P104-2 (2011)
P104-2 (2012)
P104-2 (2013)
Steam Chamber Thickness
Gamma Ray
Total Steam Chamber Thickness (m)
Gamma Ray
Horizontal Well Length (m)
1

(a)

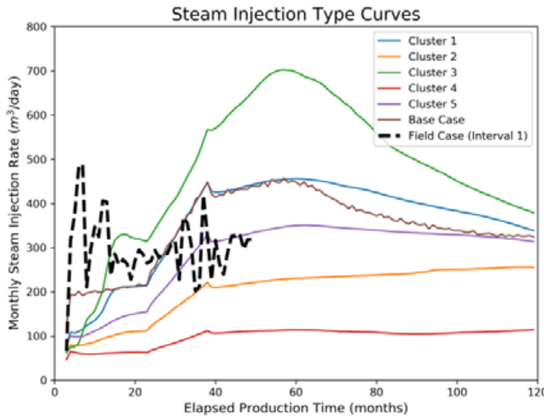


(b)

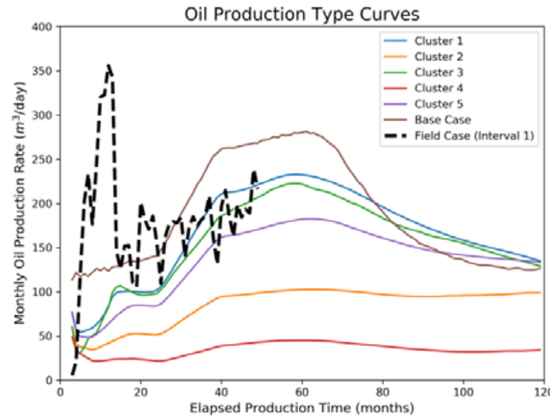
Fig. 7.2 – Gamma ray and total steam chamber thickness along the horizontal well trajectory: (a) P104-2; (b) P104-5.

Two particular intervals exhibiting distinct shale barrier configuration characteristics are selected from the well pairs. According to the steam chamber thickness and GR analysis, Interval 1 has little to no shale is observable around the well pair, while shale is present at some positions along Interval 2. These selected intervals are also marked in **Fig. 7.2**. Since well trajectory is not clearly defined in the field data, it is possible that the well goes through different layers which may lead to small errors in estimating the length of shale barriers. For the allocation purposes, the production/injection profiles from Interval 1 are scaled to the type curve corresponding to the 3D homogeneous base case (**Fig. 6.5**). On the other hand, the production/injection profiles from

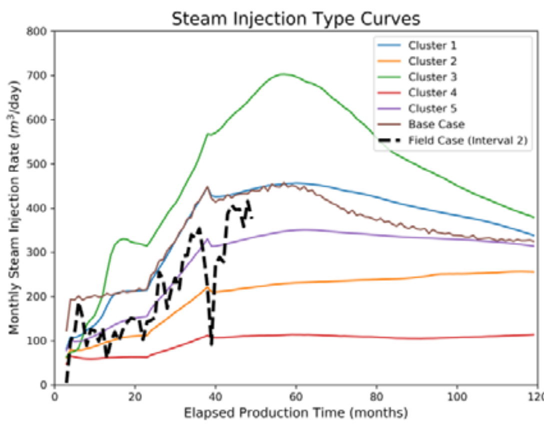
Interval 2 is scaled to the type curve corresponding to Cluster 5, where some shale barriers are present, as corroborated by the GR measurements and the relatively low oil rate of P104-5.



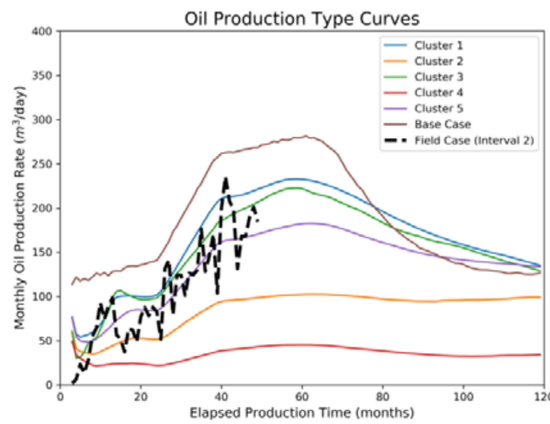
(a) Steam Injection (Interval 1)



(b) Oil Production (Interval 1)



(a) Steam Injection (Interval 2)



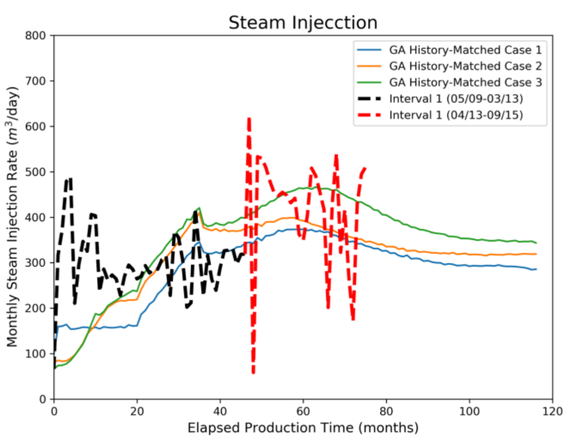
(b) Oil Production (Interval 2)

Fig. 7.3 – Comparison of the scaled and allocated field history and type curves derived from the 3D synthetic models.

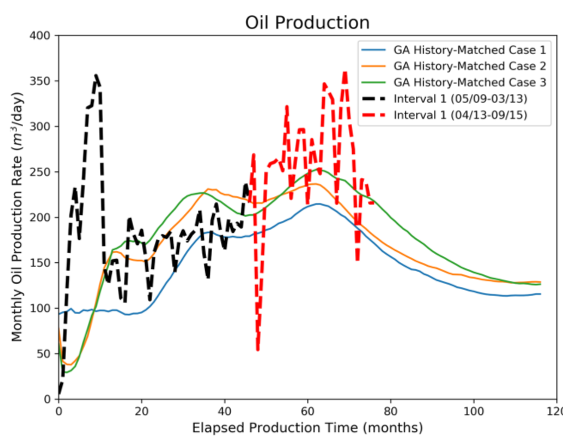
The allocated and scaled production profiles for these two intervals are compared with 5 cluster type curves in **Fig. 7.3** and they are considered as the input field history in the GA workflow. Data after 2013 is excluded in this study for two reasons: (1) the published steam chamber maps of Pad 104 after 2013 are displayed over a different color scale, and (2) infill wells are added in 2014 and interference between well pairs could be significant.

7.2 Inferring Shale Barrier Configurations from Field Histories

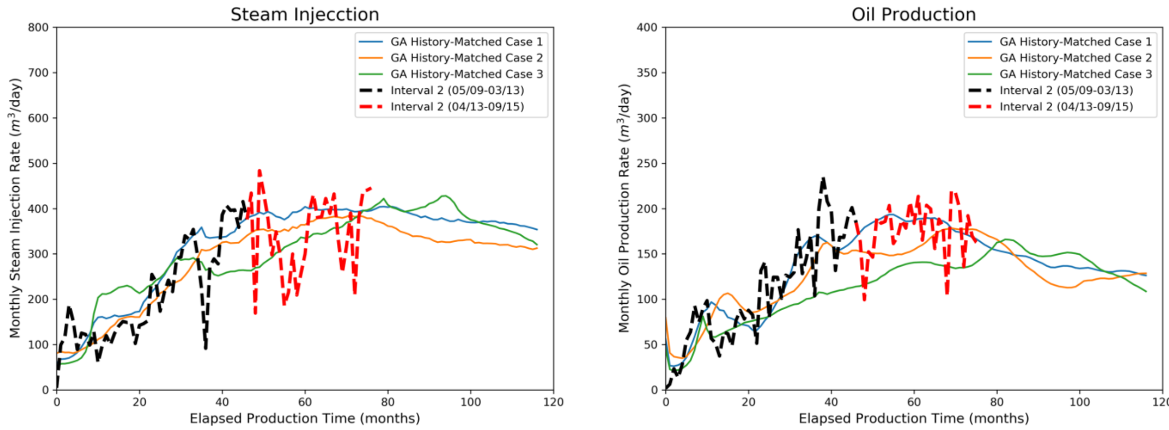
The GA workflow described in **section 6.3** is applied to infer several realizations of shale barrier configurations. A total of 3 GA history-matched shale barrier configurations are inferred for the two intervals identified in **section 7.1**. **Fig. 7.4** compares the steam injection and oil production field history and the numerical simulation predictions of the GA inferred shale barrier configurations. To assess the accuracy of the inferred shale barrier configurations, portion of the field histories from April 2013 to September 2015 have not been presented to the GA workflow but used for validation purposes instead. The results are compared in **Fig. 7.4**. The steam injection and oil production allocation ratios (**Eq. 7.1**) for the selected interval in this period are assumed to be the same as the last available steam chamber map (2013). The GA history-matched shale barrier configurations with the least mismatch with the histories are shown in **Fig. 7.5**.



Interval 1 (a)



Interval 1 (b)

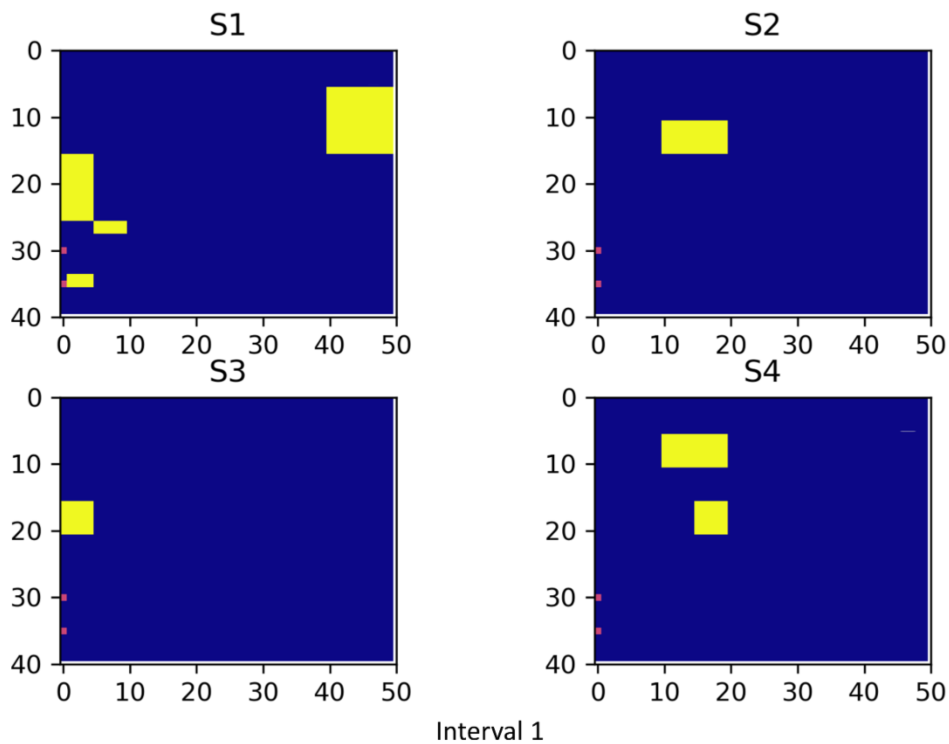


Interval 2 (a) Interval 2 (b)

Fig. 7.4 – Results of field cases: (a) steam injection profiles from numerical simulation for the GA history-matched cases and the field histories; (b) oil production profiles from numerical simulation for the GA history-matched cases and the field histories. The red portion of the field history is used for validation purposes.

For Interval 1, the GA history-matched cases generally follow the trends observed in the field history. Spikes in both steam injection and oil production profiles during the early stage of production (e.g., around month 10) are not predicted in any of the history-matched cases. In fact, none of the synthetic cases would simulate these spikes. Therefore, it is concluded that these fluctuations may be the result of various unknown deviations in the operating constraints. Very few shale barriers are inferred around the wellbore, corroborating with the steam chamber thickness and GR analysis.

For Interval 2, the GA history-matched cases are consistent with essentially the entire field history. More shale barriers are inferred for this interval, which is also consistent with the steam chamber thickness and GR analysis. Slower development of the steam chamber is observed, as compared to Interval 1 in P104-2. The GR measurements would support the inference of some shale barriers near the toe portion (segment 4) in the 80-m interval. Interestingly, the configuration in **Fig. 7.5** clearly shows a single shale barrier that is located adjacent to the injector in segment 4. For both intervals, multiple shale barriers are inferred at distances of 4-5 m laterally away from the injector in all segments; unfortunately, the presence/absence of these heterogeneities cannot be confirmed directly from the GR logs due to their limitation in the depth of investigation. However, the production behavior of these inferred shale barrier configurations are highly consistent with the field histories, as evidenced in **Fig. 7.4**.



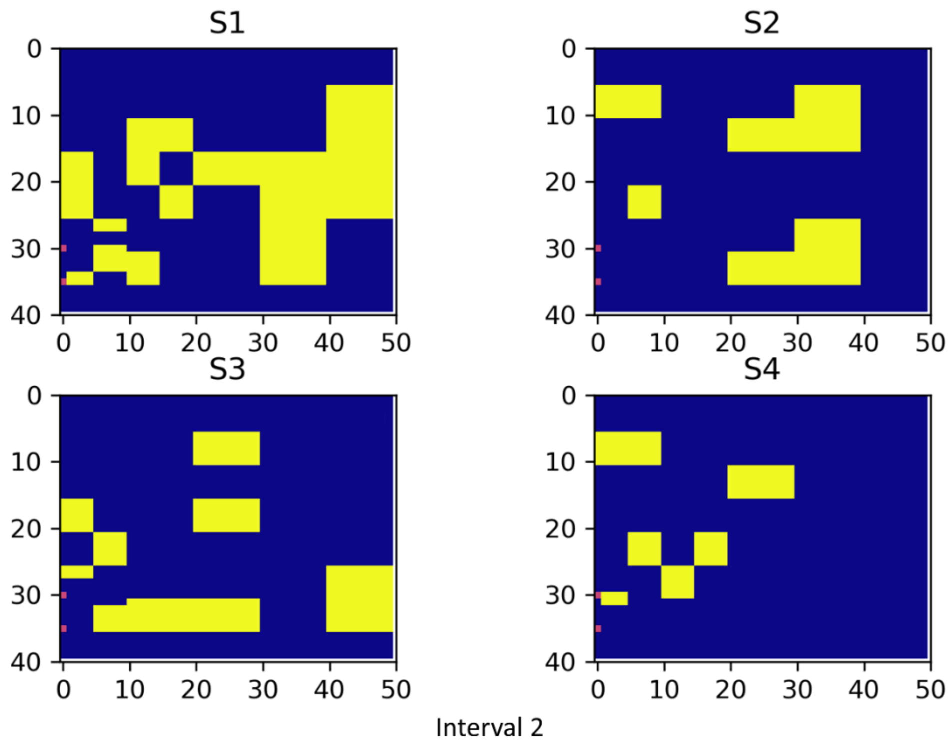


Fig. 7.5 – Examples of GA-history matched shale barrier configurations for field cases.

7.3 Summary

In this chapter, a set of field production profiles are assembled from Suncor's Firebag project. A method for allocating field production based on steam chamber volumes is introduced. The proposed workflow for inferring shale barrier configurations from SAGD production history is tested with the filed dataset. Overall, results from field cases demonstrate that the proposed AI-based workflow is effective in capturing the influences of shale barriers and generating reliable estimates of the recovery performance, despite the lack of detailed information from the field. The estimated shale barrier configurations are consistent with production histories, petrophysical log measurements and steam chamber thickness maps inferred from seismic data. Addressing the non-uniqueness in the solution of an ill-posed inverse problem, multiple shale barrier configurations can be obtained to represent the uncertainties in model parameters.

CHAPTER 8: CONCLUDING REMARKS

8.1 Summary and Conclusions

This research aims to explore the influences of shale heterogeneities on SAGD production profiles and develop a practical workflow for inferring locations of shale barriers from SAGD time-series data. A set of SAGD field data for Suncor's Firebag project is compiled from the public domain. The dataset consists of production profiles, well logs, core measurements, reservoir pressures, temperature measurements and operational constraints.

A synthetic dataset consisting of 2D and 3D cases is constructed by superimposing idealized shale barriers on a homogeneous base model, whose properties are assigned in accordance to representative petrophysical and operational parameters gathered from the Firebag project; production trends from the base case are in reasonable agreement with the actual field observations.

A novel parameterization scheme is introduced to represent the different shale barrier configurations. This is facilitated by dividing the entire reservoir domain into a number of sub-domains and formulating a set of basic shale units and binary indices. Sensitivity analysis of the basic shale units (size, geometry, position) is conducted. Synthetic datasets are constructed; binary indices of shale heterogeneity and monthly steam injection rate are considered as input variables, while the monthly oil production rate is considered as an output variable.

An AI-based workflow is proposed to facilitate the SAGD production forecast and infer the presence and distribution of heterogeneous shale barriers from production profiles (time-series data). The workflow consists of two parts. First, a forward regression model (ANN) is constructed to correlate SAGD production profiles with heterogeneous shale barrier configuration. Next, genetic algorithm (GA) is applied for inferring shale barrier configuration by matching the target production time-series data. The proposed methodology is first tested with 2D synthetic cases, and the performance of the workflow is shown to be both reliable and efficient in capturing most salient features of the production profiles and shale barrier configurations.

In order to address the increasing complexity of 3D problems, multidimensional scaling (MDS) and cluster analysis are also proposed to visualize the (dis)similarities among realizations of shale barrier configurations and to quantify their influences on SAGD production. A flow-based distance function is formulated, and MDS is applied to project the dissimilarity matrix to an n-dimensional Euclidean space, where redundancy in the original data set can be identified and cases that are too close to one another can be discarded. The reduced data set is subjected to MDS again, and K-means clustering is conducted to identify internal groupings among the data set. The workflow identifies different internal structures in 2D and 3D datasets with distinct cluster features, and new cases are successfully added to the under-sampled clusters.

With the cluster analysis methodology, the proposed workflow for inferring shale barrier configurations is optimized. The revised workflow consists of 3 steps. First, identify internal clusters among the training dataset by applying MDS and K-means analysis. Second, construct forward models (SVR) for each cluster to correlate SAGD production profiles with heterogeneous shale barrier configuration. Next, apply GA framework to estimate shale barrier configuration by matching the target production profile. A set of 3D cases are tested to examine the utility of the proposed approach for more realistic applications. The testing results are consistent with the target profiles.

Case study is conducted on two field examples assembled from the Firebag field data. In the field examples, a method for allocating field production based on steam chamber volumes is introduced. The workflow established with synthetic dataset is successfully adopted for analyzing a few field cases. Results of the field examples are highly encouraging, considering that only data from the public domain was accessed. Domain knowledge and additional data from the field operator could help to eliminate certain assumptions and reduce uncertainties in various reservoir and operational parameters.

The results of this work demonstrate the capability and flexibility of the AI-based network model, and of the parametrization technique for representing the characteristics of the shale barriers, in capturing the effects of complex heterogeneities on SAGD production. This work highlights the potential of an AI-based workflow to infer the presence and distribution of heterogeneous shale barriers from field SAGD production time-series data. The research does not aim to replace conventional reservoir characterization routines. The workflows presented in this thesis intend to provide an efficient and complementary method for exploring heterogeneous characteristics from SAGD field data, and the output shale configurations can be regarded as an initial estimate of input production profiles and subjected to further history matching for a more accurate final match.

8.2 Contributions

The primary contributions are summarized as follows:

(1) Application of data-driven modelling techniques for SAGD production analysis is promising, but not yet widely implemented in practical SAGD projects. A challenge for customizing

data-driven models for field application is to interpret reservoir mechanics from field data and integrate domain knowledge regarding the physical systems in the analysis. This research proposes a hybridization of physics-based and data-driven modelling approaches. Physic-based rules are integrated into the modelling process, while domain knowledge is applied to better understand the dataset and optimize the outputs.

(2) In order to apply AI-based data-driven techniques for examining the influence of reservoir heterogeneity on SAGD production, one particular challenge pertains to the proper parameterization of reservoir heterogeneity for model training. This research formulates a novel parameterization scheme to characterize shale heterogeneity. The scheme allows any given shale barrier configuration to be represented by a set of binary variables. This parameterization scheme can be readily extended to describe other reservoir features (e.g., lean zones).

(3) Another common challenge in data-driven model development is assembling a representative training data set. This research proposes a new workflow to visualize dissimilarities among realizations of shale barrier configurations and to quantify their influences on SAGD production. It illustrates a systematic framework for optimizing the spanning of model parameter space in a given data set, without exhaustively sampling similar realizations. It offers several potential applications in SAGD reservoir analysis; for example, the cluster features in terms of production characteristics and shale barrier configurations can be used to efficiently screen/rank realizations during history matching. The proposed technique can also be combined with other experimental design approaches to construct representative training data sets for proxy modeling. This step is particularly important if 3D cases are involved.

(4) Numerous machine learning techniques (ANN, SVR) are employed successfully to facilitate a SAGD production forecast. The performances of the ensuing models are shown to be both reliable and efficient in capturing many salient features in the production trends. It is worthwhile to mention that, once the data-driven model is trained, the computation time for the forward prediction takes only a few seconds. The developed model can serve as efficient proxy for modeling and forecasting of SAGD production time series, alleviating some aspects of the high computational and capital costs associated with repeating many reservoir simulation runs required in most typical characterization workflows.

(5) A novel AI-based workflow is proposed to infer the presence and distribution of heterogeneous shale barriers from SAGD production profiles (time-series data). The results demonstrate the feasibility of the AI-based workflow in correlating complex reservoir heterogeneities with SAGD production profiles. The inferred shale configuration can be regarded as an initial estimate and subjected to further history matching for a more accurate final match.

(6) The workflow also provides a basis for developing operating strategies to reduce the impact of shale barriers. For example, in most of the well pads in Firebag, infill wells are drilled a couple years after the original wells have commenced production. The production data obtained from the initial few years can be used to infer potential shale barrier distribution between two original well pairs, facilitating the selection of infill wells.

8.3 Limitations and Recommendations

(1) Testing results of the field examples are encouraging. However, it should be noticed that the proposed data-driven models are generated based on synthetic cases only. Although efforts

have been made to create models representative of field projects, certain field behaviors cannot be fully captured because of limited data availability (e.g., lack of bottom-hole flowing pressures). Therefore, there are uncertainties for directly applying the proposed workflow with field application.

(2) The parameterization approach presented in this research considers only two specific rock types: sands or shale. Although the results are useful for representing the characteristics of shale barriers, further study is needed to extend the approach to incorporate more rock types, as well as other heterogeneous features such as lean zones.

(3) Although sensitivity analysis regarding the size, geometry, and position of the shale barriers is conducted, there are, undeniably, choices that are made arbitrarily with respect to the basic shale units, in order to achieve a balance between model accuracy and computational cost. Further optimization of the resolution of the basic shale units could potentially improve the efficiency and model predictability.

(4) For the synthetic cases employed in this study, their shale configurations are generated without the consideration of spatial correlation. It is recommended that future study may take into consideration of spatial correlation (e.g. variogram) and other conditioning information to improve the robustness and flexibility of the workflow. Also, due to the limitation of synthetic model set-up, the model cannot look at shale continuity beyond 80 m along y-direction and 50 m along x-direction. Although its impact might be trivial as the horizontal GR logs show that seldom shale barriers exceed this length, it is recommended that future study take this into consideration.

(5) In this research, multiple machine learning and data mining techniques (e.g. artificial neural network, support vector machine, K-means, decision tree, etc.) have been successfully

implemented for the purpose of regression, classification, and clustering. There are many other techniques (e.g. deep learning algorithms) available for the similar problems, but their utilities have not been examined in this study. It would be interesting to investigate other options to further optimize the workflow.

(6) For the cluster analysis workflow proposed in Chapter 5, future studies may consider formulating or utilizing other static distance functions, including the Hausdorff distance (Dubuisson & Jain, 1994) or the connected hydrocarbon volume (CHV), as proposed by (Wilde & Deutsch, 2012). Kernel transformation, principal component analysis, and other clustering techniques can also be applied.

(7) The field data employed in this study are all gathered from the public domain. Although much effort has been made to interpret and understand the data, there are inevitable errors and unprecise estimations regarding the model setup and field case testing. It is anticipated that domain knowledge and additional data from the field operator could help to eliminate certain assumptions and reduce uncertainties in various reservoirs and operational parameters. In addition, further testing is needed to validate the utility of the proposed methodology for practical reservoir management and decision-making.

BIBLIOGRAPHY

Akin, S. (2005). Mathematical modeling of steam assisted gravity drainage. *SPE Reservoir Evaluation & Engineering*, *8*(05), 372-376.

Alberta Energy Regulator (2012). *Suncor Firebag 2012 ERCB Performance Presentation*. Alberta Energy Regulator, Calgary, AB, Canada, accessed Mar. 15, 2016, https://www.aer.ca/documents/oilsands/insitu-presentations/2012AthabascaSuncorFirebagSAGD8870.pdf.

Alberta Energy Regulator (2013). *Suncor Firebag 2013 ERCB Performance Presentation*. Alberta Energy Regulator, Calgary, AB, Canada, accessed Mar. 15, 2016, https://www.aer.ca/documents/oilsands/insitu-presentations/2013AthabascaSuncorFirebagSAGD8870.pdf.

Alberta Energy Regulator (2014). *Suncor Firebag 2014 AER Performance Presentation*. Alberta Energy Regulator, Calgary, AB, Canada, accessed Mar. 15, 2016, https://www.aer.ca/documents/oilsands/insitu-presentations/2014AthabascaSuncorFirebagSAGD8870.pdf.

Alberta Energy Regulator (2015). *Suncor Firebag 2015 AER Performance Presentation*. Alberta Energy Regulator, Calgary, AB, Canada, accessed Mar. 15, 2016, https://www.aer.ca/documents/oilsands/insitu-presentations/2015AthabascaSuncorFirebagSAGD8870.pdf.

Alberta Energy Regulator (2018). *ST98: 2018 Alberta's Energy Reserves & Supply/Demand outlook*. Alberta Energy Regulator, Calgary, AB, Canada, accessed Oct. 21, 2018, https://www.aer.ca/documents/sts/ST98/ST98-2018_Executive_Summary.pdf.

Al-anazi, A. F., & Gates, I. D. (2010). Support-vector regression for permeability prediction in a heterogeneous reservoir: A comparative study. *SPE Reservoir Evaluation & Engineering*, *13*(03), 485-495.

Amirian, E., Leung, J. Y., Zanon, S., & Dzurman, P. (2015). Integrated cluster analysis and artificial neural network modeling for steam-assisted gravity drainage performance prediction in heterogeneous reservoirs. *Expert Systems with Applications*, *42*(2), 723-740.

Arpat, G. B. (2005). *Sequential simulation with patterns*. Doctoral dissertation, Stanford University, USA.

Ballester, P. J., & Carter, J. N. (2007). A parallel real-coded genetic algorithm for history matching and its application to a real petroleum reservoir. *Journal of Petroleum Science and Engineering*, *59*(3), 157-168.

Birrel G. E., & Putnam P. E. (2000). A study of the influence of reservoir architecture on SAGD steam chamber development at the Underground Test Facility, northeaster Alberta, Canada, using a graphical analysis of temperature profiles. Paper presented at the *Petroleum Society's Canadian International Petroleum Conference*, Calgary, AB, Canada.

Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). New York, NY, USA: Springer Science & Business Media.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory,* 144-152, ACM.

Box, J. F. (1980). RA Fisher and the design of experiments, 1922–1926. *The American Statistician*, *34*(1), 1-7.

Brun, B., Gosselin, O., & Barker, J. W. (2004). Use of prior information in gradient-based history matching. *SPE Journal*, *9*(01), 67-78.

Butler, R., McNab, G., & Lo, H. (1981). Theoretical studies on the gravity drainage of heavy oil during in-situ steam heating. *The Canadian Journal of Chemical Engineering*, *59*(4), 455-460.

Butler, R. M. (1985). A new approach to the modelling of steam-assisted gravity drainage. *Journal of Canadian Petroleum Technology*, *24*(03), 42-51.

Butler, R. M. (1991). *Thermal Recovery of Oil and Bitumen*, Englewood Cliffs, NJ, USA: Prentice Hall.

Butler, R. M. (2004). The behaviour of non-condensible gas in SAGD-A rationalization, *Journal of Canadian Petroleum Technology*, *43*(01), 28-34.

Cenovus Energy (2011), *Environmental Assessment - Cenovus TL ULC Telephone Lake In-Situ Project - EIA Report and application for approval,* Cenovus Energy, Calgary, AB, Canada, accessed Mar. 15, 2016, accessed Feb. 17, 2019, https://open.alberta.ca/dataset/4871f71a-c541-47f5-9d79-58ef4a80851a/resource/e528644a-1eb7-47f7-9f10-d4d706e93abc/download/v1sec4.pdf.

Chen, Q., Gerritsen, M. G., & Kovscek, A. R. (2008). Effects of reservoir heterogeneities on the steam-assisted gravity-drainage process. *SPE Reservoir Evaluation & Engineering*, *11*(05), 921-932.

CMG (2015). *STARS: Users' Guide, Advanced Processes & Thermal Reservoir Simulator (Version 2015)*, Computer Modeling Group, Calgary, AB, Canada.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297.

Cox, T. F., & Cox, M. A. (2001). *Multidimensional scaling* (2nd ed.). London, UK: Chapman and Hall/CRC.

Crone, S. F., Guajardo, J., & Weber, R. (2006). The impact of preprocessing on support vector regression and neural networks in time series prediction. Paper presented at *2006 International Conference on Data Mining*, Las Vegas, NV, USA.

Dang, T. Q. C., Chen, Z., Nguyen, T. B. N., Bae, W., & Mai, C. L. (2013). Numerical simulation of SAGD recovery process in presence of shale barriers, thief zones, and fracture system. *Petroleum Science and Technology*, *31*(14), 1454-1470.

Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical software library and user's guide*. New York, NY, USA: Oxford University Press.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., & Vapnik, V. (1997). Support vector regression machines. In *Advances in Neural Information Processing Systems*, 155-161.

Dubuisson, M. P., & Jain, A. K. (1994). A modified Hausdorff distance for object matching. In: *Proceedings of 12th international conference on pattern recognition*, 566-568.

de Sousa, S. H. G. (2007). Scatter search metaheuristic applied to the history matching problem. Paper presented at the *SPE Annual Technical Conference and Exhibition*, Anaheim, CA, USA.

Edmunds, N. R. (2000). Investigation of SAGD steam trap control in two and three dimensions. *Journal of Canadian Petroleum Technology*, *39*(01), 30-40.

Edmunds, N. R., Haston, J., & Best, D. A. (1988). Analysis and Implementation of the steam assisted gravity drainage process at the AOSTRA UTF. Paper presented at the *4th UNITAR/UNDP International Conference on Heavy Crude and Tar Sands,* Edmonton, AB, Canada.

Elkatatny, S. (2018). Application of Artificial Intelligence Techniques to Estimate the Static Poisson's Ratio Based on Wireline Log Data. *Journal of Energy Resources Technology*, *140*(7), 072905.

Elliot, K. T., & Kovscek, A. R. (2001). A numerical analysis of the single-well steam assisted gravity drainage process (SW-SAGD). *Petroleum Science and Technology*, *19*(7-8), 733-760.

Fedutenko, E., Yang, C., Card, C., & Nghiem, L.X. (2014). Time-dependent neural network based proxy modeling of SAGD process. Paper presented at the *SPE Heavy Oil Conference-Canada*, Calgary, AB, Canada.

Feizabadi, S. A., Zhang, X. K., & Yang, P. (2014). An integrated approach to building history-matched geomodels to understand complex long lake oil sands reservoirs, part 2: Simulation. Paper presented at the *SPE Heavy Oil Conference-Canada*, Calgary, AB, Canada.

Fisher, R. A. (1936). *Statistical Methods for Research Workers* (6th ed.), Edinburgh, England: Oliver & Boyd.

Fisher, R. A. (1937). *The Design of Experiments* (2nd ed.). Edinburgh, England: Oliver & Boyd.

Fortin, F. A., Rainville, F. M. D., Gardner, M. A., Parizeau, M., & Gagné, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, *13*(Jul), 2171-2175.

Gao, G., Li, G., & Reynolds, A. C. (2007). A stochastic optimization algorithm for automatic history matching. *SPE Journal*, *12* (2), 196-208.

Good, W. K., Rezk, C., & Felty, B. D. (1997). *Possible effects of gas caps on SAGD performance*. Alberta Energy and Utilities Board, Department of Energy, Calgary, Alberta.

Government of Alberta (2017). *Oil Sands Facts and Stats*. Report prepared by Government of Alberta, Edmonton, AB, Canada, accessed Oct. 21, 2018, https://open.alberta.ca/dataset/b6f2d99e-30f8-4194-b7eb-76039e9be4d2/resource/063e27cc-b6d1-4dae-8356-44e27304ef78/download/fsoilsands.pdf.

Gu, F., Rismyhr, O., Kjosavik, A., & Chan, M. (2013). Optimization of SAGD wind-down and blow-down for Athabasca oil sands. Paper presented at the *SPE Heavy Oil Conference-Canada*. Calgary, AB, Canada.

Guyon, I., Boser, B., & Vapnik, V. (1993). Automatic capacity tuning of very large VC-dimension classifiers. In *Advances in Neural Information Processing Systems*, 147-155.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique, In: *Princeton University Bulletin*, 49-52.

Hajizadeh, Y., Christie, M., & Demyanov, V. (2011). Ant colony optimization for history matching and uncertainty quantification of reservoir models. *Journal of Petroleum Science and Engineering*, *77*(1), 78-92.

Haykin, S.S. (2008). *Neural networks and learning machines* (3rd ed.). Upper Saddle River, NJ, USA: Pearson.

Hiebert, A. D., Morrish, I. C., Card, C., Ha, H., Porter, S., Kumar, A., Sun, F., Close, J. C. (2013). Incorporating 4D seismic steam chamber location information into assisted history matching for A SAGD simulation. Paper presented at the *SPE Heavy Oil Conference-Canada*, Calgary, AB, Canada.

Honarkhah, M., & Caers, J. (2010). Stochastic simulation of patterns using distance-based pattern modeling. *Mathematical Geosciences*, *42*(5), 487-517.

IHS Energy, 2015. *AccuMap Software*. 321 Inverness Drive South Englewood, CO, accessed Apr. 23, 2018, http://www.ihsenergy.com.

Ito, Y., & Suzuki, S. (1999). Numerical simulation of the SAGD process in the Hangingstone oil sands reservoir. *Journal of Canadian Petroleum Technology*, *38*(09), 27-35.

Ito, Y., & Chen, J. (2010). Numerical history match of the Burnt Lake SAGD Process. *Journal of Canada Petroleum Technology*, *49*(5), 40-49.

Jacquard, P. (1965). Permeability distribution from field pressure data. *Society of Petroleum Engineers Journal*, *5*(04), 281-294.

Jafarpour, B., & Tarrahi, M. (2011). Assessing the performance of the ensemble Kalman filter for subsurface flow data integration under variogram uncertainty. *Water Resources Research*, *47*(5), 1-26.

Jain, T., Patel, R. G., & Trivedi, J. (2017). Application of polynomial chaos theory as an accurate and computationally efficient proxy model for heterogeneous steam - assisted gravity drainage reservoirs. *Energy Science & Engineering*, *5*(5), 270-289.

Jia, X., Cunha, L., & Deutsch, C. (2009). Investigation of a stochastic optimization method for automatic history matching of SAGD processes. *Journal of Canadian Petroleum Technology*, *48*(01), 14-18.

Kabanikhin, S. I. (2008). Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-Posed Problems*, *16*(4), 317-357.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence*, *24*(7), 881-892.

Kingsford, C., & Salzberg, S. L. (2008). What are decision trees?. *Nature biotechnology*, *26*(9), 1011-1013.

Klie, H. (2015). Physics-based and data-driven surrogates for production forecasting. Paper presented at the *SPE Reservoir Simulation Symposium*, Houston, TX, USA.

Kisman K. E., & Yeung K. C. (1995). Numerical study of the SAGD process in the Burnt Lake Oil Sands Lease. Paper presented at the *SPE International Heavy Oil Symposium*, Calgary, AB, Canada.

Kruger, W. D. (1961). Determining areal permeability distribution by calculations. *Journal of Petroleum Technology*, *13*(07), 691-696.

Le Ravalec, M., Morlot, C., Marmier, R., & Foulon, D. (2009). Heterogeneity impact on SAGD process performance in mobile heavy oil reservoir. *Oil & Gas Science and Technology – Rev. IFP*, *64* (4), 469-476.

Le Van, S., & Chon, B. H. (2018). Effective Prediction and Management of a CO2 Flooding Process for Enhancing Oil Recovery Using Artificial Neural Networks. *Journal of Energy Resources Technology*, *140*(3), 032906.

Lee, H., Jin, J., Shin, H., & Choe, J. (2015). Efficient Prediction of SAGD Productions Using Static Factor Clustering. *Journal of Energy Resources Technology*, *137*(3), 032907.

Lee, K., Lim, J., Choe, J., & Lee, H. S. (2017). Regeneration of channelized reservoirs using history-matched facies-probability map without inverse scheme. *Journal of Petroleum Science and Engineering*, *149*, 340-350.

Li, P. (2006). *Numerical simulation of the SAGD process coupled with geomechanical behavior*. Doctoral dissertation, University of Alberta, Canada.

Liu, F., Wang, J., Morris, P. (2013). A practical approach to history-matching large, multi-well SAGD simulation models: a Mackay River case study. Paper presented at the *SPE Heavy Oil Conference-Canada*, Calgary, AB, Canada.

Liu, J., Jaiswal, A., Yao, K. T., & Raghavendra, C. S. (2015). Autoencoder-derived features as inputs to classification algorithms for predicting well failures. Paper presented at *SPE Western Regional Meeting*, Garden Grove, CA, USA.

Liu, N., & Oliver, D. S. (2003). Evaluation of Monte Carlo methods for assessing uncertainty. *SPE Journal*, *8*(02), 188-195.

Liu, N., & Oliver, D. S. (2005). Ensemble Kalman filter for automatic history matching of geologic facies. *Journal of Petroleum Science and Engineering*, *47*(3), 147-161.

Liu, Y., & Horne, R. N. (2012). Interpreting pressure and flow-rate data from permanent downhole gauges by use of data-mining approaches. *SPE Journal*, *18*(01), 69-82.

Liu, Y., Yao, K., Raghavenda, C. S., Wu, A., Guo, D., Zheng, J., Olabinjo, L., Balogun, O., & Ershaghi, I. (2013). Global model for failure prediction for rod pump artificial lift systems. Paper presented at the *SPE Western Regional & AAPG Pacific Section Meeting*, Monterey, CA, USA.

Ma, Z., Leung, J. Y., Zanon, S., & Dzurman, P. (2015). Practical implementation of knowledge-based approaches for steam-assisted gravity drainage production analysis. *Expert Systems with Applications*, *42*(21), 7326-7343.

Ma, Z., Leung, J. Y., & Zanon, S. (2018). Integration of artificial intelligence and production data analysis for shale heterogeneity characterization in SAGD reservoirs. *Journal of Petroleum Science and Engineering*, *163*, 139-155.

McDonald, J. H. (2009). *Handbook of biological statistics*. Baltimore, MD, USA: Sparky House Publishing.

Maschio, C., Vidal, A. C., & Schiozer, D. J. (2008). A framework to integrate history matching and geostatistical modeling using genetic algorithm and direct search methods. *Journal of Petroleum Science and Engineering*, *63*(1), 34-42.

Mojarad, M., & Dehghanpour, H. (2016). Analytical modeling of emulsion flow at the edge of a steam chamber during a steam-assisted-gravity-drainage process. *SPE Journal*, *21*(02), 353-363.

Moussa, T., Elkatatny, S., Mahmoud, M., & Abdulraheem, A. (2018). Development of New Permeability Formulation From Well Log Data Using Artificial Intelligence Approaches. *Journal of Energy Resources Technology*, *140*(7), 072903.

Nielsen, M. A. (2015). *Neural Network and Deep learning*, Determination Press.

Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, *24*(12), 1565-1567.

Qu, X., & Wu, C. J. (2005). One-factor-at-a-time designs of resolution V. *Journal of statistical planning and inference*, *131*(2), 407-416.

Oloso, M. A., Hassan, M. G., Bader-El-Den, M. B., & Buick, J. M. (2018). Ensemble SVM for characterisation of crude oil viscosity. *Journal of Petroleum Exploration and Production Technology*, *8*(2), 531-546.

Ouenes, A., Brefort, B., Meunier, G., & Dupere, S. (1993). A new algorithm for automatic history matching: application of simulated annealing method (SAM) to reservoir inverse modeling. Paper SPE 26297.

Park, K., & Caers, J. History Matching in Low-Dimensional Connectivity Vector Space, Paper presented at the *EAGE Petroleum Geostatistics 2007 Conference*, Cascais, Portugal.

Patel, R. G., Rahim, S., & Li, Z. (2015). Initial sampling of ensemble for steam-assisted-gravity-drainage-reservoir history matching. *Journal of Canadian Petroleum Technology*, *54*(06), 424-441.

Preece, D. A. (1990). RA Fisher and experimental design: a review. *Biometrics*, *46*, 925-935.

Pooladi-Darvish, M., & Ali, S. M. (1994). Steam heating of fractured formations containing heavy oil: basic premises and a single-block analytical model. Paper presented at the *SPE Annual Technical Conference and Exhibition*, New Orleans, LA, USA.

Popa, A., Ramos, R., Cover, A., & Popa, C. (2005). Integration of artificial intelligence and lean sigma for large-field production optimization: application to Kern River Field. Paper presented at the *SPE Annual Conference and Exhibition*, Dallas, TX, USA.

Reis, J. C. (1992). A steam-assisted gravity drainage model for tar sands: linear geometry. *Journal of Canadian Petroleum Technology*, *31*(10), 14-20.

Rose, P. E. (1993). *The steam-assisted gravity drainage of oil sand bitumen*. Doctoral dissertation, The University of Utah, USA.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

Scheidt, C., & Caers, J. (2009). Representing spatial uncertainty using distances and kernels. *Mathematical Geosciences*, *41*(4), 397-419.

Saltuklaroglu, M., Wright, G. N., Conrad, P. R., McIntyre, J. R., & Manchester, G. J. (2000). Mobil's SAGD experience at Celtic, Saskatchewan. *Journal of Canadian Petroleum Technology*, *39*(04), 45-51.

Shahkarami, A., Mohaghegh, S. D., Gholami, V. & Haghighat, S. (2014). Artificial intelligence (AI) assisted history matching. Paper presented at the *SPE Western North American and Rocky Mountain Joint Regional Meeting*, Denver, CO, USA.

Sharma, J., & Gates, I. D. (2010). Multiphase flow at the edge of a steam chamber. *The Canadian Journal of Chemical Engineering*, *88*(3), 312-321.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199-222.

Sun, Q., & Ertekin, T. (2016). Structuring an artificial intelligence based decision making tool for cyclic steam stimulation processes. *Journal of Petroleum Science and Engineering*. *154*, 564-575.

Suzuki, S., & Caers, J. (2008). A distance-based prior model parameterization for constraining solutions of spatial inverse problems. *Mathematical Geosciences*, *40*(4), 445-469.

The MathWorks Inc., *MATLAB and Global Optimization Toolbox Release 2014b*, Natick, Massachusetts, United States.

TOP Analysis, 2015. *TOP Analysis Software*. TOP Analysis, Calgary, AB, Canada, accessed Nov. 17, 2015, http://www.topanalysis.com.

Vapnik, V., & Lerner, A. (1963). Pattern recognition using generalized portrait method. *Automation and remote control*, *24*, 774-780.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc..

Wang, C., & Leung, J. (2015). Characterizing the effects of lean zones and shale distribution in steam-assisted-gravity-drainage recovery performance. *SPE Reservoir Evaluation & Engineering*, *18*(03), 329-345.

Wang, J., & Wan, W. (2009). Experimental design methods for fermentative hydrogen production: a review. *International Journal of Hydrogen Energy*, *34*(1), 235-244.

Wickelmaier, F. (2003). An introduction to MDS. *Sound Quality Research Unit, Aalborg University, Denmark*. Accessed Dec. 30, 2017, http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=22FC374B48253FEAEBAA2FA675A85A2C?doi=10.1.1.495.4629&rep=rep1&type=pdf .

Wilde, B., & Deutsch, C. (2012). Calculating an improved connected hydrocarbon volume with line-of-sight for ranking realizations by SAGD performance. *CCG Annual Report*, *14*, 205.

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, *4*(2), 65-85.

Wu, X. & Nyland, E. (1986). Well log data interpretation using artificial intelligence technique. Paper presented at the *SPWLA 27th Annual Logging Symposium*, Houston, TX, USA.

Yang, C., Nghiem, L. X., Card, C., & Bremeier, M. (2007). Reservoir model uncertainty quantification through computer-assisted history matching. Paper presented at the *SPE Annual Technical Conference and Exhibition*, Anaheim, CA, USA.

Yang, G., & Butler, R. (1992). Effects of reservoir heterogeneities on heavy oil recovery by steam-assisted gravity drainage. *Journal of Canadian Petroleum Technology*, *31*(08), 37-43.

Yee, C. T., & Stroich, A. (2004). Flue gas injection into a mature SAGD steam chamber at the Dover project (formerly UTF). *Journal of Canadian Petroleum Technology*, *43*(01), 54-61.

Yu, T., Wilkinson, D., & Castellini, A. (2008). Constructing reservoir flow simulator proxies using genetic programming for history matching and production forecast uncertainty analysis. *Journal of Artificial Evolution and Applications*, *2008*, 263108.

Yu, H. & Wilamowski, B. M. (2011). “Levenberg-Marquardt Training”, *The Industrial Electronics Handbook, Vol. 5 –Intelligence Systems*, Boca Raton, FL, USA: CRC Press.

Zhang W., Youn S., & Doan Q. (2005). Understanding reservoir architectures and steam chamber growth at Christina Lake, Alberta, by using 4D seismic and crosswell seismic imaging. Paper presented at the *SPE/PS-CIM/CHOA International Thermal Operations & Heavy Oil Symposium*, Calgary, AB, Canada.

Zhang, X. K., Feizabadi, S. A., & Yang, P. (2014). An integrated approach to building history-matched geomodels to understand complex long lake oil sands reservoirs, part 1: Geomodeling. Paper presented at the *SPE Heavy Oil Conference-Canada*, Calgary, AB, Canada.

Zhao, L., Law, D. S., Nasr, T. N., Coates, R., Golbeck, H., Beaulieu, G., & Heck, G. (2005). SAGD wind-down: lab test and simulation. *Journal of Canadian Petroleum Technology, 44*(01), 49-53.

Zheng, J., Leung, J. Y., Sawatzky, R. P., & Alvarez, J. M. (2018a). A proxy model for predicting SAGD production from reservoirs containing shale barriers. *Journal of Energy Resources Technology*, *140*(12), 122903.

Zheng, J., Leung, J.Y., Sawatzky, R.P., and Alvarez, J.M. (2018b). An AI-based workflow for estimating shale barrier configurations from SAGD production histories. *Neural Computing and Applications*, https://doi.org/10.1007/s00521-018-3365-9.

Zheng, J., Leung, J.Y., Sawatzky, R.P., and Alvarez, J.M. (2018c). A cluster-based approach for visualizing and quantifying the uncertainty in the impacts of uncertain shale barrier configurations on SAGD production, Paper presented at *SPE Canada Heavy Oil Technical Conference*, Calgary, AB, Canada.