A Novel Approach for the Prediction of Real-Time Rate of Penetration in Drilling for Petroleum by Combining the Attention-Based Bidirectional-Long Short-Term Memory and Long Short-Term Memory

Lyu, Z.; Wei, F.; Pan, Y.; Li, L.; Song, C.; Song, G.; Zhu, Z.; Song, X.

DOI
10.2523/IPTC-24898-MS

Publication date
2025

Document Version
Final published version

Published in
International Petroleum Technology Conference

Citation (APA)
Lyu, Z., Wei, F., Pan, Y., Li, L., Song, C., Song, G., Zhu, Z., & Song, X. (2025). A Novel Approach for the Prediction of Real-Time Rate of Penetration in Drilling for Petroleum by Combining the Attention-Based Bidirectional-Long Short-Term Memory and Long Short-Term Memory. In *International Petroleum Technology Conference* Article IPTC-24898-MS International Petroleum Technology Conference (IPTC). https://doi.org/10.2523/IPTC-24898-MS

IPTC-24898-MS

A Novel Approach for the Prediction of Real-Time Rate of Penetration in Drilling for Petroleum by Combining the Attention-Based Bidirectional-Long Short-Term Memory and Long Short-Term Memory

Z. Lyu, F. Wei, and Y. Pan, PetroChina Oil, Gas & New Energies Company, Beijing, China; L. Li and C. Song, Research Institute of Petroleum Exploration & Development, PetroChina, Beijing, China; G. Song, Department of Geoscience and Engineering, Delft University of Technology, Stevingweg, CN, Delft, the Netherlands; Z. Zhu and X. Song, State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing, Beijing, China

This paper was prepared for presentation at the International Petroleum Technology Conference held in Kuala Lumpur, Malaysia, 18 - 20 February 2025.

Abstract

Rate of penetration has been considered as an important factor in the entire drilling industry, which can largely determine the overall costs of drilling a well. This paper proposed a novel real-time prediction of rate of penetration by combining the Attention-based Bidirectional-Long Short-Term Memory and Long Short-Term Memory (Att-Bi-LSTM-LSTM). Eight parameters, which are total vertical depth, weight on bit, revolutions per minute, mud flow rate, density, viscosity, drill-bit outer-diameter, lithology, and rate of penetration, are adopted as datasets. The drilling speed of the well is trained and validated through the drilling data while a sliding window is introduced for the real-time update. In addition, the presented prediction model is compared with other traditional prediction methods. Finally, the prospect of field application and further study is discussed and suggested. The results indicate that the proposed model shows good accuracy and robustness. Moreover, compared with the traditional methods, the model exhibits good superiority with smaller absolute and relative errors. For field applications, the model proposed in this paper attempts to provide a solution to the prediction of real-time rate of penetration. The results are expected to provide guidance for the further study on the increase of drilling speed and reduction of well costs.

Introduction

As the development of oil and gas industry, more and more hydrocarbons are extracted from deeper formations. The cost of oil and gas production has been paid more attention by engineers. Among all factors, drilling has a large influence on the overall cost. Thus, drilling optimization is an important area studied by researchers (Eren and Ozbayoglu, 2010). One of the main goals of drilling optimization is to increase the rate of penetration (ROP). Barbosa et al. (2019) reviewed machine learning method applied to drilling rate of penetration prediction and optimization. The methods were classified as traditional models, statistical models and machine learning methods. It was concluded that machine learning techniques could potentially outperform in terms of prediction accuracy compared with traditional or statistical methods. It was also pointed out that even though good and reasonable results could be achieved, there was a lack of implementation of machine learning techniques in the petroleum industry.

The traditional model focuses on the analytical solution to the prediction of ROP. One of the well-known traditional models was proposed by Bourgoyne and Young (1974), which was based on a multiple regression analysis. The applied drilling data included formation strength, depth, compaction, pressure differential across the hole bottom, bit diameter and weight, rotary speed, bit wear and hydraulics. Hareland and Rampersad (1994) presented a drag bit model considering single-cutter rock interaction, lithology coefficients and bit wear. Motahhari et al. (2010) developed a drilling optimization procedure for the positive displacement motors and PDC bits. Optimization was also obtained for weight on bit and surface revolutions per minute. Recently, a theoretical model for roller cone bit is proposed by Deng et al. (2016), in which dynamic compressive strength of the rock was applied rather than static compressive strength. In addition, a ROP model based on the regression analysis was presented by Al-abduljabbar (2019), which was shown to outperform other traditional models. Other researchers also focused on the fitting methods of these traditional models. Multiple regression is often used to obtain empirical coefficients from the ROP models (Bourgoyne and Young, 1974). Furthermore, the optimization methods are attempted to determine the parameters (Barbosa et al., 2019).

The statistical model rarely focuses on the physical interaction between the bit and the rock, but select the most important features to obtain good prediction results (Eskandarian et al., 2017; Hegde et al., 2017; Ashrafi et al., 2019). Seifabad and Ehteshami (2013) used the information from 50 oil wells in Ahvaz oil field. Different regression equations were tested and ROP models were tailed for different formations. Moraveji et al. (2016) investigated the simultaneous effect of different variables. The relationship between penetration rate and variables was established using the response surface methodology. Then the bat algorithm (BA) was used to optimize the range of factors to maximize drilling rate of penetration. The results showed that the penetration rate can be estimated accurately at 95% confidence interval with cumulative probability distribution. Besides, Hedge et al. (2015) first identified both linear and non-linear relationships between ROP and different factors to save computational power. The ensemble techniques were also used, which illustrates that it can reduce error as well as increase computational efficiency. Linear techniques with bootstrapped residuals showed a low error rate in the results.

In recent years, machine learning models have been included in the prediction of penetration rate (Elkatatny et al., 2018; Hegde and Gray, 2018; Hegde et al., 2018a). Among them, artificial neural network (ANN), support vector machines (SVM), fuzzy inference systems, neuro-fuzzy, and ensemble models are the five kinds of methods that are often applied. Diaz et al. (2018) used an ANN to predict the ROP with drilling data from a 4.2 km-deep well at an enhanced geothermal system project. Three data training scenarios, i.e. accumulative data, different amounts of data and square root resampling. The results showed that the accuracy decreases in deeper sections. It was suggested that accumulative data and data resampling could be used for ROP prediction. Ashrafi et al. (2019) used petrophysical logs and drilling data from a vertical well in Marun oil field of Iran. Savitzky-Golay (SG) smoothing filter was implemented to reduce overall noise. Then eight hybrid ANN were developed and the data were trained by genetic algorithm (GA), particle swarm optimization (PSO), biogeography-based optimizer (BBO), and imperialist competitive algorithm (ICA). Results showed that PSO-multi-layer perception (PSO-MLP) and PSO-radial basis function (PSO-RBF) neural networks yielded the highest performance compared to results of other developed models. Soares and Gray (2019) predicted real-time ROP with both analytical and machine learning models. The ROP model was assumed to be able to learn and adapt in real-time. Results showed that machine learning models performed better than analytical models. Cross-validation was also carried out to select the best performing ROP model. Ahmed et al. (2019) explored four computational intelligence techniques, which are ANN, extreme learning machine, support vector regression and least-square support vector regression (LS-SVR). Comparative performance of four techniques was carried out and the effect of reduced number of predictors was analyzed. Results showed that the LS-SVR had the best predictive performance while ANN had the best testing execution time. Besides, the specific energy concept was introduced to optimize drilling parameters. Abbas et al. (2018) developed a model to predict ROP in deviated wells with ANN. The influence of well trajectory (azimuth and inclination) was considered. Anemangely et al. (2018) proposed a hybrid model composed of a multilayer perceptron (MLP) neural network with either a PSO algorithm or a cuckoo optimization algorithm (COA). The Savitzky-Golay filter was used for petrophysical logs and drilling data. Validation was performed via the multilinear regression method and the denoising step was crucial for the better performance of the training. Yavari et al. (2018) employed Hareland-Rampersad (HR) model, Bourgoyne and Young (BY) model, and an adaptive-neuro-fuzzy inference system (ANFIS) to predict the drilling rate in the South Pars gas field offshore of Iran. It was shown that based on a large amount of data, the ANFIS was more accurate. Similarly, Brenjkar and Delijani (2022) compared four machine learning methods (i.e. multilayer perceptron neural network (MLPNN), radial basis function neural network (RBFNN), adaptive neuro-fuzzy inference system (ANFIS), and support vector regression (SVR)) and two traditional ROP models (i.e. Bourgoyne and Young (BYM) and Bingham). Results indicated that PSO-MLPNN achieved the highest performance. It was also concluded that machine learning methods were more efficient and reliable. Usually, the ROP is predicted when the well is finished. Part of the drilling data is adopted for training while the rest is used for validation (Soares and Gray, 2019). Zhang et al. (2022) combined attention-based gated recurrent unit network and fully connected neural networks to predict real-time ROP. Besides, Wallace et al. (2015) introduced a new system to optimize ROP and provide real-time performance increases and closed-loop control. The framework for phased development and integration into drilling rig operations was presented.

In summary, it can be seen that, in previous studies, various machine learning methods have been studied to predict ROP. However, the real-time ROP prediction is rarely studied. The real-time ROP prediction is important because the prediction can be performed in the whole drilling process, which can provide the basis for the drilling parameter optimization and the increase of drilling speed especially when unknown geological risks are encountered. Another advantage of the real-time ROP prediction is that the used dataset is relatively small compared with conventional artificial intelligence prediction methods.

In this paper, we attempt to propose a novel way for the prediction of ROP by combining the Attention-based Bidirectional-Long Short-Term Memory and Long Short-Term Memory (AttBi-LSTM-LSTM). This paper is organized as follows: first, the parameters of drilled wells are presented and adopted as datasets. Second, the input features of the model are filtered and missing values are imputed. Third, the ROP of the well is trained and validated through the drilling data while a sliding window is introduced for the real-time update. Finally, the novel presented prediction method is compared with other traditional methods. The results are also analyzed and the prospect of field application is discussed.

Methodology

Datasets

The ROP data from Xinjiang oil field, China are adopted (i.e. G101, G102 and G103). These wells are the first three appraisal wells in an anticline of a piedmont thrust belt, which means that unlike production wells, there are more uncertainties and the risk is much higher. The depths of G101, G102 and G103 are 7000 m, 6100 m and 6212 m, respectively. The entire dataset consists of 1910 data samples. Eight parameters are selected of the wells G101, G102 and G103, which are total vertical depth (TVD), weight on bit (WOB), revolutions per minute (RPM), mud flow rate, density (Dens), viscosity (Visco), drill-bit outer-diameter (DB_OD), lithology (Litho), and rate of penetration (ROP). Fig.1 shows the drilling rates of well G101, G102 and G103. It can be seen that although these three wells are drilled in the same anticline, the drilling rates change quiet differently between each other. The various trends make it difficult to be predicted with the same method. In addition, part of the recorded parameters from well G101 is listed in Table 1.

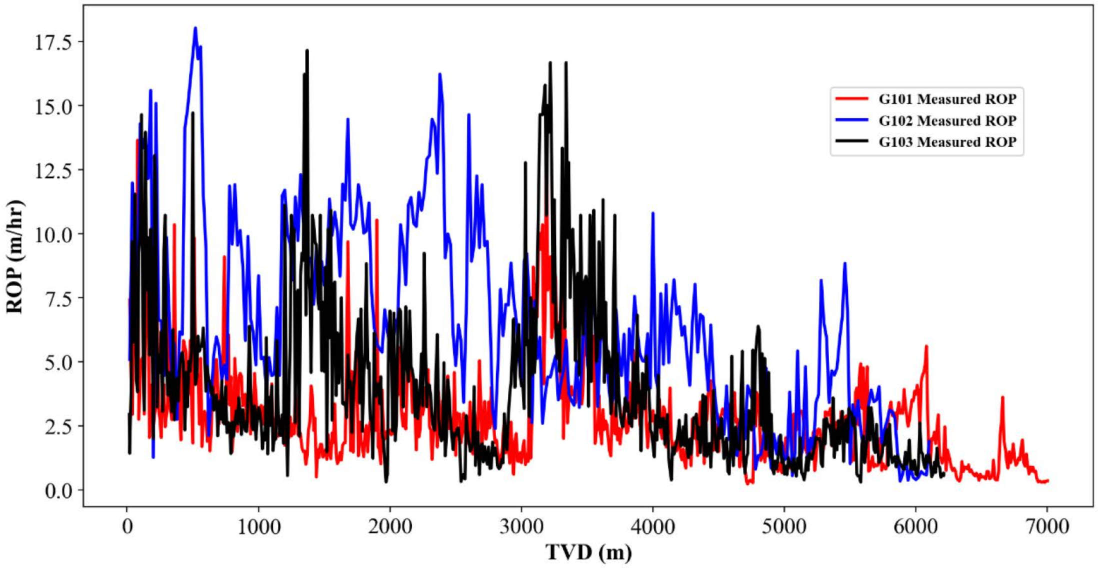

Figure 1—Rate of penetration of well G101, G102, and G103

Table 1—Part of parameters of the well G101

Depth (m)	WOB (kN)	RPM (rpm)	Mud Flow Rate (L/s)	Density (g/cm^3)	Viscosity (s)	DB (mm)	ROP (m)
20	35	55	26.5	1.26	180	660.4	6.00
30	30	55	26.5	1.26	180	660.4	4.31
…	…	…	…	…	…	…	…
6990	83	40	13.3	2.27	105	215.9	0.34
7000	82	40	15.5	2.27	98	215.9	0.35

Algorithm

In 1977, Hochreiter and Schmidhuber brought up the Long short-term memory (LSTM) method, which is part of the Recurrent Neural Network (RNN). The advantage of LSTM over RNN is that it can overcome the vanishing gradient and suitable for learning that involves long-term dependence (Sunjaya et al., 2022). The LSTM network is composed of input gate, forget gate and output gate. A detail description of the LSTM unit and its variant under stacked format applied for production forecasting is illustrated in many pilot research (Pan et al., 2019; Wang et al., 2019; Song et al., 2020 and Fan et al., 2021).

LSTM Unit. Mathematically, the gates are defined as: input gate i_t, forget gate f_t, output gate o_t, the formula for each gate is presented below.

Block input activation:

$$a_t = g\left(W_a \cdot x_t + U_a \cdot h_{t-1} + b_a\right) \tag{1}$$

Input gate:

$$i_t = \sigma\left(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i\right) \tag{2}$$

Forget gate:

$$f_t = \sigma\left(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f\right) \tag{3}$$

Output gate:

$$o_t = \sigma\left(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o\right) \tag{4}$$

Updating iteratively and passing internal states:

$$c_t = a_t \odot i_t + f_t \odot c_{t-1} \tag{5}$$

$$h_t = g(c_t) \odot o_t \tag{6}$$

Define H as the combination from equation (1) to (6):

$$h_t = H\left(x_t, h_{t-1}\right) \tag{7}$$

x_t is the input vector at time t, N is the number of LSTM units, and M is the number of inputs.

Input weights: W_a, W_i, W_f, W_o) $\in R^{N \times M}$.

Recurrent weights: U_a, U_i, U_f, $U_o \in R^N$.

Bias weights: b_a, b_i, b_f, $b_o \in R^N$.

σ, g are the nonlinear activation functions, more specifically in this paper, σ is the sigmoid function $\left(\sigma(x) = \frac{1}{1+e^{-x}}\right)$, g is the tanh function $\left(\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}\right)$. In the above equations, $\odot$ represents the pointwise multiplication of two vectors and inner productions will be denoted as.

Attention Mechanism. Attention acts as the indispensable cognitive function of human beings that selectively concentrate on the important fragment of information when and where it is needed. The attention mechanism has been proved to be effective and greatly enhancing the accuracy and efficiency in perceptual information processing. Two categories are defined, one with the bottom-up unconscious attention is called saliency-based attention, which is similar to the max-pooling and gating mechanism. The other one with the top-down conscious attention is the focused attention that has the predetermined purpose and relies on specific tasks. In this paper, the attention mechanism we referred to is the focused attention which can be used as the resource allocation scheme. Particularly in drilling ROP prediction scenario, the focused attention mechanism further provides performance improvements in addition to allocating the abundant downhole sequential information.

We follow the notation from the first proposed attention model, the *RNNsearch* (Bahdanau et al. 2014):

$$C_t = \sum_{j=1}^{T} \alpha_{tj} \hat{h}_j \tag{8}$$

The attention weight α_{tj} of each annotation

$\hat{h}_j$ can be calculated based on the following equations:

$$e_{tj} = \mathbf{a}\left(s_{t-1}, \hat{h}_j\right) \tag{9}$$

$$\alpha_{tj} = \frac{\exp\left(e_{tj}\right)}{\sum_{k=1}^{T} \exp\left(e_{tk}\right)} \tag{10}$$

Obviously, α_{tj} is determined using *softmax* function where the intermediate attention scores e_{tj} is calculated using a learnable function ***a***, specifically tanh is adopted and reflects the importance of the annotation h_j to the next hidden state s_t based on the previous state s_{t-1}.

The LSTM output at the current step is then expressed as:

$$\hat{y}_t = \mathrm{LSTM}\left(\hat{y}_{t-1}, C_t, s_{t-1}\right) \tag{11}$$

The above formulas of attention mechanism enable the LSTM networks to concentrate on the critical components other than the redundant features.

Bidirectional LSTM Structure. Bi-LSTM was developed with an analogy of bidirectional RNN (Schuster and Paliwal, 1997). Unlike unidirectional LSTM, the data sequence is processed in both forward and backward directions with two distinct hidden layers connected to the same output layer. Graves and Schmidhuber (2005) first applied Bi-LSTM to phoneme classification and further demonstrate the applicability in speech recognition (Graves et al., 2013). And in many fields, it has been widely accepted that the Bi-LSTM substantially outperforms the unidirectional ones. However, Bi-LSTM has not been well explored and non rigorous workflow has been established in drilling ROP prediction. The aforementioned two separate hidden layers consists of a forward LSTM and backward LSTM layer which is illustrated in Fig. 2. In a Bi-LSTM structure, we define the forward hidden sequence $\overrightarrow{h}$ as Eq. (8) and the backward hidden sequence $\overrightarrow{h}$ as Eq. (9) by updating the backward layer from $t=T$ to 1, the forward layer from $t=1$ to T.

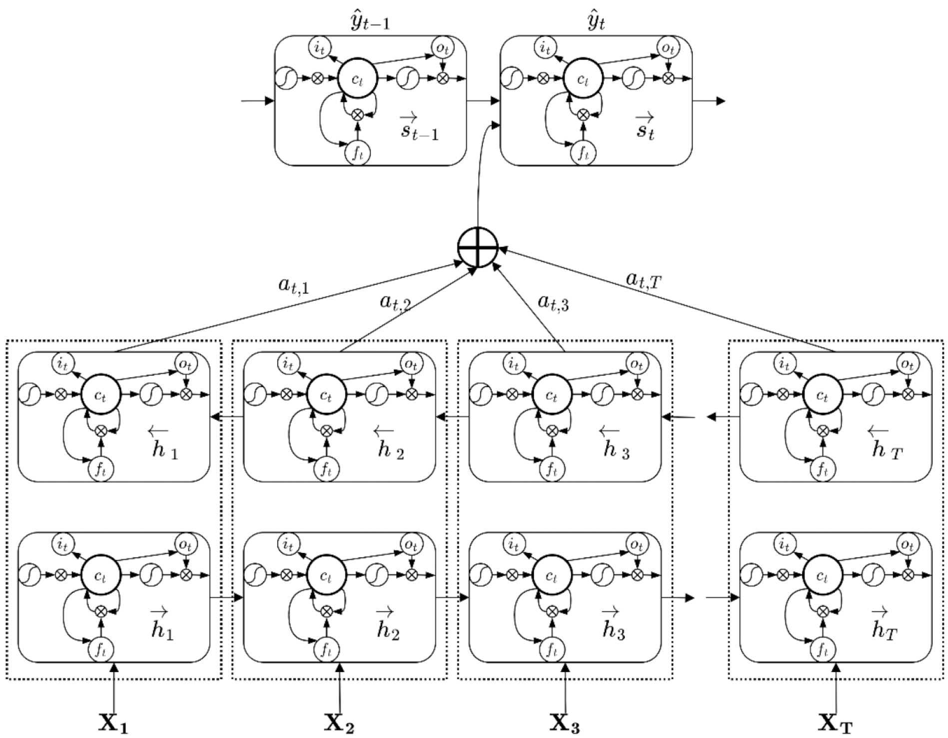

Figure 2—Illustration of the proposed stacked attention based Bi-LSTM and LSTM model for real-time drilling ROP forecast given downhole information (X_1, X_2, X_3,…, X_T)

$$\overrightarrow{h}_t = H\left(x_t, \overrightarrow{h}_{t-1}\right) \tag{12}$$

$$\overleftarrow{h}_t = H\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{13}$$

Finally, the hidden state connected to the output layer is linearly transformed to derive the estimation $\hat{h}_j$ at time t as Eq. (14) shown.

$$\hat{h}_t = W_y \bigodot \left(\overrightarrow{h}_t, \overleftarrow{h}_t\right) + b_y \tag{14}$$

The BiLSTM units are stacked, the attention layer is embedded. In forward pass, the input sequence is fed into the forward BiLSTM unit from the first time-step to the end. Simultaneously in backward pass, the input sequence is fed into the backward BiLSTM unit in reverse time-step.

Both forward and backward LSTM computes its hidden state and updates its memory cell based on the current input and the previous hidden state and memory cell. Once the above passes are completed, the hidden states from both are concatenated or applied mathematical transformations. Then the attention layer employes a weight matrix to calculate the weighted sum of Bi-directional hidden states. The summation after the attention layer is transferred to another uni-directional LSTM layer as C_t to approach the outcome $\hat{y}_t$.

Workflow. Progressively built up higher level of representation of sequence data, the hidden layer architecture of the stacked attention based Bi-LSTM and LSTM networks is composed by a bidirectional LSTM layer, an attention layer and a unidirectional LSTM layer. The basic idea of stacking is the mimic of a full scale of decoding in the attention-based neural machine translation (Bahdanau et al. 2014). The spatial-temporal information of the downhole data and the spatial dependencies of the drilling features can be captured during the training process. The purpose of this study is to predict real-time ROP. The workflow of training and validating the proposed model is illustrated in Fig. 3. The process starts from real-time data preparation with outlier screening and missing data imputation. Next, the correlation analysis is performed to evaluate the nonlinear dependencies of each two variables. The coefficient value close to 1 represents the high dependencies. Unlike other research that only part of the features has been selected, all eight available input features have been fed in during model training procedure. The weights are optimized using MSE (mean square error) metrics as shown in Eq. (15) and the Adam algorithm (the details can be found in Kingma et al. 2014) at each training and validating epoch. The best one is selected as the sliding-window validation scheme. The reason for employing sliding-window method instead of random split is the real-time forecast paradigm that the model is updated dynamically as training data is continuously expanding. In the workflow, we split the historical data into K folds so that the sliding-window is moved forward for training and validating for N_k times.

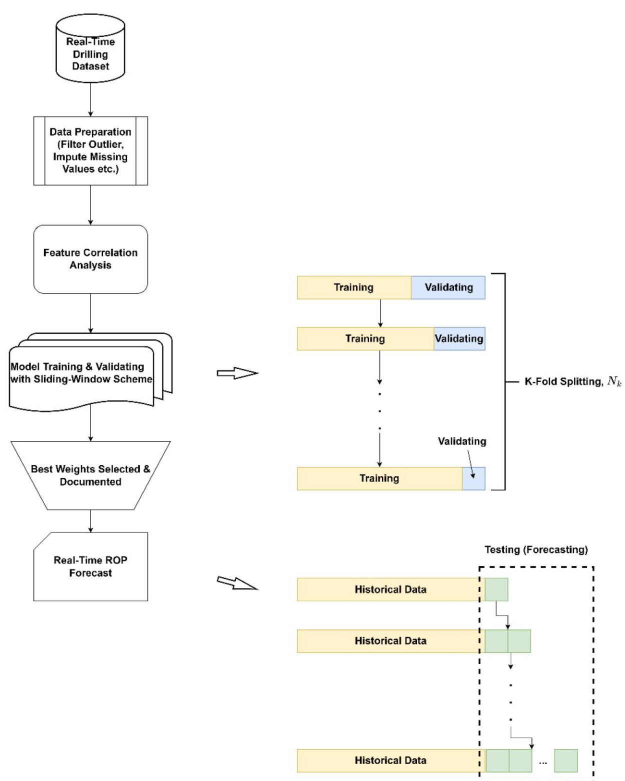

Figure 3—Workflow of the real-time ROP prediction

$$\operatorname{argmin}\frac{1}{N_k}\sum_{k=1}^{N_k}\left[\left|\frac{1}{N_{T,k}}\sum_{i=1}^{N_{T,k}}\left(\hat{y}_i - y_i\right)^2 - \frac{1}{N_{V,k}}\sum_{j=1}^{N_{V,k}}\left(\hat{y}_j - y_j\right)^2\right| + \frac{1}{N_{T,k}}\sum_{i=1}^{N_{T,k}}\left(\hat{y}_i - y_i\right)^2 + \frac{1}{N_{V,k}}\sum_{j=1}^{N_{V,k}}\left(\hat{y}_j - y_j\right)^2\right] \quad (15)$$

Results and discussion

In this section, the Att-BiLSTM-LSTM model is evaluated by the well datasets. The accuracy of the model is also discussed. Additionally, the presented model is compared with traditional methods.

Results analysis

Fig. 4 shows the variable relationships of TVD, ROP, WOB, RPM, Mud Flow Rate, Dens, Visco, DB_OD, Litho. It can be seen that about half of the values are higher than 0.5, which means that the two variables have strong correlations. From the first column of TVD to Litho, we can see that the correlation value between WOB and TVD is 0.5. Other variables with a correlation value higher than 0.5 are mud flow rate, density, DB_OD and Litho. Physically, in order to achieve the geological goal, drill-bits with different diameters need to be used at different depths. Thus, the DB_OD has a strong correlation with TVD. Accordingly, the mud flow rate varies with the DB_OD to carry the debris, resulting in the relationship with depth. Similarly, the lithology changes with the depth due to the sedimentation of different ages, which has a correlation value of 0.71. The mud density also has to be adjusted based on the pressure gradient at different depths. Thus, it is physically reasonable to use these variables as input parameters.

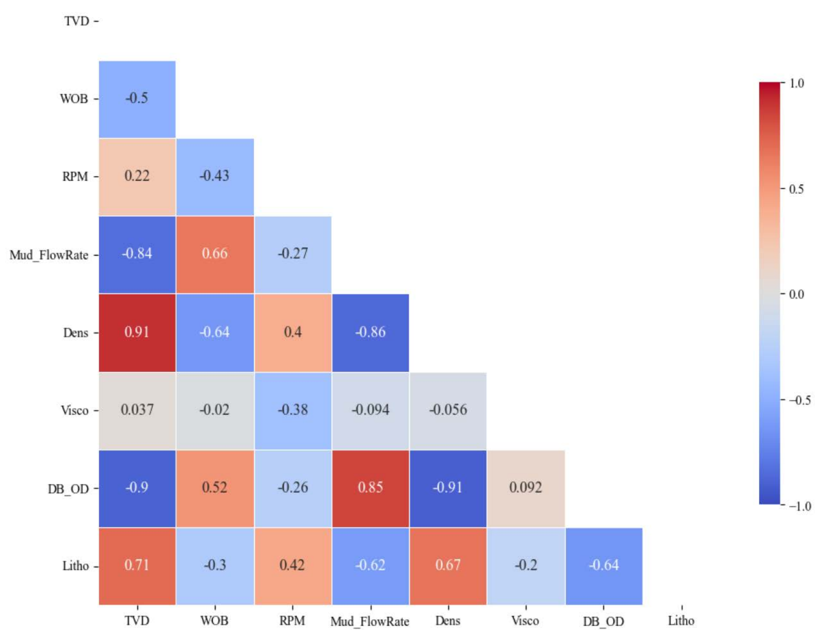

Figure 4—The heatmap illustrating the variable dependencies

Fig. 5 shows the comparison between the predicted ROP and measured ROP of well G101. The whole well dataset is divided into three sections, which are training section, validation section and test section. In Fig. 5, the ground truth ROP changes very sharply at nearly the entire drilling process, which makes it difficult to train. In this paper, by applying the Att-BiLSTM-LSTM model, it can be seen that from 0 to about 5000 m, the trained values can catch the relation between the true ROP and depth, which means that the model has been trained well. Then, in the validation section from 5000 m to 6000 m, the modeled red line may miss some of the measured ROP points, but it generally follows the variation. Finally, in the test or predicted section, the predicted values can capture the rising and dropping properties, which shows the effectiveness of the proposed method.

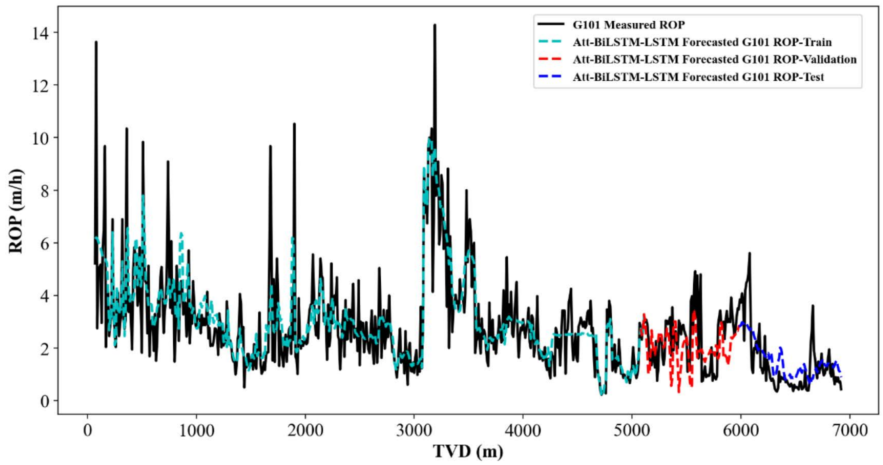

Figure 5—Comparison of the predicted ROP and ground truth ROP of well G101

Fig. 6 shows the absolute error and relative error of well G101 of the test section. The third quartile and first quartile absolute error are 0.97 m/h and 0.26 m/h respectively, while the corresponding relative error are 0.44 and 0.11 accordingly. In addition, the upper limit of the absolute and relative error are smaller than 2.0 and 1.0, respectively, which means that most of predicted results are in reasonable range. The relationship between the predicted ROP and measured ROP is shown in Fig. 7. It can be seen that the two parameters are closely related with a determination coefficient of 0.8.

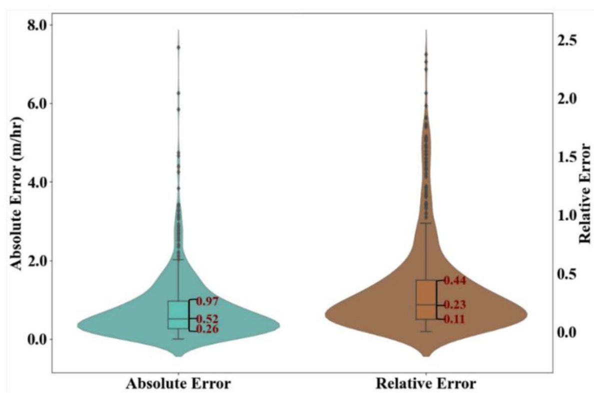

Figure 6—Absolute error and relative error of well G101

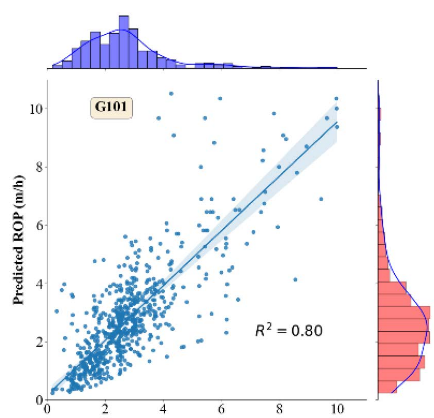

Figure 7—Relationship between the predicted ROP and measured ROP

Model applications

To test the robustness of the proposed Att-BiLSTM-LSTM model, the method is also applied to the rest two wells of G102 and G103. As before, the two well datasets are divided into three sections (i.e., training section, validation section and test section). As shown in Fig. 8 and Fig. 9, the predicted ROP can capture the changing trend of the measured ROP similarly as applied in well G101.

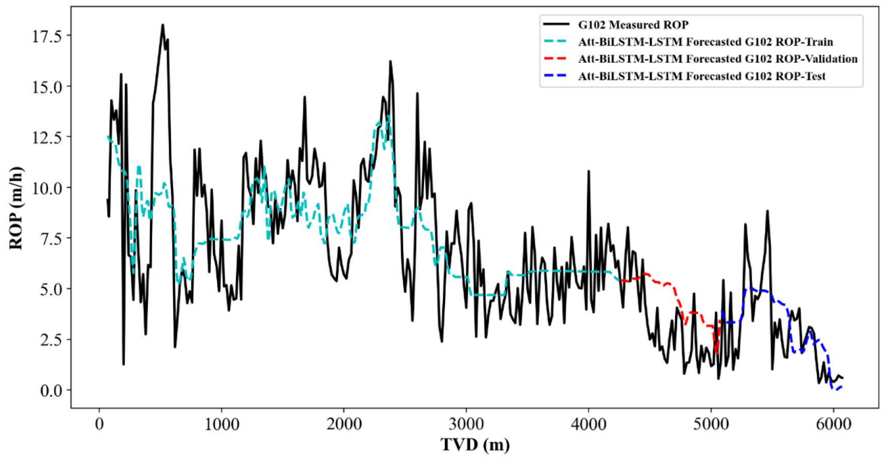

Figure 8—Comparison of the predicted ROP and ground truth ROP of well G102

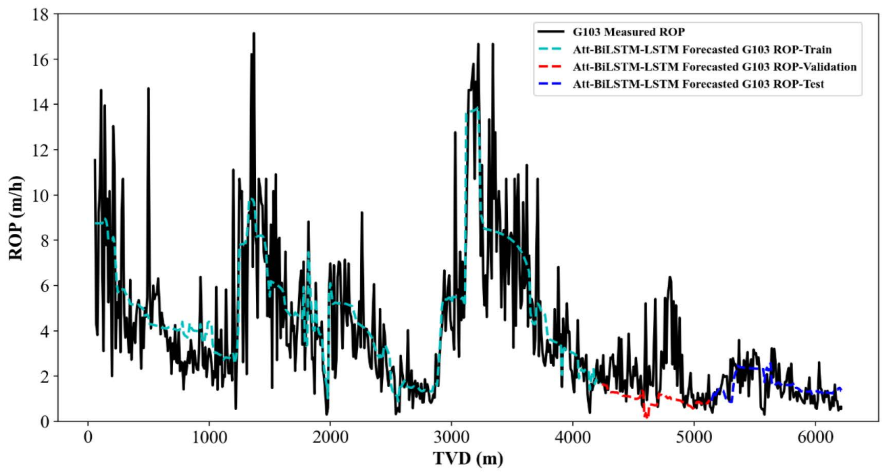

Figure 9—Comparison of the predicted ROP and ground truth ROP of well G103

Absolute error and relative error of well G102 and G103 are shown in Fig. 10. For well G102, the absolute error ranges most likely from 0.71 m/h to 2.42 m/h, while the boundaries of interquartile range are within the bracket of 0 m/h to 5.0 m/h. For well G103, the absolute error most likely ranges from 0.37 m/h to 1.65 m/h, while the interquartile values are from 0 m/h to 4.0 m/h. Fig. 11 gives the correlations of the predicted ROP and measured ROP of well G102 and G103. It can be seen that the correlation coefficients are 0.80 and 0.81, respectively, which indicates that the Att-BiLSTM-LSTM model presented in this paper can fit the nonlinear relationship between the ROP and selected variables in real time.

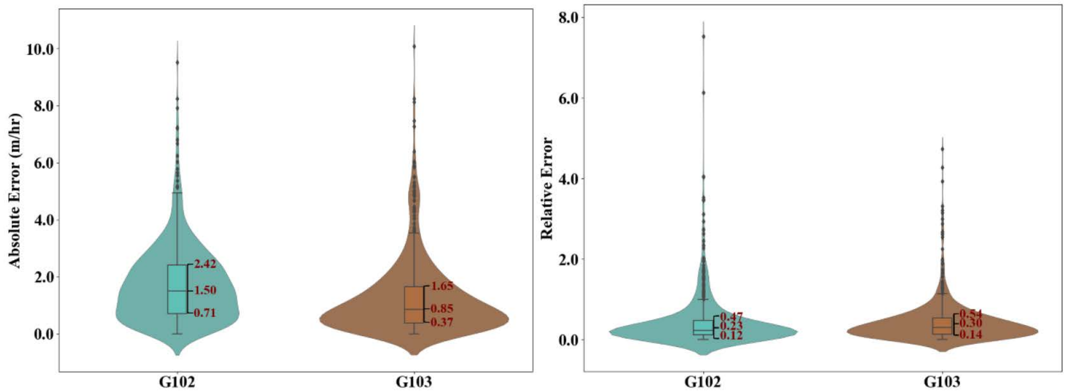

Figure 10—Absolute error and relative error of well G102 and G103

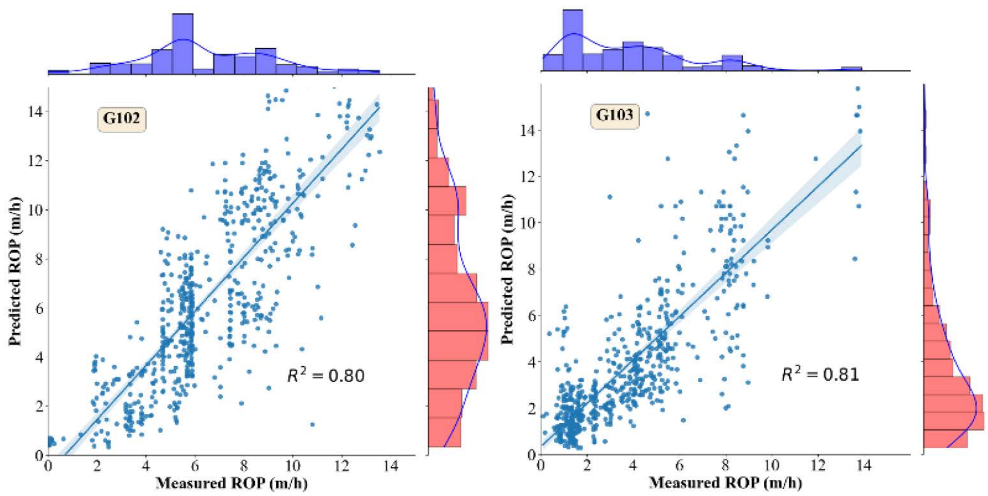

Figure 11—Relationships between the predicted ROP and measured ROP of well G102 and G103

Model comparison

In this section, the performance of the proposed Att-BiLSTM-LSTM model is compared with Modified Bourgoyne&Young (MBY), Bingham model and LSTM model.

The MBY model proposed by Soares and Gray (2019) is adopted in this section for comparison. The new formulation for the Bourgoyne and Young model is simplified to the core real-time drilling variables:

$$ROP = a_1 D^{a_2} WOB^{a_5} RPM^{a_6} q^{a_8} \tag{16}$$

where a_1, a_2, a_5, a_6, a_8 are constants, WOB is weight on bit, D is the well depth, RPM represents the rotational speed, q is the flow rate (Soares and Gray, 2019).

The Bingham model added an empirical exponent to the WOB term to include the rock formation properties (Bingham, 1964). The model is described as the following equation:

$$ROP = a\left(\frac{WOB}{d_b}\right)^b RPM \tag{17}$$

where a and b are constants, d_b is the bit diameter (Soares and Gray, 2019).

Fig. 12 shows the ROP comparison of the proposed Att-BiLSTM-LSTM model and Modified Bourgoyne&Young (MBY), and Bingham model. It can be seen that in the history match section, all three models can roughly capture the variations of the ROP of well G101. However, in the forecast section, the Att-BiLSTM-LSTM model is the most related model with the ROP. The MBY model predicts slightly higher than the measured ROP, while the Bingham model predicts slightly lower. As shown in Fig. 13, the absolute error of the Att-BiLSTM-LSTM model is smaller than the MBY model and Bingham model, with an average value of 0.76 m/h. The Bingham model exhibits the largest absolute error of 1.33 m/h. As also shown in Fig. 14, the relative error of the Att-BiLSTM-LSTM model is the smallest with an average value of 27.53%. The Bingham model also shows the largest average relative error of 47.86%.

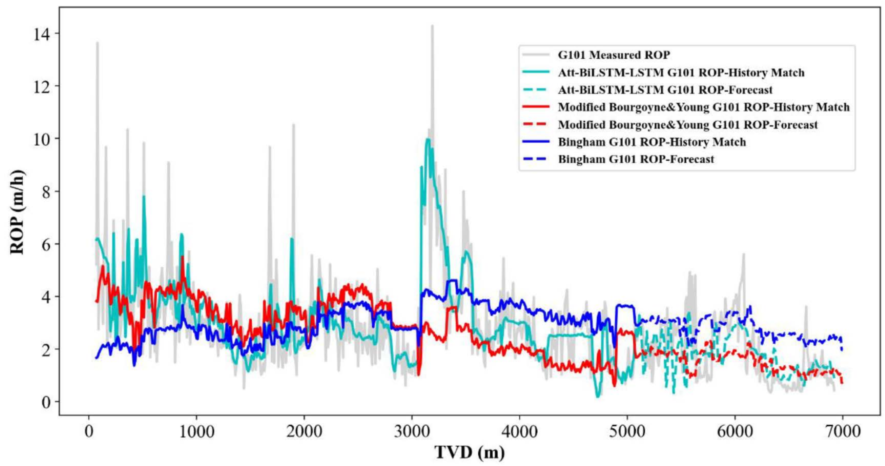

Figure 12—ROP comparison of Att-BiLSTM-LSTM model and Modified Bourgoyne&Young and Bingham model of well G101

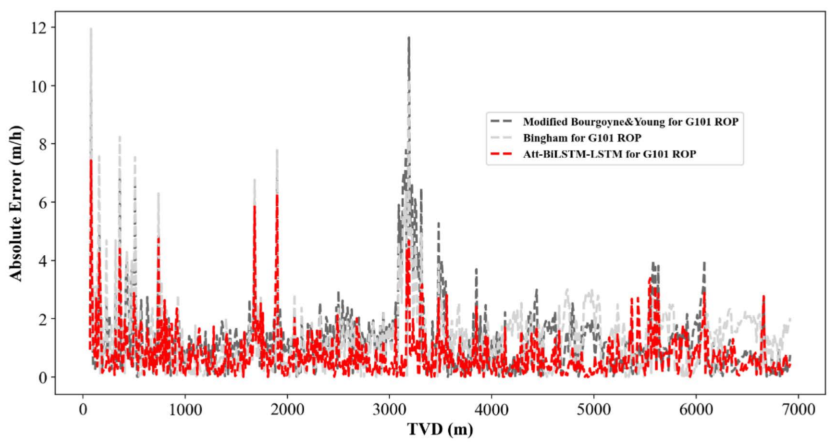

Figure 13—Absolute error comparison of Att-BiLSTM-LSTM model and Modified Bourgoyne&Young and Bingham model of well G101

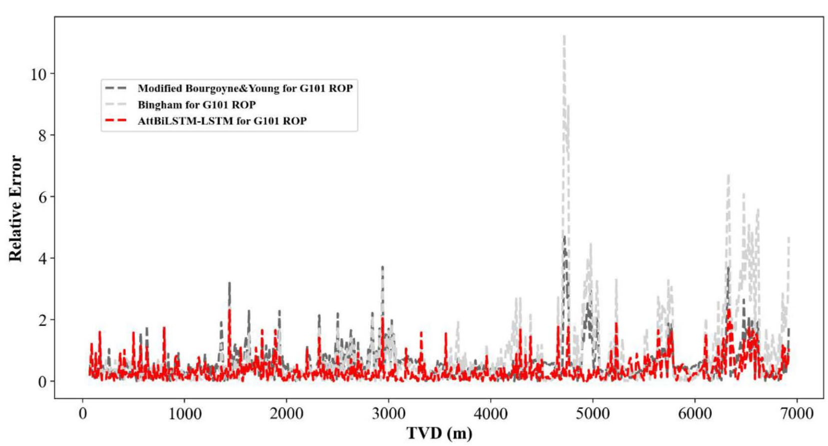

Figure 14—Relative error comparison of Att-BiLSTM-LSTM model and Modified Bourgoyne&Young and Bingham model of well G101

These three models are also applied in well G102. As shown in Fig. 15, compared with MBY model and Bingham model, the proposed Att-BiLSTM-LSTM model is the most related results in the forecast section. As described before, for well G102, the absolute error of the Att-BiLSTM-LSTM model is smaller than the MBY model and Bingham model, especially from the depth of 4500 m to 6000 m, with an average value of 1.30m/h (Fig. 16). The Bingham model gives the largest absolute error of 7.53 m/h. Fig. 17 shows the relative error of three models. The Att-BiLSTM-LSTM model is the smallest with an average value of 27.20%. The MBY model produces a relative error of 41.45%. The Bingham model also shows the largest relative error of 60.37%.

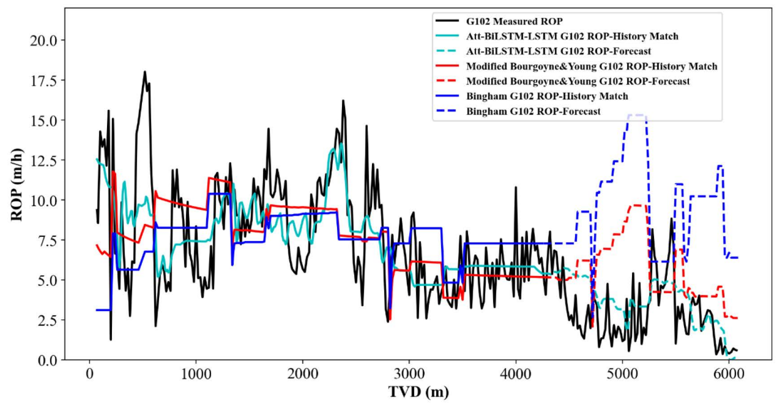

Figure 15—ROP comparison of Att-BiLSTM-LSTM model and Modified Bourgoyne&Young and Bingham model of well G102

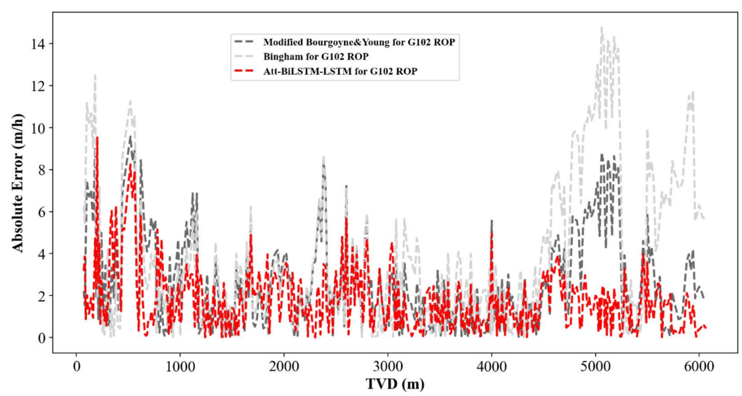

Figure 16—Absolute error comparison of Att-BiLSTM-LSTM model and Modified Bourgoyne&Young and Bingham model of well G102

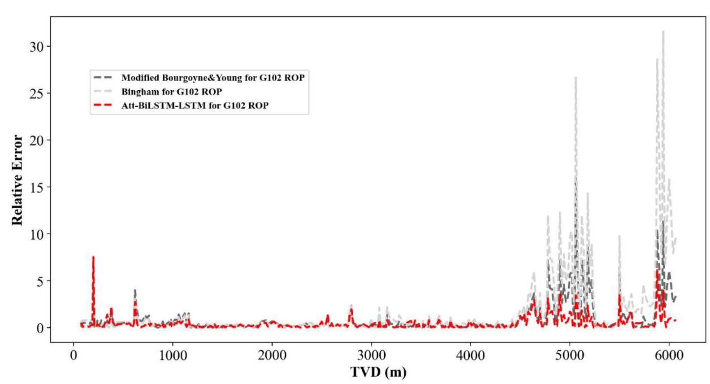

Figure 17—Relative error comparison of Att-BiLSTM-LSTM model and Modified Bourgoyne&Young and Bingham model of well G102

Fig. 18 shows the ROP comparison of the proposed Att-BiLSTM-LSTM model and LSTM model. It can be seen that although the LSTM model can be trained well in the history match section, the forecast section is obviously higher than the measured value. As shown in Fig. 19 and Fig. 20, the absolute error and relative error of the LSTM model is larger with an average value of 0.95m/h and 57.6%, respectively.

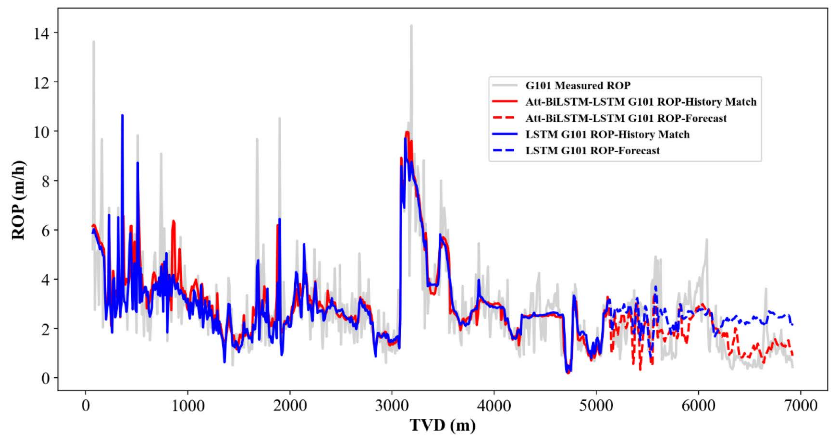

Figure 18—ROP comparison of Att-BiLSTM-LSTM model and LSTM model of well G101

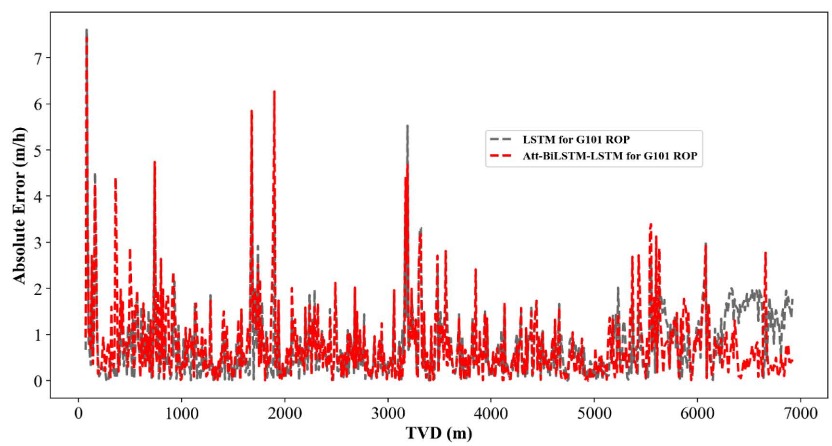

Figure 19—Absolute error comparison of Att-BiLSTM-LSTM model and LSTM model of well G101

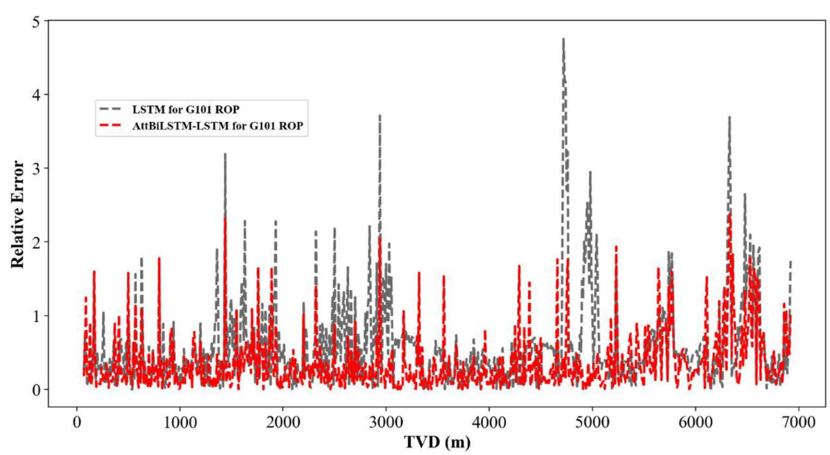

Figure 20—Relative error comparison of Att-BiLSTM-LSTM model and LSTM model of well G101

The same comparison between the Att-BiLSTM-LSTM model and LSTM model is also applied in well G102 as shown in Fig. 21, Fig. 22 and Fig. 23. The results indicate that the Att-BiLSTM-LSTM model still performs better than the LSTM model.

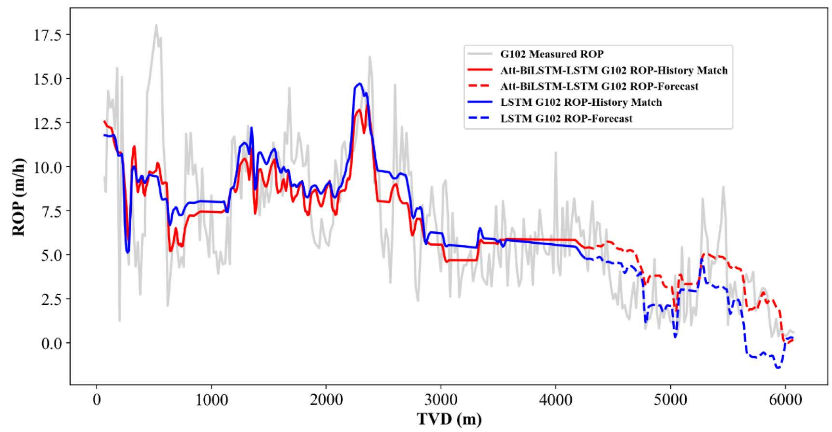

Figure 21—ROP comparison of Att-BiLSTM-LSTM model and LSTM model of well G102

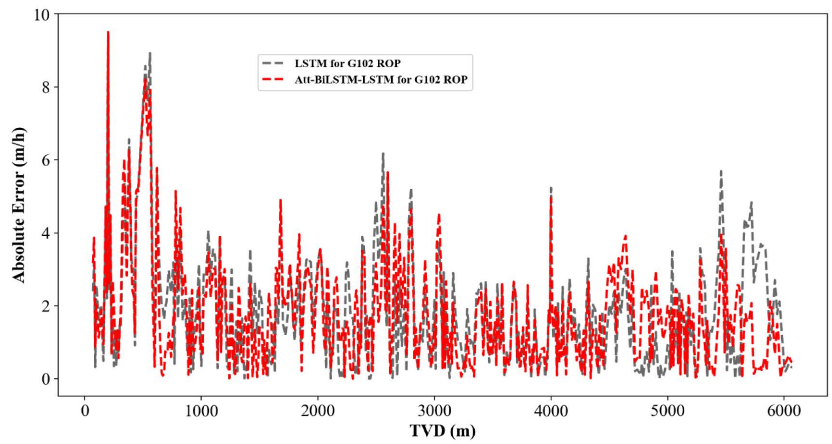

Figure 22—Absolute error comparison of Att-BiLSTM-LSTM model and LSTM model of well G102

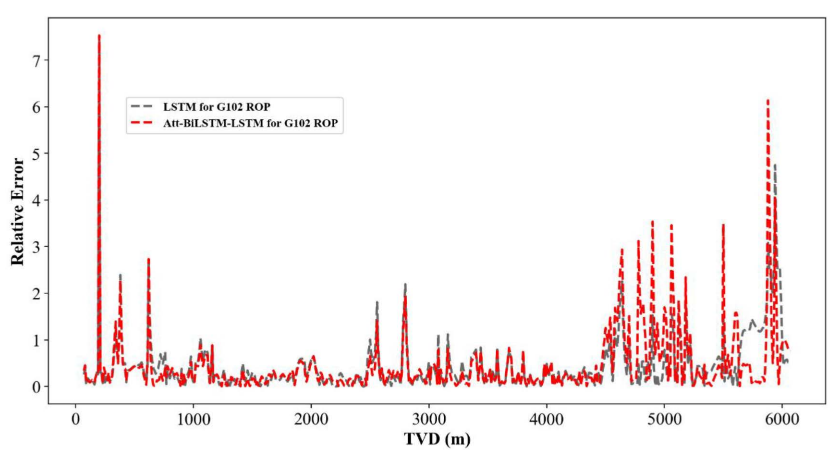

Figure 23—Relative error comparison of Att-BiLSTM-LSTM model and LSTM model of well G102

Conclusions

In this paper, a novel method for ROP prediction is built based on the Attention-based Bidirectional-Long Short-Term Memory and Long Short-Term Memory (Att-Bi-LSTM-LSTM). The model is applied in three deep wells and compared with other traditional methods. The results indicate that the model can capture the rising and dropping, exhibiting the effectiveness of the proposed method. In addition, the new model shows superiority over the modified Bourgoyne&Young and Bingham model.

In the future study, more variables, such as bit wear and formation properties, can be considered, which may help improve the accuracy of prediction. Furthermore, variable optimization is an important study topic for real-time ROP prediction. Optimized variables or parameters are obtained to increase the ROP in deep wells, which is good for the reduction of well costs.

Acknowledgments

The authors would like to acknowledge the Xinjiang oil field for providing the datasets of rate of penetration.

Nomenclature

$\boldsymbol{a}$	*tanh activation function in attention unit*
a	*Coefficient of Bingham model*
a_t	*Input activation*
a_1, a_2, a_5, a_6, a_8	*Coefficient of modified Bourgoyne and Young model*
b	*Coefficient of Bingham model*
b_a, b_i, b_f, b_o	*Bias weights vector,* $b_a, b_i, b_f, b_o \in R^N$
b_y	*Bias weights vector at the output layer*
C_t	*Context Vector in Attention Unit*
c_t, c_{t-1}	*Cell state at time t and t-1 respectively*
D	*Well depth, ft*
d_b	*Bit diameter, in*
$e(\cdot)$	*Exponential function*
e_{tj}, e_{tk}	*Intermediate attention scores*
f_t	*Forget gate*
g	*tanh activation function in LSTM unit*
H	*Recurrent function for linking hidden state*
h_t, h_{t-1}	*Annotation vector at time t and* t-1
$\vec{h}_t, \vec{h}_{t-1}, \vec{h}_t, \vec{h}_{t+1}$	*Annotation in forward component and backward component at time t, t-1,*t+1
$\hat{h}_j$	*Annotation in attention unit*
i_t	*Input gate*
$\boldsymbol{M}$	*The number of inputs*
$N, N_k, N_{T,k}, N_{v,k}$	*The number of LSTM units, the number of k-fold rolling-window validation, the number of T training samples, the number of V validation samples*
o_t	*Output gate*
q	*The flow rate*, gal/min
s_t, s_{t-1}	*Hidden state at time t and t-1*
U_a, U_i, U_f, U_o	*Recurrent weights,* $U_a, U_i, U_f, U_o \in R^N$
W_a, W_i, W_f, W_o	*Input weights,* $W_a, W_i, W_f, W_o \in R^{N\times M}$
x_t	*Input features*
$y_t, y_i, \hat{y}_t, \hat{y}_i$	*The vector of target and predicted outputs respectively*

α_{tj} *Weights of attention unit generated using Softmax fuction*

References

Abbas, A.K., Rushdi, S., Alsaba, M., 2018. Modeling rate of penetration for deviated wells using artificial neural network. In: Abu Dhabi International Petroleum Exhibition & Conference, 12-15 November, Abu Dhabi. UAE, *Society of Petroleum Engineers*.

Ahmed, O.S., Adeniran, A.A., Samsuri, A., 2019. Computational intelligence based prediction of drilling rate of penetration: a comparative study. *Pet J. Sci. Eng*. **172**, 1–12.

Al-abduljabbar, A., 2019. A robust rate of penetration model for carbonate formation. *Energy J. Resour. Technol*. **141**, 0429031–042903–9.

Anemangely, M., Ramezanzadeh, A., Tokhmechi, B., Molaghab, A., Mohammadian, A., 2018. Drilling rate prediction from petrophysical logs and mud logging data using an optimized multilayer perceptron neural network. *Geophys J. Eng*. **15**, 1146–1159.

Ashrafi, S.B., Anemangely, M., Sabah, M., Ameri, M.J., 2019. Application of hybrid artificial neural networks for predicting rate of penetration (ROP): a case study from Marun oil field. *Pet J. Sci. Eng*. **175**, 604–623.

Bahdanau, D., Cho, K., Bengio, Y., 2014. *Neural machine translation by jointly learning to align and translate*. arXiv preprint arXiv:1409.0473.

Barbosa L.F., Nascimento, A., 2019. Machine learning methods applied to drilling rate of penetration prediction and optimization - A review. *Petrol J. Sci. Eng*. **183**, 1–20.

Bingham, M.G., 1964. How rock properties are related to drilling. *Oil Gas J*. 94–101.

Bourgoyne, A., Young, F.S., 1974. A multiple regression approach to optimal drilling and abnormal pressure detection. *Soc. Pet. Eng. J*. **14**, 371–384.

Brenjkar E., Delijani E.B., 2022. Computational prediction of the drilling rate of penetration (ROP): A comparison of various machine learning approaches and traditional models. *Pet J. Sci. Eng*. **210**, 110033.

Deng, Y., Chen, M., Jin, Y., Zhang, Y., Zou, D., Lu, Y., 2016. Theoretical and experimental study on the penetration rate for roller cone bits based on the rock dynamic strength and drilling parameters. *Nat J. Gas Sci. Eng*. **36**, 117–123.

Diaz, M.B., Kim, K.Y., Kang, T.H., Shin, H.S., 2018. Drilling data from an enhanced geothermal project and its pre-processing for ROP forecasting improvement. *Geothermics* **72**, 348–357.

Elkatatny, S., 2018. New approach to optimize the rate of penetration using artificial neural network. *Arabian Sci J. Eng*. **43**, 6297–6304.

Eren, T., Ozbayoglu, M.E., 2010. Real time optimization of drilling parameters during drilling operations. In: SPE Oil and Gas India Conference and Exhibition. OnePetro.

Eskandarian, S., Bahrami, P., Kazemi, P., 2017. A comprehensive data mining approach to estimate the rate of penetration: application of neural network, rule based models and feature ranking. *Pet J. Sci. Eng*. **156**, 605–615.

Fan, D., Sun, H., Yao, J., Zhang, K., Yan, X., Sun, Z., 2021. Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy*, **220**, 119708.

Hareland, G., Rampersad, P., 1994. Drag - bit model including wear. In: SPE Latin America/Caribbean Petroleum Engineering Conference. *Society of Petroleum Engineers*, Buenos Aires, Argentina.

Kingma D, Ba J., 2014. Adam: A Method for Stochastic Optimization. *Computer Science*, 2014.DOI:10.48550/arXiv.1412.6980.

Moraveji, M.K., Naderi, M., 2016. Drilling rate of penetration prediction and optimization using response surface methodology and bat algorithm. *Nat J. Gas Sci. Eng*. **31**, 829–841.

Motahhari, H., Hareland, G., James, J., 2010. Improved drilling efficiency technique using integrated PDM and PDC bit parameters. *Can J. Pet. Technol*. **49**, 45–52.

Graves A., Schmidhuber J., 2005. Framewise phoneme classification with bidirectional LSTM networks. Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005., Montreal, QC, Canada, 2005, pp. 2047–2052 **vol. 4**, doi: 10.1109/IJCNN.2005.1556215.

Graves, A., Mohamed, A., Hinton, G.E, 2013. Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 6645–6649.

Hegde, C.M., Wallace, S.P., Gray, K.E., 2015. Use of regression and bootstrapping in drilling inference and prediction. In: SPE Middle East Intelligent Oil and Gas Conference and Exhibition, 15-16 September, Abu Dhabi. UAE, *Society of Petroleum Engineers*.

Hegde, C., Daigle, H., Millwater, H., Gray, K., 2017. Analysis of rate of penetration (ROP) prediction in drilling using physics-based and data-driven models. *Pet J. Sci. Eng*. **159**, 295–306.

Hegde, C., Gray, K., 2018. Evaluation of coupled machine learning models for drilling optimization. *Nat J. Gas Sci. Eng*. **56**, 397–407.

Hegde, C., Daigle, H., Gray, K.E., 2018a. Performance comparison of algorithms for realtime rate-of-penetration optimization in drilling using data-driven models. *SPE J*. **23**, 1706–1722.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-term Memory. *Neural computation*. **9**. 1735–80. 10.1162/neco.1997.9.8.1735.

Pan, Y., Bi, R., Zhou, P., Deng, L., Lee, J., 2019. An effective physics-based deep learning model for enhancing production surveillance and analysis in unconventional reservoirs. In: Paper Presented at the Unconventional Resources Technology Conference, Denver, Colorado, 22-24 July.

Schuster M., Paliwal K. K., 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, vol. **45**, no. 11, pp. 2673–2681, Nov. 1997, doi: 10.1109/78.650093.

Seifabad, M.C., Ehteshami, P., 2013. Estimating the drilling rate in Ahvaz oil field. *Pet J. Explor. Prod. Technol*. **3**, 169–173.

Soares, C., Gray, K., 2019. Real-time predictive capabilities of analytical and machine learning rate of penetration (ROP) models. *Petrol J. Sci. Eng*. **172**, 934–959.

Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., … Cheng, Z., 2020. Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *Petrol J. Sci. Eng*. **186**, 106682.

Sunjaya, B.A., Permai, S.D., Gunawan, A.A.S., 2023. Forecasting of Covid-19 positive cases in Indonesia using long short-term memory (LSTM). In: 7th International Conference on Computer Science and Computational Intelligence, *Pro. Com. Sci*. **216**, 177–185.

Wallace, S.P., Hegde, C.M., Gray, K.E., 2015. A system for real-time drilling performance optimization and automation. In: SPE Middle East Intelligent Oil and Gas Conference and Exhibition, 15-16 September, Abu Dhabi. UAE, *Society of Petroleum Engineers*.

Wang, S., Chen, S., 2019. Application of the long short-term memory networks for well-testing data interpretation in tight reservoirs. *Petrol J. Sci. Eng*. **183**, 106391.

Yavari, H., Sabah, M., Khosravanian, R., Wood, D.A., 2018. Application of an adaptive neuro-fuzzy inference system and mathematical rate of penetration models to predicting drilling rate. *Iran. Oil J. Gas Sci. Technol*. **7**, 73–100.

Zhang, C.K., Song, X.Z., Su, Y.N., 2022. Real-time prediction of rate of penetration by combining attention-based gated recurrent unit network and fully connected neural networks. *Pet J. Sci. Eng*. **213**, 110396.