Energy 320 (2025) 135313

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Development and comparison of reduced-order models for CO_2-enhanced oil recovery predictions

Haoming Ma [a,b,*], Sean T. McCoy [a], Zhangxin Chen [c,a,**]

[a] Department of Chemical and Petroleum Engineering, University of Calgary, Calgary, AB, Canada, T2N1N4
[b] Hildebrand Department of Petroleum and Geosystems Engineering, University of Texas at Austin, Austin, TX, USA, 78712
[c] Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, Zhejiang, 315200, China

ARTICLE INFO

Keywords:
Reduced-order model
CO_2-Enahanced oil recovery
Geological carbon storage
Reservoir simulation
Net-zero emissions

ABSTRACT

CO_2-enhanced oil recovery (CO_2-EOR) has been a mature and promising technology since the 1970s, offering a dual solution for energy production and carbon sequestration. Recent advances in reduced-order models (ROMs) using empirical analysis and artificial intelligence (AI) tools handle complex data efficiently. However, existing ROMs for CO_2-EOR often lack validation due to data confidentiality or are too case-specific for broader application. This paper introduces a framework to close these gaps, enabling the development and consistent comparison of generalized ROMs for CO_2-EOR with carbon capture and storage (CCS), even with traditional tools. A synthesis dataset (~3000 runs) was established to develop ROMs, which were validated using field data from both EOR and CCS perspectives. Three key findings are revealed. First, normalizing outputs with respect to CO_2 utilization illustrated a direct relationship between CCS and EOR. Second, generalized statistics-based ROMs reduced input complexity and validated field data but predicted fewer outputs. Machine learning-based ROMs predicted more outputs, supporting field operational decision-makings. Last, ROMs were particularly suitable for early-stage, large-scale CO_2-EOR assessments. This study extended the boundaries of developing generalized ROMs for CO_2-EOR and identified pros and cons across modeling approaches, contributing to net-zero goals and advancing sustainable and affordable energy future.

Nomenclature

Abbreviations	
AI	Artificial intelligence
ANN	Artificial neural network
BHP	Bottomhole pressure
CCS	Carbon capture and storage
CCUS	Carbon capture, utilization and storage
CGI	Continuous gas injection
CMG-GEM	Compositional reservoir simulator by Computer Modeling Group
CO_2-EOR	CO_2-enhanced oil recovery
GHG	Greenhouse gases
HCPV	Hydrocarbon pore volume
IEA	International Energy Agency
ML	Machine learning
ML-ROM	ML based ROM
Mtpa	Mt per annum
OOIP	Original oil in place
ROM	Reduced-order models
SI	Supporting information
SPE5	Fifth comparative solution project by society of petroleum engineers
stats-ROM	Statistical based ROM
STB	Standard barrel
SVM	Support vector machine
SWAG	Simultaneous water and gas
TRL	Technology readiness level
WAG	Water-altering-gas
XGBoost	Extreme gradient boosting

(*continued on next column*)

(*continued*)

Symbols	
B_o	Oil formation volume factor
$HCPV_{CO_2,inj}$	Injected CO_2 in HCPV scale
$HCPV_{CO_2,prod}$	Produced CO_2 in HCPV scale
M_{CO_2t}	Theoretical CO_2 storage capacity in mass
MAPE	Mean absolute percentage error
MSE	Mean squared error

(*continued on next page*)

* Corresponding author. Department of Chemical and Petroleum Engineering, University of Calgary, Calgary AB, Canada T2N1N4.
** Corresponding author. Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, Zhejiang, 315200, China.
E-mail addresses: haoming.ma@ucalgary.ca (H. Ma), zhachen@ucalgary.ca (Z. Chen).

https://doi.org/10.1016/j.energy.2025.135313
Received 3 December 2024; Received in revised form 16 February 2025; Accepted 26 February 2025
Available online 27 February 2025

(*continued*)

Abbreviations	
R^2	Coefficient of determination
$R_{CO_2,ret}$	CO_2 retention factor
RF_{cum}	Cumulative recovery factor
RF_{in}	Incremental oil recovery factor in percentage of OOIP
RMSE	Rooted mean squared error
RMSPE	Rooted mean square percentage error
$UF_{CO_2,net}$	Net CO_2 utilization factor
$UF_{CO_2,gross}$	Gross CO_2 utilization factor
$V_{CO_2,inj,rc}$	Volumetric CO_2 injection at reservoir condition
$V_{CO_2,prod}$	Volumetric CO_2 production at reservoir condition

1. Introduction

Limiting global warming to 1.5 °C necessitates a major decrease in greenhouse gas (GHG) emissions and is important in reaching net-zero globally emissions by around 2050 [1–3]. An all-time high of 37.7 gigatons of energy-related CO_2 ($GtCO_2$) emissions was reported in 2023, representing 85 % of total global CO_2 emissions [4]. At the same time, global energy consumption has grown by 1.7 % in 2023 and is expected to continue to grow apace [4], meaning that we must meet a growing demand for secure and reliable energy while also mitigating GHG emissions. Carbon capture, utilization, and storage (CCUS) is a set of emissions reduction technologies, the use of which has rapidly grown in the past decade and will need to continue to grow to overcome the dual challenges for the energy sector [3,4]. Successful deployment of planned CCUS projects could total the annual CO_2 capture capacity to 435 Mt and storage capacity to 615 Mt by 2030 in future energy scenarios [2,3], highlighting the substantial potential of these technologies in mitigating global warming concerns, such as reducing the carbon footprint for the hydrogen supply chains [5]. In the oil and gas industry, CO_2-enhanced oil recovery (CO_2-EOR) is a commercial-scale technology that simultaneously addresses the dual challenges of improving hydrocarbon recovery, permanently sequestrating CO_2, and mitigating CO_2 emissions [6]. As of 2017, there were approximately 166 field applications of CO_2-EOR projects all over the world, of which 136 were in the U.S. and 9 in Canada [7]. Therefore, assessment of CO_2-EOR can benefit from a sound understanding of this process from both the oil recovery and the carbon sequestration perspectives, especially at the early stages of project investments.

Traditional reservoir performance forecasting methods for CO_2-EOR include analytical/semi-analytical (i.e., fractional flow equations and the Koval theory [8]), decline curve analysis (i.e., the Hubbert theory [9,10]), and reservoir modeling (i.e., multiphase and multicomponent compositional models) approaches [11,12]. In recent years, reduced-order models (ROMs) have been established and discussed based on empirical statistics and the development of artificial intelligence (AI), such as examples from Azzolina et al. [13], Middleton et al. [14], and Bachu [15]. As a subfield of AI, machine learning (ML) has gained significant attention from researchers and practitioners due to its ability to handle large and complex data sets and to predict the reservoir performance [16,17]. Applications of ROMs in CO_2-EOR are characterized to three perspectives: (1) to develop screening criteria to clarify suitable reservoirs for miscible CO_2-EOR [15,18–20], (2) to statistically understand the relationships between reservoir performance and operational conditions at scales [13,21,22], and (3) to establish the foundational predictions for the economic and environmental assessments at the earliest stage in decision-making [14,22–24]. In summary, the ROMs are deployed to provide quick and efficient estimation in a short timeframe with less field data accessible. These approaches were subsequently used to assess life-cycle CO_2 emissions of a CO_2-EOR process [24].

Evidence in the current petroleum industry suggests that ROMs are promising techniques being increasingly articulated to study the development of CO_2-EOR as well as CO_2 sequestration. As noted, the development of ROMs in the CO_2-EOR and CCS areas mainly lies in two different categories based on the statistical approach and ML techniques. The progression, however, is still in a preliminary stage due to a limited number of existing field datasets in the peer-review literature and the uncertain pedigree for those that do exist, owing to the confidentiality reasons.

In summary, there are three challenges to develop ROMs in CO_2-EOR. First, insufficient reliable data is available for statistical analysis or ML training purpose. Second, applying ROMs contains difficulties, as the stats-ROMs are not validated properly, and ML-ROMs are mostly validated only for case-specific studies. Thus, no generalized ROMs can be widely applied. Third, the role of stats- and ML-ROMs has not been well-understood in reservoir simulation nor the circumstances of their applicability in broader analysis, for example, in the field of techno-economic and life cycle analysis.

This paper seeks to answer two specific research questions in order to address these challenges. First, how does one create, validate and consistently compare the generalized ROMs based on different methods? Second, what are the capabilities of the ROMs given them an advantage over conventional approaches? This study, for the first time, creates generalized ROMs based on statistical and ML approaches to forecast the performance of CO_2-EOR associated with CCS. Synthetic and field data have been developed and collected such that a consistent comparison between different type ROMs can be outlined with proper validation. Additionally, by comparing the ROMs and conventional methods (e.g., compositional reservoir simulation), this study identified the conditions under which different predictive approaches can be applied for CO_2-EOR.

The rest of the paper is structured as follows: In Section 2, mathematical indicators and a framework for the ROMs and statistical and ML approaches are introduced. Section 3 discusses the performance of generalized stats- and ML-ROMs in CO_2-EOR individually and applies both ROMs to the Weyburn oil field. A comparison between two approaches as well as with numerical simulation are outlined. Implications and outlooks of the ROMs in CO_2-EOR are discussed accordingly. Section 4 summarizes this work. Supporting information (SI) is provided along with this paper.

2. Methods

2.1. Overview framework

As stated, the objective of this study is to develop the generalized predictive ROMs using the statistical method (stats-ROMs) and the machine learning algorithms (ML-ROMs) of CO_2-EOR and consistently compare between two approaches. Fig. 1 depicts the schematic framework of this research. As highlighted, this research comprised of five key procedures including.

1. Applying reservoir simulation techniques to create a theoretical base CO_2-EOR case using a commercial compositional reservoir simulator (CMG-GEM) from Computer Modeling Group Ltd. and the fifth comparative solution project (SPE5) by the Society of Petroleum Engineers and our previous work [25–34].
2. Synthesizing a large production dataset (3000 equal-probability simulation runs) by iteratively running the model with a variety of input distributions based on worldwide historical CO_2-EOR operations [35].
3. Processing the simulation results and collected field data based on approaches in Section 2.1.
4. Developing the generalized ROMs by either the statistical or ML approaches.
5. Validating the ROMs with the field production data from the Weyburn oil field.

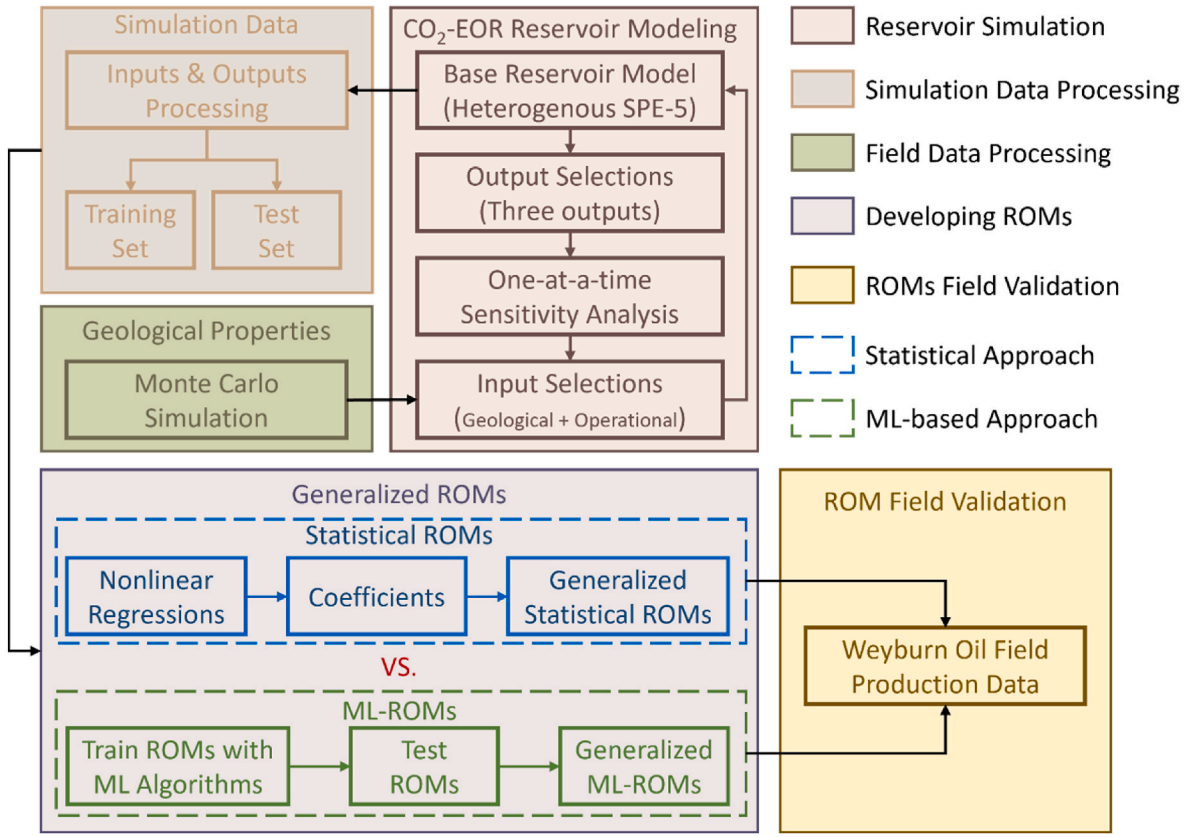

Fig. 1. The overview framework of this study. A compositional reservoir simulator is used to synthesizing the input datasets for the stats-ROMs and ML-ROMs. Geological properties of CO_2-EOR reservoirs are sourced from the world EOR survey published in 2014 [35]. Monte Carlo simulation is used in this study to create the heterogenous reservoirs (i.e., porosity and permeability profile, see SI section S1.1) and featured geological inputs for the reservoir simulations. A total of ~3000 CO_2-EOR were simulated according to existing statistics of CO_2-EOR field (see Fig. S3), which is used to create a wide range of inputs by Monte Carlo simulation (see Fig. S4). The monthly gas injection and oil production data for the Weyburn oil field [36] was processed for the purpose of validating stats-ROMs and ML-ROMs.

2.2. Computational indicators

i) The total injected CO_2 volume is expressed as a fraction (or multiple) of the hydrocarbon pore volume (HCPV), which is defined as the volume of injected CO_2 over the total HCPV at reservoir conditions, as denoted in Equation (1) [37,38].

$$HCPV_{CO_2,inj} = \frac{V_{CO_2,inj,rc}}{OOIP \times B_o} \tag{1}$$

Where, $HCPV_{CO_2,inj}$ is a dimensionless quantity that commonly varies from 0 to 3 at the field scale; $V_{CO_2,inj,rc}$ is the total volumetric amount of injected CO_2 at the reservoir condition with a unit of [L^3]; $OOIP$ is the original oil in place in the dimension of [L^3]; B_o is the oil formation volume factor, a dimensionless variable, usually expressed in units of reservoir volume over standard volume [RB/STB] [38]. This equation is also valid to covert produced CO_2 to HCPV scale ($V_{CO_2,prod} \rightarrow HCPV_{CO_2,prod}$) by substituting the subscript and plug-in corresponding values.

ii) The incremental oil recovery factor represents the additional volume of oil produced during CO_2-EOR. We adopted the dimensionless variable expressed in percentage of OOIP as Equation (2) presented [38].

$$RF_{in} = \frac{P_{EOR} - E[P_{secondary}]}{OOIP} \tag{2}$$

Where, RF_{in} is a fraction between 0 and 1, inclusive; $E[P_{secondary}]$ represents the expected cumulative oil production from the previous stage (i.e., typically the waterflooding of the secondary recovery), P_{EOR} refers to the cumulative oil production during the CO_2-EOR. Both P_{EOR} and $E[P_{secondary}]$ are in the volumetric unit of [L^3].

iii) The gross and net CO_2 utilization factor ($UF_{CO_2,gross}$, $UF_{CO_2,net}$) described in Equations (3) and (4) [13], defines the total injected or net consumed CO_2 in mass per standard barrel (STB) of incremental oil recovered, respectively. The units of both factors are [kgCO_2/STB]. In addition, the net CO_2 utilization factor can be derived by multiplying the gross CO_2 utilization factor with the CO_2 retention factor, which is given in the later text.

$$UF_{CO_2,gross} = \frac{M_{CO_2,inj}}{RF_{in} \times OOIP} \tag{3}$$

$$UF_{CO_2,net} = \frac{M_{CO_2,inj} - M_{CO_2,prod}}{RF_{in} \times OOIP} = UF_{CO_2,gross} \times R_{CO_2,ret} \tag{4}$$

iv) The CO_2 retention factor ($R_{CO_2,ret}$), a dimensionless variable, is given by taking either the mass or volumetric ratio of total CO_2 retained in the reservoir over the total injected CO_2 (at consistent reservoir or surface conditions), as Equation (5) demonstrates [13].

$$R_{CO_2,ret} = \frac{M_{CO_2,inj} - M_{CO_2,prod}}{M_{CO_2,inj}} = \frac{V_{CO_2,inj} - V_{CO_2,prod}}{V_{CO_2,inj}} = HCPV_{CO_2,inj} - HCPV_{CO_2,prod} \tag{5}$$

v) The WAG ratio for WAG and SWAG operations can be calculated according to Equation (6) by taking the ratio of water over CO_2 volumetric injection rate at the reservoir condition [38]. For the CGI, since water is not injected into the reservoir, the WAG ratio is equal to zero.

$$WAG\ Ratio = \frac{V_{w,inj,rc}}{V_{CO_2,inj,rc}} \tag{6}$$

vi) Regarding the efficiency of water usage of the WAG and SWAG process, the water production ratio is given by Equation (7) [39], which is also a dimensionless variable and expressed as the percentage of total volumetric water injection at the surface condition.

$$R_{w,prod} = \frac{V_{w,prod}}{V_{w,inj}} \tag{7}$$

2.3. Basic reservoir modeling

As mentioned, the SPE5 model was selected as the foundational reservoir model to create the theoretical heterogenous reservoir model using the CMG-GEM [26–28]. The SPE5 model is a well-known comprehensive generalized test dataset for miscible flood simulation studies for its completeness and wide applicability [34]. Both homogeneous and heterogeneous SPE5 models are used in this study, as shown in Fig. 2. The base values of SPE5 reservoir properties are listed in Table S1. There are three layers with a total thickness of 30 m. The oil components were modified according to Kamali and Cinar, which contains 35 % of C_6 and 65 % of C_{10} [30]. The minimum miscible pressure (MMP) of CO_2 and oil was obtained from CMG-WinProp (a phase behavior simulator) via the Peng-Robinson EOS and is equal to 1750 psi (or 12 MPa). A single porosity model was utilized, and the CO_2 solubility in water was neglected. For the heterogeneous model, the porosity for each block was simulated using the Monte Carlo method with a normal distribution of an average value equal to 0.3 with a 30 % standard deviation (porosity in a range from 0.1 to 0.5 for each block), the same as the average constant porosity of the homogeneous reservoir. Since SPE5 is a theoretical study case, the porosity is assumed not spatially correlated. The permeability for each block was generated based on the Kozeny-Carman equation (see SI Equation S(1)), which describes a relationship between porosity and permeability [40].

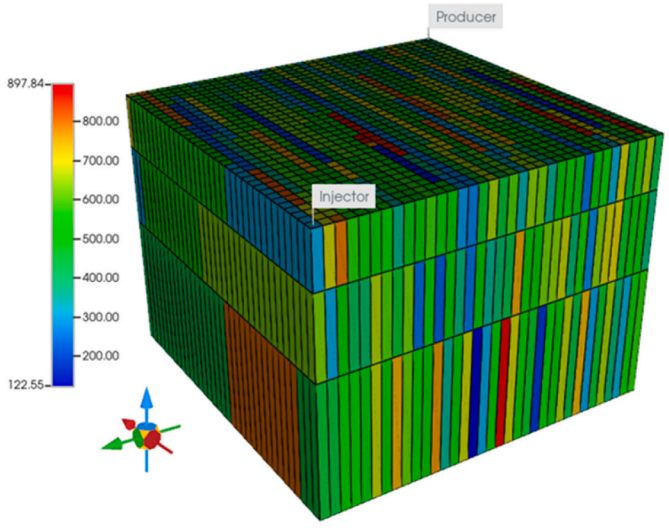

Fig. 2. Demonstration of heterogeneous reservoir creation by horizontal permeability profile of a theoretical heterogeneous reservoir model based on SPE5, where the vertical permeability is equal to 0.1 times the horizontal permeability. The relative permeability curves can be found in SI Figure S1.

There is one injector and one producer in the simulation with operational conditions following scenario three (designed for miscible gas flooding) in the original SPE5 model [25]. The primary production lasts one year due to the reservoir pressure depletion plus one year of secondary recovery of waterflooding. The oil production rate of the producer is at 12,000 STB/day with a minimum bottomhole pressure (BHP) of 1000 psi (7 MPa). A constant pressure production is assumed with the no-flow boundary condition. For the base case, the BHP for the injector is fixed at 4500 psi (31 MPa) with a water injection rate of 12,000 STB/day and a gas injection rate of 20,000 MCF/day (5.6×10^5 m^3/day). The basic homogeneous reservoir model was validated with the original solution of the SPE5 model and the results from the refined SPE5 model by Kamali and Cinar [25,30].

2.4. Statistical method (nonlinear regressions)

A nonlinear regression method was utilized in this research to explain the relationship between dependent and independent variables. A successful application of the nonlinear regression method generally contains two steps: (1) selecting a fundamental statistical model and (2) fitting modal parameters. The logistic growth regression model is described by Equation (8) below, which has been used to describe the statistical behaviors of CO_2-EOR with respect to the total injected fluid as the dependent variable [13,21]. Equation (9) is the Michaelis-Menten model that has also been investigated to describe the incremental oil behavior of CO_2-EOR in recent years [21]. The power regression described by Equation (10) is used to describe the behavior of $UF_{CO_2,net}$ and $HCPV_{CO_2,inj}$, which has not been investigated to the CO_2-EOR before this research [41].

$$y = A + \frac{(B-A)}{1+\exp\left[\frac{x-C}{D}\right]} \tag{8}$$

$$y = \frac{Ax}{(B+x)} + \text{constant} \tag{9}$$

$$y = A \times x^B \tag{10}$$

In Equations (8)–(10), $\boldsymbol{x}$ is the independent variable; $\boldsymbol{y}$ is the dependent variable; A, B, C, D are regression coefficients. In Equation (8), $\boldsymbol{x}$ refer to the total water and CO_2 gas injected in HCPV scale and $\boldsymbol{y}$ refer to the RF_{in} as discussed in later sections. Additionally, we also deploy Equation (10) to study the relationship between injected and produced CO_2, where $\boldsymbol{x}$ is the injected CO_2 and $\boldsymbol{y}$ is the produced CO_2 both in HCPV scale. In Equations (9) and (10), both $\boldsymbol{x}$ represent injected CO_2 in HCPV scale, where $\boldsymbol{y}$ denotes RF_{in} and $UF_{CO_2,net}$, respectively. The model accuracy is evaluated based on the Coefficient of determination (R^2, Equation (11)) and the mean squared error (MSE, Equation (12)) from the Minitab software [42]. The nonlinear regressions are carried out by Minitab software [42].

Selection of proper statistical models following the procedures of (1) comparing with the previous analytical/semi-analytical approaches, (2) observing the behavioral shapes and testing potential statistical models, and (3) applying relevant statistical models from existing literature. Based on this approach, we select the Michaelis-Menten model for the evaluation of EOR performance because it shares a similar format with the conventional semi-analytical fractional flow equation (see SI Section S1.5) by Koval, which explains the fluid displacement during CO_2 immiscible and miscible flooding [8,43].

2.5. Machine learning method

In this paper, we developed three ML-ROMs with different combinations of inputs and outputs based on the synthetic dataset generated

by numerical simulations. The inputs include six sensitive geological features and six operational conditions, while there are six outputs for each ROM. The detailed variables of input and outputs are documented in SI Table S2. Three binary variables were used to represent the operational schemes for WAG, SWAG and CGI, where 1 indicate the corresponding operation and 0 indicates not belonging to that category. For example, a CO_2-EOR with WAG equal to 1, stated that the CO_2-EOR is operating by WAG, thus, the value of SWAG and CGI will be equal to 0 automatically, because a project can only operate under one scheme. By creating three ML-ROMs, the results can be cross-validated with each other to assure prediction robustness, and the ROMs can be further applied to support the development of techno-economic and life cycle analysis. For instance, Model 1 applied the WAG ratio as the input, while Models 2 and 3 incorporated expected incremental oil recovery factor and net CO_2 utilization factor, respectively, to predict the WAG ratio of CO_2-EOR. Therefore, the WAG ratio predicted by Model 2 can be validated with the value predicted by Model 3 to ensure that the predictions are effective. Further, the prediction value (i.e., WAG ratio) can be used as an input to Model 1 to predict the incremental oil recovery factor so that it can match with the expectation value for Model 2 and 3 as the optional verification step. Another rationale is that a model can be selected to make predictions according to different input features from the reservoir properties as it is always challenging to collect all required input features for a ML-ROM. Regarding the selection of candidate algorithms, this study deployed supervised learning as the foundation tool to develop ML-ROMs. The input features and corresponding targets are well-labeled in this study. Typically, supervised learning can include seven steps: data collection, data preprocessing, model selection, feature selection, training model, model evaluation, and testing model. In the data preprocessing step, the synthetic dataset has been normalized to a standard range, and then, the training and testing set are split by a ratio of 8:2. A data normalization treatment is commonly used to avoid the underfitting or overfitting problems in the ML model developments [44]. The "scaling to a range" method has been used for normalization, which converts floating-point feature values from their natural range into a standard range [44]. Three algorithms are used in this study to develop each ML-ROM, including the extreme gradient boosting (XGBoost), support vector machine (SVM), and artificial neural network (ANN, we used the multilayer perceptron model specifically, herein used "ANN" as the abbreviation in this paper). The Bayesian optimization method has been used for hyper-parameter optimization for the three algorithms, following our previous work [45,46], where the optimal hyper-parameter combination is listed in Table S4. Detail descriptions of the three ML algorithms plus the Bayesian hyperparameters optimization are documented in SI Section S1.3.

2.6. ROM evaluation matrix

The consistent comparison across stats-ROMs and ML-ROMs necessitated a robust evaluation for each model regarding to the prediction performance. The coefficient of determination (R^2) is a statistical measurement of goodness of fit describing the fitness of the regression accuracy. In general, the range of R^2 varies from 0 to 1, where a higher value of R^2 of approximately one stands for better accuracy. The mathematical presentation is given in Equation (11). where y_i is the prediction of a sample (i.e., predictions from ROMs); $\widehat{y_i}$ is the corresponding actual value (i.e., calculated results from numerical model); $\overline{y_i}$ is the mean of the actual value of a target; n is the number of the samples [44].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{11}$$

Conventional measurements of the variation between predictive value and true value include the mean squared error (MSE, Equation (12)) and the root mean squared error (RMSE, Equation (13)). For stats-ROMs, the MSE are calculated by Minitab to verify the prediction accuracy and RMSE is calculated to compare with ML-ROMs. However, neither of the methods can normalize the multiple compared parameters to the same scale to make the comparison between all predictive outputs. Therefore, for the ML-ROMs, we adopted the rooted mean square percentage error (RMSPE, Equation (14)) and the mean absolute percentage error (MAPE, Equation (15)) to compare the model robustness across the three ML-ROMs, because the RMSPE and MAPE give the error in percentage so that it is easy to compare all predictive variables [44].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2 \tag{12}$$

$$RMSE = \sqrt{MSE} \tag{13}$$

$$RMSPE = \frac{100\%}{n}\sqrt{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2 / y_i} \tag{14}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}| / y_i \tag{15}$$

3. Results and discussion

3.1. Performance of stats-ROMs

In this section, the results of generalized stats-ROMs are presented according to the nonlinear regressions from both the oil recovery and CO_2 sequestration perspectives. We focus on RF_{in} to describe reservoir performance from the EOR perspective and $R_{CO_2,net}$ to characterize geological storage performance. In contrast, $UF_{CO_2,net}$ balances both perspectives, as it is effectively a ratio of oil production to CO_2 storage. As mentioned, the reservoir performance is normalized to the scale of CO_2 injection in *HCPV*. Three generalized stats-ROMs are shown by the solid curves in Fig. 3, and the 25th and 75th percentile estimates (P25/P75) are shown by the dashed lines. For each of the three stats-ROM, the coefficients from nonlinear regression results are listed in the basic equations on Fig. 3, such that the generalized stats-ROMs in this study can be further extend and applied to other relevant research.

The incremental oil recovery factor (RF_{in}). All simulation runs (~2156 runs) were calculated and compiled to a single data pair of $HCPV_{CO_2,inj}$ as the independent variable and RF_{in} as the dependent variable. The average fitting curve (generalized stats-ROM for RF_{in}) with the P25/P75 estimates are obtained as shown in Fig. 3(a) using the Michaelis-Menten model, which illustrate the empirical relationship between RF_{in} and total injected CO_2 during the CO_2-EOR operation. Although the log-logistic regression can be used to interpret this relationship according to Azzolina et al. [13], the Michaelis-Menten equation can further reduce the model complexity as it only contains two fitting coefficients instead of four, but with the same capability in forecasting this type curves [21]. In addition, the coefficients of the Michaelis-Menten equation have been endowed the physical meanings, where coefficient ***A*** represents the maximum RF_{in} at a given $HCPV_{CO_2,inj}$ and ***B*** is the time at which the RF_{in} is half of its max value at the given $HCPV_{CO_2,inj}$ [21]. The maximum RF_{in} is diminishing as more CO_2 is injected, owing to the declining residual oil saturation as the operation approaching late stage. It is reported CO_2-EOR can effectively recover 7–22 % OOIP for an $HCPV_{CO_2,inj}$ up to around 2.0 HCPV in field applications [15]. We obtained that the maximum RF_{in} can achieve to 12.1 % on average when injecting 2.0 HCPV CO_2 for conventional CO_2-EOR operations, whereas the P75 estimate can reach to 14.5 % OOIP on average. This finding is consistent with field observations, demonstrating this generalized stats-ROM can be applied as an empirical estimation tool. Moreover, in most CO_2-EOR operations, it is usually taking time for oil and displacement, therefore, a time gap is typically noticed after CO_2 injection and bounce on oil recovery, the oil rates declined at the end of waterflooding and increasing

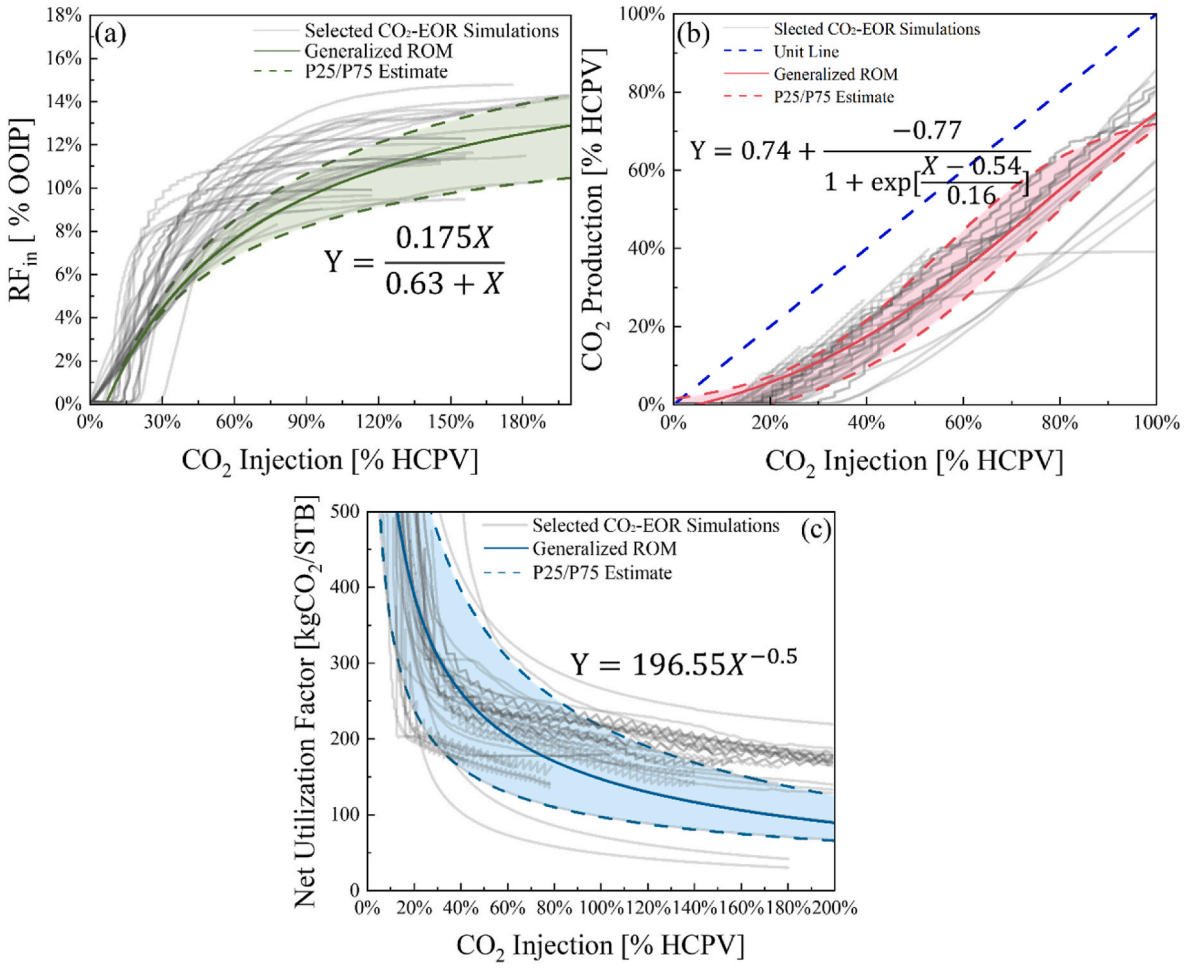

Fig. 3. Generalized stats-ROMs of CO_2-EOR according to synthetic datasets. The nonlinear regressions are deployed with respect to aggregated data of 2156 simulations. Only 100 randomly selected simulation runs are shown in this figure to maintain the high-resolution and visibility. For this figure, the ***x***-axis presents the total injected CO_2 in HCPV scale calculated by (Eq. (1)) across (a) to (c). The ***y***-axis represents (a) The incremental oil recovery (RF_{in}) in the unit of percentage calculated by (Eq. (2)), where the stats-ROM is obtained by applying the Michaelis-Menten regression(Eq. (9)); (b) The CO_2 production ($V_{CO_2,prod}$) in unit of percentage of HCPV ($HCPV_{CO_2,prod}$) calculated by (Eq. (1)), where the stats-ROM is established by logistic regression(Eq. (8)); (c) The net utilization factor ($UF_{CO_2,net}$) in the unit of $kgCO_2$/STB calculated by (Eq. (4)), where the stats-ROM is developed by power regression (Eq. (10)).

a few months after the CO_2 injected [47]. According to our observation, we noticed oil rate is increasing after 4.6 % HCPV of CO_2 is injected on average. For the cases with lower WAG ratios, typically less than 1, and completely oil depletion during the waterflooding stage, where the oil rate approach to zero by the end of waterflooding, the soaking time is usually longer and result in a greater accumulative incremental oil recovery factor. In contrast to previous studies which used the total injection (i.e. water + CO_2) as the independent variable, this work attempted to split the presence of water in the empirical analysis such that the in-depth understanding on effectiveness of CO_2 injection can be established accordingly. A limitation is observed for this generalized stats-ROM based on Michaelis-Meten equation according to Fig. 3(a), which is that the ROM is under-estimating the RF_{in}, especially during the gas breakthrough period. This is because the CO_2 breakthrough time is highly uncertain depending on the field geological variabilities, which cannot be interpreted by the generalized stats-ROM.

The CO_2 retention ($R_{CO_2,ret}$) factor. Based on the results from numerical simulation, CO_2 was produced after a gas breakthrough point. Depending on the impact of different injection schemes, the breakthrough points can be expected to be at from 10 % to 35 % HCPV of the total CO_2 injected and the average breakthrough point is approximate to 23 % HCPV after CO_2 is injected. The result in Fig. 3(b) indicates that the relationship of cumulative CO_2 production and cumulative CO_2 injection can be interpreted based on the logistic regression for all CO_2 secondary and EOR simulations. Hence, the $R_{CO_2,ret}$ can be estimated by Equation (5). The generalized stats-ROM illustrates that the maximum of 74 % injected CO_2 will be produced on average, which means that the $R_{CO_2,ret}$ is approximate to 26 % on average for the CO_2-EOR project.

The net CO_2 utilization ($UF_{CO_2,net}$) *factor*. In this part, the generalized stats-ROM is presented, whereas the $UF_{CO_2,gross}$ is discussed in SI section S1.4 and Fig. S8. As shown in Fig. 3(c), a power function was adopted to describe a correlation between the $UF_{CO_2,net}$ vs. the CO_2 injection in HCPV scale for the CO_2-EOR process. As a result, the $UF_{CO_2,net}$ declines from infinite to around 70 to 250 $kgCO_2$/STB for all simulation scenarios owing to a low increasing oil rate at the beginning stage of CO_2 injection. Then the $UF_{CO_2,net}$ rapidly declines to a breakthrough point, as in this period the CO_2 is not produced but the incremental oil recovery is significantly increased. After the breakthrough point, the $UF_{CO_2,net}$ smoothly declines to a constant value. As mentioned, the $UF_{CO_2,net}$ not only measures the efficiency of CO_2 utilization in terms of the marginal usage of CO_2 per barrel but also demonstrates the potential amount of CO_2 that can be stored in the residual oil zones in during the CO_2-EOR. There is no effective model to describe the gross utilization factor as evidenced in SI Figure S8 and section S1.4. As observed, producing one barrel of oil in this study has the capability to store 196 kg of CO_2 on average regardless of the variations of changing operational conditions.

The $UF_{CO_2,net}$ is significantly influenced by the WAG ratio, as increasing the water fraction inherently reduces the volume of CO_2 injected, thereby lowering the amount of CO_2 available for recycling. This relationship underscores the importance of optimizing the WAG ratio to balance EOR performance with $UF_{CO_2,net}$, warranting further investigation into the interplay between injection strategies and CO_2 recycling dynamics.

3.2. Performance of ML-ROMs

In this section, the performance of ML-ROMs is discussed by comparing the three models described in Section 2.5 with different combinations of inputs and outputs, which can be found in Table S2. As mentioned, after filtering the dataset, a total of 2156 and 1421 data points have been used for the ML-ROMs (where Model 1 used 2156 points, Model 2 and 3 used 1421 points). The data points have been split into training and test sets with the ratio of 8:2 and normalized to standard range before applying to ML algorithms. The difference in outputs between Model 2 and 3 is caused by the expectation of RF_{in,CO_2} or $UF_{CO_2,net}$ where Model 2 predicts the WAG ratio with the targeted RF_{in}, and Model 3 predicts the WAG ratio with the targeted $UF_{CO_2,net}$. According to Fig. 4(b) and (c), the WAG ratio can be accurately forecasted for Models 2 and 3 with R^2 less than 90 %, which is not as good as other five predicted outputs. It also shows that Model 3 can achieve better performance with the optimal ANN model to predict the WAG ratio, where R^2 can achieve 96 %. The R^2 results (Fig. 4) presented that the ANN is the best algorithm after the Baysian hyperparameter optimization. The R^2 of each output is about 95 %, where the greater covered area of the radar plot indicates a better performance of the prediction. Although the SVM also performed well in Model 2 and 3, it is limited by predicting one output for each run, resulting in a longer computational time, which is quantitively evidenced in SI section S1.7. Another observation is that all algorithms have limited capabilities in predicting the WAG ratio according to the results from Model 2 and 3. The reason is that the WAG ratios are a human-determined variable, which influences the reservoir performance substantially, and therefore, adding the uncertainty to the reservoir performance accordingly. The forecast performance of each output can be also analyzed by comparing the predictive value with the numerical simulation results as shown in Fig. 5, where each output has been compared. The closer the data point to the unit line means better forecasting performance. In short, by comparing the R^2 and results of evaluation matrix (documented in SI section S1.6), all selected algorithms can achieve high prediction accuracy with R^2 greater than 90 % after the Bayesian search of optimal hyperparameters. Among that, the optimal ANN model can predict all six outputs with R^2 greater than 96 %, which is the most accurate among all three algorithms in developing the ML-ROMs.

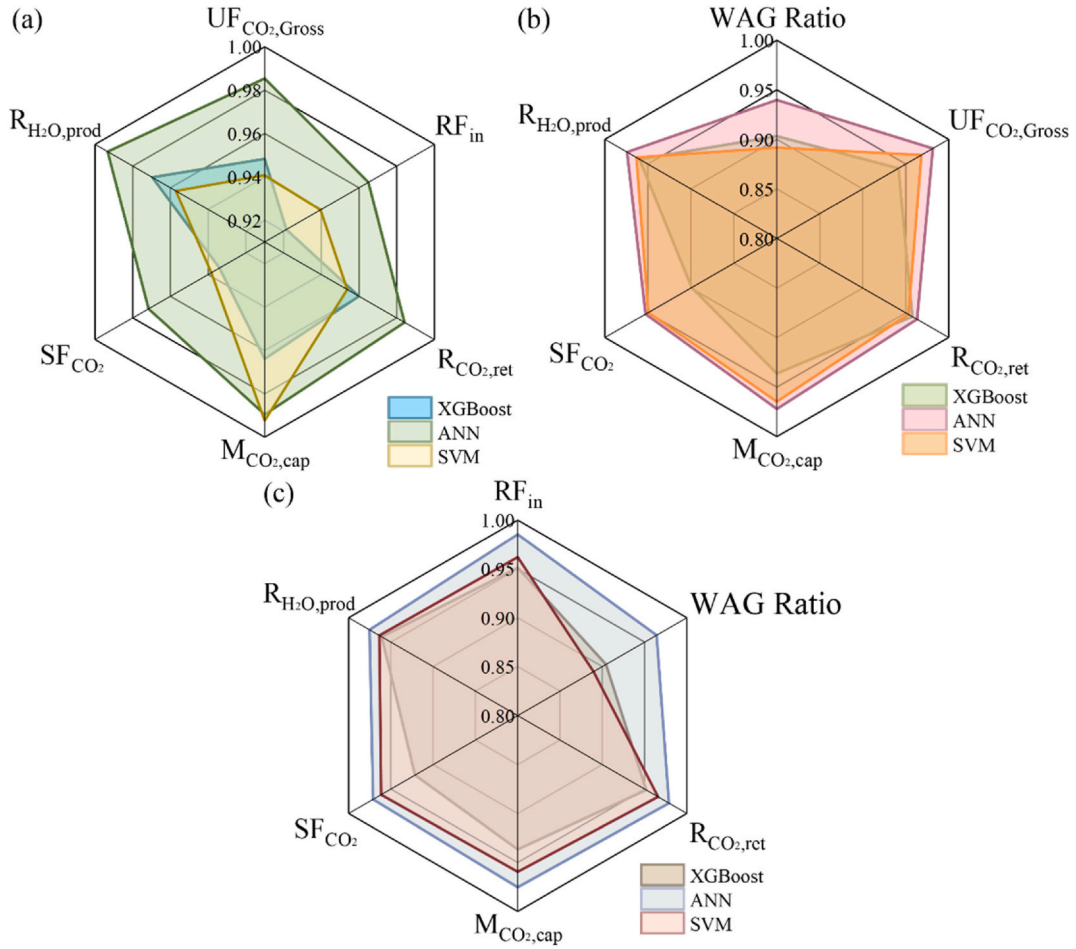

Fig. 4. Performance ML-ROMs by evaluating the R^2 of six outputs for each model. Data of this figure includes 432, 284, and 284 simulation runs (test sets) for (a) Model 1, (b) Model 2 and (c) Model 3, respectively. The three models comprised of different inputs and outputs, where the outputs (equation to calculate CO_2 storage capacity can be found in SI Section S1.5) are shown at each corner in (a) to (c) and detailed inputs can be found in SI Table S2. Axis of radar graph stands for the value of R^2 calculated by Eq. (11), where the maximum value is 1, indicating the best fitness. In brief, the greater covered area demonstrates the better performance of the selected ML algorithms.

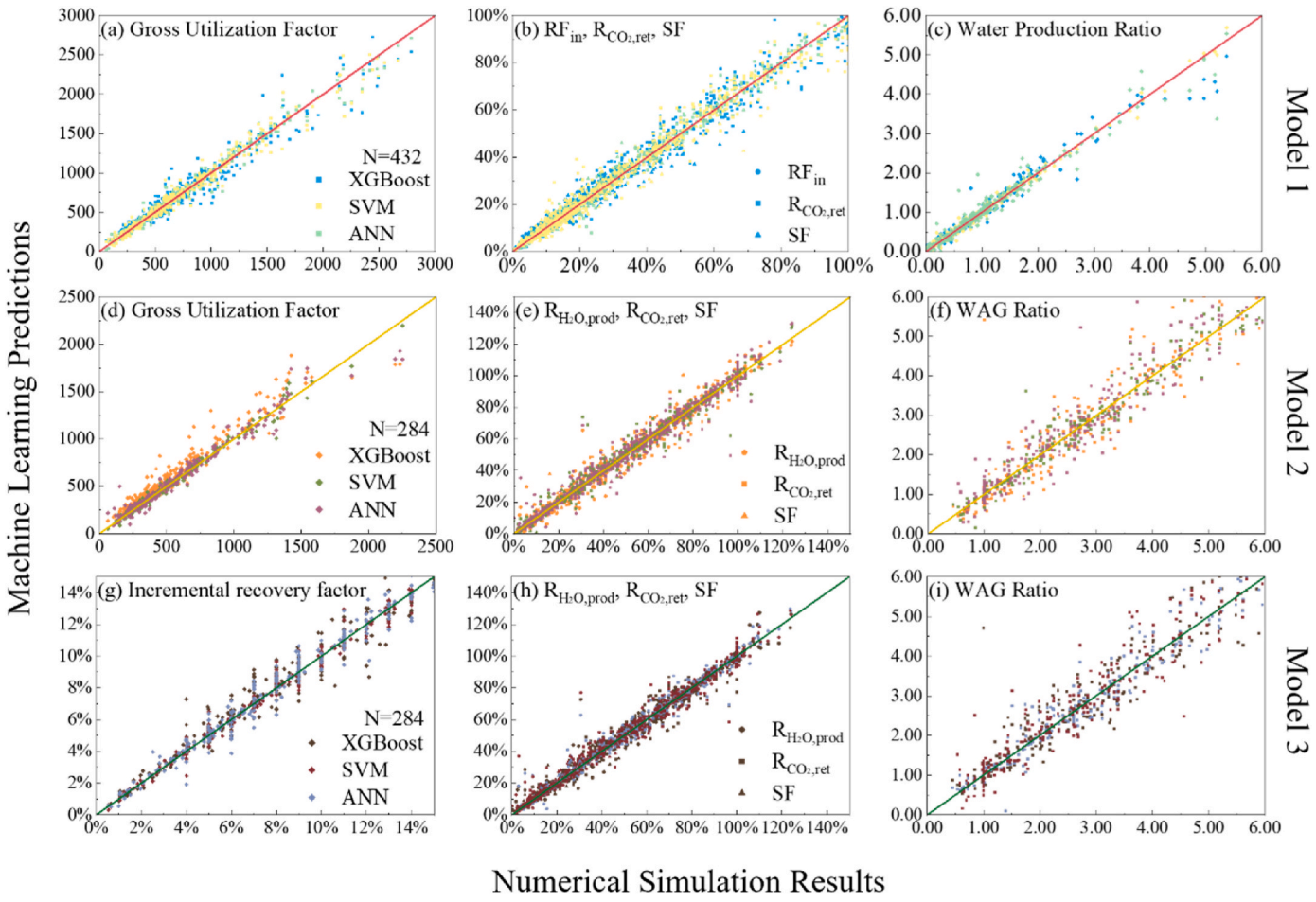

Fig. 5. Performance comparison (test sets) of ML-ROMs with the reservoir simulation by CMG-GEM. The ***x***-axis represents the values collected from the reservoir models and ***y***-axis represents the predictions by ML-ROMs. The closer to unit line in each figure indicates a better predictive performance. The color of data points stand for the ML algorithms and the shape represents the outputs for each data point. From top to bottom rows, the performance of three models is evaluated in each row. For those outputs in percentage, the integrated plots are presented by differentiating the shape of data points in (b), (e), and (h).

3.3. Application of ROMs to the Weyburn oil field

The Weyburn oil field is one of the largest medium gravity crude oilfields in Canada and has undergone CO_2-EOR since 2000, with most of the CO_2 sourced from a synfuels plant in North Dakota. As of May 2021, over 34 million tonnes of CO_2 have been stored (calculation of CO2 retention by Eq. (5)) to the reservoir with more than 155 million gross barrels of incremental crude oil has been recovered, which made the Weyburn the world's largest geological CO_2 storage project [48]. Table S6 listed the collected geological properties from previous studies of Weyburn oilfield that was required by the ML-ROM but not for the stats-ROM. Since the whole field contains multiple well patterns and operational schemes, it is extremely challenging to accomplish the reservoir simulation at the entire field scale. We assumed the operational strategy is WAG as most of the well patterns are operating under WAG. In this section, we validated and compared both stats- and ML-ROMs by deploying the monthly production data collected from SER. The corresponding results are presented and compared in Fig. 6 and Table S7.

Fig. 6(a) describes the result of RF_{in} of Weyburn with adopting the Michaelis-Menten regression. The reference cumulative recovery factor of the waterflooding was estimated based on the previous field level numerical simulation such that the RF_{in} can be calculated as the black solid curve shows [49]. As calculated, a total of 0.75 HCPV CO_2 has been injected to the reservoir. The real RF_{in} is calculated at 9.14 % (about 123 million STB), where the generalized stats-ROM based on Michaelis-Menten model is predicted the RF_{in} at 8.31 %. To make the predictions more accurate, we incorporated portions (e.g. a quarter or a half of the production data as shown in Fig. 6) of the production data to the generalized data and regenerated the Weyburn-specific stats-ROM. Fig. 6(a) clearly shows that the accuracy of the prediction can be improved with the penetration of the production data. For example, using a half of production data and the Weyburn-specific stats-ROM estimated the RF_{in} at 8.98 %, which is closer to the reality compared to the generalized stats-ROM. Although the accuracy of the generalized stats-ROM can be improved, half of the production data corresponds to 10 years of real time in the case of the Weyburn. In contrast, the ML-ROM forecast a RF_{in} of 9.40 % is an overestimate, and the incorporation of field data does not significantly affect the ML-ROMs in terms of improving the accuracy.

Regarding the $HCPV_{CO_2,prod}$, which is a necessary element to calculate the $R_{CO_2,ret}$, Fig. 6(b) demonstrates the predictive performance from generalized stats-ROM and ML-ROM, whereas the Weyburn-specific stats-ROM has been developed as above. We only applied half of the field data in this case because CO_2 breakthrough is at 18 % HCPV injection. Therefore, the $HCPV_{CO_2,prod}$ is 0 before the breakthrough, which adds noise to the stats-ROM. It is observed that 27.76 %, and 34.41 % injected CO_2 is stored to the reservoir from the generalized stats and ML-ROMs. By incorporating the 10-year field data, the $R_{CO_2,ret}$ is forecasted at 30.08 %, which is closer to the field estimation. The appearance of breakthrough adds complexity to the ROM development, which is not being to forecast by the ROMs in this investigation.

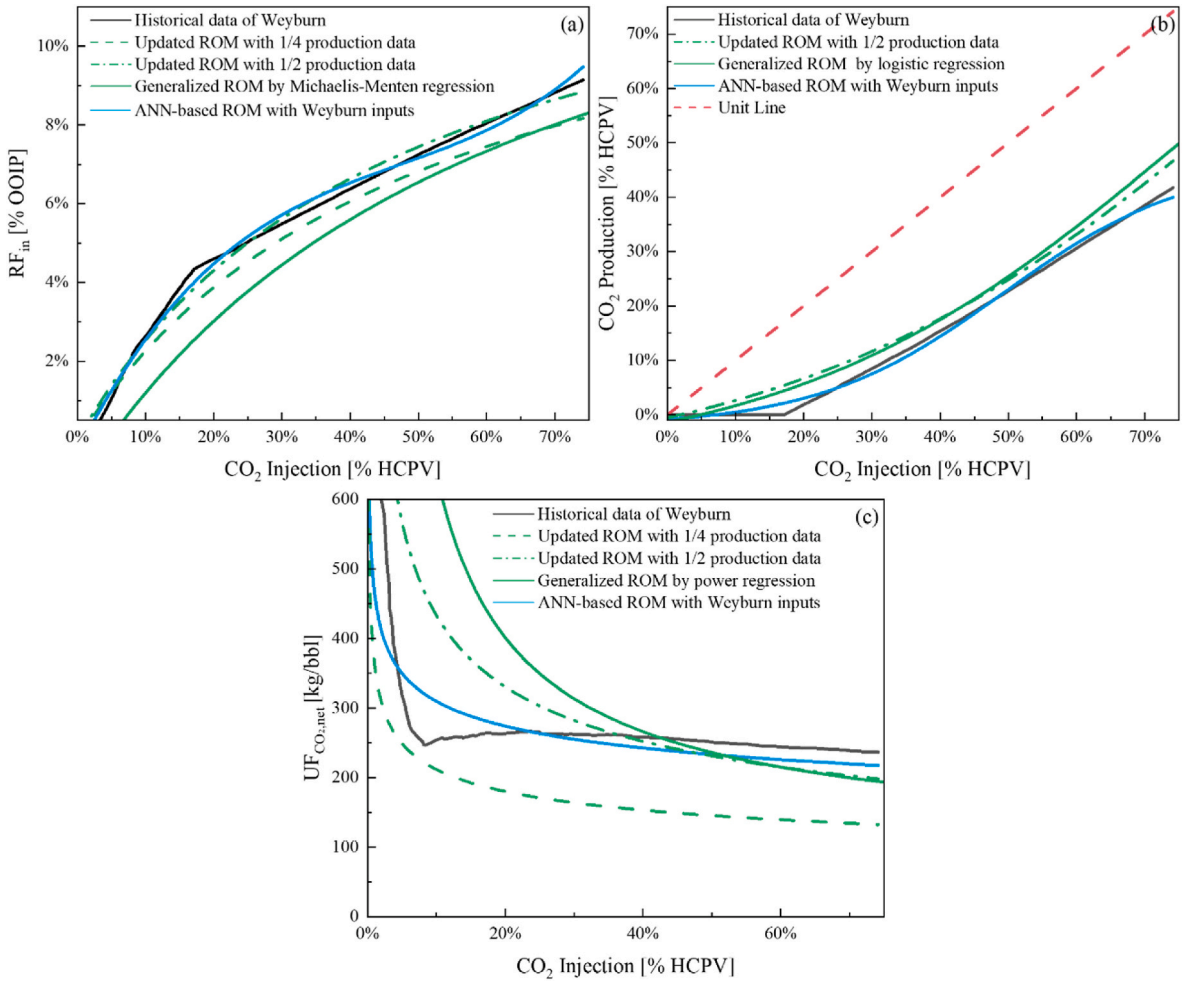

Fig. 6. Applications of stats-ROMs and ML-ROMs to the Weyburn oil field. The ***x***-axis represents the total injected CO_2 in HCPV scale and ***y***-axis represents (a) RF_{in}, (b) $HCPV_{CO_2,prod}$, and (c) $UF_{CO_2,net}$. The historical data was collected from SER and preprocessed according to equations in Section 2.1.

The calculated $UF_{CO_2,net}$ from the field data of the Weyburn is 237.6 $kgCO_2$/STB as shown in Fig. 6(c). The generalized stats- and ML-ROM predicts $UF_{CO_2,net}$ at 226.5 and 231.4 $kgCO_2$/STB. As illustrated by the field reality, the $UF_{CO_2,net}$ declines sharply before the CO_2 breakthrough as CO_2 is not being produced and approaching to a constant number after the breakthrough. However, inclusion of 1/4 production data to the stats-ROM is driving the prediction even worse because most of the data are before CO_2 breakthrough, where the $UF_{CO_2,net}$ declines drastically and leading a poor prediction. Updating the generalized stats-ROM with half of the production data also does not significantly improve the prediction performance as it is still strongly influenced by the CO_2 breakthrough. According to the observations, the ML-ROM forecasts the $UF_{CO_2,net}$ more accurately for the Weyburn oil field than the stats-ROM and both ROMs are affected by the CO_2 breakthrough, which can lead to a poor prediction. The $UF_{CO_2,net}$ can be used to benchmark the effectiveness of carbon storage associate with oil production during the CO2-EOR, which can be used to perform the economic and environmental assessment. In the case of the Weyburn, 237.6 kg of CO_2 were stored to recover one marginal barrel of oil theoretically during the CO_2-EOR operation.

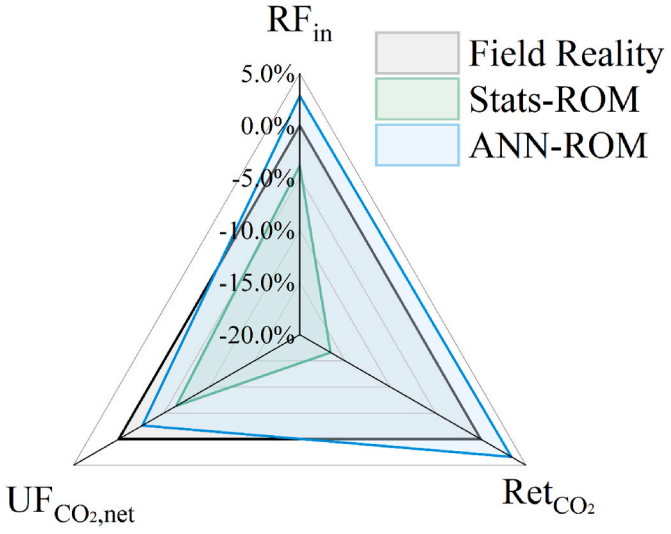

Fig. 7. Performance comparison between stats-ROMs and ANN-ROMs with the field data from the Weyburn oil filed. The axis of radar diagram represents the percent difference of three key indicators (RF_{in}; $UF_{CO_2,net}$; Ret_{CO_2}), where the value 0 stand for the same value of prediction and reality. The value greater and less than 0 stand for over- and under-estimated. A greater absolute value indicates a worse forecasting performance.

3.4. Comparison of methods

As shown, ROMs can reasonably predict reservoir performance in CO_2-EOR. In this section, we compare the strengths and weaknesses of ROMs with the reservoir simulation result as well as the comparing

between the Stats-ROM and ML-ROMs. Fig. 7 summarizes the application of stats- and ML-ROMs comparing with the reality of the Weyburn oil field. While both ROMs demonstrate a capability to predict CO_2-EOR performance, the stats-ROM tend to underestimate the outputs and the predictions are not as accurate as the ML-ROMs, especially in predicting Ret_{CO_2}. This is caused by the high uncertainty of the CO_2 breakthrough time and the geological variations across CO_2-EOR, which cannot be interpreted in the ROM training. Further, the stats-ROM in this study can only be used to forecast the reservoir performance in a more general way with respect to CO_2 injection, as it does not incorporate geological properties of the reservoir or engineering decisions (e.g., the WAG ratio). However, the ML-ROM demonstrated a powerful feature in two ways as it can interpret the sensitivity of geospatial uncertainties and infer the engineering requirements to achieve the certain expectations. In contrast, the stats-ROM can provide intuitive sense to describe the future performance of a CO_2-EOR at the early stage, where less properties are visible. This feature can lead stats-ROM a tool to benchmark the project evaluation at the early-stage investment. In summary, by comparing stats-with ML-ROMs, the stats-ROMs further reduced the complexity to forecast the performance of CO_2-EOR and endowed the ability to rectify the predictive accuracy along with the project development. The ML-ROM is sensitive to the geological uncertainties and able to predict more outputs than the stats-ROM. Moreover, the ML-ROM can be used to infer the operational decisions, but the accuracy cannot be further improved as adding more data does not affect the ML-ROM prediction. The stats-ROM is recommended to be employed at the very early stage, where less field descriptive data is accessible. Both the two types of ROMs can be used simultaneously during the project development as they can be used to benchmark with each other.

Comparing with traditional approaches, the ML-ROMs in this study takes 5 s to complete a single run and can speed up the prediction of CO_2-EOR by 2900 times compared to a reservoir model. Also, the input requirements are much less than the numerical tools, which only required 1 for the stats-ROM and 12 for ML-ROMs comparing to reservoir simulation with over 50 inputs. On the other hand, the outputs that we interested are as accurate as the results from the reservoir modeling. This improvement demonstrates the great potential on filtering CO_2-EOR in a timely manner. However, limitations are associated with the ROMs as we noticed from this work. First, temporal effects cannot be interpreted by the generalized ROMs. For example, we adopted CO_2 injection in HCPV in this work for all types of ROMs, but we could only use the time-series data as the independent variable if the injection rate is constant over the project lifespan, which is not normally the case in reality. Second, we found that the project scale really matters when adopting the ROMs. In other words, ROMs can have a better performance at a large scale, whereas the numerical simulation can have great performance from pore scale to field scale at the cost of computational time. In summary, numerical modeling can surely provide more accurate results than ROMs across scales, but ROMs can be time-efficient and accurate at the field scale. The ML-ROMs improve upon the stats-ROM by incorporating the field geological properties, but both approaches cannot interpret the impact of market variability on the operational decisions in time.

3.5. Discussion and implications

In this section, we discuss the ROMs involved in this study comparing with the existing literature and outlined the implications and outlooks of ROMs in CO_2-EOR associated with the geological carbon storage. As mentioned, Azzolina et al. [13] proposed the statistical models based on 31 existing CO_2-EOR industrial datasets and characterized the relationships between incremental oil recovery and total injection in HCPV scale, CO_2 retention and total injection in HCPV, and the utilization factor and total injection based on a statistical analysis. In 2018, Peck et al. [21] expanded this statistical model with an application to western Texas San Andres dolomite reservoirs. However, the statistical models were improved in this research from three aspects. First, owing to the dominating mechanism of CO_2-EOR, the CO_2 injection in HCPV scale is the only independent variable for both stats- and ML-ROMs in this study, so that field operational uncertainties were separated from the statistical model (e.g., impacts of changing WAG ratios and well patterns). The Michael-Meten model further reduced the complexity of log-log type curve in predicting the oil recovery performance of the CO_2-EOR. Second, by summarizing previous models, the unit of utilization factor was in the units L^3/STB in this work, and we adapted the units of $kgCO_2$/STB of the utilization factor (gross and net) were proposed in ROMs so that the marginal carbon storage associated with per barrel of oil recovery can be captured, which can be further applied to quantify the economic and environmental impacts directly. Third, the previous statistical models are mostly based on confidential industrial datasets, which makes them difficult to independently validate. Development of ML approaches are mostly site-specific and cannot easily be transferred to other similar cases. In this work, we closed the loop of developing the generalized ROMs based on both approaches including, datasets generation, model development and validation.

Although we find limitations, the generalized ROMs can still be widely applied to the early stages of CO_2-EOR associated with CCS into two categories. Firstly, the generalized ROMs can be used as a screening criterion to conduct the estimation for a potential CO_2-EOR from both oil recovery and CCS perspectives. For example, Bachu established a set of 14 criteria based on an Alberta reserves database, with previous statistical models, which estimated the cumulative incremental oil recovery and CO_2 storage [15]. The generalized ROMs in this work can further extend the previous work by improving the accuracy of screening criteria, especially across the geospatial regions at the large scale. Secondly, the generalized ROMs provide the foundation in conducting the economic and environmental assessments of a typical CO_2-EOR and other subsurface CO_2 utilization technological innovations at the early stages of the project evaluation, with less computational efforts than the numerical tools [50–54].

Future development of ROMs in this area lies in three directions from our viewpoints. First, physics-informed ROMs are necessary to further substituting the conventional numerical approaches. As mentioned, one common challenge for ROMs is that they fail to interpret the physical impacts (e.g., sweep efficiency, mobility) associated with the ROMs. Therefore, incorporating the physical information can improve the ROM performance such that the ROM can be widely applied across scales to fully substitute the conventional reservoir simulation. Secondly, applications of generalized ROMs to the site-specific cases need to be further refined in terms of quantifying the field and market uncertainties. Future development of generalized model needs to focus on establishing ways to adopt the generalized ROMs to the site-specific cases, such that the uncertainties can be quantified accordingly. Thirdly, researchers can build on these approaches to better understand the role for CO_2-EOR as an emission reduction option and the impacts of policy options.

4. Conclusion

In this work, we created synthetic datasets using conventional reservoir simulation and used them to develop both stats- and ML-ROMs to predict the reservoir performance of CO_2-EOR associated with geological carbon storage. Hence, the generalized ROMs are developed and compared based on the statistical and machine learning approaches. Three and six features are predicted by stats- and ML-ROMs, respectively, which are the critical elements to evaluate the performance of CO_2-EOR from both oil recovery and geological carbon storage perspectives. The generalized stats-ROMs in this work improved the previous approach by adding transparency of publicly accessible data, and ML-ROMs are generalized rather than only validated to the site-specific cases. The Weyburn monthly production data has been collected and validated the generalized ROMs. As a result, both approaches demonstrate an outstanding capability in predicting the reservoir performance

in a time-efficient and accurate manner. The stats-ROM further reduced the complexity of previous log-log type curves and eliminated the noise of water appearance such that the outputs can be directly related to the CO_2 utilization. We also proposed an approach to rectify the generalized stats-ROMs by incorporating the field production data as the case of the Weyburn oil field. The ML-ROMs improved the stats-ROMs by incorporating the sensitive geological reservoir properties and including more outputs. In comparison, the ROMs are more appropriate at the early stages of the CO_2-EOR development with less inputs at the field-scale than the numerical techniques, which perform better but require more inputs and longer computational time. However, the ML-ROMs fail to address the impact of economic uncertainties on operational decisions, where the stats-ROMs fail to incorporate both economic and field variabilities in the time horizon. Deployment of ROMs can provide technical support for economic and environmental assessments at the earliest stages of the project investment and be used to inform development of climate policies at the regional or national level.

CRediT authorship contribution statement

Haoming Ma: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sean T. McCoy:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Zhangxin Chen:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Data availability

Synthetic data is provided along with SI. Production data of the Weyburn oil field will be made available upon request. The code for developing the ML-ROMs was implemented using scikit-learn in Python and can be made available upon request. There are no known conflict interests.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research has been made possible by contributions from the Natural Sciences and Engineering Research Council (NSERC)/Energi Simulation Industrial Research Chair in Reservoir Simulation, the Alberta Innovates (iCore) Chair in Reservoir Modeling, and the Energi Simulation/Frank and Sarah Meyer Collaboration Centre. H. Ma would like to express the gratitude to the Alberta Graduate Excellence Scholarship, SPE Canadian Educational Foundation for the generous financial support during his Ph.D. study, and Drs. Joule Bergerson, Steven Bryant, Christopher Clarkson at the University of Calgary and Xiaoling Yin at the Colorado School of Mines for their valuable comments and support. S. McCoy was supported by the Canada First Research Excellence Fund.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.energy.2025.135313.

Data availability

I have shared my data at the attach files

References

[1] Intergovernmental Panel on Climate Change. Summary for policymakers. In: AR6 synthesis report climate Change; 2023. 2023, https://www.ipcc.ch/report/ar6/syr.
[2] International Energy Agency. Net Zero by 2021:2050. https://www.iea.org/reports/net-zero-by-2050.
[3] International Energy Agency. Energy technology perspectives, vol. 2024; 2024. https://www.iea.org/reports/energy-technology-perspectives-2024.
[4] International Energy Agency. World energy outlook 2024. https://www.iea.org/reports/world-energy-outlook-2024; 2024.
[5] Ma H, Sun Z, Xue Z, Zhang C, Chen Z. A systemic review of hydrogen supply chain in energy transition. Front Energy 2023;17:102–22. https://doi.org/10.1007/s11708-023-0861-0.
[6] Bui M, Adjiman CS, Bardow A, Anthony EJ, Boston A, Brown S, Fennell PS, Fuss S, Galindo A, Hackett LA, Hallett JP, Herzog HJ, Jackson G, Kemper J, Krevor S, Maitland GC, Matuszewski M, Metcalfe IS, Petit C, Puxty G, Reimer J, Reiner DM, Rubin ES, Scott SA, Shah N, Smit B, Trusler JPM, Webley P, Wilcox J, Mac Dowell N. Carbon capture and storage (CCS): the way forward. Energy Environ Sci 2018;11(5):1062–176. https://doi.org/10.1039/c7ee02342a.
[7] International Energy Agency. Number of EOR projects in operation globally, 1971-2017. https://www.iea.org/data-and-statistics/charts/number-of-eor-projects-in-operation-globally-1971-2017; 2022.
[8] Koval EJ. A method for predicting the performance of unstable miscible displacement in hetrogeneous media. SPE J 1963;3(6):145–54. https://doi.org/10.2118/450-PA.
[9] Hubbert MK. The theory of ground-water motion. J Geol 1940;48(1):785–944. https://doi.org/10.1029/TR021i002p00648-1. 8.
[10] Brandt AR. Testing Hubbert. Energy Policy 2007;35(5):3074–88. https://doi.org/10.1016/j.enpol.2006.11.004.
[11] Chen Z. Reservoir simulation mathematical techniques in oil recovery. Society for Industrial and Applied Mathematics; 2007.
[12] Balhoff M. An introduction to multiphase, multicomponent reservoir simulation. Elsevier 2022.
[13] Azzolina NA, Nakles DV, Gorecki CD, Peck WD, Ayash SC, Melzer LS, Chatterjee S. CO_2 storage associated with CO_2 enhanced oil recovery: a statistical analysis of historical operations. Int J Greenh Gas Control 2015;37:384–97. https://doi.org/10.1016/j.ijggc.2015.03.037.
[14] Middleton RS, Ogland-Hand JD, Chen B, Bielicki JM, Ellett KM, Harp DR, Kammer RM. Identifying geologic characteristics and operational decisions to meet global carbon sequestration goals. Energy Environ Sci 2020;13(12):5000–16. https://doi.org/10.1039/d0ee02488k.
[15] Bachu S. Identification of oil reservoirs suitable for CO_2 -EOR and CO_2 storage (CCUS) using reserves databases, with application to Alberta, Canada. Int J Greenh Gas Control 2016;44:152–65. https://doi.org/10.1016/j.ijggc.2015.11.013.
[16] Wang H, Chen S. Insights into the application of machine learning in reservoir engineering: current developments and future trends. Energies 2023;16(3):1392. https://doi.org/10.3390/en16031392.
[17] Ng CSW, Nait Amar M, Jahanbani Ghahfarokhi A, Imsland LS. A survey on the application of machine learning and metaheuristic algorithms for intelligent proxy modeling in reservoir simulation. Comput Chem Eng 2023;170:108107. https://doi.org/10.1016/j.compchemeng.2022.108107.
[18] Shaw J, Bachu S. Screening, evaluation, and ranking of oil reservoirs suitable for CO_2-flood EOR and carbon dioxide sequestration. J Can Petrol Technol 2002;41(9). https://doi.org/10.2118/02-09-05.
[19] Yin MF, Wei MZ, Bai BJ. Data analysis and screening guidance for field CO_2 flooding projects in the United States. In: 78th EAGE conference and exhibition. European Association of Geoscientists & Engineers; 2016. https://doi.org/10.3997/2214-4609.201601502.
[20] Kovscek AR. Screening criteria for CO_2 storage in oil reservoirs. Petrol Sci Technol 2002;20(7–8):841–66. https://doi.org/10.1081/LFT-120003717.
[21] Peck WD, Azzolina NA, Hamling JA, Gorecki CD, Ayash SC, Doll TE, Nakles DV, Melzer LS. How green is my oil? A detailed look at greenhouse gas accounting for CO_2-enhanced oil recovery (CO_2-EOR) sites. Int J Greenh Gas Control 2016;51:369–79. https://doi.org/10.1016/j.ijggc.2016.06.008.
[22] Middleton RS, Keating GN, Stauffer PH, Jordan AB, Viswanathan HS, Kang QJ, Carey JW, Mulkey ML, Sullivan EJ, Chu SP, Esposito R, Meckel TA. The cross-scale science of CO_2 capture and storage: from pore scale to regional scale. Energy Environ Sci 2012;5(6). https://doi.org/10.1039/c2ee03227a.
[23] Middleton RS, Yaw SP, Hoover BA, Ellett KM. SimCCS: an open-source tool for optimizing CO_2 capture, transport, and storage infrastructure. Environ Model Software 2020;124. https://doi.org/10.1016/j.envsoft.2019.104560.
[24] El-Houjeiri HM, Brandt AR, Duffy JE. Open-source LCA tool for estimating greenhouse gas emissions from crude oil production using field characteristics. Environ Science & Technology 2013;47(11):5998–6006. https://doi.org/10.1021/es304570m.
[25] Killough JE, Kossack CA. Fifth comparative solution project: evaluation of miscible flood simulators. In: SPE reservoir simulation conference san antonio; 1987. https://doi.org/10.2118/16000-MS. Texas 1–4 February.
[26] Computer Modelling Group. CMG-GEM [Computer software]. Calgary, AB, Canada. 2020.
[27] Xue Z, Zhang Y, Ma H, Lu Y, Zhang K, Wei Y, Yang S, Wang M, Chai M, Sun Z, Deng P, Chen Z. A combined neural network forecasting approach for CO_2-enhanced shale gas recovery. SPE J 2024;29(8):4459–770. https://doi.org/10.2118/219774-PA.
[28] Ma H, Yang Y, Zhang Y, Li Z, Zhang K, Xue Z, Zhan J, Chen Z. Optimized schemes of enhanced shale gas recovery by CO_2-N_2 mixtures associated with CO_2

sequestration. Energy Convers Manag 2022;268. https://doi.org/10.1016/j.enconman.2022.116062.
[29] Ma H, Chen Z, McCoy ST. Motivating CO_2-enhanced oil recovery to geologic carbon storage by integrating techno-economic and life-cycle assessments. In: 17th international conference on greenhouse gas control technologies, calgary, Canada; October, 2024. p. 20–4. https://doi.org/10.2139/ssrn.4271583 (GHGT-17).
[30] Kamali F, Hussain F. Field-scale simulation of CO_2 enhanced oil recovery and storage through SWAG injection using laboratory estimated relative permeabilities. J Petrol Sci Eng 2017;156:396–407. https://doi.org/10.1016/j.petrol.2017.06.019.
[31] Kamali F, Hussain F, Cinar Y. A laboratory and numerical-simulation study of co-optimizing CO_2 storage and CO_2 enhanced oil recovery. SPE J 2015;20(6):1227–37. https://doi.org/10.2118/171520-PA.
[32] Ma H, McCoy S, Chen Z. Economic and engineering co-optimization of CO_2 storage and enhanced oil recovery. In: 16th greenhous gas control technologies conference lyon, France; October 2022. p. 23–4. https://doi.org/10.2139/ssrn.4271583. GHGT-16).
[33] Kovscek AR, Cakici MD. Geologic storage of carbon dioxide and enhanced oil recovery. II. Cooptimization of storage and recovery. Energy Convers Manag 2005;46(11–12):1941–56. https://doi.org/10.1016/j.enconman.2004.09.009.
[34] Islam AW, Sepehrnoori K. A review on SPE's comparative solution projects (CSPs). Journal of Petroleum Science Research 2013;2(4). https://doi.org/10.14355/jpsr.2013.0204.04.
[35] Koottungal L. Worldwide EOR survey, data report. Koottungal, L., 2014. Worldwide EOR survey, data report. Oil Gas J 2014;112:79–91. https://www.ogj.com/drilling-production/production-operations/ior-eor/article/17210637/2014-worldwide-eor-survey.
[36] Government of Saskatchewan. Monthly production of weyburn unit CO_2 miscible flood project [-], Regina, SK, Canada. 2021.
[37] Lake LW, Johns R, Rossen B, Pope GA. Fundamentals of enhanced oil recovery. Society of Petroleum Engineers; 2014.
[38] Dake LP. Fundamentals of reservoir engineering. Elsevier 1983.
[39] Green DW, Willhite GP. Enhanced oil recovery. second ed. Society of Petroleum Engineers; 2018.
[40] Costa A. Permeability-porosity relationship: a reexamination of the Kozeny-Carman equation based on a fractal pore-space geometry assumption. Geophys Res Lett 2006;33(2). https://doi.org/10.1029/2005gl025134.
[41] Pyrcz MJ, Deutsch CV. Geostatistical reservoir modeling. USA: Oxford University Press; 2014.
[42] Minitab LLC. Minitab [computer software], state college, PA, United States. 2021.
[43] McCoy ST. The economics of CO_2, transport by pipeline and storage in saline aquifers and oil reservoirs. Sean T. McCoy; 2008.
[44] Belyadi H, Haghighat A. Machine learning guide for oil and gas using Python. Elsevier 2021.
[45] Xue Z, Zhang K, Zhang C, Ma H, Chen Z. Comparative data-driven enhanced geothermal systems forecasting models: a case study of Qiabuqia field in China. Energy 2023:280. https://doi.org/10.1016/j.energy.2023.128255.
[46] Xue Z, Yao S, Ma H, Zhang C, Zhang K, Chen Z. Thermo-economic optimization of an enhanced geothermal system (EGS) based on machine learning and differential evolution algorithms. Fuel 2023;340. https://doi.org/10.1016/j.fuel.2023.127569.
[47] Lake LW, Lotfollahi M, Bryant SL. Fifty years of field observations: lessons for CO_2 storage from CO_2 enhanced oil recovery. In: 14th greenhouse gas control technologies conference melbourne; October 2018. p. 21–6. https://doi.org/10.2139/ssrn.3366254 (GHGT-14).
[48] Uddin M, Coombe D, Ivory J. Quantifying physical properties of Weyburn oil via molecular dynamics simulation. Chem Eng J 2016;302:249–59. https://doi.org/10.1016/j.cej.2016.05.050.
[49] Whittaker S, Rostron B, Hawkes C, Gardner C, White D, Johnson J, Chalaturnyk R, Seeburger D. A decade of CO_2 injection into depleting oil fields: monitoring and research activities of the IEA GHG Weyburn-Midale CO_2 Monitoring and Storage Project. Energy Proc 2011;4:6069–76. https://doi.org/10.1016/j.egypro.2011.02.612.
[50] Ma H, Yang Y, Xue Z, Clarkson CR, Chen Z. Transitioning from emission source to sink: economic and environmental trade-offs for CO_2-enhanced coalbed methane recovery of Mannville coal Canada. Sci Total Environ 2025;966:178721. https://doi.org/10.1016/j.scitotenv.2025.178721.
[51] Cooney G, Jamieson M, Marriott J, Bergerson J, Brandt A, Skone TJ. Updating the U.S. life cycle GHG petroleum baseline to 2014 with projections to 2040 using open-source engineering-based models. Environmental Science & Technolology 2017;51(2):977–87. https://doi.org/10.1021/acs.est.6b02819.
[52] Yang Y, Liu S, Ma H. Impact of unrecovered shale gas reserve on methane emissions from abandoned shale gas wells. Sci Total Environ 2024;913:169750. https://doi.org/10.1016/j.scitotenv.2023.169750.
[53] Masnadi M, El-Houjeiri H, Schunack D, Li Y, Englander J, Badahdah A, Monfort J-C, Anderson J, Wallington T, Bergerson J, Gordon D, Koomey J, Przesmitzki S, Azevedo I, Bi X, Duffy J, Heath G, Keoleian G, McGlade C, Meehan DN, Yeh S, You F, Wang M, Brandt A. Global carbon intensity of crude oil production. Science (AAAS) 2018;361(6405):851–3. https://doi.org/10.1126/science.aar6859.
[54] Cooney G, Littlefield J, Marriott J, Skone TJ. Evaluating the climate benefits of CO_2-enhanced oil recovery using life cycle analysis. Environ Science & Technol 2015;49(12):7491–500. https://doi.org/10.1021/acs.est.5b00700.