2014

On the optimization of CO2-EOR process using surrogate reservoir model

Vida Gholami

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

Recommended Citation
https://researchrepository.wvu.edu/etd/5668

This Dissertation is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Graduate Theses, Dissertations, and Problem Reports by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.
ON THE OPTIMIZATION OF CO$_2$-EOR PROCESS USING SURROGATE RESERVOIR MODEL

Vida Gholami

DISSERTATION SUBMITTED
TO THE STATLER COLLEGE OF ENGINEERING AND MINERAL RESOURCES
AT WEST VIRGINIA UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN
PETROLEUM AND NATURAL GAS ENGINEERING

Approved by
Shahab D. Mohaghegh, PhD., Chair
Sam Ameri, M.S.
Jitendra Kikani, PhD.
Marcello Napolitano, PhD.
Ebrahim Fathi, PhD.

DEPARTMENT OF PETROLEUM AND NATURAL GAS ENGINEERING

Morgantown, West Virginia
2014

Keywords: Reservoir Simulation, Reservoir Modeling, CO$_2$ Sequestration, CO$_2$-EOR, Miscible CO$_2$-EOR, SACROC Field, Optimization, Artificial Intelligence, Proxy Modeling, Reduced Order Modeling, Surrogate Reservoir Modeling, Coupled Surrogate Reservoir Modeling, Grid-Based Surrogate Reservoir Modeling, Well-Based Surrogate Reservoir Modeling

© Copyright 2014, Vida Gholami
ABSTRACT

ON THE OPTIMIZATION OF CO₂-EOR PROCESS USING SURROGATE RESERVOIR MODEL

VIDA GHOLOMI

CO₂-EOR projects are becoming increasingly popular. Since enhanced recovery processes are applied to the mature fields, it usually involves a large number of wells. While the large number of wells leads to a better geological model, it results in very large flow models that are hard to manage, history match, and use as an optimization base. Nevertheless, injection-production optimization remains the core of all modeling efforts in CO₂-EOR projects.

The objective of this work is to investigate the feasibility of using state-of-the-art data-driven proxy models to facilitate injection-production optimization in a CO₂-EOR process. The use of coupled grid-based – SRMg and well-based – SRMw Surrogate Reservoir Model (as a proxy that runs in seconds) will be investigated as a tool to achieve the objectives of this study. The coupled SRM is built based on a reservoir simulation model that is developed for this purpose. The coupled SRM will be able to identify the dynamic reservoir properties (pressure, saturations, and component mole fraction at gridblock level) throughout the reservoir, along with the production characteristics at each well. It can be used to identify the optimum injection strategy (volume, rate, etc.) that would result in increased oil production.

The EOR technique that is attracting the most new market interest is CO₂-EOR. First tried in 1972 in Scurry County, Texas, CO₂ injection has been used successfully throughout the Permian Basin of West Texas and eastern New Mexico. The SACROC field, a depleted oil field located in western Scurry County in Texas, is the subject of this study.

A high resolution geological model was built for the northern platform. The model is based on a comprehensive geological study including 3D seismic survey and well logs. The porosity and permeability data for the fine grids were obtained from the Bureau of Economic Geology (BEG). The very long run-time of the reservoir simulation model that is the result of complexity of the reservoir makes it impractical to perform any sensitivity analysis, uncertainty analysis, or optimization study on the model. In order to overcome this problem, developing a surrogate reservoir model based on Artificial Intelligence and Data Mining techniques was planned. The coupled SRM provides the means for performing a large number of simulation runs, in short period of time, to be used for uncertainty quantification, and search of solution space for optimization.

Multiple injection scenarios were designed and run using a numerical reservoir simulator. The results were used in order to build a comprehensive spatio-temporal dataset, which includes all aspects of the reservoir model that is needed to train, calibrate, and validate the coupled SRM. From the parameters assimilated to form the comprehensive spatio-temporal dataset, Key Performance Indicators were identified and
ranked. These KPIs helped to determine the dimensionality of the input space used to develop the SRMs (SRM\textsubscript{W} and SRM\textsubscript{G}).

Optimization may be identified by two focus areas. Building an efficient evaluation function and finding the quickest path to global minima. In this work, we focus on the efficiency of the evaluation function. The integrated SRM was built by coupling the two aforementioned SRMs. This SRM can be used to identify the optimal injection strategy (volume, rate, etc.) that would result in increased oil production while keeping an eye on the flood front.
To my parents
ACKNOWLEDGEMENTS

First and foremost, I would like to gratefully and sincerely thank my advisor Dr. Shahab Mohaghegh for his guidance, wisdom, and, most importantly, his friendship during my graduate studies at West Virginia University. His mentorship was paramount in providing experiences and insights essential to my long-term career goals. He encouraged me to not only grow as an engineer, but also as an instructor and an independent thinker.

My thanks also go to my committee members, Professor Sam Ameri, Dr. Jitendra Kikani, Dr. Ebrahim Fathi, and Dr. Marcello Napolitano for their constructive comments and suggestions during the course of this work. I also would like to thank Dr. Grant Bromhal for his useful and valuable remarks.

I gratefully acknowledge the Computer Modeling Group, Schlumberger and Intelligent Solutions, Inc., for providing CMG, Petrel, and Data-Driven Modeling software suits used during different stages of this study.

My gratitude goes to Dr. Razi Gaskari and Mohammad Maysami at Intelligent Solutions, Inc. for their technical comments, insights, and help with the surrogate reservoir modeling aspects of this work.

I would also like to thank all members of the PEARL group for providing a lively environment, especially my officemates during the past couple of years, Yasaman Khazaeni, Saeed Zargari, Alireza Shahkarami, Alireza Haghighat, and Shohreh Amini.
I obviously could not make it without the support of all my friends whom I met in Morgantown and made my time at WVU memorable and joyful. My heartfelt thanks go to my dearest friends who became like my second family in Morgantown. Yasaman Khazaeni, Mohammad Maysami, Hossein Radmard, Maryam NaghshNejad, and Sheida Teimouri, who were there for me in happy and sad moments, good and bad times, even after each of us pursued our own journey of life.

Most importantly, I would like to give a wholehearted, deep thanks to my family, especially my mom and dad for all their love, encouragement, and sacrifices. Although they are miles away, they never let me down by their unconditional love and support in every stage of my personal and academic life.
# Table of Contents

Abstract ........................................................................................................................................ ii

Acknowledgements ...................................................................................................................... v

List of Tables ............................................................................................................................... x

List of Figures ............................................................................................................................... xii

CHAPTER 1.  Introduction ............................................................................................................. 1

  1.1.  Problem Statement ........................................................................................................... 1

  1.2.  Dissertation Outline ........................................................................................................ 3

CHAPTER 2.  Why Surrogate Models? ......................................................................................... 5

  2.1.  General Review of Proxy Models ................................................................................... 6

  2.1.1.  Response Surface Methodology (RSM) ................................................................... 6

  2.1.2.  Reduced Physics Models .......................................................................................... 5

  2.1.3.  Reduced Order Models (ROMs) ............................................................................... 14

  2.1.4.  Data Driven Modeling (DDM) ................................................................................ 32

  2.2.  Proxy Modeling in Literature ......................................................................................... 32

  2.3.  Surrogate Reservoir Model ............................................................................................ 51

CHAPTER 3.  CO₂ - Storage and EOR Application .................................................................... 54

  3.1.  Geologic Carbon Sequestration ..................................................................................... 54
3.1.1. Capture ........................................... 55
3.1.2. Transportation ...................................... 56
3.1.3. Storage ............................................. 57
3.2. Enhanced Oil Recovery (EOR) ......................... 62
  CO$_2$-EOR ............................................. 63
3.3. CO$_2$ EOR-Sequestration ............................... 74
  3.3.1. Potential Benefits .................................. 74
  3.3.2. Uncertainties in CO$_2$-EOR and Sequestration Flood Design Parameters 76
CHAPTER 4. Reservoir Modeling .................................. 77
  4.1. Field Background ...................................... 77
    4.1.1. Reservoir ......................................... 79
    4.1.2. Geologic Description ............................... 81
    4.1.3. Performance History ................................. 84
  4.2. Analyses of Field Data for Model Preparation ....... 89
    Quantification of Natural Heterogeneity in SACROC Northern Platform ....... 89
  4.3. Geo-Cellular Model .................................... 94
  4.4. Upscaling ............................................. 96
  4.5. Numerical Reservoir Model development ............. 101
    4.5.1. Initial and Boundary Condition .................... 101
    4.5.2. Relative Permeability and Capillary Pressure ........ 103
    4.5.3. Wells- Constraints and Completions ............... 104
  4.6. Results ............................................. 111
  4.7. Capturing the Physics .................................. 115
    Miscible flooding ....................................... 115
CHAPTER 5. SRM Development .................................. 132
  5.1. Design of Experiments .................................. 134
WAG Scenarios .................................................................................................................. 135

5.2. Grid-Based SRM (SRM\textsubscript{G}) ........................................................................... 142

  5.2.1. SRM\textsubscript{G} Spatio-Temporal Dataset Generation .......................................... 142

  5.2.2. Key Performance Indicator (KPI) Analysis and Input Selection .......................... 146

  5.2.3. Data partitioning ..................................................................................................... 151

  5.2.4. Network Design and Training ................................................................................. 151

  5.2.5. SRM\textsubscript{G} Deployment Results and Robustness Analysis .............................. 155

5.3. Well-Based SRM (SRM\textsubscript{W}) ............................................................................ 167

  5.3.1. SRM\textsubscript{W} Spatio-Temporal Dataset Generation .......................................... 167

  5.3.2. Key Performance Indicator (KPI) Analysis and Input Selection .......................... 170

  5.3.3. Data partitioning ..................................................................................................... 173

  5.3.4. Network Design and Training ................................................................................. 173

  5.3.5. SRM\textsubscript{W} Deployment Results and Robustness Analysis .............................. 175

5.4. Coupled Surrogate Reservoir Model ......................................................................... 184

CHAPTER 6. Conclusion and Discussion ............................................................................ 202

  6.1. Lessons Learned ....................................................................................................... 205

    Time Resolution ............................................................................................................ 205

    Sampling methods ........................................................................................................ 206

    Similar Governing Physics .......................................................................................... 207

    Speed improvement ...................................................................................................... 207

    Error Propagation ........................................................................................................ 208

    Input Data Selection ..................................................................................................... 208

  6.2. SRM vs Other Proxy Modeling Techniques ............................................................... 209

References .......................................................................................................................... 215

Appendix – Cascading and Coupling Results .................................................................... 233
LIST OF TABLES

Table 1. Major U.S. CO₂ Operators (OGJ Biennial EOR Survey 2008) [116] .................. 68
Table 2. The major injection schemes developed in the last 30 years............................ 73
Table 3. CO₂ Utilization and Potential in EOR Projects (Source of U.S. data: National Energy Technology Laboratory, Carbon Sequestration Atlas, 2007) ......................... 75
Table 4. Basic Reservoir Data- SACROC [133] .................................................................. 80
Table 5. SACROC Reservoir Fluid Properties [133] ........................................................ 80
Table 7. Initial oil composition in the upscaled geocellular model [133] ....................... 102
Table 8. Solubility data for SACROC fluid ................................................................. 102
Table 9. Numerical Control Values set in the simulator ............................................... 112
Table 10. Compositional Analyses of Crude Oils Tested by GC Weight, % [177] ...... 125
Table 11. The primary and injection fluid composition after splitting the plus fraction ........................................................................................................ 128
Table 12. Scenarios selected for building SRM based on Design of Experiments .... 137
Table 13. Information required for developing Grid-Based SRM ................................. 143
Table 14. Geological zones used for lumping and their corresponding simulation layers. ........................................................................................................ 143
Table 15. A sample of selected inputs for predicting Pressure ........................................ 149
Table 16. R-squared coefficients of all the neural networks built for Grid Based SRM. ........................................................................................................ 153
Table 17. Blind scenario for verifying the SRM .............................................................. 161
Table 18. Information required for developing Well-Based SRM ................................. 168
Table 19. A sample of selected inputs for predicting oil production rate. .................. 172
Table 20. R-squared coefficients of all the neural networks built for grid-based SRM. 174
Table 21. The properties of some of the proxy models in literature. ......................... 210
LIST OF FIGURES

Figure 1. A biological neuron (Left) and a general artificial neuron (Right) [41]........ 34
Figure 2. Notation for the three-layered network [44]........................................... 38
Figure 3. The three components involved in the System Theory, Input, System and Output................................................................. 53
Figure 4. Effect of gravity during WAG injection [99]............................................. 60
Figure 5. Two-phase relative permeability diagram................................................. 61
Figure 6. Phase diagram of CO₂ [111]...................................................................... 64
Figure 7. Variation of CO₂ density as a function of temperature and pressure [110]..... 65
Figure 8. Variation of CO₂ viscosity as a function of temperature and pressure [110]... 65
Figure 9. Yearly evolution of crude oil production (b/d) by CO₂ EOR in the USA [116].................................................................................................................. 68
Figure 10. Cross-section illustrating how carbon dioxide and water can be used to flush residual oil from a subsurface rock formation between wells [116]......................... 71
Figure 11. The Permian Basin outlined in green covers parts of western Texas and southeastern New Mexico. SACROC oilfield, identified in red, resides on the northeastern edge of the Basin. The map also identifies both natural and anthropogenic sources of CO₂ as well as the CO₂ pipelines in the region [134]................................. 78
Figure 12. SACROC Unit at the Horseshoe Atoll in west Texas and structural contours map of top of carbonate reef modified from Stafford. Contours are in m scale [138]. .............................................................................................................. 81
Figure 13. A structural and stratigraphic cross-section of profile A-A', located within the SACROC northern platform [137]. See Figure 12 for a location of profile A-A'.

Figure 14. Well logs representing the SACROC northern platform and summary of previous studies estimating carbonate rock properties in Cisco and Canyon Groups [142].

Figure 15. Contour and dotted lines, respectively, indicate the bottom hole pressure (psi) and the location of center-line water injection wells [133](a) Five months before starting water injection (September 1954). Only 1% of reservoir volume is above bubble point pressure (1805 psig=12.45 MPa). (b) Seventeen years later after starting water injection (September 1970), 77% of the reservoir volume is above bubble point pressure.

Figure 16. Frontal advance of centerline water injection, SACROC unit, Kelly Snyder field [133].

Figure 17. The steep production decline at the SACROC unit leveled and began to increase due to the successful implementation of miscible CO$_2$ floods. The blue line represents the best-producing field project that is being replicated (3).

Figure 18. SACROC Performance History.

Figure 19. (A) SACROC unit at the Horseshoe Atoll in western Texas and CO$_2$ supply system from natural CO$_2$ reservoirs, (B) Magnified map of the SACROC Unit within the Horseshow Atoll with indication of paleo-wind direction, (C) Well locations of SACROC Unit with the estimated water-flooding fronts at the end of water-flooding period in 1973 (Kane, 1979) [142].

Figure 20. Spatial heterogeneity of porosity determined by wireline log [138].

Figure 21. Spatial heterogeneity of porosity determined by both wire-line log and seismic data [138].

Figure 22. Permeability can be estimated from wireline logs using the global transformation and rock fabric number [158].

Figure 23. Porosity-Permeability correlation in SACROC Northern platform based on the core data [158].

Figure 24. The locations of wells in SACROC fields and northern platform top map based on available data.

Figure 25. The area of 3-D model of SACROC-Northern platform [158].

Figure 26. Top, Porosity and Permeability distribution in SACROC-Northern platform (Axis Scale is in m).
Figure 27. SACROC Northern Platform High Resolution Geological Model - Left: Porosity Distribution Histogram - Right: Permeability Distribution Histogram.

Figure 28. SACROC High Resolution Geological Model - Section for This Study (Axis Scale is in m) - Left: Porosity Distribution Map and Histogram - Right: Permeability Distribution Map and Histogram.

Figure 29. A cross section showing the vertical resolution in the fine scale geo-cellular model (Northern platform-Top Section).

Figure 30. Geology zones in SACROC field-Northern platform.

Figure 31. Defining geological zones and corresponding horizons and simulation layers.

Figure 32. SACROC Low Resolution Geological Model - Section for This Study (100 * 110 * 36 Grids) (Axis Scale is in m) - Left: Porosity Distribution Map and Histogram - Right: Permeability Distribution Map and Histogram.

Figure 33. (a) Fitted relative permeability curves with experimental data from Bennion and Bachu [165] (b) Fitted capillary curve with experimental data from Bennion and Bachu [172].

Figure 34. (a) Fitted relative permeability curves between brine and oil; measurements from Rohan and Haggerty [170]. (b) Fitted capillary pressure curve between brine and oil; measurements from Rohan and Haggerty [170] [138].

Figure 35. Final relative permeability and capillary pressure curves used for numerical modeling.

Figure 36. 3-D Reservoir Simulation Model Showing the Reservoir Properties (Top, Porosity, and Permeability) and well location.

Figure 37. Production/Injection well locations and the extended area added to mimic the open flow boundary.

Figure 38. Oil Production History in 130 Wells of the SACROC-Northern Platform from 1948 to 2003.

Figure 39. Water Injection History in 50 Wells of the SACROC-Northern Platform from 1948 to 2003.

Figure 40. CO₂ Injection History in 46 Wells of the SACROC-Northern Platform from 1948 to 2003.

Figure 41. Oil, water and gas production from the wells located in the SACROC section under study (Left), and water/gas injection and average pressure reservoir in this section (Right).
Figure 42. Change in oil saturation, water saturation and pressure in layer 1 during the life of reservoir.

Figure 43. Change in oil saturation, water saturation and pressure in layer 36 during the life of reservoir.

Figure 44. Mole fraction of CO$_2$ in water, oil and global distribution before (1974) and at the end of gas injection history (2004) in the first layer.

Figure 45. Mole fraction of CO$_2$ in water, oil and global distribution before (1974) and at the end of gas injection history (2004) in the last layer.

Figure 46. One-dimensional schematic showing how CO$_2$ becomes miscible with crude oil [99].

Figure 47. Typical Equilibrium of CO$_2$-Crude Oil System [177].

Figure 48. P-X diagram used to estimate first contact miscibility pressure [179].

Figure 49. P-X diagram based on the injection gas type [179].

Figure 50. ternary diagram [181].

Figure 51. Tie line in ternary diagram [181].

Figure 52. Ternary diagram construction [179].

Figure 53. P-X diagram in condensing drive [179].

Figure 54. P-X diagram in vaporizing drive [179].

Figure 55. The condensing/vaporizing process characterized by a two-phase pseudo-ternary envelope [179].

Figure 56. Slim tube simulation results show the increase in RF up to MMP.

Figure 57. Slim tube simulation results estimate a MMP greater than 2500 psi.

Figure 58. Pseudo Ternary plots at 1500 psi and 2840 psi (MMP).

Figure 59. Variation of gas mole fraction in two gridblocks.

Figure 60. Stripping of lighter fractions during the time in two GBs.

Figure 61. Higher concentration of lighter fraction in the front due to miscible CO$_2$ injection.

Figure 62. Concentration of CO$_2$ at different times corresponding to Figure 61.

Figure 63. Grid based / well-based development, General methodology flowchart.

Figure 64. A schematic of the injection cycles in the sampled (training) scenarios (Yellow: Gas, Blue: Water).

Figure 65. The BHP range for all the wells in all the training scenarios.

Figure 66. Range of WAG Ratio in training scenarios (Top-Left), Range of gas slug volume in the field in each half cycle for training scenarios (Top-Right), The WAG
ratio distribution with respect to water and gas injection rate (Bottom-Left), The relationship between injected water and gas slug volumes in each cycle (Bottom-Right).

Figure 67. Oil production rate and cumulative oil production in the training scenarios. ................................................................. 139

Figure 68. Cumulative gas and water injection in training scenarios. ................................................................. 140

Figure 69. Average reservoir pressure, field oil production rate, and field gas and water injection rate in WAG scenario 2 (Left) and scenario 7 (Right). ................................................................. 141

Figure 70. HCPV injected gas and water for Scenarios 2 and 7. ................................................................. 141

Figure 71. Structure Top (ft), Thickness (ft) (top), Porosity and Permeability (md) (bottom) distribution in the top lumped layer. ................................................................. 144

Figure 72. From left to right, Tier 1 to 8. Green block is the main blocks, Orange blocks show the ones used for tiering. ................................................................. 145

Figure 73. Sampling methods comparison. ................................................................. 146

Figure 74. Sample KPI analysis for Pressure. ................................................................. 147

Figure 75. The neural network architecture for Pressure modeling at 1/1/2009. ................................................................. 151

Figure 76. The results of training, calibration, verification and the frequency of error for pressure network at 1/1/2009. ................................................................. 152

Figure 77. The parameters having the most influence on Pressure at 1/1/2009 in the training process. ................................................................. 154

Figure 78. Grid based cascading flow-chart. ................................................................. 156

Figure 79. Global mole fraction of CO$_2$ distribution map and error histogram at the end of the first WAG cycle (2006) - SRM$_G$ deployment results on scenario 5. ................................................................. 157

Figure 80. Global mole fraction of CO$_2$ distribution map and error histogram at the end of the third WAG cycle (2012) - SRM$_G$ deployment results on scenario 5. ................................................................. 157

Figure 81. Pressure distribution map and error histogram at the end of the first WAG cycle (2006) - SRM$_G$ deployment results on scenario 5. ................................................................. 158

Figure 82. Pressure distribution map and error histogram at the end of the third WAG cycle (2012) - SRM$_G$ deployment results on scenario 5. ................................................................. 158

Figure 83. Oil saturation distribution map and error histogram at the end of the first WAG cycle (2006) - SRM$_G$ deployment results on scenario 5. ................................................................. 159

Figure 84. Oil saturation distribution map and error histogram at the end of the third WAG cycle (2012) - SRM$_G$ deployment results on scenario 5. ................................................................. 159
Figure 85. Water saturation distribution map and error histogram at the end of the first WAG cycle (2006) - SRM deployment results on scenario 5......................... 160

Figure 86. Water saturation distribution map and error histogram at the end of the third WAG cycle (2012) - SRMG deployment results on scenario 5.................. 160

Figure 87. Schematic of injection design in blind set........................................... 161

Figure 88. The BHP range for all the wells the blind scenario.................................. 161

Figure 89. Global mole fraction of CO$_2$ distribution map and error histogram at the end of the first WAG cycle (2006) - SRM deployment results on the blind scenario...... 162

Figure 90. Global mole fraction of CO$_2$ distribution map and error histogram at the end of the third WAG cycle (2011) - SRM deployment results on the blind scenario .... 162

Figure 91. Pressure distribution map and error histogram at the end of the first WAG cycle (2006) - SRM deployment results on the blind scenario ......................... 163

Figure 92. Pressure distribution map and error histogram at the end of the third WAG cycle (2011) - SRM deployment results on the blind scenario ......................... 163

Figure 93. Oil saturation distribution map and error histogram at the end of the first WAG cycle (2006) - SRM deployment results on the blind scenario.................... 164

Figure 94. Oil saturation distribution map and error histogram at the end of the third WAG cycle (2011) - SRMG deployment results on the blind scenario..................... 164

Figure 95. Water saturation distribution map and error histogram at the end of the first WAG cycle (2006) - SRM deployment results on the blind scenario.................... 165

Figure 96. Water saturation distribution map and error histogram at the end of the third WAG cycle (2011) - SRMG deployment results on the blind scenario.............. 165

Figure 97. Tier 1 and 2 location w.r.t well block; in one layer (Left) and Upscaled format (Right).......................................................................................... 169

Figure 98. The Voronoi cells overlaid on Top map in Layer 1 (Left) and Tier map (Right), White cells represent the injectors, Dark Blue: Tier 1, Green : Tier 2, Red : Tier 3........................................................................................................ 169

Figure 99. Samples of basic statistical analysis showing the correlation between the oil production rate and different parameters................................................... 170

Figure 100. A part of KPI analysis result with oil production rate at time step (t) as output........................................................................................................ 171

Figure 101. The results of training, calibration, verification and the frequency of error for oil production rate........................................................................ 174
Figure 102. The parameters having the most influence on Oil Production in the training process. ............................................................... 175
Figure 103. The wells corresponding to samples shown in ............................................................... 176
Figure 104. The SRM$^W$ deployment results in Scenario 5 - Oil production in sample wells and entire field. ............................................................... 177
Figure 105. The SRM$^W$ deployment results in Scenario 5 - CO$_2$ production of sample wells and entire field. ............................................................... 178
Figure 106. The SRM$^W$ deployment results in Scenario 5 - Water production of sample wells and entire field. ............................................................... 179
Figure 107. The SRM$^W$ deployment results in blind scenario - Oil production in sample wells and entire field. ............................................................... 181
Figure 108. The SRM$^W$ deployment results in blind scenario - CO$_2$ production of sample wells and entire field. ............................................................... 182
Figure 109. The SRM$^W$ deployment results in blind scenario - Water production of sample wells and entire field. ............................................................... 183
Figure 110. The coupled SRM flowchart. ...................................................................................... 185
Figure 111. Global mole fraction of CO$_2$ distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on scenario 5. ...................................................................................... 187
Figure 112. Global mole fraction of CO$_2$ distribution map and error histogram at the end of the third WAG cycle (2012) – Coupled SRM deployment results on scenario 5. ...................................................................................... 187
Figure 113. Pressure distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on scenario 5 ............................................................... 188
Figure 114. Pressure distribution map and error histogram at the end of the third WAG cycle (2012) – Coupled SRM deployment results on scenario 5 ............................................................... 188
Figure 115. Oil saturation distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on scenario 5 ............................................................... 189
Figure 116. Oil saturation distribution map and error histogram at the end of the third WAG cycle (2012) – Coupled SRM deployment results on scenario 5 ............................................................... 189
Figure 117. Water saturation distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on scenario 5 ............................................................... 190
Figure 118. Water saturation distribution map and error histogram at the end of the third WAG cycle (2012) – Coupled SRM deployment results on scenario 5.......... 190
Figure 119. The Coupled SRM deployment results in Scenario 5 - Oil production in sample wells and entire field. .......................................................... 191
Figure 120. The Coupled SRM deployment results in Scenario 5 - CO₂ production in sample wells and entire field. .......................................................... 192
Figure 121. The Coupled SRM deployment results in Scenario 5 - Water production in sample wells and entire field. .......................................................... 193
Figure 122. Global mole fraction of CO₂ distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on the blind scenario .................................................................................. 195
Figure 123. Global mole fraction of CO₂ distribution map and error histogram at the end of the third WAG cycle (2011) – Coupled SRM deployment results on the blind scenario .................................................................................. 195
Figure 124. Pressure distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on the blind scenario .......... 196
Figure 125. Pressure distribution map and error histogram at the end of the third WAG cycle (2011) – Coupled SRM deployment results on the blind scenario .......... 196
Figure 126. Oil saturation distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on the blind scenario .......... 197
Figure 127. Oil saturation distribution map and error histogram at the end of the third WAG cycle (2011) – Coupled SRM deployment results on the blind scenario .......... 197
Figure 128. Water saturation distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on the blind scenario .... 198
Figure 129. Water saturation distribution map and error histogram at the end of the third WAG cycle (2011) – Coupled SRM deployment results on the blind scenario. . 198
Figure 130. The SRMₓ deployment results on the blind scenario – Oil production in sample wells and the entire field .......................................................... 199
Figure 131. The SRMₓ deployment results on the blind scenario – CO₂ production in sample wells and the entire field .......................................................... 200
Figure 132. The SRMₓ deployment results on the blind scenario - Water production in sample wells and the entire field .......................................................... 201
Figure 133. Recommended time resolution for the first three years compared with the resolution used in this study.
CHAPTER 1. INTRODUCTION

1.1. PROBLEM STATEMENT

Decision making is the basis of any business model. Intelligent decisions based on well
described optimal solutions to real life problems are crucial for a successful business
model in the world of tight competition and scarcity. Obviously being one of the
world’s largest industries, there is no exception to this rule in the oil industry. In any
stage of reservoir development, the ultimate goal of the managing/engineering teams is
to develop the most accurate or “optimal” decisions.

Starting from very initial exploration to late development of the field, every stage
involves important decisions that shall outline the probability of success of the project.
It is not always easy to reach intuitive optimal decisions, as problems at hand are
normally too complex. Also, the fact that decision surface may be steep yields to
situations where slightly better decisions result in remarkably better results. As an
example, consider a problem where a new production well is to be drilled in a mature,
fluvial depositional oil reservoir. The reservoir is made of high porosity sand channels
laid in non-permeable mud formations. This is a common reservoir type, especially in
North American fields. In this case, it is of vital importance to place the well in the
correct position, as off shooting the channel will yield a non-producing well. In such a
case, there is no question that engineers will want to base their decision on numerical
models. Eventually, they would like to study all possible scenarios or consider another
case, in which tertiary recovery is an option. Different properties such as injection
design, well placement, etc. will impact the effectiveness of injection and oil recovery.
Unfortunately, numerical models that are used in the oil industry based on partial differential equations (PDEs), are computationally intensive, even with today's supercomputers. The PDEs in these applications depend on a set of parameters. These parameters are bound in the range of a parameter space. “Parameter space” refers to a hypothetical space where a “location” is defined by the values of all parameters. As used herein, the parameter space may be designated as a collection of all the parameters, in consort with the ranges of values that the parameters can take, considered at any stage of history matching, uncertainty or sensitivity analysis, and optimization. The dimension of the parameter space depends on the number of independent parameters. Each point in this space belongs to a particular model. This space is dealt in different ways depending on the engineering task at hand, i.e. a subgroup of the parameter space is searched for in order to match model responses with observations or other specific functions (parameter identification or history matching), the model responses are evaluated over the entire parameter space (uncertainty quantification), or multiple model realizations are used in order to determine parameters in the parameter space that provide optimal model responses with respect to predefined functions (optimization). In the process of field developments or oil and gas production, these applications are highly useful. The models used for simulating these processes are built based upon many geologic or engineering parameters. Some of these parameters are uncertain [1].

To reach an optimal decision in the aforementioned situations, numerous approaches could be applied [2]:

1. Tackle the bottleneck of exorbitant numerical calculations, i.e. try to run less simulation
2. Develop less costly algorithms (numerically or analytically)
3. Increase the processing power

Exhaustive and comprehensive evaluation of the solution space for designing field development strategies, as well as quantification of uncertainties associated with the static model, are the type of analyses that require a large number of simulation runs in order to provide meaningful and practical results. When a numerical simulation model takes hours for a single run, performing such analyses becomes unfeasible and the engineers have to compromise by designing and running a much smaller number of runs in order to make decisions.
Decreasing the number of simulations will produce less informed solution space, which may increase the probability of missing the global optimum. Special care should be taken to avoid local extremes and this may not be trivial. Developing less costly algorithms is not feasible, as it will decrease the accuracy of the solution and provide more assumptions, which may not be true for different cases. After all, we would like to capture the most realistic physical behavior of the field. Increasing the processing power is a viable option, but even with today’s supercomputers, it is not possible to run hundreds of simulations with a high number of gridblocks. The main focus is on developing a framework to reach more informed decisions, and tackling the bottleneck created by doing exhaustive simulations.

The objective of this research is to investigate the feasibility of using state-of-the-art data-driven proxy models to facilitate injection-production optimization in a CO₂-EOR process. This will be achieved by building an integrated SRM by coupling two SRMs. This SRM will be used as a framework to reach more informed decisions and tackling the impasse of doing exhaustive simulations.

1.2. DISSERTATION OUTLINE

General proxy modeling techniques pertinent to this study are reviewed in Chapter 2, which introduces the concept of proxy modeling and different methods used. This is followed by a literature review on proxy modeling approaches in the petroleum industry and their application.

Chapter 3 is on the application of CO₂ injection. It introduces geological carbon sequestration and different criteria important in this process. It then describes the enhanced oil recovery process and CO₂-Enhance Oil Recovery (EOR) in particular. CO₂-EOR can also be used for the sequestration purposes as explained in the rest of the chapter. The potential benefits of this method and the uncertainties related to this process are explained after that.

The SACROC field background is explained in Chapter 4. It includes some information about the geological description and performance history of the field. The analysis of the field data for developing the model and the process for building the high resolution geocellular model is also explained in this chapter. The upscaling procedure and developing the numerical reservoir model is then explained in detail. The results of the
flow model are listed and explained. It is important to assure that the created model is capturing the physics of the phenomenon of interest (miscible EOR). An introduction to miscible flooding and the process for proving this fact and the results make for the rest of the chapter.

Chapter 5 introduces different types of surrogate reservoir models (SRM\(_G\) and SRM\(_W\)) and discusses the appropriate workflow for building these SRMs. Different data handling and mining techniques used in this study are explained in detail. The results of the surrogate models showing the success of this feasibility study are presented and discussed in this chapter.

In chapter 6, the work is summarized and areas that might need further study are outlined. Some recommendations for model improvement are made. The SRM is compared to other proxy modeling techniques used in the petroleum industry and pros and cons of each are listed.
CHAPTER 2. WHY SURROGATE MODELS?

"Habit is habit, and not to be flung out of the window by any man, but coaxed downstairs one step at a time."

― Mark Twain

The traditional methodology for modeling the reservoir behavior for predicting its performance under different conditions is numerical reservoir simulation. With increasing computer capacities, reservoir models have become more complex and consist of an increasing number of variables (typically in the order of $10^4$-$10^6$).

In spite of the great advances in reservoir modeling tools and the advent of high-performance computing (HPC), high-fidelity physics based numerical simulation still remains a challenging step in understanding the physics of the reservoir and the relationship between changes in the model parameter to changes in control inputs for improved recovery efficiency. History matching, optimization, and uncertainty quantification, which are instances of computationally intensive frameworks, become impractical to be performed in a timely manner if real-time data needs to be integrated into the model. Also, the computational time of such large-scale models becomes a bottleneck of fast turnarounds in the decision-making process.

Our objective in this work is to describe the SRM procedure that can be used in the context of computational production optimization. A strategic requirement for any such technique is the ability to provide very fast, yet sufficiently accurate, simulation results.
2.1. General Review of Proxy Models

Within the context of reservoir simulation, a surrogate model is a mathematical model that can approximate the output of a simulator for a set of input parameters or alternate a given physical model. A useful surrogate model should be sufficiently accurate for the problem at hand, while requiring significantly less computational effort than the full-order simulation.

A surrogate can be made using different methods such as data regression, machine learning, reduced order modeling, reduced physics modeling, which includes coarsening of the model in spatial and time dimensions, and lower order discretization methods [1].

In all the proxy modeling methods briefly explained in this section, an approximation or simplification is taking place. Proxy modeling techniques can be divided into different categories based on different criteria. These techniques can be divided into statistical and mathematical models based on their approach and each will have subcategories. Examples of proxy models based on a statistical approach are response surfaces or the methods that are based on correlations/causalities or predetermination of functional form. Reduced order models and upscaling techniques fall within the category of the mathematical proxy models. Current model reduction techniques can be categorized into intrusive and non-intrusive methods depending on whether construction of the reduced order model requires knowledge and modification of the governing equations and numerical implementation of the full-fledged simulator. Intrusive ROMs are not applicable when the source code is inaccessible, which is the case for most commercial simulators. The other problem with most current proxy models is that by increasing the number of inputs the number of trainings needs to increase sharply, as sampling is required to cover the input parameter space. The “curse of dimensionality” challenges the common ROMs used [3].

The other approaches for categorizing the proxy modeling applied in the realm of reservoir simulation are essentially based on three categories of response surface models, reduced-physics models, and reduced-order models [4].

2.1.1. Response Surface Methodology (RSM)

The response surface method is based on the results of a series of reservoir realizations, usually picked based on an experimental design (Design of Experiments or DoE)
Design of experiments refers to techniques for identifying points for sampling variables or input parameters to be used in building a proxy or surrogate modeling system. For instance, engendering a set of equations that characterize the values of parameters at specific points in an uncertainty space. Classical systems and space-filling methods are some examples of this methodology. Factorial designs, space-filling designs, full factorial, D-Optimal design, and Latin hypercube designs and so on, are specific examples of the DoE technique [1].

Some predefined parameterized functions (response surfaces) are fit using the results of the simulations (responses). Some examples of the predefined parameterized functions are Kriging models, least squares, or thin plate splines. The response surface is used in place of the simulator for determining the output using a set of new inputs. Response surface essentially uses a series of mathematical and statistical techniques, which helps to find a functional relationship between a group of inputs or controls \((x_1, x_2, ..., x_k)\) with a response of interest \((y)\). This relationship is generally unknown, but it can be approximated using a low-degree polynomial model of the form

\[
y = f'(x)\beta + \varepsilon
\]

*Equation (1)*

where \(x = (x_1, x_2, ..., x_k)'\), \(f(x)\) is a vector function of \(p\) elements that consists of powers and cross-products of powers of \(x_1, x_2, ..., x_k\) up to a certain degree denoted by \(d (\geq 1)\), \(\beta\) is a vector of \(p\) unknown constant coefficients referred to as parameters, and \(\varepsilon\) is a random experimental error assumed to have a zero mean. This is conditioned on the belief that abovementioned model provides a satisfactory representation of the response. The quantity \(f(x)\beta\) is the expected value of \(y\) and is the mean response which is represented by \(\mu(x)\). A series of \(n\) experiments should first be performed. In each of these experiments, a specified setting of inputs or control points is used to obtain (observe) the response \(y\). These settings all together establish the so-called response surface design, or just design, which can be represented by a matrix, indicated by \(D\), of order \(n \times k\) called the design matrix.
In this matrix, $x_{ui}$ represents the $u$th design setting of $x_i (i = 1, 2, \ldots, k; u = 1, 2, \ldots, n)$. Each row of $D$ denotes a point, referred to as a design point in a $k$-dimensional Euclidean space. The appropriate design is very important for any response surface and the quality of prediction depends on it [5].

For instance, the following equation is an example of the response surface. The response of interest (for example, cumulative oil production at 30 years) is represented by value $y$, while the uncertain parameters that are different in each simulation run are shown by $x$. The indices $i$ and $j$ are the run indices for the $k$ simulations. Different methods, such as least squares optimization, are used to calculate the weights $\beta$ in the response surface created using $k$ simulation [6].

\[
y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_i^2 x_i^2 + \sum_{i=1}^{k} \sum_{j>i} \beta_{ij} x_i x_j
\]

Equation (3)

As mentioned before, there are different response surface generation methods, each having different properties. The following is a brief explanation of some of these methods.

**Least Squares**

Least squares are used with the intention of constructing a representative function composed of simple known functions, such as polynomials, which minimize the sum of the squared residuals between observed/simulated values and the function values. Least square methods are often not data exacts. If the number of prior data equals the number of unknown functional coefficients, then the representative surface generated by least squares interpolation may traverse all the prior data points. However, on the whole, to generate a smooth surface, the number of coefficients is chosen to be smaller than the number of data points. Strictly speaking, the least squares algorithm is typically applied to solve over-determined problems [7].
**Kriging**

Kriging postulates a combination of a global model plus departures,

\[ y(x) = f(x) + Z(x) \]

*Equation (4)*

where \( y(x) \) is the unknown function of interest, \( f(x) \) is a known (usually polynomial) function of \( x \), and \( Z(x) \) is the realization of a stochastic process with a mean of zero, variance \( \sigma^2 \), and non-zero covariance. The \( f(x) \) term in this equation is similar to the polynomial model in a response surface and provides a “global” model of the design space [8]. Kriging can be considered as a least squares linear regression technique that is easily generalized to multiple dimensions and arbitrarily sampled points. It assumes that the points are spatially correlated to each other. Covariance function (or a variogram model, \( \gamma \)) is used to express the extent of correlation or spatial continuity. The spatial correlation between the sampled and unsampled points is estimated using the covariance model; it also determines the weight of each sampled point on the estimation. The more spatially correlated a previously sampled value with the estimation location, the more weight it will have on this location. Therefore, Kriging is data exact, such that it will reproduce the observed value at a sampled location. For an experiment of \( N \) observations, Kriging requires the inversion of an \((N+1) \times (N+1)\) matrix, making it not practical for large \( N \) [7].

**Thin Plate Splines (TPS)**

The name thin plate spline refers to a physical analogy involving the bending of a thin sheet of metal. Just as the metal has rigidity, the TPS fit resists bending also, implying a penalty involving the smoothness of the fitted surface. Given a set of data points, a weighted combination of thin plate splines centered about each data point gives the interpolation function that passes through the points exactly while minimizing the so-called "bending energy." In 2-D cases, given a set of \( K \) corresponding points, the TPS warp is described by \( 2(K+3) \) parameters, which include 6 global affine motion parameters and 2\( K \) coefficients for correspondences of the control points. These parameters are computed by solving a linear system. In other words, TPS has a closed-form solution [7, 9].
Response surfaces have been used for optimization [10], history matching [11], and uncertainty quantification [12]. They are easy to implement and after the initial training, obtaining a model response will be quite rapid. However, in order to generate a reasonably accurate response surface, a lot of reservoir realizations (runs) will be needed. This number is directly proportional to the number of parameters involved. Hence, response surface methods are usually a good alternative to reservoir simulators only in the cases with a limited number of parameters (5 to 20) [4]. The required accuracy of the response surface depends on the type of the problem at hand. For instance, the accuracy required for a sensitivity analysis problem is much less than what is required for a history matching or optimization problem.

Despite the mentioned advantages of RSM, there are some potential limitations that do not make it alluring in a lot of cases:

1. Each response surface is designed to model only one response of interest with respect to a series of uncertain parameters. If more than one response is compulsory, new response surface models (potentially required more flow simulations) are required to be built.

2. The general assumption is that the response is smoothly varying with the change of the parameters.

3. Response surfaces are developed to work with the continuous parameters; however, in the oil and gas industry, engineers are dealing with a lot of discrete parameters, such as depositional environments, unconformities, faults, facies probability maps, and permeability curves.

4. Most reservoir models are generated using the geostatistical methods, which results in stochastic “noise” in response surfaces. While customary response surface methods assume that for a given fixed set of input values, the response is uniquely determined [6].

In addition, because the underlying physics of the problem is not incorporated in the response surface, significant error can be observed when the test simulations contain effects that are not included in the training set.
Polynomial Chaos Expansion

Polynomial chaos expansions have been initially introduced to characterize random fields. Recently, they have been used for solving stochastic partial differential equations (SPDE) [13].

PCE methods mainly fall into two categories of intrusive and non-intrusive approaches. In the intrusive approach, PCE approximations are substituted in the governing equations, and the Glarekin method and a discretization scheme are used to obtain the coefficients. In a non-intrusive approach, one can compute the coefficients through a set of deterministic calculations using a small number of model simulations and without any alteration in the governing equations (PDEs). Projection and regression are two different methods for non-intrusive approaches with the latter one having more popularity due to being less computationally prohibitive for many system inputs (uncertain parameters) and/or high-order expansion cases.

Assuming that we have a physical model \( y = \text{model}(x) \), where all input parameters gather in a vector \( x = \{x_1, ..., x_M\}^T \in \mathbb{R}^M, M \geq 1 \), and the model responses (output of interest) are shown by \( y = \{y, ..., y_N\}^T \in \mathbb{R}^N, N \geq 1 \). If the parameter vector \( x \) is uncertain and can be characterized with some probability density function (PDF), then it could be said that each model output is also a random variable. Assuming a particular model output \( y \) has a finite variance, it can be characterized by the following equation (polynomial chaos expansion).

\[
y = P(x) = \alpha_0 B_0 + \sum_{j=1}^{M} \alpha_j B_1(x_j) + \sum_{j=1}^{M} \sum_{k=1}^{j} \alpha_{jk} B_2(x_j, x_k) + \sum_{j=1}^{M} \sum_{k=1}^{j} \sum_{h=1}^{k} \alpha_{jkh} B_3(x_j, x_k, x_h) + ... \\
Equation (5)
\]

Where the \( \alpha' \) s and \( B' \) s are coefficients and the multivariate polynomial basis functions that are orthogonal with respect to the joint PDF of \( x \), correspondingly.

When it comes down to it, classical families of orthonormal polynomials can be associated with usual continuous random variables. If one of the most commonly known
PDFs, such as a normal or uniform distribution can characterize \( x \), then the corresponding orthogonal polynomials can be chosen through the Askey scheme or generalized options. For example, a Hermite polynomial sequence corresponds to a Gaussian random variable \([13, 14, 15]\).

In practice, PCE is truncated and only a finite number of basis terms \( (N_t) \) are retained. This number of terms in expansion can be calculated as follows.

\[
N_t = \frac{(M + d)!}{M! \cdot d!}
\]

*Equation (6)*

Where \( M \) is the number of model inputs (length of \( x \)), and \( d \) is the degree of the expansion, i.e., the highest degree of the polynomial basis functions.

Let us consider a set \( X = \{x^1, \ldots, x^{N_p}\} \) of realizations of the input random vector that is an experimental design. Having \( B \), the coefficient vector \( (\alpha \in \mathbb{R}^{N_t}) \) which provides the best approximation of \( y_p = P(X) \) in the least squares sense by a truncated PC expansion containing a fixed number \( (N_t) \) of terms can be calculated as follows.

\[
B\alpha = y
\]

*Equation (7)*

Where the \( N_p \times N_t \) matrix \( B \), is formed by evaluating the polynomial basis functions at the \( N_p \) selected points, and is normalized to avoid ill-conditioning due to power operations.

The vector \( y = \{y^1, \ldots, y^{N_p}\}^T \) contains \( N_p \) values of a model output \( y \), which are obtained by running the numerical simulation model with the selected points as input values. In order to make the problem well-posed \( N_p \) should be greater than \( N_t \), e.g. in case of Latin Hypercube or random selection the rule of thumb is \( N_p = 2 \cdot N_t \).
Then, $\alpha$ is solved for by minimizing the training error $\| B\alpha - y \|_2^2$, which provided that $(B^T B)^{-1}$ exists will give:

$$
\alpha = (B^T B)^{-1} B^T y
$$

Equation (8)

The approximation error of a PC expansion can be quantified by the coefficient of determination $R^2$, which might under-predict the genuine approximation error and cause overfitting. To keep track of overfitting, the leave one-out cross validation error $Q^2$ is used which is a more robust error estimate [13, 15].

As mentioned before the number of points $N_p$ (simulation runs) for building the PCE depends on the number of polynomial terms $N_t$. The number of polynomial terms, in turn, grows very fast with increase in number of uncertain inputs ($M$) and/or the degree of expansion ($d$). For example, $N_t = 8008$ with $d = 6$ and $M = 10$. Therefore in order to save the computational effort some methods such as forward and backward stepwise regression technique or best subset selection based on mixed-integer quadratic programming (MIQP) can be used to determine the best subset of basis terms in the expansion (optimum $d$), through an iterative process, by setting a target value for $Q^2$ [13, 14, 15].

2.1.2. **REDUCED PHYSICS MODELS**

Reduced-physics model, as implied by its name, reduces the run time by simplifying the physics. In general, the reduced physics model is a simplified version of the full-physics model, dealing with less physical effects or gridblocks. Considering the full-physics model as the reference solution, the parameters associated with the reduced physics model can be determined using a tuning procedure. Usually, an amount of numerical experimentation is needed to lead the engineer toward the level of physics that needs to be used for simplifying the original numerical reservoir model.

In a work done by Wilson on the application of reduced-physics models for numerical simulation of shale reservoirs, a reduced physics model has been used for optimizing the shale reservoir development [16]. In this work, Wilson has reduced the physics by considering the system as a single porosity model and ignoring desorption, non-Darcy
effects, or local grid refinement, but with tuned multipliers for permeability and porosity in the stimulated zone. Using an optimization algorithm, the multipliers that result in the minimum error between the gas production of full and reduced physics models were determined. Instead of explicitly modeling fractures, each fracture in the reduced-physics model is represented through a completion along the wellbore in the stimulated region, bringing about fewer numbers of gridblocks. One drawback of this methodology is that every time a new field design is proposed, the reduced-physics model has to be retrained, by periodically computing an updated multiplier.

Another example of the reduced-physics model is the streamline method. Streamline methods decouple the flow and transport equations, and then solve the transport equations as a series of one-dimensional problems along each streamline. This simplification can lead to substantial speedups relative to traditional simulation for some problems. Streamline methods have been applied to a wide range of problems including production optimization [17] and history matching [18]. Some data are obtained using the streams that are beneficial. Streamlines can outline the drainage and irrigation volumes, which are associated with the producers and injectors, correspondingly. Also, adding up the volumetric flow rates associated with each streamline can help allocating the rate from each well to its offset well. These approaches approximate many key effects, and though they have been widely used for water flooding applications, they are not commonly applied for compositional problems. In addition, the overall speedup using streamline methods is still limited because of the need to solve the full-order flow equations at some time steps. The improved computational speed and memory efficiency apply to problems that are specifically tailored for streamline simulation (SLS). SLS can be applied only to the slightly compressible systems in which the principle physics is displacing the resident oil. In cases of gas cap formation, or repressurization by water injection, SLS will not be applicable. The main drawbacks of the SLS are related to its architectural features: the dual grid and the assumption of independency of the streamlines from one another. These features make SLS non-mass conservative and unable to capture the physics that is transverse to the main direction of flow [19].

\subsection{Reduced Order Models (ROMs)}

Reduced order models are basically the projection of the full order numerical description into a low-dimensional subspace, which reduces the number of unknowns
that must be computed at each time step. These models can be classified into grid-based, system-based, and snapshot methods.

In the grid-based technique, the dimension of the problem is reduced by changing the dimension of the grids, and then, solving the problem based on the coarser grids. Some examples of this technique include upscaling and multi-scale methods. Upscaling can be performed using upscaling of flow by focusing on calculating porosity, permeability, and transmissibility, or upscaling of transport, in which, coarse scale properties, such as relative permeabilities or capillary pressure, are also calculated. There are many works in literature specifically on different types of upscaling [20, 21]. Since, the upscaled model can still be solved using the reservoir simulator, it makes it appealing as a reduced order model; however, it can at best resolve the coarse-scale model and it cannot provide the fine-scale solution. In very heterogeneous models, this can result in some problems. Apart from heterogeneity, in case of compositional models, upscaling has proved not to be a good technique. Despite some recent works in this area [22], an upscaling technique has not yet been developed addressing these issues.

The basis functions that capture fine-scale effects are constructed for each coarse gridblock in multi-scale methods, instead of computing coarse-scale parameters as in upscaling methodologies. Ram’e and Killough [23] introduced a dual-grid method, which has some multi-scale features, and Hou and Wu [24] suggested a multi-scale finite element method (MSFEM), that proved the applicability of multi-scale treatments for Darcy flow. Jenny et al. [25] and Tchelepi et al. [26] established multi-scale methods based on finite volume methods (MSFVM) for the first time within the perspective of reservoir simulation. In these approaches, the pressure equation is solved on the coarse-scale, using specially computed coarse-scale transmissibilities. The transport equation can be solved on the fine-scale (explicitly or implicitly), by reconstructing the fine-scale velocity field using the coarse-scale pressure and the basis functions. They have reported that this approach has resulted in speedups of about a factor of 10 to 20 [26]. Zhou [27] developed a competent two-stage algebraic multi-scale solver that has been shown to deliver better performance than current linear solvers. Besides, he provided a detailed review of various multi-scale formulations. The main problem with these approaches is that the degree of speed up might be very limited due to the essential calculations on the fine-grid in most multi-scale methods. On the other hand, as in upscaling, the multi-scale approach has not been used widely for
compositional models, although some recent works have been performed in this area [28]. Overall, grid-based methods are not the best options in the realm of reservoir management, due to their limited speedup level.

System-based methods are derived from system control theory and are the second category of reduced order models. Krylov subspace and balanced truncation methods (projection-based method) are two more common examples of this group. The main idea in these techniques is that not all outputs of a model are evenly relevant and important for a particular problem at hand. For instance, while dealing with an optimization problem, the “good-enough” approximation of a small part of the model output is of more interest, rather than the accurate approximation of all outputs. These methods are derived by considering a linear time invariant (LTI) system and converting the full-order (n-dimensional) LTI into a reduced-order (l-dimensional) LTI, which can be solved much more efficiently as \( l \ll n \). Since Krylov subspace methods only involve multiplication of matrices that are typically sparse, they are very fast. Their main restrictions are that they tend to generate larger projection subspaces than are requisite, and the resulting reduced-order model has no guaranteed error bound. Generally, single-phase flow (slightly compressible) can be described by LTI dynamic models; however, these methods cannot be used for multiphase flow, which involves different physical phenomena and fluid front interfaces moving all over the spatial domain [29]. Balanced truncation models are based on the concepts of controllability and observability. Controllability refers to the sensitivity of a state to the input or minimal energy required to get to a specific state, and observability defines the influence a state has on the output or the output energy that is generated by the specific state. Controllability and observability are enumerated by the Hankel singular value decomposition (HSVD) of the system [29]. This technique has a provable error bound for the reduced order model and the preservation of stability of the original system is guaranteed by using them; nonetheless, they suffer from very high computational cost. On the other hand, as they are only applied to the linear input-output systems, their application in the perspective of reservoir simulation has been restricted to very small models (usually less than 3,000 gridblocks) with very simple physics [4].
The main difference between snapshot-based methods and system-based methods is that the basis matrices in snapshot-based methods are derived from the snapshots, which refer to the full-state information at a particular instant in time of training simulations. In other words, snapshots are the state vectors generated and saved during the training simulations, which are used for constructing the basis matrix. Proper orthogonal decomposition (POD), also known as principal component analysis, Karhunen-Loeve decomposition, or the method of empirical orthogonal functions, falls in this category. POD is a model reduction technique to produce low-order models using snapshots from a forward simulation with the original high-order model. As other methods explained above, this method has its limitation when applied to compositional models, and the number of gridblocks plays an important role [4].

A more detailed explanation about some of the reduced order models follows.

2.1.3.1. **Proper Orthogonal Decomposition (POD)**

By projecting the high dimensional states into the optimal-lower dimension, POD is able to reduce the dimension of the problem. The simulation used for providing the snapshots is called “training” cases, and the ones used for prediction are called “test” cases.

**High Order Reservoir Model**

To generate a reduced-order model, it is first needed to run a full-order simulation and produce snapshots. Assuming a two-dimensional ($m \times n$ gridblocks), two-phase reservoir model, that has no-flow boundaries at all sides, the behavior of the reservoir model in time can be described by a set of differential equations. This set of model equations can be derived from material balances for oil and water, Darcy’s law, and two closure equations for saturation and pressure, $S_w + S_o = 1$ (where $S_w$ is water saturation, $S_o$ is oil saturation) and $p_c = p_o - p_w$ (where $p_c$ is capillary pressure, $p_o$ is pressure in the oil phase, and $p_w$ is pressure in the water phase) [30].

A spatial finite difference discretization using a five-point block-centered grid definition leads to the following matrix equation in continuous time [31]:
\[
\begin{bmatrix}
\dot{P}_o(t) \\
\dot{S}_w(t)
\end{bmatrix} =
\begin{bmatrix}
\Psi_{11} & \Psi_{12} \\
\Psi_{21} & \Psi_{22}
\end{bmatrix}^{-1}
\begin{bmatrix}
V_{11} & 0 \\
0 & V_{22}
\end{bmatrix}^{-1}
\begin{bmatrix}
T_{11} & T_{12} \\
T_{21} & 0
\end{bmatrix}
\begin{bmatrix}
P_o(t) \\
S_w(t)
\end{bmatrix} +
\begin{bmatrix}
\Psi_{11} & \Psi_{12} \\
\Psi_{21} & \Psi_{22}
\end{bmatrix}^{-1}
\begin{bmatrix}
q_o(t) \\
q_w(t)
\end{bmatrix} = f(P_o, S_w, q_o, q_w)
\]

Equation (9)

The matrix \( \Psi_{2m\bar{n} \times 2m\bar{n}} \) represents the pore fluid compressibility and porosity, \( V_{2m\bar{n} \times 2m\bar{n}} \) is a mass matrix (grid cell volume times density of the fluids), and \( T_{2m\bar{n} \times 2m\bar{n}} \) is constructed by the transmissibilities (containing permeabilities and fluid viscosities) at the gridblock boundaries. The vector with the oil pressures \( P_o \) and water saturations \( S_w \) is called the state vector and has the length of \( 2m\bar{n} \). This state vector has a “twofold” character, since the pressure and the saturation are behaving very differently. The pressure transient is a fast moving signal which also decays fast (highly damped). The saturation, on the other hand, forms a slow moving signal, which is only very lightly damped. This fact causes better prediction in terms of pressure rather than saturation. The vector containing the imposed oil and water flow rates, \( q_o \) and \( q_w \), in the well segments can be written as \( K_{2m\bar{n} \times 2m\bar{n}} u_{2m\bar{n} \times 1} \). \( K \) is a matrix selecting only the elements that contain a well segment, \( u \) is called the input vector and is comprised of total flow rates. Equation (9) is nonlinear since the matrices \( V, \Psi, T \) and the flow rates \( q_o \) and \( q_w \) are functions of \( P_o \) and \( S_w \), and can be rewritten in general space notation as:

\[
\dot{x}(t) = f(x(t), u(t)) = A_c x(t) + B_c u(t) , \quad y(t) = C_c x(t)
\]

Equation (10)

\( A_c \) is the system matrix, and \( B_c \) is the input matrix, and their parameters are still functions of \( x \), \( y \) is called the output vector, and \( K^T \) is represented by output matrix \( C_c \). Here, \( x \) is the n-dimensional state vector. \( n \) is the total number of state variables, i.e. the pressure and saturation at all the gridblocks. Therefore \( n = 2m\bar{n} \).

**Reduction of a high-order model to a linear low order model**

Since most mathematical reduction methods are applicable to linear models, the nonlinear high order reservoir model should be linearized. The linearized high order model
can then be reorganized using different techniques to generate the most dominant and important information \[31, 32\].

Linearizing Equation (10) in point \(x=x^*\) will lead to:

\[
\dot{x}(t) = [f]_{x^*} + \left[ \frac{\partial f}{\partial x} \right]_{x^*} (x(t) - x_0) + \cdots \approx \left[ \frac{\partial f}{\partial x} \right]_{x^*} x(t) + [B_c]_{x^*} u(t) + \left[ A_c - \frac{\partial f}{\partial x} \right]_{x^*} x^*,
\]

\[y(t) = C_c x(t)\]

*Equation (11)*

Considering the condition that \(\left[ \frac{\partial f}{\partial x} \right]_{x^*}\) exists, and \(\bar{x} = x + \left[ \frac{\partial f}{\partial x} \right]_{x^*}^{-1} \left[ A_c - \frac{\partial f}{\partial x} \right]_{x^*} x^*\) and \(\bar{y} = y + C_c \left[ \frac{\partial f}{\partial x} \right]_{x^*}^{-1} \left[ A_c - \frac{\partial f}{\partial x} \right]_{x^*} x^*,\) in general state space notation in continuous time, Equation (11) can be rewritten as:

\[
\dot{\bar{x}}(t) = \bar{A}_c \bar{x}(t) + \bar{B}_c \bar{u}(t), \bar{y}(t) = \bar{C}_c \bar{x}(t)
\]

*Equation (12)*

Where:

\[
\bar{A}_c = \left[ \frac{\partial f}{\partial x} \right]_{x^*}, \quad \bar{B}_c = [B_c]_{x^*}, \text{ and } \bar{C}_c = K^T
\]

Semi-implicit Euler discretization of Equation (12) by treating the state vector \(x(t)\) implicitly, and the input vector \(u(t)\) explicitly, for the unknown state \(x(k+1)\) results in:

\[
x(k + 1) = [I - \Delta t \bar{A}_c]^{-1} [x(k) + \Delta t \bar{B}_c \bar{u}(k)], \quad y(k) = \bar{C}_d \bar{x}(k)
\]

*Equation (13)*

Defining \(\bar{A}_d = [I - \Delta t \bar{A}_c]^{-1}\) and \(\bar{B}_d = [I - \Delta t \bar{A}_c]^{-1} \Delta t \bar{B}_c,\) and \(\bar{C}_d = \bar{C}_c,\) makes it possible to write the general state space system in discrete time, where \(k\) is discrete time:
\[ x(k + 1) = \bar{A}_d x(k) + \bar{B}_d u(k) \quad , \quad y(k) = \bar{C}_d x(k) \]

*Equation (14)*

In the rest of this section, the overbar sign will be omitted in order to simplify the notation.

**POD Development**

The mathematical theory of POD relies upon results from functional analysis, in particular the properties of Hilbert (or more generally, Sobolev) spaces, i.e. complete inner product spaces. While the theory exists for both infinite and finite dimensional systems, in practical applications, it generally comes down to manipulation of finite-dimensional data patterns (vectors) [29].

By projecting the original \( n \)-dimensional state space onto an \( l \)-dimensional subspace, an approximation of the system dynamics is obtained as follows. By running the \( n \)-dimensional system (reservoir model), a total of \( M \) snapshots (state vectors) \( \{x_i\}_{i=1}^{M} \), are generated. \( n \) is generally much larger than the dimension of snapshots, i.e. \( n \gg M \)

\[ X := [x_1 \ x_2 \ ... \ x_M] \]

*Equation (15)*

The objective of POD is to find \( l \) \((l \leq M \ll n)\) orthogonal eigenvectors, represented by \( \{\phi_i\}_{i=1}^{l} \), such that the total square distance,

\[ Q(\phi) := \frac{1}{M} \sum_{i=1}^{M} \|x_i - \phi \phi^T x_i\|^2 \]

*Equation (16)*

between the snapshots and its projections on the subspace defined by the \( n \times l \) matrix

\[ \Phi := [\phi_1 \ \phi_2 \ ... \ \phi_M] \]

*Equation (17)*
is minimized for any \( l \). Essentially, given the data matrix \( X \), which is constructed from state vectors \( (x) \), it is desired to find a transformation as follows:

\[
x \simeq \phi_l z \quad \text{or} \quad x \simeq \phi_l z + \bar{x}
\]

*Equation (18)*

Where \( \phi \) is a \( n \times l \) transformation matrix (as explained above) and \( z \) is a reduced state vector of length \( l \).

The optimization problem is solved by \( \{\phi_l\}_{l=1}^1 \) being the eigenvectors of snapshot correlation matrix \( R \), which has a dimension of \( n \times n \).

\[
R := \frac{1}{M} XX^T \equiv \frac{1}{M} \sum_{i=1}^{M} x_i x_i^T
\]

*Equation (19)*

corresponding to the largest eigenvalues \( \{\lambda_j\}_{j=1}^l \).

The number of dominant eigenvectors \( l \) is usually chosen as:

\[
l := \min_k \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i} \lambda_i} \geq \alpha < 1,
\]

*Equation (20)*

which stands for the smallest number of modes needed to capture the dominant information or desired amount of variability in the data. The value of \( \alpha \) (fraction of energy we want to be captured) is generally close to one, and this means that the first \( l \) eigenvectors carry almost all of the information of the variability in the snapshot set. If the singular values, ordered by magnitude, display a clear drop, the system apparently has a natural set of dominant singular values. Otherwise, the choice of \( \alpha \) becomes somewhat arbitrary. Frequently used cut-off levels are \( 0.9 < \alpha < 1.0 \).

The snapshot covariance matrix is sometimes used instead of \( R \).
\[ \tilde{R} := \frac{1}{M} \tilde{X} \tilde{X}^T \equiv \frac{1}{M} \sum_{i=1}^{M} (x_i - \bar{x}_i)(x_i - \bar{x}_i)^T \]

*Equation (21)*

\( \bar{x} \) represents the mean of snapshots \((\bar{x} = \frac{1}{M} \sum_{i=1}^{M} x_i)\) and \( \tilde{X} := [x_1 - \bar{x} \quad x_2 - \bar{x} \quad \ldots \quad x_M - \bar{x}] \). Subtraction of mean refers to moving the center of mass of the set to the origin of coordinate system. A potential benefit of the subtraction of the mean is an increased level of detail in the reduced-order description, in the case of near parallel snapshot vectors \( x_i \).

Eigenvectors and eigenvalues of symmetric matrices can be solved in an iterative fashion using some forms of the basic Lanczos recursion procedure. When the number of snapshots are less than the dimension of snapshot \((M < n)\), the ranks of the correlation matrix \( R \) and covariance matrix \( \tilde{R} \), are \( M \) and \( M - 1 \), respectively. Hence, there are at most \( M \) (or \( M - 1 \)) eigenvectors \( \phi_i \), which correspond to a non-zero eigenvalue. Instead of dealing with the \( n \times n \) matrix \( XX^T \), one can use the \( M \times M \) matrix, which is much smaller, and calculate the eigenvectors \( (v_i) \) at this dimension, and relate them to eigenvectors \( \phi_i \) in the original dimension.

Consider the following eigenvalue problem

\[ XX^T \phi_i = \lambda_i \phi_i \quad \text{or} \quad R \phi_i = \lambda_i \phi_i \]

*Equation (22)*

It can be proved that Equation (22) is equivalent to: [29]

\[ \phi_i = \frac{1}{\sqrt{\lambda_i}} X v_i \]

*Equation (23)*

It can be observed that the POD modes actually are linear combinations of the collected snapshots. Thereby, POD modes will in general be different whenever the
snapshots set is different, e.g. the cases with different applied boundary or initial conditions.

**Singular Value Decomposition**

Instead of solving the eigenvalue problem, involving product of large and intense matrices, a singular value decomposition (SVD) can be performed directly on the snapshot matrix $X$. Knowing that the dimension of $X$ is $n \times M$, it can be decomposed as:

$$X = U \Sigma V^T$$

*Equation (24)*

where $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{M \times M}$ are orthogonal matrices ($U^T U = I_n$) and ($V^T V = I_M$), and $\Sigma$ is a $n \times M$ pseudo-diagonal matrix. The singular values of $X$ are placed at the main diagonal of $\Sigma$ in non-increasing order, i.e. $\sigma_1 \geq \ldots \geq \sigma_l \geq \sigma_{l+1} \geq \ldots \geq \sigma_n \geq 0$.

These $\sigma$ values are the square roots of the eigenvalues $\lambda_i$, $i = 1, 2, \ldots, M$.

$$\Sigma = \begin{bmatrix}
\sigma_1 & 0 & \ldots & 0 \\
0 & \sigma_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \sigma_M \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0
\end{bmatrix}$$

*Equation (25)*

Matrix $XX^T$, can now be expressed as

$$XX^T = U \Sigma \Sigma^T U^T$$

*Equation (26)*

where $\Sigma \Sigma^T = diag\{\sigma_1^2, \ldots, \sigma_M^2, 0, \ldots, 0\}$. As explained above, these values stand for eigenvalues $\lambda_i, i = 1, 2, \ldots, M$. The number of non-zero values equal to the rank of $X$. The POD modes $\phi_i$ can be computed as “left singular vectors” (elements of $U$) of $X$. In the same way, the “right singular vectors” (elements of $V$) of $X$ are the eigenvectors $v_i$ of $X^TX$ [29].
Having the eigenvalues and using Equation (20), \( l \) which is the number of POD basis functions that is going to be kept, can be determined.

**POD applied on two-phase reservoir simulation**

During the reservoir simulation run, the pressure and saturation at each gridblock and each time step is saved in a vector (state vectors \( x_p, x_s \)). Each vector has a dimension of \( (n \times 1) \), with \( n \) being the total number of gridblocks. These vectors then construct the snapshot matrix \( \mathbf{X}_p, \mathbf{X}_s \), and each have the dimension of \( (n \times M) \), with \( M \) being the total number of time steps or snapshots.

In order to apply POD procedure on a two-phase reservoir simulation, Equation (18) can be replaced in Equation (14). After simplification, the following equation will be obtained.

\[
\begin{align*}
  z(k+1) &= \phi^T [\bar{A}_d x(k) + \bar{B}_d u(k)], \quad y(k) = \bar{C}_d \phi z(k) \\
  &\text{Equation (27)}
\end{align*}
\]

The full order state vector can be calculated by multiplying the reduced vector by the transformation (projection) function.

The number of state variables is reduced from \( n \) to \( l \), \( (l = l_p + l_s) \), and the matrix dimensions are reduced from \( n \times n \) to \( l \times l \). Because only \( l \) equations exist instead of \( n \) equations to solve in semi-implicit discretization, the simulation time is decreased. For a fully implicit simulation, where more than one system of equations has to be solved during every time step, the reduction in simulation time is expected to be more. However, on the bad side, the original penta-diagonal matrix structure (hepta-diagonal for three-dimensional systems) is altered to a full matrix, because it is multiplied by a full matrix \( \phi \). This will hinder the computational advantage of simulation time decrease due to vector size reduction [33].
POD Drawbacks

Potential Pitfalls for Multi-phase Reservoir Simulation

The movement of fluid-front is a challenge for any projection-based methods including POD. POD is found by time-averaging, which will result in smoothing and loss of the information that is expressed over small intervals of time. The accurate representation of the saturation behavior requires many more snapshots to be kept for defining the projection function, than the pressure [29].

Computational Complexity

There are two main parts accounting for the computational cost of POD, namely pre-processing and solving.

Pre-processing involves generating the “good representative snapshots” from the high order reservoir simulation. Due to the lack of the design of experiment methodologies for flow scenarios, a more heuristic approach is usually taken, in which, the well flow rates or bottom hole pressures are excited to create a range covering the expected scenarios. However, it should be considered that both the values and time of update should be very close to the expected scenario. This might require a computationally too demanding process of generating a lot of snapshot data matrices, which are generally dense.

Although solving $l$ dimensional reduced linear system has a cost of $O(l^2)$, which is much less than $n$-dimensional sparse linear system cost of $O(n^{1.1-1.5})$, this speedup will result in no or insignificant computational gain in calculation of the reduced order solution. The only way to overcome this is by trying to calculate the reduced order vector by some other cheaper methods, which often necessitate modification to the reservoir simulator code, as will be explained in next section [29].

2.1.3.2. Trajectory Piecewise Linearization

Achieving a higher level of speedups requires efficient treatment of the nonlinearities involved with POD. There are two main approaches addressing this problem, namely the discrete empirical interpolation method (DEIM) and trajectory piecewise linearization (TPWL). The latter combines POD with an interpolation method for calculating the reduced nonlinear terms. The basis matrix for each nonlinear term is
constructed using the snapshot matrix during training simulation. The nonlinear terms of the test simulation are approximated first by evaluating them at a (small) set of gridblocks, and then by interpolating them using their basis matrices over the entire domain [34]. However, the DEIM approach is intrusive, since it requires the computation of full-order nonlinear terms by the simulator at certain gridblocks during the reduced-order runs. Accordingly, this method does not deliver a surrogate model that can be run outside of the full-order simulator, which is a shortcoming for reservoir simulation applications [4]. Using the TPWL method, the solution at each time step of the test simulation is represented in terms of a linearization around the “closest” saved state computed during the training simulation.

TPWL was first introduced by Rewienski and White [35], where it was combined with a Krylov subspace order reduction approach and applied to circuit design problems. Since the basis from the Krylov subspace has to be updated every time the piecewise linear coefficient matrices are changed, this method has been modified in order to make it applicable to reservoir simulation applications. In the modified approach, POD is used for order reduction rather than a Krylov method, giving rise to a POD-TPWL procedure.

The partial differential equations governing subsurface flow are derived by combining statements of mass conservation for oil and water with Darcy’s law for each phase. These equations can be written as:

$$\frac{\partial}{\partial t} (\varphi \rho_j S_j) - \nabla \cdot \left[ \rho_j \lambda_j k (\nabla p_j - \rho_j g \nabla D) \right] + q_j^w = 0$$

*Equation (28)*

Where the subscript $j$ designates component/phase ($j = o$ for oil and $w$ for water), $k$ stands for the (diagonal) absolute permeability tensor, $\lambda_j = k_{rj}/\mu$ is the phase mobility, with $k_{rj}$ as the relative permeability to phase $j$, and $\mu$ as the phase viscosity. $p_j$ is phase pressure, $\rho_j$ is the phase density, $g$ is gravitational acceleration, $D$ is depth, $t$ is time, $\varphi$ is porosity, $S_j$ is saturation, and $q_j^w$ is the source/sink term [36].
The general oil-water model is completed by enforcing the saturation constraint \((S_o + S_w = 1)\), and by specifying a capillary pressure relationship \(p_c(S_w) = p_o - p_w\). Equation (28) is a non-linear equation. The two-phase flow description entails four equations and four unknowns \((p_o, p_w, S_o, S_w)\). \(p_o\) and \(S_o\) are selected as the primary unknowns (which are generally referred to simply as \(p\) and \(S\)). Once these are computed, \(p_w\) and \(S_w\) can be readily determined from the capillary pressure relationship and saturation constraint. Equation (28) is solved numerically using a fully-implicit finite volume procedure. Discrete representations, based on standard finite-volume treatments and fully implicit schemes, are introduced. The discrete system for the fully-implicit formulation can be written as [37]:

\[
g^{n+1}(x^{n+1}, x^n, u^{n+1}) = A(x^{n+1}, x^n) + F(x^{n+1}) + Q(x^{n+1}, u^{n+1}) = 0
\]

Equation (29)

Where \(x = [p_o, S_w]\) is the state vector, \(u\) represents the (specified) well control parameters (e.g. BHP), \(n\) and \(n+1\) designate the time level and \(g^{n+1}\) is the residual vector we seek to drive to zero, and \(A, F,\) and \(Q\) are the discretized accumulation, flux, and source/sink terms, respectively.

Equation (29) is a non-linear fully implicit system which is solved using Newton’s method. This requires construction of Jacobian matrix \(J = \frac{\partial g^{n+1}}{\partial x^{n+1}}\) and solution of \(J\delta x = -g\), at each iteration, where \(\delta x\) is the solution update vector. Considering that the practical reservoir models might contain up to the order of \(10^6\) gridblocks, this solution can be computationally expensive. The POD-TPWL approach provides an approximate solution with less computational cost for this Equation (25).

The idea behind TPWL is to linearize the residual equation around states saved from training simulations. Superscripts \(i\) and \(i+1\) refer to consecutive time steps in the training simulations, and \(n\) and \(n+1\) refer to time steps in test simulations.

Given the solution at time step \(n\), the solution at time step \(n+1\) is expressed as a linearization around the training solution \((x^{i+1}, x^i, u^{i+1})\) as follows:
\[ g^{n+1} \approx g^{i+1} + \frac{\partial g^{i+1}}{\partial x^{i+1}}(x^{n+1} - x^{i+1}) + \frac{\partial g^{i+1}}{\partial x^i}(x^n - x^i) + \frac{\partial g^{i+1}}{\partial u^{i+1}}(u^{n+1} - u^{i+1}) \]

Equation (30)

Where \( g^{n+1}(x^{n+1}, x^n, u^{n+1}) = 0 \) and \( g^{i+1}(x^{i+1}, x^i, u^{i+1}) = 0 \).

Defining Jacobian matrix as \( J^{i+1} = \frac{\partial g^{i+1}}{\partial x^{i+1}} \) and having \( B^{i+1} = \frac{\partial g^{i+1}}{\partial x^i} \) and \( C^{i+1} = \frac{\partial g^{i+1}}{\partial u^{i+1}} \) will result in:

\[ J^{i+1}(x^{n+1} - x^{i+1}) = -[B^{i+1}(x^n - x^i) + C^{i+1}(u^{n+1} - u^{i+1})] \]

Equation (31)

The dimensions of \( J^{i+1}, B^{i+1}, \) and \( C^{i+1} \) are \( 2N \times 2N, 2N \times 2N, \) and \( 2N \times 2n_u \), with \( N \) being the number of gridblocks and \( n_u \) being the number of well control parameters in the problem (recall that there are two unknowns per block in an oil-water problem).

Once \( u^{n+1} \) is set, the only unknown in Equation (31) is \( x^{n+1} \). As this is a linear equation, it can be solved without going through the hassle of any nonlinear iteration. However, this equation is in high dimensional space, i.e. the dimension of \( x^{n+1} \) is \( 2N \). Some methods can be applied to construct an orthonormal basis that can be used to project the high-fidelity linearized models to a lower-dimensional representation. These methods include optimal Hankel model, balanced truncation, Krylov subspace methods, and proper orthogonal decomposition (POD) [36].

As explained in the last section, the aim of POD is to construct a basis matrix \( \Phi \) such that:

\[ x = \begin{bmatrix} P \\ S_w \end{bmatrix} \approx \Phi z = \begin{bmatrix} \phi_p & 0 \\ \phi_s & I \end{bmatrix} \begin{bmatrix} z_p \\ z_s \end{bmatrix} \]

Equation (32)

The dimension of \( \Phi \) is \( 2N \times l \), where \( l = lp + ls \). In general, \( l \ll 2N \), thus substantial reduction is achieved.
Replacing Equation (32) into Equation (31) will result in:

\[ J^{i+1} \phi (z^{n+1} - z^{i+1}) = -[B^{i+1} \phi (z^{n} - z^{i}) + C^{i+1}(u^{n+1} - u^{i+1})] \]

Equation (33)

which contains \( N \) equations but only \( l \) unknowns.

Pre-multiplying both sides of Equation (33) by \( \phi^T \) gives:

\[ \phi^T J^{i+1} \phi (z^{n+1} - z^{i+1}) = -\phi^T [B^{i+1} \phi (z^{n} - z^{i}) + C^{i+1}(u^{n+1} - u^{i+1})] \]

Equation (34)

To simplify the notation the following is defined:

\[ J^{i+1}_r = \phi^T J^{i+1} \phi , \quad B^{i+1}_r = \phi^T B^{i+1} \phi , \quad C^{i+1}_r = \phi^T C^{i+1} \]

Equation (35)

These terms only need to be constructed (and saved) once during the training simulation. The dimensions of \( J^{i+1}_r \) and \( B^{i+1}_r \) are now both \( l \times l \), and the dimension of \( C^{i+1}_r \) is now \( l \times N \), with \( l \sim O(100 - 1000) \). Consequently, all three matrices are now in low-dimensional space.

Equation (34) can now be written as:

\[ z^{n+1} = z^{i+1} - (J^{i+1}_r)^{-1} B^{i+1}_r (z^{n} - z^{i}) + C^{i+1}_r (u^{n+1} - u^{i+1}) \]

Equation (36)

Or considering Equation (29):

\[ z^{n+1} = z^{i+1} - (J^{i+1}_r)^{-1} \left[ \left( \frac{\partial A^{i+1}}{\partial x^i} \right)_r (z^{n} - z^{i}) + \left( \frac{\partial Q^{i+1}}{\partial u^{i+1}} \right)_r (u^{n+1} - u^{i+1}) \right] \]

Equation (37)
Equation (36) or Equation (37) represent the POD-TPWL system that should be solved at each time step in a test simulation. This equation is linear and low dimensional in contrast to the original full order model system.

In order to solve Equation (37), first a saved state \((z^i, z^{i+1}, u^{i+1})\) should be determined as a “point” to expand around. This is accomplished by defining a distance \(d^{n,j}\) between the test solution at time step \(n\) and any saved point \(j\) in the training simulation. The point that gives the minimum distance is then selected as the linearization point.

Solving Equation (36) gives reduced variable \(z^{n+1}\), from which the full order state matrices \((P, S_w)\) can be reconstructed at the selected time and location. For example, to determine pressure \(p\) in block \(k\) at time step \(n + 1\), we can simply multiply row \(k\) of \(\phi_p\) by \(z_p^{n+1}\).

In order to obtain the flow rates at the wells, only pressure and saturation in well blocks need to be reconstructed. The flow rate into or out of the well block \(k\) for phase \(j\), can be calculated as:

\[
q_{kj}^{n+1} = \lambda_{kj}^{n+1} W_{Ik} (p_k^{n+1} - p_k^w)
\]

*Equation (38)*

where \(q_{kj}^{n+1}\) is the volumetric flow rate of phase \(j\) from block \(k\) into or out of the well at time \(n + 1\) , \(\lambda_{kj}^{n+1}\) is the mobility of phase \(j\) at block \(k\) at time \(n + 1\) , which is a function of saturation in that block. \(W_{Ik}\) is the well index for block \(k\), which is dependent on permeability and thickness values at that gridblock, \(p_k\) is the gridblock pressure, and \(p_k^w\) is the wellbore pressure for block \(k\).

POD-TPWL procedure can be divided into online and offline processes.
Offline Processing:

1. Run reservoir simulation to prepare the state vector and matrices \( \mathbf{X} \). Save the derivative terms \( f^i, B^i, \) and \( C^i \) in the training runs.
2. By performing SVDs on snapshot matrices \( \mathbf{X}_p, \mathbf{X}_s \), construct the basis matrix \( \mathbf{\Phi} \).
3. Compute the reduced states \( z^i \) and reduced derivative matrices \( f^i_r, B^i_r, C^i_r \) from the training runs or runs to be used for linearization (note that some training runs may be used only to generate snapshots).

Inline Processing:

1. Given \( z^n \), find the training point \( i \) to linearize around by minimizing \( d^{n,j} \).
2. Apply Equation (36) to obtain \( z^{n+1} \).
3. Reconstruct the full-order state \( \mathbf{x}^{n+1} \) from \( z^{n+1} \) at required locations.
4. Calculate well flow rates (using Equation (38)) and any other required information.
5. Repeat Steps 1–4 until the final simulation time is reached [4, 37].

POD-TPWL Drawbacks

This method is an intrusive method as the reservoir simulator code should be modified to output the Jacobian information from the training runs at each time step of the high fidelity model, as required by the TPWL procedure. As theses matrices can be quite large, they occupy substantial disk space. Apart from the storage problems, as the snapshot and all the reduced matrices should be constructed following the training runs, the computational cost can be significant and comparable to the time required to perform the training simulations. Because of the overhead and storage requirement for saving the matrices, sometimes the solution is to construct and save required matrices only for a part of training simulations.

If material balance error is of interest, the high fidelity matrices should be reconstructed and the calculations should be performed using them, which adds major computational costs, and is usually skipped.
If the distance calculation for selecting the “closest” saved state to the previous time step is performed using the high fidelity model state vectors, it would be time consuming. To avoid this problem, the reduced states can be used instead. Since the problem is linear in pressure, but nonlinear in saturation, the distance calculation is performed on saturation rather than on pressure states. The limitations imposed by these simplifications might lead to selecting a linearization point that does not necessarily represent the test state. The control values (e.g., BHP) of the test cases should be in the range of the training values, otherwise large errors can be observed for some quantities (e.g., water production rate) [36, 4, 37].

**2.1.4. DATA DRIVEN MODELING (DDM)**

Recent developments in computational intelligence have greatly expanded the capabilities of data-driven modeling, which comprises a new category of proxy modeling. As the name suggests, DDM is founded on scrutiny of the relation of system state variables [38]. The most common mistake in interpretation of DDM is the belief that no knowledge of the physical behavior of the system is required. Many DDMs are built based on this belief. Although in most cases they might end up with acceptable results, they fail miserably in others. This failure has been attributed to DDM rather than the procedure through which DDM is applied.

The following overlapping fields contribute in developing DDMs.

- Artificial Intelligence (AI), which is incited from the human brain, and tries to incorporate human intelligence into computers.
- Computational Intelligence (CI), which includes neural networks, fuzzy systems, and evolutionary computing in addition to other areas within AI and machine learning.
- Soft Computing (SC), which is similar to CI, but with special emphasis on fuzzy rule-based systems induced from data.
- Machine Learning (ML), which concentrates on the theoretical foundations used by CI and SC and was once a sub-area of AI.
- Data Mining (DM) and Knowledge Discovery in Databases (KDD) are mostly concentrated on huge databases and are concomitant with banking, financial services, and customer resources management applications. DM is understood as
a part of a wider KDD. Methods used in DM are generally from statistics and ML.

- Intelligent Data Analysis (IDA), which generally focuses on data scrutiny in medical research and integrates methods from statistics and ML.

A ML algorithm is used to figure out the relationship between the input and output of a system using a training set that is representative of the system behavior. The trained model can be tested on an independent dataset to test its generalization ability.

Some of the most popular computational intelligence techniques for modelling purposes are neural networks (NN), fuzzy rule based systems (FRBS), genetic algorithms (GA), chaos theory and non-linear dynamics, genetic programming, and evolutionary regressions and support vector machines (SVM) [38].

Artificial neural network (ANN) is a technique that has been used in the majority of proxy models, including the work performed in this study, and has been expounded more in the following section.

**Artificial Neural Network**

While dealing with complex problems, following the lemma “divide and conquer,” is an efficient way for solving them. The complex system can be divided into smaller parts to be handled; on the other hand, simple elements can be put together to produce a complex system [39]. Using networks is one way to approach this.

There are many types of networks, which are all characterized by a set of nodes and connections between them. The nodes can be seen as computational units. They take inputs, and process them to attain an output. The connections can be unidirectional or bidirectional. The communications of nodes through the connections bring about a global behavior of the network, which cannot be detected in the elements of the network. This global behavior is said to be emergent, which means that the abilities of the network supersede the ones of its elements, making networks a very powerful tool [40].

In one type of network, the nodes are seen as artificial neurons. This type of network is called artificial neural network (ANN). Neural networks are biologically inspired computational models founded on the way the human brain functions. Neural networks
are designed to represent the association between the input and output and the system behavior inherent in the data. An artificial neuron is a computational model inspired by the natural neurons. In natural neurons, signals are received through synapses located on the dendrites or membrane of the neuron. If received signals are strong enough that exceed a certain threshold, then the neuron is activated and emits a signal through the axon. This signal might be sent to another synapse, and might activate other neurons.

The complexity of a natural neuron is highly diminished to model an artificial neuron. It is essentially comprised of inputs (like synapses), which are multiplied by weights (standing for strength of received signals), and then undergoes a mathematical function which determines the activation of the neuron, and finally the output is computed. In general, each neuron takes in inputs either externally or from other neurons and passes it through an activation or transfer function such as a logistic or sigmoid curve. Data enters the network through the input units organized in what is called an input layer. These data are then fed forward through successive layers including the hidden layer in the middle to emerge from the output layer on the right. The inputs can be any combination of variables that are thought to be important for predicting the output; that is why the knowledge of the system being modeled is important. ANN combines artificial neurons in order to process information. The higher the weight is, the more impact the associated input will have. By adjusting the weights, a desired output for a specific set of inputs can be reached. This process is called learning or training [38, 40].

Figure 1. A biological neuron (Left) and a general artificial neuron (Right) [41]
Backpropagation Algorithm
Since the first neural model by McCulloch and Pitts [42], hundreds of types of ANNs have been developed. Their main differences are the activation functions, accepted values, topology, and the learning algorithm. Back propagation algorithm is one of the most popular ANN types for training multilayer perceptron (MLP) configurations [43]. In these configurations, the artificial neurons are organized in layers, and send their signals “forward.” After computations are performed, the errors are propagated backwards. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons in an output layer. There may be one or more intermediate hidden layers. This algorithm uses supervised learning, meaning a set of input and output data is provided. The backpropagation algorithm is a variation of a gradient descent optimization algorithm that minimizes the error between the predicted and actual output values. The dataset is divided into two or three categories of training, calibration, and validation. The weighted connections between neurons are adjusted after each training epoch until the error in the calibration data set begins to rise. The calibration data set is a second data set that is given to the network to evaluate during training. Having a calibration set helps the network not to get over-trained. If overtraining happens, then the network will not be generalized and will not work the same on the test sets. An important way to help generalization of the network is making sure that the training set is representative of the behavior in the whole dataset, e.g. all training, calibration, and validation datasets having similar statistical properties [38, 40]. The following is a brief explanation of backpropagation algorithm.

The weighted sum of the values in the input neurons ($S$) goes through an activation function in the hidden layer neuron. The Sigmoid function is the most common activation function. Where $x$ and $w$ represent the input and weight correspondingly, these functions can be shown as:

$$S_j(\bar{x}, \bar{w}) = \sum_{i=0}^{n} x_i w_{ji}$$

Equation (39)
The sigmoid function output is very close to zero for values close to negative infinity, and gets close to one as the input value gets very large. At zero, the function output is 0.5. The objective of the training process is to obtain a desired output when given a set of inputs. The error is the difference between the actual output and the one calculated by the network and is dependent on the weights and can be defined as:

\[ E_j(\bar{x}, \bar{w}, d) = \sum_j (A_j(\bar{x}, \bar{w}, d) - d_j)^2 \]

Equation (41)

where \( d \) is the actual output.

\( E \) can be minimized using an iterative process of gradient decent. Each weight can be updated using the increment

\[ \Delta w_{ji} = -\gamma \frac{\partial E}{\partial w_{ji}} \]

Equation (42)

where \( \gamma \) represents a learning constant, i.e., a proportionality parameter which defines the step length of each iteration in the negative gradient direction. According to this formula, the adjustment of each weight \( \Delta w_{ji} \) will be the negative of a constant \( \gamma \) multiplied by the dependence of the previous weight on the error of the network, which is the derivative of \( E \) in respect to \( w_j \).

Consider a network with a single real input \( x \) and network function \( F \). The derivative \( F'(x) \) is computed in two phases:
• Feed-forward: the input $x$ is fed into the network. The primitive functions at the nodes and their derivatives are evaluated at each node. The derivatives are stored.

• Backpropagation: the constant 1 is fed into the output unit and the network is run backwards. Incoming information to a node is added and the result is multiplied by the value stored in the left part of the unit. The result is conveyed to the left of the unit. The result collected at the input unit is the derivative of the network function with respect to $x$.

In the learning procedure using the backpropagation algorithm, the feed-forward step is computed in the usual way, and the output of each node is stored in its right side. Considering Equation (41), the error dependence on the output can be defined as:

$$\frac{\partial E}{\partial A_j} = 2(A_j - d_j)$$

*Equation (43)*

And the effect of weights on the output can be obtained by:

$$\frac{\partial A_j}{\partial w_{ji}} = A_j(1 - A_j)x_i$$

*Equation (44)*

Equation (43) and Equation (44) yield to:

$$\frac{\partial E_j}{\partial w_{ji}} = 2(A_j - d_j)A_j(1 - O_j)x_i$$

*Equation (45)*

And the adjustment to each weight will be:
\[ \Delta w_{ji} = -2\gamma(A_j - d_j)A_j(1 - A)x_i \]

*Equation (46)*

What was explained so far applies for a network having two layers. If more layers are added, then the dependence of the error on the weight of neurons in those layers should be considered as well [40].

In a general case, having a network with \( n \) inputs, \( k \) hidden, and \( m \) output units, the weight between input site \( i \) and hidden unit \( j \) will be called \( w_{ij}^{(1)} \), and the weight between hidden unit \( i \) and output unit \( j \) will be represented by \( w_{ij}^{(2)} \). The bias of each unit is implemented as the weight of an additional edge. The weight between constant 1 (bias) and the hidden unit \( j \) is denoted by \( w_{n+1,j}^{(1)} \), and the weight between constant 1 and the output unit \( j \) is signified by \( w_{k+1,j}^{(2)} \).

*Figure 2. Notation for the three-layered network [44]*

Having the weighted sum of the inputs in the \( j \)-th hidden unit as \( \text{net}_j \), and a sigmoid activation function will yield:
\[ \text{net}_j = \sum_{i=1}^{n+1} w_{ij}^{(1)} \delta_i \]

*Equation (47)*

\[ o_j^{(1)} = s \left( \sum_{i=1}^{n+1} w_{ij}^{(1)} \delta_i \right) \]

*Equation (48)*

The excitation of all units in the hidden layer can be computed with the vector-matrix multiplication \( \delta W_1 \). The components of vector \( o^{(1)} \) are the outputs of the hidden units.

\[ o^{(1)} = s(\delta W_1) \]

*Equation (49)*

\( o^{(2)} \) is an \( m \)-dimensional vector which is the output of the network.

\[ o^{(2)} = s(\delta^{(1)} W_2) \]

*Equation (50)*

The above formulas can be generalized for any number of layers [44].

In the backpropagation step, \( W_1 \) and \( W_2 \) which are \( n \times k \) and \( k \times m \) matrices are defined. These matrices do not have the bias row. Vector \( e \) is defined as the stored derivatives of the quadratic deviations. The \( m \)-dimensional vector of backpropagated error up to the output units \( \delta^{(2)} \) and the \( k \)-dimensional vector of the backpropagated error up to the hidden layer \( \delta^{(1)} \) are shown by

\[ \delta^{(2)} = D_2 e \]

*Equation (51)*
\[ \delta^{(1)} = D_1 W_2 \delta^{(2)} \]

*Equation (52)*

The corrections for the matrices \( W_1 \) and \( W_2 \) are then specified by

\[ \Delta W_2^T = -\gamma \delta^{(2)} \delta^{(1)} \]

*Equation (53)*

and

\[ \Delta W_1^T = -\gamma \delta^{(1)} \delta^{(1)} \]

*Equation (54)*

If the connection matrix between layer \( i \) and \( i + 1 \) is symbolized by \( W_{i+1} \), then the backpropagation error to the output layer can be obtained by

\[ \delta^{(i)} = D_i e \]

*Equation (55)*

And the backpropagated error to the \( i \)-th computing layer can be computed by

\[ \delta^{(i)} = D_i W_{i+1} ... W_{l-1} D_{l-1} W_{l-1} D_l e \]

*Equation (56)*

The corrections to the weight matrices are performed as in the two-layer network [44].

**Network Parameters**

When designing a neural network, there are some parameters that should be decided upon in advance, such as the number of layers, the number of neurons per layer, the number of training iterations, etc. The number of hidden neurons, the learning rate, and the momentum parameter are some of the more important parameters in terms of training a network, which are explicated more below [45, 46].
Number of Neurons in the Hidden Layer

The neurons that are located neither in the input nor in the output layers are called hidden neurons. Increasing the number of hidden layers might enable greater processing power and system flexibility, but on the downside the complexity in the training algorithm rises. The situations in which performance improves with a second (or third, etc.) hidden layer are very small. One hidden layer is adequate for the large majority of problems. An analogy to a network with many neurons is a system of equation with way more equations than free variables. This can cause the system to be over specified, which makes it incapable of generalization. Contrariwise, having a network with too few hidden neurons can reduce the robustness of the system and lead to under-fitting, i.e. a system that cannot properly fit the input data. There are some empirically-derived rules-of-thumb on how to select the number of nodes in the hidden layer. The most commonly relied on is “the optimal size of the hidden layer is usually between the size of the input and size of the output layers.” Jeff Heaton offers a few more [47]. In general, for most problems, one could probably get decent performance (even without a second optimization step) by setting the hidden layer configuration using just two rules: setting the number of hidden layers to one; and considering the number of nodes in that layer as the mean of the neurons in the input and output layers [45, 48].

Learning Rate

The learning rate is a network training parameter that specifies the amount of change of weight and bias during the training. It applies a greater or lesser portion of the respective adjustment to the old weight. If the factor is set to a large value it may cause a quicker learning process; nonetheless, under the circumstances that there is a high variability in the input data, the network may not learn very well or at all. Usually, the best approach is to set the learning rate to a small value and edge it upward if the learning rate seems slow.

Momentum

Momentum simply adds a portion of the previous weight update to the current one. It is used for preventing the convergence of the system to a local minima or saddle point. A high momentum can also cause faster training. However, setting momentum to a very high value can cause overshooting the minimum and system instability.
Conversely, a very small value of momentum might cause getting stuck in the local minima and slowing the training down.

**Epoch**
During iterative training of a neural network, an epoch is a single pass through the entire training set, followed by testing of the calibration set.
2.2. **Proxy Modeling in Literature**

It is not uncommon to make hundreds of numerical simulation runs for model updating or optimization steps. For real field problems, a single simulation run can take hours or even days. Recently, proxy models have been used quite intensively to alleviate the computational cost associated with many problems in the petroleum industry, including field development, history matching, uncertainty analysis, optimization, etc.

While there have been some methods used for well-based proxy modeling, there are very limited techniques available in the literature focused on grid-based proxy modeling for a black-oil reservoir model, let alone the compositional simulations. These methods have stern pitfalls and drawbacks. Attempts have been made to shed some light on these methods and studies in the following paragraphs.

Lerlertpakdee et al. [49] used the relationship between the output performance of the reservoir with the input parameters and production strategy to develop a reduced order model in a two-phase flow system. Their model was a 1-D flow network, which essentially used the injection/production information. In this method, instead of considering all the gridblocks and solving the PDE, only the connections of the wells are considered and divided to course blocks and the derived system of equation is solved for the new 1-D flow network model. The calibration (training) process includes calculating the output of multiple high fidelity model runs followed by using an objective function, which minimizes the misfit between the high and reduced order model results by changing the width (representing PV) and permeability of each gridblock. Although the model has shown a close match between the high order and the reduced order model NPV results, it should be noted that the only variables which are altering in the 10 training sets are BHP of the wells in a specific range. The models used for this study were small (maximum 13,200 GBs) and 2-D. Potential computational cost increase or accuracy loss due to gross upscaling might result if the technique is used for multiphase flow and a more complex geological model with multiple layers and production/injection wells.

Zhang and Sahinidis [14, 15, 50] used Polynomial Chaos Expansion (PCE) to build a proxy model, which is used for uncertainty quantification and injection optimization in a carbon sequestration system. The defined model is a 2-D homogenous and isotopic saline formation, with 986 gridblocks. Quantifying the impact of the uncertain
parameters (porosity and permeability) on the model outputs (pressure and gas saturation through the formation) were of interest. 100 simulation runs were performed in order to build PCEs. The PCEs are of order of \((d = 3 \text{ or } 4)\). Since the coefficients in the PCE model are dependent on space and time, this process involves generating a large number of PCE models, i.e. for each output (e.g. pressure) the number of PCEs needed will be the product of the number of gridblocks and the number of time steps. As the number of inputs (uncertain parameters) increases, the number of numerical runs needed sharply grows which makes this approach impractical, e.g. at least 8,008 runs are required in the case of having 8 inputs and PCE of order 6. The runtime of seconds for PCE and 15 minutes for numerical simulation has been reported, while it should be noted that each PCE provides the information only in one gridblock and a specific time whilst the numerical simulation provides the results at all the time steps and gridblocks. In gas saturation calculations, some of the outputs are negative values, which are numerical errors due to approximations using the polynomial terms. For gas saturation, after 30 days of injection, an average error of 6% has been reported. Although the results have been smoothed by using contours, there are some wiggling on the pressure contour boundaries, which might be due to the scillating feature of polynomial terms in the expansion.

Van Doren et al. proposed reduced order modeling for production optimization in waterflooding process by use of POD [33]. They constructed the ROM using the data obtained from many snapshots of the model states (pressure and water saturation distribution) from the full order simulation model. The POD was used to summarize the dynamic variability of the full order reservoir model in a reduced subspace. Although the number of state vectors (containing oil pressures \(P_o\) and water saturations \(S_w\) for each gridblock) is decreased using POD, change in the matrix structure from penta-diagonal (or hepta-diagonal in 3-D systems) to a full matrix, counteracts the computational advantage obtained by vector size reduction. The methodology was tested on a waterflooding scenario in a 2-D, 2-Phase model having 2,025 gridblocks and 2 horizontal (injection and production) wells with control valves in each gridblock (90 segments). The resulting reduction in computing time for NPV optimization for this sample model was maximum 35%. When the ROM was simulated with the same controls (rates of wells) as the original full-order model, the states were almost identical; however, if the controls were altered (blind set), then the states of the full
order model were less represented by the ROM. In general, although the POD methodology yields reduced-order models with low complexity, the actual speed up on the simulation is modest as compared with the size of the models. This is due to the fact that the nonlinear function for estimating the state vectors will still be evaluated at the full order number of states, which can be computationally expensive and inefficient. The main drawbacks of the POD stem from the fact that the projection basis is dependent upon the training inputs and the time scale that the snapshots are taken.

Cardoso and Durlofsky [51] emphasized that for nonlinear problems the POD procedure is limited in terms of the speed up it can achieve, because it targets only the linear solver and the computation effort for some of the operations (constructing the full residual and Jacobian matrices at every iteration of every time step) is not reduced. These authors applied a linearization process (trajectory piecewise linearization) to the governing equations in addition to the reduced order model obtained from the POD projection, and incorporated it in production optimization. The process under the study was waterflooding in a 3-D model having 20,400 gridblocks and six wells (4 producers and 2 injectors), in which the well BHPs were altering. The Trajectory Piecewise Linear (TPWL) algorithm was first developed in the electrical engineering framework in order to extract reduced-order models in circuit simulation. Although the final calculations for estimating the new state vectors can be performed fast using this method, the preprocessing calculations, including running some number of high fidelity training simulations, the construction of the reduced basis function, construction and inversion of reduced Jacobian matrix for all saved states (which needs modification in simulator code), construction of reduced representation of states, accumulation matrix, and source/sink term matrix of derivatives is still computationally costly. On the other hand, the accuracy of TPWL solution is sensitive to the number of basis vectors used in the projection matrix; therefore, some amount of numerical experimentation may be required to establish these numbers for saturation and pressure. Instabilities and deterioration of accuracy due to application of the procedure to control inputs far away from the training trajectories are one of the main drawbacks of TPWL [34].

One application of proxy models is for dealing with the compositional simulation. These simulations are required for modeling the EOR, CO₂ sequestration, or gas injection. Due to the intrinsic nonlinearity of these models and the potentially large system of
unknowns, they can demand high computational powers. Especially for production optimization, hundreds or thousands of simulation runs should be performed, which accentuates how imperative it is that a model runs efficiently.

Proxy models sometimes have been used as a term just to represent correlations or simple mathematical models to provide single outputs. In one of these applications, oil recovery of a reservoir after carbon dioxide flooding has been modeled as a hint for reservoir screening. The reservoir models used in this study are simple homogeneous models (except in permeability), which differ in type of reservoir fluid [52].

Yang et al. have used a hybrid modeling technique for reservoir development using both full-physics and proxy simulations [53]. The proxy model used in their work is a profile generator. It simplifies the reservoir into material balance tanks and well source/sink terms into a set of tables known as type curves, which relates production GOR, water cut, etc. to EUR and other parameters. Their objective is to reduce the computational cost associated with the modeling of huge reservoirs. The profile generation data should be provided for each well, which is not using the simulation grid. The main assumption is that the reservoir is to be operated in a way not significantly different from the base case for which the type curve data is produced, which might not be practical in most real field problems.

In another attempt in using proxy models, NRAP (National Risk Assessment Partnership) has focused on using this tool for long-term quantitative risk assessment of carbon storage [54]. This is performed by dividing the carbon storage system into components (reservoir, wells, seals, groundwater, and atmosphere), using proxy models for each component and integrating all the models to assess the success probability of carbon storage using the Monte Carlo simulation. Different proxy models are used including look-up table (LUT), response surface, PCE, and AI-based surrogate reservoir models. The look up table methodology is very simple but requires hundreds of runs of the high-fidelity model based on different inputs. The table is built based on the inputs, the results of the simulation runs, and a third dimension representing the time step. The outcome of a new scenario can be obtained using an interpolation-based approach from the created table. Although this method is quite rapid, the problem lies in the number of the full-physics model simulations needed to build the table. In the work performed by NRAP, for a 2-D, 2-phase (saline formation) model with 10,000 (100 by
and only three variable parameters of reservoir permeability, reservoir porosity, and seal permeability, more than 300 runs were needed to build a table for predicting the pressure and saturation at each gridblock. A heterogeneous field cannot be used in this approach and permeability must be varied through a scalar multiplier. Different time snapshots were selected in the interval of 1,000 years of post-injection. The size of the look up table and accuracy of the model depends on the selected snapshots and the time span between them.

He et al. have used POD-TPWL, which is a combination of trajectory piece wise linearization and proper orthogonal decomposition, in order to build a reduced order model [55]. Using POD alone will result in high order complexity due to the need for construction and projection of nonlinear terms; thereby TPWL has been used to address this problem. One of the limitations with this method is that the system stability is highly dependent on the type of projection scheme used. This method needs a lot of offline processing for constructing the POD-TPWL model. In order to get more accurate results, more training runs are needed which results in storage and computational problems. Modifications should be done to the reservoir simulator, hence it is not applicable using the conventional reservoir simulators. The number of variables is equal to the multiplication of the number of gridblocks and fluid components. The pre-processing (offline computations) involve running the full-order training simulations, saving and reading the states and derivatives, constructing the basis matrices, and reducing the states and derivatives, which considering only two or three training simulations, requires the same amount of time as one full-order simulation run. This technique has been tested on a small synthetic reservoir with a few wells (less than 10). The primary variables have been set to be pressure and component mole fraction. In order to obtain the flow rate, the full order primary variables should be reconstructed at a specific time and locations, and secondary variables, such as saturations, should be calculated through performing flash calculations with the primary variables. Increasing the number of the wells will introduce more variability in the states, which may have a considerable effect on the model results and computational expenses. Besides, since the solution is made based on the linearization around the generated states, it is extremely important that the test case is in the range and close to the training runs. For instance, the variable parameter in the test cases (BHP) has been selected to be very similar to the training cases.
Zhang and Pau developed a ROM for CO₂ storage in brine reservoirs [56]. Their objective was to use the ROM for risk assessment of geological carbon sequestration. Their study was based on building a response surface from a set of high fidelity forward simulations for selected parameter values. The approaches used included Gaussian process (GP) regression, radial basis function, and a lookup table combined with linear interpolation. Their ultimate goal was to predict the pressure value at a specific location and a specific time in the reservoir. A relative error was used to quantify the accuracy of approximation. Only three parameters (permeability and porosity of a sand layer and permeability of the cap rock) have been used for building the realizations. While the highest number of simulation models used for this approach is 57, which is much less than a typical response surface approach, the key limitation of this method is that the ROM will be valid for predicting only one output of interest (e.g. for a specific time and location). Thus, in order to predict each parameter through the reservoir over a span of time interval, the number of ROMs required to be built will be equal to the multiplication of the number of gridblocks and number of time steps, which makes this approach prohibitive.

In another work, Hejin et al. studied the application of five different methods in order to derive low-order models of two-phase (oil-water) reservoir flow [31]. Their study was performed on a simple synthetic model having 8 gridblocks in X and Y direction and only one layer. Two wells (one injection and one production) were considered. Modal decomposition, balanced realization, a combination of the last two methods and subspace identification resulted in linear low-order models, which were only valid for a limited time span during which the linearization was valid (only 10 days in their study). Proper orthogonal decomposition (POD) reduced the high-order model to a non-linear low-order model; however, it did not reduce the simulation time.

Gildin et al. combined the discrete empirical interpolation method (DEIM) with POD and proposed using this method to overcome the issues with the nonlinear projections [34]. DEIM is based on the approximation of the nonlinear terms by way of an interpolatory projection of a few selected snapshots of nonlinear terms. The authors applied this model on a 2-D, 2-phase, and 5-spot reservoir model having about 2,000 gridblocks and compared the results in terms of oil pressure and water saturation, with POD-TPWL. POD-TPWL gave a faster methodology and smoother results for approximating the full non-linear behavior of the reservoir in this case. Oscillations
were obtained in the production profiles of oil and water for the 5-spot pattern studied that incites more study on the interpolation algorithm.

Data driven modeling sometimes has been used in combination with other reduced order modeling techniques. Klie used a non-intrusive model reduction approach based on POD, DEIM, and Radial Basis Functions (RBF) networks to predict the production of oil and gas reservoirs [57]. POD and DEIM helped projecting the matrices from a high dimensional space to a low dimensional one. The DEIM method has a complexity proportional to the number of variables in the reduced space. In contrast, POD shows a complexity proportional to the number of variables in the high-dimensional space. Both POD and DEIM allow the collapse of a large number of spatio-temporal correlations. POD and DEIM can work as a sampling method for preparing the RBF network input.

In this work, a space-filling strategy has been used to build the realizations. Permeability in one and permeability and injection rate in another model are the control variables of the model.

Chen et al. used a non-intrusive ROM to predict the space-time pressure solutions [58]. This method is called the Black-box Stencil Interpolation Method (BSIM). Stencil locality is a key assumption made in the work, which enables the significant reduction of the input parameter space and thus, deduce the global solution from local mass conservation principles. Based on this assumption, most controls and uncertainties are typically assumed to have local support. Even though POD helps dimensionality reduction, the computational complexity is still proportional to the dimension of the high dimensional problem; thereby, DEIM is used to reduce the dimension even more. DEIM can be thought of as a sampling method in this approach. Assuming the saturation solver is available, using saturation at each time step they solved the pressure. Two approaches have been used in this study. In the first one, a Laplacian model has been used to represent the simplified physics. The pressure difference between the reservoir simulator outputs (true physics) and simplified Laplacian model (simplified physics), at each gridblock and time is calculated. An artificial neural network (ANN) model is built with the stencil model properties as input and the aforementioned difference as output. POD and DEIM are used for calculating reduced pressure values at new time steps. Having pressure values, simplified physics, and the ANN model the pressure at each time step and the gridblock is calculated. This methodology was performed on a synthetic 2-D, 2-phase model with only 900
gridblocks. In another approach, POD-DEIM is used only as a sampling method, and the pressure values are estimated using an ANN model. The most important deficiencies of this work are that the PDE has been simplified too unrealistically, considering only permeability and pressure in the formulation. Besides, the saturation is assumed to be known in all the time steps and is used as an input for calculating pressure, which is not convincing.

Fedutenko et al. used a RBF based proxy model to predict the cumulative oil production, water injection, and steam to oil ratio for the entire field in steam-assisted gravity drainage (SAGD) process of a synthetic reservoir model, having three pairs of wells [59]. The reason behind using RBF has been explained to be the necessity for less training full-fledged simulation runs. However, the number of simulation runs needed for this relatively small model is between 50 and 200. A close match is not obtained unless 200 training runs are used for proxy model development, which defeats the purpose of building proxies.

Guyagular et al. used neural network based proxy models as a part of a hybrid optimization process to pinpoint the best well positions in the Gulf of Mexico waterflooding project [60]. The proposed hybrid optimization technique was based on the genetic algorithm (GA), polytope algorithm, Kriging algorithm, and neural networks. The net present value of the waterflooding project was used as the objective function. Their work focused on improving the efficiency of the optimization itself, rather than using speedier evaluation of the objective function. The points evaluated during the progression of the GA were used to estimate the unvisited points in the search space by calibrating and using a proxy. Kriging based proxy resulted in better results compared to NN. The authors related this to the generalization incapability of the built network.

A workflow for screening/optimization of Cyclic Pressure Pulsing in naturally fractured reservoirs was studied in a work performed by Artun et al [61]. Different proxy models were developed using ANN for two different gas injection types (CO₂ and N₂), and injection scheme. The reservoir characteristics, fluid properties, along with the design parameters were changed to build the proxy models with an output of production rate, cumulative oil production, number of cycles, and duration of production. The genetic algorithm has been used on the proxy model to find the optimized scenario. Although
the method is called universal, it has been developed with a single well model having a single layer.

In a similar work, Parada and Ertekin implemented a multilayer cascade feedforward back propagation artificial network algorithm to develop proxy models, which help screening for Improved Oil Recovery (IOR) methods [62]. The rock and fluid properties along with the design parameters for 200 reservoir simulation runs were used as an input to the proxy model to obtain the expected total recovery and project lifetime. These values help in selecting the best fitting IOR scenario for the provided reservoir. The models should be 4, 5, 7, or 9 spots and symmetric in terms of all their properties. The screening tool is designed based on homogenous reservoir models.

As explained above, many of the proxy models require modification to the reservoir simulation tools; these models are in the intrusive model category. Nonetheless, commercial simulation tools do not provide open access to the mathematical model, which makes this category of proxy models inapplicable. In many cases, the available mathematical models used in reservoir simulators may not be sufficient to explain all the physics involved in the phenomena going on under surface, e.g. production from shale plays, EOR process, fracturing, etc. Data driven models give the privilege of using the data to convey what is occurring without going into the detailed mathematical formulation of the physics of that phenomenon.

2.3. **Surrogate Reservoir Model (SRM)**

Proxy models that mimic reservoir models can theoretically offer an opportunity to reduce the number of simulation runs and allow timely reservoir-management decisions to be made. The methods commonly used in our industry for developing proxy models were explained previously. The conventional approach in our industry is mainly based on geostatistics. One such method that is used quite often is response surfaces [63, 64, 65]. Reduced models are another category of proxy models. Reduced models are approximations of full three dimensional numerical simulation models that essentially approach an analytical model for tractability [66]. Not enough coverage of the parameter space and the requirement for many training simulation runs are the main drawbacks for most of these methods. The best techniques for more efficiently creating a surrogate in a parameter space would decrease the number of simulation runs used while preserving accuracy.
During analyses that are categorized as response surface, hundreds of combinations of the input parameters (realizations) are created and used as input to the full field model. Upon completion of these runs, hundreds of outputs (production from wells in the field or dynamic properties distribution in the reservoir) are generated. These outputs are then used to generate a surface of all the possible responses that can result from the predetermined realizations [67]. Selection of the realizations is usually made in a way that maximizes the coverage of the anticipated range of input parameters while requiring the minimum number of simulation runs. Usually techniques of Design of Experiments [68], such as the Latin Hyper Cube [69], are used to optimize this process. Nevertheless, most of the serious studies require hundreds of runs to provide meaningful coverage. Furthermore, once the hundreds or thousands of required simulation runs are made and the response surface is generated, the input parameters no longer play any role in the process. In other words, the approach mentioned in the system theory will not be in effect upon completion of the simulation runs.

Neural networks are also used as the main algorithm to generate a proxy of the field [70]. Usage of neural networks has gained considerable popularity in the last decade in the oil and gas industry. Their first use was powerful interpolation techniques and recently research has shifted toward cognitive science. Tolerance to not exactly certain (i.e. noisy) data and ability to respond to complex result sets are very useful for many fuzzy or not exactly defined systems. In the petroleum industry, recent research concentrates on using neural networks as fairly accurate replacements for simulations.

Surrogate reservoir models (SRM) are approximations of the high fidelity reservoir models that are capable of accurately mimicking the behavior of the full field models as a function of changes to all the involved input parameters (reservoir characteristics and operational constraints) in seconds. SRM integrates reservoir engineering and reservoir modeling with machine learning and data mining. The approach used during the development of the SRM fits more appropriately within the approach summarized in the system theory [71] (as shown in Figure 3), rather than the approach commonly used in our industry that is essentially based on geostatistics. Considering the full field reservoir model within the context of the system theory, different reservoir parameters, such as permeability and porosity, among others, from the geologic model are input to the system, while the production from the well would be the system output (system being the full-field reservoir model).
Mohaghegh presented SRM for the first time in 2006, for solving the problem of time-consuming runs for an uncertainty analysis of a giant oil field with 165 horizontal wells in the Middle East [72]. The reservoir simulation model included about one million gridblocks and took 10 hours to run using a cluster of 12, 3.2 GHz processors. SRM was used as an objective function for a Monte Carlo Simulation to build thousands of simulation runs in a very short time compared to numerical simulators. A single SRM run would take only a fraction of a second. The developed SRM was used to analyze the entire solution space (all possible combinations of production scenarios) while quantifying the uncertainties associated with the static model that was used in the flow simulator. After hundreds of thousands of SRM runs, the results were analyzed and recommendations on which wells should be selected for the rate relaxation.

Mohaghegh describes SRM as an "ensemble of multiple, interconnected neuro-fuzzy systems that are trained to adaptively learn the fluid flow behavior from a multi-well, multilayer reservoir simulation model, such that they can reproduce results similar to those of the reservoir simulation model (with high accuracy) in real-time" [73]. Since 2006, SRM as a rapid replica of a numerical simulation model with quite high accuracy has been applied and validated in different case studies [74, 75, 76, 77, 78, 79, 80]. SRM can be categorized in well-based [75, 76, 77, 79, 81] or grid-based types [74, 78, 82] depending on the objective or the output of the model. In a well-based SRM the objective is to mimic the reservoir response at well location in terms of production (or injection). The grid-based SRM, on the other hand, makes it possible to simulate any dynamic properties such as pressure, phase saturations, or the composition of fluid components at any specific time or location (gridblock) of the reservoir.
CHAPTER 3. CO$_2$ - STORAGE AND EOR APPLICATION

3.1. GEOLOGIC CARBON SEQUESTRATION

The geologic carbon sequestration is a quite recent concept and is comprised of capture, transportation, and geologic storage of CO$_2$. Due to carbon capture and geologic sequestration’s (CCGS) contribution to the sustainable processes of climate change mitigation, it is becoming more strategic.

Although the CCGS is a technology with the high potential of reduction of greenhouse gases effects, the main obstacles for its operation are the high expenditures of capital (CAPEX) and operation (OPEX) [83]. In order to aggregate value and minimize costs, there are certain actions possible: geologic sequestration as a clean development methodology (CDM) letting the carbon credits concession, development of lower cost and higher efficiency CO$_2$ capture technologies and performing CO$_2$-EOR. The latter one may be more attractive, not only due to the additional oil production, but also because the oil and gas fields having longer time eras of trapping, thus having better public acceptance [84, 85].

Even though there are over 125 CO$_2$-EOR field projects in operation, fewer than 10% use CO$_2$ currents captured from anthropogenic releases; most use CO$_2$ natural sources or buy CO$_2$ in the local commerce. In order that these field projects may be cogitated as carbon sequestration, it is required that they “close the cycle,” meaning they collect
the CO₂ that starts being produced together with the gas and oil. This CO₂ should be captured and re-injected in the reservoir [83].

In this day and age, there are reasonable numbers of field projects for carbon sequestration, mostly pilot and demonstration scales. Commercial strategies for large-scale field projects are being studied and developed by oil companies. There are numerous technological and political defies, that need special consideration:

- Capture technologies with lower cost (70% of the geologic sequestration cost is in the capture process),
- More compact systems, especially for offshore implementations,
- Improve the process eco-efficiency,
- Novel transport modals,
- Lower-priced sub-surface monitoring techniques,
- Definite legislations,
- Raise the public perception,
- Increase the international cooperation, etc. [83]

Positive aspects of a CO₂ injection process are good oil responses, gas injectivity, and gas production within planned bounds. At the same time, early CO₂ breakthrough has been considered as the key issue in the projects performed, which compromises gas processing facilities [86].

3.1.1. Capture

The carbon capture technologies may be used in several processes: post-combustion, pre-combustion, oxi-combustion, and industrial gases separation. These technologies should allow not only natural gas specification, but also separation and re-injection of CO₂ in the reservoir, for EOR purposes. These technologies primarily include absorption (chemical and physical), adsorption, membranes separation, and cryogenic distillation [87]. To make these techniques efficient both technically and economically is the main challenge.

For low to moderate CO₂ partial pressures, chemical absorption is implemented. Since CO₂ is an acid gas, chemical absorption of CO₂ from gaseous streams, such as flue
gases, hinges on acid base neutralization reactions using basic solvents. Most common commercial solvents for neutralizing CO$_2$ are alkanolamines such as monoethanolamine (MEA), diethanolamine (DEA), and methyldiethanolamine (MDEA). Ammonia and hot potassium carbonate are other chemical solvents in use. Weakly bonded intermediate compounds are formed when CO$_2$ reacts with chemical solvent. These bonds can be broken by applying heat, and the original solvent can be regenerated for reusing and producing a CO$_2$ stream. This process is called separation of CO$_2$ by chemical absorption [88].

Pure CO$_2$ can be sequestered; however, the gas generated by power plants is usually diluted with nitrogen as flue gas. Current capture technologies are not worthwhile when considered in the setting of CO$_2$ storage from power plants [86].

**3.1.2. TRANSPORTATION**

Pipelines and tanks are two methods for transporting CO$_2$ to injection sites. In the pipeline, CO$_2$ is kept at a pressure above its critical pressure that is 7.4 MPa. The operating range of pipeline pressure is usually between 8 and 17 MPa. CO$_2$ is compressed at pressures above 8 MPa to achieve a single-phase flow before transporting to prevent the problems associated with a two-phase flow [89]. Suitable operating pressure and temperature lies in between 8,619 KPa at 4°C and 15,300 KPa at 38°C. These limits are set by the ASME-ANSI 900# flange rating and ambient condition coupled with the phase behavior of CO$_2$ [90].

Using tanks and trucks or rails as an option for CO$_2$ transporting was previously considered to be more expensive than a pipeline. Nonetheless, as stated by a recent study, shipping CO$_2$ by customized LPG gas vessels is more flexible and less costly [91]. Particularly in the case of ocean disposal at distances more than about 300 km from shore, tankers carrying dry ice or supercritical liquid carbon dioxide are more cost-effective than pipelines, because pipelines need depressurization at regular intervals. The distance that the carbon dioxide is to be carried does not have a significant effect on the costs of transporting CO$_2$ by tanker [92].
3.1.3. Storage

After capturing and transportation of CO₂, it should be stored such that its emission into the atmosphere is not possible or at least greatly delayed. CO₂ storage can be done either in porous or non-porous media.

Nonporous Medium

Deep Ocean (Hydrates)

Oceans cover approximately 70% of the Earth’s surface. This makes them the largest sinks available for carbon dioxide. Ocean disposal is of interest for the countries that have coastal zones and access to ocean depths of more than 3,000 m since transportation costs rise with distance traveled [92].

The carbon disposal in the ocean carries out as follows. First, dry ice is discharged to the ocean surface from a ship. The liquid CO₂ is then injected at a depth of about 1,000 m via a pipe towed by a moving ship. This forms a rising droplet plume. A dense carbon dioxide seawater mixture is shaped at a depth of between 500 and 1,000 m, which forms sinking bottom gravity current. Introducing liquid carbon dioxide to a sea floor depression forms a stable “deep lake” at a depth of about 4000 m.

Salt Cavern

Even though salt caverns are not at the top of the sequestration sink list, they are appropriate for storing CO₂ permanently (more than 1,000 years) or temporarily (decades). Regardless of the cost and other potential environmental issues related to cavern mining, there are some benefits in storing into salt. Filling or emptying salt caverns is not restrained by the capabilities of porous media. More storage capacity is provided by storing CO₂ in shallow, cool salt caverns in a supercritical form than storing CO₂ in solution, as a free gas, or through adsorption onto coal or oil shale.

Porous Medium

Aquifers, depleted oil or gas reservoirs, and coal beds are considered as geological structures for CO₂ to be stored. Upon the injection of the CO₂ in the formation, three major steps are followed. One of these steps is the hydrodynamic trapping, through which the injected CO₂ dissolves and diffuses in oil and water and flows according to the existing pressure gradient in the porous medium. Solution trapping happens when
CO₂ reaches thermal equilibrium in all fluid phases (oil, water, and gas) depending on fluid, pressure and temperature conditions. The next step is mineral trapping, which occurs when the dissolved CO₂ reacts with the minerals within the formation and causes dissolution/precipitation reactions. Some parameters, such as size of the reservoir, effective porosity, the net fraction of the reservoir that can be filled, and the density of the CO₂ define the capacity of a reservoir to store CO₂ [93]. Different types of porous storing sites include aquifer (deep saline), depleted oil and gas reservoirs, and enhanced coal bed methane.

**Aquifer (Deep Saline)**

There is high potential to find a suitable aquifer with a large capacity or close to a CO₂ source, since they are the most equally spread sink over the world. An aquifer is suitable for underground storage of gases or liquids since it is a reservoir with porosity, permeability, and a sealing cap rock [94]. Large volumes of CO₂ can be injected into aquifers without raising the pressure to a high extent. CO₂ will dissolve in the brine once injected, and result in brine/CO₂ mixture, which is denser than the brine alone. It is estimated that hundreds to thousands of years will be necessary to trap all the CO₂ [95].

**Depleted Oil and Gas Reservoirs**

Presently injection into depleted oil reservoirs is the sequestration policy. This process is an intricate issue covering a broad scope of scientific, technological, economic, safety, and regularity issue [96]. There are several factors that make oil and gas reservoirs attractive targets for carbon sequestration, including:

- Assuming increased pressure does not change the integrity of the seals and create any new pathways to the surface, structural traps can be counted on to contain carbon dioxide as they have enclosed the oil or gas over geological timescales. The reservoir will not be environmentally degraded by the CO₂, as the reservoir has already contained hydrocarbons.
- There is a significant description of the geologic structure and physical properties of most oil and gas fields.
- Computer models have been built and they can be utilized in order to predict the displacement behavior and trapping of CO₂ for EOR [92].
• Some production wells may be converted to gas injection wells, while the others may be used to monitor the behavior of the CO$_2$ within the reservoir. A CO$_2$ sequestration plan can be implemented to increase oil production, if the field is still producing [93].

**Enhanced Coal Bed Methane (ECBM)**
Coal has more affinity toward CO$_2$ than methane, so it desorbs the methane as free gas for recovery by adsorbing and storing CO$_2$. However, injectivity in methane is low due to the low value of permeability and it will require both reservoir treatment and a larger number of wells [89]. CH$_4$ or CO$_2$ are adsorbed on the surfaces of coal particles at high pressure. According to the adsorption curve hysteresis, once CO$_2$ is adsorbed, a large amount of it will go on as adsorbed even if the pressure is reduced in the future. CO$_2$ can be used for enhancing CH$_4$ recovery since the flow in coal beds occur primarily in the fracture network, diffusing into matrix blocks and replacing adsorbed CH$_4$ [97].

**Parameters Affecting CO$_2$ Storage Process**
There are some parameters that play a key role in the storage process. Some of these factors can be named as heterogeneity of the reservoir, relative permeability, natural fracture, and reservoir fluids.

**Reservoir Heterogeneity**
Reservoir heterogeneity, especially in the vertical direction, has a strong influence on the gas/oil displacement process. CO$_2$ performance can be highly affected by this factor. Formations with higher vertical permeability, such as naturally fractured reservoirs, are influenced by cross-flow perpendicular to the bulk flow direction [98]. Due to heterogeneity, cross-flow is frequently presented in water alternate gas (WAG) projects, and may increase the vertical sweep, but in general, the oil recovery is low due to the gravity segregation and decreased flood velocity in the reservoir, as shown in Figure 4. As CO$_2$ flows preferentially toward the top portion of thick, high permeable zone, injected water may flow preferentially toward the lower portion of the zone.
Relative Permeabilities

Relative permeability includes rock wetting characteristics, heterogeneity of reservoir fluids, and rock and fluid saturations, and is an important petrophysical parameter. Relative permeability usually changes during alternate water/CO₂ injection, so that water injectivity is considerably reduced after the first gas injection cycle owing to the influence of CO₂ on water relative permeability. Understanding the relative permeability curves used in the reservoir simulator is very important in order to comprehend and forecast the storage process [99].

Laboratory experiments have shown hysteresis effects in the water relative permeability between the drainage and imbibition curves. Irreducible water saturations after drainage cycles were 15 to 20% higher than the initial connate water saturation [100]. Hysteresis refers to the directional saturation phenomena demonstrated by many relative permeability and capillary pressure curves, when a given fluid phase saturation is increased or decreased [101]. This phenomenon is illustrated in Figure 5.
Natural Fractures
Fractures, fracture networks, and faults are among the structures that can influence permeability and therefore fluid flow within an aquifer or petroleum reservoir. Fractures generally result in an overall permeability increase within a given rock volume. The rock matrix can be drained better due to significant interaction between the fracture surface and the matrix. This matrix/fracture interaction could allow for a substantial increase in recoverable hydrocarbon reserves. On the contrary, mineralized fractures and deformation bands (i.e., small displacement faults characterized by tight cataclasis and/or pore reduction through compaction) are normally categorized by significant permeability reduction. Within a given rock volume containing mineralized fractures and/or deformation bands, there will be an overall permeability decrease and possible reservoir compartmentalization. Either mineralized or partially mineralized fractures could have the effect of reducing the total amount of recoverable reserves. Fractures introduce permeability anisotropy and heterogeneity by increasing or decreasing permeability in certain directions. From a production viewpoint, it is important that they be modeled accurately. Most regional fractures are sub-vertical, and are accordingly unlikely to be sampled in vertical boreholes [86].

Reservoir Fluids
The miscible process between the reservoir fluid and injected CO₂ is believed to be controlled by reservoir fluid composition controls. CO₂ is less dense and viscous than
reservoir fluids. Complete dissolution of injected CO$_2$ takes place in a scale of hundreds to thousands of years; this depends on the gas migration and fluids reaction [102].

As carbon dioxide is injected in the formation, it mobilizes oil, dissolves into brine, and promotes dissolution of carbonate cements [96]. Brine can become supersaturated with dissolved solids and when pressure drops as it advances through the reservoir, precipitates, such as gypsum, can form [103].

3.2. ENHANCED OIL RECOVERY (EOR)
Up to three distinct phases can be considered for crude oil development and production in oil reservoirs including primary, secondary, and tertiary (or enhanced) recovery. During primary recovery, the main drive is the natural pressure of the reservoir or gravity, combined with artificial lift techniques (such as pumps). But typically during this phase, only about 10 percent of original oil in place in the reservoir is produced. During the second phase, some techniques including water and gas injection, are used to improve the field’s productive life by displacing oil and driving it to a production wellbore. The resultant recovery factor is about 20 to 40 percent of the original oil in place [104].

However, after the first two phases some attempts for producing 30 to 60 percent or more of the reservoir’s original oil in place is done by the means of several tertiary, or enhanced oil recovery (EOR) techniques. Three major categories of EOR that have been commercially proved successful to varying degrees include:

- Gas injection uses gases such as natural gas, nitrogen, or carbon dioxide. Injected gas improves oil flow rate by expanding in the reservoir and pushing additional oil to a production wellbore, or dissolving in oil. Gas injection accounts for nearly 50 percent of EOR production in the United States.
- Thermal recovery, which involves the introduction of heat by conventional steam floods with injectors and producers, drilled in tight spacing patterns. These techniques are usually applied to relatively shallow reservoirs with high viscous heavy oil. Thermal lessens the viscosity and makes the flow easier and accounts for over 50 percent of U.S. Enhanced Oil Recovery (EOR) production, primarily in California.
• Chemical injection, which can involve the use of long-chained molecules called polymers or detergent-like surfactants to increase the efficiency of waterflood or to help lower the interfacial tension between the oil and water or change the wettability of the rock. Less than one percent of U.S. EOR production falls in this category [105, 106].

**CO₂-EOR**

CO₂ has been described as a dislocating fluid for enhanced oil recovery (EOR) in lab bench since 1916. At the beginning, this method was not considered as a feasible technique due to the small CO₂ volumes available and the process high expenses. In the 1950s, the industry considered the potential of miscible methods [107].

There are advantages to CO₂ flooding over the other EOR methods, such as polymer recovery, steam flooding, etc. To begin with, efficiency of displacement is high in miscible cases and the process is supported by a solution gas drive mechanism. As well, it is favorable over a broader range of crude oils than other hydrocarbon injection methods. The miscibility can be regenerated if it is lost, which can be listed as another advantage of this method [108].

CO₂ is highly soluble in hydrocarbon fluids and depending on the PVT conditions, a large amount of CO₂ will dissolve in a barrel of oil. For instance, over 700 scf of CO₂ will dissolve in 1 bbl of a 17° API oil, giving a 10-30% increase in its volume. The swelling of oil results in less residual oil being left behind in the reservoir. Viscosity reduction also happens as CO₂ dissolves into the crude oil. The CO₂ saturated oil has 10 to 100 times less viscosity than the original oil. This makes the mobilization of the oil in the reservoir easier in addition to increasing the efficiency of the sweeping effect [108].

Oil displacement deeply depends on factors, which are connected to the phase behavior of CO₂ – crude oil mixtures. CO₂ is compressed to a supercritical state with the intention of avoiding its separation into gas and liquid phases during transportation and injection processes. At normal atmospheric conditions, CO₂ is a thermodynamically very stable gas with a density of 50% greater than air. As displayed in Figure 6, the phase diagram of pure CO₂ shows a critical temperature of 31°C and a critical pressure of 7.4 MPa (1074 psi). Below this temperature and/or pressure the CO₂ is either in a
liquid or vapor phase, and above the critical values the pure CO$_2$ is in a supercritical state. At these conditions, CO$_2$ behaves still like a gas by filling all the available volume, but has a “liquid” density that rises, contingent on pressure and temperature, from 200 to 900 kg/m$^3$ (Figure 7). This is related with the phase change from gas to supercritical fluid. Consequently, the space CO$_2$ occupies is much less in the subsurface than at the surface. At 0°C and 1 atm, 1 tonne of CO$_2$ occupies 509 m$^3$. One tonne of CO$_2$ at a density of 700 kg/m$^3$ occupies 1.43 m$^3$ or less than 6 m$^3$ of a rock with 30% porosity, if 80% of the water in the pore space could be displaced [109]. The viscosity of CO$_2$ is a strong function of pressure and temperature (Figure 8). As pressure increases at a constant reservoir temperature, gas viscosity increases. So CO$_2$ has considerably stronger sweep efficiency [110].

![Phase diagram of CO$_2$](111)

**Figure 6.** Phase diagram of CO$_2$ [111]
Figure 7. Variation of CO$_2$ density as a function of temperature and pressure [110]

Figure 8. Variation of CO$_2$ viscosity as a function of temperature and pressure [110]
CO₂ has two features that make it a good choice for tertiary recovery in oil reservoirs: it is miscible with crude oil, and, compared to other similarly miscible fluids, it is less expensive. Miscibility is the ability of fluids to mix in all proportions. There are other solvents that can be used for miscible flooding too. Nonetheless, since they are the products of crude oil refinery, they can be relatively expensive and not economically beneficial. The same goes for enriched gas with heavier hydrocarbons, which is a valuable commodity. On the other hand, CO₂ underground deposits can be inexpensive to comparatively inexpensive to acquire. The human produced CO₂ can be used as another source apart from the naturally occurring CO₂ [112].

There are some mechanisms that play an important role in CO₂ EOR, regardless of the injection type or scheme. These mechanisms have been itemized below.

- Oil viscosity reduction: As the crude oil becomes saturated with CO₂ at increasing pressures, its viscosity declines markedly. This is even more noticeable in the more viscous fluids and causes the increase in mobility ratio.
- Oil swelling: The high solubility of CO₂ in hydrocarbon oil causes swelling.
- Increase in oil density: CO₂ has an influence on the water or brine in the reservoir when displacement processes are happening. Due to the expansion of water when CO₂ is injected, the densities of the oil and water become closer to each other, which diminish the chances for gravity segregation of these fluids and the ensuing overriding of the CO₂-water mixture.
- Extraction and vaporization of oil: CO₂ can vaporize and extract portions of crude oil. This takes place at low temperatures where CO₂ is a liquid, as well as at higher temperatures above the critical temperature.
- Miscibility effects: CO₂ is highly soluble in water and in hydrocarbon oils.
- CO₂ reduces the interfacial tension between water and oil.
- Increase in the injectivity (acidic effect): the acidic effect of CO₂ on the rock has been shown to increase the injectivity of water by direct action on carbonate portions of the rock and through stabilizing action on clays in the rock.

The mechanisms, which have been listed above, are essentially important depending on whether the CO₂ displacement is miscible or immiscible. For instance, the vaporization of crude oil, development of miscibility, and reduction of interfacial tension are very
important with the miscible CO$_2$ process, whereas decrease in crude oil viscosity and its swelling have more significant effects in the immiscible CO$_2$ displacement [113].

**US CO$_2$ EOR Demographics**

Although it has been over 40 years that CO$_2$ has been used for EOR purposes, it was not until recently that its sequestration potential got highlighted and investigated. Currently, approximately 37 percent of all EOR being performed in the United States is CO$_2$ EOR projects [114].

Out of over 48 million metric tons (tonnes) per year of CO$_2$, currently used for EOR purposes, near 25 percent (12 million tonnes) is anthropogenic in origin that is produced by human activities, for instance oil refining or fertilizer manufacturing. Extraction from naturally occurring deposits comprises the rest [114]. The production from EOR projects was a total of 650,000 barrels of oil per day in 2006 [115], which is almost 13 percent of the total U.S. production. CO$_2$-EOR alone was reported to make the production in the U.S. grow to 240,000 barrels per day in 2008, according to the *Oil & Gas Journal's* biennial survey [72].

A small group of major oil companies, namely Amerada Hess, Amoco, ARCO, Chevron, Exxon, Mobil, Shell, and Texaco were undertaking almost all the CO$_2$ injection projects prior to the early 1990s. This was changed, due to a technology transfer program led by the DOE's National Energy Technology Laboratory in the 1990s and an alteration in major company investment overseas. Currently independent producers take over the roster of CO$_2$ EOR operators. One of these independent companies is Kinder Morgan, the second largest producer of oil in Texas and one of the nation’s largest owners and transporters of CO$_2$, which is presently in charge of SACROC EOR project and has more than tripled SACROC production since attaining a majority interest in the unit in 2000 [116].
Table 1. Major U.S. CO₂ Operators (OGJ Biennial EOR Survey 2008) [116]

<table>
<thead>
<tr>
<th>Company</th>
<th>Miscible Projects</th>
<th>Locations</th>
<th>Incremental Production (MBO/D*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occidental</td>
<td>29</td>
<td>TX, NM</td>
<td>90.2</td>
</tr>
<tr>
<td>Hess</td>
<td>6</td>
<td>TX</td>
<td>25.3</td>
</tr>
<tr>
<td>Kinder Morgan</td>
<td>1</td>
<td>TX</td>
<td>24.2</td>
</tr>
<tr>
<td>Chevron</td>
<td>4</td>
<td>CO, TX, NM</td>
<td>21.3</td>
</tr>
<tr>
<td>Denbury Resources</td>
<td>13</td>
<td>MS, LA</td>
<td>17.8</td>
</tr>
<tr>
<td>Merit Energy</td>
<td>7</td>
<td>WY, OK</td>
<td>13.6</td>
</tr>
<tr>
<td>ExxonMobil</td>
<td>2</td>
<td>TX, UT</td>
<td>11.7</td>
</tr>
<tr>
<td>Anadarko</td>
<td>4</td>
<td>WY</td>
<td>9.0</td>
</tr>
<tr>
<td>Whiting Petroleum</td>
<td>3</td>
<td>TX, OK</td>
<td>6.9</td>
</tr>
<tr>
<td>ConocoPhillips</td>
<td>2</td>
<td>TX, NM</td>
<td>5.5</td>
</tr>
<tr>
<td>12 other independents</td>
<td>28</td>
<td>TX, OK, UT, KS, MI</td>
<td>14.9</td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td></td>
<td>240.4</td>
</tr>
</tbody>
</table>

* thousand barrels of oil per day

Figure 9 presents the yearly average production by CO₂ EOR in USA.

Figure 9. Yearly evolution of crude oil production (b/d) by CO₂ EOR in the USA [116]
**CO₂ Flood Design Variations**

Depending on different parameters several designs can be considered for a CO₂ flooding process. The work in this dissertation is based on WAG process, which has been explained below along with some other schemes [99].

**Continuous CO₂ Injection**

In this type of injection, a predetermined CO₂ slug volume is injected continuously with no other interjected fluid or chase fluid. The candidates for this approach are usually gravity-drainage or non-waterfloodable reservoirs directly following their primary depletion. Occasionally a different gas is used for driving the CO₂ through the reservoir; for instance, in vertical downward CO₂ displacement project a lighter gas may tail CO₂ to maximize gravity tonguing and channeling.

**Continuous CO₂ Chased With Water**

The only difference between this design with the continuous CO₂ injection is that in this method the continuous CO₂ slug is followed by chase water that immiscibly displaces the mobile miscible CO₂ oil bank. The candidates for this approach normally are reservoirs with low heterogeneity. Homogenous reservoirs help minimize gas production rates because they retain more of the injected gas. If this plan is applied to relatively thick layers in homogeneous floods, then an unfavorable gravity tongue may occur.

**Conventional Alternating CO₂ and Water Chased With Water**

A predetermined slug of CO₂ is injected in cycles in which equal volumes of gas and water alternate (known as water alternating gas or WAG) at a constant water gas ratio, or WAG ratio. After the total CO₂ slug has been injected, a chase of continuous water is started. The WAG process assists in improving the sweeping efficiency by mitigating the tendency for the lower viscosity CO₂ to finger its way ahead of the displaced oil. Once the CO₂ breakthrough happens any gas injected afterwards will follow that path, decreasing the total efficiency of the injected fluids to sweep the oil from the reservoir rock [112]. Highly stratified heterogeneous reservoirs are the best candidates for applying this method since it minimizes gas cycling in high velocity (k/φ) layers by reducing the fraction of injected CO₂ entering those layers. This design improves areal and vertical sweep efficiencies.
WAG ratios in the CO₂ flood processes have ranged from 0.5 to 4.0 volumes of water per volume of CO₂ at reservoir conditions. The sizes of the alternate slugs range from 0.1 percent to 2 percent of the reservoir pore volume. Cumulative injected CO₂ volumes vary, but usually range between 15 and 30 percent of the hydrocarbon pore volume of the reservoir. Historically, the efforts in CO₂ EOR are focused on minimizing the amount of injected CO₂ per incremental barrel of oil recovered [112].

**Tapered Alternating CO₂ and Water (Sometimes Chased With Water)**

In the tapered WAG process, CO₂ slugs are injected interchangeably with ever-increasing water cycle intervals in tapered or unequal cycle volume until the total CO₂ slug volume has been injected. Sometimes the taper is followed by a chase of continuous water. The aim of a tapered method is to reduce the CO₂ utilization factor. CO₂ utilization factor is defined as the ultimate surface CO₂ injection volume Mscf per ultimate stock tank barrel of incremental oil recovered. This is done by intermittently adjusting the WAG ratio and/or slug sizes in flood patterns where utilization is extravagant.

This scheme can manage short-term flood profitability by reducing the purchase cost for CO₂ and need for gas recycling. It also may reduce near-term revenues. The near-term oil rate is as low as a consequence of reduced CO₂ throughput in the faster, mobile, oil bearing layers. Yet the oil rate usually faces an increase later in the flood due to the improved areal sweep accomplished by decreased total mobility ratios. In other words, when the CO₂ does not sweep through as fast, it does a better job of recovering all the oil.

**Alternating CO₂ and Water Chased With Gas**

The difference between this scheme and conventional WAG is in the chased fluid, which is a volume of less expensive gas after the total CO₂ slug volume has been injected. In some cases, water injection may follow the chase gas or be alternated with it. The main reason for using the chase gas is to preserve miscible displacement of the trailing edge of the CO₂ slug while reducing total CO₂ requirements. In water-sensitive lithologies, gas is chosen as the chase fluid to sidestep using water.
Figure 10. Cross-section illustrating how carbon dioxide and water can be used to flush residual oil from a subsurface rock formation between wells [116].

**Reservoir Screening**

Not all reservoirs are amenable to CO₂ flooding. There should be a screening procedure for different technological aspects before the process of CO₂ injection can be implemented. Through the screening process, reservoirs that are not suitable can be effectively eliminated. The screening parameters can be divided into oil properties or reservoir characteristics. The API gravity, the viscosity, and the composition of the oil are categorized in oil properties and the oil saturation, the formation type, net thickness and depth, average permeability, and the temperature are considered under the reservoir characteristics.
The viscosity of the fluid has to be less than 12 cp in case of miscible flooding and between 100-1,000 cp in case of immiscible flooding. In the case of miscible flooding, the lower the oil viscosity the better it is. Less viscous crude causes mobility to increase and assists in achieving higher sweep efficiency. Oil API has to be greater than 22 for immiscible and higher than 30 for miscible floods. A higher API gravity specifies that the oil include lighter carbon molecules. This will help CO₂ to become easily miscible. Moreover, the already light oils will have easily reduced viscosities. A high percentage of C5-C12 hydrocarbons is needed for performing CO₂ flood.

The oil saturation should be higher than 20%. The higher this value the better it is for CO₂ flooding. Current field projects have an average of 55%. CO₂ flooding is well applicable in a carbonate or sandstone formation. The formation depth should be greater than 2,300 ft for immiscible and 3,000 ft for miscible flooding. The original bottom-hole flowing pressure has to be 1,000 psi for immiscible and higher than 1,500 for miscible CO₂ flooding. The porosity is to some extent accounted for in oil saturation, but permeability has no effect when it comes to the screening criteria [117].

Reservoir temperature should be between 32fC and 121fC, and the ratio of the CO₂ pressure to MMP should be higher than 0.95 [118].

**CO₂-EOR Optimization**

Even though CO₂-EOR has been practiced in the industry for more than 30 years, its design and operation is always dependent on the oil price. Especially when the oil price is low all the attempts are with the objective of minimizing the cost and maintaining the revenue.

Coordinating recovery efficiency and project finances is a challenge for gravity stable injection design. Zhou et al., have suggested that the existence of high permeability vertical conduits causes the CO₂ flow rates in the reservoir to be significantly higher than the critical gravity-stable flow rate resulting in high gas recycling. They have proposed a tapered WAG design to improve operational flexibility and recovery efficiency in a mature field under their study [119].

Continuous injection design has a simpler injection system and higher injectivity than WAG design. Improved initial oil responses can be obtained at early time, which helps the project economics at an early stage of floods. Nonetheless, when CO₂ breaks
through, the produced GOR increases quickly, causing high CO₂ recycling and facility max out [120]. Continuous injection design offers limited flexibility for reservoir management and operations. It tends to have higher cumulative net CO₂ utilization factor and requires more CO₂ purchase than WAG design for the same oil produced. TWAG (Tapered WAG) is designed to inject a large slug of CO₂ at the beginning to gain quick oil response and high injectivity and to start WAG injection when GOR reaches a designed level. The WAG ratio can be varied based on the maturity of the fields, which provides an effective tool for CO₂ utilization and operational optimization.

More than 90% of the projects implemented in the U.S. are WAG or TWAG because of the improved performance control and operational flexibility. Cyclic (Huff n’ Puff) CO₂ injection has been used since the early stage of CO₂ EOR in the 1960s and 1970s for injectivity and production pilots. It was proven to be incompetent for highly connected reservoirs [121] but satisfactory for developing low permeability reservoirs [122].

How frequent the injecting fluid in a WAG process should be alternated can vary significantly from a few days to several months depending on the oil reservoir, injection and production volumes, well location, and residual oil. A useful rule-of-thumb is based upon the time at which the volume of breakthrough gas or water-cut abruptly rises compared with the volume of produced oil [123].

<table>
<thead>
<tr>
<th>Field Case</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WAG (TWAG)</strong></td>
<td>Improved control on mobility and CO₂ utilisation factors</td>
<td>Extensive field surveillance program and intensive operational adjustment</td>
</tr>
<tr>
<td><strong>Continuous Injection</strong></td>
<td>Higher recovery at the early stage</td>
<td>Field operation is limited by possibly higher CO₂ purchase and recycling at the late stage of the flood; Poor conformance control</td>
</tr>
<tr>
<td><strong>Cyclic Injection</strong></td>
<td>Tight formations; lower investment</td>
<td>Loss of productivity due to well’s shut-in for CO₂ to dissipate and dissolve.</td>
</tr>
</tbody>
</table>

Table 2. The major injection schemes developed in the last 30 years.
3.3. CO\textsubscript{2} EOR-Sequestration

It is believed that one of the key sources of global change is carbon dioxide release because of the combustion of fossil fuels. Fossil fuels provide over 85 percent of the world’s energy needs and this demand is growing [129]. Many analysts believe that the best way to resolve this growth in the use of fossil fuel with restrictions on carbon dioxide emission is carbon sequestration.

Injection of CO\textsubscript{2} into oil reservoirs can improve oil recovery and at one go alleviate the issue of increased CO\textsubscript{2} concentration in the atmosphere. The fact that in most reservoir conditions, CO\textsubscript{2} is in a supercritical state makes it a great choice as an EOR fluid, due to its high solvency power to extract hydrocarbon components and displace oil miscibly in that state.

Performing the CO\textsubscript{2} -EOR process in depleted oil reservoirs is both economically and environmentally beneficial. It can work as an immediate option for reducing CO\textsubscript{2} emissions into the atmosphere combined with enhancing the oil recovery. The aim in the CO\textsubscript{2} -EOR process is minimizing the required CO\textsubscript{2} for producing a barrel of oil. While for a sequestration purpose, maximizing the amount of CO\textsubscript{2} stored is the objective [130]. An optimization should be performed, such that the maximum amount of CO\textsubscript{2} is stored in the reservoir while incremental oil recovery is maximized.

3.3.1. Potential Benefits

CO\textsubscript{2}-EOR was initially tried in 1972 in Scurry County, Texas. Since then, this approach has been used successfully throughout the Permian Basin of West Texas and eastern New Mexico, and is now being pursued to a limited extent in Kansas, Mississippi, Wyoming, Oklahoma, Colorado, Utah, Montana, Alaska, and Pennsylvania.

Many factors including different wettabilities, gravitational effects, and reservoir heterogeneities, among others, affect the premature breakthrough, which in turn strongly influences the recovery factors. Furthermore, the operations onshore and offshore have their provisions, being offshore more complicated, with higher costs and advanced technologic requirements [74].
The potential benefits of CO₂ EOR are not limited to increasing the yield of depleted or high viscosity fields. It can also be considered as a method for sequestering carbon dioxide that would normally be released into the atmosphere. If CO₂ is injected at a pressure below MMP, then it can enhance pumping by swelling the oil and reducing its viscosity.

Several reasons make CO₂ EOR a favorable method of sequestration. First, given the integrity of geologic structures holding the oil and gas originally maintained, they should also permanently contain the injected CO₂. Since the operating depleted oil and gas fields are widespread, the likelihood of them being close to a CO₂ source increases. As a final point, carbon sequestration from CO₂ EOR projects can generate counterbalances occasioning in trades in the evolving greenhouse gas market [114].

According to the U.S. Department of Energy (DOE), depleted oil and gas wells in the United States and Canada can potentially sequester over 82 billion tonnes of carbon in total [131].

Table 3: CO₂ Utilization and Potential in EOR Projects (Source of U.S. data: National Energy Technology Laboratory, Carbon Sequestration Atlas, 2007)

<table>
<thead>
<tr>
<th>United States (2006)</th>
<th>(million tonnes/yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon Dioxide use for EOR</td>
<td>48</td>
</tr>
<tr>
<td>• Naturally occurring</td>
<td>36</td>
</tr>
<tr>
<td>• Anthropogenic</td>
<td>12</td>
</tr>
<tr>
<td>Estimated CO₂ sequestered by EOR operations</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Worldwide</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential CO₂ EOR sequestration</td>
<td>130 billion tonnes</td>
</tr>
<tr>
<td>Total CO₂ accumulated in atmosphere</td>
<td>3-4 billion tonnes/yr</td>
</tr>
</tbody>
</table>
3.3.2. Uncertainties in CO$_2$-EOR and Sequestration Flood Design Parameters

Both of the objectives for coupled EOR and sequestration projects (maximum recovery and maximum storage) depend on a large number of parameters and on the strategy used to flood the oil reservoir. Different mobility control techniques, such as WAG injection, can increase sweep efficiency and oil recovery but may jeopardize the storage objective. In addition, parameters such as type of injection and production wells and their operating constraint can affect the results significantly.

For instance, gravity stable displacement of CO$_2$ (which can be accomplished by a relatively small injection and production rates in some dipping reservoirs) can greatly enhance oil recovery and the amount of stored CO$_2$. But, in most cases, only reservoirs with very high permeability and substantial dip are good candidates for these types of floods.

Well spacing is another factor affecting a CO$_2$ flooding project. Optimum well spacing is required to meet both goals (maximum oil recovery and CO$_2$ storage). Higher well spacing delays the CO$_2$ breakthrough time in the production wells, which leads to an increase in CO$_2$ storage. On the contrary, it causes a lag in the oil production peak.

Reservoir properties can also have an influence on the best design as well as project objectives. For instance, vertical to horizontal permeability ratio has an important effect on cross flow and gravity override. One of the most important reservoir characteristics is reservoir heterogeneity. The ultimate amount of stored CO$_2$ can be limited in presence of high permeability paths. Gas breakthrough in the producers will bring about early recycling and lower sweep efficiency. It is apparent that produced CO$_2$ can be re-injected, but at the very high cost of the recycling plant in the injection site. Accordingly, reservoir heterogeneity needs to be considered closely in the design of any coupled CO$_2$-EOR and sequestration project [132].
CHAPTER 4. RESERVOIR MODELING

4.1. FIELD BACKGROUND

The Kelly-Snyder field, discovered in 1948, is one of the major oil reservoirs in the US. The original estimate of its original oil in place (OOIP) was approximately 2.73 billion bbls. The primary production mechanism was indicated as merely solution gas drive, based on the early performance history of the field, which would probably result in an ultimate recovery of less than 20 percent of the OOIP. The SACROC Unit was formed in 1953, and in Sept. 1954, a massive pressure maintenance program was started. Based on this plan, water was injected into a center-line row of wells along the longitudinal axis of the reservoir [133].

In 1968, a technical committee, studying potential substitutes, suggested that a water-driven slug of CO$_2$ be used to miscibly displace the oil in the non-water-invaded portion of the reservoir. It was also recommended that a pattern injection program be developed in this area, to apply the slug process and enhance ultimate oil recovery. CO$_2$ injection was begun in early 1972. An inverted nine-spot miscible flood program, consisting of injecting CO$_2$, driven by water was decided to be the most effective and economical alternative method for improving recovery in the SACROC unit based on the investigations. Comparison of the predicted ultimate recovery of the original water injection program and the recommended scheme, showed that the new pattern would result in about 230 million bbls increases [133].
The Scurry Area Canyon Reef Operations Committee Unit (SACROC), within the Horseshoe Atoll, has undergone CO$_2$ injection since 1972 and is the oldest continuously operated CO$_2$-EOR operation in the United States. The amount of injected CO$_2$ until 2005 was about 93 million tones (93,673,236,443 kg), out of which about 38 million tones (38,040,501,080 kg) had been produced. Accordingly, a simple mass balance suggests that about 55 million tones (55,632,735,360 kg) of CO$_2$ have been accumulated in the site [135]. Initially, Chevron Corp. operated the unit. Later, Pennzoil took over before its upstream operations were spun off into PennzEnergy Co. and merged with Devon Energy Corp. In June 2000, Kinder Morgan CO$_2$ bought Devon’s interest in SACROC. At the time of Kinder Morgan’s purchase, the production, which once was more than 200,000 bbls/day, had declined to 8,200 bbl/day. SACROC continues to be flooded by the current owner/operator, Kinder Morgan CO$_2$. 

Figure 11. The Permian Basin outlined in green covers parts of western Texas and southeastern New Mexico. SACROC oilfield, identified in red, resides on the northeastern edge of the Basin. The map also identifies both natural and anthropogenic sources of CO$_2$ as well as the CO$_2$ pipelines in the region [134].
This project was the first large scale CO₂ EOR and a lot of lessons were learned during it. Chevron did not have access to sufficient CO₂ in order to perform flooding throughout the entire field during a reasonable time frame. The parts of the reservoir, which performed poorly during the waterflood recovery and had the highest percent of OOIP, were prioritized. It took years of experience in multiple projects to reach an industry consensus that the areas that have performed the best under the waterflood process will also have the best response to CO₂ flood. Variation in well-to-well connectivity was found to be the main reason for difference in performance. The higher amount of remaining oil in the formation in some areas was of little use if the injected fluid never contacted it [136].

4.1.1. Reservoir

Among four contiguous fields along the 35 × 5 mile Canyon Reef formation, the Kelly-Snyder field, in Scurry County (west Texas) is the most important unitized field. The formation is a Pennsylvanian age limestone occurring at an average depth of 6,700 ft. It has a northeast-southwest trending massive reef buildup with thinner, gently sloping flanks. Formation thickness varies from an average of 800 ft on the crest of the structure to less than 50 ft on the flanks, and averages 213 ft overall. The SACROC unit is just less than 50,000 acres and constitutes about 98 percent of the Kelly-Snyder field. The unitized reservoir is generally bounded on the east and west by porosity development and on the northeast and southwest by offsetting units. An oil-water contact, although poorly defined in some areas, occurs throughout most of the Unit area. Only very limited water influx is detected during pressure depletion, indicating a relatively small aquifer in the immediate area [133].

Reservoir oil was initially undersaturated at an original pressure of 3,122 psig, and had a solution gas content of slightly under 1,000 scf/STB and a bubble point pressure of 1,805 psig. The reservoir oil is rich in intermediate components (31.5 mole percent C₂ – C₄) [133].

Original oil in place (OOIP) within the SACROC Unit area was estimated at 2.73 billion STB. Cumulative production through 1971 was 536 million STB, or approximately 19.7 percent of OOIP [133].
### Table 4. Basic Reservoir Data - SACROC [133]

<table>
<thead>
<tr>
<th>Physical Features of the Reservoir</th>
<th>Approximate Depth, ft.</th>
<th>6,700</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate WOC (subsea), ft.</td>
<td>-4,500</td>
<td></td>
</tr>
<tr>
<td>Avg. Gross Thickness, ft.</td>
<td>213</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Properties of Reservoir Rock (Gross Reef)</th>
<th>Avg. Porosity, %</th>
<th>7.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Permeability, md.</td>
<td>19.4</td>
<td></td>
</tr>
<tr>
<td>Avg. Interstitial S_w, % pore space</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Avg. Residual S_o, % pore space</td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>

| Reservoir Temperature (-4300 ft.), ͦF | 130 |
| Initial Reservoir Pressure (-4300 ft.), psig | 3,122 |
| Water Injection Commenced              | Sep. 21, 1954 |
| Unit MER (Aug 1972), B/D               | 243,978 |
| Unit Surface Area, Acres               | 49,900 |
| Total Wells in Unit                    | 1,256 |
| Unit Hydrocarbon Pore Volume           |         |
| Acre-ft.                               | 517,434 |
| Barrels                               | 4,014,253,000 |
| Approximate Gross Acre-ft. in Unit     | 10,668,000 |
| Original Stock Tank Oil in Place (B_o=1.4720), bbl | 2,727,000,000 |

### Table 5. SACROC Reservoir Fluid Properties [133].

<table>
<thead>
<tr>
<th>Fluid Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bubble Point Pressure (at 130 ͦF), psia</td>
<td>1,820</td>
</tr>
<tr>
<td>Reservoir Fluid Viscosity at 1,820 psi and 130 ͦF, cp</td>
<td>0.38</td>
</tr>
<tr>
<td>Reservoir Fluid Density at 1,820 psi and 130 ͦF, lb/cft</td>
<td>41.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flash Separation Data, First Stage Separator Conditions:</th>
<th>25 psia and 95 ͦF</th>
<th>31 psia and 75 ͦF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution GOR, scf/STB</td>
<td>990</td>
<td>910</td>
</tr>
<tr>
<td>Stock Tank Oil Gravity , ͦAPI</td>
<td>41</td>
<td>42.7</td>
</tr>
<tr>
<td>Casing head Gas Gravity</td>
<td>1.087</td>
<td>1.030</td>
</tr>
<tr>
<td>Formation Volume Factor at 3,137 psia, bbl/STB</td>
<td>1.528</td>
<td>1.472</td>
</tr>
<tr>
<td>Formation Volume Factor at 1,820 psia, bbl/STB</td>
<td>1.557</td>
<td>1.5</td>
</tr>
</tbody>
</table>
4.1.2. **Geologic Description**

SACROC is located in the southeastern segment of the Horseshoe Atoll within the Midland basin of west Texas (Figure 12). SACROC covers an area of 356 km² with a length of 40 km and a width of 3–15 km, within the Horseshoe Atoll [137]. Geologically, massive amounts of bedded bioclastic limestone and thin shale beds representing the Strawn, Canyon, and Cisco formations of the Pennsylvanian, and the Wolfcamp Series of the Lower Permian comprise the carbonate reef complex at SACROC [137]. Among these formations, most of the CO₂ for EOR is injected into Cisco and Canyon formations, which were deposited during the Pennsylvanian age.

![Figure 12. SACROC Unit at the Horseshoe Atoll in west Texas and structural contours map of top of carbonate reef modified from Stafford. Contours are in m scale [138].](image)

The Strawn Formation in the carbonate reef complex started to form in the early Desmoinesian period, while the basin was on the equator [139]. Carbonate sedimentation of the Canyon and Cisco formations continued during the Missourian and Virgilian periods. Drastic influx of fine-grained clastics ended accumulation of carbonate sediments on the SACROC during the Wolfcampian period. Although the Canyon and Cisco formations are mostly composed of limestone, minor amounts of anhydrite, sand, chert, and shale are present locally [140]. Based on an analysis done by Carey et al. on core samples from wells 49-5 and 49-6 in the SACROC field, the limestone is mostly calcite with minor ankerite, quartz, and thin clay lenses [141].
The Wolfcamp shale of the lower Permian, acts as a seal above the Canyon and Cisco Formations [140]. According to X-ray diffraction analysis results, the shale is mostly illite/smectite and quartz with minor feldspar, carbonate, and pyrite. Carey et al. concluded that the CO$_2$ had not interacted with shale, based on mineralogical analysis [141].

A lot of investigations have been performed by petroleum geologists to understand the spatial (aerial) distribution of carbonates in the SACROC unit, along with the variation of carbonates with depth. Based on these analyses, this unit has been divided into Lower Canyon, Middle Canyon, Upper Canyon, and Cisco Groups through detailed analyses of cores, logging data, and biostratigraphy (Fusulinid age). These investigations suggested that the Cisco group unconformably overlies the Canyon Group [142].

Both porosity and permeability show a large variability due to changes in the depositional environment. The values of SACROC porosity and permeability reported in the previous studies can be seen in Figure 14.
Figure 14. Well logs representing the SACROC northern platform and summary of previous studies estimating carbonate rock properties in Cisco and Canyon Groups [142].
4.1.3. Performance History

The Standard Oil Company of Texas discovered the SACROC Unit in November 1948. The OOIP in this unit was estimated to be approximately 2.73 billion stock tank barrels (STB) in the Canyon Reef limestone formation [133].

The rapid development of the field is obvious from the oil production rate buildup during the first 2 years. Over 1,200 producing wells with 81 operators were drilled in the Canyon Reef complex, from 1948 to 1951 [143]. A study of this early performance showed that the solution gas drive was the primary reservoir mechanism and no effective water drive existed. During the first 5 years, only less than 5 percent of the original oil in place was produced; however, GOR increased and the average reservoir pressure was reduced by 50-percent (i.e. from 3122 psi to 1,560 psi). It became obvious that some form of pressure maintenance was needed to prevent very low oil recovery. Relying on only a solution gas drive would result in an ultimate recovery of only 19 percent of oil in place, based on performance predictions. The field was unitized in March 1953 and has since been operated as the SACROC unit. The unit name was appropriately selected to represent the Scurry Area Canyon Reef Operators’ Committee, whose efforts led to its formation and to the design of an optimum plan of operation [133].

Waterflood Evaluation

In September 1954, a full-scale program of pressure maintenance by water injection throughout the Unit area began in SACROC. The first 72 water injection wells, with an injection rate of 132,000 barrels of water per day (BWPD), were selected generally along the longitudinal crest of the structure. These wells in the thicker part of the reservoir were selected, bearing in mind the critical pressure and gas saturation conditions that existed and the obvious necessity for the speedy restoration of pressure describing a “center-line pattern” injection [133, 143].

As shown in Figure 15, before water injection began, only 1% of the reservoir was above bubble point pressure (1805 psi). Within less than two years, 45% of the reservoir rose above bubble point pressure. Finally, 77% was above bubble point pressure after 17 years of water injection [133].
Figure 15. Contour and dotted lines, respectively, indicate the bottom hole pressure (psi) and the location of center-line water injection wells [133].
(a) Five months before starting water injection (September 1954). Only 1% of reservoir volume is above bubble point pressure (1805 psig=12.45 MPa).
(b) Seventeen years later after starting water injection (September 1970), 77% of the reservoir volume is above bubble point pressure.

Figure 16. Frontal advance of centerline water injection, SACROC unit, Kelly Snyder field [133].
Enhanced Oil Recovery Methods

Enriched Gas Miscible Process
This process did not look attractive due to the high cost of the injected material and risks involved with gravity override and viscous fingering. Thereby, the CO₂ slug process was selected.

Carbon Dioxide Process
Laboratory works and field studies [144, 145, 146] had demonstrated that, under favorable conditions, CO₂ miscibly displaces certain reservoir oils. Laboratory research also showed that CO₂ would successfully generate a miscible displacement of SACROC oil at bubble point conditions. Successive laboratory work pointed out that the injected CO₂ gas could be diluted with as much as 18 percent methane without seriously affecting the miscibility-generating capability of the injected gas mixture. This was a key point, since CO₂ in the volumes required was not available in the pure form, but only as a methane-diluted waste by-product from CO₂ removal plants designed to produce marketable hydrocarbon residue gas. The injection pattern of the inverted nine spots is designed to process approximately 49 percent of the original unit hydrocarbon pore volume (Figure 19) [133]. It was decided that the EOR process be performed based on different phases, due to supply rate limitation.

Figure 17. The steep production decline at the SACROC unit leveled and began to increase due to the successful implementation of miscible CO₂ floods. The blue line represents the best-producing field project that is being replicated (3).
Figure 18. SACROC Performance History
Figure 19. (A) SACROC unit at the Horseshoe Atoll in western Texas and CO₂ supply system from natural CO₂ reservoirs, (B) Magnified map of the SACROC Unit within the Horseshoe Atoll with indication of paleo-wind direction, (C) Well locations of SACROC Unit with the estimated water-flooding fronts at the end of water-flooding period in 1973 (Kane, 1979) [142].
4.2. **Analyzes of Field Data for Model Preparation**

Both porosity and permeability vary naturally as a result of complex geological processes such as structural deformation, sedimentary deposition, and diagenesis. The changeability of these rock properties significantly affects the subsurface migration of injected CO₂.

Although quantification of natural heterogeneity using diverse techniques has been discoursed in detail in the hydrology and petroleum literature [147, 148, 149, 150, 151], the application of these methods to CO₂-EOR and sequestration processes has not yet been widely performed. Thus, a high-resolution geocellular model describing reservoir heterogeneity was adapted to facilitate a detailed numerical modeling study of the CO₂-EOR process in this work. The applied high-resolution geocellular model in this study is a model consisting of a broad field of expertise in geological data (core data, well-log data, sedimentologic, and stratigraphic interpretation) and geophysical data (seismic attributes and rock physics data). Therefore, the high resolution geocellular model comprises detailed heterogeneity and helps us to better comprehend the subsurface migration of injected CO₂.

Porosity distributions implemented in this work were determined from both seismic survey and wire log analyses. Permeability distributions were calculated from seismically calibrated porosity values using empirical equations derived from rock fabric classification. The original high-resolution geocellular model consisted of over 9.4 million elements. Therefore, it was necessary to upscale it in order to acquire a reasonable number of elements for simulating CO₂-EOR process [138].

**Quantification of Natural Heterogeneity in SACROC Northern Platform**

A high-resolution geocellular model of the SACROC northern platform was constructed by the Texas Bureau of Economic Geology using porosity from wire-line logs and three-dimensional seismic data [138]. Inversion and time-depth conversion are two of the foremost methods of integrating three-dimensional seismic data into a model. Inversion transforms the reflection amplitude into acoustic impedance and time-depth conversion inverts amplitude and impedance data from time space to depth space.

**Estimation of Porosity from both Wire-line Logs and Seismic Survey**

Shallow water skeletal and oolitic limestone cycles formed the Canyon formation, whereas the Cisco formation encompasses mainly deep-water crinodal mounds and debris wedges [152]. The porosity data from 368 wire-line logs were available. Reservoir quality in the geological model created by BEG was better developed in the Canyon than the Cisco. According to this model, porosity is continuous and increases upward
from the mudstone base to the grain-rich top in the Canyon formation. Nonetheless, the reservoir lying in the Cisco formation is a highly heterogeneous formation, including low porosity zones compartmentalized within one to two well spaces. Seismic mapping and inversion can help to resolve the heterogeneity in the Cisco formation [152].

Figure 20. Spatial heterogeneity of porosity determined by wireline log [138].

The wire-line logs data were enough to deterministically map the porosity distribution in the Canyon interval. However, seismic and wire-line log porosity were combined to determine the porosity distribution in the heterogeneous Cisco interval (Figure 21). The porosity distribution significantly increases in this model compared to the one without considering the seismic data. This could not be achieved using only the wire-line log data.

Figure 21. Spatial heterogeneity of porosity determined by both wire-line log and seismic data [138].
Estimation of Permeability from Porosity

The prediction of permeability from porosity in heterogeneous carbonates is a challenging and complicated problem. The relation between permeability and porosity cannot be described by a simple correlation. Babadagli and Al-Salmi [153] provided an extensive review of existing correlations between porosity and permeability in carbonate rock. The relationship between porosity and permeability in the high-resolution geocellular model developed by the Texas Bureau of Economic Geology, which is used in this study, is based on the rock-fabric approach. As Lucia [154, 155] has discussed this approach, it is based on the idea that permeability in carbonate rock is controlled by pore size distribution related to rock fabric. To determine the relationship between rock fabric and petrophysical parameters, Lucia [154, 155] categorized pore space into two divisions: inter-particle porosity and vuggy porosity. Inter-particle porosity comprises inter-grain and inter-crystal pore space. Vuggy pore space is partitioned into separate and touching vugs. Separate vugs are connected through the inter-particle pore space (grain molds) and touching vugs form an interconnected pore system independent of the inter-particle pore space (caverns and fracture).

Lucia [154, 155] suggested that permeability could be related to rock fabrics, including inter-particle porosity, geologic descriptions of particle size and sorting, and proposed that three classes could be defined for rock-fabrics. Inter-particle size from 100 to over 500 µm is in Class I and is dominant in grainstones, dolostones, and large crystalline dolostones. Class II comprises inter-particle size from 20 to 100 µm and is dominant in grain-dominated packstones, fine and medium crystalline, grain-dominated dolopackstones, medium crystalline, and mud-dominated dolostones. Finally, Class III is dominant in mud-dominated limestones and fine crystalline, mud-dominated dolostones, and consists of inter-particle sizes from 5 to 20 µm.

The classification of carbonate rock by inter-particle size is generalized after presenting the rock-fabric number (λ), by applying reduced major axis transformation [155]. Lucia [155] and Jennings and Lucia [156] developed the global transforming equation, which determines permeability from the rock-fabric number and porosity:

\[
\log(k) = (9.7982 + 12.0838\log(\lambda)) + (8.6711 + 8.2965\log(\lambda))\phi
\]

*Equation (57)*

where, λ, k, and φ are, rock-fabric number, permeability, and inter-particle porosity, respectively. Lucia and Kerans [157] used this equation for estimating the permeability
distribution in the SACROC northern platform. Using porosity from wire-line logs and a stratigraphically defined rock-fabric number, permeability is approximated from global transformation (Figure 22).

The process starts with calculating the average apparent rock-fabric number from core data for each stratigraphic layer. A permeability profile is then calculated for each well using apparent rock-fabric number from stratigraphy and total porosity from logs as input into global permeability transform. Thereby, permeability is interpolated between wells, constrained by seismic stratigraphy.

Figure 22. Permeability can be estimated from wireline logs using the global transformation and rock fabric number [158].
Figure 23. Porosity -Permeability correlation in SACROCO Northern platform based on the core data [158].

Table 6. Permeability estimation using porosity and rock-fabric number [157]

<table>
<thead>
<tr>
<th>Formation</th>
<th>Sequence</th>
<th>Rock Fabric Number ($\lambda$)</th>
<th>Transforming Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisco</td>
<td>Late</td>
<td>$k = 2.1625 \times 10^6 \times \phi^{3.8844}$</td>
<td>Highest portion of Cisco (Implication of karsting)</td>
</tr>
<tr>
<td></td>
<td>Late</td>
<td>$k = 1.031 \times 10^7 \times \phi^{6.7592}$</td>
<td>Late Cisco (wide variety of rock fabrics)</td>
</tr>
<tr>
<td></td>
<td>Early</td>
<td>$k = 2.69 \times 10^6 \times \phi^{6.3584}$</td>
<td>Early Cisco (Characterized by fusulinid/crinoidal/peloid grain dominated packstones, grainstones, and wackestones)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$k = 97628 \times \phi^{5.3696}$</td>
<td>Early Canyon: (Characterized by moldic ooid grainstone, grain-dominated packstone, and mud-dominated fabrics)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$k = 38520 \times \phi^{5.0923}$</td>
<td>Late Canyon: (Characterized by crinoidal/fusulinid/peloid grain-dominated packstones, and mud-dominated fabrics having vuggy porosity)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>Below reservoir</td>
</tr>
<tr>
<td>Canyon</td>
<td></td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.75</td>
<td></td>
</tr>
</tbody>
</table>


4.3. **Geo-Cellular Model**

The model used in this work represents the Cisco and Canyon groups in the SACROC northern platform. The model is approximately 4 Km wide and 10 Km long. Figure 24 shows the location of the wells in the SACROC field and the northern platform. The software used for developing the geocellular model is Petrel\(^1\).

![Image of wells in SACROC fields and northern platform top map based on available data.](image)

*Figure 24. The locations of wells in SACROC fields and northern platform top map based on available data.*

The top of the geo-cellular model shows the upper configuration of the Cisco Group, which is approximately 1,200 m below sea level, and the bottom of the model shows the lower configuration of the Canyon group, which is at a depth of almost 1,400 m below sea level. The thick and high porous zone represents the upper Canyon group. The Cisco group is located above the upper Canyon group and middle/lower Canyon groups are below the upper Canyon group. The middle and lower Canyon groups comprise several high porosity zones separated by low porosity carbonate mud. The distribution of porosity and permeability, along with the elevation depth of the top of the formation, has been illustrated in Figure 26.

\(^1\) Petrel E&P Software Platform - Schlumberger
Figure 25. The area of 3-D model of SACROC-Northern platform [158].

Figure 26. Top, Porosity and Permeability distribution in SACROC-Northern platform (Axis Scale is in m).
Porosity values range from 1 to 30 percent in the northern platform with an average of 8 percent. Since the formation is carbonate, the permeability changes from very low values close to 0 mD to high values up to 4 D.

**Figure 27.** SACROC Northern Platform High Resolution Geological Model - Left: Porosity Distribution Histogram - Right: Permeability Distribution Histogram

### 4.4. UPSCALING

One goal of this study is building a meaningful numerical simulation model based on the high-resolution data provided by the Texas Bureau of Economic Geology. The high-resolution geocellular model incorporating heterogeneity comprises 9,450,623 (149 × 287 × 221) elements for describing the Cisco and Canyon Groups.

The results of this study are going to be used by Department of Energy (DOE-NETL) in software, which takes the input in the (100 × 100) format. The top section of the high resolution model has been used for the works performed in this study. This section consists of 3,138,200 (149 × 100 × 221) fine scale gridblocks (Figure 28).
As explained above, our purpose is to build a model having 100 gridblocks in each X and Y directions. So the model should go through an areal upscaling by re-dimensioning the gridblocks. On the other hand, the geo-cellular model has a very high vertical resolution (Figure 29). Although the fine scale geocellular model delivers a quantified characterization of natural heterogeneity in the SACROC northern platform, due to computational limitations, the model has to be upscaled to develop a multi-phase and -species reactive transport model. The upscaling by nature results in less resolution of heterogeneity; however, attempts were made to preserve as much of the heterogeneity as possible.
There are different upscaling techniques especially for permeability. These techniques can be divided into local and global techniques depending on whether intrinsic permeability is an explicit function of cell conductivities or not [159]. In local techniques, it is assumed that average permeability is inherent to the cell permeability in denser grids. Therefore, average permeability is generally expressed as an explicit function of the fine grid permeability. Nonlocal techniques assume that that upscaled permeability, not only depends on the fine grid permeability values, but also the boundary conditions. Generally, a solution based on Laplace’s equation falls within this category. We adapted a local technique for upscaling the high-resolution geocellular model of SACROC.

The areal upscaling is performed by changing the dimension of each gridblock from \((100 \times 100 \, ft)\) to \((142 \times 100 \, ft)\). This will result in having 100 gridblocks in the X direction. Also using this technique we make sure we have no flow boundaries on the east and west parts of the reservoir, which will be explained more in the numerical simulation model development section. The high-resolution geocellular model has 221 layers. This excessive resolution should be decreased in a methodical approach, which preserves as much geological information as possible.

As it can be seen in Figure 30, there are five distinct geological zones in the northern platform. Using the wells shown in this cross section as a guide, the corresponding horizons were used and the model was upscaled from 221 to 5 layers. Each geological zone has in turn been subdivided to multiple simulation layers.
Figure 30. Geology zones in SACROC field-Northern platform.

Figure 31. Defining geological zones and corresponding horizons and simulation layers.
The volume-weighted algorithm was used and the arithmetic averaging was performed for upscaling the porosity values. The volume weighted method averages the values of the source cells contributing to the target cell and weights by volume. This will preserve the overall volume between the two grids. For permeability upscaling, the cell-count arithmetic averaging was used. The cell-count algorithm averages contributing source cell values with equal weight to each cell (Figure 32).

Figure 32. SACROC Low Resolution Geological Model-Section for This Study (100 * 110 *36 Grids) 
(Axis Scale is in m) - Left: Porosity Distribution Map and Histogram - Right: Permeability Distribution Map and Histogram.
4.5. **Numerical Reservoir Model Development**

The built geocellular model was transformed to a fluid flow simulation environment, along with the properties maps and wells data. The reservoir model was developed using a commercial reservoir simulator (CMG’s GEM). CMG’s GEM (Generalized Equation of State Model) simulator is a multi-dimensional, finite-difference, isothermal compositional simulator that can simulate three-phase (oil, water, gas) and multicomponent fluids [160].

4.5.1. **Initial and Boundary Condition**

For evaluating the CO₂-EOR process a coarse-scale geological model was developed as mentioned in the previous section. Characterization of the porosity and permeability was described formerly. Both porosity and permeability were predicted from a combination of data from the seismic survey and wireline log interpretations and upscaled.

The reservoir is initially under-saturated with a water oil contact level of 4,500 ft. The initial reservoir pressure at the depth of 4,300 ft is 3,122 psi and the water saturation below the water oil contact is 36% [133].

In a previous study, Han interpreted the Wolfcamp shale above the Cisco and Canyon formations as an effective seal, based on the analyses of both water chemistry and δ¹³C data [138]. In addition, based on the experiment by Carey et al., carbonate in the Wolfcamp shale appears to derive from primarily diagenetic processes and its isotope compositions show typical marine-originated δ¹³C values [141]. Thereby, the upper boundary is set as a no flow boundary. The eastern, western, and northern boundaries are treated as no-flow boundaries, because the Wolfcamp Shale Formation meets these boundaries. The carbonate reef complex, the Cisco and Canyon formations, is prism-shaped (Figure 13). The bottom boundary is also designated as a no flow boundary because the Strawn formation below the Cisco and Canyon formations is a low permeability unit [138].

The southern boundary is connected to the middle part of the reservoir, so assigning a no flow boundary to this part might not be the best idea. In a similar case, Han has used a constant pressure boundary [138]. In order to be more realistic, an extended section was considered while selecting the geological model section for this study, i.e. the selected portion of the northern platform has (149 × 110 × 221) gridblocks. The aforementioned upscaling process was performed on this model. This margin of ten grid cells helps to preserve the properties of the flow boundary and its influence on the
blocks located in the lower part of the reservoir. This extended section will be removed while performing the surrogate reservoir modeling.

The oil phase is regarded as a mixture of 11 different gas components. The initial oil composition used is from Dicharry et al. [133], summarized in Table 7.

<table>
<thead>
<tr>
<th>Oil Composition</th>
<th>Mole</th>
<th>Molecular Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO₂</td>
<td>0.0032</td>
<td>44.01</td>
</tr>
<tr>
<td>N₂</td>
<td>0.0083</td>
<td>28.01</td>
</tr>
<tr>
<td>C₁(Methane)</td>
<td>0.2865</td>
<td>16.04</td>
</tr>
<tr>
<td>C₂(Ethane)</td>
<td>0.1129</td>
<td>30.07</td>
</tr>
<tr>
<td>C₃(Propane)</td>
<td>0.1239</td>
<td>44.10</td>
</tr>
<tr>
<td>I-C₄(I-Butane)</td>
<td>0.0136</td>
<td>58.12</td>
</tr>
<tr>
<td>N-C₄(N-Butane)</td>
<td>0.0646</td>
<td>58.12</td>
</tr>
<tr>
<td>I-C₅(I-Pentane)</td>
<td>0.0198</td>
<td>72.15</td>
</tr>
<tr>
<td>N-C₅(N-Pentane)</td>
<td>0.0251</td>
<td>72.15</td>
</tr>
<tr>
<td>FC₆(Hexane)</td>
<td>0.0406</td>
<td>86.00</td>
</tr>
<tr>
<td>C₇+(Heptanes plus)</td>
<td>0.3015</td>
<td>197.4</td>
</tr>
</tbody>
</table>

The CO₂ density and fugacity coefficients are calculated from the Peng-Robinson equation of state [161]. Viscosity is estimated from Jossi et al. [162] and solubility is calculated with Henry's law, adjusted for the effects of salinity using scaled particle theory [163].

The water salinity reported in the literature for SACROC is 59,000 ppm [164] or 159,000 mg/L [138]. Table 8 enumerates the component reference, Henry's constant, and molar volume at infinite dilution calculated for SACROC fluid.

<table>
<thead>
<tr>
<th>Component</th>
<th>Henry's Constant (psi)</th>
<th>Ref. Pressure (psi)</th>
<th>Vinfinity (m³/kgmole)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO₂</td>
<td>100961</td>
<td>3122</td>
<td>3.5393849E-02</td>
</tr>
<tr>
<td>N₂</td>
<td>4.27984e+006</td>
<td>3122</td>
<td>3.2186653E-02</td>
</tr>
<tr>
<td>C₁(Methane)</td>
<td>2.35937e+006</td>
<td>3122</td>
<td>3.5547162E-02</td>
</tr>
<tr>
<td>C₂(Ethane)</td>
<td>2.52363e+006</td>
<td>3122</td>
<td>5.2213230E-02</td>
</tr>
<tr>
<td>C₃(Propane)</td>
<td>4.379e+006</td>
<td>3122</td>
<td>7.1621926E-02</td>
</tr>
<tr>
<td>I-C₄(I-Butane)</td>
<td>5.3053e+006</td>
<td>3122</td>
<td>9.1191677E-02</td>
</tr>
<tr>
<td>N-C₄(N-Butane)</td>
<td>5.30626e+006</td>
<td>3122</td>
<td>9.1211599E-02</td>
</tr>
<tr>
<td>I-C₅(I-Pentane)</td>
<td>1.55369e+007</td>
<td>3122</td>
<td>1.1027798E-01</td>
</tr>
<tr>
<td>N-C₅(N-Pentane)</td>
<td>1.58955e+007</td>
<td>3122</td>
<td>1.1275443E-01</td>
</tr>
<tr>
<td>FC₆(Hexane)</td>
<td>1.46959e+021</td>
<td>3122</td>
<td>1.2470150E-01</td>
</tr>
<tr>
<td>C₇+(Heptanes plus)</td>
<td>1.46959e+021</td>
<td>3122</td>
<td>3.1399264E-01</td>
</tr>
</tbody>
</table>
4.5.2. **Relative Permeability and Capillary Pressure**

Residual CO₂ saturation in the relative permeability among the phases determines if CO₂ will migrate or be trapped. Relative permeability in numerical simulation is applied through functional relationships that include hysteretic effects. In detail, the drainage curve is used to describe the process when a non-wetting phase (CO₂) displaces a wetting phase (brine or oil), while the imbibition curve is used to describe the process when a wetting phase (brine or oil) displaces a non-wetting phase (CO₂). Hysteresis indicates that the drainage and imbibition processes are not identical. In general, the relative permeability value of the non-wetting phase (CO₂) in the imbibition curve reaches zero even though its saturation in the imbibition curve is still greater than zero. When the relative permeability value of the non-wetting phase (CO₂) in the imbibition curve becomes zero, its saturation is defined as the residual CO₂ saturation. [138].

A relative permeability function (Figure 33), developed by extrapolating relative permeability data measured in similar carbonate rock [165] has been used. Bennion and Bachu [165] measured the relative permeability of supercritical CO₂ and brine in a 0.16 m core of low-permeability carbonate rock collected from Wabamun Lake, with experimental conditions of 41 °C and 22.4 MPa. Previous workers evaluated hysteresis effects using a modified Land equation, which had been adapted to account for residual CO₂ trapping mechanisms [166, 167]. To quantify the residual-trapped CO₂, hysteretic effects in SACROC, Han et al. implemented a modified Land equation [168] in the relative permeability curve [138] (Figure 33-a). Adapting the Land equation requires the determination of maximum residual saturation (S\(\text{grmax}\)). An empirical formulation determining S\(\text{grmax}\) from porosity was proposed by Holtz [169]. Using this empirical formulation, S\(\text{grmax}\) has been calculated as ~0.4 in their model. Han et al. have used this fitted relative permeability curve to define the relative permeability relationship between oil and CO₂ in SACROC field [138].

A relative permeability curve for oil measured from SACROC cores (Core128V) at 50 °C and 1.4 MPa [170] was used to calibrate the oil relative permeability curve in Figure 34-a. The relative permeability of brine was estimated with irreducible water saturation as 0.1 [138].

A capillary pressure function between supercritical CO₂ and liquid, measured in carbonate rock (Figure 33-b) by Bennion and Bachu [165], was adapted using Parker’s function [171].
Capillary pressure between oil and brine was measured by Rohan and Haggerty, [170] from SACROC cores (Core128V, 136V, and 191V) at 50°C and 1.4MPa. Han et al. calibrated capillary pressure functions by Parker et al. [171] using the Rohan and Haggerty [170] data (Figure 34-b).

4.5.3. **WELLS- CONSTRAINTS AND COMPLETIONS**

More than 300 wells have been used for CO₂ injection and production over the past 35 years in the SACROC northern platform. The total number of production and injection wells in the section under this study is 130. Some of these wells have been converted to injectors over time. Among these wells, 50 of them have been used as a water injector (either to perform water flooding or WAG) and 46 of them as CO₂ injectors. The total amount of oil production in the section under study is 185 MMSTB without considering the production from the hypothetical production wells which were added to the extended section. The water injection and CO₂ gas injection amount through this period is about 519 MMSTB and 199 BSCF, correspondingly.
Figure 33. (a) Fitted relative permeability curves with experimental data from Bennion and Bachu [165] (b) Fitted capillary curve with experimental data from Bennion and Bachu [172]. [138]
Figure 34. (a) Fitted relative permeability curves between brine and oil; measurements from Rohan and Haggerty [170]. (b) Fitted capillary pressure curve between brine and oil; measurements from Rohan and Haggerty [170] [138].
The production and injection rates were set as the constraints. Figure 38, Figure 39, and Figure 40 display the schematic of duration of injection and production of SACROC wells in the studied section. The minimum BHP in the production wells were set as 28 psi, and the max BHP in injection wells were fixed as 5,000 psi. Since the injectivity was less than what it should be in the reservoir model, not all the wells could inject at the assigned rate values. One of the reasons is that the built model is a closed model so the pressure in the reservoir increases due to injection and gets to the maximum allowable value of 5,000 psi, beyond which injection stops. On the other hand, production data from some of the wells in the modeled section are not available, which makes the depletion in the modeled reservoir and consequent injection amount slower. In order to resolve this issue some hypothetical production wells have been added in the extended section (lower 10 gridblocks) (Figure 37). These wells mimic the flow boundary. Based on reports in the literature during the first 5 years, only less than 5 percent of the original oil in place was produced; however, the average reservoir pressure was reduced by 50 percent, (from 3122 psi to 1,560 psi) [133]. The average reservoir pressure was monitored and its value after 5 years of production was used as a yardstick to build the base model.
The volume modifier and aquifer were tested as the potential solutions, but none of them worked as well as the hypothetical production wells.

**Figure 36.3-D Reservoir Simulation Model Showing the Reservoir Properties (Top, Porosity, and Permeability) and well location.**

**Figure 37. Production/Injection well locations and the extended area added to mimic the open flow boundary**
Figure 38. Oil Production History in 130 Wells of the SACROC-Northern Platform from 1948 to 2003
Figure 39. Water Injection History in 50 Wells of the SACROC-Northern Platform from 1948 to 2003.

Figure 40. CO₂ Injection History in 46 Wells of the SACROC-Northern Platform from 1948 to 2003.
4.6. **RESULTS**

The cumulative production and injection and the average reservoir pressure during the life of the reservoir, for the section under study has been demonstrated in Figure 41.

![Figure 41. Oil, water and gas production from the wells located in the SACROC section under study (Left), and water/gas injection and average pressure reservoir in this section (Right)](image)

The modeled process is a complex process involving a lot of changes due to water and miscible gas injection occurring concurrently with the oil, water, and gas production from multiple wells. This complexity shows itself while running the simulation. The compositional models by nature take much longer pertaining to the required calculations in comparison with the black oil models. This complexity rises as water injection and later miscible gas injection period starts. The time step size is altering to reach convergence, and gets as small as seconds. The total run time for running the model from the beginning of the production (1948) to the end of injection, based on available data (2004), is about 1 month on a machine with 24 GB of RAM and 3.6 GHz CPU.

The water saturation distribution at the beginning depends on the relative permeability end points and the initial conditions. The water saturation below the water oil contact (WOC) is 36%. As time passes, even before the water injection starts, the water saturation in lower parts of the reservoir starts increasing due to the pressure drop and influence of WOC.
The pressure and saturation maps were extracted from the results of the abovementioned model at 1/1/1974, which is right before the gas injection starts in this section. A new model was built using these data. In order to reduce the running time and resolve some simulation issues, a cut-off value for thickness and pore-volume was considered. The gridblocks with less than 2 ft thickness or 200 cu ft pore volume were set as the pinch-out blocks. All the properties in this model are kept the same as the last model; however, the changes in the starting date and pinch-out blocks reduce the running time required for computing the properties from 1948 to 1974 and resolve some convergence problems. On the other hand, since the results of the full run are not easily handled, cutting the model helps with tackling this problem. On top of the aforementioned alteration, some numerical control values should be set in the simulator. These values are listed in Table 9.

Table 9. Numerical Control Values set in the simulator.

<table>
<thead>
<tr>
<th>Keyword Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum number of time steps</td>
<td>100,000</td>
</tr>
<tr>
<td>Maximum time step size</td>
<td>3 days</td>
</tr>
<tr>
<td>Minimum time step size</td>
<td>1e-4 day</td>
</tr>
<tr>
<td>First time step size</td>
<td>0.001 day</td>
</tr>
<tr>
<td>Normal variation of pressure per time step</td>
<td>500 psi</td>
</tr>
<tr>
<td>Normal variation of saturation per time step</td>
<td>0.1</td>
</tr>
<tr>
<td>Normal variation of global composition per time step</td>
<td>0.1</td>
</tr>
<tr>
<td>Normal variation of aqueous component per time step</td>
<td>0.3</td>
</tr>
<tr>
<td>Maximum pressure change per time step</td>
<td>5000</td>
</tr>
<tr>
<td>Maximum saturation change per time step</td>
<td>0.8</td>
</tr>
<tr>
<td>Normal variation of global composition per time step</td>
<td>0.8</td>
</tr>
<tr>
<td>Maximum Newton iterations</td>
<td>30</td>
</tr>
<tr>
<td>Linear solver iterations</td>
<td>200</td>
</tr>
<tr>
<td>Linear solver orthogonalizations</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 42 and Figure 43 display the change in reservoir dynamic properties in the first and last layers over the time. The snapshots taken relate to the time before the production begins, right before the start of water injection (1954), before the initiation of gas injection (1974), and at the end of the gas injection period based on available data (2003).
Figure 42. Change in oil saturation, water saturation and pressure in layer 1 during the life of reservoir.

Figure 43. Change in oil saturation, water saturation and pressure in layer 36 during the life of reservoir.
The distribution of CO₂ mole fraction in water and oil in consort with its global distribution can be seen in Figure 44 and Figure 45.

**Figure 44.** Mole fraction of CO₂ in water, oil and global distribution before (1974) and at the end of gas injection history (2004) in the first layer.

**Figure 45.** Mole fraction of CO₂ in water, oil and global distribution before (1974) and at the end of gas injection history (2004) in the last layer.
4.7. Capturing the Physics

The procedure introduced in this work mainly focuses on the CO$_2$ miscible flooding. In this section, it will be shown that the reservoir simulation model that is built for this dissertation captures the physics of the phenomenon being modeled (Miscible CO$_2$-EOR), but first the miscible process and the factors affecting it will be expounded.

Miscible Flooding

After CO$_2$ is injected into an oil reservoir, as light hydrocarbons from the oil dissolve in the CO$_2$ and CO$_2$ dissolves in the oil, it becomes mutually soluble with the residual crude oil. When the CO$_2$ density is high (when it is compressed) and when the oil has low-density, this process takes place most readily. CO$_2$ and oil will no longer be miscible if the pressure is lower than a minimum value. With a decrease in CO$_2$ density, due to increase in temperature or increase in oil density, the minimum pressure needed to reach oil/CO$_2$ miscibility increases. That is why the pressure of a depleted oil reservoir should be considered by oil field operators while evaluating its suitability for CO$_2$ EOR. Low pressured reservoirs may need to be re-pressurized by injecting water.

When the injected CO$_2$ and residual oil are miscible, the physical forces holding the two phases apart (interfacial tension) effectually fades, which enables CO$_2$ to move the oil from the rock pores, pushing it toward a producing well. As CO$_2$ dissolves in the oil, it swells the oil and lessens its viscosity and aids to raise the effectiveness of the displacement process [112].

The high density CO$_2$ (compressed at high pressure) initially condenses into the oil, making it lighter and taking out methane from the oil bank. The lighter components of the oil at that point vaporize into the CO$_2$-rich phase, making it denser, and consequently more easily soluble in the oil. Mass transfer between the CO$_2$ and the oil continues until the two mixtures become almost alike in terms of fluid properties. Figure 46 illustrates the condensing/vaporizing mechanisms for miscibility [99].

![Figure 46. One-dimensional schematic showing how CO$_2$ becomes miscible with crude oil [99].](image)

In a reservoir with shallow depth, it is not possible to reach the pressure required for miscible flood; therefore, immiscible process is used in such reservoirs. Miscible CO$_2$
flooding has a higher appeal over immiscible flooding. It increases the recovery from 6% to 8% of the OOIP as compared to water; nonetheless, immiscible flooding is beneficial in reservoirs circumscribed by shallow depth [173].

Several conditions must be met in order to have a competitive miscible process in a given reservoir. These conditions include availability of an adequate volume of CO₂ at a rate and cost allowing satisfactory economics, attainable reservoir pressure compulsory for miscibility between the solvent and reservoir fluid, and adequately large and timely incremental oil recovery for project economics to withstand the added cost [174].

**Minimum Miscibility Pressure**

One of the most important factors in the CO₂ process is believed to be the minimum miscible pressure (MMP) at which a miscible displacement takes place. Experiments designed to determine crude oil-CO₂ MMP have been reported in a number of publications.

Yellig and Metcalfe [175] proposed a slim-tube method for determining CO₂ MMP for reservoir oils. The criteria for MMP are the maximum recovery at 1.2 pore volume of CO₂ injected in a series of tests, and the appearance of a transition zone in the sight glass at the exit of the slim tube. Based on experimental work, they have concluded that the MMP for any reservoir oil is primarily a function of temperature or bubble point pressure; oil composition has a little or no significant effect on the MMP.

Holm and Josendal [176] defined the MMP as the pressure at which more than 80% of oil-in-place is recovered at CO₂ breakthrough and more than 94% at the GOR of 40,000 SCF/bbl. They found, in contrast to Yellig and Metcalfe, that the composition of crude oil, especially the C₅+ fractions, has a substantial effect on the CO₂ MMP. The MMP determined from the above-described methods is based upon oil recovery efficiency. But recovery efficiency is affected not only by pressure and temperature, but also by a number of other factors, such as injection rate, the length and diameter of the slim tube, packing material, fingering, phase segregation, override, etc. Therefore, the MMP determined from the slim tube method does not represent the true MMP, but the pressure at which an optimum oil recovery may be achieved for a given set of experimental conditions. In a true physical definition, the MMP should be interpreted as the lowest pressure at which the phase boundary, between the displacing CO₂ and the displaced crude oil, disappears [177].
In the following figure, the four regions are:

1. **Region I** - at high pressures and low levels of CO₂, there exists a single phase. This phase is composed of crude oil under saturated with CO₂ and will hence be referred to as an oil-rich liquid (L₁).

2. **Region II** - at lower pressures and higher levels of CO₂, the oil-rich liquid phase (L₁) becomes saturated with CO₂ and is in equilibrium with CO₂ vapor that has become enriched through the extraction of lighter and intermediate hydrocarbons. This vapor phase will hence be referred to as a CO₂-rich vapor phase (V).

3. **Region III** - at moderate pressures and high levels of CO₂, there exists a 3-phase region. In this region, a second liquid phase has emerged. This second liquid phase is composed of CO₂ and light and intermediate hydrocarbons and is lighter in color and density than the oil-rich liquid phase (L₁). This light liquid will hence be referred to as a CO₂-rich liquid phase (L₂). In this region, an oil-rich liquid phase (L₁), a CO₂-rich liquid phase (L₂), and a CO₂-rich vapor phase (V) are all in equilibrium.

4. **Region IV** - at higher pressures and higher levels of CO₂, an oil-rich liquid phase (L₁) is in equilibrium with a CO₂-rich liquid phase (L₂).

*Figure 47. Typical Equilibrium of CO₂-Crude Oil System [177]*
According to Wang et al [177], the MMP increases with a higher temperature at an average rate of about 10.3 psi/oF. The miscibility is developed between the CO₂-rich liquid, L₂, and the CO₂ vapor. The L₂ is composed predominantly of CO₂ and a small portion of hydrocarbons extracted from crude oil; in other words, all the compositions of L₂ are similar regardless of which crude oil is used. The composition of crude oil, however, would affect the amount of CO₂-rich liquid. Crude oil rich in C₅-C₂₀ would generate a large volume of L₂ liquid.

An ample supply of CO₂ to L₂ is of major importance for maintaining a stable and high quality transition zone. For a low C₅-C₂₀ content crude oil, the amount of L₂ generated would be very small. Consequently, the transition zone formed would be unstable, and thus the displacement would be primarily of the immiscible type.

A study by Orr and Jensen [178] on the pressure composition phase diagram of CO₂ – crude system indicates that for a low temperature system (below about 120 °F), the extrapolated CO₂ vapor pressure plus a suitable safety margin of 200-300 psi, can be used as a rough estimate of the MMP. The CO₂ vapor pressure can be estimated by the equation:

\[ P = \left(101.3 \exp\left(-\frac{2015}{T}\right)\right) + 10.91 \]

Where P is the vapor pressure in KPa, and T is the temperature in °K.

**Miscibility Type and Ternary Diagram**

First-contact miscible solvents mix directly with reservoir oils in all proportions, and their mixtures always remain single phase.

The P-X diagram is used to estimate first contact miscibility pressure; therefore, the miscibility pressure estimated by this method will be higher than multi-contact (slim tube/rising bubble) methods.
Figure 48. P-X diagram used to estimate first contact miscibility pressure [179].

In the P-X diagram method, the location of the critical point is very important for determining the mechanism of recovery by gas injection.

Figure 49. P-X diagram based on the injection gas type [179].

Other solvents are not directly miscible with reservoir oils, but under appropriate conditions of pressure and solvent composition these solvents can achieve miscibility in-situ by the mass transfer of oil and solvent components through repeated contact with the reservoir oil. Miscibility achieved in this manner is termed multiple-contact or dynamic miscibility [180].

During the injection of gas into an oil reservoir, there will be a gradual exchange of components between the two fluids and their composition will become more alike. Eventually part of the gas phase and part of the oil phase will no longer be separated by an interface and will thus become miscible.

The phase exchange is governed by the equilibrium constant $K_i$ for each component, where $K_i = f(P, T, P_K)$ and $P_K$ is the convergence pressure [181].
Dynamic miscibility (Multiple Contact) can be more easily described if the complex mixture of HC is represented by a combination of three arbitrary components made up of groups of HC with similar thermodynamic properties:

1. The light components, methane $C_1$ and possibly $N_2$
2. The intermediate components which play a major role in thermodynamic equilibrium
3. The heavy components

Some researchers believe that dynamic miscibility with $CO_2$ does not require the presence of intermediate-molecular-weight HC in the reservoir fluid [181].

Having chosen the three components, we can draw an equilibrium triangle of which each apex represents one of the components.

At any given pressure and temperature, the point (M) may represent either a single phase or a diphase fluid according to its location inside the triangle.

For a given combination of P & T, the curves bounding the diphase region, the bubble point curve, and the dew point curve is unique for a specific mixture. Critical point (C) is the point at which the mixture is at the critical pressure and temperature [181].

![Figure 50. ternary diagram [181].](image)

For any given saturated liquid $A$, there is a corresponding saturated vapor $B$ with which it is in equilibrium, the line $AB$ is known as tie line [181].
Starting with a solvent (injected gas which might be mixed with lean gas) and oil, the ternary diagram construction consists of three steps [179]:

1. At each point solvent is added to the oil such that the solvent to oil molar ratio increases by a fixed amount until it becomes pure solvent. Flash calculations are performed at each point to determine the liquid/vapor split. If no two-phase region is detected, it is judged first contact miscible and calculation terminates; otherwise, it proceeds to steps 2 and 3.
2. Using the first point, A, in the two-phase region detected above, all liquid is removed. The remaining gas is combined with the original oil in the gas-oil ratio form B1. Flash calculation is performed and liquid removed. Again, it is contacted with original oil to form B2 and flash calculation performed. The procedure is repeated and simulates the vaporizing-gas drive process. This constructs the phase envelope marked as B.
3. Again using the first point, A, in the two-phase regions detected in step 1, all vapor is removed. The remaining liquid is combined with the original solvent in the solvent liquid ratio to form C1. Flash calculation is performed and vapor removed. The procedure is repeated until the oil cannot be enriched further. This simulates the condensing gas drive, and constructs the phase envelope marked C.
Mechanisms of recovery by gas injection during a miscible flooding can be categorized as below.

- One method uses a slug of liquefied petroleum gas (LPG) of about 5% PV, such as propane, followed by lean gas.
- Condensing drive (enriched gas drive)
  - Intermediate hydrocarbons from the injected gas condense into oil, and after multiple contacts the oil becomes miscible with the injected gas.
  - Miscibility occurs at the rear of the transition zone between the oil and injected gas.
  - For a condensing process, the critical point on the pressure-composition diagram will be to the right of the cricondenbar.

- High pressure vaporizing drive
  - A lean gas (high methane content) is injected at high pressure and vaporizes intermediates from the oil.
- After multiple contacts, the vapor phase becomes rich enough in intermediates to become miscible with the oil. Miscibility is generated by forward contacts of the vapor with fresh oil.
- For a vaporizing process, the critical point on the pressure-composition diagram will be to the left of the cricondenbar.

![Figure 54. P-X diagram in vaporizing drive [179].](image)

Miscible displacement by CO₂ is similar to that in the vaporizing gas drive mechanism. The only difference is that a wider range of components, C₂ to C₃₀, is extracted. As a result, the CO₂ flood process is applicable to a wider range of reservoirs at lower miscibility pressure values compared to those having a vaporizing gas drive. As miscibility is approached as a result of multiple contacts, both the oil phase and the CO₂ phase (containing intermediate oil components) can flow together because of the low interfacial tension. One of the requirements of the development of miscibility between the oil and CO₂ is the reservoir pressure [182].

The two regions are separated by a transition zone, where there is reduced IFT and a minimum tie-line length. The condensing/vaporizing process is characterized by a two-phase pseudo-ternary envelope which does not close and has an hour-glass shape. The pinching in the phase envelope indicates a region of reduced interfacial tension (IFT) [179].
The following table contains compositional analyses obtained by a gas chromatograph. It can be seen that C$_5$-C$_{20}$ components account for 68% by weight in SACROC crude. The oil tested was stock tank oil and solution gas was not recombined.

Investigations [176, 183] have shown that the presence of methane in reservoir oil can cause an appreciable decrease in oil recovery efficiency (i.e. higher pressures are required to achieve comparable recoveries in the displacement of reservoir oil than for stock tank oil). This behavior has been interpreted as follows.
1. CO₂ originally displaces the methane from the crude and this methane-rich bank moves ahead of the displacement front. [175, 176, 184]

2. After methane has been removed, CO₂ begins to extract heavier hydrocarbon components.

Apparently the increase of MMP of live oil is caused by the dilution of methane, which reduces the extracting capability of CO₂. Orr and Jensen [178] suggest that to estimate the MMP for live oil, 200-300 psi should be added above the MMP for the dead oil.

**Slim-Tube Simulation**

Slim-tube test is a laboratory test used to estimate the minimum miscibility pressure (MMP) or minimum miscibility concentration (MMC) of a given injection solvent and reservoir oil. The slim-tube is a long coiled tube filled with sand of a specific mesh size or similar porous media. The tube is saturated at the beginning of each test with reservoir fluid at a given temperature. Solvent injection is performed at several test pressures. Effluent production, density, and composition are measured as functions of the injected volume.

<table>
<thead>
<tr>
<th>Carbon No.</th>
<th>SACROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.67</td>
</tr>
<tr>
<td>6</td>
<td>4.41</td>
</tr>
<tr>
<td>7</td>
<td>8.56</td>
</tr>
<tr>
<td>8</td>
<td>7.11</td>
</tr>
<tr>
<td>9</td>
<td>5.96</td>
</tr>
<tr>
<td>10</td>
<td>5.65</td>
</tr>
<tr>
<td>11</td>
<td>4.66</td>
</tr>
<tr>
<td>12</td>
<td>4.07</td>
</tr>
<tr>
<td>13</td>
<td>4.17</td>
</tr>
<tr>
<td>14</td>
<td>4.21</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>3.05</td>
</tr>
<tr>
<td>17</td>
<td>2.35</td>
</tr>
<tr>
<td>18</td>
<td>3.06</td>
</tr>
<tr>
<td>19</td>
<td>2.01</td>
</tr>
<tr>
<td>20</td>
<td>2.11</td>
</tr>
<tr>
<td>21</td>
<td>1.98</td>
</tr>
<tr>
<td>22</td>
<td>1.92</td>
</tr>
<tr>
<td>23</td>
<td>1.86</td>
</tr>
<tr>
<td>24</td>
<td>1.81</td>
</tr>
<tr>
<td>25</td>
<td>1.70</td>
</tr>
<tr>
<td>26</td>
<td>1.13</td>
</tr>
<tr>
<td>27</td>
<td>1.64</td>
</tr>
<tr>
<td>28</td>
<td>1.1</td>
</tr>
<tr>
<td>29</td>
<td>1.56</td>
</tr>
<tr>
<td>30</td>
<td>1.02</td>
</tr>
<tr>
<td>31</td>
<td>0.999</td>
</tr>
<tr>
<td>32</td>
<td>1.47</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>1.03</td>
</tr>
<tr>
<td>35</td>
<td>1.05</td>
</tr>
<tr>
<td>36</td>
<td>1.07</td>
</tr>
<tr>
<td>37+</td>
<td>9.44</td>
</tr>
</tbody>
</table>
Oil recovery after injection of a specific number of pore volumes (PV), such as 1.2 PV of solvent, is the test criterion for miscibility. Two trend lines appear on a plot of recovery versus pore pressure for several slim-tube tests. The point of intersection of those trend lines is the estimated MMP for the given oil-solvent system. The data from a slim-tube test can also be used as input to fine-tune a fluid equation of state for reservoir simulation [185]. In experimental slim-tube determination of MMP, it is typically assumed to be the pressure at which there is a “break” in the curve on a graph of recovery versus pressure. Thus, it is the pressure above that very little additional recovery occurs. CO₂ flooding processes are not first contact miscible with most crude oils at reservoir conditions and the miscibility process is often very analogous to a vaporizing gas displacement process [186].

Slim-tube simulation runs were conducted to establish a value for the minimum miscibility pressure using GEM. GEM is CMG’s advanced equation-of-state compositional simulator, which includes various equation-of-state options to simulate CO₂, miscible gases, volatile oil, gas condensate, and many other processes that have complex phase behavior and many more [160].

A 1-D model was constructed consisting of 292 gridblocks in “I” direction and 1 gridblock in each “J” and “K” directions of which 290 were 0.2 ft in length, and the two grid cells at the either end of the slim-tube model were 1 foot in length. The cross-section of the slim-tube was ½ inch by ½ inch. The porosity and permeability was set to be 25% and 2,500 md. The value of porosity and permeability will not have any effect on the minimum miscibility pressure. They will only influence the value of hydrocarbon pore volume (HCPV) and the time needed to inject a predetermined amount of gas.

Injection was modeled at a low constant rate of 0.0001bbl/day (0.011cc/min) into the simulation model, and production at the other end was controlled by a minimum bottom-hole pressure constraint. The injection gas is mainly composed of CO₂. The bottom-hole pressure was varied for each run from 1,500 psia to 3,500 psia, in increments of 500 psia.

Figure 56 compares the oil recovery factor at HCPV of 1.2 at different pressures. As illustrated in the figure, an increase in pressure causes an increase in the oil recovery factor. However, there is threshold value beyond which a minimal increase will be seen in recovery. That value can be interpreted as the minimum miscibility pressure (MMP).
**Figure 56.** Slim tube simulation results show the increase in RF up to MMP.

A graph of the oil recovery factor after injecting 1 and 1.2 Hydrocarbon Pore Volumes (HCPV) of CO\(_2\) versus the pressure in the slim-tube model [186] is shown in Figure 57. This figure shows an increase in oil recovery with pressure until 2,500 psia, and flattens after 2,500 psia, which shows that the MMP is greater than 2,500 psia.

**Figure 57.** Slim tube simulation results estimate a MMP greater than 2500 psi.
In order to perform more analyses, the plus fraction ($C_7^+$) was split into more fractions using CMG-WinProp [187]. The new composition, as well as the injection gas composition, is shown in Table 11.

<table>
<thead>
<tr>
<th>Component</th>
<th>Primary Fluid Composition</th>
<th>Injection Gas Composition</th>
<th>Pseudo-Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO2</td>
<td>0.32</td>
<td>98.2</td>
<td>2</td>
</tr>
<tr>
<td>N2</td>
<td>0.83</td>
<td>1.6</td>
<td>1</td>
</tr>
<tr>
<td>CH4</td>
<td>28.65</td>
<td>0.2</td>
<td>1</td>
</tr>
<tr>
<td>C2H6</td>
<td>11.29</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C3H8</td>
<td>12.39</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>IC4</td>
<td>1.36</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>NC4</td>
<td>6.46</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>IC5</td>
<td>1.98</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>NC5</td>
<td>2.51</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>FC6</td>
<td>4.06</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C07-C10</td>
<td>12.24</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C11-C15</td>
<td>8.57</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C16-C19</td>
<td>3.79</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C20-C24</td>
<td>2.65</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C25+</td>
<td>2.89</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

WinProp uses an analytical tie-line method to calculate MMP by constructing a pseudo-ternary diagram [187]. According to WINPROP results, the first contact miscibility (FCM) does not happen even at 4,500 psi. However, the multiple-contact miscibility happens and the minimum miscibility pressure is calculated to be 2,840 psi, which is in agreement with the results presented earlier in this section. The pseudo ternary graphs are plotted and shown in the following figures.

Moreover, WinProp reported a vaporizing drive as the mechanism by which miscibility was achieved. The vaporizing-gas drive process achieves dynamic miscibility by in-situ vaporization of intermediate molecular weight hydrocarbons from the reservoir oil into the injected gas [180]. Vaporizing-gas drive mechanism prevails when reservoir temperature is greater than about 120°F (49°C) [180]. However, CO$_2$ is a much more powerful vaporizer of hydrocarbons than natural gas or flue gas. Hydrocarbons, as heavy as the gasoline and gas/oil fractions, are vaporized into the CO$_2$ front in addition to intermediate-molecular-weight hydrocarbons, and, because of this, development of
vaporizing-gas drive miscibility with CO$_2$ can occur with little or no C$_2$ through C$_6$ components present in the crude oil [188].

Stripping of lighter fractions of oil is observed in Miscible EOR as one of the mechanisms of recovery. Figure 59 depicts the variation of gas mole fraction for a light (C$_4$) and heavy (C$_{25}^+$) in two gridblocks. As shown in this figure the lighter components strip as a result of miscible CO$_2$ injection in the slim-tube model.
Figure 60. Stripping of lighter fractions during the time in two GBs

Figure 61 illustrates the accumulation of lighter components. Comparing Figure 61 and Figure 62 shows high concentration of light components in the gridblocks located in front of the flood due to injection of CO₂ and stripping from the oil.

Figure 61. Higher concentration of lighter fraction in the front due to miscible CO₂ injection
According to these analyses, as long as the pressure of the reservoir is above 2,800 psi, the process can be considered as a Miscible EOR and catches the physics of a miscible process.
CHAPTER 5. SRM DEVELOPMENT

The first step in developing SRM is defining a concrete objective. The objective determines the type and scale of SRM. Preparing a spatio-temporal dataset comes next. This dataset is used to teach the SRM (AI-based proxy model) the principles of fluid flow through porous media and the complexities of the heterogeneous reservoir represented by the geological model and its impact on the fluid production and pressure changes in the reservoir. The approach in preparing the dataset should be based on a “tailored” problem specific approach rather than “one size fits all”. This will be explained more in detail in succeeding sections.

The spatio-temporal dataset for SRM is built based on the results from handful realizations of the high-fidelity model. Taking a methodical tactic for designing the runs helps extracting the maximum information while contracting the computational time.
Figure 63. Grid based / well-based development, General methodology flowchart
5.1. DESIGN OF EXPERIMENTS

As it was explained in the previous sections, depending on the purpose of proxy modeling different realizations/scenarios should be built. The parameters changing in these realizations are problem specific. For instance, in an uncertainty quantification problem for CO$_2$ sequestration projects, the geological parameters might get more highlighted than an optimization case.

By definition optimization is the selection of the best element (with regard to some criteria) from some set of available alternatives [189]. In petroleum industry the optimization process can include ranking candidate opportunities for workovers, infill drilling, artificial lift, and enhanced oil recovery. To perform this, apart from screening the historical data, modeling is usually used in some cases. These studies can be used for asset management.

This work tries to build a base for optimization in a specific field; thereby no change in geological realizations will be considered. This would differ if the objective was history matching or uncertainty analysis. Since infill drilling is out of picture in the miscible CO$_2$ EOR process in this work, the only factors that could be changed and affect the production amount are the operational constraints. The drawdown pressure at each producer is dependent on its bottom-hole pressure. The production of each well is influenced by this value. The amount of the gas or water injected is another important factor. Apart from the amount of injected fluid, its duration in each half cycle plays an important role. All this can be added up in the definition of WAG ratio.

In the WAG process, gas and water slugs are alternately injected in a predetermined ratio called the WAG ratio. As stated by Wu [190], WAG ratio can be also defined as the ratio of the volume of water injected within the reservoir compared to the volume of injected gas. This value is usually in terms of the pore volume (PV) injected. The optimal WAG ratio is reservoir dependent as the response to each WAG scheme depends deeply on the distribution of permeability and reservoir heterogeneity as well as features that determine the impact of gravity segregation (fluid densities, viscosities, and reservoir flow rates) as well as reservoir wettability and the injected gas availability. High WAG ratio might lead to water oil trapping by water blocking or prohibit sufficient solvent-oil contact. Under this situation, the production performance ends up behaving like a water flood. On the contrary, a very small WAG ratio might cause channeling. This might cause the performance to behave like a gas flood. Rapid pressure declines, early gas breakthroughs, and high decline of production rate can be the consequences.
Experimental design method is an alternative to traditional sensitivity analysis. In an experiment, one or more process variables (or factors) are deliberately changed so as to perceive the influence the changes have on one or more response variables. In order to make sure that the data obtained can be analyzed to yield valid and objective conclusions, the (statistical) design of experiments (DOE), which is an efficient procedure for planning experiments, can be used [191].

Two main issues in designing the experiments are designing a parameter space sampling strategy and carrying out experiments, and analyzing the results of the experiments.

Some of the major classes of designs used in industrial experimentation include factorial designs, Latin square designs, Taguchi methods, constructing D- and A-optimal designs for surfaces and mixtures, etc.

In the setting of statistical sampling, a square grid encompassing sample positions is a Latin square if (and only if) there is only one sample in each row and each column. A Latin hypercube is the generalization of this idea to an arbitrary number of dimensions, whereby each sample is the only one in each axis-aligned hyperplane containing it. For sampling a function of \( N \) variables, the range of each variable is divided into \( M \) equally probable intervals. \( M \) sample points are then located to fulfill the Latin hypercube requirement. This forces all variables to have the same number of divisions \( M \). One of the main advantages of Latin Hypercube Sampling (LHS) over other sampling schemes is its independence of the number of samples from the dimensions (variables). Another advantage is that random samples can be taken one at a time, remembering which samples were taken so far. LHS was used for performing the sampling in this work.

**WAG Scenarios**

The production and injection data have been available since 1948 until 2004. The objective is to build the surrogate reservoir model which can be used for prediction and optimization purposes for more than a decade from the last date of available production data.

Studying the historical data through the reservoirs showed some of the wells were inactive for a long period of time. In order to build the realizations only 87 of producers were considered. The number of WAG injectors is 39. These wells will alternatively inject water and gas in different half cycles. There is a single well, which has been injecting water continuously and has been kept the same.

The number of required scenarios is decided upon based on lessons learnt from previous experiences using SRM. The WAG scenarios were designed based on sampling of the
duration and amount of injection of water and gas as well as BHP of production wells. In order to perform the LHS, for generating 11 scenarios, using four variables (water/gas injection rate, injection half cycle duration of water and gas), a minimum and maximum was decided for each variable. The field injection rate is between 10 Mbbl/d and 50 Mbbl/d. The minimum and maximum of the gas injection rate for the field (all WAG wells together) are 10 MMScf/d and 80 MMScf/d. The injection rate was allocated to the wells as their maximum water/gas injection rate constraint, based on their recent performance. Although the constraint was kept constant through the cycles of injection, its value was different in each well. This means that not only the constraint assigned to each well is different from other wells in that scenario, but also it differs from the assigned value of the same well in other scenarios.

Each injection cycle starts with a half cycle of gas injection followed by a half cycle of water injection. The range of these cycles is considered to be 6 to 24 months.

It was made sure that the generated scenarios cover the WAG ratio interval of 0.5 to 4.

The BHP at each well at the date of 1/1/2004 was recorded; a random pressure drop between 5% and 30% was applied to this value. The calculated pressure was assigned as the operational constraint (Min BHP) to the production wells in the built scenarios (Figure 65). Using this scheme ensures that the BHP value is unique for each well and it is always greater than the MMP. The maximum allowable injection pressure for all the injectors were kept at the constant value of 5000 psi.

On top of 11 scenarios defined based on the WAG process, 2 more cases were developed for the situations having exclusively water or gas for the entire cycle at the highest injection rate.

The described procedure was used to build the scenarios required for developing the SRM. These scenarios differ from each other in terms of the injection scheme of each well (amount and duration of injection) and the operating constraint of the producers. Table 12 lists these realizations and their corresponding WAG parameter.
Table 12. Scenarios selected for building SRM based on Design of Experiments.

<table>
<thead>
<tr>
<th>Training Scenarios</th>
<th>Field Gas Inj. (MMscf/d)</th>
<th>Gas Half Cycle (yr)</th>
<th>Field Water Inj. (Mbbls/d)</th>
<th>Water Half Cycle (yr)</th>
<th>WAG Ratio</th>
<th>Total Injected Gas (MMscf)</th>
<th>Total Injected Water (Mbbls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>1.0</td>
<td>50</td>
<td>2.0</td>
<td>4.0</td>
<td>21,645</td>
<td>182,500</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>0.7</td>
<td>40</td>
<td>1.7</td>
<td>3.0</td>
<td>22,447</td>
<td>165,467</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>0.5</td>
<td>30</td>
<td>0.9</td>
<td>2.0</td>
<td>23,479</td>
<td>110,413</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>2.0</td>
<td>45</td>
<td>1.5</td>
<td>1.0</td>
<td>48,702</td>
<td>98,550</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>1.8</td>
<td>25</td>
<td>1.3</td>
<td>0.5</td>
<td>48,975</td>
<td>60,833</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>0.5</td>
<td>10</td>
<td>0.7</td>
<td>0.5</td>
<td>25,550</td>
<td>32,850</td>
</tr>
<tr>
<td>7</td>
<td>70</td>
<td>1.1</td>
<td>50</td>
<td>2.0</td>
<td>0.5</td>
<td>153,300</td>
<td>182,500</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>0.6</td>
<td>30</td>
<td>1.8</td>
<td>1.2</td>
<td>44,713</td>
<td>130,488</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>1.2</td>
<td>42</td>
<td>1.9</td>
<td>0.5</td>
<td>117,104</td>
<td>146,913</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
<td>0.5</td>
<td>45</td>
<td>2.0</td>
<td>0.9</td>
<td>102,200</td>
<td>205,313</td>
</tr>
<tr>
<td>11</td>
<td>23</td>
<td>0.7</td>
<td>49</td>
<td>1.6</td>
<td>2.0</td>
<td>41,460</td>
<td>200,178</td>
</tr>
<tr>
<td>12</td>
<td>80</td>
<td>2.0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>467,200</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0.0</td>
<td>50</td>
<td>2.0</td>
<td>0.0</td>
<td>292,000</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 64 shows the scheme of injection. As it can be seen the half cycle life is different for water and gas in each scenario.

Figure 64. A schematic of the injection cycles in the sampled (training) scenarios (Yellow: Gas, Blue: Water).
Figure 65. The BHP range for all the wells in all the training scenarios.

Figure 66 shows the statistics of the sampled (training) WAG scenarios. As it can be seen, the scenarios are selected in such a way that the minimum and maximum of the variable ranges are covered. Per definition, the slug volume is the volume of injected fluid, which is dependent on the amount and length of the injection. Each cycle is comprised of a half cycle of gas injection followed by a half cycle of water injection. The higher the values of water slug volumes with respect to the gas slug volume, the greater the WAG ratio will be and vice versa. The combination of upper limits of gas rate range with lower limits of water injection generally results in very low impractical values of WAG; that is why the lower diagonal of the charts shown in Figure 66 does not include any scenarios.

The initial condition of all the scenarios is the same at 1/1/2004. The built models were run using the numerical simulation. Figure 67 shows how different injection schemes affect the oil production. Although the scenario with continuous gas injection might result in high oil production, because of the gas breakthrough, the amount of CO₂ production will be very high.
Comparing the results of different scenarios shows even though each of the operational constraints are important in EOR performance, the combination of them is the factor defining the efficiency of this process, therefore a decision on the optimum scenario cannot be made by studying each factor individually. The amount of CO₂ production should be considered as a pivotal factor, along with the incremental oil production, while defining the objective function for optimization process.

The WAG ratio for each scenario can be calculated by dividing the amount of injected water in PV by the amount of injected gas in PV \( \frac{HCPV_{WINj}}{HCPV_{GINj}} \). This value is almost the same in each cycle for a specific scenario. Figure 70 illustrates the injected pore volume of gas and water in scenarios 2 and 7 as an example.
Figure 67. Oil production rate and cumulative oil production in the training scenarios.

Figure 68. Cumulative gas and water injection in training scenarios.
Figure 69. Average reservoir pressure, field oil production rate, and field gas and water injection rate in WAG scenario 2 (Left) and scenario 7 (Right).

Figure 70. HCPV injected gas and water for Scenarios 2 and 7.
5.2. **Grid-Based SRM (SRM$_G$)**

One of the most important applications of Grid based SRM is to estimate dynamic properties in a space and time domain. In another words, using grid-based SRM depending on the solicitation of the model, one can estimate properties such as pressure, phase saturation, or composition of reservoir fluid components at any desired time step or gridblock of the reservoir. Grid based SRM is used when there is a specific interest in determining the flood front [192]. For example, risk assessment and uncertainty quantifications in CO$_2$ sequestration entail real-time knowledge of plume front location. In EOR projects involving a fluid injection the same analogy holds.

In the current study, the reservoir is going through an alternate water and gas injection, thereby the saturations are drastically changing. The pressure and flood front can be analyzed by generating SRMs that predict pressure, phase saturation, and mole fraction of CO$_2$ at any specific gridblock and time step of the reservoir life. This is vastly useful in performing the optimization and selecting the best injection scenario.

5.2.1. **SRM$_G$ Spatio-Temporal Dataset Generation**

A lot of weight in developing an SRM goes to the dataset preparation. This dataset is what will be used for teaching the SRM about the reservoir and essentially it should convey the principles of the “physics” of the reservoir and the specific problem at hand. The integrity of output from the SRM is dependent on the integrity of input. Faulty and erroneous data used in a model will result in distorted information. One cannot expect SRM to deliver good results if it has not been taught properly. That is why a reservoir engineering insight is absolutely compulsory for developing an SRM apart from the data mining knowledge.

The structure of the dataset is totally contingent upon the type and the objective of the reservoir. Table 13 lists the information needed for building the spatio-temporal database used in this study for building the grid-based SRM.
Table 13. Information required for developing Grid-Based SRM.

<table>
<thead>
<tr>
<th>Type</th>
<th>Property</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static/Dynamic</td>
<td>Type of GB (Boundary, Null, Well, Inj/Prd)</td>
<td>Grid</td>
</tr>
<tr>
<td>Static</td>
<td>Location (i,j,k, Long, Lat)</td>
<td>Grid</td>
</tr>
<tr>
<td>Static</td>
<td>Thickness</td>
<td>Grid /Tiers</td>
</tr>
<tr>
<td>Static</td>
<td>porosity</td>
<td>Grid /Tiers</td>
</tr>
<tr>
<td>Static</td>
<td>Permeability</td>
<td>Grid /Tiers</td>
</tr>
<tr>
<td>Static</td>
<td>Top</td>
<td>Grid /Tiers</td>
</tr>
<tr>
<td>Static</td>
<td>Euclidean distance to boundary</td>
<td>Grid</td>
</tr>
<tr>
<td>Static/Dynamic</td>
<td>Euclidean distance to closest offsets (Prod/Inj)</td>
<td>Grid</td>
</tr>
<tr>
<td>Dynamic</td>
<td>COW Amount of Prod/Inj (Rate/Cum)</td>
<td>Well</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Amount of Prod/Inj (Cum)</td>
<td>Field</td>
</tr>
<tr>
<td>Dynamic</td>
<td>COW BHP</td>
<td>Well</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Global CO₂ Mole Fraction at any time</td>
<td>Grid /Tiers</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Phase saturation (Oil/Water) at any time</td>
<td>Grid /Tiers</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Pressure at any time</td>
<td>Grid /Tiers</td>
</tr>
</tbody>
</table>

A lumping process has been performed to be able to handle the data better. This has been performed based on the geological zones defined in the geological model. Based on this process, 5 layers were defined as follows:

Table 14. Geological zones used for lumping and their corresponding simulation layers.

<table>
<thead>
<tr>
<th>Geological Zone used for Lumping</th>
<th>Corresponding Simulation Layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone 1</td>
<td>1 to 15</td>
</tr>
<tr>
<td>Zone 2</td>
<td>16 to 20</td>
</tr>
<tr>
<td>Zone 3</td>
<td>21 to 28</td>
</tr>
<tr>
<td>Zone 4</td>
<td>29 to 32</td>
</tr>
<tr>
<td>Zone 5</td>
<td>33 to 36</td>
</tr>
</tbody>
</table>

All the static and dynamic properties that are going to be used in the final database for SRM modeling have been averaged based on different methods depending on their nature (e.g. weighted arithmetic for saturation, arithmetic for mole fraction, etc.). The size of records in the dataset of one scenario decreases from 396,000 to 55,000, out of which 40,458 are active. The maps of some of the static properties in the first lumped layer are shown in Figure 71.
Figure 71. Structure Top (ft), Thickness (ft) (top), Porosity and Permeability (md) (bottom) distribution in the top lumped layer.

As seen in Table 13 attempts were made to transfer all the information carrying the physics to the networks. The injectivity or productivity of each well defines the potential of that well for injection and production. The injectivity can have a direct effect on the amount of pressure raise around the well, so the distance of each gridblock from the production/injection wells is a determining factor in the rate of change in the dynamic properties of that grid. The three closest production and injection wells have been considered in dataset preparation and the changes in those wells in terms of their operational constraints have been monitored. What has happened in the entire field to the day in terms of volumetric changes (cumulative production or injection) can have high influence on local grids’ dynamic properties.

In order to consider the impact of surrounding blocks, a tiering system was defined. Each gridblock has 8 corresponding tiers for all the static and dynamic properties. These values vary in time. The values at each tier are the average of properties of the blocks belonging to those tiers using different methods subject to the type of properties being averaged. These methods include arithmetic, weighted arithmetic, and harmonic
averaging. The schematic of the tiering system is illustrated in Figure 72. The blocks assigned to each tier are as follows:

- Tier 1: Average of Immediate adjacent blocks in the same layer.
- Tier 2: Immediate adjacent block in the top layer.
- Tier 3: Immediate adjacent block in the bottom layer.
- Tier 4: Average of Immediate non-adjacent blocks in same layer.
- Tier 5: Average of Immediate non-adjacent blocks in the top layer.
- Tier 6: Average of Immediate non-adjacent blocks in the bottom layer.
- Tier 7: Average of Tier 1, Tier 2, and Tier 3 values (all adjacent blocks).
- Tier 8: Average of Tier 4, Tier 5, Tier 6 values (all non-adjacent blocks)

![Figure 72](image.png)

*Figure 72. From left to right, Tier 1 to 8. Green block is the main blocks, Orange blocks show the ones used for tiering*

**Data Sampling**

After excluding the extended part (explained in numerical reservoir modeling), the database created from all the training cases will have 650,000 records and 1,716 features. Due to limitation of software for handling such a huge dataset, a sampling process should be performed. Through this process, a part of the spatio-temporal database is selected to represent the whole dataset and will be used for the SRM development purpose. Almost 6% of the whole data records will be extracted using sampling. Different sampling methods were tried on one of the properties and the one having the best results was selected for this study. Some of these methods were as follows:

- Method A: 10% of data was selected from well locations, 90% of data was randomly selected from the whole data pool.
- Method B: The data was ranked according to their pressure values (high pressure to low pressure). Data was then divided in uneven pressure ranges in such way to have equal records in each created bins (20% of data for each bin).
- Method C: The data was ranked according to their pressure values (high pressure to low pressure) and then divided in even pressure ranges (450 psi difference).

Method C was used as the base of sampling. Most to all of the data in the tiers having the least number of records were selected (Bins1 to 3).

Since there will be one neural network trained for each property per time step, the database was parsed after sampling to generate the final database for each year including all the static data and dynamic information for only up to three previous time steps.

### 5.2.2. **Key Performance Indicator (KPI) Analysis and Input Selection**

In order to address the “Curse of Dimensionality” one must understand the behavior and contribution of each of the parameters to the outcome being modeled. This is not a simple and straightforward task. Having a lot of features emphasizes the necessity of a procedure guiding through selecting the best inputs for neural networks. The inputs selected for networks can make or break it. Statistical analyses along with calculation of key performance indicators were applied on the dataset to mine the best inputs for each network. Key performance analysis is based on fuzzy cluster analysis and helps to study the relative influence of different reservoir properties as well as operational
constraints. The degree of influence spans between 0 and 100, with 100 showing the most influence of the attribute on the model output. Combination of KPI and reservoir engineering knowledge directs one toward selecting the most prominent features as an input to the network.

The KPI analysis result for pressure in one of the time steps is displayed in Figure 75 as an example.

![Figure 74: Sample KPI analysis for Pressure.](image)

It is really important to make the network learn and distinguish between different scenarios. As the geological realizations are unique in all the cases in this study, it is essential to express the operational differences in the best manner. Apart from the amount of injection at each step, the cumulative injection to the time at which modeling is performed is decisive.

Close investigation of the results leads to defining new parameters for each scenario. One of these parameters is the well productivity index (PI). As productivity or
injectivity capability of a well is directly dependent on permeability, thickness and drawdown pressure, the following equation was used and PI was calculated for all the wells.

\[ \text{Well PI} = K \cdot h \cdot (P_{\text{res}} - BHP) \]

*Equation (58)*

Where \( K \), \( h \), and \( P_{\text{res}} \) are the weighted average of permeability, thickness, and reservoir pressure at well location and \( BHP \) is the bottom hole pressure of the well. This value for all the COWs was counted in the dataset.

The amount and extent of recent gas or water injection is another pivotal factor in predicting the rate. These values along with the slug volume of water and gas were normalized in one cycle and incorporated in the dataset in a dynamic fashion. For each scenario and time step, the percentage of gas/water injection in the past 12 and 18 months was calculated and considered in the dataset.

Table 15 shows the inputs selected out of 310 features for predicting pressure at one of the time steps as an example.
### Table 15. A sample of selected inputs for predicting Pressure.

<table>
<thead>
<tr>
<th>Category</th>
<th>Input</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static/GB</td>
<td>X/Longitude</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y/Latitude</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min D to Boundary</td>
<td>Minimum Euclidean distance from the GB to the no flow boundary</td>
</tr>
<tr>
<td></td>
<td>PrdBlock</td>
<td>Determining if the GB belongs to a producer</td>
</tr>
<tr>
<td></td>
<td>InjBlock</td>
<td>Determining if the GB belongs to an injector</td>
</tr>
<tr>
<td></td>
<td>Min D to P 1</td>
<td>Minimum Euclidean distance from the GB to the first closest producer offset well (COPW)</td>
</tr>
<tr>
<td></td>
<td>Min D to P 2</td>
<td>Minimum Euclidean distance from the GB to the second COPW</td>
</tr>
<tr>
<td></td>
<td>Min D to Inj 1</td>
<td>Minimum Euclidean distance from the GB to the first COIW</td>
</tr>
<tr>
<td></td>
<td>Min D to Inj 2</td>
<td>Minimum Euclidean distance from the GB to the second COIW</td>
</tr>
<tr>
<td>Static/Res. Characteristics</td>
<td>Top</td>
<td>Top of the structure at the gridblock</td>
</tr>
<tr>
<td></td>
<td>PERM</td>
<td>Permeability of the block</td>
</tr>
<tr>
<td></td>
<td>PERM T1</td>
<td>Permeability of tier 1 of the selected block</td>
</tr>
<tr>
<td></td>
<td>PERM T7</td>
<td>Permeability of tier 7 of the selected block</td>
</tr>
<tr>
<td></td>
<td>PERM T8</td>
<td>Permeability of tier 8 of the selected block</td>
</tr>
<tr>
<td></td>
<td>PORO T7</td>
<td>Porosity of tier 7 of the selected block</td>
</tr>
<tr>
<td></td>
<td>PORO T8</td>
<td>Porosity of tier 8 of the selected block</td>
</tr>
<tr>
<td>Static/Field</td>
<td>NormalizedGasInjDur</td>
<td>Normalized duration of gas injection in one cycle</td>
</tr>
<tr>
<td></td>
<td>NormalizedWaterInjDur</td>
<td>Normalized duration of water injection in one cycle</td>
</tr>
<tr>
<td></td>
<td>NormGasInjSlugVol-1Cycle</td>
<td>Normalized slug volume of gas injection in one cycle</td>
</tr>
<tr>
<td></td>
<td>NormWaterInjSlugVol-1Cycle</td>
<td>Normalized slug volume of water injection in one cycle</td>
</tr>
<tr>
<td>Dynamic/GB</td>
<td>PRESS (t-1) T1</td>
<td>Pressure of the block in the previous time step (Output of the last model)</td>
</tr>
<tr>
<td></td>
<td>PRESS (t-1) T2</td>
<td>Pressure of the tier 1 of the selected block in the previous time step</td>
</tr>
<tr>
<td></td>
<td>PRESS (t-1) T3</td>
<td>Pressure of the tier 2 of the selected block in the previous time step</td>
</tr>
<tr>
<td></td>
<td>PRESS (t-1) T4</td>
<td>Pressure of the tier 3 of the selected block in the previous time step</td>
</tr>
<tr>
<td></td>
<td>PRESS (t-1) T5</td>
<td>Pressure of the tier 4 of the selected block in the previous time step</td>
</tr>
<tr>
<td></td>
<td>PRESS (t-1) T6</td>
<td>Pressure of the tier 5 of the selected block in the previous time step</td>
</tr>
<tr>
<td></td>
<td>PRESS (t-1) T7</td>
<td>Pressure of the tier 6 of the selected block in the previous time step</td>
</tr>
<tr>
<td></td>
<td>PRESS (t-1) T8</td>
<td>Pressure of the tier 7 of the selected block in the previous time step</td>
</tr>
<tr>
<td></td>
<td>PRESS (t-1) T1</td>
<td>Pressure of the tier 8 of the selected block in the previous time step</td>
</tr>
<tr>
<td>Category</td>
<td>Input</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Dynamic/Field</td>
<td>SO (t-1) T7</td>
<td>Saturation of the tier 7 of the selected block in the previous time step</td>
</tr>
<tr>
<td>Dynamic/Field</td>
<td>SO (t-1) T8</td>
<td>Saturation of the tier 8 of the selected block in the previous time step</td>
</tr>
<tr>
<td>Dynamic/Field</td>
<td>Field CumGasInj (t-1)</td>
<td>The cumulative gas injection in the field until the previous time step</td>
</tr>
<tr>
<td>Dynamic/Field</td>
<td>Field CumOilPrd (t-1)</td>
<td>The cumulative oil production in the field until the previous time step</td>
</tr>
<tr>
<td>Dynamic/Field</td>
<td>Field CumWaterInj (t-1)</td>
<td>The cumulative water injection in the field until the previous time step</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>CumGasInj(t-1)(1I)</td>
<td>The cumulative gas injection in the first closest injection offset (COIW) until the previous time step</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>CumGasInj(t-1)(2I)</td>
<td>The cumulative gas injection in the second COIW until the previous time step</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>CumWaterInj(t-1)(1I)</td>
<td>The cumulative water injection in the first COIW until the previous time step</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>CumWaterInj(t-1)(2I)</td>
<td>The cumulative water injection in the second COIW until the previous time step</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>G1 Well PI(Y-1)</td>
<td>Well PI in the first COIW (Gas)</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>G2 Well PI(Y-1)</td>
<td>Well PI in the second COIW (Gas)</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>W1 Well PI(Y-1)</td>
<td>Well PI in the first COIW (Water)</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>W2 Well PI(Y-1)</td>
<td>Well PI in the second COIW (Water)</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>Min BHP Const(1P)</td>
<td>Minimum BHP constraint in the first COPW</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>Min BHP Const(2P)</td>
<td>Minimum BHP constraint in the second COPW</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>Max Gas Rate Const(1I)</td>
<td>Maximum gas rate constraint in the first COIW</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>Max Gas Rate Const(2I)</td>
<td>Maximum gas rate constraint in the second COIW</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>Max Water Rate Const(1I)</td>
<td>Maximum water rate constraint in the first COIW</td>
</tr>
<tr>
<td>Dynamic/COW</td>
<td>Max Water Rate Const(2I)</td>
<td>Maximum water rate constraint in the second COIW</td>
</tr>
</tbody>
</table>
5.2.3. DATA PARTITIONING
Selecting the inputs is followed by data partitioning. Partitioning yields mutually exclusive datasets: a training dataset, a calibration dataset, and a validation dataset. Partitioning of data prior to model development is one of the most important parts of the model development process that many take for granted. Many times the success or failure of a project would hinge on how the data is segmented into these portions, or how the entire data set is partitioned. Different strategies can be used to perform the partitioning depending on the nature of data. The random partitioning was performed on the dataset to assign 80% of the data (equivalent to 25,600 records) to training, 10% (3,200 records) to calibration, and 10% (3,200 records) to validation partitions.

5.2.4. NETWORK DESIGN AND TRAINING
For each property of pressure, oil saturation, water saturation, and CO₂ mole fraction at each year a neural network was designed. Each network has three layers. The input layer is corresponding to the input features; hence their nodes are as many as the selected inputs are. The output layer consists of one node denoting the property being modeled. The hidden layer entails the nodes in which the computations happen. The number of nodes selected in the hidden layer was set to be more than the input layer nodes.

![Neural Network Architecture](image)

Figure 75. The neural network architecture for Pressure modeling at 1/1/2009

Different parameters such as momentum and learning rate can be changed for trying to train a better network. The "Learning Rate" is an indication of how fast the network
will learn the information presented. This is usually a moderate to low number (between 0 and 1). A large learning rate may cause the network to miss the global minimum in the search space and could cause a problem during convergence during training. A small learning rate value may prolong the learning process and slow it down to a crawl. Momentum is an extra push to the learning process that serves two purposes. First, it may accelerate the learning process, and second, it has the potential to kick the solution out of the local minima, that usually exists in the search space and causes the solutions to converge pre-maturely. Furthermore, different seed value can be assigned for the random number generator. The random numbers are used to initialize the weights on the neural network prior to the training [193].

The results of network training for pressure at 1/1/2009 are demonstrated in Figure 76. The frequency of error can be plotted to analyze and verify the accuracy of the network.

![Figure 76. The results of training, calibration, verification and the frequency of error for pressure network at 1/1/2009.](image)
The stoppage criterion in all the networks is reaching the best calibration set in order to prevent an over-fitting problem.

Training a network is an iterative process. The network might show a very good behavior based on the training partition results; however, if the verification partition does not follow more or less the same trend, it can imply the overtraining, meaning that the network has “memorized” the data and over fits. Having enough records in the dataset and considering the best calibration set as the final result, rather than the best training, prohibits the overtraining.

If the verification results are showing higher errors in comparison with training results, it also can imply that the samples in verification dataset do not have enough representatives (similar cases) in the training set. To solve this problem, the part of the data which is located in the verification partition and is showing bad behavior can be transferred to the training partition.

R-squared can be used as a relative yardstick for verifying the model integrity while training and determining how well it might predict the results of a new dataset. Table 16 lists the R-squared coefficients of the networks trained for each property at different years.

<table>
<thead>
<tr>
<th>Date</th>
<th>Pressure</th>
<th>So</th>
<th>Sw</th>
<th>GCO₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1/2006</td>
<td>98.21%</td>
<td>99.62%</td>
<td>99.50%</td>
<td>99.74%</td>
</tr>
<tr>
<td>1/1/2007</td>
<td>98.41%</td>
<td>99.58%</td>
<td>99.57%</td>
<td>99.83%</td>
</tr>
<tr>
<td>1/1/2008</td>
<td>98.18%</td>
<td>99.58%</td>
<td>99.67%</td>
<td>99.81%</td>
</tr>
<tr>
<td>1/1/2009</td>
<td>98.77%</td>
<td>99.51%</td>
<td>99.54%</td>
<td>99.83%</td>
</tr>
<tr>
<td>1/1/2010</td>
<td>98.78%</td>
<td>99.70%</td>
<td>99.71%</td>
<td>99.90%</td>
</tr>
<tr>
<td>1/1/2011</td>
<td>98.48%</td>
<td>99.54%</td>
<td>99.61%</td>
<td>99.61%</td>
</tr>
<tr>
<td>1/1/2012</td>
<td>98.42%</td>
<td>98.96%</td>
<td>99.16%</td>
<td>99.56%</td>
</tr>
<tr>
<td>1/1/2013</td>
<td>99.62%</td>
<td>99.10%</td>
<td>99.64%</td>
<td>99.64%</td>
</tr>
<tr>
<td>1/1/2014</td>
<td>99.13%</td>
<td>99.71%</td>
<td>99.55%</td>
<td>99.76%</td>
</tr>
<tr>
<td>1/1/2015</td>
<td>93.12%</td>
<td>99.74%</td>
<td>99.72%</td>
<td>97.52%</td>
</tr>
<tr>
<td>1/1/2016</td>
<td>99.27%</td>
<td>99.85%</td>
<td>99.84%</td>
<td>99.81%</td>
</tr>
<tr>
<td>1/1/2017</td>
<td>96.56%</td>
<td>99.64%</td>
<td>99.76%</td>
<td>99.90%</td>
</tr>
<tr>
<td>1/1/2018</td>
<td>98.34%</td>
<td>99.72%</td>
<td>99.77%</td>
<td>99.89%</td>
</tr>
<tr>
<td>1/1/2019</td>
<td>99.02%</td>
<td>99.88%</td>
<td>99.81%</td>
<td>99.88%</td>
</tr>
<tr>
<td>1/1/2020</td>
<td>98.77%</td>
<td>99.69%</td>
<td>99.82%</td>
<td>99.89%</td>
</tr>
</tbody>
</table>
Upon completion of training, the parameters that have had the most weight in training process can be listed alongside their relative influence percentage. These parameters for Pressure at 1/1/2009 have been listed in Figure 1. As it can be seen volumetric changes in the field in terms of injected and produced fluid, injection scheme properties and pressure in the last time step have the highest importance among the dynamic properties. Permeability and porosity are the most influential static parameters.

![Figure 1](image-url)

*Figure 1.* The parameters having the most influence on Pressure at 1/1/2009 in the training process.
5.2.5. **SRM<sub>G</sub> Deployment Results and Robustness Analysis**

The total of 60 neural network models was combined to create an ensemble forming the SRM<sub>G</sub> for each individual property (15 models per property). The SRM uses a cascading feature. Having an initial dataset including the reservoir dataset at the beginning, the trained models at each year are used to predict the output at each time step. The output of the model at each time step encompasses the pressure, oil saturation, water saturation, and the global mole fraction of CO<sub>2</sub> at each gridblock of the reservoir. The output of each individual model along with the calculated tier values is imported into the database of the next time. This process continues until the last year is reached. The total time taken for deploying the SRM and performing the cascading calculations using 60 networks is almost 800 seconds.

In the cascading part, the injection values can either be estimated using a separate SRM specific to that purpose or be calculated using the injection rate constraints if the design values are closely matching the simulation injection.

Figure 79 to Figure 86 illustrate the SRM deployment results, at the end of the first and the third WAG cycles (2006 and 2012), for one of the scenarios (scenario 5) as an example. In each cycle, 14 MMScf/d of gas is injected in the field for 22 months, followed by 16 months of 25 Mbbl/d of water. The results are shown at the end of the second WAG cycle.

It should be noted that the network has been trained with only 6% of the data (equivalent to 2,461 records) and 94% of the data (corresponding to 37,997 records) in each case is totally blind and has not been used in any part of the training process. The actual maps standing for the results from the commercial reservoir simulator have been showcased alongside the results predicted by the grid-based surrogate reservoir model. The error has also been shown proving the high accuracy of the SRM.

Layers 1 and 3 are more permeable than the rest and consequently have higher injectivity. This explains the more spread out distribution of CO<sub>2</sub> mole fraction in these layers. Layer 5 is the least permeable layer, experiencing higher pressure due to injection.
• Static data
• Dynamic data at time step 0
• Info about the Injection scheme

Input Dataset @ Initial timestep

Results @ step 1
• Pressure at all GBs
• So at all GBs
• Sw at all GBs
• Mole fraction of CO2 at all GBs

Results @ Time Step (t-1)
• Pressure
• So
• Sw
• GCO2

Results @ Time step t

Figure 78. Grid based cascading flow-chart.
Figure 79. Global mole fraction of $\text{CO}_2$ distribution map and error histogram at the end of the first WAG cycle (2006) - SRM$_c$ deployment results on scenario 5.

Figure 80. Global mole fraction of $\text{CO}_2$ distribution map and error histogram at the end of the third WAG cycle (2012) - SRMG deployment results on scenario 5.
Figure 81. Pressure distribution map and error histogram at the end of the first WAG cycle (2006) - SRM\(_C\) deployment results on scenario 5.

Figure 82. Pressure distribution map and error histogram at the end of the third WAG cycle (2012) - SRM\(_C\) deployment results on scenario 5.
Figure 83. Oil saturation distribution map and error histogram at the end of the first WAG cycle (2006) - SRM$_c$ deployment results on scenario 5.

Figure 84. Oil saturation distribution map and error histogram at the end of the third WAG cycle (2012) - SRM$_c$ deployment results on scenario 5.
Figure 85. Water saturation distribution map and error histogram at the end of the first WAG cycle (2006) - SRM$_G$ deployment results on scenario 5

Figure 86. Water saturation distribution map and error histogram at the end of the third WAG cycle (2012) - SRMG deployment results on scenario 5
As aforementioned, the robustness of the model is tested via different experiments. First, a part of dataset was put aside (validation partition) while training, and used for testing the integrity of the created network. The network showed high accuracy in the validation part. This was ensued by deploying the built SRM on different scenarios. Although part of the data has been used in building the model, 94% of the records have not been used in any part of the model generation. As shown in this section, the accuracy of the model was approved in this test as well. Apart from this test, it was decided to build a new scenario based on a totally new injection scheme and operational constraints. This information is not included in any part of the spatio-temporal database used for SRM development and is called “blind set”. Figure 87, Figure 88, and Table 17 show the injection schematic and operational constraints of the blind test.

<table>
<thead>
<tr>
<th>Training Scenarios</th>
<th>Field Gas Inj. (MMscf/d)</th>
<th>Gas Half Cycle (yr)</th>
<th>Field Water Inj. (Mbbls/d)</th>
<th>Water Half Cycle (yr)</th>
<th>WAG Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>11</td>
<td>0.8</td>
<td>45</td>
<td>1.8</td>
<td>3.5</td>
</tr>
</tbody>
</table>

*Figure 87. Schematic of injection design in blind set.*

*Table 17. Blind scenario for verifying the SRM.*

*Figure 88. The BHP range for all the wells the blind scenario.*

The developed SRM was deployed on the blind set to test and verify its robustness.
The results after one and three cycles of WAG (at the end of 2006 and 2011) are illustrated in Figure 89 to Figure 96.

**Figure 89.** Global mole fraction of CO$_2$ distribution map and error histogram at the end of the first WAG cycle (2006) - SRM$_G$ deployment results on the blind scenario

**Figure 90.** Global mole fraction of CO$_2$ distribution map and error histogram at the end of the third WAG cycle (2011) - SRM$_G$ deployment results on the blind scenario
Figure 91. Pressure distribution map and error histogram at the end of the first WAG cycle (2006) - SRM$_c$ deployment results on the blind scenario

Figure 92. Pressure distribution map and error histogram at the end of the third WAG cycle (2011) - SRM$_c$ deployment results on the blind scenario
Figure 93. Oil saturation distribution map and error histogram at the end of the first WAG cycle (2006) - SRM\textsubscript{C} deployment results on the blind scenario.

Figure 94. Oil saturation distribution map and error histogram at the end of the third WAG cycle (2011) - SRMG deployment results on the blind scenario.
Figure 95. Water saturation distribution map and error histogram at the end of the first WAG cycle (2006) - SRMG deployment results on the blind scenario

Figure 96. Water saturation distribution map and error histogram at the end of the third WAG cycle (2011) - SRMG deployment results on the blind scenario
As it can be seen from the results, the SRM can predict the borders of the flood front with a very high accuracy. The error is mostly initiated from the local difference in the intensity of the predicted values, which can be improved by different methods such as a time/property-based sampling technique.

The frequency of error for different properties was shown in the above figures. It can be observed that the error occurring in all the gridblocks of the reservoir in predicting pressure is dominantly under 10%. Global mole fraction of CO₂ shows more error distribution due to local abrupt changes around the injection wells; however it can quite accurately predict the fronts. The oil and water saturation prediction error in most of the gridblocks is less than 0.1.
5.3. **Well-Based SRM (SRM$_W$)**

Well-based SRMs are used when the objective is to simulate the response of the reservoir at well level (rate) to different reservoir parameters or operational constraints. Since the run time of a well-based SRM is much smaller than the traditional reservoir models, it becomes extremely useful in engineering tasks requiring many simulation runs such as sensitivity analysis, uncertainty quantification, history matching, and optimization, which all are a part of master development plans or asset-level decision making.

Since the objective of this study is to build the basis of optimization, building a well-based SRM is important. Using the well-based SRM, the best injection scenario can be decided upon. This decision is based on the amount of oil production from each well with respect to changes in operational constraints. CO$_2$ utilization factor, which is the ratio of CO$_2$ used to volume of oil produced, is a common factor for comparing different scenarios in an optimization course.

The procedure for developing a well-based SRM has been comprehensively explained in the following sections.

**5.3.1. SRM$_W$ Spatio-Temporal Dataset Generation**

Input to the network must be designed in a way to capture the complexity of the reservoir. The comprehensive spatio-temporal dataset generation is the step through which it is made sure that the physics of the reservoir and the phenomenon being modeled is captured and honored.

The structure of the spatio-temporal dataset might be different depending on the scheme under the study.

The required data will be extracted from the reservoir simulation model in a specific format. This dataset will be reorganized and prepared for SRM development application. A time span and a resolution (yearly in this study) are determined for data extraction task.

The physics of the reservoir should be well thought out while arranging the dataset. As each phenomenon or scheme occurring in reservoir follows its own rules of nature, having a reservoir engineering vision will be immensely important in getting together a well-representative dataset. After all you can only expect the network to provide you with what it has been taught about. Properties such as thickness and porosity define the storage capacity of the reservoir, while permeability defines the ease of fluid flow and reservoir conductivity and the depth of reservoir is a defining factor for reservoir
pressure. How far a well is from a no flow boundary or another offset well is quite
important for determining the influence they have on the production profile. Some
properties do not alter in time, while others such as operational constraints do. Aside
from the operational constraints, the saturation of different phases or pressure in the
area surrounding the wells influences the potential of each well for fluid production.

Table 18. Information required for developing Well-Based SRM.

<table>
<thead>
<tr>
<th>Type</th>
<th>Property</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Time</td>
<td>Time</td>
</tr>
<tr>
<td>Static</td>
<td>Drainage Area</td>
<td>Well/Offsets</td>
</tr>
<tr>
<td>Static</td>
<td>Location (i,j,k, Long,Lat)</td>
<td>Well/Offsets</td>
</tr>
<tr>
<td>Static</td>
<td>Thickness</td>
<td>Grid (Tiers)</td>
</tr>
<tr>
<td>Static</td>
<td>porosity</td>
<td>Grid (Tiers)</td>
</tr>
<tr>
<td>Static</td>
<td>Permeability</td>
<td>Grid (Tiers)</td>
</tr>
<tr>
<td>Static</td>
<td>Top</td>
<td>Grid (Tiers)</td>
</tr>
<tr>
<td>Static</td>
<td>Euclidean distance to boundary</td>
<td>Well</td>
</tr>
<tr>
<td>Static/Dynamic</td>
<td>Euclidean distance to closest offsets (Prod/Inj)</td>
<td>Well</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Amount of Prod/Inj (Rate/Cum)</td>
<td>Well</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Amount of Prod/Inj (Rate/Cum)</td>
<td>Field</td>
</tr>
<tr>
<td>Dynamic</td>
<td>BHP</td>
<td>Well</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Global CO₂ Mole Fraction at any time</td>
<td>Grid (Tiers)</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Phase saturation (Oil/Water) at any time</td>
<td>Grid (Tiers)</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Pressure at any time</td>
<td>Grid (Tiers)</td>
</tr>
</tbody>
</table>

As explained before, in order to be able to handle the dataset and convey as much
information as possible, an integration process was performed, resulting in 5 different
layers. Different averaging approaches (simple, harmonic weighted or arithmetic
weighted) were pursued contingent to the each specific feature.

Volumetric analyses data can be used for expressing more statistics of the reservoir. To
do so, reservoir delineation and tiering were performed. Reservoir delineation provided
uniform spatial distributed pattern based on the Voronoi graph theory. In mathematics,
a Voronoi diagram is a way of dividing space into a number of regions. A set of points
(called seeds, sites, or generators) is identified in advance and for each seed there will
be a corresponding region entailing all points closer to that seed than to any other. The
regions are called Voronoi cells [194]. In a reservoir delineation application, the wells
play the role of the aforementioned seeds. It is assumed that each well is contributing
in production based on its assigned drainage area and relative reservoir characteristics.
Polygon (Voronoi) based information can also be encompassed in the final dataset
which carries a higher information content in comparison with values at each specific well block.

![Image](image1.png)

**Figure 97.** Tier 1 and 2 location w.r.t well block; in one layer (Left) and Upscaled format (Right)

![Image](image2.png)

**Figure 98.** The Voronoi cells overlaid on Top map in Layer 1 (Left) and Tier map (Right). White cells represent the injectors, Dark Blue: Tier 1, Green: Tier 2, Red: Tier 3

As most of the dynamic changes take place around the wells, it is important to keep the resolution in this area. Thereby a specific number of gridblocks were assigned to Tier 1 and Tier 2 based on their distance to each well. Tier values were calculated for each property at each layer. Since the model is heterogeneous in terms of structure, a weighted value was calculated for the final Tier corresponding to each well-based on the single layer properties. After determining Tier 1 and 2, the remaining gridblocks in each well polygon (Voronoi cell) were used to calculate Tier 3 for the corresponding well (Figure 97 and Figure 98).
The dataset is built in a chronological order. This format can build timelines that illuminate the concepts within a domain. The dynamic data chronicled in up to two previous time steps are also used. This part of dataset preparation was performed using commercial software [195] to create the dataset in a sequential format.

The integrated dataset includes the information of all the training cases, comprising 15,600 records of 586 features. Various analyses of data mining techniques should be performed in order to obtain “not apparent data” that will be useful in revealing hidden patterns before building the SRM.

5.3.2. **Key Performance Indicator (KPI) Analysis and Input Selection**

Using basic statistical analysis the relationship between different parameters including the output can be studied. Some examples of the statistical analyses are shown in Figure 99. This figure displays how the oil production data at each time varies by change in gas/water injection slug volume in one cycle, or the duration of gas injection in the previous 18 months. The correlation between pressure value at Tier 1 surrounding the well block and time is also shown. These plots are based on the comprehensive spatio-temporal dataset and might be useful. Yet, it will be really time consuming to study the mutual effect of each parameter on each other or the output value. On the other hand, there is a lot of information hidden in the data, which may not be revealed using basic statistical analysis. Therefore, KPI analysis can be used as a powerful alternative. Figure 100 shows a part of KPI results performed to detect the most influential parameters for predicting oil production.

![Figure 99: Samples of basic statistical analysis showing the correlation between the oil production rate and different parameters.](image-url)
Figure 100. A part of KPI analysis result with oil production rate at time step \((t)\) as output.

Although the statistical analysis and KPI are great tools for guiding through input selection, it should be noted that a reservoir engineering sense should be accompanied with them to select the best input set.

Not only should the values of injection and production be considered for individual wells and their closest offset wells (COWs), but also for the entire field. The quantity of injected/produced fluid in well or field scale will definitely be decisive in modeling the production rate.

The normalized injection amount and duration for water and gas, along with the duration of gas in percentage during the previous interval of 12 and 18 month, were added to the dataset. This is an important feature since the production pick is usually happening with a lag after initiation of the injection. Well PI was also considered for all the wells. Table 19 displays a sample list of the inputs used for modeling oil production.
**Table 19. A sample of selected inputs for predicting oil production rate.**

<table>
<thead>
<tr>
<th>Category</th>
<th>Input</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Static / GB</strong></td>
<td>Time and X-Y Coordinates</td>
<td></td>
</tr>
<tr>
<td></td>
<td>h-Well</td>
<td>Thickness at well gridblock</td>
</tr>
<tr>
<td></td>
<td>Perm-T3</td>
<td>Permeability at Tier 3</td>
</tr>
<tr>
<td></td>
<td>Perm-Well</td>
<td>Permeability at well gridblock</td>
</tr>
<tr>
<td></td>
<td>Polygon Pore Vol(cft)</td>
<td>Pore volume at the drainage area of the well</td>
</tr>
<tr>
<td></td>
<td>Top-Well</td>
<td>Top at well gridblock</td>
</tr>
<tr>
<td></td>
<td>Min BHP Const</td>
<td>Minimum bottom hole pressure constraint</td>
</tr>
<tr>
<td></td>
<td>PercentGasInj-Last18M(t)</td>
<td>The length of gas injection in the past 18 months</td>
</tr>
<tr>
<td></td>
<td>NormalizedGasInjDur</td>
<td>Normalized duration of gas injection in one cycle</td>
</tr>
<tr>
<td></td>
<td>NormalizedWaterInjDur</td>
<td>Normalized duration of water injection in one cycle</td>
</tr>
<tr>
<td></td>
<td>NormGasInjVol-1Cycle</td>
<td>Normalized slug volume of gas injection in one cycle</td>
</tr>
<tr>
<td></td>
<td>NormWaterInjVol-1Cycle</td>
<td>Normalized slug volume of water injection in one cycle</td>
</tr>
<tr>
<td></td>
<td>P-Well(t)</td>
<td>Pressure at well gridblock and time step t</td>
</tr>
<tr>
<td></td>
<td>SO-Well(t)</td>
<td>Water saturation at well gridblock and time step t</td>
</tr>
<tr>
<td>Dynamic / Field</td>
<td>Well PI(t)</td>
<td>Well productivity index at time step t</td>
</tr>
<tr>
<td></td>
<td>q(t)-Oil</td>
<td>Oil production at previous time step</td>
</tr>
<tr>
<td>Dynamic / Grid</td>
<td>q(t)-Oil</td>
<td>Oil production at two previous time step</td>
</tr>
<tr>
<td>Dynamic / Well</td>
<td>Distance(1P)</td>
<td>Minimum Euclidean distance from the first closest producer offset well (COPW)</td>
</tr>
<tr>
<td></td>
<td>Well PI(t)(1P)</td>
<td>Well productivity index at time step t for the first COPW</td>
</tr>
<tr>
<td></td>
<td>Min BHP Const(1P)</td>
<td>Minimum bottom hole pressure constraint for the first COPW</td>
</tr>
<tr>
<td></td>
<td>q(t)-Oil(1P)</td>
<td>Oil production at previous time step for the first COPW</td>
</tr>
<tr>
<td></td>
<td>Distance(2P)</td>
<td>Minimum Euclidean distance from the second COPW</td>
</tr>
<tr>
<td></td>
<td>Min BHP Const(2P)</td>
<td>Minimum bottom hole pressure constraint for the second COPW</td>
</tr>
<tr>
<td></td>
<td>q(t)-Oil(2P)</td>
<td>Oil production at previous time step for the second COPW</td>
</tr>
<tr>
<td></td>
<td>q(t)-Oil(2P)</td>
<td>Oil production at two previous time step for the second COPW</td>
</tr>
<tr>
<td>Dynamic / COWs</td>
<td>Distance(1I)</td>
<td>Minimum Euclidean distance from the first closest injector offset well (COIW)</td>
</tr>
<tr>
<td></td>
<td>CumGasInj(t-1)(1I)</td>
<td>The cumulative gas injection in the first COIW until the previous time step</td>
</tr>
<tr>
<td></td>
<td>CumWaterInj(t-1)(1I)</td>
<td>The cumulative water injection in the first COIW until the previous time step</td>
</tr>
<tr>
<td></td>
<td>Max Gas Rate Const(1I)</td>
<td>Maximum gas injection rate constraint in the first COIW</td>
</tr>
<tr>
<td></td>
<td>Max Water Rate Const(1I)</td>
<td>Maximum water injection rate constraint in the first COIW</td>
</tr>
<tr>
<td></td>
<td>Distance(2I)</td>
<td>Minimum Euclidean distance from the second COIW</td>
</tr>
<tr>
<td></td>
<td>CumGasInj(t-1)(2I)</td>
<td>The cumulative gas injection in the second COIW until the previous time step</td>
</tr>
<tr>
<td></td>
<td>CumWaterInj(t-1)(2I)</td>
<td>The cumulative water injection in the second COIW until the previous time step</td>
</tr>
<tr>
<td></td>
<td>Max Gas Rate Const(2I)</td>
<td>Maximum gas injection rate constraint in the second COIW</td>
</tr>
<tr>
<td></td>
<td>Max Water Rate Const(2I)</td>
<td>Maximum water injection rate constraint in the second COIW</td>
</tr>
</tbody>
</table>
5.3.3. **DATA PARTITIONING**

Investigation of the data shows three specific patterns in production. Intelligent data partitioning was used rather than random data partitioning to divide the dataset into three categories of training, calibration, and validation, which makes sure the original dataset is distributed such that all three partitioned datasets are statistically representative.

It was decided to include 80% of the data (equivalent to 12,480 records) in the training set. Each calibration and validation set incorporates 10% of the data (1,560 records). The records in the validation set are not used during training or calibration and can be used to test model generalization ability and robustness.

5.3.4. **NETWORK DESIGN AND TRAINING**

Three different networks are designed for oil, water, and CO₂ production, each of them having three layers. The nodes in the input layers are selected based on the results of KPI and reservoir engineering sense. They can be interchangeably modified to enhance the network results. The networks have only one hidden layer, usually having more nodes than the input layer. The output layer consists of only one node pertaining to the desired output.

The backpropagation algorithm is used for training the network. Training a network is an art. In general, the network architecture is not a crucial factor in making it a success or failure. On the contrary, the type of inputs can have a substantial effect on the results.

The momentum and learning rate can be modified for reaching a better result.

For modeling CO₂ production only 12 scenarios were used and scenario 12 was eliminated from the dataset. Since scenario 12 is the extreme case of continuous high volume gas injection, it changes the range of the data and does not allow a good learning process.

Figure 101 shows the training results for oil production. The criterion in training all the networks is the best calibration result in order to prevent an over-fitting problem. This means even if the network is showing improvement based on training data partition, it will not be picked unless it is showing the same trend in the calibration set.
As explained in the SRM_g section, coefficient of determination, denoted by $R^2$, is used for indicating how well data is represented by the built model. This coefficient is listed for the oil, water, and gas production models shown in Table 20.

**Table 20. R-squared coefficients of all the neural networks built for grid-based SRM.**

<table>
<thead>
<tr>
<th>Model</th>
<th>R-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oil Prod. Rate</td>
<td>98.57%</td>
</tr>
<tr>
<td>Water Prod. Rate</td>
<td>99.53%</td>
</tr>
<tr>
<td>Gas Prod. Rate</td>
<td>99.24%</td>
</tr>
</tbody>
</table>

At the end of the training process, the parameters with the most influence can be listed. The ranking of these parameters is not necessarily the same as the results of KPI analysis. Oil production in the previous time step, BHP, injection duration, and permeability have the highest weights.
Figure 102. The parameters having the most influence on Oil Production in the training process.

5.3.5. **SRM\textsubscript{w} Deployment Results and Robustness Analysis**

After the networks were studied in a standalone fashion to determine their individual quality, they were applied in a cascading feed-forward design in order to obtain a fully-fledged well-based surrogate model (SRM\textsubscript{w}) as the replica of the former numerical reservoir model.

For performing the SRM deployment only the information at the initial time step are compulsory. The SRM is applied at each time step and the output and corresponding information are calculated. This set of information includes the designated output (oil, water production, or CO\textsubscript{2} mass rate) and the values calculate based on the output, such
as the cumulative in wells and entire field, gas oil ratio, and water cut, etc. The calculated output and its byproducts at each step are used as an input to the next step using this cascading feed forward feature.

The visualized output of SRM_W for oil, water, and CO_2 production of the entire field and wells corresponding to 10-1, 15-6, 20-11, 27-11, and 8-1 (shown in Figure 103), in scenario 5 are demonstrated in Figure 104 to Figure 106. These figures show a very close match between the result of commercial reservoir simulator and the SRM_W, denoting high accuracy of the surrogate model.

Figure 103. The wells corresponding to samples shown in Figure 104 to Figure 106.
Figure 104. The $SRM_W$ deployment results in Scenario 5 - Oil production in sample wells and entire field.
Figure 105. The $\text{SRM}_W$ deployment results in Scenario 5 - CO$_2$ production of sample wells and entire field.
Figure 106. The SRM\textsubscript{w} deployment results in Scenario 5 - Water production of sample wells and entire field.
**Blind Set Verification**

To investigate the robustness of the model a new scenario was built. The BHP value of each production well in this blind case is unique in the sense that it has not been repeated in any of the wells of this case or any training scenario (Figure 88). The same is true for the injection scheme and constraints (Figure 88 and Table 17). In each cycle, 11 MMScf/day of gas is injected in the field for 10 months, followed by 22 months of 45 Mbbl/day of water.

Following the aforementioned procedure the spatio-temporal database of the so-called blind set was prepared. This dataset was used to measure the goodness of the SRM and identify the degree of confidence one can have on the predictions that are made by the SRM. Akin to the procedure followed in deploying SRM in training cases, the output of each time step in terms of production rates in conjunction with other values calculated based on them, will be the input of the next time step.

The outcomes of SRM for oil, water, and CO$_2$ production of the entire field and sample wells are presented in Figure 107 to Figure 109. These results show the generalization ability of the SRM as none of the data were seen by the model before.
Figure 107. The SRM_w deployment results in blind scenario - Oil production in sample wells and entire field.
Figure 108. The $\text{SRM}_W$ deployment results in blind scenario- CO$_2$ production of sample wells and entire field.
Figure 109. The SRM$_W$ deployment results in blind scenario—Water production of sample wells and entire field.
5.4. COUPLED SURROGATE RESERVOIR MODEL

The process of coupling the grid-based and well-based SRM has been implemented in this work for the first time. In an optimization process the flood front monitoring is as important as the production rate estimation.

To this end, we propose to assimilate the SRM\textsubscript{G} and SRM\textsubscript{W} by feeding the data from one model to another one. The coupled SRM will be a great tool in accomplishing the following tasks.

- Monitoring the interaction of wells with one another.
- Monitoring flood front during water flood or gas injection operations.
- Optimizing injection rates for maximum sweep efficiency.
- New well placements to optimize enhance recovery.
- Developing new strategies for field development.

Figure 110 shows the flowchart of performing coupled surrogated reservoir modeling. Two distinct databases are used for the integration course. The grid-based spatio-temporal database encompasses the static and dynamic data at the initial time step before the injection scheme initiates. This part of dataset is unique no matter what injection scheme is in play. The only data required apart from this common database, is the information about the injection scheme and operational constraints based on the desired WAG ratio. The same holds for the well-based spatio-temporal dataset.

The process starts with the grid-based database. Upon running the SRM\textsubscript{G}, the pressure, phase saturation, and CO\textsubscript{2} mole fraction are calculated for all the gridblocks in the model for the succeeding time step. This information is processed, well PI calculation and Tiering computations pertaining to grid-based and well-based systems are performed. The results calculated values accompanied by the well-based initial information are fed to the SRM\textsubscript{W}. Implementing this model will yield the oil, water, and CO\textsubscript{2} production of the wells and entire field at the first time step. Accordingly, the cumulative values in the wells and entire field, in consort with water cut and gas oil ratio, in each well is calculated. The database which is going to be used for the SRM\textsubscript{G} and SRM\textsubscript{W} in the next step is updated correspondingly, and the SRM\textsubscript{G} is run again to produce the dynamic data of the subsequent time step. The sequence is continued until the last time step (year) is reached.
Figure 110. The coupled SRM flowchart.
Figure 111 to Figure 118 presents the results of the coupled SRM in terms of pressure, phase saturation, and component mole fraction accompanied by an error histogram, for one of the cases (scenario 5) after the first and the third cycles of WAG (end of 2006 and 2012) as an example.

The production rates obtained from the coupled SRM is demonstrated in Figure 119 to Figure 121.

Although the results have been slightly deteriorated compared to the standalone grid-based or well-based SRMs, it can be observed that coupled SRM still show relatively high accuracy. It should be noted that coupling the grid-based and well-based SRM is logically propagating some errors since the results of each time step of one model is not only used for the next time step of that model, but also is imported to a model of a different nature (grid-based or well-based).

Despite some differences in the exact quantity of the properties in grid-based SRM, it should be noted that what matters in most of the injection cases (e.g. CO₂ sequestration) is the plume extension and its boundary location at each time rather than the exact value. SRM does an excellent job in qualitative representation of the fronts and presents acceptable results in predicting the quantities. In reduced order models however, the prediction of the exact value highly matters since the production rate is very much dependent on the saturation values according to the relative mathematical formulations.
Figure 111. Global mole fraction of CO₂ distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on scenario 5.

Figure 112. Global mole fraction of CO₂ distribution map and error histogram at the end of the third WAG cycle (2012) – Coupled SRM deployment results on scenario 5.
Figure 113. Pressure distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on scenario 5.

Figure 114. Pressure distribution map and error histogram at the end of the third WAG cycle (2012) – Coupled SRM deployment results on scenario 5.
Figure 115. Oil saturation distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on scenario 5.

Figure 116. Oil saturation distribution map and error histogram at the end of the third WAG cycle (2012) – Coupled SRM deployment results on scenario 5.
Figure 117. Water saturation distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on scenario 5.

Figure 118. Water saturation distribution map and error histogram at the end of the third WAG cycle (2012) – Coupled SRM deployment results on scenario 5.
Figure 119. The Coupled SRM deployment results in Scenario 5 - Oil production in sample wells and entire field.
Figure 120. The Coupled SRM deployment results in Scenario 5 – CO₂ production in sample wells and entire field.
Figure 121. The Coupled SRM deployment results in Scenario 5 - Water production in sample wells and entire field.
Although the integrity of coupled SRM was tested on one of the cases partially used in training, it was also deployed on a blind set. As explained before the results will show a slight deviation from the standalone cascading results due to propagation of error in two types of SRMs. The results are visualized in Figure 122 to Figure 132 and prove the applicability of coupled SRM. Once again it should be noted that none of the information regarding the constraints or injection design has been experienced by the network before.
Figure 122. Global mole fraction of CO₂ distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on the blind scenario.

Figure 123. Global mole fraction of CO₂ distribution map and error histogram at the end of the third WAG cycle (2011) – Coupled SRM deployment results on the blind scenario.
Figure 124. Pressure distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on the blind scenario.

Figure 125. Pressure distribution map and error histogram at the end of the third WAG cycle (2011) – Coupled SRM deployment results on the blind scenario.
Figure 126. Oil saturation distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on the blind scenario.

Figure 127. Oil saturation distribution map and error histogram at the end of the third WAG cycle (2011) – Coupled SRM deployment results on the blind scenario.
Figure 128. Water saturation distribution map and error histogram at the end of the first WAG cycle (2006) – Coupled SRM deployment results on the blind scenario.

Figure 129. Water saturation distribution map and error histogram at the end of the third WAG cycle (2011) – Coupled SRM deployment results on the blind scenario.
Figure 130. The SRM deployment results on the blind scenario – Oil production in sample wells and the entire field.
Figure 131. The SRM$_W$ deployment results on the blind scenario – CO$_2$ production in sample wells and the entire field.
Figure 132. The SRM_w deployment results on the blind scenario - Water production in sample wells and the entire field.
CHAPTER 6. CONCLUSION AND DISCUSSION

Like many other real world problems, the issues dealt with in oilfield management and reservoir engineering, are not straightforward. There have been many efforts to minimize the uncertainties associated with measurement by combining numerous types of information gathered from many sources in modern reservoir models. The increase in the number of constraints and parameters does not necessarily guarantee an easier way for reaching the most feasible solution.

Performing many scientific and engineering tasks such as optimization, parameter identification, uncertainty analysis, etc. require dealing with numerous models usually built based on partial differential equations (PDEs).

Solving PDEs to a high degree of accuracy is often time consuming. The engineering practices (history matching, uncertainty quantification, sensitivity analysis, and optimization) usually necessitate the PDE-based reservoir model realizations to run hundreds or thousands of times. The larger and more complicated the model is, the longer the processing time will be. Hence, it is obvious what a considerable strain each of these tasks will put on the resources of an asset team. Performing these processes is impractical in some situations especially with a real-time decision making in mind. The number of such high-fidelity simulations required for these analyses makes the computational demands even more challenging if not prohibitive.

In spite of the fact that many techniques have been made available for production optimization in the upstream oil and gas industry, it is still a challenging task to optimize reservoir performance in the presence of physical and/or financial
Economic success of an EOR project principally depends on the efficiency of the selected method in the reservoir of interest and the way it is deployed. Owing to the heavy computational work compulsory for a large number of simulations, it is rather a laborious and tedious task to determine optimum design schemes for a given project.

Proxy modeling has been introduced as an alternative for traditional reservoir modeling for conducting a high number of runs in the petroleum industry in recent years, although there has been a resistance toward shifting from traditional reservoir modeling to proxy modeling. One of the most important reasons for this defiance is getting used to traditional reservoir modeling. As Mark Twain says "Nothing so needs reforming as other people's habits." Bad presentation of the proxy models and their application is another reason for making this paradigm shift even harder. It should be specified when the proxy models are of the best use and under what circumstances they can be used in conjunction with numerical reservoir simulation tools or independent from them.

The procedure for building a surrogate reservoir model (SRM) was explained in detail through the course of this work. The field under study is called SACROC field, located in Permian basin which has been undergoing a CO₂-EOR project for more than 30 years. This study was performed to provide a base for optimization by investigating the feasibility of using state-of-the-art data-driven proxy models through development of a surrogate reservoir model. The workflow started with building the fine resolution geological model followed by upscaling and developing the reservoir simulation model.

The intricacy of simulating multiphase flow, especially in the case of miscible CO₂ injection, with a large number of time steps required to study the reservoir response to the injection, dealing with a highly heterogenous reservoir and numerous number of wells all add up to the complexity of problem. The necessity of the SRM has to do with the fact that massive potentials of the existing numerical reservoir simulation models go unrealized attributable to the long time required to make a single run. Even on a cluster of parallel CPUs, numerical models that are built to simulate complex reservoirs with multi-million gridblocks entail substantial run-time.

The simulation time for a single run of the built model for the total life of the reservoir took almost a month. Using design of experiments techniques, multiple WAG scenarios were planned and corresponding models were created and run. The spatio-temporal database was created based on the generated data. Data mining and artificial intelligence techniques were employed for devising two different types of surrogate reservoir models, with the ability of estimating the potential of wells in terms of
production and dynamic distribution of different properties in every single gridblock of the reservoir. These models were coupled to provide an integrated SRM which can predict the reservoir dynamic response both at the well and grid level.

In some cases, it might seem the grid-based model generates better results than the well-based. This can be explained with the fact that the amount of representative samples in the database records plays a very important role. As seen in the grid based model, the network had a better prediction for grids that has more representative or similar samples in the spatio-temporal database, because it had learned their behavior better. In the well-based models the samples and consequently the data records are less compared to gridblocks.

By defining an objective function (e.g. net present value or CO₂ utilization factor), the coupled SRM can be used to run and compare hundreds of scenarios based on various injection schemes and WAG ratios and determine the optimal scenario using an optimization algorithm.

It takes more than 48 hours for one run (the life span used for optimization design) to be completed using a commercial reservoir simulator on a machine with 24 GB RAM and 3.47 GHz processor. The run time of each model in the SRM on the same machine is 15 seconds out of which 10 seconds pertains to SRM deployment at each step. If 100 runs are required for performing an optimization process it will take almost 7 months to carry out all the runs; this duration increases up to almost 6 years for 1000 runs, making the whole procedure impractical and unattainable. Using the SRM, on the other hand, will only take a bit more than a day or less than two weeks to perform 100 or 1,000 runs correspondingly, which is a huge accomplishment.

Coupled SRM was developed in this study for the first time. At the time of this study, none of the works performed in this research had ever been performed in this scale of complexity neither in well-based nor grid-based level. Similar works in the well-based level have been done on the order of maximum 10 wells, and the main focus has never been on grid-based values and they have been used only as a means for rate calculation. The grid-based prediction has been performed typically in 2-D and two-phase reservoirs. This highlights the achievements of this study which is performed on a compositional heterogeneous high order 3-D, 3-phase reservoir model having 3,622,190 geocellular blocks (396,000 gridblocks), 40 injection and 75 production wells and undergoing miscible CO₂-EOR process.

Although in the majority of the cases very good results were attained, there were cases with less positive outcomes. The objective of the next section is to capitalize on what
was done well so it can be used for embarking on similar projects, and to discuss what needs to be modified or studied more in future works.

6.1. Lessons Learned
The topics listed here are results of epiphanies throughout the life of the project with hopes of presenting and featuring the best workaround for the future.

Time Resolution
It is extremely important to convey enough information to the neural networks while training. Some information might be misleading and may not result in the desired outcome. The time resolution of the model that is going to be built should be decided in advance and contingent upon the type of problem. If the problem at hand is a continuous production with minimum change in operational constraints, time resolution might not be a key element, howbeit if the alteration in the system is frequent, especially when it causes a change in physics, time resolution will be a focal point in building the representative SRM.

In a miscible WAG process the governing physics is different between each half cycle of gas and water injection. It is immensely critical to assure that the network catches this difference by tutoring it. In our study, we have considered an annual time resolution while in the designed scenarios the interchange of injection fluid in the WAG cycles does not necessarily follow this resolution (Figure 64). This will cause confusion in the network in predicting the exact values when the interchange happens between two time steps. Although some factors, such as cumulative injection of different fluids or duration of injection, were considered to circumvent this issue, it is recommended to consider a monthly or biannual resolution or define the resolutions based on the changes in the training scenarios (Figure 133).

The fact that the response of the producers to the injection (especially gas injection) occurs with a lag is another factor that accentuates the necessity of a finer time resolution.

While dealing with a cascading feedforward network, the result at each time step is directly dependent on the accuracy of the result in the preceding time step. If the model is designed based on a coarse time resolution, the abrupt changes might not be caught, resulting in generating an error, which will be fed to the next time step and might get aggregated thereafter.
**Sampling Methods**

There is no doubt that the more information is taught to network, the higher expectation one can have from it in terms of prediction. Hence, in most cases when the model is small, the results might be better since the network inherits and learns about all elements (gridblocks) of the system. This is not a general statement, because if the variability of the system is high enough and the system is heterogeneous the same task can be accomplished by performing “sampling.” Sampling is concerned with the selection of a subset of gridblocks (records) from the whole system to estimate characteristics of the whole population.

*Figure 133. Recommended time resolution for the first three years compared with the resolution used in this study.*
Random sampling will not assign a specific weight to different clusters and might not be a good idea when sampled values are not evenly distributed in the population. Sampling based on static values, such as porosity and permeability, might exist as one option, but it might not be plausible to select one of these values as the core of the sampling procedure. On the other hand, the combination of different factors causes a dynamic response in the reservoir, and when the objective is to model those dynamic responses it is best to perform the sampling based on those values.

In this work, different methods of sampling were tried, and finally one was selected based on the histogram of pressure values at initial time. Five clusters (bins) were assorted. Higher weights were appointed to the bins with lower samples to establish a good representative of the whole data records.

Although not tried in this work, it is recommended to perform the same sampling method on different properties (based on the output of the SRM) or even at different time steps. This will generate a different database for each property or time step and will facilitate transferring more information to the network.

**Similar Governing Physics**

Good thought should go to building the database. It is crucial not to “confuse” the network. In the current work the main idea is to build an optimization base for WAG process by replicating the numerical reservoir simulator. The physics of miscible CO$_2$ EOR in a WAG process is different from a continuous gas or water injection; thereby the response of the reservoir is totally different for each of them. The idea behind considering two training scenarios based on continuous water and gas injection was to take the extreme limits into account; be that as it may, this does not seem to be a good idea because it can confuse the network by presenting values that are out of the range of other training scenarios (e.g. gas production rate).

**Speed improvement**

The focus of this work was a feasibility study of coupled SRM. The bulk of the time spent on this project was related to building the numerical reservoir model, dataset preparation, and building the required networks. Making the coupled SRM speedier was not the target of this study. As mentioned before, the run time of each model in the SRM is 15 seconds, out of which 10 seconds is related to SRM deployment at each step. If time is an issue in optimization it can simply be addressed while building the networks used in the SRM, by considering more nodes in the output layer of the neural network. In this study, it is possible to build only one neural network that has pressure, global mole fraction, and water and oil saturation at each step. This will cut down the
number of networks in the SRM ensemble and make the SRM almost 4 times faster (the well-based SRM is much faster in comparison to grid-based). Doing so can result in almost 8 hours to a bit more than 3 days for running 100 to 1,000 models.

**ERROR PROPAGATION**

One more decisive factor in the accuracy of the results is the inputs of each network.

There are two different methods for training the networks. In this study, the cascading method was used, i.e. the output of each network is the input of the succeeding one. The other method is to train standalone networks based on the information from a single time step. In this method, the information of the initial time step is fed to each network and the data at the specific time step is calculated. There is a trade-off between a low training and blind set deployment results while using any of these methods.

In cases that the changes in the system are frequent and abrupt, the standalone model might not work due to deficiency in information transformation.

While cascading might end in highly accurate results in one case, it might diminish the accuracy in another case depending on the inputs. The more dynamic inputs selected while training the networks, the lower the training error will be but this is not necessarily true during deployment. Every time SRM is deployed a percentage of error is introduced in the results. This initial error depends on the goodness of the trained network. This initial error will be propagated in the consequent network as the outputs of the preceding networks are the input of the next. This error can be aggregated as the sequence of the times gets longer. Henceforth, while selecting the inputs it should be kept in mind to select the most influential dynamic properties.

While performing the SRM coupling, this might get even more accentuated. In the coupling process the error is not only propagated through one network to another in each type (grid-based or well-based) it is transferred from one type to another as well. It cannot be emphasized enough that the selected inputs out of the spatio-temporal database have the key effect in this process.

If coupled SRM does not result in a good output due to error propagation, separate grid-based and well-based SRMs can be considered.

**INPUT DATA SELECTION**

Data selection can be a challenging and intricate task. In spite of everything, a neural network is only as good as the input data used to train it. Missing important data can have a very significant effect on neural network's performance. On the other hand,
using excessive amount of information or redundant data might end in confusing the network. Developing a feasible neural network application can be substantially more difficult without a solid understanding of the problem domain. When selecting input data, the physics of the phenomenon should be kept in mind.

6.2. **SRM vs Other Proxy Modeling Techniques**

Surrogate reservoir model has proved to outperform other proxy models in many cases and can be considered as a paradigm shift. The tutor to SRM is a numerical reservoir simulator. This ensures that SRM is trained to honor the physics and the first principles of fluid flow through porous media even in the case of heterogeneous and complex models. The production of fluids and the pressure gradients throughout the reservoir is inculcated in the education of SRM, making it a smart replica of a full-fledged reservoir simulation model [197].

Going through the literature one faces some studies in which AI-based models have been referred to as techniques which will not work under circumstances or at the best level have been specified as black boxes. Most of these arguments initiate from misusing this tool or the privation of enough understanding of the technique and its engineering application.

Zubarev [198] performed one of these studies. As presented by Mohaghegh et al. [79], this is not the first time that a technology has been misused, consequently misjudged, and prematurely dismissed. Going through the way neural network has been used for building proxy model in Zubarev’s study [198], based on the brief explanation provided, explains a lot regarding why it did not work. There is a common mistake to treat neural networks merely as a regression tool, which obviously sets it for failure. It should be noted that neural networks should be treated as part of a larger toolset, which attempts to observe, learn, and generalize. Competent use and deployment of artificial intelligence and data mining (as a predominant discipline), requires a certain comprehension of machine learning activities.
Table 21. The properties of some of the proxy models in literature.

<table>
<thead>
<tr>
<th>Proxy Model</th>
<th>Reservoir Model</th>
<th>Proxy Type</th>
<th>Dimension</th>
<th>Phase</th>
<th>High-Order Model Size</th>
<th>Variables</th>
<th>Application</th>
<th>Runs</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow Network</td>
<td>Black Oil</td>
<td>WB</td>
<td>1-D, 2-D</td>
<td>2-P</td>
<td>max 13,200 GBs</td>
<td>BHP</td>
<td>Optimization (NPV)</td>
<td>10</td>
<td>[49]</td>
</tr>
<tr>
<td>(ROM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>max: 2 Winj, 6 P Wells, 400 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCE</td>
<td>Compositional</td>
<td>GB</td>
<td>Homogenous</td>
<td>2-D</td>
<td>986 gridblocks, 30 days</td>
<td>φ, K</td>
<td>Uncertainty Quantification</td>
<td>100</td>
<td>[14, 15, 50]</td>
</tr>
<tr>
<td>POD (ROM)</td>
<td>Black Oil</td>
<td>GB, WB</td>
<td>2-D</td>
<td>2-P</td>
<td>2,025 gridblocks, 2 horizontal</td>
<td>Inj. Rate allocation at each well segment (grid)</td>
<td>Optimization (NPV)</td>
<td>1</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>W (1 Winj &amp; 1 P), 949 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POD/TPWL (ROM)</td>
<td>Black Oil</td>
<td>GB, WB</td>
<td>3-D</td>
<td>2-P</td>
<td>20,400 gridlocks, 4 P Wells, 3,000 days</td>
<td>BHP</td>
<td>Optimization (NPV)</td>
<td>2</td>
<td>[51]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POD/TPWL (ROM)</td>
<td>Compositional</td>
<td>GB, WB</td>
<td>3-D</td>
<td>2-P</td>
<td>max 52,800 GBs</td>
<td>BHP</td>
<td>Optimization (NPV)</td>
<td>3</td>
<td>[55]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>max: 2 G Winj, 6 P Wells, 1,000 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Surface</td>
<td>Compositional</td>
<td>GB</td>
<td>3-D</td>
<td>2-P</td>
<td>64,000 GBs</td>
<td></td>
<td></td>
<td>54</td>
<td>[56]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Proxy for 1 GB at one time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUT</td>
<td>Compositional</td>
<td>GB</td>
<td>Hom.</td>
<td>2-D</td>
<td>10,000 GBs, 1000 yrs Post-Inj.</td>
<td>φ, K</td>
<td>Risk Assessment</td>
<td>300</td>
<td>[54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POD (ROM)</td>
<td>Black Oil</td>
<td>GB</td>
<td>2-D</td>
<td>2-P</td>
<td>8 GBs, 1 Winj, 1 P</td>
<td>Inj. Rate allocation at each well segment (grid)</td>
<td>P, Sw Prediction</td>
<td>1</td>
<td>[31]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proxy Model</td>
<td>Reservoir Model</td>
<td>Proxy Type</td>
<td>Dimension</td>
<td>Phase</td>
<td>High-Order Model Size</td>
<td>Variables</td>
<td>Application</td>
<td>Runs</td>
<td>Ref.</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>----------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------</td>
<td>-----------------------</td>
<td>------------------------------------------------</td>
<td>--------------------------------------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>POD/DEIM (ROM)</td>
<td>Black Oil</td>
<td>GB,WB</td>
<td>2-D</td>
<td>2-P</td>
<td>2,000 GBs</td>
<td>Oil rate, Sw estimation</td>
<td>1,000 days</td>
<td>1</td>
<td>[34]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black-box Stencil Interpolation Method (BSIM)</td>
<td>Black Oil</td>
<td>GB</td>
<td>2-D</td>
<td>2-P</td>
<td>13,200 GBs</td>
<td>P estimation</td>
<td>1WInj, 4P, 50 days</td>
<td>1</td>
<td>[57]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBF (SAGD)</td>
<td>Thermal</td>
<td>GB</td>
<td>2-D</td>
<td>3-P</td>
<td>17,181 GBs</td>
<td>BHP Field Cum Prd.</td>
<td>3 pair of wells, 10 years</td>
<td>200</td>
<td>[59]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>Compositional</td>
<td>WB</td>
<td>2-D</td>
<td>3-P</td>
<td>1 well (I or P)</td>
<td>Rate, Vol. Period</td>
<td>screening/optimization</td>
<td></td>
<td>[61]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td>Black Oil</td>
<td>WB</td>
<td>3-D</td>
<td>3-P</td>
<td>4,5,7,9 spot</td>
<td>Rock, fluid properties, constraints</td>
<td>screening/optimization</td>
<td>200</td>
<td>[62]</td>
</tr>
<tr>
<td>SRM</td>
<td>Compositional</td>
<td>GB,WB</td>
<td>3-D</td>
<td>3-P</td>
<td>3,622,190 GBs, 40WInj, 39GInj, 75 Prd</td>
<td>W &amp; G Inj,Rate,Inj. Duration, (Vol. &amp; WAG Ratio), BHP</td>
<td>Optimization, Uncertainty Quantification</td>
<td>11 to 13</td>
<td>Current Work</td>
</tr>
</tbody>
</table>
Table 21, lists the properties of some of the proxy models used in the literature. All of these models have been applied on synthetic models that at best were partially based on some real data and none were performed on a real field scale. Especially when the focus is on grid-based modeling, the models are at most 2-D and deal with 2 phases. Some of the models have not been tested for robustness, and the rest have been tested on cases very similar to the training scenarios.

The SRM developed in this study has numerous advantages over other proxy modeling techniques explained in the literature review section of this study.

- The typical procedure for building some of the proxy models (e.g. reduced physics models), start with simplifying the physics. In general, the reduced physics model tends to lessen physical effects by reducing the complexities or heterogeneities of the full-physics model. Per contra, there is no need to approximate the existing simulation model to develop an AI-based proxy model. SRM honors the principle of the physics of the phenomenon and does not necessitate any changes in the physics or geology of the reservoir.

- While many of the proxy modeling techniques focus on simplifying the problem by making the reservoir more homogenous or decreasing the uncertainty, or at best using scalar heterogeneity, SRM benefits from the existence of the heterogeneity in the reservoir model. Higher heterogeneity results in requiring a smaller number of training runs, as each record of data is conveying a large amount of information to the network.

- Many of the proxy models (e.g. POD-TPWL) are intrusive and require modification to the reservoir simulation tools, which makes them impractical in combination with commercial reservoir simulators that do not provide open access to mathematical models. SRM is a non-intrusive model that can be used in conjunction with any kind of reservoir simulation tool.

- Proxy modeling methods, such as response surface, require several hundreds of simulation runs for their development. This number directly depends on the number of uncertain parameters. The number of simulation runs required for the development of an AI-based proxy model is at least an order of magnitude less than traditional proxy models by using the spatio-temporal database. Depending on the problem at hand, the number of runs for building an SRM can vary from 5 to 15.

- When it comes to flexibility in terms of input variables to the model, SRM overtakes other proxy models in the number of the parameters that it can handle. In fact, the more the variables, the more information content in the
dataset. Contrarily, the reported proxy models in the literature are able to handle a limited number of variables and are not as flexible.

- The computational expenses and storage constraints related to some proxy models (e.g. POD) make them impractical for real field applications. SRM has a very small computational footprint and does not suffer from creating huge files, which cause storage problems.

- Beyond representing the production profiles at each well individually, an AI-based proxy model can replicate, with high accuracy, the pressure and saturation changes at each gridblock. There are a few techniques that make this possible. These techniques have been tried only on very small, mostly 2-D, 2-phase models. They (e.g. POD) usually use a handful of simulation runs (1 to 3) due to computational and storage restraints, which leads to minimal information based on snapshots and can cause high errors in predicting the flood fronts.

- Time discretization is a procedure that should be performed in many cases while performing proxy modeling. The outputs are then acquired by interpolation between the discretized times. The SRM can go around this procedure since the discretization is not a must in SRM development, and can be eliminated through building the model for the whole time interval being modeled.

- The simplicity of application in SRM and the complexity of the math used to accomplish results in most ROMs highlight the cleverness of the approach used in SRM.

- Almost all proxy modeling methodologies presented in the literature have been tried on small 2-D and 2-phase models, mostly synthetic, with a few wells. The fact that SRM can be applied on huge 3-D real field reservoir models, with multiphase flow makes it practical and superior to other techniques.

That being said, it should be noted that the effective use of pattern recognition and AI techniques in the petroleum industry is not a trivial process. It requires vision in both the domain of reservoir engineering, as well as a substantial application of pattern recognition techniques; otherwise, the results could be quite disappointing [198].

There are some downsides to the surrogate reservoir modeling. The key part in developing an SRM is the spatio-temporal database. The more information content included in this database, the better the network training will be. In heterogeneous models, each record of data includes a large amount of information representing several runs in statistical models; homogeneity of the system diminishes this information content and consequently causes lower accuracy of SRM, unless the number of runs
increases. In general, the number of runs and heterogeneity of the model are inversely correlated.

Each SRM is case specific, i.e. the SRM built for SACROC cannot be used in another field with different geology, although the procedure can be repeated for making a new SRM. This rule applies to almost all of the proxy models.

The range of variables used in training cases is quite important. The SRM might not give a very good result, if the case it is tested on has properties out of the range of training cases. That is why in the scope of a given study, it is suggested to estimate the required ranges of all included parameters in advance and arrange the training cases accordingly. For instance, in this study the WAG injection schemes should be designed such that it covers the length and amount of water and gas injection in any of the cases to be studied in the optimization process. This is true for ROM and other proxy models as well.

Another limitation of SRM goes back to the software rather than the SRM itself. With the current versions of the software [199] 32,000 (sometimes up to 300,000) data records can be used for surrogate modeling. While this is adequate in many cases, if data records are more, then sampling on the data should be performed. The sampling, if not carried out intelligently, can bring about deficient information transfer to the neural network, preventing a meticulous training process.

It might be counter-intuitive to have a model that does not use the mathematical models of physics as we have formulated. That is the reason that some people refer to proxy models as black boxes. However, this is a plus in many cases, because instead of imposing our understanding of the physics, we let it be deducted from the data. This is quite helpful, specifically in the cases where enough knowledge is not available about the ongoing phenomena.

In general, to get a good result out of SRM, it should be assured that it has been taught properly.
REFERENCES


Simulation Symposium, Houston, Texas, 2005.


[61] E. Artun, ”Development of Universal Proxy Models for Screening and Optimization of Cyclic Pressure Pulsing in Naturally Fractured Reservoirs”.


[71] "What is Systems Theory?".


[75] S. Mohaghegh, "Converting detail reservoir simulation models into effective reservoir management tools using SRMs; case study – three green fields in Saudi


[121] R. Boomer, "CO2 Huff-n-Puff in a Light Oil shallow Shelf Carbonate Reservoir".


[132] Y. Ghomian, K. Sepehrnoori and G. Pope, "Efficient Investigation of Uncertainties in Flood Design Parameters for Coupled CO2 Sequestration and


282–324, April, 2010.


[165] B. B. Bennion, "Relative Permeability Characteristics for Supercritical CO2 Displacing Water in a Variety of Potential Sequestration Zones in the Western


[174] Y. Pamukcu, ”Simulating oil recovery during CO2 sequestration into a mature oil reservoir”.


[177] G. Wang, E. Knight and C. Fu, ”Determination Of Miscibility Pressure Direct


[181] A. Ghareeb, A. Nasr, Y. E.E., F. E.M. and M. H.M., ”Miscible Drive & Carbon Dioxide Flooding”.


[190] X. Wu, D. Ogbe, T. Zhu and S. Khatanair, ”Critical Design Factors and


[195] Intelligent Solutions Inc., "IMagine™".


[199] Intelligent Solutions Inc., "IDEA™".
APPENDIX – CASCADING AND COUPLING RESULTS

Scenario 2 - Grid Based Cascading
Scenario 5-Grid Based Cascading
Scenario 14 (Blind Set) - Grid Based Cascading
Scenario 2 - Grid Based Coupling
Scenario 14 (Blind Set) - Grid Based Coupling
Scenario 5-Well-Based Coupling
Well-Based Cascading
Well-Based Coupling