



OPEN

# Simultaneous multi-well production forecasting and operational strategy awareness in heterogeneous reservoirs: a spatiotemporal attention-enhanced multi-graph convolutional network

Shuiqingshan Lu<sup>1,2</sup>, Chuanzhi Cui<sup>1,2</sup>✉, Zhongwei Wu<sup>3</sup>, Yin Qian<sup>1,2</sup>, Jian Wang<sup>4</sup> & Yingsong Huang<sup>4</sup>

Accurate production prediction in the ultra-high water cut stage is crucial for oilfield development. However, uncertainties from operational adjustments, reservoir heterogeneity, and inter-well interference pose significant challenges. Traditional reservoir-engineering methods rely on idealized assumptions and involve high computational costs in numerical simulations. Existing spatiotemporal graph models often separate spatial and temporal modeling, failing to represent reservoir dynamics. Meanwhile, multi-graph frameworks with fixed weights struggle to capture spatial changes during production adjustments. To address these issues, we propose a Spatiotemporal Attention-Enhanced Multi-Graph Convolutional Network (STA-MGCN) for simultaneous multi-well production forecasting. The method first constructs four graphs that encode Euclidean/non-Euclidean features of the well pattern and applies graph convolution to capture spatial interactions. To enrich the temporal context, a hybrid temporal sequence is constructed by integrating historical lag features with real-time controllable variables, and an attention-enhanced Long Short-Term Memory module is designed to dynamically weight each time step in this sequence. Finally, through layer-wise alternating updates of spatial and temporal features and multi-graph fusion, the model achieves true spatiotemporal coupling and adaptively weights the contributions of each graph. The model is evaluated on heterogeneous reservoir data from a composite five-spot pattern. Experiments show that STA-MGCN outperforms baselines, achieves high prediction accuracy and physical consistency in liquid lifting and well shut-in, and captures reservoir heterogeneity and inter-well interference effectively. Its inference speed is three orders of magnitude faster than numerical simulations, enabling efficient real-time decision-making and complex scenario forecasting.

**Keywords** Multi-well production forecasting, Multi-graph convolutional network, Spatiotemporal coupling, Inter-well interference, Reservoir heterogeneity, Liquid lifting, Well shut-in

## Abbreviations

STA-MGCN	Spatiotemporal attention-enhanced multi-graph convolutional network
CNN	Convolutional neural network
LSTM	Long short-term memory

<sup>1</sup>College of Petroleum Engineering, China University of Petroleum (East China), Qingdao 266580, China. <sup>2</sup>National Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao 266580, China. <sup>3</sup>College of Petroleum Engineering, Yangtze University, Wuhan 434023, China. <sup>4</sup>Exploration and Development Research Institute, Shengli Oilfield Company, SINOPEC, Dongying 257015, China. ✉email: ccz2008@126.com

D-LSTM	Deep LSTM
A-LSTM	Attention-based LSTM
DFW	Dynamic floating window
GA	Genetic algorithm
TCN	Temporal convolutional network
Seq2Seq	Sequence-to-sequence
GNNs	Graph neural networks
AST-GCN	Attribute-augmented spatiotemporal graph convolutional network
GRU	Gated Recurrent Unit
AG-GRU	Adaptive Graph-Gated Recurrent Unit
GCN-LSTM	Graph convolutional network-LSTM
MGCN-LSTM	Multi-graph convolutional network-LSTM
STGAN	Spatiotemporal graph attention network
GAT	Graph attention network
AST-GraphTrans	Spatiotemporal graph attention network-transformer
MOPSO	Multi-objective particle swarm optimization
EOR	Enhanced oil recovery
STA-LSTM	Spatiotemporal attention-enhanced LSTM
$G_{\text{topo}}$	Geometric topology graph
$G_{\text{bin}}$	Injection-production binary connectivity graph
$G_{\text{cond}}$	Fluid conductance graph
$G_{\text{dyn}}$	Dynamic similarity graph
BHP	Bottom-hole pressure
DTW	Dynamic time warping
RMSE	Root mean square error
MAE	Mean absolute error
$R^2$	Coefficient of determination
CCD	Central composite design
BBD	Box-Behnken design
ANOVA	Analysis of variance
RSM	Response surface methodology
CMG	Computer Modelling Group (numerical simulator)

Production forecasting is vital in petroleum engineering, acting as a foundation for optimizing production strategies and ensuring sustainable reservoir development. An oilfield's operating environment significantly increases in complexity when it moves into the ultra-high water cut stage. The heterogeneity of the reservoir increases. The uneven sweep of residual oil in low-permeability zones coexists with water channeling along high-permeability streaks. Schemes for injection-production are frequently adjusted. Ineffective cycling and inter-well interference further amplify production fluctuations. When combined, these factors impart stronger spatiotemporal coupling and nonlinearity to production data<sup>1,2</sup>. Consequently, traditional forecasting methods have significant accuracy constraints and struggle to capture the underlying dynamics. Reservoir-engineering approaches, such as Arps' Decline Curve Analysis, rely on simplified assumptions that inadequately capture complex seepage dynamics<sup>3</sup>. Full-physics numerical simulators can model this complexity but require costly, iterative computations that delay decision-making<sup>4</sup>. Classical statistical models also falter, as they cannot effectively handle the nonlinear data typical of modern digital oilfields<sup>5</sup>. Shallow machine learning methods handle nonlinear patterns but rely heavily on manual feature engineering, which limits robust generalization under dynamic conditions<sup>6,7</sup>.

By contrast, deep learning models overcome the limits of manual feature design. Convolutional Neural Networks (CNN) automatically extract multi-level representations. Long Short-Term Memory (LSTM) networks<sup>8</sup> learn long-range temporal dependencies. Both models have become common in oil production forecasting and markedly reduce error. Attention mechanisms and Transformer architectures<sup>9</sup> further improve long-term accuracy and model robustness<sup>10</sup>. Sagheer and Kotb<sup>11</sup> proposed a Deep LSTM (D-LSTM) model. They applied it to time-series data from the Huabei Oilfield and India's Cambay Basin. The model outperformed shallow machine learning methods. Kumar et al.<sup>12</sup> developed an Attention-based LSTM (A-LSTM) model. They optimized it with a Dynamic Floating Window (DFW) and Genetic Algorithm (GA). This improved long-term time-series prediction efficiency and accuracy. As model complexity increased, Pan et al.<sup>13</sup> combined Temporal Convolutional Networks (TCN) with Self-Attention Mechanisms for oil production prediction. The model enhanced injection-production data analysis and showed strong adaptability and stability. Lu et al.<sup>14</sup> designed a Transformer-based Sequence-to-Sequence (Seq2Seq) model for ultra-high water cut stage prediction. It integrated liquid lifting optimization to support oilfield management decisions.

The above literature focuses on modeling dynamic injection-production data and static geological parameters as time series curves, a method that has proven effective for forecasting oil production. However, oilfield development is inherently a coupled spatiotemporal process. Oil well production is affected not only by historical data but also by the spatial arrangement of the well pattern<sup>15</sup>. Time-series-based models alone are insufficient, as they cannot account for spatial factors such as reservoir heterogeneity, well interference, and the relationships between injection and production<sup>16,17</sup>. Therefore, developing deep learning models that capture both temporal dynamics and spatial relationships is essential for approximating reservoir dynamic behavior and enhancing prediction accuracy.

Against this backdrop, Graph Neural Networks (GNNs) offer a new pathway for oil production forecasting. The method treats each entity as a node and each relationship as an edge. It passes time-varying features along the edges, allowing the network to learn spatial structure and temporal change at once<sup>18</sup>. GNNs already support spatiotemporal tasks such as weather prediction, traffic flow monitoring<sup>19</sup>, and recommendation systems, and they are proving equally effective for oil-field production forecasts<sup>20</sup>. Within this line of work, we concentrate on integrated frameworks that combine GNNs with recurrent decoders. In these frameworks, spatial encoding and temporal decoding are modeled separately.

A representative case is the Attribute-Augmented Spatiotemporal Graph Convolutional Network (AST-GCN) introduced by Gao et al.<sup>21</sup>, which combines a Graph Convolutional Network (GCN) with a Gated Recurrent Unit (GRU) and delivers high-accuracy forecasts for a carbonate reservoir in the Middle East. The model builds its adjacency matrix from physical distance and static geological attributes and embeds dynamic injection data as weighted features of producer nodes. Because injectors appear only as feature modifiers rather than separate nodes, the model cannot explicitly learn injector-producer interactions.

Du et al.<sup>22</sup> later introduced the Adaptive Graph-Gated Recurrent Unit (AG-GRU) to capture changing spatial relations in reservoirs. The model constructs a static adjacency matrix based on geological factors and incorporates a self-learning edge correction matrix that updates edge weights in real time. Accurate production and water–oil ratio forecasts are produced by this adaptive design, which also efficiently tracks the interactions between water and oil wells. Although the model still adopts a separate modeling paradigm that combines GCN and GRU, the GRU lacks a dedicated memory unit. As a result, when processing long sequences, important information is easily diluted or overwritten by new inputs, which limits the model's ability to capture long-term dependencies.

Building upon this foundation, Du et al.<sup>23</sup> proposed a hybrid GCN-LSTM framework that handles reservoir spatial and temporal dynamics in parallel. The model constructs an adjacency matrix based on inter-well distances and control radii to capture spatial interference among production wells, while the LSTM component decodes temporal dependencies in production sequences using gated memory units. The introduction of LSTM partly compensates for the limitations of long-sequence modeling. However, the adjacency matrix, built solely on inter-well distance, fails to capture the influence of reservoir properties and inter-well connectivity on production performance.

Subsequently, Du et al.<sup>24</sup> introduced a Multi-Graph Convolutional Network-LSTM (MGCN-LSTM) framework address strong reservoir heterogeneity. The model builds four separate relation graphs to describe well pattern structures and injection-production links, giving a complete picture of inter-well interactions. Experiments show that this design captures more complex spatiotemporal features and outperforms earlier models in heterogeneous reservoirs. However, the study uses preset static weights to integrate multi-graph information. These weights remain fixed after training. As a result, the model cannot adjust the contribution of each graph when well relationships or operational conditions change.

Li et al.<sup>25</sup> further broadened graph modeling by introducing the Spatiotemporal Graph Attention Network (STGAN). The model combines Graph Attention Network (GAT) with self-attention mechanisms and introduces directed edges. Benchmarks show clear gains in capturing complex spatial layouts and fluid-migration patterns. Concurrently, Zhuang et al.<sup>26</sup> presented AST-GraphTrans, a model that combines a GAT with a Transformer for multi-objective tasks. It employs Multi-Objective Particle Swarm Optimization (MOPSO) to optimize well placement strategies, control parameters, CO<sub>2</sub>-enhanced oil recovery (EOR), and geological sequestration while effectively predicting cumulative oil production and CO<sub>2</sub> storage capacity. Both studies replace GCN with GAT for spatial encoding and use attention mechanisms or Transformers instead of LSTM for temporal decoding. This design allows the attention weights to learn adaptively, highlight the contribution of key wells to production, and better capture long-term dependencies. However, both models still follow a separate modeling paradigm that handles spatial and temporal features independently. As a result, they struggle to capture fine-grained spatiotemporal dependencies when handling dynamic changes in reservoir behavior.

In summary, production forecasting methods that combine GNNs with time-series models have continuously evolved. In spatial modeling, approaches have progressed from GCN to GAT and further to multi-graph structures. In temporal modeling, methods have shifted from GRU and LSTM toward attention mechanisms and Transformers. This evolution confirms that most current approaches still follow a separate modeling paradigm, focusing on module-level replacements but lacking fundamental breakthroughs in coupling mechanisms<sup>27</sup>. Current methods still face two major challenges<sup>28</sup>. The first is the separate modeling of spatial and temporal features. Most approaches still treat spatial encoding and temporal decoding as distinct steps. They simply concatenate the outputs, ignoring the coupling between spatial and temporal dependencies. The second challenge is that graph structures lack dynamic adaptability. Many studies employ a single, fixed weight adjacency matrix that cannot adapt to changes in inter-well relationships caused by operational adjustments.

We propose a STA-MGCN for simultaneous multi-well production forecasting. The method aims to address the limitations of existing studies, where spatial and temporal modeling are separated and graph structures rely on fixed weights. By alternating spatial and temporal updates within each layer and introducing an adaptive multi-graph fusion module, the model captures dynamic changes in inter-well relationships more effectively, thereby improving its prediction accuracy and generalization ability.

In the research design, the method first constructs four graphs that describe Euclidean and non-Euclidean relations in the well pattern. GCN modules are applied to these graphs to aggregate multi-order neighborhood information. For temporal modeling, we construct a hybrid temporal sequence by integrating historical lag features (e.g., oil production and bottom-hole pressure) with real-time controllable variables (e.g., injection rate and liquid lifting). A Spatiotemporal Attention-Enhanced LSTM (STA-LSTM) module is designed to dynamically assign weights to each time step, allowing the network to focus on critical stages of operational adjustments, such as liquid lifting or well shut-in. Subsequently, the model alternates spatial and temporal updates within each

layer, achieving true spatiotemporal coupling. Finally, a learnable attention mechanism adaptively weights the spatiotemporal features from each graph, and a fully connected layer maps the fused features to the predicted oil production of each well.

The proposed model is evaluated on data from a heterogeneous reservoir with a composite five-spot well pattern. Experimental results demonstrate that STA-MGCN outperforms benchmark models (LSTM, GCN-LSTM, MGNCN-LSTM) across standard metrics. It accurately captures inter-well interference during changes in production strategy, such as liquid lifting and well shut-in, and maintains physical consistency. It also achieves a speedup of more than three orders of magnitude compared to numerical simulations, highlighting its potential for real-time decision-making.

## Methodology

### Problem description

For a reservoir with  $N$  wells, spatial relationships can be represented by a weighted graph  $G = (V, E, W)$ , where each node  $v_{ij} \in V$  represents well  $i$ , edge  $e_{ij} \in E$  indicates a connection between wells  $i$  and  $j$ , and the corresponding weight  $w_{ij} \in W$  quantifies the strength of their interaction<sup>29</sup>.

In the spatial dimension, we define four graphs to encode different inter-well relationships. These include the Geometric Topology Graph ( $G_{\text{topo}}$ ), the Injection-Production Binary Connectivity Graph ( $G_{\text{bin}}$ ), the Fluid Conductance Graph ( $G_{\text{cond}}$ ), and the Dynamic Similarity Graph ( $G_{\text{dyn}}$ ).

In the temporal dimension, each well's time-series data is organized as a 3D tensor  $\mathbf{X} \in \mathbb{R}^{N \times M \times D}$ , where  $M$  is the number of time steps and  $D$  is the number of features. Given the tensor  $\mathbf{X}$  and the spatial graphs, the multi-well production forecasting task can be described as a spatiotemporal learning problem, where the goal is to learn the following mapping:

$$f: \mathbf{X} \in \mathbb{R}^{N \times M \times D} \times \{G_{\text{topo}}, G_{\text{cond}}, G_{\text{dyn}}, G_{\text{bin}}\} \rightarrow \widehat{\mathbf{Y}} \in \mathbb{R}^{N_P \times T} \quad (1)$$

where  $f$  is a learnable deep learning function,  $N_P \subseteq N$  denotes the set of production wells,  $T$  is the number of forecast steps, and  $\widehat{\mathbf{Y}}$  is the predicted oil production sequence.

### Model architecture overview

To address the problem defined above, we propose a STA-MGCN. Figure 1 displays the overall architecture. The model follows a modular structure with three main stages: input feature construction, spatiotemporal modeling, and multi-graph fusion.

In the first stage, we transform well dynamic data into hybrid temporal sequences and construct multiple weighted spatial graphs based on the actual well pattern (sections “[Hybrid temporal sequence construction](#)”, “[Spatial graph generator](#)”). In the second stage, the model builds four parallel spatiotemporal pipelines, each corresponding to a different graph. Within each pipeline, GCN and STA-LSTM modules are alternated to jointly encode spatial neighborhoods and temporal dependencies. Importantly, the parameters of these modules are not shared across graphs, allowing each graph to contribute unique spatiotemporal representations (sections “[Spatial dependency modeling with graph convolutional networks](#)”, “[Temporal dependency modeling with spatiotemporal attention-enhanced LSTM](#)”). Finally, a learnable attention mechanism integrates each graph's features, and a fully connected layer generates the oil production forecast (section “[Output prediction with alternating spatiotemporal modeling and multi-graph fusion](#)”).

Overall, STA-MGCN implements a “hybrid temporal sequence + multi-graph” dual-driven framework for coupled spatiotemporal learning and adaptive multi-graph feature fusion, significantly enhancing its ability to capture multi-scale spatial structures and dynamic evolution in complex reservoir systems. The following sections will detail the design and implementation of each module.

### Hybrid temporal sequence construction

As previously introduced, we organize the well's dynamic data into a 3D tensor  $\mathbf{X} \in \mathbb{R}^{N \times M \times D}$ . A sliding window<sup>30,31</sup> with length  $S$  and stride  $\Delta t$  is applied along the time axis  $\{1, 2, \dots, M\}$  to generate temporal input sequences. The total number of sequences is  $B$ , calculated as:

$$B = \left\lfloor \frac{M - S}{\Delta t} \right\rfloor + 1 \quad (2)$$

Each sequence centers on an observation point  $\tau \in \{1, 2, \dots, M\}$ . It covers the previous  $S_{\text{input}}$  time steps and the following  $S_{\text{output}}$  steps, so that  $S = S_{\text{input}} + S_{\text{output}}$ .

To capture reservoir behavior more accurately, The feature dimension  $D$  is further divided into two complementary subsets: lag features ( $L$ ) and current features ( $C$ ). Lag features emphasize historical dynamic effects, such as oil production and bottom-hole pressure (BHP), which reflect the reservoir's inertia and delay. In contrast, current features incorporate real-time control variables, such as injection rates and liquid lifting adjustments, enhancing the model perspective and better aligning with actual production management requirements<sup>32,33</sup>.

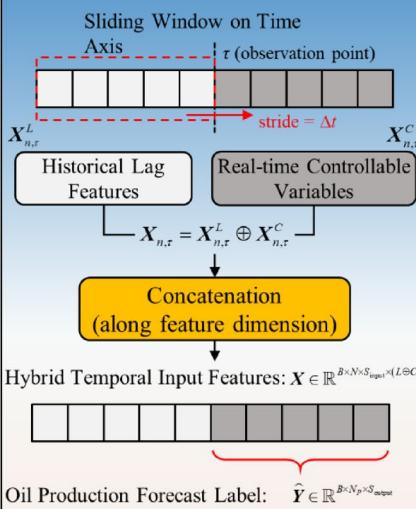
At prediction time  $\tau$ , the lag feature matrix of the  $n$ -th well is defined as:

$$\mathbf{X}_{n,\tau}^L = [x_{n,\tau-1}^L, x_{n,\tau-2}^L, \dots, x_{n,\tau-S_{\text{input}}}^L], \quad n \in \{1, 2, \dots, N\} \quad (3)$$

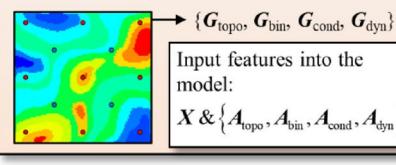
where  $x_{n,\tau-j}^L$  denotes the lag feature at the time  $\tau-j$ , and  $j \in \{1, 2, \dots, S_{\text{input}}\}$ .

## Input Feature Construction

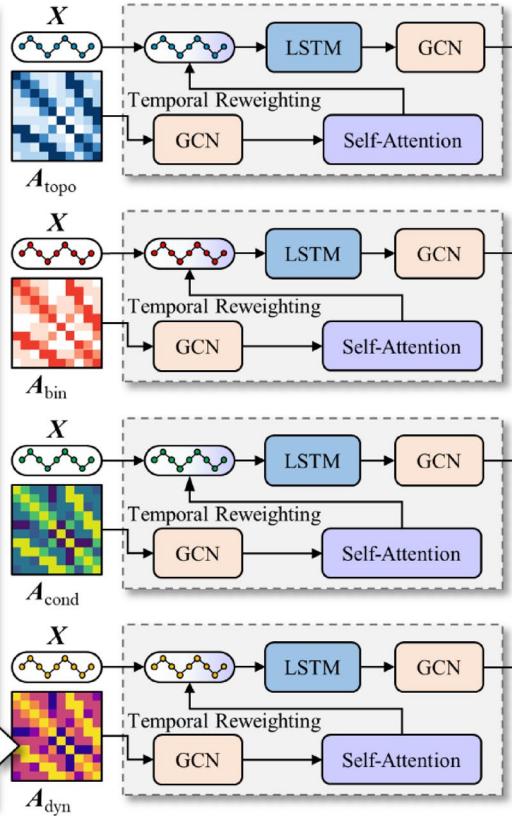
### • Hybrid Temporal Sequence Construction



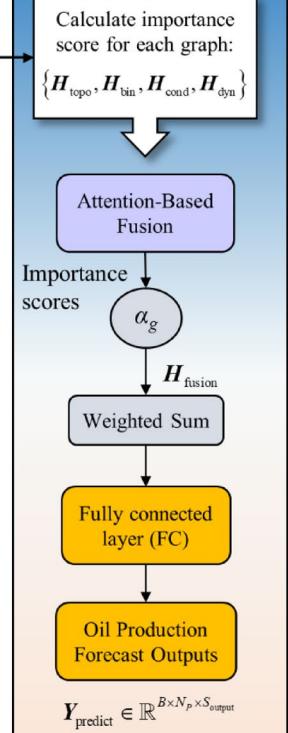
### • Weighted Spatial Graph Construction



## Spatiotemporal Modeling (STA-LSTM & GCN)



## Multi-Graph Fusion



**Fig. 1.** Overview of the STA-MGCN architecture. The model is composed of three main stages. The left panel shows the data input stage, where well dynamic data are transformed into hybrid temporal sequences and multiple weighted spatial graphs are constructed based on the actual well pattern. The middle panel represents the spatiotemporal modeling stage, where four parallel pipelines correspond to different graphs. Each pipeline contains its own GCN and STA-LSTM modules with independent parameters, ensuring that distinct graph-structured spatial dependencies are separately captured before being fused. The right panel illustrates the fusion and prediction stage, where the attention mechanism adaptively integrates multi-graph features and generates oil production forecasts. Black solid lines indicate the flow of data. Red dashed boxes highlight the construction of sliding windows and observation points.

When real-time control variables are observable, the current feature matrix of the  $n$ -th well is written as:

$$\mathbf{X}_{n,\tau}^C = [x_{n,\tau}^C, x_{n,\tau+1}^C, \dots, x_{n,\tau+S_{\text{output}}}^C] \quad (4)$$

where  $x_{n,\tau+k}^C$  is the current feature at time  $\tau+k$ , and  $k \in \{1, 2, \dots, S_{\text{output}}\}$ .

When the lengths of the input and output windows are equal (i.e.,  $S_{\text{input}} = S_{\text{output}}$ ), the lag and current features are concatenated along the feature dimension to form the hybrid temporal input:

$$\mathbf{X}_{n,\tau} = \mathbf{X}_{n,\tau}^L \oplus \mathbf{X}_{n,\tau}^C \quad (5)$$

where  $\oplus$  denotes feature-wise concatenation.

Within the same time slice, the prediction target for the  $p$ -th production well is defined as:

$$\hat{\mathbf{Y}}_{p,\tau} = [\hat{y}_{p,\tau}, \hat{y}_{p,\tau+1}, \dots, \hat{y}_{p,\tau+S_{\text{output}}}], p \in \{1, 2, \dots, N_p\} \quad (6)$$

where  $\hat{y}_{p,\tau+k}$  is the predicted oil production at time  $\tau+k$ , and  $k \in \{1, 2, \dots, S_{\text{output}}\}$ .

In summary, the overall tensor transformation is:

$$\text{Hybrid Temporal Input Features: } \mathbf{X} \in \mathbb{R}^{B \times N \times S_{\text{input}} \times (L \oplus C)} \rightarrow \text{Prediction Targets: } \hat{\mathbf{Y}} \in \mathbb{R}^{B \times N_p \times S_{\text{output}}} \quad (7)$$

### Spatial graph generator

Oil–water migration in porous media is influenced by both Euclidean factors, such as spatial distance, and non-Euclidean factors, including reservoir properties and development operations<sup>34,35</sup>. To capture these effects, we construct two types of graphs based on the actual well pattern. The  $G_{\text{topo}}$  and  $G_{\text{bin}}$  encode the Euclidean information of the well pattern, while the  $G_{\text{cond}}$  and  $G_{\text{dyn}}$  capture its non-Euclidean information.

#### Geometric topology graph

Based on the principles of graph theory<sup>36,37</sup>, the  $G_{\text{topo}}$  is constructed to represent spatial proximity between wells. The graph is defined as  $G_{\text{topo}} = (V, E, W_{\text{topo}})$ , where the weight  $w_{\text{topo}}(i, j)$  between well  $i$  and well  $j$  is calculated as the reciprocal of the shortest path length  $l_{ij}$  between them. The shortest path is determined by the minimum number of edges connecting the two wells. Therefore, wells that are geometrically close have higher connection weights, and distant wells have smaller weights.

These connection weights are assembled into the adjacency matrix  $A_{\text{topo}} = [w_{\text{topo}}(i, j)]_{N \times N}$  which captures the geometric relationship between all well pairs. Figure 2 presents a heatmap of  $A_{\text{topo}}$ : darker cells indicate stronger connections (i.e., shorter paths) between spatially adjacent wells, reflecting the actual well pattern, while lighter cells indicate weaker connections among distant wells. This distribution reveals a spatial structure that is locally dense but globally sparse. To follow standard graph-theory conventions and avoid self-loops, all diagonal elements of  $A_{\text{topo}}$  are set to zero.

#### Injection-production binary connectivity graph

To model operational relationships between injection and production wells, we construct a binary connectivity graph<sup>38</sup>. The graph is defined as  $G_{\text{bin}} = (V, E, W_{\text{bin}})$ , where the weight  $w_{\text{bin}}(i, j) = 1$  indicates a direct connection between wells  $i$  and  $j$ , and zero otherwise. These binary weights are assembled into the adjacency matrix  $A_{\text{bin}} = [w_{\text{bin}}(i, j)]_{N \times N}$ . As shown in Fig. 3, this adjacency matrix clearly highlights which injector-producer pairs are linked.

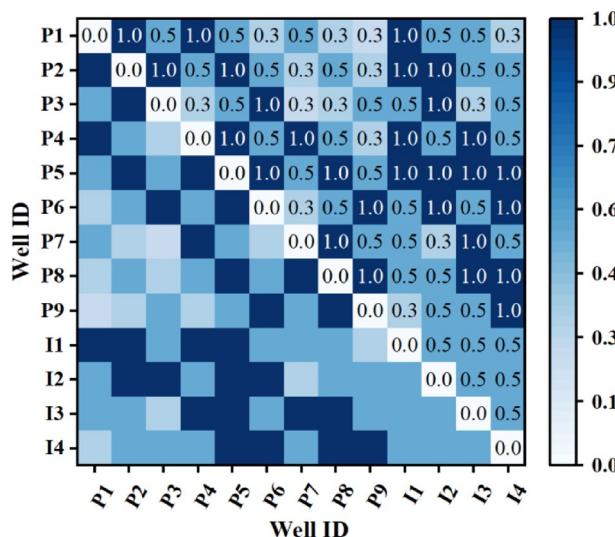
#### Fluid conductance graph

Drawing inspiration from electrical models in highly conductive grids, we adopt the concept of conductance to define “fluid conductance” for characterizing fluid flow capacity between wells<sup>39,40</sup>. We represent it as a graph  $G_{\text{cond}} = (V, E, W_{\text{cond}})$  and set the edge weight  $w_{\text{cond}}(i, j) = k_{i,j} / d_{i,j}^{\eta}$ . Here,  $k_{i,j}$  denotes the equivalent permeability between wells  $i$  and  $j$ ,  $d_{i,j}$  is the inter-well distance, and  $\eta$  is the distance decay exponent. This formulation captures both geological connectivity and geometric attenuation, and it ensures that the weight decreases as the spacing between wells increases.

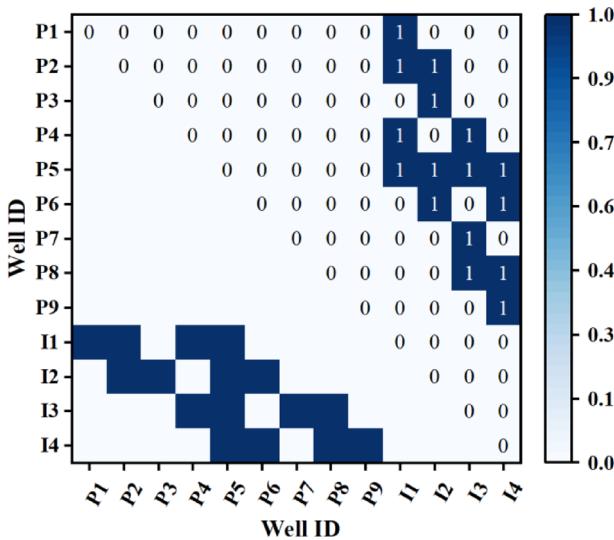
For the equivalent permeability, we idealize the dominant inter-well flow path as a one-dimensional, steady-state, single-phase channel composed of two segments in series. Based on this setup, the path can be divided into two parts. One part represents the near-well region of well  $i$ , with length  $l_i$  and permeability  $k_p$ , and the other corresponds to the near-well region of well  $j$ , with length  $l_j$  and permeability  $k_p$ . According to Darcy’s law, the total pressure drop is the sum of the pressure drops across the two segments.

$$\Delta p = \Delta p_i + \Delta p_j = \frac{\mu Q}{A} \left( \frac{l_i}{k_i} + \frac{l_j}{k_j} \right) \quad (8)$$

In this expression,  $\Delta p$  represents the total pressure drop along the flow direction, defined as positive.  $\Delta p_i$  and  $\Delta p_j$  are the pressure drops across the two segments.  $\mu$  is the fluid viscosity,  $Q$  is the volumetric flow rate, and  $A$



**Fig. 2.** Geometric topology graph.



**Fig. 3.** Injection-production binary connectivity graph.

denotes the effective cross-sectional area. Under the assumptions of steady-state, single-phase flow, both  $Q$  and  $A$  are the same for each segment. The entire path can then be treated as a homogeneous segment with total length  $L = l_i + l_j$  and an equivalent permeability  $k_{i,j}$ . From this approximation, we can derive the following:

$$\Delta p = \frac{\mu Q}{A} \frac{L}{k_{i,j}} \Rightarrow \frac{1}{k_{i,j}} = \frac{l_i}{L} \cdot \frac{1}{k_i} + \frac{l_j}{L} \cdot \frac{1}{k_j} \quad (9)$$

When the thickness or length of the near-well regions on both sides are nearly equal, the formula reduces to:

$$k_{i,j} = \frac{2}{\frac{1}{k_i} + \frac{1}{k_j}} = \frac{2k_i \cdot k_j}{k_i + k_j} \quad (10)$$

This shows that using the harmonic mean emphasizes that the inter-well flow path is controlled by the lower-permeability segment. It makes the equivalent permeability more sensitive to low values. This behavior avoids the systematic overestimation that arises when using the arithmetic mean in highly heterogeneous reservoirs.

These weights form the adjacency matrix  $A_{\text{cond}} = [w_{\text{cond}}(i, j)]_{N \times N}$ . Figure 4a illustrates a reciprocal-distance graph based on  $(1/d_{ij})$ . Figure 4b presents the proposed fluid conduction graph. A comparison between the two shows that  $G_{\text{cond}}$  better captures strong connectivity in high-permeability zones and flow resistance in low-permeability areas. This observation aligns with the actual physical characteristics of the reservoir and highlights how heterogeneity affects inter-well relationships. Self-loops are retained in the diagonal elements of  $A_{\text{cond}}$  to preserve each well's self-information.

#### Dynamic similarity graph

To quantify dynamic similarity between production and injection wells, we constructed a graph  $G_{\text{dyn}} = (V, E, W_{\text{dyn}})$ . The weight  $w_{\text{dyn}}(i, j)$  represents the similarity between the rate of change curves for wells  $i$  and  $j$ . For production wells, the time series is the rate of liquid production change. For injection wells, it is the rate of water injection change. We first compute the Dynamic Time Warping (DTW) distance<sup>41</sup> between each pair of time series, and then convert it to a similarity via a Gaussian kernel:

$$w_{\text{dyn}}(i, j) = \exp \left( -\frac{(\text{DTW}(x'_i, x'_j))^2}{2\sigma^2} \right) \quad (11)$$

where  $\sigma$  denotes the Gaussian kernel's standard deviation, controlling how fast similarity decays. Figure 5 displays the distribution of these rate of change curves. The differences between wells are not significant because the data are selected from the high water cut stage of reservoir development. During this period, liquid production rates exhibit a smooth declining trend with limited variation in slope. This setup provides the data-driven model with a consistent and stable decline pattern. It facilitates the model's ability to learn typical production dynamics and accurately identify abrupt changes caused by liquid lifting operations or well shut-in.

These weights populate an adjacency matrix  $A_{\text{dyn}} = [w_{\text{dyn}}(i, j)]_{N \times N}$ . As shown in Fig. 6, wells of the same type tend to have higher similarity. Some producer-injector pairs also show strong similarity, such as P3 and I1. Although the overall inter-well similarity is at a high level, the similarity between P5 and other wells becomes relatively low after normalization. This suggests that P5, located at the center, has more pronounced fluctuations

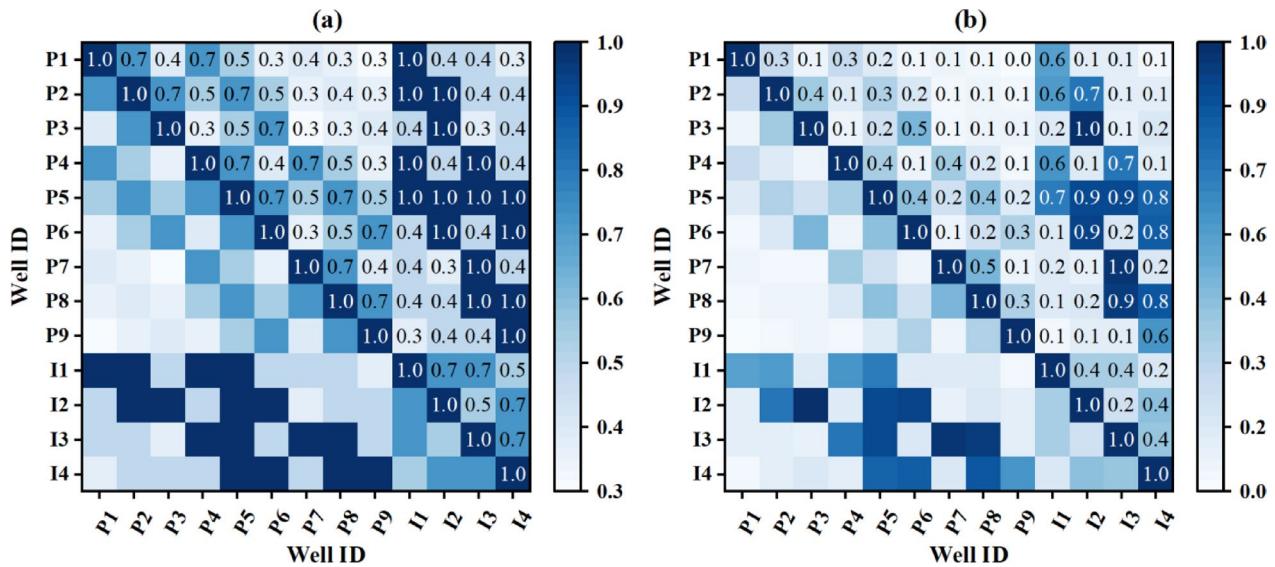


Fig. 4. Inter-well distance reciprocal Graph (a) vs. fluid conductance graph (b).

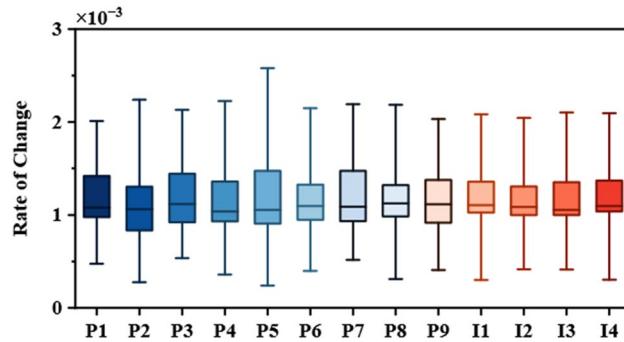


Fig. 5. Distribution of time series rate of change values across wells.

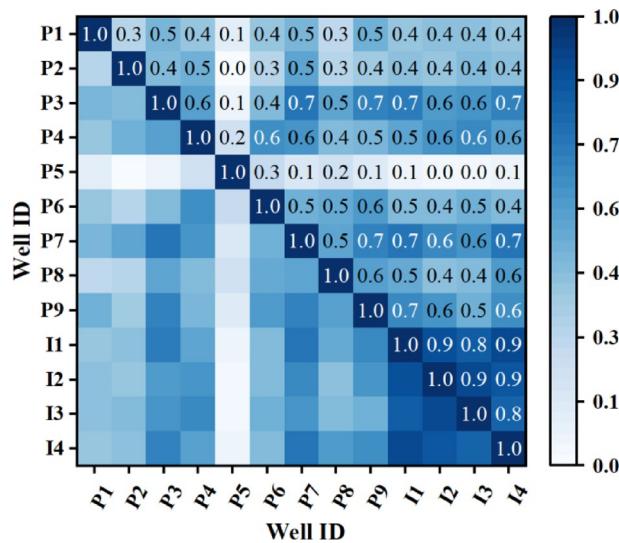


Fig. 6. Dynamic similarity graph.

in production and injection. Self-loops in  $A_{\text{dyn}}$  preserve nodal autocorrelation features and help prevent information loss during graph propagation.

### Spatial dependency modeling with graph convolutional networks

To model spatial dependencies between wells, we apply GCN<sup>42,43</sup>. The input includes multiple adjacency matrices  $\mathbf{A} \in \{\mathbf{A}_{\text{topo}}, \mathbf{A}_{\text{bin}}, \mathbf{A}_{\text{cond}}, \mathbf{A}_{\text{dyn}}\}$ , each describing a type of inter-well relationship.

At each GCN layer  $l$ , hidden features are updated by aggregating information from neighboring nodes. The operation is defined as:

$$\mathbf{H}_{\text{GCN}}^{(l)} = \text{ReLU} \left( \mathbf{W}_{\text{GCN}} f \left( \mathbf{A}; \theta^{(R)} \right) \mathbf{H}_{\text{GCN}}^{(l-1)} \right) \quad (12)$$

where  $\mathbf{H}_{\text{GCN}}^{(l-1)} \in \mathbb{R}^{N \times D_{l-1}}$  and  $\mathbf{H}_{\text{GCN}}^{(l)} \in \mathbb{R}^{N \times D_l}$  are the input and output node features at  $l-1$  and  $l$ , respectively.  $\mathbf{W}_{\text{GCN}} \in \mathbb{R}^{D_l \times D_{l-1}}$  is the learnable weight matrix for feature transformation. The function  $f(\mathbf{A}; \theta^{(R)}) \in \mathbb{R}^{N \times N}$  performs graph-based feature aggregation based on adjacency matrix  $\mathbf{A}$  with trainable parameters  $\theta^{(R)}$ .

We employ a Chebyshev polynomial approximation<sup>44</sup> to define  $f(\mathbf{A}; \theta^{(R)})$ :

$$f \left( \mathbf{A}; \theta^{(R)} \right) = \sum_{r=0}^R \theta_r^{(R)} T_r \left( \tilde{\mathbf{L}} \right) \quad (13)$$

where  $\tilde{\mathbf{L}} = 2\mathbf{L}/\lambda_{\max} - \mathbf{I} \in \mathbb{R}^{N \times N}$  is the rescaled Laplacian matrix,  $\lambda_{\max}$  is the largest eigenvalue of  $\mathbf{L}$ , and  $T_r(\cdot)$  denotes the Chebyshev polynomial. The polynomials are defined recursively:

$$T_0 \left( \tilde{\mathbf{L}} \right) = \mathbf{I}, T_1 \left( \tilde{\mathbf{L}} \right) = \tilde{\mathbf{L}}, T_{r+1} \left( \tilde{\mathbf{L}} \right) = 2\tilde{\mathbf{L}}T_r \left( \tilde{\mathbf{L}} \right) - T_{r-1} \left( \tilde{\mathbf{L}} \right) \quad (14)$$

The polynomial order  $R$  controls the receptive field size. When  $R=1$ , the model only aggregates information from immediate neighbors. Increasing  $R$  enables the model to capture long-range dependencies in well pattern. allows the model to incorporate longer-range dependencies in the well pattern. All adjacency matrices used in this study are processed through the same GCN module. Notably, this spatial modeling approach is also adaptable to other types of spatial relationships and spatiotemporal forecasting tasks in oilfield applications<sup>45</sup>.

### Temporal dependency modeling with spatiotemporal attention-enhanced LSTM

Oil production is affected by various factors, including historical production rates, operational strategies, and inter-well interactions<sup>46,47</sup>. To achieve more fine-grained temporal modeling, we design a STA-LSTM to capture both spatial context and temporal dynamics in multi-well data (Fig. 7).

At each time step  $t \in \{1, 2, \dots, S_{\text{input}}\}$ , the original hybrid temporal sequences  $\mathbf{X}_{n,t} \in \mathbb{R}^{N \times (L \oplus C)}$  are first enriched with spatial information via GCN:

$$\hat{\mathbf{X}}_{n,t} = [\mathbf{X}_{n,t} \oplus \text{GCN}(\mathbf{A}, \mathbf{X}_{n,t})] \quad (15)$$

where  $\oplus$  denotes feature concatenation and  $\mathbf{A}$  represents the set of adjacency matrices.

The sequences  $\hat{\mathbf{X}}_{n,t}$ , enriched with spatial context, are subsequently passed through a self-attention mechanism<sup>48</sup> to adjust the weight of each time step. First, the global feature signatures  $\mathbf{z}_{n,t}$  are obtained through global average pooling, which summarizes the fused features across all nodes at each time step:

$$\mathbf{z}_{n,t} = \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{X}}_{n,t} \quad (16)$$

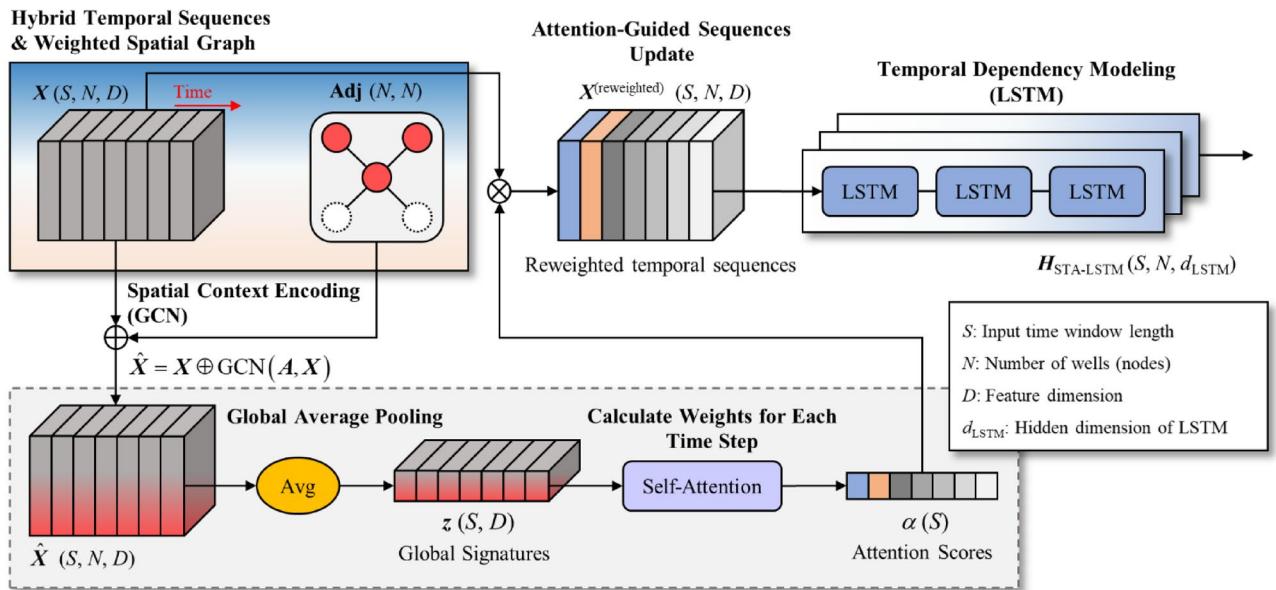
The attention matrix is computed based on the dot-product similarity of global feature signatures across time steps. Softmax normalization is subsequently applied, followed by temporal averaging over all  $S_{\text{input}}$  steps. This process yields attention weights that highlight the most informative time steps. The original input sequences are then reweighted as follows:

$$\mathbf{X}_{n,t}^{(\text{reweighted})} = \mathbf{X}_{n,t} \odot \frac{1}{S_{\text{input}}} \sum_{t=1}^{S_{\text{input}}} \left( \text{softmax} \left( \frac{\mathbf{z}_{n,t} \cdot \mathbf{z}_{n,t}^T}{\sqrt{d_k}} \right) \right) \quad (17)$$

where  $d_k$  is a scaling factor used to stabilize gradient computation and normalize attention scores across time.

Finally, the attention-reweighted sequences  $\mathbf{X}_{n,t}^{(\text{reweighted})}$  are then input to an LSTM layer, which models both short-term variations and long-term trends. Formally, at layer  $l$ , the LSTM updates its hidden state as:

$$\mathbf{H}_{\text{STA-LSTM}}^{(l)} = \text{LSTM} \left( \mathbf{X}_{n,t}^{(\text{reweighted})}, \mathbf{H}_{\text{LSTM}}^{(l-1)}, \mathbf{W}_{\text{LSTM}} \right) \quad (18)$$



**Fig. 7.** Workflow of the STA-LSTM module. Hybrid temporal sequences are first enriched with spatial context via a GCN. Next, global average pooling produces a summary vector at each time step. A self-attention layer then computes attention scores for every time point. These scores are applied to reweight the original features. Finally, an LSTM processes the attention-weighted sequence to model both short-term fluctuations and long-term trends.

where  $H_{\text{LSTM}}^{(l-1)}$  denotes the previous hidden state and  $W_{\text{LSTM}}$  represents the trainable parameters. The output  $H_{\text{STA - LSTM}}^{(l)}$  integrates spatial context, attention-guided weighting, and temporal dependency into a unified representation.

#### Output prediction with alternating spatiotemporal modeling and multi-graph fusion

To generate accurate oil production forecasts, we design an output prediction module that alternates between spatial and temporal modeling and integrates features from multiple graphs using an adaptive fusion mechanism.

##### Alternating spatiotemporal modeling

Most existing spatiotemporal models adopt a serial structure, such as applying GCN before RNN, or the reverse. These structures treat spatial and temporal features independently and combine them only at the output stage. This separation weakens the interaction between spatial and temporal information, making it difficult for the model to capture the coupling between production dynamics and inter-well interference<sup>49</sup>. To achieve true spatiotemporal feature coupling, we alternate the STA-LSTM and GCN modules within each layer. First, the STA-LSTM module receives the spatial features  $H_{\text{GCN}}^{(l-1)}$  from the previous GCN module (Eqs. 12–14), which already incorporate graph-structured spatial context based on the adjacency matrix  $A$  (e.g.,  $A_{\text{topo}}$ ), and then learns temporal dependencies with attention weights (Eqs. 15–18). It applies temporal attention and recurrent updates to these GCN-enhanced features together with the temporal states  $H_{\text{LSTM}}^{(l-1)}$  to update the current temporal representation:

$$H_{\text{STA - LSTM}}^{(l)} = \text{STA - LSTM} \left( A, H_{\text{GCN}}^{(l-1)}, H_{\text{LSTM}}^{(l-1)} \right) \quad (19)$$

Then, the output of STA-LSTM  $H_{\text{STA - LSTM}}^{(l)}$  is once again fed into a GCN module. This step re-applies the adjacency matrix  $A$  to aggregate neighborhood information and extract updated spatial features:

$$H_{\text{GCN}}^{(l)} = \text{GCN} \left( A, H_{\text{STA - LSTM}}^{(l)} \right) \quad (20)$$

where the output  $H_{\text{GCN}}^{(l)}$  represents the updated hidden features at layer  $l$ . By stacking multiple layers, the model gradually captures long-term temporal dependencies and multi-order spatial correlations in the well pattern.

##### Adaptive multi-graph fusion

In most multi-graph convolution models, the outputs from different graphs are often combined using simple summation or averaging. This approach ignores the heterogeneity between graphs and treats each graph as equally important.

After several spatiotemporal modeling layers, we obtain spatial features  $\mathbf{H}_{\text{GCN},g}^{(L)} \in \left\{ \mathbf{H}_{\text{GCN},\text{topo}}^{(L)}, \mathbf{H}_{\text{GCN},\text{bin}}^{(L)}, \mathbf{H}_{\text{GCN},\text{cond}}^{(L)}, \mathbf{H}_{\text{GCN},\text{dyn}}^{(L)} \right\}$  from four graphs. To adaptively integrate these features, we propose an attention-based<sup>50</sup> fusion strategy. This strategy uses a learnable weight vector  $v$  to calculate the importance score for each graph. During training, the vector is updated through backpropagation. It interacts with the hidden feature representations  $\mathbf{H}_{\text{GCN},g}^{(L)}$  and uses a softmax function to adaptively assign importance scores  $\alpha_g$ :

$$\alpha_g = \text{softmax} \left( v^T \tanh \left( \mathbf{H}_{\text{GCN},g}^{(L)} \right) \right) \quad (21)$$

The final spatiotemporal representation is computed as a weighted sum of the outputs from each graph. In this way, the model automatically emphasizes the graphs that contribute more to prediction accuracy rather than relying on manually predefined weights.

$$\mathbf{H}_{\text{fusion}} = \sum_g \alpha_g \left\{ \mathbf{H}_{\text{GCN},\text{topo}}^{(L)}, \mathbf{H}_{\text{GCN},\text{bin}}^{(L)}, \mathbf{H}_{\text{GCN},\text{cond}}^{(L)}, \mathbf{H}_{\text{GCN},\text{dyn}}^{(L)} \right\} \quad (22)$$

This fused representation  $\mathbf{H}_{\text{fusion}}$  is then passed into a fully connected layer to generate the final production prediction:

$$\mathbf{Y}_{\text{predict}} = FC(\mathbf{H}_{\text{fusion}}) \quad (23)$$

## Experiments and results

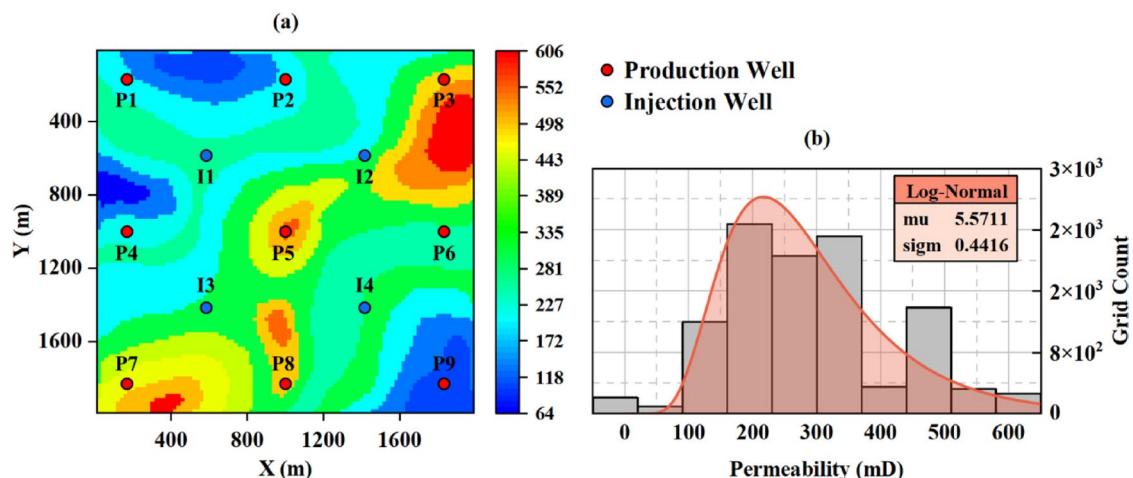
### Oilfield spatiotemporal dataset construction

#### Data overview

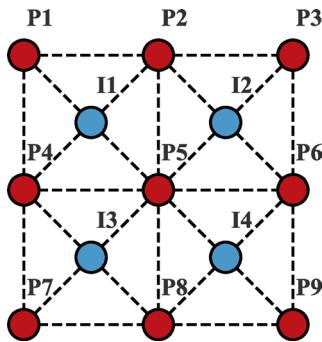
The dataset used in this study was obtained from a numerical simulation of a highly heterogeneous reservoir. As shown in Fig. 8, permeability followed a left-skewed log-normal distribution, with a peak range between 180 and 260 mD and a long right tail extending to 605 mD. High-permeability streaks (greater than 350 mD) appeared along the diagonal of the reservoir model. These streaks created strong spatial contrasts with surrounding low-permeability zones (less than 200 mD).

The reservoir adopted a composite five-spot well pattern composed of four adjacent injection-production units (Fig. 9). The average spacing between wells was 840 m, and the distance between injectors and producers was 594 m. Injection wells were set to maintain an initial pressure of 28.5 MPa, forming a constant pressure gradient of 3 MPa. Table 1 provides the permeability values and initial pressures for both injectors and producers.

Dynamic data were generated using CMG simulation over a 20-year production period. Liquid lifting operations were set when the water cut reached 87% to simulate artificial interventions. Based on the average grid permeability within each well's control radius, Table 1 reports the well-control permeability, ranging from 186 to 379 mD for producers and from 221 to 340 mD for injectors. As shown in Table 2, daily oil production varied significantly across wells, while bottomhole pressure remained highly stable.



**Fig. 8.** Spatial heterogeneity and permeability distribution in the reservoir; (a) the spatial distribution of grid permeability; (b) the histogram of permeability values with a fitted log-normal distribution. Red circles indicate production wells P1-P9, and blue circles indicate injection wells I1-I4.

**Fig. 9.** Schematic diagram of the composite five-spot well-pattern.

Well index	Type	Permeability (mD)	Initial pressure (MPa)
P1-P9	Production Well	186.61, 211.03, 379.50, 216.12, 306.97, 325.96, 365.82, 332.79, 187.34	25.5
I1-I4	Injection Well	221.37, 340.20, 335.89, 275.30	28.5

**Table 1.** Well-control permeability and initial pressure for injection and production wells.

Well ID	Mean $\pm$ SD of Oil/water Rate (m <sup>3</sup> /d)	Mean $\pm$ SD of BHP (MPa)	Mean $\pm$ SD of Liquid production (m <sup>3</sup> /d)
P1	12.107 $\pm$ 5.992	26.605 $\pm$ 0.375	90.045 $\pm$ 29.180
P2	14.551 $\pm$ 7.213	26.698 $\pm$ 0.358	102.197 $\pm$ 34.039
P3	15.133 $\pm$ 8.210	26.533 $\pm$ 0.380	148.741 $\pm$ 47.318
P4	13.537 $\pm$ 6.733	26.618 $\pm$ 0.372	96.039 $\pm$ 31.837
P5	25.059 $\pm$ 12.996	26.705 $\pm$ 0.354	211.818 $\pm$ 72.109
P6	14.950 $\pm$ 7.583	26.645 $\pm$ 0.363	115.326 $\pm$ 38.187
P7	14.434 $\pm$ 8.111	26.490 $\pm$ 0.386	152.141 $\pm$ 48.545
P8	19.352 $\pm$ 10.563	26.635 $\pm$ 0.365	183.466 $\pm$ 61.449
P9	9.542 $\pm$ 4.654	26.644 $\pm$ 0.370	64.604 $\pm$ 21.071
I1	263.280 $\pm$ 88.240	28.163 $\pm$ 0.095	–
I2	295.630 $\pm$ 95.170	28.154 $\pm$ 0.095	–
I3	320.360 $\pm$ 101.890	28.127 $\pm$ 0.102	–
I4	288.910 $\pm$ 96.500	28.163 $\pm$ 0.096	–

**Table 2.** Dynamic data statistics for each well.*Time-series data preprocessing*

We preprocessed the well production time-series data to eliminate dimensional discrepancies and reduce noise interference. First, the raw data were standardized using Z-score normalization, computed as  $x_{\text{norm}} = (x - \mu) / \sigma_x$ , where  $\mu$  and  $\sigma_x$  denote the mean and standard deviation of raw data  $x$ , respectively.

Next, an adaptive bilateral Gaussian filter was used for data smoothing and noise reduction<sup>51</sup>, with core equations:

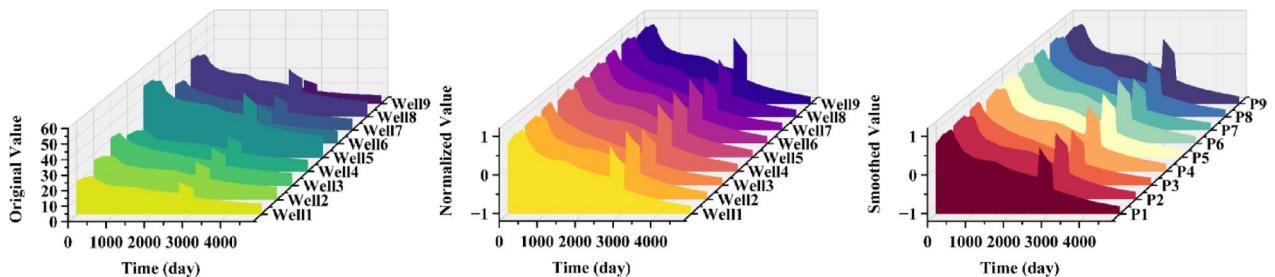
$$x_{\text{smoothed}}(\beta) = \frac{\sum_{\alpha \in \Omega(\beta)} \omega_d(\alpha, \beta) \cdot \omega_r(\alpha, \beta) \cdot x_{\text{norm}}(\alpha)}{\sum_{\alpha \in \Omega(\beta)} \omega_d(\alpha, \beta) \cdot \omega_r(\alpha, \beta)} \quad (24)$$

Spatial and range weights are defined as:

$$\omega_d(\alpha, \beta) = \exp \left( -\frac{(\alpha - \beta)^2}{2\sigma_d^2(\beta)} \right) \quad (25)$$

$$\omega_r(\alpha, \beta) = \exp \left( -\frac{(x(\alpha) - \mu_\beta)^2}{2\sigma_r^2(\beta)} \right) \quad (26)$$

where  $x_{\text{smoothed}}(\beta)$  represents smoothed data at timestep  $\beta$ . The local window around timestep  $\beta$  is denoted as  $\Omega(\beta)$ . The spatial weight  $\omega_d(\alpha, \beta)$  uses a dynamic variance  $\sigma_d(\beta)$ . A larger variance expands the smoothing range



**Fig. 10.** Comparison of original (left), normalized (middle), and smoothed (right) well production data. Only production wells are shown here as an example. Data from other features were processed similarly but are omitted for clarity.

Well type	Feature type	Feature description	Tensor dimension
Production Wells	Lag features	Historical oil production and BHP (past 30 days)	$\mathbb{R}^{N_P \times 30 \times 2}$
	Current features	Planned liquid production (current 30 days)	$\mathbb{R}^{N_P \times 30 \times 1}$
	Concatenated features	Lag and current features concatenated along feature dimension	$\mathbb{R}^{N_P \times 30 \times 3}$
Injection Wells	Lag features	Historical BHP (past 30 days)	$\mathbb{R}^{N_I \times 30 \times 1}$
	Current features	Water injection and cumulative injection (current 30 days)	$\mathbb{R}^{N_I \times 30 \times 2}$
	Concatenated features	Lag and current features concatenated along feature dimension	$\mathbb{R}^{N_I \times 30 \times 3}$

**Table 3.** Feature composition and tensor dimensions.

to reduce noise, while a smaller variance preserves important details. The range weight  $\omega_r(\alpha, \beta)$  compares local mean values  $\mu_\beta$  at timesteps  $\alpha$  and  $\beta$ . Its variance  $\sigma_r(\beta)$  decreases adaptively near sudden changes (e.g., liquid lifting operations), preventing excessive smoothing at boundaries.

As shown in Fig. 10, the normalized data (middle) enhanced the visibility of production trends and provided a consistent data scale compared to the raw data (left). The smoothed data (right) effectively suppressed short-term fluctuations while preserving key step responses caused by production operations, such as liquid lifting measures.

#### Spatiotemporal dataset construction

Following the hybrid temporal sequence construction described in section “[Hybrid temporal sequence construction](#)” and the spatial graph generation introduced in section “[Spatial graph generator](#)”, we constructed a multi-Source spatiotemporal dataset. Specifically, we used a sliding window approach with a historical input window ( $S_{\text{input}}$ ) of 30 days, a prediction output window ( $S_{\text{output}}$ ) of 30 days, with a default stride  $\Delta t$  of 1 day. Historical lag features and real-time controllable variables were separately built for production wells and injection wells (see Table 3).

The global input tensor  $X \in \mathbb{R}^{13 \times 30 \times 3}$  was constructed by concatenating the above features from 9 production wells and 4 injection wells along the well dimension, forming a hybrid temporal representation. The prediction target was the oil production rates of the 9 production wells over the prediction output window, resulting in an output tensor  $Y \in \mathbb{R}^{9 \times 30 \times 1}$ . Spatial relationships were embedded via four adjacency matrices  $\{A_{\text{topo}}, A_{\text{bin}}, A_{\text{cond}}, A_{\text{dyn}}\}$ , which represented prior knowledge about inter-well connectivity (detailed in section “[Spatial graph generator](#)”).

Finally, we split the prepared input and output feature tensors in chronological order. The core dataset (2009–2024) contained 5,844 daily timesteps, producing 5,785 sequences (Eq. 2). Among these sequences, 70% (4,049 sequences) were used for training. Validation and testing sets each included 15% (868 sequences per set). Additionally, an independent inference dataset (2025–2029) contained 1,826 daily timesteps, forming 1,767 sequences. This dataset was reserved for evaluating model robustness under new injection-production strategies.

## Model performance evaluation

#### Evaluation metrics

To evaluate the predictive performance of the STA-MGCN model, we use three commonly adopted metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination ( $R^2$ ).

The RMSE measures how much the predicted values deviate from the actual values. It is sensitive to large errors, such as production fluctuations caused by liquid lifting operations<sup>52</sup>. The formula is:

$$\text{RMSE} = \sqrt{\frac{1}{B \cdot N_P \cdot S_{\text{output}}} \sum_{b=1}^B \sum_{p=1}^{N_P} \sum_{k=1}^{S_{\text{output}}} (y_{b,p,k} - \hat{y}_{b,p,k})^2} \quad (27)$$

where  $y_{b,p,k}$  and  $\hat{y}_{b,p,k}$  are the actual and predicted oil production values on the  $k$ -th day in the prediction window for the  $p$ -th production well in the  $b$ -th sample.

The MAE calculates the average of absolute differences between predictions and actual values. Unlike RMSE, it is more robust to outliers. The formula is:

$$\text{MAE} = \frac{1}{B \cdot N_P \cdot S_{\text{output}}} \sum_{b=1}^B \sum_{p=1}^{N_P} \sum_{k=1}^{S_{\text{output}}} |y_{b,p,k} - \hat{y}_{b,p,k}| \quad (28)$$

The  $R^2$  quantifies how well the predictions explain the variance of the observed data, thereby reflecting the model's ability to capture production trends and characterize reservoir behavior. The formula is:

$$R^2 = 1 - \frac{\sum_{b=1}^B \sum_{p=1}^{N_P} \sum_{k=1}^{S_{\text{output}}} (y_{b,p,k} - \hat{y}_{b,p,k})^2}{\sum_{b=1}^B \sum_{p=1}^{N_P} \sum_{k=1}^{S_{\text{output}}} (y_{b,p,k} - \bar{y})^2} \quad (29)$$

where  $\bar{y}$  is the global average of all actual values. An  $R^2$  close to 1 indicates a strong ability to follow trends. An  $R^2$  close to 0 means the model cannot explain the observed variations.

#### *Hyperparameter optimization*

Response Surface Methodology (RSM)<sup>53</sup> was applied to systematically optimize three key hyperparameters of the STA-MGCN model: the forecasting window length ( $S_{\text{output}}$ ), the sliding window stride ( $\Delta t$ ), and the batch size. The forecasting window determines the length of the future period the model needs to predict. The sliding window stride controls how much historical data is included in the input. The batch size sets the amount of data used for each gradient update. These parameters work together to shape the temporal dimension of the input sequences, directly affecting the complexity of model training.

First, a hybrid experimental design method based on Central Composite Design (CCD) and Box-Behnken Design (BBD) was employed. Each hyperparameter was set at five discrete levels (Table 4), generating 17 experimental combinations that covered the key regions of the parameter space<sup>54</sup>. Each combination was repeated five times, and the average RMSE and MAE were calculated to ensure the stability of the results (see Table 5). To ensure comparability of the results, all parameters, except for the forecasting window length ( $S_{\text{output}}$ ), the sliding window stride ( $\Delta t$ ), and the batch size, were kept the same across all experimental groups. The model uses four GCN layers for spatial modeling. Each layer has a hidden dimension of 128 and a second-order Chebyshev polynomial kernel. For temporal modeling, the model uses four STA-LSTM modules, each with an LSTM hidden layer of size 64. The Adam optimizer is used for all experiments, with mean squared error as the loss function. The number of training epochs is fixed at 60.

Next, second-order polynomial regression equations were constructed for each dataset to describe the relationship between hyperparameters and error metrics, as shown below:

$$E = \gamma_0 + \sum_{i=1}^3 \gamma_i V_i + \sum_{i=1}^3 \gamma_{ii} V_i^2 + \sum_{i < j} \gamma_{ij} V_i V_j + \varepsilon \quad (30)$$

where  $E$  is the response variable (RMSE or MAE),  $V$  represents hyperparameters,  $\gamma$  denotes regression coefficients, and  $\varepsilon$  is the error term.

Then, an ANOVA analysis was conducted to examine the significance of each term in the regression model (see Table 6). The results indicated that the linear term of the sliding window stride had a significant positive effect on RMSE and MAE across all datasets, while its quadratic term showed a negative effect. In addition, the quadratic term of the forecasting window length was significant only in the training dataset. However, the batch size and all interaction terms did not show any statistically significant effects.

After removing insignificant items, batch size was fixed at its mean value. A two-dimensional response surface was then plotted using the sliding window stride and forecasting window length, as shown in Figs. 11 and 12. The optimal configuration was identified by finding the lowest points on the response surface, where the forecasting window length was 30, the sliding window stride was 1, and the batch size was 32.

#### *Prediction performance analysis*

Through quantitative metrics and visual analysis, a comprehensive comparison was conducted between the proposed STA-MGCN model and benchmark models (LSTM, GCN-LSTM, and MGNCN-LSTM) to evaluate its predictive capability.

Variable	Level - 2	Level - 1	Centre (0)	Level + 1	Level + 2
Forecasting Window (days)	7	15	30	45	60
Sliding Stride (days)	1	7	15	30	45
Batch Size	8	16	32	64	128

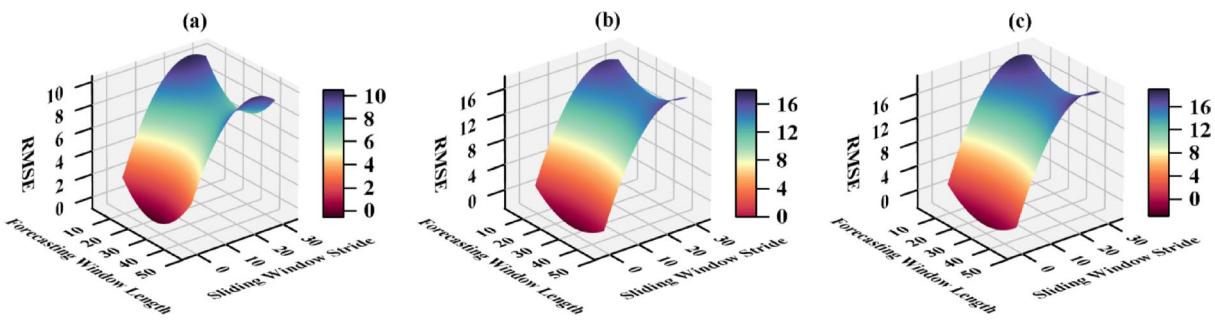
**Table 4.** Design levels for hyperparameter optimisation (CCD + BBD).

Run ID & Configurations	Train-RMSE	Val-RMSE	Test-RMSE	Train-MAE	Val-MAE	Test-MAE
Run 1 ( $S_{\text{output}} = 7, \Delta t = 1, \text{batch size} = 8$ )	1.558	0.564	0.648	0.933	0.520	0.580
Run 2 ( $S_{\text{output}} = 7, \Delta t = 1, \text{batch size} = 128$ )	0.781	0.409	0.557	0.625	0.365	0.496
Run 3 ( $S_{\text{output}} = 7, \Delta t = 45, \text{batch size} = 8$ )	1.038	1.255	1.551	0.741	1.120	1.425
Run 4 ( $S_{\text{output}} = 7, \Delta t = 45, \text{batch size} = 128$ )	5.520	10.163	11.342	4.326	9.710	10.857
Run 5 ( $S_{\text{output}} = 60, \Delta t = 1, \text{batch size} = 8$ )	10.683	0.979	2.190	8.419	0.866	2.040
Run 6 ( $S_{\text{output}} = 60, \Delta t = 1, \text{batch size} = 128$ )	6.534	7.856	8.950	5.137	7.187	8.311
Run 7 ( $S_{\text{output}} = 60, \Delta t = 45, \text{batch size} = 8$ )	7.208	11.546	12.901	5.853	11.079	12.385
Run 8 ( $S_{\text{output}} = 60, \Delta t = 45, \text{batch size} = 128$ )	6.870	12.592	13.949	5.428	12.071	13.378
Run 9 ( $S_{\text{output}} = 15, \Delta t = 15, \text{batch size} = 32$ )	6.668	9.989	11.194	5.472	9.576	10.746
Run 10 ( $S_{\text{output}} = 45, \Delta t = 15, \text{batch size} = 32$ )	6.859	12.664	13.966	5.427	12.139	13.510
Run 11 ( $S_{\text{output}} = 30, \Delta t = 7, \text{batch size} = 32$ )	2.162	2.445	3.154	1.502	1.950	2.940
Run 12 ( $S_{\text{output}} = 30, \Delta t = 30, \text{batch size} = 32$ )	7.324	11.674	13.031	5.926	11.233	12.510
Run 13 ( $S_{\text{output}} = 30, \Delta t = 15, \text{batch size} = 16$ )	5.923	8.153	9.224	4.875	7.716	8.755
Run 14 ( $S_{\text{output}} = 30, \Delta t = 15, \text{batch size} = 64$ )	7.336	11.836	13.147	5.965	11.341	12.606
Run 15 ( $S_{\text{output}} = 30, \Delta t = 15, \text{batch size} = 32$ )	7.334	11.746	13.060	5.969	11.264	12.533
Run 16 ( $S_{\text{output}} = 30, \Delta t = 15, \text{batch size} = 32$ )	7.305	11.480	12.786	5.966	11.013	12.274
Run 17 ( $S_{\text{output}} = 30, \Delta t = 15, \text{batch size} = 32$ )	7.306	11.514	12.820	5.964	11.046	12.307

**Table 5.** Mean RMSE and MAE for Each Experimental Run. Run 16 and Run 17 are repeated experiments. Their purpose is to reduce random errors and noise in the experiments, ensuring the stability and reliability of the results.

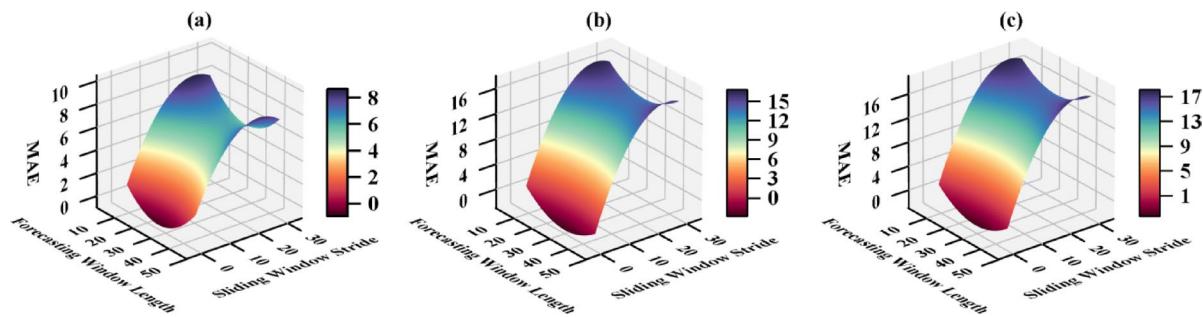
Dataset	Significant predictors	Key effects	Model fit
Train-RMSE	Sliding Stride (+, $p = 0.016$ )	Sliding Stride <sup>2</sup> (-, $p = 0.017$ ) Forecasting Window <sup>2</sup> (+, $p = 0.045$ )	$R^2 = 0.896$
Val-RMSE	Sliding Stride (+, $p = 0.013$ )	Sliding Stride <sup>2</sup> (-, $p = 0.017$ )	$R^2 = 0.899$
Test-RMSE	Sliding Stride (+, $p = 0.011$ )	Sliding Stride <sup>2</sup> (-, $p = 0.015$ )	$R^2 = 0.904$
Train-MAE	Sliding Stride (+, $p = 0.014$ )	Sliding Stride <sup>2</sup> (-, $p = 0.015$ ) Forecasting Window <sup>2</sup> (+, $p = 0.049$ )	$R^2 = 0.889$
Val-MAE	Sliding Stride (+, $p = 0.011$ )	Sliding Stride <sup>2</sup> (-, $p = 0.015$ )	$R^2 = 0.900$
Test-MAE	Sliding Stride (+, $p = 0.011$ )	Sliding Stride <sup>2</sup> (-, $p = 0.015$ )	$R^2 = 0.906$

**Table 6.** Significant predictors and model-fit statistics for the polynomial regression. Note: “+” indicates a positive linear; “-” indicates a negative relationship. Superscript “<sup>2</sup>” denotes a second-order term. p represents the significance level of the predictor in the polynomial regression model. All predictors shown are statistically significant at  $p < 0.05$ .

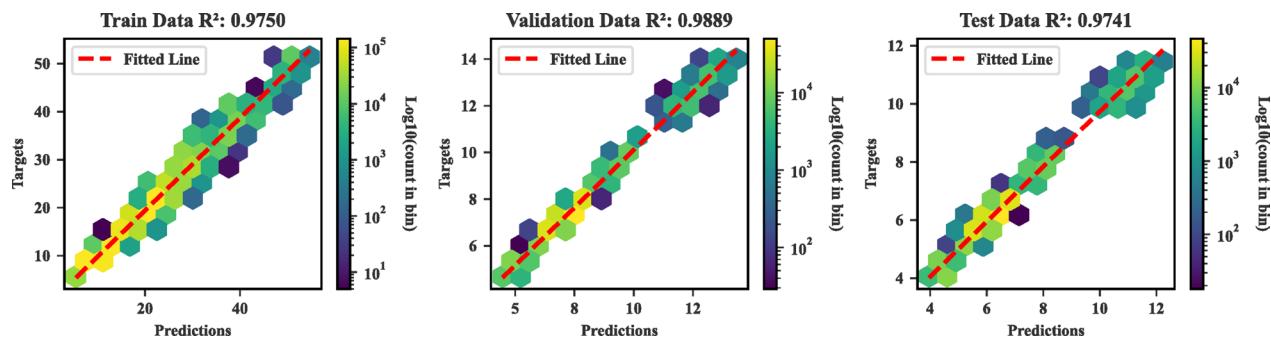


**Fig. 11.** RMSE response surface of sliding window stride and forecasting window length. Subfigures (a), (b), and (c) respectively correspond to the training set, validation set, and test set.

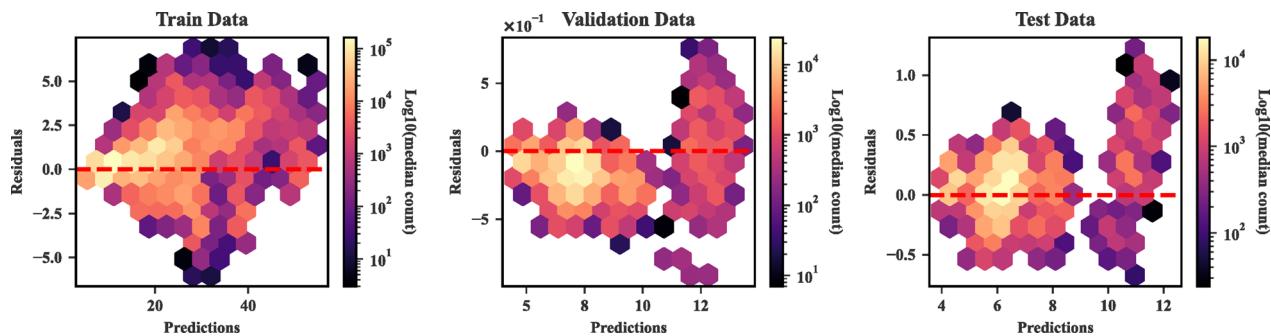
First, in terms of prediction accuracy, Figs. 13 and 14 presented scatter plots of predicted versus actual values, along with residual distributions for each dataset, color-coded by Log10(count). The STA-MGCN achieved high  $R^2$  values of 0.975, 0.988, and 0.974 for the training, validation, and test sets, respectively, demonstrating excellent predictive performance. In dense, bright-colored areas representing conventional production ranges (0–25 m<sup>3</sup>/d for training, 0–10 m<sup>3</sup>/d for validation, and 0–9 m<sup>3</sup>/d for testing), the points closely aligned along the



**Fig. 12.** MAE response surface of sliding window stride and forecasting window length. Subfigures (a), (b), and (c) respectively correspond to the training set, validation set, and test set.



**Fig. 13.** Scatter plots of predicted vs. actual values across training, validation, and test sets.



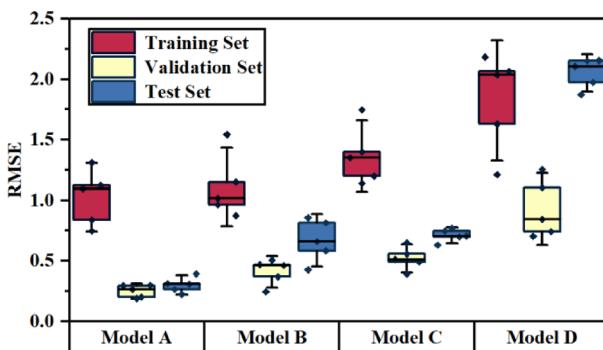
**Fig. 14.** Residual distributions across training, validation, and test sets.

Model	RMSE	MAE	R <sup>2</sup>
LSTM	2.052 ± 0.124	1.793 ± 0.087	0.897
GCN-LSTM	0.711 ± 0.053	0.586 ± 0.049	0.937
MGCN-LSTM	0.676 ± 0.175	0.552 ± 0.071	0.952
STA-MGCN	0.237 ± 0.063	0.215 ± 0.026	0.974

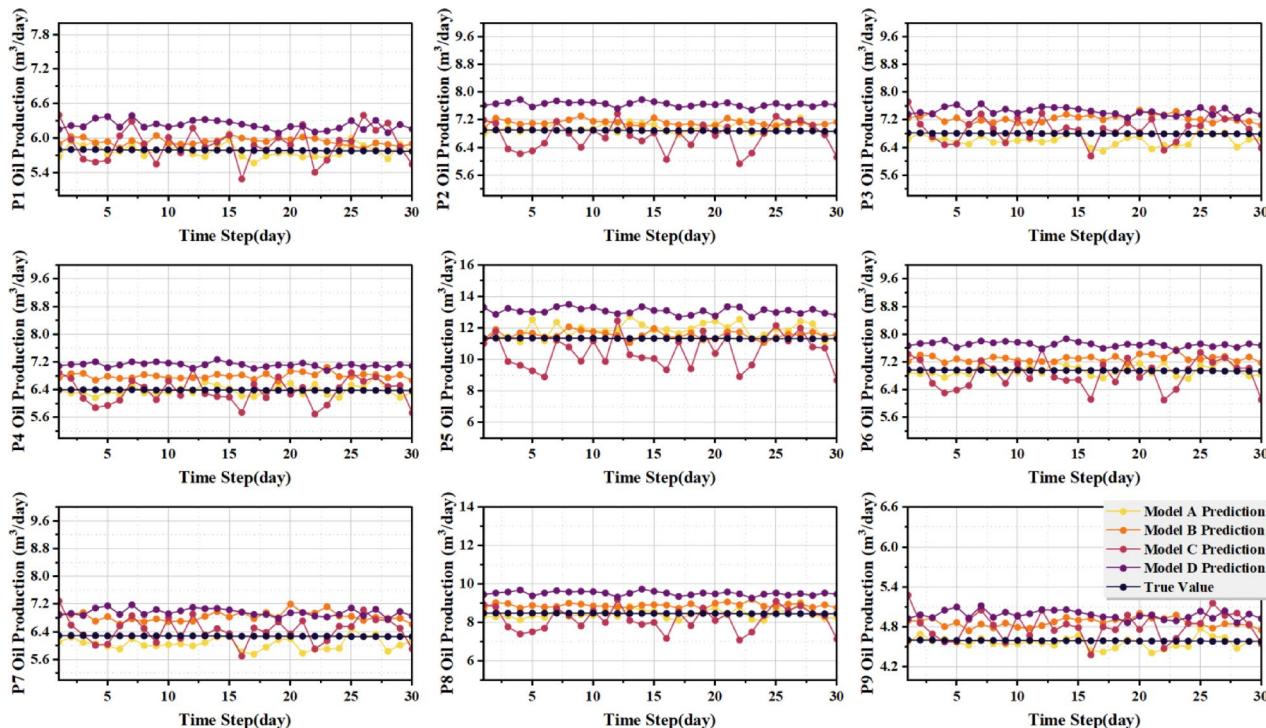
**Table 7.** Test set performance of STA-MGCN and baseline models.

diagonal, and residuals remained small. In contrast, sparse, dark-colored areas representing higher production ranges ( $> 25 \text{ m}^3/\text{d}$  for training,  $> 10 \text{ m}^3/\text{d}$  for validation, and  $> 9 \text{ m}^3/\text{d}$  for testing) exhibited greater scatter and larger residuals due to fewer available samples. Overall, the STA-MGCN showed strong prediction accuracy under normal conditions but slight deviations under high-production conditions.

In terms of prediction stability and generalization, Table 7 presented the performance comparison of STA-MGCN and benchmark models on the test set, and Fig. 15 showed the RMSE boxplots from five repeated runs.



**Fig. 15.** RMSE boxplots of STA-MGCN and baseline models on three datasets. Model A is the proposed STA-MGCN, while Models B, C, and D correspond to the benchmark methods MGCN-LSTM, GCN-LSTM, and LSTM, respectively.

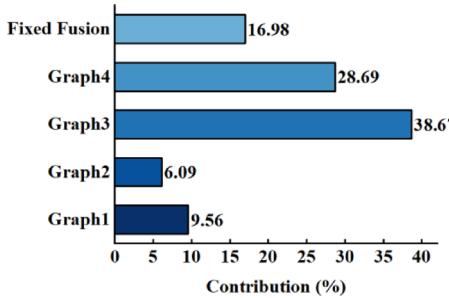


**Fig. 16.** Oil production forecasts on the test set for wells P1-P9, comparing the proposed STA-MGCN (Model A, yellow) with the three benchmark models MGCN-LSTM (Model B, orange), GCN-LSTM (Model C, red), and LSTM (Model D, purple). The true values are shown as solid black lines.

The results indicated that STA-MGCN achieved the lowest RMSE ( $0.237 \pm 0.063$ ) and MAE ( $0.215 \pm 0.026$ ), along with the highest  $R^2$  (0.974), demonstrating both accuracy and consistency. Compared to the LSTM model, GCN-LSTM and MGCN-LSTM reduced RMSE by 65.4% and 67.0%, respectively, yet their RMSE values remained 66.7% and 64.9% higher than STA-MGCN. The LSTM model recorded the highest RMSE and the widest interquartile range, indicating that the single-temporal model exhibited the greatest prediction variability.

In terms of temporal-trend prediction, Fig. 16 compares the true production series (black line) with the forecasts of STA-MGCN and the benchmark models on the test set. The results indicated that the STA-MGCN predictions closely matched the actual production curves, accurately tracking fluctuations within the 30-day forecasting window. Among the nine wells (P1-P9), STA-MGCN maintained mean absolute errors consistently below  $0.3 \text{ m}^3/\text{d}$ . In contrast, the errors of GCN-LSTM and MGCN-LSTM exceeded  $1 \text{ m}^3/\text{d}$  for high-production wells P5 and P7, while the LSTM model showed the largest fluctuations and errors. These findings confirmed that STA-MGCN outperformed other models in capturing both overall trends and local variations.

Model variant	RMSE	MAE	R <sup>2</sup>	Configuration description
Full Model (All 4 Graphs)	0.237 ± 0.063	0.215 ± 0.026	0.974	All four graphs with adaptive fusion
w/o $G_{\text{topo}}$ (Graph 1)	0.375 ± 0.172	0.308 ± 0.130	0.964	$G_{\text{bin}}, G_{\text{cond}}, G_{\text{dyn}}$ adaptively fused
w/o $G_{\text{bin}}$ (Graph 2)	0.325 ± 0.080	0.290 ± 0.032	0.968	$G_{\text{topo}}, G_{\text{cond}}, G_{\text{dyn}}$ adaptively fused
w/o $G_{\text{cond}}$ (Graph 3)	0.795 ± 0.075	0.661 ± 0.058	0.921	$G_{\text{topo}}, G_{\text{bin}}, G_{\text{dyn}}$ adaptively fused
w/o $G_{\text{dyn}}$ (Graph 4)	0.651 ± 0.085	0.610 ± 0.035	0.930	$G_{\text{topo}}, G_{\text{bin}}, G_{\text{cond}}$ adaptively fused
Fixed Fusion (All 4 Graphs)	0.482 ± 0.068	0.457 ± 0.030	0.952	All four graphs with fixed-weight fusion

**Table 8.** Test set performance of ablated models and fusion strategies.**Fig. 17.** Contribution of ablated models to prediction accuracy. The bar chart depicts the normalized prediction accuracy contributions from ablation experiments, measured via relative RMSE degradation.  $G_{\text{cond}}$  (Graph 3) and  $G_{\text{dyn}}$  (Graph 4) account for the largest improvements, at 38.67% and 28.69%, respectively, while  $G_{\text{topo}}$  (Graph 1) and  $G_{\text{bin}}$  (Graph 2) contribute 9.56% and 6.09%. The fixed fusion approach explains 16.98% of the overall accuracy gain.

#### Ablation study

To quantify the contribution of each graph ( $G_{\text{topo}}, G_{\text{bin}}, G_{\text{cond}}, G_{\text{dyn}}$ ), we conducted two ablation experiments: (1) single-graph ablation, where one graph type was removed in turn while the remaining graphs retained the original Adaptive Multi-Graph Fusion, and (2) fusion-strategy comparison, where all four graphs were retained but combined using fixed weights. Table 8 and Fig. 17 presented the test set results. The full model achieved an RMSE of 0.237 m<sup>3</sup>/d. Removing  $G_{\text{cond}}$  increased the RMSE to 0.795 m<sup>3</sup>/d, a 235% rise; removing  $G_{\text{dyn}}$  raised it to 0.651 m<sup>3</sup>/d, or 175%. Dropping  $G_{\text{topo}}$  and  $G_{\text{bin}}$  had smaller effects, raising the RMSE to 0.375 m<sup>3</sup>/d (+58%) and 0.325 m<sup>3</sup>/d (+37%), respectively. Using fixed-weight fusion for all graphs resulted in an RMSE of 0.482 m<sup>3</sup>/d, 103% higher than that of the adaptive approach. These findings showed that  $G_{\text{cond}}$  and  $G_{\text{dyn}}$  contributed the most to prediction accuracy and that attention-based adaptive fusion clearly outperformed fixed weighting.

#### Cross-scenario inference of production operations

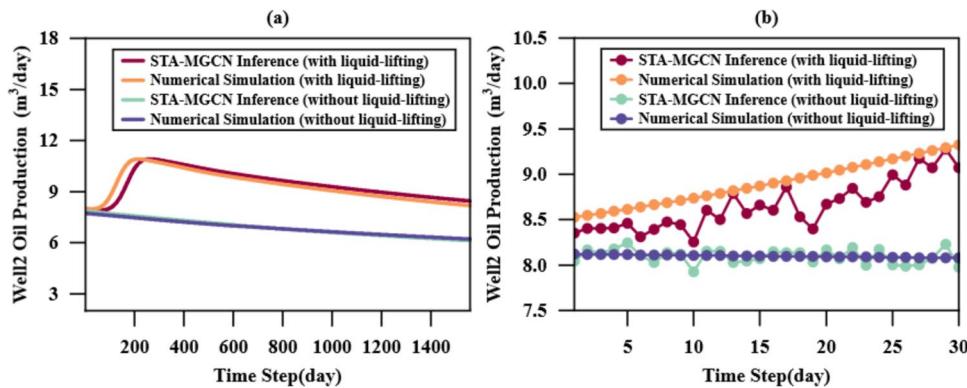
##### Inference of production enhancement with liquid lifting measures

To evaluate the inference ability of the STA-MGCN model under operational adjustment scenarios, we used a reserved inference dataset covering the period from 2025 to 2029 (1,826 days, 1,767 sequences). At this stage of development, the reservoir had already entered the ultra-high water cut period. This scenario simulates a liquid lifting operation on well P2. The liquid production of well P2 was increased from 117.32 to 167.32 m<sup>3</sup>/d and kept constant until the end of 2029. We compared the model's predictions with numerical simulation results to assess its accuracy and robustness. We retained the previously optimized sliding window setting, which uses the past 30 days of the hybrid temporal sequence as input to predict oil production for the next 30 days. This ensures consistency across scenarios and reproducibility of the results.

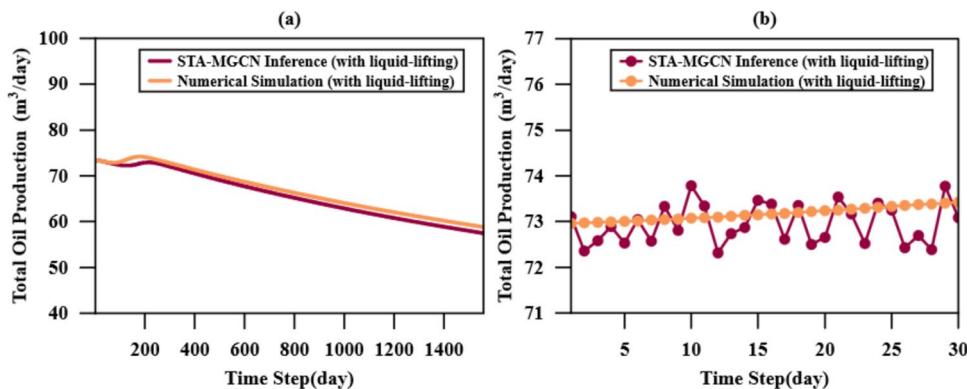
Figures 18 and 19 present the full time series comparison and the short-term response details. Panel (a) shows the complete inference cycle from January 1, 2025, to April 7, 2029 (day 1 to day 1,559), with the horizontal axis indexed by days. To highlight the dynamic changes before and after the liquid lifting intervention, the stable section in the later part of the inference cycle, where production remained unchanged, was omitted. Panel (b) presents a randomly selected 30-day forecast window within one year after the liquid lifting. This was used to test the consistency of the model's response across different intervention stages and to avoid conclusions depending on a fixed time point.

In Fig. 18, the STA-MGCN forecasts for well P2 almost overlapped with the numerical simulation results. This indicated that the model followed long-term trends and quantified the incremental production achieved through liquid lifting. The short-term forecast error was within 0.5 m<sup>3</sup>/d. Although minor fluctuations emerged, the model still reproduced the short-term lift in production.

Regarding total production for the well pattern, Fig. 19 presented the same analysis when only well P2 received liquid lifting treatment. Over the entire series and within a 30-day forecast window, the predicted curve nearly coincided with the simulation, with a maximum error below 2 m<sup>3</sup>/d. The close match confirmed that the model maintained accuracy even after the operational change in a single well.



**Fig. 18.** Comparison between STA-MGCN forecasts and numerical simulation results under the liquid lifting scenario at well P2. (a) Complete time series with and without liquid lifting intervention. (b) 30-day forecast window comparing the short-term production response with and without liquid-lifting.



**Fig. 19.** Comparison between STA-MGCN forecasts and numerical simulation results for total production of the well pattern under the liquid lifting scenario at well P2. (a) Complete time series with and without liquid lifting intervention. (b) 30-day forecast window comparing the short-term production response with and without liquid lifting.

#### Inference of shut-in impact in high-permeability zone

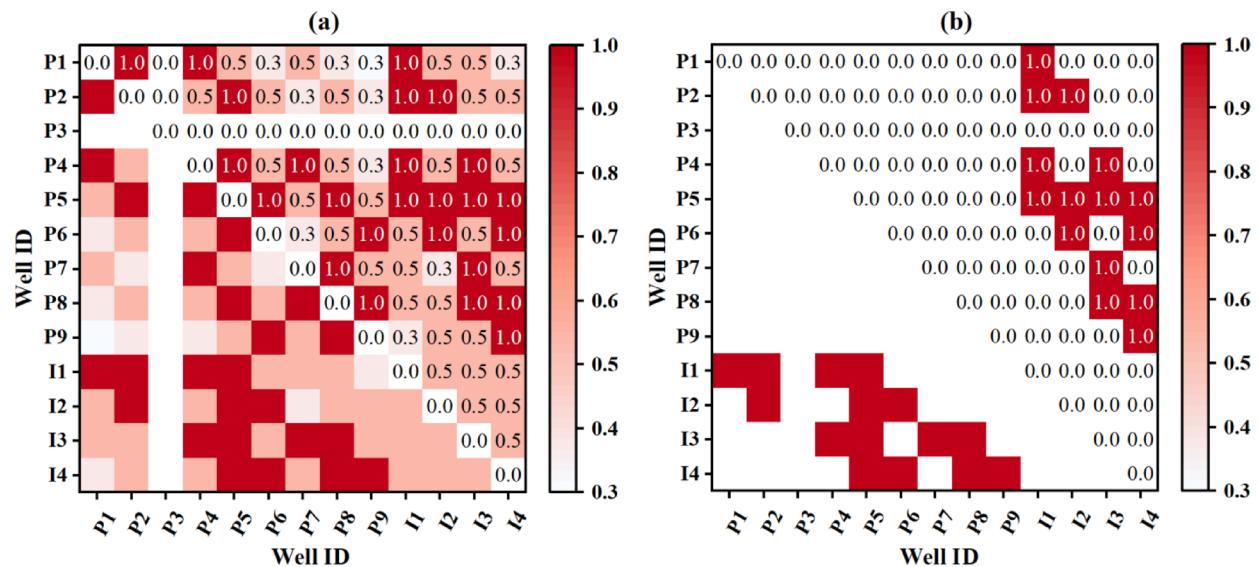
To evaluate the cross-scenario generalization ability of STA-MGCN under an unknown shut-in operation, we used the same inference dataset to simulate the shut-in of well P3, which is located in a high-permeability zone. At the same time, liquid lifting at well P2 was maintained. The prediction results were then compared with numerical simulation outcomes. This experiment also applied the previously optimized sliding window setting, consistent with the configuration described in section “[Inference of production enhancement with liquid lifting measures](#)”.

The shut-in operation required targeted adjustments to the multi-graph input framework, as shown in Fig. 20. In the  $G_{\text{topo}}$ , the spatial connections between P3 and all other wells were set to zero, indicating that P3 no longer influenced the pressure distribution of nearby wells. In the  $G_{\text{bin}}$ , the injection-production links related to well P3 were also set to zero, meaning the well stopped participating in reservoir fluid exchange. In contrast, the  $G_{\text{cond}}$  retained the reservoir's heterogeneity information and remained unchanged. Similarly, the  $G_{\text{dyn}}$  preserved the dynamic similarity of production and did not require modification. It should be noted that Figs. 21, 22, 23 and 24 in this section follow the same sub-figure configuration as Figs. 18 and 19, and the details are not repeated here.

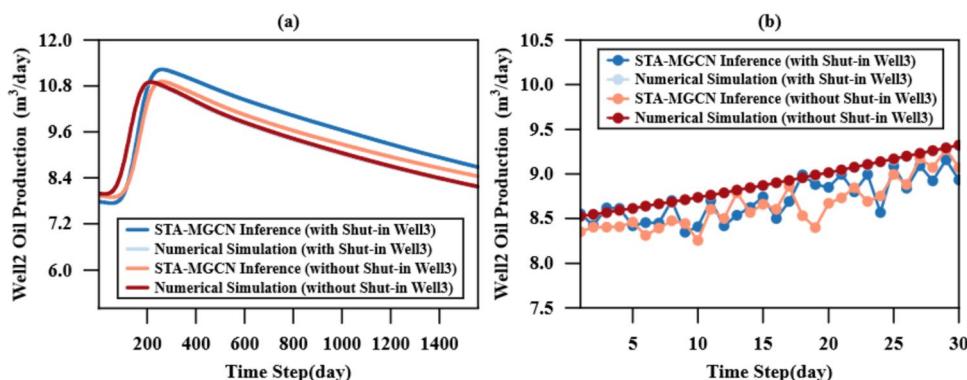
Figure 21 compared STA-MGCN forecasts with numerical simulation results for well P2 following the shut-in of well P3. Panel (a) showed that the predicted and simulated curves matched closely throughout the full 2025–2029 production period. After well P3 was shut in, the oil production of well P2, which implemented the liquid-lifting measure, increased compared to the case without a shut-in of well P3. This increase was caused by the redistribution of injected water from the high-permeability zone to surrounding low-permeability areas.

Panel (b) presented a 30-day prediction window. During this period, the STA-MGCN forecast remained within 0.5 m<sup>3</sup>/d of the simulation result. The overall trend of the predicted curve aligned well with the simulator. These results confirmed that the model accurately captured inter-well interference effects in oil production forecasting.

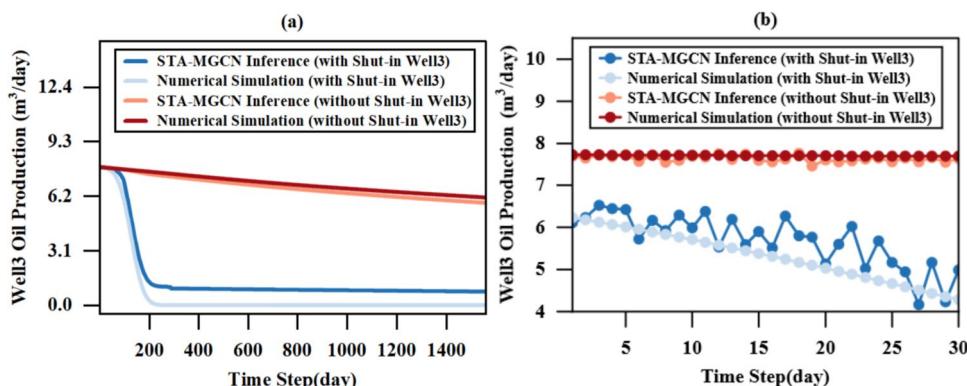
Figure 22 compared STA-MGCN forecasts with numerical simulation results for well P3 after the shut-in. Panel (a) showed that the model reproduced the sharp decline in production but slightly overestimated the tail,



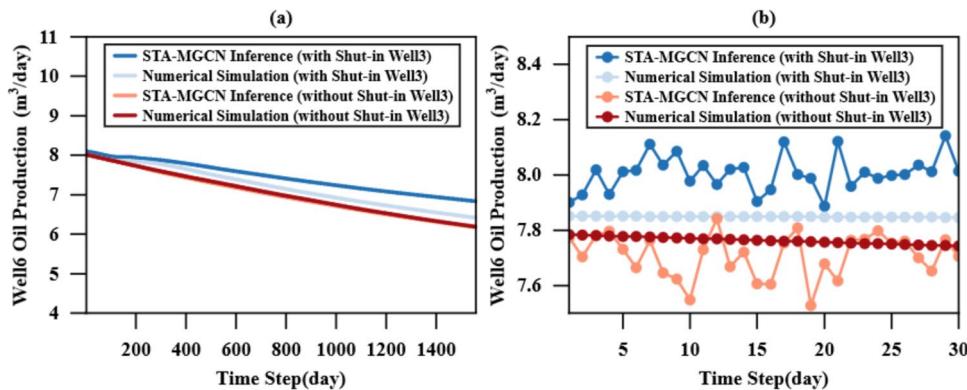
**Fig. 20.** Geometric topological graph (a) and injection-production binary connectivity graph (b) after P3 shut-in.



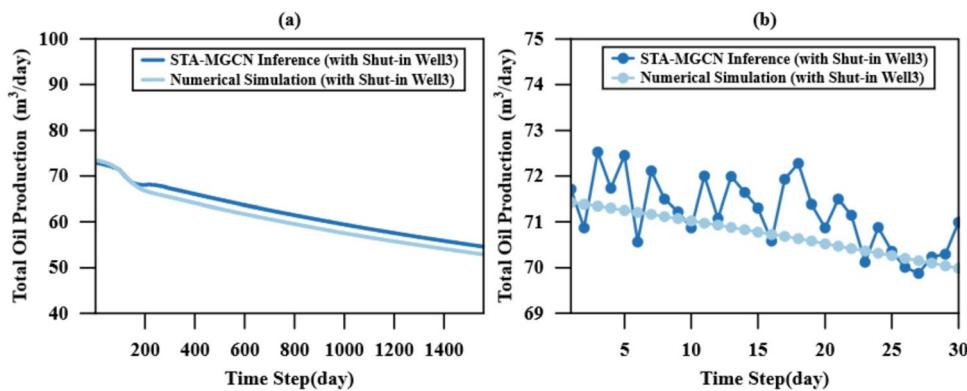
**Fig. 21.** Comparison between STA-MGCN forecasts and numerical simulation results for well P2 under the well P3 shut-in scenario. (a) Complete time series with and without the well P3 shut-in intervention. (b) 30-day forecast window comparing the short-term production response with and without well P3 shut-in.



**Fig. 22.** Comparison between STA-MGCN forecasts and numerical simulation results for well P3 under the shut-in scenario. (a) Complete time series with and without the well P3 shut-in intervention. (b) 30-day forecast window comparing the short-term production response with and without well P3 shut-in.



**Fig. 23.** Comparison between STA-MGCN forecasts and numerical simulation results for well P6 under the well P3 shut-in scenario. (a) Complete time series with and without the well P3 shut-in intervention. (b) 30-day forecast window comparing the short-term production response with and without well P3 shut-in.



**Fig. 24.** Comparison between STA-MGCN forecasts and numerical simulation results for total production of the well pattern under the well P3 shut-in scenario. (a) Complete time series with and without the well P3 shut-in intervention. (b) 30-day forecast window comparing the short-term production response with and without well P3 shut-in.

stopping short of zero. This result indicated that the model responded to the shut-in operation, although it did not fully capture the complete cessation of production.

Panel (b) displayed a 30-day prediction window. During this period, the deviation between the model predictions and the simulation results remained below 1 m<sup>3</sup>/d, with only minor fluctuations. These results indicated that the model demonstrated strong generalization ability, despite not being explicitly trained on shut-in operations.

Figure 23 compared STA-MGCN forecasts with numerical simulation results for well P6 following the shut-in of well P3. Panel (a) illustrates that the STA-MGCN prediction curves closely align with the simulator curves throughout the entire production time series. After well P3 was shut in, well P6, which did not implement liquid lifting, exhibited an increase in oil production. Panel (b) presented a 30-day prediction window. During this period, the difference between the model and the simulation remained within 0.4 m<sup>3</sup>/d. The predicted trend was consistent with the simulation curve. These findings further confirmed that STA-MGCN effectively captured inter-well fluid redistribution in oil production forecasting.

Figure 24 compared STA-MGCN forecasts with numerical simulation results for the total production of the well pattern after the shut-in of well P3. Panel (a) showed a consistent decline in total production. The model captured this downward trend but slightly overestimated the values compared to the simulation and exhibited minor fluctuations at certain time steps.

Panel (b) showed a 30-day prediction window. During this period, the difference between the STA-MGCN forecast and the numerical simulation stayed below 2 m<sup>3</sup>/d. The predicted curve closely followed the simulation trend and showed an overall downward pattern. The model slightly overestimated the total production mainly because it did not fully reduce the output of well P3 to zero. In addition, the shut-in of P3 slightly increased the production of adjacent wells P2 and P6, which together led to a small rise in total production.

Scenario Description	STA-MGCN Time	CMG Numerical Simulator Time	Speedup Factor
Historical data fitting	302 (Training)	8160 (15-Year Daily Simulation)	27×
Liquid-lifting effect prediction	1.58 (Inference)	1680 (5-Year Daily Simulation)	1063×
Shut-in impact prediction	0.897 (Inference)	1500 (5-Year Daily Simulation)	1672×

**Table 9.** Computational efficiency comparison between STA-MGCN and numerical simulator.

#### Computational efficiency comparison

To evaluate the engineering practicality of STA-MGCN, we compared its computational efficiency with that of the CMG numerical simulator using the same hardware configuration (Intel Core i7-11700 CPU, 48 GB RAM, NVIDIA GTX 1660 SUPER GPU). Table 9 presents the run times for each case.

STA-MGCN completed model training on 15 years of daily production data in 302 s, while CMG required 8,160 s, resulting in a 27× speed-up. For the 5-year liquid lifting inference, STA-MGCN generated results in 1.58 s, compared to 1,680 s for CMG, achieving a 1,063× speed-up. In the shut-in inference scenario, STA-MGCN took only 0.897 s, whereas CMG took 1,500 s, yielding a 1,672× improvement.

This computational advantage comes from STA-MGCN's end-to-end neural architecture, which avoids the iterative solving of partial differential equations. Although training takes longer than a single inference, it is still about two orders of magnitude faster than traditional simulators. More importantly, STA-MGCN can perform multi-scenario forecasting within seconds. This supports real-time decision-making and enables rapid adjustments in reservoir management, showing strong potential for practical applications.

#### Discussions

In terms of prediction accuracy (see Figs. 13, 14, 15, 16 and Table 7), the proposed STA-MGCN model performed better than the benchmark models LSTM, GCN-LSTM, and MGCN-LSTM. It achieved high accuracy, low error, and strong stability within the normal oil production range. Minor deviations and fluctuations appeared only at time points with relatively high production values in the time series. This was mainly due to the natural decline in oil production over time, which resulted in fewer high-value samples. However, the overall prediction error remained within an acceptable range.

In contrast, the LSTM model relied only on time series data and failed to account for inter-well interference or reservoir heterogeneity. It predicted each well separately and produced the highest error. GCN-LSTM and MGCN-LSTM enabled multi-well prediction using graph convolution, but their separate modeling of space and time, along with fixed-weight graph fusion, ignored spatiotemporal coupling and graph-specific contributions. Therefore, their accuracy was still lower than that of STA-MGCN. These results highlighted the strengths of STA-MGCN. The model updated temporal and spatial features alternately within each layer and utilized adaptive fusion for multiple graphs. This design effectively captured detailed spatiotemporal dependencies, supported simultaneous production prediction for multiple wells, and significantly improved prediction accuracy.

After confirming the overall performance advantage, the ablation study (Table 8, Fig. 17) further revealed the differential contributions of various spatial graphs to prediction accuracy. Among them, the  $G_{\text{cond}}$  and the  $G_{\text{dyn}}$  played the most significant roles, whereas the  $G_{\text{topo}}$  and the  $G_{\text{bin}}$  made smaller individual contributions. This difference was due to variations in how the graphs captured inter-well relationships.  $G_{\text{cond}}$  combined geological properties such as equivalent permeability with spatial information such as distance decay to approximate inter-well transmissibility.  $G_{\text{dyn}}$  used DTW to capture similarities in injection-production curves and to describe the dynamic responses between wells. Together, these two graphs revealed oil–water migration in porous media from a non-Euclidean perspective.

In contrast,  $G_{\text{topo}}$  and  $G_{\text{bin}}$  only provided structural information in Euclidean space.  $G_{\text{topo}}$  defined adjacency through the geometric shortest path, and  $G_{\text{bin}}$  only indicated whether wells were connected. Neither of them quantified the strength of connectivity.  $G_{\text{cond}}$  and  $G_{\text{dyn}}$  had clear advantages because they directly reflected reservoir heterogeneity and production dynamics. Removing  $G_{\text{cond}}$  caused the loss of key physical flow features, while removing  $G_{\text{dyn}}$  reduced the ability to respond to injection-production adjustments. As a result, the absence of either graph led to a significant decline in prediction accuracy. Although their individual contributions were limited,  $G_{\text{topo}}$  and  $G_{\text{bin}}$  still offered essential structural priors.  $G_{\text{topo}}$  described locality and sparsity in the well pattern, and  $G_{\text{bin}}$  clarified the pairing between injection and production wells. Finally, the adaptive multi-graph fusion mechanism effectively integrated the four types of information.  $G_{\text{cond}}$  and  $G_{\text{dyn}}$  provided physical and dynamic features, while  $G_{\text{topo}}$  and  $G_{\text{bin}}$  constrained spatial structure and inter-well logic. Together, they complemented each other and improved prediction performance.

In terms of inference capability (see Figs. 18, 19, 20, 21, 22, 23 and 24), STA-MGCN demonstrated strong generalization across different scenarios. On the inference dataset, the model accurately inferred the oil production changes of each well following liquid lifting at well P2. This scenario had been included in the training data. It also successfully inferred the impact of shutting in well P3 on the oil production of adjacent wells, despite the shut-in condition not being present during training.

This generalization capability was mainly attributed to the joint effect of the four graph structures.  $G_{\text{bin}}$  and  $G_{\text{topo}}$  provided editable adjacency matrices. When well P3 was shut in, its related edge weights were manually set to zero. In this way, the shut-in operation was mapped to a structural change in the graph. Based on the updated adjacency, the GCN redistributed the information flow between nodes, which caused the model to shift attention to active wells such as P2 and P6, with increased weights assigned to these neighboring wells. At the same time,  $G_{\text{cond}}$  and  $G_{\text{dyn}}$  provided physical and dynamic priors. The former preserved the connectivity of

heterogeneous reservoirs, while the latter maintained the dynamic similarity of injection-production changes between wells. Under this dual mechanism, the predicted production increase of P2 and P6 was not accidental. Instead, it was the natural result of structural editing combined with prior constraints. The adaptive fusion of the four graphs ensured that STA-LSTM could effectively reproduce the fluid redistribution process triggered by the shut-in. It also allowed the model to maintain stable and accurate inference even when facing unseen operational adjustments.

However, there were still limitations in STA-MGCN. The model relied only on graph structures to learn physical patterns implicitly and lacked direct physical hard constraints. Therefore, the model was able to capture the overall declining trend in oil production, but it failed to reduce the prediction of well P3 to zero after shut-in. This limitation indicated that the model's reliability was still insufficient under extreme operating conditions.

It should be noted that this study relies entirely on numerically simulated reservoir data. This choice allowed us to control reservoir heterogeneity and injection-production adjustments for model development and comparison, but it also limits the direct generalizability of the findings to real oilfields. Nevertheless, previous studies have demonstrated that the integration of geological and engineering priors can enhance the applicability of data-driven models in real oilfield development<sup>55</sup>. This suggests that the proposed STA-MGCN model has the potential to be transferred to field datasets. Future research will incorporate real production data from actual reservoirs to further evaluate the robustness and engineering value of data-driven models.

## Conclusions

This study proposed a STA-MGCN. By integrating hybrid temporal sequences with multiple spatial graphs, the model achieved simultaneous and accurate production forecasting for multiple wells in a heterogeneous reservoir. We constructed four types of spatial graphs: geometric topology, injection-production binary connectivity, fluid conductance, and dynamic similarity. The model alternated GCN and STA-LSTM modules to jointly learn spatial neighborhoods and temporal dependencies. An adaptive attention mechanism was applied to fuse the features from each graph, which allowed the model to fully capture the individual contribution of each one. We evaluated the proposed model using production data from a composite five-spot well pattern. It was compared with benchmark models, including LSTM, GCN-LSTM, and MGCN-LSTM. The main conclusions are as follows:

- (1) On the test dataset, STA-MGCN showed strong performance, with RMSE reduced to  $0.237 \text{ m}^3/\text{d}$  and  $R^2$  reaching 0.974. These results were significantly better than those of the benchmark models, confirming the effectiveness of the model architecture.
- (2) The model demonstrated strong generalization ability and physical consistency. Under various scenarios such as liquid lifting and well shut-in, the prediction results closely aligned with reservoir physical behavior. This confirmed that the model effectively learned inter-well interference mechanisms and accurately captured the spatiotemporal features of the reservoir.
- (3) STA-MGCN maintained high prediction accuracy under unseen production scenarios and on independent inference datasets. Compared with numerical simulation methods, its computational efficiency improved by a factor of 10 to 1000. These advantages confirmed the model's strong potential for real-time decision support and practical engineering applications.

Future research should focus on improving the model's generalizability by incorporating physical hard constraints. It should also aim to expand applications to a wider range of reservoir types and development modes. In addition, exploring dynamic graph representation methods will be important for enhancing the model's potential for field deployment.

## Data availability

The data used to support the findings of this study are available from the corresponding author upon request.

Received: 11 June 2025; Accepted: 29 October 2025

Published online: 28 November 2025

## References

1. Wang, H. L., Mu, L. X., Shi, F. G. & Dou, H. G. Production prediction at ultra-high water cut stage via Recurrent Neural Network. *Pet. Explor. Dev.* **47**(5), 1084–1090. [https://doi.org/10.1016/s1876-3804\(20\)60119-7](https://doi.org/10.1016/s1876-3804(20)60119-7) (2020).
2. Zhang, X., Du, F., Zhao, H., Ge, L. & Deng, J. Optimization of injection-production relationship of offshore unconsolidated sandstone reservoirs based on inter-well connectivity. *J. Energy Resour. Technol. Part B Subsurface Energy Carbon Capture* **1**(1), 1–10. <https://doi.org/10.1115/1.4066279> (2025).
3. Thompson, J. M., M'Angha, V. O. & Anderson, D. M. Advancements in shale gas production forecasting—A Marcellus case study. In *North American Unconventional Gas Conference and Exhibition*, 2011, vol. North American Unconventional Gas Conference and Exhibition, SPE-144436-MS. <https://doi.org/10.2118/144436-ms>.
4. Liu, W. B. & Zhang, G. M. Enhancing oil-water flow simulation in shale reservoirs with fractal theory and meshless method. *Front. Environ. Sci.* **11**, 1244543. <https://doi.org/10.3389/fenvs.2023.1244543> (2023).
5. Tripathi, B. K., Kumar, I., Kumar, S. & Singh, A. Deep learning-based production forecasting and data assimilation in unconventional reservoir. *SPE J.* **29**(10), 5189–5206. <https://doi.org/10.2118/223074-pa> (2024).
6. Fan, D. Y. et al. Review of machine learning methods for steady state capacity and transient production forecasting in oil and gas reservoir. *Energies* **18**(4), 842. <https://doi.org/10.3390/en18040842> (2025).
7. Rahmanifard, H. & Gates, I. A Comprehensive review of data-driven approaches for forecasting production from unconventional reservoirs: best practices and future directions. *Artif. Intell. Rev.* **57**(8), 213. <https://doi.org/10.1007/s10462-024-10865-5> (2024).
8. Liu, W., Liu, W. D. & Gu, J. Forecasting oil production using ensemble empirical model decomposition based Long Short-Term Memory neural network. *J. Pet. Sci. Eng.* **189**, 107013. <https://doi.org/10.1016/j.petrol.2020.107013> (2020).

9. Huang, T. et al. A time patch dynamic attention transformer for enhanced well production forecasting in complex oilfield operations. *Energy* **309**, 133186. <https://doi.org/10.1016/j.energy.2024.133186> (2024).
10. Alali, Z. H. & Horne, R. N. A Comparative study of deep learning models and traditional methods in forecasting oil production in the volve field. In *SPE Annual Technical Conference and Exhibition*, 2023, Vol. SPE Annual Technical Conference and Exhibition, D031S032R003, <https://doi.org/10.2118/214881-ms>.
11. Sagheer, A. & Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **323**, 203–213. <https://doi.org/10.1016/j.neucom.2018.09.082> (2019).
12. Kumar, I., Tripathi, B. K. & Singh, A. Attention-based LSTM network-assisted time series forecasting models for petroleum production. *Eng. Appl. Artif. Intell.* **123**, 106440. <https://doi.org/10.1016/j.engappai.2023.106440> (2023).
13. Pan, S. et al. 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> (2023).
14. Lu, S., Cui, C., Qian, Y. & He, J. Well production forecast post-liquid lifting measures: a transformer-based Seq2Seq method with attention mechanism. *Energy Fuels* **38**(15), 14072–14084. <https://doi.org/10.1021/acs.energyfuels.4c02123> (2024).
15. Gurbanov, R. S., Musayeva, S. A., Gurbanov, R. S. & Ahmedov, Z. M. Advanced well spacing system application in the development of oil and gas fields. *Procedia Comput. Sci.* **102**, 446–452. <https://doi.org/10.1016/j.procs.2016.09.425> (2016).
16. Guo, Y., Peng, Y., Hao, R. & Tang, X. Capturing spatial-temporal correlations with Attention-based Graph Convolutional Network for network traffic prediction. *J. Netw. Comput. Appl.* **220**, 103746. <https://doi.org/10.1016/j.jnca.2023.103746> (2023).
17. Xin, X., Liu, S., Ma, R., Yu, G. & Lei, Z. production optimization based on inter-well connectivity analysis using multi-source data in X Oilfield, China. In *ADIPEC*, 2023, Vol. Day 1 Mon, October 02, 2023, D011S026R004, <https://doi.org/10.2118/215940-ms>.
18. Taccari, M. L., Wang, H., Nuttall, J., Chen, X. & Jimack, P. K. Spatial-temporal graph neural networks for groundwater data. *Sci. Rep.* **14**(1), 24564. <https://doi.org/10.1038/s41598-024-75385-2> (2024).
19. Jiang, W., Luo, J., He, M. & Gu, W. Graph neural network for traffic forecasting: the research progress. *ISPRS Int. J. Geo-Inf.* **12**(3), 100 (2023).
20. Liu, R. et al. Federated graph neural networks: Overview, techniques, and challenges. *IEEE Trans. Neural Netw. Learn. Syst.* **36**(3), 4279–4295. <https://doi.org/10.1109/tnnls.2024.3360429> (2025).
21. Gao, M. et al. production forecasting based on attribute-augmented spatiotemporal graph convolutional network for a typical carbonate reservoir in the middle east. *Energies* **16**(1), 407. <https://doi.org/10.3390/en16010407> (2022).
22. Du, L., Liu, Y., Xue, L. & You, G. A deep learning framework using graph convolutional networks for adaptive correction of interwell connectivity and gated recurrent unit for performance prediction. *SPE Reservoir Eval. Eng.* **25**(04), 815–831. <https://doi.org/10.2118/210575-pa> (2022).
23. Du, E. et al. Production forecasting with the interwell interference by integrating graph convolutional and long short-term memory neural network. *SPE Reservoir Eval. Eng.* **25**(02), 197–213. <https://doi.org/10.2118/208596-pa> (2022).
24. Du, E., Liu, Y., Xue, L., Zhu, X. & Song, L. A data-driven model for production prediction of strongly heterogeneous reservoir under uncertainty. *Geoenergy Sci. Eng.* **223**, 211542. <https://doi.org/10.1016/j.geoen.2023.211542> (2023).
25. Li, J., Liu, W., Yu, M. & Xu, W. Reservoir production prediction based on improved graph attention network. *IEEE Access* **12**, 50044–50056. <https://doi.org/10.1109/ACCESS.2023.3344756> (2024).
26. Zhuang, X., Wang, W., Su, Y., Li, Y., Li, L. & Hao, Y. Co-optimization of CO<sub>2</sub>-eor strategies considering the spatio-temporal sequence prediction of CO<sub>2</sub> flooding and sequestration. 2024 2024: SPE, 2024, <https://doi.org/10.2118/218284-ms>.
27. Liu, X., Sun, M., Lin, B. & Gu, S. SGP-GCN: A spatial geological perception graph convolutional neural network for long-term petroleum production forecasting. *Energy Eng* **122**(3), 1053–1072. <https://doi.org/10.32604/ee.2025.060489> (2025).
28. Bentsen, L. Ø., Warakagoda, N. D., Stenbro, R. & Engelstad, P. a unified graph formulation for spatio-temporal wind forecasting. *Energies* **16**(20), 7179 (2023).
29. Luo, C. et al. MG-ASTN: Multigraph Framework With Attentive Spatial-Temporal Networks for Crowd Mobility Prediction. *IEEE Internet Things J.* **10**(21), 19054–19061. <https://doi.org/10.1109/IOT.2023.3281648> (2023).
30. Alrassas, A. M. et al. Advance artificial time series forecasting model for oil production using neuro fuzzy-based slime mould algorithm. *J. Pet. Explor. Prod. Technol.* **12**(2), 383–395. <https://doi.org/10.1007/s13202-021-01405-w> (2022).
31. Tao, Z., Xu, Q., Liu, X. & Liu, J. An integrated approach implementing sliding window and DTW distance for time series forecasting tasks. *Appl. Intell.* **53**(17), 20614–20625. <https://doi.org/10.1007/s10489-023-04590-9> (2023).
32. Nagao, M., Sun, W. & Sankaran, S. Data-driven discovery of physics for reservoir surveillance. 2022 2022: SPE, <https://doi.org/10.2118/209300-ms>.
33. Yang, L. et al. Predicting oil production after enhancement techniques using multidimensional feature representation learning: A case study of profile control technique. *SPE J.* **29**(08), 3876–3891. <https://doi.org/10.2118/221461-pa> (2024).
34. Elmorsy, M., El-Dakhakhni, W. & Zhao, B. Rapid permeability upscaling of digital porous media via physics-informed neural networks. *Water Resour. Res.* <https://doi.org/10.1029/2023wr035064> (2023).
35. Yu, D., Cao, R. & Li, R. Data-driven intelligent evaluation of injection-production well pattern. In *2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS)*, 946–951 (2019). <https://doi.org/10.1109/DDCLS.2019.8909041>.
36. Schvaneveldt, R. W., Durso, F. T. & Dearholt, D. W. Network structures in proximity data. In *Psychology of Learning and Motivation* Vol. 24 (ed. Bower, G. H.) 249–284 (Academic Press, Cambridge, 1989).
37. Yu, J., Jahandideh, A. & Jafarpour, B. A Neural network model with connectivity-based topology for production prediction in complex subsurface flow systems. *SPE J.* **27**(06), 3426–3445. <https://doi.org/10.2118/209831-pa> (2022).
38. Khodaei, A., Lee, H., Banaei-Kashani, F., Shahabi, C. & Ershaghi, I. A mutual information-based metric for identification of nonlinear injector producer relationships in waterfloods. In *SPE Western Regional Meeting*, Vol SPE Western Regional Meeting, SPE-121395-MS (2009). <https://doi.org/10.2118/121395-ms>.
39. Moghadam, J. N., Salahshoor, K. & Kamp, A. M. Evaluation of waterflooding performance in heavy oil reservoirs applying capacitance-resistive model. *Pet. Sci. Technol.* **29**(17), 1811–1824. <https://doi.org/10.1080/10916461003645435> (2011).
40. Schmetterling, R., Burghi, T. B. & Sepulchre, R. Adaptive conductance control. *Ann. Rev. Control* **54**, 352–362. <https://doi.org/10.1016/j.arcontrol.2022.07.005> (2022).
41. Li, M., Zhu, Y., Zhao, T. & Angelova, M. Weighted dynamic time warping for traffic flow clustering. *Neurocomputing* **472**, 266–279. <https://doi.org/10.1016/j.neucom.2020.12.138> (2022).
42. Danel, T. et al. Spatial graph convolutional networks. In *Neural Information Processing* (eds Yang, H. et al.) 668–675 (Springer International Publishing, Cham, 2020).
43. He, L., Bai, L., Yang, X., Du, H. & Liang, J. High-order graph attention network. *Inf. Sci.* **630**, 222–234. <https://doi.org/10.1016/j.ins.2023.02.054> (2023).
44. He, M., Wei, Z. & Wen, J.-R. Convolutional neural networks on graphs with chebyshev approximation, revisited. Presented at the Proceedings of the 36th International Conference on Neural Information Processing Systems, New Orleans, LA, USA (2022).
45. Wang, J. et al. STHGCN: A spatiotemporal prediction framework based on higher-order graph convolution networks. *Knowl. Based Syst.* **258**, 109985 (2022).
46. Geng, X. et al. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. presented at the Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, Hawaii, USA, 2019. Available: <https://doi.org/10.1609/aaai.v33i01.33013656>.

47. Wu, Y., Ding, Y., Zhu, Y., Feng, J. & Wang, S. Complexity to forecast flood: Problem definition and spatiotemporal attention LSTM solution. *Complexity* **2020**, 1–13. <https://doi.org/10.1155/2020/7670382> (2020).
48. Fu, E., Zhang, Y., Yang, F. & Wang, S. Temporal self-attention-based Conv-LSTM network for multivariate time series prediction. *Neurocomputing* **501**, 162–173. <https://doi.org/10.1016/j.neucom.2022.06.014> (2022).
49. Lin, L., Li, W., Bi, H. & Qin, L. Vehicle trajectory prediction using LSTMs with spatial-temporal attention mechanisms. *IEEE Intell. Transp. Syst. Mag.* **14**(2), 197–208. <https://doi.org/10.1109/MITS.2021.3049404> (2022).
50. Xie, J., Zhao, Z., Lin, Z. & Shen, Y. Multimodal graph learning for cross-modal retrieval. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, 145–153.
51. Zhang, D., Zhang, Z., Chen, N. & Wang, Y. Dynamic convolutional time series forecasting based on adaptive temporal bilateral filtering. *Pattern Recogn.* **158**, 110985. <https://doi.org/10.1016/j.patcog.2024.110985> (2025).
52. Hodson, T. O. Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev.* **15**(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022> (2022).
53. Pannakkong, W., Thiw-Anont, K., Singthong, K., Parthanadee, P. & Buddhakulsomsiri, J. Hyperparameter tuning of machine learning algorithms using response surface methodology: A case study of ANN, SVM, and DBN. *Math. Probl. Eng.* **2022**, 1–17. <https://doi.org/10.1155/2022/8513719> (2022).
54. Szpisják-Gulyás, N. et al. Methods for experimental design, central composite design and the Box–Behnken design, to optimise operational parameters: A review. *Acta Aliment.* **52**(4), 521–537. <https://doi.org/10.1556/066.2023.00235> (2023).
55. Qi, N. et al. Machine learning-based research for predicting shale gas well production. *Symmetry* **16**(5), 600 (2024).

## Author contributions

S.L. and C.C. wrote the main manuscript text. Z.W. and Y.Q. collected reference materials and carried out data visualization. J.W. and Y.H. provided the underlying data sources. All authors reviewed and approved the manuscript.

## Funding

The authors acknowledge the financial support from the National Natural Science Foundation of China (No. 51974343) and the National Science and Technology Major Project of the Ministry of Science and Technology of China (No. 2016ZX05011-002-003).

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to C.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025