

SURROGATE-BASED PREDICTION AND OPTIMIZATION OF
MULTILATERAL ICV FLOW PERFORMANCE

A THESIS
SUBMITTED TO THE DEPARTMENT OF
ENERGY RESOURCES ENGINEERING
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Mohammad Jawad Aljubran

June 2020

© Copyright by Mohammad Jawad Aljubran 2020
All Rights Reserved

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as partial fulfillment of the degree of Master of Science in Energy Resources Engineering.



(Roland N. Horne) Principal Adviser

Abstract

Smart completions enable physical measurements over space and time, which provides large volumes of information at unprecedented rates. However, the optimization of inflow control valve (ICV) settings of smart multilateral wells is a challenging task. Traditionally, ICV field tests, evaluating well performance at different ICV settings, are conducted to observe flow behavior and configure ICV's, however this is often suboptimal. This study investigated a surrogate-based prediction and optimization algorithm that minimizes the number of ICV field tests required, predicts well performance of all unseen combination of ICV settings, and determines the optimal ICV setting and net present value (NPV).

To achieve the study objective and capture the extent and variation of the problem, five numerical reservoir models were considered: Case-A, Case-B, Case-C, Case-D, and Case-E. These models differ in many aspects, including geological settings, reservoir rock and fluid properties, number of fluid phases, etc. While four are synthetic models, Case-E is based on a real offshore field located in Saudi Arabia. To optimally determine candidate ICV field tests, several sampling techniques were investigated and applied, including random, space-filling, and adaptive sampling. Predictive surrogates were trained using cross-validated feedforward neural networks. Optimization was achieved using direct enumeration and mesh adaptive direct search (MADS). Both deterministic and stochastic optimization tasks were considered. Deterministic optimization tackles cases where operators are optimizing a real well drilled and completed in the field or optimizing an exact numerical reservoir model. Meanwhile, stochastic optimization is concerned with uncertain numerical reservoir models. The utility framework was utilized to account for decision makers' risk aversion along

with the uncertainty and monetary value associated with different decisions of ICV settings.

Algorithm performance was evaluated based on the number of ICV field tests required to: 1) surpass R^2 thresholds of 80% and 90% on all unseen scenarios, and 2) match the optimal ICV settings and NPV. To determine the diminishing value of additional ICV field tests, the triangulation sampling loss was used as a stoppage criterion. Surface and downhole oil and water flow prediction and optimization were achieved successfully in the different reservoir models using this approach. Considering the real reservoir model of Case-E with multiple producers at the crest or periphery, the algorithm only required six and ten ICV field tests on average to achieve 80% and 90% R^2 across the different scenarios of this real reservoir model. The algorithm achieved robust prediction results for different Case-E scenarios involving different well locations (crest and periphery) and wellbore configurations (forked and fishbone). Furthermore, uniform and Gaussian noise types were introduced at levels of 10%, 30%, and 50% to the training production profiles to evaluate the model robustness to noise. The algorithm achieved similar results to the noise-free counterpart, hence it is robust to these common types of noise. The resultant surrogate was also used to deterministically decide on the optimal settings of ICV devices and predict NPV effectively. Further improvement was accomplished through adaptively sampling and fitting the surrogate to rather predict NPV explicitly where NPV predictions were generated with nearly 95% R^2 given 20 ICV field tests.

Stochastic optimization was investigated by applying the utility framework coupled with MADS on the proposed surrogate-based prediction. With respect to risk aversion, different decision makers were considered (risk-averse, neutral, and risk-prone). The exponential utility was used as an analytical expression to evaluate the expected utility of a given decision of ICV settings. Stochastic optimization was demonstrated for a single-well scenario, where 10,240 simulation runs would be required to accomplish exhaustive analysis. Instead, the proposed algorithm achieved comparable optimal results with only 600 simulation runs, yielding 94% saving in simulation time. Another scenario involved the stochastic optimization of two wells, where 5,242,880 simulation runs would be required to accomplish exhaustive analysis.

Instead, the proposed algorithm achieved comparable optimal results with only 2,000 simulation runs, yielding 99.96% saving in simulation time.

Using adaptive sampling and machine learning proved effective in the prediction and optimization of surface and downhole flow profiles of smart wells. This method further allows for dynamically optimizing field strategy in a reinforcement learning setting where production data are used continuously to further improve the prediction performance.

Acknowledgments

I firstly would like to acknowledge Professor Roland Horne for being my advisor during my MS journey. His patience, support, and candid supervision were incomparable. He set a great example in professionalism and technical competence, and I am very grateful for the opportunity to work with him. I am also thankful to all other faculty and staff at the ERE Department. Particularly, I would like to address the SUPRI-D and SFC research groups who always shared useful knowledge and hacks. I am also lucky to have met and worked alongside my ERE student colleagues who steepened my learning curve and put me on my toes in a very productive and challenging learning environment. I also thank my officemates and roommates for the high spirit and intellectual conversations, and special thanks go to Anthony Boukarim. My journey at Stanford has been full of great friendships and memorable moments.

I will forever owe my lovely family for their continuous support throughout my studies. I am grateful to my sister, Huda, big brother, Maitham, and little brother, Hussain. They were always present during the ups and downs, making my life bright and happy. Special thanks go to my sister's husband, Adel, for always sharing and caring. Last but not least, I express my deepest love and appreciation to my fiancée, Alaa, whose presence I could feel despite the long distance. Her joyful spirit, generous personality, and thoughtful surprises always make me smile.

Thank you God for your fathomless blessings, I am forever your servant.

Contents

Abstract	v
Acknowledgments	viii
1 Introduction	1
1.1 Smart Multilateral Technology	1
1.2 Statement of the Problem	2
1.3 Literature Review	3
1.3.1 Deterministic Optimization	5
1.3.2 Stochastic Optimization	6
1.4 Scope of Work and Thesis Outline	8
2 Methodologies	9
2.1 Numerical Reservoir Models	10
2.1.1 Case-A: Homogeneous, Single-Producer Model	12
2.1.2 Case-B: Heterogeneous, Single-Producer Model	13
2.1.3 Case-C: Fluvial Channel, Single-Producer Model	14
2.1.4 Case-D: SPE-9, Two-Producers Model	16
2.1.5 Case-E: Real Reservoir Model with Multiple-Producers	16
2.2 Sampling Techniques	19
2.2.1 Space-Filling Latin Hypercube Sampling (LHS)	20
2.2.2 Adaptive Sampling	22
2.3 Machine/Deep Learning Algorithms	26
2.3.1 Feedforward Neural Networks	26

2.4	Stochastic Optimization	29
2.4.1	Distance-Kernel Method (DKM)	30
2.4.2	Utility Framework	34
2.5	Optimization Solvers	36
2.5.1	Mesh Adaptive Direct Search (MADS)	36
2.6	Summary	37
3	Results and Discussion	38
3.1	Prediction Task	38
3.2	Deterministic Optimization	52
3.3	Stochastic Optimization	55
4	Conclusions and Future Work	66
	Bibliography	68

List of Tables

2.1	Case-A: Rock, fluid, and well properties at initial reservoir conditions.	13
2.2	Case-C: Rock, fluid, and well properties at initial reservoir conditions.	15
2.3	Case-D: Rock, fluid, and well properties at initial reservoir conditions.	16
2.4	Case-E: Rock, fluid, and well properties at initial reservoir conditions.	18
3.1	Comparison of the enumerated and MADS-calculated optimal settings based on the ground truth and predicted utility. Red values indicate mismatching optimal ICV settings between the ground truth and surrogate predictions.	62
3.2	Comparison of the enumerated and MADS (in brackets) optimal settings based on the utility framework for Case-E1 with two trilateral producers.	65

List of Figures

1.1	OnePetro publication quantification over time of two search queries: “inflow control valve” and “machine learning”	4
2.1	Summary of the algorithm major components: design of experiment, ICV field tests, adaptive sampling, surrogate model training, and optimization.	9
2.2	Common multilateral well designs with varying levels of complexity [24]. These configurations alter the path and interaction of fluid flow across the different well laterals, which induces different pressure regimes and affects the lateral-to-lateral interdependence.	10
2.3	Case-A: Cartesian reservoir model at the initial state where the oil reservoir (red) is 100% saturated with oil and occupies the upper five grid layers while the water aquifer (blue) is 100% saturated with water and occupies the lower four grid layers. A single trilateral producer is located near the center with three ICV devices located at the tie-in points to the motherbore.	12
2.4	Case-B: Permeability distribution of a heterogeneous reservoir model which exhibits identical characteristics to Case-A, except for the permeability distribution across the oil reservoir layers. We varied permeability, such that it ranges between 0.2 – 1.8 md in layers 1 and 3 and between 2 – 18 md in layers 2 and 4.	14
2.5	Case-C: Synthetic, two-dimensional, fluvial channel reservoir model with a five-spot pattern where the forked trilateral producer is placed at the center of the Cartesian model grid.	15

2.6	Case-D: Synthetic, three-dimensional, heterogeneous reservoir model with two forked trilateral producers and two horizontal injectors with a high-permeability path connecting injectors to specific producing laterals which induces early water breakthrough. Figure from [2].	17
2.7	Case-E: Saturation and permeability maps of a real model located offshore of Saudi Arabia. Figure from [2].	18
2.8	Oil and water cumulative flow rates generated using the Case-A reservoir model, where N profiles (green) are used to train and validate the machine learning surrogate while the remaining $T - N$ profiles (red) are used to evaluate its performance.	20
2.9	Three illustrative examples of LHS designs in two dimensions, where each row and column contains a single sample point. Figure from [92].	21
2.10	Two-dimensional demonstration of Delaunay triangulation [49].	24
2.11	Demonstration of the Watson retriangulation algorithm where a new point \vec{x}_i is added to an existing Delaunay triangulation (left), which is then reconstructed using Watson algorithm (right).	25
2.12	Most commonly used activation functions that introduce nonlinearity to the neural network. Because this is a regression problem, we used ReLU across the hidden layers as it is not prone to the vanishing gradient problems when compared to sigmoid and hyperbolic tanh. Note that the final output layer is always activated with the linear function.	28
2.13	DKM approach: a) calculation of distance between realization simulation responses, b) construction of distance matrix, c) mapping realization responses to Euclidean space, d) projection to feature space, e) preimage construction and percentile statistics (P10, P50, and P90). Figure from [74].	31
2.14	The ICV configuration decision tree. Figure from [44].	34
2.15	The ICV configuration decision tree with expected utility value for each decision. Figure from [44].	35
2.16	Effect of risk aversion factor of the shape of the exponential utility function.	36

3.1	Comparison between four sampling techniques: random sampling, LHS, adaptive sampling with four corners, and adaptive sampling with all cube corners (eight corners). Left, middle, and right columns refer to sampling of five, 20, and 30 field tests in total.	40
3.2	R^2 test performance of the machine learning surrogate in the prediction of the Case-A surface flow quantities with respect to the different sampling techniques.	41
3.3	R^2 test performance of the machine learning surrogate in the prediction of the Case-A surface and downhole flow quantities with respect to the different sampling techniques.	42
3.4	Average simplex hypervolume and gradient as a function of the number of ICV field tests using four-corner adaptive sampling to predict both surface and downhole flow quantities of Case-A.	43
3.5	R^2 test performance of the machine learning surrogate in the prediction of the Case-B surface and downhole flow quantities with respect to the different sampling techniques.	44
3.6	R^2 test performance of the machine learning surrogate in the prediction of the Case-C surface and downhole flow quantities with respect to the different sampling techniques.	45
3.7	R^2 test performance of the machine learning surrogate in the prediction of the Case-D surface and downhole flow quantities with respect to the different sampling techniques. This scenario is only concerned with the prediction of a single producer (Producer-1) while keeping the other producer without ICV devices.	45
3.8	R^2 test performance of the machine learning surrogate in the prediction of the Case-D surface and downhole flow quantities with respect to the different sampling techniques. This scenario is only concerned with the prediction of a single producer (Producer-2) while keeping the other producer without ICV devices.	46

3.9	R^2 test performance of the machine learning surrogate in the prediction of the Case-D surface and downhole flow quantities with respect to the different sampling techniques. This scenario demonstrates the prediction of ICV flow performance in both forked trilaterals simultaneously.	47
3.11	R^2 test performance of the machine learning surrogate in the prediction of the Case-E1 surface and downhole flow quantities with respect to the different sampling techniques.	48
3.10	Case-E well placement and configuration choices are demonstrated on the reservoir oil saturation map. These four scenarios are 1) Case-E1: fishbone multilaterals placed at the periphery (top left), 2) Case-E2: fishbone multilaterals placed at the crest (top right), 3) Case-E3: forked multilaterals placed at the periphery (bottom left), and 4) Case-E4: forked multilaterals placed at the crest (bottom right). Note that the target well in each case, where the algorithm is to be applied, is circled in red.	49
3.12	R^2 test performance of the machine learning surrogate in the prediction of the Case-E2 surface and downhole flow quantities with respect to the different sampling techniques.	50
3.13	R^2 test performance of the machine learning surrogate in the prediction of the Case-E3 surface and downhole flow quantities with respect to the different sampling techniques.	50
3.14	R^2 test performance of the machine learning surrogate in the prediction of the Case-E4 surface and downhole flow quantities with respect to the different sampling techniques.	51
3.15	Example of a randomly selected ICV setting where surface and down-hole flow rates are shown with various uniform and Gaussian noise levels.	51
3.16	Algorithm R^2 results as applied to the Case-E1 reservoir model after adding uniform and Gaussian noise. The proposed algorithm shows to be robust to noise.	52

3.17 Demonstration of NPV surface at ICV-3 setting values of 3 and 7 for the ground truth (left), four-corner flow surrogate (middle), and eight-corner flow surrogate (right). Note that the 4-corner surrogate is less accurate at mapping the NPV surface when ICV-1 and/or ICV-3 settings are set to 7 (fully open). This is because some corners are omitted in this case, which leaves these regions of the design domain less defined to the surrogate. Meanwhile, the eight-corner surrogate shows more accurate surface map as its training points are sampled to capture all design domain corners.	54
3.18 Demonstration of NPV surface at ICV-3 slices three and seven of the ground truth (left), four-corner NPV surrogate (middle), and eight-corner NPV surrogate (right). Note that the four-corner surrogate is less accurate at mapping the NPV surface when ICV-3 setting is set to seven (fully open). Meanwhile, the eight-corner surrogate shows a more accurate surface map as its training points are designed to capture all design domain corners.	55
3.19 Permeability trend along depth as used to construct the Case-E geo-statistical realizations using SGeMS.	56
3.20 Example permeability realizations (left) of the Case-E payzone as generated using SGeMS in comparison to the history-matched permeability map (right). Note that the history-matched permeability field is shown based on the irregular Cartesian grid of Case-E where it involves elongated grid cells at the boundaries.	57
3.21 Demonstration of DKM clustering and P10, P50, and P90 fit results using water cut profiles of Case-E. Squares represent the medoids of the different clusters.	58
3.22 Sample profiles of oil production rate, water production rate, and NPV based on exhaustively simulating 10,240 reservoir models using the 20 most representative permeability fields.	59
3.23 Histogram of NPV based on production of a single trilateral for a total of two years in all 10,240 simulation runs.	59

3.24 R ² histogram based on the evaluation of the 20 surrogates each trained on 30 adaptively sampled ICV field tests and tested on the remaining 511 – 30 = 481.	60
3.25 Visualization of NPV standard deviation across the different ICV settings based on the predictions of the trained surrogates.	61
3.26 Ground truth utility histograms for different risk aversion factor, r, values.	62
3.27 Ground truth utility surfaces for values of $ICV - 3 = 7$ (fully open).	63
3.28 Normalized negative utility values of MADS surrogate function evaluations for different decision makers (different risk aversion factors).	63
3.29 Evaluated points by MADS throughout the optimization of the single-producer, Case-E surrogates using utility.	64
3.30 The 20 most representative realizations as per DKM in comparison to the history-matched permeability map. Two trilaterals are drilled and completed close to the grid center.	65

Chapter 1

Introduction

1.1 Smart Multilateral Technology

Over the past few decades, horizontal wells have been adopted at growing rates as they have significantly enhanced well productivity and ultimate recovery. Furthermore, extensions of horizontal designs, i.e. multilateral wells, have come to be considered mature as the technological level has increased while the associated drilling and completion risks have decreased. Multilateral well complexity introduced control challenges which led to the development of advanced downhole technologies, namely intelligent completions.

In 1997, the first intelligent well was successfully completed in the Norwegian sector of the North Sea with inflow control valve (ICV) and permanent downhole gauge technologies [6]. Smart well technology has flourished since then, and oil and gas operators are now past the technology adoption phase. Recent research and development of smart well hardware and software led to an increasing use of this technology in field development to improve reservoir management [28]. Prominence of smart completions was evident especially in multilateral wells where they enabled spatiotemporal flow monitoring, troubleshooting without the need for workover operations, and lateral-by-lateral control. Hence, these designs allowed for maximizing reservoir contact, dynamic zonal monitoring and control, reduced well intervention frequency and cost, and increased return on drilling investment.

Intelligent wells are superior to their conventional counterparts as they enable downhole monitoring and control, and thus compartmentalization in complex multilateral wells. Intelligent design and instrumentation commonly consist of several downhole components, including main production packer, secondary zonal isolation packer/s, pressure/temperature dual gauge, and ICV devices. These designs allow for monitoring and control capabilities at each lateral tie-in. Hence, water/gas breakthrough can be monitored, tracked down to the problematic lateral, and controlled with ICV devices accordingly. Consequently, intelligent multilateral wells allow for maximizing reservoir contact, increasing return to drilling investment, minimizing intervention cost and frequency, delaying water/gas breakthrough, and isolating and monitoring reservoir zones [2].

Practical usefulness of smart multilaterals is established and multiple successful stories stand out, such as the South Shaybah and Haradh Fields in Saudi Arabia, Minagish Field, Kuwait, Agbami Field in Nigeria, Brent and Tern Fields in the UK sector of the North Sea, Gullfaks Field in the Norway sector of the North Sea, and Nakika Field in the Gulf of Mexico [4, 8]. Furthermore, reports indicate that overall hydrocarbon recovery could potentially improve by 25% on average if smart completions were installed across a field [6].

In the presence of ICV devices and downhole sensors, performance optimization of smart multilateral wells goes beyond well placement and becomes dimensionally complex. ICV control strategies can generally be divided into two types: reactive and proactive. Reactive strategies aim to correct hydrocarbon production deviations from corporate targets, hence they are focused on short-term optimization [3, 24, 33]. Proactive strategies, which are advantageous for optimal hydrocarbon recovery, aim to predict and optimize long-term well flow performance for different ICV configurations [64].

1.2 Statement of the Problem

Consider the problem of using a reservoir numerical simulator to decide simultaneously on the optimal placement and ICV configurations of a proposed multilateral

well with given configuration and specifications. This also involves uncertainty in reservoir rock and fluid parameters which requires the integration of multiple geological realizations to correctly quantify the mismatch between the numerical model and the actual reservoir in the field. The simulation tools available in the oil and gas industry are computationally expensive, hence it quickly becomes infeasible to couple numerical reservoir models with common optimizers in such highly dimensional problems. Generally, there are two approaches that alleviate this computational challenge: reduction of the necessary number of simulations, and approximation of the numerical reservoir model with a less expensive alternative.

On the other hand, as an improvement over simulation, consider the problem of ICV setting optimization in an existing smart multilateral well. In this case, the measured production data at the field pose an attractive alternative to the uncertain numerical reservoir model. Raw field data enable the training of an accurate predictive proxy and allow for robust optimization. This requires a representative model to describe the subsurface flow behavior with respect to the fluid and rock properties, well configuration, and pressure losses across the ICV constriction area.

The main objective of this research was to investigate the use of raw field multilateral well production data in developing a machine learning surrogate that predicts and optimizes lateral-by-lateral flow performance while minimizing the number of ICV field tests required. Initially, the optimization considered a deterministic model. Subsequently, this research explored stochastic ICV optimization under geological uncertainty based on the utility framework. Given the high dimensionality of ICV settings, uncertainty in reservoir rock and fluid physical quantities, and nonlinearity of porous flow, machine learning algorithms were examined to achieve this task.

1.3 Literature Review

In order to investigate the importance of ICV devices to operators and also the rapid adoption of machine learning algorithms, publications were quantified over the years. To achieve this task, OnePetro was used to search publications as it is the main library for upstream oil and gas articles. Using “inflow control valve” and “machine learning”

as independent search queries, we retrieved publications relevant to upstream research and application efforts on ICV devices and machine learning. This process resulted in 3,269 and 4,684 articles in total on ICV devices and machine learning, respectively. As seen in Fig. 1.1, publications on ICV devices record a rapid increase starting during the 1990’s. This coincides with the spring of multilateral well designs, and also with the first successful intelligent well equipped with ICV devices and permanent downhole gauge technologies [6]. Furthermore, Fig. 1.1 indicates that the application of machine learning algorithms in the upstream oil and gas industry attracted much attention over the past few years. Hence, these observations emphasize the significance of ICV devices and machine learning research, which is the focus of this thesis.

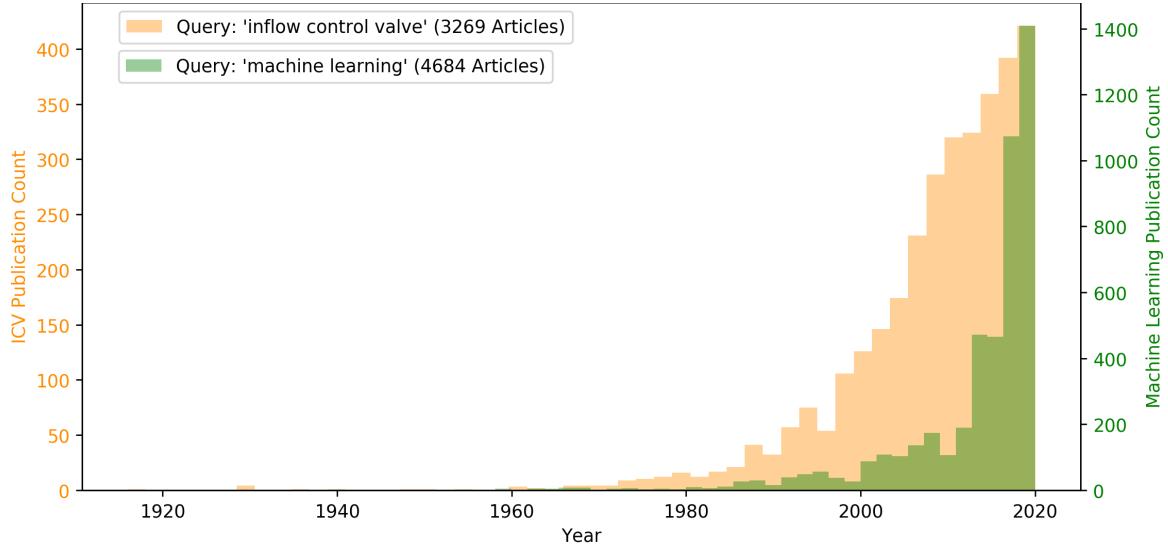


Figure 1.1: OnePetro publication quantification over time of two search queries: “inflow control valve” and “machine learning”.

In this section, we will summarize some of the existing literature on smart multilateral ICV setting optimization that is relevant to this study. First, we will highlight those research efforts focused on deterministic optimization through the coupling of common optimizers with numerical and/or analytical reservoir models. We also highlight research efforts using statistical surrogates to achieve prediction and deterministic optimization. Second, we will shed light on stochastic optimization of ICV settings under uncertainty. This section will refer to a recent publication of part of

this work by the author to highlight the most relevant and significant literature [12].

1.3.1 Deterministic Optimization

Numerical and Analytical Methods

Multiple studies focused on deterministic production and control optimization of smart wells, only considering an exact numerical reservoir model [25, 77, 61, 31, 62, 90, 30]. Many of these efforts investigated the optimization of smart well ICV settings by coupling numerical reservoir simulators with common optimizers based on steepest ascent algorithms [29]. For instance, a group of researchers proposed a field strategy to optimize ICV devices in commingled production scenarios using sequential linear programming [61], while others optimized waterflood sweep efficiency by optimizing ICV configurations under operational and facility constraints over a streamline-based simulator [31].

Meanwhile, other efforts aimed to avoid deterministic optimization over numerical reservoir models which are prone to uncertainty in fluid and rock properties, potentially leading to unrepresentative simulation results. Instead, they explored deterministic optimization of ICV settings using raw well flowback and/or production data by fitting parametric analytical models, which were subsequently optimized. One such research effort involved a two-level analytical algorithm that aims to minimize the required field ICV tests and optimize either physical quantity prediction reliability (monitoring) or cumulative oil production (control) [57, 56]. This effort successfully demonstrated this algorithm on a synthetic reservoir model with a single-lateral, multizonal smart well using the generalized reduced gradient nonlinear optimizer.

More recent efforts focused on surrogate-based optimization by training statistical proxies. While different statistical algorithms have been explored in the literature, machine learning has been found to provide sufficient nonlinearity to model the various phenomena of underground statics and dynamics. These efforts have included modeling crude oil phase behavior, heterogeneous rock permeability, multiphase flow, wire-line log measurements, and others [47, 36, 7, 5, 9]. More rigorous efforts involved using machine learning to build statistical twins of numerical reservoir models

by learning historical patterns based on real and/or numerically generated synthetic data without explicitly programming the reservoir dynamics. Various implementations involved pressure, temperature, and hydrocarbon flow rate prediction over time and space in homogeneous/heterogenous, and single/multiphase reservoirs using feature-based kernel ridge regression, support vector machines, decision trees, random forest, AdaBoost, gradient boosting, k-nearest neighbors, and nonlinear autoregressive exogenous models. Gaussian kernel ridge regression, support vector machines, and NARX were found successful in modeling single-phase flow, multiphase (oil and water) flow rate in multiwell fields, and feature-free scenarios when knowledge about the driving physics laws is limited, respectively [83, 85, 86, 84, 67, 87, 10].

Machine learning algorithms were also found useful in intelligent multilateral well prediction and deterministic optimization problems. Efforts involved developing neural network proxies to multilateral producers (without ICV), achieving an average mean absolute error (MAE) of 8% [37]. Saudi Aramco engineers trained a neural network to predict oil flow rate based on multilateral well data acquired from an oil reservoir in the Middle East [26]. Using the number and effective length of laterals, open-hole size, choke size, reservoir pressure, wellhead pressure, average permeability, they predicted oil surface flow rate with 91% R^2 on testing data. Other researchers trained a ten-neuron, one-layer fully connected neural network to predict surface and downhole total oil flow rate (single-valued prediction after a given period of production) in a heterogeneous reservoir with two multilateral producers and one injector [2]. They showed that training on ten and 80 ICV combination samples yielded 83% and 99% R^2 , respectively, on surface measurement predictions, and 93% and 99% R^2 , respectively, on downhole measurement predictions.

1.3.2 Stochastic Optimization

Due to subsurface uncertainties, the spatial distribution of reservoir flow-relevant quantities is not exactly determined. Recognizing that uncertainty represents a major drawback to the deterministic optimization of numerical reservoir models, other efforts considered stochastic optimization where realizations are generated to account

for the underlying geological uncertainties. Efforts in stochastic ICV setting optimization generally follow either reactive strategies (short-term optimization) [45, 43] or proactive strategies (long-term optimization) [53, 89]. While each offers different advantages, stochastic optimization is more essential to the success of proactive strategies as they are much more sensitive to geological uncertainties in the reservoir model. However, this requires a range of models to evaluate multiple realizations which is computationally expensive. Hence, a robust and efficient optimization algorithm is essential [76].

Research efforts addressed this problem using three stochastic optimization methods: gradient-based [41, 38, 25, 71, 15, 80], derivative-free [51, 65, 34, 16], and meta-heuristic [95, 8, 13, 64] optimization. Using the adjoint of the forward function is a common gradient-based optimization technique because it approximately costs one additional simulation run only to compute the gradient, hence it is efficient at calculating local optima in highly dimensional control optimization problems [71, 53]. Given the difficulty of complementing a commercial software with adjoint gradient programs, many researchers treat the full-physics numerical reservoir model as a black box and alternatively estimate the gradient using ensemble-based methods. However, gradient-based algorithms are generally sensitive to the initial guess and could potentially get stuck in a local optimum [15]. Unlike gradient-based algorithms, derivative-free algorithms (i.e. Hooke-Jeeves, Nelder-Mead, Line-Search, etc.) are less likely to get stuck in local optima but they become rapidly inefficient as the control optimization problem grows dimensionally [76]. Meanwhile, meta-heuristic algorithms (i.e. simulated annealing, genetic algorithm, particle swarm optimization, etc.) are common methods with more capability in capturing the global optimum, yet they rapidly become less efficient depending on the problem dimensionality and required degree of convergence.

1.4 Scope of Work and Thesis Outline

Chapter 1 set the stage and defined the challenges associated with the prediction and optimization of smart multilateral wells. Chapters 2 and 3 will focus on methodologies and study results, respectively. Chapter 2 will define five different numerical reservoir models that encompass various scenarios encountered in practice. Methodologies will also highlight and compare random, space-filling, and adaptive sampling techniques which are critical to training accurate surrogates. In addition, feedforward neural networks will be discussed as means of training surrogates on the candidate ICV field tests. Furthermore, several reduction and optimization algorithms will be introduced and integrated to achieve stochastic optimization. Chapter 3 focuses on the results obtained by applying these methodologies to the five reservoir models. Multiple deterministic and stochastic scenarios will be demonstrated to confirm the observations with regards to the surrogate-based prediction and optimization algorithm. In Chapter 4, we will ultimately make conclusions and draw insights based on the study findings.

Chapter 2

Methodologies

This chapter presents the methodologies used to achieve the research objectives. The chapter consists of five major parts: numerical reservoir models, sampling techniques, machine/deep learning algorithms, optimization solvers, and stochastic optimization. Multiple numerical reservoir models, both synthetic and real, are outlined and discussed. Different genres of sampling techniques are also compared. Several machine/deep learning algorithms are visited and evaluated with the premise of developing an accurate predictive surrogate. Lastly, deterministic and stochastic optimization approaches are described. These techniques are integrated to perform robust optimization of multilateral well ICV devices under both geologically certain and uncertain circumstances. Generally, the algorithm steps for a deterministic model or a single realization of a stochastic model can be summarized as seen in Fig. 2.1.

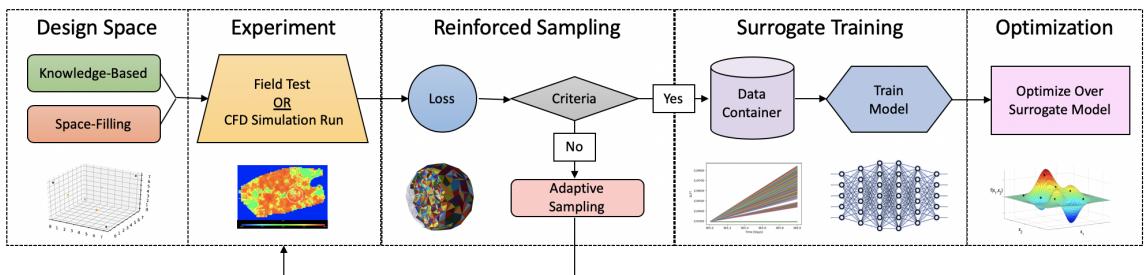


Figure 2.1: Summary of the algorithm major components: design of experiment, ICV field tests, adaptive sampling, surrogate model training, and optimization.

2.1 Numerical Reservoir Models

Schlumberger ECLIPSE was used to perform the numerical reservoir simulations throughout this work. ECLIPSE is capable of simulating multiple functionalities (reservoir properties and conditions, well configurations, production constraints, etc.) using a collection of keywords [40]. ECLIPSE facilitates the implementation of multilateral wells using the keyword WSEGS, which can simulate multilateral wells with multiple laterals in various configurations based on segments. As seen in Fig. 2.2, multilateral well designs vary depending on several factors, such as the target payzones, mechanical feasibility of drilling trajectories, expenditure, etc. We evaluated the proposed algorithm on forked and fishbone trilateral well designs in order to demonstrate the impact of well configuration on ICV performance prediction and optimization.

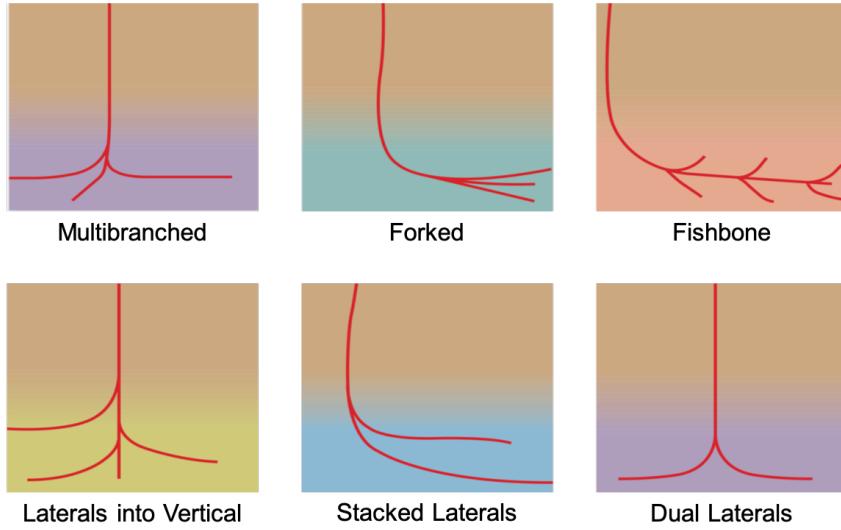


Figure 2.2: Common multilateral well designs with varying levels of complexity [24]. These configurations alter the path and interaction of fluid flow across the different well laterals, which induces different pressure regimes and affects the lateral-to-lateral interdependence.

ECLIPSE further offers multiple keywords to represent fluid flow restriction and flow through ICV devices by applying additional pressure drop across the sandface. To simulate pressure drop across ICV devices, we utilized the keyword WSEGVALVE which imposes frictional loss in the form of flow through a constriction with a given

cross-sectional area. This quantity consists of pressure drop due to supercritical flow through a constriction and pressure drop due to flow along the lateral, which is given by [40]:

$$\Delta P_{ICV} = \Delta P_c + \Delta P_f = \frac{\rho_m q_m^2}{2 A_c^2 C_v^2} + 2f \frac{L}{D} \rho_m q_m^2 A_p^2 \quad (2.1)$$

where ΔP_c and ΔP_f represent the pressure drop due to cross-sectional constriction and lateral friction, respectively, q_m is the fluid mixture volumetric flow rate, ρ_m is the fluid mixture density, A_c and A_p are the areas of the ICV constriction and pipe, respectively, f is the Fanning friction factor, and C_v is the dimensionless valve geometry coefficient.

Although the state-of-the-art technology offers continuous ICV device settings, the discrete ICV device counterparts are more commonly installed and encountered in the field. Hence, we considered ICV devices with both eight and ten discrete settings. We use ICV indices of “0”, and “7” (eight discrete settings design) or “10” (11 discrete settings design) to indicate fully closed and fully open choke configurations, respectively. Note that the constriction area is not linearly correlated with the discrete settings, rather ICV flow area is case-sensitive and depends on the well design and reservoir capacity [21]. In case the installed ICV flow area size is too large, choking ICV configurations will not effectively influence the flow regime, especially at larger ICV openings. Meanwhile, if the ICV is designed with too small flow areas for a given case, then the reservoir capacity is larger than that of the ICV and flow optimization could be physically impossible. Hence, we selected discrete ICV choke settings that are compatible with each reservoir model to ensure the ICV input constriction areas fully span the domain of interest. Note that all simulation runs in this study were performed assuming no flow or pressure constraints on the target well.

This section will present five different cases of numerical reservoir models used in this study. These models differ in many aspects, including geological settings, reservoir rock and fluid properties, number of fluid phases, etc. While four are synthetic models, one of these is based on a real offshore field located in Saudi Arabia. These models will be used to demonstrate and evaluate the subsequent prediction

and optimization algorithms.

2.1.1 Case-A: Homogeneous, Single-Producer Model¹

A geologically simple reservoir model was first constructed for algorithm development and preliminary validation. As seen in Fig. 2.3, this model involves a homogenous, light-oil reservoir with two-phase flow (oil and water), mainly driven by an underlying Carter-Tracy water aquifer [27]. This aquifer method assumes constant water influx rates rather than constant oil production rates over finite intervals of time. Table 2.1 briefly describes the primary rock and fluid properties along with other characteristics of this homogeneous reservoir model.

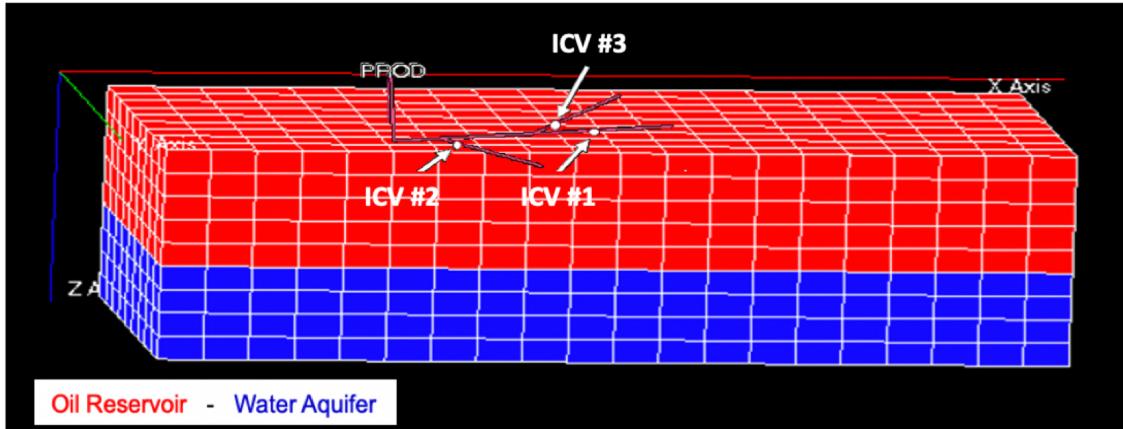


Figure 2.3: Case-A: Cartesian reservoir model at the initial state where the oil reservoir (red) is 100% saturated with oil and occupies the upper five grid layers while the water aquifer (blue) is 100% saturated with water and occupies the lower four grid layers. A single trilateral producer is located near the center with three ICV devices located at the tie-in points to the motherbore.

Case-A includes a fishbone trilateral oil producer that is located near the grid center and completed across the second layer of the oil reservoir with three ICV devices located at the tie-in points. Each ICV device is capable of 11 fixed choke settings, where “0” and “10” indicate fully shut-in and fully open choke settings,

¹Available on GitHub: https://github.com/aljubrmj/MS_Thesis/tree/master/ECLIPSE_Simulations/CaseA.

Table 2.1: Case-A: Rock, fluid, and well properties at initial reservoir conditions.

Property	Value
Reservoir Size (ft ³)	2000 × 900 × 270
Payzone Thickness (ft)	150
Porosity (%)	10
Horizontal Permeability (md)	10
Vertical Permeability (md)	1
Reservoir Pressure (psi)	6000
Aquifer Thickness (ft)	120

respectively. While discrete settings narrow down the possible scenarios, they pose a potentially more challenging optimization problem because integer programming is required. Note that ICV choke settings are represented using restriction area to simulate the induced pressure drop, which makes the corresponding maximum restriction area where the ICV is fully open dependent on the actual hole size. For this reservoir model, the ICV devices are fully shut-in and fully open at restriction area values of 0 and 0.001 ft², respectively. The production period for each combination of ICV settings was set to 4,000 days, such that the surrogate has to predict extended production profiles for different ICV settings in this case.

2.1.2 Case-B: Heterogeneous, Single-Producer Model²

To introduce geological complexity, we altered the permeability field of Case-A to have a heterogeneous reservoir model, namely Case-B. This reservoir model is identical to Case-A in all aspects, except for the heterogeneous permeability distribution across the oil reservoir layers. As seen in Fig. 2.4, we varied permeability, such that it ranges between 0.2 - 1.8 md in layers 1 and 3 and between 2 - 18 md in layers 2 and 4. Note that permeability across these layers increases monotonically along the y-axis while it is constant along the x-axis. ICV-2 and ICV-3, seen in Fig. 2.3, control flow from the most and least permeable reservoir regions, respectively.

²Available on GitHub: https://github.com/aljubrmj/MS_Thesis/tree/master/ECLIPSE_Simulations/CaseB.

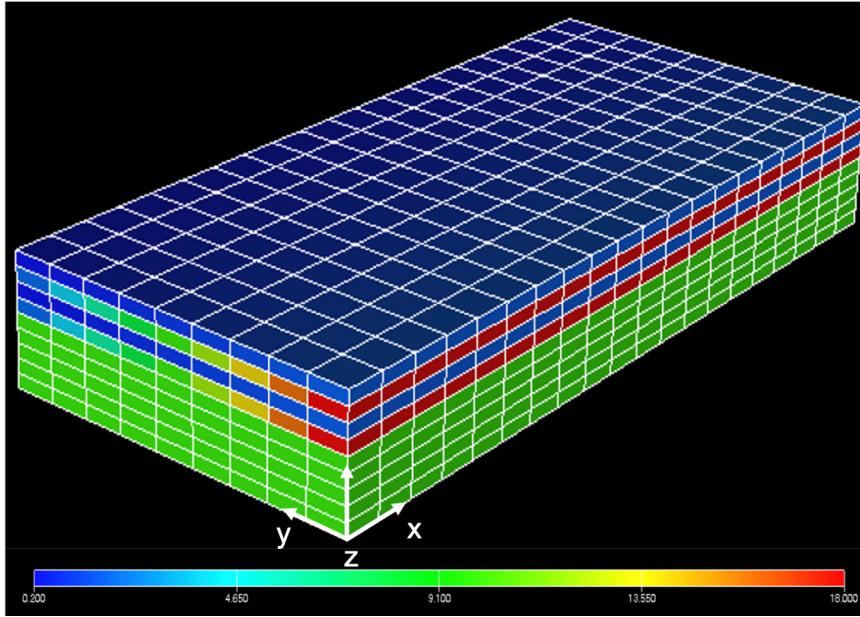


Figure 2.4: Case-B: Permeability distribution of a heterogeneous reservoir model which exhibits identical characteristics to Case-A, except for the permeability distribution across the oil reservoir layers. We varied permeability, such that it ranges between 0.2 – 1.8 md in layers 1 and 3 and between 2 – 18 md in layers 2 and 4.

2.1.3 Case-C: Fluvial Channel, Single-Producer Model³

This case involves a two-dimensional, heterogeneous, isotropic, fluvial channel reservoir model [8]. As seen in Fig. 2.5, the permeability distribution distinguishes this scenario when compared to Case-A and Case-B. Table 2.2 summarizes the rock, fluid, and well properties under initial reservoir conditions. This reservoir model is completed with a five-spot pattern where a forked trilateral producer is located at the center while four vertical power water injectors (PWI's) are placed at the corners. The producer is equipped with three ICV devices, such that each is capable of eight fixed choke settings, where “0” and “7” indicate fully shut-in and fully open choke settings, respectively. For this reservoir model, the ICV devices are fully shut-in and fully open at restriction area values of 0 and 0.00057 ft², respectively. The production period for each combination of ICV settings was set to a single day, such that

³Available on GitHub: https://github.com/aljubrmj/MS_Thesis/tree/master/ECLIPSE_Simulations/CaseC.

the surrogate model has to predict short-term production rates. This is analogous to building a surrogate to predict production flow rates across different ICV settings based on raw data of an existing well at the field.

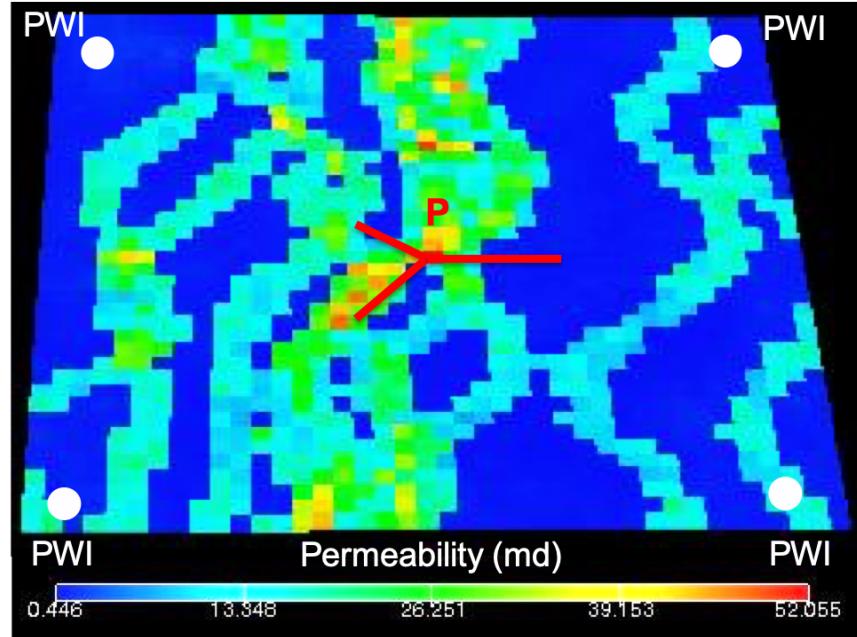


Figure 2.5: Case-C: Synthetic, two-dimensional, fluvial channel reservoir model with a five-spot pattern where the forked trilateral producer is placed at the center of the Cartesian model grid.

Table 2.2: Case-C: Rock, fluid, and well properties at initial reservoir conditions.

Property	Value
Reservoir Size (ft ³)	40 × 40 × 1
Payzone Thickness (ft)	50
Porosity (%)	30
Horizontal Permeability (md)	0.44 – 52
Vertical Permeability (md)	0.44 – 52
Reservoir Pressure (psi)	6000
Max PWI BHP (psi)	8000

2.1.4 Case-D: SPE-9, Two-Producers Model⁴

For this case, a three-dimensional, heterogeneous oil reservoir model represented by SPE-9 was used [2, 40]. The model oil saturation and permeability maps, and properties are shown in Fig. 2.6 and Table 2.3, respectively. The model involves two forked trilateral producers and two horizontal injectors. Injector laterals are drilled across the middle and bottom layers, such that they connect to a highly permeable path which feeds into one lateral of each producer. This results in early water breakthrough across those laterals which intersect the high permeability channel. Each producer is equipped with three ICV devices, such that each is capable of six fixed choke settings, where “0” and “5” indicate fully shut-in and fully open choke settings, respectively. For this reservoir model, the ICV devices are fully shut-in and fully open at restriction area values of 0 and 0.00057 ft^2 , respectively. The production period for each combination of ICV settings was set to a single day, such that the surrogate model has to predict short-term production rates. This is analogous to building a surrogate to predict production flow rates across different ICV settings based on raw data of an existing well at the field.

Table 2.3: Case-D: Rock, fluid, and well properties at initial reservoir conditions.

Property	Value
Reservoir Size (ft^3)	$24 \times 25 \times 15$
Payzone Thickness (ft)	20
Porosity (%)	12.6
Average Horizontal Permeability (md)	108
Average Vertical Permeability (md)	10.8
Reservoir Pressure (psi)	3000
Max PWI BHP (psi)	8000

2.1.5 Case-E: Real Reservoir Model with Multiple-Producers

Unlike the other reservoir models, Case-E comes from an offshore carbonate reservoir located in Saudi Arabia where oil is accumulated in a dome stratigraphic trap [1],

⁴Available on GitHub: https://github.com/aljubrmj/MS_Thesis/tree/master/ECLIPSE_Simulations/CaseD.

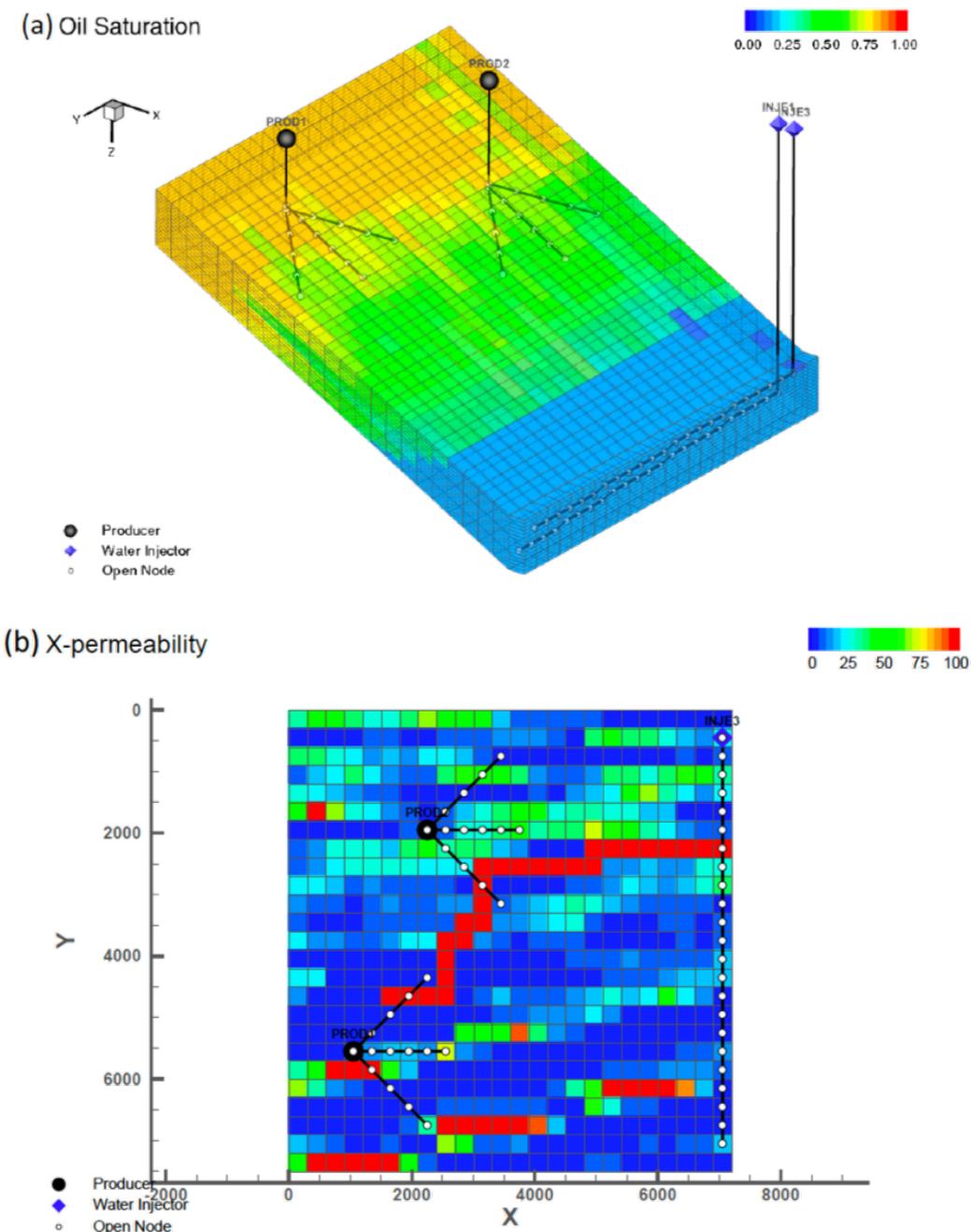


Figure 2.6: Case-D: Synthetic, three-dimensional, heterogeneous reservoir model with two forked trilateral producers and two horizontal injectors with a high-permeability path connecting injectors to specific producing laterals which induces early water breakthrough. Figure from [2].

seen in Fig. 2.7. This model was constructed with $69 \times 122 \times 14$ grid blocks of different sizes while rock and fluid properties and conditions, seen in Table 2.4, are determined based on many core samples and open hole log measurements, which indicated that permeability is mildly heterogeneous across the field. Though the average permeability is around 300 md, 1-2 darcy permeability zones are present around the field. The model is spatially isotropic in permeability, with 0.05 vertical-to-horizontal permeability ratio.

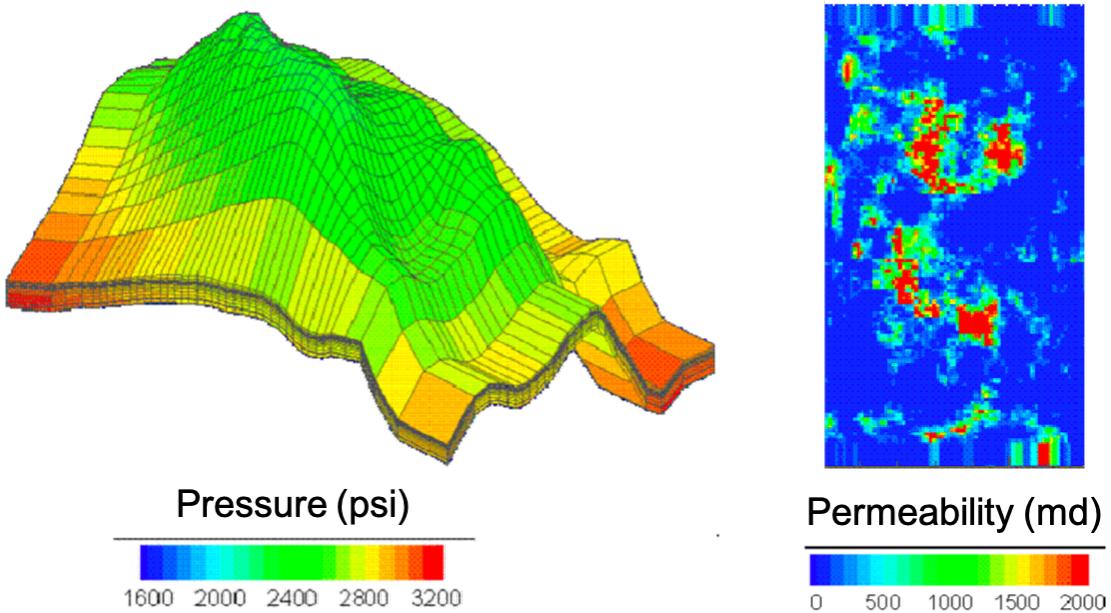


Figure 2.7: Case-E: Saturation and permeability maps of a real model located offshore of Saudi Arabia. Figure from [2].

Table 2.4: Case-E: Rock, fluid, and well properties at initial reservoir conditions.

Property	Value
Reservoir Size (ft^3)	$69 \times 121 \times 14$
Average Payzone Thickness (ft)	151
Average Porosity (%)	12.6
Average Horizontal Permeability (md)	300
Average Vertical Permeability (md)	15
Reservoir Pressure (psi)	1738

Literature shows studies where exact optimization on this real model was performed for well placement and configuration [1]. The evaluated designs included five well configurations: deviated, single-lateral, dual-lateral, trilateral, and quad-lateral. Findings indicate that trilateral configurations are optimal in this reservoir model [1]. Hence, we will examine both forked and fishbone trilateral configurations in this case. Each producer is equipped with three ICV devices, such that each is capable of eight fixed choke settings, where “0” and “7” indicate fully shut-in and fully open choke settings, respectively. For this reservoir model, the ICV devices are fully shut-in and fully open at restriction area values of 0 and 0.01 ft^2 , respectively. The production period for each combination of ICV settings was set to a single day, such that the surrogate model has to predict short-term production rates. This is analogous to building a surrogate to predict production flow rates across different ICV settings based on raw data of an existing well at the field.

2.2 Sampling Techniques⁵

In newly drilled intelligent multilateral wells, optimal ICV settings are traditionally chosen based on a few evaluative well performance tests with different choices of ICV setting combinations, ultimately to maximize oil/gas production and delay water/gas breakthrough. However, this optimization problem suffers from the curse of dimensionality, where each ICV device installed adds another dimension. Given n ICV devices installed in one or multiple wells with S discrete settings per device, the total possible ICV configuration combinations is $T = S^n$. Hence, the problem quickly grows dimensionally, implying that optimal sampling is critical. In other words, exponentially more sample points in the ICV setting space are required to provide consistent coverage as dimensionality increases.

We first define the term “ICV field test” as the evaluation of a smart well surface and downhole flow performance at a given combination of ICV configurations. Operationally, we only conduct a limited number of ICV field tests as the well needs to

⁵Available on GitHub: https://github.com/aljubrmj/MS_Thesis/tree/master/Surrogate_Utils_%26_Example.

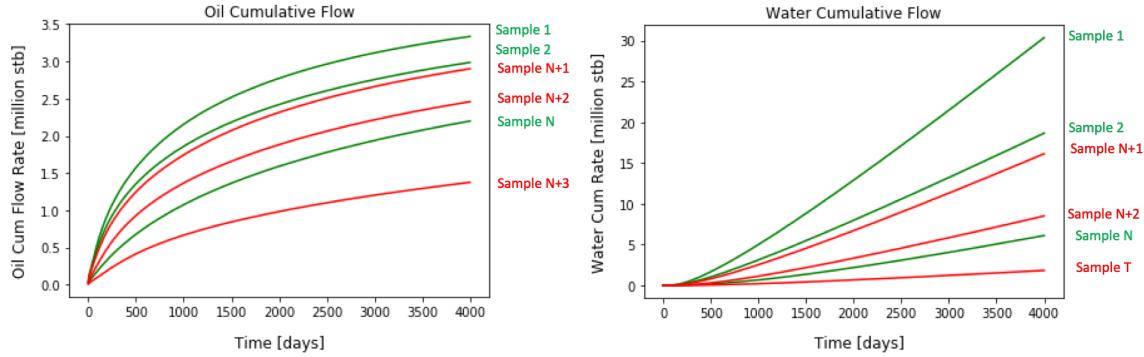


Figure 2.8: Oil and water cumulative flow rates generated using the Case-A reservoir model, where N profiles (green) are used to train and validate the machine learning surrogate while the remaining $T - N$ profiles (red) are used to evaluate its performance.

start production to meet corporate plans. Hence, we should carefully choose these test ICV configurations to be as representative of the well surface and downhole flow quantities as possible. Fig. 2.8 illustrates this challenge using the Case-A reservoir model as the ground truth, where only the surface oil and water flow rates are considered in this demonstration. Recall that Case-A involves a single trilateral producer with three ICV devices, each with 11 possible settings S ; hence, the total number of ICV settings combinations T is 1,331 in this case. Given N field tests, or sample points, we train and validate a machine learning surrogate to predict the well flow performance for the remaining or unseen combinations of ICV settings, $(T - N)$.

Given the importance of optimally sampling ICV field tests, this section will introduce two sampling techniques: space-filling Latin Hypercube Sampling, and adaptive sampling. These are compared to random sampling to illustrate their advantages and disadvantages.

2.2.1 Space-Filling Latin Hypercube Sampling (LHS)

LHS is a classical design of experiment technique based on statistical stratification [35, 59]. LHS is commonly used in Monte Carlo simulations to reduce variance given a fixed number of sample points in a multidimensional probability distribution [96]. A latin square is a square grid with samples positioned, such that each occupies a

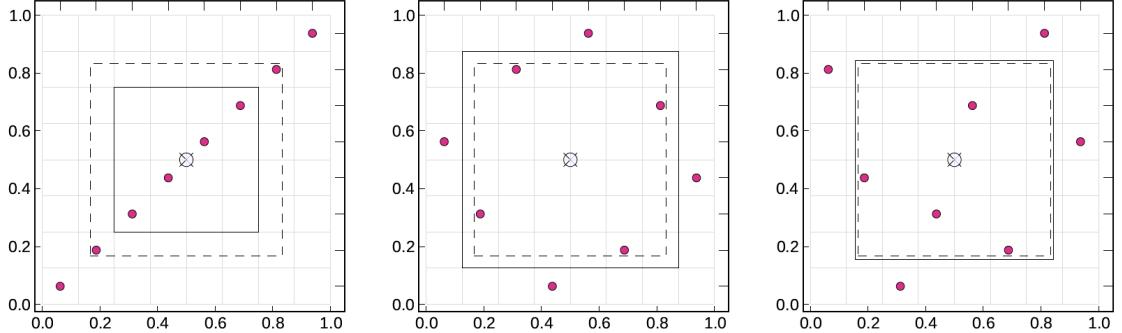


Figure 2.9: Three illustrative examples of LHS designs in two dimensions, where each row and column contains a single sample point. Figure from [92].

unique combination of rows and columns. Similarly, this concept is generalized to a latin hypercube, where sample points are positioned such that they occupy unique axis-aligned hyperplanes.

To sample N field tests, LHS is based on independently stratifying each dimension in the input space $\mathbb{D} \subset \mathbf{N}^n$ into N equi-possible intervals of probability $1/N$. For each input dimension of \mathbb{D} , one sample x_i is generated in each interval according to a selected joint probability function which yields N scalar samples (x_1, \dots, x_N) [58]. In this application, we randomly choose each scalar sample within its designated interval. The random matching of these scalar samples across \mathbb{D} results in multiple LHS realizations each with a set of sample points, W . This construction guarantees perfectly uniform reflections along any of the input dimensions, which is a desirable property for many applications [70]. However, Fig. 2.9 shows that the LHS property does not guarantee a space-filling design nor offer control over the distance to boundary [92].

Hence, optimization is required to find the most space-filling LHS design based on a chosen criterion. We use a greedy heuristic which starts with a single random point \vec{x}_i and sequentially adds optimal points from a random set until the LHS design is complete [20]. In each sampling iteration i , $f_{dup}(N - i)$ candidate points are nominated, where $f_{dup} \in \mathbf{N}$ is a duplication factor that is heuristically set to be less than ten. To be chosen, a sample point must have the LHS design property. Let us

define $d_{nn}(\vec{x}_i, W)$ as the Manhattan distance between the newly sampled point \vec{x}_i and its nearest neighbor in the existing point set W . In a problem with n -dimensional input domain \mathbb{D} , the optimization objective is to minimize $|d_{nn}(\vec{x}_i, W) - d_{opt}|$, the deviation from a heuristically optimal distance $d_{opt} = N/\sqrt[n]{N}$ to the existing points W . This greedy LHS approach yields space-filling and improved results compared to the conventional design [70]. However, LHS is not sequential and rather requires the determination of N apriori which subsequently influences the greedy search and optimization.

2.2.2 Adaptive Sampling

Adaptive sampling is a family of sampling techniques based on evaluating a function at optimal regions of its input domain \mathbb{D} to retrieve maximum information. The adaptive selection process depends on the behavior of the target function [14]. This characteristic allows for efficient sampling of functions that exhibit strong nonlinear variations within \mathbb{D} [94]. These techniques allocate a larger number of samples in subdomains with higher nonlinearity while regions with mild function variations are not sampled as frequently [19]. Adaptive sampling has been used in many applications, i.e. energy conservation and computational load control via optimizing sensor sampling frequency, surveying wildlife populations, efficient development of statistical surrogates, and others [11, 52, 14, 54].

Unlike LHS, adaptive learning does not require the knowledge of N and also considers the function evaluation $f(\vec{x}_i)$ at each sampled point \vec{x}_i . Hence, adaptive learners are useful in situations where the evaluation of $f(\vec{x}_i)$ is expensive, i.e. ICV field tests. While there exist multiple methods to sample adaptively, this study evaluated an adaptive learner based on triangulation. The method is structured to allow for sampling a function $f : \mathbb{D} \rightarrow \mathbb{G}$ with domain \mathbb{D} and range \mathbb{G} . This level of flexibility is required as we are interested in the evaluation of surface and downhole flow profiles of oil and water (output) with respect to the configurations of multiple ICV devices (input). This adaptive sampling algorithm consists of four major steps: initialization, Delaunay triangulation, simplex loss computation and subsequent sampling,

and Watson retriangulation.

Let us consider a smart multilateral prediction or optimization problem with n ICV devices. Define a function $f(\vec{x})$ that maps any combination of ICV settings $\vec{x} = (x_1, \dots, x_n)^T$ with domain \mathbb{D} , to M flow quantities in a point of time with range $\mathbb{G} \subset \mathbf{R}^M$. Sampling requires an initiation that is both suitable to the sampling algorithm and also beneficial to training a surrogate to learn the $f(\vec{x})$ hypersurface with the least number of sample points N_{opt} . This adaptive sampling algorithm starts by evaluating the input domain corner points (extremes). This is advantageous as it encourages interpolation and discourages extrapolation when using the machine learning surrogates for prediction and optimization. Note that the case where all ICV devices are closed, intuitively, outputs a zero vector of flow quantities. Next, an additional point \vec{x}_i is sampled closer to the centroid of \mathbb{D} . Given n ICV devices, this step requires the evaluation of 2^n sample points.

Next, we use Delaunay triangulation to construction simplices based on the evaluated 2^n initial points. Delaunay triangulation, equivalent to the connections of points in a Voronoi diagram, is unique if and only if every simplex circumcircle contains no points within its vicinity, seen in Fig. 2.10 [32, 49]. Assuming that we are evaluating a scalar flow quantity as an output, triangulation is limited to a single hypersurface with $n + 1$ dimensions. For instance, in cases with $n = 1, 2, 3$, the triangulation simplices take the form of *line*, *triangle*, *tetrahedral*, respectively. In case of M output flow quantities, we perform M Delaunay triangulations on hypersurfaces with $n + 1$ dimensions. Note that the resultant M triangulations are identical with respect to the input domain \mathbb{D} , but differ in terms of the output quantity range \mathbb{G} .

Using the M Delaunay triangulations, we decide on the next sample point \vec{x}_i that explains the most nonlinear subregion (simplex) of $f(\vec{x})$. Considering the j^{th} ($n + 1$)-dimensional triangulation for $j = 1, 2, \dots, M$, define $\vec{v}_{ij} = (\vec{x}_i, f_j(\vec{x}_i))^T, i = 1, 2, \dots, n+2$ to be $n + 2$ points which form a simplex $V(\vec{v})$. The simplex hypervolume $|V(\vec{v})|$ is then computed [79]. Next, we reduce the simplex hypervolumes to the sampling space $\mathbb{D} \subset \mathbf{N}^n$ using an arithmetic mean. As a result, the subsequent sample points is located at the centroid of the simplex with the largest volume, indicating the subregion of the largest variability in the objective function. The following equations

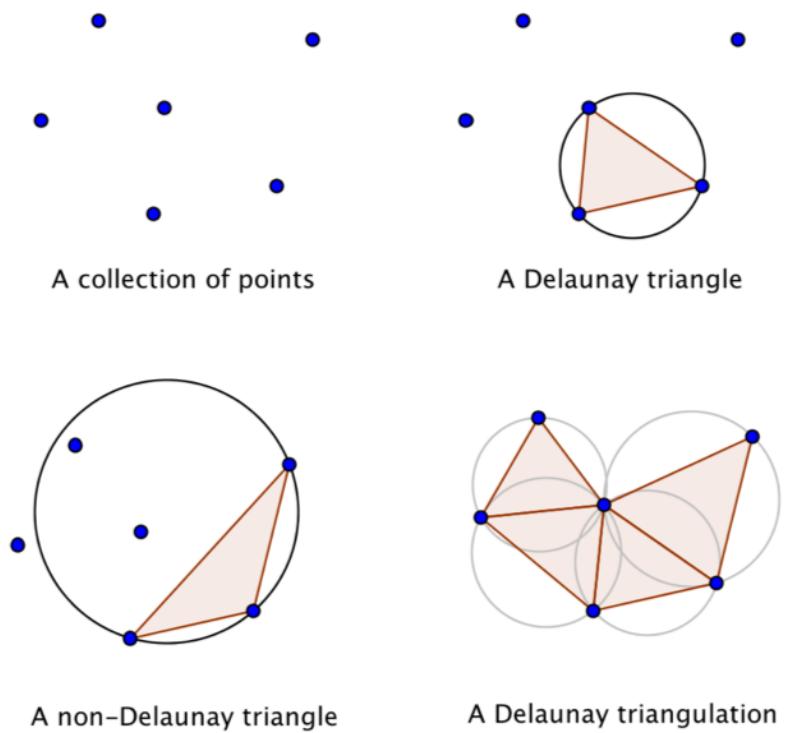


Figure 2.10: Two-dimensional demonstration of Delaunay triangulation [49].

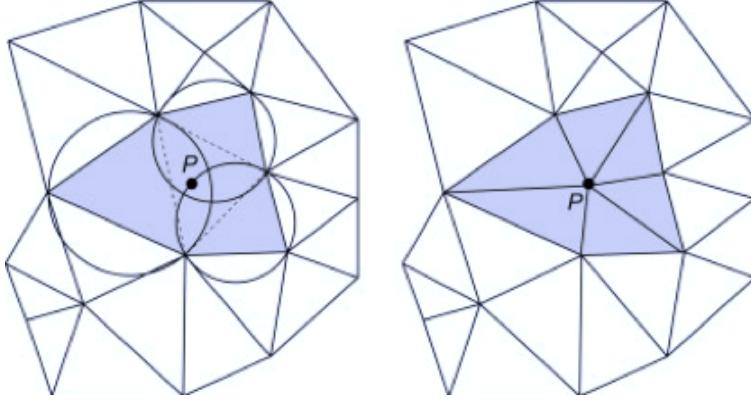


Figure 2.11: Demonstration of the Watson retriangulation algorithm where a new point \vec{x}_i is added to an existing Delaunay triangulation (left), which is then reconstructed using Watson algorithm (right).

demonstrate hypervolume calculation and reduction across triangulations:

$$|V(\vec{v}_j)| = \pm \frac{1}{(n+2)!} \begin{vmatrix} \vec{v}_{1j} & \dots & \vec{v}_{(n+2)j} \\ 1 & \dots & 1 \end{vmatrix} \quad (2.2)$$

$$|V(\vec{v})| = \frac{1}{M} \sum_{j=1}^M |V(\vec{v}_{ij})| \quad (2.3)$$

After introducing a new point based on the simplex hypervolumes, the preexisting triangulation is reconstructed using Watson retriangulation [91]. As seen in Fig. 2.11, this procedure involves the following steps: a) locate newly inserted point, b) identify all simplices whose circumcircles contain this point, c) delete these simplices, and d) form new simplices by connecting the vertices of the now empty convex cavity to the newly added point [22].

The sampling and retriangulation process is repeated until a certain criterion is met. There are multiple criteria that oil and gas operators could use to stop ICV field tests and proceed with production to meet corporate targets. These criteria could involve a budget which sets a maximum allowable number of days or ICV field tests for this operation. In addition, the adaptive sampling approach actually introduces a practical criterion, the average volume of all simplices at any sampling

step. This quantity decreases continuously with a plateauing gradient, indicating the diminishing return on further ICV field tests. Hence, the quantity can be used as a stoppage criterion as we will demonstrate in the upcoming chapters.

2.3 Machine/Deep Learning Algorithms

The application of artificial intelligence and machine learning has proliferated rapidly over the past few years and witnessed significant advancement and use cases across many scientific fields. A widely referenced description of machine learning is given by Tom Mitchell (1997): “A computer program is said to learn from an *experience* with respect to some *task* and *performance measure*, if its *performance* on the said *task*, as per the said *measure*, improves with that *experience*” [60]. In our problem, the task is to predict multilateral well flow performance for unseen ICV setting combinations while the experience is the seen well flow performance at given ICV settings. Meanwhile, R^2 is used as the performance measure for the algorithm to learn the correlation between the inputs and outputs.

2.3.1 Feedforward Neural Networks⁶

There exist many machine learning algorithms that can map the problem input to output in a regression setting where the outputs are continuous quantities. Feed-forward neural networks represent a foundation for many other architectures which target unstructured data, i.e. images, audio recordings, text, etc. Neural networks were loosely inspired by the human brain, incorporating complexity, nonlinearity, and architectural parallelism [46]. A neural network is an organized set of nodes, called neurons, complimented with nonlinear activation functions that are combined across hidden layers in a mathematically linear setup to map out a given input into a desired output. A neural network graph is characterized by breadth (neurons per layer) and depth (total number of layers).

⁶Available on GitHub: https://github.com/aljubrmj/MS_Thesis/tree/master/Surrogate_Utils_%26_Example.

Even though each node conveys nonlinearity, the literature shows that deeper networks display more representational power than their shallower counterparts, which gave the rise to a machine learning subdomain, called deep learning. Hence, each node at each layer trains on a different set of features which are numerical representations of the input variables and allow for capturing more complex correlations as they are aggregated depth-wise. This is known as feature hierarchy which describes the complex abstraction of deep networks that enable algorithms to draw correlations and insights from large- and high-dimensional data [82].

Neural networks accept an input vector (training example) and sequentially perform a set of algebraic computations complemented with nonlinear activation functions to estimate the output quantities. While there exist a myriad of such algebraic architectures, we investigated a $100 \times 100 \times 100 \times 100$ (four layers with 100 neurons each) fully-connected neural network to predict the well performance (note that the input and output vector sizes differ based on the problem configuration). Fig. 2.12 shows the most common activation functions. We adopted the activation function Rectified Linear Unit (ReLU) across all hidden layers as it is computationally inexpensive, and does not saturate at the tails which alleviates the vanishing gradient problem where a network stops learning due to small gradient during optimization [93]. Note that the output layer is always activated with a purely linear identity function.

Forward and backward propagation are the major two steps in neural network training. A forward pass means performing algebraic operations from input to map out outputs while a backward pass involves computing and propagating the weight gradients with respect to the mean-squared error (MSE) cost function [68]. Eq. 2.4 shows the network forward propagation for layer l , where Θ is the weight matrix, \vec{b} is the bias vector, \vec{z} is the linear output, and \vec{a} is the activated output of layer l . Eq. 2.5 demonstrates the application of ReLU as means of nonlinearity. At the start of model training, the bias vector is initiated to zeros while the network weights are initiated with zero mean and variance seen in Eq. 2.6 [42]. We aim to minimize the regularized MSE cost function seen in Eq. 2.7.

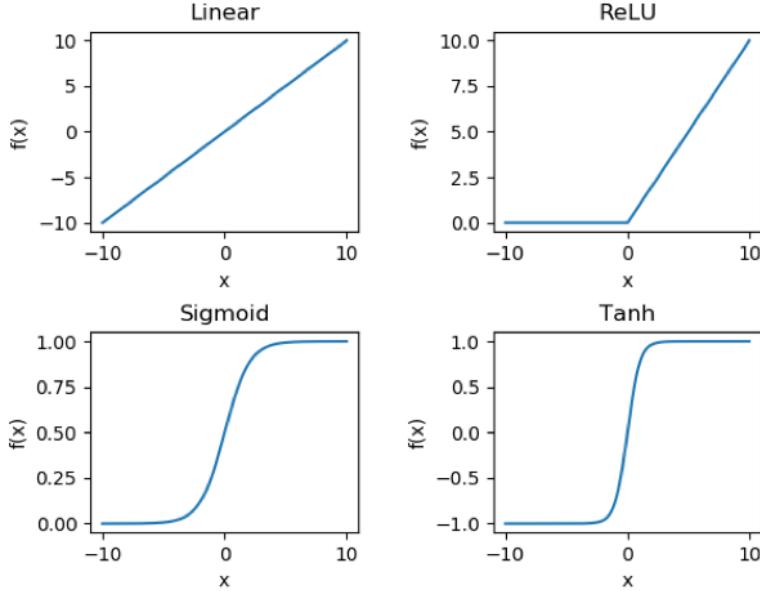


Figure 2.12: Most commonly used activation functions that introduce nonlinearity to the neural network. Because this is a regression problem, we used ReLU across the hidden layers as it is not prone to the vanishing gradient problems when compared to sigmoid and hyperbolic tanh. Note that the final output layer is always activated with the linear function.

$$\vec{z}^{[l]} = \Theta^{[l]T} \vec{a}^{[l-1]} + \vec{b}^{[l]} \quad (2.4)$$

$$\vec{a}^{[l]} = \text{ReLU}(\vec{z}^{[l]}) = \max(0, \vec{z}^{[l]}) \quad (2.5)$$

$$\text{Var}(\Theta^{[l]}) = \frac{2}{(n_{in}^{[l]} + n_{out}^{[l]})} \quad (2.6)$$

$$J(\Theta) = \frac{1}{2K} \sum_{i=1}^K \sum_{j=1}^M (y_j^{(i)} - \hat{y}_j^{(i)})^2 + \frac{\lambda}{2K} \sum_{l=1}^{L-1} \sum_{i=1}^{S_l} \sum_{j=1}^{S_{l+1}} (\Theta_{j,i}^{(l)})^2 \quad (2.7)$$

where n_{in} is the number of input neurons, n_{out} is the number of output neurons, J is the cost function, $y_j^{(i)}$ is output value of quantity j in the i^{th} training example, $\hat{y}_j^{(i)}$ is the neural network model prediction of quantity j in the i^{th} training example, K is the number of training examples, M is the number of output variables, λ is the regularization term (L2 weight decay), L is the number of layers, S_l is the number of

neurons in the l^{th} layer.

Note that the MSE cost function is usually nonconvex in the presence of the non-linear ReLU activation function. Literature offers multiple optimization algorithms, yet the Adam optimizer proved to be superior in most tasks [55]. Model training is conducted in minibatches which is generally advantageous compared to stochastic and fullbatch approaches. Minibatch requires less memory during training and adds moderate noise which is ideal for nonconvex optimization as it helps escape saddle points and local minima [39]. Hence, we trained the model with batch size of 200 and 200 epochs (Adam gradient descent optimization steps).

Bias-variance tradeoff is an important aspect as too-simple neural networks can underfit while too complex neural networks can overfit the training dataset. Hence, we chose a neural network to provide sufficient complexity to fit the dataset at hand, and further performed two forms of regularization to prevent overfitting. We used L2 regularization which is represented in the second term in Eq. 2.7, and also adopted three-fold cross validation. Tuning the model hyperparameters was done by varying the L2 regularization coefficient λ along an interval of $[0, 0.001]$.

2.4 Stochastic Optimization

While field production data is certain and require deterministic optimization, reservoir models are geologically uncertain and require stochastic treatment. These models are commonly used to plan well placement, configuration, production schedule, etc. Uncertainty plays a major role in corporate, engineering, and operational decision making. Operators cannot make decisions to sequentially evaluate uncertainty, but rather make integrated and sound decisions based on incomplete information. In other words, a decision can be defined as a conscious, irrevocable allocation of resources that aims to optimize an objective of interest [50].

Numerical reservoir simulation allows for the generation of multiple reservoir models (realizations) to evaluate and characterize uncertainty in field data (seismic, well integrity, etc.) and subsurface settings (rock properties, fluid properties, etc.). If a

sufficiently large number of reservoir simulations is performed, uncertainty quantification allows for computing percentile statistics (P10, P50, and P90) of the flow profiles. However, these flow simulations are computationally expensive. Subsequently, operators must choose a smaller subset of simulations to be performed for that matter. Hence, we discuss a distance-kernel method (DKM) which aims to reduce the computational expense of stochastic optimization while maintaining the representation of the field measurements. In addition, decision making under uncertainty is influenced by both the physical uncertainty and also decision makers attitudes towards risk. Thus, we discuss an economic utility framework which aims to maximize the expected utility of a decision with respect to the decision makers risk aversion.

2.4.1 Distance-Kernel Method (DKM)

DKM aims to select the most representative realizations to cover the spread of uncertainty with respect to specific response quantities. DKM nominates only a few such realizations that accurately capture the response percentile statistics (P10, P50, and P90). DKM is based on dissimilarity of realizations based on their corresponding response flow profiles. The most representative realizations are chosen, such that they have the same percentile statistics as the original full set of realizations. As seen in Fig. 2.13, the DKM approach consists of five main steps: a) calculation of distance between realization simulation responses, b) construction of distance matrix, c) mapping realization responses to Euclidean space, d) projection to feature space, e) preimage construction and percentile statistics (P10, P50, and P90)[74].

Consider N_R realizations generated and simulated based on a reservoir model with geological uncertainty in one or more parameters. Steps (a) and (b) of the DKM approach involve the construction of an $N_R \times N_R$ dissimilarity matrix using the simulation flow responses. For each two realizations i and j , this matrix is based on a distance metric δ_{ij} that describes the similarity between these realizations. This matrix could be based on all or one desired flow response. There exist multiple distance metrics commonly used for reservoir and petrophysical models, i.e. flow-based distance [72, 73], time-of-flight-based distance [63], and Hausdorff distance

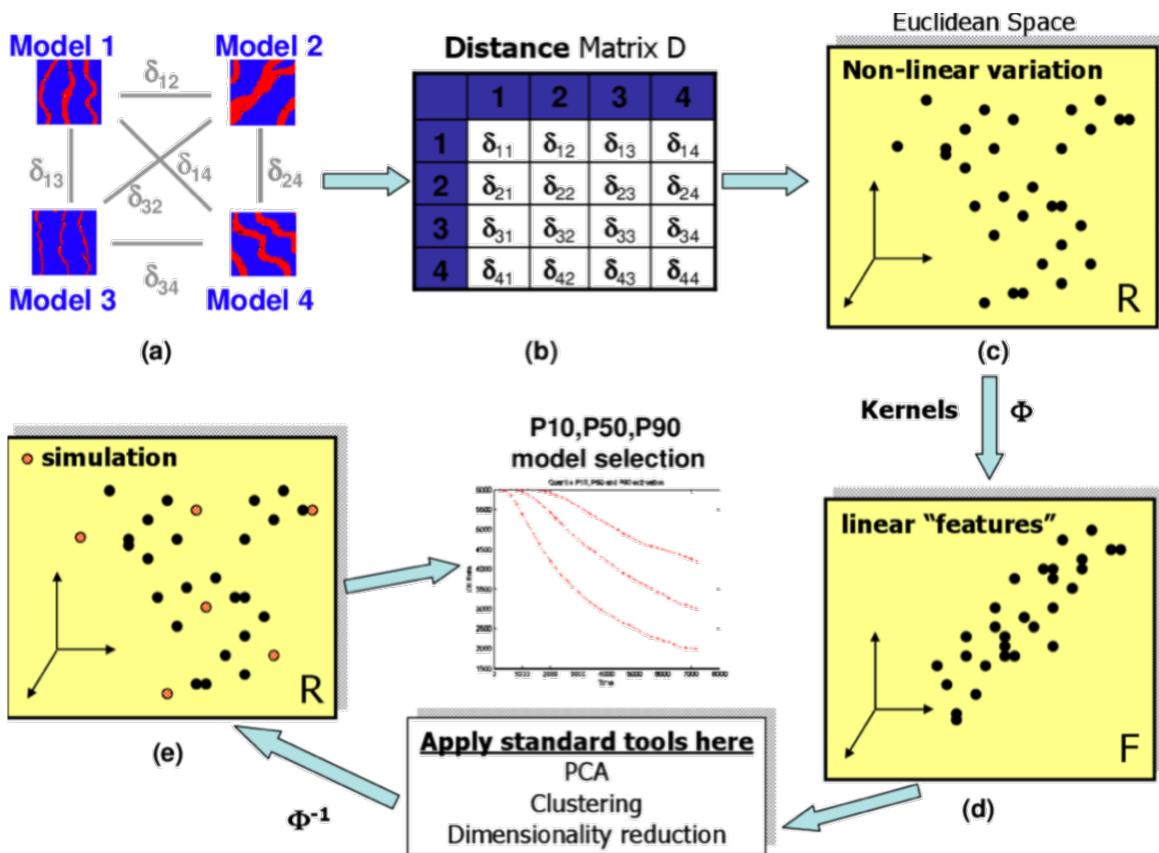


Figure 2.13: DKM approach: a) calculation of distance between realization simulation responses, b) construction of distance matrix, c) mapping realization responses to Euclidean space, d) projection to feature space, e) preimage construction and percentile statistics (P10, P50, and P90). Figure from [74].

[81]. In this application, we use the Hausdorff distance to construct the dissimilarity matrix. For any two sets A and B with points a and b , respectively, whose similarity is measured using the Euclidean distance $d(a, b)$, the Hausdorff distance $H(A, B)$ can be defined as follows:

$$H(A, B) = \max\{h(A, B), h(B, A)\} \quad (2.8)$$

$$h(A, B) = \max_{a \in A} \{\min_{b \in B} \{d(a, b)\}\} \quad (2.9)$$

In step (c), the dissimilarity matrix is then mapped into an α -dimensional Euclidean space \mathbf{R}^α using multidimensional scaling (MDS) [23]. MDS translates each realization flow response into a point in \mathbf{R}^α (typically small values of $\alpha = 2, 3, \dots, 10$). MDS aims to represent these points such that their Euclidean distance is closely representative of the correlational dissimilarity between realizations as established in step (b). We first construct a matrix $A : a_{ij} = -\frac{1}{2}\delta_{ij}^2$. Using the centering matrix $C = \mathbf{I} - \frac{1}{\alpha}\mathbf{1}^T\mathbf{1}$, construct a Gram matrix $B = CAC$. Extract the p largest positive eigenvalues Λ_p and eigenvectors E_p of matrix B . The p -dimensional Euclidean space configuration of the N_R realizations is then derived from the coordinate matrix $X = E_p\Lambda_p^{1/2}$. Using Euclidean distances is advantageous as MDS provides an efficient analytical solution, requiring no iterations. Effectively, the dissimilarity matrix is conditioned to be positive-definite; however, negative eigenvalues of small magnitudes can be ignored. In addition, it is important to note that the resultant realization coordinates in this p -dimensional Euclidean space are relative, but not absolute.

While the selection of most representative realizations is possible at this point, it is suboptimal because the standard component reduction and clustering techniques assume linear structures in the nonlinear Euclidean space \mathbf{R}^α . In step (d) of Fig. 2.13, we apply kernel transformation to make the relationship between these points more linear in a new feature space \mathbb{F} [75]. One advantage of kernel methods is that there is no need to explicitly map points from \mathbb{F} to \mathbf{R}^α , where computations can be performed using a high-dimensional transformation function $\Phi : \mathbf{R}^\alpha \rightarrow \mathbb{F}$. As seen in Eq. 2.10, the kernel function k can be defined for any two vectors \vec{u} and \vec{w} . We use the Gaussian kernel, seen in Eq. 2.11, with kernel width σ of 20% of the range of

distances between realization responses. This value is recommended in the literature where it usually shows robust results [78, 74].

$$k(\vec{u}, \vec{w}) = \langle \Phi(\vec{u}), \Phi(\vec{w}) \rangle \quad (2.10)$$

$$k(\vec{u}, \vec{w}) = \exp\left(\frac{\|\vec{u} - \vec{w}\|^2}{2\sigma^2}\right) \quad (2.11)$$

Step (e) involves the application of the classical k -medoids clustering algorithm in the feature space \mathbb{F} . This algorithm requires the specification of number of clusters k desired as an output. While the algorithm is similar to k -means, k -medoids computes cluster centroids based on preexisting realization response points which can be used directly as the most representative realizations. The k -medoids algorithm assigns points each to a cluster Ψ_i by minimizing the mean squared distance J between points and their corresponding cluster median, $\tilde{\mu}_i$:

$$J = \sum_{i=1}^k \sum_{\vec{x}_j \in \Psi_i} |\vec{x}_j - \tilde{\mu}_i|^2 \quad (2.12)$$

The k -medoids algorithm is a partitioning technique that initially randomly splits points into k clusters Ψ_i for $i = 1, 2, 3, \dots, k$. Then, the algorithm determines the median $\tilde{\mu}_i$ of each cluster Ψ_i . Each point is then assigned to the closest cluster. The median determination and cluster assignments steps are then performed repeatedly until convergence. Convergence is when points stop shifting from one cluster to another and cluster medians Ψ_i do not change. Note that clustering occurs in the high-dimensional feature space \mathbb{F} , where the Euclidean distances can be calculated using the inner product as follows:

$$\|\Phi(\vec{u}) - \Phi(\vec{w})\|^2 = k(\vec{u}, \vec{u}) + k(\vec{w}, \vec{w}) - 2k(\vec{u}, \vec{w}) \quad (2.13)$$

Finally, the cluster medians $\tilde{\mu}_i$ are chosen as the most representative subset to allow for uncertainty quantification. Evaluation of the result is possible by comparing the percentile statistics (P10, P50, and P90) of the most representative subset versus those of the full realization set.

2.4.2 Utility Framework

Given a reservoir model under geological uncertainty with a smart multilateral well equipped with ICV devices, the decision analysis framework is concerned with the selection of the optimal ICV settings based on a set of geological realizations. As seen in Fig. 2.14, each decision of ICV configuration leads to multiple possibilities, or events, each with probability of occurrence P_i and reward NPV_i . Based on the decision maker's level of risk aversion, each event is assigned a utility function, U_i , which reflects their degree of satisfaction. This utility function depends on the magnitude and variability of NPV_i across all events subsequent to each decision. In other words, the utility framework is based on the maximization of the expected utility value rather than the monetary value of a decision [44].

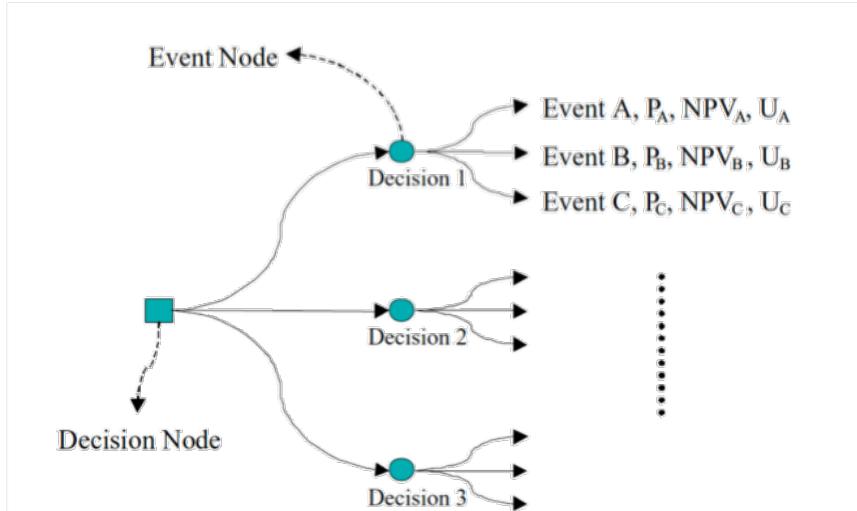


Figure 2.14: The ICV configuration decision tree. Figure from [44].

Based on the investigation of this study, the expected value was found to be a suitable statistical measure to evaluate the expected utility of each decision based on its events N_{event} , seen in Eq. 2.14. As seen in Fig. 2.15, each decision is associated with a single expected utility value. The optimization aims to find the decision with the maximum utility value.

$$E\{U(\text{decision})\} = \sum_{i=1}^{N_{\text{event}}} U_i P_i \quad (2.14)$$

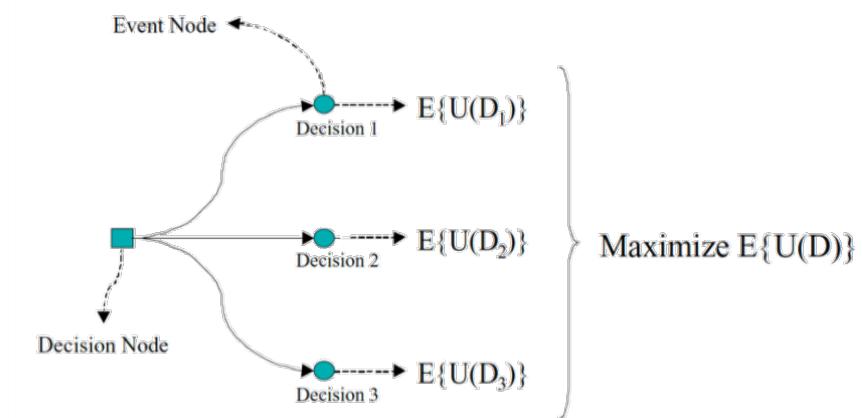


Figure 2.15: The ICV configuration decision tree with expected utility value for each decision. Figure from [44].

Next, we define the utility function which is the means to quantify the decision maker's attitude towards risk [69]. Decision makers can be either risk-averse, risk-neutral, or risk-prone. A risk-averse person tends to avoid decisions characterized with uncertainty despite the possibly high monetary return. A risk-neutral person makes decisions purely based on NPV. A risk-prone person is more of a gambler who leans towards making uncertainty decisions in sake of higher monetary rewards [48]. This work will utilize a classical analytical exponential function, seen in Eq. 2.15, where r is the risk aversion coefficient which represents the decision maker's level of risk aversion. Fig. 2.16 shows the behavior of the exponential utility function depending on r sign and magnitude.

$$U(x) = \begin{cases} \frac{1-e^{-rx}}{r} & r \neq 0 \\ x & r = 0 \end{cases} \quad (2.15)$$

For the utility theory to be used, five axioms must apply to the decision maker. These are orderability of outcomes, continuity, substitutability, monotonicity, and

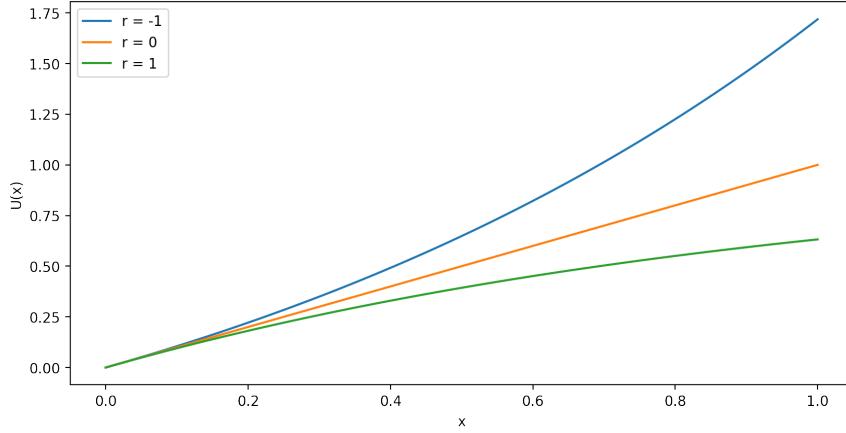


Figure 2.16: Effect of risk aversion factor of the shape of the exponential utility function.

decomposability [88]. Only if applicable, the consequent decision with the maximum expected utility value can be used to describe a decision maker’s attitude towards risk.

2.5 Optimization Solvers

While optimization by enumeration eliminates errors stemming out of common optimization solvers, it becomes infeasible to optimize over surrogates as the input domain grows large. Hence, it is necessary to evaluate common optimization solvers on the trained machine learning surrogates. However, this study was not concerned about the evaluation of optimization solvers; hence, this particular discussion will be brief.

2.5.1 Mesh Adaptive Direct Search (MADS)

This study explored the use of MADS over the trained surrogates which are treated as a blackbox. MADS is a derivative-free, direct-search algorithm which is based on the iteration of a series of meshes with various sizes and with controlled refinement. These meshes represent a discretization of the input domain. Each MADS iteration aims to find a point on the current mesh that improves the current solution. If

this is infeasible, a finer mesh is constructed [17]. Because the input variables in this study are discrete, it is important to consider integer programming. Hence, we adopt NOMAD which offers an implementation of MADS modified for granular and discrete variables [18]. While there are multiple hyperparameters that control the mesh search and refinement along with other properties, we will only demonstrate MADS optimization results based on the default parameters.

2.6 Summary

This chapter introduced several algorithms utilized in this study. The chapter covered reservoir models, sampling, machine learning, and optimization techniques. Five numerical reservoir models were highlighted, including a real field case. Sampling techniques involved LHS and adaptive sampling. Feedforward neural networks were highlighted as the machine learning algorithm of choice. Meanwhile, optimization included both deterministic and stochastic optimization. This involved optimization of exact reservoir models or data and also optimization under uncertainty. The next chapter will highlight and discuss the results of these techniques on the prediction and optimization of ICV device in trilateral wells.

Chapter 3

Results and Discussion

In this chapter, we will describe the performance of the proposed prediction and optimization algorithms in both deterministic and stochastic settings. We firstly focused on the prediction task and verified the algorithm precision on the different reservoir models. In addition, various levels of Gaussian and uniform noise were introduced to further verify the effectiveness of the proposed algorithm. Secondly, deterministic optimization was demonstrated on the real reservoir model, Case-E. Thirdly, the effectiveness of the algorithm in stochastic optimization using the utility framework was investigated.

3.1 Prediction Task

As seen in Fig. 2.1, the first step in the proposed algorithm is to sample the desired number of field tests, N . Fig. 3.1 demonstrates sampling results for N values of five (left), 20 (middle), and 30 (right). For each cube, sampling is repeated from scratch. The trivial sample point $[0, 0, 0]$, where all three ICV devices ($n = 3$) are shut-in, is included by default without being counted into N . While increasing N allows for capturing more subregions of the input space, it is economically infeasible. We visually observe that random sampling is suboptimal and highly stochastic upon repetition. LHS is space-filling and less stochastic compared to random sampling. Meanwhile, the proposed adaptive sampling algorithm requires the sampling of all input space

corners, eight samples in Case-A. While this is superior algorithmically, it requires exponentially more corner points with increasing numbers of installed ICV devices. Hence, we also examined the use of adaptive sampling where we heuristically sample $n + 1$ corners only. This techniques results in tetrahedrons along triangular hyperplanes defined by the selected corner points; hence, the machine learning surrogate has to extrapolate to predict the abandoned subregions.

Next, we used the sampled points of each technique to perform ICV field tests while the remaining points in $\mathbb{D} \subset \mathbb{N}^n$ are used for testing. We first used Case-A to simulate the ICV field tests and investigate the performance of the proposed feedforward neural network with respect to each sampling technique. Note that we focused on cumulative flow quantities as the proposed machine learning algorithm was found to be superior in predicting monotonically increasing quantities. For each N value of each sampling technique, we sampled and trained machine learning models ten times to account for stochasticity due to sampling and machine learning model weight, Θ , initialization. This is important to account for uncertainty in model prediction. Each model was then tested on the unseen ICV settings to compute R^2 and evaluate its performance in the prediction task. We first considered the prediction of surface quantities only, oil and water cumulative flow rates. Practically, prediction of surface quantities is sufficient to achieve the prediction and optimization tasks, although it does not provide knowledge of downhole flow quantities. Fig. 3.2 shows the surrogate R^2 value (bold line) and uncertainty (transparent band) test results on Case-A. Except for eight-corner adaptive sampling, all sampling algorithms require four and six ICV field tests to achieve R^2 of over 80% and 90%, respectively. Meanwhile, eight-corner adaptive sampling requires eight ICV field tests to achieve similar results as the first seven ICV field tests to cover the corners (recall that the trivial case where all ICV devices are shut-in is discarded). Minor differences between random and LHS sampling were observed in this case. Generally, uncertainty decreases with increasing ICV field tests. In addition, adaptive sampling techniques result in lower uncertainty compared to the other sampling methods.

Similarly, we can evaluate the proposed algorithm in the prediction of both surface (oil and water) and downhole (total liquid at each ICV device) flow rates. Note that

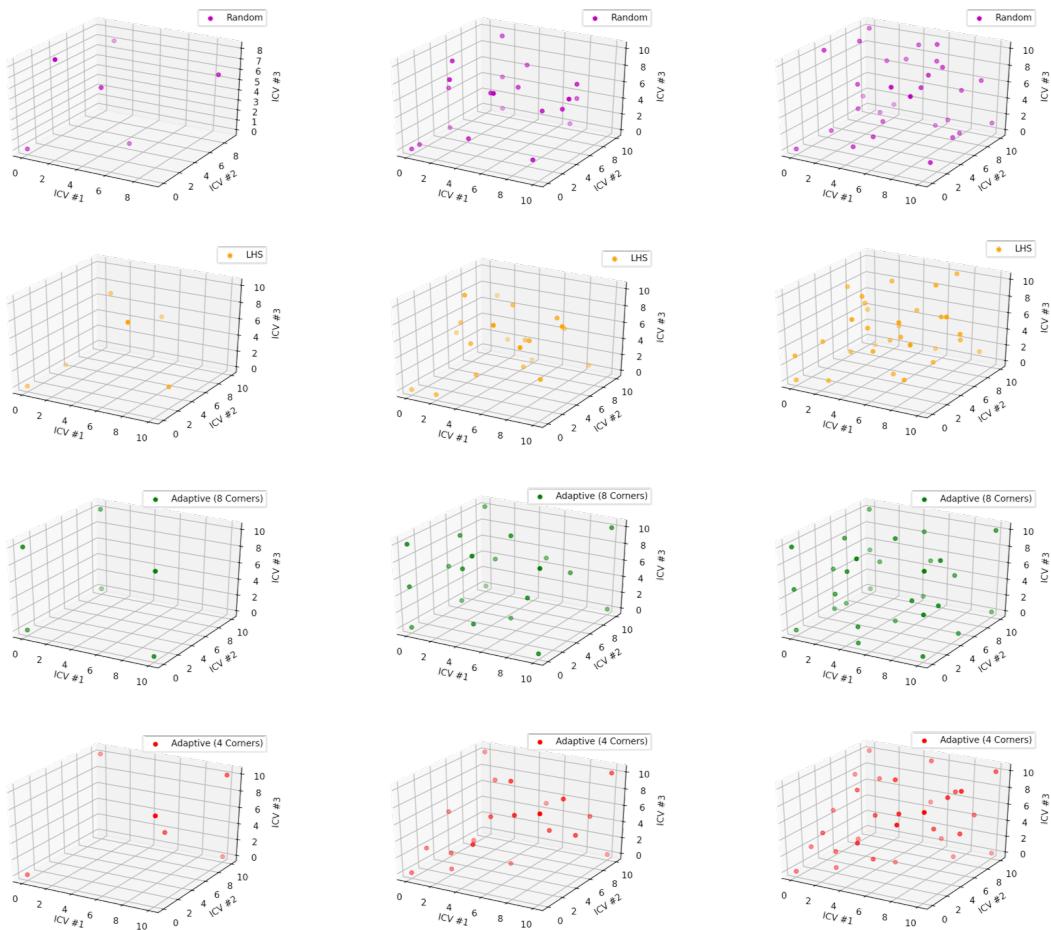


Figure 3.1: Comparison between four sampling techniques: random sampling, LHS, adaptive sampling with four corners, and adaptive sampling with all cube corners (eight corners). Left, middle, and right columns refer to sampling of five, 20, and 30 field tests in total.

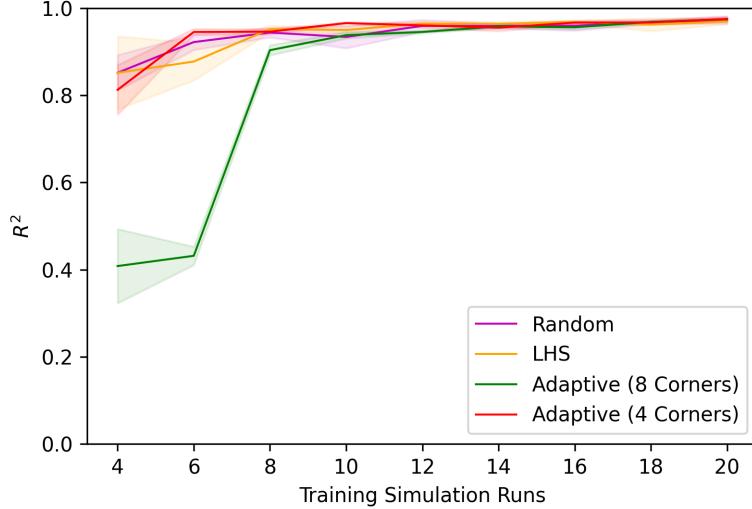


Figure 3.2: R^2 test performance of the machine learning surrogate in the prediction of the Case-A surface flow quantities with respect to the different sampling techniques.

we only predict total liquid downhole flow rates without differentiating oil and water rates as the state-of-the-art technology does not measure such quantities. Fig. 3.3 shows that the prediction of both surface and downhole flow quantities is a more challenging task than the prediction of only surface flow quantities. Note that negative R^2 values indicate that the machine learning surrogate is worse than predicting the mean. In this case, it is evident that adaptive sampling is ultimately superior to random and LHS sampling. It is important to note that four-corner adaptive sampling outperforms eight-corner adaptive sampling, especially at limited numbers of ICV field tests. However, four-corner adaptive sampling R^2 plateaus early as it omits several subregions where it has to extrapolate. Overall, four-corner adaptive sampling requires nearly nine and 14 ICV field tests to achieve 80% and 90% R^2 , respectively.

While it is theoretically reassuring to use the unseen ICV field test flow profiles to evaluate the surrogate performance, it is impractical. The proposed adaptive sampling offers a practical alternative using Eq. 2.3. Fig. 3.4 shows the average simplex hypervolume (green) and corresponding gradient (red) using second-order central difference in the interior, and forward and backward differences at the ends.

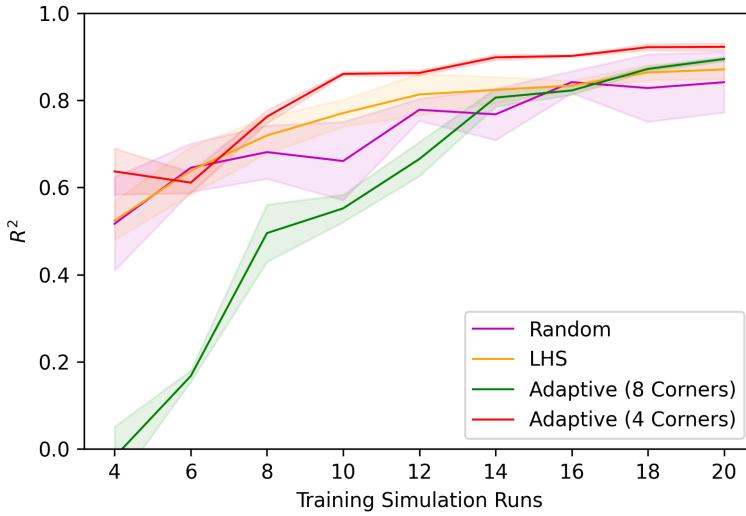


Figure 3.3: R^2 test performance of the machine learning surrogate in the prediction of the Case-A surface and downhole flow quantities with respect to the different sampling techniques.

These are generated based on the application of the four-corner adaptive sampling technique on Case-A. While the average simplex hypervolume requires picking the elbow, we can directly approximate the point of diminishing return (insignificant return upon further sampling) using the gradient quantity. This closely resembles the results of Fig. 3.3 where 80% and 90% R^2 values were achieved with only nine and 14 ICV field tests. Hence, we conclude that the proposed adaptive sampling technique offers robust prediction results and further enables oil and gas operators to decide when to stop ICV field tests and set wells onto production with an accurate model at hand.

Case-B was introduced as a heterogeneous replica of the Case-A reservoir to evaluate the proposed algorithm. As seen in Fig. 3.5, the adaptive sampling algorithm records improved results after adding heterogeneity. While the point of diminishing return is approximately at 14 ICV field tests again, the four-corner adaptive sampling achieves 80% and 90% R^2 with only four and ten ICV field tests, respectively, in Case-B compared to nine and 14 in Case-A. This is a result of the optimization behavior of the neural network where the induced heterogeneity allowed for ICV-2 to yield

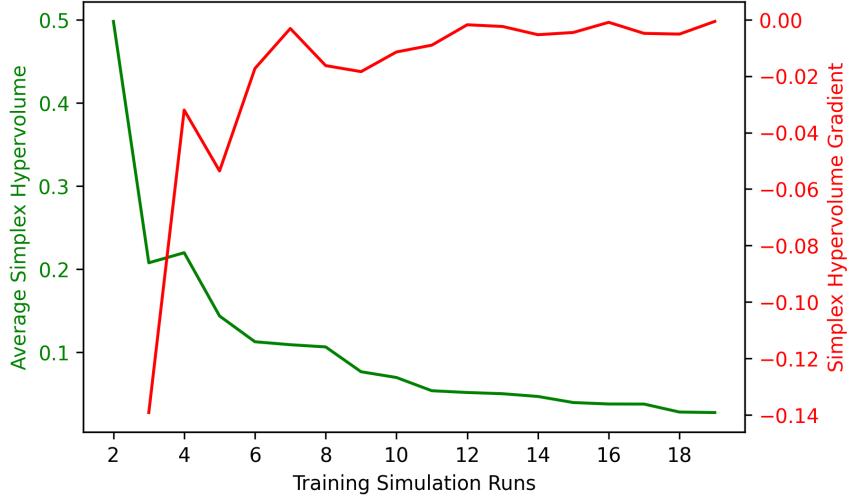


Figure 3.4: Average simplex hypervolume and gradient as a function of the number of ICV field tests using four-corner adaptive sampling to predict both surface and downhole flow quantities of Case-A.

higher flow rate as it is choking the lateral across the advantageous reservoir region while ICV-3 yielded the lowest flow rates as it is choking the lateral within the lowest permeability zones. Asymmetry reshapes the cost function hypersurface resulting in prominent minima, and thus an easier optimization process. On the other hand, homogeneity results in symmetry and a consistently bumpy cost function hypersurface across the decision variable domain with less prominent minima, and hence a harder optimization problem.

Similarly, the algorithm was also applied to Case-C which involves a complex fluvial channel system completed with a five-spot pattern, with a conductive path between injectors and some of the producing laterals. Applying the surrogate-based algorithm, prediction performance on Case-C is shown in Fig. 3.6. We observe that the geological complexity of this model induces flow profiles which are harder for the proposed algorithm to capture. Again, the four-corner adaptive sampling technique is superior at smaller numbers of ICV field tests. However, we observe that using eight-corner adaptive sampling results in ultimately superior performance as it starts by capturing all corner points, leaving less margin for extrapolation. Based on the four-corner adaptive sampling method, the algorithm requires 15 and 27 ICV field

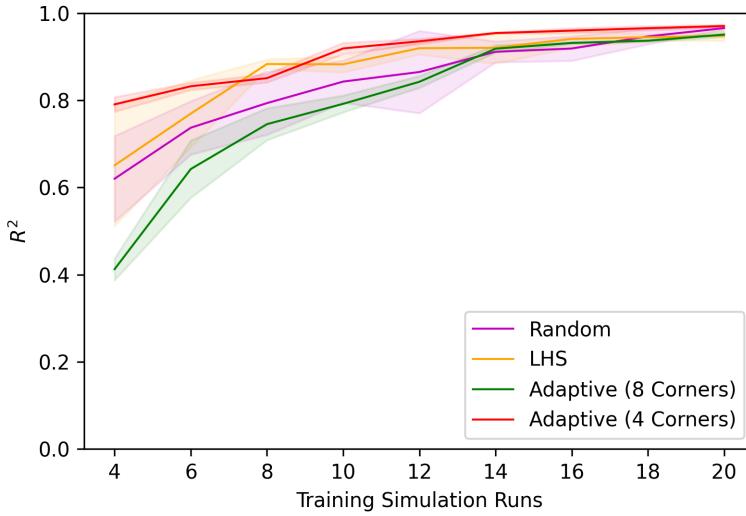


Figure 3.5: R^2 test performance of the machine learning surrogate in the prediction of the Case-B surface and downhole flow quantities with respect to the different sampling techniques.

tests to achieve R^2 of 80% and 90%, respectively.

Case-D is based on the SPE-9 reservoir model with two forked producers and two horizontal injectors, such that injectors connect to some of the producing laterals with a highly permeable path. We first investigated the scenario where one producer has no ICV devices while the other has three ICV devices. This scenario allows for exploring the effect of well interference on the proposed algorithm. Applying the surrogate-based algorithm, prediction performance on Case-D, given that only Producer-1 is equipped with ICV devices, is shown in Fig. 3.7. The results confirm the superiority of four-corner and eight-corner adaptive sampling techniques given relatively small and large numbers of ICV field tests, respectively. Using the four-corner adaptive sampling method, the algorithm requires four and eight ICV field tests to achieve R^2 of 80% and 90%, respectively. Similarly, prediction performance on Case-D, given that only Producer-2 is equipped with ICV devices, is shown in Fig. 3.8. The results are very close in absolute and relative trends across the different sampling methods.

Using Case-D, let us consider the task of surface and downhole ICV flow performance of both producers simultaneously. In this scenario, each producer is equipped

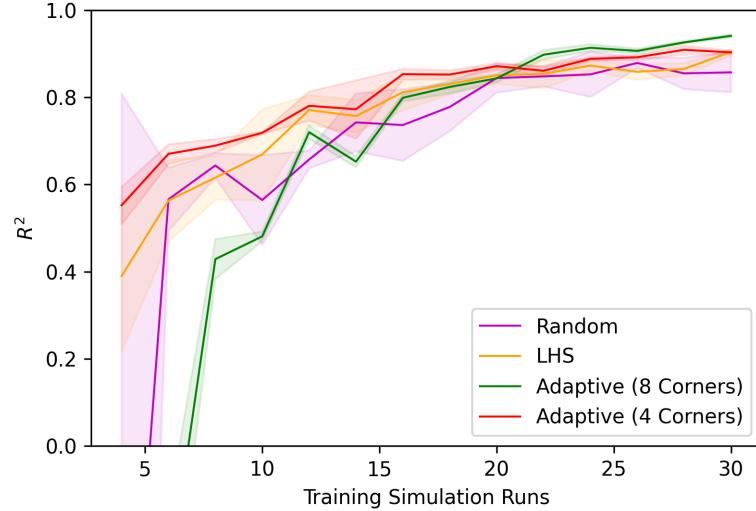


Figure 3.6: R^2 test performance of the machine learning surrogate in the prediction of the Case-C surface and downhole flow quantities with respect to the different sampling techniques.

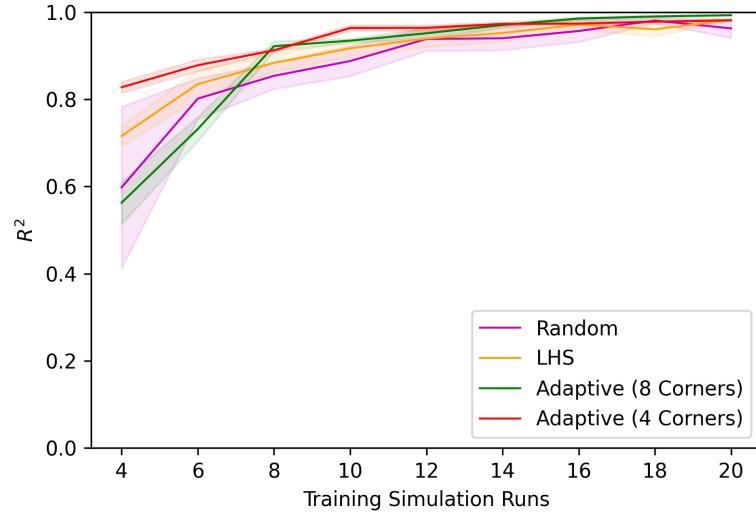


Figure 3.7: R^2 test performance of the machine learning surrogate in the prediction of the Case-D surface and downhole flow quantities with respect to the different sampling techniques. This scenario is only concerned with the prediction of a single producer (Producer-1) while keeping the other producer without ICV devices.

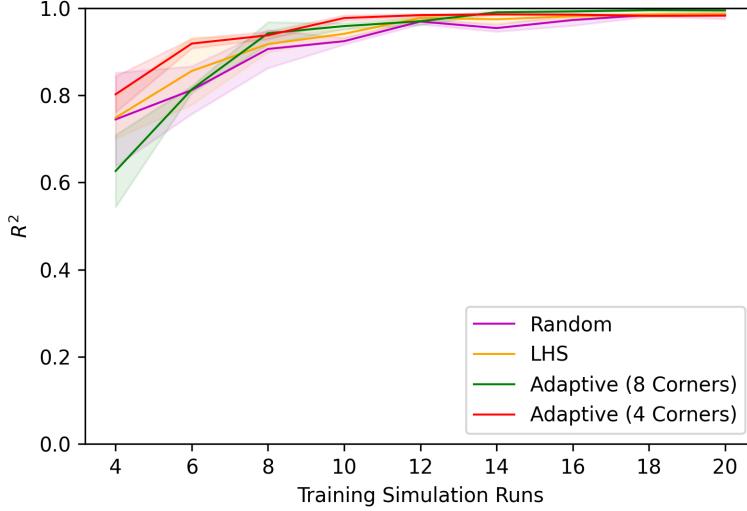


Figure 3.8: R^2 test performance of the machine learning surrogate in the prediction of the Case-D surface and downhole flow quantities with respect to the different sampling techniques. This scenario is only concerned with the prediction of a single producer (Producer-2) while keeping the other producer without ICV devices.

with three ICV devices, such that each device is capable of four fixed choke settings, where “0” and “3” indicate fully shut-in and fully open choke settings, respectively. Note that we have six ICV devices in total as input variables, which implies that we heuristically sample only $n = 6 + 1 = 7$ corner points out of the $4^6 = 64$ possible corners. Applying the surrogate-based algorithm, prediction performance of both producers in Case-D is shown in Fig. 3.9. We observe that the proposed 64-corner (all possible corners) adaptive sampling does not perform well at smaller numbers of ICV field tests, N . This is because $N < 64$, hence it is still sampling at the corners which only provides limited knowledge at the borders to the machine learning surrogate and omits sampling the interior of the input domain. It further does not improve monotonically with increasing N nor does its uncertainty band decrease monotonically for the same reason. Meanwhile, the heuristic seven-corner adaptive sampling alternative results in a surrogate that achieves 80% R^2 with only $N = 8$. However, it only improves slowly with increasing N . This is attributed to the curse of multidimensionality, where this sampling algorithm leaves out many subregions when

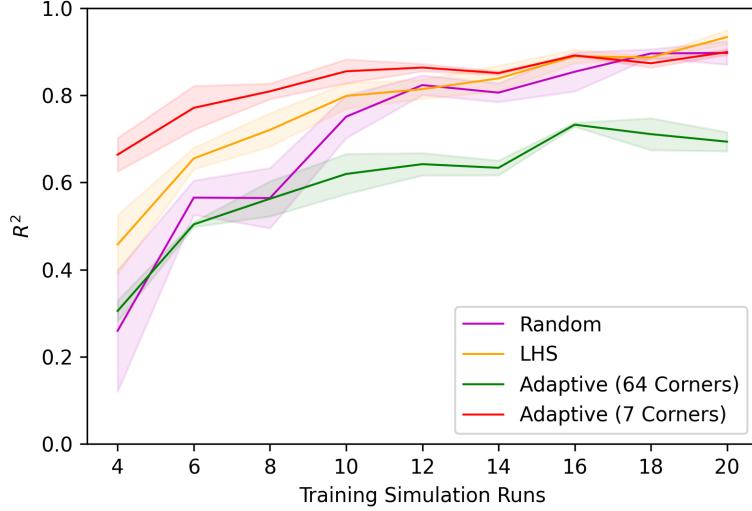


Figure 3.9: R^2 test performance of the machine learning surrogate in the prediction of the Case-D surface and downhole flow quantities with respect to the different sampling techniques. This scenario demonstrates the prediction of ICV flow performance in both forked trilaterals simultaneously.

constructing triangulations based on $7/64 = 11\%$ of all the available corners. On the other hand, LHS R^2 improves at relatively faster rates in comparison because it is a space-filling technique. While adaptive sampling still offers the best performance in achieving high R^2 values with a small number of ICV field tests, further improvement is possible by introducing a more space-filling criterion in cases where dimensionality is high.

The Case-E reservoir model is based on a real field, which we used to further verify the generalizability of the proposed algorithm in achieving the ICV performance prediction task. As seen in Fig. 3.10, combination of well types (forked and fishbone) and locations (crest and periphery) were used to generate four scenarios: Case-E1, Case-E2, Case-E3, and Case-E4. In addition dummy multilateral wells are also drilled and completed in each scenarios to introduce well-to-well complexity. Prediction results of scenarios Case-E1, Case-E2, Case-E3, and Case-E4 are demonstrated in Figs. 3.11, 3.12, 3.13, and 3.14, respectively (Note that we dropped random sampling as it resulted in much worse than average surrogates in these scenarios). Again,

adaptive sampling techniques resulted in more monotonic and superior R^2 results compared to LHS. Although wells at the crest did not produce any water, there were no significant differences in the algorithm performance. This is because the prediction complexity mainly stems out of the downhole, lateral-to-lateral flow prediction. In addition, the algorithm resulted in a fairly close performance in modelling both forked and fishbone well designs.

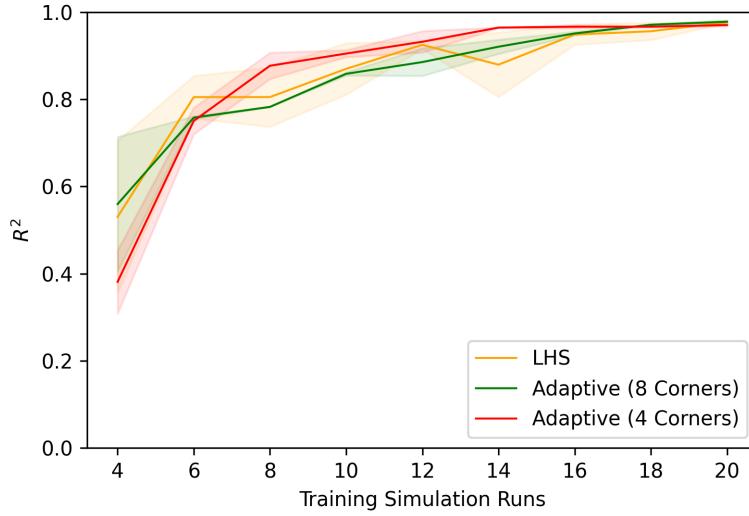


Figure 3.11: R^2 test performance of the machine learning surrogate in the prediction of the Case-E1 surface and downhole flow quantities with respect to the different sampling techniques.

To make Case-E even closer to reality, we considered the addition of uniform and Gaussian noise. Case-E1 is used to demonstrate the algorithm performance in this task. Note that we will use flow rate instead of cumulative flow quantities since the former is the raw and noisy data that is typically recorded in the field. As seen in Fig. 3.15, noise is added at standard deviation of 10%, 30%, and 50% of each training point. Note that the testing set is left noise-free because the premise is to test the algorithm's ability in capturing the true reservoir physics rather than noise. As seen in Fig. 3.16, the proposed algorithm was robust to noise where only minor variations in performance were observed after introducing uniform and Gaussian noise.

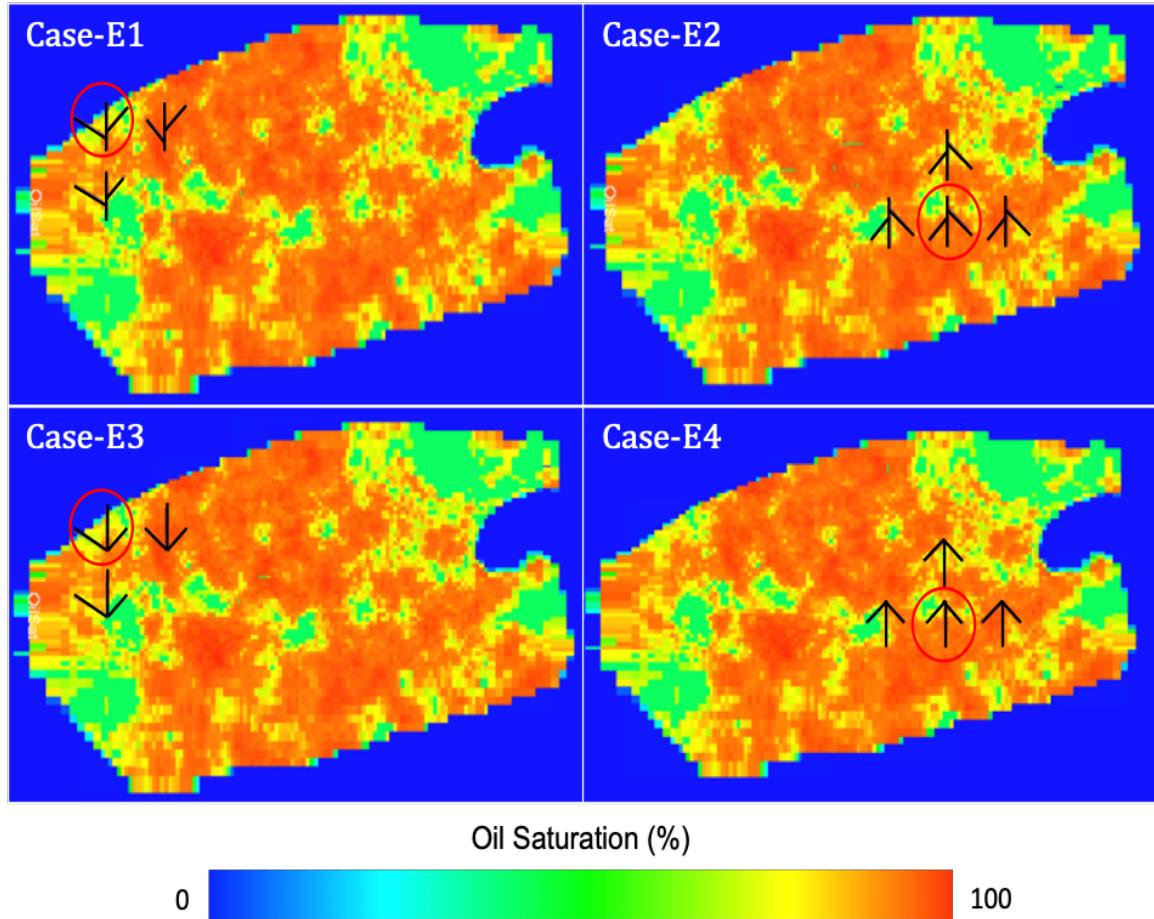


Figure 3.10: Case-E well placement and configuration choices are demonstrated on the reservoir oil saturation map. These four scenarios are 1) Case-E1: fishbone multilaterals placed at the periphery (top left), 2) Case-E2: fishbone multilaterals placed at the crest (top right), 3) Case-E3: forked multilaterals placed at the periphery (bottom left), and 4) Case-E4: forked multilaterals placed at the crest (bottom right). Note that the target well in each case, where the algorithm is to be applied, is circled in red.

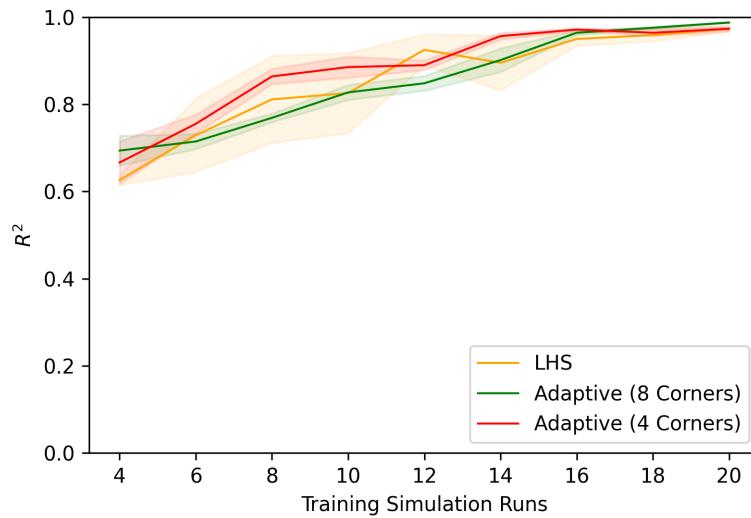


Figure 3.12: R^2 test performance of the machine learning surrogate in the prediction of the Case-E2 surface and downhole flow quantities with respect to the different sampling techniques.

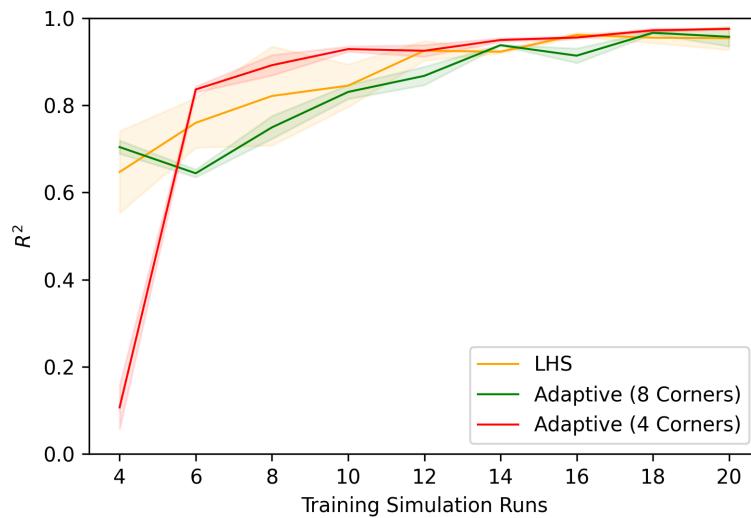


Figure 3.13: R^2 test performance of the machine learning surrogate in the prediction of the Case-E3 surface and downhole flow quantities with respect to the different sampling techniques.

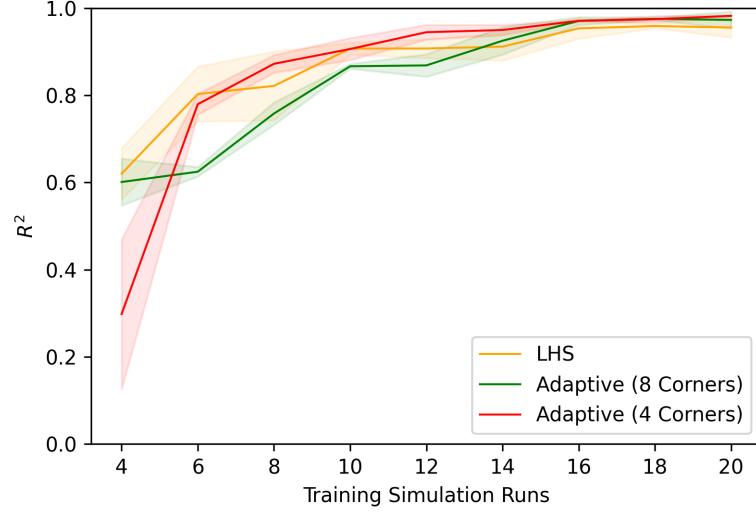


Figure 3.14: R^2 test performance of the machine learning surrogate in the prediction of the Case-E4 surface and downhole flow quantities with respect to the different sampling techniques.

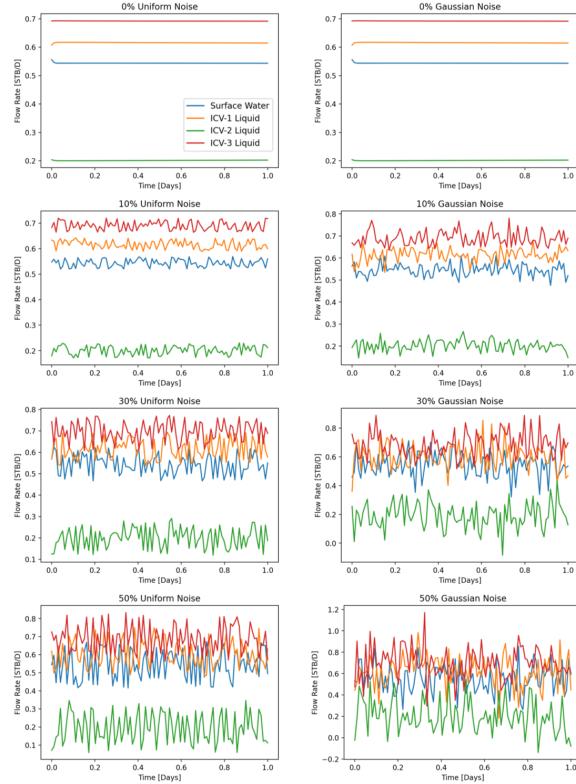


Figure 3.15: Example of a randomly selected ICV setting where surface and downhole flow rates are shown with various uniform and Gaussian noise levels.

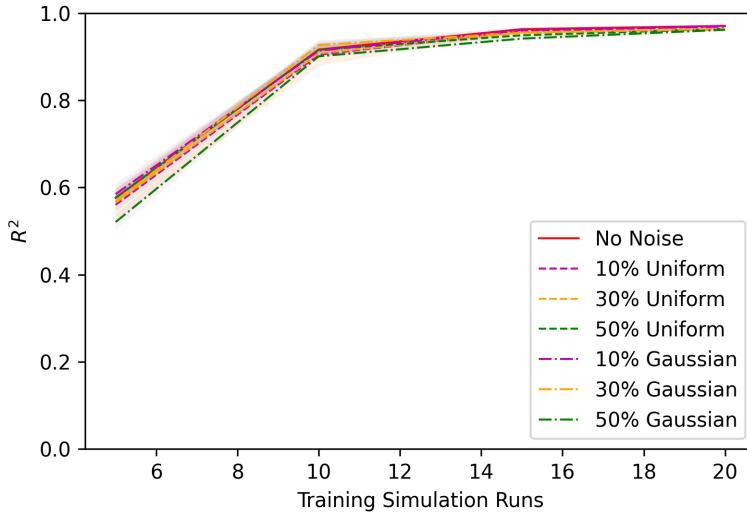


Figure 3.16: Algorithm R^2 results as applied to the Case-E1 reservoir model after adding uniform and Gaussian noise. The proposed algorithm shows to be robust to noise.

3.2 Deterministic Optimization

Given an accurate surrogate, optimization becomes possible. To illustrate the feasibility of optimization using the proposed algorithm, the objective function is simply set to maximizing the short-term NPV as the surrogate is only trained to predict current flow performance while forecasting is not considered. As seen in Eq. 3.1, NPV is computed simply using oil price p_{oil} , water cost c_{water} , cumulative oil production rate q_{oil} , and water production rate q_{water} . Throughout this study, we assumed p_{oil} and c_{water} of 60 USD and 20 USD, respectively. In this deterministic optimization, we only used enumeration for locating the optimal of both the real and predicted models. This is reasonable given only 512 possible combinations of the decision variables in a trilateral with ICV devices that can take eight possible discrete settings. Enumeration eliminates optimization underperformance due to potential inadequacy in common optimization solvers.

$$NPV = p_{oil}q_{oil} - c_{water}q_{water} \quad (3.1)$$

To illustrate the optimization potential of this algorithm, we will refer to Case-E1 as it involves undesired water production. The target trilateral well of this scenario, seen in Fig. 3.10, involves two laterals crossing the oil-water contact, lateral-1 (motherbore) and lateral-2, while lateral-3 is completed successfully across an oil-prolific zone. Note that ICV-1, ICV-2, and ICV-3 are installed at the tie-in points of lateral-1, lateral-2, and lateral-3, respectively. As seen in Fig. 3.17, the ground-truth NPV response surface of this well is first constructed using Eq. 3.1, where the optimal ICV settings and NPV values are enumerated as [1, 7, 7] and 249,215 USD per day, respectively. To enable visualization, we chose to fix ICV-3 to be fully open and visualize the NPV surface as a function of the other two ICV devices. We observe that NPV increases significantly as ICV-1 is choked further while the effect of ICV-2 is relatively minor. This is because the water cut (WC) is high across lateral-1 (70% when ICV-1 is fully open) while it is smaller across lateral-2 (40% when ICV-2 is fully open).

Using the proposed algorithm while training the surrogate surrogate on the flow quantities while calculating NPV implicitly, we can evaluate the algorithm performance in achieving both the prediction and optimization tasks simultaneously. As seen in Fig. 3.17, this is demonstrated by using four-corner and eight-corner adaptive sampling for comparison. Both surrogates are trained on a total of $N = 20$ ICV field tests with over 95% R^2 in predicting both surface and downhole flow quantities. The four-corner and eight-corner surrogates yield optimal ICV settings and NPV values of [0, 6, 7] and 213,143 USD per day, and [0, 7, 7] and 229,468 USD per day. Though both are matching and are close to the optimal ICV setting of [1, 7, 7], it is evident that the eight-corner surrogate is superior in capturing the optimal NPV magnitude and overall surface response. Furthermore, the four-corner and eight-corner surrogates achieve overall NPV R^2 of 87% and 95%, respectively. This shows the importance of capturing the corner points when field operations allow for a sufficient number of ICV field tests, especially in such problems where the optimal point is located in extreme

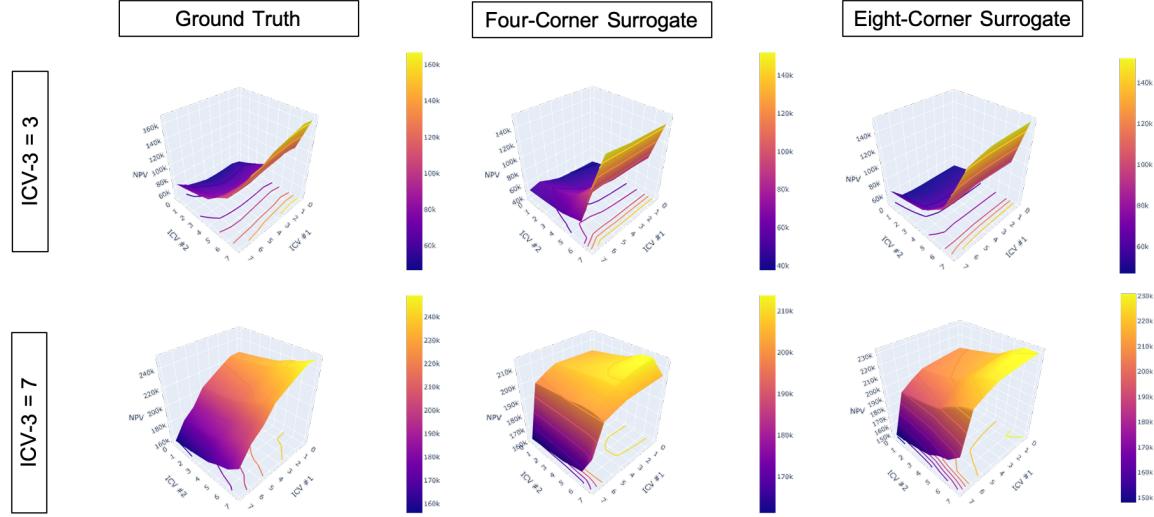


Figure 3.17: Demonstration of NPV surface at ICV-3 setting values of 3 and 7 for the ground truth (left), four-corner flow surrogate (middle), and eight-corner flow surrogate (right). Note that the 4-corner surrogate is less accurate at mapping the NPV surface when ICV-1 and/or ICV-3 settings are set to 7 (fully open). This is because some corners are omitted in this case, which leaves these regions of the design domain less defined to the surrogate. Meanwhile, the eight-corner surrogate shows more accurate surface map as its training points are sampled to capture all design domain corners.

regions of the design domain.

In order to further improve NPV prediction and optimization, we considered the situation where the operator is mainly interested in capturing the optimal ICV settings and closely estimating NPV without the need for predicting the flow quantities. In other words, we deployed adaptive sampling based on NPV and also fit the machine learning surrogate to predict NPV directly while ignoring physical quantities. As seen in Fig. 3.18, the four-corner and eight-corner NPV surrogates yield optimal ICV settings and NPV values of [0, 3, 7] and 207,455 USD per day, and [0, 7, 7] and 231,258 USD per day. This result confirms that predicting NPV explicitly is superior when the sole goal is to optimize ICV settings and NPV. However, we observe that using eight-corner adaptive sampling is still critical to capturing the case where ICV-3 is set to setting seven (fully open ICV device).

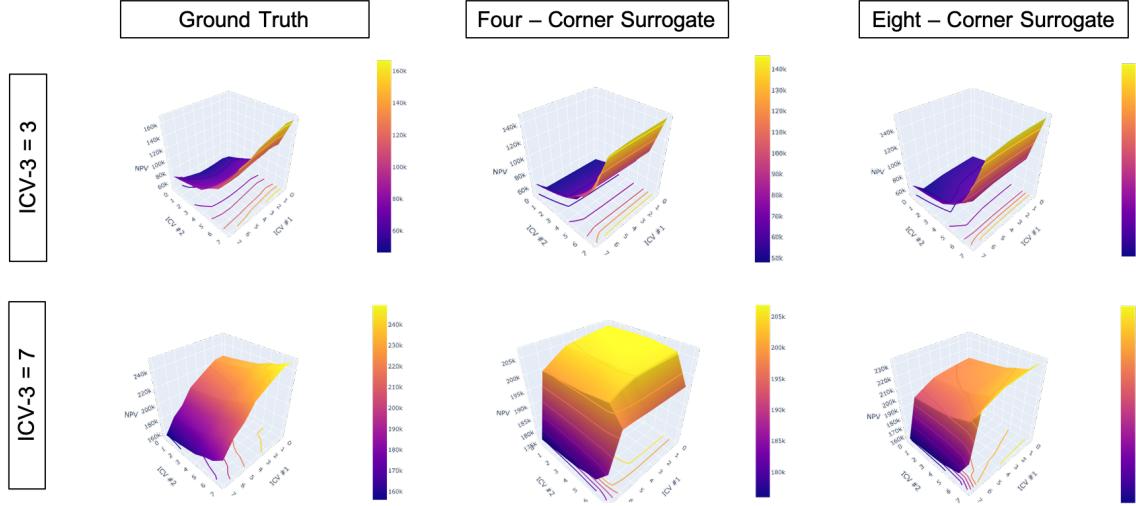


Figure 3.18: Demonstration of NPV surface at ICV-3 slices three and seven of the ground truth (left), four-corner NPV surrogate (middle), and eight-corner NPV surrogate (right). Note that the four-corner surrogate is less accurate at mapping the NPV surface when ICV-3 setting is set to seven (fully open). Meanwhile, the eight-corner surrogate shows a more accurate surface map as its training points are designed to capture all design domain corners.

3.3 Stochastic Optimization

This section describes the optimization of a geologically uncertain reservoir model where exhaustive analysis is infeasible. The Case-E reservoir model was used to investigate the application of the proposed algorithm alongside the utility framework in accomplishing this task. In addition, a water Carter-Tracy water aquifer was introduced across the underlying, water-saturated layers. This aims to induce early water breakthrough at producers driven by this aquifer. The Stanford Geostatistical Modelling Software (SGeMS) was used to generate the geological realizations [66]. Uncertainty in permeability was considered as the sole source of geological uncertainty, and it was used to generate geostatistical realizations. Hard data were also included based on permeability measurements retrieved from six real observation wells drilled across the different Case-E model layers. Because Case-E is based on a $69 \times 122 \times 14$ Cartesian grid, hard data included a total of $6 \times 14 = 84$ grid cells. Based on the history-matched model, an average-permeability trend along depth was adopted to

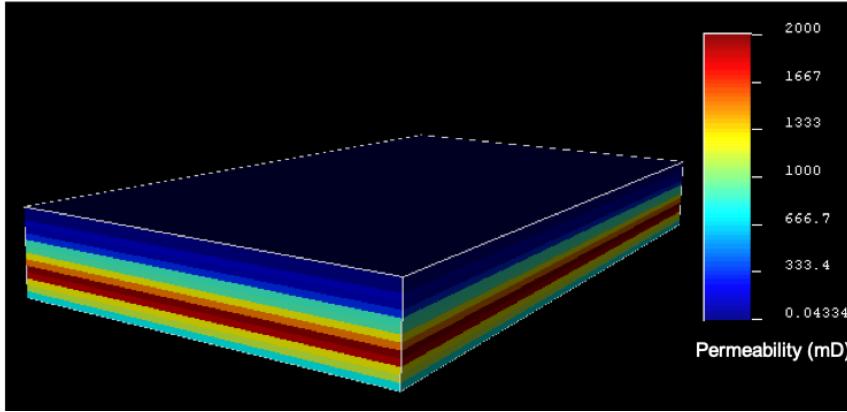


Figure 3.19: Permeability trend along depth as used to construct the Case-E geostatistical realizations using SGeMS.

generate realizations, seen in Fig. 3.19. A total of 500 permeability realizations were then generated using SGeMS, seen in Fig. 3.20.

Next, we applied DKM to find the most representative realizations. Note that these must be the most representative with respect to the different ICV settings as well. Hence, we placed a trilateral well at the center of the Case-E Cartesian grid to be produced for two years. Then, we adaptively chose 20 ICV settings. A total of 10,000 simulation runs were conducted and DKM was applied on water cut profiles to nominate the most representative realizations. As seen in Fig. 3.21, we evaluated different numbers of clusters (number of most representative realizations) to sufficiently match the P10, P50, and P90 statistics of water cut with the least number of clusters. Note that clustering is performed at a higher dimension then clusters are projected onto two dimensions for visualization purposes. We found that using 20 clusters was sufficient in this case.

Using these 20 most representative permeability realizations, we placed a trilateral at the center of the Cartesian grid to be produced for two years. This well was equipped with three ICV devices, each with eight discrete settings. Subsequently, a total of $20 \times 8^3 = 10,240$ numerical reservoir simulations were conducted to exhaustively perform stochastic optimization at first. NPV was also computed afterwards, and sample profiles of oil production rate, water production rate, and NPV were plotted in Fig. 3.22. Fig. 3.23 shows a histogram of NPV earned after two years of

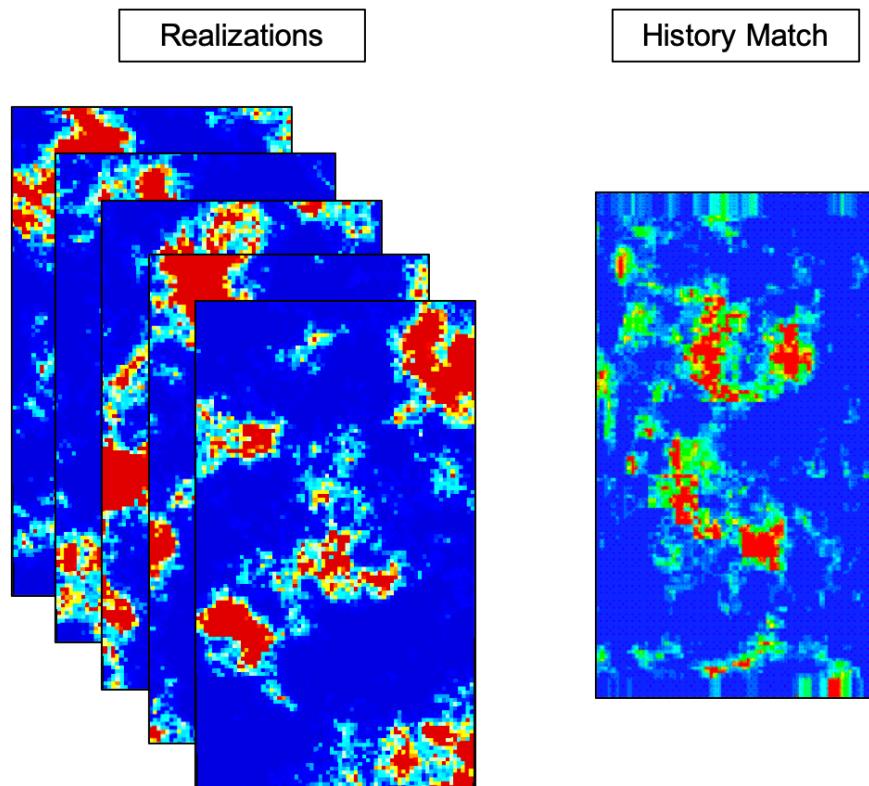


Figure 3.20: Example permeability realizations (left) of the Case-E payzone as generated using SGeMS in comparison to the history-matched permeability map (right). Note that the history-matched permeability field is shown based on the irregular Cartesian grid of Case-E where it involves elongated grid cells at the boundaries.

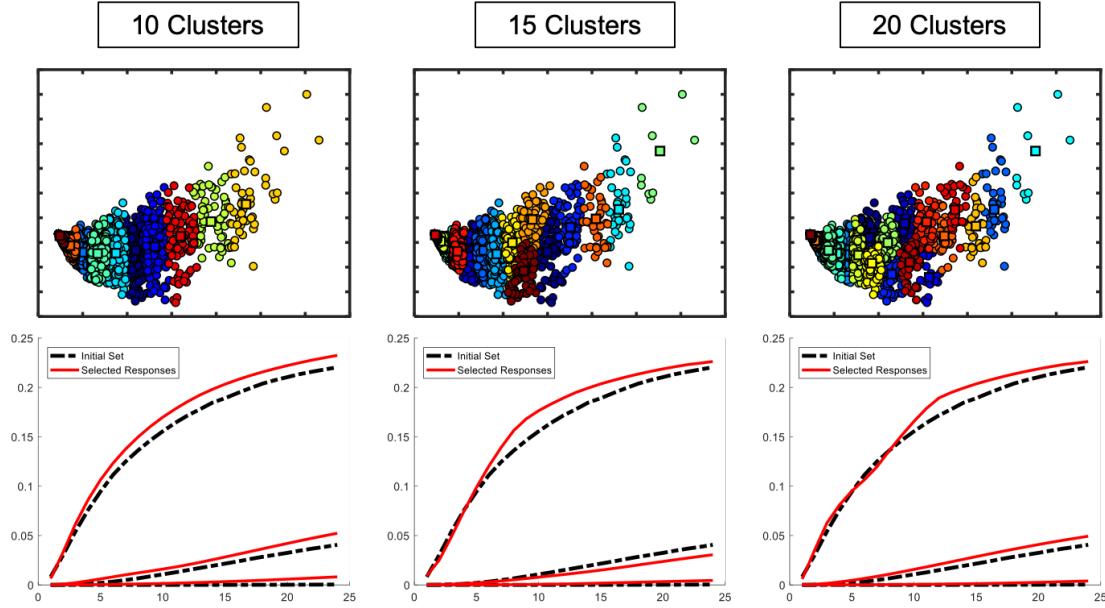


Figure 3.21: Demonstration of DKM clustering and P10, P50, and P90 fit results using water cut profiles of Case-E. Squares represent the medoids of the different clusters.

production based on each of the 10,240 realizations.

For each ICV setting, we used the arithmetic mean and standard deviation across realizations to quantify the NPV value and its uncertainty. The exhaustive analysis indicates that the maximum NPV is when ICV settings are all fully open [7, 7, 7] with 5.847×10^8 USD. However, it is computationally expensive to perform all 10,240 simulation runs exhaustively. Instead, we used the proposed algorithm to train a surrogate for each realization. As seen in Fig. 3.24, we performed this process using 20, 30, and 40 ICV field tests independently. We eventually chose 30 simulation runs as this results in surrogates whose R^2 is over 95% across all realizations. Using these surrogates, the maximum NPV is also when ICV settings are all fully open at [7, 7, 7] with 5.784×10^8 USD. This confirms that the algorithm successfully captured the optimal ICV setting based on the monetary value, NPV. Note that it took only 600 simulation runs to train the surrogates, resulting in a significant saving ($10,240 - 600$)/ $10,240 = 94\%$ of the total simulation runs.

However, this is not applicable to all decision makers as we need to also consider

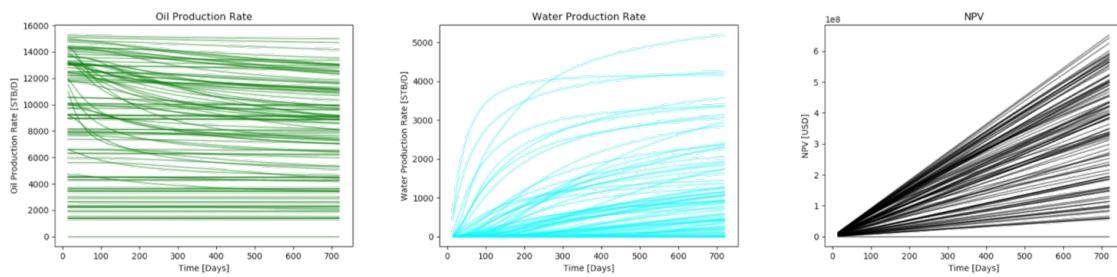


Figure 3.22: Sample profiles of oil production rate, water production rate, and NPV based on exhaustively simulating 10,240 reservoir models using the 20 most representative permeability fields.

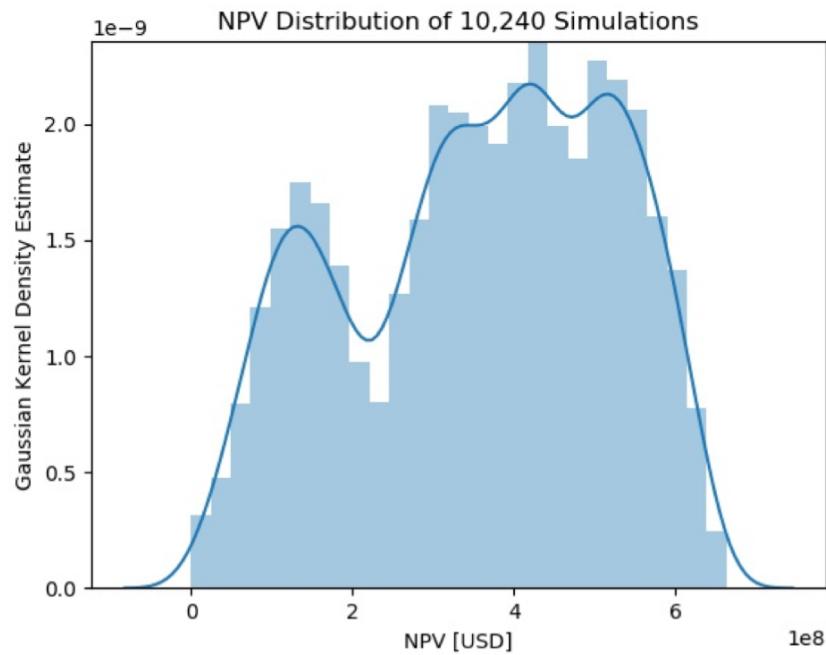


Figure 3.23: Histogram of NPV based on production of a single trilateral for a total of two years in all 10,240 simulation runs.

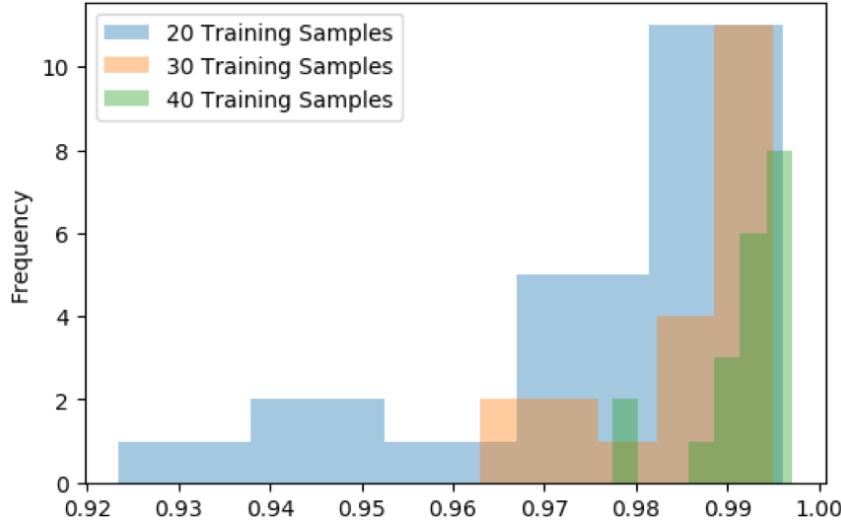


Figure 3.24: R^2 histogram based on the evaluation of the 20 surrogates each trained on 30 adaptively sampled ICV field tests and tested on the remaining $511 - 30 = 481$.

the uncertainty in NPV and the decision makers' risk aversion. This can be accomplished using the utility framework. We first visualized the surrogates' ability in capturing the uncertainty in NPV based on standard deviation. As seen in Fig. 3.25, the trained surrogates' predictions across realizations were used to quantify the NPV standard deviation for each ICV setting. We visually observed a close fit in trend with an overall R^2 of over 95%. Hence, we used Eq. 2.15 to perform the utility framework analysis for risk aversion value of $r = \{4, 2, 0, -2, -4\}$. Recall that $r < 0$, $r = 0$, and $r > 0$ indicate risk-averse, neutral, and risk-prone decision makers, respectively.

Enumeration was used again to determine the optimal ICV settings based on the utility framework for both the ground truth and model prediction, seen in Table 3.1. We observed a close match between the ICV settings of the ground truth and surrogate predictions of utility. Mismatches are found in cases of $r \geq 0$, where the decision maker is either risk-averse or neutral. These mismatches also result in NPV errors of 7.43%, 2.02%, 0.45% at $r = 4$, $r = 2$, and $r = 0$, respectively. These are relatively small errors that are monotonically decreasing with decreasing risk aversion factor, r , values. To explain this phenomenon, we can visualize the ground truth utility histograms for different decision makers, i.e. different risk aversion factor

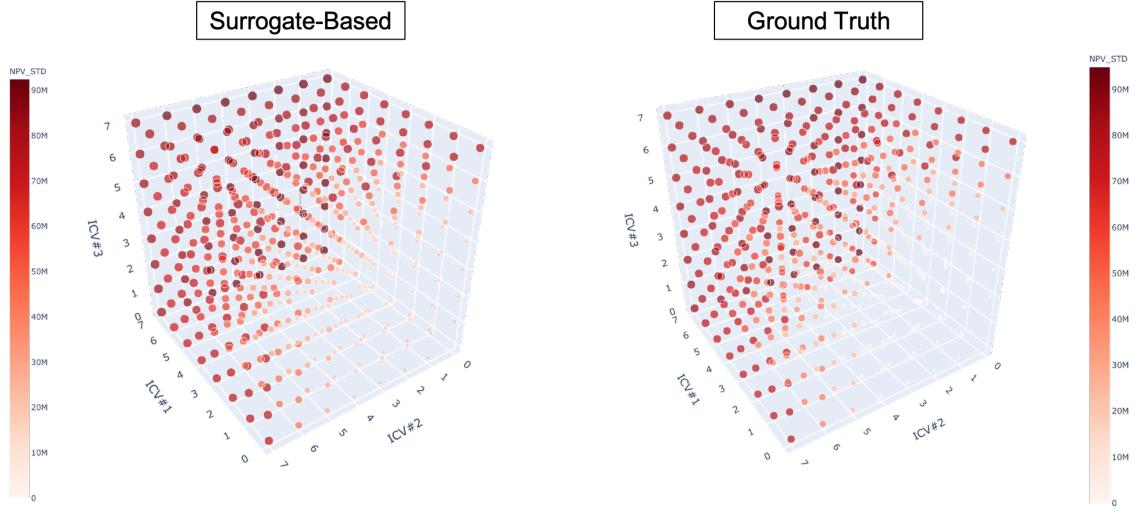


Figure 3.25: Visualization of NPV standard deviation across the different ICV settings based on the predictions of the trained surrogates.

values, seen in Fig. 3.26. We observe that when a decision maker is risk-averse $r > 0$, the distribution is left-skewed. Hence, there exist multiple decisions (combinations of ICV settings) that result in close-to-optimal utility values. Meanwhile, when a decision maker is risk-prone $r < 0$, the distribution is right-skewed. Hence, the optimal decision is more defined in comparison. Furthermore, we can visualize the utility surface for by fixing $ICV - 3 = 7$ (fully-open). This choice is reasonable because all ground-truth optimal decisions involved $ICV - 3 = 7$, fully open. Fig. 3.27 shows a similar conclusion to 3.26 where the global optimum is more defined in cases when the decision maker is risk-prone $r < 0$.

In addition, it is important to achieve computationally efficient optimization of the trained surrogates rather than using enumeration. Hence, we applied MADS to optimize over utility based on the surrogate response surface. In the MADS optimization process, we systematically initialize all ICV devices at setting four, the closest to the center of the input domain. Fig. 3.28 shows the surrogate utility as a function of the MADS iterations for risk aversion factor of $r = (4, 2, 0, -2, -4)$. These results indicate early convergence with as low as 35 surrogate function evaluations, which is 15 times faster than using enumeration. Comparison of enumeration and MADS optimization results over the trained surrogates are shown in Table 3.1. We observe

Table 3.1: Comparison of the enumerated and MADS-calculated optimal settings based on the ground truth and predicted utility. Red values indicate mismatching optimal ICV settings between the ground truth and surrogate predictions.

Risk Aversion (r)	Ground Truth			Enumeration (MADS)		
	ICV-1	ICV-2	ICV-3	ICV-1	ICV-2	ICV-3
4	3	6	7	3(3)	6(6)	4(4)
2	4	7	7	6(0)	6(7)	7(7)
0	6	7	7	7(7)	7(7)	7(7)
-2	7	7	7	7(7)	7(7)	7(7)
-4	7	7	7	7(7)	7(7)	7(7)

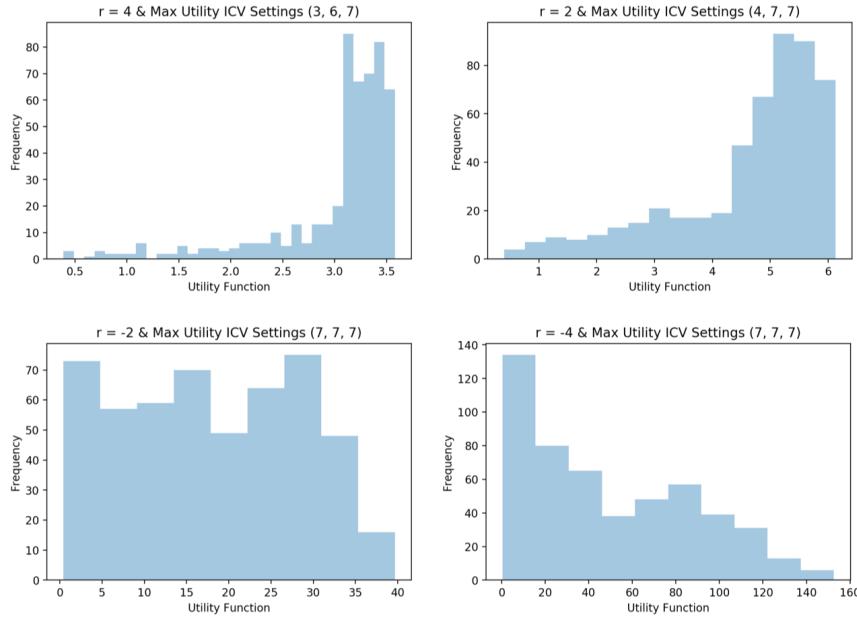


Figure 3.26: Ground truth utility histograms for different risk aversion factor, r , values.

a close match in most cases with small errors. Again, this mismatch is attributed to the left-skewed utility distribution in the cases or risk-averse decision makers where multiple local optima exist in a single subregion. Fig. 3.29 shows the points evaluated by MADS in the three-dimensional input domain. The optimization solver is clearly capturing the desired subregions of the input domain where utility is optimal.

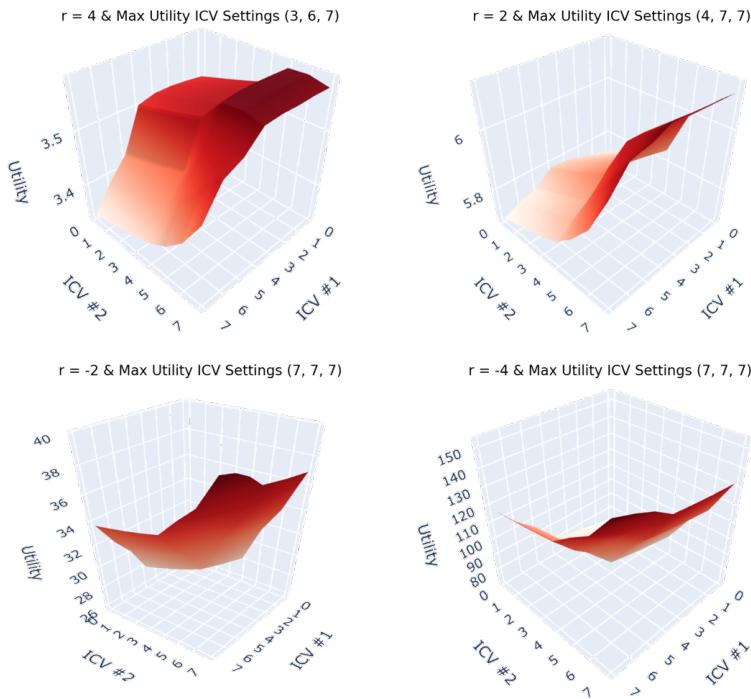


Figure 3.27: Ground truth utility surfaces for values of $ICV - 3 = 7$ (fully open).

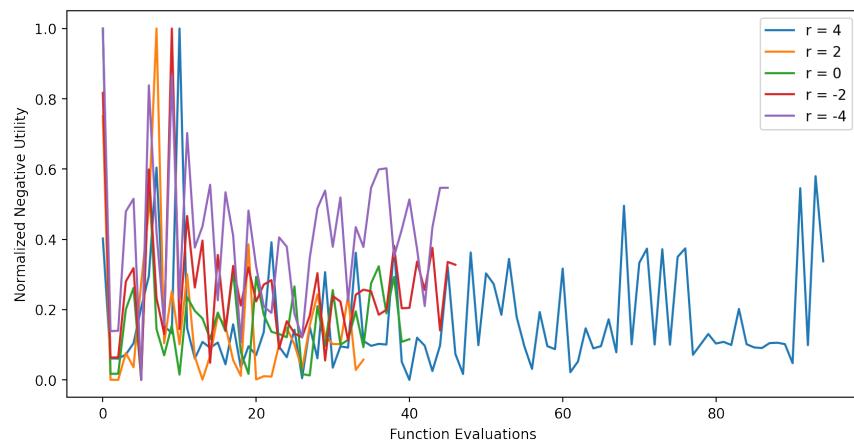


Figure 3.28: Normalized negative utility values of MADS surrogate function evaluations for different decision makers (different risk aversion factors).

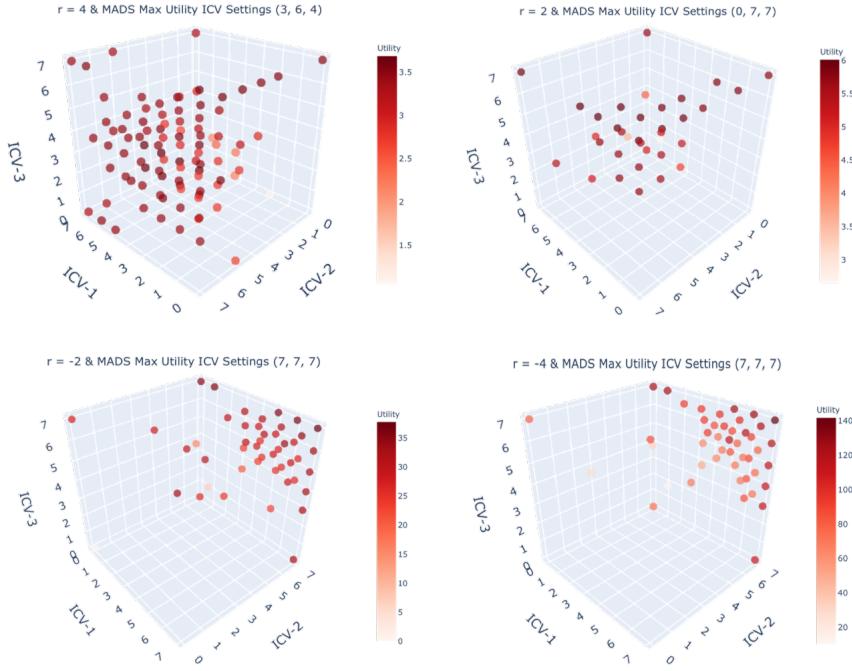


Figure 3.29: Evaluated points by MADS throughout the optimization of the single-producer, Case-E surrogates using utility.

To further examine the significance of the proposed methodology, we investigated a situation where we aim to optimize the ICV settings of two trilaterals, each with three eight-discrete-settings ICV devices. Given the 20 most representative permeability field realizations, seen in Fig. 3.30, this case would require a total of $20 \times 8^6 = 5,242,880$ simulation runs to accomplish exhaustive analysis; however, this is infeasible. Instead, we only sampled 100 ICV settings adaptively for each realization and subsequently trained surrogates. This only requires a total of $20 \times 100 = 2,000$ realizations, which results in avoiding $(5,242,880 - 2,000)/5,242,880 = 99.96\%$ of the total number of simulation runs. Enumeration was used again to determine the optimal ICV settings based on the utility framework for both the ground truth and model prediction, seen in Table 3.2. Note that we did not perform exhaustive analysis in this case due to the large number of simulation runs required; hence, the ground-truth optimization results are not shown. Similar to the single-trilateral optimization problem, we observe that risk-prone decision makers $r < 0$ have higher tendency to

keep ICV devices open when compared to their risk-averse counterparts.

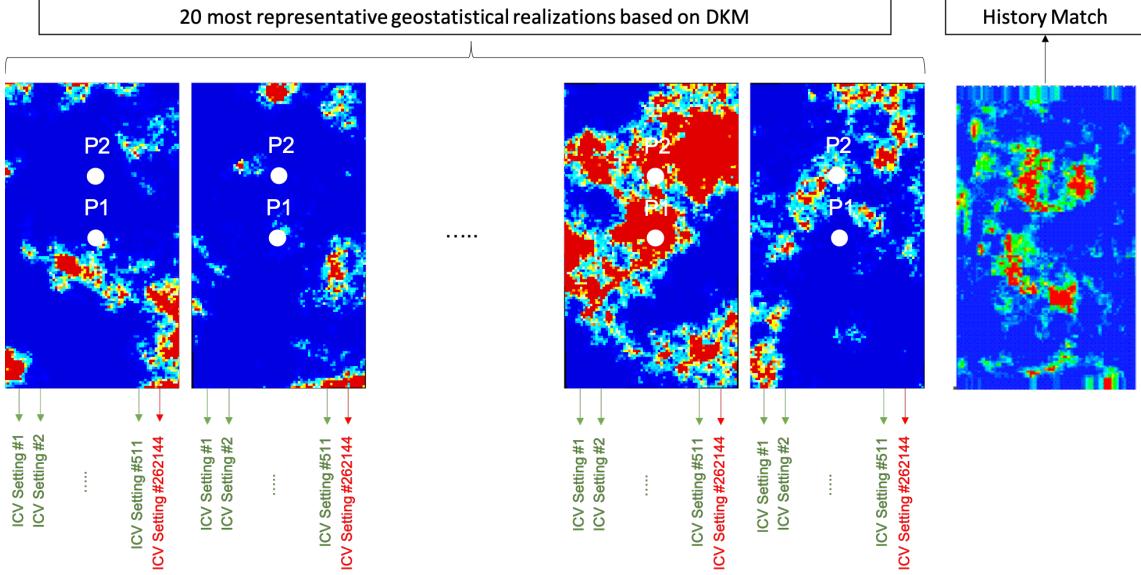


Figure 3.30: The 20 most representative realizations as per DKM in comparison to the history-matched permeability map. Two trilaterals are drilled and completed close to the grid center.

Table 3.2: Comparison of the enumerated and MADS (in brackets) optimal settings based on the utility framework for Case-E1 with two trilateral producers.

Well	Risk Aversion (r)	4	2	0	-2	-4
Producer-1	ICV-1	7(7)	7(7)	7(7)	7(7)	7(7)
	ICV-2	0(4)	4(0)	4(4)	7(7)	7(7)
	ICV-3	0(7)	7(6)	7(7)	7(7)	7(7)
Producer-2	ICV-1	6(0)	0(0)	2(2)	0(0)	6(6)
	ICV-2	6(6)	6(7)	6(6)	7(7)	7(7)
	ICV-3	7(7)	7(0)	7(7)	7(7)	7(7)

In addition, we optimized this scenario using MADS and achieved close matches to enumeration as seen in Table 3.2. In this case, all MADS optimization runs terminated with an average of nearly 1,500 utility function evaluations, which indicates optimization time saving of $(5,242,880 - 1,500)/5,242,880 = 99.97\%$. Again, mismatches are attributed to the left-skewed utility distribution in the cases or risk-averse decision makers where multiple local optima exist in a single subregion.

Chapter 4

Conclusions and Future Work

This study investigated a surrogate-based prediction and optimization algorithm that minimizes the number of ICV field tests required, predicts well performance of all unseen combination of ICV settings, and determines the optimal ICV setting and NPV.

This algorithm was found effective when applied to five numerical reservoir models: Case-A, Case-B, Case-C, Case-D, and Case-E. Considering the real reservoir model of Case-E with multiple producers at the crest or periphery, the algorithm only required six and ten ICV field tests on average to achieve 80% and 90% R^2 across the different scenarios of this reservoir model. Furthermore, the model achieved similar results when trained on production profiles with uniform and Gaussian noise at levels of 10%, 30%, and 50%. The resultant surrogate was also used to decide on the optimal settings of ICV devices and also predict NPV effectively.

Stochastic optimization was achieved for a single-well scenario, where 10,240 simulation runs would be required to accomplish exhaustive analysis. Instead, the proposed algorithm achieved comparable optimal results with only 600 simulation runs with 94% saving in simulation time. Another scenario involved the stochastic optimization of two wells, where 5,242,880 simulation runs would be required to accomplish exhaustive analysis. Instead, the proposed algorithm achieved comparable optimal results with only 2,000 simulation runs with 99.96% saving in simulation time.

Several steps may be considered in the future in order to further expand on this work. Enhancement of sampling techniques, such that both the space-filling and adaptive sampling could be integrated into a more optimal optimizer. Other statistical and/or physical models can also be considered to improve the resultant surrogate. It would be useful to develop long-term NPV prediction and optimization, where forecasting is considered. Finally, in-depth investigation of the different utility functions is required to achieve stochastic optimization that is more compatible with the interest of the decision makers.

Bibliography

- [1] A. Abukhamsin. Optimization of well design and location in a real field. Master's thesis, Stanford University, 2009.
- [2] A. Y. Abukhamsin. *Inflow profiling and production optimization in smart wells using distributed acoustic and temperature measurements*. PhD thesis, Stanford University, 2017.
- [3] E. Addiego-Guevara, M. Jackson, and M. Giddins. Insurance value of intelligent well technology against reservoir uncertainty. Paper SPE-113918-MS, presented at the SPE Symposium on Improved Oil Recovery, Tulsa, Oklahoma, 20-23 April, 2008.
- [4] O. A. Adekunle. Intelligent well applications in production wells. Master's thesis, The University of Aberdeen, 2008.
- [5] M. Ahmadi, A. Ebadi, M. Shokrollahi, and S. Majidi. Evolving artificial neural network and imperialist competitive algorithm for prediction oil flow rate of the reservoir. *Applied Soft Computing*, 13(2):1085–1098, 2013.
- [6] A. Ajayi and M. Konopczynski. Intelligent wells technology-fundamental and practical applications. A Course Introduced at the SPE Annual Technical Conference and Exhibition (ATCE), Denver, Colorado, 20-21 September, 2008.
- [7] A. Al-Anazi and I. Gates. A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. *Engineering Geology*, 114(3-4):267–277, 2010.

- [8] Z. M. Al-Ghareeb. Monitoring and control of smart wells. Master's thesis, Stanford University, 2009.
- [9] M. D. AlAjmi, S. A. Alarifi, and A. H. Mahsoon. Improving multiphase choke performance prediction and well production test validation using artificial intelligence: A new milestone. Paper SPE-173394-MS Presented at the SPE Digital Energy Conference and Exhibition, The WoodLands, Texas, 3-5 March, 2015.
- [10] A. Alakeely and R. Horne. Simulating the behavior of reservoirs with convolutional and recurrent neural networks. Paper IPTC-20191-MS Presented at the International Petroleum Technology Conference, Dhahran, Saudi Arabia, 13-15 Januray, 2020.
- [11] C. Alippi, G. Anastasi, C. Galperti, F. Mancini, and M. Roveri. Adaptive sampling for energy conservation in wireless sensor networks for snow monitoring applications. Paper Presented at IEEE International Conference on Mobile Ad-hoc and Sensor Systems, 2007.
- [12] M. J. Aljubran and R. N. Horne. Prediction of multilateral inflow control valve flow performance using machine learning. Paper SPE-196003-MS Presented at the SPE Annual Technical Conference and Exhibition (ATCE), Calgary, Alberta, 30 September – 2 October, 2019.
- [13] L. F. Almeida, M. M. Vellasco, and M. A. Pacheco. Optimization system for valve control in intelligent wells under uncertainties. *Journal of Petroleum Science and Engineering*, 73(1-2):129–140, 2010.
- [14] R. Arnab. Survey sampling theory and applications. Academic Press, 2017.
- [15] M. Asadollahi and G. Naevdal. Waterflooding optimization using gradient based methods. Paper SPE-125331-MS Presented at the SPE/EAGE Reservoir Characterization and Simulation Conference, Abu Dhabi, UAE, 19-21 October, 2009.
- [16] M. Asadollahi, G. Naevdal, M. Dadashpour, and J. Kleppe. Production optimization using derivative free methods applied to brugge field case. *Journal of Petroleum Science and Engineering*, 114:22–37, 2014.

- [17] C. Audet and J. Dennis Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 17(1):188–217, 2006.
- [18] C. Audet, S. Le Digabel, C. Tribes, and V. Montplaisir. The nomad project. 2015.
- [19] M. Bakr and M. Negm. Modeling and design of high-frequency structures using artificial neural networks and space mapping. Elsevier BV, 2012.
- [20] B. Beachkofski and R. Grandi. Improved distributed hypercube sampling. Paper 2002-1274 Presented 43rd American Institute of Aeronautics and Astronautics AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2002.
- [21] T. Behrouz, M. Rasaei, and R. Masoudi. A novel integrated approach to oil production optimization and limiting the water cut using intelligent well concept: Using case studies. *Iranian Journal of Oil Gas Science and Technology*, 5(1):27–41, 2015.
- [22] J. Blazek. Computational fluid dynamics: Principles and applications. Butterworth-Heinemann, 2015.
- [23] I. Borg and P. Groenen. Modern multidimensional scaling: Theory and applications. Springer Series in Statistics, 1997.
- [24] S. Bosworth, H. S. El-Sayed, G. Ismail, H. Ohmer, M. Stracke, C. West, and A. Retnanto. Key issues on multilateral technology. *Oil Field Review*, 10(4):14–28, 1998.
- [25] D. Brouwer and J. Jansen. Dynamic optimization of water flooding with smart wells using optimal control theory. *SPE Journal*, 9(4):391–402, 2004.
- [26] A. Buhulaigah, A. Al-Mashhad, S. Al-Arif, M. Al-Kadem, and M. Al-Dabbous. Performance prediction for advanced well structure in unconventional oil and gas reservoirs using artificial expert systems. Paper SPE-183688-MS Presented

- at the SPE Middle Easter Oil Gas Show and Conference, Manama, Kingdom of Bahrain, 6-9 March, 2017.
- [27] R. Carter and G. Tracy. An improved method for calculating water influx. *Transactions of AIME*, 219:415–417, 1960.
 - [28] G. Carvajal, M. Maucec, and Cullick S. Intelligent digital oil and gas fields: concepts, collaboration, and right-time decisions. Gulf Professional Publishing, 2018.
 - [29] B. Chen and A. Reynolds. Optimal control of icv's and well operating conditions for the water-alternating-gas injection process. *Journal of Petroleum Science and Engineering*, 149:623–640, 2010.
 - [30] B. Chen and A. Reynolds. Optimal control of icv's and well operating conditions for the water-alternating-gas injection process. *Journal of Petroleum Science and Engineering*, 149:623–640, 2017.
 - [31] A. Datta-Gupta, A. Alhuthali, B. Yuen, and J. Fontanilla. Field applications of waterflood optimization using optimal rate control with smart wells. *SPE Reservoir Evaluation Engineering*, 13(3):406–422, 2010.
 - [32] B. Delaunay. On the empty sphere. *Bulletin of the Academy of Sciences of the USSR*, 6:793–800, 1934.
 - [33] F. Ebadi and D. R. Davies. Should "proactive" or "reactive" control be chosen for intelligent well management? Paper SPE-99929-MS Presented at the Intelligent Energy Conference and Exhibition, Amsterdam, The Netherlands, 11-13 April, 2006.
 - [34] D. Echeverria Ciaurri, O. Isebor, and Durlofsky L. Application of derivative-free methodologies to generally constrained oil production optimization problems. *International Journal of Mathematical Modelling and Numerical Optimisation*, 2(2):134–161, 2011.

- [35] V. Eglajs and P. Audze. New approach to the design of multifactor experiments. *Problems of Dynamics and Strengths*, 35:104–107, 1979.
- [36] E. A. El-Sebakhy. Forecasting pvt properties of crude oil systems based on support vector machines modeling scheme. *Journal of Petroleum Science and Engineering*, 64(1-4):25–34, 2009.
- [37] C. Enyioha and T. Ertekin. Performance prediction for advanced well structure in unconventional oil and gas reservoirs using artificial expert systems. Paper SPE-187037-MS to Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 9-11 October, 2017.
- [38] G. V. Essen, P. V. Hof, and J. D. Jansen. Hierarchical long-term and short-term production optimization. *SPE Journal*, 16(1):191–199, 2011.
- [39] R. Ge, F. Huang, C. Jin, and Y. Yuan. Escaping from saddle points—online stochastic gradient for tensor decomposition. Paper Presented at the Conference on Learning Theory., 2015.
- [40] GeoQuest. ECLIPSE technical description. Schlumberger, 2013.
- [41] M. Giles and N. Pierce. An introduction to the adjoint approach to design. *Flow, Turbulence and Combustion*, 65:393–415, 2000.
- [42] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. Paper Presented at the thirteenth international conference on artificial intelligence and statistics, 2009.
- [43] I. M. Grebenkin and D. R. Davies. A novel optimisation algorithm for inflow control valve management. Paper SPE-154472-MS Presented at the SPE Europec/EAGE Annual Conference, Copenhagen, Denmark, 4-5 June, 2012.
- [44] B. Guyaguler. *Optimization of well placement and assessment of uncertainty*. PhD thesis, Stanford University, 2002.

- [45] B. Guyaguler and T. Byer. A new rate-allocation-optimization framework. *SPE Production and Operations*, 234(4):448–457, 2008.
- [46] S. Haykin. Neural networks: A comprehensive foundation. Prentice Hall, 1994.
- [47] H. Helle, A. Bhatt, and B. Ursin. Porosity and permeability prediction from wireline logs using artificial neural networks: A north sea case study. *Applied Soft Computing*, 49(4):431–444, 2001.
- [48] C. Holloway. Decision making under uncertainty. Prentice Hall, 1979.
- [49] B. Hoover. Moving toward an optimal and automated geospatial network for ccus infrastructure. *Los Alamos National Laboratory*, 2016.
- [50] R. Howard. Information value theory. *IEEE Transactions on Systems, Man, and Cybernetics*, 2(2):22–26, 1966.
- [51] A. Huell, C. Whitson, and M. Hoda. Model-based integration and optimization – gas cycling benchmark. *SPE Journal*, 15(2):646–657, 2010.
- [52] A. Jain and E. Chang. Adaptive sampling for sensor networks. Paper Presented at 1st International Workshop on Data Management for Sensor Networks, New York, 2004.
- [53] J. Jansen, R. Brouwer, and S. Douma. Closed loop reservoir management. Paper SPE-119098-MS Presented at the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 2-4 February, 2009.
- [54] Y. Jin, J. Li, W. Du, and F. Qian. Adaptive sampling for surrogate modelling with artificial neural network and its application in an industrial cracking furnace. *The Canadian Journal of Chemical Engineering*, 94(2):262–272, 2015.
- [55] D. Kingama and J. Ba. Adam: A method for stochastic optimization. Paper presented at the International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May., 2015.

- [56] R. Malakooti, A. Ayop, B. Maulianda, K. Muradov, and D. Davies. Integrated production optimization and monitoring of multi-zone intelligent wells. *Journal of Petroleum Exploration and Production Technology*, 20(1):159–170, 2020.
- [57] R. Malakooti, K. Muradov, and D. Davies. Flow control optimization to maximize the accuracy of multi-phase flow rate allocation. Paper SPE-173873-MS Presented at the SPE Bergen One Day Seminar, Bergen, Norway, 22 April, 2015.
- [58] J. Mario and M. Balesdent. Estimation of rare event probabilities in complex aerospace and other systems: a practical approach. Woodhead Publishing, 2016.
- [59] M. McKay, R. Beckman, and W. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *American Statistical Association*, 21(2):239–245, 1979.
- [60] T. Mitchell. Machine learning. McGraw Hill, 1997.
- [61] M. Naus, N. Dolle, and J. Jansen. Optimization of commingled production using infinitely variable inflow control valves. *SPE Production Operations*, 21(2):293–301, 2006.
- [62] D. F. Oliveira and A. C. Reynolds. An adaptive hierarchical multiscale algorithm for estimation of optimal well controls. *SPE Journal*, 19(5):909–930, 2014.
- [63] K. Park and J. Caers. History matching in low-dimensional connectivity-vector space. Paper Presented at the EAGE Petroleum Geostatistics Conference, Cascais, Portugal, 10–14 September, 2007.
- [64] M. A. Pinto, C. E. Barreto, and D. J. Schiozer. Optimization of proactive control valves of producer and injector smart wells under economic uncertainty. Paper SPE-154511-MS Presented at the SPE Europec/EAGE Annual Conference, Copenhagen, Denmark, 4–7 June, 2012.
- [65] S. Rahmawati, B. Whitson, C. and Foss, and A. Kuntadi. Field asset integrated optimization benchmark. Paper SPE-130768-MS Presented at the SPE

- EUROPEC/EAGE Annual Conference and Exhibition, Barcelona, Spain, 14-17 June, 2010.
- [66] N. Remy, A. Boucher, and J. Wu. Applied geostatistics with sgems. Cambridge University Press, 2011.
 - [67] T. Ristanto. Machine learning applied to multiphase production problems. Master’s thesis, Stanford University, 2018.
 - [68] R. Rojas. Neural networks: A systematic introduction. Springer Science Business Media, 2013.
 - [69] M. Rubinstain. Patterns of problem solving. Prentice Hall, 1975.
 - [70] Y. Saka, M. Gunzburger, and J. Burkardt. Latinized, improved lhs, and cvt point sets in hypercubes. *International Journal of Numerical Analysis and Modeling*, 4(3-4):729–743, 2007.
 - [71] P. Sarma, W. Chen, L. Durlofsky, and K. Aziz. Production optimization with adjoint models under nonlinear control-state path inequality constraints. *SPE Reservoir Evaluation Engineering*, 11(2):326–339, 2008.
 - [72] C. Scheidt and J. Caers. Using distances and kernels to parameterize spatial uncertainty for flow applications. Paper Presented at the EAGE Petroleum Geostatistics Conference, Cascais, Portugal, 10–14 September, 2007.
 - [73] C. Scheidt and J. Caers. Representing spatial uncertainty using distances and kernels. *Mathematical Geosciences*, 141(4):397–419, 2008.
 - [74] C. Scheidt and J. Caers. Uncertainty quantification in reservoir performance using distances and kernel methods—application to a west africa deepwater turbidite reservoir. *SPE Journal*, 14(4):680–692, 2009.
 - [75] B. Schoelkopf and A. Smola. Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT Press, 2002.

- [76] M. H. Sefat, A. H. Elsheikh, K. M. Muradov, and D. R. Davies. Reservoir uncertainty tolerant, proactive control of intelligent wells. *Computational Geosciences*, 20:655–676, 2015.
- [77] P. Serma, L. Durlofsky, and K. Aziz. Implementation of adjoint solution for optimal control of smart wells. Paper SPE-92864-MS Presented at the SPE Reservoir Simulation Symposium, The Woodlands, Texas, 31 January – 2 February, 2005.
- [78] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [79] P. Stein. A note on the volume of a simplex. *The American Mathematical Monthly*, 73(3):299–301, 1966.
- [80] E. Suwartadi, S. Krogstad, and B. Foss. On state constraints of adjoint optimization in oil reservoir waterflooding. Paper SPE-125557-MS Presented at the SPE/EAGE Reservoir Characterization and Simulation Conference, Abu Dhabi, UAE, 19-21 October, 2009.
- [81] S. Suzuki and J. Caers. History matching with an uncertain geological scenario. Paper SPE-102154-MS Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, 24–27 September, 2006.
- [82] M. Telgarsky. Benefits of depth in neural networks. *Journal of Machine Learning Research*, 49:1–23, 2016.
- [83] C. Tian. Applying machine learning and data mining techniques to interpret flow rate, pressure and temperature data from permanent downhole gauges. Master’s thesis, Stanford University, 2014.
- [84] C. Tian. *Machine learning approaches for permanent downhole gauge data interpretation*. PhD thesis, Stanford University, 2018.
- [85] C. Tian and R. Horne. Machine learning applied to multiwell test analysis and flow rate reconstruction. Paper SPE-175059-MS Presented at the SPE Annual Technical Conference and Exhibition, Houston Texas, 28-30 September, 2015.

- [86] C. Tian and R. Horne. Recurrent neural networks for permanent downhole gauge data analysis. Paper SPE-187181-MS Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 9-11 October, 2017.
- [87] C. Tian and R. Horne. Applying machine-learning techniques to interpret flow-rate, pressure, and temperature data from permanent downhole gauges. *SPE Reservoir Evaluation Engineering*, 22(2):386–401, 2019.
- [88] J. von Neumann and O. Morgenstern. Theory of games and economic behavior. Princeton University Press, 1947.
- [89] C. Wang, G. Li, and A. Reynolds. Production optimization in closed-loop reservoir management. *SPE Journal*, 14(3):506–523, 2009.
- [90] D. Wang, Y. Li, B. Chen, Y. Hu, B. Li, D. Gao, and B. Fu. Ensemble-based optimization of interwell connectivity in heterogeneous waterflooding reservoirs. *Journal of Natural Gas Science and Engineering*, 38:245–256, 2016.
- [91] D. Watson. Computing the n-dimensional delaunay tessellation with application to voronoi polytopes. *The Computer Journal*, 24:167–72, 1981.
- [92] S. Wessing. *Two-stage methods for multimodal optimization*. PhD thesis, Universitätsbibliothek Dortmund, 2015.
- [93] B. Xu, R. Huang, and M. Li. Revise saturated activation functions. Paper presented at the International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May., 2016.
- [94] Q. Zhang and K. Gupta. Neural networks for rf and microwave design. Artech House, 2000.
- [95] D. Zingg, M. Nemec, and T. Pulliam. A comparative evaluation of genetic and gradient-based algorithms applied to aerodynamic optimization. *European Journal of Computational Mechanics*, 17(1-2):103–126, 2008.

- [96] A. Zolan, J. Hasenbein, and D. Morton. Optimizing the design of a latin hypercube sampling estimator. Paper Presented at the 2017 Winter Simulation Conference, Las Vegas, NV, 3-6 December, 2018.