Saumitra Dwivedi

Application of Artificial Intelligence and Data Science Methods in Oil and Gas Reservoir Simulation

Master's in Simulation and Visualization
Supervisors: Ricardo Da Silva Torres
Co-Supervisors: Ibrahim A. Hameed, Guillaume Suzanne

June 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of ICT and Natural Sciences

NTNU
Innovation and Creativity

NTNU

Norwegian University of Science and Technology

Master's in Simulation and Visualization

Faculty of Information Technology and Electrical Engineering
Department of ICT and Natural Sciences

© 2020 Saumitra Dwivedi. All rights reserved

ISBN (printed version)
ISBN (electronic version)
ISSN 1503-8181

Master's thesis at NTNU,

Printed by NTNU-trykk

Summary

The oil and gas industry works with a few of the most complex and challenging industrial processes across several domains of engineering. Such industrial processes usually require precision, in order to suppress the outcome error magnitude and volatility, thus ensuring effective and safe hydrocarbon production. To properly monitor and control such processes, this industry deals with vast amounts and various sources of data. For instance, it is highly important for an oil and gas company to accurately predict the oil and gas fluid flow inside the subsurface reservoir, so as to make profitable business decisions (e.g. where and how deep to drill an oil producing well or a water injection well? with which kind of well characteristics?). For this purpose, reservoir data like pressure, amounts of production fluids (oil, water and gas) are frequently monitored at the wells.

Conventionally, such oil and gas reservoir data are used to build complex and computationally expensive, physics-based numerical simulation models to account for geological heterogeneity while representing the overall reservoir fluid flow dynamics. However, these physics-based numerical models are based on several assumptions, made to simplify the mathematical representation of fluid flow physics. Moreover, being computationally intensive, the accuracy of such models is usually challenged by the limitation of computation resources. To address these challenges, this thesis aims to explore modern techniques based on artificial intelligence (AI) and data science, in order to produce data-driven workflows to analyze, model, and simulate reservoir pressure dynamics.

In particular, two independent aspects associated with reservoir engineering and simulation were studied in this thesis. Firstly, it was investigated a data-driven workflow to model reservoir pressure at any point in space and time from sparse pressure data observed at wells, without building a physics-based numerical model. This workflow was termed as spatiotemporal modelling of reservoir pressure. Secondly, a methodology was formulated to build an AI proxy model in order to produce machine-learning-based reservoir pressure simulations as a fast alternative to con-

ventional physics-based numerical simulations.

Spatiotemporal modelling of reservoir pressure was based on a three step workflow including multivariate analysis of pressure data, pressure modelling and spatiotemporal interpolation. The overall workflow provided a comprehensive method to understand and map the reservoir fluid flow dynamics using data science tools. Several modelling techniques like generalized additive models, machine learning, spatiotemporal kriging were investigated for its applicability and accuracy as a part of this thesis. Moreover, the workflow was applied to a real oil and gas reservoir case, for which the reservoir pressure prediction accuracy was optimized through a few experiments. The optimum experiment produced highly accurate prediction with a mean absolute error of 26.85 psi. Moreover, a portion of data used was kept to evaluate blind test accuracy, which amounted to a mean absolute error of 55 psi, for the optimum case.

On the other hand, the workflow for machine-learning-based reservoir simulations was based on the idea of building a proxy model from a few sensitivity physics-based numerical simulations, using machine learning. This AI-based proxy model was then used to simulate pressure profiles, thereby replacing the physics-based numerical simulation model. The overall applicability of this workflow was aimed towards the objective of history matching. Such machine-learning-based simulations provided a fast alternative to conventional simulations, thereby providing an opportunity to simulate several scenarios to find a better history match. This would be in contrast to the conventional practice of using gradient-based optimization methods (with fewer simulations), which usually end up providing a local optimum as a history match. Additionally, several experiments were conducted to optimize for the AI-based model characteristics, based on comparing AI-based simulations and physics-based numerical simulations. The most optimized model produced a mean absolute error of 13.37 bars. Similar to ST modelling workflow described above, the AI-based proxy model's accuracy was also evaluated on test dataset which yielded a mean absolute error of 15.33 bars, for the most optimized experiment.

The proposed data-driven workflows were aimed to improve current methods of reservoir engineering and simulation. The suggested workflows showed high accuracy in reservoir pressure predictions and simulations with high efficiency in use of computational resources and time. Additionally, the proposed workflows were developed using open source libraries which pose no additional cost to computation, in contrast to extremely expensive industry standard physics-based reservoir simulation software. Moreover these workflows could also be used to model other reservoir properties like production ratios (Water cut, and Gas-Oil Ratio), contacts (Water-Oil contact and Gas-Oil contact), among others.

Acknowledgements

I would like to express my deepest gratitude to my supervisors: Prof. Ricardo da Silva Torres, Prof. Ibrahim A. Hameed, and Guillaume Suzanne. They have provided me with much needed guidance and support throughout the course of this research.

I would also like to give my special thanks to Beicip Franlab. They have provided me with access to confidential data, vital for the success of this research work. Moreover, they provided me with academic license of their oil and gas geoscience software (OpenFlow^{TM}), which formed to be an important ingredient to this thesis. Also, they coordinated and arranged for my travel to their headquarters at Paris to hold necessary discussions and presentations.

In the end, I would like to thank my family and friends for all their support and motivation.

Preface

This masters thesis is a submitted as a part of Master of Science degree in Simulation and Visualization program at Norwegian Institute of Science and Technology (NTNU), Department of ICT and Natural Science. This research work was carried out during the final semester (spring 2020) of the master program in collaboration with Beicip Franlab.[1] Beicip Franlab is an international oil and gas consulting and software solutions provider based in Paris. It provided access to all the necessary data required as a part of this research work.

This thesis explores data-driven methodologies in regards to oil and gas reservoir engineering and simulation. Several techniques based on machine learning and data science have been investigated for their applicability and overall contribution to customized workflows while dealing with oil and gas reservoir data. Myself, being a reservoir engineer, have dealt with several reservoir studies and was motivated to pursue this topic while studying modern AI, data science, and simulation methods during the course of my masters.

---

[1] `http://www.beicip.com/` (As of June 2020).

Contents

Summary **iii**

Acknowledgements **v**

Preface **vii**

Contents **xi**

List of Tables **xiii**

List of Figures **xxiii**

1 Introduction **1**

1.1 Background 1

1.2 Problem Overview 2

1.3 Objectives 3

1.4 Confidentiality 5

1.5 Thesis Structure 6

2 Background Concepts **7**

2.1 Introduction to Oil and Gas Reservoir Engineering 7
2.1.1 Reservoir Fluid Flow Simulation 10
2.1.2 Conventional Practices in the Oil and Gas Industry 11
2.2 Generalized Additive Models 13
2.3 Spatiotemporal Kriging 14
2.4 Machine Learning 16
2.4.1 Feed Forward Neural Network 16
2.4.2 Regularization 19
2.4.3 Z-score Scaling 19
2.5 Evaluation Measures 20
2.5.1 Mean Absolute Error (MAE) 20
2.5.2 R-squared (R^2) 20

3 Related Work 21
3.1 Conventional Oil and Gas Reservoir Simulation 21
3.2 Spatiotemporal Interpolation 23
3.3 Proxy Modeling for Fluid Flow in Porous Media 23

4 Spatiotemporal Modelling of Oil and Gas Reservoir Data 25
4.1 Methodology and Applied Workflow 26
4.1.1 Feature Engineering 26
4.1.2 Dynamics Modeling 27
4.1.3 Spatiotemporal Interpolation 28
4.2 Tools and Techniques 28
4.3 Applied Case: Undersaturated Oil Reservoir with Water and Gas Injection 29
4.3.1 Data Description 30
4.3.2 Experiments 31

4.3.3 Results and Discussion 32

5 Machine-Learning-based Oil and Gas Reservoir Simulation 55

5.1 Methodology and Applied Workflow 56

5.1.1 Database preparation 57

5.1.2 Machine Learning 57

5.1.3 Machine-learning-based simulation 57

5.2 Tools and Techniques 58

5.3 Applied Case: Undersaturated Oil Reservoir with Water Injection 58

5.3.1 Data Description 59

5.3.2 Experiments 60

5.3.3 Results and Discussion 61

6 Conclusion 71

6.1 Thesis Contribution 73

6.2 Future Work 74

References 85

A ML-based reservoir simulation: Experiment results 87

A.1 Results for experiment 2 87

A.2 Results for experiment 3 87

A.3 Results for experiment 4 87

A.4 Results for experiment 5 88

A.5 Results for experiment 6 88

List of Tables

4.1 Modelling accuracy for ST Modeling experiments 32

5.1 Uncertain Parameters. 59
5.2 Numerical Simulations for training dataset. 59
5.3 Numerical Simulations for test dataset. 60
5.4 Training and test accuracy of ML based simulations (at wells) . . 60

List of Figures

1.1 Research scope composed by integration of artificial intelligence, data science and oil and gas reservoir engineering. The research scope formed two parts of the thesis i.e. 1) Spatiotemporal (ST) modeling of oil and gas reservoir data and 2) Machine learning (ML) based oil and gas reservoir simulation 5

1.2 Thesis structure describing each chapter's content (the figure outline is adapted from Murthy [50]). The colored squares in the figure indicate the use of concept introduced in Chapter 2. The numbered circles indicate towards which RQ, the chapter addresses. 6

2.1 A typical petroleum trap [28]. 8

2.2 Figure on the left shows typical reservoir pressure decline profile in a reservoir with oil production (X axis idicates % of oil produced from original oil in place (OOIP)) for different drive mechanisms (water drive, gas cap drive and solution gas drive). Figure on the right shows typical production gas-oil ratio (GOR) profile, with oil production for different drive mechanisms (water drive, gas cap drive and solution gas drive) [34] 9

2.3 Supplementing natural reservoir energy by injecting water into water zone and injection gas into the gas cap [3]. 9

2.4 Integrated reservoir modelling and simulation [27]. 11

2.5 Upscaling a geological model to dynamic model [57]. 12

2.6 Figure shows approaches to history matching [53]. The plot on the left shows a classical history match workflow, which considers only fluid flow simulation parameters for history matching. The plot on the right shows an integrated approach to history matching considering geological as well as fluid flow parameters for history matching. 12

2.7 GAM basis functions [72]. 13

2.8 Generalized Additive Model complexity. 14

2.9 A typical spatiotemporal variogram [68]. 15

2.10 A perceptron unit. 17

2.11 A three layered MLP [7]. 18

2.12 Common activation functions. 18

2.13 Swish and LeakyReLU activation functions. 19

4.1 Spatiotemporal modelling workflow goal of modelling pressure observed at sparse observation locations (i.e. wells), on the left, to predict reservoir pressure in 3D space and time, on the right. . . 26

4.2 Spatiotemporal modelling workflow. 26

4.3 XY location feature engineered to a quasi-cylindrical coordinate system. 27

4.4 Top view of reservoir depth map. 29

4.5 3D model of reservoir. 30

4.6 Measured well pressure data for three sample wells (A01, A02, A03). 31

4.7 Base Case: Correlation matrix indicating Pearson correlation coefficients of pressure data against spatial and temporal features. . . . 33

4.8 Base Case: Feature multivariate analysis of reservoir pressure against spatial and temporal features evaluating heterogeneity in reservoir dynamics. 33

4.9 Figures indicating the accuracy of GAM model for Base Case in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by GAM model. The figure on the right shows predicted and actual pressures vs. date. 34

4.10 Figure shows hyperspace of pressure predicted by GAM model for Base Case against Elevation and Date, for four values of Theta (representing directions of East, North, South, and West from the centre of the oil and gas field). 35

4.11 Base Case: Correlation matrix indicating Pearson correlation coefficients of GAM modelling residuals against spatial and temporal features. 36

4.12 Figures indicating the accuracy of MLP model for Base Case in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by MLP model. The figure on the right shows predicted and actual pressures vs. Date. 37

4.13 Base Case: Correlation matrix indicating Pearson correlation coefficients of MLP modelling residuals against spatial and temporal features. 37

4.14 Base Case: Figure shows observed pressure (measured data), GAM pressure prediction (in red), and final pressure model (in blue), incorporating GAM predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in training dataset. 39

4.15 Base Case: Figure shows observed pressure (measured data), GAM pressure prediction (in red), and final pressure model (in blue), incorporating GAM predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in test dataset. 40

4.16 Base Case: Figure shows observed pressure (measured data), MLP pressure prediction (in red), and final pressure model (in blue), incorporating MLP predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in training dataset. 41

4.17 Base Case: Figure shows observed pressure (measured data), MLP pressure prediction (in red), and final pressure model (in blue), incorporating MLP predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in test dataset. 42

4.18 Figures indicating the accuracy of the GAM model for Temporal scale partitioning (pre-October 1981) in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by the GAM model. The figure on the right shows predicted and actual pressures vs. date. 43

4.19 Figure shows hyperspace of pressure predicted by the GAM model for Temporal scale partitioning (pre-October 1981) against Elevation and Date, for four values of Theta (representing directions of East, North, South and West from the centre of the oil and gas field). 44

4.20 Temporal scale partitioning (pre-October 1981): Correlation matrix indicating Pearson correlation coefficients of GAM modelling residuals against spatial and temporal features. 45

4.21 Figures indicating the accuracy of the MLP model for Temporal scale partitioning (pre-October 1981) in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by the MLP model. The figure on the right shows predicted and actual pressures vs. Date. 45

4.22 Temporal scale partitioning (pre-October 1981): Correlation matrix indicating Pearson correlation coefficients of MLP modelling residuals against spatial and temporal features. 46

4.23 Figures indicating the accuracy of the GAM model for Temporal scale partitioning (post-October 1981) in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by GAM model. The figure on the right shows predicted and actual pressures vs. Date. 47

4.24 Figure shows hyperspace of pressure predicted by the GAM model for Temporal scale partitioning (post-October 1981) against Elevation and Date, for four values of Theta (representing directions of East, North, South and West from the centre of the oil and gas field). 47

4.25 Temporal scale partitioning (post-October 1981): Correlation matrix indicating Pearson correlation coefficients of GAM modelling residuals against spatial and temporal features. 48

4.26 Figures indicating the accuracy of MLP model for Temporal scale partitioning (post-October 1981) in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by MLP model. The figure on the right shows predicted and actual pressures vs. Date. 48

4.27 Temporal scale partitioning (post-October 1981): Correlation matrix indicating Pearson correlation coefficients of MLP modelling residuals against spatial and temporal features. 49

4.28 Temporal scale partitioning: Figure shows observed pressure (measured data), GAM pressure prediction (in red), and final pressure model (in blue), incorporating GAM predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in training dataset. 50

4.29 Temporal scale partitioning: Figure shows observed pressure (measured data), GAM pressure prediction (in red), and final pressure model (in blue), incorporating GAM predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in test dataset. 51

4.30 Temporal scale partitioning: Figure shows observed pressure (measured data), MLP pressure prediction (in red), and final pressure model (in blue), incorporating MLP predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in training dataset. 52

4.31 Temporal scale partitioning: Figure shows observed pressure (measured data), MLP pressure prediction (in red), and final pressure model (in blue), incorporating MLP predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in test dataset. 53

5.1 Concept behind machine-learning-based reservoir simulation. . . . 56

5.2 Figure shows the idea behind recursive algorithm used to produce Machine-learning-based reservoir simulation. At every timestep, uncertain parameter values and pressure values simulated at previous timestep are fed into the trained ML-based surrogate reservoir model to simulate pressure. 56

5.3 Machine-learning-based reservoir simulation. 57

5.4 3D model of reservoir. 58

5.5 Experiment 1: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset 62

5.6 Experiment 1: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot). . . . 63

5.7 Experiment 1: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells. 64

5.8 Experiment 2: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 2 (training case). 65

5.9 Experiment 2: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 21, on top and No. 22, on bottom (test cases). . 66

5.10 Experiment 3: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N3) for simulation No. 20, on top and No. 21, on bottom (test cases). . . . 66

5.11 Experiment 4: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 2 (training case). 67

5.12 Experiment 4: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 21, on top and No. 22, on bottom (test cases). . 68

5.13 Experiment 5: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 20 (test case). 68

5.14 Experiment 6: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 2 (training case). 69

5.15 Experiment 6: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 21, on top and No. 22, on bottom (test cases). . 69

A.1 Experiment 2: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset. 89

A.2 Experiment 2: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot). . . . 90

A.3 Experiment 2: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells. 91

A.4 Experiment 3: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset. 92

A.5 Experiment 3: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot). . . . 93

A.6 Experiment 3: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N3) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells. 94

A.7 Experiment 4: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset. 95

A.8 Experiment 4: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot). . . . 96

A.9 Experiment 4: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells. 97

A.10 Experiment 5: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset. 98

A.11 Experiment 5: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot). . . . 99

A.12 Experiment 5: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells. 100

A.13 Experiment 6: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset. 101

A.14 Experiment 6: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot). . . . 102

A.15 Experiment 6: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells. 103

Chapter 1

Introduction

This chapter structures the motivation and background of the research domain and provides an overall understanding to the research objectives and scope.

1.1 Background

With the advent of Industrial Revolution in the early nineteenth century, fossil fuels have been a major source of energy to run machines all around the world. Essentially, development of internal-combustion engines and other devices marked the importance of petroleum and its derivatives. Oil and gas industry was born out of an accelerated growth in demand of crude oil and natural gas [3]. Consequently, over the years, exploration and production of oil and natural gas have found itself as a major research subject across several domains of specialty.

With the increase in demand for oil and gas, a field of specialization termed "Petroleum Reservoir Engineering" was developed. According to Craft and Hawkins [18], petroleum engineers have been invested in deriving gas-energy relationships and recognised the need to gather data with good precision to map the properties and conditions of the subsurface reservoirs. In overall, the main applications with regards to reservoir engineering activities can be summarized to reserves estimation, development planning, and production operations optimization [26].

The hydrocarbon reserves form the primary assets of an oil company and hence, its quantification is of paramount importance. According to Ross [62], the main goals of reserves estimation are (a) to derive the best estimate of hydrocarbon recovery through the life of oil and gas field until abandonment, and (b) to assess what is the uncertainty in that estimate. Such uncertainties are usually related to the gathered data specifically in terms of reservoir geological properties. Recent publications

[60] form a good example of effective workflows to quantify the risks, posed by such uncertainties, to the estimated reserves.

Moreover, in order to plan the development of an oil and gas field, a reservoir engineer needs to formulate optimized development strategies based on the characteristics of the oil field defined by limited available data, before the start of drilling and production in the oil field [70]. Nowadays, usually such field development strategies are based on simulating several scenarios using physics-based oil and gas reservoir numerical simulation models [4]. Such numerical simulation models describe 3D multiphase fluid flow in heterogeneous media by solving large sets of finite differential equations [17]. The current widespread use of these numerical simulation models in the reservoir engineering community can be accounted to the applicability and versatility of these models to tackle a variety of problems, ease of use (usually facilitated by industry standard software) and acceptance in the oil and gas industry as a standard practice, by not only engineers, but also high level management [17].

1.2 Problem Overview

However, these numerical models should only be employed when there is an expected worth in pursuing such a computationally expensive method [17]. According to Coats [13],

"In reservoir simulation, the question is not whether, but how and how much.
The complexity of the questions being asked, and the amount and reliability of the data available, must determine the sophistication of the system to be used."

In spite the widespread popularity of conventional reservoir simulation methods, these methods can have limitations and can be easily misused [13]. This thesis attempts to address the following problems associated with conventional physics-based oil and gas reservoir simulation:

- **Accuracy of simulation results:** The conventional reservoir numerical simulation methods are based on several numerical and analytical assumptions which in-turn engender additional uncertainties [69]. Such assumptions can lead to highly counterproductive numerical errors, thus compromising the accuracy of the simulation results.
- **Uncertainty:** Reservoir simulation studies are usually subject to a degree of uncertainty, arising from the incomplete/misrepresented knowledge of reservoir geology, validity of input data, choice of numerical parameters, etc [17].

- **Available Resources:** Such conventional physics-based reservoir numerical simulations are highly expensive in terms of computational resources and technical expertise [17].

This thesis aims to use data-driven techniques, which are not based on physics-based assumptions, and are capable of analyzing and modelling a dynamic system using only the observed data. In this thesis, techniques related to artificial intelligence and data science have been explored, by taking into account their applicability and accuracy in order to model the pressure data gathered from the oil field. Using techniques related to artificial intelligence and data science with regards to reservoir engineering and simulation has a huge potential in terms of analysis and modelling capability of a dynamic system. Such techniques have already proven themselves in other domains, such as temperature modelling [38], air pollution modelling [65] among others. However, rarely these techniques have been applied to reservoir engineering problems [23, 2]. But such studies deal with different dataset in terms of geological setting and reservoir fluid flow dynamics.

This thesis offers a more comprehensive application of AI and data science techniques, with a quantified comparison between various modelling technique configurations. According to Dake [21], reservoir engineering being one of the geosciences, attempts to map the wide spaces of the reservoir between the sparse points of observation (i.e., the wells). The proposed data-driven methods in this thesis help to better understand the reservoir behaviour away from the wells without constructing a 3D physics-based numerical model of the reservoir.

1.3 Objectives

The scope of this thesis is encompassed by three disciplines (oil and gas reservoir engineering, artificial intelligence, and data science) coming together to create data-driven methods of analyzing, modelling, and simulating oil and gas reservoir dynamics (see Fig. 1.1).

According to Dake [21], accurately modelling the dynamics of the reservoir between the wells is the distinctive characteristic of petroleum reservoir engineering. In general, this thesis aims to explore data driven methods in order to best exploit the data measured (specifically reservoir pressure data) across the wells in the oil field and also to find ways to improve the current conventional practices of reservoir engineering.

In support of the above scope, the following research questions were studied and addressed:

- **Q1.** Would data science methods be effective to identify and analyze well pressure heterogeneity and associated geological uncertainties, in spatial and temporal context for a long time ranged (in order of decades) and heterogeneous oil and gas reservoir dynamics?
  Analyzing well pressures by taking into account its heterogeneity in spatial and temporal contexts can be done through feature engineering followed by multivariate correlation analysis. The evolution of well pressures can be a function of space (or different volumes of reservoir) or/and time. This function can be mapped by engineering spatial and temporal information from the wells and then analyzing their (feature's) individual impact through computation of correlations (or correlation matrix). This kind of analysis help to identify and map reservoir geological and fluid flow heterogeneity and associated geological uncertainties.

- **Q2.** How data science methods could be used to simulate pressure in 3D space and time (within the spatial and temporal domain of observed pressure data at oil and gas wells) without constructing a physics-based reservoir numerical simulation model?
  Spatiotemporal (ST) modelling of reservoir pressure from sparse well bottom hole pressure data can be done by modelling pressure as a function of spatial and temporal features. Modeling techniques like generalized additive models (GAM) or machine learning can be used to do the same. Moreover, spatiotemporal kriging [19, 78] can be used to model the pressure residuals generated by the above mentioned modelling techniques in order to capture localised effects in space and time. In this way, reservoir pressure can be accurately mapped for entire oil and gas reservoir space and at any time, without constructing a physics-based simulation model.

- **Q3.** Would machine learning based techniques like multi-layer perceptron (MLP) improve the current reservoir simulation practices, specifically for the scope of history matching of an oil and gas reservoir simulation model?
  The current practices of reservoir fluid flow simulation is based on solving partial differential equations over several timesteps, which are highly time consuming and computationally expensive. This can be improved by using machine learning to build surrogate reservoir model from a few physics-based numerical simulations, to obtain accurate (if not precise) and very fast fluid flow simulations. Such a fast alternative to physics based numerical simulations, provides an opportunity to simulate several scenarios to obtain a global optimum as the history match, instead using conventional optimization methods like gradient based optimization which usually provide local optimum.

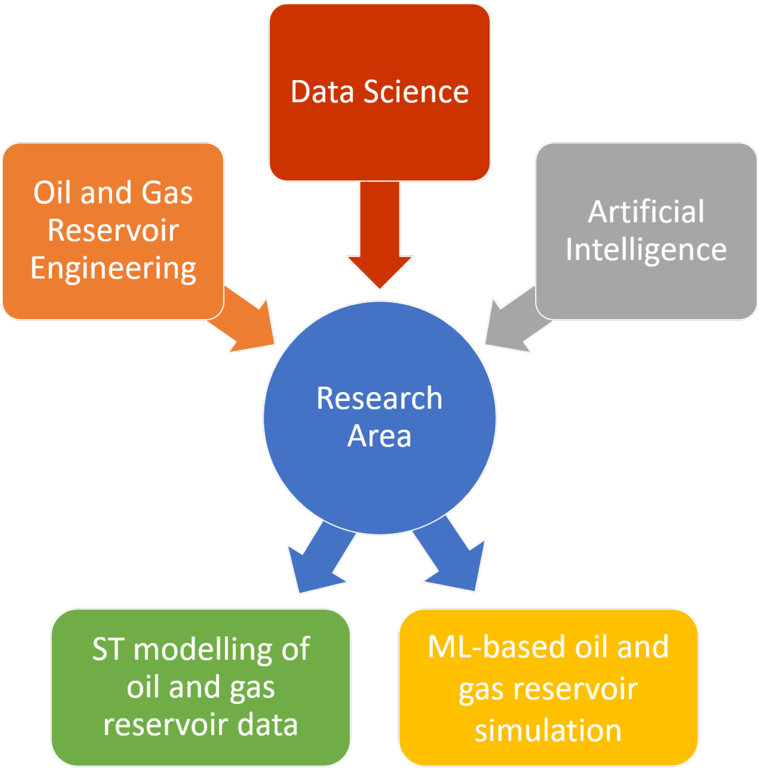
Data Science
Oil and Gas Reservoir Engineering
Artificial Intelligence
Research Area
ST modelling of oil and gas reservoir data
ML-based oil and gas reservoir simulation

Figure 1.1: Research scope composed by integration of artificial intelligence, data science and oil and gas reservoir engineering. The research scope formed two parts of the thesis i.e. 1) Spatiotemporal (ST) modeling of oil and gas reservoir data and 2) Machine learning (ML) based oil and gas reservoir simulation

1.4 Confidentiality

The project was carried out as a collaboration between Norwegian University of Science and Technology (NTNU) and Beicip Franlab. A confidentiality agreement (non-disclosure agreement) was signed between all parties concerned regarding the privacy of the data used in this research.

The oil and gas field data provided by Beicip Franlab cannot be fully disclosed in order to protect the company's proprietary information and knowledge. Consequently, the information on geo-location and the well names is not provided. However, none of the changes in data ordinals and redaction affected the versatility of research methodology and accuracy of modelling results.

1.5 Thesis Structure

This thesis's structure is aimed towards a seamless information flow as outlined by Figure 1.2. The same can be described as follows:

Chapter 2 - Background Concepts: This chapter describes the background concepts related to the thesis. It includes theory related to oil and gas reservoir engineering and simulation, generalized additive modelling, spatiotemporal interpolation, and machine learning.

Chapter 3 - Related Work: This chapter summarizes the research related to oil and gas reservoir simulation, spatiotemporal modeling, and surrogate reservoir modeling (or proxy modeling) in various journals and other sources of literature.

Chapter 4 - Spatiotemporal Modelling of Oil and Gas Reservoir Data: This chapter describes the project methodology and results in regards to ST Modelling of oil and gas reservoir pressure data from a real oil and gas field.

Chapter 5 - Machine learning based Oil and Gas Reservoir Simulation: This chapter describes method and results of surrogate reservoir modeling and its use to simulate fluid flow in a real oil and gas field.

Chapter 6 - Conclusion: This chapter concludes the research work by reporting answers to the research questions and thesis contributions. Also, it sheds light on possible future work regarding the use of data-driven methods on oil and gas reservoir simulation.

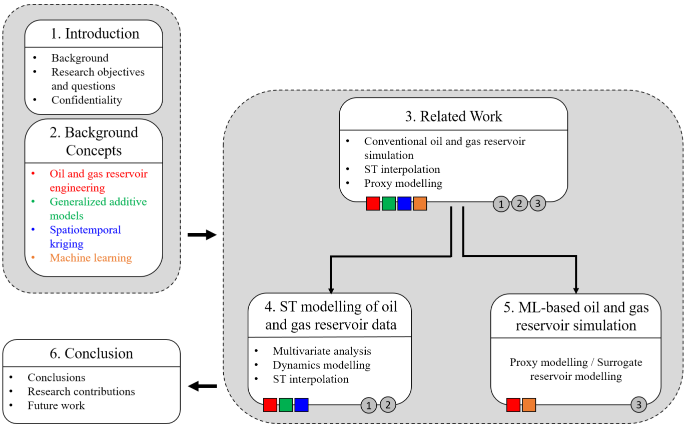
1. Introduction
Background
Research objectives and questions
Confidentiality
2. Background Concepts
Oil and gas reservoir engineering
Generalized additive models
Spatiotemporal kriging
Machine learning
3. Related Work
Conventional oil and gas reservoir simulation
ST interpolation
Proxy modelling
1 2 3
4. ST modelling of oil and gas reservoir data
Multivariate analysis
Dynamics modelling
ST interpolation
1 2
5. ML-based oil and gas reservoir simulation
Proxy modelling / Surrogate reservoir modelling
3
6. Conclusion
Conclusions
Research contributions
Future work

Figure 1.2: Thesis structure describing each chapter's content (the figure outline is adapted from Murthy [50]). The colored squares in the figure indicate the use of concept introduced in Chapter 2. The numbered circles indicate towards which RQ, the chapter addresses.

Chapter 2

Background Concepts

This chapter provides an overview of technical concepts used in this thesis. The explained concepts include oil and gas reservoir engineering, spatiotemporal kriging, generalized additive models, and machine learning.

2.1 Introduction to Oil and Gas Reservoir Engineering

Petroleum is a mineral produced from the earth. The word ‘petroleum’ comes from the Latin roots of petra, meaning “rock” and oleum meaning “oil" [3]. It is found in liquid or gaseous form, deep beneath the earth surface, contained in the pore spaces or interstices of rock materials. These rock materials are referred to as reservoir rocks [3].

The oil and gas accumulations are highly localised, formed under a specific geological environment, i.e., petroleum deposits do not underlie all the surface of the earth but only a few places usually well deep inside the earth’s crust [3]. Crude oil and natural gas are produced by several well bores, drilled to such reservoir rocks [3].

A typical oil and gas reservoir is a result of entrapment of hydrocarbons in a reservoir rock with overlying rock formations of low permeability also known as cap rock as shown in Figure 2.1 [3].

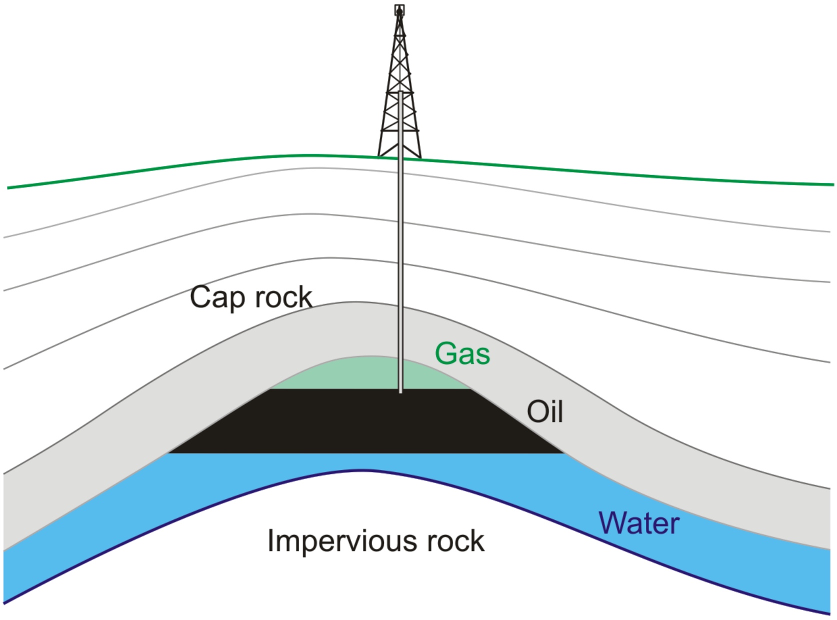

Figure 2.1: A typical petroleum trap [28].

Such oil and gas formations are usually associated with high pressure environment due to the considerable depth which naturally spurts the hydrocarbons from well bore to the surface. Additionally, there are several other energy sources in play bringing the hydrocarbons to the wellbore from farther reaches of the reservoir. It is one of a major role of a reservoir engineer to study various aspects of such energy sources and determine reservoir recovery mechanisms and performance.

Several energy sources exist in the reservoir formation, one of which is expansive energy of the hydrocarbon fluid. The wellbore acts as a pressure sink and the expansive energy of the hydrocarbon fluid at higher pressure leads to migration/flow of fluid towards the wellbore through the permeable reservoir rock [3]. Additionally, hydrocarbon bearing formations are associated with surrounding deeper water bearing formations. This leads to the force of gravity promoting segregation of the various fluids, i.e., water, oil, and gas [3].

For a proper understanding of reservoir behavior and predicting future performance, it is necessary to have knowledge of the driving mechanisms that control the behavior of fluids within reservoirs. The overall performance of oil reservoirs is largely determined by the nature of the energy, i.e., driving mechanism, available for moving the oil to the wellbore [3]. Moreover, the various energies associated with reservoir drive mechanisms are quantified and monitored by studying reservoir pressure depletion profile and production profile of hydrocarbons. Figure 2.2 illustrates pressure and production profiles of different drive mechanisms showing the effect of different energy sources [3].

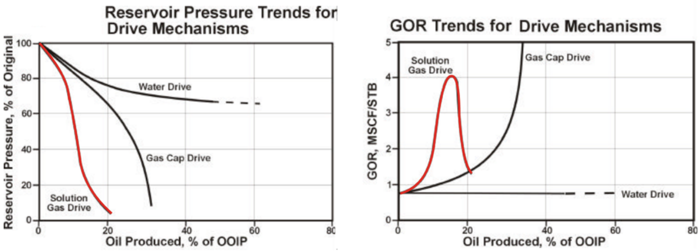

Figure 2.2: Figure on the left shows typical reservoir pressure decline profile in a reservoir with oil production (X axis idicates % of oil produced from original oil in place (OOIP)) for different drive mechanisms (water drive, gas cap drive and solution gas drive). Figure on the right shows typical production gas-oil ratio (GOR) profile, with oil production for different drive mechanisms (water drive, gas cap drive and solution gas drive) [34]

It is highly important not to deplete too much, reservoir pressure or natural energy of the reservoir for higher recovery of hydrocarbons. Based on the reservoir rock and fluid characteristics, the natural energy of the reservoir can also be maintained or supplemented by injecting water or gas back into the reservoir via injection wells [3]. Figure 2.3 illustrates how water injection and gas injection wells can maintain or increase reservoir pressure.

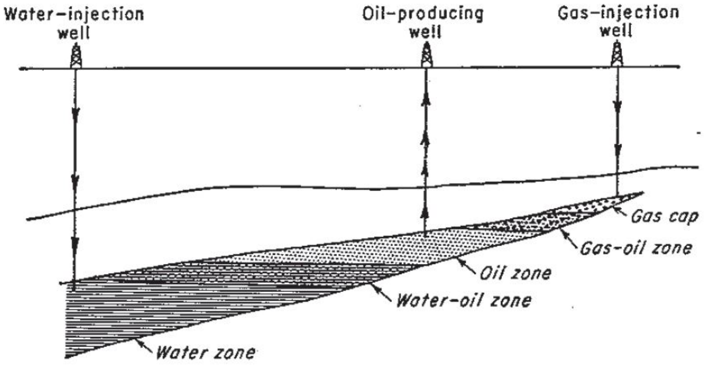

Figure 2.3: Supplementing natural reservoir energy by injecting water into water zone and injection gas into the gas cap [3].

It is paramount for an oil and gas company to be able to predict accurately oil and gas fluid flow in a reservoir, in order to make profitable business decisions and mitigate risk. With time, several analytical methods were formulated to accurately predict production profiles across all wells in an oil and gas field.

During the 1960s, the terms “reservoir simulation” and “reservoir mathematical modelling” became popular. These synonymous terms refer to the ability to use mathematical formulations to predict reservoir performance via sophisticated numerical methods. The overall idea was to be able to solve large number of equations using finite-difference or finite-element techniques [14].

Reservoir simulation became especially popular with rise of high performance computing, thus enabling reservoir engineers to solve partial differential equations for several timesteps in a reasonable duration and computation cost. Section 2.1.1 further describes details about reservoir fluid flow simulation.

2.1.1 Reservoir Fluid Flow Simulation

A reservoir simulation modelling essentially integrates data from several sources, such as seismic, well-logs, outcrop analog data, rock core analysis, and fluid composition analysis, to build a dynamic mathematical model, accurately characterizing complex geological features and fluid flow mechanisms. In overall, the applications of reservoir simulation model can be described by (but are not limited to) the following:

- Formulate business plans and mitigate corporate risks. Due to volatile oil and gas markets, the production forecasts help to evaluate cash flows and margin policies [5].

- Approach reservoir development with multi-granular domains. There are a number of ways by which we can model a reservoir. For example, a reservoir simulation model may consider a single well, a group of wells, or several wells interacting as a complex system [22].

- Estimate oil and gas reserves with high accuracy. The capability of a reservoir simulation model to effectively capture spatial heterogeneity provides a precise estimation of oil and gas volumes in place.

- Achieve effective reservoir management. Simulation model facilitates accurate interpretation of reservoir behavior through study of mechanics of fluid flow in porous media.

- Optimize field development plans. This is achieved by simulating several possible forecast scenarios and choosing the best while mitigating the controllable risks [5].

- Apply advanced techniques for pressure maintenance and enhanced oil recovery.

2.1.2 Conventional Practices in the Oil and Gas Industry

Currently, conventional practices revolve around using data from several sources to form an integrated approach to build a comprehensive simulation model. Figure 2.4 depicts a comprehensive approach to integrated reservoir modelling and simulation [15], consisting of several steps including seismic data processing, structural modelling, stratigraphic modelling, geological modelling, and fluid flow simulation. However, data uncertainties have to be accounted at every step building up to fluid flow simulation model. Mostly these uncertainties are addressed during history matching process as explained in the next paragraph.

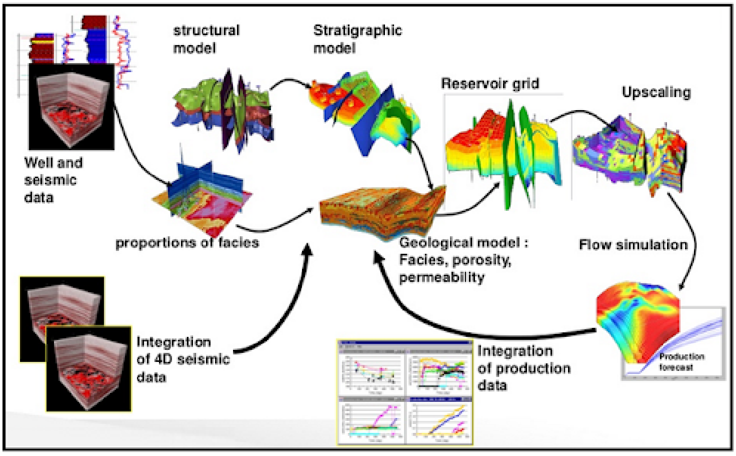
structural model
Stratigraphic model
Reservoir grid
Upscaling
Well and seismic data
proportions of facies
Geological model : Facies, porosity, permeability
Flow simulation
Integration of 4D seismic data
Integration of production data
Production forecast

Figure 2.4: Integrated reservoir modelling and simulation [27].

History Matching

According to Okotie and Ikporo [54], *"To develop a model that cannot accurately predict the past and present performance of a reservoir within a reasonable engineering tolerance of error is not a good model for predicting the future performance of the same reservoir. Hence, history matching is a process of adjusting key properties of the reservoir model to fit or match the actual historic or field data. It helps to identify the weaknesses in the available field data, it improves the reservoir description and forms the basis for the future performance predictions."*

However, these adjustments should be made in a geologically consistent manner [59]. History matching is a highly challenging task. For instance, upscaling a geological model (to reduce the computational cost) [16], from a fine model to a coarser model (see Figure 2.5) can produce challenges in regards to losing fine geological features around the wells making it difficult to accurately match the observed reservoir behaviour around the same well [22].

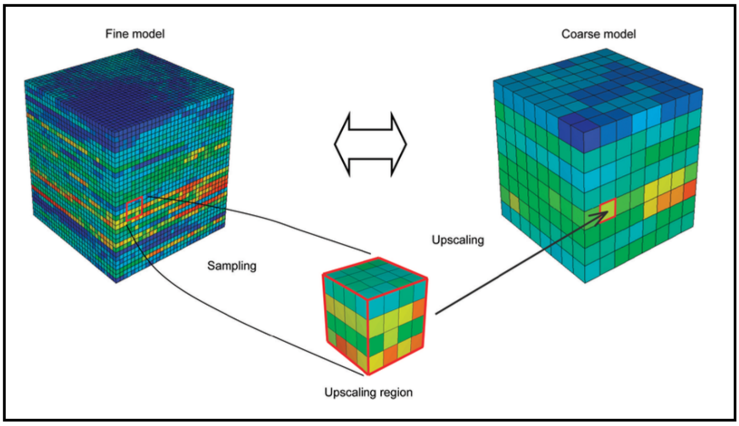
Fine model
Coarse model
Upscaling
Sampling
Upscaling region

Figure 2.5: Upscaling a geological model to dynamic model [57].

With time, the reservoir engineers have progressed in formation of more sophisticated workflows to achieve better history match. For instance, while the classical history matching workflow takes into account only the uncertainties from dynamic data sources while keeping the geological model unchanged during the process of history matching; more integrated workflows includes geological model into the loop (see Figure 2.6) [15].

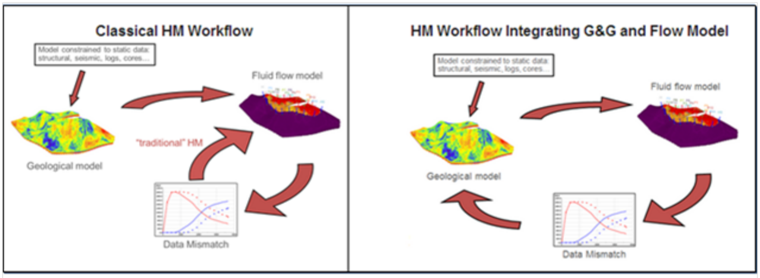
Classical HM Workflow
Model constrained to static data: structural, seismic, logs, cores...
Fluid flow model
"traditional" HM
Geological model
Data Mismatch
HM Workflow Integrating G&G and Flow Model
Model constrained to static data: structural, seismic, logs, cores...
Fluid flow model
Geological model
Data Mismatch

Figure 2.6: Figure shows approaches to history matching [53]. The plot on the left shows a classical history match workflow, which considers only fluid flow simulation parameters for history matching. The plot on the right shows an integrated approach to history matching considering geological as well as fluid flow parameters for history matching.

All such conventional practices (classical as well as integrated) aim to achieve better match to the historical data at wells.

2.2 Generalized Additive Models

Generalized additive model (GAM) was invented by Trevor Hastie and Robert Tibshirani in 1986 [35]. GAM is a flexible statistical modelling method, which effectively, captures the effects of each input variables on an outcome variable. This modelling technique is a form of generalized linear models [40] where the linear predictors are a parametric or non-parametric function of individual or transformed predictor variables [35].

Equation 2.1 accounts for modelling i^{th} data point to a univariate response/prediction y_i with relation to predictor variables x_{ij}. The f_j functions can be non-parametric smooth functions, or can be a group of specified parametric functions (e.g., polynomial, or spline) [35].

$$y_i = \alpha + \sum_j f_j(x_{ij}) + \epsilon_i \tag{2.1}$$

where

$$\epsilon_i \sim N(0, \sigma^2)$$

In general, a smooth function $s(x)$ is a sum of number of basis functions $\beta_j(x)$ weighted by regression coefficients γ_j as given in Equation 2.2. These regression coefficients are determined by penalised regression [44]. Figure 2.7 depicts the first ten basic basis functions.

$$s(x) = \sum_j \beta_j(x)\gamma_j \tag{2.2}$$

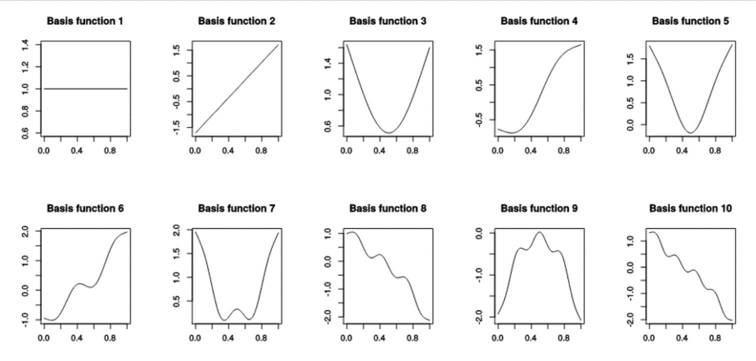

Figure 2.7: GAM basis functions [72].

The advantages of using GAM for modelling can be summarized to interpretability, flexibility/automation, and regularization. The GAM framework includes several smooth functions linked to an additive function. Such additive models provide accurate interpretation and quantification of impact from individual predictor variables, to the response. Thus, the relations of variables and predictions can be understood very easily, contrary to those in complex machine learning models. Additionally, GAM models can be changed for its degree of non-linearity by defining number of basis functions in the smooth functions, thus providing flexibility while modelling. Moreover, the nature of smooth functions can also be changed to parametric functions. Regularization techniques like L2 regularization and Bayesian regression, penalize the smoothness and prevent overfitting [44].

Thus GAM provides a modelling technique in middle of simple linear regression and highly complex "black box" machine learning (see Figure 2.8) [44].

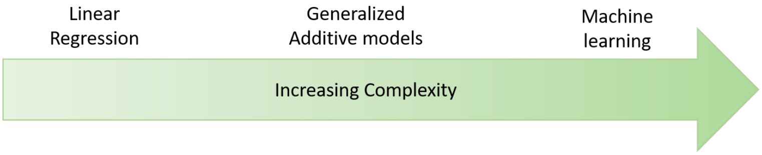

Figure 2.8: Generalized Additive Model complexity.

2.3 Spatiotemporal Kriging

Spatial random fields can be predicted based on observations at sparse spatial points using methods like Kriging. Kriging is an interpolation method devised by Georges Matheron in 1960, based on the Master's thesis of Danie G. Krige [46]. Kriging methods have been routinely used for interpolation in geostatistics, giving its uses in geology, mining, hydrology, and atmospheric sciences [39]. Such methods are based on assumption that processes which are spatially closer and more strongly related than processes far apart (Tobler's law) [66]. Kriging is based on the principles of variography which revolves around the calculation of variogram and co-variogram based on the equations 2.3 and 2.4 [39].

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{(i,j)\in N(h)} (Z(x_i) - Z(x_j))^2 \tag{2.3}$$

$$C(h) = \frac{1}{|N(h)|} \sum_{(i,j)\in N(h)} (Z(x_i) - m(h))(Z(x_j) - m(h)) \tag{2.4}$$

where Z is the property to be interpolated, $m(h) = \frac{1}{|N(h)|}\sum_{(i,j)\in N(h)}(Z(x_i) + Z(x_j))$, and $N(h)$ denotes the set of pair of observations i, j such that $|x_i - x_j| = h$ and $|N(h)|$ is the number of pairs in the set.

For processes that vary in space as well as time, spatiotemporal interpolation is more effective that purely spatial interpolation. In spatiotemporal interpolation, observations are considered in spatial as well as temporal context. Publications like Cressie and Wikle [19] and Sherman [63] form the ground work for spatiotemporal statistics.

Spatiotemporal Kriging variography is quite similar to purely spatial kriging, with one major difference. Time is also considered as a dimension for variogram and co-variogram computation as given by equations 2.5 and 2.6 [78, 32].

$$\gamma(h,u) = \frac{1}{2|N(h,u)|}\sum_{(i,j)\in N(h,u)}(Z(x_i,t_i) - Z(x_j,t_j))^2 \tag{2.5}$$

$$C(h,u) = C_s(h) \otimes C_t(u) \tag{2.6}$$

The co-variance model used as per Equation 2.6 is also known as separable model which simply incorporates the separability of spatial and temporal components, composed together by Kronecker product [29, 32]. A Kronecker product is an operation on two arbitrary sized matrices, in order to produce a tensor product block matrix [74]. Figure 2.9 depicts a typical spatiotemporal variogram showing typical variance profiles with different spatial distances and temporal time-lags. The left part of the figure shows computed variances on observed data and the right part shows the fitted variogram model on the observed data.

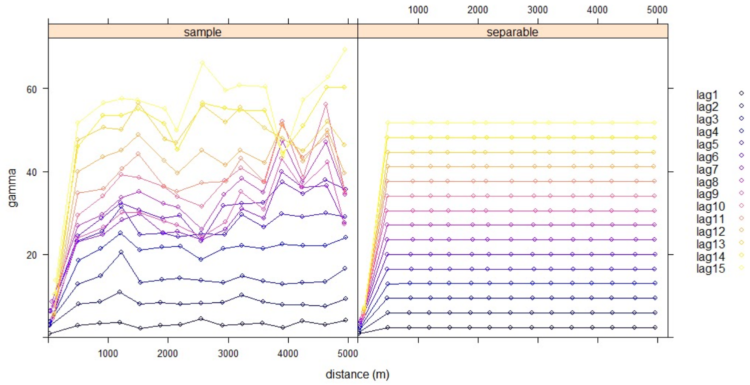

Figure 2.9: A typical spatiotemporal variogram [68].

2.4 Machine Learning

Machine learning is a part of Computer Science which is concerned with building algorithms based on a collection of data examples of some phenomenon [7]. These algorithms are adaptive in nature, in order to find patterns in the data which later can be used to make accurate predictions of the given phenomenon [42]. This process of adapting the algorithms to produce correct outputs for the collection of inputs is typically called "learning" or "training" [7].

Following describes the types of machine learning [7]:

- **Supervised Learning:** In supervised learning, the algorithm learns the relationship between a set of input features (or input feature vector) and a target output label. The dataset used to train the model is a collection of labeled examples (or instances) $(x_i, y_i)_{i=1 \to N}$ where each x_i is the input feature vector and y_i is the label. After training the model, the same can be used to make predictions on a new set of instances, hence establishing the worth of the algorithm.
- **Unsupervised Learning:** In unsupervised learning, the training data is a set of unlabeled instances $(x_i)_{i=1 \to N}$ which the algorithm uses to find patterns in terms of clustering, dimensionality reduction or outlier detection.
- **Reinforcement Learning:** In reinforcement learning, the learning is based on a system of reward or punishment. The algorithm is provided by a score to every prediction and consequently the algorithm adapts to learn a policy.

In this thesis, only supervised learning was used for different problems. Application of supervised learning is broadly to two types of problems [45]:

- **Regression:** In such problems, the algorithm is supposed to predict a number/real value to a given set of input feature vector.
- **Classification:** In this case, the algorithm tries to predict outputs, discrete in nature, i.e., a set of classes, to the given inputs.

2.4.1 Feed Forward Neural Network

Widely, machine learning algorithms are built on artificial neural networks (ANN) which is based on biological neurons connected together to communicate and process information [45]. Feed forward neural network is one of the simplest type of ANN which is a mathematical function $y = f_{NN}(x)$. The function f_{NN} is a nested function, characterized by the number of layers. For instance, a three layer network can be given as [7]:

$$y = f_{NN}(x) = f_3(f_2(f_1(x))) \tag{2.7}$$

where

$$f_i(z) = g_i(W_i z + b_i) \tag{2.8}$$

Here, i denotes the layer index and the function g_i is called an activation function [7]. W_i is referred to as weight matrix and b_i as bias vector. Such a network is also called multi layer perceptron (MLP).

The above equations also define the characteristic of individual perceptron units which has inputs x_1, x_2, x_3, ... , x_m and outputs y as shown in Figure 2.10.

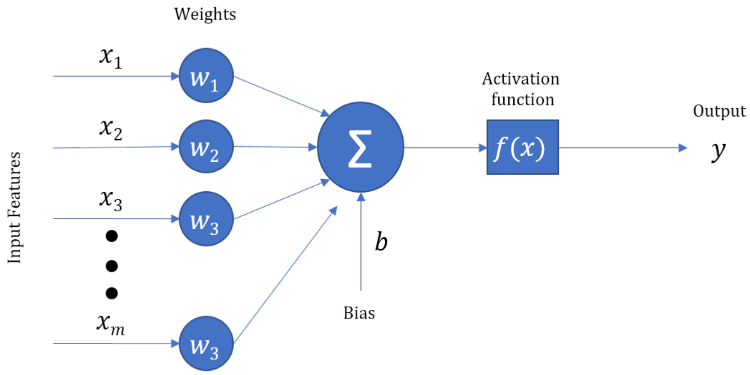
Weights
x_1
x_2
x_3
x_m
w_1
w_2
w_3
w_3
Input Features
Σ
b
Bias
Activation function
$f(x)$
Output
y

Figure 2.10: A perceptron unit.

Considering a multi-layer perceptron, the ANN is built by several units (perceptrons) organised into layers. Considering a MLP of 3 layers, Figure 2.11 depicts the internal computations and communications of the network [7].

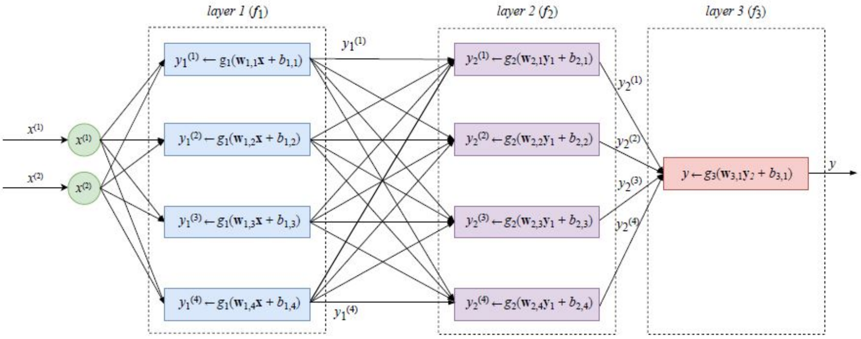
layer 1 (f_1)
layer 2 (f_2)
layer 3 (f_3)
$x^{(1)}$
$x^{(2)}$
$y_1^{(1)} \leftarrow g_1(\mathbf{w}_{1,1}\mathbf{x} + b_{1,1})$
$y_1^{(2)} \leftarrow g_1(\mathbf{w}_{1,2}\mathbf{x} + b_{1,2})$
$y_1^{(3)} \leftarrow g_1(\mathbf{w}_{1,3}\mathbf{x} + b_{1,3})$
$y_1^{(4)} \leftarrow g_1(\mathbf{w}_{1,4}\mathbf{x} + b_{1,4})$
$y_1^{(1)}$
$y_1^{(4)}$
$y_2^{(1)} \leftarrow g_2(\mathbf{w}_{2,1}\mathbf{y}_1 + b_{2,1})$
$y_2^{(2)} \leftarrow g_2(\mathbf{w}_{2,2}\mathbf{y}_1 + b_{2,2})$
$y_2^{(3)} \leftarrow g_2(\mathbf{w}_{2,3}\mathbf{y}_1 + b_{2,3})$
$y_2^{(4)} \leftarrow g_2(\mathbf{w}_{2,4}\mathbf{y}_1 + b_{2,4})$
$y_2^{(1)}$
$y_2^{(2)}$
$y_2^{(3)}$
$y_2^{(4)}$
$y \leftarrow g_3(\mathbf{w}_{3,1}\mathbf{y}_2 + b_{3,1})$
y

Figure 2.11: A three layered MLP [7].

There are several types of activation functions used when dealing with MLP. Equations 2.9, 2.10, 2.11 and Figure 2.12 summarize the most common activation functions [7]:

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{2.9}$$

$$relu(x) = max(0, x) \tag{2.10}$$

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.11}$$

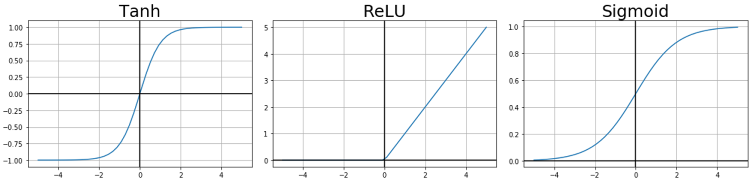

Figure 2.12: Common activation functions.

Recently, activation functions like LeakyReLU and Swish functions have proved to be even more effective. Equations 2.12, 2.13 and Figure 2.13 define those functions.

$$swish(x) = \frac{x}{1 + e^{-x}} \tag{2.12}$$

$$LeakyReLU(x) = \left\{ \begin{matrix} x\ if\ x >= 0 \\ alpha \times x\ if\ x < 0 \end{matrix} \right\} \tag{2.13}$$

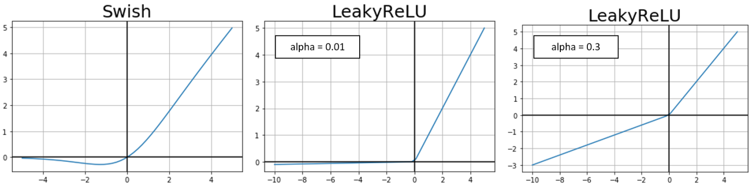

Figure 2.13: Swish and LeakyReLU activation functions.

For regression problems, the last layer of MLP has linear activation functions.

2.4.2 Regularization

While dealing with machine learning models, one of the biggest challenge is the problem of overfitting. It is referred to as the model being too complex to correctly represent the pattern of the phenomenon. In practice, dropout method is widely used as a regularization technique to prevent overfitting. By using dropout the network is made to train different input representations by "switching off" a share of perceptrons in each training stage while the remaining perceptron units shall be updated [7].

2.4.3 Z-score Scaling

An important aspect in feature engineering while dealing with machine learning, is scaling of the feature values. This scaling is done in order to change the feature values range to a common scale. The reason behind the necessity of such scaling is to increase the speed of learning while training of the machine learning model. The training with unscaled features shall put more weight on higher ordered features [7].

One of such scaling methods is standardization method, also called Z-score scaling method. In this type of scaling, the features are scaled in such a manner that, the scaled values have a mean = 0 and standard deviation = 1. The following equation describes the scaling method formulation:

$$\hat{x} = \frac{x - \mu}{\sigma} \tag{2.14}$$

where x is the feature, μ is the mean of the feature, σ is the standard deviation of the feature, and $\hat{x}$ is the scaled feature values.

2.5 Evaluation Measures

In this thesis, the modelling accuracies have been quantified by statistical measures including mean absolute error (MAE) and R-squared values, as described below:

2.5.1 Mean Absolute Error (MAE)

The mean absolute error (MAE) evaluates the accuracy of a model prediction by providing a measure of error between two the model predictions and ground truth data [75]. It is calculated based on the Equation 2.15 [75].

$$MAE = \frac{\sum\limits_{i=1}^{n} |y_i - x_i|}{n} \tag{2.15}$$

where, y_i is the model prediction, x_i is the ground truth data and n is the number of observations.

2.5.2 R-squared (R^2)

The R-squared value (R^2) or coefficient of determination provides a measure of a model's ability to replicate ground truth data, as prediction, based on the proportion of total variation of ground truth data explained by the model predictions [73]. It is calculated as given in Equation 2.16 [73].

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{n} (y_i - x_i)^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \tag{2.16}$$

where, y_i is the model prediction, x_i is the ground truth data and n is the number of observations.

Chapter 3

Related Work

Oil and gas reservoir modelling and simulation is a widely researched topic, as it addresses one of the most challenging industrial processes in the world. In this chapter, research carried out in the industry aimed at reservoir simulation has been summarized. Moreover, research related to data science and artificial intelligence techniques, relevant to this thesis are discussed as well.

The aim of this chapter is to summarize the contributions from various researchers across the above mentioned domains, with a prospect to understand the scope of amalgamation of these techniques, which in turn forms the basis of this thesis.

3.1 Conventional Oil and Gas Reservoir Simulation

The task of understanding and predicting reservoir behaviour has been an ever present challenge in oil and gas reservoir engineering. Till early 1960's, only analytical techniques like Material balance equation (MBE) [21] was the conventional workflow towards understanding the oil and gas reservoir. This method was based on the law of conservation of matter, applied to hydrocarbon reservoirs with the prospect of quantitative analysis of reservoir behaviour. The main purpose of this technique was to understand the main drive mechanisms acting in the reservoir and also determine relative impact of each drive [15]. Additionally, it also provided an insight to the consistency of the data points.

With favourable conditions, such analytical techniques can provide in-depth insight to reservoir dynamics analysis, such as water and gas front advance [15]. Havela et al. [37] provide a good example of workflows, implementing additional constraints to the MBE to ensure the validity of such favourable conditions. In it, a straight line method has been introduced to implement an additional constraint to the MBE in order to inculcate a much more dynamic interpretation of reservoir behaviour to the MBE. Recent publications like Mosobalaje et. al. [49], introduced new methods to solve the implicit MBE problem of calculation cumulative oil production with declining pressure. Despite the oldness and even with rise of new numerical methods, such analytical methods are still in use as evident by publications like Esor et. al. [25]. In that study, for example, the MBE was used to understand and better quantify the uncertainties which in turn was input to the process of history matching, thereby enhancing the numerical reservoir simulation.

However, the availability of computing resources in early 1960's gave rise to numerical reservoir simulation, which then became the mainstream methodology in reservoir engineering. These numerical simulations were based on partial differential equations, governing the conservation of components as published by Aziz and Settari [4] and Peaceman [55]. These partial differential equations were solved by the methods of finite difference, finite volume and discontinuous Galerkin techniques [11, 9, 61].

One of the main tasks while building a numerical reservoir simulation model is history matching. Only an effectively history matched model can accurately predict the production forecasts. The traditional history matching methods focus on the calibration of dynamic parameters only, however recent efforts in the industry have been towards more innovative and integrated approaches for the same. Elrafie et al. [24], for example, presented a good example of an innovative history match workflow. This workflow employed an assisted history match engine while characterizing and classifying the static and dynamic uncertainties based on its impact on history match and prediction. Such a methodology ensured a better history match while keeping the uncertainty in check.

Very recent efforts have been focused towards data-driven techniques to effectively characterize and build reservoir models. Artificial Intelligence and Data Science technology engender these novel methods to produce effective analysis and modelling of complex and multi-disciplinary data. Such methods have been used for well test interpretation [41], reservoir characterization [48], formation evaluation [47] and CO_2 sequestration and coal bed methane studies [30].

This thesis work is inspired by publications like El-Faidouzi et. al. [23], exhibiting a fine amalgamation of data science techniques and complex reservoir dynamics modelling. However, in this thesis a more comprehensive application of modern AI and data science techniques have been made to model reservoir fluid flow dynamics of a real oil and gas reservoir. Moreover, quantified comparison has been made between different modelling techniques (i.e., Generalized Additive Modelling and Machine learning) for their accuracy in such data driven workflows.

3.2 Spatiotemporal Interpolation

Prediction of random spatial fields is usually required in the areas of geostatistics, based on sampled data at sparse observation points. Kriging methods [39] are usually applied for spatial interpolations. However, for dynamic properties which vary in space and time, spatiotemporal modelling pose a much more effective modelling tool. Publications by Cressie and Wikle [19] lay the foundations for spatiotemporal interpolation. There have been many applications of spatiotemporal temporal interpolation across various domains of research [38, 65, 31].

For instance, Hengl et al. [38] published daily temperature variations across several observation points in Croatia, by applying spatiotemporal kriging. In it, the temperature was modelled as a function of several factors like geo-location (latitude and longitude), elevation, distance from the sea, time etc. Consequently the modelling residuals were subjected to spatiotemporal kriging to map localised effects in space and time. A similar workflow was proposed by Szpiro et al. [65]. Their approach predicts the air pollution levels in the area of Los Angeles by using spatiotemporal kriging for modelling residuals, as a result of modelling of nitrogen oxides (NO_x) levels using a combination of basis functions. The work of Graeler et al. [31], in turn, concerns spatiotemporal interpolation of air quality data across Germany, shows various approaches taken to model variography used in spatiotemporal kriging.

For this thesis, lessons learnt from the above examples have been used exhaustively, in order to achieve good modelling accuracy of oil and gas reservoir data.

3.3 Proxy Modeling for Fluid Flow in Porous Media

Due to the fact that numerical reservoir simulations can be computationally expensive, there have been several publications to replace the numerical simulation model by a proxy model or surrogate model.

Slotte et al. [64] modelled a response surface from sensitivity simulations for the objective function, to achieve a better history match. Narayanan et al. [51] used proxy modelling to upscale geological models by modeling response surfaces to represent different pseudo-functions for different reservoir properties. Carreras et al. [8] employed proxy models to produce field development plans representing P10, P50, and P90 probabilistic oil recoveries and economic indicators like net present values (NPV). This workflow allowed exploring the uncertainty space by use of experimental designs and polynomial proxy models to predict the oil recoveries.

In recent years, artificial neural networks have been used to develop such proxy

/ surrogate reservoir models. Chen et al. [10] introduced a non-intrusive reduced order simulation methods, which aim towards predicting space-time pressure solutions. This was achieved by using a combination of Discrete Empirical Interpolation (DEIM) method and artificial neural network. Shohreh et al. [2], in turn, developed a proxy model for a CO_2 sequestration project, to predict reservoir behaviour with varying concentrations of CO_2 injection.

For this thesis, several technique components from the above research publications, especially based on machine learning, have been derived and effectively used. However, the application of these techniques has been kept more focused towards building a surrogate reservoir model, in order to replace conventional physics-based reservoir simulation model. The built surrogate reservoir model has been used to perform ML-based simulations with the prospect of a better history match, in contrast to the above publications.

Chapter 4

Spatiotemporal Modelling of Oil and Gas Reservoir Data

In an oil and gas field, there are several sources and types of data, collected over the different parts of the field life cycle. These data are measured to capture the field dynamics across various aspects and procedures dedicated to produce oil and gas. One of the most important data pertains to represent subsurface reservoir dynamics evolving over years of hydrocarbon production.

As hydrocarbons are produced through oil wells, the subsurface reservoir pressure changes, based on the production and injection strategy and natural reservoir characteristics. As explained in Section 2.1, the evolution of subsurface reservoir pressure through the reservoir life cycle proves to be a strong indicator of natural and induced energy sources at play inside the reservoir, which in turn is responsible for effective hydrocarbon recovery. Consequently, acquisition and analysis of reservoir pressure data is of high importance to an oil company and to reservoir engineers.

In this chapter, an effective way of analyzing and modelling reservoir data (specifically reservoir pressure data) is introduced; its description covers the proposed methodology and the applied workflow. The devised data-driven workflow models reservoir pressure without constructing a physics-based numerical simulation model. The effectiveness of this workflow has been quantified through reservoir pressure prediction accuracy. The goal of this workflow was to use the concepts of spatiotemporal modelling to model reservoir pressure measured at sparse observation locations, i.e., wells, in order to predict reservoir pressure across any point in space and time, within the domain of observation data, as depicted by Figure 4.1.

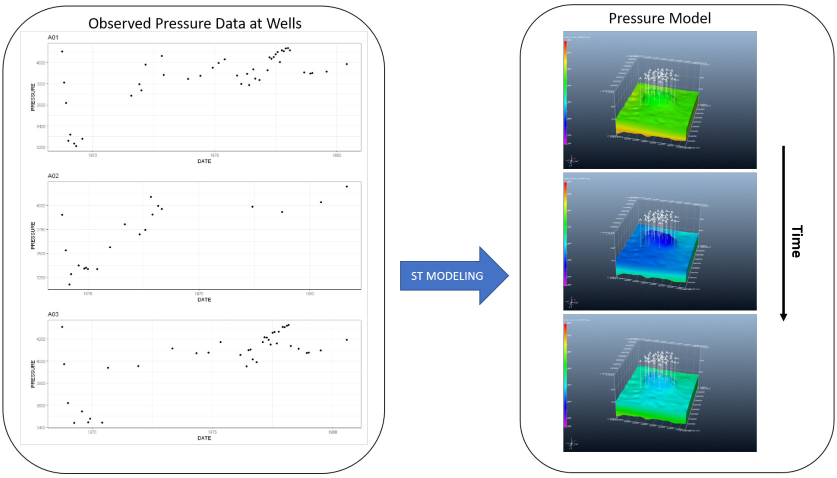

Figure 4.1: Spatiotemporal modelling workflow goal of modelling pressure observed at sparse observation locations (i.e. wells), on the left, to predict reservoir pressure in 3D space and time, on the right.

4.1 Methodology and Applied Workflow

As stated above, the goal of this workflow was to take reservoir pressure data measured at wells and model them across space and time. This workflow comprises a three-step process, as illustrated in Figure 4.2:

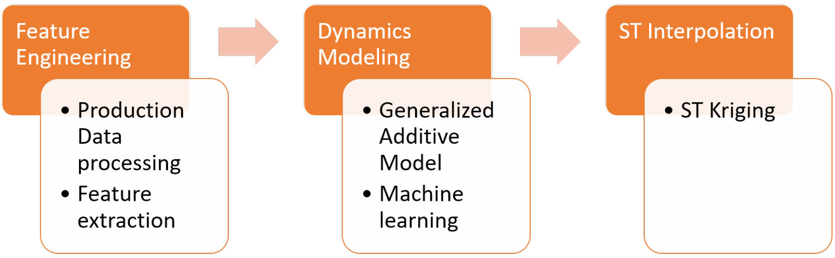
Feature Engineering
Production Data processing
Feature extraction
Dynamics Modeling
Generalized Additive Model
Machine learning
ST Interpolation
ST Kriging

Figure 4.2: Spatiotemporal modelling workflow.

4.1.1 Feature Engineering

The well pressure was analyzed based on its profile defined in terms of its geo-location (X and Y coordinates), depth, and time. The aim of this analysis was to identify patterns of pressure evolution based on the location in the reservoir space and time, by performing a multivariate analysis of pressure against the engineered spatial and temporal features. Tools like Pearson correlation matrix [76] were used to perform such multivariate analysis. For feature engineering, the input features were transformed in order to produce synthetic features, providing a better sense of reservoir dynamics. For instance, it was decided to engineer the location features (X, Y coordinates) to a transformed feature "$Theta$", in order to indicate the direction of the measurement point from the centre of field as depicted by Figure 4.3. In overall, the geo-location features (X, Y, Z) were transformed to a quasi-cylindrical coordinate system $(Theta, Z)$.

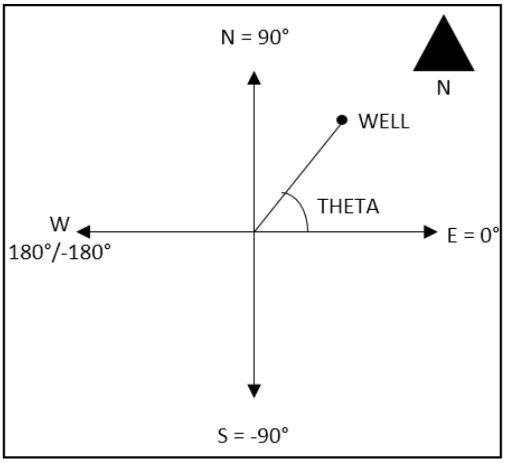

Figure 4.3: XY location feature engineered to a quasi-cylindrical coordinate system.

4.1.2 Dynamics Modeling

In this step, modelling techniques like Generalized Additive Modeling (see Section 2.2) and Machine learning (see Section 2.4) were employed to capture the global trends of pressure in the reservoir. The overall goal was to be able to predict accurately, reservoir pressure for a given location in space and time. This was achieved by taking reservoir space (x represented by $theta$, and $elevation$) and time (t) as the basis for the finalized input features (x_{theta}, x_{elev}, and t). The measured pressure values p were taken as output labels and were mapped as a function of the input features according to Equation 4.1. ϵ represents the residual errors as a result of modelling.

$$p = f(x_{theta}, x_{elev}, t) + \epsilon \tag{4.1}$$

For generalized additive modelling (GAM), the pressure was modelled by a set of additive smooth functions of the input features mentioned above including their tensor interactions as given by Equation 4.2.

$$p = f_1(x_{theta}) + f_2(x_{elev}) + f_3(t) + f_4(x_{theta} \otimes x_{elev} \otimes t) + \epsilon \tag{4.2}$$

For machine learning, a multi layer perceptron (MLP) was trained for regression with above input features and pressure values as labels (see Equation 4.3).

$$p = f_{MLP}(x_{theta}, x_{elev}, t) + \epsilon \tag{4.3}$$

4.1.3 Spatiotemporal Interpolation

As a consequence of dynamics modelling, the pressure predictions were obtained with some residual errors. These residuals were analyzed for correlation to x_{theta}, x_{elev}, and t. Being uncorrelated to overall space and time domains, these residuals were accounted for localized processes in space as well as time. Consequently, spatiotemporal kriging method (see Section 2.3) was employed to interpolate/model these residuals, as the residuals were of stationary nature in terms of mean and variance across spatial and temporal contexts.

In the end, the final pressures were modelled as a result of Equation 4.4:

$$p_{total} = p_{model} + p_{residual} \tag{4.4}$$

where, p_{total} is the final modelled pressure, p_{model} is the pressure predictions obtained from the dynamics modelling step and $p_{residual}$ is the ST krigged residuals.

4.2 Tools and Techniques

In order to ensure seamless and effective execution of the above workflow, following implementation tools and techniques were used:

- Jupyter notebook R and Python: Jupyter notebooks [56] provide a seamless platform for data science projects, especially due to the ease of switching between Python (version 3.7.6) [67] and R programming (version 3.6.1) [58] language kernels.
- Spatiotemporal Modelling: gstat library (version 2.0-4) [32].
- Generalized Additive Models: mgcv library (version 1.8-31) [77].

- Machine Learning: keras library (version 2.2.5.0) [1].
- Parallel computing: parallel library in R (version 3.6.1) [58].
- OpenFlow^{TM} [6]: Oil and gas industry geoscience software (version 2019.1 Update 8).

4.3 Applied Case: Undersaturated Oil Reservoir with Water and Gas Injection

In this thesis, spatiotemporal modelling of oil and gas reservoir data was performed for a real oil and gas field. The field had about 50 years (from August 1968 – January 2019) of production history comprising of 7 gas injector wells, 17 water injector wells, and 41 oil producer wells. The reservoir top depth was around 8475ft with a thickness of about 1600ft with initial water oil contact at 9460ft. Figures 4.4 and 4.5 depict a 2D and 3D model of the reservoir, respectively.

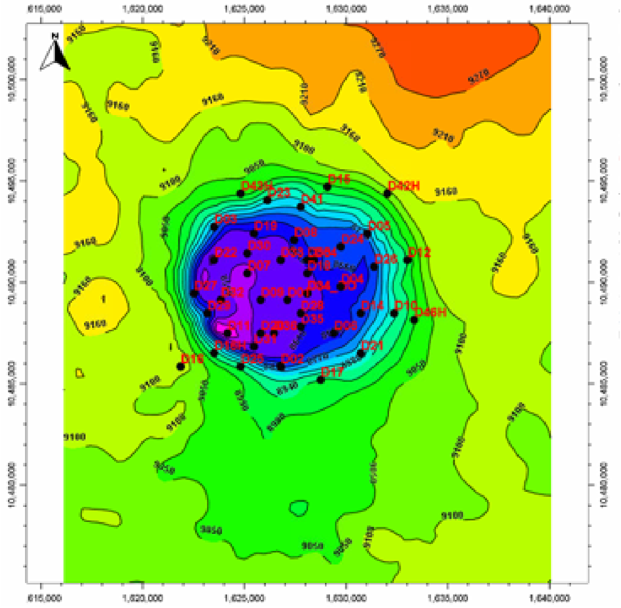
1,615,000
1,620,000
1,625,000
1,630,000
1,635,000
1,640,000
10,500,000
10,495,000
10,490,000
10,485,000
10,480,000
N
D42H
D23
D15
D41
D43H
D03
D19
D08
D05
D24
D22
D30
D33
D12
D07
D18
D26
D27
D32
D09
D34
D04
D29
D28
D14
D10
D46H
D11
D35
D06
D18H
D31
D21
D16
D25
D02
D17
9160
9210
9270
9100
9050

Figure 4.4: Top view of reservoir depth map.

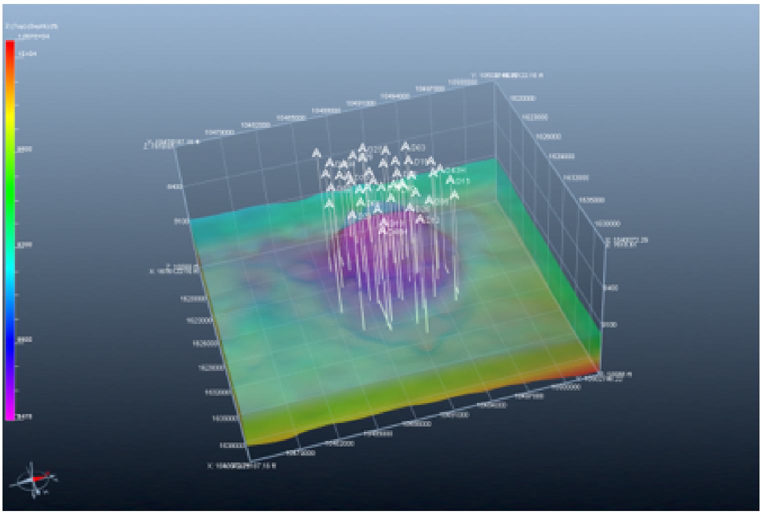

Figure 4.5: 3D model of reservoir.

4.3.1 Data Description

The database available for this project is comprised of the following data entities:

- Oil and Gas Production History: It comprised of oil and gas production data from 41 producer wells.
- Gas and Water Injection History: It comprised of gas and water injection data from 7 gas injectors and 17 water injectors.
- Well locations: Well head X, Y coordinates and Z positions.
- Well perforation data: Perforation depth and date.
- Well pressure data: Measured pressure data as depicted by Figure 4.6.

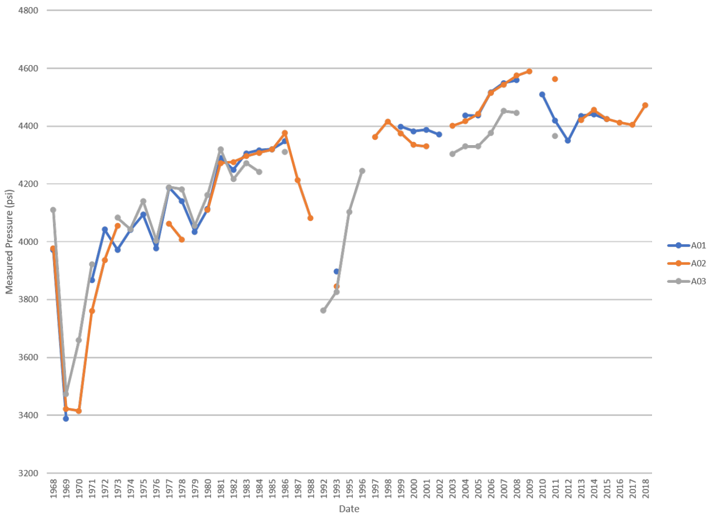

Figure 4.6: Measured well pressure data for three sample wells (A01, A02, A03).

4.3.2 Experiments

In order to qualitatively assess the potential and quantify the effectiveness of the workflow (see Section 4.1), the following experiments were conducted:

- **Experiment 1: Base case**
  In this experiment, the overall pressure data was included for ST modeling in accordance with the workflow. Section 4.3.3 shows the results of the experiment.

- **Experiment 2: Temporal scale partitioning**
  In this experiment, the time scale was split into two parts representing two different set of reservoir dynamics at play. This was supported by the fact that the injection strategy was changed post October 1981 from water injection to gas injection. Section 4.3.3 shows the results of the experiment.

Table 4.1 summarizes the accuracy of the above experiments.

Table 4.1: Modelling accuracy for ST Modeling experiments

Model Accuracy				
	Base Case		Temporal Scale Partitioning	
	GAM	MLP	GAM	MLP
Number of Modelling Wells	32	32	32	32
Number of Test Wells	3	3	3	3
Modelling Error in psi (MAE)	42.65	34.4	26.85	31.5
Test Error in psi (MAE)	54	65.4	55	71.7

4.3.3 Results and Discussion

The workflow described in Section 4.1 was successfully executed on a real oil and gas field data. The experiments described in the section above, were carried out as given below.

Base Case

As a first experiment (Base Case), all pressure data were considered for modelling the reservoir pressure dynamics.

- **Feature analysis and engineering:**

  Firstly, the pressure data was analyzed for its variability against spatial and temporal features ($Theta$, Z, and $Date$). As evident from Figures 4.7 and 4.8, the pressure evolution was uniformly consistent across all values of $Theta$, as indicated by a low Pearson correlation coefficient of 0.06. This formed to be a good indicator of homogeneous reservoir dynamics in all parts of the reservoir. Moreover, the correlation of pressure against elevation$/Z$ (Pearson correlation coefficient of -0.48) honored the reservoir pressure gradient based on the fluid gravity. In overall, the feature analysis provided a rudimentary understanding of the reservoir dynamics across space and time.

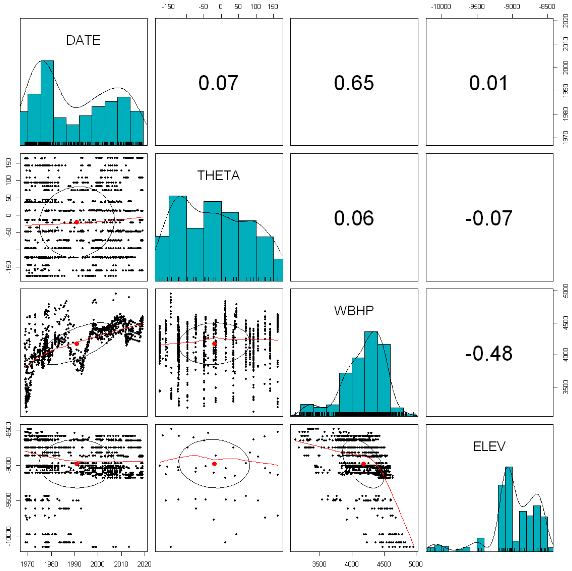

Figure 4.7: Base Case: Correlation matrix indicating Pearson correlation coefficients of pressure data against spatial and temporal features.

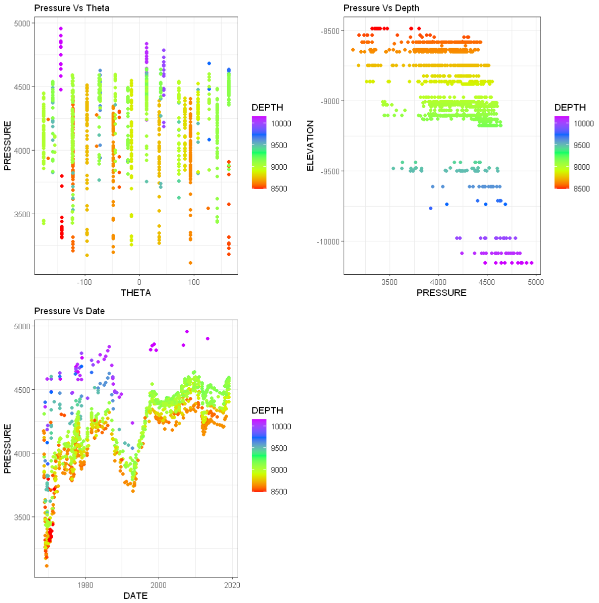

Figure 4.8: Base Case: Feature multivariate analysis of reservoir pressure against spatial and temporal features evaluating heterogeneity in reservoir dynamics.

- **Dynamics Modelling with GAM:**

  In order to model the overall dynamics of the reservoir pressure (second part of the workflow), two approaches were considered, i.e., Generalized additive modeling and Machine learning. Both approaches were applied to model pressure as a function of spatial and temporal features (Theta, Z, and Date). As depicted by Figure 4.9, the GAM model was able to capture the data variability with high accuracy, producing a mean absolute error (MAE) of about 53 psi and a high R^2 value of 0.937.

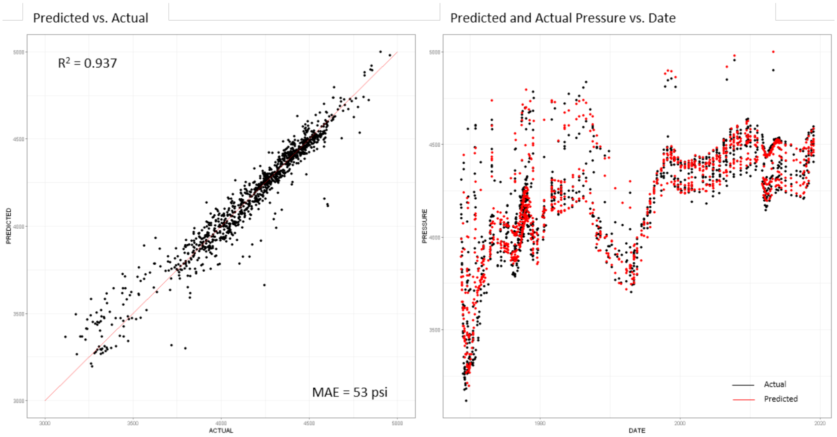

Figure 4.9: Figures indicating the accuracy of GAM model for Base Case in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by GAM model. The figure on the right shows predicted and actual pressures vs. date.

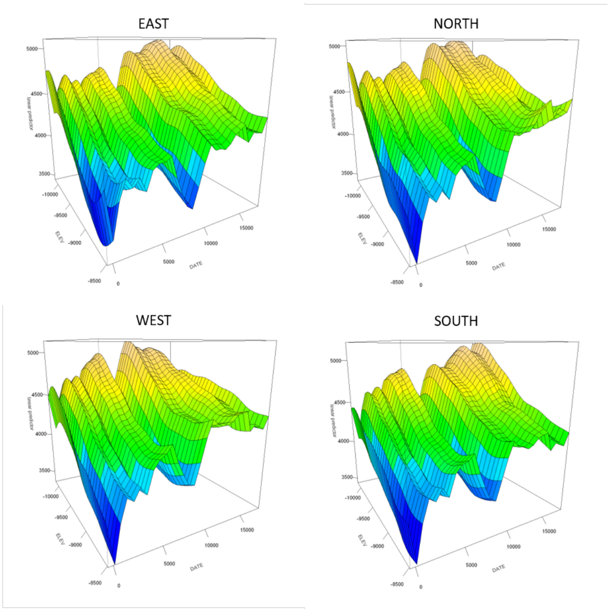

Figure 4.10: Figure shows hyperspace of pressure predicted by GAM model for Base Case against Elevation and Date, for four values of Theta (representing directions of East, North, South, and West from the centre of the oil and gas field).

In overall, none of the modelling residuals showed any correlation to the spatial and temporal features (see Figure 4.11). However, it was observed that most of the error was originating from pressure data before October 1981, thus leading to a non-stationary residual variance. Non-stationarity in modelling residual variance was a challenge, especially for the third step of the workflow (Spatiotemporal kriging) as the ordinary kriging algorithm works best for a random field with stationary mean and variance [43].

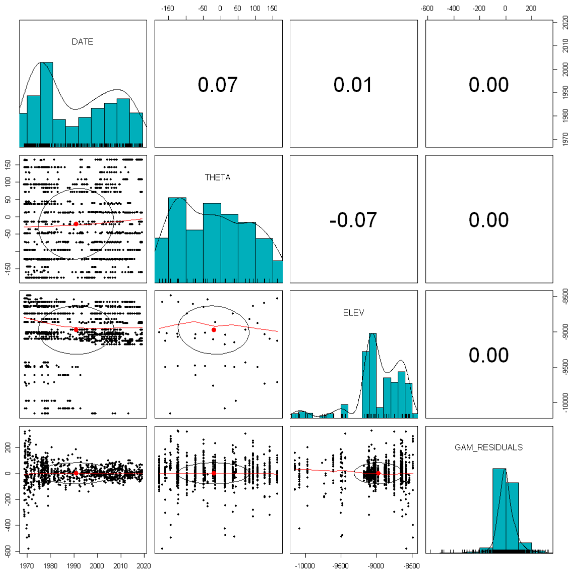

Figure 4.11: Base Case: Correlation matrix indicating Pearson correlation coefficients of GAM modelling residuals against spatial and temporal features.

- **Dynamics Modelling with MLP:**

  On the other hand, the MLP model produced a mean absolute error of about 54.5 psi and R^2 value of 0.937 (see Figure 4.12). Similar to the GAM modelling residuals, the MLP modelling residuals did not show any overall correlation to the spatial and temporal features, but they were of higher variance before October 1981 (see Figure 4.13).

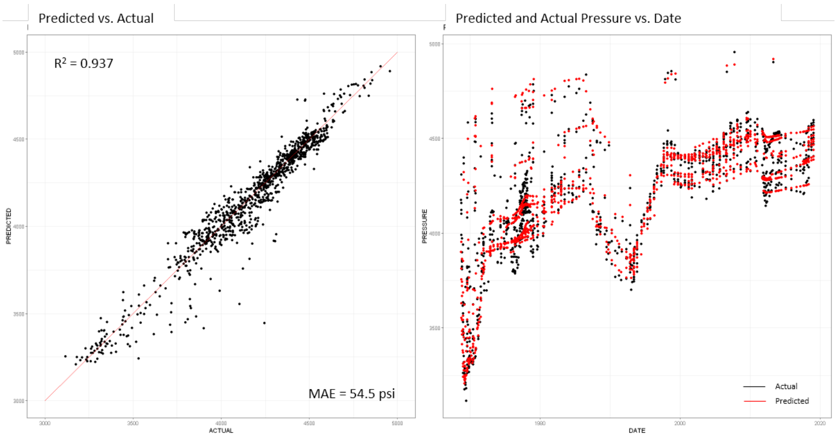

Figure 4.12: Figures indicating the accuracy of MLP model for Base Case in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by MLP model. The figure on the right shows predicted and actual pressures vs. Date.

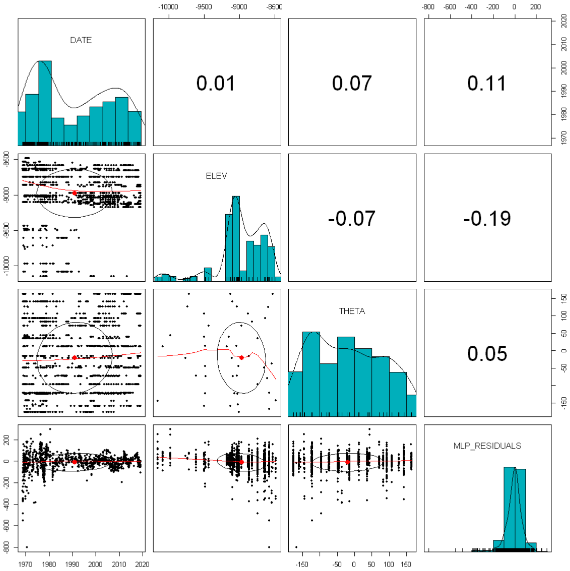

Figure 4.13: Base Case: Correlation matrix indicating Pearson correlation coefficients of MLP modelling residuals against spatial and temporal features.

- **Final Model:**

  Despite the presence of non-stationarity of modelling residual variance, the third step of the workflow (spatiotemporal kriging) was executed and thereafter final models, composed of GAM/MLP model and spatiotemporally interpolated residuals (as described by Section 4.1), were calculated. The final models were quantified for their accuracy as given by Table 4.1. Figures 4.16 and 4.17 show the finally modelled pressure at the training and test wells for MLP model; and Figures 4.14 and 4.15 show the same for GAM model. It can be concluded that the Generalized additive modelling is much better at capturing the non-linearity than the Machine learning model. Nevertheless, spatiotemporal kriging ensures high accuracy, regardless of modelling technique used.

From the above mentioned results, in was also inferred that in order to get accurate predictions at spatial and temporal points away from the ones at training wells, it is highly important to have a model capable of capturing the overall reservoir dynamics before implementing ST kriging. This is evident by machine learning model producing a final model with high error for the test wells.

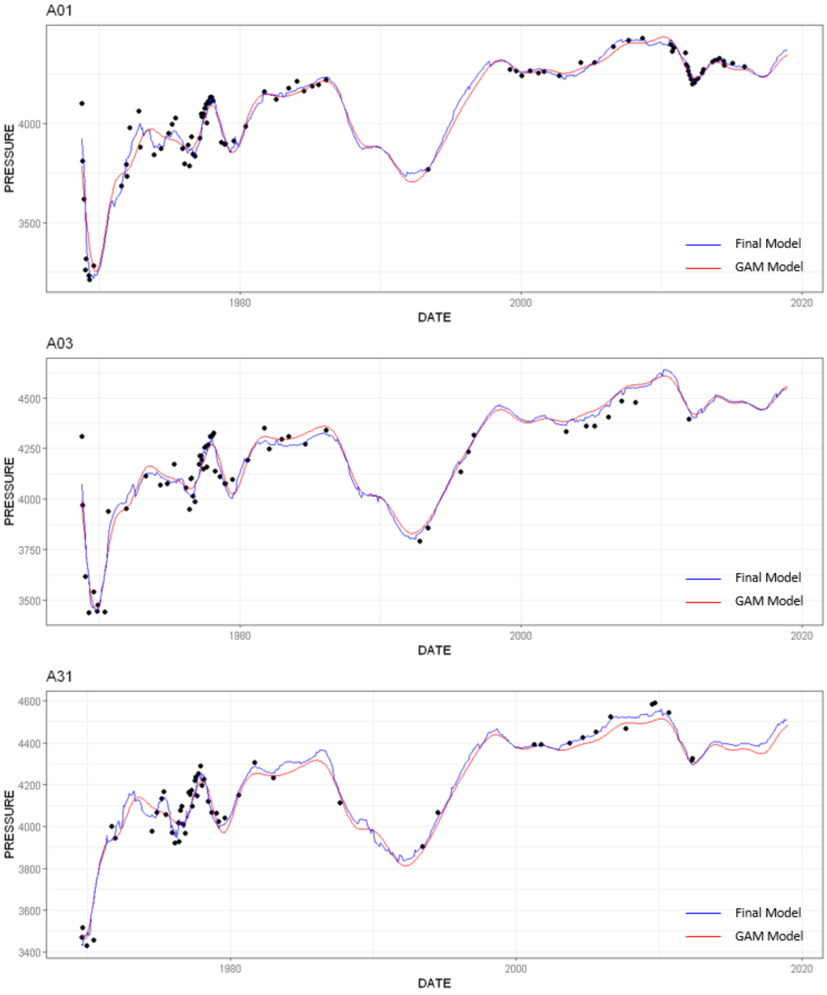

Figure 4.14: Base Case: Figure shows observed pressure (measured data), GAM pressure prediction (in red), and final pressure model (in blue), incorporating GAM predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in training dataset.

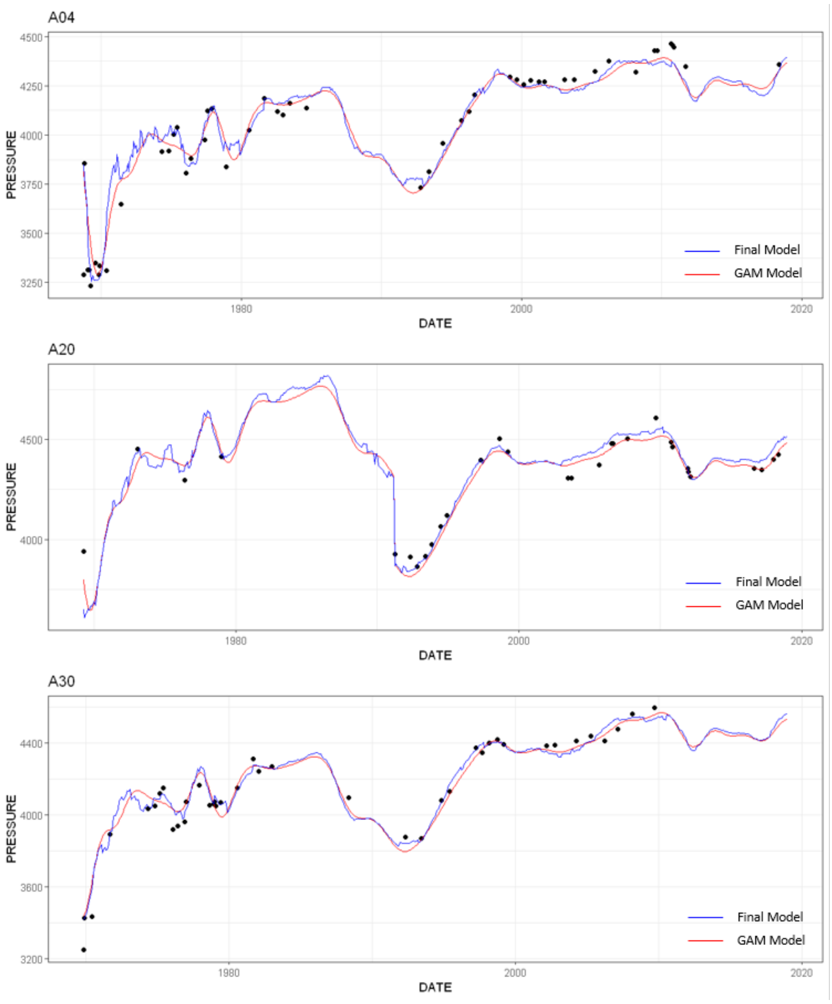

Figure 4.15: Base Case: Figure shows observed pressure (measured data), GAM pressure prediction (in red), and final pressure model (in blue), incorporating GAM predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in test dataset.

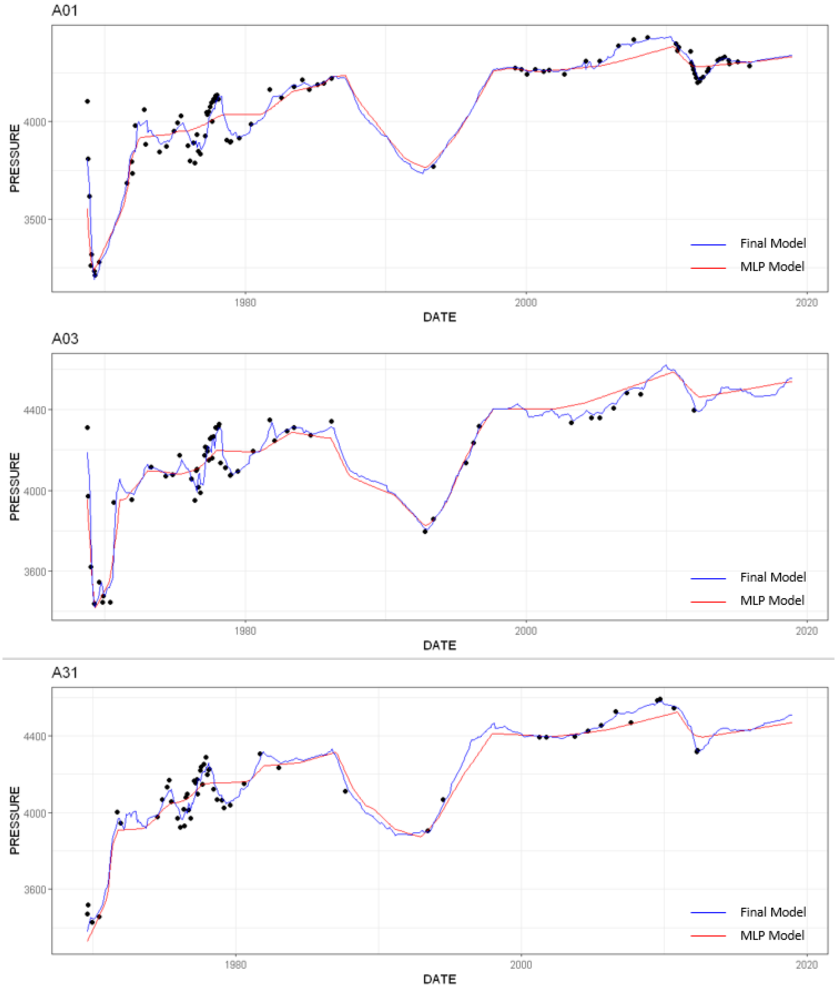

Figure 4.16: Base Case: Figure shows observed pressure (measured data), MLP pressure prediction (in red), and final pressure model (in blue), incorporating MLP predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in training dataset.

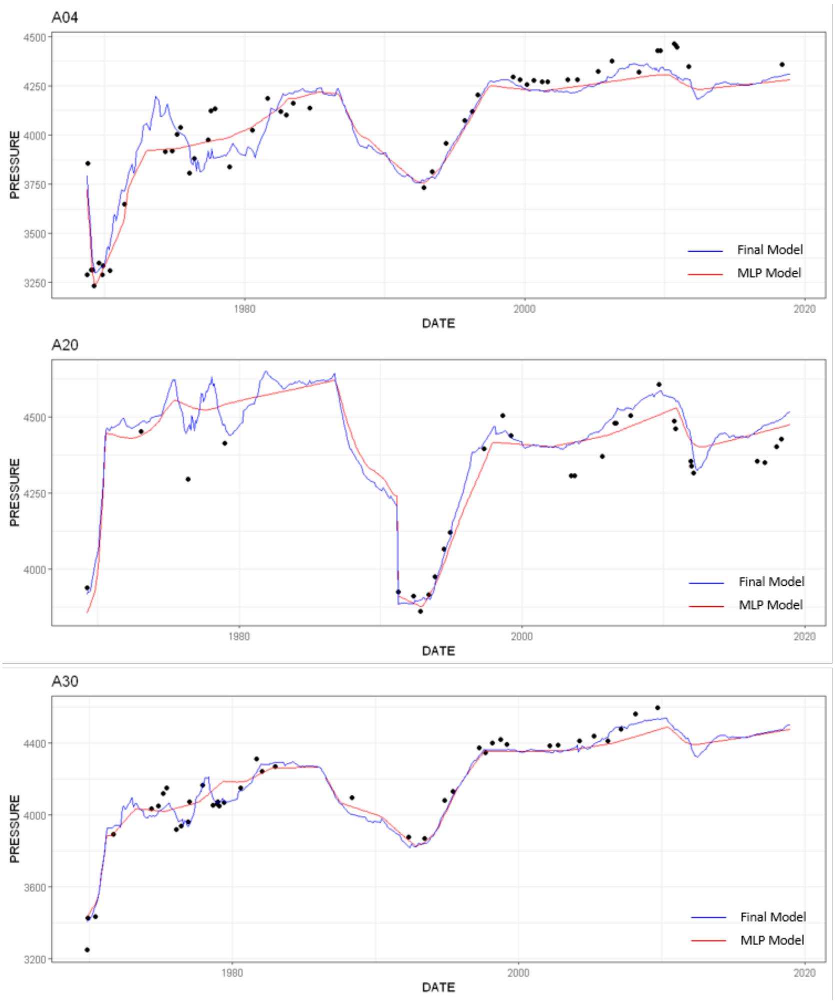

Figure 4.17: Base Case: Figure shows observed pressure (measured data), MLP pressure prediction (in red), and final pressure model (in blue), incorporating MLP predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in test dataset.

Temporal scale partitioning

In order to tackle the above challenge of non-stationary residual variance, reason behind the same was investigated. It was known from the field development history, that the injection strategy changed post October 1981 from water injection to gas injection, which in-turn led to a change in reservoir dynamics. Thus it was necessary to implement independent models to capture the dynamics of the two time duration (pre-October 1981 and post October 1981). This engendered a second experiment of Temporal Scale Partitioning, in which the GAM and MLP models were built for the two time duration.

- **Dynamics Modelling pre-October 1981:**

  For the time duration before October 1981, Figures 4.18, 4.19 show the accuracy (MAE = 55 psi, R^2 = 0.944) and pressure prediction hyperspace of the GAM model and Figure 4.21 shows the the accuracy (MAE = 110.3 psi, R^2 = 0.80) of the MLP model. Again, the GAM model showed better accuracy to model the pressures. However, the MLP and GAM modelling residuals showed stationary mean and variance and no correlation to spatial and temporal features (see Figures 4.22 and 4.20, respectively).

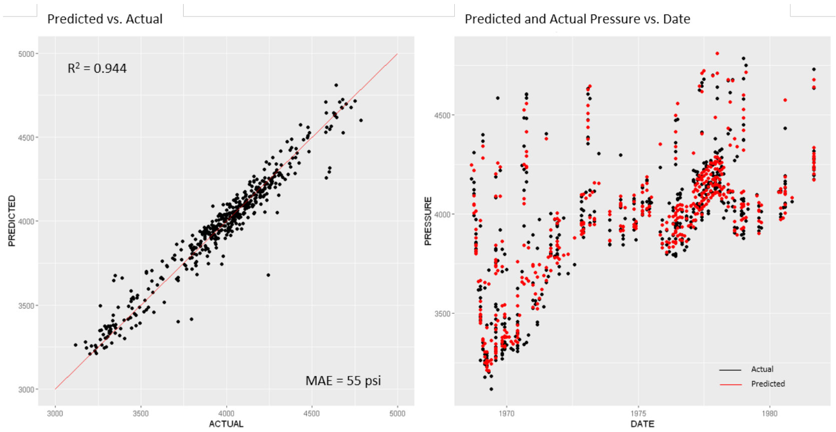

Figure 4.18: Figures indicating the accuracy of the GAM model for Temporal scale partitioning (pre-October 1981) in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by the GAM model. The figure on the right shows predicted and actual pressures vs. date.

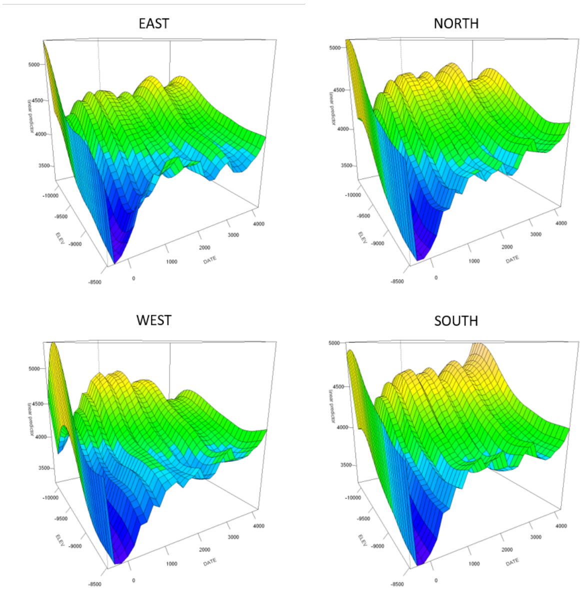

Figure 4.19: Figure shows hyperspace of pressure predicted by the GAM model for Temporal scale partitioning (pre-October 1981) against Elevation and Date, for four values of Theta (representing directions of East, North, South and West from the centre of the oil and gas field).

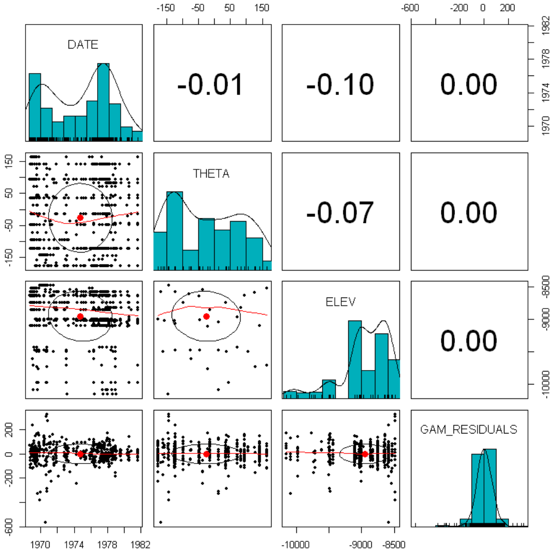

Figure 4.20: Temporal scale partitioning (pre-October 1981): Correlation matrix indicating Pearson correlation coefficients of GAM modelling residuals against spatial and temporal features.

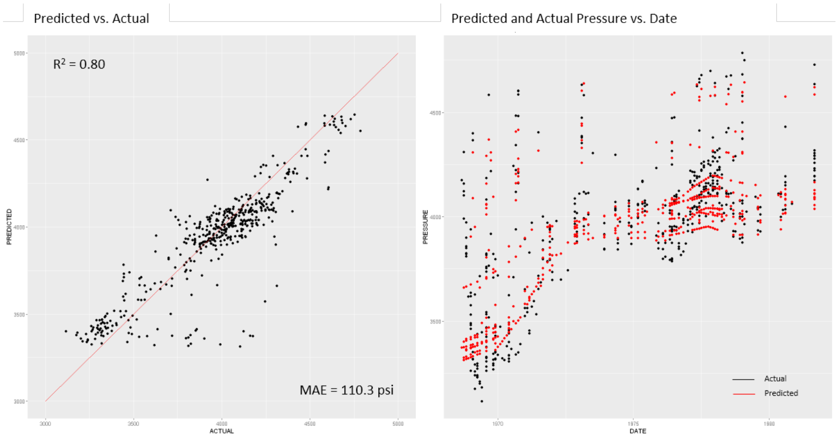

Figure 4.21: Figures indicating the accuracy of the MLP model for Temporal scale partitioning (pre-October 1981) in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by the MLP model. The figure on the right shows predicted and actual pressures vs. Date.

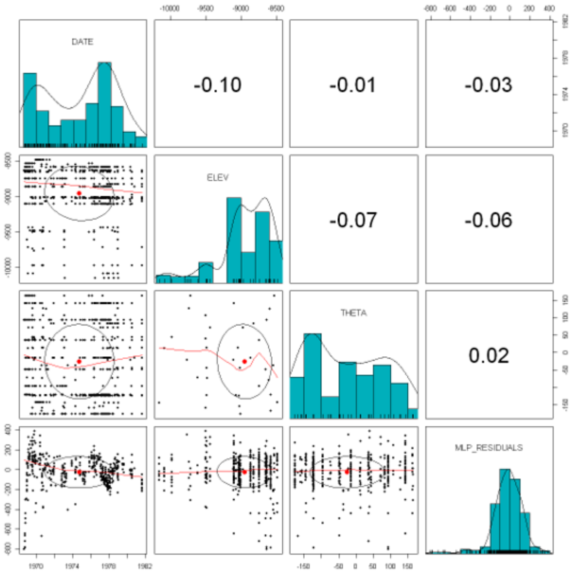

Figure 4.22: Temporal scale partitioning (pre-October 1981): Correlation matrix indicating Pearson correlation coefficients of MLP modelling residuals against spatial and temporal features.

- **Dynamics Modelling post-October 1981:**

  For time duration after October 1981, Figures 4.23, 4.24 show the accuracy (MAE = 22 psi, R^2 = 0.98) and pressure prediction hyperspace of the GAM model and Figure 4.26 shows the the accuracy (MAE = 39.4 psi, R^2 = 0.94) of the MLP model. Consistently, GAM had outperformed MLP in the capability of capturing the non-linear dynamics. The MLP and GAM modelling residuals for this time duration too, showed stationary mean and variance and no correlation to spatial and temporal features (see Figures 4.27 and 4.25 respectively).

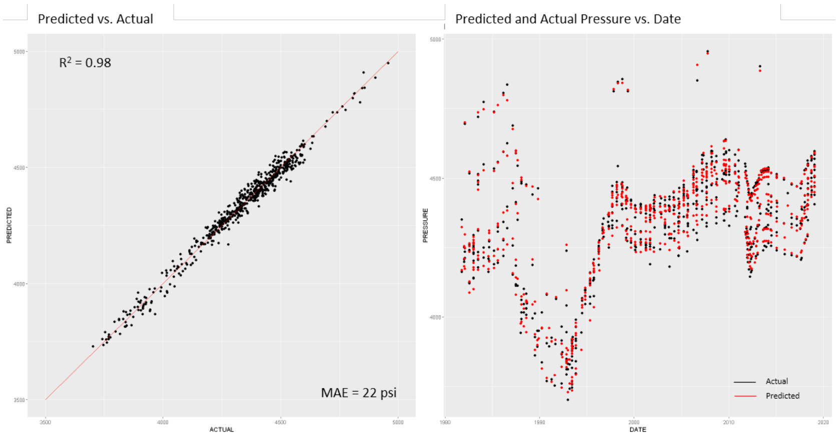

Figure 4.23: Figures indicating the accuracy of the GAM model for Temporal scale partitioning (post-October 1981) in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by GAM model. The figure on the right shows predicted and actual pressures vs. Date.

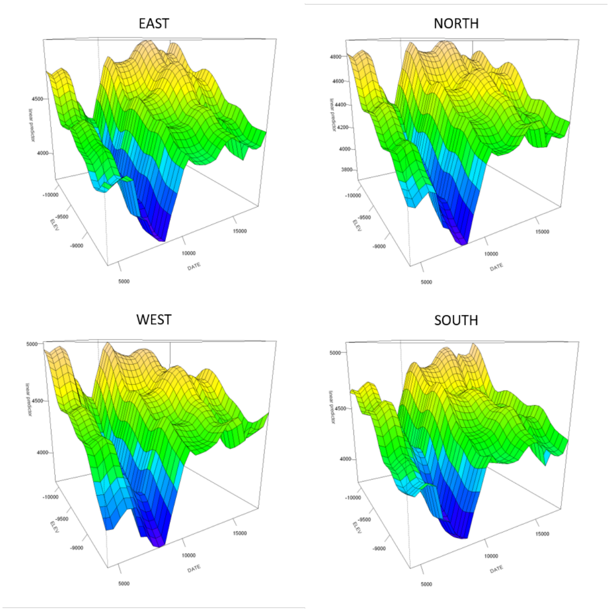

Figure 4.24: Figure shows hyperspace of pressure predicted by the GAM model for Temporal scale partitioning (post-October 1981) against Elevation and Date, for four values of Theta (representing directions of East, North, South and West from the centre of the oil and gas field).

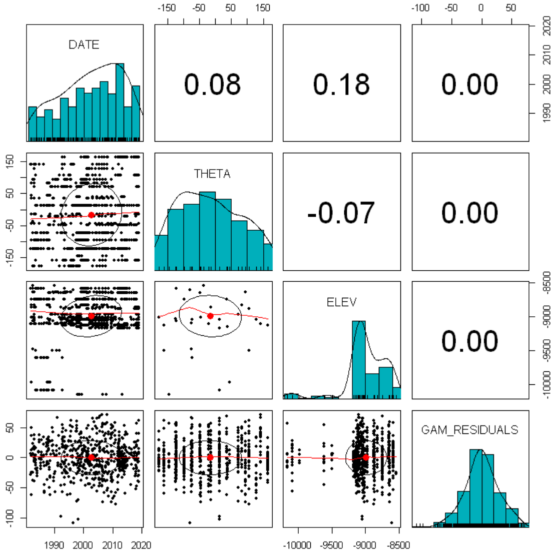

Figure 4.25: Temporal scale partitioning (post-October 1981): Correlation matrix indicating Pearson correlation coefficients of GAM modelling residuals against spatial and temporal features.

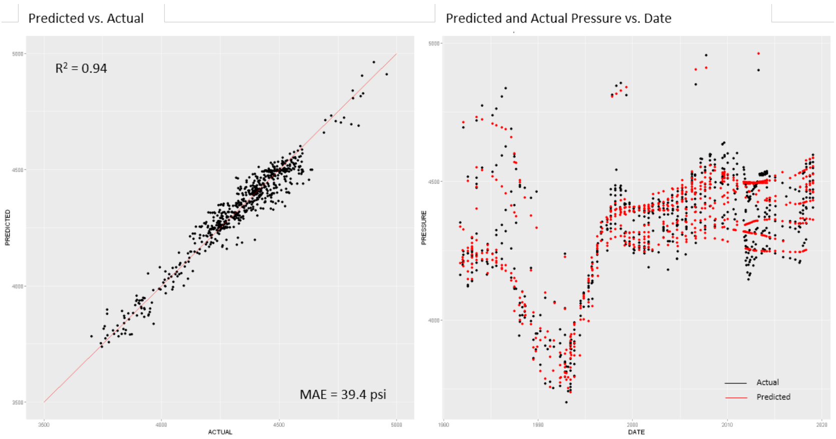

Figure 4.26: Figures indicating the accuracy of MLP model for Temporal scale partitioning (post-October 1981) in capturing the pressure data variance. The figure on the left shows a scatter plot of actual pressure values vs. pressure predicted by MLP model. The figure on the right shows predicted and actual pressures vs. Date.

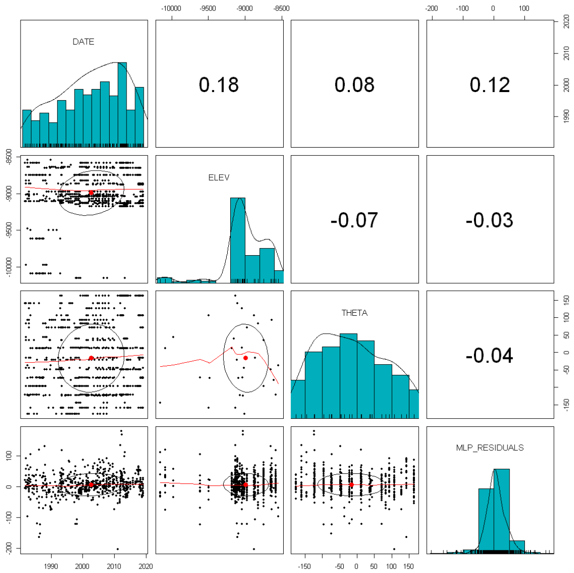

Figure 4.27: Temporal scale partitioning (post-October 1981): Correlation matrix indicating Pearson correlation coefficients of MLP modelling residuals against spatial and temporal features.

- **Final Model:** The modelling residuals were interpolated through spatiotemporal kriging for both the models (i.e., GAM and MLP model). Final pressure model were then calculated taking into account the respective interpolated residuals. Final models with GAM and MLP were then evaluated for accuracy as tabulated in Table 4.1. Figures 4.28 and 4.29 show the final pressure profiles at the training and test wells respectively for GAM model. Figures 4.30 and 4.31 show the same for MLP model.

  The temporal scale partitioning yielded better modelling accuracy for both GAM and MLP models (see Table 4.1), in comparison to the base case. The experiment was especially effective while using GAM model, which produced the final modelling MAE of 26.85 psi on training dataset and MAE of 55 psi on test dataset.

  In overall, the best modelling and test accuracy was obtained by Generalized additive model.

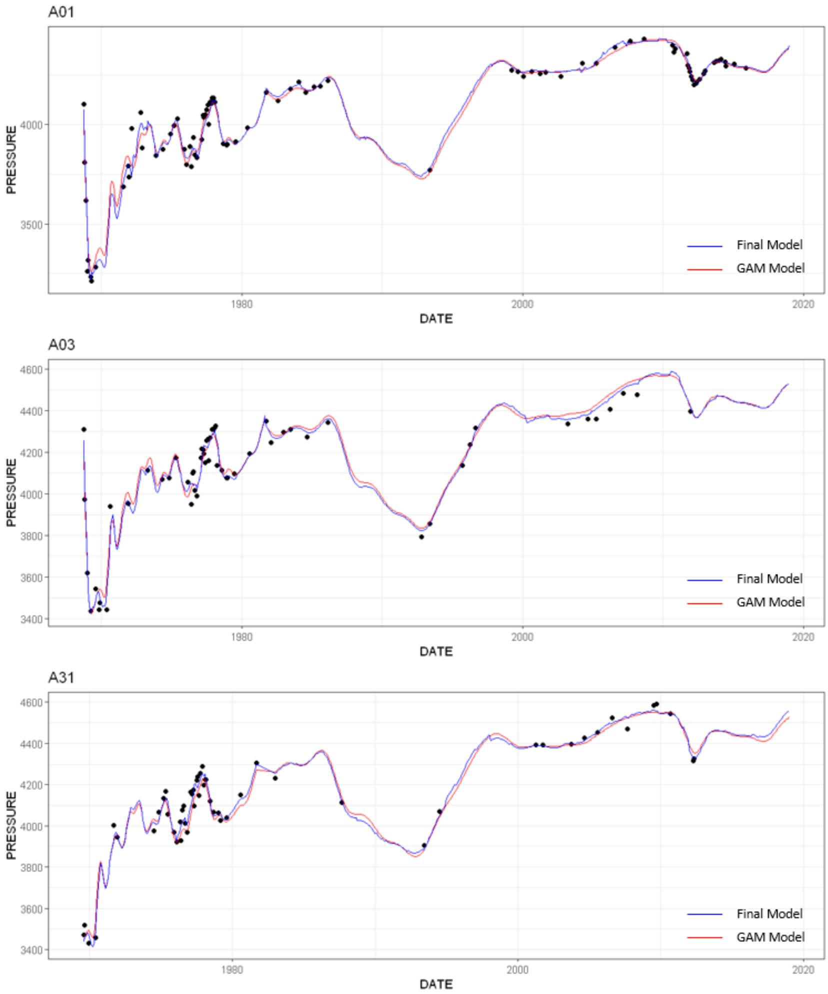

Figure 4.28: Temporal scale partitioning: Figure shows observed pressure (measured data), GAM pressure prediction (in red), and final pressure model (in blue), incorporating GAM predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in training dataset.

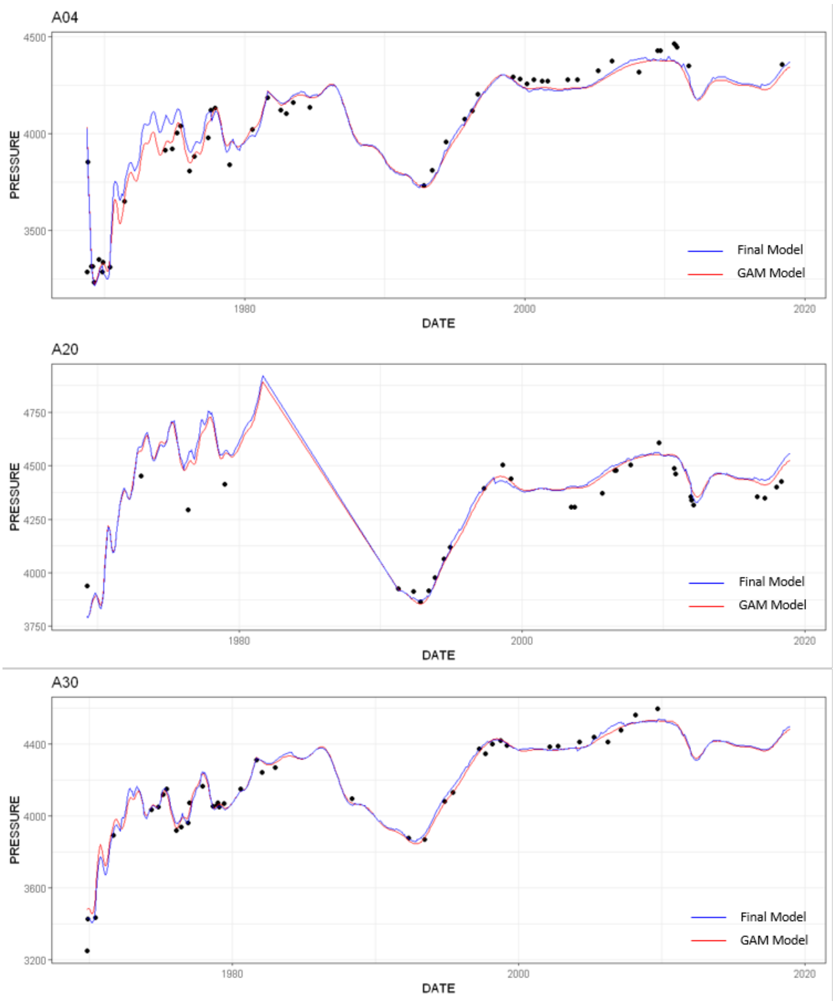

Figure 4.29: Temporal scale partitioning: Figure shows observed pressure (measured data), GAM pressure prediction (in red), and final pressure model (in blue), incorporating GAM predictions and spatiotemporally kriged/interpolated modelling residuals, for sample wells in test dataset.

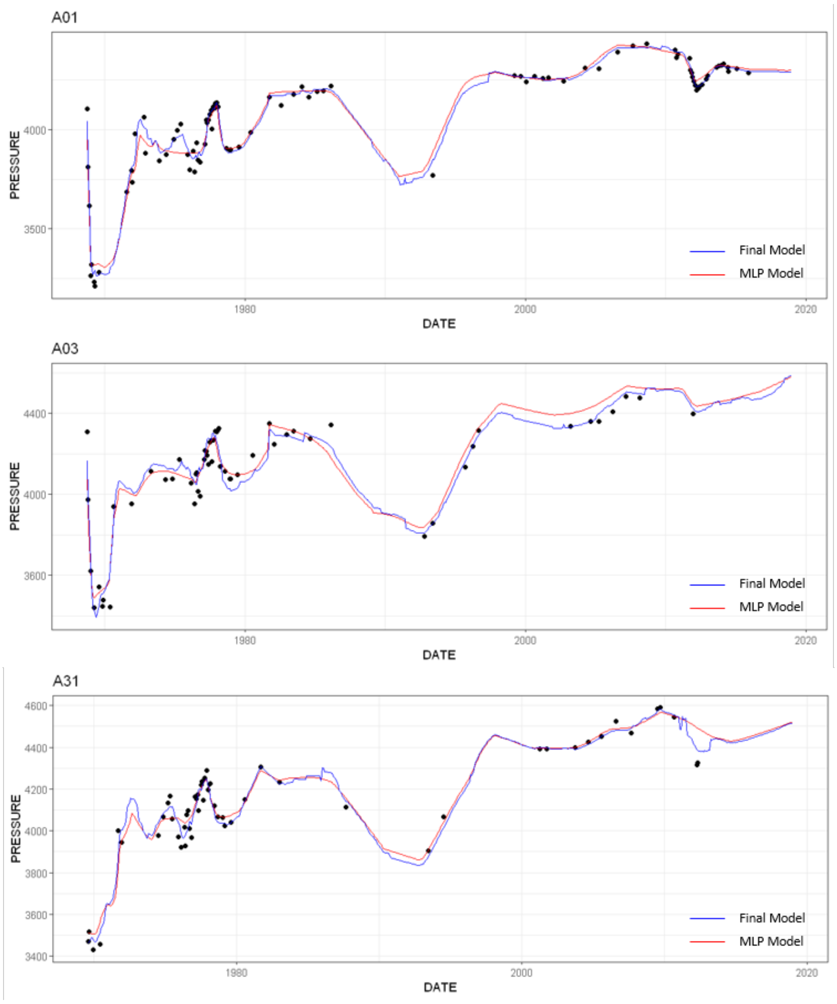

Figure 4.30: Temporal scale partitioning: Figure shows observed pressure (measured data), MLP pressure prediction (in red), and final pressure model (in blue), incorporating MLP predictions and spatiotemporally krigged/interpolated modelling residuals, for sample wells in training dataset.

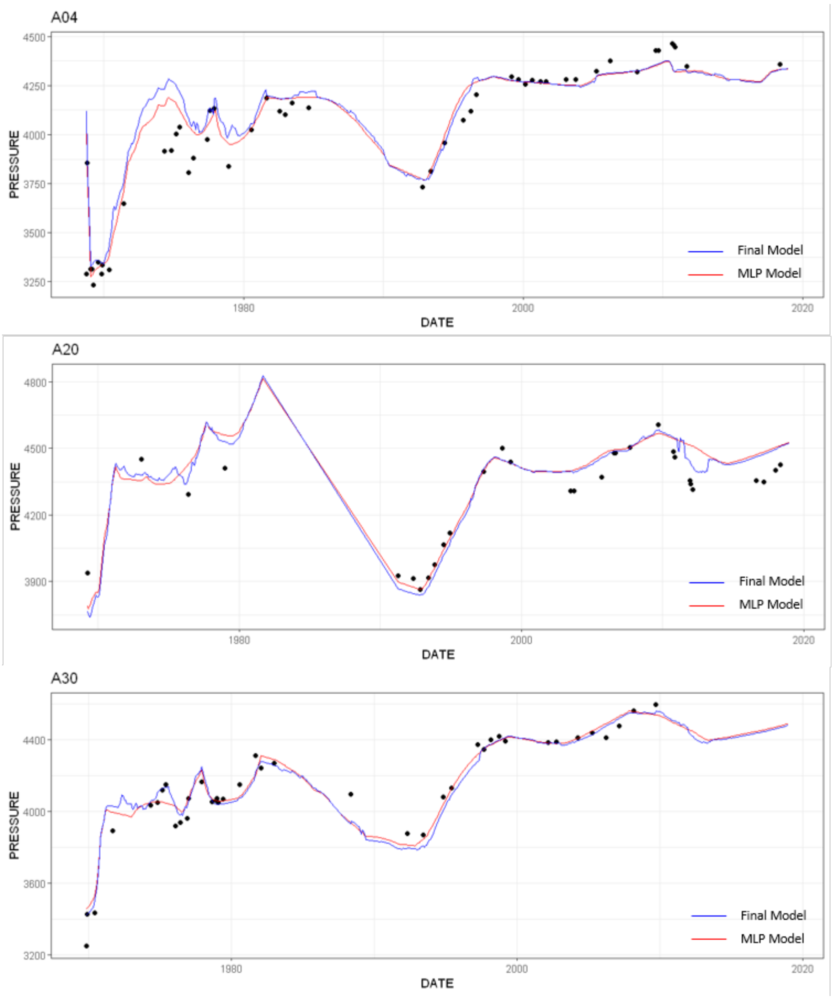

Figure 4.31: Temporal scale partitioning: Figure shows observed pressure (measured data), MLP pressure prediction (in red), and final pressure model (in blue), incorporating MLP predictions and spatiotemporally kriged/interpolated modelling residuals, for sample wells in test dataset.

Computational Efficiency

In terms of computational efficiency, the spatiotemporal modelling of reservoir pressure data provided pressure predictions in a matter of minutes (less than 20 minutes), in contrast to more than 4 hours of simulation time while using physics-based numerical simulation of the real data case.

Ultimately, pressure modelling was performed on a grid mesh consisting of 852348 cells over 606 timesteps (every month from August 1968 to February 2019) leading to about half a billion computation for every pressure ordinal (GAM/MLP, ST kriging and final model). This was facilitated by implementing parallel computing into the framework. Parallel computing refers to breaking down entire computational problem into smaller parts and processing it simultaneously by multiple processors.

The entire framework developed to provide accurate pressure predictions through the workflow of spatiotemporal modelling of oil and gas reservoir data, used open source libraries, thus expending no additional resources. This was achieved without using conventional physics-based reservoir simulation software, which in contrast are extremely expensive (in some cases, in the order of hundreds of thousands of dollars).

Chapter 5

Machine-Learning-based Oil and Gas Reservoir Simulation

As explained in Section 2.1, reservoir simulation models are currently used as a vital part of conventional practices to understand and forecast the behaviour of oil and gas reservoirs. These models are constructed by bringing together data from various sources, such as seismic, well-logs, outcrop analogs, rock core analysis, and fluid composition analysis. As the geological heterogeneity of reservoir increases, such reservoir simulation models become increasingly computationally expensive. This is due to the fact that, current simulation techniques rely on solving partial differential equations over several time steps using methods like finite element, finite difference etc. Moreover, with challenging tasks like history matching, reservoir engineers have to run several iterations of such simulation models in order to tune the uncertain parameters to the historically measured reservoir data. Eventually, engineers have to tackle this problem with a trade-off between accuracy and model complexity.

In this chapter, an innovative methodology has been devised to replace the reservoir simulation model by a machine learning model using a few numerical reservoir simulations. The concept behind the workflow is to use a machine learning model as a function approximation (as per universal approximation theorem [20]). As depicted by Figure 5.1, a machine learning model can approximate the function mapped for solving differential equations from every timestep $t-1$ to timestep t, in a conventional physics-based numerical simulator. The trained machine learning model then can be used as a surrogate reservoir simulation model. In the scope of this thesis, only pressure simulations have been taken into account. Thus, post machine learning, the trained surrogate reservoir model was able to predict/simulate pressures from a given initial conditions to consequent timesteps recursively, as illustrated by Figure 5.2.

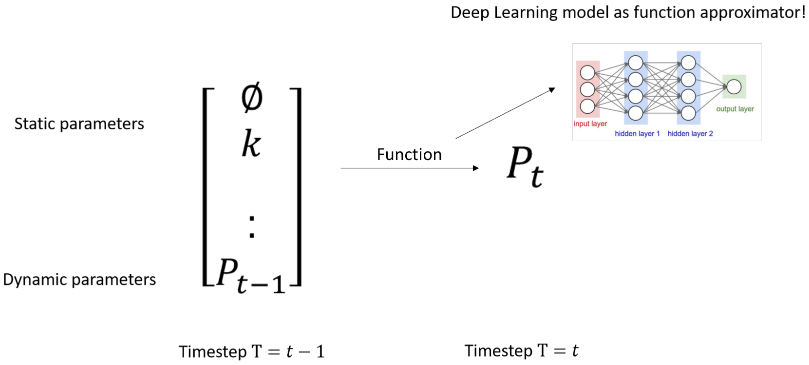
Deep Learning model as function approximator!
input layer
hidden layer 1
hidden layer 2
output layer
Static parameters
Dynamic parameters
$\emptyset$
k
$:$
P_{t-1}
Function
P_t
Timestep T $= t-1$
Timestep T $= t$

Figure 5.1: Concept behind machine-learning-based reservoir simulation.

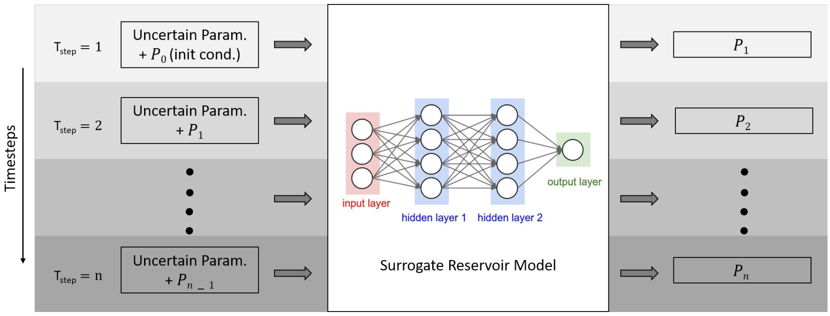
Timesteps
$T_{step} = 1$
Uncertain Param. + P_0 (init cond.)
$T_{step} = 2$
Uncertain Param. + P_1
$T_{step} = n$
Uncertain Param. + P_{n-1}
input layer
hidden layer 1
hidden layer 2
output layer
Surrogate Reservoir Model
P_1
P_2
P_n

Figure 5.2: Figure shows the idea behind recursive algorithm used to produce Machine-learning-based reservoir simulation. At every timestep, uncertain parameter values and pressure values simulated at previous timestep are fed into the trained ML-based surrogate reservoir model to simulate pressure.

5.1 Methodology and Applied Workflow

With the above stated goal to build a surrogate reservoir model and to use it to simulate reservoir pressure, a three step process workflow was devised as given by Figure 5.3.

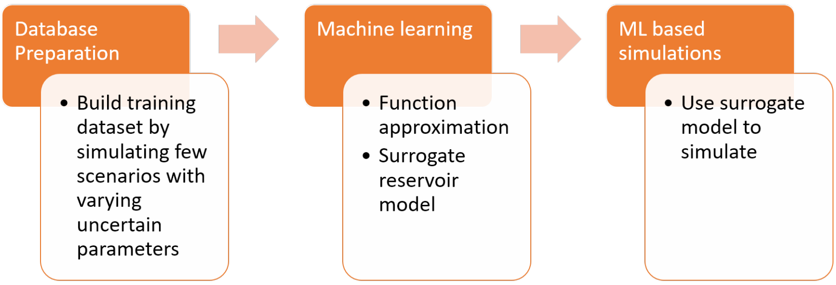
Database Preparation
Build training dataset by simulating few scenarios with varying uncertain parameters
Machine learning
Function approximation
Surrogate reservoir model
ML based simulations
Use surrogate model to simulate

Figure 5.3: Machine-learning-based reservoir simulation.

5.1.1 Database preparation

A training dataset was created by taking few numerical reservoir simulations with varying combinations of uncertain parameters. These uncertain parameters were supposed to be the parameters to be used for history matching. To prepare the training dataset, reservoir pressure grid results from the numerical fluid flow simulations were wrangled for every grid cell as a training instance. The input features were compiled for each cell's location (X, Y, Z coordinates), uncertain parameters values and pressure values for timestep $t - 1$. The corresponding labels were compiled for each cells pressure values for timestep t. Additionally, a separate set of numerical simulations were run (with same set of uncertain parameters but different combinations, than the training dataset) to generate a test dataset to be later used to check the accuracy of machine-learning-based models.

Only the cells above water-oil contact (WOC) were accounted into the training dataset. This was done to map more relevant fluid flow dynamics (confined in hydrocarbon bearing zone) into the surrogate reservoir model.

5.1.2 Machine Learning

With dataset prepared for training, a regularized multi-layer perceptron (see Section 2.4) was trained.

5.1.3 Machine-learning-based simulation

After training the MLP, machine-learning-based simulations were computed based on the concept illustrated in Figure 5.2. Moreover, the same MLP model was evaluated based on its test accuracy for test simulations.

5.2 Tools and Techniques

The tools used to execute various parts of the workflow can be documented by the following:

- Jupyter notebook with Python (version 3.7.6) [56] [67].
- Machine Learning: keras library (version 2.3.1) [12].
- OpenFlow^{TM} [6]: Oil and gas industry geoscience software (version 2019.1 Update 8). PumaFlow^{TM} (part of OpenFlow^{TM} platform) was used to perform reservoir fluid flow numerical simulations.

5.3 Applied Case: Undersaturated Oil Reservoir with Water Injection

The workflow of machine learning based reservoir simulation was executed on an oil and gas field situated in the North Sea. The field had about 4 years (from December 1987 – January 1992) of production history comprising of 4 water injector wells and 7 oil producer wells. The reservoir top depth was around 9960ft with a thickness of about 2130ft with initial water oil contact at 10600ft. Figure 5.4 depicts a 3D model of the reservoir.

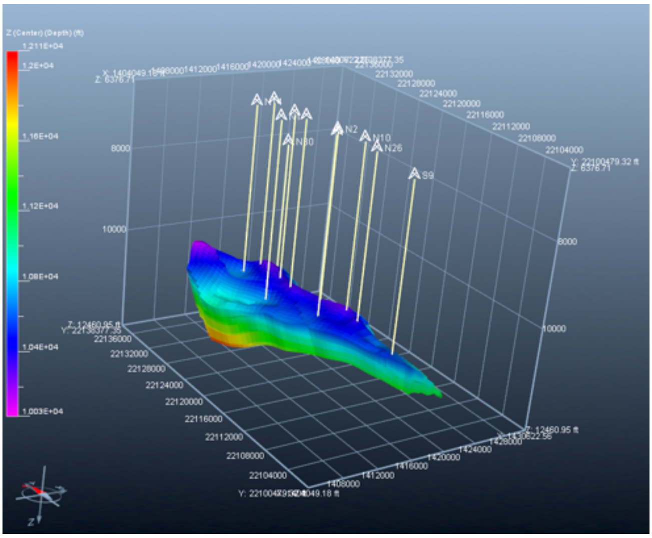
Z (Center) (Depth) (ft)
1.211E+04
1.2E+04
1.16E+04
1.12E+04
1.08E+04
1.04E+04
1.003E+04
X: 1404049.18 ft
Z: 6376.71
1412000
1416000
1420000
1424000
22132000
22128000
22124000
22120000
22116000
22112000
22108000
22104000
Y: 22100479.32 ft
Z: 6376.71
8000
10000
N30
N2
N10
N26
S9
Z: 12460.95 ft
Y: 22138377.35
22136000
Z: 12460.95 ft
X: 1430622.56
1428000
1408000
Y: 22100479.32
Z

Figure 5.4: 3D model of reservoir.

5.3.1 Data Description

The workflow execution for North Sea case was based on 18 numerical simulations (run in PumaFlow^{TM}) for training dataset and 4 simulations for the test dataset. Table 5.1 enumerates the uncertain parameters taken with the prospect of history match. Tables 5.2 and 5.3 document the values of uncertain parameters for the training and test dataset, respectively.

Table 5.1: Uncertain Parameters.

Uncertain Parameter	Min	Max	Unit
Rock Compressibility	1.38E-06	5.52E-06	1/psi
Fault (FLT1) Transmissivity	0	1	
Fault (FLT2) Transmissivity	0	1	
KRWM RockType-1 formation	0.05	0.5	
KRWM RockType-2 formation	0.05	0.5	
Aquifer volume	1.887E+07	1.887E+13	stb

Table 5.2: Numerical Simulations for training dataset.

Sim	Compress	FLT1	FLT2	KRWM_RT1	KRWM_RT2	Volume
1	1.87E-06	0.588	0.412	0.076	0.447	4.44E+12
2	2.35E-06	0.824	0.176	0.341	0.341	1.33E+13
3	5.03E-06	0.882	0.000	0.474	0.500	1.11E+12
4	3.08E-06	0.176	0.882	0.156	0.103	5.55E+12
5	3.57E-06	0.294	0.059	0.368	0.394	1.22E+13
6	5.27E-06	0.412	0.824	0.315	0.474	1.89E+07
7	4.30E-06	0.765	0.941	0.103	0.076	8.88E+12
8	3.81E-06	0.000	0.235	0.182	0.315	6.66E+12
9	4.54E-06	0.353	0.588	0.050	0.050	1.55E+13
10	5.52E-06	0.118	0.765	0.421	0.288	1.78E+13
11	2.11E-06	0.529	0.647	0.500	0.262	1.44E+13
12	1.38E-06	0.647	0.471	0.394	0.421	3.33E+12
13	2.60E-06	0.941	0.706	0.235	0.368	1.89E+13
14	4.06E-06	0.706	0.118	0.288	0.182	1.11E+13
15	3.33E-06	1.000	1.000	0.129	0.156	1.66E+13
16	4.79E-06	0.235	0.529	0.262	0.129	9.99E+12
17	2.84E-06	0.059	0.353	0.447	0.235	2.22E+12
18	1.62E-06	0.471	0.294	0.209	0.209	7.77E+12

Table 5.3: Numerical Simulations for test dataset.

Sim	Compress	FLT1	FLT2	KRWM_RT1	KRWM_RT2	Volume
1	4.14E-06	0.667	0.333	0.500	0.050	6.29E+12
2	1.38E-06	1.000	0.000	0.350	0.500	1.26E+13
3	5.52E-06	0.333	1.000	0.050	0.200	1.89E+07
4	2.76E-06	0.000	0.667	0.200	0.350	1.89E+13

5.3.2 Experiments

Several experiments were performed in order to derive the best set of hyperparameters, to train the machine learning model. Following summarizes various configurations of machine learning models experimented with:

- **Experiment 1 (Base Case):** 3 layered MLP with Sigmoid activation functions
- **Experiment 2:** 3 layered MLP with ReLU activation functions
- **Experiment 3:** 2 layered MLP with ReLU activation functions
- **Experiment 4:** 1 layered MLP with ReLU activation function
- **Experiment 5:** 1 layered MLP with Swish activation function
- **Experiment 6:** 1 layered MLP with Leaky ReLU activation function

Table 5.4 summarizes the accuracy of the above experiments measured at the wells.

Table 5.4: Training and test accuracy of ML based simulations (at wells)

Experiments	Training Error (MAE)	Test Error (MAE)
1	13.60	25.88
2	14.39	42.17
3	13.12	19.59
4	12.23	21.83
5	12.15	38.63
6	13.37	15.33

5.3.3 Results and Discussion

The machine learning based reservoir simulation workflow was enacted on the case of North Sea oil and gas field. The goal of the workflow was to build a machine learning based surrogate reservoir model, replacing the numerical simulation model.

In order to achieve the above goal, firstly the reservoir fluid flow numerical simulation model of the North Sea oil and gas reservoir was subjected to sensitivity runs by varying 6 parameters as given by Table 5.1. These uncertain parameters were chosen in order to aim for history matching. Table 5.2 presents the 18 sensitivity simulations with varying values of the said uncertain parameters. The goal of these sensitivity simulations was to build a training data set, to be used for machine learning (as described by section 5.1). Additionally, another four numerical simulations were performed in order to build a test database, with the same uncertain parameters (see Table 5.3). The training and test dataset were prepared for the 18 and 4 sensitivity numerical simulations respectively, considering the uncertain parameter values, cell X, Y, Z locations and pressure values at every time step. These feature values were also scaled as per Z-score scaling (see Section 2.4).

Experiment 1 (Base Case): MLP with 3 layers activated by Sigmoid functions

To achieve the proof of concept for machine learning based simulations as described in Section 5.1, a base case (also referred to as Experiment 1) MLP neural network with 3 layers activated by Sigmoid functions was trained on the training dataset. The trained model was then used to perform machine-learning-based simulations. Figure 5.5 shows the accuracy of the simulations on the training dataset, and also forms a proof of concept, indicating that this workflow is able to produce accurate (if not precise) pressure profiles by building a surrogate reservoir model, on various sensitivity parameter configurations. Moreover, Figure 5.6 shows different pressure profiles per simulation and per well respectively.

The model was then evaluated on test parameter configurations in order to quantify the ability of the workflow to produce coherent machine learning based simulation results for blind test parameter configurations. As depicted by Figure 5.7, few pressure profiles were incoherent with the numerical simulation model behaviour, hence decreasing the test accuracy of the model. The reason behind such incoherent results was hypothesized to MLP model being very complex.

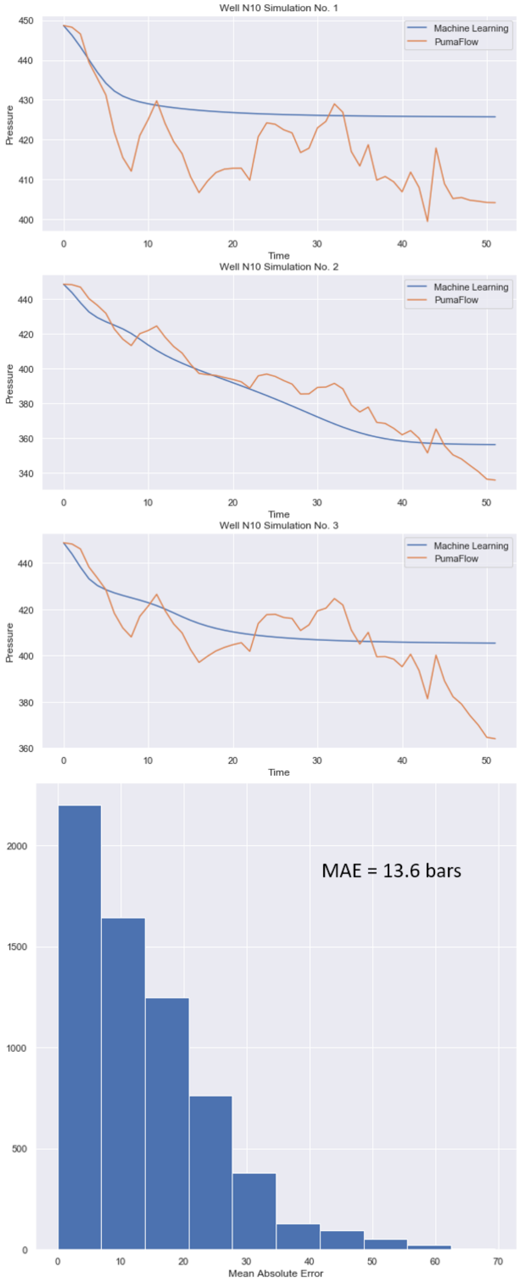

Figure 5.5: Experiment 1: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset

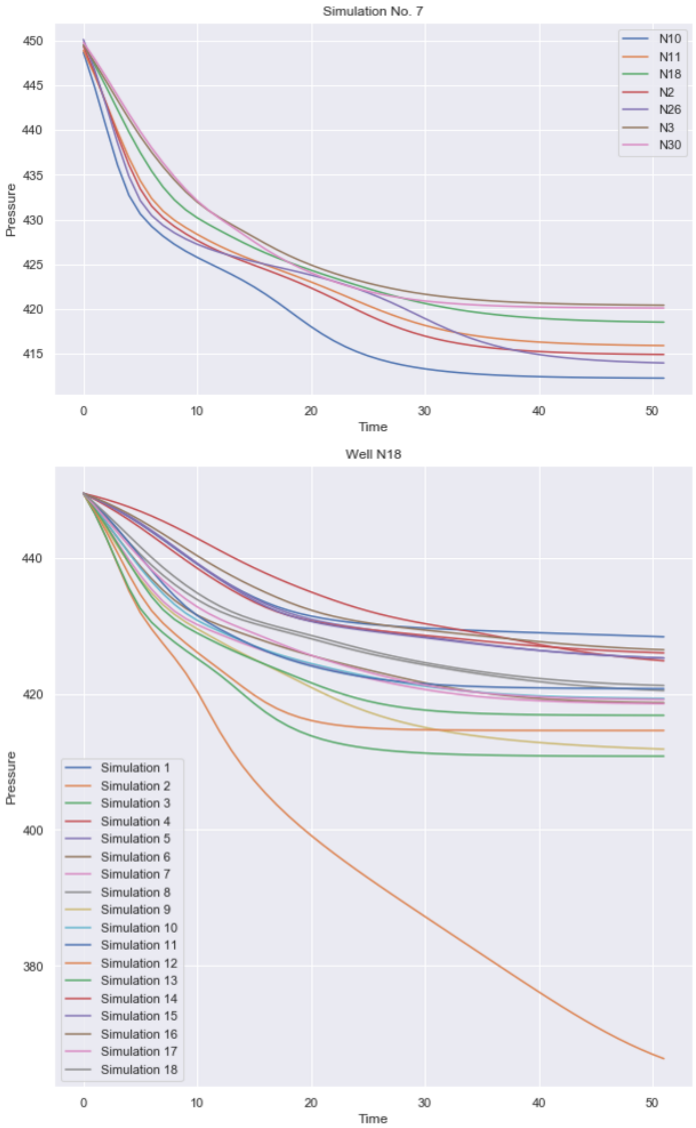

Figure 5.6: Experiment 1: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot).

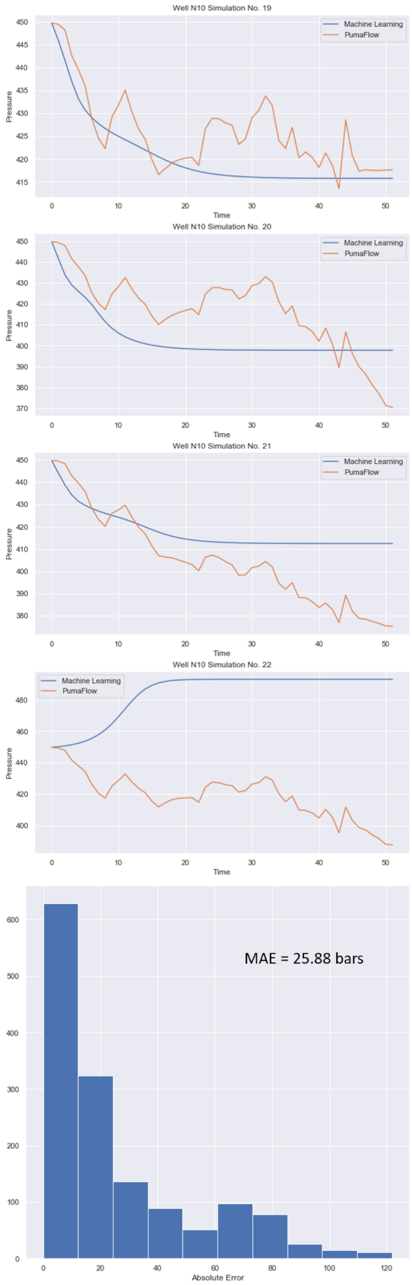

Figure 5.7: Experiment 1: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells.

Experiments 2 and 3

Hence, additional experiments were performed to optimize for the size of the MLP and activation functions involved. The evaluation of these experiments were based on the accuracy of machine-learning-based simulations on these training and test cases. Table 5.4 summarizes the training and test accuracy of six experiments conducted by varying the size and activation functions involved in the MLP neural network.

To start with, Experiment 2 was conducted with a 3 layered MLP activated by ReLU functions. Figures 5.8 and 5.9 show a sample comparison between ML-based simulation and the corresponding physics-based (PumaFlow^{TM}) simulation. Section A.1 describes more results of the experiment. The model showed a good training accuracy but the test accuracy was even worse (as compared to Experiment 1). Next, Experiment 3 was conducted with a 2 layered MLP activated by ReLU functions. Several test cases for MLP based simulations were incoherent as illustrated by Figure 5.10 (see Section A.2 for more results). Thus the size of the MLP was further reduced to 1 layered models with varying activation functions as described in experiments 4, 5 and 6.

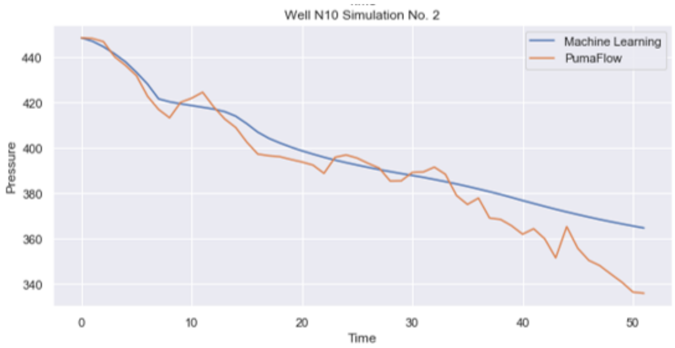

Figure 5.8: Experiment 2: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 2 (training case).

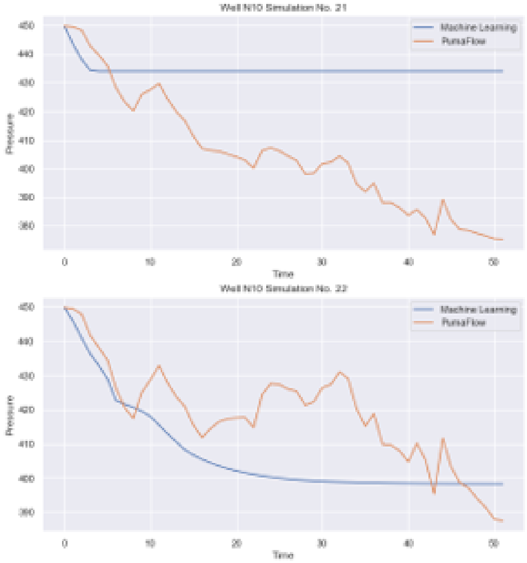

Figure 5.9: Experiment 2: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 21, on top and No. 22, on bottom (test cases).

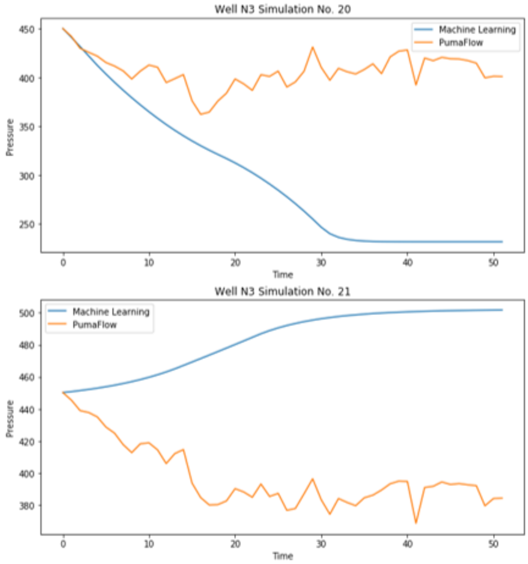

Figure 5.10: Experiment 3: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N3) for simulation No. 20, on top and No. 21, on bottom (test cases).

Experiments 4 – 6

Experiment 4 considered a 1 layered MLP activated with ReLU function. This resulted in high training accuracy and coherent test pressure profiles as depicted by Figures 5.11 and 5.12 (see Section A.3 for more results). Also, Experiment 5 considered 1 layered MLP activated with Swish activation function. Experiment 5 showed good training accuracy and test accuracy, except for one test simulation case as given by Figure 5.13 (see Section A.4 for more results). It is quite possible, that the 18 sensitivity simulations were not able to exhaustively map the uncertainty space produced by the 6 parameters. In that case, it is quite normal to get low accuracy for few test simulations. Nevertheless, the produced pressure profiles were of coherent behaviour.

Finally, Experiment 6 was conducted with 1 layered MLP activated by LeakyReLU function. The model produced high training and test accuracy as depicted by Figures 5.14 and 5.15 (see Section A.5 for more results). Due to highly accurate ML-based pressure simulations, Experiment 6 was concluded as the optimized MLP configuration for the North Sea case.

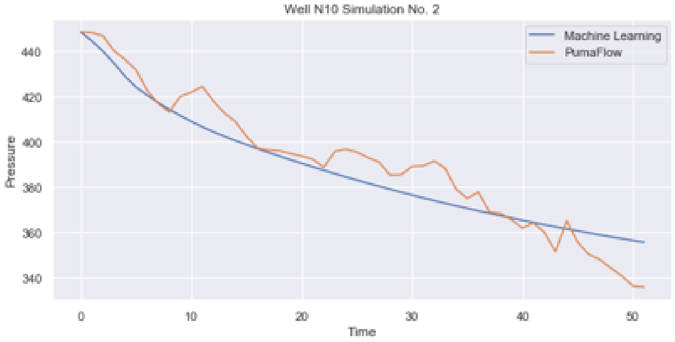

Figure 5.11: Experiment 4: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 2 (training case).

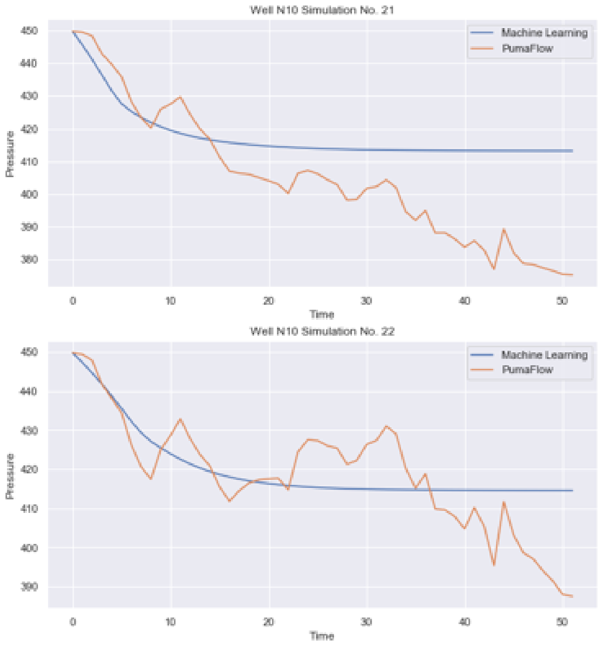

Figure 5.12: Experiment 4: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 21, on top and No. 22, on bottom (test cases).

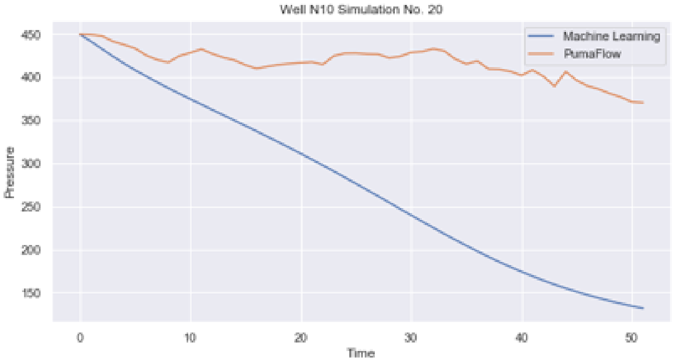

Figure 5.13: Experiment 5: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 20 (test case).

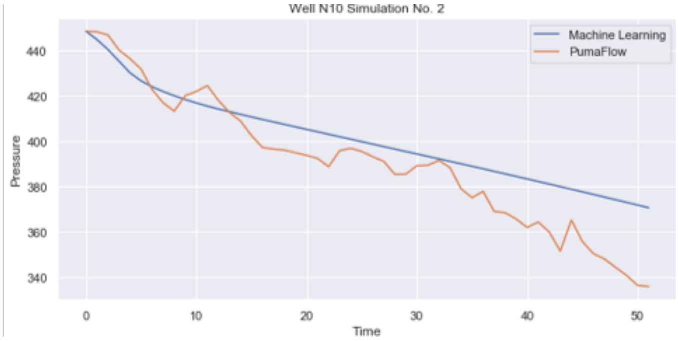

Figure 5.14: Experiment 6: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 2 (training case).

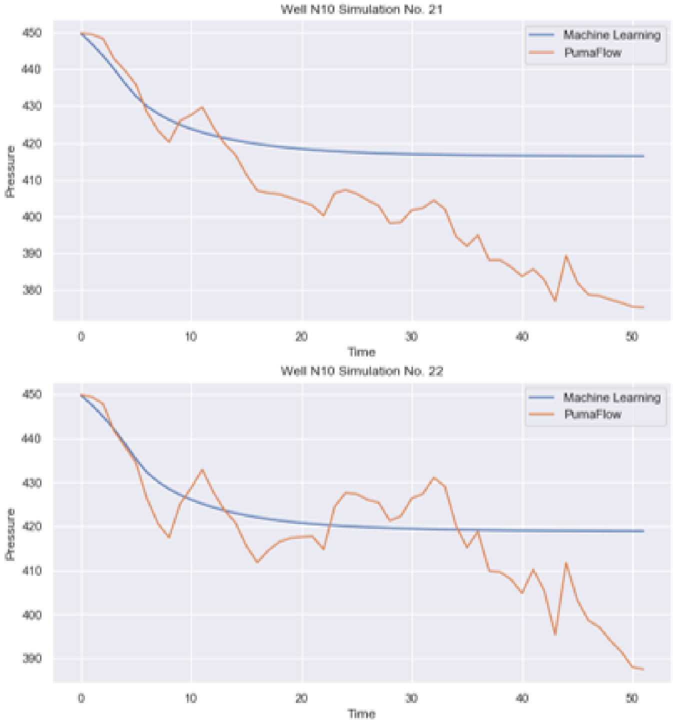

Figure 5.15: Experiment 6: This figure compares the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for simulation No. 21, on top and No. 22, on bottom (test cases).

Computational Efficiency

The largest part of required computation in this workflow was limited to the training of machine learning model which was achieved in a few minutes (less than 10 minutes). The model training was performed on a Nvidia RTX 2080Ti GPU [33]. Moreover, the trained surrogate reservoir model provided instantaneous pressure simulations. This speaks to a huge advantage of using this workflow, as one can simulate reservoir pressures with varying uncertain parameter configurations, instantaneously. This is in contrast to computationally expensive physics-based simulations which can take hours of computation.

Additionally, such surrogate reservoir models can improve the conventional approach of history matching, where gradient-based optimization methods are used to minimize the objective function (error between simulation results and observed data), due to the computationally expensive nature of physics-based simulations.

The main disadvantage of using gradient-based optimization is its tendency to provide local optimum rather than global optimum [36]. In contrast, ML-based surrogate reservoir models, being very fast, can work very well for this challenge, by simulating several scenarios to find a global optimum as the history match.

Chapter 6

Conclusion

In spite the widespread popularity of physics-based oil and gas reservoir simulation, its applications in oil and gas industry can prove to be highly counterproductive. Several limitations of this method originate from underlying analytical and numerical assumptions, modelling uncertainties and computational resources. The aim of this thesis was to explore the application of Artificial Intelligence and Data Science methods in oil and gas reservoir simulation, in order to analyze, model and simulate reservoir dynamics using data driven workflows, without constructing the said physics-based simulation models. Based on the experiments conducted on spatiotemporal modelling of reservoir data in Chapter 4 and machine learning based reservoir simulations in Chapter 5, it can be fairly concluded that the application of such modern techniques possesses a huge potential to improve the current conventional practices and provide highly productive insight on the fluid flow dynamics of the reservoir.

The research on spatiotemporal modelling on reservoir data showed various possibilities to use data science methods to analyze and model reservoir data. Multivariate data analysis of reservoir pressure showed the capability of mapping reservoir dynamics against space and time and evaluate the heterogeneity of the same. Moreover, it was proved that modelling techniques like generalized additive models and machine learning can capture the overall dynamic behaviour of oil and gas reservoir across the overall domain of space and time. Techniques of spatiotemporal interpolation has proved its utility to model spatially and temporally localized processes accurately, thus contributing towards a much more accurate spatiotemporal model. On the other hand, the research on machine learning based reservoir simulation provided the proof of concept of building a surrogate reservoir model in order to replace the reservoir numerical simulation model. The machine-learning-based simulations are proved to be accurate enough to be used for applications like history matching. Such simulations being computationally very cheap, can provide basis for simulating several scenarios to perform history matching in order to achieve a global optimum.

The three research questions studied for in this thesis are reiterated and answered as below:

- **Q1.** Would data science methods be effective to identify and analyze well pressure heterogeneity and associated geological uncertainties, in spatial and temporal context for a long time ranged (in order of decades) and heterogeneous oil and gas reservoir dynamics?
  The reservoir pressures represented by a few observation points (i.e., at wells) and sparse time stamps, were analyzed for their heterogeneity in space and time using multivariate analysis methods and data modelling methods. Firstly the spatial and temporal features were engineered to more meaningfully represent the reservoir spatial and temporal contexts. Next, the pressure data was mapped against the engineered spatial features (Theta and Z) and temporal features (date). Such multivariate analysis provided a rudimentary understanding of the heterogeneity in reservoir dynamics. Tools like multivariate correlation matrix quantified the influence of individual feature to the pressure variance.

  Furthermore, modelling techniques like generalized additive models were used to model the said pressure heterogeneity in space and time. The pressure prediction hyperspace was computed against the spatial and temporal features to get a better understanding of pressure evolution in various parts of the reservoir volume through time. Additionally, analysis of modelling residuals/errors provided an insight to possible multi-dynamics at play at different time duration i.e. temporal heterogeneity.

- **Q2.** How data science methods could be used to simulate pressure in 3D space and time (within the spatial and temporal domain of observed pressure data at oil and gas wells) without constructing a physics-based reservoir numerical simulation model?
  Spatiotemporal (ST) modelling of reservoir pressure from sparse well bottom hole pressure data was performed by a multi step workflow (see Chapter 4). After the analysis of pressure variance against space and time, the pressure was modelled as a function of the spatial and temporal features (i.e. Theta, Z and date). In order to achieve accurate modelling, methods of generalized additive modelling and machine learning were investigated. Such

modelling methods produced accurate pressure predictions for any point in space and time.

Moreover, the modelling residuals were analyzed for their correlation to space and time. If uncorrelated residuals, the local processes in space and time were hypothesized for the reason behind such residuals. To model such localized dynamic processes, methods of spatiotemporal interpolation like spatiotemporal kriging were investigated. ST interpolation mapped the residuals to all points in space and time. Consequently, the residuals were added to the modelled pressure predictions thus contributing towards a more accurate pressure spatiotemporal model.

- **Q3.** Would machine learning based techniques like multi-layer perceptron (MLP) improve the current reservoir simulation practices, specifically for the scope of history matching of an oil and gas reservoir simulation model? The current/conventional practices of reservoir fluid flow simulation are based on solving partial differential equations over several timesteps and several spatial locations (cells), which can prove to be highly time consuming and computationally expensive. This was improved by using machine learning to build surrogate reservoir model to obtain accurate (if not precise) and very fast fluid flow simulations, thereby replacing the numerical simulation model.

  As described in Chapter 5, the above machine learning model was trained by using a few numerical simulations as training data, in order to approximate the fluid flow dynamics. Next, the trained machine learning model was used to simulate the pressures starting from initial conditions and to subsequent time steps using a recursive algorithm.

  Such fast alternative to conventional physics-based reservoir simulations provide an opportunity to improve the current approach of history matching which use gradient-based optimization methods, due to computationally expensive nature of physics-based simulations. Gradient-based optimization methods usually produce local optimum thus comprising the quality of history match. ML-based simulations can be used to simulate several scenarios very quickly, to find globally optimized history match.

6.1 Thesis Contribution

The oil and gas industry works with the most challenging industrial processes across several domains of engineering. In order to properly monitor and control such processes, this industry deals with vast amounts and sources of data. As such, data-driven methods and workflows can prove to be natural procedures for more effective and more safe production of oil and gas.

This thesis provides a comprehensive exploration of such data-driven techniques for the domain of oil and gas reservoir simulation. The importance of this research is highlighted by the magnitude of the problem while dealing with accurate mapping of oil and gas reservoir dynamics. The applied workflows provide predictions with average errors less than 60 psi for average reservoir pressures more than 4000 psi. This work shows high prediction accuracy for a dynamic system several thousands of feet below the surface, thus highly intractable to accurately measure and map. Following encapsulates the research domains where this thesis makes major contributions:

- **Reservoir Data Analysis:** Proposal of multivariate analysis framework, implemented on a real oil and gas reservoir data
- **Spatiotemporal modelling of reservoir data:** Proposal of multi-step workflow to model sparse reservoir data gathered at wells, across the domain of space and time. This workflow was implemented with real oil and gas reservoir data, and evaluated for several workflow component configurations.
- **Surrogate reservoir modelling:** Proposal of a proxy modelling workflow based on machine learning, also implemented and optimized for a real data case.
- **Reservoir Big Data Framework Development:** Development of reservoir big data analysis and modelling framework incorporating different platforms (like R and Python), implementing techniques like machine learning, generalized additive modelling, parallel computing, spatiotemporal interpolation, etc., using open source libraries thus posing no additional cost.

6.2 Future Work

This thesis has provided the proof of concept to use data-driven techniques in the scope of oil and gas reservoir engineering and simulation. However, additional research can be performed on several fronts as listed below:

- In this thesis, reservoir data analysis has been performed for a real oil field data case with spatial and temporal features. However, additional features like production and injection density (indicating scaled amount of production or injection in the vicinity of the observation well) can be included to make the analysis more accurate and comprehensive.

- The spatiotemporal modelling workflow has been applied to model reservoir pressures in this thesis. However, it can be applied to model other properties like production ratios (water cut, gas-oil ratio), gas-oil contact, water-oil contact, etc. In general, this workflow can be used to model any property varying over space and time.

- The spatiotemporal modelling, in this research work, was performed using a quasi-cylindrical coordinate system (Theta and Z) to represent space, as the applied case consisted of an oil and gas reservoir which was relatively homogeneous, with no reservoir compartmentalization (i.e. no major geological heterogeneity). However, a cylindrical coordinate system (Theta, R and Z) can also be used, which can be useful to capture heterogeneity like sealing faults, compartmentalization, etc.

- This thesis has aimed to provide proof of concept for machine learning based reservoir simulations and has implemented the same on a real oil and gas field data. However, the same can be carried forward to implement optimization methods like Genetic algorithm (GA) and Particle swarm optimization (PSO) [36] to conduct history matching and obtain a global optimum as a history match.

References

[1] J. Allaire and F. Chollet. *keras: R Interface to 'Keras'*. R package version 2.2.4.1. 2019. URL: `https://CRAN.R-project.org/package=keras`.

[2] S. Amini and S. Mohaghegh. "Application of Machine Learning and Artificial Intelligence in Proxy Modeling for Fluid Flow in Porous Media". In: *Fluids* 4.3 (2019). ISSN: 2311-5521. DOI: `10.3390/fluids4030126`. URL: `https://www.mdpi.com/2311-5521/4/3/126`.

[3] J. W. Amyx. "Petroleum reservoir engineering : physical properties". eng. In: New York: McGraw-Hill, 1960, pp. 1–35. ISBN: 0070016003.

[4] K. Aziz, K. Aziz and A. Settari. *Petroleum reservoir simulation*. London: Applied Science Publishers, 1979. ISBN: 9780853347873. URL: `https://books.google.no/books?id=GJ5TAAAAMAAJ`.

[5] K. Aziz and L. Durlofsky. *Notes on Reservoir Simulation*. Stanford University. Stanford, 2004.

[6] Beicip-Franlab. *OpenFlow Suite*^{TM}. 2019. URL: `http://www.beicip.com/openflow-suite` (visited on 01/06/2020).

[7] A. Burkov. *The Hundred-Page Machine Learning Book*. Quebec City: Andriy Burkov, 2019. ISBN: 9781999579517. URL: `https://books.google.no/books?id=0jbxwQEACAAJ`.

[8] P. E. Carreras, S. E. Turner and G. T. Wilkinson. "Tahiti: Development Strategy Assessment Using Design of Experiments and Response Surface Methods". In: Anchorage, Alaska, USA: Society of Petroleum Engineers, Jan. 2006, p. 11. ISBN: 978-1-55563-234-2. DOI: `10.2118/100656-MS`. URL: `https://doi.org/10.2118/100656-MS`.

[9] G. Chavent and J. Jaffré. *Mathematical Models and Finite Elements for Reservoir Simulation: Single Phase, Multiphase and Multicomponent Flows through Porous Media*. ISSN. Amsterdam: Elsevier Science, 1986. ISBN: 9780080875385. URL: `https://books.google.no/books?id=sPVl-ZofdvkC`.

[10] H. Chen, H. Klie and Q. Wang. "A Black-Box Interpolation Method To Accelerate Reservoir Simulation Solutions". In: The Woodlands, Texas, USA: Society of Petroleum Engineers, Feb. 2013, p. 16. ISBN: 978-1-61399-233-3. DOI: `10.2118/163614-MS`. URL: `https://doi.org/10.2118/163614-MS`.

[11] Z. Chen, G. Huan and Y. Ma. *Computational Methods for Multiphase Flows in Porous Media*. Computational Science & Engineering. Philadelphia: Society for Industrial and Applied Mathematics, Jan. 2006, p. 523. ISBN: 978-0-89871-606-1. DOI: `10.1137/1.9780898718942`. URL: `https://doi.org/10.1137/1.9780898718942`.

[12] F. Chollet. *keras*. `https://github.com/fchollet/keras`. 2015.

[13] K. H. Coats. "Use and Misuse of Reservoir Simulation Models". In: *Journal of Petroleum Technology* 21.11 (Nov. 1969), pp. 1391–1398. ISSN: 0149-2136. DOI: `10.2118/2367-PA`. URL: `https://doi.org/10.2118/2367-PA`.

[14] K. Coats. "Reservoir Simulation: State of the Art (includes associated papers 11927 and 12290)". In: *Journal of Petroleum Technology* 34.08 (Aug. 1982), pp. 1633–1642. ISSN: 0149-2136. DOI: `10.2118/10020-PA`. URL: `https://doi.org/10.2118/10020-PA`.

[15] L. Cosentino. "Integrated Reservoir Studies". In: Institut français du pétrole publications. Paris: Editions Technip, 2001, pp. 2–5. ISBN: 9782710807971. URL: `https://books.google.no/books?id=vGWeCgAAQBAJ`.

[16] L. Cosentino. "Integrated Reservoir Studies". In: Institut français du pétrole publications. Paris: Editions Technip, 2001, pp. 264–267. ISBN: 9782710807971. URL: `https://books.google.no/books?id=vGWeCgAAQBAJ`.

[17] L. Cosentino. "Integrated Reservoir Studies". In: Institut français du pétrole publications. Paris: Editions Technip, 2001, pp. 247–288. ISBN: 9782710807971. URL: `https://books.google.no/books?id=vGWeCgAAQBAJ`.

[18] B. C. Craft. *Applied petroleum reservoir engineering*. eng. 2nd ed. revised by Ronald E. Terry. Englewood Cliffs, NJ: Prentice-Hall, 1991. ISBN: 0130398845.

[19] N. Cressie and C. Wikle. *Statistics for Spatio-Temporal Data*. CourseSmart Series. New Jersey: Wiley, 2011. ISBN: 9780471692744. URL: `https://books.google.no/books?id=-kOC6D0DiNYC`.

[20] B. Csaji. “Approximation with Artificial Neural Networks”. MA thesis. Budapest, Hungary: Dept. Science, Eotvos Lorand University, 2001. URL: `https://ci.nii.ac.jp/naid/20001716508/en/`.

[21] L. P. Dake. *The practice of reservoir engineering*. eng. Vol. 36. Developments in petroleum science. Amsterdam: Elsevier, 1994. ISBN: 0444885382.

[22] V. I. Dzyuba et al. “Application of Sector Modeling Technology for Giant Reservoir Simulations”. In: Moscow, Russia: Society of Petroleum Engineers, Jan. 2012, p. 10. ISBN: 978-1-61399-214-2. DOI: `10.2118/162090-MS`. URL: `https://doi.org/10.2118/162090-MS`.

[23] M. M. El Faidouzi and D. E. Ouzzane. “Data-Driven Analytics: A Novel Approach to Performance Diagnosis Using SpatioTemporal Analysis in a Giant Field Offshore Abu Dhabi”. In: Abu Dhabi, UAE: Society of Petroleum Engineers, Nov. 2018, p. 16. ISBN: 978-1-61399-632-4. DOI: `10.2118/192759-MS`. URL: `https://doi.org/10.2118/192759-MS`.

[24] E. A. Elrafie et al. “Innovated Simulation History Matching Approach Enabling Better Historical Performance Match and Embracing Uncertainty in Predictive Forecasting”. In: Amsterdam, The Netherlands: Society of Petroleum Engineers, Jan. 2009, p. 15. ISBN: 978-1-61399-427-6. DOI: `10.2118/120958-MS`. URL: `https://doi.org/10.2118/120958-MS`.

[25] E. Esor, S. Dresda and C. Monico. “Use of Material Balance to Enhance 3D Reservoir Simulation: A Case Study”. In: Houston, Texas: Society of Petroleum Engineers, Jan. 2004, p. 6. ISBN: 978-1-55563-151-2. DOI: `10.2118/90362-MS`. URL: `https://doi.org/10.2118/90362-MS`.

[26] P. L. Essley Jr. “What is Reservoir Engineering?” In: *Journal of Petroleum Technology* 17.01 (Jan. 1965), pp. 19–25. ISSN: 0149-2136. DOI: `10.2118/920-PA`. URL: `https://doi.org/10.2118/920-PA`.

[27] A. Fornel and A. Estublier. “To A Dynamic Update Of The Sleipner CO2 Storage Geological Model Using 4d Seismic Data”. In: *Energy Procedia* 37 (Dec. 2013), pp. 4902–4909. DOI: `10.1016/j.egypro.2013.06.401`.

[28] E. Generalic. *Petroleum*. 2018. URL: `https://glossary.periodni.com` (visited on 01/06/2020).

[29] M. G. Genton. “Separable approximations of space-time covariance matrices”. In: *Environmetrics* 18.7 (2007), pp. 681–695. DOI: `10.1002/env.854`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.854`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/env.854`.

[30] F. B. Gorucu et al. “A Neurosimulation Tool for Predicting Performance in Enhanced Coalbed Methane and CO2, Sequestration Projects”. In: Dallas, Texas: Society of Petroleum Engineers, Jan. 2005, p. 14. ISBN: 978-1-55563-150-5. DOI: `10.2118/97164-MS`. URL: `https://doi.org/10.2118/97164-MS`.

[31] B. Graeler, E. Pebesma and G. Heuvelink. “Spatio-Temporal Interpolation using gstat”. In: *The R Journal* 8 (Jan. 2016), pp. 204–218. DOI: `10.32614/RJ-2016-014`.

[32] B. Gräler, E. Pebesma and G. Heuvelink. “Spatio-Temporal Interpolation using gstat”. In: *The R Journal* 8.1 (2016), pp. 204–218. DOI: `10.32614/RJ-2016-014`. URL: `https://doi.org/10.32614/RJ-2016-014`.

[33] *Graphics Reinvented: NVIDIA GeForce RTX 2080 Ti Graphics Card*. [Online; accessed 15-June-2020]. URL: `https://www.nvidia.com/en-us/geforce/graphics-cards/rtx-2080-ti/`.

[34] P. Grover. “Formation Evaluation MSc Course Notes”. In: Aberdeen: Aberdeen University, 2001, pp. 19–34.

[35] T. Hastie and R. Tibshirani. *Generalized Additive Models*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Abingdon: Taylor & Francis, 1990. ISBN: 9780412343902. URL: `https://books.google.no/books?id=qa29r1Ze1coC`.

[36] R. Haupt and S. Haupt. *Practical Genetic Algorithms*. Wiley InterScience electronic collection. New Jersey: Wiley, 2004. ISBN: 9780471671756. URL: `https://books.google.no/books?id=k0jFfsmbtZIC`.

[37] D. Havlena and A. S. Odeh. “The Material Balance as an Equation of a Straight Line”. In: *Journal of Petroleum Technology* 15.08 (Aug. 1963), pp. 896–900. ISSN: 0149-2136. DOI: `10.2118/559-PA`. URL: `https://doi.org/10.2118/559-PA`.

[38] T. Hengl et al. “Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images”. In: *Theoretical and Applied Climatology* 107.1 (Jan. 2012), pp. 265–277. ISSN: 1434-4483. DOI: `10.1007/s00704-011-0464-2`. URL: `https://doi.org/10.1007/s00704-011-0464-2`.

[39] A. Journel and C. Huijbregts. *Mining Geostatistics*. New Jersey: Blackburn Press, 2003. ISBN: 9781930665910. URL: `https://books.google.no/books?id=Id1GAAAAYAAJ`.

[40] P. K. Dunn and G. K. Smyth. *Generalized Linear Models With Examples in R*. New York: Springer-Verlag New York, 2018. ISBN: 978-1-4419-0117-0. DOI: 10.1007/978-1-4419-0118-7. URL: https://link.springer.com/book/10.1007/978-1-4419-0118-7.

[41] A. U. Al-Kaabi and W. J. Lee. “Using Artificial Neural Networks To Identify the Well Test Interpretation Model (includes associated papers 28151 and 28165)”. In: *SPE Formation Evaluation* 8.03 (Sept. 1993), pp. 233–240. ISSN: 0885-923X. DOI: 10.2118/20332-PA. URL: https://doi.org/10.2118/20332-PA.

[42] J. Kelleher, B. Namee and A. D’Arcy. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. The MIT Press. Massachusetts: MIT Press, 2015. ISBN: 9780262029445. URL: https://books.google.no/books?id=uZxOCgAAQBAJ.

[43] R. M. Lark. “Kriging a soil variable with a simple nonstationary variance model”. In: *Journal of Agricultural, Biological, and Environmental Statistics* 14.3 (Sept. 2009), pp. 301–321. ISSN: 1537-2693. DOI: 10.1198/jabes.2009.07060. URL: https://doi.org/10.1198/jabes.2009.07060.

[44] K. Larsen. *GAM: The Predictive Modeling Silver Bullet*. Aug. 2015. URL: https://multithreaded.stitchfix.com/blog/2015/07/30/gam/ (visited on 01/06/2020).

[45] S. Marsland. *Machine Learning: An Algorithmic Perspective, Second Edition*. Chapman & Hall. Abingdon: CRC Press, 2014. ISBN: 9781466583337. URL: https://books.google.no/books?id=6GvSBQAAQBAJ.

[46] G. Matheron. *Traité de géostatistique appliquée*. Memoires v. 1. Paris: Éditions Technip, 1962. URL: https://books.google.no/books?id=88YKAQAAMAAJ.

[47] S. Mohaghegh et al. “Design and Development of An Artificial Neural Network for Estimation of Formation Permeability”. In: *SPE Computer Applications* 7.06 (Dec. 1995), pp. 151–154. ISSN: 1064-9778. DOI: 10.2118/28237-PA. URL: https://doi.org/10.2118/28237-PA.

[48] S. Mohaghegh et al. “Petroleum reservoir characterization with the aid of artificial neural networks”. In: *Journal of Petroleum Science and Engineering* 16.4 (1996), pp. 263–274. ISSN: 0920-4105. DOI: https://doi.org/10.1016/S0920-4105(96)00028-9. URL: http://www.sciencedirect.com/science/article/pii/S0920410596000289.

[49] O. O. Mosobalaje, C. Y. Onuh and I. Seteyeobot. “A New Solution Methodology to the Material Balance Equation, for Saturated Reservoirs”. In: Lagos, Nigeria: Society of Petroleum Engineers, Aug. 2015, p. 12. ISBN: 978-1-61399-434-4. DOI: `10.2118/178392-MS`. URL: `https://doi.org/10.2118/178392-MS`.

[50] U. Murthy et al. *Digital Libraries with Superimposed Information: Supporting Scholarly Tasks that Involve Fine Grain Information*. May 2.

[51] K. Narayanan et al. “Response Surface Methods for Upscaling Heterogeneous Geologic Models”. In: Houston, Texas: Society of Petroleum Engineers, Jan. 1999, p. 2. ISBN: 978-1-55563-374-5. DOI: `10.2118/51923-MS`. URL: `https://doi.org/10.2118/51923-MS`.

[52] M. Negnevitsky. *Artificial Intelligence: A Guide to Intelligent Systems*. London: Pearson Education Limited, 2011. ISBN: 9781408225752. URL: `https://books.google.no/books?id=8mmpBwAAQBAJ`.

[53] Oil and G. Portal. *Integrated Reservoir Modeling*. URL: `http://www.oil-gasportal.com/reservoir-management/integrated-reservoir-modeling/` (visited on 01/06/2020).

[54] S. Okotie and B. Ikporo. “History Matching”. In: *Reservoir Engineering: Fundamentals and Applications*. Cham: Springer International Publishing, 2019, pp. 355–364. ISBN: 978-3-030-02393-5. DOI: `10.1007/978-3-030-02393-5_10`. URL: `https://doi.org/10.1007/978-3-030-02393-5_10`.

[55] D. Peaceman. *Fundamentals of Numerical Reservoir Simulation*. ISSN. Amsterdam: Elsevier Science, 2000. ISBN: 9780080868608. URL: `https://books.google.no/books?id=-DujQRDF4kwC`.

[56] F. Pérez and B. E. Granger. “IPython: a System for Interactive Scientific Computing”. In: *Computing in Science and Engineering* 9.3 (May 2007), pp. 21–29. ISSN: 1521-9615. DOI: `10.1109/MCSE.2007.53`. URL: `https://ipython.org`.

[57] Petroleum-Geology-Forums. *Upscaling from Static to Dynamic Models*. URL: `http://www.epgeology.com/gallery/image.php?album_id=12&image_id=190` (visited on 01/06/2020).

[58] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2019. URL: `https://www.R-project.org/`.

[59] R. Raghavan et al. "Integration of Geology, Geophysics, and Numerical Simulation in the Interpretation of a Well Test in a Fluvial Reservoir". In: *SPE Reservoir Evaluation & Engineering* 4.03 (June 2001), pp. 201–208. ISSN: 1094-6470. DOI: `10.2118/72097-PA`. URL: `https://doi.org/10.2118/72097-PA`.

[60] F. V. Rahimov, A. S. Eminov and R. M. Huseynov. "Risk and Uncertainty Assessment While Estimation of Reserves". In: Baku, Azerbaijan: Society of Petroleum Engineers, Nov. 2017, p. 5. ISBN: 978-1-61399-588-4. DOI: `10.2118/189019-MS`. URL: `https://doi.org/10.2118/189019-MS`.

[61] B. Riviere. *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*. Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 2008. ISBN: 9780898717440. URL: `https://books.google.no/books?id=JklTMrPZT3gC`.

[62] J. G. Ross. "The Philosophy of Reserve Estimation". In: Dallas, Texas: Society of Petroleum Engineers, Jan. 1997, p. 7. ISBN: 978-1-55563-411-7. DOI: `10.2118/37960-MS`. URL: `https://doi.org/10.2118/37960-MS`.

[63] M. Sherman. *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. Wiley Series in Probability and Statistics. New Jersey: Wiley, 2011. ISBN: 9780470974926. URL: `https://books.google.no/books?id=ebvy63s%5C_f6YC`.

[64] P. A. Slotte and E. Smorgrav. "Response Surface Methodology Approach for History Matching and Uncertainty Assessment of Reservoir Simulation Models". In: Rome, Italy: Society of Petroleum Engineers, Jan. 2008, p. 9. ISBN: 978-1-55563-178-9. DOI: `10.2118/113390-MS`. URL: `https://doi.org/10.2118/113390-MS`.

[65] A. A. Szpiro et al. "Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies". In: *Environmetrics* 21.6 (2010), pp. 606–631. DOI: `10.1002/env.1014`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.1014`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/env.1014`.

[66] W. R. Tobler. "A Computer Movie Simulating Urban Growth in the Detroit Region". In: *Economic Geography* 46 (1970), pp. 234–240. ISSN: 00130095, 19448287. URL: `http://www.jstor.org/stable/143141`.

[67] G. Van Rossum and F. L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN: 1441412697.

[68] F. Veronesi. *Spatio-Temporal Kriging in R*. URL: `https://www.r-bloggers.com/spatio-temporal-kriging-in-r/` (visited on 01/06/2020).

[69] M. C. Vincent. "Examining Our Assumptions – Have Oversimplifications Jeopardized Our Ability to Design Optimal Fracture Treatments?" In: The Woodlands, Texas: Society of Petroleum Engineers, Jan. 2009, p. 51. ISBN: 978-1-55563-208-3. DOI: `10.2118/119143-MS`. URL: `https://doi.org/10.2118/119143-MS`.

[70] R. F. Volz et al. "Field Development Optimization of Eastern Siberian Giant Oil Field Development under Uncertainty". In: Moscow, Russia: Society of Petroleum Engineers, Jan. 2008, p. 18. ISBN: 978-1-55563-203-8. DOI: `10.2118/116831-MS`. URL: `https://doi.org/10.2118/116831-MS`.

[71] J. W. Watts. "A Compositional Formulation of the Pressure and Saturation Equations". In: *SPE Reservoir Engineering* 1.03 (May 1986), pp. 243–252. ISSN: 0885-9248. DOI: `10.2118/12244-PA`. URL: `https://doi.org/10.2118/12244-PA`.

[72] M. Wieling. *Generalized additive modeling*. URL: `http://www.let.rug.nl/wieling/Statistics/GAM-Intro` (visited on 01/06/2020).

[73] Wikipedia. *Coefficient of determination — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Coefficient%20of%20determination&oldid=953904846`. [Online; accessed 22-June-2020]. 2020.

[74] Wikipedia. *Kronecker product — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Kronecker%20product&oldid=950806654`. [Online; accessed 12-June-2020]. 2020.

[75] Wikipedia. *Mean absolute error — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Mean%20absolute%20error&oldid=961700965`. [Online; accessed 22-June-2020]. 2020.

[76] Wikipedia. *Pearson correlation coefficient — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Pearson%20correlation%20coefficient&oldid=962708395`. [Online; accessed 16-June-2020]. 2020.

[77] S. Wood. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Texts in Statistical Science. Abingdon: Taylor & Francis, 2006. ISBN: 9781584884743. URL: `https://books.google.no/books?id=hr17lZC-3jQC`.

[78] V. van Zoest et al. “Spatio-temporal regression kriging for modelling urban NO2 concentrations”. In: *International Journal of Geographical Information Science* 34.5 (2020), pp. 851–865. DOI: `10.1080/13658816.2019.1667501`. eprint: `https://doi.org/10.1080/13658816.2019.1667501`. URL: `https://doi.org/10.1080/13658816.2019.1667501`.

Appendix A

ML-based reservoir simulation: Experiment results

This appendix accounts for the experiment results (for Experiments 2 – 6) performed for Machine-learning-based reservoir simulations as described in Chapter 5.

A.1 Results for experiment 2

- Training accuracy of machine-learning-based simulation at wells: see Figure A.1
- Machine-learning-based simulations: see Figure A.2
- Test accuracy of machine-learning-based simulation at wells: see Figure A.3

A.2 Results for experiment 3

- Training accuracy of machine-learning-based simulation at wells: see Figure A.4
- Machine-learning-based simulations: see Figure A.5
- Test accuracy of machine-learning-based simulation at wells: see Figure A.6

A.3 Results for experiment 4

- Training accuracy of machine-learning-based simulation at wells: see Figure A.7

- Machine-learning-based simulations: see Figure A.8
- Test accuracy of machine-learning-based simulation at wells: see Figure A.9

A.4 Results for experiment 5

- Training accuracy of machine-learning-based simulation at wells: see Figure A.10
- Machine-learning-based simulations: see Figures A.11
- Test accuracy of machine-learning-based simulation at wells: see Figure A.12

A.5 Results for experiment 6

- Training accuracy of machine-learning-based simulation at wells: see Figure A.13
- Machine-learning-based simulations: see Figures A.14
- Test accuracy of machine-learning-based simulation at wells: see Figure A.15

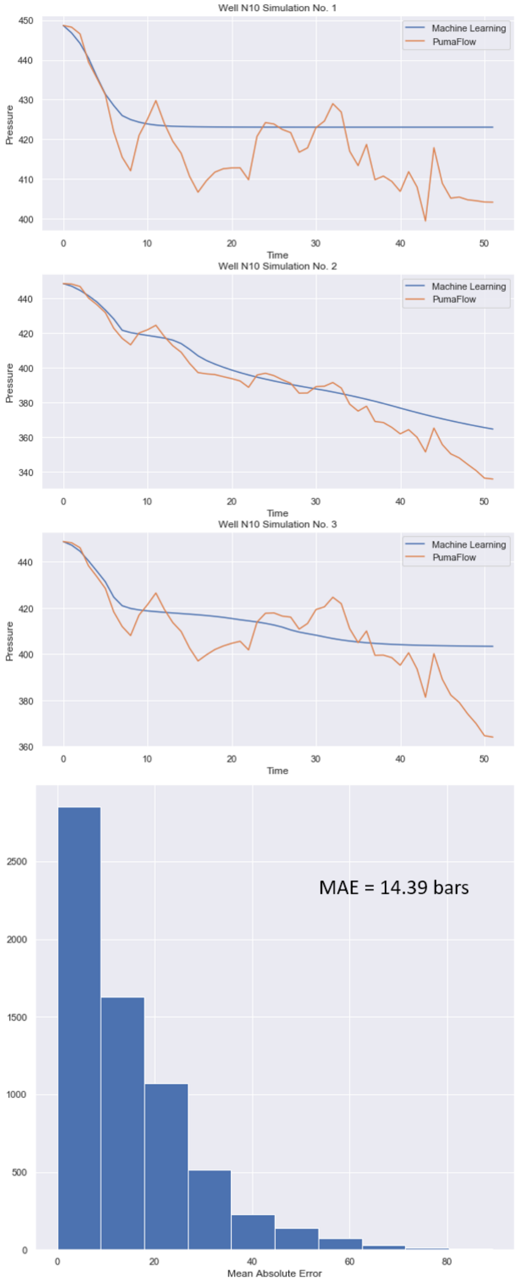

Figure A.1: Experiment 2: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset.

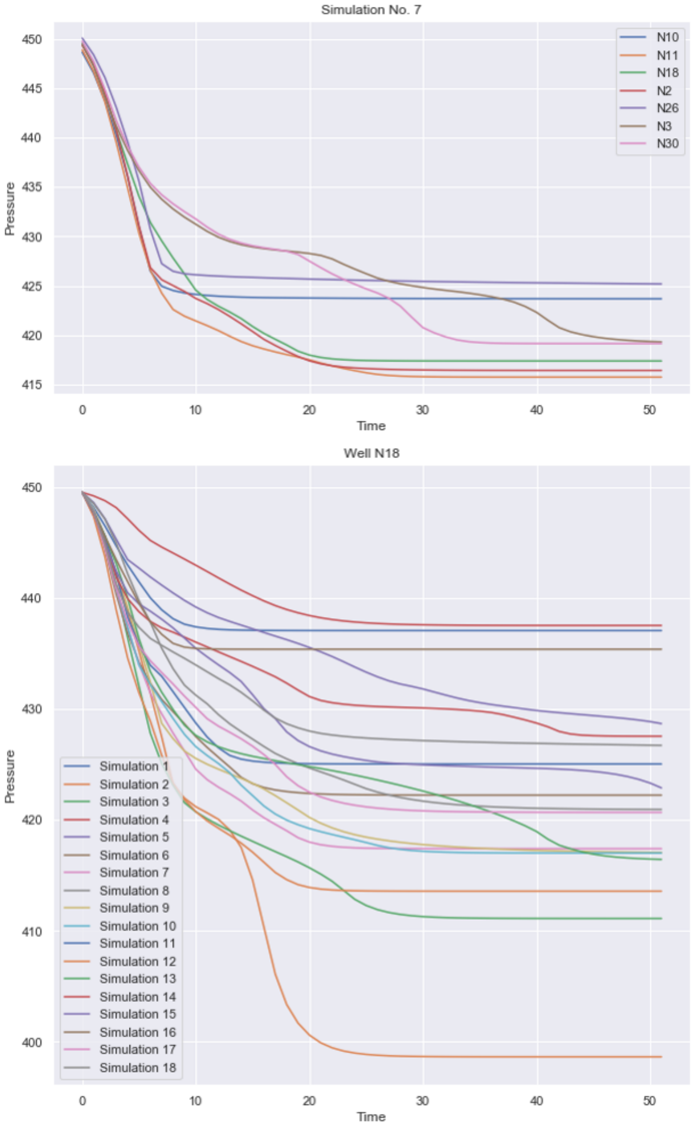

Figure A.2: Experiment 2: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot).

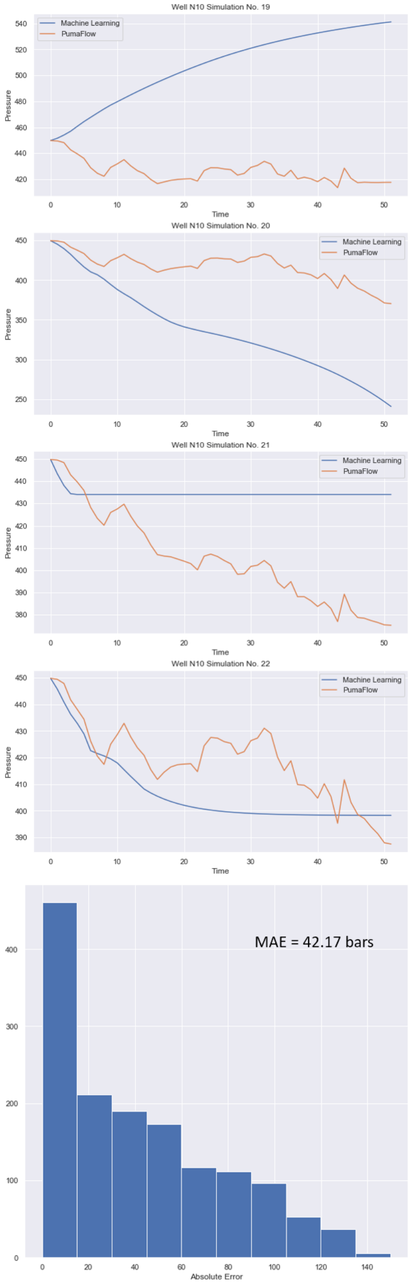

Figure A.3: Experiment 2: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells.

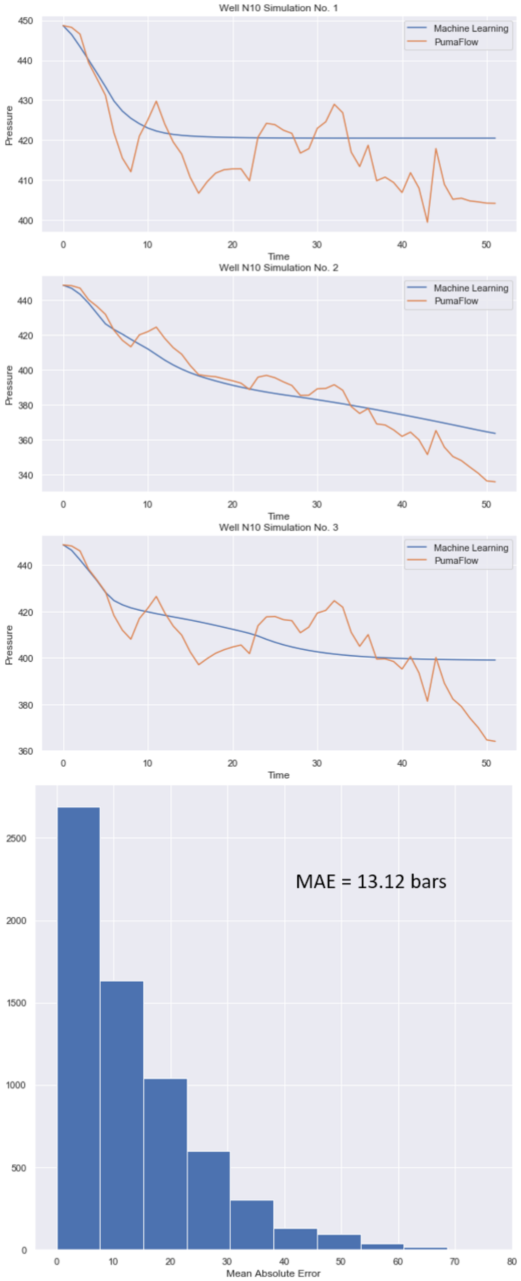

Figure A.4: Experiment 3: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset.

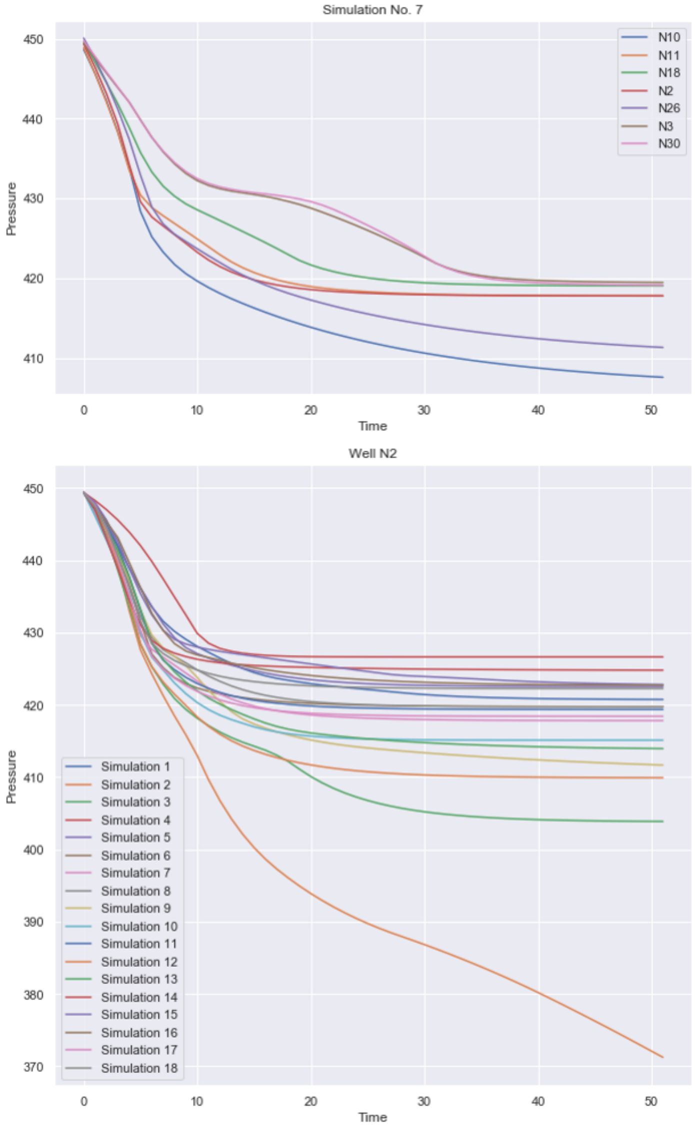

Figure A.5: Experiment 3: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot).

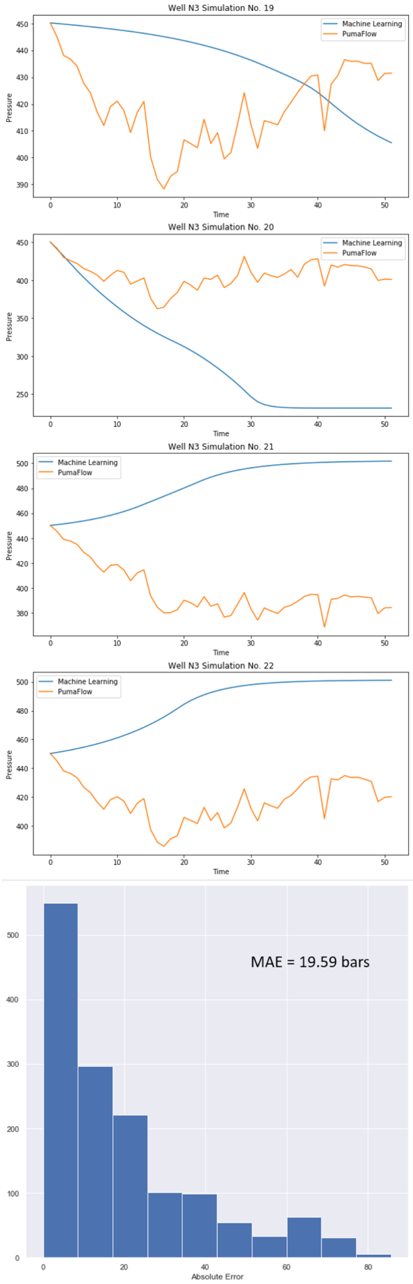

Figure A.6: Experiment 3: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N3) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells.

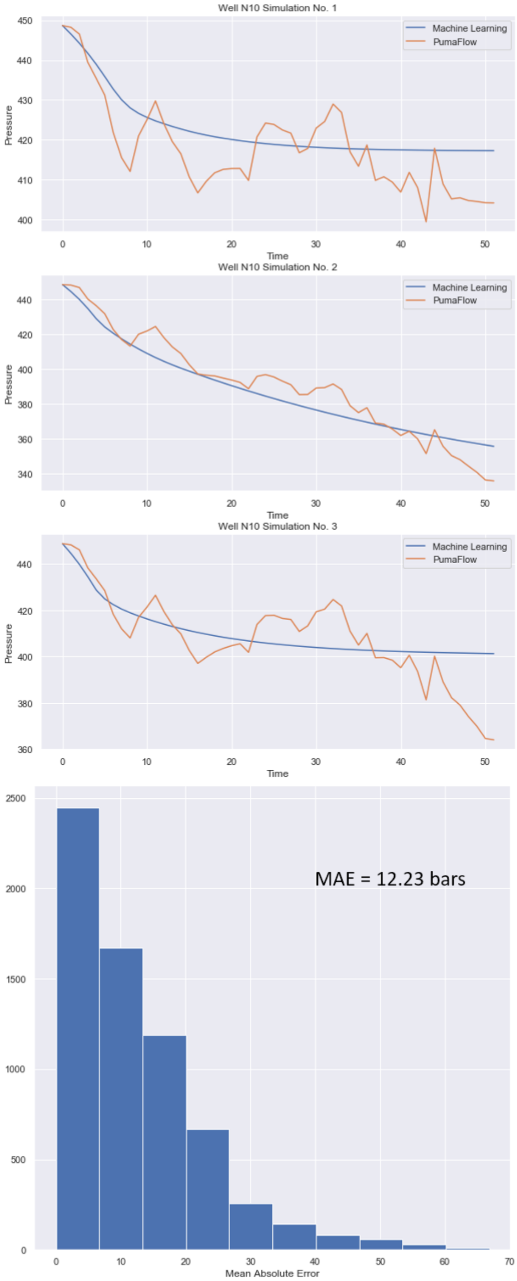

Figure A.7: Experiment 4: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset.

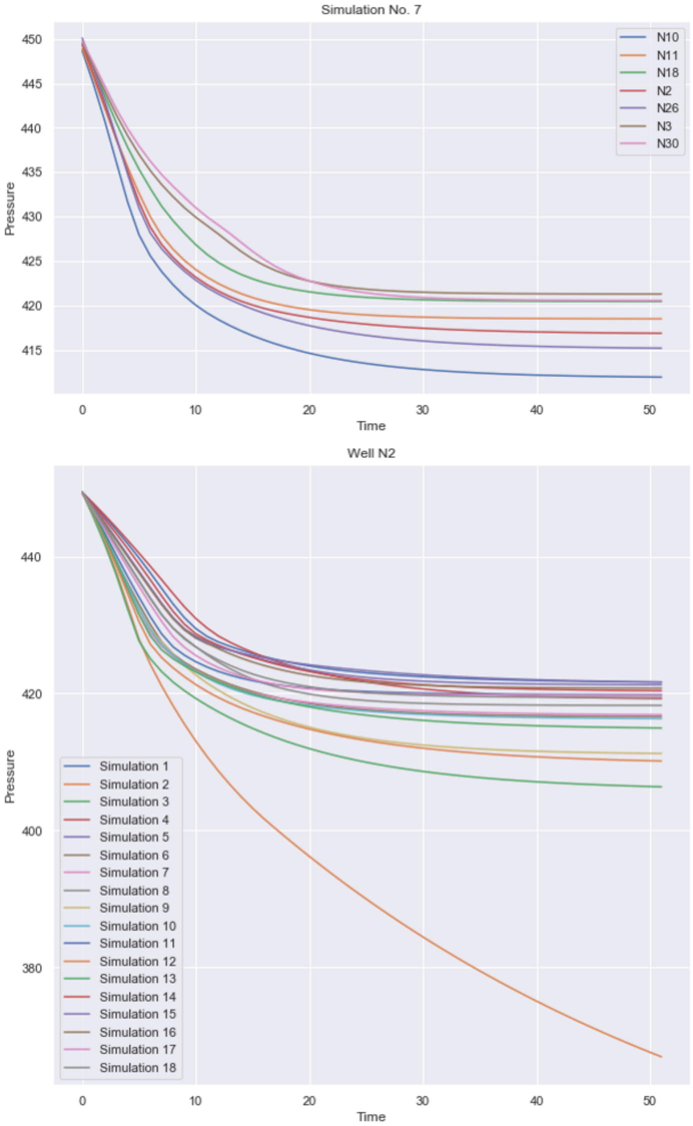

Figure A.8: Experiment 4: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot).

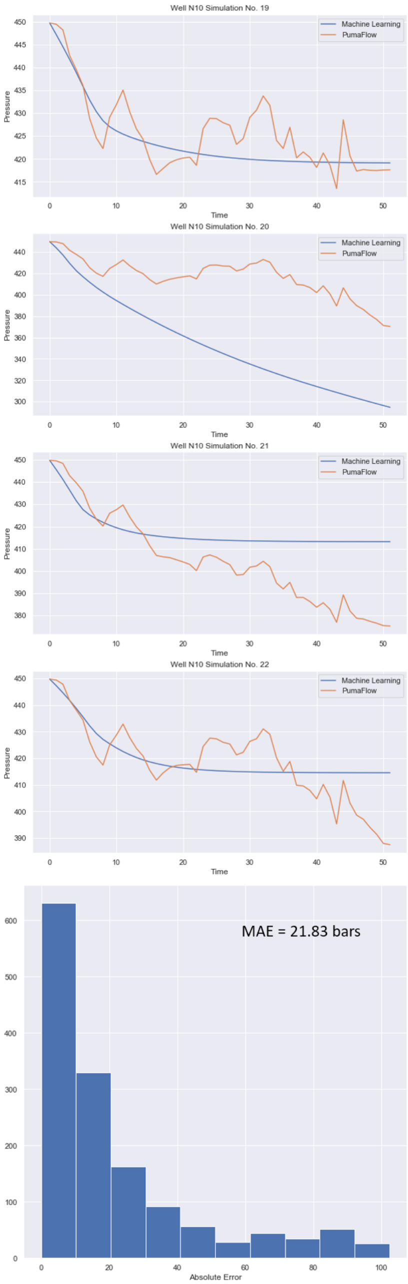

Figure A.9: Experiment 4: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells.

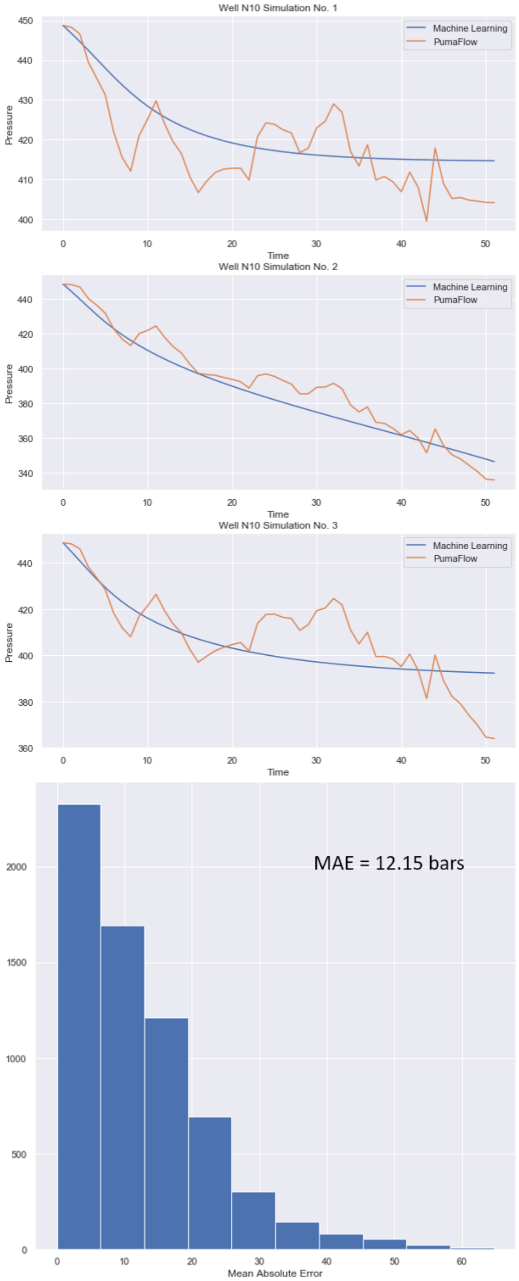

Figure A.10: Experiment 5: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset.

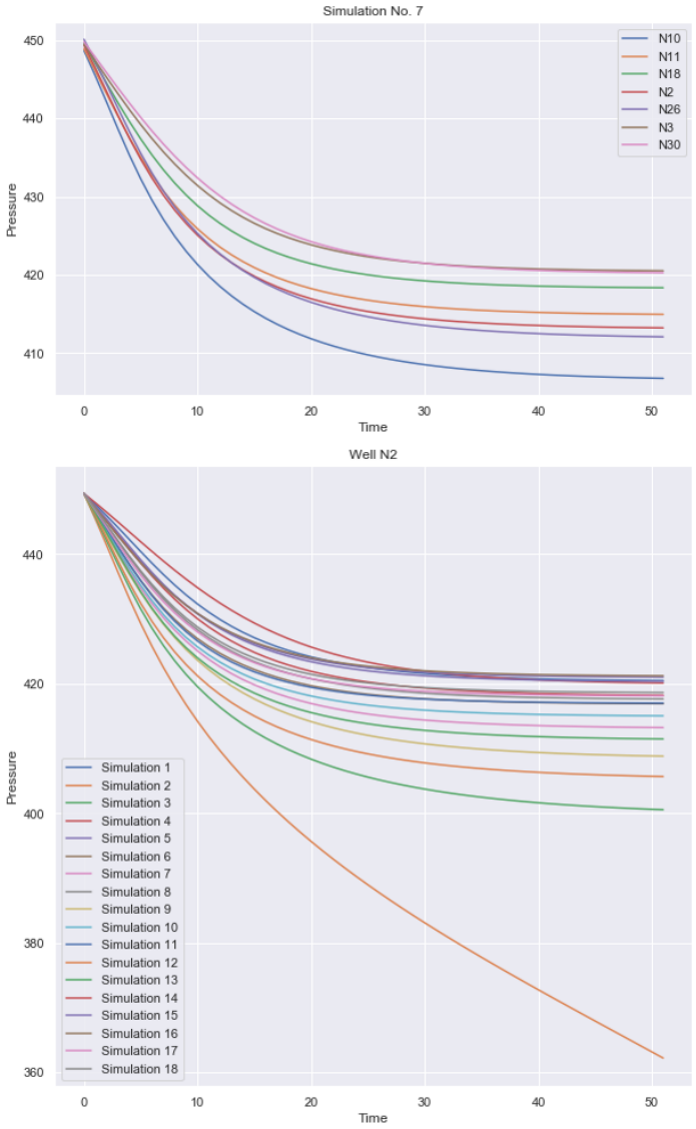

Figure A.11: Experiment 5: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot).

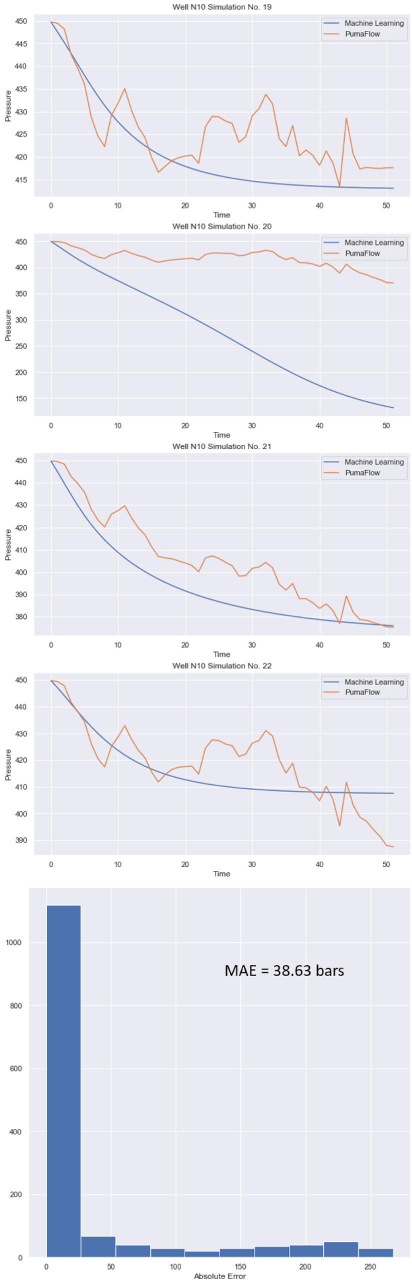

Figure A.12: Experiment 5: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells.

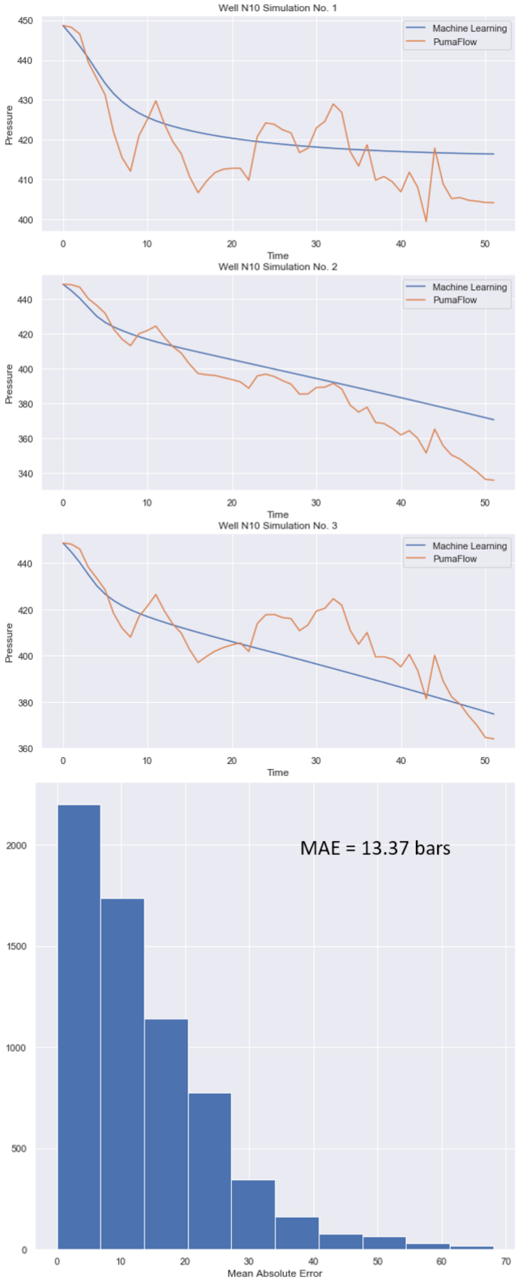

Figure A.13: Experiment 6: This figure consists of plots to effectively understand ML-based simulation accuracy. The top three plots compare the ML-based simulation and PumaFlow^{TM} physics-based simulations at sample wells and simulations from the training dataset. The bottom plot shows the error distribution while comparing the ML-based simulation to physics-based simulation at all wells and simulation from training dataset.

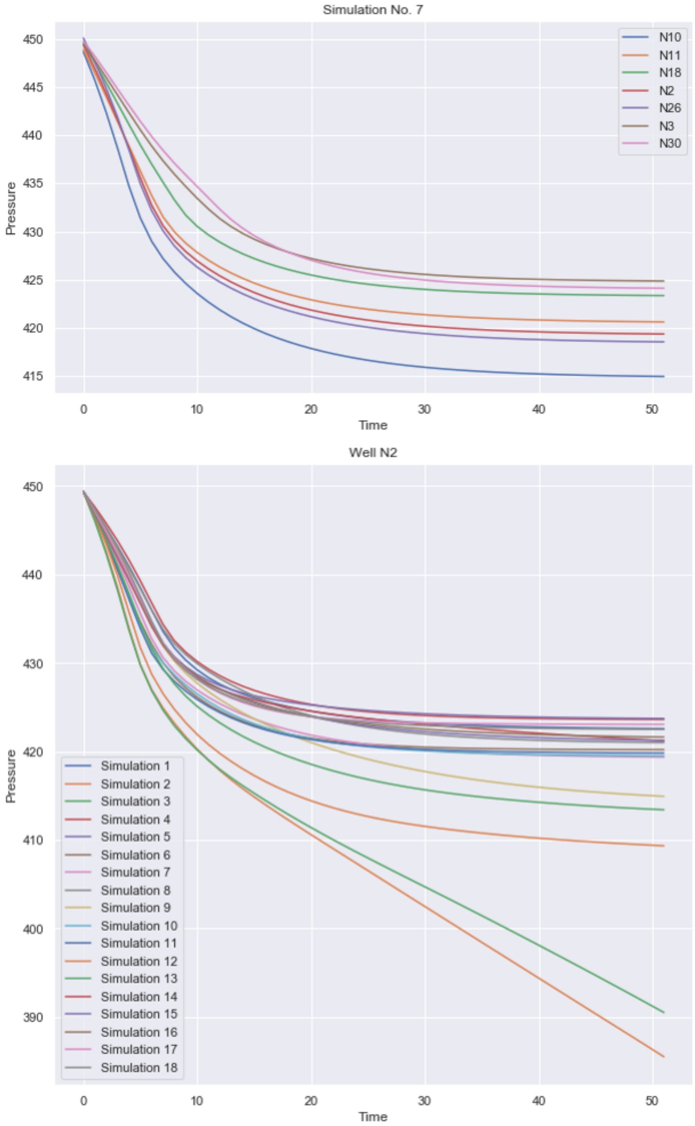

Figure A.14: Experiment 6: This figure shows pressure profiles generated by ML-based simulation at sample well, for all training simulations (bottom plot) and sample simulation, for all wells (top plot).

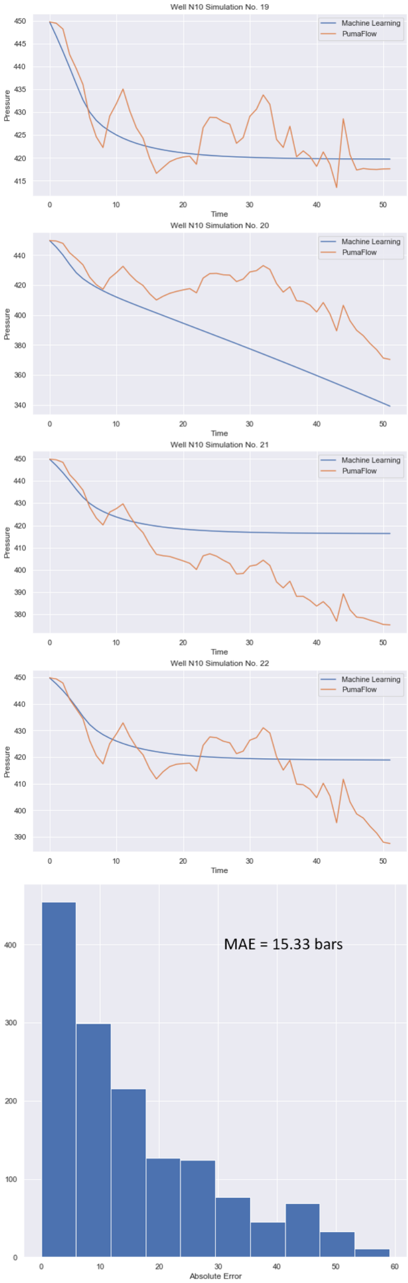

Figure A.15: Experiment 6: This figure consists of plots to understand the ML model's ability to accurately simulate pressure for test cases. The top four plots compare the ML-based simulations and PumaFlow^{TM} physics-based simulations at a sample well (N10) for all (four) test cases. The bottom plot shows the error distribution of the same for all wells.