Petroleum Science 23 (2026) 643–655

Contents lists available at ScienceDirect

Petroleum Science

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

Original Paper

A machine learning method for evaluating shale gas production based on the TCN-PgInformer model

Hao-Yu Zhang [a,b,c], Wen-Sheng Wu [a,*], Zhang-Xin Chen [d,e,**], Benjieming Liu [d,e]

[a] State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum (Beijing), Beijing, 102249, China
[b] Frontiers Science Center for Rare Isotopes, Lanzhou University, Lanzhou, 730000, Gansu, China
[c] School of Nuclear Science and Technology, Lanzhou University, Lanzhou, 730000, Gansu, China
[d] Eastern Institute of Technology, Ningbo, 315100, Zhejiang, China
[e] Department of Chemical and Petroleum Engineering, University of Calgary, Calgary, T2N 1N4, Canada

ARTICLE INFO

Article history:
Received 22 August 2024
Received in revised form
16 May 2025
Accepted 9 November 2025
Available online 13 November 2025

Edited by Meng-Jiao Zhou

Keywords:
Shale production forecasting
Informer
TCN
Machine learning
Daily gas production

ABSTRACT

Since shale gas is a valuable energy resource, effective planning for its extraction and utilization depends on precise forecasting of gas well production. Conventional models need long computation time, a wide range of geological and fluid data, and suffer from unstable predictions. To develop a low-cost, intelligent, and reliable forecast system for shale gas production, a hybrid Temporal Convolutional Network-Policy Gradient Informer (TCN-PgInformer) model was constructed for multivariate production prediction research. This model is based on the Informer model of its own unique self-attention mechanism, which lowers the temporal complexity of conventional self-attention technique while increasing the model's accuracy. Meanwhile, to completely avoid the gradient vanishing problem, the dilated convolutions of TCN structure are employed to extract the long-term dependency relationships. Ultimately, a policy gradient (Pg) algorithm is introduced to enhance the parameter training speed. The results indicate that the daily gas production may be accurately predicted by TCN-PgInformer model. A detailed performance comparison was carried out among TCN-PgInformer, CNN, GRU and CNN-LSTM models in the literature. The comparison demonstrates that the suggested TCN-PgInformer model outperforms existing techniques. For four different gas production stages, the MAPE/RMSE error of other models is 2–12 times higher than that of the TCN-PgInformer model, while the R^2 accuracy of TCN-PgInformer model can be as high as 1 time higher than other models. Therefore, the designed model has excellent applicability, which offers reference and guidance for shale gas development.

1. Introduction

Recently, thanks to innovative breakthroughs of theory and rapid technological progress, China has made leapfrog development in the exploration and development fields of shale gas (Yuan et al., 2017). Accurately predicting shale gas production is meaningful in geological terms. Reliable production forecasts allow for a deeper insight into reservoir permeability and porosity, and help identify spatial variations in natural gas distribution. In this way, not only can drilling and fracturing designs be optimized to enhance recovery rates, but geological risks can also be avoided, ensuring the safety and economy of the development process. To accurately calculate the recoverable shale gas reserves of single well, it is particularly important to select suitable methods for the production prediction of shale gas (Nguyen and Shin, 2022).

The early method for predicting gas field production was the water-drive curve method (Omoniyi and Adeolu, 2014). But it can only forecast recovery rates under various water contents, and cannot consider the time factor (Boah et al., 2018). With the time passage, numerical simulation methods and formula derivation methods have developed into conventional shale gas production forecasting methods, both of which can predict the daily gas production. Numerical simulation methods primarily rely on computers to simulate the reservoir production status and seepage laws, requiring complete geological and development data.

* Corresponding author.
** Corresponding author.
E-mail addresses: wwsheng@cup.edu.cn (W.-S. Wu), zhachen@ucalgary.ca (Z.-X. Chen).
Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2025.11.022

Nevertheless, for gas reservoirs with short production cycles or in the early development stages, it is impossible to accurately simulate their production conditions. The production decline curve method based on statistical analysis can predict production dynamics by analyzing long-term production data of oil and gas (Fan et al., 2021; Henrik et al., 2017). The fundamental idea behind decline curve analysis is to match historical production and time data with "models" like hyperbolic, harmonic, and exponential models. As ideal curves, all of these models are unable to account for the actual geological conditions. This method makes it hard to guarantee proper performance (Pang et al., 2023). Thus, a more efficient and practical method should be developed, which takes into account the internal decrease of different geological and fluid elements while remaining highly accurate.

With the rapid growth of artificial intelligence and the advancement of on-site big data technology in oil fields, substantial foundations have been laid for machine learning to anticipate oil field production. Scholars are increasingly becoming more interested in machine learning for estimating oil and gas well production (Bhattacharya and Mishra, 2018; Broni-Bediako et al., 2019; Gupta et al., 2014). An example of a time series structure is production data, which is impacted by both internal and external variables. Hence, numerous researchers utilize time series exploration techniques to uncover latent data from historical time series to forecast the performance of oil well output in the future. Zhu et al. (2015) combined the Support Vector Regression (SVR) method with the reconstruction features of time series to create a Support Vector Regression Linear Programming (SVRLP) model for forecasting natural gas reserves. Li and Wang (2017) established a novel grey prediction model, which demonstrated excellent applicability in oil fields. Vos et al. (2018) presented the Convolutional Neural Network (CNN) model for predicting daily gas production, which utilizes the automatic feature extraction capability of CNN to obtain excellent accuracy even in noisy input. Li et al. (2022b) introduced a novel Time Convolutional Network (TCN) approach that may estimate production purely based on wellhead pressure. Machine learning techniques are coupled with other techniques to further increase the prediction model's stability and accuracy. Wang et al. (2024) used a physical information-based neural network (PINN) method for identifying shale gas well decline curves, and further verified that the model can effectively predict the production performance of test wells. On the one hand, machine learning techniques are combined with genetic (GA) algorithms (Sagheer and Mostafa, 2019), particle swarm optimization (PSO) algorithms (Ding et al., 2022), and imperial competition algorithms (ICA) (Gao et al., 2023) to optimize the neural network model's hyperparameters (Fan et al., 2024). Zha et al. (2022) predicted the monthly production of gas fields by Convolutional Neural Network-Long Short-Term Memory Network (CNN-LSTM) model on the basis of bagging strategy and partially unknown recursive prediction strategy (PURPS). Cuthbert et al. (2021) integrated PSO with recurrent neural network (RNN), feedback neural network (FNN), and support vector machines (SVM) to estimate oil well production. On the other hand, adopting multivariate inputs increases the information content of the data itself. Through causal extension convolution, TCN can obtain more sensory fields with fewer network layers to solve longer time series. Although the above models perform well in continuous time series prediction, there is a problem with unstable prediction during the prediction process. These models leave a lot of opportunity for improvement. Informer model is a relatively new model to predict time series data (Ma et al., 2023). Zhuang et al. (2024) utilized Graph Convolutional Network-Informer (GCN-Informer) model to achieve high-precision photovoltaic power generation prediction. Bommidi et al. (2024) created a new hybrid framework called ICEEMDAN-Informer-GWO (ICEEMDAN-Informer-Grey Wolf Optimizer) model, which combines three components to enhance the accuracy of wind speed prediction. In the energy field, Informer is also employed to predict electricity consumption and photovoltaic output coefficient, which helps to achieve optimal allocation of energy resources. But the model has been yet to be applied to shale gas production forecast.

For the prediction issue of gas production time series, an Informer model based on its own special attention mechanism is used to lower the temporal complexity of traditional self-attention mechanisms and improve the accuracy of the model. Meanwhile, to completely avoid the gradient vanishing problem, the dilated convolutions of TCN structure are employed to extract the long-term dependency relationships. Ultimately, a policy gradient (Pg) algorithm is added to enhance the parameter training speed of Informer model. Combining encoding and decoding network of Informer model to process contextual information, thus achieving prediction of shale gas daily production. The predicted results of the proposed TCN-PgInformer model were compared with those of other models in the literature.

2. Methodology

2.1. Informer model

A time series prediction model called Informer model is derived using the standard Transformer model as a basis (He and Xiong, 2024; He et al., 2024; Ren et al., 2023). Informer model has made numerous improvements to improve Transformer-like models' predictive power in long sequence time-series forecasting (LSTF) situations (Wang et al., 2023). Similar to the standardized attention-based Transformer model, the Informer model is also composed of the Encoder and the Decoder. The definition of input at time t is shown in Eq. (1):

$$X^t = \left\{ x_1^t, x_2^t, \cdots, x_{L_x}^t \middle| x_i^t \in R^{d_x} \right\} \tag{1}$$

where L_x is the current input length, and the output at time t is

$$Y^t = \left\{ y_1^t, y_2^t, \cdots, y_{L_y}^t \middle| y_i^t \in R^{d_y} \right\} \tag{2}$$

where L_y is the length of the current output. When dealing with LSTF problems, a longer output length L_y is required.

Secondly, traditional self-attention mechanism vectorizes the input through an algorithm and embeds positional and temporal information. It then scales the dot product of the Query, Key, and Value components, i.e.:

$$\text{Attention}(Q, K, V) = Soft\max\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_K}}\right)V \tag{3}$$

In the equation, $Q \in R^{L_Q \times d}$, $K \in R^{L_K \times d}$, $V \in R^{L_V \times d}$. d_K denotes the input dimension, and the Attention coefficient for the i-th Query component takes the following probability form:

$$\text{Attention}(q_i, K, V) = \sum_j \frac{k(q_i, k_i)}{\sum_l k(q_i, k_l)} v_j = E_{p(k_j|q_i)}\left[v_j\right] \tag{4}$$

where $p\left(k_j \middle| q_j\right) = \frac{k(q_i, k_i)}{\sum_l k(q_i, k_l)}$ and $k(q_i, k_i)$ adopts asymmetric index $\exp\left(\frac{q_i k_j^{\mathrm{T}}}{\sqrt{d_K}}\right)$.

In the previous equation, when calculating probabilities $p\left(k_j \middle| q_j\right)$ by the traditional self-attention mechanism, a dot product with quadratic temporal complexity is employed, and its memory usage needs to be computed. Hence, the predictive ability of the model is limited. Furthermore, studies demonstrated that the self-attention probability distribution exists a certain sparsity, and that each probability $p\left(k_j \middle| q_j\right)$ can be assigned a "specific" counting strategy without substantially altering the mechanism. Consequently, it is essential to assess the classic "sparsity" self-attention learning mode qualitatively first. Self-attention is distributed with a long tail, which makes sense given that only a few dot products would significantly contribute to the primary attention. The contributions of additional dot products are insignificant and may be disregarded (Yan et al., 2023).

This paper introduces KL divergence to calculate the sparsity of the Query component, where the evaluation formula for the sparsity of the i-th Query component is (Liao et al., 2022):

$$M(q_i, K) = \mathrm{In} \sum_{j=1}^{L_K} \mathrm{e}^{\frac{q_i k_j^{\mathrm{T}}}{\sqrt{d_K}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^{\mathrm{T}}}{\sqrt{d_K}} \tag{5}$$

where the first term refers to the process of taking the maximum value for each Key component, and the arithmetic mean is the second. According to the above evaluation equation, the equation for probabilistic sparse (Probsparse) self-attention can be obtained, which is

$$\mathrm{Attention}(Q, K, V) = Soft\ \max\left(\frac{\overline{Q}K^{\mathrm{T}}}{\sqrt{d_K}}\right)V \tag{6}$$

where $\overline{Q}$ is an identically sized sparse matrix to q and only included in the top-u Queries components under sparse evaluation $M(q, K)$. Let $u = c \cdot \mathrm{In}L_Q$ (c represents the constant sampling factor) reduce the computational complexity of Probsparse self-attention mechanism. For each Query-Key component, only $O(\mathrm{In}L_Q)$ dot product operation needs to be calculated. Here, $O(\mathrm{In}L_Q)$ is mainly described the growth level of computational complexity.

Fig. 1 depicts the overall structure of Informer model. It is split into two sections: training part on the left and prediction part on the right. The training data is fed into the Encoder module, which extracts features using the multi-head Probsparse self-attention structure in the wine green section. The retrieved data has a substantially smaller network size, and the robustness of Encoder module can be dramatically increased by combining multi-layer convolution, pooling, and the multi-head Probsparse self-attention structure (Dreher and Krska, 2021). The prediction part inputs the predicted data into the Decoder module. It selects 0 elements to fill the required predicted data, and the multi-head attention module quickly generates decoding elements under the weighting of the feature map. In the end, the decoding elements are output through the fully connected layer.

The Informer decoder is made up of two similar multi-head attention units, and the input vector is:

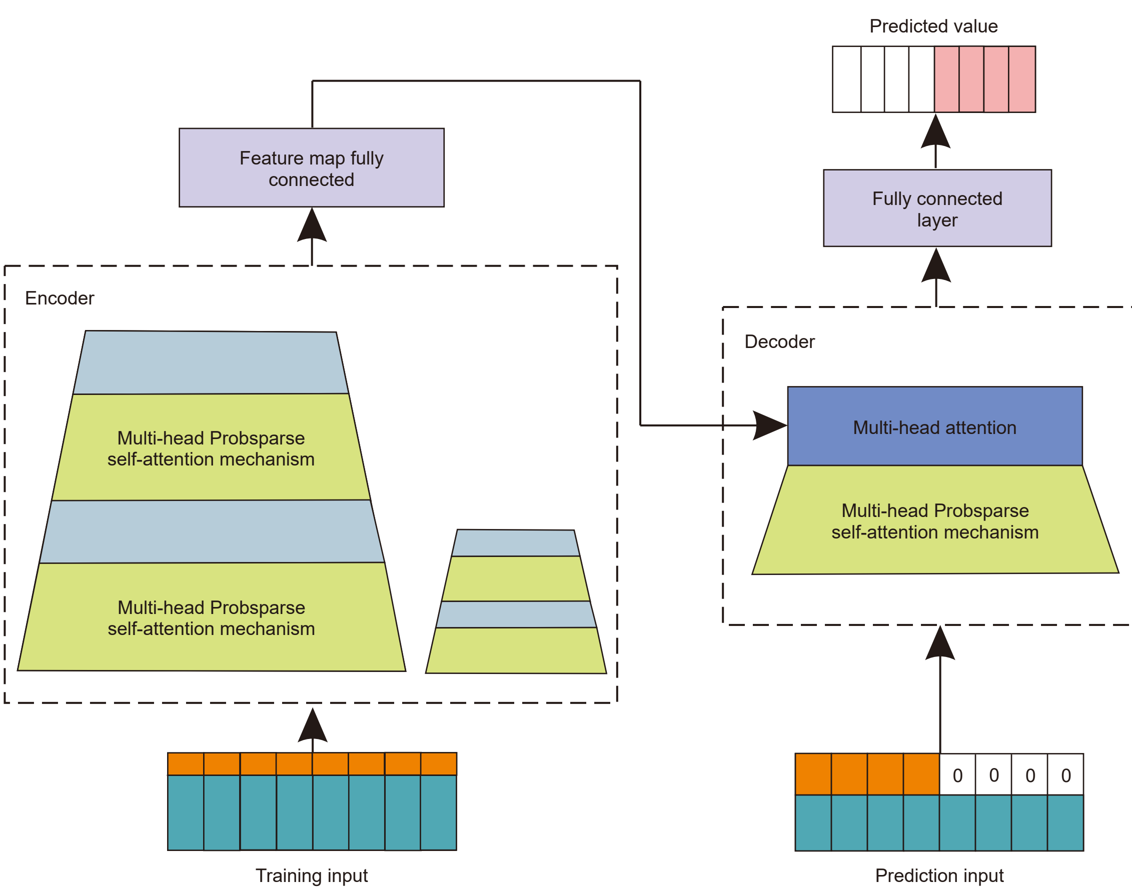

Fig. 1. Schematic diagram of Informer model structure.

$$X_{\mathrm{de}}^t = Concat\left(X_{\mathrm{token}}^t, X_0^t\right) \in R^{(L_{\mathrm{token}}+L_y)\times d_{\mathrm{model}}} \tag{7}$$

where $Concat(\cdot)$ function connects X_{token}^t and X_0^t to form the input vector X_{de}^t of the decoder at time t. $X_{\mathrm{token}}^t \in R^{L_{\mathrm{token}}\times d_{\mathrm{model}}}$ refers to the start token technique, which is derived from the encoder input data depending on the target prediction sequence's length L_y. For example, if the data length of the encoder input is 5D and the target prediction sequence length is 2D, then $L_y = 2$ and $L_{\mathrm{token}} = 3$. X_{token}^t represents the top 3d data in the encoder input. $X_0^t \in R^{L_y\times d_{\mathrm{model}}}$ is the target prediction sequence, and d_{model} indicates the prediction dimension.

2.2. Temporal convolutional network (TCN)

TCN is mainly composed of various residual model stacks (Aslan and Kozat, 2022; Liu et al., 2022; Wang et al., 2022). Two convolution units and one nonlinear mapping unit make up the residual module. Each convolution unit relies heavily on one-dimensional dilated causal convolution operation (Fig. 2). Its mathematical form is described in Eq. (8), where y^t is the predicted variable for n features. There is a relationship that establishes the connection between x_t and y^t, and f is a function that determines what that relationship is. Only the data from before time t is used for processing the value at time t. The aim is to guarantee the order of data processing. When supplying a one-dimensional sequence, the dilated causal convolution operation $F_{(S)}$ and *filter* f:{0, …, $k-1$}*xgboost* on its elements are defined as Eq. (9). In the equation, the input sequence information is denoted by s, the filter size is denoted by k, and the location of specific historical information is indicated by (s-d·i). d= (1, 2, 4) denotes the dilated coefficient. Both the amount of received information and the perceptual field of TCN expand as d increases.

$$y^t = f(x_1, x_2, x_3, \cdots, x_t) \tag{8}$$

$$F_{(s)} = (x{*}df)(s) = \sum\nolimits_{i=0}^{k-1} f(i)\cdot x_{s-d\cdot i} \tag{9}$$

Temporal convolutional network (TCN) is utilized to accomplish feature extraction from the input data of the encoding and decoding layers, hence improving the Informer model's feature extraction capacity. To maintain temporal order, TCN adopts one-dimensional causal convolution to retrieve historical data.

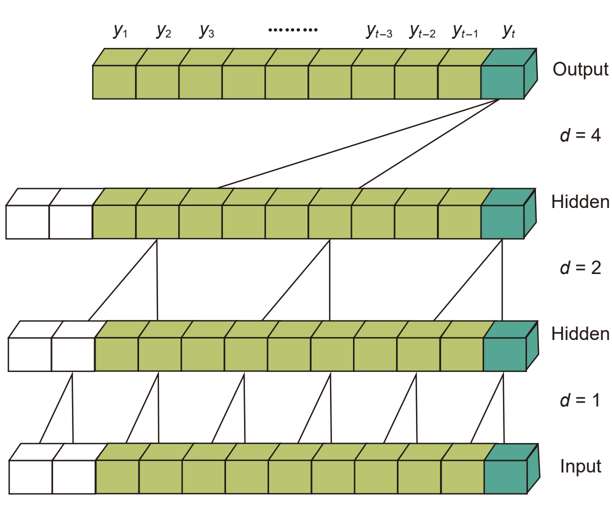

Fig. 2. Schematic diagram of the dilated causal convolution (Liao et al., 2022).

Residual connections accelerate convergence, and dilated convolutions enable the extraction of temporal features, which can gather more extended information dependencies (Dreher and Krska, 2021).

2.3. Policy gradient (Pg) algorithm

In long-term time series prediction tasks, policy gradient algorithm can be used to efficiently maximize Informer model's performance and help them better learn how to extract key features from large amounts of historical data (Wu et al., 2022). Through this reinforcement learning method, the Informer model can dynamically adjust its strategy to achieve higher accuracy in predicting future time points. In this model, the task objective is to minimize prediction error. That is to ensure that the model can accurately predict future time series values as much as possible. For this purpose, the objective function $J(\theta)$ is defined, where θ denotes the parameter set of Informer model (Uchibe and Doya, 2021). The value of $J(\theta)$ is positively correlated with prediction accuracy, representing the overall performance of the model. The core idea of the policy gradient algorithm is to maximize $J(\theta)$ and optimize model parameters (Qian et al., 2021).

For the policy gradient algorithm, the policy is usually represented as $\pi_\theta(a|s)$. It is the probability of the model selecting action a in a given state s. In the Informer model, this strategy can be understood as how to allocate the resources of the self-attention mechanism. That is, how the model selects features at different historical time points to maximize prediction accuracy. By this dynamic allocation, the Informer model can better adapt to the complexity of long series time series data. To optimize the strategy, the gradient $\nabla_\theta J(\theta)$ of objective function $J(\theta)$ is needed to calculate with respect to parameter θ. The policy gradient algorithm provides an effective method to approximate the gradient values using the following equation:

$$\nabla_\theta J(\theta) = E[\nabla_\theta \log \pi_\theta(a|s) \cdot R] \tag{10}$$

where $\nabla_\theta \log \pi_\theta(a|s)$ is the derivative of strategy, which represents the gradient direction under a given strategy. R is a reward signal used to measure the quality of predictions (usually expressed as negative prediction error).

Due to the complexity of directly calculating gradient expectations, Monte Carlo sampling method is utilized for estimation, which involves repeatedly sampling from the strategy to obtain a series of prediction results. Calculate the reward R based on these sampling results, and then obtain an approximate value of the gradient. Next, the gradient rise method is employed to update the parameter θ:

$$\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta) \tag{11}$$

where α is the learning rate used to control the update step size. Through multiple iterations, the strategy parameters of the Informer model gradually tend to be optimal. The policy gradient algorithm is able to track the historical update information of parameter θ, thereby accelerating the convergence of Informer model parameters. In other words, it improves the training speed of Informer model parameter set θ (Xiong et al., 2022).

2.4. TCN-PgInformer model

TCN-PgInformer model is presented to better incorporate the benefits of Informer model, TCN and Pg algorithm. Fig. 3 illustrates its basic structure. The Probsparse self-attention mechanism of Informer model may overlook some long-term sequence

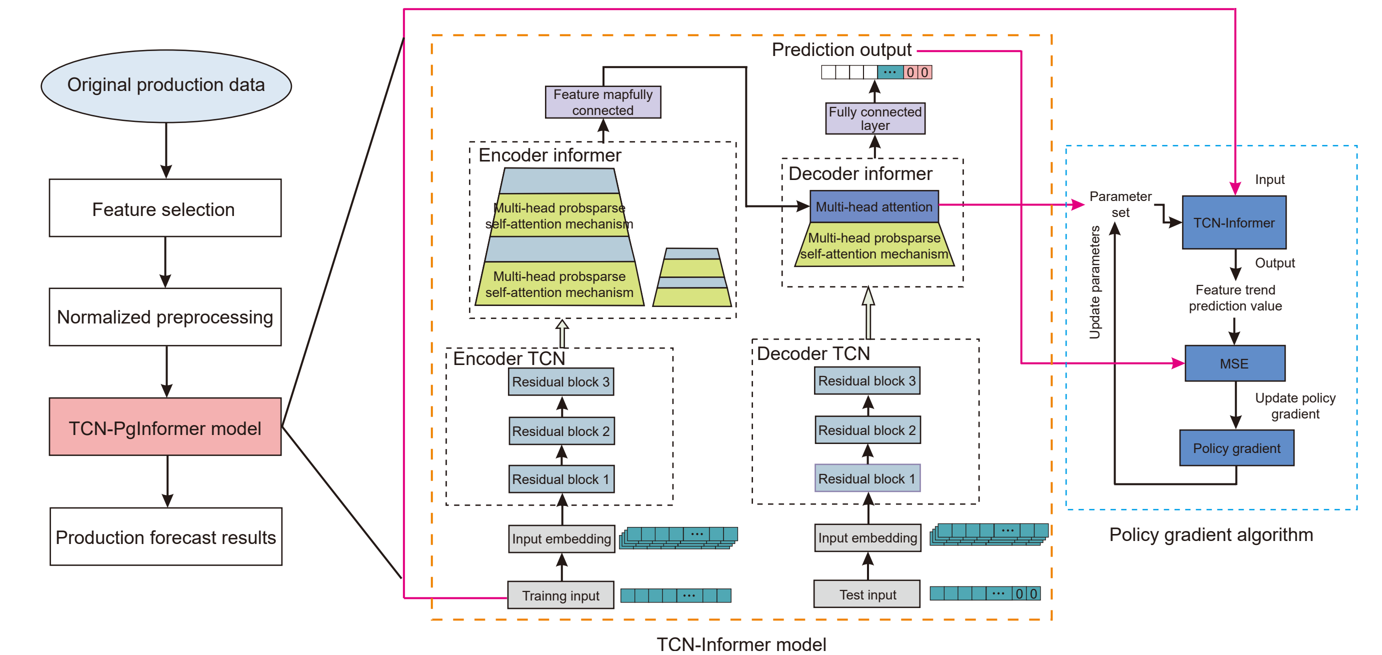

Fig. 3. TCN-PgInformer model structure diagram.

fluctuations, while the advantage of TCN is that it can extract historical information through causal convolution and achieve temporal feature extraction through dilated convolution. Compared to CNN, it increases the receptive field and allows each convolution output to contain a larger range of information, thereby extracting more extended information dependencies and compensating for the shortcomings of Informer model. Moreover, the policy gradient algorithm accelerates the training speed of Informer model parameters.

The prediction procedure of TCN-PgInformer model is described below:

Step 1: Perform feature selection and normalization on the input dataset, and then construct the sliding window. The sliding window method is popular because it meets model training requirements, fully utilizes data, and addresses data challenges in processing time series predictions like shale gas production. The group of sliding window is depicted in Fig. 4. Each training and test unit forms a short time series, including input and output information for the first n time steps. Take the value of n as 5 (Li et al., 2022a), select the first 5 time steps to predict the next time step, that is:

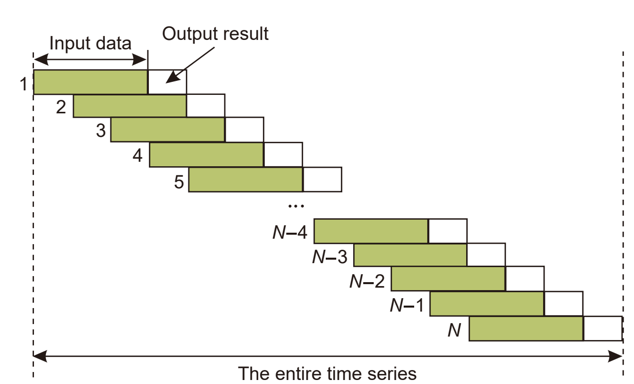

Fig. 4. Schematic diagram of sliding window method.

$$x_{t,\text{pre}} = f(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}) \tag{12}$$

where $x_{t,\text{pre}}$ are the predicted values of the time series at time t. x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5} are the input values related to the first 5 steps.

Step 2: Input the training set into the TCN-PgInformer model, find the optimal model hyperparameters, and achieve the best model performance.

Step 3: Finally, the test set data is placed into the trained prediction model, and the predicted daily production data of test set is output. The forecasting error of shale gas daily production is calculated based on the evaluation indicators.

3. Experimental design

3.1. Data source

To examine the predictive performance of TCN-PgInformer model, historical daily data of two production wells in the same production layer of Qinshui Basin over the past five years were collected through an oil, gas, and water well production data management system (Table 1). Data from eight dimensions, including daily gas production (DGP), daily water production (DWP), production duration (PD), transmission pressure (TP), maximum casing pressure (MCP), average casing pressure (ACP), maximum oil pressure (MOP) and average oil pressure (AOP), were summarized while accounting for the impact of artificial production adjustments during the self-spraying production process. Based on the above 8 dimensions, perform feature selection and choose the data group that has the biggest influence on daily gas production. Calculate by Shapely Additive exPlanations (SHAP) method (Lundberg and Lee, 2017) and plot the feature importance figure based on SHAP score (Fig. 5).

The score values are also displayed in Fig. 5, where the feature importance and difference provide a clearer representation. As can be observed, DGP is positively correlated with other feature parameters except MSP and MOP. Based on SHAP, the DGP is mostly

Table 1
Statistical details of 8 dimensions data.

	DWP, m^3/d	PD, h	TP, MPa	MCP, MPa	ACP, MPa	MOP, MPa	AOP, MPa	DGP, 10^4 m^3/d
Mean	6.2	12.2	25.2	16.7	120.4	18.2	13.4	8.4
Min	1.1	1.4	12.2	5.3	6.1	7.5	8.2	1.5
Median	5.4	8.5	26.3	13.4	17.4	25.5	15.6	5.2
Max	10.2	24	61.6	37.9	28.8	41.8	35.4	14.6

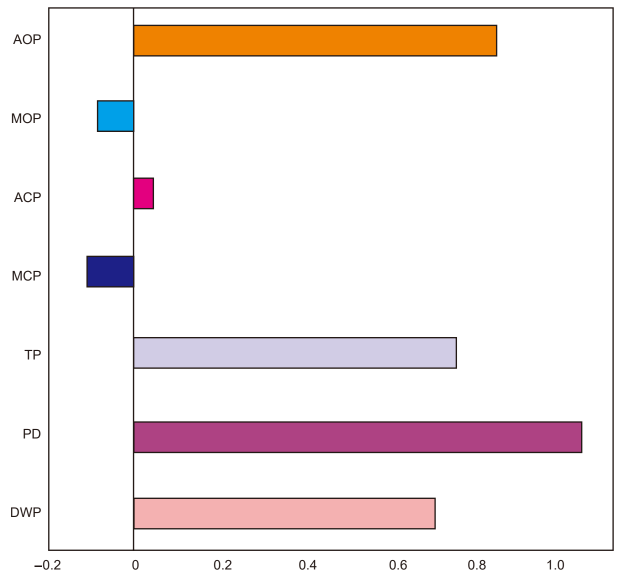

Fig. 5. Feature importance values related to DGP according to the SHAP scores.

influenced by DWP, PD, TP and AOP data. From a theoretical standpoint, daily water production (DWP) affects the fluidity and output capacity of gas, the production duration (PD) determines the total gas output, appropriate transmission pressure (TP) can help ensure smooth gas flow and increase production, and the average oil pressure (AOP) is closely related to the fluidity and production capacity of shale gas. In summary, DWP, PD, TP and AOP are associated with changes in shale gas production, and are also key parameters that need to be monitored and managed in the shale gas production process. As a result, when performing multivariate prediction, DWP, PD, TP and AOP are selected as input feature data for training to predict gas production. For Well I, 1876 d of observations were recorded from March 19, 2016 to May 8, 2021. During the shale gas extraction process, it can be roughly broken into four stages: exploration period (0–325 d), early development period (326–712 d), mid development period (713–992 d) and mature development period (993–1876 d). The first 80% of the samples in each mining operation are utilized as the training set, with the remaining sample data being utilized for testing. The gas production time of Well II is 1245 d, which can also be divided into four stages: exploration period (0–185 d), early development period (186–489 d), mid development period (490–725 d) and mature development period (726–1245 d). The relevant feature data during each mining process is also separated into a 4:1 training to test set ratio.

3.2. Data preprocessing

Before utilizing the dataset, missing value filling and data normalization are necessary steps in the data preprocessing process (Bhattacharya et al., 2019; Rahmanifard et al., 2020).

(1) Missing value filling. Due to equipment wear and tear in 2018 and 2019, 180 sets of data were missing, which accounts for approximately 2% of the preprocessed data. When the amount of missing data is small, the historical data from the same period were chosen to fill in. Historical data of the same period refers to data from the past few years on the same day, which offers consistency or similarity in statistical distribution and trend with the currently missing data. In 2020 and 2021, the oilfield experimental equipment was brand new and improved, and there were 15 sets of missing data. The missing parts only account for 0.17% of the preprocessed data. Because the amount of missing data is fewer and discontinuous (with large missing intervals), they are filled in by averaging the sum of the 50 sets of data before the missing values and the 50 sets of data after the missing values.
(2) Data normalization. To increase the model's prediction performance and reduce dimensional discrepancies across features, normalize the input data prior to training. This paper employs maximum minimal normalization method to convert the original data into the range of [0, 1], with the following normalization equation:

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the input data, respectively.

3.3. Evaluation index

To highlight the impact of TCN-PgInformer model on the production prediction of shale gas well, three commonly used time series prediction methods, CNN (Amin et al., 2022), GRU (Al-Shabandar et al., 2021), CNN-LSTM (Zha et al., 2022) were used as a comparison. The reasons why these three models are selected as benchmark models are as follows. CNN is commonly adopted to process spatial data, such as local patterns in images or time series data. For shale gas data, CNN can identify short-term patterns and local correlations in production time series. As a variant of RNN, GRU performs excellently in processing time series data, especially suitable for handling long-term dependencies. Shale gas production data may contain long-term trends and seasonal characteristics, and GRU is suitable for capturing these complex dynamic changes. CNN-LSTM combines the feature extraction capability of CNN with the time-dependent modeling capability of LSTM, making it suitable for tasks with both spatial structures and temporal dependencies. In shale gas production prediction, CNN-LSTM helps to capture both local and long-term features simultaneously. The evaluation index assesses the model's accuracy of the yield prediction results in this work. The selected evaluation indicators contain determination-coefficient (R^2), mean-absolute-percentage-error (MAPE), and root-mean-square-error (RMSE), which are obtained using the following equation (Klie and Florez, 2020; Mahzari et al., 2022).

Table 2
The optimal parameter settings for different models during the exploration period of shale gas.

Model	Parameter settings
CNN	Filters = 64, kernel_size = (3, 1), units = 128, activation = "relu", epochs = 4000, learning_rate = 0.001.
GRU	Filters = 64, units = 100, Dropout rate = 0.3, activation = "tanh", epochs = 200, learning_rate = 0.01.
CNN-LSTM	$T = 4$, units = 18,16,12, …,10,10,2, epochs = 400, times = 25 (Zha et al., 2022).
TCN-PgInformer	Kernel_width = 5, max pool_size = 3, $\gamma = 0.7$, $\alpha = 0.001$, $n_{\max} = 250$, $\mathrm{MSE}_{\min} = 5 \times 10^{-4}$.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(x_{i,\mathrm{pred}} - x_{i,\mathrm{real}}\right)^2}{\sum_{i=1}^{N}\left(x_{i,\mathrm{pred}} - \overline{x}\right)^2} \tag{14}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|x_{i,\mathrm{pred}} - x_{i,\mathrm{real}}\right|}{x_{i,\mathrm{real}}} \tag{15}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{i,\mathrm{pred}} - x_{i,\mathrm{real}}\right)^2} \tag{16}$$

where n is the scattered point number in the time series. $x_{i,\mathrm{pred}}$ and $x_{i,\mathrm{real}}$ are the predicted and real production values at any given time point.

RMSE reflects the overall difference between predicted and true values, MAPE focuses more on the percentage error between predicted results and true values, and R^2 score is employed to measure how well the model fits the data. When measuring model performance against these benchmarks, the model's predictions are compared with real data. By calculating the three indicators, the prediction accuracy and model stability can be intuitively understood. For example, if RMSE value of the model are small and the MAPE value is within the range of 0–5%, it indicates that the predicted results of the model are close to the real data and the prediction error is low. At the same time, the R^2 score of the model will also be considered to evaluate its fitting effect on the data. If the R^2 score is close to 1, it shows that the model can fit the data well and offers excellent predictive performance.

4. Results

The development process of shale gas can be broadly divided into four stages. The daily production of shale gas shows different trends in each stage. To test the daily gas production prediction ability of TCN-PgInformer model, research can be conducted on shale gas daily production at different stages. Meantime, CNN, GRU and CNN-LSTM models are compared with TCN-PgInformer model. To obtain the optimal hyperparameters for each model, Bayesian optimization method is adopted. By constructing a surrogate model, the optimal parameter combination can be found with fewer experiments. Furthermore, cross validation is used to evaluate the performance of each hyperparameter combination. Based on "Accuracy" performance indicators, it helps to objectively compare the performance of different hyperparameter combinations for various models, ensuring the optimal setting selection.

4.1. Shale gas exploration period

During the exploration period of shale gas development, the primary goal is to evaluate the gas reservoir potential and economic feasibility of underground shale layers. In shale gas Well I, 0–325 d belong to the exploration period of shale gas development. We utilized the feature data from the first 260 d as the training set to understand the gas production changes during this stage, and the feature data from the last 65 d as the test set. In Well II, 0–185 d belong to the exploration period of shale gas development. We used the feature data from the first 148 d as the training set and the feature data from the last 27 d as the test set. In the network structure of TCN-PgInformer, the convolution kernel width of the one-dimensional convolutional filter is 5, and the max pooling layer stride is 3. In the policy gradient algorithm, the discount coefficient γ is 0.7, the learning rate α is 0.001, the maximum training step size $n_{\max}$ is 250, and the MSE training threshold $\mathrm{MSE}_{\min}$ is 5×10^{-4}. After completing the parameter settings of TCN-PgInformer model, the daily gas production trend may be obtained through the training process and prediction process of TCN-PgInformer model.

Table 2 summarizes the best settings for every model in the exploration period. The forecast outcomes of several models based on ideal parameters are displayed in Fig. 6. These are represented by colorful lines that show the predicted outcomes of several

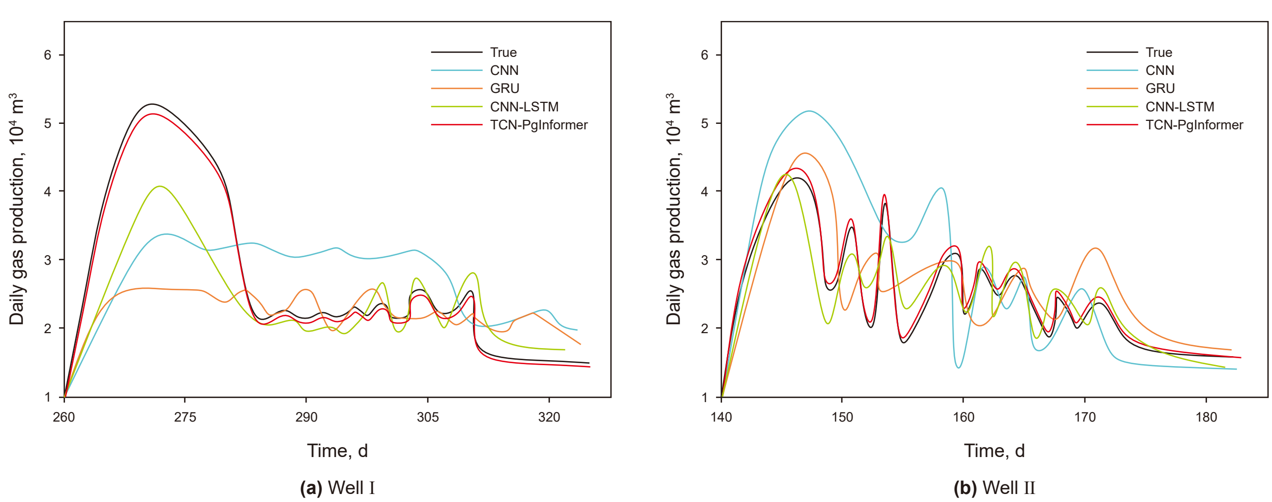

Fig. 6. The prediction results of daily gas production by various models during the exploration period of shale gas well.

Table 3
Evaluation indicators for predicting daily gas production of two wells using different methods during the exploration period of shale gas.

	Model	R^2	MAPE	RMSE
Well I	CNN	0.384	0.564	0.627
	GRU	0.528	0.451	0.453
	CNN-LSTM	0.725	0.247	0.254
	TCN-PgInformer	0.947	0.047	0.044
Well II	CNN	0.426	0.487	0.664
	GRU	0.654	0.417	0.428
	CNN-LSTM	0.812	0.207	0.271
	TCN-PgInformer	0.953	0.031	0.072

models, while the black line shows the actual data. It is evident from the actual daily gas production curve that the initial production during the exploration period is usually higher. This is because the formation pressure is higher after fracturing and production has not yet had an impact on the fracture network, allowing gas to escape rapidly. But as time goes by, shale gas rapidly decreases and eventually tends towards a relatively stable low value.

Fig. 6 illustrates that during the exploration period of shale gas development, the predicted values of TCN-PgInformer model (red lines) and the actual values (black lines) are very well fitted. The predicted result of each measuring point is basically compatible with the actual deformation trend, and TCN-PgInformer model has the greatest prediction effect, indicating that the model provides excellent applicability. To further quantify the prediction accuracy, we evaluated the fitting degree of the neural network model through R^2, MAPE, and RMSE. According to Table 3, CNN model performs the worst. Compared with simple CNN and GRU models, CNN-LSTM model shows a corresponding improvement in accuracy. Meanwhile, TCN-PgInformer model with temporal convolutional neural network exhibits the best prediction accuracy (the maximum R^2 value), and the MAPE value and RMSE value are also smaller.

4.2. Early development period of shale gas

In the early development period of shale gas, the main goal is to validate reservoir potential and optimize production technology by establishing production wells and starting large-scale production. For shale gas Well I, 326–712 d belong to the early development period of shale gas. We adopted 326–635 d feature data as the training set to understand the gas production trend, and 636–712 d feature data as the test set. For shale gas Well II, 186–489 d belong to the early development period of shale gas. We used 186–428 d feature data as the training set and 429–489 d feature data as the test set. TCN-PgInformer model parameter settings are as follows: the convolution kernel width of the one-dimensional convolutional filter is 3, the stride of the one-dimensional max pooling layer is 2, the discount coefficient γ is 0.9, the learning rate α is 0.001, the maximum training steps n_{max} is 700, and the MSE training threshold MSE_{min} is 5×10^{-4}. The training process and prediction process of TCN-PgInformer model can yield the prediction results of daily gas production trend after the parameter settings of TCN-PgInformer model have been completed.

The ideal parameters for every model at the early development period of shale gas are listed in Table 4. Fig. 7 depicts the prediction results for several models based on ideal parameters. It is clear from the actual daily gas production curve that production changes exhibit characteristics of initial stability and rapid peak. At this point, there is a noticeable fluctuation in production tail. Evidently, the predicted daily gas production values by TCN-PgInformer model (red lines) and the actual values (black lines) are extremely similar, and the two changes' trends are quite consistent.

Fig. 7 compares the predicted daily gas production of two wells. The curve fluctuations of TCN-PgInformer model (red lines) for each well are relatively small compared to the actual daily gas production curve, while the predicted curves of the other three models have larger changes. Under TCN-PgInformer model,

Table 4
The optimal parameter settings for different models in the early development period of shale gas.

Model	Parameter settings
CNN	Filters = 64, kernel_size = (3, 1), units = 128, activation = "relu", epochs = 300, learning_rate = 0.01.
GRU	Filters = 64, units = 50, Dropout rate = 0.5, activation = "tanh", epochs = 400, learning_rate = 0.001.
CNN-LSTM	$T = 6$, units = 18,16,12, …,10,10,2, epochs = 500, times = 50 (Zha et al., 2022).
TCN-PgInformer	Kernel_width = 3, max pool_size = 2, $\gamma = 0.9$, $\alpha = 0.001$, $n_{max} = 700$, $MSE_{min} = 5 \times 10^{-4}$.

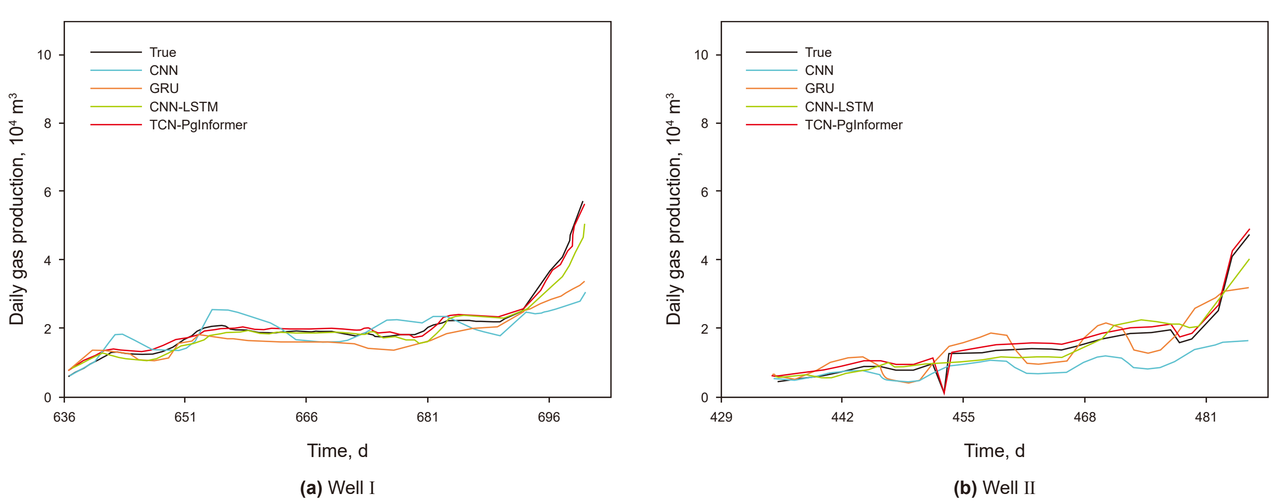

Fig. 7. The prediction results of daily gas production by various models during the early development period of shale gas.

Table 5
Evaluation indicators for predicting daily gas production of two wells using different methods in the early development period of shale gas.

	Model	R^2	MAPE	RMSE
Well I	CNN	0.554	0.357	0.228
	GRU	0.671	0.241	0.156
	CNN-LSTM	0.801	0.169	0.124
	TCN-PgInformer	0.923	0.038	0.042
Well II	CNN	0.593	0.314	0.287
	GRU	0.704	0.275	0.223
	CNN-LSTM	0.825	0.146	0.118
	TCN-PgInformer	0.963	0.031	0.044

nonlinear fluctuations induced by human operation can also be effectively predicted. For example, the well was shut down due to human error on the 454th production day of Well II, which results in a production decrease. However, TCN-PgInformer model displays clear grooves from 453 to 455 d, indicating well shut in operations. Due to the fact that sudden fluctuations are mainly caused by manual operations, CNN (baby blue lines), GRU (orange lines) and CNN-LSTM (wine green lines) models are unable to indicate this abnormal situation, while TCN-PgInformer model (red lines) is able to capture the drastic change. Table 5 shows the evaluation indicators of two wells under four different models. As can be shown, TCN-PgInformer model reflects the lowest error and the maximum accuracy. Hybrid TCN-PgInformer model offers the lowest MAPE, RMSE values and the highest R^2 value, with Well II having an R^2 value as high as 0.963. The daily gas production error values of additional models are 3–10 times greater than those of TCN-PgInformer model. From this, it is obvious that TCN-PgInformer model can adapt well to the prediction of gas well production time series.

4.3. Mid development period of shale gas

The main goal in the mid development period is to maximize economic returns and stabilize production through continuous production and optimized management. For shale gas Well I, the period from 713 to 992 d belongs to the mid development period of shale gas. We selected 713–936 d feature data as the training set to understand the gas production trend in this stage, and the 937–992 d feature data as the test set. For Well II, 490–725 d are considered to be in the mid development period of shale gas. To comprehend gas production trends, we employed training set consisting of 490–678 feature data and test set consisting of 679–725 d feature data. Because of the different input feature data, the optimal parameters for every model are also various. Table 6 presents the ideal parameters of each model.

The predictions made by various models with optimal parameters are displayed in Fig. 8. Black line is the measured data, whereas colored lines show the outcomes of other models' predictions. In general, it is evident that daily gas production fluctuates greatly in the mid development period and has not yet reached a steady level. Gas production often steadily declines when production times are extended. This is because the easily exploitable gas has already been extracted, and the fluidity of the remaining gas is relatively low.

According to Fig. 8, the predicted daily gas production by TCN-PgInformer model (red lines) is more consistent with the actual daily gas production (black lines), and the overall trend is almost identical. CNN model (baby blue lines) and GRU model (orange lines) appear to provide poor performance in predicting daily gas production, with low and unstable overall prediction accuracy. For example, in Well I, the period of daily gas production increase is from 956 to 958 d, but CNN and GRU models show opposite prediction trends, which is an unreasonable performance for daily gas

Table 6
The optimal parameter settings for different models in the mid development period of shale gas.

Model	Parameter settings
CNN	Filters = 32, kernel_size = (3, 1), units = 64, activation = "relu", epochs = 250, learning_rate = 0.01.
GRU	Filters = 64, units = 50, Dropout rate = 0.5, activation = "tanh", epochs = 300, learning_rate = 0.001.
CNN-LSTM	$T = 9$, units = 84, 82, …, 40, 10, 2, epochs = 800, times = 50 (Zha et al., 2022).
TCN-PgInformer	Kernel_width = 2, max pool_size = 4, $\gamma = -4$.

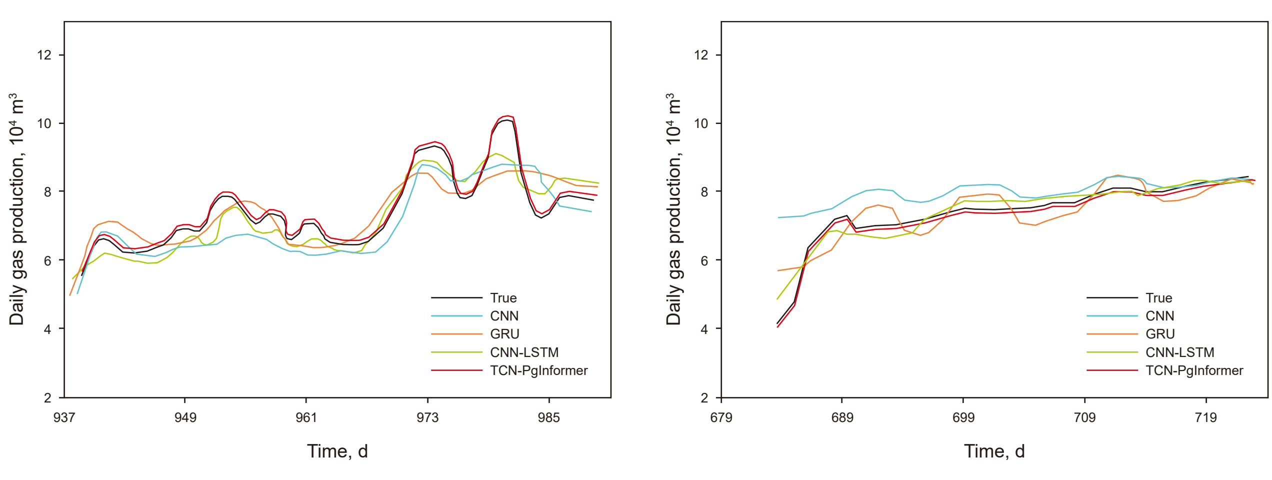

Fig. 8. The prediction results of daily gas production by various models during the mid development period of shale gas.

Table 7
Evaluation indicators for predicting daily gas production of two wells using different methods in the mid development period of shale gas.

	Model	R^2	MAPE	RMSE
Well I	CNN	0.482	0.366	0.339
	GRU	0.592	0.307	0.244
	CNN-LSTM	0.785	0.134	0.138
	TCN-PgInformer	0.952	0.044	0.055
Well II	CNN	0.510	0.409	0.337
	GRU	0.702	0.211	0.238
	CNN-LSTM	0.842	0.128	0.157
	TCN-PgInformer	0.972	0.033	0.039

production prediction. Compared to CNN and GRU models, CNN-LSTM model (wine green lines) exhibits improved accuracy, but its prediction stability is poor and cannot sustain its high accuracy in the test set. Table 7 displays the evaluation indicators of two wells under four different models. TCN-PgInformer model offers the R^2 accuracy that is up to one time more than other models. The MAPE error of other models is 3–12 times larger than that of TCN-PgInformer model, as well as RMSE that is 2–9 times higher. As a result, TCN-PgInformer model is more reliable for estimating daily gas production.

4.4. *Mature development period of shale gas*

The primary purpose of the mature development period is to sustain economic production, extend well service life, and eventually conduct safe well shut in and abandonment. For shale gas Well I, the period from 993 to 1876 d belongs to the mature development period. We selected 993–1699 d feature data as the training set to understand the gas production trend in this stage, and the 1700–1876 d feature data as the test set. For Well II, 726–1245 are considered to be in the mature development period of shale gas. 726–1141 d feature data is regarded as the training set and 1142–1245 d feature data is the training set. Table 8 shows the optimal parameters of each model to ensure the best performance.

The predictions made by several models with optimal parameters are plotted in Fig. 9. Although there may be some upward and downward phenomena during the mature stage, the curve changes gradually tend to stabilize. The production at this stage is significantly reduced compared to the early and middle stages of development. The yield decline curve seems to be linear, which shows a relatively gentle downward trend. In the later stage of maturity, the yield tends to be very low and stable, with a very slow decline rate.

In Fig. 9, the prediction trends of CNN model (baby blue lines) and GRU model (orange lines) fluctuate greatly, which exhibits a more unstable state during the mature development period of shale gas. For example, gas production should tend to stabilize in the later stage of maturity, but CNN and GRU models present an upward trend, which is clearly not in line with reality. The forecasting performance of CNN-LSTM model (wine green lines) has enhanced. It is roughly consistent with the trend variations of gas production, but the gas production values vary greatly. Notably, the daily gas production data predicted by TCN-PgInformer model (red lines) closely matches the actual data (black lines), demonstrating a high degree of consistency. After comparison, it is found that TCN-PgInformer model performs the best. Table 9 shows the evaluation indicators of two wells under four different models (mature development period of shale gas). The R^2 accuracy of TCN-PgInformer model can be as high as 0.964, and the MAPE error (0.045/0.041) of TCN-PgInformer's daily gas production prediction is smaller than CNN (0.618/0.559), GRU (0.382/0.257), and CNN-LSTM (0.144/0.113). The RMSE error of other models is 3–8 times greater than that of TCN-PgInformer model. For this reason, TCN-PgInformer model can precisely pinpoint the moment of stable gas production in the later stage of shale gas development.

Table 8
The optimal parameter settings for different models in the mature development period of shale gas.

Model	Parameter settings
CNN	Filters = 64, kernel_size = (3, 1), units = 64, activation = "relu", epochs = 400, learning_rate = 0.01.
GRU	Filters = 32, units = 50, Dropout rate = 0.4, activation = "tanh", epochs = 400, learning_rate = 0.001.
CNN-LSTM	T = 12, units = 84, 48, …, 100, 50, 12, 2, epochs = 1000, times = 50 (Zha et al., 2022).
TCN-PgInformer	Kernel_width = 3, max pool_size = 5, γ = 0.8, α = 0.001, n_{max} = 600, $MSE_{min} = 4 \times 10^{-4}$.

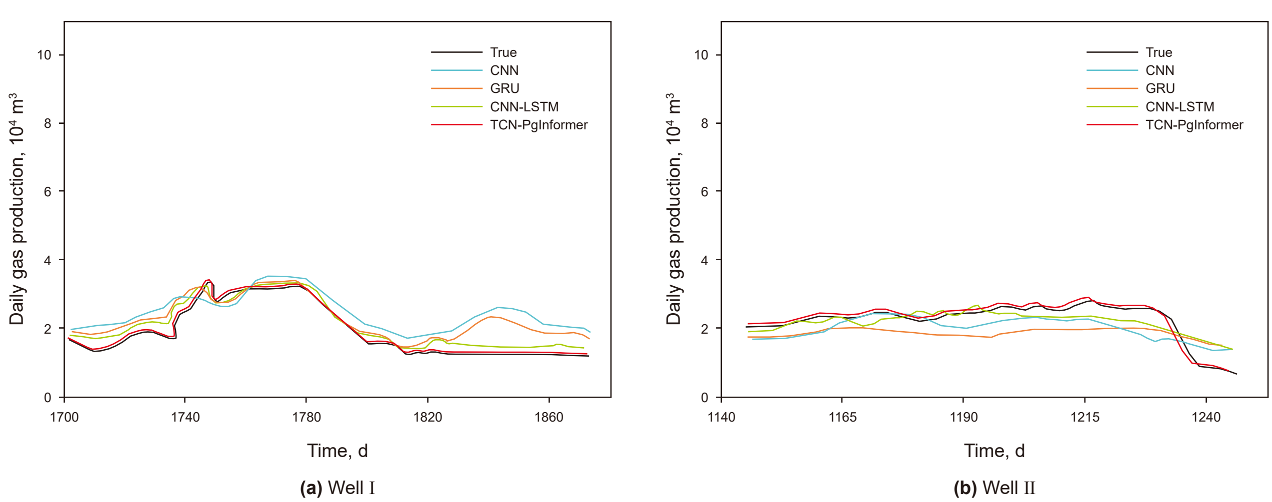

Fig. 9. The prediction results of daily gas production by various models in the mature development period of shale gas.

Table 9
Evaluation indicators for predicting daily gas production of two wells using different methods in the mature development period.

	Model	R^2	MAPE	RMSE
Well I	CNN	0.435	0.618	0.317
	GRU	0.661	0.382	0.218
	CNN-LSTM	0.839	0.144	0.119
	TCN-PgInformer	0.964	0.045	0.041
Well II	CNN	0.482	0.559	0.342
	GRU	0.694	0.257	0.218
	CNN-LSTM	0.822	0.113	0.148
	TCN-PgInformer	0.948	0.041	0.051

5. Discussion

5.1. The generalization performance analysis of TCN-PgInformer model

The generalization performance of TCN-PgInformer model refers to its ability to perform on new data. In other words, it is whether the model can still retain strong prediction accuracy and robustness when dealing with unseen data. The generalization performance of neural network is crucial for their effectiveness in practical applications. The production data from 14 shale gas wells of multiple blocks in Sichuan Basin was applied to this performance validation. This process covers geological exploration, drilling, completion, fracturing, and production data, with a total time span of 3150 d. Here is a summary of the statistics pertaining to the five main characteristic dimensions: average oil pressure, production duration, transmission pressure, daily gas output, and daily water production. Unlike the above experiment, a thorough daily projection of gas output for the shale gas well across various development stages was carried out. Similarly, test set was done on the sample data that remains after the first 80% were utilized as the training set.

Fig. 10 displays the daily gas production prediction results of various models in the new test set. In addition to the comparative models mentioned above, TCN, Informer, TCN-Attention and PINN models have also achieved certain results in predicting shale gas production of Sichuan Basin. Thus, these four models were also added to the comparison sequence in actual case verification. Bayesian optimization method was selected to find the optimal hyperparameter combination for each model. CNN, GRU, TCN and Informer models both predict that the trend of daily gas production fluctuates greatly and cannot reflect the true values very well. The most obvious phenomenon is that the daily gas production rapidly decreased due to human shutdown during the period of 2847–2853 d. But under the prediction of CNN and GRU models, the opposite growth phenomenon occurred. The TCN and Informer models only exhibit a downward trend after this time period, and the prediction results have a lag. These phenomenons reflects the irrationality of the corresponding models. Although CNN-LSTM and TCN-Attention models provide no significant difference in predicting the sharp increase and decrease of daily gas production, they cannot grasp the specific numerical values of daily gas production well. The newly added PINN model performs much better in predicting daily gas production than other comparison models, but its ability to process data at extreme values is not strong. There is a certain gap between the predicted values and the true values. Relative to other models, TCN-PgInformer model shows a stable production time series prediction trend, which is consistent with the true daily gas production trend.

The evaluation index between various models is also demonstrated in Fig. 11. In the absence of non-linear manual operations, TCN-PgInformer model performs well in all evaluation indicators, with R^2, MAPE and RMSE values of 0.968, 0.049 and 0.031 respectively. This model maintains a significant improvement over all the single CNN model (R^2, MAPE, and RMSE of 0.451, 0.564 and 0.611), GRU model (R^2, MAPE, and RMSE of 0.582, 0.411 and 0.430), TCN model (R^2, MAPE, and RMSE of 0.601, 0.388 and 0.417) and Informer model (R^2, MAPE, and RMSE of 0.677, 0.302 and 0.297). In comparison with the hybrid CNN-LSTM model (R^2, MAPE and RMSE of 0.782, 0.237 and 0.231) and TCN-Attention model (R^2, MAPE and RMSE of 0.812, 0.138 and 0.152), TCN-PgInformer model shows 23.8% and 19.2% increases in R^2 respectively. MAPE decreases by 79.3% and 64.5% respectively. RMSE decreases by 67.1% and 79.6% respectively. As an advanced time series model, PINN model provides the R^2, MAPE and RMSE values of 0.968, 0.124 and 0.098. The predicted daily gas production index is satisfactory, but not as accurate as the TCN-PgInformer model. The reason is that TCN-PgInformer model utilizes the benefits of linear and nonlinear units to outperform previous models. Table 10 describes the training time of eight models. It is clear that although the suggested new model is a hybrid model, the training time is only 184 s due to the addition of Pg algorithm, and its time cost is acceptable. TCN-PgInformer network gives reliable model training efficiency while achieving strong generalization performance.

In summary, the hybrid TCN-PgInformer model works well regarding generalization and prediction accuracy. It extracts more extended information through TCN's dilated causal convolution to compensate for the problem of the Informer model's Probsparse self-attention mechanism being hard to grasp long-term series fluctuations. The Pg algorithm enhances the training speed of Informer model parameter set. Especially when the production

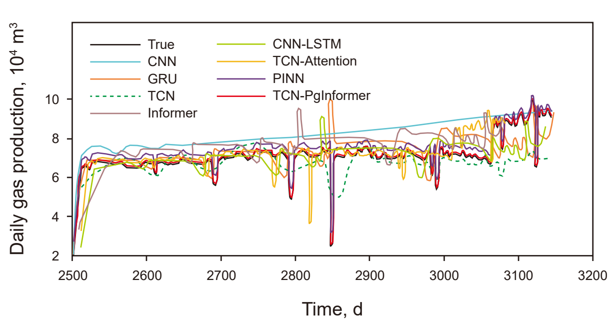

Fig. 10. The performance of CNN, GRU, TCN, Informer, CNN-LSTM TCN-Attention, PINN and TCN-PgInformer models for predicting daily gas production in the new test set.

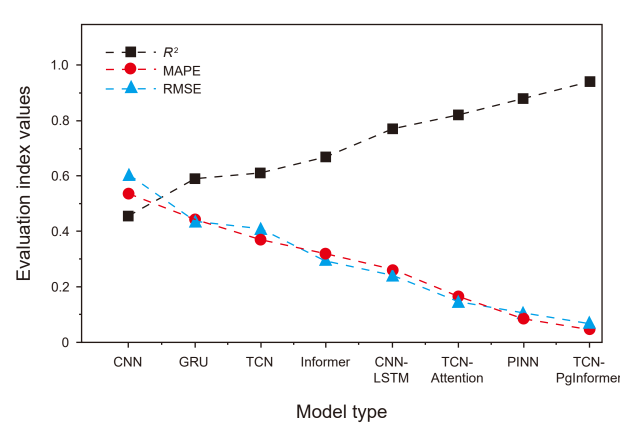

Fig. 11. Evaluation indicators for the performance of eight models.

Table 10
Training time comparison for eight models.

Model	CNN	GRU	TCN	Informer	CNN-LSTM	TCN-Attention	PINN	TCN-PgInformer
Training time, s	52	61	64	71	189	328	204	184

Table 11
Model variants with different structures.

Model variants	Add TCN network?	Add Pg algorithm?
Informer	No	No
TCN-Informer	Yes	No
PgInformer	No	Yes
TCN-PgInformer	Yes	Yes

time series is non-stationary and there is manual operation interference during production, TCN-PgInformer model can better capture this trend. On the contrary, other models perform poorly on nonlinear fluctuations and possess a bigger influence from manual operations. Overall, TCN-PgInformer coupled model offers stronger adaptability and higher efficiency, which can guide engineers in predicting shale gas production.

5.2. Ablation analysis

To test the effectiveness of various modules introduced in the hybrid TCN-PgInformer model, ablation experiments are designed by modifying the model structure to establish different model variants (Table 11). Specifically, by calculating the RMSE values of each model variant for predicting daily gas production within a standard time window, the predictions accuracy of different model variants is compared. Besides, the impact of each module on daily gas production prediction is examined.

As illustrated in Fig. 12, the ablation experiment results reveal that compared to the original Informer model, TCN-Informer and PgInformer models have improved prediction accuracy under different training set ratios. It is worth noting that incorporating both TCN network and Pg algorithm into Informer model can sensibly enhance prediction accuracy. Relative to the original Informer, the *RMSE* value of TCN-PgInformer model decreases by at least 46.1%. Relative to the Informer model with only TCN network or the Informer model with only Pg algorithm, the error of TCN-PgInformer is reduced by 37.3%–57.1%. The experiment shows

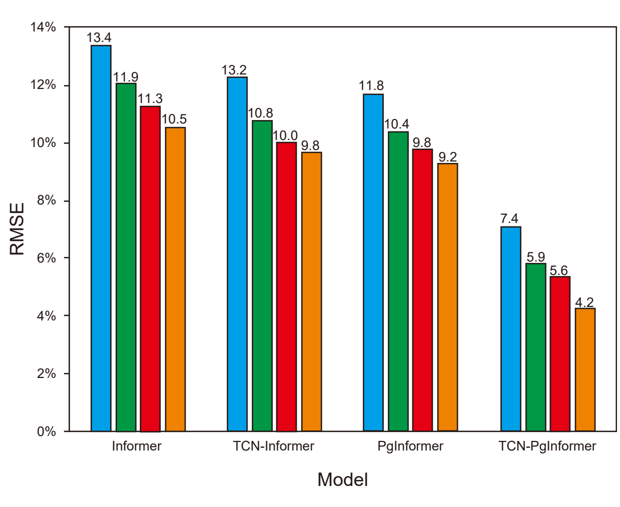

Fig. 12. Ablation analysis result comparison of TCN-PgInformer module variants.

that the addition of TCN module and Pg module plays a crucial part in the performance of the model.

Additionally, for the TCN-PgInformer module, the number increase of training set will generate a favorable impact on the prediction results of daily gas production. The prediction effect is somewhat impacted by the comparatively short production time of shale gas wells at the moment. The improvement path after this study is how to increase yield prediction accuracy with few samples and substantial data variations.

6. Conclusions

To solve the production prediction difficulties of shale gas, such as complex influencing factors and dynamic variability, and enhance the production prediction accuracy, a correlation analysis was conducted on production dimensions. Daily water production (DWP), production duration (PD), transmission pressure (TP), and average oil pressure (AOP) were selected as independent variables, and daily gas production (DGP) was input as the dependent variable into the prediction model. A TCN-PgInformer production prediction model integrating TCN network and Pg algorithm was established. This model is based on the Informer model of its own unique self-attention mechanism, which lowers the temporal complexity of conventional self-attention technique while increasing the model's accuracy. Meanwhile, to completely avoid the gradient vanishing problem, the dilated convolutions of TCN structure are employed to extract the long-term dependency relationships. Ultimately, a policy gradient (Pg) algorithm is introduced to enhance the parameter training speed. The experiments demonstrate the excellent prediction performance and applicability of TCN-PgInformer model. There is no significant difference in predicting the sharp increase and decrease of daily gas production, which reveals the model's excellent stability in predictions. Compared with the single models (CNN, GRU, TCN and Informer) and hybrid models (CNN-LSTM, TCN-Attention and PINN), the forecast accuracy of gas well production in actual case also increased. Besides, the training time is also relatively short. TCN-PgInformer model offers reference and guidance for shale gas development. The improvement path after this study is how to increase yield prediction accuracy with few samples and substantial data variations.

CRediT authorship contribution statement

Hao-Yu Zhang: Writing – review & editing, Writing – original draft, Software, Formal analysis. **Wen-Sheng Wu:** Writing – review & editing, Methodology. **Zhang-Xin Chen:** Writing – review & editing, Supervision. **Benjieming Liu:** Methodology, Writing – review & editing, Software.

Conflict of interest

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

Acknowledgment

We would like to thank PetroChina Xinan Company for providing us with the well logging data. We appreciate the editor and the anonymous reviewers for their constructive comments and suggestions, which significantly improved the quality of this paper.

References

Al-Shabandar, R., Jaddoa, A., Liatsis, P., et al., 2021. A deep gated recurrent neural network for petroleum production forecasting. Mach. Learn. Appl. 3, 100013. https://doi.org/10.1016/j.mlwa.2020.100013.

Amin, S., Narges, Y.G., Amirhossein, N., 2022. Application of rough neural network to forecast oil production rate of an oil field in a comparative study. J. Petrol. Sci. Eng. 209, 109935. https://doi.org/10.1016/j.petrol.2021.109935.

Aslan, F., Kozat, S.S., 2022. Handling irregularly sampled signals with gated temporal convolutional networks. Signal Image Video Process 17 (3), 817–823. https://doi.org/10.1007/s11760-022-02292-2.

Bhattacharya, S., Ghahfarokhi, P.K., Carr, T.R., et al., 2019. Application of predictive data analytics to model daily hydrocarbon production using petrophysical, geomechanical, fiber-optic, completions, and surface data: a case study from the Marcellus Shale, North America. J. Petrol. Sci. Eng. 176, 702–715. https://doi.org/10.1016/j.petrol.2019.01.013.

Bhattacharya, S., Mishra, S., 2018. Applications of machine learning for facies and fracture prediction using Bayesian network theory and random forest: case studies from the Appalachian Basin, USA. J. Petrol. Sci. Eng. 170, 1005–1017. https://doi.org/10.1016/j.petrol.2018.06.075.

Boah, E.A., Borsah, A.A., Brantson, E.T., 2018. Decline curve analysis and production forecast studies for oil well performance prediction: A case study of reservoir X. Int. J. Eng. Sci. 7 (11), 22–30. https://doi.org/10.9790/1813-0711012230.

Bommidi, B.S., Teeparthi, K., Mallesham, V.D., 2024. ICEEMDAN-Informer-GWO: A hybrid model for accurate wind speed prediction. Environ. Sci. Pollut. Res. 31 (23), 34056–34081. https://doi.org/10.1007/s11356-024-33383-x.

Broni-Bediako, E., Fuseini, N.I., Akoto, R.N., et al., 2019. Application of intelligent well completion in optimising oil production from oil rim reservoirs. Adv. Geo-Energy Res. 3 (4), 343–354. https://doi.org/10.26804/ager.2019.04.01.

Cuthbert, S., Ashkan, J.G., Menan, N.A., 2021. Well production forecast in volve field: application of rigorous machine learning techniques and metaheuristic algorithm. J. Petrol. Sci. Eng. 208, 109468. https://doi.org/10.1016/J.PETROL.2021.109468.

Ding, S.W., Xi, Y., Liu, Q., et al., 2022. An automatic optimization method of CO_2 injection for enhanced oil recovery and storage in low permeability reservoirs based on particle swarm optimization algorithm. J. China Univ. Pet. Ed. Nat. Sci. 46 (4), 109–115. https://doi.org/10.3969/j.issn.1673-5005.2022.04.013 (in Chinese).

Dreher, S.D., Krska, S.W., 2021. Chemistry informer libraries: conception, early experience, and role in the future of cheminformatics. Acc. Chem. Res. 54 (7), 1586–1596. https://doi.org/10.1021/acs.accounts.0c00760.

Fan, D., Sun, H., Yao, J., et al., 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, D.Y., Yang, C., Sun, H., et al., 2024. Production prediction of shale gas well based on time series similarity and machine learning method. J. China Univ. Pet. Ed. Nat. Sci. 48 (3), 119–126. https://doi.org/10.3969/j.issn.1673-5005.2022.04.013 (in Chinese).

Gao, Y.J., Tang, L.H., Wang, Z.P., 2023. Oil field production prediction method based on recurrent neural network and data differential processing. China Offshore Oil Gas 35 (3), 126–136. https://doi.org/10.11935/j.issn.1673-1506.2023.03.013 (in Chinese).

Gupta, S., Fuehrer, F., Benin, C., 2014. Production forecasting in unconventional resources using data mining and time series analysis. In: The SPE/CSUR Unconventional Resources Conference. Alberta, Canada.

He, Z., He, Z., Li, S., et al., 2024. A ship navigation risk online prediction model based on informer network using multi-source data. Ocean Eng. 298, 117007. https://doi.org/10.1016/j.oceaneng.2024.117007.

He, Z.H., Xiong, Z.Q., 2024. Research on mining pressure prediction of working face based on informer neural network. Min. Res. Dev. 44 (7), 142–148. https://doi.org/10.13827/j.cnki.kyyk.2024.07.015 (in Chinese).

Henrik, W., Linnea, L., Kjell, A., 2017. Production decline curves of tight oil wells in Eagle ford shale. Nat. Resour. Res. 26 (3), 365–377. https://doi.org/10.1007/s11053-016-9323-2.

Klie, H., Florez, H., 2020. Data driven prediction of unconventional shale reservoir dynamics. SPE J. 25 (5), 2564–2581. https://doi.org/10.2118/193904-PA.

Li, D., Wang, T., 2017. Single-well oil yield prediction method based on grey theory and genetic algorithm. Tianjin Sci. Technol. 44 (8), 48–51. https://doi.org/10.14099/j.cnki.tjkj (in Chinese).

Li, D., Wang, Z., Zha, W., et al., 2022. Predicting production-rate using wellhead pressure for shale gas well based on temporal convolutional network. J. Petrol. Sci. Eng. 216, 110644. https://doi.org/10.1016/J.PETROL.2022.110644.

Li, X., Xiao, K., Li, X., 2022. A well rate prediction method based on LSTM algorithm considering manual operations. J. Petrol. Sci. Eng. 210, 110047. https://doi.org/10.1016/J.PETROL.2021.110047.

Liao, X.C., Zhu, C.H., Zhao, H.Y., et al., 2022. Iron ore pellet process prediction model based on MTCN-informer. Comput. Technol. Dev. 34 (9), 188–194. https://doi.org/10.20165/j.cnki.ISSN1673-629X.2024.0155 (in Chinese).

Liu, T.H., Qiao, X.M., Jin, C.F., et al., 2022. TCN short term wind power prediction with attention mechanism and parameter optimization. Proc. CSU-EPSA 56 (4), 273–286. https://doi.org/10.19635/j.cnki.csu-epsa.000761 (in Chinese).

Lundberg, S.M., Lee, S., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4768–4777.

Ma, Z., Luo, W., Jiang, J., 2023. Spatial and temporal characteristics analysis and prediction model of PM 2.5 concentration based on spatio temporal-informer model. PLoS One 18 (6), 125–137. https://doi.org/10.1371/journal.pone.0287423.

Mahzari, P., Emambakhsh, M., Temizel, C., et al., 2022. Oil production forecasting using deep learning for shale oil wells under variable gas-oil and water-oil ratios. Petrol. Sci. Technol. 40 (4), 445–468. https://doi.org/10.1080/10916466.2021.2001526.

Nguyen, L.V., Shin, H., 2022. Artificial neural network prediction models for Montney shale gas production profile based on reservoir and fracture network parameters. Energy 244, 123150. https://doi.org/10.1016/J.ENERGY.2022.123150.

Omoniyi, O.A., Adeolu, S., 2014. Decline curve analysis and material balance, as methods for estimating reserves (A case study of D4 and E1 fields). Int. J. Innov. Res. Dev. 3 (10), 250–260.

Pang, L.S., Wang, Y., Jiang, W., et al., 2023. Research on short production cycle carbonate gas well production prediction based on machine learning. Special Oil Gas Reservoirs 30 (2), 134–141. https://doi.org/10.3969/j.issn.1006-6535.2023.02.017 (in Chinese).

Qian, Y.M., Xiong, F.Z., Liu, Z.Y., 2021. Zero-shot policy generation in lifelong reinforcement learning. Neurocomputing 446, 65–73. https://doi.org/10.1016/j.neucom.2021.02.058.

Rahmanifard, H., Alimohamadi, H., Gates, I., 2020. Well performance prediction in montney formation using machine learning approaches. In: SPE/AAPG/SEG Unconventional Resources Technology Conference.

Ren, S., Wang, X., Zhou, X., et al., 2023. A novel hybrid model for stock price forecasting integrating encoder forest and informer. Expert Syst. Appl. 234, 215–230. https://doi.org/10.1016/j.eswa.2023.121080.

Sagheer, A., Mostafa, K., 2019. Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 323 (1), 203–213. https://doi.org/10.1016/j.neucom.2018.09.082.

Uchibe, E., Doya, K., 2021. Forward and inverse reinforcement learning sharing network weights and hyperparameters. Neural Netw. 144, 138–153. https://doi.org/10.1016/j.neunet.2021.08.017.

Vos, M., Bender-Saebelkampf, C., Albayrak, S., 2018. Residential short-term load forecasting using convolutional neural networks. In: IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (Smart Grid Comm).

Wang, H., Wang, M.M., Chen, S.N., et al., 2024. A novel governing equation for shale gas production prediction via physics-informed neural networks. Expert Syst. Appl. 248, 123387. https://doi.org/10.1016/j.eswa.2024.123387.

Wang, H., Yang, J., Shi, L., et al., 2022. Remaining useful life prediction based on adaptive SHRINKAGE processing and temporal convolutional network. Sensors 22 (23), 9088–9098. https://doi.org/10.3390/s22239088.

Wang, X.H., Xia, M.C., Deng, W.W., 2023. MSRN-Informer: time series prediction model based on multi-scale residual network. IEEE Access 11, 65059–65065. https://doi.org/10.1109/ACCESS.2023.3289824.

Wu, T., Peng, J.J., Zhang, W.Q., et al., 2022. Video sentiment analysis with bimodal information-augmented multi-head attention. Knowl. Base Syst. 235, 107676. https://doi.org/10.1016/j.knosys.2021.107676.

Xiong, J.H., Li, F., Tang, B.P., et al., 2022. Remaining life prediction of rolling bearings based on policy gradient informer model. Adv. Eng. Sci. 56 (4), 273–286. https://doi.org/10.12454/j.jsuese.202300136 (in Chinese).

Yan, J.W., Yi, C.C., Huang, T., et al., 2023. Remaining life prediction of rolling bearings based on CSPA-informer model. Mod. Mach. Tool. Auto. Manuf. Tech. 10, 85–90. https://doi.org/10.13462/j.cnki.mmtamt.2023.10.019 (in Chinese).

Yuan, Y.S., Jin, Z.J., Zhou, Y., et al., 2017. Burial depth interval of the shale brittle–ductile transition zone and its implications in shale gas exploration and production. Pet. Sci. 14 (4), 637–647. https://doi.org/10.1007/s12182-017-0189-7.

Zha, W., Liu, Y., Wan, Y., et al., 2022. Forecasting monthly gas field production based on the CNN-LSTM model. Energy 260, 124889. https://doi.org/10.1016/J.ENERGY.2022.124889.

Zhu, L., Li, M.S., Wu, Q.H., et al., 2015. Short-term natural gas demand prediction based on support vector regression with false neighbours filtered. Energy 80, 428–436. https://doi.org/10.1016/j.energy.2014.11.083.

Zhuang, W., Li, Z.H., Wang, Y., et al., 2024. GCN–Informer: a novel framework for mid-term photovoltaic power forecasting. Appl. Sci. 14 (5), 2181–2187. https://doi.org/10.3390/app14052181.