Petroleum Research 11 (2026) 399–414

Contents lists available at ScienceDirect

Petroleum Research

journal homepage: http://www.keaipublishing.com/en/journals/petroleum-research/

Full Length Article

Research on segmented modelling and prediction of shale gas production based on time series analysis

Feifei Fang [a,b], Hongjian Chen [a,*], Sijie He [b], Jie Zhang [b,c], Wei Guo [d], Weixiang Jin [b], Yue Gong [b], ChuXiang Xia [b]

[a] School of Mathematical and Physical Sciences, Chongqing University of Science and Technology, Chongqing 401331, China
[b] School of Petroleum Engineering, Chongqing University of Science and Technology, Chongqing 401331, China
[c] Chongqing Unconventional Oil & Gas Development Research Institute, Chongqing University of Science and Technology, Chongqing 401331, China
[d] Research Institute of Petroleum Exploration and Development, PetroChina, Beijing 100083, China

ARTICLE INFO

Article history:
Received 8 May 2025
Received in revised form
16 September 2025
Accepted 28 September 2025
Available online 15 October 2025

Keywords:
Shale gas
Segmented modeling
Full-cycle modeling
IPSO–CNN-BiGRU-attention
Production prediction

ABSTRACT

At present, the world is experiencing an unprecedented period of change, and the third wave of energy conversion is also emerging. Natural gas is widely regarded as an ideal bridge between traditional fossil fuels and new energy transformations due to its significant low-carbon advantages. Forecasting shale gas output each day is key for secure natural gas provision. However, the high dimension and nonlinear characteristics of shale gas production data present significant challenges for prediction. Therefore, this study proposes a segmented modeling method that divides the production cycle into unstable and stable periods and combines the IPSO–CNN-BiGRU-Attention hybrid model to model these two cycles respectively. Comparative results indicate segmented modeling offers superior predictive performance compared to whole production cycle modeling. Furthermore, when modeling different periods, the IPSO–CNN-BiGRU-Attention hybrid model demonstrates a superior predictive effect compared to traditional single time series models.

1. Introduction

The globe now faces unparalleled transformation, coinciding with the rise of a third-generation power shift. Eco-friendly restructuring now significantly propels worldwide advancement. Advancing eco-friendly, enduring progress presents novel prospects alongside significant hurdles. With its significant low-carbon advantage, natural gas is regarded as an ideal bridge between traditional fossil fuels and new energy transformation. It plays an irreplaceable role in cooperative work with energy storage technology, especially in gas-electric peak shaving, which greatly promotes the development of new energy (Zou et al., 2021, 2024). Shale gas output relies on numerous elements, including complex, non-linear, and temporally dynamic geological attributes. Consequently, predicting its output, particularly via time series analysis, remains significantly challenging given its intricate nature and varied determinants (Wang et al., 2020).

At present, shale gas production prediction mainly uses numerical simulation, statistical methods, machine learning and deep learning. Numerical simulations offer crucial insights for forecasting shale gas production. They've proven their worth in fine-tuning well design and sussing out how to maximize shale gas extraction (Kim et al., 2017; Wang et al., 2017). Numerical simulations rely on the bedrock principles of mass and energy conservation, and they require a whole host of inputs, everything from drilling and completion data to geological and mining parameters (Fan et al., 2010; McGlade et al., 2013) utilized reservoir modeling to examine gas generation and well-stimulation performance in the Haynesville shale formation. The findings indicate that the intricate web of fractures is instrumental in boosting the initial output of wells. Moreover, they validate the efficacy of this approach for refining construction blueprints and enhancing overall production efficiency. When running computer simulations of shale gas behavior, researchers have really gone the whole nine yards in accounting for all sorts of complicated movement and flow processes (Wu et al., 2019; Zhang et al., 2021). These traits aid researchers in comprehending shale gas hydraulic fracturing features (You et al., 2019). That said, numerical simulations often demand the development of intricate geological frameworks

* Corresponding author.
E-mail address: chenhongjian@cqust.edu.cn (H. Chen).
Peer review under the responsibility of Editorial Office of Petroleum Research.

https://doi.org/10.1016/j.ptlrs.2025.09.009

and the resolution of extensive Navier-Stokes equations, both of which can be resource-intensive and prone to introducing systemic inaccuracies. In addition, previous published reservoir numerical simulation results for predicting shale gas production are inconsistent, showing challenges in model accuracy and reliability in this field (Liu et al., 2019). Statistical methods such as the autoregressive moving average (ARMA) models are highly effective at crunching past data to forecast very short-term trends. However, such methodologies struggle to model time-series data dynamic, erratic nature.

Fueled by accelerated progress in AI and machine learning, data-centric methodologies have seen substantial evolution in recent decades (Cichos et al., 2020; Kudithipudi et al., 2022). Employing machine learning to assess extraction logs, diverse technical metrics, and shale gas yields proves critical for gauging evolving efficacy and boosting output. To get a handle on the dynamic performance of shale gas (Liang and Zhao, 2019; Mudunuru et al., 2020), leveraged the random forest algorithm. They dug into the production history and a whole host of engineering parameters from the Eagle Ford shale formation to really get under the hood (Kocoglu et al., 2020; Liu et al., 2021). pioneered an artificial neural network methodology to forecast shale gas well Estimated Ultimate Recovery (EUR). With machine learning showing real promise for assessing shale gas production, oil and gas professionals are increasingly keen to explore how various machine learning techniques can be put to use in actual production wells on the ground. Consequently, research now emphasizes developing tailored algorithms for particular datasets (Guo et al., 2021; Han et al., 2020; Niu et al., 2022). explored a range of machine learning techniques, constructing predictive models using a shared dataset to identify the optimal algorithm tailored to particular challenges, thereby enhancing real-world outcomes. While conventional machine learning approaches—including regression trees (Hui et al., 2023), multi-layer perceptrons (Wang et al., 2021a), support vector machines (Niu et al., 2022), and random forests (Xue et al., 2021)—have proven effective across various applications, they fall short when it comes to managing high-dimensional shale gas datasets and tackling intricate nonlinear time series forecasting challenges. These models, despite their strengths, struggle to fully capture the complexity inherent in such data. These models do not perform well in capturing the systematic characteristics and nonlinear relationships of data, which limits their representation ability and application effect.

As deep learning continues to evolve as a specialized field within machine learning, it has proven to be a game-changer in tackling intricate technical challenges and managing high-dimensional data with remarkable efficiency (Lyu and Liu, 2021). In contrast to conventional machine learning techniques, deep learning approaches excel at uncovering intricate, nonlinear patterns within vast datasets, demonstrating superior accuracy in forecasting time series data (Tang et al., 2021). Deep learning detects patterns in high-dimensional data, significantly enhancing prediction precision (Liang et al., 2023). Consequently, numerous studies employ neural network architectures like CNN, RNN, LSTM, and GRU for shale gas production forecasting (Li et al., 2022a; Yang et al., 2021). CNN extracts feature through convolutional layers, and its global sharing feature allows weights to be shared at different locations, thereby improving computational efficiency and reducing the complexity of the model and training time (LeCun et al., 2015). Both LSTM and GRU are variants of RNN, which perform well in processing time and long-term dependent data, overcoming the problem of gradient explosion of RNN (Li et al., 2022b; Alimohammadi et al., 2020; Huang et al., 2022; Sun et al., 2018) analyzed the link between historical output data and shale gas well productivity using LSTM networks. The results show that LSTM performs well in capturing the complex relationship between production data. While deep learning frameworks with a singular structure often fall short in handling the intricacies of natural gas data, this underscores the critical need to refine and enhance these models for more accurate predictions in shale and sandstone gas applications (Wang et al., 2021b). By integrating diverse deep neural networks, the hybrid model leverages the unique strengths of each to efficiently extract pertinent temporal features. This approach optimizes the advantages of a unified network while also improving the capacity to identify fundamental data characteristics in predicting shale gas production (Fan et al., 2021; Zha et al., 2022) forecasted monthly gas field output using a combination of CNN and LSTM models. The findings indicate that the hybrid model is capable of capturing both temporal and spatial characteristics with greater precision, leading to a marked enhancement in prediction accuracy. Furthermore, the attention mechanism has proven effective in forecasting shale gas production as well (Kumar et al., 2023; Wang et al., 2023).

While numerous studies have yielded encouraging outcomes in forecasting shale gas output, notable challenges persist, and the specific problems to be addressed in this study are systematically summarized as follows:(1) Inadequacy of full-cycle modeling: Existing full-cycle modeling methods do not distinguish between the data characteristics of the unstable and stable periods of shale gas production. The high-frequency, large-amplitude fluctuation data of the unstable period and the low-noise, gentle decline data of the stable period interfere with each other, leading to significant prediction errors—especially for wells with severe early-stage fluctuations. (2) Limitations of traditional models in adapting to multi-stage data: Single models and ordinary hybrid models cannot balance the core needs of different production stages. For example, CNN fails to characterize long-term decline trends in the stable period, while RNN/LSTM struggle to capture short-term fluctuations in the unstable period, resulting in insufficient extraction of high-dimensional and nonlinear production data features. (3) Lack of systematic hyperparameter optimization: Most existing hybrid models rely on empirical hyperparameter settings without scientific optimization mechanisms. This leads to unstable model performance and inability to adapt to inter-well differences in geological-engineering conditions, limiting prediction accuracy. (4) Information loss and insufficient focus on key nodes in long-time series: Single temporal models suffer from information loss when processing long-term production data. Additionally, they cannot prioritize key production nodes, resulting in limited capture of critical features.

To address the aforementioned four key issues, this research proposes a targeted technical framework with three core components, each aligned to solve specific defects of existing methods and verified by the subsequent experimental design: (a) A data-driven adaptive segmented modeling approach integrated with Variational Mode Decomposition (VMD) and Pruned Exact Linear Time (PELT) algorithm, dedicated to resolving the inadequacy of full-cycle modeling. (b) An Improved Particle Swarm Optimization (IPSO) algorithm for systematic hyperparameter optimization, addressing the instability of empirically set hyperparameters. (c) An IPSO–CNN-BiGRU-Attention hybrid model designed to tackle the multi-stage data adaptation and long-time series key node focus issues.

The rest of this study is organized as follows: Section 2 offers a concise summary and explanation of the research methodologies employed. Meanwhile, Section 3 elaborates on the segmented modeling approach utilized for shale gas production. Section 4 examines and evaluates the outcomes of both the full-cycle and segmented modeling techniques, confirming the efficacy of the segmented model. In conclusion, Section 5 highlights the main findings and suggests avenues for upcoming research.

2. Methodogy

2.1. Convolutional neural network (CNN)

The convolutional neural network (CNN) is an effective model for handling data with spatial structure characteristics. It effectively extracts spatial hierarchical information and feature information in time series data through hierarchical structure (Feng et al., 2022). A basic convolutional neural network includes an input layer, several convolutional layers, a pooling layer, and a fully connected layer, as illustrated in Fig. 1.

The output from each layer is referred to as feature mapping and serves as the input for the subsequent layer. The process in the convolution layer is referred to as the convolution operation. For 1D-CNN, the formula for the convolution operation is as follows:

$$v_i^p = \sum_{n=1}^{N}\sum_{m=0}^{M-1} k_m^p u_{(i+m)}^n + b^p (i = 1, 2, ..., I) \tag{1}$$

where, u denotes the input data; v signifies the feature mapping; k represents the convolutional filter; b refers to the bias component; N indicates the total input channels; M t specifies the dimensions of the convolutional filter applied to the input data; p signifies the total number of convolutional filters, which matches the number of channels used for feature mapping. The convolutional kernel is utilized on the input data with a variable stride length. Zero-padding is added to the input's edges to prevent edge information loss. As a result, the dimensions of the feature map I rely on the padding width and stride.

The pooling layer performs a process known as pooling operation, creating feature maps v by utilizing the convolution kernel on the input u. For 1D-CNN, the formula for the pooling operation is as follows:

$$v_i^p = F\left(u_{(i+0)}^n, u_{(i+1)}^n, ..., u_{(i+M-1)}^n\right)(i = 1, 2, ..., I) \tag{2}$$

where, F denotes the pooling function. Typical operations within the pooling function are maximum pooling and average pooling. The count of feature mapping channels p matches the count of input channels n. The convolutional kernel of the pooling operation can be applied to the input data using various stride lengths. The pooling operation can utilize zero padding to retain edge information from the input, while the feature map I size is determined by the padding width and stride.

In the fully connected layer, inputs are flattened into a vector and undergo linear transformation. The fully connected layer produces either a vector or scalar. Besides the convolutional and pooling layers, activation functions are also commonly utilized in CNN architectures. The activation function is usually applied element-wise across all channels in the feature map following each layer's operation. This process does not alter the number or size of the feature map's channels but enhances the model's expressive power by introducing nonlinear characteristics.

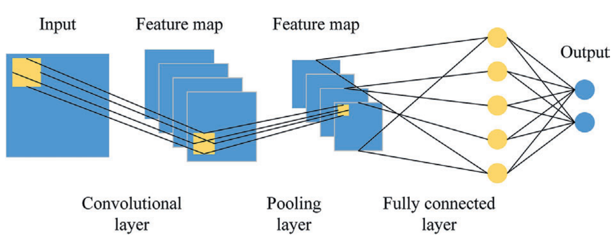

Fig. 1. Structure of CNN.

2.2. Bi-directional gate recurrent unit (BiGRU)

The gated recurrent unit (GRU) streamlines the conventional long short-term memory network (LSTM) by consolidating the forget, input, and output gates into a more straightforward configuration featuring just the reset and update gates, as illustrated in Fig. 2.

The GRU effectively addresses the challenges of gradient vanishing and explosion, enhancing the model's ability to handle long-term dependencies (Niu, 2022). Fig. 2 shows the update door z_t, reset the door r_t. The reset gate determines how much information from the previous time step h_{t-1} is discarded and influences the candidate hidden state accordingly. The update gate regulates the amount of prior information h_{t-1} carried into the current state.

The formulas for the reset door and update door calculations are as follows:

$$\begin{cases} r_t = \sigma(x_t W_{xr} + h_{t-1} W_{hr} + b_r) \\ z_t = \sigma(x_t W_{xz} + h_{t-1} W_{hz} + b_z) \\ \tilde{h}_t = \tanh(x_t W_{xh} + (r_t \odot h_{t-1}) W_{hh} + b_h) \\ h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \end{cases} \tag{4}$$

where, W_{xr} and W_{hr} refers to the weight matrix for the reset gate; b_r denotes the bias for the reset gate; W_{xz} and W_{hz} signifies the weight matrix for the update gate; b_z indicates its corresponding bias; W_{xh} and W_{hh} represents the weight matrix associated with the hidden state; b_h refers to the bias of the hidden state.

In the GRU neural network architecture, the output of the neuron state relies solely on past information, excluding any future data. However, while the present condition may connect to data from both the past and future, the capacity to identify intricate patterns within the actual sequence remains restricted. To address this limitation (Niu, 2022), introduced a bidirectional gated recurrent unit (BIGRU). This model enhances the ability to capture intricate features within time series data by taking into account input information from both forward to backward and backward to forward during sequence processing, as illustrated in Fig. 3.

The BiGRU has two independent hidden layers. The initial layer of GRU forwards information from front to back in a chronological sequence, whereas the subsequent layer transmits information from back to front in reverse chronological order. The result is influenced by the conditions of the two GRUs. This means that the existing hidden layer state is influenced by the forward hidden

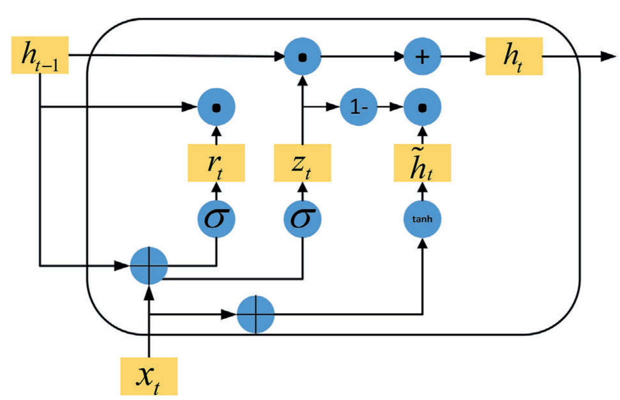

Fig. 2. Structure of GRU

layer state $\overrightarrow{h_{t-1}}$ and the backward hidden layer state $\overleftarrow{h_{t-1}}$. The formula for calculation is as follows:

$$\begin{cases} \overrightarrow{h}_t = G\left(x_t, \overrightarrow{h_{t-1}}\right) \\ \overleftarrow{h}_t = G\left(x_t, \leftarrow h_{t-1}\right) \\ h_t = w_t \overrightarrow{h_t} + v_t \leftarrow h_t + b_t \end{cases} \tag{5}$$

where, $G(\cdot)$ denotes the nonlinear transformation applied to the input data; w_t and v_t signifies the state of the forward hidden layer $\overrightarrow{h_t}$ and the reverse hidden layer $\overleftarrow{h_t}$, respectively; b_t represents the bias associated with the hidden state at a given moment t.

2.3. Attention mechanism

The attention mechanism is a resource allocation system that mimics the human brain's attention processes. By focusing on key information and ignoring irrelevant information, the efficiency of extracting important information is improved (Niu, 2022). In forecasting shale gas production, the attention mechanism creates feature codes that reflect the probability distribution of attention by assessing its probabilities. Ultimately, this process produces feature vectors that underscore the significance of various time steps within the input data. This is illustrated in Fig. 4.

In Fig. 4, x_t denotes the input for the BiGRU model, s_t signifies the output from the hidden layer of the BiGRU model, a_t indicates the weight associated with the force probability distribution, and Y refers to the output that has been refined by the attention mechanism. The attention layer receives as input the output vector S from the BiGRU activation layer, and the weight coefficient is calculated using the following formula:

$$\begin{cases} e_t = \tanh(w_t s_t + b_t) \\ \alpha_t = \exp(e_t) \Big/ \sum_{i=1}^{t} e_i \\ Y = \sum_{t=1}^{n} \alpha_t s_t \end{cases} \tag{6}$$

where, e_t denotes the energy value derived from the state vector s_t of the feature vector t; w_t signifies the weight coefficient matrix linked to the feature vector t; b_t represents the offset associated with the feature vector t; α_t indicates the weight of the features.

2.4. Particle swarm optimization (PSO) algorithm

The particle swarm optimization (PSO) algorithm mimics the social dynamics of bird flocks by simulating the actions of massless particles. Each particle is characterized by its velocity and location. In the search space, particles autonomously discover the optimal solution, identified as the current individual extremum (Kennedy' and Eberhart, 1995). This ideal solution is then communicated to other particles, allowing the best individual extremum in the group to evolve into the global optimal solution for the current group. Utilizing the present local extremum and the global optimum, the particle can autonomously modify its speed and location to consistently progress towards the optimal solution.

The velocity update formula for each particle i at time $t+1$:

$$v_i(t+1) = \omega \cdot v_i(t) + c_1 r_1 [p_i(t) - x_i(t)] + c_2 r_2 [g(t) - x_i(t)] \tag{7}$$

where, the inertia weight ω serves to harmonize both global and local search abilities. The factors c_1 and c_2 indicate the coefficients for individual and social cognition, respectively; r_1 and r_2 are random numbers drawn from a uniform distribution within the range of [0,1], adding an element of unpredictability. The term $p_i(t)$ refers to the individual best position of particle i at time t, while $g(t)$ signifies the best overall position of all particles at that same time. $x_i(t)$ represents the position of particle i at time t.

The formula for updating the position x_i of each particle i at time $t+1$:

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{8}$$

Nevertheless, the shift in position of the particle swarm optimization (PSO) algorithm during the optimization process carries certain constraints, making it susceptible to premature convergence or settling into a local optimum. This study enhances inertia weight updating and integrates adaptive mutation probability, boosting global exploration and optimization efficacy in the algorithm.

Inertia weight update formula:

$$w = w_{max} - \frac{w_{max} - w_{min}}{1 + e^{-t/t_{max}}} \tag{9}$$

where, w_{max} and w_{min} denote the maximum and minimum values of w respectively; t indicates the current iteration count; t_{max} represents the total allowed iterations. As t rises, w will steadily decline, with a greater weight in the initial phase and a smaller weight in the later phase.

Adaptive mutation probability formula:

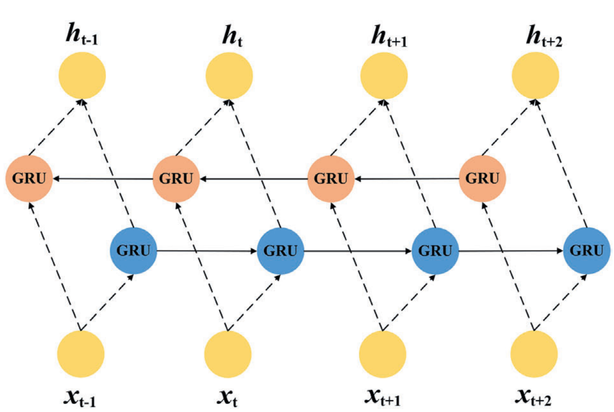

Fig. 3. Structure of BiGRU.

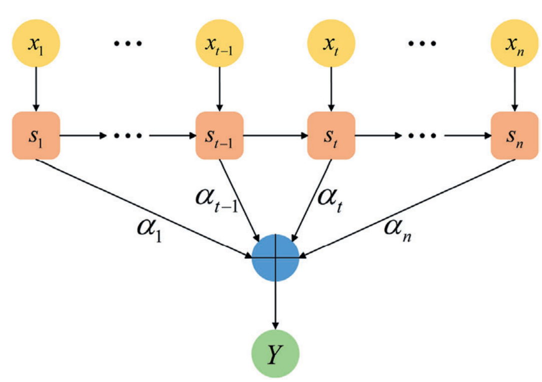

Fig. 4. Structure of attention mechanism.

$$rand > \frac{1}{2}\left(1+\frac{t}{t_{max}}\right) \tag{10}$$

The right side of the inequality represents the adaptive mutation function. When t is small, the adaptive mutation function is small, and the probability of particle mutation is high, which helps the particle to leave the local range and carry out a global search. As t increases, the adaptive mutation function increases, the probability of particle mutation decreases, and gradually turns to local search. In the initial phase, this process is marked by a broad, global search that, as the number of iterations rises, shifts to a more focused local search. This transition signifies a strategic optimization journey, evolving from extensive exploration to meticulous local refinement.

As an evaluation metric for the PSO algorithm, a smaller fitness function indicates that the particle is nearer to the global optimum, necessitating a more accurate local search. Conversely, a higher fitness value indicates a greater distance from the global optimal solution, necessitating an enhanced global search. The fitness function in this study is the weighted mean of the average relative error for both the training and test sets. The formula for calculation is as follows:

$$F=\frac{1}{2}\left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\widehat{y}_i}{y_i}\right| + \frac{1}{m}\sum_{j=1}^{m}\left|\frac{y_i-\widehat{y}_i}{y_i}\right|\right] \tag{11}$$

where, y_i denotes the actual value of the sample i; $\widehat{y}_i$ signifies the actual value of the sample i; n and m indicate the number of training and test data samples, respectively. Utilizing the weighted average approach allows the fitness function to thoroughly evaluate the model's efficacy on both the training and test datasets, steering the PSO algorithm toward an optimal balance between broader searching and more localized exploration.

2.5. IPSO–CNN-BiGRU-attention model

The IPSO–CNN-BiGRU-Attention model represents a sophisticated hybrid framework that combines particle swarm optimization (IPSO), convolutional neural networks (CNN), bidirectional gated recurrent units (BiGRU), and an attention mechanism, as illustrated in Fig. 5.

In this research, the model is utilized for analyzing time series data to forecast shale gas production. Firstly, a CNN extracts spatial characteristics from input shale gas production time series. The input data is fed through a series of one-dimensional convolutional layers. Each layer employs a variety of convolution kernels and window sizes, the aim being to snag the data's multi-scale features. Following pooling layer operation, the convolution layer's output is reduced to a fixed-size feature vector.

These feature vectors are then fed into the BiGRU for processing. The BiGRU can process sequence data bidirectionally, that is, from front to back and from back to front, so as to capture the bidirectional dependence of data and stitch the hidden states of the two directions. Ultimately, the concealed state following splicing is fed into the attention layer. The attention mechanism highlights key features and information according to context information by dynamically allocating weights, thereby enhancing the model's expression ability. Consequently, the IPSO–CNN-BiGRU-Attention model adeptly captures and leverages the multifaceted characteristics of time series data, resulting in a notable enhancement in both the performance and precision of shale gas production forecasts.

3. Development of the production prediction model

3.1. Data source and evaluation indicators

This study is based on actual production data of shale gas wells, with production data from 3 target wells selected to construct a production prediction model and these three wells adopt distinct production strategies and well completion methods. As of the statistical time point, the production cycles of the 3 wells commenced on January 9, 2015, December 3, 2015, and July 1, 2015, respectively, and all concluded on March 23, 2023, with a cumulative production duration exceeding 2400 days. Their production variation characteristics are illustrated in Fig. 6. In accordance with the production decline law of shale gas wells, the initial production stage is affected by the coupled effects of multiple factors, such as reservoir heterogeneity, differences in fracturing effectiveness, and bottom-hole flowing pressure fluctuations. This leads to significant production fluctuations, making it difficult to identify a clear decline law. With the progression of production, the output gradually stabilizes, the decline law becomes increasingly prominent, and it presents a gentle decline curve pattern. Based on this observation, this study divides the production data into two phases: the unstable production phase and the stable

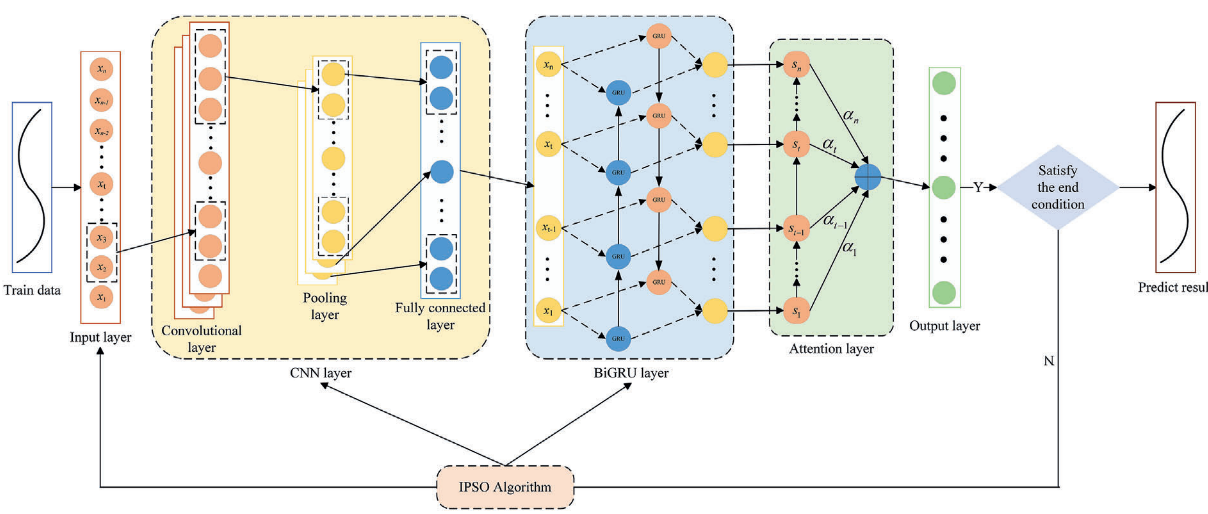

Fig. 5. Structure of IPSO–CNN-BiGRU-Attention.

production phase, and constructs corresponding prediction models for each phase respectively.

Affected by differences in inter-well reservoir properties, well completion process parameters, and production system, there are significant variations in the critical transition points from the unstable production phase to the stable production phase among different shale gas wells. To address this issue, this study proposes a data-driven adaptive production cycle division method: First, Variational Mode Decomposition (VMD) is applied to the production data of shale gas wells to extract the core production trend and eliminate random interference factors; subsequently, the Pruned Exact Linear Time (PELT) algorithm is used to solve for the critical division points between the unstable phase and the stable phase; finally, the accurate two-phase division of production data is achieved. The specific division process for the 3 target wells is shown in Fig. 7.

As can be seen from Fig. 7, the critical division points between the unstable production phase and the stable production phase of the three shale gas wells (Well 1, Well 2, and Well 3) exhibit significant differences. Specifically, the division point of Well 1 falls on the 265th production day, that of Well 2 is on the 240th production day, and that of Well 3 is on the 365th production day. This result indicates that the proposed adaptive production cycle division algorithm can successfully achieve the phase division of the production cycles for the three wells, and is capable of accurately completing the demarcation between the unstable production phase and the stable production phase based on the exclusive critical division point of each individual well.

To assess the precision and efficacy of segmented modeling, we chose root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) as our key performance metrics. The relevant calculation formulas are as follows:

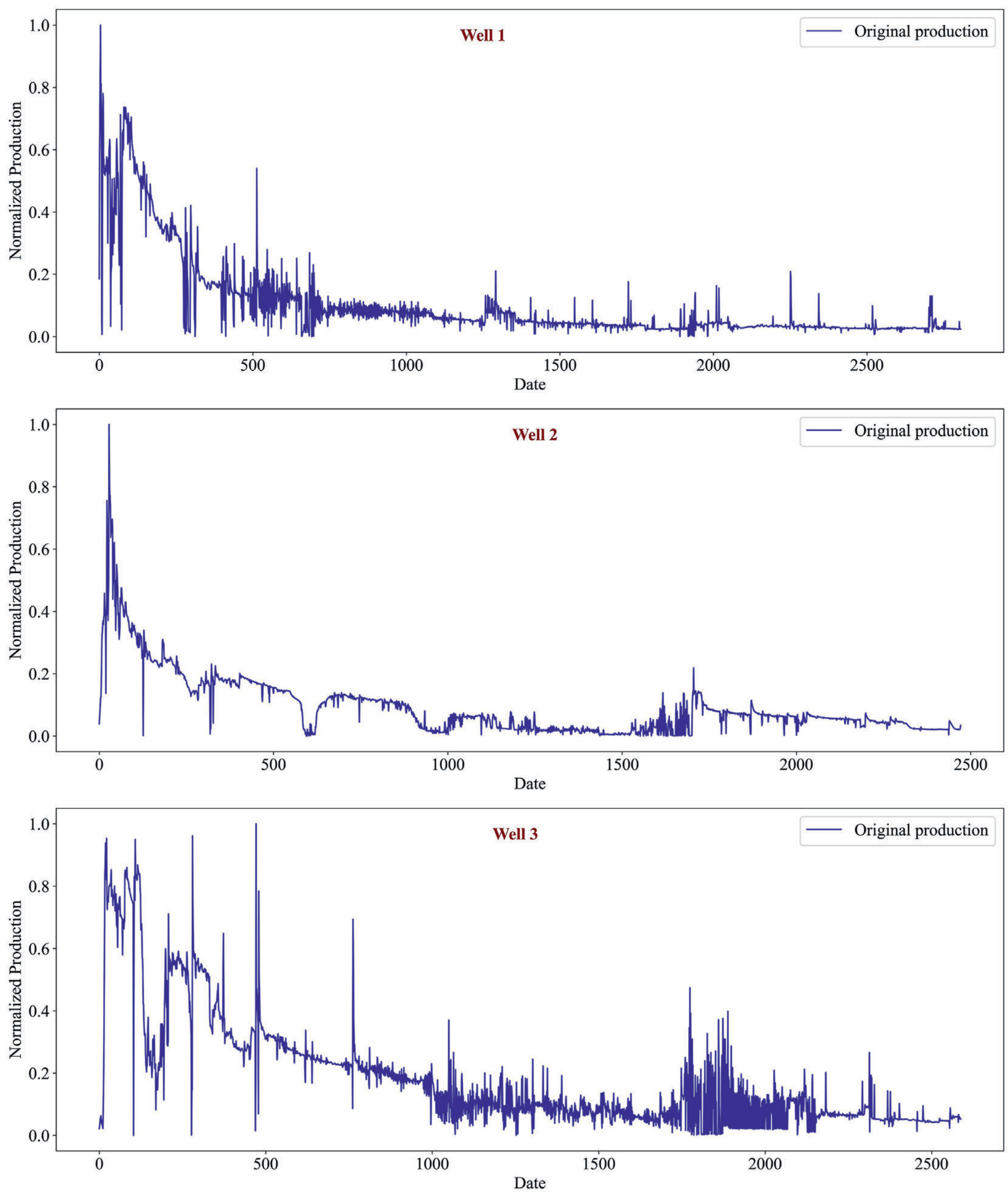

Fig. 6. Overview of shale gas production from three wells.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{1}{n} [y_i - \hat{y}_i]^2} \tag{12}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{13}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{14}$$

where, y_i signifies the actual value; $\hat{y}_i$ denotes the predicted value; n refers to the total number of data samples. The authors have chosen the mean square error (MSE) as the loss function, with its calculation expressed by the following formula (Wang et al., 2021c):

$$MSE = \sum_{i=1}^{n} \frac{1}{n} [y_i - \hat{y}_i]^2 \tag{15}$$

The segmented model is compared with the full-cycle model. By uniformly modeling the data of the entire production cycle, the model may ignore the differences in the characteristics of each stage. In contrast, the segmented model can capture subtle changes in different stages. By assessing the predictive accuracy and practicality of both models, we can effectively measure the strengths and weaknesses of the segmented modeling approach in analyzing production data for shale gas wells.

There are several common problems in shale gas well data, including missing values and anomalous values. If untreated data is directly used, it may lead to a significant decline in model performance. Thus, data preprocessing is a crucial phase, encompassing data cleaning and normalization. For data cleaning, a hybrid method combining the three-sigma rule and production engineering sense is adopted to handle anomalous values: first, the mean (μ) and standard deviation (σ) of daily production data for each well are calculated, and values outside the range [μ - 3σ, μ + 3σ] are marked as initial anomalies; then, false anomalies caused by legitimate engineering events. For missing values, a 5-day sliding window mean-filling method is applied: the missing value is filled with the average of 2 valid values before and 2 after the missing point; if the missing point is at the start or end of the time series, 5 consecutive valid values from the nearest side are used for averaging. For data normalization, considering the significant differences in production ranges among the three wells, normalization is performed individually for each well to avoid data distortion caused by cross-well range differences, the calculation formula is shown in formula (16). After normalization, the data range of all wells is strictly limited to [0,1], this ensures no significant bias in the normalized data, which could otherwise affect the stability of model training.

The formula for data normalization is outlined below (Wang et al., 2021a):

$$X_{scaled} = \frac{\left(X - X_{min(axis=0)}\right)}{\left(X_{max(axis=0)} - X_{min(axis=0)}\right)} \cdot (max - min) + min \tag{16}$$

where, $axis = 0$ indicates the equivalent characteristics of the time series; X_{max} and X_{min} signify the utmost maximum and minimum values, respectively; X denotes the original input time series; X_{scaled} stands for the normalized value.

3.2. Sliding window and hyperparameter configuration

In time series prediction, the setting of the sliding window size is very significant in the model's data set division—it directly determines the range of historical production information the model can use to predict future output, and its selection must align with the temporal characteristics of shale gas production data, production cycle features, and the model's dependency-capturing capabilities. This study sets the sliding window range to [1,10] days: pre-experiments confirm windows of 11–15 days increase training time by 30%–50% and raise MAPE by 1.2%–2.5%. This range also adapts to production cycle characteristics: small windows (1–3 days) capture high-frequency fluctuations in the unstable period but miss underlying weak trends, while medium-to-large windows (7–10 days) smooth short-term noise in the stable period to highlight long-term decline trends without obscuring inflection points. For model dependency capture, small windows make RNN/LSTM/GRU struggle to grasp long-term correlations, while a 10-day window balances the correlation of historical information and model adaptability. Generally, a smaller sliding window results in reduced prediction smoothness and a greater emphasis on short-term forecasting, while a larger one focuses more on long-term prediction. Therefore, this study uses the control variable method to determine the optimal sliding window size: for RNN, LSTM, and GRU architectures, the learning rate is set at 0.001, with a single network layer and 40 neurons; the sliding window range is established at [1,10], and the training process is

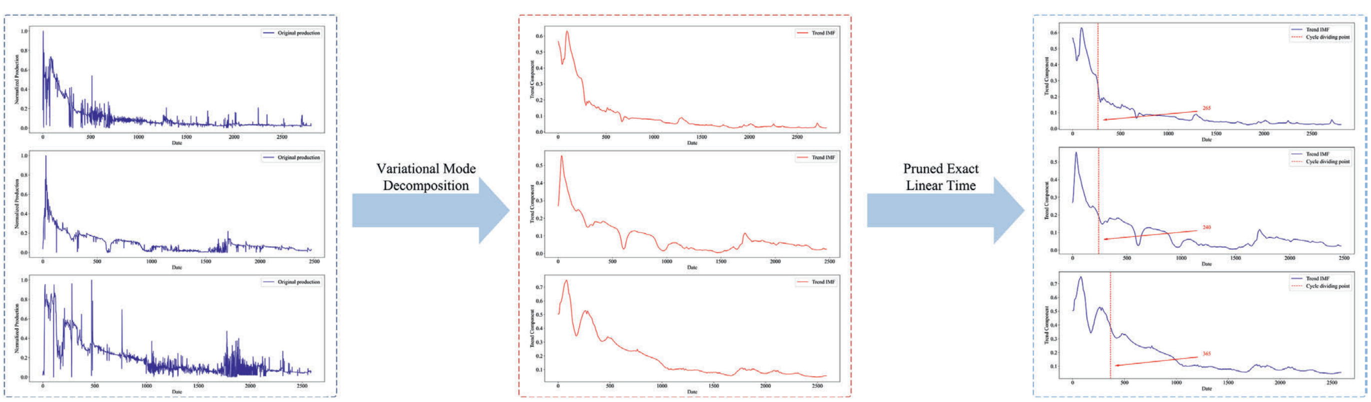

Fig. 7. Data-driven adaptive partitioning algorithm.

performed sequentially, with the MAPE trend illustrated in Fig. 8.

Fig. 8 illustrates that for RNN, LSTM, or GRU models, a sliding window size of 10 achieves the minimum MAPE, signifying the model's error is at its lowest during this condition. Thus, this research establishes the sliding window size at 10 and the time step interval at 1. In other words, we begin by leveraging the historical data from the last 10 days to forecast the results for the 11th day. Next, we utilize the data from the preceding 9 days along with the previously predicted value to estimate the output for the 12th day, and we continue this process onward.

The complete production cycle generates yield data, which is segmented into data sets based on a sliding window size of 10. These sets are further split into 80% for the training dataset and 20% for the testing dataset. In the CNN-BiGRU-Attention model, its model architecture and training process are jointly characterized by multiple sets of key hyperparameters, which specifically include the number of convolutional layers, the number of convolutional kernels, convolutional kernel size, the number of BiGRU network stacking layers, the number of neurons in the BiGRU layer, and the initial learning rate. To obtain the optimal hyperparameter combination, the Improved Particle Swarm Optimization (IPSO) algorithm is adopted for hyperparameter optimization and fine-tuning of the model, with the number of iterations set to 1000; the specific value ranges of each hyperparameter are detailed in Table 1.

3.3. Modeling of unstable production period

For unstable-phase production time-series data, it is split into a training set and a test set in a ratio of 80% for the training set and 20% for the test set. This ratio design ensures both the sample sufficiency during model training and the reliability of generalization ability evaluation results. In the hyperparameter optimization phase, to balance the comprehensiveness of the model's global optimization and the effectiveness of the optimal parameter combination, the number of iterations of the Improved Particle Swarm Optimization (IPSO) algorithm is set to 1000, The final optimal hyperparameter combination is shown in Table 2. Through global efficient optimization, the optimal hyperparameter combination suitable for the target model is obtained. Based on the model with optimized hyperparameters, quantitative evaluation is performed on its prediction performance, yielding key evaluation metrics, the specific evaluation results are shown in Table 3.

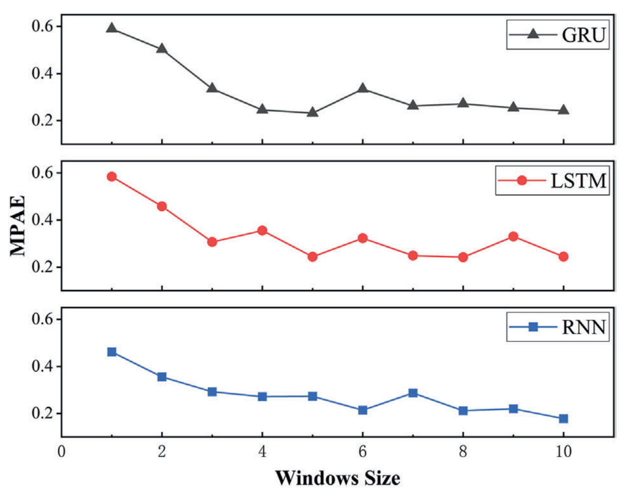

Fig. 8. Sliding window change trend.

Table 3 systematically compares the performance of various prediction models for the unstable production period of three target shale gas wells (Well 1: 265 days, Well 2: 240 days, Well 3: 365 days). The core feature of production data during this phase is high-frequency and intense fluctuations, which are mainly regulated by reservoir heterogeneity, fracturing fluid flowback efficiency, and sudden changes in bottom-hole flowing pressure. Consequently, the production exhibits characteristics including large fluctuation amplitude, multiple peaks, and no obvious regularity. CNN, relying on the local sliding feature extraction mechanism of 1D convolution kernels, can efficiently capture short-term dynamic features such as production peaks during the fracturing flowback period and production recovery fluctuations after temporary well shutdowns. It thus performs optimally among all single models in the tests on the three target wells. However, the global weight sharing mechanism of CNN makes it difficult to effectively distinguish the contribution differences between key fluctuation nodes and random noise. As a result, its prediction performance significantly degrades during the unstable production period of Well 3 (with greater fluctuation intensity), and the MAPE rises to 9.10%. RNN suffers from the defect of gradient vanishing, which prevents it from effectively transmitting fluctuation correlation information in long-time series during the unstable period. Its MAE reaches as high as 7296 m^3/d in the prediction for Well 3. Although LSTM and GRU have been improved based on RNN through gating mechanisms, they still

Table 1
Optimization range of hyperparameters.

Hyperparameters	Min	Max
Convolutional Kernel Size	3	5
Number of Convolutional Layers	2	4
Number of Convolutional Filters	16	64
Dropout Rate	0.1	0.3
Number of Hidden Layer Neurons	16	64
Number of BiGRU Layers	1	10
Attention Mechanism Unit Rate	16	64
Learning Rate	0.0001	0.005

Table 2
Hyperparameters during the production unstable period.

Well Number	Hyperparameters	value
Well 1	Convolutional Kernel Size	3
	Number of Convolutional Layers	2
	Number of Convolutional Filters	52
	Dropout Rate	0.2
	Number of Hidden Layer Neurons	35
	Number of BiGRU Layers	3
	Attention Mechanism Unit Rate	38
	Learning Rate	0.0005
Well 2	Convolutional Kernel Size	5
	Number of Convolutional Layers	3
	Number of Convolutional Filters	57
	Dropout Rate	0.2
	Number of Hidden Layer Neurons	28
	Number of BiGRU Layers	1
	Attention Mechanism Unit Rate	42
	Learning Rate	0.0001
Well 3	Convolutional Kernel Size	5
	Number of Convolutional Layers	4
	Number of Convolutional Filters	26
	Dropout Rate	0.2
	Number of Hidden Layer Neurons	23
	Number of BiGRU Layers	6
	Attention Mechanism Unit Rate	58
	Learning Rate	0.0002

struggle to achieve an effective balance between short-term fluctuation capture and long-term trend memory, leading to suboptimal prediction accuracy for production during the unstable period. Accordingly, CNN demonstrates significant advantages in capturing short-term production peaks, while GRU performs better in characterizing long-term production trend changes.

The CNN-BiGRU-Attention hybrid model achieves accurate adaptation to the laws of production data during the unstable period through the synergistic effect of three mechanisms: it first extracts local fluctuation features of production data via CNN; the bidirectional temporal encoding mechanism of BiGRU can simultaneously transmit forward short-term fluctuation information and backward correlation information, effectively overcoming the memory limitations of single temporal models; the Attention mechanism dynamically assigns feature weights, enabling the model to focus on key information features related to target production prediction while reducing the weight proportion of random noise and avoiding interference from invalid information. The IPSO algorithm conducts hyperparameter optimization for inter-well heterogeneity, dynamically adjusting parameters according to the production characteristics of each well to ensure the model adapts to the production fluctuation intensity of different wells. Test results show that the IPSO–CNN-BiGRU-Attention model performs optimally in the prediction for all three target wells, indicating that this model has good adaptability and prediction performance for complex scenarios with intense local fluctuations and significant inter-well heterogeneity during the unstable production period.

Fig. 9 presents the comparison results between production prediction values and actual values of five models during the unstable production period of 3 wells. Among them, the locally enlarged view of the test set clearly shows that the IPSO–CNN-BiGRU-Attention hybrid model has the highest degree of fit between its prediction curve and the variation trend of actual production, with significantly reduced prediction deviation in these critical intervals. By contrast, single-structure models such as RNN and LSTM exhibit more prominent prediction deviations in the aforementioned fluctuation intervals. Fig. 10 show that the IPSO–CNN-BiGRU-Attention model has the narrowest width of the error band, and most data points are concentrated in the low-error interval of 3%–4%; the RNN model has the widest width of the error band, with the error of some data points (e.g., RNN prediction results for Well 3) exceeding 15%; the error level of the CNN model lies between that of the RNN and IPSO–CNN-BiGRU-Attention models. The results of the two figures above mutually confirm that during the unstable production period, the IPSO–CNN-BiGRU-Attention hybrid model exhibits significant advantages in both production prediction accuracy and error control capability.

3.4. Modeling of the stable production period

For the time-series data of the stable-phase production process, a division strategy of 80% for the training set and 20% for the test set is adopted to split the dataset into training and test subsets. During the hyperparameter optimization phase, to balance the comprehensiveness of the model's global search for stable-phase-adapted parameters and the effectiveness of the optimal parameter combination, the number of iteration rounds of the IPSO algorithm is uniformly set to 1000, The final optimal hyperparameter combination is shown in Table 4. Based on the model with optimized hyperparameters, key evaluation metrics are derived; the specific evaluation results are presented in Table 5.

As shown in Table 5, a comparative analysis of the performance of various production prediction models for 3 target shale gas wells during the stable production period has been completed. The core characteristics of production data during this period are gentle decline and low noise, mainly controlled by the slow consumption of reservoir energy and the stability of seepage laws, ultimately presenting phenomena of production decline, large fluctuation amplitude, and occasional decline rate inflection points. The prediction requirement shifts from short-term fluctuation capture to long-term time-dependent modeling and trend correction, leading to significant differences in the adaptability of different models to stable-period data. CNN relies on the local sliding feature extraction mechanism of convolution kernels, lacks temporal memory capability, and can only extract local segment features of production data. It is difficult to correlate long-term production decline trends, resulting in poor performance in the tests of all 3 target wells. RNN still has the defect of gradient vanishing and cannot effectively transmit long-term decline trend information in long sequences during the stable period; it achieves relatively better prediction performance in Well 2, which is attributed to the simple production decline law of Well 2, and its overall adaptability to stable-period scenarios remains limited. LSTM and GRU are improved on the basis of RNN through gating mechanisms, enabling more accurate characterization of long-term decline trends and achieving good prediction performance in Well 1. Their unidirectional temporal encoding mechanism cannot use subsequent gentle decline features to correct current prediction results. When encountering decline rate inflection points, they cannot distinguish whether temporary production fluctuations are accidental events or trend-driven, leading to decreased prediction performance. For example, the MAPE of LSTM in Well 2 rises to 16.56%.

The CNN-BiGRU-Attention model achieves accurate matching with the laws of stable-period production data through the synergy of three mechanisms: CNN extracts local features of

Table 3
Evaluation indicators of models in unstable production period.

Well Number	Model	RMSE(10^4 m^3/d)	MAE(m^3/d)	MAPE(%)
Well 1	CNN	0.2974	2089	4.08
	RNN	0.3721	3077	6.38
	LSTM	0.4550	3970	8.28
	GRU	0.3408	2816	5.80
	IPSO–CNN-BiGRU-Attention	**0.2470**	**1632**	**3.21**
Well 2	CNN	0.5759	4961	6.43
	RNN	0.5952	5037	6.68
	LSTM	0.6151	5348	6.95
	GRU	0.6471	5661	7.25
	IPSO–CNN-BiGRU-Attention	**0.3960**	**2707**	**3.51**
Well 3	CNN	0.5131	3970	9.10
	RNN	0.7893	7296	15.97
	LSTM	0.6913	6298	13.43
	GRU	0.6548	5940	12.85
	IPSO–CNN-BiGRU-Attention	**0.2960**	**1753**	**3.99**

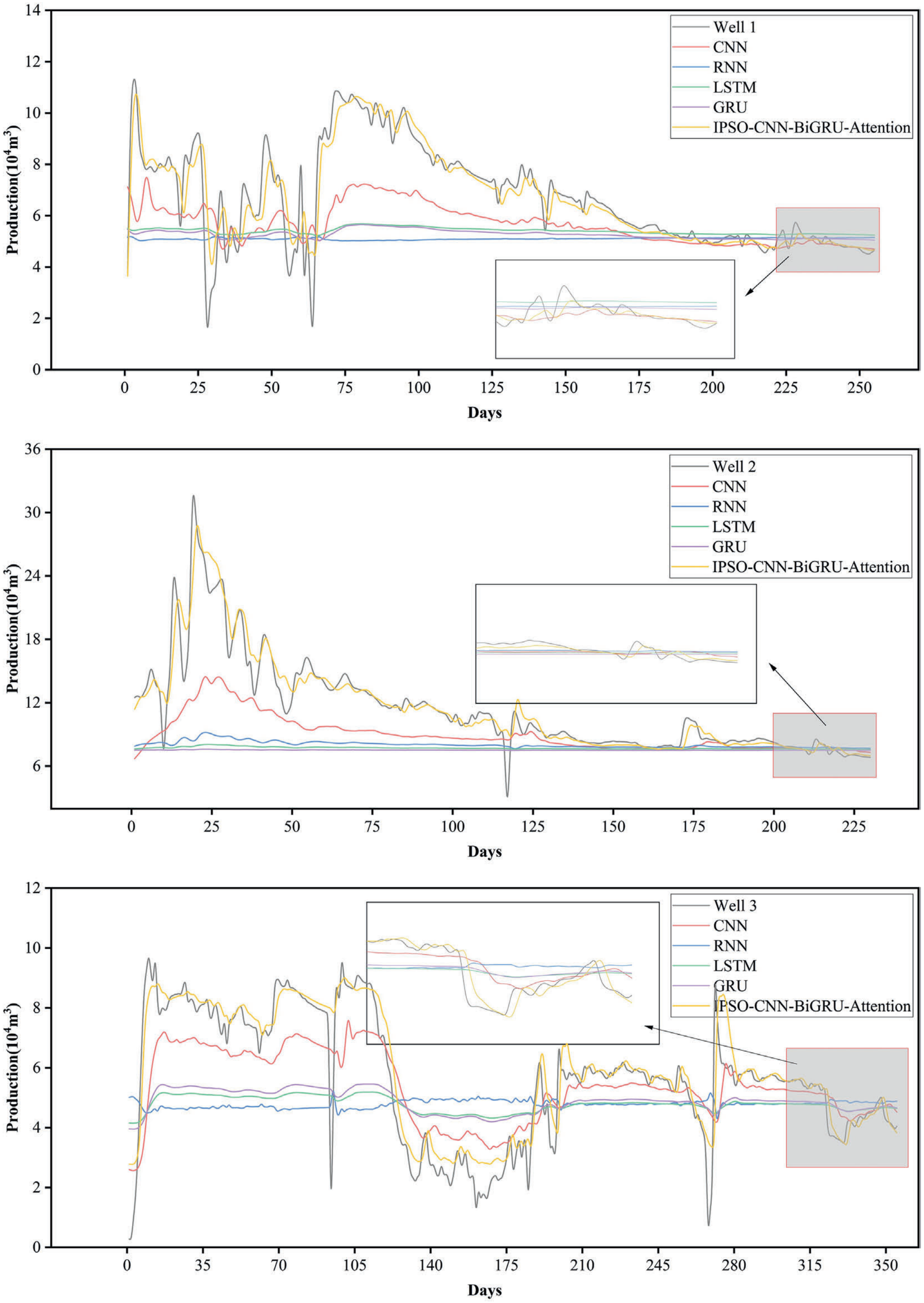

Fig. 9. Comparison of the predicted results of the unstable production model.

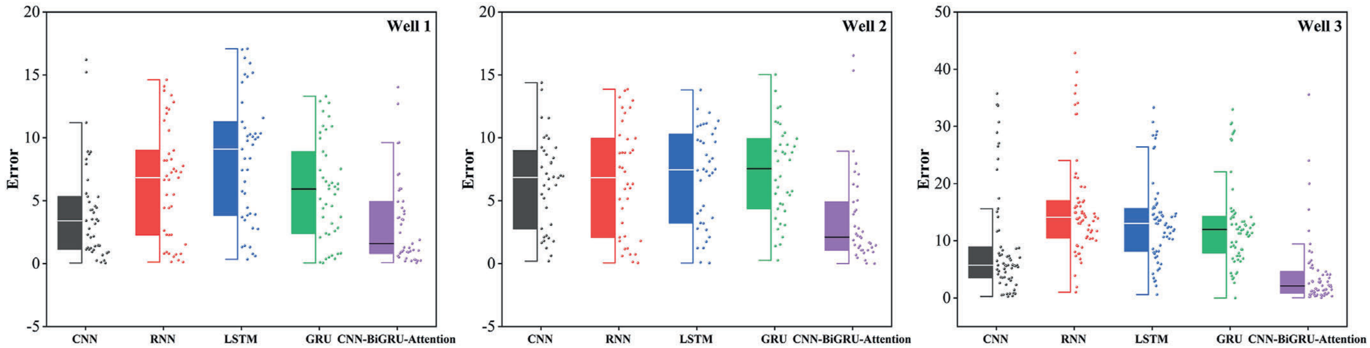

Fig. 10. The relative error of each model in the unstable production model.

production data and filters residual noise, laying a foundation for subsequent long-term trend modeling; the bidirectional temporal encoding mechanism of BiGRU synchronously transmits forward historical decline trend information and backward future gentle change information, significantly improving the accuracy of long-term trend representation; the Attention mechanism focuses on key information such as decline rate inflection points, while reducing the weight proportion of residual noise, ensuring the model can effectively capture core trend features. The IPSO algorithm conducts hyperparameter optimization based on the stable-period production characteristics of each well. Test results show that the IPSO–CNN-BiGRU-Attention model achieves the best performance in the prediction of all 3 target wells. The results indicate that this model has good adaptability and prediction performance in stable-period scenarios with long-term gentle production decline, occasional trend turning points, residual noise, and significant differences in inter-well decline characteristics.

Fig. 11 presents the comparison between predicted values and actual values of five models during the stable production period of 3 wells. Among them, the enlarged test set segment shows that the prediction curve of the IPSO–CNN-BiGRU-Attention model has the highest degree of fit with the gentle declining trend of actual production. However, due to the lack of temporal memory capability, CNN exhibits relatively more obvious prediction deviations in the long-term declining interval; similarly, LSTM also has such deviations because of its one-way encoding limitation. As shown in Fig. 12, the IPSO–CNN-BiGRU-Attention model has the narrowest error band, with most data points concentrated in the low error range of 5%–14%. For Well 3, the CNN model cannot distinguish between real production decline and residual noise, resulting in the widest error band, and the MAPE of some data points exceeds 30%. The error performance of the RNN, LSTM, and GRU models falls between the above two extremes.

Table 4
Hyperparameters during the production stable period.

Well Number	Hyperparameters	value
Well 1	Convolutional Kernel Size	5
	Number of Convolutional Layers	4
	Number of Convolutional Filters	26
	Dropout Rate	0.2
	Number of Hidden Layer Neurons	36
	Number of BiGRU Layers	4
	Attention Mechanism Unit Rate	40
	Learning Rate	0.0011
Well 2	Convolutional Kernel Size	5
	Number of Convolutional Layers	3
	Number of Convolutional Filters	49
	Dropout Rate	0.2
	Number of Hidden Layer Neurons	26
	Number of BiGRU Layers	8
	Attention Mechanism Unit Rate	18
	Learning Rate	0.0001
Well 3	Convolutional Kernel Size	3
	Number of Convolutional Layers	2
	Number of Convolutional Filters	57
	Dropout Rate	0.3
	Number of Hidden Layer Neurons	18
	Number of BiGRU Layers	3
	Attention Mechanism Unit Rate	57
	Learning Rate	0.0018

4. Discussion

4.1. Performance comparison of segmented and full-cycle modeling

Advancements in deep learning and computational power have enabled the training of more sophisticated models. The IPSO–CNN-BiGRU-Attention model introduced in this study has achieved notable advancements in predicting shale gas production, considering temporal and spatial factors. This model successfully overcomes the limitation of complex features in the actual shale gas well production data set. The findings indicate that the IPSO–CNN-BiGRU-Attention model substantially enhances shale gas production prediction performance over the traditional time series model. After simulating the boom-and-bust, as well as the steady-state phases of shale gas well production, the findings suggest that the IPSO–CNN-BiGRU-Attention model really shines when it comes to handling the unpredictable nature of shale gas output. The model excels in accurately predicting shale gas production by seamlessly combining the effective parameter optimization of the IPSO algorithm, the feature extraction capabilities of CNN, the time series modeling strength of BiGRU, and the feature weighting provided by the Attention mechanism, resulting in performance that markedly surpasses that of conventional models.

Table 6 reveals the differences in prediction performance between the segmented modeling and full-cycle modeling of the IPSO–CNN-BiGRU-Attention model for three target shale gas wells during the unstable production period and stable production period. It clearly illustrates the adaptability differences of the two modeling methods to the data characteristics of different production stages, as well as the core advantages of segmented modeling. During the unstable production period, under the coupled influence of factors such as reservoir heterogeneity and fracturing fluid flowback efficiency, the production data exhibit characteristics of high-frequency and large-amplitude fluctuations. As shown in Table 6, when segmented modeling is adopted for Well 1 in this stage, its RMSE is 0.2470×10^4 m^3/d, MAE is 1632 m^3/d, and MAPE is 3.21%. In contrast, when full-cycle modeling is used, the aforementioned indicators increase to 0.8257×10^4 m^3/d, 7736 m^3/d, and 15.46%, respectively, with the

Table 5
Evaluation indicators of models in stable production period.

Well Number	Model	RMSE(10^4 m^3/d)	MAE(m^3/d)	MAPE(%)
Well 1	CNN	0.1791	697	14.24
	RNN	0.1630	741	16.20
	LSTM	0.1758	686	13.95
	GRU	0.1644	626	13.37
	IPSO–CNN-BiGRU-Attention	**0.1620**	**625**	**13.14**
Well 2	CNN	0.2116	1350	10.07
	RNN	0.1635	975	9.76
	LSTM	0.2317	1660	16.56
	GRU	0.1984	1178	10.76
	IPSO–CNN-BiGRU-Attention	**0.1492**	**649**	**5.57**
Well 3	CNN	0.2369	1711	30.93
	RNN	0.2030	970	16.27
	LSTM	0.2121	1358	24.21
	GRU	0.1997	991	16.92
	IPSO–CNN-BiGRU-Attention	**0.1988**	**922**	**14.93**

reduction rates of RMSE and MAPE reaching 69.9% and 79.2%, respectively. The core reason for this difference is that segmented modeling can effectively isolate the high-frequency fluctuation characteristics of the unstable period, avoiding the "data interference effect" caused by the mixing of stable-period data in full-cycle modeling. For Well 2, since the production fluctuation law in its unstable period is relatively gentle, the advantage margin of segmented modeling is slightly reduced, but significant performance improvement is still achieved: RMSE decreases from 0.4642×10^4 m^3/d to 0.3960×10^4 m^3/d, MAE decreases from 3573 m^3/d to 2707 m^3/d, and MAPE decreases from 4.61% to 3.51%. As the well with the longest unstable production period and the most severe production fluctuations among the three wells, Well 3 exhibits the most prominent performance improvement with segmented modeling, with the reduction rates of RMSE, MAE, and MAPE reaching 55.2%, 70.6%, and 68.2%, respectively. This result further verifies the law that "the stronger the data heterogeneity in the unstable production period, the more significant the performance advantage of segmented modeling". However, full-cycle modeling generally suffers from overfitting issues due to its inability to effectively distinguish between key fluctuation nodes and random noise.

During the stable production period, the production data show characteristics of gentle decline and low noise, with occasional inflection points in the decline rate. The data in Table 6 indicate that segmented modeling still outperforms full-cycle modeling in this stage, and its core advantage lies in the fact that segmented modeling can focus on learning the long-term decline trend of the stable period, effectively avoiding the interference of noise from the unstable production period. Specifically, when segmented modeling is applied to Well 1, the RMSE is 0.1620×10^4 m^3/d, MAE is 625 m^3/d, and MAPE is 13.14%; the corresponding indicators of full-cycle modeling are 0.1842×10^4 m^3/d, 898 m^3/d, and 19.81%, respectively, with a MAPE reduction rate of 34%. The core reason for this is that full-cycle modeling is affected by the residual effect of "fluctuation memory" from the unstable production period, leading to deviations in the prediction of long-term trends. Due to the obvious inflection point in the decline rate of Well 2 during the stable period, the targeted advantage of segmented modeling is more prominent: RMSE decreases from 0.1899×10^4 m^3/d to 0.1492×10^4 m^3/d, MAE decreases from 1123 m^3/d to 649 m^3/d, and MAPE decreases from 8.68% to 5.57%, which effectively solves the technical bottleneck of full-cycle modeling in accurately identifying the inflection point of the decline rate due to the mixing of multi-stage data. After adopting segmented modeling for Well 3, the MAPE decreases from 26.87% to 14.93%, further confirming that segmented modeling has better noise filtering capability; in contrast, full-cycle modeling tends to misjudge noise fluctuations as real decline trends, resulting in a decrease in prediction accuracy.

Based on the comprehensive analysis above, segmented modeling demonstrates superior prediction performance compared to full-cycle modeling in both the unstable and stable production periods of shale gas wells, and its advantage is more significant in the unstable production period. The core mechanism lies in the fact that segmented modeling can accurately match the unique data characteristics of each production stage, while full-cycle modeling treats the entire production cycle as a homogeneous data sequence, which is prone to causing systematic prediction errors. In addition, the IPSO–CNN-BiGRU-Attention model achieves the lowest prediction errors for the three wells in both production stages under the segmented modeling framework. This result further confirms that the collaborative application of "data-driven production stage division" and "multi-mechanism hybrid model" is of great effectiveness in improving the prediction accuracy of shale gas production.

4.2. Computational efficiency analysis

To ensure the comparability of computational efficiency evaluations, all experiments were conducted under a unified hardware and software environment: the hardware included an Intel i9-13900K CPU, an NVIDIA RTX 4070, and 128 GB RAM; the software adopted Python 3.9 and PyTorch 2.7.1; training epochs were set to 1000 for all models. Additionally, all efficiency tests were conducted on the production data of Well 1 to ensure consistent data input conditions across models.

As shown in Fig. 13, whether in the stable period or unstable period, the training time of the IPSO–CNN-BiGRU-Attention hybrid model is significantly longer than that of traditional single models. The main reason is that the hybrid model integrates a multi-module architecture and introduces the IPSO algorithm for parameter optimization. This complex structural design and multi-module collaborative operation characteristics lead to a significant increase in the computational load during the training process. However, the inference time of this model is in the millisecond (ms) order of magnitude, which has little difference compared with most single models, and it has higher prediction accuracy. In engineering practice scenarios, accurate production prediction is of crucial significance for optimizing mining strategies and ensuring production stability. Therefore, prediction accuracy should take priority over mere computational efficiency considerations.

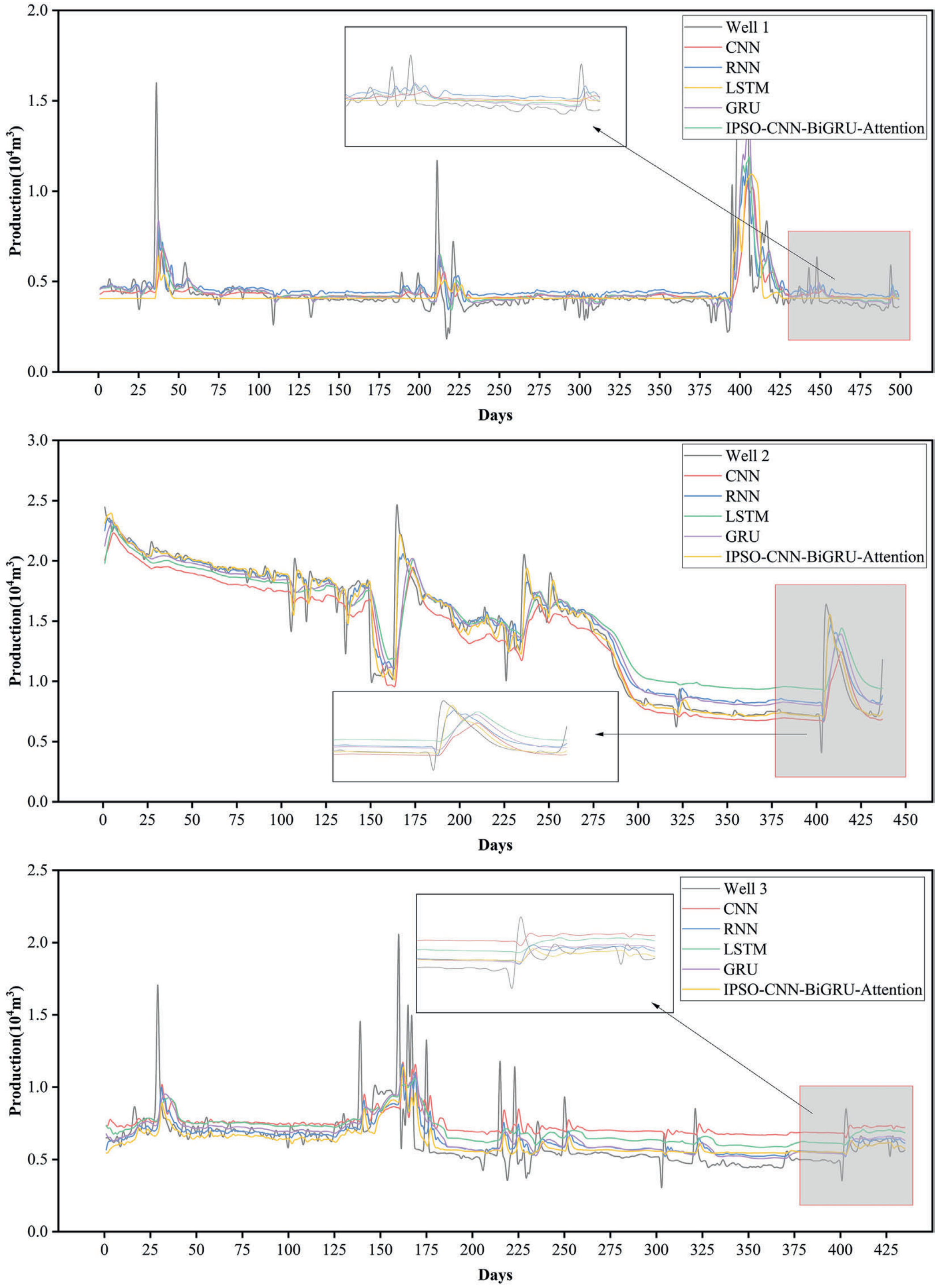

Fig. 11. Comparison of the predicted results of the stable production model.

4.3. *Uncertainty analysis of predictions*

Uncertainty quantification is critical to verifying the reliability of production prediction models, as it determines the credibility of outputs in guiding engineering decisions. This section uses the residual Bootstrap sampling method (1000 resamples) to calculate the average confidence interval (CI) width under 85%, 90%, and 95% confidence levels for all models across three wells and two production stages, with results summarized in Table 7.

As shown in Table 7, the confidence interval (CI) width of all models increases with the elevation of confidence levels (85% < 90% < 95%), and this regularity is consistent with the basic logic of statistical inference. In addition, the CI width in the unstable production period is significantly larger than that in the stable period, which is consistent with the inherent characteristics of shale gas production: high-frequency fluctuations in production during the unstable period aggravate prediction uncertainty, while the gentle decline in production during the stable period

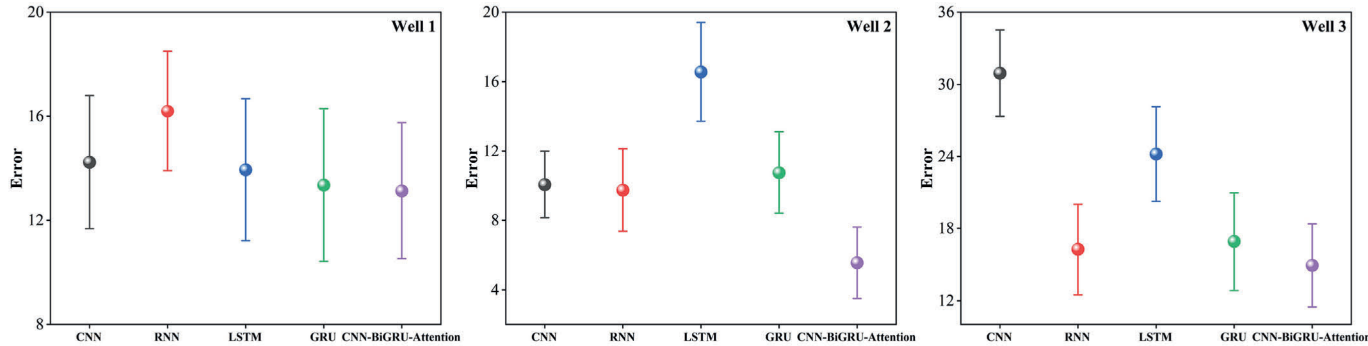

Fig. 12. The relative error of each model in the stable production model.

Table 6
Evaluation indicators of segmented and full-cycle modeling.

Well Number	Producation period	Segmented model			Full-cycle model		
		RMSE (10^4 m^3/d)	MAE (m^3/d)	MAPE	RMSE (104 m^3/d)	MAE (m^3/d)	MAPE
Well 1	Unstable period	0.2470	1632	3.21	0.8257	7736	15.46
	Stable period	0.1620	625	13.14	0.1842	898	19.81
Well 2	Unstable period	0.3960	2707	3.51	0.4642	3573	4.61
	Stable period	0.1492	649	5.57	0.1899	1123	8.68
Well 3	Unstable period	0.2960	1753	3.99	0.6613	5967	12.56
	Stable period	0.1988	922	14.93	0.2241	1489	26.87

effectively reduces this uncertainty. The IPSO–CNN-BiGRU-Attention model exhibits the narrowest CI width under all confidence levels across all well conditions and production stages. In the unstable production period, the advantages of this hybrid model are particularly prominent: taking Well 3 as an example, its 95% CI width is only 53.5% of that of the RNN model, 61.9% of the LSTM model, and 72.1% of the GRU model. Even for Well 1, its 95% CI width is still 10.8% narrower than that of the CNN model and 14.0% narrower than that of the RNN model. In the stable production period, the hybrid model maintains its advantages. Taking Well 2 as an example, its 95% CI width is 38.5% narrower than that of the CNN model and 15.0% narrower than that of the RNN model.

The performance advantage of this model stems from its multi-module synergy mechanism. In engineering practice scenarios, the hybrid model can still maintain a narrow CI width at high confidence levels, thereby effectively reducing decision-making risks. In summary, the IPSO–CNN-BiGRU-Attention model possesses both high prediction accuracy and low prediction uncertainty, which effectively addresses the dilemma of traditional single models struggling to balance prediction accuracy and reliability.

4.4. Application potential and implementation path

The segmented modeling framework integrated with VMD-PELT algorithm and the IPSO–CNN-BiGRU-Attention model proposed in this study demonstrates prominent application potential in shale gas engineering practice. In early-stage fracturing effectiveness evaluation, the model's accurate identification of unstable-stable period boundaries and low prediction error enables precise assessment of fracturing flowback peaks, guiding proppant and injection parameter optimization. For real-time production management, its millisecond-level inference time and narrow confidence intervals allow integration into field monitoring systems, supporting timely diagnosis of production deviations. Additionally, the IPSO algorithm's adaptive hyperparameter optimization makes the model scalable across wells with different reservoir properties.

The engineering implementation can follow a streamlined path: first, standardize data preprocessing by collecting production/pressure data, eliminating anomalies via the three-sigma rule and engineering judgment, filling missing values with 5-day sliding window averaging, and segmenting cycles using VMD-PELT; second, train the model with 80% of segmented data, optimize hyperparameters via 1000-iteration IPSO, and validate with 20% of data; finally, deploy the model as lightweight API services for field use, retaining multi-confidence CI outputs, and update monthly with new data to adapt to reservoir dynamic changes. This path balances modeling accuracy and engineering operability, providing practical support for intelligent shale gas production management.

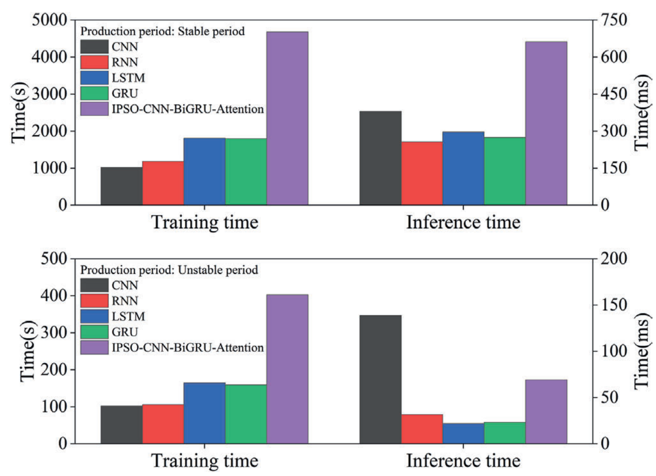

Fig. 13. Comparison of computational efficiency indicators of different models.

Table 7
Comparison of CI widths across multi-confidence levels and stages.

Well number	Producation period	Model	85% CI avg width (10^4 m^3/d)	90% CI avg width (10^4 m^3/d)	95% CI avg width (10^4 m^3/d)
Well 1	Unstable period	CNN	0.7560	0.9259	1.1733
		RNN	0.9053	1.0978	1.2142
		LSTM	0.8182	1.0722	1.1862
		GRU	0.8176	1.0509	1.1966
		IPSO–CNN-BiGRU-Attention	**0.7322**	**0.8288**	**1.0454**
	Stable period	CNN	0.1639	0.2776	0.5555
		RNN	0.1490	0.2415	0.4814
		LSTM	0.1554	0.2789	0.5856
		GRU	0.1657	0.2326	0.4879
		IPSO–CNN-BiGRU-Attention	**0.1476**	**0.2201**	**0.4525**
Well 2	Unstable period	CNN	1.5375	1.7171	2.0611
		RNN	1.6513	1.7450	1.9264
		LSTM	1.7028	1.7819	1.9537
		GRU	1.7141	1.8148	1.9626
		IPSO–CNN-BiGRU-Attention	**1.0450**	**1.1560**	**1.8401**
	Stable period	CNN	0.3410	0.4628	0.8044
		RNN	0.2469	0.3289	0.5212
		LSTM	0.4550	0.5799	0.8719
		GRU	0.3602	0.4729	0.8259
		IPSO–CNN-BiGRU-Attention	**0.2041**	**0.2757**	**0.4944**
Well 3	Unstable period	CNN	1.4939	1.6253	1.7390
		RNN	2.1571	2.2311	2.3453
		LSTM	1.9227	1.9628	2.0290
		GRU	1.8438	1.8783	1.9711
		IPSO–CNN-BiGRU-Attention	**0.5809**	**0.8576**	**1.2547**
	Stable period	CNN	0.2998	0.3975	0.6264
		RNN	0.2405	0.3183	0.6814
		LSTM	0.2495	0.3425	0.6073
		GRU	0.2397	0.3361	0.6697
		IPSO–CNN-BiGRU-Attention	**0.2300**	**0.3111**	**0.6020**

5. Conclusions

Against the backdrop of the global energy transition, accurate shale gas production prediction is critical for ensuring stable natural gas supply. However, the high-dimensional, nonlinear characteristics of shale gas production data and the significant differences in data features between different production stages have long posed challenges to traditional prediction methods. To address these issues, this study proposes a segmented modeling approach based on time series analysis and constructs an IPSO–CNN-BiGRU-Attention hybrid model, achieving targeted improvements in prediction accuracy and adaptability. The main conclusions of this research are as follows.

(1) The data-driven adaptive segmented modeling method effectively solves the problem of low prediction accuracy caused by the "homogenization" of full-cycle modeling. This study instead combines Variational Mode Decomposition (VMD) and the Pruned Exact Linear Time (PELT) algorithm to realize adaptive division of shale gas production cycles. For three target wells with different reservoir properties, fracturing effects, and production histories, the method accurately identified unique critical transition points: the 265thproduction day for Well 1, the 240th for Well 2, and the 365th for Well 3. By modeling the unstable and stable periods separately, this approach avoids the "data interference effect" of full-cycle modeling. Comparative results show that segmented modeling significantly reduces prediction errors: in the unstable period, the RMSE and MAPE of Well 1 are reduced by 69.9% and 79.2% respectively compared with full-cycle modeling; even for Well 2 with relatively gentle fluctuations, the MAPE is still reduced by 23.9%; for Well 3 with the most severe fluctuations, the reduction rates of RMSE, MAE, and MAPE reach 55.2%, 70.6%, and 68.2% respectively. In the stable period, segmented modeling also maintains advantages, with the MAPE of Well 2 reduced by 35.8%, effectively solving the problem of full-cycle modeling's inability to accurately identify inflection points.

(2) The IPSO–CNN-BiGRU-Attention hybrid model demonstrates superior performance in capturing the complex features of shale gas production data compared with traditional single models. The IPSO algorithm excels in parameter optimization, working hand in hand with the feature extraction capabilities of CNN, the time series modeling strength of BiGRU, and the feature weight allocation provided by the Attention mechanism. This collaboration allows each component to leverage its unique strengths effectively. Comparative experiments with traditional models (CNN, RNN, LSTM, GRU) show that the IPSO–CNN-BiGRU-Attention model achieves the lowest errors in both production periods: in the unstable period, its MAPE for the three wells is 3.21%, 3.51%, and 3.99% respectively; in the stable period, its MAPE for Well 2 is only 5.57%.

The research provides a feasible technical framework for shale gas production prediction, but there is room for further optimization. Future studies can be expanded in two directions: first, integrate more models with strong fitting capabilities to further enhance the model's ability to capture complex temporal and spatial features; second, refine the production cycle division and match each sub-period with a more targeted sub-model, maximizing the adaptation between the model and the dynamic changes of production data.

CRediT authorship contribution statement

Feifei Fang: Formal analysis, Data curation. **Hongjian Chen:** Methodology, Formal analysis. **Sijie He:** Writing – review &

editing, Software. **Jie Zhang:** Writing – review & editing. **Wei Guo:** Writing – review & editing. **Weixiang Jin:** Writing – review & editing, Writing – original draft. **Yue Gong:** Writing – review & editing. **ChuXiang Xia:** Writing – review & editing.

Declaration of competing interest

Here we confirm that there is no conflict of interest.

References

Alimohammadi, H., Rahmanifard, H., Chen, N., 2020. Multivariate time series modelling approach for production forecasting in unconventional resources. In: Proceedings of the SPE Annual Technical Conference and Exhibition. Virtual. https://doi.org/10.2118/201571-MS. SPE-201571-MS.

Cichos, F., Gustavsson, K., Mehlig, B., Volpe, G., 2020. Machine learning for active matter. Nat. Mach. Intell. 2 (2), 94–103. https://doi.org/10.1038/s42256-020-0146-9.

Fan, D., Sun, H., Yao, J., Zhang, K., Yan, X., Sun, Z., 2021. Well production forecasting based on arima-lstm model considering manual operations. Energy 220, 119708. https://doi.org/10.1016/j.energy.2020.119708.

Fan, L., Thompson, J.W., Robinson, J.R., 2010. Understanding gas production mechanism and effectiveness of well stimulation in the haynesville shale through reservoir simulation. In: Proceedings of the Canadian Unconventional Resources and International Petroleum Conference. https://doi.org/10.2118/136696-MS. Calgary, Alberta, Canada. SPE-136696-MS.

Feng, C., Zhang, J., Zhang, W., Hodge, B.-M., 2022. Convolutional neural networks for intra-hour solar forecasting based on sky image sequences. Appl. Energy 310, 118438. https://doi.org/10.1016/j.apenergy.2021.118438.

Guo, Z., Wang, H., Kong, X., Shen, L., Jia, Y., 2021. Machine learning-based production prediction model and its application in duvernay formation. Energies 14 (17), 5509. https://doi.org/10.3390/en14175509.

Hoover, K., Mudunuru, M., 2020. Machine Learning Models for real-time Forecasting of Shale Gas Production. Los Alamos National Laboratory (LANL) Report, Los Alamos, NM, USA. https://doi.org/10.13140/rg.2.2.36170.59843.

Han, D., Jung, J., Kwon, S., 2020. Comparative study on supervised learning models for productivity forecasting of shale reservoirs based on a data-driven approach. Appl. Sci. 10 (4), 1267. https://doi.org/10.3390/app10041267.

Huang, R., Wei, C., Wang, B., Yang, J., Xu, X., Wu, S., Huang, S., 2022. Well performance prediction based on long short-term memory (lstm) neural network. J. Pet. Sci. Eng. 208, 109686. https://doi.org/10.1016/j.petrol.2021.109686.

Hui, G., Chen, Z., Wang, Y., Zhang, D., Gu, F., 2023. An integrated machine learning-based approach to identifying controlling factors of unconventional shale productivity. Energy 266, 126512. https://doi.org/10.1016/j.energy.2022.126512.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948. https://doi.org/10.1109/icnn.1995.488968. Perth, Australia.

Kim, T.H., Cho, J., Lee, K.S., 2017. Evaluation of co_2 injection in shale gas reservoirs with multi-component transport and geomechanical effects. Appl. Energy 190, 1195–1206. https://doi.org/10.1016/j.apenergy.2017.01.047.

Kocoglu, Y., Wigwe, M.E., Sheldon, G., Watson, M.C., 2020. Machine learning based decline curve—spatial method to estimate production potential of proposed wells in unconventional shale gas reservoirs. In: Proceedings of the 8th Unconventional Resources Technology Conference. https://doi.org/10.15530/urtec-2020-3108. Online. URTeC-2020-3108.

Kudithipudi, D., Aguilar-Simon, M., Babb, J., et al., 2022. Biological underpinnings for lifelong learning machines. Nat. Mach. Intell. 4 (3), 196–210. https://doi.org/10.1038/s42256-022-00452-0.

Kumar, I., Tripathi, B.K., Singh, A., 2023. 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.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

Li, X.C., Ma, X., Xiao, F., Xiao, C., Wang, F., Zhang, S., 2022a. A physics-constrained long-term production prediction method for multiple fractured wells using deep learning. J. Pet. Sci. Eng. 217, 110844. https://doi.org/10.1016/j.petrol.2022.110844.

Li, X.L., Xiao, K., Li, X.B., Yu, C., Fan, D., Sun, Z., 2022b. A well rate prediction method based on lstm algorithm considering manual operations. J. Pet. Sci. Eng. 210, 110047. https://doi.org/10.1016/j.petrol.2021.110047.

Liang, B., Liu, J., You, J., Jia, J., Pan, Y., Jeong, H., 2023. Hydrocarbon production dynamics forecasting using machine learning: a state-of-the-art review. Fuel 337, 127067. https://doi.org/10.1016/j.fuel.2022.127067.

Liang, Y., Zhao, P., 2019. A machine learning analysis based on big data for eagle ford shale formation. In: Proceedings of the SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/196158-MS. Calgary, Alberta, Canada. SPE-196158-MS.

Liu, W., Zhang, Q., Zhu, W., 2019. Numerical simulation of multi-stage fractured horizontal well in low-permeable oil reservoir with threshold pressure gradient with moving boundary. J. Pet. Sci. Eng. 178, 1112–1127. https://doi.org/10.1016/j.petrol.2019.04.033.

Liu, Y.Y., Ma, X.H., Zhang, X.W., Guo, W., Kang, L.X., Yu, R.Z., Sun, Y.P., 2021. A deep-learning-based prediction method of the estimated ultimate recovery (eur) of shale gas wells. Pet. Sci. 18 (6), 1450–1464. https://doi.org/10.1016/j.petsci.2021.08.007.

Lyu, W., Liu, J., 2021. Artificial intelligence and emerging digital technologies in the energy sector. Appl. Energy 303, 117615. https://doi.org/10.1016/j.apenergy.2021.117615.

McGlade, C., Speirs, J., Sorrell, S., 2013. Methods of estimating shale gas resources – comparison, evaluation and implications. Energy 59, 116–125. https://doi.org/10.1016/j.energy.2013.05.031.

Niu, D., 2022. Short-term multi-energy load forecasting for integrated energy systems based on cnn-bigru optimized by attention mechanism. Appl. Energy. https://doi.org/10.1016/j.apenergy.2022.120063 (in press).

Niu, W., Lu, J., Sun, Y., Guo, W., Liu, Y., Mu, Y., 2022. Development of visual prediction model for shale gas wells production based on screening main controlling factors. Energy 250, 123812. https://doi.org/10.1016/j.energy.2022.123812.

Sun, J., Ma, X., Kazi, M., 2018. Comparison of decline curve analysis dca with recursive neural networks rnn for production forecast of multiple wells. In: Proceedings of the SPE Western Regional Meeting. https://doi.org/10.2118/190104-MS. Garden Grove, California, USA. SPE-190104-MS.

Tang, J., Fan, B., Xiao, L., Tian, S., Zhang, F., Zhang, L., Weitz, D., 2021. A new ensemble machine-learning framework for searching sweet spots in shale reservoirs. SPE J. 26 (2), 482–497. https://doi.org/10.2118/204224-PA.

Wang, H., Chen, L., Qu, Z., Yin, Y., Kang, Q., Yu, B., Tao, W.Q., 2020. Modeling of multi-scale transport phenomena in shale gas production — a critical review. Appl. Energy 262, 114575. https://doi.org/10.1016/j.apenergy.2020.114575.

Wang, H., Wang, S., Chen, S., Hui, G., 2023. Predicting long-term production dynamics in tight/shale gas reservoirs with dual-stage attention-based ten-seq2seq model: a case study in duvernay formation. Geoenergy Sci. Eng. 223, 211495. https://doi.org/10.1016/j.geoen.2023.211495.

Wang, K., Li, H., Wang, J., Jiang, B., Bu, C., Zhang, Q., Luo, W., 2017. Predicting production and estimated ultimate recoveries for shale gas wells: a new methodology approach. Appl. Energy 206, 1416–1431. https://doi.org/10.1016/j.apenergy.2017.09.119.

Wang, S., Qin, C., Feng, Q., Javadpour, F., Rui, Z., 2021a. A framework for predicting the production performance of unconventional resources using deep learning. Appl. Energy 295, 117016. https://doi.org/10.1016/j.apenergy.2021.117016.

Wang, T., Wang, Q., Shi, J., Zhang, W., Ren, W., Wang, H., Tian, S., 2021b. Productivity prediction of fractured horizontal well in shale gas reservoirs with machine learning algorithms. Appl. Sci. 11 (24), 12064. https://doi.org/10.3390/app112412064.

Wang, Y., Zou, R., Liu, F., Zhang, L., Liu, Q., 2021c. A review of wind speed and wind power forecasting with deep neural networks. Appl. Energy 304, 117766. https://doi.org/10.1016/j.apenergy.2021.117766.

Wu, K., Chen, Z.X., Li, J., Lei, Z., Xu, J., Wang, K., Li, R., Dong, X., Peng, Y., Yang, S., Zhang, F., Chen, Z.L., Gao, Y., 2019. Nanoconfinement effect on n-alkane flow. J. Phys. Chem. C 123 (28), 16456–16461. https://doi.org/10.1021/acs.jpcc.9b03903.

Xue, L., Liu, Y.T., Xiong, Y., Liu, Y.L., Cui, X., Lei, G., 2021. A data-driven shale gas production forecasting method based on the multi-objective random forest regression. J. Pet. Sci. Eng. 196, 107801. https://doi.org/10.1016/j.petrol.2020.107801.

Yang, R., Liu, W., Qin, X., Huang, Z., Shi, Y., Pang, Z., Zhang, Y., Li, J., Wang, T., 2021. A physics-constrained data-driven workflow for predicting coalbed methane well production using a combined gated recurrent unit and multi-layer perception neural network model. In: Proceedings of the SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/205903-MS. Dubai, UAE. SPE-205903-MS.

You, X., Liu, J., Jia, C., Li, J., Liao, X., Zheng, A., 2019. Production data analysis of shale gas using fractal model and fuzzy theory: evaluating fracturing heterogeneity. Appl. Energy 250, 1246–1259. https://doi.org/10.1016/j.apenergy.2019.05.049.

Zha, W., Liu, Y., Wan, Y., Luo, R., Li, D., Yang, S., Xu, Y., 2022. Forecasting monthly gas field production based on the cnn-lstm model. Energy 260, 124889. https://doi.org/10.1016/j.energy.2022.124889.

Zhang, X., Ma, F., Yin, S., Wallace, C.D., Soltanian, M.R., Dai, Z., Ritzi, R.W., Ma, Z., Zhan, C., Lü, X., 2021. Application of upscaling methods for fluid flow and mass transport in multi-scale heterogeneous media: a critical review. Appl. Energy 303, 117603. https://doi.org/10.1016/j.apenergy.2021.117603.

Zou, C., Lin, M., Ma, F., Liu, H., Yang, Z., Zhang, G., Yang, Y., Guan, C., Liang, Y., Wang, Y., Xiong, B., Yu, H., Yu, P., 2024. Development, challenges and strategies of natural gas industry under carbon neutral target in China. Pet. Explor. Dev. 51 (3), 476–497. https://doi.org/10.1016/s1876-3804(24)60038-8.

Zou, C., Xue, H., Xiong, B., Zhang, G., Pan, S., Jia, C., Wang, Y., Ma, F., Sun, Q., Guan, C., Lin, M., 2021. Connotation, innovation and vision of "carbon neutrality". Nat. Gas Ind. B 8 (6), 523–537. https://doi.org/10.1016/j.ngib.2021.08.009.