



## OPEN Forecasting deep shale gas production using a ROA-optimized Transformer–Mamba hybrid network

Weikang He<sup>1,2,3</sup>, Xizhe Li<sup>1,2,3</sup>✉, Yujin Wan<sup>1</sup>, Nan Wang<sup>1</sup>, Honming Zhan<sup>1,2,3</sup>, Xiangyang Pei<sup>1</sup>✉, Longyi Wang<sup>1,2,3</sup>, Wenxuan Yu<sup>1</sup> & Yuhang Zhou<sup>1,2,3</sup>

Based on continuous production data from 112 wells in a deep shale gas block, comprising approximately 160,000 production days, this study proposes a hybrid deep learning framework integrating the Rabbit Optimization Algorithm (ROA), Transformer, and Mamba architectures for daily shale gas production forecasting. Average casing pressure, daily water production rate, and flowback ratio are employed as input features, while ROA is used to globally optimize model hyperparameters. Model performance and robustness are evaluated using five-fold cross-validation. Compared with the standalone Mamba model, the proposed ROA-Transformer–Mamba framework reduces RMSE from 0.0418 to 0.0328 (approximately 21.5%) and MAE from 0.02438 to 0.0174 (approximately 28.6%), while increasing the coefficient of determination to 0.938, demonstrating superior prediction accuracy and generalization capability. To the best of the authors' knowledge, this is the first study to integrate ROA, Transformer, and Mamba architectures for shale gas production forecasting, providing an effective data-driven solution for capturing complex production dynamics under multi-factor coupling conditions.

**Keywords** Shale gas, Production forecasting, Transformer, Mamba

Amid the ongoing restructuring of the global energy system and the escalating pursuit of low-carbon energy sources, natural gas—especially unconventional forms like shale gas—has emerged as a pivotal component in safeguarding national energy supply and promoting the evolution toward a cleaner energy mix.

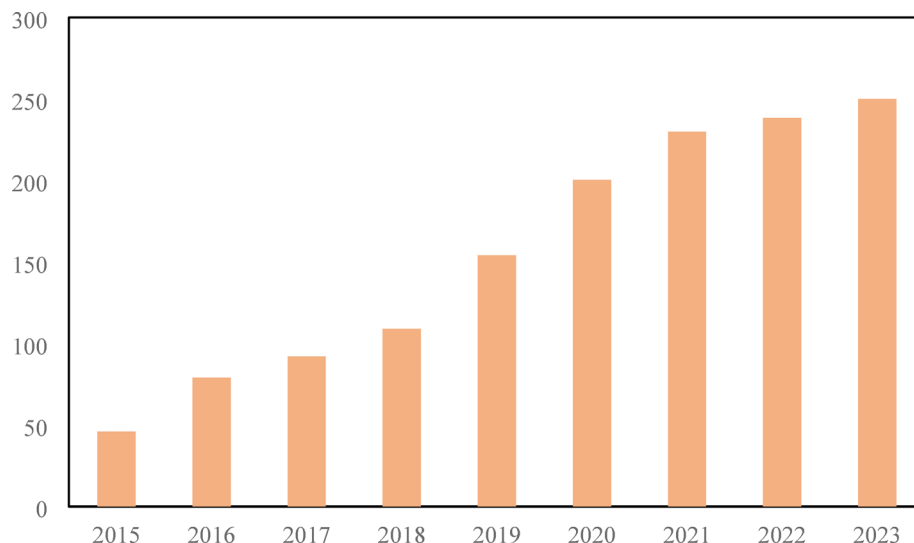
In recent years, China has accelerated the exploration and development of shale gas, achieving significant breakthroughs in regions such as the Sichuan Basin, Fuling, Changning–Weiyuan, and Zhaotong<sup>1–4</sup>. According to statistics from the National Energy Administration, China's annual shale gas production exceeded 25 billion cubic meters in 2023, with its share in total domestic natural gas output continuing to increase. It has become a primary contributor to natural gas production growth during the 14th Five-Year Plan period (see Fig. 1).

However, shale gas resources are typically characterized by deep burial, complex geological structures, and strong reservoir heterogeneity<sup>5</sup>, which makes their production behavior significantly different from that of conventional gas reservoirs and poses substantial challenges for accurate production forecasting.

Traditional shale gas production forecasting methods are largely based on decline curve analysis (DCA), such as Arps models and SEPD curves, whose theoretical foundations rely on statistical or empirical formulations. While these approaches provide useful guidance during the early stages of reservoir development, they generally assume stable flow regimes and simplified production mechanisms. As a result, their prediction accuracy deteriorates significantly under conditions involving strong production fluctuations, gas–water co-production, or frequent well control adjustments, which are commonly observed in deep shale gas wells.

To overcome these limitations, several studies have attempted to enhance early-stage production evaluation by incorporating production dynamics and statistical analysis. Bu Tao et al.<sup>6</sup> proposed a rapid estimation method for ultimate recoverable reserves (EUR) based on flowback-phase dynamic data, enabling preliminary assessment of shale gas potential at an early stage. Zhu Yuancong et al.<sup>5</sup>, through large-scale statistical feature

<sup>1</sup>Research Institute of Petroleum Exploration and Development, Beijing 100089, China. <sup>2</sup>Institute of Porous Flow and Fluid Mechanics, Chinese Academy of Sciences, Langfang 065000, China. <sup>3</sup>School of Engineering Science, University of the Chinese Academy of Sciences, Beijing 101400, China. ✉email: lxz69@petrochina.com.cn; peixy@petrochina.com.cn



**Fig. 1.** Annual shale gas production in China (2015–2023), in billion cubic meters.

analysis, investigated the relationships between shale gas productivity and production parameters, highlighting the importance of data-driven methodologies in capturing complex production behaviors.

In parallel, data-driven machine learning approaches have been increasingly adopted in gas engineering applications. Recent studies published in *Energy* and *Applied Energy* demonstrated that machine learning models can effectively predict key production-related parameters—such as deliverability in underground natural gas storage systems—offering competitive accuracy with substantially reduced computational cost compared to physics-based simulation<sup>7,8</sup>. However, most of these approaches rely on static or shallow learning models and fail to explicitly account for temporal dependency and long-range dynamics inherent in gas production processes.

With the rapid advancement of artificial intelligence, deep learning-based time-series prediction models<sup>9–12</sup> have gained increasing attention in oil and gas production forecasting. Among them, Long Short-Term Memory (LSTM) networks have shown strong capability in modeling nonlinear temporal dependencies and have been successfully applied to daily shale gas production prediction across multiple blocks. To address data scarcity and generalization issues, Alolayan et al.<sup>9</sup> introduced a transfer learning framework, while Nguyen-Le et al.<sup>10</sup> developed a multivariate input strategy combining early production data to improve forecasting performance in the Barnett shale reservoir.

More recently, hybrid deep learning architectures have been explored to further enhance sequence modeling capability. Qiao Songbo et al.<sup>13</sup> proposed a hybrid REMD–CNN–Transformer–LSTM framework for complex time-series prediction, demonstrating the advantages of integrating multiple representation mechanisms. Liang et al.<sup>14</sup> introduced a BiLSTM–RF–MPA model tailored to the nonstationary characteristics of shale gas production, achieving improved robustness and prediction accuracy. In summary, LSTM-based models are effective in capturing short-term temporal dependencies but often struggle with long-range dynamics, CNN-based approaches focus on local pattern extraction with limited temporal context, whereas Transformer-based models excel at global dependency modeling at the cost of increased computational complexity.

Meanwhile, the success of the Transformer architecture in natural language processing and time-series modeling has provided new opportunities for capturing long-range dependencies in complex sequential data. In addition, state-space models (SSMs), such as the recently proposed Mamba architecture, have shown strong potential in efficiently modeling long sequences with linear computational complexity, offering a promising alternative to attention-based mechanisms.

Motivated by these developments, this study proposes a hybrid prediction framework that integrates the global dependency modeling capability of Transformer with the efficient long-range state-space representation of Mamba, while employing the Rabbit Optimization Algorithm (ROA) for automated hyperparameter optimization. The proposed model is systematically evaluated against Transformer–LSTM and standalone Mamba baselines using real production data from deep shale gas wells.

The main objective of this study is to develop a Transformer–Mamba hybrid model for shale gas production forecasting and to systematically evaluate its predictive accuracy and generalization capability using large-scale field production data, with particular emphasis on its ability to capture production dynamics under complex operating conditions. The remainder of this paper is organized as follows: Sect. 2 describes the data and methodology, Sect. 3 presents the results and discussion, and Sect. 4 summarizes the main conclusions and outlines directions for future research.

## Research methods

### Problem definition and modeling objective

Daily shale gas production forecasting is essentially a multivariate time-series regression problem, which aims to predict future daily gas production trends based on historical production and operational parameters. At time step  $t$ , the production state of a shale gas well is represented by a multidimensional input vector

$$\mathbf{x}_t = [x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(n)}] \quad (1)$$

which includes key production variables such as pressure, water production rate, and flowback ratio. The corresponding prediction target is the daily gas production at the next time step, denoted as  $y_{t+1}$ .

In this study, a sliding time-window approach is adopted, where a historical sequence of length  $L$ ,

$$\{\mathbf{x}_{t-L+1}, \dots, \mathbf{x}_t\} \quad (2)$$

is used as the model input to perform regression and predict future daily gas production. This formulation enables the model to learn the dynamic mapping between historical operational states and subsequent production responses.

### Transformer–Mamba hybrid model architecture

To simultaneously capture short-term fluctuations and long-term evolutionary trends in shale gas production data, a hybrid deep learning model integrating Transformer<sup>15–19</sup> and Mamba<sup>20–25</sup> modules is developed in this study. The overall architecture of the proposed model is illustrated in Fig. 2. The model first employs a one-dimensional convolutional layer to project the original input features into a higher-dimensional space, thereby enhancing the representation of local temporal patterns. Subsequently, a Transformer encoder is introduced, in which the multi-head self-attention mechanism is used to model global dependencies among different time steps and to capture the coupling effects of production parameters across multiple temporal scales.

Following the Transformer module, a Mamba module is incorporated to perform state-space modeling of the time series. Unlike attention-based architectures, Mamba leverages a selective state update mechanism to compress and propagate historical information, enabling efficient modeling of dynamic state evolution in long sequences while maintaining linear computational complexity. This property makes the Mamba module particularly suitable for shale gas production data, which are often characterized by long-term gradual decline superimposed with localized perturbations. Finally, a fully connected layer maps the high-dimensional temporal features to the predicted future daily gas production.

The Mamba component is implemented as a lightweight, Mamba-inspired selective state-space modeling block to capture long-range temporal dependencies efficiently. The state dimension  $N$  is set equal to the Transformer embedding dimension ( $N = 32$ ), ensuring consistent feature representation across modules. Continuous-time state evolution is approximated through discrete convolutional operators, providing an efficient discretisation of state transitions. Selective information propagation is realized via gated nonlinear transformations and residual connections, enabling adaptive compression and transmission of historical production dynamics.

To clearly describe the data flow within the model and the interconnections among different modules, the tensor structure of the proposed Transformer–Mamba hybrid model is detailed as follows. Let  $B$  denote the batch size,  $L$  the length of the sliding time window, and  $F$  the number of input features (in this study,  $F = 3$ , corresponding to average casing pressure, daily water production, and flowback ratio). The input tensor of the model can thus be expressed as

$$\mathbf{X} \in \mathbb{R}^{B \times L \times F} \quad (3)$$

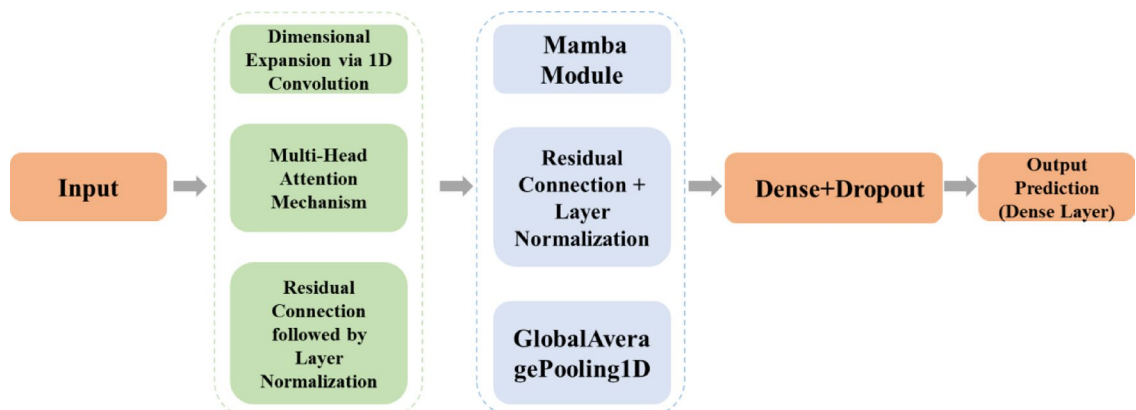


Fig. 2. Hybrid model architecture.

First, a one-dimensional convolutional layer (Conv1D) is applied to perform linear mapping and feature expansion on the input sequence, thereby enhancing the representation of local temporal patterns. The resulting output tensor is given by

$$\mathbf{X}_c \in \mathbb{R}^{B \times L \times D_c} \tag{4}$$

where  $D_c$  denotes the dimensionality of the convolutional feature space.

Subsequently, the convolutional output is fed into a Transformer encoder module. Through the multi-head self-attention mechanism, the Transformer models global dependencies across different time steps while preserving the temporal dimension, yielding an output tensor expressed as

$$\mathbf{X}_t \in \mathbb{R}^{B \times L \times D_t} \tag{5}$$

where  $D_t$  represents the hidden dimension of the Transformer.

On this basis, a Mamba module is introduced to perform state-space modeling of the time series. By means of a selective state update mechanism, the Mamba module compresses and propagates dynamic information over long sequences, producing an output tensor of the form

$$\mathbf{X}_m \in \mathbb{R}^{B \times L \times D_m} \tag{6}$$

where  $D_m$  denotes the state dimension of the Mamba module.

Finally, the temporal features are aggregated along the time dimension (e.g., by selecting the last time step or applying global average pooling) to obtain a fixed-dimensional representation, which is then passed through a fully connected (Dense) layer to generate the prediction. The final model output is expressed as

$$\hat{\mathbf{y}} \in \mathbb{R}^{B \times 1} \tag{7}$$

corresponding to the predicted daily gas production at the future time step.

### Benchmark model configuration

To evaluate the effectiveness of the proposed model, two commonly used models are selected as benchmarks: a Transformer–LSTM hybrid model and a standalone Mamba model. The Transformer–LSTM model combines the global attention mechanism of the Transformer with the recurrent memory structure of LSTM<sup>14,26,27</sup> to capture both short-term and long-term dependencies in time-series data. The Mamba model is employed to assess the performance of a pure state-space architecture in the shale gas production forecasting task.

The basic structure of the LSTM unit and its gate mechanisms are illustrated in Fig. 3.

The commonly used formulation for the LSTM model is given as follows:

$$cf_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f) \tag{8}$$

$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i) \tag{9}$$

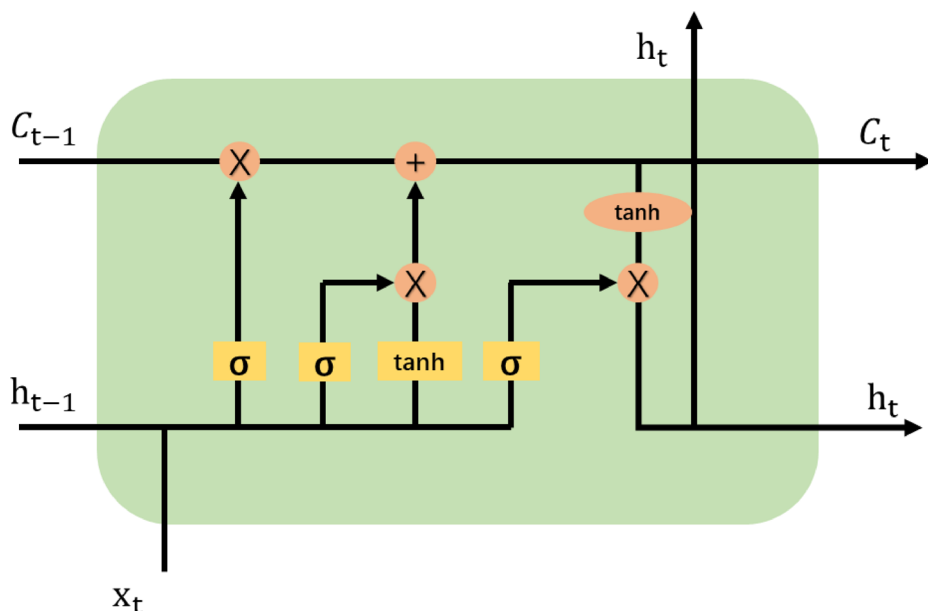


Fig. 3. Basic unit of the LSTM network.

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (10)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (11)$$

$$(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12)$$

$$h_t = o_t * \tanh(C_t) \quad (13)$$

Where  $x_t$  denotes the input at time step  $t$ , I, F, and O represent the input gate, forget gate, and output gate, respectively. The symbol  $\sigma$  denotes the sigmoid activation function, and  $\tanh$  denotes the hyperbolic tangent function.

For a fair comparison, all models adopt the same input features, data preprocessing procedures, and training-testing split strategy.

### ROA-based hyperparameter optimization strategy

To mitigate the uncertainty introduced by manual hyperparameter tuning and to further enhance prediction accuracy and generalization performance, the Rabbit Optimization Algorithm (ROA)<sup>28–30</sup> is employed to perform global hyperparameter optimization for the proposed model. In ROA, each candidate hyperparameter configuration is treated as an individual “rabbit,” and the optimal solution is searched through an adaptive process that balances global exploration and local exploitation.

In this study, ROA is used to optimize key model hyperparameters, including the learning rate, hidden layer dimensionality, number of attention heads in the Transformer module, and the state dimension of the Mamba module. The optimization objective is defined as the root mean squared error (RMSE) on the validation set. During the iterative optimization process, the positions of individual rabbits are updated based on fitness evaluation, and the algorithm terminates when either the maximum number of iterations is reached or the objective function converges. The optimal hyperparameter configuration obtained by ROA is then adopted for final model training.

### Model training and evaluation

During model training, a five-fold cross-validation strategy is adopted to evaluate model stability and robustness. In each fold, the dataset is divided into training and validation subsets, with the training subset used for parameter learning and the validation subset used for performance assessment and hyperparameter optimization. After cross-validation, the model is retrained using the optimal hyperparameters and evaluated on an independent test set to assess its generalization capability.

Model performance is quantitatively evaluated using several commonly adopted regression metrics, including mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination ( $R^2$ )<sup>31</sup>. By comparing these metrics across different models, the predictive accuracy, stability, and generalization performance of the proposed Transformer–Mamba model are systematically assessed.

To eliminate scale differences among wells and improve the stability of model training, all input variables and the target variable (daily gas production) were normalized using Min–Max scaling prior to model training. As a result, the model was trained and evaluated in a normalized space. The predicted and observed values of daily gas production are therefore dimensionless and range between 0 and 1. Accordingly, the evaluation metrics reported in this study, including mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE), are dimensionless quantities that reflect relative prediction accuracy rather than absolute production rates. The coefficient of determination ( $R^2$ ) remains invariant to linear scaling and thus retains its conventional interpretation.

The corresponding formulas are as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (17)$$

Where  $y_i$  is the actual value,  $\hat{y}_i$  is the predicted value,  $\bar{y}$  is the mean of the actual values, and  $n$  is the total number of samples.

### Ethical approval and compliance

All methods were carried out in accordance with relevant guidelines and regulations. No human participants or animals were involved in this study. The use of shale gas production data was approved by Southwest Oil & Gasfield Company under a data use agreement.

## Results and discussion

### Experimental data

The data used in this study are obtained from the production data management system of an oil and gas company and consist of continuous daily production records from multiple deep shale gas wells within a key development block. To ensure data consistency and comparability, all selected wells are completed in the same reservoir formation and adopt similar hydraulic fracturing designs and production schemes. The dataset covers the entire production history since well commissioning and is characterized by a long time span and complex dynamic variations.

The original dataset includes daily gas production, daily water production, average casing pressure, average tubing pressure, and flowback ratio, among other production and operational parameters, with daily gas production serving as the prediction target. After performing data preprocessing procedures—including outlier removal, missing-value imputation, and unit normalization—a sliding time-window approach is employed to construct supervised learning samples, thereby preserving the temporal evolution of production parameters. Flowback ratio is used to characterize the intensity of the post-fracturing flowback process. It is defined as the ratio of the volume of returned (flowback) fluids to the total injected fluid volume during stimulation. In this study, the flowback ratio is defined as the cumulative flowback fluid volume from the start of production at time  $t_0$  to time  $t$ , divided by the total injected fracturing fluid volume:

$$FR(t) = \frac{V_{fb}(t)}{V_{inj}} = \frac{\int_{t_0}^t q_{fb}(\tau) d\tau}{V_{inj}} \quad (18)$$

where  $q_{fb}$  is the flowback-fluid volumetric rate,  $V_{fb}(t)$  is the cumulative flowback volume from the start of production/flowback  $t_0$  to time  $t$ , and  $V_{inj}$  is the total injected fluid volume.

### Data preprocessing

The production data used in this study were obtained from a field production management system of a deep shale gas block. Daily gas production, expressed in  $10^4 \text{ m}^3 \cdot \text{d}^{-1}$ , is taken as the prediction target, while pressure- and water-related parameters are used as candidate inputs. Prior to modeling, the raw data undergo outlier removal, missing-value interpolation, and unit normalization. To preserve the temporal evolution characteristics of production dynamics, supervised learning samples are constructed using a sliding time window approach.

The early production period is divided into a flowback phase and a subsequent stable production phase. The flowback phase is identified as the period immediately following hydraulic fracturing, during which fracturing fluids are continuously recovered and water production remains relatively high. This phase is characterized by a rapid increase in the flowback ratio and strong fluctuations in gas and water production. Only continuous production data after the onset of stable flowback behavior are retained to ensure consistency in time-series modeling.

In terms of input feature selection, a two-stage strategy combining engineering knowledge and statistical analysis is adopted. Based on shale gas production mechanisms, six candidate features are initially selected, including pressure-related parameters as well as flowback- and water-production-related variables. These features describe shale gas well performance from the perspectives of reservoir energy release, fracture conductivity evolution, and gas–water two-phase flow behavior.

Subsequently, the Spearman rank correlation coefficient is employed to evaluate the monotonic relationships between candidate features and daily gas production, thereby reducing redundancy and multicollinearity among inputs. The Spearman coefficient is defined as:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (19)$$

where  $d_i$  denotes the difference between the ranks of two variables, and  $n$  is the number of samples. Unlike Pearson correlation, Spearman correlation measures monotonic dependence rather than linear dependence, making it more suitable for nonlinear production data.

In this study, the correlation strength is classified as follows:

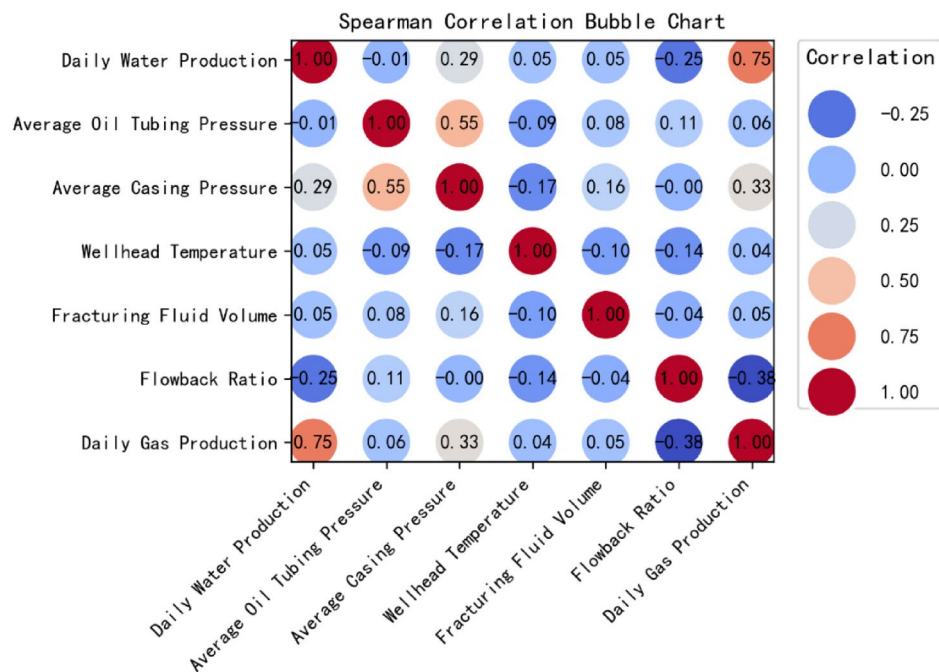
$|\rho| \geq 0.7$  indicates strong correlation,  $0.3 \leq |\rho| < 0.7$  indicates moderate correlation, and  $|\rho| < 0.3$  indicates weak correlation.

The complete correlation matrix and classification thresholds are reported in Table 1 and visualized in Fig. 4.

The analysis results show that average casing pressure, daily water production, and flowback ratio exhibit stable and significant correlations with daily gas production and are therefore selected as the final model inputs.

$\rho$ value	Correlation
$0 <  \rho  \leq 0.3$	Low correlation
$0.3 \leq  \rho  \leq 0.7$	Moderate correlation
$0.7 <  \rho  \leq 1$	High correlation

**Table 1.** Correlation strength.



**Fig. 4.** Spearman correlation analysis.

From a physical perspective, casing pressure reflects reservoir–wellbore pressure transmission and energy depletion, daily water production characterizes fracture flowback behavior and multiphase flow effects, and the flowback ratio represents fracture cleanliness and early-stage conductivity recovery. The combination of these parameters allows the model to capture both production dynamics and underlying physical mechanisms, thereby improving prediction reliability and engineering interpretability.

The weak correlations between fracturing fluid volume and both daily water and gas production ( $\rho \approx 0.05$ ) should be interpreted in the context of the data. The selected wells were completed under similar completion programs, resulting in a relatively narrow distribution of injected fluid volumes, which reduces its statistical leverage. Moreover, instantaneous daily water production during the early production period is affected by choke management, drawdown strategy, and transient wellbore/storage effects, which may decouple daily water rate from total injected volume. Note that min–max normalization to  $[0,1]$  does not change the rank ordering of samples and therefore does not affect Spearman correlation coefficients.

After data preprocessing and sliding-window sequence construction, all samples from all wells were combined into a unified dataset. A five-fold cross-validation strategy was employed for model training and evaluation, in which approximately 80% of the samples were used for training and the remaining 20% for validation in each fold. The final performance metrics were reported as the mean and standard deviation across the five folds. Unlike a fixed train–validation–test split (e.g., 70–20–10), this k-fold cross-validation scheme was adopted to reduce sampling bias and to obtain more reliable and statistically robust performance estimates under limited data conditions. Similar cross-validation strategies have been successfully applied in data-driven reservoir characterization studies, where ensemble learning models were shown to achieve stable generalization performance when predicting lithofacies and permeability for unseen intervals and wells based on well-log data alone<sup>32</sup>.

### Prediction results and analysis

The preprocessed data were fed into the constructed Transformer–Mamba model, and the model’s hyperparameters were optimized using the Rabbit Optimization Algorithm. The predicted daily gas production was then compared with the actual production data to evaluate the model’s effectiveness. As shown in Fig. 5, the Transformer–Mamba model demonstrates good fitting performance and is capable of accurately forecasting future daily gas production.

In the comparative experiments, three model architectures were selected for production forecasting: the baseline Mamba model, a hybrid model (Transformer–LSTM), and the proposed hybrid model (Transformer–Mamba). To ensure fairness in comparison, all three models utilized the same input data format and underwent hyperparameter tuning using the Rabbit Optimization Algorithm.

Figures 6, 7 and 8 collectively present the training and validation loss convergence behaviors of the Transformer–Mamba, Mamba, and Transformer–LSTM models, respectively. For all three models, the training loss decreases monotonically with increasing epochs and stabilizes within approximately 8–10 epochs, indicating efficient and stable optimization. The validation loss follows a consistent downward trend and remains close to the training loss without systematic divergence or late-epoch escalation, providing quantitative evidence against severe overfitting. Among the three models, the Transformer–Mamba model achieves faster convergence

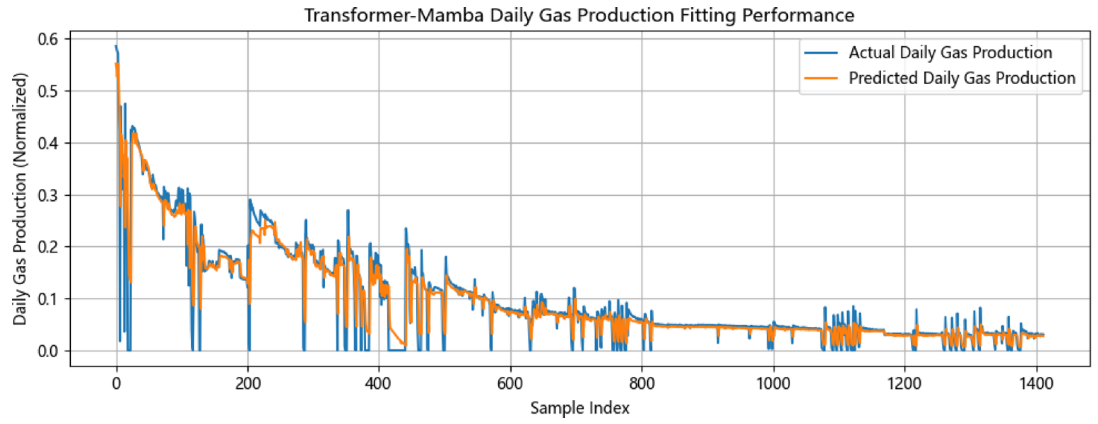


Fig. 5. Prediction results of the Transformer-Mamba model for well X.



Fig. 6. Loss function evolution of the Transformer-Mamba model.

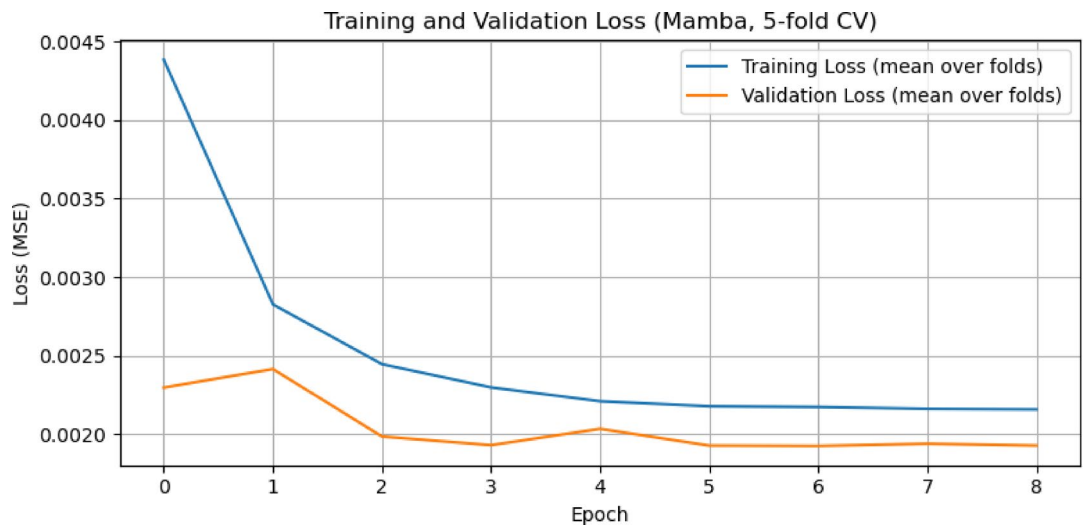
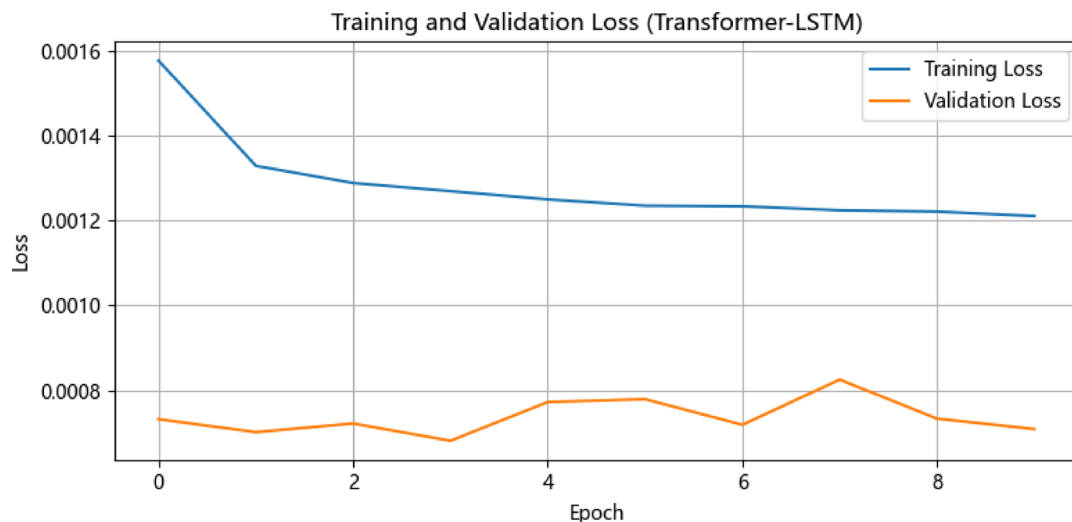
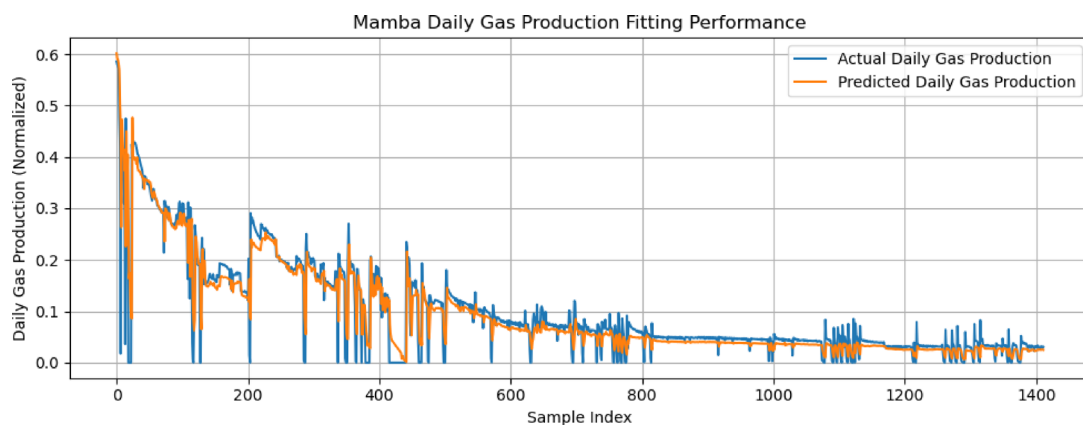


Fig. 7. Loss function evolution of the Mamba model.



**Fig. 8.** Loss function evolution of the Transformer-LSTM model.



**Fig. 9.** Prediction results of the Mamba model.

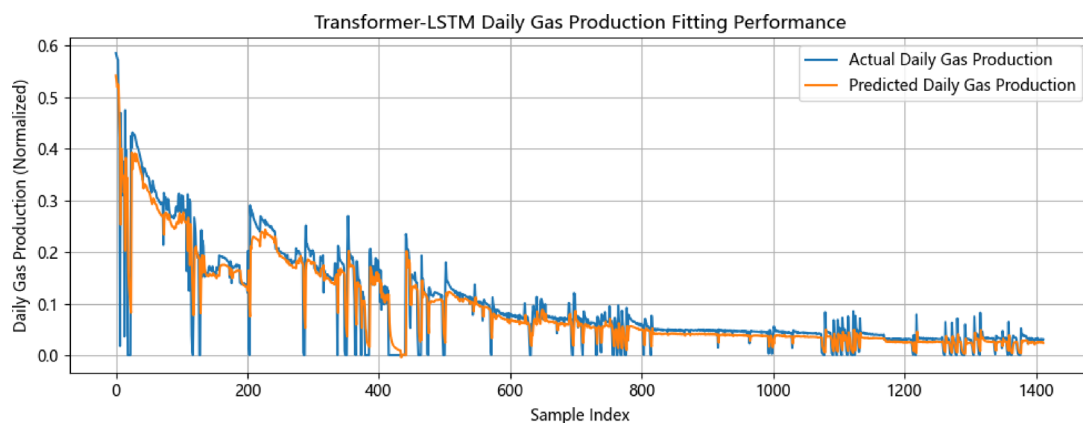
and a lower stabilized validation loss, while the standalone Mamba model exhibits slightly larger early-epoch fluctuations, and the Transformer-LSTM model shows smoother convergence but a higher final loss level. Overall, the comparable convergence patterns and bounded training-validation loss gaps demonstrate stable learning behavior and reasonable generalization performance across all models.

Figures 9 and 10 show the comparison between the predicted results and the actual production data for the Mamba model and the Transformer-LSTM model. It can be observed that both models are capable of effectively capturing the overall trends in the production time series, with their prediction curves closely aligning with the actual data. Among the three models compared, the Transformer-Mamba model demonstrates the best performance in terms of peak prediction, inflection point detection, and fluctuation amplitude reconstruction, followed by the Transformer-LSTM model.

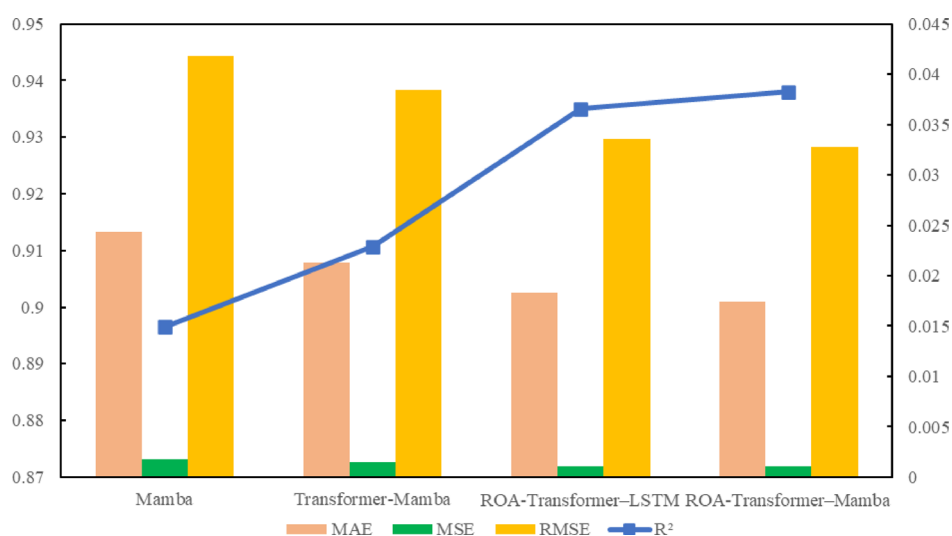
To rigorously evaluate predictive accuracy and robustness, four statistical indicators—MAE, RMSE, MSE, and  $R^2$ —were employed for model comparison. As shown in Fig. 11; Table 2, the Transformer-Mamba model consistently achieves lower MAE, RMSE, and MSE values and a higher coefficient of determination than the Transformer-LSTM baseline, indicating improved predictive accuracy. This improvement suggests that incorporating Transformer-based global attention enhances the effectiveness of the Mamba-based temporal modeling framework.

The proposed Transformer-Mamba architecture serially integrates the global dependency modeling capability of the Transformer with the dynamic state-space representation of a Mamba-like module, making it well suited for shale gas production forecasting. To further examine whether the observed performance gains are statistically significant, a Wilcoxon signed-rank test was conducted on fold-wise results obtained from five-fold cross-validation. The test results confirm that the proposed model achieves statistically significant improvements over baseline models in terms of MAE, RMSE, and  $R^2$  ( $p < 0.05$ ).

Quantitatively, compared with the standalone Mamba model, the ROA-Transformer-Mamba framework reduces RMSE from 0.0418 to 0.0328 (approximately 21.5%) and MAE from 0.02438 to 0.0174 (approximately



**Fig. 10.** Prediction results of the Transformer-LSTM model.



**Fig. 11.** Comparison of evaluation metrics for the three models.

Model	MSE	MAE	RMSE	$R^2$
Mamba	0.0018	0.02438	0.0418	0.8965
Transformer-Mamba	0.0015	0.0213	0.0384	0.9107
ROA-Transformer-LSTM	0.0011	0.0183	0.0336	0.935
ROA-Transformer-Mamba	0.0011	0.0174	0.0328	0.938

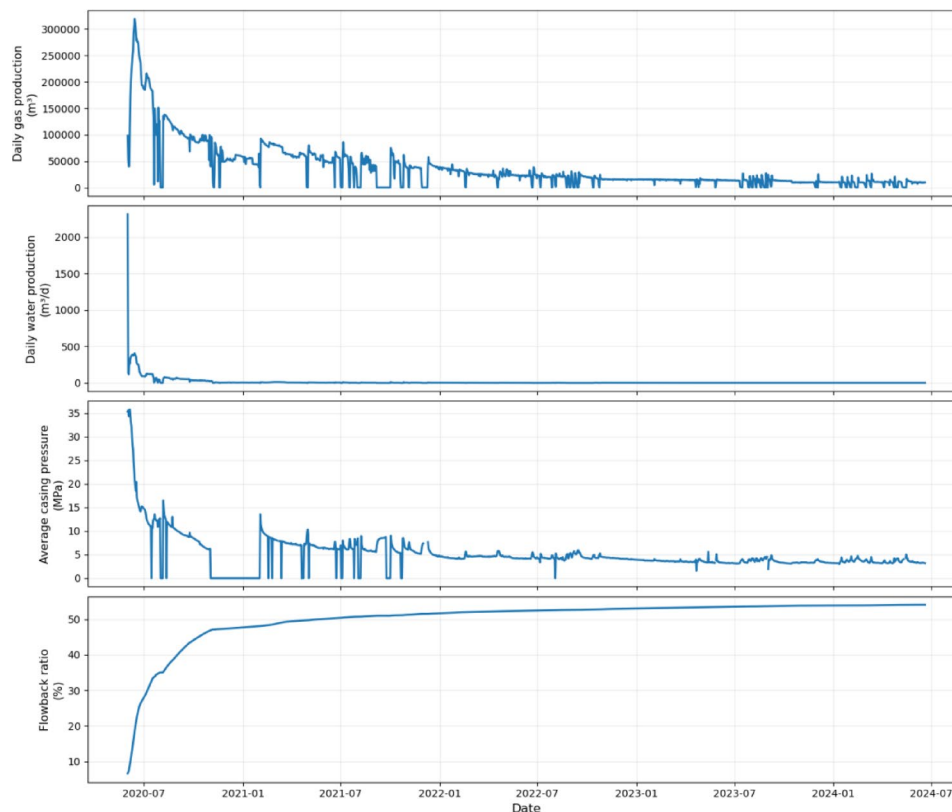
**Table 2.** Comparison of evaluation metrics.

28.6%), while increasing  $R^2$  from 0.8965 to 0.938. Relative to the Transformer-Mamba model without ROA, ROA-based hyperparameter optimization further reduces RMSE by 14.6% and MAE by 18.3%, demonstrating the effectiveness of ROA in improving generalization performance.

These performance gains are consistently observed across five-fold cross-validation and independent test wells, indicating that the proposed model not only achieves higher prediction accuracy but also exhibits a smaller generalization gap when applied to unseen production data. Overall, the results highlight the strong potential of the proposed framework for engineering applications and its value in supporting data-driven decision-making in shale gas field operations.

## Discussion

To mitigate the bias introduced by a single random split, model performance was first assessed using five-fold cross-validation on the training set, followed by independent testing on the held-out dataset. Based on the preprocessed shale gas well production data, a systematic evaluation of the predictive performance of the



**Fig. 12.** Input parameter variation curves.

proposed Transformer–Mamba hybrid model was conducted. The time-series data were divided into training and test sets with a ratio of 8:2, where the training set was used for model parameter learning and the test set was employed to assess generalization performance on unseen data. The benchmark models include Mamba, Transformer–LSTM, and the proposed Transformer–Mamba model. All three models adopt identical input features and data-splitting strategies, and their key hyperparameters are uniformly optimized using the Rabbit Optimization Algorithm (ROA). The evaluation results indicate that the Transformer–Mamba model achieves the lowest RMSE and MAE values and the highest coefficient of determination ( $R^2$ ) on the test set, demonstrating superior prediction accuracy and stability compared with the benchmark models. Notably, the proposed model exhibits enhanced capability in tracking production trends and capturing peak behavior during periods of pronounced production fluctuations.

A comparison between the predicted production curves and the measured daily gas production further shows that the model not only accurately reproduces the overall production decline trend, but also responds reasonably to local fluctuations and inflection points.

As shown in Fig. 12, when average casing pressure decreases continuously, the predicted daily gas production exhibits an accelerated decline, reflecting the controlling effect of reservoir energy depletion on production performance. During periods characterized by relatively high or strongly fluctuating daily water production, the predicted gas rate shows corresponding suppression or delayed response, indicating that the model effectively learns the influence of gas–water two-phase flow on effective gas transport pathways. In addition, during stages when the flowback ratio increases, the predicted production curve displays production recovery or a reduced decline rate, suggesting that the model is able to identify the positive impact of fracture cleanliness improvement on the recovery of fracture conductivity. These predictive behaviors are consistent with the typical production mechanisms observed during the flowback and early stable production stages of shale gas wells.

It should be noted that, although the proposed approach is fundamentally data-driven, its predictions are not merely the result of mathematical curve fitting. Instead, the model reflects, to a certain extent, the combined effects of key physical processes in shale gas production, including pressure transmission, fracture conductivity evolution, and multiphase flow effects. Therefore, the proposed model can be regarded as a surrogate representation of complex physical processes, providing an effective engineering tool for shale gas production forecasting in scenarios where detailed physical parameters or high-fidelity numerical simulations are unavailable. Nevertheless, prediction accuracy may be affected under extreme operating conditions, such as prolonged shut-ins, aggressive choke adjustments, or abnormal pressure drops. Addressing these scenarios by incorporating operational-state indicators or uncertainty analysis constitutes an important direction for future research.

## Conclusion

Based on long-term production data from multiple wells in a deep shale gas block, this study develops a hybrid deep learning framework integrating the Rabbit Optimization Algorithm (ROA), Transformer, and Mamba modules for daily shale gas production forecasting. Compared with benchmark models such as Mamba and Transformer-LSTM, the proposed Transformer-Mamba model achieves lower RMSE and MAE values and a higher coefficient of determination ( $R^2$ ) on the test set, demonstrating improved prediction accuracy and generalization capability, particularly during periods of pronounced production fluctuations and inflection points.

From an engineering perspective, the model's predictive behavior is physically interpretable. The selected input features—average casing pressure, daily water production, and flowback ratio—represent reservoir energy depletion, fracture flowback dynamics, and multiphase flow effects, respectively. The model responses to variations in these parameters are consistent with the typical production mechanisms observed during the flowback and early stable production stages of shale gas wells. This indicates that, although the approach is data-driven, it captures key physical controls on shale gas production rather than relying solely on numerical fitting.

The proposed framework provides a practical data-driven tool for short- to mid-term production forecasting in situations where detailed geological characterization or full-physics numerical simulation is unavailable. Prediction performance may degrade under extreme operating conditions, such as prolonged shut-ins or aggressive choke adjustments. Future work will focus on incorporating operational-state indicators, uncertainty quantification, and transfer learning to further enhance model robustness and cross-block applicability.

## Data availability

The data supporting the findings of this study are available from Southwest Oil & Gasfield Company, but restrictions apply to the availability of these data, which were used under license for the current study, and are therefore not publicly available. However, data may be obtained from the authors upon reasonable request and with permission from Southwest Oil & Gasfield Company.

Received: 23 August 2025; Accepted: 17 March 2026

Published online: 03 April 2026

## References

- Guo, T. L. Review and reflections on shale gas development in China—From the Silurian to the Cambrian. *Oil Gas Reservoir Eval. Dev.* **15**(03), 339–348. <https://doi.org/10.13809/j.cnki.cn32-1825/te.2025.03.001> (2025).
- Dong, D. Z. et al. Strategic recommendations for shale gas development in China. *Nat. Gas Geosci.* **27**(03), 397–406 (2016).
- Nie, H. K. et al. 20 years of shale gas research and development in China: Review and prospects. *Nat. Gas. Ind.* **44** (03), 20–52 (2024).
- Liu, H. Y. et al. Profitable shale gas development in China: Theoretical logic, practical logic, and outlook. *Nat. Gas. Ind.* **43** (04), 177–183 (2023).
- Zhu, Y. C. et al. Shale gas productivity prediction based on big data. *Oil Gas Well Test.* **28**(01), 1–6. <https://doi.org/10.19680/jcnki1004-4388201901001> (2019).
- Bu, T. et al. Rapid evaluation method of shale gas EUR based on flowback period dynamic data. *Unconv. Oil Gas.* **10** (03), 74–79. <https://doi.org/10.19901/jfcgyq20230310> (2023). 102.
- Ali, A. et al. Towards more accurate and explainable supervised learning-based prediction of deliverability for underground natural gas storage. *Appl. Energy* **327**, 120098. <https://doi.org/10.1016/j.apenergy.2022.120098> (2022).
- Ali, A. Data-driven based machine learning models for predicting the deliverability of underground natural gas storage in salt caverns. *Energy* **229**, 120648. <https://doi.org/10.1016/j.energy.2021.120648> (2021).
- Alolayan, O. S., Raymond, S. J., Montgomery, J. B. & Williams, J. R. Towards better shale gas production forecasting using transfer learning. *Upstream Oil and Gas Technology* **9**, 100072. <https://doi.org/10.1016/j.jupstre2022100072> (2022).
- Nguyen-Le, V., Kim, M., Shin, H. & Little, E. Multivariate approach to the gas production forecast using early production data for Barnett shale reservoir. *J. Nat. Gas Sci. Eng.* **87**, 103776. <https://doi.org/10.1016/j.jngse2020103776> (2021).
- Meng, J. et al. Hybrid data-driven framework for shale gas production performance analysis via game theory, machine learning, and optimization approaches. *Pet. Sci.* **20**(1), 277–294. <https://doi.org/10.1016/j.petsci202209003> (2023).
- Chen, Y. et al. Estimation of shale adsorption gas content based on machine learning algorithms. *Gas Sci. Eng.* **127**, 205349. <https://doi.org/10.1016/j.jgsce2024205349> (2024).
- Qiao, S. B. et al. Carbon trading price prediction based on REMD-CNN-Transformer-LSTM hybrid model. *J. Xi'an Univ. Technol.* 1–12 <http://kns.cnki.net/kcms/detail/611294N202504161617010.html> (2025).
- Liang, B. et al. A novel framework for predicting non-stationary production time series of shale gas based on BiLSTM-RF-MPA deep fusion model. *Pet. Sci.* **21**(5), 3326–3339. <https://doi.org/10.1016/j.petsci202405012> (2024).
- Liu, W. T. & Lu, X. M. Research progress of Transformer based on computer vision. *Comput. Eng. Appl.* **58**(06), 1–16 (2022).
- Li, X. et al. A review of Transformer research in computer vision. *Comput. Eng. Appl.* **59**(01), 1–14 (2023).
- Mousa, R., Rezaei, B., Mahmoudi, L. & Abdollahi, J. Multi-modal wound classification using wound image and location by Swin Transformer and Transformer. *Expert Syst. Appl.* **280**, 127077. <https://doi.org/10.1016/j.eswa2025127077> (2025).
- Xu, L., Gao, Z., Li, Y. & Gulliver, T. A. Cross-domain intelligent cooperative spectrum sensing algorithm based on Federated Learning and Swin-Transformer neural network. *Eng. Appl. Artif. Intell.* **157**, 111370. <https://doi.org/10.1016/j.engappai2025111370> (2025).
- Zhang, D., Wang, C., Wang, H., Fu, Q. & Li, Z. An effective CNN and Transformer fusion network for camouflaged object detection. *Comput. Vis. Image Underst.* **259**, 104431. <https://doi.org/10.1016/j.cviu2025104431> (2025).
- Wang, Y. Z. et al. RMFKAN: A cyber army detection method based on improved graph Mamba. *J. Comput. Sci. Explor.* **19** (05), 1365–1378 (2025).
- Wei, X. J., Liu, X. Y. & Zhou, H. L. Seismic random noise suppression method driven by convolutional Mamba model. *Coal Geol. Explor.* **53**(05), 196–206 (2025).
- Wang, M. H., Gao, Y. & Li, J. Image deblurring network based on Mamba and frequency domain fusion. *Comput. Meas. Control* **33**(06), 264–271 (2025).
- Ma, Z., Li, J., Jiang, K. & Wong, W. K. Integrating local and global correlations with Mamba-Transformer for multi-class anomaly detection. *Knowl. Based Syst.* **324**, 113740. <https://doi.org/10.1016/j.knosys2025113740> (2025).

24. Wu, Z., Lu, T., Zhang, Y. & Chai, X. CMANet: A CNN-Mamba aggregation network for face super-resolution. *Pattern Recognit.* **168**, 111859. <https://doi.org/10.1016/j.patcog.2025.111859> (2025).
25. Zhang, J., Shi, X., Feng, Z., Gui, Y. & Wang, J. TMCN: Text-guided Mamba-CNN dual-encoder network for infrared and visible image fusion. *Infrared Phys. Technol.* **149**, 105895. <https://doi.org/10.1016/j.infrared.2025.105895> (2025).
26. Zhou, Q. et al. Shale oil production prediction based on an empirical model-constrained CNN-LSTM. *Energy Geosci.* **5**(2), 100252. <https://doi.org/10.1016/j.jengeos.2023.100252> (2024).
27. Fargalla, M. A. M. et al. TimeNet: Time2Vec attention-based CNN-BiGRU neural network for predicting production in shale and sandstone gas reservoirs. *Energy* **290**, 130184. <https://doi.org/10.1016/j.energy.2023.130184> (2024).
28. Xiang, B.-W., Xiang, Y.-X. & Zhang, T.-Y. Rabbit algorithm for global optimization. *Appl. Math. Model.* <https://doi.org/10.1016/j.apm.2024.115860> (2024).
29. Alsaiani, A. O., Moustafa, E. B., Alhumade, H., Abulkhair, H. & Elsheikh, A. A coupled artificial neural network with artificial rabbits optimizer for predicting water productivity of different designs of solar stills. *Adv. Eng. Softw.* **175**, 103315. <https://doi.org/10.1016/j.advengsoft.2022.103315> (2023).
30. Gülmez, B. Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. *Expert Syst. Appl.* **227**, 120346. <https://doi.org/10.1016/j.eswa.2023.120346> (2023).
31. Shi, X., Xu, M. & Du, J. Max-sum test based on Spearman's footrule for high-dimensional independence tests. *Comput. Stat. Data Anal.* **185**, 107768 (2023).
32. Al-Mudhafar, W. J. & Wood, D. A. *Tree-Based Ensemble Algorithms for Lithofacies Classification and Permeability Prediction in Heterogeneous Carbonate Reservoirs, Presented at the Offshore Technology Conference*, <https://doi.org/10.4043/31780-MS> (2022).

## Author contributions

Weikang He: Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Writing – original draft. Xizhe Li: Supervision, Project administration, Writing – review & editing. Yujin Wan: Validation, Investigation, Data acquisition. Nan Wang: Data curation, Resources, Writing – review & editing. Honming Zhan: Supervision, Funding acquisition, Writing – review & editing. Xiangyang Pei: Methodology, Validation, Writing – review & editing. Longyi Wang: Formal analysis, Visualization, Writing – review & editing. Wenxuan Yu: Data acquisition, Software support. Yuhang Zhou: Resources, Investigation, Writing – review & editing.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to X.L. or X.P.

**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) 2026