A Machine Learning-Based Forecasting Tool for Carbon Dioxide Enhanced Gas Recovery Associated with Carbon Storage in Shale Gas Reservoirs

Zhang, Yuming

https://hdl.handle.net/1880/116541

Downloaded from PRISM Repository, University of Calgary
A MACHINE LEARNING-BASED FORECASTING TOOL FOR CARBON DIOXIDE ENHANCED GAS RECOVERY ASSOCIATED WITH CARBON STORAGE IN SHALE GAS RESERVOIRS

by

Yuming Zhang

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER SCIENCE

GRADUATE PROGRAM IN CHEMICAL ENGINEERING

CALGARY, ALBERTA
MAY, 2023

© Yuming Zhang 2023
Abstract

The use of machine learning (ML) in the study of CO\textsubscript{2} storage is a hot topic, but the feasibility of applying ML to complex reservoir conditions, such as those with CO\textsubscript{2} enhanced shale gas recovery (CO\textsubscript{2}-ESGR), remains untested. This research aims to develop a novel multi-task ML workflow using simulation data to provide an effective and timesaving approach for the evaluation of CO\textsubscript{2} storage and enhanced gas recovery during CO\textsubscript{2}-ESGR. A three-stage reservoir model, including primary recovery, CO\textsubscript{2}-ESGR, and CO\textsubscript{2} storage, using the Barnett Shale formation as the basis of research was developed to determine the sensitivities of various parameters and provide the data for subsequent ML modeling. A composite ML model, which consists of a combination of Artificial Neural Network (ANN) and Long Short-term Memory (LSTM) algorithms, was used to provide better forecasting performance. The two parts of this ML model include the prediction of gas production during CO\textsubscript{2}-ESGR and CO\textsubscript{2} storage during the CO\textsubscript{2} storage stage; its other functionalities include a switching criterion and prediction of CO\textsubscript{2} storage during CO\textsubscript{2}-ESGR. These two parts were optimized by tuning the algorithm hyperparameters and further combined into one model to validate the simulation results. The results show that both gas recovery during CO\textsubscript{2}-ESGR and CO\textsubscript{2} storage during the CO\textsubscript{2} storage stage were mainly affected by seven key features, which include matrix porosity, fracture porosity, pressure gradient, geothermal gradient, hydraulic fracture conductivity, and adsorption properties of CO\textsubscript{2} and CH\textsubscript{4}. By applying the ML model, the results on gas recovery during CO\textsubscript{2}-ESGR were matched with simulation results with an accuracy of 99.9986\% for CH\textsubscript{4} production and an accuracy of 99.995\% for CO\textsubscript{2} production. Additionally, the monthly CO\textsubscript{2} storage during the CO\textsubscript{2} storage process achieved a match to the simulation results with an accuracy of 99.7\%. Furthermore, the ML model was economized 99.9\% on computational efficiency compared to reservoir simulations. The greatest advantage of this novel ML model is that it can accurately predict production and storage in both CO\textsubscript{2}-ESGR and CO\textsubscript{2} storage without the need of historical production data.
Preface

This thesis is an original work by the author. No part of this thesis has been previously published.
Acknowledgements

With my greatest gratitude, I would like to thank my thesis advisor, Professor Dr. Zhangxing (John) Chen for his unwavering support, guidance, and encouragement throughout my graduate studies and the completion of this thesis. Thanks to his tireless explanations and insightful feedback, I was able to understand the complex concepts and come up with innovative approaches to this research. His expertise and patience have been indispensable in shaping my research and helping me achieve my academic goals.

This research has been made possible by contributions from the Faculty of Graduate Studies and the Department of Chemical and Petroleum Engineering, which provided me with a learning platform, making my graduate degree possible.

Special thanks go to Haoming Ma, Zhenqian Xue, Dr. Kai Zhang, and Ziteng Huang for their guidance, valuable comments, and discussions on this research, which helped to complete it.

Finally, I would like to express my appreciation to my family, my wife Jiaqi Zhuang, my in-laws, and my friends. Thanks for their love, endless encouragement, and understanding, supporting me through difficult times and pushing me forward.

I am deeply grateful to all these individuals and institutions for their contribution to my academic and personal development. Without their support, this research wouldn’t have been possible.
# Table of Contents

ABSTRACT .......................................................................................................................... II
PREFACE ............................................................................................................................. III
ACKNOWLEDGEMENTS ....................................................................................................... IV
TABLE OF CONTENTS .......................................................................................................... V
LIST OF FIGURES ................................................................................................................. VI
LIST OF TABLES .................................................................................................................. VII
LIST OF SYMBOLS, ABBREVIATIONS, AND NOMENCLATURE ...................................... VIII

1. INTRODUCTION ............................................................................................................... 9

2. METHODOLOGY ............................................................................................................. 11
   2.1 WORKFLOW .................................................................................................................. 11
   2.2 RESERVOIR MODEL .................................................................................................... 12
   2.3 MUTUAL ADSORPTION IN SHALE GAS RESERVOIR .................................................. 16
   2.4 DATA PREPARATION AND PREPROCESSING ............................................................ 18
      2.4.1 Sensitivity Analysis and Data Preparation ............................................................. 18
      2.4.2 Data Preprocessing ............................................................................................... 19
   2.5 MACHINE LEARNING MODELS .................................................................................. 20
      2.5.1 Artificial Neural Network (ANN) .......................................................................... 20
      2.5.2 Long Short-term Memory (LSTM) .................................................................... 22
      2.5.3 Combined Model ................................................................................................. 26
      2.5.4 Optimization ....................................................................................................... 27
      2.5.5 Evaluation Metrics ............................................................................................... 28

3. RESULTS AND DISCUSSION ......................................................................................... 29
   3.1 RESERVOIR SIMULATION .......................................................................................... 29
   3.2 SENSITIVITY ANALYSIS ............................................................................................ 31
   3.3 OPTIMIZATION OF THE ML MODEL ....................................................................... 35
      3.3.1 Optimization of Learning Rate .......................................................................... 36
      3.3.2 Optimization of the Number of Hidden Layers and Hidden Units .................. 38
   3.4 ML MODEL PERFORMANCE ....................................................................................... 49
      3.4.1 Comparison of Each Individual Component and The Composite Model ............ 50
      3.4.2 Performance of ML Model with and without History Data Provided ................. 54

3.3.1 CONCLUSIONS ......................................................................................................... 56

REFERENCE ...................................................................................................................... 59
List of Figures

Figure 1: ML based Workflow for CO₂-ESGR and CO₂ Storage ................................................................. 12
Figure 2: Reservoir Model Heterogeneity (Color bar standing for the horizontal porosity) .................... 14
Figure 3: Relative Permeability ................................................................................................................ 15
Figure 4: Data Preprocessing .................................................................................................................... 20
Figure 5: Typical ANN Structure and Computation .................................................................................. 21
Figure 6: General Construction of a LSTM Unit ......................................................................................... 23
Figure 7: Back Propagation Computation in a LSTM Unit ....................................................................... 23
Figure 8: Composite Model Construction ................................................................................................. 27
Figure 9: Gas Recovery Factor for Primary Stage of Recovery in Reservoir Simulation ......................... 30
Figure 10: Gas Mass Rate during CO₂-ESGR Process ............................................................................. 30
Figure 11: Gas Mass Rate during CO₂ Storage Stage .............................................................................. 31
Figure 12: Sensitivity Analysis on CH₄ Mass Production Rate during CO₂-ESGR Process .................. 33
Figure 13: Sensitivity Analysis on CO₂ Mass Production Rate during CO₂-ESGR Process .................... 33
Figure 14: Sensitivity Analysis on CO₂ Mass Storage Rate during CO₂ Storage Process ..................... 34
Figure 15: ANN Learning Rate Performance on CO₂ Production Prediction during CO₂-ESGR Process .... 37
Figure 16: ANN One Hidden Layer Performance on CO₂ Production Prediction during CO₂-ESGR Process ......................................................................................................................... 39
Figure 17: ANN Modified Two Hidden Layers Performance on CO₂ Production Prediction during CO₂-ESGR Process ....................................................................................................................... 40
Figure 18: ANN Four Hidden Layers Performance on CO₂ Production Prediction during CO₂-ESGR Process ............................................................................................................................... 42
Figure 19: ANN Hidden Layer’s Optimum Condition on Prediction of CO₂ Production during CO₂-ESGR Process with Minimum Values Evaluation ............................................................................ 44
Figure 20: ANN Hidden Layer’s Optimum Condition on CO₂ Production Prediction during CO₂-ESGR Process with Average Values Evaluation ......................................................................... 45
Figure 21: Optimized ANN Model Performance on CO₂ Production Prediction during CO₂-ESGR Process. 48
Figure 22: Optimized ANN Model Train and Test Set Performance on CO₂ Production Prediction during CO₂-ESGR Process ......................................................................................................... 49
Figure 23: Solely LSTM Component Performance with History Data Provided on CO₂ Production Prediction during CO₂-ESGR Process ........................................................................................... 51
Figure 24: ANN Component Predictions with History Data Provided ....................................................... 52
Figure 25: Composite Model Performance with History Data Provided on CO₂ Production Prediction during CO₂-ESGR Process ........................................................................................................... 53
Figure 26: Composite Model Predictions with History Data Provided ..................................................... 54
Figure 27: Composite Model Predictions with History Data Provided during CO₂-ESGR Process ........ 55
Figure 28: Composite Model Predictions without History Data Provided ......................................... 56
Figure 29: Composite Model Predictions without History Data Provided during CO₂-ESGR Process ........ 56
List of Tables

Table 1: Reservoir Input Parameters........................................................................................................16
Table 2: Candidate Parameters used in Sensitivity Analysis.........................................................................18
Table 3: Hyperparameters Involved in Optimization and Range Selections....................................................28
Table 4: ML Input Features and Target Values. ..............................................................................................35
Table 5: Pre-setting of ANN Part Structure and Hyperparameters.................................................................35
Table 6: Pre-setting of LSTM Part Structure and Hyperparameters...............................................................36
Table 7: Optimum Conditions for ML Model Part Predicting CO₂ Production.................................................46
Table 8: Optimum Conditions for ML Model Part Predicting CH₄ Production..............................................46
Table 9: Optimum Conditions for ML Model Part Predicting CO₂ Storage....................................................46
Table 10: Optimized Epochs for Each Part of Composite Model ..................................................................49
Table 11: Errors for the Optimized Model.....................................................................................................49
Table 12: Candidate Parameters used in Sensitivity Analysis.......................................................................50
# List of Symbols, Abbreviations, and Nomenclature

<table>
<thead>
<tr>
<th>Nomenclature</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEA</td>
<td>International Energy Agency</td>
</tr>
<tr>
<td>$\text{CO}_2$</td>
<td>Carbon Dioxides</td>
</tr>
<tr>
<td>GCS</td>
<td>Geological CO$_2$ Sequestration</td>
</tr>
<tr>
<td>IPCC</td>
<td>Intergovernmental Panel on Climate Change</td>
</tr>
<tr>
<td>CO$_2$-ESGR</td>
<td>CO$_2$ Storage with Enhanced Shale Gas Recovery</td>
</tr>
<tr>
<td>BCM</td>
<td>Billion Cubic Meters</td>
</tr>
<tr>
<td>Tcf</td>
<td>Trillion Cubic Feet</td>
</tr>
<tr>
<td>Gt</td>
<td>Gigatonne</td>
</tr>
<tr>
<td>NGL</td>
<td>Natural Gas Liquid</td>
</tr>
<tr>
<td>CMG</td>
<td>Computer Modelling Group</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>FWU</td>
<td>Farnsworth Unit</td>
</tr>
<tr>
<td>MOPSO</td>
<td>Multiple Swarm Optimization Algorithms</td>
</tr>
<tr>
<td>NPV</td>
<td>Net Present Value</td>
</tr>
<tr>
<td>LSTM</td>
<td>Long Short-Term Memory</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean Absolute Error</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Squared Error</td>
</tr>
<tr>
<td>$R^2$</td>
<td>Coefficient of Determination</td>
</tr>
<tr>
<td>OGIP</td>
<td>Original Gas In Place</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>Length of Model Dimension</td>
</tr>
<tr>
<td>$W$</td>
<td>Width of Model Dimension</td>
</tr>
<tr>
<td>$H$</td>
<td>Height of Model Dimension</td>
</tr>
<tr>
<td>$D$</td>
<td>Depth of Reservoir</td>
</tr>
<tr>
<td>$h$</td>
<td>Thickness of Grid</td>
</tr>
<tr>
<td>$T$</td>
<td>Reservoir Thickness</td>
</tr>
<tr>
<td>$P$</td>
<td>Reservoir Pressure</td>
</tr>
<tr>
<td>$\phi_m$</td>
<td>Average Matrix Porosity</td>
</tr>
<tr>
<td>$K_m$</td>
<td>Average Matrix Permeability</td>
</tr>
<tr>
<td>$\phi_f$</td>
<td>Average Fracture Porosity</td>
</tr>
<tr>
<td>$K_f$</td>
<td>Average Fracture Permeability</td>
</tr>
</tbody>
</table>
1. Introduction

As is widely acknowledged within the scientific community, climate change presents an urgent challenge for the world at large and requires a concerted worldwide effort to curtail greenhouse gas emissions to limit its scale. The sum of fossil fuel consumption and escalating energy demands account for the vast majority of global emissions and have yielded irreversible consequences for the climate, compelling nations worldwide to prioritize an emission reduction. A recent report from the International Energy Agency (IEA) revealed that numerous countries have pledged to achieve net-zero carbon dioxide (CO$_2$) emission by 2050. A CO$_2$ emission reduction has become one of the primary focuses of research in recent years as it accounts for a significant portion of greenhouse gas emissions, and geological CO$_2$ sequestration (GCS) has emerged as a promising technology in this regard and has been endorsed by the Intergovernmental Panel on Climate Change (IPCC) [1,2]. Furthermore, the IEA forecasted a 0.8% annual rise in the consumption of natural gas, with a projected total of 4240 BCM by 2025. To address the rising demand for comparatively clean natural gas and concurrently meet CO$_2$ geo-storage objectives, a novel approach called CO$_2$ Storage with Enhanced Shale Gas Recovery (CO$_2$-ESGR) has been introduced. CO$_2$-ESGR entails pressurization and injection of captured CO$_2$, which is a product of industrial operations, into shale formations, allowing for the stabilization of CO$_2$ in a solid or dissolved form in the formations rather than being released into the atmosphere. Moreover, the storage of massive amounts of solid or dissolved CO$_2$ in shale formations expedites the displacement of natural gas as CO$_2$ presents with preferential adsorption compared to CH$_4$ in organic formations, which promotes their exchange in the formations and enables greater productivity relative to conventional depleted oil reservoirs [2,3,4,5,6]. Research has demonstrated that a range of applications of CO$_2$-ESGR are global in nature, wherein Godec [8] suggested that approximately 2505 trillion cubic feet (Tcf) of shale gas resources can be economically produced and 281 gigatons (Gt) of CO$_2$ can be stored worldwide. In the United States, it has been estimated that approximately 134 Gt of CO$_2$ can be stored in 19 gas shale formations [9]. Heller and Zoback [10] further supported the efficacy of CO$_2$-ESGR through their findings that indicate that CO$_2$ has roughly two to three times the adsorptive capacity of methane in shale reservoirs at 40°C [10]. Despite these promising results, commercial utilization of CO$_2$-ESGR is still in its preliminary stages, with only a limited number of field tests having been conducted. Louk et al. [11,12] conducted a small-scale CO$_2$ injection test in the Chattanooga Shale Formation and their results indicate that CO$_2$ injection led to a
significant boost in gas flow rates and helped alleviate gas condensate blockage in the reservoir, which ultimately resulted in the successful storage of 59% of the injected CO₂. Zhan et al. [13] and Liu et al. [14] numerically simulated CO₂-ESGR and their results demonstrated the efficacy of this technique in the enhancement of shale gas production and CO₂ sequestration. However, owing to a complex relationship between the adsorption of CO₂ and CH₄ and the geological complexity of shale formations, the computational requirements for numerical models are quite substantial. To contend with the computational requirements of modelling and simulation, recent studies have focused on an application of artificial intelligence (AI) and machine learning (ML) in numerical simulations, wherein their advantages lie in rapid response speeds, ease of use, cost optimization, and robust generalization capacity. ML has demonstrated enormous potential in solving a wide range of reservoir engineering problems, including the simulation of CO₂ storage in enhanced oil recovery [15-17]. Li et al. [18] used a random forest regression algorithm to predict oil production and CO₂ storage in a CO₂-WAG project based on data from the ShengLi Oilfield in China and their results were shown to be of high accuracy and computational efficiency. Hung et al. [19] developed an Artificial Neural Network (ANN) model that accurately predicts oil recovery and CO₂ storage capacity in residual oil zones. You et al. [20] used ML to optimize a CO₂-WAG project in the Farnsworth Unit (FWU) in Texas, USA and compared the performance of neural networks with a single hidden layer and those with multiple layers; they found that a multi-layer neural network model performed better and used it to develop a robust optimization framework that incorporated multiple objective particle swarm optimization (MOPSO) algorithms. Furthermore, their results demonstrated the efficacy of MOPSO in calculating cumulative oil production, CO₂ sequestration, and project net present value (NPV) [21]. These studies demonstrate the potential of ML for the optimization of CO₂ storage for enhanced oil recovery. However, thus far, the application of ML in CO₂-ESGR processes remains largely unexplored. Moataz et al. [22] designed an ANN model to predict incremental enhanced CH₄ during an EGR process, but their calculations focused on the forecasts at a certain time and not on the amount of CO₂ storage during and after CO₂-ESGR. Based on the current information, no study has been conducted on time-series forecasting in this area. To achieve higher performance in the application of ML models and to better capture a correlation among data, this study proposes a multi-task ML-based workflow for the CO₂-ESGR process. The proposed workflow aims to forecast CO₂ production, CH₄ production, and CO₂ storage during the CO₂-ESGR stage as well as to predict CO₂ storage in the CO₂ storage stage; moreover, this methodology considers time-series data, allowing for a more accurate and efficient completion of the entire process. Furthermore, this workflow also enables the prediction of reference time, allowing for the alteration of an ML model from a CO₂-ESGR process to a pure CO₂ storage stage. In the following sections, the workflow, methodology including the reservoir model, ML model input dataset, ML model algorithms, optimization methods, evaluation metrics, results, discussions, and
conclusions are presented.

2. Methodology

2.1 Workflow

A two-component, ML input and model construction, comprehensive workflow, with multiple objectives and optimization procedures, is discussed herein, as shown in Fig. 1.

The decision to include an input stage, which was primarily designed to provide data for model construction, was made due to a lack of CO2-ESGR field data; thus, a CMG-GEM (compositional simulator) reservoir simulation approach was utilized to generate an artificial dataset in the place of field data. The first step of the input stage entails the establishment of a three-stage reservoir model that can completely simulate the primary recovery of CH4 after hydraulic fracturing, CO2-ESGR, and pure CO2 storage. Primary recovery is only simulated to advance the reservoir simulation to the point where reservoir conditions are such that CO2-ESGR should be logically applied; thus, the primary recovery production data is not included in the input dataset for an ML model. Subsequently, a series of candidate parameters, including reservoir (formation porosity and permeability), CO2 and CH4 adsorption, and operational properties (hydraulic fracturing parameters), were selected from the parameters in the reservoir model. A sensitivity study was conducted to determine the parameters that most significantly affected the output values; afterwards these parameters were assigned varying values and iteratively input into the CMG-GEM model to obtain a wide range of results. The parameters with the highest sensitivities were selected as the input features of the ML model, and the results of the CMG-GEM model were used as targets of the ML model.

After the selection of the input features, the raw data was cleaned and preprocessed, an ML model was established and an optimization plan for its hyperparameters was designed. The ML model consists of two main components, whose functions are training and then forecasting production data for CO2 and CH4 during CO2-ESGR, and predicting the quantity of CO2 storage throughout the CO2 storage stages, respectively. The performance of the ML model is constantly evaluated according to a series of evaluation metrics, and optimization occurs when the model results do not satisfy the evaluation metrics in terms of accuracy and reliability. Optimization entails iterative looping of varied hyperparameter values using trial-and-error until the results are satisfactory, which allows for the identification of a combination of hyperparameters that resulted in optimal model performance. A switching criterion
was implemented in the ML model to determine the specific timestep in which the mode of a reservoir should be altered from CO\textsubscript{2}-ESGR to CO\textsubscript{2} storage based on a ratio between the produced CO\textsubscript{2} and CH\textsubscript{4} to prevent the negative effects of excess CO\textsubscript{2} production. CO\textsubscript{2} storage during CO\textsubscript{2}-ESGR was calculated by using the difference between CO\textsubscript{2} injection and predicted CO\textsubscript{2} production. The prediction of CO\textsubscript{2} storage throughout the pure CO\textsubscript{2} storage stage in the second component of the ML model was achieved via similar optimizing procedures to those in the first component of the ML model, which can achieve optimized results. Upon completion of the optimization, the first and second components were integrated into a single comprehensive model. During the testing phase, the ML model initially received inputs into the first component and proceeded with the prediction of CO\textsubscript{2}-ESGR production until the predetermined switching criterion was met. At this point, the model switches from the first component to the second component in order to predict the quantity of CO\textsubscript{2} storage in the CO\textsubscript{2} storage stage.

To verify the performance of the ML model, a random set of parameter values was used to output two sets of results from the reservoir and the ML models.

![Figure 1: ML based Workflow for CO\textsubscript{2}-ESGR and CO\textsubscript{2} Storage.](image)

### 2.2 Reservoir Model

A dual-permeability reservoir model was developed based on the parameters of the Barnett shale formation using CMG-GEM, which is a commercial compositional simulator that can be used to consider a multi-component system [24, 25, 26, 27]. GEM modelling was shown to be able to take into account a natural fracture effect, model assumptions
include initial gas composition being composed of only CH$_4$, a reservoir presents with only connate water, and injected gas components are defined as CO$_2$ only. The initial model was validated using work published by Yu et al. [28]. Subsequently, the parameters were modified, and heterogeneity was added to fit this study. The Kozeny-Carman equation, as shown in Eq. (1), is used to correlate porosity and permeability in both the matrix and natural fracture media [29]; therefore, reservoir heterogeneity can be determined by assuming that the literature values for porosity and permeability are the average values in this study [28, 30].

$$k = \alpha \frac{\phi^3}{(1-\phi)^2}$$  \hspace{1cm} (1)

$$\alpha = k \frac{(1-\phi)^2}{\phi^3}$$  \hspace{1cm} (2)

where $k$ denotes permeability [mD]; $\alpha$ is a regression coefficient; $\phi$ represents porosity. To calculate $\alpha$, Eq. (1) can be modified into Eq. (2) using known average porosity and permeability values. The porosity values for each block were generated using Monte Carlo simulations [31], with a normal distribution and preset average values. The distributions of porosity are shown in Figs. 2 (a) and (b), and the corresponding permeabilities of the matrix and fracture media were calculated using Eq. (1). The well development and hydraulic fracture system for this reservoir were designed as follows: two horizontal wells were set 900 ft apart, as shown in Figs. 2 (a) and (b). Staggered fractures were created, with 100 ft of stage spacing and half-length set at 200 ft. Hydraulic fracture conductivity was calculated based on the provided fracture width and intrinsic permeability. The operating duration of this simulation was set to 100 years, wherein primary recovery lasted for 35 years, until the incremental increase in the recovery factor became negligible. Subsequently, CO$_2$-ESGR was initiated by altering well 1 from a producer into an injector well, with the injection pressure set to a value that did not exceed the reservoir pressure in order to avoid potentially fracturing the reservoir matrix. The potential OGIP in this reservoir was estimated to be approximately 1.1425x10$^6$ MMSCF, including CH$_4$ in both free and adsorbed forms. The relative permeability curves for the matrix and fractures are shown in Figs. 3 (a) and (b), respectively, and the detailed input parameter values for the reservoir model are shown in Table 1.
Figure 2: Reservoir Model Heterogeneity (Color bar standing for the horizontal porosity)
Figure 3: Relative Permeability

a) Matrix

b) Fracture
Table 1: Reservoir Input Parameters.

<table>
<thead>
<tr>
<th>Reservoir Properties [28]</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Dimension</td>
<td>10000 (L) * 10000 (W) * 100 (H)</td>
<td>ft³</td>
</tr>
<tr>
<td>Depth, D</td>
<td>6481</td>
<td>ft</td>
</tr>
<tr>
<td>Grid Thickness, h</td>
<td>100</td>
<td>ft</td>
</tr>
<tr>
<td>Reservoir Temperature, T</td>
<td>150</td>
<td>°F</td>
</tr>
<tr>
<td>Reservoir Pressure, P</td>
<td>3500</td>
<td>psi</td>
</tr>
<tr>
<td>Average Matrix Porosity, ϕₘ,i</td>
<td>0.041</td>
<td>–</td>
</tr>
<tr>
<td>Average Matrix Permeability, Kₘ,i</td>
<td>0.0003</td>
<td>mD</td>
</tr>
<tr>
<td>Average Fracture Porosity, ϕ₇,i</td>
<td>0.0004</td>
<td>–</td>
</tr>
<tr>
<td>Average Fracture Permeability, K₇,i</td>
<td>1.4</td>
<td>mD</td>
</tr>
<tr>
<td>Anisotropy</td>
<td>0.1</td>
<td>–</td>
</tr>
<tr>
<td>Nature Fracture Spacing</td>
<td>0.5</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operational Properties</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom Hole Pressure</td>
<td>500</td>
</tr>
<tr>
<td>Injection Pressure</td>
<td>3500</td>
</tr>
<tr>
<td>Injection rate</td>
<td>4000000</td>
</tr>
<tr>
<td>Half Length</td>
<td>200</td>
</tr>
<tr>
<td>Hydraulic Fracture Conductivity</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adsorption Properties</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Langmuir Pressure CH₄</td>
<td>702</td>
</tr>
<tr>
<td>Langmuir Volume CH₄</td>
<td>55</td>
</tr>
<tr>
<td>Langmuir Pressure CO₂</td>
<td>927</td>
</tr>
<tr>
<td>Langmuir Volume CO₂</td>
<td>35.5</td>
</tr>
</tbody>
</table>

As excess CO₂ production, which is the eventual result of CO₂-ESGR after CO₂ breakthrough from the production well, damages operational facilities and is costly to separate, this study contends that a cessation criterion must be established. Kidnay et al. [32] suggested that the maximum CO₂ weight in a production stream should be 5.2%, and this study used this criterion in the inspection of the ratio of produced CO₂ to CH₄. After reaching the maximum CO₂ weight, the production well was then also converted into an injection well to transition from CO₂-ESGR to pure CO₂ storage. The period for the switching condition varied depending on the values of the input parameters for the different simulation cases created using CMG-GEM.

### 2.3 Mutual Adsorption in Shale Gas Reservoir

The mechanism of mutual adsorption of CO₂ and CH₄ and the method of determination of adsorption properties are discussed herein. Natural gas in shale reservoirs is generally stored as free gas in inner pore spaces, dissolved gas in kerogen, and adsorbed gas in organic matter and mudstone. Previous studies have shown that, depending on the reservoir conditions, adsorbed shale gas can account for 20-85% of the original shale gas in place [33, 34, 35]. During primary recovery in CO₂-ESGR (enhanced shale gas recovery) projects, hydraulic fracturing is typically used to
produce the majority of the free shale gas, leaving a significant amount of adsorbed gas in a formation. Previous experimental results have suggested that pre-adsorbed CH\textsubscript{4} can be replaced in the shale matrix by injected CO\textsubscript{2} as CO\textsubscript{2} presents with stronger adsorption ability than CH\textsubscript{4}, which aids in the enhancement of CH\textsubscript{4} recovery during CO\textsubscript{2}-ESGR [36, 37]. To model a monolayer gas adsorption volume in a shale gas reservoir at a certain pressure, \( P \), the most commonly used Langmuir model [38] was applied, as shown in Eq. (3).

\[
V(P) = \frac{V_L P}{P + P_L} \tag{3}
\]

where \( V(P) \) is the gas adsorption volume at pressure \( P \); \( V_L \) represents the Langmuir volume or the maximum volume of gas that can be adsorbed at infinite pressure; \( P_L \) denotes the Langmuir pressure, corresponding to the pressure at which half of the Langmuir volume can be adsorbed. Although the Langmuir model is simple to use, it has a limitation because it only considers the calculations of a single layer. To overcome this limitation, Brunauer et al. [39] proposed the following BET model for multi-layer adsorption:

\[
V(P) = \frac{V_m C P}{(P_o - P)[1 + (C - 1)P/P_o]} \tag{4}
\]

where \( P_o \) represents the gas saturation pressure; \( V_m \) is the maximum gas adsorption volume; \( C \) is the constant of the net heat of adsorption. In CMG-GEM, the extended Langmuir isotherm for multicomponent adsorption is described [40,41] as follows:

\[
\omega_i = \frac{\omega_{i,max} B_i y_{i,aqu}}{1 + \sum_j B_j y_{j,aqu}} \tag{5}
\]

where \( \omega_i \) is the number of moles of the adsorbed component \( i \) per unit mass of rock; \( \omega_{i,max} \) is the maximum number of moles of the adsorbed component \( i \) per unit mass of rock; \( B_i \) is a parameter for the Langmuir isotherm relation; \( y_{i,aqu} \) is the molar fraction of the adsorbed component \( i \) in the aqueous phase. The final calculations are obtained by summing over the adsorbed components, as described in reference [28]. In this study, the aqueous phase was modeled as a gas phase due to the absence of other phases in the reservoir.
2.4 Data Preparation and Preprocessing

2.4.1 Sensitivity Analysis and Data Preparation

The initially established reservoir model provided the learning dataset for the ML model; however, in the model, a variety of parameters can affect the production of CO$_2$ and CH$_4$ during CO$_2$-ESGR as well as the amount of CO$_2$ stored during the subsequent CO$_2$ storage stage. Although myriad parameters can affect productivity and CO$_2$ storage, this study contends that only the most significant parameters that production and storage are sensitive to should be focused upon in order to economize on computational resources. To aid in the selection of the above-mentioned parameters, a sensitivity analysis was conducted to screen the important parameters that will be used as input for the ML model. A total of 15 parameters were selected and summarized for testing in the sensitivity analysis, as shown in Table 2. These parameters were chosen from four categories, including reservoir, operational, hydraulic fracturing, and adsorption properties, such that they can affect various aspects of production values. Moreover, the dataset was enriched via many methods, including the merging of pressure and temperature with depth in order to calculate gradient values, the assumption of a positive relationship between porosity and permeability, and the consideration of only matrix and fracture porosities. Additionally, multipliers or proportions between the adsorption properties of CO$_2$ and CH$_4$ were utilized.

Table 2: Candidate Parameters used in Sensitivity Analysis.

<table>
<thead>
<tr>
<th>Candidate Parameters</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure Gradient</td>
<td>0.54004</td>
<td>psi/ft</td>
</tr>
<tr>
<td>Geothermal Gradient</td>
<td>0.023144444</td>
<td>F/ft</td>
</tr>
<tr>
<td>Matrix Porosity</td>
<td>0.041</td>
<td></td>
</tr>
<tr>
<td>Fracture Porosity</td>
<td>0.0004</td>
<td></td>
</tr>
<tr>
<td>Anisotropy on K Direction</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Nature Fracture Spacing</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Relative Permeability Curve Exponent Krg</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Relative Permeability Curve Exponent Krw</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Relative Permeability Curve Exponent Krow</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Relative Permeability Curve Exponent Krog</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Bottom Hole Pressure</td>
<td>500</td>
<td>psi</td>
</tr>
<tr>
<td>Half Length</td>
<td>200</td>
<td>ft</td>
</tr>
<tr>
<td>Hydraulic Fracture Conductivity</td>
<td>10</td>
<td>ft. mD</td>
</tr>
<tr>
<td>Max Adsorbed Mass Multiplier</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Adsorption Constant Multiplier</td>
<td>0.761290323</td>
<td></td>
</tr>
</tbody>
</table>
This study elects to add or subtract 5% to each parameters’ base value during the iterative sensitivity analyses rather than adding or subtracting fixed values so that the differences in the magnitudes of the variables can be accounted for. Simulations of varying CMG-GEM models, with varied input parameter values, were conducted and the simulation outputs were used as a reference to rank the input parameters in terms of significance/sensitivity. Specifically, the percentage of change in simulation outputs driven by the variation of each parameter was examined and ranked by their magnitudes. To ensure simulation consistency, the seven most sensitive parameters, which accounts for approximately half of the total, were selected as the input features of the ML model. Each feature was assigned a range of values, with several random values selected, various CMG-GEM models were created and were used to generate massive amounts of data for the input dataset of the ML model.

### 2.4.2 Data Preprocessing

After sensitivity analysis and data preparation, the previously obtained data was then cleaned, normalized, and split.

The first step involved conducting data cleaning to simplify and improve data quality. This included combining datasets from various simulation cases into one and removing duplicate entries and missing values that were produced during the process. This was followed by a data normalization step, as shown in Eq. (6) [42,43].

\[ X_{\text{norm},i} = \frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \]  

(6)

where \( X_{\text{norm},i} \) is the scaled value of \( i \)th observation after being normalized; \( X_{\text{max}} \) represents the largest value among features; \( X_{\text{min}} \) denotes the minimum value among features. Prior to scaling, the parameters in the raw data presented with large differences in terms of scale, therefore the parameters with large values may be assigned a higher weight during the calculation by the algorithm, which may be harmful to the accuracy of the prediction. As such, the value of each parameter was normalized in the range of 0 to 1, which correspond to \( X_{\text{min}} \) to \( X_{\text{max}} \), respectively, in order to maintain fairness for the algorithm’s judgment in terms of assigned weights, as shown in Fig. 4 (a). Finally, the data was then split as ML models require sufficient data to train and collect information from the dataset, however some of the data is required for validation purposes. Therefore, in this study, the dataset was split into three subsets: the training, validation, and test sets, as shown in Fig. 4 (b). The training set was the set of data used for training the ML models, which required the largest amount of data, the test set was used to test the model's performance, and the validation set
was used in the validation stages during the training process to overcome potential overfitting. Considering these requirements, the processed dataset was further split in a ratio of 8:1:1 for the training, validation and test sets [45].

![Data Normalization and Data Split Strategy](image)

**Figure 4: Data Preprocessing.**

### 2.5 Machine Learning Models

The different types of ML models that were utilized in this study along with the processing of model inputs are discussed herein. ML models are comprised of a single or an assembly of algorithms that allow for the learning of apparent or unseen patterns, sequences, or clusters in a given dataset without explicit programming, which can be further used for the predictions for the dataset. In this study, two models: an Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) were utilized [44,45].

#### 2.5.1 Artificial Neural Network (ANN)

ANNs are a type of computational model that are designed to imitate the structure and functions of the human brain and are particularly adept at identifying nonlinear relationships within multivariate datasets [45]; wherein, typical ANN architecture consists of interconnected layers of neurons, which communicate with each other to perform computations and generate predictions. The three primary layers in ANN are the input, output, and hidden layers, where the input layer receives external data and determines the input shape, the output layer provides the final results of the computation and shapes them according to the desired number of target values, and the hidden layer, which contains multiple nodes, performs the bulk of the computation and provides the network with its processing power.
Neurons in the input layer are connected to neurons in the hidden layer via weighted connections, with bias values also being provided as initial inputs that are further updated during the training process. Activation functions are placed at the connections between neurons to facilitate the computations performed by the network. Although ANNs are well-suited for analyzing the relationships and correlations among multiple parameters and target values, their effectiveness is limited when it comes to learning time-series data. Specifically, ANNs are only able to provide predictions at singular instances, without considering the influence of previous or subsequent data.

![Figure 5: Typical ANN Structure and Computation](image)

$$y_{i}^{(t)}(x) = g(\sum_{j=1}^{N} W_{i,j}^{(t)} x_{j} + b^{(t)})$$ (7)

$$W_{i,j}^{(t+1)} = W_{i,j}^{(t)} - \alpha \frac{\partial \text{loss}}{\partial W_{i,j}^{(t)}}$$ (8)

$$b^{(t+1)} = b^{(t)} - \alpha \frac{\partial \text{loss}}{\partial b^{(t)}}$$ (9)

where $y_{i}^{(t)}$ represents the results for layer $i$ at the current time step and considering the result for the last computation as the prediction for this time step, noting $y_{i}^{(t+1)}(x)$; $x_{j}$ is the $j$th value of input; $g()$ is the activation function used in this case; $W_{i,j}^{(t)}$ and $W_{i,j}^{(t+1)}$ represent the weights for layer $i$ and entry $j$ at the current and next time steps, respectively; $b^{(t)}$ and $b^{(t+1)}$ are the biases for the current and next time steps, respectively; $\alpha$ denotes the selected learning rate; and $\frac{\partial \text{loss}}{\partial W_{i,j}^{(t)}}$ and $\frac{\partial \text{loss}}{\partial b^{(t)}}$ represent the partial derivatives of selected loss functions with respect to $W_{i,j}^{(t)}$ and $b^{(t)}$. To train an ANN model, a series of steps must be followed. Firstly, the weights and biases must be randomly initialized, subsequently
forward propagation must be applied using Eq. (9) to determine the values of the neurons and make the first prediction from left to right, as depicted by the black arrow in Fig. 4. Once forward propagation is completed, backward propagation is performed, which calculates updated gradient values for weight and bias using Eq. (8) and (9), respectively, from right to left, as depicted by the red arrow in Fig. 4. This step is crucial as it enables the network to learn from its errors and adjust the weights and biases to improve its predictions. This process of forward propagation, backward propagation, and weight and bias updating is repeated until the loss function reaches a satisfactory level.

Once the ANN has been trained on the available data, it can be used to make predictions on unseen data by saving the final values of the weights and biases and repeating the forward propagation process.

### 2.5.2 Long Short-term Memory (LSTM)

Sequential or time-series data is the predominant type of data for the oil and gas industry as it can be used to capture historical, current, and even forecasted production or injection data [45]. Recurrent Neural Networks (RNNs) are purposefully designed to handle sequential data and are capable of using the previous time step's output to generate the current time step's result. However, their effectiveness in capturing long-term dependencies in data is limited due to the issues attributed to vanishing gradients, which refers to the gradient shrinkage induced improper updating of weights and biases [23]. To address this issue, Hochreiter and Schmidhuber proposed an advanced version of RNN, known as Long Short-Term Memory (LSTM) [46, 47], whose structure is shown in Fig. 6 [48]. In a single LSTM unit, in addition to the standard hidden cell state for information flow, there are unique modules: a connected memory cell state which operates as a conveyor belt for preserving and controlling information flow, and three operating units (Forget gate, Update gate, Output gate) used for removing, updating, and storing data. These modules ensure stable error propagation paths by enforcing constant error carousels, thereby overcoming the problem of vanishing gradients. With the aid of these modules, LSTM becomes capable of preserving or discarding information from data, retaining it in the cell state, and relating it to other information [23, 45].
The detailed calculations are presented in Eq. (10)-(32), with Eq. (10)-(16) specifically describing the forward propagation calculations and Eq. (17)-(32) describing being the calculations for backward propagation.

\[
\begin{align*}
\Phi_f^{(t)} &= \sigma(W_f [a^{(t-1)}, x^{(t)}] + b_f) \\
\Phi_u^{(t)} &= \sigma(W_u [a^{(t-1)}, x^{(t)}] + b_u)
\end{align*}
\]
\[ c^{(t)} = \tanh (W_c [a^{(t-1)}, x^{(t)}] + b_c) \]  
(12)

\[ \Pi_o^{(t)} = \sigma (W_o [a^{(t-1)}, x^{(t)}] + b_o) \]  
(13)

\[ c^{(t)} = \Pi_f^{(t)} \star c^{(t-1)} + \Pi_u^{(t)} \star \hat{c}^{(t)} \]  
(14)

\[ a^{(t)} = \Pi_o^{(t)} \star \tanh (c^{(t)}) \]  
(15)

\[ y_{pred}^{(t)} = g(W_y a^{(t)} + b_y) \]  
(16)

where \( a^{(t-1)} \) and \( a^{(t)} \) are the hidden states at a previous and current time step, respectively; \( x^{(t)} \) is the input at the current time step; \( c^{(t-1)} \) and \( c^{(t)} \) are memory cell states at a previous and current time step, respectively; \( \Pi_f^{(t)} \) represents the forget gate, having a value in range of 0 to 1; \( \Pi_u^{(t)} \) represents the update gate, having a value in a range of 0 to 1; \( \Pi_o^{(t)} \) represents the output gate, having a value in a range of 0 to 1; \( \hat{c}^{(t)} \) denotes the candidate value having a value in a range of -1 to 1; \( W \) represents the weight; \( b \) represent the bias; \( y_{pred}^{(t)} \) outputs or prediction values; \( g(), \sigma(), \tanh() \) are different activation functions used for different cases; In Fig. 5, a forward calculation occurs in a single LSTM cell from left to right, as shown in Eq. (10)-(16). During this process, the previous time step's information is evaluated to determine whether to add or ignore changes to the current state. The forget gate, represented by \( \Pi_f^{(t)} \), monitors the correlated data structure from the hidden state \( a^{(t-1)} \). When the forget gate value approaches 0, it indicates that the stored state from the previous cell state \( c^{(t-1)} \) should be deleted to match the new feature. The candidate value \( \hat{c}^{(t)} \) contains information from the current time step that may be stored in the current cell state \( c^{(t)} \), depending on whether the update gate \( \Pi_u^{(t)} \) passes its values. The update gate decides which aspects of the candidate \( \hat{c}^{(t)} \) to add to the cell state \( c^{(t)} \). When the update gate value approaches 0, it prevents the corresponding value from the candidate being added onto the hidden state. Finally, the output gate \( \Pi_o^{(t)} \) determines what is sent as the prediction (output) of the time step.

\[ dy_o^{(t)} = da^{(t+1)} \star \tanh (c^{(t+1)}) \star \Pi_o^{(t)} \star (1 - \Pi_o^{(t)}) \]  
(17)

\[ dy_c^{(t)} = (dc^{(t+1)} \star \Pi_u^{(t)} + \Pi_o^{(t)} \star (1 - \tanh^2 (c^{(t+1)})) \star \Pi_u^{(t)} \star da^{(t+1)}) \star (1 - (\hat{c}^{(t)})^2) \]  
(18)
\[ dy_u^{(t)} = (dc^{(t+1)} \cdot \tilde{c}^{(t)} + \Pi_o^{(t)} \cdot (1 - \tanh^2(c^{(t+1)})) \cdot c^{(t)} \cdot da^{(t+1)}) \cdot \Pi_u^{(t)} \cdot (1 - \Pi_u^{(t)}) \]  \tag{19}

\[ dy_f^{(t)} = (dc^{(t+1)} \cdot c^{(t-1)} + \Pi_o^{(t)} \cdot (1 - \tanh^2(c^{(t+1)})) \cdot c^{(t-1)} \cdot da^{(t+1)}) \cdot \Pi_f^{(t)} \cdot (1 - \Pi_f^{(t)}) \]  \tag{20}

\[ dW_f = dy_f^{(t)} \left[ \frac{a^{(t-1)}}{x_i} \right]^T \]  \tag{21}

\[ dW_u = dy_u^{(t)} \left[ \frac{a^{(t-1)}}{x_i} \right]^T \]  \tag{22}

\[ dW_c = dy_c^{(t)} \left[ \frac{a^{(t-1)}}{x_i} \right]^T \]  \tag{23}

\[ dW_o = dy_o^{(t)} \left[ \frac{a^{(t-1)}}{x_i} \right]^T \]  \tag{24}

\[ db_f = \sum_{batch} dy_f^{(t)} \]  \tag{25}

\[ db_u = \sum_{batch} dy_u^{(t)} \]  \tag{26}

\[ db_c = \sum_{batch} dy_c^{(t)} \]  \tag{27}

\[ db_o = \sum_{batch} dy_o^{(t)} \]  \tag{28}

\[ W^{(t+1)} = W^{(t)} - dW^{(t)} \]  \tag{29}

\[ b^{(t+1)} = b^{(t)} - db^{(t)} \]  \tag{30}

\[ da^{(t-1)} = dx^{(t)} = W_f^T dy_f^{(t)} + W_u^T dy_u^{(t)} + W_c^T dy_c^{(t)} + W_o^T dy_o^{(t)} \]  \tag{31}

\[ dc^{(t-1)} = dc^{(t+1)} \cdot \Pi_f^{(t)} + \Pi_o^{(t)} \cdot (1 - \tanh^2(c^{(t+1)})) \cdot \Pi_f^{(t)} \cdot da^{(t+1)} \]  \tag{32}

where \( dy_o^{(t)}, dy_c^{(t)}, dy_u^{(t)} \) and \( dy_f^{(t)} \) represents the gradient values of the output gate, candidate value, update gate and forget gate respective of the cost function, respectively. \( dW^{(t)} \) is the gradient of weight and \( db^{(t)} \) is the gradient of bias.
Similar to the ANN approach, in order to optimize the values for weight and bias, it is necessary to calculate the partial derivatives with respect to selected loss functions, followed by Eq. (17)-(32). The calculations proceed from right to left in Fig. 7, resulting in the values of gradients for weight and bias. To achieve optimization, the entire process of forward and backward propagations must be looped through, and the values for weight and bias must be updated using the appropriate equation until the loss reaches an acceptably low level.

2.5.3 Combined Model

When predicting the storage of CO$_2$ or productions of CH$_4$ and CO$_2$ over a given period of time, time series forecasting must be utilized to provide more realistic predictions. Though LSTM models are well-suited for the modeling of long-term dependencies in sequential data, they are not effective at capturing the non-linear relationships between multiple parameters and their respective target values. To address this limitation, Yu Shi et al. [23] proposed a model that combines both ANN and LSTM models; wherein, data must first be input into the LSTM layers subsequently they are then transferred to the ANN model. This approach is advantageous in that it allows LSTM to focus on the time-series properties of the data while leaving ANN to assist with learning relationships among features and target values, resulting in a more accurate prediction. However, a major limitation of the application of LSTM in reservoir simulations is that its training requires historical production data to support further predictions. LSTM can be applied for existing wells that have production data; however, it cannot be utilized for new production wells. This study proposes a novel approach that entails the reversal of the sequential relationship between ANN and LSTM, which will allow the novel model to take maximal advantage of the benefits of both ANN and LSTM while concurrently being able to avoid LSTM’s requirement for historical data.
Fig. 8 depicts this study’s modified approach to the composite model of ANN and LSTM, which is used for the prediction of CH4, CO2 production, or CO2 storage based on values for pressure, temperature gradient, and other related features. The proposed model initially inputs the features and target values into the ANN’s input layer, where a non-linear mapping relationship between them is determined. The number of hidden layers and units in the ANN were then optimized to achieve the best performance in terms of the selected evaluation metrics, afterwards temporary predictions were generated using the optimized ANN, which were then converted into a feature and fed into the LSTM along with historical data. The number of hidden layers and units in the LSTM were also optimized to produce the final prediction values. Once the optimization work on this model is complete, it can use the provided feature values to generate historical data for internal use in the LSTM model to carry out the prediction results.

2.5.4 Optimization

In the previous section, a composite model that combines ANN and LSTM models was introduced to predict future values of a target variable. The composite model consists of a combination of ANN and LSTM models; however, these two models must be separately optimized then merged into the optimized composite model. During the optimization process, three key hyperparameters were focused upon: the learning rate of the selected optimizer, the number of hidden layers, and the number of hidden units. The optimizer, which is an algorithm that alters the attributes of the ML model, used for this case employed the gradient descent algorithm to train the model and minimize the value.
of the loss function. The learning rate of the optimizer determines the step size of this process; wherein, if selected learning rate values are too large or too small, the model may easily overshoot local minima or may present with significant decreases in computational efficiency. Thus, a range of values for learning rate were initially randomly selected on a logarithmic scale and the ML model was iteratively utilized to determine optimal learning rate, afterwards the ML model was applied iteratively beginning with 1 hidden layer and adding 1 hidden layer each time until performance ceased to improve. Thereafter, the optimal learning rate and number of hidden layers were applied in the ML model to determine the optimal number of hidden units in a similar manner to that of the learning rate. Each optimization was conducted by looping the training process and monitoring preset evaluation metrics, and each loop was terminated when the values of these metrics dropped low enough. The optimization details were summarized, as shown in Table 3, including the range of values tested for each hyperparameter. By optimizing these hyperparameters for both the ANN and LSTM models separately, and then combining them into a composite model, improved performance was achieved compared to the individual models alone.

Table 3: Hyperparameters Involved in Optimization and Range Selections.

<table>
<thead>
<tr>
<th>Hyperparameters</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimizer Learning Rate</td>
<td>0.00000001</td>
<td>0.1</td>
</tr>
<tr>
<td>Hidden Units</td>
<td>2</td>
<td>1001</td>
</tr>
<tr>
<td>Hidden Layers</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

2.5.5 Evaluation Metrics

In order to assess the predictive accuracy and performance of the machine learning (ML) models, four evaluation metrics were selected: mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination ($R^2$) [49]. Specifically, MSE was calculated using Eq. (34), MAE was computed using Eq. (35), RMSE was obtained using Eq. (36), and $R^2$ was determined using Eq. (37).

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - y_{i}^{pred})^2
\]  \hspace{1cm} (33)

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y_{i}^{pred}|
\]  \hspace{1cm} (34)

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_{i}^{pred})^2}
\]  \hspace{1cm} (35)

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - y_{i}^{pred})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}
\]  \hspace{1cm} (37)
where $y_i$ represents the true values; $y_i^{pred}$ is the prediction values; $y_i^{avg}$ denotes the mean of true values; $n$ is sample quantity. Lower values of MSE, MAE, and RMSE indicate more precision predictions. $R^2$ is a coefficient that ranges from 0 to 1, with values approaching 1 indicating higher performance. In this study, MAE was used to measure the average residual of the dataset, while MSE and RMSE were used to measure the variance and standard deviation of the residual, respectively. Furthermore, $R^2$ was used to assess how well the predictor’s features could explain the variation in target values. To provide a more comprehensive assessment, the time cost associated with making predictions using machine learning models was also considered, which allows for a more thorough evaluation of model performance.

3. Results and Discussion

3.1 Reservoir Simulation

The simulation results based on three target values were analyzed from the preset base case of the study, providing corresponding interpretations. Production for the simulation began in 2000 and the results of the three stages, including primary production, CO$_2$-ESGR, and CO$_2$ storage, in terms of recovery factor, production rate, and storage rate are shown in Figs. 9, 10, and 11, respectively.

In the initial phase of primary recovery, the gas production rate remained unchanged at its maximal production rate from 2000-2021, as shown in Fig. 10, and recovery factor maintained its maximal rate of increase in the same period, as shown in Fig. 9, which demonstrates the sufficient pressure support from the pressure differences created by hydraulic fracturing. Subsequently, from 2021-2035, the gas production rate began to rapidly decline, as shown in Fig. 10, and, at the same time, the incremental increase in recovery factor also began to significantly decrease, as shown in Fig. 9, which demonstrates the rapid depletion of pressure support. In 2035, the gas production rate and incremental increase in recovery factor becomes negligible, which indicates the need for additional driving forces in the reservoir.

CO$_2$-ESGR was implemented in 2035, which entails the alteration of well #1 from producer to injector, and immediately induced a significant increase in gas production rate, as shown in Fig. 10, owing to the additional driving force provided by the injection of CO$_2$ as well as the superior adsorption of CO$_2$ compared to CH$_4$ which quickly take the place of the originally adsorbed CH$_4$. This trend continued until 2041, which was when CO$_2$ broke through and first
appeared in the production stream and CO$_2$-ESGR was discontinued; wherein, the amount of produced CO$_2$ satisfied the pre-set criteria for the shift from CO$_2$-ESGR to CO$_2$ storage as the separation of CO$_2$ and CH$_4$ is costly and difficult. The breakthrough of CO$_2$ is due to the combination of the effects of reaching maximum saturation of adsorbed CO$_2$ and the overpassing of CH$_4$ by CO$_2$ due to the differences in density between the two in the formation. Although CO$_2$ storage began in 2041, this study elected to continue to display predicted production data past 2041, as shown in Fig. 10, to show that the production of CO$_2$ quickly outpaced that of CH$_4$ after CO$_2$ breakthrough. As the CO$_2$ injection rate of each well was set to a constant rate, the CO$_2$ storage rate throughout CO$_2$-ESGR was immediately doubled throughout CO$_2$ storage because well #2 was also changed into an injector, as shown in Fig. 12. Throughout the CO$_2$ storage process, the amount of injected CO$_2$ was assumed to be the amount of stored CO$_2$ as no more producing wells remained. As reservoir pressure approaches the original value, the amount of stored CO$_2$ will rapidly decline because of the saturation of CO$_2$ storage and a lack of difference between injection and reservoir pressures, as shown in Fig. 12.

![Gas Recovery Factor](image1)

Figure 9: Gas Recovery Factor for Primary Stage of Recovery in Reservoir Simulation.

![Gas Mass Rate](image2)

Figure 10: Gas Mass Rate during CO$_2$-ESGR Process.
3.2 Sensitivity Analysis

The results of the sensitivity analyses of the input parameters that most significantly affect three target values, including CH₄ and CO₂ mass production rates and CO₂ mass storage rate, are herein presented using tornado charts, as shown in Figs. 12-14.

The sensitivity analysis revealed that the seven most important features were consistent across all three target values, including the maximum adsorbed mass multiplier, adsorption constant multiplier, matrix properties, fracture properties, pressure gradient, geothermal gradient, and hydraulic fracture conductivity. Adsorption properties, including maximum adsorbed mass multiplier and adsorption constant multiplier, indicate the extent of competitive adsorption of CO₂ over CH₄; wherein, higher values indicate more exchange of CO₂ for adsorbed CH₄ and consequently greater production of CH₄. Matrix and fracture porosity represent similar concepts; however, greater matrix porosity results in more trapped CH₄ as there is more impermeable space, while greater fracture porosity results in more produced CH₄ as there are more production channels within the reservoir matrix. Additionally, the mechanism of hydraulic fracture conductivity is similar to that of fracture porosity. Higher geothermal gradients indicate higher temperatures at specific depths, which results in increased activity and desorption of adsorbed CO₂ molecules, while higher pressure gradients indicate higher pressures at specific depths, which results in more driving force for CH₄ production and stable adsorption of CO₂ in the reservoir. The impact of rock and fluid properties and wellbore conditions on the target values was found to be negligible due to the low density of gas.

Fig. 12 displays the sensitivity of CH₄ production to the 15 selected parameters; wherein, the seven most significant parameters, in order of most to least significant, are maximum adsorbed mass multiplier, hydraulic fracture connectivity, matrix porosity, geothermal gradient, adsorbed constant multiplier, pressure gradient, fracture porosity, and anisotropy. A 5% increase in maximum adsorbed mass multiplier resulted in a 6% increase in CH₄ production,
while a 5% decrease in matrix porosity resulted in a 3.125% increase in CH₄ production, as shown in Fig. 12. Of these seven most significant parameters, only a decrease in matrix porosity resulted in increased CH₄ production, while increases in the other parameters resulted in increased CH₄ production.

Fig. 13 shows the sensitivity of CO₂ production to the selected features; wherein the two most dominant features are maximum adsorbed mass multiplier and adsorbed constant multiplier, which are both adsorption properties. 5% increases in maximum adsorbed mass multiplier and adsorbed constant multiplier result in 63.76% and 48.16% decreases in produced CO₂, respectively, while 5% decreases in maximum adsorbed mass multiplier and adsorbed constant multiplier result in 181.52% and 97.31% increases in produced CO₂, respectively. Therefore, it is clear that CO₂ production is more sensitive to negative changes in the value of adsorption properties because CO₂ production occurs more rapid than CO₂ adsorption with decreased adsorption capabilities.

In Fig. 14, the seven most significant features, in terms of CO₂ storage during the CO₂ storage stage, in order of most to least significant, are matrix porosity, maximum adsorbed mass multiplier, geothermal gradient, hydraulic fracture conductivity, adsorbed constant multiplier, fracture porosity, and pressure gradient. CO₂ storage is more sensitive to positive changes in matrix porosity and maximum adsorbed mass multiplier; wherein, 5% increases in matrix porosity and maximum adsorbed mass multiplier resulted in 52.03% and 43.57% increases in CO₂ storage, respectively. Geothermal gradient is the only feature for which CO₂ storage increased with negative changes in feature value.

For CH₄ production in CO₂-ESGR, adsorption properties are the most dominant features, which indicates that the competitive adsorption between CO₂ and CH₄ is very significant in this stage. Additionally, CH₄ production is also affected by features that affect the porosity of the reservoir and channels of production, including hydraulic fracture conductivity and matrix and fracture porosity. For CO₂ storage in the CO₂ storage stage, it was found that competitive adsorption did not contribute as much as it did in the CO₂-ESGR stage, which was caused by the absence of CH₄ production in the process, rather it depends more on features that affect space in the reservoir and the stability of adsorbed CO₂, which include matrix porosity and geothermal gradient, respectively.
Figure 12: Sensitivity Analysis on CH$_4$ Mass Production Rate during CO$_2$-ESGR Process.

Figure 13: Sensitivity Analysis on CO$_2$ Mass Production Rate during CO$_2$-ESGR Process.
Figure 14: Sensitivity Analysis on CO$_2$ Mass Storage Rate during CO$_2$ Storage Process.

In the sensitivity analysis, seven features were identified and assigned 21 random values within a specified range, as shown in Table 4. Each value range was collected from various sources across different geographical locations worldwide [28,50-56]. This approach was taken to consider a wide range of reservoir conditions and to provide greater diversity in the resulting dataset. Using these feature values, multiple simulation cases were generated by modifying the base case created with CMG, as described in the previous section. The results obtained from CMG-GEM were then used as inputs for the machine learning model; wherein, the resulting input dataset contained 510,774 rows.
Table 4: ML Input Features and Target Values.

<table>
<thead>
<tr>
<th>Features</th>
<th>Maximum</th>
<th>Minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Langmuir Adsorption Constant Multiplier</td>
<td>3</td>
<td>0.5</td>
</tr>
<tr>
<td>Maximal Adsorbed Mass Multiplier</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Average Matrix Porosity</td>
<td>0.1</td>
<td>0.01</td>
</tr>
<tr>
<td>Average Fracture Porosity</td>
<td>0.0005</td>
<td>0.0003</td>
</tr>
<tr>
<td>Pressure Gradient</td>
<td>1.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Geothermal Gradient</td>
<td>0.026</td>
<td>0.012</td>
</tr>
<tr>
<td>Hydraulic Fracture Conductivity</td>
<td>50</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Target Values

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CO$_2$ Production</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>CH$_4$ Production</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>CO$_2$ Storage</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

3.3 Optimization of the ML Model

The composite model, which is composed of ANN and LSTM components, for CO$_2$-ESGR and CO$_2$ storage was optimized according to the methodology in Section 2.5.4 and its performance is herein thoroughly analyzed. To validate model accuracy, a random set of parameter values were utilized in the model and its results were compared with those obtained from CMG-GEM simulation. The ML models and corresponding plots were developed using the Keras API [59] and the matplotlib library, both of which are open-source Python libraries. The structural and hyperparameter parameters of the base case are listed in Tables 5 and 6, respectively.

Table 5: Pre-setting of ANN Part Structure and Hyperparameters.

<table>
<thead>
<tr>
<th>ANN Basic Information</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation Function</td>
<td>ReLU</td>
</tr>
<tr>
<td>Optimizer Algorithm</td>
<td>Adam</td>
</tr>
<tr>
<td>Number of Input Layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of Dropout Layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of Output Layers</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANN Base Case Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hidden Layers</td>
</tr>
<tr>
<td>Number of Hidden Units</td>
</tr>
</tbody>
</table>
Table 6: Pre-setting of LSTM Part Structure and Hyperparameters.

<table>
<thead>
<tr>
<th>LSTM Basic Information</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation Function</td>
<td>ReLU</td>
</tr>
<tr>
<td>Optimizer Algorithm</td>
<td>Adam</td>
</tr>
<tr>
<td>Number of Input Layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of Dropout Layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of Output Layers</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LSTM Base Case Setting</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hidden Layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of Hidden Units</td>
<td>50</td>
</tr>
</tbody>
</table>

Many hyperparameters have been considered in the design of the optimal composite ML model for the prediction of CO$_2$ mass production rate during CO$_2$-ESGR. This study has elected to utilize the Rectified Linear Unit (ReLU) activation function to avoid the vanishing gradient effect in hidden layers and the ‘Adam’ optimizer algorithm was applied due to its superior performance in the training processes. To minimize the effects of overfitting, both ANN and LSTM components were equipped with a dropout layer with a 0.2 percentage level, which regulates the data by randomly terminating the computation in 20% of the units during training. To ensure efficient data analysis and to avoid potential overfitting, the training process was limited to 50 epochs with a preset early stop condition to terminate the training when MSE no longer improves. The study also focused on the prediction of CO$_2$ mass production rate during CO$_2$-ESGR process to avoid redundancy in discussions. Findings for other variables were summarized in tables.

### 3.3.1 Optimization of Learning Rate

The evaluation of error and prediction for the optimization of ANN learning rate in the prediction of CO$_2$ production during CO$_2$-ESGR is shown in Fig. 15. Several evaluation metrics, including MSE, RMSE, and MAE, their respective validation results, and prediction time are evaluated in unison to determine the value of the optimal learning rate. An ideal learning rate should be one that balances desirable values for all evaluation metrics as well as prediction time. In Fig. 15 (a), it can be seen that the evaluation metrics exhibit a concave trend with minimum errors occurring at approximately 0.001, however 0.0001 was selected instead of 0.001 as 0.0001 presented with a far more favorable prediction time than that of 0.001 though they had similar errors. All errors were kept at a low level, with values lower than 0.075, indicating the reliability of the model. Fig. 15 (b), which shows the $R^2$ of this ANN model, indicates that 0.001 learning rate corresponds to the maximum $R^2$ value, however due to the significant difference between the prediction times of 0.001 and 0.0001, as shown in Fig. 15 (a), 0.0001 considered the most optimal learning rate, providing the model with both high accuracy and a reliable $R^2$ score.
a) Evaluation on Errors and Prediction Time

b) Evaluation on R2 Score

Figure 15: ANN Learning Rate Performance on CO₂ Production Prediction during CO₂-ESGR Process.
3.3.2 Optimization of the Number of Hidden Layers and Hidden Units

Using the previously determined optimal learning rate (0.0001), the number of hidden layers and hidden units were optimized using the previously used evaluation metrics. For one hidden layer, the magnitude of error decreases with increases in the number of hidden units; however, the errors were still decreasing even with the maximum number of hidden units, which indicates that prediction performance can still be improved, as shown in Fig. 16 (a). Similarly, $R^2$ still exhibits an increasing trend at the highest number of hidden units, as shown in Fig. 16 (b); thus it is clear that one hidden layer is not the optimal configuration.

![Graph showing the relationship between number of hidden units and error metrics.](image)

a) Evaluation on Errors and Prediction Time
In examining the performance of models with more than one hidden layer, it was observed that strong fluctuations occurred when the number of hidden units in the previous layer was lower than that in the subsequent layer. To stabilize the results, this study elected to limit the number of hidden units in subsequent sets to between 0.73-1 times that of the preceding set. At greater numbers of hidden units, the errors of two hidden layers began to stabilize at lower than 0.02, while those of one hidden layer did not, as shown in Figs. 17 (a) and 16 (a), respectively. For two hidden layers, the errors stabilized after the numbers of hidden units reached approximately [155,114], which suggests that the optimal number of hidden layers is close to 2. Similarly, the $R^2$ reached a maximum value of 0.99, which was greater than those of one hidden layer, and began to plateau after the numbers of hidden units reached [45,33], as shown in Fig. 17 (b). Though the performance of 2 hidden layers is quite desirable, the performance of more hidden layers must be verified prior making the decision for the optimal number of hidden layers.

Figure 16: ANN One Hidden Layer Performance on CO$_2$ Production Prediction during CO$_2$-ESGR Process.
**Figure 17: ANN Modified Two Hidden Layers Performance on CO$_2$ Production Prediction during CO$_2$-ESGR Process.**

Figs. 18 (a) and (b) show the performance of a modified version of a four-hidden-layer ANN used to predict CO$_2$
production during the CO$_2$-ESGR process. It was observed that errors reached the lower part of the trends much earlier, and all trends began to fluctuate again, as seen in Fig. 18 (b). This effect was caused by the algorithm's loss function reaching a local minimum and bouncing back and forth in the nearby region. Based on these findings, it was determined that the optimal number of hidden layers for this model should be between 2-4.
b) Evaluation on $R^2$ Score

Figure 18: ANN Modified Four Hidden Layers Performance on CO\textsubscript{2} Production Prediction during CO\textsubscript{2}-ESGR Process.

The decision to select the optimal number of hidden layers for predicting CO\textsubscript{2} production using an ANN was challenging due to the ambiguity in relying solely on error values and $R^2$ scores. To overcome this issue, the minimum and average values of the evaluation metrics were considered for each hidden layer. Furthermore, the potential overfitting during the process was checked by assessing the difference between the minimum errors of the training and validation sets.

Figs. 19 (a) and (b) depict the results based on the minimum errors calculated for 1 to 4 hidden layers in the ANN model. Specifically, Fig. 19 (a) illustrates that the largest reduction in errors occurred when transitioning from 1 to 2 hidden layers, with a minor increase in prediction time. This finding was further supported by Fig. 19 (b), where the R2 score reached its maximum value of over 0.99 for the two-hidden-layer situation. Fig. 19 (c) shows the difference between the minimum errors of the test and validation sets, which was minimal at 2 hidden layers and overfitting did not occur. As 2 hidden layers was the most optimal point for all three charts, 2 was selected as the most optimal number of hidden layers.
a) Evaluation on Minimum Errors and Prediction Time

b) Evaluation on Difference of Minimum Errors between Test and Validation Set
c) Evaluation on Minimum $R^2$ Scores

Figure 19: ANN Hidden Layer’s Optimum Condition on Prediction of CO$_2$ Production during CO$_2$-ESGR Process with Minimum Values Evaluation.

In Figs. 20 (a) and (b), the red bars and columns represent the standard deviations and values of the evaluation metrics, including MSE, MAE, RMSE, and prediction time, respectively. As number of hidden layers increased, prediction time increased while errors decreased. There were rapid drops in errors from one to two hidden layers and prediction time rose slightly in both Figs. 20 (a) and (b), thus 2 hidden layers was selected as most optimal. Furthermore, the bars representing the standard deviation in both sub-figures are relatively short, suggesting that the results obtained are reliable and robust.
a) Evaluation on Average Errors and Standard Deviation

![Graph showing average errors and standard deviation for different number of hidden layers.]

b) Evaluation on Difference of Average Errors between Test and Validation Set

![Graph showing difference in average errors for different number of hidden layers.]

Figure 20: ANN Hidden Layer’s Optimum Condition on CO₂ Production Prediction during CO₂-ESGR Process with Average Values Evaluation.
Based on the analysis conducted, it is recommended that a two-hidden-layer implementation be used as the optimal condition for the ANN component of the model to predict CO$_2$ production during the CO$_2$-ESGR process. Furthermore, in Fig. 17 (a), [212,212] corresponds to a trough in the prediction time after reaching stable errors for the evaluation metrics. Although the prediction time was not at its lowest level at this point, the 0.2-second difference between the times was deemed insignificant. Furthermore, the R2 score reached its maximum value in Fig. 17 (b) at [212,212]. Thus, the optimum setting for this component of the model is to employ two hidden layers and apply \[212,212\] hidden units for each corresponding layer. Similar procedures were applied to analyze other parts of the model, and the results are summarized in Tables 7, 8, and 9.

Table 7: Optimum Conditions for ML Model Part Predicting CO$_2$ Production.

<table>
<thead>
<tr>
<th>ANN</th>
<th>Optimizer Learning Rate</th>
<th>Hidden Units</th>
<th>Hidden Layers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0001</td>
<td>[212, 212]</td>
<td>2</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.01</td>
<td>114</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 8: Optimum Conditions for ML Model Part Predicting CH$_4$ Production.

<table>
<thead>
<tr>
<th>ANN</th>
<th>Optimizer Learning Rate</th>
<th>Hidden Units</th>
<th>Hidden Layers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0001</td>
<td>212, 212</td>
<td>2</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.0001</td>
<td>114</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 9: Optimum Conditions for ML Model Part Predicting CO$_2$ Storage.

<table>
<thead>
<tr>
<th>ANN</th>
<th>Optimizer Learning Rate</th>
<th>Hidden Units</th>
<th>Hidden Layers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0001</td>
<td>61, 45, 33</td>
<td>3</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.001</td>
<td>114</td>
<td>1</td>
</tr>
</tbody>
</table>
After completing the analysis to obtain optimum conditions for predicting CO₂ production during the CO₂-ESGR process using an ANN, the final construction of the model was executed. The final step in the optimization of ANN is to determine the optimal number of epochs to maximize both performance and computational efficiency without overfitting. The errors were calculated based on the ANN under the optimal condition of hidden layers and hidden units in the model construction, as depicted in Figs. 21 (a), (b), and (c). It was determined that 9 epochs fulfilled this condition for the MSE in Fig. 21 (a), MAE in Fig. 21 (b), and RMSE in Fig. 21 (c). The model was trained with the optimized number of epochs, resulting in final performance on MSE, MAE, and RMSE values of 0.00018, 0.005, and 0.01, respectively, which are considered very low. The performance of the optimized model on the test and train sets is presented in Figs. 22 (a) and (b), and a good match was observed. Notably, there was no significant distinction between the match in Fig. 22(a) and Fig. 22(b), suggesting that no overfitting occurred in this case. Furthermore, the R² scores for Fig. 22 in both the test set and train set cases remained at a high level of approximately 0.992, which verifies the reliability of these results. Similar approaches were also applied to other parts of the ML model, and detailed information is provided in Tables 10 and 11.
b) MAE Performance

Figure 21: Optimized ANN Model Performance on CO₂ Production Prediction during CO₂-ESGR Process.

a) RMSE Performance

a) Performance on Comparison of Test Data Values
Figure 22: Optimized ANN Model Train and Test Set Performance on CO₂ Production Prediction during CO₂-ESGR Process.

Table 10: Optimized Epochs for Each Part of Composite Model.

<table>
<thead>
<tr>
<th>ML Model Component</th>
<th>CO₂ Production</th>
<th>CH₄ Production</th>
<th>CO₂ Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>9</td>
<td>28</td>
<td>49</td>
</tr>
<tr>
<td>LSTM</td>
<td>10</td>
<td>15</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 11: Errors for the Optimized Model.

<table>
<thead>
<tr>
<th>Evaluation Metrics</th>
<th>CO₂ Production</th>
<th>CH₄ Production</th>
<th>CO₂ Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.00012</td>
<td>0.000040</td>
<td>0.0025</td>
</tr>
<tr>
<td>MAE</td>
<td>0.0036</td>
<td>0.0029</td>
<td>0.020</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.011</td>
<td>0.0063</td>
<td>0.05</td>
</tr>
</tbody>
</table>

### 3.4 ML Model Performance

This section presents the results of two comparisons conducted to evaluate the performance of different ML models: (1) a comparison of ANN, LSTM, and a composite model combining both, and (2) a comparison of the composite ML model's performance when trained with and without historical data. All comparisons were based on the verification of results between numerical simulation and ML models, with the numerical simulation results considered the reference or "true value". To generate the results presented in this section, a set of random feature values was used,
as shown in Table 12.

<table>
<thead>
<tr>
<th>Candidate Parameters</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure Gradient</td>
<td>0.75</td>
<td>psi/ft</td>
</tr>
<tr>
<td>Geothermal Gradient</td>
<td>0.023</td>
<td>F/ft</td>
</tr>
<tr>
<td>Matrix Porosity</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Matrix Permeability</td>
<td>0.0004</td>
<td>mD</td>
</tr>
<tr>
<td>Fracture Porosity</td>
<td>0.0004</td>
<td></td>
</tr>
<tr>
<td>Fracture Permeability</td>
<td>1.4</td>
<td>mD</td>
</tr>
<tr>
<td>Hydraulic Fracture Conductivity</td>
<td>13</td>
<td>ft.mD</td>
</tr>
<tr>
<td>Max Adsorbed Mass Multiplier</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Adsorption Constant Multiplier</td>
<td>1.2</td>
<td></td>
</tr>
</tbody>
</table>

### 3.4.1 Comparison of Each Individual Component and The Composite Model

In this section, the results of each individual model component and those of the composite model with history production data were individually compared with the results of the numerical simulation and one another. LSTM was used to explore data relationships at the previous, current, and subsequent time steps, while ANN was employed to determine internal relationships among features and target values. In Figs. 23 (a), (b), and (c), MSE, MAE, and RMSE values of LSTM predicted CO$_2$ production all demonstrated sufficiently low levels of error, but they all presented with overfitting as LSTM tends to ignore correlations among features and target values, which results in its poor performance and reliability when dealing with unseen data. In Figs. 21 (a), (b), and (c), the ANN performance presents with very little overfitting, however the magnitudes of error are higher than those of LSTM; this occurred because ANN only considers relationships between features and target values and ignores internal correlations within the time-series. In comparing the true and predicted CH$_4$ production, CO$_2$ production, and CO$_2$ storage, it is evident that the predicted ANN results exhibit clear deviations in some regions and accurate predictions in others, as shown in Figs. 24 (a), (b), and (c). The predictions of LSTM did not match the true production at all. As the switching criteria for CO$_2$-ESGR to CO$_2$ storage requires extremely precise predictions, the individual utilization of either LSTM or ANN cannot satisfy the criteria’s requirements.
Figure 23: Solely LSTM Component Performance with History Data Provided on CO₂ Production Prediction during CO₂-ESGR Process.
Figure 24: ANN Component Predictions with History Data Provided.

In order to accurately predict production data for the switching criteria, a composite model, consisting of both ANN and LSTM components, was utilized. As illustrated in Figs. 25 (a), (b), and (c), the composite model combines the advantages of both ANN and LSTM as it exhibits both limited or nonexistent overfitting and low error, which were the individual advantages of ANN and LSTM, respectively, as shown in Figs. 21 and 23, respectively. For example, the average MSE of the composite model was 70% lower than that of ANN and the entirety of the MSE did not exhibit any overfitting.

Figure 25: MSE Performance
The performance of predicting three selected target values using the composite model is illustrated in Fig. 26. The results demonstrate a clear improvement in predictions with the assistance of the combined model as there were no apparent deviations from the true value. To further illustrate the efficiency of the composite model, the time required for both numerical simulation and the ML model to generate results was compared; wherein, the traditional reservoir modelling completed by CMG-GEM took more than 2 hours, while the proposed composite ML model took only 12 seconds to complete the task. This significant difference in time spent highlights the extraordinary efficiency of the proposed ML model compared to the traditional simulator.
To determine the capability of the composite ML model to accurately predict production and storage data without the aid of production history, the results of the ML model both with and without production history were compared. Fig. 27 presents comparisons between true and predicted CO₂ production, CO₂ storage, and CH₄ production for CO₂-ESGR with history data provided. The CO₂-ESGR process involved the injection of CO₂ into the formation, displacing CH₄. Since the production of CO₂ was initially low, the storage rate of CO₂ remained constant during this period. Subsequently, as CO₂ broke through, more CO₂ was produced rather than stored in the formation, resulting in a decline in CO₂ storage. As previously discussed, owing to the high accuracy achieved by the proposed model, the trends of CO₂ and CH₄ production during the CO₂-ESGR process were well-matched. Consequently, the composite model exhibited high performance in predicting the switching date, with only a 3-day deviation from the true value. Additionally, the prediction of CO₂ retention during the CO₂-ESGR process was well-matched with the application of the composite model.
This study evaluated the performance of the combined model in predicting CO\textsubscript{2} and CH\textsubscript{4} production during the CO\textsubscript{2}-ESGR process, as well as CO\textsubscript{2} storage, without providing historical data, as shown in Figs. 28 (a), (b), and (c). This study found that the composite model's predictions for CO\textsubscript{2} production were consistently underestimated or delayed, resulting in inaccuracies in the prediction of CO\textsubscript{2} storage during the CO\textsubscript{2}-ESGR process as presented in Fig. 29. The overall performance of the composite model without the use of historical data was satisfactory, but the predicted switching date was delayed by 30 days. Furthermore, the performance of the composite model was influenced by the ANN component, as seen in Fig. 25, as the performance of the LSTM component relied heavily on the ANN-generated results.
3.3.1 Conclusions

This study presents a systematic multi-task ML based workflow for the investigation of the CO₂-ESGR and CO₂ storage process. A numerical simulation model was developed using CMG software and served as the source of artificial data for the ML model. To identify the most influential parameters, a sensitivity analysis was performed using pre-selected candidate parameters, and the results obtained from the numerical model were processed to generate the
input dataset. The input dataset was used to train a composite ML model composed of a combination of ANN and LSTM components. The hyperparameters of the model were tuned to optimize predictions. The performance of the model was validated using the numerical model's results, and its performance was compared with that of models using solely ANN or LSTM components. The study's detailed analysis revealed several key findings, including:

- The sensitivity analysis conducted in this study using the reservoir model provided, revealed the seven most influential parameters among the 15 candidates on the target values of CO₂ and CH₄ production and CO₂ storage. These parameters were the maximum adsorbed mass multiplier, adsorption constant multiplier, matrix porosity, fracture porosity, pressure gradient, geothermal gradient, and hydraulic fracture conductivity. In all cases, the adsorption properties showed their significance and demonstrated their contribution as the driving force for all three target values in both CO₂-ESGR and CO₂ storage.

- In this study, an optimized composite model was developed to predict CO₂ and CH₄ production and CO₂ storage. The optimized configuration of the ANN component for CO₂ production included 2 hidden layers and 212 hidden units in each layer, while the LSTM component for CO₂ production included 1 hidden layer with 114 hidden units. The same configurations were used for the ANN and LSTM for predicting CH₄ production. For CO₂ storage, the optimal number of hidden layers for the ANN and LSTM were 3 and 1, respectively, and the number of optimal hidden units in each layer were [61,45,33] and 114, respectively.

- The composite model proposed in this study has exhibited exceptional predictive performance in terms of CO₂ production, CH₄ production in CO₂-ESGR processes, and CO₂ storage during storage stages. Specifically, this model has successfully reduced the MSE values to approximately 0.000014, 0.00005, and 0.003, respectively, while attaining an R² score exceeding 0.99 for each category.

- When given a set of randomly generated feature values and historical data from the provided reservoir model, the composite model exhibited superior forecasting performance compared to the individual ANN and LSTM models, albeit with comparatively minor deviations in trend matching. The composite model accurately predicted the termination date for the CO₂-ESGR process, with an error of only three days. Furthermore, the proposed composite model demonstrated the ability to make predictions without relying on historical data, albeit at the cost of some accuracy. This reduction in accuracy can be attributed to the dominance of the ANN component in the composite model's results. Specifically, the error in predicting the switching condition of the CO₂-ESGR process was approximately 30 days, which represents a significant inaccuracy compared to that achieved with the support of historical data.

- In comparing the composite model to the reservoir model, the former demonstrated superior performance in
predicting CO\textsubscript{2} and CH\textsubscript{4} productions as well as CO\textsubscript{2} storage, as evidenced by the models' highly accurate results. The composite model's exceptional forecasting capabilities were further supported by its successful performance in a random case study, indicating its potential for generalization across various applications. Furthermore, the composite model significantly outperformed the reservoir model in terms of computational efficiency, effectively minimizing the time required to complete these tasks. Specifically, the reservoir simulation took 1.833 hours to generate the results, whereas the ML model only took 4 seconds, highlighting the significant improvement provided by the composite model.

- This workflow utilized a novel termination criterion for the CO\textsubscript{2}-ESGR process. To deal with the inherent inadequacies of composite ANN-LSTM ML models in terms of prediction without historical data, this study novelly reversed the sequential order of the ANN-LSTM components; this study utilized a general case study to prove the feasibility and generality of the ML model and the accuracy of the results.

- There is still room for improvement in both the workflow and machine learning model used. To increase the versatility of the system and better generalize it, this study recommends the consideration of cases with various CO\textsubscript{2} injection rates applied in the dataset. Furthermore, the inclusion of more field production data in the dataset can improve the model’s ability to accommodate real-life scenarios. Finally, optimizing the amount of data and architecture of the machine learning model could improve its ability to accurately predict outcomes without historical data.
Reference

2) IPCC. Climate change 2022: mitigation of climate change; 2022.


26) [CMG-GEM. [Computer Software]. Calgary, AB Canada; 2020.]


49) DeepLearning.AI by Andrew Ng


