Petroleum Science 21 (2024) 3326–3339

Contents lists available at ScienceDirect

Petroleum Science

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

Original Paper

A novel framework for predicting non-stationary production time series of shale gas based on BiLSTM-RF-MPA deep fusion model

Bin Liang [a, *], Jiang Liu [b], Li-Xia Kang [c], Ke Jiang [a, c], Jun-Yu You [d], Hoonyoung Jeong [e, f], Zhan Meng [a, **]

[a] State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu, 610500, Sichuan, China
[b] Xinjiang Oilfield Corporation, PetroChina, Karamay 834000, Xinjiang, China
[c] Research Institute of Petroleum Exploration & Development, Beijing, 100086, China
[d] Chongqing University of Science and Technology, Chongqing, 401331, China
[e] Department of Energy Resources Engineering, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea
[f] Institute of Engineering Research, Seoul National University, Seoul, 08826, Republic of Korea

ARTICLE INFO

Article history:
Received 29 January 2024
Received in revised form
30 March 2024
Accepted 20 May 2024
Available online 22 May 2024

Edited by Meng-Jiao Zhou

Keywords:
Production forecasting
Shale gas
BiLSTM-RF-MPA model
Nonstationary production time series
Deep learning

ABSTRACT

Shale gas, as an environmentally friendly fossil energy resource, has gained significant commercial development and shows immense potential. However, accurately predicting shale gas production faces substantial challenges due to the complex law of decline, nonlinear and non-stationary features in production data, which greatly repair the robustness of current models in predicting shale gas production time series. To address these challenges and improve accuracy in production forecasting, this paper introduces a novel and innovative approach: a hybrid proxy model that combines the bidirectional long short-term memory (BiLSTM) neural network and random forest (RF) through deep learning. The BiLSTM neural network is adept at capturing long-term dependencies, making it suitable for understanding the intricate relationships between input and output variables in shale gas production. On the other hand, RF serves a dual purpose: reducing model variance and addressing the concept drift problem that arises in non-stationary time series predictions made by BiLSTM. By integrating these two models, the hybrid approach effectively captures the inherent dependencies present in long and nonstationary production time series, thereby reducing model uncertainty. Furthermore, the combination of BiLSTM and RF is optimized using the recently-proposed marine predators algorithm (MPA) to fine-tune hyperparameters and enhance the overall performance of the proxy model. The results demonstrate that the proposed BiLSTM-RF-MPA model achieves higher prediction accuracy and demonstrates stronger generalization capabilities by effectively handling the complex nonlinear and non-stationary characteristics of shale gas production time series. Compared to other models such as LSTM, BiLSTM, and RF, the proposed model exhibits superior fitting and prediction performance, with an average improvement in performance indicators exceeding 20%. This innovative framework provides valuable insights for forecasting the complex production performance of unconventional oil and gas reservoirs, which sheds light on the development of data-driven proxy models in the field of subsurface energy utilization.
© 2024 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Shale gas is a kind of unconventional natural gas resource with enormous potential for development (Wang et al., 2014). Due to its characteristics of clean and high energy density, it meets the needs of global energy consumption structure in the transition process of green, low-carbon, and sustainable development. Therefore, the development of shale gas resources has been highly valued by

* Corresponding author.
** Corresponding author.
E-mail addresses: binliang@swpu.edu.cn (B. Liang), zhan.meng@swpu.edu.cn (Z. Meng).

https://doi.org/10.1016/j.petsci.2024.05.012
1995-8226/© 2024 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

energy-producing or consuming countries such as the United States, Canada, and China (Yuan et al., 2015; Lyu et al., 2021; You et al., 2020). Take the case of the United States as an example, thanks to the successful commercial application and continuous development and improvement of multi-fractured horizontal well (MFHW) technology, the total production of shale gas in the United States has grown approximately 14 times from 2007 to 2021 (Wang et al., 2021; U.S. EIA, 2021). Unlike conventional oil and gas resources, shale reservoirs typically have ultra-low permeability (reaching the nanodarcy level, 10^{-15}–10^{-3} μm^2) and extremely strong heterogeneity (containing a large number of natural fractures with significant differences in distribution and arrangement) (Wang et al., 2020; Gale et al., 2014). The extreme reservoir properties are the main reason for the lack of natural productivity of shale gas wells. Large-scale horizontal drilling and hydraulic fracturing operations are necessary ways to improve shale reservoir permeability and increase single-well production (Wang et al., 2020; Dong et al., 2016). However, in the actual production process, due to the comprehensive effects and influences of various complex factors such as multi-scale fracture-matrix systems, complex gas diffusion and transport mechanisms, and macroscopic production measures, shale gas production declines rapidly and has strong uncertainties (i.e., production sequences have complex nonlinear and non-stationary characteristics), which poses significant challenges to shale gas production forecasting (Kocoglu et al., 2021; Liang et al., 2020; Sagheer and Kotb, 2019). Reasonable and accurate shale gas production forecasting methods are important supports to enhance decision-making level, measure economic benefits for specific development plans, and improve production measures, which is of great significance for the efficient development of shale gas reservoirs.

Currently, various methods have been proposed for the dynamic prediction of shale gas production, which can be roughly divided into three categories: theoretical formula method, numerical simulation method, and data-driven method (Liang et al., 2023). The decline curve analysis technique is a commonly used theoretical formula method in the petroleum industry. Traditional empirical decline models mainly describe the changing relationship between production and time by fitting the decline pattern in production dynamic data. Due to its simplicity and ease of implementation, this method has been widely used in various oil and gas reservoir developments capturing the key trends in production decline. The modern decline analysis method is an improved mathematical model that incorporates the theory of unstable seepage with production decline analysis and is more suitable for describing the complex production dynamics of unconventional reservoirs such as shale (Sun et al., 2017). Sun et al. (2018) established a modern decline analysis model that can consider the fracture variable conductance capacity for hydraulically fractured tight gas wells, which achieves higher accuracy than conventional models. Lu et al. (2019) developed a semi-analytical model suitable for analyzing the production dynamics of multi-stage fractured shale gas horizontal wells under variable bottom hole pressure conditions based on the Blasingame decline curve, which fits well with field data. However, these theoretical formula methods generate smooth and idealized production curves, which can only fit a small amount of non-linear variation and have limited prediction accuracy (Li et al., 2022a). Reservoir numerical simulation is a robust and reliable production evaluation technique for shale reservoirs, which can incorporate important processes such as gas adsorption and desorption, stress sensitivity, and large-scale hydraulic fracturing in the production process (Kalantari-Dahaghi et al., 2015). Benefiting from the comprehensive application of multiple interdisciplinary and technological fields, this simulation technique based on commercial software has been able to increasingly reflect the complex seepage processes of multi-scale and multi-physics fields in shale gas reservoirs (Shen et al., 2016; Chen et al., 2019; Zhong et al., 2020). However, numerical simulation techniques have the drawbacks of complex modeling processes, large computational workloads, and huge time consumption. Its prediction accuracy not only relies heavily on a large amount of logging, seismic, and well-test data but also requires high-quality historical fitting of the model (Li et al., 2022a; Zhao et al., 2015; Hutahaean et al., 2015).

Data-driven methods are emerging production dynamic prediction methods that integrate artificial intelligence (AI) technology with relevant theories in oil and gas development. With the push of the big data era and the continuous breakthroughs and innovations in the AI industry, data-driven methods have been widely researched and applied in the field of oil and gas production (Abad et al., 2022; Zhong et al., 2019, 2021; Kim et al., 2023; Jeong et al., 2018; Nwachukwu et al., 2018; Jo et al., 2022). The results of numerous studies have shown that data-driven methods based on machine learning/deep learning have strong non-linear mapping ability, good robustness, and the ability to avoid complicated modeling processes, making them a reliable alternative tool for the dynamic prediction of oil and gas production. Traditional data-driven predictive models mostly rely on machine learning algorithms to construct multiple regression proxy models that can explore the correlation between dynamic and static production parameters and oil and gas production (Liang et al., 2023; Mahdaviara et al., 2022). Wang et al. (2019) identified the main controlling factors affecting the cumulative oil production of Bakken shale reservoirs after fracturing through sensitivity analysis and established a prediction model using a feedforward neural network with multiple hidden layers. Xue et al. (2021) generated a dataset containing geological and hydraulic fracturing characteristics using numerical simulation techniques and used a multi-objective random forest method to extract the quantitative relationship between the single well shale gas production and these static physical parameters, and the established model obtained good predictive ability. Liu et al. (2021) constructed a shale ultimate recovery prediction model with a deep feedforward neural network that can comprehensively consider various factors such as geology, engineering, and production, and found that the quality of input data has an important influence on the model accuracy. Li et al. (2022b) developed a small-sample production prediction framework for fractured wells using a multi-task learning model, which can extract relevant features from unstructured data more effectively. In general, these multiple regression proxy models have a wide range of applications but require higher quality input feature data, and researchers need to have substantial field development experience or professional physical mechanism analysis ability to obtain key factors affecting the target reservoir production (Vikara et al., 2020; Liu et al., 2020; Mohd et al., 2021). Although there are some feature engineering techniques (common algorithms such as principal component analysis, random forest, etc.) or deep learning models that can assist researchers in extracting more important feature parameters, it is difficult to collect a comprehensive and high-quality feature dataset due to the complexity of production mechanisms in shale reservoirs and the immaturity of intelligent data monitoring facilities in the field (Zhou et al., 2014; Mohammadpoor and Torabi, 2020; Desai et al., 2021).

In recent years, due to the outstanding performance of recurrent neural networks with long short-term memory (LSTM) in time series prediction, more and more scholars have introduced it into the field of dynamic prediction of unconventional reservoir production, such as shale, to capture the dependency relationship of production sequence data on the time dimension. Kocoglu et al. (2021) used a bidirectional LSTM (BiLSTM) neural network to

establish a prediction model for monthly shale gas production data of multi-horizontal wells in the Marcellus formation and optimized the network structure and hyperparameters using the Bayesian algorithm. The results showed that compared with the power law exponent, Duong model, LSTM, gated recurrent unit (GRU), and other methods, BiLSTM had a better capturing ability for production patterns. Unfortunately, the changing trend of production data used in this work was relatively simple and could not fully highlight the learning ability of BiLSTM in dealing with production data with complex nonlinear features. Li et al. (2022a) took the oil nozzle size and well shut-in time related to shale oil daily production changes into consideration and established a multivariate production time series prediction framework. The framework mainly consisted of the sparrow search algorithm and BiGRU, achieving a one-day-ahead production prediction. Li et al. (2022c) combined bidirectional GRU with a deep hybrid neural network to establish a multi-step ahead prediction model that can simultaneously consider various physical constraints and production time series of fractured wells. Although these physical constraints can improve the accuracy and interpretability of the model to some extent, they also require high-quality feature information, similar to multiple regression proxy models. Aranguren et al. (2022) attempted to introduce the Seq2Seq architecture into this field and built a Seq2Seq-LSTM model based on LSTM neural network that can predict shale oil production for the next 10 days in one shot. The results showed that although the model performed poorly in various error metrics, it was better than conventional methods such as decline curve analysis and numerical simulation. In summary, current research on production time series prediction can be divided into two types. One is to purely explore the correlation of production sequences themselves by using deep recurrent neural networks and other data preprocessing methods, and the other is a multivariate time series model that integrates other time-related physical constraint features. Essentially, their goals are to deal with and capture the complex nonlinear changes in non-stationary production time series, and achieve accurate production prediction over longer periods. Kuznetsov and Mohri (2020) proposed a data-dependent learning boundary that can be used to guide and design new non-stationary time series prediction algorithms. The learning boundary is represented by a data dependence measure of sequence complexity and a difference measure, where the difference measure can be estimated from the data under some modest assumptions. Arik et al. (2022) proposed a novel adaptive prediction method to address the problem that actual time series data often violate the assumption of standard supervision (its distribution will drift over time), which improves the accuracy of the model in severely non-stationary data prediction tasks to a certain extent. However, current work still shows great uncertainty in dealing with time series with high non-stationary levels while requiring numerous subsidiary and hard-to-get conditions to guide their models.

To move the step forward, we established a hybrid model based on BiLSTM, random forest (RF), and recently-proposed marine predators algorithm (MPA) for ultra-short-term multiple-step ahead prediction of actual shale gas production time series. Using the daily production data of a shale gas well as an example, we compared and validated the adaptability of the model in non-stationary environments and its capability to extract complex nonlinear temporal information, and successfully achieved multiple-step ahead prediction. Results showed that the BiLSTM-RF-MPA model, by balancing the variance and bias between models, demonstrated significant superiority in ultra-short-term multiple-step ahead prediction. The whole work in this study is arranged as follows: Section 2 explains the theoretical background and principles of the methods involved in this study (mainly including BiLSTM, RF, and MPA), as well as the motivation for the proposed BiLSTM-RF-MPA model. Section 3 provides a detailed case analysis of the practical application of the proposed model. Section 4 summarizes and discusses the results of the whole work and possible future research directions.

2. Methods

2.1. Long short-term memory model

LSTM (long short-term memory) neural network is a type of recurrent neural network (RNN) specifically designed to handle sequence data, developed by Hochreiter and Schmidhuber (1997). As a high-performance deep learning model, LSTM has been widely used in hydrocarbon production dynamics forecasting in recent years (Sagheer and Kotb, 2019; Liu et al., 2020; Song et al., 2020). A standard RNN only has a simple tanh activation layer inside to process information (Lu et al., 2021). However, during learning and training, gradients often vanish during the update of network weight parameters, making it difficult to capture the long-term dependencies between sequence data. Unlike standard RNNs, LSTM introduces more complex gate structures to selectively store and update important temporal information, thus effectively alleviating the problem of gradient vanishing (Liang et al., 2023). A large number of studies have shown that production dynamic prediction models based on LSTM networks not only can avoid complex physical modeling processes, but also have strong fitting and learning abilities, fast convergence speed, and high prediction accuracy when dealing with complex nonlinear regression problems (Sagheer and Kotb, 2019; Liu et al., 2020; Song et al., 2020).

As shown in Fig. 1, the LSTM network is composed of multiple LSTM cells connected horizontally, and each LSTM cell mainly controls the update of the cell state $\boldsymbol{c}_t$ and the transmission and output of the hidden state $\boldsymbol{h}_t$ through three gate control units: input gate, forget gate, and output gate. The calculation process of parameters in time t of an LSTM cell can be expressed in mathematical formulas as follows:

(1) Calculation process of the forgetting gate:

$$\boldsymbol{f}_t = \sigma\left(\boldsymbol{W}_{xf}\boldsymbol{x}_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f\right) \tag{1}$$

(2) Calculation process of the input gate and the candidate cell state $\tilde{\boldsymbol{c}}_t$:

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_{xi}\boldsymbol{x}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_i) \tag{2}$$

$$\tilde{\boldsymbol{c}}_t = \tanh(\boldsymbol{W}_{x\tilde{c}}\boldsymbol{x}_t + \boldsymbol{W}_{h\tilde{c}}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{\tilde{c}}) \tag{3}$$

(3) Renewal process of cell state $\boldsymbol{c}_t$:

$$\boldsymbol{c}_t = \boldsymbol{f}_t * \boldsymbol{c}_{t-1} + \boldsymbol{i}_t * \tilde{\boldsymbol{c}}_t \tag{4}$$

(4) Calculation process of the output gate and the hidden state $\boldsymbol{h}_t$:

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_{xo}\boldsymbol{x}_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o) \tag{5}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t * \tanh(\boldsymbol{c}_t) \tag{6}$$

where $\boldsymbol{x}_t$ denotes the input at the time t; $\boldsymbol{f}_t$, $\boldsymbol{i}_t$ and $\boldsymbol{o}_t$ are the output values obtained after activation by the sigmoid function σ in the

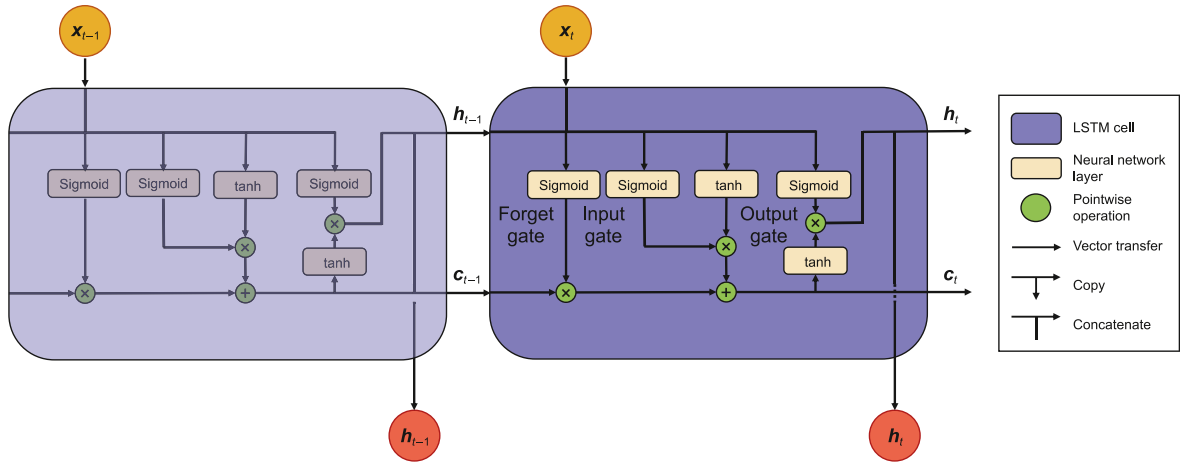


Fig. 1. LSTM network structure diagram.

forget gate, input gate, and output gate respectively; $\boldsymbol{W}_{xf}$, $\boldsymbol{W}_{hf}$, $\boldsymbol{W}_{xi}$, $\boldsymbol{W}_{hi}$, $\boldsymbol{W}_{x\tilde{c}}$, $\boldsymbol{W}_{h\tilde{c}}$, $\boldsymbol{W}_{xo}$ and $\boldsymbol{W}_{ho}$ are the weight coefficient matrices; $\boldsymbol{b}_f$, $\boldsymbol{b}_i$, $\boldsymbol{b}_{\tilde{c}}$ and $\boldsymbol{b}_o$ are the corresponding bias vectors.

2.2. Bi-directional long short-term memory model

The BiLSTM network is a novel data-driven model developed from the traditional LSTM network, and is more proficient in time series regression analysis. This model focuses more on extracting past and future information in the input sequence to more accurately capture the temporal dependencies of data information (Peng et al., 2021). While keeping the internal unit structure of the original long and short-term memory unchanged, the BiLSTM network adjusts the network structure on the hidden layer to a forward LSTM layer and a backward LSTM layer, whose network structure is shown in Fig. 2. The forward LSTM layer updates the forward hidden state $\boldsymbol{h}_t$ iteratively in time sequence from front to back, while the backward LSTM layer updates the backward hidden state $\boldsymbol{h}_t'$ at time t in the opposite order (Peng et al., 2021; Kulshrestha et al., 2020; Singla et al., 2022). It is this clever adjustment and modification that allows the output results of the BiLSTM network to not only be influenced by important feature information over a long time range but also to consider information between the past and future (Kulshrestha et al., 2020; Singla et al., 2022). The computational and updating process of information in this network can be represented through mathematical expressions, as follows:

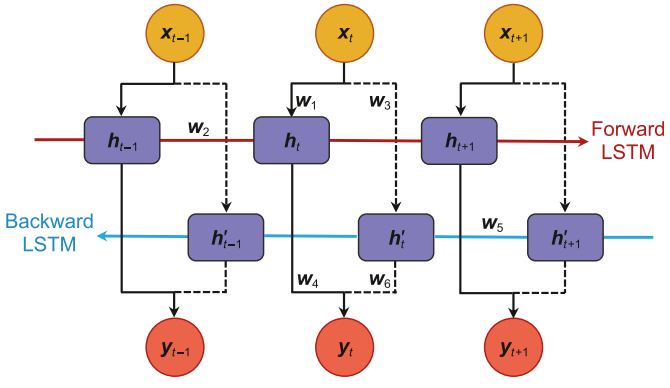


Fig. 2. BiLSTM network structure diagram.

$$\boldsymbol{h}_t = \overrightarrow{\mathrm{LSTM}}\left(\boldsymbol{W}_1\boldsymbol{x}_t + \boldsymbol{W}_2\boldsymbol{h}_{t-1} + \boldsymbol{b}\right) \quad t \in [1, T] \tag{7}$$

$$\boldsymbol{h}_t' = \overleftarrow{\mathrm{LSTM}}\left(\boldsymbol{W}_3\boldsymbol{x}_t + \boldsymbol{W}_5\boldsymbol{h}_{t+1}' + \boldsymbol{b}'\right) \quad t \in [T, 1] \tag{8}$$

$$\boldsymbol{y}_t = \boldsymbol{W}_4\boldsymbol{h}_t + \boldsymbol{W}_6\boldsymbol{h}_t' + \boldsymbol{b}_y \tag{9}$$

where $\overrightarrow{\mathrm{LSTM}}(\cdot)$ and $\overleftarrow{\mathrm{LSTM}}(\cdot)$ represent the trained forward LSTM and backward LSTM, respectively (Sun et al., 2019); $\boldsymbol{W}_i\,(i = 1,2,\ldots,6)$ is the weighting coefficient matrix; $\boldsymbol{b}$, $\boldsymbol{b}'$ and $\boldsymbol{b}_y$ are the corresponding bias vectors; T is the length of the input sequence.

2.3. Random forest model

Random forest is an ensemble learning model based on the combination of decision trees and bagging, which effectively transforms the decision tree from a weak to a strong learner (Liang et al., 2023; Breiman, 1996, 2001). A single decision tree algorithm is prone to overfitting during supervised learning, which results in a poor generalization capability of the decision tree on the testing set. Unlike the decision tree model, the RF model generates multiple training sample subsets using bootstrapping during the training and learning process, and employs a strategy of randomly selecting feature subsets during the decision tree node splitting process (Ho, 1995; Svetnik et al., 2003). The weak correlation between multiple decision trees trained by the above method allows the RF model to have low variance in the final output result calculated based on majority voting or averaging (Breiman, 2001; Wang and Chen, 2019). In addition, the ensemble learning algorithm is also used to study concept drift learning from non-stationary distributions (Zhang and Ma, 2012; Mejri et al., 2013). Currently, several improved RF algorithms for non-stationary data have been proposed, but there is no widely accepted and used version of the improved algorithms (Zhukov et al., 2017; Zhong et al., 2020, 2021). Zhukov et al. (2017) proposed a method based on random forest algorithm to deal with concept drift in wind power generation time series prediction. By adopting the weighted majority voting integrated aggregation rule, the model can get better results when dealing with progressive concept drift. Zhong et al. (2021) designed an online random forest regression model with adaptive memory activation mechanism to identify hidden changes when concept drift occurs in non-stationary data streams. Numerical experiments show that the prediction accuracy of the

model on multiple real data sets has been significantly improved.

2.4. Marine predator optimization algorithm

MPA is a relatively novel metaheuristic global optimization algorithm proposed in 2020 by Afshin Faramarzi and others, inspired by the foraging strategies of marine predators in the natural world (Faramarzi et al., 2020). This algorithm belongs to a class of stochastic optimization algorithms called swarm intelligence, which also includes the particle swarm optimization (PSO) algorithm. They mainly construct mathematical models of the intelligent behavior of biological populations in nature and seek solutions to complete global searches of parameter spaces that may have optimal solutions (Halim et al., 2021). Currently, many researchers have combined metaheuristic optimization algorithms with established production prediction proxy models to solve complex hyperparameter combination optimization problems. This not only greatly improves the optimization efficiency of the proxy model, but also synchronously enhances the predictive and generalization abilities of the model (Liu et al., 2020; Song et al., 2020; Zhang et al., 2022; Han and Bian, 2018).

In the mathematical model constructed based on the interaction rules between marine predators and prey, predators and prey mainly imitate the foraging behavior and rules in the real marine environment through a combination and switching between Lévy flight and Brownian motion. Lévy flight is a foraging strategy adopted by predators when they are in areas where food is scarce. When food is abundant in the area, predators generally switch to a Brownian-type search mode (Humphries et al., 2010). Depending on the value of the ratio of the movement speed ν of prey and predators, this mathematical model divides the iterative optimization process into three stages: in the high-velocity ratio ($\nu \geq 10$), in the unit-velocity ratio ($\nu \approx 1$), and in the low-velocity ratio ($\nu = 0.1$). The predatory behavior in each stage can be described by the following mathematical expressions (Faramarzi et al., 2020; Alqaness et al., 2022; Abdel-Basset et al., 2021).

In the initial search phase ($\nu \geq 10$), the prey is distributed relatively evenly throughout the entire search space in this stage. In order to effectively explore the surrounding areas rich in food, the strategy of predators remaining stationary while prey moves in a Brownian motion manner is adopted.

$$\boldsymbol{Prey}_i = \boldsymbol{Prey}_i + P \cdot \boldsymbol{R} \otimes \boldsymbol{S}_i \quad t < \frac{1}{3} t_{\max} \tag{10}$$

$$\boldsymbol{S}_i = \boldsymbol{R}_B \otimes (\boldsymbol{Elite}_i - \boldsymbol{R}_B \otimes \boldsymbol{Prey}_i) \quad i \in [1, n] \tag{11}$$

where $\boldsymbol{Prey}_i$ and $\boldsymbol{Elite}_i$ are the i-th vector in the predator matrix and prey matrix, respectively. P is a constant (=0.5); $\boldsymbol{R}$ is a set of random vectors generated from [0,1]; $\otimes$ denotes multiplication by term; $\boldsymbol{S}_i$ is the calculated moving step vector; $\boldsymbol{R}_B$ is a random vector generated based on Brownian motion; n is the population size, while t and $t_{\max}$ respectively refer to the current iteration number and the maximum iteration number.

In the mid-search phase ($\nu \approx 1$), predators begin to search for prey using a Brownian foraging strategy. After the initial foraging stage, food in the vicinity becomes scarce, so the prey switches its foraging strategy to Lévy flight mode, allowing it to search for food over a greater distance. In this stage, we use the first half of the population to simulate the movement strategy of the prey, and the second half to simulate the movement strategy of the predators.

$$\boldsymbol{Prey}_i = \boldsymbol{Prey}_i + P \cdot \boldsymbol{R} \otimes \boldsymbol{S}_i \quad \frac{1}{3} t_{\max} < t < \frac{2}{3} t_{\max} \tag{12}$$

$$\boldsymbol{S}_i = \boldsymbol{R}_L \otimes (\boldsymbol{Elite}_i - \boldsymbol{R}_L \otimes \boldsymbol{Prey}_i) \quad i \in [1, n/2] \tag{13}$$

where R_L is a random vector generated based on Lévy flights.

$$\boldsymbol{Prey}_i = \boldsymbol{Elite}_i + P \cdot CF \otimes \boldsymbol{S}_i \quad \frac{1}{3} t_{\max} < t < \frac{2}{3} t_{\max} \tag{14}$$

$$\boldsymbol{S}_i = \boldsymbol{R}_B \otimes (\boldsymbol{R}_B \otimes \boldsymbol{Elite}_i - \boldsymbol{Prey}_i) \quad i \in [n/2, n] \tag{15}$$

$$CF = \left(1 - \frac{t}{t_{\max}}\right)^{\left(\frac{2t}{t_{\max}}\right)} \tag{16}$$

where CF is the adaptive defined convergence factor of the predator movement step at time t.

In the post-search phase ($\nu = 0.1$), the main strategy is to switch the predator's foraging strategy to Lévy flight in order to efficiently complete the search for prey in a certain area.

$$\boldsymbol{Prey}_i = \boldsymbol{Elite}_i + P \cdot CF \otimes \boldsymbol{S}_i \quad t > \frac{2}{3} t_{\max} \tag{17}$$

$$\boldsymbol{S}_i = \boldsymbol{R}_L \otimes (\boldsymbol{R}_L \otimes \boldsymbol{Elite}_i - \boldsymbol{Prey}_i) \quad i \in [1, n] \tag{18}$$

2.5. BiLSTM-RF-MPA combination model

Considering the complex nonlinear relationships and non-stationary characteristics in the real shale gas production time series, we propose a hybrid proxy model that combines BiLSTM, RF, and MPA based on three considerations. Firstly, while BiLSTM neural networks are good at capturing long-term dependencies between inputs and outputs, they tend to overfit when faced with input data with complex nonlinear features (Hammami et al., 2020; Xiao et al., 2019; Fortmann-Roe, 2012). In other words, although recurrent neural networks improve the fitting ability of complex data by increasing the complexity of the model, they become more sensitive to disturbances in input data, leading to increased model variance. In addition, as the distribution of data in non-stationary time series drifts with time, most deep neural network prediction models based on this experience significant differences in performance between training and testing stages (Kuznetsov and Mohri, 2020; Arik et al., 2022; Montero-Manso and Hyndman, 2021). This supervised learning process in a non-stationary environment exacerbates the overfitting phenomenon of recurrent neural network models when faced with complex nonlinear input data (Naug et al., 2022). On the other hand, RF is a strong learning model that is robust and representative of ensemble learning. With the ensemble learning strategy of Bagging and random feature subspaces, RF not only reduces the model's variance but also has a certain ability to solve the concept drift problem in non-stationary time series prediction (Zhukov et al., 2017; Genuer, 2012). Therefore, based on the strengths and weaknesses of BiLSTM and RF in complex non-stationary time series prediction problems, we use data preprocessing operations and the integration of these two strong heterogeneous learning models to achieve relatively more accurate prediction performance of shale gas ultra-short-term dynamic production. In addition to building the BiLSTM-RF shale gas production prediction hybrid model, we also introduce MPA to combine and optimize the hyperparameters of the proxy model to

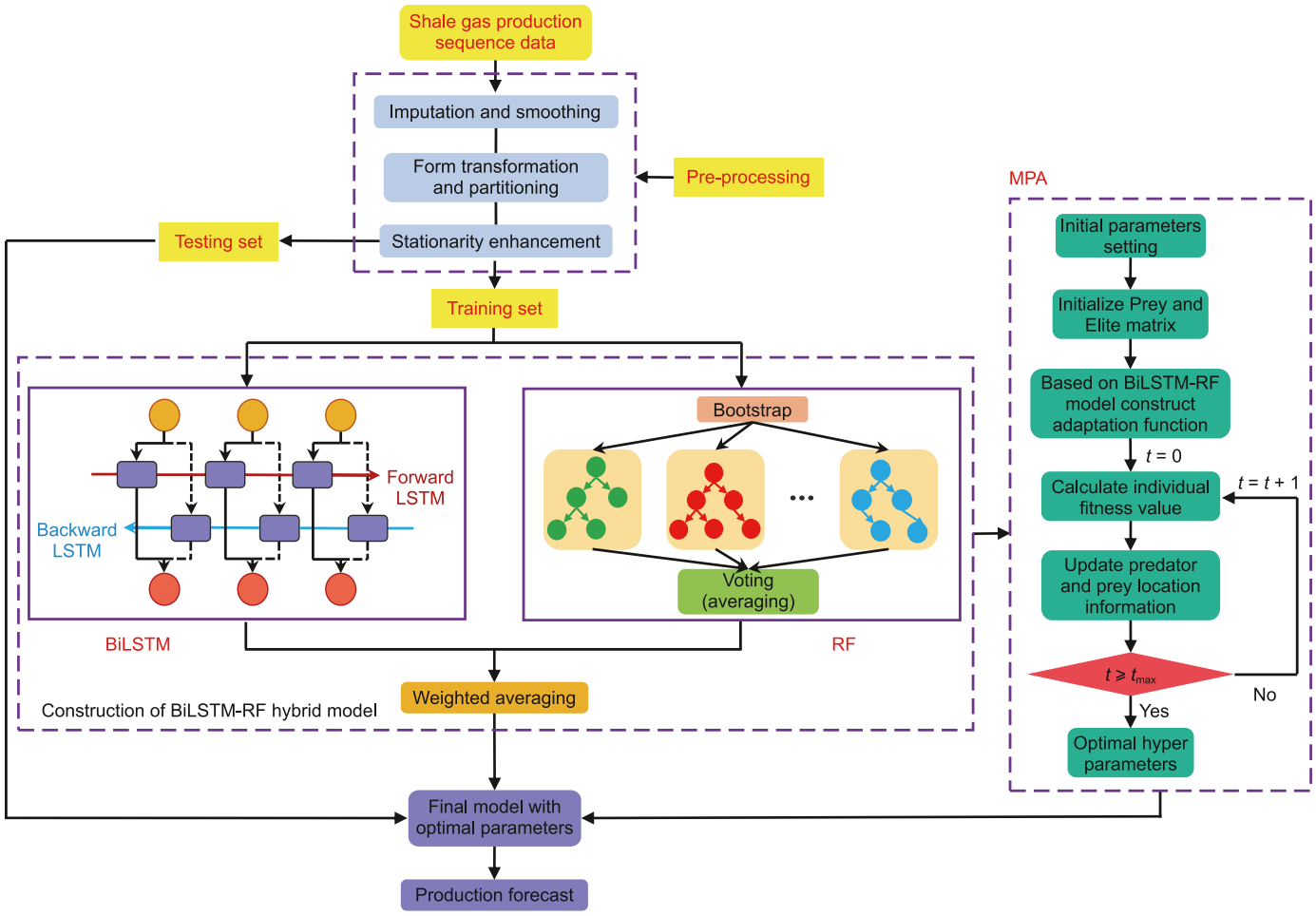


Fig. 3. Structure diagram of BiLSTM-RF-MPA model.

improve and enhance the model's generalization ability. The overall structure and prediction process of the proxy model is shown in Fig. 3.

3. Results and discussion

3.1. Data pre-processing

In this study, we selected the daily production dynamic data of a single well from the Zhejiang shale gas reservoir, consisting of 1297 days (from February 19, 2019 to September 28, 2022) as the input data for the proxy model. The data recorded the process of the gas well from starting production to gradually approaching a stable phase of production. The data contains multiple abnormal zero values and noise data. Moreover, during the early and middle stages of the production decline, the fluctuation of production change is significant, exhibiting highly non-linear characteristics. In addition, because the data is derived from the actual production development process, the sequence inevitably contains some non-stationary characteristics, and the change in production will display certain natural trends, with statistical properties and distributions changing over time (Sagheer and Kotb, 2019; Liu et al., 2022). Therefore, facing the various adverse factors mentioned above, we took the following preprocessing operations on the original data, which can significantly improve the quality of input data and enable the model to efficiently and stably complete learning and prediction tasks.

Step 1 Data imputation and smoothing

In the production process of gas wells, due to special operations such as temporary well shut-in, the recorded shale gas production for the day may be 0. These abnormal 0 values may cause significant interference with the training and fitting of data-driven models (Rathnayake et al., 2022). Therefore, we use piecewise cubic Hermite interpolating polynomial (PCHIP) to interpolate and fill in the positions where 0 values exist in the data. In addition, we also use a three-point moving average filter to smooth the noise contained in the original production data. This method mainly generates the smoothed sequence by sequentially taking the mean of the continuous three data items in the original sequence, which can significantly reduce the high-frequency noise in the sequence data. The processing procedure is shown in Fig. 4.

Step 2 Data form transformation and partitioning

To adapt the production time series data processed in the previous step to the form suitable for self-supervised learning tasks and enable the proxy model to better cope with various drift phenomena in non-stationary environments, we use the sliding window method to reorganize the data (Arik et al., 2022; Hammami et al., 2020). Specifically, we select the historical production data of the previous m days (where m is the length of the sliding window) to predict the production value of the next time step and use the actual production value of the next time step as the

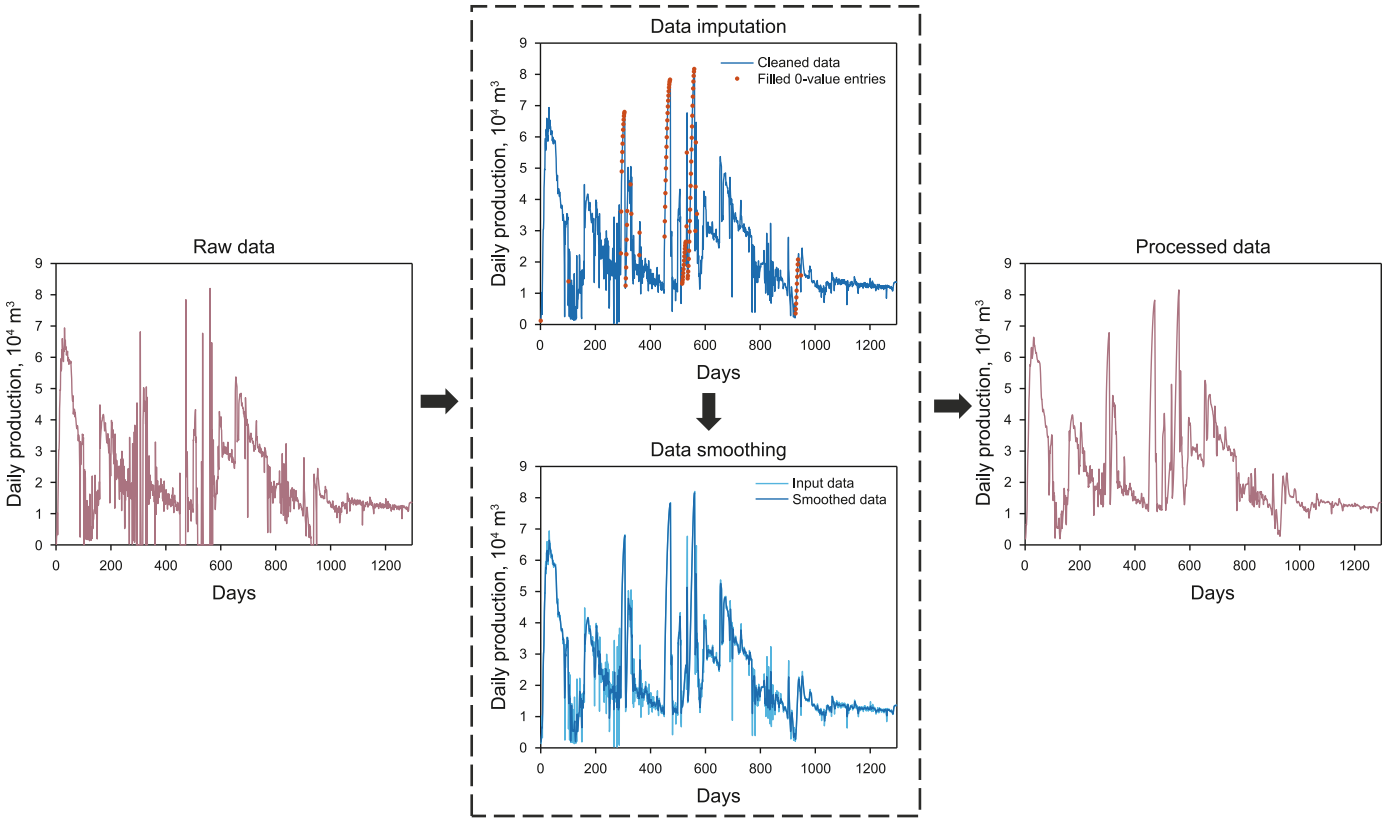


Fig. 4. Filling and smoothing of shale gas production data.

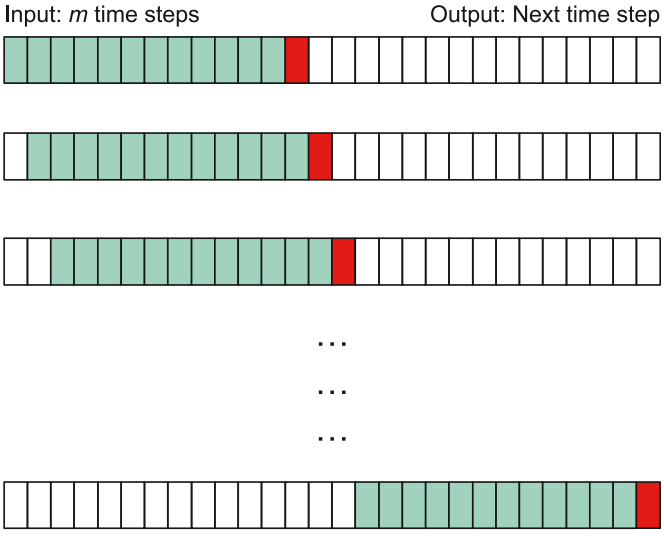


Fig. 5. Window sliding diagram.

label value (the window sliding strategy is shown in Fig. 5). It is worth noting that when the time step of inputting shale gas production data is week, month or year, the model will still be applicable. At the same time, we adopt the strategy of constantly sliding the window forward to construct the input and output data sets of the model. After multiple trial-and-error processes, we set the window length to 12. In addition, we divide the production sequence data into training and test sets in an 8:2 ratio. 5% of the training data is split into a validation set for hyperparameter optimization of the proxy model.

Step 3 Enhancement of data stationarity

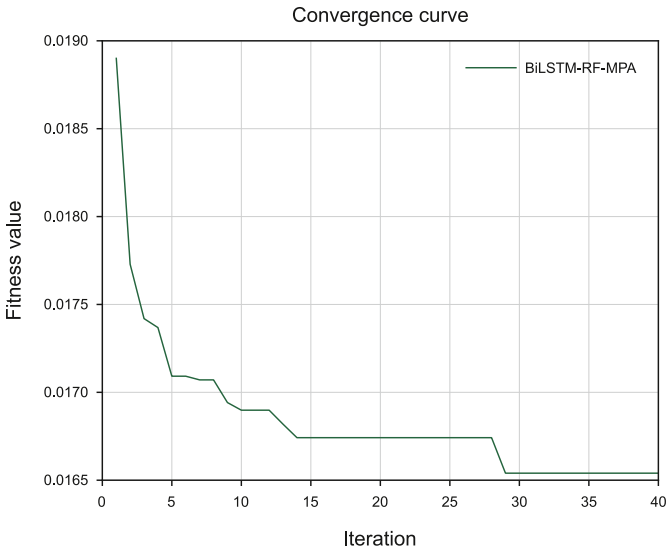
Given that direct prediction of non-stationary time series remains a very challenging problem, preprocessing operations can be taken on the input data to provide a more stable data distribution for the proxy model. In previous studies, sequence decomposition methods (such as wavelet transform, empirical mode decomposition, etc.) or detrending operations have typically been used to obtain more stationary production time series data (Liu et al., 2012, 2020; Aranguren et al., 2022). However, traditional methods of enhancing data stationarity tend to strip almost all non-stationarities from time series data, leading to excessively high levels of data stationarity and a drastic reduction in information content. This, in turn, makes it difficult for models to effectively distinguish and capture time-dependent changes during training, thereby limiting the models' ability to predict and provide guidance on real-world events (Liu et al., 2022). To avoid the phenomenon of over-stabilization, we use Z-score standardization to transform the data into a distribution with a mean of 0 and a standard deviation of 1, thereby attenuating the non-stationarity of the input data. The following algorithm is used to implement Z-score standardization:

$$y = (x - \mu_x) \odot \frac{\sigma_y}{\sigma_x} + \mu_y \tag{19}$$

where $\mu_y = 0$, $\sigma_y = 1$; x and y are the data to be standardized and the data after standardization; μ_x and σ_x are the mean and variance of each row in the input matrix x; $\odot$ is the element product operator.

3.2. Model parameter settings

In the proposed BiLSTM-RF-MPA hybrid model, we combine the



Parameters	Value
Optimal fitness value	0.01654
Number of hidden units	109
Initial learning rate	0.0049742
Min leaf size	5
Number of regression trees	199

Fig. 6. Fitness convergence curve and parameter optimization results.

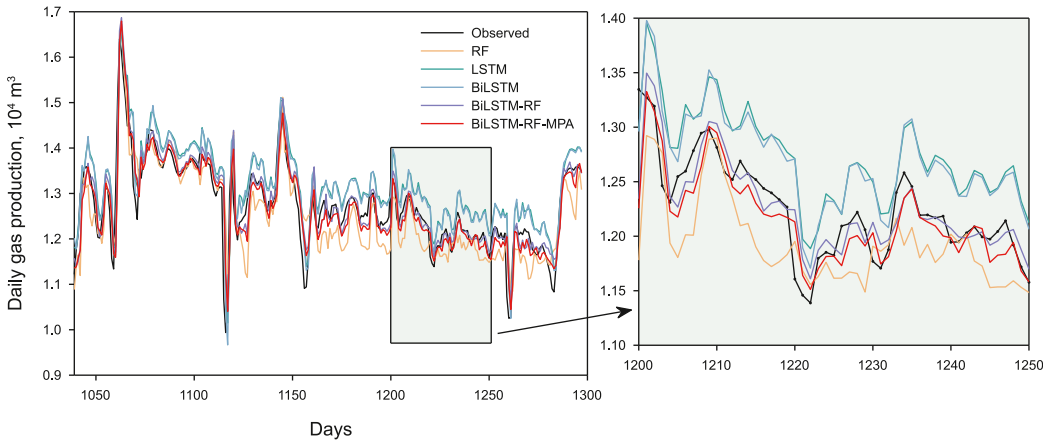


Fig. 7. Comparison of daily shale gas production ahead of one-step forecast results.

prediction results of BiLSTM and RF models using a weighted average with the weight coefficient set to 0.5 for each model as an example. BiLSTM is a 5-layer deep neural network structure model consisting of an input layer, a bidirectional LSTM layer, a dropout layer, a dense layer, and an output layer. We set the number of features in the input layer and the size of the dense layer to 1, and the maximum training epoch number and mini-batch size to 250 and 500, respectively. To prevent overfitting of the BiLSTM model due to its complexity, we introduce the dropout regularization technique to break the cooperative effect between neurons caused by backpropagation learning in the neural network model (Srivastava et al., 2014; Krizhevsky et al., 2017). To achieve the strongest regularization effect of the dropout technique, we set the dropout rate to 0.5 (Ba and Frey, 2013). In addition, the adaptive moment estimation (Adam) optimizer is used to adjust the weight values between each neuron in the BiLSTM neural network.

In addition to the design of the hybrid model structure and hyperparameters mentioned above, we also adopt MPA to globally search and optimize the number of hidden units and initial learning rate for BiLSTM, as well as the minimum leaf node number and number of regression trees for RF. The upper and lower bounds of the 4-dimensional search space, which are formed sequentially by these four important hyperparameters, are set to [20, 0.001, 5, 20] and [200, 0.01, 50, 200], respectively. For the initial parameters of MPA, we manually tune the number of search proxies and the maximum iteration number t_{max} to 20 and 40, respectively, to reduce the computational time cost during hyperparameter optimization. The optimization objective of the algorithm is to find the minimum mean square error on the fitness function constructed based on the BiLSTM-RF model for the training and validation sets, shown as follows:

$$Obj_{\text{fitness}} = \min\left(MSE_{\text{training}} + MSE_{\text{validation}}\right) \tag{20}$$

where MSE_{training} and $MSE_{\text{validation}}$ are the mean square error of training and verification respectively; Obj_{fitness} is the target fitness value.

In order to highlight the effectiveness of the BiLSTM-RF-MPA hybrid proxy model proposed in this work for ultra-short-term dynamic prediction of non-stationary shale gas production sequences, we compare it with models constructed based on LSTM, BiLSTM, RF, and BiLSTM-RF methods.

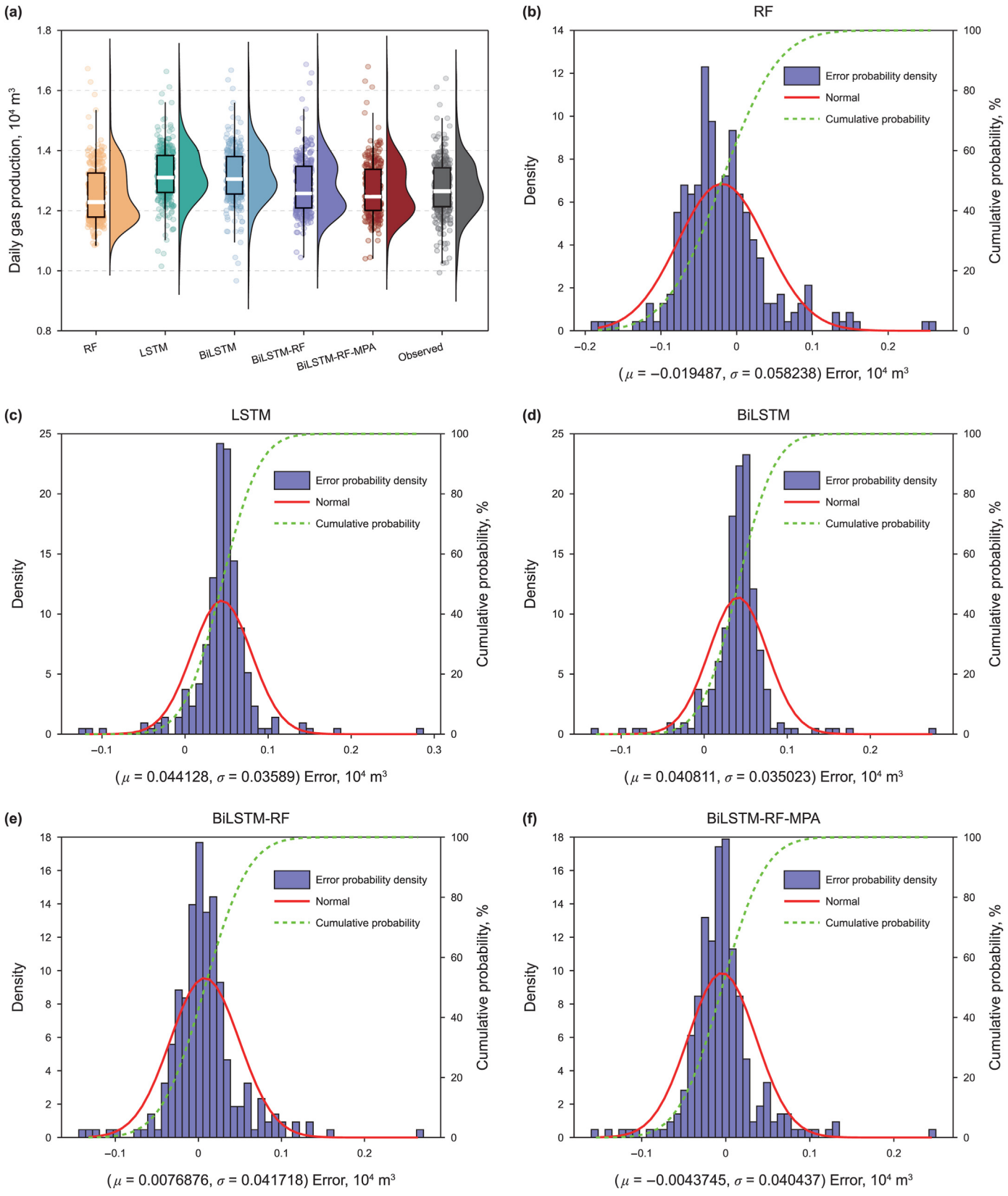


Fig. 8. **(a)** Shows the raincloud plot of the one step ahead prediction results of the five models, namely, RF, LSTM, BiLSTM, BiLSTM-RF, and BiLSTM-RF-MPA. **(b)**–**(f)** are used to evaluate the cumulative probability and probability density distribution of the prediction errors of the above five models. The red solid line represents the probability density curve fitted using a normal distribution on the probability density histogram. The mean and standard deviation of the prediction errors of each model are represented by μ and σ, respectively.

3.3. Performance metrics

To evaluate the performance differences between the BiLSTM-RF-MPA and all the other models used for comparison, we used five performance indicators, namely, the coefficient of determination (R^2), root mean square error (*RMSE*), mean absolute error (*MAE*), mean absolute percentage error (*MAPE*), and Pearson correlation coefficient (r) to assess the predictive results of the models (Yang et al., 2023). Additionally, we also evaluated the p-value corresponding to r and only when $p < 0.05$, the calculated r value is statistically significant (Xiao et al., 2019).The calculation formulas of R^2, RMSE, MAE, MAPE and r are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2} \tag{21}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \tag{22}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - O_i| \tag{23}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - O_i}{O_i}\right| \times 100\% \tag{24}$$

$$r = \frac{\sum_{i=1}^{N}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}\sqrt{\sum_{1}^{N}(P_i - \overline{P})^2}} \tag{25}$$

where O_i is the observed shale gas production value, while P_i is the predicted shale gas production value; $\overline{O}$ and $\overline{P}$ are the average of O and P; N is the number of samples in the test set.

3.4. Result discussions

Taking the dynamic daily production data of a shale gas reservoir in Zhejiang as an example, we calculated and compared the learning and prediction performance of the proposed proxy model when facing complex production sequence input data. Based on the MPA global optimization algorithm, we obtained the fitness iteration convergence curve of the objective function and gave the optimization results of four key hyperparameters, including the number of hidden units, initial learning rate, minimum leaf node number, and regression tree number, as well as the optimal fitness value (as shown in Fig. 6). Then, based on the optimized hyperparameter values, we set up and used the BiLSTM-RF model to predict the daily shale gas production for one-step ahead for the sample data in the test set.

Fig. 7 shows the shale gas daily production prediction results for the five models RF, LSTM, BiLSTM, BiLSTM-RF, and BiLSTM-RF-MPA one step ahead, and their comparison with the actual observations. It can be found that the RF model can accurately capture the production trend in the early prediction, but it produced a significant underestimation in the middle and late stages, especially at the trough. The predictions of the LSTM and BiLSTM models, on the other hand, were generally overestimated and became more severe with time. In comparison, the BiLSTM-RF and BiLSTM-RF-MPA hybrid models performed better in completing the prediction task. Their predicted values were much closer to the observed values than those of the RF, LSTM, and BiLSTM models, which can be seen more intuitively in Fig. 8(a). Fig. 8(b)–(f) describes the prediction performance of the above five models from the perspective of error distribution. After comprehensive comparison, it is found that the probability density distribution of RF prediction errors was relatively scattered and the cumulative probability density function curve increased more slowly. In contrast, LSTM and BiLSTM showed the opposite trend, with the probability density distribution of their prediction errors displaying a higher and more concentrated trend from a holistic perspective. Additionally, compared with RF, LSTM, and BiLSTM, the mean prediction errors of BiLSTM-RF and BiLSTM-RF-MPA were closer to 0, and their distribution was denser near 0, indicating that their distribution was closer to the standard normal distribution.

To further accurately evaluate the prediction performance of each model, Table 1 summarizes the quantitative results of the R^2, *RMSE*, *MAE*, *MAPE*, and r performance indicators (including the p-value corresponding to r). In addition, a radar chart was used for more intuitive comparison and display, as shown in Fig. 9. Based on Table 1 and Fig. 9, we can clearly notice that the BiLSTM-RF model achieved significant improvements in the R^2, *RMSE*, *MAE*, and *MAPE* performance indicators compared to the RF, LSTM, and BiLSTM models. Exceptionally, although the r value of BiLSTM-RF showed extremely strong correlation, it was slightly lower than that of LSTM and BiLSTM. This situation is similar to the comparison results of the standard deviation σ of the prediction errors of each model calculated in Fig. 8(b)–(f). By analyzing the p-value corresponding to r, it can be found that the p-values of all models were 0, which means that the r values of all models' prediction results have statistical significance. Based on the above analysis, we believe that the occurrence of this exception is due to the fact that this hybrid model sacrifices some model bias to enhance the generalization ability of the model when balancing the variance and bias of the

Table 1
The quantitative results of the prediction performance index of the five models in one time step ahead of time are counted.

Models	R^2	*MAPE*, %	*MAE*	*RMSE*	r	p-value
RF	0.5775	3.73	0.0465	0.0613	0.8151	0
LSTM	0.6368	3.92	0.0492	0.0568	0.9260	0
BiLSTM	0.6754	3.64	0.0456	0.0537	**0.9308**	0
BiLSTM-RF	0.7984	2.20	0.0274	0.0423	0.9037	0
BiLSTM-RF-MPA	**0.8147**	**2.13**	**0.0267**	**0.0406**	0.9080	0

Note: Bold values indicate the calculated optimal performance metrics values.

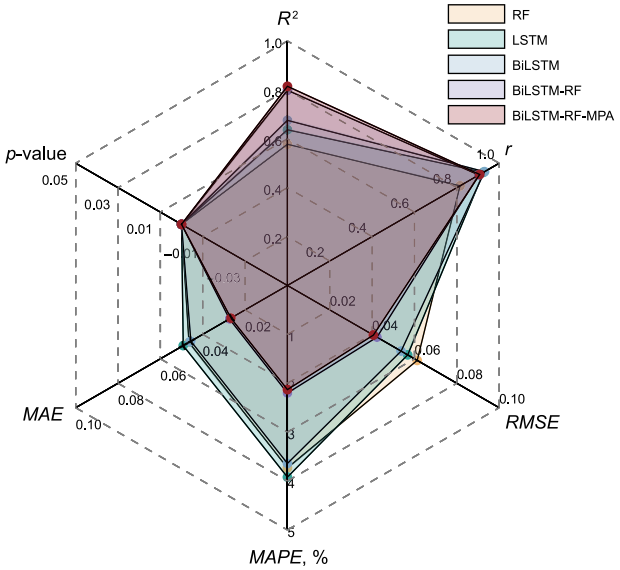


Fig. 9. Comparison and display of quantitative results of prediction error performance indicators of different methods.

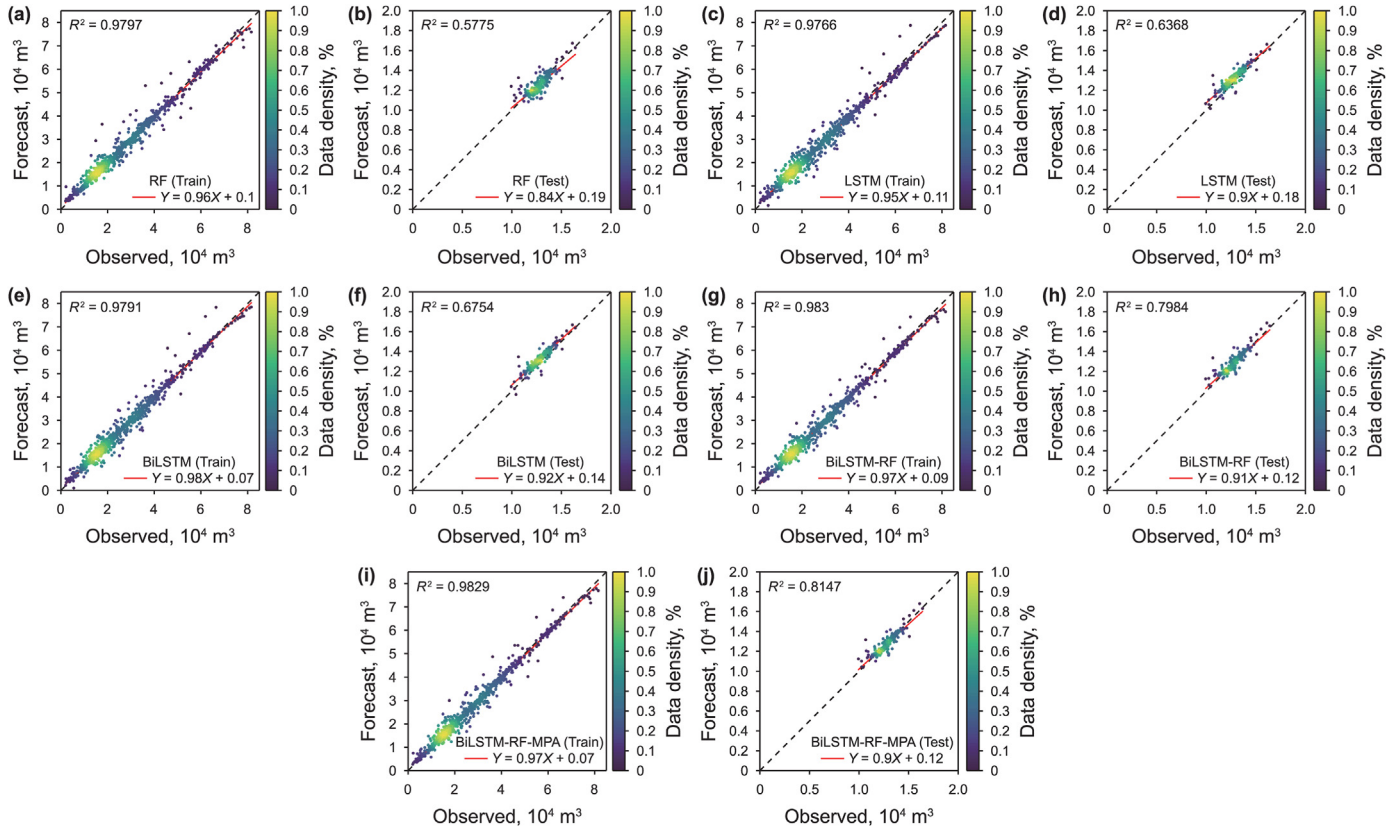


Fig. 10. Five methods, RF ((**a**), (**b**)), LSTM ((**c**), (**d**)), BiLSTM ((**e**), (**f**)), BiLSTM-RF ((**g**), (**h**)) and BiLSTM-RF-MPA ((**i**), (**j**)), were used to plot the density scatter plots of the one step ahead prediction results on the training and test sets, respectively. The color of the data points in the graph changes from dark blue to light green and then to bright yellow, reflecting the gradual increase in the density of the data points. The slope of the black dotted line in the subgraph is 1 and the intercept on the *y*-axis is 0, indicating that the predicted value is equal to the observed value, while the red solid line represents the linear regression result of all the scatters in the subgraph.

Table 2
Statistics of production multi-step ahead prediction results based on BiLSTM-RF-MPA model.

Models	2-day			3-day		
	MAE	*RMSE*	*MAPE*, %	*MAE*	*RMSE*	*MAPE*, %
RF	0.0677	0.0921	5.44	0.1041	0.1462	8.31
LSTM	0.0844	0.0976	6.75	0.1049	0.1241	8.41
BiLSTM	0.0891	0.1017	7.11	0.1185	0.1365	9.49
BiLSTM-RF	0.0486	0.0716	3.93	0.0904	0.1200	7.31
BiLSTM-RF-MPA	**0.0476**	**0.0698**	**3.83**	**0.0829**	**0.1153**	**6.67**

Note: Bold values indicate the calculated optimal performance metrics values.

model. Moreover, it is worth noting that the performance of the BiLSTM-RF model after MPA optimization has been improved in all performance indicators, especially the coefficient of determination R^2, which exceeded 0.8.

To test the generalization ability of the model proposed in this paper, we drew the density scatter plots of the results predicted one step ahead of each model on the training set and the test set, and carried out linear fitting on them respectively, as shown in Fig. 10. By comprehensively comparing the scatter density distribution and the fitting regression performance, we observed that the BiLSTM-RF and BiLSTM-RF-MPA models showed better generalization ability. The scatter plots of these two models on the training and testing sets converged along the black dashed line, and the fitted lines were closest. Additionally, it is found that compared with RF, LSTM, and BiLSTM, the above two hybrid models also had improved R^2 values on the training set. Furthermore, the difference in R^2 values between the training and testing sets was the smallest for the BiLSTM-RF-MPA model. This fully demonstrates that the proposed model not only improves its fitting effect but also has a strong generalization ability.

Overall, we analyzed the prediction effect of five models, RF, LSTM, BiLSTM, BILSTM-RF and BILSTM-RF-MPA, when the production was forecasted one step ahead from a variety of different perspectives. Exploratory data analysis methods such as probability density histograms, radar charts, and density scatter plots were used to compare and evaluate the performance of different models. Based on the comprehensive analysis results, we found that the BiLSTM-RF-MPA model can achieve higher prediction accuracy and stronger generalization ability by better handling the complex nonlinear and non-stationary properties of shale gas production time series. Therefore, we also realized the production prediction two time steps ahead and three time steps ahead based on BiLSTM-RF-MPA model for the shale gas production data of this well. As shown in Table 2, compared with other models, this model still obtained good prediction performance in terms of the MAE, RMSE, and MAPE performance indicators. This further highlights the excellent generalization ability of the proposed hybrid model in increasing the ultra-short-term shale gas production prediction range.

4. Conclusions

This work mainly focuses on the ultra-short-term dynamic prediction of shale gas production and develops and evaluates new models for this task. To address the complex nonlinear and non-

stationary properties of actual production time series, we combined the unique prediction performance advantages of two strong heterogeneous learning models BiLSTM and RF to construct a hybrid proxy model for ultra-short-term prediction of shale gas production. To improve the relationship between the accuracy and consistency of the proxy model, we used a metaheuristic global optimization algorithm called MPA to perform combination optimization of multiple hyperparameters on the training and validation sets. This allowed the hybrid model to have minimum generalization error by balancing the bias and variance between models.

To verify the performance of the surrogate model proposed in this paper, we use real gas well production data to compare and evaluate the production prediction of one time step in advance from multiple perspectives, and also realize the prediction of 2 and 3 time steps in advance. The model will still be applicable when the time steps for input shale gas production data are weeks, months, or years. The case studies showed that the model not only captured the nonlinear features of the production time series better but also adapted well to non-stationary environments and had strong fitting and generalization abilities. From the results, this study achieved some effectiveness in trying to solve the prediction problem of non-stationary production time series by introducing a sliding window strategy, Z-score standardization, and constructing a hybrid proxy model by combining RF and BiLSTM.

Future work can emphasize on dealing with more complex concept drift phenomena (such as simultaneous, gradual, and repetitive drifts), the model's active adaptation ability to non-stationary environments (including the self-adaptivity of model structure and parameters) can be enhanced by using dynamically sized window management strategies (Hammami et al., 2020). Additionally, in practical applications, the proposed model can be more quickly and conveniently applied to the production prediction of other shale gas wells by leveraging model/parameter-based transfer learning techniques (Zhuang et al., 2020; Odi et al., 2021; Yin et al., 2020).

CRediT authorship contribution statement

Bin Liang: Writing – original draft, Methodology. **Jiang Liu:** Software, Investigation. **Li-Xia Kang:** Resources, Data curation. **Ke Jiang:** Writing – review & editing. **Jun-Yu You:** Visualization, Formal analysis. **Hoonyoung Jeong:** Writing – review & editing, Conceptualization. **Zhan Meng:** Validation, Project administration.

Declaration of competing interest

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

Acknowledgements

This work is supported by Sichuan Natural Science Foundation (Grant No. 2023NSFSC0423), CNPC Innovation Found (Grant No. 2022DQ02-0207) and Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202201510). Hoonyoung Jeong was supported by a grant from the Human Resources Development program (No. 20216110100070) of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), funded by the Ministry of Trade, Industry, and Energy of the Korean Government.

References

Abad, A., Ghorbani, H., Mohamadian, N., et al., 2022. Robust hybrid machine learning algorithms for gas flow rates prediction through wellhead chokes in gas condensate fields. Fuel 308, 121872. https://doi.org/10.1016/j.fuel.2021.121872.

Abdel-Basset, M., El-Shahat, D., Chakrabortty, R., et al., 2021. Parameter estimation of photovoltaic models using an improved marine predators algorithm. Energy Convers. Manag. 227, 113491. https://doi.org/10.1016/j.enconman.2020.113491.

Al-qaness, M., Ewees, A., Fan, H., et al., 2022. Boosted ANFIS model using augmented marine predator algorithm with mutation operators for wind power forecasting. Appl. Energy 314, 118851. https://doi.org/10.1016/j.apenergy.2022.118851.

Aranguren, C., Fragoso, A., Aguilera, R., 2022. Sequence-to-sequence (Seq2Seq) long short-term memory (LSTM) for oil production forecast of shale reservoirs. In: Unconventional Resources Technology Conference. https://doi.org/10.15530/urtec-2022-3722179.

Arik, S., Yoder, N., Pfister, T., 2022. Self-adaptive forecasting for improved deep learning on non-stationary time-series. arXiv preprint, arXiv:2202.02403. https://doi.org/10.48550/arXiv.2202.02403.

Ba, J., Frey, B., 2013. Adaptive dropout for training deep neural networks. Adv. Neural Inf. Process. Syst. 26.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140. https://doi.org/10.1007/BF00058655.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/a:1010933404324.

Chen, X., Tang, C., Du, Z., et al., 2019. Numerical simulation on multi-stage fractured horizontal wells in shale gas reservoirs based on the finite volume method. Nat. Gas. Ind. B 6 (4), 347–356. https://doi.org/10.1016/j.ngib.2018.12.004.

Desai, J., Pandian, S., Vij, R., 2021. Big data analytics in upstream oil and gas industries for sustainable exploration and development: a review. Environ. Technol. Innov. 21, 101186. https://doi.org/10.1016/j.eti.2020.101186.

Dong, D., Wang, Y., Li, X., et al., 2016. Breakthrough and prospect of shale gas exploration and development in China. Nat. Gas. Ind. B 3 (1), 12–26. https://doi.org/10.1016/j.ngib.2016.02.002.

Faramarzi, A., Heidarinejad, M., Mirjalili, S., et al., 2020. Marine predators algorithm: a nature-inspired metaheuristic. Expert Syst. Appl. 152, 113377. https://doi.org/10.1016/j.eswa.2020.113377.

Fortmann-Roe, S., 2012. Understanding the bias-variance tradeoff. http://scott.fortmann-roe.com/docs/BiasVariance.html.

Gale, J., Laubach, S., Olson, J., et al., 2014. Natural fractures in shale: a review and new observations. AAPG Bull. 98 (11), 2165–2216. https://doi.org/10.1306/08121413151.

Genuer, R., 2012. Variance reduction in purely random forests. J. Nonparametric Statistics 24 (3), 543–562. https://doi.org/10.1080/10485252.2012.677843.

Halim, A., Ismail, I., Das, S., 2021. Performance assessment of the metaheuristic optimization algorithms: an exhaustive review. Artif. Intell. Rev. 54, 2323–2409. https://doi.org/10.1007/s10462-020-09906-6.

Hammami, Z., Sayed-Mouchaweh, M., Mouelhi, W., et al., 2020. Neural networks for online learning of non-stationary data streams: a review and application for smart grids flexibility improvement. Artif. Intell. Rev. 53, 6111–6154. https://doi.org/10.1007/s10462-020-09844-3.

Han, B., Bian, X., 2018. A hybrid PSO-SVM-based model for determination of oil recovery factor in the low-permeability reservoir. Petroleum 4 (1), 43–49. https://doi.org/10.1016/j.petlm.2017.06.001.

Ho, T., 1995. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. IEEE, pp. 278–282. https://doi.org/10.1109/ICDAR.1995.598994, 1.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Humphries, N.E., Queiroz, N., Dyer, J.R.M., et al., 2010. Environmental context explains Lévy and Brownian movement patterns of marine predators. Nature 465 (7301), 1066–1069. https://doi.org/10.1038/nature09116.

Hutahaean, J., Demyanow, V., Christie, M., 2015. Impact of model parameterisation and objective choices on assisted history matching and reservoir forecasting. In: SPE Asia Pacific Oil and Gas Conference and Exhibition. https://doi.org/10.2118/176389-MS.

Jeong, H., Sun, A., Lee, J., et al., 2018. A learning-based data-driven forecast approach for predicting future reservoir performance. Adv. Water Resour. 118, 95–109. https://doi.org/10.1016/j.advwatres.2018.05.015.

Jo, S., Jeong, H., Min, B., et al., 2022. Efficient deep-learning-based history matching for fluvial channel reservoirs. J. Petrol. Sci. Eng. 208, 109247. https://doi.org/10.1016/j.petrol.2021.109247.

Kalantari-Dahaghi, A., Mohaghegh, S., Esmaili, S., 2015. Coupling numerical simulation and machine learning to model shale gas production at different time resolutions. J. Nat. Gas Sci. Eng. 25, 380–392. https://doi.org/10.1016/j.jngse.2015.04.018.

Kim, Y., Min, B., Sun, A., et al., 2023. Novel learning-based pattern-data-driven forecast approach for predicting future well responses. Available at SSRN 4429169. https://doi.org/10.2139/ssrn.4429169.

Kocoglu, Y., Gorell, S., McElroy, P., 2021. Application of Bayesian optimized deep Bi-LSTM neural networks for production forecasting of gas wells in unconventional shale gas reservoirs. In: Unconventional Resources Technology Conference, pp. 2176–2196. https://doi.org/10.15530/urtec-2021-5418.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90. https://doi.org/10.1145/3065386.
Kulshrestha, A., Krishnaswamy, V., Sharma, M., 2020. Bayesian BiLSTM approach for tourism demand forecasting. Ann. Tourism Res. 83, 102925. https://doi.org/10.1016/j.annals.2020.102925.
Kuznetsov, V., Mohri, M., 2020. Discrepancy-based theory and algorithms for forecasting non-stationary time series. Ann. Math. Artif. Intell. 88 (4), 367–399. https://doi.org/10.1007/s10472-019-09683-1.
Li, X., Ma, X., Xiao, F., et al., 2022a. Time-series production forecasting method based on the integration of bidirectional gated recurrent unit (Bi-GRU) network and sparrow search algorithm (SSA). J. Petrol. Sci. Eng. 208, 109309. https://doi.org/10.1016/j.petrol.2021.109309.
Li, X., Ma, X., Xiao, F., et al., 2022b. Small-sample production prediction of fractured wells using multitask learning. SPE J. 27 (3), 1504–1519. https://doi.org/10.2118/209231-PA.
Li, X., Ma, X., Xiao, F., et al., 2022c. A physics-constrained long-term production prediction method for multiple fractured wells using deep learning. J. Petrol. Sci. Eng. 217, 110844. https://doi.org/10.1016/j.petrol.2022.110844.
Liang, B., Liu, J., You, J., et al., 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, H., Zhang, L., Zhao, Y., et al., 2020. Empirical methods of decline-curve analysis for shale gas reservoirs: review, evaluation, and application. J. Nat. Gas Sci. Eng. 83, 103531. https://doi.org/10.1016/j.jngse.2020.103531.
Liu, W., Liu, W.D., Gu, J., 2020. Forecasting oil production using ensemble empirical model decomposition based long short-term memory neural network. J. Petrol. Sci. Eng. 189, 107013. https://doi.org/10.1016/j.petrol.2020.107013.
Liu, Y., Ma, X., Zhang, X., et al., 2021. A deep-learning-based prediction method of the estimated ultimate recovery (EUR) of shale gas wells. Petrol. Sci. 18 (5), 1450–1464. https://doi.org/10.1016/j.petsci.2021.08.007.
Liu, Y., Wu, H., Wang, J., et al., 2022. Non-stationary transformers: exploring the stationarity in time series forecasting. Adv. Neural Inf. Process. Syst. 35, 9881–9893.
Liu, Z., Wang, Z., Wang, C., 2012. Predicting reservoir production based on wavelet analysis-neural network. Advances in Computer Science and Information Engineering 535–539. https://doi.org/10.1007/978-3-642-30126-1_84.
Lu, T., Li, Z., Lai, F., et al., 2019. Blasingame decline analysis for variable rate/variable pressure drop: a multiple fractured horizontal well case in shale gas reservoirs. J. Petrol. Sci. Eng. 178, 193–204. https://doi.org/10.1016/j.petrol.2019.03.036.
Lu, W., Li, J., Wang, J., et al., 2021. A CNN-BiLSTM-AM method for stock price prediction. Neural Comput. Appl. 33, 4741–4753. https://doi.org/10.1007/s00521-020-05532-z.
Lyu, Q., Tan, J., Li, L., et al., 2021. The role of supercritical carbon dioxide for recovery of shale gas and sequestration in gas shale reservoirs. Energy Environ. Sci. 14 (8), 4203–4227. https://doi.org/10.1039/D0EE03648J.
Mahdaviara, M., Sharifi, M., Ahmadi, M., 2022. Toward evaluation and screening of the enhanced oil recovery scenarios for low permeability reservoirs using statistical and machine learning techniques. Fuel 325, 124795. https://doi.org/10.1016/j.fuel.2022.124795.
Mejri, D., Khanchel, R., Limam, M., 2013. An ensemble method for concept drift in nonstationary environment. J. Stat. Comput. Simulat. 83 (6), 1115–1128. https://doi.org/10.1080/00949655.2011.651797.
Mohammadpoor, M., Torabi, F., 2020. Big data analytics in oil and gas industry: an emerging trend. Petroleum 6 (4), 321–328. https://doi.org/10.1016/j.petlm.2018.11.001.
Mohd, S., Cornelio, J., Jahandideh, A., et al., 2021. Integrating deep learning and physics-based models for improved production prediction in unconventional reservoirs. In: SPE Middle East Oil & Gas Show and Conference. https://doi.org/10.2118/204864-MS.
Montero-Manso, P., Hyndman, R., 2021. Principles and algorithms for forecasting groups of time series: locality and globality. Int. J. Forecast. 37 (4), 1632–1653. https://doi.org/10.1016/j.ijforecast.2021.03.004.
Naug, A., Quinones-Grueiro, M., Biswas, G., 2022. Deep reinforcement learning control for non-stationary building energy management. Energy Build. 277, 112584. https://doi.org/10.1016/j.enbuild.2022.112584.
Nwachukwu, A., Jeong, H., Pyrcz, M., et al., 2018. Fast evaluation of well placements in heterogeneous reservoir models using machine learning. J. Petrol. Sci. Eng. 163, 463–475. https://doi.org/10.1016/j.petrol.2018.01.019.
Odi, U., Ayeni, K., Alsulaiman, N., et al., 2021. Applied transfer learning for production forecasting in shale reservoirs. In: SPE Middle East Oil & Gas Show and Conference. https://doi.org/10.2118/204784-MS.
Peng, T., Zhang, C., Zhou, J., et al., 2021. An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting. Energy 221, 119887. https://doi.org/10.1016/j.energy.2021.119887.
Rathnayake, S., Rajora, A., Firouzi, M., 2022. A machine learning-based predictive model for real-time monitoring of flowing bottom-hole pressure of gas wells. Fuel 317, 123524. https://doi.org/10.1016/j.fuel.2022.123524.
Sagheer, A., Kotb, M., 2019. Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 323, 203–213. https://doi.org/10.1016/j.neucom.2018.09.082.
Shen, W., Xu, Y., Li, X., et al., 2016. Numerical simulation of gas and water flow mechanism in hydraulically fractured shale gas reservoirs. J. Nat. Gas Sci. Eng. 35, 726–735. https://doi.org/10.1016/j.jngse.2016.08.078.
Singla, P., Duhan, M., Saroha, S., 2022. An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. Earth Sci. Inform. 15 (1), 291–306. https://doi.org/10.1007/s12145-021-00723-1.
Song, X., Liu, Y., Xue, L., et al., 2020. Time-series well performance prediction based on long short-term memory (LSTM) neural network model. J. Petrol. Sci. Eng. 186, 106682. https://doi.org/10.1016/j.petrol.2019.106682.
Srivastava, N., Hinton, G., Krizhevsky, A., et al., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.
Sun, H., Ouyang, W., Zhang, M., 2017. Advanced production decline analysis and performance forecasting of gas wells based on numerical model. Acta Pet. Sin. 38 (10), 1194. https://doi.org/10.7623/syxb201710010.
Sun, H., Ouyang, W., Zhang, M., et al., 2018. Advanced production decline analysis of tight gas wells with variable fracture conductivity. Petrol. Explor. Dev. 45 (3), 472–480. https://doi.org/10.1016/S1876-3804(18)30052-1.
Sun, J., Shi, W., Yang, Z., et al., 2019. Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems. IEEE Trans. Veh. Technol. 68 (11), 10348–10356. https://doi.org/10.1109/TVT.2019.2925562.
Svetnik, V., Liaw, A., Tong, C., et al., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. 43 (6), 1947–1958. https://doi.org/10.1021/ci034160g.
U.S. EIA, 2021. Natural gas gross withdrawals and production. https://www.eia.gov/dnav/ng/ng_prod_sum_dc_NUS_mmcf_a.htm.
Vikara, D., Remson, D., Khanna, V., 2020. Machine learning-informed ensemble framework for evaluating shale gas production potential: case study in the Marcellus Shale. J. Nat. Gas Sci. Eng. 84, 103679. https://doi.org/10.1016/j.jngse.2020.103679.
Wang, H., Chen, L., Qu, Z., et al., 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, Q., Chen, X., Jha, A.N., et al., 2014. Natural gas from shale formation—the evolution, evidences and challenges of shale gas revolution in United States. Renew. Sustain. Energy Rev. 30, 1–28. https://doi.org/10.1016/j.rser.2013.08.065.
Wang, S., Chen, S., 2019. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. J. Petrol. Sci. Eng. 174, 682–695. https://doi.org/10.1016/j.petrol.2018.11.076.
Wang, S., Chen, Z., Chen, S., 2019. Applicability of deep neural networks on production forecasting in Bakken shale reservoirs. J. Petrol. Sci. Eng. 179, 112–125. https://doi.org/10.1016/j.petrol.2019.04.016.
Wang, S., Qin, C., Feng, Q., et al., 2021. 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.
Xiao, C., Chen, N., Hu, C., et al., 2019. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. Remote Sens. Environ. 233, 111358. https://doi.org/10.1016/j.rse.2019.111358.
Xue, L., Liu, Y., Xiong, Y., et al., 2021. A data-driven shale gas production forecasting method based on the multi-objective random forest regression. J. Petrol. Sci. Eng. 196, 107801. https://doi.org/10.1016/j.petrol.2020.107801.
Yang, Q., Zhang, J., Zhou, J., et al., 2023. A hybrid data-driven machine learning framework for predicting the performance of coal and biomass gasification processes. Fuel 346, 128338. https://doi.org/10.1016/j.fuel.2023.128338.
Yin, Z., Strebelle, S., Caers, J., 2020. Automated Monte Carlo-based quantification and updating of geological uncertainty with borehole data (AutoBEL v1. 0). Geosci. Model Dev. (GMD) 13 (2), 651–672. https://doi.org/10.5194/gmd-13-651-2020.
You, J., Ampomah, W., Sun, Q., 2020. Development and application of a machine learning based multi-objective optimization workflow for CO_2-EOR projects. Fuel 264, 116758. https://doi.org/10.1016/j.fuel.2019.116758.
Yuan, J., Luo, D., Feng, L., 2015. A review of the technical and economic evaluation techniques for shale gas development. Appl. Energy 148, 49–65. https://doi.org/10.1016/j.apenergy.2015.03.040.
Zhang, C., Ma, Y., 2012. Ensemble Machine Learning: Methods and Applications. Springer Science & Business Media. https://link.springer.com/book/10.1007/978-1-4419-9326-7.
Zhang, L., Dou, H., Wang, T., et al., 2022. A production prediction method of single well in water flooding oilfield based on integrated temporal convolutional network model. Petrol. Explor. Dev. 49 (5), 1150–1160. https://doi.org/10.1016/S1876-3804(22)60339-2.
Zhao, X., Rui, Z., Liao, X., et al., 2015. A simulation method for modified isochronal well testing to determine shale gas well productivity. J. Nat. Gas Sci. Eng. 27 (2),

479–485. https://doi.org/10.1016/j.jngse.2015.08.035.
Zhong, Y., Yang, H., Zhang, Y., et al., 2020a. Online random forests regression with memories. Knowl-Based Syst 201, 106058. https://doi.org/10.1016/j.knosys.2020.106058.
Zhong, Y., Yang, H., Zhang, Y., et al., 2021a. Long short-term memory self-adapting online random forests for evolving data stream regression. Neurocomputing 457, 265–276. https://doi.org/10.1016/j.neucom.2021.05.026.
Zhong, Z., Sun, A.Y., Wang, Y., et al., 2020b. Predicting field production rates for waterflooding using a machine learning-based proxy model. J. Petrol. Sci. Eng. 194, 107574. https://doi.org/10.1016/j.petrol.2020.107574.
Zhong, Z., Sun, A., Jeong, H., 2019. Predicting CO_2 plume migration in heterogeneous formations using conditional deep convolutional generative adversarial network. Water Resour. Res. 55 (7), 5830–5851. https://doi.org/10.1029/2018WR024592.
Zhong, Z., Sun, A., Ren, B., et al., 2021b. A deep-learning-based approach for reservoir production forecast under uncertainty. SPE J. 26 (3), 1314–1340. https://doi.org/10.2118/205000-PA.
Zhou, Q., Kleit, A., Wang, J., et al., 2014. Evaluating gas production performances in marcellus using data mining technologies. In: Unconventional Resources Technology Conference. https://doi.org/10.15530/urtec-2014-192021.
Zhuang, F., Qi, Z., Duan, K., et al., 2020. A comprehensive survey on transfer learning. Proc. IEEE 109 (1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555.
Zhukov, A., Sidorov, D., Foley, A., 2017. Random forest based approach for concept drift handling. In: Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016. Springer International Publishing, pp. 69–77. https://doi.org/10.1007/978-3-319-52920-2_7.