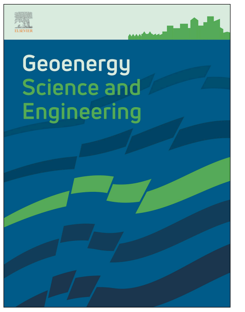
Journal Pre-proof



A Review of Intelligent Decision-Making Strategy for Geological CO_2 Storage: Insights from Reservoir Engineering

Jinjie Mao, Ashkan Jahanbani Ghahfarokhi

PII: S2949-8910(24)00321-X

DOI: https://doi.org/10.1016/j.geoen.2024.212951

Reference: GEOEN 212951

To appear in: *Geoenergy Science and Engineering*

Received Date: 26 January 2024

Revised Date: 11 May 2024

Accepted Date: 20 May 2024

Please cite this article as: Mao, J., Ghahfarokhi, A.J., A Review of Intelligent Decision-Making Strategy for Geological CO_2 Storage: Insights from Reservoir Engineering, *Geoenergy Science and Engineering*, https://doi.org/10.1016/j.geoen.2024.212951.

A Review of Intelligent Decision-Making Strategy for Geological CO_2 Storage: Insights from Reservoir Engineering

Jinjie Mao, Department of Geoscience and Petroleum, Norwegian University of Science and Technology, Trondheim, Norway.

Ashkan Jahanbani Ghahfarokhi, Department of Geoscience and Petroleum, Norwegian University of Science and Technology, Trondheim, Norway.

Highlights

1. State-of-the-art techniques are offered to address the technical barriers in Geological CO_2 Storage (GCS) from the perspective of reservoir engineering.
2. An intelligent modeling-optimization paradigm is proposed as a general solution for the decision-oriented GCS implementation.
3. Applications of the intelligent modeling-optimization paradigm is surveyed, showcasing the current deficiencies and future trends in GCS/GCSU (Geological CO_2 Storage and Utilization) practice.

Abstract

In a world characterized by a heavy reliance on fossil fuels, it becomes imperative to strike a harmonious balance between energy demands and carbon mitigation. This article delves into the practice of injecting carbon dioxide (CO_2) into subsurface formations as a potent strategy for mitigating climate change. It underscores the critical role of dynamic modeling in addressing the challenges related to CO_2 leakage throughout the life cycle of Geological CO_2 Storage (GCS) projects, spanning pre-operational, operational, and post-operational phases. Barriers to implementing GCS are discussed, including challenges in high-fidelity modeling, multi-scale simulation, and economic justifications. State-of-the-art techniques with regard to numerical simulation, Data-Driven Modeling (DDM), and multi-objective optimization are comprehensively reviewed. Moreover, an intelligent modeling-optimization paradigm using artificial intelligence and machine learning (AI&ML) is proposed to formulate the optimal development plan during field-scale GCS/GCSU (Geological CO_2 Storage and Utilization) projects. Successful case studies from the literature are surveyed, providing insights into the execution of the paradigm in real-world circumstances. Lastly, the paper concludes by outlining the existing challenges, emerging opportunities, and future directions for integrating intelligent modeling-optimization techniques into the decision-making processes of GCS/GCSU and conveying its potential application to the marching towards sustainable energy transition.

Keywords

Geological CO_2 Storage; Data-Driven Modeling (DDM); Intelligent Proxies; Multi-objective Optimization; Reservoir Management.

1 Introduction

The Paris Agreement stated that by the end of the 21st century, the global temperature rise should be restricted to a maximum of 2°C (preferably 1.5°C) above pre-industrial levels (IPCC Special Report, 2022). Based on the International Energy Agency (IEA, 2022), fossil fuels will remain the primary energy source for the next 50 years, with the amount of CO_2 released into the atmosphere from fossil fuel combustion reaching 3.63 billion ton by 2021. In alignment with international endeavors aimed at mitigating greenhouse gas (GHG) emissions, numerous countries and regions across the globe, encompassing the European Union, the United States, Canada, China, Australia, Japan and etc., have undertaken varied scales of Carbon Capture and Storage (CCS) experiments and demonstrations (Fragkos et al., 2021).

Geological CO_2 Storage (GCS) and its extension, Geological CO_2 Storage and Utilization (GCSU) are both crucial CCS technologies aiming to isolate CO_2 from the atmosphere. Ensuring secure storage without leakage risk is paramount, requiring precise evaluation and strategic design in large-scale commercial endeavors. To this end, the complicated subsurface system necessitates use of sophisticated numerical simulations to monitor dynamic CO_2 migration. However, the traditional numerical modeling tools involving multi-components, phases, time and space scales, and physical phenomenon are computationally intensive when thousands of simulations may be performed during the decision-making process. This computational challenge is addressed by AI&ML technologies. Data-driven models (DDMs) constructed via AI&ML methods demonstrate the capability to replicate outcomes generated by numerical simulations or real-field data, offering fast and accurate predictions to achieve efficient decision-making process. The promising performance of intelligent proxy models leads to a compelling prospect for practical decision support, efficiently handling optimization problems to achieve optimum results within reasonable calculation time and desirable accuracy levels.

Facing these multifaceted challenges and incorporating innovative techniques aligned with the cutting-edge trends, this paper embarks on a comprehensive exploration of the decision-making process inherent to CO_2 underground storage from the perspective of reservoir engineering. Zubarev (2009) provided a comprehensive review of typical proxy modeling applications in reservoir engineering, which is sensitivity analysis of uncertain variables, probabilistic forecasting and risk analysis, history matching and field development planning and production optimization. Ng et al. (2022) recently summarized a paradigm for ML-based proxy development, featuring the utilization of metaheuristic algorithms for training and optimization. They showcased several cutting-edge applications of this coupled ML-metaheuristic paradigm, constructing intelligent proxy models as exemplars. Meanwhile, Yao et al. (2023) directed their focus toward the application of ML methods in CCS, particular in applications ranging from physical properties prediction to success probability assessment of CCS project from a geoscience perspective. However, it becomes evident that a research gap exists in the systematic consolidation of the ML-based decision-making workflow and the classification of related applications into GCS and GCSU while addressing the challenges associated with CO_2 leakage risk assessment and prevention from a reservoir modeling perspective. This paper is oriented by CO_2 leakage mitigation, which covers a full spectrum of decision-making implementations supported by ML-based reservoir modeling. It probes into the upsides and downsides of new-generation intelligent CO_2 behavior modeling with respect to techno-economic considerations, while highlighting the vital nexus with pioneering optimization techniques.

This review follows a structured framework. It commences by presenting the background information behind the research in the Section 1. Section 2 elaborates on the important role of dynamic modeling in evaluating and mitigating the risks associated with CO_2 storage. The technical barriers concerning the dynamic modeling of CO_2 storage behavior within subsurface settings are surveyed in Section 3, while Section 4 delves into the ongoing research endeavors aimed at overcoming these barriers. Subsequently, Section 5 introduces a systematic approach tailored to address the multifaceted aspects of decision-making problem in CO_2 injection scenarios derived from the insights gleaned in Section 4, serving as a comprehensive solution to the issues posed in Section 3. To underscore the effectiveness of this paradigm, Section 6 presents an extensive case study that spotlights the state-of-art application associated with challenges encountered across various phases of both GCS and GCSU projects. Finally, Section 7 encapsulates the review with a summary of key findings and Section 8 provides a glimpse into future prospects.

2 The Role of Dynamic Modeling in Reducing CO2 Storage Risks

The primary objective of geological CO_2 storage is to store CO_2 securely and durably in geological formations. However, preventing leakage poses a formidable challenge, given its wide-ranging implications for climate change (Mikhaylov et al., 2020), groundwater resources (Xiao et al., 2016), ecosystems (Ko et al., 2016), and human health (Damen et al., 2006).

The GCS system consists of two main components: the wellbore system and the reservoir-caprock system (Fig.1). The wellbore system plays an important role in injecting CO_2 underground, serving as the exclusive conduit between the surface and subsurface. Yet, it also represents a prominent avenue for potential CO_2 escaping. The wellbore system is represented by a simplified model consisting of casing-cement-formation assemblage, with the addition of a cement plug to secure the integrity of the wellbore (Bachu & Bennion, 2009). Besides, the reservoir within the reservoir-caprock system functions as the repository for CO_2 storage, while the cap acts as a protective barrier, impeding the upward migration of CO_2. Notably, these components exhibit discernible disparities in terms of petrophysical characteristics.

Over time, the integrity of both the wellbore system and the reservoir-caprock system gradually weaken due to factors such as fluid flow dynamics, variations in temperature and pressure, mechanical changes, and chemical corrosion. Consequently, pathways emerge through which CO_2 can leak from the storage system. Within the wellbore system, critical pathways for potential leakage exist at the internal interface between casing and cement and external interface between cement and surrounding formation. Likewise, in the reservoir-caprock system, open faults and fractures serve as significant conduits for potential leakage of CO_2.

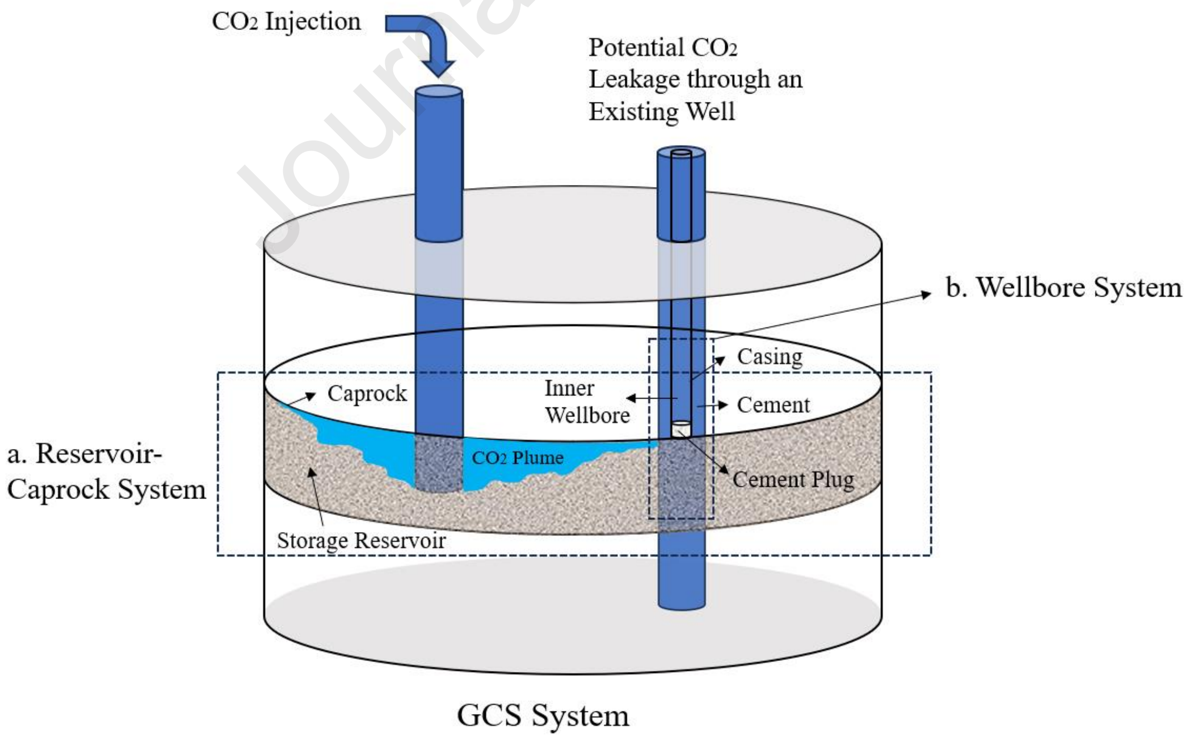


Fig.1. A portrait of GCS system consisting of a. the reservoir-caprock system and b. the wellbore system.

In order to ensure the safety of CO_2 storage, it is imperative to comprehensively understand the integrity variations within the storage system during downward injection, simultaneous upward and lateral migration, and the continuous and stable process of sequestration. During the injection phase, the escalating pore pressure will potentially compromise the integrity of the wellbore cement and result in the reactivation of faults and opening of fractures in the cap layer near the wellbore. Additionally, the acidization of the formation water may compromise the bonding strength of cement-rock assembly. The migration and sequestration phases also pose the risk of CO_2 leakage through injection wells, abandoned oil and gas wells, as well as pre-existing and undetected faults or fractures. Also, the changes in the stress field resulting from increased pore pressure may trigger seismic activity.

However, the main concern for policy makers and the public involves the substantial risk of leakage associated with CO_2 storage, raising questions about the exact possibility for injected CO_2 or expelled fluids to escape to the surface, water resources, or active petroleum reservoirs. Simultaneously, the operator, investors, and government agencies that take technical, financial, environmental, and societal risk are driven to maximize the CO_2 storage capacity while minimizing undesirable costs (Lie, 2016). To address these concerns, the only viable way to conduct assessments upfront is through dynamic model studies that aim to quantitatively investigate the likely outcomes of a storage operation. Through dynamic simulation, the complexity of CO_2 behavior and rock nature in reservoirs can be forecasted throughout the operation duration and beyond (Fig.2). The recent industrial storage projects and pilot experiments give valuable insights towards the importance of involving dynamic modeling into safe and long-term storage operations, encompassing either pre-operational, operational, or post-operational activities.

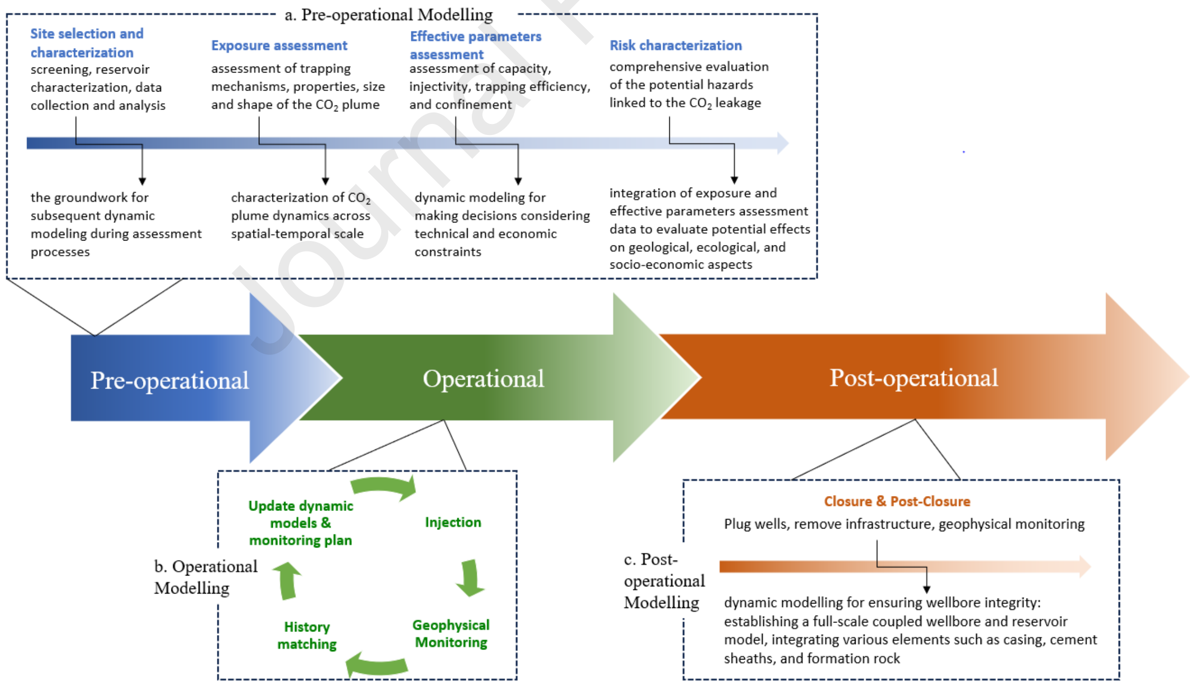


Fig.2 Development stages of the GCS lifecycle: (a) pre-operational stage; (b) operational stage; (3) post-operational stage.

2.1 Pre-operational Modeling

When CO_2 is initially injected into a subsurface accommodation space, the risk of leakage from the storage site is considerably elevated considering the inherent geological complexity and limited availability of data (Benson, 2007). Even in the case of depleted oil and gas reservoirs that have undergone extensive exploration and development over the years, their own set of obstacles is still present due to the geochemical interaction of CO_2 with overburden rocks and resulting petrophysical and geomechanical alterations as well as the presence of surrounding abandoned wells which have been in disrepair for a long time. Consequently, meticulous pre-injection preparation plays a crucial role within the GCS project life cycle, serving as an important measure to guarantee the safety and security of the storage site, particularly during the early stages of GCS implementation. We outline four essential stages involved in pre-injection preparation, namely site selection and characterization, exposure assessment, effective parameters assessment, and risk characterization. As illustrated in Fig.2 (a), after a detailed site investigation, dynamic modeling is used to characterize the CO_2 plumes migration during exposure assessment (Zapata et al., 2020), determine the efficiency of capacity, injectivity, trapping and confinement during effective parameters assessment (Rubin & De Coninck, 2005; Raza et al., 2016) and evaluate the potential hazards in geological, ecological and socio-economic aspects.

2.2. Operational Modeling

After the operation commencement, the continuous injection of CO_2 initiates the alteration of reservoir fluids composition, as well as the changes to reservoir rock properties. Ensuring the effective sweeping and sequestration of underground CO_2 faces substantial challenges. Monitoring the migration of CO_2 is essential throughout the CO_2 injection period and beyond. This entails timely comprehension of CO_2 propagation, fine-tuning of displacement strategies, and the detection and early warning of CO_2 leakage to ascertain escape pathways.

Once CO_2 is injected, active sensors need to be deployed to monitor the lateral movement of CO_2 and investigate changes in overlying formations. However, in field conditions, each CO_2 storage site is unique. CO_2 storage monitoring differs from fluid monitoring in oil and gas reservoirs, bringing difficulties to the conventional rock physics for seismic interpretation, specifically in establishing the relationship between rock properties and seismic response (Hao & Yang, 2012). Additionally, the time scale of storage ranges from decades to centuries. Therefore, accurate detection of leaks and subtle changes proves challenging due to measurement or interpretation errors as well as scarcity or poor resolution of data. Uncertainty often lies in the information derived from monitoring data, demanding the utilization of models to fill in the gaps (Harp et al., 2019).

Several large-scale GCS projects worldwide, such as the Sleipner project in Norway (Arts et al., 2018), the Otway project in Australia (Jenkins et al., 2012), Ketzin in Germany (Würdemann et al., 2010), and Frio in USA (Hovorka et al., 2006), have incorporated forward modeling with threshold approaches to design monitoring plans. This consideration is particularly crucial as certain faults or precursors to faults may remain undetected, thereby directly influencing the design of monitoring activities and potentially the feasibility of the storage site (Habert et al., 2016). As is shown in Fig.2 (b), we follow a cyclical process workflow to keep consistency between dynamic flow models and actual data. Mismatches between dynamic models and monitoring data may require model calibration, updated monitoring plans, or corrective measures.

2.3. Post-operational Modeling

After the completion of operations, the closure phase of geological storage begins, involving the abandonment of wellbores, the removal of surface facilities and geophysical monitoring (Fig.2 (c)). A proper well abandonment procedure is crucial for maintaining the long-term containment of injected CO_2. Wellbores have been identified as the most significant pathways for potential CO_2 leakage during storage, especially since they often penetrate the sealing caprock (Kutchko et al., 2007). The well completion and plugging methods should be carefully formulated to inhibit potential physical hazards associated with the well, the migration of contaminants between different formations and hydrological communication among initially isolated aquifer systems.

The petroleum industry gives primary attention to the integrity of wells throughout their lifecycle, from production till abandonment after depletion, lasting for several decades. However, when it comes to GCS, longer timeframes of hundreds to thousands of years after site abandonment need also to be included. The introduction of CO_2 with its distinct properties presents challenges to traditional well completion and plug techniques, as the geological and geochemical responses of wellbores during post-abandonment periods can have detrimental effects on their integrity in the long term. The main reason of wellbore failure are chemical corrosion and thermal and mechanical damage.

Therefore, it is essential to gain a comprehensive understanding of the long-term behavior of the chemical, thermal, and mechanical system consisting of steel casing, cement sheaths, and formation rock. The integrity of the casing-cement and cement-rock interfaces significantly impacts the performance of wellbore systems for CO_2 storage reservoirs.

This entails establishing a full-scale coupled wellbore and reservoir model. It is a challenging task due to the need to evaluate numerous parameters and uncertainties, as well as the limited understanding of the processes and factors governing wellbore integrity. Also, various elements such as casing, cement sheaths, and formation rock occur at different scales in terms of width and length, which may pose numerical challenges in discretization methods. The entire loading history, including drilling, completion, injection/production phases, and final abandonment, should be incorporated into the model during consecutive operations.

3 Identifying Barriers: Current Modeling Challenges in Sustainable Decision Making

In the context of GCS, sustainability refers to the safe and long-term sequestration of CO_2 by injecting it into saline aquifers, depleted natural gas and oil reservoirs, volcanic rock formations and underground caverns. The practice of GCSU focuses on geological storage while extracting commercial value by injecting CO_2 into oil and gas reservoirs, shale formations, unmineable coal seams and gas hydrates. When scrutinizing the decision-making process for sustainable GCS/GCSU projects, the myriad of literature overwhelmingly puts emphasis on environmental friendliness, intergenerational resource availability, and price affordability (Ramírez et al., 2008) as illustrated in Fig.3. Whereby, "environmental friendliness" refers to measures taken to curtail risks associated with health, safety, and the environment (HSE). "Intergenerational resource availability" means adopting a broad spatial and time scale spanning tens of miles and thousands of years to ensure accessibility of sufficient resources for future generations. "Price affordability" involves cost management of land storage, taking into account economically viable prices that align with present and future socio-economic conditions.

Fig.3. Ternary plot showing the distribution of the literatures with respect to three dimensions of environmental friendliness, intergenerational resource availability, and price affordability in sustainable GCS/GCSU projects.

Simulation models are indispensable in supporting and informing decision analysis. Through simulations conducted under various scenarios, the potential outcomes and risks of different strategies can be assessed, guiding decision-makers toward informed and prudent decisions. However, performing dynamic modeling of CO_2 flow in geological formations faces several challenges including barriers to build physically sound numerical simulations of CO_2 flow behavior, barriers to represent multi-temporal and spatial scales, barriers to seek balance between economic and sustainability dimensions.

3.1 Barriers to Develop High-fidelity Physics-based Models

GCS constitutes a highly intricate and multi-faceted process involving thermal, hydrodynamic, mechanical, and chemical (THMC) phenomena (Kolditz et al., 2012). CO_2 exhibits diverse forms, including supercritical (free CO_2), aqueous (CO_2 dissolved in water), complex ionic (formed through hydrolysis), oleic (CO_2 dissolved in oil), and solid (precipitated as carbonate minerals and hydrates). Consequently, quantifying the distribution of CO_2 in different phases subsequent to its injection into subsurface would need the consideration of the intricate physical and chemical processes governing these phase interactions.

During the initial stages, experimental methodologies act as the primary avenue for comprehending and predicting the intricate interactions among CO_2, crude oil, brine, and minerals constituents. Although extensive experimental investigations have been conducted to explore CO_2-brine interactions, a significant research gap persists in our understanding of the dynamic interplay between CO_2-(oil)-water and rock processes after CO_2 injection, with specific emphasis on the alterations in grain and pore geometry (Peter et al., 2022). Numerical simulation emerges as a powerful tool for quantitatively characterizing the intricate reaction processes and elucidating the distribution of CO_2 across diverse phases. Numerical simulation overcomes the limitations of experimental techniques, including temporal constraints and inherent uncertainties associated with natural analogies, particularly in the realm of CO_2-brine-rock interactions at both the field-scale and long-term timescales.

The numerical simulation of CO_2 storage relies on fundamental equations consisting of mass, momentum, and energy conservation. These equations are complemented by constitutive relationships, supplemented by additional physics-based equations that aim to capture geomechanical effects and geochemical reactions (Ajayi and Gupta, 2019). However, due to the limited availability of field and laboratory data and an incomplete understanding of underlying mechanisms, current numerical models tend to simplify certain physical and chemical equations. Especially when it comes to long-term and field-scale decision-making problems, these models encounter even greater trade-offs between the fidelity and the computational expenses. As a decision-making tool, the compromise in model fidelity can be summarized to four key assumptions: (1) temperature changes during CO_2 storage are often treated as negligible so that the energy equation can be omitted; (2) flow processes are approximated as two-dimensional or one-dimensional problems within homogeneous media; (3) geomechanical processes are assumed to exert minimal influence on rock properties, and their effects are approximated through the incorporation of rock compressibility coefficients; (4) geochemical reactions, which can induce significant carbonate precipitation and thereby alter porosity and permeability, are largely excluded from the model.

Upon the investigation of reservoir simulation implementations in GCS/GCSU, a comparison of the most renowned simulators has been drawn as illustrated in Fig.4. The open-source numerical simulators, such as the TOUGH suite of simulators (Jung et al., 2017; Ma et al., 2017; Xu et al., 2006), PFLOTRAN (Lu & Lichtner, 2007), MUFTE (Ebigbo et al., 2006), STOMP (Nguyen et al., 2016), MRST-CO_2lab (Nilsen et al., 2015), OPM (Sandve et al., 2018), GEOS (Fu et al., 2014), offer great flexibility and are readily available for scholars worldwide for extended development. However, they tend to be less convenient in pre- and post-processing. The commercial simulators are used extensively in the petroleum industry, such as CMG-GEM (Ranganathan et al., 2011), ECLIPSE (Class et al., 2009), t-Navigator (Hassani et al., 2024). They can quantitatively characterize CO_2 in gaseous, dissolved liquid, and dissolved oil phases, but often simplistically consider aqueous components equilibrium and water-rock dissolution and mineralization. Since each simulator has its own drawbacks, there is always a substantial discrepancy between actual field data and numerical simulation findings (Lu et al., 2011), and different numerical simulation tools can produce wildly divergent outcomes when applied to the same problem (Jiang, 2011).

Simulator	THMC				Wellbore Coupling Capability	Pre- and Post-Processing
	Thermal	Hydrodynamic	Mechanical	Chemical		
TOUGH3	☑	☑			☑	
TOUGH-REACT				☑	☑	
TOUGH-FLAC	☑		☑		☑	
PFLOTRAN	☑	☑		☑		☑
MUFTE	☑	☑				
STOMP	☑	☑				
MRST-Co2lab	☑	☑				☑
OPM	☑	☑			☑	☑
GEOS	☑	☑	☑		☑	
ECLIPSE	☑	☑				☑
CMG-GEM		☑		☑	☑	☑
t-Navigator	☑	☑	☑			☑

Fig.4. A comparison matrix illustrating the performance of simulators with respect to six aspects: each of the THMC processes, coupling of borehole flow, and pre- and post-processing. The performance of each simulator is qualitatively assessed from the literature and user guide, with the strong aspects highlighted with a tick symbol in the checkbox.

3.2 Barriers to the Multi Temporal-Spatial Scale Simulation

The barriers associated with physics-based modeling and numerical simulation in the context of CO_2 flow in the subsurface are closely intertwined. When it comes to simulating the physical aspects of GCS, we must deal with the complexities arising from multiphase, multicomponent and multiscale systems, which create the need of efficient numerical algorithms capable of solving a substantial number of governing equations. Also, it should be highlighted that both the primary variables (pressure, saturation, composition, enthalpy, etc.) and auxiliary variables (fluid and rock parameters) are not independent of each other. Therefore, a properly-designed framework is required during the iterative process to construct nested and sequential mathematical relationships among them.

Once the mathematical model for CO_2 flow in porous media is determined, the next step involves the transformation of the partial differential equations (PDEs) into a discrete form of nonlinear equations. The complexity of simulators is closely linked to the selection of an appropriate discretization scheme. In realistic storage projects, a huge number of spatial grids and demanding computational resources are typically required to achieve the desired numerical resolution. Furthermore, simulations must span significant time steps to capture physical phenomena over thousands of years. The updating of primary and auxiliary variables during each time iteration and grid incurs substantial computational effort.

Linearization methods are then applied to obtain a linear system of equations, which could be solved by iterative methods, least squares methods, conjugate gradient methods, etc. The fully implicit method (FIM) is the most extensively-used linearization method in practice due to its robust stability and relaxed time step constraints. FIM exhibits efficacy in addressing strongly nonlinear problems, making it particularly suitable for numerical simulations of GCS. Within the FIM framework, the computational time dedicated to solving the Jacobian linear algebraic equations often exceeds 80% of the total simulation time (Zhao et al., 2022). Consequently, the development of cost-saving techniques for solving large-scale, strongly coupled, and approximately singular Jacobian linear algebraic equations in a high level of accuracy remains an ongoing research pursuit.

Overall, as the demand for refined CO_2 sequestration simulations continues to surge, the new generation of simulators encounters a series of challenges as summarized below:

(1) The solvers applied to the large-scale, fine-resolution and highly ill-conditioned Jacobian linear algebraic equations still lack accuracy and robustness (developing algorithms aligned with a parallel computing environment could be the future trend).
(2) The mathematical models employed for reservoir simulations are becoming increasingly intricate, necessitating the inclusion of comprehensive coupling with THMC processes and wellbore.
(3) There is a growing need for high-resolution grid discretization, especially the adoption of localized and adaptive mesh refinement near wellbore (Jackson et al., 2015).

3.3 Barriers to Economic and Sustainable Justification

The economic feasibility is an important aspect in evaluating the sustainability of potential geological sites for CO_2 storage. During the storage period, the capital costs include geological exploration, CO_2 injection infrastructure, drilling (both new well construction and old well remediation) and other on-site requirements. The operational costs encompass the monitoring network, maintenance and human labor costs.

The cost of GCS depends on the storage plan, location, depth, reservoir characteristics, as well as the benefits and additional costs generated from associated by-products. Onshore storage costs are more influenced by geographical factors such as location, and topography. For offshore applications that require platforms or subsea infrastructure, the unit costs are typically higher. In this paper, facility costs in the phase of storage can be summarized into two categories: capacity-dependent and capacity-independent costs. As is shown in Fig.5, the former type of costs received the majority of attention while the latter type of costs was often overlooked in the literature. On the one hand, the economic viability of CO_2 storage is majorly influenced by CO_2 capacity. According to Rubin et al. (2015), the cost of storing one ton of CO_2 in depleted oil and gas fields approximately ranges from \$3 to \$10. As the scale of CO_2 to be stored expands, there will be a corresponding escalation in expenses related to the activities associated with the operation, monitoring, and maintenance of the storage system. On the other hand, pre-FID (final investment decision) modeling and logging costs as well as injection tests, are capacity-independent factors, with emphasis on specific considerations related to saline aquifer storage site (ZEP, 2011). Additionally, the capacity-independent cost increments associated with well completion, including the use of corrosion-resistant tubing, casing, and cementing materials (EPA, 2008), should be taken into account.

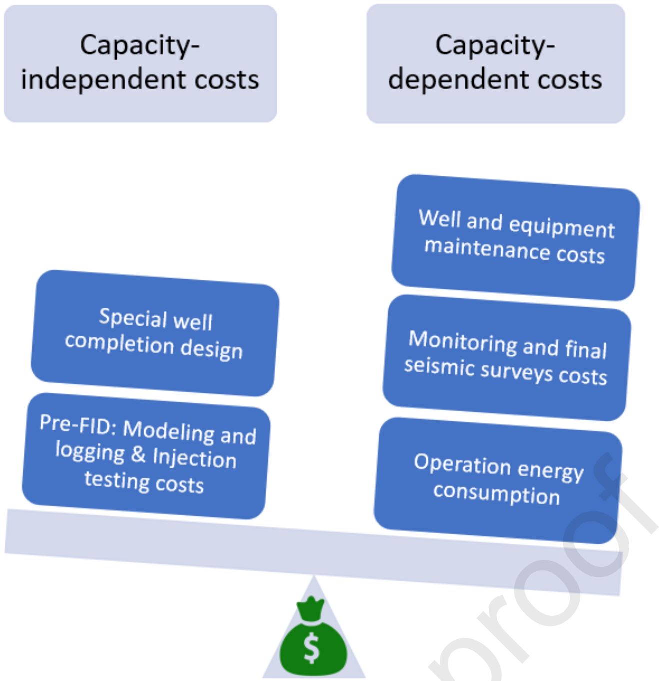


Fig.5. The literature survey with respect to the costs term in GCS objective function.

The economic incentives for the sole CO_2 sequestration technologies may be somewhat limited. Economic feasibility can only be quantified based on the existing impact of carbon taxes, meaning that the cost of carbon sequestration must be lower than the imposed carbon tax. Therefore, most GCS designs mainly focus on storage capacity along with the key performance index (KPI) quantification of storage effectiveness, without thorough economic analysis. It can be observed that in defining the evaluation criteria for storage performance, the emphasis is often on a single objective function, such as the final storage capacity within a certain time frame or the ratio of immobile CO_2 to mobile CO_2 (Pham et al., 2013). Within these current practices, the emphasis with regard to economics is often placed on capacity-dependent facility costs, while the investigation of facility costs that are irrelevant to capacity remains largely unexplored.

In the field of GCSU, which are represented by various applications such as CO_2-enhanced oil recovery (CO_2-EOR), CO_2-enhanced coal bed methane (CO_2-ECBM), CO_2-enhanced shale gas recovery (CO_2-ESGR), as well as CO_2 utilization in combustible ice development, the economic aspect often takes precedence. However, the costs regarding storage, whether capacity-related or not, also present a blank gap. The profit-driven nature leads to a preference for setting the net present value (NPV) of oil and gas as the primary objective function. Only a few studies formulate the NPV as the integration of petroleum production and CO_2 storage, with the balance leaning towards the maximization of profits. In recent years, it is worth noting there is a small subset of GCSU projects that adopt multiple objective functions to strike a balance between two key aspects: optimizing flood performance to enhance the profitability of the EOR project and ensuring efficient storage for the long-term and safe reduction of CO_2 emissions (Balch & McPherson, 2016). In this context, the GCSU process can be quantified through three responses of incremental recovery factor (oil or gas that is produced from CO_2 flooding), net CO_2 utilization (the amount of purchased CO_2 used to recover a cubic meter of hydrocarbon resources), and CO_2 retention factors (ratio of CO_2 retained in the reservoir to total injected CO_2) (Melzer, 2012; Mao & Jahanbani Ghahfarokhi, 2023).

4 Addressing the Barriers: Techniques Used in Previous Work

The state-of-the-art research work endeavors to overcome barriers that impede the attainment of an optimal development strategy. These barriers include the deficiency in implementing the proper physics, unaffordable computational demand when simulating high-resolution models and optimization challenges considering both economic and sustainability aspects. Overall development strategy is the basis of any successful business model in the world of tight competition and scarcity. In any stage of GCS/GCSU project, the goal of the managing/engineering teams is to develop the most accurate or optimal decisions that require real-time modeling tools including CO_2 plume prediction, monitoring, leakage and remediation. Unfortunately, the numerical models based on PDEs, are computationally intensive, even with today's supercomputers. As a result, binding all the underlying factors of the complicated physics in the large spatial and time scale is extremely hindered in devising an economically sustainable plan. All the aforementioned technical and economical (Section 3) constraints have led a great part of the petroleum community to investigate new alternatives which enable the same problems to be solved with considerable computation speed and precision.

This contributes to an overwhelming number of papers dedicated in avoiding the compromise of accuracy and reliability of results with the computational costs. In light of this, numerous attempts have been undertaken, which can be categorized with the three orientations: (1) the advancement of cost-effective numerical simulation algorithms; (2) the mitigation of computational bottlenecks by minimizing the number of required simulations by data analytics; (3) the adoption of multi-objective optimization techniques to facilitate more comprehensive decision support.

4.1 Numerical Simulation of GCS/GCSU

The multiphase-flow and transport problems encountered in GCS/GCSU projects involve a complex system of nonlinear equations that are tightly coupled and exhibit strong spatial and temporal dependencies. To ensure the attainment of stable solutions during practical simulation of GCS/GCSU, a number of modern reservoir simulators come out incorporating a wide range of gridding methods, numerical algorithms, and heterogeneous hardware systems.

The first approach incorporates the implementation of advanced gridding and upscaling methods in both the temporal and spatial dimensions. Syed et al. (2022) utilized a sophisticated 3D reservoir model within the CMG-GEM software, featuring a hydraulically fractured single horizontal well and employing local grid refinement (LGR). The objective of their study was to evaluate the impact of hydraulic fracture parameters and designs of Huff-n-Puff (HnP) operations on the performance of EOR and CO_2 trapping efficiency in tight oil reservoirs. Kamashev and Amanbek (2021) introduced LGR into the ECLIPSE 300 simulator with the CO_2 storage option, employing finer grids in areas where CO_2 plume migration may occur. They conducted a comprehensive sensitivity analysis using supervised ML algorithms and Monte Carlo sampling to investigate the feasibility of CO_2 storage. In a separate study, Suriano et al. (2022) employed an advanced unstructured Voronoi grid with refined resolution near the wellbore, enabling a sensitivity analysis of four different gridding schemes with varying progressive ratios. The objective was to investigate the impact of gridding resolution on the characterization of overpressure phenomena during CO_2 injection. The gridding resolution not only affects the understanding of well bottom-hole pressure profiles but also influences the estimation of the amount of CO_2 permanently trapped in the aquifer through residual and solubility trapping, especially in the few hundred years following injection. Zhang et al. (2021) reviewed the application of upscaling methods for fluid flow and mass transport, providing an exhaustive comparison of deterministic and stochastic upscaling methods. They also provided guidance on appropriate upscaling methods for transport and reactive processes, formation heterogeneity, and the desired level of coarsening in the model. These practices all share the common ground that significant computational expenses due to the fine-scale modeling would be incurred as an unignorable byproduct while scheduling long-term and field-scale problems Consequently, it becomes crucial to intelligently and adaptively "downscale" or "upscale" the accuracy level to strike a balance between computational expenses and desired outcomes.

The second approach involves the implementation of efficient numerical model solvers to handle the complex nature of the equations. The literature has witnessed the emergence of numerous strategies aimed at reducing the computational time needed for modeling by simplifying the formulation of mathematical model (including the reduction of dimensions, variables, and physical equations). Examples of these strategies include the utilization of black oil simulator rather than compositional simulator (Iogna et al., 2017), an invasion percolation (IP) simulator (Krishnamurthy et al., 2017), vertical equilibrium model (Arkai et al., 2021), and streamline simulation methods (Park et al., 2019). There are also various approaches by integrating advanced spatial discretization and linearization schemes proposed to reduce computational time while maintaining accuracy. For instance, Ahusborde et al. (2021) devised a fully coupled FIM finite volume (FV) technique to tackle the strongly-coupled nonlinear system of two-phase flows merging with geochemical reactions on a reservoir scale. The methodology adopted Newton's method addressing nonlinear algebraic equations, while an algebraic multigrid method was applied to solve linear equations in parallel. Even though this work has not been tested by trustworthy benchmarks with the same focus on bonding geochemistry and fluid flow in porous media, it shares high similarities with the sequential implicit approach that decouples the problem into two-phase flow and reactive transport. Voskov (2017) presented another novel approach known as Operator-Based Linearization (OBL) that largely simplifies the computation of fully-implicit method. This approach defined each discretized conservation equation component as the product of two operators: state-dependent that dynamically parameterized over physical space and space-dependent operators used in the traditional manner. The state-dependent operators were generated during simulations using multilinear interpolation of nonlinear parameters. By decoupling the computation of nonlinear physics from the conventional discretization terms, this methodology yielded remarkable advancements in the performance of Jacobian assembly (Khait and Voskov, 2017). Likewise, Li et al. (2021) developed a reservoir simulation framework amalgamated the mimetic finite difference (MFD) spatial discretization with the OBL scheme, which resulted in improved accuracy and efficiency in reservoir simulations. The OBL scheme provided adaptability and extensibility, while the MFD scheme facilitated the implementation of a multipoint scheme. The integration of these schemes was accomplished within a fully-implicit parallel framework utilizing high-performance computing through Message Passing Interface (MPI).

Lastly, parallel reservoir simulation has gained significant attention, highlighting the need for specialized algorithms for target parallel architectures. Cai et al. (2022) introduced GPSGLOW, which employs a distributed compressed sparse row format for storing the Jacobian matrix and right-hand side vector, enabling flexibility in utilizing third-party supercomputers with shared memory, GPU, or hybrid parallel computing. The implementation utilizes domain decomposition with ParMETIS for load balancing and employs OPENMP for multi-thread parallel computing simulations. Advanced communication schemes using MPI ensure thread safety and efficient computation which is implemented by using a large loop for the assembly of a Jacobian matrix and EOS computation. Gross and Mazuyer (2021) proposed GEOSX, an open-source multi-physics and level physics simulation tool designed for scalability on multiple CPUs and GPUs. GEOSX offers a suite of easily-extended physical solvers with the focus on multi-physics simulations involving geomechanics, flow, and transport mechanics. For different issues, such as mesh deformation or fluid propagation with pressure and saturation, GEOSX offers access to optional multilayer physical solvers. Zhou (2012) established AD-

GPRS that supports OpenMP parallelization on multicore platforms and a Nested Factorization linear solver for systems with multiple GPUs. Equipped with these capabilities, AD-GPRS can simulate challenging problems in a flexible and efficient way, which enabled the modeling of long-term behavior of CO_2 sequestration processes and fluid flow in reservoirs with complex geological features, such as fractures and faults.

4.2 Data-Driven Proxy Techniques

Within the realm of geo-energy engineering, proxy modeling, which is also known as surrogate modeling, holds the aim to capture the complex essence of fluid flow dynamics in porous media. These models can be constructed based on numerical simulation data, enabling precise replication of simulator responses within seconds (Zubarev, 2009). The establishment of accurate proxy models facilitates faster processing of decision-making problems in reservoir management, particularly when frequent updates to management plans are necessary. Over the past two decades, these models have found widespread application in various domains, encompassing sensitivity analysis of uncertain variables, probabilistic forecasting and risk analysis, history matching, and field development planning and production optimization (Gu et al., 2021). Ng et al. (2022) categorized proxy modeling methods into reduced-order modeling (ROM) and DDM which can be further classified into statistics-based and ML-based approaches. Among these, ML-based approaches can be divided into Smart Proxy Modeling (SPM) and Top-Down Modeling (TDM). The distinction of them lies in the data sources, with TDM utilizing field data or a combination of real field data and numerical simulation data, while SPM relies solely on data generated through numerical simulations. Bahrami et al. (2022) provided a comprehensive overview of proxy modeling classifications, considering different simplification principles, coupled optimization algorithms, and types of objective functions. In this paper, they derived a new categorization method, encapsulating Multi-Fidelity Modeling (MFM), ROM, traditional proxy modeling (TPM), and SPM under the umbrella of proxy models.

As famously quote by Sondergaard (2011), "Information is the oil of the 21st century, and analytics is the combustion engine." Likewise, Mohaghegh et al. (2011, 2020) firmly believed that the oil industry is ongoing a shift towards the fourth scientific paradigm, characterized by data-intensive science, and introduces the concept of subsurface data analytics (SDA). With the growing popularity of data analysis techniques empowered by AI&ML in the field of geo-energy engineering, DDM has emerged as the most prevalent approach. DDMs replicate the outcomes generated by numerical models or real-field data with an acceptable level of accuracy. These models can be conceptualized as advanced interpolation tables, allowing for rapid interpolation of nonlinear data ranges based on a few simulation runs, thus yielding quick approximate solutions as substitutes for high-fidelity numerical models (Bahrami et al., 2022). The key difference between DDM and traditional data-driven methods, such as decline curve analysis, lies in the inclusion of a Neuro-fuzzy system. The fuzziness of data arises from both the inherent randomness of the field data and the uncertainty surrounding the relationships among the data. When uncertainties are present, statistical methods and probability theory are employed to make deterministic observations. When dealing with complex GCS/GCSU systems, it becomes evident that most uncertainties stem from data scarcity, a lack of expertise, and imprecise representations. Fuzzy logic provides a modeling approach for uncertain systems, obviating the need for an excessive focus on inherent correlations between the data, whether they are static or dynamic. In this way, the emphasis is placed on analyzing the existing data itself without introducing new concepts, intermediate calculations, or excessive secondary data. Moreover, DDM outperforms traditional proxy modeling approaches, such as reduced-order and reduced-physics modeling, (Mohaghegh et al., 2015) which heavily rely upon the sophistication of the model, size of the design space, and quality of input data (Zubarev, 2019). By teaching reservoir engineering to a machine, or for now, to a computer program (due to the benefits of decoupling the equations, restrictions, and complexity of mathematical problems into numerical datasets, as well as providing more precision at the grid scale for estimating output parameters), they can produce proxy models at high-level accuracy without sacrificing the physics and order of the original system, and the spatial-temporal resolution of the model.

With the prevailing data sciences, a surge of studies addressing GCS/GCSU modeling have also emerged in recent years (Leach et al., 2011). AI&ML technologies possess the capacity to extract intricate nonlinear features embedded within data and establish complex mappings that link input parameters to designed outputs in the application. As is illustrated in Fig.6, the input parameters consist of operational variables typically associated with economic and safety considerations (e.g., injection well location, patterns, rates, modes, paths, supply limits, bottom-hole pressure), as well as geological uncertainties (e.g., the distribution of petrophysical parameters, fluid and rock properties, and parameters of physical mechanisms). The output of DDMs in these practices can be classified into two main categories: static modeling which captures spatial phenomena within a single timestep (e.g., physical properties and economic indicators at specific timestep), and dynamic modeling which addresses temporal phenomena within specific grids, wells, or models (e.g., history matching and production dynamics).

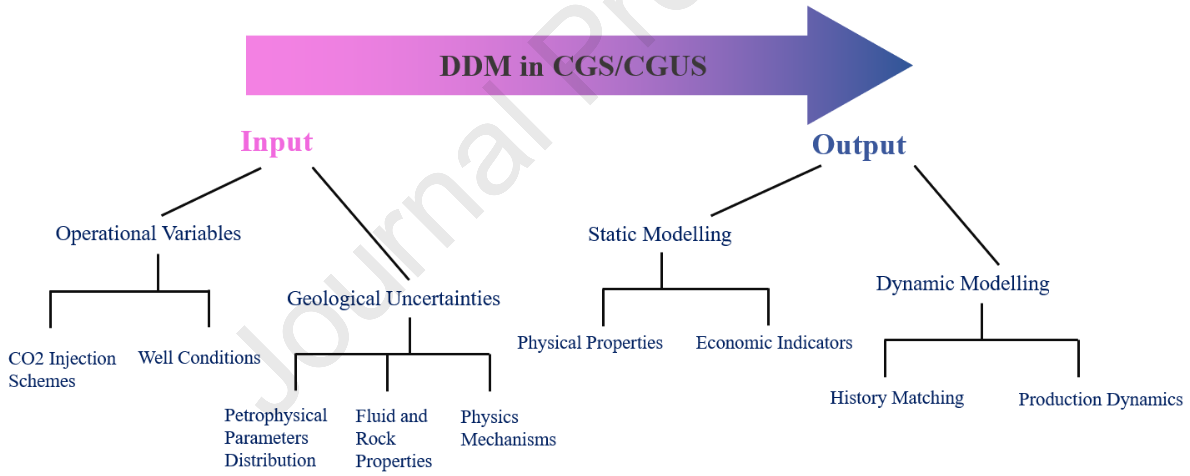


Fig.6. The hierarchical classification of input and output variables within DDM framework for GCS/GCSU.

Several examples of static proxy models that address both operational and geological uncertainties have been documented in the literature. For instance, Nwachukwu et al. (2018a) employed the Extreme Gradient Boosting (XGBoost) to develop a static proxy model for CO_2 flooding, with the well location as the input and the NPV as the output. In addition to conducting CMG simulations to generate reservoir simulation data, they incorporated well connectivity coefficients obtained through the Capacitance Resistance Model (CRM) and connectivity coefficients between points calculated by Fast Marching Method (FMM). To evaluate the influence of varied training dataset volumes on accuracy of proxy model, they created five synthetic cases with complexities ranging from low to high. In another study, Nait Amar et al. (2019) utilized Multilayer Perceptron (MLP) and Radial Basis Function Neural Networks (RBFNN) to construct proxies for predicting the solubility of CO_2 in brine. Notably, they employed the Levenberg-Marquardt algorithm to train the MLP, while Genetic

Algorithms (GA), Artificial Bee Colony (ABC), and Particle Swarm Optimization (PSO) were utilized to train the RBFNN. Among the range of models evaluated, the RBFNN-ABC model performed exceptionally well, surpassing the other models in terms of accuracy and predictive capability. Matthew et al. (2023) used Artificial Neural Network (ANN) to create a dynamic proxy model for CO_2-WAG (Water Alternating Gas) displacement. They accurately forecasted field CO_2 production profiles over a ten-year period, with an average error of less than 2% between the proxy model and numerical simulation data. Proxy models were subsequently employed to optimize the half-cycle length and gas/water injection rate using Non-dominated Sorting Genetic Algorithm II (NSGA-II), with the goal of maximizing both stored CO_2 and oil recovery. More applications utilizing data-driven proxies are further categorized and summarized in the Section 6.

4.3 Multi-Objective Optimization

The optimization of CO_2 injection strategies within GCS/GCSU, including well placement and control, well pattern design, and EOR measures, is of paramount importance in reservoir management, seeking the optimal efficiency of CO_2 storage and recovery rates of by-products while satisfying a variety of constraints. Reservoir management entails a complex endeavor that demands a thorough consideration of parameters such as permeability, porosity, and fluid viscosity to determine the optimal well placement and pattern design, injection-production regulations, EOR mechanisms, specifications of wells and associated facilities, as well as various economic factors. Consequently, this gives rise to a multi-dimensional optimization problem involving multiple objectives and parameters. Traditional optimization methods typically involve manually setting up multiple well deployment strategies based on reservoir geological characteristics, remaining oil saturation distribution, and production potential, followed by numerical simulations to compare their performance in terms of storage and displacement efficiency, liquid production, net present value, etc. However, such techniques rely substantially on reservoir engineers' experience, are prone to subjective biases, and fail to discover the optimum solutions in practice. In modern optimization methods, this complex undertaking is conceived as a mathematical problem seeking an optimal solution. The well location, injection rate, and other relevant parameters are treated as the variables, guided by objective functions and constraints. Through the integration of optimization algorithms and numerical reservoir simulators, iterative calculations are performed to identify the optimal deployment scheme. When compared to manual optimization, this approach dramatically reduces the computation costs and increases the possibility of obtaining optimal solutions. Despite this, numerical simulation-based optimization strategies remain faced with problems such as lengthy calculation times during an abundance of simulation runs. It is widely accepted that combining powerful proxy models with optimization methods is a superior choice to solve nonlinear and multidimensional issues (Onwunalu et al., 2008). With the advent of SDA technology, this integration has become viable and has emerged as a promising option for accelerating optimization procedures.

This advanced intelligent optimization technique coupled with DDM is the most important consideration of this paper. The unique characteristics of CO_2 sequestration in comparison with conventional petroleum industries bring about specific challenges for AI&ML-based optimization in the context of GCS/GCSU. These challenges can be summarized as follows: problem formulations need to involve multiple or numerous objectives to reflect the attitude of the management team towards giving priority to GCS or GCSU; the treatment of discrete and uncertain variables during the static modeling stage significantly impacts decision support outcomes; optimal control of multi-objective problem incorporates physical and economic constraints, which can be linear or nonlinear, dependent or independent with one another, contributing to complex constraints design domains; both gradient-based and gradient-free optimizers are prone to converge to local optimum, making the attainment of a global optimal solution uncertain; real-time CO_2 monitoring, calibration, and optimization generate new data that require self-adaptive automated modeling and optimization techniques. Nowadays, specialists and scholars all over the world are striving to discover solutions to these challenges, and the most forefront techniques corresponding to each of these challenges are reviewed below.

4.3.1 Dealing with the Favor of the Management Team to Conflicting Objectives

The initial step involves the formulation of mathematically generic mono- and multi-objective functions encompassing technical, operational, computational, and economic considerations, along with the inclusion of additional parameters for penalizing non-environmental scenarios (e.g., contaminated water production) and CO_2 emissions tax to prioritize storage scenarios. The field of multi-objective intelligent optimization can be classified into two primary methods: Pareto-based methods and non-Pareto-based methods. Non-Pareto-based methods aim to prioritize and convert multi-objective problems into single-objective problems through techniques such as weighted sum or prioritized ranking. On the other hand, Pareto-based methods explicitly generate a set of solutions satisfying different optimization priorities or weights, forming a range of optimal trade-offs for decision-making, known as Pareto front.

Pan et al. (2014) considered mass fraction of stored CO_2 by dissolution and residual entrapment and the maximum bottomhole pressure (BHP) of wells as the output and the CO_2 injection rate at four wells as the input. A Multi-Objective Evolutionary Algorithm (MOEA) was run on the ANN-based approximators to generate solutions of the multi-objective optimization's Pareto front. The resulting Pareto front obtained from the ANN-based solutions closely aligned with the Pareto front obtained from the ECLIPSE-based solutions. A multi-objective optimization problem was formulated that the optimal injector location was determined to maximize CO_2 storage while minimizing geomechanical risks, where NSGA-II is used to construct and evaluate a Pareto-front for the decision space (Zheng et al., 2021). You et al. (2019) proposed a robust computational framework that combines ANN and multi-objective particle swarm optimization (MOPSO) algorithm for co-optimizing oil recovery, CO_2 storage, and project NPV in a CO_2-WAG project. Additionally, the work was implemented in the Morrow-B formation, underscoring the advantages of Pareto front solutions over aggregate equation methods. The Pareto dominance-based criterion is ineffective for optimization problems with the number of objectives exceeding three, and the diversity estimator tends to favor Dominance Resistance Solutions (DRSs), hindering the diversity of population. To address this issue, a new Pareto-based algorithm was proposed by Liu et al. (2020), which included an interquartile range method to eliminate DRSs and a penalty mechanism for balancing convergence and diversity.

4.3.2 Dealing with the Geological Uncertainties

Uncertainties during optimization arise primarily from the highly heterogeneous subsurface conditions and the limited geological information, attributed to the uncertain, noisy and incomplete data of seismics, core samples, and borehole logs. Geological models established through interpolation with low-quality data result in significant uncertainty and inadequate representation of true reservoir conditions. Thus, optimization under uncertainty becomes a vital technique in mitigating development risks arising from geological model uncertainty. Three popular optimization methodologies which incorporate uncertainty into the mathematical models are Robust Optimization (RO), Certainty Equivalence (CE), or Stochastic Programming (SP). Since it is hard to determine the probability distribution of field-scale reservoir parameters, the prevailing optimization methods considering uncertainty are RO and CE optimization (Capolei et al., 2015) during the reservoir engineering practice. RO utilizes stochastic variables to describe uncertainty and employs probability theory for problem analysis, incorporating techniques such as Monte Carlo simulation, convolution, point estimation, and scenario-based analysis. Conversely, CE optimization transforms uncertain optimization into deterministic optimization by deriving a deterministic objective function based on the expected values of uncertain parameters. To illustrate with a simple example, when considering uncertain reservoir parameters like permeability fields, RO considers the entire ensemble of permeability fields, whereas CE optimization focuses solely on the average of permeability field.

To fix the issue of RO methods that might not fully capture the interrelationships among multiple criteria, Capolei et al. (2015) proposed the Mean-Variance (MV) method to describe the trade-off between expected NPV and risk in the objective function, achieving favorable optimization results. Their approach considered an ensemble of 100 permeability realizations, showing that RO is a special case within the MV framework. Ampomah et al. (2017) conducted research on a multi-objective optimization model to maximize oil recovery and CO_2 storage. A comprehensive uncertainty quantification model was constructed by Latin Hypercube sampling, Monte Carlo simulation, and sensitivity analysis of uncertain variables on the defined objectives. A risk aversion factor, assumed to follow a normal distribution, was used to compute a combined objective function, facilitating decision-making by providing results at different confidence levels. Nevertheless, the optimization procedure only included two geological realizations, with vertical permeability anisotropy (Kv/Kh) as the unknown parameter. An extended work was done by Nwachukwu et al. (2018b) in which XGBoost-based proxies were made to offer reservoir responses corresponding to well locations and control during WAG under geological uncertainty. They built a 20-model ensemble incorporating well block properties (porosities, permeabilities and initial saturations) as input variables during the proxy training.

4.3.3 Dealing with the Complex and Massive Constraints

In the practical optimization process of reservoir development, satisfying boundary constraints are crucial to ensure the practicality of the optimization solutions. Boundary constraints encompass constraints regarding input variables such as geological features, faults, well spacing, operation, regulation restrictions, process variables like pressure, temperature and fluid composition and output variables including the magnitude of gain or loss from individual profit or risk events. Evolutionary algorithms, known for their high search efficiency, robustness, and resistance to local optimum, are better suited for tackling complex optimization problems in the GCS/GCSU domain compared to classical optimization algorithms. However, it is important to note that evolutionary algorithms are essentially unconstrained optimization methods iteratively exploring the solution space, inspired by natural selection and genetic evolution, using a fitness measure to guide the search towards desirable solutions (Fonseca & Fleming, 1998). Given the numerous equality or inequality constraints in real-world engineering applications, which stem from adhering to processing facility capacities during the life-cycle storage optimization process, the choice of appropriate constraints handling methods significantly influences the performance of constrained optimization algorithms. Effective constraint handling methods should be capable of transforming constrained optimization problems into unconstrained ones while fully harnessing the searching advantages of evolutionary algorithms. Within the field of evolutionary computation, researchers have proposed diverse constraint handling methods and constraint optimization algorithms, which can be categorized as repair algorithms, penalty functions, decoder functions, feasibility-preserving representations and operators, and multi-objective optimization-based methods (Kramar, 2010).

Volkov and Bellout (2018) introduced a repair procedure in a 3D model with deviated wells, simplifying the satisfaction of geometric constraints. Salehian et al. (2021) utilized a similar repair procedure to bring infeasible solutions to the closest feasible solution through multilevel optimization frameworks, transferring close-to-optimum solutions from one level to the next. Zou and Durlofsky (2023) presented a comprehensive framework for optimizing monobore well locations and injection rates in GCS, incorporating appropriate linear and nonlinear constraints treated with PSO and Differential Evolution (DE) algorithms, along with preprocessing repair, penalty, and filter methods for handling various constraints. In another study by Nguyen et al. (2023), the performances of Augmented Lagrangian Method (ALM), Line-Search Sequential Quadratic Programming (LS-SQP), and trust-region SQP (TR-SQP) were compared for nonlinearly constrained RO problems with Stochastic Simplex Approximated Gradients (StoSAG), showing the superiority of SQP over ALM. Therefore, this article selected the LS-SQP workflow with a focus on GCSU by CO_2-EOR process, successfully solving both mono-objective and multi-objective optimization problems while minimizing constraint violations.

Obviously, the penalty function method stands as one of the prevailing techniques for constraints handling in the domain of GCS/GCSU applications. Nonetheless, the outcomes derived from this method typically offer only approximate satisfaction of the constraint conditions. Moreover, the penalty coefficients significantly influence the optimization's quality, and identifying appropriate values for this parameter can pose challenges. Another commonly used approach, the death penalty method, represents a simple yet stringent treatment that outright rejects solutions failing to meet the constraint conditions, potentially causing algorithmic stagnation. On top of that, a variety of ways to deal with restrictions have been investigated in the field, albeit only to a limited extent. Both Cihan et al. (2015) and Nait Amar et al. (2020) explored the integration of the 'Three feasibility rules' method introduced by Deb (2000), to effectively handle constraints in their optimization approaches. Cihan et al. (2015) focused on optimizing formation pressure during CO_2 sequestration by strategically determining optimal well placement and brine extraction rates using the DE algorithm. Meanwhile, Nait Amar et al. (2020) formulated the CO_2-WAG problem as a non-linear constrained optimization task and successfully incorporates the 'Three feasibility rules' method into the GA algorithm.

4.3.4 Dealing with the Local Optima

The optimization problem of well settings and operational decisions is a research hotspot in the GCS/GCSU development, involving high dimensionality, multiple local optima, and severe nonlinearity. Algorithms for solving this problem can be categorized into gradient-based optimization algorithms and gradient-free optimization algorithms, depending on whether gradient information of the objective function is required during the optimization process. However, optimization problems often exhibit discontinuous and nonlinear characteristics, making it challenging to obtain the gradients of the objective functions using analytical methods. To address this, adjoint methods, finite difference numerical computation, and stochastic approximation methods are commonly employed to obtain gradient information, but they each have their limitations of complex solving algorithms, model-scale restrained, slow convergence speed. In response to the challenges posed by gradient-based algorithms, gradient-free optimization methods have proven effective for solving well placement optimization problems without relying on gradient information from the objective function, instead utilizing evolution strategies inherent to the algorithms. The increasing penetration of AI&ML has demonstrated the robustness of metaheuristic algorithms in conventional reservoir development, with GA being widely applied in approximately 60% of well placement optimizations (Al Qahtani et al., 2012). The GA mainstream was also seen in GCS/GCSU optimization. Several novel gradient-free optimization algorithms, including Covariance Matrix Adaptation Evolution Strategy (CMA-ES) (Miyagi et al., 2018), Imperialist Competition Algorithm (ICA) (Karkevandi-Talkhooncheh, 2018), Spider Monkey Optimization (SMO) (Bansal et al., 2014), Thermal Exchange Optimization (TEO) (Kaveh & Dadras, 2017), and Atom Search Optimization (ASO) (Sun et al., 2021), have emerged over the past decade, and show great promise in addressing GCS/GCSU optimization challenges. However, gradient-free optimization techniques often confront issues in balancing global and local optimization objectives, with their performance impacted by algorithm control parameters, resulting in a complex process in determining optimal parameter values. Furthermore, they may be trapped into local optima if the positions of global best and local best coincide over a number of iterations (Niknam et al., 2009 & Freisleben & Merz, 1996).

To improve the precision and efficiency of solving the optimization problem during CO_2 injection, it is essential to utilize effective optimization algorithms. However, no individual algorithm can efficiently address all optimization problems. Researchers frequently blend the strengths of multiple optimization algorithms to create hybrid approaches, aiming to enhance the optimization capabilities. Three types of hybrid methods can be summarized from the literature: (1) amalgamating two optimization algorithms, (2) integrating additional mathematical techniques into basic optimization algorithms, and (3) incorporating evolutionary strategies, such as selection, mutation, and crossover strategies, into basic optimization algorithms. There are very few examples of hybrid optimization algorithm being applied to GCS/GCSU.

Nwankwor et al. (2013) introduced a hybrid method that integrated PSO perturbation into the overall framework of DE, aiming to maintain population diversity and significantly improve the performance of well placement optimization compared to using basic PSO and DE individually. Ding et al. (2014) developed modified PSO (MPSO) through adjustments to the inertia weight factor, velocity update strategy, and the incorporation of a flying time factor. The MPSO method combined with the Quality Map (QM) outperformed the standard PSO, MPSO, and center-progressive PSO in achieving optimal well placement. The application of a global optimizer (e.g., nature-inspired algorithms) as coarse approximation and a local optimizer as fine tuning (e.g., direct search method), provides flexibility. Aliyev (2011) proposed a novel hybrid approach by combining PSO with the Hooke-Jeeves Directed Search Algorithm (HJDS). Initially, the PSO algorithm was utilized to discover an initial solution, which was then employed as the starting point for HJDS, effectively leveraging the global optimization capability of PSO and the local optimization ability of HJDS to enhance the process of well placement optimization. In another study by Isebor et al. (2014), PSO was incorporated as a step in the Mesh Adaptive Direct Search (MADS) algorithm, resulting in the PSO-MADS hybrid algorithm. The research outcomes indicated that the PSO-MADS algorithm exceeded the efficiency of individual basic algorithms, PSO, and MADS, in addressing well placement optimization problems. In 2014, Humphries and Haynes developed the PSwarm algorithm, which combined PSO with the Generalized Pattern Search Algorithm (GPS). The optimization process involved an alternating strategy of PSO and GPS to enhance the algorithm's performance in solving well placement problems. Furthermore, a cutting-edge approximation algorithm called BPGrad was introduced (Zhang, 2018), specifically designed for locating global optimality in Deep Learning (DL) through Branch and Pruning (BP) techniques. This algorithm effectively reduces the gap between the lower and upper bounds of the global optimum by efficiently branching and pruning the parameter space.

However, there are few implementations that include these optimization algorithms in GCS/GCSU systems, with the representatives listed as follows. Dehghani et al. (2008) pointed out a hybrid GA coupled with BPNN to automate the hyperparameters of ANN. This gene-evolved Neural Network (NN) was proved to be efficient in predicting the minimum miscibility pressure (MMP) required for gas-injection EOR. Chen et al. (2010) optimized a CO_2-WAG process using a hybrid GA that incorporates the orthogonal array (OA) and Tabu approaches, resulting in a faster convergence speed than classic GA.

4.3.5 Dealing with Real-time Modeling and Optimization

The principal mechanisms governing GCS/GCSU operations display time-varying behavior, including trapping processes and related physical and chemical phenomena. These operations often present formidable challenges, such as strong nonlinearity, time-varying parameters, and significant disturbances, which complicate the design of control algorithms. To effectively address these challenges, the adoption of adaptive control methods capable of accommodating the inherent complexities and accurately describing the key processes within a specific time frame is recommended. Currently, adaptive algorithms are primarily utilized to achieve real-time update of nonlinear modeling in GCS/GCSU operations.

In the research conducted by González-Nicolás (2019), a pioneering adaptive modeling approach was implemented, wherein model parameters are continuously updated through monitoring, calibration, and optimization processes. The study demonstrates the efficacy of adaptive optimization methods in efficiently planning brine extraction activities and thoroughly investigates the impact of initial site characterization data quality and the utilization of recently obtained monitoring data, such as pressure readings from observation wells, on the overall optimization performance. Chen et al. (2020) employed Bagging MARS (BMARS), a variation of the Multivariate Adaptive Regression Splines (MARS) algorithm, to construct a proxy model simulating CO_2 injection and migration. BMARS is built upon an ensemble of MARS models, utilizing different types of local models adaptively in various regions of the data space to capture the effects and interactions between input variables.

Karkevandi-Talkhooncheh et al. (2017) developed an intelligent model based on the Adaptive Neuro-Fuzzy Interface System (ANFIS) to predict MMP values for different reservoir conditions using experimental data. ANFIS, coupling the capabilities of fuzzy logic and NN, constructs a fuzzy inference system and tunes its membership function parameters through neuro-adaptive learning. Both works by Ng & Jahanbani Ghahfarokhi (2022) and Ng et al. (2023) implemented adaptive training in conjunction with optimization in MLP-based proxy modeling. The DDMs were adaptively re-trained by applying an updated training database via the addition of extra samples retrieved from optimization with the proxy models. The aforementioned methodologies can be regarded as the integration of NN into Generalized Predictive Control (GPC) process, where the prediction errors are computed by NN and used to calibrate NN model parameters simultaneously during GPC. In which, real-time learning is essential for adaptive NN control, therefore learning speed becomes a critical concern. Existing adaptive neural controllers suffer from relatively slow execution speeds, making the enhancement of online adaptive neural control speed a current research emphasis.

5 An Intelligent Modeling-Optimization Paradigm Using AI&ML Approach

As discussed in Section 2, the dynamic modeling of CO_2 injection operations aim to predict and optimize storage efficiency and safety, ensuring the long-term sustainability of large-scale underground CO_2 storage. Current practices for optimizing field development strategies differ by approach of adopting the reservoir models in the optimization framework including direct optimization using numerical simulation models and optimization based on proxy models to accelerate the tasks. However, numerical simulation-based optimization algorithms are hindered in need of many simulations to present precise results. Moreover, statistic-based proxy, such as Response Surface Model (RSM) suffers from lack of efficiency to reach the optimal solutions where the response surface is non-smooth and highly multi-modal, whereas reduced-physics and reduced-order proxies may result in certain extent of accuracy and resolution sacrifice, owing to the physics and order degradation of the original system. Hence, intelligent modeling (also known as smart proxy modeling) leveraging AI&ML capabilities to perceive data relationships has become known as the most promising alternative reservoir modeling practice to overcome both the computational limitations of numerical models and efficiency and accuracy degradation in traditional proxy models. AI&ML provide the possibility to synergize traditional and intelligent modeling to develop more powerful computational protocols.

In this section, we will provide a brief overview of the integrated intelligent modeling-optimization paradigm for GCS/GCSU within the background of reservoir engineering. This framework is an amalgamation of the state of the art (Section 4) that serves as the foundation of methodology design, taking a techno-economic perspective into consideration. As shown in Fig.7, three components make up the complete workflow: problem formulation, ML-based proxy model training, and optimization.

5.1 Paradigm Framework

5.1.1 Problem Formulation

Given the extensive time frame of implementation, spanning before, during, and after the GCS/GCSU operation, and the intricate physical and chemical changes among rock and fluids, each step and time-varying variables in the proxy modeling and optimization process demand meticulous consideration, specifically tailored to the unique characteristics of subsurface CO_2 storage. To ensure a rational and scientifically robust approach to tackle this intricate optimization problem, the initial paramount step involves precisely defining the objectives of both the proxy and optimization models and providing a comprehensive description of the relevant aspects. This step is crucial in determining what information should be generated or extracted from the numerical simulation model. Following the precise definition of the optimization problem, reservoir engineers gain a deeper comprehension of the necessary database to develop the corresponding proxy model, while also maintaining a comprehensive awareness of the target optimization objectives, allowing for the effective construction of the four fundamental elements of the proxy and optimization models: input parameters, process variables, output parameters, and constraints on either of the other three. Although the proxy model's input and output parameters are normally equivalent to the optimization model's decision variables and objective functions, there are a few cases where a secondary transformation of the proxy model's inputs and outputs is required before embedding into the optimization model, or vice versa, where the proxy model's inputs are process variables between the decision variables and ultimate objective functions.

Input parameters can be broadly categorized into two types: uncertain and control variables. At the initial stage of developing a representative reservoir simulation model, precise characterization of the input parameters is crucial, encompassing formation properties, rock and fluid behavior, physical and geochemical trapping mechanisms, well conditions, and operating circumstances. Uncertainties often arise due to limited knowledge of petrophysical properties, insufficient rock and fluid data, as well as scarce experimental evidence regarding the physical and chemical effects on trapping and recovery. Additionally, there are a number of operational (CO_2 injection methods, control mode) and technical (well type, perforation length, and well pattern) factors that need to be addressed in large-scale GCS/GCSU projects (Dai et al., 2014). In industrial practice, it is common to treat operational conditions as control variables and view the physical and chemical properties of rocks and fluids as sources of uncertainty. However, there are occasions when the focus shifts towards investigating reservoir geological conditions and rock and fluid properties as variables controlling CO_2 storage capacity and safety. Factors affecting CO_2 storage capacity primarily include reservoir depth, thickness, pressure gradient, porosity, permeability, and fluid properties such as composition, viscosity, and density. On the other hand, factors influencing CO_2 storage safety encompass caprock thickness, lithology, sedimentary sequence, sealing index of caprock, and the presence of abandoned wells.

The outputs of reservoir simulation can be classified into static and dynamic results. Static modeling involves capturing single attributes or spatial phenomena within a single time step, while dynamic modeling addresses time-dependent phenomena within specific grids, wells, or models (Section 4.2). Static outputs include rock and fluid properties (Nielsen et al., 2012) and physical and geochemical mechanism parameters (Li & Jiang, 2020). Temporal variability of static outputs may also be observed under different pressure and temperature conditions. Additionally, the ultimate recovery factor (Safi et al., 2016) or NPV (Rodrigues et al., 2022) achieved within a certain time interval is also classified as static outputs. Dynamic outputs comprise variations over time in CO_2, water/brine, and oil and gas production rates. In addition to these, intermediate states such as BHP and phase saturation/component molar fraction are considered as process variables, which act as important indicators reflecting the input-output relationship. In some cases, process variables can also serve as outputs for subsequent proxy and optimization models (Allen et al., 2017 & Nghiem et al., 2010).

After determining the inputs, states and outputs, the proactive definition and clear specification of constraints based on the specific project background are of paramount importance. These constraints are critical in narrowing down the feasible design domain and increasing the likelihood of finding optimal solutions, thereby ensuring that the resulting models are well-aligned with the project's objectives, safety requirements, and environmental considerations. Not only could we impose constraints on the input and output, but we could also add restriction on the process variables, such as water saturation in grid blocks close to producer wells (Suwartadi et al., 2009).

In the process of formulating a scientific problem, a prerequisite is to ascertain that strong correlation coefficient exists between inputs and outputs, otherwise the problem will be meaningless and insignificant. To achieve this, extra simulation runs will be needed to conduct sensitivity analysis of variables and quantification of uncertainties before embarking on data generation. Remarkably, this crucial step is largely overlooked in numerous proxy models and optimization endeavors. The underlying reason for this omission may stem from the exhaustive elucidation of the causal relationship between input and output variables in preceding research. Nonetheless, this article recommends integrating this step in the problem definition process. Even though the robust linkage between input and output variables has been beyond dispute, varying scenarios across different cases, with even a single precondition changing, can yield diverse degree of sensitivity with the upper and lower bounds of feasible solutions to be adjusted. Moreover, these experimental simulations contribute to a heightened understanding of both the individual and overlapping effects among the variables under investigation. By subjecting the model to preliminary runs, it becomes possible to discern unique influences that variables wield over the output in a particular case. This process, in turn, narrows down the range of feasible solutions, paving the way for an informed and highly efficient optimization process.

5.1.2 ML-based Proxy Model Training

After the comprehensive problem description, we proceed with an organized data generation process before the time-consuming simulation runs. Design of Experiments (DoE) are commonly employed as the first step of data generation, serving the purpose of randomly selecting samples to obtain a representative dataset from the entire population. This allows for a full coverage of cross-correlations and potential causal relationships among multiple variables. Theoretically, as the sampling size of numerical simulation scenarios increase, the proxy model holds enhanced proficiency in digging correct input-output mappings and increase the likelihood of covering the optimal solution. However, the constraints of available computational resources must be considered. In order to attain the most accurate fitting law with minimal numerical computations, it is essential to effectively plan the execution of experiments set under statistically optimal conditions. Modern DoE approaches are the mainstream techniques applied in computational engineering design studies (Giunta et al., 2003), which include space-filling algorithms (e.g., Latin hypercube design (LHD), optimum LHD, maximin LHD), Monte Carlo and stratified Monte Carlo sampling, Quasi-Monte Carlo sampling by quasi-random low discrepancy (QRLD) sequence (Halton, 1960; Sobol',1967; Hammersley and Handscomb, 1964), OA sampling and novel adaptive sampling (Joseph, 2016; Gramacy, 2020; Giunta, 2003). Subsequently, the "raw database" is generated after the sampled simulation runs. Data partitioning can be either applied to the raw dataset following the DoE or to the processed data after the next step of data preprocessing, which involves the splitting of training, validation, and test sets. While there are no strict rules for the exact proportion of each partition, most literature (Ahmadi et al., 2018; Vida et al., 2019; Shahkarami and Mohaghegh, 2020; You et al., 2022) adopts training data that accounts for more than 70% of the total dataset, with validation and test data sets accounting for 10% to 15%.

Data preprocessing is a second step of data collection, which significantly influences the predictive accuracy. This paper proposes the data refinement workflow consisting of data cleaning, normalization, and secondary data extraction before the ANN model training. For instance, the upstream well data assume an unstructured form or contain missing parameters, which is unsuitable for direct use. Therefore, data cleansing and normalization along with appropriate data transformation and integration are necessary to improve the data quality. Zhou & Lascaud (2019) transformed highly heterogenous data to a more fit-for-purpose form, including well spacing, stacking and infill timing data. Nwachukwu et al. (2018a) also proved that the integration of transformed (inter-well and grid-grid connectivity) and first-hand data (well coordinates, porosity, and permeability) would enhance pattern recognition during ANN training. Feature selection serves as another critical aspect of preprocessing, meticulously aimed at refining model accuracy by judiciously eliminating irrelevant or redundant features from the original dataset. While the incorporation of more parameters potentially facilitates the integration of more information, the inclusion of redundant information may inadvertently slow down the training process and even compromise the predictive model's accuracy. Ng et al. (2022) summarized three widely-used feature selection methods. The first method relies upon reservoir engineering expertise, leveraging professional insights to guide the selection process. Additionally, statistical-based techniques such as Z-score, principal component analysis (PCA), isometric feature mapping (ISOMAP), and locally linear embedding (LLE) (Bird et al., 2021), would be effective to identify and retain indispensable features. Fuzzy pattern recognition, as an example of the third method grounded in AI&ML, has been proved to outperform statistical approaches in feature selection (Mohaghegh, 2018).

After the procedure of data generation, the next step is to develop an ML-based proxy model. In order to construct an intelligent proxy model that accurately approximates the mapping between inputs and outputs, a process known as machine self-learning seeks to reduce the error between the desired output (target values) and the expected output (predicted values). These errors are attributed by the dual effect of improper model parameters and hyperparameters. Model parameters are what the machine learns from the data, such as the weights and biases in the NN, the support vectors in the Support Vector Machine (SVM), and the coefficients in Linear Regression (LR). These model parameters undergo continuous updates and automatic optimizations throughout the training process. The initial weight and threshold will affect the distribution of the activation value of the hidden layer, whereas inadequate settings may lead to the gradients vanishing or exploding. Model hyperparameters do not require machines to learn from data. They are pre-set external configurations of the model, such as the learning rate for training NN, C and sigma parameter for SVM, and K value for K-Nearest Neighbor (KNN). Taking a typical ANN based on Back Propagation training (BP-ANN) as an example, the training process consists of 4 steps:

1. Initialization of weights and biases.
2. Forward propagation: The data is propagated through the network in a forward direction, starting from the input layer. At each hidden layer, the data is processed using the activation function before being passed on to the next layer. This step requires careful set-ups of initial weights, biases, and activation functions to prevent the gradient from either exploding or vanishing due to the chain rule multiplication during the subsequent backward propagation.
3. Computation of the loss function: The error is quantified by a loss function such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Average Percent Relative Error (APRE), or Average Absolute Percent Relative Error (AAPRE).
4. Backpropagation: The error initiates from the output layer and is propagated backward from one layer to the preceding layer. Using an appropriate gradient descent algorithm (depending on the choice of different optimizers), the NN's weights are automatically optimized layer by layer, aligning with the direction of error reduction.

These steps are iteratively repeated a specified number of times (i.e., epoch) until it is believed that the loss function has been minimized without overfitting the training data at the same time. During the training of BP-ANN, the learning rate along with the number of hidden layers and neurons, constitute a typical set of hyperparameters to be tuned. However, the rest of hyperparameters, such as batch size, epoch, activation function, and optimizers, are usually optimized through trial and error or general rule of thumb. In general, various optimization algorithms are commonly employed in intelligent systems, including manual tunning, grid and random search, Bayesian Optimization (BO), gradient-based optimizers, and evolutionary optimization. Among these, evolutionary optimization algorithms can be seamlessly integrated with established intelligent proxies, exhibiting strong robustness and extensive applicability in both parameter and hyperparameter optimization. Currently, among the ML-based training practices in the discipline of reservoir engineering, the optimization of initialized model parameters is relatively less explored compared to the optimization of hyperparameters (Nait Amar et al., 2018). Furthermore, there has been limited research on simultaneous optimization of model parameters and hyperparameters, possibly due to concerns that curse of dimensionality may arise as the size of design variables increases.

Lastly, the evaluation of proxy model performance involves three distinct phases: training, validation, and blind testing. In each phase, we employ various metrics, including the Coefficient of determination (R2) along with multiple loss functions to quantify the model's effectiveness. In essence, after determining the model hyperparameters using the validation set and adjusting model parameters based on the training set, the model is subjected to the blind testing set to examine the model's generalizability. During blind testing, it is critical to keep the blind data separate from the training and validation datasets. Besides, due to the fact that intelligent proxies are typically more capable of interpolation tasks than extrapolation tasks (Xu et al., 2020), it is strongly advised to generate the blind testing dataset within the bounds of the previously generated database. Furthermore, the poor performance of proxy model can be diagnosed by the differences in accuracy between the training set and the validation set (or blind testing set). Higher validation or blind testing accuracy compared to training accuracy may result from factors such as uneven dataset partition, excessive model regularization, or superfluous data preprocessing that alters the true distribution of the training data. On the other hand, if the training set possesses higher accuracy than the validation or testing set, it suggests that the machine has not fully captured the underlying physics. Possible reasons could be inappropriate settings of hyperparameters leading to local optima, insufficient training and validation data, or poorly-handled data preprocessing causing substantial perturbations in the results. By comparing the accuracy across the training, validation, and testing sets, we can uncover solutions targeted to these disparities, ultimately guiding data reconstruction and improving the success rate of the retrained model.

5.1.3 Optimization

Numerous research efforts have demonstrated establishing intelligent proxy models that achieve high predictive accuracy. However, more recent studies have gone beyond mere accuracy improvement and aimed to expand the application domain of proxy models. Several studies have pursued the integration of proxy model techniques with advanced optimization algorithms, ushering in new possibilities for the parallel development of optimal decision-making workflow with high precision and speed. To delve into the technical details, the four core elements of the optimization model, consisting of objective functions, design variables, uncertainties, and constraints, must be tightly matched with ML-based proxies.

In the optimization phase (the third component of our paradigm), we begin by developing mathematically generic mono- and multi-objective functions, which encompass a comprehensive range of considerations, including technical, operational, computational, environmental, and economic aspects. To construct the hierarchy of these objectives, Pareto-based methods are employed to provide decision support under uncertainty, allowing for the exploration of optimal trade-offs without predetermined weights. The decision variables undergo prefiltering based on sensitivity and KPI analysis to enhance their relevance and effectiveness as described in Section 5.1. To address the challenges posed by uncertainty, RO emerges as an ideal and easy-to-implement choice. It considers uncertainty as a set of deterministic scenarios, eliminating the need for distribution models or fuzzy membership functions. Also, the constraints in RO need to be strictly satisfied in the worst-case scenario, which guarantee the solutions remain feasible. As for the articulation of constraints, the objective function could be confined by penalty and repair strategy. To avoid the local optima, the crucial part is to select the most competitive optimization algorithm to seek the global optimum. Thus, the application of a global optimizer (with constraints) as a coarse approximation, such as nature-inspired algorithms, and a local optimizer as a fine-tuning aid, such as the GPS method, provides flexibility. The last, we adopt the dynamic modeling-optimization by adding adaptive training management during the optimization iterations, where SPMs are retrained along with the update of the training data during the optimization. This incremental approach progressively strengthens the proxy model's accuracy in critical regions, accelerating the convergence towards the optimal solution.

5.2 Limitations

Although the proposed paradigm represents a significant advancement in addressing numerous bottlenecks inherent in the traditional modeling-optimization framework, it is imperative to recognize that certain limitations of data-driven proxies also become entwined with optimization applications. Firstly, real-world field development planning (FDP) designs are of great complexity, requiring the optimization of a vast number of temporal and spatial variables that often conflict or intersect with each other within a limited computing timeframe. The adoption of sophisticated network architectures, deep learning, or hybrid ML approaches may enable simultaneous training, with incomplete and incorrect mappings to be learned and exponentially increased model training time to be spent. Moreover, building one single proxy model for simultaneous multi-variable and multi-objective prediction also demands an exponential growing amount of reservoir simulation runs to cover different scenarios.

Secondly, the limitations of data-driven approaches are linked to the applicability of proxy models (Ng et al. 2021). ML with multiple input variables proves more effective when capturing attributes that share high correlation as well as similarities in terms of data type and quantity. Fundamentally, DDMs are only relevant to the specific reservoir being studied and lack transferability. For instance, proxy models designed for aquifer storage prove to be inadequate when applied to carbon storage in depleted oil and gas reservoirs. The distinct storage sites introduce a multitude of variables, many of which demonstrate high sensitivity to the objective function yet hold little correlation to one another. Similarly, if a model is aimed at the water-flooding oil reservoirs, it cannot be readily extended to EOR methods, such as CO_2-WAG injection. Though the incorporation of Cartesian coordinates, deviation angles, and fixed well lengths into a single proxy model has been extensively studied and shown to be feasible, the introduction of a Boolean-type variable representing well type (vertical or horizontal), or varying numbers of wells would significantly complicate the fitting process. Additionally, the two control methods for CO_2 injection well operations, namely injection rate and BHP control, cannot be simultaneously trained within the same model due to their convoluted interactions. As an input, the injection rate alters the BHP as a state variable, while the bottomhole pressure, when used as an input, changes the injection rate as a state variable. Moreover, the automatic transition between different control mechanisms in realistic operations poses challenge in determining the optimal well control framework (considering both injection rate and BHP) as the design variables during optimization will not automatically switch from injection rate to BHP.

Indeed, data-driven intelligent proxies are not the master key to unlock all kinds of situations. Prediction and optimization of complex scenarios still require further fine-tuned strategy of building proxies. In a study by Kim et al. (2021), a new Long Short-Term Memory (LSTM) and Convolution Neural Network (CNN)-based sequential proxy modeling procedure which comprises the well pattern and operation optimizations was successfully developed. LSTM optimizes the number, location and type of wells, using time-of-flight (TOF) maps reparametrized from permeability model as the input. CNN-based proxies' input is BHP matrix (time series × the number of wells) with fixed well pattern parameters from LSTM. PSO algorithm is used for both optimizations to consider the geological uncertainty in the permeability field. Matthew et al. (2023) established 6 proxy models to solve a three-variable and dual-objective CO_2-WAG optimization problem using NSGA-II. The entire 10-year production simulation data was segmented by CO_2 and water injection. Each scenario was further divided into three proxies by the timely behavior of CO_2 flow in the subsurface: 1st year, 2nd-5th year, and 6th–10th year to enhance the quality of proxy models. These techniques catered to specific scenarios serve as the icing on the cake of basic paradigm proposed in this article, enriching its capabilities in dealing with intricate optimization scenarios.

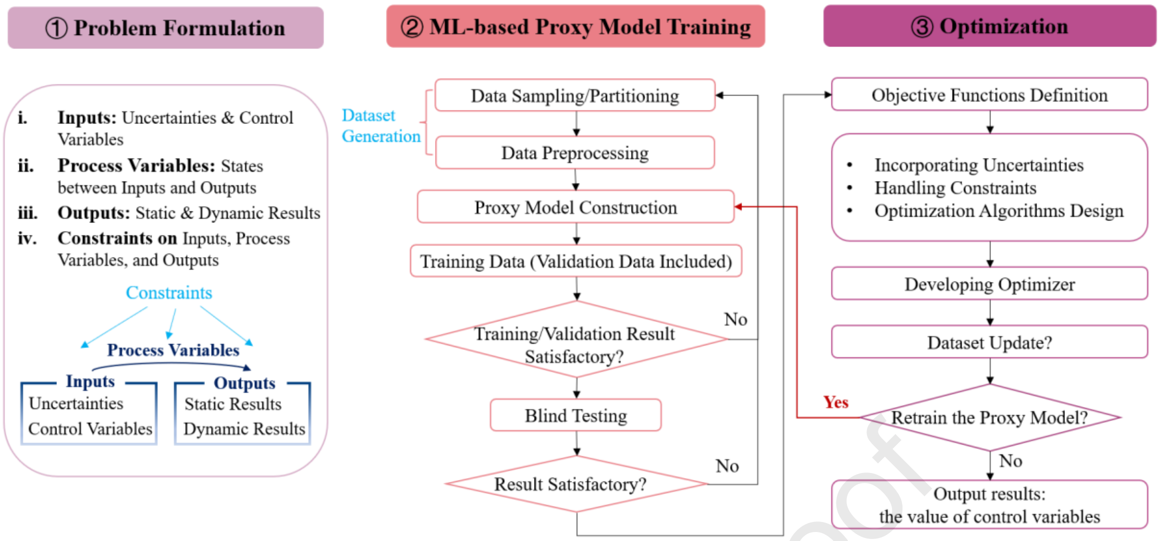


Fig.7. The advanced modeling-optimization paradigm using intelligent proxy approach.

6 Applications of Intelligent Modeling-Optimization in GCS/GCSU

The concerns surrounding CO_2 leakage raised by policy makers and the general public, coupled with the desire of operators, investors, and government agencies to optimize CO_2 storage capacity while minimizing costs, highlight the significance of intelligent modeling-optimization applications in GCS and GCSU initiatives. Dynamic model studies serve as essential forecasting tools, providing quantitative insights into CO_2 behavior and reservoir characteristics, enabling the identification of the optimal approach to maximize storage capacity while simultaneously minimizing the risk of plume leakage The integration of dynamic modeling with optimization techniques further enhances the efficiency and precision of decision-making processes, elevating the overall performance and success of such carbon storage endeavors.

In this section, our focus lies in summarizing recent industrial storage projects and pilot experiments, with particular emphasis on the ML-based proxy model coupling with optimization applications at different stages of GCS/GCSU: pre-operational, operational, and post-operational phases. After searching through the case studies, it can be observed that an overwhelming number of papers on GCSU manage to encompass the entirety of prediction and optimization whereas the relevant applications of GCS primarily pertain to prediction. Moreover, given that GCS and GCSU applications differ in their goals, we have further divided the pre-operational, operational and post-operational applications into two categories —GCS and GCSU—for clear distinction. Fig.8 illustrates the domains investigated in this section, primarily centered around the discipline of reservoir engineering. Each subsection will present a comprehensive collection of relevant literature, outlining the data collected, methods utilized, main results achieved in each work. By delving into the applications of intelligent modeling-optimization in diverse scenarios, we aim to shed light on the possibilities, advancements, and challenges in GCS/GCSU implementations, ultimately guiding the progression and successful deployment of this paradigm in real-world applications.

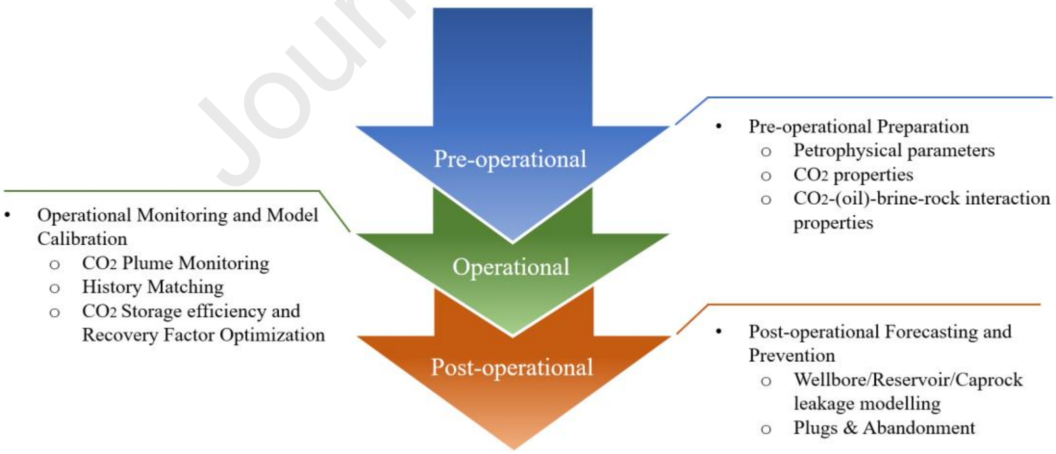


Fig.8. An overview of intelligent modeling-optimization applications across pre-operational, operational, and post-operational stages of GCS/GCSU projects.

6.1 Pre-operational Preparation

6.1.1 GCS

Within the framework of the pre-injection preparation stages (as detailed in Section 2.1), the conclusive stage of risk characterization will be initially completed during the pre-operational modeling. Then, the risk indicators for GCS candidate assessment will be updated during the preceding operational and post-operational stages, utilizing fine-tuned models that increasingly approximate reality through field data calibration. As discussed in effective assessment, a viable candidate for GCS must have adequate storage capacity to hold desired amount of CO_2, sufficient well injectivity to inject CO_2 at acceptable rates through as few wells as possible, high-security trapping mechanisms to retain the CO_2 effectively and efficiently and robust confinement to permanently isolate the sequestered gas from the environment. This section will extensively cover the pre-operational prediction attempts across four dimensions of capacity, injectivity, trapping, and confinement.

Low-viscosity CO_2 will face less resistance and higher mobility in a reservoir with high permeability, which will be advantageous in terms of storage capacity and well injectivity. On the other hand, injecting gas of high transmissibility will lead to gas fingering, which severely impairs the efficacy of trapping and confinement. Thus, the CO_2 storage performance is predominantly influenced by both petrophysical parameters and supercritical CO_2 properties. The introduction of supercritical CO_2 triggers chemical, physical, and mechanical reactions within the geological formations of these sites, which contribute to the alternation in both porosity and permeability of rock matrix. For one thing, CO_2 engages in dissolution and precipitation interactions with various multiphase fluids, encompassing water, oil, CH_4, and minerals that reside within the rock matrix and pores. For another thing, the supercritical CO_2 injection results in effective stress reduction of a rock formation, accompanied by matrix expansion and changes in deformation modulus. Yan et al. (2020) studied the role of CO_2 injection pressure, buried depth, temperature, and coal mechanical characteristics on coal permeability evolution. This was accomplished through the implementation of SVM-based proxies hybrid with six evolutionary optimization algorithms. In parallel, the work carried out by Mardhatillah et al. (2022) represented an extension of the SVM-hybrid framework, designed to predict the permeability change during CO_2 injection in saline aquifer. To examine the influence of salt precipitation and the fine migration on the change of CO_2 injectivity throughout the CO_2 sequestration process, this research included brine salinity, CO_2 injection rate, particle sizes, and particle concentration to the input datasets for the proxy model. Understanding the transfer of mass and energy during CO_2 transportation in subsurface relies heavily on its thermophysical properties. Abdolbaghi et al. (2019) indicated that attempts to predict viscosity and thermal conductivity by equations of state is not successful. Extensive demonstrations have shown that AI provides more dependable computational approaches than conventional empirical methods. Nait Amar et al. (2020a, 2020b) focused on using data-driven methods to predict CO_2 viscosity and thermal conductivity across varying thermodynamic conditions, with MLP-LM model yielding the best fit compared to MLP, Gene Expression Programming (GEP), and Group Method of Data Handling (GMDH). They also developed innovative methods using MLP and RBFNN models, further enhanced by Committee Machine Intelligent Systems (CMIS). The research by Talebi et al. (2023) positioned Decision Tree (DT) and Random Forest (RF) techniques as the top-performing algorithms among the methodologies considered, offering improved predictions for CO_2 viscosity.

The sequestration of CO_2 mainly involves four trapping mechanisms: structural trapping, residual trapping, solubility trapping, and mineral trapping. From structural trapping to mineral trapping, the security of long-term storage arises. Nowadays, more and more research has shed light on residual trapping, solubility trapping, and mineral trapping with higher reliability. Among these trapping mechanisms, the interaction nature among CO_2, brine and rock plays an important role. Due to capillary forces and interfacial tension, some of the CO_2 gets trapped in the pores of intermediate rocks. Dehaghani & Soleimani (2019) introduced a novel Stochastic Gradient Boosting (SGB) tree algorithm for predicting CO_2-aquifer brine interfacial tension (IFT), accounting for temperature, pressure, and brine salinities, which outperforms various other ML models, including RBF, MLP, Least Squares Support Vector Machine (LSSVM), and adaptive neuro fuzzy inference system. Safaei-Farouji et al. (2022) demonstrated precise IFT forecasting with the RF model emerging as particularly effective. Tariq et al. (2023) implemented the Feed-Forward Neural Network (FFNN) to model the contact angles in a ternary system of rock minerals (quartz and mica), CO_2, and brine under different pressure and temperature. Another portion of CO_2 gets trapped within brine due to solubility and diffusion processes. In the work by Nait Amar et al. (2019), CO_2 solubility in brine was modeled using MLP and RBFNN, with inputs encompassing molality of NaCl, pressure, and temperature. Mohammadian & Riazi (2022) compared XGBoost, MLP, KNN, and GA to predict solubility based on pressure, temperature, and salinity. GA has been proved to hold the best performance in terms of all the metrics. Ratnakar et al. (2023) estimated the CO_2 solubility in brine employing a ML-based workflow, integrating physics-based understanding with DT, RF, and ANN. Ion properties and their molar concentrations, pressure, and temperature are the selected input variables. As for CO_2 diffusion coefficient prediction, the method introduced by Feng et al. (2018) provided a fast and precise prediction of CO_2 diffusivity in brine under reservoir conditions. Bemani et al. (2020) took a distinctive approach, employing the Adaptive Neuro-Fuzzy Inference System (ANFIS) in tandem with five diverse evolutionary algorithms. Additionally, effective outcomes are achieved through white-box ML techniques by Nait Amar et al. (2020), where the GMDH and GEP were employed to establish relationships between the diffusivity coefficient of CO_2 in brine and key parameters including pressure, temperature, and solvent viscosity. The most reliable trapping mechanism involves geochemical reactions between CO_2 and minerals within the reservoir. Ahmed et al. (2021) analyzed reactive-transport simulation data of GCS system using Non-negative Matrix Factorization (NMF), which is an unsupervised ML algorithm. From a dataset containing 19 attributes, NMF model identifies four reaction stages and the dominant attributes in each stage, such as calcite, dolomite, and ion concentration. Tariq et al. (2023) have developed two CNN-based model with different architectures to effectively predict the dissolution and precipitation of various essential minerals, including anorthite, kaolinite, and calcite, during injection into deep saline aquifers.

6.1.2 GCSU

GCS initiatives face unique challenge posed by stored CO_2's lack of inherent commercial value, which conventionally renders pure storage projects seemingly less attention-grabbing than GCSU from a business perspective. As the industrial landscape gravitates towards solutions that align environmental consciousness with economic gain, GCSU stands out as a transformative avenue to bridge the gap between sustainability and profitability. In this context, it is crucial to extend the investigation beyond the conventional scope of physical and chemical interactions within the CO_2-brine-rock system. This entails paying extra attention to the interplay of crude oil with CO_2, brine, and rock.

The physics of CO_2-oil system has garnered significant attention with the increasing use of CO_2 as an enhanced recovery agent since 1980s (Alvarado & Manrique, 2020). However, accurately determining the properties of CO_2-oil mixtures is quite challenging as direct measurements are often elusive and existing empirical models are tailored to specific assumptions and simplified hydrocarbon compound complexities. Addressing this issue, ML-based models emerge as efficient and dependable tools. The RBF model provided fast prediction of density of CO_2-oil mixture providing accurate predictions as well as describing the density crossover phenomenon (Moradkhani et al. 2023). The effective CO_2 injection for miscible flooding relies heavily on MMP, a parameter that has gathered substantial research interest. It was found that precise estimates of MMP can be obtained using AdaBoost integrated with classification and regression trees (AdaBoost-CART) and a hybrid-adaptive neuro-fuzzy inference system (Hybrid-ANFIS) (Ghiasi et al., 2021). Tree-based and DL algorithms like CatBoost, XGBoost, and Deep Neural Network (DNN) were implemented for MMP prediction in oil-CO_2 streams (Lv et al., 2023). Huang et al. (2023) predicted MMP of pure/impure CO_2 and crude oil systems based on a conditional Generative Adversarial Network (cGAN) and BO algorithm. In addition, Bagalkot & Keprate (2021) employed ML to characterize CO_2 diffusion in hydrocarbons, assessing algorithms such as Gradient Boosting, Gaussian Process Regression, KNN, and DT to calculate CO_2 diffusion coefficients in n-decane. Salehi et al. (2022) modeled IFT between a CO_2/N2 mixture and n-alkanes at different pressure, temperature, n-alkane carbon number, and N2 mole fraction with RBF model exhibiting exceptional precision than six other ML methods.

Pointed to phase flow behavior of CO_2-oil-brine system, Jiang et al. (2021) conducted investigation of viscous coupling effects in pore-scale three-phase flow under different saturation, geometric parameters, wettability and viscosity ratios. ANN model is applied to predict the permeability of three-phase flow in pore network connected by tube-shaped throats. ML has also been applied to address the phenomena at the interface between oil and brine, where the underlying correlation with multicomponent hydrocarbon and brine containing varying concentration of cations remain ambiguous. Results by Nait Amar et al. (2019) revealed that Gradient Boosting Decision Tree (GBDT) provide very satisfactory predictions for IFT determination in crude oil/brine systems. Kirch et al. (2020) combined ML approaches including RF, ET, GB, and EN with conventional molecular dynamics simulations (MD) to estimate oil/brine IFT efficiently.

With the goal of GCSU in tight reservoirs, the special physics of CO_2 adsorption on clay minerals should be remarked. Bemani et al. (2020) introduced LSSVM optimized through PSO. This approach aimed to learn and forecast methane and CO_2 adsorption capacity in Jurassic shale samples across various gas mixtures, with the input factors encompassing pressure, temperature, gas composition, and TOC. Similarly, the correlations of adsorbed methane in shale formations based on pressure, temperature, moisture, and TOC have been established by two rigorous data-driven techniques, namely GEP and GMDH (Nait Amar et al., 2021).

6.2 Operational Monitoring and Model Calibration

6.2.1 GCS

The ongoing injection of CO_2 into reservoirs induces a cascade of transformations, not only in fluid composition but also in the characteristics of the reservoir rocks. Monitoring the real-time migration of CO_2 during operational stage is imperative to unravel the implicit complexities of underground CO_2 storage and thereby facilitating accurate forecast of CO_2 migration behaviors. This necessitates a two-fold approach of forward modeling and inverse modeling.

Forward modeling serves as our lens into the reservoir's behavior over time, enabling us to comprehend the dynamic interplay of these variables as CO_2 injection progresses. An effective forward monitoring model usually involves the spatial and temporal evolution of various indexes such as pressure, saturation, trapping performance as well as injection and production rates. There are two main streams of ML-based modeling techniques to address computational challenges associated with CO_2 injection modeling: one is common pool of ML techniques such as ANN, SVM, RF and LR to solve the final-step prediction problem and the other is temporal DL models that exhibit strong performance in predicting the migration of CO_2 plume across the entire duration of the simulation. To solve the spatial distribution of CO_2 at the final timestep, opting for simpler ML techniques is better suited than sophisticated CNN model since it skips over abundant training of intermediate timesteps. Wu et al.

(2021) developed an ANN emulator with fluid pressure distribution and CO_2 saturation as target responses, highlighting the influence of capillary entry pressure on CO_2 mineralization, permeability alteration, and fluid mobility. In the study by Alali (2023), the widely-used ML techniques including KNN, RF, Multi-output regression were employed to estimate the CO_2 saturation map at the end of 2000-day period. The main physics encountered in the process were gravity segregation and capillary trapping. Davoodi et al. (2023) employed three robust ML and one DL algorithms to model the solubility trapping index (STI) and residual trapping index (RTI) of CO_2 in saline aquifers. The findings highlight the LSSVM model as the most accurate, even better than CNN model. Al-Qaness et al. (2023) presented the use of an optimized LSTM for predicting STI and RTI in deep saline aquifers. The authors discussed the increased computational cost associated with incorporating LSTM in AOSMA, a hybrid of the Aquila Optimizer (AO) and the Slime Mold Algorithm (SMA). However, there is an absence of statistics data regarding computational time of proposed LSTM framework.

It is undoubtful that there is continuous progress in utilizing DL techniques to establish proxy models. Extensive research efforts have been directed towards harnessing DL techniques for predicting the dynamic distribution of subsurface properties over time. CNNs, renowned for their proficiency in image-based data processing, offer great convenience when converting real-time tracking of CO_2 plume to an image-to-image task. Mo et al. (2019) applied a deep convolutional encoder-decoder approach to model multiphase CO_2-water flow systems, with the inclusion of injection duration as an additional input parameter for time-dependent predictions. Likewise, a conditional Deep Convolutional Generative Adversarial Network (cDC-GAN) as a proxy model was proved to be capable of learning complex mappings between high-dimensional permeability field and changing CO_2 saturation fields through the time (Zhong et al. 2019). Sun (2020) proposed a Deep Multitask Learning (DeepMTL) approach based on the U-net architecture to forecast CO_2 sequestration dynamics within brine aquifers. This predictive model takes permeability and time-varying injection rates as inputs parameters, yielding temporal changes in pressure and saturation as outputs. A Residual U-Net (R-U-Net) based CNN model was proposed by Wen et al. (2021) to predict the migration patterns of CO_2 plumes away from an injection well. This model was designed to accommodate variations in permeability fields, as well as various injection parameters, encompassing injection duration, rates, and depth. For the purpose of imaging and visualization of the evolving CO_2 plume using routine pressure/temperature measurements, Nagao et al. (2022) presented an autoencoder-decoder network combined with Multi-Dimensional Scaling (MDS) to handle the inefficiency caused by high-dimensional training outputs. The approach utilizes field measurements and an encoder network to predict latent variables, which are then fed into decoder network to generate 3D onset time images. In a recent development, the U-LSTM-net NN, as demonstrated by Lin et al. (2024), surpassed the performance of both U-net and Attention-U-net models. It adeptly integrates spatio-temporal information, showcasing robust memory capabilities, and proficiently handles the simultaneous learning and prediction of multiple flow fields. Furthermore, it leverages techniques like Transfer Learning and Gradient Normalization (GradNorm) method, resulting in remarkable predictive performance across dynamic attributes with different hierarchies.

Simultaneously, inverse modeling is undertaken to enhance the fidelity of the forward model which involves a meticulous calibration of the dynamic model to minimize the gap between simulated data and monitored behavior. In particular, ML techniques have found valuable applications in reducing the computational cost of data assimilation processes, enabling researchers and practitioners to efficiently manage the large amount of forward simulations involved. MARS was used in the filtering-based data assimilation process to quantify the uncertainty of CO_2 leakage (Chen et al., 2018). In this application, the measurement locations and data type were optimized based on the extent of uncertainty reduction. Tang et al. (2022) developed 3D recurrent R-U-Net proxy model capable of predicting flow and geomechanical responses in CO_2 storage operations. The inverse problem was resolved by using the synthetic surface deformation data obtained from the proxy model (only small portion of observation data) to determine the precise permeability and porosity fields in the aquifer. In another study conducted by Tang et al. (2021), an Ensemble Smoother Multiple Data Assimilation (ES-MDA) framework was built for concurrent estimation of pressure history and the spatial extent of a CO_2 plume, where wide residual network and R-U-Net architecture was used to forecast the CO_2 plume pressure and saturation maps respectively based on permeability distribution and well locations. Chen et al. (2022) implemented the advanced version of ES-MDA with geometric inflation factors (ES-MDA-GEO) as the history matching model to update the permeability field, where Fourier Neural Operator (FNO)-based proxies after feature coarsening were applied as forward model. Furthermore, Han et al. (2023) gained posterior estimates of meta-parameters in storage systems by Markov chain Monte Carlo (MCMC)-based history matching methods, with the R-U-Net model functioning as the accelerated tool. Most recently, Physics-Informed Neural Networks (PINN) have emerged as a transformative approach that imbues DDMs with physical consistency (Raissi et al., 2019), which demonstrate significant promise in overcoming the limitation of sparse observed data and mitigating geological and fluid flow uncertainties by inverse-solving the PDE parameters. Most PINN implementations have demonstrated efficacy in petroleum engineering studies, providing physically plausible predictions of oil production, water flooding and Enhanced Oil Recovery (EOR), as well as inverse estimation of Buckley-Leverett equation parameters and decline curve analysis (DCA) model parameters (Maniglio et al., 2021; Tadjer et al., 2022; Liu et al., 2023; Gladchenko et al., 2023; Manasipov et al., 2023). Despite its potential, PINN is still in its early stages (Franklin et al., 2022; Almajid & Abu-AI-Saud, 2022) and requires careful considerations for broader application in GCS studies, with current research primarily focusing on simplified scenarios. Zhang et al. (2023) extended the PINN investigation from water flooding (Fraces et al., 2020) to CO_2 storage in aquifers by introducing a gravity term into the governing equation. They applied a two-shock Buckley-Leverett model as their PDEs to model miscible CO_2-brine displacements. Du at al. (2023) embedded Fourier feature in the PINN architecture to model density-driven flow under the context of carbon sequestration. These works shared the same simplification that the mass of CO_2 was transferred in a 1D or shoebox-shaped homogeneous subsurface porous formation. The work by Wang et al. (2023) served as the first try for extending hybrid Physics-Informed Data-Driven Neural Network (HPDNN) in a non-homogeneous application. Their PDEs encompass the CO_2 adsorption, diffusion, dissolution, Darcy flow, and slip flow at multi-scale shale reservoir coupled with wellbore.

6.2.2 GCSU

In terms of the CO_2 monitoring over time, GCS operations have given most of consideration to mitigate geological uncertainties including permeability and porosity fields during prediction. However, optimization of operational strategies such as injection schemes, rates, duration, well control, pattern, location, depth, etc. has been widely researched in monitoring efforts of GCSU because it is crucial to ensure not only effective CO_2 storage but also a high economic return. Wherein, ML-based proxies emerge as powerful tool to accelerate the optimization of dual or multiple objectives.

While operating GCSU in oil and gas fields, CO_2-WAG stands out as one of the most promising and extensively explored EOR technique due to its unique ability to mitigate challenges such as CO_2 fingering and premature breakthrough. Nwachukwu et al. (2018b) employed the XGBoost to develop the proxies and MADS to optimize well locations as well as water, gas injection rates and gas-water slug ratio under geological uncertainty. You et al. (2020) proposed an ANN-based PSO optimizer to strike a balance between flooding and storage performance. Various WAG parameters were optimized simultaneously, including the gas/oil ratio, WAG cycle, production rates, fluid injection rates, and BHP of producers and injectors as well as placement of new infill wells. Bocoum and Rasaei (2023) contributed to optimize CO_2-gas-water injection time ratio (GWITR), producers BHP, and fluid injection rates using a hybrid algorithm combining ANN and NSGA-II. The NSGA-II algorithm identifies a diverse and convergent Pareto Front, providing multiple decision-making solutions for operators. Other than CO_2-WAG injection, the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm was proposed for optimizing multi-well production-injection rates in CO_2 flooding, effectively improving NPV and sweep efficiency (Rongtao et al., 2022). Besides, Rezk et al. (2023) utilized ANN to develop predictive models for the intermittent CO_2-assisted gravity drainage (CO_2-GAGD) process.

The other potential geologic formations for GCSU include shales, coal seams, etc. These formations provide prospects for CO_2 injection to enhance coal bed methane and shale gas recovery. Hamdi et al. (2019) presented a Bayesian workflow to optimize CO_2-HnP control variables including CO_2 injection rate, huff and soaking time, as well as the BHP during puff under uncertainty, in which kriging-based proxy model is used to accelerate MCMC estimation of posterior distributions of uncertainties. Enab (2023) emphasized the importance of well design and injection strategy when dealing with tight shale reservoirs characterized by abundant fractures. This optimization process takes into account various elements, including well design, hydraulic fracture design, gas injection rates, the number of injection cycles, and the duration of injection-soaking-production periods. Production forecasting of CBM wells is crucial for seeking an optimal development strategy to reach maximum economic benefits. In the study by Du et al. (2023), a data-driven methodology was employed, utilizing dynamic production data and geological information from 530 CBM wells. This approach integrated the XGBoost and BO algorithm to construct a predictive model for methane production capacity.

Finally, it is worth noting a novel approach referred to as CO_2 injection combined with saline water/brine recovery (CO_2-WR). Though there is no utilization of brine, we categorize CO_2-WR among GCSU applications due to the recovery of byproducts other than CO_2. CO_2-WR mostly functions as reservoir pressure management tool to create additional space for aquifer CO_2 injection. There is a lack of investigations exploring inclusion of brine production as the objective function at aquifer sites. As a demonstration, Omosebi et al. (2021) trained DL-based proxy models utilizing MLP, CNN, LSTM and GRU algorithms for predicting reservoir pressure, CO_2 saturation, and well extraction rate. Musayev et al. (2023) addressed the challenge of determining optimal CO_2 injection and brine extraction well locations for pressure management in the Pohang Basin by proposing the use of an ANN-based proxy model and GA optimization algorithm. However, as the production of saline water can be a costly endeavor that accounts for almost 50% of the overall costs during oilfield development (Farajzadeh et al., 2019), the target functions shift towards the maximization of CO_2 storage while minimizing the production of saline water when storing CO_2 in depleted hydrocarbon fields. The production of saline water from aquifers can also incur high expenses, yet the costs are often overlooked during CO_2 storage in aquifers. To fill in the gap that simultaneously considering the roles of produced brine and stored CO_2 in aquifers, Vaziri and Sedaee (2023) explored the coupling of MARS with NSGA-II as a computationally efficient choice for CO_2 injection rate and duration optimization.

6.3 Post-operational Forecasting and Prevention

6.3.1 GCS

The assessment of post-operation phase risks in GCS systems heavily relies on wellbore leakage models. Although full-physics numerical simulations of wellbore leakage may provide enhanced predictive accuracy (Section 2.3), their substantial computational demands make them unsuitable when optimizing storage in large-scale depleted fields with numerous potential leaking pathways through the existing wells (Middleton et al., 2020). To address this computational challenge, SPM emerges as a pragmatic strategy, which enables an implicit auto-examination and learning of the primary processes governing well leakage, thus affording a notable reduction in computational workload.

In GCS applications, the literature usually followed the system-level modeling approach by decoupling wellbore component from the specific reservoir model. Harp et al. (2016) concentrated on reduced-order modeling using the MARS algorithm, specifically designed for a system-level model that separates reservoir, wellbore, and shallow aquifer components. In this setup, the reservoir model computes pressure and saturation data at the wellbore entry point, that sequentially pass to the wellbore model to calculate the leakage rates to aquifer. DL-based proxy models have been a popular way to predict reservoir responses during CO_2 storage post-operations, serving as input data for sequentially modeled wellbore response (Zhang et al., 2018). They utilized reduced-order cemented and open wellbore models from Well Leakage Analysis Tool (WLAT) with the simulated data from TOUGH2 as the input. The influence of factors such as depth, location relative to injection point, and effective near-wellbore permeability (the surrounding area of abandoned wells) on CO_2 and brine leakage rates were investigated within the above-seal monitoring interval. Gan et al. (2021) integrated ROMs derived from reservoir simulations into the NRAP-IAM-CS toolset to assess the potential for CO_2 and brine leakage via wellbores. Furthermore, they developed an additional ROM that utilizes leakage rates calculated from the previous ROM to predict the geochemical impacts on a shallow aquifer. In a contrasting approach, Baek et al. (2023) pioneered AutoKeras algorithm-based proxy to predict CO_2 and brine leakage rate from the wellbore and then coupled the wellbore leakage model to a reservoir model. Within the coupling process, one more sub-proxy applying classification ML algorithms was built to identify the presence of CO_2 leakage through the well and across the caprock layer.

An alternative approach involves constructing a single proxy model that combines both geological and wellbore aspects. Nevertheless, this approach may experience a certain extent of simplification towards wellbore simulation. In the workflow for real-time CO_2 leakage rate forecasting under uncertainty (He et al., 2021), a LSTM-based proxy model was utilized with input-output pairs collecting from one single black-oil model built by MRST, whereas specific wellbore details were not included in this study. In the study by Raad et al. (2022), sensitivity analysis of wellbore transient behavior during CO_2 injection was conducted using RBFNN and Polynomial Regressions. A synthetic 2D radial model coupled wellbore and reservoir was constructed by CMG's STARS^{TM} + FlexWell, revealing that the most influential factors on wellbore BHP were CO_2 injection rate and permeability of the target layer. Meguerdijian et al. (2023) used DNN-based ROM to estimate brine and CO_2 leakage rates through a fault. FEHM (Finite Element Heat and Mass) simulator considering fluid mechanics and thermal effects was applied to generate the training data, where most focus has been attached to properties of faults and underlying aquifers as well as the well operation parameters rather than the wellbore information.

The mystery of THMC coupling processes involved in wellbore leakage damages the confidence in simulation results for wellbore integrity, even when employing analytical, semi-analytical, experimental, numerical, or data-driven techniques. Therefore, some studies assess the risk of wellbore failure by skipping the full physics and instead analyzing existing wellbore data, considering well conditions like construction year, location, construction materials, well pattern density, extent of CO_2 exposure to the wellbore, etc. Using well design and operational data from over 500 CO_2-exposed wells (Li et al., 2018), a computerized statistical model with NN algorithm was developed to predict the LPI (Leakage-safe Probability Index). Plug and Abandonment (P&A) operational data from established wells, sourced from the Alberta Energy Regulator (AER) in Canada, have been analyzed to assess the potential for GHG emissions through these wellbores (Ugarte and Salehi, 2023). Their research shows RF as the most effective method for classifying wells and prioritizing those at greater risk of leakage. Bai et al. (2023) proposed FBN model based on Bayesian Directed Acyclic Graphs (BDAG) to evaluate the probability of CO_2 leakage and identified potential risk sources, with a case study highlighting cementing quality and corrosion as significant contributors.

6.3.2 GCSU

The widespread use of CO_2 in EOR has proven economically beneficial by enhancing hydrocarbon extraction from various reservoirs. However, as these reservoirs become depleted and filled with CO_2, a new and growing concern emerges: the potential for leaks through abandoned wells. This transformation of once-active oil reservoirs into repositories of CO_2 poses a significant environmental risk. Chen et al. (2022) investigated the potential for CO_2 and oil leakage from abandoned wellbores in a CO_2-EOR field, employing MARS-based reduced-order models created through numerical modeling consisting of aquifer, caprock, reservoir and cemented wellbore components. The Monte Carlo simulations revealed that the number of uncertain parameters affecting oil leakage—including reservoir depth, abandoned reservoir pressure multiplier, caprock thickness, near-wellbore permeability, reservoir permeability, initial oil saturation, residual CO_2 saturation, and fraction of oil components—is greater than the factors influencing CO_2 leakage, which are limited to reservoir depth, wellbore pressure, and near-wellbore permeability. In a similar vein, Mehana et al. (2022) pursued the same objective by developing reduced-order models using the Light Gradient Boosting Machine (LGBM) technique, which demonstrated superior performance compared to MARS. Monte Carlo simulations were also conducted to quantify parameter uncertainties, highlighting the significance of near-wellbore permeability in leakage profiles of both oil and CO_2. In continuation, the findings presented by Lei et al. (2022) underscored the superior performance of the NN model compared to both MARS and Gradient Boosting. Additionally, their results emphasized the advantages of employing a set of sub- reduced-order models to improve prediction accuracy across diverse scenarios.

Upon reviewing the available literature, there is a noticeable scarcity of ML-related research devoted to the examination of wellbore integrity within the post-production phase of CO_2 flooding projects compared to the post-operational phase of pure GCS projects. For CO_2-flooded depleted hydrocarbon reservoirs that have completed their production lifecycles or CO_2-flooded depleting hydrocarbon reservoirs actively operating which serve as potential storage sites, a comprehensive appraisal of wellbore integrity is imperative. This assessment must encompass not only abandoned wells but also extend to currently operational injection and production wells, which attribute to evaluating leakage risks and determining uncertainties associated with all drilling and production activities in these brownfields. In situations where historical information on well drilling, completion, operation, maintenance, P&A schemes and the impact of CO_2-induced corrosion on the structural integrity of the cement-casing-formation composite remain undisclosed, the direct inheriting of existing infrastructures for storage purpose can significantly heighten the likelihood of CO_2 exposure. Thus, it is evident that further research in this area is warranted.

7 Conclusions

Currently, the utilization of subsurface for CO_2 storage, including both GCS and GCSU, is still in the research and demonstration phase. It has been barely three decades since the commencement of world's first GCS project, Sleipner, with its long-term environmental impacts yet to be seen. Underground storage has faced opposition particularly from groups like Greenpeace (Rochon et al., 2008), who argue that subsurface storage technology may not be enough mature to combat the climate change due to the main concerns regarding potential CO_2 leakage from proposed storage sites. Additionally, in practical engineering applications, criticisms have arisen as "too expensive" and "unproven." Inevitably, various stochastic variables such as carbon prices/taxes, oil prices, reservoir characteristics, fluid properties, and THMC coupling physics are inherent in this context. The oscillation and scarcity of database and insufficiency of understanding of underlying mechanisms escalate the computational costs of dynamic simulations by introducing uncertainty and complexity into decision-making reservoir models. Relying on traditional modeling tools based on highly ill-posed PDEs will hinder the optimal design of GCS system operations due to their limitations in speed and reliability. In response to these challenges, this paper investigates the entire fast decision-making process for CO_2 underground storage from the perspective of reservoir engineering. The main findings are summarized as follows:

1) The main technical barriers surfaced as a result of the complexities of modeling and simulating CO_2 behavior in subsurface, particularly the multiphase, multicomponent, and multiscale interactions. Additionally, the economic and sustainability dimensions of such projects require a delicate equilibrium, yet current models tend to lack comprehensive consideration of all aspects.
2) To address these barriers, state-of-the-art attempts have focused on three main step-by-step orientations: (a) advancing cost-effective gridding algorithms, numerical solvers, and hardware architecture; (b) data-driven proxy techniques and (c) multi-objective optimization techniques for solving non-linear, global, and multidimensional optimization problems associated with CO_2 storage reservoir management.
3) An innovative modeling-optimization paradigm integrating the above-mentioned techniques offers a general solution for the diverse landscape of decision-oriented CO_2 injection implementation. Our enthusiasm for this paradigm lies in its possibility to complement, modernize, and even revolutionize the traditional reservoir management, empowering decision-makers with not only fast and accurate solutions but adaptive and intelligent tools. It outlines three key components: problem formulation, SPM training, and optimizer development. The optimizer and SPM are coupled in an auto-updated pattern, with the SPM continuously calibrated by the new data obtained from real-time monitoring or optimization iterations. However, even the advanced paradigm will confront limitations when exposed to the high-complexity real-world problems, calling for fine-tuned strategies to unlock the full function of intelligent proxies in decision support.
4) An extensive survey of available cases was carried out to justify the paradigm, with the focus on specific challenges at different phases of GCS/GCSU projects, from pre-operational data preparation, operational plume monitoring to post-operational leakage prevention. The application survey demonstrated the deficiency of the real-world practice, particularly in the area of GCS decision making with respect to storage performance optimizations (capacity, injectivity, trapping, and confinement) and wellbore integrity assessment during the post-production phase of CO_2 flooding projects. We hope that this survey can convey some insights to both the academia and industry to contribute to filling in the gaps of CGS/CGUS practice through the proposed paradigm.

8 Recommendations

The GCS system presents a complex decision-making problem involving multi-components, phases, scales, and physical phenomenon, which is computationally intensive. When we compare the paradigm introduced in Section 5 with the case studies in Section 6, it becomes clear that current research tends to simplify this complex problem to varying degrees. This tendency is likely the reason why we find in most of current studies that applying simplified versions of the paradigm can also yield satisfactory results. Given the existing gap between state-of-the-art techniques and innovative implementation, the integration of AI&ML with GCS/GCSU as well as the transition from innovative research work to real-world practice is still in progress. Considering the urgency of the energy transition, it's crucial to broaden our horizons beyond the discipline of reservoir engineering. Instead, we should amalgamate the forefront progress spanning disciplines such as Applied Mathematics/Statistics, Data Science, Computer Engineering, Petroleum Engineering, Drilling and Production Engineering to continually push the boundaries of the proposed paradigm. This will enable real-time monitoring of CO_2 plume migration and early prediction of potential CO_2 leakage from the formation, ultimately advancing decision-making process for CO_2 storage initiatives.

In the first phase of this paradigm, which constitutes the problem formulation phase, there remain numerous research blanks to be filled, as outlined in Section 7.1. Our objective is to establish efficient decision-making processes for field-scale GCS systems, particularly under the more intricate conditions encountered in real-world scenarios. As the fundamental cornerstone of ML-based proxy model training, the second phase involves the collection of numerical simulation data. In the future, concerted efforts should be directed towards optimizing the computational aspects of numerical simulations in parallel with the advancements in computer hardware. This optimization can be achieved through high-performance parallelization strategies based on parallel numerical solvers.

While training proxy models, relying solely on implicit mapping relationships between the inputs and outputs of numerical simulations can introduce significant discrepancies due to the cascading of errors inherent in the simulations. To devise more comprehensive and rationalized intelligent proxies, it is important to amalgamate traditional reservoir flow dynamics, the principles of mass, energy, and momentum conservation, as well as the wealth of field experience with ML algorithms. Encouraging endeavors like HPDNN/PINN (Wang et al., 2023) have already introduced the loss of PDE system into the ML implementation, demonstrating notable improvements in accuracy and reduced demands for extensive training samples. Additionally, there have been a multitude of initiatives aiming to implement Transfer Learning techniques within the domain of CO_2 sequestration, expanding the potential of pre-trained NN models without necessitating extensive data accumulation. The subsequent stride entails harnessing Transfer Learning to enhance the resolution, account for heterogeneity, and encompass broader spatial and temporal scales within the generalized SPM. This extension of knowledge can even encompass subsurface reservoirs subjected to CO_2 injection and extend to hydrogen, natural gas, and geothermal reservoirs, thereby acquiring a versatile solution for energy storage project with minimal computational expenses for modeling, effectively contributing to the attainment of energy transition objectives. In the concluding phase which involves optimization of decision-making procedure, the new generation of SPM-based optimization environment will offer self-adaptive and automatic capabilities when managing versatile uncertainties, objectives, dimensions, constraints, and variable number and types. These optimal-seeking solutions will also incorporate innovative dynamic navigation strategy to overcome the challenges posed by local extrema.

Acknowledgement

This research is a part of BRU21 – NTNU Research and Innovation Program on Digital Automation Solutions for the Oil and Gas Industry (www.ntnu.edu/bru21) and Gemini Centre - CO_2 Enhanced Oil Recovery & Storage (CEORS) (www.ntnu.edu/ceors).

Reference

[1] Masson-Delmotte, V., Zhai, P., Pörtner, H. O., Roberts, D., Skea, J., & Shukla, P. R. (2022). Global Warming of 1.5° C: IPCC Special Report on Impacts of Global Warming of 1.5° C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty. Cambridge University Press.

[2] IEA (2022). Global Energy Review: CO2 Emissions in 2021. IEA. https://www.iea.org/reports/global-energy-review-co2-emissions-in-2021-2.

[3] Fragkos, P., van Soest, H. L., Schaeffer, R., Reedman, L., Köberle, A. C., Macaluso, N., ... & Iyer, G. (2021). Energy system transitions and low-carbon pathways in Australia, Brazil, Canada, China, EU-28, India, Indonesia, Japan, Republic of Korea, Russia and the United States. Energy, 216, 119385.

[4] Mohaghegh, S. (2018). Data-driven analytics for the geological storage of CO2. CRC Press.

[5] Zubarev, D. I. (2009, October). Pros and cons of applying proxy-models as a substitute for full reservoir simulations. In SPE Annual Technical Conference and Exhibition. OnePetro.

[6] Ng, C. S. W., Nait Amar, M., Jahanbani Ghahfarokhi, A., & Imsland, L. S. (2022). A Survey on the Application of Machine Learning and Metaheuristic Algorithms for Intelligent Proxy Modeling in Reservoir Simulation. Computers & Chemical Engineering, 108107.

[7] Yao, P., Yu, Z., Zhang, Y., & Xu, T. (2023). Application of machine learning in carbon capture and storage: An in-depth insight from the perspective of geoscience. Fuel, 333, 126296.

[8] Mikhaylov, A., Moiseev, N., Aleshin, K., & Burkhardt, T. (2020). Global climate change and greenhouse effect. Entrepreneurship and Sustainability Issues, 7(4), 2897.

[9] Xiao, T., McPherson, B., Pan, F., Esser, R., Jia, W., Bordelon, A., & Bacon, D. (2016). Potential chemical impacts of CO2 leakage on underground source of drinking water assessed by quantitative risk analysis. International Journal of Greenhouse Gas Control, 50, 305-316.

[10] Ko, D., Yoo, G., Yun, S. T., & Chung, H. (2016). Impacts of CO2 leakage on plants and microorganisms: A review of results from CO2 release experiments and storage sites. Greenhouse Gases: Science and Technology, 6(3), 319-338.

[11] Damen, K., Faaij, A., & Turkenburg, W. (2006). Health, safety and environmental risks of underground CO2 storage–overview of mechanisms and current knowledge. Climatic Change, 74(1-3), 289-318.

[12] Bachu, S., & Bennion, D. B. (2009). Experimental assessment of brine and/or CO2 leakage through well cements at reservoir conditions. International Journal of Greenhouse Gas Control, 3(4), 494-501.

[13] Lie, K. A., Nilsen, H. M., Andersen, O., & Møyner, O. (2016). A simulation workflow for large-scale CO2 storage in the Norwegian North Sea. Computational Geosciences, 20(3), 607-622.

[14] Benson, S. A. (2007). Geological Storage of CO2: Analogues and Risk Management" presentation to Carbon Sequestration Leadership Forum-CSLF, 7 May 2007. Pittsburgh PA.

[15] Zapata, Y., Kristensen, M. R., Huerta, N., Brown, C., Kabir, C. S., & Reza, Z. (2020). CO2 geological storage: Critical insights on plume dynamics and storage efficiency during long-term injection and post-injection periods. Journal of Natural Gas Science and Engineering, 83, 103542.

[16] Rubin, E., & De Coninck, H. (2005). IPCC special report on carbon dioxide capture and storage. UK: Cambridge University Press. TNO (2004): Cost Curves for CO2 Storage, Part, 2, 14.

[17] Raza, A., Rezaee, R., Gholami, R., Bing, C. H., Nagarajan, R., & Hamid, M. A. (2016). A screening criterion for selection of suitable CO2 storage sites. Journal of Natural Gas Science and Engineering, 28, 317-327.

[18] Hao, Y., & Yang, D. (2012). Research progress of carbon dioxide capture and geological sequestration problem and seismic monitoring research: Progress in Geophysics.

[19] Harp, D. R., Oldenburg, C. M., & Pawar, R. (2019). A metric for evaluating conformance robustness during geologic CO2 sequestration operations. International Journal of Greenhouse Gas Control, 85, 100-108.

[20] Arts, R. J., Chadwick, A., Eiken, O., Thibeau, S., & Nooner, S. (2008). Ten years' experience of monitoring CO2 injection in the Utsira Sand at Sleipner, offshore Norway. First break, 26(1).

[21] Jenkins, C. R., Cook, P. J., Ennis-King, J., Undershultz, J., Boreham, C., Dance, T., ... & Urosevic, M. (2012). Safe storage and effective monitoring of CO2 in depleted gas fields. Proceedings of the National Academy of Sciences, 109(2), E35-E41.

[22] Würdemann, H., Möller, F., Kühn, M., Heidug, W., Christensen, N. P., Borm, G., ... & CO2Sink Group. (2010). CO2SINK—From site characterisation and risk assessment to monitoring and verification: One year of operational experience with the field laboratory for CO2 storage at Ketzin, Germany. International Journal of Greenhouse Gas Control, 4(6), 938-951.

[23] Hovorka, S. D., Benson, S. M., Doughty, C., Freifeld, B. M., Sakurai, S., Daley, T. M., ... & Knauss, K. G. (2006). Measuring permanence of CO2 storage in saline formations: the Frio experiment. Environmental Geosciences, 13(2), 105-121.

[24] Harbert, W., Daley, T. M., Bromhal, G., Sullivan, C., & Huang, L. (2016). Progress in monitoring strategies for risk reduction in geologic CO2 storage. International Journal of Greenhouse Gas Control, 51, 260-275.

[25] Kutchko, B. G., Strazisar, B. R., Dzombak, D. A., Lowry, G. V., & Thaulow, N. (2007). Degradation of well cement by CO2 under geologic sequestration conditions. Environmental science & technology, 41(13), 4787-4792.

[26] Ramírez, A., Hoogwijk, M., Hendriks, C., & Faaij, A. (2008). Using a participatory approach to develop a sustainability framework for carbon capture and storage systems in The Netherlands. international journal of greenhouse gas control, 2(1), 136-154.

[27] Kolditz, O., Görke, U. J., Shao, H., & Wang, W. (2012). Thermo-hydro-mechanical-chemical processes in porous media: benchmarks and examples (lecture notes in computational science and engineering).

[28] Peter, A., Yang, D., Eshiet, K. I. I. I., & Sheng, Y. (2022). A review of the studies on CO2–brine–rock interaction in geological storage process. Geosciences, 12(4), 168.

[29] Ajayi, T., & Gupta, I. (2019). A review of reactive transport modeling in wellbore integrity problems. Journal of Petroleum Science and Engineering, 175, 785-803.

[30] Jung, Y., Pau, G. S. H., Finsterle, S., & Pollyea, R. M. (2017). TOUGH3: A new efficient version of the TOUGH suite of multiphase flow and transport simulators. Computers & Geosciences, 108, 2-7.

[31] Ma, T., Rutqvist, J., Oldenburg, C. M., & Liu, W. (2017). Coupled thermal–hydrological–mechanical modeling of CO2-enhanced coalbed methane recovery. International Journal of Coal Geology, 179, 81-91.

[32] Xu, T., Sonnenthal, E., Spycher, N., & Pruess, K. (2006). TOUGHREACT—a simulation program for non-isothermal multiphase reactive geochemical transport in variably saturated geologic media: applications to geothermal injectivity and CO2 geological sequestration. Computers & geosciences, 32(2), 145-165.

[33] Lu, C., & Lichtner, P. C. (2007, July). High resolution numerical investigation on the effect of convective instability on long term CO2 storage in saline aquifers. In Journal of Physics: Conference Series (Vol. 78, No. 1, p. 012042). IOP Publishing.

[34] Ebigbo, A., Bielinski, A., Kopp, A., Class, H. & Helmig, R. (2006). Numerical modeling of CO2 sequestration with MUFTE-UG. CMWR XVI, Copenhagen.

[35] Nguyen, B. N., Hou, Z., Bacon, D. H., Murray, C. J., & White, M. D. (2016). Three-dimensional modeling of the reactive transport of CO2 and its impact on geomechanical properties of reservoir rocks and seals. International Journal of Greenhouse Gas Control, 46, 100-115.

[36] Nilsen, H. M., Lie, K. A., & Andersen, O. (2015). Analysis of CO2 trapping capacities and long-term migration for geological formations in the Norwegian North Sea using MRST-co2lab. Computers & Geosciences, 79, 15-26.

[37] Sandve, T. H., Rasmussen, A., & Rustad, A. B. (2018, October). Open reservoir simulator for CO2 storage and CO2-EOR. In 14th Greenhouse Gas Control Technologies Conference Melbourne (pp. 21-26).

[38] Fu, P., Settgast, R. R., Johnson, S. M., Walsh, S. D., Morris, J. P., & Ryerson, F. J. (2014). GEOS. User Tutorials (No. LLNL-TR-665515). Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States).

[39] Ranganathan, P., van Hemert, P., Rudolph, E. S. J., & Zitha, P. Z. (2011). Numerical modeling of CO2 mineralisation during storage in deep saline aquifers. Energy Procedia, 4, 4538-4545.

[40] Class, H., Ebigbo, A., Helmig, R., Dahle, H. K., Nordbotten, J. M., Celia, M. A., ... & Wei, L. (2009). A benchmark study on problems related to CO2 storage in geologic formations: summary and discussion of the results. Computational geosciences, 13, 409-434.

[41] Hassani, H., Shahbazi, A., Shahbalayev, E., Hamdi, Z., Behjat, S., & Bataee, M. (2024, April). Machine Learning-Based CO2 Saturation Tracking in Saline Aquifers Using Bottomhole Pressure for Carbon Capture and Storage CCS Projects. In SPE Norway Subsurface Conference? (p. D011S009R002). SPE.

[42] Lu, J., Wilkinson, M., Haszeldine, R. S., & Boyce, A. J. (2011). Carbonate cements in Miller field of the UK North Sea: a natural analog for mineral trapping in CO 2 geological storage. Environmental Earth Sciences, 62, 507-517.

[43] Jiang, X. (2011). A review of physical modelling and numerical simulation of long-term geological storage of CO2. Applied energy, 88(11), 3557-3566.

[44] Farshidi, S. F. (2016). Compositional Reservoir Simulation-Based Reactive-Transport Formulations, with Application to CO 2 Storage in Sandstone and Ultramafic Formations. Stanford University.

[45] Zhao, L., Feng, C., Zhang, C. S., & Shu, S. (2022). Parallel multi-stage preconditioners with adaptive setup for the black oil model. Computers & Geosciences, 168, 105230.

[46] Jackson, M. D., Percival, J. R., Mostaghimi, P., Tollit, B. S., Pavlidis, D., Pain, C. C., ... & Blunt, M. J. (2015). Reservoir modeling for flow simulation by use of surfaces, adaptive unstructured meshes, and an overlapping-control-volume finite-element method. SPE Reservoir Evaluation & Engineering, 18(02), 115-132.

[47] Rubin, E. S., Davison, J. E., & Herzog, H. J. (2015). The cost of CO2 capture and storage. International Journal of Greenhouse gas control, 40, 378-400.

[48] Zero Emissions Platform (ZEP). (2011, July). The Costs of CO2 Capture, Transport and Storage. Post-demonstration CCS in the EU. http://www.zeroemissionsplatform.eu/downloads/887.html.

[49] EPA (United States Environmental Protection Agency). (2008). Geologic CO2 Sequestration Technology and Cost Analysis (EPA 816-B-08-009). https://www.epa.gov/sites/default/files/2015-07/documents/support_uic_co2_technologyandcostanalysis.pdf.

[50] Pham, V. T. H., Riis, F., Gjeldvik, I. T., Halland, E. K., Tappel, I. M., & Aagaard, P. (2013). Assessment of CO2 injection into the south Utsira-Skade aquifer, the North Sea, Norway. Energy, 55, 529-540.

[51] Balch, R., & McPherson, B. (2016, May). Integrating enhanced oil recovery and carbon capture and storage projects: a case study at farnsworth field, Texas. In SPE Western Regional Meeting. OnePetro.

[52] Melzer, L. S. (2012). Factors involved in adding carbon capture, utilization and storage (CCUS) to Enhanced Oil Recovery. Melzer consulting, 1-18.

[53] Mao, J., & Jahanbani Ghahfarokhi, A. (2023, June). Impact of Uncertainties and Decision Variables on CO2 Enhanced Oil Recovery and Storage: A Numerical Investigation. In 84th EAGE Annual Conference & Exhibition. European Association of Geoscientists & Engineers.

[54] Syed, F. I., Muther, T., Van, V. P., Dahaghi, A. K., & Negahban, S. (2022). Numerical trend analysis for factors affecting EOR performance and CO2 storage in tight oil reservoirs. Fuel, 316, 123370.

[55] Kamashev, A., & Amanbek, Y. (2021). Reservoir simulation of co2 storage using compositional flow model for geological formations in frio field and precaspian basin. Energies, 14(23), 8023.

[56] Suriano, A., Peter, C., Benetatos, C., & Verga, F. (2022). Gridding Effects on CO2 Trapping in Deep Saline Aquifers. Sustainability, 14(22), 15049.

[57] Zhang, X., Ma, F., Yin, S., Wallace, C. D., Soltanian, M. R., Dai, Z., ... & Lü, X. (2021). Application of upscaling methods for fluid flow and mass transport in multi-scale heterogeneous media: A critical review. Applied energy, 303, 117603.

[58] Iogna, A., Guillet-Lhermite, J., Wood, C., & Deflandre, J. P. (2017, June). CO2 storage and enhanced gas recovery: Using extended black oil modelling to simulate CO2 injection on a North Sea depleted gas field. In SPE Europec featured at 79th EAGE Conference and Exhibition. OnePetro.

[59] Krishnamurthy, P. G., Senthilnathan, S., Yoon, H., Thomassen, D., Meckel, T., & DiCarlo, D. (2017). Comparison of Darcy's law and invasion percolation simulations with buoyancy-driven CO2-brine multiphase flow in a heterogeneous sandstone core. Journal of Petroleum Science and Engineering, 155, 54-62.

[60] Akai, T., Kuriyama, T., Kato, S., & Okabe, H. (2021). Numerical modelling of long-term CO2 storage mechanisms in saline aquifers using the Sleipner benchmark dataset. International Journal of Greenhouse Gas Control, 110, 103405.

[61] Park, J., Iino, A., Datta-Gupta, A., Bi, J., & Sankaran, S. (2019, July). Rapid modeling of injection and production phases of hydraulically fractured shale wells using the fast marching method. In SPE/AAPG/SEG Unconventional Resources Technology Conference (p. D023S052R004). URTEC.

[62] Ahusborde, E., Amaziane, B., & Moulay, M. I. (2021). High performance computing of 3D reactive multiphase flow in porous media: Application to geological storage of CO2. Computational Geosciences, 25(6), 2131-2147.

[63] Voskov, D. V. (2017). Operator-based linearization approach for modeling of multiphase multi-component flow in porous media. Journal of Computational Physics, 337, 275-288.

[64] Khait, M., & Voskov, D. V. (2017). Operator-based linearization for general purpose reservoir simulation. Journal of Petroleum Science and Engineering, 157, 990-998.

[65] Li, L., & Abushaikha, A. (2021). A fully-implicit parallel framework for complex reservoir simulation with mimetic finite difference discretization and operator-based linearization. Computational Geosciences, 1-17.

[66] Cai, Z., Zhang, K., & Guo, C. (2022). Development of a novel simulator for modelling underground hydrogen and gas mixture storage. International Journal of Hydrogen Energy, 47(14), 8929-8942.

[67] Gross, H., & Mazuyer, A. (2021, October). GEOSX: A multiphysics, multilevel simulator designed for exascale computing. In SPE Reservoir Simulation Conference? (p. D011S010R007). SPE.

[68] Zhou, Y. (2012). Parallel general-purpose reservoir simulation with coupled reservoir models and multisegment wells (Doctoral dissertation, Stanford University).

[69] Gu, J., Liu, W., Zhang, K., Zhai, L., Zhang, Y., & Chen, F. (2021). Reservoir production optimization based on surrogate model and differential evolution algorithm. Journal of Petroleum Science and Engineering, 205, 108879.

[70] Bahrami, P., Sahari Moghaddam, F., & James, L. A. (2022). A Review of Proxy Modeling Highlighting Applications for Reservoir Engineering. Energies, 15(14), 5247.

[71] Sondergaard, P. (2011, October). Speech presented at Gartner Symposium/ITxpo, Orlando, Florida.

[72] Mohaghegh, S. D. (2011). Reservoir simulation and modeling based on artificial intelligence and data mining (AI&DM). Journal of Natural Gas Science and Engineering, 3(6), 697-705.

[73] Mohaghegh, S. D. (2020). Subsurface analytics: Contribution of artificial intelligence and machine learning to reservoir engineering, reservoir modeling, and reservoir management.

[74] Mohaghegh, S. D., Abdulla, F., Abdou, M., Gaskari, R., & Maysami, M. (2015, November). Smart proxy: an innovative reservoir management tool; Case study of a giant mature oilfield in the UAE. In Abu Dhabi international petroleum exhibition and conference. OnePetro.

[75] Leach, A., Mason, C. F., & van't Veld, K. (2011). Co-optimization of enhanced oil recovery and carbon sequestration. Resource and Energy Economics, 33(4), 893-912.

[76] Nwachukwu, A., Jeong, H., Pyrcz, M., & Lake, L. W. (2018a). Fast evaluation of well placements in heterogeneous reservoir models using machine learning. Journal of Petroleum Science and Engineering, 163, 463-475.

[77] Nait Amar, M., Hemmati-Sarapardeh, A., Varamesh, A., & Shamshirband, S. (2019). Predicting solubility of CO2 in brine by advanced machine learning systems: Application to carbon capture and sequestration. Journal of CO2 Utilization, 33, 83-95.

[78] Matthew, D. A. M., Jahanbani Ghahfarokhi, A., Ng, C. S. W., & Nait Amar, M. (2023). Proxy Model Development for the Optimization of Water Alternating CO2 Gas for Enhanced Oil Recovery. Energies, 16(8), 3337.

[79] Onwunalu, J., Litvak, M., Durlofsky, L. J., & Aziz, K. (2008, November). Application of statistical proxies to speed up field development optimization procedures. In Abu Dhabi international petroleum exhibition and conference (pp. SPE-117323). SPE.

[80] Pan, I., Babaei, M., Korre, A., & Durucan, S. (2014). Artificial Neural Network based surrogate modelling for multi-objective optimisation of geological CO2 storage operations. Energy Procedia, 63, 3483-3491.

[81] Zheng, F., Jahandideh, A., Jha, B., & Jafarpour, B. (2021). Geologic CO2 storage optimization under geomechanical risk using coupled-physics models. International Journal of Greenhouse Gas Control, 110, 103385.

[82] You, J., Ampomah, W., & Sun, Q. (2020). Development and application of a machine learning based multi-objective optimization workflow for CO2-EOR projects. Fuel, 264, 116758.

[83] Liu, Y., Zhu, N., & Li, M. (2020). Solving many-objective optimization problems by a Pareto-based evolutionary algorithm with preprocessing and a penalty mechanism. IEEE transactions on cybernetics, 51(11), 5585-5594.

[84] Capolei, A., Suwartadi, E., Foss, B., & Jørgensen, J. B. (2015). A mean–variance objective for robust production optimization in uncertain geological scenarios. Journal of Petroleum Science and Engineering, 125, 23-37.

[85] Ampomah, W., Balch, R. S., Cather, M., Will, R., Gunda, D., Dai, Z., & Soltanian, M. R. (2017). Optimum design of CO2 storage and oil recovery under geological uncertainty. Applied energy, 195, 80-92.

[86] Nwachukwu, A., Jeong, H., Sun, A., Pyrcz, M., & Lake, L. W. (2018b, April). Machine learning-based optimization of well locations and WAG parameters under geologic uncertainty. In SPE improved oil recovery conference. OnePetro.

[87] Fonseca, C. M., & Fleming, P. J. (1998). Multiobjective optimization and multiple constraint handling with evolutionary algorithms. I. A unified formulation. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 28(1), 26-37.

[88] Kramer, O. (2010). A review of constraint-handling techniques for evolution strategies. Applied Computational Intelligence and Soft Computing, 2010, 1-19.

[89] Volkov, O., & Bellout, M. C. (2018). Gradient-based constrained well placement optimization. Journal of Petroleum Science and Engineering, 171, 1052-1066.

[90] Salehian, M., Sefat, M. H., & Muradov, K. (2021). A multisolution optimization framework for well placement and control. SPE Reservoir Evaluation & Engineering, 24(04), 923-939.

[91] Zou, A., & Durlofsky, L. J. (2023). Integrated Framework for Constrained Optimization of Horizontal/Deviated Well Placement and Control for Geological CO2 Storage. SPE Journal, 1-20.

[92] Nguyen, Q. M., Onur, M., & Alpak, F. O. (2023). Multi-objective optimization of subsurface CO 2 capture, utilization, and storage using sequential quadratic programming with stochastic gradients. Computational Geosciences, 1-16.

[93] Cihan, A., Birkholzer, J. T., & Bianchi, M. (2015). Optimal well placement and brine extraction for pressure management during CO2 sequestration. International Journal of Greenhouse Gas Control, 42, 175-187.

[94] Nait Amar, M., Zeraibi, N., & Jahanbani Ghahfarokhi, A. (2020). Applying hybrid support vector regression and genetic algorithm to water alternating CO2 gas EOR. Greenhouse Gases: Science and Technology, 10(3), 613-630.

[95] Deb, K. (2000). An efficient constraint handling method for genetic algorithms. Computer methods in applied mechanics and engineering, 186(2-4), 311-338.

[96] AlQahtani, G., Vadapalli, R., Siddiqui, S., & Bhattacharya, S. (2012, April). Well optimization strategies in conventional reservoirs. In SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition (pp. SPE-160861). SPE.

[97] Miyagi, A., Akimoto, Y., & Yamamoto, H. (2018, July). Well placement optimization for carbon dioxide capture and storage via CMA-ES with mixed integer support. In Proceedings of the Genetic and Evolutionary Computation Conference Companion (pp. 1696-1703).

[98] Karkevandi-Talkhooncheh, A., Rostami, A., Hemmati-Sarapardeh, A., Ahmadi, M., Husein, M. M., & Dabir, B. (2018). Modeling minimum miscibility pressure during pure and impure CO2 flooding using hybrid of radial basis function neural network and evolutionary techniques. Fuel, 220, 270-282.

[99] Bansal, J. C., Sharma, H., Jadon, S. S., & Clerc, M. (2014). Spider monkey optimization algorithm for numerical optimization. Memetic computing, 6, 31-47.

[100] Kaveh, A., & Dadras, A. (2017). A novel meta-heuristic optimization algorithm: thermal exchange optimization. Advances in engineering software, 110, 69-84.

[101] Sun, P., Liu, H., Zhang, Y., Tu, L., & Meng, Q. (2021). An intensify atom search optimization for engineering design problems. Applied Mathematical Modelling, 89, 837-859.

[102] Niknam, T., Amiri, B., Olamaei, J., & Arefi, A. (2009). An efficient hybrid evolutionary optimization algorithm based on PSO and SA for clustering. Journal of Zhejiang University-SCIENCE A, 10, 512-519.

[103] Freisleben, B., & Merz, P. (1996, September). New genetic local search operators for the traveling salesman problem. In International Conference on Parallel Problem Solving from Nature (pp. 890-899). Berlin, Heidelberg: Springer Berlin Heidelberg.

[104] Nwankwor, E., Nagar, A. K., & Reid, D. C. (2013). Hybrid differential evolution and particle swarm optimization for optimal well placement. Computational Geosciences, 17, 249-268.

[105] Ding, S., Jiang, H., Li, J., & Tang, G. (2014). Optimization of well placement by combination of a modified particle swarm optimization algorithm and quality map method. Computational Geosciences, 18, 747-762.

[106] Aliyev, E. (2011). Use of hybrid approaches and metaoptimization for well placement problems (Doctoral dissertation, Stanford University).

[107] Isebor, O. J., Durlofsky, L. J., & Echeverría Ciaurri, D. (2014). A derivative-free methodology with local and global search for the constrained joint optimization of well locations and controls. Computational Geosciences, 18, 463-482.

[108] Humphries, T. D., Haynes, R. D., & James, L. A. (2014). Simultaneous and sequential approaches to joint optimization of well placement and control. Computational Geosciences, 18, 433-448.

[109] Zhang, Z., Wu, Y., & Wang, G. (2018). Bpgrad: Towards global optimality in deep learning via branch and pruning. In Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (pp. 3301-3309).

[110] Dehghani, S. M., Sefti, M. V., Ameri, A., & Kaveh, N. S. (2008). Minimum miscibility pressure prediction based on a hybrid neural genetic algorithm. chemical engineering research and design, 86(2), 173-185.

[111] Chen, S., Li, H., Yang, D., & Tontiwachwuthikul, P. (2010). Optimal parametric design for water-alternating-gas (WAG) process in a CO2-miscible flooding reservoir. Journal of Canadian petroleum technology, 49(10), 75-82.

[112] González-Nicolás, A., Cihan, A., Petrusak, R., Zhou, Q., Trautz, R., Riestenberg, D., ... & Birkholzer, J. T. (2019). Pressure management via brine extraction in geological CO2 storage: Adaptive optimization strategies under poorly characterized reservoir conditions. International Journal of Greenhouse Gas Control, 83, 176-185.

[113] Chen, M., Abdalla, O. A., Izady, A., Nikoo, M. R., & Al-Maktoumi, A. (2020). Development and surrogate-based calibration of a CO2 reservoir model. Journal of Hydrology, 586, 124798.

[114] Karkevandi-Talkhooncheh, A., Hajirezaie, S., Hemmati-Sarapardeh, A., Husein, M. M., Karan, K., & Sharifi, M. (2017). Application of adaptive neuro fuzzy interface system optimized with evolutionary algorithms for modeling CO2-crude oil minimum miscibility pressure. Fuel, 205, 34-45.

[115] Ng, C. S. W., & Jahanbani Ghahfarokhi, A. (2022). Adaptive proxy-based robust production optimization with multilayer perceptron. Applied Computing and Geosciences, 16, 100103.

[116] Ng, C. S. W., Jahanbani Ghahfarokhi, A., & Wiranda, W. (2023). Fast Well Control Optimization with Two-Stage Proxy Modeling. Energies, 16(7), 3269.

[117] Dai, Z., Viswanathan, H., Middleton, R., Pan, F., Ampomah, W., Yang, C., ... & White, M. (2016). CO2 accounting and risk analysis for CO2 sequestration at enhanced oil recovery sites. Environmental science & technology, 50(14), 7546-7554.

[118] Nielsen, L. C., Bourg, I. C., & Sposito, G. (2012). Predicting CO2–water interfacial tension under pressure and temperature conditions of geologic CO2 storage. Geochimica et Cosmochimica Acta, 81, 28-38.

[119] Li, D., & Jiang, X. (2020). Numerical investigation of convective mixing in impure CO2 geological storage into deep saline aquifers. International Journal of Greenhouse Gas Control, 96, 103015.

[120] Safi, R., Agarwal, R. K., & Banerjee, S. (2016). Numerical simulation and optimization of CO2 utilization for enhanced oil recovery from depleted reservoirs. Chemical Engineering Science, 144, 30-38.

[121] Rodrigues, H. W., Mackay, E. J., & Arnold, D. P. (2022). Multi-objective optimization of CO2 recycling operations for CCUS in pre-salt carbonate reservoirs. International Journal of Greenhouse Gas Control, 119, 103719.

[122] Allen, R., Nilsen, H. M., Andersen, O., & Lie, K. A. (2017). On obtaining optimal well rates and placement for CO2 storage. Computational Geosciences, 21(5-6), 1403-1422.

[123] Nghiem, L., Shrivastava, V., Kohse, B., Hassam, M., & Yang, C. (2010). Simulation and optimization of trapping processes for CO2 storage in saline aquifers. Journal of Canadian Petroleum Technology, 49(08), 15-22.

[124] Suwartadi, E., Krogstad, S., & Foss, B. (2009, October). On state constraints of adjoint optimization in oil reservoir water-flooding. In SPE/EAGE Reservoir Characterization & Simulation Conference (pp. cp-170). European Association of Geoscientists & Engineers.

[125] Giunta, A., Wojtkiewicz, S., & Eldred, M. (2003, January). Overview of modern design of experiments methods for computational simulations. In 41st Aerospace Sciences Meeting and Exhibit (p. 649).

[126] Halton, J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik, 2, 84-90.

[127] Sobol', I. Y. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 7(4), 784-802.

[128] Hammersley, J. M., & Handscomb, D. C. (1961). Monte Carlo Methods. In Proceedings (Vol. 7, p. 17). US Army Research Office.

[129] Joseph, V. R. (2016). Space-filling designs for computer experiments: A review. Quality Engineering, 28(1), 28-35.

[130] Gramacy, R. B. (2020). Surrogates: Gaussian process modeling, design, and optimization for the applied sciences. Chapman and Hall/CRC.

[131] Ahmadi, M. A., Zendehboudi, S., & James, L. A. (2018). Developing a robust proxy model of CO2 injection: Coupling Box–Behnken design and a connectionist method. Fuel, 215, 904-914.

[132] Vida, G., Shahab, M. D., & Mohammad, M. (2019). Smart proxy modeling of SACROC CO2-EOR. Fluids, 4(2), 85.

[133] Shahkarami, A., & Mohaghegh, S. (2020). Applications of smart proxies for subsurface modeling. Petroleum Exploration and Development, 47(2), 400-412.

[134] You, J., Ampomah, W., Tu, J., Morgan, A., Sun, Q., Wei, B., & Wang, D. (2022). Optimization of Water-Alternating-CO2 Injection Field Operations Using a Machine-Learning-Assisted Workflow. SPE Reservoir Evaluation & Engineering, 25(02), 214-231.

[135] Zhou, H., & Lascaud, B. (2019, October). An Integrated Machine Learning Framework for Optimizing Unconventional Resources Development. In Unconventional Resources Technology

Conference, Denver, Colorado, 22-24 July 2019 (pp. 3079-3087). Unconventional Resources Technology Conference (URTeC); Society of Exploration Geophysicists.

[136] Bird, G. D., Gorrell, S. E., & Salmon, J. L. (2021). Dimensionality-reduction-based surrogate models for real-time design space exploration of a jet engine compressor blade. Aerospace Science and Technology, 118, 107077.

[137] Nait Amar, M., Zeraibi, N., & Redouane, K. (2018). Bottom hole pressure estimation using hybridization neural networks and grey wolves optimization. Petroleum, 4(4), 419-429.

[138] Xu, K., Zhang, M., Li, J., Du, S. S., Kawarabayashi, K. I., & Jegelka, S. (2020). How neural networks extrapolate: From feedforward to graph neural networks. arXiv preprint arXiv:2009.11848.

[139] Ng, C. S. W., Jahanbani Ghahfarokhi, A., & Nait Amar, M. (2021). Application of nature-inspired algorithms and artificial neural network in waterflooding well control optimization. Journal of Petroleum Exploration and Production Technology, 11(7), 3103-3127.

[140] Kim, J., Choe, J., & Lee, K. (2022). Sequential field development plan through robust optimization coupling with CNN and LSTM-based proxy models. Journal of Petroleum Science and Engineering, 209, 109887.

[141] Yan, H., Zhang, J., Rahman, S. S., Zhou, N., & Suo, Y. (2020). Predicting permeability changes with injecting CO2 in coal seams during CO2 geological sequestration: A comparative study among six SVM-based hybrid models. Science of the Total Environment, 705, 135941.

[142] Mardhatillah, M. K., Md Yusof, M. A., Sa'id, A. A., Mohammad Fuad, I. I., Sokama Neuyam, Y. A., & Md Akhir, N. A. (2022, March). Predictive Modelling of Carbon Dioxide Injectivity Using SVR-Hybrid. In Offshore Technology Conference Asia. OnePetro.

[143] Abdolbaghi, S., Barati-Harooni, A., & Najafi-Marghmaleki, A. (2019). Improving the prediction ability of reference correlation for viscosity of carbon dioxide. Journal of CO2 Utilization, 31, 106-114.

[144] Nait Amar, M., Ghriga, M. A., Ouaer, H., Seghier, M. E. A. B., Pham, B. T., & Andersen, P. Ø. (2020a). Modeling viscosity of CO2 at high temperature and pressure conditions. Journal of Natural Gas Science and Engineering, 77, 103271.

[145] Nait Amar, M., Jahanbani Ghahfarokhi, A., & Zeraibi, N. (2020b). Predicting thermal conductivity of carbon dioxide using group of data-driven models. Journal of the Taiwan Institute of Chemical Engineers, 113, 165-177.

[146] Talebi, S., Gohari, M., & Mirmohamadi, A. (2023). Enhancement of CO2 viscosity prediction using advanced intelligent methods: Application to carbon capture and storage. Geoenergy Science and Engineering, 225, 211727.

[147] Dehaghani, A. H. S., & Soleimani, R. (2019). Estimation of interfacial tension for geological CO2 storage. Chemical Engineering & Technology, 42(3), 680-689.

[148] Safaei-Farouji, M., Thanh, H. V., Dashtgoli, D. S., Yasin, Q., Radwan, A. E., Ashraf, U., & Lee, K. K. (2022). Application of robust intelligent schemes for accurate modelling interfacial tension of CO2 brine systems: Implications for structural CO2 trapping. Fuel, 319, 123821.

[149] Tariq, Z., Ali, M., Yan, B., Sun, S., Khan, M., Yekeen, N., & Hoteit, H. (2023, March). Data-driven machine learning modeling of mineral/co2/brine wettability prediction: Implications for CO2 geo-storage. In Middle East Oil, Gas and Geosciences Show. OnePetro.

[150] Nait Amar, M., Hemmati-Sarapardeh, A., Varamesh, A., & Shamshirband, S. (2019). Predicting solubility of CO2 in brine by advanced machine learning systems: Application to carbon capture and sequestration. Journal of CO2 Utilization, 33, 83-95.

[151] Mohammadian, E., Liu, B., & Riazi, A. (2022). Evaluation of different machine learning frameworks to estimate CO2 solubility in NaCl brines: implications for CO2 injection into low-salinity formations. Lithosphere, 2022(Special 12), 1615832.

[152] Ratnakar, R. R., Chaubey, V., & Dindoruk, B. (2023). A novel computational strategy to estimate CO2 solubility in brine solutions for CCUS applications. Applied Energy, 342, 121134.

[153] Feng, Q., Cui, R., Wang, S., Zhang, J., & Jiang, Z. (2019). Estimation of CO2 diffusivity in brine by use of the genetic algorithm and mixed kernels-based support vector machine model. Journal of Energy Resources Technology, 141(4), 041001.

[154] Bemani, A., Baghban, A., & Mosavi, A. (2020). Estimating CO2-Brine diffusivity using hybrid models of ANFIS and evolutionary algorithms. Engineering Applications of Computational Fluid Mechanics, 14(1), 818-834.

[155] Nait Amar, M., & Jahanbani Ghahfarokhi, A. (2020). Prediction of CO2 diffusivity in brine using white-box machine learning. Journal of Petroleum Science and Engineering, 190, 107037.

[156] Ahmmed, B., Karra, S., Vesselinov, V. V., & Mudunuru, M. K. (2021). Machine learning to discover mineral trapping signatures due to CO2 injection. International Journal of Greenhouse Gas Control, 109, 103382.

[157] Tariq, Z., Yildirim, E. U., Gudala, M., Yan, B., Sun, S., & Hoteit, H. (2023). Spatial–temporal prediction of minerals dissolution and precipitation using deep learning techniques: An implication to Geological Carbon Sequestration. Fuel, 341, 127677.

[158] Alvarado, V., & Manrique, E. (2010). Enhanced oil recovery: an update review. Energies, 3(9), 1529-1575.

[159] Moradkhani, M. A., Hosseini, S. H., & Ranjbar, K. (2023). Universal intelligent models for liquid density of CO2+ hydrocarbon mixtures. Fuel, 334, 126642.

[160] Ghiasi, M. M., Mohammadi, A. H., & Zendehboudi, S. (2021). Use of hybrid-ANFIS and ensemble methods to calculate minimum miscibility pressure of CO2-reservoir oil system in miscible flooding process. Journal of Molecular Liquids, 331, 115369.

[161] Lv, Q., Zheng, R., Guo, X., Larestani, A., Hadavimoghaddam, F., Riazi, M., ... & Li, J. (2023). Modelling minimum miscibility pressure of CO2-crude oil systems using deep learning, tree-based, and thermodynamic models: Application to CO2 sequestration and enhanced oil recovery. Separation and Purification Technology, 310, 123086.

[162] Huang, C., Tian, L., Zhang, T., Chen, J., Wu, J., Wang, H., ... & Zhang, K. (2022). Globally optimized machine-learning framework for CO2-hydrocarbon minimum miscibility pressure calculations. Fuel, 329, 125312.

[163] Bagalkot, N., & Keprate, A. (2021, June). Machine Learning Approach to Estimate the Diffusion Coefficient of CO2 in Hydrocarbons. In International Conference on Offshore Mechanics and Arctic Engineering (Vol. 85208, p. V010T11A036). American Society of Mechanical Engineers.

[164] Salehi, E., Mohammadi, M. R., Hemmati-Sarapardeh, A., Mahdavi, V. R., Gentzis, T., Liu, B., & Ostadhassan, M. (2022). Modeling interfacial tension of N2/CO2 mixture+ n-alkanes with machine learning methods: application to eor in conventional and unconventional reservoirs by flue gas injection. Minerals, 12(2), 252.

[165] Jiang, F., Yang, J., Boek, E., & Tsuji, T. (2021). Investigation of viscous coupling effects in three-phase flow by lattice Boltzmann direct simulation and machine learning technique. Advances in Water Resources, 147, 103797.

[166] Nait Amar, M., Shateri, M., Hemmati-Sarapardeh, A., & Alamatsaz, A. (2019). Modeling oil-brine interfacial tension at high pressure and high salinity conditions. Journal of Petroleum Science and Engineering, 183, 106413.

[167] Kirch, A., Celaschi, Y. M., de Almeida, J. M., & Miranda, C. R. (2020). Brine–oil interfacial tension modeling: assessment of machine learning techniques combined with molecular dynamics. ACS applied materials & interfaces, 12(13), 15837-15843.

[168] Bemani, A., Baghban, A., Mohammadi, A. H., & Andersen, P. Ø. (2020). Estimation of adsorption capacity of CO2, CH4, and their binary mixtures in Quidam shale using LSSVM: Application in CO2 enhanced shale gas recovery and CO2 storage. Journal of Natural Gas Science and Engineering, 76, 103204.

[169] Nait Amar, M., Larestani, A., Lv, Q., Zhou, T., & Hemmati-Sarapardeh, A. (2022). Modeling of methane adsorption capacity in shale gas formations using white-box supervised machine learning techniques. Journal of Petroleum Science and Engineering, 208, 109226.

[170] Wu, H., Lubbers, N., Viswanathan, H. S., & Pollyea, R. M. (2021). A multi-dimensional parametric study of variability in multi-phase flow dynamics during geologic CO2 sequestration accelerated with machine learning. Applied Energy, 287, 116580.

[171] Alali, A. (2023, March). Modeling CO2 Geologic Storage Using Machine Learning. In SPE Middle East Oil and Gas Show and Conference (p. D021S066R004). SPE.

[172] Davoodi, S., Thanh, H. V., Wood, D. A., Mehrad, M., Rukavishnikov, V. S., & Dai, Z. (2023). Machine-learning predictions of solubility and residual trapping indexes of carbon dioxide from global geological storage sites. Expert Systems with Applications, 222, 119796.

[173] Al-Qaness, M. A., Ewees, A. A., Thanh, H. V., AlRassas, A. M., Dahou, A., & Elaziz, M. A. (2023). Predicting CO2 trapping in deep saline aquifers using optimized long short-term memory. Environmental Science and Pollution Research, 30(12), 33780-33794.

[174] Mo, S., Zhu, Y., Zabaras, N., Shi, X., & Wu, J. (2019). Deep convolutional encoder‐decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. Water Resources Research, 55(1), 703-728.

[175] Zhong, Z., Sun, A. Y., & Jeong, H. (2019). Predicting CO2 plume migration in heterogeneous formations using conditional deep convolutional generative adversarial network. Water Resources Research, 55(7), 5830-5851.

[176] Sun, A. Y. (2020). Optimal carbon storage reservoir management through deep reinforcement learning. Applied Energy, 278, 115660.

[177] Wen, G., Tang, M., & Benson, S. M. (2021). Towards a predictor for CO2 plume migration using deep neural networks. International Journal of Greenhouse Gas Control, 105, 103223.

[178] Nagao, M., Yao, C., Onishi, T., Chen, H., & Datta-Gupta, A. (2022, September). An Efficient Deep Learning-Based Workflow for CO2 Plume Imaging Using Distributed Pressure and Temperature Measurements. In SPE Annual Technical Conference and Exhibition. OnePetro.

[179] Lin, J., Zhang, K., Zhang, L., Liu, P., Peng, W., Zhang, H., ... & Yao, J. (2024). Towards efficient and accurate CO2 sequestration optimization: Integrating hierarchical spatio-temporal information into deep-learning-based surrogate models. Fuel, 356, 129343.

[180] Chen, B., Harp, D. R., Lin, Y., Keating, E. H., & Pawar, R. J. (2018). Geologic CO2 sequestration monitoring design: A machine learning and uncertainty quantification-based approach. Applied energy, 225, 332-345.

[181] Tang, M., Ju, X., & Durlofsky, L. J. (2022). Deep-learning-based coupled flow-geomechanics surrogate model for CO2 sequestration. International Journal of Greenhouse Gas Control, 118, 103692.

[182] Tang, H., Fu, P., Sherman, C. S., Zhang, J., Ju, X., Hamon, F., ... & Morris, J. P. (2021). A deep learning-accelerated data assimilation and forecasting workflow for commercial-scale geologic carbon storage. International Journal of Greenhouse Gas Control, 112, 103488.

[183] Chen, B., Yan, B., Kang, Q., Harp, D., & Pawar, R. (2022). Deep Learning Accelerated Inverse Modeling and Forecasting for Large-Scale Geologic CO2 Sequestration. Available at SSRN 4283252.

[184] Han, Y., Hamon, F. P., Jiang, S., & Durlofsky, L. J. (2023). Surrogate Model for Geological CO2 Storage and Its Use in MCMC-based History Matching. arXiv preprint arXiv:2308.06341.

[185] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. arXiv preprint arXiv:1711.10561.

[186] Maniglio, M., Fighera, G., Dovera, L., & Stabile, C. C. (2021, December). Physics Informed Neural Networks Based on a Capacitance Resistance Model for Reservoirs Under Water Flooding Conditions. In Abu Dhabi International Petroleum Exhibition and Conference (p. D021S027R001). SPE.

[187] Tadjer, A., Hong, A., & Bratvold, R. (2022). Bayesian Deep Decline Curve Analysis: A New Approach for Well Oil Production Modeling and Forecasting. SPE Reservoir Evaluation & Engineering, 25(03), 568-582.

[188] Liu, J., Jing, H., & Pan, H. (2023, March). New fast simulation of 4d (x, y, z, t) co2 eor by fourier neural operator based deep learning method. In SPE Reservoir Simulation Conference? (p. D010S001R006). SPE.

[189] Gladchenko, E. S., Illarionov, E. A., Orlov, D. M., & Koroteev, D. A. (2023, January). Physics-informed neural networks and capacitance-resistance model: Fast and accurate oil and water production forecast using end-to-end architecture. In SPE Middle East Intelligent Oil and Gas Symposium (p. D021S004R001). SPE.

[190] Manasipov, R., Nikolaev, D., Didenko, D., Abdalla, R., & Stundner, M. (2023, January). Physics informed machine learning for production forecast. In SPE Reservoir Characterisation and Simulation Conference and Exhibition? (p. D021S007R003). SPE.

[191] Franklin, T. S., Souza, L. S., Fontes, R. M., & Martins, M. A. (2022). A physics-informed neural networks (PINN) oriented approach to flow metering in oil wells: an ESP lifted oil well system as a case study. Digital Chemical Engineering, 5, 100056.

[192] Almajid, M. M., & Abu-Al-Saud, M. O. (2022). Prediction of porous media fluid flow using physics informed neural networks. Journal of Petroleum Science and Engineering, 208, 109205.

[193] Zhang, J., Braga-Neto, U., & Gildin, E. (2024). Advancing Petroleum Engineering Solutions: Integrating Physics-Informed Neural Networks for Enhanced Buckley-Leverett Model Analysis. arXiv preprint arXiv:2404.12502.

[194] Fraces, C. G., Papaioannou, A., & Tchelepi, H. (2020). Physics informed deep learning for transport in porous media. Buckley Leverett problem. arXiv preprint arXiv:2001.05172.

[195] Du, H., Zhao, Z., Cheng, H., Yan, J., & He, Q. (2023). Modeling density-driven flow in porous media by physics-informed neural networks for CO2 sequestration. Computers and Geotechnics, 159, 105433.

[196] Wang, Y. W., Dai, Z. X., Wang, G. S., Chen, L., Xia, Y. Z., & Zhou, Y. H. (2023). A hybrid physics-informed data-driven neural network for CO2 storage in depleted shale reservoirs. Petroleum Science.

[197] You, J., Ampomah, W., Sun, Q., Kutsienyo, E. J., Balch, R. S., Dai, Z., ... & Zhang, X. (2020). Machine learning based co-optimization of carbon dioxide sequestration and oil recovery in CO2-EOR project. Journal of Cleaner Production, 260, 120866.

[198] Bocoum, A. O., & Rasaei, M. R. (2023). Multi-objective optimization of WAG injection using machine learning and data-driven Proxy models. Applied Energy, 349, 121593.

[199] Rongtao, L., Liao, X., Wang, X., Zhang, Y., Mu, L., Dong, P., & Tang, K. (2023). A multi-agent deep reinforcement learning method for co2 flooding rates optimization. Energy Exploration & Exploitation, 41(1), 224-245.

[200] Rezk, M. G., Ibrahim, A. F., & Adebayo, A. R. (2023). Uncertainty quantification for CO2 storage during intermittent CO2-EOR in oil reservoirs. International Journal of Coal Geology, 266, 104177.

[201] Hamdi, H., Clarkson, C. R., Esmail, A., & Costa Sousa, M. (2019, June). A bayesian approach for optimizing the huff-n-puff gas injection performance in shale reservoirs under parametric uncertainty: A duvernay shale example. In SPE Europec featured at 81st EAGE Conference and Exhibition. OnePetro.

[202] Enab, K. (2023, June). Machine Learning Based Approach to Optimize Huff-N-Puff Gas Injection in Naturally Fractured Shale Oil Reservoirs. In SPE/AAPG/SEG Unconventional Resources Technology Conference (p. D031S073R007). URTEC.

[203] Du, S., Wang, M., Yang, J., Zhao, Y., Wang, J., Yue, M., ... & Song, H. (2023). A novel prediction method for coalbed methane production capacity combined extreme gradient boosting with bayesian optimization. Computational Geosciences, 1-10.

[204] Omosebi, O. A., Oldenburg, C. M., & Reagan, M. (2022). Development of lean, efficient, and fast physics-framed deep-learning-based proxy models for subsurface carbon storage. International Journal of Greenhouse Gas Control, 114, 103562.

[205] Musayev, K., Shin, H., & Nguyen-Le, V. (2023). Optimization of CO2 injection and brine production well placement using a genetic algorithm and artificial neural network-based proxy model. International Journal of Greenhouse Gas Control, 127, 103915.

[206] Farajzadeh, R., Zaal, C., Van den Hoek, P., & Bruining, J. (2019). Life-cycle assessment of water injection into hydrocarbon reservoirs using exergy concept. Journal of Cleaner Production, 235, 812-821.

[207] Vaziri, P., & Sedaee, B. (2023). A machine learning‐based approach to the multiobjective optimization of CO2 injection and water production during CCS in a saline aquifer based on field data. Energy Science & Engineering.

[208] Middleton, R. S., Ogland-Hand, J. D., Chen, B., Bielicki, J. M., Ellett, K. M., Harp, D. R., & Kammer, R. M. (2020). Identifying geologic characteristics and operational decisions to meet global carbon sequestration goals. Energy & Environmental Science, 13(12), 5000-5016.

[209] Harp, D. R., Pawar, R., Carey, J. W., & Gable, C. W. (2016). Reduced order models of transient CO2 and brine leakage along abandoned wellbores from geologic carbon sequestration reservoirs. International Journal of Greenhouse Gas Control, 45, 150-162.

[210] Zhang, L., Dilmore, R., Huerta, N., Soong, Y., Vasylkivska, V., Namhata, A., ... & Li, X. (2018). Application of a new reduced‐complexity assessment tool to estimate CO2 and brine leakage from reservoir and above‐zone monitoring interval (AZMI) through an abandoned well under geologic carbon storage conditions. Greenhouse Gases: Science and Technology, 8(5), 839-853.

[211] Gan, M., Nguyen, M. C., Zhang, L., Wei, N., Li, J., Lei, H., ... & Stauffer, P. H. (2021). Impact of reservoir parameters and wellbore permeability uncertainties on CO2 and brine leakage potential at the Shenhua CO2 Storage Site, China. International Journal of Greenhouse Gas Control, 111, 103443.

[212] Baek, S., Bacon, D. H., & Huerta, N. J. (2023). Enabling site-specific well leakage risk estimation during geologic carbon sequestration using a modular deep-learning-based wellbore leakage model. International Journal of Greenhouse Gas Control, 126, 103903.

[213] He, X., Zhu, W., Santoso, R., Alsinan, M., Kwak, H., & Hoteit, H. (2021, September). CO2 Leakage Rate Forecasting Using Optimized Deep Learning. In SPE Annual Technical Conference and Exhibition? (p. D021S023R001). SPE.

[214] Raad, S. M. J., Lawton, D., Maidment, G., & Hassanzadeh, H. (2022). Sensitivity analysis of operating parameters affecting CO2 wellbore transient flow at the CaMI Field Research Station, Alberta, Canada. International Journal of Greenhouse Gas Control, 114, 103594.

[215] Meguerdijian, S., Pawar, R. J., Chen, B., Gable, C. W., Miller, T. A., & Jha, B. (2023). Physics-informed machine learning for fault-leakage reduced-order modeling. International Journal of Greenhouse Gas Control, 125, 103873.

[216] Li, B., Zhou, F., Li, H., Duguid, A., Que, L., Xue, Y., & Tan, Y. (2018). Prediction of CO2 leakage risk for wells in carbon sequestration fields with an optimal artificial neural network. International Journal of Greenhouse Gas Control, 68, 276-286.

[217] Ugarte, E. R., & Salehi, S. (2023, May). An Uncertainty Risk Evaluation Tool for Wellbore Leakage Prediction for Plug and Abandonement (P&A). In SPE Western Regional Meeting (p. D041S015R001). SPE.

[218] Bai, M., Zhang, Z., Yang, E., & Du, S. (2023). A fuzzy bayesian network based method for CO2 leakage risk evaluation during geological sequestration process. Geoenergy Science and Engineering, 222, 211423.

[219] Chen, B., Mehana, M. Z., & Pawar, R. J. (2022, April). Towards Quantitative Approach to Evaluating Greenhouse Gas Leakage from CO2 Enhanced Oil Recovery Fields. In SPE Improved Oil Recovery Conference? (p. D021S028R003). SPE.

[220] Mehana, M., Chen, B., & Pawar, R. (2022, June). Reduced-order models for wellbore leakage from depleted reservoirs. In SPE/AAPG/SEG Unconventional Resources Technology Conference (p. D031S053R003). URTEC.

[221] Liu, L., Chen, B., Mehana, M., Prodanovic, M., Pyrcz, M., & Pawar, R. (2022, December). Reduced-Order Models for the Prediction of Greenhouse Gas Leakage from Depleted Hydrocarbon Reservoirs Using Machine Learning Methods. In AGU Fall Meeting Abstracts (Vol. 2022, pp. GC12E-0492).

[222] Rochon, E., Kuper, J., Bjureby, E., Johnston, P., Oakley, R., Santillo, D., ... & Von Goerne, G. (2008). False Hope. Why carbon capture and storage won't save the climate.

Declaration of interests

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: