

Graduate Theses, Dissertations, and Problem Reports

2023

Leveraging Artificial Intelligence and Geomechanical Data for Accurate Shear Stress Prediction in CO2 Sequestration within Saline Aquifers (Smart Proxy Modeling)

Munirah Alawadh ma0108@mix.wvu.edu

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

Part of the Petroleum Engineering Commons

Recommended Citation

Alawadh, Munirah, "Leveraging Artificial Intelligence and Geomechanical Data for Accurate Shear Stress Prediction in CO2 Sequestration within Saline Aquifers (Smart Proxy Modeling)" (2023). *Graduate Theses, Dissertations, and Problem Reports.* 12191. https://researchrepository.wvu.edu/etd/12191

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Leveraging Artificial Intelligence and Geomechanical Data for Accurate Shear Stress Prediction in CO2 Sequestration within Saline Aquifers

(Smart Proxy Modeling)

Munirah Alawadh

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

In partial fulfillment of the requirements for the degree of

Master of Science in Petroleum and Natural Gas Engineering

Shahab D. Mohaghegh, Ph.D., Chair Samuel Ameri, M.S., P.E. Kashy Aminian, Ph.D.

Department of Petroleum and Natural Gas Engineering

Morgantown, West Virginia 2023

Keywords: Artificial Intelligence, Neural Network, Data Science, Smart Proxy Models, Reservoir Simulation, Petroleum Engineering, Geomechanics.

Copyright 2023 Munirah Alawadh

Abstract

"Leveraging Artificial Intelligence and Geomechanical Data for Accurate Shear Stress Prediction in CO2 Sequestration within Saline Aquifers (Smart Proxy Modeling)" Munirah Alawadh

This research builds upon the success of a previous project that used a Smart Proxy Model (SPM) to predict pressure and saturation in Carbon Capture and Storage (CCS) operations into saline aquifers. The Smart Proxy Model is a data-driven machine learning model that can replicate the output of a sophisticated numerical simulation model for each time step in a short amount of time, using Artificial Intelligence (AI) and large volumes of subsurface data.

This study aims to develop the Smart Proxy Model further by incorporating geomechanical datadriven techniques to predict shear stress by using a neural network, specifically through supervised learning, to construct Smart Proxy Models, which are critical to ensuring the safety and effectiveness of Carbon Capture and Storage operations. By training the Smart Proxy Model with reservoir simulations that incorporate varying geological properties and geomechanical data, we will be able to predict the distribution of shear stress.

The ability to accurately predict shear stress is crucial to mitigating the potential risks associated with Carbon Capture and Storage operations. The development of a geomechanical Smart Proxy Model will enable more efficient and reliable subsurface modeling decisions in Carbon Capture and Storage operations, ultimately contributing to the safe and effective storage of CO2 and the global effort to combat climate change.

Dedication

" To my Future Family, Thank you for being my inspiration even before our paths have crossed."

Acknowledgements

I begin with gratitude to Allah, the Lord of all creation, whose blessings have illuminated my path throughout this research journey.

My heartfelt appreciation goes to Dr. Shahab Mohaghegh, my esteemed advisor, for extending a warm welcome to the Laboratory for Engineering Application of Data Science (LEADS) group at WVU. His mentorship and unwavering support guided me through the complexities of my master's program, making this endeavor possible. Without him, I would not have become the first Kuwaiti female to pioneer the use of Artificial intelligence "Smart Proxy Modeling" in the oil and gas industry, turning my dream into reality.

I am deeply thankful to Prof. Samuel Ameri and Dr. Kashy Aminian for their kind reception into the master's program and their continual guidance and encouragement throughout my academic journey. Their expertise has been a source of inspiration.

To my twin sister, Moudhi Alawadh, I owe immense gratitude for being my constant companion through the highs and lows of this journey. Your unwavering support is a cherished gift. I extend my appreciation to my mother, Mona Almoharb, and my siblings Hadeel, Mohammed, Jacob, and Ali for their understanding and resilience during my periods of absence. Your sacrifices and love have been a pillar of strength.

Special recognition goes to the Computer Modeling Group, Inc., and Python for their invaluable contributions, providing essential resources and technical insights that significantly enhanced the quality of this research. Their collaboration was pivotal to the success of this study.

Lastly, I want to express my heartfelt thanks to all those who believed in my capabilities and contributed in various ways to the successful completion of this research. Your faith in me has been instrumental in reaching this milestone.

With profound gratitude,

Table of Contents

Abstracti
List of Figuresvvii
List of Tables
List of Symbols / Nomenclaturex
Chapter 1: Introduction 1
1.1 Problem Statement 1
1.2 Research Objective
1.3 Dissertation Outline
Chapter 2: Proxy Models Literature Review
2.1 Proxy Modeling Classifications6
2.1.1 Traditional Proxy Modeling6
2.1.2 Smart Proxy Modeling10
2.2 Exploring Smart Proxy Modeling in the Literature12
Chapter 3: Introduction to Artificial Intelligence15
3.1 History of Artificial Intelligence
3.1 History of Artificial Intelligence 15 3.2 Artificial Neural Networks 17
3.1 History of Artificial Intelligence 15 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17
3.1 History of Artificial Intelligence 15 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18
3.1 History of Artificial Intelligence 16 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 19
3.1 History of Artificial Intelligence 16 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 19 3.6 Datasets 20
3.1 History of Artificial Intelligence 16 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 19 3.6 Datasets 20 Chapter 4: Why Carbon Capture and Storage? 22
3.1 History of Artificial Intelligence 16 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 18 3.6 Datasets 20 Chapter 4: Why Carbon Capture and Storage? 22 4.1 General Information About CCS 23
3.1 History of Artificial Intelligence 16 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 18 3.6 Datasets 20 Chapter 4: Why Carbon Capture and Storage? 22 4.1 General Information About CCS 23 4.2 Storage process 23
3.1 History of Artificial Intelligence 15 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 18 3.6 Datasets 20 Chapter 4: Why Carbon Capture and Storage? 22 4.1 General Information About CCS 23 4.2 Storage process 23 4.3 Techniques for Storing CO2 24
3.1 History of Artificial Intelligence 16 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 18 3.6 Datasets 20 Chapter 4: Why Carbon Capture and Storage? 22 4.1 General Information About CCS 23 4.2 Storage process 23 4.3 Techniques for Storing CO2 24 4.4 Type of CO2 Storage 25
3.1 History of Artificial Intelligence 16 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 18 3.6 Datasets 20 Chapter 4: Why Carbon Capture and Storage? 22 4.1 General Information About CCS 23 4.2 Storage process 23 4.3 Techniques for Storing CO2 24 4.4 Type of CO2 Storage 25 4.4.1 Saline formations 26
3.1 History of Artificial Intelligence 15 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 18 3.6 Datasets 20 Chapter 4: Why Carbon Capture and Storage? 22 4.1 General Information About CCS 23 4.2 Storage process 23 4.3 Techniques for Storing CO2 24 4.4 Type of CO2 Storage 25 4.4.1 Saline formations 26 4.4.2 Oil and natural gas reservoirs 27
3.1 History of Artificial Intelligence 16 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 17 3.5 Type of Neural Network 18 3.6 Datasets 20 Chapter 4: Why Carbon Capture and Storage? 22 4.1 General Information About CCS 23 4.2 Storage process 23 4.3 Techniques for Storing CO2 24 4.4 Type of CO2 Storage 25 4.4.1 Saline formations 26 4.4.2 Oil and natural gas reservoirs 27 4.4.3 Unmineable coal seams 28
3.1 History of Artificial Intelligence 16 3.2 Artificial Neural Networks 17 3.3 The Structure of a Neural Network 17 3.4 Mechanics of Neural Networks 18 3.5 Type of Neural Network 18 3.6 Datasets 20 Chapter 4: Why Carbon Capture and Storage? 22 4.1 General Information About CCS 23 4.2 Storage process 23 4.3 Techniques for Storing CO2 24 4.4 Type of CO2 Storage 25 4.4.1 Saline formations 26 4.4.2 Oil and natural gas reservoirs 27 4.4.3 Unmineable coal seams 28 4.4.4 Basalt formations 28

4.5 CO2 Storage Characteristics	29
4.6 Trapping CO2 in The Subsurface	29
Chapter 5: GEOMECHANICS	30
5.1 Importance of Geomechanics in CO ₂ Sequestration	30
5.1.2 Key Objectives and Challenges in Geomechanics	31
5.2 Types of stresses	31
5.2.1 Mohr's circle and stress transformation	32
5.2.2 Relationship between stress and strain (Hooke's law)	33
5.3 Rock Mechanics	34
5.3.1 Overview of rock properties and classification	34
5.3.2 Failure criteria of rocks (Mohr-Coulomb criteria)	35
5.3.3 Rock behavior under different loading conditions	35
5.3.4 Laboratory testing techniques for rock characterization	35
5.3.5 Young's modulus and Poisson's ratio	35
5.4 Geomechanical considerations for site selection and well design	36
5.5 Risk assessment and mitigation strategies for geomechanical hazards	36
Chapter 6: Reservoir Simulation Modeling	37
6.1 Overview	37
6.2 Description of Reservoir Geological Structures	37
6.3 Implementation of Geological Reservoir Realizations	
6.4 CO2 Injection Design (Number of Injector, Injection Time)	41
Chapter 7: Methodology	42
7.1 Review	42
	43
7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps	
7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps	44
7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps 7.3 Preparing Data for Artificial Neural Networks 7.4 Data Processing	44 45
 7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps 7.3 Preparing Data for Artificial Neural Networks 7.4 Data Processing 7.5 Features Generation and Engineering 	44 45 45
 7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps 7.3 Preparing Data for Artificial Neural Networks 7.4 Data Processing 7.5 Features Generation and Engineering 7.6 Spatio-Temporal Dataset Construction 	44 45 45 48
 7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps 7.3 Preparing Data for Artificial Neural Networks	44 45 45 48 49
 7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps 7.3 Preparing Data for Artificial Neural Networks	44 45 48 49 50
 7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps 7.3 Preparing Data for Artificial Neural Networks 7.4 Data Processing 7.5 Features Generation and Engineering 7.6 Spatio-Temporal Dataset Construction 7.7 Data Partitioning 7.8 Training Process of Artificial Neural Network and SPM. 7.9 Error Measurements 	44 45 45 48 49 50 52
 7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps 7.3 Preparing Data for Artificial Neural Networks	

Chapter 8: Results and Discussions	55
8.1 Overview	55
8.2 Reservoir Shear Stress Distribution Results	55
8.2.1 Training Results of Shear Stress IJ, IK, and JK	56
8.2.2 Blind Results of Shear Stress IJ, IK, and JK	86
8.3 Mean Absolute Error Output Plots Comparison	265
Chapter 9: Conclusions and Recommendations	266
9.1 Conclusions	266
9.2 Recommendations	266
Appendix	267
Reference	272

List of Figures

Figure 1: Roller Coaster of Success and Setbacks of Artificial Intelligence16
Figure 2: Neurons in the human brain17
Figure 3: Artificial Neural Networks
Figure 4: Land and Ocean Temperature Departure from Average Jan-Dec 2022 (with respect to
1991-2022 base period)
Figure 5: Carbon capture and storage three major process
Figure 6: Illustration of Pressure Effects on CO2 (based upon image from CO2CRC). The blue
numbers show the volume of CO2 at each depth compared to a volume of 100 at the surface25
Figure 7: Stratigraphic column showing tertiary and cretaceous formations at Citronelle field27
Figure 8: Illustration of Normal or Shear Stresses Acting on Rocks beneath the Earth's Crust32
Figure 9: Representation of Normal Fault Regime Formation, Vertical and Horizontal Effective
Stress
Figure 10: presenting a three-dimensional perspective of the reservoir model's geometry and the
relative locations of the injection wells. Keep in mind that the thickness of each layer exhibits
substantial variations along the depth
Figure 11: Shear stress IJ Direction 3D view from GEM results40
Figure 12: the simulation study illustrates the designated timesteps for injection and post-injection
phases. These time intervals were thoughtfully chosen to capture the critical stages of the
reservoir's response to injection and to monitor its behavior over an extended period after injection
concludes41
Figure 13: The general workflow utilized for SPM in this study42
Figure 14: Flowchart shows the design of the three ANNs of the Smart Proxy Model44
Figure 15: Distance to the boundary and injection wells from the Focal Cell47
Figure 16: Schematics showing the implementation of tier cells48
Figure 17: displays a comparison between the predicted Smart Proxy Model Shear Stress results
for Model IJ and the actual CMG results, along with the corresponding error percentages, across
all layers for train run-01 on January 1, 2050. The arrangement from left to right includes the CMG
result, SPM result, and Percentage Error66
Figure 18: displays a comparison between the predicted Smart Proxy Model Shear Stress results
for Model IK and the actual CMG results, along with the corresponding error percentages, across
all layers for train run-01 on January 1, 2050. The arrangement from left to right includes the CMG
result, SPM result, and Percentage Error76

Figure 19: displays a comparison between the predicted Smart Proxy Model Shear Stress results for Model JK and the actual CMG results, along with the corresponding error percentages, across all layers for train run-01 on January 1, 2050. The arrangement from left to right includes the CMG Figure 20: Comparison of Predicted SPM Shear Stress IJ Results with Actual CMG Results and Corresponding Error Percentage Across All Layers for Blind Run-13 on January 1, 2050. (Displayed from left to right: CMG Result, SPM Result, and Percentage Error).......95 Figure 21: Comparison of Predicted SPM Shear Stress IK Results with Actual CMG Results and Corresponding Error Percentage Across All Layers for Blind Run-13 on January 1, 2050. (Displayed from left to right: CMG Result, SPM Result, and Percentage Error).....104 Figure 22: Comparison of Predicted SPM Shear Stress JK Results with Actual CMG Results and Corresponding Error Percentage Across All Layers for Blind Run-13 on January 1, 2050. Figure 23: Cross Section i Shear Stress model IJ results for all layers for timestep 01-01-2050......140 Figure 24: Cross Section i Shear Stress model IK results for all layers for timestep 01-01-Figure 25: Cross Section i Shear Stress model JK results for all layers for timestep 01-01-Figure 26: Cross Section j Shear Stress model IJ results for all layers for timestep 01-01-Figure 27: Cross Section j Shear Stress model IK results for all layers for timestep 01-01-Figure 28: Cross Section j Shear Stress model JK results for all layers for timestep 01-01-Figure 29: Mean Absolute Error for Geological Realization 13 Shear Stress IJ, IK, JK Output

List of Tables

Table 1: Typical range for Young's modulus and Poisson ratio for different types of rock	39
Table 2 Reservoir Parameters and Properties	.40
Table 3: Typical hyperparameters used in this study for one of the ANNs	51
Table 4: A list of selected features as input to the neural network a particular SPM's	.54

List of Symbols / Nomenclature

AI: Artificial Intelligence

- **ANN:** Artificial Neural Networks
- BHP: Bottom Hole Pressure
- **CMG:** Computer Modelling Group gLtd.
- CO2: Carbon Dioxide
- DARPA: Defense Advanced Research Projects Agency
- FGCP: Fifth Generation Computer Project
- **GEM:** Compositional and Unconventional Simulator
- PCA: Principal Component Analysis
- PCE: Polynomial Chaos Expansion
- PDO: Proper Orthogonal Decomposition
- ROM: Reduced Order Models
- **RPM:** Reduced Physics Models
- **RSM:** Response Surface Methodology
- SPM: Smart Proxy Modeling
- SRM: Surrogate Reservoir Modeling
- SWP: Southwest Regional Partnership
- SSIJ: Shear Stress IJ
- SSIK: Shear Stress IK
- SSJK: Shear Stress JK

Chapter 1: Introduction

This part of the study focuses on providing an overview of the research. It includes the problem statement, research objective, and dissertation outline.

1.1 Problem Statement

Over the past couple of decades, there has been significant development and enhancement in various geologic sequestration technologies aimed at mitigating greenhouse gas emissions. Among these technologies, the injection of carbon dioxide into subsurface formations has emerged as one of the most viable and promising approaches to significantly reduce anthropogenic CO2 emissions into the atmosphere. To achieve this important objective, deep saline formations have been recognized as particularly suitable for large-scale CO2 storage. When CO2 is injected into underground saline aguifers, various factors come into play to ensure the smooth and successful implementation of the process without facing any damageable failure. These factors include determining the geomechanics properties and structure of the storage area, determine the number of injection wells to be used, calculating the optimal amount of CO2 to be injected into each well, adjusting the choke sizes of individual wells to regulate the flow rate, and monitoring the overall production or injection from the field. Usually these crucial decisions are typically made by a reservoir engineer team, relying on numerical simulation models that enable them to examine and evaluate multiple production and injection scenarios simultaneously and efficiently. Reservoir simulation is considered one of the most essential tools in oil and gas reservoir development and management. It provides valuable insights and predictions regarding the future performance of the reservoir, as well as optimization strategies to enhance hydrocarbon recovery under various operating conditions. Oil and gas organizations invest significant resources, including substantial financial investments, in reservoir simulation to maximize the value of their assets. However, it is important to acknowledge that numerical reservoir simulation models can be computationally intensive, particularly in complex projects involving uncertainty assessment. The process of predicting shear stress behavior often takes up to approximately one week to obtain results. This computational complexity poses challenges, especially when timely decision-making is required. In response to these challenges and the increasing demand for efficient and accurate modeling of CO2 sequestration scenarios, researchers have explored alternative techniques that can replicate the roles and functionalities of numerical reservoir simulation models within a shorter time frame, while maintaining the same level of accuracy. One such technique is the utilization of artificial intelligence (AI) models, particularly the smart proxy model, which has been chosen as the technological model for this specific project. The smart proxy model leverages AI algorithms and techniques to provide rapid and accurate predictions of various CO2 sequestration scenarios. By employing advanced data-driven approaches and machine learning algorithms, the smart proxy model aims to achieve comparable accuracy to traditional numerical simulation models but within a significantly reduced time frame. This alternative modeling approach holds great promise in meeting the increasing modeling needs for CO2 sequestration, particularly as the volume of CO2 emissions continues to rise. In recent years, the research group led by Mohaghegh at West Virginia University has made significant progress in developing a range of Smart Proxy Models. These models have emerged as viable alternatives to traditional numerical reservoir simulation models.

1.2 Research Objective

The primary objective of this research is to develop an artificial intelligence model capable of effectively emulating the functionality of numerical reservoir simulation models. The main objective is to precisely forecast the spatial distribution of shear stress throughout the simulation period, encompassing every grid block within the reservoir. This will be accomplished through the utilization of injection and post-injection procedures. The findings of this study showcase the remarkable capabilities of Smart Proxy Models in addressing the practical challenges encountered in numerical reservoir simulation workflows. These models serve as rapid and precise tools for simulating reservoir behavior. Moreover, it possesses the capability to replicate the fluid flow dynamics across various time intervals throughout the injection and post-injection cycles of CO2 sequestration into a saline aquifer. The accomplishments of this dissertation, however, go beyond demonstrating the modeling approach for shear stress distribution in saline aquifer formations. The findings have larger ramifications since they might be applied to other time-consuming processes in reservoir management workflows. This dissertation proposes a unique technique for developing proxy models, which serve as helpful companions in the research and effective modeling of CO2 storage in geological and saline formations by leveraging the power of artificial intelligence and machine learning algorithms. This feature provides a comprehensive understanding of the system's behavior, facilitating more accurate predictions and analysis.

1.3 Dissertation Outline

This dissertation is structured into nine chapters, which are arranged in the following manner:

- Chapter 1 is dedicated to introducing the problem statement and elucidating the objectives of this study.
- In Chapter 2, an extensive literature review of reservoir proxy models is presented. The chapter begins by examining the fundamentals of proxy modeling and their categorizations. It then introduces the Smart Proxy Model, discussing prior research conducted in the field and outlining the respective advantages and disadvantages of each study.
- Chapter 3 provides a foundational overview of the field of artificial intelligence and delves into its various sub-field categories.
- Chapter 4 explores Carbon Capture and Storage (CCS) as a vital approach to combat global warming and reduce CO2 emissions. It covers the storage process, techniques, and various options for storing CO2 in different geological formations. The chapter emphasizes the importance of understanding subsurface conditions and geomechanical aspects for effective CO2 sequestration.
- Chapter 5 gives a detailed overview of geomechanics associated with the storage of CO2 deep down in sedimentary formations. Foremost, the chapter introduces the concept of CO2 storage deep down within rock layers, processes of geomechanics and the resulting operational issues. Next, the chapter addresses in detail the geomechanical aspects such as stress and strain of the reservoir, micro-seismicity and integrity of the well. Lastly, the chapter integrates geomechanical observations at some geological carbon storage field deployments.
- Chapter 6 focuses on the development of numerical reservoir simulation, which serves as the essential groundwork for implementing Smart Proxy Models. The chapter extensively covers the methodologies employed to generate additional geological models, incorporating geomechanical properties crucial for the project's success. Furthermore, it concludes by shedding light on the configuration of data ingestion, injection constraints, and the adoption of diverse time steps to ensure accurate and efficient modeling.

- Chapter 7 focuses on the methodology employed in the study, which outlines the essential
 procedures and steps taken to ensure the study's success. This methodology involves
 several key steps to achieve accurate predictions using the developed smart proxy. It
 begins with selecting a specific number of realizations and extracting relevant data from
 them while adding essential features to enhance prediction accuracy. The study relies on
 methods such as training, calibration, and blind validation to attain the expected results
 for the smart proxy model.
- In Chapter 8, the study presents an analysis and discussion of Shear Stress. This chapter specifically zooms in on a singular blind result and a solitary training result, both encompassing all layers at a particular time step. Additionally, the chapter provides insights into the cross-sectional view and offers a comparison of Mean Absolute Error for Shear Stress Outputs. To maintain clarity and streamline the presentation, the remaining blind and training run results for the initial 20 layers have been exclusively reserved for the Appendix.
- Chapter 9 serves as the conclusion of this study and offers final remarks and suggesting enhancements for the model.

Chapter 2: Proxy Models Literature Review

Over the past few decades, scientists and engineers in the oil industry have developed and implemented several proxy modeling techniques to address the challenge of execution timing for numerical reservoir simulations. These techniques aim to overcome the timing challenge by utilizing mathematical, statistical, or data-driven approaches to replicate the outputs of reservoir simulation models. In subsurface modeling and related earth science disciplines, proxy models serve as facilitators for numerical physics-based approaches, such as the CMG model. Reservoir simulation plays a crucial role in various aspects of petroleum reservoir management, including forecasting and analyzing reservoir communication. Consequently, reservoir development decisions nowadays heavily rely on feedback from reservoir simulation models. Reservoir simulation offers oil and gas companies a convenient and effective means to assess saline aquifer formations, monitor CO2 plumes during and after injection, and aid decision-making processes for optimization scenarios such as monitoring wells, injection volume, constraints, number of injection wells, and pressure maintenance. It involves constructing a mathematical representation of physical phenomena based on relevant observations and simplified assumptions. The evolution of numerical reservoir simulation from a basic decision-making tool to a crucial technology for complex production and development scenarios has been remarkable in the realm of reservoir engineering over the past few decades.

However, obtaining practical solutions for uncertainty analysis or optimization schemes requires developing reservoir simulation models and running them thousands of times under various operational and geological scenarios, considering the variability of uncertain parameters. This process consumes significant time and computational resources. In the oil and gas industry, where the concept of "time is money" holds significant importance, finding efficient solutions is essential. The Smart Proxy Model addresses this challenge by leveraging its unique capabilities. It combines reduced order modeling techniques, which simplify the complex physics of stress distribution, with data-driven modeling approaches that harness the power of available reservoir data which is the primary focus of this study. By capturing the underlying patterns and relationships between input parameters and stress distribution, the Smart Proxy Model offers an efficient means of shear stress prediction. This enables efficient exploration of large solution spaces for optimal or near-optimal solutions, as well as performing Monte Carlo simulations to quantify uncertainties associated with geological models (Mohaghegh, p: 19, 2023). The objective

is to evaluate the model's effectiveness, accuracy, and reliability in capturing the shear stress distribution phenomena. By successfully achieving this goal, the Smart Proxy Model can offer valuable insights into stress-related issues, assist in decision-making processes, and contribute to the optimization of reservoir development strategies.

In recent decades, the petroleum industry has witnessed the development and introduction of various proxy model techniques. These computational approaches encompass a wide spectrum, ranging from statistical methods like response surface modeling to mathematical techniques such as reduced order models, and even include data mining and artificial intelligence methodologies like smart proxy models. In the forthcoming sections, a more comprehensive examination of these techniques will be provided, delving into their intricacies. Moreover, the strengths and limitations of these techniques will be assessed.

2.1 Proxy Modeling Classifications

Proxy modeling can be categorized into different groups based on various criteria. These techniques can be further subdivided into statistical or mathematical models, depending on their respective approaches. Each of these techniques will have its own set of subcategories. The main two groups are Traditional proxy modeling and Smart Proxy Modeling.

2.1.1 Traditional Proxy Modeling

It simplifies the problem being solved in order to minimize computational time. This is in contrast to numerical simulation, which aims to approximate the solution rather than simplifying the problem. Traditional proxy modeling can be further divided into Reduced Order Models (ROM) and Response Surface Methodology (RSM).

a) Reduced Order Models (ROM):

It is an alternative solution to reduce the order of models and it is one of the traditional types of proxy models. The mathematical equations governing the physical phenomenon addressed in Numerical Simulation are simplified by establishing more fundamental and simplified relationships. That means the resolution of models is reduced in both time and space. However, it is important to acknowledge that within a specific paradigm, any reduction in computational time comes at a cost and the price paid by reduced order models (ROMs) is the accuracy of the models (Mohaghegh et al., 2015). There exist two primary methods to decrease the order of a model: reduce the resolution in time and space and reduce physics models (RPM).

1-Reduce the resolution in time and space:

In this method, the model undergoes a significant upscaling process to the extent that, in certain instances, the solution for each component of the model approximates the analytical solution despite its inherent limitations. To achieve an optimal balance, workflows have been devised to gradually enhance the spatial resolution by iteratively refining the initial coarse grid configuration (Mohaghegh, p:15, 2023).

2-RPM: Reduce Physics Models:

In this method, the model will concentrate on the physics of the problem in order to circumvent the computational time and it represents a mathematical proxy model. The application of reduced order model (ROM) approaches in numerical modeling of shale production has been prevalent in recent years by Wilson and Durlofsky (2012). In their work, they presented a reduced physics technique that initially incorporated dual porosity, dual permeability, and gas desorption. By employing the reduced physics model (RPM) approach along with tuning adjustments to align with gas production data, they simplified the complex full-physics model into a single porosity and permeability model. These approaches rely on physical proxies as substitutes for the detailed understanding of the underlying physics. As a result, ROM approaches address a different problem than the original intention of the numerical simulation model (Mohaghegh et al, 2015).

b) RSM: Response Surface Methodology

These models require a large number of simulation runs to be effective and it is known as Statistics-based proxy models. These models use statistical techniques to analyze and predict the responses generated by numerical simulation models. Examples of this specific type of methodology comprise of space designs and classical systems. Examples of the experimental design techniques consist of full factorial, factorial designs, and space-filling designs. The "curse of dimensionality" poses a problem for commonly used ROMs (Gholami et al. 2019; He, 2011; Chen et al. 2013). Because of the nature of the ROM models, the time required to develop it is often comparable to the time required to execute a numerical simulation. As a result, ROMs can be computationally expensive at times.

However, they face challenges associated with traditional statistics, especially when applied to problems with well-defined underlying physics. One common problem is the issue of correlation versus causality. Just because two variables are correlated, it doesn't mean that one variable is

causing the other. Another problem with statistics-based approaches is that they impose a predefined functional form, usually a polynomial, to analyze or model the data. While different functional forms can be tested to find the best fit, it becomes problematic when the data doesn't conform to a predefined form and exhibits multiple changing behaviors. Response surfaces struggle to establish well-defined and robust relationships between the variables in a numerical simulation model and the model's responses (Mohaghegh, p:16, 2023). Most of these techniques have only been applied to academic problems with a small number of cells. The real challenge will arise when they are used to build proxies for large-scale industry-based numerical models with millions of cells. Recently, some new statistics-based proxy models have emerged that use principal component analysis (PCA) as their core technology. Principal component analysis is a dimensionality reduction technique that is frequently used to reduce the dimensionality of large data sets. This analysis can be broken down into five steps. (Jaadi, 2021).

1-Standardize the range of continuous initial variables: This step standardizes the range of the variables.

2-Compute the covariance matrix to identify correlations: The goal of this step is to understand how the variables of the input data set are varying from the mean with respect to each other, or in other words, to see if there is any relationship between them.

3-Compute the eigenvectors and eigenvalues of the covariance matrix to identify the principal components: To find the primary components of the data, we must compute the linear algebra concepts of eigenvectors and eigenvalues from the covariance matrix.

4-Create a feature vector to decide which principal components to keep: The feature vector is simply a matrix with the eigenvectors of the components that has been chosen to keep as columns, making it the first step towards dimensionality reduction. In this step it has been chosen whether to keep all these components or discard those of lesser significance (of low eigenvalues), and form with the remaining ones a matrix of vectors that we call Feature vector.

5-Recast the data along the principal components axes: By multiplying the transpose of the original data set by the transpose of the feature vector, it is possible to use the feature vector created using the eigenvectors of the covariance matrix to reorient the data

from the original axes to the ones represented by the principal components (hence the name Principal Components Analysis).

These models, such as proper orthogonal decomposition (PDO) and polynomial chaos expansion (PCE), have been used to develop proxies for numerical reservoir models (Mohaghegh, p:16, 2023) PCE methods are categorized into intrusive and non-intrusive approaches. In the intrusive approach, approximations are substituted into the governing equation to obtain accurate coefficients. In the non-intrusive approach, coefficients can be computed through a small number of model simulations without altering the governing equations. The non-intrusive approach includes two main methods: projection and regression. Regression is more popular due to its lower computational requirements for multiple system inputs (Zhang and Sahinidis, 2013a, b, c). The POD method, which offers ways to find the best lower-dimensional approximations for the given data set, essentially provides an orthogonal basis for representing a given set of data in a certain least-squares optimal sense (Luo & Chen, 2019). However, these techniques are likely to converge with traditional response surfaces, as they are developed within the same computational framework.

c) Data Driven Models

Significant advancements and breakthroughs in computational intelligence, particularly in the realm of machine learning, have played a crucial role in expanding the possibilities and capabilities of empirical modeling. These developments have greatly enhanced our ability to model and understand complex systems based on empirical data. On the other hand, data-driven modeling involves the thorough examination of system data to uncover relationships among various variables, including inputs and outputs. Data-driven modeling serves as an inclusive field that encompasses various emerging approaches. This approach encompasses several key areas in its methodology, including:

 Data mining and knowledge discovery in databases are specialized fields that primarily deal with large databases and have significant applications in domains such as banking and financial services. Data mining is considered a subset of the broader field of knowledge discovery in databases. These disciplines rely on statistical methods and machine learning techniques to extract valuable insights, patterns, and knowledge from extensive datasets (Stedman & Hughes, 2021).

- Machine learning, originally classified within the realm of artificial intelligence, focuses on the fundamental principles shared by computational intelligence and soft computing.
- Artificial intelligence is a distinct and fascinating field that explores the integration of human intelligence into computers, aiming to understand how computers can mimic and incorporate human-like intelligence.
- Soft computing, a field closely related to computational intelligence, primarily emphasizes the application and development of fuzzy rule systems. It is a subfield that deals specifically with the utilization of fuzzy logic and fuzzy rule-based systems to address complex problems and make intelligent decisions (Soft Computing, 2023).
- Computational intelligence encompasses a range of techniques and approaches, including fuzzy systems, neural networks, and evolutionary computing. It extends beyond these specific areas to include other aspects of artificial intelligence and machine learning as a comprehensive field of study (What Is Computational Intelligence? - IEEE Computational Intelligence Society, n.d.).

Data-driven modeling techniques involve using a training set to establish the connection between system inputs and outputs, effectively capturing the system's behavior. After generating this relationship, a separate data set that was not part of the training process is employed to assess the model's predictive capabilities.

2.1.2 Smart Proxy Modeling

In 2006, Prof. Shahab Mohaghegh at West Virginia University initially developed the design for smart proxy models. Smart Proxy Modeling has been extensively studied and developed to deliver fast and accurate results, typically within seconds. It has been validated for both steady-state and transient problems, serving as a potential replacement for traditional approaches. Smart proxy models, also known as surrogate reservoir models, are advanced reservoir approximations that can accurately mimic the behavior of complete field models when subjected to changes in all input parameters such as reservoir characteristics and operating constraints. Smart proxy modeling is a combination of machine learning technologies that are trained to understand and replicate the complex physics of fluid flow in a hydrocarbon reservoir. It significantly reduces computational time without changing how numerical simulation is used to generate solutions for complex physical phenomena. It differs from other proxy models by not simplifying the physics or reducing the resolution of the simulation model, and it does not rely on pre-defined functional forms. It

preserves the physics and resolution of the original simulation model by utilizing the data extracted from the Numerical Simulation model during the training process. The model follows the principles of system theory, with an Input-System-Output structure, rather than being a statistical best-fit of the simulator responses. They integrate reservoir engineering, modeling techniques, and employ data mining and machine learning methods within the framework of surrogate reservoir modeling (SRM). According to Mohaghegh et al. (2012), SRM is a collection of interconnected neuro-fuzzy systems that are trained to adaptively learn fluid flow behavior from a reservoir simulation model comprising multiple wells and layers. The objective is to replicate the results of the reservoir simulation model with a high degree of accuracy in real-time operations.

In order to construct a smart proxy model, multiple reservoir simulation scenarios were required, taking into account different operational constraints and geological conditions. Several modeling scenarios were developed to capture these variations. The spatial-temporal database was created by gathering geological parameters and results from the desired simulation runs. The selection of input parameters and acquisition of data required for developing the smart proxy model relied on key performance indicators and expertise in petroleum engineering. The trained model was then employed in a blind run to validate its performance as an intelligent proxy model. Additionally, another grid-based proxy model was created to construct the well-production profile, enabling a more comprehensive assessment of field performance. Remarkably, both smart proxy models yielded highly precise results.

In order to train the Smart Proxy Model, we integrate the features of the geological and flow models, along with the operational limitations employed in the numerical simulator, with the corresponding output from the simulator. This amalgamation produces a comprehensive spatio-temporal dataset that encompasses vital fluid flow information specific to the field. This assimilated dataset serves as the foundation for training and fine-tuning the Smart Proxy Model. To ensure its accuracy, the model undergoes validation through blind simulation runs, where it is tested against novel data that it hasn't encountered previously. After successfully completing the validation process, the Smart Proxy Model becomes a valuable tool for reservoir management and strategic planning. Its application in these domains enables efficient decision-making and enhances the overall management of the reservoir. With its validated accuracy and reliable predictions, the Smart Proxy Model empowers stakeholders to make informed choices regarding the optimal utilization of resources, ensuring effective reservoir management and facilitating strategic planning initiatives.

Smart Proxy Model can be characterized as well-based (Mohaghegh et al. 2012; Shahkarami et al. 2014) or cell-based (Mohaghegh et al. 2012). (Gholami et al. 2019; Alenezi and Mohaghegh 2017). Numerous research studies have focused on well-based proxy modeling, there is a scarcity of literature discussing dynamic cell-based proxy modeling specifically for compositional simulations, such as CO2 sequestration into saline aquifers. The primary objective of a well-based SRM is to replicate reservoir production specifically at the well location, whether it involves production or injection activities. In contrast, a cell-based SRM offers the flexibility to simulate various dynamic reservoir parameters, including phase saturations, or fluid component composition, and pressure, at any desired location or time within the grid block. In contrast, the cell-based SRM provides the capability to model a wide range of dynamic reservoir parameters, including pressure, phase saturations, and fluid component composition, at any desired time and location within the reservoir grid block.

2.2 Exploring Smart Proxy Modeling in the Literature

The Laboratory for Engineering Application of Data Science (LEADS) at the Department of Petroleum and Natural Gas Engineering (PNGE) in West Virginia University has conducted numerous studies (Jalali, Mohaghegh, 2009; Kalantari-Dahaghi, Mohaghegh, et al. 2011; Mohaghegh, Amini et al. 2012; Amini, Mohaghegh, et al, 2014; Shahkarami, Mohaghegh, et al, 2014; Gholami, Mohaghegh, et al, 2014; Mohaghegh, Gaskari, et al. 2014; Haghighat, Mohaghegh, 2015; Alenezi, Mohaghegh, 2017; Alnuaimi, Mohaghegh, 2022) that specifically delve into the application of this technology. The following sections will discuss all studies that applied this technology into their models.

The smart proxy model has been applied to various field and reservoir challenges. An example of its utilization is demonstrated by Jalalai (2009), who employed the smart proxy model for reservoir simulation and uncertainty analysis in coal bed methane production. To conduct uncertainty analysis, Jalalai employed a Monte Carlo Simulation, which typically necessitates a large number of simulation runs. However, the developed smart proxy model enabled the generation of these simulation runs in a significantly shorter timeframe. Additionally, compared to conventional statistical techniques, a considerably smaller number of simulation runs were required to construct the response surface using the smart proxy model. In a pioneering study by Kalantari et al. (2011), the groundbreaking technology of the Smart Proxy Model was introduced and applied to shale gas reservoirs. This innovative approach marked a pivotal moment, revolutionizing the estimation of shale reservoir production profiles with a commendable level of

accuracy. Another study by Mohaghegh et al. (2012), an innovative SPM approach which revolutionizes reservoir analysis. Departing from traditional practices, this groundbreaking model was implemented at the gridblock level of a substantial oil field situated in the Middle East, specifically in Saudi Arabia, renowned for its vast array of wells. By leveraging the power of gridbased computations, the Smart Proxy Model showcased exceptional prowess in accurately capturing intricate changes in pressure and saturation across the reservoir, surpassing the capabilities of conventional simulation methodologies. This seminal research marked a significant milestone in reservoir engineering, paving the way for more advanced and precise reservoir characterization techniques. In a groundbreaking study conducted by Gholami (2014), a smart proxy model was developed to optimize reservoir injection processes. The research concluded that smart proxy modeling offers a computationally feasible alternative to numerical reservoir simulation. What sets this approach apart is its ability to accurately replicate pressure and saturation distributions throughout the reservoir, at the grid block level and at each time step, without compromising the underlying physics or resolution of the original numerical simulation model. Achieving a comprehensive understanding of pressure and saturation variations across the geological formation, particularly beyond the injection wells, is crucial for effective CO2 storage and various reservoir engineering procedures. This necessitates the integration of reservoir engineering, reservoir modeling, machine learning, and data mining within the framework of smart proxy modeling. By utilizing smart proxy modeling, the technology aims to learn the intricacies of fluid flow in porous media from numerical simulation model data, enabling the replication of results for different scenarios. This ability to replicate and reproduce diverse outcomes is a key advantage of smart proxy modeling. In Amini et al. (2014) study, they introduced a grid-based smart proxy model for a CO2 sequestration project. The aim of their research was to accurately estimate the gas saturation output at the grid level, specifically in the context of reservoirs undergoing CO2 injection. The results demonstrated that the smart proxy model achieved this goal with minimal error when compared to the numerical model. By employing this innovative approach, researchers were able to effectively simulate and predict the behavior of gas saturation during CO2 injection in the reservoir, providing valuable insights for CO2 sequestration projects. Also, in Alenezi and Mohaghegh (2017)study, they embarked on an exploration of the remarkable world of smart proxy modeling. Venturing into the SACROC unit field in Scurry County, Texas, they conducted a captivating investigation that unfolded at both the grid block and well level. What set their research apart was the ingenious employment of a cascade training and validation method, a unique approach where each time step's inputs were generated based on the output of the preceding time step, creating a dynamic flow of data that

fueled the smart proxy model. Implementing a cascade 22 training and validation method, they derived the inputs for each time step from the output of the preceding time step, excluding the first time step, until reaching the final time step. This ingenious feature endowed the model with the ability to self-feed from carefully selected dynamic data sources, granting it the power to evolve and adapt as the study progressed. The complexity of the production performance and the intricacies of its geological classification made the SACROC unit field an optimal testing ground to evaluate the prowess and effectiveness of the smart proxy model. These models are far more valuable now that they can be created in seconds rather than hours or days. In a recent study conducted by Alnuaimi (2022), an artificial intelligence model was developed to replicate the functions of a numerical reservoir simulation model. The primary objective of this model was to predict the pressure distribution and CO2 plume within a saline aguifer reservoir at different time points and grid blocks. The model consisted of a grid with dimensions of 211 x 211 in the X and Y directions, and 30 layers in the Z direction, with 28 layers comprising a mixture of conductive and non-conductive rock types. The top two layers were designated as shale barrier layers. The prediction of pressure distribution and CO2 plume involved injection and post-injection processes, with four injectors aimed at injecting approximately two million metric tons per year over a tenyear period. However, due to limited training data available for the neural network, the model faced challenges in accurately predicting saturation across future time steps, resulting in accumulation of errors and poor predictions.

Chapter 3: Introduction to Artificial Intelligence

Undoubtedly, you have likely encountered the terms "Artificial Intelligence" and may have even incorporated them into your daily life. However, have you ever contemplated the precise definition of artificial intelligence and its historical origins? This chapter aims to provide a clear understanding of "Artificial Intelligence".

Artificial intelligence involves emulating human intelligence by replicating the cognitive abilities of the human brain. Artificial intelligence is activated through Machine learning which is a scientific and technological discipline that involves the use of algorithms to enable computers to perform tasks without the need for explicit programming. The objective of Artificial intelligence is to empower machines to perform tasks such as decision making, analysis, and modeling. It encompasses various domains, including machine vision and expert systems, all driven by the common goal of developing machines capable of "thinking".

3.1 History of Artificial Intelligence

The foundations of what we now recognize as artificial intelligence and machine learning were established in the early 1950s when these concepts and ideas first emerged (Gold, 2023). According to the information provided on the Science in the News website by Rockwell Anyoha, the field of artificial intelligence (AI) experienced significant growth from 1957 to 1974, as demonstrated in (Figure 1). During the period from 1957 to 1974, the field of artificial intelligence (AI) experienced substantial growth. Advances in computer technology, including improved storage capacity, speed, and affordability, made computers more accessible. Researchers made progress in developing machine learning algorithms and understanding their application to specific problems. Early demonstrations, such as the General Problem Solver and ELIZA, showcased advancements in problem-solving and natural language interpretation. These achievements, along with the support of influential researchers, led government agencies like DARPA to fund AI research at various institutions. The government's focus was on developing a machine capable of transcribing, translating spoken language, and efficiently processing large amounts of data. Despite the optimism surrounding Al's potential, challenges remained in achieving goals like natural language processing, abstract thinking, and self-recognition. One significant obstacle was the limited computational power of computers, which hindered their ability

to store and process vast amounts of information. Consequently, AI research experienced a slowdown as funding decreased and enthusiasm waned for about a decade (SITNFlash, 2020).

In the 1980s, AI experienced a resurgence due to two key factors: the expansion of the algorithmic toolkit and increased funding. Deep learning techniques popularized by John Hopfield and David Rumelhart enabled computers to learn through experience. Edward Feigenbaum introduced expert systems that mimicked human decision-making, providing advice in specific domains and finding application in various industries. The Japanese government invested heavily in AI and expert systems through the Fifth Generation Computer Project (FGCP), aiming to revolutionize computer processing, logic programming, and AI development. Although not all ambitious goals of the FGCP were fully realized, it inspired a new generation of engineers and scientists (SITNFlash, 2020).

From the 1990s to the 2000s, AI continued to make significant progress even without extensive government funding or widespread public attention. Landmark achievements included IBM's Deep Blue defeating chess grandmaster Gary Kasparov in 1997, representing a significant milestone in artificial decision-making. Speech recognition software developed by Dragon Systems was implemented on Windows, further advancing the interpretation of spoken language. The exploration of human emotion also became a focus, exemplified by Cynthia Breazeal's robot, Kismet, capable of recognizing and displaying emotions. These developments demonstrated the potential of AI to tackle a wide range of problems (SITNFlash, 2020).



Figure 1: Roller Coaster of Success and Setbacks of Artificial Intelligence

3.2 Artificial Neural Networks

Artificial Neural Networks (ANNs) are an intriguing field of study, one that has been growing in popularity since its inception. ANNs are a powerful tool that have revolutionized the field of machine learning and artificial intelligence. They are used in a variety of applications, such as image recognition, speech recognition, natural language processing, and predictive analytics. The oil and gas industry has applied artificial intelligence (AI) to tackle a range of challenges, such as characterization of reservoirs, modeling of reservoir simulation, identification of wells suitable for stimulation, interpretation of well logs, drilling, and analysis of pressure transients. The successful implementation of artificial intelligence and machine learning in engineering applications began in 2004 in the field of petroleum engineering, specifically in numerical simulation, pioneered by Intelligent Solutions, Inc. Similarly, in 2016, the field of mechanical engineering, particularly computational fluid dynamics, witnessed the effective integration of artificial intelligence and machine learning techniques, spearheaded by West Virginia University (Mohaghegh, 2023, p. ix).

3.3 The Structure of a Neural Network

The concept of ANNs is rooted in the biological structure of neurons in the human brain by McCulloch and Pitts back to 1943 (Mohaghegh, 2000). The brain is a complex network of interconnected neurons that communicate with each other through electrochemical signals, forming the basis of neural networks. In (Figure 2) It is shown that a neuron typically consists of a cell body, axon, and dendrite that connect to other neurons through synaptic connections. Communication between neurons occurs via electrochemical signals that enter the cells through dendrites. When a neuron is stimulated by a characteristic input, it releases an output signal that travels through its axon. This process is known as "firing a signal," and it occurs when the electrical potential reaches a threshold value (Mohaghegh, 2018 p. 12).



Figure 2: Neurons in the human brain

3.4 Mechanics of Neural Networks

An artificial neural network (ANN) is composed of layers of neurons that are arranged in specific formations, as depicted in (Figure 3). These layers include input layer, hidden layer(s), and output layer. The input layer serves as the entry point for data and it is the initial raw information that is inputted into the network. In other words, the number of neurons in the input layer corresponds to the number of parameters presented to the network as input, while the same is true for the output layer. The hidden layer is influenced by the activities of the input units and the weights assigned to the connections between the input and hidden units. It performs intermediate computations and processes the information. Although hidden layers are sometimes called a "black box" within the network system, this does not mean that one cannot examine the function of those layers (Mohaghegh, 2018, p. 15). The output layer's outputs are determined by the activities of the hidden units and the weights associated with the connections between the hidden and output units. It produces the final results or predictions based on the processed information from the hidden layer.



Figure 3: Artificial Neural Networks

Furthermore, the mathematical description of the neuron model can be expressed by the following equation.

$$y = F(\sum_{i=0}^{m} w_i * x_i + b)$$
 Eq. 1

Where:

 x_i : is input value.

 w_i : is the weight value.

b: is bias.

- F: is a transfer function.
- y : is the output value.

3.5 Type of Neural Network

Neural networks can be classified in two major categories: supervised and unsupervised. Supervised learning is a widely adopted machine learning approach that relies on labeled datasets to train algorithms. In this method, the algorithm is guided by labeled input and output data to classify data or make predictions. On the other hand, unsupervised learning involves the analysis and clustering of unlabeled data without the presence of output or labeled data. This approach utilizes algorithms that can uncover hidden patterns or categories within the input data, even without prior knowledge of the expected output. They can be classified further as feed-forward networks and feedback networks (E, 2010). In a feed-forward ANN, information solely flows in a forward direction. The inputs and outputs are predetermined, and there are no connections or loops between units that could form. Data travels from the input layer to the hidden layer and finally reaches the output layer.On the other hand, feedback ANN models incorporate loops within the system. These loops establish connections from outputs directly back to inputs and hidden layers. The entire process can be categorized into five stages, namely:

- Transfer Function: The transfer function in a neural network calculates the output by taking the dot product or sum of products of the inputs and weights. It's important to note that even if the input values are the same, the output values will vary based on the assigned weights. These weights determine the strength of the connections between neuron layers(Artificial Neural Network, n.d).
- Activation Function: After the transfer function, a non-linear and differentiable activation function is applied to the result. This activation function introduces nonlinearity to the system, allowing the neural network to learn complex patterns and handle negative values in the data.

$$\theta = \sum x_1 w_{1j} + x_2 w_{2j} + x_3 w_{3j} + \dots + x_n w_{nj}$$
 Eq. 2

Different activation functions have different behaviors. For example, the tanh function outputs values between -1 and 1, while the leaky ReLU function retains some negative values. The logistic sigmoid function produces values between 0 and 1 based on the transfer function output. (Pramoditha, 2022)

$$f(\theta) = \frac{1}{1 + e^{-\theta}}$$
 Eq. 3

 Error Function: The output value calculated at the last hidden layer is compared to the actual output from the target dataset to calculate the total error using an error function like Error Sum of Squares (Course Hero, 2018).

$$ESS = \sum_{i=1}^{n} \frac{1}{2} (Y_{Target} - Y_{output})^2$$
 Eq. 4

- 4. Backpropagation: It is an algorithm used in neural networks for supervised learning. It calculates the gradient of the error function with respect to the network's weights, enabling the adjustment of weights to minimize the error. This algorithm operates by propagating the error backward through the network, updating the weights layer by layer. Backpropagation has gained popularity in deep learning applications, such as image and speech recognition, due to its efficiency and effectiveness in training neural networks (*Backpropagation | Brilliant Math & Science Wiki*, n.d.).
- 5. **Weight Optimization**: The weights symbolize the magnitude or significance of the information being conveyed, while the layer represents the pathway or direction through which the information is transmitted.

3.6 Datasets

The process of neural data processing involves dividing the database into three sets: training, calibration, and verification or validation dataset.

 Training: It refers to the dataset used to train the ANN. It consists of input samples (reservoir simulation data) along with their corresponding target outputs (e.g., desired reservoir behavior). During the training phase, the ANN learns from this data, adjusting its internal parameters to minimize the discrepancy between its predictions and the target outputs.

- 2. **Calibration:** It is a dataset which is majorly a watchdog, it is a separate dataset used to evaluate and analyze the ANN's performance during the training phase.
- 3. Verification or Validation dataset: Once the training phase is complete, the trained ANN needs to be evaluated on an independent dataset to assess its generalization capabilities and ensure it has not overfit the training data. It contains samples that were not used during training or calibration. The process helps gauge the ANN's reliability and provides insights into its real-world applicability.

During the training process, the weights between the neurons or processing elements are adjusted using an iterative process, but it's important to avoid over-training or memorization, which can lead to inaccurate generalization. To prevent this, it's common practice to periodically stop the training process and evaluate the network's performance on the calibration dataset. Each artificial neuron handles input signals, calculates a total for the combined input signals, and uses an activation function to determine the output. The weight of each input represents its connection strength to the neuron in the next layer, and the weight is adjusted during the training process. The inner product of the input and weight vectors determines the total input signal.

Chapter 4: Why Carbon Capture and Storage?

According to the National Centers for Environmental Information, the average temperature across global surfaces in 2022 was 1.55°F (0.86°C) higher than the 20th-century average of 57.0°F (13.9°C). This is the sixth-highest temperature among all years in the recorded period from 1880 to 2022. Moreover, it marks the 46th consecutive year since 1977 where global temperatures have been above the 20th-century average. These findings also indicate that the last nine years from 2014 to 2022 have been among the ten warmest years on record. In fact, all ten of the warmest years have occurred since 2010, highlighting a concerning trend of rising temperatures across the globe. These findings underscore the urgent need for immediate action to curb greenhouse gas emissions and address the ongoing climate crisis. In (Figure 4) shows the quantity of heat stored in the top 2000 meters of the ocean, known as the upper ocean heat content, reached an all-time high in 2022, exceeding the previous record established in 2021. The last four years, from 2019 to 2022, have witnessed the highest levels of upper ocean heat content, with the top four positions being held by them (*Assessing the Global Climate in 2022*, 2023).



Figure 4: Land and Ocean Temperature Departure from Average Jan-Dec 2022 (with respect to 1991-2022 base period)

As a result of human activities, such as burning fossil fuels, deforestation, and transportation, there has been an increase in carbon dioxide emissions, leading to global warming and climate change. Given this evidence, the significance of Carbon capture and storage (CCS) technology

is becoming more apparent in mitigating the impacts of climate change, as it can effectively reduce greenhouse gas emissions from industrial activities while keeping these activities operational. Nonetheless, the safety, sustainability, and cost-effectiveness of CO2 storage must be ensured to make Carbon capture and storage a viable solution in the fight against climate change.

4.1 General Information About CCS

Carbon capture and storage (CCS) is a vital process for reducing carbon emissions from industrial activities by capturing carbon dioxide (CO2) and storing it in deep underground geological formations. This chapter aims to provide general information about CCS projects and their process. It outlines the three main steps involved in the CCS process, and the five types of underground formations for CCS projects and storage characteristics. Additionally, this chapter will discuss the trapping of CO2 in the subsurface, as well as the Regional Carbon Sequestration Partnership (RCSP) Initiative projects as examples.

4.2 Storage process

Carbon capture and storage (CCS) is a critical technology that has gained significant attention as a solution to reducing greenhouse gas emissions. The Carbon capture and storage process involves three steps: capturing, transporting, and storing carbon dioxide (CO2) (Figure 5). Each step of the process is essential to ensure the safe and efficient storage of CO2.



Figure 5: Carbon capture and storage three major process

a) Capturing CO2

Capturing CO2 is the first step in the CCS process. This step involves separating CO2 from other gases produced during industrial processes such as power generation, cement manufacturing, and steel production. There are several methods for capturing CO2, including post-combustion capture, pre-combustion capture, and oxyfuel combustion. Post-combustion capture involves removing CO2 from flue gases after the combustion of fossil fuels. Pre-combustion capture involves burning fossil fuels in pure oxygen to produce a flue gas that is mainly CO2.

b) Transporting CO2

The second step in the carbon capture and storage process is transporting CO2 from the capture site to the storage site. CO2 is transported using pipelines, road transport, or ships. Pipelines are the most common method of transporting CO2, especially for large-scale CCS projects. Road transport and ships are used for smaller-scale projects or in areas where pipelines are not feasible. CO2 is transported in a supercritical state, which means it is in a fluid state under high pressure and high temperature and more details will be discussed in this chapter.

c) Storing CO2

The final step in the carbon capture and storage process is storing CO2. CO2 is stored in geological formations deep underground, including depleted oil and gas reservoirs, saline formations, and unmineable coal seams. The storage process involves injecting CO2 into these formations, where it is stored permanently. CO2 is stored in a supercritical state, which ensures that it remains dense and does not escape into the atmosphere.

4.3 Techniques for Storing CO2

According to the information provided by the National Energy Technology Laboratory. One method of storing CO2 is by injecting it into underground rock formations. CO2 can be stored in different states, including as a supercritical fluid, which is a state where it is subjected to high temperature and pressure. Supercritical CO2 is a state where CO2 is at a temperature above 31.1°C (88°F) and a pressure in excess of 72.9 atm (about 1,057 psi) (*Carbon Storage FAQs*, n.d.). Under these conditions, CO2 has properties of both gases and liquids, where it is dense like a liquid but has a viscosity like a gas. The main advantage of storing CO2 in the supercritical state is that it requires significantly less storage volume than storing it under standard temperature
and pressure conditions. The natural temperature and pressure in the Earth's crust increase with depth, as does the pressure of fluids such as oil, brine, and gas. At depths below 800 meters (about 2,600 feet), the natural temperature and fluid pressures are usually above the critical point of CO2 for most places on Earth (*Carbon Storage FAQs*, n.d.). Therefore, when CO2 is injected into rock formations at this depth or deeper, it remains in the supercritical state given the temperatures and pressures present. Storing CO2 as a supercritical fluid has several benefits. It reduces the amount of storage space required, which can reduce the overall cost of CCS. Additionally, it ensures that the CO2 will remain stored underground for long periods, as it is less likely to leak out of the storage formation. However, storing CO2 in rock formations requires careful selection of suitable sites to ensure that storage is safe, environmentally sustainable, and cost-effective.



Figure 6: Illustration of Pressure Effects on CO2 (based upon image from CO2CRC). The blue numbers show the volume of CO2 at each depth compared to a volume of 100 at the surface.

4.4 Type of CO2 Storage

Storing CO2 is the most critical step in the CCS process, and it must be done safely and permanently. Geological formations are the most promising storage sites, but they require careful selection and monitoring to ensure that CO2 does not escape into the atmosphere. The storage process also requires ongoing monitoring and verification to ensure the long-term safety of the storage site. There are several types of geologic formations that can be used for carbon storage, including saline formations, oil and natural gas reservoirs, unmineable coal seams, basalt

formations, and organic-rich shales. In this chapter, we will discuss each of these storage types in detail with real examples of CCS Project of each type of storage.

4.4.1 Saline formations

Saline formations are underground rock formations that contain highly saline water. These formations are widespread and can be found all over the world. The porosity of the rock formations enables them to store large amounts of CO2. When CO2 is injected into the saline formations, it dissolves in the saline water, forming a weak acid that reacts with the minerals in the rock, forming stable mineral compounds. This process ensures that the CO2 is permanently stored underground. Saline formations are considered an ideal storage type for CO2 because they are abundant, widely distributed, and have a low risk of leakage.

According to the information provided on the website (SECARB Phase III Site: SECARB Citronelle Projects (Anthropogenic Test), n.d.). The Citronelle Project is an example of carbon capture and storage (CCS) project that involves the injection of captured CO2 into a saline formation located in the Southern Mississippi region. In (Figure 7) shows a Stratigraphic column showing tertiary and cretaceous formations at Citronelle field. The project aims to demonstrate the technical and economic viability of saline storage for CCS. The CO2 is captured from a nearby fertilizer plant and transported via pipeline to the injection site. The project is part of the Southeast Regional Carbon Sequestration Partnership (SECARB) program, which is funded by the U.S. Department of Energy.

System	Series	Stratigraphic Unit	Major Sub Units		Potential Reservoirs and Confining Zones	
Tertiary	Plic- Pliccene		Citronelle Formation		Freshwater Aquifer	
	Miccone	Undifferentiated			Freshwater Aquifer	
	Oligocene	Vicksburg Group	Chicasawhay Fm. Bucatunna Clay		Base of USDW	
		Jackson Group			Minor Saline Reservoir	
	8	Claiborne Group	T	alahatta Fm.	Saline Reservoir	
	ve Paleco	Wilcox Group	Hatchetigbee Sand Bashi Marl Salt Mountain LS Porters Creek Clay		Saline Reservoir	
	ana a	Midway Group			Confining Unit	
	Upper	Selma Group			Confining Unit	
0		Eutaw Formation			Minor Saline Reservoir	
retac		Upper		Union Tune		Minor Saline Reservoir
8			Tuscaloosa Group	The R	Marine Shale	Confining Unit
sn			Lower Tune	Pilot Sand Massive sand	Saline Reservoir	
	Lower	Washita-	Dantzler sand		Saline Reservoir	
		Fredericksburg	E	Basal Shale	Primary Confining Unit	
Cretaceous		Paluxy Formation	'Upper' 'Middle' 'Lower'		CO ₂ injection	
		Mooringsport Formation			Confining Unit	
		Ferry Lake Anhydrite	1		Confining Unit	
		Donovan Sand	Rodessa fm. 'Upper' 'Middle'		Oil reservoir Minor Saline Reservoir	
				'Lower'		Oil Reservoir

Figure 7: Stratigraphic column showing tertiary and cretaceous formations at Citronelle field

4.4.2 Oil and natural gas reservoirs

Oil and natural gas reservoirs are underground rock formations that contain oil or natural gas. These reservoirs can also be used for CO2 storage. When CO2 is injected into these reservoirs, it mixes with the oil or natural gas, reducing the viscosity of the fluids and increasing their flow rate. This process is known as Enhanced Oil Recovery (EOR) and has been used for many years to extract more oil or gas from reservoirs. However, the use of oil and gas reservoirs for CO2 storage is limited to the amount of oil or gas present in the reservoirs, making it a less reliable storage option (*Carbon Storage FAQs*, n.d.).

According to the National Energy Technology Laboratory, the Cranfield Project is a carbon capture and storage (CCS) project located in Mississippi, USA is focused on storing carbon dioxide (CO2) in depleted oil reservoirs, which are classified as oil and natural gas reservoirs in geological storage formations. The project is a part of the Southeast Regional Carbon Sequestration Partnership (SECARB) initiative, which is a collaboration between industry, government agencies, and academic institutions to evaluate the feasibility of carbon capture and

storage in the region. The goal of the Cranfield Project is to inject CO2 into the depleted oil reservoirs to enhance oil recovery and store the remaining CO2 in the reservoirs. This project aims to demonstrate the effectiveness of carbon storage in oil and gas reservoirs and evaluate the potential for long-term CO2 storage in such formations (SECARB Phase III Site: SECARB Cranfield Project (the Early Test), Mississippi, n.d.)

4.4.3 Unmineable coal seams

Unmineable coal seams are underground coal formations that are not suitable for coal mining due to their depth or poor quality. These seams can also be used for CO2 storage. When CO2 is injected into the coal seams, it adsorbs onto the coal surface, displacing the methane gas that is present in the coal. This process is known as Enhanced Coalbed Methane Recovery (ECBM) and has the potential to increase the amount of methane that can be extracted from the coal seams. Unmineable coal seams are considered an attractive option for CO2 storage because they are abundant, widely distributed, and can potentially enhance coalbed methane recovery (Shi & Durucan, 2005).

4.4.4 Basalt formations

Basalt formations are underground rock formations that are composed of volcanic rock. When CO2 is injected into the basalt formations, it reacts with the minerals in the rock, forming stable mineral compounds. This process is known as mineral carbonation and ensures that the CO2 is permanently stored underground. Basalt formations are considered an attractive option for CO2 storage because they are abundant, widely distributed, and have a high storage capacity (Alzayani et al., 2023).

4.4.5 Organic-rich shales

Organic-rich shales are sedimentary rock formations that contain significant amounts of organic matter, such as kerogen. These formations can also be used for CO2 storage. When CO2 is injected into the organic-rich shales, it reacts with the organic matter, converting it into stable mineral compounds. This process is known as mineralization and ensures that the CO2 is permanently stored underground. Organic-rich shales are considered an attractive option for CO2 storage because they are abundant, widely distributed, and have a high storage capacity (Rahman et al., 2020).

4.5 CO2 Storage Characteristics

In order to ensure safe and effective long-term carbon dioxide (CO2) storage, several reservoir characteristics of a subsurface storage complex are studied. These characteristics include storage resource, injectivity, integrity, and depth. The storage resource refers to the space available to store large amounts of compressed CO2, and injectivity relates to the rate at which CO2 can be injected into the subsurface. The integrity of a storage complex refers to its ability to confine CO2 safely within a predetermined volume without any breaches, and the depth of the storage zone is important to ensure that CO2 can be injected as a supercritical fluid. All of these characteristics are carefully examined to determine whether a potential storage complex is suitable for CO2 storage.

4.6 Trapping CO2 in The Subsurface

Trapping is the process of confining the injected carbon dioxide (CO2) in the subsurface storage formation. There are four main mechanisms involved in trapping CO2 underground, each playing a crucial role in keeping it confined. The first mechanism is Residual Trapping, where CO2 remains trapped in the pore spaces between rock grains as the CO2 plume migrates through the rock. The second in Solubility Trapping, a portion of the injected CO2 dissolves into the brine water that is present in the rock's pore spaces. Thirdly, the mechanism is Structural Trapping which physically traps CO2 in rock formations. The rock layers and faults surrounding the storage formation prevent CO2 from moving out of the injection zone. Structural Trapping is the most effective trapping mechanism as it prevents CO2 from migrating laterally or vertically. Finally, Mineral Trapping refers to a reaction that can occur when CO2 dissolved in the rock's brine water reacts with the minerals in the rock. Over extended periods, this weak acid can react with the minerals in the surrounding rock to form solid carbonate minerals, permanently trapping and storing that portion of the injected CO2. Therefore, these four mechanisms are essential for keeping the CO2 trapped in the subsurface storage formation.

Chapter 5: GEOMECHANICS

Geomechanics is a multidisciplinary field that investigates the deformation behavior of rocks and soils in response to changes in temperature, stress, and pressure. It encompasses the study of pressure transmission through fluids within rocks and the transfer of effective stress across rock particle interfaces. To gather reliable data for geomechanical models various methods used like coring, seismic analysis, well testing (e.g., transient pressure analysis, hydraulic fracturing stress testing), and geophysics (e.g., acoustic emission) (Wikipedia contributors, 2023b). This scientific discipline plays a pivotal role in understanding a wide range of geological processes and engineering applications. In geomechanics, the behavior of rocks and soils is extensively studied to facilitate various critical activities, such as mineral excavation from subsurface layers, assessment of borehole stability, design of stimulation programs for reservoir engineering, and analysis of fluid flow beneath the earth's surface (Cook, 2023). By comprehending the mechanical responses of geologic materials to environmental and engineering factors, geomechanics provides valuable insights into the safe and efficient utilization of subsurface resources while ensuring the stability and sustainability of infrastructure projects.

5.1 Importance of Geomechanics in CO₂ Sequestration

Increasing greenhouse gases into the atmosphere is a growing concern as they are immensely contributing to global climate change. In this regard, research for both environmentally and economically sustainable ways of reducing CO2 emissions are underway. One of the effective ways of fighting against global warming and climate change is through Geological Carbon Sequestration which entails capturing and injecting CO_2 into geological formations for permanent storage. Even though this strategy has been proven efficient in preventing health and environmental threats, it has both economic and security implications. Geomechanics is a very critical consideration during the injection and long-term storage of CO_2 . Projects of sequestering CO_2 entail injecting CO_2 in a supercritical state into certain depths in highly porous and permeable formations (Feng & Zhang, 2022). It is thus important to eliminate or reduce leakages for a prolonged period of at least 1000 years. Understanding geomechanics is key herein since the trapping of CO_2 below the rock layers depends highly on the geological structure. For instance, during formation deposition, a complex geological structure forms. Besides, locations with low and high permeability determine how the fluid flows within the formation. The CO_2 injected into

the formation will either flow laterally or vertically until it reaches an impermeable or a low permeable caprock since the density of fluid formation is higher than that of CO_2 (Feng & Zhang, 2022). Low permeability structures or formations make up the physical traps of CO_2 . The varying density between CO_2 and groundwater results in an upward movement (Rutqvist, 2012). Afterward, in the wetting phase, the groundwater re-enters the pore spaces in the rock to replace the CO_2 in the weak wetting phase. As CO_2 gets replaced with groundwater, the percentage of CO_2 in the rock pores reduces significantly. Capillary or residual trapping is the mechanism that holds isolated CO_2 in a stable phase (Feng & Zhang, 2022).

5.1.2 Key Objectives and Challenges in Geomechanics

One of the key objectives of geomechanics is its use in mining science. It entails the investigation of deformation processes in the ground surface and rock mass for efficient management of subsoil and sustainable operation of underground and subsurface facilities. Subsoil facilities are permanent structures used for mineral mining. In addition, transportation lines as well as high-rise structures, water bodies, power generation points as well as city and industrial infrastructures are established in subsoils. For all these establishments, the underground rock masses and ground surface components are integral. In case of either man-made or natural calamities, ground and rock masses get subjected to deformations; consequently, the rock layers below the earth's surface are prone to instability risks. Advanced knowledge and understanding of stress-strain analysis of the rocks and the relevant technologies help a big deal in offering insight into the origin and causes of stress and strain patterns, as well as their variability (Sashurin, 2020).

5.2 Types of stresses

The accumulated weight of overlying rocks and plate collisions exert substantial forces on the rocks at a given depth. The size of the force varies depending on the area of exposure. When the force gets adjusted for the area over which it is distributed, then it is known as stress. Rocks are said to be under strain when applied stress causes the rocks to change in shape ("Chapter 13. Stress and Strain," n.d.). Rocks are acted upon by two categories of forces namely; normal stress and shear stress. The normal stresses act on the rocks at right angles while shear stresses act parallel to the rock surfaces. The normal stresses further fall under two forms of compression and tension. Compression is a force that squeezes on the rocks while tension is a force that pulls the rocks apart. As outlined in "Chapter 13 of Stress and Strain," n.d.), compression in the rock layers takes place when the plates collide with each other or get buried under other rock layers. On the other hand, when the rocks diverge, they are said to be experiencing tensional forces. For

instance, during continental drift, a continent begins to pull apart from the other due to tensional forces. When rock plates move side by side, under the characteristics of transform plate boundaries they experience shear stress. Rocks maintain an equilibrium state in a certain face when equal shear stresses act in the opposite direction. However, a positive shear force acting in an X phase and X direction causes the rocks to move in the X direction. Similarly, an extreme shear force acting in the negative Y plane in the negative X direction results in the rocks moving in the negative X direction. A positive Y plane of rocks being acted on by forces in the X negative X direction results in the rocks moving in the negative Y direction results in the rocks to move in the positive Y plane being pushed in the negative Y direction results in the rocks to move in the positive Y direction (Viewd Mechanical, 2020).



Figure 8: Illustration of Normal or Shear Stresses Acting on Rocks beneath the Earth's Crust.

5.2.1 Mohr's circle and stress transformation

Mohr's circle is a representation of the triaxial stress state in the core. In triaxial compression tests, the maximum principal stress gets applied along the cylindrical rock specimen's axis while the minimum overall stresses get applied on the specimen's lateral surface. When shear stress is equivalent to the shear strength, then shear damage takes place (Feng & Zhang, 2022). In a normal formation of fault regimes, the maximum overall stress represents the vertical stress whereas the overall horizontal stress represents the minimum stress as shown in figure 9 (a) below.



Figure 9: Representation of Normal Fault Regime Formation, Vertical and Horizontal Effective Stress.

The initial formation's Mohr's circle is indicated in Figure 9 (b) above and the changes in the Mohr's circle during the injection of CO_2 are represented by Figure 9 (c) above. During an increase in pore pressure, there is a greater variation in the maximum effective stress in comparison to the minimum effective stress. Eventually, the radius of the Mohr's circle reduces and slightly moves to the left which means that the shear stresses have reduced. Thus, during the initial phase of CO_2 injection, the Mohr's circle tends to move away from the damaged area hence increasing the stability of the faults and caprock. Afterward, the shear stress undergoes a continuous decrease with a corresponding injection of CO_2 until a balance between maximum and minimum effective stresses is reached (Feng & Zhang, 2022).

5.2.2 Relationship between stress and strain (Hooke's law)

In structural geology, stress and strain are closely related concepts. This relationship can get described using the elasticity concept. Perhaps, elasticity is the ability of the material to undergo deformation when subjected to stress and then return to the original size and shape when the stress gets released. The stress and strain relationship in an elastic material is linear; the resulting deformation is directly proportional to the induced stress. Hooke's law is used to describe this relationship. The law states that; the stress is proportional to the strain in a material with the elastic modulus as the proportionality constant.

Where;

 $\sigma = Stress$

E= Elastic (A measure of material's stiffness)

 $\varepsilon = Strain$

Nonetheless, this stress and strain linear relationship applies until a certain point is referred to as the yield point. The rocks begin to undergo plastic deformation beyond this point hence the nonlinear relationship. The level of plastic deformation depends on the amount and type of stressinduced, rock structure and material composition.

5.3 Rock Mechanics

5.3.1 Overview of rock properties and classification

On the basis of their mode of formation, rocks fall under three main classes namely; igneous, sedimentary, and metamorphic. Primarily, igneous rocks form by cooling and solidification of the magma from deep in the earth's layers when it reaches the earth's surface. Sedimentary rocks on the other hand form due to the accumulation of rock sediments in topographical sinks and tectonic basins. Lastly, metamorphic rocks form from exposure to extreme temperatures and pressures on pre-existing rock layers. This means that metamorphic rocks form deep down the earth's layers but get exposed due to epeirogenic movements or erosion (Mibei, 2014).

Rocks can further get classified based on various properties. Arguably, the properties of a rock refer to either lithologic or measurable aspects of the rock materials that can either get examined in hand specimens or through laboratory tests. Such properties include but are not limited to mineralogy, primary porosity, hardness, particle size, color, and permeability. Different rocks have different levels of hardness; the resistance of the rock to permanent deformation. Categories of hardness provide rational estimations of the rock material strengths for the classification of earth material as rock. This characteristic is important for designers of rock excavations such as auxiliary spillways when doing estimations on the location and alignment ("Chapter 4 Engineering Classification of Rock Materials," 2012). Different rocks also have different textures. For instance, crystalline is a textural term for rocks consisting of entirely interlocking crystals while pegmatitic rocks are coarse-grained crystals of greater than 10 mm diameter and glassy rocks are extrusive igneous rocks that cooled quickly without discrete crystallization. The permeability of the rock is another aspect that determines the classification of rocks. It is highly interlinked with the pore spaces of a particular rock. Perhaps, rocks with coarse gravel have a high rate of permeability as opposed to those with silt, clay, or shale particles ("Chapter 4 Engineering Classification of Rock Materials," 2012).

5.3.2 Failure criteria of rocks (Mohr-Coulomb criteria)

Shear failure remains to be the most common type of failure which occurs when shear stress acting on a caprock plane exceeds the shear strength at the plane of fault. This type of rock failure is identified using the Mohr-Coulomb criterion. It is because Mohr-Coulomb's criterion considers both minimum and maximum principal stresses, and underestimates the strength of the rock hence appropriate during designs.

5.3.3 Rock behavior under different loading conditions

As argued by Zhang et al., (2021), when rocks get subjected to an increased strain as the stress is kept constant, they undergo creeping. On the other hand, when the stress reduces at constant strain, the rocks undergo slackening. Whenever the induced stress exceeds the crack initiation stress, then the micro-cracks within the rocks expand at a steady rate. Stress exceeding damage stress results in accelerated inelastic strain which triggers unstable growth of micro-cracks.

5.3.4 Laboratory testing techniques for rock characterization

Laboratory testing of the rocks gets undertaken to determine the elastic properties and strength of intact specimens. It also aids in evaluating the degradation and disintegration potential of rocks. Such information is thus helpful in the design works of rock fills, both shallow and deep foundations, cut slopes, and rip-raps used in protecting shores. The mostly conducted laboratory tests for determining the strength of intact rocks include; compressive strength, direct shear, point load index, and the Brazilian test. The durability tests on the other hand include abrasion and slaking tests ("Laboratory testing and rock properties," 2011).

5.3.5 Young's modulus and Poisson's ratio

Young's modulus refers to the measure of stiffness of an elastic material. Perhaps, it is the ratio of stress to strain.

$$E = \frac{Stress}{Strain} = \frac{\sigma}{\varepsilon}$$
 Eq. 6

Rocks that have low Young's modulus have ductility characteristics whereas those with high Young's modulus have brittle characteristics. Brittle rocks always depict better targets of hydraulic fracturing as well as completion quality (Ma et al., 2016). When rocks as any other material get

stretched in one direction, they tend to compress in the perpendicular direction to the phase at which the force gets applied. Perhaps, poisson's ratio refers to the change in the width per the material's unit width to the change in length per unit length when subjected to strain.

Poisson's Ratio = $\frac{Transverse Strain}{Longitudinal Strain}$ Eq. 7 $v = \frac{-\varepsilon t}{\varepsilon l}$ $\varepsilon t = Transverse or Lateral Strain$ $\varepsilon l = Axial or Longitudinal Strain$ v = Poisson's Ratio

The Poisson's ratio gets represented by a negative sign whereas normal materials have a positive ratio. Nonetheless, Poisson's ratio is positive for tensile deformations and negative for compressive deformations ("Poisson's Ratio," n.d.). Both Poisson's ratio and Young's modulus are critical parameters in the design of hydraulic fractures (Ma et al., 2016).

5.4 Geomechanical considerations for site selection and well design

Modeling of fracture propagation extends its benefits to site selection and well design. Fracture length information obtained from the modeling techniques is useful in designing well patterns (Lee et al., 2022). Ground surface uplifts and micro seismic events are also useful in monitoring subsurface fluid flow which is critical in siting of wells.

5.5 Risk assessment and mitigation strategies for geomechanical hazards

A high injection rate of CO2 can cause the downhole pressure to exceed the pressure at the fracture hence inducing tensile fracture. If the pressure builds up and exceeds the normal range, the response can easily extend to the top of the target reservoir and sea bed surface. When the cap rock is impermeable, then shear failure at the interface is highly possible. Thus, the stability of the caprock becomes insecure. Eventually, CO₂ leakages resulting from buoyancy are highly probable hence posing an environmental threat (Song et al., 2023). Besides, CO₂ traps induce a variation in temperature or pressure which eventually results in buildup of stress or strain around the injection area hence causing either deformation, change in injectivity and permeability of rock layers. Also, the build-up of pore pressure due to CO₂ induction act as pathways for CO₂ and induce micro-seismic waves that can result in earthquakes and ground surface uplifting. Mohr-Coulomb failure however can be used to predict fault reactivation through evaluating the stability of existing faults (Song et al., 2023).

Chapter 6: Reservoir Simulation Modeling

6.1 Overview

It is necessary to construct a reservoir simulation that takes into account geological and petrophysical factors in order to comprehend the movement of fluids in subsurface reservoirs that contain porous materials. Numerical reservoir modeling is a widely used technique among petroleum engineers to simulate hydrocarbon reservoirs. By creating a model that accurately reproduces past performance, engineers can frequently predict the future behavior of the reservoir. In this study, numerical reservoir simulation was utilized to generate the essential data for the development and testing of an artificial neural network. Furthermore, these models are employed to devise the best field development strategies by simulating a variety of scenarios. In this study, numerical reservoir simulation was utilized to generate the essential data for the development and testing of an artificial neural network. The subsequent sections provide a detailed description of the process used to develop the reservoir model in this study.

6.2 Description of Reservoir Geological Structures

Computer Modeling Group software was the chosen tool for building the reservoir models, which were designed to mimic a hypothetical reservoir featuring spatially varying properties, exhibiting differences both within and between its geological layers. The Computer Modeling Group (CMG) is responsible for developing the Compositional and Unconventional Simulator (GEM), a sophisticated compositional simulator based on a general equation of state. The remarkable aspect of GEM lies in its all-encompassing feature set, which incorporates dual porosity modeling, support for horizontal wells, comprehensive well management tools, the ability to handle complex phase behavior, CO2 and miscible gas simulation, representation of gas condensate behavior, and the representation of volatile oil behavior (GEM, Compositional & Unconventional Reservoir Simulator Unconventional Reservoir Training, 2016). The CO2 injection and sequestration into an aquifer formation were replicated in this study using the simulator's dedicated CO2 module. The structural components of the reservoir model, namely the top, bottom, and layer thickness, were extracted from a previously constructed history match model at WVU (Haghighat, 2015) for the Citronelle field, which serves as a saline reservoir in Mobile County, Alabama, USA. The reservoir structure was obtained from a model created previously at LEADS at WVU in 2021 by Alaboodi.

The model's design comprises a grid layout of 125 x 125 blocks along both the X and Y axes. Its geological representation involves a total of 65 layers in the Z direction. Within these layers, 51 are composed of sand formations, interspersed with 14 shale or impermeable layers strategically positioned at the reservoir model's upper, middle, and lower segments. Importantly, layers 5 to 28 define the Upper Aquifer, while layers 35 to 61 pertain to the Lower Aquifer in the simulation model. For a more illustrative understanding of the reservoir's structure and the placement of wells, (Figure 10) serve as informative visual aids. In addition, the reservoir simulation model takes into account the heterogeneity caused by thickness variations within individual layers and across different layers. This authenticity enhances the applicability and value of the study's outcomes.



Figure 10: presenting a three-dimensional perspective of the reservoir model's geometry and the relative locations of the injection wells.

6.3 Implementation of Geological Reservoir Realizations

The process of incorporating geological reservoir realizations is discussed for training artificial neural networks (ANN) in constructing the Smart Proxy Model. To achieve this, a set of 20 geological models with similar broad porosity distribution patterns but distinct and varied distributions were created. The objective was to develop a realistic, heterogeneous reservoir with different characteristics across layers. Through the analysis of porosity and permeability data from

(Haghighat, 2015), it was identified that the reservoir model contained four permeability groups: very conductive perm, conductive perm, average perm, and tight perm. To generate meaningful permeability data, porosity-permeability correlations were used to populate permeability values in each grid cell of the model. The top portion of the Upper Aquifer exhibited a highly conductive permeability correlation in the top 8 sand layers, while the lower 16 sand layers had an average conductive permeability correlation. The layers in the Lower Aquifer were divided into Conductive Perm and Tight Perm correlations. As a result, the complex 65-layer model was simplified into two main reservoir intervals: the top interval with a high-permeability upper layer and an intermediate-permeability middle layer, and the bottom interval with a single relatively low-permeability layer. These intervals were separated by three very low-permeability shale formations, acting as barriers to vertical gas migration.

After creating the 20 geological models with distinct permeability distributions, the next step involved enhancing these realizations with additional geological properties. These additional properties, such as rock types, Young's modulus, Poisson's ratio, and stress values, were incorporated to gain a comprehensive understanding of shear stress in the reservoir before feeding the data into the Smart Proxy Model. As shown in Table 1, shale and sandstone exhibit a range of Young's modulus and Poisson ratio values, and these are the values that have been selected. The values for these geological properties were meticulously gathered through in-depth research and examination experiments, including insights from tutorials and default values from CMG (Computer Modelling Group) software. By incorporating these crucial parameters, the geological realizations were enriched, enabling a more accurate representation of the reservoir's behavior under different conditions. A detailed presentation of all these geological properties and parameters will be provided in (Table 1 and 2) will be followed with a 3D view of shear stress IJ direction from GEM results (Figure 11). This table will serve as a valuable reference for the Smart Proxy Model, contributing to its robustness and effectiveness in simulating the reservoir's complex behavior.

Lithology	Young's modulus $(GPa = 1 \times 10^3 MPa)$	Poisson ratio (-)	
Soft sandstone	0.7-6.9	0.2 - 0.35	
Medium sandstone	13.8-34.5	0.15 - 0.25	
Hard sandstone	41.4-68.9	0.1 - 0.15	
Limestone	55.1-82.7	0.3 - 0.35	
Coal	0.7 - 6.9	0.35 - 0.45	
Shale	6.9-68.9	0.28 - 0.43	

Table 1: Typical range for Young's modulus and Poisson ratio for different types of rock

Parameter	Value	Parameter	Value
Finite element	3D	Stress x,y,z	5000
Coupling	One way	Stress gradient x,y,z	-12
Rock type	Elasto Plastic: Mohr-Coulomb	Thermal expansion coefficient	1e-5 1/F

Table 2: Reservoir Parameters and Properties



Figure 11: Shear stress IJ Direction 3D view from GEM results

6.4 CO2 Injection Design (Number of Injector, Injection Time)

In the reservoir simulation model (as shown in Figure 10), there are four vertical injection wells. These wells were perforated and completed in all 51 sand layers, encompassing both the upper and lower aquifer formations. The injection strategy was designed to adhere to two main constraints: primary constraints related to the Bottom Hole Pressure (BHP) at the well level and secondary constraints associated with the maximum injection rate. To maintain a secure operational environment, a maximum BHP limit of 5,500 psi was established, with a prudent 10% safety margin from the reservoir fracture pressure. This precautionary measure aims to safeguard the integrity of the reservoir and prevent potential risks during the injection process. As illustrated in (Figure 12), the injection activities commenced on January 1, 2020, and they are planned to span 30 years, lasting until January 1, 2050. Following the injection phase, a post-injection monitoring phase is scheduled to begin on January 1, 2050, and will continue for an extensive period of 270 years, concluding on January 1, 2320. In total, the entire injection and post-injection monitoring scheme will extend over a span of 300 years. This long-term approach ensures a comprehensive assessment of the reservoir's behavior and provides valuable insights into its response to the injection activities over an extended period.



Figure 12: the simulation study illustrates the designated timesteps for injection and post-injection phases. These time intervals were thoughtfully chosen to capture the critical stages of the reservoir's response to injection and to monitor its behavior over an extended period after injection concludes.

Chapter 7: Methodology

7.1 Review

This chapter focuses on the methodologies used to design artificial neural networks, particularly in Smart Proxy Modeling for Shear Stress. It covers strategies for training development and blind deployment phases, aiming to clarify these processes. Additionally, it provides an overview of the procedures used to create a Smart Proxy tailored to the project's needs. This foundation supports understanding subsequent discussions. The chapter offers guidelines for building a predictive tool using Artificial Neural Networks and devising features, based on insights from the progression and training processes. A blind simulation scenario was executed for primary evaluation. Smart Proxy Modeling moves to the next stage if it is successful, involving validation with an unseen dataset. This particular phase holds significant importance as it serves as the primary yardstick for measuring the performance prowess of the SPM approach. The research followed the subsequent procedure outlined in (Figure 13).



Figure 13: The general workflow utilized for SPM in this study

To simplify the previous flowchart of Smart Proxy Model (SPM) development, there are four major stages for SPM in this study:

- Reservoir Simulation Scenarios: We start with a number of reservoir simulation scenarios that are run based on the study's purpose and this study will be focusing on Shear Stress in three different directions.
- Data Retrieval and Structuring: The next step involves the retrieval and structuring of relevant static and dynamic data. This process is crucial for producing a comprehensive spatio-temporal dataset, which simplifies the overall process. This dataset is then used to construct the input data set for neural network training.
- 3. Neural Network Construction and Training: Multiple neural network models are constructed and trained. This training process is vital as it simplifies the neural networks' ability to interpret the dataset effectively.
- 4. Model Testing: Finally, the trained neural network models are subjected to testing by running them through a series of blind scenarios. This evaluation phase, although an integral part of the process, simplifies the assessment of the models' accuracy and their capability to make predictions under various conditions.

7.2 Create a SPM for Predicting Dynamic Behavior Across Multiple Time Steps

The project's focal point involves crafting a dynamic smart proxy model for predicting Shear Stress in three distinct directions at each simulation time step. To fulfill this objective, three distinct Artificial Neural Networks Models were constructed. The accompanying flowchart provides a visual representation of the three Artificial Neural Networks Models (refer to Figure 14). The first ANN Model (referred to as ANN 1) takes input comprising Static Parameters and Shear Stress in the IJ direction at time step (t - Δ t), yielding Shear Stress in the IJ direction at time step (t) as its output. The Static Parameters, detailed in subsequent sections, encompass constants unaffected by time step changes. The second ANN Model (denoted as ANN 2) employs the predicted Shear Stress in the IK direction from time step (t - Δ t) as input, generating Shear Stress in the IK direction at time step (t) as output. Similarly, the third ANN Model (referred to as ANN 3) utilizes forecasted Shear Stress in the JK direction from time step (t - Δ t) as input and produces Shear Stress in the JK directions at time step (t) as output. Shear Stress values in the IJ, IK, and JK directions at time step (t - Δ t), coupled with the Static Parameters, serve as inputs to the ANNs for forecasting Shear Stress in those respective directions at time step (t).



Figure 14: Flowchart shows the design of the three ANNs of the Smart Proxy Model

Amidst the deluge of data from numerical reservoir simulations during injection and post injection phases, a strategic move was made. Time step data was neatly partitioned into five categories: monthly, yearly, decade, and 50 year intervals for both injection and post injection scenarios (Alaboodi, 2021). SPM were crafted and trained for each category, functioning as interconnected puzzle pieces and here's where it gets innovative. Think of it as a relay race. The latest Smart Proxy Models output takes the baton and becomes the input for the next Smart Proxy Models in line. And within each Smart Proxy Models, not only does the previous output tag along, but static data also joins the ride. This cascade trains the neural network to predict simulation outcomes one step at a time, without the traditional need for previous simulation outputs. Rather than playing catch-up, the SPM confidently anticipates the next move using its own projections as cues. The real beauty? It conjures up simulation results for any step, without hunting for prior numerical reservoir outputs. A game-changer compared to the more rigid methods of other studies.

7.3 Preparing Data for Artificial Neural Networks

The process of dataset preparation holds immense significance in the development of a Smart Reservoir Model. This dataset will serve as the educational material for the Smart Reservoir Model, imparting knowledge about the reservoir's characteristics and the specific scenarios at hand. The soundness of the Smart Reservoir Models output relies on the integrity of its input. Using flawed or inaccurate data in the model will inevitably lead to misguided outcomes. It's unreasonable to anticipate positive results from an SPM that hasn't been adequately trained. As mentioned earlier, following the execution of the twenty numerical reservoir simulation models, a thorough analysis and quality checks were conducted at each time step and layer for every individual run. This comprehensive scrutiny aims to comprehend the intricate interplay between various parameters. Furthermore, some crucial features have been generated, substantially enhancing the accuracy of the Smart Proxy Models. The next step involves extracting vital data from each of these runs. This extracted data serves as the fundamental building blocks for the spatio-temporal database, forming the very foundation on which the smart proxy model relies. The process of creating the smart proxy entails designing and executing these twenty simulation runs, resulting in a substantial collection of records that contribute to the richness of the dataset. In the quest to build a robust Artificial Neural Network model, it's imperative to employ domain knowledge when constructing the dataset. This essentially means that the data needs to be carefully prepared to effectively reveal the underlying structure of the problem to the machine learning algorithm. Thus, comprehending the physics and fluid flow dynamics within the reservoir model becomes the linchpin for generating an appropriate dataset for the Smart Proxy.

7.4 Data Processing

The numerical reservoir simulation runs conducted using CMG software generated a significant volume of raw data, both in terms of outputs and inputs. This raw data underwent analysis and was then transformed into a format suitable for further use. Python was employed to initiate this process, which involves reading the simulation outputs and inputs, followed by the extraction and storage of relevant data according to predefined criteria. This process demands a substantial amount of computational time and a comprehensive understanding of the critical data derived from reservoir simulations. The task of selecting input parameters for training a neural network from the pool of variables assimilated in the database is anything but simple. Project databases often encompass a multitude of parameters, each of which could potentially serve as an input for the neural networks that will facilitate data-driven model training. Within these databases, both static and dynamic attributes are provided, alongside comparable data for different offset wells. These values are organized as columns in a flat file, destined to be utilized for the training of data-driven models, particularly neural networks.

7.5 Features Generation and Engineering

The process of feature engineering involves the skillful application of domain expertise in selecting and transforming critical variables from raw data during the construction of predictive models using machine learning techniques. This practice not only empowers machine learning algorithms but also enhances their performance substantially. Within a particular project, this approach goes beyond mere enhancement; it imparts reservoir engineering insights by leveraging data from a numerical simulation model. However, the feature engineering undertaken in this context is no ordinary endeavor. It entails a thoughtful provision of pertinent information to machine learning algorithms, enabling them to comprehend the intricate dynamics of fluid flow within porous media. Beyond the extraction of static and dynamic attributes from the CMG model, a deeper dive has occurred. Novel features have been crafted following a thorough understanding of the data structure. This step is undertaken with the goal of elevating the learning capacity of the models, thereby enabling them to generate highly precise predictions. The dataset operates in a cellbased framework, where each row represents a focal cell within the model. This landscape is composed of 125 cells along the "I" axis, an equal number along the "J" axis, and 65 layers along the "K" axis, resulting in a total of 1,015,625 cells for each realization. The genesis of each new feature lies in meticulous calculations, with individual columns being dedicated to each focal cell. This multifaceted approach contributes to a comprehensive mosaic of characteristics. The concept of distance emerges as a pivotal element. The space between each cell and the injectors takes the spotlight, with a cluster of attributes shedding light on proximity. This information is not just an accompaniment; it becomes an integral part of the ANN Model. These cells cease to be mere coordinates; they transform into entities intertwined with numerical indexes (I, J, and K). Further, each cell is uniquely identified, bestowing upon it an emblem that aids recognition by the ANN. As this symphony of data unfolds, fresh features come to life, expanding the horizons of comprehension and predictive prowess. In the orchestration of learning, these features act as guiding notes for algorithms, allowing them to unravel the intricacies of reservoir behavior with heightened clarity. In visual terms, these features serve as brushstrokes that collaboratively paint the comprehensive picture of reservoir intricacies and behavioral patterns. Each realization, encompassing an impressive 1,015,625 cells, contributes to the intricate dance of information, with each new feature thoughtfully composed as a column aligning with individual focus cells. This orchestration not only yields an expansive reservoir of data but also highlights the meticulous craftsmanship behind the creation of predictive models.



Figure 15: Distance to the boundary and injection wells from the Focal Cell

In the context of a numerical reservoir simulation, the characteristics of a specific cell, such as its Shear Stress, are not solely influenced by its own attributes but are also impacted by its neighboring cells. Hence, when training the ANN Model, it becomes crucial for the model to gain insights from not just the primary cell but also the surrounding ones. To achieve this, a tier system is introduced, where each central cell is enveloped by three concentric layers termed tiers: tier 1, tier 2, and tier 3. These tiers are characterized by their arrangement into surface, line, and point contacts, respectively. In this methodology, both tier 1 and tier 2 play a significant role in shaping the dataset for the ANN Model. Tier 1 consists of cells in direct face-to-face interaction, encompassing six cells encircling the central one. Conversely, tier 2 includes cells in line contact, forming a configuration of 12 cells around the central cell. As a result, attributes from a total of 18 neighboring cells are utilized as input data to train the ANN Model. This arrangement is visually illustrated in Figure 16, depicting the circular arrangement of tier 1 cells and their relationship with the central cell, while also demonstrating the placement of both tier 1 and tier 2 cells.



Figure 16: Schematics showing the implementation of tier cells

In order to effectively guide the ANN model in comprehending the behavior of each cell within the system, it becomes crucial to establish a unique identification for each cell's position and employ it as input for the model. This is achieved by determining the precise location of each cell within the reservoir simulation model. This location is represented by calculating the distance of the cell from the model's boundaries comprising top, bottom, east, west, north, and south dimensions. Simultaneously, the three indexing values (i, j, and k) are utilized to pinpoint the exact coordinates of each cell. Additionally, to distinctly identify each cell's location, a unique cell ID has been assigned within the reservoir simulation model and introduced as input to the ANN, thereby enhancing the model's grasp of individual cell positions. To further augment the ANN's understanding, the distances between cell positions and reservoir model boundaries are combined with information regarding the cell's proximity to injection wells. This entails incorporating the order of proximity to the injection sources (e.g., distance to the closest injector, second closest injector, and so forth). This supplementary data serves as an additional feature, enabling the ANN to discern the relative significance of each cell's location concerning CO2 injection sources.

7.6 Spatio-Temporal Dataset Construction

The term "spatio-temporal dataset" refers to a collection of data that combines both spatial and temporal information relevant to a particular research subject. In this study, various geological characteristics, as well as additional features and outcomes from simulations, were gathered to form a comprehensive spatio-temporal dataset. This dataset is then utilized to teach the Smart Proxy about the fundamental principles of fluid movement through porous materials. Moreover, it helps the Smart Proxy understand the intricate nature of the reservoir's heterogeneity as depicted in the geological model, and how this heterogeneity influences the flow of fluids and fluctuations

in pressure within the reservoir. To efficiently gather information and reduce computational time, a systematic approach is employed during the design and execution of simulations. The study involves constructing multiple artificial neural networks, each dedicated to aspects such as Shear Stress in different directions (IJ, IK, JK). These networks consist of a hidden layer and an output layer. There is one hidden layer and one output in each network and it has been shown that single layer feedforward networks with backpropagation works well. In this context, the rows within the spatio-temporal database correspond to individual records or samples within the dataset, while the columns represent specific attributes or characteristics. Each time a numerical reservoir simulation generates a geological scenario, it produces over a million records associated with grid cells due to the complexity of the model. As the simulation program comprises 115 time-steps, the resulting spatio-temporal database encompasses the records. Each of these records encapsulates the static and dynamic attributes of a specific grid block within a given simulation run and time-step.

7.7 Data Partitioning

In the course of each Smart Proxy Model's (SPM) progression, its developmental data undergoes a methodical division into three distinct segments: training, calibration, and validation. This allocation follows a meticulous structure: 80% designated for training, 10% for validation, and an additional 10% for calibration, a configuration tailored specifically to the Shear Stress models within the IJ, IK, and JK directions. Among these segments, the training dataset takes precedence, notably due to its substantial size. It serves as the crucible where the neural network refines its capabilities, intricately weaving connections between input and output elements. In essence, the training dataset acts as a repository of fundamental elements that a data-driven model seeks, ultimately governing the breadth of its capabilities. Stepping into the scene, the calibration data assumes a more deliberate presence, functioning as an assessor of the artificial neural network's (ANN) performance throughout its training spectacle. Concurrently, the training data assumes the role of an educator for the ANN. Following the culmination of the training phase, the educated ANN embarks upon the validation stage. Visualize the calibration dataset as a vigilant sentinel, meticulously overseeing the training performance and proclaiming a halt when needed. The rationale behind this intervention is rooted in the fact that the network's efficacy is inherently tethered to its calibration dataset, a shrewdly selected dataset purposefully held in the dark. However, there's more to unveil. Amidst this didactic spectacle, four reservoir simulation runs are handpicked to remain untouched, designated as enigmatic "blind runs." They stand resolute, untouched by the training process, awaiting their moment to contribute to the verification

phase by assuming the guise of novel simulation data. And then, in a grand culmination, enters the validation dataset. This segment emerges as the final act but commands profound significance. Diverging from its counterparts, it refrains from molding the network or steering its calibration. Instead, it maintains its distance from the outset, reserved for a singular purpose: to be the masked dataset. It maintains a patient presence throughout the proceedings, anticipating its role or in other words, it is being highlighted that the validation dataset is attentively awaiting the moment when it will be used for its intended purpose in evaluating the trained neural network's performance. The adeptly trained neural network encounters its initial litmus test through this dataset, assessing its capacity for generalization. This marks the preliminary stage preceding the ultimate examination, the blind validation test.

7.8 Training Process of Artificial Neural Network and SPM

In the realm of training neural networks, simulation runs illustrate how the model interacts with data, reflecting its ability to oversee CO2 injection across diverse scenarios. Proxy modeling, a data-driven representation, captures the reservoir system's essence and internal dynamics, demanding insights from various domains. Static data maintains constancy, including reservoir structure details. Supplementary data covers distances between blocks, encapsulating initial values and proportions. Dynamic data spans well and blocks aspects, encompassing changing pressures, rates, and variables like shear stress content. Crafting a network for a specific purpose initiates training, beginning with random initial weights. This iterative learning process refines the network's weights for accurate input-output mapping. During training, the network's hidden layers process data sequentially, aligning generated outputs with desired outcomes. Errors trigger adjustments, refining connections over time. During the development of neural networks, once construction and refinement are complete, the training phase commences with designated data. The hyperparameters of each artificial neural network (ANN) are tailored individually. The training halts upon noticeable improvements in predicted performance aligned with predefined stopping criteria. The calibration dataset guides the neural network to determine the optimal training endpoint for deploying the best-fit model. An excessively accurate neural network trained solely on the training data is termed an over-trained model, memorizing records without genuine predictive capacity referred to as over-training or over-fitting, a condition to avoid. Such tendencies are mitigated by the calibration dataset. Monitoring the network's behavior during training provides valuable insights into its progression. This observation is a crucial gauge for effective convergence toward a solution.

In this study, neural networks with a single hidden layer were used for analysis, as mentioned previously. Table 3 below provides a sample of the hyperparameters used for the neural networks in this study. The optimization of hyperparameters for each neural network was based on factors such as the size of its spatio-temporal dataset and initial observations from the training process. The hyperparameters, encompassing factors like the learning rate, the quantity of neurons within hidden layers, and the selection of an activation function, have been recognized as influential determinants of the neural networks' training performance. It's essential to emphasize that hyperparameter tuning stands as a pivotal component of the process. Its particulars can fluctuate depending on the inherent nature and intricacy of the problem being addressed, as well as the distinctive attributes of the available data. Additionally, the proficiency of the machine learning engineer plays a significant role in shaping the effectiveness of hyperparameter tuning.

Hyperparameter	Values / type
Number of Hidden Layers	1
Number of Neuron in Hidden Layer	1000
Learning Rate	0.001
Activation Function (Input Layer-Hidden Layer)	ReLU
Activation Function (Hidden Layer-Output Layer)	Tanh
Number of Epoch	10,000
Batch size	20,000

Table 3: Typical hyperparameters used in this study for one of the ANNs

In the architecture of a neural network, there's a hidden layer positioned between the input and output layers. This hidden layer comprises a defined number of neurons, often referred to as "Neurons in the Hidden Layer." The neural network's learning rate is a pivotal parameter that governs the size of each step taken during the iterative process aimed at minimizing the loss function. Simultaneously, the activation function assumes a crucial role in determining whether a neuron should activate or remain dormant. It accomplishes this by calculating a weighted sum and adding a bias. The primary objective of the activation function is to introduce nonlinearity into a neuron's output using basic mathematical operations. This nonlinearity is essential for neural

networks to effectively model and comprehend intricate relationships within data. Now, when we discuss the training process, it involves processing all the data records in a training set and measuring the error between the neural network's output and the actual desired output. A full cycle of this training process is commonly referred to as an "epoch." During an epoch, the entire training dataset is propagated through the neural network, both forward and backward, once. Typically, specific criteria are established to determine when the training process should conclude. This conclusion occurs when the error between the predicted output and the actual output drops below a predefined threshold. To facilitate the training process, the training data can be divided into smaller subsets known as the "batch size." This division serves a dual purpose. First, it accommodates situations where the training dataset is too extensive to be processed all at once by the neural network. Second, it introduces an additional parameter that contributes to the neural network's training dynamics.

7.9 Error Measurements

The validity of the proposed smart proxy model hinges on its assessment using a distinct validation dataset. Crucially, the model's precision in comparison to the blind dataset must be established. In this context, precision is determined by quantifying the difference between the output of the numerical simulator and the smart proxy model. The evaluation of error must occur at every grid block since the smart proxy generates output at each of these blocks. A low error signifies favorable performance, whereas a high error is clearly undesirable. The method of error calculation varies based on the nature of the output data, leading to the utilization of diverse error formulas. Specifically, for Shear Stress IJ, IK, and JK directions output, the subsequent error formulas are employed:

Absolute Error Percentage = [(absolute (Artificial Neural Network Output - CMG Output)) / CMG Output] * 100

In the following error formula the difference between what the ANN predicts and what the numerical reservoir simulation model shows tells us how accurate the SPM is.

Absolute Error (or Prediction Accuracy) = absolute (Artificial Neural Network Output - Numerical Simulator Output)

7.10 Ensuring Neural Network Proficiency with Validation Datasets

Upon reaching a level of satisfaction in the Smart Proxy neural network training, the process advances to a crucial validation stage. Here, the validation datasets come into play, which remain untapped during the training phase. These datasets are exclusively introduced after the training concludes and just prior to the deployment of the Smart Proxy on entirely novel simulation runs, a process referred to as blind validation. The final touch in the SPM development journey involves subjecting the evolved model's outcomes to scrutiny through blind validation. This assessment entails running simulations that have never been employed for neural network training, thereby encapsulating the essence of a blind validation scenario.

7.11 Smart Proxy Deployment

In the context of Smart Proxy Deployment, the training procedures for the Smart Proxy Models related to Shear Stress (IJ, IK, and JK models) were similarly followed by a calibration and validation phase aimed at assessing their internal performance. This validation process ensured the internal performance of the trained models, and deployment was initiated upon achieving the desired performance according to the established stopping criteria. This deployment leveraged a blind validation dataset comprising four numerical simulations: Blind Run 5, Blind Run 9, Blind Run 13, and Blind Run 17, all of which were entirely distinct from the neural network's training data. It's noteworthy that throughout the entire training and development stages, the neural network had not encountered these blind-validation simulations at any point.

Shear Stress IJ Features					
id	inj_well_based_bhp_1	permeability_NT	top_SB	porosity_TE	
i	inj_well_based_bhp_2	thickness_NT	pr_initial_SB	permeability_TE	
j	inj_well_based_bhp_3	top_NT	porosity_SW	thickness_TE	
k	inj_well_based_bhp_4	pr_initial_NT	permeability_SW	top_TE	
x	inj_well_based_rate_1	porosity_NB	thickness_SW	pr_initial_TE	
у	inj_well_based_rate_2	permeability_NB	top_SW	Shear_Stress_IJ_t_dt	
porosity	inj_well_based_rate_3	thickness_NB	pr_initial_SW	initial_Shear_Stress_IJ	
permeability	inj_well_based_rate_4	top_NB	porosity_SE		
thickness	inj_cell_based_bhp_1	pr_initial_NB	permeability_SE		
top	inj_cell_based_bhp_2	porosity_NW	thickness_SE		
bottom	inj_cell_based_bhp_3	permeability_NW	top_SE		
paydepth	inj_cell_based_bhp_4	thickness_NW	pr_initial_SE		
Distance to East	inj_cell_based_rate_1	top_NW	porosity_BW		
Distance to West	inj_cell_based_rate_2	pr_initial_NW	permeability_BW		
Distance to North	inj_cell_based_rate_3	porosity_NE	thickness_BW		
Distance to South	inj_cell_based_rate_4	permeability_NE	top_BW		
Distance to Top Seal	porosity_E,W,N,S	thickness_NE	pr_initial_BW		
Distance to Bottom Seal	permeability_E,W,N,S	top_NE	porosity_BE		
Young_Modulus	thickness_E,W,N,S	pr_initial_NE	permeability_BE		
Poissons_Ratio	top_E,W,N,S	porosity_ST	thickness_BE		
pr_initial	pr_initial_E,W,N,S	permeability_ST	top_BE		
pr_t_dt	porosity_T,B	thickness_ST	pr_initial_BE		
sg_t_dt	permeability_T,B	top_ST	porosity_TW		
closest Dist 1	thickness_T,B	pr_initial_ST	permeability_TW		
closest Dist 2	top_T,B	porosity_SB	thickness_TW		
closest Dist 3	pr_initial_T,B	permeability_SB	top_TW		
closest Dist 4	porosity_NT	thickness_SB	pr_initial_TW		

Shear Stress IK Features					
id	inj_well_based_bhp_1	permeability_NT	top_SB	porosity_TE	
i	inj_well_based_bhp_2	thickness_NT	pr_initial_SB	permeability_TE	
j	inj_well_based_bhp_3	top_NT	porosity_SW	thickness_TE	
k	inj_well_based_bhp_4	pr_initial_NT	permeability_SW	top_TE	
x	inj_well_based_rate_1	porosity_NB	thickness_SW	pr_initial_TE	
у	inj_well_based_rate_2	permeability_NB	top_SW	Shear_Stress_IK_t_dt	
porosity	inj_well_based_rate_3	thickness_NB	pr_initial_SW	initial_Shear_Stress_IK	
permeability	inj_well_based_rate_4	top_NB	porosity_SE		
thickness	inj_cell_based_bhp_1	pr_initial_NB	permeability_SE		
top	inj_cell_based_bhp_2	porosity_NW	thickness_SE		
bottom	inj_cell_based_bhp_3	permeability_NW	top_SE		
paydepth	inj_cell_based_bhp_4	thickness_NW	pr_initial_SE		
Distance to East	inj_cell_based_rate_1	top_NW	porosity_BW		
Distance to West	inj_cell_based_rate_2	pr_initial_NW	permeability_BW		
Distance to North	inj_cell_based_rate_3	porosity_NE	thickness_BW		
Distance to South	inj_cell_based_rate_4	permeability_NE	top_BW		
Distance to Top Seal	porosity_E,W,N,S	thickness_NE	pr_initial_BW		
Distance to Bottom Seal	permeability_E,W,N,S	top_NE	porosity_BE		
Young_Modulus	thickness_E,W,N,S	pr_initial_NE	permeability_BE		
Poissons_Ratio	top_E,W,N,S	porosity_ST	thickness_BE		
pr_initial	pr_initial_E,W,N,S	permeability_ST	top_BE		
pr_t_dt	porosity_T,B	thickness_ST	pr_initial_BE		
sg_t_dt	permeability_T,B	top_ST	porosity_TW		
closest Dist 1	thickness_T,B	pr_initial_ST	permeability_TW		
closest Dist 2	top_T,B	porosity_SB	thickness_TW		
closest Dist 3	pr_initial_T,B	permeability_SB	top_TW		
closest Dist 4	porosity_NT	thickness_SB	pr_initial_TW		

Shear Stress JK Features					
id	inj_well_based_bhp_1	permeability_NT	top_SB	porosity_TE	
i	inj_well_based_bhp_2	thickness_NT	pr_initial_SB	permeability_TE	
j	inj_well_based_bhp_3	top_NT	porosity_SW	thickness_TE	
k	inj_well_based_bhp_4	pr_initial_NT	permeability_SW	top_TE	
x	inj_well_based_rate_1	porosity_NB	thickness_SW	pr_initial_TE	
у	inj_well_based_rate_2	permeability_NB	top_SW	Shear_Stress_JK_t_dt	
porosity	inj_well_based_rate_3	thickness_NB	pr_initial_SW	initial_Shear_Stress_JK	
permeability	inj_well_based_rate_4	top_NB	porosity_SE		
thickness	inj_cell_based_bhp_1	pr_initial_NB	permeability_SE		
top	inj_cell_based_bhp_2	porosity_NW	thickness_SE		
bottom	inj_cell_based_bhp_3	permeability_NW	top_SE		
paydepth	inj_cell_based_bhp_4	thickness_NW	pr_initial_SE		
Distance to East	inj_cell_based_rate_1	top_NW	porosity_BW		
Distance to West	inj_cell_based_rate_2	pr_initial_NW	permeability_BW		
Distance to North	inj_cell_based_rate_3	porosity_NE	thickness_BW		
Distance to South	inj_cell_based_rate_4	permeability_NE	top_BW		
Distance to Top Seal	porosity_E,W,N,S	thickness_NE	pr_initial_BW		
Distance to Bottom Seal	permeability_E,W,N,S	top_NE	porosity_BE		
Young_Modulus	thickness_E,W,N,S	pr_initial_NE	permeability_BE		
Poissons_Ratio	top_E,W,N,S	porosity_ST	thickness_BE		
pr_initial	pr_initial_E,W,N,S	permeability_ST	top_BE		
pr <u>t</u> dt	porosity_T,B	thickness_ST	pr_initial_BE		
sg_t_dt	permeability_T,B	top_ST	porosity_TW		
closest Dist 1	thickness_T,B	pr_initial_ST	permeability_TW		
closest Dist 2	top_T,B	porosity_SB	thickness_TW		
closest Dist 3	pr_initial_T,B	permeability_SB	top_TW		
closest Dist 4	porosity NT	thickness SB	pr initial TW		

Table 4: A list of selected features as input to the neural network a particular SPM's

Chapter 8: Results and Discussions

8.1 Overview

The purpose of this chapter is to assess the performance of the smart proxy modeling technique developed in this study in replicating the Shear Stress IJ, IK, and JK distribution results obtained from numerical simulations during blind runs conducted using SPM. To confirm the accuracy and reliability of the Smart Proxy's predictions, CMG was also used to generate results for the blind runs. A comparative analysis is then presented between the results produced by the Smart Proxy and CMG for each layer of the reservoir model. Visual representations in the form of two-dimensional distribution maps are utilized to showcase the quality of the results generated by both the Smart Proxy and the numerical reservoir simulation model. Additionally, mean absolute error output plots will be shared to provide a clearer understanding of the effectiveness of SPM in capturing Shear Stress patterns. In each blind run, there are a total of 153 two-dimensional distribution plots and 125 cross-section plots for both i and j directions. As there are 51 layers in each simulation run, this chapter presents results for only one training and blind realization for each direction of Shear Stress, with the remaining results for training and blind runs provided in the Appendix section, focusing on specific reservoir layers for brevity.

8.2 Reservoir Shear Stress Distribution Results

Utilizing the Smart Proxy method, the distribution of reservoir shear stress was accurately predicted, achieving the highest degree of accuracy among various runs. These results consistently matched the outcomes obtained from CMG simulations at all time intervals, with an error rate of less than 10%. All of these neural networks were subsequently deployed in sequence to predict reservoir shear stress across all time steps in the reservoir simulation model. Importantly, the greatest error was observed when utilizing data generated by previous neural networks as input for subsequent ones. This can be attributed to the cumulative nature of forecasting errors in the sequential neural network approach.

To gain a more comprehensive understanding of the variations in Shear Stress within individual layers, 2D results were generated and examined for all 51 layers in the reservoir model. For brevity, the 2D plots displaying Shear Stress outcomes for blind-13 on January 1, 2050, appear on the following pages, along with shared training results for train-01. These plots are arranged as follows: CMG data on the left, SPM results in the center, and prediction accuracy on the far

right. The color scale used for visualization was selected based on the minimum and maximum Shear Stress values observed in each layer.

8.2.1 Training Results of Shear Stress IJ, IK, and JK

In this section, we will delve into the training outcomes for each of the Model IJ, IK, and JK, specifically focusing on No. train-01. The remaining training results are reserved for the Appendix section, where only a select few runs and reservoir layers are included, ensuring brevity. Upon scrutinizing the training results of Model IJ, we observe outcomes that are less than optimal compared to those of Model IK and JK in Figures 17 to 19. Analyzing the results of Shear Stress IK and JK, it becomes evident that the predictive performance is highly satisfactory, indicating a more effective learning process compared to the initial model Shear Stress IJ.
























Figure 17: displays a comparison between the predicted Smart Proxy Model Shear Stress results for Model IJ and the actual CMG results, along with the corresponding error percentages, across all layers for train run-01 on January 1, 2050. The arrangement from left to right includes the CMG result, SPM result, and Percentage Error.



[Training Run-01 Results for Shear Stress IK direction 01-01-2050]



















Figure 18: displays a comparison between the predicted Smart Proxy Model Shear Stress results for Model IK and the actual CMG results, along with the corresponding error percentages, across all layers for train run-01 on January 1, 2050. The arrangement from left to right includes the CMG result, SPM result, and Percentage Error.



[Training Run-01 Results for Shear Stress JK direction 01-01-2050]



















Figure 19: displays a comparison between the predicted Smart Proxy Model Shear Stress results for Model JK and the actual CMG results, along with the corresponding error percentages, across all layers for train run-01 on January 1, 2050. The arrangement from left to right includes the CMG result, SPM result, and Percentage Error.

8.2.2 Blind Results of Shear Stress IJ, IK, and JK

As observed in the blind-validation run-13 outcomes, the performance graphs for Shear Stress in the Smart Proxy Model for IJ, IK, JK direction (Figures 20 through 22) reveal that certain layers yielded exceptionally favorable results, while others were reasonably satisfactory. However, it's noteworthy that the Smart Proxy Model still demonstrates a relatively consistent ability to approximate the results of numerical reservoir simulation across nearly all the layers. Given that each simulation run comprises 51 layers and, consequently, a substantial volume of blind results, the chosen presentation includes only one blind result for Shear Stress in each direction. The remaining blind results can be found in the Appendix section to streamline the presentation.



[Blind Run-13 Results for Shear Stress IJ direction 01-01-2050]

















Figure 20: Comparison of Predicted SPM Shear Stress IJ Results with Actual CMG Results and Corresponding Error Percentage Across All Layers for Blind Run-13 on January 1, 2050. (Displayed from left to right: CMG Result, SPM Result, and Percentage Error)



[Blind Run-13 Results for Shear Stress IK direction 01-01-2050]
















Figure 21: Comparison of Predicted SPM Shear Stress IK Results with Actual CMG Results and Corresponding Error Percentage Across All Layers for Blind Run-13 on January 1, 2050. (Displayed from left to right: CMG Result, SPM Result, and Percentage Error)



[Blind Run-13 Results for Shear Stress JK direction 01-01-2050] Blind_Run_13: layer_5 Smart Proxy Perc. Error Plot



















Figure 22: Comparison of Predicted SPM Shear Stress JK Results with Actual CMG Results and Corresponding Error Percentage Across All Layers for Blind Run-13 on January 1, 2050. (Displayed from left to right: CMG Result, SPM Result, and Percentage Error)

8.2.2.1 Cross Section i and j Direction Results of Shear Stress IJ, IK, and JK

In the following sections, you'll discover 2D graphs depicting the results of Shear Stress IJ, IK, and JK reservoir cross-section analysis for i and j direction. These graphs pertain to shear stress within the context of the blind-13 case and can be found in Figures 23 through 28. As there are 125 indexes for each blind case, to maintain consciousness, only half of the indices will be included in the Appendix section. Each row within these figures corresponds to data collected at a specific time step, precisely on January 1, 2050. In each row, you'll notice three plots: the

leftmost one showcases CMG data, the middle plot displays Smart Proxy method results, and the rightmost plot highlights the percentage error. The blind cross section case results indicate that the Smart Proxy Model consistently reproduces the results of numerical reservoir simulation across all layers, confirming the successful training of the ANN algorithm.






















































Figure 23: Cross Section i Shear Stress model IJ results for all layers for timestep 01-01-2050

[Blind Run-13 Cross Section i Results for Shear Stress IK direction 01-01-2050]





















































Figure 24: Cross Section i Shear Stress model IK results for all layers for timestep 01-01-2050

[Blind Run-13 Cross Section i Results for Shear Stress JK direction 01-01-2050]




















































Figure 25: Cross Section i Shear Stress model JK results for all layers for timestep 01-01-2050



[Blind Run-13 Cross Section j Results for Shear Stress IJ direction 01-01-2050]


















































Figure 26: Cross Section j Shear Stress model IJ results for all layers for timestep 01-01-2050

[Blind Run-13 Cross Section j Results for Shear Stress IK direction 01-01-2050]





















































Figure 27: Cross Section j Shear Stress model IK results for all layers for timestep 01-01-2050






















































Figure 28: Cross Section j Shear Stress model JK results for all layers for timestep 01-01-2050 8.3 Mean Absolute Error Output Plots Comparison

In the context of mean absolute error output comparisons for geological realization 13, as illustrated in Figure 29, it is evident that certain layers consistently demonstrate exceptional performance, while others exhibit commendable results. Across various blind runs, there are instances where the match is reasonably good, and there are instances where it is exceptionally strong. For instance, when considering layer 5 for Shear Stress IJ, it exhibits lower error rates when compared to Shear Stress IJ and IK. Similarly, in the case of layer 28 for Shear Stress IJ, it shows lower error rates compared to Shear Stress IK and JK. These observations collectively affirm the Smart Proxy Model's capability to consistently replicate the outcomes of numerical reservoir simulations across all layers. This achievement underscores the successful training of the ANN algorithm. The remaining output plots will be available in the appendix.



Figure 29: Mean Absolute Error for Geological Realization 13 Shear Stress IJ, IK, JK Output Comparison results

Chapter 9: Conclusions and Recommendations

9.1 Conclusions

In the oil and gas industry, the primary aim is to boost revenue while cutting operational costs. Reservoir simulation plays a pivotal role in field planning, encompassing tasks like history matching and optimization. However, modern geological models, though more precise, demand hefty computational resources, driving up expenses. This study spotlights the time and cost savings achievable through the smart proxy model. Unlike traditional simulations, smart proxies are fast and computationally efficient. They can run on regular personal computers, eliminating the need for expensive hardware. The smart proxy database, though extensive, can be streamlined through sampling. Through the adoption of the smart proxy model, one can achieve a significant reduction in simulation costs and a notable acceleration in obtaining results. As evidence a single conventional simulation may typically consume seven hours, but with smart proxies, the same task can be completed in a matter of seconds. The Shear Stress model outcomes not only demonstrated its ability to closely replicate the CMG model with remarkable precision but also showcased its efficiency in achieving this at a significantly faster pace compared to traditional simulation methods, all the while maintaining affordability.

9.2 Recommendations

Presented below are recommendations for future research in the domain of advancing smart proxy development for reservoir simulation:

- Based on the comparison output plot highlighting in the results the suboptimal performance of the Smart Proxy Shear Stress IK model, it is strongly recommended to incorporate both Shear Stress IJ and JK model results as inputs when reconstructing the Shear Stress IK model. This approach is anticipated to yield significantly improved results.
- For Carbon Capture and Storage (CCS) projects, it is advisable to employ a two-way coupling approach rather than relying solely on one-way coupling. This two-way coupling strategy is expected to enhance the efficiency and effectiveness of CCS initiatives.

Appendix

Appendix 10.1: Training Shear Stress Results (IJ, IK, JK) on 01-01-2050 for Selected Reservoir Layers. In maintaining consistency throughout this section, we have chosen specific reservoir layers to remain constant across all models and time steps. These selected layers are denoted as follows: #5, #20, #40, #50, and #60.









Appendix 10.2: Result for Blind Run shear stress at 01-01-2050 for selected reservoir Layers
Shear Stress Results IJ
Shear Stress Results IK







Appendix 10.3: Result for cross section i & j Blind Run 09 shear stress at 01-01-2050 for selected reservoir Layers





Shear Stress Results IK





Shear Stress Results JK







Appendix 10.4: Mean Absolute Error for Geological Realization 5,9,17 Shear Stress IJ, IK, JK Output Comparison results



Reference

- Alnuaimi, Marwan, "Application of artificial intelligence for CO2 storage in saline aquifer (SMART proxy for snap-shot in time)" (2022).Graduate Theses, Dissertations, and Problem Reports.11278
- Alenezi and S. Mohaghegh, "A Data-Driven Smart Proxy Model for A Comprehensive Reservoir Simulation," IEEE, 2016
- Assessing the global climate in 2022. (2023, February 8). National Centers for Environmental Information (NCEI). https://www.ncei.noaa.gov/news/global-climate-202212
- Alzayani, A., AlYousef, Z., & Almajid, M. M. (2023). CO2 geological sequestration in basalt formations. *TWA*. https://jpt.spe.org/twa/co2-geological-sequestration-in-basalt-formations
- Amirmasoud Kalantari Dahaghi, Shahab D Mohaghegh, et al. Numerical simulation and multiple realizations for sensitivity study of shale gas reservoirs. In SPE Production and Operations Symposium. Society of Petroleum Engineers, 2011.
- Alenezi, Faisal, and Shahab Mohaghegh. "Developing a smart proxy for the SACROC waterflooding numerical reservoir simulation model." SPE western regional meeting. OnePetro, 2017.

ArtificialNeuralnetwork.(n.d.).https://www.saedsayad.com/artificial_neural_network.htm#:~:text=The%20transfer%20function%20translates%20the,%2C%20piecewise%20linear%2C%20and%20Gaussian.&text=The%20output%20is%20set%20at,less%20than%20some%20threshold%20value.

Backpropagation | Brilliant Math & Science Wiki. (n.d.). https://brilliant.org/wiki/backpropagation/

- Chapter 13. Stress and Strain. (n.d.). OPENPRESS.USASK.CA An Open Academic Publishing Platform. https://openpress.usask.ca/physicalgeology/part/chapter-4-plate-tectonics-3rded/https://openpress.usask.ca/physicalgeology/chapter/13-1-stress-andstrain2/#:~:text=Stress%20is%20force%20adjusted%20for,rocks%20are%20deformed%20by%2 0stress
- Chapter 4 Engineering Classification of Rock Materials. (2012). In *Part 631 Geology National Engineering Handbook* (55th ed.). United States Department of Agriculture Natural Resources Conservation Service.
- Cook, J. (2023). Optimizing development of new reservoirs and revitalizing mature fields. SLB. https://careers.slb.com/inside-slb/transformative-technology/defininggeomechanics#:~:text=Geomechanics%20is%20the%20study%20of,%2C%20temperature%2C %20and%20other%20parameters
- Course Hero, "The Math Behind Neural Networks (01)," 18 November 2018. [Online]. Available: https://www.youtube.com/watch?v=RiqWATOoos8.
- Carbon Storage FAQs .(n.d.). netl.doe.gov. https://netl.doe.gov/carbon-management/carbonstorage/faqs/carbon-storage-faqs

- Dong, L., Gong, F., Xu, H., Fan, P., & Wu, Z. (2021, February 4). On the experimental determination of Poisson's ratio for intact rocks and its variation as deformation develops. Publishing Open Access research journals & papers | Hindawi. https://www.hindawi.com/journals/ace/2021/8843056/
- E, Alpaydın, Introduction to Machine Learning Second Edition. Massachusetts Institute of Technology, 2010.
- F. Alenezi and S. Mohaghegh, "Developing a Smart Proxy for the SACROC Water-Flooding Numerical Reservoir Simulation Model," Society of Petroleum Engineers, p. 12, 2017
- Feng, Y., & Zhang, S. (2022). Geomechanics of geological carbon sequestration. *Carbon Sequestration*. https://doi.org/10.5772/intechopen.105412
- Gholami, V.; Mohaghegh, S. and Maysami, M. (2019). Smart Proxy Modeling of SACROC CO2-EOR. https://www.mdpi.com/2311-5521/4/2/85
- Gholami, V.; Mohaghegh, S. and Maysami, M. 2019. Smart Proxy Modeling of SACROC CO2-EOR. Fluids 2019, 4(2), 85; https://doi.org/10.3390/fluids4020085.
- Gold, E. (2023). The History of Artificial Intelligence from the 1950s to Today. *freeCodeCamp.org*. https://www.freecodecamp.org/news/the-history-of-ai/
- "GEM, Compositional and Unconventional Reservoir Simulator Unconventional Reservoir Training." CMG GEM | Compositional and Unconventional Reservoir Simulation, 31 Aug. 2016, <u>www.cmgl.ca/gem</u>.
- Haghighat, S. A., et al. "Reservoir Simulation of CO2 Sequestration in Deep Saline Reservoir, Citronelle Dome, USA." Twelfth Annual Conference on Carbon Capture, Utilization and Sequestration. 2013.
- Haghighat, S. A., Mohaghegh, S. D., Borzouie, N., Moreno, D., Shahkarami, A., & A. (2013, May). Reservoir Simulation of CO2 Sequestration in Deep Saline Reservoir, Citronelle Dome, USA. In Twelfth Annual Conference on Carbon Capture, Utilization and Sequestration
- Haghighat, S. Alireza, and Shahab D. Mohaghegh. "Interpretation of real-time pressure measurements to detect CO2 leakage." SPE Annual Technical Conference and Exhibition. OnePetro, 2015.
- He, J., Sætrom, J., Durlofsky, L.J., 2011a. Enhanced linearized reduced-order models for subsurface flow simulation. Journal of Computational Physics 230, 8313–8341.
- Jalali, Jalal, and S. Mohaghegh. "Comprehensive analysis of enhanced CBM production via CO2 injection using a surrogate reservoir model." Eighth Annual Conference on Carbon Capture and Sequestration. 2009.
- Jalali, Jalal, and Shahab D. Mohaghegh. "Reservoir simulation and uncertainty analysis of enhanced cbm production using artificial neural networks." SPE Eastern Regional Meeting. OnePetro, 2009.
- Jaadi, Z. (2021). A Step-by-Step Explanation of Principal Component Analysis (PCA). *Built In.* https://builtin.com/data-science/step-explanation-principal-component-analysis
- Luo, Z., & Chen, G. (2019). Foreword and Introduction. In *Elsevier eBooks* (pp. xi–xvi). https://doi.org/10.1016/b978-0-12-816798-4.00005-x

- Laboratory testing and rock properties. (2011). *Harmonising Rock Engineering and the Environment*, 315-426. https://doi.org/10.1201/b11646-10
- Lee, Y., Park, Y., Song, I., & Lee, H. (2022). undefined. *EAGE Asia Pacific Workshop on CO*₂ *Geological Storage*. https://doi.org/10.3997/2214-4609.202275004
- Mohaghegh, S. D., et al. "Smart proxy: an innovative reservoir management tool; Case study of a giant mature oilfield in the UAE." Abu Dhabi International Petroleum Exhibition and Conference. OnePetro, 2015.

Mohaghegh, Shahab D., et al. "Development of Surrogate Reservoir Model (SRM) for fast track

analysis of a complex reservoir." International Journal of Oil, Gas and Coal Technology

2.1 (2009): 2-23.

- Mohaghegh, Shahab D., Amini, Shohreh, Gholami, Vida, Gaskari, Razi, and Grant Bromhal. "GridBased Surrogate Reservoir Modeling (SRM) for Fast Track Analysis of Numerical Reservoir Simulation Models at the Grid block Level." Paper presented at the SPE Western Regional Meeting, Bakersfield, California, USA, March 2012. doi: https://doi.org/10.2118/153844-MS.
- Mohaghegh, Shahab D. "Reservoir simulation and modeling based on pattern recognition." SPE

Digital Energy Conference and Exhibition. OnePetro, 2011.

- Mohaghegh, S. Smart proxy modeling for numerical reservoir simulations big data analytic in e and p, 2015.
- Mohaghegh, Shahab D. "Quantifying uncertainties associated with reservoir simulation studies

using a surrogate reservoir model." SPE Annual Technical Conference and Exhibition.

OnePetro, 2006.

Mohaghegh, Shahab D., et al. "Uncertainty analysis of a giant oil field in the middle east using

surrogate reservoir model." Abu Dhabi International Petroleum Exhibition and

Conference. OnePetro, 2006.

Mohaghegh, Shahab D. "Converting detail reservoir simulation models into effective reservoir

management tools using SRMs; case study-three green fields in Saudi Arabia."

International Journal of Oil, Gas and Coal Technology 7.2 (2014): 115-131.

Mohaghegh, S. D., et al. "Smart proxy: an innovative reservoir management tool; Case study of a giant mature oilfield in the UAE." Abu Dhabi International Petroleum Exhibition and Conference. OnePetro, 2015.

Mohaghegh, Shahab D "Top-Down Intelligent Reservoir Modeling (Tdirm); A New Approach In

Reservoir Modeling By Integrating Classic Reservoir Engineering With Artificial

Intelligence & Data Mining Techniques".AAPG,2009

- Mohaghegh, S. D. (2000). Virtual-Intelligence Applications in Petroleum Engineering: Part 1—Artificial neural networks. *Journal of Petroleum Technology*, 52(09), 64–73. https://doi.org/10.2118/58046-jpt
- Mohaghegh, Shahab D. Data-Driven Analytics for the Geological Storage of Co2. CRC Press, Taylor & Francis Group, 2018. INSERT-MISSING-DATABASE-NAME, INSERT-MISSING-URL.
- Mohaghegh, Shahab D. Smart proxy modeling : artificial intelligence and machine learning in numerical simulation, 2023.
- Mibei, G. (2014). INTRODUCTION TO TYPES AND CLASSIFICATION OF ROCKS. *Exploration for Geothermal Resources*. https://orkustofnun.is/gogn/unu-gtp-sc/UNU-GTP-SC-19-0302.pdf
- Ma, Y. Z., Sobernheim, D., & Garzon, J. R. (2016). Glossary for unconventional oil and gas resource evaluation and development. Unconventional Oil and Gas Resources Handbook, 513-526. https://doi.org/10.1016/b978-0-12-802238-2.00019-5
- Pramoditha, Rukshan. "How to Choose the Right Activation Function for Neural Networks." *Medium*, 26 Jan. 2022, towardsdatascience.com/how-to-choose-the-right-activation-function-for-neural-networks-3941ff0e6f9c.
- Poison's Ratio. (n.d.). Instron. Https://www.instron.com/en-us/resources/glossary/p/poissonsratio#:~:text=The%20equation%20for%20calculating%20Poisson's,direction%20of%20the%20a pplied%20force
- Rahman, M. J., Fawad, M., & Mondol, N. H. (2020). Organic-rich shale caprock properties of potential CO2 storage sites in the northern North Sea, offshore Norway. *Marine and Petroleum Geology*, 122, 104665. https://doi.org/10.1016/j.marpetgeo.2020.104665
- Rutqvist, J. (2012). The Geomechanics of CO2 storage in deep sedimentary formations. *Geotechnical and Geological Engineering*, *30*(3), 525-551. https://doi.org/10.1007/s10706-011-9491-0
- Stedman, C., & Hughes, A. (2021). data mining. *Business Analytics*. https://www.techtarget.com/searchbusinessanalytics/definition/datamining?Offer=abt_pubpro_AI-Insider

Soft computing. (2023, August 2). Springer. https://www.springer.com/journal/500

- Shahkarami, Alireza , Mohaghegh, Shahab D., Gholami, Vida , and Grant Bromhal. "Application of Artificial Intelligence and Data Mining Techniques for Fast Track Modeling of Geologic Sequestration of Carbon Dioxide – Case Study of SACROC Unit." Paper presented at the SPE Digital Energy Conference and Exhibition, The Woodlands, Texas, USA, March 2015. doi: https://doi.org/SPE-173406-MS.
- SITNFlash. (2020, April 23). *The History of Artificial intelligence Science in the News*. Science in the News. https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

- SECARB Phase III Site: SECARB Citronelle Projects (Anthropogenic Test). (n.d.). netl.doe.gov. https://netl.doe.gov/coal/carbon-storage/atlas/secarb/phase-III/citronelle-projects
- SECARB Phase III Site: SECARB Cranfield Project (The Early Test), Mississippi. (n.d.). netl.doe.gov. https://netl.doe.gov/coal/carbon-storage/atlas/secarb/phase-III/cranfield-project#geoDetails
- Shi, J., & Durucan, S. (2005). CO2Storage in Deep Unminable Coal Seams. *Oil & Gas Science and Technology Revue D'IFP Energies Nouvelles*, *60*(3), 547–558. https://doi.org/10.2516/ogst:2005037
- Sashurin, A. (2020). Current problems and objectives in geomechanics. *Mining informational and analytical bulletin*, 188-198. https://doi.org/10.25018/0236-1493-2020-31-0-188-198
- Song, Y., Jun, S., Na, Y., Kim, K., Jang, Y., & Wang, J. (2023). Geomechanical challenges during geological CO₂ storage: A review. *Chemical Engineering Journal*, 456, 140968. https://doi.org/10.1016/j.cej.2022.140968
- Vida, Gholami, Mohaghegh D. Shahab, and Maysami Mohammad. "Smart proxy modeling of SACROC CO2-EOR." Fluids 4.2 (2019): 85.
- Viewd Mechanical. (2020). What suffix of shear stress represent?/Shear stress sign convention/ Positive and negative shear [Video]. YouTube. https://www.youtube.com/watch?v=I5EM5LYRMcA
- What is Computational Intelligence? IEEE Computational Intelligence Society. (n.d.). <u>https://cis.ieee.org/about/what-is-</u> <u>ci#:~:text=Computational%20Intelligence%20(CI)%20is%20the,Fuzzy%20Systems%20and%20E</u> <u>volutionary%20Computation</u>.
- Wilson, K. C., and Louis J. Durlofsky. "Computational optimization of shale resource development using reduced-physics surrogate models." SPE Western Regional Meeting. OnePetro, 2012.
- Wikipedia contributors. (2023b, October 20). Geomechanics. Wikipedia. https://en.wikipedia.org/wiki/Geomechanics
 Zhang, Y. Modeling Uncertainty and Risk in Carbon Capture and Storage. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2013a.
- Zhang, Y.; Sahinidis, N. Uncertainty Quantification in CO2 Sequestration Using Surrogate Models from Polynomial Chaos Expansion. Ind. Eng. Chem. Res. 2013b, 52, 3121–3132.
- Zhang, Y.; Sahinidis, N. Developing Surrogate Models for CO2 Sequestration Using Polynomial Chaos Expansion; NRAP-TRS-III-004- 2013; NRAP Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, USA, 2013c; p. 20.
- Zhang, H., Tang, Y., Okubo, S., Peng, S., & Chen, C. (2021). undefined. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.769621