

Original Paper

Deep learning-assisted optimization for enhanced oil recovery and CO₂ sequestration considering gas channeling constraints

Xin-Yu Zhuang^{a, b}, Wen-Dong Wang^a, Yu-Liang Su^{a, *}, Zhen-Xue Dai^{c, d}, Bi-Cheng Yan^e

^a State Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao, 266580, Shandong, PR China

^b Department of Civil and Environmental Engineering, University of Alberta, Edmonton, T6G 1H9, Canada

^c College of Construction Engineering, Jilin University, Changchun, 130026, Jilin, PR China

^d School of Environmental and Municipal Engineering, Qingdao University of Technology, Qingdao, 266520, Shandong, PR China

^e Physical Science and Engineering (PSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

ARTICLE INFO

Article history:

Received 3 September 2024

Received in revised form

25 April 2025

Accepted 28 April 2025

Available online 28 April 2025

Edited by Yan-Hua Sun

Keywords:

Spatio-temporal sequence prediction

Multi-objective optimization

Enhanced oil recovery

CO₂ sequestration

Gas channeling

ABSTRACT

Carbon dioxide Enhanced Oil Recovery (CO₂-EOR) technology guarantees substantial underground CO₂ sequestration while simultaneously boosting the production capacity of subsurface hydrocarbons (oil and gas). However, unreasonable CO₂-EOR strategies, encompassing well placement and well control parameters, will lead to premature gas channeling in production wells, resulting in large amounts of CO₂ escape without any beneficial effect. Due to the lack of prediction and optimization tools that integrate complex geological and engineering information for the widely used CO₂-EOR technology in promising industries, it is imperative to conduct thorough process simulations and optimization evaluations of CO₂-EOR technology. In this paper, a novel optimization workflow that couples the AST-GraphTrans-based proxy model (Attention-based Spatio-temporal Graph Transformer) and multi-objective optimization algorithm MOPSO (Multi-objective Particle Swarm Optimization) is established to optimize CO₂-EOR strategies. The workflow consists of two outstanding components. The AST-GraphTrans-based proxy model is utilized to forecast the dynamics of CO₂ flooding and sequestration, which includes cumulative oil production, CO₂ sequestration volume, and CO₂ plume front. And the MOPSO algorithm is employed for achieving maximum oil production and maximum sequestration volume by coordinating well placement and well control parameters with the containment of gas channeling. By the collaborative coordination of the two aforementioned components, the AST-GraphTrans proxy-assisted optimization workflow overcomes the limitations of rapid optimization in CO₂-EOR technology, which cannot consider high-dimensional spatio-temporal information. The effectiveness of the proposed workflow is validated on a 2D synthetic model and a 3D field-scale reservoir model. The proposed workflow yields optimizations that lead to a significant increase in cumulative oil production by 87% and 49%, and CO₂ sequestration volume enhancement by 78% and 50% across various reservoirs. These findings underscore the superior stability and generalization capabilities of the AST-GraphTrans proxy-assisted framework. The contribution of this study is to provide a more efficient prediction and optimization tool that maximizes CO₂ sequestration and oil recovery while mitigating CO₂ gas channeling, thereby ensuring cleaner oil production.

© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Injecting CO₂ into reservoirs to enhance hydrocarbons (oil and gas) recovery is a conventional tertiary recovery technique, often

referred to as CO₂-EOR technology (Li et al., 2021; Wang et al., 2023; Yan et al., 2023). When injected into subsurface reservoirs, CO₂ serves a dual purpose: enhancing recovery of hydrocarbons by miscible flooding and concurrently facilitating the substantial sequestration of CO₂ within the interstitial voids of the reservoir (Yuan et al., 2022). However, the intricate geological and engineering uncertainties blur the potential for both CO₂ flooding and sequestration (Ampomah et al., 2017a, 2017b; Zhang et al., 2019;

* Corresponding author.

E-mail address: suyuliang@upc.edu.cn (Y.-L. Su).

Zhuang et al., 2025). It is imperative to develop effective methods to accurately predict the production dynamics of CO₂ flooding and efficiently optimize CO₂-EOR strategies, which generally include well placement and well control parameters, ensuring the successful implementation of CO₂-EOR technology.

In the CO₂ flooding development of hydrocarbon reservoirs, commonly employed methods for production dynamics prediction include the analytical model, production decline curve analysis, and reservoir numerical simulation (Vo et al., 2020). The analytical model establishes mathematical models based on fluid seepage principles to forecast well productivity (Ding et al., 2020; Dutt and Mandal, 2012; Ji et al., 2017). However, this method tends to be overly idealized, considering numerous assumptions and ideal conditions. It faces challenges in accurately describing the intricate seepage processes of strongly heterogeneous hydrocarbon reservoirs, leading to suboptimal prediction accuracy. Decline curve analysis is a statistical method that predicts future production by fitting historical production data (Tan et al., 2018; Tang et al., 2024; Yehia et al., 2023). This method involves fitting production data with a mathematical model. Although this method is straightforward and practical, it is limited to forecasting production under stable conditions. Furthermore, it does not account for the dynamic influences on well dynamics, often resulting in significant disparities between predicted and actual production outcomes (Tadjer et al., 2022). Reservoir numerical simulation can replicate complex underground fluid flow processes, but their predictive accuracy excessively relies on detailed geological descriptions and high-quality history matching results (Jiang, 2011; Xiao et al., 2024; Zhou et al., 2019). Consequently, simulation costs are elevated, hindering the feasibility of large-scale prediction tasks. In recent years, the application of artificial intelligence technologies in dynamic production prediction for hydrocarbon reservoirs has gained increasing prominence (Al-Khafaji et al., 2023; Choubey and Karmakar, 2021; Van and Chon, 2018; Zhuang et al., 2024a). Various machine learning and deep learning methods can analyze potential dynamic production patterns from intricate historical data (Panja et al., 2018). Some methods even consider multidimensional information, such as the temporal and spatial aspects influencing reservoirs, to achieve precise prediction of production dynamics (Thanh et al., 2023; Wu et al., 2021; Zhuang et al., 2024b). However, selecting appropriate methods tailored to distinct geological conditions and application scenarios remains crucial to ensure the maximization of computational resource utilization.

In the realm of reservoir production optimization, traditional methods primarily rely on reservoir numerical simulation for comparing production scenarios and selecting the one with the highest cumulative oil production as the recommended option (Chen et al., 2010; Dai et al., 2014; Pancholi et al., 2020). This method falls short of genuine optimization, as it merely involves selecting the best option among those available (Al-Mudhafar et al., 2018; Kamali et al., 2015; Qiu et al., 2020). In recent years, optimization algorithms based on mathematical theories have seen increasing application in the field of CO₂ flooding (He et al., 2021; Menad and Nouredine, 2019; Safi et al., 2016). Ampomah et al. (2017a, 2017b) utilized a genetic algorithm with a mixed-integer capability optimization approach to maximize the oil recovery and CO₂ sequestration in the target reservoir. Kashkooli et al. (2022) defined a weighted combined objective function to simultaneously optimize CO₂ flooding and sequestration effects, explicitly highlighting differences in optimization results before and after CO₂ channeling. You et al. (2020a, 2020b) employed a coupling method of artificial neural network proxy models and particle swarm optimization, considered economic (net present value) and physical (reservoir pressure) constraints for optimizing oil recovery and CO₂ sequestration volume in a CO₂-WAG project. However, the

significant disparity in mobility between CO₂ and oil makes CO₂ flooding prone to gas channeling. Additionally, with increasing reservoir heterogeneity, CO₂ tends to advance along high-permeability layers, leading to a significant decline in oil displacement efficiency (Gao et al., 2014; Hao et al., 2016; Zhao et al., 2020). Remarkably, the mentioned studies did not account for the impact of gas channeling on CO₂-EOR during the optimization process.

In this paper, a novel deep learning proxy-assisted optimization workflow is proposed for the collaborative optimization of CO₂-EOR strategies. The workflow comprises two main components: the construction of a proxy model and an optimization method. Initially, a proxy model is established by coupling Graph Attention Network (GAT) with Transformer to achieve rapid responses in the dynamics of CO₂ flooding and sequestration. GAT is employed to extract spatial features of geological properties and the injection-production well pattern, while Transformer handles the temporal correlations of well control parameters across various time steps. Additionally, a joint optimization method for CO₂-EOR strategies is introduced, considering balanced CO₂ flooding. It achieves multi-objective collaborative optimization of cumulative oil production and CO₂ sequestration volume while mitigating the occurrence of CO₂ channeling phenomena based on the Multi-objective Particle Swarm Optimization (MOPSO) algorithm. Finally, the proposed workflow is validated for its performance and effectiveness using a 2D synthetic model and a 3D field-scale reservoir model.

The rest of this paper is structured as follows: Section 2 elucidates the fundamental principles and contributions of the methods employed in this study. Section 3 outlines the main framework of the proposed method and elaborates on the functions of each component. Section 4 validates the optimization workflow by employing various reservoir models to assess its performance and effectiveness. Finally, Section 5 presents conclusions drawn from the study and outlines avenues for future research.

2. Related techniques

2.1. Graph attention network

Convolutional Neural Networks (CNNs) are widely employed in the field of spatial feature extraction (Zhang et al., 2023). CNNs are specifically designed for processing Euclidean data. However, practical engineering problems often involve non-Euclidean data, with graph structures being a prominent example of such non-Euclidean or topological data, such as chemical molecular structures and interwell connectivity. The inherent variability in the number of adjacent nodes for each vertex poses a challenge, rendering traditional CNNs incapable of performing local convolution operations on non-Euclidean data. Graph Neural Networks (GNNs) exhibit excellent performance in extracting crucial complex features from non-Euclidean data, uncovering deep and effective topological information, and achieving rapid processing of massive datasets (Huang et al., 2023). In the process of reservoir development, the injection-production well pattern can be equivalently represented as a graph structure, with strong correlations between well points. Therefore, dynamic prediction of CO₂ flooding and sequestration using GNNs becomes imperative, as it incorporates spatial information like well distribution and temporal information pertaining to well control parameters.

In recent years, numerous scholars have conducted innovative research on GNNs and derived a series of variants. Among them, the GAT has demonstrated outstanding performance (Chen et al., 2024; Lin et al., 2023; Pan et al., 2023; Zhen et al., 2022). In the application of GNNs, the computation of parameter values typically occurs in an iterative fashion, which may result in a relatively high number of

iterations and reduced efficiency. Additionally, deeper levels of GNNs can lead to excessive similarity in features among all nodes, causing oversmoothing issues. To address these challenges, a growing number of scholars have conducted extensive research using GAT, which is a more flexible and expressive graph convolutional network structure (Do et al., 2021; Jiang and Luo, 2022). GAT introduces a self-attention mechanism, allowing the model to dynamically assign varying weights to each node. This enables each node to focus on its neighboring nodes to varying degrees, thereby enhancing the expressiveness of the model. The core elements of GAT are similar to GNNs, and they can be described using the following equation.

$$G = (V_N, E_{ij}, A) \quad (1)$$

where V_N is the set of nodes; E_{ij} is the set of edges characterizing the strength of connections between nodes; A is the weighted adjacency matrix describing the connection relationships between nodes.

Based on the above, GAT incorporates attention weight which focuses on the dynamic relationships between nodes and the contributions of different nodes to the target node. This implies that the model can discern the contribution of each neighboring node to the target node based on the particular task and dataset during the aggregation of information from neighboring nodes to update a node's representation. This adaptability is achieved without relying on a fixed weight matrix.

The graph attention layer serves as the core layer in GAT, employing attention mechanisms to dynamically compute the weights between nodes, thereby enabling a more adaptable aggregation of information from neighboring nodes. The computation process of a graph attention layer is as follows. Initially, linear transformations are applied to the input node features and adjacency matrix to obtain node representations, as shown in the following equation:

$$h_i^{(l)} = X_i \cdot W^{(l)} \quad (2)$$

where $h_i^{(l)}$ is the node representation; X_i is the node feature matrix; $W^{(l)}$ is the weight matrix; l is the number of layers.

Subsequently, attention weights need to be computed for each node. This is achieved by calculating the weights using the learned weight vector within the attention mechanism. The computation of attention weights involves evaluating the relationships between nodes to determine the contribution of each node when aggregating information from neighbors. The equation is as follows:

$$e_{ij} = \text{LeakyReLU}(\vec{a}^{(l)\top} [Wh_i^{(l)} \parallel Wh_j^{(l)}]) \quad (3)$$

$$\alpha_{ij}^{(l)} = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})} \quad (4)$$

where e_{ij} is the transition parameter characterizing the contribution of node i to node j ; $\vec{a}^{(l)\top}$ is the learned attention parameter vector; LeakyReLU is the activation function; $\parallel$ is vector concatenation; $\alpha_{ij}^{(l)}$ is the attention weight obtained through softmax normalization.

Finally, it is necessary to aggregate the representations of neighboring nodes with attention weights to obtain the updated representation for each node. The equation is as follows:

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in N(i)} \alpha_{ij}^{(l)} \cdot Wh_j^{(l)} \right) \quad (5)$$

where $h_i^{(l+1)}$ is the updated node representation; σ is the activation function.

The aforementioned process outlines the computation of a basic graph attention layer and the schematic diagram of which is shown in Fig. 1. By stacking multiple graph attention layers, GAT is capable of progressively learning more complex graph structural features, thereby enhancing its accuracy in handling prediction or classification tasks.

2.2. Transformer model

The Transformer has achieved tremendous success in the fields of Natural Language Processing (NLP) and Computer Vision (CV) (Carpenter, 2022; Zhang et al., 2023). In recent years, due to its advantages such as parallel computation and fewer trainable parameters, Transformer has gradually replaced convolutional neural networks and recurrent neural networks in engineering applications dealing with time series data. It particularly excels in scenarios involving time series data, such as reservoir production forecasting.

Transformer primarily consists of an encoder and a decoder. Furthermore, the encoder is composed of multiple identical encoding layers, with each encoding layer incorporating a self-attention mechanism and a feedforward neural network. In the self-attention mechanism, the representation at each position can depend on other positions in the sequence. To enhance the expressive power of the model, Transformer introduces multiple attention heads. The calculation of attention weights for each attention head is as follows:

$$\delta^h = \text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (6)$$

where δ^h is the attention weights for each attention head; Q , K , and V are matrices obtained by projecting the query, key, and value through weight matrices; d_k is the feature dimension of the attention head. The output of multi-head attention is the concatenation of the outputs from each attention head, as illustrated below:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\delta^1, \delta^2, \dots, \delta^h) W_O \quad (7)$$

where W_O is the projected weight matrix of multi-head attention.

In the feedforward neural network section, following the processing by the self-attention mechanism, the representation at each position undergoes a feedforward neural network layer which enhances the nonlinear expressive capabilities of the model. The operations in this part are similar to the computation process in an artificial neural network and would not be reiterated here. Additionally, both layer normalization and residual connection are applied after the multi-head self-attention layer and the feedforward neural network layer. This ensures a smooth training process for the model while better capturing long-range dependencies and complex spatio-temporal features in the input sequence. The structure of decoder mirrors that of the encoder. These layers are stacked multiple times to construct a deep Transformer model, enabling the capturing of time dependencies in the output of GAT. The Transformer architecture in an encoder setting is shown in Fig. 2.

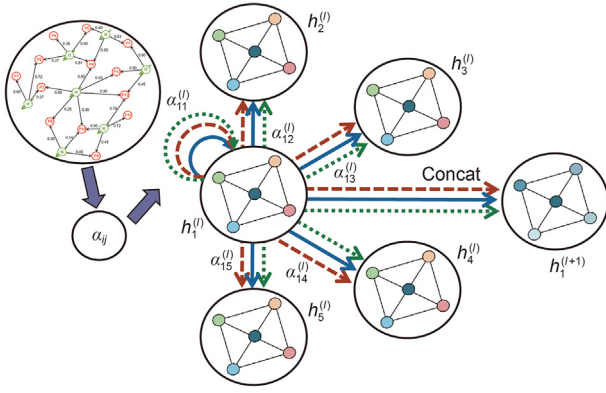


Fig. 1. Schematic diagram of the graph attention layer.

2.3. Multi-objective optimization algorithm

In practical engineering problems, one often encounters optimization issues involving the interaction of multiple objectives. Single-objective optimization algorithms struggle to address such problems effectively, leading to the emergence of multi-objective optimization algorithms. It proves challenging in multi-objective optimization problems to identify solutions that simultaneously satisfy each objective, given the interdependence, mutual constraints, and even contradictions among the objectives. Two methods exist for handling this challenge. One involves assigning different weight coefficients based on the importance of each objective. Using a weighted method, multiple objectives are transformed into a single objective function incorporating distinct weights, thereby converting the problem into a single-objective optimization one. Nevertheless, this method exhibits limitations by neglecting the conceivable competition among objectives and the subjective nature inherent in the assignment of weight values to each objective. The alternative method employs Pareto

optimization methods, leveraging Pareto dominance strategies to evaluate solution quality and ascertain the Pareto front, thereby yielding an optimal set of solutions. Subsequently, constraints are introduced in accordance with the decision-maker's specifications to identify compromised solutions. The Pareto front can be construed as comprising optimal solutions.

To align with the development trend of multi-objective optimization, incorporating corresponding optimization strategies into the basic Particle Swarm Optimization (PSO) algorithm becomes imperative for directly addressing multi-objective problems. The MOPSO algorithm harnesses collective intelligence within the particle swarm, where cooperation and competition among particles guide the optimization search (Farahi et al., 2021; You et al., 2020a, 2020b). During each iteration, particles update their positions and velocities based on two parameters: the individual best solution (P_{best}) for each particle and the global best solution (G_{best}) within the entire swarm. The update equations for the velocity v_i^{k+1} and position s_i^{k+1} of the i -th particle at the $(k+1)$ -th iteration are expressed as follows:

$$v_i^{k+1} = \omega v_i^k + c_1 r_1 (P_{best} - s_i^k) + c_2 r_2 (G_{best} - s_i^k) \quad (8)$$

$$s_i^{k+1} = s_i^k + v_i^{k+1} \quad (9)$$

where k is the number of iterations; ω is inertia weight; c_1 and c_2 are acceleration coefficients used to adjust the proportion of individual best positions and neighborhood best positions in the velocity update process; r_1 and r_2 are random numbers uniformly distributed in the range $[0, 1]$.

Incorporating an external archive space into the standard PSO algorithm proves instrumental for recording the Pareto optimal solution set. During each optimization iteration, the MOPSO algorithm determines the Pareto front of the particle swarm, updating the optimal solution set. To accommodate a full archive space, an adaptive grid method is employed for optimizing the archive space. This augmentation results in the formulation of MOPSO algorithm. The application of MOPSO to the optimization process in this study is illustrated in Fig. 3.

3. Methodology

In this study, we propose a deep learning proxy-assisted optimization workflow that integrates three key components. First, the graph structure captures both dynamic and static reservoir information within the CO₂ flooding well pattern. Second, the AST-GraphTrans-based proxy model enables precise dynamic prediction of CO₂ flooding and sequestration. Third, the constrained multi-objective optimization ensures optimal performance across multiple development objectives while delaying gas channeling. Through the collaboration of these three components, the prediction and optimization tasks involved in CO₂-EOR technology can be effectively addressed.

3.1. Main framework

Considering that the input and output data designed for this study encompass multidimensional data such as graph structures, graph matrices, and time series, traditional single machine learning or deep learning model may not be suitable for this task. Therefore, this paper proposes an Attention-based Spatial-temporal Graph-Transformer Network (AST-GraphTrans) proxy-assisted optimization workflow. It primarily consists of three components. Firstly, the graph structure is constructed using complex topological

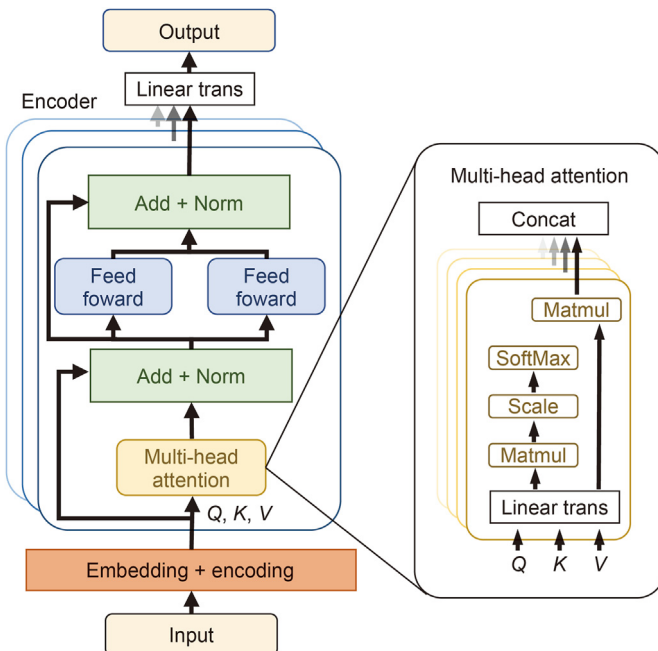


Fig. 2. Schematic diagram of the Transformer architecture in an encoder setting.

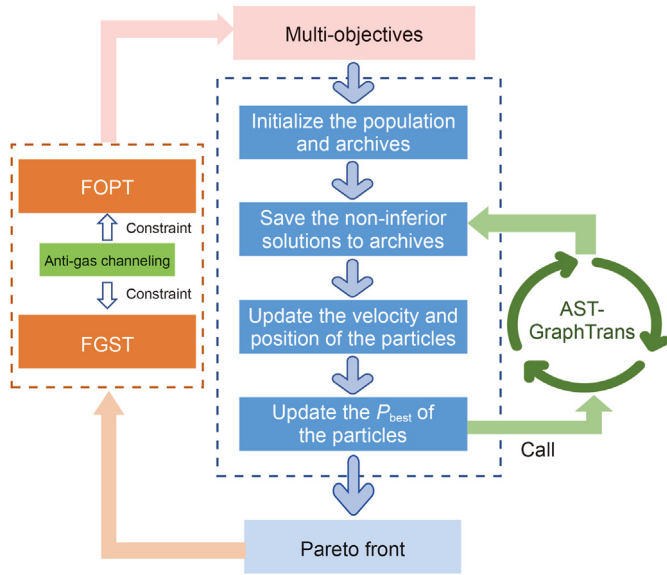


Fig. 3. Schematic diagram of the MOPSO workflow.

relationships to extract the spatio-temporal features of geological properties, well distribution, and development strategies. In this study, special attention is given to the coordinates of each well and CO₂ flooding connectivity data. Secondly, an AST-GraphTrans-based proxy model is utilized to analyze the inherent temporal dependencies in the CO₂ flooding and sequestration process through the multi-layer stacking of GAT and Transformer. Integrated with fully connected layers, the model can generate an output time series that includes cumulative oil production, CO₂ sequestration volume, and CO₂ plume front distribution at different time steps. Finally, the proxy-assisted optimization framework facilitates the joint optimization of well placement and well control parameters. On the optimized Pareto front, cumulative oil production and CO₂ sequestration volume are emphasized to optimize the scheme, thus providing guidance for the scheme decision of reservoir engineers. The overall architecture of the AST-GraphTrans proxy-assisted optimization workflow is illustrated in Fig. 4.

3.2. Construction of the graph structure

The graph structure serves as the nucleus of GAT. If the graph structure fails to effectively encode node features, it leads to an ambiguity in the correlation among nodes. It hampers the accurate extraction of potential spatio-temporal information between nodes, potentially diminishing predictive performance. In this study, the graph structure primarily characterizes the spatial correlations among different wells within the CO₂ flooding well pattern. Wells are treated as nodes, the existence of a connection between two wells is regarded as an edge, and the strength of connectivity between injection and production wells is considered as the weight of the edge (commonly used to establish the adjacency matrix). The process of abstracting the well pattern into the graph structure is shown in Fig. 5.

Furthermore, after determining the connectivity between different wells, it is essential to specify the connectivity strength among injection and production wells by characterizing the weight coefficients of the edges. The robustness of connectivity between injection and production wells is impacted not solely by the distance between two wells but also by reservoir geological parameters, such as permeability and porosity. For instance, the Euclidean

distance between injection well I1 and production well P3 is 300 m, with an average permeability between wells of 10 mD; whereas the distance between I1 and P1 is 500 m, with an average permeability of 100 mD. In practical production scenarios, the intricate determination of connectivity between I1 and P1 (or between I1 and P3) poses a significant challenge. Therefore, this paper proposes a quantitative evaluation index for the CO₂ flooding connectivity between injection and production wells, expressed by the following equation:

$$Z(I1, P1) = \frac{K \left(\frac{K_{ro}(S_o)}{\mu_o} + \frac{K_{rg}(S_o)}{\mu_g} \right) H / D_{I1,P1}}{\sum_{c=1}^n \left[\left(\frac{K_{ro}(S_o)}{\mu_o} + \frac{K_{rg}(S_o)}{\mu_g} \right) H / D_{I1,Pc} \right]_c} \quad (10)$$

where $Z(I1, P1)$ is the CO₂ flooding connectivity coefficient between I1 and P1; K is the average permeability between I1 and P1; $K_{ro}(S_o)$ and $K_{rg}(S_o)$ are the relative permeability of the oil phase and gas phase at the average oil saturation between wells; μ_o and μ_g are the viscosity of the oil phase and gas phase, respectively; H is the reservoir thickness; $D_{I1,P1}$ is the well spacing between I1 and P1; c is the production well index in I1 well group; n is the number of the production wells in I1 well group. Since this paper concentrates on the oil production capacity of production wells and the CO₂ injection capacity of injection wells, the equation accounts solely for the influence of the flow ability of oil and gas on the connectivity between wells. Using this equation, the CO₂ flooding connectivity matrix, i.e., the adjacency matrix, can be computed. This yields the initial connection weights between different wells, which are subsequently normalized within [0, 1].

In this paper, the input data for AST-GraphTrans consists of the adjacency matrix and the node feature matrix. The construction of the adjacency matrix is guided by domain-specific knowledge in the petroleum engineering field. The inclusion of physically grounded connection relationships can further enrich the AST-GraphTrans-based proxy model with additional reservoir information, thereby enhancing predictive performance. Additionally, we organize the time series of well control parameters, which characterize human-operated dynamics, into the node feature matrix. Integrating the Transformer model into this framework facilitates the incorporation of temporal correlations inherent in injection and production data. The amalgamation of these latent spatial and temporal insights empowers the proposed framework to consistently and precisely forecast the potential of CO₂ flooding and sequestration across diverse time intervals.

The dataset employed in this paper is characterized by its multidimensional nature. The input data comprises the graph structure of the target reservoir model, which includes the node feature matrix consisting of well control parameters and the adjacency matrix formed by the CO₂ flooding connectivity coefficients. The output data encompasses cumulative oil production, CO₂ sequestration volume, and the spatial distribution of the CO₂ plume front at various time steps. To optimize the efficiency of model training and alleviate potential numerical errors arising from variations in data dimensions, preprocessing of the input data is executed. Specifically, the well control data for each well undergoes scaling to the interval [0, 1] through the application of the min-max normalization method, as articulated by the following equation:

$$X_s = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (11)$$

where X_s is the scaled value; X is the original data; $\max(X)$ and $\min(X)$ are the maximum and minimum values in the dataset,

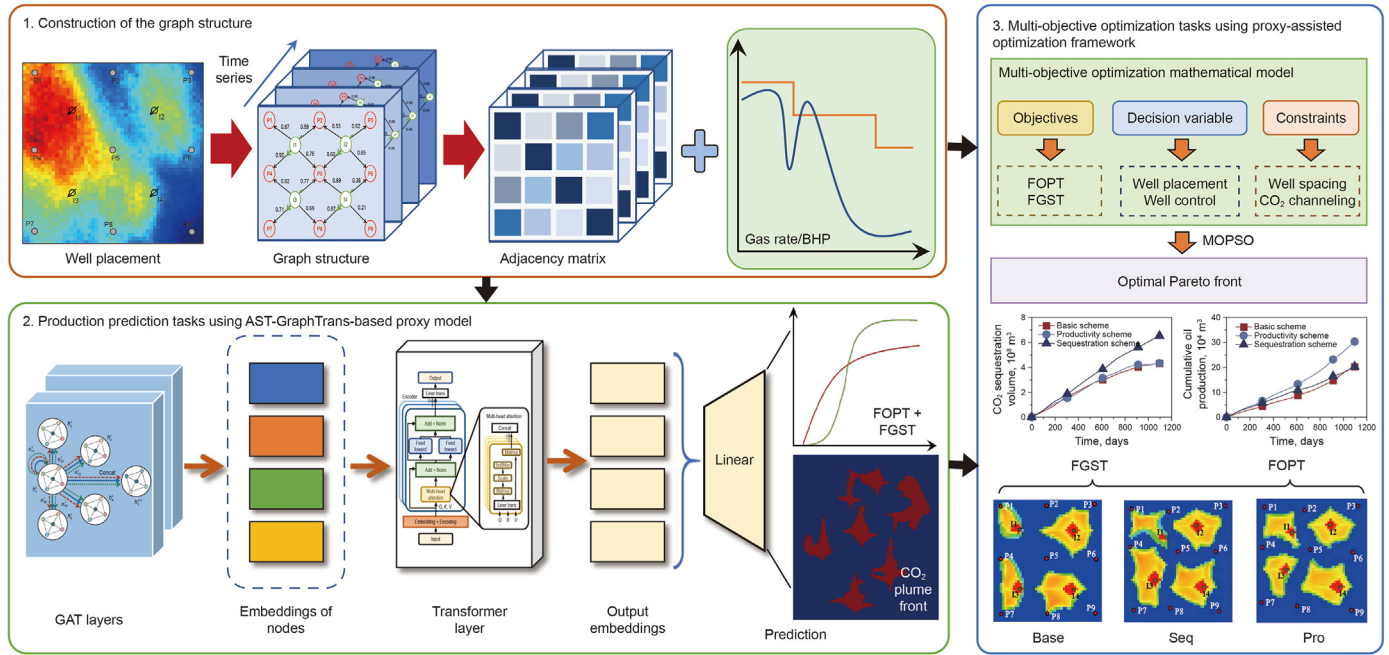


Fig. 4. Schematic diagram of AST-GraphTrans proxy-assisted optimization workflow.

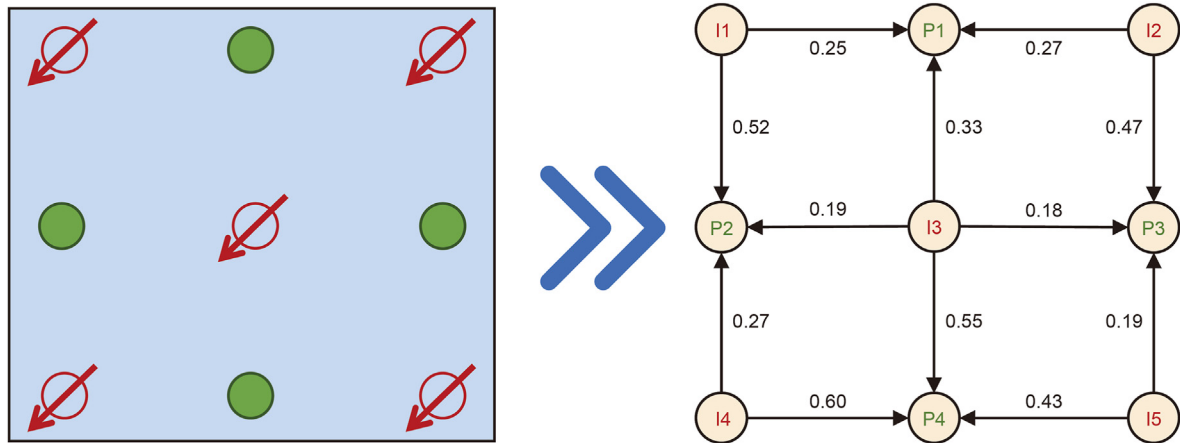


Fig. 5. Schematic diagram of abstracting the well pattern into the graph structure.

respectively.

Furthermore, the dataset is partitioned into training, validation, and testing sets in a ratio of 6:2:2. Given that well control data constitutes a time series characterized by temporal correlations and influences, the partitioning is conducted in accordance with the chronological order of the time series. The initial 60% of the data is allocated as the training set for AST-GraphTrans, thereby supporting the training process. The subsequent 20% is assigned as the validation set, serving the purpose of assessing the model's generalization accuracy and preventing overfitting. Lastly, the remaining 20% comprises the testing set, employed for evaluating the predictive performance of AST-GraphTrans.

3.3. Production prediction tasks using AST-GraphTrans-based proxy model

In this paper, the AST-GraphTrans-based proxy model is established to accomplish production prediction tasks, serving as a cost-effective alternative tool to numerical simulations for subsequent

optimization tasks. The inputs to the proxy model consist of two components: the well control data for each well and the CO₂ flooding connectivity relationships. The well control data primarily includes bottomhole pressures for the production wells and injection rates for the injection wells. These data are sequentially arranged according to the well order to construct the node feature matrix. The CO₂ flooding connectivity relationships are established based on Eq. (10) to build the graph structure, taking into account the intricate relationships between different injection and production wells in constructing the adjacency matrix. Leveraging GAT to process both the node feature matrix and adjacency matrix, we derive a feature matrix that effectively captures the spatial interactions among injection and production wells. Subsequently, feeding the feature matrix at various time steps into the Transformer model allows for the exploration of the temporal correlations of well control parameters, building upon the extracted spatial features of the CO₂ flooding well pattern. Finally, cumulative oil production, CO₂ sequestration volume, and CO₂ plume front at different time steps are output through fully connected and flatten

layers.

The fundamental architecture of the AST-GraphTrans comprises two graph attention layers stacked on top of each other along with a Transformer module. The stacking of two graph attention layers can learn more in-depth feature representation, and take into account the further connection relationships between nodes, thus it has stronger expression ability for capturing CO₂ flooding connectivity relationships. The Transformer encoder is employed to perform self-attention operations on the input, enabling the model to capture the internal structure of the data. Details regarding the successive architecture and the corresponding output sizes of the AST-GraphTrans are provided in Table 1.

The table provides definitions for the variables used: N is the number of samples; V is the number of nodes, which corresponds to the quantity of wells in this study; F is the number of features per node, referring to the well control parameters for each well; F_G is the feature dimension of each attention head, set to 128 in this study; H is the number of attention heads, set to 8 in this study; D_{Trans} is the dimension of the hidden layers in the Transformer, set to 256 in this study; T is the sequence length, determined based on the time steps in this study; O_1 , O_2 , and O_3 respectively represent the dimensions of the three output layers, where O_1 and O_2 are one-dimensional sequences, and O_3 is a two-dimensional image.

This proxy model leverages the analysis of graph structures incorporating reservoir information and engineering uncertainties. It excavates latent patterns in the spatial distribution of geological properties and temporal variations in well control parameters, capturing their influence on the underground CO₂ plume pattern. This enables rapid responses to the CO₂ flooding and sequestration dynamics under different geological and engineering parameter combinations. The AST-GraphTrans-based proxy model achieves prediction accuracy equivalent to numerical simulations while ensuring lower time costs, thus providing a robust foundation for efficient optimization execution.

Two widely recognized metrics within the industry, R^2 (coefficient of determination) and MRE (mean relative error), are chosen in this paper to evaluate the prediction accuracy of the proxy model. Due to the relatively large numerical values of cumulative oil production and CO₂ sequestration volume, traditional metrics like MSE may not effectively gauge predictive performance. Therefore, R^2 and MRE are selected. The equations are as follows:

$$R^2 = 1 - \frac{\sum (y - y')^2}{\sum (y - y_{\text{ave}})^2} \quad (12)$$

$$\text{MRE} = \frac{1}{n} \sum \frac{|y - y'|}{|y|} \quad (13)$$

where y is the actual value; y' is the predicted value; y_{ave} is the average of the actual values. The two metrics are employed to assess the performance of the trained proxy model.

To validate the hyper-parameter configuration's rationality for

predictions as evaluation metrics. Initially, for the feature dimension of each attention head (F_G), we comparatively analyzed three configurations: $F_G = 64$, 128, and 256. Results demonstrate that $F_G = 128$ achieves the optimal balance between performance and efficiency, with significantly superior predictive performance compared to $F_G = 64$, while simultaneously reducing training time by 65% and GPU memory consumption by 32% relative to $F_G = 256$. Subsequently, optimization of the number of attention heads (H) revealed $H = 8$ as the optimal choice, delivering performance nearly equivalent to $H = 16$ (differing by merely 0.5%), while dramatically reducing training time and memory consumption by 82% and 52%, respectively. Further investigation of the hidden layers' dimension (D_{Trans}) showed that $D_{\text{Trans}} = 256$ significantly enhances predictive accuracy while maintaining computational efficiency. Additionally, noise robustness experiments indicated stable performance of the model under low noise ($\sigma = 0.05$) and moderate noise ($\sigma = 0.1$) conditions. Although performance marginally decreased under high noise ($\sigma = 0.2$), it remained within acceptable ranges, with memory consumption maintaining stability. These experimental results comprehensively substantiate that the parameter configuration of $F_G = 128$, $H = 8$, and $D_{\text{Trans}} = 256$ achieves an optimal balance between predictive performance and computational efficiency.

3.4. Multi-objective optimization tasks using proxy-assisted optimization framework

In the process of CO₂ flooding within the hydrocarbon reservoir, the primary challenge arises from factors such as reservoir heterogeneity, imperfect well pattern, and unreasonable injection-production systems. The most significant issue faced during CO₂ flooding is gas channeling, exacerbated by reservoir heterogeneity and inadequate well pattern. Early occurrences of gas channeling lead to inefficient injection gas circulation. Addressing these challenges, this study establishes the interwell CO₂ flooding diffusion coefficient to quantitatively evaluate the balanced CO₂ flooding level among injection and production wells. This coefficient is utilized as a constraint for optimizing well placement and well control parameters jointly, with the multi-objectives of maximizing cumulative oil production and CO₂ sequestration volume. Consequently, the multi-objective optimization mathematical model for CO₂-EOR strategies is established.

In this mathematical model, three fundamental elements: the objective function, decision variables, and constraints are essential. Firstly, the objective function is a critical component. Given that CO₂-EOR technology seeks not only improved oil recovery but also aims to maintain a substantial amount of CO₂ underground for large-scale geological storage, this study establishes a hybrid function with the multi-objectives of maximizing cumulative oil production and CO₂ sequestration volume. The equation of this hybrid objective function is expressed as follows:

$$\begin{cases} \text{MaxFOPT} = (q_{l1} + q_{l2} + q_{l3} + \cdots + q_{lm}) \times (1 - f_w) \\ \text{MaxFGST} = (I_{g1} + I_{g2} + I_{g3} + \cdots + I_{gn}) - (q_{g1} + q_{g2} + q_{g3} + \cdots + q_{gm}) \end{cases} \quad (14)$$

the AST-GraphTrans architecture, this study conducted systematic ablation experiments and noise robustness analysis, with results presented in Table 2. The experiments employed the average value of R^2 and MRE of cumulative oil production and CO₂ sequestration

where m and n are the number of production wells and injection wells; q_{lm} is the liquid production rate of the m -th production well; f_w is the water cut; I_{gn} is the CO₂ injection rate of the n -th injection well; q_{gm} is the gas production rate of the m -th production well.

Table 1
Fundamental architecture of the AST-GraphTrans.

Layer		Parameter	Output size
Input		/	(N, V, F)
GAT layer-1 (attention heads H)	Nonlinearity	LeakyReLU	$(N, V, F_G \times H)$
	Dropout	$p = 0.05$	$(N, V, F_G \times H)$
GAT layer-2 (attention heads H)	Nonlinearity	LeakyReLU	$(N, V, F_G \times H)$
	Dropout	$p = 0.05$	$(N, V, F_G \times H)$
Flatten		/	$(N, V \times F_G \times H)$
Transformer encoder layer	Position-wise feed forward	/	(N, T, D_{Trans})
	Normalization	/	(N, T, D_{Trans})
	Dropout	$p = 0.05$	(N, T, D_{Trans})
Output layer-1 (FC)		Dense	(N, O_1)
Output layer-2 (FC)		Dense	(N, O_2)
Output layer-3 (FC)		Dense + reshape	(N, O_3)

Table 2
Results of ablation experiment and noise robustness analysis.

Hyper-parameter	Value	Average R^2	Average MRE	Training time, h	GPU memory, GB
F_G	64	0.939	0.188	3.2	5.6
	128	0.960	0.163	4.1	7.8
	256	0.963	0.157	6.8	10.3
H	4	0.930	0.199	2.8	5.2
	8	0.957	0.183	4.5	7.8
	16	0.962	0.165	8.2	11.9
D_{Trans}	64	0.936	0.193	3.0	5.4
	128	0.940	0.188	4.7	7.8
	256	0.968	0.118	5.2	8.5
AST-GraphTrans (Gaussian noise $\sigma = 0.05$)		0.955	0.133	5.3	7.8
AST-GraphTrans (Gaussian noise $\sigma = 0.1$)		0.941	0.190	5.4	7.8
AST-GraphTrans (Gaussian noise $\sigma = 0.2$)		0.913	0.212	5.4	7.8

Subsequently, the decision variables are the indispensable components in the optimization process during each iteration. The engineering outcomes of the CO₂-EOR technology, such as oil recovery and sequestration efficiency, are contingent upon both the inherent conditions of the undeveloped reservoir and the chosen development strategy. Reservoir conditions encompass geological structure, distribution of permeability fields, and hydrocarbon properties. Optimization of the field development primarily involves the design of well pattern and injection-production schemes. In this paper, the decision variables for optimization include the well coordinates for each well, the injection rates for injection wells, and the bottomhole pressures for production wells. These decision variables in the mathematical model are chiefly represented by the following vectors:

$$X_{\text{Coor}} = [X_1, X_2, X_3, \dots, X_{m+n}] \quad (15)$$

$$Y_{\text{Coor}} = [Y_1, Y_2, Y_3, \dots, Y_{m+n}] \quad (16)$$

$$I_{\text{inj}} = [q_{g1}, q_{g2}, q_{g3}, \dots, q_{gn}] \quad (17)$$

$$\text{BHP}_{\text{pro}} = [P_{p1}, P_{p2}, P_{p3}, \dots, P_{pm}] \quad (18)$$

where X_{Coor} and Y_{Coor} are the coordinates for each well; I_{inj} is the CO₂ injection rate of injection wells. BHP_{pro} is the borehole pressure of production wells.

The final element is the constraint, which constitutes a crucial aspect of our study. We categorize the constraints into two main aspects, with the first one pertaining to the constraints on well coordinates. In the optimization process, the selection of well placement cannot be arbitrary. Instead, it must adhere to the

constraint of minimum well spacing. Deviating from this constraint could exacerbate the occurrence of gas channeling, leading to premature decommissioning of production wells. This constraint can be defined by the following equation:

$$\sqrt{(X_A - X_B)^2 + (Y_A - Y_B)^2} \geq D^{\min} \quad (19)$$

where X_A , X_B , Y_A , and Y_B are the coordinates for well A and well B, respectively; $D^{\min}$ is the minimum well spacing between any two wells, which is set at 650 ft in this study.

The second aspect pertains to the constraints associated with well control parameters, specifically addressing anti-gas channeling constraints. Thoughtful adjustments to well control parameters across various time steps can play a positive role in delaying the onset of gas channeling. It is essential to establish reasonable relationship between different injection and production wells to ensure that the injected CO₂ is primarily employed for displacing hydrocarbon and retained underground. This prevents the preferential migration of CO₂ channeling that might adversely impact the efficiency of production wells. In this paper, since the proxy model adeptly predicts the dynamic distribution of CO₂ plume fronts between injection and production wells, we aim to constrain the range of well control parameters based on the distribution of CO₂ plume fronts from the previous time step. This constraint enables the restraint of CO₂ plume underground effectively. We define a metric representing the uniformity of CO₂ flooding diffusion between different injection and production wells, termed the interwell CO₂ flooding diffusion coefficient λ and the relative CO₂ flooding diffusion degree η . The equations are presented below

$$\lambda_h = \frac{d_h}{D_h} \quad (20)$$

$$\eta_h = \frac{\lambda_h - \left(\sum_{h=1}^m \lambda_h \right) / m}{\left(\sum_{h=1}^m \lambda_h \right) / m} \quad (21)$$

where λ_h is the interwell CO₂ flooding diffusion coefficient between the h -th production well and injection well within a particular well group; d_h is the Euclidean distance between the CO₂ plume front of the h -th production well and the injection well; D_h is the well spacing between the injection well and the h -th production well; η_h is the relative CO₂ flooding diffusion degree between the h -th production well and injection well; m is the number of production wells within a specific well group.

In essence, the aforementioned constraint conditions are established based on the relative error between the interwell CO₂ flooding diffusion degree of each production well and the average interwell CO₂ flooding diffusion degree at the previous time step. This relative error is utilized to constrain the parameter selection ranges for the injection rate and the bottomhole pressure at the current time step. We conducted a sensitivity analysis of the relative CO₂ flooding diffusion degree to further elucidate how this parameter can be utilized as a dynamic constraint condition in guiding well control parameter optimization. Upon obtaining the relative CO₂ flooding diffusion degree from the previous time step, it becomes necessary to integrate a constraint multiplier for determining the current time step's well control parameters, as illustrated in Eq. (22).

$$W_{h,t} = \sigma \times W_{h,t-1} \quad (22)$$

where $W_{h,t}$ is the well control parameter range at the current time step; σ is the constraint multiplier, indicating the adjustment magnitude of well control parameters. We conducted a sensitivity analysis of different relative CO₂ flooding diffusion degrees from the previous time step, as illustrated in Fig. 6.

The analysis reveals several key insights. When the relative CO₂ flooding diffusion degree from the previous time step is negative, it indicates that the interwell CO₂ flooding diffusion coefficient is below the well group average. Consequently, the relative CO₂ flooding diffusion degree of the current time step increases with the constraint multiplier's expansion, with the rate of increase gradually decelerating. Conversely, when the relative CO₂ flooding

diffusion degree from the previous time step is positive, it signifies that the interwell CO₂ flooding diffusion coefficient exceeds the well group average, causing the relative CO₂ flooding diffusion degree of the current time step to decrease as the constraint multiplier increases, with the rate of decrease progressively accelerating. When the relative CO₂ flooding diffusion degree equals zero, it confirms that the interwell CO₂ flooding diffusion coefficient has reached the average value, which is an optimal balanced displacement performance that we aim to achieve. Therefore, we adjust well control parameters to approach a relative CO₂ flooding diffusion degree of zero as closely as possible. From the analysis, we observe that a constraint multiplier of 1.14 enables the interwell CO₂ flooding diffusion coefficient to reach the average value when the relative CO₂ flooding diffusion degree of the previous time step is negative. Conversely, when the relative CO₂ flooding diffusion degree of the previous time step is positive, a constraint multiplier of 1.30 achieves the same equilibrium. Consequently, we can further refine the equation as follows:

$$W_{h,t} = \begin{cases} (1 - 1.14 \times \eta_{h,t-1}) \times W_{h,t-1}, & \eta_{h,t-1} \leq 0 \\ (1 - 1.30 \times \eta_{h,t-1}) \times W_{h,t-1}, & \eta_{h,t-1} > 0 \end{cases} \quad (23)$$

The multi-objective optimization mathematical model established based on the above three elements, after being solved by MOPSO, effectively addresses the joint optimization problem of well placement and well control parameters. This method achieves optimal results by considering the CO₂ plume front in a manner that prioritizes safety while maximizing the effect of CO₂ flooding and sequestration.

4. Case study

In this section, we apply the proposed optimization workflow to two CO₂ flooding reservoir models to evaluate its feasibility for broader application. The first model is a 2D synthetic model used to verify the performance of the workflow in conventional reservoirs. The second model is a 3D field-scale reservoir model located in the Shengli Oilfield, chosen to assess the performance of the workflow in reservoirs with complex geological conditions. All simulations in this study are carried out using the commercial software Eclipse (version 2021.10.2, Schlumberger).

4.1. Case 1: 2D synthetic model

4.1.1. Data pre-processing and analysis for 2D synthetic model

The established AST-GraphTrans proxy-assisted optimization workflow is applied to the 2D synthetic reservoir model, and the prediction and optimization performance on conventional reservoirs is verified. The model comprises 2500 grid points, with grid increments set to 100 ft in the x and y directions and 20 ft in the z direction. The well distribution in the model follows a five-spot pattern, consisting of 4 injection wells and 9 production wells. Injection wells inject CO₂ at fixed rates constrained by injection capacity, while production wells maintain fixed bottomhole pressures for oil production. To approximate the complex geological conditions of actual reservoirs, the permeability distribution of this model exhibits significant heterogeneity. The permeability field and initial well placement are illustrated in Fig. 7. Further detailed modeling information is provided in Table 3. Leveraging the computational efficiency of 2D modeling, the production period of the model is set at 5 years, with recording intervals of 3 months, resulting in a total of 20 time steps.

The process of constructing the graph structure described above is applied to the 2D synthetic model. The CO₂ flooding connectivity coefficients are computed by utilizing Eq. (10) and subsequently

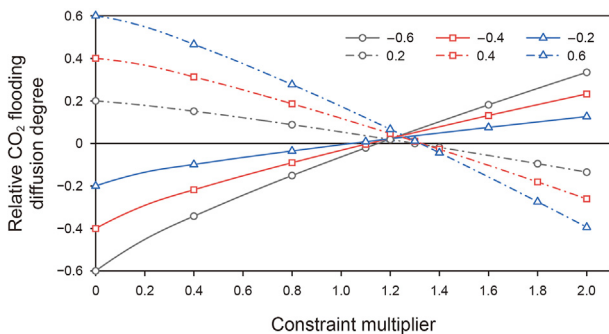


Fig. 6. The relationship between the relative CO₂ flooding diffusion degrees and the constraint multiplier under different CO₂ flooding diffusion degrees from the previous time step.

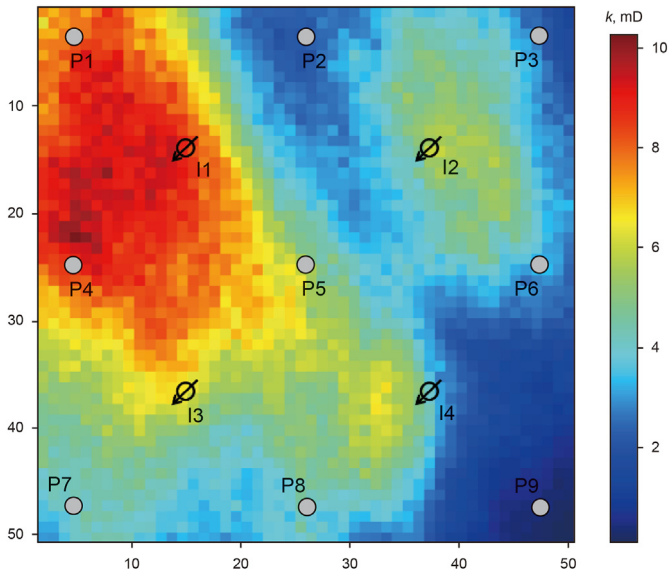


Fig. 7. Geological property and initial well placement distribution of 2D synthetic model.

organized to establish the graph structure and adjacency matrix for the 2D synthetic model, as illustrated in Fig. 8. It is evident that the interwell CO₂ flooding connectivity coefficients are influenced by geological properties and injection-production intensities, thereby effectively reflecting the CO₂ flooding capacity between injection and production wells. Consequently, the graph structure encapsulates a greater wealth of static reservoir information and dynamic production characteristics.

4.1.2. Production prediction tasks for 2D synthetic model

The proposed AST-GraphTrans-based proxy model is employed to address the production prediction task of the 2D synthetic model. This involves constructing the graph structure using the coordinates of 13 wells, their well control parameters, and CO₂ flooding connectivity coefficients. Leveraging GAT for graph structure information extraction and Transformer for temporal correlation analysis, the proposed proxy model can output cumulative oil production, CO₂ sequestration volume, and CO₂ plume front at 20 different time steps. For the training process, we selected the Stochastic Gradient Descent (SGD) optimizer. The learning rate is dynamically adjusted during training using the ReduceLROnPlateau learning rate scheduler. Its parameters are configured as follows: The factor is set to 0.5, representing the multiplicative factor for learning rate reduction. The patience parameter is set to 10, indicating the number of epochs with no improvement before reducing the learning rate. The minimum learning rate is set to 1e-6 as the lower bound. The initial learning rate is set to 0.0001. In addition, we implemented L2 regularization with a regularization parameter of 0.0001 and applied a dropout rate of 0.05. The training is conducted over 100 epochs with a batch size of 32, utilizing Mean Absolute Error (MAE) as the loss function. The proxy model of 2D synthetic model is established based on the trained AST-GraphTrans.

Table 3
Properties of the 2D synthetic model.

Depth, ft	Porosity	Initial pressure, psi	Initial water saturation	Oil viscosity, cP
10000	0.3	4500	0.25	2.2

To assess the predictive accuracy of the AST-GraphTrans-based proxy model, we randomly select 50 simulations from the testing set for comparison between prediction and simulation results. Each simulation run consists of 20 time steps, resulting in a total of 1000 samples for validation analysis. The comparison between prediction and simulation results of cumulative oil production and CO₂ sequestration volume for the 50 simulation runs is illustrated in Figs. 9 and 10, respectively. The proxy model demonstrates considerable predictive accuracy for both indexes within the 2D synthetic model. Notably, it shows slightly higher accuracy in predicting cumulative oil production compared to CO₂ sequestration volume. Indeed, the higher predictive accuracy for cumulative oil production can be attributed to the typical development trend, characterized by rapid initial increases followed by gradual stabilization. This trend is effectively captured by the proxy model, facilitating accurate predictions. However, fluctuating trends in CO₂ sequestration volume stemming from differences in injection rates may result in slightly diminished predictive performance.

Furthermore, in order to verify the superiority of AST-GraphTrans-based proxy model in handling this prediction task in the 2D synthetic model, we perform a comparative analysis with various deep learning methods. Specifically, we utilize the base learners of this model (GAT and Transformer) as a comparison. As mentioned above, we introduce R^2 and MRE to evaluate the predictive performance of these methods for cumulative oil production and CO₂ sequestration volume. Table 4 showcases the predictive performance of various deep learning methods acting as proxy models. Despite the incorporation of extensive spatial sequential data into the inputs, the predicted cumulative oil production and CO₂ sequestration volume distinctly display temporal patterns. Additionally, owing to the simplified spatial characteristics of the 2D synthetic model compared to its 3D model, the influence of reservoir spatial information on this prediction task is diminished.

To ensure a fair comparison, we created two hybrid models by combining Convolutional Neural Network (CNN) with Long Short-Term Memory (LSTM), and Graph Convolutional Network (GCN) with LSTM for comparison with ST-ResNet. The shared fundamental parameters across these models were configured identically. Through evaluative metric analysis of CNN-LSTM, GCN-LSTM, and ST-ResNet, we observed several key insights: Both CNN-LSTM and GCN-LSTM models effectively analyze reservoir attributes and spatial geological information through convolutional operations, simultaneously utilizing LSTM to extract temporal sequence features of engineering parameters and achieve information fusion. This approach enables more accurate capture of CO₂ migration and diffusion patterns, significantly improving prediction accuracy compared to single base learners. However, due to the processed input data structure being more conducive to graph neural networks and their variants, GCN-LSTM demonstrates superior performance in handling such datasets. Although both models fundamentally employ convolution operations, GCN-LSTM exhibits greater proficiency in processing the graph-structured data proposed in this study. Consequently, in predicting cumulative oil production and CO₂ sequestration volume, GCN-LSTM's accuracy notably surpasses that of CNN-LSTM. The AST-GraphTrans model, leveraging its multi-layer stacked Transformer architecture, outperforms LSTM-based models in processing time series output data

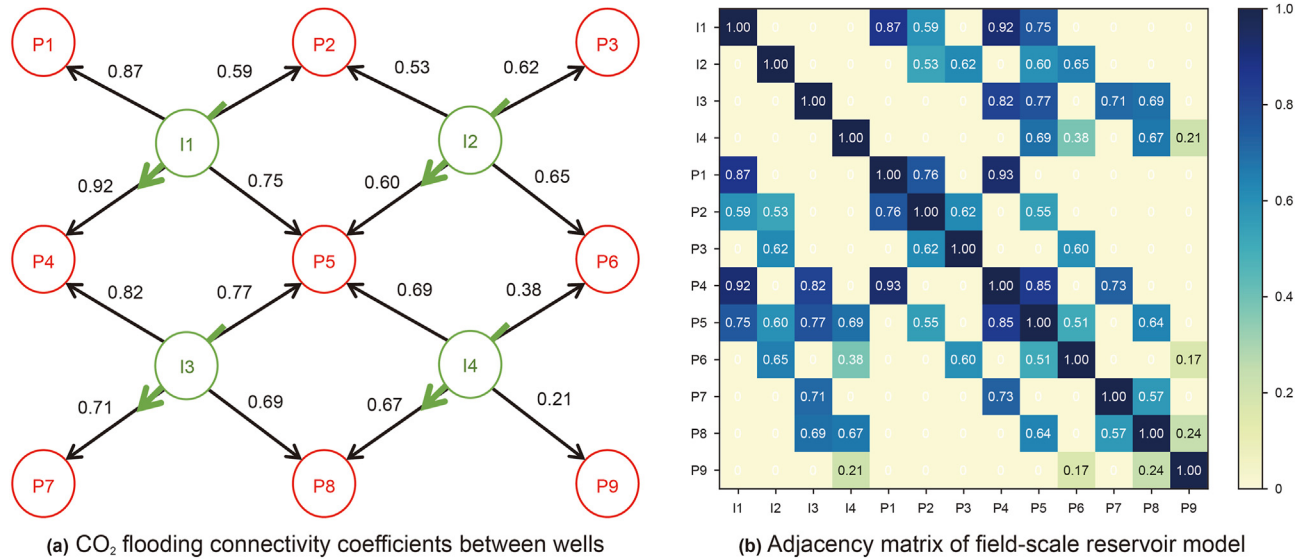


Fig. 8. CO₂ flooding connectivity coefficients and adjacency matrix of 2D synthetic model.

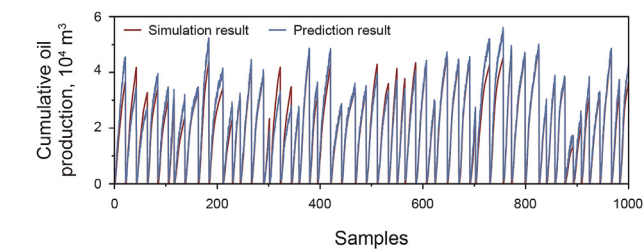


Fig. 9. Comparison between the prediction and simulation results for cumulative oil production in 2D synthetic model.

such as cumulative oil production and CO₂ sequestration volume. Moreover, the multi-head attention mechanism in Transformer enables simultaneous consideration of feature correlations across different time steps and spatial positions, achieving more comprehensive spatio-temporal dependency modeling. It is an advantage unattainable by LSTM with its gated structure. As a result, the Transformer model demonstrates superior performance over the GAT, yielding more satisfactory outcomes. However, owing to its robust capability in handling diverse data types, the AST-GraphTrans-based proxy model significantly outperforms other deep learning methods in this production prediction task. Consequently, the AST-GraphTrans proposed in this paper can effectively serve as a proxy model to aid subsequent optimization processes.

Furthermore, a specific scheme is chosen from the testing set to assess the performance of the proxy model in predicting the

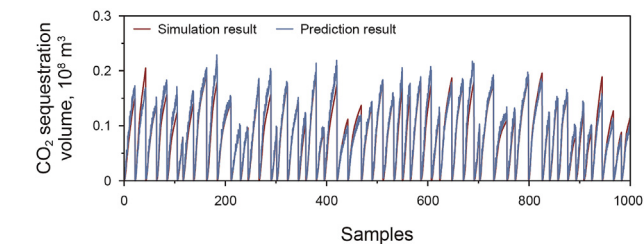


Fig. 10. Comparison between the prediction and simulation results for CO₂ sequestration volume in 2D synthetic model.

subsurface CO₂ plume front. The prediction and simulation results of the subsurface CO₂ plume front at time steps 10 and 20 are illustrated in Fig. 11. Additionally, the Structural Similarity Index (SSIM) is employed as an evaluation metric to assess the similarity between the prediction and simulation results (Sara et al., 2019). The equation of SSIM is shown below

$$SSIM(x,y)=\frac{\left(2\mu_x\mu_y+c_1\right)\left(2\sigma_{xy}+c_2\right)}{\left(\mu_x^2+\mu_y^2+c_1\right)\left(\sigma_x^2+\sigma_y^2+c_2\right)}$$

(24)

where x and y are the data bodies transformed into prediction and simulation images, respectively; μ_x and μ_y are their respective means; σ_x and σ_y are their respective variances; σ_{xy} is the covariance between x and y ; c_1 and c_2 are constant. Generally, a higher SSIM value indicates greater similarity between two images. Since the 2D synthetic model only consists of a single layer grid in the vertical direction, a comparative analysis across different layers is not conducted. It can be observed that the SSIM increases with the progression of time steps. This is mainly attributed to the influence of previous time step predictions on the contour of the current CO₂ plume front. And the temporal correlation is effectively captured by the Transformer model, which leads to continuous enhancement in prediction accuracy.

4.1.3. Multi-objective optimization tasks for 2D synthetic model

Based on the geological property and well pattern of the 2D synthetic model, the objective function, decision variables, and constraint conditions of the model are determined. Through the utilization of the MOPSO algorithm, joint optimization process

Table 4

Evaluation of prediction effect of proxy model in 2D synthetic model.

Model	Cumulative oil production		CO ₂ sequestration volume	
	R ²	MRE	R ²	MRE
GAT	0.879	0.244	0.885	0.204
Transformer	0.950	0.142	0.931	0.167
CNN-LSTM	0.955	0.140	0.941	0.159
GCN-LSTM	0.969	0.128	0.949	0.155
AST-GraphTrans	0.973	0.102	0.957	0.135

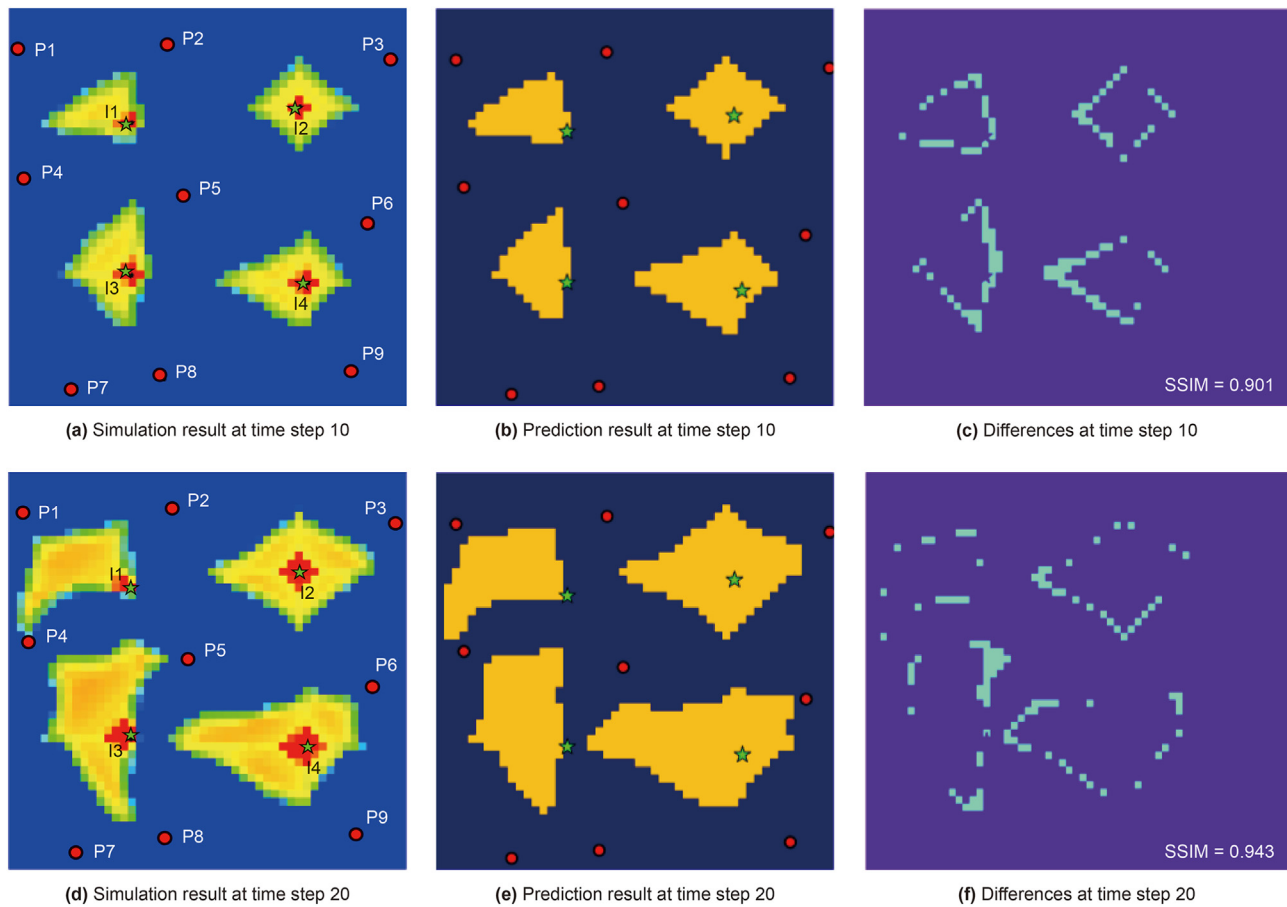


Fig. 11. Comparison of the simulation results and prediction results of CO₂ plume front, as well as the corresponding differences at different time steps in 2D synthetic model.

involving both well placement and well control parameters is undertaken to assess the reliability of the proposed optimization workflow in conventional reservoir models. A standard five-spot well pattern is employed in the model, with constant well control parameters (injection rates for injection wells and bottomhole pressures for production wells) set as a comparative benchmark. The basic parameters of MOPSO algorithm are set as follows: the number of particles is 100; the number of iterations is 200; inertia weight is 1.2; the acceleration coefficient is 1.4. The Pareto front of the 2D synthetic model is obtained by solving the established multi-objective optimization mathematical model using the MOPSO algorithm, as illustrated in Fig. 12.

In condition, the constraints for the decision variables are also important in the optimization problem. The constraint range of decision variables used in this optimization process is shown in Table 5.

In this Pareto front, it encompasses the optimal outcomes achievable by balancing multiple objectives under the current algorithm parameter settings. As decision-makers, we select schemes with the same cumulative oil production as the benchmark (referred to as the productivity scheme) and schemes with the same CO₂ sequestration volume (referred to as the sequestration scheme) for comparative analysis of optimization effectiveness. In the Pareto front, the region to the left of the sequestration scheme represents the FGST-dominated region, while the region below the productivity scheme represents the FOPT-dominated region. The performance of these mentioned schemes in terms of cumulative oil production and CO₂ sequestration volume is illustrated in Table 6 and Fig. 13.

The optimization results underscore the advantages of the workflow proposed in this study. The two obtained optimization schemes exhibit improvements in both hydrocarbon recovery and sequestration effectiveness compared to the benchmark. The productivity scheme demonstrates the same sequestration volume as the benchmark, yet achieves an increase of $2.47 \times 10^4 \text{ m}^3$ in cumulative oil production. Similarly, the sequestration scheme

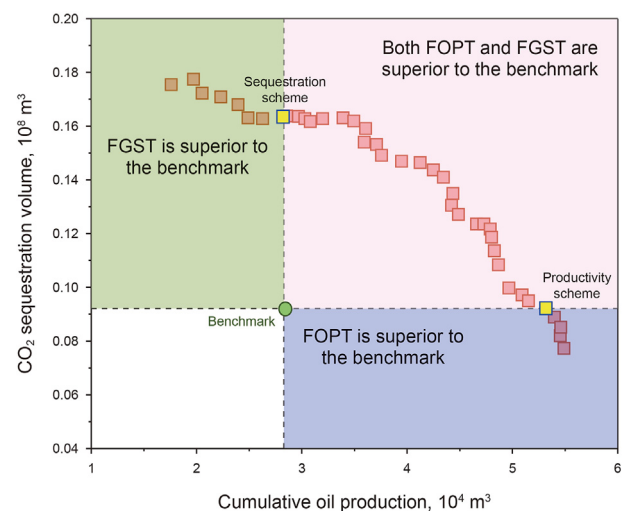


Fig. 12. Pareto front with multiple objectives in 2D synthetic model.

Table 5
Constraints for the decision variables in 2D synthetic model.

Decision variable	Constraint lower bound	Constraint upper bound
X_{Coor}	0	50
Y_{Coor}	0	50
I_{inj} , m ³ /d	2000	10000
BHP _{pro} , psi	2000	4000

matches the benchmark in cumulative oil production while increasing the sequestration volume by $7 \times 10^6 \text{ m}^3$. These two optimized schemes, through the reconfiguration of the well placement and well control parameters, are tailored to emphasize different development objectives, thereby maximizing both cumulative oil production and CO₂ sequestration volume.

Moreover, it is evident from the comparison of CO₂ plume front distributions between the optimized schemes and the benchmark, as depicted in Fig. 14, that the optimized schemes demonstrate smoother and more balanced CO₂ plume fronts. Specifically, the sequestration scheme only conducts constrained optimization for the I2 and I4 well groups (i.e., the two well groups on the right side in Fig. 14(b)), while optimization for the I1 and I3 well groups does not consider anti-gas channeling constraints. In contrast, the productivity scheme encompasses optimization for all well groups, taking into account this constraint. Comparing Fig. 14(a) and (b) reveals that the sequestration scheme achieves a greater CO₂ plume area and more CO₂ remained underground by dynamically adjusting well control parameters to enable collaboration between injection and production wells. Furthermore, optimization considering anti-gas channeling constraints results in a more balanced and stable distribution of CO₂ underground for the I2 and I4 well groups compared to the I1 and I3 well groups. Comparative

Table 6
Comparison of the results of the benchmark and optimization schemes in 2D synthrtic model.

Scheme	Cumulative oil production, 10 ⁴ m ³	CO ₂ sequestration volume, 10 ⁸ m ³	Variation
Benchmark	2.85	0.09	—
Sequestration scheme	2.84	0.16	FGST ↑
Productivity scheme	5.32	0.09	FOPT ↑

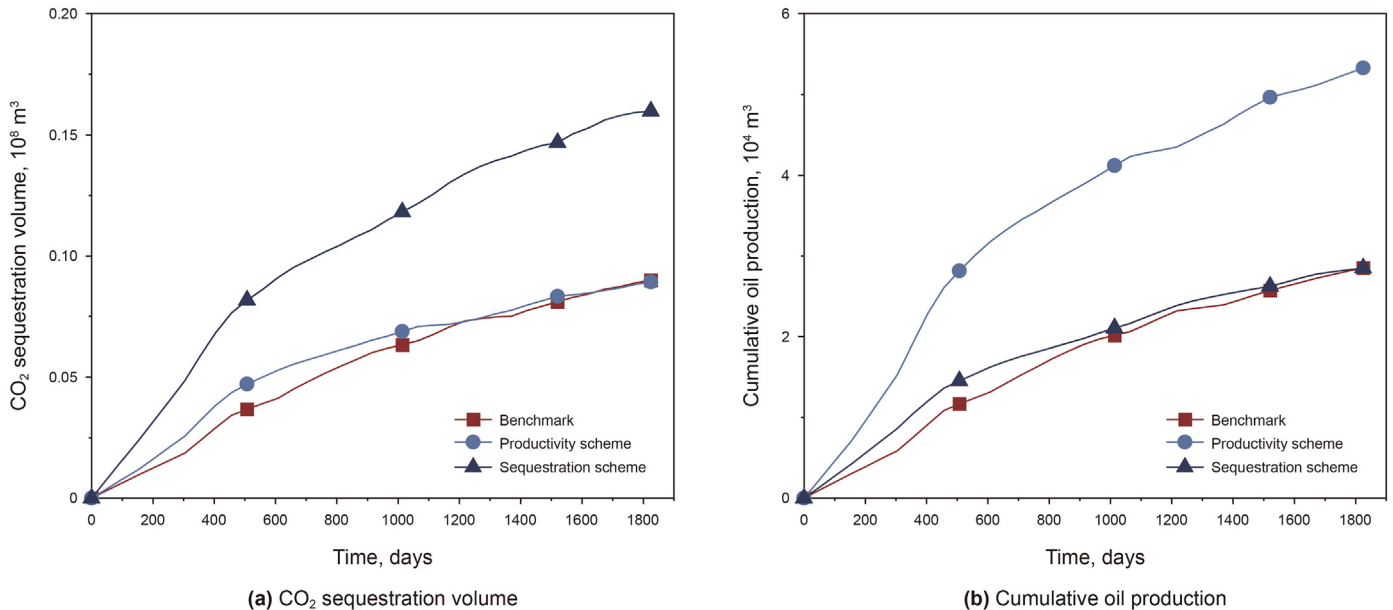


Fig. 13. Comparison of the multiple objectives of the benchmark and optimization schemes in 2D synthetic model.

analysis of Fig. 14(a) and (c) reveals that the primary challenge to achieve higher productivity in the productivity scheme is gas channeling. Premature gas channeling in production wells leads to a rapid decline in oil production. Therefore, each production well avoids significant gas channeling by rearranging well placement and well control parameters, resulting in increased production. Optimization considering anti-gas channeling constraints is implemented for all well groups in the productivity scheme, resulting in a smoother distribution of CO₂ underground compared to the other two schemes. This achieves the goal of balanced displacement, demonstrating that the proposed optimization workflow in this study can enhanced hydrocarbon recovery and CO₂ sequestration effectiveness while preventing gas channeling.

4.2. Case 2: 3D field-scale reservoir model

4.2.1. Data pre-processing and analysis for 3D field-scale reservoir model

The AST-GraphTrans proxy-assisted optimization workflow is applied to validate its prediction and optimization performance under complex geological conditions on a field-scale reservoir model. This 3D model comprises a total of 126,500 grid points, with grid increments in the x, y, and z directions set at 130, 130, and 5 ft, respectively. Furthermore, the model encompasses seven well groups, consisting of 7 injection wells and 13 production wells. The injection wells maintain fixed injection rates, while the production wells operate at constant bottomhole pressures. Fig. 15 illustrates the permeability, porosity, and initial oil saturation of the model, while detailed geological properties are presented in Table 7. The entirety of the production period spans 3 years, with a recording interval set at 3 months, resulting in a total of 12 time steps.

The application of the graph structure is extended to the field-

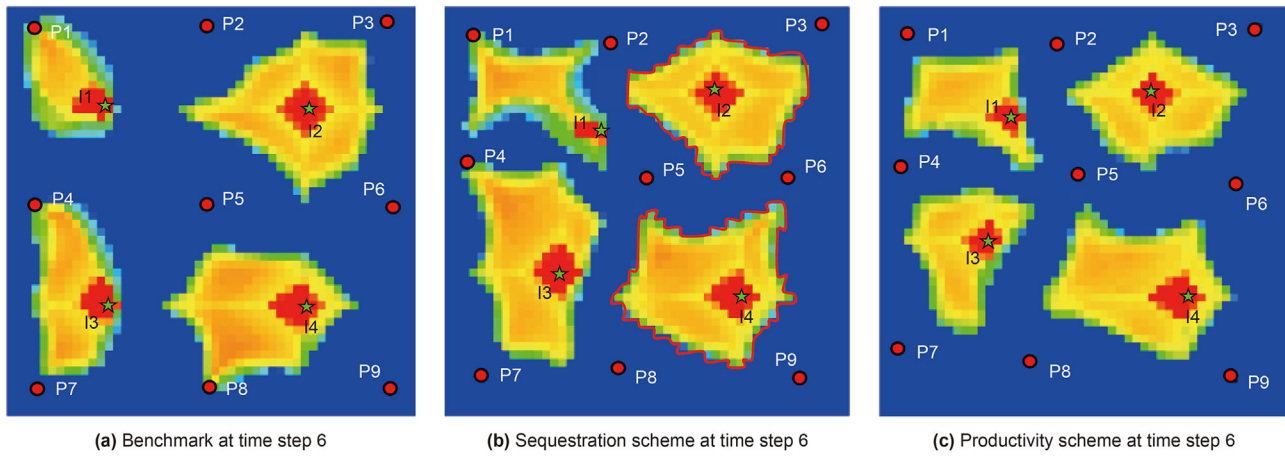


Fig. 14. Comparison of CO₂ saturation distribution for the benchmark and optimization schemes in 2D synthetic model.

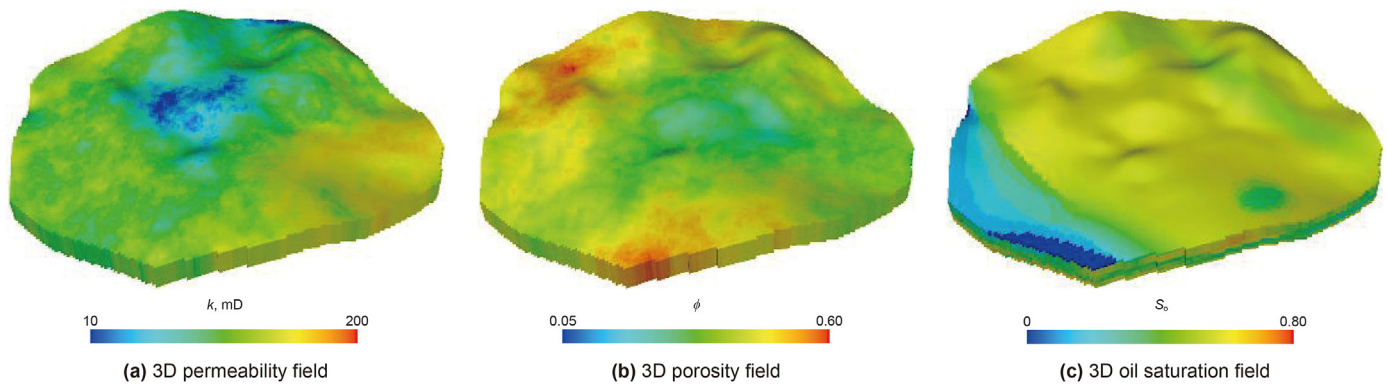


Fig. 15. Geological property distribution of 3D field-scale reservoir model.

Table 7
Properties of the 3D field-scale reservoir model.

Depth, ft	Porosity	Initial pressure, psi	Oil viscosity, cP
12500	0.15–0.55	5600	2.4

scale reservoir model utilized in this study, with its distribution of initial well placement depicted in Fig. 16. Based on the distances between injection and production wells, the preliminary construction of the graph structure for this model is illustrated in Fig. 16(b). From the illustration, it becomes apparent that the graph structure exhibits high flexibility in connecting wells. Production wells that are notably affected by the proximity of the injection well are typically linked to the injection well node (e.g., I1 and P1, P3), represented by edges. Additionally, wells not directly connected by edges but linked through a shared node may also interfere each other (e.g., P3 and P8 indirectly connected through I1).

After the connection relationship between different wells is defined, it is necessary to further determine the strength of the edge which is characterized by the CO₂ flooding connectivity coefficients. This coefficient is calculated by Eq. (10) for the field-scale reservoir model is depicted in Fig. 17(a), resulting in the adjacency matrix shown in Fig. 17(b). Notably, this matrix is symmetric.

4.2.2. Production prediction tasks for 3D field-scale reservoir model

The AST-GraphTrans proposed in this study is served as the proxy model to tackle the production prediction task of the field-

scale reservoir model. The input data comprises the CO₂ flooding connectivity relationships of 20 wells (used to construct the adjacency matrix) and the well control parameters of these 20 wells (the bottomhole pressures of 13 production wells and the injection rates of 7 injection wells, utilized for building the dynamic node feature matrix). The output data includes cumulative oil production, CO₂ sequestration volume, and the CO₂ plume front at various time steps. For the training process, we also selected the SGD optimizer to better handle the increased complexity of the 3D field-scale reservoir model. The ReduceLRonPlateau learning rate scheduler is configured to dynamically adjust the learning rate during training. Unlike the 2D synthetic model, we increased the patience parameter from 10 to 15 epochs. The initial learning rate is set to 0.0001, with the total number of training epochs established at 200 to allow more time for convergence. Additionally, we reduced the batch size from 32 to 12 to accommodate the larger memory requirements of 3D field-scale reservoir data. The loss function is set as the MAE.

Furthermore, to validate the disparities between the proxy model and numerical simulation methods, a comparative analysis is conducted on 50 selected schemes from the testing set. Each data point represents a sample, denoting the results corresponding to a specific time step, totaling 600 samples. The comparison between the prediction and simulation results for cumulative oil production and CO₂ sequestration volume in these 50 schemes is illustrated in Figs. 18 and 19. This demonstration showcases that during the testing process, the proxy model adeptly fits the temporal evolution curves of cumulative oil production and CO₂ sequestration

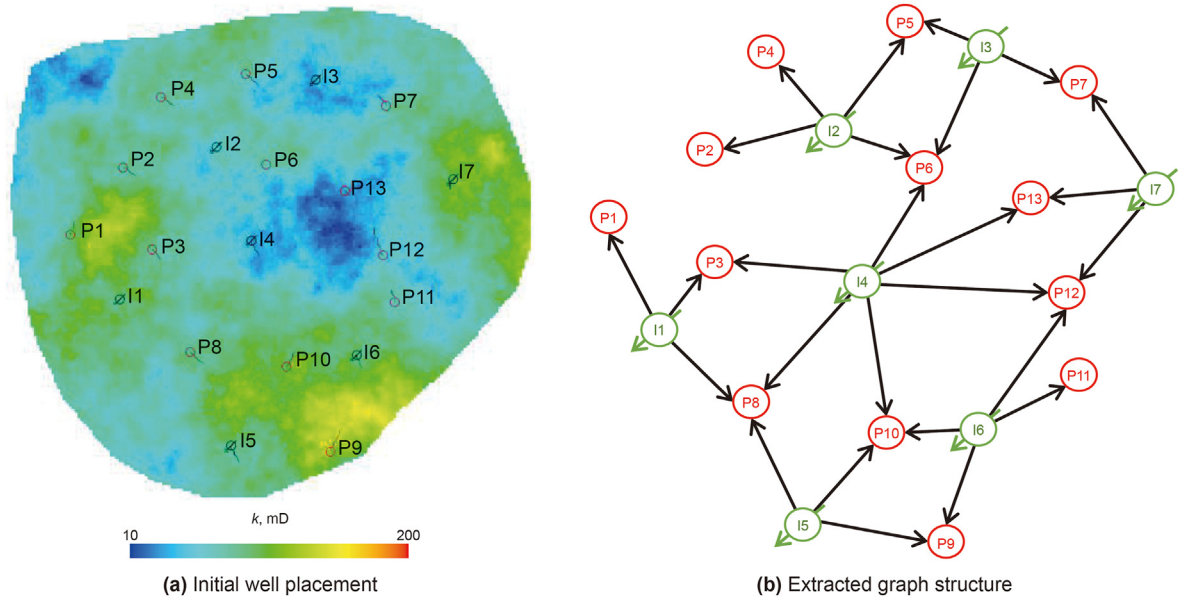


Fig. 16. Well placement distribution and extracted graph structure of 3D field-scale reservoir model.

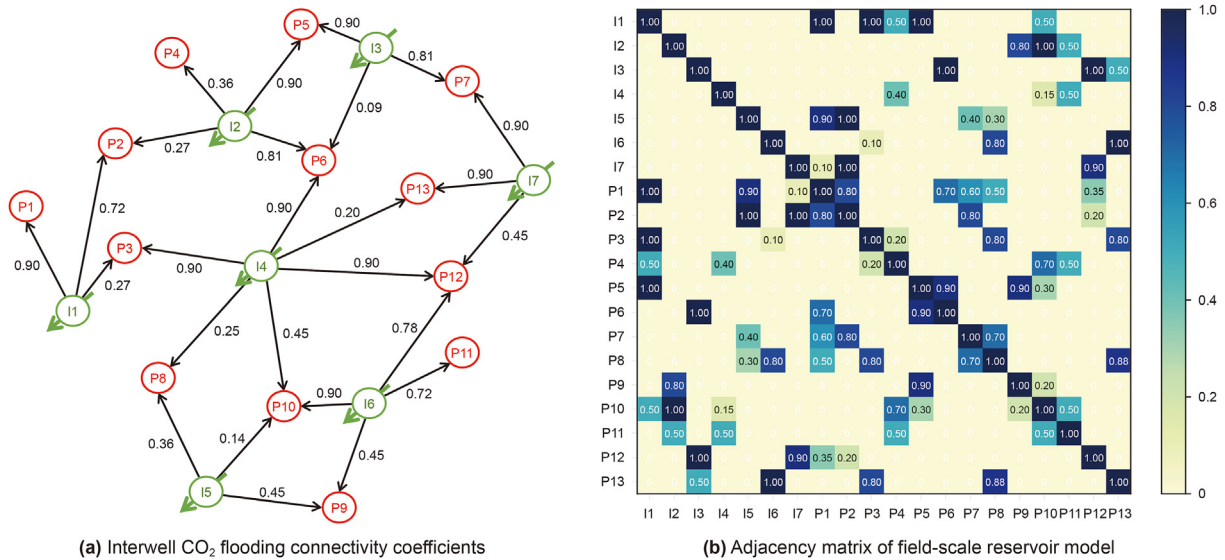


Fig. 17. CO₂ flooding connectivity coefficients and adjacency matrix of 3D field-scale reservoir model.

volume obtained through numerical simulation. The errors between prediction and simulation results fall within a reasonable range, indicating that the proxy model constructed in this case can

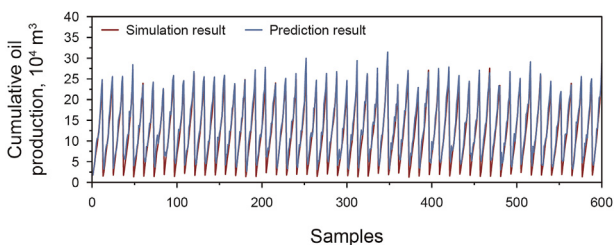


Fig. 18. Comparison between the prediction and simulation results for cumulative oil production in 3D field-scale reservoir model.

effectively forecast the development effect of CO₂ flooding under complex geological and well pattern conditions.

Furthermore, to demonstrate the advantages of AST-GraphTrans in handling the production prediction task on a 3D field-scale reservoir, GAT and Transformer are selected as comparison models, as in Case 1. Additionally, we employ R^2 and MRE to assess the comprehensive performance of various deep learning methods in predicting cumulative oil production and CO₂ sequestration volume. Table 8 shows the predictive performance of various deep learning methods as proxy models. Similar to their performance on the 2D synthetic model, base learners such as GAT and Transformer exhibit unique strengths in handling spatial and temporal sequence data, respectively. However, they cannot simultaneously extract features from both types of data. The proposed AST-GraphTrans-based proxy model, on the other hand, leverages graph structures to extract spatial information of well patterns and geological

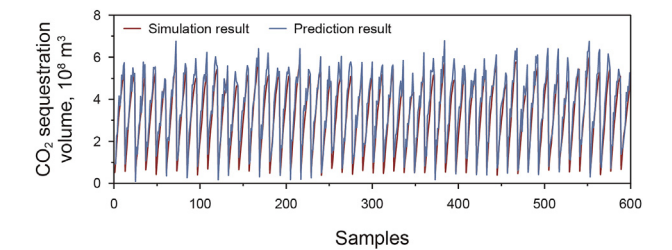


Fig. 19. Comparison between the prediction and simulation results for CO₂ sequestration volume in 3D field-scale reservoir model.

Table 8
Evaluation of prediction effect of proxy model in 3D field-scale reservoir model.

Model	Cumulative oil production		CO ₂ sequestration volume	
	R ²	MRE	R ²	MRE
GAT	0.845	0.272	0.842	0.277
Transformer	0.911	0.198	0.855	0.250
CNN-LSTM	0.915	0.196	0.866	0.231
GCN-LSTM	0.933	0.190	0.869	0.226
AST-GraphTrans	0.936	0.187	0.872	0.202

properties, while employing an encoder architecture to analyze the temporal correlation of reservoir production dynamics, thereby achieving more precise predictions of cumulative oil production and CO₂ sequestration volume. Compared to GAT and Transformer, the AST-GraphTrans-based proxy model demonstrates superior performance due to its better generalization and robustness in predicting the dynamic CO₂ flooding and sequestration processes involving high-dimensional reservoir information. We conducted the comparative analysis of the AST-GraphTrans against CNN-LSTM and GCN-LSTM models in terms of predictive performance. The inherent spatial complexity of the 3D field-scale model progressively elevates the significance of geological spatial information. Given that GCN and GAT demonstrate exceptional capabilities in processing graph-structured data rich in spatial information, their evaluation metrics significantly outperform CNN-based models. Furthermore, considering the Transformer's superior handling of temporal sequence data compared to LSTM, AST-GraphTrans remains the top-performing model.

For the prediction of the CO₂ plume front, we randomly select one schemes from the testing set and conducted predictions using both the proxy model and numerical simulation. The simulation and prediction results at various time steps and layers are depicted in Fig. 20. Additionally, to quantitatively analyze the disparities between the simulation and prediction results of the CO₂ plume front, we employ the SSIM for evaluation. It is evident that as the time steps increase, SSIM gradually rises, affirming that the proxy model adeptly captures the dynamic changes in the CO₂ plume front at various time steps. Furthermore, considering geological, fluid, and well control parameters, the proxy model exhibits accurate prediction regarding the shape of the CO₂ plume front. In practical engineering applications, this proxy model demonstrates the ability to accurately predict the distribution of CO₂ plumes among various injection and production wells, leveraging intricate geological and well pattern information. Consequently, it offers valuable guidance for on-site gas channeling warning and prevention in oilfield operations.

4.2.3. Multi-objective optimization tasks for 3D field-scale reservoir model

Taking into consideration the three core elements mentioned in Section 3.4, we establish the multi-objective optimization mathematical model for 3D field-scale reservoir model. The joint optimization of both well placement and well control parameters for this model is carried out through the utilization of the MOPSO algorithm. Prior to initiating the optimization process, a series of baseline scenarios comprising the coordinates and control parameters for each well are established as a comparative benchmark. The basic parameters of MOPSO algorithm are set as follows: the number of particles is 100; the number of iterations is 100; inertia weight is 0.8; the acceleration coefficient is 2.0. The Pareto front resulting from the MOPSO optimization process is depicted in Fig. 21. The constraint range of decision variables used in this optimization process is shown in Table 9.

Following the screening process, the resulting Pareto front is represented as a relatively smooth curve, comprising 16 Pareto optimal solutions. The cumulative oil production among these optimal solutions ranges from 10×10^4 to 33×10^4 m³, while the CO₂ sequestration volume is centered between 2.5×10^8 and 7.0×10^8 m³. Each solution within the Pareto optimal set, derived through MOPSO optimization, can be considered as genuinely optimal. In practical engineering applications, decision-makers can selectively choose the well placement and well control parameter combination from the Pareto front that aligns with their specific needs. This selection is based on achieving the desired engineering objectives, whether emphasizing hydrocarbon recovery or sequestration efficiency. Referring to the benchmark, we selected schemes from the Pareto front with equivalent CO₂ sequestration volumes (referred to as the productivity scheme) and equivalent cumulative oil production (referred to as the sequestration scheme). A detailed comparison of different optimization schemes in terms of objective function values is presented in Table 10 and Fig. 22.

The productivity scheme demonstrates a notable increase in cumulative oil production, surpassing the benchmark by 10.12×10^4 m³, while maintaining the same CO₂ sequestration volume. This enhancement stems from strategic adjustments in well placement and the optimization of injection and production relationships. These adjustments enlarge the CO₂ sweep area and promote a greater intermixing with oil, thereby enhancing hydrocarbon recovery. If reservoir engineers prioritize higher production, the productivity scheme would meet their requirements.

Similarly, the sequestration scheme achieves the same cumulative oil production as the benchmark but exhibits a larger CO₂ sequestration volume, elevated by 2.17×10^8 m³. This outcome is attributed to the multi-objective optimization of CO₂-EOR strategies, facilitating the extensive underground CO₂ plume. While ensuring the same production, this scheme effectively mitigates CO₂ escape through dominant channels, thereby increasing the volumetric fraction of CO₂ retained underground.

As a result of incorporating intricate anti-gas channeling constraints into this optimization process, the shape of the CO₂ plume front becomes more balanced across various time steps, thereby delaying gas channeling in production wells. A comparison of the CO₂ plume front distribution for the benchmark, sequestration scheme, and productivity scheme at time steps 6 and 12 is illustrated in Fig. 23. Each CO₂ plume front distribution depicted in the figures is derived from numerical simulations. These clearly illustrate that both the productivity and sequestration schemes achieve a more balanced CO₂ plume front shape following constrained

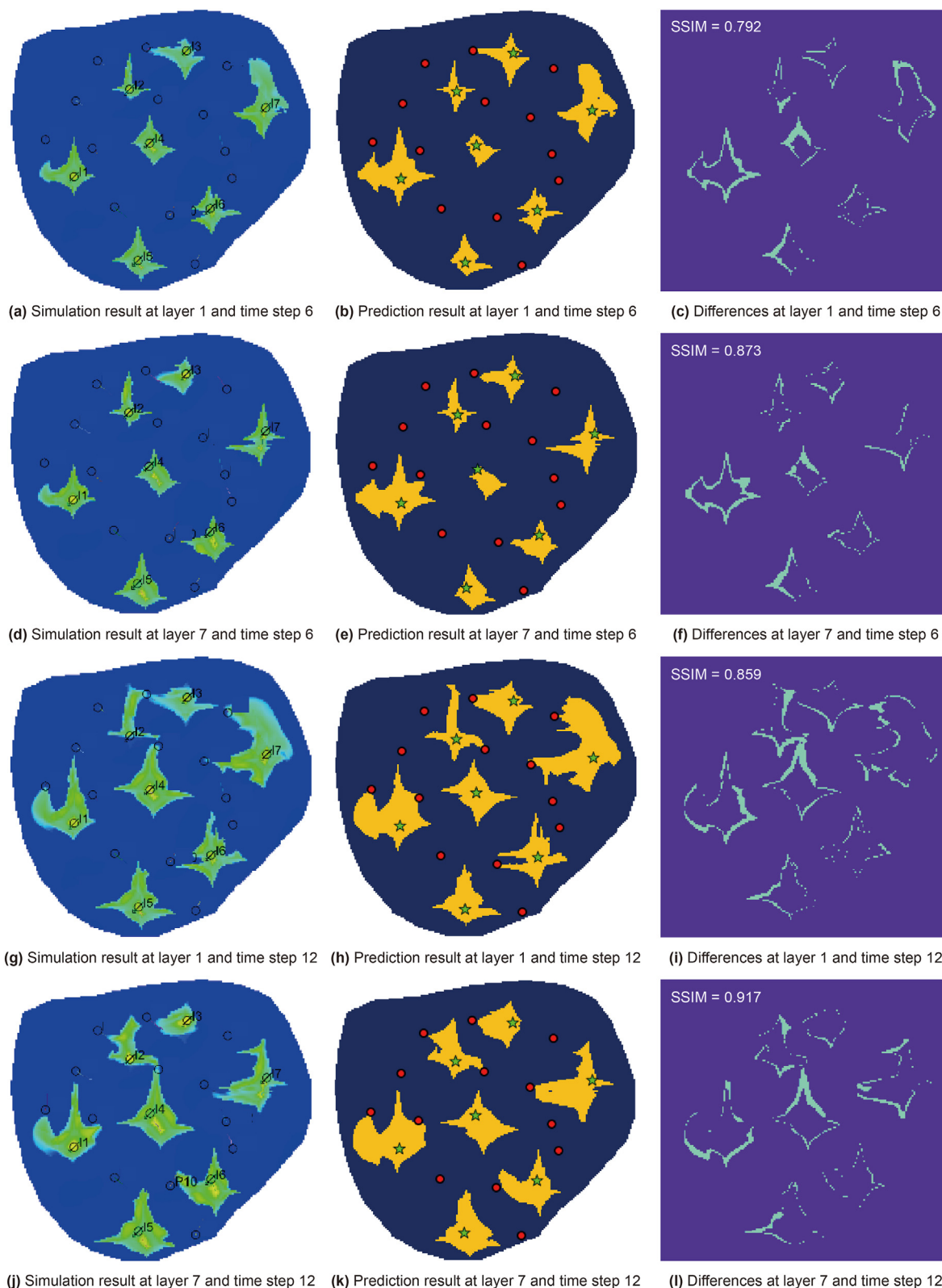


Fig. 20. Comparison of the simulation and prediction results of CO₂ plume front, as well as the corresponding differences at different time steps and layers in 3D field-scale reservoir model.

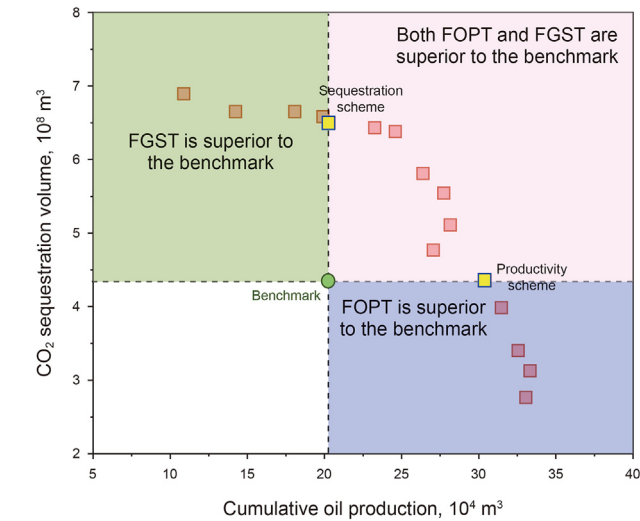


Fig. 21. Pareto front with multiple objectives in 3D field-scale reservoir model.

Table 9
Constraints for the decision variables in 3D field-scale reservoir model.

Decision variable	Constraint lower bound	Constraint upper bound
X_{Coor}	0	178
Y_{Coor}	0	167
I_{inj} , m ³ /day	2×10^4	12×10^4
BHP_{pro} , psi	2500	5000

Table 10
Comparison of the results of the benchmark and optimization schemes in 3D field-scale reservoir model.

Scheme	Cumulative oil production, 10 ⁴ m ³	CO ₂ sequestration volume, 10 ⁸ m ³	Variation
Benchmark	20.29	4.34	—
Sequestration scheme	20.30	6.51	FGST ↑
Productivity scheme	30.41	4.34	FOPT ↑

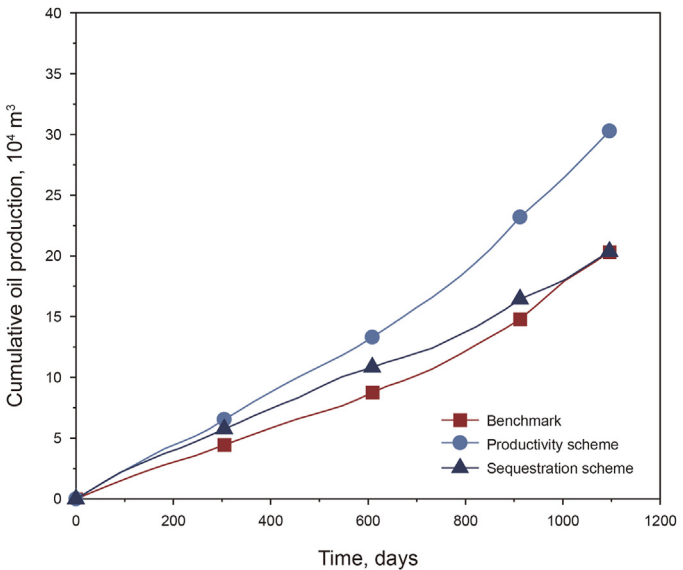
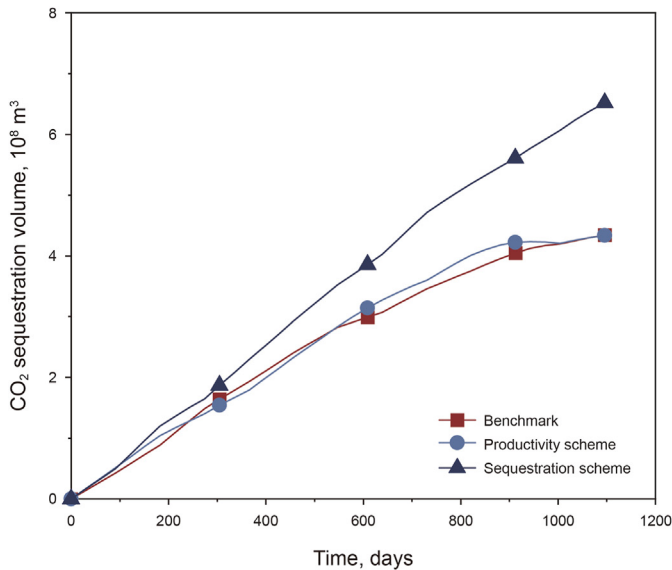


Fig. 22. Comparison of the multiple objectives of the benchmark and optimization schemes in 3D field-scale reservoir model.

optimization under the specified conditions elucidated in this paper. Furthermore, the gas channeling time in each production well is prolonged. In Fig. 23(a), it is apparent that in the benchmark at the time step 6, the CO₂ plume front of I7 is approaching P7 and P13 wells. This indicates that P7 and P13 wells are on the verge of experiencing channeling by CO₂, potentially resulting in a significant increase in CO₂ production. Conversely, in both the productivity and sequestration schemes at time step 6, the CO₂ plume front of I1 has not yet reached P7 and P13 wells. Particularly in the sequestration scheme, the CO₂ plume front between I7 and P7, as well as P13, has only covered approximately half of their well spacing. This effectively restrains the advancement of the CO₂ plume front compared to the base scheme, thereby delaying the gas channeling time for P1. This comparison underscores the effectiveness of the proposed optimization workflow in achieving a more judicious distribution of subsurface CO₂, thereby enhancing both hydrocarbon recovery and sequestration efficiency. Moreover, it has the capacity to delay the onset of gas channeling, offering valuable technical insights for the design of well placement and well control parameters in CO₂ flooding operations.

Based on the multidimensional characteristics of data involved in CO₂ flooding and sequestration, we design an optimization workflow that integrates a proxy model for spatio-temporal sequence prediction and a multi-objective optimization algorithm considering anti-gas channeling constraints in CO₂ EOR strategies. The successful performance of the workflow across various types of reservoirs demonstrates its capability to effectively tackle the prediction and optimization tasks inherent in CO₂-EOR technology.

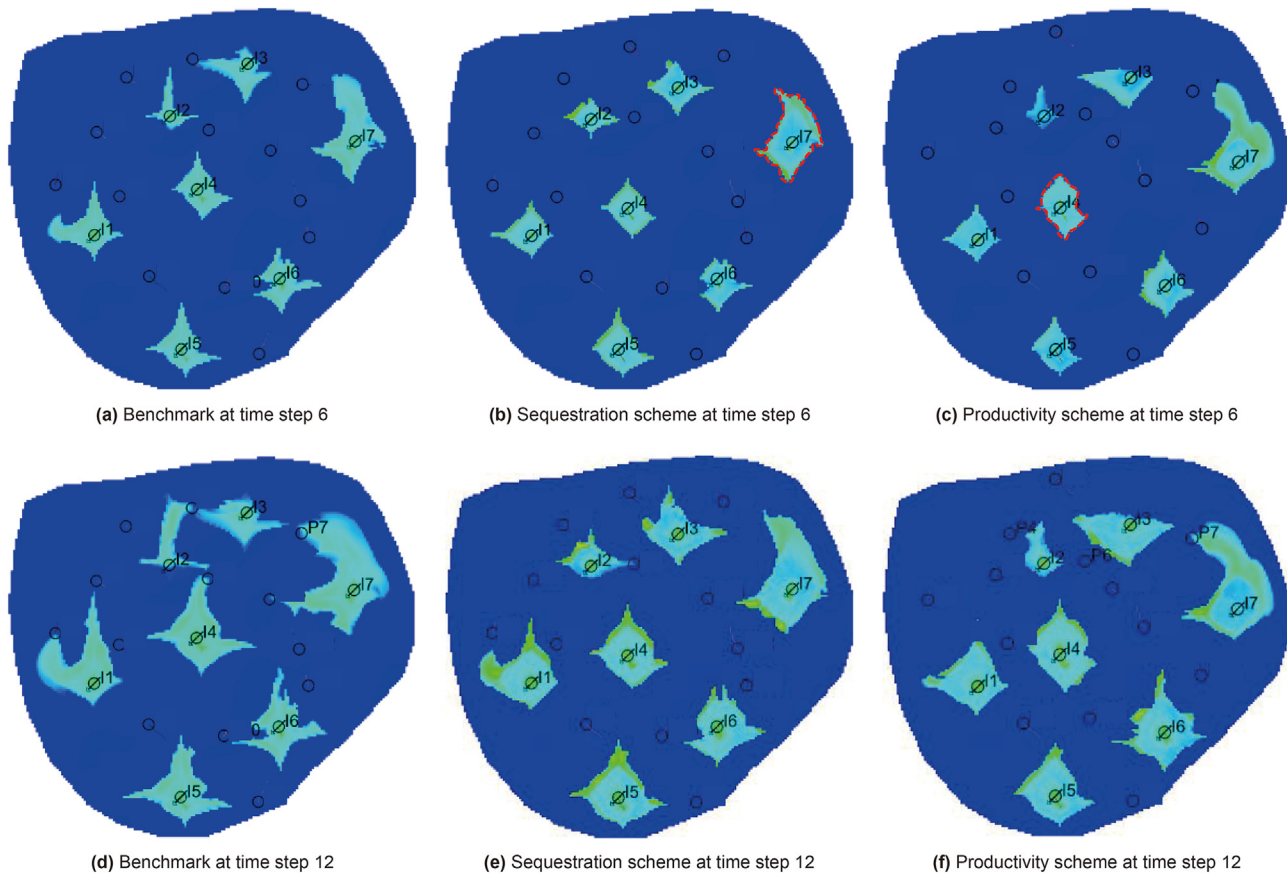


Fig. 23. Comparison of CO₂ saturation distribution for the benchmark and optimization schemes at different time steps in 3D field-scale reservoir model.

5. Summary and conclusion

The CO₂-EOR technology ensures substantial underground CO₂ sequestration while also boosting the production capacity of subsurface hydrocarbons. Rational well placement and well control parameters are crucial for the successful implementation of CO₂-EOR technology. Traditional optimization workflows fail to simultaneously coordinate the impacts of well placement and well control parameters on the effectiveness of CO₂ flooding and sequestration. These workflows typically do not take into account high-dimensional reservoir information and dynamic production characteristics, limiting their effectiveness in optimizing for enhanced oil recovery and carbon storage. Moreover, most existing optimization workflows overlook the detrimental effects of gas channeling on production. Therefore, a novel deep learning proxy-assisted optimization workflow is proposed for the collaborative optimization of CO₂-EOR strategies. This workflow primarily encompasses three exciting aspects. Firstly, the constructed graph structure takes into account the intricate well distribution and dynamic well control parameters, proficiently characterizes the spatio-temporal correlation among different wells within the CO₂ flooding reservoir. It encapsulates the spatio-temporal evolution patterns of injection and production behavior, thereby endowing the proxy model with a more comprehensive understanding of reservoir dynamics during its construction. Subsequently, the AST-GraphTrans-based proxy model employs GAT and Transformer modules to accurately extract multidimensional features, accounting for geological and engineering uncertainties. This model

enables swift and responsive prediction of the spatio-temporal sequences of well placement and well control parameters across various time steps, facilitating dynamic forecasting of CO₂ flooding and sequestration. Finally, a novel proxy-assisted optimization workflow is introduced, integrating the AST-GraphTrans-based proxy model with multi-objective optimization algorithms. This workflow not only facilitates the training of a proxy model capable of substituting physics-based simulators but also tackles joint optimization challenges in CO₂-EOR strategies. The effectiveness of the proposed workflow is validated on a 2D synthetic model and a 3D field-scale reservoir model. The results demonstrate significant increases in cumulative oil production by 87% and 49%, and CO₂ sequestration volume enhancement by 78% and 50% across various reservoirs. This ensures the maximization of cumulative oil production and CO₂ sequestration volume while constraining and mitigating gas channeling. Thus, this study provides a more efficient prediction and optimization tool that maximizes CO₂ sequestration and oil recovery while mitigating gas channeling, thereby ensuring cleaner oil production. However, the present research has not explored the influence of various CO₂-EOR strategies on economic benefits such as NPV. Therefore, it lacks the ability to handle three or more objectives simultaneously in optimizing well placement and control parameters, which may make the optimization time much longer. This aspect holds significant importance for the comprehensive evaluation of CO₂-EOR projects. In addition, this workflow has been exclusively validated in medium and high permeability reservoir environments currently, demonstrating promising performance. However, the methodology

remains untested for compatibility with more complex geological formations, including unconventional reservoirs and those characterized by intricate fault and fracture networks. Nonetheless, the proposed workflow represents an innovative and extensible analytical approach. By strategically re-optimizing inputs, outputs, and specific model hyperparameters, the workflow shows considerable potential for addressing prediction and optimization challenges across diverse and complex reservoir conditions. These nuanced considerations will be subject to further exploration in forthcoming investigations.

CRedit authorship contribution statement

Xin-Yu Zhuang: Writing – original draft. **Wen-Dong Wang:** Methodology, Formal analysis. **Yu-Liang Su:** Supervision, Resources. **Zhen-Xue Dai:** Writing – review & editing, Project administration. **Bi-Cheng Yan:** Software, Conceptualization.

Declaration of competing interest

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

Acknowledgement

The authors acknowledge the computation resources supported by the National Natural Science Foundation of China (Nos. 52374064, 52274056) and China Scholarship Council (No. 202406450086).

References

- Al-Khafaji, H.F., Meng, Q., Hussain, W., et al., 2023. Predicting minimum miscible pressure in pure CO₂ flooding using machine learning: Method comparison and sensitivity analysis. *Fuel* 354, 129263. <https://doi.org/10.1016/j.fuel.2023.129263>.
- Al-Mudhafar, W.J., Rao, D.N., Srinivasan, S., 2018. Robust optimization of cyclic CO₂ flooding through the gas-assisted gravity drainage process under geological uncertainties. *J. Petrol. Sci. Eng.* 166, 490–509. <https://doi.org/10.1016/j.petrol.2018.03.044>.
- Ampomah, W., Balch, R.S., Cather, M., et al., 2017a. Optimum design of CO₂ storage and oil recovery under geological uncertainty. *Appl. Energy* 195, 80–92. <https://doi.org/10.1016/j.apenergy.2017.03.017>.
- Ampomah, W., Balch, R.S., Grigg, R.B., et al., 2017b. Co-optimization of CO₂-EOR and storage processes in mature oil reservoirs. *Greenhouse Gases: Sci. Technol.* 7 (1), 128–142. <https://doi.org/10.1002/ghg.1618>.
- Carpenter, C., 2022. Transformer-based models aid prediction of transient production of oil wells. *J. Petrol. Technol.* 74 (10), 91–93. <https://doi.org/10.2118/1022-0091-JPT>.
- Chen, S., Li, H., Yang, D., 2010. Optimization of production performance in a CO₂ flooding reservoir under uncertainty. *J. Can. Petrol. Technol.* 49 (2), 71–78. <https://doi.org/10.2118/133374-PA>.
- Chen, Z., Zhang, D., Li, J., et al., 2024. Prediction of production indicators of fractured-vuggy reservoirs based on improved Graph Attention Network. *Eng. Appl. Artif. Intell.* 129, 107540. <https://doi.org/10.1016/j.engappai.2023.107540>.
- Choubey, S., Karmakar, G.P., 2021. Artificial intelligence techniques and their application in oil and gas industry. *Artif. Intell. Rev.* 54 (5), 3665–3683. <https://doi.org/10.1007/s10462-020-09935-1>.
- Dai, Z., Middleton, R., Viswanathan, H., et al., 2014. An integrated framework for optimizing CO₂ sequestration and enhanced oil recovery. *Environ. Sci. Technol. Lett.* 1 (1), 49–54. <https://doi.org/10.1021/ez4001033>.
- Ding, L., Wu, Q., Zhang, L., et al., 2020. Application of fractional flow theory for analytical modeling of surfactant flooding, polymer flooding, and surfactant/polymer flooding for chemical enhanced oil recovery. *Water* 12 (8), 2195. <https://doi.org/10.3390/w12082195>.
- Do, T.H., Nguyen, D.M., Bekoulis, G., Munteanu, A., et al., 2021. Graph convolutional neural networks with node transition probability-based message passing and DropNode regularization. *Expert Syst. Appl.* 174, 114711. <https://doi.org/10.1016/j.eswa.2021.114711>.
- Dutt, A., Mandal, A., 2012. Modified analytical model for prediction of steam flood performance. *J. Pet. Explor. Prod. Technol.* 2, 117–123. <https://doi.org/10.1007/s13202-012-0027-9>.
- Farahi, M.M.M., Ahmadi, M., Dabir, B., 2021. Model-based water-flooding optimization using multi-objective approach for efficient reservoir management. *J. Petrol. Sci. Eng.* 196, 107988. <https://doi.org/10.1016/j.petrol.2020.107988>.
- Gao, Y., Zhao, M., Wang, J., et al., 2014. Performance and gas breakthrough during CO₂ immiscible flooding in ultra-low permeability reservoirs. *Petrol. Explor. Dev.* 41 (1), 88–95. [https://doi.org/10.1016/S1876-3804\(14\)60010-0](https://doi.org/10.1016/S1876-3804(14)60010-0).
- Hao, H., Hou, J., Zhao, F., et al., 2016. Gas channeling control during CO₂ immiscible flooding in 3D radial flow model with complex fractures and heterogeneity. *J. Petrol. Sci. Eng.* 146, 890–901. <https://doi.org/10.1016/j.petrol.2016.07.034>.
- He, R., Ma, W., Ma, X., et al., 2021. Modeling and optimizing for operation of CO₂-EOR project based on machine learning methods and greedy algorithm. *Energy Rep.* 7, 3664–3677. <https://doi.org/10.1016/j.egy.2021.05.067>.
- Huang, H., Gong, B., Sun, W., 2023. A deep-learning-based graph neural network-long-short-term memory model for reservoir simulation and optimization with varying well controls. *SPE J.* 28 (6), 2898–2916. <https://doi.org/10.2118/215842-PA>.
- Ji, J., Yao, Y., Huang, S., et al., 2017. Analytical model for production performance analysis of multi-fractured horizontal well in tight oil reservoirs. *J. Petrol. Sci. Eng.* 158, 380–397. <https://doi.org/10.1016/j.petrol.2017.08.037>.
- Jiang, W., Luo, J., 2022. Graph neural network for traffic forecasting: A survey. *Expert Syst. Appl.* 207, 117921. <https://doi.org/10.1016/j.eswa.2022.117921>.
- Jiang, X., 2011. A review of physical modelling and numerical simulation of long-term geological storage of CO₂. *Appl. Energy* 88 (11), 3557–3566. <https://doi.org/10.1016/j.apenergy.2011.05.004>.
- Kamali, F., Hussain, F., Cinar, Y., 2015. A laboratory and numerical-simulation study of co-optimizing CO₂ storage and CO₂ enhanced oil recovery. *SPE J.* 20 (6), 1227–1237. <https://doi.org/10.2118/171520-PA>.
- Kashkooli, S.B., Gandomkar, A., Riazi, M., et al., 2022. Coupled optimization of carbon dioxide sequestration and CO₂ enhanced oil recovery. *J. Petrol. Sci. Eng.* 208, 109257. <https://doi.org/10.1016/j.petrol.2021.109257>.
- Li, H., Yang, Z., Li, R., et al., 2021. Mechanism of CO₂ enhanced oil recovery in shale reservoirs. *Pet. Sci.* 18 (6), 1788–1796. <https://doi.org/10.1016/j.petsci.2021.09.040>.
- Lin, X., Li, Z., Han, Y., et al., 2023. Novel spatiotemporal graph attention model for production prediction and energy structure optimization of propylene production processes. *Comput. Chem. Eng.* 181, 108507. <https://doi.org/10.1016/j.compchemeng.2023.108507>.
- Menad, N.A., Noureddine, Z., 2019. An efficient methodology for multi-objective optimization of water alternating CO₂ EOR process. *J. Taiwan Inst. Chem. Eng.* 99, 154–165. <https://doi.org/10.1016/j.jtice.2019.03.016>.
- Pan, S., Yang, B., Wang, S., et al., 2023. Oil well production prediction based on CNN-LSTM model with self-attention mechanism. *Energy* 284, 128701. <https://doi.org/10.1016/j.energy.2023.128701>.
- Pancholi, S., Negi, G.S., Agarwal, J.R., et al., 2020. Experimental and simulation studies for optimization of water-alternating-gas (CO₂) flooding for enhanced oil recovery. *Petrol. Res.* 5 (3), 227–234. <https://doi.org/10.1016/j.ptlrs.2020.04.004>.
- Panja, P., Velasco, R., Pathak, M., et al., 2018. Application of artificial intelligence to forecast hydrocarbon production from shales. *Petroleum* 4 (1), 75–89. <https://doi.org/10.1016/j.petlm.2017.11.003>.
- Qiu, R., Zhang, H., Zhou, X., et al., 2020. A multi-objective and multi-scenario optimization model for operation control of CO₂-flooding pipeline network system. *J. Clean. Prod.* 247, 119157. <https://doi.org/10.1016/j.jclepro.2019.119157>.
- Safi, R., Agarwal, R.K., Banerjee, S., 2016. Numerical simulation and optimization of CO₂ utilization for enhanced oil recovery from depleted reservoirs. *Chem. Eng. Sci.* 144, 30–38. <https://doi.org/10.1016/j.ces.2016.01.021>.
- Sara, U., Akter, M., Uddin, M.S., 2019. Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study. *J. Comput. Commun.* 7 (3), 8–18. <https://doi.org/10.4236/jcc.2019.73002>.
- Tadger, A., Hong, A., Bratvold, R., 2022. Bayesian deep decline curve analysis: A new approach for well oil production modeling and forecasting. *SPE Reservoir Eval. Eng.* 25 (3), 568–582. <https://doi.org/10.2118/209616-PA>.
- Tan, L., Zuo, L., Wang, B., 2018. Methods of decline curve analysis for shale gas reservoirs. *Energies* 11 (3), 552. <https://doi.org/10.3390/en11030552>.
- Tang, H., He, G., Ni, Y., et al., 2024. Production decline curve analysis of shale oil wells: A case study of Bakken, Eagle Ford and Permian. *Pet. Sci.* 21 (6), 4262–4277. <https://doi.org/10.1016/j.petsci.2024.07.029>.
- Thanh, H.V., Dashtgoli, D.S., Zhang, H., et al., 2023. Machine-learning-based prediction of oil recovery factor for experimental CO₂-foam chemical EOR: Implications for carbon utilization projects. *Energy* 278, 127860. <https://doi.org/10.1016/j.energy.2023.127860>.
- Van, S.L., Chon, B.H., 2018. Effective prediction and management of a CO₂ flooding process for enhancing oil recovery using artificial neural networks. *J. Energy Resour. Technol.* 140 (3), 032906. <https://doi.org/10.1115/1.4038054>.
- Vo, Thanh H., Sugai, Y., Sasaki, K., 2020. Application of artificial neural network for predicting the performance of CO₂ enhanced oil recovery and storage in residual oil zones. *Sci. Rep.* 10 (1), 18204. <https://doi.org/10.1038/s41598-020-73931-2>.
- Wang, F., Liao, G., Su, C., et al., 2023. Carbon emission reduction accounting method for a CCUS-EOR project. *Petrol. Explor. Dev.* 50 (4), 989–1000. [https://doi.org/10.1016/S1876-3804\(23\)60444-6](https://doi.org/10.1016/S1876-3804(23)60444-6).
- Wu, H., Lubbers, N., Viswanathan, H.S., et al., 2021. A multi-dimensional parametric study of variability in multi-phase flow dynamics during geologic CO₂ sequestration accelerated with machine learning. *Appl. Energy* 287, 116580. <https://doi.org/10.1016/j.apenergy.2021.116580>.
- Xiao, C., Zhang, S., Hu, Y., et al., 2024. Robust optimization of geoenery production using data-driven deep recurrent auto-encoder and fully-connected neural network proxy. *Expert Syst. Appl.* 242, 122797. <https://doi.org/10.1016/j.eswa.2024.122797>.

- j.eswa.2023.122797.
- Yan, L., Hu, J., Fang, Q., et al., 2023. Eco-development of oil and gas industry: CO₂-EOR technology. *Front. Earth Sci.* 10, 1063042. <https://doi.org/10.3389/feart.2022.1063042>.
- Yehia, T., Naguib, A., Abdelhafiz, M.M., et al., 2023. Probabilistic decline curve analysis: State-of-the-art review. *Energies* 16 (10), 4117. <https://doi.org/10.3390/en16104117>.
- You, J., Ampomah, W., Sun, Q., et al., 2020a. Machine learning based co-optimization of carbon dioxide sequestration and oil recovery in CO₂-EOR project. *J. Clean. Prod.* 260, 120866. <https://doi.org/10.1016/j.jclepro.2020.120866>.
- You, J., Ampomah, W., Sun, Q., 2020b. Development and application of a machine learning based multi-objective optimization workflow for CO₂-EOR projects. *Fuel* 264, 116758. <https://doi.org/10.1016/j.fuel.2019.116758>.
- Yuan, S., Ma, D., Li, J., et al., 2022. Progress and prospects of carbon dioxide capture, EOR-utilization and storage industrialization. *Petrol. Explor. Dev.* 49 (4), 955–962. [https://doi.org/10.1016/S1876-3804\(22\)60324-0](https://doi.org/10.1016/S1876-3804(22)60324-0).
- Zhang, C., Zhang, R., Zhu, Z., et al., 2023. Bottom hole pressure prediction based on hybrid neural networks and Bayesian optimization. *Pet. Sci.* 20 (6), 3712–3722. <https://doi.org/10.1016/j.petsci.2023.07.009>.
- Zhang, F., Nghiem, L., Chen, Z., 2023. Evaluating reservoir performance using a transformer based proxy model. *Geoenergy Sci. Eng.* 226, 211644. <https://doi.org/10.1016/j.geoen.2023.211644>.
- Zhang, N., Yin, M., Wei, M., et al., 2019. Identification of CO₂ sequestration opportunities: CO₂ miscible flooding guidelines. *Fuel* 241, 459–467. <https://doi.org/10.1016/j.fuel.2018.12.072>.
- Zhao, F., Wang, P., Huang, S., et al., 2020. Performance and applicable limits of multi-stage gas channeling control system for CO₂ flooding in ultra-low permeability reservoirs. *J. Petrol. Sci. Eng.* 192, 107336. <https://doi.org/10.1016/j.petrol.2020.107336>.
- Zhen, Y., Fang, J., Zhao, X., et al., 2022. Temporal convolution network based on attention mechanism for well production prediction. *J. Petrol. Sci. Eng.* 218, 111043. <https://doi.org/10.1016/j.petrol.2022.111043>.
- Zhou, X., Yuan, Q., Zhang, Y., et al., 2019. Performance evaluation of CO₂ flooding process in tight oil reservoir via experimental and numerical simulation studies. *Fuel* 236, 730–746. <https://doi.org/10.1016/j.fuel.2018.09.035>.
- Zhuang, X., Wang, W., Su, Y., et al., 2024a. Spatio-temporal sequence prediction of CO₂ flooding and sequestration potential under geological and engineering uncertainties. *Appl. Energy* 359, 122691. <https://doi.org/10.1016/j.apenergy.2024.122691>.
- Zhuang, X., Wang, W., Su, Y., et al., 2024b. Multi-objective optimization of reservoir development strategy with hybrid artificial intelligence method. *Expert Syst. Appl.* 241, 122707. <https://doi.org/10.1016/j.eswa.2023.122707>.
- Zhuang, X., Wang, W., Su, Y., et al., 2025. Life-cycle prediction and optimization of sequestration performance in CO₂ mixture huff-n-puff development for tight hydrocarbon reservoirs. *Appl. Energy* 388, 125618. <https://doi.org/10.1016/j.apenergy.2025.125618>.